3-years engineer position:

Improvement and integration of the PDI library for High-performance code coupling

Maison de la Simulation offers a 2-year engineer position in the framework of the EoCoE-II European project to work on the PDI library. This position can be extended for a 3rd year in the framework of the AMPERE ERC project.

Contact:	Julien Bigot, CEA Researcher, julien.bigot@cea.fr, Tel: 01 69 08 01 75
Location:	Maison de la Simulation, bât. 565, CEA Saclay, 91191 Gif-sur-Yvette Cedex
Employer:	CEA (year 1 & 2), then CNRS (year 3)
Salary:	Based on CEA and CNRS salary grid
Duration:	2 or 3 years starting as early as possible from November 1st 2019

Overview

The Portable Data Interface (PDI) is a library designed and developed in-house at Maison de la Simulation for efficient in-process code coupling of high-performance simulation codes. PDI has been selected as a core library in the EoCoE-II European project to interface between the application developers and Input/Output (IO) specialists. In this framework, you will be co-responsible for the maintenance of library, including the following responsibilities:

- implementation of new features required in the project,
- integration of the library in one specific code of the project (Gysela),
- support to the other users of the library.

In addition, a third year will be dedicated to the integration of the library in the Metalwalls code.

The integration in both Gysela and Metalwalls will be done in close collaboration with the respective code development teams.

Skills

The successful candidate will master the following skills and knowledge:

- team-work and integration in an international team (English language interactions)
- software engineering and library design,
- proficiency in `C++-11+` & knowledge of Fortran,
- parallel computing (including the MPI library).

In addition, the following skills and knowledge will be considered a plus:

- programming languages: `C` & Python,
- HPC IO libraries: HDF5, NetCDF
- HPC in-situ processing libraries: ParaView, VisIt & Sensei,
- experience with library build and packaging tools (CMake, spack, etc.).

Context

Maison de la Simulation (MdlS – http://www.maisondelasimulation.fr/) is a joint laboratory located on the Saclay plateau between: the French Alternative Energies and Atomic Energy Commission (CEA), the French National Center for Scientific Research (CNRS), the University of Versailles Saint-Quentin (UVSQ) and the University of Paris-Sud. The laboratory groups activities around high performance computing (HPC): research in computer science and applied mathematics, engineering and development for high-performance simulation applications. Of specific interest at MdlS are software engineering aspects of HPC, especially regarding separation of concern.

The **Energy oriented Center of Excellence** (EoCoE – https://eocoe2.eu/) is a European project lead by MdlS at the cross-road of the energy and digital revolutions. EoCoE develops and applies cutting-edge computational methods in its mission to accelerate the transition to the production, storage and management of clean, decarbonized energy. EoCoE is anchored in the HPC community and targets research institutes, key commercial players and SMEs who develop and enable energy-relevant numerical models to be run on exascale supercomputers, demonstrating their benefits for low-carbon energy technology. EoCoE drives its efforts into 5 scientific Exascale challenges in the low-carbon sectors of energy: Meteorology, Materials, Water, Wind and Fusion.

The **AMPERE** (http://www.phenix.cnrs.fr/spip.php?article816) ERC project aims to fill the gap between theory and experiment in the field of electrochemistry by introducing a whole set of new methods for simulating electrochemical systems. Applied electrochemistry plays a key role in many technologies, such as batteries, fuel cells, supercapacitors or solar cells. It is therefore at the core of many research programs all over the world. Yet, fundamental electrochemical investigations remain scarce. From the computational point of view, there is no molecular dynamics (MD) software devoted to the simulation of electrochemical systems while other fields such as biochemistry (GROMACS) or material science (LAMMPS) have dedicated tools. The AMPERE ERC will provide a cutting-edge MD software adapted to supercomputers to the computational electrochemistry community.

The **Portable Data Interface** (PDI – https://pdi.julien-bigot.fr/) is a library designed and developed at Maison de la Simulation for process-local loose coupling in high-performance simulation codes. It offers a reference system similar to Python or `C++ shared_ptr` with locking so as to ensure coherent access by coupled modules. It provides a global namespace (the data store) to share references and implements the *Observer* pattern to enable modules to react to data availability and modifications. It implements a metadata system that can be used to specify a dynamic type for references based on the value of other data (eg. array size based on the value of a shared integer). Codes using PDI declarative API expose the buffers in which they store data and notify it of significant steps in the simulation. Third-party libraries such as HDF5, SIONlib or FTI are wrapped in PDI plugins. A YAML file is used to to interleave application code with plugins use and additional code without having to modify the original application.

The **GYrokinetic SEmi-LAgrangian** (Gysela – http://gyseladoc.gforge.inria.fr/) code models the electrostatic branch of the Ion Temperature Gradient turbulence in tokamak plasmas. It has been developed at CEA -IRFM in collaboration with many partners. Gysela is a 5D full-f and flux-driven gyrokinetic code and evolves in time the five-dimensional (3D in configuration space, 2D in velocity space) ion distribution function. From the numerical point of view, Gysela is based on a semi-lagrangian scheme and couples two main solvers: a Vlasov solver and a Poisson solver. The code has demonstrated a good scalability up to millions of threads in parallel and uses millions of compute hours every year. In order to correctly simulate plasma turbulence at the scale in tokamaks such as ITER (being built in Cadarache), Exascale is a requirement in this code.

Metalwalls is a molecular dynamics package aiming at performing molecular simulation of electrochemical devices, mostly in the framework of energy storage applications. It is written in modern Fortran and parallelized using MPI. Its main originality with respect to conventional codes is that it allows to simulate electrode at fixed potentials, which mimics the experimental conditions.

Work-plan and environment

The candidate will join the PDI team including two engineers working on the library, a Post-Doc working on extension to a parallel data-store as well as a PhD student working on smart workflows for HPC. In collaboration with the other engineers in the team, the candidate will work on the improvement of the library along multiple axis identified in the EoCoE project:

- development of new plugins to support the NetCDF IO library, Sensei in-situ post-processing interface and Melissa ensemble-run tool,
- overall improvement, maintenance and support to users of the library.

To carry this work, the candidate will have access to highly parallel computing resources including the MdlS dedicated cluster, French national Tier-1 and European Tier-0 supercomputers.

An additional goal during the first two years of the project will be to integrate PDI in the Gysela extreme-scale plasma simulation code. This code currently embeds in a monolithic `Fortran` application both data pre-processing, the simulation per-se, in-line and in-situ post-processing and a checkpoint-restart mechanism and additionally relies on Python post-processing scripts for data analysis. The main objective of this integration will be to modularize the code so as to make it possible to integrate complex post-processing in the code including Python or machine-leaning based approaches.

During the third year, focus will shift to the integration of PDI in the Metalwalls computational electrochemistry code. This code is currently used in complex workflows that involve multiple steps interconnected by accesses to the file-system. The main objective of this integration will be to support more efficient coupling between these steps thanks to the modularity offered by PDI.