



Identification/Selection of E-CAM Electronic Structure Codes for Development

E-CAM Deliverable 2.2
Deliverable Type: Report
Delivered in March, 2018



E-CAM
The European Centre of Excellence for
Software, Training and Consultancy
in Simulation and Modelling



Funded by the European Union under grant agreement 676531

Project and Deliverable Information

Project Title	E-CAM: An e-infrastructure for software, training and discussion in simulation and modelling
Project Ref.	Grant Agreement 676531
Project Website	https://www.e-cam2020.eu
EC Project Officer	Juan Pelegrín
Deliverable ID	D2.2
Deliverable Nature	Report
Dissemination Level	Public
Contractual Date of Delivery	Project Month 24 (1 st October, 2017)
Actual Date of Delivery	28 th March, 2018
Description of Deliverable	Special report on the state of the art codes and methods in Quantum Monte Carlo and Density Functional Theory (DFT) and beyond DFT methods. It will contain a review of the basic features that the majority of these codes have in common with a view to modularisation. It will also contain a review of software in this area and a list of new modules to be developed by E-CAM.

Document Control Information

Document	Title:	Identification/Selection of E-CAM Electronic Structure Codes for Development
	ID:	D2.2
	Version:	As of 28 th March, 2018
	Status:	Accepted by WP leader
	Available at:	https://www.e-cam2020.eu/deliverables
Review	Review Status:	Reviewed
Authorship	Written by:	Michael Payne (University of Cambridge)
	Contributors:	Emilio Artacho (University of Cambridge), Stefano de Gironcoli (SISSA), Carlo Pierleoni (University of L'Aquila), Leon Petit (STFC)
	Reviewed by:	Rodolphe Vuilleumier (ENS Paris)
	Approved by:	Michael Payne (University of Cambridge)

Document Keywords

Keywords:	E-CAM, HPC, CECAM, Electronic Structure, ...
-----------	--

28th March, 2018

Disclaimer: This deliverable has been prepared by the responsible Work Package of the Project in accordance with the Consortium Agreement and the Grant Agreement. It solely reflects the opinion of the parties to such agreements on a collective basis in the context of the Project and to the extent foreseen in such agreements.

Copyright notices: This deliverable was co-ordinated by Michael Payne¹ (University of Cambridge) on behalf of the E-CAM consortium with contributions from Emilio Artacho (University of Cambridge), Stefano de Gironcoli (SISSA), Carlo Pierleoni (University of L'Aquila), Leon Petit (STFC). This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>.



¹mcp1@cam.ac.uk

Contents

Executive Summary	1
1 Introduction	2
2 State of the Art: Codes and Methods	3
2.1 Density functional theory	3
2.2 Beyond DFT methods	4
2.3 Quantum Monte Carlo	5
3 Scope for Modularisation	7
3.1 Common Features	7
3.2 Modularisation Targets	8
4 Modules to be developed in WP2 and other Electronic Structure related developments	9
4.1 Modules	9
4.1.1 MatrixSwitch	9
4.1.2 Automated forcefield parameterisation	9
4.1.3 Wannier90	9
4.1.4 QMCPack	9
4.2 ESL Bundle	9
4.3 Wannier90 library	10
4.4 Exascale ensemble electronic structure calculations	10
4.5 Scaling QMC towards the Exascale	10
5 Conclusion	11
References	12

Executive Summary

Electronic structure is generally taken to refer to the application of quantum mechanical based methods to condensed matter systems. It is a field in which, historically, Europe has excelled and it is still home to many of the leading codes which are used worldwide. Electronic structure calculations are computationally expensive and so it is important that they run efficiently on large scale computers. Hence, under the conventional model of code development considerable effort is expended in continuously adapting each code to new computer hardware and new computing paradigms. The Electronic Structure Library (ESL) was launched to provide shared high quality utilities to many electronic structure codes. The aim of this initiative is not only to reduce the total effort expended in adapting each individual code to new computers - by instead simply porting the utilities to new computers - but also to reduce the possibility of human error when implementing new code. The ESL is now well developed, partly as a result of E-CAM efforts to date, and future plans are primarily focussed around deployment. The ESL model has been adopted for much of the work in other Work Packages in E-CAM. Interestingly, this model has been copied elsewhere. For instance, the US based [Molecular Sciences Institute](#), which is focussed on Quantum Chemistry methods, has adopted the ESL model for the part of its own work and clearly acknowledges the intellectual input from the ESL.

The [Wannier 90](#) code currently provides post-processing capability for electronic structure codes. Previous work in E-CAM has added new modules to the Wannier90 code and the primary target for future work on Wannier90 is to make it possible to access Wannier90 during calculations rather than purely as a post-processing tool which will open up new areas of 'on-the-fly' application.

Many electronic structure methods are mature and the computational methodologies are well established and they have become the workhorse for understanding complex materials problems. The critical scientific and societal challenges are now often related to exploring complex, multi-dimensional configuration spaces - for instance to identify intermediate structures encountered during charging and discharging batteries. We will carry out proof of concept calculations to show how to exploit electronic structure calculations on exascale computing resources to address such profound societal challenges.

Another important area for electronic structure calculations is to abstract the critical information from quantum mechanical calculations to be used in other methodologies and, as an example of this approach, we will develop modules to automate the process of obtaining accurate force fields.

Finally, work will be carried out to extend both the capability of Quantum Monte Carlo (QMC) methods and the scalability of QMC codes towards the exascale.

1 Introduction

Electronic structure has become the standard terminology to designate quantum mechanical based approaches to the determination of the properties of condensed matter systems. There are many scientific and technological challenges in this field including the fact that there are a very large number of diverse physical and chemical properties that may be interest in condensed matter systems as well the fact that there are an extremely large number of such systems. It is then, perhaps, remarkable that we are now able to determine so many of these properties for so many different systems. Those familiar with quantum mechanics might argue that this is not so surprising given that it is a parameter free theory that, in principle, allows us to predict the properties of any system of atoms. In practice, however, the computational cost of quantum mechanical calculations increases extremely rapidly with the number of electrons in the system and calculations rapidly become intractable. The rise of electronic structure methods over the last few decades has been driven by: (i) the development of approximate methods that retain most of the accuracy of exact quantum mechanical methods but with significantly reduced computational cost; (ii) the development of efficient numerical techniques to implement these methods, often associated with reduced scaling of computational effort; and (iii) the continuous increase of computational power driven by Moore's Law. It is notable that Europe has played a leading role in the development of electronic structure codes and has a large number of researchers involved in developing and maintaining codes.

Although increased computational power has been an important contributor to the rise of electronic structure it is easy to identify periods during which reformulation of the methodology and use of more efficient numerical algorithms produced computational speed-ups of a factor of 100 for every factor of 10 provided by Moore's Law. In many cases these changes were enabled by changes in computational paradigms. For instance, in the early 1990s when total energy pseudopotential calculations first moved to parallel computers these machines did not offer faster computational speeds than existing vector supercomputers but they did offer much larger amounts of memory which allowed much larger calculations to be performed than on previous high-end machines.

It is clear that future increases of computational power to the Exascale will introduce new computational paradigms. However, this is both an opportunity and a risk for Europe's electronic structure codes as a large amount of time will have to be expended to move all these codes to larger machines. This issue has been a primary driver for the creation of CECAM's ESL. The aim of the ESL is to provide a tool box of utilities used by many electronic structure codes. These utilities can be ported to any new computer hardware and/or architecture thus allowing any code that uses these utilities to exploit future machines without requiring a substantial code re-write. Given the large number of electronic structure codes, many of them with a European base, this has the potential to produce a cost saving of many person-years and allow skilled software developers to focus on the more important task of developing enhanced capability within the codes. Further work on the ESL will continue in E-CAM with further development of modules that will provide a greater range of capability in the ESL and providing a 'bundle' of ESL capabilities provided in a form that is simple to use by non-experts. In addition, an E-CAM [Extended Software Development Workshop](#) will take place in early 2019 and will focus on the scalability of electronic structure applications.

An alternative mechanism for enhancing the capability of electronic structure methods without depending on skilled expertise to implement the capability in each individual code is to provide a post-processing capability that takes the output of a large number of electronic structure codes and uses this to compute a wider range of functionality than is available in any single code. This approach provides a cost saving of many person-years as any new functionality only has to be developed once in the post-processing tool. In the case of electronic structure, such a post-processing tool is already available and is provided by the [Wannier90](#) code. The ESL already contains enhancements of Wannier90 generated by E-CAM and further work on this code will be a part of E-CAM's future effort on electronic structure.

It is increasingly understood that most electronic structure methods will not scale to the exascale when run as a single monolithic calculation. However, accessing the science that researchers and industry require to address the societal challenges we currently face by exploiting the unique predictive capabilities of electronic structure calculations will need exascale compute resources. However, rather than these resources being used for a single monolithic calculation the exascale resource must instead be used for multiple instances of terascale level computations. This is the approach needed to sample multiple configurations or compositions or alchemical mutations or, indeed, combinations of all three. Work-package 2 (WP2) of E-CAM (Electronic Structure) is firmly committed to participating in and exploiting E-CAM's development of this technology under WP7 (Hardware considerations and the PRACE relationship) to exploit exascale compute resources and will carry out proof of technology applications using Ab Initio Random Structure Searching and Nested Sampling.

The one class of electronic structure methodology which has historically proven capacity for extreme scaling are QMC methods. E-CAM will develop modules to extend the capability and useability of QMC codes by providing a wider range of interfaces to electronic structure codes. The future Extended Software Development Workshop focussing on scaling of electronic structure codes will work on further enhancing the scaling of QMC methods towards the exascale.

2 State of the Art: Codes and Methods

There are many electronic structure codes and a wide variety of methods and this review cannot include every code and method though does attempt to include the most widely used methods and codes. The most significant descriptors that differentiate different electronic structure codes are: (i) the basis set used to represent the electronic wavefunctions, the most commonly used being either plane waves or atomic basis functions; (ii) the representation of the ionic potentials, the most commonly used being pseudopotentials and a smaller number of codes using the full Coulomb potential in all-electron calculations; (iii) the level of description of electronic exchange and correlation effects, by far the most common at present being density functional theory. There are a large number of codes that are based on plane wave basis sets, pseudopotentials and density functional theory and while there will be similarities between such approaches it is often the case that different codes offer a specialised capability for a number of applications which are either unique to that code or are of higher quality than in alternative codes. A list of the codes that have been interfaced to the NoMAD CoE repository [NOMAD - Major Codes](#) provides a reasonably comprehensive list of the major electronic structures codes. This listing provides information on the different methodologies used in each code and also includes links to code home pages, including most of the major electronic structure codes.

2.1 Density functional theory

Density functional theory (DFT) currently dominates electronic structure. Indeed, papers based on DFT dominate the list of the 20 most highly cited papers published in American Physics Society journal and the current rate of publication of papers based on DFT exceeds 30,000 papers per annum. These are quite remarkable statistics given that only a handful of DFT papers before the early 1980s and the level remained relatively low during the 1980s before beginning to grow rapidly from the early 1990s. This increase in the number of papers published per year has been driven by many factors. One significant factor has been the introduction of new functionals that provided more accurate predictions of properties. For instance, the initial developments of DFT was confined almost entirely to the physics community because the local density approximation (LDA), used almost universally at that time, gave wildly inaccurate predictions of reaction energies and activation energies which are probably the two most significant properties that chemists would want to obtain from any calculation. Despite the fact that geometries and, remarkably, vibrational energies (given that they depend on the second derivatives of the energy with respect to atomic positions) were predicted very accurately in the local density approximation, DFT was almost universally ignored, if not despised, in the chemistry community at that time. Generalised gradient approximations (GGAs) which were introduced in the late 1980s, provided much more accurate predictions of energies and this, combined with the much larger system sizes accessible to DFT methods than conventional quantum chemistry methods spurred a widespread adoption of DFT by the chemistry community. The introduction of GGAs which, unlike the LDA, included empirical fitting parameters opened a Pandora's Box of creating functionals. Indeed, one of the challenges facing any researcher embarking on a DFT calculation is deciding which functional is appropriate for any particular calculation.

The majority of the most highly used electronic structure codes either use DFT in conjunction with plane wave basis sets and pseudopotentials (including important generalizations such as the projector augmented wave method), or all-electron methods that calculate the eigenstates using the full potential, i.e without approximating core region of the ionic potential, an no approximation to the basis set. There has been a long ongoing argument about the accuracy of pseudopotentials compared to all-electron methods, however, this argument has now been largely settled by Cottenier's Delta Project [Delta Project](#) which reveals a remarkable consistency of results from a very large number of solid state DFT based electronic structure codes irrespective of the representation of the ionic potential or the basis set used (provided that it is large enough). The web page for the [Delta Project](#) also provides a useful listing of DFT codes and a fascinating insight into the improvement in the accuracy of codes over time. Most of these improvements have been due to improvements in the accuracy of pseudopotentials - some of these were a result of ongoing work to enhance the accuracy of pseudopotentials but others were driven by the Delta test itself as it provided a mechanism by which poor pseudopotentials, say for a particular element, could be identified and improved. The test also identified the effects of numerous bugs and numerical inaccuracies in many codes, which were subsequently corrected and, as such, the Delta project as a whole has had a significant positive benefit on the entire electronic structure community.

The initial Delta test focused on energies as a function of volume and there is a pressing need for this work to be extended to other properties. However, it was a significant conclusion that a wide variety of electronic structure codes provided a consistent description of the energy-volume behaviour of a large number of elemental crystals. This, in turn, perhaps raises the question of why there are so many different electronic structure codes, each of which requires continuous effort to maintain and port to new computer architectures. While the desire to 'own' a code and maintain control over its development does partially explain this diversity, it is certainly the case that each code tends to focus on a particular range of properties - at least to the extent that the relevant functionality tends to appear in a particular code first. However, once a new functionality has been developed in one particular code it is common for it to be

implemented in a number of other codes. Though this still requires a significant software development effort which does little to progress the entire research field.

2.2 Beyond DFT methods

After years where large scale applications were dominated by DFT, we are now seeing the broader adoption of post-DFT methodologies. These techniques offer higher accuracy than is possible with DFT and can address a wide range of problems where DFT has been shown to struggle. These advanced methods can, in some cases, reach absolute accuracy significantly higher than the apocryphal 'chemical accuracy' of 1 kcal/mol. Going beyond DFT, our description of electronic structure is no longer solely based on the electron density, instead it is increasingly expressed in terms of the electronic wave function, or Green's functions. One approach to this is to add explicit many body interactions on top of a DFT calculation - this is not a new methodology since the first level implementation of this approach can be traced back to the 1980s and the first implementations of GW techniques to compute more accurate bandgaps than standard DFT. Developments since this time allow (i) GW calculations to be performed routinely for many systems and (ii) have included higher order terms in the many body expansion. One issue of such approaches is that they only work well if the underlying DFT description of the system is 'accurate' enough - for instance that the DFT bandstructure of the material correctly gives the materials as an insulator, say, rather than a metal. As DFT tends to underestimate bandgaps then this is not always the case. This problem may be solved by incorporating exact exchange - or some fraction of exact exchange in the initial calculation either by initially performing a Hartree Fock calculation or using density functional that includes some fraction of exact exchange.

Another approach for going beyond DFT is Dynamical Mean Field Theory (DMFT). In this approach a set of electronic states are extracted from the underlying, say, DFT calculation and these are used to construct an Anderson Hamiltonian which describes these orbitals, including the interactions between electrons in them, linked to a reservoir which describes the other electrons in the system. There are good numerical methods available for solving this Anderson Hamiltonian which give the full many body properties of the system. These methods have been very successful in describing atoms which have strongly correlated 'd' or 'f' electrons. As in the case of GW and higher order many body perturbation theory approaches, there are issues about the accuracy of the underlying electronic structure calculation for the orbitals used in the DMFT calculation though in this approach it is possible, in principle, to take the output electronic density of the DMFT calculation and incorporate this in the underlying electronic structure calculation to generate a new set of orbitals, etc to use in the DMFT calculation and iterate to self-consistency. Utilities that can be used to perform DMFT calculations have been developed as part of several projects including [LISA](#) which is primarily focussed on DMFT, [TRIQS](#) which is a toolkit to among others solve a generic quantum impurity model or to run a complete LDA+DMFT calculation, and [ALPS](#) which covers a wider range of methods for many body calculations and [TRIQS](#). These utilities allow a user to perform a DMFT calculation but do require significant effort and expertise and are far from an automated methodology that could be used in, say, high throughput calculations.

One emerging method that aims to both avoid the issue about the accuracy of the description of the underlying electronic structure methods and to improve the accuracy of low order many body perturbation theory is the Quasi Self-Consistent GW method (QSGW) method. In this approach, the underlying electronic structure calculation is constructed in such a way as to minimise the difference between the true interacting system and the underlying non-interacting system. The [Questaal](#) code uses this methodology and has recently been released to the community. One final point on this approach is that, as a Green function, based methodology it links seamlessly to current DMFT methods, in contrast to DFT methods where the different mathematical formulations in terms of densities (for DFT) and Green functions (for DMFT) means that there is an underlying issue of matching the two methodologies.

The main strength of Greens function based approaches such as GW and/or [BSE](#) (using many body perturbation theory to solve the Bethe Salpeter equation) is their description of excited-state properties enabling the highly accurate modelling of spectroscopic properties. Within DFT, the time dependent (TD)DFT, based on calculating the linear response function, has been similarly succesful at describing the exited state properties of molecules and solids. These theoretical spectroscopies have been implemented in an increasing number of electronic structure codes, being constantly improved for applications requiring higher precision and larger system sizes, in order to interface with an experimental environment of ever increasing capabilities.

Although they are not usually included within the classification of Electronic Structure, in this rapid survey of methodologies it is worth pointing out the significant recent progress in Quantum Chemistry methods. There was rather little method development in Quantum Chemistry during the 1990s when the community instead focussed on the development of DFT (or if we are honest, playing catch up having poured scorn on DFT and ignored it during the 1980s). However, this situation has now changed and significant progress has been made with these approaches over the last decade to the point that they can now be applied to systems containing many tens or even hundreds of atoms. While Quantum Chemistry methods are not well suited to calculation of the properties of bulk systems, as it is known that very large clusters, of the order of thousands of atoms, are needed before bulk properties begin to emerge, such

methods are of relevance to nanoscale systems and it will be interesting to see to what extent they displace standard electronic structure methods in particular fields.

2.3 Quantum Monte Carlo

QMC methods apply stochastic sampling methods to sample the many body wavefunction and compute properties associated with this. These methodologies are, in principle, exact and thus overcome many of the limitations of other Electronic Structure approaches.

There are a wide variety of different methodologies but in the context of Electronic Structure applications to materials then variational Monte Carlo and diffusion Monte Carlo within the fixed node approximation are the most commonly used. Historically, it is interesting that QMC played an important role in the rise of DFT by providing accurate exchange-correlation energies for the uniform electron gas - particularly for intermediate electronic densities to connect between the analytic expressions for the limits of very high electron densities for which exchange dominates and very low electron densities where correlation dominates. There are a number of codes available for QMC calculations of materials. The longest established one is [CASINO](#) developed originally in Cambridge but there is also [QWalk](#) and [QMCPACK](#). It is probably fair to say that many QMC codes struggle to maintain a significant user community over the long term and it is not unusual that after a very impressive start many new codes do not manage to maintain this initial momentum. This problem might be addressed by the community focussing on a single code and all agreeing to develop and maintain a single code but, to date, there has been no move in this direction

A recent development in QMC methods, again more closely related to the quantum chemistry community, is the introduction of the Full Configuration Interaction Quantum Monte Carlo Methods ([FCIQMC](#)). In this approach, stochastic sampling is applied within the space of Slater determinants which contribute to the many electron wavefunction. Although this space is exponentially large, it has been shown that very high accuracy can be achieved, significantly higher than so called 'chemical accuracy' of 1 kcal/mol through this sampling as only a small fraction of the Slater determinants contribute significantly to the many electron wavefunction. Codes for applying [FCIQMC](#) calculations can be found on the [NECI](#) GitHub repository. Although applications have primarily focussed on atomic and molecular systems to date, there is strong interest amongst the developers of the methodology to apply the technique to extended materials system and given the progress of this new method to date and the clear momentum behind the project, it is likely that this goal will be achieved.

There is a particular interest in QMC techniques given the future exascale challenge as this is the only Electronic Structure methodology which has already shown almost perfect scaling to over half a million cores. Within the E-CAM project we intend to focus our effort on the code [QMCPACK](#) which presently has the main advantage over the competitors (see above) to be well supported and developed by a rather large (on the QMC scale) community. QMCPACK appears among the applications supported by the Exascale Computing Project financed by the Department of Energy (DOE) of the United States of America (U.S.) and, as such, has become a common project among four National Labs (Argonne, Oak Ridge, Livermore and Sandia), with about 15 people, among researcher and engineers, working at its development and maintenance. On the long term we intend to participate in this common effort by developing our own branch of the code. Within this branch we will develop new algorithms for simulation of nuclear dynamics with QMC accuracy, the Coupled Electron-Ion Monte Carlo method, and for computing electronic band gaps with QMC accuracy.

As stated above the QMC solution of the electronic problem is based on a Monte Carlo sampling of the many-body wave of the system. The most widely employed form for the latter is usually built by an antisymmetric (under particle-exchange) part, comprising a linear combination of "few" determinants of single electron orbitals, times a suitable symmetric correlation factor which comprises single-body (electron-nucleus), two-body (electron-electron) and three-body (electron-electron-nucleus) terms. The single electron orbitals entering the determinants are generally obtained from a less fundamental theory such as DFT and often are reformulated in terms of quasi-particle coordinates (back-flow).

It is therefore of paramount importance to develop interfaces between the QMC code and DFT solvers which allow to transfer the orbitals efficiently. At present this is accomplished by running the DFT solver first (presently, the Plane-Wave Self-Consistent Field ([PWSCF](#)) from the [QuantumEspresso](#) (QE) suite), storing the orbitals coefficients to disk, and read those coefficients within the QMC code for later use. This strategy is not efficient enough to be used within the Coupled Electron-Ion method, and we need to develop new modules implementing this exchange on the fly. Moreover the present interface couples QMCPACK and QE5.3.0 while in the most recent version of QE (6.2.1) the memory structure for the orbital coefficient has changed and is incompatible with the old interface. We need to develop a module, to be integrated both in QMCPACK and in QE, that allow to link [PWSCF](#) as a library to QMCPACK and perform the transfer of the orbitals and of DFT energies and forces.

This new module will be necessary for Coupled Electron-Ion Monte Carlo but will also simplify the other application of QMC to static nuclei without the need to store data on the disk with the obvious slow down of the calculation, in particular for large systems.

A second module that we intend to develop within QMCPACK is to perform Grand-Canonical QMC calculations. We are presently developing a method for computing the fundamental band gap in insulators and semiconductors with QMC accuracy which requires to run in parallel many QMC calculations with different twist boundary conditions and with different number of electrons. This can be achieved with the present version of the code but we need an automaton to speed up the workflow and the data analysis. This can be accomplished by a specific module.

3 Scope for Modularisation

3.1 Common Features

There are numerous areas of commonality between the electronic structure codes. A number of these have been identified and the most important of those have been the focus of utilities to include in the [ESL](#). Here we briefly review those utilities and their capabilities.

There are various bundles of the solvers which are needed in all density-functional theory codes:

- Kohn-Sham eigensolvers: this is the focus of the “Electronic Structure Infrastructure” ([ELSI](#)) project which connects U.S. based researchers to the wider ESL effort. ELSI will bring together several different approaches to solving or circumventing the Kohn-Sham eigenvalue problem, including a density matrix solver Orbital Minimization Method Library (**LibOMM**), a resolvent-based method (PEXSI), and a scalable diagonalization method (ELPA), and will also be open to integrating other open source solvers that are compatible and/or complementary;
- Poisson solvers: this implements several different algorithms in a single library - **Poke** - which, in turn, allows different Fast Fourier Transform (FFT) back ends to be connected;
- Atomic solvers: these provide one of the most notable examples of replicated development in different code bases with no modern independent library available. The aim is, therefore, to create a new library with a coherent list of capabilities including spherical Hartree-Fock, semi-local and hybrid Density Functional Theory (DFT), and different optional relativistic levels of theory. It will be connected to LibXC (a library that contains the most widely used density functionals and provided to the ESL by authors of [Octopus](#)). It provides access to many different exchange-correlation functionals. These capabilities are provided through the Quantum Atomic Radial Equations (**SQARE**) modules, a library of utilities which deal with functions discretized on radial meshes, wave-equations with spherical symmetry and their corresponding quantum states. The utilities are segregated into three levels: radial grids & functions, ODE solvers, and states. SQARE is linked to LibXC which is a library that contains the most widely used density functionals.

Allowing outputs to be passed between different electronic structure codes also offers a considerable saving of time and effort, a reduction of potential errors and access to a wider variety of functionality than that provided by a single code. Within the ESL, the following utilities are provided for data transfer:

Libescdf is a library containing tools for reading and writing massive data structures related to electronic structure calculations, following the standards defined in the Electronic Structure Common Data Format. It is a library created to exchange electronic-structure-related data in a platform-independent and efficient manner. It is based on the Electronic Structure Common Data Format Specifications, as well as Hierarchical Data Format (HDF5).

Libpspio is a pseudopotential I/O library for DFT calculations. It can both read and write pseudopotential data, which makes it suitable for use with pseudopotential generators and electronic structure codes. It allows any DFT code to access or produce pseudopotential information without worrying about file formats and it is also helpful in improving file format specifications.

Finally, the aim of modularisation is to allow the research scientist to develop novel functionality rapidly with their work driven by the science without unnecessary regard to low level hardware and software issues. However, this approach will only be successful if researchers have a utility that deals with this interface between the science layer and the numerical compute layer. In the ESL, part of this capability is provided by **MatrixSwitch** which is a module which acts as an intermediate interface layer between high-level routines for physics related algorithms and low-level routines dealing with matrix storage and manipulation. This allows the high-level routines to be written in a way which is physically transparent, and enables them to switch seamlessly between different software implementations of the matrix operations.

Modules LibOMM, MatrixSwitch, Libpspio, Libescdf, Poke, SQARE radial grid & function, SQARE ODE solvers and SQARE states, have recently been reported in E-CAM deliverable D2.1: Electronic structure E-CAM modules I [1], and are stored in the E-CAM [Electronic Structure Library GitLab page](#). Documentation is accessible from the Electronic Structure repository at [this location](#). The development of these libraries started at an Extended Software Development Workshop in Zaragoza organized by the ESL developers team, revolved around the broad theme of solvers.

3.2 Modularisation Targets

The level of development of the [ESL](#) means that significant steps have already been taken to providing a modularisation of many common features of electronic structure codes though further modules will be added to the ESL to extend its current capability. We note contributions to the ESL include those from the following authors, where the codes these authors are principally associated with are shown in brackets: Micael Oliveira, University of Liège, Yann Pouillon, University of the Basque Country, Matthieu Verstraete, University of Liège ([ABINIT](#)); Joseba Alberdi, University of the Basque Country, David Strubbe, University of California, Miguel Marques, Martin-Luther-Universität ([Octopus](#)); Volker Blum, Duke University, William Huhn, Duke University ([FHI-AIMS](#)); Fabiano Corsetti, QuantumWise (and other authors [ELSI](#)); Jose M. Soler, U. Autónoma Madrid, Alberto García, ICMAB-CSIC, Raúl de la Cruz, Barcelona Supercomputer Centre ([SIESTA](#)); Stefano de Gironcoli ([QuantumEspresso](#)). The next step in the progress of the ESL is to bundle a set of modules into an easy to use package for non-experts. However, the high level of activity around electronic structure means that the ESL will continue to expand as further modules are contributed to the library.

Further work on new modules for the ESL will include additions to module [MatrixSwitch](#) to provide more capabilities for sparse matrix operations.

Further modules will be added to [Wannier90](#) to deal with higher levels of automation, to provide the capability to deal with spinors and symmetry-adapted Wannier functions. These capabilities will allow the study of new topological phases of matter and to allow the use of the Wannier90 in Python based workflows. Further work will create a library version of Wannier90 which will allow it to be used 'on-the-fly' within electronic structure calculations - for instance providing access to properties during a long molecular dynamics simulation for which it would be too expensive to output the entire electronic structure data from the entire simulation for post-processing, as would currently be necessary with Wannier90.

There is strong demand for force field parameters generated from quantum mechanical calculations which can, for instance, be used with the methodologies developed in [WP1](#) "Classical MD". However, it is not straightforward to generate interatomic potentials that do not have undesirable features beyond, or even sometime between, the fitting points. Furthermore, those who would wish to use such potentials will generally not be experts in electronic structure calculations or in generating forcefields. There is a strong motivation for providing a set of modules that automate the fitting procedure, in part by choosing the optimum set of configurations to use as fitting points for any particular system, but which are also well behaved such that they do not introduce unphysical behaviour in the resulting potential. Given that the only outputs needed for the fitting process from the electronic structure calculations are energies and forces then these modules could be used with any electronic structure code. This methodology will be applied to obtain force fields for metal atoms in biological systems, for which existing biochemical forcefields are often inaccurate or lacking entirely. We note that there are encouraging developments in the use of machine learning and neural networks to generate interatomic potentials. To date these methods have not been applied to systems as complex as biological molecules, though this situation will certainly change over the next few years.

4 Modules to be developed in WP2 and other Electronic Structure related developments

The following software development targets for WP2 of E-CAM are a summary of what has been previously stated in Sections 1-3 of this deliverable. For an extended discussion of why we chose them, please consult those sections. The effort devoted to these software development tasks will derive from the [ESL](#) developers team; the participants of our extended software development workshops (so far focused on the [Wannier 90](#) code and developments for the [ESL](#), but in the future also focused on scaling [QMC](#) and electronic structure applications towards the Exascale); and E-CAM funded postdoctoral researchers working directly on industrially oriented problems or developing software with high potential interest for industry. Note that in addition to the modules listed below additional modules can also be developed based on the needs arising from state-of-the-art workshops, extended software development workshops and the industrial collaborations of [WP2](#).

4.1 Modules

4.1.1 MatrixSwitch

The Distributed Block Compressed Sparse Row ([DBC SR](#)) package for sparse matrix operations will be added to the MatrixSwitch environment, an existing E-CAM module in E-CAM (see documentation in our electronic structure repository [here](#)). It will be tested using the [SIESTA](#) code which has already implemented the MatrixSwitch environment.

4.1.2 Automated forcefield parameterisation

A number of modules have already been developed, and future work will complete a further 5 modules in the context of the [pilot project](#) developed in collaboration with our industrial partner [BiKi Technologies](#) on Quantum Mechanical Parameterisation of Metal Ions in Proteins. Further work will be performed to increase the level of the automation of the procedure by testing it on a number of test cases, initially for metal ions in biological systems but later applying the methodology in other areas of application including the parameterisation of force fields of solvents and anions.

4.1.3 Wannier90

Further modules will be developed for the [Wannier90](#) code. These will focus on automation of the generation of Maximally localised Wannier functions and interfacing the Wannier90 code with python-based workflows, such as [AiiDA](#), to perform High-Throughput calculations. New modules will be developed to deal with spinors and symmetry-adapted Wannier Functions.

4.1.4 QMCPack

Modules will be developed to interface [QMCPack](#) with the [PWSCF](#) electronic structure codes and to allow greater automation of QMCPack calculations thus opening up [QMC](#) calculations to a wider user community.

4.2 ESL Bundle

The [ESL](#) Bundle will provide a set of utilities that can be used by any existing electronic structure code and will also allow rapid development of new codes that can take advantage of any new computer hardware and/or architecture since the basic operations required will be provided by the ESL bundle.

The aim of providing a bundle is to make the ESL accessible to non-experts and so the bundle will be developed to be easy to use. The initial ESL bundle will contain the following ESL packages:

- FUTILE (Fortran utilities)
- FLOOK (Fortran interface for the Lua scripting language)
- FDF (input file parser)
- PSPIO (Pseudopotential parser)

- LIBXC (Exchange-correlation functionals)
- LIBGRIDXC (XC evaluation on a grid)
- PSOLVER (Poisson solver)
- ELSI (High level interface for eigensolvers)
- ELPA (Eigensolver)
- LIBOMM (Eigensolver)
- PEXSI (Eigensolver)

All these packages will be able to be built with a single command when using the bundle, both with and without [MPI](#), thus considerably simplifying access to the capabilities of the ESL.

4.3 Wannier90 library

A library version of [Wannier90](#) will be developed so that it can be used 'on-the-fly' within a run of any electronic structure code.

4.4 Exascale ensemble electronic structure calculations

The electronic structure contribution to this E-CAM project will be proof of concept applications of the technology on Ab Initio Random Structure Searching and Nested Sampling.

4.5 Scaling QMC towards the Exascale

An extended software development workshop will focus on further extending the scaling of [QMC](#) calculations so that they can exploit exascale systems.

5 Conclusion

The [ESL](#) has reached a level of maturity and breadth of functionality which, combined with the auto-build capability to be introduced, should see it adopted by an increasing number of codes. This will give competitive advantage to those codes by freeing up expert software developers to more productive tasks, such as adding new functionality, and hence we expect a snow ball effect of increasing uptake by the community. Given the high risk associated with any disruptive technology then it is remarkable that the [ESL](#) has developed to this point so quickly and it is also notable that the model has been copied by our international competitors. The re-engineering of [Wannier90](#) will allow its capability, which is already widely used across the electronic structure community, to be exploited through a different use model, in particular in high-throughput approaches. Given that such approaches represent a significant and growing fraction of the use of electronic structure methods then we expect this development to have a considerable impact on the size of the [Wannier90](#) user base. Finally, while electronic structure calculations have challenges in moving to the exascale, the approaches to be developed will allow non-expert users to address critical societal challenges through the combination of electronic structure calculations and exascale compute resources and we will extend the scaling of Quantum Monte Carlo methods towards the exascale.

References

Acronyms Used

WP Work-package

CECAM Centre Européen de Calcul Atomique et Moléculaire

HPC High Performance Computing

DFT Density functional theory

ESL Electronic Structure Library

QMC Quantum Monte Carlo

LDA local density approximation

GGAs Generalised gradient approximations

DMFT Dynamical Mean Field Theory

QSGW Quasi Self-Consistent GW method

BSE Bethe Salpeter equation

FCIQMC Full Configuration Interaction Quantum Monte Carlo Methods

PWSCF Plane-Wave Self-Consistent Field

DBCSR Distributed Block Compressed Sparse Row

MPI Message Passing Interface

URLs referenced

Page ii

<https://www.e-cam2020.eu> ... <https://www.e-cam2020.eu>

<https://www.e-cam2020.eu/deliverables> ... <https://www.e-cam2020.eu/deliverables>

mcp1@cam.ac.uk ... <mailto:mcp1@cam.ac.uk>

<http://creativecommons.org/licenses/by/4.0> ... <http://creativecommons.org/licenses/by/4.0>

Page 1

Molecular Sciences Institute ... <https://www.molsci.org/>

Wannier 90 ... <http://www.wannier.org/>

Page 2

Extended Software Development Workshop ... <https://www.cecama.org/workshop-1640.html>

Wannier90 ... www.wannier.org

Page 3

NOMAD - Major Codes ... <https://www.nomad-coe.eu/externals/codes>

Δ Project ... <https://molmod.ugent.be/deltacodesdft>

Δ Project ... <https://molmod.ugent.be/deltacodesdft>

Page 4

LISA ... <http://dmft.rutgers.edu/LISA/>

TRIQS ... <https://triqs.ipht.cnrs.fr/1.x/>

ALPS ... http://alps.comp-phys.org/mediawiki/index.php/Main_Page

TRIQS ... https://triqs.ipht.cnrs.fr/applications/dft_tools/

Questaal ... <https://www.questaal.org/>

Page 5

CASINO ... <https://vallico.net/casinoqmc/>

QWalk ... <http://qwalk.github.io/mainline/>

QMCPACK ... <http://qmcpack.org/>

NECI ... https://github.com/ghb24/NECI_STABLE

QMCPACK ... <http://qmcpack.org/>

QuantumEspresso ... <http://www.quantum-espresso.org/>

Page 7

ELSI... <https://wordpress.elsi-interchange.org/>
Octopus... http://octopus-code.org/wiki/Main_Page
Electronic Structure Library GitLab page... <https://gitlab.e-cam2020.eu/es1>
this location... <http://e-cam.readthedocs.io/en/latest/Electronic-Structure-Modules/modules/MatrixSwitch/readme.html>

Page 8

ABINIT... <https://www.abinit.org/>
Octopus... <http://octopus-code.org/>
FHI-AIMS... <https://aimsclub.fhi-berlin.mpg.de/index.php>
ELSI... <https://wordpress.elsi-interchange.org/>
SIESTA... <https://departments.icmab.es/leem/siesta/>
QuantumEspresso... <http://www.quantum-espresso.org/>
MatrixSwitch... <http://e-cam.readthedocs.io/en/latest/Electronic-Structure-Modules/modules/MatrixSwitch/readme.html>
Wannier90... <http://www.wannier.org/>

Page 9

Wannier 90... <http://www.wannier.org/>
here... <http://e-cam.readthedocs.io/en/latest/Electronic-Structure-Modules/modules/MatrixSwitch/readme.html>
SIESTA... <https://departments.icmab.es/leem/siesta/>
pilot project... <https://www.e-cam2020.eu/pilot-project-biki-2/>
BiKi Technologies... <http://www.bikitech.com/>
Wannier90... <http://www.wannier.org/>
AiiDA... <http://www.aiida.net/>
QMCPack... <http://qmcpack.org/>

Page 10

Wannier90... <http://www.wannier.org/>

Page 11

Wannier90... <http://www.wannier.org/>
Wannier90... <http://www.wannier.org/>

Citations

- [1] F. C. Liang Liang, Micael Oliveira and Y. Pouillon, "Electronic structure E-CAM modules I," Sep. 2017. [Online]. Available: <https://doi.org/10.5281/zenodo.1185170>