The present demand relies on a research project in the IPREM institute which is described below. This project occurs at a special date for the university of Pau which has just succeeded in 2017 to the I-site competition. This success has led to the creation of the E2S project: Energy Environment Solutions\footnote{http://e2s-uppa.eu/en/index.html}. The core scientific domain of the project focuses on Environment and Energy and relies on strongly recognized laboratories supported by state-of-the-art equipment. This is a new challenge for the university of Pau which will define its future position at the European and international level. The current demand will allow me to take a full advantage of this high dynamic configuration.

The E2S project, provides a series of call for proposals in order to increase the scientific excellence and the international visibility of the laboratories. In particular, several chairs of excellence are supposed to start in the coming year in close interaction with my research activities. Simultaneously, I will apply for the current call of proposals of the young researcher program of the French National Research Agency (ANR JCJC 2018). The teaching vacation demanded will allow me to play a full part in this project and to contribute actively to the scientific program of the chairs. Finally, this will permit me to broaden my research skills and reach the require experience in order to apply for an HDR (Accreditation to Supervise Research).

The current project focus on the investigation of the chemical reactivity using molecular dynamics simulations. The long term objective is to be able to describe dynamically the whole chemical-physics process by considering both the thermodynamic and kinetic aspect of the reactivity. To reach this objective, this project focuses on the development of new methods in order to describe the potential energy surface of the system in condensed matter.

**\subsection{Context and objectives of the project}**

Molecular dynamics simulation is an essential method to gain insight into dynamical molecular processes of condensed matter at the atomic resolution and is nowadays widely applied to a huge variety of systems from biological systems to oil mixtures. The accuracy of the simulations is directly associated to the parametrization of robust and reliable force-fields associated to the potential energy surface of the system. Nonetheless, force-fields always require a trade-off between accuracy and computational cost. In consequence, the parametrization of a new force-fields needs to follow a robust procedure with well identified target properties and applications to ensure the quality of the model.

A variety of classical force-fields have been developed and devoted to biological applications: GROMOS\cite{gromos}, AMBER\cite{amber} or OPLS\cite{opls1996}, among others. The development of these force-fields followed different strategies which lead to a large discrepancy in the parameters although the potential energy functions used are similar. This illustrates the difficulty of force-field parametrization. Indeed, because of the large number of parameters and the amount of target data the solution is highly underdetermined and parameters strongly correlated. Nonetheless, biological force-fields are successful and are able to describe numerous properties of biological systems. A reason of this success is the fact that a core set of a limited number of molecules such as amino acids, common solvent or nucleotides can clearly be identified and the available force-fields reproduce accurately the properties of these molecules.

In order to address the investigation of, for example, structure-based drug-design techniques, throw the screening of large pharmaceutical database, new general force-fields were developed\cite{gaff} and automatic force-field parametrization programs or platforms\cite{atb, prodrg}, provides facilities to obtain the parameters of a custom drug or substrate. This can be achieved following two strategies. The more usual, is based on the determination of atom types from the local environment of each atom of the molecule. Thanks to robust databases associating atom types and force constants, the parameters of the

force-field are extracted for each term of the chosen potential energy function\cite{atb, prodrg}. The second strategy consists in determining the force-field parameters directly from \textit{ab initio} calculations. Recently \citet{zheng2016} proposed a new software, based on the Seminario method\cite{nilsson2003} which compute directly the required parameters from a subset of the Hessian matrix of the system.

Nevertheless, classical force-fields failed to describe the chemical reactivity. Indeed, the chemical bonds are defined at the beginning of the simulation and bond breaking are not allowed. To circumvent this issue, the ReaxFF force-field was developed by Adri van Duin, William A. Goddard, and co-workers at the California Institute of Technology\cite{reaxff2001}. ReaxFF replaces the explicit definition of bonds by a bond orders criteria, which allows continuous bond forming/breaking. It is developed to be as general as possible and has been parameterized and tested for hydrocarbon reactions, high-energy materials, or catalytic systems with transition metals. As a results, ReaxFF is an intermediate computational method between \textit{ab initio} and classical molecular dynamics.

During the last ten years, a new class of potentials was proposed by \citet{behler2007} which take full advantages of the emerging technologies about machine learning algorithms. The aim of this new potentials is to provide an efficient method in order to compute the energy and the forces acting on a system with a comparable accuracy as \textit{ab initio} calculations. In order to do that, \citet{behler2007} introduce a neural-network representation of the \textit{ab initio} potential energy surfaces. Using this approach, it is possible to compute molecular dynamics simulations of large systems (or long time scale), with high accuracy. A recent article\cite{behler2017} review a wide range of applications of this method. This new potential is intrinsically a reactive potential because it is only based on \textit{ab initio} calculations and does not need to define chemical bonds. Nevertheless, the new bottleneck of the calculations is now the construction of the training set from which the neural network \textit{"learns"} the potential energy surface and consists in a large number of \textit{ab initio} calculations.

In this project, we would like to develop new force field using the above mentioned strategies

about the automatic generation, the possibility to describe the reactivity and the implementation of modern algorithms. The aim is to be able to describe various properties of a system such as transport, thermodynamic or catalytic properties. The target system will be hydrocarbon system in condensed matter with heteroatoms (O, N, S) and transition metal (Ni, V). In order to reach this objective, three steps are considered at more or less long term.

**\subsubsection{Classical force-field}**

***\textbf{Task 1:}***

This first step focuses on the development of a global strategy in order to obtain reliable and robust classical force-fields using an automated procedure. Several aspect will be combine to set up this strategy. First, the input data to build the force-field will be extracted from \textit{ab initio} calculations and in particular the Seminario method\cite{nilsson2003} will be used to compute the needed force constants. Next, we will take advantage of the strong and recognized experience of the laboratory concerning vibrational properties and IR or Raman spectra will be the target properties for the fit of the force-field parameters. Using, home made code, the vibrational properties would be considered at the harmonic or anharmonic level.

This task will be achieve by the validation of the force-fields against experimental data. We could use the data base of IR spectra built by the Pr John Shaw (University of Alberta) and Dr Michaelian Kirk (University of Alberta) who collaborate with our laboratory on the investigation of crude oil systems.

***\textbf{Task 2:}***

The validated force-fields parameters will be store in a consolidated data-base. As a result, the force-field parameters of a canonical basis set of small molecules could be constructed and then, used to compute parameters for larger or more complex molecules. This second step could be achieved from the implementation of modern statistical algorithm based on machine learning or deep learning algorithm. Two approaches could be considered. The first one would be based on the link between the topology of the molecule and the force-field parameters and compute interpolated parameters for a new molecules. The second one would be based on the molecular fragmentation methods, in order to get the electronic structure and the Hessian matrix of the large molecules from the calculations of smaller ones\cite{he2014, collins2014}.

**\subsubsection{Reactive and quantum force-field}**

Using the force-fields obtained from Task 1 and 2, we will be able to describe the initial and final state of a chemical reaction. We need now to switch on reactive force-fields. Here the aim is to be able to describe bonds forming/breaking accurately but with cheap calculations. Neural Network Potentials were recently used to investigate organic reactions\cite{gastegger20015} but the \textit{ab initio} calculations needed to build the training set are still a strong bottleneck.

***\textbf{Task 3:}***

Following the strategies of classical force-fields, we propose here to develop quantum force-fields, based on semi-empirical methods. The force constants used as parameters in classical force-field will be replaced by the parameters of the semi-empirical method which describe the ground state of the system. This approach is a trade-off between the possibility to describe the efficiently the reactivity and the accuracy of semi-empirical methods. The strategy followed to achieve Task 1 and 2 and the acquired experience will be applied in this Task with the new kind of parameters.

***\textbf{Task 4:}***

In order to go further in the description of the quantum properties along the dynamic of the system, several improvements of the quantum force-fields developed in Task 3 could be done. In particular, in the case of open shell systems such as a metalloporphyrins or a metallic catalyst, a multi-reference or valence-bond approach, based on the semi-empirical wavefunction would be implemented. This will allow us to investigate a wide field of applications by considering excited states dynamics or electronic transfer.

**\subsection{Project organization and means implemented}**

The coordinator of the project will be Germain \textsc{Vallverdu}. In the IPREM institute, the coordinator is involved in two projects in line with the research priorities of the institute: The french research network electrochemical energy storage (RS2E) and the Complex Matrices Molecular Characterization (C2MC), the current project positioning itself, in this later one. Moreover, the present project topic is totally in line with the project under way of the university about Energy and Environment Solutions (E2S project).

The coordinator's team will be completed by several people of the IPREM instituteand will benefit from the researcher of the LMA laboratory of the UPPA university (Laboratory of mathematics and their applications). Moreover, external collaborationswith Pr John Shaw (University of Alberta, Canada) and Dr Michaelian Kirk (University of Alberta, Canada), will be helpful for the comparison with experimental data.

In the IPREM institute, Hugo \textsc{Santos-Silva} (Postdoc) expert in molecular dynamic simulations and specialist of crude oil investigations, will contribute to the new methods for the force-field parameters determination both on the development point of view and on the needed computational effort to obtain and validate the parameters. Didier \textsc{Bégué} and Isabelle \textsc{Baraille}, full professor in theoretical chemistry will contribute by their deep and recognized experience in\textit{ab initio} calculations and vibrational properties which will be used as target properties for the development of the new force-fields. Finally, Brice \textsc{Bouyssiere}, full professor in analytic chemistry and specialist of crude oil and molecular characterization will provide precious insight and valuable experimental data in order to investigate complex matrices.

The LMA laboratory, part of the IPRA institute (multidisciplinary research institute applied to petroleum engineering), will bring to the team the needed skills in statistical and probabilistic approaches in order to implement relevant machine learning algorithms.

The UPPA university provides various local and regional high performance calculations (HPC) and data storage facilities which will be used in this project. Moreover, the members of the team are accustomed to answer to the calls concerning national HPC facilities (several hundred of thousand hours by year on the four last years and one million hours last year). This expertise in HCP will be advantageously exploited to reach the objectives of the current project.

**\subsection{Impact and benefits of the project}**

The methods developed in the scope of this project will provide powerful tools in order to investigate complex systems applied to the energetic transition.

The first application of the method will be devoted to the investigation of complex matrices particularly in the case of petroleum chemistry. The aim is to establish a clear cartography of the molecule existing in crude oil and to determine the composition of such complex mixtures in order to get a better knowledge and understanding of their chemical and physical properties. In that scope, using our classical force-field we will be able to investigate complex mixture or to do a screening on a wide range of molecule. Moreover, the reactive and quantum force-field will be useful to consider the role of transition metal on a structural point of view and about the chemical reactivity of these compounds.

More generally, the force-field will be designed to investigate accurately the molecular interactions between organic systems and between organic and inorganic systems. This will open a wide field of research. This is of high interest in order to develop new catalyst, for example for the dehydrogenation of \ce{CH4} or \ce{H2S} ; or to improve the efficiency of matrices for \ce{CO2} storage.

This project is in line with the ANR challenge about the energetic transition: A sustainable, green, secure and efficient energy. Its theoretical aspect agrees with the objectives of the sub-thematic about fundamental, exploratory and breaking research.