MARVEL

NATIONAL CENTRE OF COMPETENCE IN RESEARCH

Materials' Revolution: Computational Design and Discovery of Novel Materials

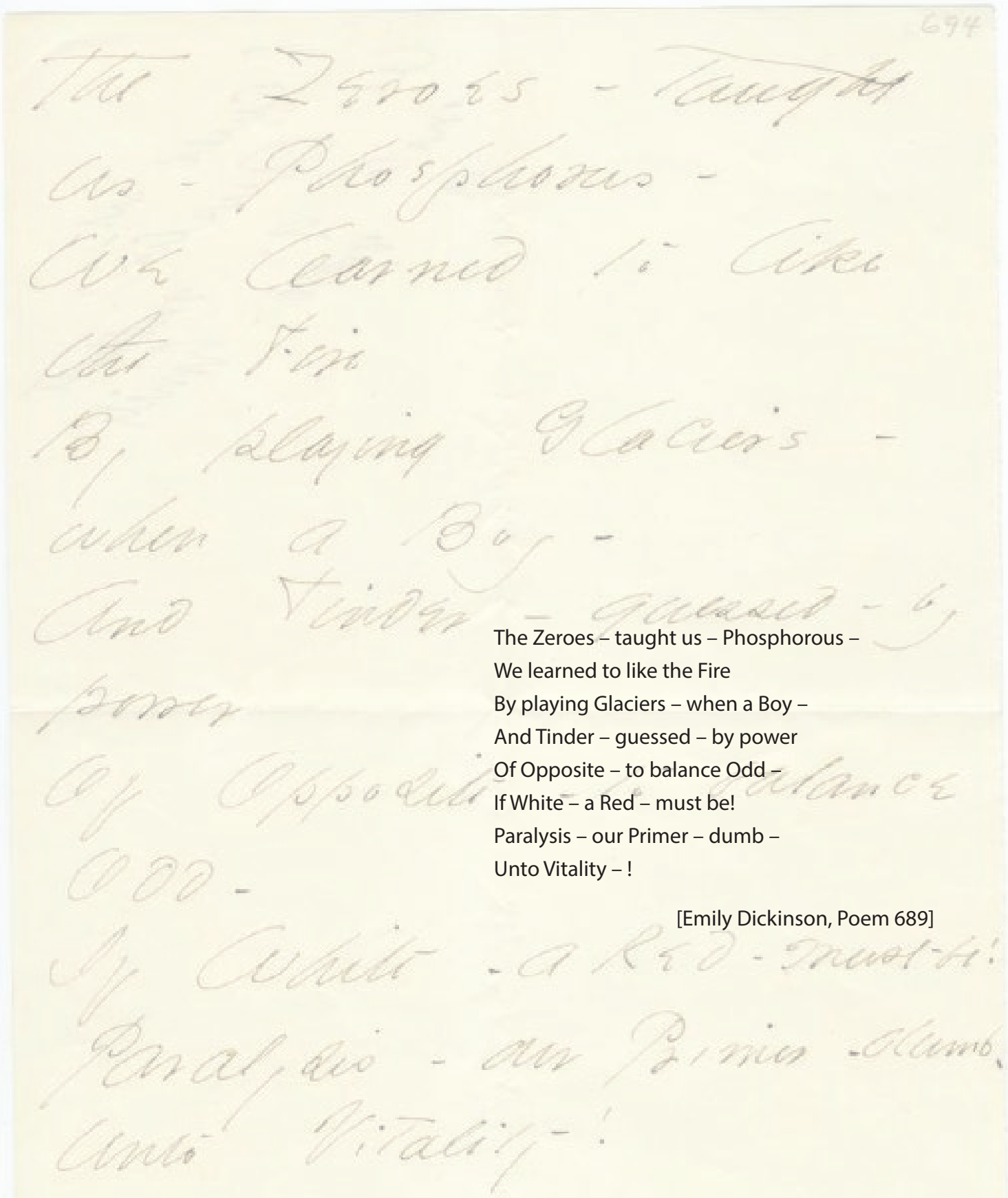
Progress Report

Year 1

May 2014 - January 2015

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

Les Pôles de recherche nationaux (PRN) sont un instrument d'encouragement du Fonds national suisse
Die Nationalen Forschungsschwerpunkte (NFS) sind ein Förderungsinstrument des Schweizerischen Nationalfonds
I Poli di ricerca nazionali (PRN) sono un mezzo di sostegno del Fondo Nazionale Svizzero
The National Centres of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation



The Zeroes – taught us – Phosphorous –
We learned to like the Fire
By playing Glaciers – when a Boy –
And Tinder – guessed – by power
Of Opposite – to balance Odd –
If White – a Red – must be!
Paralysis – our Primer – dumb –
Unto Vitality – !

[Emily Dickinson, Poem 689]

SWISS NATIONAL SCIENCE FOUNDATION

www.snf.ch
Wildhainweg 3, P.O. Box 8232, CH-3001 Berne

Programmes division
National Centres of Competence in Research

NCCR: 1st Progress Report - Cover Sheet

Title of the NCCR	Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL)
NCCR Director Name, first name Institution	 Marzari, Nicola EPFL, Lausanne
Date of submission	28 February 2015

Contents

1 Executive summary

MARVEL is a National Centre of Competence in Research (NCCR) dedicated to the *Computational Design and Discovery of Novel Materials*; MARVEL members belong to the two federal Institutions EPFL (leading house) and ETHZ, the five Universities of Basel, Fribourg, Geneva, Svizzera Italiana, and Zurich, the Swiss supercomputing centre CSCS, IBM Research, the Paul Scherrer Institute (PSI) and the Swiss Federal Laboratories for Materials Science and Technology (EPA). The start date of the project is May 1, 2014, with its funding (18 M CHF in cash from the SNSF, 6.4 M CHF in cash and kind from EPFL, and 9.8 M CHF in kind from the other institutions) allocated in the contract between EPFL and the SNSF of July 15, 2014.

A NCCR is a research instrument of the the Swiss National Science Foundation to promote long-term research networks in areas of strategic importance for Swiss science, economy and society, and MARVEL is one of the 8 that were announced by the Federal Department of Economic Affairs, Education and Research on December 17, 2013, belonging thus to the 4^{th} series of the program. Each NCCR is funded for up to 3 phases of 4 years each, and there are currently 21 NCCRs in operation (5 launched in 2005, 8 launched in 2010, and 8 from 2014).

The **core research topic** of MARVEL is the accelerated design and discovery of materials displaying novel physics or improved properties, via a materials' informatics platform of database-driven, high-throughput quantum simulations. This effort is powered by advanced electronic-structure capabilities that deliver verified-and-validated accuracy, by innovative sampling methods and machine-learning techniques able to explore compositional and configurational space, and by the application of computer science concepts and algorithms to the vast data resulting from materials modeling. The search is targeted to the most urgent societal needs, with a focus on materials for energy harvesting, storage, and conversion, materials for information-and-communication technologies, and organic crystals/pharmaceuticals.

The **goal** of MARVEL is to radically transform and accelerate invention and discovery in science and technology, and especially to transform and accelerate the design and discovery of novel materials in order to achieve improved properties and performance, or to witness the emergence of original physical properties. This goal will be achieved by exploiting the predictive accuracy that quantum mechanical simulations have now reached for realistic, complex systems, the treasure trove of ever-increasing computational power ideally suited to intrinsically parallel problems, and the powerful synergies arising with the computer science of heterogeneous data management, data mining, and machine learning.

By combining accurate, realistic, and predictive simulations of materials with an infrastructure able to run or interrogate thousands to hundreds of thousands of calculations at a time, we will be able to explore phases, structures, and compositions in materials' space, searching for the emergence of novel and fundamental physical properties or for the optimal combination of, e.g., performance, price, and durability, in the quest for better, cheaper, or more environmentally sound materials. We will thus develop a layer of materials' informatics where we can run queries in a database-driven mode, where we provide a natural platform to move the field of quantum simulations from the narrower picture of case-by-case explorations to a broader vista of questions posed for entire classes of materials, where we can rapidly respond to the discovery of novel materials or novel physical properties, and where data mining and machine learning complement our established approaches based on intuition and intelligence. Much needed verification and validation will be achieved by code interoperability and by extensive benchmarking. We will apply this research platform to key strategic areas for societal development and well-being, where improved materials translate directly into improved performance: from the harvesting, storage, and conversion of en-

ergy, to information-and-communication technologies, to synthesis and stability of pharmaceuticals, with deliverables that will focus in the first four years on novel multifunctional materials, materials with novel physical properties, novel materials for energy applications, and organic molecular crystals. Synthesis and characterization will be provided by the two Swiss national materials laboratories Empa and PSI, engaged at the highest level, while the Swiss supercomputing center, CSCS, will provide the high-performance ecosystem for this entire effort. At EPFL, the ACCES Initiative, the Center for Digital Education, and CECAM will provide the educational springboards towards the Swiss community and the rest of the world, while IBM will spearhead leadership for first-principles simulations in industry and materials' technology transfer, in collaboration with our industrial advisories. In the process, we will create powerful synergies between all groups involved, allowing rapid transfer and broad dissemination of new or state-of-the-art capabilities, and sustaining our extensive track-record in education and open-source software dissemination in Switzerland, in Europe, and in the developed and developing world.
Our ultimate, long-term goal is an infrastructure able to simulate and discover or invent materials and devices with an accuracy often comparable or even exceeding that of experiments, and with a speed that mirrors the speed at which computation, data storage, and processing improve, rather than mirroring the constraints of our physical laboratories, manpower, and creativity.

Research

The research efforts of MARVEL are organized around two vertical projects, on novel materials physics and novel materials applications, empowered by three horizontal projects, on advanced quantum simulations, advanced sampling methods, and materials informatics, and supported by two platform projects on the supercomputing and software infrastructure, and the experimental synthesis and characterization. These in the following are the objectives of these seven projects.

Vertical Project 1: Novel Materials Physics. In Vertical Project 1 (VP1) we focus on the design of materials in which the fundamental physics and resulting functionalities are determined by strong interactions between individual electrons, or by new and emerging phenomena, such as topological properties of their band structures. As an additional challenge, the accurate description of these materials requires the development of methods by the Horizontal Project 3 that could also later be used in high-throughput material design. As early accomplishments for the first year we have developed a software package to identify topological properties of materials and have identified β-Bi_4I_4 as a strong Z_2 topological insulator. We have identified a new mechanism for multiferroism close to room temperature in the class of layered rare-earth barium-iron-copper-oxide materials. We have assessed the effect of strain on the metal-insulator transition in $LaVO_3$ and $LaTiO_3$ and proposed a consistent theoretical description of the metal-insulator transition of nickelates. And, finally, we have made an *ab initio* design for a material whose Hamiltonian is that of the single-band Hubbard model, and computed the resulting properties.

Vertical Project 2: Novel Materials Applications. Vertical Project 2 (VP2) is devoted to the application of quantum simulations, enhanced sampling methods and materials informatics to the high-throughput design of materials with novel properties, in the domain where techniques like density functional theory provide already predictive accuracy. Major objectives of VP2 are the development of novel materials for energy and the environment, information-and-communcation technologies, and the characterization of the properties of organic crystals relevant to the electronics or pharmaceutical industry. Some of the recent highlights include photovoltaics, where in a joint experimental and theoretical investigation it has been possible to develop a dye-sensitized solar cell (DSSC) with record efficiency based on novel computationally designed sensitizers, to provide a molecular interpretation of the first *in situ* atomic force microscopy measurements of a device under realistic working conditions and to identify a coverage-dependent phase transition of the absorbed dyes. In a further joint experimental/theoretical project, graphene nanoribbon heterojunctions were produced by combining hydrocarbon precursors with nitrogen substituted analogues and the electronic properties of these and related systems were calculated showing that the electronic levels of the graphene nanoribbon can be controlled via chemical substitution of the monomers. In energy storage, we were able to characterize the conduction properties of zirconium-containing lithium-lanthanum oxide, a promising candidate for use as solid state electrolyte in Li-based batteries, and could assess the effect of dop-

ing on the stabilization of the different polymorphs. For CO_2 sequestration, we have been able to give a detailed computational characterization of the chemical reactions involved in CO_2 capture and release in monoethanolamine solutions.

Horizontal Project 3: Advanced Quantum Simulations. In Horizontal Project 3 (HP3) we develop and implement the novel theories and algorithms that are needed for systems with a complex electronic structure. These are often strongly correlated systems with novel and interesting quantum states that are of interest, for example, in the field of quantum information and computing, or systems where standard density functional theory approaches fail qualitatively or quantitatively. New methods that are being developed include continuous time quantum Monte Carlo (QMC), self-consistent GW+DMFT, and density functionals based on non-local correlations (MP2 and random phase approximation (RPA)) or orbital densities. On-going work involves the extension of continuous time quantum Monte Carlo to the case of long-range Coulomb forces, enabling its use in the simulation of realistic materials, the development of a GW+DMFT self-consistency loop for multi-orbital systems and self-consistent calculations within a small subspace of orbitals corresponding to the lowest-energy bands, and adding features to RPA and MP2 for the description of electrochemical systems, such as molecular dynamics capabilities and unrestricted calculations.

Horizontal Project 4: Advanced Sampling Methods. Horizontal Project 4 (HP4) aims to push the frontiers for structure-prediction, phase space and composition sampling. Earlier work of these groups laid the foundations for future ambitious achievements, from the development of metadynamics and well-tempered metadynamics, basin-hopping methods, and sketch maps. In an extensive structure prediction of silicon allotropes, we have considerably extended the scope of known allotropes and found numerous structures with useful properties for photovoltaic applications. We were able to simulate, with unprecedented detail and in agreement with experiments, the nucleation of organic molecules from solution. In addition, fundamental work is taking place in developing novel variational approaches for the calculation of free energies. Other achievements include the identification of zeolite structures with optimal properties for carbon dioxide or methane capture, the development of machine-learning tools for the automatic recognition of hydrogen bonds, and further developments on alchemical derivatives method to predict the change of properties in alchemical transformations.

Horizontal Project 5: Materials Informatics. The work in the Horizontal Project 5 (HP5) concentrated on two main fronts: the design of a big data system to handle the needs of the MARVEL applications and the design of domain-specific language frameworks that will allow the easy development of high-performant applications in the context of MARVEL. We made significant progress for this first year especially at the design and planning levels. Early algorithmic advances already show very good potential and hold promise for significant impact in the near future.

Platform Project 6: Informatics. The Informatics Platform (PP6) is focused on providing the hardware and software infrastructure for the activities of MARVEL, and on populating and disseminating materials data to the scientific community at large. On the hardware side, these activities involved the procurement and purchase of high-performance computing to support the numerical efforts of all the groups, and the procurement and purchase of a Petabyte-class storage facility for the research activities in data-mining and machine learning, and to act as a public repository of data and workflows. On the software side, key efforts are going into the development of the materials informatics platform AiiDA, that implements our vision for future activities in computational science to support high-throughput calculations, data preservation and storage, scientific workflows, and public dissemination.

Platform Project 7: Experiments. Turning the results of the computational approach into real materials requires the intense collaboration between theory and experimental groups. The Platform Project 7 (PP7) is the linking point and provides funding for experimental verification projects. These research projects will start in the second year of the NCCR and in the first year a proposal system for the distribution of the funding has been developed and launched, following very successful one-day workshop with more than 30 presentations showing the high interest and potential for these collaborations, and forming the basis for new projects.

Other aspects

We started developing the long-term strategies for *knowledge and technology transfer*, based on our extensive track-record of development and dissemination of open-source codes, and the development of the materials informatics platform AiiDA; for *education and training*, building on the expertise on the organization of tutorials, workshops, and schools, in the developed and developing world, and on distant-education learning; on *equal opportunities*, by working closely with the relevant offices at EPFL and elsewhere, and targeting new activities both for young girls (8- to 10-year old, and 11 to 13), and workshops and seminars on the challenges for female PhD students, junior professors, and more established researchers; and on *communication*, developing our website for communication to broader audiences.

On the financial side, we are pleased to report that we secured cash matching funds for all our experimental activities (up to 1'500'000 CHF from PSI, and up to 900'000 CHF from Empa); 820'000 CHF in-kind funds from ETHZ to support the administration, maintenance and running expenses of our supercomputing infrastructure; 416'000 CHF in an industrial project to develop materials with novel properties; and a PhD fellowship from Sciex to fund a student from Vilnius, Lithuania, to work with us for 13 months on the interface between MARVEL data and their Crystallography Open Database.

2 Research

2.1 Scientific development of the NCCR and status of integration

2.1.1 Structure of the NCCR

The creation of the National Centre of Competence in Research (NCCR) MARVEL — Materials' Revolution: Computational Design and Discovery of Novel Materials has been decided by the Federal Department of Economic Affairs, Education and Research on 17 December 2013 together with the creation of seven other centres. This new NCCR, hosted at EPFL, started on 1^{st} May 2014. The final contract was signed on 15 July 2014.

At the scientific level, the NCCR MARVEL is structured around (Fig. 1) two vertical projects of materials design and discovery, with a focus, respectively, on

VP1 novel materials physics,

VP2 novel materials applications,

supported by three horizontal projects on

HP3 advanced quantum simulations,

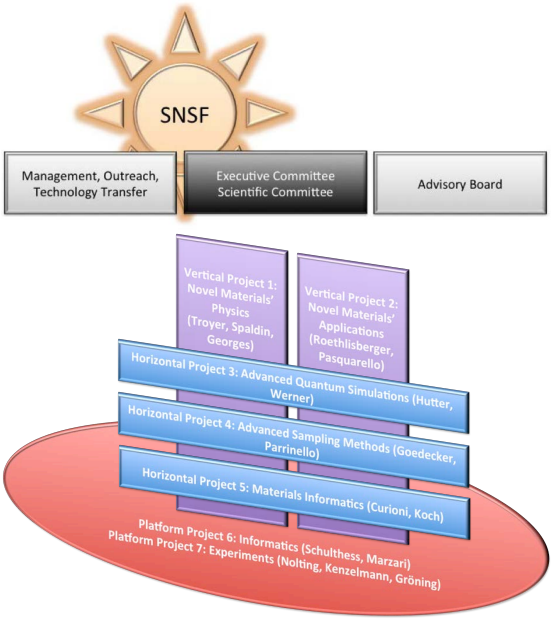


Figure 1: *Administrative and scientific structure of MARVEL*

HP4 advanced sampling methods,

HP5 materials informatics,

interfaced to two platform projects on

PP6 the informatics and supercomputing infrastructure, and

PP7 experiments: synthesis, characterization, and testing.

Its home institution is EPFL and participating scientists are coming from a wide range of Swiss academic and industrial institutions, forming a strong network in the field of computational design and discovery of novel materials. MARVEL is organized around different bodies responsible of its operation, the Executive Committee, the Scientific Committee and the Scientific Advisory Board.
The general organization of the NCCR and the composition of its bodies is presented in the following sections.

Academic and industrial institutions participating in MARVEL

- Federal Institute of Technology, Lausanne (EPFL), Home institution
- Federal Institute of Technology, Zurich (ETHZ)
- University of Geneva (UniGE),
- University of Fribourg (UniFR)
- University of Basel (UniBas)
- University of Zurich (UniZH)

- Università della Svizzera italiana (USI)
- Swiss National Supercomputing Centre (CSCS)
- Paul Scherrer Institute (PSI)
- Materials Science and Technology Research Institute (Empa)
- IBM Research GmbH, Rüschlikon

Executive Committee

The Executive Committee is composed by the director, the deputy director, and a third member from a different institution.

- Nicola Marzari, EPFL, director
- Alfredo Pasquarello, EPFL, deputy director
- Nicola Spaldin, ETHZ

MARVEL Project Leaders

The Project Leaders are in charge of coordinating and overseeing the activities of the seven projects.

- Matthias Troyer, ETHZ, VP1
- Ursula Röthlisberger, EPFL, VP2
- Jürg Hutter, UniZH, HP3
- Stefan Goedecker, UniBas, HP4
- Alessandro Curioni, IBM, HP5
- Thomas Schulthess, CSCS and ETHZ, PP6
- Frithjof Nolting, PSI, PP7

Scientific Committee

The Scientific Committee is composed by the Executive Committee and the seven Project Leaders.

MARVEL Group Leaders

The Group Leaders receive funding from the NCCR MARVEL to pay PhD students or postdoctoral associates, whose research they supervise.

Principal Investigators (PIs)

- Alessandro Curioni, IBM
- Antoine Georges, UniGE
- Stefan Goedecker, UniBas
- Pierangelo Gröning, Empa
- Jürg Hutter, UniZH
- Michel Kenzelmann, PSI
- Christoph Koch, EPFL
- Nicola Marzari, EPFL
- Frithjof Nolting, PSI
- Michele Parrinello, USI and ETHZ
- Alfredo Pasquarello, EPFL
- Ursula Röthlisberger, EPFL
- Thomas Schulthess, CSCS and ETHZ
- Nicola Spaldin, ETHZ
- Matthias Troyer, ETHZ
- Philipp Werner, UniFR

Agility members

- Wanda Andreoni, EPFL
- Michele Ceriotti, EPFL
- Clémence Corminboeuf, EPFL
- Daniele Passerone, Empa
- Berend Smit, EPFL
- Joost VandeVondele, ETHZ
- Anatole von Lilienfeld, UniBas
- Oleg Yazyev, EPFL

Scientific Advisory Board

The Scientific Advisory Board has the formal role of suggesting research lines that need to be strengthened within the NCCR, or that have run their course, and act as an independent reviewer of the activities of the NCCR in addition to the official Review Panel of the SNSF. Its members are:

- Giulia Galli, Liew Family Professor, Institute for Molecular Engineering, University of Chicago IL, USA
- Gian-Luca Bona, Director, Empa, Dübendorf, CH
- Gerbrand Ceder, R. P. Simmons Professor of Materials Science and Engineering, MIT, Cambridge MA, USA
- Karsten Jacobsen, Center for Atomic-scale Materials Design, Technical University of Denmark, Lyngby, DK
- Boris Kozinsky, Robert Bosch Research and Technology Center, Cambridge MA, USA
- Alexander Lichtenstein, Institute of Theoretical Physics, University of Hamburg, DE
- Joël Mesot, Director, Paul Scherrer Institute, Villigen, CH
- Sadasivan Shankar, Materials Design, Intel Corporation, Santa Clara CA, USA
- Risto Nieminen, Chairman, Psi-K network and Dean, Department of Applied Physics, Aalto University School of Science, Helsinki, FI
- Dominic Tildesley, Director, CECAM - Centre Européen de Calcul Atomique et Moléculaire, EPF Lausanne, CH

Management

A team of administrative and scientific collaborators has been formed to carry out management and financial duties of MARVEL and take care of the other responsibilities of the NCCR, such as communication, knowledge and technology transfer, education and equal opportunities.

- Nicola Marzari, director
- Alfredo Pasquarello, deputy director
- Nicola Spaldin, executive committee
- Lidia Favre-Quattropani, scientific manager, communication, equal opportunity
- Elizabeth Gueniat, administrative assistant
- Valérie Le Dreau, financial assistant
- Farnaz Moser, EPFL Equal Opportunities Office, equal opportunities
- Clémence Corminboeuf, equal opportunities
- Michele Ceriotti, education and training
- Pascale Van Landuyt, Alliance, knowledge and technology transfer, from spring 2015
- A program and communication manager is being recruited

Scientific projects

Vertical projects

VP1 — Novel Materials Physics

Project leader:

- Matthias Troyer, ETHZ

Members:

- Matthias Troyer, ETHZ
- Nicola Spaldin, ETHZ
- Antoine Georges, UniGE
- Oleg Yazyev, EPFL

VP2 — Novel Materials Applications

Project leader:

- Ursula Röthlisberger, EPFL

Members:

- Ursula Röthlisberger, EPFL
- Alfredo Pasquarello, EPFL
- Jürg Hutter, UniZH
- Nicola Marzari, EPFL
- Alessandro Curioni, IBM
- Stefan Goedecker, UniBas
- Wanda Andreoni, EPFL
- Clémence Corminboeuf, EPFL
- Daniele Passerone, Empa

Horizontal projects

HP3 — Advanced Quantum Simulations

Project leader:

- Jürg Hutter, UniZH

Members:

- Jürg Hutter, UniZH
- Matthias Troyer, ETHZ
- Philipp Werner, UniFR
- Joost VandeVondele, ETHZ

HP4 — Novel Materials Physics

Project leader:

- Stefan Goedecker, UniBas

Members:

- Stefan Goedecker, UniBas
- Michele Parrinello, USI and ETHZ
- Berend Smit, EPFL
- Michele Ceriotti, EPFL
- Anatole von Lilienfeld, UniBas
- Nicola Marzari, EPFL
- Ursula Röthlisberger, EPFL

HP5 — Materials Informatics

Project leader:

- Alessandro Curioni, IBM

Members:

- Alessandro Curioni, IBM
- Christoph Koch, EPFL

Platform projects

PP6 — Informatics

Project leader:

- Thomas Schulthess, CSCS and ETHZ

Members:

- Thomas Schulthess, CSCS and ETHZ
- Nicola Marzari, EPFL

PP7 — Experiments

Project leader:

- Frithjof Nolting, PSI

Members:

- Frithjof Nolting, PSI
- Michel Kenzelmann, PSI
- Pierangelo Gröning, Empa

2.1.2 Scientific development

This NCCR was initiated with a 31-page preproposal submitted by Nicola Marzari and the other 12 computational principal investigators to the SNSF in January 2012, in response to the call for the 4^{th} series of NCCRs. Funding was secured by this same group (180'000 CHF) in October 2012 through HP2C, the Swiss Platform for High-Performance and High-Productivity Computing (now succeeded by the Platform for Advanced Scientific Computing (PASC)), for a number of preparatory actions that focused on code development between different PIs, but included also a community workshop to brainstorm about needs and opportunities, and a junior retreat of all the computational students and postdocs. In November 2013, the NCCR preproposal was evaluated and rated A, and invited for a full proposal that was submitted in February 2013, following a number of brainstorming meetings between all the PIs involved. A team of five was interviewed in Bern in June 2013, and a final presentation involving the EPFL Direction was held in October 2013. This extensive refinement of the ideas contained in the core of the preproposal, the very collaborative spirit of all the PIs involved, and the geographical proximity of all the institutions involved allowed for a careful and balanced development of all the scientific themes that are now contained in the full proposal and have started being developed in this initial few months.

2.1.3 Status of integration

It is been a policy of the NCCR to stress the importance of hiring the right people for the right job, and to employ them fully (i.e. 100%) in a NCCR project. This is done both to insure a sense of identity, dedication, and commitment, but also to make sure that PIs are not compelled to spend all their funds in year 1, given the obvious challenge of having to hire on the domestic and international market a considerable number of individuals without making any compromise on quality.
The organic subdivision of the group leaders in the seven projects of the NCCR already provide a clear vision and match between skills and opportunities. In particular, one of our metrics of success is provided by the novel synergies that emerge out of this NCCR, and the collaborative projects that are empowered by sharing of state-of-the-art skills — the goal, at the end, is to obtain a result that is far greater than the sum of the parts.
Many new collaborations have been ignited, that are exclusively attributable to this effort — to mention just a few, there are new and powerful collaborations that include the work on

- topological insulators between the groups of Troyer and Yazyev, that relies also on the calculation of Chern invariants with the community code WANNIER90,
- structure prediction for novel metal-organic perovskites between the groups of Röthlisberger and Goedecker,
- graph databases between the groups of Koch, Curioni, and Marzari,
- novel software to drive agnostically quantum engines and perform structural optimizations via basin-hopping, molecular dynamics with colored-noise thermostats, and path-integral molecular dynamics, by Ceriotti and Goedecker.

These are just a few early examples of collaborations that are organically starting at all levels, and will constitute one of the main metrics of evaluation and success for the Centre.

2.2 Objectives of the research units, first achievements

Vertical Project 1 — VP1 — Novel Materials Physics

Project leader: Matthias Troyer (ETHZ)

Participating members: Matthias Troyer (ETHZ), Nicola Spaldin (ETHZ), Antoine Georges (UniGE), Oleg Yazyev (EPFL)

Summary and highlights: In Vertical Project 1 (VP1) we focus on the design of materials in which the fundamental physics and resulting functionalities are determined by strong interactions between individual electrons, or by new and emerging phenomena, such as topological properties of their band structures. The challenge here is the development of methods that can later be used in high-throughput material design. In the first year we have developed a software package to identify topological properties of materials and have identified β-Bi_4I_4 as a strong Z_2 topological insulator. We have identified a new mechanism for multiferroism close to room temperature in the class of layered rare earth barium iron copper oxide materials. We have assessed the effect of strain on the metal-insulator transition in $LaVO_3$ and $LaTiO_3$ and proposed a consistent theoretical description of the metal-insulator transition of nickelates. And finally, we have made an *ab initio* design of a material whose Hamiltonian is that of the single-band Hubbard model and computed the resulting properties.

General view of the project

Major research questions

Strongly correlated materials, and especially transition metal (TM) oxides exhibit a wide range of fascinating and potentially technologically important phenomena, including exotic superconductivity, metal-insulator transitions, unusual magnetic ordering and coupled charge/spin/orbital interactions, as a result of their coupled spin, orbital, electronic and lattice orders. These materials display a number of competing phases, with small energy differences between them. An outstanding challenge is to *control these materials* in order to "guide" them towards a specific phase with a desired functionality, determined by their chemical composition and crystal structure. How to achieve control of these materials by acting on these "knobs" is the big research question which underlies the present project.

The discovery of topological insulators, materials realizing topologically non-trivial band structures, opened a new direction of research in condensed matter physics with high potential for technological applications, not the least for quantum devices. The first "practical" Z_2 topological insulators, bismuth and antimony chalcogenides, were discovered approximately five years ago in the well-known class of thermoelectric materials. Topological materials seem to be common among heavy element compounds featuring strong spin-orbit interactions. The search for new members of the family of Z_2 topological insulators and materials belonging to other topological classes, such as topological crystalline insulators, Chern insulators, Dirac and Weyl semimetals, is another priority direction in this emerging field of research.

Short- and long-term goals

The field of correlated and topological materials is very wide and we thus focus on a few highly promising topics, which we will quickly summarize here and then develop in more detail in the individual subproject reports.

A particularly promising class of materials on which we focus currently are the rare earth nickelates, which display a phase transition between a high-temperature metallic phase and low-temperature insulating phases. This transition, which can be close to room temperature, is highly sensitive to chemical composition, as well as structural constraints and strain (tunable by changing the substrate or by heterostructuring). This could find applications to switches or to the recently proposed piezoelec-

tric transistors. Deciphering the basic mechanism of the metal-insulator transition (MIT) in nickelates and clarifying to what extent orbital-engineering and the realization of a single-band material are possible for these materials is one of the short and medium term goals of this project.

Oxide heterostructures often exhibit emerging properties that are different from those of the constituent materials. One example is the observation of metallic behavior in heterostructures containing two otherwise insulating oxides. The complex nature of the insulator-to-metal transitions in such heterostructures mainly stems from the simultaneous appearance of several new features compared to the bulk forms of the constituent materials. As a first step we aim to understand the role of bare epitaxial strain in the insulator-to-metal transitions for some prototypical Mott insulators, will then move to an explicit description of interface effects which take place at the boundary between two different oxides, with a long-term goal to achieve a concise and comprehensive picture of oxide heterostructures composed of Mott insulators.

Multiferroic materials, that exhibit simultaneous ferroelectricity and ferromagnetism, are of fundamental interest because of the chemical contraindication between these phenomena, and of potential technological importance because of the cross-coupling between them. Currently, practical materials that are multiferroic at room temperature do not exist. In the first year of the project we have focused on a class of materials — the RBaFeCuO$_5$ (R^{3+} = Y^{3+} and 4f rare earth ions) series of bilayered perovskites — that is promising for achieving room-temperature multiferroism and studied the origin of the unusual magnetic order that is believed to be responsible for the polar state.

While most subprojects of this vertical project are concerned with understanding new phenomena in a quantitative manner in realistic material simulations instead of effective models, here also follow the opposite path. Inspired by ultracold atomic gases, we propose to engineer materials with precisely (and only) the bands and interactions contained in model Hamiltonians (such as the Hubbard model). Learning to design "simple" materials will be an important first step and will give us model materials that exhibit physical phenomena known from effective models in simple materials, not encumbered by many complications of more complex materials.

For topologically non-trivial materials we plan to set up a fully automated procedure for calculating topological invariants of band structures starting from first-principles calculations and use it for generic scanning of existing materials databases for identifying materials candidates for quantum spin Hall, quantum anomalous Hall effects, as well as 3D topological insulators and Weyl and Dirac semimetals. In parallel, we search for optimal candidate materials for hybrid systems, such as heterostructures and nanowires that are predicted to realize Majorana fermions.

First achievements and next steps

On the topic of nickelates we proposed a consistent theoretical description of the metal-insulator transition [1] using only the strongly hybridized O-Ni e_g states which form the active degrees of freedom close to the Fermi level. Our next goals are to calculate optical spectra of nickelates based on this theoretical description above, allowing for direct comparison to ongoing experiments in van der Marel's group (UniGE) and to systematically investigate the key energy scales involved in this description over the whole RNiO$_3$ series, to work towards "orbital engineering".

In the strained films, we have studied the role of epitaxial strain in the insulator-to-metal transitions for two prototypical Mott insulators with different formal bulk electron configurations, LaTiO$_3$ (formal t_{2g} occupation d^1) and LaVO$_3$ (d^2). We now have a basis to develop a simple and, possibly, general model explaining the strain-induced changes of the metallic properties through a limited number of parameters (e.g. bandwidth, on-site energies, filling). The next step will be the explicit description of interface effects which take place at the boundary between two different oxides, such as in a thin film grown epitaxially on an insulating substrate (e.g. SrTiO$_3$). Here we have begun discussions with researchers at the Paul Scherrer Institute, spawned by the first theory-experiment MARVEL workshop held at the PSI in October.

In YBaFeCuO$_5$ we showed that the Fe^{3+} and Cu^{2+} ions are disordered, but that the most energetically favorable configurations always have ferromagnetic coupling between two planes in the same bilayer and antiferromagnetic between bilayers. Any additional disorder can introduce strong antiferromagnetic coupling in two planes of the same bilayer and locally induce frustration. We proposed a model in which this frustration can lead to the experimentally observed spiral; the next step is to simulate the model to determine whether it

is indeed consistent with the experimental behavior.
The first phase of automated search for topological materials is now completed. Software has been developed that determines Chern numbers of band structures along with Z_2 topological invariant of time-reversal symmetric systems. This code, which is planned to be integrated with WANNIER90, allows one to search for topological insulators with time-reversal symmetry as well as for Weyl and Dirac semimetals. During our first attempts to search for novel topological materials, we identified a promising family of candidate materials: the quasi-one-dimensional bismuth halides and especially the β-phase of bismuth iodide Bi_4I_4. It exhibits a number of novel properties which makes it very different from layered bismuth and antimony chalcogenides and is less prone to structural defects [2]. This work has been a new collaboration enabled by MARVEL, leveraging the skills of different groups within the NCCR.
Recent experimental observation of what is believed to be a Majorana fermion in a hybrid semiconductor/superconductor structure poses a clear demand for theoretical/computational guidance of such experiments. We are developing realistic effective models of InSb and other zincblende semiconducting wires and heterostructures, to guide the design of new quantum devices.
The "standard model" of condensed matter physics is widely regarded to be the Hubbard model which augments independent-electron band theory through a single parameter to account for electron-electron correlations. However, beyond one dimension the problem is intractable and so much current research aims at finding appropriate approximations for obtaining the Hubbard model phase diagram. Here we take a new approach — we used *ab initio* methods to design a material whose Hamiltonian is exactly that of the Hubbard model so that its solution can be found by measuring the material properties. After identifying a crystal class and several appropriate chemistries, we used density functional theory to screen for the desired electronic band structure, and dynamical mean field theory to study the Mott transition. Following this, we took the most promising candidate and addressed its structural stability, possibilities for doping and exotic superconductivity, and its similarity with the cuprate high-temperature superconductors. We plan to continue the search for such "model materials" with a focused workshop involving all relevant teams of MARVEL in the first half of 2015, to identify promising material classes and computational approaches.

Research contributions to the overall goals of the NCCR and to the existing literature

This project works towards a quantitative understanding of the electronic structure of strongly correlated materials with exotic properties and closely competing phases, as well as the identification and prediction of materials with non-trivial topological electronic structure. These are important steps towards both a quantitative understanding of the properties of these materials and future high-throughput screening of existing materials, and the design of new materials.

Collaborative components

All subprojects in this project are collaborative, either sharing and developing common computational tools as in the projects on strongly correlated materials (groups of Georges, Spaldin and Werner), or directly collaborating on the detection of topological properties and identification of new topological materials (groups of Marzari, Troyer and Yazyev). Collaborations focus on regular meetings and discussions about research directions and algorithmic developments and go far beyond joint publications. Theory postdoc Andrea Scaramucci (Spaldin group) is now located full-time at the Paul Scherrer Institute with the explicit goal of forging and consolidating theory/experiment collaborations.

1 Topological materials (Matthias Troyer — ETHZ)

1.1 Research summary

- Topological invariants computed *ab initio*.
- Realistic modeling of topological superconductivity in nanowires.
- Wannier functions for topological materials.

1.2 Research question

The aim of our project is the ability to predict and design topological properties of materials from first-principles calculations and symmetry analysis. Recent theoretical and experimental advances in the field of topological properties of band structures of metals, insu-

lators and superconductors reveal possibilities for major technological advances, such as realization of experimentally realizable platforms for quantum computation. The ability to identify distinct topologies of band structures of known materials, as well as find routes to design nanostructures with desirable topological properties becomes of crucial importance for further development of this field. We establish ways to identify topological properties of materials from *ab initio* calculations, significantly simplifying search for potential candidate materials. We also develop new techniques for understanding properties of topological materials and identify best candidate materials for realization of topological superconductivity in nanostructures.

1.3 Scientific goals

We plan to set up fully automated procedure for calculating topological invariants of band structures starting from first-principles calculations and use it for generic scanning of existing materials databases for identifying materials candidates for quantum spin Hall, quantum anomalous Hall effects, as well as 3D topological insulators and Weyl and Dirac semimetals. In parallel, we search for optimal candidate materials for hybrid systems that are predicted to realize Majorana fermions. These are semiconducting wires proximity coupled to an *s*-wave superconductor.

a) *Short-term*

- Implementation of automated search for topological insulators and Dirac and Weyl semimetals.
- Construction of maximally localized Wannier functions for a real 3D topological insulator Bi_2Te_3 and using this representation to study surface states and spin/charge distribution in topological materials.
- Developing automated procedures for finding crystalline symmetry protected topological order in materials. Identifying crystalline topological insulators is often a challenge even in simple models.
- Exploring optimal material candidates for realization of Majorana fermions in hybrid superconductor/semiconductor devices.

b) *Long-term*

- Classifying known narrow gap semiconductors and semimetals according to the topology of the material band structure.
- Setting up novel approach to model heterostructures of zincblende materials using a mixture of first-principles calculations with tight-binding modeling and symmetry analysis.
- Optimizing heterostructures for topology-related applications. For example, finding optimal experimental setup to get best possible edge channels in InAs/GaSb quantum spin Hall system.
- Understanding of observable consequences of Dirac and Weyl nodes present at the Fermi level of the semimetal. In particular, explanation of peculiar magnetoresistance data.

1.4 Plans and results: May 2014 – Jan 2015

The first phase of automated search for topological materials is now completed. Software has been developed that determines Chern numbers of band structures along with Z_2 topological invariant of time-reversal symmetric systems. This allows one to search for topological insulators with time-reversal symmetry as well as for Weyl and Dirac semimetals.

Topological invariants, when non-zero, represent obstructions for a smooth choice of Bloch states throughout the Brillouin zone. The question is whether this has observable consequences in real space. In other words, how different is spin/charge distribution of topological insulators compared to the ordinary ones. In order to answer this question we developed a procedure to construct maximally localized Wannier functions for topological insulators in 2D, paving the way to make such a construction viable in 3D real materials.

Recent experimental observation of what is believed to be a Majorana fermion in a hybrid semiconductor/superconductor structure posed a clear demand for theoretical/computational guidance of such experiments. We made first such attempt by modeling realistic wires of InSb, subject to external electric and magnetic fields in proximity to an *s*-wave superconductor. Hybrid functionals were used to extract a high-precision multiband tight-binding Hamiltonian for InSb nanowire. Superconductivity was included on a mean field level, but with a full account of symmetries. We are able to identify optimal experimental setups to realize Majorana fermions in zincblende semiconducting wires.

1.5 Planned research: Feb 2015 – Jan 2016

The work on topological invariants of *ab initio* obtained band structures will be continued. Open question remains about how to automate the search for crystalline topological insulators, where non-trivial metallic surface states are protected by point group symmetries. Some promising approaches are already developed on the tight-binding models, but implementing and testing them on real materials calculation is still to be done.

Several materials were identified by us as being potentially more promising for Majorana fermion realization than currently used InSb and InAs wires. We will explore the possibility of using these materials in real experiment. If successful, this can lead to a significant increase of spin-orbit coupling that is the major ingredient in cooking Majorana fermions.

We will continue our so far successful path of constructing exponentially localized Wannier functions for occupied states of topological insulators. This is a promising tool for real space study of surfaces and interfaces of topological materials, as well as a valuable contribution towards materials with large magnetoelectric response.

We plan to continue modeling zincblende materials using hybrid functional calculation as a starting point. Being computationally hard these functionals are used to extract high-precision multiband tight-binding models. We will make an attempt to build up on this method to obtain tight-binding models of interfaces and heterostructures of zincblende materials, which would be of great importance for experimental applications enabling one to find optimal parameters of the system prior to growing it.

1.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Search for topological materials is done in collaboration with the group of Oleg Yazyev at EPFL. The development of codes is done in collaboration with the groups of Nicola Marzari and Oleg Yazyev.

b) *Outside MARVEL* Majorana fermion related work is done in collaboration with R. Lutchyn, C. Nayak and B. Bauer at Microsoft Station Q, Santa Barbara, CA, USA.
Topological semimetal related work is done in collaboration with B. A. Bernevig and N. P. Ong at Princeton University, Princeton, NJ, USA.

1.7 Personnel

NCCR personnel

- Alexei Soluyanov, postdoc, 100%, from September 2014.
- Georg Winkler, PhD student, 100%, from September 2014.

Matching personnel

- Lei Wang, postdoc, 50%, from May 2014.
- Mauro Iazzi, postdoc, 100%, from May 2014.
- Hiroshi Shinaoka, postdoc, 50%, from May 2014.
- Guglielmo Mazzola, postdoc, 50%, from January 2015.

2 High-throughput search of novel topological materials (Oleg Yazyev — EPFL)

2.1 Research summary

Our project aims at the discovery of novel materials realizing topologically non-trivial electronic phases by means of high-throughput database screening. One candidate material, the quasi-one-dimensional bismuth iodide β-Bi_4I_4, is theoretically predicted and experimentally confirmed as the strong Z_2 topological insulator with a number of novel properties.

2.2 Research question

The discovery of topological insulators, materials realizing topologically non-trivial band structures, opened a new direction of research in condensed matter physics with high potential for technological applications. The first "practical" Z_2 topological insulators, bismuth and antimony chalcogenides, were discovered approximately five years ago in the well-known class of thermoelectric materials. It is now realized that topological materials are rather common among the heavy element compounds featuring strong spin-orbit interactions. The search for new members of the family of Z_2 topological insulators and materials belonging to other topological classes, such as topological crystalline insulators, Chern insulators, Dirac and Weyl semimetals, is the priority direction in this emerging field of re-

search. The present project aims at discovering novel topological materials by means of high-throughput screening of the databases of known compounds followed by in-depth investigations of candidate materials.

2.3 Scientific goals

a) *Short-term*

- Development of a computational methodology for high-throughput screening of databases of known compounds in order to identify candidate topological materials

b) *Long-term*

- Discovery of topological materials realizing novel topological phases or displaying novel physical properties via in-depth theoretical studies and experiments performed on promising candidate materials

2.4 Plans and results: May 2014 – Jan 2015

a) *Methodological developments* Our methodology for performing the high-throughput search of novel topological materials relies on public databases of known materials, such as the International Crystal Structure Database (ICSD). The critical step of the search process consists in calculating the respective topological invariants from first principles for a set of materials sampled from the database according to a number of predefined criteria. Evaluation of Z_2 topological indices for centrosymmetric materials is a trivial task as it requires only the knowledge about wave function parities for the occupied states at the time-reversal invariant momentum points. In case of materials lacking inversion symmetry, or when other topological invariant are relevant, a different approach is required. In order to tackle this problem, we established a collaboration with Dr. Alexey Soluyanov (group of Matthias Troyer, ETHZ). The approach originally proposed by Dr. Soluyanov and implemented in the Z2PACK code evaluates the Chern numbers and Z_2 invariants by tracking the position of hybrid Wannier centers in high-symmetry planes of the Brillouin zone. The Z2PACK package is now being tested on a class of known topological materials in view of a public release of the code [3]. These tests show that the Z2PACK method can identify successfully Z_2 topological insulators, Dirac semimetals and Weyl semimetals. We are currently investigating the possibility of extending its capability to the search of crystalline topological insulators.

In addition, we are currently studying various ways of defining the search protocol for screening materials databases. Initial selection of materials can be performed using simple criteria such as chemical composition, size of the band gap, strength of the spin-orbit coupling and crystallographic space group. The topological invariants of selected materials are computed using Quantum-ESPRESSO and Z2PACK. For materials presenting promising results, a further refinement of the band gap and band ordering is performed using the GW approximation methodology or hybrid functional calculations. The final candidate materials will be proposed to experimental researchers for eventual realization.

b) *Quasi-one-dimensional bismuth halides* During our first attempts to perform search of novel topological materials, we identified a promising family of candidate materials — the quasi-one-dimensional bismuth halides. One of these materials, the β-phase of bismuth iodide Bi_4I_4, was predicted to be a (0;001) weak topological insulator according to our density functional theory (DFT) calculations. The crystal structure of β-Bi_4I_4 is composed of one-dimensional molecular fragments, narrow nanoribbons of bismuth bilayer terminated by iodine atoms, held together by weaker non-covalent interactions (Fig. 1a, b). We further proceeded towards establishing experimental collaborations that would allow us to confirm this remarkable prediction. The ARPES measurements on the samples of β-Bi_4I_4 synthesized at TU Dresden have been performed at the Advanced Light Source synchrotron facility of the Lawrence Berkeley National Laboratory. These early measurements confirmed the presence of Dirac fermion surface states while revealing intriguing discrepancies with our theoretical predictions. Indeed, while DFT predicts the band inversion occurring at the M and Y time-reversal invariant momentum points in the Brillouin zone, more accurate calculations performed at the level of GW approximation show that the band inversion takes place only at the Y point (Fig. 1c). This results in the (1;110) strong topological insulator phase, for which we find a very good agreement between the simulated (Fig. 1d) and measured (Fig. 1e, f) ARPES spectra of the (001) surface.
The discovered material exhibits a number of novel properties which make it very different from layered bismuth and antimony chalcogenides. Compared to these "second generation" topological insulators, β-Bi_4I_4 is less prone to structural defects as it is chemically and structurally well-defined. Even though

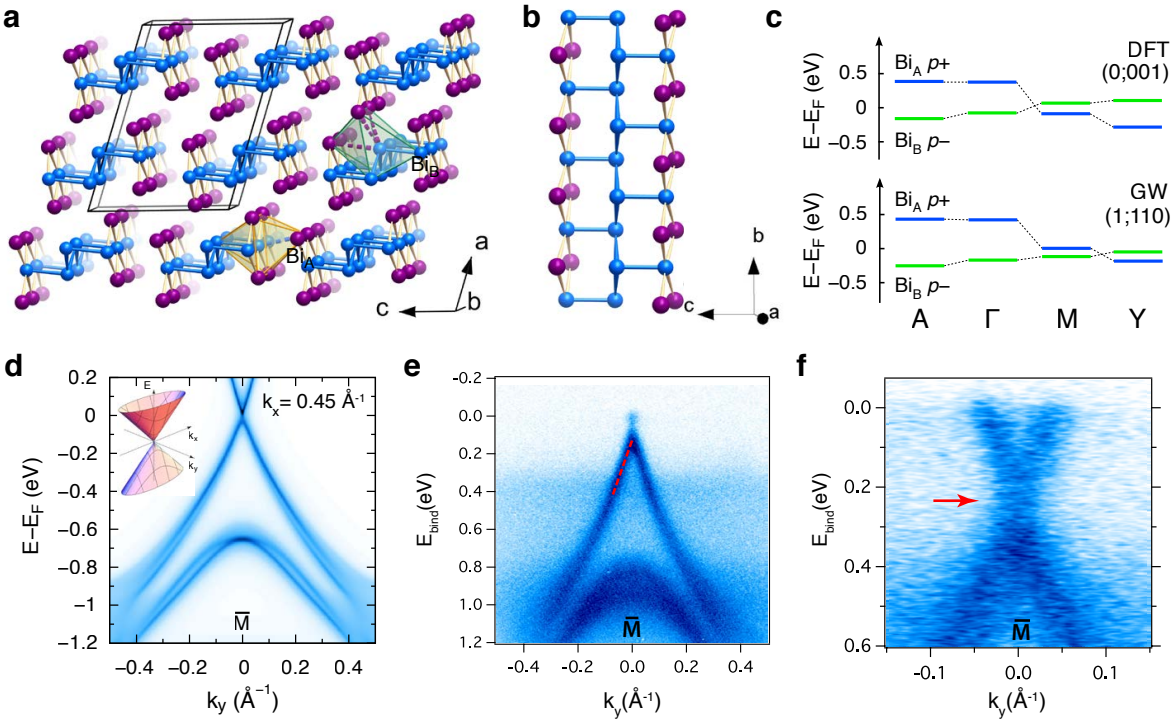


Figure 1: *Atomic and electronic structure of quasi-one-dimensional topological insulator β-Bi_4I_4. (a) Crystal structure of β-Bi_4I_4 viewed along the chain direction (lattice vector b). Bismuth and iodine atoms are shown in blue and purple, respectively. Coordination polyhedra of two non-equivalent Bi atoms, Bi_A and Bi_B, and the conventional unit cell are indicated. (b) Atomic structure of an individual chain-like building block of β-Bi_4I_4. (c) Schematic diagram of band inversion at the time-reversal invariant momentum points A, Γ, M and Y obtained from DFT and GW approximation calculations, leading to the weak (0;001) and strong (1;110) topological phases, respectively. (d) Momentum-resolved local density of states at the (001) surface along k_y ($k_x = 0.45$ Å^{-1}) calculated for the (1;110) phase of β-Bi_4I_4. The inset shows a schematic drawing of the highly anisotropic surface-state band dispersion. (e) Band dispersion around the $\bar{M}$ point along the y direction observed in ARPES experiments. The extracted Fermi velocity (red line) is $v_{Fy} = 0.6 \times 10^6$ m/s. (f) Close-up of band dispersion after potassium evaporation reveals a clear crossing. The Dirac point energy is marked by a red arrow.*

the bulk band gap is small, the absence of defects results in a low concentration of intrinsic charge carriers, and hence, in the proximity of the Fermi level to the Dirac point energy of the topological surface-state band. Importantly, the quasi-1D structure of β-Bi_4I_4 results in highly anisotropic surface-state Dirac fermions and suggests the possibility of combining topological order with other types of ordering characteristic to one-dimensional systems, such as the charge density wave. Finally, being a (1;110) strong TI, this material is placed in proximity to two distinct topological phase transitions, leading to a (0;001) weak TI and a (0;000) trivial insulator, which presents further opportunities for exploring topological physics. There is a fair chance that our finding opens the new class of "third generation" of topological insulators. The manuscript reporting both the theoretical prediction and experimental confirmation is currently under review [2].

c) $(Sb_2)_m$–$(Sb_2Te_3)_n$ *natural superlattices* Another joint experiment-theory project on novel topological insulators has been performed within the established collaboration with the group of Marco Grioni (EPFL). Antimony, Sb, and antimony telluride, Sb_2Te_3, are Z_2 topological semimetal and Z_2 topological insulator, respectively. These two materials form a homologous series of layered superlattices $(Sb_2)_m$–$(Sb_2Te_3)_n$ composed by stacking Sb_2 bilayers and Sb_2Te_3 quintuple layers in ordered arrangements. This family of materials is ideal for investigating the evolution of topological properties upon altering the stacking sequence of the constituent building blocks. Our DFT calculations explain the results of conventional and pump-probe ARPES measurements performed on the two members of this family — Sb_2Te_3 ($n = 1$; $m = 0$) and Sb_2Te ($n = 1$; $m = 2$) — and further complement the measurements with predictions for other materials in this series. We show that all materials

in this family are topological with alternating Z_2 indices (1;000) and (1;111) as the parameter m increases. Our calculations reveal systematic changes in the Fermi velocity, the degree of electron-hole asymmetry and the hexagonal warping parameter of the topological surface states, along the series. The manuscript reporting these results is currently in preparation [4].

2.5 Planned research: Feb 2015 – Jan 2016

a) *Methodological developments*

- Extension of the high-throughput search methodology to other topological phases, such as the topological crystalline insulators in which topological protection is associated with different symmetry applications.
- Realization of database-wide search of candidate topological materials and classification of the results of this search.

b) *In-depth studies of candidate materials*

- Further investigation of the discovered quasi-one-dimensional bismuth halide topological insulators, including the studies of bismuth bromide Bi_4Br_4 and mixed halides, pressure- and strain-induced topological phase transition and the development of tight-binding model for performing transport simulations.
- Eventual in-depth investigation of other promising candidate materials, including collaborative efforts involving experimental groups within MARVEL.

2.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* The described research program implies close collaboration with Dr. Alexey Soluyanov (group of Matthias Troyer, ETHZ) on developing methodologies for calculating topological invariants and efficient high-throughput search of novel topological materials.

b) *Outside MARVEL* Our project involves a growing number of groups with expertise in experimental techniques that can be used for synthesis, characterization and investigation of candidate topological materials. To date, the following connections have been established to confirm our prediction of the novel topological phase in the quasi-one-dimensional bismuth iodide β-Bi_4I_4:

- materials synthesis with emphasis on bismuth compounds: Dr. Anna Isaeva (group of Prof. Michael Ruck, TU Dresden);
- angle-resolved photoemission spectroscopy (ARPES): Dr. Jens C. Johannsen, Prof. Marco Grioni (EPFL);
- ARPES investigations at synchrotron facilities: Dr. Luca Moreschini (groups of Dr. Eli Rotenberg and Dr. Aaron Bostwick, Lawrence Berkeley National Laboratory);
- transport measurements: Andrea Pisoni, Prof. László Forró (EPFL);
- transmission electron microscopy (TEM): Dr. Wouter Van den Broek (group of Prof. Ute Kaiser, Ulm University).

2.7 Personnel

NCCR personnel

- Gabriel Autès, postdoc, 100%, from July 2014.

Matching personnel

- Hyungjun Lee, postdoc, 50%, from May 2014.
- Gabriel Autès, postdoc, 100%, from May to June 2014.

3 Theory/experiment collaborations on multiferroics (Nicola Spaldin — ETHZ, Michel Kenzelmann — PSI)

3.1 Research summary

Understand the origin of unusual magnetic ordering in multiferroic materials.

3.2 Research question

The RBaFeCuO$_5$ (R^{3+} = Y^{3+} and $4f$ rare earth ions) series of compounds with bilayered perovskite structure provides an interesting set of materials where a (multifunctional) multiferroic magnetic spiral structure arises at high temperatures [7]. We recently found that Fe^{3+} and Cu^{2+} in the Y^{3+} member of this series are disordered [5]. However, the most energetically favorable configurations are such that the sign of the relevant magnetic couplings is homogeneous (e.g. ferromagnetic coupling be-

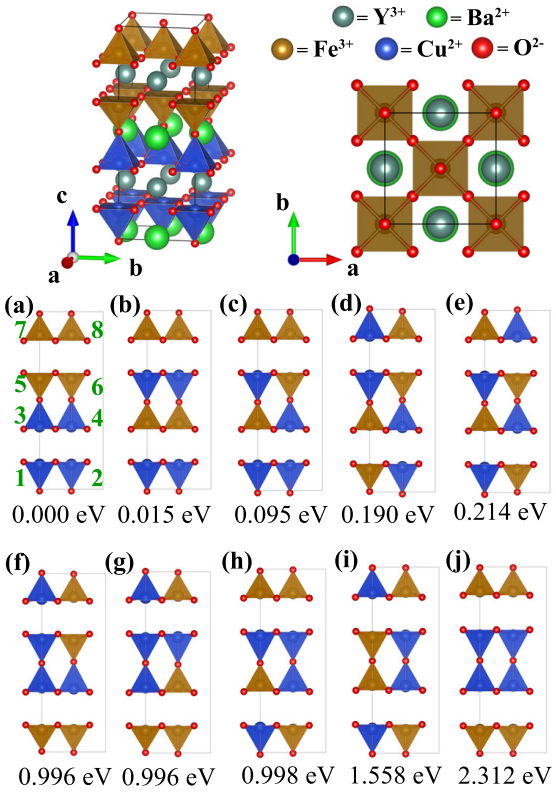


Figure 2: *Three-dimensional structure (left upper panel) and its projection on the ab-plane (right upper panel) of* $YBaCuFeO_5$*. Panels from (a) to (j) show the projection on the ac-plane of the considered structures with different orderings of* Fe^{2+} *ions (golden square pyramids) and* Cu^{3+} *(blue square pyramids) and their energies relative to that of structure (a). For clarity* Ba^{2+} *and* Y^{3+} *ions are not shown.*

tween two planes in the same bilayer and antiferromagnetic between bilayers). Nonetheless, a small concentration of strong magnetic bonds with opposite sign can be present (e.g. strong antiferromagnetic coupling between two Fe^{3+} in two planes of the same bilayer) and locally induce frustration (Fig. 2).

3.3 Scientific goals

The primary aim of this project is to check if the results of *ab initio* calculations together with the model of spiral induced by disorder can explain the experimental measurements on the compounds with R^{3+} different from Y^{3+} currently in progress at the Paul Scherrer Institute (PSI).

a) *Short-term* Our ongoing work suggests that, under certain conditions, such a type of disorder can induce a spin-spiral ordering. This implies that the spiral states might appear at different temperatures depending on the preparation methods, which is in agreement with the relatively broad range of T_c's present in the literature. The study of the theoretical relationship between disorder and spiral magnetism in a model system will be completed and written up in the next reporting period.

b) *Medium-term* Using *ab initio* calculations, we will next investigate the role that R ions play in the energetics of Fe^{3+}/Cu^{2+} disorder and on the size of exchange coupling constant.

c) *Long-term* Our long term goals are to identify ways, such as preparation procedure, pressure or strain, in which the transition temperature can be increased. We will also perform a survey of other material types where we expect to observe the behavior.

3.4 Plans and results: May 2014 – Jan 2015

See above

3.5 Planned research: Feb 2015 – Jan 2016

Mn_2GeO_4 is one of the few multiferroic materials where electric polarization is induced by a magnetic ordering which also gives rise to a net macroscopic magnetic moment [8]. At low temperature, this compound presents a complex magnetic state in which a commensurate and incommensurate magnetic order coexist. Elastic neutron scattering experiments show that, in this phase, magnetic ordering has non-vanishing wave vector components only in the *ab*-plane [8]. Considering uniform spin order along the *c* direction we find that ground state magnetic ordering might be studied by using a two-dimensional Heisenberg model in the *ab*-plane. Interestingly, considering the magnetic couplings involving only Mn-O-Mn bonds, in this model magnetic ions form a triangular lattice with four inequivalent exchanges between nearest neighbor spins. Using Monte Carlo simulations, we will study the low-temperature phase diagram in the exchange coupling space, for such a Heisenberg model.
The aim of the project is to investigate if the complex magnetic order of Mn_2GeO_4 originates from geometrical frustration, to obtain a set of values for the effective couplings in this compound, to check whether they can reproduce the observed magnetic excitations measured at PSI and the finite temperature phase diagram (using a three-dimensional model) observed. Finally, the study will be relevant for other compounds with olivine structure.

3.6 Personnel

NCCR personnel

- Andrea Scaramucci, postdoc, 50%, from September 2014.

Matching personnel

- Marisa Medarde, senior researcher, 100%, from May 2014.
- Andrea Scaramucci, postdoc, 100%, from May to August 2014.
- Andrea Scaramucci, postdoc (PSI), 50%, from September 2014.

4 DFT+DMFT study of strain and interface effects in d^1 and d^2 t_{2g}-perovskites (Nicola Spaldin — ETHZ)

4.1 Research summary

- Assess the effect of strain on metal-insulator transition in $LaVO_3$ and $LaTiO_3$.
- Electronic reconstruction at oxide-oxide interfaces.

4.2 Research question

Oxide heterostructures often exhibit emerging properties that are different from those of the constituent materials. One example is the observation of metallic behavior in heterostructures containing two otherwise insulating oxides. The complex nature of these insulator-to-metal transitions is nowadays the subject of intense investigation and debate, both because of its fundamental physical relevance, and because of its potentially ground-breaking technological applications. The complexity mainly stems from the simultaneous appearance of several new features in the heterostructure with respect to the bulk forms of the constituent materials. These include for example the epitaxial strain due to the lattice mismatch, the formation of interfaces, the possible development of structural defects, or local changes in stoichiometry. Understanding the specific role of each of theses factors is of paramount importance, also in view of tailoring heterostructures with desired properties. Such task can be accomplished via atomistic simulations based on density functional theory (DFT) plus dynamical mean field theory (DMFT), which can properly account for both the specific atomic structure of the heterostructure, and the correlated electron nature of these materials.

4.3 Scientific goals

a) *Short-term* As a first step it is desirable to understand the role of bare epitaxial strain in the insulator-to-metal transitions for some prototypical Mott insulators with different formal bulk electron configurations, such as $LaTiO_3$ (formal t_{2g} occupation d^1) and $LaVO_3$ (d^2). This provides the basis to develop a simple and, possibly, general model explaining the strain-induced changes of the metallic properties through a limited number of parameters (e.g. bandwidth, on-site energies, filling). Basically, this model should answer the question why some oxide thin-films become completely metallic under epitaxial strain whereas others are metallic only next to the interface with the substrate.

b) *Medium-term* The next step will be the explicit description of interface effects which take place at the boundary between two different oxides, such as in a thin film grown epitaxially on an insulating substrate (e.g. $SrTiO_3$). This is especially important for those material which develop a metallic state at the interface, but do not show any strain-induced bulk metallicity throughout the film (e.g. $LaVO_3$).

c) *Long-term* The long-term goal is to achieve a concise and comprehensive picture of oxide heterostructures composed of Mott insulators (such as $LaTiO_3$ and $LaVO_3$), and to clearly identify the source of the observed metallic properties. Within a material design perspective, this knowledge will enable us to predict and engineer new heterostructures displaying the desired properties, as well as to optimize the characteristics of already existing systems.

4.4 Plans and results: May 2014 – Jan 2015

The role of epitaxial strain in $LaTiO_3$ and $LaVO_3$ was addressed through a series of so-called "strained-bulk" calculations. The in-plane lattice constant is clamped to that of the substrate, while the out-of-plane one is relaxed, as shown schematically in Fig. 3. The evolution of the Mott phase as a function of

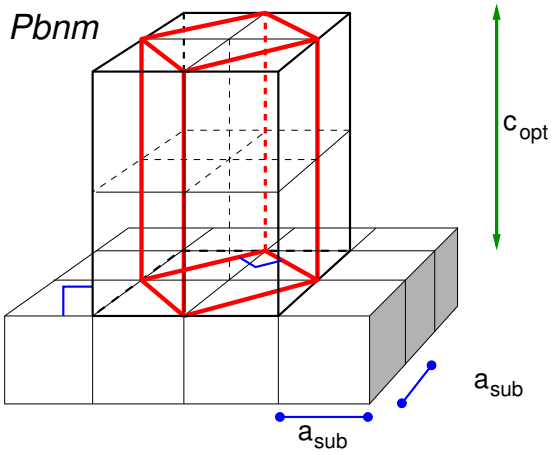


Figure 3: *Schematic picture of the Pbnm unit cell used to study the effect of epitaxial strain in $LaTiO_3$ and $LaVO_3$. The dimensions of the basal plane are clamped to the lattice constant of the substrate a_{sub}, whereas the out-of-plane vector c_{opt} and the internal degrees of freedom are optimized.*

strain was studied and interpreted. Our calculations show that mere epitaxial strain suffices to induce an insulator-to-metal transition in $LaTiO_3$, but not in $LaVO_3$ (Fig. 4), in agreement with recent experiments. We show that this difference can be explained by the combined effect of strain-induced changes in the crystal field splitting of t_{2g} orbitals and the different orbital filling in these two materials. To complete the picture in its generality, we also included the d^1 correlated-metal $SrVO_3$, which also serves as a benchmark of our methodology. We predict that this material will undergo

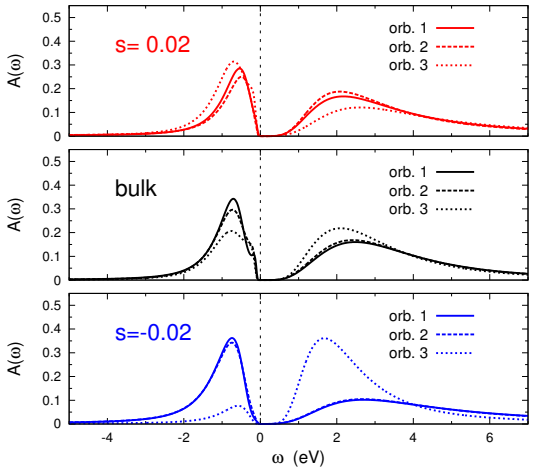


Figure 4: *Diagonal elements of the spectral function $A(\omega)$ for the three t_{2g} orbitals of epitaxially strained $LaVO_3$. The persistence of a band gap in $A(\omega)$ indicates that the Mott insulating phase of $LaVO_3$ is robust against both tensile strain ($s = 0.02$) and compressive strain ($s = -0.02$).*

a metal-insulator transition when a sufficiently large tensile strain is applied by the substrate. The inclusion of electronic correlations beyond the DFT level is mandatory to correctly capture the Mott phase, as well as the crossover to a metallic phase. DFT+DMFT provides an efficient methodology that fulfills this requirement, and it will be our method of choice. It leaves however some arbitrariness on how to treat some aspects, such as double counting term, charge self-consistency, size of the correlated Hamiltonian, choice and localization of the basis set of correlated orbitals. Therefore, the influence of these methodological aspects will be taken into careful consideration at all stages of the project, in order to ensure that the physical description of our systems is indeed sound.

4.5 Planned research: Feb 2015 – Jan 2016

The role of the oxide-oxide interface will be investigated through DFT+DMFT simulations of $LaVO_3/SrTiO_3$ heterostructures with varying superlattice periodicities and substrate terminations. The first aim is to assess whether the metallicity observed at the $LaVO_3/SrTiO_3$ interface could be driven by pure electronic reconstruction effects, rather than structural or stoichiometric reasons (such as, e.g., O-related defects). Several aspects need to be taken into account at the modeling level, such as possible surface terminations, structural relaxations at the interface, correct description of the electrostatic boundary conditions, and the proper description of the charge screening at the interface.
We also plan to extend these studies to different material combinations, with the long-term goal of building up a comprehensive understanding on electronic reconstruction at interfaces.

4.6 Personnel

NCCR personnel

- Gabriele Sclauzero, postdoc, 100%, from January 2015.
- Andrea Scaramucci, postdoc, 50%, from September 2014.

Matching personnel

- Claude Ederer, senior scientist, 10%, from July 2014.
- Gabriele Sclauzero, postdoc, 100%, from July to December 2014.
- Andrea Scaramucci, postdoc, 100%, from May to August 2014.

- Andrea Scaramucci (PSI), postdoc, 50%, from September 2014.

5 Design of a single-band Hubbard model material (Nicola Spaldin — ETHZ)

5.1 Research summary

Ab initio design of a new material whose Hamiltonian is that of the single-band Hubbard model.

5.2 Research question

The "standard model" of condensed matter physics is widely regarded to be the Hubbard model. It augments independent-electron band theory through a single parameter to account for electron-electron correlations and has been remarkably successful at addressing a range of correlation effects in transition metal materials. However, beyond one dimension, the problem is intractable and so much current research aims at finding appropriate approximations for obtaining the Hubbard model phase diagram. Here we explore a new approach: whether *ab initio* methods can be used to design a material whose Hamiltonian is exactly that of the Hubbard model so that its solution can be found by measuring the material properties.

5.3 Scientific goals

a) *Short-term* The first step is to identify a suitable crystal class and appropriate chemistries then use density functional theory to screen for the desired electronic structure, and dynamical mean field theory to study the Mott transition.

b) *Medium-term* For the most promising candidates, the structural stability will be checked as well as possibilities for doping and properties such as exotic superconductivity.

c) *Long-term* The long-term goal is to achieve an experimental synthesis and characterization of a proposed compound.

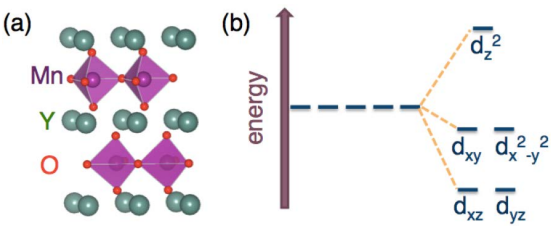


Figure 5: *(a) The hexagonal manganite crystal structure and (b) associated crystal field splitting. Note the isolated single d_{z^2} band.*

5.4 Plans and results: May 2014 – Jan 2015

Our strategy has been to engineer a non-degenerate d-manifold using crystal field considerations. This is manifested in a variety of ionic environments — pentagonal bipyramidal, square antiprismatic, square planar, square pyramidal and trigonal bipyramidal. To date we have explored the final option — trigonal bipyramidal — which is found in the hexagonal manganite structure with which we already have considerable experience (Fig. 5). The other crystal field options are possibilities for future exploration.

First we selected ion combinations that yield half-filling of this isolated d band and allow for doping across the range from empty to filled band. To half-fill the d_{z^2} band, we require a d^9 configuration, suggesting Cu^{2+}. Next we identified three promising anionicities — oxides, fluorides and sulphides — and the corresponding chemistries: $ZrCuX_3$ and $SnCuX_3$ for the oxides and sulphides (X = O and S) and $LiCuF_3$ and $NaCuF_3$ for the fluorides.

Next we performed structural optimizations and band structure calculations for our candidate materials using density functional theory (Fig. 6). We identified $LiCuF_3$ and $NaCuF_3$ as promising candidates based on their likely stability and electronic properties. We then explored the feasibility of chemical doping and

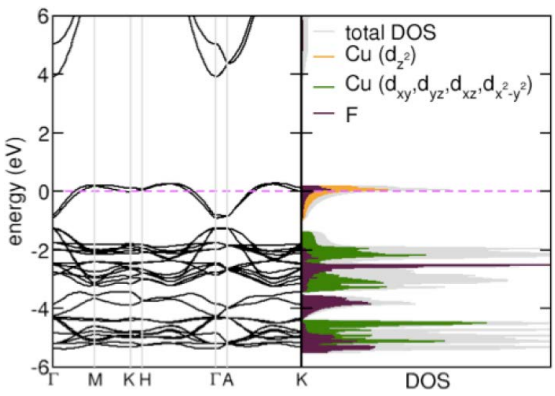


Figure 6: *Calculated band structure and density of states for paramagnetic $NaCuF_3$ in the hexagonal manganite structure (this work). The Fermi level is set to 0 eV and is marked by the dashed line. The electronic structure is particularly promising for achieving single-band Hubbard model behavior.*

reproduced the metal-insulator transition using dynamical mean field theory.

5.5 Planned research: Feb 2015 – Jan 2016

We are in active discussion with a number of experimental groups regarding the synthesis and characterization of our proposed compounds. In the meantime, in collaboration with the group of Thomas Schulthess, we begin to explore theoretically the possibility of exotic superconductivity. In addition we plan a MARVEL workshop to discuss other promising material classes and model Hamiltonians.

5.6 Personnel

Matching personnel

- Sinead Griffin, PhD student, 100%, from May to June 2014, postdoc, 100%, from July to December 2014.

6 Controlling the electronic state of transition metal oxides (Antoine Georges — UniGE)

6.1 Research summary

- Control and "orbital engineering" of transition metal oxides.
- Mechanism underlying the metal-insulator transition of nickelates.

6.2 Research question

Transition metal (TM) oxides present a large diversity of electronic properties, which offer potential promises for electronics or energy-related applications. Examples are the colossal magnetoresistance of manganites, the large thermopower of cobaltates, the metal-insulator transition of vanadates or nickelates, superconductivity of cuprates, etc. These materials, along with all materials involving strong electron correlations, display a number of competing phases, with small energy differences between them. An outstanding challenge is to *control these materials* in order to "guide" them towards a specific phase, hence selecting a desired functionality. These functionalities are determined by both chemical composition, crystal structure and electronic structure. The latter is determined by key ingredients such as: which orbitals form the low-energy electronic bands ("active orbitals"), what are their crystal field splitting and bandwidths, as well as orbital populations and interaction matrix elements (such as the screened Coulomb repulsion, Hund's and spin-orbit coupling). How to achieve control of these materials by acting on these "knobs" is the big research question which underlies the present project.

6.3 Scientific goals

A particularly promising class of materials on which we focus currently is the rare earth nickelates $R\mathrm{NiO}_3$. These materials display a phase transition (MIT) between a high-temperature metallic phase and low-temperature insulating phases (paramagnetic or magnetic). This transition, which can be close to room temperature, is highly sensitive to chemical composition (e.g. changing the rare earth ion R), as well as structural constraints and strain (tunable by changing the substrate or by heterostructuring). This could find applications to switches or to the recently proposed piezoelectric transistors. Furthermore, it has been proposed [9, 10] that strain control can be used in order to "orbital engineer" the nickelates, stabilizing the $d_{x^2-y^2}$ component of the e_g doublet at the expense of the $d_{3z^2-r^2}$ one. If full orbital polarization could be reached, this would lead to a "single-band" material, with an electronic structure very similar to that of a cuprate, and hence possibly to high-temperature superconductivity. A single active band is favorable because of: (i) the absence of competing orbital fluctuations, and (ii), importantly, the large antiferromagnetic superexchange expected in this case.

a) *Short-term* Deciphering the basic mechanism of the MIT in nickelates and clarifying to what extent orbital-engineering and the realization of a single-band material are possible for these materials are the basic short and medium term goals of this project. This is being done by combining first-principles electronic structure calculations with dynamical mean field theory (DFT+DMFT) using the most recent state-of-the-art techniques in this field (many of them developed in our own group). This is also done in close collaboration with experimental groups at UniGE: with Jean-Marc Triscone's group (to which our group is strongly linked through the ERC-Synergy grant QMAC) as far as strain control is concerned, and with Dirk van der Marel's group regarding optical spectroscopy. Exchanges with the experimental group of Bernhard Keimer at Max Planck Institute (Stuttgart)

are also taking place.

b) *Long-term* In the longer run, we plan to broaden our search to other TM oxides in which "orbital engineering" would prove feasible. Possible candidates are oxides of titanium and iron. This search will be a topic for exchanges and collaboration with the MARVEL ETHZ group of Nicola Spaldin. Other possible directions for the control of electronic functionalities of TM oxides will also be explored.

6.4 Plans and results: May 2014 – Jan 2015

In a work published just before the beginning of MARVEL [6], we established that the orbital polarization that can be reached under tensile strain is sizable, but not large enough to turn these materials into a single-band system (at least when considered in bulk or thin film forms, or for (nm) heterostructures with n, $m > 1$). Our results are in excellent agreement with recent experiments of Keimer's group at Max Planck Institute, Stuttgart [11]. Among the key reasons opposing full orbital polarization, as also pointed out by other authors [12], is the effect of Hund's coupling and of ligand holes, leading to an important spin-1 orbital compensated $d^8\underline{L}$ component of the ground state.
Our most significant work within MARVEL for this period is a second article, recently submitted [1] in which we propose a consistent theoretical description of the MIT of nickelates using only the strongly hybridized O-Ni e_g states which form the active degrees of freedom close to the Fermi level. The correlations are described by on-site interactions U (screened Coulomb repulsion) and J (Hund's coupling).

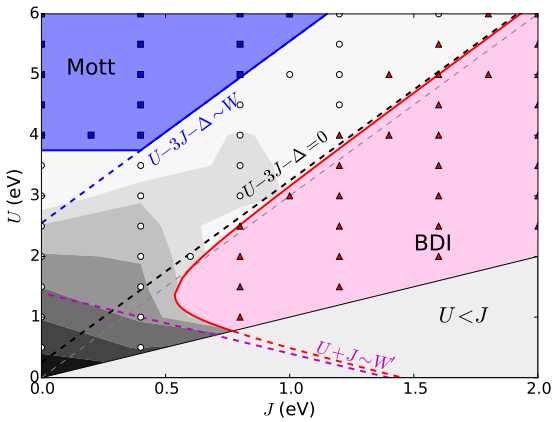


Figure 7: *Phase diagram of the DFT+DMFT e_g-only description of $LuNiO_3$ (see text), as a function of Hund's rule coupling J and on-site repulsion U.*

Using DFT+DMFT, we found that there is a wide region in the parameter space (U,J), in which the system reveals a strong sensitivity to even a slight bond disproportionation (BDI phase, Fig. 7). In particular, this region is characterized by the orthorhombic phase being a bad metal, while the distorted monoclinic phase is an insulator with a sizable modulation of the e_g occupancy and bond disproportionation (BDI). This picture validates previous proposals emphasizing the importance of ligand holes and "negative charge-transfer" [13], Hund's coupling [14] and "site-selective Mott transition" [15], and provides a new theoretical description of these effects from a low-energy perspective.

6.5 Planned research: Feb 2015 – Jan 2016

Our immediate goals are (i) to calculate optical spectra of nickelates based on the theoretical description above, allowing for direct comparison to ongoing experiments in van der Marel's group (UniGE) and (ii) to systematically investigate the materials dependence of the key energy scales involved in this description over the whole $RNiO_3$ series. These goals are likely to be reached by mid-2015. From then on, we plan to broaden our research related to the control of TM oxides and "orbital engineering" by considering other classes of materials, hopefully identifying or proposing materials in which a large degree of control can be achieved.

6.6 Synergies with other computational and experimental efforts

a) *Within MARVEL*

- Ongoing collaboration with Claude Ederer and collaborators (group of Nicola Spaldin, ETHZ), aimed at porting the interface between WANNIER90 and DMFT codes into the open-source software library TRIQS (http://ipht.cea.fr/triqs/1.1/index.html). This project should be completed during the first semester of 2015.
- Synergy with the group of Philipp Werner (UniFR) on extensions and generalizations of DMFT.

b) *Outside MARVEL* Experimental synergies with the groups of Jean-Marc Triscone and Dirk van der Marel (UniGE), as detailed above, as well as the Max Planck Institute, Hamburg, group of A. Cavalleri (within ERC Synergy QMAC).

6.7 Personnel

NCCR personnel

- Oleg Peil, postdoc, 100%, from September 2014.

Matching personnel

- Oleg Peil, postdoc, 100%, from May to August 2014.
- Alaska Subedi (based in Paris), 50%, from May to September 2014, on Ecole Polytechnique (France) funds.

MARVEL-related publications

List of publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.

[1] A. Subedi, O. E. Peil, and A. Georges, *A low energy description of the metal-insulator transition in the rare-earth nickelates*, arXiv:1410.2830 (2014).

[2] G. Autès, A. Isaeva, L. Moreschini, J. C. Johannsen, A. Pisoni, T. G. Filatova, A. N. Kuznetsov, L. Forró, W. Van den Broek, Y. Kim, J. D. Denlinger, E. Rotenberg, A. Bostwick, M. Grioni, and O. V. Yazyev, *A Novel Quasi-One-Dimensional Topological Insulator in Bismuth Iodide β-Bi_4I_4*, *submitted* (2014).

[3] D. Gresch, A. Soluyanov, D. Vanderbilt, D. Ceresoli, G. Autès, O. V. Yazyev, and M. Troyer, *Z2pack: a tool for computing topological invariants from band structure calculations*, *in preparation* (2015).

[4] J. C. Johannsen, A. Crepaldi, G. Autès, S. K. Moser, C. Tournier-Colletta, H. Berger, P. Bugnon, A. Magrez, J. Avila, M. C. Ascensio, F. Parmigiani, O. V. Yazyev, and M. Grioni, *Controlling Topological Surface States in the Natural Superlattice Series $(Sb_2)_m(Sb_2Te_3)_n$*, *in preparation* (2015).

[5] M. Morin, A. Scaramucci, M. Bartkowiak, E. Pomjakushina, G. Deng, D. Sheptyakov, L. Keller, J. Rodriguez-Carvajal, N. A. Spaldin, M. Kenzelmann, K. Conder, and M. Medarde, *Incommensurate magnetic structure, Fe/Cu chemical disorder, and magnetic interactions in the high-temperature multiferroic $YBaCuFeO_5$*, Physical Review B **91**, 064408 (2015).

[6] O. E. Peil, M. Ferrero, and A. Georges, *Orbital polarization in strained $LaNiO_3$: Structural distortions and correlation effects*, Physical Review B **90**, 045128 (2014).

Other references

[7] Y. Kawamura, T. Kai, E. Satomi, Y. Yasui, Y. Kobayashi, M. Sato, and K. Kakurai, *High-Temperature Multiferroic State of $RBaCuFeO_5$ (R = Y, Lu, and Tm)*, Journal of the Physical Society of Japan **79**, 073705 (2010).

[8] J. S. White, T. Honda, K. Kimura, T. Kimura, C. Niedermayer, O. Zaharko, A. Poole, B. Roessli, and M. Kenzelmann, *Coupling of Magnetic and Ferroelectric Hysteresis by a Multicomponent Magnetic Structure in Mn_2GeO_4*, Physical Review Letters **108**, 077204 (2012).

[9] J. Chaloupka and G. Khaliullin, *Orbital Order and Possible Superconductivity in $LaNiO_3/LaMO_3$ Superlattices*, Physical Review Letters **100**, 016404 (2008).

[10] P. Hansmann, X. Yang, A. Toschi, G. Khaliullin, O. Andersen, and K. Held, *Turning a Nickelate Fermi Surface into a Cupratelike One through Heterostructuring*, Physical Review Letters **103**, 016401 (2009).

[11] M. Wu, E. Benckiser, M. Haverkort, A. Frano, Y. Lu, U. Nwankwo, S. Brück, P. Audehm, E. Goering, S. Macke, V. Hinkov, P. Wochner, G. Christiani, S. Heinze, G. Logvenov, H.-U. Habermeier, and B. Keimer, *Strain and composition dependence of orbital polarization in nickel oxide superlattices*, Physical Review B **88**, 125124 (2013).

[12] M. Han, X. Wang, C. Marianetti, and A. Millis, *Dynamical Mean-Field Theory of Nickelate Superlattices*, Physical Review Letters **107**, 206804 (2011).

[13] T. Mizokawa, D. Khomskii, and G. Sawatzky, *Spin and charge ordering in self-doped Mott insulators*, Physical Review B **61**, 11263 (2000).

[14] I. Mazin, D. Khomskii, R. Lengsdorf, J. Alonso, W. Marshall, R. Ibberson, A. Podlesnyak, M. Martínez-Lope, and M. Abd-Elmeguid, *Charge Ordering as Alternative to Jahn-Teller Distortion*, Physical Review Letters **98**, 176406 (2007).

[15] H. Park, A. Millis, and C. Marianetti, *Site-Selective Mott Transition in Rare-Earth-Element Nickelates*, Physical Review Letters **109**, 156402 (2012).

Vertical Project 2

VP2 — Novel Materials Applications

Project leader: Ursula Röthlisberger (EPFL)

Participating members: Ursula Röthlisberger (EPFL), Alfredo Pasquarello (EPFL), Jürg Hutter (UniZH), Nicola Marzari (EPFL), Alessandro Curioni (IBM), Stefan Goedecker (UniBas), Wanda Andreoni (EPFL), Clémence Corminboeuf (EPFL), Daniele Passerone (Empa)

Summary and highlights: Vertical Project 2 (VP2) is devoted to the application of quantum simulations, enhanced sampling methods and materials informatics for the high-throughput design of solid materials and liquids. The specific objectives of VP2 are the development of novel materials for energy harvesting and storage and to a smaller extent to the capture of CO_2. Here are some of the highlights that have emerged during these first months.
1) *Photovoltaics:* in a joint experimental and theoretical investigation, it has been possible to develop a dye-sensitized solar cell (DSSC) with record efficiency based on novel computationally designed sensitizers [1]. In another experimental/theoretical collaboration, Röthlisberger *et al.* [2] have used molecular dynamics simulations of DSSCs to provide a molecular interpretation of the first *in situ* atomic force microscopy (AFM) measurements of a device under realistic working conditions and were able to identify a coverage-dependent phase transition of the absorbed dyes. In a further joint experimental/theoretical project, graphene nanoribbon heterojunctions were produced by combining hydrocarbon precursors with nitrogen substituted analogues and the electronic properties of these and related systems were calculated by Passerone *et al.* [3, 4] showing that the electronic levels of the graphene nanoribbon can be controlled via chemical substitution of the monomers.
2) *Energy storage:* Curioni *et al.* [5] were able to characterize the conduction properties of zirconium-containing lithium-lanthanum oxide, a promising candidate for use as solid state electrolyte in Li-based batteries, and could assess the effect of doping on the stabilization of the different polymorphs.
3) *CO_2 capture:* Andreoni *et al.* [17] have been able to give a detailed computational characterization of the chemical reactions involved in CO_2 capture and release in monoethanolamine solutions and are now extending the same type of study to novel liquid absorbents.

General view of the project

Major research questions

The high energy demands of modern society and the global warming of our planet are some of the most urgent areas where creative materials innovation with the power to translate into improved device performance is most urgently needed for a sustainable development. As a first step, VP2 thus aims at the fundamental atomistic/electronic understanding of processes and materials involved in energy harvesting and storage and CO_2 capture with the goal of designing novel materials and more efficient devices through quantum-based computational screening. To this end, VP2 will exploit the unique combination of expertise available in the NCCR MARVEL that integrates advances in quantum simulation methods (developed in HP3) with enhanced sampling techniques (developed in HP4) and materials informatics approaches (HP5). Optimal fostering of these complementary approaches will likely allow VP2 to make some revolutionary progress in the applications of computational materials design.

VP2 has started with a focus on three related and interlinked subprojects: the design of (i) photovoltaic materials (dye-sensitized and organic/inorganic perovskite-based solar cells), (ii) materials for energy storage (photocatalytic splitting of water and lithium/air batteries), and (iii) systems for efficient CO_2 capture. All subprojects share similar computational challenges in describing the structural and the electronic properties of materials and in overcoming difficulties associated to the efficient sampling of configurational and chemical space. In particular, most subprojects face the need for an accurate description of the relative position of energy bands and localized energy levels, as well as, for most, the accurate calculation of electrochemical properties. In addition, many of them also share the common feature of simulating complex solid/liquid or solid/gas interfaces. Due to this, several collaboration within VP2 as well as between VP2 and HP3,

HP4 and HP5 have already been initiated during the first few months of the NCCR MARVEL. The specific research questions for these different subprojects are detailed below.

a) *Photovoltaics* Several types of light harvesting materials and devices are currently under investigation in VP2 ranging from dye-sensitized solar cells to (main group) organic/inorganic perovskites and nanostructured carbon materials.
Dye-sensitized solar cells are a well established and cost effective technology that starts to become a real alternative to traditional silicon-based devices. However, photoelectric conversion efficiencies are still relatively low due to the limited spectral absorption range and competing recombination and loss processes. The latter are hard to control since almost no information is available about the detailed atomistic structure and mechanisms in realistic operating devices consisting of the dye-covered semiconductor surface, the solvent, electrolyte and various additives. One of the subprojects of VP2 aims at a better fundamental understanding of these issues through the use of complementary computer simulation techniques ranging from quantum mechanical electronic structure calculations, empirical force field and first-principles based molecular dynamics (MD) and excited state dynamics on comprehensive realistic systems. This information will in turn be used for the design of higher-efficiency dye-sensitized solar cells (DSSCs) with optimal absorption properties and minimal losses.
During the last year, also mixed organic/inorganic and fully inorganic halide perovskites, ABX_3 (A = organic or inorganic monovalent cation, B = bivalent cation, mostly Pb^{2+} or Sn^{2+}, X = halogen anion), have emerged as particularly promising novel light harvesting and charge transport materials and power conversion efficiencies as high as 24% have been reported for solid state perovskite-based solar cells in 2014. A significant research effort of VP2 is thus devoted to the investigation of perovskite materials to characterize the fundamental structure-property relationships, understand the role of defects and help in the search of novel environmentally friendly perovskite materials with even further enhanced efficiencies and improved stabilities.
In addition, surface supported carbon-based nanostructures are studied as new candidates for photovoltaic applications. In close collaboration with experimental groups, the effects of the nature and position of dopants will be systematically evaluated.

b) *Energy Storage* Besides optimal light harvesting, the development of efficient ways of storing the gained energy is of central importance for the generation of sustainable energy sources. A series of VP2 subprojects are thus dedicated to energy-storage materials for electrochemical cells. Identification of optimal materials which efficiently catalyze the photochemical splitting of water is particularly attractive, since the produced hydrogen fuel can recombine with oxygen giving nontoxic water as only waste product. Four subprojects within VP2 are focussing on this topic and one is dedicated to the search of novel materials for electrolytes, catalysts and anode/cathode electrodes for lithium/air electrochemical cells.
Hutter *et al.* study the basic mechanisms in water oxidation and reduction using molecular cobalt-cubane and pyridine/TiO_2 based catalytic systems, respectively. Corminboeuf *et al.* investigate the same process with exfoliated layered transition metal oxide materials (β-$Co(OH)_2$ or Birnesites ($CaMnO_x$)) and Marzari *et al.* use descriptor-based high-throughput quantum mechanical screening to identify rare earth perovskites as new candidate electrocatalysts for water splitting photoelectrochemical cells. Pasquarello *et al.* work on the development of computational tools, which are sufficiently accurate to screen inorganic materials against their catalytic functionality in the photochemical water splitting process. While in this initial phase these four subprojects are targeting distinct aspects of this chemical process, it is foreseen that, as the research proceeds, significant cross-fertilization will occur from which all these studies will benefit.
Curioni *et al.* investigate the fundamental properties of a specific class of solid state inorganic electrolytes, zirconium-containing lithium-lanthanum oxide (LLZO), as promising candidate for use as solid state electrolyte in Li-based batteries.

c) *CO_2 Capture* Industrial post-combustion carbon capture mainly uses monoethanolamine aqueous solutions as chemical absorbents. These solutions have a unique high reactivity with CO_2 and a relatively low cost but are known to be corrosive and to require high regeneration energy, which has promoted a large worldwide effort to identify and assess novel solvents that could reduce the high energy penalty and decrease the rate of amine degradation due to oxidation, corrosion and salt formation. The VP2 subproject of Andreoni aims at a

quantitative characterization of the chemical reactions involved in the capture and release of CO_2 to obtain useful information for a rational design of novel solvents and/or improved processes. To this end, they are employing extensive solvent screening via advanced *ab initio* simulations.

Short- and long-term goals

The short-term goals for all subprojects within VP2 are a thorough characterization and understanding of specific materials involved in either energy harvesting, energy storage or CO_2 capture with the long-term goal of using these insights for an efficient rational and high-throughput design of systems and devices with enhanced performance properties that will be validated in close collaboration with experimental groups. Specific short and long-term objectives for each activity are detailed below.

d) *Photovoltaics* The short-term goal of these subprojects are a characterization of the basic properties of light harvesting materials (DSSC, perovskites and nanocarbon materials) and a determination of the atomistic mechanisms involved in light conversion and charge transport. The long-term aim is the rational and/or high-throughput design and experimental verification of novel materials for more efficient photovoltaic applications.

e) *Energy Storage*
Water oxidation — The short-term goal of Corminboeuf's subproject is an understanding of the effects of exfoliation and a detailed characterization of the mechanisms involved in the oxygen evolution reaction with the ultimate aim of predicting new electrocatalysts. Hutter *et al.* pursue an analogous strategy using molecular transition metal catalysts and pyridine-decorated TiO_2 surfaces for water oxidation and reduction, respectively, and Marzari *et al.* will apply their high-throughput screening approach to investigate a class of lanthanide perovskites (with general formula ABO_3) as potential photocatalytic materials.
In the subproject of Pasquarello, band edges and the water redox levels will be aligned with respect to a common scale through the explicit modeling of the solid-water interface at first at the generalized-gradient approximation (GGA) density functional and later at the hybrid density functional theory (DFT) and GW level. In a later step, they will focus on the mechanisms of the charge transfer processes occurring at the solid-water interface, which might lead to additional criteria affecting the overall efficiency of photochemical splitting of water.
Lithium/air batteries — Curioni *et al.* plan to test the capabilities of different machine learning algorithms (in direct conjunction with Anatole von Lilienfeld and HP4) on screening different types of known solid state Li-ion conductors with the long-term goal of machine learning and data-driven design of novel classes of solid state inorganic electrolytes.

f) *CO_2 capture* The short and long-term goals of this project are the characterization of the basic reaction mechanisms involved in CO_2 capture and release in monoethanolamine solutions and the computational prediction of novel solutions with improved CO_2 capture capabilities.

First achievements

In spite of the short time since the start of MARVEL, there are already a number of first achievements.
Röthlisberger *et al.* were able to design a new dye sensitizer which led to a new DSSC with record efficiency [1] providing a first proof of principles that computationally guided materials search can indeed lead to the development of devices with significantly improved performance. In view of the fact that the photoconversion efficiency of DSSCs has leveled off since many years, the newly achieved performance jump is especially impressive.
Also in close collaboration with experiment, Passerone *et al.* calculated and predicted the properties of surface supported carbon nanostructures containing nitrogen heteroatoms [3, 4].
Hutter *et al.* studied the stability of many different configurations of a Co(II)-based cubane catalyst through comparison of electronic energy averages obtained from Born-Oppenheimer molecular dynamics sampling of the respective molecules in a periodic box of water.
Pasquarello *et al.* studied liquid water through isobaric molecular dynamics simulations based on a nonlocal van der Waals (vdW) density functional, demonstrating an overall improved description compared to standard gradient-corrected density functionals.
Curioni *et al.* were able to characterize the conduction properties of zirconium-containing lithium-lanthanum oxide, a promising candidate for use as solid state electrolyte in Li-based batteries, and could assess the effect of doping [5] on the stabilization of the different polymorphs.

Using *ab initio* simulation, Andreoni *et al.* have studied the chemical reactions leading to CO_2 capture and release in monoethanolamine solutions, have quantitatively characterized the entire cycle and have proposed a novel and well defined pathway that is consistent with available experimental data [17].

Next steps

In a next step, Röthlisberger *et al.* will study the aggregation behavior of ruthenium and porphyrin-based DSSCs with a comprehensive model including the TiO_2 surface, different dye coverages, acetonitrile solution and the electrolyte. They will also continue the characterization of the fundamental optical and transport properties of halide perovskites with a special focus on the effect of dopants. Corminboeuf, Hutter and Marzari will continue the characterization and design of molecular, exfoliated and crystalline rare earth perovskite photocatalysts for water oxidation. Pasquarello *et al.* will determine the redox levels with two different functionals which yield different structural properties, such as a standard gradient-corrected functional and a nonlocal vdW functional.
Curioni *et al.* will model the ion-transport process in solid state electrolytes, and select and evaluate good descriptors that characterize the transport properties of these materials to generate a dense database of known structures to be used in the machine learning.
Andreoni *et al.* intend to study CO_2 capture in a primary amine other than monoethanolamine and later in ammonia solutions. They will focus on the absorption process of CO_2 with the specific goal to investigate the formation of carbamates versus bicarbonates as primary products, and to explore pathways for carbamate conversion into bicarbonates.

Research contributions to the overall goals of the NCCR and to the existing literature

The development of a computationally designed DSSC with record efficiency has already contributed to the ultimate goal of the NCCR MARVEL, namely the revolutionary design of novel materials with enhanced performance. In addition, the numerous interactions and collaborations that have been initiated since the start of MARVEL, both within VP2 and between VP2 and the other projects of MARVEL, already start to contribute to the overall goal of the NCCR to pursue new lines of research and produce scientific results that are truly synergistic and more than the mere sum of individual contributions.

Collaborative components

The projects in VP2 share many common methodological aspects and address similar theoretical and computational challenges in the rational design and high-throughput screening of materials and solvents based on electronic structure calculations, and the challenge of efficient phase space and compound space sampling. Collaborative efforts within VP2 have been initiated (e.g. Andreoni/Pasquarello on the simulation of water and aqueous solutions) and will be further intensified during the second year. There are also tight links between the application module VP2 and the novel electronic structure methods developed in HP3 and the materials informatics' infrastructure and the efficient sampling techniques of HP4, and multiple collaborations have been started (e.g. Curioni/von Lilienfeld, Röthlisberger/von Lilienfeld, Röthlisberger/Goedecker).

1 Novel materials for solar cells (Ursula Röthlisberger — EPFL)

1.1 Research summary

This project aims at the computational characterization and design of molecules and solid state materials for application in solar cells. Specific goals are (i) to identify the relation between the chemical composition of halide perovskites and their band gap and their hole/electron effective masses; (ii) to understand the finite temperature dynamical behavior of the organic monovalent cation in methyl ammonium lead iodide and its effect on the electronic spectrum in the infrared range; and (iii) to investigate the interaction between dyes and the TiO_2 surface, and its effect on the spectrum of the sensitizer.

1.2 Research question

Solar cells are developed on the basis of several approaches, including single and multi-junction solar cells, semiconductor quantum dots, organic polymers, sensitized wide band gap solar cells and many others. Here, we focus on sensitized solar cells, and in this field investigate two different kinds of solar cell mate-

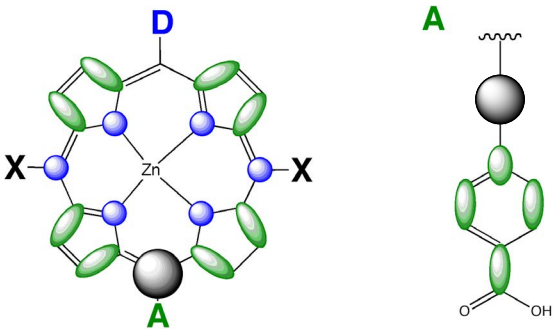


Figure 1: *General structure of a porphyrinic dye. In green is reported the density isosurface of the HOMO and in blue of the LUMO orbitals, representing the single particle states of the excited electron and hole.*

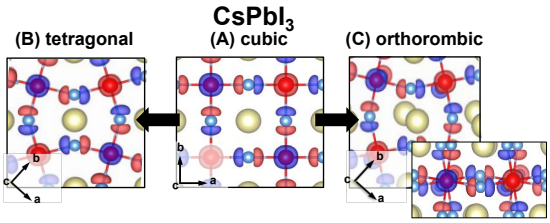


Figure 2: *$CsPbI_3$ in three in crystal phases: tetragonal (left), cubic (center), orthorhombic (right). Isosurfaces of the HOMO orbital (red positive and blue negative) highlight its antibonding nature.*

rials: dye-sensitized solar cells and perovskite-based photovoltaics.

a) *Dye sensitized solar cells* Dye sensitized solar cells (DSSCs) are a well established technology, which recently reached an efficiency of $\sim 13\%$ [1]. Traditionally, ruthenium complexes are used as sensitizers but more recently porphyrin derivatives have emerged as a promising alternative (Fig. 1). The objective of research in this field is to find dyes that can maximize solar light absorption resulting in excited states, in which the exited electron and the hole are spatially separated, so as to minimize recombination.
The best performing dyes are complex molecules containing long alkyl chains in order to prevent the contact between the electrolyte, used for dye regeneration, and the TiO_2 semiconductor surface, which would annihilate the injected electrons, equivalent to short-circuiting the solar cell from the inside. Thus, a further objective in the development of novel dyes is to develop molecules, which can reduce this phenomenon. This is achieved by combining sterical hindrance by the branches mentioned above and a high surface coverage, which in addition also increases light absorption efficiency. Steric hindrance and high packing of dye molecules, required by the high surface coverage, are somewhat conflicting characteristics. We are investigating how all these features can be optimized in a single dye molecule while maintaining favorable optical and redox properties.

b) *Halide perovskite-based solar cells* (Hybrid organic-inorganic) lead iodide perovskite, $APbI_3$ (A^+ = methyl ammonium, formamidinium, cesium (Fig. 2), etc.), have recently shown a photovoltaic efficiency close to 18%. Furthermore, these materials can easily be prepared in a cost effective manner. This promising results have initiated an intense research effort to characterize and understand the fundamental properties of these materials and to develop systems with even higher efficiency. One research direction is to replace lead with less polluting elements, e.g. tin. However, tin-halide perovskites are unstable and decompose when getting into contact with air. Moreover, they contain a significant amount of intrinsic defects, which makes their performance lower than the one of their lead-based analogues. Thus, another active research line is devoted to the development of preparation protocols that result in more defect-free crystals. One strategy is doping or alloying perovskites with other elements, thus producing compounds of the type $ASn_{1-x}B_xI_3$, where A is a generic monovalent cation, organic or inorganic, and B is a dopant that can replace Sn in the perovskite lattice.

How preparation protocols affect the chemical composition and structural characteristics of halide perovskites is unknown. Not known is also how much optical and charge carrier transport properties of alloys depart from those of the pure $ASnI_3$. We will use a range of computer simulation techniques to gain an understanding of the fundamental relation between stoichiometry/structure of perovskites and their optical and transport properties. Once these basic relations are understood, they can be used to design novel, more efficient materials. Simulations will also be used to predict the most likely defects present in the material, and their effects on the optical and transport properties.

To implement this ambitious plan, we use a hierarchical approach, in which we combine simpler but more efficient *ab initio* techniques with advanced search algorithms to identify candidate materials with improved properties, and more accurate quantum mechanical methods to study in detail the properties of the best candidate systems. This approach allows to focus the effort of our experimental collaborators

(the group of Michael Graetzel at EPFL) to a restricted list of candidate samples with an accurate previous evaluation of their properties via computer simulations.
In particular, we are using density functional theory (DFT) for an identification of stable structures and a preliminary screening of optical properties. GW/Bethe-Salpeter calculations are used for further refining the optical property predictions of promising novel materials. As for the transport properties, we use DFT in conjunction with standard charge carrier trapping models to identify defects that can limit the transport efficiency of novel materials, which identify neutral or low-charged deep state defects as possible hole-electron recombination centers. For selected systems, we also use excited state dynamics approaches to confirm the qualitative conclusions achieved on the basis of the more heuristic models just mentioned.

1.3 Scientific goals

a) *Dye sensitized solar cells* During the first year, we have been studying the coverage-dependent packing of dye molecules on the (101) surface of TiO_2 with the help of force field based molecular dynamics simulations. The results of these simulation were used to interpret low-resolution *in situ* atomic force microscopy (AFM) measurements [2]. We have also started to investigate the effect of the presence of the TiO_2 surface on the optical properties of dye molecules. Despite the intense computational research in this field, realistic simulations, with dye molecules absorbed on the TiO_2 surface in presence of solvent and electrolyte are still lacking. Understanding this effect is crucial for further improving the contribution of advanced simulations on the design of novel dye molecules.

b) *Halide perovskite-based solar cells* Our first aim is to study the fundamental optical and transport properties of both known and novel halide perovskite materials. We focus on $APbX_3$ and $ASnX_3$ systems (X = Cl, Br, I; A = Cs, Rb, K), which are prototypical classes of halide perovskites used to investigate the fundamental structure-property relationships of these materials to systematically identify the factors that influence, e.g., the band gap and transport properties. We will also investigate modified Sn-based systems, which are candidate materials for lead-free perovskite-based solar cells. In this case, we aim at identifying modifications (e.g. doping) that can reduce the concentration of defects and limit the oxidative degradation sensitivity of these systems.

c) *Long-term* On a long term perspective, we want to develop a comprehensive strategy for the automated characterization and design of dye sensitizers and perovskite-based solar cells. To this end, we will combine electronic structure calculations, force field based and *ab initio* molecular dynamics, and time-dependent DFT (TDDFT) based excited state dynamics with enhanced sampling techniques for the identification of the most stable crystalline phases and the application of evolutionary algorithms for efficient exploration of compound and materials design space.

1.4 Plans and results: May 2014 – Jan 2015

a) *Dye sensitized solar cells* During this period, we have developed force fields for ruthenium and porphyrin dye sensitizers and performed classical molecular dynamics simulations to gain insights into the coverage dependent packing of the dyes on the titanium surface and the solvent and electrolyte accessibility. The results of the ruthenium-based dyes were used to interpret the first *in situ* AFM measurements of these systems [2]. We also performed simulations on a fully solvated porphyrin dye bound to a TiO_2 surface with two different binding modes, one in which the O-H bond of the carboxy anchoring group is dissociated, and one in which it is not. We first relaxed the system classically and then run relatively long ($5-10$ ps) first-principles molecular dynamics (MD) simulations. We are now computing the electronic spectra of the two systems, duly averaged over an ensemble of configurations sampled along the MD. These spectra will be compared with that of a dye molecule in vacuum and continuum solvent to understand the effect of the surface, electrolyte and solvent.

b) *Halide perovskite-based solar cells* In this initial part of the project we focused on the identification of parameters that determine and control the band gap and hole/electron effective masses of a wide set of halide perovskites, and the dependence of these properties on the chemical composition and crystalline structure of the system. Two manuscripts are in preparation, and at least one [6] will be submitted within the deadline of this first-year report.

1.5 Planned research: Feb 2015 – Jan 2016

During the second year, we will complete the analysis of our classical molecular dynamics simulations of DSSCs and identify general guiding principles for the design of efficient donor-bridge acceptor type dyes. For the halide perovskite systems, we will focus on a qualitative and quantitative characterization of the nature and the effects of defects in both lead- and tin-based perovskites. We will also investigate their excited state and transport behavior.

1.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* This project is linked to the collaborative projects with the Goedecker group on minima hopping (HP4) and the von Lilienfeld one on machine learning (HP4).

b) *Outside MARVEL* We are developing our theoretical work in collaboration with the experimental group of Prof. Graetzel at EPFL.

1.7 Personnel

NCCR personnel

- Simone Meloni, postdoc, 80%, from May 2014.
- Nicholas John Browning, PhD student, 10%, from May 2014.

Matching personnel

- Negar Ashari Astani, PhD student, 90%, from May 2014.
- Giulia Palermo, PhD postdoc, 50%, from May 2014.
- Ariadni Boziki, PhD student, 100%, from December 2014.
- Martin Bircher, PhD student, 10%, from May to December 2014.

2 Phase discovery and phase diagram of halide perovskite via minima hopping (Stefan Goedecker — UniBas, Ursula Röthlisberger — EPFL)

2.1 Research summary

In this project, we apply the minima hopping method to identify the possible phases of the prototypical halide perovskite $CsPbI_3$. As a first outcome, new phases of cesium lead iodide were identified.

2.2 Research question

Hybrid organic-inorganic lead iodide perovskite, $APbI_3$ (A^+ = methyl ammonium or formamidinium), have shown a photovoltaic efficiency close to 18%. More striking is that these results have been achieved with materials produced by a cheap solution process. These two factors contribute to make halide perovskites a very promising material for photovoltaic applications.
However, due to the liquid processing, there is only limited control of the crystal phase that is formed. For example, in the case of the relatively well known methyl ammonium lead halide perovskites, experiments evidence the possible presence of multiple phases (cubic, tetragonal and orthorhombic) in different and, sometimes, the same device. The crystalline phases can crucially affect the optical and transport properties of the material. Thus, protocols for the *in silico* design of novel materials require the investigation of the various possible polymorphs.

The identification of the phases of a crystal is not a simple task. In 1988, the then editor of *Nature*, John Maddox, began an editorial with the following statement: "One of the continuing scandals in the physical sciences is that it remains impossible to predict the structure of even the simplest crystalline solids from a knowledge of their composition." A reliable structure prediction requires two ingredients: a method that is able to provide an accurate potential energy surface, and an algorithm which can find all the low-energy local minima on this potential energy surface. The global minimum gives the ground state structure and the other low-energy minima correspond to metastable structures. The potential energy surface is a high-dimensional function which gives the energy of the system as a function of the atomic coordinates of all its constituent atoms. Force fields and similar schemes are fast but are in general not highly accurate. Density functional theory gives a much higher accuracy and is therefore the method of choice for structure prediction if it is computationally affordable. Some of the latest exchange correlation functionals, such as the PBE functional [18], are nowadays the gold standard for simulations of bulk materials and give highly reliable structural properties. Be-

cause of the high cost of density functional calculations, it is of utmost importance that the algorithm which is used to locate the global minimum and other low-lying minima is highly efficient and requires the smallest possible number of energy and force evaluations. In the last decade, a lot of progress has been made along these lines. In particular, several algorithms have been proposed to find low-energy minima on the potential energy surface in an efficient way. One of these methods is the minima hopping approach [19] developed in the group of Goedecker. It has already allowed to solve several challenging structural problems. Initially it was applied to the structure prediction of clusters and fullerenes. Later on, it was extended to predict the structure of crystalline materials and surfaces [20]. The minima hopping method is universally applicable to any condensed matter system since it is built on the elementary operations of molecular dynamics and local geometry optimizations.

2.3 Scientific goals

a) *Short-term* The short-term objective of this project is to develop a reliable and efficient approach for the identification of the polymorphs of novel materials. In this context, we are testing the minima hopping algorithm for an exploration of the low-energy crystalline structures of the prototypical halide perovskite $CsPbI_3$. To perform this search, we have coupled the Quantum-ESPRESSO code with the black minima hopping algorithm using a communication scheme based on sockets (see also the project description in HP4).

b) *Long-term* The long term goal of this project is to crucially assist the computational design of efficient halide perovskite-based solar cell materials. The identification of the possible crystalline phases and the existence of phase transitions is of primordial importance for the overall design process. The first tests with the minima hopping method provided very promising results that we will extend to construct a reliable phase diagram for $CsPbI_3$ and apply the same procedure to the most promising materials identified in the VP2 subproject "Novel materials for solar cells" (section 1).

2.4 Plans and results: May 2014 – Jan 2015

In a first step of the project, we investigated the possible crystalline phases of $CsPbI_3$, which is often used as a simplified prototype material

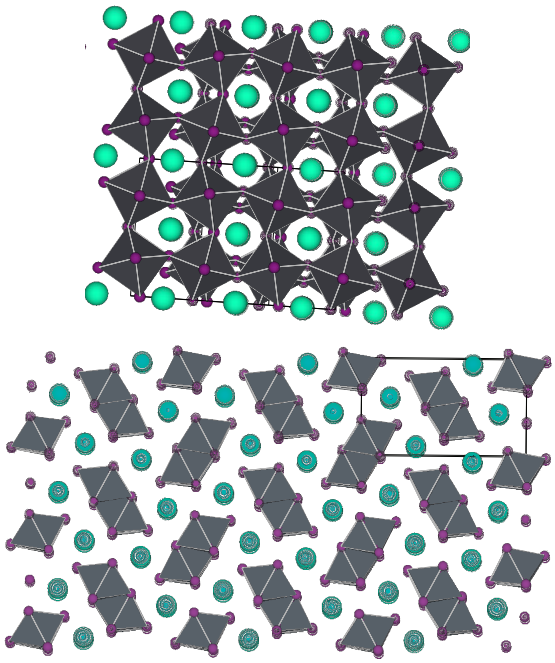


Figure 3: *One of the several 3D perovskite-like (top) and 1D δ-like (bottom) phases of $CsPbI_3$ identified along our minima hopping simulations.*

with properties similar to the more often employed methyl ammonium system. $CsPbI_3$ is experimentally not well characterized and theoretical structure predictions are therefore very important. During the first few months of this project, we obtained already an initial set of structures for $CsPbI_3$. This set contains several quite distinctive structural motifs; in fact the structures identified can be divided in two classes: 3D perovskite-like structures (Fig. 3, top), and 1D δ-like ones (Fig. 3, bottom). Some of these structures have not been identified before. We plan to extend this search and to compute the free energy of the various phases. This will allow us to build a phase diagram of $CsPbI_3$.

2.5 Planned research: Feb 2015 – Jan 2016

In the second year of this project, we will establish a full phase diagram and try to get experimental verifications of the newly predicted phases. We will also extend the same type of approach to tin-based perovskites.

2.6 Synergies with other computational and experimental efforts

a) *Outside MARVEL* We are developing our theoretical work in collaboration with the experimental group of Prof. Graetzel at EPFL.

b) *Within MARVEL* This work is a collaboration between the groups of Stefan Goedecker at UniBas and Ursula Röthlisberger at EPFL

2.7 Personnel

NCCR personnel

- Li Zhu, postdoc, 100%, from August 15 2014.
- Simone Meloni, postdoc, 20%, from May 2014.
- Nicholas John Browning, PhD student, 10%, from May 2014.

Matching personnel

- Max Amsler, postdoc, 100%, from August 2014.
- Negar Ashari Astani, PhD student, 10%, from May 2014.

3 Systematic doping of molecular precursors: toward a high-throughput scanning of carbon-based nanomaterial properties (Daniele Passerone — Empa)

3.1 Research summary

- Carbon-based structures. Controlled doping of molecular precursors.
- Nanoheterojunctions. Interplay with experiments.

3.2 Major research question

The major research question asked within this scientific project concerns the systematic theoretical investigation of surface-supported carbon-based nanostructures by a bottom-up approach based on molecular precursors with different compositions. Specific to our group is the proximity to a laboratory where an efficient bottom-up strategy for the synthesis of a class of novel promising low-dimensional nanomaterials for photovoltaics and electronics has been successfully developed.
Systematically changing the doping nature, position and amount in the precursors induces regular modifications in the extended structure, and simulation allows scanning a large amount of putative precursors, thus limiting the actual number of experiments to be performed.

3.3 Scientific goals

The present agility grant supports the extension to low-dimensional nanostructures of one of the main MARVEL scientific questions, namely identifying suitable materials with specific functionalities through computational design.
In a recent publication [3] the laboratory nanotech@surfaces was able to produce graphene nanoribbon heterojunctions by combining pristine hydrocarbon precursors with nitrogen-substituted equivalents. In that work, we computed electronic properties of those structures extending the analysis to cases not explicitly considered experimentally, and we showed for example the control of the graphene nanoribbon electronic levels via monomer chemical substitution (Fig. 4). The mid-term scientific goal is to extend the present systematic analysis to a broader class of materials with applications in the domains of interest for VP2,

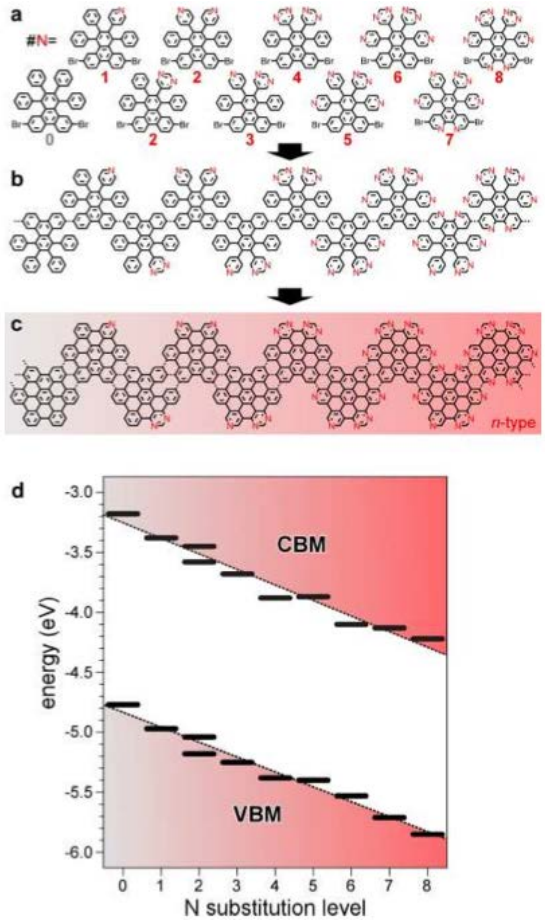


Figure 4: *Graphene nanoribbon electronic level control via monomer chemical substitution (from [3]).*

in particular for photovoltaics. The short and long term scientific goals include the following steps.

a) *Short-term* Dope the molecules already available to the experimentalists with species of different chemical nature, as an example borazine-based units to increase the band gap in the resulting nanoribbon, applied on already available precursors.

b) *Long-term* We propose to augment the success rate of the well-established bottom-up surface synthesis strategy through the application of the MARVEL concepts of high-throughput screening of materials candidates. The parameters to scan are in this case:

- shape of the undoped precursor, in accordance to the feasibility of the organic synthesis;
- dopant species (atomic doping, molecular fragments like borazine to enhance the gap, different edge terminations);
- amount of doping and position of the dopants within the molecule.

3.4 Plans and results: October 2014 – Jan 2015

The hires of this agility grant consist in the assignment of Dr. Carlo A. Pignedoli to MARVEL-specific research projects starting from October 2014. During the reporting period, Dr. Pignedoli showed his potential as material designer in interplay with the experiment through several MARVEL-related publications [3, 7, 4, 8, 9, 10, 11, 12]. In [3] the ability to control electronic properties of atomically precise graphene-based heterojunctions was demonstrated experimentally. Dr. Pignedoli and collaborators could drive the laboratory analysis through extensive targeted simulations on differently doped structures. The study includes the electronic structure analysis of realistic, extended nanostructures with resolution along the interface profile with different nitrogen content in the molecular precursors; the connection to electrical conductance of the putative electronic device; the demonstration of the possibility of fully tuning the electronic level via monomer chemical substitution.
Already in this initial phase of MARVEL, the peculiar role of our group in the community has appeared clear.

- Given the fact that the group is embedded in an experimental laboratory, with which a fruitful collaboration has led in the last years to important joint results in the fields of surface catalysis, surface chemistry and carbon-based nanostructures, we can act as bridging facility between MARVEL projects and platforms and the applied science world.
- We promote initial contacts and scientific plans between experimental groups at Empa and different research entities within MARVEL. As an example, we cite the contacts promoted between Lars Jeurgens (Empa laboratory of Joining Technology and Corrosion) and Alfredo Pasquarello, as well as the plan of collaboration involving our group, Nicola Marzari and the nanotech@surfaces laboratory about one-dimensional polar discontinuities.
- By sticking to a class of materials that have already shown potential for an industrial application (Empa projects and contacts with BASF and IBM), we enter the pipeline of material design vs experimental realization at an advanced stage, namely at the point where it is already clear that the *in silico* material scan will have a good chance to be translated into the lab for the most promising cases.

3.5 Planned research: Feb 2015 – Jan 2016

The main research lines we plan to follow during the next year are:

- the systematic doping of molecular precursors will be extended to other classes of molecules — and thus, to other classes of extended nanostructures; we plan interactions with Jürg Hutter, Alfredo Pasquarello and Joost VandeVondele for the application of hybrid functionals and novel linear scaling algorithms to the simulation of extended nanostructures in the presence of a metallic substrate; we also plan to apply the concepts and tools of AiiDA (PP6) to the high-throughput scanning and databasing of structures with different kind of doping;
- in a recent work [13], the group of Nicola Marzari was able to demonstrate the possibility of engineering polar discontinuities in honeycomb lattices by creating interfaces between differently functionalized nanostructures. The plan of the research of the next year concerns thus the application of this idea to materials that can be realized in the laboratory. If possible, this activity will be blended with

parallel experimental projects approved through the next project call for PP7.

3.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Our group has a long-standing collaboration with the developers of the CP2K code at UniZH (Jürg Hutter) and at ETHZ (Joost VandeVondele). Given our proximity to the laboratory, our role is to propose realistic testbeds for novel features of the code, such as computational X-Ray spectroscopy. An example is the recent implementation of the X-ray absorption near-edge structure (XANES) code in CP2K, inspired by the experiments performed at the novel facilities of the Nano-Lab at IBM. A further role of our group is to act as a connection and MARVEL proximity group to Empa laboratories working on themes related to the NCCR. We also recently started a collaboration with Marco Gibertini in the group of Nicola Marzari (one-dimensional polar discontinuities).

b) *Outside MARVEL* Concerning theory, the group has open collaborations with Prof. E. Tosatti (SISSA/ISAS, Trieste), with Dr. A. Ferretti, Dr. D. Prezzi and Prof. E. Molinari (University of Modena, Italy), Prof. V. Meunier (Troy, USA). Concerning experiments, we have collaborations with Prof. T. Jung (PSI/UniBas), Prof. E. Meyer (UniBas), Prof. K. Mullen (Max Planck Institute for Polymer Research, Mainz), Dr. R. Denk (Institute of experimental physics, Linz, Austria).

3.7 Personnel

NCCR personnel

- Carlo A. Pignedoli, senior scientist, 60%, from October 2014.

4 Photo-catalyst design for water oxidation and reduction (Jürg Hutter — UniZH)

4.1 Research summary

- Co(II)-based cubane molecules in solution.
- Water oxidation catalyst.
- Pyridine-based molecules on rutile(110) surface.
- Water reduction catalyst.

4.2 Research question

We are investigating the basic mechanisms in water oxidation and reduction catalysis. Insight gained from detailed atomistic studies should allow to device design strategies for optimized molecular structures for catalysis.

4.3 Scientific goals

In this project we investigate the structure and dynamics of a family of Co(II)-based cubane molecules as water oxidation catalysts (WOC) and pyridine-based molecules on TiO_2 as water reduction photo-catalysts. The calculations will help, together with experiments, to deepen the understanding of the mechanisms involved in the reactions, and to identify the crucial structural and electronic motifs that determine the catalytic efficiency.

a) *Short-term* We will simulate Co(II)-based cubane molecules in solution, understand the geometric and electronic structure and make connection to experimental observation for these systems. Further, we will establish simulation protocols for the investigation of water reduction catalysts based on ligand molecules with a cobalt atom center adsorbed on a rutile surface. We will investigate the effect of the adsorbate on the electronic structure of the TiO_2 and guide synthesis of novel ligand derivatives.

b) *Long-term* The general goal of our work is to understand the details and energetics of complete reaction mechanisms, determine the most important factors influencing water oxidation and reduction catalysts activity and, with this knowledge at hand, to finally design efficient and robust catalysts.

4.4 Plans and results: May 2014 – Jan 2015

a) *Water oxidation catalyst* For a Co(II)-based cubane **1** and also a novel $Co_3^{II}ErO_4$-based cubane **2**, we studied the stability of many different configurations through comparison of electronic energy averages obtained from Born-Oppenheimer molecular dynamics sampling of the respective molecules in a periodic box of water. Our findings point toward far higher ligand mobility, corresponding with higher WOC activity, for **2** compared to **1**. For **1** it could be shown that, once a mono-dentate acetate ligand is lost, the mobility of the bidentate acetate ligands is drastically increased. Furthermore, we elaborate on the issue of buffer anions present in solution

hampering the catalytic activity by demonstrating that phosphate binds thermodynamically favorably to **1**. We considered entropic contributions in our calculations to be negligible, which we rationalized by performing vibrational analysis and calculations of thermochemical properties. All our findings agree well with experiments obtained from FT-IR and X-ray absorption techniques.

b) *Water reduction catalyst* To modify the band gap of TiO_2 and increase its photocatalytic activity, a pyridine-based molecule is used as photo-sensitizer and a cobalt atom as metal center. We performed density functional calculations on the rutile(110) surface. Our strategy for the investigation of adsorption on the TiO_2 surface is first optimizing the structure at the PBE level of theory, afterward obtaining more accurate electronic structure information of the optimized systems with the Heyd-Scuseria-Ernzerhof (HSE06) hybrid functional. We have determined the preferential adsorption geometry and adsorption energies of the target molecule on the TiO_2 surface. The most stable TiO_2-molecule configuration has been investigated further by calculating the band structure and projected electronic density of states. Moreover, the preferential adsorption sites of photo-sensitizer and metal center on the TiO_2 surface and effect of the molecule on surface modification of TiO_2 have been analyzed.

4.5 Planned research: Feb 2015 – Jan 2016

a) *Water oxidation catalyst* We will further investigate the effects of solvation on the cubane systems by carrying out calculations using different solvation models, namely a periodic box with explicit water molecules, vacuum calculation (no solvation), and with implicit solvent models. In addition, we plan to analyze the first proton coupled electron transfer (PCET) step of the catalytic cycle of **1** and a simpler model system consisting of an octahedrally coordinated Co^{2+}-ion. To this end, we will employ and test different methods ranging from simple electronic energy differences to calculations of free energies and barriers.

b) *Water reduction catalyst* We will further investigate the interaction of pyridine-based molecules with a TiO_2 surface. We will use the rutile(110) surface in our model. The effect of metal atoms added to the inner core of the molecule and the structural properties of this system will be studied. The modification of the band gap by the absorption of the molecule on the surface will be investigated. The simulation protocols will be refined. Especially the choice of density functional will be verified. We will then focus on the mechanism of water reduction, possible intermediate products, and energy barrier for hydrogen production.

4.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* There are currently no fully established collaboration within MARVEL for this project. However, we are in contact with the groups of Alfredo Pasquarello and Clémence Corminboeuf working in VP2 regarding the exchange of technical aspects for the simulation of this type of systems.

b) *Outside MARVEL* Our efforts in this project are coordinated with the UniZH Research Priority Program (URPP) LightChEC (`http://www.lightchec.uzh.ch/`). The final goal of this program is to create an architecture of catalysts and antennas to mimic the natural photosynthesis model. The individual steps: light harvesting, oxidation and reduction of water, are combined into one single functional system.

4.7 Personnel

Matching personnel

- Sandra Luber, postdoc, 100%, from May 2014.
- Yeliz Gürdal, PhD student, 100%, from May 2014.
- Florian Hodel, PhD student, 100%, from May 2014.

5 Water oxidation at exfoliated transition metal oxides — from concepts to new electrocatalysts (Clémence Corminboeuf — EPFL)

5.1 Research summary

This project aims at elucidating central aspects of water oxidation at transition metal oxides. A key aspect will be to understand the mechanisms responsible for the catalytic improvements found upon exfoliation of NiFe and CoNi mixed oxides. The influence of exfoliation will be tested on β-$Co(OH)_2$ by consider-

ing different active sites and geometrical distortions of the monolayer system. The effect of doping with non-inert ions will be investigated as well.

5.2 Research question

Splitting of water into hydrogen and oxygen is a central question in electrochemistry already for several decades. During the last years, this reaction gained even further attention due to its importance in the electrochemical production of renewable energy carriers such as hydrogen. During this process the oxygen evolution reaction (OER) can not be avoided and is indeed also the main source of the overpotential [21]. State-of-the-art electrocatalyst for the oxidation of water to oxygen at the anode are dimensionally stable anodes (DSAs) which were developed already in the 60s [22]. DSAs consist of a mixture of IrO_2, RuO_2 and TiO_2 painted on titanium. These electrodes combine high activity with good conductivity and long term stability even under extreme conditions [22]. While DSAs indeed combine a number of favorable electrochemical properties, rendering them central in todays electrochemical industry, their usage at even large scales as required for the production of green energy carriers is hampered by the need to rely on scarce and expensive Ir and Ru oxides. This is in contrast to biological systems where oxygen is evolved inside photosystem II (PS II) relying solely on a Mn_4CaO_x tetramer to catalyse the same reaction [23]. Inspired by PSII and due to the need to develop electrocatalysts based on more abundant materials a large number of transition metal (TM) oxides were studied during the last years. Among those cobalt [24, 25] and manganese [26] oxides were found to be promising candidate catalysts. Further enhancement of the pure oxides' activity was found when mixing with various other transition metal oxides such as in NiFe [25, 27], CoFe [24], CoMn [28] or by adding traces of gold to cobalt [29] or manganese [30] oxides. Explanations for the enhancements based on mechanistic insights and first-principles calculations have been proposed [27].

A special class of OER catalysts are layered TM oxides such as β-$Co(OH)_2$ [25] or Birnessites ($CaMnO_x$) [26]. While these materials are already promising electrocatalysts in their bulk structure [26, 25], an even lower overpotential was found upon exfoliation [25]. Details on the reasons for these improvements are yet to be discovered.

Considering the importance of the water oxidation reaction, it is the aim of this subproject to shade light on central questions by means of first-principles calculations and to generate a basic understanding allowing to predict new candidate electrocatalysts. An important question within this framework will be to understand the mechanisms responsible for the improvements found upon exfoliation of nanosheets and mixing of transition metals.

5.3 Scientific goals

a) *Short-term* In the short term, the research will be focused on understanding the improvements resulting from exfoliation of layered transition metal oxides studied by Hu *et al.* [28, 25]. Both the influence of the formation of single monolayers on the catalyst as well as the reasons responsible for the improvements resulting from mixing with other transition metal oxides will be considered. In detail the following questions will be addressed by means of density functional theory (DFT) calculations.

- Evaluation of the differences between exfoliated layers and comparable multilayered electrocatalysts. Here both the relative stability of the intermediates as well as possible changes in the OER mechanism will be considered.

- Search for the active sites at exfoliated oxide layers. A large number of differently bound oxygen species can be found at exfoliated TM oxides which may all contribute to the overall activity of the catalyst. Thus, detailed knowledge of the activity of the different sites is central, since it may allow for tailoring the catalyst such that the number of the most active sites is maximized.

- Evaluation of different OER mechanisms that may be of importance at layered oxides. During the last years, several water oxidation mechanisms were proposed and studied in detail. Among these, the most important one is the mononuclear mechanism [31], assuming the reaction to proceed through a series of four electrochemical H^+/e^- abstraction steps resulting in the formation of *-OH, *=O and *-OOH intermediates at a single catalytic site.

$$* + H_2O \longrightarrow * - OH + H^+ + e^- \quad (1)$$

$$* - OH \longrightarrow * = O + H^+ + e^- \quad (2)$$

$$* = O + H_2O \longrightarrow * - OOH + H^+ + e^- \quad (3)$$

$$* - OOH \longrightarrow * + O_2 + H^+ + e^- \quad (4)$$

Based on linear scaling relationships, it was proposed that any catalyst proceeding through this mechanism requires a minimal overpotential of at least 0.4 eV [32, 21]. Alternative mechanisms which are not subject to this restriction are the bi-nuclear mechanism [33, 34] and the hydrogen-transfer mechanism [35]. Both mechanisms require — similar to the mononuclear mechanism — the formation of *=O oxo intermediates but differ in the subsequent O–O bond formation step. In case of the bi-nuclear mechanism the critical O–O bond is formed by recombination of two adjacent oxo groups

$$2 * = O \longrightarrow * - O - O - * \quad (5)$$

while the hydrogen-transfer mechanism assumes instantaneous transfer of a hydrogen from *-OOH to an adjacent Hydrogen acceptor site A resulting in the direct formation of O_2 from the TM=O intermediate and H_2O.

$$* = O + H_2O + A \longrightarrow * + O_2 + A - H + H^+ + e^- \quad (6)$$

Depending on the activity and the detailed structure of the surrounding of the active site different mechanisms may be favorable.

- Change of structural properties which may influence the stability of the OER intermediates. Upon formation of single monolayer sheets, the structural properties of the surface layer are no longer determined by the bulk. Accordingly both the lattice parameters as well as the shape of the sheet are no longer fixed. This may allow for changes of the lattice parameters depending on the coverage and type of intermediates covering the surface layer. Additionally the sheet may start to bend resulting in a wavelike shape. This phenomenon is well established for graphene sheets [36]. The influence of both parameters will be studied in detail.

b) *Long-term* On the long term it is aimed at generalizing the understanding obtained from studying experimentally known systems. Based on this more general understanding the following goals shall be achieved.

- Prediction of new electrocatalysts. Based on a general understanding of the mechanistic requirements and the limitations of exfoliated and classical water oxidation catalysts new OER catalysts shall be proposed.
- Refinement of the methods used to address the water oxidation reaction. The state-of-the-art approach in studying electrocatalysts relies solely on the knowledge of the intermediates' stability for assessing the activity of different electrocatalysts and to obtain trends. Additionally solvent effects are generally neglected. While this ansatz indeed proofed to be successful in the past [21, 31, 35], a more detailed picture derived by taking into account explicit solvent effects as well as activation barriers may be beneficial, since it allows to cover effects that may be hidden in the simplified ansatz. In the long term we aim at studying these effects for selected model systems.
- Generalization of the insights obtained for water oxidation to related reactions. Systems considered as potential OER catalysts may also be of interest for closely related reactions such as the oxygen reduction reaction (ORR) which is the main source for the overpotential in fuel cells [37]. Considering the significant similarities between both reactions in terms of reaction mechanisms and intermediates, materials considered as OER catalysts may also be of interest for the ORR.

5.4 Plans and results: May 2014 – Jan 2015

During the first reporting period, a thorough study of the influence of exfoliation is planed and currently performed. This study is performed focusing solely on β-$Co(OH)_2$ as test system. The project is in close collaboration with the experimental efforts of Prof. Xile Hu (EPFL). All calculations are performed at the generalized-gradient approximation (GGA) level of theory using the PBE functional [18] in combination with a Γ-point k-point set and a DZVP basis set. Core electrons are approximated by Goedecker-Teter-Hutter (GTH) pseudopotentials [38, 39, 40].
The current work is aimed at understanding the differences between single- and multilayered β-$Co(OH)_2$ as well as comparing their ac-

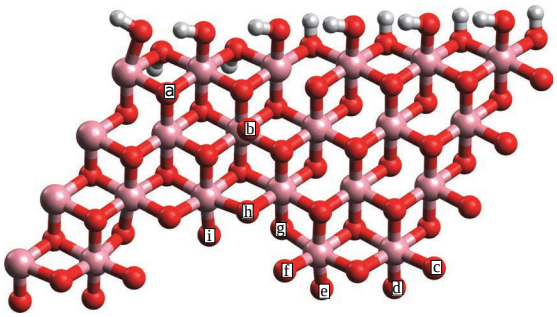


Figure 5: *A schematic overview over potential active sites at fully oxidized exfoliated β-Co(OH)$_2$ is depicted (color code: Co: pink; O: red; H: white).*

tivity with that of different cobalt oxides with a more compact bulk structure such as CoO or Co_3O_4. The latter are added solely for comparison. In order to be able to compare these very different systems, a thorough search for the catalytically active sites at β-Co(OH)$_2$ is required. For this purpose all potentially active positions are taken into account. This includes terrace sites, edge sites and corner sites. In case of terrace sites, two distinct situations need to be considered. First of all an oxo group bound in a threefold position may be active towards O – O bond formation and oxygen evolution (Fig. 5a). This situation is evaluated for an exfoliated β-Co(OH)$_2$ sheet. Assuming a fully oxidized surface, i.e. all Co ions are in a formal +IV oxidation state, the oxidation of a threefold bound Co_3 – OH group to Co_3 – O is found to require a potential of approximately 1.2 eV versus the normal hydrogen electrode (NHE), which is indeed close to the thermodynamic onset potential of the OER. The subsequent formation of the Co_3 – OOH intermediate is however found to require a potential of 2.4 eV versus NHE which corresponds to an overpotential of approximately 1.17 eV rendering these sites inert towards water oxidation. Besides threefold sites, an additional hydroxo group may be adsorbed in a top position (Fig. 5b). This possibility is studied assuming a partially reduced surface where half the threefold oxo species are replace by hydroxo moieties and for a fully oxidized surface. In both cases a reconstruction of the surface is envisaged. Already in case of the partially reduced surface a potential of 1.8 eV versus NHE is needed to split water into hydrogen and an adsorbed OH moiety. This is indeed significantly larger than what would be required to oxidize a Co_3 – OH unit to Co_3 – O. Thus, these sites can also be exclude. Calculations addressing the activity of edge (Fig. 5c and d) as well as corner sites (Fig. 5e to i) are currently performed.

In parallel to the search for active sites, structural effects are studied focusing mainly on the potential influence of the formation of a wavelike structure similar to that found for graphene sheets [36]. Currently the possibility of the appearance of this sheet shape and its relative stability compared to a flat surface is tested. This will be followed by evaluating the activity of potentially active terrace sites assuming the oxidation of water to oxygen via Co – OH, Co = O and Co – OOH intermediates.

5.5 Planned research: Feb 2015 – Jan 2016

Based on the results obtained for pure β-Co(OH)$_2$ the activity of other experimentally studied materials, such as layered NiFe oxides and CoNi oxides, will be addressed. As a first step, it is planned to study exfoliated Ni(OH)$_2$ similarly to the study performed for pure Co(OH)$_2$. While Ni oxides are generally not considered as water oxidation catalysts, these results are required as basis to understand the beneficial effects of Fe. Employing the results obtained for the pure oxides mixed oxides such as CoNi and NiFe will be considered. Here two main situations need to be taken into account: the additive may either act as a spectator influencing only the electronic properties of the active site indirectly or it may be involved directly into the OER mechanism. In case of the latter, it may either act as an oxygen donor in case of a bi-nuclear mechanism [33, 34], or it may provide oxo groups that can be used as hydrogen acceptor sites within the framework of the hydrogen-transfer mechanism [35]. Similarly the additive may also act as active site for water oxidation. All situation will be studied in detail including a thorough study of all potentially active sites similar to that performed for the pure oxides.

5.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Within MARVEL the following collaborations are planned.

- Jürg Hutter, UniZH, HP3: a collaboration concerning the treatment of explicit solvent effects for water oxidation catalysts using a DFTB/MM ansatz is planned.

- Michele Ceriotti, EPFL, HP4: a collaboration concerning interfacing i-PI with electronic structure codes is ongoing. The objective is to perform metadynamics simulations.

- Nicola Marzari, EPFL, VP2: it is planned to assist his group with developing new suitable descriptors based on mechanistic insights.

b) *Outside MARVEL* The project is in close collaboration with the experimental efforts of the group of Prof. Xile Hu at EPFL.

5.7 Personnel

NCCR personnel

- Michael Busch, postdoc, 100%, from October 2014.

6 Novel rare earth perovskites for visible light harvesting applications (Nicola Marzari — EPFL)

6.1 Research summary

With this project, we aim at the identification of novel rare earth perovskites with potential for light harvesting and water splitting applications. We will use *ab initio* quantum mechanical simulations within a density functional theory framework and a screening approach based on fundamental descriptors, including phase stability, band gaps, and redox levels. In a second stage, selected materials will be studied with a more realistic model of the electrochemical environment.

6.2 Research question

The main objective of this investigation is to discover novel rare earth perovskites for light harvesting applications using high-throughput quantum mechanical techniques. The screening is based on a combination of criteria that include phase stability and corrosion in water, a band gap in the visible range to achieve high conversion efficiency, and its position with respect to the redox levels of water for the catalytic activity [41, 42]. These descriptors would select a handful of rare earth perovskites with potential for light harvesting and water splitting applications, to be targeted for experimental synthesis and investigation.

6.3 Scientific goals

a) *Short-term* A preliminary part of this project involves the investigation of the better performing, both in terms of reliability and computational cost, among all the pseudopotentials available, and that will be used in the project. In addition, due to the well-known limitations of density functional theory (DFT) in evaluating band gaps, different approaches will be tested. These include many-body G_0W_0 and the GLLB-SC functional, which has been recently proven to give good results, within 15%, with respect to the eigenvalue-self-consistent GW [43]. As soon as the most appropriate method has been validated, we will start the production runs of rare earth perovskites using first the simplest 5-atom unit cells.

b) *Long-term* The long-term objective is to refine the results of the screening by means of more expensive and reliable methods, and to study possible lower-symmetry structures for candidate materials resulting from the first phase of the project. In addition, the catalytic properties of the most promising compounds will be studied. This involves the study of thermodynamic phase stability, modeling surfaces in ultrahigh vacuum and in a real electrochemical environment, calculating the positions of the band edges when the solvent is present and the reaction barriers for the chemical reactions. Due to the large amount of calculations, the project will take advantage of the AiiDA platform, developed within MARVEL.

6.4 Plans and results: May 2014 – Jan 2015

In the last years, many efforts have been made to find novel materials with high sun-to-hydrogen conversion efficiency in a photoelectrochemical (PEC) cell. During his PhD, Dr. Ivano Castelli has broadly investigated the class of oxide and oxynitride perovskites, in the cubic and layered phases [41, 42, 44]. Lanthanides offer a vast space for computational discovery due to the more limited amount of prior experimental work and because only recently the necessary tools to describe accurately f electrons have been developed and tested. Based on the screening approach already established, we will investigate the lanthanides perovskite class (with general formula ABO_3, where A is a lanthanide element, B is an appropriate cation which can fit into the perovskite structure and makes the total number of electrons and the global valence of the compound equal to zero [45]). It has been shown that a nitrogen substitution has often a beneficial effect on the electronic and catalytic properties of the materials [46] and we will thus investigate oxynitride perovskites as well. In the project we will use two *ab initio* DFT code: Quantum-ESPRESSO [47] for the struc-

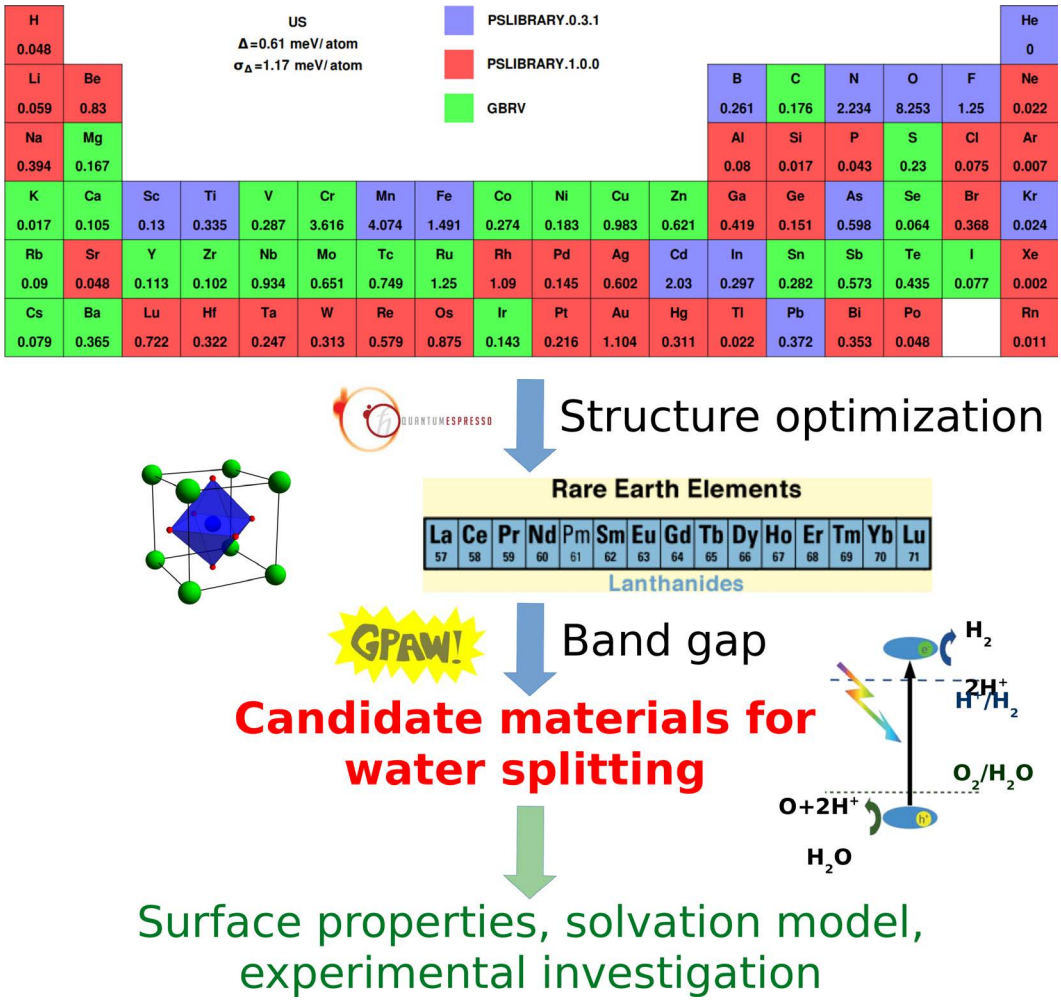


Figure 6: *Summary of the project. Using the optimal set of ultra-soft pseudopotentials, we will investigate the lanthanide perovskites by means of the* Quantum-ESPRESSO *(for the structure optimization) and of the* GPAW *codes (for the calculation of the band gap). A handful of candidates will be suggested for further investigation.*

ture optimization and GPAW [48, 49] for the evaluation of the band gaps. In the last few months, we started investigating systematically the quality of the pseudopotentials used in these codes, i.e. performing the *verification* step that insures that our calculations reproduce accurately all-electron results. Lejaeghere *et al.* [50] have established a quality factor, called Delta, measuring the discrepancies between equations of state for elemental compounds, and provided a set of reference data to test calculations against.

We have systematically tested the more comprehensive libraries available in Quantum-ESPRESSO, and Fig. 6 shows the optimal set of ultra-soft pseudopotentials (including two pseudopotentials libraries, pslibrary.0.3.1 [51] in blue and pslibrary.1.0.0 [52] in red, and the set from Garrity, Bennett, Rabe, and Vanderbilt (GBRV) [53], in green) for each element, and the corresponding Delta factor. This set has a Delta factor of 0.61 meV/atom compared to the WIEN2k all-electron code, representing the state-of-the-art in pseudopotential calculations. Among all elements, oxygen shows a significant error. Since the importance of oxygen in our screening effort, we are now generating and testing other pseudopotentials aiming at an improved description. This part of the project, as well as the next steps, have used extensively the AiiDA framework. In particular, a workflow has been implemented for fast calculation and analysis of each pseudopotential, while another workflow to check the convergence of the cutoffs parameters, using phonons calculations, is currently being developed and tested. All these workflows will be made available for MARVEL and for the broader scientific community. We believe that these simple steps are actually of key importance for all the computational materials design community because they implement and allow a robust evaluation of all the necessary parameters and quality of any pseudopotential used. For

the lanthanides, not included in the figure, we will use the PAW datasets recently generated by Topsakal *et al.* [54], which have been tested and shown to reproduce with excellent agreement the all-electron data.

As sketched in Fig. 6, we will proceed then with the structure optimization of the lanthanide perovskites in 5-atom unit cell. Once that structural and chemical-based rules are applied to speed up the screening process [45], we will end up in all likelihood with a few hundred oxides (ABO_3) and oxynitrides (ABO_2N and $ABON_2$). The stability of the optimized compounds will be then evaluated with respect to solid and dissolved phases (for the corrosion in water) in which the material can separate [41, 55].

6.5 Planned research: Feb 2015 – Jan 2016

In the next months, we will test different methods for evaluating the band gaps, namely G_0W_0 and GLLB-SC, using GPAW. GLLB-SC has been widely used in previous works [41, 42, 44, 43] but it has never been tested for lanthanides. Comparisons between GLLB-SC, G_0W_0 and the available experimental data are thus required. In particular, we will first study rare earth nitrides for their simplicity and, in a second time, a few perovskite examples. The band gaps will be then calculated using the GLLB-SC functional (or G_0W_0 in case of failures of GLLB-SC). The goal will be to narrow our search criteria to identify a reliable set of materials for further investigation, that includes modeling of the surface in a realistic electrochemical environment.

6.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* The project will benefit from two collaborations that are just being established. On one side, together with Prof. Thomas Schmidt and Dr. Emiliana Fabbri (PSI), we are investigating realistic electrochemical environments of oxo-perovskites with potential for running the oxygen evolution reaction. On the other side, with Prof. Thomas Lippert and Dr. Daniele Pergolesi (PSI), we are studying few oxynitride perovskites suggested by our previous screening efforts [41, 42]. Novel oxynitrides will thus be of interest for both the experimental groups.

b) *Outside MARVEL* Rare earths are relatively less explored in materials' design efforts — albeit some of them being not particularly rare or expensive. In addition, *f* electrons have been much less studied within the pseudopotential approximation. Recently, a set of PAW datasets has been generated with remarkable accuracy in structure optimization [54]. Additional efforts have to be made to use these for band gap calculations. This will be done in collaboration with the developers of the GPAW code, in which the GLLB-SC functional is implemented, and in particular with Prof. Karsten Jacobsen's group (DTU, Kgs. Lyngby, DK).

6.7 Personnel

NCCR personnel

- Ivano E. Castelli, postdoc, 50%, from September 2014.

7 Photocatalytic water splitting (Alfredo Pasquarello — EPFL)

7.1 Research summary

- We studied liquid water through isobaric molecular dynamics simulations based on a nonlocal van der Waals density functional, demonstrating an overall improved description compared to standard gradient-corrected density functionals.
- We calculated ionization potentials at surfaces with various advanced electronic structure methods, finding that the hybrid functional rather than the GW results achieved the best overall agreement with the measured values.
- We integrated the use of the CP2K code within the project and started a set benchmark calculations for comparisons with plane-wave codes.

7.2 Research question

Several inorganic materials have already been identified as catalysts for the photochemical splitting of water. In order to properly function as catalysts, such materials need to satisfy constraints as far as the size of their band gap and the positioning of their band edges with respect to reduction and oxidation levels of water are concerned. However, to identify the most efficient catalyst, it is recognized that a wide range of materials needs to be screened. The ultimate goal of this project is to develop a computational tool which is sufficiently accurate

to screen inorganic materials against their catalytic functionality in the photochemical water splitting process.

7.3 Scientific goals

a) *Short-term* In the first stage of the process, we will mainly focus on the constraints on the energy levels. We will align band edges and the water redox levels with respect to a common scale through the explicit modeling of the solid-water interface at the density functional level. However, we will turn to more advanced methods, such as hybrid density functionals and many-body perturbation methods at the GW level, for accurately determining the band gaps and the relative positions between the band edges and the water redox levels building on analogous concepts developed for defect levels.

b) *Long-term* In a further stage of this project, we will focus on the detailed charge transfer processes occurring at the solid-water interface which might lead to additional criteria affecting the overall efficiency of the photochemical splitting of water.

7.4 Plans and results: May 2014 – Jan 2015

In this initial phase, the project has made progress mainly due to the participation of matching personnel. The work has proceeded in two directions. On the one hand, we have set up simulations of liquid water which corresponds to one of the interface components. On the other hand, we have focused on accurate electronic structure methods at the GW level in order to identify the positions of the band edges. Finally, we have started using the CP2K code, which is expected to play a leading role in speeding up the required calculations, an essential aspect in the screening protocol.

a) *Liquid water with van der Waals interactions* Recent isobaric molecular dynamics simulations have shown that the equilibrium density of liquid water achieved with a functional based on a standard generalized gradient approximation is excessively low [56]. This result draws attention to the structure of liquid water which is achieved with such functionals and questions whether the resulting electronic properties would be accurately described. This final aspect is of course highly critical for the purpose of our project. The excessively low density stems from the fact that the gradient-corrected functional excessively favors the formation of the highly directional hydrogen bonds. We therefore considered to include van der Waals (vdW) interactions and turned to a nonlocal functional which carries the advantage of not implying any significant overhead. With this functional, we carried out first-principles molecular dynamics simulations of liquid water in the NpH ensemble at near ambient conditions [14]. First, we tested the performance of the nonlocal vdW functional on the water dimer and on the I_h phase of ice, only a slight increase in the strength of the hydrogen bond in those systems compared to semilocal functionals. In our molecular dynamics simulations of liquid water, we then found that the nonlocal vdW functional provides a noticeably improved description compared to standard semilocal functionals. Not only the equilibrium density was recovered, but also the structural and dynamical properties were found to be much closer to experimental observations. In Fig. 7, the oxygen-oxygen pair correlation function achieved under isobaric conditions with the standard gradient-corrected functional (GGA) and with the nonlocal vdW functional are compared with experiment. The improvement is noticeable and directly relates to the fact that the correct equilibrium density is reproduced in the simulation based on the nonlocal vdW functional. We also focused on the local order and on an analysis of the hydrogen bond network. For instance, the average number of hydrogen bonds per molecule achieved with the nonlocal vdW functional (3.59) agrees much better with the estimate derived from experimental data (3.58) than the gradient-corrected functional (3.73). Finally, also the diffusion coefficient is found to improve going from the gradient-corrected to the nonlocal vdW func-

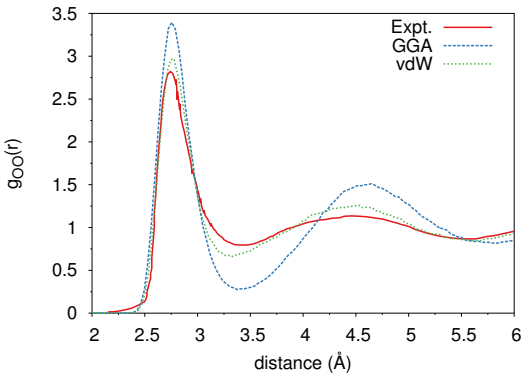


Figure 7: *Oxygen-oxygen pair correlation function as obtained with isobaric molecular dynamics simulations with gradient corrected (GGA) and nonlocal van der Waals (vdW) functionals, compared to experiment (Expt).*

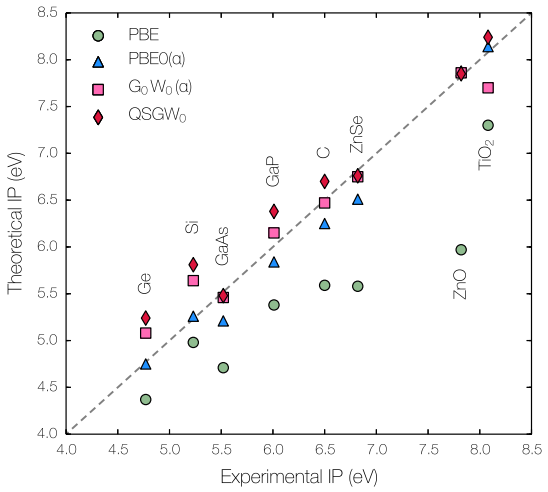


Figure 8: *Calculated vs measured ionization potentials for various surfaces as obtained with a standard nonlocal functional (PBE), a hybrid functional (PBE0(α)), the one-shot G_0W_0 scheme (G_0W_0(α)), and a self-consistent QSGW$_0$ scheme.*

tional. Hence, we conclude that the adopted vdW functional allows for a good description of liquid water at a computational cost equivalent to gradient-corrected density functionals and thus appears particularly attractive for the continuation of the present project.

b) *Band edges through self-consistent GW* We have pursued our studies on the band edge positions of semiconductors and oxides. The scope of this study is to identify accurate electronic-structure methods by which such band edges can reliably be determined. This will ultimately play a critical role in positioning the semiconductor band edges with respect to the water redox levels at our solid-liquid interfaces. To assess the accuracy by which the band edges are determined, we focused on ionization potentials at semiconductor surfaces [15]. We compared the results achieved with a standard nonlocal functional (PBE), a hybrid functional (PBE0(α)), the one-shot G_0W_0 scheme on top of a hybrid functional starting point (G_0W_0(α)), and a self-consistent QSGW$_0$ scheme for various surfaces (Fig. 8). Since we noticed that the comparison is directly affected by the quality by which the band gap is reproduced, we used the fraction α of nonlocal exchange in the hybrid functional as a fitting parameter to reproduce the experimental band gap. From the comparison in Fig. 8, we notice that the hybrid functional results produces the best overall description [15].

c) *Use of the* CP2K *code* For the continuation of the project, it appeared important to be able to integrate within our project the use of the

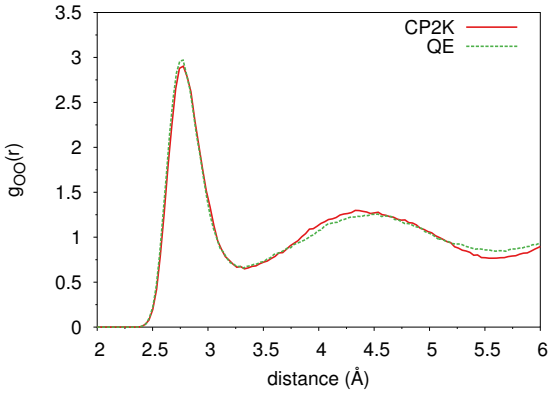


Figure 9: *Oxygen-oxygen pair correlation function as obtained with isobaric molecular dynamics simulations with* CP2K *and* Quantum-ESPRESSO *(QE) under identical conditions.*

CP2K code developed in Jürg Hutter's group. Indeed, this code not only appears to be significantly faster than the plane-wave codes generally used in our group but also to scale better with the number of processors used. It is thus expected that the use of this code can offer a significant speed up both in consumed CPU time and in wall time, an advantage that might become critical in the screening phase. Giacomo Miceli thus spent one week in Jürg Hutter's group to acquire to required expertise for using the CP2K code. Applied to liquid water under the same conditions as discussed in section 7.4 a, we see that the achieved structural properties only show minor differences (Fig. 9), a quite promising result in view of the systematic use of the CP2K code in the screening procedure.

7.5 Planned research: Feb 2015 – Jan 2016

In November 2014, two personnel members hired within the NCCR joined our group: the postdoc Francesco Ambrosio and the PhD student Zhendong Guo. With their arrival, the proper NCCR activity is expected to enter in the operative phase. The work will be subdivided as follows. Francesco Ambrosio is going to focus on the determination of reference electronic levels in water, such the water redox levels. Zhendong Guo will focus on the the construction of solid-liquid interfaces and will focus on their stability against hydroxylation. The research will be complemented by the activity of the matching personnel (Giacomo Miceli and Wei Chen).

a) *Redox levels in liquid water* The first priority is to determine the redox levels of liquid water. We envisage to determine these

electronic levels with two different functionals which yield different structural properties, such as a standard gradient-corrected functional and the nonlocal vdW functional discussed in section 7.4 a. This study is expected to reveal to what extent these electronic levels are sensitive to the molecular arrangement. As secondary target, we would like to determine these redox levels with two different electronic structure codes, i.e. a plane-wave code, such as either Quantum-ESPRESSO or CPMD, and the CP2K code. If successful, this comparison will provide confidence for using CP2K also for electronic properties.

b) *Hydroxylation of Al_2O_3* In electrochemical cells, the structure of the semiconductor/oxide degrades with use. As the negative charge is collected across the semiconductor/oxide, hydroxylation occurs which modifies the electronic properties of the photocatalist. This process critically affects the performance of the photocatalist and needs to be considered before proceeding blindly with the screening analysis. In order to investigate this process, we intend to study the stability of the Al_2O_3/ℓ-H_2O interface against hydroxylation. An OH^- unit will be inserted in the oxide or in the near-interfacial liquid and its effects on the oxide structure will be analyzed. Upon neutralization, the process can be repeated with additional OH^- units. This study allows us to establish a direct link with experimental activity at Empa. Furthermore, it will constitute our first experience in constructing solid/liquid interfaces, which will turn useful in the continuation of the project.

c) *Excitation spectrum of water* To complete the study of the electronic properties of water, we envisage to use GW calculations for achieving the excitation spectrum of liquid water. We are particularly interested to confront the results achieved with different liquid structures as achieved with standard gradient-corrected functional and the nonlocal vdW functional discussed in section 7.4 a. This study involves a statistical treatment of configurations as well as the solution of the Bethe-Salpeter equation in order to account for excitonic effects.

d) *Comparison between different codes* We intend to carry out a more stringent comparison between the CP2K code and a plane-wave code such as Quantum-ESPRESSO or CPMD. We are going to focus on liquid water in the NVE ensemble. We are in contact with Stefan Goedecker, who is an expert in pseudopotential generation. In order to make the comparison more stringent, we intend to use identical pseudopotentials in both types of code. This will allow us to refer the differences solely to the different methodological aspects associated to the solution of the underlying equations. The comparison shall involve physical properties, such as (i) structural properties (pair correlation functions, various local order indices, parameters characterizing the hydrogen bond network), (ii) dynamical properties (diffusion coefficient), and (iii) electronic properties (band edge positions), as well as indices of computational performance.

7.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* We have established contact with the group of Jürg Hutter in order to get familiar with the use of the CP2K code, which is expected to be heavily used in the course of this project. For this purpose, Giacomo Miceli spent one week in Zurich within the group of Jürg Hutter. We have also established preliminary contacts with Stefan Goedecker in relationship with the new pseudopotentials that his group is developing. The use of these pseudopotentials might turn advantageous as they will be designed to be compatible with both CP2K and Quantum-ESPRESSO, two computer codes, which will be heavily relied upon in the course of this project. Finally, we also had an interesting interaction with an experimental group at Empa (Lars Jeurgens, Claudia Cancellieri, and Patrick Schmutz) addressing the stability of oxides in photochemical cells.

b) *Outside MARVEL* In view of using GW calculations for determining the band edges of our semiconductor components, we maintain fruitful interactions with G.-M. Rignanese in Louvain-La Neuve. The group to which G.-M. Rignanese belongs is at the origin of the ABINIT code, which we use for our GW calculations.

7.7 Personnel

NCCR personnel

- Francesco Ambrosio, postdoc, 100%, from November 15, 2014.
- Zhendong Guo, PhD student, 100%, from January 2015.

Matching personnel

- Giacomo Miceli, postdoc, 100%, from May 2014.

- Wei Chen, postdoc, 100%, from May 2014.
- Zhendong Guo, PhD student, 30% from October, 100%, from November to December 2014.

8 Material screening for metal/air batteries (Alessandro Curioni — IBM)

8.1 Research summary

- Solid state conductors.
- Doping.
- Garnets.
- Metal/air batteries.
- Machine learning algorithms.
- Database driven materials design.

8.2 Research question

During the first year of the project we engaged on studying and understanding few fundamental properties of a specific class of solid state inorganic electrolytes. In fact, the characterization and development of new materials for efficient energy storage technologies is one of the most active research areas in which the usage of combined simulation and data-analytics will be a game–changer in the forthcoming years. The zirconium-containing lithium-lanthanum oxide (LLZO) constitutes a promising candidate for use as solid state electrolyte in Li-based batteries, being a material that combines desired properties such as chemical stability, high safety (non-flammable), ease of device fabrication, and low cost. For these reasons, we considered LLZO as a prototypical example to start the investigation of novel material design.

8.3 Scientific goals

a) *Short-term* We plan to test the capabilities of different machine learning algorithms (in direct conjunction with Anatole von Lilienfeld, HP4) on screening materials on the base of specific properties, enlarging the investigation to different types of known solid state Li-ion conductors. The goal in the short term is the set up of a computational platform to investigate in a more automatic way the ion conductivity (for Li and Na) in a large number of different solid state materials, with a strong synergistic interaction with both HP4 and HP5.

b) *Long-term* We plan to do machine learning and data-driven design of novel classes of solid state inorganic electrolytes, based on the knowledge acquired on available data and simulations.

8.4 Plans and results: May 2014 – Jan 2015

We undertook a thorough analysis of the conduction properties of the zirconium-containing lithium-lanthanum oxide (LLZO) [16] and concurrently studied in details the effect of few elemental doping [5] on the stabilization of the different polymorphs.

8.5 Planned research: Feb 2015 – Jan 2016

In the second year, we plan to:

- model the ion-transport process in solid state electrolytes;
- select and evaluate good descriptors that well characterize the transport properties of these materials;
- generate a dense database of known structures to be used in the machine learning process.

8.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* We plan to test of the capabilities of different machine learning algorithms in direct conjunction with Anatole von Lilienfeld, HP4, and strong interaction with HP5 for the design of novel data-driven material design approaches.

8.7 Personnel

NCCR personnel

- Ivano Tavernelli, research staff member, 100%, from November 2014.

Matching personnel

- Teodoro Laino, research staff member, 75%, from May 2014.

9 The industrial way for post-combustion capture of CO_2: optimizing the solvent (Wanda Andreoni — EPFL)

9.1 Research summary

Post-combustion carbon capture mainly employs solvent wet scrubbing, primarily using alkanolamine aqueous solutions as chemical absorbents. The crucial need for more efficient solvents has promoted a widespread empirical effort toward process and material optimization. We intend to use computer simulations to explore the properties of different solutions and especially to quantitatively characterize the chemical reactions involved in the capture and release of CO_2. In this way we should obtain useful information for a rational design of novel solvents and/or improved processes. Our proposal, namely solvent screening via advanced *ab initio* simulations, is unprecedented.

9.2 Research question

The disastrous impact of anthropogenic carbon dioxide (CO_2) emissions on the environment is one of the main concerns of our era. Among the options under examination for the reduction of these emissions and the mitigation of global climate change, post-combustion capture (PCC) and subsequent sequestration/storage is considered one of key importance for a viable and near-term solution [57, 58, 59]. The most mature technology for CO_2 capture, currently in use in the chemical industry, exploits a cyclic process, in which CO_2 is selectively and reversibly absorbed in an amine (aqueous) solution.

Alkanolamine aqueous solutions are most frequent in industrial applications and primarily monoethanolamine (MEA) at 30 wt% concentration. MEA solutions exhibit several unique advantages, e.g. high reactivity with CO_2, relatively low solvent cost and special ease of reclamation, but are known to be corrosive and to require high regeneration energy. These disadvantages still inhibit its large-scale deployment in power plants. A large empirical effort is ongoing worldwide to identify and assess novel solvents, primarily meant to reduce the high energy penalty required for amine regeneration (which is responsible for more than 90% of the costs) and decrease the rate of amine degradation due to oxidation, corrosion and salt formation (e.g. [60, 61, 62, 63, 64, 65]).

Modified amines and salt additions have been tried, on the basis of phenomenological models of the reaction kinetics. An extended screening based on measurements of the CO_2 absorption capacity and initial absorption rates was performed for 76 aqueous amine solutions [60]. Out of these, seven emerged as exhibiting outstanding absorption capacities, and some also showed initial absorption rates comparable to the industry standard monoethanolamine (MEA). In the attempt to rationalize this behavior, some common unusual structural features have been identified in these special amines, whose role however remains unexplained. More recently high-throughput screening was applied to a set of 50 monoamines aqueous solutions [61], aiming at assessing their global performance, namely not only the absorption capacity but also the regeneration capacity. The results reveal no correlation between these two properties and are still awaiting an interpretation.

There is a general need for a fundamental understanding of the mechanisms involved in these processes and for a characterization of the relative importance of the different steps in determining the bottlenecks. The motivation of this project stems from this awareness and the not uncommon belief that only large-scale accurate simulations can help to obtain deep insight into the relative role of the fundamental chemical reactions, and thus lead to new ideas about improvements or alternative routes.

Using *ab initio* simulations, we have recently studied the chemical reactions leading to CO_2 capture and release in MEA solutions (Fig. 10) [66, 17, 67]. In particular, it was possible to quantitatively characterize the entire cycle and to propose a novel and well defined pathway that is consistent with available experimental data.

For each solvent, the fundamental questions to be tackled refer to the nature and dynamics of the reactions accompanying both the uptake of CO_2 via absorption and its release. In particular, our simulations will uncover which are the steps dominating the activation barriers, how the characteristics of the solution influence these reactions, and in particular the role of water and of concentration.

The difficulty to answer any of such questions with targeted experiments is a severe obstacle for the application of rational design to the search of alternative solvents. Hence the need for computer simulations based on robust approaches.

9.3 Scientific goals

The goal of our work is fourfold: from our simulations we can identify and characterize (i) the

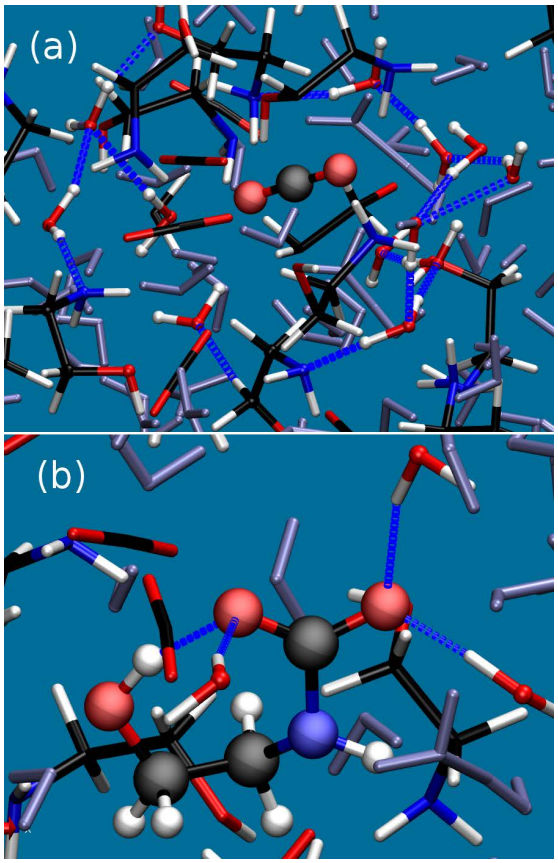


Figure 10: *CO_2 solvated (a) and captured (in a carbamate) (b) in MEA solution. From [17].*

amine or ammonia aqueous solution at equilibrium, (ii) the main chemical reactions leading to CO_2 uptake, (iii) the peculiar response of the different reaction products in solution to diverse probes, and (iv) the main chemical reactions leading to CO_2 release in solution.
The results allow us to compare the mechanisms and the relative (multiple) reaction barriers in different solvents and under different temperature conditions, and to provide the needed ingredients for speciation of the components of the solution before and after absorption from, e.g., vibrational spectra.

9.4 Planned research: Feb 2015 – Jan 2016

We intend to study CO_2 capture in a primary amine other than MEA — chosen out of the outstanding candidates identified in [60] — for a direct comparison with MEA and — later — in ammonia solutions. The latter appear to provide a few advantages relative to MEA, especially higher loading capacity and lower thermal energy for the regeneration process from CO_2-rich solutions.
We will focus on the absorption process of CO_2 with the specific goal to investigate the formation of carbamates versus bicarbonates as primary products, and to explore pathways for carbamate conversion into bicarbonates. Indeed, uptake into carbamates is relatively fast for primary amines and ammonia but is also an undesirable part of the interaction, as it decreases the cyclic capacity of the amine solution (e.g. [68]).
Our approach will use the same methods as for MEA, namely DFT-based molecular dynamics empowered by enhanced sampling. As initial steps, it will include a series of test calculations to validate the DFT scheme and classical molecular dynamics of the ammonia solution with solvated CO_2, which requires the tuning of the force-field.

9.5 Synergies with other computational and experimental efforts

a) *Within MARVEL* We intend to interact with Alfredo Pasquarello on the application of DFT — with different approximations for the exchange-correlation functional — to simulate water. Depending on the results, we may eventually connect to the effort of Berend Smit on CO_2 capture.

b) *Outside MARVEL* We intend to collaborate with experimentalists in ETHZ working on the characterization and optimization of solvents for CO_2 capture (group of Prof. Mazzotti).

9.6 Personnel

NCCR personnel

- Changru Ma, postdoc, 100%, from March 2015.

Matching personnel

- Changru Ma, postdoc, 100%, from May 2014 to February 2015.

MARVEL-related publications

List of publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.

- [1] S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B. F. E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M. K. Nazeeruddin, and M. Grätzel, *Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers*, Nature Chemistry **6**, 242 (2014).
- [2] K. Voïtchovsky, N. Ashari-Astani, I. Tavernelli, N. Tetreault, U. Röthlisberger, F. Stellacci, M. Grätzel, and H. A. Harms, *In-situ mapping of the molecular arrangement of amphiphilic dye molecules at the TiO_2 surface of dye sensitized solar cells, submitted* (2014).

[3] J. Cai, C. A. Pignedoli, L. Talirz, P. Ruffieux, H. Söde, L. Liang, V. Meunier, R. Berger, R. Li, X. Feng, K. Müllen, and R. Fasel, *Graphene nanoribbon heterojunctions*, Nature Nanotechnology **9**, 896 (2014).

[4] R. Denk, M. Hohage, P. Zeppenfeld, J. Cai, C. A. Pignedoli, H. Söde, R. Fasel, X. Feng, K. Müllen, S. Wang, D. Prezzi, A. Ferretti, A. Ruini, E. Molinari, and P. Ruffieux, *Exciton-dominated optical response of ultra-narrow graphene nanoribbons*, Nature Communications **5**, 4253 (2014).

[5] K. Meier, T. Laino, and A. Curioni, *Why Al doping stabilizes the cubic phase of garnet-like LLZO - Conclusive answers from an exhaustive computational study, submitted* (2014).

[6] S. Meloni, G. Palermo, N. Ashari Astani, B. F. E. Curchod, M. Graetzel, and U. Roethlisberger, *Valence and conduction bands engineering in halide perovskites for solar cell applications*, arXiv:1412.3659 (2014).

[7] D. J. Adams, S. Chappellet, F. Lincker, M. Ibn-Elhaj, B. Watts, M. Iannuzzi, D. S. Jung, C. A. Pignedoli, and D. Passerone, *Identifying Photoreaction Products in Cinnamate-Based Photoalignment Materials*, The Journal of Physical Chemistry C **118**, 15422 (2014).

[8] R. Gaspari, R. Erni, Y. Arroyo, M. Parlinska-Wojtan, J. Dshemuchadse, C. A. Pignedoli, D. Passerone, P. Schmutz, and A. Beni, *Real space crystallography of a complex metallic alloy: high-angle annular dark-field scanning transmission electron microscopy of o-Al_4(Cr,Fe)*, Journal of Applied Crystallography **47**, 1026 (2014).

[9] J. Prinz, R. Gaspari, Q. S. Stöckl, P. Gille, M. Armbrüster, H. Brune, O. Gröning, C. A. Pignedoli, D. Passerone, and R. Widmer, *Ensemble Effect Evidenced by CO Adsorption on the 3-Fold PdGa Surfaces*, The Journal of Physical Chemistry C **118**, 12260 (2014).

[10] R. Jaafar, C. A. Pignedoli, G. Bussi, K. Aït-Mansour, O. Groening, T. Amaya, T. Hirao, R. Fasel, and P. Ruffieux, *Bowl Inversion of Surface-Adsorbed Sumanene*, Journal of the American Chemical Society **136**, 13666 (2014).

[11] H. Söde, L. Talirz, O. Gröning, C. A. Pignedoli, R. Berger, X. Feng, K. Müllen, R. Fasel, and P. Ruffieux, *Electronic band dispersion of graphene nanoribbons via Fourier-transformed scanning tunneling spectroscopy*, Physical Review B **91**, 045429 (2015).

[12] L. Talirz, P. Shinde, D. Passerone, and C. A. Pignedoli, *Synthesis of Atomically Precise Graphene-Based Nanostructures: a Simulation Point of View, submitted* (2014).

[13] M. Gibertini, G. Pizzi, and N. Marzari, *Engineering polar discontinuities in honeycomb lattices*, Nature Communications **5**, 5157 (2014).

[14] G. Miceli, S. de Gironcoli, and A. Pasquarello, *Isobaric first-principles molecular dynamics of liquid water with nonlocal van der Waals interactions*, The Journal of Chemical Physics **142**, 034501 (2015).

[15] W. Chen and A. Pasquarello, *Band-edge positions in GW: Effects of starting point and self-consistency*, Physical Review B **90**, 165133 (2014).

[16] K. Meier, T. Laino, and A. Curioni, *Solid-State Electrolytes: Revealing the Machanisms of Li-Ion Conduction in Tetragonal and Cubic LLZO by First-Principles Calculations*, The Journal of Physical Chemistry C **118**, 6668 (2014).

Other references

[17] C. Ma, F. Pietrucci, and W. Andreoni, *Capturing CO_2 in Monoethanolamine (MEA) Aqueous Solutions: Fingerprints of Carbamate Formation Assessed with First-Principles Simulations*, The Journal of Physical Chemistry Letters **5**, 1672 (2014).

[18] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Physical Review Letters **77**, 3865 (1996).

[19] M. Amsler and S. Goedecker, *Crystal structure prediction using the minima hopping method*, The Journal of Chemical Physics **133**, 224104 (2010).

[20] S. Goedecker, *Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems*, The Journal of Chemical Physics **120**, 9911 (2004).

[21] I. C. Man, H. Su, F. Calle-Vallejo, H. A. Hansen, N. G. Martînez, J. I. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov, and J. Rossmeisl, *Universality in Oxygen Evolution Electrocatalysis on Oxide Surfaces*, ChemCatChem **3**, 1159 (2011).

[22] S. Trasatti, *Electrocatalysis: understanding the success of DSA*, Electrochimica Acta **45**, 2377 (2000).

[23] Y. Umena, K. Kawakami, J. Shen, and N. Kamiya, *Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å*, Nature **473**, 55 (2011).

[24] M. Hamdani, R. N. Singh, and P. Chartier, *Co_3O_4 and Co-Based Spinel Oxides Bifunctional Oxygen Electrodes*, International Journal of Electrochemcal Science **5**, 556 (2010).

[25] F. Song and X. Hu, *Exfoliation of layered double hydroxides for enhanced oxygen evolution catalysis*, Nature Communications **5**, 4477 (2014).

[26] M. M. Najafpour, T. Ehrenberg, M. Wiechen, and P. Kurz, *Calcium-Mangan (III)-Oxide ($CaMn_2O_4 \cdot xH_2O$) als biomimetische Katalysatoren für die Sauerstoffbildung*, Angewandte Chemie **122**, 2281 (2010).

[27] M. Busch, E. Ahlberg, and I. Panas, *Validation of binuclear descriptor for mixed transition metal oxide supported electrocatalytic water oxidation*, Catalysis Today **202**, 114 (2013).

[28] F. Song and X. Hu, *Ultrathin Cobalt-Manganese Layered Double Hydroxide Is an Efficient Oxygen-Evolution-Catalyst*, Journal of the American Chemical Society **136**, 16481 (2014).

[29] B. Yeo and A. Bell, *Enhanced Activity of Gold-Supported Cobalt Oxide for the Electrochemical Evolution of Oxygen*, Journal of the American Chemical Society **133**, 5587 (2011).

[30] A. M. Mohammad, M. I. Awad, M. S. El-Deab, T. Okajima, and T. Ohsaka, *Electrocatalysis by nanoparticles: Optimization of the loading level and operating pH for the oxygen evolution at crystallographically oriented manganese oxide nanorods modified electrodes*, Electrochimica Acta **53**, 4351 (2008).

[31] J. Rossmeisl, A. Logadottir, and J. K. Nørskov, *Electrolysis of water on (oxidized) metal surfaces*, Chemical Physics **319**, 178 (2005).

[32] M. T. M. Koper, *Thermodynamic theory of multi-electron transfer reactions: Implications for electrocatalysis*, Journal of Electroanalytical Chemistry **660**, 254 (2011).

[33] M. Busch, E. Ahlberg, and I. Panas, *Electrocatalytic oxygen evolution from water on a Mn(III-V) dimer model catalyst - A DFT perspective*, Physical Chemistry Chemical Physics **13**, 15069 (2011).

[34] J. O. Bockris and T. Otagawa, *Mechanism of Oxygen Evolution on Perovskites*, The Journal of Physical Chemistry **87**, 29601 (1983).

[35] N. Bendtsen Halck, V. Petrykin, P. Krtil, and J. Rossmeisl, *Beyond the Volcano Limitations in Electrocatalysis - Oxygen Evolution Reaction*, Physical Chemistry Chemical Physics **16**, 13682 (2014).

[36] J. Meyer, A. Geim, M. Katsnelson, K. Novoselov, T. Booth, and S. Roth, *The structure of suspended graphene sheets*, Science **446**, 60 (2007).

[37] S. Park, Y. Shao, J. Liu, and Y. Wang, *Oxygen electrocatalysts for water electrolyzers and reversible fuel cells: status and perspective*, Energy & Environmental Science **5**, 9331 (2012).

[38] S. Goedecker, M. Teter, and J. Hutter, *Separable dual-space Gaussian pseudopotentials*, Physical Review B **54**, 1703 (1996).

[39] C. Hartwigsen, S. Goedecker, and J. Hutter, *Relativistic separable dual-space Gaussian pseudopotentials from H to Rn*, Physical Review B **58**, 3641 (1998).

[40] M. Krack, *Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals*, Theoretical Chemistry Accounts **114**, 145 (2005).

[41] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, *Computational screening of perovskite metal oxides for optimal solar light capture*, Energy & Environmental Science **5**, 5814 (2012).

[42] I. E. Castelli, D. D. Landis, K. S. Thygesen, S. Dahl, I. Chorkendorff, T. F. Jaramillo, and K. W. Jacobsen, *New cubic perovskites for one- and two-photon water splitting using the computational materials repository*, Energy & Environmental Science **5**, 9034 (2012).

[43] I. E. Castelli, F. Hüser, M. Pandey, H. Li, K. S. Thygesen, B. Seger, A. Jain, K. A. Persson, G. Ceder, and K. W. Jacobsen, *New Light-Harvesting Materials Using Accurate and Efficient Bandgap Calculations*, Advanced Energy Materials **5**, 1400915 (2015).

[44] I. E. Castelli, J. M. García-Lastra, F. Hüser, K. S. Thygesen, and K. W. Jacobsen, *Stability and bandgaps of layered perovskites for one- and two-photon water splitting*, New Journal of Physics **15**, 105026 (2013).

[45] I. E. Castelli and K. W. Jacobsen, *Designing rules and probabilistic weighting for fast materials discovery in the Perovskite structure*, Modelling and Simulation in Materials Science and Engineering **22**, 055007 (2014).

[46] R. Aguiar, D. Logvinovich, A. Weidenkaff, A. Rachel, A. Reller, and S. G. Ebbinghaus, *The vast colour spectrum of ternary metal oxynitride pigments*, Dyes and Pigments **76**, 70 (2008).

[47] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, *Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials*, Journal of Physics: Condensed Matter **21**, 395502 (2009).

[48] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, *Real-space grid implementation of the projector augmented wave method*, Physical Review B **71**, 35109 (2005).

[49] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Hakkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiotz, K. S. Thygesen, and K. W. Jacobsen, *Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method*, Journal of Physics: Condensed Matter **22**, 253202 (2010).

[50] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, *Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals*, Critical Reviews in Solid State and Materials Sciences **39**, 1 (2014).

[51] A. Dal Corso, *Projector augmented wave method with spin-orbit coupling: Applications to simple solids and zincblende-type semiconductors*, Physical Review B **86**, 085135 (2012).

[52] A. Dal Corso, *Pseudopotentials periodic table: From H to Pu*, Computational Materials Science **95**, 337 (2014).

[53] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, *Pseudopotentials for high-throughput DFT calculations*, Computational Materials Science **81**, 446 (2014).

[54] M. Topsakal and R. Wentzcovitch, *Accurate projected augmented wave (PAW) datasets for rare-earth elements (RE = La – Lu)*, Computational Materials Science **95**, 263 (2014).

[55] I. E. Castelli, K. S. Thygesen, and K. W. Jacobsen, *Calculated Pourbaix Diagrams of Cubic Perovskites for Water Splitting: Stability Against Corrosion*, Topics in Catalysis **57**, 265 (2014).

[56] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, *Isobaric-Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions*, The Journal of Physical Chemistry B **113**, 11959 (2009).

[57] J. Wilcox, *Carbon Capture* (Springer, New York, 2012).

[58] G. T. Rochelle, *Amine Scrubbing for CO_2 Capture*, Science **325**, 1652 (2009).

[59] Ö. Yildirim, A. A. Kiss, N. Hüser, K. Lessmann, and E. Y. Kenig, *Reactive Absorption in Chemical Process Industry: A Review on Current Activities*, Chemical Engineering Journal **213**, 371 (2012).

[60] G. Puxty, R. Rowland, A. Allport, Q. Yang, M. Bown, R. Burns, M. Maeder, and M. Attalla, *Carbon Dioxide Postcombustion Capture: a Novel Screening Study of the Carbon Dioxide Absorption Performance of 76 Amines*, Environmental Science & Technology **43**, 6427 (2009).

[61] F. Porcheron, A. Gibert, F. Mougin, and A. Wender, *High Throughput Screening of CO_2 Solubility in Aqueous Monoamines Solutions*, Environmental Science & Technology **45**, 2486 (2011).

[62] F. Bougie and M. C. Iliuta, *Sterically Hindered Amine-Based Absorbents for the Removal of CO_2 from Gas Streams*, Journal of Chemical & Engineering Data **57**, 635 (2012).

[63] F. P. Gordesli, C. S. Ume, and E. Alper, *Mechanism and Kinetics of Carbon Dioxide Capture Using Activated 2-Amino-2-methyl-1,3-propanediol*, International Journal of Chemical Kinetics **45**, 566 (2013).

[64] W. Conway, D. Fernandes, Y. Beyad, R. Burns, G. Lawrance, G. Puxty, and M. Maeder, *Reactions of CO_2 with Aqueous Piperazine Solutions: Formation and Decomposition of Mono- and Dicarbamic Acids/Carbamates of Piperazine at 25.0°C*, The Journal of Physical Chemistry A **117**, 806 (2013).

[65] W. Conway, X. Wang, D. Fernandes, R. Burns, G. Lawrance, G. Puxty, and M. Maeder, *Toward the Understanding of Chemical Absorption Processes for Post-Combustion Capture of Carbon Dioxide: Electronic and Steric Considerations from the Kinetics of Reactions of CO_2(aq) with Sterically hindered Amines*, Environmental Science & Technology **47**, 1963 (2013).

[66] C. A. Guido, F. Pietrucci, G. A. Gallet, and W. Andreoni, *The Fate of a Zwitterion in Water from* ab Initio *Molecular Dynamics: Monoethanolamine (MEA)-CO_2*, Journal of Chemical Theory and Computation **9**, 28 (2013).

[67] C. Ma, F. Pietrucci, and W. Andreoni, *Capture and Release of CO_2 in Monoethanolamine Aqueous Solutions: The "Whole Story" Revealed with First-Principles Reaction Dynamics*, *submitted* (2014).

[68] X. Wang, W. Conway, D. Fernandes, G. Lawrance, R. Burns, G. Puxty, and M. Maeder, *Kinetics of the Reversible Reaction of CO_2(aq) with Ammonia in Aqueous Solution*, The Journal of Physical Chemistry A **115**, 6405 (2013).

Horizontal Project 3 — HP3 — Advanced Quantum Simulations

Project leader: Jürg Hutter (UZH)

Participating members: Jürg Hutter (UZH), Matthias Troyer (ETHZ), Philipp Werner (UniFR), Joost VandeVondele (ETHZ)

Summary and highlights: In the Horizontal Project 3 (HP3) on advanced quantum simulations, we develop new theories and algorithms for systems with a complex electronic structure. These are strongly correlated systems with novel and interesting quantum states that are of interest, for example, in the field of quantum information and computing. The other systems of interest are electrochemical solid/liquid interfaces at finite temperature. New methods developed include continuous time quantum Monte Carlo (QMC), self-consistent GW+DMFT, and non-local correlation (MP2 and RPA) based density functionals.

General view of the project

Major research questions

For a the class of strongly correlated systems, that give raise to extremely interesting but scarcely known states of matter, the most commonly used methods for materials based on density functional theory (DFT) fail. Recently, a quantum Monte Carlo (QMC) algorithm that is at the same time free from systematic errors and optimally performing has become available. This continuous time approach (CTQMC) has to be further developed to allow to study realistic systems. We aim to predict observables adequate for experimental verification and measures of interest in the field of quantum information and computation. The long-term goal is the realization of a scheme for studying strongly correlated states of matter with satisfactorily predictive power.

Another method to address strongly correlated systems is the combination of density functional based *ab initio* calculations with dynamical mean field theory (LDA+DMFT). Replacing the LDA input by GW allows to put the theory (GW+DMFT) onto a firm diagrammatic basis and avoids conceptual problems related to LDA+DMFT. It has the potential to become a true *ab initio* method for strongly correlated compounds. At present, self-consistent GW+DMFT calculations have been implemented only for very simple models. The goal is to implement the GW+DMFT approach for real multiband materials.

The useful work that can be extracted from an electrochemical system is described by the free energy of electrons. Being able to compute, rationalize, and manipulate this quantity in complex chemical systems is important for a sustainable energy future. The capability to accurately compute this quantity in realistic systems is still limited. The challenge is formidable, as it requires to describe finite temperature systems, interfaces between ordered solids and disordered aqueous systems, and the balance between localized and delocalized electrons. It is essential to obtain a qualitatively correct picture for such systems related to research planed in Vertical Project 2. The search for density functionals that yield reliable results for a wide range of such systems has lead to methods that are based on an explicit treatment of correlation. Functionals based on the random phase approximation (RPA) have the potential to provide both, high accuracy and general applicability. We are looking for algorithms to calculate the energy at these levels of theory while still providing a computational efficiency that allows application to large and complex systems.

Short- and long-term goals

Short-term

- Extend CTQMC for long-range Coulomb forces, enabling its use in the simulation of realistic materials, and develop a general scheme to calculate fidelity in CTQMC.
- Using an imaginary frequency GW code, develop a GW+DMFT self-consistency loop for multi-orbital systems and perform self-consistent GW+DMFT calculations within a small subspace of orbitals corresponding to the lowest-energy bands.

- Improve the auxiliary density matrix methods (ADMM), make them more robust and applicable to a wider range of problems. Apply the computational frameworks of ADMM and resolution-of-the-identity (RI) to GW-based methods. Add features to RPA and MP2 required for the description of electrochemical systems, such as molecular dynamics (MD) capability and unrestricted calculations.

Long-term

- Apply the CTQMC scheme with long-range forces for the *ab initio* simulation of large scale molecules and realistic materials. Implement the extraction of phenomenological parameters from CTQMC to describe the dynamical responses of strongly correlated and low-temperature phases and apply quantum entanglement and fidelity measures to quantify the correlation and to characterize novel ground state phase diagrams.

- Extend the self-consistent GW+DMFT calculations to a larger subspace of orbitals, covering an energy window of $\sim$ 20 eV, while restricting the impurity calculation to a few strongly correlated bands.

- Implement more advanced forms of RPA with higher-order corrections that improve the accuracy of the electronic structure method. Increase the performance of the methods to enable simulations of large systems, and to enable sampling.

- Develop a solver for quasi-particle equations defined by orbital-dependent and non-local potentials. Transform this into a general scheme for property calculations. Develop a solver for quasi-particle equations with frequency-dependent self-energies. Explore a wide range of approximations for the self-energy (G_0W_0 to coupled cluster methods).

First achievements

We have recently developed a new CTQMC algorithm for fermions which surpasses the performance scaling of any previously known continuous time method and matches that of the best conventional ones. For the first time we can combine optimal performance with accuracy and flexibility of CTQMC. The performance gain makes the goal of simulating materials at finite and zero temperature directly using an unbiased scheme and extracting new important measures that were previously out of reach for other simulations.

As a warm-up problem for GW+DMFT simulations, and as a continuation of a long-term effort to incorporate the dynamically screened nature of the Coulomb interaction into *ab initio* simulations of correlated materials, we have performed simulations of a prototypical high-T_c material, La_2CuO_4, using both the *ab initio* band structure and the frequency-dependent interaction parameters computed with the constrained random phase approximation. This study showed that the dynamical screening effect in cuprates is very strong, and that it is important to treat this effect in order to obtain a consistent description of the low-energy electronic structure.

We have implemented alternative methods to constrain the auxiliary density matrix in ADMM methods and have interfaced them to work together with MP2 and RPA. A framework to calculate nuclear gradients for post-HF methods has been established and the MP2 gradient successfully implemented and tested. The post-HF methods have been optimized for massively parallel computing and the use of CPU accelerators [1].

Next steps

We aim at extracting phenomenological parameters from CTQMC, that can be used both for the *ab initio* simulation of realistic systems, and to calculate the large-scale dynamical responses of strongly correlated systems using effective theories such as the time-dependent Ginzburg-Landau theory employed to describe superconductors. We will use CTQMC to study quantum phase transitions in systems involving Dirac fermions (like graphene) and multi-flavor Dirac fermions. Further applications include using entanglement entropy and fidelity to search for and characterize exotic quantum states without a local order parameter description, such as the topological Mott insulators.

The imaginary-axis GW code is currently being developed by an outside collaborator, Fredrik Nilsson, in the group of Ferdi Aryasetiawan (Lund University). At the same time, a multi-band GW+DMFT code for model systems is being developed at UniFR. The plans are to complete the tests for the 3-band $SrVO_3$ model and to automatize the exchange of data between the GW code and the (extended) DMFT self-consistency loop. We will test the effect of the long-range interaction terms in a model context, and try to identify appropriate simplified

schemes with a finite interaction range.
The implementation of GW-based methods into the CP2K code will be started. A first milestone will be the G_0W_0 method working. We will then explore possible connections to related developments within the quantum chemistry community. Another development will be based on methods defined by a non-local, but frequency independent, potential. We will develop a general framework for such methods based on a solver for the quasi-particle equation. The implementation of forces in unrestricted MP2 will be done and we will work on extending the RPA code to deal with more advanced correlation schemes, such as, e.g., AXK or SOSEX.

Research contributions to the overall goals of the NCCR and to the existing literature

The class of strongly correlated systems is at the center of interest in Vertical Project 1. These systems give raise to extremely interesting states of matter. However, the most commonly used methods based on density functional theory fail. In HP3, two lines of research are followed to develop methods that allow for a first-principles determination of properties of strongly correlated systems. One method is the continuous time approach to quantum Monte Carlo (CTQMC). It is at the same time free from systematic errors and optimally performing and promises to allow strongly correlated states of matter with predictive power. Another promising method is the combination of the GW method with DMFT in a self-consistent way. This method goes beyond and overcomes many of the conceptual problems of the currently used LDA+DMFT. Developing GW+DMFT into a tool that can be applied to real materials is the main goal.
The systems studied and materials developed in VP2 (solar cells, water splitting) have a complex heterogenous interface in common. The calculation of a figure of merit for such systems requires a balanced description of the free energy of electrons at finite temperature. Methods to simulate such systems at a sufficient level of accuracy have to be based on modern density functionals. Such functionals make use of non-local correlation methods and require extensive computational resources. In HP3 algorithms and computer programs are developed that allow to apply such new methods to complex systems at finite temperature.
All developments in this project are state-of-the-art and at the forefront of current research.

Collaborative components

Collaborations within HP3 are expected between the strong correlation groups (Troyer and Werner) and are fully established with the DFT groups (Hutter and VandeVondele). Further collaborations are planned between the Hutter and Marzari groups on orbital-dependent functionals. The strong correlation groups have connections to VP1, notably the Troyer and Spaldin groups are closely collaborating. The work on DFT methods will be beneficial for the studies proposed in VP2, e.g. dye-sensitized solar cells, and photo-catalytic water splitting systems (groups of Ursula Röthlisberger and Alfredo Pasquarello). We also expect to contribute to other projects, e.g. molecular packing and nucleation (Goedecker). The code developments within DFT depend strongly on the basic high-performance computing modules available within CP2K. There is a co-design effort, supported by the Platform for Advanced Scientific Computing (PASC), to improve these basic libraries. Besides the groups mentioned above, this effort includes contributions from CSCS, Cray, and NVIDIA. Guidance with the use of all new methods can be provided to MARVEL team members.

1 Algorithm development for correlated fermion simulations (Matthias Troyer — ETHZ)

1.1 Research summary

- Develop efficient continuous time quantum Monte Carlo methods for lattice fermion simulations.
- Calculations of topological invariances and quantum information measures for correlated materials.

1.2 Research question

Despite being fundamentally quantum in nature, materials and other large atomic systems are traditionally and most commonly studied with approximate methods such as the local density approximation (LDA) and density functional theory (DFT). However such methods are not able to satisfactorily simulate the wide class of *strongly correlated systems* where the quantum phenomena play an important role and give raise to extremely interesting but

scarcely known states of matter. While quantum Monte Carlo (QMC) algorithms would be the most obvious choice of method for studying such models, it wasn't until recently that an algorithm that is at the same time free from systematic errors (the so-called *continuous time approach*) and optimally performing has become available.
Our aim is to develop new methods to describe the strongly correlated phases of materials without resorting to approximations and assumptions on the nature of the systems under study. Building on recent progress we aim at expanding the field of applicability of *exact methods* beyond the unrealistically simple or small. We also aim at extending the current state-of-the-art algorithms to predict observables adequate for experimental verification and measures of interest in the field of quantum information and computation. The long-term goal is the realization of a scheme for studying strongly correlated states of matter with satisfactorily predictive power.
Besides their efficiency and accuracy, continuous time QMC methods (CTQMC) are very flexible and allow better access to interesting observables. For example, one can directly extract the topological index of correlated topological insulators and related quantum information measures such as entanglement entropy and the fidelity. These tools are indispensable to study quantum phases and phase transitions of complex materials.

1.3 Scientific goals

a) *Short-term*

- Extend the efficient CTQMC for long-range Coulomb force, enabling its use in the simulation of realistic materials and molecules
- Develop a general scheme to calculate fidelity in CTQMC

b) *Long-term*

- Apply the CTQMC scheme with long-range forces for the *ab initio* simulation of large scale molecules and realistic materials.
- Implement the extraction of phenomenological parameters to describe the dynamical responses of strongly correlated and low-temperature phases
- Apply quantum entanglement and fidelity measures to correlated materials to quantify the correlation and to characterize novel ground state phase diagrams.

1.4 Plans and results: May 2014 – Jan 2015

We have recently developed a new CTQMC algorithm for fermions which surpasses the performance scaling of any previously known continuous time method and matches that of the best conventional ones [2]. For the first time we can combine optimal performance with accuracy and flexibility of CTQMC. The performance gain makes the goal of simulating materials at finite and zero temperature directly using an unbiased scheme and extracting new important measures that were previously out of reach for other simulations.
We have investigated the origin of the negative sign problem in QMC simulations and could relate it to topological invariants of the configurations [3].
We have realized a new scheme for calculating the Renyi entanglement entropy of interacting fermions using our CTQMC [4]. The algorithm only samples the interaction correction of the entanglement entropy, which by design ensures the efficient calculation of weakly interacting systems. Combined with Monte Carlo reweighting, the algorithm also performs well for systems with strong interactions.
We have developed a new QMC algorithm for dynamical mean field theory (DMFT) simulations, combining QMC updates with matrix product state calculations to evaluate configuration weights [5].

1.5 Planned research: Feb 2015 – Jan 2016

We intend to use the efficient CTQMC method to extend and improve the understanding and knowledge of materials in several ways.
On the one hand, we aim at extracting phenomenological parameters from these exact simulations, that can be used both for the *ab initio* simulation of realistic systems with LDA, DFT or Car-Parrinello algorithms, and to calculate the large-scale dynamical responses of strongly correlated systems using effective theories such as the time-dependent Ginzburg-Landau theory employed to describe superconductors.
On the other hand, we plan to use the efficient CTQMC to study quantum phase transitions in systems involving Dirac fermions (like graphene) and multi-flavor Dirac fermions. In the later case, we can systematically approach to the large-N limit, conventionally studied by the field theory, and test its validity.
Further applications include using the entanglement entropy and fidelity to search for and characterize exotic quantum states without a

local order parameter description, such as the topological Mott insulators.

1.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Synergies with the groups of Philipp Werner and Nicola Spaldin are expected.

b) *Outside MARVEL* We plan to pursue high-performance CTQMC implementations in collaboration with Miguel Morales at the Lawrence Livermore National Laboratory.

1.7 Personnel

NCCR personnel

- Alexei Soluyanov, postdoc, 100% from September 2014.
- Georg Winkler, PhD student, 100%, from September 2014.

Matching personnel

- Lei Wang, postdoc, 50%, from May 2014.
- Mauro Iazzi, postdoc, 100%, from May 2014.
- Hiroshi Shinaoka, postdoc, 50%, from May 2014.
- Guglielmo Mazzola, postdoc, 50%, from January 2015.

2 Dynamical mean field based methods – GW+DMFT (Philipp Werner — UniFR)

2.1 Research summary

- Multi-orbital impurity solvers.
- Frequency dependent interactions.
- Extended DMFT.
- Self-consistent GW+DMFT.

2.2 Research question

Calculating the electronic structure of strongly correlated materials from first principles is a challenging problem for which reliable and accurate methods still need to be developed. The current state of the art is the combination of density functional based *ab initio* calculations with dynamical mean field theory (LDA+DMFT) [9]. While this scheme has provided important insights, it has conceptual problems such as the double-counting of interaction energies. Replacing the LDA input by GW allows to put the theory (GW+DMFT [10]) onto a firm diagrammatic basis. This scheme should avoid some of the double counting problems of LDA+DMFT and enable a fully self-consistent computation of the dynamically screened interaction parameters. It therefore has the potential to become a true *ab initio* method for strongly correlated compounds. At present, self-consistent GW+DMFT calculations have been implemented only for very simple models [11, 12][6]. The goal of this project is to implement and test the self-consistent GW+DMFT approach for real multi-band materials.

2.3 Scientific goals

a) *Short-term* DMFT maps a lattice model onto a self-consistently determined quantum impurity model [13], so that the numerically challenging step becomes the solution of an impurity model. While there are efficient and numerically exact Monte Carlo techniques to treat impurity models with up to 7 orbitals [14], these methods work on the imaginary-time axis. A self-consistent GW+DMFT calculation is therefore most conveniently implemented on the imaginary-time axis, and an important short-term goal is to get hold of an imaginary frequency GW code.
Furthermore, since realistic applications of GW+DMFT will involve multi-orbital systems with long-range interactions, we have to develop a GW+DMFT self-consistency loop for such multi-orbital systems. Once these two ingredients are available, the goal will be to perform self-consistent GW+DMFT calculations within a small subspace of orbitals corresponding to the lowest-energy bands.

b) *Long-term* The long-term goal will be to extend the self-consistent GW+DMFT calculations to a larger subspace of orbitals, covering an energy window of $\sim$ 20 eV, while restricting the impurity calculation to a few strongly correlated bands. Once the optimal scheme, in terms of accuracy and computational effort, has been identified, the method may then be applied to a broad range of materials. The hope is that this scheme will allow quantitative predictions for the electronic structure of three di-

mensional correlated compounds.

2.4 Plans and results: May 2014 – Jan 2015

The imaginary-axis GW code is currently being developed by an outside collaborator, Fredrik Nilsson, in the group of Ferdi Aryasetiawan (Lund University). At the same time, a multi-band GW+DMFT code for model systems is being developed at UniFR by Lewin Boehnke, who started as a postdoc in October 2014. The interfaces between the two codes have been defined, so that it should be easy to read in the parameters and self-energies from the *ab initio* GW code into the model code and vice versa. First test calculations for $SrVO_3$, in which the self-consistent calculation is restricted to the three t_{2g} bands near the Fermi level, are expected to be running by the end of 2014.

In the mean time, as a warm-up problem for GW+DMFT simulations, and as a continuation of a long-term effort to incorporate the dynamically screened nature of the Coulomb interaction into *ab initio* simulations of correlated materials [15], we have performed simulations of a prototypical high-T_c material, La_2CuO_4, using both the *ab initio* band structure and the frequency-dependent interaction parameters computed with the constrained random phase approximation [7]. This study showed that the dynamical screening effect in cuprates is very strong, and that it is important to treat this effect in order to obtain a consistent description of the low-energy electronic structure. In particular, it was found in a three band calculation that the spectral function of the stochiometric compound does not have a gap if the interaction parameters are set equal to static interaction values, while the simulation which treats the full frequency dependence produces a gap of about 2 eV, which is in good agreement with photoemission experiments. It was also clarified that the strong screening effect in cuprates originates from a low-energy collective excitation at 9 eV, which is localized on the Cu $d_{x^2-y^2}$ orbital. This follows from the fact that the corresponding peak in the imaginary part of the frequency dependent interaction shows up only in the *d*-*d* interaction, but not in the *p*-*p* and *d*-*p* interaction (Fig. 1). We furthermore found that it is essential to treat the interatomic *p*-*d* interactions at least at the Hartree level in order to stabilize a correct relative splitting between the *d* and *p* bands and to open a Mott gap.

The GW+DMFT scheme will go beyond such an LDA+DMFT calculation with frequency-dependent $U(\omega)$ in several respects: (i) the frequency dependent interaction will be self-consistently computed, taking into account at least the renormalization or splitting of the lowest-energy bands, (ii) the GW+DMFT calculation will take into account long-range Coulomb interactions, and their effect on screening and the *p*-*d* level splitting, and (iii) the GW+DMFT scheme introduces a momentum dependence of the self-energy. The latter is expected to compensate to some extent the effect of the frequency dependence, because the k dependence tends to widen the bands, while the frequency dependence of the interaction leads to an effective narrowing of the bands [16]. The GW+DMFT formalism may also enable a more transparent definition of a "double counting" term, which is needed when only a subset of orbitals is treated within DMFT.

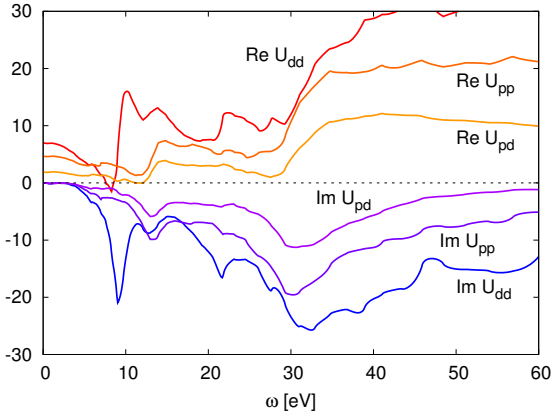


Figure 1: *Dynamically screened interactions in the three-band model of La_2CuO_4. Note that the sharp low-energy peak at 9 eV is present only in the d-d interaction (from [7]).*

2.5 Planned research: Feb 2015 – Jan 2016

The plans for the next year are to complete the tests for the 3-band $SrVO_3$ model and to automatize the exchange of data between the GW code and the (extended) DMFT self-consistency loop. Since the proper treatment of the long-range Coulomb interaction is a numerically subtle issue, we will furthermore test the effect of the long-range interaction terms in a model context, and try to identify appropriate simplified schemes with a finite interaction range. We will also systematically increase the energy range of the subspace in which the GW self-energy is self-consistently computed, while keeping the number of correlated orbitals (treated in DMFT) fixed. This will give important information about the quantitative effect of the self-consistent renormalization

of the intermediate energy electronic structure. On the basis of these results, one can then define a "multiscale" approach which is both accurate and numerically feasible.

2.6 Personnel

NCCR personnel

- Lewin Boehnke, PhD student, 100%, from October to November 2014, postdoc, 100%, from December 2014.

Matching personnel

- Denis Golez, postdoc, 100%, from May 2014.

3 Density functional theory and beyond (Jürg Hutter — UniZH)

3.1 Research summary

- Hartree-Fock (HF) exchange.
- Auxiliary density matrix methods (ADMM).
- MP2 method.
- Random phase approximation (RPA).
- GW method.
- High-performance computing.
- massively parallel computers.

3.2 Research question

The search for density functionals that yield reliable results for a wide range of systems (molecules and solids, metals, semiconductors, and insulators, strong and weak interactions) has led to methods that are based on an explicit treatment of correlation. Functionals based on the family of methods connected to the random phase approximation (RPA) have the potential to provide both, high accuracy and general applicability. We are looking for algorithms to calculate the energy at these levels of theory while still providing a computational efficiency that allows application to large systems.
Another line of research follows from the general form of the Dyson quasi-particle equation. This non-linear single particle Schrödinger equation can be found in mean-field methods, like Hartree-Fock (HF) and Kohn-Sham density functionals, in special forms of coupled cluster theories, and with a full frequency dependent self-energy in GW theory.

$$\left[-\frac{1}{2}\nabla^2(\mathbf{r}) + V_{\mathrm{H}}(\mathbf{r})\right]\Phi_i(\mathbf{r}) + \int d^3r'\,\Sigma(\mathbf{r},\mathbf{r},\varepsilon_i)\Phi_i(\mathbf{r}') = \varepsilon_i\Phi_i(\mathbf{r})$$

Based on this similarity it is intriguing to investigate the underlying connection of these methods and to device new approaches that bridge the gap between them.

3.3 Scientific goals

The goal of this project is to develop new high-level correlation methods for electronic structure calculation of complex materials, implement them into efficient software programs (CP2K) and apply them on innovative materials using modern massively parallel computer architectures.

a) *Short-term*

- Improvement of the auxiliary density matrix methods (ADMM), make them more robust and applicable to a wider range of problems.

$$E_x^{\mathrm{HF}}[\mathbf{P}] \approx E_x^{\mathrm{HF}}[\tilde{\mathbf{P}}] + E_x^{\mathrm{local}}[\mathbf{P}] - E_x^{\mathrm{local}}[\tilde{\mathbf{P}}]$$

- Apply the computational framework of the ADMM method together with the resolution-of-the-identity method (already used together for MP2 and RPA calculations) to GW-based methods.

b) *Long-term*

- Further develop high-level correlation methods using efficient algorithms for large systems. This will be based on the mean-field (HF and KS-DFT) kernel and a posteriori correlation methods. Builds on our RPA and MP2 engines.

- Develop a solver for quasi-particle equations defined by orbital-dependent and non-local potentials. Transform this into a general scheme for property calculations.

- Develop a solver for quasi-particle equations with frequency-dependent self-energies. Explore a wide range of approximations for the self-energy (G_0W_0 to coupled cluster methods).

3.4 Plans and results: May 2014 – Jan 2015

In the first period of this project we are focusing on the treatment of Hartree-Fock exchange. We build our efforts on previous work in our group that resulted in new schemes for exact exchange calculations. These auxiliary density matrix methods (ADMM) are very flexible and leave a lot of room for improvements. We first concentrate on alternative methods to constrain the auxiliary density matrix and to improve the flexibility for the density correction exchange functional.
As a side project we are interfacing the TELNES program with CP2K. TELNES is a versatile program for electron energy loss near-edge structure (ELNES) calculations, including treatment of orientation, relativistic interaction potential, quadrupole or arbitrary transitions, convergence / collection / scattering angle, broadening schemes for core hole lifetime / final state lifetime / instrumentation. Most of these factors strongly influence the calculated spectrum and improve it beyond comparing ELNES measurements to X-ray absorption near-edge structure (XANES) calculations.

3.5 Planned research: Feb 2015 – Jan 2016

We expect that a second PhD student will start in spring 2015 working within this project. It is also our goal to find an experienced postdoc that can join the project. Implementation of GW-based methods into the CP2K code will be the primary focus. We will make use of the combination of resolution-of-the-identity approaches with the Gaussian and plane-waves method. The exact exchange will be treated using ADMM methods. The implementation will be based on our high-performance sparse matrix multiplication library (DBCSR) [8]. A first milestone will be to have the G_0W_0 method working. We will then explore possible connections to related developments within the quantum chemistry community; namely, Dyson orbitals and propagator methods. This will allow us to further develop possible approximations that could lead towards total energy methods.
A second line of development will be based on methods defined by a non-local, but frequency independent, potential. We will develop a general framework for such methods based on a solver for the quasi-particle equation. We will explore different energy functional definitions and derive properties (gradients, multipoles) for these schemes.

3.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* We do have a close collaboration on the development of these types of methods with Joost VandeVondele (ETHZ). We anticipate collaborations on orbital-dependent functionals in general with the group of Nicola Marzari (EPFL). There will be synergies from the applications of these functionals within VP2.

b) *Outside MARVEL* These developments depend strongly on the basic high-performance computing modules available within CP2K. There is a co-design effort, supported by the PASC initiative, to improve these basic libraries. Besides the groups mentioned above, this effort includes contributions from CSCS, Cray, and Nvidia.

3.7 Personnel

NCCR personnel

- Jan Wilhelm, PhD student, 100%, from November 2014.

Matching personnel

- Kevin Jorissen, postdoc, 100%, from 17 September 2014.

4 (Photo-)electochemisty beyond GGA DFT (Joost VandeVondele — ETHZ)

4.1 Research summary

- Random phase approximation.
- MP2 theory.
- Hybrid functionals.
- Redox potentials.
- Aqueous chemistry.
- TiO_2.

4.2 Research question

Being able to compute, rationalize, and manipulate the free energies of electrons in complex chemical systems is key for a sustainable energy future. Indeed, the useful work that can be extracted from an electrochemical system is described by this quantity. Yet, our capability to accurately compute this quantity in realistic systems is recent and limited [17, 18]. The challenge is indeed formidable, as it requires

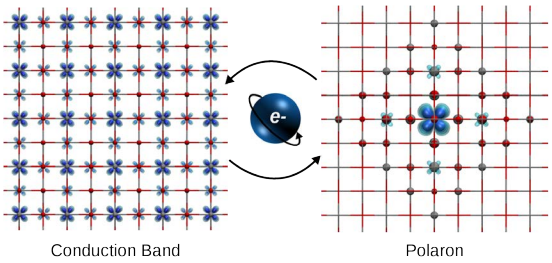


Figure 2: *Shown is an excess electron in TiO_2 (anatase), as either a conduction band electron (left) or a localized polaron (right) (from [20]). Getting the energetic balance between these states right requires 4th and 5th rung functionals. For the latter, more recent, class of functionals, computing properties (e.g. forces, spin densities, etc.) is still not possible in the condensed phase, especially for systems with unpaired electrons. To fully describe these systems at the 5th rung is a target of this project.*

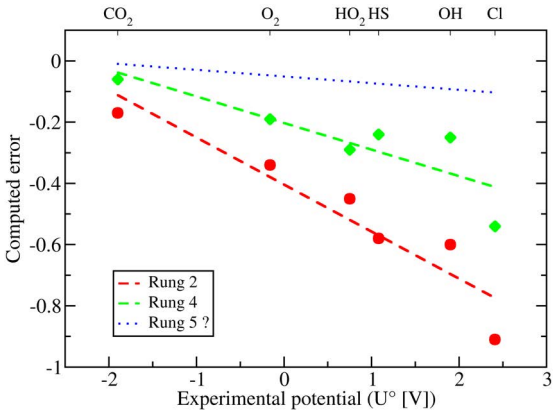


Figure 3: *Shown is the error in computed redox potential vs the experimental redox potential for a number of aqueous species (see e.g. [17, 18]). For those species that have a redox level close to the valence band of the liquid, the error is significant. The error is very large with rung 2 (GGA, red) functionals, and reduced with rung 4 (hybrid, green) functionals. However, quantitative and predictive behavior needs a next level of electronic structure theory. In this project, the performance of so called 5th rung functionals will be investigated, anticipated performance is shown in blue.*

conceptual advances as well as state-of-the-art computational methodologies, to describe finite temperature systems, interfaces [19, 20] between ordered solids [20] and disorded aqueous systems [21, 22], and the balance between localized and delocalized electrons. Striking this balance is essential to obtain a qualitatively correct picture of a variety of phenomena related to this NCCR. Shown in Fig. 2 is the example of the excess electron in anatase, that can either be delocalized as a conduction band electron, or localized to form a polaron. Which state is formed depends sensitively on the level of theory and the model employed, yet resolving this question is key to understand transport in a variety of devices based on TiO_2, including sensors and dye sensitized solar cells. The incorrect balance between localized and delocalized states is a known weakness of the theory almost uniquely used for the formerly mentioned systems, i.e. density functional theory (DFT) in the generalized-gradient approximation (GGA). The impact of this is shown in Fig. 3 for the redox potentials of a number of aqueous species. The redox potential of those species that have a redox level close to the valence band of liquid water can not be computed accurately. This includes, for example, the OH radical, which is an important intermediate. In this project, we will go beyond GGA DFT, and establish more advanced electronic structure methods, for these systems.

4.3 Scientific goals

The project goal is to establish advanced electronic structure methods such as RPA and MP2 for aqueous electrochemistry and ultimately solid/liquid interfaces. This includes a wide variety of aspects, ranging from applications to developments. In no particular order: a) add features required for the description of electrochemical systems, such as molecular dynamics (MD) capability, unrestricted calculations (deal with an odd number of electrons), and further properties (e.g. spin density); b) implement more advanced forms of RPA, i.e. kernels or higher-order corrections that improve the accuracy of the electronic structure method; c) increase the performance of the method to enable simulations of large systems, and to enable sampling; d) apply this theory to aqueous and interfacial systems.

4.4 Plans and results: May 2014 – Jan 2015

No results have been produced so far, as this project has been staffed starting December 1^{st}. Now that a suitable postdoc has been successfully hired, we expect to make progress with this project quickly.

4.5 Planned research: Feb 2015 – Jan 2016

The new postdoc will be trained to work with CP2K [23], perform *ab initio* molecular dynamics simulations, and study the existing MP2 and RPA code [24, 25][1]. As first tasks he will extend the implementation of the forces in MP2 to deal with unrestricted calculations. Following this, he will work on extending the RPA

code to deal with more advanced correlation schemes, such as, e.g., AXK or SOSEX. During this time, he will work on fundamental aqueous systems containing redox active species.

4.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* The planned work will be beneficial for the studies proposed in VP2, i.e. provides some of the essential theoretical tools to deal with systems such as dye-sensitized solar cells, and photocatalytic water splitting systems (groups of Ursula Röthlisberger and Alfredo Pasquarello, VP2). The development work will take place in the CP2K code, which I co-develop with Jürg Hutter (HP3). This project will be performed in close collaboration as the theoretical ingredients are broadly useful. With these advanced electronic structure techniques, that for example provide proper treatment of van der Waals interactions, we expect to contribute to other projects, e.g. molecular packing and nucleation (Goedecker). Guidance with the use of these methods can be provided to all MARVEL team members.

b) *Outside MARVEL* Collaborations on this topic are ongoing with M. Sprik (Cambridge) and J. Cheng (Aberdeen).

4.7 Personnel

NCCR personnel

- Vladimir Rybkin, postdoc, 90%, from December 2014.

MARVEL-related publications

List of publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.

[1] M. Del Ben, O. Schütt, T. Wentz, P. Messmer, J. Hutter, and J. VandeVondele, *Enabling simulation at the fifth rung of DFT: Large scale RPA calculations with excellent time to solution*, Computer Physics Communications **187**, 120 (2015).

[2] M. Iazzi and M. Troyer, *Efficient continuous-time quantum Monte Carlo algorithm for fermionic lattice models*, arXiv:1411.0683 (2014).

[3] M. Iazzi, A. A. Soluyanov, and M. Troyer, *Topological Origin of the Fermion Sign Problem*, arXiv:1410.8535 (2014).

[4] L. Wang and M. Troyer, *Renyi Entanglement Entropy of Interacting Fermions Calculated Using the Continuous-Time Quantum Monte Carlo Method*, Physical Review Letters **113**, 110401 (2014).

[5] H. Shinaoka, R. Sakuma, M. Troyer, and P. Werner, *Accuracy of downfolding based on the constrained random phase approximation*, arXiv:1410.1276 (2014).

[6] L. Huang, T. Ayral, S. Biermann, and P. Werner, *Extended dynamical mean-field study of the Hubbard model with long-range interactions*, Physical Review B **90**, 195114 (2014).

[7] P. Werner, R. Sakuma, F. Nilsson, and F. Aryasetiawan, *Dynamical screening in La_2CuO_4*, arXiv:1411.3952 (2014).

[8] U. Borštnik, J. VandeVondele, V. Weber, and J. Hutter, *Sparse matrix multiplication: The distributed block-compressed sparse row library*, Parallel Computing **40**, 47 (2014).

Other references

[9] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, *Electronic structure calculations with dynamical mean-field theory*, Reviews of Modern Physics **78**, 865 (2006).

[10] S. Biermann, F. Aryasetiawan, and A. Georges, *First-Principles Approach to the Electronic Structure of Strongly Correlated Systems: Combining the GW Approximation and Dynamical Mean-Field Theory*, Physical Review Letters **90**, 086402 (2003).

[11] T. Ayral, S. Biermann, and P. Werner, *Screening and nonlocal correlations in the extended Hubbard model from self-consistent combined GW and dynamical mean field theory*, Physical Review B **87**, 125149 (2013).

[12] P. Hansmann, T. Ayral, L. Vaugier, P. Werner, and S. Biermann, *Long-Range Coulomb Interactions in Surface Systems: A First-Principles Description within Self-Consistently Combined GW and Dynamical Mean-Field Theory*, Physical Review Letters **110**, 166401 (2013).

[13] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions*, Reviews of Modern Physics **68**, 13 (1996).

[14] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, *Continuous-time Monte Carlo methods for quantum impurity models*, Reviews of Modern Physics **83**, 349 (2011).

[15] P. Werner and A. J. Millis, *Dynamical Screening in Correlated Electron Materials*, Physical Review Letters **104**, 146401 (2010).

[16] R. Sakuma, P. Werner, and F. Aryasetiawan, *Electronic structure of $SrVO_3$ within GW+DMFT*, Physical Review B **88**, 235110 (2013).

[17] J. Cheng, X. Liu, J. VandeVondele, M. Sulpizi, and M. Sprik, *Redox Potentials and Acidity Constants from Density Functional Theory Based Molecular Dynamics*, Accounts of Chemical Research **47**, 3522 (2014).

[18] C. Adriaanse, J. Cheng, V. Chau, M. Sulpizi, J. VandeVondele, and M. Sprik, *Aqueous Redox Chemistry and the Electronic Band Structure of Liquid Water*, The Journal of Physical Chemistry Letters **3**, 3411 (2012).

[19] J. Cheng, J. VandeVondele, and M. Sprik, *Identifying Trapped Electronic Holes at the Aqueous TiO_2 Interface*, The Journal of Physical Chemistry C **118**, 5437 (2014).

[20] C. Spreafico and J. VandeVondele, *The nature of excess electrons in anatase and rutile from hybrid DFT and RPA*, Physical Chemistry Chemical Physics **16**, 26144 (2014).

[21] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, *Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory*, The Journal of Physical Chemistry Letters **4**, 3753 (2013).

[22] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, *Correction to "Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory" (**4**, 3753 (2013))*, The Journal of Physical Chemistry Letters **5**, 3066 (2014).

[23] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, *CP2K: atomistic simulations of condensed matter systems*, WIREs Computational Molecular Science **4**, 15 (2014).

[24] M. Del Ben, J. Hutter, and J. VandeVondele, *Second-Order Moller-Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach*, Journal of Chemical Theory and Computation **8**, 4177 (2012).

[25] M. Del Ben, J. Hutter, and J. VandeVondele, *Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme*, Journal of Chemical Theory and Computation **9**, 2654 (2013).

Horizontal Project 4 HP4 — Advanced Sampling Methods

Project leader: Stefan Goedecker (UniBas)

Participating members: Stefan Goedecker (UniBas), Michele Parrinello (USI and ETHZ), Berend Smit (EPFL), Michele Ceriotti (EPFL), Anatole von Lilienfeld (UniBas), Nicola Marzari (EPFL), Ursula Röthlisberger (EPFL)

Summary and highlights: The Horizontal Project 4 (HP4) gathers several internationally leading groups in the area of potential energy surface sampling. Earlier work of these groups laid the foundations for future ambitious achievements. The work in the Parrinello group on the nucleation and growth of organic crystals from solution was pioneering in this field. The methodological work of the Goedecker group on structure prediction methods allows to predict a large number of crystalline structures with density functional accuracy. These foundations and achievements will allow to reach the next ambitious objectives.

General view of the project

Major research questions

The potential energy surface, which gives the energy of a system as a function of its atomic coordinates, is the central quantity to understand the structure and dynamics of any condensed matter system. The global minimum gives the structural ground state, other local minima metastable states and the properties of the saddle points determine the speed of chemical reactions as well as other dynamical processes. For a system containing N atoms the potential energy surface is a $3N$ dimensional function. This high dimensionality makes the exploration of the potential energy surface extremely difficult. The main goal of HP4 is to develop methods to sample and to understand this potential energy surface for various systems. The group of Goedecker focuses on methods for finding low-energy local minima as well as reaction pathways on this potential energy surface. The group of Parrinello rather focuses on the free energy which is typically not given as a function all the $3N$ coordinates of the system but as a function of a smaller subset of order parameters. The group of Parrinello is also interested in kinetic processes occurring on the free energy surface. The group of von Lilienfeld focuses on the chemical compound space, i.e is interested how the potential energy surfaces and resulting properties change if the identity of atoms changes. Finally several groups are interested in developing methods for describing and quantifying the potential energy surface and configurations corresponding to certain points on the potential energy surface. The group of Ceriotti has developed for instance sketch maps (in collaboration with Michele Parrinello) as well as other methods to characterize certain bonding patterns such as hydrogen bonds. The group of Smit is applying topological analysis to discover sets of similar materials. As a consequence, all aspects of sampling and characterizing configurational spaces are well covered in this horizontal project, and plenty of common points of interest exist.

Short- and long-term goals

All the groups within HP4 have developed novel methods for sampling the potential energy surface. Applying and improving these methods will be the main activity in the short term. The group of Goedecker will do structure predictions for materials that are of interest to application oriented projects within MARVEL. The group of Parrinello will mainly concentrate on applying metadynamics to systems with a significant conformational flexibility. They will also start applying their new variational approach for enhanced sampling to more complex systems. The group of Smit will use more sophisticated tools to characterize zeolite pore topologies. The group of Ceriotti will further improve their methodology to construct sketch maps and make these developments available within the PLUMED software package. The group of von Lilienfeld will apply "alchemical" methods to alkali halides.
The long term goals of all the groups in HP4 are rather convergent. A better prediction of structures as well as dynamical processes such as nucleation and growth, the development

of better descriptors to characterize configurations and as a starting point for machine learning methods.

First achievements

In an extensive structure prediction of silicon allotropes, the group of Goedecker has considerably extended the scope of known silicon allotropes and has found numerous structures with useful properties for photovoltaic applications. The group of Parinello was able to simulate, with unprecedented detail and in agreement with experiment, the nucleation of organic molecules from solution. In addition the group has developed a promising variational approach for the calculation of free energies. The methods developed by the group of Smit have allowed to identify zeolite structures with optimal properties for carbon dioxide or methane capture. The group of Ceriotti has developed machine learning tools for the automatic recognition of hydrogen bonds. The group of von Lilienfeld has further developed their alchemical derivatives method to predict the change of properties in alchemical transformations.

Next steps

The NCCR has just started and many groups have not yet terminated their hiring phase. Once the full manpower of the NCCR will be available, a further acceleration of the scientific output is to be expected.

Research contributions to the overall goals of the NCCR and to the existing literature

The availability of efficient sampling methods is a key ingredient for the more application oriented vertical projects of MARVEL. First collaborations between HP4 and other subprojects have been established and the improved sampling methods will be used within these collaborations. The integration of various softwares has also well taken off and will open up new possibilities. Virtually all publications of the involved groups were published in high-quality journals, demonstrating the top quality of this NCCR.

Collaborative components

Numerous collaborative efforts are under way both within HP4 as well as with other MARVEL projects outside HP4. In a collaboration between the groups of Ceriotti and Goedecker, an interface was developed to couple the minima hopping structure prediction method with the Quantum-ESPRESSO code to perform structure predictions for solids. This was done using a technology developed by the Ceriotti group. The Goedecker group has also started a collaboration with the group of Ursula Röthlisberger to perform structure predictions on perovskite materials that are promising for photovoltaic applications. The group of Ceriotti is in close contact with the Parrinello group for the further development of sketch maps and with the group of Marzari for the incorporation of their software developments into AiiDA. The group of von Lilienfeld is analyzing recent results on alanates of the Goedecker group with its machine learning tools and is collaborating with the Röthlisberger group on optimizing the composition of training sets for machine learning models.

1 Low-density allotropes of silicon (Stefan Goedecker — UniBas)

1.1 Research summary

A systematic structural search reveals that there is a very large number of low-density silicon allotropes with potentially highly interesting properties.

1.2 Research question

Most materials have several allotropes, i.e. different structures for a given stochiometry. These allotropes can have vastly different properties. Silicon clathrates are already known for a long time as low-density silicon allotropes and they can also be synthesized experimentally. Some other low-density structures have theoretically been proposed as well. By doing a systematic structural search, we were however able to discover a large number of hitherto unknown low-density structures.

1.3 Scientific goals

As silicon is one of the most abundant materials on earth and since a large amount of experience is available for the production and technological handling of this material, it is of great interest to discover the large variety of structures that can be obtained from this material. Using the minima hopping structure prediction method [8], we have already pre-

Figure 1: *Polyhedral structure of the cages in low-density allotropes of silicon.*

viously shown that even some normal density allotropes have quite different properties compared to the standard diamond structure, and that in particular some of these have a much higher light absorption which would make them ideally suited for photovoltaic applications. It turns out that some of these new low-density allotropes have also quite surprising properties.

a) *Short-term* We will finish the study of the properties of the structures we have found and publish the results. Fig. 1 gives an overview over the structures found.

b) *Long-term* For the structures with very most promising properties, we will try to give guidance to the synthesis process of these structures. In this context, we will perform structure predictions for low-density structures, where the cages are filled with alkaline or earth alkaline metals, and we will map out the stability of various structures as a function of pressure and temperature. We will also explore whether the metal atoms can be removed easily by chemical post-processing steps to obtain pure silicon structures.

1.4 Plans and results: May 2014 – Jan 2015

We have developed a modified version of minima hopping that is more efficient than the standard version in finding low-density structures. In this version, we include empty Lennard-Jones spheres which are repulsive with respect to the real atoms in the system. In this way, we find low-energy structures of a artificial binary system consisting of the real atoms and the fake Lennard-Jones atoms. After removing the fake atoms, we check carefully whether the structure remains stable. This is in general the case and voids are formed at the places where the fake atoms were positioned. Neighboring voids frequently form channel-like structures, which could be useful for gas storage.

1.5 Planned research: Feb 2015 – Jan 2016

We will further study silicon structures with voids and we will in particular examine whether real atoms can be placed in the voids. This would be important to know for synthesis efforts, since, in the standard synthesis procedure for clathrates, two-component structures of silicon and other elements are first formed before the other species of atom is removed. We will also further explore potentially useful properties of these novel materials.

1.6 Personnel

Matching personnel

- Max Amsler, postdoc, 100%, from August 2014.

2 Software integration of minima hopping with Quantum-ESPRESSO and other quantum engines (Michele Ceriotti — EPFL, Stefan Goedecker — UniBas, Nicola Marzari — EPFL)

2.1 Research summary

The integration of various software packages that are available in different groups participating in MARVEL is an important aspect within this NCCR and will allow new applications.

2.2 Research question

The minima hopping structure prediction method developed in the group of Goedecker [8] is a powerful tool. It requires a quantum engine to calculate energies and forces as well as stress tensors, and we want to use the Quantum-ESPRESSO package in this context. Until recently the coupling of the two codes had to be done via files, which can considerably slow down the simulations and prevent a high degree of parallelization.

2.3 Scientific goals

In a collaboration between the groups of Michele Ceriotti and Stefan Goedecker, we succeeded to couple the two codes by using a communication scheme that is based on sockets. This eliminates the cumbersome communication by files and will allow to do structure predictions much more rapidly on large parallel computers such as the computer of the Swiss National Supercomputing Center (CSCS) in Lugano. To do so, we leveraged the use of i-PI, an interface that has originally been designed to perform path integral simulations, but that can be more in general used to evolve nuclear configurations, obtaining the energy and forces from an external program, the communication happening over TCP/IP sockets based on a client-server paradigm (Fig. 2) [9]. A one-week visit of Max Amsler to EPFL has made it possible to implement a small change to the i-PI interface, that is designed to use an external program to propagate the atomic configurations. Therefore i-PI can now serve as an intermediate agent that makes it possible for the structure search code developed in Basel to easily communicate with the growing set of electronic structure codes that provide an i-PI interface. In the next few months we will also develop an interface between i-PI and the PLUMED code for advanced free-energy sampling, to enable the use of such techniques also with simulation packages that are not directly supported with PLUMED patches.

a) *Short-term* Minima hopping coupled to Quantum-ESPRESSO will be tested on the CSCS machines and be used in first applications.

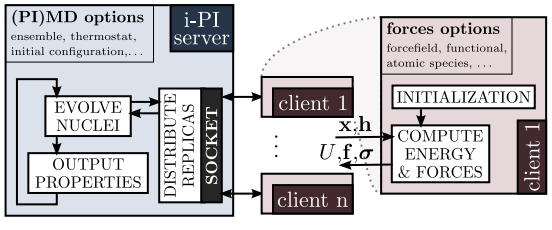


Figure 2: *Schematic illustration of the i-PI communication scheme*

b) *Long-term* The long term goal is a large scale integration of all the algorithms and software packages used within the NCCR.

2.4 Plans and results: May 2014 – Jan 2015

After having accomplished the integration of the methods, we started to use this combination of methods in a first application which was the structure prediction of perovskites that are of interest for photovoltaic applications. Runs on several computer architetures were successful and some interesting new physical results could be obtained.

2.5 Planned research: Feb 2015 – Jan 2016

The integration of these various software developments will allow to tackle new challenging physical problems. It is to be expected that most of the future structure prediction work for crystalline structures will be done based on this combination of minima hopping with Quantum-ESPRESSO.

2.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* This work is a collaboration of the groups of Michele Ceriotti and Nicola Marzari at EPFL and the group of Stefan Goedecker.

2.7 Personnel

Matching personnel

- Max Amsler, postdoc, 100%, from August 2014.

3 Development of a library to include the effects of an aqueous environment on electronic structure calculations (Stefan Goedecker — UniBas, Nicola Marzari — EPFL)

3.1 Research summary

The study of quantum simulations in the presence of wet environments plays a key role in many scientific fields. If quantum simulation methods are applied to these processes and systems, it becomes imperative to have a qualitatively and quantitatively correct modeling of the effects of the environment. The present research will considerably extend the range of applications of such methods.

3.2 Research question

The computational study of chemical reactions in complex, wet environments is critical for applications in many fields, and of cross-disciplinary interest to physics, chemistry, materials science, chemical engineering, and biology. It is often essential to study chemical reactions in the presence of an applied electrochemical potential, establishing the correct relation between charge and potential, and taking into account the complex electrostatic screening coming from the solvent and the electrolytes. Simulations of various systems in an aqueous, or a generic wet environment are of great interest, and powerful libraries are needed which allow to solve the generalized Poisson's equation with a spatially varying dielectric constant as well as the Poisson-Boltzmann equation.

3.3 Scientific goals

a) *Short-term* The library to solve the generalized Poisson equation will be integrated in the existing electronic structures codes of MARVEL. Minima hopping coupled to Quantum-ESPRESSO will be tested on the CSCS machines and be used in first applications.

b) *Long-term* The core objective of this library will be to describe complex electrostatic environments where an explicit solvent becomes implicit, with a position-dependent dielectric constant, or where mobile ions can shield the charge as well as higher multipoles of the system of interest; it will embed the quantum simulation engines into a robust and efficient Poisson-Boltzmann solver that has been extensively verified and validated. The Poisson-Boltzmann equation is more difficult to solve than the generalized Poisson equation due to the fact that it leads to a non-linear problem and due to the possibly strong variation of the ion concentration with respect to the electric potential. Hence, in contrast to the standard or generalized Poisson equation, there is no general consensus on what flavor of the Poisson-Boltzmann equation is most appropriate for various problems. As a consequence the Poisson-Boltzmann library should allow to solve various flavors of this equation. It should also be possible for users to add easily to the library new variants.

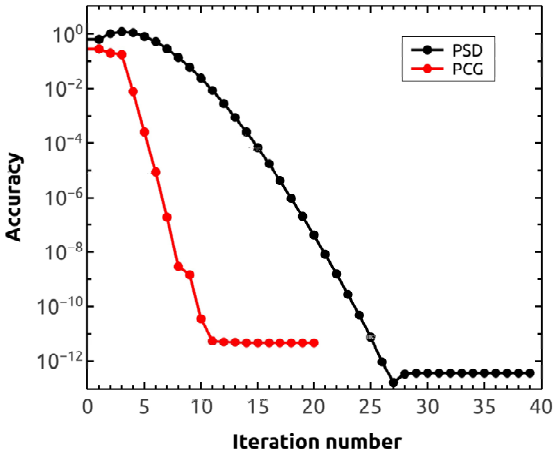


Figure 3: *Accuracy of the preconditioned steepest descent (PSD) and Preconditioned conjugate gradient (PCG). Gaussian functions have been used for the potential and dielectric function.*

3.4 Plans and results: May 2014 – Jan 2015

To handle the generalized Poisson equation an algorithm based on the preconditioned conjugate gradient (PCG) method has been implemented. It allows to solve iteratively the minimization problem with some ten iterations. The algorithm takes advantage of a preconditioning procedure which exactly represents the operator in the limit of a slowly varying dielectric constant, and is based on the BigDFT Poisson solver for the standard Poisson equation. Since the preconditioner is the exact inverse of the associated Hessian in the limit of a slowly varying dielectric constant, a very small number of iterations is required. Fig. 3 shows the accuracy of two different minimization approaches: the preconditioned steepest descent (PSD) and preconditioned conjugate gradient (PCG). Gaussian functions have been used as analytic test functions for the potential and dielectric function. In both approaches the preconditioner is the BigDFT Poisson solver for the standard Poisson equation. Both methods reach an order of accuracy of $10^{-11} - 10^{-12}$, where the implemented PCG needs only a number of iteration of $\sim$ 10. The code turns out to be no-memory consuming and fast. The use of the BigDFT solver as preconditioner allows to manage free, surface, wire and periodic conditions, and concurrently fixes the output accuracy. To test and demonstrate the efficiency of the method, applications of the generalized solver and comparison to other approaches have been carried out. The method used for solving generalized Poisson's equation with a variable dielectric constant will be published soon.

3.5 Planned research: Feb 2015 – Jan 2016

The integration of these various software developments will allow to tackle new challenging physical problems. The library will be highly optimized for massively parallel computer architectures using both traditional CPUs as well as GPUs. Extensive testing, validation and verification of all the components of the new library will be performed. Development, validation and testing of the Poisson-Boltzmann solver will be done for different electrostatic environments.

3.6 Synergies with other computational and experimental efforts

Several systems that are of interest to VP2 require the simulation of systems in an aqueous environment.

a) *Within MARVEL* This work is done in collaboration with the group of Nicola Marzari at EPFL who has developed the physical model on which our numerical implementation is based.

3.7 Personnel

Matching personnel

- Giuseppe Fisicaro, postdoc, 100%, from May 2014.

4 Nucleation of organic molecules from solution (Michele Parrinello — USI and ETHZ)

4.1 Research summary

Computational investigation of organic compounds nucleation processes in solution.

4.2 Research question

The aim of this project is the investigation of nucleation processes using advanced molecular simulation tools. The study of the early stages in the formation of a stable crystal phase from solution represents a crucial, long standing challenge in physical chemistry, connected to a number of applications in material, chemical and pharmaceutical engineering. The mechanism of crystal nucleation is still not completely understood in all its complex-

ity, due to limitations of both the available experimental and computational tools of investigation. The time and size scales involved in crystallization physics represent a huge obstacle for direct experimental observation of nucleating crystallites. On the other hand, the numerical simulation of nucleation through classical molecular dynamics (MD) encounters the opposite problem, since the typically available computational resources access space and timescales which are not large enough to sample nucleation events, even for small atomic or molecular crystals. In this project we aim at overcoming the usual limitations encountered by molecular simulation methods through the application of enhanced sampling approaches. In particular we are interested in tackling the challenges associated with the nucleation from solution of organic compounds, a topic of great relevance in the production of active pharmaceutical ingredients and fine chemicals.

4.3 Scientific goals

a) *Short-term* In the initial months of this project, we have devised a metadynamics-based approach to the simulation of nucleation from solution of small, relatively rigid organic molecules. In the short term, we aim at applying this approach to molecular systems of increasing practical relevance, such as small active pharmaceutical ingredients or fine chemicals. The main challenge in this regard is the extension to systems possessing a significant conformational flexibility.

b) *Long-term* In the long term we aim at studying crystallization processes in realistic systems providing reliable tools for the identification of competing polymorphs, as well as directly computing nucleation rates from enhanced sampling simulations. To this aim, extending the array of available computational tools is very important. In this regard we foresee an increasing synergistic cooperation with other theoretically oriented projects within the MARVEL initiative.

4.4 Plans and results: May 2014 – Jan 2015

In order to devise an enhanced sampling protocol based on metadynamics [10, 11] that could allow to investigate nucleation processes from solution, we have focused our attention on a simple organic molecule: urea. Urea exhibits a rich crystallization behavior. From an investigation of urea crystal growth it has emerged that different crystal faces grow following different mechanisms [12], and that solvents and additives may substantially affect urea crystal morphologies that can range from needle-like particles to compact tetrahedra [13].
We have focused our work in three main areas: (i) the assessment of finite size effects due to confinement in molecular simulations of nucleation proecesses from solution, (ii) the investigation of nucleation mechanisms of urea in different solvents, and (iii) the development of a simulation scheme that allows to mimic constant supersaturation conditions in a finite sized simulation box.

a) *Confinement effects in a finite sized simulation* Conventional molecular dynamics (MD) simulations of crystallization processes in solution are performed in the isothermal-isobaric ensemble, in which temperature, pressure and number of atoms are constant. In this ensemble, when a phase transition, such as nucleation, takes place, the chemical potential of the mother phase (i.e. the solution) inherently couples with the size of the embryo of the new phase [1]. This effect is reflected in the free energy profile associated with the formation of a finite sized crystal nucleus of size N_c, that can be written as:

$$\begin{aligned}\Delta G_{\ell\to c} = &-N_c\, k_B T \ln\left(\frac{\gamma\, x}{\gamma^* x^*}\right) + \sigma' N_c^{\frac{2}{3}} \\ &+ N_{tot} k_B T \ln\left(\frac{\gamma\, x}{\gamma_0 x_0}\right) \\ &+ N_s k_B T \ln\left(\frac{\gamma_s\,(1-x)}{\gamma_{s,0}\,(1-x_0)}\right)\end{aligned} \tag{1}$$

where N_c is the number of molecules belonging to the crystal phase, k_B the Boltzmann constant, T the temperature, γ_i are the activity coefficients, x is the actual molar fraction of the solute, x^* the molar fraction of the solute at equilibrium with the solid, x_0 its molar fraction in absence of any crystal-like nuclei, N_s the number of solvent molecules, N_{tot} the total number of solute molecules in the system, and σ' the effective surface energy that takes into account in an effective way both the surface tension and the shape of the nucleus. From the simulation standpoint, the most relevant consequence of finite size effects is that solution does not depend only on the supersaturation of the initial solution $S = \gamma_0 x_0/\gamma^* x^*$, but also on the volume of the simulation box [1]. The finite size dependence of the free energy profile associated to nucleation is reassumed in Fig. 4.
This rational interpretation of the effects of confinement on the reversible work of nucleation, as well as the verification that the free energy of nucleation computed with metadynamics can be understood using Eq. 1, allow to de-

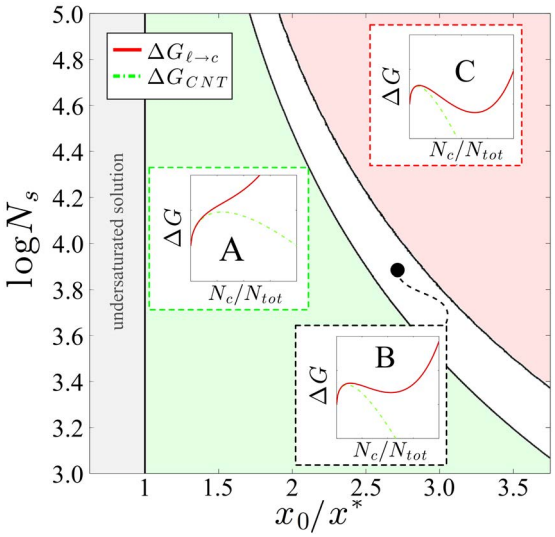


Figure 4: *Reported from [1]. Dependence of the free energy of nucleation features from the initial composition x_0 and the volume expressed as the number of solvent molecules N_s ($V \propto (N_{tot}+N_s) = (1-x_0)^{-1} N_s$). In region A, even if the system is supersaturated, either a stable or a metastable state in which a crystal is present in solution is not present. In region B, a finite sized crystal represents a metastable state with respect to the solution, while in region C a finite sized crystal represents the most stable state for the system.*

vise a systematic correction strategy in order to reweigh the free energy obtained from a finite sized simulation to estimate the free energy in an infinitely large system at constant supersaturation. The correction term for the free energy in the finite sized system can be written as the difference between free-energy in the infinite system ΔG_{CNT} and that in the finite sized simulation box reported in Eq. 1:

$$\begin{aligned}\Delta G_{corr}(N_c) = \Delta G_{CNT} - \Delta G_{\ell\to c} = \\ - (N_{tot} - N_c)\, k_B T \ln\left(\frac{\gamma x}{\gamma_0 x_0}\right) \\ - N_s k_B T \ln\left(\frac{\gamma_s\,(1-x)}{\gamma_{s,0}\,(1-x_0)}\right)\end{aligned} \quad (2)$$

b) *Urea nucleation mechanism* Tacking advantage of the analysis described in the previous paragraph, and the WT metadynamics strategy described in detail in [14, 15], to sample nucleation events of organic molecules with WT metadynamics, we have investigated the nucleation process of urea from different solutions [1, 2]. The solvents considered in our analysis correspond to those taken into account in a previous work on crystal growth [13]: water, methanol, ethanol, and acetonitrile.
The description of the nucleation process emerging from our analysis actually provides an interesting insight into the mechanisms of nucleation from solution. In particular we find that, while in methanol and ethanol the most probable nucleation pathway can be interpreted as a classical single-step nucleation process, in water and acetonitrile nucleation proceeds predominantly through a two-step process, whereby embryonic crystal nuclei emerge from dense, disordered urea clusters.
Furthermore, in the early stages of nucleation, we observe a competition between two different polymorphic structures, that is instead only marginally affected by the solvent choice and agrees with the phenomena identified in previous studies on urea nucleation from the melt [14].

c) *Mimicking constant supersaturation conditions: CμMD* As mentioned in the previous paragraphs, MD study of crystallization is hampered by the size limits of the simulated systems. This limitation determines that the solution is depleted as more and more molecules crystallize. As a result, the solution supersaturation is coupled with the crystal size determining the finite size effects previously described.
In order to reach a deeper understanding of crystallization dynamics it would be of extreme interest to perform MD in the grand canonical (GC) ensemble, that is by simulating an open system in contact with an infinite reservoir of molecules. In the GC ensemble the finite size effects are indeed not present, and crystallization occurs at constant supersaturation, as in realistic macroscopic systems.
However, the classical methods for GC sampling require on-the-fly insertion and removal of particles, which is highly inefficient in a dense fluid. For this reason we have proposed an alternative method to simulate crystallization in a constant-supersaturation environment, avoiding particle insertion. In our scheme, named CμMD, the system containing the crystallite and its solution environment is in contact with a finite size reservoir of liquid. The exchange of molecules between these two regions is regulated by an artificial force, which acts to maintain the crystal environment at constant chemical potential. The artificial force and the volume of the reservoir adapt themselves to the evolution of the system, so that the crystallization is not affected by finite size limitations.
The method is valid as long as the reservoir, which is finite, contains enough solute or solvent molecules to allow the control of the solution in the vicinity of the crystal. The CμMD method could be used to study urea crystal

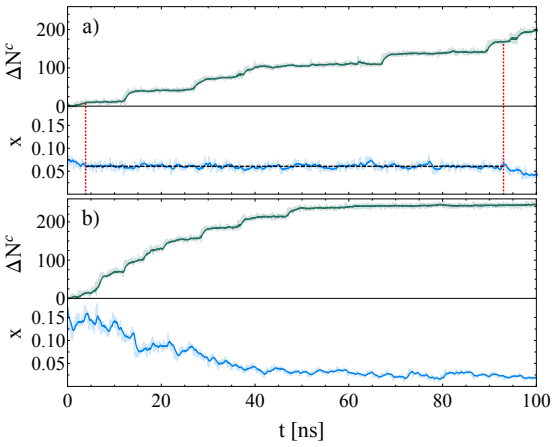


Figure 5: *Crystal size increase ΔN^c (green curves) and solution mole fraction x (blue curves) in (a) CμMD and (b) ordinary MD simulations. In (a) the red dashed lines delimit the validity range of CμMD method and the black dashed line indicates the average x within this range. The mole fraction is measured at a fixed distance from the surface of the crystal.*

growth in aqueous solution, obtaining relevant insights on its dynamics in constant supersaturation conditions. As an example, in Fig. 5, we report the growth dynamics of a urea crystal slab obtained with the CμMD method (a) and ordinary MD (b). As one can note, the CμMD method maintains a stable mole fraction for a relevant time range, while ordinary MD exhibits a continuous solution depletion.

4.5 Planned research: Feb 2015 – Jan 2016

The first two problems that we plan to tackle in the next year are natural extensions of the results obtained in this first stint, while the third is related to the extension to nucleation problems of metadynamics-based methods.

a) *Investigation of polymorph selection in para-aminobenzoic acid nucleation* In order to provide an insight into the effects of polymorph selection due to the solvent choice, we plan to investigate with the array of tools used in the case of urea to tackle the nucleation of para-aminobenzoic acid (PABA). This molecule is known to exhibit two polymorphs that dominate in different crystallization conditions. The aim of our work will be to simulate PABA nucleation in solution in order to extract information about the eventual coexistence of the two structures in small nuclei, and rationalize the effect of the mother solution on the process of polymorph selection.

b) *Investigation of nucleation processess at constant supersaturation* In parallel with the investigation of PABA nucleation, we will extend the *CμMD* method to the investigation of nucleation events. To this aim we will initially investigate urea nucleation. In a second phase we will eventually extend the use of this approach to the case of PABA.

c) *Nucleation rates from metadynamics* Tacking advantage of the recent developments in the calculation of transition rates from metadynamics [16, 17], and of the insight gained in nucleation processes in finite sized simulation boxes [14][1, 2], we will compute nucleation rates from metaynamics, trying to extend rates estimates to *low-supersaturation* regimes, where established methods based on unbiased MD fall short due to the intrinsically long simulation times required.

4.6 Synergies with other computational and experimental efforts

a) *Outside MARVEL* The Parrinello group has an active collaboration with Prof. Marco Mazzotti's group at ETHZ, aimed at mutually benefitting from complimentary work in fields of crystal nucleation and growth.

4.7 Personnel

NCCR personnel

- Pablo Piaggi, PhD student, 100%, from November 2014.

Matching personnel

- Matteo Salvalaglio, postdoc, 50%, from May 2014.
- Claudio Perego, postdoc, 40%, from October 2014.

5 Variational approach to enhanced sampling (Michele Parrinello — USI and ETHZ)

5.1 Research summary

Developing the next generation of enhanced sampling techniques.

5.2 Research question

Advanced sampling techniques are nowadays an integral part of any studies involving atomic-based computer simulations. One

most successful of these methods is metadynamics [18, 10] which was developed in the Parrinello group and has been applied to a wide range of problems from the domain of the chemistry, biophysics, and materials science [11]. The aim of this project is to develop the next generation of advanced sampling techniques based on a novel variational approach [3] that we have recently introduced. The work will furthermore build on top of more than 10 years worth of experience with metadynamics. In the project, considerable attention will be on kinetics as we intend in the long term to develop a well-defined procedure for obtaining rates and other kinetic information from computer simulations in faster manner than currently possible.

5.3 Scientific goals

a) *Short-term* In the short-term the main work is on improving the general performance of the variational approach. We furthermore aim to improve and extend the code for the variational approach and integrate it in the PLUMED 2 enhanced sampling plug-in.

b) *Long-term* One of the long-term goals is to explore the option of developing physics-based models for free energies of molecular system and use these models within our variational approach. Another big emphasis will be on kinetics as we intend to develop a well-defined procedure for getting rates from molecular simulations.
All these developments will be included in the PLUMED 2 enhanced sampling plug-in and made available to other researchers in the MARVEL project and the wider scientific community.

5.4 Plans and results: May 2014 – Jan 2015

Similar to metadynamics our recently introduced variational approach [3] is based on enhancing the sampling in low-dimensional space of some coarse grained descriptors of the system that are called collective variables (CVs). Having defined these CVs, the main premises of our variational method is the following functional $\Omega[V]$ of a bias potential $V(\mathbf{s})$ that acts on the system

$$\Omega[V] = \frac{1}{\beta}\log\frac{\int d\mathbf{s}\, e^{-\beta[F(\mathbf{s})+V(\mathbf{s})]}}{\int d\mathbf{s}\, e^{-\beta F(\mathbf{s})}} + \int d\mathbf{s}\, p(\mathbf{s})V(\mathbf{s}), \tag{3}$$

where $\mathbf{s}$ are the biased CVs and $p(\mathbf{s})$ is the so-called target CV distribution. The bias potential that minimizes $\Omega[V]$ is given up to a constant as

$$V(\mathbf{s}) = -F(\mathbf{s}) - (1/\beta)\log p(\mathbf{s}), \tag{4}$$

where $F(\mathbf{s})$ is the free energy surface (FES) associated with the CVs. $\Omega[V]$ is a convex functional so this is also the global minimum. The CVs will be sampled according to $p(\mathbf{s})$ when this optimal bias potential is acting on the system, hence the name target CV distribution. Most straightforward choice is to consider a uniform target distribution, $p(\mathbf{s}) \propto k$, although this is not always the most efficient choice, as we see below.
The variational principle provided by Eq. 3 provides a rather interesting and novel viewpoint for tackling the sampling problem: we can enhance the sampling of the CVs and determine the FES $F(\mathbf{s})$ by minimizing the functional $\Omega[V]$ with respect to the bias potential $V(\mathbf{s})$. The method also allows for considerable flexibility as we can tailor the desired enhancement in sampling by choosing the target CV distribution $p(\mathbf{s})$.
In practice the minimization of $\Omega[V]$ is performed by assuming some functional form and associated parameter set $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ for the bias potential $V(\mathbf{s};\boldsymbol{\alpha})$ and minimizing the function $\Omega(\boldsymbol{\alpha}) = \Omega[V(\boldsymbol{\alpha})]$ with respect to the parameters. It is normally most suitable to consider a linear basis set expansion of the bias potential, $V(\mathbf{s};\boldsymbol{\alpha}) = \sum_k \alpha_k \cdot f_k(\mathbf{s})$, where $f_k(\mathbf{s})$ is some set of orthogonal basis functions, for example plane waves or Chebyshev polynomials. In this case the gradient $\Omega'(\boldsymbol{\alpha})$ needed for the optimization just involves averages of the basis functions over the biased ensemble and the target CV distribution, $\frac{\partial\Omega(\boldsymbol{\alpha})}{\partial\alpha_i} = -\langle f_i(\mathbf{s})\rangle_{V(\boldsymbol{\alpha})} + \langle f_i(\mathbf{s})\rangle_p$. A small number of terms in the linear expansion normally suffice as the FESs are generally rather smooth functions. This can be compared to metadynamics where the bias potential is represented as a sum of Gaussians and a much larger number of Gaussians are needed to give a comparably accurate description of the FES. Another interesting idea is to assume some physics-based model for the FES that depends on a small number of parameters and use that model as bias potential. Exploring this option will be a significant part of the project going forward.
For the optimization we have found most optimal so far to use an averaged stochastic steepest descent algorithm [19] where we consider two sets of parameters, the instantaneous ones

that are updated according to

$$\boldsymbol{\alpha}^{(n+1)} = \boldsymbol{\alpha}^{(n)} - \mu \left[\Omega'(\bar{\boldsymbol{\alpha}}^{(n)}) + \Omega''(\bar{\boldsymbol{\alpha}}^{(n)})[\boldsymbol{\alpha}^{(n)} - \bar{\boldsymbol{\alpha}}^{(n)}]\right], \quad (5)$$

and the averaged ones $\bar{\boldsymbol{\alpha}}^{(n)} = (n)^{-1} \sum_{k=0}^{n} \boldsymbol{\alpha}^{(k)}$, that are used to define the bias potential $V(\mathbf{s}; \bar{\boldsymbol{\alpha}}^{(n)})$ that acts on the system, and are used to obtain the gradient $\Omega'(\bar{\boldsymbol{\alpha}}^{(n)})$ and Hessian $\Omega''(\bar{\boldsymbol{\alpha}}^{(n)})$.

In the initial phase of the project, the main emphasis has been on improving the convergence behavior of the optimization procedure. For this, we have considered a well-tempered sampling of the CVs where the target distribution is taken as $p(\mathbf{s}) = \frac{e^{-\beta' F(\mathbf{s})}}{\int d\mathbf{s}\, e^{-\beta' F(\mathbf{s})}}$, where $\beta' = [k_B(T + \Delta T)]^{-1}$ and $F(\mathbf{s})$ is the FES at temperature T. This corresponds to sampling the CVs at a higher temperature $T + \Delta T$ and is the same distribution as asymptotically sampled in well-tempered metadynamics [10]. Here, the limit $\Delta T \to \infty$ corresponds to the uniform distribution. The well-tempered distribution is expected to improve the convergence as it focuses the sampling of the CVs to the most relevant regions in the phase space. For the variational approach, $F(\mathbf{s})$ is a priori unknown so, in order to achieve this target distribution, we have implemented a procedure where the $p(\mathbf{s})$ is iteratively updated, based on the latest estimate of $F(\mathbf{s})$. As shown in Fig. 6, the well-tempered target distribution leads to an impressive improvement in the convergence of the FES as compared to using a uniform target distribution. We believe based on this initial results that the usage of the well-tempered target distribution, and associated iterative procedure, will be highly advantageous when studying complex systems with the variational approach. A manuscript discussing the details of the iterative procedure for the well-tempered target distribution is in preparation and will be submitted soon to the *Journal of Chemical Theory and Computation*.

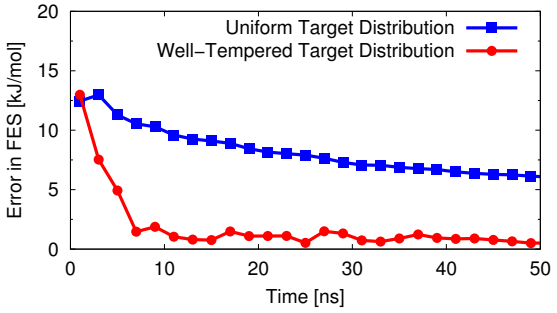


Figure 6: *Error in the free energy surface (FES) for a benchmark system of alanine tetrapeptide. The results are obtained with the variational approach using either an uniform (blue line) or a well-tempered (red line) target distribution. Employing an uniform distribution leads to a very slow convergence while convergence is significantly improved when employing a well-tempered distribution.*

We have also started to work on improving and extending our code for the variational approach with the aim of making it publicly available within the PLUMED 2 [20] enhanced sampling plug-in in the coming future. This involves making the code more robust and easier to use for the general practitioner. This also involves implementing new basis sets and functional forms for the bias potential. This work will continue into the next reporting period. Regarding this, we would like to note that our code is completely independent of the molecular dynamics code used as PLUMED 2 is compatible with a wide range of molecular dynamics codes, both classical and *ab initio*.

5.5 Planned research: Feb 2015 – Jan 2016

In the next phase of the project we will continue the work on extending and improving the code for the variational approach. We also aim to make publicly available in the PLUMED 2 plug-in the initial version of our code.

We also intend to explore the possibility of designing physics-based models for the FES and use these models as bias potentials within the variational approach. One of the research areas where we plan to pursue this option is for nucleation and growth of a crystal from solution and this will be done in collaboration with other members from the Parrinello group working on the MARVEL project (Pablo Piaggi, Claudio Perego, and Matteo Salvalaglio)

We furthermore plan to investigate how to obtain kinetics and rates from molecular simulations using the variational approach. This will be based on a recent work from our group where it was shown how to obtain kinetic information from metadynamics simulations [16, 17]. One of the main conditions for achieving this is that no bias is acting on the transition states between metastable states. This can be achieved within the variational approach by intelligently using the variational property of the method and the flexibility in the choice of the target distribution. We believe that this should allow us to obtain rates from molecular simulations in a shorter time than currently possible.

5.6 Personnel

NCCR personnel

- Omar Valsson, postdoc, 100%, from September 2014.

6 Topological data analysis for nonporous materials (Berend Smit — EPFL)

6.1 Research summary

The first steps have been made to apply topological data analysis methods to identify similarities between the best performing materials.

6.2 Research question

The objective of our research effort is to develop efficient data-mining techniques to discover sets of similar materials. While we have been quite successful in large-scale screening of these materials for different applications, we have made little progress in the data-mining aspects. For example, in our screening studies for materials for carbon capture or methane storage, our database of zeolite materials gives us the best performing zeolites for these applications. As these materials are in the all-silica form of the material, from a chemical point of view all materials have the same chemical composition and only differ in their pore topology. Once we have identified an optimally performing structure, we should be able to screen the database for structures with similar topologies. Initial studies have shown that simple descriptors such as maximum included sphere, maximum pore diameter, density, pore volume are not very successful in identifying structures that have similar performances. In this work we aim to develop alternative computational techniques that are based on topological data analysis methods. The idea is to use an image of the pore volume or energy landscape and use the Mapper methodology to reduce these high-dimensional data sets into so-called simplicial complexes with far fewer points, which can capture topological and geometric information at a specified resolution. These simplicial complexes are used, for example, in image recognition to identify similarities between objects.

6.3 Scientific goals

a) *Short-term* The short-term aims are to develop topological data analysis methods adapted to the study of nanoporous materials.

b) *Long-term* We intend to appy the developed topological data analysis methods to different screening studies and to extend the methodology to other systems.

6.4 Plans and results: May 2014 – Jan 2015

We started this project in November 2014. As a first step we will apply the Voronoi holograms method as developed by Martin *et al.* [21] to detect promising pore topologies. Fig. 7 shows an example of these representations. Viewing these holograms as points in a high-dimensional space, we will apply methods of topological data analysis. First of all, we will construct a set of projections of those high-dimensional vectors to $\mathbb{R}^2$ and $\mathbb{R}^3$ in order to verify if we can cluster the zeolites with respect to their ability to store gases. For this we will use standard Laplace-Beltrami operator as well as the Mapper software [22] to obtain various projections. If time permits, we will also try to use some standard machine learning techniques to analyze the obtained point cloud.

6.5 Planned research: Feb 2015 – Jan 2016

In the next period we plan to test whether the topological data analysis methods, in particular more sophisticated approaches to persis-

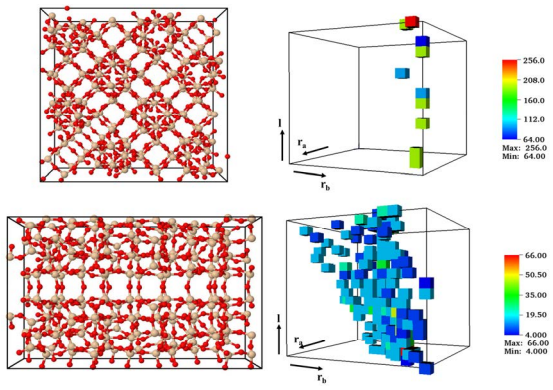


Figure 7: *Top left: the zeolite FAU structure. Top right: the Voronoi hologram for the zeolite FAU. Bottom left: the zeolite TUN structure. Bottom right: the Voronoi hologram for the zeolite TUN. In the Voronoi holograms, features are assigned a color based on the frequency of occurrence. r_a and r_b denote the radii of Voronoi nodes, and l represents the length between nodes.*

tent homology [23] such as persistence landscapes [24], applied to Voronoi holograms or directly to the IZA database of zeolite structures, will allow us to characterize different zeolite pore topologies. In addition, we will apply the method to different screening studies for materials that are optimal for carbon capture and other separations, as well as gas storage.

6.6 Synergies with other computational and experimental efforts

a) *Outside MARVEL* Part of this project will be carried out in collaboration with Pawel Dlotko (University of Pennsylvania) and Maciej Haranczyk (Lawrence Berkeley National Laboratory).

6.7 Personnel

NCCR personnel Dr. Yongjin Lee will start to work on this project as of February 2015.

Matching personnel

- Matthew Witman (based at UC Berkeley), PhD student, 50% from December 2014.

7 Machine-learning analysis of molecular materials (Michele Ceriotti — EPFL)

7.1 Research summary

This project endeavors to develop and apply enhanced sampling algorithms to explore the configuration space of molecular materials. We will start by developing and implementing analysis tools based on machine-learning, which we will use to automatically recognize bonding patterns and to build a coarse-grained representation of the accessible configuration space. In a second stage, these techniques will be applied to organic-inorganic interfaces, molecular materials and pharmaceutical compounds.

7.2 Research question

The main objective of this line of research is to investigate how machine-learning techniques can be used to simplify the description of complex bonding patterns in materials that exhibit a complex structural landscape. This includes both the automatic recognition of bonding patterns in the material and the coarse-grained representation of the configurational landscape based on non-linear dimensionality reduction of a high-dimensional representation of the actual structure. We will focus in particular on molecular materials, with the final goal of providing the rational foundations for controlling the crystal structure of drugs and small organic molecules [25, 26].

7.3 Scientific goals

a) *Short-term* The objective for the first phase of this project involves the development of a set of tools that can be used to streamline the analysis of atomistic simulations of complex materials, with a particular but non-exclusive focus on molecular compounds. These tools will be made available in open-source packages such as PLUMED [20], so as to simplify their use by other work packages in MARVEL, and by the scientific community at large. We expect that these developments will be highly coordinated with other work within HP4, and there is room for constructive synergies with HP5 and PP6.

b) *Long-term* The long-term objective of this project involves using this set of tools to study structural transitions in polymorphic materials, in particular (relatively) small molecular compounds such as those that are often used as pharmaceuticals.

7.4 Plans and results: May 2014 – Jan 2015

Non-linear dimensionality reduction techniques that we have been developing in the last few years (sketch-map [27, 28, 29]) can be very useful to describe in a coarse-grained but very detailed way the configurational stability of a complex material. However, a few aspects limit the applicability of these techniques:

- the fact that they have not yet been implemented in an efficient and robust way;
- the fact that they are not easily accessible by the users of different codes, particularly those based on electronic-structure methods that have limited capabilities in terms of free-energy methods;
- the fact that sketch-map variables do not have a well-defined physical meaning, and so still have to rely on a preliminary, physically-motivated, representation of atomic configurations.

In this first reporting period we have begun to address these challenges. With the help of

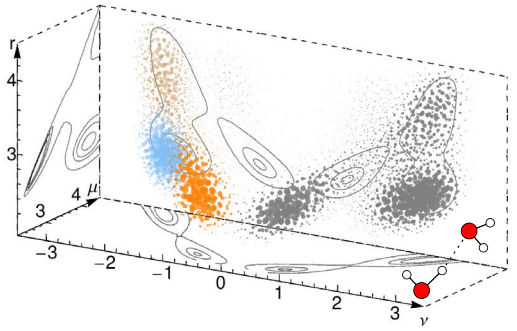


Figure 8: *Probability distribution for putative hydrogen-bonded configurations involving two oxygen and one hydrogen atom, in an* ab initio *simulation of liquid water at 300 K, using a combination of correlated noise and path integral molecular dynamics to simulate the quantum nature of light nuclei [30]. Data points are colored according to the PAMM clustering: the orange hue corresponds to configurations that have been identified as hydrogen-bonded.*

Piero Gasparotto, a PhD student in the group who has been dedicated to this project while the hiring of a postdoctoral researcher is being pursued, we have developed a probabilistic analysis of molecular motifs (PAMM) framework, that can be used to recognize recurring atomic patterns in a simulation [4].

The idea behind PAMM is to extract descriptors for groups of atoms from an atomistic simulation (e.g. the pairwise distances between groups of atoms of the chosen species), and analyze how often each atomic arrangement occurs in the simulation. A kernel-density estimate of the probability distribution is used to recognize the most frequent (therefore stable) arrangements, and to partition configuration space in disjoint clusters that can be associated with each mode of the distribution. Finally, based on such partitioning, one can work out an unbiased, fuzzy order parameter that probabilistically assigns each molecular motif to one of the clusters.

Fig. 8 demonstrates the application of PAMM to the definition of a structural descriptor of hydrogen bonding that can also be used in the context of the (transient) bond-breaking events that are observed in an *ab initio* simulation of water that includes nuclear quantum effects. Each triplet of two oxygen atoms and one hydrogen is considered as a putative hydrogen bond, and is described based on the combination of distances ($(\nu = d(O,H) - d(O',H)$; $\mu = d(O,H) + d(O',H)$; $r = d(O,O')$). Within the probability distribution of these coordinates, one can recognize clearly multiple recurring patterns, one of which can be clearly associated with the hydrogen bond. Since the shape of the probability distribution depends on the potential energy model, the thermodynamic conditions, and the kind of bond considered, this definition can adapt to the broad range of conditions that are characteristic of the hydrogen bond.

This application demonstrates how machine learning can be used to construct *physically motivated* order parameters, that can describe local order in a complex material. A combination of these local indicators could then be used as the input to non-linear dimensionality reduction techniques, a line of research that will be further investigated in the near future.

7.5 Planned research: Feb 2015 – Jan 2016

The procedure for obtaining a work permit is currently underway to hire a postdoctoral researcher (Dr. Sandip De) to focus on the development of sketch-map and its application to molecular materials. It is expected that Dr. De will join COSMO in early March, and will immediately begin working on an improved code to perform sketch-map analyses, possibly collaborating with Prof. Gareth Tribello (Queen Mary University Belfast) and with Prof. Volkan Cevher (EPFL - IEL) on fundamental improvements to the optimization algorithms that underlie different parts of the methodology.

Piero Gasparotto will continue to work on the development of PAMM and on the study of molecular materials, albeit probably he will focus on different aspects of the problem, most notably the role of quantum nuclei in modulating the stability of hydrogen bonds in such compounds.

7.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Collaboration with Dr. Max Amsler, from the group of Stefan Goedecker, has already led to a — literal — interface between the work in this project and the structure search endeavor which is underway in Basel. There is ongoing discussion with the group of Nicola Marzari to ensure that the software infrastructure developed within this project will be fully compatible with the AiiDA platform. Collaboration with the group of Michele Parrinello is focussing on the application of sketch-map analysis to polypeptides [5].

7.7 Personnel

NCCR personnel

- Piero Gasparotto, PhD student, 100%, from May 2014.

8 "Alchemical" coupling and predictions (Anatole von Lilienfeld — UniBas)

8.1 Research summary

Predicting property trends across chemical compound space is at the heart of the computational design of new materials. We have carried out numerical calculations for various crystalline systems to establish a basis for a series of computational experiments involving "alchemical" coupling between materials. If successful, derivatives with respect to coupling parameters might offer glances at materials compound space with unprecedented efficiency.

8.2 Research question

We investigate how to exploit the use of "alchemical" coupling approaches for the efficient prediction of property trends across chemical compound space. This approach holds great promise to dramatically enhance the sampling of materials space, thereby enabling us to more rapidly zoom into those materials that exhibit particularly interesting combinations of properties.

8.3 Scientific goals

a) *Short-term* The short-term goal consists of assessing the accuracy and predictive power of first-order Hellmann-Feynman derivatives [31, 32] of first-principles properties in the compositional space of materials [33]. To this end, we first plan to investigate simple materials, such as alkali-halogenide crystals. Thereafter we will proceed to more challenging systems.

b) *Long-term* In the long run, we plan to use first-order derivatives, and possibly also higher order derivatives if necessary, to simultaneously explore multitudes of compounds, and to iteratively converge towards optimal properties through gradient-based optimization methods.

8.4 Plans and results: May 2014 – Jan 2015

a) *Hires* A postdoctoral student, Alisa Solovyeva was hired, and started her work in September 2014. Alisa graduated in the group of Prof. J. Neugebauer at the University of Münster in Germany. Her PhD work deals with the development of embedded electron density methods.

b) *Theory* The potential energy of a material, the most basic property one would like to control, can be Taylor expanded in the space of all possible compounds and materials,

$$E(\lambda) = E(\lambda_0) + \left.\frac{\partial E}{\partial \lambda}\right|_{\lambda_0} \Delta\lambda + \frac{1}{2}\left.\frac{\partial^2 E}{\partial \lambda^2}\right|_{\lambda_0} \Delta\lambda^2 + HOT, \quad (6)$$

where we have truncated the higher order terms (HOT) after the second order derivative term. The coupling parameter λ linearly relates any two iso-electronic Hamiltonians. This includes systems with identical number of valence electrons, such as NaCl and KCl. In this case, the Hamiltonian reads,

$$\hat{H}(\lambda) = \hat{H}_{\mathrm{NaCl}} + \lambda(\hat{H}_{\mathrm{KCl}} - \hat{H}_{\mathrm{NaCl}}), \quad (7)$$

and its first-order derivative is given analytically by virtue of the Hellmann-Feynman theorem [31, 32],

$$\frac{\partial E}{\partial \lambda} = \left\langle \hat{H}_{\mathrm{KCl}} - \hat{H}_{\mathrm{NaCl}} \right\rangle, \quad (8)$$

where $\langle ... \rangle$ corresponds to the Bra-Ket notation of the quantum mechanical observable. In practice, energies and derivatives for the interpolating Hamiltonian can be calculated by simply interpolating the external potential, i.e. through pseudopotential parameter interpolation.

c) *Energies of alkali-chloride salt crystals* As first step towards the implementation of derivative guided design of materials, we have carried out density functional theory calculations for the fcc (8 atoms supercell) and bcc (16 atoms supercell) crystals of NaCl, KCl, CsCl, and RbCl at the Γ-point only. Basis-set convergence runs of energies as a function of lattice constants have been performed using the pseudopotential plane-wave based CPMD code [34], the PBE functional [35], and Goedecker-Teter-Hutter pseudopotentials, as parameterized by Krack [36]. All alkalide pseudopotentials have

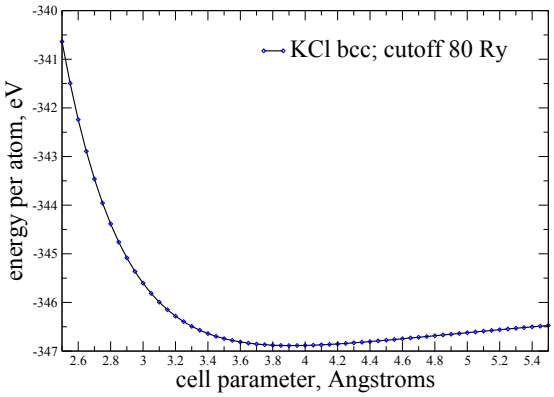


Figure 9: *Calculated total energy of KCl as a function of its bcc lattice constant. The energy corresponds to a plane-wave cutoff of 80 Ry.*

been used with nine valence electrons, while the chloride pseudopotential has seven valence electrons.

Plane-wave convergence — In order to interpolate materials with different chemical composition it is important to ensure that the employed basis-set is converged throughout. For bcc KCl, for example, a smooth and converged energy curve has been obtained already at 80 Ry (Fig. 9).

For bcc NaCl, however, convergence is much slower, with cutoffs of neither 80 Ry, nor 180 Ry, nor 220 Ry (not shown) being sufficient, as shown in Fig. 10. Only after increasing the cutoff to 250 Ry, similar smoothness is found as in the case of KCl at 80 Ry. Slower convergence is to be expected for atoms occurring earlier in the periodic table: Their repulsive core region in the pseudopotential is smaller, therefore requiring more plane waves to account for the more localized electronic oscillations near the core.

For CsCl and RbCl, one expects even faster convergence than for KCl. Also in the case of the fcc crystals, faster convergence has been observed for all systems, including NaCl. For all future numerical studies, we therefore plan to use at least a 250 Ry cutoff in order to ensure converged results.

Compute needs — Due to the slow plane-wave basis convergence of systems with hard pseudopotential for small ionic cores our CPU needs are significant. For example, in the case of the 250 Ry calculations, necessary to reach convergence for NaCl, a single point on the bottom curve in Fig. 10 cost $\approx$ 36 CPU hours on a modern compute node.

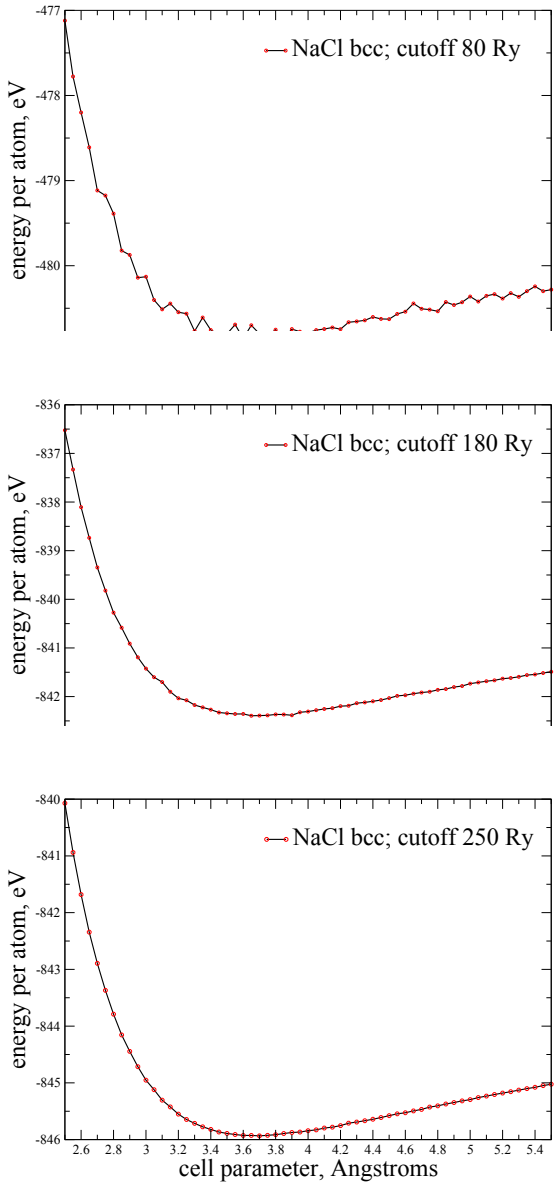


Figure 10: *Calculated total energy of NaCl as a function of its bcc lattice constant, for various plane-wave cutoff values in increasing order from top to bottom. Convergence is reached at 250 Ry (bottom).*

8.5 Planned research: Feb 2015 – Jan 2016

a) *First-order derivatives* We plan to investigate the predictive power of above discussed first order derivatives for predicting energy differences between bcc and fcc-phases. This promises to be useful for the rapid scanning of materials phase diagrams.

b) *Other crystals* We plan to extend our calculations to also include other chemical elements such as halides. We will also consider Zincblende crystal structures, and possibly Wurtzite.

8.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* In collaboration Stefan Goedecker (UniBas), we have started to perform preliminary test calculations for alanate crystals, recently explored using enhanced sampling methods developed in the Goedecker group [37].
In collaboration with Ursula Röthlisberger (EPFL), we have started numerical explorative studies of optimizing the composition of training sets for obtaining more efficient machine learning models of molecular properties [38].

b) *Outside MARVEL* In collaboration with Dennis Andrienko at the Max Planck Institute for Polymer Research (Mainz) we have started to perform preliminary test calculations for the application of the alchemical coupling approach towards the design of superior molecular electronic devices.

8.7 Personnel

NCCR personnel

- Alisa Solovyeva, postdoc, 100%, from September 2014.

Matching personnel

- Diana N. Tahchieva, PhD student, 100%, from January 2015.

9 Development of an evolutionary algorithm tool set for chemical compound and materials design (Ursula Röthlisberger — EPFL)

9.1 Research summary

We are developing and implementing evolutionary algorithms for an efficient search of chemical space enabling accelerated materials and chemical compound design. During the first year, we have implemented a first version of a flexible and modular package called EVOLVE. As first test and performance validation studies, we have applied EVOLVE to the design of a helical peptide [6] and, in collaboration with the group of Anatole von Lilienfeld, we have worked for the optimization of machine learning training sets [7].

9.2 Research question

Genetic algorithms (GAs) are a well established family of evolutionary algorithms. Selection, mutation and crossover (reproduction) operators iteratively act upon the genetic encoding of a population of trial solutions and are able to identify near-optimal elements in huge optimization spaces. Although there are a vast number of successes in the application of GAs, there are relatively few applications of such optimization techniques in the context of electronic structure applications. With the advances in computer power, it has now become feasible to apply GAs even with quantum mechanical fitness evaluations, enabling, e.g., the design of molecular catalysts and materials with tailored properties. In this project, we are developing a general and highly modular GA engine with the goal of automatizing the computational search for optimal molecular and perovskite-based light absorbers in solar cells.

9.3 Scientific goals

a) *Short-term* During the first year, we have implemented an efficient and highly flexible GA toolbox (EVOLVE) and have tested its performance of two vastly different optimization problems. As a first validation study, we have applied a single-objective GA code within EVOLVE to a single 20 amino acid long *alpha* helix in different dielectric environments. We have optimized the sequence of the central eight amino acids using a library of 177 conformational rotamers, resulting in a space of $177^8 \sim 10^{17}$ possibilities. The fitness function in this case was based on a molecular mechanics evaluation of the helical stability. We also introduced a form of genetic memory that led to a large convergence acceleration. Our results show that EVOLVE is achieving fast optimization, is converging and is proficient in finding low-energy sequences (Fig. 11).

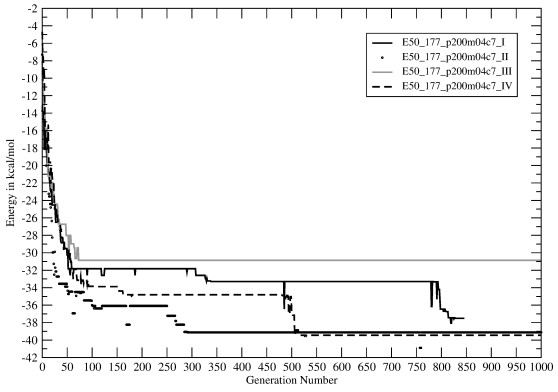


Figure 11: *Fittest individual as a function of generation for four independent runs.*

In a second application, in a collaboration with the group of von Lilienfeld, we used EVOLVE to optimize training sets for a machine learning (ML) model to solve the electronic Schrödinger equation. The initial training pool from which training sets were chosen contained 6095 molecules. Finding an optimal set of, e.g., 1000 molecules from this pool corresponds to a huge optimization space (i.e. $4.5 \cdot 10^{1179}$ combinations!). We find that application of the GAs to a training set of given size improves the performance of the ML model substantially, when compared to a ML model trained on randomly selected molecules coming from the GDB-17 subspace of organic molecules [39].

b) *Long-term* We will further develop EVOLVE to include multi-objective GAs and diverse particle swarm optimization algorithms, and apply it to the design of optimal dye-sensitizers with tailored light absorption and redox properties. We will also apply it to the optimization of perovskite-based solar cells to accelerate the search for materials with ideal band gap and transport properties. The predicted systems will be tested experimentally in the group of Michael Graetzel at EPFL.

9.4 Plans and results: May 2014 – Jan 2015

In this first phase, we have started the development of a highly flexible artificial intelligence toolset with the ultimate goal of facilitating the search for compounds and materials with optimal properties. This first implementation has been evaluated on two very different search problems, the design of optimal biological molecules and the choice of training molecules for machine learning. For both applications, papers are currently in preparation [6, 7].

9.5 Planned research: Feb 2015 – Jan 2016

In the second year, we will generalize EVOLVE for the genetic optimization of general chemical compounds and materials and extend it to multi-objective optimizations that will be especially useful for, e.g., design of dye sensitizers where various properties (that often mutually exclusive) have to be optimized. We will also implement the possibility of particle swarm optimizations and test the relative merits of these two categories of evolutionary algorithms. They will be applied to the design of porphyrin-based dyes with optimal optical and redox properties. To this end, EVOLVE will be coupled with a quantum mechanical fitness evaluation for a direct solution of the electronic Schrödinger equation, or as a continuation of the ongoing collaboration with the von Lilienfeld group with a ML model. In this phase of the project, we will also build up the chemical libraries for possible compound modifications.

9.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Part of this project is a collaboration with Raghunathan Ramakrishnan and Anatole von Lilienfeld (UniBas).

b) *Outside MARVEL* The project is done in close collaboration with the experimental group of Michael Graetzel at EPFL.

9.7 Personnel

NCCR personnel

- Nicholas John Browning, PhD student, 80%, from May 2014.

Matching personnel

- Marta Andreia Da Silva Perez, postdoc, 50%, from May 2014.

MARVEL-related publications

List of publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.

- [1] M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello, *Molecular-dynamics simulations of urea nucleation from aqueous solution*, Proceedings of the National Academy of Science of the USA **112**, E6 (2015).
- [2] M. Salvalaglio, M. Mazzotti, and M. Parrinello, *Urea homogeneous nucleation mechanism is solvent dependent*, Faraday Discussions (2015), doi:10.1039/C4FD00235K.
- [3] O. Valsson and M. Parrinello, *Variational Approach to Enhanced Sampling and Free Energy Calculations*, Physical Review Letters **113**, 090601 (2014).
- [4] P. Gasparotto and M. Ceriotti, *Recognizing molecular patterns by machine learning: An agnostic structural definition of the hydrogen bond*, The Journal of Chemical Physics **141**, 174110 (2014).
- [5] A. Ardevol, G. Tribello, M. Ceriotti, and M. Parrinello, *Probing the unfolded configurations of a beta hairpin using sketch-map, submitted* (2014).
- [6] M. A. S. Perez, N. J. Browning, E. Brunk, P. Anthi, and U. Rothlisberger, *A Genetic Algorithm for the Design of Peptides and Proteins, in preparation* (2015).
- [7] N. J. Browning, R. Ramakrishnan, A. O. von Lilienfeld, and U. Rothlisberger, *Addressing the selection bias: Genetically optimized molecular data sets for the training of machine learning models of molecular enthalpies of atomization, in preparation* (2015).

Other references

[8] S. Goedecker, *Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems*, The Journal of Chemical Physics **120**, 9911 (2004).

[9] M. Ceriotti, J. More, and D. E. Manolopoulos, *i-PI: A Python interface for ab initio path integral molecular dynamics simulations*, Computer Physics Communications **185**, 1019 (2014).

[10] A. Barducci, G. Bussi, and M. Parrinello, *Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method*, Physical Review Letters **100**, 020603 (2008).

[11] A. Barducci, M. Bonomi, and M. Parrinello, *Metadynamics*, WIREs Computational Molecular Science **1**, 826 (2011).

[12] M. Salvalaglio, T. Vetter, F. Giberti, M. Mazzotti, and M. Parrinello, *Uncovering molecular details of urea crystal growth in the presence of additives*, Journal of the American Ceramic Society **134**, 17221 (2012).

[13] M. Salvalaglio, T. Vetter, M. Mazzotti, and M. Parrinello, *Controlling and Predicting Crystal Shapes: The Case of Urea*, Angewandte Chemie International Edition **52**, 13369 (2013).

[14] F. Giberti, M. Salvalaglio, M. Mazzotti, and M. Parrinello, *Insight into the nucleation of urea crystals from the melt*, Chemical Engineering Science **121**, 51 (2014).

[15] M. Salvalaglio, F. Giberti, and M. Parrinello, *1, 3, 5-Tris (4-bromophenyl) benzene prenucleation clusters from metadynamics*, Acta Crystallographica C **70**, 132 (2014).

[16] P. Tiwary and M. Parrinello, *From Metadynamics to Dynamics*, Physical Review Letters **111**, 230602 (2013).

[17] M. Salvalaglio, P. Tiwary, and M. Parrinello, *Assessing the Reliability of the Dynamics Reconstructed from Metadynamics*, Journal of Chemical Theory and Computation **10**, 1420 (2014).

[18] A. Laio and M. Parrinello, *Escaping free-energy minima*, Proceedings of the National Academy of Science of the USA **99**, 12562 (2002).

[19] F. Bach and E. Moulines, *Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$*, in *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, eds. (Curran Associates, Inc., 2013), p. 773.

[20] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *PLUMED 2: New feathers for an old bird*, Computer Physics Communications **185**, 604 (2014).

[21] R. L. Martin, B. Smit, and M. Haranczyk, *Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials*, Journal of Chemical Information and Modeling **52**, 308 (2012).

[22] G. Singh, F. Mémoli, and G. Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, in *Eurographics Symposium on Point-Based Graphics*, M. Botsch and R. Pajarola, eds. (2007).

[23] H. Edelsbrunner and J. L. Harer, *Computational Topology, An Introduction* (American Mathematical Society, 2010).

[24] P. Bubenik, *Statistical topological data analysis using persistence landscapes*, arXiv:1207.6437 (2014).

[25] S. S. L. Price, *Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism*, Accounts of Chemical Research **42**, 117 (2009).

[26] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, and D. E. Williams, *A test of crystal structure prediction of small organic molecules*, Acta Crystallographica B **56**, 697 (2000).

[27] M. Ceriotti, G. A. Tribello, and M. Parrinello, *From the Cover: Simplifying the representation of complex free-energy landscapes using sketch-map*, Proceedings of the National Academy of Science of the USA **108**, 13023 (2011).

[28] G. A. Tribello, M. Ceriotti, and M. Parrinello, *Using sketch-map coordinates to analyze and bias molecular dynamics simulations*, Proceedings of the National Academy of Science of the USA **109**, 5196 (2012).

[29] M. Ceriotti, G. A. Tribello, and M. Parrinello, *Demonstrating the Transferability and the Descriptive Power of Sketch-Map*, Journal of Chemical Theory and Computation **9**, 1521 (2013).

[30] M. Ceriotti and D. E. Manolopoulos, *Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei*, Physical Review Letters **109**, 100604 (2012).

[31] R. P. Feynman, *Forces in Molecules*, Physical Review **56**, 340 (1939).

[32] O. A. von Lilienfeld, *Accurate* ab initio *gradients in chemical compound space*, The Journal of Chemical Physics **131**, 164102 (2009).

[33] O. A. von Lilienfeld, *First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties*, International Journal of Quantum Chemistry **113**, 1676 (2013).

[34] J. Hutter *et al.*, *Computer code* CPMD, *v 3.17*, © IBM Corp. 1990-2008, © MPI-FKF Stuttgart 1997-2001.

[35] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Physical Review Letters **77**, 3865 (1996).

[36] M. Krack, *Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals*, Theoretical Chemistry Accounts **114**, 145 (2005).

[37] T. D. Huan, M. Amsler, M. A. L. Marques, S. Botti, A. Willand, and S. Goedecker, *Low-Energy Polymeric Phases of Alanates*, Physical Review Letters **110**, 135502 (2013).

[38] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Fast and accurate modeling of molecular atomization energies with machine learning*, Physical Review Letters **108**, 058301 (2012).

[39] L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, *Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17*, Journal of Chemical Information and Modeling **52**, 2864 (2012).

Horizontal Project 5 HP5 — Materials Informatics

Project leader: Alessandro Curioni (IBM)

Participating members: Alessandro Curioni (IBM), Christoph Koch (EPFL)

Summary and highlights: The work in Horizontal Project 5 (HP5) focused on two main goals: the design of a big-data system to handle the needs of the MARVEL applications and the design of domain-specific language frameworks that will allow the easy development of high-performant applications in the context of MARVEL. We made significant progress for this first year especially at the design and planning levels. Early algorithmic advances already show very good potential and hold promise for significant impact in the near future.

General view of the project

Major research questions

The key research question for the first year of the project is to set the outline and the general objectives of how can we leverage computer science research, and specifically big-data-systems research, to support the efforts of the MARVEL community. In particular, we focused on two main fronts:

1. The design aspects of a system for big data based material discovery, in synergy with AiiDA. In particular, to goal was to identify a pipeline for the data flow with clear stages that allow us to focus on key aspects of the system. The goal is to integrate this pipeline within the material discovery platform and test/validate it on the project of Material Screening for Metal/Air Batteries [1, 2] part of the Vertical Project 2.

2. The design of specialized high-level programming and scripting languages for the computational materials science community that make MARVEL teams more productive in their coding efforts. How can we make programs developed in such languages run highly efficiently on high-performance computing platforms.

Short- and long-term goals

In terms of the big data system, the short-term goals focus on the development of a data acquisition and data curation layer that is capable of handling large scale data (i.e. thousands of documents). Long term goals focus on the knowledge creation layer and advanced analytics algorithms. In particular, emphasis is given on graph analytics algorithms and query implementation. In addition, we have commenced work on scalable and low-cost machine learning methods that are based one primal-dual methods (an improved form of the augmented Lagrangian method) and on neural networks techniques.

Short-term goals in the context of the second axis target at the creation of an extendible compiler framework for domain-specific languages (DSLs). Long term goals deal with the application and tuning of this framework on the particular case of the MARVEL partners. In particular, offering DSL support for Python is deemed crucial.

First achievements

For the big data system, in terms of the data input stage, we used standard text annotators and content extraction softwares based on the Apache UIMA framework. For the data curation and storage phase we will use relational databases such as IBM DB2 and NoSQL DBs such as Titan which is a graph DB. We have already working on implementations of the first two stages based on the above technologies. In addition, we are working on high-performance implementations both leveraging multicore machines and for data center architectures (using the Hadoop MapReduce and Apache Spark frameworks). Optimal solutions and the experience from this project could then be incorporated into the AiiDA framework.

With respect to the DSL work we have been developing a compilation framework for DSLs that allows to plug in very powerful domain-specific compiler optimizations to produce fast code.

Next steps

The analytics research is focusing on the development of low-complexity graph analytics algorithms. In particular, we are developing close to linear cost methods for the estimation of node centrality measures such as sugbraph node centralities, alpha-centrality and in betweenness centrality.

The compiler research will build upon previous work on the popular Scala framework. We expect finalizing the core compiler framework will take most of the coming reporting period. For this first compiler release, the input language is Scala, and the output is highly optimized C code; supporting further front-end languages such as Python will be future work.

Research contributions to the overall goals of the NCCR and to the existing literature

Key publications with regards to the analytics work include [4, 5, 6, 7]. The paper [3], currently under submission, describes the design of our prototype DSL compiler infrastructure.

Collaborative components

The analytics work has benefited from collaborations with the FP7 EU projects EXA2GREEN and NANOSTREAMS for energy aware analytics algorithms and streaming analytics. The DSL work is already engaged with other members of the MARVEL community to define the specifics of the high-level applications. Moreover, the team has reached out to other groups within EPFL. An effort to build a database management system using the DSL developments is already underway.

1 Big data based material discovery (Alessandro Curioni — IBM)

1.1 Research summary

- Data driven knowledge discovery.
- Big data.
- Graph analytics.

1.2 Research question

During the first year of the project we engaged on the design aspects of a system for big data based material discovery. In particular, we identify a pipeline for the data flow with clear stages that allow us to focus on key aspects of the system. The goal is to integrate this pipeline within the material discovery platform and test/validate it on the project of Material Screening for Metal/Air Batteries [1, 2] part of the Vertical Project 2.
The pipeline consists of the following four phases:

- **Data input:** this stage consists of systems that are able to gather data from a variety of sources, such as scientific publications, patents, and in various formats, structured and unstructured, such as text, images or simulation results.

- **Data storage and curation:** this stage uses big data storage systems to store and curate the data. Both relational (SQL) and not only relational (NoSQL) databases need to be used as the nature of the data is such that not any one solution is adequate on its own.

- **Knowledge creation:** this stage uses advanced machine learning algorithms and graph analytics to link the data and create knowledge graphs that hold the knowledge in the data. This stage creates the data model in such a way that it can support intelligent questions (queries) from the users.

- **Query formation:** finally, intelligent questions (i.e. high-level questions close to the user nomenclature and lingo) are mapped to algorithms that run on the data model. The results are then analyzed and transformed in such a way that they are easily consumable by the users.

1.3 Scientific goals

We have conducted a first analysis and design search for all four phases. However, since the knowledge creation stage is the one with the highest complexity and computational cost, we gave particular emphasis to it. Our goal is to come up with a first implementation of a prototype for the full pipeline as soon as possible.

a) *Short-term* To develop a data acquisition and data curation layer that is capable of handling large scale data (i.e. thousands of documents).

b) *Long-term* Develop the knowledge creation layer and advanced analytics algorithms. In particular, we will focus on graph analytics algorithms and query implementation.

1.4 Plans and results: May 2014 – Jan 2015

For the data input stage we used standard text annotators and content extraction software based on the Apache UIMA framework. For the data curation and storage phase we will used relational databases such as IBM DB2 and NoSQL DBs such as Titan which is a graph DB. We have already working on implementations of the first two stages based on the above technologies.

1.5 Planned research: Feb 2015 – Jan 2016

a) *Data model* We use knowledge graphs to represent the knowledge in the input data. The nodes of the knowledge graph represent entities such as particular materials, processes, simulations, chemical elements and documents. Each node has several attributes that can hold structured and unstructured data (e.g. chemical composition, results of experiments). The exact number and nature of the attributes are parametrizable for the application at hand. The edges that connect nodes on the graph represent relationships such as chemical similarity, co-mention in a publication, or use of similar production methods. Edges can have weights to encode the different importance of relations or confidence in the data that supports the relation.

b) *Graph analytics and low-complexity kernels* In order to implement intelligent queries, we need to perform advanced graph analytics such as graph similarity, graph simplifications and node importance [4, 5]. In addition, we worked and continue to work on developing low-complexity algorithms for the fundamental linear algebra kernels that are heavily used in these applications [6, 7]. To this end, we developed novel algorithms for:

- subgraph centrality, which is a powerful generalisation of the page rank principle. We are working on a novel algorithm that reduces the $O(N^3)$ cost of the original method to close to linear order. Thus, we can now treat very large scale problems or a large number of smaller ones in parametric studies;
- graph comparison. We developed a linear cost method for the comparison of any two graphs. The method calculates what is an equivalent of a density of states for the adjacency matrix of the graph. Then, the density of states plots can easily be compared.

For the second year of the project our goal is to develop a complete end to end implementation of the knowledge discovery pipeline and apply it to the search of potential new solvents for metal-air batteries, eventually comparing this approach with the machine learning approach proposed by Anatole von Lilienfeld (HP4). This will require work on the fourth stage (query implementation and intelligence consumption).

1.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Collaboration with HP4 and VP2 for the development of machine learning algorithms for materials discovery. Collaboration with PP6 to introduce optimal solutions in the AiiDA framework.

b) *Outside MARVEL* Collaborations with the FP7 EU projects EXA2GREEN and NANOSTREAMS for energy aware analytics algorithms and streaming analytics.

1.7 Personnel

NCCR personnel

- Peter Staar, postdoc, 75%, from January 2015.

Matching personnel

- Costas Bekas, research staff member, 25%, from May 2014.

2 Domain-specific languages and compilation (Christoph Koch — EPFL)

2.1 Research summary

- Build a prototype extensible compiler infrastructure for domain-specific languages.
- Allow to easily add domain-specific compiler optimizations for generating highly efficient low-level code.

2.2 Research question

How can we design specialized high-level programming and scripting languages for the computational materials science community that make MARVEL teams more productive in their coding efforts? How can we make programs developed in such languages run highly efficiently on high-performance com-

puting platforms?

2.3 Scientific goals

Software libraries for a particular domain are traditionally written as either object-oriented components or subroutine libraries. However, an alternative development methodology, which has recently obtained much attention, is to define a new programming language for the particular problem domain. This category of languages is known as domain-specific languages (DSLs). By using DSLs, programs can be written using the idioms of their particular domain, a fact which improves the productivity of developers as well as the maintainability of software programs themselves. However, this comes at a price, as developing DSLs has numerous disadvantages such as the cost of educating users and the initial cost of designing and implementing a DSL itself.

Like every other programming language, a DSL needs the development of a full-fledged compiler for it. A compiler consists of two main components: 1) a front end which is responsible for checking the syntactic and semantic correctness of the given DSL program, and 2) a back end which is responsible for optimizing the program and generating either machine code or code in another programming language.

The front end of host language compiler can be reused as it is for developing the DSL by re-using the syntax of a host programming language. A DSL developed using this approach is known as an embedded DSL (EDSL). The programming language Scala, developed at EPFL and currently becoming extremely popular all over the world, provides appropriate infrastructure for developing EDSLs inside it. By using Scala as the host language, EDSLs are benefiting from the advanced software compositional features provided by this multi-paradigm programming language. The programs written using these EDSLs use the same runtime system as the Scala programming language, i.e. the Java virtual machine (JVM). Hence, the performance of such a DSL is highly dependent on the JVM. More specifically, our experience shows that, in most cases, this dependence causes overhead that is simply not acceptable from system developers.

Extensible optimizing compilers are compiler back ends that are used to resolve this performance issue. Such compilers provide interfaces in order to specify optimizations that can be applied to a certain DSL. In addition, extensible optimizing compilers provide interfaces for generating the code of a different programming language. This way, a program written in a high-level DSL is compiled into a program in a low-level language (such as C). Furthermore, the generated program can reuse the existing libraries written in its own language (for example, if the generated program is in the Python programming language, we can reuse the numerous libraries written in Python for the generated code).

In our experience, existing extensible optimizing compilers are not yet adequate for high-performance computing. The fundamental reason for this is that such systems basically force developers to become compiler experts before developing any software. Thus, in this project, we investigate the infrastructure needed for easy development of DSLs. This considers both the frameworks for facilitating the embedding of DSLs as well as the frameworks for optimizing DSLs.

Many people involved in MARVEL develop code in one way or the other. Increasingly it is code in scripting languages that has the purpose of doing pre- and post-processing or providing glue code between several existing code bases. Scripting languages are known to be rather inefficient, and their purpose is often not just to call code in more efficient languages such as C and Fortran, but to move substantial amounts of data between these more efficient and tuned code bases, and to transform this data. Such computations should be made efficient in a high-performance computing (HPC) environment. While there exists expertise on optimizing and tuning HPC code in this community, it is a work-intensive and difficult task. General-purpose compilers cannot be relied upon to generate low-level code competitive in performance with expertly hand-optimized code. However, there is recent progress that suggests this can be done well for more restricted languages.

a) *Short-term* We are developing an extensible compiler framework for DSLs.

b) *Long-term* We want to study the use of domain-specific languages (DSLs) and scripting languages in computational materials science, and develop specialized high-level and high-productivity programming languages for the computational materials science community. This will create real value for the MARVEL community in that it will make the various teams more productive in their coding efforts.

Some MARVEL teams, such as the AiiDA team and the group at UniGE, are already now using

Python as a host language for embedded DSL development.
Some of the code bases are performance-critical, so it is important to have compilers that automatically transform DSL programs into highly efficient code to run on high-performance computing platforms. We will embark on a long-term collaborative effort involving all interested MARVEL teams to co-design DSLs and compiler optimizations for our compiler framework, as their needs and challenges develop.
We will extensively evaluate the potential of domain-specific compilation to automatically generate HPC code that is competitive in performance with the best code achievable by expert human programmers. If another major effort to develop a performance critical quantum simulation or machine learning code base is started during the duration of MARVEL, we will attempt to get involved in this effort to see whether our techniques can lead to more high-level implementations at higher productivity, without a performance penalty.
One risk factor to the success of the DSL project is supporting the Python scripting language as a front-end language. Python is an untyped language, from which is difficult to automatically produce highly optimized code. This is a substantial and difficult research challenge. However, it is of wider interest in computer science. For example, JavaScript engines in web browsers and efficient implementations of the JVM (such as HotSpot) face the same problem that great performance benefits could be obtained if unavailable type information could be effectively discovered. In the case of Java, the type information is eliminated by the compiler and not available in the Java bytecode received by the JVM; in the case of JavaScript, the type information never existed. For that reason, work on this problem is currently starting internationally at a number of places.

2.4 Plans and results: May 2014 – Jan 2015

Since May 2014, Amir Shaikhha has been developing a compilation framework for DSLs that allows to plug in very powerful domain-specific compiler optimizations to produce code that is competitive in performance with the most efficient handwritten low-level code developed by HPC experts. The design of our prototype DSL compiler infrastructure is described in [3], currently under submission. We want to make this framework easy to use and to work with programming languages and scripting languages that members of the MARVEL community want to use.

2.5 Planned research: Feb 2015 – Jan 2016

Building a production-quality compiler usually is a multiple person-year effort. In this case we can profit from prior work at EPFL and the presence of arguably the most successful academic compilers research group in the world, the creators of Scala.
We expect finalizing the core compiler framework will take most of the coming reporting period. For this first compiler release, the input language is Scala, and the output is highly optimized C code; supporting further front-end languages such as Python will be future work. In the coming year, we will also start interacting closely with HPC computing experts from CSCS, to build domain-specific HPC code optimizations and test our framework against the code quality of expert HPC programmers.

2.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* The goal of this work is to make other MARVEL teams more productive in their coding efforts, while still allowing their codes to be efficient. This will have impact on all computational efforts in MARVEL, as long as teams are willing to work with us, and is particularly relevant to (HPC and other) software infrastructure and materials informatics projects, such as AiiDA, and our advanced analytics and machine learning efforts. We have been interacting on this with Thomas Schulthess, Nicola Marzari's team, and several of the machine learning teams. These interactions will have to be intensified in the coming months.

b) *Outside MARVEL* Our compiler infrastructure will be evaluated by other computer science groups at EPFL for developing DSLs and productively building efficient software systems, such as database management systems and software-defined networking systems. An effort to build a database management system in this way is already underway.

2.7 Personnel

NCCR personnel

- Amir Shaikhha, PhD student, 100%, from May 2014.

3 Materials databases and advanced analytics (Christoph Koch — EPFL, Volkan Cevher — EPFL)

3.1 Research summary

- First efforts to create a team on materials analytics and a pipeline of capable students into MARVEL.
- Started research on fast parallel implementations of machine learning algorithms.

3.2 Research question

How can we leverage computer science research, and specifically big data systems and machine learning research, to support the efforts of the MARVEL community?

3.3 Scientific goals

Machine learning bears the promise of opening up new opportunities for computational materials science, and several groups inside MARVEL are interested in it, including those of Alessandro Curioni, Anatole von Lilienfeld, Michele Ceriotti, and ours. In addition, Volkan Cevher from EPFL is a machine learning expert and has become interested in the problems posed within MARVEL through discussions with all the above-mentioned people. Although currently not funded by MARVEL, he has started collaborations with several of us.

3.4 Plans and results: May 2014 – Jan 2015

Since September, Cevher and Koch are working on getting two EPFL first-year PhD students, Edo Collins and Radu Ionescu, up to speed. We want to co-advise them on research in machine learning related to the MARVEL project. We are currently evaluating these two students, and might hire neither, one, or both of them. They are currently reading up on the research literature and taking an advanced machine learning course.
We have tasked them with implementing scalable versions of two different machine learning problems, a primal-dual method (an improved form of the augmented Lagrangian method), and a stochastic gradient descent method for neural network training. The techniques are of quite general interest, but we want to apply them to problems in computational materials science in the future. The students are working on scalable high-performance implementations, both leveraging multicore machines and for data center architectures (using the Hadoop MapReduce and Apache Spark frameworks). Radu Ionescu has recently finalized a working prototype of a parallel version of a primal-dual algorithm developed by Volkan Cevher.
These two students are on fellowships for their first year and thus cannot be paid yet by MARVEL.
Cevher and Koch are also using this opportunity to test a very close collaboration where Cevher contributes the cutting edge machine learning expertise and Koch contributes on the systems and scalability side. Koch's senior PhD student, Milos Nikolic, has helped coach the new students on implementing parallel algorithms.

a) *Attracting Students Through Integrated Teaching and Research.* In the spring term of 2015, Koch will be teaching the graduate course *Big data* at EPFL, which comes with a significant student team project. He is currently evaluating whether MARVEL-related machine learning, analytics, or database problems (including challenges in the AiiDA project) can be given as course projects, which might be a worthwhile investment to attract suitable students for several of the MARVEL teams.

b) *Publications* So far, our team is in the process of formation and has not published MARVEL papers. The two full-time PhD students are part of the incoming class of fall 2014 and are just getting started. The publications [8, 9, 10, 11, 12], however, are work that creates foundations for the machine learning work we intend to do in the MARVEL project.

3.5 Planned research: Feb 2015 – Jan 2016

Machine learning techniques provide an important computational alternative to simulation-based techniques in computational materials science. The general idea is quite elementary. Given a training set of compounds with pre-calculated quantum mechanical properties, we seek to construct supervised machine learning models that accurately infer the corresponding properties for similar materials. On this front, a variety of techniques have been applied so far, including kernel ridge regression and neural network training, which show great promise of these techniques ([13] and the references therein).
In the short run, we would like to improve the computational scalability of these techniques so that they can be efficiently applied to increasingly larger system sizes. This initially re-

quires additional work on scalable algorithms, such as applying new primal-dual algorithms for the kernel regression problems, or new stochastic gradient descent method for neural network training problems.

On both fronts, we will leverage our recent mathematical results. For instance, in the NIPS 2014 conference, we provided the first fully rigorous algorithmic framework for first-order primal-dual methods with optimal convergence rates on the primal objective residual and the primal feasibility gap of their iterates separately. Through a dual smoothing and proxy-center selection strategy, our framework subsumes the augmented Lagrangian, alternating direction, and dual fast-gradient methods as special cases. In a manuscript currently under review, we have developed a new stochastic gradient descent method, called the stochastic spectral descent, which seems to significantly improve the neural network training.

In the long run, we would like to focus on the correctness of such approaches and provide an analysis framework with approximation guarantees. We have already laid down the foundations of such work in our earlier work in NIPS 2012 and NIPS 2013, and ACHA 2014. The main challenge in this setting is to tailor a basic functional model that can incorporate translation and rotation invariance of the electron density. While the translation invariance appears the low hanging fruit, rotation invariance requires a second order modeling of the functional, which is of great theoretical interest by itself.

As a result, on the computational front, we would like to go well beyond the current literature by rigorously characterizing the tradeoff between the training data size and the actual computational effort required to obtain a level of approximation guarantee. We plan to leverage our recent time-data trade-off work on this front, which was published at NIPS 2014. Only an approach such as this one can help us computationally navigate an increasingly massive database of training information without decreasing the quality of the learning results.

The work on efficient and scalable machine learning techniques and software infrastructure for computational materials science has just started and will continue through the coming year. It is hard to make precise predictions of our progress over the coming year since our students are on fellowships, and neither are we committed to continuing working with them after January 2015, nor are they bound or committed to continuing to work with us.

We have mutually expressed our interest to collaborate with the IBM group on advanced analytics and machine learning problems, but this has to be intensified. One of the goals, there, is to use a very significant analytics and machine learning infrastructure developed inside IBM (known as Watson technology, after the system that got into the news for defeating the world's best players in a televised game of Jeopardy in the USA) for automatically "reading" the materials science literature and extracting new insights from it. For this collaboration to get really started, we most likely need to find a suitable PhD student to coadvise between IBM and EPFL.

3.6 Synergies with other computational and experimental efforts

Within MARVEL Synergies with the teams of Michele Ceriotti (EPFL), Alessandro Curioni (IBM), and Anatole von Lilienfeld (UniBas) were mentioned above.

3.7 Personnel

Matching personnel

- Milos Nikolic, PhD student, 10%, from May 2014.

MARVEL-related publications

List of publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.

- [1] K. Meier, T. Laino, and A. Curioni, *Solid-State Electrolytes: Revealing the Machanisms of Li-Ion Conduction in Tetragonal and Cubic LLZO by First-Principles Calculations*, The Journal of Physical Chemistry C **118**, 6668 (2014).
- [2] K. Meier, T. Laino, and A. Curioni, *Why Al doping stabilizes the cubic phase of garnet-like LLZO - Conclusive answers from an exhaustive computational study, submitted* (2014).
- [3] A. Shaikhha, Y. Klonatos, C. Koch, V. Jovanovic, and L. Brown, *SysC: A Framework for Systems-Compiler Co-Design, under double-blind submission* (2015).

Other references

[4] F. E. Faisal, Y. Ineichen, A. C. I. Malossi, P. Staar, C. Bekas, and A. Curioni, *Massively Parallel and near Linear Time Graph Analytics*, in *Proceedings of SC14, The International Conference for High Performance Computing, Networking, Storage and Analysis, New Orleans* (2014).

[5] Y. Ineichen, C. Bekas, and A. Curioni, *Large Scale Subgraph Node Centrality Estimation, submitted* (2014).

[6] P. Klavík, A. C. I. Malossi, C. Bekas, and A. Curioni, *Changing computing paradigms towards power efficiency*, Philosophical Transactions of the Royal Society A **372**, 20130278 (2014).

[7] C. Malossi, Y. Ineichen, C. Bekas, A. Curioni, and E. S. Quintana-Ortí, *Performance and energy-aware characterization of the sparse matrix-vector multiplication on multithreaded architectures*, in *Proceedings of the 43rd International Conference on Parallel Processing (ICPP-2014), Minneapolis* (2014).

[8] H. Tyagi and V. Cevher, *Active Learning of Multi-Index Function Models*, in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2012), p. 1466.

[9] J. Djolonga, A. Krause, and V. Cevher, *High-Dimensional Gaussian Process Bandits*, in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2013), p. 1025.

[10] Q. Tran-Dinh and V. Cevher, *Constrained convex minimization via model-based excessive gap*, in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds. (Curran Associates, Inc., 2014), p. 721.

[11] H. Tyagi and V. Cevher, *Learning non-parametric basis independent models from point queries via low-rank methods*, Applied and Computational Harmonic Analysis **37**, 389 (2014).

[12] J. J. Bruer, J. A. Tropp, V. Cevher, and S. R. Becker, *Time–Data Tradeoffs by Aggressive Smoothing*, in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds. (Curran Associates, Inc., 2014), p. 1664.

[13] L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.-N. Niranjan, P. Duncan, M. Rupp, K.-R. Müller, and K. Burke, *Understanding Machine-learned Density Functionals*, arXiv:1404.1333 (2014).

Platform Project 6 PP6 — Informatics

Project leader: Thomas Schulthess (ETHZ and CSCS)

Participating members: Thomas Schulthess (ETHZ and CSCS), Nicola Marzari (EPFL)

Summary and highlights: The Informatics Platform (PP6) is focused on providing the hardware and software infrastructure for the activities of MARVEL, and on populating and disseminating materials data to the scientific community at large. On the hardware side, these activities involve in the first 4-year phase the procurement and purchase of high-performance computing to support the numerical efforts of all the groups, and the procurement and purchase of a Petabyte-class storage facility for the research activities in data-mining and machine learning, and to act as a public repository of data and workflows. On the software side, key efforts are going into the development of the materials informatics platform AiiDA (http://www.aiida.net), that implements our vision for future activities in computational science (we call this the ADES model: Automation, Data, Environment, and Sharing, Fig. 1) to support high-throughput calculations, data preservation and storage, scientific workflows, and public dissemination. We also list here the first computational efforts that implement the paradigm of being database-driven/database-filling.

General view of the project

Major research questions

The accuracy and predictive power of materials simulations based on first-principles calculations allows nowadays a paradigm shift for computational design and discovery, in which massive HPC (high-performance computing) and HTC (high-throughput computing) efforts can be launched to identify novel materials or materials with improved or designed properties and performance; where behavior of ever-increasing complexity can be addressed in computational experiments; where data and workflows sharing can greatly enhance the synergies between different communities and efforts; and where services can be provided in the form of data, codes, expertise, workflows, turnkey solutions, and a liquid market of computational resources.

The impact of materials modeling on the overall scientific and economic progress will depend on our ability to promptly identify and take up the emerging scientific, technological, and societal challenges and to provide swift solutions to them. Coping with the latter task will require the generation, storage, retrieval, and analysis of ever increasing amounts of data, and the management of ever more complex codes. The complexity of the corresponding workflows calls for the implementation of a middleware infrastructure — what we label here a materials informatics ecosystem — to organise and coordinate systematically thousands to hundreds of thousands of simulations, searching for optimal properties and performance, while acquiring a variety of heterogeneous microscopic data from the *ab initio* calculations.

This infrastructure should ideally allow for an automated design and implementation of complex workflows and task tracking, based on a scripting interface for job creation and submission, and aimed at the creation and feed of heterogeneous databases of structures and properties that will in turn drive further simulations. The data thus generated will be used, e.g., for data-mining/machine learning, or to build classical neural networks to further ramp up the time and length scales accessible to numerical modeling.

Complementing this software infrastructure, it is essential to deploy the hardware infrastructure that supports the activities of all the groups, and that is amenable to optimal performance for high-throughput simulations, and integrating and supporting it through the informatics ecosystem alluded to above.

Last, open-access data dissemination and sharing needed to be supported by appropriate hardware that guarantees the storage, preservation, provenance, and dissemination of the data.

Short- and long-term goals

- Delivery of the materials informatics ecosystem for the automation of high-throughput calculations, the automatic

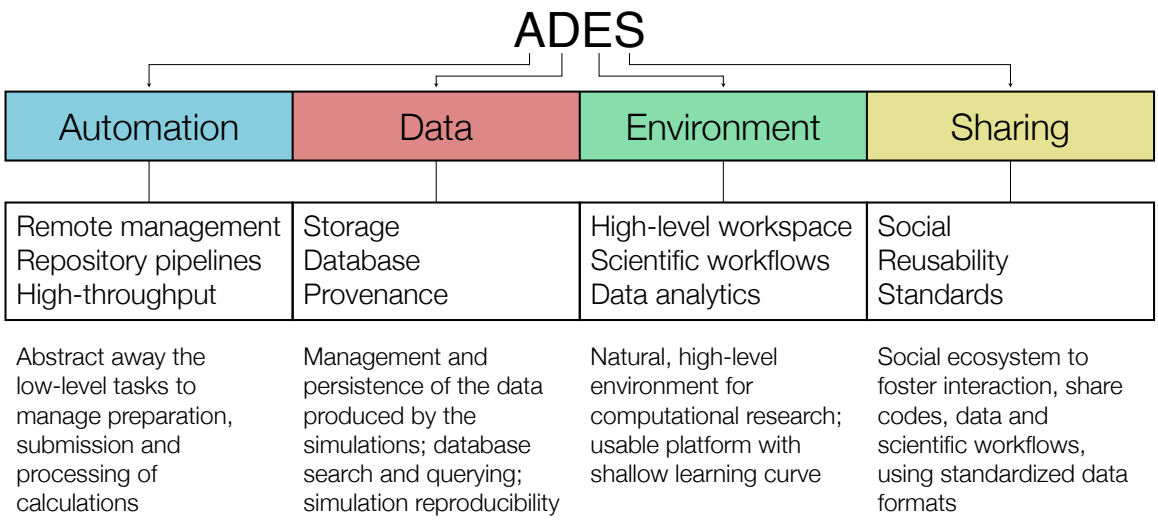


Figure 1: *The ADES model [1]: four pillars of an infrastructure for computational science. At the lower level, an automation framework and an efficient data management solution are needed to abstract away the tasks of calculation management, to persist the generated data and to query it. At the user level, a natural high-level environment allows researchers to analyse the results without the need of technical skills of database management; this is coupled together with a social ecosystem to stimulate the sharing of codes, data and workflows.*

storage of data in graph databases, a working environment where workflows and turnkey solutions are enabled, and sharing of data and workflows is made possible, while provenance, storage and preservation, reproducibility, and reuse are guaranteed.

- Delivery of the hardware infrastructure that sustains the computational needs of the MARVEL groups and the long-term storage, preservation, and dissemination needs.

First achievements

We have signed a collaboration agreement between EPFL, ETHZ, and CSCS, leading to the procurement of the hardware infrastructure for MARVEL, in the form of a Cray XC30 with 180 compute nodes of 24 cores each (Dual-socket Intel Haswell 12-Core, E5-2680v3, 2.5 GHz), where each core has 64 GBytes of memory (DDR4@2133 MHz), and a scratch capacity of 2.7 PetaBytes (shared with Piz Daint). SNSF funding for this have been matched by 820'000 CHF from ETHZ to cover the costs of operating the machine for four years (electricity, cooling, maintenance, system administration, etc.) and 500'000 CHF from EPFL to provide PetaByte class storage for the long-term dissemination of the MARVEL data.

On the software side, the first open-source release of the AiiDA software (`http://www.aiida.net`) will take place on February 27, 2015.

Next steps

We are working with CSCS to complete deployment of the supercomputing resources by April 1st, 2015, and are preparing the website to host the open-source/open-access codes, data and workflows that will be distributed by MARVEL.
Also, next year will be fully dedicated to the continuous development of the structure of AiiDA.

Research contributions to the overall goals of the NCCR and to the existing literature

The informatics platform supports all computational activities of the NCCR. On the software side, particular care is being put to insure integration with ASE (the Atomistic Simulation Environment, `https://wiki.fysik.dtu.dk/ase/`) and data compatibility with the Materials Project (`https://www.materialsproject.org` [2]) and the Crystallography Open Database (`http://www.crystallography.net` [3]).

Collaborative components

A visit to Lawrence Berkeley Laboratory has taken place in August 2014 to start the discussion on data compatibility with the Materials Project.

1 The AiiDA Materials' Informatics Platform (Nicola Marzari — EPFL)

1.1 Research summary

Providing the MARVEL community and the scientific community at large with the tools to deal most efficiently with the complexity needed and empowered by computational science, including the automation of high-throughput simulations on remote resources; the storage, preservation, provenance, and efficient searching of data; the development of a working environment where workflows connecting different and complex codes are constructed; the sharing and dissemination of data and workflows.

1.2 Research question

Besides the identification and implementation of a sustainable growth model for computer simulations, the impact of materials modeling on the overall scientific and economic progress will depend on our ability to promptly identify and take up the emerging scientific, technological, and societal challenges, and to provide swift solutions to them. Coping with the latter task will require the generation, storage, retrieval, and analysis of ever increasing amounts of data, and the management of ever more complex codes. The complexity of the corresponding workflows calls for the implementation of a middleware infrastructure — what we label here a **materials' informatics ecosystem** — to organise and coordinate systematically hundreds of thousands of simulations, searching for optimal properties and performance, while acquiring a variety of heterogeneous microscopic data from the *ab initio* calculations. This infrastructure should ideally allow for an automated design and implementation of complex workflows and task tracking, based on a scripting interface for job creation and submission, and aimed at the creation and feed of heterogeneous databases of structures and properties that will in turn drive further simulations.

1.3 Scientific goals

The four pillars of a materials' informatics platform (Fig. 1) can be summarized as:

- **Automation:** automate job execution, in particular by creating the input files for a given structure or task, automatically copying files to the remote supercomputer, submitting the calculation to the job scheduler and waiting for the job to finish, and finally retrieving the relevant files and parsing the results.
- **Data:** store calculations, their inputs and their results (either parsed, extracted from XML outputs, or from text files with the appropriate dictionaries) in a database, in a format suitable for computational materials science. Such database can be easily browsed and, most importantly, queried to retrieve in an efficient way results that match specific criteria. In AiiDA, a tailored hybrid SQL-NoSQL database solution is chosen that is targeted at atomistic simulations. In particular, any data object and calculation is represented as a node in a directed acyclic graph; links between nodes represent their causal relationship (inputs and outputs of a calculation). The backend is SQL (we support MySQL, PostgreSQL and SQLite) to provide the stability, reliability and fast querying of modern SQL databases. On the other hand, full NoSQL flexibility allows to store any kind of property that we want to attach to any node in the graph, by means of suitably designed entity-attribute-value (EAV) tables. A transitive closure table (automatically updated via SQL triggers) makes it possible to discover with a single query whether two given nodes are logically connected by a path in the graph, with any number of intermediate hoppings. This database structure is designed not only to guarantee the reproducibility of calculations, but also the *provenance* of the data: by just following the links in the database, a user can easily identify all the calculations that were needed to produce the final result, together with the input parameters of each calculation.
- **Environment:** much more than the data, the future of computational materials science will rely on preparing, storing, and sharing the workflows that generate the data, and that combine the different computational engines in the complex sequences that are needed to calculate advanced properties. For this reason a workflow engine has been developed within AiiDA: by defining the calculations that need to be run, together with their logical dependencies, AiiDA will take care of the submission process, so that children calculations start to run only when the parents have completed. Workflow steps can run either in parallel, if inde-

pendent, or sequentially, if dependencies are set. Moreover, a given step can be a sub-workflow itself (and one can define sub-sub-workflows to any depth level) so that existing stable workflows can be encapsulated into more complex or general ones in a modular structure. Many workflows are already available for Quantum-ESPRESSO, to manage for instance the steps needed for the calculation and plotting of a band structure, the evaluation of the density of states, the calculation of phonon properties, thermodynamic and thermo-mechanical properties (including thermal conductivities, in a multi-scale coupling with the Boltzmann transport equation for phonons), or complex thermostatting sequences of first-principles molecular dynamics simulations. These workflows also take care of automatically monitoring the calculations, with the ability to recover from typical errors (network failures, calculations to be restarted due to the wall time limit, code crashes that can be fixed by simple changes to the input file, etc.).

- **Sharing:** users (or trusted group of users, e.g. within the same institution) work within their own database, ensuring, e.g., that their data is kept private, or automatically uploading successful calculations to the public area. Import/export features allow different researchers to share selected portions of the database (e.g. with other researchers, when working within a collaboration). Such a sharing model has the advantage of guaranteeing that the data is private for those users that are not willing to share their results (especially in the first phases of their research). At the same time, however, the data is automatically stored in a uniform format in the database, and therefore sharing is made straightforward as soon as the user is willing to do it. Moreover, thanks to the uniform structure of the data stored, and the easiness of data retrieval from the database, the creation of public web repositories of calculation results becomes straightforward.

1.4 Plans and results: May 2014 – Jan 2015

During this year, we all worked at the development of the core structure of the software, that was first released to the MARVEL members in October 2014, and to the broader community in February 2015. A full documentation is available at http://www.aiida.net/wp-content/uploads/aiidadocs/index.html.

1.5 Planned research: Feb 2015 – Jan 2016

- Development of graphical web user interface.
- REST API for programmatic data access, query and retrieval.
- Improved, scalable database backend.
- Improved workflow engine, e.g. with automatic recognition of existing calculations, to avoid multiple executions of the same calculation.
- Addition of new plugins for other simulation software (WANNIER90, ...).
- Development of complex scientific workflows (equation of state, band structure, phonons, thermoelectric properties,...) for Quantum-ESPRESSO, including also auto-recovery capabilities in case of common errors easier querying capabilities for common query types, without the need to know the querying language.
- Release of virtual machines or docker images for easy automatic and deployment installation of AiiDA.
- Improved multiuser support.
- Improved sharing capabilities.
- Realization of an open web portal of materials' data calculated using AiiDA.

1.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* A first tutorial has been held at ETHZ in October 2014 ($\sim$ 30 attending) and at the Psi-k/CECAM conference in Berlin in February 2015 ($\sim$ 40 attending).

b) *Outside MARVEL* We are collaborating closely with the Crystallography Open Database (http://crystallography.net [3]) in importing and exporting structures automatically through AiiDA (Andrius Merkys is visiting the group for a year, supported by Sciex), and started discussing on how to maintain data compatibility with the Materials Project (https://www.materialsproject.org [2]). We plan also to start working with the Open Quantum Materials Database (http://oqmd.org/) and with SUNCAT and the Computational Materials Repository (CMR, https://wiki.fysik.dtu.dk/cmr/) — Nicola Marzari visited Stanford in February 2015 and Ivano Castelli is a former member and key contributor to CMR.

1.7 Personnel

NCCR personnel

- Giovanni Pizzi, postdoc, 100%, from September 2014.
- Martin Uhrin, research associate, 100%, from January 15, 2015.
- Nicolas Mounet, postdoc, 100%, from February 2015.

Matching personnel

- Andrius Merkys, visiting PhD student, 100%, from October 2014.
- Nicolas Mounet, visiting postdoc from CERN, 50%, from July 2014 to January 2015.
- Giovanni Pizzi, postdoc, 100%, from May to August 2014.
- Andrea Cepellotti, PhD student, 100%, from May 2014.
- Philippe Schwaller, master student, 50%, from September 2014.

2 High-throughput vibrational, dielectric and thermomechanical properties of novel materials (Nicola Marzari — EPFL)

2.1 Research summary

This project aims at providing a database of vibrational properties of crystals computed from first principles. A set of density functional perturbation theory calculations will be performed to obtain the phonon frequencies of known bulk crystals, as well as common or potential two-dimensional materials that can be exfoliated from existing three-dimensional ones. Then, all the quantities that can be derived from phonons, such as the vibrational free energies, the evolution of structural and mechanical properties with temperature, and solid-solid phase diagrams, will be computed within the quasi-harmonic approximation. In addition, for insulators, dielectric properties and effective charges will be obtained as a by-product of the phonon computations.

2.2 Research question

In the past decades, density functional theory (DFT) has become a popular tool to explore the properties of materials from first principles, with an accuracy that is often comparable to experiments. Recently, DFT-based high-throughput computations of materials have become increasingly feasible on standard computer clusters, enabling the calculation of, e.g., structural properties, ground state energetics, and band structures of large classes of materials, from molecules to three-dimensional crystals [4, 2]. This research project is aiming at achieving the natural next step in this high-throughput paradigm: providing temperature-dependent properties computed from first principles, thanks to density functional perturbation theory (DFPT). The focus here is the calculation of phonon frequencies, which then allow, e.g. through the quasi-harmonic approximation, to compute the vibrational free energy and all the quantities that can be derived from it. One first goal is to provide a complete database of phonon dispersions for two- and three-dimensional crystals, together with by-products such as the dielectric properties in the case of insulators. This effort will be followed by the high-throughput computation of all the related thermal properties, enabling to provide a complete picture of the thermomechanical behavior of a very large set of materials, ultimately almost as complete as typical crystallographic databases.

2.3 Scientific goals

a) *Short-term* A first phase of the project involves the realization of a set of tools enabling the high-throughput computation of phonon dispersions in a systematic way. These tools are being developed within the AiiDA computational platform [1]. To this purpose, AiiDA is interfaced with the Quantum-ESPRESSO distribution [5], in particular its DFPT code enabling the computation of dynamical matrices and phonon frequencies. In the short term, the first class of materials being studied are two-dimensional crystals, of which graphene, boron-nitride and molybdenum-disulfide are typical examples. In the spirit of the high-throughput band structures and ground state properties computed in [6], we are computing phonon dispersions over the full Brillouin zone for all possible 2D materials that can be predicted from existing databases. This will be followed by the prediction of the thermomechanical properties of those 2D compounds. Layered 3D crystals and superlattices will then be studied in the same way, using an extension of

Quantum-ESPRESSO that allows the use of van der Waals exchange-and-correlation functionals within DFPT [7].

b) *Long-term* In a second phase, the set of tools developed and tested on 2D and layered materials will be applied more generally to three dimensional crystals, as available in structural databases such as the Inorganic Crystal Structure Database (ICSD) [8] and the Crystallography Open Database (COD) [3]. Phonon frequencies and the vibrational thermomechanical properties will be computed and made available. Also, all kinds of quantities that can be derived easily from the same calculations will be added, including dielectric tensors, effective charges, Raman intensities and infrared spectra, powder diffraction spectra, Debye-Waller factors and mode Grüneisen parameters.

2.4 Plans and results: May 2014 – Jan 2015

a) *Set-up of the AiiDA tools* Several preliminary steps are needed before performing a high-throughput study of phonons in 2D materials. As for any high-throughput study that requires handling large amounts of data and calculations, it has to be performed in a well controlled computational architecture, allowing to efficiently launch calculations as well as to retrieve and store their results, together with their full provenance data, in a queryable database. The AiiDA platform [1], developed in the group, provides such an environment, and we will both use and improve it in this research line.
The first step is obviously to find the structures to study. In the specific case of 2D materials, since only a few of them have actually been synthesized up to now, one cannot extract many from current databases. So the strategy is rather to find potential 2D compounds from the existing 3D bulk crystals, more precisely from a subclass of those that are found to be layered according to a simple and general criterion based on atomic positions in the 3D structure (namely the fact that one can extract from the bulk a set of 2D layers separated by a large enough space such that no covalent bonds can be present between layers). A similar but slightly less general procedure was adopted in [6] to select 2D compounds. Moreover, plugins are also developed within AiiDA to import automatically structures from databases such as ICSD [8] and COD [3].
Then, within Quantum-ESPRESSO the computation of a full phonon dispersion, given an initial structure, requires four successive steps: (i) a self-consistent computation of the DFT ground state and relaxed structure, (ii) a self-consistent computation of dynamical matrices on a relatively coarse grid of $\mathbf{q}$ vectors of the Brillouin zone, (iii) a Fourier transform to obtain real-space interatomic force constants, and (iv) a Fourier interpolation of the latter to obtain dynamical matrices and their eigenvalues, the phonon frequencies, on a fine mesh of $\mathbf{q}$ vectors. Each of these steps is performed by a different code, and for each of them we designed a specific plugin to interface it with AiiDA. Also, since these steps need to be performed one after another, workflows are being developed to automatise the full computational sequence. The workflows include several levels of subworkflows, in order to automatically check the code outputs, identify the most common failures and launch again the calculations with updated parameters when possible. This limits human intervention to the rarest failures and greatly improves the efficiency of the full process.
The numerical parameters in the two first self-consistent steps described above need to be sufficiently converged. For this, we will use several sets of carefully tested pseudopotentials, for which energy cutoffs are found from a series of convergence tests performed both for the ground state energy and for the phonon frequencies. These tests are done as part of another research project in VP2.
Finally, the outcome of the last step has to be handled properly to get readily usable results, such as phonon dispersion curves. Again, several different cases have to be foreseen to encompass the full variety of 2D crystal structures investigated. This is done within AiiDA through the use of a special class handling in particular automatic path determination along the Brillouin zone, for any kind of crystal and any dimensionality.

b) *Vibrational properties of 2D materials* The set of tools that have been developed are now in the stage where they can be tested and used on a large scale study of phonon frequencies of 2D materials. As examples, we show in Fig. 2 the phonon dispersions of several 2D materials.
Knowing the phonon frequencies for several lattice parameters allows to compute the vibrational free energy as a function of lattice constants, which in turn can be minimized at any given temperature to get the structural parameters, in the framework of the quasi-harmonic approximation. An example of the success of such a procedure is shown in Fig. 3, extracted from [9], with the calculated coefficient

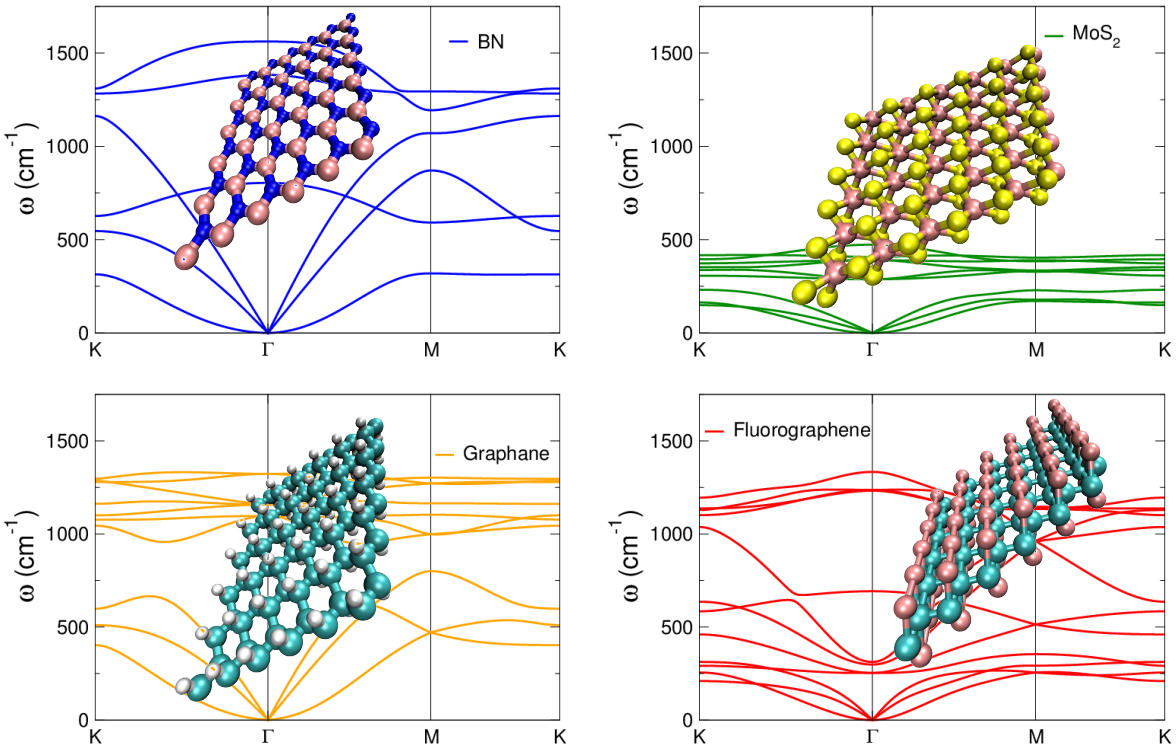


Figure 2: *Phonon dispersions of boron nitride, molybdenum-disulfide, graphane and fluorographene along high-symmetry lines of the Brillouin zone, together with their 2D crystal structure. For graphane, the two highest optical branches exhibit frequencies above* 2800 cm^{-1} *and have been omitted.*

of in-plane thermal expansion for graphite and graphene, together with experimental results for graphite. The agreement between experiments and theory is in this case very good, justifying the approximations used. As another example, we also show the coefficient of thermal expansion of 2D molybdenum-disulfide vs temperature in Fig. 4.

Many other thermomechanical properties can be obtained from the vibrational free energy, in particular heat capacities, and the dependence on temperature of the elastic constants. Moreover, from the vibrational Gibbs energy one can also compute pressure-temperature phase transitions between different 2D crystal orderings. This will require the computation of additional phonon dispersions for different phases, which can be included in our initial

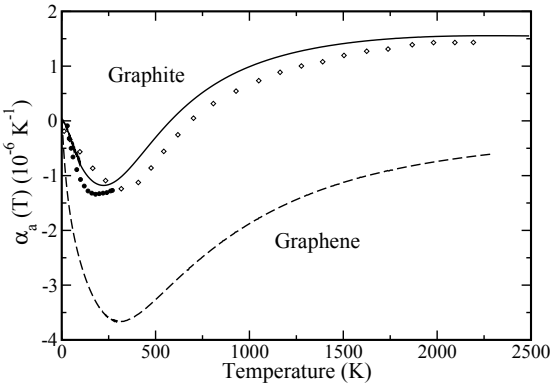


Figure 3: *In-plane coefficient of thermal expansion as a function of temperature for graphite and graphene. Two sets of experimental results for graphite are also shown (filled circles and open diamonds). From [9].*

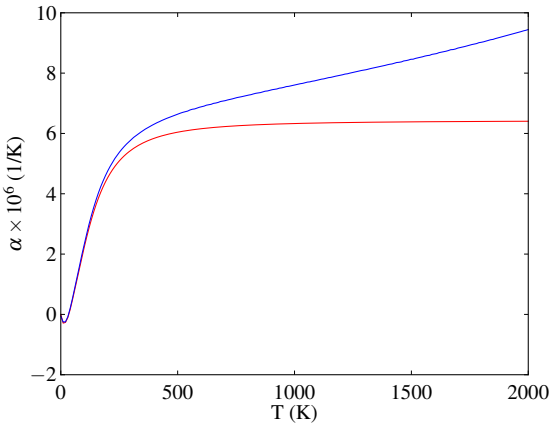


Figure 4: *In-plane coefficient of thermal expansion as a function of temperature for 2D molybdenum-disulfide. The blue curve was computed using a direct minimisation of the free energy, while the red one was obtained using the Grüneisen approximation.*

high-throughput phonon study.

c) *Vibrational properties of layered materials* In this case, the layers are typically weakly bonded to each other and we need to resort to van der Waals non-local exchange-correlation functionals, such as vdW-DF1, vdW-DF2 and rVV10, which are available in an extension of the Quantum-ESPRESSO DFPT code ([7] and references therein). For this study, an intermediate goal will be to obtain a set of tests on, e.g., elastic constants or phonon frequencies, to select the best performing functionals compared to the available experimental data. The materials tested will be essentially the layered 3D compounds that were identified in section 2.4a).

2.5 Planned research: Feb 2015 – Jan 2016

The study of 2D and 3D layered materials will continue by computing and storing all the quantities that can be derived from the vibrational frequencies, in particular the Raman intensities and infrared spectra, the powder diffraction spectra, Debye-Waller factors and mode Grüneisen parameters. Dielectric tensor and effective charges will also be studied, and possibly ferroelectric or highly polarized 2D materials screened.
Moreover, several structures made of an heterogenous stacking of 2D layers will be constructed, and their vibrational properties predicted as well, in collaboration with the group of Prof. Andrea Ferrari in Cambridge University that will be able to experimentally test such structures.
Finally, the 2D and layered-3D materials study can be seen as a first test bench that will foster improvements and optimisation of the AiiDA platform. The latter will then be fully mature for an even larger scale project: building an as complete as possible database of phonon frequencies of all 3D bulk materials, by systematically extracting the structures in the main crystal databases and performing DFPT calculations on them.

2.6 Synergies with other computational and experimental efforts

a) *Within MARVEL* Synergies with efforts in VP2 are taking place, in particular on the testing of pseudopotentials and the convergence of numerical parameters.

b) *Outside MARVEL* Collaboration with Prof. Andrea Ferrari from the University of Cambridge is foreseen to compare theoretical predictions on layered materials with experiments, in particular the elastic constants. On a similar subject, a collaboration with Dr. Daniele Stradi from the Technical University of Denmark has been initiated to compute phonon dispersions of boron-nitride and graphene heterostructures, and compare them with experiments.

2.7 Personnel

NCCR personnel

- Nicolas Mounet, postdoc, 100%, from February 2015.

Matching personnel

- Nicolas Mounet, visiting postdoc from CERN, 50%, from July 2014 to January 2015.
- Andrea Cepellotti, PhD student, 100%, from May 2014.
- Nicola Varini, postdoc, 100%, from October 2014.
- Philippe Schwaller, master student, 50%, from September 2014 to December 2014.

MARVEL-related publications

List of publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.

⬢ [1] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, *AiiDA: an Infrastructure for Computational Material Science, submitted* (2015).

Other references

[2] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, *Commentary: The Materials Project: A materials genome approach to accelerating materials innovation*, APL Materials **1**, 011002 (2013).

[3] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quirós, N. R. Serebryanaya, P. Moeck, R. T. Downs, and A. Le Bail, *Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration*, Nucleic Acids Research **40**, D420 (2012).

[4] R. Armiento, B. Kozinsky, M. Fornari, and G. Ceder, *Screening for high-performance piezoelectrics using high-throughput density functional theory*, Physical Review B **84**, 014103 (2011).

[5] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, *Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials*, Journal of Physics: Condensed Matter **21**, 395502 (2009).

[6] S. Lebegue, T. Bjorkman, M. Klintenberg, R. M. Nieminen, and O. Eriksson, *Two-Dimensional Materials from Data Filtering and Ab Initio Calculations*, Physical Review X **3**, 031002 (2013).

[7] R. Sabatini, *Non-local correlation in Density Functional Theory*, Ph.D. thesis, SISSA (2012).

[8] http://www.fiz-karlsruhe.com/icsd.html.

[9] N. Mounet and N. Marzari, *First-principles determination of the structural, vibrational and thermodynamic properties of diamond, graphite, and derivatives*, Physical Review B **71**, 205214 (2005).

Platform Project 7 PP7 — Experiments

Project leader: Frithjof Nolting (PSI)

Participating members: Frithjof Nolting (PSI), Michel Kenzelmann (PSI), Pierangelo Gröning (Empa)

Summary and highlights: Turning the results of the computational approach into real materials requires the intense collaboration between theory and experimental groups. The Platform Project 7 (PP7) is the linking point and provides funding for experimental verification projects. These research projects will start in the second year of the NCCR and in the first year a proposal system for the distribution of the funding has been developed and launched. A very successful one day workshop showed the high interest and potential for these collaborations and formed the basis for possible projects.

1 Major research questions

The goal of the experimental verification platform on experimental synthesis and characterization within the NCCR MARVEL is to harness the vast and diverse experimental capabilities in materials' synthesis, analysis and characterization of Swiss laboratories to create a tight synergy with the computational activities of this NCCR, with the objectives of

- providing experimental synthesis, characterization, and measurement of the properties and performance for the materials' design effort;
- help directing this effort to support some of the strategic goals of the experimental materials' laboratories.

The research activities in PP7 are foreseen to start in the second year of the NCCR and, in the first year, the focus was on establishing the first contacts between theory and experimental groups and to define the boundaries for the collaborative projects.

2 Short- and long-term goals

In order to take into account the large variety of research topics addressed in MARVEL, the diverse experimental capabilities and to be able to react on the developments in the field, a proposal system for the financing of research projects has been implemented rather than assigning the funding to specific groups. The funding is for focused projects, ranging from six months up to three years, to link the (mostly theoretical) MARVEL group with the experimental efforts. In the long term it would be focused to synthesis and characterization of materials suggested by the simulations, but at this early stage it is also to nucleate core synergies, and involve the computational people more in the experimental landscape in Switzerland.
A first call with a deadline end of January 2015 has been announced for experimental projects with a strong connection to numerical theory in the following fields:

- materials with novel multifunctional properties (multiferroics and artificial heterostructures);
- materials with novel physical properties (topological insulators, model Hamiltonian, dynamically excited materials);
- novel materials for energy applications (dye-sensitized solar cells, photocatalytical water splitting, metal/air batteries, thermoelectrics);
- organic molecular crystals.

3 First achievements

A one day workshop on "NCCR MARVEL — Experimental verification" has been held on 17th October 2014 at the Paul Scherrer Institute. The aim was to get people in contact, make experimental groups aware of NCCR MARVEL, make theory groups in MARVEL aware of experimental possibilities and to initiate discussions leading to a proposal submission. The workshop consisted of short introductions into MARVEL and PP7, longer perspectives talks, and talks about examples of successful collaborations. The central part of the workshop was a session with 25 short 5-minutes presentations about possible projects ideas followed by a poster session for detailed discussion (Fig. 1). The workshop was well attended with over 50

Figure 1: *Left: Nicola Marzari gave a short introduction and overview of the topics of the NCCR MARVEL. Right: the short talks were followed by discussions at the poster session and coffee break.*

participants and, in particular, the short presentations gave a very good stimulus for discussions and future collaborations.
A web-based proposal submission platform was designed which enables the submission and handling of the proposals in an efficient manner. Through these proposals, it is possible to receive funding for the salary for postdoc or PhD students and a small amount of consumables. The proposal will be reviewed by a committee consisting of the PP7 project members (Frithjof Nolting, Michel Kenzelmann, Pierangelo Gröning, Maksym Kovalenko) and representative of the NCCR theory PIs (Nicola Spaldin and Nicola Marzari). PP7 project members may also seek the expert help of external reviewers, in rather exceptional cases. The evaluation criteria are:

- link to topics of NCCR MARVEL, connections to theory group in MARVEL;
- scientific merit;
- feasibility.

4 Next steps

The deadline for the proposal submission is end of January 2015. The proposals will be evaluated in February 2015 and the selected projects can start from May 2015 on. The development of these projects will be followed and support by PP7. The discussions and interactions with the members of the NCCR and experimental groups in Switzerland will be continued. An experimental verification workshop will be held and a new call for proposal will be launched at the end of 2015.

5 Research contributions to the overall goals of the NCCR and to the existing literature

Since the funding for research projects within PP7 starts in the second year of the NCCR, the contribution of PP7 is mainly on the networking side between theory and experimental groups. This already resulted in closer interaction between some groups and the initiating of new joint projects.

6 Collaborative components

Two major criteria for the evaluation of the proposals are (i) scientific merit and feasibility of the planned project and (ii) the link to topics and groups in MARVEL and a group leader of MARVEL is required as co-proposer. Therefore the collaboration between different groups is at the heart of this project. Besides the concrete research projects an important contribution of PP7 is to provide a contact point between the different MARVEL theory groups and the experimental groups in Switzerland.

2.3 New projects

The group of Volkan Cevher, from EPFL Laboratory for Information and Inference Systems (LIONS), has been involved in the NCCR activities, initially through the collaboration and funding assigned to Prof. Koch. Prof. Cevher is an expert in machine learning, among other things, and his expertise will be very valuable for our HP5 Materials Informatics Platform. Synergies are also being developed between the group of Cevher and those of von Lilienfeld and Ceriotti.

3 Knowledge and technology transfer

As defined in the Strategy for Knowledge and Technology Transfer, in Annexe 1, knowledge and technology transfer in the NCCR will happen at several levels, from the dissemination of open-source materials' simulation codes, to the training in the use of those codes and of the newly developed materials informatics framework, to the sharing of all results from materials simulations in a network of open-access servers, to the verification and validation of calculations. Training and education will not only target the traditional area of academic computational groups, but also experimental groups, embedding the use of first-principles spectroscopies and microscopies into the toolbox of those researchers, and a special effort will be made to transitioning these tools to local and global industries.

A number of activities have taken place in these early stages of our collaboration, and are listed in the following.

Software

AiiDA (Automated Interactive Infrastructure and Database for Atomistic simulations, http://www.aiida.net) is the infrastructure associated to the Informatics Platform PP6. In August 2014 Nicola Marzari and Giovanni Pizzi visited the Materials Project at Lawrence Berkeley Laboratory, to present the AiiDA platform, and to discuss possible synergies, collaborations, and efforts at data standardization. The first official open-source release of the AiiDA code (internal version number 0.4.0), with a MIT-BSD, took place on February 27, 2015. Two tutorials on AiiDA were already organized before this first release:

- on October 31, 2014 in ETHZ for members of MARVEL, gathering about 30 participants;
- on February 5, 2015 in Berlin, for the whole community, held after the CECAM/Psi-k research conference, gathering about 40 participants.

A PhD student (Andrius Merkys) is visiting EPFL for 13 months, to work on the seamless integration of AiiDA with COD (http://crystallography.net), and supported by Sciex (supporting interactions between Switzerland and EU-10 countries, in this case Lithuania). We also concluded the agreement with Robert Bosch LLC, that has collaborated on the development of the platform, for a full release of it under the open-source licence.

On October 31, 2014, a collaborative meeting took place initiated by Antoine Georges at UniGE. The focus was pushing forward the electronic structure tools in the TRIQS library, especially the interface with WANNIER90 (Cf. scientific report on VP1). This joint effort at open-source software design was very constructive and productive, and gathered members of the group of Nicola Spaldin at ETHZ and of the group of Antoine Georges at UniGE as well as those participating in similar efforts in Paris and Saclay (Fig. 1).

Figure 1: *Collaborative meeting on the* TRIQS *library at UniGE on October 31, 2014.*

The development of the capability to compute Chern invariants with the Wannier code using spin-orbit coupling and ultrasoft pseudopotentials is taking place, led by the Spaldin group and in collaboration with the groups of Troyer and Yazyev.

Hardware

The collaboration agreement between EPFL, ETHZ, and CSCS was signed by all in December 2014, providing the framework for the deployment of the 1.5 M CHF Cray XC30 supercomputing facility at CSCS, securing 0.82 M CHF matching funds from ETHZ to support administration, maintenance, and running costs, and to prepare for the 0.5 M CHF PetaByte storage facility. A meeting in ETHZ took place in January 2015, with Thomas Schulthess and all the key CSCS personnel, and Nicola Marzari and the AiiDA team, to discuss planning, management, deployment, and administration of the facility.

Conferences and collaborations

A number of conferences and workshops have been organized or are planned. These include:

- the Psi-k/CECAM annual research conference, on "Frontiers of Electronic-structure Simulations: Materials Design and Discovery", held in Berlin on Feb 1-5 2015 (Marzari co-organizer);
- a CECAM workshop on "Future Technologies in Automated Atomistic Simulations", to be held in EPFL on Jun 8-10 2015 (Marzari co-organizer);
- a session organized by Nicola Marzari on "Materials Design" at the quinquennial Psi-k General Conference Psi-k2015, to be held in San Sebastian on Sep 6-10 2015;
- a Psi-k "International Symposium and Workshop on Electronic Structure Theory for the Accelerated Design of Structural Materials", to be held in Moscow on Oct 26-30 2015 (Marzari co-organizer).

MARVEL management will participate at the Platform for Advances Scientific Computing PASC15 conference on June $1-3$, 2015 at ETHZ. An information stand will be set-up promoting MARVEL activities. Prof. Galli, head of the Scientific Advisory Board of MARVEL, will provide a plenary lecture on the topic of "Materials Discovery and Scientific Design by Computation: a Revolution Still in the Making".

Dr. Nongnuch Artrith, from the Mechanical Engineering Department of MIT, will visit MARVEL for four months, April to July 2015. She will work on solvation models in collaboration with Stefan Goedecker and Nicola Marzari, and on neural networks in collaboration with Michele Ceriotti.

A meeting is planned by Nicola Spaldin in scientific project VP1, to continue the search for "model materials" with a focused effort involving all relevant teams of MARVEL in the first half of 2015, to identify promising material classes and computational approaches/model Hamiltonians.

Technology transfer

A contract was signed between Nicola Marzari and an industrial partner with a financial contribution to the NCCR for four years, to support a PhD student in novel materials design in connection with the project VP2.

The NCCR Chemical Biology is organizing in March 2015 a Pre-seed Workshop "specifically targeted towards high-technology entrepreneurial ideas within Switzerland". MARVEL students or other researchers were invited, through the MARVEL group leaders, to submit a "champion idea". At least, one group of MARVEL students was selected to present their project, in connection with the materials informatics platform AiiDA.

Contacts were taken with EPFL technology transfer office (TTO) and Alliance, the industrial liaison program of EPFL, and an agreement is being negotiated for a 20% appointment of an industrial liaison officer whose responsibilities would be:

- support MARVEL, principally the EPFL-labs, in the implementation of their joint strategy for technology transfer and external fundraising;
- develop partnerships and business development activities with external private partners in Switzerland and abroad in the research fields related to computational approach to the discovery and development of materials applied to the fields of energy, ICT and pharmaceuticals;
- identify companies potentially interested by the use of modeling to accelerate or

optimize materials development in synergy with Vice-Presidency for Innovation and Technology Transfer Corporate relationships and relevant EPFL Centers;

- organize the contacts with labs affiliated to NCCR Marvel and give support to build project and to formalize the contractual and IP aspects in collaboration with the TTO EPFL;
- manage and coordinate of joint research projects between MARVEL's research laboratories and leading industrial players;
- develop close interactions with EPFL scientists in MARVEL's research laboratories as well as with the management teams of MARVEL.

Pascale Van Landuyt (PhD Materials Science) should be this industrial liaison officer and should begin to work for MARVEL in spring 2015.

4 Education and training

As defined in the strategy for education and training, in Annexe 2, one central activity of MARVEL is to contribute to the formation of doctoral and postdoctoral students. This is achieved through various actions such as scientific meetings, lectures, schools and workshops. This first year was also dedicated to the analysis of the existing offers. In particular, as mentioned in the strategy, all computational classes at the Master and PhD level in the participating institutions have been identified. This offer might benefit from a coordination effort and has to be advertised to the students members of MARVEL and used by them. MARVEL will organize its own school in 2016, but encourages already participates in and supports the attendance of young researchers to existing schools and workshops. Regarding to the younger generation, it is part of the MARVEL responsibilities to increase their interest in science, technology, engineering, and mathematics (STEM).

At the level of PhD and postdoctoral students

In the MARVEL community

Existing courses

A list of the existing offer of computational classes at the Master and PhD level in the participating institutions has been compiled in the strategy for education and training, in Annexe 2. This offer has to be advertised to the students members of MARVEL and used by them.

Participation to schools

In this first year, MARVEL encouraged students (PhD and postdocs) associated to the NCCR to participate in various schools and workshops in the field of computational design and discovery of novel materials, or more generally in topics of current interest in theoretical condensed matter physics and materials science. The following schools and workshops were advertised to all the MARVEL group leaders in the different institutions, asking them to inform their PhD students and postdocs of the opportunity to participate, with, in most cases, some financial support from MARVEL.

- The 6th MaNEP Winter School, "Shedding light on correlated electrons?", Saas Fee (VS), January 18 to 23, 2015 (http://www.manep.ch/saasfee15/) was advertised, with the mention that the registration (including travel) was free for PhD students from the *Conférence universitaire de Suisse occidentale* (CUSO) member institutions.

- The international conference *Nothing is Perfect – The Quantum Mechanics of Defects*, Ascona (TI), April 26 to 29, 2015 (http://nothingisperfect.epfl.ch) has been advertised with the possibility to benefit from MARVEL sponsoring for the registration fees. Seven PhD students and postdocs will participate with a total MARVEL financial support of 5'000 CHF.

- As mentioned in the strategy for education and training, in Annexe 2, close contacts exist between MARVEL and the "Simons Collaboration on the Many Electron Problem" (https://www.simonsfoundation.org/mathematics-and-physical-science/many-electron-collaboration/) directed by Andrew Millis. As a first action, MARVEL students are invited to the annual Simons Many Electron Collaboration Summer School held at the Simons Center for Geometry and Physics at Stony Brook from June 8 to 12, 2015, focusing on electronic structure and its many body extensions. Four MARVEL students will participate; the Simons Collaboration will support all of the local expenses, and the student travel to and from the school will be payed by MARVEL, up to 5'000 CHF.

- The *MolSim-2015* school was held at CECAM, January 5 – 16, 2015, and organized by Berend Smit and others. This school provided training in the field of simulation techniques for the study of many-particle (molecular) systems. The content constituted a recap of the statistical mechanics relevant for molecular simulation, and subsequently covered basic and advanced simulation techniques. These including Monte Carlo, molecular dynamics, free energy calculations, rare events, coarse graining. The theory and techniques covered in the lectures were accompanied by illustrative examples of present-day research.

Junior retreat

In July 2013, a materials simulation junior retreat was organized by a group of postdocs with the financial support of the Swiss Platform for High-Performance and High-Productivity Computing (HP2C) / Platform for Advanced Scientific Computing (PASC), as part of the preparatory actions of the present group of PIs. The retreat gathered young researchers in Swiss institutions "working in the field of materials simulation where physics, chemistry, mathematics and materials science meet high performance computing".
In a similar line, a first MARVEL junior retreat is to be organized from July 8 to 10, 2015. The organizing committee consists of three postdocs, in three different MARVEL groups, supervised by a junior professor, Michele Ceriotti (EPFL), the MARVEL group leader in charge of education and training at the scientific level.
The aim of this retreat is to bring about 50 young researchers together, to get to know each other, and share their experiences in their vast and highly challenging research field. It will be held at the Hotel Boldern, in Mannedorf near Zurich.

Outside MARVEL

MARVEL takes also a responsibility to support education of young researchers outside Switzerland, in particular with a computational science support in emerging regions. In this framework, MARVEL has supported the costs for one lecturer (Dr. Matteo Cococcioni in Nicola Marzari's group at EPFL) and one mentor and tutor (Florian Thöle in Nicola Spaldin's group at ETHZ) for the third African School on Electronic Structure Methods and Applications (ASESMA) in Johannesburg, South Africa, from January 19 to 30, 2015

Figure 1: *Group picture at the third African School on Electronic Structure Methods and Applications (ASESMA) in Johannesburg in January 2015.*

(Fig 1). This school is organized in collaboration with the International Center for Theoretical Physics (ICTP) in Trieste. MARVEL also financially supported the participation of Dr. Sinead Griffin (Nicola Spaldin's group at ETHZ) to lecture in electronic materials at the satellite school in Sudan, the Khartoum Workshop for the Advancements in Material Sciences (KWAMS-15), from February 1 to 10, 2015. ASESMA was created in 2010 (Nicola Spaldin and Nicola Marzari both in the advisory board) to promote electronic structure methods in Sub-Saharan Africa, following a first 2-weeks workshop at the University of Addis Ababa in September 2008, organized by Nicola Marzari.

For the younger generation

As described in the equal opportunity chapter (chapter 5, p. 114), a special workshop related to MARVEL fields for young girls 7 to 10 years old was developed and took place on three Wednesdays in January 2015. The goal of this workshop, entitled *Diamant, alu, caoutchouc, ils sont fous ces matériaux !*, is to let young girls discover the world of material sciences having fun. Another similar workshop for slightly older girls (11 to 13 years old) will be proposed on one week during the summer holidays in August 2015.
Boys will not be forgotten from MARVEL educational plans and MARVEL will participate at events for the general public, as stated in the communication chapter (chapter 6, p. 119). Some hands-on experiments and demonstrators need to be developed for such events, as well as for activities that could be proposed more generally for schools. A collaboration with the scientific mediators of the Science Outreach Department at EPFL is planned.
For the next months, the scientific manager, together with the new program manager that will be hired, and interested researchers of the MARVEL community, will deepen the reflex-

ion on educational materials and activities for the younger generation.

5 Equal opportunities

MARVEL defined a strategy in order to contribute to a better awareness concerning the gender issues, promote MARVEL-related fields among girls to encourage them to choose scientific and engineering fields, recruit female scientists and provide them with information and tools and offer them an adequate environment to develop their career and to access to higher positions, and tend to a gender balance at all levels of MARVEL.
In this first year, MARVEL developed some specific measures in the field of MARVEL and ran shared actions in the framework of this strategy together with the Synapsy and Robotics, the two other NCCRs hosted at EPFL, and the EPFL Equal Opportunities Office.

Numbers

MARVEL has a large number of female professors in key positions. The head of the Scientific Advisory Board is a woman, Prof. Giulia Galli). One of the three Executive Committee members is Prof. Nicola Spaldin and one of the seven project leaders is Prof. Ursula Röthlisberger, both being part of the Scientific Committee. Prof. Clemence Corminboeuf is in charge of representing the interests and goals of young researchers and women in the NCCR administration. Our young researchers therefore have local access to female role models and mentors.
At present, five female PhD students, five female postdocs, two senior researchers and four group leaders are member of MARVEL. Efforts will be focused on recruiting more female scientists in following years.

Mentoring, networking and career planning actions for female PhD students and scientific staff

Events

In the framework of common actions of the Equal Opportunities Office and NCCRs hosted at EPFL, the following events were organized in order to support female PhD students as well as the postdoctoral fellows to continue building their career (Table 5.1). These activities are advertised via the management of the NCCRs to all the PhD students and postdocs hired in the different institutions where the NCCRs are present.
Discussions are also taking place between different NCCRs present at ETHZ in order to organize "Professional Skills Development" Workshops for women within the framework of NCCRs and the Women Professors Forum of ETHZ. Nancy Houfek, who runs workshops hosted by the American Physical Society was contacted. She is enthusiastic to offer such workshops in Switzerland, and proposed three 1-day workshops at the PhD/postdoc, junior faculty/researcher, and senior faculty levels. Two could be hosted at ETHZ and one at EPFL, to offer the opportunity to participate to more female scientists, at all levels. Discussions are ongoing to find the best solution taking in to account the efficiency and the costs.

Collaborations concerning mentoring and coaching programs

One of the goals of MARVEL announced in the strategy for equal opportunity is to support female scientists with mentoring and coaching programs. Thanks to the collaboration between the *Suisse latine* universities, female PhD students and Postdocs of MARVEL, employed by UniGE, UniFR, and EPFL have access to two mentoring programs named *StartingDoc* and *Réseau romand de mentoring pour femmes*. Those from EPFL, Empa, ETHZ and PSI can take part in the program "Fix the leaky pipeline!". Those from Basel, Zurich and Fribourg universities have access to the *Mentoring Deutschschweiz* program. All these programs are briefly presented hereafter.

October 31, 2014	*Google @ EPFL* Networking event with women in industry	This 2-hours informal lunch meeting was organized at EPFL with engineers and employees of Google. 50 female PhD students and Postdocs participated at this event (Fig. 1).
November 13 and 14, 2014	*"Self-Marketing Skills" — indispensable to boost your career* Course by Dr. Monika Clausen "Fix the leaky pipeline!" program	The workshop was focusing on four main topics: reflection on self-confidence, development of communication skills, practicing self-presentation, and discussion of tips and tricks for improved self-marketing.
November 18, 2014	*Visit CISCO* Company Visit and networking	25 scientific women could visit CISCO at EPFL Innovation Square during a day and have discussions and lunch together with the CISCO staff. A journalist of the French speaking television of Switzerland RTS Un was present and produced a video-reportage (http://www.rts.ch/play/tv/-/video/les-femmes-dirigeantes-dans-les-societes-de-nouvelles-technologies-restent-rares?id=6365566).
December 2, 2014	*Developing a Comprehensive Skills' Profile* One day workshop by Dr.Pamela Alean-Kirkpatrick "Fix the leaky pipeline!" program	The workshop focused on increasing the awareness of skills developed in an academic context as a researcher — Development of a personal skills profile: identifying strengths, detecting gaps and drawing up a personal development plan.
January 22, 2015	*Women at work: same difference?* Two hours workshop by Jill Székely, organized by the NCCR Chemical Biology	This workshop was set up to bring solutions in the following fields: - Freedom and Power: the magical union. - Time... flies: how to manage a double career: professional and biological CV? - who am I and what am I worth in public & private industry?

Table 5.1: *Events organized in order to support female PhD students and postdocs to continue building their career.*

- *StartingDoc* — http://www.unil.ch/mentoring — is a collaboration of *Suisse latine* universities and EPFL. The goal of *StartingDoc* is to enable doctoral students at the beginning of their thesis to acquire skills of the academic world. A new meeting of this program with several workshops was organized at EPFL Equal Opportunities Office in November 2014 by the coordination team of the program.

- The *Réseau romand de mentoring pour femmes* — http://www.unifr.ch/f-mentoring — is a collaboration of *Suisse latine* universities and EPFL for advanced PhD students and Postdocs.

Figure 1: *Google lunch meeting at EPFL on October 31, 2014.*

Through networking, this program seeks to enhance the position of female researchers in academia. The next meeting of the *Réseau romand de mentoring pour femmes* takes place on January 30 and 31, 2015 with six workshops on efficient communication, how to prepare the thesis defense, the nomination process rules and procedure (for advanced postdoctoral fellows), skills' profile, academic career and its different steps, and networking.

- "Fix the leaky pipeline!" — http://www.fix-the-leaky-pipeline.ch — is a collaboration of the institutions of the ETH domain, financed by the institutions and the ETH board. This program offers young female scientists (PhD students, Postdocs, and others) the opportunity to reflect on their professional situations, develop a strategy for embarking or continuing on their career paths, receive targeted further training, and extend their personal and scientific networks. Workshops, courses and coaching groups were organized during 2014. Several courses and coaching groups in ETHZ and EPFL area are planned for 2015.

- MARVEL female students and staff have access also to a portal and a forum gathering the necessary information for pursuing an academic career developed as a collaboration of equal opportunities offices of the *Suisse latine* Universities and EPFL. The portal: http://www2.unine.ch/releve/lang/en/Homepage provides also information about specific opportunities for women (e.g. mentorships, funding, etc.). A link on this site will be put on the MARVEL website when available.

Female PhD students and Postdoctoral fellows are encouraged to participate at one of these mentoring and coaching programs. The professors of MARVEL, men and women, will be encouraged to volunteer as mentors for the next editions of one of these programs. Information about these programs will be directly sent to female scientists of MARVEL and put forward on the MARVEL website.

Daycare facilities and work-life balance

MARVEL students and staff have access to daycare facilities and advice in the different institutions of MARVEL. This information will also be edited on MARVEL website. In addition it is planned to organize a meeting with female scientists of MARVEL to determine their needs concerning work/life reconciliation as well as daycare facilities. This may take place during the MARVEL Review & Retreat in September 2015. Appropriate measures could be implemented as a function of the needs.

Actions concerning young girls

Polythèmes

Polythèque is a place for children to discover and share ideas on science through adapted books (more then 700), videos, CDs and games. Children can meet in a place where parents are welcome and where staff is present to answer kids' questions. Children can register to participate at scientific workshops, *Polythèmes*.
The goal of these workshops is to teach how to tackle a scientific problem and how to learn scientific concepts by experimentation. Children discover how interesting and amazing sciences are and why they are useful for the society.

Figure 2: Polythèmes Diamant, alu, caoutchouc, ils sont fous ces matériaux ! *on January 7, 2015. Left: all participating girls are playing a game to sort materials between metals, polymers, ceramics, and composites. Right: young girls are growing alum or borax crystals.*

The content of a special workshop related to MARVEL fields was developed. The goal of this workshop, entitled *Diamant, alu, caoutchouc, ils sont fous ces matériaux !*, is to let young girls discover the world of material sciences having fun. Different types of material and their properties (color, hardness, thermal, electric and magnetic properties, etc.) are presented to young girls and they learn about these materials by performing simple hands-on experiments, such as synthesis of crystals, slime, bouncing balls, or galalith, observation of properties of super-absorbing materials or non-newtonian fluids (cornstarch and water), etc. (Fig. 2). The first edition of the workshop took place on January 7, 14 and 21, 2015. This workshop met a huge success and there are enough registrations on the waiting list for organizing a second workshop, on the second semester of the year. This workshop will be organized twice a year during the first phase of MARVEL.

A special scientific week for 11 to 13 years old girls

The content of a new scientific week having for subject different fields related to MARVEL is under preparation. The first edition of this workshop will be organized in August 2015. Besides hands-on experiments and theoretical explanations young girls will discover the research running in different labs of MARVEL. The relation between what they learn during the scientific week and what is done in the labs will be explained. During the week, the participants will also prepare a presentation about the lab they visited. As they are divided in different groups and visit different labs, at the end of the week, each group will present a different research project to the other participants and to the parents. The participants will also receive a certificate at the closing ceremony in presence of their parents.

Mathematics workshops Maths en jeu

Studies show that in many industrial countries, girls do not have enough confidence in their capacities concerning mathematics. For example, the PISA survey 2012[1] reports that the students' attitudes towards mathematics are already well-formed by the time students are 15. Even when girls and boys perform equally well, girls are more likely to feel anxious towards mathematics, and have less confidence than boys in their own mathematical skills and in their ability to solve mathematics problems. This contributes later strongly to the non-choice of scientific fields by young girls.
In order to overcome this problem, targeted actions are necessary. This is the case of the mathematics workshops *Maths en jeu* aimed at girls from 11 to 13 years old. The main purpose of these courses is to discover mathematics from a common sense point of view. Mathematics is not presented as a hindrance but as a useful tool for a future scientific career. Beginning with one course, following the enthusiasm of young girls to continue such workshops, advanced mathematical workshops have been developed. These workshops are organized within the framework of the collaboration program of NCCRs Synapsy and MARVEL and the EPFL Equal Opportunities Office.
Activities organized for young girls to increase their interest in these fields are very success-

[1] "Are boys and girls equally prepared for life", Organization for Economic Co-operation and Development (OCDE) 2014. Document available on http://www.oecd.org/pisa/pisaproducts/PIF-2014-gender-international-version.pdf

ful and a big demand is expressed by children and parents to organize more activities. It is planned to develop also workshops and activities for 13 – 15 years old girls in the MARVEL fields.

MARVEL webpage for equal opportunities

The next important goal of MARVEL is to build the webpage for equal opportunities in order to communicate with group leaders and the whole community of MARVEL on this issue. It is also very important to gather the necessary information concerning measures in the different institutions participating in MARVEL and to announce events on this website.

6 Communication

As stated in the communication strategy (Annexe 4), the communication effort of MARVEL has two trends. The first one is oriented to internal communication, between the members of the NCCR. The second one wants to reach various publics outside the NCCR, scientists in the domain of material simulation at the international level, media, industries, as well as the general public.

Internal communication

These first months, the communication action was mainly directed to internal communication. The first goal has been to build the MARVEL community and network. This was done on the one hand by organizing various meetings of the group leaders, and on the other hand by communicating to them different actions supported by the NCCR, such as distinguished lectures, support and sponsoring of conferences and workshops, as well as all the explanations concerning the reporting process and the rules for financial reimbursements.

Meetings

Kick-off meeting

The kick-off meeting took place at the very beginning of the project on May 15 and 16, 2014 at EPFL, bringing together almost all the project leaders — PIs and Agility members — and senior researchers. The meeting was also attended by Boris Kozinsky from the Scientific Advisory Board and Andrew Millis (Columbia University and "Simons Collaboration on the Many Electron Problem"). The financial management of the NCCR was briefly introduced and explained. The seven scientific projects were presented by their respective project leaders and discussed, and the last afternoon was used for brainstorming in groups. 22 persons were present the first day and 15 the second.

Review and retreat

Every year, the "MARVEL Review & Retreat", an annual meeting of all MARVEL participants, project leaders, group leaders, senior researchers as well as postdoctoral and doctoral students, is planned at the beginning of September. The first meeting took place on two and a half days on September 8 – 10, 2014, at EPFL and gathered more than fifty participants. Almost all the group leaders came, at least for part of the meeting, and about thirty participants were either postdocs or PhD students.

The program planned two or three presentations for each of the seven projects, given by the project leaders as well as participating group leaders or senior researchers (Fig. 1). Parts to the two afternoon were planned for brainstorming in small or medium groups. All the present group leaders (18 on 24) met for dinner on the Monday evening. A poster session from the attending students was organized on the Tuesday evening with food and wine and further discussions for everybody. The first meeting of the Scientific Committee took also place during the Review and Retreat.

Experimental cooperation

The platform project 7 on experiments will only start in the second year of MARVEL. Some preliminary communication work started already this year. Some information meetings were organized at PSI and Empa in August, as well as a one day workshop on "NCCR MARVEL — Experimental verification" on October 17, 2014, at PSI. This meeting gathered over fifty participants, either from the computational side, or from the experimental side of MARVEL. More details can be found in the PP7 project description, p. 100.

Figure 1: *First MARVEL Review & Retreat on September* 8 – 10*, 2014 at EPFL, with presentations on ongoing research and a lot of discussions.*

NCCR MARVEL distinguished lectures

We have planned to regularly invite top scientists in the field of computational design and discovery of novel materials to give "NCCR MARVEL Distinguished Lectures" either at EPFL or in one of the other participating institutions. In this first year, two distinguished lectures were already organized at EPFL.

- On Thursday October 9, 2014, Prof. Alex Zunger from the University of Colorado in Boulder gave a very interesting lecture on "The Inverse Problem in materials theory: Find the system that has a given target property" to a public of about 70 persons (Fig. 2).

- On Wednesday January 14, 2015, Prof. Gustavo Scuseria from Rice University (Houston, Texas) presented "The strong correlation problem: A quantum chemistry perspective" and about 50 persons attended to this presentation.

The idea in the future would be to telecast or record these lectures, so that they become available to the other MARVEL institutions.

Figure 2: *First "NCCR MARVEL Distinguished Lectures" by Prof. Alex Zunger from the University of Colorado on October 9, 2014, at EPFL.*

Other meetings

Other actions, such as meetings and tutorials, inside the MARVEL community, are reported in the knowledge and technology transfer an education and training chapters.

External communication

Website

For general communication, both for internal and external purposes, one important work is the development of a website. As MARVEL is a big project with partners all over Swiss academic and industrial institutions, the task of the website is mostly communication of this project to other academics and to the public, but should also allow a member-only communication platform for the core members, in a flexible way. It should also be able to advertise workshops, conferences and other events to the interest of the MARVEL community and, in a short term, a newsletter should be set up. For this development, we have chosen a small company in Lausanne, *ttree* (`http://ttree.ch`), proposing a multistep approach. The first step will propose an informative web site, with an approach that gives entries to various publics, from the specialists from the material simulation community, to the general public, going through other interested public, such as media and industry partners. A first English version of the site should be available in March 2015. A first idea of the homepage is visible on Fig. 3. It will be developed in the next months, including other functions, such as inscription to conference and a newsletter, as well as a French version, with a priority to the industry, media and general public pages.

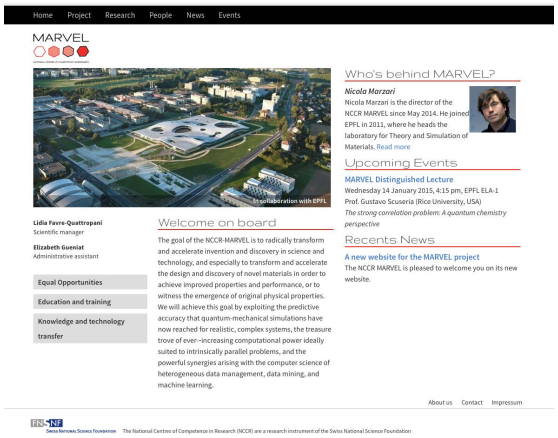


Figure 3: *Preview of the provisional homepage of the MARVEL website.*

Artistic work

A way to bring science to the general public is to go through art and this approach is further developed in the communication strategy in annexe 4.
In this direction, Nicola Spaldin, at ETHZ, initiated preliminary contacts with Julie Birenbaum, an artist whose work has included conveying the scientific process through light and images taken in a scientific environment, such as blackboards with quantum mechanics formulas (see for example `http://www.juliehbirenbaum.com/#!galerie-off-the-wall/`). Our plan is to develop an art exhibit with accompanying scientific explanations which will first be shown at EPFL and will then tour the other MARVEL sites
At the beginning of 2015, a visualization contest (`http://acces.epfl.ch/contest`) is organized by the "Application-Centered Computational Engineering Science" (ACCES) collaboration at EPFL for Bachelor and Master students enrolled in a section of the School of Engineering (STI) or of the School of Architecture, Civil and Environmental Engineering (ENAC), and from PhD students and post-docs employed by a laboratory fully or partially affiliated with these schools. This contest aims to foster interest in advanced visualization by encouraging the insightful and visually engaging presentation of scientific and engineering data. Entries must have a scientific/engineering content that is computer generated (not a photo), and be based on the contributors' own numerical simulation or experimental data. Winning entries will be exhibited in public display areas at EPFL and on the ACCES web site. EPFL-based MARVEL students can participate at this contest.

Such visualisation contest could also be organized in the framework of MARVEL for doctoral and postdoctoral students to propose innovative ways of visually presenting their data, with an exhibition of the resulting works and prize giving during the MARVEL Review & Retreat. This first edition might take place already this year in September.

Events

In the future MARVEL has planned to participate at events for the general public. Some hands-on experiments and demonstrators should be developed for such events. As MARVEL specificity is the *computational* design and discovery of novel materials, the ideas is to find a nice way to show to the public how computer based research brings its contribution to material research. A collaboration with the scientific mediators of the Science Outreach Department at EPFL is planned.

Other communication-related actions

Since August 2014, Lidia Favre-Quattropani is participating at the monthly meeting of all EPFL Communication officers.
She also attended the ScienceComm'14 conference in Beromünster/Sursee on September 18 and 19, 2014. This national conference organized by the *Science et Cité* foundation brings together Swiss science communication experts and offers them a platform for exchanging views and information. This edition was the fourth one and addressed topics such as relations between science, media and politics, science and art, and science visualization.

MARVEL in the EPFL news and in the press

Since the decision of the eight new National Centres of Competence in Research by the Federal Department of Economic Affairs, Education and Research in December 2013, MARVEL was mentioned several times in the electronic news of EPFL, some of them taken up in the local press or the EPFL internal *Flash* newspaper.

- 17.12.2013, "EPFL has become the Swiss capital for research on new materials" by Emmanuel Barraud, EPFL (http://actu.epfl.ch/news/epfl-has-become-the-swiss-capital-for-research-on-/). The announce of the eight new NCCRs was largely taken up by the Swiss press, and that particular news on MARVEL was developed in *24 Heures* on December 20,

2013 as well as in the *Flash* newspaper of January 2014.

- 23.06.2014, "EPFL to participate in EU Graphene Flagship project" by Nik Papageorgiou, EPFL (http://actu.epfl.ch/news/epfl-to-participate-in-eu-graphene-flagship-projec/). This news mentions the participation of Nicola Marzari, director of MARVEL, to this European research initiative.
- 12.11.2014, "An Electrical Wire a Few Atoms Wide" by Laure-Anne Pessina, EPFL (http://actu.epfl.ch/news/an-electrical-wire-a-few-atoms-wide/). An article published by the group of Nicola Marzari in *Nature Communications* is presented here to the general public and the local stakeholders. This news was also taken up in the *Flash* newspaper of December 2014.
- 12.01.2015, "Nicola Marzari is named Fellow of the American Physical Society" by Laure-Anne Pessina (http://actu.epfl.ch/news/nicola-marzari-named-fellow-of-the-american-physic/).

7 Structural aspects

We detail below the structural measures that have been implemented at EPFL

Professorships

In March 2013, Prof. Michele Ceriotti was hired as a tenure-track assistant professor in materials science in the Faculty of Engineering.
In December 2014, Prof. Oleg Yazyev, currently SNF-funded professor at EPFL, was hired as a tenure-track assistant professor in theoretical physics in the Faculty of Basic Sciences.

Infrastructure

New building

MARVEL will move in the new ME building (Fig. 1) in early spring 2016, where space has been reserved for all the personnel overseen by Nicola Marzari (Fig. 2). NCCR Robotics will also move in the same building, on the first floor, greatly facilatating inter-NCCR synergies.

Distant education classroom

A distant education classroom is being set up in the space contiguous to the new MARVEL location, in ME D2 1124. This room will have 28 seats, 4 large screens (two behind the lecturer, two in front of the lecturer), electronic whiteboard, computer facilities and a large electronic tablet, plus microphones and cameras following automatically the lectures.

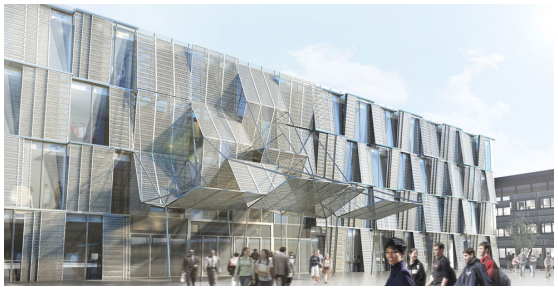


Figure 1: *Rendering of the new ME building, a low-energy-consumption (Minergie) building designed by Dominique Perrault (http://www.perraultarchitecte.com/en/homepage/), EPFL professor (among his other projects, the French National Library in Paris, and Ewha Womens University in Seoul).*

Visualization facility

As an effort of the home institution, EPFL financed a collaborative visualisation facility (ACCES CoViz, Fig. 3), available to all EPFL MARVEL members, as well as other ACCES (Application-Centered Computational Engineering Science) members. This facility is intended to facilitate collaborative examination of computational engineering datasets, essential as a research, teaching and outreach tool for scientific discovery and communication. It

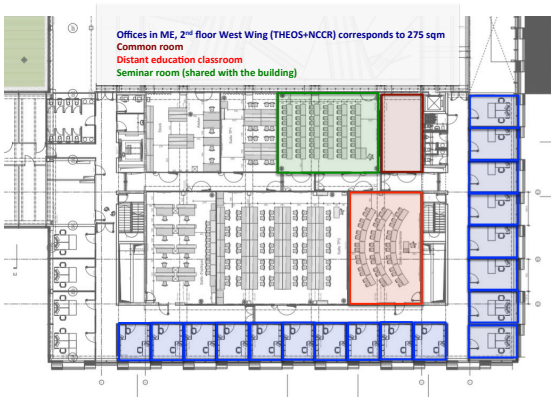


Figure 2: *Layout of the west wing of the second floor of ME building, dedicated to MARVEL and THEOS. Office space is in blue, the distant education classroom in red, the seminar room in green, the common room in brown. Close proximity with the group of Prof. Curtin (an expert in multiscale simulations and mechanical properties of materials) on the east wing will increase synergies, and sharing of other common spaces, dedicated to a library and reading room, and a meeting room.*

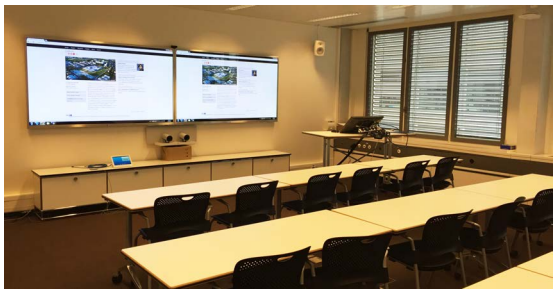


Figure 3: *The ACCES collaborative visualisation facility at EPFL in MXC 320, contiguous to the offices where the current MARVEL personnel works.*

is also useful for distant education. It is comprised of the following elements:

- two 84" UltraHD displays (LG 84LM960V);
- display server (dual 8-core Intel Xeon E5-2660, 128 GB, 2 Nvidia Quadro K5000);
- render server (dual 4-core Intel Xeon E5-2609, 64 GB, 4 Nvidia GeForce Titan);
- presentation server (15" MacBook Pro, Intel i7, 16 GB, 1 TB SSD, retina screen, external BluRay writer);
- video/audio processor + BluRay writer (OPPO BDP-105EU);
- 4 high-quality active speakers;
- quality videoconferencing system (Cisco SX80 with SpeakerTrack cameras plus additional PTZ camera, multiple microphones);
- 22" pen touch display (Wacom DTH-2242);
- 10.1" control tablet (Samsung Galaxy Note);
- advanced control devices (e.g. 3D gyroscopic mouse, 3D space mouse);
- a standard full-wall whiteboard;
- cabled (Gigabit Ethernet) connection to internet;
- wireless connectivity with participants' laptops for interactive sharing of visual data;
- flexible seating for up to 20 participants.

Additional details can be found on the ACCES web site: `http://acces.epfl.ch/coviz`.

Database cluster

As discussed in PP6, a collaborative agreement between EPFL, ETHZ and CSCS has allowed the co-location of all the computational infrastructure shared by all Swiss members at CSCS, with in-kind matching funds from ETHZ of 0.82 M CHF for the cash contribution of EPFL of 0.5 M CHF, dedicated to the storage facilities for the MARVEL database.

8 Management

Since May 2014, and the official start of the NCCR MARVEL, the management of MARVEL has been set up gradually. A team of administrative and scientific collaborators has been formed to carry out management and financial duties and take care of the other responsibilities, such as communication, knowledge and technology transfer, education and equal opportunities.

Organizational and procedural changes

MARVEL is hosted at EPFL. As for the other NCCRs hosted at EPFL, the financial aspect of the project is taken care directly by the National Funding Group of EPFL Research Office. To this end, it hired Cheikhna Mangassouba as financial assistant in May 2014. Unfortunately Cheikhna Mangassouba had to leave Switzerland for family reasons at the end of September. He was replaced by Valérie Le Dreau who was already working at the EPFL Reseach Office, taking care of the financial reporting and participating in the budget process of the NCCR Synapsis. Her knowledge of the processes and of the operation of the NCCRs allowed her to rapidly take care effectively of all the duties in relation with her charge.

Since June 23, 2014, Dr. Lidia Favre-Quattropani is the scientific manager of the NCCR. She is in charge of all the scientific questions regarding the management and the communication of the NCCR. She also took care in 2014 of other more administrative questions, in particular the organization of some events, until the arrival on December 1, 2014 of an administrative assistant dedicated to MARVEL, Elizabeth Gueniat, and the upcoming hiring of a program manager (hopefully for April or May 2015). Before the arrival of Elizabeth Gueniat, Irène Laroche, the administrative assistant of the research group of Nicola Marzari, took care of a lot of administrative duties of MARVEL, such as funds opening, invoices payement, events or travel organization. It is planned for spring 2015 to complete the team with the involvement dedicated one day a week for MARVEL of Pascale Van Landuyt from Alliance, the industrial liaison program of the French part of Switzerland, supported by EPFL.

The MARVEL management, together with the executive committee, took also care of the implementation of the various rules and procedures specific to the NCCRs. In collaboration with EPFL Research Office legal counsel, it prepared the "NCCR-MARVEL Regulations" (MARVEL internal rules), the regulations with IBM, MARVEL industrial partner, "regarding intellectual property rights and technology transfer", as well as an agreement between EPFL and CSCS for the purchase of a supercomputer at CSCS.

Some contacts and first agreements were also taken with external partners. At a local level, a contract was signed between Nicola Marzari and an industrial partner with a financial contribution to the NCCR for four years. At an international level, contacts are taken with different scientific partners such as the "Simons Collaboration on the Many Electron Problem", with a sketch of memorandum of understanding between its director Andrew Millis and Nicola Marzari. Contributions were payed from MARVEL to the Psi-k Network and the Quantum-ESPRESSO Foundation as group memberships.

Concerning the organization of the scientific projects of MARVEL, a change in the structure appears concerning Prof. Wanda Andreoni. As she will be retiring end of February 2015 and pursue her activities in MARVEL as a professor emeritus, the budget dedicated to cover the scientific personal working for her agility project is formally allocated to her through Prof. Alfredo Pasquarello as a group leader.

Activities and measures

Hiring activities

When a project is starting, hiring personnel is an important task. This was the case in all research groups. This was also the case for the director, as far as the management is concerned. He hired the scientific manager in June and an administrative assistant in December. A program and communications manager is currently being recruited.

NIRA database

As the NCCR is just beginning, filling in NIRA, the SNSF database dedicated to the NCCRs, is a time consuming task, in particular as far as the personnel is concerned, and the management has been extremely busy collecting the data and consolidating them into NIRA.

Events organisation

In 2014, the MARVEL management organized two scientific meetings that are described in more details in the communication chapter (chapter 6).

- The kick-off meeting took place on May 15 and 16, 2014 at EPFL, with almost all the project leaders and some senior researchers. 22 persons were present the first day and 15 the second.
- The first "MARVEL Review & Retreat" took place on September 8 – 10, 2014, at EPFL, with more than fifty participants. Almost all the group leaders came, at least for part of the meeting, and about thirty participants were either postdocs or PhD students.

It also organized two NCCR MARVEL distinguished lectures at EPFL with the venue on October 9, 2014 of Prof. Alex Zunger from the University of Colorado in Boulder and on January 14, 2015 of Prof. Gustavo Scuseria from Rice University (Houston, Texas). Both events are described in more details in the communication chapter (chapter 6).

As educational activities, the management team has just begun to organize the first MARVEL junior retreat, that will take place at the Hotel Boldern, in Mannedorf near Zurich from July 8 to 10, 2015.
MARVEL management will participate at the Platform for Advances Scientific Computing PASC15 conference on June 1 – 3, 2015 at ETHZ. An information stand will be set-up promoting MARVEL activities.

Other activities

The following activities are also part of the management duties.

- MARVEL is preparing its website with the help the company *ttree* (`http://ttree.ch`) in Lausanne. It should be online in March 2015.
- MARVEL management is sponsoring through its education and training budget the international conference *Nothing is Perfect – The Quantum Mechanics of Defects*, in Ascona (TI) on April 26 – 29 2015 and the participation of MARVEL-related students to the annual Simons Many Electron Collaboration Summer School at the Simons Center for Geometry and Physics at Stony Brook on June 8 – 12 2015 through the payment of their travel to and from the school.
- On September 5, 2014, Nicola Marzari, Nicola Spaldin and Lidia Favre-Quattropani participated at the "Leadership and stimulation of transcending collaboration in research networks" workshop organized by the SNSF for the newest NCCRs. On February 3 – 5, 2015, Alfredo Pasquarello and Lidia Favre-Quattropani will take part in the Management Workshop also organized by the SNSF for the NCCRs of the third and fourth series.

9 Lists

9.3 Publications since the beginning of MARVEL

All publications have been entered in NIRA, and are also listed below for convenience, sorted by group leaders. We list publications either resulting directly from the NCCR (marked with a red hexagon) or with minor contributions from the NCCR.
The following lists cover the period from May 1st, 2014 to January 31st, 2015.

9.3.1 Scientific articles in journals with peer review

Group of Michele Ceriotti

- P. Gasparotto and M. Ceriotti
  Recognizing molecular patterns by machine learning: An agnostic structural definition of the hydrogen bond
  The Journal of Chemical Physics **141**, 174110 (2014).
  Group(s): Ceriotti / Project(s): HP4

Group of Alessandro Curioni

- K. Meier, T. Laino, and A. Curioni
  Solid-State Electrolytes: Revealing the Mechanisms of Li-Ion Conduction in Tetragonal and Cubic LLZO by First-Principles Calculations
  The Journal of Physical Chemistry C **118**, 6668 (2014).
  Group(s): Curioni / Project(s): VP2

Group of Antoine Georges

O. E. Peil, M. Ferrero, and A. Georges
Orbital polarization in strained $LaNiO_3$: Structural distortions and correlation effects
Physical Review B **90**, 045128 (2014).
Group(s): Georges / Project(s): VP1

D. Stricker, J. Mravlje, C. Berthod, R. Fittipaldi, A. Vecchione, A. Georges, and D. van der Marel
Optical Response of Sr_2RuO_4 Reveals Universal Fermi-Liquid Scaling and Quasiparticles Beyond Landau Theory
Physical Review Letters **113**, 087404 (2014).
Group(s): Georges / Project(s): VP1

Group of Jürg Hutter

U. Borštnik, J. VandeVondele, V. Weber, and J. Hutter
Sparse matrix multiplication: The distributed block-compressed sparse row library
Parallel Computing **40**, 47 (2014).
Group(s): Hutter, VandeVondele / Project(s): HP3

M. Del Ben, O. Schütt, T. Wentz, P. Messmer, J. Hutter, and J. VandeVondele
Enabling simulation at the fifth rung of DFT: Large scale RPA calculations with excellent time to solution
Computer Physics Communications **187**, 120 (2015).
Group(s): Hutter, VandeVondele / Project(s): HP3

Group of Michel Kenzelmann

- M. Morin, A. Scaramucci, M. Bartkowiak, E. Pomjakushina, G. Deng, D. Sheptyakov, L. Keller, J. Rodriguez-Carvajal, N. A. Spaldin, M. Kenzelmann, K. Conder, and M. Medarde
  Incommensurate magnetic structure, Fe/Cu chemical disorder, and magnetic interactions in the high-temperature multiferroic $YBaCuFeO_5$
  Physical Review B **91**, 064408 (2015).
  Group(s): Kenzelmann, Spaldin / Project(s): VP1, PP7

Group of Nicola Marzari

G. Borghi, A. Ferretti, N. L. Nguyen, I. Dabo, and N. Marzari
Koopmans-compliant functionals and their performance against reference molecular data
Physical Review B **90**, 075135 (2014).
Group(s): Marzari / Project(s): HP3, VP2

I. Dabo, A. Ferretti, and N. Marzari
Piecewise Linearity and Spectroscopic Properties from Koopmans-Compliant Functionals
Topics in Current Chemistry **347**, 193 (2014).
Group(s): Marzari / Project(s): HP3, VP2

A. Ferretti, I. Dabo, M. Cococcioni, and N. Marzari
Bridging density-functional and many-body perturbation theory: Orbital-density dependence in electronic-structure functionals
Physical Review B **89**, 195134 (2014).
Group(s): Marzari / Project(s): HP3, VP2

S. Bruzzone, G. Iannaccone, N. Marzari, and G. Fiori
An Open-Source Multiscale Framework for the Simulation of Nanoscale Devices
IEEE Transactions on Electron Devices **61**, 48 (2014).
Group(s): Marzari / Project(s): PP6

A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari
An updated version of `wannier90`: A tool for obtaining maximally-localised Wannier functions
Computer Physics Communications **185**, 2309 (2014).
Group(s): Marzari / Project(s): PP6

G. Pizzi, D. Volja, B. Kozinsky, M. Fornari, and N. Marzari
An updated version of BoltzWann: A code for the evaluation of thermoelectric and electronic transport properties with a maximally-localized Wannier functions basis
Computer Physics Communications **185**, 2311 (2014).
Group(s): Marzari / Project(s): PP6

G. Pizzi, D. Volja, B. Kozinsky, M. Fornari, and N. Marzari
BoltzWann: A code for the evaluation of thermoelectric and electronic transport properties with a maximally-localized Wannier functions basis
Computer Physics Communications **185**, 422 (2014).
Group(s): Marzari / Project(s): PP6

C.-H. Park, N. Bonini, T. Sohier, G. Samsonidze, B. Kozinsky, M. Calandra, F. Mauri, and N. Marzari
Electron-Phonon Interactions and the Intrinsic Electrical Resistivity of Graphene
Nano Letters **14**, 1113 (2014).
Group(s): Marzari / Project(s): PP6, VP2

T. Sohier, M. Calandra, C.-H. Park, N. Bonini, N. Marzari, and F. Mauri
Phonon-limited resistivity of graphene by first-principles calculations: Electron-phonon interactions, strain-induced gauge field, and Boltzmann equation
Physical Review B **90**, 125414 (2014).
Group(s): Marzari / Project(s): PP6, VP2

O. Andreussi and N. Marzari
Electrostatics of solvated systems in periodic boundary conditions
Physical Review B **90**, 245105 (2014).
Group(s): Marzari / Project(s): VP2

N. Bonnet, I. Dabo, and N. Marzari
Chemisorbed Molecules under Potential Bias: Detailed Insights from First-Principles Vibrational Spectroscopies
Electrochimica Acta **121**, 210 (2014).
Group(s): Marzari / Project(s): VP2

N. Bonnet and N. Marzari
Static Dielectric Permittivity of Ice from First Principles
Physical Review Letters **113**, 245501 (2014).
Group(s): Marzari / Project(s): VP2

M. Gibertini, F. M. D. Pellegrino, N. Marzari, and M. Polini
Spin-resolved optical conductivity of two-dimensional group-VIB transition-metal dichalcogenides
Physical Review B **90**, 245411 (2014).
Group(s): Marzari / Project(s): VP2

M. Gibertini, G. Pizzi, and N. Marzari
Engineering polar discontinuities in honeycomb lattices
Nature Communications **5**, 5157 (2014).
Group(s): Marzari / Project(s): VP2

G. Fugallo, A. Cepellotti, L. Paulatto, M. Lazzeri, N. Marzari, and F. Mauri
Thermal Conductivity of Graphene and Graphite: Collective Excitations and Mean Free Paths
Nano Letters **14**, 6109 (2014).
Group(s): Marzari / Project(s): VP2, PP6

Group of Michele Parrinello

- M. Salvalaglio, M. Mazzotti, and M. Parrinello
  Urea homogeneous nucleation mechanism is solvent dependent
  Faraday Discussions (2015), doi:10.1039/C4FD00235K.
  Group(s): Parrinello / Project(s): HP4

- O. Valsson and M. Parrinello
  Variational Approach to Enhanced Sampling and Free Energy Calculations
  Physical Review Letters **113**, 090601 (2014).
  Group(s): Parrinello / Project(s): HP4

- M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello
  Molecular-dynamics simulations of urea nucleation from aqueous solution
  Proceedings of the National Academy of Science of the USA **112**, E6 (2015).
  Group(s): Parrinello / Project(s): HP4

Group of Alfredo Pasquarello

W. Chen and A. Pasquarello
Band-edge positions in GW: Effects of starting point and self-consistency
Physical Review B **90**, 165133 (2014).
Group(s): Pasquarello / Project(s): VP2

- G. Miceli, S. de Gironcoli, and A. Pasquarello
  Isobaric first-principles molecular dynamics of liquid water with nonlocal van der Waals interactions
  The Journal of Chemical Physics **142**, 034501 (2015).
  Group(s): Pasquarello / Project(s): VP2

Group of Daniele Passerone

R. Gaspari, R. Erni, Y. Arroyo, M. Parlinska-Wojtan, J. Dshemuchadse, C. A. Pignedoli, D. Passerone, P. Schmutz, and A. Beni
Real space crystallography of a complex metallic alloy: high-angle annular dark-field scanning transmission electron microscopy of o-Al_4(Cr,Fe)
Journal of Applied Crystallography **47**, 1026 (2014).
Group(s): Passerone / Project(s): VP2

R. Jaafar, C. A. Pignedoli, G. Bussi, K. Aït-Mansour, O. Gröning, T. Amaya, T. Hirao, R. Fasel, and P. Ruffieux
Bowl Inversion of Surface-Adsorbed Sumanene
Journal of the American Chemical Society **136**, 13666 (2014).
Group(s): Passerone / Project(s): VP2

J. Prinz, C. A. Pignedoli, Q. S. Stöckl, M. Armbrüster, H. Brune, O. Gröning, R. Widmer, and D. Passerone
Adsorption of Small Hydrocarbons on the Three-Fold PdGa Surfaces: The Road to Selective Hydrogenation
Journal of the American Chemical Society **136**, 11792 (2014).
Group(s): Passerone / Project(s): VP2

R. Denk, M. Hohage, P. Zeppenfeld, J. Cai, C. A. Pignedoli, H. Söde, R. Fasel, X. Feng, K. Müllen, S. Wang, D. Prezzi, A. Ferretti, A. Ruini, E. Molinari, and P. Ruffieux
Exciton-dominated optical response of ultra-narrow graphene nanoribbons
Nature Communications **5**, 4253 (2014).
Group(s): Passerone / Project(s): VP2

J. Cai, C. A. Pignedoli, L. Talirz, P. Ruffieux, H. Söde, L. Liang, V. Meunier, R. Berger, R. Li, X. Feng, K. Müllen, and R. Fasel
Graphene nanoribbon heterojunctions
Nature Nanotechnology **9**, 896 (2014).
Group(s): Passerone / Project(s): VP2

D. J. Adams, S. Chappellet, F. Lincker, M. Ibn-Elhaj, B. Watts, M. Iannuzzi, D. S. Jung, C. A. Pignedoli, and D. Passerone
Identifying Photoreaction Products in Cinnamate-Based Photoalignment Materials
The Journal of Physical Chemistry C **118**, 15422 (2014).
Group(s): Passerone / Project(s): VP2

J. Prinz, R. Gaspari, Q. S. Stöckl, P. Gille, M. Armbrüster, H. Brune, O. Gröning, C. A. Pignedoli, D. Passerone, and R. Widmer
Ensemble Effect Evidenced by CO Adsorption on the 3-Fold PdGa Surfaces
The Journal of Physical Chemistry C **118**, 12260 (2014).
Group(s): Passerone / Project(s): VP2

- H. Söde, L. Talirz, O. Gröning, C. A. Pignedoli, R. Berger, X. Feng, K. Müllen, R. Fasel, and P. Ruffieux
  Electronic band dispersion of graphene nanoribbons via Fourier-transformed scanning tunneling spectroscopy

Physical Review B **91**, 045429 (2015).
Group(s): Passerone / Project(s): VP2

Group of Ursula Röthlisberger

- S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B. F. E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M. K. Nazeeruddin, and M. Grätzel
  Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers
  Nature Chemistry **6**, 242 (2014).
  Group(s): Röthlisberger / Project(s): VP2

Group of Thomas Schulthess

P. Staar, T. Maier, and T. C. Schulthess
Two-particle correlations in a dynamic cluster approximation with continuous momentum dependence: Superconductivity in the two-dimensional Hubbard model
Physical Review B **89**, 195133 (2014).
Group(s): Schulthess / Project(s): PP6

I.-H. Chu, A. Kozhevnikov, T. C. Schulthess, and H.-P. Cheng
All-electron GW quasiparticle band structures of group 14 nitride compounds
The Journal of Chemical Physics **141**, 044709 (2014).
Group(s): Schulthess / Project(s): PP6

Group of Nicola Spaldin

- M. Morin, A. Scaramucci, M. Bartkowiak, E. Pomjakushina, G. Deng, D. Sheptyakov, L. Keller, J. Rodriguez-Carvajal, N. A. Spaldin, M. Kenzelmann, K. Conder, and M. Medarde
  Incommensurate magnetic structure, Fe/Cu chemical disorder, and magnetic interactions in the high-temperature multiferroic $YBaCuFeO_5$
  Physical Review B **91**, 064408 (2015).
  Group(s): Kenzelmann, Spaldin / Project(s): VP1, PP7

Group of Matthias Troyer

- L. Wang and M. Troyer
  Renyi Entanglement Entropy of Interacting Fermions Calculated Using the Continuous-Time Quantum Monte Carlo Method
  Physical Review Letters **113**, 110401 (2014).
  Group(s): Troyer / Project(s): VP1, HP3

Group of Joost VandeVondele

M. Del Ben, O. Schütt, T. Wentz, P. Messmer, J. Hutter, and J. VandeVondele
Enabling simulation at the fifth rung of DFT: Large scale RPA calculations with excellent time to solution
Computer Physics Communications **187**, 120 (2015).
Group(s): Hutter, VandeVondele / Project(s): HP3

U. Borštnik, J. VandeVondele, V. Weber, and J. Hutter
Sparse matrix multiplication: The distributed block-compressed sparse row library
Parallel Computing **40**, 47 (2014).
Group(s): Hutter, VandeVondele / Project(s): HP3

Group of Anatole von Lilienfeld

K. Y. S. Chang and O. A. von Lilienfeld
Quantum Mechanical Treatment of Variable Molecular Composition: From 'Alchemical' Changes of State Functions to Rational Compound Design
CHIMIA International Journal for Chemistry **68**, 602 (2014).
Group(s): von Lilienfeld / Project(s): HP4

Group of Philipp Werner

L. Huang, T. Ayral, S. Biermann, and P. Werner
Extended dynamical mean-field study of the Hubbard model with long-range interactions
Physical Review B **90**, 195114 (2014).
Group(s): Werner / Project(s): HP3

9.3.2 Scientific articles in journals without peer review

Group of Antoine Georges

- A. Subedi, O. E. Peil, and A. Georges
  A low energy description of the metal-insulator transition in the rare-earth nickelates
  arXiv:1410.2830 (2014).
  Group(s): Georges / Project(s): VP1

Group of Ursula Röthlisberger

- S. Meloni, G. Palermo, N. Ashari Astani, B. F. E. Curchod, M. Graetzel, and U. Roethlisberger
  Valence and conduction bands engineering in halide perovskites for solar cell applications
  arXiv:1412.3659 (2014).
  Group(s): Röthlisgerger / Project(s): VP2

Group of Matthias Troyer

- M. Iazzi, A. A. Soluyanov, and M. Troyer
  Topological Origin of the Fermion Sign Problem
  arXiv:1410.8535 (2014).
  Group(s): Troyer / Project(s): HP3, VP1

- M. Iazzi and M. Troyer
  Efficient continuous-time quantum Monte Carlo algorithm for fermionic lattice models
  arXiv:1411.0683 (2014).
  Group(s): Troyer / Project(s): HP3, VP1

- H. Shinaoka, R. Sakuma, M. Troyer, and P. Werner
  Accuracy of downfolding based on the constrained random phase approximation
  arXiv:1410.1276 (2014).
  Group(s): Troyer, Werner / Project(s): HP3, VP1

Group of Philipp Werner

P. Werner, R. Sakuma, F. Nilsson, and F. Aryasetiawan
Dynamical screening in La_2CuO_4
arXiv:1411.3952 (2014).
Group(s): Werner / Project(s): HP3

- H. Shinaoka, R. Sakuma, M. Troyer, and P. Werner
  Accuracy of downfolding based on the constrained random phase approximation
  arXiv:1410.1276 (2014).
  Group(s): Troyer, Werner / Project(s): HP3, VP1

Annexe 1 Strategy for Knowledge and Technology Transfer

Knowledge and technology transfer in the NCCR will happen at several levels, from the dissemination of open-source materials' simulation codes, to the training in the use of those codes and of the newly developed materials informatics framework, to the sharing of all results from materials simulations in a network of open-access servers, to the verification and validation of calculations. Training and education will not only target the traditional area of academic computational groups, but also experimental groups, embedding the use of first-principles spectroscopies and microscopies into the toolbox of those researchers, and a special effort will be made to transitioning these tools to local and global industries, thanks also to the leadership role that our partner IBM can take, and the industrial partners in our Scientific Advisory Board. Last, we will actively engage key stakeholders in the public sector, from supercomputing centers to policy makers.

Assessment of the current state

The impact of quantum simulations to science and technology can be, however cursorily, gauged even by simple citation counts — to give a few examples (i) the top 16 most cited papers in the entire history of the American Physical Society are on electronic structure simulations (Fig. 1), or, similarly, (ii) there are 13 papers on electronic structure simulations in the top 100 most cited paper in the entire history of the scientific literature worldwide [1]. These citation counts are ultimately the result of the wide availability of user-friendly and robust quantum simulation codes (many distributed under a free, open-source model), able to run state-of-the-art calculations even on a single computer (and thus pursuing a genuinely democratic model for science that is available to researchers around the world without constraints on financial resources). The impact of such a computational laboratory, improving in throughput capacity at the current sustained speed (doubling every 14 months), and shared worldwide by a very broad scientific community cannot be underestimated.

The relevance of quantum simulations towards scientific advance, technological innovation and industrial competitiveness is becoming more broadly acknowledged both by funding agencies and policy makers — most notably in the case of the Materials Genome Initiative, launched in 2011 by the Office of Science and Technology Policy of the White House (http://www.whitehouse.gov/mgi) for a "new, multistakeholder effort to develop an infrastructure to accelerate advanced materials discovery and deployment" with a "vision of how the development of advanced materials can be accelerated through advances in computational techniques, more effective use of standards, and enhanced data management." Among the many initiatives pioneering these efforts, we should mention the Materials Project (http://www.materialsproject.org), the SUNCAT center for interface science and catalysis (http://suncat.slac.stanford.edu), the Center for Inverse Design (http://www.centerforinversedesign.org).

Against this backdrop, Europe and Switzerland provide a very healthy and active community in the development and application of these techniques to the most relevant scientific, technological, and societal problems. In Europe, three organizations have been at the forefront of dissemination and knowledge transfer in materials simulations — CECAM, Psi-k, and ICTP. CECAM (the *Centre Européen de Calcul Atomique et Moléculaire*), founded in Paris in 1969, has moved its headquarters to EPFL in 2008, with Prof. Wanda Andreoni as its first appointed director (also, the current director, Prof. Dominic Tildesley, is in the Scientific Advisory Board of MARVEL), and provides a healthy activity of 25+ workshops every year in its headquarters, together with activities in the 18 national nodes. Psi-k is a

	Journal	# cites	Title	Author(s)
1	PRB (1988)	39190	Development of the Colle-Salvetti Correlation-Energy ...	Lee, Yang, Parr
2	PRL (1996)	25452	Generalized Gradient Approximation Made Simple	Perdew, Burke, Ernzerhof
3	PRA (1988)	22904	Density-Functional Exchange-Energy Approximation ...	Becke
4	PR (1965)	20142	Self-Consistent Equations Including Exchange and Correlation ...	Kohn and Sham
5	PRB (1996)	13731	Efficient Iterative Schemes for Ab Initio Total-Energy ...	Kresse and Furthmuller
6	PRB (1976)	13160	Special Points for Brillouin-Zone Integrations	Monkhorst and Pack
7	PRB (1992)	10876	Accurate and Simple Analytic Representation of the Electron ...	Perdew and Wang
8	PRB (1999)	10007	From Ultrasoft Pseudopotentials to the Projector Augmented ...	Kresse and Joubert
9	PRB (1990)	9840	Soft Self-Consistent Pseudopotentials in a Generalized ...	Vanderbilt
10	PR (1964)	9789	Inhomogeneous Electron Gas	Hohenberg and Kohn
11	PRB (1981)	9787	Self-Interaction Correction to Density-Functional Approx. ...	Perdew and Zunger
12	PRB (1992)	9786	Atoms, Molecules, Solids, and Surfaces - Applications of the ...	Perdew, Chevary, ...
13	PRB (1986)	9313	Density-Functional Approx. for the Correlation-Energy ...	Perdew
14	PR (1934)	9271	Note on an Approximation Treatment for Many-Electron Systems	Moller and Plesset
15	PRB (1994)	9100	Projector Augmented-Wave Method	Blochl
16	PRL (1980)	7751	Ground-State of the Electron-Gas by a Stochastic Method	Ceperley and Alder
17	PRL (1987)	7663	Inhibited Spontaneous Emission in Solid-State Physics ...	Yablonovitch
18	PRL (1986)	7589	Atomic Force Microscope	Binnig, Quate, Gerber
19	PRB (1991)	7425	Efficient Pseudopotentials for Plane-Wave Calculations	Troullier and Martins
20	PRB (1993)	6925	Ab initio Molecular Dynamics for Liquid Metals	Kresse and Hafner
21	PR (1961)	6467	Effects of Configuration Interaction on Intensities and Phase Shifts	Fano
22	PR (1957)	6260	Theory of Superconductivity	Bardeen, Cooper, Schrieffer

Figure 1: *Most cited papers in the history of the American Physical Society (1893–2013) — those marked in red are on electronic structure simulations.*

network of electronic structure groups, and also coordinates and funds around 20 workshops every year, often in collaboration with CECAM, including an annual high-profile Psi-k/CECAM research conference, and a general conference every 5 years, with 1'000+ participants. ICTP, the Abdus Salam International Centre of Theoretical Physics, founded in Trieste in 1964, has a target of "providing scientists from developing countries with the continuing education and skills that they need to enjoy long and productive careers." Within this context, it provides a very rich calendar of workshops, schools and colleges, both in Trieste and worldwide, that are aimed at scientists from both the developing and developed world — including, e.g., the "Total energy and forces" conference, held biennially in Trieste, that is the longest running series in electronic structure simulations for materials (it started in 1984). It is also notable to mention the experience of the Thomas Young Centre (http://www.thomasyoungcentre.org), i.e. the London Centre for the Theory and Simulation of Materials, that has been able to catalyze the geographical presence of many researchers in the field in the London area, and has developed a very healthy outreach effort of conferences, schools, and seminars.

These research activities are being complemented ever more often by dissemination efforts in the data that have been produced by the simulations. Examples include again some of the centers described above and more, such as

- the Materials Project (http://www.materialsproject.org),
- the Crystallography Open Database (http://www.crystallography.net),
- the Open Quantum Materials Database (http://oqmd.org),
- the Automatic-FLOW for Materials Discovery (http://www.aflowlib.org),
- the Harvard Clean Energy Project (http://cleanenergy.molecularspace.org),
- the Computational Materials Repository (https://wiki.fysik.dtu.dk/cmr/),
- the Electronic Structure Project (http://gurka.fysik.uu.se/ESP/),
- the Novel Materials Discovery Repository (http://nomad-repository.eu/cms/),
- the ESTEST framework (http://estest.ucdavis.edu).

In order to develop common standards and favor dissemination and interoperability, Nicola Marzari has organized a workshop at CECAM in June 2015 (together with Dane Morgan, Claudia Draxl, and Kristin Persson) on "Future Technologies in Automated Atomistic Simulations", that will be attended by almost all of the players discussed above.
It is worth commenting that, at variance with most databases mentioned above, MARVEL AiiDA materials' informatics infrastructure, is an entire framework able to manage high-throughput calculations, to provide a model for provenance and preservation, not only of data, but of the entire complex workflow of calculations, and to automatically couple the calculations to the storage and dissemination of data (see the description in PP6 and the documentation at http://www.aiida.net), and it shares some capabilities with Pymatgen (the Python materials genomics library for materials analysis, http://pymatgen.org) and the Atomistic Simulation Environment (ASE, https://wiki.fysik.dtu.dk/ase/), with which is fully compatible.
The economic impact of modeling has been recently assessed by Goldbeck consulting, in a 2012 study commissioned by the CECAM Maxwell node [2], and points to

- "the strong contribution of modeling to academic research and to value creation in industry" with which "the macroeconomic impact has been estimated on the basis of data for the contribution of chemistry research to the UK economy. The preliminary figures suggest a value add equivalent to 1% of GDP."
- It is found that molecular modeling forms a small but very important part of workflows that have produced very considerable returns on investment.
- **E-infrastructures such as high-throughput computing, materials informatics systems and high performance computing act as multipliers of impact**.

It is also refreshing to look at some numbers to assess the balance between investment in research and fundamental activities that support quality control and sharing of resources in the field: if one were to limit oneself to papers published using periodic-boundary codes, as is typical for materials applications, the order of magnitude is of 8'000 to 10'000 papers/year — and with a conservative estimate of 1 author/paper/year (or, equivalently, 8 authors producing 8 papers per year), the research costs, only for salaries, are of the order of between one hundred million and one billion euros per year (considering the salary and research costs for one person to be between 10'000 and 100'000 euros/year). Against this backdrop, it is worrisome the small effort that has been put into verification and validation of these electronic structure codes, and we will try to provide more systematic testing and standards. As part of this effort, we are developing a comprehensive test suite for many pseudopotential datasets, for which a first version has been completed, based on the accuracy of the equations of states and the vibrational properties of elemental solids and selected binary compounds (SSSP, Standard Solid State Pseudopotentials). Further refinements will cover binary compounds like oxides and nitrides spanning a wide variety of oxidation states for *d* and *f* elements, and the inclusion of band structures (especially conduction bands) and logarithmic derivatives to exclude ghost states.

Knowledge transfer towards the academic sector

The activities of all the group leaders in MARVEL have already a very robust component of knowledge transfer, both with the development or co-development of key open-source or licensed codes such as CP2K, CPMD, QUANTUM-ESPRESSO, ABINIT, BIGDFT, ALPS, TRIQS, WANNIER90, Piglet, and with a sustained activity of training workshops and conferences. In addition, and thanks to the collaboration with the Swiss Platform for Advanced Scientific Computing (PASC), key algorithmic, methodological, and computational improvements are taking place, especially within CP2K, QUANTUM-ESPRESSO, and BIGDFT/ABINIT — from optimal performance of sparse matrix libraries, to GPU optimization, to development of new multi-scale code-agnostic efforts.

Areas of intervention

There is still potential to increase the knowledge transfer between the different groups, and, e.g., the annual junior retreat (see the strategy for education and training, Annexe 2) will act as a catalyst for this. The larger community that MARVEL is creating makes it easier to organize regular tutorials for the key codes discussed above. A longer-term goal is to empower also the experimental community to use the current state-of-the-art codes

in the calculation of first-principles spectroscopic data, for which the experimental platform project PP7 will create the key synergies.
The AiiDA platform will empower the sharing of materials data — structures, workflows to calculate complex properties, and algorithms for machine learning and data mining.
We will keep fostering our international collaborations, in particular with the research groups participating in the Materials Genome Initiative, to insure data compatibility and synergies; natural collaboration arise with the Materials' Project, the Open Quantum Materials Database, and the SUNCAT Center. We will also keep working with the Crystallography Open Database (a PhD student supported by Sciex, Andrius Merkys, is visiting MARVEL for 13 months).

Action plan

We have released (Februaray 27, 2015) the first public, open-source version of the AiiDA framework (Fig. 2). In the first phase, we will keep working at increasing its capabilities (as detailed in PP6), and we will complete soon the release of the website for the dissemination of the MARVEL workflows and data, starting with the Standard Solid-State Pseudopotential library (Fig. 3).
We will provide a regular calendar of AiiDA tutorials, continuing on those organized on October 31, 2014 at ETHZ (30 participants) and on February 5, 2015 at the Psi-k/CECAM conference on Materials Design and Discovery (40 participants).
We are organizing the CECAM workshop in Lausanne in June 2015 on "Future Technologies for Automated Atomistic Simulations" and the Psi-k International Symposium and Workshop in Moscow in October 2015 on "Electronic Structure Theory for the Accelerated Design of Structural Materials"; we will keep working on a sustained program of tutorials and activities in this field.

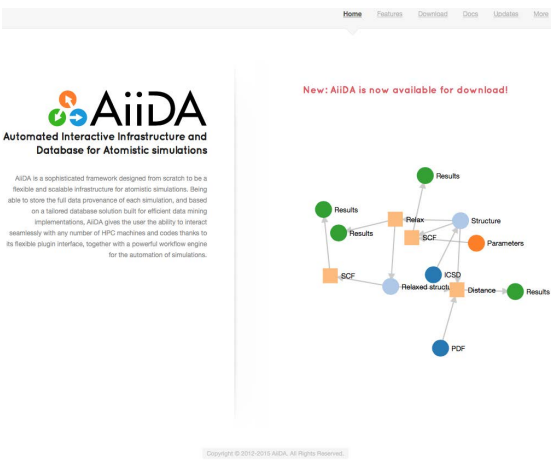


Figure 2: *AiiDA webpage at* `http://www.aiida.net`*, with the release of v.0.4.0 (February 2015).*

Knowledge transfer towards the private and public sector

Areas of intervention

Our priority in this first phase is to interact closely with five main stakeholders:

- supercomputing centers in Switzerland (CSCS) and elsewhere (CINECA Italy, Jülich Germany, Oak Ridge National Laboratory, Lawrence Berkeley Laboratory),
- major experimental facilities, with priority to the Paul Scherrer Institute, and the Empa materials laboratory,
- independent software vendors,
- small and medium enterprises, and large companies,
- policy makers.

Action plan

- We have prepared and submitted a Horizon 2020 proposal for a "e-infrastructure — Centre of Excellence for Computing Events" on *MAterials design at the eXascale (MAX)*, federating the activities of the supercomputing centers CINECA (Italy), CSCS (Switzerland), BSC (Spain), Jülich (Germany) and KTH (Sweden), with letters of support from the Department of Energy supercomputing centers in Oak Ridge and Lawrence Berkeley Laboratory (NERSC).
- We have engaged the two national laboratory PSI and Empa at the highest levels, with one-day meetings to present the MARVEL activities, to brainstorm for joint proposals, and to prepare and submit experimental synthesis and characterization efforts, through the vehicle of PP7. Eight new experimental projects will be funded in year 2 and year 3, to be followed by another round of eight in year 3 and year 4. We have secured cash matching funds from the directors of PSI and of Empa, up to 1'500'000 CHF (PSI) and 900'000 CHF (Empa).

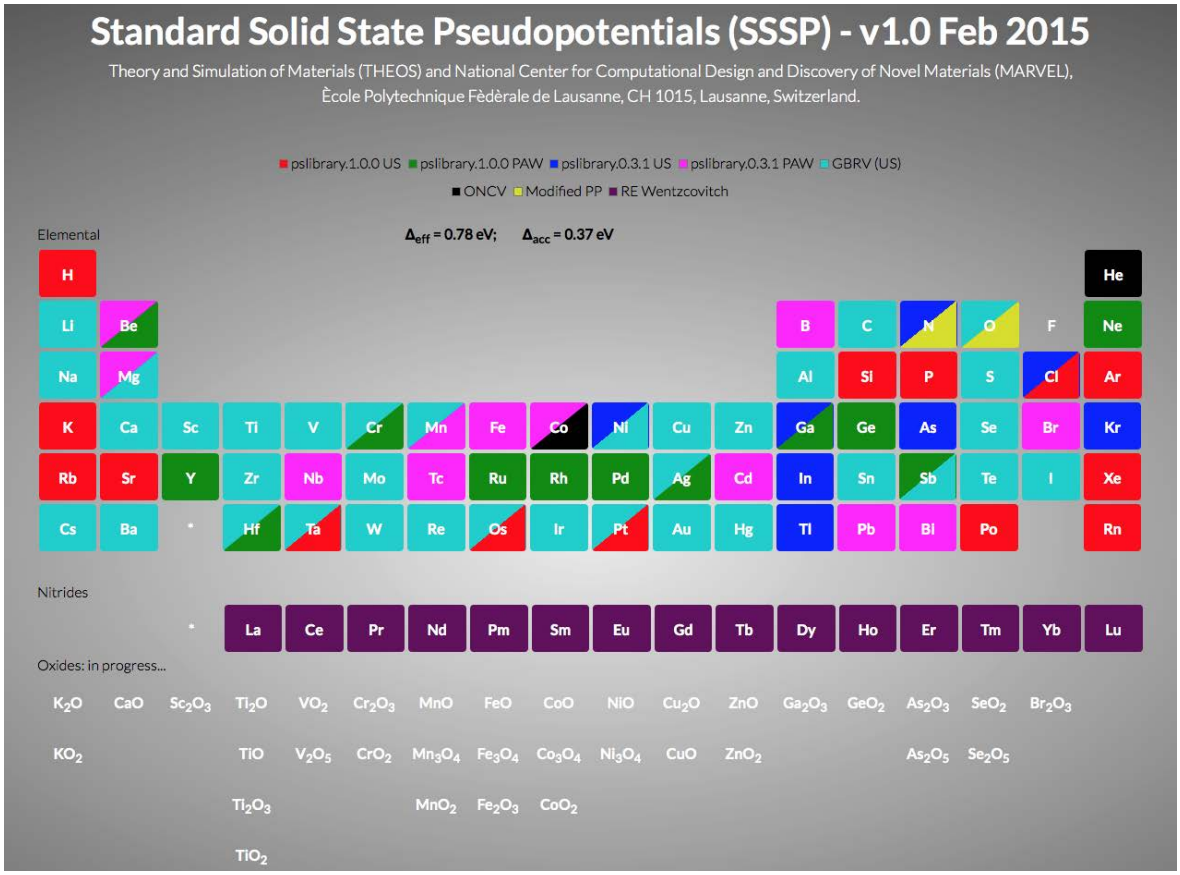


Figure 3: *SSSP (Standard Solid State Pseudopotential) library, obtained by selecting the most accurate or most efficient pseudopotentials chosen against a number of reference calculations (for now, equations of state and phonons).*

- We have established a strategic connection with the Swiss Centre for Competence in Energy Research (SCCER) on Heat and Electricity Storage, through the funded projects in PP7.
- We have been working with Schroedinger (http://www.schroedinger.com) to define XML input/output schemas for the inclusion of the Quantum-ESPRESSO open-source software in their platform. We have been working with QuantumWise (http://www.quantumwise.com) for the release of Virtual NanoLab (VNL) for Quantum-ESPRESSO (http://quantumwise.com/item/3-what-is-atk).
- We plan to have an Industrial Advisory Board and organize an annual Industry day, in collaboration with other EPFL entities (e.g. CCMX or the Energy Center), to provide an active platform of dissemination of MARVEL activities. Nicola Marzari will chair a session together with Carlos Tavares (Chairman, Peugeot Citroën) on Materials Breakthroughs for 2050 at the World Materials Forum (http://www.worldmaterialsforum.com/program.html) in June 2015.
- We have submitted an application for a Latsis symposium at EPFL — a high-profile reach-out activity to highlight the vision of MARVEL. This vision requires the creation of new policies with the cooperation of academic researchers, the private sector, and governmental agencies. High profile exponents of these sectors will be invited to debate about these issues in the symposium, organized around the themes of energy and environment, future of computing, science policy, health and biomedical materials, electronics and new materials, industrial applications, complemented by two high-profile public events (a Nobel lecture, and the IgNobel prizes). The primary targeted audience of this event will be industry and policy makers, but we will take at advantage of this symposium to involve the media and bring these ideas to the public at large.

Technology transfer towards the private sector

Areas of intervention

A long-term engagement with the industry is one of the key strategic goals of the NCCR. At inception, the NCCR included the laboratories of IBM Rüschlikon as a partner, and Robert Bosch on the Scientific Advisory Board. In addition, this group of investigators has a track record of collaborating with industry, that include partners in Switzerland such as Nestlé, Roche, Novartis, ABB, ROLIC, Oerlikon Balzers, Debiopharma, Firmenich, Gaba, and Casale Group, and worldwide inlcuding Intel, Dupont, BASF, Hoechst, Rolls Royce, HP, Semiconductor Research Corporation, ENI, Portland Cement Association, DE Shaw Research, NEC, Ford, Mitsubishi, Schott, Unilever, AkzoNobel.

Action plan

- To encourage strongly the MARVEL community to develop contacts with industrial partners, and proactively initiate these. To identify also the collaborations that could be strengthened, including CTI projects. With this goal, we are hiring Pascale Van Landuyt (PhD, Materials Science), from Alliance, the industrial liaison program of EPFL, to be dedicated to this task.
- To make MARVEL known in the industrial world, with the obvious focus of materials development with specific properties (e.g. started several contact and collaborations with the Swiss high-precision and watchmaking industry). Close interactions with CCMX (http://www.ccmx.ch/home/), the Competence Centre for Materials Science and Technology, based in the Institute of Materials at EPFL, are also planned.
- To have close contacts with the technology transfer offices of EPFL and of the other institutions participating to MARVEL, to proactively establish and reinforce links between researchers and relevant industries.
- To inform and propose to the MARVEL students the courses, formations, and grants available in the technology transfer offices of the different institutions participating to MARVEL.
- A special entry for industrial partners is planned on the MARVEL website (more details in the communication strategy, Annexe 4).
- To educate PIs on intellectual property rights.
- To be present with MARVEL booths at conferences or fairs (e.g. starting with PASC2015 at ETHZ).

Resource allocation and organization

Within the management team, the KTT actions will be carried out mainly by the program and communication manager (currently being recruited). Also, from spring 2015, Pascale Van Landuyt from Alliance, the industrial liaison program of EPFL, will be hired at 20% to proactively work on technology transfer. A close contact with the Technology Transfer Office (TTO) of EPFL and at other participating institutions will be assured through her.
At the scientific level, Nicola Marzari will be responsible for the overall view of KTT actions and recommendations. Nicola Spaldin will lead the effort for the closer interaction within MARVEL of computational and experimental groups.
A Industrial Advisory Board will be created with representatives from the industries participating in MARVEL and other key Swiss and European players.
For the four years of the first phase a budget of 90'000 CHF, outside salaries, is planned for the implementation of KTT actions.

Other references

[1] R. Van Noorden, B. Maher, and R. Nuzzo, *The top 100 papers*, Nature **514**, 550 (2014).

[2] G. Goldbeck, *The economic impact of molecular modelling*, Report, Psi-k network, http://www.psi-k.org (2012).

Annexe 2 Strategy for Education and Training

Computational design and discovery of novel materials is a discipline straddling the frontier between many diverse fields, including physics, materials, chemistry, and computer science, and with relevance to disciplines as diverse as earth sciences, mechanical engineering, or biology — NCCR MARVEL draws students from programs in all these fields. Through the University curriculum, computational science is present at all levels (Bachelor, Master, PhD). MARVEL has the opportunity to be present and inform of coherent offerings in these programs, as well as initiate, co-organize, or sponsor doctoral schools, summer schools, tutorials, and workshops, and also develop and provide key online educational material. Last, with computers embedded in current lifestyles, computational science can also be an attractive way to increase the interest of young people to science, technology, engineering, and mathematics (STEM).

Assessment of the current state

Junior researchers

Within the faculties of physics, materials, chemistry, biology, and computer science, in the field of MARVEL, the following existing courses at the Master level in various institutions participating to MARVEL are particularly relevant for many of the junior researchers.

EPFL

- *Quantum simulations of materials: Properties and spectroscopies*, Nicola Marzari
  This course presents the theory and application of quantum simulations to model, understand, and predict the properties of real materials, including electronic structure and first-principles approaches, temperature and thermodynamic averages, or how to obtain materials' properties from simulations.

- *Computer simulation of physical systems I*, Alfredo Pasquarello
  The main topics covered by this course are ordinary differential equations, classical molecular dynamics, random variables, random walks, and Monte Carlo integration.

- *Computer simulation of physical systems II*, Alfredo Pasquarello
  The course treats the model of diffusion limited aggregation, density functional theory and its applications, and the solution of the Schrödinger equation by variational and diffusion Monte Carlo.

- *Computational physics III*, Oleg Yazyev
  This course teaches the students practical skills needed for solving modern physics problems by means of computation. A number of examples illustrate the utility of numerical computations in various domains of physics. This course deals with Fourier series and transforms, linear systems and matrix manipulation and eigenvalues problems.

- *Physical and computational organic chemistry*, Clémence Corminboeuf
  This course introduces computational organic electronic structure methods as well as physical organic concepts to illustrate the stability and reactivity of organic molecules and rationalise reaction mechanisms.

- *Applied molecular quantum chemistry*, François Rotzinger
  This course is an introduction of the bases allowing the computation of chemical reactions and spectroscopic properties of molecules or ions in the gas phase and in solution using quantum chemical methods.

- *Molecular quantum dynamics*, Jiri Vanicek
  The course covers several exact, approxi-

mate, and numerical methods to solve the time-dependent molecular Schrödinger equation, and applications including calculations of molecular electronic spectra. More advanced topics include introduction to the semiclassical methods and Feynman path integral.

- *Introduction to electronic structure methods*, Ursula Röthlisberger
  This course gives a repetition of the basic concepts of quantum mechanics and main numerical algorithms used for practical implementations, and basic principles of electronic structure methods: Hartree-Fock, many body perturbation theory, configuration interaction, coupled-cluster theory, density functional theory.

- *Molecular dynamics & Monte-Carlo simulations*, Ursula Röthlisberger
  This course is an introduction to molecular dynamics and Monte-Carlo simulation methods.

- *Computational methods in molecular quantum mechanics*, Ursula Röthlisberger and Ivano Tavernelli
  This lecture teaches the basic notions of density functional theory and its time-dependent version (TDDFT) in the context of adiabatic and nonadiabatic quantum dynamics.

- *Big Data*, Christoph Koch
  This course is intended for students who want to understand modern large-scale data analysis systems and database systems. It covers a wide range of topics and technologies, and will prepare students to be able to build such systems as well as read and understand recent research publications.

- *Mathematics of data: from theory to computation*, Volkan Cevher
  This course reviews recent advances in convex optimization and statistical analysis in the wake of Big Data. It provides an overview of the emerging convex formulations and their guarantees, describes scalable solution techniques, and illustrates the role of parallel and distributed computation.

- *Statistical methods in atomistic computer simulations*, Michele Ceriotti
  This course gives an overview of simulation techniques that are useful for the computational modeling of materials and molecules at the atomistic level. The students will learn about basic and advanced methods to evaluate thermodynamic averages by molecular dynamics, including accelerated sampling for the study of rare events, and non-linear dimensionality reduction to study structurally-complex systems.

ETHZ

- *Molecular and materials modeling*, Joost van de Vondele and Daniele Passerone
  This course introduces the basic techniques to interpret experiments with contemporary atomistic simulation. These techniques include force fields or density functional theory (DFT) based molecular dynamics and Monte Carlo. Structural and electronic properties, thermodynamic and kinetic quantities, and various spectroscopies will be simulated for nanoscale systems.

- *Introduction to computational physics*, Hans Herrmann
  This course offers an introduction to computer simulation methods for physics problems and their implementation on PCs and super computers: classical equations of motion, partial differential equations (wave equation, diffusion equation, Maxwell's equation), Monte Carlo simulations, percolation, phase transitions

- *Programming techniques for scientific simulations I*, Matthias Troyer
  This lecture provides an overview of programming techniques for scientific simulations. The focus is on advances C++ programming techniques and scientific software libraries. Based on an overview over the hardware components of PCs and supercomputer, optimization methods for scientific simulation codes are explained.

- *Computational quantum physics*, Matthias Troyer
  This course provides an introduction to simulation methods for quantum systems, starting with the one-body problem and finishing with quantum field theory, with special emphasis on quantum many-body systems. Both approximate methods (Hartree-Fock, density functional theory) and exact methods (exact diagonalization, quantum Monte Carlo) are covered.

UniZH

- *Condensed matter electronic structure theory*, Jürg Hutter and Marcella Mauri
  This lecture presents computational methods and theoretical approaches commonly used to determine many properties of materials. The presented techniques are based on the fundamental equations for the electrons and can provide insight into the physics and chemistry of real systems and observed phenomena. It focuses on density functional theory, which is the most widely used approach in the field of condensed matter electronic structure calculations.

UniBas

- *Computational design of drugs and materials*, Anatole von Lilienfeld
  How to computationally design new molecules and materials is the focus of this course. First, it discusses the underlying fundamentals that relate structure to function. Functions covered include biological, chemical and physical properties. Thereafter, several optimization strategies such as high-throughput screening, stochastic approaches (genetic algorithms), and gradient based methods are explained. Finally, successful application examples from the literature are reviewed.

USI

- *Molecular dynamics*, Matteo Salvalaglio
  This course is dedicated to the introduction of the basic principles of molecular dynamics simulations and their application in a variety of scientific areas ranging from computational biology to material science. The course will be focused on methods and applications of classical molecular dynamics. An introduction to the physical foundations of the method, as well as an overview of the algorithms typically used in molecular dynamics, will be given. Applications of molecular dynamics to problems of interest in biomolecular chemistry, material science and engineering will be presented in order to highlight challenges and opportunities offered by molecular dynamics in state-of-the-art research. Small projects and hands-on examples will be carried out during the course using open source software packages.

In addition to these classes, relevant courses are held regularly at:

- the Swiss National Supercomputing Center (CSCS) in Lugano. An updated list is maintained at http://www.cscs.ch/events/index.html;

- the European Center for the Calculation of Molecular and Atomic properties, CECAM, which is based in Lausanne, at EPFL. A list is found at http://www.cecam.org;

- related to the above, the *MolSim-2015* school was held at CECAM, January 5 – 16, 2015, and organized by Berend Smit and others. This school provides training in the field of simulation techniques for the study of many-particle (molecular) systems. The content constitutes a recap of the statistical mechanics relevant for molecular simulation, and will subsequently cover basic and advanced simulation techniques. These including Monte Carlo, molecular dynamics, free energy calculations, rare events, coarse graining. The theory and techniques covered in the lectures will be illustrated by illustrative examples of present day research.

Undergraduates

At the undergraduate level, a large amount of consolidated experience exists in the education of students in computational sciences both at EPFL and ETHZ. Indeed, an interdisciplinary diploma program in Computational Science and Engineering (CSE) was pioneered at ETHZ in 1997; after the Bologna declaration, this was transformed into Bachelor and Master programs that started in 2003 and 2005, respectively. At EPFL several courses for undergraduate students are given on topics related to methods of simulations, and a new Master in Computational Science and Engineering was created in 2009. The university of Basel has a computational sciences curriculum which gives to the students the necessary knowledge in simulation methods to solve scientific problems. The students can specialize in computational physics, chemistry, biology and mathematics.

As seen the offer is very rich. The goal of the NCCR is to coordinate and cleverly use it in order to propose an appropriate formation in computational material science to the students participating to MARVEL and, through this offer, to attract new students.

Areas of intervention

In our NCCR proposal, we proposed to build on this strong background to develop joint courses via tele-teaching, particularly at the Master level, and to develop a set of web-based e-learning materials. In particular, we proposed to establish a common doctoral program in computational materials science between ETHZ and EPFL with shared summer/winter schools.
Besides the academic categories, projects with primary school and high school students are also planned to encourage the young people to choose studies in the field of science, technology, engineering, and mathematics (STEM).

Action plan

Towards a joint doctoral program

The first step is the establishment of tele-teaching links between ETHZ and EPFL and then to the other NCCR sites. We aim at first to broadcast the MARVEL Distinguished Lectures (see communication strategy, Annexe 4) across sites, to develop a familiarity with the distant-education technologies (it should be stressed that as part of the structural commitments of EPFL to MARVEL, a visualization facility has been build in the MX building, and a distant-education enabled classroom will be built in the new ME building where the NCCR is going to move in early spring 2016).
The second step involves offering short, intensive classes that are taught at one site, and that are re-broadcast at other sites. Ideally, these could become mini-MOOCs that are advertised more broadly to the community at large. The communication manager will be in charge of exploring these possibilities together with the Center for Digital Education at EPFL.
In addition to the formal courses, we plan one junior retreat every year, which will be organized by our PhD students and postdocs with guidance from the PIs. It will combine few visionary talks from guest scientists selected by the students, together with a substantial body of student presentations and discussions, and special activities to broadcast the offerings of SNSF (starting from the Doc.Mobility program, http://www.snf.ch/en/funding/careers/doc-mobility/Pages/default.aspxi, but also the PostDoc.Mobility, Ambizione, SNSF Professorship).
The first Junior Retreat is planned for July 2015, over three days. It will be supervised by Prof. Michele Ceriotti (EPFL), but organized by Dr. Hiroshi Shinaoka (Troyer/Werner groups), Dr. Ivano Castelli (Marzari group) and Dr. Sandip De (Ceriotti group). This activity builds on an earlier and very successful junior retreat that took place in 2013, as part of the preparatory MARVEL actions as funded by HP2C and PASC (https://sites.google.com/site/pascjuniors/home).

Education platform

A key long-term deliverable will be the preparation of an education platform collecting a library of video lectures and associated material on advanced topics in electronic structure simulations, to act as a permanent repository given, by the most notable practitioners in the field. Such activity will benefit from synergies with the MARVEL Distinguished Lectures and with the constant stream of high-profile visitors that attend CECAM's activities.

Participation of PhD students and post-docs in workshops

MARVEL sponsors workshops, conferences, and tutorials that focus on topics associated to the MARVEL NCCR by providing a financial contribution to young MARVEL collaborators, such as PhD students and post-docs. In this way, through the contact with world-experts attending such conferences, MARVEL encourages high-level education and training of its members. In this context, the international conference *Nothing is Perfect – The Quantum Mechanics of Defects*, which will take place in the Center Stefano Franscini in Monte Verità, Ascona (TI, Switzerland) from April 26 to 29, 2015, is already benefiting from such support.
Close contacts have been established between MARVEL and the "Simons Collaboration on the Many Electron Problem" (https://www.simonsfoundation.org/mathematics-and-physical-science/many-electron-collaboration/) directed by Andrew Millis. The two centers have complementary goals and common needs and joint efforts are planned, through a memorandum of understanding. A first action is taking place via the annual Simons Many Electron Collaboration Summer School held at the Simons Center for Geometry and Physics at Stony Brook from June 8 to 12, 2015, focusing on electronic structure and its many body extensions. The Simons Collaboration will support all of the local expenses for MARVEL students, and the student travel to and from the school will be payed by MARVEL.

The sponsorship of workshops, conferences, and tutorials has a twofold target. On one hand it contributes to the education of young MARVEL collaborators, on the other hand it contributes to the world-wide visibility of MARVEL. Indeed, through the financial support provided, MARVEL officially appears among the sponsors of the event.

Primary school and high school students

The group leaders participate in many programs already in place at our institutions for facilitating the involvement of school children in our science activities. For example, at IBM and PSI as well as at EPFL and ETHZ, high school students are invited for one-day visits, and EPFL and ETHZ have touring Science buses. At EPFL, several group leaders actively participate in a special day, *Oser tous les métiers*, designed to remove gender barriers and expose primary school aged children to a variety of professions. In addition, professional outreach experts at our institutions organize special events such as *Treffpunkt Science City* at ETH Hönggerberg, the *Nuit de la science* in Geneva. We will continue to participate in these programs, and where necessary the NCCR will provide support for their further expansion. At all levels, girls and young women will be particularly targeted. The installation of a 3D visualization facility at EPFL will be particularly useful for outreach activities towards these groups.

Mentoring training; computational sciences in emerging regions

In collaboration with the International Center for Theoretical Physics (ICTP) in Trieste, we continue our on-going involvement in the ASESMA (African School series on Electronic Structure Methods and Applications) program, which we plan to expand within this NCCR. ASESMA is a series of schools providing an introduction to the theory of electronic structure, with an emphasis on the computational methods for practical calculations. A forerunner to this series was a two-week school on computational materials science organized by Nicola Marzari in Addis Ababa (Ethiopia) in 2008, with sponsorship from ICTP and the NSF International Center for Materials Research, then headed by Nicola Spaldin. The 2010 school was held at the African Institute for Mathematical Sciences, Cape Town, South Africa and the 2012 school was held at Chepkoilel University College, Eldoret, Kenya. Both Nicola Spaldin and Nicola Marzari are members of the ASESMA Advisory Panel, and they and/or their students have taught at the schools. An important component of ASESMA is the inclusion of mentors: advanced students and postdocs who assist with tutorials and provide ongoing consultation support between the workshops. The mentors at the first school hailed mainly from Spaldin's group, and they proved to be an essential ingredient as they forged close connections with the African participants. At the second school, many of the mentors were African students from the first school. Within the NCCR we will train and support our local PhD students to travel as mentors to future ASESMA schools. In addition to the obvious benefit to the students that they teach, this is an exciting training opportunity for own students in teaching and mentoring, as well as exposure to different cultures. In January 2015, Spaldin-group PhD student Florian Thöle was attending an ASESMA School in Johannesburg as a mentor, and Marzari-group senior scientist, Matteo Cococcioni was attending as a lecturer. A new component of the program that we propose for the NCCR is to coordinate with the ICTP to provide a visitors program for African students to our laboratories. For a recent discussion of the ASESMA program, including the motivation for establishing such a program in Africa, by ASESMA web coordinator and former Spaldin-group student Alison Hatt, see [1].

Resource allocation and organization

Within the management team, the education and training actions will be carried out mainly by the program and communication manager (being recruited), while at the scientific level, Michele Ceriotti and Nicola Marzari will be responsible for the overall view of education and training actions and recommendations.

Commitment of the group leaders in teaching short doctoral schools, and providing lectures and organization of workshops and tutorials will be sought.

For the four years of the first phase a budget of 150'000 CHF is planned for the implementation of education and training actions, including a yearly sponsoring of schools and workshops of the order of 10'000 CHF. 30'000 CHF have also been allocated to the budget for the Junior Retreat, typically covering $50-60$ students for 3 days.

Other references

[1] A. Hatt, *Networking at the heart of African workshop on computational materials science*, MRS bulletin **38**, 12 (2013).

Annexe 3 Strategy for Equal Opportunity

Assessment of the current state

As in many other scientific and engineering fields, the number of female students entering MARVEL-related curricula and, in particular, continuing an engineering career in industry and academia is low. The main fields related to MARVEL are physics, chemistry, material sciences and computer science. In physics and computer science (Tables 3.1 and 3.2), the proportion of women is far less than the average of female students in all fields, at EPFL (PhD 30%, and undergraduate 27%), as well as at ETHZ (PhD 31% and undergraduate 30%). At EPFL, 26% of the scientific staff is constituted of women, and this ratio is 25.5% at ETHZ. Significant efforts have been made in recent years to increase the number of female professors at EPFL. The percentage of female professors has risen from (3%) in 2000 to (12%) in 2013. However this ratio is still very low. The percentage of associate and full professors is even less, with 8% for EPFL and 9% for ETHZ. The numbers are very close to each other comparing ETHZ and EPFL and show that they reflect a general situation in these fields, which is also the one found in the other institutions participating in MARVEL. These ratios demonstrate that further efforts are essential in order to increase the number of women at all levels.

There are multiple reasons for the underrepresentation of women in these fields. Gender stereotypes, the lack of confidence of girls in their capacity in mathematics and scientific branches, and absence of role models are the most important obstacles for young girls to choose scientific and engineering fields. Specific actions are necessary in order to overcome the cultural issues constructing barriers against the choice of some fields by girls. This level is very important and needs a special attention because it prepares also the reservoir for the upper levels. In order to make cultural changes in advanced industrial countries, it seems to be necessary to develop targeted programs for young girls in STEM (Science, Technology, Engineering, and Mathematics) fields to encourage them to choose scientific and engineering fields. It is essential to give confidence to girls in their capacities in mathematics and scientific fields and to work with teachers, schools, and parents in order to shatter stereotypes and increase the number of girls choosing scientific and engineering fields.

Unfortunately there is also a systematic loss of women during the academic path, called the

Field	INF	CH	MX	PH	AF
Bachelor and master students	11	42	25	19	27
PhD students	22	35	31	23	30

Table 3.1: *Ratio (in %) of female students at EPFL in 2013 as a function of the field of study. INF = Computer Science, CH = Chemistry and Chemical Engineering, MX = Material Science and Engineering, PH = Physics, and AF = all fields.*

Field	INF	CHB	MA	PH	AF
Bachelor and master students	13.5	44.5	30.5	15.5	30
PhD students	17.5	33.5	28	16.5	31

Table 3.2: *Ratio (in %) of female students at ETHZ in 2013 as a function of the field of study. INF = Computer Science, CHB = Chemistry and Applied Biosciences (curricula: chemistry, chemical engineering, interdisciplinary sciences, pharmaceutical sciences), MA = Material Sciences, PH = Physics, and AF = all fields.*

leaky pipeline. A lack of tools and skills as well as the visibility and adequate networks act as obstacles for developing careers of scientific women. Gender unconscious bias as well as lack of general awareness about the gender issues are also barriers against the access of women to higher positions in academia. Scientific women at the age of postdoctoral phase are also at the age when having children becomes urgent. Absence of flexibility within the academic system as well as the deficiency of structures to help to reconcile family life and work push many women to quit academia. Procedures should be established and measures undertaken at all levels in order to overcome these problems and guarantee equal opportunities and advancement of women in scientific and engineering fields.

Goals

The goal of MARVEL is to develop a program to increase the number of female students in MARVEL-related branches, to increase the general awareness about the gender issues, to recruit female PhD students and researchers, and to offer them an environment and tools to continue their studies and to consider an academic career.
The main goals of the program are:
1) to enable young girls to discover the world of engineering and research by providing them the possibility to have MARVEL-related workshops adapted to their age, meeting female scientists and engineers in these fields as role models, and visit research labs to encourage them to choose scientific fields for their future studies;
2) to recruit female scientists and to offer them an adequate environment to develop their career by

- providing information and documentation about opportunities for women;
- offering support to build careers, providing tools such as courses and workshop for improving communication, leadership and negotiating skills, mentoring and coaching programs, to answer to their questions, to help them identifying their strengths, detecting gaps and drawing up a personal development plan;
- increasing their networks by organizing networking activities; academic and industrial networks;
- increasing their visibility and promoting them to higher positions whenever possible.

The program run in collaboration with the Equal Opportunities Office of EPFL will be composed of common actions to the three NCCRs hosted at EPFL — Synapsy, Robotics and MARVEL — and of specific actions in the field of MARVEL.

Measures and actions

The program of MARVEL for equal opportunities is summarized in the four following tables (Tables 3.3, 3.4, 3.5, and 3.6). In each table, the objective, the measures, possible indicators, and the expected outcomes are given.

Resource allocation and organization

The Equal Opportunities Office of EPFL and the NCCRs Synapsy, Robotics and MARVEL put together their resources in order to be able to conduct a coherent program with concrete actions. During phase I, MARVEL will contribute to this program with 50'000 CHF per year.
Moreover MARVEL will provide more funding, for example for one-shot actions such as invitation of female professors to give seminars, for organization of skill workshops, or for master internships in the MARVEL labs. In case of reallocation of unspent funds in the groups, MARVEL will strive to launch specific funding to promote research opportunities for female members at all levels (PhD students, postdocs, senior scientists, professors). These funds would be in addition to the regular project specific allocation to the group leaders.
Concerning personal allocation, within the management team, the equal opportunity actions and the contact with the Equal Opportunities Office of EPFL will be carried out by the scientific manager.
At the scientific level, Clémence Corminboeuf (EPFL) will be the representative for equal opportunity.

Table 3.3: *Raising the gender awareness.*

Objective	⬡ Raising the awareness concerning the gender issue, gender mainstreaming through the NCCR.
Measures	⬡ A guideline for group leaders for implementation of equal opportunities measures. ⬡ Workshops for increasing the gender awareness. ⬡ Special attention for the recruitment of women at all levels and their presence in committees. ⬡ Exhaustive website.
Indicators	⬡ Numbers.
Outcome	⬡ Behavior changes, better recruitment of female scientists, better gender balance within the NCCR and in key positions.

Table 3.4: *Female scientists.*

Objective	⬡ Encourage academic careers for all categories of female academic staff, enable career advancement in MARVEL-related fields and career building (tools and information), and increase the visibility.
Measures	⬡ Implement events for female scientists and PhD students of MARVEL (conferences, networking, coaching, mentoring, trainings). ⬡ Organize specific workshops to improve communication and leadership skills. ⬡ Collaborate with partners for joint national and international projects for career advancement of women ("Fix the leaky pipeline!", mentoring with *StartingDoc*, *Réseau romand de mentoring pour femmes*, *Mentoring Deutschschweiz*). ⬡ Support scientific women in their participation to conferences as invited speakers. ⬡ Support junior scientific women in their participation to workshops and conferences. ⬡ Prepare nominations of women for awards.
Indicators	⬡ Survey and testimonies, number and type of activities.
Outcome	⬡ Empowerment of female scientists. ⬡ More female scientists continue an academic career in MARVEL-related fields. ⬡ Access of women to leader positions.

Table 3.5: *Young girls, next generation of female scientists.*

Objective	⬡ Promotion of MARVEL-related fields specially among girls, to encourage them to choose scientific and engineering fields.
Measures	⬡ Implement various actions for young girls of 8 to 15 years old. ⬡ Young girls: learning by doing, having fun, refering to female models. ⬡ Teachers and schools: open up a dialogue. ⬡ Parents: discussion and dialogue, workshops.
Indicators	⬡ Testimonies, number of requests and participants.
Outcome	⬡ Behavior changes, less stereotypes and obstacles related to gender. ⬡ Increased knowledge and interest of young girls for MARVEL-related fields; increased knowledge of parents about the place of women in these fields. ⬡ Increased confidence of young girls in their capacities in STEM fields. ⬡ Building a reservoir of female scientists and engineers in Switzerland.

Table 3.6: *Work/life balance.*

Objective	⬡ Contribute to a better reconciliation of family life and professional activity, at all levels and for both men and women.
Measures	⬡ Information to MARVEL members concerning measures already undertaken in the different institutions. ⬡ Implement the SNSF encouragement measures, 120% support grant. ⬡ Meetings with MARVEL members to identify their specific needs. ⬡ Development of new measures as a function of these needs. ⬡ Meeting and conference hours fixed so as to let a reconciliation of family life and work.
Indicators	⬡ Testimonies of women and men; survey.
Outcome	⬡ The family responsibility is not an obstacle for women career development.

Annexe 4 Communication Strategy

General context and goals

One of the general purposes of the NCCR MARVEL is to promote and install a new paradigm in the discovery and deployment of novel advanced materials. The core of this paradigm consists in developing an infrastructure in which advanced computational modeling, data archiving and data mining, and physical experimentation are strongly intertwined and cross-linked. The success of this paradigm rests on the creation of an open platform, in which access to advanced computational modeling software and data transparency are guaranteed to all stakeholders playing an active role from material discovery to product manufacturing. The ultimate goal of the NCCR is to introduce a cultural evolution in which computational modeling for predicting materials properties are routinely relied upon when new materials and products are designed. To achieve such an evolution, communication must play a key role in our strategy.

More specifically, the NCCR MARVEL has multiple objectives which require an effective communication strategy to be achieved.

The primary objective consists in restructuring academic research in the area of materials science in order to fully take advantage of the continuously increasing computational resources as they become available. Such a paradigm shift is expected to accelerate the discovery of materials with novel or improved functionalities, reducing in this way the overall time required to bring the benefits of such discoveries to the society. The first transformation within the research academic community required to achieve this goal consists in making internationally recognized research leaders in computational sciences collaborate and coordinate their efforts. This collaboration is not only expected to overcome specific research issues but also to lead to the construction of a background network, in which research instruments such as computer codes as well as research results (in the form of digital data) are made available and shared within the computational community. Developed methods and achieved results should be capitalized and made available in a friendly way to possible users. This requires an infrastructure with common standards which can only be built up through an effective communication among the computational members of the NCCR. The second reorganization within the academic research community involves the interaction between the computational and experimental communities. To strengthen the links between the two communities, the NCCR includes experimental researchers among its members. Once the most promising materials have been identified through a computational screening procedure, their functionalities need to be experimentally verified. This verification step is critical towards the overcoming of all kind of practical issues which may prevent materials identified *in silico* to be useful in practice. The NCCR thus strongly encourages exchange of expertise and knowledge between the two communities. Ideally, such exchange of information could be turned into selection criteria and it is thus critical to allow this exchange process to occur as early as possible in the search procedure.

A secondary more long-term objective concerns the interactions between academic research and the private sector. In the spirit of accelerating the development of a final product based on the discovery of novel materials, it is necessary that the private sector be first of all informed about the above developments taking place within the NCCR. However, a full exploitation of the novel ideas proposed within the present NCCR requires not only awareness of the private sector, but also transfer of expertise and knowledge in order to optimally access the required data or to generate new data if necessary. The communication

Strengths	Weaknesses
○ Well defined goal ○ Cutting edge research ○ Project leaders of high international reputation ○ Project leaders know each other well ○ Complementary and diversified expertise ○ Experimental platform included in the program ○ Top-quality hardware expertise	○ Geographically distributed in Switzerland ○ Collaborative projects are not common practice ○ The groups involved have their own favorite computer codes ○ Most groups are active on various research fronts other than just the NCCR
Opportunities	**Threats**
○ Facilitate the interaction between simulation groups ○ Access to computer codes developed within the simulation community ○ Involve the experimental community in an organic way ○ Development of common standards for computer simulations, data sharing and exchange, and data accumulation ○ Training new generations in taking advantage of the computational infrastructure developed in the NCCR ○ Establish contact between the private sector and computational science	○ Disinvestment from the private sector in absence of coherent policies ○ Ambiguous perception of big data by the public at large and the media ○ Lack of collaboration with similar existing or future efforts running in the United States or in Europe.

Table 4.1: *SWOT analysis.*

strategy of the NCRR in this regard is twofold. On the one hand, the education and training of young PhD students and postdocs within the NCCR will contribute in exporting the required knowledge and expertise to all sectors of the society. On the other hand, the NCCR would like to accelerate these changes by encouraging the creation of direct links with the private sector. It is envisaged that existing financial instruments could be used to establish such links. The communication strategy should thus be designed to actively involve not only agents in the private sector but also governmental agencies which have the objective of promoting the transfer of knowledge from the academic sector to the private sector (see also the strategy for KTT, Annexe 1).

As a final objective, the public at large needs to be informed about the activities of the NCCR and about its associated vision. The awareness of the public is fundamental in gaining long-term support for the introduction of new policies. Hence, the communication strategy will need to be designed accordingly.

In order to evaluate the strengths, weaknesses, opportunities and threats of the communication involved in a NCCR such as MARVEL, a SWOT analysis is presented in Table 4.1.

Targeted audiences and related core messages

Internal audiences

MARVEL Community

- We have the opportunity to discover new materials and to affect society.
- The NCCR MARVEL is going to occupy a central place in Switzerland and in the world as far as the area of computational materials science is concerned.

PhD students, Postdocs

- Cutting edge research activity will be performed within MARVEL.
- MARVEL provides an ideal context for learning computational science techniques.
- MARVEL is going to be visible and recognized throughout the world.
- MARVEL naturally connects to the private sector offering new job opportunities.

Group leaders/scientists

- *The whole is greater than the sum of its parts* – our individual research is going to be more effective in the presence of a shared computational infrastructure of research instruments.
- MARVEL will enhance the impact of your research on experimental research in particular and on society in general.
- MARVEL will strengthen the importance of computational materials science in academics and in society at large.

Experimental partners

- The role of experimental expertise is critical.
- Experimental research can be improved through indications from simulations.

Executive and Scientific Committees

- Work in a collegial manner guarantees the success of the NCCR.
- You lead an important paradigm shift in which computational materials science plays a key role.
- MARVEL influences the way computational materials science impacts on research, the private sector, and society at large.

Scientific Advisory Board

- You advise the Executive and Scientific Committees to develop the potential available in MARVEL.
- You provide a channel by which the NCCR impacts on computational communities external to MARVEL.
- You provide scientific guidance to young students and postdocs involved in MARVEL.
- You advise MARVEL collaborators when searching for future positions.

External audiences

Computational community in materials science

- MARVEL is a new way of performing science based on computational resources.
- MARVEL is supporting cutting edge science carried out by top scientists.
- MARVEL will deliver results which are directly useful for society.
- MARVEL builds up a scientific infrastructure required for the discovery of new functional materials, with the dissemination of open-source materials simulation codes.

Experimental community in materials science

- MARVEL is leading a new effort to reorganize interactions between simulation and experiment in view of accelerating the discovery of new materials.
- Experimental research can be made more effective by taking advantage of the infrastructure developed within MARVEL.

Private sector and industry

- MARVEL builds a new scientific infrastructure from which the private sector can benefit.
- MARVEL facilitates contacts between the private sector and top scientists active in materials science.

- MARVEL accelerates the discovery and development of new materials useful for society.
- MARVEL builds up a scientific infrastructure which can be used to identify new functional materials.

Governmental and funding agencies

- MARVEL is proposing a new way of bringing the benefits of computer simulation to experimental research and to the private sector.
- MARVEL is initiating a long-term process which needs to be supported not only through funding, but also through targeted policies.

Public at large

- MARVEL is accelerating the discovery of new functional materials and reducing the the time to bring the related benefits to society.
- MARVEL is reorganizing the way simulation and experiment interact in the search of new materials useful for society.
- MARVEL offers opportunities to meet specialists in materials science simulation.
- Big data is part of MARVEL (e.g. big data based material discovery) and the ambiguous perception of this thematics has to be taken into account when communicating with the public at large.

Girls and young women

- Materials science and materials science simulation are not reserved to men.

Media

- Scientific journalists are an ideal link between scientists and the public at large and should be specifically targeted.
- Big data is part of MARVEL (e.g. big data based material discovery) and the ambiguous perception of this thematics has to be taken into account when communicating with the media.

PhD students

- MARVEL provides a top-class and diversified education in the area of materials science simulation.
- MARVEL is nationally and internationally well connected.
- MARVEL is ideal for starting an academic career in computational materials science.
- MARVEL delivers expertise which will be useful in the private sector of the future.

Communication measures

Internal measures

The main goal of internal communication is to create the best possible conditions for a successful achievement of the tasks set up by the NCCR through the fostering of synergies and the definition of standards for data accumulation and exchange. Apart from the transmission of general information about the evolution of the project, the role of internal communication will be particularly emphasized. Indeed, the expertise available among the group leaders is particularly rich and diverse and will be invoked upon in the advancement of the individual projects. The following measures are envisaged.

Website

The MARVEL website will be a dynamic site on which all the news and developments occurring within the NCCR will be updated in a continuous manner. The website will contain both organizational and research issues. In particular, the website will also contain sections devoted to the various projects including lay summaries and progress reports. The latter ones will allow all MARVEL members to remain updated concerning the overall status of the NCCR. A professional company has already been contacted to realize such a website and its contents will be made accessible as soon as possible.

Newsletters

A newsletter will be sent to all MARVEL members by email on bi-monthly basis. The newsletter will be used to ensure that all MARVEL members are aware of the latest developments. The e-newsletters will make reference to the MARVEL webpage as much as possible. This newsletter will also be available to all

interested person, also outside MARVEL and registration will be made possible on the website.

Contact with experimental partners

MARVEL has associated from the very beginning experimental partners to researchers working on advanced numerical algorithms and data sharing systems. The experimental expertise associated with MARVEL will provide in-house advice to optimize the strategy towards the physical experimentation of materials. The facilities of our experimental partners will be involved directly once a promising new material is identified. In order to ensure that the experimental partners within MARVEL be well connected with the simulation activity and to identify the most suitable experimental activities deserving financial support, "Experimental verification workshops" involving simulation and experimental activities are organized within MARVEL on a regular basis. One of such workshops has already taken place at the Paul Scherrer Institute on October 17, 2014.

Annual plenary meetings

Such meetings generally take place on an annual basis. On this occasion all MARVEL members are convened. The plenary meetings, also called "Review & Retreat", consist of several parts. The first part features various presentations which represent the ongoing activity in the various projects of MARVEL. In the second part, the project leaders and their main collaborators sit around tables in small groups in order to prepare strategic action plans. In the third part, the action plans are presented to the general audience and are made open to discussion and adjustment. The plenary meeting will not only foster even closer collaborations between the group leaders but also ensure that the various projects converge towards the overall purpose of the NCCR. It is also the occasion to meet and discuss with the Scientific Advisory Board. In this initial phase, two plenary meetings already took place. The MARVEL kick-off meeting, took place at EPFL on May 15 – 16, 2014, only a few days after the official starting date of the NCCR. This meeting was followed by the first "Review & Retreat" on September 8 – 10, 2014, still at EPFL.

Scientific Committee meetings

These meetings will take place at least once a year, during the annual plenary meeting, to discuss the management of the research projects and take important decisions such as major budget reallocations or establishment of new projects. They will also discuss and propose plans and layouts for the general scientific reporting.

Executive Committee meetings

Executive Committee meetings are organized on a continuous basis per email/skype or direct meetings as often as necessary to take care of day-to-day decisions and management. Specific group leaders or members of the management team are invited to attend those meetings when necessary. All potential issues affecting the NCCR are handled by this committee, in particular the preparation of the site visit.

Management team meetings

These meeting are organized by the Director whenever necessary and have the purpose of handling all organizational issues of the NCCR.

Scientific Advisory Board meetings

These meetings will take place once a year, during the annual plenary meeting. The Scientific Advisory Board has access to the advancement of the NCCR research, in particular to the annual progress reports, to be able to suggest research lines and to act as an independent reviewer of the activities of MARVEL.

External measures

External communication is envisaged for disseminating the scope and the results achieved by the NCCR, but also for promoting the new paradigm supported by the NCCR. Indeed, it is envisaged that, in order to accelerate the introduction of novel disocoveries into the market, all stake-holders, such as industrial partners, private investors, and governmental agencies, should be properly informed about the novel possibilities enabled by the NCCR MARVEL.

Scientific communications

The scientific results will be conveyed to the general scientific community through publications in scientific journals following the usual practice. The scientific activity of the NCCR will further be publicized by pursuing the publication of dedicated special issues and through oral presentations at international conferences.

Distinguished Lectures

In order to enhance the visibility of the NCCR but also to bring high-profile lecturers in contact with MARVEL members, the NCCR sponsors a series of "NCCR MARVEL Distinguished Lecturers". It is expected to have a couple of such lectures per year. Two of such lectures have already been organized. On October 9, 2014, MARVEL has hosted Prof. Alex Zunger from the University of Colorado in Boulder, and Prof. Gustavo Scuseria from Rice University (Houston, Texas) has been invited on January 14, 2015. These lectures can take place either at EPFL or in one of the other participating institutions. The could also be telecasted or recorded, so that they become available in all the MARVEL institutions.

Website

The website will not only be used as a practical instrument for internal communication but will also have general pages by which MARVEL can be showcased to the rest of the world. In particular, it will propose entries to various publics, from the scientific community to the public at large, going through other interested publics, such as media and industries, with presentations of the MARVEL project and its results in a way adapted to the respective targeted audience.

Brochures

Brochures containing a description of the NCCR in lay terms will be prepared to advertise the activities of MARVEL. These brochures will be made available in all special events sponsored by the NCCR, and to the public at large in events where MARVEL will be present. The group leaders will distribute these brochures broadly when they give invited talks at workshops and conferences.

Media/press-releases

MARVEL will work hand in hand with EPFL and other institutions' public relation offices to establish and maintain contacts with specific medias. In particular they will prepare together press releases. This instrument will be used on two kinds of occasions, either to advertise important research achievements (e.g. publications in high-profile scientific journals) or to publicize widely important dedicated events organized by MARVEL (e.g. the reach-out symposium, as developed in the strategy for knowledge and technology transfer).
Members of the management team will participate at meetings or events dedicated to communication in sciences, such as the ScienceComm congress (http://www.sciencecomm.ch) organized every years by the *Science et Cité* foundation and which brings together Swiss science communication experts and offers them a platform for exchanging views and information.

Actions dedicated to the public at large

Events The institutions involved in MARVEL organise or participate at a lot of internal or external events dedicated to the presentation of science to the public at large (e.g. *Treffpunkt Science City* at ETH Hönggerberg, the *Nuit de la science* in Geneva). MARVEL will encourage its members to take part in them. Moreover hands-on experiments and demonstrators should be developed for such events. As MARVEL specificity is the *computational* design and discovery of novel materials, the ideas is to find a nice way to show to the public how computer based research brings its contribution to material research. A collaboration with the scientific mediators of the Science Outreach Department at EPFL is planned.

Works of art, exhibitions and movies A way to bring science to the public at large is to go through art. Art is indeed a way to render complex notions such as materials science simulations and quantum mechanics more accessible by using more visual and tangible supports. Moreover it is a way to bring people normally not interested in science, but interested in art, to meet science and scientists.
In this direction, Nicola Spaldin, at ETHZ, initiated preliminary contacts with Julie Birenbaum, an artist whose work has included conveying the scientific process through light and images taken in a scientific environment, such as blackboards with quantum mechanics formulas (see for example http://www.juliehbirenbaum.com/#!galerie-off-the-wall/), as stated in the communication chapter of the progress report.
Another idea to use art to bring science to the public at large is to organize visualisation contests for doctoral and postdoctoral students to propose innovative ways of visually presenting their data. The resulting works can then be presented to the public through an exhibition.
Due to its name, the NCCR MARVEL often initiates contacts with various publics

through some well known superheroes. Superheroes are also in the spotlight of the "Comictanium^{TM}: The Super Materials of the Superheroes" traveling exhibition (http://www.tms.org/comictanium/) developed by The Minerals, Metals and Materials Society (TMS), the TMS Foundation and the ToonSeum in Pittsburgh, PA, USA. It is an educational exhibition which presents real world materials by the use they might have for well-known comic book heroes, like Iron Man, Captain America, Spider-Man, Batman and others. The goals of this exhibition is to encourage the young people to choose studies in the field of science, technology, engineering, and mathematics (STEM). MARVEL would like to try to bring this exhibition in Switzerland.

The EPFL Energy Center is one of the co-organiser of the *Festival International du Film sur l'Energie de Lausanne* (FIFEL) through its participation to the FIFEL foundation. This biennial festival proposes a selection of documentary films on the problematics of energy and environment. As the new director of the Energy Center, Berend Smit, is also a MARVEL group leader, and as materials for energy harvesting (such as photovoltaic) and storage, as well as CO_2 capture are part of the MARVEL priorities, synergies could be developed between MARVEL and the FIFEL festival. Contact could also be taken with the renowned documentary film festival *Vision du Réel* in Nyon, putting together two local initiatives.

Resource allocation and organization

The communication strategy is directly under the supervision of the director and the executive committee, with a special attention from Alfredo Pasquarello, the deputy director. The NCCR is currently hiring a program manager who will take responsibility for the communication work together with the scientific manager. Links have already been established with the public relation office of EPFL, and in the next months they will be developed with the main purpose of setting up a joint communication strategy as far as interactions with the media are concerned.
The budget allocation for implementing the communication strategy in the first year is about 40'000 CHF, with an important part dedicated to the website development. For the following years, the tentative budget allocation will be about 30'000 CHF per year.
We have planned to contact Prof. Paola Rodari, in charge of the master program in scientific communication of the SISSA International School of Advanced Studies in Trieste, to assess and evaluate our communication strategies and activities, and to identify the evaluation criteria that we should monitor.

Cover picture

Polyhedral structure of the cages in low-density allotropes of silicon (from Stefan Goedecker, UniBas, Horizontal Project 4).

Acknowledgements

We would like to thank MaNEP, Material with Novel Electronic properties, an NCCR of the first series, and in particular Christophe Berthod at UniGE, for the MaNEP LaTeX style that served as a base to the present MARVEL style.