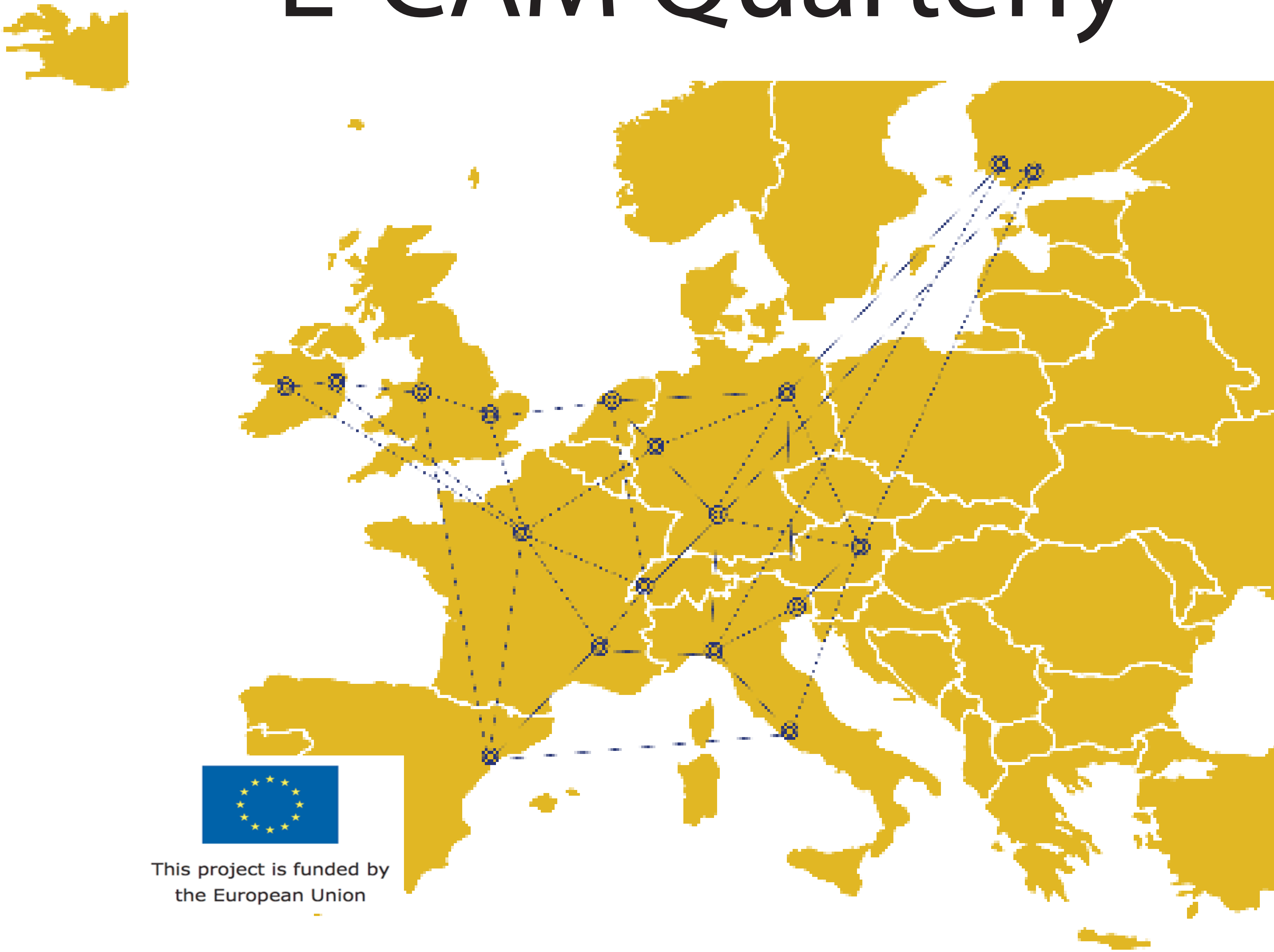
E-CAM Quarterly

Supporting HPC simulation in industry and academia



This project is funded by the European Union

October 2016 -January 2017

Table of Contents

E-CAM Update: October 2016 - January 2017

E-CAM has ramped up its output since October 2016.

The Second E-CAM General Assembly took place in November 2016 at the Maison de la Simulation, Saclay.

An Extended Software Development Workshop (ESDW) on Trajectory Sampling was held in Traunkirchen, Austria 14^{th} - 25^{th} November 2016. The photograph above was taken from the location of the meeting.

Two key changes in E-CAM management have take place with the appointments of Ana Catarina F. Mendonça as Project Coordinator, and Ignacio Pagonabarraga as the Technical Manager, due to the retirement of Dominic Tildesley, who, will continue to lead our interactions with industry.

A multitude of deliverables have been submitted to the European Union, and some 30 sofware modules have been uploaded to the E-CAM repository.

Read on for further details, in this fifth issue of the E-CAM newsletter.

In addition, two pieces have been specially written. The first argues that the E-CAM ESDW model is a big deal, seeking as it does to address 3 key challenges to computational science. And the second piece, written by David Ceperley, poses the question "Does our simulation community need EXASCALE", and offers his perspective.

E-CAM EVENTS

E-CAM Extended Software Development Workshops

Meso and multiscale methods, 3-14 July, 2017 , CECAM-ES

Quantum MD, 17-28 July, 2017 July, CECAM-IRL

Classical MD, 14-25 August, 2017, CECAM-NL

Meso and multiscale modeling 18-29, September 2017, CECAM-DE-MMS

E-CAM State of the Art Workshops

Meso and Multiscale Modelling, 29-31 May, 2017, CECAM-IRL

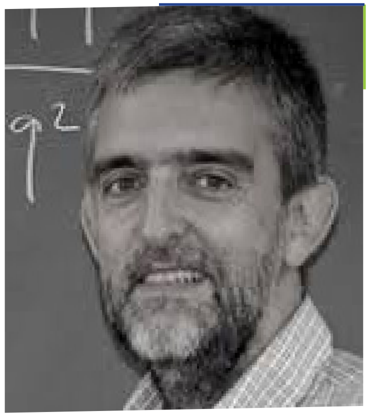
E-CAM Industry Scoping Workshops

Solubility Prediction, 10-12 July, 2017, CECAM-FR-RA

From the atom to the molecule, 18-20 Sept. 2017, CECAM-UK-JCMAXWELL

E-CAM NEW MANAGEMENT BIOGRAPHIES

Since January 2017, Professor-Ignacio-Pagonabarraga is the E-CAM Technical Manager, and director of CECAM HQ at EPFL. Before assuming these posts, he was Full Professor in Condensed Matter Physics at the University of Barcelona. He has developed and exploited mesoscopic computational methods to model the dynamics of soft matter and complex fluids. Recently he has also extended his interests to study the behavior of biological systems at molecular and cellular scales. He has expertise in the development of computational codes for their use in supercomputing facilities. He has gained continued, competitive access to Mare Nostrum and PRACE projects. He has secured funding from the Catalan Government, the Spanish Government, and the European Union, as well as from industrial companies and private foundations. He has led 54 scientific projects competitively funded. Ignacio is a member of: the Scientific PRACE Committe; the Scientific Panel of the Regional Government of Castilla-Leon (Spain); the External Council Board of the School of Mathematics and Physics of the University of Lincoln (UK).



Since November 2016, Dr. Ana Catarina Mendonça is the E-CAM Project Co-ordinator. She obtained a degree in chemistry in 2008 from the Technical University of Lisbon - Instituto Superior Técnico. In 2009 she started a Marie Curie PhD fellowship in the European Project MINILUBES, where she investigated the structure and interactions of ionic liquids at metallic surfaces using MD codes. In 2013 she was engaged as a post-doc for two years at CECAM HQ (EPFL) under the supervision of Prof. Dominic Tildesley, the CECAM Director at the time, where she focused on Monte-Carlo simulations of grafted polymer brushes. She then decided to turn her career around, and in 2015 was appointed to the position of CECAM Administrative Manager. At the same time, she also pursued a programe of advanced studies in project management, and was awarded a diploma on that topic in February 2017.

E-CAM ESDW's : a Pragmatic Response to Challenges facing Computational Science

Donal Mac Kernan
University College Dublin

Three Challenges that face computational science

The challenges that face computational science are well known and include:

- the increasing difficulty of creating simulation software fit for extreme scale/exascale applications and multiscale/multi-paradigm methods;
- the lack of highly skilled computaional scientists well versed both in science and in modern advanced software engineering; and,
- working with the large number of codes written by the indvidual scientist or small group, frequently poorly documented and lacking proper unit testing and other attributes associated with modern software engineering[1].

Four elements of a pragamatic response

E-CAM extended software development workshops[2] seek to respond pragmatically to these challenges by combining software module generation with "training by doing" bringing modern programming standards and techniques into the work practices of the participants, in its four core scientific areas: classical MD, electronic structure, quantum dynamics, and meso/multiscale modeling.

1 Definition of an E-CAM software module

The first element of this response is the very definition of an E-CAM software module[3], and goes beyond the traditional concept of software modules (i.e. of a piece of code possibly including data performing logically discrete operations, interacting through interfaces). In essence E-CAM extends the definition to include inter alia workflow scripts, analysis tools, and test suites as well as traditional subroutines and functions. Obviously, an E-CAM software module should be of potential use to the community, encapsulate some additional functionality, enhanced performance or improved usability for performing computational simulations in the domain areas of interest to E-CAM. Strictly speaking, an E-CAM sofware module should be certified as such by E-CAM to become part of its sofware repository[3]. At a practical level, an E-CAM sofware module must be developed using the E-CAM Gitlab service and include three elements in addition to the code itself:

- documentation;
- a recipe for building the software and any dependencies it may have;
- unit tests (including possibly regression tests);

aswell as basic adherence to an E-CAM code writing style (basically a simple naming convention for variables names, maximum line length etc).

The modules share with the traditional computer science definition the concept of hiding the internal workings of a module behind simple and well-defined interfaces. It is likely that many modules will result from the abstraction and refactoring of useful ideas from existing codes, rather than being written entirely from scratch. Moreover, due to a heterogenous community, the E-CAM software focus is not to enforce a specific language, although the ability to use C as an interfacing language is prefered, but instead to enforce good software programming practise allowing anticipated hardware developments in the near future to be easily exploited.

2 Assembly of a dedicated team of programmers

The second element of this response is the assembly by E-CAM of a small but dedicated team of programmers/software architects and a software manager who provide the general training and quality control infrastructure for ESDW's, and a set of postdoctoral associates (pda's) focused on the development and use of E-CAM modules and other methods on specific scientific questions and applications.

3 Planning, structure, and running of each ESDW

The third element is the planning, structure, and running of each ESDW, which starts typically a year ahead of the event with the submission of an ESDW proposal to CECAM for peer review on a reasonably focused scientific theme. If accepted, the E-CAM software manager, programmers, and corresponding Workpackage coordinators and ESDW organisers will identify the software skills needed for the meeting, and associated training material and documentation specific to individual modules. Participants are given training material well in advance of the ESDW, so as to prepare themselves for the meeting, and select the module(s) they consider the most interesting or relevant to their work.

Each ESDW typically co-locates 12-15 trainees aswell as programmers and others to work on between 6 and 9 modules for about two weeks. At the start of the ESDW participants are introduced to a workflow and tools that facilitate the creation of modules using programming best practices, including the demands and constraints likely to be required to achieve their optimal performance on future hardware. The rest of the meeting includes scientfic lectures associated with the modules to be created, but is largely focused on their generation, testing and documentation in small teams of 2-4 participants working together, with progammers and more seniour scientists passing from team to team to provide additional support and guidance. Typically once per day a 20-30 minute open discussion takes place in which each team outlines the present status of their module and work plan, including any possible difficulties they may have. This provides an opportunity for members of different teams to help each other, including the temporary transfer/addition of members. In addition, if a module is completed early, a participant may choose to work on another module either with the same team or another one. After the EWDW, participants disperse to their home labs and continue module development in remote collaboration for several months, until the teams meet again for a 2-3 day workshop to complete,and upload the software modules, documentation, and unit tests to the E-CAM repositry.

E-CAM ESDW in Traunkirchen, Austria: a walk between coding sessions

4 Online training infrastructure

A fourth element partially completed and still under development is an appropriate online training infrastructure. The E-CAM software repository is the principal access point for users wishing to interact with E-CAM, including training. There, they are encouraged to download and upload software, through a structured scheme of quality control and what is effectively a support infrastructure. This is facilitated through an extensive set of E-CAM services: Redmine, Etherpad, ShareLatex, and in particular Gitlab. The provision, use and further development of these services is an integral part of ESDW's, and one of the principle means by which E-CAM will deliver online material. E-CAM is establishing strong partnerships with PRACE and leading HPC centres in Europe to connect to appropriate training content that can bring the E-CAM user communities to the exa-scale.

Promoting Gender Equality

E-CAM has a policy of promoting gender equality in all of its activities, and is examining how this can be best implemented on a practical level. CECAM activities taking place at CECAM HQ have implemented a policy (which is being extended to E-CAM) regarding child care for parents with yound children attending workshops as a first step in this important direction.

How to collaborate

Collaboration on the development of E-CAM software modules is open to anyone with a serious interest, attending an ESDW is not a pre-requisite. To collaborate, one first should email info@e-cam2020.eu and request an account on the E-CAM Community GitLab Service. Once an account is open, simply click on the webpage and then click on request access on the webpage of the corresponding repository E-CAM/Classical-MD-Modules;

E-CAM/ Electronic-Structure-Modules; E-CAM/Meso-Multi-Scale-Modelling-Modules; E-CAM / Quantum-Dynamics-Modules. An E-CAM ESDW newsgroup also has recently been created.

2016 E-CAM ESDWs and software modules

Four E-CAM ESDW's were held in 2016, 3 of which followed the E-CAM model; and a fourth which followed a different format. Those having an E-CAM format included sessions on software carpentry including coding style and structure; source code documentation and module testing and its documentation; use of the E-CAM GitLab repository; theory, and coding/ languages (e.g. Python) etc. specific to the theme of each ESDW.

Electronic Structure Library coding workshop: solvers

The objective of the first workshop was to develop three libraries focusing on Kohn Sham eigensolvers, Poisson solvers, and atomic solvers, and 9 corresponding software modules: LibOMM, MatrixSwitch, Libpspio, Libescdf, Poke, SQARE radial grid & function, SQARE ODE solvers, SQARE states and FDF. They included codes for solvers of localised orbitals, for computing on a grid and their documentation. In addition to the solvers, data format modules, pseudopotential data file I/O operation module and intermediary interface layer module were also developed[4].

Quantum Mechanics and Electronic Structure

In the the second workshop 6 software modules SodLib, ChebLib, PhysConst, PotLib, AuXMod, ClassMC were created for exact integrators of quantum dynamics: for low dimensional systems, user defined potentials, calculation of quantum time correlation functions[5].

Trajectory Sampling

The goal of the third workshop was to to address using existing md engines the computational challenges caused by rare events through the creation of a library of python modules for path sampling and analysis consisting of 6 modules: Basic shooting and shifting algorithm; Biased path sampling; Aimless shooting algorithm; Reactive flux algorithm; Calculation of the transition state ensemble, and Maximum likelihood optimization of the reaction coordinate.[6].

Wannier Functions

The fourth workshop was built around the Wannier 90 electronic structure community code for generating maximally-localized Wannier functions and using them to compute with high efficiency and accuracy a host of advanced materials properties. Wannier90 is a paradigmatic example of an interoperable software tool, achieved by ensuring that the quantities that need to be input into it are entirely independent of the underlying electronic structure code from which they are obtained. All of the major electronic structure codes in the world have an interface to Wannier90. This workshop was instrumental in catalysing the transition of Wannier90 from a code developed by a small handful of developers to a community code with a much wider developer base. This has been achieved in two principal ways through the workshop: (i) situating the source code and associated development efforts on a public GitHub repository; and (ii) building a community of connected Wannier90 developers by facilitating new and hopefully lasting personal interactions between individuals at the workshop[7].

Software languages

Due to the constraints of working with pre-existing code, software modules were produced in 2016 with an assortment of software languages: Fortran 2008, C with Fortran 2003 bindings, Fortran 95, and Fortran 90. While E-CAM does not wish to be overly prescriptive, some rationalization for future modules is envisiged. For languages, where possible C, C++ and Fortran 2008 (i.e. versions of Fortran that are interoperable with C).

2017 E-CAM ESDW's and software modules

Four workshops are planned for 2017

- The objective of the first meeting, Meso- and Multiscale Modelling, is to generate 6 software modules for: Model and dimension reduction; and, Markov state modelling of open boundary MD[8].
- The objective of the second meeting, Quantum MD, is to generate six modules for: Eigensolvers using iterative schemes; Sparse grids for exact dynamics; Path Sampling, and Quantum-Classical Propagation[9]
- The focus of the third meeting will be on the creation of modules for statistical and machine learning tools analysis of rare events. The

following software modules are planned:Bootstrapping paths; Rate constant calculations via transition interface sampling (TIS); Optimal placement of interfaces; (Single) replica exchange TIS; Sampling multiple state networks; Multiple interface sets; Reweighting schemes; Interface with external order parameter module; Analysis tools for path ensembles[10].

- The fourth meeting is also on Mesoscale/multiscale modelling to make up for the lack of a meeting in 2016. Several modules are envisiged to address the following issues/themes : account for solvent polarizability at a mesoscopic level;analysis of the interplay of polarizability models merged into electro-kinetic software packages to address critically the potential of such modules and identify their limitations; long range interactions in highly parallel environments; multiscale simulation of polymer suspensions simulation of heterogeneous polymeric materials with nanometric inclusions; modeling of active suspensions[11].

For the latter two meetings, the precise software modules that will be developed are still under dicussion.

References/Sources

[1] CECAM's Role in the Development and Support of Scientific Software, 2010.

[2] E-CAM Deliverable 5.2 ESDW guidelines and programme II, Alan Ó Cais, Ana Mendonça, Donal MacKernan, November 2016.

[3] E-CAM Deliverable 6.1 ESDW Technical Software Guidelines, Alan Ó Cais, Dominic Tildesley, Yann Pouillon, and Matthieu Haefele, March 2016.

[4] E-CAM Deliverable 2.1 Electronic structure E-CAM modules I,Liang Liang, Micael Oliveira, Fabiano Corsetti, Yann Pouillon, October 2016.

[5] E-CAM Deliverable 3.1 Quantum Dynamics E-CAM modules I, Sara Bonella, Momir Mališ , Graham Worth, David Lauvergn, and Liang Liang, December 2016 .

[6] E-CAM Deliverable 1.2 Classical MD E-CAM Modules I, David W.H. Swenson, Christoph Dellago, Peter G. Bolhuis, and Jony Castagna, January 2017.

[7] CECAM E-CAM workshop report Wannier functions, Arash A. Mostofi and Mike Payne, November 2017.

[8] Extended Software Development Workshop in meso and multiscale methods, 3-14 July 2017 , CECAM-ES, organized by Ignacio Pagonabarraga Mora

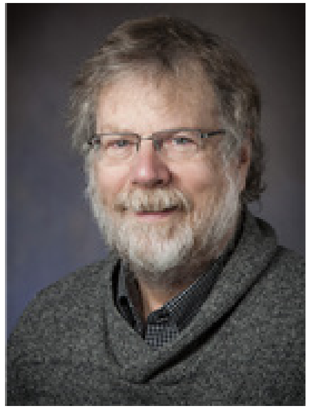
[9] Extended Software Development Workshop: Quantum MD 17-28 July, 2017 July, CECAM-IRL, organized by Sara Bonella, Graham Worth and Donal MacKernan

[10] Extended Software Development Workshop Classical MD, 14-25 August, 2017, CECAM-NL, organized by Peter Bolhuis.

[11] Extended Software Development Workshop in Meso and multiscale modeling 18-29 September 2017, CECAM-DE-MMS, organized by Carsten Hartmann and Luigi Delle Site.

Does our simulation community need EXASCALE ?

David Ceperley
University of Illinois Urbana-Champaign



The computer simulation of electrons, atoms, molecules, and their assemblies in soft and hard matter is foundational for many scientific disciplines and important commercially. Exascale computing is coming and our community should take part as are our colleagues in lattice gauge theory, climate modeling, cosmology, genomics and other disciplines.

These resources present a great opportunity. The proposed exascale machines will be massively parallel with millions of interconnected processors. They will likely require new programming and languages to use optimally. Their memory will be such that one will need to explicitly control how data is stored and accessed since the interconnect bandwidth and latency will limit the use of global memory. The funders of the new machines will want the machines to be used to their full potential and to generate science commensurate with their costs. Clearly not every simulation algorithm will be useful in this environment.

Applications need to promise more than incremental progress

We have to think which applications warrant this resource and which are doable in the next decade. Applications need to promise more than incremental progress, e.g. not just another decimal place on the correlation energy of the homogeneous electron gas. Most of our current projects should stay on their current platforms. But every so often, there is a paradigm shift, where something that was once thought too difficult or impossible becomes routine. Such occurred in 1985 when R. Car and M. Parrinello [1] linked Molecular Dynamics with a DFT evaluation of the Born-Oppenheimer forces. We must look for such opportunities provided by the increased computer power of exascale machines.

Material design by computer

One candidate for an application (really a whole family of applications) is material design by computer. The US Material Genome Initiative [2] and other similar efforts worldwide have been funded to work towards this goal. Materials design, currently done experimentally, for example with a "shake and bake" procedure, is technologically and commercially very important -- but costly. There has been much progress during the past decade but its promise is still largely in the future. Our community, for example as is present in E-CAM, is focused on accurate predictions of semi-empirical models of physical systems. I believe that materials design will require calculations that can be run without experimental input with a reasonable probability of a successful prediction (i.e. that a certain structure can be made, that it will be stable and have predicted properties). While it is true that computer design has made a small impact to date, eventually it will, even though we don't know when.

Computer design of materials is a good candidate for an exascale application because one can do searches in parallel; each simulation/electronic structure calculation occupying a small part of the machine. The power of the machine can allow thousands or millions of candidate structures to be examined in parallel. But some things need to happen first.

Accuracy of the electronic structure a fundamental consideration

One consideration is the fundamental accuracy of the electronic structure calculation. A key property that we need is the Born-Oppenheimer surface of the ions to determine the stable crystal structure, and its electronic properties such as the response to electromagnetic fields. To be accurate one needs to resolve the energy differences between different structures and energy barriers, since without knowing the structure one cannot even begin to describe its properties. Room temperature provides a typical energy scale. However, since (100 K = 0.3mH = 8meV), we are interested in very small energy differences relative to typical electronic energies (1 Hartree=27.2eV=315 775K). Although we require quite accurate energies, current methods are getting close for many physical systems!

Currently, for structure searching one often uses DFT calculations for stability. However, there is now a multitude of DFT functionals, and it's not obvious which one is best. Without empirical information one cannot decide. A recent article [3] suggests that the current semi-empirical approach to improving functionals does not lead to systematic path toward the exact functionals. Today's best functionals do not typically meet the accuracy criterion without empirical tuning and selection. For example, we cannot even make confident predictions of the ground state structure of solid hydrogen[4], the first element and one of the simplest elements. DFT is good for interpolating between materials where the accuracy has been confirmed experimentally. However, the space of potential materials is so vast that one cannot rely on a semi-empirical method. It is likely that the best material for a given application will be made of a particular combination of elements that has not been looked at with high quality experiments or high accuracy electronic structure methods and would not be in the data base that is used to construct the functional or model.

Quantum Monte Carlo methods

I am an advocate of Quantum Monte Carlo (QMC) methods [5]. These methods are the generalizations of Molecular Dynamics (MD) and Monte Carlo (MC) to quantum many-body systems and are particularly needed when mean-field based methods fail. For some systems, QMC methods are exact in the sense that classical MD and MC are exact, but to simulate electrons one runs into the fermion sign problem. No algorithm has yet been demonstrated that gives a controlled error in polynomial computer time as the number of electrons goes to infinity. But because the fixed-node or fixed-phase methods give upper bounds for the energy in polynomial time, we have unambiguous internal information about the accuracy of the fixed-node estimate, and we know when we have an improvement. In addition, exact (controlled) estimates can be performed for small electron systems. Path integral methods can treat non-zero temperatures and quantum nuclear effects. QMC is the most general, robust algorithm for solving the equilibrium electronic structure problem and can be shown to reach the requisite accuracy in many cases. In addition, the stochastic nature of its procedure can be incorporated together with a classical MC simulation without seriously impacting the computational effort. This allows one to study a disordered system such as dense liquid hydrogen [6] with a higher and better-controlled accuracy than would be obtained using DFT forces.

I mention QMC here because it is a leading exascale electronic structure algorithm. Diffusion Monte Carlo works by evolving independent walkers using the Hamiltonian. Evaluation of the trial wave function, its gradient and Laplacian dominate the computer effort. The independence of the walker's evolution results in nearly perfect parallelism for ten of thousands of processors. In most applications one does not need millions of walkers but more parallelism can be achieved by simultaneously looking at different compounds or boundary conditions. Of course, to make the QMC method more applicable there are many technical problems to solve in addition to the sign problem: e.g., the elimination of core electrons in a more accurate way, and better scaling to more electrons. Certainly other methods are indispensable but QMC provides a benchmark of their accuracy if experiment is not available.

Difficulty of going from the nano to the mesoscale

A second problem, I want to mention is the difficulty of going accurately, robustly and automatically from the nanoscale to the mesoscale. Since the goal is to fabricate materials simulation techniques must be able to handle the typical complexity of real materials that include defects, impurities, and dynamical processes. Material designers need to consider effects of non-zero temperature, entropy, electronic and optical properties and formation routes. It is not realistic to think that an expert in the theory and practice of electronic structure will be always involved in a materials design project; one needs a "black box" solution to the multiscale problem. The accuracy of an electronic structure calculation needs to be extended to large systems in an automatic way.

One approach is to use accurate electronic structure methods that are limited to small systems (say fewer than a thousand electrons) to generate data that can be used to generate potential energy surfaces that can be used in MD [7] or with another model. The MD simulation can then be run for millions or billions of atoms and to make estimates of some of the needed properties at a much larger length or longer time scale. This needs to be done routinely but tailored for a particular application. An important obstacle is to find a basis set appropriate to describe complex molecular interactions that is universal, accurate and reasonably compact. Whether this can be achieved in general and still maintain the required accuracy is an open question.

An organizational problem of material design is that research and development cuts across different research communities: it needs a larger, longer-scale effort such as could be provided by E-CAM. Porting and optimization of a single code to the new machines could take several person-years of work; global cooperation will be advantageous.

Changes in computer architecture are annoying and upsetting. In my career I have seen more than a dozen of such shifts each requiring a big investment to stay current. However, I am confident that the simulation community will stay involved with high performance computing and, as a consequence, our community will reap the intellectual, scientific and financial rewards.

References

[1] R. Car and M. Parrinello, Phys. Rev. Letts. 55, 2471 (1985).
[2] https://www.whitehouse.gov/mgi
[3] M. G. Medvedev et al. Science 355, 49 (2017); S. Hammes-Schiffer, Science 355, 28 (2017).

[4] M. A. Morales, Phys. Rev. B 87, 184107 (2013).
[5] R. M. Martin, L. Reining and D. M. Ceperley, Interacting Electrons, Theory and Computational Approaches Cambridge (2016); L. K. Wagner and D. M. Ceperley, Rep. Prog. Phys. 79 094501 (2016).
[6] C. Pierleoni et al., Proc. Nat. Acad. Science (US) 113, 4953–4957 (2016).
[7] M. J. Gillan et al , J. Chem. Phys. 139, 114101 (2013).

Second E-CAM General Assembly

Luke Drury[1,2] and Donal Mac Kernan[2]
[1] Dublin Insitute for Advanced Studies [2] University College Dublin

The second General Assembly of E-CAM was held in the Maison de la Simulation, Saclay, Paris on the 7th and 8th of November 2016. The meeting was attended by representatives of all E-CAM partners, all postdoctoral researchers employed on the project, the two scientific programmers and the software manager, and a number of invited guests. The meeting marked a significant stage in the evolution of E-CAM with some high-level personnel changes, a move of the coordination office, and the start of a strategic discussion about how E-CAM should respond to the changing environment for HPC in Europe.

Le Plateau de Saclay- a Leading Technology-Academic Hub

Before delving into the details of meeting, it is worth remarking on the splendid location provided by its hosts, "La Maison de la Simulation", and the context in which they work. La Maison de la Simulation is a large research laboratory on numerical simulation founded jointly by the French Atomic Energy Commission (CEA), the French National Centre of Research (CNRS), the Institut National de Recherche en Informatique et Automatique (INRIA), and the Université Paris-Saclay, located in Saclay about 20 km South West of Paris. Its activities cover transverse domains (applied mathematics, computing, visualization, ...) as well as high performance computing applications. While Saclay in the past would have been known simply as a location on a Roman road, later as place of feudal unrest, and later still as a source of water for the palace of Versailles, today it, or rather "le Plateau de Saclay" has become a major technology-academic hub, through a sustained effort by the French state starting shortly after the end of the second world war. This transformation has included most recently the formation of the University Paris-Saclay in 2015 through in part the fusion of multiple leading research centres, with the ambition to be in the top ten in the world of academic institutions, and thereby facilitate the realization of "le Plateau de Saclay" as a technological centre similar to Silicon Valley. Much has already happened. The academic insititions located there already boast several Nobel prizes in physics and field medals, while the hub includes more than 50 leading technolgical companies, many of which increasingly use simulation in the design process. This focus on the use of simulation to enhance the competiveness of European industry is a goal shared by E-CAM.

Second E-CAM General Assembly, Maison de la Simulation, Saclay, Paris

Changes at the Top

Opening the meeting the project Chair, Luke Drury, paid tribute to the "heavy lifting" that had been done by Kate Collins as project administrator in Dublin as well as to the engagement and energy invested in the project by the outgoing technical coordinator and Director of CECAM, Dominic Tildesley. The general assembly supported these thanks by warm acclamation. The chair observed that while the management of E-CAM had been effective, it had become clear that a closer coordination between E-CAM and CECAM, a sharper distinction between governance and management, and a more centralised management would be beneficial. The regrettable departure of Kate Collins now presented an opportunity to address these issues at the same time as the hand-over from Dominic to the incoming Director of CECAM and ex officio technical manager of E-CAM, Ignacio Pagonabarraga. This was accepted by the assembly which voted unanimously to move the management and coordination of the project from Dublin to Lausanne subject to the agreement of the European Commission. The meeting reviewed progress in the various work-packages with particular emphasis on industrial engagement (a particular strength of E-CAM). There were lively and useful discussions on the various types of meetings used by E-CAM (state-of-the-art workshops, scoping workshops, extended software development workshops) as well as on the collaborative software tools used to support the project (redmine, github, sharelatex).

European Technology Platform for High Performance Computing

The meeting took advantage of being in Saclay to invite presentations from Jean-Phillipe Nomine of ETP4HPC, the European Technology Platform for High Performance Computing and Edouard Audit, the chair of the Energy Orientated Centre of Excellence EoCoE. ETP4HPC is the major public-private partnership between the EU and Industry to promote uptake of High Performance Computing as part of the overall digital agenda. The Centres of Excellence (CoEs) are one of the instruments that fall under the umbrella of ETP4HPC as well as a broad range of Future and Emerging Technology projects (FETs) and in the near future the Extreme Scale Demonstrator projects (ESDs). Jean-Phillip gave a valuable overview of the work of ETP4HPC and emphasised that there is a short window of opportunity for the CoEs to be involved in the strategic planning of the next call. The need for a greater engagement with the ESDs in particular, and more generally in the work of ETP4HPC including our sister CoEs, was one strong message that came from the informal "check" held by the commission in May 2017 in Brussels and attended by the Chair and the Technical Manager the other being the need to adopt a more rigorous approach to the quality control and format of deliverables.

Apart from these management and strategic aspects, the meeting was also a great opportunity for participants to network and associate faces with names. The pleasant environment of the Maison de la Simulation, the excellent French catering, and a very enjoyable social dinner all contributed to this important aspect of the meeting for which the chair thanked the local hosts in closing the meeting. The postdoctoral researchers, the software manager and the programmers stayed on for an extra day to discuss in detail the development of software modules for the E-CAM repository, and software carpentry.

E-CAM DELIVERABLES OCTOBER 2016 - JANUARY 2017

The output rate of E-CAM measured in terms of deliverables produced has greatly accelerated since October 2016. 12 Deliverables were submitted to European Commission: D1.1, D1.2, D2.1, D3.1, D4.1, D5.2, D6.3, D7.2, D8.1, D9.1, D11.1, D11.5 . All but D8.1 (relating to industry collaboration is confidential) and D11.1 (relating to internal management also confidential) were published at https://www.e-cam2020.eu/deliverables/
They are summarised below.

D1.1
Identification/selection of E-CAM MD codes for development

Summary
Many processes in nature and technology are characterized by rare but important events, which occur on time scales orders of magnitudes longer than basic molecular motions. Such processes, which, for instance, include chemical reactions, protein folding and first order phase transitions, are difficult to simulate with classical molecular dynamics (MD) simply because of the extreme time scales involved. The main goal of Work Package 1 (WP1) is to develop software tools capable of dealing with rare events and complex free energy properties, thus extending the time scales accessible with regular MD. In this report, we will first briefly review current algorithms for the simulation of rare events and related algorithms for the computation of free energies. We will then discuss software packages that make these methods available. Based on this information, we will then give an overview of the software modules to be developed within WP1 of E-CAM. Finally, we will describe how we will benchmark some popular molecular dynamics engines on which the modules to be developed in WP1 will be based.

D1.2
Classical MD E-CAM modules I

Summary
In this report, 9 software modules in classical dynamics are presented. Of the 9 modules, 7 have been incorporated into the core code of OpenPathSampling (OPS). They are: path density; direct (on-the-fly) flux and rate calculation; improved input for OPS networks; new WHAM code for OPS; flux/rate from existing trajectories; OPS snapshot features; and two-way shooting . The other 2 modules build on OpenPathSampling, but remain separate. They are: annotated_trajectories; and ops_piggybacker,
Together, these 9 modules represent improvements and new features in

software for trajectory sampling and for studying the thermodynamics and kinetics of rare events.

Each module is thoroughly tested with unit tests, and includes in-code documentation as well as external documentation in the form of Jupyter notebook examples. In this report, a short description is written for each module, followed by a link to the respective Merge-Request on the GitLab service of E-CAM. These merge requests contain detailed information about the code development, testing and documentation of the modules.

D2.1
Electronic structure E-CAM modules I

Summary
Work Package 2 of E-CAM focuses on selecting software functionalities that are common to many electronic structure implementations, important for the coding and efficiency of codes, and mature enough to allow for a good definition of standards and interfaces. E-CAM collaborates with the Electronic Structure Library Project (ESL) whose goal is to build a community-maintained library of software of use for electronic structure simulations. Starting from the Extended Software Development Workshop at Zaragoza, the development of new libraries revolved around the broad theme of solvers. In this report, 9 software modules in electronic structure, which are related to the ESDW held by E-CAM at Zaragoza in June 2016, are presented. The 9 modules are respectively named: LibOMM, MatrixSwitch, Libpspio, Libescdf, Poke, SQARE radial grid & function, SQARE ODE solvers, SQARE states and FDF. They include codes for solvers of localised orbitals, for computing on a grid and their documentation. In addition to the solvers, data format modules, pseudo-potential data file I/O operation module and intermediary interface layer module are also developed. In this report, a short description is written for each module, followed by a link to the respective Merge-Request on the GitLab service of E-CAM. These merge requests contain detailed information about the code development, testing and documentation of the modules.

D3.1
Quantum dynamics E-CAM modules I

Summary
Software development in quantum dynamics has so far been less systematic than in other fields of modeling, such as classical molecular dynamics or electronic structure. Thus, E-CAM WP3 will also provide and environment to stimulate the transition from in-house codes, often developed and used by single groups, to the development of modular, community-based, packages capable of multiple functionalities and adopting common benchmarks.

In this report 6 software modules in quantum dynamics which are related to the ESDW held by E-CAM at Maison de la Simulation Saclay, in July 2016, are presented. The 6 modules are respectively named: SodLib, ChebLib, PhysConst, PotLib, AuXMod, ClassMC. They include codes for exact integrators of quantum dynamics for low dimensional systems, potentials, calculation of quantum time correlation functions, and their documentation. In this report, a short description is written for each module, followed by a link to the respective Merge-Request on the GitLab service of E-CAM. These merge requests contain detailed information about the code development, testing and documentation of the modules.

D4.1
Identification/selection of E- CAM meso and multi-scale modeling codes for development

Summary
The present report has analyzed a variety of existing modeling approaches to analyze the equilibrium and non-equilibrium properties of complex systems at a coarse grained and multi-scale level. The analysis has shown the viability to exploit existing expertise within E-CAM and use this expertise to select a number of multi scale and coarse grained codes. Developing these codes, and producing new modules associated to them and that can also work transversally with other computational packages provides a fruitful perspective to make progress. E-CAM activities, such as the state of the art

workshops, will constitute the natural forum to expose and confront the selected codes to complementary approaches. The outcome of such activities will help to decide the need to enlarge the palette of codes and/or promote the development of transverse modules that can interface selected codes with complementary coarse-grained and multi-scale software packages.

The report also shows that all existing codes are amenable to be run on supercomputing environments. We will exploit E-CAM activities to probe more accurately their scalability within Partnership for Advanced Computing in Europe (PRACE). These scaling tests will serve to provide the community with a deeper understanding of the potential of the software within High Performance Computing (HPC). Not all codes are equally supported on CPUs and GPUs. We identify a general need to be able to use both type of hardware. This fact will be taken into account when developing modules. The report shows that the selected codes will be developed for specific new uses according to industrial interest. The outcomes of these projects in terms of modules associated to the mentioned codes will be of wider use.

D5.2
ESDW Guidelines and Programme II

Summary
E-CAM delivers on average four Extended Software Development Workshops every year each focused on software development in one of its four core scientific areas: classical MD;electronic structure; quantum dynamics; and meso and multi-scale modeling. The purpose of an ESDW is twofold. On the one hand they are a mechanism for generating software modules for inclusion in the E-CAM repository. On the other, they are an integral part of the E-CAM training programme and represent the primary "training by doing" component.

The present deliverable is an updated version of deliverable D5.1 on the current guidelines for Extended Software Development Workshop events. These guidelines for content, structure and output help to ensure that the workshops are run consistently across the scientific Work Packages and meet the quality standards for E- CAM software. In addition to refining the guidelines of D5.1, this deliverable defines:

- the scope of training at Extended Software Development Workshop (ESDW) events;
- the online material accessible through the E-CAM software repositories and website;
- the role of the programmers;
- the concept of module in E-CAM and its acceptance criteria;
- a day-to-day set of recommendations by previous ESDW participants and the certification of workshop attendees.

The programme of ESDWs for the second year of the project is also defined within this document. These guidelines are intended to be a living document which evolves to reflect experience gained in running the ESDWs and thus they are subject to further revision based on the outcomes of each year's activities.

D6.3
E-CAM Software Platform I

Summary
This deliverable describes the provision of online services in the E-CAM project which together form the E-CAM web platform. The primary landing point for information about the resources of the project is the E-CAM project website. This site alone covers the basic requirements of the E-CAM User Portal:
E-CAM library of software modules and interfaces
The software modules of E-CAM are linked through the website, in addition to the rendered documentation that result from them.
Access to E-CAM's resources
All of E-CAM resources are described and available through the E-CAM website. This includes our upcoming E-CAM events and the E-CAM online services.
Make requests for software developments
We would like to deal with development requests directly on a case-by-case basis with the relevant Work Package (WP) leader and the Software Manager being in direct contact with the person making the request. For this

reason we have created a very simple technical first contact page in order to channel users to the correct WP.

Register for events

All E-CAM events are managed through Centre Européen de Calcul Atomique et Moléculaire (CECAM) with registration for events happening through them. On the E-CAM website we provide detailed descriptions of the E-CAM events and links to the registration process of CECAM.

Web infrastructure for teaching tools

An E-CAM training page is under development to provide a list of training tools for the project beneficiaries, participants and the wider community in the High Performance Computing skills space. More complex teaching infrastructures are planned but will take considerable time to reach maturity.

However, E-CAM has delivered a number of additional online resources and capabilities that include:

- Software modules are contributed to E-CAM through the documentation repository of the relevant research related Work Package. The sources for the documentation are stored on the E-CAM GitLab service with rendered documentation available through ReadTheDocs.org.
- A Kanban service has been made available to facilitate a lower setup overhead and direct interaction with the issue reporting features of GitLab.
- The Redmine service is used to manage larger software projects and track related issues. It allows users to manage multiple projects and associated sub-projects. It features project wikis and forums, issue tracking, time tracking, and flexible, role-based access control.
- An Etherpad service has been provided for a number of participants to simultaneously add to meeting notes and minutes during an online collaborative meeting.
- A ShareLatex service has been added to facilitate the collaborative production of publication-quality papers using LATEX.

Over the lifetime of the project these online services will mature and expand, particularly in the case of online learning.

D7.2

E-CAM software porting and benchmarking data I

Summary

The purpose of this deliverable is to deliver a joint technical report on results of porting and optimisation of at least 8 new modules out of those developed in the ESDW events to massively parallel machines and the benchmarking and scaling of at least 8 new modules out of those developed in the ESDW events on a variety of architectures.

This deliverable also prescribes a work-flow to ensure that, going forward, the porting effort is efficiently integrated into the ESDW events and effectively communicated to the end-user community. This work-flow includes:

- creating reproducible and efficient software builds using EasyBuild,
- a benchmarking work-flow using JUBE,
- application optimisation with Scalasca.

Timing issues related to the schedule of ESDW events did not permit us to completely follow this template. Instead, the Software Manager requested two representative applications from each research Work Package that are in common use and likely to be components of ESDW events. These 8 applications were then ported, optimised and scaled on the HPC resources available to the project. The purpose of the exercise was to create reference performance levels for the E-CAM users for these applications, to gain experience on the hardware infrastructures available to the project and to expose the E-CAM programmers to the tools to be used within the proposed work-flow.

The particular applications that were investigated were for:

- Classical Molecular Dynamics: GROMACS and LAMMPS,
- Electronic Structure: Quantum ESPRESSO and CP2K,
- Quantum Dynamics: PIM and Quantics,
- Meso- and Multi-scale Modeling: Ludwig and DL_MESO_DPD.

The scaling behaviour of these applications on cluster systems with and without the use of accelerating co-processors (GPUs and Xeon Phi) was investigated

D8.1
Industrial Collaboration
Confidential

D9.1
E-CAM Public Wiki and newsletters I

Summary

The objective of the present deliverable is to report on the dissemination activities of E-CAM during the last 4 quarters regarding (a) the generation and updating of a Wiki (or equivalent) describing E-CAM's activities and, (b) E-CAM newsletters; published in previous 4 quarters, which are download-able at the E-CAM website.
The pdf version of this report includes extensive use of hyper referencing to online items, and are visible as blue clickable text.

The primary access point to E-CAM is its website E-CAM. It is rendered using WordPress. The home page includes 4 primary button links on how to: attend a workshop; become a partner; access the E-CAM software library; and, ask a technical question.

It's main menu bar has links for more detailed material.

General information is at the webpage about E-CAM which includes links to the 4 E-CAM scientific workpackages, brief descriptions of the E-CAM software repository, and upcoming and past E-CAM Events

Four issues of the E-CAM newsletter have been published in this reporting period. Each newsletter typically includes a list of upcoming events, an editorial or commentary on an important scientific/technical/industrial topic, a list of deliverables published in the last quarter, brief reports on E-CAM events in the last quarter, and recent news. That said, the format is not rigid, and additional items are added on occasion.

D11.1
Management Report
Confidential

D11.5
Data Management Plan

Summary

E-CAM activities can be divided into three complementary areas and associated types of data object: 1) software and algorithm development; 2) advanced training in the production, documentation and use of scientific software; 3) outreach to industry and academia to identify evolving scientific software needs and opportunities. The objective of the present deliverable is to describe at a high level how we plan to manage the data generated from these activities.

All code developed will follow, where practical, the Extended Software Development Workshop (ESDW) Technical Software Guidelines, and will be subject to the quality acceptance criteria defined by the Guidelines for the ESDW's. Software development will be guided by requests from end-users through the E-CAM website, through industry scoping workshops and through direct collaborative projects with industry. The associated data relating to the project will be maintained in the form of software and metadata repositories (version-controlled through Git).

E-CAM will create software modules rather than complete packages to allow for the rapid inclusion of new algorithmic ideas and their effective dissemination. The project will interface these modules to existing software codes either directly, or through translators. Where a software module is standalone and generated entirely within E-CAM project, the source code software repository will be created and maintained on the E-CAM GitLab service. Where modules relate to externally maintained software packages, appropriate links will be provided in the metadata repository as well as patch files that detail the changes/additions to the source code.

Training material generated directly by the E-CAM project will be made publicly available through a training material repository on our GitLab service following the best practices and guidelines of the Software Carpentry Foundation. The training material will be created using the Software Carpentry approach, meaning that the content will itself be stored in a version-controlled repository.

Our State of the Art and Industry Scoping workshops are all required to write reports of their activities. Templates to facilitate community exploitation of data produced through these events have been defined, and are being updated to maximise their impact.

There is also a quarterly newsletter distributed to all project partners and the E-CAM mailing list as well as other reports scheduled for distribution to that list, including information regarding the E-CAM ESDW and industry pilot project outcomes. All of these reports will be stored at EPFL and be accessible through the E-CAM website.

Editor:
Donal Mac Kernan
Contributions:
David Ceperley, Luke Drury and Donal Mac Kernan

info@e-cam2020.eu
www.e-cam2020.eu