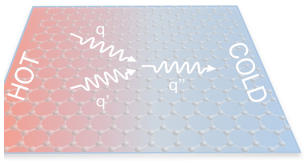
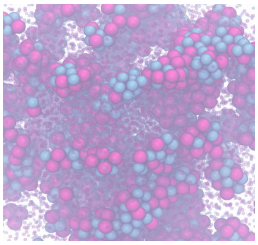
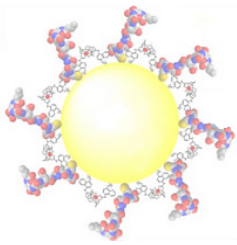
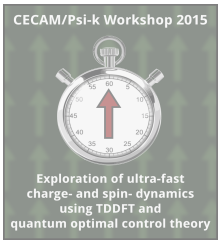
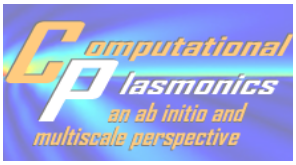
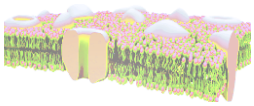


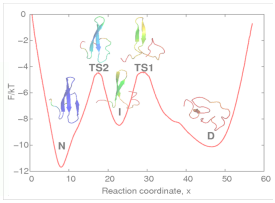



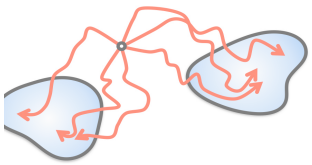


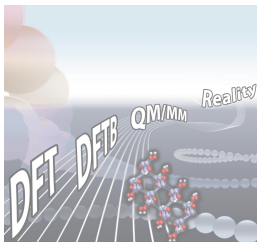



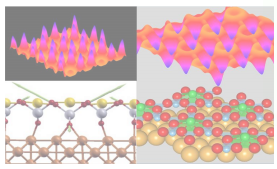



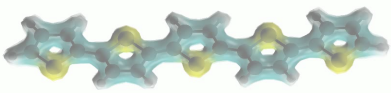



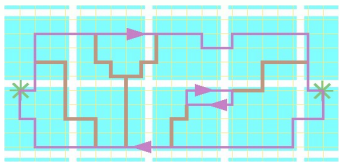


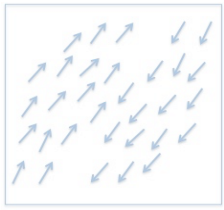



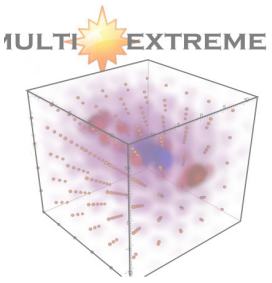


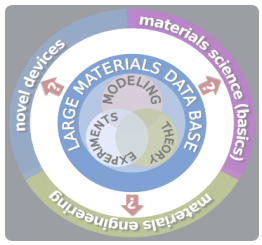




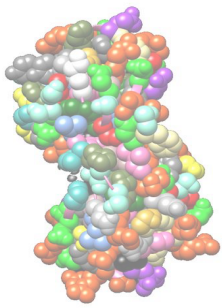


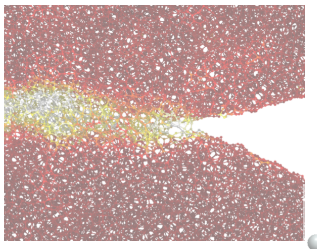


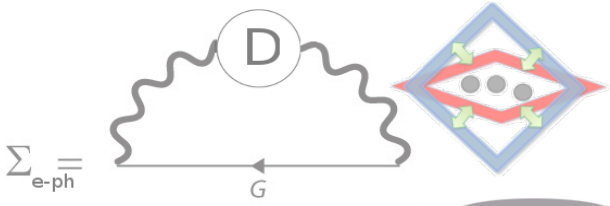

# Workshop Reports 2015

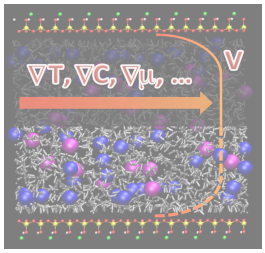


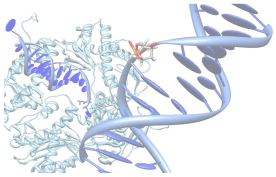

cecam
Centre Européen de Calcul Atomique et Moléculaire

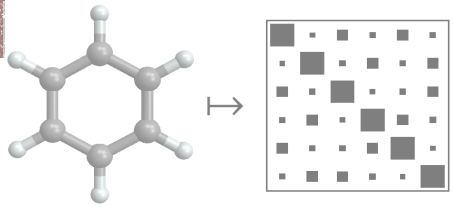


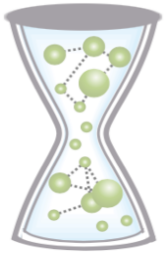


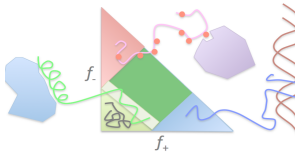


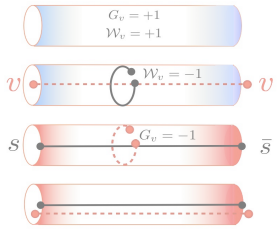


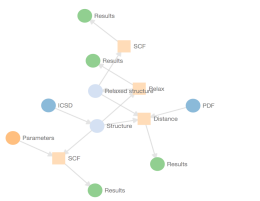


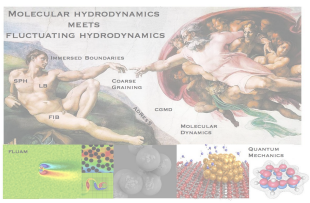


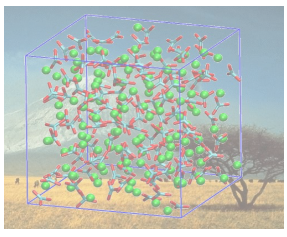


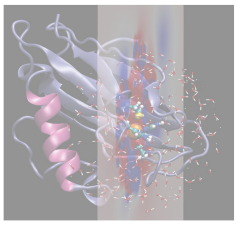


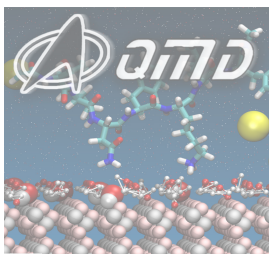


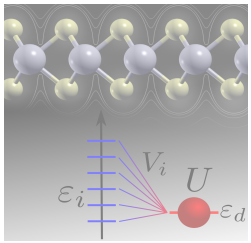



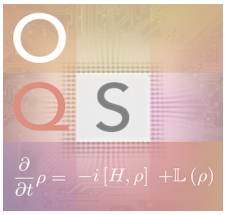



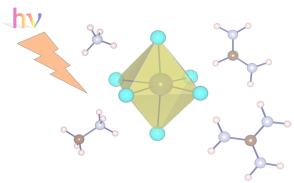



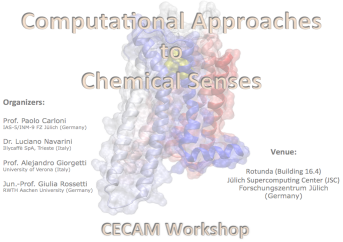


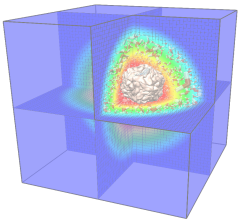


Macromolecular Simulation Software Workshop

# Table of Contents

# Introduction

In 2015 CECAM continued its mission to promote fundamental research on advanced computational methods. A total of 48 workshops were hosted at our Headquarters in Lausanne and in the other 18 Nodes in the CECAM Network. This represents a substantial number of events, with a slightly upward trend compared to last year. It is interesting to note that, while the number of workshops at HQ remained essentially constant, the events at the Nodes have grown, reflecting both the larger number of Nodes in the Network and an overall increase in their activities.

The 2015 workshops covered the core areas of interest for CECAM, in particular:

- Electronic structure calculations and Density Functional Theory.
- Classical and Quantum Monte Carlo.
- Molecular Dynamics.
- First principles dynamics.
- Coarse-grained simulation methods.
- Computational fluid dynamics.
- Multi-scale modelling.

The contents of the workshops emphasized different aspects or applications of simulations that can be broadly classified as follows:

- Biological (10 events).
- Materials (20 events).
- Methods, quantum or classical (18 events).

More than 2300 scientists participated in these workshops. They were at different stages of their careers and came from from 51 countries (31 of these in Europe). 25% of the attendees were women. The contributions of all attendees, in the talks and in the lively CECAM discussion sessions, created an interesting and productive environment for scientific exchange and research, ensuring the success of the 2015 program.

In this on-line booklet, we collect the scientific reports produced by the organizers at the end of each workshop. They contain an overview of the state of the art in the topic addressed, and describe the main outcomes of the meeting. In many cases, they also indicate the infrastructural and funding needs of the communities involved in the different fields and point to possible interdisciplinary collaborations. Taken as a whole, this collection provides an in depth overview of the advancements and challenges in computational science in CECAM's fields. We offer it to the community and to the funding agencies as a useful guide to the developing world of simulation and modelling. Hopefully, it will contribute by spreading awareness of the potential of these methods and by indicating new and exciting directions for our research.

# Charge Transfer Modelling in Chemistry: new methods and solutions for a long-standing problem

Organisers: Masahiro Ehara, Rika Kobayashi, Carlo Adamo, Ilaria Ciofini
Location: CECAM-FR-MOSER, Paris (France)
Date: 7 – 10 April, 2015

**State of the Art:** Light-induced charge-transfer (CT) phenomena have long been known to be a crucial step in several processes ranging, for instance, from those governing photosynthetic routes of plants and bacteria, to those controlling molecular switches or organic dyes functioning. However, despite their experimental interest these processes are still not satisfactorily modelled or predicted from a computational chemistry viewpoint. Even if wavefunction (WF)-based methods are able to provide an accurate description of such excited state phenomena, their computational cost is still too high to make them suitable tools for the investigation of complex molecular compounds. In this respect, Density Functional Theory (DFT)-based methods due to their valuable accuracy/computation-cost ratio become an appealing alternative for a large number of applications, especially when the studied systems contain a significant number of atoms. Unfortunately, standard DFT approaches have failed to model charge-transfer phenomena correctly until relatively recently, where more modern approaches, such as range-separated hybrids provide at least the right trends. The situation is even more complex when the approximations for the excited states treatments (TD-DFT) are summed up, highlighting further issues. WF-based methods are still necessary to reach high accuracy or, more simply, as references for approximate TD-DFT approaches. Recently, methods stemming from many-body perturbation theory showed interesting potential for applications in this domain, albeit the actual example of application is still limited. Furthermore, beside the methodological problems related to the correct choice of the computational procedure, CT phenomena are extremely dependent on environment so that efficient models to simulate structural and electronic effects of the surroundings on CT processes should also be considered.

**Outcomes:** Regarding methodology, discussion of the latest developments in density functionals particularly by Don Truhlar, who is an active developer in this field, concurred that a universal functional may not be readily found in the near future and there would still be work needed in this area. Consequently, more researchers are turning to alternative methods, mostly hybrid methods, anticipating more active developments. An emergent theme in the open discussions was a need to define benchmarking strategies to evaluate reliability. It was highlighted that correct modelling of environment was still problematic with Benedetta Mennucci demonstrating the need of a different treatment of solvent effects in the excited state. To do this requires the as yet unavailable correct response properties for the excited state and a frequency-dependent DFT kernel. Work is underway on the former but as yet work on the latter is at an early stage. The major developments in the next few years would appear to be heading towards tackling large-systems by method development (embedding/fragmentation) or applications. Especially as the discussion of methodology did not promise anything radically new there is more focus on using the current tools in novel ways.

The major obstacles to success were identified foremost as the shortage of skilled computational chemists, especially in programming. Encouragement came in reports from groups that were large enough and well-supported enough to be able to train students in software development and science. As software development can be difficult

and time-consuming the best strategy appeared to be to continue encouraging networks of collaboration.

**Requirements:** Computational chemistry is notoriously resource-intensive stressing compute, memory and disk. Software in the field has been developed across a myriad of program packages, ranging from the easily parallelisable molecular mechanics/dynamics to more complicated ab initio wavefunction codes with density functional theory lying somewhere in between. Whereas the former class of codes, such as NAMD, can run easily in parallel given a decent interconnect, the limitation of the method in providing many high-accuracy properties are leading researchers back to high-accuracy wavefunction methods. The best-known program in this category is Gaussian, which uses OpenMP, and which scales reasonably well within a shared memory machine. There is a trend towards only supporting massively parallel workhorses at supercomputer centres but it is also being found in other fields that large-scale SMP boxes will still be required. For such machines interconnects are not relevant unless as a compromise vSMPs e.g. ScaleMP are deployed, as increasingly being adopted by some computer centres. For training, across all disciplines, there is a perceived need for more discipline-specific scientific programmers, and many institutions no longer provide such training.

# Emergent structural and electronic phenomena at interfaces of nanoscale oxides

Organisers: Jacek Goniakowski, Alessandro Fortunelli. Henrik Groenbeck, Alex Shulger
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 8 – 10 April, 2015

**State of the art**: Novel phenomena, not found in the bulk constituents, such as interface charge, spin, and orbital reconstructions, have been shown to occur at nanoscale interfaces involving oxide materials. These phenomena have stimulated a flurry of activity aiming at creating and manipulating novel electronic properties of oxide interfaces. Polarity of the oxide component adds an additional level of complexity, which may dramatically impact the interface charge state and electronic characteristics. The reduced thickness of oxide films brings additional features, since interface effects may strongly alter the characteristics of both film and substrate. This leads to a sub-critical regime at low thickness, where nano-oxides can be stabilized in a polar uncompensated state or/and with a structure and/or stoichiometry different from the bulk ones. The existence of interface electronic states, their nature, charge state, and character of localization are principally driven by the band character and alignment at the interface. However, ab initio computational studies of interfaces suffer from drawbacks of DFT-based methods. Indeed, it is found that (semi-) local exchange-correlation functionals (GGA or LDA) fail to reproduce the band-gap and absolute positions of valence and conduction bands of the oxide constituents. It is not yet clear how effective are the improvements in the calculated interface band alignments brought about by the use of non-local (hybrid) functionals. The atomic structures of low-dimensional oxides and interfaces have been fully established only in a limited number of cases. It is well known that low dimensional systems may adopt structural motifs, which differ from the corresponding bulk system. Due to the lack of long-range order, it is practically impossible to determine these structures using experimental techniques alone and deep insight and understanding can only be achieved by combining the multitude of different experimental techniques with extensive calculations.

**Outcomes:** Several parallel routes of developments have been clearly evidenced during the workshop. (1) It is becoming increasingly clear that the semi-local functionals commonly used in DFT are not accurate enough to describe various properties of oxide surfaces and interfaces. Because of this, there most likely will be a change of "standard" calculations to include a part of Hartree-Fock exchange (hybrid functionals). However, there are issues regarding the performance of hybrid functionals for metallic systems, which causes problems in describing multi-component systems. One possible route to avoid these problems that most likely will be advanced during the next few years is creating functionals that are space-dependent, i.e. different functionals are used for different parts of the system. (2) The increased use of hybrid-functionals, will be paralleled by an increased use of functionals that include van der Waals interaction. There is clear evidence that accurate treatment of van der Waals interaction is required for a proper description of metal/oxide interfaces as well as adsorption of molecules on oxide surfaces. It is likely that van der Waals functionals will replace the presently often-used semi-empirical corrections. (3) A computationally more efficient alternative is offered by GGA+U approaches, and numerous examples of their successful application to a variety of oxide materials have been discussed during the presentations. However, as stressed during the discussion, in the case of nano-scale objects, surfaces, and interfaces it may become particularly difficult to take into account the effect of local environment on the value of U. Elaboration of (semi-)empirical rules governing the behaviour of U appears as

a promising opening for making this efficient approach more robust. (4) Another growing area concerns powerful tools for structural optimization required for describing the structural and chemical complexity of nano-oxides and their interfaces. Efficient algorithms exist (GA, basin hopping, etc.) and have been successfully applied principally to oxide nano-particles and unsupported thin films. Coming years will see further improvement and extension of these optimization techniques towards supported films and interfaces, with a focus on an efficient handling of strain-induced effects. Also large models for amorphous systems will probably also grow in importance.
(5) Alternatively, application of global optimization methods for a search/optimization of a given property rather than merely for energy minimization has been evoked as an interesting opening towards a more efficient material engineering. (6) Yet another emerging research area that will certainly receive considerable attention over the next few years is that of the oxide/liquid interfaces.
The major obstacles to success have also been identified as: (1) *accuracy of description of the electronic structure of oxides* and, in particular, oxide surfaces and interfaces. Although hybrid functionals have now been used for several years, the computational cost has limited the benchmarking for oxide surfaces and interfaces. The uncertainly regarding the actual performance is to some extent hindering the field. This obstacle will probably be removed during the next years by extensive benchmarking. (2) *Structural complexity*. Very large systems are still out of range for DFT, and these will probably still be handled by model potentials. Indeed, the computational cost of global optimization techniques makes them hardly applicable at the ab initio level. Several examples of successful applications of empirical atomic potentials at the early stages of optimization, followed by further ab initio refinement have been given. While particularly promising, difficulties in an efficient account for alternative metal oxidation states have been discussed and apparent limitations of the empirical potentials in description of more covalent oxides (e.g. titanium oxides) have been stressed. (3) *Validation of the theoretical methods* is of paramount importance. Accurate experimental data on well-defined systems are required to assess the performance of theoretical methods.

**Community needs:** The main requirement of the field is the widespread availability of powerful computer resources in order to perform calculations at sufficient scale and accuracy. "Big Data" concept of a large database for storage of the calculated particle structures, and properties has been evoked in the context of small oxide nano-particles. Aside from difficulties with book-keeping of such a large amount of data, questions related to possible protocols of their production, selection, and verification have been discussed during the workshop.

**Funding:** One important component to advance the understanding and design of oxide-based systems is coordination of theoretical as well as experimental efforts at the European level. Over the years, COST Actions have shown to particularly well serve this purpose and it is desirable that the area will continue to receive support by the funding council. However, the applied-research character of most H2020 calls represents a serious hindrance to the development of a field such as that of nanoscale oxides which is extremely promising in terms of technological applications in the medium term, but which is not yet always at a TRL level sufficiently high to successfully compete with other, less promising but more mature topics. We believe it is of the highest importance for the future competitiveness of EU research in the medium/long term to devise funding schemes/rules in which fields as promising as the present one of nanoscale-thin oxide can be supported.

**Networking & training:** Support in organising and funding workshops promoting international joint experimental and theoretical projects addressing important technological issues are required.

**Will these developments bring societal benefits?** Oxide surface and metal/oxide interfaces are ubiquitous and have extremely wide and important applications in modern technologies ranging from heterogeneous catalysis for sustainable energy conversion to sensors, capacitors, micro-electronic and opto-electronic devices. For example, many catalysts consist of small metal particles dispersed on an oxide substrate. Moreover, many oxides may have an intrinsic catalytic activity, which can be additionally enhanced at the nano-scale (e.g. CO oxidation by chemical looping combustion) and efficiently tuned by the interaction with a metal or an oxide substrate (e.g. SMSI oxide catalysts). Catalysis is an enabling technology that is crucial to various parts of our society. Catalytic materials are used to produce 90-95% of all used chemicals and are vital for emission control and sustainable energy systems. The European industry includes catalyst manufacturers as well as catalyst users. As the computational community is moving into catalyst design, the links to industry will be further strengthened. Issues governing the design of multi-component, nanostructured interfaces for controlling reactivity apply also in a new generation of high performance solid electrochemical devices and advanced nano-electronics. Interfaces of semiconductors and metals with oxide thin films govern the performance of capacitors, actuators, thermoelectrics and memory cells with a total market at estimated 1.5 trillion USD.

## Computational many-body physics in the era of artificial gauge fields

Organisers: Lode Pollet, Andrea Trombettoni, Massimo Capone
Location: Physics Department, Ludwig Maximilian University, Munich (Germany)
Date: 8 – 10 April, 2015

**State of the Art:** In one dimension the DMRG method provides a virtually exact answer for the ground state, but it is limited to low dimensions. It gives limited information in higher dimensions, but can in 1d also give information on real time evolution up to quite long times.

For fermionic systems, in case of time-reversal symmetry (Kramers doublet) or particle-hole symmetry in combination with bipartiteness of the lattice, sign-free determinant Monte Carlo simulations can be done in any dimension. Bosons can also be studied, usually, in a sign-free way. In other cases approximations are usually needed, or methods such as DMFT and DCA are used. With DCA and related methods one can typically study systems with a temperature not lower than the hopping (order of magnitude). Methods that use imaginary time usually need to be combined with analytic continuation in order to get access to real frequencies. This problem is mathematically ill-defined. Systems with spin-orbit coupling (Rashba or Dresselhaus) have in general a sign problem.

**Outcomes:** Experimentally one can expect further developments on systems with strong spin-orbit coupling and the detection of Majorana zero modes with the first attempts to braiding. In atomic and molecular physics more experiments will be performed on topological systems in the presence of interactions. Systems with artificial magnetic and gauge fields will be studied intensively, especially ladders with different rungs and fluxes. The realization of dynamical gauge fields may pave the way to have a physical realization of lattice gauge theories. Also a microscope for fermionic systems is near completion in a number of labs. Other hot topics will be many-body localization and novel types of insulating transitions due to disorder to name just a few.

In numerics, one can expect further developments with tensor network states for studying quantum link models. Systematic studies of the Hubbard model and the electron gas with diagrammatic Monte Carlo. Bosonic dynamical mean-field theory can be extended to deal with current experimental systems with synthetic magnetic fields.

**Obstacles:** For certain problems the current algorithms are advanced enough and they can be implemented straightforwardly. However, for most problems in strongly correlated systems the bottlenecks are novel algorithms. Monte Carlo simulations suffer from the sign problem whereas DMRG based methods work only in low dimensions. In the latter case radically new ideas are needed. The effort for the physical implementation of gauge potentials and fields in ultra cold experiments may also provide a new playground in which the quantum simulation of lattice gauge theories is possible, with an expected further cross-fertilization of ideas between the community working in the numerical simulation of strongly correlated systems and the ones working in ultra cold gases and condensed matter.

We believe the largest obstacle is training and networking. Newcomers in the field must have knowledge of programming, hardware, algorithms but also of field-theory, information science, group theory, superconductivity etc. It is not possible to expect that all these areas are covered in the Master education, nor can one expect that courses and

training are available in each university. Computational work is labor intensive and changing fast. At the same time, one has to make sure that the most talented people can stay in the system by offering clear career paths, realistic grow opportunities, appropriate remuneration, and maintaining a fine balance between academic freedom and the growing influence of third-party funding.

**Requirements:** The community in the first place needs to have a system where it can keep attracting the very best people to work on the world's hardest and most prominent problems in physics. The community needs further funding into the fields of strong spin-orbit coupling in condensed matter physics and atomic and molecular physics, superconducting nano-wires, support for developing algorithms and methods for strongly-correlated systems. Funding is needed for access to computer power at the supercomputers but also an infrastructure where novel ideas can be tried out. This is particularly relevant for the topics discussed in this workshop. Furthermore, greater acceptance is needed for ideas that involve hiring programmers and information scientists in order to integrate them in the team that is mostly physics oriented.

**Will the results of these developments be of societal benefit?** The focus of the workshop was on fundamental theoretical physics. Historically, developments in theoretical physics require about 30 years to reach industrial application. Assessing the needs of European industry on such a long timescale seems a dangerous extrapolation. Nevertheless, certain topics that we have discussed have the potential of understanding strongly correlated materials better. The most prominent example is the issue of Majorana fermions (talk by C. Marcus) and the possibility of observing non-abelian anyons, which are the crucial ingredient for topological quantum computing. The hype created by the company d-wave on the possible demonstration of a quantum annealer led to massive investments by companies such as Microsoft and Google in this field in the USA. Via Station-Q Microsoft is financially supporting several experiments on the fractional quantum Hall effect and the superconducting nanowires (including experiments in Europe). We are not aware of such initiatives by European companies. In our conference we witnessed to a fruitful exchange of competences and fresh ideas between people working in different fields (numerical simulations, strongly correlated systems, condensed matter, AMO), the formation of such common interest being one of the key ideas of our proposal. It is our opinion that events as the Munich CECAM conference we organized may catalyze and focus the interest of a vast European community to these rapidly evolving fields of research, and the outcome of our Workshop and the raised interest fully confirmed our expectation.

# Perspectives of many-particle methods: total energy, spectroscopy and time-dependent dynamics

Organisers: Thomas Frauenheim, Tim Wehling, Thorsten Klüner, Johannes Lischner, Andreas Savin
Location: CECAM-DE-MM1P, University of Bremen (Germany)
Date: 20 – 24 April 2015

**State of the Art:** The field of computational chemistry, physics and material science made recently tremendous steps toward a first-principle description of excited state and correlated electronic systems including solids, nanostructures and macromolecular resp. bio-molecular systems. The impact of the applications strongly depends on the underlying methods and the achieved accuracy by accurately treating non-local interactions, electron correlations and time-dependent processes in electronic ground and excited states. This not only concerns accurate total energies and forces but also a quantitative description of spectroscopic signals and related time-dependent electron and coupled electron-ion dynamics. Despite considerable progress in the field various challenges remain to be solved in the future, addressing

- Many-body interactions – new XC functionals
- Perturbation treatment of quasi-particles beyond DFT
- Strong electron correlations, dynamical mean field theory and beyond
- Time-dependent DFT beyond the adiabatic approximation, two particle excitations

Of prime importance is the development of next-generation realistic many-body computational tools, which are fast, reliable and are able to describe non-trivial quantum dynamics of complex systems. In order to address these problems, new integrated software tools for realistic quantum simulations of correlated systems need to be developed for a broad scientific community [1]. One possible direction to open new fields of applications is the combination of approaches developed in different communities, which shall be fostered at this workshop.

The program consisted of 30 invited talks of 40 minutes (35+5) each and one poster session presenting 29 posters. In addition, many social events (reception and conference dinner) to allow for informal exchange were held. The invited talks were given by well-established scientists from the different theoretical communities, which acted as platform for interesting cross-/interdisciplinary discussions. The invited talks were followed by a poster session where the younger participants could show their scientific work and exchange of ideas with a broad knowledge in computational chemistry, solid-state physics and computational materials science. The organization was very compact with the scientists accommodated in the same hotel fostering exchange and discussion between the participants also outside the meeting room.

**Outcomes:** This workshop brought together people from different correlated electron communities, solid state physics, computational material science and quantum chemistry to discuss a possible synergies and new ideas in quantum many-body methods. The workshop became a forum to brainstorm ideas about solutions to important correlated-electron problems and identify new directions for many-body method development and challenging applications. The presentations have shown, there is already active on-going exchange between many-body solid-state theoretical physics and wave function based correlated quantum chemistry. Developers of different methods and codes are initiating

design of a new generation software tools for many-body quantum modelling of realistic complex systems. The delivery of this technology to a broad community will facilitate future breakthroughs on high-impact materials science problems in magnetic nano-science, transitions-metal biophysics and new energy storage, unconventional high-temperature superconductivity, functional material design.

Computational materials sciences are outstanding growth areas of research. In the future an increasingly larger part of our technological development will depend on computer applications, in particular in materials, nano and bio-nano sciences. Ab-initio calculations based on the density functional theory, wave function based quantum chemistry in combination with dynamical mean-field theory and various quantum-embedding schemes can make a considerable progress in the field of nanostructures with d- and f-elements which usually means strong electron correlations. Moreover, wave function-based schemes offering the systematic pathway for approximating the electronic Schrödinger equation will become more and more efficient, so that also highly complex systems will be and are already accessible today. Prospects of method unifications to approach challenging problems such as electron correlations in complex systems, non-local interactions and dynamical correlations, dynamics and non-equilibrium phenomena have been discussed throughout the workshop.

The most difficult problem in physics, chemistry, biology and materials science is the seamless bridging of different length and time scales. Generally, accurate theories for specific time and length scales exist, but these theories break down in the aforementioned systems because it is no longer possible to treat the different time and length scales independently. For example, highly accurate theories for the electronic structure of small molecules have been developed by chemists, while the solid-state physics community has developed successful descriptions of extended systems. However, the description of localized defects in solids or adsorbed molecules on surfaces or nanostructure requires a joint description of extended and localized aspects.

To overcome this challenge, it is necessary that a dialogue and exchange of ideas between researchers from different communities is established. While such an exchange has been achieved in this conference, the conference has also brought to light that progress in one area of science often remains unknown in closely related areas. To overcome this obstacle, we recommend that further, possible regular interdisciplinary exchanges in the form workshops and conferences are organized.

**Requirements:** As described in the previous section, we believe that the scientific exchange of researchers from different disciplines is key to make progress in the most challenging problems of theoretical materials modelling. This exchange can be achieved by interdisciplinary conferences and workshops. Moreover, it would be desirable to enable exchange not only at the level of researchers, but also at the level of PhD and Masters students.

The field of theoretical materials modelling relies strongly on the availability of state-of-the-art computer code and computing resources. Therefore, the continued funding for the development of modelling software is an important infrastructure requirement: Only when codes are freely available, can new ideas be tested and developed by a broad community of researchers. Since new codes that address the most challenging materials systems will push the limits of current supercomputing resources, it is also commendable that funding of these resources remains significant.

**Will these developments bring societal benefits?** Progress in the field of many body physics and wave function based correlated quantum chemistry is fundamental to many

European industries connected to high-tech materials design and device applications. Examples are

- Advanced hybrid photovoltaics
- Photo-catalytic processes in energy storage and pollutant degradation
- Hybrid nano/bio-systems for medical applications.
- Single-defect-based quantum optical and spintronic devices

Such directions can be strengthened by focused research projects for the development of new materials and devices in key enabling technologies. The field of nanodevices is currently opening to new materials, especially 2D. The EU flagship on graphene and 2D materials is indeed expected with the aid of computational predictions to produce several new outcomes. However, technological innovation is not limited to these novel materials.

# Stochastic Wavefunction Methods in Quantum Chemistry, Electronic Structure Theory and Condensed Matter Physics

Organisers: James J. Shepherd, William Matthew, Colwyn Foulkes
Location: CECAM-HQ, EPFL Lausanne (Switzerland)
Date: 21 – 24 April, 2015

**State of the Art:** Electronic structure theory faces many exciting new challenges in the coming years. These are dominated by the challenge of solving the many-body problem to high accuracy, the need to achieve this consistently for large systems, and the requirement to adapt software to high-performance computing in the petascale era. Monte Carlo methods are immediately suited to meet these challenges, and a wealth of new stochastic algorithms for the numerical solution of the many-particle Schrödinger equation in quantum chemistry and condensed-matter physics have arisen in recent years.

At this very diverse workshop, we heard leaders in the application and development of a wide range of different stochastic approaches to electronic structure theory discuss the state of the art. These include some substantial successes:

1. Studying metal-insulator transitions with diffusion Monte Carlo
2. Using diffusion Monte Carlo to improve DFT simulations of condensed phase water and materials in semiconductor physics
3. The use of auxiliary field quantum Monte Carlo in condensed matter physics
4. Obtaining exact wave functions for chemistry and the solid state with full configuration interaction quantum Monte Carlo
5. The phase transition between a mixed molecular-atomic quantum liquid and a completely dissociated fluid has been observed with quantum Monte Carlo molecular dynamics
6. Exotic heavy-element phenomena in bulk cerium were understood using variational Monte Carlo

These successes represent a considerable growth in the utility of quantum Monte Carlo methods within the last 3-4 years, and the number of high impact papers published by the attendees of this conference has been substantial. This workshop sought to consolidate this growth and discuss ways to build on recent successes.

**Outcomes:** There is a great sense of achievement in the community, especially where individual groups are concerned. If these trajectories continue, the futures of these research programs are very bright indeed.

Taking the community as a whole, though, there are various major advances that would be desirable. Those marked (*) should be achieved in the next 2-3 years; others are longer-term goals.

1) Open source codes; fewer codes with more developers *
2) Making codes more black box; reducing the human time to run simulations *
3) Better interfacing/GUI with quantum chemical codes like Gaussian *
4) Lower accuracy settings for faster calculations
5) Coalescing around solutions to the forces problem
6) Better pseudopotentials (diffusion Monte Carlo) or no pseudopotentials (auxiliary field QMC)

7) Better support for the training and retention of postdocs

**Major obstacles**: QMC offers exact or almost exact answers, but at high resource cost. The biggest bottleneck of these is probably human time. The current practical reality of running a QMC calculation seems to be the biggest barrier to success. A typical "flagship" QMC calculation is run by a PhD student who has trained full time for more than a year in a recognised QMC group. Such calculations typically take around 18 months of human time and substantial supercomputer time. Most calculations involve some degree of coding, typically optimization of algorithms or development of new methods.

In particular, then:

1) QMC codes are cutting-edge software and require cutting-edge computers to demonstrate their value.
2) QMC has a large learning/technical barrier of entry. This means that the number of expert researchers in the field is relatively low.
3) QMC grant writing is perceived to be challenging, but success seems relatively abundant.
4) Having a method that is in principle exact means that publications are now expected, under review to be more rigorous. This further delays publication and can lead to infighting in the field.
5) QMC development is still needed, although less than commonly perceived. There are some problems, for example cusps, pseudopotentials and forces, which are in the process of being overcome but are either absent or easily dealt with in less accurate methods like DFT.

The progress made in the next five years will likely hinge on whether those groups are able to coalesce and work together to build the community.

In spite of these apparent drawbacks, QMC is currently flourishing. In the last 3-4 years, however, a number of Faculty jobs in the US and elsewhere have been taken by PIs who intend to run QMC-focused groups. This has greatly augmented the number of people now doing QMC. The complexity of the challenges mentioned above can attract some of the best students to our community. Furthermore, two codes are scheduled for open source release in the upcoming year. This will add auxiliary field QMC and full configuration interaction QMC family of methods to the list of QMC codes easily available.

The standard of QMC has never been higher and the recent increase in the number of flagship papers on important applications has enabled us to encourage more students to sign up for QMC at post-graduate level. In a few years, we will also have more postdocs. The main things we need to do to overcome the obstacles above are to continue to meet and talk, to continue to do good work, and to foster the careers of young people.

In this field, we look for anything that would lower the labour cost of running a QMC group, or running QMC calculations. In terms of physical resource this is: faster computers; reliable compilers; black box and robust parallel libraries; consistent parallel architectures; and software engineers having an awareness of QMC requirements. In terms of human resource and advocacy this includes: mentoring for young people; small-scale intense consultancy from experts in other communities (to answer the question of what can we do for them?); resources for advocacy and outreach to other communities; and recognition of the importance of software engineering in academia. In terms of training this means: small-group, two-day PI retreats with substantial travel support; workshops between communities; and summer schools for student.

**Community needs:** Our research community needs: programme managers who understand the role of high accuracy electronic structure calculations, both in historical and modern research; programme managers who understand the long-term importance of continuing to develop new methods and algorithms; the development of new, rather than the application of existing tools to new problems; and the support for large-scale (e.g. exascale) parallel architectures.

**Will these developments bring societal benefits?** Quantum Monte Carlo is the basis for a user-friendly, widely used, post-DFT method. The work presented in this workshop shows that the community as a whole is making leaps and bounds towards working directly with European industry.

The challenge for QMC is to find problems that would benefit from having a small number of extremely high accuracy calculations performed some 18 months to 2 years down the line. We can expect to be cautiously optimistic at the rise in data science and machine learning techniques entering the field. This is just such the case where a small number of highly reliable data points can have a substantial impact.

With that in mind, there are some potentially promising lines of enquiry:

- Battery design
- Catalysis design
- The polymorph problem and drug discovery
- Drug delivery design
- Chemical/chemical biology mechanism elucidation

## Green's function methods: the next generation II

Organisers: Arjan Berger, Pina Romaniello, Francesco Sottile
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 4 – 7 May, 2015

**State of the art:** Green's functions have always played a prominent role in many-body physics. In particular the one-body Green's function (GF) delivers a wealth of information about a physical system, such as ground-state energies, excitation energies, densities and other measurable quantities. Therefore the development of approximate methods to calculate the one-body GF has been an active research topic in many-body physics since the 60's, and many routes have been explored in order to find increasingly accurate GFs. A very popular class of methods is based on the iterative solution of an integral equation for the GF containing an effective potential, the so-called self-energy, which needs to be approximated. The well-known GW approximation belongs to this class; this approximation is the method of choice for calculating band structures, but it also shows several shortcomings, such as the wrong description of satellites in photoemission spectra, in particular in so-called strongly-correlated materials. Therefore more refined levels of approximations are needed to keep the pace with the advances made in experiment. Recently much progress has been made in this direction both by going beyond standard methods and also exploring completely novel routes to calculate GF. A new wave of original ideas, understanding, and solutions, has pervaded the field and therefore we think it was timely to gather these new concepts in a workshop.

**Outcomes:** It is clear that state-of-the-art approximations are not sufficient anymore to describe new phenomena and emerging new physics that the advances in the experimental techniques allow us to explore.
The so-called strongly correlated materials, for example, exhibit remarkable electronic and magnetic properties, such as metal-insulator transitions, half-metallicity, or unconventional superconductivity, which make them among the most attractive and versatile materials. Describing these systems requires to go beyond today's state-of-the-art approaches and represents nowadays one of the greatest challenge for condensed-matter theory. New strategies start to emerge which revisit fundamental equations and reformulate them in such a way to make clearer the physics they describe. New physical constraints are searched in order to obtain physical approximations. It is also clear that collaboration with other fields is essential, since similar problems are common to various domains and solutions might already exist.
The four most important scientific advances that are required in the next 2 to 3 years are:

1) Fully self-consistent GW calculations for realistic periodic materials to compare to the non-selfconsistent results (role of vertex and excitonic effects included)
2) Dynamical mean-field calculations from first principles in order to assess its predictive power
3) Solving the Kadanoff-Baym equation numerically for real materials by calculating the functional derivative of the Green function with respect to an external potential
4) A systematic approach to obtain the physical solution amongst a multitude of mathematical solutions that one obtains when solving the functional differential equations for the Green functions

The main goal is to have a method which is accurate and computational efficient at the same time. The main obstacle of going beyond the state-of-the-art GW approximation, for example, is that, when using the standard hierarchy, the equations to deal with quickly

become impossible to treat numerically. Moreover, it is not guaranteed that results will improve. In this respect the study and understanding of the physics underlying the approximations used is essential and will allow us to overcome this possible roadblock.
There is also much effort in reducing the scaling of the calculations. For example self-consistent GW calculations are becoming standard in finite systems, but they are not yet standard for extended systems. Therefore, besides fundamental development we also need numerical development to optimize the codes.

**Community needs:** Many of the concepts here reported will benefit from an increase support of fundamental research. However, for what concerns specifically software and hardware, we have both good and bad news. Many groups have already partnerships with local software engineers, few others decided to ask support to the PRACE facilities. This is a crucial step: it is indeed true that the computer power available today (both in national and European calls) is good news to tackle bigger and more complicate systems. However without a proper writing (and often, re-writing) of our computer codes (and this is the bad news), to take full advantage of the modern paradigms (blue-gene low memory nodes, GPUs accelerators, mics, and all hybrid architectures), there won't be any possibility to tackle such complex materials and behaviours. The scalability required to take full advantage of this new kind of resources is in fact such (1-1000000 cores) that only a professional approach can succeed.

Finally, in order to make these new codes and capabilities widely available, training strategies have to be devised, at both hard (summer dedicated schools) and soft (e-learning platforms, MOOCS) levels.

**Funding:** The research area on many-body Green's functions is a fast-developing research field that is attracting a lot of attention, especially in recent years, not only for the fundamental research but also for its application to tackle important technological problems.
*Ab initio* methods based on Green's functions are indeed the methods of choice for band-structure and photoemission-spectra calculations, and therefore they are essential tools for identifying, for example, candidates for photovoltaic and photocatalytic applications, topological insulators, phase transitions, etc. In order to improve the state-of-the art methods and to be able to treat materials and physics of growing complexity, progress in fundamental research is crucial. This is why the field offers a playground for technological as well as intellectual challenges.
Therefore, it is of utmost importance that funding agencies continue or even increase their funding of *fundamental research* as this will both increase knowledge and advance technology.

The workshop brought together many established experts in the field of many-body Green's function methods as well as young promising talented researchers. Therefore there are several EU programs that might fund these researchers such as

- ERC starting grant
- ERC consolidator grant
- ERC advanced grant
- ERC synergy grant

**Will these developments bring societal benefits?** Progress in Green's function methods is of great importance to European Industry. Many-body Green's functions contain a wealth of information and this information can be used to develop novel materials for various innovating technological applications.

For example, the single-particle Green's function contains information on the electron addition and removal energies, which allows for an accurate description of band gaps, band alignments etc.. This is very important information in the development of photovoltaic devices. The single-particle Green's function also contains information on the current density, which can be used to describe accurately all kinds of phenomena related to electron transport.

From the two-particle Green's function we can obtain information about the interaction between two particles such as the interaction of an excited electron with the hole it leaves behind. This is crucial information for the development of new photovoltaic and photocatalytic devices, for which its success largely depends on the ability of the material to split the electron and the hole after their formation.

However, this applies to the field globally. Concerning the more restricted focus of the workshop in particular, it is about fundamental research preparing the long term future on a scale of, say, 10 years which is not of direct interest for a company now.

# Molecular hydrodynamics meets fluctuating hydrodynamics

Organisers: Ignacio Pagonabarraga Mora, Rafael Delgado-Buscalioni, Pep Español
Location: CECAM-ES, Residencia La Cristalera, Miraflores de la Sierra, Madrid (SPAIN)
Date: 10 – 14 May, 2015

**State of the Art:** Structure and dynamics of soft matter are governed by molecular interactions at short scales and the emerging collective behaviour due to these molecular interactions at large scales. These large scales are described by hydrodynamics, which seems to apply well down to the nanoscale. At these scales one needs to include also thermal fluctuations that are the responsible of the Brownian motion of suspended structures (nanoparticles, colloids, polymers). While there is a clear conceptual understanding of the issues involved, there are several important fronts, particularly at the computational level, that require further work.

**Outcomes:** Over the 5 days workshop about 25 speakers and 50 participants in total discussed several topics related to multiscale modelling between nano and micron scales. This is a brief list of the major discussions.

1. Nanofluids: Nanofluids are made of nanoscopic colloids and are becoming the subject of intensive research due to its potential industrial and technological applications. They are one of the systems in which really MD meets FD. On one hand, the physicochemical interactions between the nanoparticles and the solvent are crucial in their macroscopic performance while at the same time MD is no longer able to deal with a system of just a few nanoparticles in suspension. Methods for reliably coarse graining the particles, making them a bead, for example, are required and this implies both, CG nanoparticle-nanoparticle interactions but also CG nanoparticle-fluid interactions. We expect to see much work coming in this field.

2. How to Coarse Grain: There is now a large body of knowledge about how to coarse grain a complex molecular system in terms of simpler representations, usually in the form of blobs or united atoms in order to reproduce *structural* properties. The different methods (Boltzmann inversion, force matching, relative entropy) for obtaining the effective interactions between the CG units are well understood and its pros and cons have been assessed. The effective interactions respect the structure at the CG level, but usually the *dynamics* is not well captured by the effective interactions and phenomenological fitting (time scaling) is used. The treatment of friction and its associated thermal noise is crucial to capture the dynamics and much less is known in this respect. Note that one aspect that immediately enters when considering dynamics is the issue of memory. It is expected that only if the degree of coarse-graining is large (large CG units) there will be separation of time scales and the possibility of describing the dynamics with usual Stochastic Differential Equations (i.e. Langevin like equations). But this goes at the expense of loosing molecular detail, which may be too important to loose in some cases. The dilemma faced by a dynamic coarse-grainer is the following: do I coarse-grain sufficiently to ensure the validity of the time scale separation with the danger of loosing to much detail or do I not?

3. Memory: When we go down scales towards the nanoscopic world, hydrodynamics not only starts to fluctuate but also is less and less Markovian. In fact, memory in the form of viscoelastic effects starts to play a role even for simple fluids at nanoscales. It is possible to deal with these effects by including more non-hydrodynamic variables into the

description, along well-established generalized hydrodynamics, but its implementation in actual algorithms for fluctuating hydrodynamics is rare.
Methods for computing the memory kernel for simple liquids out of MD have been discussed in the workshop, and ways of simulating the resulting equations. Another interesting talk discussed the origin of hydrodynamic Green's functions at molecular leve (Oseen-type tensors created by a single molecule).

4. Boundary conditions: The issue of boundary conditions of fluids at interfaces is closely related to the problem of coarse-graining solid-liquid interactions. There have been theoretical advances in how to do this in order to predict the degree of slip or thermal resistance at interfaces out of the molecular interactions. However, not much is known about how to introduce these macroscopic concepts (slip, thermal resistance) in boundary conditions for *fluctuating* hydrodynamics. This is, a derivation from the microscopic dynamics of the solid-liquid interaction of the appropriate boundary conditions to be applied in fluctuating-hydrodynamics is still missing. Also, the very concept of boundary condition may need to be abandoned when the size of the interface is in the nanoscale. While typically one uses "hydrodynamic radii" and "slip", both are just fitting parameters whose connection with the molecular dynamics is unclear.

5. Polymers Hierarchy (problem of friction, empirical time scaling). The problem of friction and characteristic times in polymer melts is a difficult one. T One session was devoted to analyse results from full atomistic MD of polymer melts. Although the coarse-graining of friction between polymer "blobs" is tractable for short chains, it becomes difficult in the case of entanglements. A discussion was open to bring about ideas about how to coarse-grain entanglements effects.

6. Hybrid schemes: these are computational schemes in which some region of space that needs detailed molecular information (near a solid surface or a protein) the system is described with Molecular Dynamics, while in regions where the collective large scale behaviour is relevant (and molecular details irrelevant) the system is described with a computationally much cheaper model like for example a coarse-grained molecular representation in terms of blobs or fluctuating hydrodynamics. There have been a lot of methodological developments in past years, but these methods face always the challenge of the "different clocks" required by each dynamics. It is foreseen that work in the direction of irregular adaptive grids for solving fluctuating hydrodynamics may lead to interesting developments, although the time integrators will require as much attention as the space discretization, precisely to account for the different characteristic times associated with different resolutions. The treatment of the "sub-grid" suspended particles and its interaction with the grid resolved hydrodynamic fields seems to be another area of development where theoretical work will be crucial in order to coarse-grain the molecular interactions.

**Requirements:** The area is experiencing several kinds of advances, which were discussed over the workshop. First, we need to advance in the coarse-graining theoretical models, starting from atomistic detail up. Examples of this works are based on extended Mori-Zwanzig formalisms (Español), and also force matching and relative entropy formalisms (Dama) and density functional theory extension to deal with dynamics. Advances on the role of memory in complex fluids, such as those presented by Rotemberg and Taras Brik are also promising. A more difficult issue is to deal with entalglements in polymers (Harmandaris), where coarse-graining is still ellusive. Adaptive resolution (Paprotknik, Kremer) and flexible hybrid models connecting fluctuating hydrodynamics (Karabasov) and molecular dynamics are a second front line. A third line of research is the extension of fluctuating hydrodynamics to deal with multiphase flow, mixtures, reactions (Donev), also including the immersed boundary formalism for solutes, with extended properties

such as arbitary compressibility (Balboa), surface tension, electrostatic interactions (Voulgarakis) energy exchange, surface properties (Harting), etc. The applicability of such methods (either based on Lattice Boltzmann or on finite volume schemes) is quite large and is under fast development. On the software and hardware side, Graphical Proccessor Units (GPU's) are now extending over this domain, and over the workshop several codes based on GPU's were presented (LB, Hybrids and Fluctuating hydrodynamics). Extensions of Adaptive Resolution to GPU's are still missing, probably due to difficulties in parallelization, but would be much appreciated. Training or working Networks to connect the several parts of this, relatively sparse, community are also starting to appear (COST, Marie-Curie) but are still scarce. More connection with industry would also benefit the community, because industry is certainly in a need to solve real multiscale problems.

This question is a generic one. Indeed, the community would benefit from funds to start multidisciplinary projects, including a sound theoretical and numerical perspective. We would insist on first funding doctorate students and secondly postdoctoral fellows.

**Will these developments bring societal benefits?** Multiscale research is key for industrial problems and (it is now the time to start approaching industry) the approach of the community to industry is happening steadily. Over the workshop we had a presentation about nanoparticles in solar collectors (Pietro Asinari), which was carried out in collaboration with industry. We were also in contact with Abengoa Research in other related lines. Companies such as REPSOL were also present in the workshop, due to their great interest in these multiscale methodologies. And finally XFLOW, a company on scientific modelling was also present at the workshop to learn new techniques in Lattice Boltzmann and multiphase flows. The support needed to contact industry should mostly come from local and national institutions, which should foster these sorts of connections. Academic work is however time-consuming, so these kind of connections (academy-industry) should probably have a deeper bottom-line and be more flexible in time, something that is not in the portfolio of most industries. Despite all of this, we see that multiscale modelling is now attracting many large and medium-size industrial R&D departments.

# From Many-Body Hamiltonians to Machine Learning and Back

Organisers: Matthias Rupp, Alexandre Tkatchenko
Location: CECAM-DE-MM1P, Dahlem, Berlin (Germany)
Date: 11 – 13 May, 2015

**State of the art**: An approach to significantly speed up atomic and molecular calculations is to combine quantum mechanics (QM) with machine learning (ML). Such hybrid QM/ML models use ML to interpolate between QM reference calculations, based on the observation that running the same calculation on similar inputs implies some form of redundancy that can be exploited. Understanding of what makes a representation well suited is still incomplete, in particular when comparing different systems. Parts of the field, in particular potential energy surface interpolation for molecular dynamics simulations, are comparatively mature, and are being applied to real systems. This is done using know-how and technology inside individual research groups. Steps towards integrating existing technology into widely available software packages for quantum chemistry would be desirable. Other parts of the field, such as predictions across chemical compound and materials space, are less mature, with a lot of model development work still required. Despite this, it would be highly desirable to see an actual compound or materials design study facilitated by quantum mechanics/machine learning models, ideally in collaboration with experimentalists, in the next two years.

**Outcomes:** The workshop contributed towards more systematic, constructive development based on theoretical understanding by providing the opportunity for side-by-side comparison and active discussion of successful representations developed so far, including symmetry functions, smooth overlap of atomic positions, and the Coulomb matrix. This included strengths and weaknesses of the different approaches, e.g., discontinuities in sorted Coulomb matrices. Specifically, while general requirements such as invariance to translation, rotation, and nuclear permutations are undisputed, the interplay between representations and interpolated property is less clear (e.g., property-specific invariances, gradual coarsening depending on target accuracy). It was noted that several successful representations could be interpreted as simultaneously viewing the represented systems at different spatial resolutions. Other questions discussed at the workshop included, among others, domain of applicability and confidence of predictions, interpretability of models, coupling to physical models of long-range interactions, and availability of data. Some of the major obstacles to success were also identified as: (a) *Lack of benchmarks.* To reliably assess quality and performance of new proposed models standardized protocols and benchmark suites are needed. While individual datasets and guidelines have been published, an agreed-upon standard is missing. The community needs to agree on and develop this. (b) *Lack of data.* Large-scale databases of computational results are needed. While projects like the materials genome initiative, aflowlib, and NoMaD have been set up, they need to be further developed and established (e.g., NoMaD has become an official repository for the Scientific Data journal), similar to mandatory deposition of new protein structures into PDB repository. This is being actively pursued. (c) *Lack of dual-education researchers.* There are only very few researchers with dual degrees in chemistry/physics and computer science, but these people are needed due to the interdisciplinary nature of QM/ML models. In their absence, an alternative is to push for interdisciplinary workshops, summer schools, and exchange visits.

**Community needs:** The need to establish data infrastructure in the form of public repositories has been recognized by funding agencies (e.g., recent EU large-scale funding

for the NoMaD laboratory). It is now up to the community to establish these repositories, both technically and organizationally, e.g., by encouraging deposition of calculations used in publications. Recent conferences and workshops focusing on QM/ML models (IPAM long research program on Navigating Chemical Compound Space for Materials and Bio Design, 2011; CECAM workshop on Machine Learning in Atomistic Simulations, 2012; IPAM workshop on Machine Learning for Many-Particle Systems, 2015; this workshop) have been of a technical and research-oriented nature. It is now time for training-oriented events such as summer schools, workshops with tutorial sessions, student and postdoc visits between groups, and similar endeavors.

**Networking & training:** The workshop brought together many of the researchers actively working on this topic, providing a cross-section of the current state of the art, with a focus on how to numerically represent molecules or materials as input for the ML algorithms, a crucial step in the design of QM/ML models. Many participants explicitly and strongly endorsed the format of the workshop (limited number of participants, equal number of junior and senior researchers, focus on discussions (15 min talk, 30 min discussion), scheduled discussion breaks between talks, scheduled blocks of time for discussion in smaller groups).

**Will these developments bring societal benefits?** Models combining quantum mechanics with machine learning have the potential to benefit areas such as health care and green energy indirectly by widening the scope in terms of system size, time scales, etc. of current calculations, facilitating computational design of, e.g., new chemical entities, better solar cells, and batteries.

## Simulation of systems under thermodynamic-like gradients

Organisers: Ignacio Pagonabarraga Mora, Benjamin Rotenberg, Emmanuel Trizac
Location: CECAM-ES, University of Zaragoza (Spain)
Date: 20 – 22 May, 2015

**State of the Art**: The first objective of the meeting is to identify and address the challenges related to the fundamental understanding and simulation of systems under thermodynamic gradients. In particular, the discussions will focus on the following aspects:

1) How to simulate thermodynamic gradients and determine the corresponding transport coefficients?

    a) Equilibrium versus non-equilibrium
    b) Technical aspects (size effects, etc.)
    c) Non-equilibrium evolution with chemical reactivity
    d) How to describe these phenomena on the molecular and mesoscopic scales
    e) Energy dissipation and entropy production

2) Role of fluctuations, of confinement

    a) From the fundamental point of view, what role do interfaces play? For example, it seems that an osmotic pressure gradient can generate flow only if an interface is present, not in the bulk.
    b) From the simulation point of view, what should be done with the walls (in particular in terms of energy transfer, thermostats, etc)?
    c) For systems under confinement, fluctuations play a more important role than in the bulk. What is the impact of these fluctuations on the convergence of the computed properties?

A second objective is to encourage discussions between experimentalists and experts in simulation and theory on these topics. We have identified a tentative list of participants with this in mind.

The third objective is to encourage cross-fertilization of ideas and challenges among diverse fields where thermodynamic gradients do play a relevant role and to identify which are the technological and fundamental issues in the different areas.

**Major outcomes:** The important of thermodynamic driving is becoming more and more appreciated, in particular due to the shift from micro to nano-fluidics and essential role played by interfaces, as well as in the context of active matter. In the next 2-3 years we expect significant work in the following directions:

- Fundamental questions to analyze non equilibrium phenomena: sampling issues, connection with continuum thermodynamics of irreversible processes, definition of probability distributions and relevant correlation functions for transport coefficients (in particular at interfaces), entropy production in microscopic models, fluctuation theorems, chemical reaction and meaning of hybrid MD/GCMC in the context of reactive systems.
- Fundamental mechanics underlying mesoscopic or thermodynamic descriptions: osmotic pressure and temperature gradients from

hamiltonian/liouvillian, Brownian motion in active colloidal systems, and dissipation in non-isothermal systems. Based on the above fundamental questions, how to determine the relevant transport parameters for a given system (solvent or solute specific effects), in particular for thermal gradients. The experimental determination of microscopic transport coefficients (eg. ion specific specific Soret coefficients in water?) will require the development of new experimental techniques.

- The issue of thermodynamic driving is encountered everywhere. Researchers from very different areas (physics, chemistry, biology, engineering) may have very different views of the same phenomena. It is therefore important to bring together different communities, who are not always aware of the limitations of their own approach.

In order to address some of the above issues, reliable experimental data is needed to validate the models or discriminate between them. Only recently have such experiments started to be developed, in particular for probing the interfaces (under confinement or at the surface of particles) on the nanoscale. The collaboration between theory/simulation and state-of-the-art experiments need to be fostered. Such links are particularly essential for materials undergoing transformations on a variety of length and time scales (an illustration discussed during the workshop can be found in cement).

**Infrastructure requirements:** Based on the above discussion, it is clear that networking and training will play a major role in making progress. Various European activities may provide such framework, in particular:

- European Training Networks, such as the NANOTRANS project that has just been selected for funding and involving some of the participants to the present workshop
- FET calls, that in addition should involve industrial partners embarked on this topic
- Future related CECAM workshops: based on the very active discussion during the present workshop, we anticipate that significant progress can be made in the next few years, so that another meeting on the same (or closely related) topic will be useful to assess the achievements in the field.

**Will these developments bring societal benefits?** The workshop allowed identifying a number of practical applications for which a fundamental understanding of thermodynamic driving is needed, pertaining to:

- Material construction, in particular the improvement of processes involving cement, which is a major source of carbon emissions at the global level.
- Durability of batteries.
- Shale gas and shale oil extraction.
- Medical applications for diagnosis exploiting thermodynamic (e.g. thermal and concentration) gradients.
- Nanotechnology in general, including nanofluidics, nanomachines and nanorobots.
- Energy harvesting (osmotic power).

# Collective dynamics in physics, biology and social sciences

Organisers: Vladimir Lobaskin, Thomas Ihle
Location: CECAM-IRL, University College Dublin (Ireland)
Date: 20 – 22 May, 2015

**State of the Art:** Collective motion is used by various living species to improve the efficiency of foraging, mating, migration, or defense. The level of consciousness of the individuals apparently plays minor role for the large scale dynamics as the same principles of self-organisation that govern the dynamics of groups of animals or cells apply to human social phenomena, traffic, robotics, and decision making. One of the best-studied collective dynamics phenomena is the onset of globally aligned motion in active swarms. It demonstrates many of the features of the paramagnetic-ferromagnetic phase transition, where the onset of orientationally ordered phase is caused by anisotropic interactions between individual dipoles. In swarms of active particles, the aligning interactions usually compete with noise at a fairly constant propulsion speed. The commonly used order parameter for the swarm, the mean particle velocity measured as a function of the noise amplitude, behaves in the same way as the magnetization vector upon the variation of temperature. As these phenomena share the type of symmetry breaking, it is not surprising that they demonstrate similarity in the behaviour of spatial correlations of the order parameter, of the susceptibility, and of the correlation radius at the critical point. It should be noted, however, that the analogy between these two types of systems is far from obvious as we attempt comparing static and dynamic phenomena. In contrast to equilibrium magnetization, the global alignment in swarms is only possible in far-from-equilibrium conditions, when the rate of the external energy supply to the particles is sufficiently high. Moreover, in contrast to molecular systems, the groups of living species are often inhomogeneous, and individual particles can play different roles in the group. It is tempting to suggest that the similarity between the equilibrium condensed matter and active swarms is not limited to static or steady state statistical properties but also extends to dynamics. Moreover, the dynamic hysteresis can provide an insight into the mechanisms of dynamic self-organisation in systems with complex interactions such as groups of animals, human crowds, traffic flow, and social groups.

**Outcomes:** The Workshop brought together speakers, including experimentalists, to discuss the common features and difference between the wide varieties of systems, which exhibit collective dynamics. Examples include living systems, such as schools of fish, swarms of birds, pedestrians and swimming microbes, which are called "active matter" as they are composed of self-propelled constituents. A common theme in several of the presentations was the intrinsic non-equilibrium character of active matter, and its capacity to exhibits a plethora of novel collective phenomena as revealed by a recent combined effort of statistical theory, computer simulations and real-space experiments. The level of consciousness of the individuals apparently plays minor role for the large scale dynamics as the same general principles that apply to groups of animals or cells seem to govern also human social phenomena, traffic, robotics, and decision making. Therefore, similar modelling approaches are employed to describe their generic dynamic properties. A number of ideas emerged from the workshop:

1. It seems that there is still space for fundamentally new models for collective dynamics. These, among many others, should address the questions of heterogeneity modularity and hierarchy observed in group behaviour
2. Most collective dynamic models work with relatively simple components (agents). Increasing the complexity of the components in tandem with the

richness of the interactions will be crucial, and there is a major challenge in finding the most useful level of description of both to best describe specific phenomena.
3. The design of synthetic systems that will allow the control of the critical features of active matter (propulsion mechanism, geometry, polarity), free from the constraints of living organisms. If successful, this will provide the opportunity to probe the limits of self-organization in dynamic systems that operate far from equilibrium.

**Obstacles:**

1. As one tries to dig deeper into the details of a living group the picture becomes more and more complex, many times getting beyond feasibility at this stage. Models and experiments should become more realistic and sophisticated.
2. Many established fields will not readily accept the kind of context-dependent modelling coming out of work on collective dynamics. For most fields, there are background assumptions about the essential nature of the components that must be questioned if we aspire to finding a phenomenon-specific level of description of both components and interactions.
3. It is important to bring together experts from diverse disciplines for extended workshops.
4. Currently, there are serious "language" barriers that prevent in-depth understanding of the problems. As an example from active matter research, there are few opportunities for theorists and experimentalists to get together to hash out the important issues in the field.

**Requirements:**

1. Groups of quadrupeds will be observed in the future by quadcopters. HD footages need storage memory and digital video image processing is quite CPU demanding

**Will these developments bring societal benefits?:**

Possible applications include

1. Societal planning
2. Animal (including fish) Conservation and stock maintenance
3. Modelling of social networks – including for commercial applications

# Electron-vibration coupling: theoretical and numerical challenges

Organisers: Andrea Marini, Xavier Gonze, Claudia Draxl
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 27 – 29 May 2015

**State of the Art:** The electron-vibration (EV) coupling is well known to play a key role in several physical phenomena. For example it affects the renormalization of the electronic bands [1], the carriers mobility in organic devices [2] or the position and intensity of Raman peaks [3]. The EV coupling is also the driving force that the transition to a superconducting phase in solids and nanostructures [5].

The EV interaction has been the subject of several fundamental studies thanks to its link with very well known phenomena like superconductivity [20] and platonic theories [6,15]. However, the study of this very fundamental interaction in an *ab-initio framework* has received very limited attention and only recently numerous works in this field have appeared.

This is inevitably creating a gap between the *ab-initio* [16,17,18] and the experimental communities as most of the experiments are performed at finite temperature and many of them directly address effects induced by the EV coupling.

The state-of-the-art of the research activity in the EV field can be easily appreciated by noticing that the first works addressing the EV interaction that also relied on an atomistic description of the materials appeared many years ago [1] by Heine, Allen and Cardona (HAC). They pointed out that the EV coupling could induce corrections of the electronic levels as large as those induced by the electronic correlation. However their calculations where based on the use of semi-empirical methods.

Their approach was based on a simple static perturbation theory approach that was neglecting all sort of dynamical effects. It is surprising to notice that this approach has remained the state-of-the-art method for many years (almost 40!) as witnessed by the many authors that used it. The HAC approach has been, indeed, used to compute the gap renormalization in carbon–nanotubes [11], the finite temperature optical properties of semiconductors and insulators [12], and to confirm a large zero–point renormalization of the band–gap of bulk diamond [4,5,13], previously calculated by Holler et al. Using semi–empirical methods [14]. Also the EV-induced lifetimes have been calculated using a simple extension of the HAC approach [10]. Also Raman spectra [8] and phonon-assisted absorption [7] have been interpreted in terms of the HAC approach.

On the experimental point of view very recent experimental results have disclosed the complex dynamical structures that appear in the spectral functions of prototypical materials, like in the case of TiO2 [21], an oxide. The results have been interpreted by using model Hamiltonians and ad-hoc parameters as an atomist theory is still lacking.

The workshop was held in the CECAM headquarters in Lausanne on May 27-29, 2015, and gathered more than 30 participants, including 14 speakers. It was structured in four sessions:

1. Foundations and spectroscopy (3 presentations)
2. Model Hamiltonians (2 presentations)
3. Ab-initio approaches (6 presentations)

4. Heat transport and thermoelectricity (3 presentations)

Most of the presentations were 30 minutes long, with 30 minutes discussion. There were also three round tables to discuss future directions, each lasting one hour or more, as well as a poster session (May 27), and a social dinner (May 28).

Overall, we had excellent talks followed by very lively discussions. The format of the workshop fostered fruitful in-depth discussions that are usually prevented if most of the time is dedicated to presentations. Open questions that were brought up after the talks were further debated in the round-table meetings. The friendly atmosphere supported possible new collaborations to be established.

References

[1] P. B. Allen and M. Cardona. Phys. Rev. B 27, 4760 (1983). [2] P. Goslar and Sang-ill Choi, Phys. Rev. 150, 529 (1966).
[3] C. Attaccalite, L. Wirtz, M. Lazzeri, F. Mauri, and A. Rubio, Nano Letters 10, 1172 (2010). [4] E. Cannuccia and A. Marini, Phys. Rev. Lett. 107, 255501 (2011).
[5] G. Antonius *et al*, Phys. Rev. Lett. 112, 215501 (2014).
[6] N. Vukmirovic, C. Bruder, and V.M. Stojanovic, Phys. Rev. Lett. 109, 126407 (2012) [7] J. Noffsinger, *et al*, Phys. Rev. Lett. 108, 167402 (2012)
[8] Y. Gillet, M. Giantomassi, and X. Gonze, Phys. Rev. B 88, 094305 (2013)
[9] B. Xu and M.J. Verstraete, Phys. Rev. Lett. 112, 196693 (2014)
[10] A. Eiguren, C. Ambrosch-Draxl, Phys. Rev. Lett. 101, 036402 (2008) [11] R.B. Capaz *et al,* Phys. Rev. Lett. 94, 036801 (2005)
[12] A. Marini, Phys. Rev. Lett. 101, 106405 (2008)
[13] F. Giustino, S. G. Louie, and M. L. Cohen. Phys. Rev. Lett. 105, 265501 (2010) [14] S. Zollner, M. Cardona, and S. Gopalan. Phys. Rev. B 45, 3376 (1992)
[15] A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W. P. Sum Rev. Mod. Phys. 60, 781 (1988)
[16] G. Onida, L. Reining, and A.l Rubio. Rev. Mod. Phys. 74, 601 (2002) [17] S. Baroni *et al,* Rev. Mod. Phys. 73, 515 (2001)
[18] X. Gonze. Phys. Rev. A 52, 1096 (1995)
[19] S. Moser, S *et al,* Phys. Rev. Lett. 110, 196403 (2013)
[20] J. Robert Schrieffer, Theory Of Superconductivity, Advanced Book Program, Perseus Books, 1999

**Outcomes:** As it is clear from the state-of-the-art section, despite the enormous interest in EV interaction and more and more scientists getting involved, this research field is still fragmented. There are many different methods and approaches to tackling different questions. This is inevitably creating a gap between the theoretical and experimental communities. Major advances are expected in the following domains, with associated major obstacles to overcome:

1. Search of a systematic way to perform a perturbative treatment of the EV interaction in a pure ab-initio manner.
2. Problems with a perturbative treatment of the EV interaction beyond the harmonic level: is it suitable and accurate?
3. One of the most used and widely spread assumption is that the EV interaction can yield only minor corrections (of the order of meV) to the electronic levels. As a consequence keeping the atoms frozen in their crystallographic positions generally performs the majority of the simulations of the electronic and optical properties of a wide class of materials. In view of the results presented at the workshop, this is not true. Corrections up to one eV or more are due to EV corrections. This information should be disseminated and also consolidated (e.g. through calculations for dozens of materials).

4. Phonons are harmonic excitations, and the harmonic approximation is known to be valid only for temperatures much lower than the Debye temperature. What happens above this temperature? Are anharmonic corrections always negligible even at, for example, room temperature?
5. Heat transport and thermoelectricity are intrinsically related to EV interaction. How can we quantitatively describe them?
6. The EV coupling can be included in the adiabatic limit, or on the contrary, taking into account dynamical (non-adiabatic) phonon effects. In which situation are those mandatory? Does the theory break down without dynamical effects?
7. For IR-active materials, the electron-phonon interaction diverges in the long wavelength limit. How to cope with this divergence, especially when small or moderate size cells are used in simulations? What is the link with Frohlich hamiltonian?

On one hand, this question has been jointly addressed in the previous section. On the other hand, one should work tightly with experimentalists, and disentangle with them, specifically the harmonic EV effects, the anharmonic EV effects, those EV effects beyond perturbative descriptions, from effects that are not related to EV effects.

**Requirements:** The computational demand (CPU, data storage) placed on EV calculation varies much with the level of theory. At present, there is a wide area of computations (exploratory or production) for which EU Tier1 or even Tier2 machines provide adequate computational resources. Development of software is, at the same time, on-going in the different group. Similarly, the theory development is on-going, and it is clear that more networking (e.g. in the form of the CECAM workshop, but perhaps EU training network) is needed. Concerning training: the field (including software applications) might be ready to organize a training tutorial in 2017. EU funding for a network focused on EV effects might be appropriate. CECAM funding for another workshop, and later, a tutorial might be appropriate.

**Will these developments bring societal benefits?** Ab initio prediction of temperature-dependent properties of materials might be very relevant, especially in the search for new materials thanks to high-throughput computations. Applications are numerous: battery materials (temperature-dependence of voltage), thermoelectric materials, LED materials (electroluminescent materials, as well as phosphors), photovoltaic materials.

# JUJOLS VIII: An European workshop on theoretical approaches of molecular magnetism

Organisers: Jean-Paul Malrieu, Hélène Bolvin
Location: Bages (France)
Date: 1 – 5 June, 2015

**State of the Art:** Areas of current interest include

1. Research of methods based either on wave-function (WFT) or on density functional theories (DFT) for the calculation and prediction of magnetic and spectroscopic properties of organic and inorganic molecules with open-shells; zero-field splitting, magnetic coupling, g factors; consistent treatment of spin-orbit coupling and dynamical correlation, spin decontamination in broken symmetry DFT approaches, spin dependent geometry relaxations and role of the vibrations;

2. Development and applications of effective hamiltonians in term of model hamiltonians, which permit the modelling of the magnetic properties of these molecules. Cases of nearly degenerate states and Jahn-Teller effect, lanthanide and actinide complexes;

3. Application of the above methods to describe magnetic anisotropy, single molecule magnets, LIESST effect (entanglement of quantum qubits, dimers of lanthanide and actinide, collective effects and hysteresis, effect of the application of an static electric field);

4. Addressing of new materials (actinide, organic, graphene nanoflakes);

5. and of new properties (magnetism and conduction, paramagnetic NMR).

**Outcomes:** we have identified the following issues:

- Improvement of the quality of the magnetic orbitals as starting information for the theoretical treatments;

- Need for progresses in the accuracy and domains of applicability of computational methods: Density Functional Theory and Wave Function Methods, correlation, relativistic effects.

These improvements are especially necessary in view of the appearance of new demands concerning heavy metals complexes, hyperfine coupling, single molecule magnets (SMM). Three topics are particularly important

- Collective effects and transport phenomena;
- Dynamical effects;
- Extension from molecules to materials.

The size dependency of the available treatments is a severe limitation, in particular in systems with several open shell atoms. Taking into account the impact of the dynamical

degrees of freedom on the observable of tiny magnitude is also computationally very demanding. The development of observable-dedicated strategies is also desirable.

**Requirements:** This workshop has gathered several groups, which are involved in coding development. (ORCA, MOLCAS, MOLPRO). The series of the Jujols meetings has favoured the emergence of a real collaborative network with exchanges of methods. This results in a dense flux of visiting stays of senior scientists and students.

During this workshop, two pedagogical lectures have been given and the problematic of the description of magnetic systems now takes a significant place in several summer schools in which the participants of this workshop are involved.

The countries of the South part of Europe need fellowships for students and post-docs.

**Will these developments bring societal benefits?** The SMM and quantum computing are considered to be major issues for the development of information storage and manipulation. Some medical application using lanthanide complexes are under development.

## Electronic Structure Library coding workshop: utilities toolbox

Organisers: Fabiano Corsetti, Micael Oliveira, Volker Blum, Arash A. Mostofi
Location: CECAM-HQ, EPFL Lausanne (Switzerland)
Date: 1 – 6 July, 2015

**State of the Art:** The aim of the Electronic Structure Library (ESL) is to create a common repository of high-quality software and data standards in the field of electronic structure simulation, which will facilitate reuse of code, interoperability between different code bases, and development of new methodologies.

At the moment it is widely recognized in the community that these are important, yet unsolved, problems. Most of the large code developments in electronic structure are entirely separate from one another, independently maintaining routines providing overlapping functionalities, and making use of non-universal data formats.

However, there are a few notable exceptions of communal software libraries specific to electronic structure, which have found some success in bridging between different codes, e.g., Wannier90 for maximally localized Wannier functions, and Libxc for exchange and correlation functionals.

**Outcomes:** The electronic structure community faces the challenge of managing codes, which are ever growing in size and complexity. Furthermore, this needs to be done whilst keeping up with developments in hardware (especially for high-performance computing) and coding practices. Notable examples include GPU architectures and standards for manipulating big data.

A growing recognition of the importance of high standards of software development for scientific research (as evidenced by, e.g., the creation of the Software Sustainability Institute in the UK) promises to have an impact on all fields, including electronic structure. This can already be seen in the fact that many of the largest codes based on density-functional theory (DFT) are entering into collaborations with computer scientists from high-performance computing facilities; some are also reserving funding for hiring researchers to work exclusively as scientific software developers, in order to improve aspects of the code such as performance and portability. However, the major challenge in the next few years will be to coordinate the effort between different coding groups in order to avoid wasting resources on replicated and incompatible developments, and thus increase research efficiency in the community as a whole.

There are a number of quite distinct obstacles. From the software development point of view, it is very difficult to predict which hardware evolutions will ultimately prove to be enduring. The strategy of the ESL, therefore, should not be too dependent on coding for particular architectures. Rather, we will try to approach the problem from the perspective of the science to be implemented, and focus on identifying functionalities that can be segregated into self-contained libraries. The aim will therefore be to develop software that helps define a clear programming interface for a particular functionality. If this is successful, it will be extremely helpful for adapting to new architectures as the need arises.

From the organizational point of view, the ESL is entirely based on contributions from the community. Continuing development is therefore crucially dependent on maintaining and developing interest from potential contributors. There has already been a good level of

interest in the project and involvement from a number of well-know DFT codes (SIESTA, FHI-AIMS, ABINIT, Octopus, BigDFT, Quantum ESPRESSO, etc.) Our aim will be to try to increase collaborations with existing codes, as well as bring in researchers not affiliated with any established code base. So far we have found that offering funded places at workshops has worked as a good incentive for attracting developers; even so, researchers' busy schedules has proved to be a difficult problem. As the ESL grows we hope to overcome this obstacle by building strong bridges between difference code bases, which will naturally lead to increased collaboration, and being able to provide concrete examples of software developed for the project that has increased research productivity in the community. This will encourage continued contributions to the ESL.

**Requirements:** CECAM is already hosting the ESL wiki website on one of its servers. However, at the moment there is no expert who can dedicate the required amount of time to technical maintenance. This is a key task that would help the development of the ESL as a whole.

Currently the software itself is hosted on external repositories. It would be important for the future strategy of the ESL to provide dedicated machines for storing code and performing nightly builds and automatic testing. This would allow for a more integrated suite of software, and would help the ESL maintainers to ensure consistent high quality amongst all contributions. Similarly, access to HPC facilities would allow for in-house development and testing of massively parallel code.

Finally, providing training in collaboration with the ESL would be useful for attracting new researchers and giving them the tools to contribute to the projects being developed. Training areas of interest would be software engineering, modern coding practices, data analysis tools, and development for HPC. On top of this, ESL maintainers could benefit from meeting HPC companies and learning about developments in hardware.

The most substantial contribution needed from national and European funding councils is the establishment of specialized Fellowships to support full-time work on these activities from experts in the field. An example of the model to follow would be the Research Software Engineer (RSE) Fellowships from the UK Engineering and Physical Sciences Research Council (EPSRC), established this year. Additionally, internship programmes would be important for supporting short-term software development goals and setting up new coding projects.

**Applications:** The aim of the ESL initiative is to develop the research infrastructure underpinning a huge amount of scientific research, spanning materials physics, physical chemistry and biology, nanotechnology and nano-medicine, Earth science, and more. Indeed, the importance of atomistic modelling from first principles electronic structure is widely recognized, and its applications outside of academia are many. In particular, the opportunity to collaborate with researchers in industry involved in such modelling projects would allow us to develop software that could drive innovation and growth.

It is also important to note that we encourage ESL contributors to make use of permissive free software licenses that are compatible with commercial applications. This means that companies will be able to utilize (although not directly benefit) from our developments; this is a standard collaboration model that has been very successful in other fields and which is generally recognized as beneficial for both parties.

# Flexibility and Disorder in Metal-Organic Frameworks

Organisers: Alain Fuchs , Thomas Bennett, François-Xavier Coudert. Anthony Cheetham
Location: CECAM-FR-MOSER, Chimie ParisTech, Paris (France)
Date: 3 – 6 June, 2015

**State of the art**: The scientific literature of the last few years has seen an exponential growth in the number of nanoporous framework materials reported, with thousands of new metal–organic frameworks (MOFs), covalent organic frameworks (COFs) and molecular framework materials. These novel families of materials open up new horizons in practically all branches of engineering, physics, chemistry, biology, and medicine. Nanoporous materials find numerous applications as selective adsorbents and catalysts, substrates for biosensors and drug delivery, membranes and films in various nanotechnologies, which involve fluids adsorbed or confined to nanoscale pores. Whilst the field is still very young, evidence is emerging of a change in emphasis from the synthesis of new structures to exploring the properties of those already known. In particular, while such nanoporous materials were traditionally seen as crystalline rigid structures, there is growing evidence that large-scale flexibility and long-range disorder are not the exception, but rather the norm, in metal-organic frameworks. Studies published in 2012 have highlighted the existence of very soft deformation modes in MOFs, both experimentally (by Brillouin scattering and nanoindentation) and by quantum chemical calculations. [1,2] Moreover, it was recently demonstrated by high-throughput computational screening of MOFs, based on simple mechanical arguments and models, that it is a much more common feature than previously believed. [3] Though flexibility and disorder are intrinsically related to one another, the latter has been afforded much less attention in the literature. A small but growing number of cases of MOFs lacking the long-range order characteristic of crystalline structures are steadily capturing scientific interest. [4,5] Disorder can range from slight variations from the ideal structure, through to amorphous materials, which have no long-range order. Indeed, a recent proof-of-concept study in the UiO-66 family of materials has shown that structural defects can be introduced and controlled as correlated nanoscale disorder, a targetable and desirable motif in the design of new materials. [6] Finally, we note that both flexibility and disorder have been observed to drastically alter the properties of MOFs, particularly those involving sorption and/or separation of guest molecules. This is of the utmost importance given that MOFs are heralded for these very abilities. Proposals for use in industrial applications are therefore dependent on knowledge and understanding of this behaviour, a unified approach to which has not yet been made.
The complex, non-rigid behaviour of MOFs is now being recognized on a much wider scale. Computational modelling is becoming an extremely valuable tool with which to study such phenomena. We expect that the definition of MOFs as rigid materials will continue to change, and that their dynamic response to external stimuli will take on an increasingly prominent role in research in the field. [7] This broadening of the scope of "flexibility", by looking at a richer array and broader range of stimuli, will make it necessary to revisit the known materials and reassess their potential. We are however, far less advanced in our understanding of disorder, and the techniques available to fully characterize it. Major advances in the characterization of the disordered systems, be it topological, static or dynamic, are required, in order for us to fully understand the potential applications of MOFs. There is also a need to establish some "community standards", or best practices, for evaluating and characterizing disorder. Finally, advances in molecular dynamics simulations, density functional theory and experimental methods are anticipated in the years to come. In particular, the computational modelling of disordered MOFs is still in its infancy. While current studies can reasonably study the

*impact* of disorder and defects on properties and function, [8] there is still a real need for computational work elucidating the fundamental origins of defects in MOFs, their collective behaviour (clustering, correlation and disorder), and the link between chemical structure and nature of the defects.

References
[1] Exceptionally Low Shear Modulus In a Prototypical Imidazole-Based Metal-Organic Framework, Jin-Chong Tan, B. Civalleri, Chung-Cherng Lin, L. Valenzano, R. Galvelis, Po-Fei Chen, T. D. Bennett, C. Mellot-Draznieks, C. M. Zicovich-Wilson, A. K. Cheetham, *Phys. Rev. Lett.* **2012**, 108, 095502
[2] Anisotropic Elastic Properties of Flexible Metal–Organic Frameworks: How Soft Are Soft Porous Crystals?, A. U. Ortiz, A. Boutin, A. H. Fuchs and F.-X. Coudert, *Phys. Rev. Lett.* **2012**, 109, 195502.
[3] On the Flexibility of Metal–Organic Frameworks, L. Sarkisov, R. L. Martin, M. Haranczyk, B. Smit, *J. Am. Chem. Soc.* **2014**, 136, 2228–2231.
[4] Structure and Properties of an Amorphous Metal-Organic Framework, T. D. Bennett, A. L. Goodwin, M. T. Dove, D. A. Keen, M. G. Tucker, E. R. Barney, A. K. Soper, E. G. Bithell, J. C. Tan, A. K. Cheetham, *Phys. Rev. Lett.* **2010**, 104, 115503
[5] Structural disorder in molecular framework materials, A. B. Cairns and A. L. Goodwin, *Chem. Soc. Rev.* **2013**, 42, 4881-4893
[6] Correlated defect nanoregions in a metal–organic framework, M. J. Cliffe, W. Wan, X. Zou, P. A. Chater, A. K. Kleppe, M. G. Tucker, H. Wilhelm, N. P. Funnell, F.-X. Coudert and A. L. Goodwin, *Nature Comm.* **2014**, 5, 4176.
[7] Responsive Metal–Organic Frameworks and Framework Materials: Under Pressure, Taking the Heat, In the Spotlight, With Friends, F. X. Coudert, *Chem. Mater.* **2015**, 27, 1905-1916
[8] Water adsorption in UiO-66: the importance of defects, P. Ghosh, Y. J. Colón and R. Q. Snurr, *Chem. Commun.* **2014**,50, 11329-11331

**Outcomes:** The workshop identified as the major obstacle to success the lack of a joint experimental-computational approach to study flexibility and disorder in MOFs. The rich behaviour of these multifunctional materials requires large-scale collaborations with access to several characterization techniques, both on the experimental side (*in situ* diffraction, adsorption, mechanical testing, etc.) and on the theoretical side (quantum chemistry calculations, *ab initio* molecular dynamics, force field-based MD or Monte Carlo, etc.). The workshop has made some progress in addressing this, in bringing together two sets of the community in a fertile, informal environment in which collaborations can be set up and discussed.

**Community needs:** In terms of experimental infrastructure, the use of *in situ* diffraction of both X-rays and neutrons is the state of the art for the study of stimuli-responsive materials, and is widely used in the literature. One element highlighted in the discussion is the need to greater access to advanced characterization of disordered MOFs, including in particular cryo-electron microscopy (cryo-TEM). In terms of computational work, high-throughput screening of materials requires powerful computational facilities as well as data storage and public interfaces (web-based APIs). These requirements will increase when flexibility and disorder are added to such databases. It is important to note that there is currently heavy duplication of efforts in this area, with several overlapping databases hosted in various academic institutions (MOFOMICS at Princeton, Hypothetical MOF at Northwestern) and on commercial services (CoRE-MOF on github, Materials Project on their own servers). This is an area where some standardization effort and central hosting would be beneficial for the community.

**Funding:** The community possesses sufficient tools and expertise to approach the problem, though, as is often the case, suffers from a shortage of available researchers. A common problem, across international boundaries and disciplines, is that numerous smaller 'pots' of money are available to set up collaborations, though there is no money

available for researchers to work on them. Though the EU has funded some flagship programs in this area, they are not all-encompassing and typically highlight very specific areas of the field. This mostly application-driven top-down funding can lead to lack of resources for fundamental studies required to develop novel methodologies.

**Will these developments bring societal benefits?** The work proposed is of great impact within the MOF field, in helping to realise the plethora of applications currently proposed. Such potential applications include drug delivery, harmful waste storage, $CO_2$ capture and industrial waste separations, and as such the workshop will play a role in achieving the huge potential societal impact of MOFs.

# Future technologies in automated atomistic simulations

Organisers: N. Marzari, D. Morgan, K. Persson, C. Draxl
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 8 – 10 June 2015

**State of the Art:** The ability to automate atomic scale simulations (AAS), particularly ab initio calculations that can treat any set of elements, is enhancing our ability generate materials and chemistry data by orders of magnitude. Presently there are significant efforts in many groups to develop the methods, tools, and databases to support and build on AAS, including workflow tools, materials data databases, benchmark databases, machine learning and data mining tools, and more robust simulation methods. This ecosystem is transforming the amount and nature of available materials data, creating new opportunities in materials design, and changing the scale and way many simulations are performed. However, there is little communication between these efforts, limited awareness of other group activities, and few established best practices. A number of large databases and data management tools have recently become available or will become available soon. These will provide many new options for using and sharing data for researchers. Furthermore, a number of high-throughput atomistic modelling studies are being completed or well underway that demonstrate the exceptional power of these techniques. These studies are expected to attract significant attention and increased participation in the field. It is also clear from these studies that machine learning techniques will play and increasing role in the field and that a strong cross-disciplinary connection with computer science and statistics will be valuable. Increasingly, the tools in the field are being developed using the python programming language, which enables easy extension. However, perl, java, C++, and other languages are still being used and must be included in the community. In general open source codes and data are being promoted, which is extremely valuable for the community.

**Outcomes:** A detailed analysis of the factors hindering development in the field was performed. The main points identified are:

1. Redundancy: There is significant duplication of tools, infrastructure, and services being developed among different groups. This is partly to be expected and perhaps even desirable so that best practices and approaches can be developed freely. However, it was felt that greater awareness of completed and ongoing efforts would benefit the community. To help establish a stronger community some of the participants expressed strong interest in a regular workshop series, perhaps incorporating CECAM, Gordon conferences, or satellite meetings or symposia within larger meetings (e.g., the MRS biannual meetings). Additionally, the already strong commitment to open source data and tools, including software support and documentation, should be encouraged to enable software and data reuse and integration.
2. Code heterogeneity: It is clear that significant extra effort is required to develop tools and databases due the their being multiple atomistic simulation codes. As unifying the simulation codes is likely impractical and undesirable, some in the community are adopting modular plug-in approaches in their code development to allow maximal reusability given this heterogeneity. This modular plug-in approach should be encouraged for the community in general when practical.
3. Interface with non-expert users: The explosion of new tools and databases is likely to make it difficult for non-experts, e.g., those not familiar with simulation tools and/or those not familiar with the community, to know where to go for the services they need

(e.g., where to post data, what data to trust for what level of accuracy, where to find tools for automating their workflows, etc.). This issue is somewhat inevitable as the field develops but its negative impact can be minimized by promoting a strong focus on documentation, error estimation, and communication to non-experts.

4. Data identification: There is a significant danger of data not being inter-relatable between databases (e.g., how do I know a structure is the same in two databases so I can compare its calculated properties?). This issue could be addressed by a community effort to establish a minimum set of information to describe each system. This could be specified in a JSON identifier, perhaps building on the existing CIF structure format. In addition, a meta-database of what is available in different databases might be valuable, particularly if it can enable search across multiple databases from a single interface. Such a structure identifier system would be particularly useful in enabling searching across databases. The data identification problem can be further reduced if databases track provinence of data, e.g., so a final calculated structural property can be related back to the original structural coordinates, which might have come from the ICSD or a publication.
5. Success stories: There is a concern that this nascent field needs to show some clear success stories to maintain its momentum. In particular, these successes will most likely focus on the impact of the tools and data on the discovery of new materials. There is already a web site from Psi-k with a list of such stories (Highlight 122 from http://www.psi-k.org/newsletters.shtml ) and extending it as new results are obtained would be valuable for the field. Some of these stories might be promoted through blogging high-impact papers, and Jan Jenson's blog "Computational Chemistry Highlights" was suggested as a potential location for such commentary.
6. Connection to industry: There is an exciting opportunity for the new capabilities of automated atomistic simulation to impact industry. To realize this goal is it essential to invite industry participants to workshops and conferences. The industry representatives at this workshop were very engaged and their perspective highly valued.

Overall, many of the above challenges can be addressed by establishing a strong community that communicates well and frequently. Specific actions taken by the workshop to promote this include

1. We have asked and obtained support of Jan Jenson to make use of his blog "Computational Chemistry Highlights" (many thanks to Alan Apuru-Guzik for his help in developing this). Conference participants have already agreed to post highlights of key papers in the field to try to develop this as resource for informing the community of topical developments.
2. We have established a Twitter hashtag #compmat to tweet about important developments.
3. We have established a twitter account @compmater to help support dissemination via Twitter.

**Networking & training:** We have initiated planning of an annual workshop in this general field of computational materials/automated atomistic simulation/materials informatics.

**Community needs:** There are extensive needs of this community for computing, cloud resources, robust simulation tools, etc. However, in most cases these resources are available and/or developing in an appropriate way through existing programs, provided the funding for general computational materials/chemistry continues to be strong. One area where a change from the present path would be beneficial is the establishment of more high-throughput computing resources at large-scale computing centers. These centers have tended to focus on developing tools for tightly coupled parallel simulation

of very large systems (e.g., exoscale computing) rather than simulation of millions of smaller systems. More support for the latter would accelerate what can be done with automated atomistic simulations. In addition, there was significant discussion of the need to improve training for computational scientists, and in particular to help them develop the interdisciplinary skills from computer science, statistics, and data science that the new automated atomistic simulations field will require. Increase in funding for cyber-infrastructure development is also key. This is sometimes underappreciated, but is essential to enabling the new software tools and databases required by this field. In addition, establishing stronger requirements for governments funded activities to be open source (both software and data) will help drive the field forward by enabling easier (re)use of the best technologies.

**Will the results of these developments be of societal benefit?** The advent of automated atomistic simulations and the associated databases and informatics developments are of potentially enormous economic impact. These approaches enable generation of materials data and exploration of materials space at an unprecedented rate. It is not unreasonable to say that for many basic materials properties (alloying energies, elastic constants, band gaps, diffusion coefficients, defect properties, etc) we can expect to generate over the next ~10 years calculated databases with accuracies comparable to experiment that exceed the collected data from all of history by ordered of magnitude. These approaches will therefore support accelerated discovery and optimization of new materials for many materials centric applications critical to society, from solar energy to new drug delivery methods.

# Industrial Challenges of Crystallization, Nucleation, and Solubility: Perspectives from Industry, Experiment and Simulation

Organisers: Donal Mac Kernan, Brian Glennon Baron, Ake Rasmuson, Zaworotko Michael, Jean-Christophe Jacquier
Location: CECAM-IRL, University College Dublin (Ireland)
Date: 9 – 12 June, 2015

**State of the Art:** Nucleation and solubility are playing an increasingly important role in fundamental research, and industry where it is exploited in for example, the development, processing and delivery of pharmaceutical active ingredients (API). Despite the enormous industrial interest and a century of efforts, the most practical nucleation theories make simplifying assumptions that prevent them from being quantitative. For several decades, molecular simulation approaches have been pursued as routes to quantitative predictions of rates, metastable zone widths, polymorph selectivity as functions of key variables like temperature, composition, density, additive concentrations, etc. Early advances pertained to idealized single-component nucleation processes, e.g. crystallization of Lennard-Jones fluids and hard spheres. Recently, some key methodological advances have dramatically advanced the state-of-the-art. These advances include:

- New order parameters and rational ways of designing order parameters for arbitrarily complex molecular and ionic crystals;
- Improvements in rare events methods for sampling free energies and computing rates ;
  Improvements in methods to compute phase diagrams, particularly polymorph coexistence lines:
- Improved understanding of two-step nucleation mechanism and the role of metastable fluid phases, gel-phases and slow diffusion effects;
- Simulation methods that control and/or correct for chemical potential variations upon solute precipitate nucleation.

These advances have recently enabled studies of complex molecular crystal nucleation processes from the melt as well as crystal nucleation from multicomponent solutions. The advances in solute precipitate nucleation in particular make it possible to simulate problems of direct interest to the pharmaceuticals industry, to materials science, to geochemistry, and biological/biomimetic self-assembly. Some of the key scientific and processing issues faced by the pharmaceutical and food for health industries involve molecular scale mechanisms, such as the purification of active pharmaceutical ingredients (API) & nutraceuticals (NU) through crystallization [22], solubility of API & NU and delivery mechanisms including:

- Their co-crystal embedding with physiological safe agents;
- The use of anti-solvents to promote nucleation/crystallization;
- Continuous rather than traditional batch manufacturing, where frequently the system is kept in a state of non-equilibrium through a flow reactor or the application of shear stress;
- Continuous measurement using methods such as real-time infrared and Raman spectroscopy and control of the system through accessible control parameters such as effective temperature, density, and flow rate.

Many of these methods can be and frequently are combined. One particular attraction of continuous manufacturing is that it gives in principle both greater control of the final

product (for example crystal size distribution important for bio-availability), and allows demand driven cost efficient production of tailored quantities of the desired product (rather than fixed batch sizes).

Co-crystals are emerging as a very important technology allowing the modification of solubility for example of target API's, nutraceuticals, and other agents through non-covalent bonding with appropriately selected conformers. Key findings concerning pharmaceutical co-crystallization have thus far tended to concentrate on gaining a better understanding of co-crystals from the materials design and structural perspectives. The role of co-crystals as potential APIs suitable for use in drug products is less well delineated and, to the best of our knowledge, common Pharmacokinetics (PK) parameters are only reported for 64 co-crystals. Nevertheless, the high potential of co-crystals to fine-tune physiochemical and PK properties for a given API is becoming crystal clear and it seems that co-crystals are poised to become a platform technology that can be adapted across the full range of drug molecules and Biopharmaceutics Classification System (BCS) classes.

At a theoretical level, the key questions that the community remains a better understanding of nucleation, and the accurate calculation of the lowest energy polymorph. Solubility is also a very important issue. New processing approaches, in particular the use of continuous processing is having a very big impact in industry and open up a new set of theoretical challenges for simulation.

**Outcomes:** The discussions highlighted the difficulties in the field, and need for more work on several points as follows

1. The difficulty of ensuring that at a given temperature the lowest free energy polymorph is identified missing it can result in catastrophic over-estimates of solubility in physiological conditions.
2. There is still a very high discrepancy between nucleation rates extracted from theoretical/simulated nucleation and those of experiment. An understanding of this discrepancy is still lacking, and very desirable, both from a theoretical, and experimental/industrial perspective.
3. The level of control of crystal size distribution that can now be achieved using non-equilibrium (i.e. driven) experimental methods is very impressive, and is already having an impact on moving industry away from traditional bulk pharmaceutical processing, notwithstanding the difficulties so entailed of ensuring high degree of pharmaceutical purity.
4. The role of solubility is central to much of pharmaceutical processing and bioavailability particularly under different ambient conditions of temperature, pH and solvent mixtures including the use of anti- solvents. Improving our ability to predict solubility for novel compounds, and in novel solvents would be of great interest to industry.
5. There remain some obstacles, including (1) there is limited PK information in the public domain and the effects of co-crystallization upon BCS class I, III and IV drug substances remain understudied; and, (2) a far greater range of coformer types remain to be subjected to systematic study, especially ionic co-crystals.

**Obstacles:** At a theoretical/simulation level, one of the key problems remains identifying the relevant order parameters required to describe the nucleation process. Although theoretical advances such as minimum free energy path methods have certainly helped clarify many issues such methods still require (a) a clear identification of the relevant initial and final end points (e.g. the chemical reactants and products) and (b) a relevant

order parameter space embedding the ultimate reaction path. In terms of polymorph stability, there are several difficulties including: from how to simply search a very complex energy landscape efficiently; (c) how to compute at an ab-initio level the energies of polymorphs accurately and efficiently-for instance how important are kinetic and polarization effects; (d) how do build an efficient force fields for molecular dynamics simulations, including polarization. (d) how to simulate efficiently systems under constant pH; (e) how to model continuous processing conditions. While there have been significant advances on most of these questions, there remains a great amount of work to do.
The use of data-mining/machine learning methods combined with sophisticated sampling methods may help in extracting useful order parameters - although it is unlikely to be a panacea. There has been quite a lot of progress regarding points (c) and (d), so it is more a question of how long it will take for these developments to be fully included into the set of standard methodological used by the community - and into main stream molecular mechanics and ab-initio codes. Continuous processing is fundamentally a non-equilibrium approach - and offers a great theoretical challenge to simulation, albeit with great industrial relevance. A key issue for the latter is also how it can be combined with methods for insuring a high degree of purity of the final pharmaceutical product

**Infrastructure requirements:** The key infrastructure needs here are not so much more powerful machines, but rather better force-fields, better ab-initio method, better approximation methods, more systematic approaches to extracting order parameters. A standardized means to compare the results from different approaches, and a database for all of the above would be very useful.

**Will these developments bring societal benefits?** The pharmaceutical industry is one of the largest creators of wealth in many of our economies. Were CECAM and simulation able to enhance significantly processing and API discovery, it would be significant economically, and in increased health due to the production of better drugs, and improved methods of delivery. It is also important to stress that the pharmaceutical industry is very close to the advanced food and food for health industry. Indeed, there is from a simulation perspective a great synergy between the two. A number of the organizers are working on developing some of these themes currently.

# Molecular Quantum Dynamics Methods: Benchmarks and State of the Art

Organisers: Federica Agostini, Sara Bonella, Ivano Tavernelli, Graham Worth, Jiri Vanicek
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 15 – 19 June, 2015

**State of the art**: Simulating the exact quantum dynamics of systems of interacting particles is presently a task beyond reach but for the smallest systems, as the numerical cost for solving the time-dependent Schrödinger equation scales exponentially with the number of degrees of freedom. Consequently, considerable effort is devoted to developing approximate algorithms to reduce the computational cost of quantum dynamical properties with an acceptable, and controlled, loss of accuracy. However, due to their intrinsic differences in the way the fundamental equations of motion are approximated the direct comparison of methods is often very difficult and in many cases unsatisfactory. Therefore, while the available approaches span a wide range of applications and formal frameworks, common goals and unifying theoretical grounds are somehow missing. Different levels of possible approximations involved in practical developments lead to a natural classification of methods in this field. There are techniques, such as multiconfiguration time-dependent Hartree (MCTDH), addressing the full quantum nature of a given set of degrees of freedom. Semiclassical methods, based on the path integral formulation of quantum mechanics, can instead be combined with a representation in terms of classical trajectories via a stationary-phase approximation and still account for quantum effects such as interference or tunneling. Moreover, these methods can be naturally combined with on-the-fly ab initio evaluation of the energies, forces, and couplings. There are also different flavors of mixed quantum-classical approaches, either considering the density matrix as basic variable (as in the partial Wigner representation of the Liouville-von Neumann equation or in the linearized version of the path integral representation of thermal correlation functions), or based on the wave function formalism (as in trajectory surface hopping, non-adiabatic Bohmian dynamics and the quantum-classical treatment of the factorized form of the electron-nuclear wave function). Finally, there exist also several successful heuristic approximate methods for dynamics at finite temperature, e.g., the centroid and ring polymer molecular dynamics. A few attempts have been reported, whose goal was analysing the connections among those methods, comparing their shortcomings and advantages, understanding the restrictions of each approach and developing schemes with more general validity. However, it seems that these have been isolated examples, limited to only some of the methods listed above. An exhaustive and comprehensive analysis should instead include a broader sample of the state-of-the-art methods, with well-defined benchmarks to assess merits and limitations of the different approaches and to identify the different areas of applicability. These issues were discussed in the workshop.

**Outcomes:** Two main strategies to address the problem of the validation and verification of the different solutions to quantum dynamics were identified. The first deals with the design of model systems that can be solved with the different approaches. The purpose of this test is not to judge the quality of the different methods, but to identify strengths and weaknesses of the approaches to design possible improvements. Having a set of models for which the 'exact' solution is available will provide clear guidelines for the future development of the field. The second strategy is based on the design of a class of realistic model systems that constitute a valid challenge for all available solutions to quantum dynamics. In this case, the aim is to design problems that each community can try to solve. The quality of the different solutions can be assessed through crosschecking or by direct

comparison with available experimental results. This challenge will provide important information about the quality of the results that can be obtained from the different approaches. We hope in this way to be able to provide guidelines for the selection of the most appropriate method for the evaluation of a desired observable (spectral function, quantum yield, nonadiabatic ratio, time scale for a given reaction). Finally, the comparison among the different solutions will allow the development of better theories and algorithms and the creation of synergies among the different approaches.
Some key problems in quantum dynamical simulations were also highlighted. First, the need for accurate potential energy surfaces. This is one of the main bottlenecks for quantum dynamics simulations. Traditional methods require global surfaces and at present it is not possible to calculate surfaces for systems with more than a few (3-4) atoms. In part this is down to the large number of computationally intensive electronic structure calculations required to cover the appropriate configuration space. A second problem is finding suitable fitting functions. Model potentials with suitable parameterisation are one way forward, but while these may be powerful in providing insight into dynamical behaviour, they are often limited in scope. For this reason the present state-of-the-art is developing "direct dynamics" methods in which the potential function is calculated on-the-fly only when required. The potential surface problem is exacerbated for non-adiabatic problems in which excited-state surfaces and non-adiabatic couplings are also required. New electronic structure methods are needed to cope with these problems in a straightforward way. Second, the field lacks user-friendly codes. Quantum dynamics is still an emerging field with much work presently on algorithmic development. Most software is still "one group" codes, not easy to use by someone outside a narrow circle of experts. For the development of the field it is important to create codes so that they can be broadly used, e.g. by experimentalists. This will move the field to a more main-stream discipline able to support a wider range of users and so gain validity. There is indeed some movement in this direction, especially surface hopping codes (CPMD, SHARC, Newton-X), and the MCTDH based Quantics package, but more developments are required.

**Community needs:** The community would benefit from a centralized web platform providing access to quantum dynamics software, potential energy surfaces, benchmark results, and description of the underlying algorithms. The first step has already been taken to achieve this goal—namely, a Wikimedia page (qdyn.cecam.org) has been set up at CECAM. The main purpose of this platform is to build a database of quantum dynamics algorithms: the contributors would either upload their software or provide a link to an external web site, from which this software is accessible. The format of the Wikimedia page allows each contributor to include a description of the algorithm, instructions for using the software, and explanation of provided examples. In addition to a software repository, the web platform will serve as a library of potential energy surfaces and benchmark results obtained for various systems by various methods. Maintaining this web platform will require a certain amount of data storage facilities, particularly for the collection of benchmark results. Yet, this effort will certainly pay off in facilitating the exchange of new ideas and accelerating progress.

**Funding:** The most common funding source so far have been COST actions. As the industrial interest towards quantum dynamics increases, however, new funding possibilities appear. For example, one of the Work-Packages of the recently established H2020 Centre of Excellence for Computation E-CAM is devoted to quantum dynamics. Public private partnerships based on E-CAM should be explored further.

**Will these developments bring societal benefits?** Quantum dynamical effects are increasingly important in many industrial sectors including hardware design (coherence and interference effects), pharmaceutics (radiation damage), and energy production

(when light is used to induce quantum physical or chemical transformations for application, for example, to solar cells). Future increase in computational power and method improvements will make it possible to extend the methods discussed in this workshop towards technologically and experimentally relevant scales and contribute to drive innovation.

# Atomistic simulations in Earth Sciences

Organisers: Rodolphe Vuilleumier, Guillaume Ferlat, Marco Saitta
Location: CECAM-FR-MOSER, Paris (France)
Date: 17 – 19 June, 2015

**State of the Art:** Atomic simulations ranging from first-principle simulations to empirical force fields have been successfully applied to a variety of systems relevant for the study of geophysical and geochemical systems. These simulations have been used to study the viscosity of high pressure fluids relevant to the deep Earth interior, the dielectric constant of high pressure water, the thermoconductivity of high pressure oxides, transport and solubility of CO2, trace element partitioning, and metal complexation in supercritical water. The simulation techniques involved range from first principle molecular dynamics simulations, that takes into account the chemical complexity and the extremes of temperature and pressure to empirical force fields and multiscale techniques. These are often combined with thermodynamic integration and other free energy sampling methods that have allowed us to establish the phase diagrams of these systems by simulation

From these early successes atomic simulation is beginning to be recognized by the Earth and planetary science community. However, a number of significant challenges remain.

**Outcomes:** During the course of the workshop, the following topics were discussed

1) Estimation of the thermodynamic and transport properties
Molecular simulation has reached a stage where computing thermodynamic quantities such as free energy differences is now the ultimate goal. This is of particular relevance for Earth sciences because these quantities are often missing from experiment, because of the extreme conditions encountered, but are key in modelling the geophysical and geochemical processes.

During the workshop, novel methods were reported for the evaluation and the design of quantities such as: equation of states of melts, element and isotope partitioning between melt and solids, phase stability, chemical potentials, conductivity and viscosity, transport properties, pKa's and charge state of surfaces.

Many of these quantities involve different phases, as they are either intrinsically interfacial quantities or free energy differences between two phases. During, the workshop it was deemed necessary to have a consistent description of these two phases. This may be the choice of a density functional for *ab-initio* calculations that are performed both for the solid phase and the liquid phase. More generally the approximation and model for free energy evaluation should be consistent; this has been illustrated in the calculation of chemical potentials and element or isotope fractionation.

2) Force-field parameterization from *ab-initio* simulations
As for all atomistic simulations, force fields are an important issue. There are a few aspects that are particular to Earth sciences. These relate to the extreme conditions that are encountered and the wide range of compositions. *Ab-initio* simulations based on density functional theory allow for the description of this variability but are still too costly for systematically sampling thermodynamic conditions and chemical composition ranges. Two trends have been observed: building force fields based on the *ab-initio* data and the new development of polarizable and versatile force fields. However, there is still a strong

need for the development of reactive force fields to describe the effect of volatiles like water and carbon dioxide in silicate melts.

3) Multi-scale approaches

Many of the materials encountered in Earth science are heterogeneous and the thermodynamic properties of their constituents in homogeneous situations is not sufficient to describe the overall material properties. Techniques are now available to describe matter at the mesoscopic scale, such as the lattice Boltzmann technique for the description of transport properties. However, an issue was raised about the correct description of the heterogeneities. For clays for example, random packing and orientation of platelets is based on electron microscopy experiments. When experiments are not available, again because of the extreme conditions under consideration, atomistic simulations can again play a role. This has been illustrated for the description of microcrystallinity and the coexistence of solid phases with small fraction of melts.

# Multi-scale modelling of matter under extreme irradiation

Organisers: J. Kohanoff, T. Apostolova, S. Fahy and A. Rivera
Location: CECAM-IRL, University College Dublin (Ireland)
Date: 17 – 19 June 2015

**State of the Art:** Intense electronic excitation produces a myriad of processes: ionization, carrier scattering, carrier diffusion, exciton formation, Auger recombination, non-radiative exciton decay, followed by thermal processes such as ablation, heat diffusion and atomic rearrangement [1]. In order to obtain a better insight into the underlying physical phenomena we need a combined theoretical and experimental approach to radiation-matter interaction in the regime dominated by strong electronic excitations. In this field we can consider a variety of radiation sources such as ultra-intense laser beams, swift heavy ion irradiation, highly charged ions and high flux ion-, X-ray- or electron-pulses. Although a full description is still missing, efforts to follow the electronic evolution and its subsequent coupling to the atomic lattice are underway, mainly within the framework of intense fs-laser irradiation, through elaborate kinetic models, semi-classical Monte Carlo approaches and hydrodynamic codes [2,3].

A more sophisticated tool for the interpretation of extreme irradiation experiments requires developing a detailed quantum-mechanical model for the description of electron dynamics ingeniously combined with appropriate models to describe lattice evolution. This problem can only be tackled by means of a multi-scale approach combining a variety of computational methods. Figure 1 (below) shows schematically the multi-scale paradigm. Note that the methods described in boxes are matured and successfully applied to describe some problems, in principle, only restricted to the nature of the method (approaches, applicability, time and space scales) and to the available computational resources. On the other hand, the lines connecting the boxes represent coupling strategies to extend the physical description of an excited material beyond the limitations of a given method. A good example, thoroughly discussed during the workshop, is the electron-phonon coupling. Serious efforts are being carried out to describe it in detail within *ab initio* formalisms. On the other hand, phenomenological two-temperature models consider it by coupling the rate equations for atoms and electrons by means of an electron-phonon constant. Despite its simplicity, a number of experiments can be successfully described in this way. In the middle, the efforts to include electron-phonon effects in molecular dynamics turn out interesting.
A successful multi-scale solution should be able to reliably describe irradiation-induced nonlinear effects including transient optical properties (i.e. opacity), reversible and irreversible modifications of physical properties (e.g. refraction index), particle emission, plume formation, ultra-fast opto-acoustic processes, optically induced phase transformations, and lattice heating and cooling phenomena. Along with the predictive design of semiconductor lasers, the developed codes could be applied to the design of detectors for high-energy physics and space satellites, for optical elements in high power lasers, for developing new applications in laser processing of transparent materials and for tailoring the properties of materials with the aid of controlled irradiation tools. Coherent control of optically-excited phase transformations and photo-catalytic processes requires a detailed knowledge of how atomic forces in highly excited materials depend on the incident radiation and on the coherent and incoherent electron-phonon coupling [4,5].

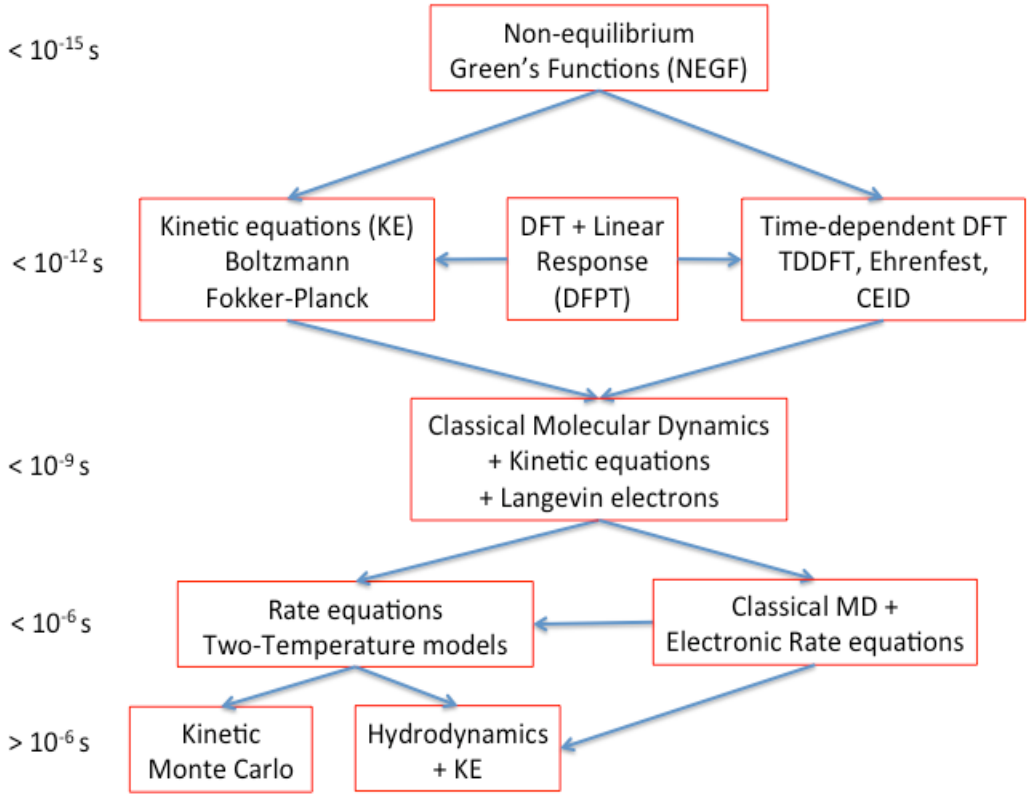



Figure 1. Schematic representation of the multi-scale paradigm applied to intense electronic excitation of materials. The boxes represent different methods applied to the study of highly excited systems. The lines connecting boxes constitute the essence of the multi-scale paradigm since they represent strategies to extend the capabilities of one particular method by coupling it to another method.

Significant progress in this field has taken place in the past few years. However, the efforts have not been conducted towards the multi-scale computational development as evidenced during the workshop. Such a development requires a high degree of cooperation between different experimental and theoretical groups, from the materials and plasma communities. This workshop is one of the first coordinated efforts aimed at bringing together all these communities: theoretical groups working on NEGF, TDDFT, perturbation theory methods, kinetic models, Monte Carlo codes, Molecular Dynamics, Radiation Hydro-Dynamics and experimentalists working with intense lasers, swift ions and X-rays.

**Outcomes:** A general multi-scale solution applicable to any situation is probably unfeasible. Moreover, it is not that desirable either, because it makes it difficult to exploit the specific features of a problem, e.g. laser irradiation is quite homogeneous in space, while swift ion irradiation is localized along a track. We advocate then that this field has to be attacked in terms of specific problem-oriented solutions. Different strategies will be determined by the relevant processes at play, by their time scales, the competition and synergy between processes and, not less importantly, by computational cost considerations. During the deliberations, we discussed a number of relevant cases where intense electronic excitation assumes a prominent role. We also analysed a variety of strategies towards multi-scale simulations. The interesting cases identified were: hadron therapy; generation of high harmonics; diamond detectors for high-energy particles;

transducers; materials modification with swift ions and intense lasers; silica based materials and devices for nuclear environments; model validation by comparison with experiment.

The field of intense electronic excitation has some unique features. It is necessary to understand these to estimate what the progress will be in the next 2-3 years. Multi-scale simulations of highly excited matter require the coupling of different methods, which cover different time and space scales. The individual methods are relatively mature and have shown success when applied to cases within their domain. The major difficulty arises just when trying to couple them. As mentioned above, a general solution is not expected (even with unlimited access to computational resources). The goal is to apply the multi-scale paradigm to selected cases of interest.

When the solution to a problem depends on different coupled codes that overlap in time and scale then the probability of success greatly increases. When the fine details of the physical evolution at a certain scale do not have a strong influence on the time of interest, the multi-scale development is easier to implement. In general, the fewer the number of codes the easier the implementation of the multiscale approach.

For instance, nowadays permanent modification of materials is described to a large extent through phenomenological models, which depend on a number of relatively arbitrary parameters, often questioned by theoreticians (as pointed out during the discussions of this workshop). Consistent multi-scale approaches are being explored to obtain a detailed quantitative description of the processes. Since the electron and lattice systems show very different characteristic times in most systems, the final fate of the highly excited material can be described, e.g., by a quantum kinetic code taking into account the most relevant processes in the electron system coupled to an atomistic code. The energy transfer from the electron to the lattice system generally occurs in the sub-picosecond range and the subsequent atomic evolution controls the final fate of the material in the range of tens to hundreds of picoseconds. As a result, coupling both codes is in most cases feasible, since very fine details in the electron system do not significantly affect the subsequent lattice evolution. If the details of energy transfer through the electron-phonon coupling play a significant role, or the time scales for electron-electron and electron-phonon processes are not clearly separated, as in the case of laser excitation of Bismuth, special care needs be taken.

In the future, we expect an important advance, with application to several of the cases mentioned above; in particular, those involving permanent damage, such as hadron therapy and materials modification. The cases in which transient effects play a role (ultra-fast optics, detectors, transducers) will unavoidably require a special effort to extend the capabilities of the existing sophisticated TDDFT codes to the picosecond level through coupling to faster codes. Important progress is expected since coupling methodologies are already under investigation and will continue to be studied and broadened, in part as a result of the workshop

References

[1] S. K. Sundaram and E. Mazur, Nature Mater. 1, 217 (2002).
[2] D. Huang, P. M. Alsing, T. Apostolova, and D. A. Cardimona, Phys. Rev. B 71, 045204 (2005).
[3] N. A. Medvedev, A. E. Volkov, N. S. Shcheblanov, and B. Rethfeld, Phys. Rev. B 82, 125425 (2010).
[4] A. Cavalleri et al., Phys. Rev. B 70, 161102 (2004).
[5] B. Arnaud and Y. Giret, Phys. Rev. Lett. 110, 016405 (2013).

# Lattice Gauge Theory Simulations Beyond the Standard Model of Particle Physics

Organisers: B. Svetitsky, M.P. Lombardo, K. Rummukainen, T. Volansky
Location: CECAM-ISR, Tel Aviv University (Israel)
Date: 22 – 26 June, 2015

**State of the Art:** The recent discovery of a light Higgs boson completes the Standard Model of particle physics, as proposed and elaborated over the last fifty years. A consistent picture of the dynamics of particle physics now describes a basic set of particles—quarks, leptons, and gauge bosons. At first glance, the simplest mechanism for giving them mass via spontaneous symmetry breaking has now been confirmed. The Higgs itself, however, presents us with serious difficulties stemming from a conceptual gap in the Standard Model. The Standard Model demands extension. Such extensions can be tested, to a degree, by future runs at the Large Hadron Collider and at future accelerators. Phenomenologists have been constructing extensions of the Standard Model in order to deal with the difficulties posed by the Higgs, as well as all the other conundrums of the model. Usually they have been limited by the necessity of analytical understanding of their models, meaning that the models must contain small parameters for approximation near a well-understood theory. Any non-perturbative (non-supersymmetric) physics is often assumed to result from some rescaling of quantum chromodynamics (QCD), regarded as a well-understood theory. At the same time, lattice gauge theorists have been applying nonperturbative methods honed for QCD to study an ever-wider variety of gauge theories. These have sometimes been chosen in order to explore certain properties in convenient theories that may not have direct relevance to the real world. This work invariably involves supercomputer simulations on a very large scale.

**Outcomes:** The workshop was successful in its purpose of bringing together lattice theorists and phenomenologists who work on a variety of approaches to physics beyond the Standard Model (abbreviated as BSM). Its main conclusions are reported in the following.

1. Composite Higgs theories are of great current phenomenological interest. Generally they are written as effective field theories, such as higher-dimensional theories; or as non-linear sigma models, generalizing the chiral lagrangian of QCD. The sigma model is only an effective low-energy description with the correct symmetry properties. A fuller understanding of the dynamics requires an underlying, ultraviolet-complete theory that gives rise to the sigma model at low energies. Some sigma models can be related to ultraviolet-complete asymptotically free gauge theories, generalizations of QCD, which can give first-principles derivations of the constants in the effective theory. This is where lattice simulations come into play. Several families of such theories were discussed in talks at the workshop. Many aspects of such theories, from the basic symmetry breaking pattern to observable violations of flavour conservation, can be determined by lattice simulations.
2. A common theme of many composite-Higgs models is a partially composite top quark. Since it is the only quark whose mass is comparable to the electroweak scale, it is believed that the top quark may play a special role in the electroweak symmetry breaking mechanism. Talks at the workshop addressed this topic both from its phenomenological aspects and from the perspective of its study on the lattice within ultraviolet-complete theories.

3. Technicolor is an older idea than Composite Higgs, and in view of the discovery of the Higgs boson, technicolor theories are being studied on the lattice in an effort to see whether they do have a light scalar particle that can be interpreted as the Higgs boson. Several talks by lattice theorists presented tentative and conflicting results on this issue, as well as the related issue of deciding where specific gauge theories lie relative to the boundary between confinement and conformality.
4. One requirement of a realistic model is that it supplies a candidate particle that can make up the dark matter of the Universe. Many Composite Higgs and technicolor models have such particles, but it might be difficult to reconcile the abundance or detectability of the dark matter with current experimental data. An alternative approach is to begin with a simpler theory that contains a hidden sector of strongly self-interacting dark matter particles, and constrain the interactions with Standard Model particles to agree with observation. This, then, is also a source of theories subject to lattice simulations.

Other discussions included the lattice simulation of gauge theories in more than four dimensions, as well as of theories that are supersymmetric in the continuum limit. Both these approaches, of great interest to BSM model builders, are still in their computational infancy.
Lattice theorists came away from the workshop with a detailed picture of what information phenomenologists need to obtain from lattice computations. The phenomenologists learned about lattice techniques and the difficulties inherent in the large-scale computations being carried out.

**Will these developments bring societal benefits?** There is considerable European investment in computational resources made available to lattice gauge theory simulations, through PRACE and other ventures. The bulk of this effort is in simulations of quantum chromodynamics, with an eye towards supplying precise calculations to compare with experiments on the precision frontier. While this is a traditional focus of lattice gauge theory computations, there is increasing awareness of the importance of calculations for theories that take us beyond the Standard Model. Facilities in the United States that were created for lattice QCD calculations, such as the USQCD collaboration, have been allocating sizable portions of their computer resources to BSM projects. In both the United States and Europe, large lattice BSM projects have received large grants of computer time at universities and national laboratories. The Large Hadron Collider is producing data concerning production and decay modes of the Higgs boson that may point towards deviations from the Standard Model. More dramatically, the current runs, at unprecedentedly high energies, may reveal new particles that form direct evidence of new interactions at higher energy scales. As our workshop has shown, there are a number of theoretical directions that demand exploration. Detailed understanding of BSM models will be needed to disentangle the possibilities. It is important to allocate computational resources to the numerical study of BSM theories, as an essential component of research at the energy frontier.

# Integrating genomics with hierarchical physical models of DNA and chromosomes

Organisers: Ralf Everaers, Sarah Harris and Cedric Vaillant
Location: CECAM-FR-RA, Centre Blaise Pascal, ENS Lyon (France)
Date: 22 – 26 June, 2015

**State of the Art:** While rapid DNA sequencing has hugely increased the amount of genetic information available, we are still far from understanding how genomes are regulated. For bacteria, this limits our ability to understand virulence, and ultimately to control infection. Understanding gene regulation is particularly challenging for multicellular organisms, whose cells all share the same genome, but as a result of a differentiation processes are highly specialized in the functions they fulfil. In particular, the epigenetic control of the gene activity through DNA- or RNA-methylation, histone modifications, or repressor proteins can be stably propagated through cell division. Recent statistical analysis of hundreds of epigenetic marks along the genome has revealed that eukaryotic chromatin is linearly organized into epigenomic domains characterized by a specific chromatin type (Thurman et al., 2012, Kharchenko et al., 2011, Filion et al., 2010) . At the same time, high-throughput chromosome conformation capture (Hi-C) experiments have shown that the bacterial nuclear material and eukaryotic chromatin is folded into what has been dubbed topologically-associated domains (TADs), which are marked by enhanced intra- and reduced inter-domain contacts (Tung et al, 2013, Sexton et al, 2012, Nora et al.., 2012, Hou et al., 2012, Dixon et al., 2012). The experimental evidence in eukaryotes suggests a strong correlation between epigenomic and topologically-associated domains and hence a coupling between epigenetic regulation and the tridimensional chromosome architecture. Importantly, aberrations in chromosome organization in higher eukaryotes have been linked to serious diseases, such as cancer. These observations indicate that the conventional bioinformatics approach (The ENCODE consortium) alone will not be sufficient for describing gene regulation, because the physical locality and temporal behaviour of the DNA itself imparts information, in addition to the genetic sequence and epigenetic chemical modifications.

There are essentially two approaches to build ensembles of 3d chromosome conformations (Dekker, Marti-Renom and Mirny, 2013): Knowledge-based modeling, where a fiber is folded imposing (distance) restraints inferred from Hi-C and related experiments, and approaches that model chromatin as a polymer. While useful for the interpretation, restraint-based 3d modelling can only hope to predict the consequences of removing loci involved in crucial interactions. The polymer approach has mostly concentrated on exploring statistical organizational features of chromosomes folding. In particular, there is a growing understanding of the key role of topological constraints (Grosberg et al., 1993, Rosa and Everaers, 2008, Halverson et al. 2013) for the crumpled (Liebermann-Eiden et al., 2009) and territorial (Cremer and Cremer, 2001) 3d organization of chromosomes. While most of the current models consider a homopolymer without entering into a detailed modeling of the interactions in real chromosomes and their dependence on the epigenetic state, there are now first attempts to include specific self-attraction or adsorption (Mukhopadhyay 2011, Barbieri et al., 2012; Jerabek and Heermann, 2012; Jerabek and Heermann, 2014) or effective interactions between transcriptionally “active” versus “inactive” genomic compartments (Mateos-langerak 2009).

In contrast to knowledge based “top down” approaches to understanding gene regulation, ab initio simulation methods attempt to use quantum mechanical or atomistic

models to parameterise the mechanical properties of DNA and chromatin from the "bottom up"to deduce the behaviour of the nuclear material; the ultimate aim is to construct a hierarchical physical model of the whole genome. The aim of this CECAM workshop was to discuss the relative merits and disadvantages of the contrasting "knowledge based" and ab initio approaches to understanding gene regulation in the context of the most recent experimental data.

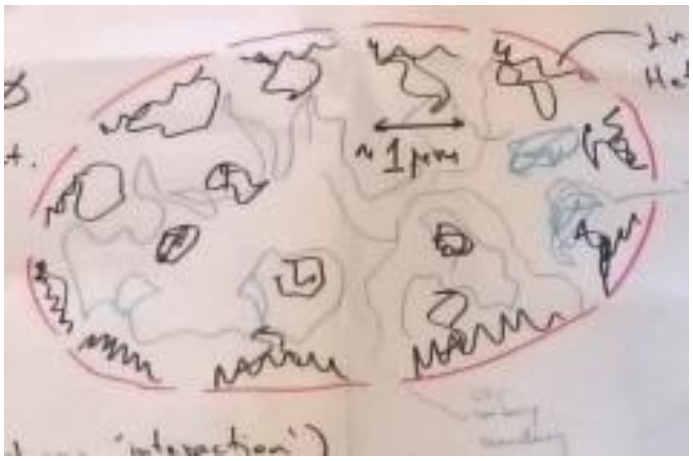



Figure 1. Image of the chromosome organisation produced in discussion session.

**Outcomes:** The following emerged as important questions in the field

1) Interpretation of experimental data and new approaches
   - How do we interpret and validate the results of HiC?
   - How do we measure and understand the changes in chromatin structure as a function of the cell cycle?
   - How do we understand the heterogeneity between individual cells?
   - How do we quantify the outputs from transcription, so we can understand how genes are regulated and monitor changes?
   - What new technologies might we develop to alter the structure of chromatin, so we can understand the affect?
   - Can we do HiC with the proteins removed from the DNA?
   - Can we do HiC and FISH at the same time?
   - Can we look at minimal organisms?
   - Can we find alternative cross-linkers for HiC?
2) Finding an approach to model building
   - What might be suitable simplified model systems to study – both theoretically and experimentally?
   - Are there any universal properties?
   - What time and length-scales are important?
   - Are there any regimes in which a "mean-field" approach can be valid?
3) Understanding the relationship between structure and function
   - Why is 98% of the genome non-coding?
   - What are nucleosomes for?
   - What are Topological Associating Domains (TADS)?
   - Does structure determine function – or function determine structure?
   - Why do chromosomes exisit in one piece in eukaryotes? Why not multiple plasmids, as in bacteria?

**The 3D Organisation of the Genome:** We need to be careful when interpreting HiC data that we do not assume that there is a simple relationship between the contact probability and the distance between two regions of the genome.

Factors that control the folded structure of the chromosome

| **Physical** | **Biochemical** | **Biological** |
|---|---|---|
| DNA/chromatin rigidity (ie phase diagram) | Protein-DNA interactions | Cell cycle |
| Topological constraints | DNA-RNA interactions | Cell shape |
| Supercoiling and plectoneme formation | Histone modifications | Species and cell type |
| Crowding | Transcription and polymerases | Shape of nucleus |
| Topoisomerases | DNA polymorphisms (eg cruciforms, open DNA) | Need to replicate and segregate |
| Interactions with lamellar | DNA looping | Haploid versus diploid |
| Polymer physics, torsional and bending modulus | Lamins, cohesion, CTCF etc | Feedback mechanism, propagation network. |
| Control by diffusion times | Chromatin remodellers | Single versus co-regulated genes |
| Non-equilibrium effects (eg non-ergodicity) | DNA methlyation | Recombination |
| | Co-operative binding and or nucleation of structural proteins | |
| | The role of noncoding RNA in regulation | |

How do we determine how much intermingling there is between different TADS and chromosomes? This is important for segregation after replication.

**The Regulation of the Genome:** The structure and regulation of the genome are controlled by the same factors – it proved impossible to order these in terms of their importance because it depends on the level of resolution and the precise question you are interested in.

These additional points were raised:

- How do we determine the dynamic nature of TADS? Are their positions encoded in the DNA sequence?
- How do we parameterise multiscale models?
- Replication may well take place from each individual TAD concurrently – so there will be one replication origin per TAD. The 3D organisation of the genome, eg the level of compactness of chromatin, enables the cell to remember where it is in the cell cycle.

**Future Directions:** We need a multi-scale, multi-mode, multi-disciplinary approach to understand the relationship between the physical organisation of the genome and its function as an information storage device (3Ms!). We need to understand changes in genome structure with time and across different cell types (to know if we can treat cell ensembles using a "mean field approach").

Our understanding is limited by non-equilibrium effects, poor spatial and temporal resolution, taking population averages, perturbations introduced by the experiments and the huge complexity of the problem. We need to take care with simplified cartoons. While these can be useful for communication of complex ideas they can also be misleading, especially when they are taken too literally by researchers from diverse disciplines (eg

physicists reading biological papers). Most importantly, we need to remember that the central function of a biological system is to replicate itself, and then to separate. Any hierarchical physical genomics approach must be constructed with this basic understanding in mind.

References


R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* **489**, 75 (Sep 6, 2012).

P. V. Kharchenko *et al.*, Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480 (Mar 24, 2011).

G. J. Filion *et al.*, Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212 (Oct 15, 2010).

Tung et al "High-resolution mapping of the spatial organisation of a bacterial chromosome." Science (2013) **342**, 731.

T. Sexton *et al.*, Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458 (Feb 3, 2012).

E. P. Nora *et al.*, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381 (May 17, 2012)..

C. Hou, L. Li, Z. S. Qin, V. G. Corces, Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular cell* **48**, 471 (Nov 9, 2012).

J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376 (May 17, 2012).

The ENCODE Project Consortium. "An integrated encyclopedia of DNA elements in the human genome." Nature (2012) **489**, 57.

J. Dekker, M. A. Marti-Renom, L. A. Mirny, Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature reviews. Genetics* **14**, 390 (Jun, 2013).

A. Rosa, R. Everaers, Structure and dynamics of interphase chromosomes. *PLoS computational biology* **4**, e1000153 (2008).

E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289 (Oct 9, 2009).

T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews. Genetics* **2**, 292 (Apr, 2001).

M. Barbieri *et al.*, Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 16173 (Oct 2, 2012).

H. Jerabek, D. W. Heermann, Expression-dependent folding of interphase chromatin. *PloS one* **7**, e37525 (2012).

H. Jerabek, D. W. Heermann, How chromatin looping and nuclear envelope attachment affect genome organization in eukaryotic cell nuclei. *International review of cell and molecular biology* **307**, 351 (2014).

S. Mukhopadhyay, P. Schedl, V. M. Studitsky, A. M. Sengupta, Theoretical analysis of the role of chromatin interactions in long-range action of enhancers and insulators. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 19919 (Dec 13, 2011).

J. Mateos-Langerak *et al.*, Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 3812 (Mar 10, 2009).

# The Physics of Protein Self-Assembly

Organisers: J. McManus, N. Asherie, P. Charbonneau, E. Zaccarelli
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 22 – 24 June, 2015

**State of the art:** Globular proteins exhibit rich phase behavior that includes crystallization, liquid-liquid phase separation, gelation and cluster formation. Higher-order protein structures also include virus capsid assemblies and condensed phases associated with protein condensation diseases, such as amyloid fibrils. The self-assembly of proteins is an important thread in soft condensed matter research both experimentally and computationally. Being both mesoscale objects in aqueous solutions, colloids and proteins indeed share many similarities. However, anisotropy is a key consideration for understanding protein self-assembly; standard colloidal coarse-grained models with isotropic interactions cannot capture the rich assembly capabilities of proteins. The growing sophistication of synthetic anisotropic colloids and of the corresponding theoretical and computational models thus open the way for a better microscopic understanding of protein self-assembly. These advances have indeed occurred alongside recognition within the protein community of how anisotropy (or "patchiness") affects protein aggregate formation, whether periodic or not. The study of several soft matter models has helped clarify how anisotropic short-ranged interactions control virus capsid assembly, protein crystallization, and cataract formation. Many open challenges remain, particularly for large and non-spherical proteins, such as antibodies. The study of colloidal models that include a varying degree of internal flexibility and shape anisotropy is thus expected to improve our physical understanding of these systems.

**Outcomes:** Key topics to be addressed for progress in the field have been identified:

1. Self-assembly of proteins into mesophases, clusters and hierarchical structures. Two main problems were explored: (1) *The formation of amyloid fibrils*. Due to the association with the pathogenesis of several diseases, the assembly of proteins into fibrils is a very active field of enquiry. During the workshop, a number of experimental, theoretical and numerical studies were reported and discussed. Some advances have been made in modelling the assembly kinetics as a two-step process, and simulations are attempting to mimic these effects both by means of simple patchy models exhibiting "cooperative polymerization" or by more specific models that account for shape or interactions switches between the two steps. In some cases, simulations even predict a three-step process. It is also important to understand the interactions of amyloid peptides with membranes. These results are promising but additional experimental and computational efforts are needed to understand and, crucially, to control protein self-assembly and its connection with the onset of diseases. (2) *Viruses and giant protein assemblies.* Viruses are higher-order protein structures that assemble in vivo. The self-assembly of these uniquely symmetrical structures has been investigated by both experiments and simulations. A two-step mechanism seems to be at play in the simple viruses consisting of single-stranded RNA and capsid proteins. Initially the RNA single-strands reversibly bind to the proteins and then capsid assembly proceeds. Other natural protein assemblies are just beginning to be appreciated. Among them is a broad class of giant, capsid-like assemblies, named bacterial microcompartments, which serve as primitive metabolic organelles. These structures have recently been studied to understand the principles of symmetry that are shared by nearly all natural self-assembly structures, and used as a paradigm to develop methods

for engineering novel proteins that assemble in a variety of complex architectures, including cages, extended two-dimensional layers and three-dimensional crystals. The ability to create sophisticated supramolecular structures from designed protein subunits opens the way to broad applications in synthetic biology.

2. Protein crystallization and phase diagrams. Strategies for crystallizing proteins were presented. One approach based on surface entropy reduction mutagenesis predicts how specific point mutations on the protein surface increase the crystallization probability. The approach has already enabled the efficient crystallization of proteins that were otherwise impossible to crystallize. This approach also allows changes in protein-protein interactions produced by these mutations to be assessed. Microfluidic techniques have been optimized to produce reactors for a single protein crystal to nucleate and grow, all measured with unprecedented resolution. Protein anisotropy, defined by the protein surface chemistry, is clearly important in defining the self-assembly of proteins. Molecular anisotropy (inherent and induced by mutation or chemical modification) leads to changes in the phase diagrams for proteins, and several examples of this effect were discussed. Given the diversity of protein behaviours, predictability of the self-assembly process remains a challenge. To take molecular anisotropy into account, new types of theoretical models are being explored. They build on existing anisotropic "patchy" models for colloidal particles, but need to be adapted to account for the specificity and heterogeneity of the protein surface. These models will be particularly useful when handling high-concentration protein solutions. In addition to well-studied proteins, such as eye-lens crystallins and lysozyme, other classes of proteins are currently being explored, such as monoclonal antibodies, in which the non-spherical protein shape is found to strongly influence the protein phase behaviour. Patchy particle models are also assessed to understand complex gradients in squid lenses. Capturing the essential ingredients that govern these complex assemblies remains a formidable challenge.

3. Where are we at with modelling? Coarse-grained versus atomistic simulations. Towards a multiscale approach. Two broad categories of modelling were discussed during the meeting. (1) *Bulk protein solutions and phase diagrams*: Isotropic "colloidal" models do not capture the full complexity of protein phase behaviour. Several scenarios require anisotropic interaction potentials. The emergence of patchy models, partly inspired by progress in colloid synthesis, has been a main theme in schematic modelling of bulk protein solutions. However, the workshop clearly highlighted the need to go beyond such simplified models. One possibility would be to devise patch-patch specific interactions models, which take into account the specificity of protein contacts that are all different from each other and highly selective. This approach could be combined by a judicious coupling with atomistic simulations, in order to obtain an estimate of the patch-patch interactions. A number of groups are working in this direction. However, the loss of universality it entails currently limits this approach to select groups of proteins. Efforts to find quantities that can be more easily generalized are required. For example, is it possible to predict what are the minimal key ingredients that prevent crystallization of a protein? If so, can we propose detailed experiments (e.g., specific mutations) that could enhance protein crystallization? (2) *Peptides, peptoids and their assemblies*: In this case, atomistic simulations are currently exploited to understand the main features of assemblies, such as the formation of sheets and other nanostructures.

In summary, the workshop highlighted the huge need for better models to describe protein behaviour, mostly at the coarse-grained level. To test these models, it will be important to devise more systematic experiments. However, such a plan is very challenging due to the inherent complexity of the problem, the specificity of each protein molecule and the wide variety of perspectives on the problem. A general consensus is developing, but has yet to emerge. Perhaps the first aspect to implement will be to align experiments, theory and simulation more closely in overlapping research themes. The CECAM workshop was the first step towards a more unified view on protein self-assembly.

**Networking & training:** The topic of protein self-assembly brings together scientists working in different fields, which usually separately attend conferences in the areas of soft matter, material science, colloids and biophysics. This workshop has brought a number of these scientists together for the first time to discuss several shared problems and solutions, and to chart the path forward.

**Funding:** Although a suitable Horizon 2020 call was not identified, given the diversity of approaches within the field, the possibility of regular meetings to discuss the topic in depth already provides a strong platform for establishing connections. The creation of a network, e.g. a training network, within the next few years would be an exciting possibility. Establishing a COST action is also particularly attractive. Because of the geographically scattered nature of the community, it is challenging for researchers to meet regularly. The strong support from CECAM, NSF (USA) and academic centres in both Europe and the US made this workshop possible and shows the benefits an international funding mechanism could bring. Since formal interactions between US and European scientists will be essential to making significant progress, better collaborative vehicles via for example H2020 and NSF are needed.

**Will these developments bring societal benefits?** Progress in understanding protein self-assembly will be extremely important for industrial and societal benefit. Protein condensation diseases, such as Alzheimer's, are directly related to protein self-assembly. More generally, if we can crystallize proteins and determine their structure more effectively, we will be able to develop better medical therapies. Also, increases in manufacturing both in the US and in the EU are likely to be driven over the next decade by advances in the biopharma sector. This rapidly growing sector notably relies on better knowledge of how to maintain protein stability (here, *prevent* their self-assembly) and on the ability to recruit highly-skilled researchers with expertise in this field.

# Modelling metal-based nanoparticles: towards realistic environments

Organisers: F. Calvo, N. Tarrat, M. Benoit
Location: CECAM-FR-GSO, CEMES, Toulouse (France)
Date: 29 June – 2 July, 2015

**State of the art**: This workshop was initially focused on objects, rather than on methods, and aimed to illustrate the current state of computational approaches to metallic nanoparticles in explicit contact with solid, liquid, or gaseous environments. The motivation originates from practical applications of metal nanoparticles, which in the context of information storage, plasmonics, catalysis or nanomedicine (imaging, therapeutics, diagnostics) are either deposited on a support or embedded in a coating medium possibly offering protection from chemical reagents or simply biocompatibility. The methods to study these systems are diverse and reflect not only the variety of situations but also the different properties of interest, ranging from structure to phase transitions, reactivity and optical spectra.

**Outcomes:** The specific topics covered in this workshop include: the accurate determination of interactions (especially long-range forces and charge transfer at contact between chemically different materials), the calculation of optical spectra and the need to overcome the natural time limitation of molecular dynamics through kinetic models. The discussion has also pointed out that taking into account the surrounding environment implies the modelling of much larger systems than for bare nanoparticles. In addition to specific results related to traditional topics of morphology, kinetics and growth, nanocatalysis or nanoplasmonics, several talks highlighted some more general ideas about taking the most of the increasingly large amount of data computed on different systems, either by reduction into general phenomena (through scaling laws) or by explicitly accounting for this diversity in the theoretical prediction (through probability distributions). Two experimental contributions also lead to interesting discussions about practical constraints and the ideal interplay with theory. In the light of the workshop, progress toward a better description of environment effects can be foreseen along the following directions:

1. Better approaches for treating dispersion and charge transfer in the description of metal nanoparticles on extended substrates or grafted with numerous ligands;
2. Better models to represent the interaction between chemically different subcomponents, still chemically realistic but prone to the development of approximate electronic structure methods or force fields;
3. Better ways to account for dynamical processes over long times, in the chemical evolution of nanocatalysts or to describe the growth of nanoparticles in their natural environment;
4. The development of multiscale and coarse-graining approaches to capture large and complex environments, either for the environment itself (implicit descriptions of a fluid or pressure transmitting medium) or the system of interest (ligands made of long polymer chains). For instance, the issues regarding parameterization and transferability are essential as one would like to describe the chemistry as best as possible;
5. The added complexity offered by multimetallic particles (nanoalloys) has become accessible owing to powerful algorithms, and in the near future it should become easier to take the environment into account for those particles as well. Some issues of nanoalloys, including the propensity for mixing and how to deal with

chemical ordering, are also relevant to metal nanoparticles interacting with specific chemical environments such as hydrides or carbides. Here again coarse-graining plays an important role through the possible simplification of atomic structure into Ising-like lattice models that only explicitly consider chemical ordering.

Some of the main obstacles to success were also discussed. The main difficulties identified are associated with the diversity of approaches and properties of interest to the various researchers. Clearly the field would benefit from greater networking, keeping track of the methodological developments that are not necessarily devoted to metal nanoparticles but have a broader scope. Some central aspects involving the vast majority of the community are the descriptions of the interactions, including electrostatic and long-range forces, the coexistence of metal-nonmetal components, and the possible simplification and integration into force fields at various levels of coarse-graining. Benchmarking large-scale electronic structure calculations requires highly efficient codes that only few groups possess, so this community should also remain in contact. Overall, one of the general difficulties faced by researchers in the field relies in the transfer of knowledge from electronic structure calculations towards force field parameterization and higher levels of coarse-graining so that larger length scales or longer time scales can be reached.

**Community needs:** No specific infrastructure seems necessary to advance the hardware and software aspects of the field. Data storage and having available repositories is also not essential at this early stage. However the existing infrastructures for networking and structuring the community should be further exploited.

**Funding:** Together with other networks such as COST actions (one application on the field of nanohybrids is presently submitted), the infrastructures provided by CECAM offer a particularly suitable way of gathering the theoreticians of the field and, as organizers; we received the warmest encouragements to extend the present workshop into future off-springs.

**Networking and training:** The workshop was a meeting place for theoretical chemists, physicists and biologists commonly working on different aspects of metal nanoparticles from various perspectives, and using methods such as force fields (at different levels of coarse-graining) or employing explicit descriptions of the electronic structure for the ground or excited states.

**Will these developments bring societal benefits?** The central theme of our workshop remains fundamental at present, but is clearly driven towards applications in various areas of materials science, chemistry and biology. So far the intrinsic difficulties associated with the description of chemically complex systems have prevented a unified picture from emerging. General algorithmic strategies to deal with ligand grafting or the self-assembly of particles, for instance, are still lacking on a quantitative level and it is highly desirable to identify more systematic routes to efficient modelling that would pave the way toward realistic modelling of large nanostructures in their natural environment.

## Modelling activity versus selectivity in metalloproteins

Organisers: Marco De Vivo, Michele Cascella, Carlo Adamo, Ursula Roethlisberger
Location: CECAM-FR-IDF, Paris (France)
Date: 29 June – 1 July 2015

**State of the Art:** This workshop was focused on the current methodological challenges that first-principles-based methods are nowadays facing when applied to metalloproteins. In particular, we focused on the progress made toward a better description of the functional role of different metal ions in catalysing enzymatic reactions. Issues related to the description of the electronic and steric effects of metal ions such as Mg, Ca, Mn, Zn, Fe etc., bound to enzymes, have been subject of discussions with opinion leaders in the field.

**Outcomes:** Over the next 5-10 years, methodological advancements of computations, such as new functionals or new multiscale schemes will further improve our ability to treat realistic metal-aided biological model systems with increased accuracy. The impact of computation on experiments will help, and guide, practical applications such as design of artificial metalloenzymes with wide applications, including industrial biosynthesis.

To advance computational methods toward this level of simulations, the community will face major challenges related to the treatment of the metal ion(s) interactions with the biological counterpart. This was the major debate at the workshop, with discussion on electronic delocalization effects during metal recognition and binding to enzymatic pockets; structural (steric or electronic) effects regulating metalloenzyme activity and selectivity (with particular emphasis on long-range coupling to protein scaffold and solvent); the accuracy of electronic structure methods in describing enzyme-bound transition metal ions.

Finally, one key point is whether first-principles-based methods will impact on the rational design of biomimetic metal-complexes with engineered functions, or small compounds able to block, or modulate, metalloenzymes. Not only from a pure academic point of view, but with a concrete applicability of computational tools for modelling metalloproteins in a realistic way, accelerating the use of these once prohibitive tools in industry, as well – for example.

In the medium/long run, studies centred on metalloproteins, with focus on how – exactly – the metal ion(s) operate to accelerate catalysis, and help achieving substrate selectivity, might generate great impact on the pharmaceutical industry, where these proteins are usually considered neglecting the fact that the metal is crucial for function. A better understanding of how these proteins work would therefore help the pharmaceutical industry, enlarging the type of mechanisms and inhibitors currently investigated to block this type of targets, which typically do not take any advantage from the presence of the metal in the targeted pocket.

Also, the design of artificial metalloenzymes represents a frontier in the field, with promising applications in biosynthesis – this could indeed represent a topic of interest in the near future, with applications where rationally designed metalloenzymes catalyse challenging industrial reaction, building selectively targeted macromolecules.

The workshop offered a list of contributors from the experimental and computational communities. In particular, talks included topics covering electronic structure theory,

protein simulations, enzymatic catalysis, structural biology, homogeneous catalysis, and novel functionalized biomaterials.

The different sessions gave the chance of debating hot topics metalloenzymatic catalysis, like Rnase H, where multiple speakers from computational and experimental groups could debate on possible mechanisms of catalysis and selectivity.

We were also pleased to give the opportunity to young researchers to present their work both in a poster session and by a number of selected contributed talks. Also thanks to the informal environment kept during the length of the workshop, it was possible to maximize the exchange between senior and early stage researchers. We hope that the experience contributes as much as possible to the scientific growth of the young researchers attending the meeting.

Finally, the possibility of starting new collaborations on multiple topics among several participants was debated, and new collaborations were established during the meeting. Being this one of the major goals of such initiatives, we can consider that our workshop has been very successful.

# Advanced thermoelectrics at nanoscale: from materials to devices

Organisers: Michele Amato, Stefano Ossicini, Riccardo Rurali, Phillipe Dolfus, Gyeong Hwang
Location: CECAM-FR-IDF, Paris (France)
Date: 7 – 10 July, 2015

**State of the Art**: The combination of thermal, electrical, and semiconducting properties of thermoelectric nanomaterials allows them to convert heat into electricity. These materials are expected to play an increasingly important role in meeting the energy challenge of the future. Major advances in this field strongly depend on our fundamental understanding of heat and charge carrier transport and on the ability to find new strategies to design and fabricate high efficiency thermoelectric devices and circuits. In spite of the substantial advances in the description of thermal and electronic dynamics in bulk materials, the extension of transport bulk theory to nanostructures, is still under development. One of the main problems in modelling the nanostructures for thermoelectrics is the fact that they usually have complex compositions and structures. To these complex structures, usually, several external elements are added to improve either the thermoelectric properties and to become functional elements of devices and circuits. The final material is hence a quite complex object whose phononic and electronic structure is unknown. Recently, the research community has consolidated its belief that more accurate theories and approaches are needed. In particular first-principles calculations (i.e. accurate simulations free of adjustable parameters like Density Functional Theory and Density Functional Perturbation Theory) have started to be considered the most reliable way of computing phonons and electrons related quantities in a material. Usually these kinds of simulations are combined with quantum (i.e. Non equilibrium Green's Functions), semi-classical (i.e. Boltzmann Transport Theory) or classical (i.e. Molecular Dynamics) approaches in order to get electronic and thermal conductance and to take into account specific effects (i.e. when different types of scattering or bigger systems have to be considered). Very recently also the coupling of the above-mentioned methods with macroscale approaches (e.g. compact models) has emerged a promising way to reproduce devices and circuit characteristics through modelling. In all these hierarchical multiscale strategies one carries out the computational simulation at one scale and extracts quantities that can be used to define the parameters of the model operative on the adjoining (usually larger) scale.

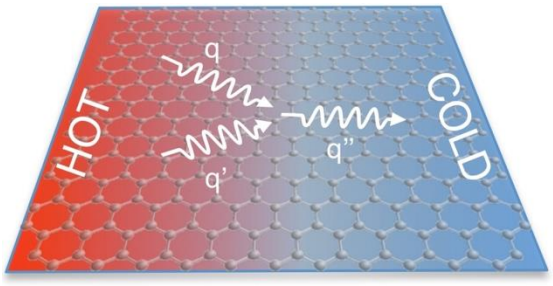



Figure 1. Pictorial view of phonon scattering in graphene [from Fugallo *et al.*, Nano Lett., **14** (11), pp 6109 (2104)].

**Outcomes:** Even though thermoelectricity is still largely governed by engineering sectors, that favour an empirical approach to the problems, rather than pursuing a

fundamental understanding of the underlying physics, encouraging signals are coming from basic research. We foresee an increasingly important role of theoretical modelling and simulations in what, rather than application oriented, we would like to name application inspired research. The key challenge for the next 3-5 years is creating a real synergy with industrial R&D divisions. For this to happen, modelling must become more predictive, and must be able to communicate that its predictive capability, rather than quantitative, can be qualitative (elucidation of trends, highlighting of physical mechanisms) and still be useful. The workshop's organisation proposed some novelties with respect to the traditional CECAM format. A world leader in the field of thermoelectricity, Prof. Mark Lundstrom (Purdue University) was in fact invited to give a tutorial lecture. Prof. Lundstrom is author of several seminal papers and a best selling book on nanoscale electron transport. At variance with ordinary invited talks of the workshop we reserved a 1h30' slot for this inaugural talk. We must say that, not only students, but also senior scientists were delighted by the pedagogical character of such an introductory lecture. Another minor novelty, already implemented by a few workshops in the past, is that we are collecting the presentation slides of all the speakers that have given their consent in this direction, and will upload them on the workshop website for dissemination. We believe that, especially in the case of tutorial presentations like the one of Prof. Lundstrom, this can gather a considerable interest.
A major output of the workshop was establishing the importance of collaborative research projects entirely or at least partially based on ab-initio calculations in the field of thermoelectricity. This is a necessary choice since more and more it is necessary to understand the properties of new materials at the nanoscale, at an atomistic level in order to guide the experimental work towards innovation.

**Funding:** It would be unfair to say that the field of thermoelectricity has not previously received funding from the regular programs of the EU. However, we strongly believe that this is the right time for a strategic increase of funding in this area. As said, modelling can be of great utility, but has been too often neglected in these large-scale projects. Specific calls directed to the theoretical community could partially restore this unbalance.

**Will the results of these developments be of societal interest?** In Horizon 2020 it is explained that the specific programme shall consist of the following parts: (a) Part I *Excellent science*; (b) Part II *Industrial leadership*; (c) Part III *Societal challenges*; (d) Part IV *Non-nuclear direct actions of the Joint Research Centre (JRC)*. The output of the workshop is directly linked with *Excellent science*, *Societal challenges*, and indirectly with *Industrial leadership*. In particular, the introduction of new nanomaterials for thermoelectric application plays an important role in the enabling technologies for information and communication and nanotechnologies. Furthermore, regarding the priority *Secure, clean and efficient energy* in Societal challenges, we envisage two key points for use of nanostructured thermoelectric materials and devices: (a) In thermoelectric application, where reliable and scalable devices are presently used in a number of applications for both turning heat into electricity, measuring temperature and using electricity to produce cooling (Peltier cooling and temperature controllers). For the design of practical energy generator devices, not only the efficiency has to be taken into account. Indeed, since the waste heat is abundant (and of course inevitable) in industrial processes, home heating and automotive exhaust, the specific output power, which in the end determines the final utilization of the generated power, turns to be more significant. Thus, nanostructured-based devices are expected to take on special importance since they have low cost and can be fully integrated; (b) the substitution of rare and toxic material, one of the main action that Europe will favour in Horizon 2020.

# Next generation quantum based molecular dynamics: challenges and perspectives

Organisers: T. Frauenheim, A. Niklasson, J. Hutter, B. Hourahine
Location: CECAM-DE-MM1P, University of Bremen (Germany)
Date: 13 – 17 July, 2015

**State of the art**: Molecular dynamics (MD) simulation is a powerful computational tool that is widely used in materials science, chemistry and biology. MD simulations provide a general approach to understand and analyse a wide range of material systems and serve as a valuable complement to support and interpret experiments. MD simulations allow us to directly "see" and manipulate phenomena at the atomic scale and evaluate how properties can be modified to optimize specific designs. The field of MD is currently dominated by classical simulation methods, where the interactions between atoms are described through classical force fields in which the complex electronic structure of molecules and solids is reduced to simple, parameterized, interactions between atoms that are fitted to experiments and/or electronic structure calculations. Unfortunately, classical MD simulations can be applied only to a limited set of materials, and are often not reliable outside the initial parameterization scope. This is especially problematic in situations when there is significant interatomic charge transfer between atoms, or there is covalent bond formation or breaking. In fact, for many important problems a meaningful parameterization may not even be possible. A quantum based MD therefore offers not only a more accurate and trustworthy replacement of classical MD, which is critically needed, but it also provides an essential path to modelling systems and phenomena that are clearly beyond the reach of existing, classical methods. Unfortunately, QMD simulations remain intractable for most systems because of their computational complexity.

**Outcomes:** The workshop identified and tackled a set of key topics for the field. These were:

1. *Fast quantum based MD methods* (Stable, energy conserving algorithms that reduce the computational overhead of SCF optimization to a minimum and that can be combined with O(N) sparse matrix schemes that can be efficiently parallelized for the current generation of high performance computers.) A key problem with the size of the computational pre-factor that so far has limited many practical applications of O(N) based QMD simulations was addressed in several of the discussions as well as how stability can be achieved even in the limit of vanishing SCF convergence.
2. *Low pre-factor Order-N electronic structure algorithms* (Methods that have a computational complexity that scales only linearly with the system size, instead of the cubic scaling of traditional methods. In this respect a low pre-factor can be achieved by applying sparse matrix algebra and new parallelization strategies.) A number of O(N) approaches to efficient parallelism, both over shared and distributed memory platforms were presented as well as different algorithms to achieve linear scaling. Some of parallel linear scaling electronic structure implementations demonstrated rapid recent advances by possibly as much as an order of magnitude in run time performance compared to state-of-the-art less than a year ago. This topic was covered by representatives of five of the most important method development groups.
3. *Approximate DFT-methods* (Self-consistent charge density functional based tight-binding methods and semi-empirical schemes that can reduce the computational

overhead compared to direct DFT methods by about 2 orders of magnitude for suitable systems.) Progress with the implementation of an extended Lagrangian Born-Oppenheimer formalism were presented that further reduces the computational overhead in approximate DFT methods by an order of magnitude. New approaches to combine approximate DFT based QMD schemes with linear scaling electronic structure theory were also presented.

4. *Quantum-derived many-body potentials* (Bond-order, Learn-on-the-fly and Neural Network potentials, machine learning for parameterization. These enable enhanced transferability over classical potentials.) New methods, for example, how neural networks can be trained to represent complex interatomic potential energy surfaces by including long range forces were presented, as well as recent progress with bond-order potentials.
5. *QM/MM hybrid approaches* (Multi-scale approaches where the effects and responses of the extended surroundings are modeled with more approximate but computationally faster classical force fields, while retaining quantum mechanical accuracy for the more complex core region(s).)
6. *Accelerated MD for bridging time scales* (Techniques to extend the effective time scale of MD simulations as well as methods to boost MD driven sampling of the phase space.)
7. *Advanced integration schemes for MD* (Multiple time step algorithms, parallelised integrators for dynamics, path-integral formulations of quantum nuclear motion)
8. *Challenges in Computational Science* / hybrid architectures (Formulation and design of algorithms for large-scale QMD simulations that take full advantage of advanced massively parallel hybrid architectures, and look towards the next generation of > 100 Petaflop parallel architectures.)

**Community needs:** There are several open source and commercial software solutions for classical and quantum mechanical based MD simulations. The workshop recommended developing repositories of highly accurate and open quantum based data for validation of more approximate methods (such as DFTB and classical potentials). Vice versa parameter sets for the approximate methods could also be stored in such a repository. The participants believe that the scientific exchange of researchers from different disciplines is key to make progress in the most challenging problems of theoretical materials and bio-systems modelling. This exchange can be achieved by interdisciplinary conferences and workshops. Moreover, it would be desirable to enable exchange not only at the level of researchers, but also at the level of PhD and Masters students.

**Funding:** The field of theoretical materials modelling relies strongly on the availability of state-of-the-art computer code and computing resources. Therefore, the continued funding for the development of modelling software is an important infrastructure requirement: Only when codes are freely available can new ideas be tested and developed by a broad community of researchers. Since new codes that address the most challenging materials and bio-systems applications will push the limits of current supercomputing resources, it is also important that funding of these resources remains significant in the future. The workshop identified, in particular, research areas that might be funded by the EU 2020 program particularly through the EINFRA and COST programs, and workshop participants would support an application involving multi-scale applications of Quantum based Molecular Dynamics to either of these programs. Modelling of materials stability, interface design and bio-molecular interactions could form areas of particular focus.

**Networking & training:** The Workshop was successful in interconnecting different communities involved in quantum-based molecular dynamics modelling working on large-scale atomistic simulations of materials and devices, regardless the specific scientific subarea of interest. The workshop focused on most recent developments of computational and theoretical methods and techniques stimulating an intense dialogue between ab-initio, semi-empirical, empirical and classical method communities. The program consisted of 30 invited talks of 40 minutes (35+5) each and a poster session presenting 28 contributed posters. In addition, social events were held (including a reception and conference dinner) to allow for informal exchanges. Well-established scientists gave the invited talks from different theoretical communities from all around the world, which acted as platform for interesting cross-/interdisciplinary discussions. The invited talks were followed by a poster session where the younger participants could show their scientific work and exchange of ideas with a broad knowledge in computational chemistry, solid-state physics and computational materials science. The organization was very compact with the scientists accommodated in the same hotel fostering exchange and discussion between the participants also outside the meeting room. The invited talks and intense discussions contributed to define the state of the art at a computational level, regarding performance, scalability, and parallelisation on different platforms and hence range of applicability of different quantum based MD techniques. Common targets for method developments between different communities have been discussed, focusing on hierarchical and domain multi-scale coupling techniques, interconnection between electronic structure and classical methods. The Workshop Speakers and Participants agreed to have a similar meeting in 2 to 3 years from now to evaluate the rapid progress in this field.

**Will these developments bring societal benefits?** Progress in the field of molecular dynamics is of great interest to Industry (representatives of Atotech Deutschland GmbH and the Fraunhofer Institute for Mechanics of Materials being present at the workshop for example). Classical force-field MD is ubiquitous in fields such as pharmaceutical design (e.g. Unilever) and more general bio-science (Procter & Gamble), targeted surfactants (e.g. BP, Schlumberger) and nano-fluidics (?). Other developing areas of industrial interest include high energy density materials (explosives) with coupling of mass and heat transfer to very fast chemical reactions; battery technologies and energy storage, specifically for applications to car manufacture; chemical corrosion driven crack propagation in engineering materials. Further developments in these areas require advancement to the quantum MD level of modelling. Transferability demands and increased chemical complexity also require robust parameter free approaches (or methods that transfer smoothly to more approximate fast methods), but the challenges of long time scales and large relevant phase spaces also require advanced integration and sampling techniques. Molecular dynamics applications are currently substantial users of present high performance computing (featuring on the roadmaps of several national HPC programs), but migration to near term platforms that require efficient use of even more processors presents immanent new challenges. Robust, open, and most importantly validated, software and methods are essential to reliably drive future developments and applications in this field.

# Recognizing the Relevance of Change: Analysis and Control of Time-evolving Networks in Epidemiology and Evolutionary Medicine

Organisers: Max Von Kleist, Natasa Conrad, Heike Siebert, Christof Schuette
Location: CECAM-DE-MMS, Freie Universität Berlin (Germany)
Date: 20 – 22 July, 2015

**State of the Art**: Typically, static networks are derived by integrating the information derived from observing a process over a period of time, determining the extent of linkage of nodes and subsequent thresholding to obtain an easier-to-interpret sparse network. Several examples presented throughout the workshop highlighted that this traditional way of analysing data lacks information about causal paths that underlie the observed process and consequently false conclusions may be drawn of how to control a process on this network. In the Life Sciences one typically aims to understand and possibly control a process. This necessitates an extension of vocabulary, methods and results from static to dynamic graph theory. Grounding work formalizing the notions started only a few years ago and has rapidly become very interdisciplinary. Analysis of time-evolving networks (TENs) focuses on characterization in terms of topological properties, statistical measures and relevant structures. While many notions can be generalized from static network theory, some relevant graph theoretical concepts cannot be transferred in a straightforward way to capture and explain evolution of network structures over time, posing challenges for current research. In the field of epidemiology and disease transmission, it is evident that a mean field approximation for disease spread (and contagion evolution at the population level) is not capturing the true dynamics. Instead, almost all infectious diseases are transmitted through contacts, which dynamically change in time. In evolutionary medicine, e.g. in applications like tracking cancer progression and drug resistance development, changes in network topology are often seen in terms of successive genetic alterations. Epistatic interactions give rise to complex and non- random traits of network evolutions that may even be reconstructed using novel experimental tools (e.g. next generation sequencing) in combinations with advanced mathematics/statistics. In general, network inference or reverse engineering approaches are a hot topic both in a theoretical and a more applied context. The difficulty of this task is amplified when considering the possibility of network changes over the time of data acquisition. Possible approaches include, e.g., the introduction of hidden states describing the networks structure at a given time point utilizing variants of Hidden Markov Models, the use of time-slices, which however introduce some bias. Particularly in public health and medical applications, control of time-evolving networks is the ultimate goal. Currently, even control strategies for static networks may utilize a targeted change of network topology to achieve some desired outcome, thus pinpointing the possibilities of being embedded in a more comprehensive theory taking variability of topology into account. In fact, many infectious diseases and cancer have a large propensity to develop drug resistance, which is driven by an evolving network structure. Control strategies for such dynamic networks pave the way to long-term suppression of infectious diseases. Rigorous methods and formalisms incorporating the time variability of networks in the context of network control are mostly missing, opening the field up for fruitful developments.

**Outcomes:** The following topics emerged as key during the workshop:

- *Network interference.* Application-tailored development of analysis methods and tools hinges naturally on the availability of high-quality network models. In

reverse engineering, approaches need to be adapted so that temporal changes in the network structure can be recognized, allowing for higher time resolution while still coping with data uncertainty. Although in a different context, similar difficulties need to be overcome when generating contact networks, which are of particular interest in epidemiological applications. Here, the high effort of conducting the corresponding studies, including the supply of the needed technology and ensuring the consent of all participants leads to only few real-world networks being available for case studies. Applicability of the developed analysis and control strategies, however, can only be ensured if the suitability and generality of the developed notion and concepts is strongly supported by extensive tests on real-world networks.

- *Scalability of analysis and simulation:* With the increasing power of high throughput technologies and storage capacities, more and more datasets from real-world systems in the form of large time-evolving networks became available. In order to understand the underlying systems, new temporal metrics and methods need to be developed. In the workshop, several approaches were discussed, focusing on two main points: what type of measures can capture the best the temporal patterns of interest and how can these be efficiently applied for understanding complex and large real-world systems generated by uncertain data. Participants of the workshop coming from various fields approached these questions from different angles, which resulted in vivid discussions and pointed out the need for more interdisciplinary between researchers working on this topic.

- *Control:* Although apparent in all talks dealing with applications, current approaches of controlling time-evolving networks are rather ad-hoc and limited to the evaluation of a few candidate strategies. A rigorous framework implementing approaches from control theory is largely missing and there is room for development. Two very interesting observations were made: There is a natural link between time-evolving networks and control theory, where the latter aims to change a process (e.g. described by a network) over time to reduce some overall cost/objective function, thus deriving some 'optimal time-evolving network'. However, in the field of optimal control, efficient methods for systems that are intrinsically stochastic are largely missing.

**Community needs:** Researchers in the field are generally willing to share their software tools and that there is already a lively exchange of simulation and visualization tools. However, there is no central software hub for solving recurring tasks in TENs (the individual solutions are currently largely deposited in GitHub) and therefore one could envision, as the specific research community grows, to create a central domain that summarizes & links available tools. It also became evident, that some approaches, particularly those aiming to bridge within host time-evolving infectious disease networks with between-host dynamics heavily rely of computing power.

**Funding:** The workshop underlined the potential of interdisciplinary projects tackling problems in different areas of the health sector. Consequently, funding opportunities are given in the EU Horizon 2020 programme not only in the Excellent Science but also in the Societal Challenges section. Relevant research areas include: Infectious diseases, pandemic threats and antimicrobial resistance. Specifically, inference, analysis and control of time-evolving networks could play an important role in finding better ways to prevent disease spread. For personalized medicine TEN could be used in the context of managing antibiotic resistance as outlined in the workshop.

**Will these developments bring societal benefits?** Time-evolving networks play an important role in capturing phenomena essential for understanding and controlling the dynamics of and on networks underlying a wide variety of applications, e.g. in the public health sector. Minimizing contagion risks, and thus spreading of diseases, and developing treatment strategies are only two scenarios that can benefit extremely from approaches and insights gained from analysis and control of time-evolving networks. In systems pharmacology these approaches may be used to avoid drug resistance by controlling the disease process. Similarly, since time-evolving contact networks shape disease dynamics and evolution at the population level, the effects of time- evolving connectivity patterns need to be taken into account when formulating contagion control strategies.

# The role of local structure in dynamical arrest

Organisers: C. Patrick Royall, Thomas Speckuly
Location: Johannes Gutenberg University Mainz (Germany)
Date: 21 – 23 July, 2015

**State of the Art:** A key theme of materials science is that the structure assumed by the constituent atoms and molecules underlies the nature of the material. Glasses challenge this notion, and indeed whether one can even distinguish glasses and liquids structurally remains a matter of heated debate [1,2,3,4]. There were two major challenges (one technical and one fundamental) addressed at the workshop. The first, technical, challenge is the gap between the structural data accessible in experiments and the dynamical range achievable with simulations. In experiments on molecular and metallic glass formers one is typically restricted to two-body correlations such as the static structure factor, which makes the identification of local structural motifs a major challenge [3]. Simulation, which provides access to all coordinates, is itself limited in the dynamic range it can access.

Second, in addition to the challenge of identifying structural change in the relevant regime, the question arises as to whether this change is responsible for, or merely a by-product of, the dynamic slowdown that characterizes the glass transition. Different communities are divided on this topic: those who study dynamical arrest from a fundamental viewpoint are divided on the role of structure [3,4] while those who study practical glass-forming materials such as metallic glasses largely accept that local structure is responsible for the slow dynamics [5]. Going even further, the exploitation of metallic and chalcogenide materials in particular requires control of the delicate balance between vitrification and crystallisation [3].

To move beyond the current impasse, the workshop brought together these two communities. The aim here is twofold: on the one hand new techniques have been developed both in simulation (such as pinning and the discovery of dynamical phase transitions [4]) and in experiment (such as nanobeam electron diffraction [3] and scanning electron nanodiffraction [6]). On the other hand, by combining the knowledge from a theoretical viewpoint with practical structural studies on glass-forming materials, we aimed for an improved consensus to emerge.

In summary, if we accept that there is some change in structure approaching dynamical arrest we are left with three major questions that were investigated in this workshop: (I) Does the development of locally favored structures (LFS), such as the canonical icosahedron, really drive dynamical arrest, or is it simply a by-product of cooling down a liquid? How to bridge the gap from simulation to experiment? (ii) What is the relation between local structure and crystallization? (iii) How universal, if any, is the role of structure across the range of dynamically arrested systems including, e.g., gels and granular systems?

**Outcomes:** In addressing the questions above, new insight concerning (i) was presented. In the simulation-accessible regime, it has become clear that some model systems (such as the Wahnstrom Lennard-Jones model) exhibit dynamics that appear to be largely controlled by locally-favoured structures. This model lies at one end of a spectrum, at the other end, LFS have little impact on the dynamics, rather it seems "mean-field amorphous order" dominates. The workshop is the first time this concept was explicitly expressed, and three speakers made the link. A major challenge going forward is to define clearly what is meant by "mean-field amorphous order": such materials are well-described by

Mode-Coupling theory, but the connection from the limited degree of supercooling accessible to simulation to the further 9-10 decades of dynamic slowing relevant to experiments remains for the future. The role of the entropy cost of LFS in the liquid as a key aspect of determining liquid stability and suppressing crystallization was introduced.

Being able to obtain some measure of 3d coordinate data from experiments is crucial. This is now becoming possible, albeit for a limited (though important) range of materials, such as the metallic glass, CuZr. Since experimental systems can be so much more deeply supercooled than simulations, we hope it may soon be possible to use data such as this to discriminate between different theories of the glass transition such as Random-First-Order Transition theory and dynamic facilitation. The workshop bought together leading experimentalists working with such techniques and theorists/simulations with the background to use such data to great effect.

What the community needs. One of the aims of the workshop was to bring together theorists and simulators working at a very fundamental level with those working on model and more applied experimental systems. We believe this was achieved; moreover the standard of the talks was very high, in particular from the point of view of reaching out to other communities. What became clear at the meeting is how diverse these communities are. It would support the field greatly, and enable us to leverage the ideas expressed in the workshop, if some kind of support could be realized for networking between these different fields. Ideally perhaps an Initial Training Network might enable theoretical insight to lower key barriers to commercial exploitation such as glass-forming ability in metallic glasses (currently the major barrier to industrial utilization of such materials) and crystallisation times in chalcogenide phase change materials [point (ii) above], which limits their current use in non-volatile memory applications. A particular example is Nokia, which is using phase change materials in mobile devices but is facing strong commercial competition from other technologies.

Scientific challenges. The role of chemical interactions in many practical glass-forming systems was emphasized. While theories and simulation seem focused on spherical particles, these are relevant predominantly to metallic glass-formers: deeper understanding of molecular and oxide glass-formers requires more sophisticated approaches.

Our understanding of crystallisation in glass-forming systems is in its infancy [3]. Some progress has been made (and was presented at the meeting) but full phase diagrams of the most popular model glass-formers remain to be calculated, let alone nucleation rates. We believe these calculations, and their counterparts in well-parameterized applied systems (such as CuZr) present key opportunities to improve glass-forming ability in metallic glasses, which is currently among the key limitations in the emergence of such materials.

Order corresponds to predictability. Thus at a more theoretical level we would like to be able to predict the structure (either LFS-based or "mean field amorphous order"). If there were some way to extend point-to-set approaches to much deeper supercooling, this might be realizable.

References

[1] "The Nature of Glass Remains Anything but Clear", New York Times, 29th July (2008)
[2] Ediger M.D. and Harrowell, P., J. Chem. Phys. 137, 080901 (2012)
[3] Royall C.P. and Williams S.W., Phys. Rep. 560, 1-75 (2015)
[4] Berthier L. and Biroli, G., Rev. Mod. Phys. 83 587-645 (2011)
[5] Cheng Y.Q. and Ma, E., Prog. Mat. Sci 56 379 (2011)
[6] Liu A.C.Y. et al. Phys. Rev. Lett. 110 205505 (2013)

# Intrinsically Disordered Proteins: Bringing Together Physics, Computation and Biology

Organisers: B. Kragelund, R. Best, K. Lindorff-Larsen, B. Schuler
Location: CECAM ETH, Zürich (Switzerland)
Date: 18 – 21 August, 2015

**State of the Art:** Intrinsically disordered proteins (IDPs) or regions of disorder interspacing folded domains (IDRs) comprise large fractions (~third) of the cellular proteomes. The pervasive structural heterogeneity and complex dynamical properties that characterize IDPs and IDRs make them challenging to study both experimentally and through simulation studies. Thus, to gain a more detailed view of IDPs it is typically needed to combine multiple different techniques including, for example, NMR spectroscopy, ensemble spectroscopy (FRET, fluorescence, CD, etc.), SAXS, single molecule spectroscopy, molecular simulations and purely theoretical considerations. Although a number of systems have now been characterized in substantial detail using these methods, only a few studies have so far linked these structural studies to the biological function of IDPs. To promote the importance of bringing together *in vivo*, *in vitro* and *in silico* approaches to studying IDPs and to provide researchers from different fields with inspiration and knowledge dissipation, this interdisciplinary CECAM workshop on IDPs was established as a forum for exchanging results, ideas and hypotheses. The meeting brought together theoreticians and experimentalists, but also gathered different disciplines, notably physics, chemistry and biology.

**Outcomes:** The interdisciplinary exchange of scientific competence and perspectives proved to be a very fruitful source of inspiration for all the participants. A total of 18 invited and 5 contributed oral presentations were given during the workshop, combined with ample time for discussion during and after every presentation, an aspect that was highly appreciated by the participants. Furthermore, approximately 20 poster presentations enriched the program through two extensive poster sessions, covering various aspects of IDP characterisation, single-molecule studies, biophysical studies of folding-upon-binding, molecular and coarse grained simulations of IDPs and IDRs, bioinformatics analyses, and biological data.

*The meeting highlighted the following key areas for current and future studies*

1. Strong prospects for integrating experiments, theory and simulations to study the structure and dynamics of IDPs
2. Role of IDPs in formation of 'granules' and separate liquid-like phases within cells. In particular, theory and simulation could play a key role in developing this area further.
3. Charge distribution, hydrophobicity and transient structures as important determinants for function by modulating affinities, binding rates and structural properties of IDPs.
4. Potential for rapid evolution of new protein functions within intrinsically disordered regions.
5. Concepts from polymer theory can provide a strong baseline for a conceptual understanding of IDPs and for interpreting experiments
6. Affinity tuning through ensemble redistribution and post-translational modifications
7. Advances in molecular simulation force fields which make the treatment of IDPs much more accurate using explicit solvent simulations. This is clearly area where further significant improvements can be achieved.

8. Coarse-grained and multi-scale models appropriate for IDPs are nonetheless essential for describing many large-scale phenomena.

The main conclusion from the workshop highlighted that the field is now at a point where computation, physics and biology are obliged to unite and cross-fertilise of benefit for the field's development. The closer interaction between these disciplines was a key theme of many of the presentations at the workshop. With IDPs emerging as increasingly central to molecular and cellular biology, discoveries of new functionalities and hitherto unknown mechanisms are likely to occur.

**Community needs:** A key outcome of the meeting was that it is clear that computational methods will continue to be an integral part of biophysical and biological studies of IDPs. Given the complexity of the systems and the long time-scales associated with IDP function, such studies will require the application of a range of methods, both existing and novel, with a broad range of computational requirements. At the more expensive end, the meeting highlighted the increasing accuracy of all-atom molecular dynamics simulations. Currently, serious applications to studies of MD simulations require access to substantial national and international computing infrastructure (e.g. PRACE), and we envisage that future developments will require access to very large computing facilities as well as the development of theories and software to make simulations run efficiently. Multiscale methods, which integrate models of different levels of complexity, as well as methods to directly integrate experiments and simulations will also play a key role in the future.

**Will the results of these developments be of societal benefit?** Progress in the understanding of and ability to manipulate IDPs has practical applications in a wide range of areas. For example, the central role of IDPs in signalling networks, as well as their involvement in the most widespread causes of death and disease (cancer, cardiovascular, neurodegenerative diseases), means that IDPs are unique targets for the treatment of an aging population. Studies of IDPs will thus be important in several of the focus areas of the 2020 programmes. To overcome the inherent problems in developing IDP drug molecules several areas for future studies were identified in this meeting:

3. Improving the accuracy of energy functions for molecular simulations of disordered proteins
4. Developing new methods for combining experimental data with molecular simulation to determine structural properties of specific IDPs
5. Development of new polymer theory to describe the structure and dynamics of IDPs
6. Detailed study of IDPs, in particular larger proteins, require access to state-of-the-art facilities for biophysical experiments including NMR, SAXS and computational facilities
7. Complexity of IDPs necessitates the integration of a wide range of disciplines to study IDP biology
8. Develop approaches for targeting IDPs with drug molecules
9. Stronger and broader biological approaches to link structure and functions of IDPs

The research in this field should ultimately impact the pharmaceutical industry, given the involvement of IDPs in many human diseases ranging from cancer (due to IDPs' role in signalling) to amyloid diseases such as Alzheimer's, because some disordered proteins have a propensity to form ordered aggregates. In addition, the newly discovered role of IDPs in 'granule' formation may have implications for a number of diseases including ALS. Indeed, our meeting included several speakers whose work addresses disease-related aspects of IDPs, including Wolfgang Peti, Joerg Gsponer, Tanja Mittag and Simon Alberti. More generally, the meeting addressed the physical mechanisms behind the function of

IDPs, which should aid our efforts to treat the diseases in which they are implicated. Further, a detailed understanding of the role protein dynamics plays in determining function is becoming increasingly important also in biotechnological applications. The studies of protein disorder will also lead to insights, as well as methodological advances, that will play a key role in this area.

# Perovskite solar cells: the quest for a theoretical description

Organisers: Paolo Umari, Edoardo Mosconi, Filippo De Angelis, Jacky Even, Giacomo Giorgi
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 25 – 28 August 2015

**State of the Art:** The workshop focused on the theoretical aspects of a novel class of materials for photovoltaic (PV) oriented applications. **Hybrid organic-inorganic halide perovskites** (OIHPs) are materials with unique characteristics in PV as testified by the very high photo conversion efficiencies already achieved when used as light harvesters in solar devices.

In proposing the workshop, we clearly stated that in order to achieve a fundamental breakthrough in PV, we would require PCEs of ~ 20%, which is not an impossible target in the next years using OIHP technology. Actually, such goal has been achieved and surpassed in the few months between the proposal submission and the workshop (certified PCEs of 20.1% have been obtained by Seok group at KRICT labs) suggesting OIHPs as a major breakthrough.

State-of-art computational methodologies applied to materials science have shown their powerful predictive capabilities and their employment in the study of OIHP based solar cells has already partly reduced the initial lack of theoretical knowledge characterizing them. Theoreticians are expected both to discover and suggest viable and optimized procedures to assemble the final devices (towards the PCE limit imposed by Shockley and Queisser) and to predict environmentally friendly alternatives to the usage of toxic Pb (many countries prohibit indeed its use for residential applications).

**Outcomes:** The workshop has focused on all the most debated theoretical aspects of OIHPs (ranging from the chemistry of defects to the excitonic nature) and in particular few fundamental points, listed below, have been raised still deserving a conclusive answer and thus representing the possible future direction of the topic in the next few years:

1. Band gap and band edges can change by tenths of eV due to functional inaccuracy. At the same time at which level of accuracy is it possible to assess that there are no deep trap states for OIHPs?
2. The phase of the fully inorganic compound ($CsPbI_3$) occurs spontaneously in $CsPbI_3$. Suggesting possible coexistence of nanoclusters of secondary phases in the organic-inorganic compound.
3. Insulators have strong bonds, and semiconductors have weak bonds (HOMO-LUMO splitting). Strong absorption always goes along with fatally weak crystal structure? Coexistence of covalent and ionic nature of the bond.
4. Closer experimental-theoretical interactions. What are theoreticians really calculating? Comparability with experimental data.
5. Can OIHPs being considered closed shells?
6. Why did fluorinated molecules show so little promise? (Chemical instability).
7. Electronic properties. Comparison of the results obtained from the unit cell towards super cells. Relevant role of *ab-initio* Molecular Dynamics.

There are thus still many directions to follow and issues that need to be fixed, such as environmental suitability and stability. Recent encouraging results have already predicted alternative, Pb-free, materials.

The state-of-the art computational approach which resulted from the workshop is clearly a combination of Molecular Dynamics (including ab initio molecular dynamics) DFT and GW methods, the latter with inclusion of spin-orbit coupling. All these methods are computationally very intensive and require different computer architectures for optimal performances. In general, however, a parallel architecture with a reasonable number of "heavy" (in terms of CPU, memory and memory bandwidth) nodes (ideally vector/parallel machines or more practically linux clusters with high speed connection) may ensure a reasonable compromise and allow high-end applications.

Networking and training should definitely be implemented in this field, to ensure "best practice" procedures and to avoid inconsistencies in the data from different laboratories which have emerged during the workshop.

**Requirements:** Perovskites solar cells in general and their modelling are a relatively young fields (started in 2012-2013), thus fund-raising at the European level has been so far limited with little success. Given the breakthrough nature of this class of materials for solar cells and optoelectronic applications in general, a dedicated H2020 call is highly desired. In the most recent H2020 workplan drafts, such call will probably be announced at the end of 2016 or the beginning of 2017 but it will probably be an experimental and device driven call. A strong cooperative project on computational modelling of perovskites would be enormously beneficial to the large community working in the field, which was mainly gathered at the workshop. In this sense, any possible lobbying activity at any European level that CECAM may exert is most welcome and would definitely support the European leadership in this field.

**Will these developments bring societal benefits?** We firmly believe that many of the results we have collected and provided during the workshop may represent a first step that the photovoltaic European industry has to follow in the development of OIHP based solar cells. The present PV technology indeed still relies in crystalline silicon, and in the technologies based on the thin-film architecture that account for about 10% of worldwide PV distribution. Several key countries like US, Spain, Italy, China, Germany, UK, and Japan have boosted in the last two years their investments in solar PV plants with a global rise of new installations that has reached ~20% in 2014. In this urgent global quest for alternative energy, the development of a cheap PV cell is the essential goal and in this direction OIHP cells have been already demonstrated to represent the future mainly because of the reported low processing temperature of such devices combined with the extremely high efficiencies already achieved: not surprisingly, the quest for OIHP cell mass production has already started, largely anticipating the usual decennial R&D-to-production path for similar devices. Furthermore there are other very important yet unexplored fields where OIHPs have been shown to be promising: still related to the renewables, (i) multi-junction solar cells (the first 2-terminal monolithically-integrated metal-halide perovskite/Si multi-junction solar cell prototype not based yet on the use of the optimal materials has been reported to have a 13.7% efficiency), and also solar thermoelectric: for hybrid organic/inorganic Pb- and Sn-based perovskites, the thermoelectric figure of merit, ZT, is predicted to be ~1. All these facts show the PV industry is already extremely active in the field and that there is still a lack of fundamental knowledge that theoreticians might supply in the next few years to improve the potential of this class of materials: we are confident that this workshop has opened the way in this direction.

# Topological phases in Condensed Matter and Cold Atoms systems

Organisers: Didier Poilblanc, Eddy Ardonne, Nicolas Regnault , Roderich Moessner
Location: Institut d'Etudes Scientifiques de Cargèse (IESC), Cargèse (Corse) (France)
Date: 31 August –12 September, 2015

**State of the art:** The study of topological phases of matter has recently experienced a tremendous intensification, with much progress on both the experimental as well as theoretical side. Most notable are the newly discovered topological insulators (or superconductors), which combine physics from the quantum Hall effect and graphene. Currently, most of the interesting physics in topological insulators emerges from combining non-interacting band theory with the notion of topology, which has led to some spectacular results.

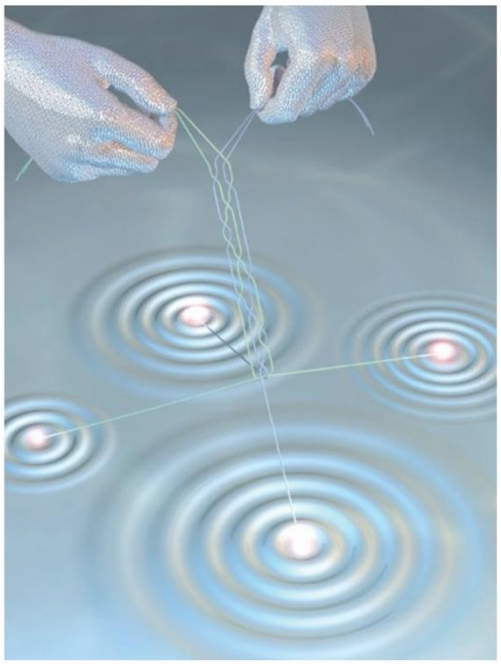


Figure 1. quantum operations by braiding quasi-particles

However, the fact that most of the developments in the field of topological insulators have focused on the effects of the topological properties alone means that consideration of the consequences of adding electron interactions are largely missing. While the latter give rise to very interesting physics in their own right, combining them with topological structures will most certainly lead to many interesting discoveries. The fractional quantum Hall effect is a prime example of where this interplay indeed has led to very exciting new physics.
For systems without interactions, the possible topological phases have been classified. Also, it is now known from some explicit examples, that taking interactions into account will change the picture. In particular, phases that appear to be distinct in the absence of interactions can become equivalent in the interacting case. Unfortunately, classifying topological phases in the presence of interactions is a daunting task, so that making even a little progress will greatly enhance our understanding of topological phases. This is the first main question that was be addressed during the workshop.

The second main question was in which realistic experimental systems interesting topological phases (be it interacting or non-interacting) can be expected to occur? Despite the recent promising results on realizing Kitaev's one-dimensional topological superconductor in several experiments, experimental realizations of topological phases with non-trivial excitations beyond the quantum Hall regime are few.

**Outcomes:** Our workshop has fully confirmed that the field of topological phases in condensed matter physics and atomic gases is an extremely active field, where theoretical (both analytic approaches and simulations) and experimental progress hand in hand. Since our previous CECAM workshop in 2013, we have observed significant progress. We managed to gather again world-class physicists with different backgrounds - numerical, theoretical and experimental - and with the common interest of topological phases of

matter. Reports of the very recent developments in the field were given and lively discussions have taken place. People left with interesting ideas from the discussions, which now being worked out in detail. The following important topics emerged:

a) Fractional Chern insulators: The classification of topological phases in the presence of interactions is progressing steadily. New phases on lattices, with no fractional quantum Hall state equivalent, have been discussed in the course of the workshop.

b) Topological phases in quantum magnetism: Rapid progress is made in the field of frustrated magnetism showing that a number of quantum spin systems in low dimensions may host topological spin liquids.

c) New phases in cold atomic systems: The versatility of cold atomic matter is continuing to supply suprises. With the recent experiments on many-body localisation in cold matter, the promise of realising various new types of orders predicated on many-body localisation (such as 'eigenstate order') is in the air.

d) New entanglement-based theoretical tools: a number of new tools borrowed from quantum information and based on entanglement (Entanglement Spectrum, Tensor Networks) have enabled advances in the understanding of correlated systems and topological orders.

**Networking & training advances:** We believe "Topo" workshops (at IES Cargèse or in a similar facility) every two years would serve the community well, as a regular platform to discuss the new developments in this rapidly growing field. We have managed to attract funding from a variety of sources, which has enabled us to put together an event allowing a considerable number of young scientists to benefit from the presence and availability of a strong roster of accomplished researchers.

**Identification of key research areas:** (for the EU Horizon 2020 program)

1. *The field of topological phases:* Topological quantum computation solves, by construction, the problem of local decoherence. Implementing topological quantum computation in realistic experimental systems is one of the holy grails of the community.

2. *Numerical simulations with theoretical guidance of these complex many-body systems*: density matrix renormalisation group and tensor network studies of cold atoms, topological spin liquids, fractional topological insulators, etc… are extremely dynamics fields currently being developed by our community.

**Will the results of these developments be of societal benefit?** The field of topological phases is still very much an area of fundamental research. While applications to the design of quantum computers have been proposed and are being pursued by research groups funded by Microsoft Corporation, it will be several years before this field and the research presented at the workshop will be of more immediate interest for commercial exploitation by European industry.

# Chemical and Structural Transformations in Materials under Mechanical Load

Organisers: James Kermode, Lars Pastewka, Gianpetro Moras, Alessandro De Vita, Mike Payne
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 1 – 4 September, 2015

**State of the Art:** Our current capability to model complex chemical-mechanical processes relevant for e.g. for tribology, fracture and plasticity in ceramics, oxides and metals was clearly identified at the workshop. Key strengths that were identified include the accuracy of quantum-mechanical (QM) simulations of realistic systems with several chemical elements, usually based on density functional theory (DFT). In parallel, accurate and efficient classical interatomic potentials exist for many unary and binary metallic and covalent systems. Important steps forward have been made in the development of free energy methods and their application to important entropically driven materials problems. The link from DFT to classical atomistic simulations is well established but there is clear scope for further improvement e.g., though the automatic generation of reactive potentials for three or more elements. Obtaining mechanistic insight was identified as a key strength of modelling. There was a strong consensus that our main focus should remain here, rather than attempting the direct simulation of systems with the full complexity typically encountered in engineering applications. Cross-fertilization between adjacent fields with similar problems is important and common, and close collaboration of experimental and theoretical teams is very encouraging. The interoperability of data and codes has improved markedly in recent years. The emergence of shared datasets is expected to capitalise on this, helping to ensure less time is spent recalculating the same properties. Mechanical properties of materials, such as friction, wear, fracture, plasticity, toughness or yield strength are still theoretical challenges and deserve future attention, in particular for complex multicomponent systems in contact with environmental chemistry. The field is moving toward advanced sampling techniques to capture phase space (free energies) and predict, e.g. unbiased phase diagram without an a priori knowledge to truly identify phase-boundaries. Computation of spectra for the interpretation of experimental data is already commonplace in the chemistry community and slowly permeating investigations of chemical-mechanical investigations.

**Outcomes:** Some relevant issues were identified. First of all, accessing realistic time-scales is still very much an issue. While time-acceleration techniques exist, these are limited and make simplifications that make them difficult to apply to system with structural heterogeneity, such as glasses (where they are most important because of slow relaxation or ageing processes and logarithmic strain-rate dependencies of driven processes). Furthermore, usefully accurate descriptions (with controlled uncertainty) of multicomponent systems that scale O(N) are still lacking, in particular for oxides and interactions with the environment (e.g., $O_2$ and $H_2O$). Expensive DFT calculations can capture chemical complexity but are typically too slow to be combined with phase-space sampling techniques. Our current capability for modelling the microstructural level material response is limited. Addressing this requires making connections to mesoscale theories such as non-equilibrium statistical mechanics. More generally, advances are inhibited by the limited knowledge that each community has of the techniques used by other communities. Cross-fertilisation of ideas from different experimental and simulation techniques should be accelerated (through workshops

and bilateral exchanges/visits) and more research projects should try to combine different simulation techniques to scale up results of small- scale techniques or validate the results of large scale methods. This is even more important considering the concern, expressed by the participants, that real systems are so complex that it is hard to disentangle individual processes in experiments. This also suggests that simpler experiments on model materials should be carried out where possible. Examples include experiments on steels, which are difficult to capture with atomistic techniques because of their chemical complexity. Experiments should, for example, start by looking at highly pure iron sample (which is not technologically relevant) and then introduce alloying elements in a controlled manner.Another concern in the emerging area of big data was the proper acknowledgment of data uploaded by a third-person and then used in research by another party, as well as indicators for the precision of simulation data. Data stored in databases should be reproducible. These questions all boil down to mutual trust with data quality and handling of data.
Some recommendations to CECAM also emerged during the workshops. In particular, the participants identified key roles that CECAM could play:

1. facilitate two-way exchanges/internships/sabbaticals between academia and industry, e.g. by maintaining lists of industrial contacts
2. host introductory materials modelling courses for experimental PhD students
3. establish challenge problems in canonical systems common to experimentalists and modellers
4. maintain a database of funding sources for researcher exchange and programs for academia/industry exchange
5. encourage workshop organisers to foster cross-fertilisation of different fields both in terms of
   experiment/theory and in terms of different techniques
6. support the sustainable development of software codes through hosting of coding efforts
7. finally, there was considerable enthusiasm for the establishment of a CECAM node in the USA.

**Community needs:** A specific concern was the development and long-term maintenance of codes. It was suggested that CECAM could play a role here in aiding development of software. Software specialists that increase the efficiency of codes would also be helpful. Another concern was ensuring a sustainable career structure for software specialists both inside and outside academia. Databases should be created to publicly save and share experimental and simulation data. It was particularly noted that providing modellers with access to direct numerical results from experiments has the potential for significant impact.

**Funding** The long-term maintenance of codes requires significant funds that are not currently provided by the major funding councils. Academic institutions and funding agencies should enforce the formation of working groups composed by theorists and experimentalists as well as a focus on mechanisms and not on specific techniques. This would encourage collaboration between experimentalists and theorists. In this context theorists should be encouraged to propose and design experiments, and experimentalist should propose and design simulations.

**Will these developments bring societal benefits?** The main societal benefits were identified in technologically driven collaborations between academy and industry. Industrial workshop attendees from ExxonMobil, BP and BMW informed a discussion on how to enhance academic engagement with industry, e.g. through network forming opportunities. It's important that these should also include PhD students to aid them in searching for jobs. There should generally be clarity on the expectations from both sides

and academia should make realistic promises of what can be achieved within the timescale of a PhD or postdoc project. Challenging industrial problems should inspire methodological development rather than being viewed as special cases of previous academic work. Another mechanism to facilitate collaboration between academia and industry should be opportunities for co-funding of small projects. The French system was cited as having a useful program where single PhD positions could be co-funded 50% by industry and the government. EU Horizon 2020 or German BMBF projects were considered too large to be useful for efficient targeted research. Industry sabbaticals were cited as opportunities for more senior scientists. However, the workshop participants expressed some concern that over-concentration of resources on industrial R&D problems could ultimately be self defeating, since people trained to solve problems actually constitute a major output of blue sky academic research. Explaining to funders and stakeholders when ideas have arisen which were unexpected at the start of projects may help.

# Molecular and Chemical Kinetics

Organisers: Frank Noé, Guillermo Pérez-Hernández, John Chodera, Vijay Pande, Christof Schütte
Location: CECAM-DE-MMS, Zuse Institute Berlin (Germany)
Date: 7 – 9 September, 2015

**State of the art**: In the past few years, experimental evidence has mounted for the complexity of molecular kinetics in macromolecules, such as the existence of multiple long-lived intermediate states in biomolecular folding or binding processes, conformational selection, etc. These findings have been accommodated by a substantial advance in molecular simulation power (especially GPU-based computing and grid computing, but also special-purpose machines such as Anton) and methods to enhance the sampling of rare events without forcing the system (Markov state modelling, transition interface sampling, etc.). The simulation is field is now able to reach the millisecond timescales using atomistic models with affordable hardware. We are thereby for the first time able to explore biologically relevant processes such as protein-ligand binding, functional conformation changes in protein, ion transduction, etc. with unbiased all-atom MD simulations and deliver a yet unseen amount of mechanistic detail that will inspire new experiments and help achieving technological goals such as drug and material design.

While reaching the biological timescales, many molecular simulation labs now turn to the question what happens on longer length-scales. How are drugs distributed in the cell, how and where are ions buffered, how do proteins interact in time and space to give rise to cellular signal transduction. Particles-based simulation methods, such as particle-based reaction-diffusion (PBRD), Greens function reaction dynamics (GRD) and others have been developed to meet this challenge computationally. In parallel, new experimental techniques have come of age that provide detailed and quantitative data on protein locations and numbers, such as particle tracking techniques and super-resolution microscopy. Such techniques heavily rely on computational modelling and analysis to yield high-quality data from the raw image data.

The time has come to seek for systematic and rigorous methods (computationally, experimentally, and on their interface) that combine molecular kinetics and chemical kinetics of the cell. How can we systematically model, simulate and analyse processes that on one hand need atomistic resolution (at least in parts of the system), but are also able to describe transport and sorting processes amongst thousands of diffusing and interacting protein copies?

**Outcomes:**

- The field of molecular kinetics modelling is rapidly expanding, mostly due to the requirement to analyse increasingly large amounts of molecular simulation data. The models are becoming more robust, we understand their mathematical and numerical properties better. The series of CECAM Molecular Kinetics conferences have significantly helped to catalyse this process.
- A number of key developments have emerged in the recent years in the field of molecular kinetics modelling. These include: (a) Markov state modelling, (b) the variational approach for conformation dynamics which allows the approximation of molecular kinetics as an eigenvalue problem, (c) time-lagged independent component analysis (TICA), a special case of the variational approach that is well-suited to do dimension reduction of high-dimensional molecular trajectories

while preserving the kinetics, (d) methods such as TRAM and DHAM that allow kinetic models to be obtained from simulations at multiple thermodynamic states. Much of the current research in the field is revolving around these key technologies.

- A few relative complete codes to analyse MD data exist and are now heavily used, in particular PyEMMA, MSMBuilder and MDTraj. These codes also represent a new and dynamic programming model (Python / C or C++, open source, github, active community development) that is likely to dominate code development in the next years.
- There are first steps in trying to combine molecular dynamics-based models with reaction kinetics models in order to bridge the scale gap between molecular and cellular or supramolecular systems. This research is still at its infancy but it is envisaged that it will have a similar key role akin to kinetic modelling itself and it is worth investing in this direction - both in terms of research projects as well as workshops.
- We are experiencing at a tighter integration between experimental techniques (both addressing the intramolecular dynamics and the supramolecular arrangement or localization) and simulation techniques (both molecular dynamics and reaction-diffusion). The aim at being quantitative and comparable is found in both theoretical and experimental groups, and there are an increasing number of groups that combine both techniques.

**Community needs:**
Community needs include the following

- Availability of more automatic, efficient and reliable codes to construct and interpret kinetic models.
- New methods for multiscale simulation in order to explore molecular effects on the cellular scales.
- Methods that allow the effect of mutations of molecular systems to be efficiently predicted, and that allow small molecules to be designed that have specific effects (e.g. inhibitors).

**Will these developments bring societal benefits?**
Potentially, yes. Applications include drug development, vaccine development, biotechnology and nanotechnology applications. Thus these methods may have a beneficial effect on health and well being.

## Computational approaches to chemical senses

Organisers: Alejandro Giorgetti, Paolo Carloni, Luciano Navarini, Giulkia Rossetti, Pablo Campomanes
Location: CECAM-DE-JUELICH (Germany)
Date: 9 – 11 September, 2015

**State of the art**: Food perception relies on the chemical senses receptors of olfactory and taste systems. When tastants and odor molecules bind to their target receptors, they activate downstream events whose final result is the production of ion currents and signalling to the brain. Computations, especially if combined with molecular biology and genetic studies, currently are the method of choice for molecular-level investigations of these fascinating systems.

**Outcomes:** The workshop fostered the meeting between simulators and experimentalists from food companies and research institutions. The field is young and many of the top researchers met here for the first time. This led to unprecedented insights and debate in the field. Several issues were vividly addressed. These include:

- Problems related to the paucity of experimental data in the field and possible solutions to them were discussed.
- The need for multiscale modelling was clearly established. Computational systems biology and mesoscale simulations need to be integrated with molecular simulation. The latter cannot provide a physical basis, at present, of the readout of the signaling process (ion current). A multi-scale approach will trigger appropriate projects and financial investments that may lead to the establishment of novel research directions capable of maximizing the impact of the field in the near future.
- The gap with the needs of food industry was identified and ideas to bridge it were discussed. The talks of industry scientists (Illy, Barilla) clearly showed that their challenges for improving food are currently very difficult to be met by the molecular simulations community. Strengthening interactions with them seems the best way to start addressing this stringent issue. At the moment, these interactions are rather weak.

**Networking & training:** A new, mixed industrial/academic community has been formed. We decided to meet on a regular basis, and possibly to apply together to EU funding

**Funding:** The workshop identified research areas that might be funded by the EU H2020 programme particularly through the MSCA joint PhDs programmes. Participants would support an application involving labs around EU and Israel able to train young researchers in combined multi-scale computational/experimental approaches applied to chemical senses.

**Will these developments bring societal benefits?** Progress in the field of computational chemoperception is of great interest to EU Industry. In fact this meeting has been co-organized together with industrial partners and some of the speakers, actively participating at the different meetings carried out during the workshop belong to industry. (e.g. Illy, ChemCom, Barilla). Specific industrial sectors identified are: food, beverage production and perfumery.

# From trajectories to reaction coordinates: making sense of molecular simulation data

Organizers: Christoph Dellago, Peter Bolhuis, Gerhard Hummer
Location: CECAM-AT, University of Vienna (Austria)
Date: 16 – 18 September, 2015

**State of the Art:** As computational science is playing an increasingly important role in all fields of science and engineering and even the social sciences and the humanities, enormous amounts of data are produced each day by computer simulations running on computers around the world. In many areas, the challenge no longer is to create data, but to make sense of them. Molecular simulation has been dealing with such problems for some time, for instance when studying rare transitions between long-lived states such as the folding of a protein or the freezing of a liquid. Given detailed observations from computer simulations or from experiment, how does one extract a reliable description of the mechanism of the process? Such mechanistic understanding is expressed in terms of low-dimensional models that capture the essential features of the process under study.

The problem of constructing dynamical models from trajectory data is tied to two classic sampling problems in statistical mechanics, i.e., the calculation of the populations of metastable states, expressed in terms of free energies, and the sampling of the rare events associated with transitions between the metastable states. For these purposes, a broad range of methods have been developed that allow us to explore conformation space and to reach time scales much more efficiently than with direct unbiased simulation. These enhanced sampling methods typically rely on advance knowledge of relevant coordinates. Order parameters, for instance, are used to delineate the boundaries of the metastable states. Reaction coordinates measure also the progress of the transition in a dynamically relevant way.

With reaction coordinates occupying such a central role in both the sampling and of simulation of trajectories and their mechanistic interpretation, it is no surprise that finding them has attracted much attention. Formally, the progress of a reaction can be measured using the so-called committor. However, this quantity is both hard to determine and difficult to make sense of. In practice, therefore, one often seeks simple functions in terms of physically meaningful quantities that approximate the committor. Over the last couple of years, a number of methods have been developed to find good reaction coordinates. These methods are based on a variety of approaches including neural networks, maximum likelihood, transition path theory, path reweighting, Markov state models as well as dimensionality reduction.

The central goal of this workshop was to define and advance the state of the art in constructing reliable and meaningful models from data obtained from molecular simulations, and in using such models to gain mechanistic insight as a guide for applications and further study. Furthermore, the workshop aimed at pooling the cumulative knowledge of diverse but related fields. Applied mathematicians interested in transition path theory, computer scientists interested in machine learning, physicists interested in phase transitions, chemists working on reaction mechanisms, and biologists elucidating the function of molecules s are representative of the many fields benefiting from an improved understanding of reaction coordinates. A main goal of the meeting therefore was to enhance communication in this diverse community, with the hope of developing new ideas and computational approaches.

**Outcomes:** The following topics emerged as central from the presentations and the Q&A sessions of the workshop. These issues were also at the focus of a round table discussion, which took place within the meeting.

*One-dimensional models, committor as reaction coordinate:* As the committor can in principle serve as the perfect reaction coordinate, the question arises, whether the dynamics in this coordinate (or in any coordinate that can be simply mapped into the committor) is Markovian such that the kinetics of the process can be easily analyzed in terms of the free energy landscape. Beyond this issue of principle, from a practical point of view the question arises if it always feasible to adopt a one-dimensional description of the dynamics based on a single (but possibly complicated) reaction coordinate or whether it is preferable to construct higher dimensional models such as Markov-State-Models. In this context an important recent publication was pointed out (A.M. Berezhovski and A. Szabo, J. Phys. Chem. B 117, 13115, 2013) and its significance was discussed at the workshop. In this paper, the authors demonstrated that if a multidimensional dynamics is projected onto the committor, the resulting diffusion equation yields the correct flux even if it does not correctly describe the dynamics. The question of the validity of one-dimensional models is of particular importance because, as pointed out by the experimentalists participating in the workshop, such models are very popular for the interpretation of experimental data.

*Machine learning methods for finding reaction coordinates*: One topic intensely discussed at the workshop was the use of machine learning tools to identify relevant variables necessary to describe the dynamics of the system. Given the rapid increase in computing power, such tools and related dimensionality reduction approaches (for instance diffusion and sketch maps) will be of crucial importance in the future. The hope is that on this point one will be able to learn from other fields in which machine learning methods are used in big data applications.

*Trajectory sampling:* A recurring issue in the discussions at the workshop concerned the possibility of understanding the dynamics of the system in the relevant variables in terms of free energies alone. In other words, when do we need to sample trajectories and when can we get away sampling solely configuration space and then adding the stochastic dynamics on top of the free energy surface? Furthermore, particular complications both for the sampling but also for the construction of simplified models arise if a reaction can proceed via several different pathways.

Resolving these issues is a major current challenge for the simulation and modeling of molecular processes.

**Funding, attendance and future plans:** The workshop, organized at the Vienna CECAM-Mode DACAM and hosted by the Erwin- Schrödinger-Institute for Mathematics and Physics (ESI) at the University of Vienna, was funded jointly by CECAM (12.000 EUR) and the ESI (6.000 EUR).

The announcement of the workshop generated much interest in the community. In fact, the organizers received many more requests for workshop participation than they could accommodate. Including local participants (mainly students and postdocs from the simulation groups at the University of Vienna and the Technical University of Vienna), the workshop was attended by about 70 people, which is probably too large but this number is due of the overwhelming interest in the workshop. About 50 applications for participation had to be turned down. In response to the great interest in the topic of the workshop, the organizers have submitted an application for a Lorentz

Centre/CECAM workshop on “Reaction coordinates from molecular trajectories”. This workshop, which is intended to focus on the central issues that emerged from the Vienna workshop, is planned to take place in 2016.

# A Scientific Roadmap for Simulation and Modelling for 2020

Organisers: Sara Bonella, Ralf Everaers, Dominic Tildesley
Location: CECAM USI-Lugano (Switzerland)
Date: 16 – 18 September, 2015

**State of the Art:** The second CECAM strategy meeting "A scientific roadmap for simulation and modelling for 2020" was held in Lugano, on September 16-18 2015. Work was organized in three main sessions: Scientific Strategy, Training, and Industry. These sessions started with a short introduction by a moderator, followed by discussions focused by prepared questions and conducted by all participants divided in four work groups. The results of these discussions were then summarized and further common discussion carried out. The final session of the meeting focused on the presentation of the E-CAM Centre of Excellence and on gathering conclusions from the previous days. The main points that emerged, and suggested actions, are reported below.

**Outcomes:** A full report of the meeting is available. The final discussion of the established the key actions for CECAM moving into 2016. Note the bullet points marked in red are a high priority for 2016.

**General**

- Develop clear communication to Council, Nodes and Funders regarding the expansion in the range of activities of CECAM and the relationship between E-CAM and CECAM
- Commission further work on the CECAM web site using some professional design support. Establish real clarity of the workshop, software, consultancy and training programme available to all interested partners through the web site.
- Simplify and clarify the methods of application for CECAM funds.
- Continue to drive gender diversity at meetings through schemes such as childcare and by setting action standards with all event organisers.
- Effect the change in the CECAM programme year from April to April

**Improving the quality of workshops**

- Establish balance between top-down strategic pushes and bottom-up events. Insist on a diverse but expert attendance at workshops from as many member organisations as possible.
- Improve scientific advise on workshops to Director by strengthening and enlarging the role of the Scientific Advisory Committee.
- Avoid too many repeated workshops by raising the bar for obvious repeats and by combining proposals where possible.
- Encourage more disruptive or alternative events under the workshop label. Continue to push the distinctive nature of the CECAM workshop as doing and discussion events.
- Try some smaller focus and discussions meetings (1 day, 1 night events).
- Continue to work with Psi-k and the Lorentz Centre to establish joint events.

**Extending the range of workshops**

- Do not extend the subject range too far into other fields (e.g. astrophysics, QCD)
- Seek more engineering and computational fluid dynamics workshops

- Continue to push into the life-science space with meetings in network and systems biology through call letter.
- Continue to drive workshops around important data-driven applications.
- Develop workshops around modern software quality assurance methods.

**Training**

- Enhance the visibility and impact of the existing CECAM annual program of ~20 tutorials and schools by improving the presentation of training on the website.
- Provide permanent access to high quality, up to date teaching material as developed for the CECAM tutorials and make this material widely available to the community.
- Centralize the advertisement of student internships to be held at CECAM-HQ and the nodes
- Foster community based initiatives towards grant applications for Erasmus Mundus Master Course / Joint Doctoral programs and Innovative Training Networks.

**Industry**

- Provide training and consultancy to industry through the E-CAM grant activities
- Look to enhance the number of active industrial partners by five per annum
- Increase the attendance of industrial researchers in simulation and modelling at CECAM workshops and schools
- Provide access for industry to the CECAM software repository under an agreed licensing model

# Exploration of ultra-fast timescales using time dependent density functional theory and quantum optimal control theory

Organisers: Sangeeta Sharma, E.K.U. Gross, Uwe Bovensiepen
Location: CECAM-HQ, EPFL Lausanne (Switzerland)
Date: 28 September – 2 October, 2015

**Abstract:** Manipulation of electrons by ultra-short laser pulses opens up the vast and largely unexplored physical landscape of ultra-short time scales. One possibility in this landscape is to use electronic spins, which can be optically manipulated (flipped) using lasers to store data as binary bits. The advantage of such a technique is obvious: reduction in the speed and size of the storage device by orders of magnitude. Recently, great advances have been made in the production of short-time laser pulses and this has lead to very fast flipping of spins and modulation of charge in materials. However, we are still far from optimal manipulation and control of spins, needed for device production. Two major reasons for this are (a) the lack of theoretical understanding of the phenomena behind this spin- light interaction induced spin-flip and (b) the lack of unambiguous experimental access to the fundamental spin-dependent processes on the ultrafast timescales of few femto- to attoseconds. The CECAM-Psi-k workshop was aimed at getting together scientists exploring this field of "ultra-short time-scales" to discuss recent advances and areas of lack of understanding.

**State of the Art:** Ultra-fast light-induced demagnetization was demonstrated in 1996, where demagnetization times (in Ni) faster than a few picoseconds were achieved using intense electromagnetic pulses. Recently, these spin-manipulation times have been reduced to a few femtoseconds, owing to great advances made in the production of short-time laser pulses. However, we are still far from achieving sufficiently controlled manipulation of spins required for production of useful devices. Two main reasons behind this are the lack of full understanding of the phenomena leading to demagnetization and limited theoretical as well as experimental tools to investigate spin dynamics on such ultra-fast time scales. This field started with two pioneering pump-probe experiments, which reported magnetization changes on specific characteristic time scales. The experiments by Beaurepaire *et al.* investigated Ni and found that the magnetization changes much faster than previously predicted by spin- lattice coupling. This opened a new path of manipulating spins, under non-equilibrium conditions, order of magnitudes faster than at thermal equilibrium. Since then several dedicated pump-probe experiments (including magneto-optical, x-ray, and photo-electron techniques) have demonstrated the same. However, the underlying fundamental physics still remains a matter of debate. One reason behind this is that the current experimental approaches are essentially limited to response of the material in a few 10s to 100fs. In this regime a complex interplay of spin, charge, and lattice vibrations implies a difficulty in unravelling the mystery behind each process separately, which is essential to achieve desired microscopic understanding of the process.
Theoretically, there have been a number of attempts at explaining this optically induced spin-dynamics– combined action of spin-orbit coupling and interaction between spins and laser photons. Super-diffusive spin transport occurs where excited electrons carry spin with them from one part of the sample to another. Elliott-Yafet mechanism where electron-phonon or electron-impurity mediated spin-flip is the major contributor. However, all these attempts to understand the physics of spin modulations by light have one thing in common– they are all based on model Hamiltonians. The problem with this is that models based on totally different underlying physics are seemingly able to explain the same experimental data, rendering it impossible to separate fact from fiction. What

one requires in such a situation is a fully *ab-initio* theoretical description. Time-dependent density functional theory (TD-DFT), which extends density functional theory into the time domain, is a formally exact method for describing real-time dynamics of charge and spin under the influence of an external field like an applied laser pulse.

Given this situation following questions needed a detailed exploration and were discussed during the workshop:

1. How TD-DFT can be used to understand the phenomenon of demagnetization?
2. How various models can be designed based on highly accurate TD-DFT calculations?
3. How these models can then be used to study large systems?
4. What kind of information can experiments provide in this endeavour?
5. How can optimal control theory be used to control and manipulate spins?
6. How can the time resolution in experiments be pushed to even shorter times scales?
7. What are the suitable materials to investigate in order to facilitate microscopic insight?

**About the workshop:** All talks were given 1 hour, however, the speakers were allowed to talk only for 35 minutes. A 25 minute long discussion time provided at the end of each talk to discuss problems associated with the field covered by the talk led to many important points being identified. Speakers were specifically requested to present not just the fantastic new results, but also problems and brick-walls encountered in their specific area. This led to many future collaborative developments in this field viable. The main topic discussed during the workshop were:

1. An overview of the experimental and theoretical developments in manipulation of charge and spins in ultra-short time scales

2. Model Hamiltonians and established experimental approaches

• What can we learn about the physics at ultra-short times scales from model Hamiltonians
• What are the crucial open questions considering the experimental information gained so far?

3. Ab-initio theory and recent experimental developments

• The ab-initio theory for study of laser induced charge and spin-dynamics, namely time-dependent density functional theory
• What is the underlying physics described by TDDFT
• What are the limitations of TDDFT and what is required to overcome these limitations
• What are the recent experimental developments and how theory can meet these experiments

4. Optimal control of charge and spins

• How can TDDFT be combined together with optimal control theory to manipulate spins using specially designed laser pulses
• Can ab-initio calculations provide parameters for model Hamiltonian for studying systems at large length scales?
• What are promising materials or material combinations and what are the suit- able experimental methods to reach controlled manipulation.

# Virus as a whole: meso- and macroscopic structure and dynamics at all atom resolution (CECAM-Lorentz Workshop)

Organisers: Dmitry Nerukh, Sergey Karabasov , Anton Markesteijn
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 6 – 10 October, 2015

**State of the Art:** State of the art computers can simulate liquid molecular systems several hundred million atoms in size (tens of nanometres across) using classical molecular dynamics (MD) methodology. Experimentally, modern x-ray crystallography can measure atomistic structure of complete parts of living cells with up to ~0.1Å resolution (for example, human ribosome [A. Ben-Shem, Science (2011)], photosystem complex [Y. Umena, Nature (2011)], entire virus [P. Plevka, Science (2012), R. Khayat, J. Virology (2011)]). These two facts together imply that it is now possible for the first time to model the structure of an entire virus at atomistic resolution. Even though at the moment there are no viruses whose complete structures are known (with the nucleic acid inside), it is a feasible task if the latest experimental data from cryo electron microscopy is used. However, simulating the dynamics of such systems over biologically interesting times is still impractical (up to 1μs simulation is reported [D.S.S. Larsson, PLoS Comput Biol (2012)]). The **goal** of this workshop was to bring together methodologies for computer modelling of viruses at their entirety in water at multiple scales in space and time: from all atom resolution of the core and critical surrounding water to macroscopic mechanical elasticity of the shell, from atomistic motion of individual atoms to hydrodynamic flows at macroscopic distances.

**Outcomes:** We have organised the workshop such that a set of "miniprojects" were under active development. These formulated the most active specific directions in the field and possible ways of solving the main problems in each one. In the appendix there is a list of participants involved in each miniproject.

1. *DNA inside PCV2.*

One of the central topics of discussion was the problem of modelling DNA inside a capsid of a small virus PCV2. Reza Hayat and Jade Forwood reported on experimental measurements of the capsid structure and partial information on the DNA inside the capsid. Dmitry Nerukh, Elvira Tarasova, and Ivan Korotkin presented the results of MD and hybrid MD/hydrodynamics simulation of the capsid. The next step in this field will be in simulating the DNA inside the capsid to complete the all-atom structure of the virus. It has been recognised, that at the moment there are no reliable tools to recreate the structure of DNA with the limited experimental information available to date. Various possible protocols of simulations have been suggested.

2. *Water/ion exchange across capsid*

Water and ion exchange across capsids relates to physical chemistry, physiology, and biology of capsids, and is partially responsible for mechanical property of capsids. Thus it is much valuable to quantify the amount of water and ion exchange per unit time using the data obtained by all-atom molecular dynamics calculation. There are a few difficulties in evaluating number of water molecules exchanging across the capsid.
Three main problems have been discussed and possible solutions are suggested (see the appendix for details).

3. *Viral Disruption of Membranes by Peptides*

Many viruses contain a membrane lytic peptide that serves to help the virus escape from the endosome. Examples include adenovirus protein VI and human papillomavirus L2

protein. Within adenovirus protein VI there is a predicted amphipathic alpha-helix that has membrane disruption properties. Within papillomavirus L2 there is a 14 amino acid stretch that has a hydrophobic region followed by a basic region. Both peptides are thought to disrupt membranes by a carpeting mechanism without formation of a pore. In the case of adenovirus, about 300 copies of protein VI are contained within a single virion, while for papillomavirus there are a maximum of 72 copies of L2 per virion. Molecular dynamics (MD) simulations have been performed on model membranes (for examples see references 1-2). Dissipative Particle Dynamics (DPD) simulations have been performed for lipid vesicles (ref 3).
At the workshop we discussed the feasibility of performing MD or DPD simulations to investigate the action of viral peptides on membranes. MD simulations would have the benefit of using full atomistic models, however this approach would limit the accessible time scale of the simulation. DPD simulations would utilize simplified molecular models, however the time step of the simulation would 2 or 3 orders of magnitude larger. Possible starting models could be either a small patch of lipid, a complete lipid vesicle, or a partial lipid vesicle with periodic boundary conditions. The desired outcome would be the observation of the disruption of the lipid bilayer. (see appendix for references).

4. *Adaptive Resolution Simulation (AdResS) and its application for DPD – coarse-grained MD hybrid simulations*

Current issues of AdResS scheme and its expansion to DPD – coarse-grained MD (Martini, in particular) hybrid simulations were discussed1–3. In particular, we considered the problems of allowing passage of a polymer chain through hybrid region. At a current stage this question requires further exploring, especially on a force-field level: how do we transfer bond-length and/or angle-bond interactions from atomistic to coarse-grained region?
An expansion of AdResS to DPD – Martini hybrid simulations was considered to be possible, as both share the same length scale. Within this frame two most important questions are put under consideration: 1) the application of a multiple timestep method (as DPD and Martini don't share the same time scale); 2) conjugation of force fields of different nature. Solving of a second problem can require the application of an external pressure to equilibrate two regions. As for the first issue, we think that multiple timestep might be implemented during splitting the system into cells. However, an exact implementation of such method would be different for standard MD packages.
We have arranged some further collaboration via email exchange.

5. *Mori Zwanzig generalised Langevin*

Dmitry Nerukh, Rafael Delgado-Buscalioni, and Anton Markesteijn have discussed the possibility of representing liquid matter between the atomistic and hydrodynamics scales using generalised Langevin equation of motion approach. It has been concluded that there are conceptual difficulties on this route, which are likely to make this idea difficult to implement.

6. *Disassembly of the capsid*

An interesting idea was put forward in connection with the problem of virus self assembly modelling. Instead of modelling the assembly of the virus, the process of disassembly can be modelled. The latter is much more feasible for modelling as it does not require subtle alignment of the virus parts without which self-assembly is impossible. At the same time the forces that keep the virus together can be investigate during the simulation.

**Software infrastructure:** One of the most important discussions of the workshop was the hardware for very large-scale molecular simulations. In particular, the K-computer, one of the fastest machines in the world was discussed, and MDGRAPE, a specialised accelerator for MD simulations. Their use for hybrid MD/hydrodynamics was discussed.

**Will these developments bring societal benefits?** Simulation of viruses as a whole has very significant potential implications to drug design industries. Understanding the atomistic details on the mechanisms of self-assembly, infection and replication opens up the routes to developing effective drug molecules that could control these processes. Moreover, the effective hybrid implementation of such simulation would provide access to these modelling methods to a wide community, when 'personal supercomputer' would be enough to perform complicated simulations of large biological systems like viruses and cell organelles.

## Appendix

### Miniprojects discussed at the workshop

| | title | participants | When discussed |
|---|---|---|---|
| 1 | **DNA inside PCV2**: | Reza Khayat<br>Daniel Buzo<br>Jade Forwood<br>Kate Smith<br>Justin Flatt<br>Phoebe Stewart<br>Dmitry Nerukh<br>Elli Tarasova<br>Makoto Taiji?<br>Teruhisa Komatsu?<br>Yoshimichi Andoh? | Wed lunch time |
| 2 | **Mori Zwanzig generalised Langevin**: | Rafael Delgado-Buscalioni<br>Dmitry Nerukh<br>Anton Markesteijn | Wed first coffee break |
| 3 | **Disassembly of the capsid**: | Urs Greber<br>Dmitry Nerukh | Wed Coffee break after lunch |
| 4 | **Water/ions flow through the pores**: | Yoshimichi Andoh<br>Elli Tarasova | Wed after talks? |
| 5 | **Hybrid AdreSS/Martini - DPD** | Alexander Muratov<br>Justin Flatt<br>Jan Peters | |
| 6 | **Viral disruption of membrane by peptides** | Daniel Buzo<br>Phoebe Stewart<br>Urs Greber<br>Alex Muratov | |

### Water/ion exchange across capsid

The first problem is too much amount of data in MD trajectory. Usually, AA-MD calculations of viruses treat over 1 million atoms in total, and output a few MB trajectory data per interval (one frame). So direct analysis of water exchange events on the outputted data consumes large computational time. To solve this, information of water should be coarse-grained: Mass centers of each water molecule are to be calculated and

saved to a separate file per frame of MD trajectory. It makes the analysis more handy, and debugging of analysis program more easily.
The second problem is the capsid rotates and diffuses in simulation time. Then, it becomes troublesome to calculate relative distances of each water molecule from capsid surface if using original Cartesian coordinates of atoms. To solve this problem, as done in the RMSD calculation, all coordinates are fit to reference structure (X-ray crystalline structure, for example). Then, analyses on water positions are performed. It is better to set the origin of new coordinates to the mass centre of capsid, because the following analyses consider polar coordinates measured from capsid mass centre.
The third one is a definition of boundary of inside capsid and outside capsid. There are many ways to define the boundary. The way we adopted is as follows: Firstly, radius, zenith angle, and azimuth angle are discretized. Then, one water molecule is selected, and its polar coordinate $(r,\theta,\phi)$ is calculated. Based on the calculated polar coordinates, the water molecule is assigned to a bin of 3D meshed polar coordinates. Doing the same analysis for all water molecules and time averaging over MD trajectories (frames), a histogram of water number of total space is obtained. Further, by dividing each bin's volume (discretized Jacobian of polar coordinates), yhe number density of water molecule is calculated. We defined a point of boundary along one radius as the minimum of water number density along the radius selected. By connecting every boundary points, boundary surfaces, which divides inside capsid and outside capsid, are determined.
Finally, the number of water molecules exchanged across the capsid is evaluated as follows. The method is very similar to the way to evaluate lifetime of hydrogen bonding between water molecules. At first, one frame of MD trajectory ($t=t_0$) is selected, and then the all water molecules are assigned to inside capsid region or outside capsid region using the boundary surface we defined above. Next, subsequent frames ($t=t_0+ndt$) are selected by assigning all water molecules to two kinds of region (inside/outside the capsid). The water molecules moved from inside to outside is defined as that it was at inside at $t=t_0$ and at outside at $t=t_0+ndt$.
The definition of the water molecules moving from outside to inside is the opposite. By taking time average of $t_0$ over equilibrated period, a plot of number of transferred water molecules as a function of time N(t) is obtained. An average of increments of N(t) with each interval, dN(t), gives the rate of water exchange as discussed in our paper and its supplementary material [JCP,141,165101(2014)]. The same kind of analysis is valid for non-equilibrated trajectories. However it should be noted that boundary of the inside/outside of the capsid is time dependent in such a case. Thus, the calculated N(t) may contain large error, which prohibits us from detailed discussion on non-equilibrium water exchange across the capsid.

**Viral Disruption of Membranes by Peptides**

References

1. M.A. Cherry et al., J Phys Chem B 118, 12462 (2014).
2. E. Drolle et al., Biochim Biophys Acta 1828, 2247 (2013).
3. J.C. Shillcock and R. Lipowsky, J Chem Phys 117, 5048 (2002).

**Adaptive Resolution Simulation (AdResS) and its application for DPD – coarse-grained MD hybrid simulations**

References

1. R.D. Groot and P.B. Warren, Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation, J. Chem. Phys. 107, 4423 (1997).
2. S.J. Marrink et al., The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, J. Phys. Chem. B 111, 7812 (2007).
3. M. Praprotnik et al., Adaptaitve Resolution Smulation of Liquid Water, J. Phys. Condens. Matter 19, 292201 (2007).

# Development next generation accurate approximate DFT/B methods

Organisers: T. Frauenheim, M. Elstner, F. Spiegelmann
Location: CECAM-DE-MM1P, University of Bremen (Germany)
Date: 11 – 16 October, 2015

**State of the art:** The past 15 years have shown that approximate quantum methods represent an essential part of computational methods for a solid atomistic understanding of a broad range of physical, chemical and biological problems. Besides the semi-empirical molecular orbital methods (MNDO, AM1, PM3, OMx), the self-consistent-charge density-functional based tight-binding method (DFTB) has been developed as an alternative approach derived from density functional theory (DFT) by careful neglect, approximation and parameterization of interaction integrals http://www.dftb.org/. Although being (usually) less accurate than DFT and ab initio methods on average, their main advantage is a greater computational efficiency, which can be 2-3 orders of magnitude faster when compared to DFT (and Hartree-Fock) using medium sized basis sets basis setsles access to both larger system sizes and/or sufficiently long sampling times for more meaningful QM or QM/molecular mechanics (QM/MM) coupling for molecular dynamics. Advanced functions in DFTB include spin-polarized DFTB for collinear and non-collinear magnetism, time-dependent TD-DFTB with both linear response and in the time domain, London dispersion corrections (LDC), LDA+U, various QM/MM-couplings and non-equilibrium Green's function DFTB for charge transport. This makes the method attractive to broad range of applications in diverse fields of sciences. At the same time further systematic corrections yield considerable improvements in accuracy for the description of ground state and excited state properties and the construction of electronic DFTB integrals throughout the periodic table, supporting numerous future applications.

**Outcomes:** The Workshop was successful in interconnecting the international leading core groups involved in the development of the DFTB-software and data generation of electronic Hamiltonians and repulsive energies but also various applications in materials physics, chemistry and molecular biology. The official Workshop program consisted of 30 invited talks of 40 minutes (30+10) each and a poster session presenting 29 contributed posters. In addition, social events were held (including a reception and conference dinner) to allow for informal exchanges. The invited talks were given by well-established scientists and DFTB-developers, which acted as platform for interesting cross-/interdisciplinary discussions. The invited talks were followed by a poster session where the younger participants could show their scientific work and progress in DFTB development and application and exchange of ideas with a broad knowledge in computational chemistry, solid state physics and computational materials science.

**During the official workshop, the following emerged as key topics:**

1. Semi-automated parameterization effort: DFTB relies on predetermined parameters, which are usually difficult to determine and rely on large molecular databases. A key issue in the future will be the handling of databases to allow for an efficient (re-) parameterization of DFTB models with growing methodological complexity. A key question is also how to combine the parameterization efforts for materials with those for molecules.
2. New chemistry/higher accuracy: A successful application of DFTB in future will require a more complex design of the basic model. This clearly involves the use of hybride and range-separeted exchange and ppRPA correlation functionals, as well

as extensions leading to a better description of non-bonding interactions or transition metal complexes. Since a modification of the Hamiltonian requires a reparameterization, efficient parameterization strategies (a) seem to be a key ingredient.

3. Excited states properties: The application of range-separated functionals and new methods (ppRPA) may open new research areas for DFTB. Recent tests have shown, that DFTB now reaches a reasonable accuracy required for large-scale applications of these methods.

4. Coupling to electron transfer/transport methodologies: DFTB is well suited for problems, where a quantum mechanical description is required for large systems, as typically occurring in electron- and exciton transfer problems. In the area, Greens-function and propagation methods have been successfully applied.

5. Linear scaling and accelerated MD schemes: For many systems, current force field methods seem to hit a wall in terms of accuracy. Hydrogen bonded networks as well as metal binding and conformational complexity seem to pose problems to the simple fixed charge models. Including polarizabilty still does not include charge transfer, however, leading to a substantial increase in computer time. DFTB, on the other hand, using new parallel computer architectures (e.g. GPU's) may shrink the gap to the more advanced empirical force field methods in terms of computational cost, providing a higher accuracy for certain chemical properties. Therefore, schemes to decrease the computational cost by avoiding or reducing the number of steps in diagonalisation (O(N), Extended Lagrangian) seem to promising for a routine application of DFTB to huge system sizes.

**Software infrastructure:** DFTB is implemented as open source code in DFTB+, where major contributions of the Karlsruhe/Bremen/Regensburg/Glasgow-groups are incorporated on a regular level. In addition DFTB implementations as module are incorporated in DeMon and in the commercial software products Materials Studio, Atomistix Toolkit and ADF. The workshop recommended developing on the one hand repositories of highly accurate and open quantum based data for validation of the more approximate DFTB. Vice versa parameter sets for the approximate methods could also be stored in such a repository. On the other hand a platform should be generated as an open source library (libDFTB), where other groups so far not contributing to the DFTB developments may add new functions for applications. The major aim will to incorporate already in early stages all new developments into a free DFTB software library, which can serve as a proof of concept for the given development, and at the same time enable both, commercial and non-commercial DFTB codes to directly use that specific feature without having to re-implement it from scratch again. To support this effort, the DFTB+ developers announced the release of an LGPL-licensed open source library (libDFTB) in about 6 month time, which would basically contain the full functionality of the current DFTB+ code (http://www.dftb-plus.info/) packaged into a modularized library which can fulfill the goal formulated above.

**Networking and training advances:** The organization was very compact with the scientists accommodated in the same hotel fostering exchange and discussion between the participants also outside the meeting room. The core developers of various DFTB implementations and some active junior researchers (in total 25) met another full day on Sunday October 11th to discuss and define future strategies for method and code developments, with special emphasis on sharing advanced implementations and challenging the DFTB data generation throughout the periodic table. Most of the time was given to present, discuss and define future parameterization strategies. Here in particular

the Heine/Vitek/Elstner and Bremen groups were involved to combine and synergize their efforts. The importance for exploration of possible collaborations combining manpower at different locations has been given strong support. Besides the general parameterization issue there is also ongoing effort in generation of specialized purpose data for example to address predictive large-scale device simulations. Special focus was also given in the remaining time to new developments in the DFTB ground state Hamiltonian (e.g. on-site correction, hybrid functionals, multipoles, ppRPA for correlation energy for Van der Walls interactions) and excited state simulations (hybrid functional and ppRPA for correct descriptions of single, double and charge transfer excitations).

The CECAM-Workshop has been run in combination with a two days hands-on-tutorial (fully compatible with the workshop program) which provided training and expertise in applications of advanced functions of DFTB+ in ground state molecular dynamics, excited states and non-equilibrium Green´s functions charge transport calculations. The tutorial on the deMon-nano code presented the possibility of doing configuration interaction calculations on the basis of DFTB, and also the basics of DFT has been demonstrated by a tutorial about the Siesta code. In addition commercial demonstrations of ADF (SCM-Amsterdam), Materials Studio (Accelrys), and Atomistix Toolkit (Quantumwise) introduced the participants into the respective implementations and functionality of their DFTB-module.

The participants believe that the scientific exchange of researchers from different disciplines is key to make progress in the most challenging problems of theoretical materials and bio-systems modelling. This exchange can be achieved by interdisciplinary conferences and workshops. Moreover, it would be desirable to enable exchange not only at the level of researchers, but also at the level of PhD and Masters students. The Workshop Speakers and Participants agreed to have a similar meeting in 2 years from now (Toulouse has been suggested) to evaluate the rapid progress in this field.

**Funding:** The field of theoretical materials modelling relies strongly on the availability of state-of-the-art computer code and computing resources. Therefore, the continued funding for the development of modelling software is an important infrastructure requirement: Only when codes are freely available can new ideas be tested and developed by a broad community of researchers. Since new codes that address the most challenging materials and bio-systems applications will push the limits of current supercomputing resources, it is also important that funding of these resources remains significant in the future. The workshop identified research areas that might be funded by the EU 2020 program particularly through the EINFRA and COST programs, and workshop participants would support an application involving multi-scale applications of Quantum based Molecular Dynamics and approximate DFT/B methods to either of these programs. Modelling of materials stability, interface design and bio-molecular interactions could form areas of particular focus.

**Will the results of these developments be of societal benefit?** Progress in the field of the approximate DFTB method is of great interest to Industry (Martin Persson, *Dassault Systtult, Cambridge being the link to he industrial and e.g. Fraunhofer users* being present at the workshop for example) and academics. This is reflected as well by the implementation and distribution of DFTB as module in ADF and ATK (Stan van Gijsbergen (SCM-Amsterdam) and Kurt Stokbro (Quantumwise, Kopenhagen) being present at the workshop too. Other developing areas of industrial interest include charge and heat transport in nanomaterials, electronic device simulations, energy and environmental technologies. Further developments in these areas require advancement to the quantum MD-DFTB level of modelling. Transferability demands and increased chemical complexity also require robust parameter free approaches (or methods that transfer smoothly to

more approximate fast methods), but the challenges of long time scales and large relevant phase spaces also require advanced integration and sampling techniques. Molecular dynamics applications are currently substantial users of present high performance computing (featuring on the roadmaps of several national HPC programs), but migration to near term platforms that require efficient use of even more processors presents immanent new challenges. Robust, open, and most importantly validated, software and methods are essential to reliably drive future developments and applications in this field.

# Free Energy Landscapes for Protein Folding: Consensus or Dissensus?

Organisers: Emanuele Paci, Sergei Krivov, Amedeo Caflisch
Location: CECAM ETH, Zurich (Switzerland)
Date: 12 – 14 October, 2015

**State of the Art:** Protein folding is the quintessential problem where biology, chemistry and physics meet, and where computation has been for four decades hailed as a decisive tool to "solve the problem". There is consensus on that protein folding can be understood as a diffusive process on a free energy landscape. Such free energy landscape is exceedingly complex, unless a small-dimensional order parameter (or reaction coordinate) is assumed. However, there is little consensus that the free energy landscapes so far produced, are correct or even meaningful.

Direct experimental determination of free energy landscapes is hindered by the limited temporal and spatial resolution of even the most advanced experimental techniques. Current experimental efforts focus on indirect determination of some "properties" of free energy landscapes. A recent highlight is the direct determination of transition path times by single photon counting.

On the other side, recent advances in computer hardware and software finally made possible atomistic simulations of reversible protein folding at equilibrium. Such simulations have potentially infinite temporal and spatial resolution. Such simulation allow in principle to directly determine the protein folding free energy landscape. The key step of such an analysis is the identification of an order parameter or reaction coordinate. It turns out that two different state of the art methods of selection of such coordinates, when applied to the same folding trajectory resulted in completely different free energy landscapes.

The aim of this workshop was to gather some of the key players in protein folding to try to learn from recent successes and narrow down possible reasons for disagreements. The workshop was a success, dominated by open and constructive discussion. A number of novel collaborations between participants were also foreseen.

**Outcomes**: During the workshop the following topics emerged as key:

a) Optimal reaction coordinates: The notion of "optimal" reaction coordinates for the description of molecular simulation trajectories emerged as one of the main discussion topics. This is definitely a general question that becomes increasingly important as fully converging simulations of folding and other functionally-relevant conformational transitions become possible for a number of systems (and was also a theme of the CECAM workshop "From trajectories to reaction coordinates: Making sense of molecular simulation data" that took place in Vienna, Sept 16-18, 2015). In protein folding, while there is agreement on what is intended with "optimal coordinate", different algorithms provide different coordinates often not optimal ones.

   In particular it was concluded that thorough investigation of the difference between the state-of-the-art algorithms for the determination of optimal reaction coordinates is a likely way to elucidate the difference for free energy landscapes obtained from simulation. In particular, it was suggested to assess the fraction of native contacts coordinate using alternative optimality criteria. There was consensus among the participants to the workshop that "optimality" of the reaction coordinate should be demonstrated, at least qualitatively, whenever free energy landscapes are presented. It was emphasized, that an optimal coordinate may not be physically intuitive and in

any way related to experimental observables, which is a main obstacle for the comparison of experiments and simulations.

b) Virtual intermediate state: The properties of the folding barriers are another related fundamental question. The two state of the art methods produce completely different results: one shows that the folding barrier is low, broad and very rugged, while the other says that there is an intermediate state with high, smooth barriers. State of the art experiments seem to support the first method. However careful consideration shows that these experiments employs a virtual intermediate state for their interpretation. It was pointed out that it is impossible to differentiate between the virtual and real intermediate state using experimental data. As a one way to avoid this issue it was suggested to focus experimental work on proteins, where, according to simulations, there is no intermediate state.

c) Pre-exponential factor for protein folding: In principle, free energy landscapes can differ in many aspects: the number of intermediate and transition states, the number of pathways, the height of the barriers, etc. It was suggested to focus first on resolving the difference in estimates of pre-exponential factor for protein folding, since this is a very well defined quantity, which can be related to many different characteristics of folding dynamics such as the height of the free energy barriers, the folding rate, the transition path times, and a number of different ways have been suggested for its determination. Current estimates of the prefactor from experiment and simulation differ by two orders of magnitude and are in the range $k_0^{-1}$~10 ns- 1us. To simplify the problem it was suggested to focus first on proteins without an intermediate state, for example on NLe/NLe mutant of HP35.

d) **Internal friction:** Another topic covered in a few presentations was the lack of understanding of the origin of internal friction in proteins.

**Challenges:** The computational aspects of the research on protein folding have benefited enormously in the past decade by the investment on optimised codes and novel dedicated hardware. The development has been fostered by the availably of open access software (NAMD, Gromacs, OpenMM, etc.) and by a collective effort into improving force fields. It is now important that trajectory analysis tools are also made broadly and freely available as well as long simulation trajectories. This will require improved storage resources (much larger and easily accessible).

**Requirements:** The workshop was the occasion for participants to foresee collaborations and consider applications for funding specific interdisciplinary projects. Useful research in this field requires state-of-the art experimental techniques such as FRET (W. Eaton and B. Schuler), triplet-triplet energy transfer (T. Kiefhaber), NMR (K. Lindorff-Larsen), millisecond detailed simulations (such as those performed by DE Shaw's group and S. Piana), and advanced methods to analyse such trajectories (as developed in the groups of A. Caflisch, S. Krivov and R. Best), that in turn can be also used to analyse experimental time-series as those measured by FRET. Setting out clear targets and benchmarks, as it was done during this workshop, is necessary to attract further funding and to raise the impact of the currently funded research.

**Will these developments bring societal benefits?** Protein folding is an exciting intellectual challenge, which summarises well the importance of synergies between experiment, theory and numerical simulation. More importantly, advances in understanding of protein folding and functional motion in proteins that lead to a consensus between experiments and simulation will provide scientists with tools suitable to design novel proteins for industrial and medical applications, and more broadly to reproduce synthetically complex biological machineries.

# Macromolecular simulation software workshop

Organisers: I. Bethune, P. Carloni, C. Clementi, P. Fowler, S. Jha, C. Laughton, J. Michell
Location: CECAM-DE-JUELICH (Germany)
Date: 12 – 15 October, 2015

**State of the Art:** As the use of biomolecular simulation grows, a corresponding boom in (decentralized) software development is taking place. Much of this is within multidisciplinary research groups and is highly focused on the specific needs of that community. This Development is often done by researchers with little or no formal training in software engineering or programming. This is not entirely a negative; it encourages practical solutions to real-life problems and 'thinking outside the box'. However, it is obvious that it is not the ideal route to the production and maintenance of high quality, flexible, sustainable, and usable software products that can be adopted and modified by others. Giving researchers the right tools and training to develop such software will reduce the amount of 'wheel-reinvention', and consequently improve research productivity.

This CECAM Software Development Workshop gave an opportunity for representatives of these projects and many others that are as yet less well-known, to interact with end users with the aim of tackling the most challenging problems in this domain, particularly those relating to how the capabilities of and opportunities presented by future generations of massive, sometimes distributed, heterogeneous, computational resources can best be leveraged in this domain of science.

This highly interactive event, targeted principally at early stage researchers, was formatted as six sequential workshops, each 2 days long. Participants were not required to sign up for the full duration of the course, but typically attended a sequence of one or more sub-workshops over the period that particularly interested them or addressed their needs. Each sub-workshop had 35-40 participants, and overall 93 researchers took part at some stage, including 15 from the USA, through additional support provided by the NSF.

**Outcomes:** Below, each of the sub-workshop leads comments on the activity they organised.

**"Software carpentry workshop'' (instructors: Philip Fowler)**

We ran a workshop covering shell scripting, python and version control using git to enable participants who were less experienced to learn these tools, which were pre-requisites for the remainder of the CECAM workshop. These workshops are interactive with one instructor live coding on the screen, whilst the other roams the room solving problems. An anonymous questionnaire was collected at the end of the workshop; all learners agreed with the statement "I enjoyed the workshop". For more details, including more feedback please visit http://philipwfowler.me/2015/10/23/software-carpentry-workshop-julich-12-13-october-2015/

**Analysing simulation data (Philip Fowler)**
The current practice in analysing simulation data involves either "black-box" compiled C programs (such as the g-tools that are part of GROMACS) or limited scripting capabilities built-in to molecular viewers (such as Tcl in VMD). A shift is coming to the use of high-level languages, such as python. This has three key advantages: (i) development is open

and the logic of an underlying algorithm can be easily understood, (ii) an analysis tool can read trajectories from any simulation code and (iii) the researcher has access to the ecology of powerful python modules, enabling analysis that otherwise would have been very difficult. The developers of two such python modules, MDAnalysis and pmx, were invited to this workshop. Each introduced either module with a short talk, followed by an extensive hands-on tutorial. At the end of the first day, the participants proposed research questions that could be solved using these tools and six teams formed. The following day, the six teams worked on their projects, under guidance from the assembled experts and academic software developers before presenting back to the group (i.e. a HackDay). From feedback, all the participants learnt something useful that will help their research and would also recommend a workshop like this one to other researchers, whilst 88% felt confident enough to now contribute to an academic open source software project. The biggest complaint by far from the participants was the poor Wifi. This made installing software and running a collaborative HackDay very difficult. For example, two teams decided to work in the onsite cafeteria on the second day was the Wifi was better. For more information see http://philipwfowler.me/2015/10/28/analysing-simulation-data-cecam-workshop-julich/

**Setting up simulations (Charlie Laughton, Nottingham)**
On the presumption of "rubbish in, rubbish out", it is vital to pay as much attention to the preparation of systems for molecular dynamics simulation as it is to work on optimizing the performance of the MD codes themselves and methods to analyze the data they generate. Automated, standardized, set-up tools offer opportunities for enhancing quality (e.g. by automatically identifying errors in input structures), reproducibility (e.g. by creating shareable, well-documented, workflows), and usability (e.g. in high-throughput simulation campaigns).

The aim of this sub-workshop was to introduce participants to alternative paradigms and toolkits for setting up molecular dynamics simulations, including validating starting structures and repairing "errors", embedding solutes in solvents and membranes, relaxation and equilibration protocols, and setting up complex simulations such as free energy perturbation calculations. Participants had the opportunity to explore the capabilities of a range of web-based (e.g. http://mmb.irbbarcelona.org/MDWeb/) and python toolkit-based solutions (e.g. http://www.hecbiosim.ac.uk/fesetup;http://sbcb.bioch.ox.ac.uk/memprotmd/beta/), and in a "hackathon" type activity, worked in small groups with the code owners to tackle problems related to unmet needs, flexibility, interoperability, and performance.

The enthusiasm and hard work of the participants made this a very successful event. Many participants were introduced to possibilities of ways of working that they had never had a chance to consider before. During the "Hackathon" day attendees from widely differing backgrounds had an opportunity to learn how to work together effectively in small teams to develop software to address genuine and pressing needs. More information is available on the https://bitbucket.org/claughton/cecam-setup/wiki/Home.

**Developing interoperable and portable molecular simulation software libraries (Julien Michel)** We ran a workshop on best practices to efficiently write molecular simulation software that will be reusable by others. Over the past few years there has been a trend towards writing molecular simulation libraries rather than new monolithic piece of software. This has the advantage that i) code can be more easily shared and re-used in other projects ii) new developers have access to communities of fellow developers. However fully embracing this model requires changes in current software engineering practices in the field. During the first session the invited speakers discussed

their perspective on software engineering for molecular simulation projects. This was followed by a group discussion. The main findings were that most code developers have a background i chemistry or physics and need to be explicitly taught the benefits of software engineering early in their studies. One major benefit is that developers should write well documented and portable code for their 'future-selves', and that the long term productivity gains compensate for the perceived additional burden in documenting code and writing unit tests. This was followed by a hands-on tutorial on how to develop with the Plumed molecular simulation library. The tutorial was unfortunately severely hampered by very poor wireless networking. On the second day the delegates followed an hands-on tutorial on the molecular simulation library OpenMM. This was followed by a 'free-form' session where delegates broke-out in three subgroups that were free to discuss topics of their interest with the invited speakers. Topics for the free-form session had been collected prior to the meeting and reviewed by the invited speakers/developers for feasibility/appropriateness. From the feedback received all delegates were satisfied/very satisfied with the workshop. The biggest complaints were related to the difficulties of using the Wifi on-site, and the difficulties of traveling to the venue.

**High Performance Distributed Computing tools for Biomolecular Simulations (Shantenu Jha,)** The objective of Workshop 5 was to introduce participants to tools that would enable them to utilize high-performance computing towards their science. There were two main parts of Workshop 5; the first part introduced students to the suite of RADICAL-Cybertools [1], the second to Copernicus [2].

Workshop 5 assumed there would be a wide diversity in the background and interests of students/participants. This would range from the "expert end-user" to the "disinterested black-box" user. Indeed, our pre-workshop survey confirmed the broad range of skills and expectations ("I just want to use this tool" to "Its interesting to know how this works under the hood"). Ultimately managing the expectation of a wide range of participants was the greatest challenge.

In anticipation, we designed Workshop 5 to have several tracks. Furthermore, In order to provide the students with hands on experience, we arranged for access to US NSF funded supercomputers: Stampede and Comet. This required a (separate) proposal for computing time (100,000 SUs), working with the XSEDE workflows team to provide an "access point" for all participants so they could initiate their work onto XSEDE.

The RADICAL-Cybertools part [3] in turn was comprised of three components: the first was an introduction to RADICAL-Pilot [4] the second a quick introduction to Replica-Exchange framework [5] and the third component an overview of Ensemble-MD toolkit [6]. The RADICAL-Pilot and the Ensemble-MD framework component in turn had three main tracks. The first was an introduction to the tool and a "walk through" a series of exercise to help understand these tools; the second was to use the tool on a meaningful problem at meaningful scale; the third track was to use the tool in an "open" exercise. The Replica-Exchange framework only had the first two parts.

The Copernicus section introduced students to fundamental dataflow network concepts and the idea of adaptive control. It then guided students through setup of Copernicus using the Stampede resource and a series of tutorials on running molecular dynamics and sampling workflows using the Copernicus package.

In the open exercises, students took the tools they had learnt and were encouraged to apply them to their own problems. Participant responses to the open exercises were mostly along the following lines: (i) "These tools are great and we'd like to use them on our home machine, can you help us install/provide them?", (ii) "These tools will be useful

if I can use it with this application, or modify to do X". Students started to think of tools/software as a way to develop "composable methods". Limited time meant that many of the "extensions" did not reach completion, but some started the process.

The participants learnt the importance and power of "advanced" that enable them to use resources more efficiently. An important consideration for all scientists is how to determine which tools to use and how much effort to invest in making "buggy(!), developing, unpolished" tools work. These were some considerations that were discussed and students were asked to respond to. [8,9]

An important lesson for the organizers of the Workshop 5 and RADICAL-Cybertools component in particular, was the significant amount of set-up to get the exercises going. This was mainly because we wanted to give students the full experience of running jobs on supercomputers. This in turn is necessary if we students are to learn the advantage of "scalable computing" as opposed to "just doing it on your laptop". If we had limited our tutorial to a laptop or a virtual machine, many of these problems would have been localized and controlled. The desire to give students an "in vivo" experience on production supercomputer/cyberinfrastructure is a non-trivial challenge.

In general, The setup of a controlled environment which would allow students to perform a set of sophisticated tasks is a non-trivial exercise. This required significant perturbation to the default software environment of "workflows.iu.xsede.org" machine (which served as the starting/access point).

http://radical-cybertools.org/
https://github.com/kassonlab/copernicus.git
https://github.com/radical-cybertools/tutorials/wiki/CECAM-2015/
https://radicalpilot.readthedocs.org/
http://repex.readthedocs.org/
http://radicalensemblemd.readthedocs.org/
https://drive.google.com/file/d/0B4XS8WEUG0G4UW9tOEZGaVlzUVE/
https://docs.google.com/spreadsheets/d/1JVc-iWjWm3Pwd2ESZq-Zu0YxRq-HxaY3VPLCK-Nv8wA/
https://docs.google.com/spreadsheets/d/1bw1XqRQ7wqWiHvhkhVFkwl1dSbRKVuMEeIj6pmfSwBI/

**Advanced Sampling and Long Timescale Molecular Dynamics (Cecilia Clementi)** The final workshop focused on software tools for advanced sampling methods. The first day covered tools developed by the ExTASY project (www.extasy-project.org) A short lecture presented the Diffusion Map-directed-MD method (based on Locally Scaled Diffusion Maps) and COCO (based on Principal Component Analysis), then the remainder of the day was devoted to guided tutorial exercises. The DM-d-MD and COCO methods are implemented as workflows built using the RADICAL EnsembleMD package, presented in workshop 5, and so the software environment was very familiar to delegates. In the first tutorial session we walked through how to run two pre-defined workflows on remote HPC resources, this was followed by a walk-through the process of implementing a new workflow with a different combination of tools, and in the final session, delegates worked on either sampling a small peptide using COCO or on modifying/extending the example workflows for their own use. We gathered qualitative feedback from delegates, which were generally positive, although they found that the software was quite 'bleeding edge' and that they would have liked more time to experiment with the individual tools as well as running the workflows. One commented that they were left a bit confused by the wide array of tools (and levels of abstraction) layered on top of RADICAL-Pilot over the previous 3 days.

The second day was taken up with a PyEMMA workshop (ca. 35 participants). Presented by Prof. Dr. Frank Noé (FU Berlin), Dr. Guillermo Pérez-Hernández (FU Berlin), and Dr. Antonia Mey (Univ. Edinburgh), this workshop was an introduction to the Markov modeling package PyEMMA, an open-source (LGPL) Python package for analyzing molecular dynamics simulations in such a way that models of the thermodynamics and kinetics of potentially complex molecular processes can be obtained, analyzed, interpreted and compared to experimental data.

After a scene-setting introduction, the workshop consisted of interactive hands-on tutorials where the participants were asked to complete certain tasks after having seen a simple example or part of the solution while using the PyEMMA online documentation. At the end of each part, the workshop organizers presented the solution.

The workshop had significant impact by introducing a selected group of interested participants to the PyEMMA software. These participants are likely to keep using the software and propagate their knowledge further by introducing it to other potential users in their working groups. Some feedback was immediately received by the fact that several participants became watchers or stargazers of the PyEMMA project on github.

**Summary:** This new format of event for CECAM has clearly been a great success. Feedback from participants has been extremely positive. Over ninety early-stage researchers, from across Europe and the US, have had an opportunity to not just learn about, but to contribute to, cutting-edge innovations in molecular simulation software and methodology. In addition the format promoted networking and group-working opportunities and skills that will help promote a secure and exciting future for the next generation of molecular software developers and users.

For the perspective of the workshop teams, there were also very useful pedagogical lessons learnt for future activities.It was very interesting to compare the different approaches to running different workshops (e.g. synchronous formats where all participants worked through an example problem step-by-step together, versus asynchronous formats where participants worked at their own pace, supported by circulating experts), and also alternative approaches to delivering software to the heterogeneous computing resources that participants brought with them (direct installation of packages, virtual machines, Dockers, etc.). In addition to the teams presenting and supporting the individual workshops, we would like to thank the administration team at Forschungzentrum Juelich for their vital assistance with the logistics of this complex event.

# Computational Modelling of Gene Expression and its Evolution

Organisers: T. Tuller, M. Linial, C. Wilke, M. Ehrenberg
Location: CECAM-ISR, Tel-Aviv University (Israel)
Date 13 – 17 October, 2015

**State of the art**: Gene expression is the fundamental process by which information encoded in the DNA (genotype) is used to produce proteins (phenotype). The process includes several stages and sub-stages (*e.g.* transcription, RNA processing, translation, transport, mRNA and protein degradation, etc.), and as it is partially encoded in the transcript it affects and relates to its evolution. The increasing *exponential* rate by which experimental data related to gene expression is generated, and the knowledge related to gene expression gained in recent years enables:

1) developing/tailoring multivariate machine learning and biophysical models of this process;
2) developing/tailoring computational models that connect gene expression to molecular evolution of transcripts.

Currently most of the researchers (and conferences) in the field focus mainly on one of the following topics/aspects/perspectives related to gene expression:

- Molecular evolution/population genetics perspective related to gene expression.
- Algorithms in molecular biology for analyzing gene expression data generally based on computer science methodologies such as graph theory and machine learning, without considering the biophysics of gene expression.
- Biophysical/ and analysis mathematical modelling of various gene expression steps without emphasis on computational, and/or gene expression, and/or evolution, aspects.

We believe that efficient gene expression computational/mathematical modeling is a crucial tool enabling encapsulation and thus understanding of the immense amounts of data and biological discoveries that exist today. Moreover, we believe that it is impossible to study and understand gene expression without considering the biophysics end evolution of this process. Thus we aimed at organizing the first workshop that will deal with computational modeling (or mathematical formalizations enabling computational modeling) of all aspects of gene expression biophysics and evolution. Specifically, we aimed at bringing together researchers studying gene expression using the different approaches mentioned above and promoting discussions about an integrated view of the field.

**Outcomes:** The specific topics discussed during the conference included:

- new approaches/techniques for computationally/mathematically modeling various biophysical aspects of gene expression (for example, stochastic simulations, mean field approximations, molecular dynamics);
- computational modeling and understanding the relation between gene expression and transcript evolution using simulations of molecular evolution and tools from fields such as population dynamics;
- computational modeling and understanding gene expression in diseases using (among others) tools from computer science and statistics;
- biotechnological applications related to computational modeling of gene expression (e.g. developing vaccines and heterologous gene expression);

- computational/statistical approaches for dealing with biases, parameter estimations, and model;
- Inference based on large scale gene expression measurements (e.g. ribosome profiling).

The lecture topics included various gene expression aspects/stages: transcription, translation initiation, translation elongation, mRNA degradation, co-translational protein folding, and more. The informal discussions included topics such as: conducting future frameworks for similar meeting/conferences on the topic of gene expression modeling in the future; debates related to various biophysical/evolutionary phenomena of gene expression, and new approaches for solving them; new topics for collaborations between the industry and academia related to gene expression analysis and modeling; and accurate analysis of some novel experimental outputs.

# Carbon at Extreme Conditions

Organisers: Razvan Caracas, Giulia Galli, Adrian Jones, Roberto Bini, Roland Winter
Location: CECAM USI-Lugano (Switzerland)
Date: 26 – 30 October, 2015

**State of the Art:** Carbon is a very peculiar element. It is the basis of life, as we know it, and of the entire field of organic chemistry. It is a remarkable reservoir for discovery of new exciting materials (fullerenes, nanotubes, graphene) with several Nobel prizes attributed to these discoveries, and it is a resource for future technologies. The carbon cycle is crucial for the global climate contributing massively to the concentration of CO2 in the atmosphere and methane, stored as hydrates on the sea beds. Carbon plays a major role in planetology: for example diamonds trap inclusions, that witness the processes taking place in the deep earth. It is also the base of the carbon matter from meteorites, constitutes the kerogenes at the surface of various icy worlds, and is a main constituent of white dwarfs. Today the study of carbon is an expanding field at the boundary between solids state physics, chemistry, and mineralogy. In our workshop we wanted to explore the high pressures and high temperatures regime and tried to push beyond the known state condition for the material

**Outcomes:** The following key topics received particular attention:

a) Characterization of low-dimensional forms of carbon:
- nanotubes: Raman spectroscopy is a powerful tool to identify the various types, including embedded molecules, stacking faults;
- carbon wires: such wires are produced using an ATM head with carbon-coated atoms; the wires are a few atoms long and their measured and computed transport properties are in perfect agreement;
- 2D diamond films: the various molecular intercalates between graphene oxide layers have distinct Raman signatures; passivation with hydrogen allows for atomistic simulations of the material;
- graphene and graphene oxide: layered materials can be filled under pressure with small molecules modifying the structural (measured by synchrotron X-ray) and chemical (FTIR, Raman) properties

b) New carbon-based materials:
- boron-carbon alloys: new techniques of ab initio structural predictions revealed a whole new class of hard and ultra-hard materials, based on B11C and C-B-C chains
- new composites have been obtained by in situ HP polymerization of C based molecules (ethylene, acetylene, CO) in synthetic zeolites

c) Mesoscopic carbon-based constructions:
- new, ultra high-pressure cell: perfect hemispheres made of nano-diamonds and cut by focused-ion beam spectrometers were glued on the two diamond surfaces inside the high- pressure diamond-anvil cell; static pressures reached more than 7 mega bar.
- biological membranes: Identification of protein response of bacteria under pressure compression and decompression, effects of temperature and pressure are different on the membrane behaviour
- new large-scale simulations using density-functional tight-binding show complex behaviour in the phase diagram of carbon, with metallic liquid phases

d) Earth and planetary science interest:
- new carbon-silicon mixed oxides: laser heated DAC is a powerful tool to reproduce the deep Earth conditions, a crystalline CO2–SiO2 solid solution has

been synthesized by reacting carbon dioxide and temperature exceeding 4,000K, simulations suggest new structures in the same system
- new carbonates: the carbon coordination increases with pressure from 3 to 4 in solid phases, mixtures with iron result in formation of various oxides and carbonates with previously unknown stoichiometries
- HP-HT behaviour of CO2: complex phases of carbon dioxide (silica analogues) can be crucial to understand the fate of this molecule, central to the carbon cycle, in the deep interior.
- water solutions and ices: water is the most common molecule on the Earth's surface being the most common environment for many carbon based molecules. The structure (neutron, x-ray and spectroscopic methods) and the dynamics (ultrafast spectroscopic methods) have been reported in combination with high pressure devices to study pure water, solutions and clathrate hydrates.

**Interdisciplinary aspects:** The workshop was at the interface between several communities: high-pressure chemistry, Earth sciences, biology, and materials science. Most participants gave talks. Their format was very flexible: we tried to have about 30 minutes of presentation out of which at least 10 minutes were supposed to be spent on a general introduction. About 15 minutes or more were left for discussions after each talk. The whole setup of the conference made possible the continuation of these discussions in smaller groups during the evenings. Out of one of these discussions was identified the need for a school on numerical methods (static and dynamic ab initio simulations, both lattice dynamics and molecular dynamics, all coupled with thermodynamic modeling) addressed to early-career experimentalists. Probably the most important outcome of the week in Lugano is the amount of information and exchange that passed around between the participants. Everybody could benefit plenty of their colleagues' experience; and it was a really appreciated opportunity for everybody to offer insight of their work to the other communities.

**Funding:** The workshop was partly sponsored by the Extreme Physics and Chemistry Directorate of the Deep Carbon Observatory (DCO, https://deepcarbon.net/) and by the European Geoscience Union (http://www.egu.eu/). As DCO is highly geoscience-oriented, one of the purposes of its presence at our workshop was to open up to other communities, which it achieved to some extent. Nevertheless this is a path to explore in seeking for funding in the future, for example for a school on atomistic numerical methods.

# Computational plasmonics: an ab initio and multiscale perspective

Organisers: A. Calzolari and S. Corni
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 2 – 4 November, 2015

**State of the art:** Plasmonics, the study of the interaction between electromagnetic radiations and free electron in condensed matter systems, is a rapidly growing field, with important applications. Simulation of plasmonics systems is currently dominated by classical electrodynamics (CED) macroscopic models, and noble metals, whose conduction electrons behave approximately as free, are the plasmonics materials *par excellence*. Yet, several aspects of plasmonics, such as transition from molecular to continuum-like behaviour, do require going beyond CED, towards atomistic descriptions. The aim of the present workshop is to address open questions such as: a) How to identify a plasmon excitation in electronic structure method results? How do excitations transform from molecular-like to plasmon-like with increasing size? b) What are the bottlenecks to be overcome to get a first principle description of plasmonics? c) Are the present multiscale approaches to molecular plasmonics complete? How can the chemical interaction be included in this framework? d) What are the computational predictions accessible to experimental verifications? e) Are non-conventional plasmonics materials such as graphene or metal oxides posing new or different challenges with respect to metal nanostructures? Is the classical electrodynamics treatment as good for these materials as for metals, or is an atomistic modelling is mandatory?

**Outcomes:**
**a) Plasmons from a microscopic point of view**. i) First-principles approaches properly describe plasmons in metal clusters. The scaling up to larger nanoparticles (NP) is not a theoretical issue, but rather it is a matter of faster/better performing algorithms. ii) In the definition of what is a plasmon a nanoscale systems it would be important to quantify the plasmonic character of an optical excitation. The problem of quantify/discriminate excitations at the molecular scale is still open, and no general consensus has been reached on the way to solve it. Coulomb potential has been proposed as the key quantity, and it is related to the near-field enhancement. Separation of short-range vs long-range Coulomb interactions emerged as a promising route to explore.

**b) Challenges in multiscale modeling of plasmonics.** One possible way to describe the quantum effects in NP plasmons is trying to link atomistic, jellium-based and quantum hydrodynamics models upon changing size. This can be reasonably done in the next few years. In many cases, however, in order to face with experimental results, such as SERS and SEIRA, the bare description of the plasmonic NP is not sufficient and the first-principles coupling of molecules with focused EM fields description is needed. The development of embedding schemes to include a metal region atomistically is also important for describing near-field effects (e.g atomic lighting rod, chemical reactivity, conductivity, hot-electron photochemistry, including phonons). Only a few attempt to include these effects have been proposed, mostly derived by the dielectric treatment of molecules/NP in solution. At present there are no available models to describe the optical forces induced by the plasmons (e.g. metal atom rearrangements under plasmon excitation). This is a computational challenge: no straightforward way to solve it has emerged. Other open and not-yet solved problems deal with the out-of-equilibrium behavior of plasmonics phenomena, which are responsible for e.g. photochemistry, heat production, plasmon-induced dynamics.

**c) Non-conventional plasmonic materials**. First-principles modeling are needed to predict the optical properties of new unconventional materials with tunable properties (such as doped oxides and nitrides), for which defects and doping are the important parameters to be included in the models. In particular, computational material discovering and the definition of proper figure of merit have emerged as a critical need of experimental partners to guide the choice of new setups. Standard ab initio techniques, based on (TD) DFT, are the tools of choice to solve these issues. However, simulations of non-linear optical properties, and the simultaneous inclusion of further physical effects (e.g., magnetic, optical and thermal) are still missing. This is a major issue, as these are often the motivations/advantages of using such materials.

**d) Quantized plasmons**. The fully quantum nature of plasmons, including simultaneously an atomistic and a quantum-electrodynamics (QED) description is almost theoretically unexplored. In particular, this appears to be due to the fact that the extension of TDDFT (the most used computational technique to simulate optical properties in nanostructures and molecules) to QED is still in its infancy. As a consequence, the treatment of quantized plasmons with an atomistic description of the NP is missing, and the strong coupling of atomistically-described molecules and quantum plasmons are not accessible from a theoretical point of view. No clear way to solve this issue from a microscopic point of view has emerged.

**Infrastructure requirements** Hardware and Software: on one side, plasmonics simulations where the entire system is treated atomistically from first principle are computationally not different from TDDFT calculations on very large molecules. Therefore, the improvements of hardware, software parallelism (e.g. the transition to exascale computing) and faster algorithms that are pursued in TDDFT are the same as required for this aspect of plasmonics. On the other hand, a multiscale model where only a region of the system is treated atomistically appears to have a great potential here. They need to be included in existing QM software, typically joining techniques from CED with first-principle ones. Network & Training: the computational plasmonics community still involves many different backgrounds, from theoretical physics interested in QED to biochemists interested in applications. Regular workshops and training events that can mix such backgrounds, create a common language and provide cross-fertilization are a necessity.

**Funding:** Computational plasmonics encompasses both fundamental and applicative aspects. Some computational applications, ancillary to systems of immediate technological interests (see below), may be funded through H2020 under NMP or Health priorities. On the other hand, fundamental studies and model/theory/software developments are difficult to be collocated in the H2020 programme. This is a very short-sighted strategy, as many of the present fundamental investigations in plasmonics (such as photochemistry in the strong coupling regime) may translate to genuinely new technologies in a few years, and always require a theoretical/computational support to be understood and rationally designed.

**Will these developments bring societal benefits?** Plasmonic effects in molecular and nanoscale systems have been long envisaged for application in a wide range of applications. Indeed, the confinement of electromagnetic energy below the diffraction limits can produce a local enhancement of the external electromagnetic field near the nano-system. When this happens, the system acts as a plasmonic nano-antenna able to amplify local electric fields at the nanoscale. This phenomenon is particularly appealing in light harvesting devices in the visible and near-IR wavelengths, energy conversion, and telecommunications in the terahertz region. Plasmonic nano-sensors have been proposed

for biomedical applications, where they allow spectroscopy in truly nanometer-scale volumes, increasing signals and resolution of spectroscopies up to single-molecule sensing. Furthermore, novel and tuneable plasmonic ceramic materials have recently been proposed as refractory while also offering biocompatibility, compatibility with CMOS devices, chemical stability, corrosion resistance, and mechanical strength and durability. This is particularly attracting for new energy conversion concept, namely solar thermo-photovoltaics, which promises efficiencies up to 85%, as well for the realization of higher-density data recording devices (e.g. heat-assisted magnetic recording) and hyperbolic meta-materials.

# DNA Damages: modeling and rationalize structure and reactivity

Organisers: Elise Dumont, Célia Fonseca Guerra, Filip Lankas and Antonio Monari
Location: CECAM-FR-RA, Centre Blaise Pascal, ENS Lyon (France)
Date: 3 – 6 November, 2015

**State of the art**: DNA is the fundamental molecule encoding and storing the genetic information of all living organisms. [1] All the fine effects governing peculiar DNA structures have been the subject of important research. The subtle balance of non-covalent interactions, namely hydrogen bonding and stacking, assures DNA stability but also a structural flexibility. Hence different DNA forms such as G-quadruplexes [2], that are usually associated to telomeres, exist and are thought to play significant biological roles in regulating cell life cycle. Furthermore, one has to consider that in cellular DNA is associated to proteic structures, histones, which gives rise to stable nucleosomes structures. [3]
Because of its biological role DNA is designed to enhance stability towards external stress. Nevertheless exogenous or endogenous agents may threaten its integrity leading to the formation of deleterious lesions [4-5]. When left unrepaired, DNA defects can trigger cellular apoptosis, or even induce mutagenicity of the cells. However the potential danger of a particular lesion also depends on the efficiency of cellular enzymes to repair it. There are new opportunities emerging from the advances in DNA simulations that could elucidate a relation between structure of damaged DNA, induction of defects and repair yield, which has not yet been settled on a firm basis. The existence of complex DNA cross-links can now be surmised from quantum calculations and proposed to the experimental community.

Most DNA lesions are induced by oxidative stress, implying radical mechanism eventually ionizing radiation can come into play, or by exposure to UV and UV-Vis light [9-11]. UVA and even visible light can be harmful because of photosensitization. Endogenous or exogenous chromophores interacting with DNA are activated by light absorption that triggers the subsequent DNA degradation [9].
Even the formation mechanisms of the most common and simple defects such as 8-oxoguanine and (oxidized) abasic sites are not fully elucidated. Also the range of DNA lesions or drugs interacting with DNA has been expanded over the last years, with increasingly complex and most harmful lesions patterns evidenced in cellular DNA: intra- and interstrand DNA cross-link [5-7], protein-DNA cross-link [8] DNA lesions have been characterized owing to the tremendous advance of analytical techniques [5-7] that allow to detect very rare lesions. Indeed, if many studies have mimicked DNA chemistry in isolated nucleobases nowadays more and more complex environments are treated including nucleosomes, while the use of EPR also allows to detect many labile radical intermediates [6,7]. Yet many of the mechanism of lesions production and repair are not completely elucidated.

Refereces


1 M. D. Franck-Kamenetski *Phys. Life Rev.* **2014**, *11* 153-170
2 S. Burge et al. *Nuc. Acid. Res.* **2006**, *34,* 5402-5415
3 C. Zhou et al. *J. Am. Chem. Soc.* **2012**, *134,* 8090-8093
3 W. P. Roos et al. *Cancer Lett.* **2013**, 332, 237-248
4 B.-B. S. Zhou et al. *Nature* **2000**, *408,* 433-439
5 P. Regulus, et al. *Proc. Natl. Ac. Science* **2007**, *104,* 14032-14037
6 C. Zhou et al. *J. Am. Chem. Soc.* **2013**, *135,* 593
7 C. Zhou et al. . *Am. Chem. Soc.* **2012**, *134,* 16734
8 K. M. Johnson et al. *J. Am. Chem. Soc.* **2013**, *135* 1015-1025

9 M. Consuelo-Cuquerella et al., *Acc. Chem. Res.* **2012**, *45,* 1558-1570
10 E. Dumont et al. *J. Phys. Chem. Lett.* **2013**, *4* 4119-4124
11 V. Vendrell-Criado et al. *Angewandte Int. Ed.* **2013**, *52,* 6476-6479

**Outcomes**: Elucidating the structure of damaged oligonucleotides as well as the mechanisms leading to lesions and promoting repair is a fundamental challenge in chemical biology. Yet the workshop as shown how the European community working of the field shares a high degree of maturity and is ready to tackle this challenge providing it will satisfy the necessity of networking and interdisciplinarity. The DNA damage workshop has brought together leading experimentalists and theoreticians to give a most comprehensive picture of these processes, at a molecular and electronic level. Furthermore, the necessity of a stronger cooperation between chemistry and biology has been underlined, in particular in order to unravel the biological significance of DNA lesions at a molecular level. In that respect it will be important to evidence how the structural modification of DNA, induced by damaged nucleobases or by sensitization, govern both the energy- and electron-transfer phenomena as well as the repair mechanism and rate. This scientific challenge needs to combine efforts of cutting-edge spectroscopies, cellular and molecular biology, and molecular modelling. In particular multiscale model going from explicit quantum description, also in the excited state, to mesoscale coarse grain level passing through atomistic classical description is fundamental to get a clear and in-depth comprehension of all these fundamental mechanisms.

Some key points that have been tackled concerns:

- Elucidate the chemical mechanism leading to radical DNA oxidation. Identification of products and key intermediates and of their main structural aspects.
- Unravel the structure and dynamics of DNA/sensitizers complexes, their dynamic evolution and the interaction modes. Identifying possible selective non covalent modes and interpret their stability.
- Modelling the time evolution of the ground and the excited states leading to possible (photo-)chemical adducts.
- Modelling structural dynamics and mechanical properties of damaged DNA to elucidate their effect on repair mechanisms, protein-DNA interactions, and genome organization
- Understand complex tandem or cluster DNA lesions, in particular concerning their high degree of mutagenicity.
- Modelling and interpreting the interaction of DNA with repair enzymes. In particular elucidating the reasons, both structural and electronic, of the lack of repair of some lesions.
- Modelling non standard DNA structures such as G-quadruplexes, single strands, or hairpins and elucidate the factors governing their stability
- Provide ideas for the rational design of novel anticancer drugs.
- Asses DNA long-time scale behaviour in particular concerning ion distribution and mobility and ionic strength effects

**Community needs:** First of all it has been agreed that the community needs to be better structured, this will be ensured by different actions that will be taken as described in the following. First of all a COST proposal will be submitted in order to enhance networking and cooperation within a more stable framework and favour exchanges of competences between the partners. Furthermore, it has been decided that the exchanges realized during the workshop are fundamental and as such those actions to be continued over the time. It has been decided to hold, with a similar workshop/conference on DNA damages and repair with a biannual periodicity. The next workshop will most probably be

organized in Amsterdam (The Netherlands). Finally it has been underlined the necessity to have a sort of web centralized repository in order to store and exchange significant results, but also in order to better communicate key results to the general public as well as to the policy makers. The website will be realized in the following months also profiting of the help from European scientific societies.

**Funding:** The most common funding source so far have been COST actions. This aspect will also be strengthened as an outcome of the present workshop since the need for networking has emerged. In addition it has been decided to enhance the capacity to respond to the European calls in the framework of H2020. The relationship with industries, start up, and private companies has also emerged as a crucial strategy to pursue. This aspect will be also favoured by the envisaged strong collaboration between theoreticians and experimentalists, coming from chemistry and biology.

**Will these developments bring societal benefits?** The study of DNA damages and lesions has an obvious societal and economic importance. Indeed it is related to the development and therapy of a number of debilitating illnesses including cancer and neurodegenerative diseases. Furthermore, the study of DNA repair pathways as well as the identification of the biological relevance of different lesions and which mechanisms are mostly involved will constitute a breakthrough shedding light on the molecular bases of a number of diseases. This will also be realized strengthening the collaboration between chemistry and biology as well as between theoreticians and experimentalists. Furthermore the possible rational design of novel therapeutic agents, notably against cancer will be advanced. As such the network will be extremely opened toward the collaborations with pharmaceutical companies.

# Understanding function of proteins in membrane by atomistic and multiscale simulations

Organisers: Vittorio Limongelli, Michael Lelimousin, Mark SP Sansom
Location: CECAM USI-Lugano (Switzerland)
Date: 10 – 12 November, 2015

**State of the art:** To understand the way a protein functions it is important to consider its cellular environment. About 25 % of genes encode membrane proteins, and furthermore membrane proteins are targets for ~50 % of marketed drugs. Signalling mechanisms of membrane receptors involve subtle conformational changes of these proteins. Therefore it is of paramount importance from a biological and pharmaceutical point of view to elucidate the dynamics of these macromolecules in their native lipid environment. The use of molecular dynamics (MD) simulations represents the first natural choice to investigate such functional dynamics, providing an atomistic description of the interactions at work. Furthermore, the growing number of experimental structures released over the last fifteen years has represented a further stimulus for computational investigations. Thus, taking also advantage of modern hardware we may enhance our understanding of membrane protein functions and the role of specific lipid/protein interactions. For instance, mechanisms of gating, permeation and selectivity could be deciphered using MD simulations. Nevertheless, many functional mechanisms, such as ligand/protein binding and large protein motion, remain inaccessible through standard calculations because of the limiting timescale of MD simulations. Furthermore, increasing the size and complexity of the systems is necessary to obtain models that more closely resemble the *in vivo* cell membrane environment. To overcome these limitations one needs to step up the computational strategy. In this view enhanced sampling methods like metadynamics, have proven to be successful to study long timescale biological phenomena, whilst coarse-grained modelling has enlarged the scales accessible to membrane protein studies. Complementary multiscale approaches have been determinant to optimize the choice of methods used to treat variety of biological and chemical problems. However the rearrangements involved in the function of complex membrane proteins (e.g. receptors, transporters, ion channels) still require the development of advanced methodologies. Allosteric mechanisms involved in the activity of signalling receptors also impose extension of the timescales reached by simulations. In addition there are increasing evidences that long-range and cooperative effects at membranes (e.g. curvature, nano-domains and clustering) play significant roles in the modulation of membrane proteins activity. Thus, today there is a tremendous need to accurately consider the complexity in size and composition of cell membranes in computational models and to reach with simulations the real-life timescale.

**Major outcomes:** This workshop drew together both the theoretical and experimental communities that study membrane proteins in order to address the challenges of developing a quantitative and predictive understanding of the relationship between membrane protein structure and function. The presentations covered areas of research ranging from the simulation of permeation events in channels and porins to the identification of functional motions in membrane proteins. Another paramount topic was the research on ligand/protein binding and drug design. New directions of research to include more complexity in the molecular models of membranes have been evoked. In particular the contribution of experimentalists was really valuable to discuss innovative approaches to bridge the gap between simulations and *in vivo* studies. In this sense, the progress achieved over recent years in the accuracy of all-atom force fields, the development of coarse-grained models, and the routine access to more efficient softwares

and hardwares (e.g. gpu-accelerated MD) were shown to significantly favor collaborations with experimentalists.

**Outcomes:** This workshop was a unique opportunity to update the state of the art of simulations and experiments on membrane protein systems. As general outcome of the meeting, it was evident from the results presented in the talks that we are now able to obtain more realistic models close to the physiological conditions (*e.g.* computational electrophysiology) as well as performing simulations with many biomolecules in interaction at cell membranes. In consequence, major advances are expected in the understanding of signalling mechanisms over the next years. In particular the characterization of allosteric mechanisms involved in the regulation of membrane receptors, including protein clustering and multi-protein complex formation, should become more and more frequent. This should have a significant impact on computer-assisted drug design. The improved accuracy in the prediction of ligand/protein binding affinities and kinetics should also go in this direction. This is expected that these results will increase the contribution of academic research to drug discovery. However one should also try integrating into this model important aspects of drug delivery, such as membrane permeation, pharmacokinetics and toxicity of a drug candidate. In this prospect, the establishment of European collaborative framework that would connect computational and experimental laboratories will be decisive.

**Funding:** The most common funding source has been COST actions so far. Additionally, the recent outstanding results achieved on membrane proteins systems might open a new era of interest of the bio/pharma companies towards simulations. Based also on the feedback of the conclusive round table discussion, new funding possibilities from the industrial world appear plausible.

**Will these developments bring societal benefits?** Progress in the field of membrane proteins is of great interest to EU industry, since they are targets for ~50 % of marketed drugs. The significantly improved accuracy in drug binding prediction and in the description of the membrane/protein interactions will raise even more the interest of the pharmaceutical companies (Novartis, Roche, Sanofi, etc). Also small biotechnological companies can benefit from this fruitful background. The on-going increase in computational power and methods development will make possible to extend the techniques discussed in this workshop towards technologically and experimentally relevant scales, thus contributing to drive innovation. For instance, the investigation of the interaction of nanomaterials with cell membranes can represent an important approach to develop new pharmaceutical vectors and prevent potential toxicity effects. These studies might have a great impact in medicine with potential benefits in public health and societal life.

# Big Data of Materials Science – Critical Next Steps

Organisers: Luca M. Ghiringhelli, Sergey V. Levchenko, and Matthias Scheffler
Location: CECAM HQ, EPFL Lausanne (Switzerland)
Date: 30 November –4 December, 2015

**State of the art**: By using first-principles electronic-structure codes, a huge number of materials have been studied in recent years. The amount of already created data is immense. Thus, the field is facing the challenges of "Big Data", which are often characterized in terms of Volume (amount of information), Variety (heterogeneity of the form and meaning of the data), and Veracity (uncertainty of the data quality).
Obviously, the computed data may be used as is: read out what was stored. However, for achieving deeper and novel scientific insight, the above "V's" should be complemented by the Big-Data Analysis. Calculating properties and functions for many materials, (e.g. efficiency of potential photovoltaic, thermoelectric, or catalytic materials) is the necessary first step. Finding the actuating mechanisms (the "causes") of a certain function is the desired science. In fact, such scientific understanding is needed for deciding what new materials should be studied next as most promising novel candidates and for identifying interesting anomalies.
The field of Big Data Analysis in Materials Science has evolved in both directions: a) constructing
large databases of internally consistent ab initio data on which simple or less simple analyses are performed and b) development of advanced analysis techniques moving from existing machine learning methods. Examples of a) are the Materials Project [1], related to the Materials Genome Initiative [2], CatApp [3] (catalytic materials), aflowlib [4] (bulk semiconductor and metal compounds), AiiDA ("Automated interactive infrastructure and Database for Ab initio design") [5], and NOMAD (Novel Materials Discovery) center of excellence [6]. Selected examples of b) are the use of kernel ridge regression [7,8,9], neural networks [10], dimensionality reduction [11], genetic programming [12], and sparsity-related techniques [13,14] for the learning and prediction of electronic, structural, elastic, and transport properties of bulk crystals [7,9,12,13,14], cohesive energies of molecules [8,10], and the design of new functional materials [11].

During the workshop, one session was devoted to topic a), while all the other 7 sessions were devoted to topic b) and in particular the crucial role of the descriptor in the big data analytics.

**Outcomes:** several general questions and issues were addressed during the workshop, thanks also to the extended (45 minutes) discussions held at the end of each session and moderated by the chairperson of the same session.

- **Are big data reliable?** In order to perform analytics on data coming from possibly heterogeneous sources, it is fundamental to be able to go back to the detail of the original source, in order to allow for the detection of possible mistakes. In facts, if some data point is identified as an anomaly, is that really a physical system behaving as an outlier or is it a mistake in the calculation? In general it was felt the worry that creation of large amounts of data, without human micro-management, can lead to uncontrolled mistakes, because electronic structure codes are in general not working as black boxes and sometime can give seemingly reasonable – to a superficial inquiry – but wrong results. A roadmap towards a solution to this problem is to build rich metadata structures (talks by Curtarolo, Mohamed, Pizzi, Ineichen) for the storage of the results of calculations,

so that complex queries can be set up, also *a-posteriori*, to detect wrong calculations. Steps in this direction are undertaken by the NOMAD center of excellence, and the MaX center of excellence (in particular the Aiida platform).
A related issue is whether the accuracy level of *ab initio* calculations is good enough in order to be able to predict properties of realistic materials (talk by Thygesen) or if algorithms are fast enough in order to perform accurate calculations on large amount of materials (talk by Levchenko). The answer is that only recently, in the last few years, the community has reached the level of performing somewhat routinely high-level calculations with reasonable CPU usage (in order to perform calculations on several hundreds to thousands of materials).

- **Descriptors: finding or designing them?** It was agreed that for a successful machine-learning type of analysis, the most important step is to identify a good descriptor (a set of descriptive parameters, identifying the data-point – the material – for the analysis), that reflects physical constraint and symmetries (e.g., invariances). Two schools of thoughts are prominent. In one, the descriptor is designed, refined step by step, by human effort (Behler, Csanyi, Rupp, von Lilienfeld, Ramprasad), in order to incorporate the physical constraint. Of particular relevance is the work of Csanyi, where the descriptor and the kernel (loosely speaking, the mathematical model used for the statistical learning) are constructed hand in hand, opposite to the mainstream tendency to select one or few kernels and then build descriptor that go together well with the selected kernel(s). The other school of thought (Ghiringhelli, Vybiral, Mueller, Hammerschmidt) is trying to learn the descriptor from the data, by constructing several candidates that are selected by means of sparsifying methodologies (e.g, compressed sensing). It was suggested that the two schools of thoughts can merge, by singling out the best of both fields.

- **What do we expect from machine learning?** A point addressed in several talks (e.g., by Rupp, on Lilienfeld, Vybiral, Ghiringhelli, Csanyi, Behler) is whether we can expect that machine learning can identify predictive models, also able to extrapolate prediction outside the region spanned by the training set. First of all, realising whether a test point is or not outside the region spanned by the training set is not a trivial task. Several machine-learning developers have addressed it and there is no general answer, due to complex topological features of high-dimensional spaces. It seems that linear models are the most promising class of models that can allow for extrapolation. The requirements of such predictive power seems to be in the selection of *causal* descriptors, i.e., such that the relationship between the descriptor and the predicted property turns out to have physical grounds. Tools for assessing whether a descriptor has or not a causal relationship with the predicted property are not yet fully developed, though.

- **Global vs local models.** An interesting new approach in statistical learning and in particular pattern discovery sub-field, was introduced by Mario Boley, who described methods that can partition the data points into subset, so that optimal local models (one for each subset) can be learned. In general these local models perform better than global ones (acting over the whole data set), at the expense of bigger computational complexities. This approach will prove useful in materials science, if the partitioning criterion (done with purely statistical criteria) can be related to a physical understanding of the reasons behind the partitioning (e.g, one set turns out to be metal and the other non-metals without having built in this distinction, etc).

- **Can we think at big data analytics as an artificial intelligence version of the Brahe ("high throughput" data storage and management) + Kepler (learning of laws from data) in history of science? So that, to the human researcher is left the role of Newton, i.e, understanding the physical law behind the statistically inferred relationships among data.** This is a somewhat provocative question that has of course no definitive answer, but reflects a way of looking at the whole big-data analytics, at least in physics and in particular materials science. Namely, differently from other fields, the physicist does not like to stop at a statistically inferred relationship, but wants to find the *reason* why such relationship holds. This last step seems to be genuinely human, but certainly machine (learning) can speed up the process but proposing more and more non-trivial relationships.

**Community needs:**
The community needs efforts in two directions: a) establishing reliable and long-term open-access infrastructures for the storage and management of large to huge amounts of data (petabytes, towards exabytes), these serve as the basis for the analysis, with the valuable outcome of disentangling data production from the analysis, i.e., groups that cannot afford access to large computer facilities, can perform analysis and do science. b) Development of *domain specific* data analysis tools, for example to address the specific need of testing the casual nature of discovered descriptor -> predicted property relationships

**Funding:** three centers of excellence, in the framework of the horizon-2020 call for e-infrastructure, have been funded in the materials science field, with several millions euros and therefore several academic positions: Materials design at the eXascale (MaX), E-CAM, and Novel Materials Discovery (NOMAD). With different focuses, all three centers share the goal of further establish high-performance computation for novel materials design and discovery.

**Will these developments bring societal benefits?** The possibility of finding and designing new materials for technological applications (e.g., catalysts, superconductors, photovoltaic materials, thermoelectric materials, materials for fuel cells, …) on the basis of computational screening and data analytics can have an enormous societal benefits. Finding new materials faster and with a fraction of the cost, as compared to the more traditional experimental screening, will advance at faster rate several technological areas [2].

References
[1] https://www.materialsproject.org/references
[2] http://www.whitehouse.gov/mgi
[3] J. S. Hummelshøj et al., Angew. Chem. 51, 272 (2012).
[4] http://aflowlib.org/
[5] http://www.aiida.net
[6] http://nomad-coe.eu
[7] A.P. Bartók et al. PRL 104 , 136403 (2010)
[8] M. Rupp et al. PRL 108, 058301 (2012).
[9] G. Pilania et al., Scientific Reports 3, 2810 (2013)
[10] J. Behler and M. Parrinello, PRL 98 146401 (2007)
[11] P. V. Balachandran et al. Proc. R. Soc. A 467, 2271 (2011)
[12] T. Mueller, E. Johlin, and J. C. Grossman, PRB 89, 115202 (2014)
[13] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozolins, PRB 87, 035125 (2013).
[14] L.M. Ghiringhelli et al., Big Data of Materials Science – Critical Role of the Descriptor, PRL 114, 105503 (2015)

# Open Quantum Systems: Computational Methods

Organisers: C. Rodriguez-Rosario, T. Frauenheim, G. Chen, O. Hod, E.K.U. Gross
Location: University of Hong Kong (China)
Date: 30 November – 4 December 2015

**State of the art:** Open Quantum Systems (OQS) is the study of quantum mechanical systems that are not isolated from their environments being most abundant. They allow for the description of quantum mechanical systems under the influence of dissipation, decoherence, quantum measurements and more. As well, they are a robust way to account for processes were particle number is not conserved, or that are coupled to thermodynamic baths. This picture allows us to average, coarse grain or approximate the environmental degrees of freedom while still retaining important features of their dynamical impact on the system, without having to fully simulate all the degrees of freedom of the environment.

Open Quantum Systems are often represented as the evolution of density matrices under the influence of superoperators knowns as dynamical maps, or in their differential forms, the master equation. In its Markovian instance, this is known as the Gorini-Sudarshan-Kossakowski-Lindblad master equation. The relationship between these and their mathematical structure has been discussed intensely throughout the workshop. Work has been presented done on the different structures of the equations that can characterize non-Markovian dynamics. The assumption of initially uncorrelated system-environment states are being challenged in many of the invited talks, and it is particularly relevant in connection to Non-Markovian dynamics. It has been shown that this leads to a significant increase in the number of parameters complicating the existing computational methods. Now these superoperators are well understood and the focus is changing towards understanding how to apply them as efficient computational methods.

From the conference, it is clear that all directions of computational quantum chemistry, quantum physics and quantum molecular biology are moving towards fully embracing the open quantum systems paradigm. However, the computational achievements are scattered, with not enough sharing of methods. It was positive to discuss all these, their strengths and weaknesses in order to provide the community the understanding to develop further robust quantum computational tools. The official Workshop program of 9 half days consisted of 39 invited talks of 40 minutes (30+10) each and a poster session presenting 27 contributed posters.

**Outcomes:** During the workshop, the following emerged as key topics

a) Similarities in methods for Molecular Junctions and Spectroscopy: an interesting development was the similarities of the open system methods used for both simulating Molecular Junctions and Spectroscopy experiments. This comes from the similarities of the physics of electron-phonon and electron-photon coupling. An important differences is that one focuses more on steady state, while the other on transient effects. Many questions were raised that suggest that methods that operate efficiently in both regimes can be devised.

b) QM/MM hybrid approaches: multi-scale approaches where the effects and responses of the extended surroundings are modeled with more approximate but computationally faster classical force fields, while retaining quantum mechanical accuracy for the more

complex core region(s) were presented to be useful and mature ways to simulate many systems.

c) Role of electronic structure: the role of efficient methods for electronic structure was shown to be very important for realistic open system dynamics in chemical, biological and solid state systems. Although the relative strengths and weaknesses of methods such as DFT and DFTB, or of wave-packet modeling and TD-DFT remain unclear.

d) Markovian and Non-Markovian approaches: One of the most important properties of open system that was discussed repeatedly in most of the talks was the property of memory in the baths. The pros and cons of different methods, and how much overhead they add to ensure physicality was discussed. An important presentation proposed a new way to characterize methods without assumptions of Markovianity. Similarly, the question of how Markovian is a process, or how effects such as system-bath correlations can make a process non-Markovian in some regimes, is one that will require further studies to develop more concrete ways to approximate the memory effects.

e) Hierarchical Equations of Motions: this approach has really been embraced by the community as a way to expand master equations in a powerful but tractable way. It can be fully non-Markovian. This was clearly one of the new dominant methods in the community, and possible future conferences could be dedicated to this topic alone. Among the important lines of research identified was an understanding HEOMs so that it can be thoroughly compared and contrasted to older master equations.

f) Control as an OQS: modeling active control systems as open quantum system is one that is leading to novel proposals for devices in many fields, from spectroscopy to heat transport and chemical reactions. There were a lnbumber of novel ideas identified during the discussion that could be further explored in future conferences.

g) Thermodynamics: a surprising connection was how important new ideas in quantum thermodynamics suggest questions regarding the approximations made by the different simulation methods. This is an area where further new collaborations were discussed in a preliminary way.

h) Computational advantages of combined methods**:** this was discussed a number of times, however, more understanding of how the methods are related to each other is needed before progress can be made.
The Workshop Speakers and Participants agreed to have a similar meeting in 2 years from now (Tel Aviv University has been suggested) to evaluate the rapid progress in this field.

**Software infrastructure:** The field of dynamical simulation of open quantum systems out of equilibrium involves complex physics that requires highly sophisticated methodologies and advanced computational schemes. To our knoweldge, the communitycurrently uses a variety of mainly home-made (and not well documented) codes that cover different computational approaches. To better support the development of this field, we recommend that publicly available code repositories should be developed. This will enable enhancement of code development and usage. Platforms such as GitHub may be appropriate for this task.

The participants believe that the scientific exchange of researchers from different disciplines is key to make progress in the most challenging problems of open quantum systems in chemistry, physics and molecular biology. This exchange can be achieved by interdisciplinary conferences and workshops. Moreover, it would be desirable to enable

exchange not only at the level of researchers, but also at the level of PhD and Masters students.

The field of open quantum systems modelling relies strongly on the availability of state-of-the-art computer codes and computing resources. Therefore, the continued funding for the development of modelling software is an important infrastructure requirement: Only when codes are freely available can new ideas be tested and developed by a broad community of researchers. Since new codes that address the most challenging materials and bio-systems applications will push the limits of current supercomputing resources, it is also important that funding of these resources remains significant in the future.

**Funding:** The workshop was very productive and further constant communication between its participants was recommended. Some areas identified there could by funded by EU 2020 program, in particular a new collaborative network in OQS Computational Methods under the COST program. Also, in researchers in Hong Kong will apply to the Research Grants Council to compliment this, and serve as a hub for connecting with other Asian researchers for a broad EU/Asia collaboration.

**Will these developments bring societal benefits?** The broad scope of the different systems of interest means the methods discussed in the conference will be an essential part of understanding the frontiers of biology, chemistry, and physics. All the systems of interests were motivate by concrete applications that have importance for society, such as molecular electronics, heat transport, energy transport, characterization of biological systems, material science, and much more.

Only a few of the computational methods discussed are mature enough for them to be used by the industry at this point. Further tests of scalability and computational costs will be needed before any methods become clear winners for the industry.

The great potential that this field holds for a variety of industrial interests ranging from renewable energy technology such as solar energy conversion to the understanding and utilization of complex biological systems for energy storage and transport calls for further investment and development. One of the suggested ways to fulfil this potential is to hold similar workshops in the future and, when the time comes, to invite relevant industrial and societal stakeholders.