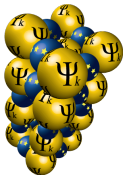

Ψ_k **Scientific Highlight Of The Month**

No. 133 **December 2016**

Three European Centers of Excellence in Computational Science

R. M. Nieminen[1], S. Bonella[2], L. Drury[3], M. Scheffler[4], and E. Molinari[5]

[1] *Department of Applied Physics, Aalto University, Helsinki, Finland*

[2] *CECAM-HQ, École Polytechnique Fédérale de Lausanne, Switzerland*

[3] *Dublin Institute for Advanced Studies, Dublin, Ireland*

[4] *The Fritz Haber Institute of the Max Planck Society, Berlin, Germany*

[5] *CNR-Insitute Nanoscience and University of Modena and Reggio Emilia, Modena, Italy*

Three European Centers of Excellence in Computational Science

R. M. Nieminen[1], S. Bonella[2], L. Drury[3], M. Scheffler[4], E. Molinari[5]

Introduction

In 2015, the European Commission funded nine new Centers of Excellence (CoEs) in Computational Science. Three of these CoEs, NOMAD, E-CAM, and MaX, originate directly from the extended Psi-K and CECAM communities. Researchers from these communities are also present in other CoEs, such as EoECO (focused on energy) and BioExcel (simulation of biological systems). This remarkable success brings about 15 million Euro over a period of five years (2015-2020) to our field and provides us with an unprecedented opportunity to develop and apply advanced simulation methods, create new software tools, develop systematic interactions with industrial partners, and play a key role in defining the European modeling and simulation ecosystem in the next five years and beyond.

In this article, the three CoEs are presented, in order to highlight what they are developing and can offer to our community. We also provide a short summary of the genesis of the call and of the overall goals set by the European Commission for the CoEs. The coordination and synergy benefits as well as collaborations enabled by the three CoEs are also emphasized.

In 2014, the European Commission decided to establish a limited number of user-driven Centers of Excellence in the application of HPC for addressing scientific, industrial or societal challenges. The decision was enforced, within the Horizon 2020 program, with the E-INFRA5 call in 2015. This call is part of a broader scheme, supported by different funding pillars at the EU level, to enhance the European research and computational infrastructure, with specific focus on fostering academic and industrial collaborations that require advanced methods and top of the line computational facilities. The concerted actions in this scheme aim at expanding the High Performance Computer (HPC) user base in industry, particularly SMEs, boosting European competitiveness, developing excellence in key scientific and technological domains, and fostering innovative solutions for grand societal challenges.

The specific role of the CoEs is to create multidisciplinary research environments in which domain expertise and HPC, software and algorithm expertise progress side by side. These Centers federate available competences across Europe, creating a network

[1] Department of Applied Physics, Aalto University, Helsinki, Finland

[2] CECAM-HQ, École Polytechnique Fédérale Lausanne, Switzerland

[3] Dublin Institute for Advanced Studies, Dublin, Ireland

[4] The Fritz Haber Institute of the Max Planck Society, Berlin, Germany

[5] CNR-Insitute Nanoscience and University of Modena and Reggio Emilia, Modena, Italy

of excellence and ensuring synergies with national/local programs, with relevant stakeholders (such as large industrial conglomerates and SMEs), and with other relevant players in HPC (such as PRACE or the ETP4HPC).

The CoEs can be 'thematic' (addressing specific application domains such as, for example, materials, energy, life sciences), address general issues in computational science (algorithms, analytics, numerical methods etc.), 'challenge-driven', addressing societal or industrial challenges (ageing, climate change, clean transport etc.), or a combination of these types. They must be user-driven and contribute to strengthen not only the European research infrastructure, but also its human resources base by providing advanced training and transfer of knowledge between academia and industry.

In addition to the general goals stated above, the CoEs' expected impacts include improving access to computing applications and expertise for academia and industry, and promoting scientific excellence through new methods and modeling strategies, better code performance and portability and better code maintenance and availability.

The three CoEs presented in this article tackle different aspects of the challenging goals set by the European program. Together, they represent an impressive effort by the electronic structure, soft matter, quantum and statistical mechanics, overlapping, communities to engage in and contribute to the success of this initiative. They also create a critical mass of resources and initiatives in our fields whose benefits can extend beyond the boundaries of the individual projects by creating opportunity for collaborations and scientific exchange.

E-CAM: supporting software, training, and discussion in simulation and modeling
www.e-cam2020.eu

The E-CAM Centre of Excellence is based around the experience and scientific community of the extended CECAM family, established over more than four decades, as well as the computational and hardware expertise of the European partnership PRACE. Since its founding in 1969 CECAM has been at the forefront of promoting computational simulation as a powerful research tool for understanding and predicting the properties of materials, including biological systems, both at European and international level. E-CAM in partnership with PRACE and the other Centers of Excellence will now build on and broaden this by adding a strong infrastructural element as part of the European Union's ambitions for HPC under Horizon 2020.

At the scientific level E-CAM has chosen to prioritize four broad areas of science that are central to the interests of CECAM. **These vertical scientific pillars are classical molecular dynamics, electronic structure calculations, quantum dynamics, and meso-scale and multi-scale modelling.** In addition to covering most of the scientific interests

of the extended CECAM community these areas were chosen because they are also of considerable interest to industrial researchers. At the institutional level E-CAM takes advantage of the distributed nodal structure of CECAM and brings together 18 institutions with a good geographic distribution across Europe including three PRACE centres.

The E-CAM team is led by experts coordinating the work in its different workpackages. The PIs of the project include: Luke Drury (coordinator), Dominic Tildesley (CECAM-HQ, EPF-Lausanne), Christoph Dellago (CECAM-AT, University of Vienna), Mike Payne (CECAM-UK-MAXWELL, University of Cambridge), Sara Bonella (CECAM-HQ, EPF-Lausanne), Ignacio Pagonabarraga (CECAM-ES, University of Barcelona), Ana Catarina Mendonça (E-CAM project manager), Alan O'Cais (FZ Jülich, E-CAM software manager), Godehard Sutmann (FZ Jülich), Donal MacKernan (CECAM-IRL, University College Dublin).

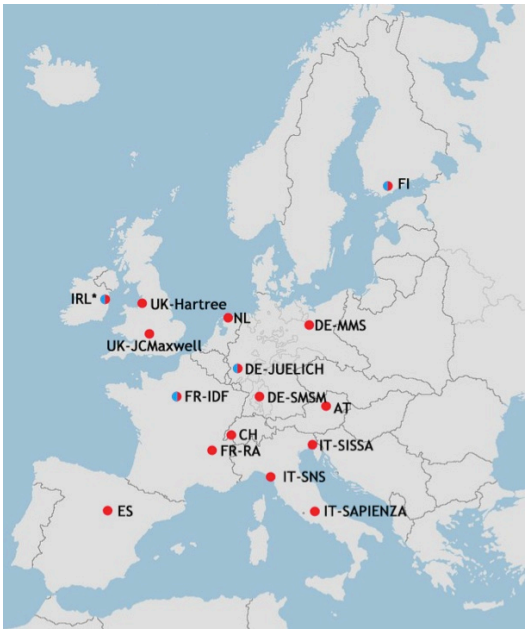

Figure 1: Distribution of E-CAM beneficiaries. In red CECAM nodes in the project, in blue HPC centers

E-CAM already has fourteen industrial partners and aims to significantly increase this number over the life of the project as part of its roadmap to long-term sustainability. In its first year of activity, 7 post-doctoral researchers and 2 software developers based at different beneficiary institutions were hired to work on scientific and software developments in a set of pilot projects specifically targeted at academic and industrials end-users.

E-CAM complements the scientific focus of CECAM by emphasizing **three transversal activities** that cut across the various scientific fields represented in CECAM and provide essential support for the work of CECAM and PRACE (and are thus infrastructural in character). These are, first, **the creation of opportunities for engagement with industry and the involvement of industrial scientists and partners in the activities of E-CAM**. Cutting edge techniques and methods used in academic computational science currently take too long to transfer to industry, and conversely, interesting and significant problems and techniques thrown up by industrial research are not efficiently communicated back to the world of academic research. E-CAM will address this by building communication channels between these two worlds at both personal and at institutional level. This is a key defining characteristic of E-CAM; it is determined to break down, or at least lower, the barriers between academic and industrial research (including both SMEs and multi-nationals) in its areas of science. Just as the distinction between basic and applied research is at root a false dichotomy, so too is that between

industrial and academic research; ultimately what matters is good and interesting science wherever it is done and excellence can be found outside the pale of academia.

The second transversal activity of E-CAM focuses on **the need for new and improved algorithms and code modules**. There is growing recognition that Moore's law is beginning to break down and that we can no longer simply expect newer and faster processors to improve the performance of our old legacy codes. Recent advances in computational power depend on either massive parallelism, or specialist hardware accelerators, or increasingly both; this means that the old legacy codes need to be rewritten to exploit these possibilities and, in many cases, that totally new algorithms have to be implemented. E-CAM will do this in close collaboration with the other Centers of Excellence and the European partnership for supercomputing PRACE. In many ways, it is not just a question of new software, but of finding new and more efficient ways to produce robust and well-documented software that is easily ported and adapted for new challenges. At present people developing new codes waste too much time on relatively routine aspects replicating what has been done before. Frameworks, tools, documentation and standards need to be developed to allow better use of the creativity of programmers and the extraordinary success of many free-software projects in using distributed networks of volunteer programmers needs to be replicated in the sphere of scientific software. This also throws up the difficult question of how the work done by scientific programmers should be adequately recognized and rewarded by career progression; we have all seen good students whose careers have been blighted by writing code and not scientific papers. It is clear that we overvalue publications (especially in the so-called high-impact journals) but it is not obvious how the system should be changed. However, for the long-term health of the field we do need to think about this problem and seek solutions. The same applies to the question of gender stereotyping and implicit biases. While a secondary focus of E-CAM these are issues that will not just go away and we need to be conscious of them. The primary output of this activity will however be the online E-CAM repository of tested and robust code modules, snippets, wrappers and tests that support modern simulation science across a range of platforms. These will be developed in association with PRACE and other FET projects to enable efficient exploitation of the current and next generation of European supercomputer facilities.

The third transversal activity is associated with the broad issue of training, mentoring and continuing professional development of the European pool of human capital available in support of computational simulation. This is crucial and a clear connection to one of the core missions of CECAM. The health of computational science depends, as any field of research, on a steady influx of new people into the field and on a continuing effort to maintain the skills of its established practitioners. E-CAM recognizes this and will devote significant effort to training and up-skilling workshops. An important aspect of these workshops is that they will address a broad audience of both academic and industrial scientists and they will combine more traditional training with practical hands-on training. In particular, extensive use will be made of "extended software

development workshops" which will have a two-fold purpose. On the one hand, they will be a key mean of generating the software modules that ultimately constitute the E-CAM repository. On the other they will have an explicit training dimension in that participants will gain direct experience of using modern software engineering techniques and open-source development tools to generate community code.

The goals of E-CAM are pursued via a set of coordinated actions and networking. Its main tasks are as follows:

- Software development targeted at specific end-users needs, and including testing of scaling and portability;
- Development of the E-CAM repository, an open source repository of software modules and packages for simulations in academy, material and life science, engineering. Modules include up to date documentation and benchmarks;
- Training and dissemination in the field of computational science via a series of workshops in collaboration with CECAM:
  - Extended software development workshops for production of modules for the repository based on input from the community and the industrial partners;
  - Scoping workshops for discussion and problem definition with industrial end-users;
  - State-of-the-art workshops for monitoring developments, identifying new directions and transferring knowledge to industry.
- Support for academic and industrial research via a set of pilot projects supervised by scientists in the team and sustained by E-CAM funded post-doctoral fellows.

The actions above should not be seen as disjointed activities. As we have seen the extended software development workshops address both the training and the algorithmic innovation transversal themes. Equally the software repository, and especially the expertise to use it, will be the key to the success of the industrial consultancy action. And ultimately it all depends on having a good pool of human capital to draw on. Nor will E-CAM operate in a vacuum. Just as the E-CAM actions enhance and complement each other, so too E-CAM will exploit synergies and overlaps with other Centers of Excellence and HPC initiatives.

NOMAD: developing a materials encyclopedia, big-data analytics and advanced graphics tools for materials science and engineering

https://nomad-coe.eu/

The NOMAD (Novel Materials Discovery) Laboratory CoE develops Big-Data Analytics tools and a Materials Encyclopedia for materials science and engineering. This is reinforced by advanced graphics and animation tools.

Eight complementary research groups in computational materials science along with four high-performance computing (HPC) centers form the synergetic core of this CoE. Specifically the following PIs represent the *NOMAD Laboratory CoE*: Arndt Bode (Leibniz Supercomputing Centre, Garching), Jose Maria Cela (Barcelona Supercomputing Centre, Barcelona), Ciaran Clissmann (Pintail Ltd., Dublin), Alessandro De Vita (King's College, London), Claudia Draxl (Humboldt-Universität, Berlin, member of the executive team), Daan Frenkel (University of Cambridge), Stefan Heinzel (Max Planck Computing and Data Facility, Garching), Francesc Illas (University of Barcelona), Kimmo Koski (CSC – IT Center for Science, Espoo, member of the executive team), Risto Nieminen (Aalto University, Espoo, member of the executive team), Angel Rubio (Max Planck Institute for the Structure of Dynamics of Matter, Hamburg), Matthias Scheffler (Fritz Haber Institute of the Max Planck Society, Berlin, coordinator), Kristian Sommer Thygesen (Technical University of Denmark, Lyngby). The composition of this consortium also reflects the strong embedding of the CoE in the Psi-k, CECAM, and ETSF communities.

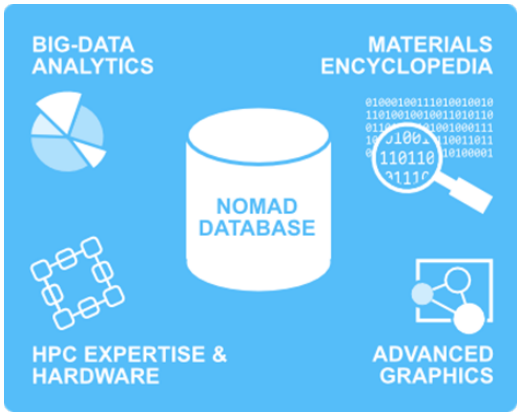

Figure 1: Sketch of the concept of the NOMAD Laboratory. This CoE establishes a code-independent data base built from calculations that employ any of the important codes used in computational materials science. The data are analyzed by novel ***big-data analytics tools****, and they are communicated to the public by an* ***extensive materials encyclopedia****. All this is reinforced by* ***advanced graphics and animation tools****, and it is enabled by the infrastructure and competence of* ***high-performance computing*** *centers.*

NOMAD's actions are motivated by the following observation. New commercial products, from smart phones to solar cells to artificial hips, are typically built from new or improved materials. Choosing the right material is difficult. Computational materials-science can help making this choice as it can predict the properties and functions of the material. Such approaches, typically based on density-functional theory, quantum chemistry, or many-body perturbation theory, can be used for known materials and for compounds that are difficult to create or handle in the experimental labs, for example because some of them may only exist under extreme conditions or are poisonous, radioactive or inherently unstable. Furthermore, the computational methodology is also powerful enough to consider materials that do not yet exist, thus developing entirely new materials which could help address fundamental issues in a number of widespread fields such as energy storage and transformation, mobility, safety, information, and health.

The *NOMAD Laboratory CoE* starts from the NoMaD Repository (http://NoMaD-Repository.eu, see also https://www.youtube.com/watch?v=L-nmRSH4NQM) which

contains input and output files of millions of high-quality calculations performed all over the word.

The Repository is designed to meet the increasing demand for storing scientific data and making them available for at least ten years. While this Repository has been and is useful for its purpose of data sharing, the data created by the many different electronic-structure codes are very heterogeneous. Thus, they are not directly useful for data analytics and extensive comparisons.

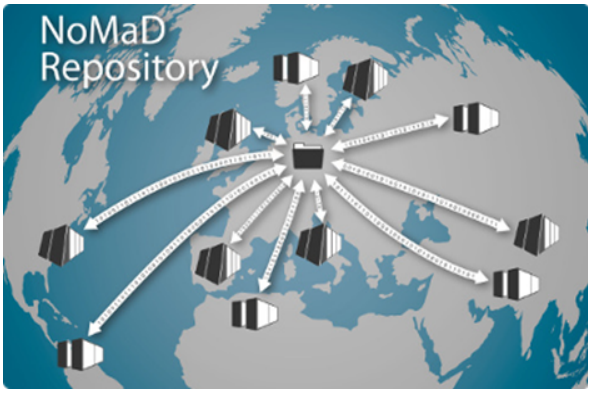

Figure 2: The NoMaD Repository is unique in the sense that it is not restricted to one or a few simulation programs but it accepts (and even requests) the full input and output files from all important codes used in computational materials science. It guarantees to store these data for at least 10 years. In spring 2016, the NoMaD Repository contained these files from more than 3 million calculations. It enables the confirmatory analysis of materials data, their reuse, and their use for new studies not intended in the original calculations.

Thus, the scientists of the *NOMAD Laboratory CoE* first build conversion layers to transform all these data into a code-independent format that forms the *NOMAD Database*. This transformation is already a significant scientific challenge as it has to address different definitions used in different fields (condensed matter physics, chemistry, biophysics, materials science), adjust energy zeros, consistently represent electron densities and wave functions, define error bars or confidence levels of methodologies, etc. (cf. Psi-k Highlight of July 2016). This goal has already been reached by developing advanced *Parsers* for 30 community codes. Part of this is an extensive hierarchical metadata scheme. Obviously, steady updates and extensions are necessary as the methodology and codes in the community evolve.

The next steps are then to “do science with the data”, and this proceeds as follows.

Big Data in materials science contain structures and correlations that may not be detectable by standard tools. *NOMAD* develops **Big-Data Analytics tools** to extract this information and use it for analyzing materials properties and functions, visualizing trends and anomalies, and possibly even predicting novel materials. Key words of the employed methodology are data mining, statistical or machine learning, in particular kernel based methods, neural networks, compressed sensing, and causal inference (see: https://analytics-toolkit.nomad-coe.eu).

Despite the practically infinite number of possible materials, we note that “the chemical compound space” is sparsely populated when the focus is on selected properties or functions. Our aim is to develop **Big-Data Analytics tools that will help to “sort” all of the available materials data to identify trends and anomalies.** The overarching topics that will be addressed in the *NOMAD Laboratory CoE* are:

- Crystal-structure prediction with the capability of quantifying the energy difference between different (metastable) structures of the same composition;
- Scanning for good thermoelectric materials;
- Finding better materials for heterogeneous catalysis, e.g. focusing on CO_2 activation and methane oxidation;
- Searching for better optoelectronic and better photovoltaic materials;
- Analyzing steels and their plasticity;
- … and more.

People who are not part of the NOMAD team, be they in industry or academia, are encouraged to collaborate on the developments and to employ the developed methods for their topics of interest. Any study can be kept fully private or shared with others.

A ***Materials Encyclopedia*** represents a user-friendly public access point to the knowledge inherent in the database. The developed search engine will comprehensively characterize materials and enable to retrieve those materials exhibiting one or more required features. By comparison of calculations based on different level of methodology, it will facilitate the assessment of, e.g., different functionals or force-fields for a given class of materials and properties.

Seeing helps understanding. We develop an infrastructure for remote visualization of the multi-dimensional NOMAD data. **Advanced graphics and a virtual-reality environment** allow for interactive data exploration, training, and dissemination.

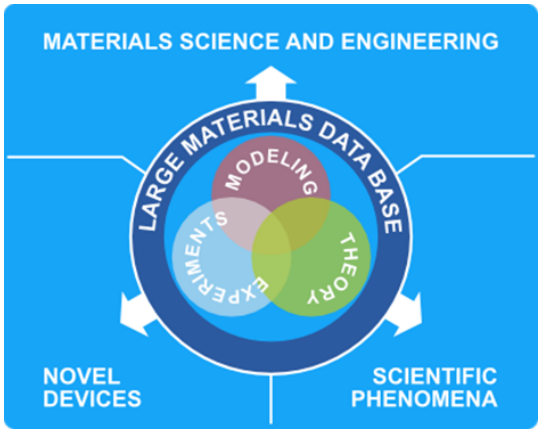

Figure 3: Presently, the NOMAD Materials Data Base considers data from theory and modeling. Later we will also add experimental data. Starting from the data base we will develop new big data analytics tool (key words are compressed sensing and other machine learning concepts) to find structure, patterns, correlations in the data that is not recognizable by standard tools. This will advance materials science and engineering, help to identify new scientific phenomena, and to design novel devices.

A sophisticated technological platform for the integrated design of the data workflow will handle the demands of *Encyclopedia*, visualization, and data analytics. This High-Performance Computing (HPC) infrastructure will support modeling of big data and provide application-enabling services. The *NOMAD Laboratory* CoE will also open up new HPC opportunities by simplifying access to existing materials data and developing new tools to search, retrieve, and manage large datasets. It will also enable new high-quality calculations for materials where important information is missing in the database.

To make sure that the *NOMAD Laboratory CoE* is valuable and relevant to end-users, its extensive network of researchers, industry representatives, students, and other stakeholders will be extended, in collaboration with Psi-k and CECAM, by organizing workshops and schools.

MaX: enabling the best use and exascale evolution of HPC and HTC for materials research

www.max-centre.eu

"MaX- Materials design at the eXascale" is a CoE established to support code developers and end-users in materials simulations, design and discovery through high-performance (HPC) and high-throughput (HTC) computing, and to prepare the exascale transition to exaFlop and exaByte architectures.

MaX coordinates a team of over 70 researchers, expert in its different domains. The consortium currently involves 5 research teams Elisa Molinari (CNR Modena coordinator); Stefano Baroni (SISSA Trieste); Nicola Marzari (EPF-Lausanne); Stefan Blugel (FZ Julich); and Pablo Ordejon (ICN2 Barcelona), 5 HPC Centers: Carlo Cavazzoni (CINECA Bologna); Thomas Schulthess (CSCS Lugano); Dirk Pleitner (FZ Julich); Erwin Laure (KTH Stockholm); and Jose Cela (BSC Barcelona), and 3 education and business partners Ivan Girotto (ICTP Trieste); Piero Altoè (E4 Computer Engineering); and Carlo Daffara (Cloudweavers Ltd).

MaX is driven by the fact that materials are crucial to scientific and technological advance, industrial competitiveness, and to tackle key societal challenges - from energy and environment, to health care, food and water, information and communication technologies, manufacturing, safety, and transportation.

The forthcoming 'exascale transition' will propel the current revolution in simulation science to the world of **quantum simulations**, making them fully viable for complex, realistic problems. Quantum simulations of materials, supported by the ever-improving outcomes of fundamental science, will become pervasive also in industrial research to sustain discovery and innovation, following on the footsteps of e.g. computational fluidodynamics or drug design.

Satisfying the rapid evolution of the software applications needed and the huge efforts required to optimize them on complex architectures **is beyond the capabilities of individual research groups or even of individual national projects.** The same is true for the effort to strengthen and expand the **communities involved in the use and development of quantum simulations of materials** through training initiatives and services.

MaX is the coordinated initiative at European level to support such **joint effort in high-performance (HPC) and high-throughput (HTC) computing of materials.**

MaX has a two-fold focus:

1. **enabling the best use and evolution of HPC technologies by creating an ecosystem**

of interoperable codes, data, workflows, and analytic tools that deliver user-oriented capabilities and services;

2. **enabling the exascale transition in the materials domain through HPC and HTC,** by developing advanced programming models, novel algorithms, domain-specific libraries, in-memory data management, software/hardware co-design actions, supported by a materials' informatics infrastructure for the automation, preservation, reproducibility, and dissemination of data and calculations.

Initially, MaX focuses on widely used **flagship codes that are open source** to work as test beds for its developments. Quantum-ESPRESSO, Siesta, Yambo, and Fleur are the quantum engines chosen to represent the variety of first-principles codes, iPI is its atomic engine, and AiiDA acts as the informatics infrastructure for the entire effort – this latter in close collaboration with MARVEL, the Swiss National Centre for Computational Design and Discovery of Novel Materials, coordinated by EPFL.

MaX is expanding the potential of these flagship codes on the present HPC platforms, by implementing new capabilities and algorithms for the study of complex materials, properties and processes in realistic condition, far beyond the current realms. MaX is working to radically enhance the performance of the flagship codes in terms of scaling, robustness, and usability, making them ready for the forthcoming exascale hardware architectures. Finally, MaX is providing users with robust and reliable turnkey solutions to calculate accurately and predictively materials properties without the need of domain expertise in first-principles simulations.

In this way, MaX is designing and implementing a sustainable approach that will be relevant **beyond its current flagship codes and core domain**.

MaX targets the needs of

- **end-users** of materials applications who work **in industry or academia**, exploring materials discovery and relying on computer experiments;
- **developers,** i.e. scientists who develop new methods, algorithms, applications and tools in materials simulations;
- **software engineers and vendors** who optimise hardware and software performance and usability together with analytical tools for increasingly efficient computer-assisted materials design;
- **HPC centres, industries and hardware manufactures** that are interested in empowering the most advanced and ambitious solutions and recognize the importance of hardware-software co-design for sustainable HPC.

MaX delivers a number of **training, dissemination** and **community-oriented activities,** often in collaboration with CECAM and Psi-k.

In addition, MaX consolidates and expands the activities of its flagship-code teams, by offering a set of direct **'on-demand' services for individual users.** They are aimed at a variety of users, including entry-level users in industry and academia, and are organized

for convenience according to the following catalogue:

- Code downloading service: high quality releases of open-source flagship codes, offering to all users the latest release and documentation, and access to a repository of benchmarks;
- Basic support to all users of its codes;
- Help-desk for code utilization, input preparation and output analysis, code personalisation involving small modifications
- Advanced support addressing specific HPC technical issues and guiding in the choice of code for specific research context;
- Advanced consulting focusing on industrial needs, including the definition of the scientific case, development of workflow, dedicated code development and/or porting, full technical support;
- Dedicated training
- Coordinated access to the MARVEL integrated environment for turnkey solutions for the calculation of material properties.

In addition, MaX is currently working at a new option, named "Quantum as a Service" (QaaS), a virtual-machine based service that will allow running MaX flagship codes – through AiiDA— on federated HPC infrastructures or commercial clouds with appropriate security and hardware configurations. This is meant as an easy-to-deploy and easy-to-execute service, favouring broader access of industrial and academic users. A number of 'Pilot Cases' of industrially relevant research are currently developed in close contact with leading industrial partners, as a test ground for the MaX working approach. The resulting packages and workflows constitute the basis for a '**market-place**', where solutions developed for a given problem are made available and tailored to other end-users.

Most services in the above catalogue are already available, a few are planned and will follow in the next months.

Conclusions

MaX, E-CAM, and NOMAD have been successfully launched and are now at the beginning of their second year of life. These three CoEs have complementary activities and foci. MaX and NOMAD target primarily materials science. MaX starts from a set of flagship codes and plans to develop a platform for turnkey solutions for end-users. This CoE also has a strong focus on scaling towards the exascale. NOMAD privileges big-data analytics and a materials encyclopedia to drive progress in identifying structure in big data in order to advance materials science and engineering, possibly to identify new scientific phenomena, and to help industry to develop new devices. E-CAM starts by identifying four scientific areas that, singularly or combined, may address the needs of

end-users in a broader range of applications: materials and engineering, but also life science. It is more strongly oriented on developing specific methods and software to respond to solicitations from a divers set of current and perspective users from academia and, in particular, industry.

Beyond these differences, however, the boundaries between the areas, expertise and approach of MaX, E-CAM and NOMAD, are not sharp. Indeed, the three CoEs share overall scientific objectives and goals. They aim at placing computational science and data analytics at the centre of scientific discovery and technological exploitation. Importantly, they originate from communities with strong overlaps and a long and successful history of collaboration of research teams and institutions such as CECAM and Psi-k. Exchange and collaboration can only increase the impact of each project and facilitate long-term sustainability. Some initiatives in this direction have already been implemented, such as the participation of representatives from the three CoEs to the first E-CAM scoping workshop, the workshop "Towards a common format for computational material science data" held in Lausanne in January 2016 and the related Psi-k highlight paper published in July 2016. New collaborations, specifically in the area of training, are also being discussed.

The considerations above are particularly important to plan for the future. The present program (EINFRA-5-2015) will in fact be followed by a second, larger one supporting successful CoEs and establishing new CoE pilots in areas not currently covered. The details of the new call are not yet known, but the upcoming program may offer the possibility for MaX, E-CAM and NOMAD to join efforts, and explore new partnerships, in a broadly-based proposal addressing the challenges and opportunities offered by the emerging exascale computing resources. Such a proposal could address thematic goals in materials and life sciences coupled with the necessary transversal objectives in computational and data-analytical capabilities. This is an important opportunity to push further the scientific boundaries and to accelerate innovation, knowledge transfer and training of the next generation of modelers, and the principal investigators of the three CoEs intend to explore it in depth.

Acknowledgement

Several people have contributed to the writing of this article. In particular, we want to thank all the PIs, team leaders and scientists involved in the exciting work at the CoEs.