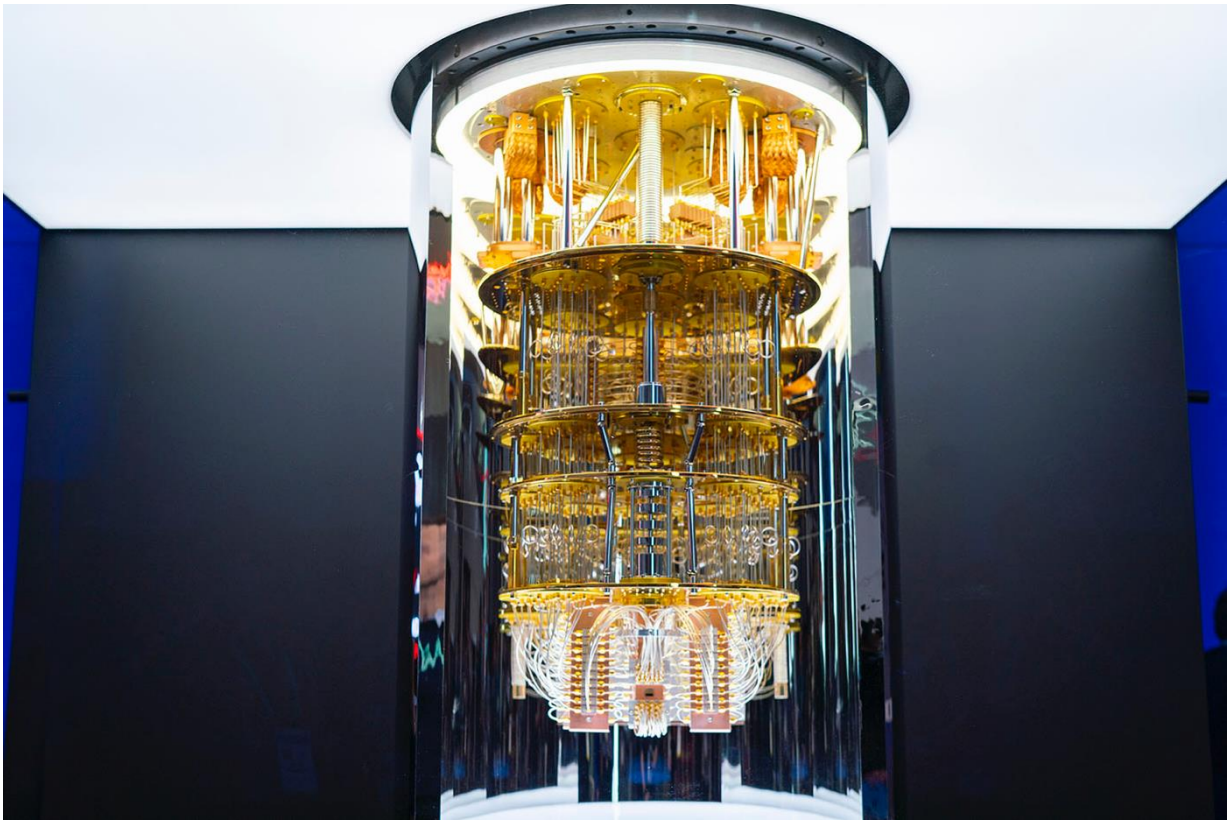
Italian National Agency for New Technologies,
Energy and Sustainable Economic Development

High Performance Computing on CRESCO infrastructure: research activities and results 2020



December 2021

High Performance Computing on CRESCO Infrastructure: research activity and results 2020

Contributions provided by a selection of users of the CRESCO infrastructure
Scientific Editors: Davide De Chiara and Simone Giusepponi, ENEA, TERIN-ICT-HPC
Cover: Amedeo Trolese, ENEA, TERIN-ICT, CR Frascati
ISBN: 978-88-8286-429-3

Contents

FOREWORD

During the year 2020, the CRESCO high performance computing clusters have provided 108 million hours of "core" computing time, at a high availability rate, to about 150 users, supporting ENEA research and development activities in many relevant scientific and technological domains. In the framework of joint programs with ENEA researchers and technologists, computational services have been provided also to academic and industrial communities.

This report, the twelfth of a series started in 2008, is a collection of 51 papers illustrating the main results obtained during 2020 using the CRESCO/ENEAGRID HPC facilities. The significant number of contributions proves the importance of the HPC facilities in ENEA for the research community. The topics cover various fields of research, such as materials science, efficient combustion, climate research, nuclear technology, plasma physics, biotechnology, aerospace, complex systems physics, geophysical flow, renewable energies, environmental issues, HPC technology. In particular, cause to the pandemic situation related to COVID-19, in 2020 part of the computational resources were devoted to work related to this considerable issue. These researchers employed about 15 % of the overall computing time and yielded 8 works. The report shows the wide spectrum of applications of high performance computing, which has become an all-round enabling technology for science and engineering.

Since 2008, the main ENEA computational resources are located near Naples, in Portici Research Centre. This is a result of the CRESCO Project (Computational Centre for Research on Complex Systems), co-funded, in the framework of the 2001-2006 PON (European Regional Development Funds Program), by the Italian Ministry of Education, University and Research (MIUR).
The CRESCO Project provided the financial resources to set up the first HPC x86_64 Linux cluster in ENEA; a major computing installation for both the Italian and the International context: it ranked 126 in the HPC Top 500 June 2008 world list, with 17.1 Tflops and 2504 cpu cores. It was later decided to keep CRESCO as the name for all the Linux clusters in the ENEAGRID infrastructure, which integrates all ENEA scientific computing systems, and is currently distributed in six Italian sites.
CRESCO computing resources were later upgraded in the framework of PON 2007-2013 with the project TEDAT and the cluster CRESCO4, 100 Tflops computing power. In 2020 the ENEAGRID computational resources consist of ~25000 computing cores and a raw data storage of about 5 PB.
In 2015 ENEA and CINECA, the main HPC institution in Italy, signed a collaboration agreement to promote joint activities and projects in HPC. In this framework, CINECA and ENEA participated successfully to a selection launched by EUROfusion, the European Consortium for the Development of Fusion Energy, for the procurement of a several PFlops HPC system, beating the competition of 7 other institutions. The new system MARCONI-FUSION started operation in July 2016 at 1 Pflops computation power level which has been increased to 5 Pflops in the summer of 2017 and thanks to a new awarded tender, the HPC services of MARCONI Fusion have been extend until 2023 with a power peak of 8 PFlops of conventional processors Intel Skylake and 2 PFlops of accelerated GPU partition of CINECA Marconi 100.
The ENEA-CINECA agreement is the key basis for ENEA HPC developments. The CRESCO6 cluster has been installed in 2018 and its own 1.4 PFlops peak computing power, ranked 420th in November 2018 Top500 list. CRESCO6 has been a challenge in HPC co-design system thanks to implement a multi-fabric network able for working Infiniband and Omni-Path on a single GPFS cluster using the same storage systems of CRESCO data centre.

The success and the quality of the results produced by CRESCO stress the role that HPC facilities can play in supporting science and technology for all ENEA activities, national and international collaborations, and the ongoing renewal of the infrastructure provides the basis for an upkeep of this role in the forthcoming years.

Dipartimento Tecnologie Energetiche e Fonti Rinnovabili,
Divisione per lo Sviluppo di Sistemi per l'Informatica e l'ICT - CRESCO Team

ENEA HPC CRESCO IN THE TIME OF COVID-19 AND NEW SUPERCOMPUTING FRONTIERS

F.Iannone* and *CRESCO team*:
D.Alderuccio, F.Ambrosino, G.Baldassarre, T.Bastianelli, R.Bertini, G.Bracco, L.Bucci, F.Buonocore, M.Caiazzo, B.Calosso, G.Cannataro, M.Caporicci, G.Carretto, M.Celino, M.Chinnici, R.Clemente, D.De Chiara, M.De Rosa, D.Di Mattia, G.Formisano, S.Ferriani, G.Ferro, A.Funel, D.Giammattei, S.Giusepponi, G.Guarnieri, M.Gusso, W.Lusani, M.Marano, A.Mariano, S.Migliori, M.Mongelli, P.Ornelli, S.Pagnutti, P.Palazzari, F.Palombi, S.Pecoraro, A.Perozziello, S.Pierattini, G.Ponti, M.Puccini, G.Santomauro, A.Scalise, F.Simoni, M.Steffè, D.Visparelli

ENEA-Energy Technologies & Renewable Sources Department - Information Communication Technologies, Lungotevere Thaon di Revel, 76, 00196 Rome Italy

ABSTRACT. High Performance Computing and Data Analysis (HPC/HPDA) infrastructures are important to face up the outbreak of SARS-COV2 in several application areas, such as: the design of new drugs, new physical barriers to reduce contagion as well as to plan strategies of containment, mitigation and suppression using predictive models for infection diseases. ENEA has made available CRESCO, its own HPC main facility with 1.4 Pflops peak computing power, for projects, requiring massive computing resources, related to mitigate the impact of COVID-19 pandemic. Within this framework, ENEA CRESCO has been carried out a huge amount of numerical simulations since spring 2020 providing to the scientific community a powerful tool for modelling the molecular dynamics (MD) of drug-receptor able to calculate actual binding free energy with accuracy as well as to carry out high-fidelity computational fluid dynamics (CFD) simulations of the transport mechanisms and related fluid dynamics of saliva droplets causing virus contagion. This paper provides the operations data of ENEA CRESCO gathered on the last year emphasising the considerable usage of the computing power in the middle of the Italian lockdown. Furthermore the paper describes the main numerical simulations carried out on ENEA CRESCO for applications relevant to counteract the pandemic as well as ENEA initiatives on the new supercomputing frontiers for the current exascale challenge and beyond for the quantum computing era.

1 Introduction

The HPC landscape is in the limelight for about two decades as hot topic both for computer scientists and end users. The level of expectations is increasing, motivated by the noticeable technical advances and what is announced at the horizon. The usage of a high fraction of the available processing power to solve real life problems is a central goal to achieve, as the gap between the theoretical performance and the sustained efficiency is more and more perceptible on modern supercomputers. From the scientific viewpoint, there are number of challenging achievements that are expected in order to come up with efficient and scalable computing solutions. Each involved topic is subject to intensive researches, with significant discoveries that are already effective. Solving large-scale problems in a short period of time using heterogeneous supercomputers is the main concern of the *exascale* challenge in the next few years, whilst the sustainability in terms of energy efficiency pave the way for new generations of computing paradigms based on the quantum mechanics. The HPC is doing round breaking work that might not be possible without them, and this has changed the rules of science and industry. With computing possibilities running up against the far edge of current technology, researchers are looking for new ways to shrink processors, combine their power, and gather enough energy to make them all work efficiently.

*Corresponding author. E-mail: francesco.iannone@enea.it

Computational capabilities are nowadays an essential part in cutting-edge scientific and engineering experiments. The capability to analyse and predict from huge amount data has incredibly improved with the use of the HPC and HPDA. Modern computational chemistry drives innovation in areas ranging from drug discovery to material design; neuroscientists can evaluate a large number of parameters in parallel to find good models for brain activity; automobile manufacturers can perform more realistic crash simulation to improve safety; astronomers can analyse different regions of the sky in parallel to search for supernovae; nuclear and particle physics are moving beyond common belief with large-scale simulations; search engines can launch parallel search across large-scale clusters of machines and instantly aggregates the results, thus reducing the latency of each request while improving relevance and accuracy; cryptography and computer systems security will benefit from the computation of gigantic prime numbers; researchers in artificial intelligence are trying to use large supercomputers to replicate (or surpass) a high-functioning human's ability to answer questions; social networking services are increasing their pervasiveness through large-scale Deep/Machine Learning processing. While keep striving to provide breath taking faster computers, designers need to contend with power and energy constraints. For decades, computers got faster by increasing their (aggregated) central processor unit. However, high processor frequency means lot of heat. This question of energy is more crucial as computing are being reported to the "Cloud", which is another innovative and affordable way to fulfil the need of high-range computing facilities. Indeed, Cloud computing offers a great alternative on mass storage, software and computing devices. Federating available computing resources, assuming a fast network, is certainly a valuable way to offer a more powerful computing system to the community. Energy, both dissipated and consumed, is also a critical concern, which is subject to active investigation from both the hardware and software standpoints.

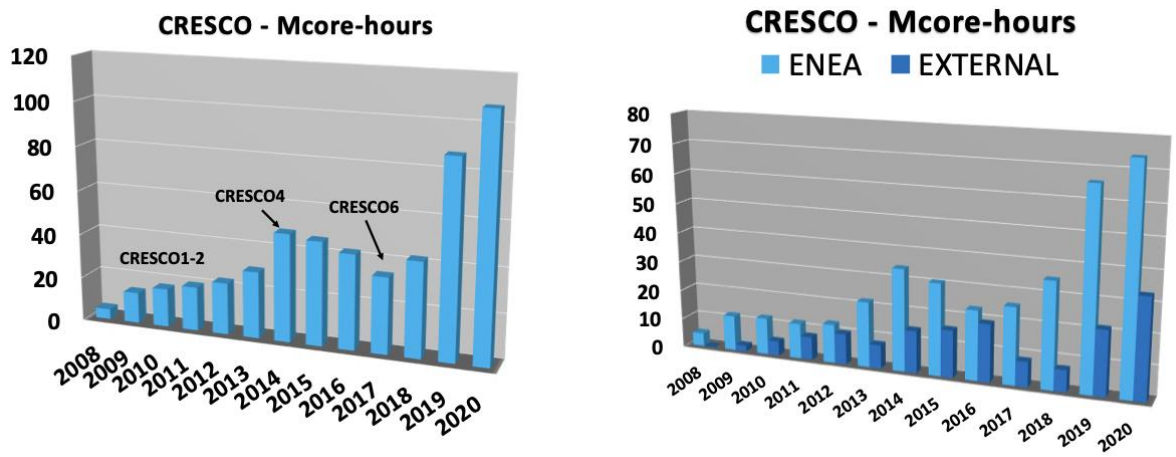
From the programming point of view, harvesting hardware advances to rich the level of cutting-edge research expectations is more challenging. Indeed, beside the ambient enthusiasm around the evolution of supercomputers, the way to peak performances is far from straightforward. In addition to algorithmic efforts to express and quantify all levels of parallelism, specific hardware and system considerations have to be taken into account when trying to provide an efficient, robust, and scalable implementation on (heterogeneous) multi-core processors. This has brought an unprecedented level of complexity in program design. Adapting a code for a given architecture or optimize it accordingly requires a complex set of program transformations, each of them addressing one or more aspects (e.g. registers, cache, instruction pipeline, data exchanges) of the target architecture. When the program is complex enough, or when the target architecture is a combination of different processing units (hybrid or accelerated computing), devising highly efficient programs becomes seriously hard. This is the price anyone should be aware of, when it comes to current and future states of HPC.

In the first part of this paper the statistic of CRESCO usage, during the 2020 pandemic year, will be shown and the second part, it faces the main topics of the new challenges of the HPC landscape and beyond.

2 CRESCO usage

The CRESCO project, started in 2008, opened the ENEA HPC era of the HPC systems based on architecture of multicore/multiprocessor nodes interconnected by a low latency and high bandwidth networks. This architecture has been held in the various stages of the HPC systems scaling up driven by the development of the CPU technology. Thanks to this development strategy, a community of users, internal ENEA and external, has grown more and more causing an exponential demand of computing time (figure 1).

The figure 1(a) shows the growth in terms of millions of core-hours over about a decade, starting in 2008 with the first HPC systems, CRESCO1-2 with a peak power of 25 Tflops, moving in 2015 to CRESCO4 with a peak power of 100Tflops, up to CRESCO6 installed at its peak power of 1.4 Pflops in 2018. The figure 1(b) shows the same growing trend of the CRESCO usage broken down for ENEA internal and external CRESCO users. The usage of CRESCO HPC systems in 2020 is ~110 Mcore-hours corresponding to ~1.4 M€ with Amazon Web Service rate of a node similar to the CRESCO ones.



(a) (b)

Fig. 1: (a) CRESCO usage since 2008; (b) CRESCO usage since 2008 broken down in ENEA internal and external users

What happened during the pandemic year is shown in the figure 2 CRESCO6 usage peak in 2020 was in the months of March and April, during the Italian lockdown whilst the usage peak in 2019 was as usual in the months of June and July.
To complete the statistics of the CRESCO6 operations in 2020, the figure 3(a) shows the availability that is 97.7 % as annual mean, whilst the figure 3(b) shows the wait main time of the jobs in the CRESCO and CRESCO6 queues.
The main numerical simulations, carried out during the year of pandemic, have used the Molecular Dynamic (MD) codes. Several of these codes belong to the open source material science community and they are undergoing the process of code-refactoring in order to run efficiently on exascale heterogeneous systems.

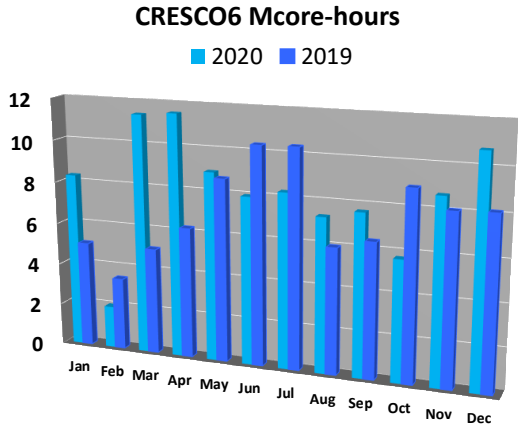
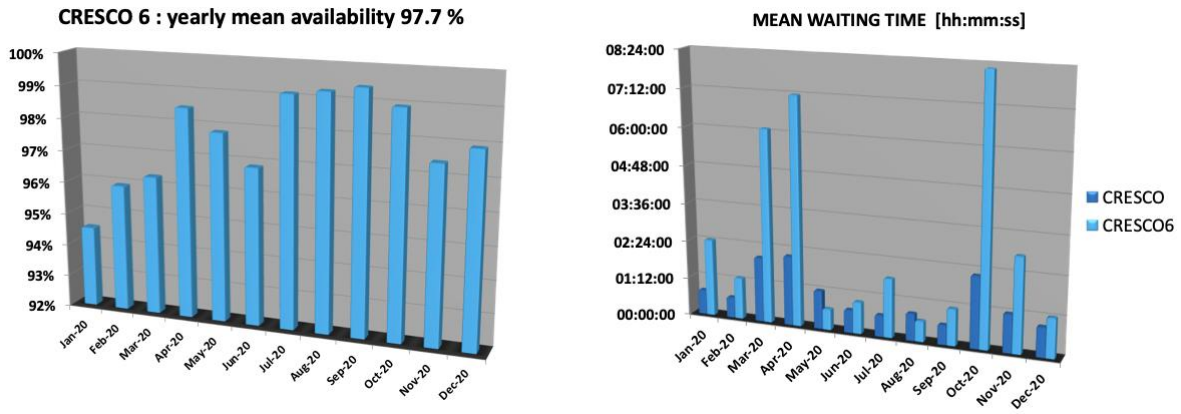


Fig. 2: CRESCO6 usage in 2020;

Among the main MD codes undergoing the code-refactoring process include the following:
Gromacs: it is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles. It is primarily designed for biochemical molecules like proteins and lipids that have a lot of complicated bonded interactions, but since GROMACS is extremely fast at calculating the nonbonded interactions (that usually dominate simulations) many groups are also using it for research on non-biological systems, e.g. polymers [1].



(a) (b)

Fig. 3: (a) CRESCO6 availability in 2020; (b) Total CRESCO clusters and CRESCO6 mean waiting time in the queues in 2020.

LAMMPS: it is a classical molecular dynamics code, and an acronym for Large-scale Atomic/Molecular Massively Parallel Simulator. LAMMPS has potentials for solid-state materials (metals, semiconductors) and soft matter (biomolecules, polymers) and coarse-grained or mesoscopic systems. It can be used to model atoms or, more generically, as a parallel particle simulator at the atomic, meso, or continuum scale. LAMMPS runs on single processors or in parallel using message-passing techniques and a spatial-decomposition of the simulation domain. The code is designed to be easy to modify or extend with new functionality [2].
CP2K: it is a quantum chemistry and solid state physics software package that can perform atomistic simulations of solid state, liquid, molecular, periodic, material, crystal, and biological systems. CP2K provides a general framework for different modelling methods such as DFT using the mixed Gaussian and plane waves approaches GPW and GAPW. Supported theory levels include DFTB, LDA, GGA, MP2, RPA, semi-empirical methods (AM1, PM3, PM6, RM1, MNDO, …), and classical force fields (AMBER, CHARMM, …). CP2K can do simulations of molecular dynamics, metadynamics, Monte Carlo, Ehrenfest dynamics, vibrational analysis, core level spectroscopy, energy minimization, and transition state optimization using NEB or dimer method. CP2K is written in Fortran 2008 and can be run efficiently in parallel using a combination of multi-threading, MPI, and CUDA. It is freely available under the GPL license. It is therefore easy to give the code a try, and to make modifications as needed [3].
Some of these codes have been used in customized implementation provided by users themselves as well as from ENEA CRESCO support. Some numerical simulations used molecular dynamics-based technique for the calculation of the absolute binding free energies in drug-receptor systems of the of the main protease ($3CL^{pro}$) of the SARS-CoV2 [4]; another one used a customized version of LAMMPS to obtain a multiscale model of the SARS-CoV2 spike protein interacting with the human ACE2 receptor. Numerical simulations were carried out on HPC CRESCO based on a thermo-fluid dynamic models of saliva droplet diffusion using a customized version of Openfoam [5].

3 The exascale challenge

MD simulations are usually used successfully on HPC systems playing a significant role in drug design. The MD model is a *N*-body classical physical problem consisting to obtain the dynamic of *N* mass particle interacting according to a given force law. The problem arises in several contexts of the classical physic from molecular scale in structural biology systems to stellar scale research in astrophysics.
A typical simulation consists of a force evaluation step, where the force law and the current configuration of the system are used to the compute the forces on each particle, and an update step, where the dynamical equations (usually Newton's laws) are numerically stepped forward in time using the computed forces. The updated configuration is then reused to calculate forces for the next time step and the cycle is repeated as many times as desired. The simplest force models are pairwise additive, that is the force of interaction between two particles is independent of all the other particles, and the individual

forces on a particle add linearly. The force calculation for such models is of complexity $O(N^2)$ as shown in the algorithm 1 of figure 4 with details in pseudo-code.
Since typical studies involve a large number of particles (10^3 to 10^6) and the desired number of integration steps is usually very large (10^6 to 10^{15}), the computational requirements often limit both the problem size as well as the simulation time and consequently, the useful information that may be obtained from such simulations. Numerous methods have been developed to deal with these issues. For molecular simulations, it is common to reduce the number of particles by treating the solvent molecules as a continuum.

```
Algorithm 1: n-body
    set initial positions and velocities
    for each time-step Δt do
        for each particle i-th do
          for each particle j-th do
              evaluate the force on j-th
            particle
          end for j-th particle
          integrate to calculate Δx for i-th
         particle
        end for i-th particle
       end for time-step
```

Fig.4: n-body algorithm.

The current petascale HPC systems are based on architectures with conventional multi-cores processors, composing a SMP compute node, sharing main memory and interconnected to other nodes by means a low latency network. Thanks to a multi-thread handling in a SMP node, hybrid parallel applications, using MPI and OpenMP (or others threading API) are still suitable on petascale HPC systems with the growing the number of cores. So far, the MD applications, based on *N*-body algorithms such as Gromacs, Lammps, have still a good speed up for simulations with 100M particles on large HPC systems running over 150k cores reaching performance of 24 ns/day. As *N*-body algorithm needs to update the whole configuration domains by means MPI all-to-all communications among nodes interconnected by low latency network, it is a biggest computational challenge for MD simulations as well as parallel applications based on numerical models strongly coupled, such us CFD, to scale-up on exascale HPC systems with tens of million cores in tens of thousands nodes consuming a huge quantity of electric energy. Hence the main challenge in the exascale transition is the containment of the power consumption to a certain value, e.g. 20 mWatt per GFlops, while improving compute performance by, say, doubling at least the current top HPC systems. The issue of the efficiency loss, major and disruptive changes in hardware architectures are taking into account. There is increasing consensus that the constraints set by power consumption can only be met by heterogeneous architectures, with specialized core processors that maximize efficiency for a specific set of instructions. This will make the overall computations more time- and energy-efficient by mapping different compute-intensive tasks to different specialized processing units, allowing performance to grow while keeping the power budget under control. The ensuing architectural complexity will set nontrivial requirements in terms of data movement, heterogeneous memory management, and fault tolerance, which will most probably require a major, possibly joint, re-design of circuits and algorithms and the adoption of different programming paradigms. In the last decade, mainly because of the continuously increasing graphics processing demands of the video game industry, Graphics Processing Units (GPUs) have evolved into massively parallel computing engines. On the other hand, FPGA solutions can also offer high throughput to numerous data-intensive applications with critical time constraints in a reconfigurable environment.
In particular the FPGA accelerator boards got 2/4 QSFP network interfaces that allow to design HPC systems based on *heterogenous* architectures with network topologies: 1D torus or 2D mesh/torus with 2/4 node degrees and diameters as $n/2$ or $n^{1/2}$.
The heterogenous architectures based on FPGA got a level of flexibility able to set dynamically mixed precision methods suitable to save energy when solving $Ax=b$ using Iterative Refinement (IR). The idea of the mixed precision method is to utilize a low precision for $O(N^3)$ LU solver, while attaining a solution accuracy through $O(N^2)$ refinement, where N is a matrix size. Further the energy efficiency can be

improved shifting thermal evaluation from the chip design phase to the run-time thermal management. In this context, accurate, fast and flexible thermal simulators help understand the power dissipation requirements, tailoring the cooling to the chip requirements to best utilize HPC infrastructures while keeping cooling costs at a minimum and enabling run-time thermal management. It can be implemented with more flexibility on programmable accelerator boards FPGA based.
Starting from European Processor Initiative activities on the stencil/tensor accelerator, boosting it using mixed-precision/trans-precision arithmetic and/or Posits and also developing an accelerator with an hardware posit processing unit for HPC computation. Alternative to floats Posits are promising to increase bandwidth and memory efficiency and hence boosting performance and saving power either for AI services (Posit8 instead of float16/32), or for scientific computation (posit16/32 instead of float64/128). In order to design HPC heterogenous system running on FPGA accelerators, the availability of a High-Level Synthesis tool and the possibility to use in the flow pre-designed computation/interfacing Intellectual Property (IP) is mandatory to lower the access barrier currently limiting the widespread adoption of FPGA devices. In this context several toolchains provide new programming models for developing on GPU/FPGA accelerators.
One of the basic guidelines in energy efficient computing is the optimization and the acceleration of algorithms and software libraries that provide a reduction of the elapsed time of HPC applications and, as consequence turn, a significant cut in energy consumption.
The new power-to-solution metrics requires a rethinking of many computational kernels of HPC applications looking for a trade-off between the reduction of the total energy and the minimization of the time-to-solution, promoting scalability also for solving ever larger problems as required by high-resolution simulations and big data applications.
Within this context, new open-source high-performance algorithms and software libraries for some, among the most widely used, kernels in numerical linear algebra and graph computation, shall be deployed. New algorithms shall be designed and optimized having as target platforms clusters of hybrid nodes, with thousands of simple computing units and a memory hierarchy that is much more exposed to the developer's control with respect to the traditional multi-level cache based systems. It is not unusual for algorithms, inefficient on traditional computing platforms, to become very much competitive on accelerators like GPU because additional computations are well tolerated and convenient when using complex memory access patterns. On the new architectures also data structures may need an in-depth revision. As for GPU, for instance, thread-locality rather than data-locality must be privileged. Single-precision floating-point arithmetic offers many advantages in terms of memory footprint and computational efficiency on GPUs (latest generation GPUs also offer a half-precision, 16 bit-based floating-point arithmetic). Although in general algorithms should be oblivious to the precision of the floating-point arithmetic, at the level of software implementation, any algorithm must be double checked under different conditions (size of the problem, range of the values that the variables can assume, presence of reduction operators), in order to guarantee that single precision and potentially half precision can be safely used, within several iterative refinement techniques for obtaining user's required accuracy.
Heterogenous architecture based on GPU/FPGA is already used to reduce the huge processing time in MD simulations. Its programming model provides a top level abstraction for low level hardware routines as well as consistent memory and execution models for dealing with massively-parallel code execution. A standard programming model is OpenCL, set from Khronos, for heterogenous parallel computing on cross-vendor and cross-platform hardware. Figure 5 shows an overview of the OpenCL architecture. One CPU-based “Host” controls multiple “Compute Devices” (for instance CPUs, GPUs & FPGAs are different compute devices). Each of these coarse grained compute devices consists of multiple “Compute Units” (akin to execution units & arithmetic processing unit groups on multi-core CPUs - think “cores”) and within these are multiple “Processing Elements”. At the lowest level, these processing elements all execute OpenCL “Kernel Functions”.

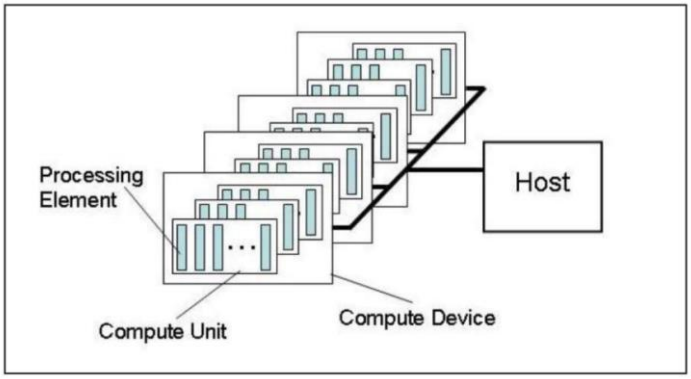


Fig. **5**: OpenCL programming model set by Khronos.

The "Kernel Functions" can be thought of as functions that transform each element of an input stream into a corresponding Processing Element of an output stream. When expressed this way, the "Kernel Function" can be applied to multiple "Processing Elements" of the input stream in parallel. Instead of blocking data to fit caches, the data is streamed into the compute units. Since streaming fetches are predetermined, data can be fetched in parallel with computation.
Since all coordinates are fixed during the force calculation, the force computation can be parallelized for the different values of *i*. In terms of streams and kernels, this can be expressed as the kernel function shown in pseudo-code of figure 6.
The kernel is applied to each element of the stream coordinates to produce an element of the forces stream. Note that the kernel can perform an indexed fetch from the coordinates stream inside the *j-loop*. An out-of-order indexed fetch can be slow, since in general, there is no way to prefetch the data. However in this case the indexed accesses are sequential. Moreover, the *j-loop* is executed simultaneously for many *i-particles*; even with minimal caching, the position of the *j-th* particle can be reused for many *N i-particles* without fetching from memory thus the performance of this algorithm would be expected to be high.

Algorithm 2: N-body kernel function

```
__kernel void kforce(global double i-th partcle)
          for each particle j-th do
              evaluate   the   force   on   j-th
            particle by i-th particle
          end for j-th particle
end __kernel kforce
```

Fig.**6**: N-body kernel function algorithm.

There are dozens of MD packages in production use, many of which have been successfully accelerated with GPUs. Scaling, however, remains problematic for the small simulations (20K - 50K particles) commonly used in critical applications, e.g., drug design, where long timescales are also extremely beneficial. FPGAs have been explored as possible MD accelerators for many years [6–12]. The first generation of complete FPGA/MD systems accelerated only the range limited (RL) force and used CPUs for the rest of the computation. While performance was sometimes competitive, high cost and lack of availability of FPGA systems meant that they were never in production use. In the last few years, however, it has been shown that FPGA clusters can have excellent for the Long Range force computation (LR) [17–20], the part of MD that is most difficult to scale.
It remains to be demonstrated, however, whether a single FPGA MD engine can be sufficiently competitive to make it worth developing such a cluster. And if so, how should it be implemented? One thing that is certain is that previous CPU-centric approaches are not viable: long timescales require ultra-short iteration times which make the cost of CPU-device data transfers prohibitive. This leads to another question: is it possible to build such an FPGA MD engine where there is little interaction with other devices? One advantage with current FPGAs is that it is now possible–for simulations of great interest (up to roughly 40K particles)–for all data to reside entirely on-chip for the entire computation. Although

this does not necessarily impact performance (double-buffering off-chip transfers still works), it simplifies the implementation and illuminates a fundamental research question: what is the best mapping among particles, cells, and force computation pipelines? Whereas the previous generation of FPGA/MD systems only dealt with a few cells and pipelines at a time, the concern now is with hundreds of each. Not only does this lead to a new version of the problem of computing pairwise forces with cutoff (see [21, 22]), it also requires orchestrating LR with the other force computations, and then all of those with motion update and particle movement.

Table 1: Forces formulas and number of float pointing operations per instructions in MD algorithms

Kernel	Formula	Flop per Interaction
Coulomb	$\frac{q_j}{r_{ij}^3}\bar{\mathbf{r}}_{ij}$	19
LJC (constant)	$\frac{q_i q_j}{r_{ij}^3}\bar{\mathbf{r}}_{ij} + \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right]$	30
LJC (linear)	$\frac{q_i q_j}{r_{ij}^4}\bar{\mathbf{r}}_{ij} + \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right]$	30
LJC (sigmoidal)	$\frac{q_i q_j}{\zeta(r_{ij})r_{ij}^3}\bar{\mathbf{r}}_{ij} + \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right]$ $\zeta(r) = e^{(\alpha r^3+\beta r^2+\gamma+\delta)}$	43

Basically the main MD algorithms are force calculation and time integration based on the *n-body* algorithm. The forces computed depend on the system being simulated and may include bonded terms, pairwise bond, angle, and dihedral; and non-bonded terms, van der Waals and Coulomb [13]. It allows to avoid $O(N^2)$ calculations treating the solvent as a continuum model. In such models, the quantum interaction of non-bonded atoms is given by a Lennard-Jones function and the electrostatic interaction is given by Coulomb's Law suitably modified to account for the solvent. The LJC(constant) kernel calculates the Coulomb force with a constant dielectric, while the LJC(linear) and LJC(sigmoidal) kernels use distance dependent dielectrics. The equations used for each kernel as well as the arithmetic complexity of the calculation are shown in Table 1.

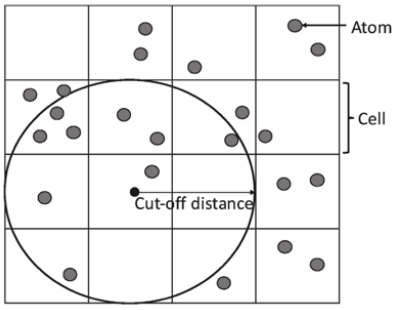


Fig. **7**: Division of the simulation box in to cells.

There are several techniques available to reduce the computation cost and to accelerate non-bonded force computation. MD simulation is done for a system that is usually represented by a box of atoms. To reduce the computation complexity, the box is divided in to multiple cells.
The figure 7 shows a 2-D representation of the cell division. A cut-off distance is set between two atoms and the neighboring cell-pairs within the cut-off distance are extracted to a cell-pair list. Non-bonded force computation is done for the atoms of the cell-pairs in the list. As a result, all atom-pair combinations for the force computation do not have to consider. Since the atoms move in the box, the

cell-pair list is updated in each iteration. A periodic boundary condition is used when an atom leaves the box. Usually the same box is replicated at the boundaries so that an atom leaves from the box reappears from the opposite direction. Using this method, a large system can be simulated by using only a small number of atoms. Even with these techniques, MD simulation takes a huge amount of processing time. Non-bonded force computation occupies most of the total processing time. Therefore, the Non-bonded force computation is accelerated using FPGA.

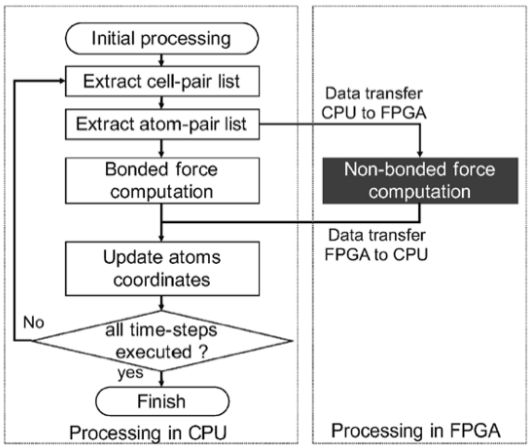


Fig.8: Flow-chart of the CPU-FPGA for MD algorithms in heterogeneous architecture.

In order to implement a parallel pipeline architecture for FPGA, the force computation is separated from the atom-pair selection. First the complete list of atom-pairs based on the cell-pair list is extracted. Then the force computation is performed for each atom-pair in the list. The atom pair-list extraction is just a searching procedure that does not contain heavy computations. On the other hand, force computation contains many multiplications and divisions. Therefore, the host computer is used for atom-pair list extraction and transfer the list to the FPGA for force computation.

Once the list is extracted, only a single loop is sufficient for the force computation of all the atom-pairs in the list. As a result, loop-pipeling can be implemented on FPGA to accelerate the computation. The figure 8 shows the flow-chart of the CPU/FPGA heterogeneous architecture. Bonded-force computation is done on CPU while non-bonded force computation is done on FPGA. After computing all the forces, the atom coordinates are updated using the Newton's equations. Then a new atom pair-list is extracted. In this method, we there are two overheads, atom-pair-list data transfer to FPGA and force data transfer from FPGA. The FPGA board is connected to the CPU through a PCI express bus. Initially, the atom-pair list and the atom coordinates are transferred from the host computer to the global memory (DRAM) of the FPGA board. After the computations are done on the FPGA board, force data are read by the host computer. This data transfer is done through the PCIe port of the host computer motherboard. The FPGA accelerator read the input data from the global memory and performs the computation. The outputs are written back to the global memory. The data read, computation and write-back is fully pipelined, so that force data are written to the global memory in every clock cycle after the pipeline is filled. Since the proposed architecture is completely designed by software, the same program code can be reused by recompiling it for any OpenCL capable FPGA board. Any future algorithm change can be also implement by just updating the software and recompiling it by using just few hours of design time.

However, the data transfers between CPU and FPGA is still a problem. This problem can be solved by future SoC based FPGA boards that contain a multicore CPU and an FPGA on the same chip. Therefore, PCI express based data transfers can be replaced by much faster on-board data transfers. To use shared memory may be also able to completely eliminate data transfers. The heterogeneous computing system can contain multiple FPGAs and they can be connected to build a scalable computing cluster.

4 The Quantum Computing era

As envisioned by the great 20th century physicist Richard Feynman, hope for simulating increasingly complex quantum systems lies in a new paradigm for information processing: quantum computation. Such computers can store and process information about simulated quantum systems natively, reducing the computational resource scaling with the size of the system to just polynomial growth, in principle. In the thirty years since Feynman's forecast, the field of quantum information science has emerged and has made tremendous strides: experimental advances in controlling quantum systems have brought quantum computers to the brink of outperforming classical computation; the synergy between quantum chemistry and quantum algorithm development has continued to equip and refine the toolbox of disruptive applications for quantum computation; and at the interface between computational chemistry and quantum information science, researchers are poised to carry Feynman's vision to fruition. Richard Feynman's original idea was to simulate quantum many-body dynamics – a notoriously hard problem for a classical computer – by using another quantum system [14].

4.1 Theory of Quantum Computing

A quantum computer, on the other hand, uses quantum bits, or qubits. A qubit is a quantum system that encodes the zero and the one into two indistinguishable quantum states. Using Dirac notation [15], a qubit can be in the state $|0\rangle$ or $|1\rangle$, or (more importantly) a superposition (linear combination) of these states. Mathematically, the state of a single qubit $|\psi\rangle$ is:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad (1)$$

such that $\alpha, \beta \in \mathbb{C}$. The coefficients also must follow a normalization condition of $|\alpha|^2 + |\beta|^2 = 1$. In the above state, the complex numbers α and β are known as *amplitudes*. The states $|0\rangle$ and $|1\rangle$ are known as basis states. Importantly, given any state $|\psi\rangle$, it is impossible to extract the amplitudes of any basis state. Commonly used is the vector notation for states. The basis states $|0\rangle$ and $|1\rangle$ are vectors that form an orthonormal basis for that qubits state space. The standard representation is:

$$|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, \qquad |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}, \qquad (2)$$

Following from this, the state $|\psi\rangle$ can be represented as a unit vector in the two-dimensional complex vector space,

$$|\psi\rangle = \begin{pmatrix}\alpha\\\beta\end{pmatrix} \qquad (3)$$

The concepts here generalize to quantum systems containing many qubits. Since a single qubit has two distinct basis states, an n qubit system has 2^n distinct basis states. In quantum computing, a multiple qubit system is known as a register. To combine the states of two individual qubits, the Kronecker/tensor product must be used. For example, to combine the states two qubits $|\psi_1\rangle$ and $|\psi_2\rangle$,

$$|\psi_1\rangle \otimes |\psi_2\rangle = \begin{pmatrix}\alpha_1\\\beta_1\end{pmatrix} \otimes \begin{pmatrix}\alpha_2\\\beta_2\end{pmatrix} = \begin{pmatrix}\alpha_1\alpha_2\\\alpha_1\beta_2\\\beta_1\alpha_2\\\beta_1\beta_2\end{pmatrix} \qquad (4)$$

When basis vectors are combined, it is convention to say $|1\rangle \otimes |0\rangle = |10\rangle$ or $|2\rangle$ (as *'10'* is *2* in binary. More generally, An *n* qubit register is described by a unit vector $|\phi\rangle$ in the 2^n dimensional complex vector space,

$$|\phi\rangle = \begin{pmatrix}\alpha_0\\\alpha_1\\\vdots\\\alpha_{2^n-1}\end{pmatrix} \qquad (5)$$

This is equivalent to a linear combination of the basis states:

$$|\psi\rangle = \sum_{j=0}^{2^n-1} \alpha_j |j\rangle \qquad (6)$$

Where $|j\rangle$ is the *jth* basis vector, and $\sum_{j=0}^{2^n-1} |\alpha_j|^2 = 1$.

There are some things note from this. Consider the vector $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. It was stated before that individual qubits can be combined using the Kronecker/ tensor product. Yet, there is no solution for the vectors $|a\rangle$ and $|b\rangle$ to the equation $|a\rangle \otimes |b\rangle = |\phi\rangle$. That is because $|\phi\rangle$ entangled, which means the state cannot be separated into individual qubit states. This is important, as it is the entanglement that makes the simulation of quantum computers hard, as it means the number of amplitudes that need to be stored grows exponentially rather then linearly.

4.2 Model of Quantum Computing

Although several theoretical models of quantum computation exist and are well studied, such as quantum circuits, topological quantum computation, dissipative quantum computing, quantum walks, and the adiabatic quantum computing model, each model has its pros and cons in the context of an actual hardware implementation.

The space of possible quantum computational models is far from fully charted, and developing models in a co-design approach with quantum hardware development may benefit both in:

– *Adiabatic Quantum Computing (AQC)*: The principles of Adiabatic QC are rooted in the so-called quantum annealing protocol, suggested for finding the global minimum of a given objective function over a given set of candidate solutions by exploiting quantum fluctuations. In quantum annealing, the system is initialized in an equal-weight superposition of all possible states and then left free to evolve according to its, usually time-dependent, Hamiltonian. Annealing is obtained introducing a slow transverse-field, slow enough for the system to stay close to the ground state of the instantaneous Hamiltonian, i.e. to evolve adiabatically. If the rate of change of the transverse field is then accelerated, the system may leave the ground state temporarily but is likely to arrive in the ground state of the final problem Hamiltonian, i.e., adiabatic evolution. The transverse field is finally switched off, and the system is expected to finally lands in the ground state of the classical model corresponding to the solution to the original optimization problem. In AQC the Hamiltonian of interest is that with a ground state describing the solution to the problem of interest. This Hamiltonian may be involved but another system with a simple Hamiltonian is prepared and initialized to the ground state. Then, the simple Hamiltonian is adiabatically evolved to the desired Hamiltonian. Since the system remains in the ground state (so the term adiabatic), at the end the state of the system describes the solution to the problem. AQC has been shown to be polynomial equivalent to conventional quantum computing in the circuit model and it is robust against dissipation since the system is always in its ground state. An experimental demonstration of the success of quantum annealing for random magnets was reported immediately after the initial theoretical proposal. Current implementations of AQC are the only commercial devices available (Dwave). They are based on Josephson junctions qubits and contain CPUs made of approximately 512 qubits in the first generation (now 2000), the number of functional qubits varying significantly from chip to chip, due to flaws in manufacturing.

  Machines performing QA at the hardware level such as the D-Wave computer minimizes the following Ising Hamiltonian function [16]:

$$\mathcal{H}(\boldsymbol{h}, \boldsymbol{J}, \boldsymbol{s}) = \sum_i h_i s_i + \sum_{i<j} J_{ij} s_i s_j \qquad (6)$$

  This is closely related to the Ising model energy function as a problem Hamiltonian, where spin variables $s_i \in \{-1, +1\}$ are subject to local fields h_i and pairwise interactions with coupling strengths J_{ij}. Each qubit's behaviour is governed by the laws of quantum mechanics, enabling qubits to be in a "superposition" state – that is, both a "-1" and a "+1" at the same time, until an outside event causes it to collapse into either a "-1" or a "+1" state. The output of an anneal is a low-energy ground state $\boldsymbol{s}$, which consists of an Ising spin for each qubit where $s_i \in \{-1, +1\}$. This is the basis upon which a quantum computer is constructed which gives the ability to quickly solve certain classes of NP-hard complex problems such as optimization, machine learning and sampling

problems. On the D-Wave device the connectivity between the binary variables s_i is described by a fixed sparse graph $G = (V,E)$ called the Chimera graph (Figure 9). Nodes in V as qubits represent problem variables with programmable weights, and edges as couplers in E have programmable connection strengths. There are weights h_i associated with each qubit s_i and strengths J_{ij} associated with each coupler between qubits (s_i and s_j). A quantum machine instruction (QMI) solves the objective function given the weights, strengths, and qubits.
Physical constraints on current D-Wave platforms such as limited precision/control error and range on weights and strengths, sparse connectivity, and number of available qubits have an impact on the problem size and performance. Embedding algorithms are required to map or fit a problem graph onto the hardware. Strictly quantum approaches are limited by the number of graph nodes that can be represented on the hardware. Larger graphs require hybrid classical-quantum approaches.

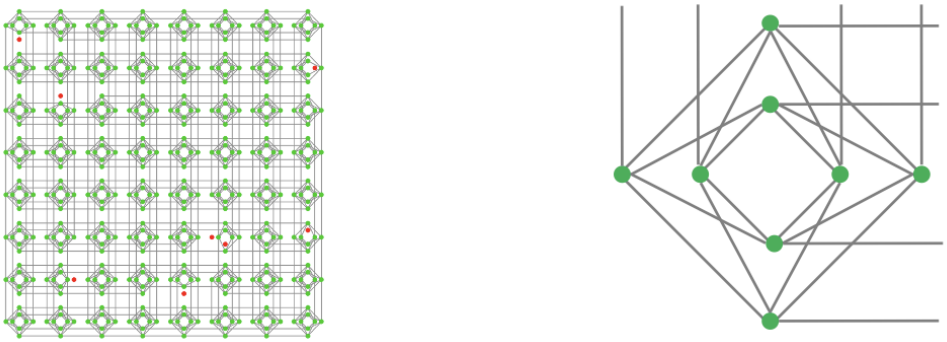


Fig. 9: Schematic representation of the "Chimera" hardware graph of the D-Wave Two X (DW2X) housed at the Information Sciences Institute (ISI) at the University of Southern California (USC). DW2X was upgrade to 1098qubits from 512 shown in the figure. Green circles represent active qubits, red circles represent inactive qubits and lines represent couplings between qubits.

– *Circuit-Based Quantum Computing (CBQC)*: to solve a particular problem, computers, be it classical or quantum, follow a precise set of instructions that can be mechanically (or quantum-mechanically) applied to yield the solution to any given instance of the problem. CBQC uses qubits, i.e. the intrinsic spin-1/2-like degree of freedom of any bi-stable quantum system, to encode information and unitary operations to process them. The interest raised by CBQC and its development had been fostered by the precise identification of criteria that should be fulfilled by any architecture suitable for a scalable quantum computer, i.e.:

  i. It should be possible to initialize an arbitrary N-qubit quantum system to a known state;
  ii. A universal set of quantum operations must be available to manipulate the initialized system and bring it to a desired correlated state;
  iii. The technology must have the ability to reliably measure the quantum system;
  iv. It must allow much longer qubit lifetimes than the time of a quantum logic gate.

The second requirement encompasses multi-qubit operations; thus, it implies that a quantum architecture must also allow for sufficient and reliable communication between physical qubits. Ordinarily, in a classical computer, the logic gates other than the NOT gate are not reversible. In CBQC quantum logic gates are reversible. However, classical computing can be performed using only reversible gates. For example, the reversible Toffoli gate can implement all Boolean functions. This gate has a direct quantum equivalent, showing that quantum circuits can perform all operations performed by classical circuits.
Just as the state of a system of qubits was defined using vectors, the way they change can be described also. The state of a qubit (or multiple qubits) is changed by quantum logic gates, or just gates. When representing the state of qubits as vectors, quantum gates are represented using matrices. These matrices must comply with certain rules in order to be valid quantum gates.
For a matrix to represent a quantum gate, it must be *unitary*. A matrix U is *unitary* if it satisfies the property that it's conjugate transpose $U^\dagger$ is also its inverse, thus $U^\dagger U = U\,U^\dagger = I$, where I is the identity matrix. In quantum computing, all gates have a corresponding *unitary* matrix, and all *unitary*

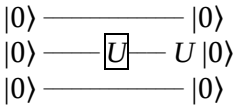
matrices have a corresponding quantum gate. Gates that act on a single qubit are represented by a *2 × 2* matrix. More generally, an *n* qubit gate is represented by a *2n × 2n* matrix. A single qubit gate can be applied to a quantum register with an arbitrary number of qubits. For a gate *U* to act on the *jth* qubit in an *n* qubit register, the full gate is formed by:

$$U = \underbrace{I \otimes I \otimes \ldots.\otimes}_{j-1\ times} U \otimes \underbrace{\ldots.\otimes I}_{n-j\ times} \qquad (7)$$

or equivalent to:

$$U_t = \bigotimes_{j=1}^{n} \begin{cases} U & j = t \\ I & otherwise \end{cases} \qquad (8)$$

In matrix form, gates are applied to registers using matrix multiplication. Multiple gates can be applied to a register. This is called a circuit. The gates being applied to a register can be detailed using a circuit diagram. In a circuit diagram, each line across represents a qubit, and each of the blocks on the lines) represents gates or other operations such as measurement. An example circuit diagram for three qubits, applying the gate U to the second qubit is shown below.



Quantum computers which are programmed using quantum circuits can be constructed out of any quantum technology that allows for defining qubit elements, and can implement single- and multi-qubit gate operations with high-fidelity. At present, architectures based on superconducting circuits (developed by IBM, figure 10), trapped-ions, semiconducting quantum-dots, photons, and neutral atoms, are actively being developed, and many are accessible to users over the internet. IBM-Q is the cloud solution of IBM to Quantum computing simulation. A IBM toolkit, *Qiskit*, [17] can compile a quantum circuit to match the entangling gate topology of a quantum device, map the circuit instructions into the native gate set of the device, and optimize the resulting quantum circuit for enhanced fidelity. The IBM toolkit, *Qiskit,* is installed on the HPC *XCRESCO*, the GPU cluster in operation.

IBM Q has successfully built and tested the most powerful universal quantum computing processors it is developing a suite of scalable, increasingly larger and better processors, with a 1,000-plus qubit device, called IBM Quantum Condor, targeted for the end of 2023. IBM Q devices are available for public use by developers, researchers and programmers via the IBM Cloud at no cost (more than hundred thousand of quantum experiments have been run by users on the IBM Cloud).

Fig. **10**: A view inside the IBM Quantum System One.

4.3 Assessing quantum computers

As quantum computers become available, it will be critical to ascertain whether, in fact, the performance of quantum computers exceeds that of conventional computers and if so, whether that gain is due to quantum effects that can be expected to scale. Developing clear metrics for assessing quantum computers now is therefore an important exercise to understand how to fairly compare resource usage between quantum and "classical" computers, and how to even compare different types of quantum computers. The main items are:

- Establish benchmarking quantum computing tests to validate and verify performance.
- Develop methods for emulating features of quantum computers with classical computers.
- Develop automatic methods for estimating resource consumption of a given quantum program (quantum algorithm), most notably in terms of quantum device implementation requirements such as number of qubits and quantum gates.

4.4 Computational Support for Quantum Algorithm Development

The quantum algorithms research lacks many of the computational tools available to conventional efforts. Future development of quantum algorithms will benefit from both exploratory mathematics research and development of computational tools for testing implementations.

It remains possible to use abstract machine models for algorithm testing purposes. The benefits derived from quantum algorithms can then be measured relative to the forecasted impact on a model HPC system. Abstract machine models can offer representations of both the quantum processor and the hybrid HPC system level. These representations can provide meaningful feedback to algorithm developers on which architectural constraints and issues must be addressed. Ultimately, how system architecture constrains algorithm implementations is likely to be a key bottleneck for quantum algorithm performance in future HPC platforms. Along side architectural issues, we expect that programming and execution models for hybrid HPC systems will also play a role in shaping quantum algorithms for applied mathematics. Programming models define the means by which end users make use of quantum algorithm implementations.

Execution models that define the order and precedence with which resources are used and the methods by which execution is negotiated must be specified. Even in an abstract setting, these execution models can provide insight into the best choices for algorithm implementation. There are currently few conventions for developing quantum algorithmic libraries but it is clear that future adoption of these libraries will need to reconcile design features with HPC system concerns.

4.5 Quantum Simulation

Quantum simulation is the emulation by a controlled quantum system of another quantum system of interest in the physical sciences. Much recent progress has been made, especially in trapped ion and trapped atom systems, in the analogue simulation of physical systems. Digital simulation then uses well-established techniques, principally Trotter formulae, to implement a simulated time evolution under a given Hamiltonian as a sequence of elementary gates. Given such a gate sequence, one can then apply quantum error correction to it, so that once the physical device has reached the error correction threshold, simulations that exceed the decoherence time of the device can be performed. There are four proposed approaches to digital quantum simulation (DQS) of physical systems: *i)* a grid to discretize space, represent the position of each particle on this grid by the binary expansion of its components, and evolve forward in time according to the Hamiltonian; *ii)* based directly on the second quantized Hamiltonian; *iii)* take the Configuration-Interaction (CI) matrix of a fermionic system as the starting point for the simulation; *iv)* different approach is needed for the simulation of the quantum dynamics of fields.

Quantum algorithms can be used more generally, for example, as subroutines that support a broad range of applications or numerical solvers. In this applied mathematics setting, quantum and classical algorithms work together, perhaps in parallel, and may exhibit non-trivial dependencies on each other. Their use is not application specific but rather driven by the varying demands of HPC end users. Quantum algorithms for applied mathematics therefore represent a very broad and revolutionary approach to algorithmic development for high performance computing. In particular, quantum algorithms achieving exponential speedup over known classical algorithms have been discovered for certain problems in linear algebra and combinatorial optimization. In addition, quantum algorithms achieving polynomial speedup have been discovered for integration and summation, extraction of

certain graph-theoretic properties, and optimization on structured objective functions. Adiabatic quantum computation and quantum annealing also show promise for optimization problems.
Quantum processors based on superconducting qubits can now perform computations in a Hilbert space of dimension $2^{53} \approx 9 \times 10^{15}$, beyond the reach of the fastest classical supercomputers available today. Quantum processors have thus reached the regime of quantum supremacy. Their computational power will continue to grow at a double-exponential rate: the classical cost of simulating a quantum circuit increases exponentially with computational volume, and hardware improvements will probably follow a quantum-processor equivalent of Moore's law, doubling this computational volume every few years.

5 Conclusions

The CRESCO HPC systems during the pandemic year have proved to be a useful tool for the users community involved in the research of solutions to fight the Covid-19, in term of drug design as well as to provide numerical simulations of virus spread by means fluid dynamics and statistic models.
Tanks to the exascale HPC class, much more computing resources is going on the availability of the community of modellers and new challenges can be achieved. The quantum computing is transitioning from a research topic to a technology that unlocks new computational capabilities and it shall be the beginning of the future generation of the computing era.

Acknowledgements

The work presented here has been carried out within the following EU projects:
H2020-EU.2.1.1.2.: Extreme scale computing and data driven technologies - Grant Agreement n. 956831;
H2020-INFR AEDI-2018-1: HPC Centres of Excellence -Grant Agreement n: 824158;
H2020-EU.1.4.1.3.: Development, deployment and operation of ICT-based e-infrastructures - Grant Agreement n. 823964.

References

[1] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen. GROMACS: fast, flexible, and free. Journal of Computational Chemistry, 26:1701–1718, 2005
[2] S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics, J Comp Phys, 117, 1-19 (1995).
[3] Thomas D. Kühne et. Al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. J. Chem. Phys. 152, 194103 (2020).
[4] Macchiagodena M, Pagliai M, Karrenbrock M, Guarnieri G, Iannone F, Procacci P. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. J Chem Theory Comput. 2020 Nov 10;16(11):7160-7172.
[5] https://www.openfoam.com
[6] N. Azizi, I. Kuon, A. Egier, A. Darabiha, and P. Chow. Reconfigurable molecular dynamics simulator. In Proceedings of IEEE Symposium on Field Programmable Custom Computing Machines (FCCM), pages 197–206, 2004.
[7] T. Hamada and N. Nakasato. Massively parallel processors generator for reconfigurable system. Proceedings of IEEE Symposium on Field Programmable Custom Computing Machines (FCCM), pages 329–330, 2005.
[8] R. Scrofano and V. Prasanna. Preliminary investigation of advanced electrostatics in molecular dynamics on reconfigurable computers. In Proceedings of ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2006.
[9] V. Kindratenko and D. Pointer. A case study in porting a production scientific supercomputing application to a reconfigurable computer. In Proceedings of IEEE Symposium on Field Programmable Custom Computing Machines (FCCM), pages 13–22, 2006.
[10] S.R. Alam, P.K. Agarwal, M.C. Smith, J.S. Vetter, and D. Caliga. Using FPGA devices to accelerate biomolecular simulations. Computer, 40(3):66–73, 2007.

[11] M. Chiu and M.C. Herbordt. Molecular dynamics simulations on high performance reconfigurable computing systems. ACM Transaction on Reconfigurable Technology and Systems (TRETS), 3(4):1–37, 2010.
[12] J. Cong, Z. Fang, H. Kianinejad, and P. Wei. Revisiting FPGA Acceleration of Molecular Dynamics Simulation with Dynamic Data Flow Behavior in High-Level Synthesis. arXiv preprint arXiv:1611.04474, 2016.
[13] J.M. Haile. Molecular Dynamics Simulation. Wiley, New York, NY, 1997.
[14] Richard P. Feynman, "Simulating physics with computers," Int. J. Theor. Phys. 21, 467–488 (1982).
[15] P. A. M. Dirac. A new notation for quantum mechanics. In Mathematical Proceedings of the Cambridge Philosophical Society, volume 35, pages 416–418. Cambridge University Press, 1939.
[16] Ushijima-Mwesigwa, H., Negre, C., Mniszewski, S. (2017). Graph Partitioning using Quantum Annealing on the D-Wave System. arXiv:1705.03082, 1–20.
[17] https://www.research.ibm.com/ibm-q/

PRACTICAL PARALLELIZATION OF A LAPLACE SOLVER WITH MPI

Marco Aldinucci[1*], Valentina Cesare[2], Iacopo Colonnelli[1], Alberto Riccardo Martinelli[1], Gianluca Mittone[1] and Barbara Cantalupo[1]

[1]*University of Torino, Computer Science Department, Corso Svizzera 185, Torino, Italy*

[2]*INAF, Osservatorio Astrofisico di Catania, Catania, Italy*

ABSTRACT. This work exposes a practical methodology for the semi-automatic parallelization of existing code. We show how a scientific sequential code can be parallelized through our approach. The obtained parallel code is only slightly different from the starting sequential one, providing an example of how little re-designing our methodology involves. The performance of the parallelized code, executed on the CRESCO6 cluster, is then exposed and discussed. We also believe in the educational value of this approach and suggest its use as a teaching device for students.
Keywords: loop parallelism, Laplace, MPI

1 Introduction

The current interest towards the parallel computing platforms is inevitably also changing the software production technologies. More and more parallel code will be needed in the future, and producing it should not be more onerous than producing traditional sequential code. Designing a parallel code from scratch is always the best solution to achieve efficiency, portability, and correctness, but this cannot always be done. Legacy code exists and is widely employed: a strategy for the practical parallelization of existing code is needed for the years to come. This work aims to develop a cost-effective methodology for the practical parallelization of existing sequential scientific code.
The proposed methodology consists of four steps:
Identify all parallelizable loops in the code, according to a depth-first search strategy;
Evaluate the potential performance gain obtainable by modifying each parallelizable loop, filtering out those that are not worth the parallelization effort;
Make each of the remaining candidate loops self-contained to remove true data dependencies among different iterations;
Use non-parallelizable loops to implement a checkpointing logic to support stop-resume behavior.
This report cannot address all the possible potentialities of this methodology, like its application with OpenMP, CUDA, and its educational implication. An extensive paper [1] provides a more comprehensive discussion with further examples.

2 Practical Parallelization of a 2-D Laplace solver with MPI

A 2D Laplace solver (Fig.1 left panel) iteratively searches for an approximate solution of the Laplace equation, i.e., a Poisson equation, with a source term ρ equal to 0, on a $N_x \times N_y$ evenly spaced cartesian grid. First of all, the value of the potential ϕ at the boundaries of the domain has to be fixed a priori (line 12), while other points are simply initialized to 0 (line 10). We extend the grid on each side by one point where the Boundary Conditions (BCs) are set (red points in Fig. 2, left panel). Then, for each point in the grid, the new value of ϕ is computed upon the values assumed by all its neighbors in the previous iteration (line 19). The algorithm iterates until ε, the sum of all the differences between the computed potentials in the last two iterations (line 20), goes below a given tolerance *tol* (line 11).

[*]Corresponding author. E-mail: marco.aldinucci@unito.it

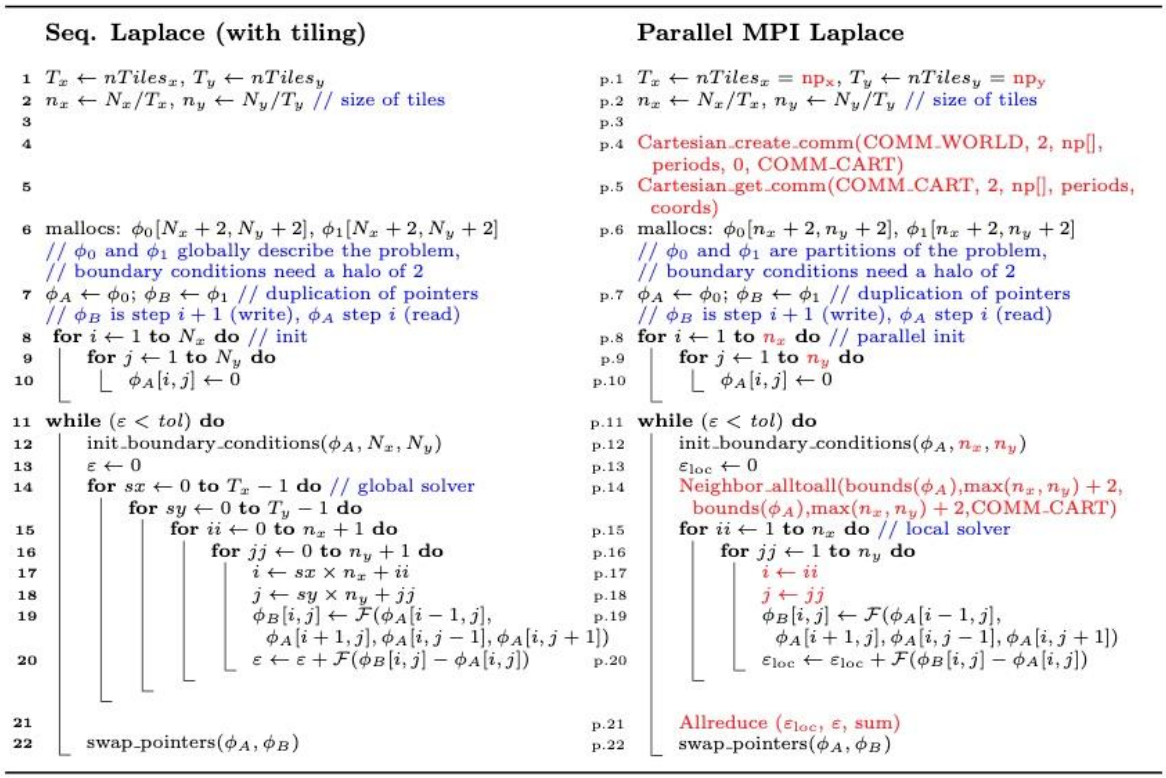


Fig.1: Sequential and Parallel MPI Laplace Algorithm

The first thing for parallelizing the Laplace solver with the proposed approach is to reorganize the serial code by dividing the global domain in $T_x \times T_y$ equally-sized sub-domains (or tiles). To compute the solution of the equation it is now necessary to firstly iterate over the tiles along each dimension (line 14) and then to iterate over the cells in each tile (lines 15-16), resulting in 4 nested for loops.

The right panel of Algorithm in Fig.1 shows in red how it is possible to parallelize with MPI the two external for loops and to combine each sub-solution with the `MPI_Allreduce` function (line 21) to derive the global solution. The decomposition of the domain is performed in parallel through the `MPI_Cart_create` function, which assigns each sub-domain to an individual process according to the topology illustrated in Fig.2. This function returns a communicator that encodes the new MPI topology. In an execution with size processes, the number of processes along each dimension of the grid, np_x and np_y, is determined through a manual *maximally squared decomposition*, where $np_x = \sqrt{size}$ and $np_y = size/np_x$. $np_x \times np_y$ is equal to size when size is a square number; if $np_x \times np_y \neq size$ then the domain cannot be decomposed and the execution does not start.
Since we compute the Laplace equation in parallel on different portions of the grid, the number of BCs increases: besides physical BCs set at the borders of the global domain, now inter-processes BCs are also present. The definition of inter-processes BCs requires communication among the different neighbor processes in the new MPI topology, since each process has to send to and to receive from its neighbors the arrays containing the potential computed in the most external interior points of the sub-grid. These arrays are then saved at the boundaries of each local domain, such that the Laplace equation can be solved on every process. This halo-swap communication is performed through the `MPI_Neighbor_alltoall` function (line 14), as shown in the right panel of Fig. 2.

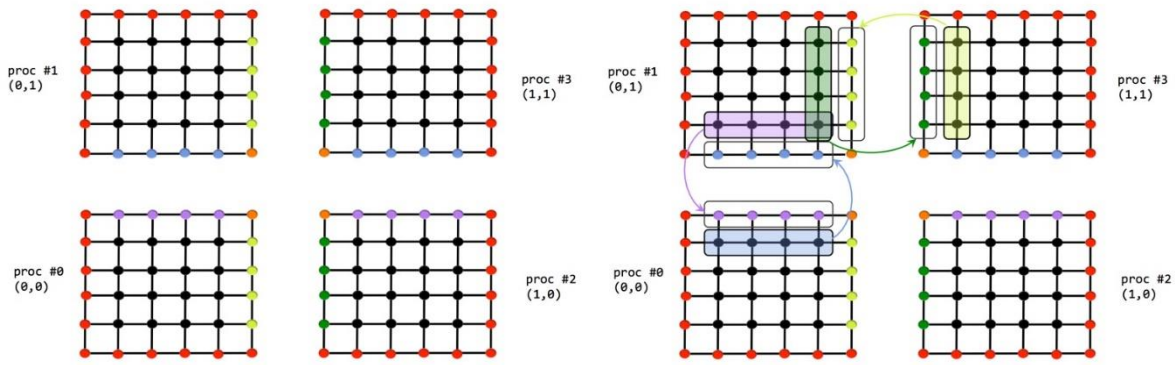


Fig.2: **Left**: Example of a Cartesian decomposition between 4 processes of the domain performed with the `MPI_Cart_create` function. Each portion of the global domain is mapped to a single process and to a couple of coordinates. Black dots show the points of the local grids where Laplace equation is computed iteratively. Red dots show the points where physical Boundary Conditions (BCs) are set whereas the dots colored in different ways show the points where inter-processes BCs are defined. **Right**: Halo-swap operation between inter-processes BCs.

3 Performance Evaluation

To demonstrate the effectiveness of the parallelization methodology we performed scaling measurement running the code on the CRESCO6 cluster.

3.1 Strong scaling

The scaling performances of the parallelized Laplace solver have been tested, using an OpenMPI version explicitly compiled for Intel processors. We evaluated the solver using a grid of size 32768×32768, which is sufficient to guarantee computational times larger than the communication ones, with a challenging tolerance of 4×10^{-7}. We decided to allocate only half of the cores available on each node (i.e. 24 cores out of 48) since a greater occupation led to performance deterioration of the node itself.
The left panel of Fig.3 shows the average speedups on ten executions: Amdahl's law holds until 16 cores but, albeit more slowly, the speedup continues growing along the entire range of considered cores, achieving a maximum value of ~778 with 2808 cores.

3.2 Weak scaling

To investigate the weak scalability of the code, we kept constant the number of grid points per core, meaning that we added 32768×32768 points to the grid for every core used in the run. This implies an increase in the global grid's resolution since its dimensions are fixed; consequently, runs with a different number of processors will converge at different speeds, thus making the results not comparable. For this reason, we did not use a fixed tolerance as the ending condition of the program, but we ran all the computations for a fixed number of iterations. As for the strong scaling, we allocate only half of the cores available on each node. The right panel of Fig.3 illustrates the weak scaling result, with the times being averaged on ten executions. The time slowly increases from ~2200 to ~2800 s, up to 624 processes, and then it remains constant around this value, showing a rather good weak scalability.

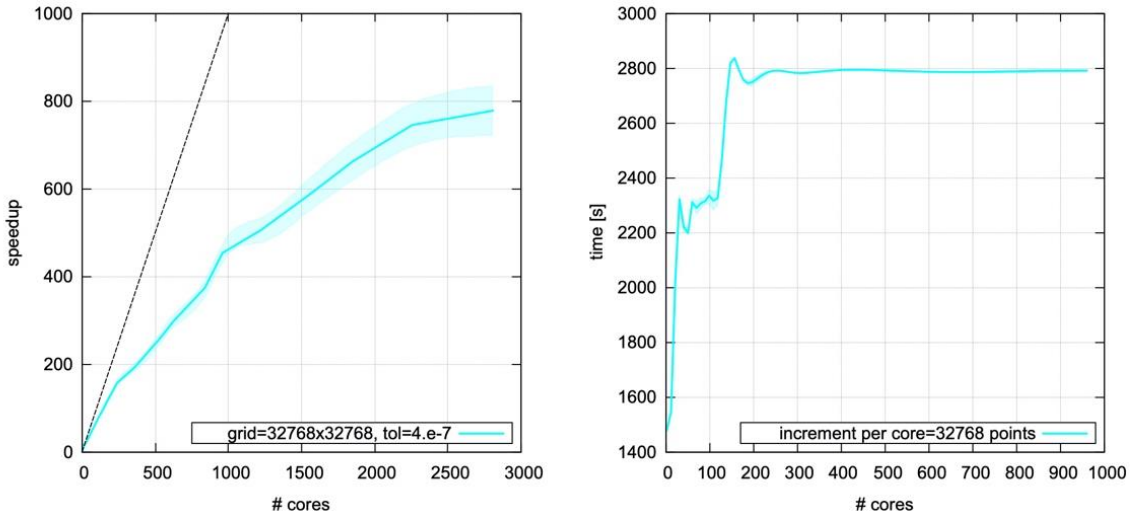


Fig.3: **Left**: strong scaling (Amdahl's law) and **Right**: weak scaling (Gustafson's law) for the Laplace solver code, averaged on ten executions on the CRESCO6 platform. The shaded areas represent the errors on the measurements and the black dashed line is the ideal Amdahl's law.

4 Conclusions

In this report, we presented a practical methodology for sequential code parallelization. Our approach differs from the many others proposed in the literature because it does not focus on single loops but whole applications. It consists of only four simple steps; it is cost-effective and straightforward to apply. Due to these characteristics, the proposed approach is particularly suitable for educational purposes: a beginner programmer can quickly start experimenting with parallel code by exploiting his/her sequential codes. The practical effects of this approach have been shown in the provided tutorial code, and its educational capabilities have been proven in years of teaching. We also executed the reference code in the CRESCO6 cluster to investigate strong and weak scaling and demonstrate the effectiveness of this methodology.

Acknowledgments

The work has been partially supported by HPC4AI [2] funded by the Region Piedmont POR-FESR 2014-20 (INFRA-P). We gratefully acknowledge the support of Francesco Iannone from ENEA and the CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [3].

References

[1] M. Aldinucci, V. Cesare, I. Colonnelli, A. R. Martinelli, G. Mittone, B. Cantalupo, C. Cavazzoni, and M. Drocco, "Practical parallelization of scientific applications with OpenMP, OpenACC and MPI," Journal of parallel and distributed computing, vol. 157, pp. 13-29, 2021. doi:10.1016/j.jpdc.2021.05.017
[2] M. Aldinucci, et al., HPC4AI, an AI-on-demand federated platform endeavour, in: ACM Computing Frontiers, Ischia, Italy, 2018. doi:10.1145/3203217.3205340.
[3] F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, F. Palombi, G. Ponti, G. Santomauro, P. Procacci, CRESCO ENEA HPC clusters: a working example of a multifabric GPFS spectrum scale layout, in: 2019 International Conference on High Performance Computing Simulation (HPCS), 2019, pp. 1051-1052.

APPLICATION DOMAINS AND CUSTOMER KNOWLEDGE IN SOCIAL MEDIA DATA CRAWLING & ANALYSIS

Daniela Alderuccio[1*] and Silvio Migliori[2]

[1]*ENEA – TERIN-ICT-HPC – Rome, Italy*

[2]*ENEA-TERIN-ICT – Rome, Italy*

ABSTRACT. The Knowledge Extraction task is influenced by application domains, which in turn influences web crawling and data analysis tasks. Each application domain hosts a specific type of knowledge. In this study we present an on-going research in the Industry 4.0 - Luxury Fashion domain, where Fashion Knowledge exists in a spaceless and accessible form as knowledge transmitted also through social media sites. In this application domain, Firm's Knowledge and Customer Knowledge intertwine in Social Media Data. ENEAGRID is equipped with a social media crawler and by using Big Data Analytics we can access these different types of knowledge, investigate firm-consumer interactions in social media and understand customer engagement. We use raw textual data from Twitter and Instagram. Results from this study suggest that: (i) Web crawling and data analytics tasks need to be scheduled and coordinated, taking into consideration the differences between application domains, also in terms of geo-locating and timing; (ii) Application domains need cross-linguistic semantic pre-analysis enabling to refine crawling strategies and data analysis; (iii) Raw data storage is at the basis of web crawling and data analytics.
Keywords: Social Media Crawling & Analysis - Knowledge Extraction - Web Crawling - Social Big Data Analytics - Industry 4.0 Luxury Fashion

1 Introduction

Social Media Data Analysis in ENEA started in 2019[1]. In this paper we present preliminary findings of a case-study in Industry 4.0 - Luxury Fashion domain. From this ongoing research we draw some general considerations with implications for all application domains, aimed at optimizing Social Media Data Crawling and refining the strategies of Analytics tasks.

Social media is an important source of up-to-date information. In order to investigate firm-consumer interactions on social media, Big Data offer an interesting insight to understand customer engagement[1] on social media and to formulate more effective brands' social media marketing efforts [4]. Over the past 10 years Luxury Brands have significantly increased their use of social media, recognizing the crucial role they play in customer engagement and luxury consumer's purchasing decisions. A Luxury 4.0[2] operating model has begun to emerge: brands use data to get closer to customers, to capture emerging preferences and streamline the process of turning ideas into new products. In the Lyst Index 2019[3] hottest brands dominated social media; the top 20 brands' names were used on Instagram. Customers consider social media to be a more trustworthy source of information than traditional instruments of marketing communications (press release or advertising).

2 Fashion (as) Knowledge

Each application domain hosts a specific knowledge. Fashion knowledge is commonly understood as an aesthetic knowledge and it is considered as a complex form of knowledge. In order to maximise its penetration of geographical space, Fashion Knowledge is purposefully 'dis-placed' or universalised,

*Corresponding author. E-mail: daniela.alderuccio@enea.it

[1] Customer engagement is the customer behavioural manifestation toward a brand or firm, beyond purchase

[2] Luxury 4.0 is inspired by Industry 4.0 (the fully digitized model for manufacturing, creating a new system that ties together self-aware machines in factories, smart logistics, customer data and design)[6]

[3] The LYST Index ranks fashion's hottest brands and products.

traversing national, cultural and social boundaries, while at the same time preserving the boundaries of its niche in aesthetic space. Fashion Knowledge is an expert knowledge and has switching centres for the transmission of ideas (Paris, Milan, Rome, New York, London). In these key locations, fashion designers work across time and space to create new ideas or fashion innovations. Fashion knowledge exists in a 'spaceless' and accessible form as knowledge transmitted through traditional mass media and social media. Fashion knowledge is institutionalized within firms[11]: designers capture and privatise Fashion Knowledge in the form of brands and trademarks, publicly identified with particular aesthetic values. Fashion Knowledge is a critical knowledge, subjected to the judgements of social peers, with issues of social inclusion or exclusion.

From customer data, Fashion Brands extract Customer Knowledge: by interpreting customer needs (and feedbacks) brands create new products and business models. On their side, Consumers are sensitive to the perception of prestige and social values of owning luxury goods (in Asian and Arab regions), of financial values (Americans), of high quality and functional values (Europeans), of demonstration of belonging to the elite class through the consumption of luxury goods[7]. Brands improve their relationship with customers, by interacting with them, restoring the authentic personal experience of luxury and inferencing customer preferences[6]. Brands have changed their strategy from one-way communication to consumer interaction with brands. Now brands accept that some aspects of their messaging will be co-created with their customer, rather than unilaterally control by their management team. Brands show a multiple presence in Social Network Sites and different account creation dates. Important luxury brands created their account in 2009 on Twitter and 2-3 years later on Instagram (Instagram started in 2010). A few brands started first on Instagram in 2012 and then created accounts in Twitter. Luxury brands have multiple presences on Twitter and Instagram (creating two or more accounts). Some fashion brands make distinct posts on Twitter and Instagram; other brands focus on the same topics across these two social media platforms. Some brands use Twitter exclusively to communicate with the customers, and Instagram for brand advertising. Brand strategies take into considerations that when users/customers post on Twitter or Instagram they are "Tweeting the Mind and Instagramming the Heart"[5]: on Instagram a post is a photo that speaks better than words. The Luxury brand has a brand identity that represents the desired image (the corporation's intended meaning) and can be expressed by images and texts in marketing communication. The brand image represents the perceived image in the consumer perception of brand communities. Official and spontaneous online brand communities have become an effective tools for luxury fashion brand, to establish stronger relationships with their consumers. Brand online communities have transformed the old one-way communication into interactive socializing and benefit from being able to communicate to each other without any restrictions of time, place or content. While in traditional marketing communications the promotion is controlled, brand marketers on social media can monitor and provide input to debates. They create word-of-mouth promotion

by implementing marketing campaigns on social media. This also allows customers to gain valuable knowledge from other members of the community and the brand itself [2]. Furthermore, various brands use both direct and indirect marketing strategies to promote their products. The economic value of Fashion Knowledge relies in its capacity to elicit emotional responses that stimulate consumption.

3 Social Media Data Crawling and Analytics in ENEA

ENEA CRESCO (Computational REsearch Centre on COmplex systems) is a High Performance Computing and Data Analysis (HPDA) infrastructure and has the capability to analyse and predict from huge amount of data. The usage of CRESCO HPC systems in 2020 was ~110 Mcorehours corresponding to ~1.4 M€ [3]. Integrated in HPC ENEAGRID/CRESCO infrastructure, ENEA created a collaborative Web Crawling Project and implemented a social media crawler, used to retrieve and analyze data from the Web. From these data we extract Knowledge, by crawling and analyzing textual data from social media (e.g. Twitter and Instagram). In this on-going research, Social Media Data are processed on a "Text Mining & Analytics" Platform (hosted in ENEAGRID environment) for the management of researches in Economic-Social domains. Since 2019 ENEA's Web crawling tasks from social media have been focused on contents from Twitter and Instagram in real case scenarios: in Finance (Cryptocurrencies) and Industry 4.0 (Luxury/ Fashion domain) [1]. On both application domains, crawling tasks are related to: text crawling from tweets and from Instagram's tags and captions; and

image crawling (forthcoming). All raw data were crawled and stored in files: info.txt (likes, comments, caption, commentMessages, dataTime), in raw.html (content message) and image.jpg.

3.1 Social Media Platforms and Application Domains

Comparing Social Media Platforms, Manikonda et al. (2016) found that different Social Platforms offer different contents: i.e. Twitter hosts a significantly higher percentage of posts on business while Instagram is used more for posting about fashion; on Instagram personal and social aspects and news dominate, whereas Twitter is mainly dominated by opinions and work-related aspects; Instagram is largely a sphere of positive personal and social information, whereas Twitter is primarily a news sharing media with higher numbers of negative emotions shared by users; on Instagram users seem to mainly share their joyful and happy moments of their personal lives while content on Twitter are mainly related to news, opinions or other general user interests; furthermore on Instagram posts receive more visibility in terms of receiving more likes (and comments) compared to the favourites of posts on Twitter. Liu et al. (2021) found that Twitter may not be the ideal social media platform in which to investigate luxury brands' customization efforts. Twitter is most often used: by luxury brands as a mean for social media marketing along with Facebook and Instagram; by customers, Twitter is useful to learn about or to comment on live events in real time, while Facebook is mostly informative about promotions [8]. Geo-referenced social media provide info on: the occurrence of a post in space; on metadata about places and users (including number of friends, likes, number of messages, citations); on keywords (often given in forms of hashtags) and on the message content itself. To geo-locate User/Firm messages and hashtags/posts, language expressions (next to/near to) and geographical names are used. Due to the idea behind Instagram (i.e. to share photographs and videos by using hashtags (#) so that other users can find the photographs)[10] by working with hashtags one may chart: how and when users adopt the same language, so noting language-alignment and densities in some place by geo-locating hashtags; whether and how hashtag characteristics, types (structural, entertaining or artistic) and intensity (frequency) reflects hashtag style in individualistic vs collectivistic cultures (i.e. the "we-feeling"); how hashtags show the dynamic of geographical and temporal diffusion from key locations.

In this ongoing research - comparing the two application domains (Finance vs. Luxury Fashion) with reference to crawling and analytics tasks - preliminary findings are as follows: both application domains need previous cross-linguistic semantic pre-analysis enabling to refine crawling strategies and data analysis; web crawling and data analytics tasks need to be scheduled, taking into consideration the following application domains' differences: (i) Financial Knowledge vs. Fashion Knowledge [1]; (ii) in financial markets the timing of tweets and Instagram posts from the centres of diffusion of influence are correlated to the opening/closing hours of financial market. In luxury fashion domain timing is related to centres of diffusion and events (e.g. fashion week), and to brand-vs-audience daily and weekly timing; (iii) the geo-locating of spreading messages on Twitter and Instagram follows a longitudinal axis for financial markets (from Eastern to Western Stock Exchanges); and a latitudinal axis for luxury markets (from Paris, New York, Milan, Rome in the Northern hemisphere to the Southern hemisphere); (iv) there are also difference in the timing and the geo-locating of user/firm hashtags and posts (influential sites create a hierarchical time-space relation in which everywhere else in the world is subordinate: i.e. the rhythms and the aesthetic direction of the world's mass-market fashion industries are shaped by fashion events and activities).

3.2 Data analysis in Social Digital Luxury: Preliminary Findings

Social Media Platforms are becoming powerful tools for direct selling not just advertising. In Luxury Fashion domain, Social Media metrics are used for: (i) growing the brand: awareness gives information on current and potential audience (followers, shares); (ii) turning customer into advocate: engagement shows how audiences interacts with your content (comments, likes, @mentions); (iii) driving leads and sales: conversions demonstrates the effectiveness of user social engagement (website clicks, email signups, etc); (iv) improving customer retention: consumer reflects how active customers think and feel about a brand (testimonials, influencer, sentiment, etc,). Social media metrics are often described as "vanity metrics"[10] measuring the self-online and vanity; "Sharing" can be considered as a particular form of wish fulfillment. Vanity metrics rely on the counting of page views and likes as indicators of success[9]. In the Luxury Fashion domain, the analysis of linguistic features enables to cluster textual messages in Twitter and Instagram, in different dimensions: entertainment, trendiness, interaction or customization (personalization)[4]. In Social Media Networks, Firm's Knowledge and Customer

Knowledge intertwine. By using Big Data Analytics we can access these different types of knowledge. In order to investigate firm-consumer interactions in social media, Big Data Analysis gain an interesting insight to understand customer engagement on social media and to formulate more effective brands' social media marketing efforts [4]. From social media sources we extract firm's and customer knowledge, crawling and analyzing raw textual data from Twitter and Instagram. In this case-study, to run the crawling activities we searched for hashtags associated with luxury items, such as words (e.g. #fashion), luxury brand names, branded products, events (#MilanFashionWeek). Due to the strategy of multiple presence in Social Network Platforms, Luxury brands have more than one accounts[4]. For this reason the simple search of the "brandname" is not able to mine all accounts, so we extended the search using semantics and found other brand account's name on social platforms. Similarly for words connected to the luxury semantic domain such as "fashion", we first searched for *#fashion* and then - in an iterative process - we found other hashtags: *#Fashion, #fashionblog, #fashionblogger, #fashionweek, #FashionWeek, #fashionart*. For the fashion event "Milan Fashion Week" we searched for *#MilanFashionWeek*, and then found other hashtags: *#milanfashionweek, #milanofashionweek, #milanofashionweek19, #MFW, #mfw, #mfw19, #mfw2019, #mfwss20, #milan, #Milan, #milan, #milanitaly, #milano, #milanocity, #milanofashion*). In scheduling crawling tasks, Brand-Audience Geo-locating and Timing have a great relevance: Fashion has a changing nature, and seasonal rhythms alternate with capsule collections (limited edition). Brand-audience relationships experience a spatio-temporal misalignment because Fashion Knowledge has a space-time positioning in the Northern hemisphere, for this reason cities in the Southern hemisphere experience in reverse both their seasons and the temporal order of knowledge transmission (in Australia, 'codified' fashion knowledges precede the local buzz of fashion ideas from the streets). Timing is an essential part of brand's social presence. Brand and audience activity (by day of the week and by hour of day) shows a timing disparity between volume of brand's posts and audience's conversations directed at luxury brands: brands are most active during working days and hours, while their audiences are active at all times and are tweeting at brands on Sunday and at times from 9:00 to11:00 p.m.[5] In this exploratory on-going study, using big data we expect to find evidence on relationships between firm's engagement activities and the resulting customer engagement, and consequently to find answers to the following research questions: (i) which social media platform luxury fashion labels most frequently use in Western Countries and in other world's regions; (ii) how luxury fashion brands are using Twitter and Instagram in terms of: types of multimedia types content (photos, video), content subjects (posts being about existing products, red carpet events, candids- and behind-the-scenes content; fashion shows, previews of products and stores); content themes (artistic, educational, in-store/special events); (iii) whether luxury fashion labels' usage of Instagram differ based on their country of origin; (iv) how direct (products) and indirect marketing (events, celebrities, etc.) mirrors in Social Media Networks; (v) whether and how user engagement relates to financial performance of brands; vi) how content category variables are distributed (hashtags, mentions or other content types): in posts with single/multiple hashtags or mentions; in posts adopting an exclusivity/ belonging/youth appeal; (vii) how brand identity and brand image reflects in hashtagging; (viii) how hashtags can show the level of customer activity: consumption, contribution or creation. (ix) In more detail, as far as hashtagging in fashion domain is concerned, we expect to find evidences on how narcissism (in Western cultures) and "Mianzi"[6] (for Eastern cultures) reflects in hashtags; (x) how and when hashtags use the entertainment, trendiness, interaction or customization dimensions [4]; (xi) and how hashtags show the dynamic of geographical and temporal diffusion. For the sake of brevity, all these findings will be reported in forthcoming publications.

4 Conclusions

From this research, some considerations can be drawn with implications for all application domains:
1. Application domain influences data extraction and analysis because of:
(i) its domain-specific knowledge (requiring domain cross-linguistic ontologies and dictionaries, in order to share meaning and language and to be correctly interpreted); for this reason, before crawling, a

[4] On Instagram Dior has 4 accounts (Dior Official, Dior Makeup, Dior Parfums, Dior Beauty Lovers); Chanel 2 accounts (Chanel Official, Chanel Beauty) – (Source: How 5 fashion brands use Social Media)
[5] Brandwatch.com, Social Insights on Luxury Fashion Industry (2015) - Research: Some Findings from Brandwatch 2015
[6] Mianzi is the reputation achieved through a publicly visible success life.

semantic cross-linguistic content analysis is required in order to find all useful domain-related words' occurrencies, hashtags (#) or mentions (@);
(ii) the timing of domain-specific activities, strictly correlated to the geo-locating of domain's centres of diffusion. It follows that Web crawling tasks need to be coordinated with the timing and the geo-locating of application domain's activities.
2. The Linguistic Alignment[7] is relevant: you can extract how and when users use the same language, aligning on the basis of a shared language (e.g. Instagram photo's captions mostly use English as the language medium; on Twitter the vocabulary from non-English language posts is higher, as compared to the captions on Instagram).
3. Raw data storage is the basis of data crawling, storage and analysis.
In more detail this on-going research in Luxury Fashion domain draws following preliminary conclusions. Social digital Luxury is multi-channel and multi-user. Luxury brand sell "The experience", "The Dream". Luxury's concept experienced a switch from "Old" to "New" Luxury (from a web 1.0 static luxury to a participatory luxury on social media) recently defined[8] as "fluid Luxury". Instagram represents the use of new channels not only as devices for creating desire but also as tools to be used in the most forward-looking way to provide a complete, efficient and fluid shopping experience. In the past, by denying the two-way conversation and maintaining a sense of exclusivity, Social media marketing of luxury brands kept a distance between brand and consumer. The Brand-Audience relationship has now changed in Social Digital Luxury: luxury is more sharable and free on social media. Users co-create and share content about the brand and gather in official and spontaneous online brand communities. Fashion knowledge flows across space and time, and brand-audience relationships experience a space-temporal mis-alignment. For this reason, Brand-Audience Geo-locating and Timing are relevant in crawling and analytics tasks, mirroring in the timing of spreading Tweets and Instagram posts and geo-locating the centres of influences[9]. Furthermore in fashion domain cultural models affect market place, retailer deliver & logistics; the timing of visibility in the Asian Social Platform is defined by governmental economic strategy not by brand; products' prices and the way of payment are different.

5 Final remarks

The Fashion Luxury domain experiences the paradox of exclusivity (rarity, dream, desire) vs. main diffusion in Social Media Platform [4] oscillating from its exclusiveness to the social media's "inclusiveness"; its customization is by definition on a one-to-one individual level, while entertainment, interaction, and trendiness can be delivered through mass communication [4]. Luxury brands strike a balance between exclusivity and inclusivity and spread brand and worth awareness far beyond the target group. In 2019, sector studies [6] forecasted that by 2025, 1/5 of luxury sales would be online. This study started in September 2019 and data reported the pre-pandemic scenario. In 2020 COVID-19 pandemic emergency changed the scenario and our previous findings will be updated. During global lockdowns online marketing and social media played a vital role to sustain sales and reach both high-segment and low-middle segments of consumers/customers. In the next CRESCO Report 2021 our research will investigate scenario's changes during and after coronavirus lockdowns and the new equilibrium in luxury fashion domain, as reflected in social media activities.

References

[1] D. Alderuccio, S. Migliori and ICT-HPC Team. Knowledge Extraction from Social Media Web Sources: Elements affecting Web Crawling and Data Analytics Tasks in ENEAGRID. *High Performance Computing on CRESCO Infrastructure: research activity and results 2019,* pp. 155-158 - ISBN: 978-88-8286-390-6 – ENEA (2020)
[2] G. Bowen, W. Ozuem. Computer-Mediated Marketing Strategies. Social Media and Online Brand Communities. IGI Global, - IGI Publisher - doi: 10.4018/978-1-4666-6595-8 (2015)

[7] The Linguistic Alignment is when multiple actors use the same language or when publics do so.
[8] Beghelli C. Il CEO Venturini: Digitale, nuovi mercati, sostenibilità: così porteremo Valentino nel futuro (2021)
[9] Fashion events and activities in world cities of fashion (Paris–Milan–Rome-London–New York) shape both the temporal rhythms and the aesthetic direction of the world's mass-market fashion industries.

[3] F. Iannone and HPC-CRESCO Team. ENEA HPC CRESCO in the time of Covid-19 and new Supercomputing Frontiers, *ENEA CRESCO in the fight against COVID-19* - ENEA - pp. 8-31 - ISBN:978-88-8286-415-6 (2021)
[4] X. Liu, H. Shin, A.C. Burns. Examining the impact of luxury brand's social media marketing on customer engagement: Using Big Data Analytics and Natural Language Processing. *Journal of Business Researc*h,**125**, pp. 815-826 – doi: 10.1016/j.busres.2019.04.042 (2021)
[5] L. Manikonda, V.V. Meduri, S Kambhampati. Tweeting the Mind and Instagramming the Heart: Exploring Differentiated Content Sharing on Social Media. *Proceedings ICWSM 10th International AAAI Conference on Web & Social Media* (2016)
[6] Mc Kynsey&Company. The Age of digital Darwinism. Experience the dynamics of digital luxury. McKinsey insights (2019)
[7] O. Naumova, S.Bilan , M. Naumova. Luxury consumers' behavior: a cross-cultural aspect. *Innovative Marketing* , **15**(4), pp. 1-13. doi:10.21511/im.15(4).2019.01 (2019)
[8] PMX Agency Trend Report: Luxury Brands online (2017)
[9]A. Rao, N. Spasojevic, Z. Li,T. Dsouza. Klout Score: Measuring Influence Across Multiple Social Networks. *IEEE International Conference on Big Data* doi:10.1109/BigData.2015.7364017 (2015)
[10] R. Rogers, Otherwise engaged: Social Media from Vanity Metrics to Critical Analytics, *International Journal of Communication* **12**, pp. 450-472 (2018)
[11] S. Weller, Fashion as viscous knowledge: fashion's role in shaping trans-national garment production. *Journal of Economic Geography*, Oxford University Press,vol.**7**(1), pp. 39-66, (2007)

MULTI-SCALE MODELING OF SALIVA DROPLETS AIRBORNE TRANSPORT IN RELATION TO SARS-COV-2 TRANSMISSION

Valerio D'Alessandro*, Matteo Falone and Renato Ricci

Università Politecnica delle Marche, Dipartimento di Ingegneria Industriale e Scienze Matematiche, Via Brecce Bianche 12, 60131, Ancona, Italy

ABSTRACT. It is well known that SARS–CoV–2, can be transmitted through airborne diffusion of saliva micro-droplets which travel into atmospheric air through a thermo-fluid dynamic interaction with it.
Social distancing is a key-element to limit or even avoid SARS-CoV-2 spread. However, it is important to emphasize that available knowledge is largely inadequate in predicting the airborne diffusion of infectious droplets emitted during extraordinary respiratory activities. Consequently, it is very difficult to achieve proper guidelines for social distancing rules based on scientific evidence. For this reason, the main aim of our research activity is to provide a contribution to numerical modelling of saliva droplets diffusion produced by coughing. Thus, a campaign of HPC computations were performed on ENEA CRESCO in order to give a better understanding on: (i) saliva droplets diffusion physics; (ii) the effectiveness of the social distancing rules adopted in Italy during the pandemic.

1 Introduction

It is well known that several viruses, as well as SARS-CoV-2, can be transmitted through airborne diffusion of saliva micro-droplets, following three typical infection mechanisms: (i) direct transfer of large droplets to the receiver's conjunctiva, mouth, or nose; (ii) physical contact with droplets deposited on the surface and subsequent absorption to the nasal mucosa of the receiver; and (iii) inhalation of respiratory ejected aerosolized droplet nuclei [1].
Saliva droplets travel because of their inertia and aerodynamic interaction with moist air. Furthermore, their mass varies due to evaporation and, in most cases, this element can produce a supersaturated solution of sodium chloride (NaCl) in water, leading to a crystallization and growth process that generates solid dry nuclei. Therefore, it is very straightforward to understand that the study of droplets' thermo-fluid dynamics is the key ingredient to determine the guidelines on social distancing, face mask wearing, and the implementation of new practices in daily social life.
In this context, one of the aims of this research work is the development of a new multi-scale computational model, relying on the well-established OpenFOAM library, for the evaluation of saliva droplets' dynamics during coughing. Specifically, the crystallization kinetics triggered from NaCl/water solution supersaturation is taken into account.
This paper is organized as follows: the governing equations are reported in Sec. 2, while the numerical approximations are discussed in Sec. 3. Numerical results are shown in Sec. 4. Finally, Sec. 5 contains the conclusions.

2 Governing Equations

The numerical computations are performed using an Eulerian–Lagrangian framework. In particular, the particle-source-in-cell (PSI-Cell) method [2] is adopted to couple Eulerian and Lagrangian phases, while Population Balance Equation (PBE) is solved within the Lagrangian frame to evaluate the effect related to the salt crystallization kinetics in each droplet.
Eulerian phase was modeled using compressible Reynolds Averaged Navier-Stokes equations, here not reported for sake of compactness, in which the unclosed terms are handled using standard constitutive equations, *i.e.* Fick law, Fourier law and rheological equation for Newtonian fluids. Turbulence

*Corresponding author. E-mail: v.dalessandro@univpm.it

modeling is performed using standard SST k-ω model, while source terms are adopted for Lagrangian and Eulerian phases coupling as discussed in D'Alessandro et al. [3].
Saliva particles are handled using a Lagrangian frame and, for efficiency reasons, they are organized in groups, called parcels. Each parcel represents, in our approach, the center of mass of a small cloud of non-collisional spherical particles having the same properties. Initial particels' diameters follow the Rosin-Rammler distribution, calibrated as in [3]. Moreover, trajectory, momentum, mass, and energy equations are solved and, due to the particles' dimension adopted, the forces acting on the parcels are gravity, aerodynamic drag, and buoyancy [4,5]. Aerodynamic drag coefficient is obtained from Putnam correlation [6], while the convective heat transfer coefficient and the mass transfer one are obtained from the Ranz-Marshall correlation for Nusselt and Sherwood numbers [7].
In this study PBE is solved at parcel level to consider droplet nuclei generation and its coupling with PSI method allows to model micro-scale particles behaviour induced by meso-scale thermo-fluid dynamic phenomena.
PBE for spatially in-homogeneous crystallization processes can be written as follows [8]:

$$\frac{\partial N_j}{\partial t} + \boldsymbol{\nabla} \cdot \left(N_j \boldsymbol{u}_p\right) - \boldsymbol{\nabla} \cdot \left(D_t \boldsymbol{\nabla} N_j\right) = -\sum_j \frac{\partial \left(G_j N_j\right)}{\partial r_j} + B \prod_j \delta\left(r_j - r_{j0}\right) + h \tag{1}$$

where, on the left-hand side, N_j is the NaCl crystals number density within a parcel, and D_t is the local turbulent diffusivity. On the right-hand side, G_j is the growth rate, r_j is the particle internal coordinate, r_{j0} is the particle internal coordinate for a crystal nucleus, δ is the Dirac function, and B is the nucleation rate. The last term, h, is the creation or destruction of particles which is neglected because only the crystallization phenomenon is considered. It is worth noting that the convective and diffusive terms for the particles disappear because each parcel is assumed to be well mixed and tracked independently in the Lagrangian frame.
Starting from semi-discrete PBE, obtained by integrating eq. (1) over r, and considering the parcel-averaged crystal mass, $N_{w,j}$, we can obtain the following formulation [8]:

$$\frac{\partial N_{w,j}}{\partial t} = -\frac{\rho_c k_v}{\Delta r}\left(r_{j+1/2}^4 - r_{j-1/2}^4\right) max\left(sign\, \Delta c, 0\right) \times \left[G_{j+1/2}\left(f_j + \frac{\Delta r}{2}(f_r)_j\right) - G_{j-1/2}\left(f_{j-1} + \frac{\Delta r}{2}(f_r)_{j-1}\right)\right] + B|_{j=0} \tag{2}$$

where $\Delta c = c - c^*$ is supersaturation, while c and c* are the NaCl concentration and its solubility in pure water. G_j and B values in eq. (2) are obtained from literature experimental data [16, 17].
From eq. (2), can be also calculated the total mass related to crystallization process, m_{cr}, and the radius of of the droplets' dry nucleus, r_N, as follows:

$$m_{cr} = \pi \frac{D_p^3}{6} \sum_j N_{w,j}\,, \qquad r_N = \frac{\sum_j N_{w,j} r_j^4}{\sum_j N_{w,j}\, r_j^3} \tag{3}$$

3 Numerical approximation

The governing equations solution relies on the unstructured, collocated, cell-centered finite volume approach available within OpenFOAM library. An implicit, three level, second-order scheme was used for the time integration. Face interpolation of convective fluxes is handled by the linear upwind scheme, whereas diffusive terms are discretized by a standard second-order central scheme. Furthermore, pressure–velocity coupling is handled through the Pressure-Implicit with Splitting Operators (PISO) procedure.
Lagrangian phase momentum and mass equations were solved using a standard Euler scheme for time-integration, while energy equation was solved analytically. A special care was devoted to PBE which was time-integrated using an explicit Strong Stability Preserving Runge-Kutta (SSPRK) having 9 stages and 5-th order of accuracy [11] to avoid blow-up of the computations. Moreover, $(f_r)_j$ terms, are approximated by the min-mod limiter [8].

In the present work, a 3D computational domain representing an open air volume starting from the mouth print of a standing coughing person was considered. A length L = 4 m, a width W = 1 m, and a height H = 3 m were adopted, in accordance with Dbouk and Drikakis [12]. A suite of three fully structured grids, having hexahedral cells were built in order to discretize the domain. All the related details as well as the space-time convergence study can be found in D'Alessandro et al. [3].

3.1 Initial and boundary conditions

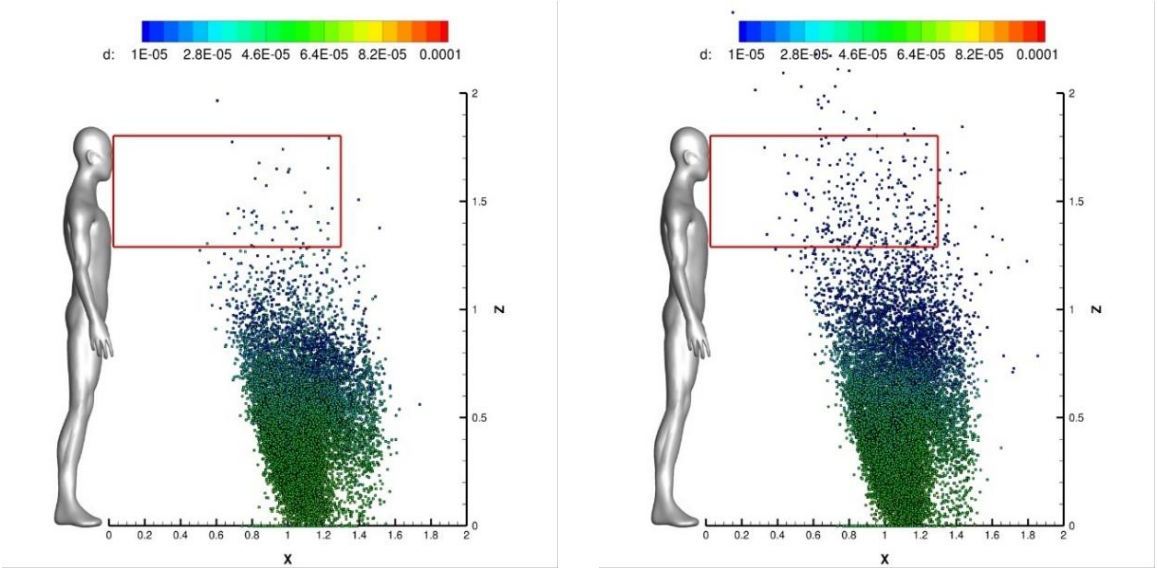
The human cough was simulated by means fluid and parcels injection at the mouth print boundary. A stepped velocity, lasting 0.12 s and equal to 8.5 m/s, was assumed both for the carrier fluid and injected parcels in accordance with the measurements carried by Scharfman et al. [13].
The initial total mass of saliva droplets laden into the domain for a single cough event is 7.7 mg; this value was deduced by previous experimental measurements [14] and CFD simulations [3,12]. Furthermore, a completely diluted NaCl/water solution was adopted for the saliva chemical composition [15]. NaCl mass is fixed equal to 1% of droplets' initial mass.
A still air volume was considered having an initial temperature of 20° C with relative humidity fixed at 50%. The ground is at 25° C, while the air and droplets ejected by coughing are at 34° C. It is important to remark that a preliminary computation was performed for the sole Eulerian phase, to achieve a field initialization consistent with atmospheric conditions [3].

4 Results

In this section the numerical results referred to the saliva droplets produced during coughing are presented. Some cloud characteristics are computed in order to investigate its diffusion, i.e., (i) the cloud center of mass, $\boldsymbol{G}$, and (ii) fraction of particles present in a reference volume, Φ_Ω, defined as:

$$\boldsymbol{G} = (x_G, y_G, z_G,) = \frac{\sum_{i=1}^{N_p(\Omega_0)} m_{P,i}\boldsymbol{x}_{P,i}}{\sum_{i=1}^{N_p(\Omega_0)} m_{P,i}}, \qquad \Phi_\Omega = \frac{\sum_{k=1}^{N_p(\Omega_i)} N_{p,k}}{\sum_{k=1}^{N_p(\Omega_0)} N_{p,k}} \tag{4}$$

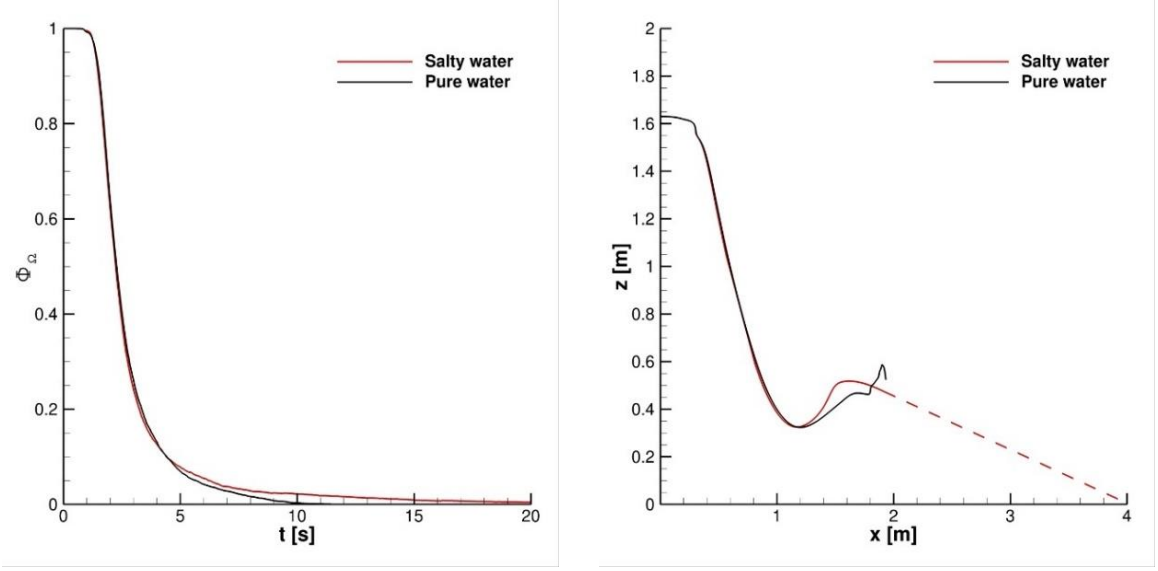
where $N_p(\Omega_0)$ is the number of parcels laden in the overall domain, Ω_0, in a given time instant. It is worth noting that x-axis is associated to stream-wise direction, y-axis represent the transverse direction, while z is the vertical one. On the other hand, Φ_Ω is used to track the droplets' population distribution in a risk volume $\Omega_i = [0, \alpha_i] \times [-0.5, 0.5] \times [1.3, 1.8]$.
The parameter α_i spans the following values: 1.0 m and 1.3 m. Ω_i stream-wise dimension was selected in order to investigate the effectiveness of the safety distance adopted in Italy, equal to1 m. The transverse direction range is considered to completely cover the domain. The z-axis interval is defined for acting on a sufficiently wide range of possible virus receivers' heights.
In this section saliva chemical composition effect is addressed. More in depth, the impact of NaCl presence within droplets, rather than pure water particles, on the related cloud space-time evolution is discussed. In this framework, looking at Fig. 1 (a) and Fig. 1 (b), it is very easy to perform a qualitative analysis about clouds behaviour. Indeed, at t = 10 s, the effect of NaCl crystallization process produces an evident impact on the parcels' space distribution. In particular, the clouds shape is clearly different, and the risk area is sensibly more populated, even for a stream-wise length greater than 1 m, when salty droplets are considered. This evidence is also expressed quantitatively from Φ_Ω time-history represented in Fig. 2 (a). It is very important to put in evidence that the particles contained into Ω_i domain are dry. They have diameters approximately around 13 μm and a fly time having an order of magnitude of 60 s is expected.



(a) Pure water droplets (b) Salty droplets

Fig. 1: Cloud representation at t = 10 s. Parcels are colored with the particle diameter.

The center of mass trajectory on the x–z plane is shown in Fig. 2 (b). For 0.35 m< x_G < 1 m, the x_G, z_G curve underlines a peculiar trend. Actually, the trajectory is almost linear due to the cancellation of the inertial term. For x_G > 1 m the impact of droplets chemical composition is evident, and it is related to the crystallization effects and the consequent droplet nuclei generation. For x_G > 2 m case the curve $z_G = f(x_G)$ is extrapolated since the background flow is extinct and the Lagrangian particles have uniform velocity. It is really evident that dry nuclei are able to reach distance sensibly longer than pure water droplets which completely evaporates from the domain in less than 20 s.



(a) Center of mass evolution (b) Fraction of particles in the risk volume Ω_i

Fig. 2: Saliva cloud parameters comparison for salty and pure water droplets

5 Conclusions

This paper addresses the development and application of an Eulerian–Lagrangian model for saliva droplets' cloud deriving from coughing.

More in depth, we have introduced a multi-scale mathematical model which adopts a coupling strategy between PSI-Cell method and PBE solution within each Lagrangian particle. This news approach allows to predict NaCl crystallization kinetics effect. In this context, it was showed that the chemical composition of saliva droplets laden into CFD domain strongly affects the obtained solution. Specifically, the NaCl crystallization produces dry particles having small diameters which can be transported for time longer than pure water droplets.

Future work will be devoted to more realistic configurations in several relative humidity conditions since this parameter strongly affects evaporation phenomena, hence crystallization kinetics.

Acknowledgments

The authors want to acknowledge "Associazione Nazionale Big Data" that awarded this research work within COVID-19–Fast access to the HPC supercomputing facilities program. We acknowledge ENEA for awarding us access to CRESCO6 based at Portici.

References

[1] R. Mittal, R. Ni, and J.-H. Seo. The flow physics of COVID-19. J. Fluid Mech. 894, F2, (2020).

[2] C. T. Crowe, M. P. Sharma, and D. E. Stock. The particle-source-in cell (PSI-Cell) model for gas-droplet flows. J. Fluids Eng. 99, 325–332, (1977).

[3] V. D'Alessandro, M. Falone, L. Giammichele, and R. Ricci. Eulerian–Lagrangian modeling of cough droplets irradiated by ultraviolet–C light in relation to SARS-CoV-2 transmission. Phys. Fluids 33, 031905, (2021).

[4] G. Busco, S. R. Yang, J. Seo, and Y. A. Hassan. Sneezing and asymptomatic virus transmission. Phys. Fluids 32, 073309, (2020).

[5] M. Abuhegazy, K. Talaat, O. Anderoglu, and S. V. Poroseva. Numerical investigation of aerosol transport in a classroom with relevance to COVID-19. Phys. Fluids 32, 103311, (2020).

[6] A. Putnam. Integratable form of droplet drag coefficient. ARS J. 31, 1467–1468, (1961).

[7] W. E. Ranz and W. R. Marshall. Evaporation from drops. Chem. Eng. Prog. 48, 141–146, (1952).

[8] X.Y. Woo, R.B.H Tan, P.S. Chow, and R.D. Braatz. Simulation of mixing effects in antisolvent crystallization using a coupled CFD-PDF-PBE approach. Cryst. Growth Des. 6, 1291–1303, (2006).

[9] J. Dedsarnau, H. Derluyn, J. Carmeliet, D. Bonn, N. Shahidzedeh. Metastability Limit for the Nucleation of NaCl Crystals in Confinement. J. Phys. Chem. Lett., 5, 5, 890–895, (2014).

[10] A. Naillon, P. Joseph, M. Prat, Sodium chloride precipitation reaction coefficient from crystallization experiment in a microfluidic device. J. Cryst. Growth, 463, 201-210, (2017).

[11] S. J. Ruuth and R. J. Spiteri. High-Order Strong-Stability-Preserving Runge-Kutta Methods with Downwind-Biased Spatial Discretizations. SIAM J. Numer. Anal., 42(3), 974–996, (2004).

[12] T. Dbouk and D. Drikakis. On coughing and airborne droplet transmission to humans, Phys. Fluids 32, 053310, (2020).

[13] B. E. Scharfman, A. H. Techet, J. W. M. Bush and L. Bourouiba. Visualization of sneeze ejecta: Steps of fluid fragmentation leading to respiratory droplets. Exp. Fluids 57, 24, (2016).

[14] X. Xie, Y. Li, H. Sun and L. Liu. Exhaled droplets due to talking and coughing. J. R. Soc. Interface 6, S703, (2012).

[15] M.E. Rosti, S. Olivieri, M. Cavaiola et al. Fluid dynamics of COVID-19 airborne infection suggests urgent data for a scientific design of social distancing. Sci Rep 10, 22426, (2020).

CFD SIMULATIONS OF CO2 METHANATION IN A SHELL AND TUBE REACTOR

Antonio Di Nardo*, Giorgio Calchetti, Claudia Bassano and Paolo Deiana

ENEA, Energy Technology Department, Casaccia Research Center, Via Anguillarese 301, Rome 00123, Italy

ABSTRACT. CO_2 methanation represents a way of storage of the energy surplus produced by renewable energy sources. Shell and tubes reactors are among the most widespread type of methanation reactors. In this work the CO_2 methanation in a cooled multi-tubular catalyst reactor has been investigated through 3D CFD simulations, using the Ansys-Fluent™ software. A sensitivity analysis has been conducted for what concerns coolant and reactants temperatures and flow rates, catalyst load and operating pressure in the tubes. Nitrogen and oil cooling have been studied and the results have been compared. Practical solutions for the high temperatures mitigation in the reactive zones have been evaluated and discussed. The model has proved to be very useful for a detailed analysis of the phenomenon and for obtaining essential data that resulted difficult to measure by experimentation.

1 Introduction

Power to Gas (PtG) technology represents one potential tool for managing renewable power intermittency and power generation and is becoming increasingly popular in recent years, as a system for accumulation and storage of the energy surplus produced by solar and wind. In fact the energy surplus can be used to produce hydrogen by electrolysis, which can be subsequently stored or converted into chemicals as methane through CO_2 methanation. Potential CO_2 feedings come from different gaseous streams as: captured CO_2 in energy intensive industry, gasification syngas, biogas, geothermic fields, soil gas and gas wells. In addition, instead of building new infrastructures for hydrogen transportation, it is possible to use the existing gas network, with considerable money savings.
The CO_2 methanation reaction:

$$CO_2+4H_2 \leftrightarrow CH_4+2H_2O \qquad \Delta H_0 298K=-165 \text{ kJ/mol}$$

is highly exothermic and this leads to high temperature zones inside the reactor. High temperatures are favourable from a kinetic point of view but thermodynamic limits CO_2 conversion and methane yield, according to Le Chatelier's principle. Higher working pressure can however counterbalance the effect of temperature. Temperatures control is essential for a correct functioning of the reactor, given the strong exothermicity of the methanation reaction.
A reliable and accurate numerical model therefore becomes fundamental in order to optimize the reactor design in all its aspects. Most of the works in literature use one-dimensional pseudo-homogenous or heterogeneous models or simplified CFD models, using an imposed heat exchange coefficient and a constant coolant temperature. The analysis of CO_2 methanation in a multi-tubular catalyst fixed-bed reactor by 3D simulations is the aim of the present work [1]. The experimental reactor installed to our laboratory [2] has been simulated by the commercial software Ansys-Fluent™ and a sensitivity analysis has been carried out in order to identify possible solutions to hot spots formation.

2 Experimental Rig

The pilot scale experimental facility is a shell and tube reactor containing seven externally cooled fixed-bed steel tubes of 21.3 mm of diameter and long 400 mm. The tubes are placed in a cylindrical insulated case of 102.1 mm of diameter and are arranged at the center and at the vertices of a hexagon of side 28.4 mm. The tubes are filled of catalyst and inert glass particles in the centre part of each tube, for a length

*Corresponding author. E-mail: antonio.dinardo@enea.it.

of about 200 mm, using a dilution ratio of 1/1 in volume, while the remaining part of the tube is filled with inert only. The coolant fluid flows in counter-current compared to the reactants. The catalyst used is a commercial 0.5% Ru wt. on Al_2O_3, made of 3.2 mm cylinders (diameter and height). Ruthenium based catalysts are more stable, offer high activity and CH_4 selectivity and high resistance to oxidation. The reactants H_2 and CO_2 are feed from cylinders, mixed in the desired proportion and heated using an electrical heater. Water is separated by a condenser and the product gas is burned in a flare.

3 Numerical Model

The computational mesh includes 5000000 elements. The mesh elements are hexahedral inside the tubes and tetrahedral outside. The total number of elements has been gradually increased until no variations have been observed in conversion and temperature values, variables of interest in this study. The ANSYS-Fluent™ code has been used to solve the conservation equations for continuity, momentum, turbulence, energy, radiation and species concentration. The internal part of the tubes has been modelled considering the catalyst and the inert bed as a porous medium. The standard k-ε model has been adopted for turbulence and the discrete ordinates (DO) radiation model to solves the radiative transfer equation. The chemical kinetic model adopted in this work was proposed by some of the authors in a previous paper [3].

The simulations have been run on the ENEA-CRESCO platform, using up to 512 cores. Thanks to this computational facility, it has been possible to execute a large number of 3D CFD simulations. The convergence criterion has been an energy residual lower than 10^{-8}, other variables residuals lower than 10^{-4} and when no variation of conversion and maximum temperature have been registered.

4 Results

In the nitrogen cooling case (Fig. 1), it has been not possible to identify operating conditions that allow to contain the high temperatures developed in the reactive zones, because the intense heat release, and that at the same time assure high conversions, unless some suitable solutions are adopted. The catalyst material subjected to excessive temperature undergoes to rapid deterioration. Water vapour reactants dilution and catalyst uneven distribution have demonstrated to be proper techniques to achieve the desired results. In particular the second one is easy to be implemented and doesn't involve the addition of a water vapour generator with the consequent complication of the entire system.

In the oil cooling case (Fig. 2), operating conditions that give low temperatures and high conversions have been more easily found for the case where the catalyst is distributed over the entire tubes, especially if high oil inlet temperatures are adopted. Those temperatures can be however very close to the maximum suggested and can cause problems like thermal oil degradation during the operation of the system. The choice of lower oil working temperatures requires instead higher catalyst loads and the adoption of solutions for temperatures containment, as for nitrogen cooling cases.

While for nitrogen cooling the reaction extinguishes rapidly for a variation of the operating conditions, as for inlet temperatures or catalyst load reduction, in the case of oil cooling the reaction is more stable and undergoes to a slow attenuation. The numerical model implemented turned out to be very useful for several issues as verify different operating conditions, design of experiments and reactor optimization.

Table 1: N2 cooling operating ranges.

WHSV=3500-11500 $Ncm^3/(g_{cat}*h)$
Catalyst= 113-161 g
Reactants pressure =1-5 barg
H_2/CO_2 mole ratio=2.5-5
H_2O reactants %vol =0-50
T_reactants inlet=200-300 °C
N_2 flow rate=8 Nm^3/h
N_2 pressure=1 bar
T_ N_2 inlet=200-300 °C

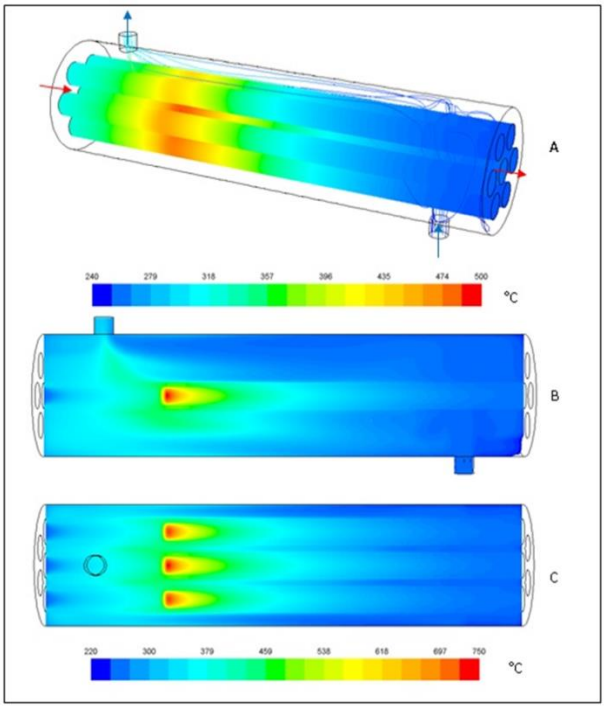


Fig.**1**: Temperature (°C) map. A) Inside view on tubes surface and coolant flow path lines. B) Cross section with coolant inlet and outlet. C) Cross section with three tubes. WHSV=7000 $Ncm^3/(gcat*h)$, P_reactants=3 barg, H_2/CO_2=4.5 mol, T_reactants inlet=250 °C, T_N_2 inlet=250 °C, N_2_flow rate=8 Nm^3/h , catalyst=161 g in the central sectors.

Table 2: Oil cooling operating ranges.

WHSV=2333-7000 $Ncm^3/(g_{cat}*h)$
Catalyst=161-483 g
Reactants pressure =3 barg
H_2/CO_2 mole ratio=4.5
T_reactants inlet=250-350 °C
Oil flow rate=11.5-34.5 l/min
T_ oil inlet=300-325 °C

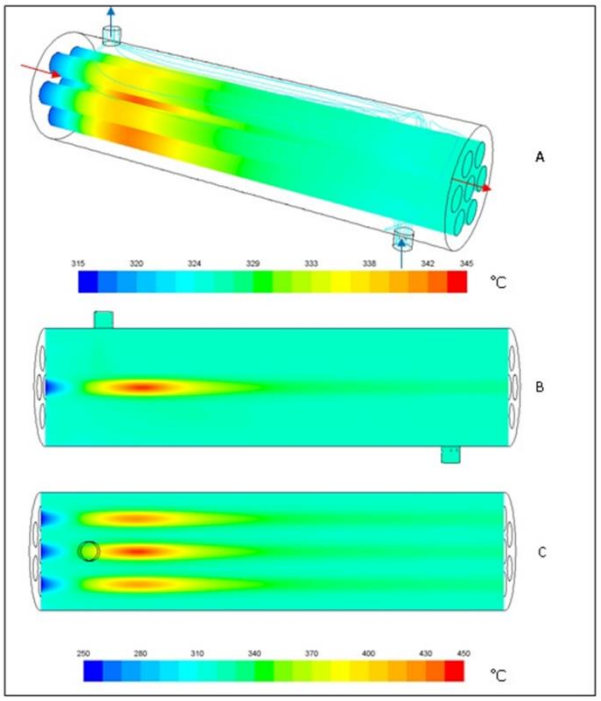


Fig.**2**: Temperature (°C) map. A) Inside view on tubes surface and coolant flow path lines. B) Cross section with coolant inlet and outlet. C) Cross section with three tubes WHSV=3500 $Ncm^3/(gcat*h)$, P_reactants=3 barg, H_2/CO_2=4.5 mol, T_reactants inlet=250 °C, T_oil inlet=325 °C, Oil_flow rate=11.5 l/min, catalyst=322 g over the entire tubes.

References

[1] A. Di Nardo, G. Calchetti, C. Bassano, P. Deiana. CO_2 methanation in a shell and tube reactor CFD simulations: high temperatures mitigation analysis. *Chemical Engineering Science* **246,** 116871, (2021).
[2] C. Bassano, P. Deiana, L. Lietti, C.G. Visconti. P2G movable modular plant operation on synthetic methane production from CO_2 and hydrogen from renewable sources. *Fuel* **253**, pp. 1071-1079, (2019).
[3] L. Falbo, M. Martinelli, C.G. Visconti, L. Lietti, C. Bassano, P. Deiana. Kinetics of CO_2 methanation on a Ru-based catalyst at process conditions relevant for Power-to-Gas applications. *Applied Catalysis B: Environmental* **225**, pp. 354-363, (2018).

MAGNETIC SYSTEMS FOR QUANTUM TECHNOLOGIES: STRUCTURAL AND MAGNETIC PROPERTIES FROM THE CRYSTAL PHASE TO THE ADSORPTION ON SURFACE

Andrea Albino[1], Francesco Buonocore[2], Massimo Celino[2] and Federico Totti[1*]

[1]*Università degli Studi di Firenze, Chemistry Department "U. Schiff", 50019, Sesto F.no, Italy*

[2]*ENEA, Casaccia Research Centre, 00123 Rome, Italy.*

ABSTRACT. The challenge of developing and implementing quantum devices is increasingly topical. Indeed, quantum computing, spintronics, and ultra-high density storages are experiencing a real boost since they appeared both as immature implementations or as next-future promising technological applications. Here we present computational studies on a vandium-based S=1/2 molecule, [$VO(acac)_2$], and several graphene derivatives as potential components of spintronics devices.

1 Introduction

One of the main motivations for work in quantum information science is that certain computational tasks might be executed exponentially faster on a quantum processor than on a classical processor. Although this is a relatively old idea [1] only recent advances in material science permitted the actual realization of this technology. In the context of molecular magnetism several research groups have proposed both systems and methodologies to create a molecular based quantum computer. Synthetic chemistry can here play a fundamental role in modulating specific interactions between microscopic constituents as phonons and spins in order to achieve the desired molecular properties. Among the open challenges, both the control of intrinsic properties of molecular systems and development of synthetic strategies that allow the control of the spatial distribution of the molecules at the surface and interphase, represents a fundamental one.

The aim of the present work is *i)* investigate the magnetization dynamics of molecular spins to engineer their properties toward the desired application – quantum computation or spintronics, *ii)* assess the bulk [2] and interphase [3] properties of materials, that are always determined by cooperative interactions between the constituent molecules, which consequently must be assembled in the lattice in such a way as to tune the macroscopic response.

The observable parameters focused for quantum computation regard the relaxation processes of magnetization of vanadyl-acetylacetonate, [$VO(acac)_2$]: the spin-lattice relaxation time, T_1, and spin-spin relaxation time, T_2. The former is connected to the life time of non-equilibrium population distribution in the spin states, the latter takes into account the life time of the coherent states. T_2 is the ultimate figure of merit to consider in the design of electronic spin molecular qubits as establishing the time limit for the exploitation of coherent states in quantum algorithms and it is strongly dependent on T_1.

The observable parameters focused for spintronics, or spin electronics, involves the study of long-range ordering between spins (ferromagnetism, antiferromagnetism) in hydrogenated graphene [4], especially in terms of exchange coupling constants [5]. Graphene is an ideal material in spintronics. Indeed, thanks to its layer structure, weak spin–orbit coupling that facilitates spin transport and modulation of electrical properties. The use of graphene for applications in electronics suffers from a major drawback: graphene is, in its pristine state, a zero-band-gap semiconductor, as pointed out in 1947 by Wallace [6]. In order to use graphene in possible applications such as photodetector, photovoltaic, sensors, organic light-emitting diodes, organic thin film transistors, supercapacitor, and catalytic applications, it is essential to precisely modulate its electronic properties. Several ways have been explored to induce a finite band gap in graphene. Moreover, because graphene does not contain d or f electrons, the production of

*Corresponding author. E-mail: federico.totti@unifi.it

magnetic moments would be difficult. Making graphene magnetic could lead to high-Curie-temperature diluted magnetism and, ultimately, thin, two-dimensional (2D) magnetic materials, which could match the demands of ever increasing magnetic information storage density. Hydrogenation of graphene is an attractive solution and it is foreseen to open a band gap [7] and to produce a ferromagnetic behaviour [8]. The calculations were performed on the CRESCO6 computing cluster of ENEA's ENEAGRID computing infrastructure [9], using around 1,000,000 CPU hours.

2 Methods

2.1 Vibrational Properties and Spin−Phonon Coupling Coefficient Calculations

In order to correctly account for the vibrational properties of the solid-state systems, where intermolecular interactions become relevant, DFT calculations are performed in the crystal phase with a Gaussian and plane-wave formalism, as implemented in the Quickstep module of the CP2K package. [10] The generalized gradient approximation (GGA) functional in the PBE approximation is chosen for all of the calculations. Van der Waals interactions are taken into account with the nonlocal rVV10 correction scheme.

Calculation of the spin and spin−phonon Hamiltonian parameters has been carried out with the ORCA package. [11] The level of theory used is DFT (PBE0), with a def2-TZVP basis set for the V, O, and S atoms and def2-SVP for the C and H atoms. Calculation of the spin−phonon coupling coefficients is performed following a tensor differentiation procedure, as described in a previous report on single-molecule magnets. [10] This procedure is here applied to the Landé g tensor that describes the coupling between the spins and an external magnetic field.

Exchange coupling parameters, *J*'s, have been calculated by broken-state calculations from the Eq. 2 [4]

$$\Delta E(S_{max} - S_{BS}) = E(S_{max} = S_A + S_B) - E(S_{BS} = |S_A - S_B|) = \sum_{i<j} J_{ij}\left(2|s_i s_j| + s_j\right)\lambda_{ij}\left(s_i \geq s_j\right) \qquad (2)$$

3 Results

Our work aims at presenting an in-depth study on how the spin-phonon coupling in $[VO(acac)_2]$ (Fig. 1), can change as a function of the temperature by using Terahertz Time Domain Spectroscopy (THz-TDS) and DFT. THz spectra of powder samples were recorded between 10 and 300 K. The temperature dependence of the vibrational frequencies was then accounted for in the periodic DFT calculations by using the unit cell parameters measured at two different temperatures and the optimized ones, as usually reported in literature. In such a way it was possible to calculate the observed THz anharmonic frequency shift with high accuracy. The overall differences in the spin-phonon coupling magnitudes as a function of the temperature were also highlighted showing that the computed trends have to be ascribed to the anisotropic variation of cell parameters.

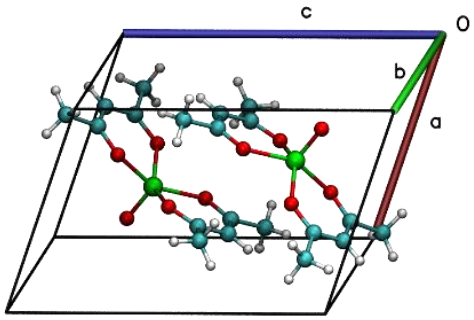


Fig.1: Primitive cell of $[VO(acac)_2]$ belonging to the triclinic P-1 space group. Colour code: Vanadium = green; Oxygen = red; Carbon = cyan; Hydrogen = white.

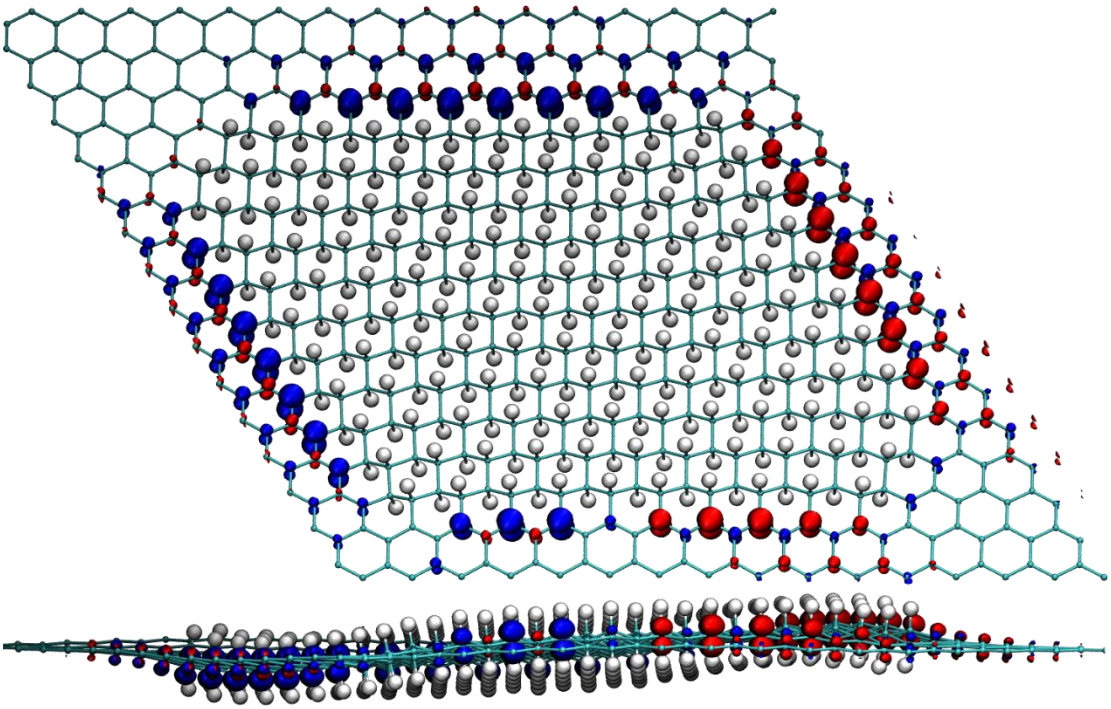


Fig.2: Top: perspective view of a graphane island showing magnetic moments on its interface with surrounding graphene structure. Bottom: side view of the graphane island, the ripples formed by hydrogenation are clearly noticeable and can further perturb the band structure. Red and blue correspond to spin densities (up and down, respectively)

The quest of magnetic materials for spintronics, on the other hand, seeks at implantation of stable magnetic moments on covalently hydrogenated graphene (Fig. 2). The formation of covalent bonds between carbon atoms and other species (e.g., hydrogen and fluorine) converts the carbon hybridization from sp_2 to sp_3, triggering significant changes in the geometric structure of graphene.
We investigated several avenues to obtain, at the *in silico* level, a stable magnetic phase for hydrogenated graphene. It is the first time that a comprehensive study on a wide variety of systems is accomplished within the same computational framework, allowing a detailed and reliable comparison of energetics and magnetic properties already studied in literature and new ones.
To overcome the magnetic instability which favours the diamagnetic states for isolated hydrogenated graphene derivatives, two main approaches have been tested: *i)* the adsorption of hydrogenated graphene layer on a substrate (gold and graphene or both) to allow the establishment of new interaction stabilizing the ferromagnetic phase; *ii)* non-long range ordering of hydrogenation by means of localized islands of covalently bonded H.

5 Conclusions

With regards to the two projects the following take-home messages can be drawn:
1) The spin-phonon coupling in vanadyl-acetylacetonate, $[VO(acac)_2]$, can change as a function of the temperature as shown by using Terahertz Time Domain Spectroscopy (THz-TDS) and Density Functional Theory (DFT) calculations. The overall differences in the spin-phonon coupling magnitudes as a function of the temperature were also highlighted showing that the computed trends have to be ascribed to the anisotropic variation of cell parameters.
2) The energy of ferromagnetic G-one can be significantly lowered by its adsorbtion on a surface even if this is not enough to make it more stable than the diamagnatic one. However, the injection of significant magnetic moment in graphene can be also achieved by the creation of a high-density of small G-ane islands to magnify the graphene/G-ane interfaces.

References

[1] R. P. Feynman, “Simulating Physics with Computers,” vol. 21, pp. 467–488, 1982.
[2] M. Atzori et al., “Quantum Coherence Times Enhancement in Vanadium(IV)-based Potential Molecular Qubits: the Key Role of the Vanadyl Moiety,” J. Am. Chem. Soc., 138, 35, 11234–11244, 2016.
[3] S. Sanvito, “Molecular spintronics: The rise of spinterface science,” Nat. Phys., 6, 562–564, 2010.
[4] F. Buonocore, A. Mosca Conte, N. Lisi, “Effects of the substrate on graphone magnetism: A density functional theory study,” Phys. E, 78, 65–72, 2016.
[5] A. Bencini, F. Totti, “A few comments on the application of density functional theory to the calculation of the magnetic structure of oligo-nuclear transition metal clusters,” J. Chem. Theo. Comp., 5, 144–154, 2009.
[6] P. R. Wallace, “The Band Theory of Graphite,” Phys. Rev., 71, 622–634, 1947.
[7] E. J. Duplock, et al. “Hallmark of perfect graphene. Phys. Rev. Lett., 92, 1–4, 2004.
[8] L. Xie, et al. “Room temperature ferromagnetism in partially hydrogenated epitaxial graphene,” Appl. Phys. Lett., 98, 1–4, 2011.
[9] F. Iannone et al., CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, 1051-1052, 2019.
[10] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. Vandevondele, “Cp2k: Atomistic simulations of condensed matter systems,” Wiley Interdiscip. Rev. Comput. Mol. Sci., 4, 1, 15–25, 2014.
[11] F. Neese, “The ORCA program system,” *Rev. Comput. Mol. Sci.*, 2, 1, 73–78, 2012.

AB INITIO MOLECULAR DYNAMICS SIMULATIONS OF AMORPHOUS GEO2 IN EXTENDED PRESSURE RANGE

G. Mancini[1*] and M. Celino[2]

[1]*Università di Camerino, Sezione di Fisica della Scuola di Scienze e Tecnologie, Via Madonna delle Carceri 9b, 62022 Camerino (MC), Italia*[1]

[2]*ENEA, Ente per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile, C. R. Casaccia, Via Anguillarese 301, 00123 Roma, Italia*

ABSTRACT. Extensive *First Principles* Molecular Dynamics Simulations were carried out for amorphous GeO_2 under pressure in the range 00-100GPa. A well marked density increase taking place at 46-48GPa resulted irreversible upon releasing the system back to 0GPa. The computed Ge-O coordination numbers are in good agreement with the experimental measurements. The simulations clearly showed the steady evolution from four-fold to seven-fold coordinated Ge-O glass configuration.

1 Introduction

Due to GeO_2 poly-amorphism and the associate properties, the determination and the study of the structure of glassy GeO_2 subjected to high pressures is of great interest in various scientific fields (condensed-matter physics, materials science, geology, etc) as well as for engineering and industrial applications. Indeed, available results for GeO_2 are quite diverging: although experimental studies highlighted a gradual transition from the four-fold to the six-fold coordinated Ge-O glass configurations, the pressure at which a fully six-coordinated state occurs was debated long, with existing experiments suggesting values in a wide interval ranging from 9GPa to 30GPa [1-6]; moreover recent experiments reported the persistence of the six-fold coordination up to 100GPa [7], this result being independently confirmed up to 133GPa [8], while others found the Ge–O coordination number (CN) remained constant at ~6 in the range 22.6-37.9GPa, then rapidly increasing with increasing pressures, up to the value the of 7.4 at 91.7GPa [9]. Molecular Dynamics [MD] simulations presented a similar scenario, yielding for the mean Ge-O coordination a value below 6.5 at pressures up to 100GPa [10-12]. Both experimental and simulations results reported a gradual increase of the coordination number (CN) with increasing pressure; in addition, no abrupt change in density was observed. A recent paper of ours [13] illustrated a series of *first principles simulations* of amorphous GeO_2 in the pressure range 0-30GPa yielding a close agreement with the above-mentioned results; highlighting that, in particular, all pressure-induced transformations were reversible upon system relaxation back to 0GPa. In this paper we present the results we obtained extending up to 100GPa our *Ab Initio Molecular Dynamics* simulations of amorphous GeO_2 subjected to hydrostatic pressures at room temperature. Two relevant facts resulted: the Ge-O coordination number increases up to around 7 in the higher pressure range, and a sudden, sharp density increase at 46-48GPa that proved to have irreversible effects causing a higher density then the initial one when the pressure was released back to 0GPa.

2 Simulation details

The simulations were performed using CPMD software, based on the self-consistent evolution of the electronic structure within the frame of density functional theory [14,15]. Our simulated system consisted in amorphous GeO_2 at room temperature, obtained quenching down a liquid sample from 4000K to room temperature. It was formed by 80 germanium and 160 oxygen atoms, contained in a 15.602Å-edged cubic simulation box to get a density ρ= 3.66 gr/cm^3. A generalized gradient approximation (BLYP-GGA) was adopted for the exchange and correlation part of the total energy [16,17] together with norm conserving Goedecker pseudo-potentials for the core-valence interactions

*Corresponding author. E-mail: giorgio.mancini@unicam.it

[18,19]. The electronic wave functions were expanded in plane waves up to the kinetic energy cut-off of 100 Ry. A value of 400a.u. was used for the fictitious electronic mass, 2000cm $^{-1}$ and 10000cm $^{-1}$ for the characteristic frequencies of ions and electrons, respectively, and 2a.u. (0.048 fs) for the time-step. The initial system was quenched down to 300K and relaxed at an initial pressure of 0Pa; it subsequently was submitted to increasing pressures up to 100GPa. Pressures were applied via the Parrinello-Raman method [20], in steps of 2GPa. At each stage the atomic system was equilibrated for 3ps according to the criterion that all the physical quantities converged. Thermostatting was obtained by Nosé Hoover chain thermostats [21] set on ions and electronic degrees of freedom, a fictitious electronic kinetic energy of 0.05au was used throughout the entire pressure range.

3 Results

Fig.1 depicts the evolution for the volume of our system. As anticipated a clear slope discontinuity occurs in the interval 46-48GPa, indicating a sudden structural change, as confirmed by the marked changes of mutual distances illustrated in Fig. 2 - note in particular the splitting of the $g_{GeGe}(r)$ peak.

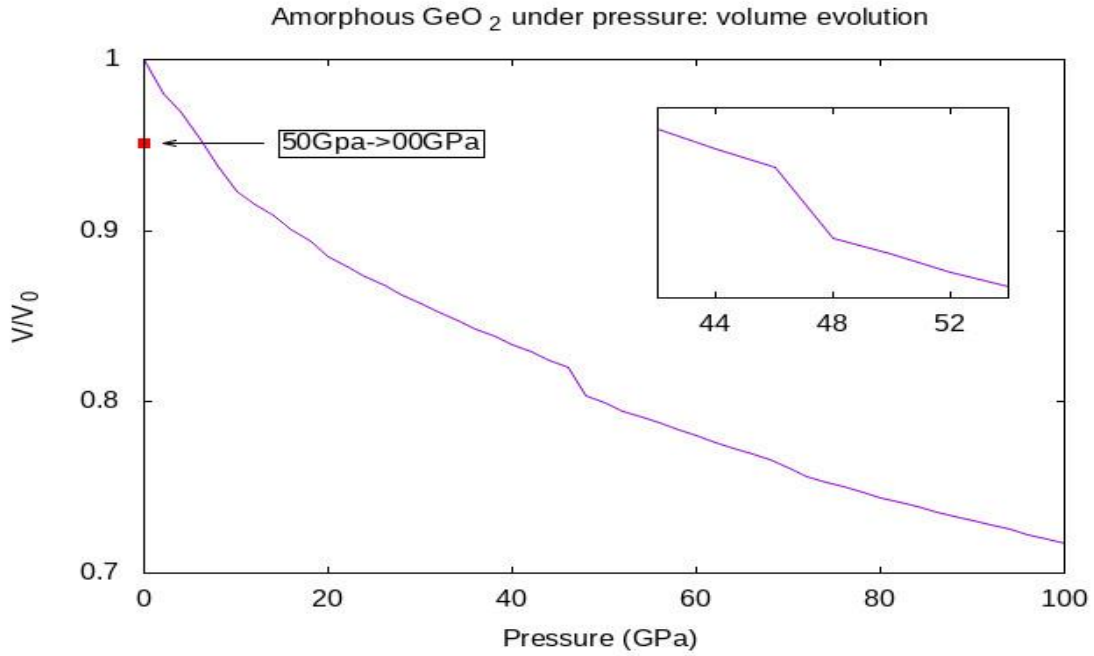


Fig.**1**: Volume evolution for amorphous GeO_2 with pressure (room temperature). A sharp slope change takes place in the interval 46-48GPa. The red box marks the final volume for the system when relaxed back to 0GPa from 50GPa, evidencing the permanent nature of the transition.

The associated changes occurring in Ge-O and O-O coordination number show a smooth transition characterised by crossing curves as shown in Fig.3 for the sole Ge-O. Finally the anticipated evolution yielding a CN growth for Ge-O up to circa 7 with increasing pressure is shown in the same figure.

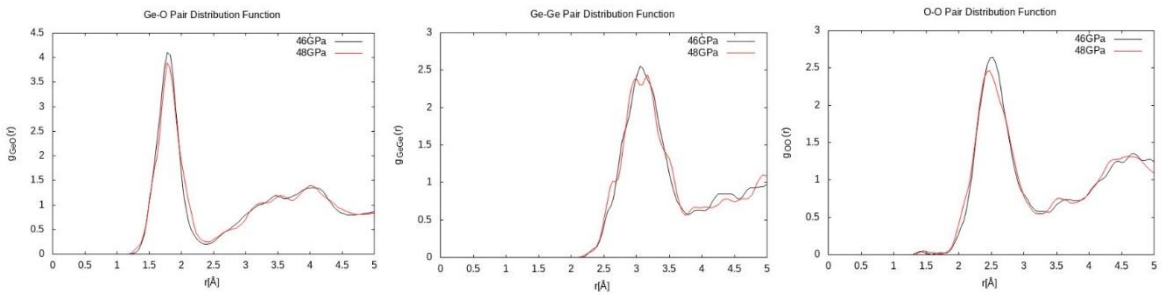


Fig.**2**: Pair distribution functions at 46GPa and 48GPa depicting the nature of the observed transition: a pronounced shortening of O-O distances and a marked split in Ge-Ge ones.

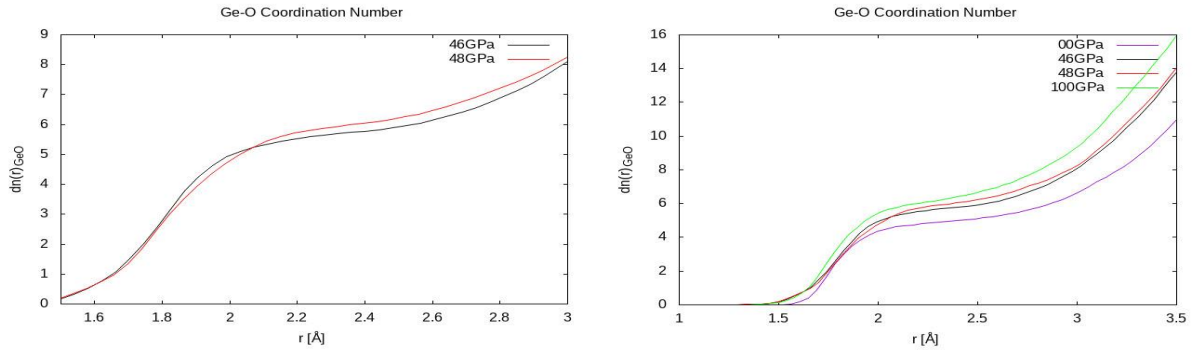


Fig.**3**: CN evolution for Ge-O with pressure: (left) curves cross at the transition – (right) its growth over the entire pressure range, reaching a value around 7.

4 Computational resources

The calculations were performed using the facilities and services available at the ENEAGRID infrastructure (Italy). Molecular Dynamics simulations have been carried out using CPMD v3.15.3 running on CRESCO4 and CRESCO6 clusters. 800GB of permanent disk storage has been granted together with the access to a 2TB scratch area.

5 Conclusions

The choice of Goedecker pseudo-potentials for Ge and O atoms allowed us to perform *ab-intio* simulations for amorphous GeO_2 in a pressure range as extended as 00-100GPa (other usual potentials resulted in divergent simulations at 36GPa [13]). The general frame obtained is in overall agreement with currently known results, both experimental and computational. A phase transition – here reported for the first time - was observed in the range 46-48GPa that induced an irreversible density increase for the sample released back to standard room conditions. In addition, a seven-fold Ge-O coordination resulted at the higher simulation pressures.

References

[1] M. Vaccari M, G. Aquilanti G, S. Pascarelli and O. Mathon. A new EXAFS investigation of local structural changes in amorphous and crystalline GeO_2 at high pressure J. Phys.: Condens. Matter 21 145403 (2009).

[2] M. Baldini, G. Aquilanti, H. K. Mao, W. Yang, G. Shen, S. Pascarelli and W.L. Mao. High-pressure EXAFS study of vitreous GeO_2 up to 44 GPa Phys. Rev. B 81 024201 (2010).

[3] M. Guthrie M, C. A. Tulk,C. J. Benmore, J. Xu, J.L. Yarger, D. D. Klug, J. S. Tse J, H. K. and R. J. Hemley. Formation and structure of a dense octahedral glass. Phys. Rev. Lett. 93 115502 (2004).

[4] P. S. Salmon, J. W. E. Drewitt, D. A. J. Whittaker, A. Zeidler, K. Wezka, C. L. Bull, M. G. Tucker, M. C. Wilding, M. Guthrie and D. Marrocchelli. Density-driven structural transformations in network forming glasses: a high-pressure neutron diffraction study of GeO_2 glass up to 17.5 GPa. J. Phys.: Condens. Matter 24 415102 (2012).

[5] Q. Mei, S. Sinogeikin, G. Shen, S. Amin, C. J. Benmore and K. Ding. High-pressure x-ray diffraction measurements on vitreous GeO_2 under hydrostatic conditions Phys. Rev. B 81 1741132010 (2010).

[6] J. P. Itie, A. Polian, G. Calas, J. Petiau, A. Fontaine and H. Tolentino. Pressure-induced coordination changes in crystalline and vitreous GeO_2. Phys. Rev. Lett. 63 398–401 (1989).

[7] G. Spiekermann, M. Harder, K. Gilmore, P. Zalden, C.J. Sahle, S. Petitgirard, M. Wilke, N. Biedermann, C. Weis, W. Morgenroth, J.S. Tse, E. Kulik, N. Nishiyama, H. Yavas and C. Sternemann. Persistent Octahedral Coordination in AmorphousGeO2 up to 100 GPa by Kβ'' X-Ray Emission Spectroscopy, Phys. Rev. X 9, 011025 (2019).

[8] S. Petitgirard, G.,Spiekermann, K.Glazyrin, J. Garrevoet and M. Murakami. Density of amorphous GeO_2 to 133 GPa with possible pyritelike structure and stiffness at high pressure. Phys. Rev. B 100, 214104 (2019).

[9] Y. Kono, C. Kenney-Benson, D. Ikuta, Y. Shibazaki, Y. Wang and G. Shen. Ultrahigh-pressure polyamorphsim in GeO 2 glass with coordination number >6, Proc. Natl. Acad. Sci. USA 113, 3436 (2016).
[10] D. Marrocchelli, M. Salanne and P. A. Madden. High-pressure behaviour of GeO_2: a simulation study. J. Phys.: Condens. Matter 22, 152102 (2010).
[11] M. Micoulaut. Structure of densified amorphous germanium dioxide. J. Phys.: Condens. Matter 16, L131–L138 (2004).
[12] X. P. Du and J. S. Tse. Oxygen packing fraction and the structure of silicon and germanium oxide glasses. J. Phys. Chem. B 121, 10726 (2017).
[13] G. Mancini and M. Celino. First principles molecular dynamics simulationsof amorphous GeO_2 under pressure in the range 0-30GPa. High Performance Computing on CRESCO infrastructure: research activities and results 2019, pages 133-136, December 2020. ISBN: 978-88-8286-403-3.
[14] CPMD v3.13.2 copyright ibm corp 1990-2008, copyright mpi für festkörperforschung stuttgart 1997–2001.
[15] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. Physical Review Letters, 55(22):2471–2474 (1985).
[16] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. Phys. Rev. A, 38:3098–3100, Sep 1988.
[17] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlationenergy formula into a functional of the electron density. Phys. Rev. B, 37:785–789, Jan 1988.
[18] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. Physical Review B, 54:1703–1710 (1996).
[19] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. Phys Rev B 58:3641 (1998)
[20] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. Phys. Rev. Lett., 45:1196–1199, Oct 1980.
[21] S. A. Nose. A molecular dynamics method for simulations in the canonical ensemble. ´ Molecular Physics, 52:255–268 (1984).
[22] S.A. Nose. A unified formulation of the constant temperature molecular dynamics methods. The Journal of Chemical Physics, 81:511–519 (1984).

GRAFTING- DEGRAFTING MECHANISM ACTING AT THE INTERFACE BETWEEN THIN FILMS AND BRUSH LAYERS

Gianmarco Munaò[1*], Antonio De Nicola[2], Antonio Pizzirusso[2] and Giuseppe Milano[3]

[1]*Dipartimento di Scienze Matematiche e Informatiche, Scienze Fisiche e Scienze della Terra, Università degli Studi di Messina, viale F. Stagno d'Alcontres 31, 98166 Messina, Italy.*

[2]*Dipartimento di Chimica e Biologia "A. Zambelli", Università di Salerno, Fisciano 84084, Italy*

[3]*Dipartimento di Ingegneria Chimica, dei Materiali e della Produzione Industriale, Università di Napoli "Federico II", Corso Umberto I, 40, 80138 Napoli, Italy.*

Abstract. In the present contribution we discuss a simulation study concerning the grafting-degrafting mechanism which acts at the interface between polymer thin films and brush layers. For such an aim we make use of the hybrid particle-field molecular dynamics method. In our system, the brush layer is constituted by poly(styrene d8-r-methyl methacrylate) statistical copolymers with different molecular weights grafted to a silicon substrate. The thin film is modelled through poly(styrene-r-methyl methacrylate) statistical copolymers deposited over the grafted substrate. We simulate the grafting-degrafting mechanism of polymer chains from the silicon substrate, showing how the reactivity of the system is related to the degree of stretching of the brush layer induced by the entering of additional polymer chains.

1 Introduction

Polymer brushes offer nowadays inspiring possibilites in a large variety of different applications, from the stabilization of colloidal dispersions [1] to the control of adhesion and wetting. [2] When coupled with polymer thin films, they can be also used for innovative optical applications. [3] In this context, a proper control of microscopic mechanisms leading a polymer chain to graft a solid surface is therefore of utmost importance. In particular, recent studies have provided evidence of a continuous exchange between grafted and incoming chains, which affect the chemical composition of the brush layer. [4] This is in contrast with the more traditional view of the grafting process as an almost self-limited reaction, where the brush thickness reaches a pseudo-plateau after a sufficient amount of time.

In this work we aim to shed light on the mechanisms underlying the "grafting to" reactions by simulating a model constituted by a silica wafer grafted by poly(styrene d8-r-methyl methacrylate) (PMMA) statistical copolymers. A random copolymer thin film constituted by the same kind of polymer chains is simulated at the interface with the brush layer. The study is performed by means of the hybrid particle-field molecular dynamics representation (MD-SCF), [5] which allows to obtain a fast relaxation of polymer chains. [6] In order to obtain accessible simulation times for a proper calculation of structural properties of polymer chains, needed to calculate the brush height and the chain stretching, we thus make use of CRESCO supercomputing resources.

2 Simulation method

n this work, hybrid particle-field molecular dynamics technique was employed to investigate in detail the effect of chain stretching during the grafting-ungrafting process. Proper coarse-grained (CG) models representing the substrate and the polymer chains were developed. According to this approach, each bead interacts with the other beads in the simulation box through an effective potential $V(r)$ given by a density field due to the presence of all other beads [5]. The mean field solution for the potential acting on a particle of type K at position r, $V(r)$ is:

*Corresponding author. E-mail: gmunao@gmail.com

$$V_k(r) = k_B T \sum_{k'} \chi_{kk'} \phi_{k'}(r) + \frac{1}{\kappa}\left(\sum_k \phi_k(r) - 1\right) \quad (1)$$

where k_B is the Boltzmann constant, T is the temperature, $\chi_{KK'}$ are the mean field parameters for the interaction of a particle of type *K* with the density field due to particles of type *K'* and the second term on the right-hand side of Eq. 1 is the incompressibility condition, κ being the compressibility. Also, $\Phi_K(\mathbf{r})$ and $\Phi_{K'}(\mathbf{r})$ are the density functions of the CG beads of type *K* and *K'*, respectively. We simulate a planar silica substrate constituted by 18720 beads, each one representing five SiO_2 groups. These beads are arranged according to a hexagonal configuration, and the lattice distance has been fixed to 0.7 nm. The silica substrate is composed by eight connected layers, with a total thickness of 4.9 nm. Copolymer chains (whose molecular weight has been se to M_n = 19500 g/mol) have been randomly bonded on the outermost substrate layer. More specifically, a deuterated copolymer chain is constituted by a random sequence of 33 beads of two different types, which represent five PMMA or six PS repeating units, respectively. The hydrogenated copolymer chains are constituted by three (M_n = 1700 g/mol) to 65 beads (M_n = 38600 g/mol). In addition, each vacuum particle has a molecular weight of 460.70 g/mol , in agreement with previous studies [7, 8].

3 Results

In order to investigate the local arrangement of polymer chains during the grafting-degrafting process, we have simulated the two extreme cases of a single polymer chain with a molecular weight of 38600 (labeled as R38.6) or 1700 g/mol (labeled as R1.7) entering the brush layer. The latter is constituted by chains of 19500 g/mol (labeled as Rd19.5) with a layer thickness of 9.0 nm. With the aim to shed light on the chain stretching in these two different situations, we have computed the end-to-end (ete) distances for the Rd19.5 chains in the brush layer that are located far from or adjacent to an incoming R1.7 or R38.6 polymer chain. The Rd19.5 chains located far from the single chain that enters the brush are referred to as unperturbed chains, since their spatial arrangement is not perturbed by the incoming chain. Instead, the Rd19.5 chains close to the incoming chain are indicated as adjacent chains. The chain stretching can be documented by computing the ete distance of unperturbed and adjacent chains. The probability distribution of the ete distance normalized by the ideal value for the unperturbed and adjacent Rd19.5 chains, when the hydrogenated R38.6 chain is grafted to the substrate and when an adjacent chain of Rd19.5 degrafts from the substrate, is reported in the top panel of Fig. 1, respectively. In Fig. 1a we observe that, when a single chain of R38.6 grafts to the brush layer of Rd19.5, the ete distribution of adjacent chains moves toward significantly higher values than the corresponding ete distribution of the unperturbed chains. This circumstance indicates the existence of a substantial stretching. In a similar way, Fig. 1b shows that, upon degrafting of one of highly stretched chains, the ete distributions for adjacent and unperturbed chains appear quite similar with values close to the ideal ones. This finding demonstrates a remarkable reduction of the chain stretching. A similar scenario is shown in the bottom panel of Fig. 1, which shows the probability distribution of the ete distance (normalized by the ideal value for the unperturbed and adjacent Rd19.5 chains) when the hydrogenated R1.7 chain is grafted to the substrate and when an adjacent Rd19.5 chain degrafts from the substrate, respectively. In the grafting phase, the probability distribution of the ete distance of the adjacent R19.5 chains is placed at slightly higher values than the one of the unperturbed chains. This indicates a slight chain stretching.

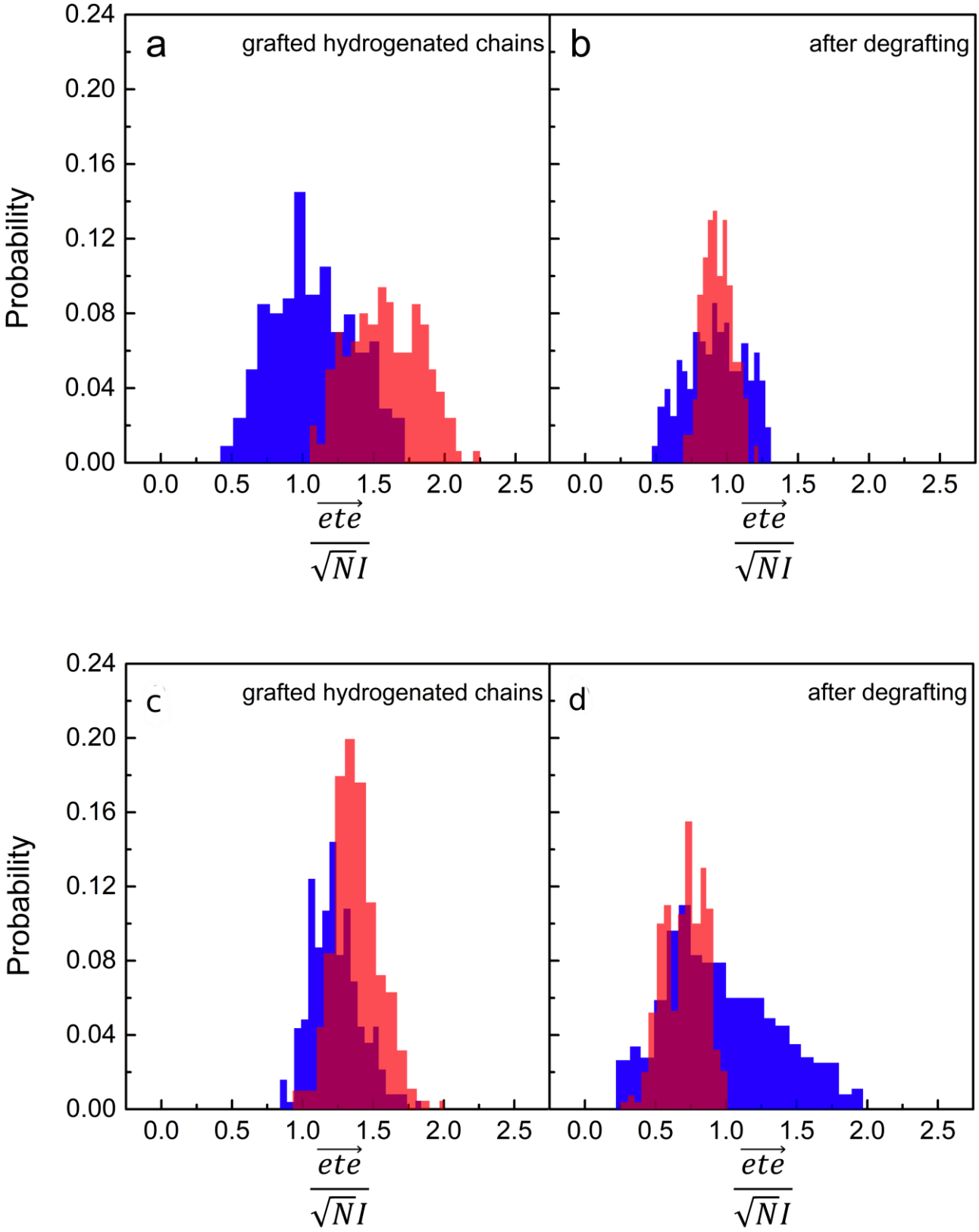


Fig. 1: Probability distribution of the end-to-end distance normalized by the ideal value for unperturbed (blue) and adjacent (red) grafted chains for R38.6 (left panel) and R1.7 (right panel) in the grafting stage (a,c) and after degrafting (b,d).

Conversely, after degrafting of a chain of Rd19.5, chain relaxation occurs, in agreement with the shift of the ete distribution of the adjacent chains toward lower values. In this frame, the unperturbed chain distribution covers a wider range of values since the number of grafted chains (271) is much higher than that of adjacent chains (4), and therefore many different configurations are possible.
The investigation of the grafting-degrafting mechanism with such a high detail has been possible thanks to the CRESCO Supercomputing resources, and for such an aim we have performed extensive molecular simulations on the cresco queues cresco6 h144, cresco4 h144 and small h144.

References

[1] T. Kreer. Polymer-brush lubrication: A review of recent theoretical advances. *Soft Matter*, **12**, pp. 3479–3501, (2016).
[2] C. Ligoure. Adhesion between a polymer brush and an elastomer: A self-consistent mean field model. *Macromolecules*, **29**, pp. 5459–5468, (1996).
[3] M. Stefik, S. Guldin, S. Vignolini, U. Wiesner, and U. Steiner. Block copolymer self-assembly for nanophotonics. *Chem. Soc. Rev.*, **44**, pp. 5076–5091, (2015).
[4] L. Wan, R. Ruiz, H. Gao, and T. Albrecht. Self-registered self-assembly of block copolymers. *ACS Nano*, **11**, pp. 7666–7673, (2017).
[5] G. Milano and T. Kawakatsu. Hybrid particle-field molecular dynamics simulations for dense polymer systems. *J. Chem. Phys.*, **130**, pp. 214106, (2009).
[6] G. Munaò, A. Pizzirusso, A. Kalogirou, A. De Nicola, T. Kawakatsu, F. Müller-Plathe, and G. Milano. Molecular structure and multi-body interactions in silica-polystyrene nanocomposites. *Nanoscale*, **10**, pp. 21656, (2018).
[7] K. Sparnacci, R. Chiarcos, V. Gianotti, M. Laus, T. J. Giammaria, M. Perego, G. Munaò, G. Milano, A. De Nicola, M. Haese, L. P. Kreuzer, T. Widmann, and P. Müller-Buschbaum. Effect of trapped solvent on the interface between ps-b-pmma thin films and p(s-r-mma) brush layers. ACS Appl. Mater. Interfaces, **12**, pp. 7777–7787, (2020).
[8] M. Laus, R. Chiarcos, V. Gianotti, D. Antonioli, K. Sparnacci, G. Munaò, G. Milano, A. De Nicola, and M. Perego. Evidence of mechanochemical control in “grafting to” reactions of hydroxy-terminated statistical copolymers. *Macromolecules*, **54**, pp. 499–508, (2021).

ENHANCING CFD SIMULATIONS OF COVID-19 DIFFUSION BY COUGHING AND SNEEZING USING DATA ASSIMILATION

Rossella Arcucci[1,3,*], César Quilodrán Casas[1], Aniket Joshi[2], Laetitia Mottet[2], Asiri Obeysekara[2], Yi-Ke Guo[1,4] and Christopher Pain[2]

[1]*Data Science Institute, Department of Computing, Imperial College London, UK*

[2]*Department of Earth Science and Engineering, Imperial College London, UK*

[3]*Leonardo Centre, Imperial College Business School, Imperial College London, UK*

[4]*Department of Computer Science, Hong Kong Baptist University, Hong Kong*

ABSTRACT. Coughing is one of the most effective methods for SARS-CoV-2, the coronavirus strain that causes COVID-19, to spread. Coughing is a natural reaction that serves to protect the lungs and airways from irritants and infections by expelling droplets at speeds of up to 50 miles per hour. Unfortunately, it's also one of the most efficient methods for infections to spread, particularly respiratory viruses that require host cells to replicate. CFD (Computational Fluid Dynamics) is an useful tool for simulating droplets ejected by the mouth and nose during coughing and sneezing. Coughing and sneezing models, like any numerical models, add uncertainty through the choice of scales and parameters. Any numerical simulation must take these uncertainties into account in order to be accepted. In the medium to long-term analysis, numerical forecasting models frequently use Data Assimilation (DA) approaches for uncertainty quantification. DA is the ideal combination of time-distributed data with a dynamic model to approximate the true state of a physical system at a particular time. In order to improve numerically forecast results, DA adds observational data into a prediction model. We use a Variational Data Assimilation model to assimilate direct observation of the physical mechanics of droplet generation at the mouth's exit during coughing in this research. We employ high-speed imaging to analyse the fluid fragmentation at the exit of the lips of healthy participants in a sneezing scenario, which we learned about from previous studies. We demonstrate the impact of the suggested approach on the accuracy of CFD simulations.

1 Introduction

Infections of the respiratory tract, like as influenza, are transmitted when a healthy person comes into touch with respiratory droplets from an infected person's cough, sneeze, or breath [1]. Coughing is one of the most effective methods for SARS-CoV-2, the coronavirus strain that causes COVID-19, to spread. Coughing is a natural reaction that protects the lungs and airway from irritants and germs by expelling droplets at speeds of up to 50 miles per hour. Unfortunately, it's also one of the most effective methods for infections to spread, particularly for respiratory viruses that require host cells to proliferate. CFD (Computational Fluid Dynamics) is an useful tool for simulating droplets ejected by the mouth and nose while coughing or sneezing. CFD models for coughing and sneezing, like any numerical models, introduce uncertainty through the choice of scales and parameters. Any numerical simulation must take these uncertainties into account in order to be accepted. In the medium to long-term analysis, numerical forecasting models frequently use Data Assimilation (DA) approaches for uncertainty quantification.

DA is the ideal combination of time-distributed data with a dynamic model to approximate the true state of a physical system at a particular time. In order to improve numerically forecasted results, DA adds observational data into a prediction model. It allows for problems like redundancy and unequal spatial and temporal data distribution to be handled, allowing models to assimilate information more efficiently.

[*]Corresponding author. E-mail: r.arcucci@imperial.ac.uk

"What can be said about the value of an unknown variable x that characterises the evolution of a system if we have some measured data y and a model M of the underlying mechanism that created the data?" DA seeks to answer questions like this. This is the Bayesian setting, in which we seek a quantification of the uncertainty in our parameter information, which, according to Bayes' rule takes the form:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Here, the physical model is represented by the conditional probability (also known as the likelihood) p(y | x), and the prior knowledge of the system by the term p(x). The denominator is considered as a normalising factor and represents the total probability of y. Many DA approaches have been created based on this formulation [2], with the majority of them being custom-built for the forecasting model with which they are integrated. The variational DA (VarDA) approaches [3] based on the minimisation of a function that estimates the discrepancy between numerical results and observations, assuming that the two sources of information, forecast and observations, have errors that are adequately described by error covariance matrices, have gained acceptance as powerful methods in the last ten years. In order to apply a DA approach to a CFD model for coughing and sneezing, real observations are needed.

1.1 Related works and contribution of the present work

The travel characteristics of evaporating droplets discharged into the vented chamber were investigated using a CFD study with an Eulerian-Lagrangian model in [4]. With the use of Bayesian Data Assimilation, this study tries to explain the transport and dispersal of droplets created by coughing in a ventilated environment. The beginning velocity and duration of a coughing burst were measured in experiments. Instead of a CFD study, [5] proposes an analytical technique. To explore the dispersion and deposition of expiratory droplets in a room during coughing, the authors analyse the detailed processes of cough jet flow, including droplet evaporation and motion, turbulent flow surrounding jet, and particle tracking. The authors report the results of a combined experimental and theoretical examination of the fluid dynamics of such violent expiratory events in [6]. Sneezing and coughing events are multiphase turbulent buoyant clouds with suspended droplets of varying sizes, according to direct observation. The creation of a theoretical model of pathogen-bearing droplets interacting with a turbulent buoyant momentum puff is guided by observations. The transport characteristics of saliva droplets produced by coughing are investigated in [7] in a peaceful indoor setting. The Lagrangian equation is used to investigate the dispersion processes of saliva droplets of various sizes expelled while coughing. The findings show that the size affects the transport characteristics of saliva droplets caused by coughing.

The authors disclose firsthand observation of the physical mechanics of droplet generation at the mouth's outflow during sneezing in [8]. They use high-speed imaging to analyse the fluid fragmentation at the mouth exits of healthy people in particular. In this paper, we use the genuine photos from [8] to improve the CFD models of coughing and sneezing. We used a 3D Variational DA model with an ideal parameter to balance the weight of the errors covariance matrices to achieve this goal.

In summary, we employ data assimilation to improve the accuracy of the CFD models to simulate droplet and aerosol size distributions [10] using real-world data. This will provide us more precise information on the evolution of the particle size distribution around the mouth.

2 Data Assimilation

Data Assimilation (DA) is an approach for fusing data (observations) with prior knowledge (e.g., mathematical representations of physical laws; model output) to obtain an estimate of the distribution of the true state of a process [9].

In order to perform DA, one needs observations (i.e., a data or measurement model), a background (i.e., a priori state or process model), and information about the distribution of the errors on these two.

DA merges the estimated state $x_t \in R^n$ of a discrete-time dynamic process at time t:

$$x_{t+1} = M_{t+1}x_t + w_t$$

with an observation $y_t \in R^m$:

$$y_t = H_t x_t + v_t$$

where M_t is a dynamic system and H_t is called observation operator. The vectors w_t and v_t represent the process and observation errors, respectively. They are usually assumed to be independent white-noise processes with Gaussian probability distributions.

$$w_t \sim \mathcal{N}(0, Q_t), \quad v_t \sim \mathcal{N}(0, R_t)$$

where Q_t and R_t are called error covariance matrices of the model and observation respectively.
DA is a Bayesian inference that combines the state x_t with y_t at each given time. The Bayes theorem conducts to the estimation of x_t^a which maximise a probability density function.
given the observation y_t and a prior from x_t. This approach is implemented in one of the most popular DA methods which is the three-dimensional Variational (3DVar) DA. The goal of 3DVar is to compute an optimal solution, x_t^a, that minimises a weighted difference between the actual measurement, y_t, and the measurement prediction.
If the error covariance matrices Q_t and R_t are designed to be correlated by a parameter α and such that $Q_t = \alpha I$ and $R_t = (1 - \alpha)I$ with $0 < \alpha < 1$ and where I is the identity matrix [10], we can decide the degree of fidelity we want to give to the observations with respect to the CFD simulation by setting a proper values of the parameter α. As the weight of the covariance matrices in the DA process is given by the inverse of the matrices [2], with this choice of covariance matrices we can chose a bigger value of α if we assume that the observations are very reliable or a smaller value to α if the CFD model is a high-fidelity model. With this choice the data assimilation process can be described as following:

$$\delta x_t^a = argmin_{\delta x} J(\delta x)$$

with

$$J(\delta x) = \frac{\delta x^T \delta x}{2\alpha} + \frac{(H_t \delta x - d_t)^T (H_t \delta x - d_t)}{2(1-\alpha)}$$

where $\mathrm{d_t} = [\mathrm{y_t} - \mathrm{x_t}]$ is the misfit and $\delta \mathrm{x} = \mathrm{x} - \mathrm{x_t}$ is the increment.

As an important issue in Data Assimilation is to provide a result in real-time, the choice of an efficient method to compute the minimum of the functional J is a fundamental topic.
In this paper, we compute the minimum of the functional J by the minimisation method proven to be faster for optimisation problems [11], i.e., the L-BFGS (Limited-Broyden Fletcher Goldfarb Shanno) method. The L-BFGS method is a Quasi-Newton method that can be viewed as extension of conjugate-gradient methods in which the addition of some modest storage serves to accelerate the convergence rate.

3 Test case

The test case represents an idealised case used to test the ability of our 3DVar to be used to assimilate real images of sneeze or cough emissions in coughing and sneezing CFD models. The images of real sneeze emissions were obtained from [8]. In order to assimilate sneezing images with a coughing CFD, we are here assuming that, in terms of velocity and mass fraction, the sneeze at later time step corresponds to a cough at earlier time step. The data was pre-processed using OpenCV. OpenCV (Open Source Computer Vision Library) is an "open source computer vision and machine learning software library built to provide a common infrastructure for computer vision applications" [12].
Two images of real observations were used: sneeze emissions after 5 ms recorded at 2000 fps; and sneeze ejecta at 8 ms recorded at 8000 fps. The images represent the observed data in the DA function. The background data from the coughing CFD model are two images of simulated sneeze emissions: at an angle of 24 degrees and horizontal. This data represents the state vector in the DA function. The images from the CFD simulation were scaled between 0 and 0.01 (for sneezing) and 0 and 0.02 (for coughing) with respect to their water mass fraction. The images from the simulations were cropped between 1.53 m and 1.63 m of height and a 0.12 m width. After this pre-process, the dimensions of the

images from the CFD simulation and the observations match. All images were set at the same resolution and the observations (interpolation) operator H_t is the identity function. The observation images were also scaled between 0 and 0.01 (for sneezing) and 0 and 0.02 (for coughing) with respect to their water mass fraction to be consistent to the simulation images.
The four images in were transformed to grayscale (ranged from 0 to 255 in one channel) using OpenCV. Additionally, a mask was drawn by hand on all images to eliminate the nose and mouth from the observations, and the inlets from the simulations. We chose a non-grayscale colour (blue) for the mask to avoid eliminating useful information. The mask allows us to perform the data assimilation only on the sneeze emissions and ejecta [10]. The backgrounds of all four images were set to white. Since the observations do not include a water mass fraction associated to them, we assumed a value of 0.01 (for sneezing) and 0.02 (for coughing) for pixels in black and 0 for pixels in white, to complement the simulations. The execution time of the algorithm for assimilating y in x is approx. 0.38 to 0.42 seconds. These values of the execution times have been computed as mean values of 50 runs of the algorithm on the same machine and the same data set and for different values of $0<\alpha<1$. Table 1 shows value of Mean Square Error (MSE) defined as

$$MSE(x) = \frac{\|x - x_C\|_2}{\|x_C\|_2}$$

where x_C denotes a control variable. The MSE is here computed with respect to the observed data before and after the assimilation process for different values of the parameter α. As expected, for bigger values of α, the result of the assimilation come closer to the observations, and it presents a smaller value of MSE.

Table 1: Values of MSE, for the sneezing CFD simulations, computed with respect to the observed data before and after the assimilation process for different values of the parameter α.

α	**couching**	**sneezing**
0.1	0.94	1.68
0.2	0.89	1.60
0.3	0.83	1.48
0.4	0.74	1.32
0.5	0.62	1.12
0.6	0.49	0.88
0.7	0.35	0.63
0.8	0.22	0.40
0.9	0.10	0.18

Fig. 1 shows how the DA technology merges the two data for different values of $0< \alpha <1$. The results confirm that for small values of α the solution x^a of the assimilation process is closer to the CFD simulation. The assimilation of these data could have a significant impact on real-world applications for determining safe distances. In the scenario at hand, the CFD estimates that the dispersion of the droplets after 5 ms has a radius of about 4 cm and a distance of nearly 6 cm from the mouth. The observation shows a smaller radius near the mouth, but the droplets reach 12 cm from the mouth. The combination of this data is critical in determining safe distances for human interactions. In fact, the technology and model we gave are generic and may be used to simulate other scenarios using other types of computational fluid dynamic systems.

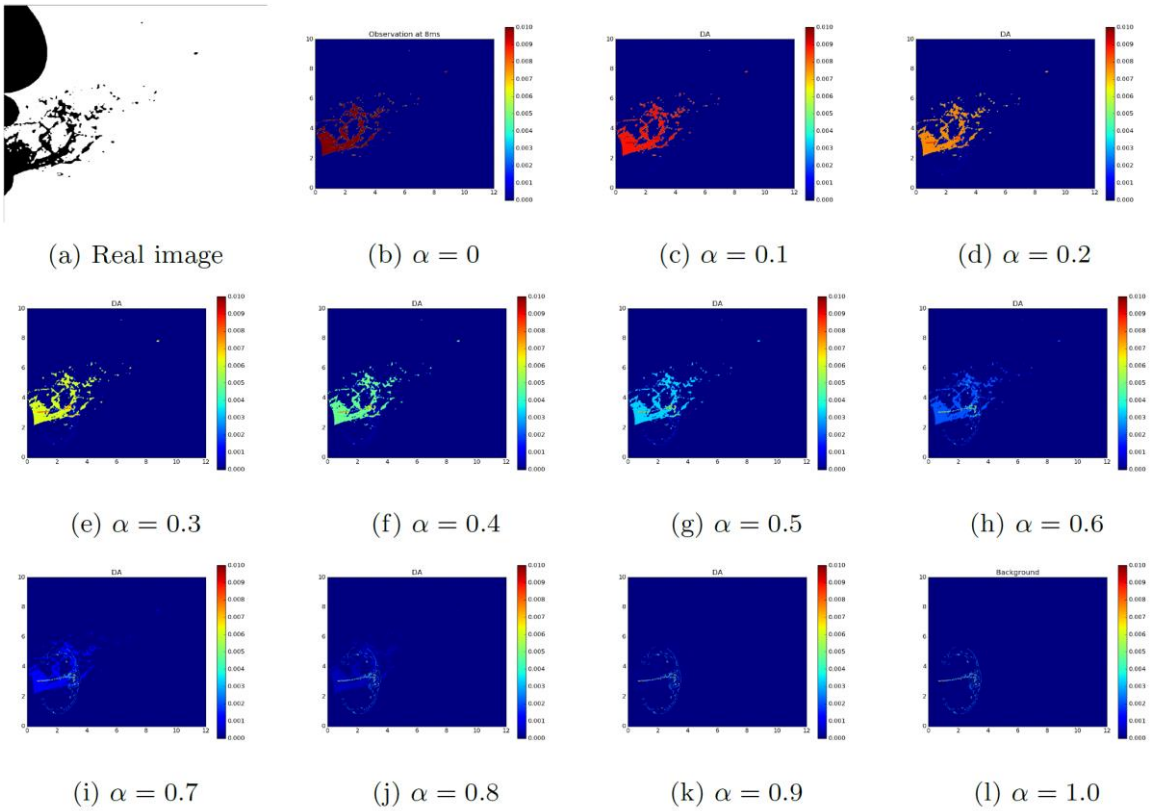


Fig.1: Results of the assimilation of the real image y in the CFD coughing simulation x for different values of α.

4 Conclusions and future work

In this study, we compare experimental data from [8] of sneezing with results from a coughing CFD simulation with various inlet geometries. We've only considered the mouth as a source of ejecta in our first test case. We also hypothesised that comparing a sneeze simulation to a cough simulation at a later timestep would yield equivalent and interchangeable findings, allowing us to assess the adaptability of the DA approaches. We used a 3D Variational DA model with an optimal parameter to balance the weight of the error covariance matrices in the assimilation function.

The adaptability of DA in efficiently using the experimental sneezing results in the coughing CFD simulation highlights the benefits of this work. This adaptability may be used to create a comprehensive technique that can aid in the modelling of various COVID-19 spread situations that incorporate ventilation airflows (indoors or outdoors), while also leaving room for modelling crowd airflow dynamics.

The CFD simulations in domain such as a room aim to show the dispersion in real case scenarios. These simulations need High Performance Computing infrastructure to run. The computing resources and the related technical support used for this future work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [13]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

Acknowledgements

This work is supported by the EP/V036777/1 Risk EvaLuatIon fAst iNtelligent Tool (RELIANT) for COVID19.

References

[1] Mahesh Jayaweera, Hasini Perera, Buddhika Gunawardana, and Jagath Manatunge. Transmission of covid-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy. Environmental Research, page 109819, 2020.

[2] Mark Asch, Marc Bocquet, and Maffelle Nodet. Data assimilation: methods, algorithms, and applications, volume 11. SIAM, 2016.

[3] Rossella Arcucci, Laetitia Mottet, Christopher Pain, and Yi-Ke Guo. Optimal reduced space for variational data assimilation. Journal of Computational Physics, 379:51-69, 2019.

[4] Wei Sun and Jie Ji. Transport of droplets expelled by coughing in ventilated rooms. Indoor and Built Environment, 16(6):493-504, 2007.

[5] Jianjian Wei and Yuguo Li. Enhanced spread of expiratory droplets by turbulence in a cough jet. Building and Environment, 93:86-96, 2015.

[6] Lydia Bourouiba, Eline Dehandschoewercker, and John WM Bush. Violent expiratory events: on coughing and sneezing. Journal of Fluid Mechanics, 745:537-563, 2014.

[7] Shengwei Zhu, Shinsuke Kato, and Jeong-Hoon Yang. Study on transport characteristics of saliva droplets produced by coughing in a calm indoor environment. Building and environment, 41(12):1691-1702, 2006.

[8] BE Scharfman, AH Techet, JWM Bush, and L Bourouiba. Visualization of sneeze ejecta: steps of fluid fragmentation leading to respiratory droplets. Experiments in Fluids, 57(2):24, 2016.

[9] Christopher K Wikle and L Mark Berliner. A bayesian tutorial for data assimilation. Physica D: Nonlinear Phenomena, 230(1-2):1-16, 2007.

[10] Rossella Arcucci, César Quilodrán Casas, Aniket Josh, Asiri Obeysekara, Laetitia Mottet, Yi-Ke Guo, Christopher Pain - Merging Real Images with Physics Simulations via Data Assimilation- EuroPar 2021, Lecture Notes in Computer Science (in print)

[11] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. Mathematical programming, 45(1-3):503-528, 1989.

[12] G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.

[13] F. Iannone et al., "CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout," 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, 2019, pp. 1051-1052, doi: 10.1109/HPCS48598.2019.9188135.

DYNAMICS OF INTERACTION OF TUMOR HOMING PEPTIDES AND THEIR RECEPTORS BY MOLECULAR DYNAMICS SIMULATIONS

Maurizio S. Podda[1], Chiara Lico[2], Selene Baschieri[2], Mariateresa Mancuso[3] and Caterina Arcangeli[1*]

[1]*Laboratory of Health and Environment. Department of Sustainability, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rome, Italy*

[2]*Laboratory of Biotechnology. Department of Sustainability, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rome, Italy*

[3]*Laboratory of Biomedical Technologies. Department of Sustainability, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rome, Italy*

ABSTRACT. In this report, part of the results published by Lico et al. [1] and obtained during the third year of activity of NANOCROSS project will be presented. In particular, the dynamics of interaction, by means of molecular dynamics (MD) simulations, of two specific tumor homing (TH) peptides, suitable for virus-derived nanoparticle (NP) functionalization, and two tumor receptors (TRs) is here reported. Simulations were used to assess whether the display of TH peptides on the surface of the plant virus NP can affect their ability to recognize and bind with good affinity the TR targets. All the simulations were performed on CRESCO6 infrastructure using GROMACS 2019 package.

1 Introduction

This work was carried out as part of NANOCROSS project, "Plant virus nanoparticles for blood-brain barrier crossing and medulloblastoma targeting" (IG 20314), granted by the "Fondazione AIRC per la Ricerca sul Cancro" to M. Mancuso. The ultimate goal of the project is the development of a theranostic platform. The platform will be composed of plant-derived viral NP genetically modified to display on their surface specific peptides able to cross the blood-brain barrier and reach the associated molecular targets on medulloblastoma. The candidate TH peptides for the functionalization of the theranostic NP-based platform have been selected and widely characterized in previous works [2,3]. Here we report the simulation of the molecular interaction between such peptides and their specific receptors. The effect of peptides conjugation on the surface of a plant-derived virus NP on their ability to recognize and bind the tumor brain-associated receptors has also been investigated. The simulations evidenced similar mechanisms in terms of recognition of the tumor-associated receptors also when the virus NP is functionalized with the peptides.

2 Computational methods

All the MD simulations were performed by using Gromacs 2019 code [4] with all-atom AMBER 99sb-ildn force field [5] for receptor and peptides in combination with TIP3P water model [6]. Periodic boundary conditions were applied and all the systems were energy minimized, gradually heated and equilibrated (details of the equilibration procedure are provided in Table 1).
Unrestrained MD were carried out in NPT ensemble. In Table 2 details of the simulations are reported. All the simulations were performed through the CRESCO/ENEAGRID High Performance Computing infrastructure [7].

*Corresponding author. E-mail: caterina.arcangeli@enea.it

Table 1: Equilibration Procedure. a = simulation time (ps); b = timestep (fs); c = temperature (K); d=positionally constraint force (kJ/mol^{-1}/nm^{-2}); e = temperature coupling constant (ps); f = pressure bath coupling (ps) ; g = τ_t for the solute; h = τ_t for the solvent and ions.

Time[a]	Δt[b]	T[c]	Fc (x10^3)[d]	τ_t[e]		τ_p[f]
2	1	50	100	0.01[g]	0.01[h]	-
4	1	100	100	0.01	0.01	-
6	1	150	100	0.01	0.01	-
8	1	200	100	0.03	0.01	-
10	1	250	100	0.03	0.01	-
12	1	300	100	0.03	0.01	-
100	2	300	30	0.03	0.01	-
150	2	300	10	0.1	0.1	-
200	2	300	3	0.1	0.1	-
400	2	300	-	0.1	0.1	2.0

Table 2: Simulation details

Simulation details	
Temperature	300 K
Neutralizating ions	NaCl (150 mM)
Dielectric constant	1
Integration time step	2 fs
Long-range electrostatic force	PME method
Short-range electrostatic and van der Waals cutoff	1 nm
Lenght of simulation	100 ns

3 Results and discussion

In order to assess the effect of conjugation of peptides to the surface of a plant-derived virus NP, six different peptide-target complexes have been prepared (Table 3). In three of them (a, b, c), the selected and well characterized TH_pep3 and TH_pep4 [2,3] are complexed with three different brain tumor receptors, namely TR1 (PDB id: 5IYY, [8]), TR2 (PDB id: 5CE4, [9]) and TR3 (PDB id: 1FE3 [10]). In the remaining three systems (d, e, f) the TRs are complexed with the same peptides conjugated to the surface of the virus NP. The starting structures for the MD simulations of the six complexes are shown in Figure 1.

MD simulations were performed to investigate the dynamics of interaction within the binding sites of the TRs and to monitor the structural stability of the peptide-TR complexes.

Table 3: Peptide-Target Complexes used for the MD simulations. Details on the docking procedure used to obtain the complexes are provided in reference [1]

Tumor Receptors (TR) in complex with free peptides (TH pep)	Tumor Receptors (TR) in complex with peptides conjugated to viral NP (NP_TH-pep)
a) TH-pep4 – TR1	d) NP_TH-pep4 – TR1
b) TH-pep3 – TR2	e) NP_TH-pep3 – TR2
c) TH-pep3 – TR3	f) NP_TH-pep3 – TR3

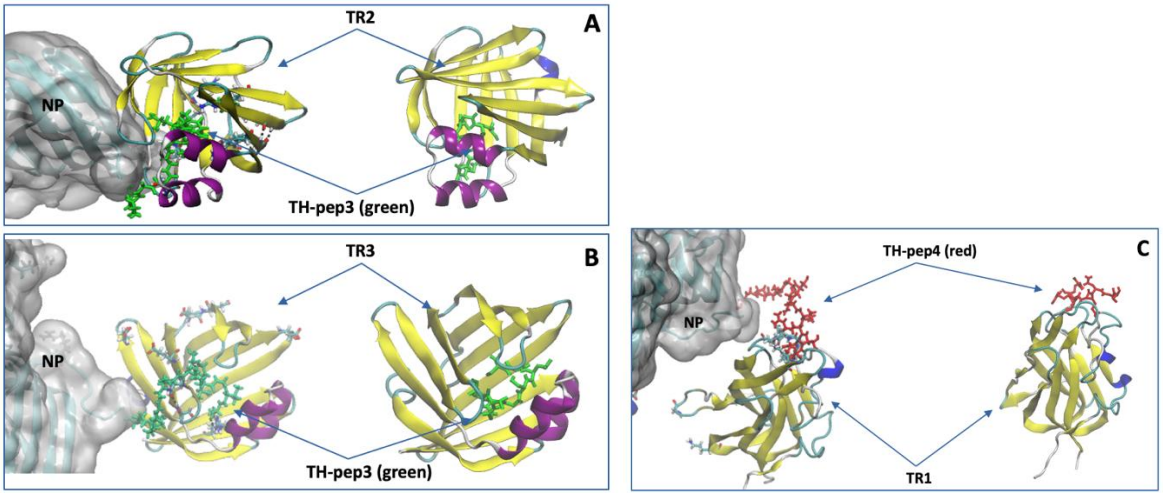


Fig. 1: Starting structures for the MD simulations of the peptide—receptor complexes. A) TH-pep3 peptide in complex with TR2 (right) and after NP functionalization (left). B) TH-pep3 peptide in complex with TR3 (right) and after NP functionalization (left). C) TH-pep4 peptide in complex with TR1 (right) and after NP functionalization (left).

The root mean square deviations (RMSDs) of the Cα atoms of the ligand-target complexes from their initial values were calculated as a function of simulation time and shown in Figure 2. The RMSD trend measure the stability of the complexes and to monitor their conformational sampling during the MD simulation and are shown in Figure 2. The RMSDs values of TH-pep4 —TR1 complex (Figure 2A) indicate that after about 60 ns of simulation the complex undergoes structural instability, whereas the RMSD values of TH-pep3 —TR2 complex (Figure 2B) seem to reach a plateau within 20 ns of simulation, suggesting a conformational stability. The RMSD trend registered for the TH-pep3—TR3 seems to suggest that this complex sample many conformational states (Figure 2C). The anchoring of both TH-pep4 and TH-pep3 peptides to the surface of the virus NP seems to improve the structural stability of all the complexes, as suggested by their RMSD time evolution (Figure 2D-F).
The dynamics of interaction can be appreciated by a visual inspection of the molecular configurations sampled by the TH peptides inside the binding site of the TR. For this reason, selected simulation snapshots, taken at selected time of MD trajectories, are also shown in Figure 2. The behaviour of TH-pep4 peptide inside the binding site of the TR1 receptor can be divided in three phases: initially the peptide stably binds the receptor, then after 60 ns it starts to move inside the binding site and after about 70 ns of the simulation it moves out from the active site of the receptor (Figure 2A). When the same peptide is anchored to the surface of the viral NP, a number of interactions between the residues of the TR1 receptor with those of viral NP contribute to stabilize the complex (Figure 2D). The dynamics of interaction between TH-pep3 peptide and TR2 receptor (Figure 2B) and with the TR3 receptor (Figure 2C) seems to be very stable, although the TH-pep3—TR3 complex appear less stable if compared to

TH-pep3—TR2 complex. Again, the TH-pep3 is able to bind both the TR2 (Figure 2E) and TR3 (Figure 2F) receptors also after the functionalization on the viral NP surface.

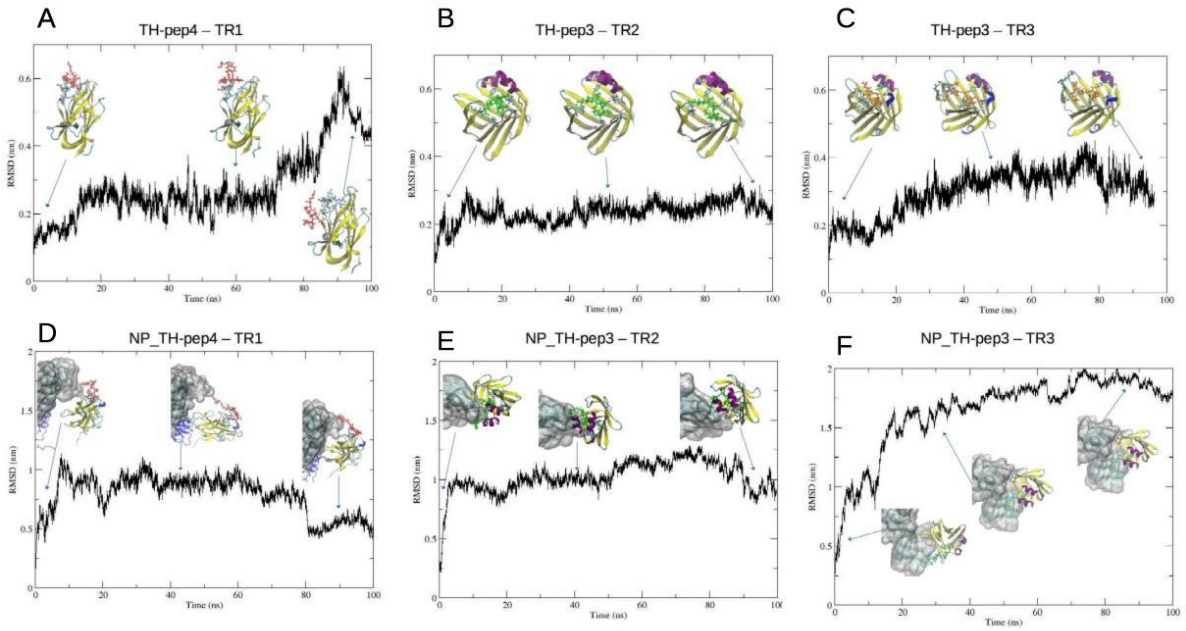


Fig. 2: Cα RMSD values with respect to the equilibrated conformations, as a function of simulation time, of the complexes in which the TH peptides are free (upper plots) and of the complexes in which the same peptides are conjugated to the surface of virus NP (lower plots). Simulation snapshots, taken at selected time of the MD trajectories, are also shown.

4 Conclusion

The NANOCROSS project aims at creating a theranostic platform capable of providing diagnostic data and therapeutics effects. The platform was thought to avoid side effects and preventing problems similar to those caused by classical therapies. In previous works [2,3] we selected, by means of a deep structural and dynamical characterization by MD simulations, two specific peptides (TH-pep3 and TH-pep4) which are known to cross the blood-brain barrier and reach the associated molecular brain tumor targets. In the present work, we have successfully tested, by means of computational techniques, the ability of these two TH peptides to recognize and bind molecular receptors, normally over expressed in brain tumors.

The outcomes obtained suggest that all the selected peptides are able to bind the tumor receptors also after the functionalization of the surface of the viral NP. Our results have provided a valid predictive contribution to the design of viral NPs before starting the complex experimental production of the NANOCROSS therapeutic platform. It would be interesting to extend this type of analysis to other peptides selected within the NANOCROSS project to evaluate the effectiveness in recognizing other medulloblastoma molecular targets.

Acknowledgments

This work was supported by the NANOCROSS project “Plant virus nanoparticles for blood-brain barrier crossing and medulloblastoma targeting” (IG 20314) granted by the “Fondazione AIRC per la Ricerca sul Cancro” to M. Mancuso. We acknowledge the ENEA-HPC team for supporting our computational activities in the ENEAGRID infrastructure and on the CRESCO high performance platform.

References

[1] Lico C. Tanno, B., Marchetti, L., Novelli, F., Giardullo P., Arcangeli C., Pazzaglia S., Podda M.S., Santi L., Bernini R., Baschieri S., Mancuso M. Tomato Bushy Stunt Virus Nanoparticles as a platform for drug to Shh-dependent medulloblastoma Submitted to *Int J Mol Sci*, (2021).

[2] Arcangeli, C., Lico, C., Baschieri, S., and Mancuso, M. Characterization Of Blood–Brain Barrier Crossing And Tumor Homing Peptides By Molecular Dynamics Simulations. *Int J Nanomed* **14**, pp 10123–10136, (2019).

[3] Arcangeli, C., Lico, C., Baschieri, S., Mancuso, M. Molecular Dynamics Simulations of Peptides for Brain Theranostic Nanosystems, in: "High Performance Computing on CRESCO Infrastructure: Research Activites and Results 2018." p. 155, (2019)

[4] Van Der Spoel, David, et al. "GROMACS: Fast, Flexible, and Free." *J Comput Chem*, **26(16)**, pp. 1701–1718, (2005).

[5] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **78**, pp. 1950–1958, . (2010).

[6] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, pp 926–935, (1983).

[7] F. Iannone et al., "CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout," 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, 2019, pp. 1051-1052, doi: 10.1109/HPCS48598.2019.9188135.

[8] Mota, F., Fotinou, C., Rana, R.R., Edith Chan, A.W., Yelland, T., Arooz, M.T., O'Leary, A.P., Hutton, J., Frankel, P., Zachary, I., et al. Architecture and hydration of the arginine-binding site of neuropilin-1. *The FEBS Journal* **285**, pp 1290–1304, (2018).

[9] Howard, E.I., Guillot, B., Blakeley, M.P., Haertlein, M., Moulin, M., Mitschler, A., Cousido-Siah, A., Fadel, F., Valsecchi, W.M., Tomizaki, T., et al. High-resolution neutron and X-ray diffraction room-temperature studies of an H-FABP-oleic acid complex: study of the internal water cluster and ligand binding by a transferred multipolar electron-density distribution. *IUCrJ* **3**, pp 115–126 , (2015).

[10] Balendiran, G. K. "Crystal Structure and Thermodynamic Analysis of Human Brain Fatty Acid Binding Protein." *Journal of Biological Chemistry*, 14 June (2000), 10.1074/jbc.m003001200.

GENERALIZED OPTIMAL PATHS AND WEIGHT DISTRIBUTIONS IN COMPLEX NETWORKS

Ricardo Gutiérrez[1*] and Carlos Pérez-Espigares[2]

[1]*Complex Systems Interdisciplinary Group (GISC), Department of Mathematics, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain*

[2]*Departamento de Electromagnetismo y Física de la Materia and Institute Carlos I for Theoretical and Computational Physics, Universidad de Granada, 18071 Granada, Spain*

ABSTRACT. Numerous scientific problems are related to finding shortest (or otherwise optimal) paths in networks, frequently in the presence of obstacles or constraints. A similar class of problems deals with finding optimal weight distributions in networks, which, for a given connection topology, minimize some flow or minimize a cost function. We develop a versatile theoretical framework to address these problems by an analysis of large-deviation functions of random walks on networks. Optimal paths and weights can be tailored to given statistical characterization of path observables, which allows for a huge variety of possibilities that exceed the cases commonly addressed in the standard literature.

1 Introduction

Shortest paths are one of the main objects of interest in network science. They are relevant in the study of transportation networks, the internet, protein-protein interaction networks, and other systems, and they enter into the definition of key structural properties, such as the closeness, the efficiency and the betweenness centrality. Their definition can be expanded as to include constraints, obstacles, to reach several targets from one or several sources, etc. Much research work in the applied discrete mathematics, theoretical computer science and statistical physics communities has been devoted to the solution of such problems, and the computational complexity of the algorithms employed to this end.

Such research work addresses problems such as how to find the shortest path between two nodes always passing, or without ever passing, through another node, but situations where a node is visited a fraction of times the target is reached are rarely considered. Moreover, most methods do not lend themselves easily to a generalization whereby a given observable (a path length, a frequency of visit, etc.) is required to take a specific (not necessarily extremal) value, or a prescribed statistical characterization. For example, one might conceivably want to extremize the variance that a given observable takes when a "particle" performs a given kind of cycle or search process using alternative routes, or request that a certain node be visited on average twice as frequently as some other node. In more practical terms, what is e.g. the shortest path that a passenger or a data packet can take without saturating certain node or link? These are optimal paths in a generalized, statistical sense, as they are chosen so as to ensure that the statistics of some observable takes some preassigned values or does not exceed certain bounds. Similarly, one could redistribute the link weights of a network, so that, for a given topology, some flow is maximized or some cost function is minimized subject to some statistical constraints. This is a related problem less intensively investigated in the literature, but important in its own right.

In Ref. [1] we propose a theoretical approach that unveils such generalized optimal paths and weight distributions in arbitrary networks by means of a study of the large deviations of random walks. A combination of probabilistic tools derived from non-equilibrium statistical mechanics and computational linear-algebraic tools and algorithms, such as eigensolvers and finite-size scalings of matrix spectra, are required. For the latter we use codes that we have written ourselves in C, as well as visualization tools from gnuplot and MATLAB. In the following we provide a qualitative description of our mathematical framework and the kind of results that one can achieve through its application.

*Corresponding author. E-mail: rigutier@math.uc3m.es

2 A simple example

We first illustrate the idea of unveiling optimal paths with a simple example based on a ring with a shortcut, as shown in the inset of the left panel of Fig 1. A particle starts from the node that is highlighted in red, and hops counterclockwise to its neighboring node, then hops again from there, and so on. When it reaches the node from which the shortcut starts, it has two options: continue along the ring with a certain probability or take the shortcut with another one (the sum of the two probabilities should of course equal 1). We study the statistics of the length of a cyclic walk that starts and ends at the red node. In the "natural dynamics", the statistics follows a binomial distribution that, as more and more cycles are considered in the trajectory, the sample mean converging to the mean of the distribution with Gaussian fluctuations, as dictated by the central limit theorem (see Fig. 1, left panel).

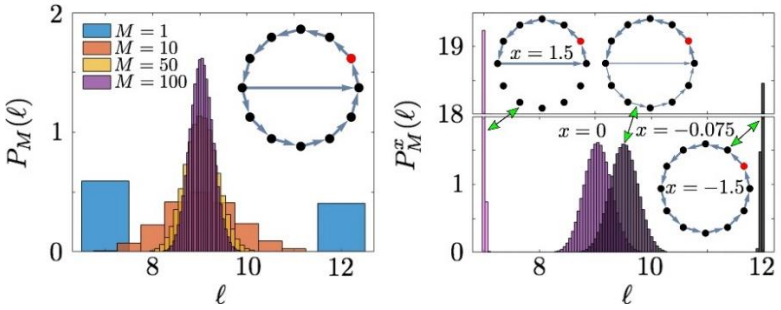


Fig. 1: Random walk on a ring with a shortcut. (*Left*) The particle starts from the node highlighted in red and returns to that node after taking ℓ steps. In a long trajectory, containing a large number of cycles, which is denoted as M, the distribution of the cyclic path length ℓ, as expected, centers around the average value (the sample mean converges towards the mean of the distribution) with Gaussian fluctuations. (*Right*) The probability distribution of the path length ℓ is biased so that the average path length is maximized ($x = -1.5$) or minimized ($x = 1.5$), or its variance is maximized ($x = -0.075$). The Doob-transformed processes giving the biased random walk that sustains such statistics appears as rings with links proportional to the jump probability.

When the probability of obtaining a given path length is biased, we obtain different probability distributions that may be chosen so as to maximize or minimize the average path length, or maximize the fluctuations (i.e. the variance of the path length), see the right panel of Fig. 1. The crucial aspect of our work is that, having achieved that, we can recover an auxiliary stochastic process, the so-called generalized Doob transform, that in fact yields the desired cycle path statistics in its stationary distribution. The link widths of the rings in the inset of Fig. 1, right panel, correspond to the probabilities of the different transitions in the Doob operator that governs such biased dynamics.

All information about the biased distribution is contained in a large-deviation function known as the scaled-cumulant generating function (SCGF), which is illustrated for the random walk in a ring in Fig. 2. The SCGF contains the full statistics of the observable under study (in this case, the cyclic path length), as the derivatives of the former correspond to the cumulants of the latter. In a sense, it plays the role of the thermodynamic potentials in equilibrium statistical mechanics. Indeed, different ensembles can be explored in order to study the statistics of dynamic observables along trajectories, which are defined by the constraints imposed on the dynamics and become equivalent in the long-time limit, much as the different equilibrium statistical mechanics ensembles (probability distributions of static observables across configurations) become equivalent in the thermodynamic limit.

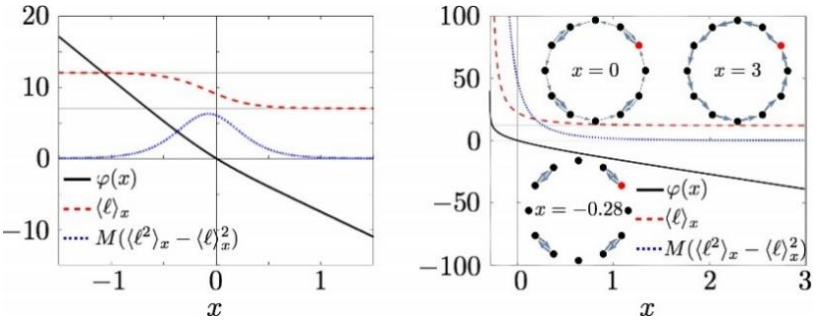


Fig. 2: SCGF $\varphi(x)$ and average and variance of the cyclic path length ℓ. The biasing of the distribution is performed via an analysis of large deviation functions, which are obtained as the numerically computed largest eigenvalue of certain matrix related to the transition matrix of the original random walk. The cumulants of the path length are derivatives of the SCGF. In the left panel, these functions of the biasing parameter *x* are shown for the ring with a shortcut of Fig. 1. In the right panel we show them for a different ring with bidirectional links, as illustrated in the insets.

3 Optimal paths in random graphs

The framework that has been applied to a ring before can be easily adapted to any network, directed or undirected, weighted or binary, which can also be spatial. That includes realizations of different well-known models (such as Erdös-Rényi random graphs, Barbási-Albert scale-free networks, etc.), but also complex systems constructed from empirical data, where some of the regularities found in mathematical models may be absent. The only practical limitation is computational in nature, and has to do with the size of the network, or more specifically with that of the transition matrix ($\boldsymbol{N} \times \boldsymbol{N}$, where $\boldsymbol{N}$ is the number of nodes): the network cannot be so large that obtaining the largest eigenvalue of the corresponding biased transition matrix, which yields the SCGF, may become computationally unfeasible. We illustrate our methodology by finding optimal paths in a small random graph of $\boldsymbol{N} = \mathbf{20}$ nodes with $\mathbf{3}\,\boldsymbol{N}$ directed links distributed uniformly at random among them, seeFigs. 4.

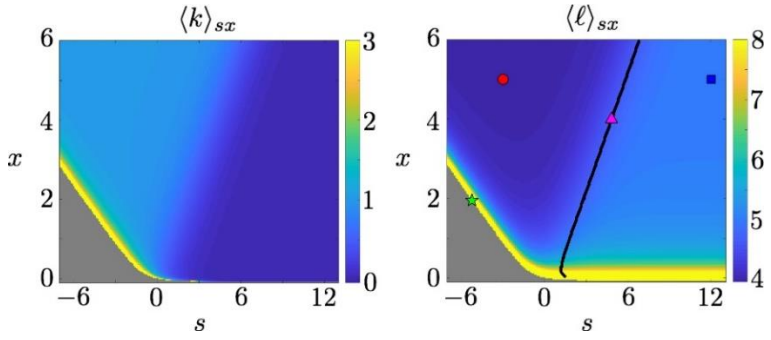


Fig. 3: Finding optimal paths in random graphs in the presence of constraints. (*Left*) Average frequency with which node 15 is visited, *k*, as a function of two biasing parameters. (*Right*) Average length of a path joining nodes 1 and 20, ℓ, as a function of the same two biasing parameters.

The graph is represented in Fig. 4. The points highlighted in the right panel correspond to the colored paths shown in Fig. 4. The black segment shows the contour line for and average value *<k>=1/3.*
The objective is to find paths of a given length that go from node 1 to node 20, passing through an "obstacle", namely node 15, with a certain frequency. A large-deviation analysis which includes two biasing parameters allows us modify the probability distribution of the two relevant observables: the path length $1 \to 20$, denoted as ℓ, and the frequency with which node 15 is visited, denoted as k. Their averages are explored for certain values of the parameters, whose role is to modify the averages of the natural dynamics by biasing their probability distributions, see Fig 3. By means of the Doob transform, we can find the trajectories that realize these biased statistics. For example, we can find paths that minimize the length to go from node 1 to node 20 without constraints (red dot in Fig. 3, red arrows in

Fig. 4), or with the constraint that node 15 should be visited once every three times that node 20 is reached (magenta triangle in Fig. 3, magenta arrows in Fig. 4). While the first problem can be addressed using Dijkstra's or other standard algorithms, the second problem and further extensions appear to be outside the reach of standard graph-theoretical algorithms and combinatorial approaches. More complex constraints and problems related to flows in spatial networks are studied in Ref. [1].

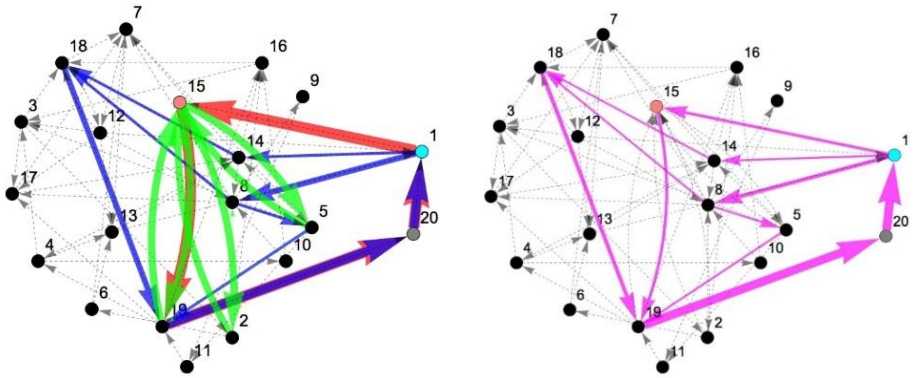


Fig. 4: Generalized optimal paths obtained via the probability fluxes of the Doob-transformed random walk. (*Left*) These paths correspond to the red circle, the green star and the blue square points shown in Fig. 3, right panel. (*Right*) These paths correspond to the magenta triangle in Fig. 3, right panel.

4 Conclusions

We have developed a versatile large-deviation framework for finding generalized optimal paths and optimal weight distributions in networks. Possible applications may range from optimizing transportation networks to the design of dissipative quantum walks.

Acknowledgments

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Cofund Programme Athenea3I Grant Agreement No. 754446, from the European Regional Development Fund, Junta de Andalucia-Consejeria de Economia y Conocimiento, Ref. A-FQM-175-UGR18, and from MINECO (Spanish Ministry of Economy), Spain, Ref. FIS2017-84151-P.

References

[1] R. Gutiérrez and C. Pérez-Espigares. *Generalized optimal paths and weight distributions revealed through the large deviations of random walks on networks. Physical Review E* **103**, 022319 (2021).

NEUTRONIC ANALYSES OF THE BFS EXPERIMENTS FOR AN "LFR ORIENTED" ADJUSTMENT OF NUCLEAR DATA LIBRARIES

Massimo Sarotto[1*], Francesco Lodi[2], Donato M. Castelluccio[2], Giacomo Grasso[2] and Vincenzo G. Peluso[2]

[1]*ENEA FSN-SICNUC-PSSN, C.R. Saluggia, Strada per Crescentino 41 – 13040 Saluggia (VC)*

[2]*ENEA FSN-SICNUC-PSSN, C.R. "E. Clementel", Via Martiri di Monte Sole 4 – 40129 Bologna*

ABSTRACT. The design - and in perspective the licensing - of the core for a new reactor concept as the Lead Fast Reactor (LFR) requires a thorough validation of neutronic calculation codes and data libraries to demonstrate the validity of the results and the appropriateness of the design methodologies and tools. Nowadays, thanks to the development of very sophisticated deterministic and stochastic neutronic codes running on high-performance computers, nuclear cross-sections represent the main source of uncertainty. By exploiting the experimental measurements available in the OECD/NEA's IRPhE database and selecting the most representative ones for the European demonstrator of the LFR technology (ALFRED), the so-called nuclear data adjustment methodology has been carried out in order to improve the accuracy of the neutronic tools currently in use in ENEA for the LFR cores design. This work reports an example of the validation activities performed for the ERANOS deterministic code ported to CRESCO HPC. The very complex geometry of the BFS fast reactor (IPPE, Russia) was accurately reproduced by means of 2D and 3D geometrical models. The Calculation-to-Experiment ratio (C/E) results here reported - successively adopted for the ENDF/B-VIII.0 data library adjustment – deal with the measurements of both integral and local neutronic parameters.

1 Introduction

The Lead-cooled Fast Reactor (LFR) is one of the six reactor concepts contemplated by the Generation IV International Forum (GIF) initiative [1]. In the EURATOM 7th Framework Program the LEADER project had as a main goal the preliminary design of the Advanced LFR European Demonstrator (ALFRED) [2]. In LEADER, ENEA coordinated the core design activities and has maintained this role within the Fostering ALFRED Construction (FALCON) consortium [3] to pursue a refined core design by detailed analyses including uncertainties propagation [4].

Since no previous LFR experiences are available for the validation of the neutronic tools used in the ALFRED core design, a thorough assessment of the calculation uncertainties is needed in the perspective of the reactor licensing by the Romanian safety authorities. Indeed, sensitivity and uncertainty (S/U) analyses are used for establishing the adequateness of the assumed safety margins that have to cope – with the aimed confidence – with the relative uncertainties.

The S/U analyses were performed for the key safety-relevant parameters of the ALFRED reactor – such as core criticality, Control Rods (CRs) worth, delayed neutrons fraction and coolant density effect – by identifying the nuclear reactions contributing the most to such parameters and their combined uncertainties [6]. The initial results showed that cross-sections represent the largest source of uncertainties: for this, the nuclear data adjustment methodology was established in the ENEA FSN-SICNUC-PSSN laboratory, and successively applied to the ENDF/B-VIII.0 evaluated library [5] to improve the accuracy of the neutronic tools in use (*e.g.*, ERANOS system code).

The adjustment requires information from integral experiments, that is added to the differential measurements at the basis of the evaluations contained in such libraries. Experiments representative of the ALFRED core – in terms of spectrum and core characteristics – were thus selected among those performed in the past in critical facilities and included in the IRPhE database [7].

*Corresponding author. E-mail: massimo.sarotto@enea.it.

This report draws on the calculations related to one of the experiments in the Big Physical Facility (BFS) at the Institute of Physics and Power Engineering (IPPE, Russia) and it describes briefly:

- in §2, the main features of the ERANOS deterministic code (ported to the CRESCO HPC);
- in §3, the main features of the BFS-62-3A experiment, a fast reactor with U and U-Pu fuel with different enrichments (15÷21%), together with its 2D and 3D ERANOS models;
- in §4, the Calculation-to-Experiment ratio (C/E) values obtained for the core criticality (expressed through the effective multiplication factor, k_{eff}), the central CR worth - that is the reactivity variation (Δk_{eff}) caused by the CR insertion - and a couple of spectral indexes.

2 The ERANOS code coupled with ENDF/B-VIII.0 library

The European Reactor ANalysis Optimised calculation System (ERANOS) is a modular system and consists of data libraries, deterministic codes and calculation procedures developed within the European collaboration on fast reactors [8]. The different modules perform several functions to analyse reactivity, fluxes, burn-up, reaction rates, etc. of the nuclear system under study. The cross-sections of the different core regions are distributed in fine energy group structures (1968 and 172 groups) for initial analysis and successive condensation at few groups (*e.g.*, 33) by the European Cell COde (ECCO) [9]. The condensed cross-sections are then used by ERANOS to analyse the system in either 2D and 3D geometries (see §3 for the BFS-62-3A experiment), in which the flux solution is obtained with finite elements and variational nodal methods [10], respectively.

It is worth stressing that, since ERANOS does not accept libraries in the original ENDF-6 format [11] used for evaluated nuclear data files, in order to use the most recent ENDF/B-VIII.0 evaluation both the library and the covariance matrices were processed to generate their analogous versions in the peculiar format required by ERANOS. For the particular application aimed at a library adjustment, it was decided to perform the ECCO condensation at 80 groups that, combined with the large amount of spatial meshes needed to model the complex geometry of the BFS-62-3A core (§3), required a wide amount of RAM for reasonable execution times.

Despite the fact that ERANOS is a serial code, its installation on the HPC clusters of the CRESCO infrastructure provides several advantages: the wide amount of RAM and fast processors, the high number of available nodes (in which different instances of the code can run in parallel *e.g.*, for parametric investigations) and the high quality assurance deriving from a shared resource (in which an official-centralized copy of ERANOS is verified and constantly updated).

3 The BFS-62-3A reactor and its ERANOS 2D and 3D models

The BFS-62-3A is a full-scale model of fast reactor cores, with blankets and in-vessel shields. Fig. 1 shows a horizontal view of the core layout together with its ERANOS model by means of a 3D hexagonal geometry. The critical assemblies are a set of vertical Stainless Steel (SS) or/and Al tubes with 5.0 cm outer diameter and 0.1 cm wall thickness, ~3 m long with ~90 cm fissile length surrounded by axial SS reflectors. As shown in Fig. 2, the cylindrical tubes (arranged in the triangular core lattice) were filled with disks (pellets) of fissile (metallic U, Pu) or structural materials (*e.g.*; SS, Na used as coolant in fast spectrum reactors): this axial geometry of the fuel rod/assembly was accurately reproduced in the ECCO cell models (§2).

Fig. 3 shows the 2D ERANOS cylindrical model of the BFS-62-3A core. The cylindrical transformation of the different fuel enrichment zones (LEZ, MEZ, PEZ, HEZ) was performed, as well as for the surrounding radial reflector made by SS (~1/3) and blanket (*i.e.*, depleted U) elements. To obtain an optimised model, a spatial discretization for all the elements was implemented by identifying, for all the assembly types, the radial regions having different neutronic features. In the transformation the overall volume of each region was preserved, by a careful choice of the radii of the interfaces between the different fuel zones, the CRs and the Safety Rods (SRs).

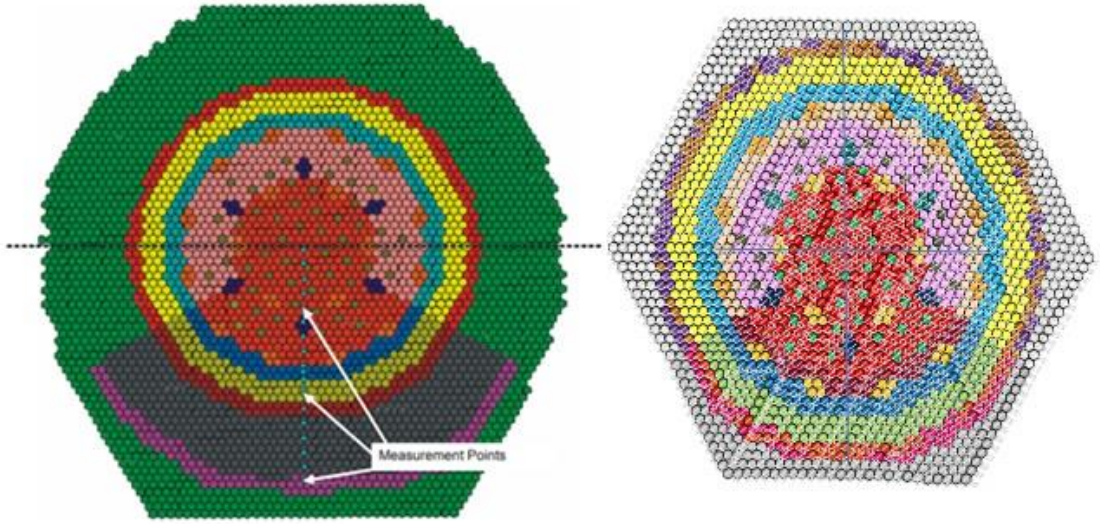


Fig.**1**: The BFS-62-3A reactor experiment (left) and its ERANOS 3D hexagonal model (right).

Fig.**2**: Cut of a BFS fuel rod/assembly with fissile and structural pellets [7].

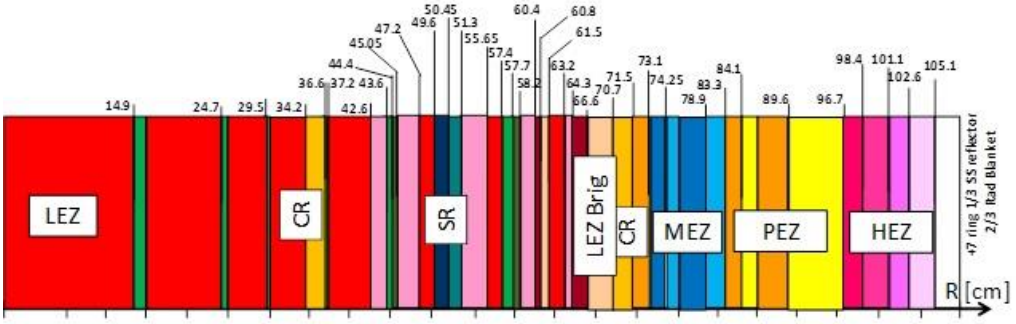


Fig.**3**: The ERANOS 2D cylindrical model of the BFS-62-3A core (LEZ/MEZ/HEZ = Low / Medium / High Enriched Zone, PEZ = Pu Enriched Zone, LEZ Brig = LEZ with different Na pellet).

4 Some C/E results for nuclear data libraries adjustment

Fig. 1 shows the measurement points in the BFS-62-3A core where fission chambers were located at the core mid-plane for the measurement of two spectral indexes. They are defined as the ratio between the fission cross sections of ^{238}U and ^{249}Pu over the ^{235}U one: f28/f25 and f49/f25, respectively. In the 80 energy-group structure they were calculated trough the ratio of reaction rates:

$$SI^{r} = \frac{\Sigma_g \sigma_g^{r} \phi_g}{\Sigma_g \sigma_g^{F25} \phi_g} \qquad (1)$$

where:

- "SI" is the spectral index value for f28/f25 and f49/f25;
- "ϕ_g" is the flux value in each energy group g;
- "σ_g^{f25}" is the ^{235}U fission microscopic cross section value in each energy group g;
- "σ_g^{r}" is the fission microscopic cross section for the nuclide "r" of interest (^{238}U and ^{249} Pu).

Table 1 reports the C/E values obtained with ERANOS / ENDF/B-VIII.0 for the two spectral indexes in the LEZ measurement point (by the 2D geometry model), the k_{eff} and central CR worth measurements (by 2D and 3D models). To be noticed that the central CR does not appear in Figs. 1 and 3: for its worth measurement, the CR bank replaced four central assemblies in the core lattice. The other columns report the experimental uncertainties (ΔE/E) and the calculation accuracy ((E-C) / C), both expressed in per cent and then directly comparable. It results evident that the calculation uncertainty is lower than the experimental one for the CR worth and the f49/f25 measurements. Since, at the same time, these (integral and local) parameters satisfy well the ALFRED representativeness established by S/U analyses performed in 2D geometries, the corresponding C/E results were adopted for the adjustment of the related cross-sections data [6].

Table 1: Experimental uncertainty ΔE/E, ERANOS C/E and (E-C) / C values for two spectral indexes, the k_{eff} multiplication factor and central CR worth (2D and 3D core geometry models).

Parameter	C / E 2D RZ	C / E 3D Hex	ΔE / E [%]	(E – C) / C 2D RZ	[%] 3D Hex
k_{eff}	0.989	0.983	0.300	1.088	1.742
CR worth	1.015	1.096	1.954	-1.436	-8.759
f28/f25	1.064		2.752	-5.984	
f49/f25	0.992		1.762	0.785	

References

[1] OECD/NEA. Technology Roadmap Update for Gen. IV Nuclear Energy Systems (2014).

[2] G. Grasso *et al.*. The core design of ALFRED, a demonstrator for the European lead-cooled reactors", *Nuclear Engineering and Design*, Vol. **278**, pp. 287-301 (2014).

[3] https://www.euronuclear.org/e-news/e-news-43/ansaldo.htm.

[4] G. Grasso *et al.*. "An improved design for the ALFRED core", *Proc. International Congress on Advances in Nuclear Power Plants ICAPP 2019*, Juan-les-pins, France (2019).

[5] ENDF/B-VIII.0 evaluated nuclear data library, https://www.nndc.bnl.gov/endf/b8.0

[6] D. M. Castelluccio *et al.*, "ENDF/B-VIII.0 nuclear data sensitivity and uncertainty analysis (S/U) of key safety-relevant reactivity coefficients for the ALFRED core", *Proc. IAEA Int. Conf. on Fast Reactors and related fuel cycles: sustainable clean energy for the future FR22*, Beijing China (2022).

[7] OECD/NEA. International Reactor Physics Experiments Database Project (IRPHE), *International Handbook of Evaluated Reactor Physics Benchmark Experiments*, (2004).

[8] G. Rimpault *et al.* The ERANOS code and data system for Fast Reactor neutronic analyses", *Proc. Int. Conf. PHYSOR 2002*, Seoul, Korea (2002).

[9] G. Rimpault *et al.* Physics documentation of ERANOS: the ECCO cell code, *CEA Tech. Report DERSPRC-LEPh-97-001* (1997).

[10] J.M. Ruggeri *et al.*. 1993. "TGV: a coarse mesh 3 dimensional diffusion-transport module for the CCRR/ERANOS code system", *CEA Tech. Report DRNR-SPCILEPh-93-209*, (1993).

[11] CROSS-SECTION EVALUATION WORKING GROUP, ENDF-6 Formats Manual, Document ENDF-102, *Internal report BNL-203218-2018-INRE*, BNL, USA (2018).

NODE DIFFERENTIATION DYNAMICS ALONG THE SYNCHRONIZATION PATH IN COMPLEX NETWORKS

Irene Sendiña-Naidal[1,2*], Inmaculada Leyva[1,2] and Christophe Letellier[3]

[1]*Universidad Rey Juan Carlos & GISC, Complex Systems Group, C/Tulipa´n s/n, Mo´stoles, Madrid, Spain*

[2]*Universidad Politécnica de Madrid, Center for Biomedical Technology, Pozuelo de Alarco´n, Madrid, Spain*

[3]*Rouen Normandie University - CORIA, Campus Universitaire du Madrillet, F-76800 Saint-Etienne du Rouvray, France.*

ABSTRACT. Synchronization has been the subject of intense research during decades mainly focused on determining the structural and dynamical conditions driving a set of interacting units to a coherent state globally stable. However, little attention has been paid to the description of the dynamical development of each individual networkedunit in the process towards the synchronization of the whole ensemble. In this work, we show how in a network of identical dynamical systems, nodes belonging to the same degree class differentiate in the same manner visiting a sequence of states of diverse complexity along the route to synchronization independently on the global network structure. We unveil how this differentiation route and howit depends on the coupling function. This study provides new insights to under- stand better strategies for network identification and control or to devise effectivemethods for network inference.

1 Introduction

Since the beginning of the research on the dynamics of complex networks, the deep relationship between topology and dynamics has been thoroughly explored with regard the synchronization between the nodes' dynamics, and the knowledge gathered so far has driven the advance in crucial applications. However, there are cases in which the system performs its activity in a partial or weakly synchronization regime to preserve the balance between functional integration and segregation, whereas full synchronization is pathological. But even in this far from synchronized state, each coupled unit is encoding inits own dynamics the signature of its structural role. We explore how this prominent feature could be used to extract information about the network, without having the need to make any reference to pairwise correlations, even in those situations where the structure is unknown. The effect of the coupling strength against the dynamical complexity of the nodes is found to be a function of their topological roles, with nodes of higher degree displaying lower levels of complexity. Importantly, our results imply that it is possible to infer the degree distribution of a network only from individual dynamical measurements in a large variety of systems [1, 2].

2 Model

Let us consider a network of N identical dynamical units which evolves according to,

$$\dot{x}_i = f(x_i) - d\sum_{i=1}^{N} \mathcal{L}_{ij} h(x_j) \quad (1)$$

*Corresponding author. E-mail: irene.sendina@urjc.es

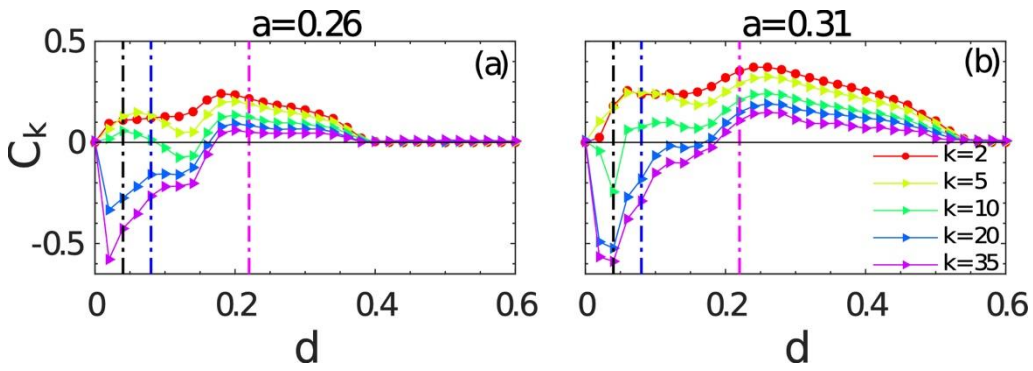


Fig 1: Average k-class relative complexity Ck for several values of degree N=300, <k>=4 SF networks of chaotic Ro¨ssler oscillators as a function of the coupling strength d. The vertical dotted lines mark the couplings analyzed in Fig. 2.

where $f(x_i) = (-y_i - z_i, x_i + ay_i, b + z_i(x_i - c))$ is the vector function governing the node dynamics, d is the coupling strength and $\mathcal{L} = \{\mathcal{L}_{ij}\}$ is the Laplacian matrix describing the coupling structure. To evaluate our hypothesis about the relationship between topological centrality and node dynamics, for each node i we calculate the modified complexity Ci = Si + Δi being Si the permutation entropy and Δi the structurality marker [2]. We chose this complexity definition because of its ability to distinguish chaotic dynamics from noise.

As we expect that nodes having the same degree k will play equivalent dynamical roles in the ensemble, we compute the evolution of $\langle C \rangle_k$ within a degree class by averaging over the nodes that have identical degree k:

$$\langle C \rangle_k = \frac{1}{N_k} \sum_{[i|k_i=k]} C_i \qquad (2)$$

As the reference value, we compute the single element complexity $\langle C \rangle_o$, equivalent to the uncoupled case d=0. Finally, our main measure is the *relative complexity*:

$$C_k = \langle C \rangle_k - \langle C \rangle_o \qquad (3)$$

3 Numerical results

3.1 Numerical methods and use of computational resources

We used the CRESCO resources for the computational study of the evolutionof dynamical networks, considering different topologies, network size and different average degrees. Home-made C codes based on fix-step 4th order Runge-Kutta algorithm were written, compiled with GCC and executed at CRESCO. Extensive distributed computations have been performed for large parameters ranges, with statistical validation of the results. Calculations were performed in CRESCO4 and CRESCO5, using the h144 queues for full evolution simulations. Homemade MATLAB scripts wereused for visualizing the results.

3.2 Rössler complex networks

In our simulation we use the paradigmatic Rössler system to exemplify our hypothesis. The vector flow taking b=0.2 and c=5.7 as fixed parameters and a as the bifurcation parameter.

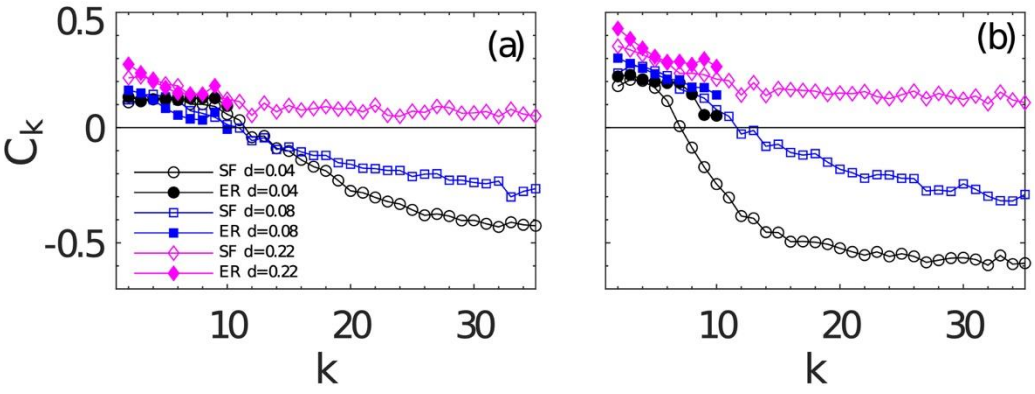


Fig 2: Relative complexity C_k as a function of the degree k, for SF (void symbols) and ER (full symbols) ensembles with the coupling strengths d marked in Fig.1 with vertical lines: $d = 0.04$ (circles), $d = 0.08$ (squares), and $d = 0.22$ (triangles). Results are averaged over 10 different network realizations.

Initial conditions are randomly selected within a small neighborhood of radius 0.5 centered at the origin of the state space. In Figure 1 we show the results for a $\langle k \rangle = 4$ scale-free ensemble of N=300 identical chaotic Rössler oscillators [1,3] and two different values of the bifurcation parameter a. We observe that immediately after coupling, the relative complexity of the node dynamics splits as a strongly hierarchical function of k, that persists for all the range of d up to the synchronization value, where all the nodes recover the original complexity of the uncoupled state. This behavior is suggesting an application to structurally rank the nodes in a network according to the complexity of their time series and, therefore, to potentially use this anti-correlation as a proxy for the degree sequence. The fact that this correlation between C_k and k persists along a large range of coupling d means that the method could be useful in natural systems, where in general the coupling is not an accessible parameter, and in addition could be different for different areas of the system. We have obtained equivalent results in a large variety of other systems as non phase-coherent chaos as Lorenz model, pulse-coupled neurons [1], delayed systems as Mackey-Glass model, higher dimensional systems as Saito [3], and has been also experimentally observed in networks of nonlinear electronic circuits [1]. The result is robust against node heterogeneity, noise and dynamical and topological changes [1, 3]

Therefore, we conclude that it is possible to correlate the node degree with the relative complexity. Furthermore, the degree centrality is the single structural parameter that affects the node behavior, while the rest of environmental topological features has no impact. This is shown in Fig. 2, where we plot the value of the relative complexity C_k as a function of k for the three representative values of d marked in Fig. 1 with vertical dotted lines, both for SF networks and ER networks with $<k>=4$. The ER and SF curves overlap and, therefore, the dependence of C_k with d is the same regardless of the topology. This is quite remarkable since the ER and SF networks have a different critical coupling: for a given d-value, their global dynamics are different, but the nodes of degree k have equivalent dynamics independently of the environment.

To better illustrate the node differentiation, we plotted the first-return maps for three different k-classes of nodes along the route to synchronization (Fig. 3). Low-degree nodes (k=2) produce first-return maps whose thickness increases with the coupling strength (top row in Fig. 3). For these nodes, the relative complexity is always positive, and the first maps resemble the map of the uncoupled dynamics but thicker. For large degree nodes (third row in Fig. 3), once the minimum complexity is reached in the weakly coupled regime, the complexity increases with the coupling strength d. Around the minimum, the maps comprise a small cloud of points in the neighborhood of the inner singular point; this is progressively transformed into a "noisy" period-1 limit cycle, which is characterized by a cloud of points elongated perpendicularly to the bisecting line and located around the center of the map. Before the onset of synchronization, large degree nodes produce a map which resembles the nominal one but is slightly thicker. As previously discussed, node differentiation is mostly a monotonous function of the coupling strength and the degree.

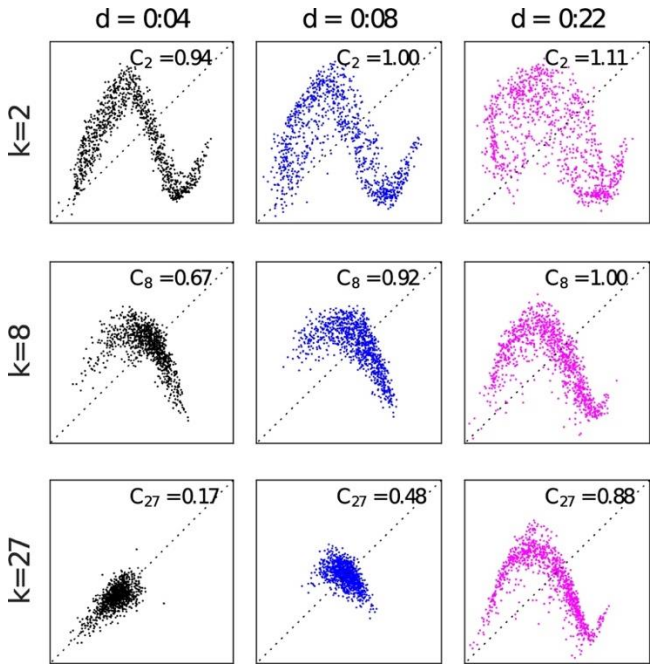


Fig 3: First-return maps to a Poincaré section for three k-classes of nodes along a route to synchronization in one realization of a N=300 (<k>= 4) SF network of Rössler nodes (a = 0.31 and other parameters as in Fig. 1).

Conclusions

We have inspected the relationship between the topological role of a node in a complex network and its dynamical behavior, represented by its relative complexity. In the weakly coupled regime, we observed a distinctive negative correlation between complexity and degree centrality in a large variety of systems, as chaotic oscillators or pulse-coupled neurons. These results suggest that the role played by the topology of a network could be unveiled by just computing the dynamical complexity associated with the time series sampled at each node. The fact that structural information of a network can be inferred without computing pairwise correlations like those commonly performed in functional networks could be exploited in diverse fields as neuroscience, econophysics or power grids.

References

[1] A. Tlaie, I. Leyva, R. Sevilla-Escoboza, V. P. Vera-Avila, I. Sendiña-Nadal. Phys. Rev. E 99, 012310 (2019).
[2] C. Letellier, I. Leyva, I. Sendiña-Nadal. Phys. Rev. E 101, 022204 (2020).
[3] C. Letellier, I. Sendiña-Nadal, L. Minati, I. Leyva. Phys. Rev. E 104, 014303 (2021).

MOLECULAR DYNAMICS SIMULATIONS TO EVALUATE THE EFFECT OF ENVIRONMENTAL POLLUTANTS ON EPIGENETIC MODULATORS

Giorgia Innamorati[1,2*], Maria Pierdomenico[1], Barbara Benassi[1] and Caterina Arcangeli[1]

[1]*Laboratory of Health and Environment, Division Health Protection Technologies, ENEA, Rome, Italy*

[2]*Università Cattolica del Sacro Cuore, Laboratory of Histology and Embriology, Rome, Italy*

ABSTRACT. Long-lasting adverse effects on human health caused by environmental contaminants have been detected, including those affecting the activity of epigenetic modulators. Here we present part of the results by Innamorati et al. 2021 [under submission, 1]. Molecular Dynamics (MD) simulations were employed to evaluate whether and how a set of selected emerging environmental pollutants can influence the level of DNA methylation. All the simulations and binding free energy calculations were performed on CRESCO6.

1 Introduction

The exposure to emerging environmental contaminants, like plastic-associated chemicals and persistent organic pollutants, have been proven to induce several toxic effects in humans. Although there is evidence of the impacts of these pollutants on the modulation of gene expression [2 and references therein], their ability to directly interact with the epigenetic machinery still has to be univocally determined. In this report the potential binding and the dynamics of interaction between persistent organic pollutants with two members of the human DNA methyltransferases (DNMTs) family has been evaluated by using MD simulations. The simulations and the binding free energy values identified a subset of chemicals which do not interact stably with the epigenetic targets, whereas most of the environmental pollutants establish strong and stable interactions with the binding domains of the epigenetic targets, thus suggesting that they might act as inhibitors of the enzymatic activity.

2 Computational Methods

The contaminant-enzymes complexes obtained by docking procedures [1,3] were used as starting configuration for the MD simulation. All the simulations were performed by using Gromacs2019 [4] with GROMOS54a7 force field [5] in combination with SPC/E water model [6]. Periodic boundary conditions were applied and after 50000 steps of steepest descent minimization [7] and 400 ps of equilibration unrestrained MD were carried out in NPT ensemble for 150 ns. Table 1 summarizes some details of the simulations. Trajectories were visualized by means VMD molecular visualization program [8]. The binding free energy were computed by using g_mmpbsa tool for Gromacs [9].

*Corresponding author. E-mail: g.innamorati1@campus.unimib.it

Table 1: Simulation details

Temperature	310 K [10]
Neutralizating ions	NaCl (150 mM)
Pressure	1 bar [11]
Integration time step	2 fs
Long-range electrostatic force	PME method [12]
Short-range electrostatic and van der Waals cut-off	1 nm
Length of simulation	150 ns

3 Results and Discussion

In order to assess whether a direct interaction of the contaminants with the epigenetic enzymes might occur, we have performed MD simulations of 18 different contaminant-enzymes complexes obtained by docking procedure [1]. The complexes are described in Table 2.

Table 2: ID of complex contaminant-enzyme

Contaminants	Epigenetic Enzyme - EpiE1[a]	Epigenetic Enzyme – EpiE2[b]
Cont1c	Cont1– EpiE1	Cont1– EpiE2
Cont2d	Cont2– EpiE1	Cont2– EpiE2
Cont3e	Cont3– EpiE1	Cont3– EpiE2
Cont4f	Cont4– EpiE1	Cont4– EpiE2
Cont5g	Cont5– EpiE1	Cont5– EpiE2
Cont6h	Cont6– EpiE1	Cont6– EpiE2
Cont7i	Cont7– EpiE1	Cont7– EpiE2
Cont8j	Cont8– EpiE1	Cont8– EpiE2
Cont9k	Cont9– EpiE1	Cont9– EpiE2

[a)] human DNA methyl transferase 1 (PDB id: 3swr.pdb, [Hashimoto et al., to be published])
[b)] human DNA methyl transferase 3 (PDB id: 5yx2.pdb, [13])
[c)] Benzyl butyl phthalate (PubChem ID: 2347)
[d)] Dibutyl phthalate (PubChem ID: 3026)
[e)] Di-(2-ethylhexyl) phthalate (PubChem ID: 8343)
[f)] Dimethyl phthalate (PubChem ID: 8554)
[g)] Monobutyl phthalate (PubChem ID: 8575)
[h)] Mono-(2-ethylhexyl) phthalate (PubChem ID: 20393)
[i)] Mono-methyl phthalate (PubChem ID: 20392)
[j)] Perfluorooctanoic acid (PubChem ID: 9554)
[k)] Perfluorooctanesulfonic acid (PubChem ID: 74483)

The structural stability and the compactness of the enzymes in complex with the contaminants, was assessed by calculating the root means square deviations (RMSDs) of the backbone atoms of the EpiE enzymes as a function of the simulation time. Both EpiE1 (Figure 1) and EpiE2 (Figure 2) with ligands in their binding sites seem to stabilize after a short time from the start of the simulation.

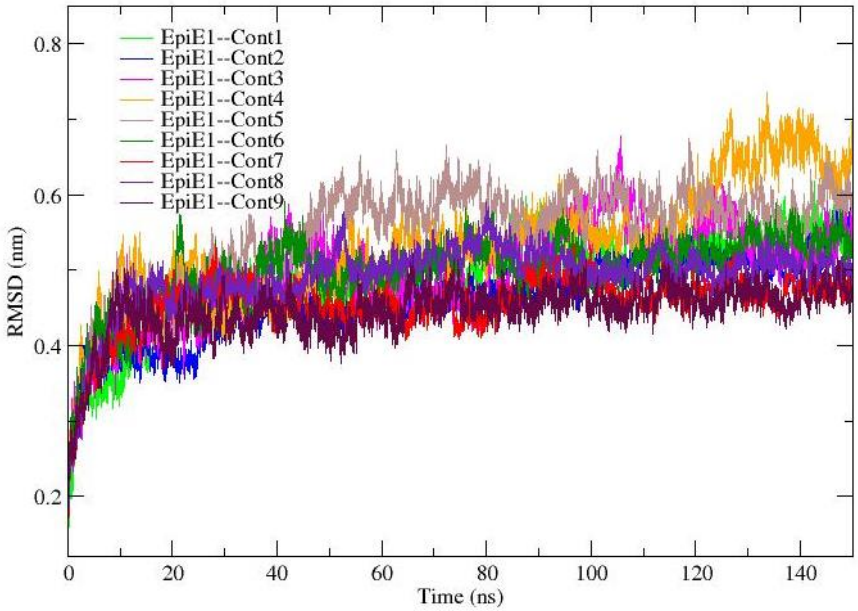


Fig. 1: RMSDs, as a function of simulation time, of the backbone atoms of the EpiE1 enzyme in complex with the contaminants.

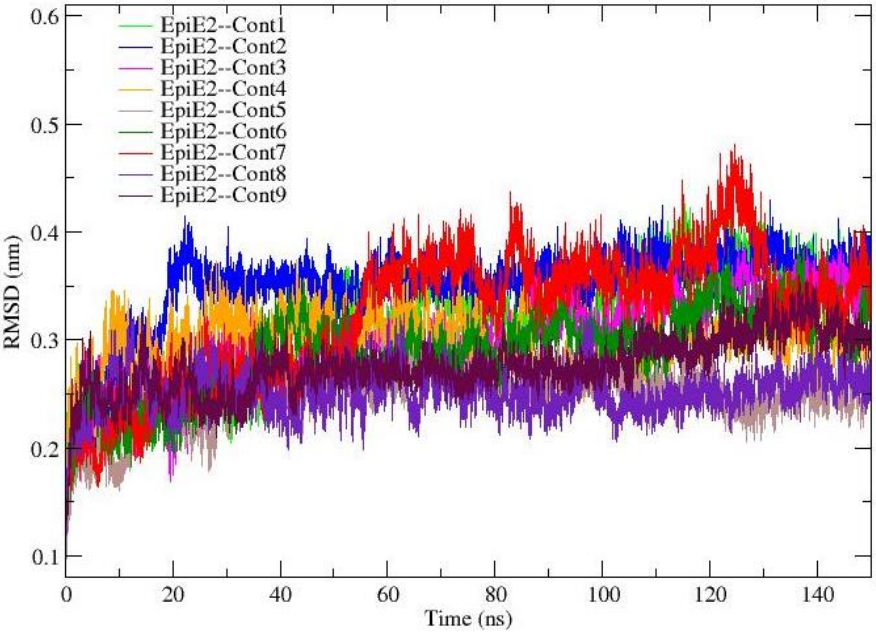


Fig. 2: RMSDs, as a function of simulation time, of the backbone atoms of the EpiE2 enzyme in complex with the contaminants.

A visual inspection of the structural configurations sampled by the contaminants in the target binding sites and an evaluation of the binding free energies (Table 3) shown that these contaminants can act differently with the chosen targets. Some of them, the Cont2, Cont3, Cont5 and Cont6 could be potential inhibitors of both EpiE1 and EpiE2 enzymes, indeed they form stable complexes (Figure 3A) and have a high affinity with the binding pockets (Table 3). Cont4, Cont7 do not stably interact with both the epigenetic enzymes (Figure 3B, Table 3), whereas Cont1, Cont8 and Cont9 show a differential behaviour when interacting with EpiE1 and EpiE2 (Figure 3 C, Table 3).

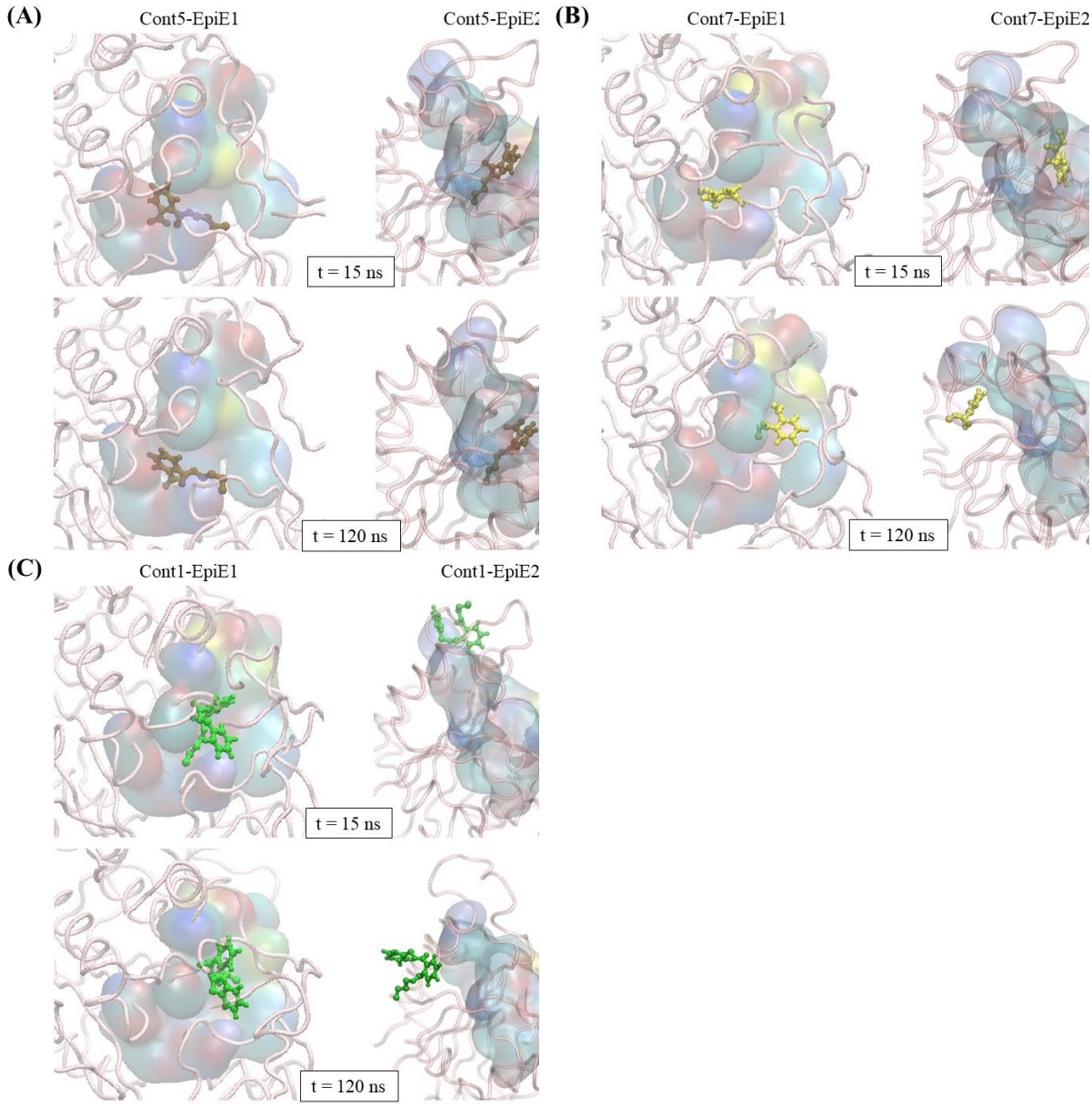


Fig. 3: Simulation snapshots taken at selected times of MD trajectories (0-45-145 ns) of selected contaminant-enzyme complexes. (A) Cont5 as example of possible inhibitor of both epigenetic enzymes; (B) Cont7 as example of contaminants that do not interact stably with both epigenetic enzymes; (C) Cont1 as example of contaminant that interact differently with the epigenetic enzymes (it inhibits EpiE1 but not EpiE2). Contaminants are represented by CPK model and epigenetic enzymes by Tube model. The binding site is represented by QuickSurf model.

Table 3: Binding free energies of EpiE1 and EpiE2 binding site in complex with the chosen contaminants.

Contaminants	MM-PSBA Binding Free Energy (kJ/mol) - EpiE1	MM-PSBA Binding Free Energy (kJ/mol) – EpiE2
Cont1	-52.854 ± 16.491	-11.725 ± 27.084
Cont2	-68.146 ± 10.133	-69.342 ± 13.043
Cont3	-67.329 ± 16.683	-94.898 ± 16.476
Cont4	-35.838 ± 20.254	-51.509 ± 14.061
Cont5	-55.032 ± 13.057	-59.570 ± 11.303
Cont6	-42.025 ± 21.987	-69.530 ± 12.940
Cont7	-36.000 ± 9.957	-30.058 ± 15.891
Cont8	-16.719 ± 11.273	-57.565 ± 13.256
Cont9	-38.297 ± 16.535	-63.778 ± 19.541

4 Conclusions

The results demonstrate that the computational approach based on the combination of MD simulation and MM-PBSA calculations represent a valuable *in silico* tool for identifying direct potential interactions of emerging contaminants with key epigenetic regulators.

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High performance Computing infrastructure and his staff [13]. CRESCO/ENEAGRID High performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

References

[1] G. Innamorati, M. Pierdomenico, B. Benassi, C. Arcangeli. The interaction of DNMT1 and DNMT3A epigenetic enzymes with phthalates and perfluoroalkyl substances: an in silico approach. J Biomol Struct Dyn. 2022 Jan 6:1-17. doi: 10.1080/07391102.2021.2023642. Epub ahead of print. PMID: 34986741...

[2] M. N. Alam, U. M. Shapla, H. Shen, and Q. Huang, "Linking emerging contaminants exposure to adverse health effects: Crosstalk between epigenome and environment," *Journal of Applied Toxicology*, vol. 41, no. 6, pp. 878–897, 2021, doi: 10.1002/jat.4092.

[3] T. M. Menchaca, C. Juàrez-Portilla, and R. C. Zepeda, "Past, Present, and Future of Molecular Docking," *IntechOpen*, vol. 14, no. 1, pp. 549–556, 2020, doi: 10.5772/intechopen.90921.

[4] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005, doi: 10.1002/jcc.20291.

[5] N. Schmid *et al.*, "Definition and testing of the GROMOS force-field versions 54A7 and 54B7," *European Biophysics Journal*, vol. 40, no. 7, pp. 843–856, 2011, doi: 10.1007/s00249-011-0700-9.

[6] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," *Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987, doi: 10.1021/j100308a038.

[7] S. S. Petrova and A. D. Solov'Ev, "The Origin of the Method of Steepest Descent," *Historia Mathematica*, vol. 24, no. 4, pp. 361–375, 1997, doi: 10.1006/hmat.1996.2146.

[8] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual Molecular Dynamics," *Journal of molecular graphics*, vol. 14, no. October 1995, pp. 33–38, 1996, doi: 10.1016/0263-7855(96)00018-5.

[9] R. Kumari, R. Kumar, and A. Lynn, "G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations," *Journal of Chemical Information and Modeling*, vol. 54, no. 7, pp. 1951–1962, 2014, doi: 10.1021/ci500020m.

[10] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *Journal of Chemical Physics*, vol. 126, no. 1, 2007, doi: 10.1063/1.2408420.
[11] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981, doi: 10.1063/1.328693.
[12] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method,” *The Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995, doi: 10.1063/1.470117.
[13] Z. M. Zhang *et al.*, “Structural basis for DNMT3A-mediated de novo DNA methylation,” *Nature*, vol. 554, no. 7692, pp. 387–391, 2018, doi: 10.1038/nature25477.

LARGE EDDY SIMULATION OF A TRANSCRITICAL H_2/O_2 NON-PREMIXED FLAME

Eugenio Giacomazzi*, Donato Cecere and Nunzio Arcidiacono

ENEA, Laboratory of Processes and Systems Engineering for Energy Decarbonisation, Department of Energy Technologies and Renewable Sources, Rome, Italy

ABSTRACT. The article is devoted to the analysis of the flame characteristics and anchoring mechanism of a turbulent non-premixed flame burning gaseous hydrogen and liquid oxygen at 60 bar. The test case was experimentally investigated in the facility MASCOTTE at ONERA. The divergent tapered injector is included in the computational domain to consider its effects. The case is simulated by means of the in-house parallel code HeaRT in the three-dimensional LES framework run on the CRESCO6 supercomputing platform. The cubic Peng-Robinson equation of state in its improved translated volume formulation is assumed. Diffusive mechanisms and transport properties are accurately modelled. A finite-rate detailed scheme involving the main radicals, already validated for high-pressure H_2/O_2 combustion, is adopted.

1 Introduction

Aerospace applications, and in particular liquid oxygen rocket engines and cooling systems, largely promoted research on real gas flows. In the last decade, high-pressure combustion of reactants exhibiting real gas behaviour has been focusing more attention not only for the renewed interest in space exploration, but also for its application in electric power generation, as in advanced supercritical CO2 gas turbine cycles (designed to operate at 300 bar), organic Rankine cycles, and diesel engines. In such applications the flow can be far away from ideal thermodynamics, and the different fluid behaviour has to be accounted for by means of real gas equations of state and specific models for molecular transport properties.

Experimental work at such high-pressure conditions may be prohibitive: the use of advanced laser diagnostics is not an easy task to achieve (few examples exist); besides, facilities themselves may be very expensive. Most of research and design in transcritical/supercritical combustion is based on numerical simulation, but also this approach is not an easy task.

Some issues deal with the modelling accuracy of the equation of state, e.g., its ability to capture huge variations of fluid properties when crossing the pseudo-boiling line; some others with its computational efficiency, i.e., too complex and accurate equations of state (EoS) cannot be used in time-consuming approaches, like LES and DNS. In fact, some EoS can be very accurate at the expense of a non-straightforward solution of the fluid state. For LES or DNS, approaches for which an analytical solution is readily available, like the Peng-Robinson EoS, would be preferable. Besides, critical uncertainties arise from the chemical kinetics mechanisms.

Furthermore, numerical schemes are stressed hardly: the high-density gradients typical of some applications (as those involving liquid injection) and the multi- species transport enhance wiggle formation in fully compressible solvers; spurious pressure oscillations are generated when a fully conservative scheme is adopted, due to the high non-linearity of the real-fluid equation of state.

2 Numerics and Modelling

The numerical simulations are performed by means of the in-house parallel code HeaRT and ENEA's supercomputing facility CRESCO6.

The compressible Navier-Stokes equations are solved for a reacting real gas flow in the Large Eddy Simulation framework. The mathematical models adopted are derived for a Newtonian and Stokesian fluid of N_s chemical species. The Peng-Robinson cubic equation of state in its improved translated

*Corresponding author. E-mail: eugenio.giacomazzi@enea.it.

volume formulation is assumed. Mass flux is modelled considering the effects of concentration, pressure and temperature gradients; heat flux considers Fourier and mass/enthalpy diffusion effects. Molecular viscosity and thermal conductivity are accurately modelled through NIST models in REFPROP with an Extended Corresponding States method and fluid-specific correlations [1]. The diffusion coefficient D_i of the i-th species into the rest of mixture is modelled according to the Hirschfelder and Curtiss expression, where the required binary diffusion coefficient is calculated by means of kinetic theory with Takahashi's correction for high-pressure. Unclosed turbulent combustion subgrid terms are modelled through the Vreman's model [2] and the authors' LTSM (Localised Turbulent Scale Model) turbulent combustion model [3]. The combustion of hydrogen and oxygen is modelled by means of the finite-rate detailed chemical mechanism derived by Boivin from the San Diego mechanism, furtherly simplified by taking the high-pressure limit of the falloff reactions (due to the high pressure of the test case) [4]. The mechanism accounts for 8 species and 12 elementary reactions.

The equations are discretised through staggered finite difference schemes. A second-order accurate centred scheme is adopted for diffusive fluxes. Convective terms are discretized through the AUS M^+-up method [5] coupled with a second-order accurate interpolation with a TVD, linear preserving limiter for non-uniform grids [6] to reduce spurious oscillations. It is noted that in simulations of transcritical liquid flows this methodology is not sufficient to avoid large spurious pressure oscillations along the liquid interface due to the compressible solver; upon the observations in [7] numerical dissipation is locally increased by forcing such a limiter to be first-order accurate when the fluid is in a compressible liquid state, i.e., the compressibility factor is less than 0.8. The low-storage third-order accurate Runge-Kutta method of Shu-Osher is used for time integration. The total energy is defined as sum of internal (thermal) and kinetic energy only. The authors found this choice mandatory to avoid, or at least reduce, unphysical energy and temperature oscillations, mainly driving to the divergence of calculation. No spurious waves were experienced in previous simulations of premixed flames, when the total energy was defined including the chemical formation contribution.

Non-reflecting boundary conditions are implemented at open boundaries in their extended form to consider variable transport properties, local heat release and real gas effects. Turbulent velocity fluctuations are superimposed to the mean inlet values by means of a synthetic turbulence generator.

3 The Test Case and Results

The reactive case simulated in this work was experimentally investigated in the MASCOTTE cryogenic combustion test facility developed by ONERA for rocket applications. It consists of a combustor having a 50 mm × 50 mm square-section, 40 cm length with a final nozzle, and an axisymmetric coaxial injector for O_2 and H_2 mounted vertically downward, as sketched in Fig. 1 (left). The particular test case numerically studied in this article has a nominal pressure of 60 bar (case C-60) and considers injection of transcritical (liquid) O_2 and gaseous H_2 [8,9]; injection data are reported in Table 1; the turbulent Reynolds numbers, Re_t, for the two jets were estimated considering the thickness of the wall separating the two jets, h = 1 mm, as the integral turbulent length scale, l_t, and the u'_{inj} forced at the inlets. In the simulation, the combustor is modelled as cylindrical (instead of its squared section) and the tapered central pipe is modelled in a staircase fashion. The computational domain is a 120°slice of the full cylindrical geometry. The simulated length of the combustor is 0.2 m and that of the injector coaxial pipes is 0.1 m. The domain is discretised by means of 18.703.853 grid nodes using 32 nodes along the azimuthal direction.

Due to the transcritical injection of oxygen, its stream exhibits huge property changes. The wide range of the compressibility factor experienced in the present simulation, [0.2-1.03], makes clear that the adoption of a real gas equation of state is mandatory. The lower value 0.2 is reached in the liquid oxygen jet core; moving towards the liquid interface, Z increases up to 0.24, then decreases before increasing again up to 1; slightly higher values are reached in the hydrogen jet. The thickness of the region between the inner Z=0.24 isoline and the Z=1 isoline ranges from 40 to 980 μm, and sometimes becomes much wider moving downstream.

Table 1: Injection data for the two reactants in the MASCOTTE C-60 test case.

Species	T_{inj}	U_{inj}	Re_t
O_2	83	8.72	2097
H_2	275	152.09	4534

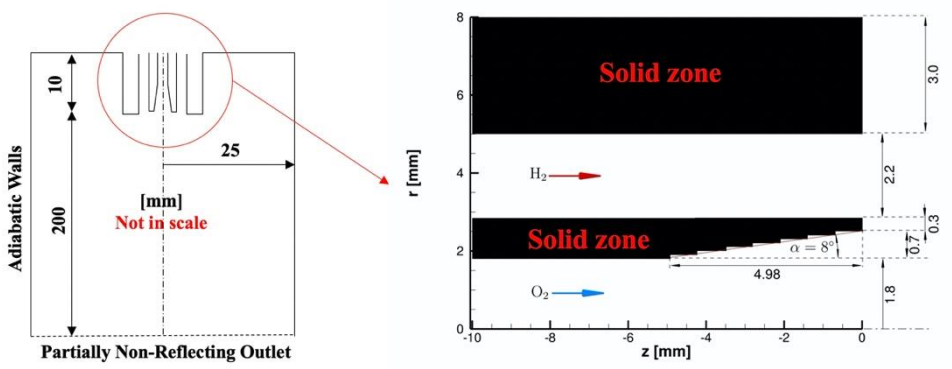


Fig.1: Sketch of the combustor and zoomed-in-view of the modelled tapered injector.

The oxygen flows into the combustion chamber through an injector with a tapered final section, where the flow exhibits separation nearly ~ 3 mm before the exit. The resulting recirculation zone captures some hydrogen and transports it within the tapered injector. Due to this recirculation, a quite stable flame establishes in it (both adiabatic walls and solid walls with heat transfer were explored).

The interaction between the shedding of large structures at the outer edge of the H_2 pipe and the synchronous shedding of small vortices at its inner edge produces a complex dynamic of the flame root. Since at supercritical conditions no atomization occurs, the flame evolves in a very thin mixing layer between the gaseous hydrogen and liquid oxygen jets. As a consequence, even for this turbulent case the flame thickness remains limited, and the flame front is confined close to the LOX jet (see Fig. 1, top); hence, due to the fluctuation of this thin reacting layer, experimentalists measured low mean temperatures.

For the 60 bar supercritical conditions, only small wrinkles appear at the LOX jet surface which do not seem to pull out from the jet, but only dissolved: these wrinkles are similar to those generated by conventional turbulent mixing layers between two compressible. gases. No ligament structures are exhibited, at least with the current spatial resolution, as also deduced from experiments.

The equivalence ratio distribution (see Fig. 2, bottom) shows a mostly rich flame. Oxygen exhibits a very quick transition from its compressed liquid state to supercritical gas (40 to 980 μm). Hence, combustion and most of the mixing process take place at nearly ideal gas conditions (consider that the compressibility factor of O_2 reaches 0.98 at 370 K).

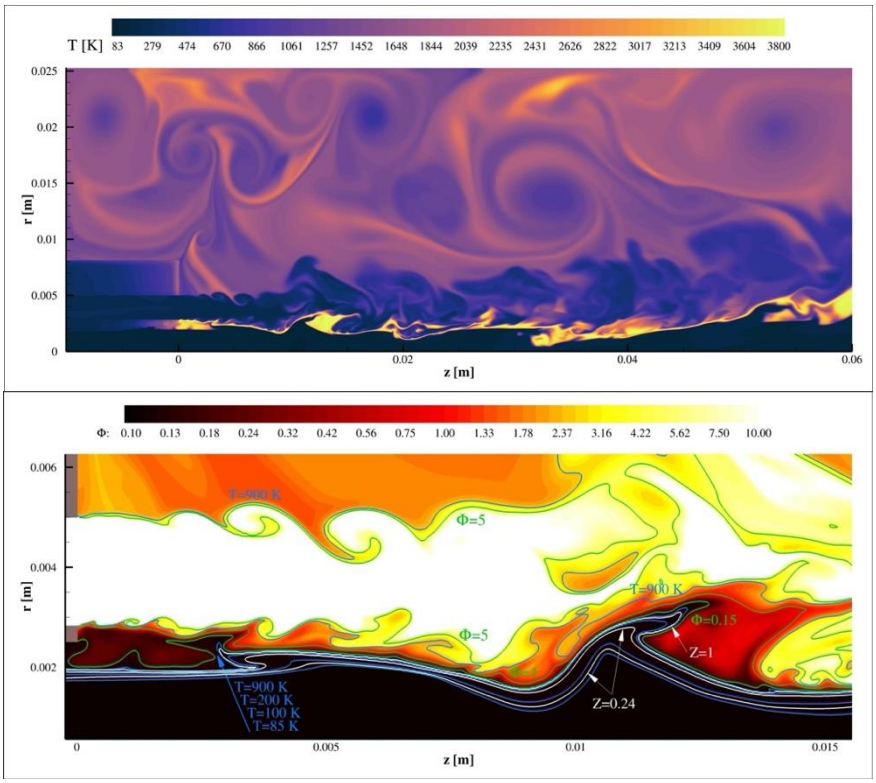


Fig.**2**: Instantaneous temperature snapshot (top); associated zoomed-in-view map of equivalence ratio, some temperature T and compressibility factor Z isolines (bottom).

References

[1] Chichester J.C., Huber, M.L., Documentation and Assessment of the Transport Property Model for Mixtures Implemented in NIST REFPROP V. 8.0. Technical Report NISTIR 6650, NIST, 2008.

[2] Vreman A.W., An eddy-viscosity subgrid-scale model for turbulent shear flow: Algebraic theory and applications. Physics of FLuids, 16(10):3670–3681, October 2004.

[3] Giacomazzi E., Cecere D., A Combustion Regime-Based Model for Large Eddy Simulation. Energies, 14(16):4934 (23), 2021. Special Issue: Combustion and Propulsion Systems.

[4] Ruiz A., Unsteady Numerical Simulations of Transcritical Turbulent Combustion in ´Liquid Rocket Engines. ED MEGEP: Energetique et transferts - Unite´ de recherche: CERFACS, Institut National Polytechnique de Toulouse (INP Toulouse), Toulouse, France, February 2012.

[5] Liou M.S., A sequel to AUSM, II: AUSM^{+}-up for all speeds. J. Comp. Phys., 214:137–170, 2006.

[6] Berger, M. and Aftosmis, M.J. and Murman, S.M. Analysis of Slope Limiters on Irregular Grids. In 43rd AIAA Aerospace Sciences Meeting, AIAA 2005-0490, 10-13 January 2005. Reno, Nevada.

[7] Masquelet M.M., Large-Eddy Simulations of High-Pressure Shear Coaxial Flows Relevant for H_2 /O_2 Rocket Engines. PhD thesis, School of Aerospace Eng., Georgia Inst. of Technology, May 2013.

[8] Vingert L., Habiballah M., Vuillermoz P., Zurbach S., Mascotte, A test facility for cryogenic research at high pressure. In 51st Int. Astronautical Congress. AIAA, 2000. Rio de Janeiro, Brazil.

[9] Habiballah M., Orain M., Grisch F., Vingert L., Gicquel P., Experimental studies of high-pressure cryogenic flames on the MASCOTTE facility. Comb. Science and Technology, 178:101–128, 2006.

CROCUS SATIVUS L. AND NUTRIEPIGENOMICS: IN SILICO STUDIES OF INTERACTION WITH HISTONE DEACETYLASES

Alessia Piergentili*, Barbara Benassi and Caterina Arcangeli

Laboratory of Health and Environment. Department of Sustainability, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rome, Italy

ABSTRACT. In this report the dynamics of interaction between saffron bioactive molecules and two specific histone deacetylases will be presented. In particular, molecular dynamics (MD) simulations were performed by exploiting the high-performance parallel computing provided by CRESCO6 and using GROMACS 2019 code. The MD simulations highlight the first nutri-epigenomic capacity of saffron, supporting its definition as promising functional food.

1 Introduction

In recent decades, the scientific evidence supporting the interaction between a living organism and its environment has progressively emerged. At molecular level, the relationship between genotype and phenotype is functionally based on the epigenetic control of gene expression [1]. "Let food be thy medicine, and let medicine be thy food" Hippocrates used to say to highlight the impact of nutrition on human health, a relationship that is today investigated by the nutri-epigenomics, a science uncovering the molecular mechanisms of functional food [2, 3].

Saffron is receiving increasing attention due to its ability to protect against chronic disease [4, 5]. Although both antioxidant and anti-inflammatory functions of saffron are well defined by experimental studies, its nutriepigenomic property is still poorly understood. In this report, a part of the studies aimed at investigating the interaction between saffron bioactive molecules and key epigenetic enzymes, are reported. The molecular interactions have been investigated by exploiting a combining approach of docking and MD simulations. This approach allowed us to study in depth the binding of the ligands to the epigenetic enzymes. The results demonstrate a direct and stable interaction between saffron biomolecules and the epigenetic targets, which -according to our findings- have been divided into two categories based on complex stability.

2 Computational methods

The ligands and proteins resulting from molecular screening were parametrized for docking procedure by AutoDock Tools 1.5.6 [6], adding Gasteiger charges to saffron's biomolecule (SB) and Kollmann charges to epigenetic targets (EpiT). The best poses of the ligands in the binding site of the enzymes obtained from docking results by VINA [7] were used as the starting structures of the MD simulations. MD simulations were performed with GROMACS 5.0.7. Table 1 shows all the complexes simulated by *in-silico* experiments. The epigenetic enzymes were parametrized using the AMBER 99sb-ildn force field [8], the ligands with ACPYPE (AnteChamber PYthonParser interfacE) [9], in AnteChamber [10]. The systems were solvated with TIP3P [11] water models and neutralized with 150 mM of NaCl. Constraints are set for bonds with hydrogen atoms using the LINCS algorithm [12] and initial velocities are randomly assigned by the Maxwell-Boltzmann distribution. Unrestrained MD were carried out in NPT ensemble for 200 ns. Table 2 summarizes some details of the simulations.

*Corresponding author. E-mail: piergiaa26@gmail.com

Table 1: List of the simulated systems

Ligand	Receptor EpiT1[e]	Receptor EpiT2[f]
SB1[a]	EpiT1:SB1	EpiT2:SB1
SB2[b]	EpiT1:SB2	EpiT2:SB2
SB3[c]	EpiT1:SB3	EpiT2:SB3
SB4[d]	EpiT1:SB4	EpiT2:SB4

[a)] Crocetin, PubChem ID: CID 5281232
[b)] Crocin, PubChem ID: CID 10368299
[c)] Safranal, PubChem ID: CID 61041
[d)] Picrocrocin, PubChem ID: CID 130796
[e)] Histone Deacetylase, HDAC2, PDB ID: 4LY1
[f)] Sirtuin, Sirt1, PDB ID: 5BTR

Table 2: MD simulation parameters.

Simulations	All-atom based
Energy Minimization	Steepest descent
Short-range electrostatic and van der Waals cut-off	1nm
Cut-off scheme	Verlet
Long-range electrostatic force	PME method [13]
Equilibration	300ps
Thermostat	Bussi-Donadio-Parrinello [14]
Temperature	300K
Barostat	Parrinello-Rahman [15]
Pressure	1bar
Dielectric constant	1
Integration step	2fs
Length of simulations	200ns

3 Results and discussion

In order to evaluate the stability of the complexes and the quality of the trajectories the RMSD (Root Mean Square Deviation) analysis was performed. This analysis measures the average deviations of the Cα atoms of the protein and all the atoms of the ligand compared to a reference structure, and gives information on the conformational sampling of the complexes as a function of simulation time.
The RMSD value's trends of the EpiT1 and EpiT2 in complex with the saffron's biomolecules are shown in Figure 1 and Figure 2, respectively. All the plots show that the RMSDs stabilize around a mean value and oscillate slightly around it, with the exception of both the epigenetic receptors in complex with SB3. In particular, the high oscillations of the RMSD values observed for EpiT1:SB3 (Figure 1C) and EpiT2:SB3 (Figure 2C) indicate that these complexes sample several conformations, suggesting high

instability in terms of intermolecular interactions. Visual inspections of the structural configurations sampled by the ligands inside the binding pocket of the epigenetic targets, taken at selected time (t = 0ns and t = 200 ns) of the MD trajectory, are also reported in Figure 1 and Figure 2. From these inspections, specific molecular events, such as the removal of the SB3 ligand from the binding pocket of EpiT1 (Figure 1C) and wide conformational changes which result in the opening of the binding pocket of EpiT2 (Figure 2C), are observed. The visual inspection also confirms the stable fitting of SB1, SB2 and SB4 ligands within the binding pocket of EpiT1 (Figure 1A, B, D) and EpiT2 (Figure 2A, B, D).
On the overall the MD results underline strong affinity between EpiT1 and EpiT2 with SB1, SB2 and SB4 molecules and poor interaction with SB3.

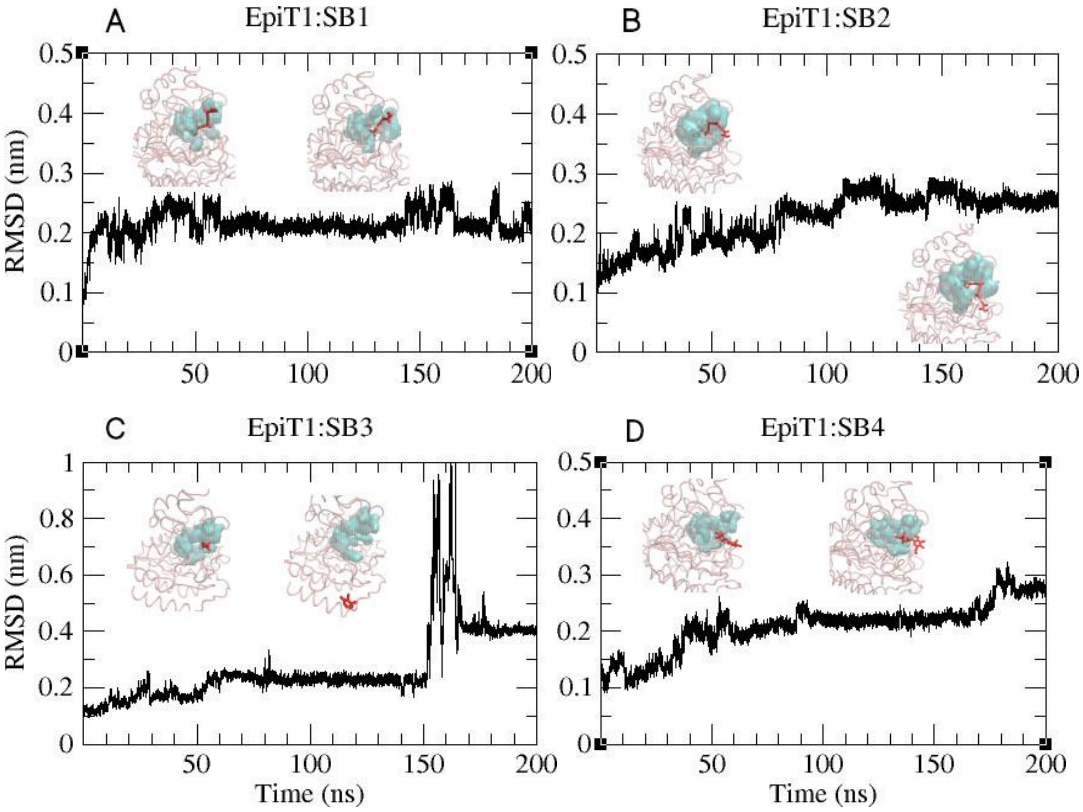


Fig.1: RMSD values, as a function of simulation time, of the systems in which EpiT1 enzyme is in complex with SB1 (A), SB2 (B), SB3 (C), and SB4 (D).
The snapshots from the trajectories are also shown; frame t = 0ns (left) and t = 200ns (right). Ligands are represented by Licorice model (red) and receptor by transparent Tube model (pink). The binding pocket is represented by transparent QuickSurf model (cyan).

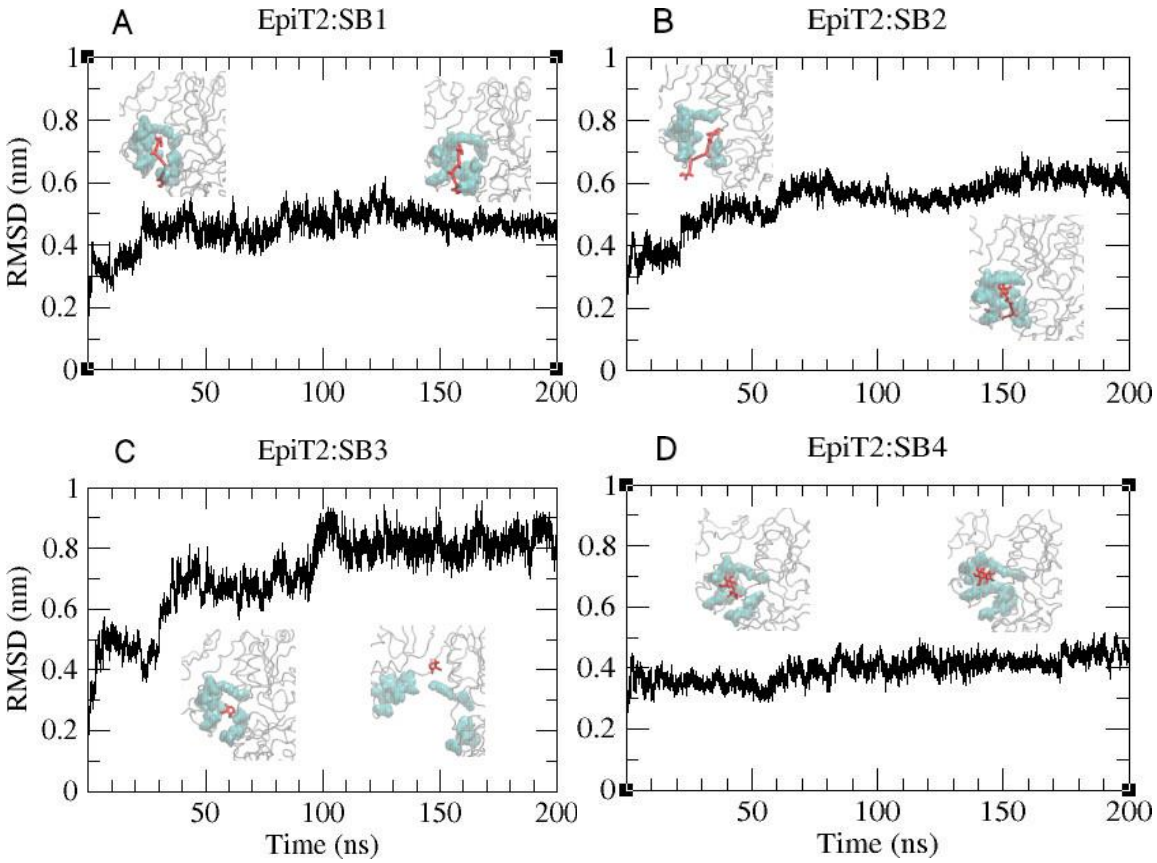


Fig.2: RMSD values, as a function of simulation time, of the systems in which EpiT2 enzyme is in complex with SB1 (A), SB2 (B), SB3 (C), and SB4 (D). The snapshots from the trajectories are also shown; frame t=0ns (left) and t=200ns (right). Ligands are represented by Licorice model (red) and receptor by transparent Tube model (silver). The binding pocket is represented by transparent QuickSurf model (cyan).

4 Conclusions

The interaction between food-derived bioactive molecules and epigenetic enzymes might alter specific gene expression, and thus affecting both human physiological and pathological processes. In this work, we exploited structural bioinformatics and MD simulations for testing the potential binding of four different bioactive molecules from saffron (ligands) with two epigenetic enzymes (targets). The binding between saffron's biomolecules and the epigenetic enzymes have been demonstrated to be stable in almost all simulated complexes, suggesting the potential nutri-epigenomic properties of saffron. However, given the limited evidences reported in the literature on the involvement of saffron's biomolecules on epigenetic mechanisms, further simulations and targeted experiments need to be performed to validate our *in silico* prediction, in the perspective of possible application in terms of therapeutic purposes as well as validation as novel functional food.

Acknowledgments

The computational work was carried out on the CRESCO (Computational RESearch Centre on COmplex systems) cluster system [16] of ENEAGRID, the infrastructure providing access to the computational resources of the ICT (Information and Communication Technologies) division, founded by the research centre ENEA (Italian National Agency for New Technologies, Energy and Sustainable Economic Development).

References

[1] Waddington, C. Organisers and genes. *Cambridge University Press*, (1940).

[2] Cencic, A. and Chingwaru, W. The Role of Functional Foods, Nutraceuticals, and Food Supplements in Intestinal Health. *Nutrients*, 2(6), pp. 611–625, (2010).

[3] McClements, D. J., Li, F. & Xiao, H. The Nutraceutical Bioavailability Classification Scheme: Classifying Nutraceuticals According to Factors Limiting their Oral Bioavailability. *Annual Review of Food Science and Technology*, 6(1), pp. 299–327, (2015).

[4] Bagur, M. J., Salinas, G. A., Jiménez-Monreal, A., Chaouqi, S., Llorens, S., Martínez-Tomé, M. and Alonso, G. Saffron: An Old Medicinal Plant and a Potential Novel Functional Food. *Molecules*, 23 (1), 30, (2017).

[5] Bathaie, S. Z., Bolhassani, A. and Tamanoi, F. Anticancer Effect and Molecular Targets of Saffron Carotenoids. *The Enzymes*, Elsevier, pp. 57–86, (2014).

[6] Michel, F. Python: A Programming Language for Software Integration and Development. *J. Mol. Graphics Mod.*, 17, pp. 57–61, (1999).

[7] Trott, O. and Olson, A. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 31, pp. 455–461, (2010).

[8] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Kleipes, J. L., Dror, R. O. and Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, 78(8), pp. 1950–1958, (2010).

[9] Da Silva, A. W. S. and Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. BMC *Res Notes*, 5, pp. 367, (2012).

[10] J. Wang, W. W., Kollman, P. A. and Case, D. A. Antechamber: An accessory software package for molecular mechanical calculations. *Journal of the American Chemical Society*, 222, pp. U403, (2001).

[11] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*., 105, pp. 1407, (1983).

[12] Hess, B., Bekker, H., Berendsen, H. J. C. and Fraaije, J. G. E. M. LINCS: a linear constraint solverfor molecular simulations. *J. Comp. Chem.*, 18, pp. 1463–1472, (1997).

[13] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. and Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103, pp. 8577–8592, (1995).

[14] Bussi, G., Donadio, D. and Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126, (2006).

[15] Parrinello, M. and Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52, 7182–7190, (1981).

[16] Iannone, F., Ambrosino, F., Bracco, G., De Rosa, M., Funel, A., Guarnieri, G., Migliori, S., Palombi, F., Ponti, G., Santomauro, G. and Procacci, P. CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 International Conference on High Performance Computing Simulation (HPCS), pp. 1051–1052, (2019).

INTEGRATED SURVEY AND 3D PROCESSING ON ENEA CRESCO: CASE STUDY OF SAN NICOLA IN CARCERE IN ROME

Marco Puccini[1*], Antonio Camassa[2], Marialuisa Mongelli[1], Samuele Pierattini[1], Silvio Migliori[1], Marco Canciani[2], Giovanna Spadafora[2] and Mauro Saccone[2]

[1]*ENEA, TERIN ICT, Lungotevere Thaon di Revel, 76 - Rome, Italy*

[2]*Department of Architecture - University Roma Tre, Largo G. B. Marzi, 10 - Rome, Italy*

ABSTRACT. This work shows the experimental activities carried out by a joint research group on the case study of the Basilica di San Nicola in Carcere in Rome, concerning the integration of different 3D reconstruction technologies in the Cultural Heritage domain. Dataset obtained from laser scanner and photogrammetry, were processed on a remote platform hosted on the powerful ENEA computational infrastructure, in order to analyse the dense clouds, to compare them and to improve the quality of the data. There are also showed the efforts to upgrade this platform and the optimisation of the processes.

1 Introduction

In the past few years, a growing interest in the field of digital reconstruction techniques and related analysis processes has marked out the Cultural Heritage studies. Thanks to those tools it is possible to share knowledge in a more efficient way, allowing both collaboration among researchers and remote access to heritage for the public. For this reason, a good set of performant software and hardware tools that permits 3D reconstruction and related information enrichment are fundamental. The activities described in this work concern the optimization of the ecosystem components dedicated to the acquisition of datasets from instrumental surveys carried out with different methodologies and the processing of three-dimensional analytical models to support research activities. The optimization of the It@cha [1] platform on ENEAGRID/CRESCO [2] infrastructure is part of the research project promoted by Enea and Roma Tre University for joint studies on Cultural Heritage. The objective of the collaboration, born within the framework of the DTC Lazio [3] in 2019, is to enable the acquisition, management and presentation of digital twins of cultural heritage assets, through advanced software services integrated in a single digital ecosystem dedicated to the knowledge of cultural heritage.

2 Methodology

It@cha "Tecnologie Italiane per applicazioni avanzate nei Beni Culturali" is the platform developed by ENEA for the project financed by the National Operational Program (PON) "Ricerca e Competitività 2007-2013". It is hosted on the ENEAGRID/CRESCO HPC infrastructure, managed and developed by TERIN-ICT, connected in a Wide Area Network. It@cha is a system that allows the remote use of software for 3Dprocessing and modelling by exploiting the computing resources of a pool of machines of the infrastructure, dedicated to graphic applications. In particular, the platform offers the possibility to obtain photogrammetric reconstructions through the software Agisoft Metashape, their manipulation through the software Meshlab (developed by ISTI-CNR) plus other open-source software for three-dimensional graphic processing. The remote control of these applications is managed by a software component of the platform that redirects the rendering by exploiting the resources of the remote servers. This therefore allows both hard processing and rendering of the graphical environment. The pool of machines serving this platform is shown, with their main characteristics, in Table 1.

[*]Corresponding author. E-mail: marco.puccini@enea.it

Table 1: It@cha platform hardware characteristics.

Hostname	CPU	Memory	GPU
cresco6-nvi1	2x Intel Xeon Platinum 8160 2.10GHz 48 cores	192 GB	Nvidia Tesla V100 32GB
cresco6-nvi2			
cresco4-nvi1	2x Intel Xeon E5-2680 2.80 GHz 20 cores	64 GB	2x Nvidia K40m 12GB

LSF (Load Sharing Facility) is the software component of the infrastructure that has the task to manage the job queues to allow a proper usage of computational resources and to avoid system overloads. The optimization of the It@cha platform concerned the workflows through the monitoring of computing resources during the massive use of 3D reconstruction and 3D modelling software. This allowed us to identify strategies to improve the performance of graphic software on the platform and consequently enhance their analytical capabilities.

3 Case Study

The Basilica of San Nicola in Carcere (Fig. 1) stands on the archaeological ruins of three temples built in the *Forum Olitorio* between the third century BC and the second century BC, dedicated to *Juno Sospita*, *Janus* and *Spes*, still visible outside the church [4].

Fig.**1**: The Basilica of San Nicola in Carcere (left), the crypt (right).

The first structure of the Basilica was built in this area in the 6th century on the ruins of the Forum Olitorio and received its name in carcere from a prison present there, the *Carcer ad Elephantum*. The latter, which can still be visited, is in a significant state of deterioration due to the thermo-hygrometric situation that compromises the preservation of the superficial layer of the masonry and does not allow a precise stratigraphic reading. Therefore, it was decided to concentrate the surveys, and the subsequent elaborations, in the crypt.

Thanks to the instrumentation provided by Riltec (Department of Architecture - University of Roma Tre - coord. Prof. Marco Canciani), the laser scanner survey of the entire church of San Nicola represented the first phase of acquisition of metric and dimensional data. It has also allowed us to understand how the volumes of the upper part of the church insisted on the underground rooms of the crypt, revealing the structural continuity of the two overlapping environments, particularly important to assess any phenomena of rising damp to the environments of the church. Subsequently, two photogrammetric survey campaigns have been carried out, with the aim of elaborating, through the SfM methodology [5], a dense cloud and detailed orthophotos of the masonry inside the crypt.

3.1.1 Monitoring and optimization of workflows

The interoperability between software in the It@cha platform has allowed us to define a pipeline for the management of point clouds obtained from the processing of data sets acquired. During the software processing phases, the metrics of the hardware computing resources used were monitored in order to

estimate the performance of each step of the SfM algorithm. In particular, data related to CPU, memory and GPU usage were monitored. Among other tests, just for example of the analysis typologies, a GPU impact evaluation was carried out. This test is particularly important due to the fact that not every step of the SfM algorithm uses the graphical acceleration. This could bring no advantages and, in some cases, could even worsen the entire processing time. To evaluate this impact, the elapsed time for each SfM step has been measured both with and without GPU, as shown in Fig. 2.

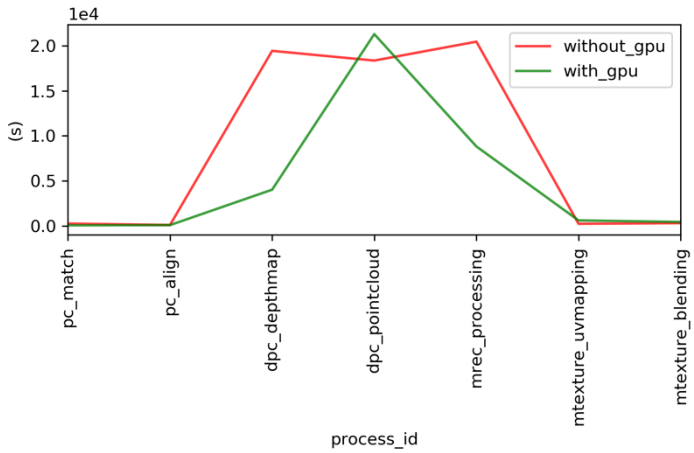


Fig.**2**: SfM single step elapsed time (s).

This measure had allowed to define an index, the *change percentage* p_c which has been calculated with the formula:

$$p_c = \frac{t_A - t_N}{t_N} \cdot 100 \qquad (1)$$

Thanks to this index and other analyses a set of optimized workflows have been designed and an on-platform jobs submission policy has been defined. With these improvements a more focused attention has been moved to dense cloud comparison.

3.1.2 Geometric analyses

The computational power of the platform has allowed it to elaborate the point cloud, starting from the data acquired during the photogrammetric survey, with the maximum accuracy allowed by the Agisoft Metashape software, considerably reducing the calculation time. The dense point cloud thus obtained was compared with the points acquired by laser scanner, in order to verify the geometric and dimensional fidelity of the survey carried out with SfM methodology. The comparison between the two dense clouds was made through the software Cloud Compare (Fig. 3), through which it was possible to calculate the square deviation between the two clouds, then display the areas where the distance between the two clouds was minimum and maximum.

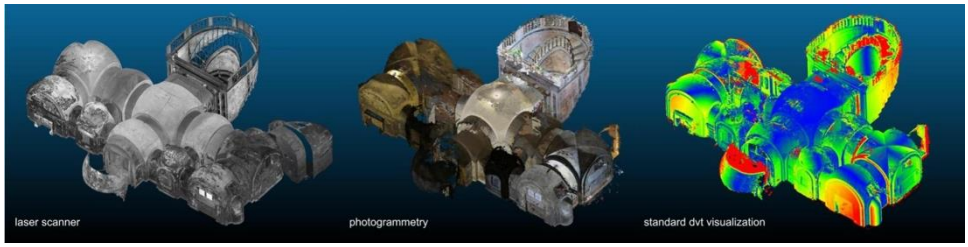


Fig.**3**: The comparison through Cloud Compare software of the clouds obtained from laser scanner and SfM (RMS difference 1,4 cm).

Subsequently, through the calculation of the normals and the analysis of their dip and direction, it was possible to perform analysis of the material consistency of the wall surfaces of the crypt. The objective was to isolate materials and patterns coherent with each other to facilitate the stratigraphic reading through the recognition of the surface characteristics of the materials (Fig. 4).

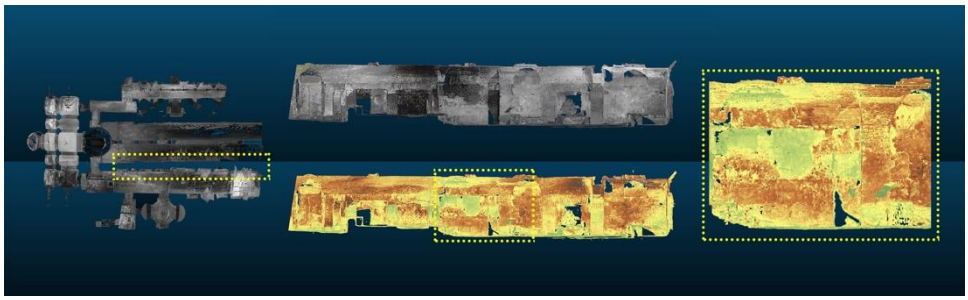


Fig.4: Computation of the normal, highlighting the differences between the masonry patterns.

3.1.3 The Open Digital Twins Ecosystem

A cloud native web app to allow data management is under development [6]. This tool is the module of a wider and more innovative ecosystem with a micro-services architecture named *Open Digital Twins*. Such ecosystem is designed to flank the It@cha platform for what concern data storage and visualization. A first Proof of Concept (Figure 6) of this module allow to upload, store and visualize data such as publications, images and 3D reconstructions models with their related metadata (including the geographical references). While these last one are stored in a modern non-relational database that allow even full text searches, the data are stored in a distributed storage system to ensure their availability and persistence. With this tool, research products can be easily stored, shared and enriched among researchers' community, students and stakeholders.

4 Perspectives

The possibility to perform all the software elaborations in a single environment that can be managed remotely, the identification of an optimised pipeline for the software elaborations and the calculation capacity ensured by the It@cha platform, allow to outline some of the future goals: through further experimentation it is foreseen the implementation of further opensource software steps [7] for the dimensional and geometric analysis, the comparison and the visualization in multi-resolution mode of the point clouds, as well as the automatic recognition of surface patterns and constructive elements.

Acknowledgments

All the activities described above are carried on the ENEAGRID/CRESCO infrastructure. Thanks to the fundamental support of ENEAGRID/CRESCO team in terms of computing resources, networking and know how it was possible to carry out this work.

References

[1] http://www.progettoitacha.it/
[2] https://www.eneagrid.enea.it/
[3] https://dtclazio.it/
[4] Squadrilli T., Vicende e monumenti di Roma (Rome: Staderini Editore) 1961 p.338
[5] Gruen A, Development and status of image matching in photogrammetry Photogrammetric Record 2012 27 (137) pp. 36-57
[6] Canciani M, Saccone M, Spadafora M, Migliori S, Mongelli M, Puccini M, Quintiliani A, Gallia A, MasettiC 2020 Modelli 3D e dati GIS: una loro integrazione per lo studio e la valorizzazione dei beni culturali, Archeomatica (Mediageo) 11pp. 18-23
[7] Remondino F et al., Low-cost and open-source solutions for automated image orientation – A critical overview Lecture Notes in Computer Science 2012 7616 pp. 40-54

MONTE CARLO SIMULATIONS FOR TDCR METHOD USED IN NUCLEAR MEDICINE APPLICATIONS.

Francesco Tortorici[1,2*], Farmesk Abubaker[1,2], Marco Capogni[3], Concetta Sutera[1,2], Vincenzo Bellini[2] and Guido Guarnieri[4]

[1]*Università di Catania, Physics Department, Via S. Sofia 64, IT-95125 Catania, Italy*

[2]*INFN Sezione di Catania, Via S. Sofia 64, Catania, Italy*

[3]*ENEA, Fusion and Technology for Nuclear Safety and Security Department*
Istituto Nazionale di Metrologia delle Radiazioni ionizzanti (INMRI), C.R. Casaccia, Roma, Italy

[4]*ENEA TERIN-ICT-HPC, C. R. Portici, Portici (NA), Italy*

ABSTRACT. This project aims at solving, by means of Monte Carlo simulation tools, detection efficiency issues of the portable Triple-to-Double Coincidence Ratio (TDCR) detector of the Italian National Institute of Ionizing Radiation Metrology (INMRI) belonging to ENEA. This detector can be used, for direct activity measurements of pure-beta emitting and/or pure electron capture radionuclides, as a traveling instrument for in-situ activity measurements with particular benefits in many applications such as in nuclear medicine. In this report we focused the attention on ^{18}F detection efficiency computation performed for the portable TDCR using GEANT4 as Monte Carlo simulator and the ENEA CRESCO computing facility. The use of such a heavy-duty facility was needed in order to collect in just a few months of runtime a high statistics (millions of events) for radionuclides emitting secondaries at high- (order of MeV) and low- or very-low- energy, an otherwise even-larger time-consuming task. The interest for this work is also due to the fact that ^{18}F is widely used radio-pharmaceuticals in nuclear medicine for Positron Emission Tomography (PET) applications and PET scanners require robust and reliable calibration which can be performed directly on site by the portable TDCR counter.

1 Introduction

The applications and developments of nuclear medicine (NM) have been an interest of nuclear science for almost half a century. In recent years, a wide variety of radionuclides have been used in NM with different and complex decay schemes. Therefore, the accuracy for activity measurements significantly requires standardization of radioactivity [1]. Precise and accurate radioactivity measurements are considerably difficult due to the complexity in the kind of radionuclides used and their decay characteristics (decay rates, type of particles emitted and their energies and intensities). In addition, one cannot use the same detector or method for measuring all radionuclides. For that reason, standardization measurement of a particular radionuclide implies a specific study of its decay scheme and the measurement technique for their interaction with matter [2]. At the Italian National Institute of Ionizing Radiation Metrology (INMRI), belonging to ENEA and located in Casaccia Research Center, the TDCR method is used for absolute activity measurements and standardization of pure electron capture (ec) or pure beta emitting radionuclides [3]. A portable TDCR detector, made up of three photomultipliers (PMTs) working in coincidence and symmetrically arranged around a specially designed optical chamber, is available at ENEA-INMRI. It allows for the measurement of radioactive material in solution of liquid scintillator, contained in a vial which is put inside the optical chamber. It is used for standardization of pure beta emitters and electron capture radionuclides [4] [5].
The portable TDCR counter was completely designed and built in 2013 at ENEA-INMRI [6] with identical performance as the TDCR counters operated for activity primary measurements in metrology

*Corresponding author. E-mail: francesco.tortorici@ct.infn.it.

[3]. This TDCR counter can be used as an on-site device for environmental measurements or directly at the hospitals to measure the activity of a radiopharmaceutical injected into patients, such as ^{18}F. To equip the portable TDCR with light and portable electronics, a collaboration with CAEN and ENEA-INMRI was established to figure out the best front-end electronic acquisition system for the counter [6]. Furthermore, the detection efficiency of the TDCR counter is mainly based, for a defined geometry, on the PMTs quantum efficiency; in particular, its computation in a classical approach relies on the theory of the free-parameter which represents the number of emitted photons per unit energy released by the ionizing particle or radiation in the scintillation solution [7][8].
The activity of the radionuclides can be measured using the triple and double coincidence between PMTs tubes based on the triple and double coincidence detection efficiency of the TDCR. To find the detection efficiencies of double and triple coincidences, a statistical model of the light emission and the experimental value of TDCR are needed. Indeed, a stochastic approach is explored as an alternative to this classical route, by implementing a complete representation and modeling of the TDCR counter drawn on the Monte Carlo simulation Geant4 code [9] [10]. Geant4 code has the ability to simulate different physical processes associated with the charged particle or photon's transportation inside matter and generate optical photons resulting from the interaction between charged particles with the scintillator in the liquid scintillation counter (LSC). This approach allows to obtain TDCR model, taking into account the different reflection and refraction processes which take place at the interfaces inside the optical chamber such as: vial to air, air to PMT-window, etc. Practically, TDCR detection efficiency is affected by an unknown quenching effect of the liquid scintillator. To take under control this effect, the influence of the change in detection efficiency on the estimated activity of the radionuclide in solution is performed by the detection efficiency variation technique carried out using coaxial grey filters, chemical quenching, and/or PMT defocusing [8]. In this project the simulated detection efficiency variation of TDCR counter has been examined for high-energy beta-emitters (^{18}F). The detection efficiency of the TDCR counting system for pure-beta emitters, such as ^{18}F experimentally measured, is compared with the GEANT4-simulated TDCR parameter.

2 Short description of the ENEA-INMRI portable TDCR counter

TDCR method is considered a powerful method for absolute activity measurement for pure-beta and pure electron-capture radionuclides in LSC. In this project the portable TDCR at ENEA-INMRI is used [11]. It consists of an optical chamber counter which has an inner prismatic shape with equilateral base (H = 73 mm and L = 60 mm). The optical chamber is made from white PTFE (Teflon) and it is surrounded by a black Teflon outer cylindrical box shape of ($\emptyset$ = 150 mm and H = 150 mm). The three Hamamatsu R7600U-200 square package type photomultipliers (PMTs) arranged symmetrically 120^{o} around the LS vial in the optical chamber [6]. The selected PMTs have a high quantum efficiency, high gain, short time of response, comparatively wide-range wavelength of about 300–650 nm, with relatively low supply voltage (about 900 V) with the grounded cathode [11]. The PMTs operated in photon counting mode [11]. The high-power electronics device is necessary to record the fast signals for the TDCR applications. For such a reason, the DT5720 CAEN digitizer is connected directly without any preamplifier to the output signals of the PMTs. The DT5720 CAEN digitizer provides a fast and precise 12 bit and sampling rate Analog Digital converter (ADC) 250 MSample/s within the time acquisition window. The CAEN digitizer is PC remote-controlled by the CAEN Control Software which runs on Windows or Linux operating systems; the data acquisition parameter (gate length, DC offset, pulse polarity, threshold level, etc.) for each channel can be set independently by the software [6].

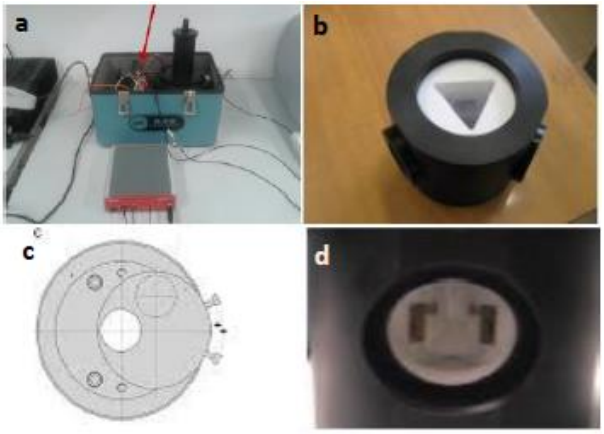


Fig. 1: (a) The portable TDCR system at ENEA. (b), (d) The external and the internal view (up and down) respectively of the optical chamber (c) Optical shutter design

3 The TDCR system setting and analysis Software

The precise adjustment of the discrimination above the noise for each one of the three PMTs and threshold under the single-electron peak (SEP) are important settings of the TDCR method [12]. The SEP signal is generated by each PMT in the portable TDCR counter. The setting of the data acquisition parameters is established and recorded by the CAEN digitizer [11]. For each individual channel, the control software of the digitizer is used to fix the energy discrimination threshold and the width of the ADC gate (typically 24 ns). This setting allows to minimize the noise and process the entire analog pulse which arrives from the photomultiplier tubes (PMT), and which is covered fully in time by the gate pulse [11]. In order to have the same gain of the three PMTs in different channels, one can check the SEP for each PMT in order to be located at the same histogram channel. To achieve this goal, the DT5720 digitizer records a typical SEP spectrum from each PMT of the TDCR counter without a vial inside the optical chamber. The CAEN digitizer was used in histogram mode [11]. The CAEN digitizers DT5720 are dead-timeless acquisition devices; the software is written and customized in C++ [6]. The dead-time is managed by the digitizer which is provided by CAEN [6]. In the portable TDCR system, signals come from each PMT, and the digitizer manages the off-line coincidence analysis between them by emulating the MAC3 logic [13]. For the off-line analysis of the portable TDCR system, a software was implemented in CERN ROOT frame [13]. When the CAEN digitizer records and organizes data, they are read event by event by the ENEA-INMRI software as “leaves" of a tree in the ROOT output file. In fact, by using the DT5720 digitizer it is possible to read the Trigger Time Tag (TTT) and integrated charge (Q) values for each event. In addition, for defined dead-time, t_{dead}(μs), and coincidence resolving time t_c(μs) values, one can compute the off-line TCDR analysis for the single counts of each PMT (A, B, and C), the coincidence counts (AB, BC, AC, D and T), the sum of them (S), the live time, t_{live}, the real time, t_{real}, and the TDCR parameter [11].

4 Experimental Measurements with TDCR

The TDCR parameter was measured by using the ENEA portable TDCR detector equipped with the CAEN DT5720 digitizer for different pure beta-emitting radionuclides allowing to characterize the counter for a set of the TDCR parameters. The selected set of pure beta emitters is interesting for radionuclide standardization and activity measurement in nuclear medicine applications and/or for other kinds of applications, such measurements on nuclear site or in the environment. The TDCR analysis was performed by recording the integrated charge Q value with an ADC gate equal to 24 ns. The long-lived beta sources used were prepared for other past experiments at the Radiochemistry laboratory of the ENEA- INMRI; gamma-impurity check was performed on the master solutions of these beta sources by means of a high energy resolution HPGe detector owned by ENEA-INMRI. The sources were made, as usual, of 10 ml Ultima Gold (UG) as liquid scintillator and approximately 10 mg of radioactive solution contained in 20 ml borosilicate glass vials. At the first, for background measurements one blank source was prepared containing only 10 ml of UG in the previously mentioned geometry. A customized data

acquisition software was implemented in order to perform a TDCR analysis on the events which are collected in a list-mode file; corresponding events from different input channels can then be compared to their TTTs when each of the acquisition channels of a digitizer are synchronized. After the run ends up, the data acquisition software scans the consequent event lists checking for coincidences. The recorded data stream is then analyzed by the ENEA-INMRI analysis software and the values of dead time (t_{dead} = 50 μs) and coincidence resolving time (t_c = 140 ns) are applied. The ratios between the background-subtracted or net count rate, $R_T = T/t_{live}$, of Triple (*T*) and the ones, $R_{AB}=AB/t_{live}$, $R_{AC}= AC/t_{live}$ and R_{BC}, $R_{AC}= BC/t_{live}$ of the double coincidences (*AB, AC* and *BC*) are calculated for each set of data. They were used to compute the efficiencies for the triple coincidences ε_T and for the logical sum of the double coincidences ε_D for each radionuclide measured with the TDCR counter. TDCR parameter can be calculated experimentally by taking into account the ratio of the net count rate of triple coincidence to the logical sum of double coincidence, $R_D = D/t_{live}$, and it is equal to the ratio of the efficiency of triple coincidence ε_T to the efficiency for the logical sum of the double coincidences ε_D for high statistics conditions of light emissions (i.e., for high-energy beta emitters) [14].

5 The TDCR-GEANT4 modeling based on LS

The Monte Carlo Simulation (MCS) technique has overcome the shortcoming of the numerical method used for simple geometries based on the radiation transport equations [15]. Although MCS has been used for a few decades in many areas of experimental physics such as elementary particles physics and high energy physics, it has only gained interest in radiation physics in recent decades due to the great advance of computing power. MCS in radiation physics simulates the transport of electrons, positrons, and photons in materials with arbitrary compositions [16]. In this project, Monte Carlo Geant4.9.6 p04 code was selected for its complete set of tools permitting accomplish modeling of the TDCR counter, simulating the different physical processes ongoing in the optical chamber and determining the TDCR parameters for different radionuclides sources in LSC. In LS counting, Monte Carlo can be used to calculate the probability for the interaction of beta or electron capture particles in a given scintillator and the absorbed energy distribution [17]. The photon absorbed energy distribution is mostly deliberated in discrete energy bins and the quenching function is applied to each bin midpoint [15]. The "low energy" package on the basis of Livermore data was selected for simulating photons and electrons which are created by Bremsstrahlung [13]. In order to achieve a realistic simulation, an explanation of the optical properties related to the materials of the optical chamber is carried out comprehensively: the ultra bialkali photocathodes, the borosilicate vial glass loaded with liquid scintillator, the fused silica window of the PMTs, and the reflecting cavity of Teflon is about 95% [18]. Regarding the TDCR detection efficiency calculation, the extension to LS counting needs the exact description of the borosilicate vial loaded with 10 ml of Ultima Gold LS cocktail. The Ultima Gold atomic composition and the dimensions of the vial are taken from the experimental set up. The borosilicate glass vial is used and the dispersive refractive index for the borosilicate material is 1.52 at wavelength of 400 nm [18]. The borosilicate vial is placed inside the optical chamber, the model of dielectric – dielectric applied for photon reflection and refraction at the inner and outer scintillation vial (borosilicate) vial surfaces [13]. In order to calculate the production of photoelectrons each time when the photons reflect at the interface with the ultra bialkali photocathode, one can consider the various optical processes happening within the fused silica of the photomultiplier window. The center of the optical chamber is positioned 17 mm from the outer surface of the photomultiplier window; for that reason, the dielectric–dielectric boundary model is used. The range of PMT spectral sensitivity from 300 nm to 650 nm has a peak at 400 nm. The fused silica window is covered by the ultra bialkali photocathode and the model of dielectric–dielectric boundary is used [18]. Quantum efficiency is provided by using binomial trials for converted photons to photoelectrons in the photocathode [18]. The maximum quantum efficiency is ~ 43% at 430 nm [19]. The double- and triple- coincidences are determined for each photomultiplier tube by counting the numbers of photoelectrons. The probability for photoelectrons to reach the first dynode depends on the defocusing parameter; it permits the detection-efficiency variation to be simulated as experimentally found by PMTs defocusing [18]. The relationship between triple- and double-coincidence detection efficiencies and TDCR value can be determined by applying the defocusing parameter, and finally the

activity of the source can be calculated.
Performing MCS in Liquid scintillation counter of ionizing radiations losing their energy in the scintillator is time consuming to obtain very high statistics, especially for simulations concerning very low beta emitters such as ^{3}H and ^{63}Ni. For these reasons the ENEA CRESCO computing facility was taken into account to carry out such a simulation. In particular, to test the run time necessary to perform the simulations and considering the importance of the use of the portable TDCR detector for nuclear medicine applications, the ^{18}F radionuclide was studied for all the tests performed with ENEA CRESCO. Hereafter a short description of the jobs run on ENEA CRESCO.

6 Jobs on CRESCO

The jobs run simulations in Geant4 of the TDCR apparatus as described above.

CRESCO system was accessed from the following front-end nodes, with username "radionuc" belonging to ENEA-INMRI:
cresco6x001.portici.enea.it
cresco6x002.portici.enea.it

Due to specific library requirements of the MonteCarlo code, it was necessary to install Geant4 version 9.6 on
/afs/.enea.it/project/metrologia/soft/geant4_cresco6
and root 6.20.04 (thanks to one of the author, i.e G.G) in
/afs/.enea.it/project/metrologia/soft/root_cresco6

The jobs, running up to about 135 hours each (depending on the number of events, and type and energy of the primary particles) are submitted to CRESCO by bsub on typically 40 cores. Their main output files, in ROOT format (of the order of tens of MB) for offline analysis, are produced on the scratch area $PFS_SCRATCH0 and then moved to the store area $PFS_POR_STORE0 using appropriate shell scripts only if successfully completed. LSF error and output files are written to AFS. In this project the TDCR parameter for radionuclide ^{18}F has endpoint energy about 634.9(5) KeV [20] as shown in Table 1, and single charged particles such as electron and positron with high energy equal to 1 MeV for different number of events starting from (1000, 10 000, and 100 000) events for each charged particles e-/e+ is computed by using the same code. The defocusing parameter and Birks constant kept the same for both particles 0.97 and 0.1 mm/MeV respectively as shown in Table 2.

Table 1: Simulated TDCR parameter for ^{18}F radionuclide

Radionuclide	Defocusing parameter	Birks Constant	No. Of events	ε_D	ε_T	TDCR ratio
^{18}F	0.95	0.1 mm/MeV	10 000	0.9999	0.9991	0.9992

Table 2: Simulated TDCR parameter for high energy 1 MeV of e-/e+ for different number of events

Particle	Energy (MeV)	No. of event	1000	10 000	100 000
Electron	1	ε_D	1	0.9999	0.99969
		ε_T	0.999	0.9996	0.99893
		TDCR ratio	0.999	0.9997	0.99924
Positron	1	ε_D	0.493	0.4999	0.49667
		ε_T	0.387	0.3901	0.38897
		TDCR ratio	0.785	0.7804	0.78316

It is clear from the Table 2, positron has a significantly low value of logical sum of double efficiency ε_D, triple efficiency ε_T and TDCR parameter compared to the electron. In other words, the unexpected

asymmetry is observed in the logical sum of double efficiency ε_D, triple efficiency ε_T, and TDCR parameter between e^- and e^+. This means that the version of the TDCR-Geant4 code used to perform the simulations does not work properly for β^+ radionuclide emitters or e^+ charged particles. For that reason, the study on ε_D, ε_T and TDCR parameter is still under investigation and some actions are required in the Physics List (G4PhysicsList) of Geant4 code used to solve this problem of the evident asymmetry registered for the two kinds of particles. New run jobs are therefore expected on ENEA CRESCO on this kind of problematic which are still under study.

Acknowledgments

This work was funded by the University of Catania cooperation with ENEA-INMRI, INFN and the CAEN S.P.A in the framework of the National Operational Program on Research and Innovation PON 2014-2020. PON was carried out with the funding of the European Union. The authors are grateful to Dr Pierino De Felice, Director of the ENEA-INMRI, for supporting the cooperation among different Institutions.

References

[1] Pochwalski, K., Broda, R., Radoszewski, T., "Standardization of pure beta emitters by liquid scintillation counting", Appl. Radiat. Isot, vol. 39, p. 165–172., 1988.
[2] M. L'Annunziata,, Handbook of Radioactivity Analysis, Press, ISBN: 0-12-436603-1., Academic press, 2003, p. Second Edition.
[3] Mo. L, in Standardization of pure β and β–γ emitting radionuclides in various physical forms, Sydney, University of Sydney, 2007, p. Degree of Doctor of Philosophy.
[4] Grau Malonda, A. and Coursey, B.M., "Calculation of beta-particle counting efficiency for liquid-scintillation systems with three phototubes," Appl. Radiat. Isot., vol. vol 39, no. p1191-1196, 1988.
[5] Broda R., Cassette P., Kossert K. "Radionuclide metrology using liquid scintillation counting", Metrologia, vol. 44, no. S36-S52, (2007).
[6] Mini, G., Pepe, F., Tintori, C., Capogni, M. "A full digital approach to the TDCR method," Applied Radiation and Isotopes, vol. 87, no. 166-170, 2014.
[7] Birks, J.B., "Scintillations from Organic Crystals: Specific Fluorescence and Relative Response to Different Radiations," Proc. Phys. Soc. A, vol. 64, no. 874, 1951.
[8] Cassette, P., Broda, R., Hainos, D., Terlikowska, T., "Analysis of detection-efficiency variation techniques for the implementation of the TDCR method in liquid scintillation counting," Appl. Radiat. Isot, vol. vol 52, no. p643-648, 2000.
[9] S. Agostinelli, J. Allison, et al., "Geant4 - a simulation toolkit," NIMA, vol. vol. 506, no. no. 3, p. pp. 250 – 303, 2003.
[10] J. Allison et al., "Geant4 developments and applications," IEEE Transactions on Nuclear Science, vol. vol. 53, p. pp. 270–278, Feb 2006.
[11] Capogni, M., DeFelice, P. "A prototype of a portable TDCR system at ENEA", Applied Radiation and Isotopes, vol. 93, pp. 45-51, 2014.
[12] Xiaolin Hou, 2018. Liquid scintillation counting for determination of radionuclides in environmental and nuclear application. Journal of Radioanalytical and Nuclear Chemistry. 318, 1597–1628(2018).
[13] Cassette, P., Bouchard, J., "The design of a liquid scintillation counter based on the triple to double coincidence method", Nucl. Instrum. Meth, vol. A 505, p. 72–75, 2003.
[14] Broda R., "A review of the triple-to-double coincidence ratio (TDCR) method for standardizing radionuclides," Appl. Radiat. Isot., vol. vol 58, pp. p585-594, 2003.
[15] D. Rodrigues, P. Arenillas, M.E. Capoulat, C. Balpardo, "General data analysis code for TDCR liquid scintillation counting," Applied Radiation and Isotopes, vol. 66, p. 1049–1054., 2008.
[16] S. Guatelli, D. Cutajar, B. Oborn and A. B. and Rosenfeld, "Introduction to the geant4 simulation toolkit 2011," institutional repository for the University of Wollongong, 2011.
[17] V. N. Ivanchenko, M. Maire and L. Urban, "Geant4 standard electromagnetic package for HEP applications,", in Proc. Conf. Rec. 2004 IEEE Nuclear Science Symposium. Rome, Italy, N33-165 (2004).
[18] Thiam, C., Bobin, C., Chauvenet, B., Bouchard, J., "Application of TDCR-Geant4 modeling to

standardization of 63 Ni", Applied Radiation and Isotopes, vol. 70, pp. 2195-2199., 2012.
[19] Hamamatsu photonics. K.K., "Photomultiplier tubes and assemblies for scintillation counting and high energy physics," (Online). Available: https://seltokphotonics.com/upload/iblock/310/310a7014996a49206f14ede96e70a917.pdf. (Accessed 24 05 2021).
[20] F-18_tables.pdf (nucleide.org)

FLUX AND ENERGETIC DISTRIBUTION CHARACTERIZATION OF NEUTRONS IN THE TAPIRO'S TANGENTIAL CHANNEL. SIMULATIONS IN SUPPORT OF EXPERIMENTAL MEASUREMENTS.

Nunzio Burgio and Alfonso Santagata*

ENEA, FSN-FISS-RNR, C.R. Casaccia, Roma, Italy

ABSTRACT. The experimental campaign AOSTA (Activation of Osmose Samples in TAPIRO) planned in the tangential channel of the ENEA's nuclear research reactor R.S.V. TAPIRO requires an accurate measure of the energy distribution and flux intensity of neutrons in some positions of the tangential channel. Preliminary simulations are mandatory for a better definition of the measurement protocol for the selected irradiation positions. Monte Carlo neutron transport simulations were performed with the parallel executable of MCNP code on the CRESCO cluster. The results of two different statistical estimators of neutron flux are compared, exhibiting a good agreement in the region of low statistical fluctuation. The simulations and their outcomes are briefly explained in the present work.

1 Introduction

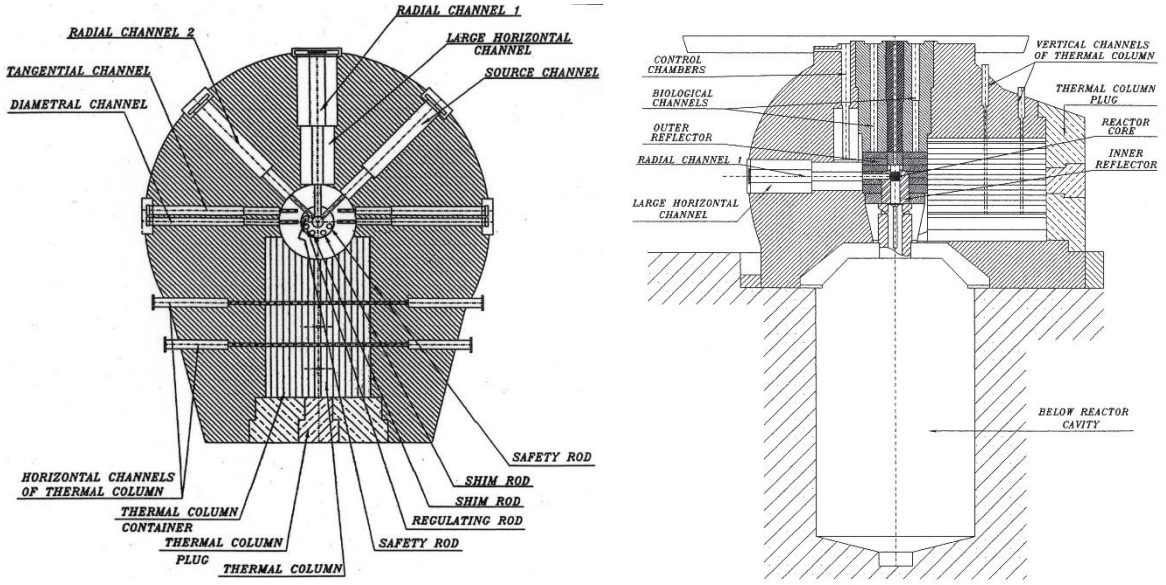
The experimental campaign, expected in the AOSTA (Activation of Osmose Samples in TAPIRO) program, activity performed in collaboration with researchers of C.E.A., is subdivided into two phases. In the first phase, several positions of the tangential channel will be selected for neutron spectrum and flux measurement with the multi-foils activation technique. In the multi-foils technique, the neutron flux and spectrum are obtained, unfolding the values of the residual gamma emission of several metallic foils activated by neutrons. The outcomes of the neutron characterization campaign will define the operative conditions of TAPIRO, power and irradiation duration of the second phase. In the second phase, some samples containing ^{241}Am, will be positioned in the tangential channel to measure the capture cross-section. The preparation of the first phase of the experimental campaign needs some preliminary simulations for selecting the foils (dimension and type) and defining the irradiation characteristics (reactor power and irradiation time). The adequacy of a specific foil is deduced by comparing the activity calculated with the estimated neutron spectrum and flux intensity and the minimum gamma emission (background subtracted) measurable with "acceptable" relative error with an HPGe (High Purity Germanium) detector. The neutron energy distributions and fluxes intensity in the irradiation positions were estimated by the Monte Carlo transport code MCNP [1] with a simplified model of the TAPIRO reactor as input. Low values of the statistical error inherent in the Monte Carlo technique are reached utilizing the CRESCO cluster, where the OpenMPI parallel executable of MCNP code version 6.2 is installed. In the following, it is reported a brief description of the TAPIRO reactor, the MCNP model and the results of calculations performed with two types of estimators. Finally, a description of the utilization of the results for applying the multi-foils technique is reported.

2 R.S.V. TAPIRO research reactor

R.S.V. TAPIRO (Reattore Sorgente Veloce TAratura PIla rapita potenza 0), located in the ENEA Casaccia Research Center of Rome, is a fast neutron irradiation facility [2]. TAPIRO has a nominal thermal power of 5 kW, and the neutron flux at the core center is about $4 \cdot 10^{12}$ n/cm^2/s. In figure 1, the main components of the reactor in a horizontal and vertical section are highlighted. The cylindrical core is constituted by a U-Mo (98.5% U, 1.5% Mo) alloy is surrounded by about 67 cm of a copper reflector and a biological shield made of heavy borate concrete with a thickness of about 170 cm. The core is cooled by Helium that flows in a gap between the fuel clad's lateral surfaces and the reflector's inner surface. The control rods system, housed in the copper reflector, is constituted by five movable

*Corresponding author. E-mail: alfonso.santagata@enea.it.

cylindrical copper rods: 2 Safety Rods (S.R.), 2 Shim Rods (C.R.) and 1 Regulating Rod (R.R.). Several irradiation channels with different distances from the core and irradiation volumes provide different flux intensities and energy distributions. The diametrical channel (D.M.) passes through the entire core diameter at mid-plane height, allowing samples irradiation in an almost pure fission spectrum. The tangential channel is parallel to the D.M. channel and passes through the reflector at 50 cm above the mid-plane height at a minimum distance from the core axis of 10.6 cm and a minimum diameter of 3 cm. There are also two radial channels, RC1 and RC2 and a source channel that hosts the reactor start-up neutron source. The "Thermal Column" cavity is approximately a parallelepiped volume of 110x100x160 cm^3. Finally, two vertical channels penetrate the biological shield from the top of the reactor down to the outer surface of the copper reflector.



a) b)

Fig. 1: TAPIRO sections. (a) Horizontal and (b) vertical

3 MCNP model

The MCNP model of the TAPIRO research reactor was developed several years ago and was subjected to validation work based on the comparison between the simulation outcomes and the experimental results of a characterization campaign performed in 1983 [3]. The evaluation of neutron flux and spectra in the tangential channel was carried out by inserting neutron flux estimators in 440 points along one half of the tangential channel separated by a distance of 0.5 cm and ranging from the center of the channel to the periphery of the biological shield. The flux results on the other half were considered symmetrical and, for this reason, not investigated in the simulation. The evaluation of the neutron flux was initially carried out with a track length mesh estimator, superimposed on the geometry of one-half of the tangential channel. Unfortunately, the reduction of the number of neutrons with the increase of the distance from the core causes an increase of the relative errors or, in some cases, null values. Therefore, point detector estimators were additionally included in the MCNP input to overcome these problems. The point detector estimator, tally F5 in MCNP syntax, uses the "next event" variance reduction technique and is available only for neutrons and photons. F5 is a deterministic statistical estimator that is expensive from the computational (memory and CPU load) point of view and can provide ambiguous results.

The relative position of the tangential channel and the reactor core is reported in figure 2, where a horizontal portion of the TAPIRO, as reproduced by MCNP plot routine, is represented.

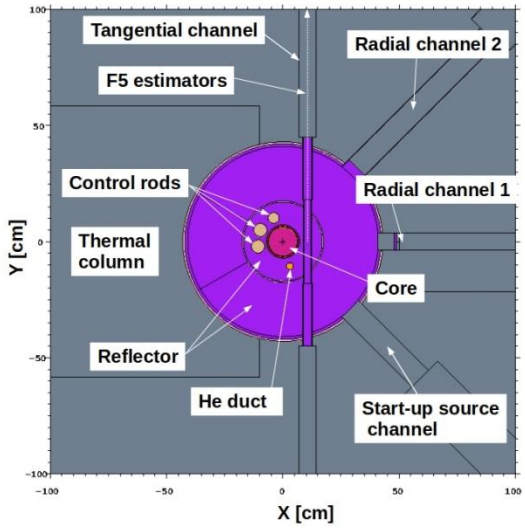


Fig.2: horizontal section of the TAPIRO's MCNP model with the indication of the main reactor parts and the line on which the point detector flux estimators are located inside the tangential channel.

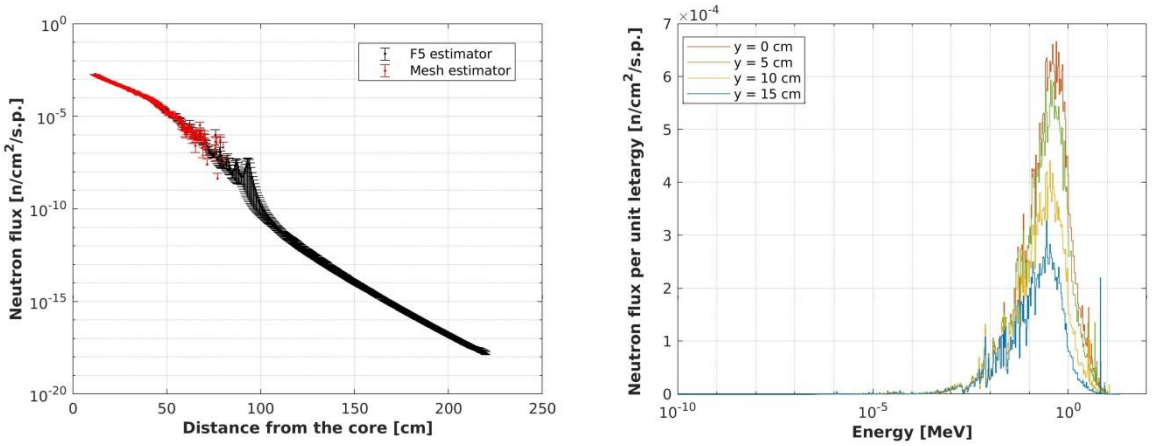
The material compositions were defined at the isotopic level with the nuclides available in the cross-section library JEFF-3.3 (Joint Evaluated Fusion File version 3.3) released by the N.E.A. Data Bank [4]. All the cross-sections were considered at room temperature (293.6 K), neglecting the spatial anisotropies of the temperatures.

4 MCNP 6.2 code on CRESCO

MCNP (Monte Carlo N-Particle) is an export-controlled code developed by Los Alamos National Laboratory and distributed by Radiation Safety Information Computational Center (RSICC) based at the Oak Ridge National Laboratory (Oak Ridge, Tennessee). This code carries out the transport of several nuclear particles in a matter through the Monte Carlo technique in three-dimensional and time-resolved phase space. The pointwise or group-wise cross-sections extract the probability distributions used in the Monte Carlo sampling of MCNP code in A.C.E. (A Compact ENDF) format obtained by the evaluated nuclear data. Various nuclear models implemented in the code can be selected for energies above the maximum energy available in the evaluated cross-sections. The statistical inference with an "acceptable" precision of the quantities estimated with the Monte Carlo technique require, in some cases, the use of parallel programming technique. The parallel executable of MCNP code on the CRESCO cluster was generated, linking the OpenMPI library version 1.6.5 to the Intel Fortran compiler version 14.0.1. A typical simulation performed in this work with 144 tasks (1 thread each) and about $2.5 \cdot 10^6$ neutrons takes approximately 4 hours.

5 Results and discussion

The results of the track length and point detector flux estimators are compared in figure 3a. Figure 3b reports the neutron spectrum evaluated with the F5 estimator at four different distances from the center of the tangential channel (y coordinate in the legend). The track length estimator does not score neutrons beyond 72 cm, where a higher number of source neutrons is clearly needed for this type of estimator. The maximum relative difference between the two estimators is 6.6% in the region of the copper reflector, comparable with the maximum relative error of the mesh tally (5.3%) in the same zone.



a) b)

Fig.**3**: neutron flux and spectrum estimation per source particle (s.p.). (a) Comparison of the track length (Mesh estimator) and point detector (F5 estimator) neutron flux in the tangential channel. (b) Neutron spectrum, in lethargy unit, estimated by point detector at four distances from the channel centre.

The agreement between the results of the two types of estimators (track length and point detector) proves the absence of ambiguities in the F5 tally. Consequently, the neutron spectra, evaluated by the F5 tally, can be considered reliable in assessing characteristics (type and dimension) of the metallic foils. Furthermore, the neutron energy distribution calculated by the F5 estimator can be reliably used as trial input in the unfolding program [5] to determine the energy spectrum from the measured saturated activities of the foils.

References

[1] C.J.Werner Editor, MCNP6 User's Manual, October 2017, Manual Rev. 0. LA-UR-17-2998;
[2] M. Cesaroni. Reattore nucleare di ricerca TAPIRO.
https://www.enea.it/it/Ricerca_sviluppo/documenti/nucleare/TAPIRO.pdf
[3] N. Burgio, L. Cretara, M. Frullini, A. Gandini, V. Peluso, A. Santagata. Monte Carlo simulation analysis of integral data measured in the SCK-CEN/ENEA experimental campaign on the TAPIRO fast reactor. Experimental and calculated data comparison. *Nuclear Engineering and Design* **273**, pp. 350–358, (2014).
[4] A.J.M. Plompen, et al. The joint evaluated fission and fusion nuclear data library, JEFF-3.3. *Eur. Phys. J. A* **56**(2020)181.
[5] INTERNATIONAL ATOMIC ENERGY AGENCY. Current Status of Neutron Spectrum Unfolding (Proceedings of C Technical Committee Meeting, Oak Ridge, 10 - 12 Oct.1977), IAEA-TECDOC-221, IAEA, Vienna (1979).

LONG TIMESCALE MOLECULAR DYNAMICS AND ONE TRILLION VIRTUAL SCREENING ON HPC5

Francesco Frigerio[1*], Silvia Pavoni[1], Alessandro Grottesi[2], Neva Bešker[2], Andrew Emerson[3], Federico Ficarelli[3], Giorgia Frumenzio[3] and Carmine Talarico[4]

[1]*Eni SpA, Physical Chemistry Department, via Maritano 26, IT-20097, San Donato Milanese (Mi), Italy*

[2]*CINECA, HPC Department, via dei Tizii 6, IT-00185 Roma, Italy*

[3]*CINECA, HPC Department, via Magnanelli 6/3, IT-40033 Casalecchio di Reno (Bo), Italy*

[4]*Dompé Farmaceutici SpA, via Campo di Pile, IT-67100, L'Aquila, Italy*

ABSTRACT. The EXSCALATE4CoV project, funded by the European Union and coordinated by Dompé Farmaceutici, set up a platform for an immediate response to COVID-19. Its main goal is the identification of the most promising safe-in-man drugs and de-novo small molecules to be active against the SARS-CoV-2 virus. The computational pipeline implemented on the Eni HPC5 GPU-accelerated supercomputing facility comprises: 1) very long Molecular Dynamics simulations of all SARS-CoV-2 proteins, with selection of a few conformational clusters for each (performed by GROMACS); 2) massively parallel Virtual Screening of a huge molecular library on the best selected protein clusters (by LIGEN). The following experimental pipeline takes on in-silico molecule candidates throughout laboratory tests and ends up with clinical trials at Italian hospitals. The key achievement so far is the identification of the osteoporosis drug Raloxifene as possibly active against COVID-19. It is already registered and generic, with known safety characteristics (e.g. dosage and side effects). Application has been made to the European Medicines Agency and clinical trials have started

1 Introduction

At the beginning of 2020 the European Union (EU) funded the EXSCALATE4CoV (E4C) project within Horizon 2020. This action set up a platform for an immediate response to COVID-19 in the search of new molecules as potential drugs against the new Coronavirus SARS-CoV-2. E4C (exscalate4cov.eu) is coordinated by Dompé Farmaceutici and brings together 18 partners among institutions and top research centers from 7 European countries, including CINECA from Italy.

Fig. **1**: the Eni HPC5.

*Corresponding author. E-mail: Francesco.Frigerio@eni.com

In March 2020 Eni offered its HPC5 facility (Figure 1) to support the big computational effort required in the project. With 58 PetaFLOPS and hybrid CPU/GPU architecture, HPC5 was the most powerful supercomputer for industrial use in the world. Since then the computing power made available by Eni joined the 32 PetaFLOPS of M100 - contributed by CINECA - in the infrastructure deployed for the E4C platform. Molecular Dynamics (MD) and Virtual Screening (VS) experiments were performed on both HPC resources following the project timeline.

2 Eni Collaboration with E4C

2.1 Main goal of the project

E4C aims at the identification of the most promising safe-in-man (SIM) drugs and de-novo small molecules to be active against the SARS-CoV-2 virus. This is to be accomplished along a clearly defined mixed pipeline (Figure 2) with computational actions, biochemical studies, in vitro and in vivo experiments and finally hospital trials.

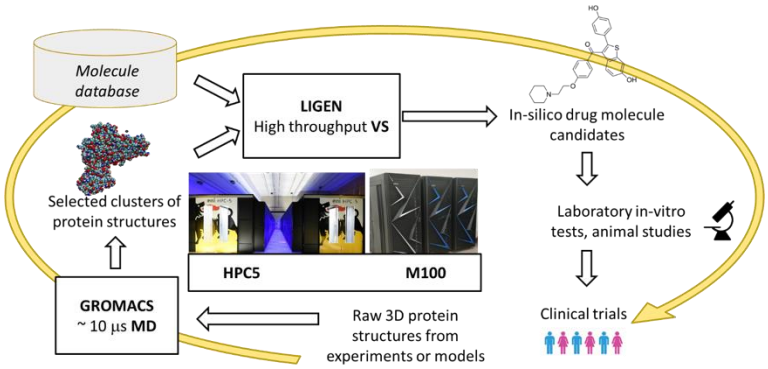


Fig. 2: the E4C computational/experimental pipeline.

Within the E4C platform, two main computational items were implemented by research scientists of CINECA and Eni on M100 and HPC5 (Table 1):

- Very long MD simulations of the experimental crystal structures and the homology models of all SARS-CoV-2 proteins. They were followed by the selection of a few representative clusters of conformations for each protein.
- Massively parallel VS of very large SIM molecular libraries through docking into carefully identified sites of the best selected protein clusters.

Table 1: Hardware resources.

System name	Eni HPC5	CINECA M100
Nodes	3400	980
Processors (node)	2x24 Intel Cascade Lake @ 2.1 GHz	2x16 IBM P9 @ 3.1 GHz
Accelerators (node)	4 x Nvidia Tesla V100	4 x Nvidia Tesla V100 (NVLink CPU-GPU)
Memory (node)	56 GB	200 GB
Peak Performance	51.7 PFlops	32PFlops
TOP500 (06/2020)	6th	9th

Details of both computational items are described in the following.

2.2 MD simulations

For every SARS-CoV-2 protein the initial structure was taken from published X-ray crystallography determinations or from available homology models (an example in Figure 3).

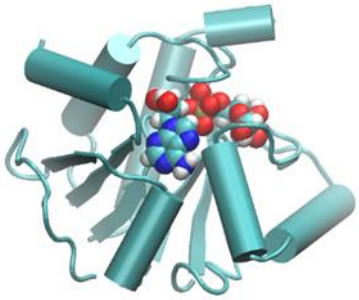


Fig. **3**: The structure of the SARS-CoV2 protein NSP3 in complex with its ligand [1].

The initial dataset contained 30 unique structures and, due to known variants and models, a total of 46 simulations were started according to a supervised MD protocol. High, medium or low priorities were assigned according to their role in viral infection and replication. The AMBER ff99 forcefield [2] was adopted and the GROMACS program suite [3] was used for simulation and analysis. The GPU acceleration characterizing HPC5 and M100 was fully applied to GROMACS MD simulations.
The system preparation involved the assignment of forcefield parameters, the protein soaking into water, the possible addition of inorganic ions for electrical neutrality, the energy minimization and structure relaxation to remove any close contacts. This was followed by long simulations in the NVT ensemble at ambient pressure and temperature with a timestep of 2 femtoseconds. For most proteins 10 microseconds long trajectories were produced, while in a few specific cases 20 microseconds were reached. All results are being periodically deposited in a publicly available cloud for the whole scientific community (see further).
The structure equilibration was checked with Root Mean Square Deviation analysis. Each trajectory was subjected to Principal Component Analysis and Cluster Analysis. These treatments allowed to identify major clusters and to extract only a few highly representative protein conformations for the subsequent VS treatment with LiGen [4].

2.3 VS experiments

The VS platform (Figure 4) is based on the adaptation of the LiGen program to the GPU accelerated HPC5 and M100 facilities. Through extremely efficient docking evaluation of protein/ligand complementarity a high-performance in silico screening of huge databases of chemical structures was made possible.

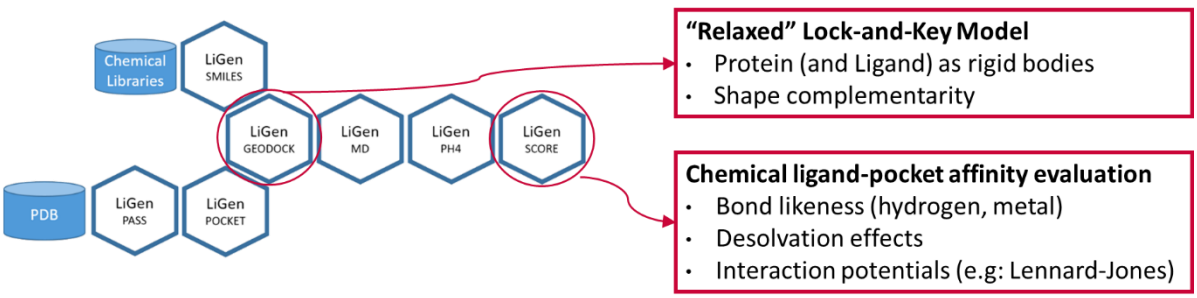


Fig. **4**: the E4C VS software.

The LiGen code base is owned by Dompè and was co-developed by CINECA and Politecnico di Milano along a production process lasting for about 20 years. Its main goal is the exploration of a tangible

chemical space, collecting known chemical structures, whose synthesis is considered achievable in one reaction step from commercial reagents.
The main effort was the need to port all relevant algorithms (Figure 5) to GPU-accelerated HPC platforms using the CUDA parallel computing platform.

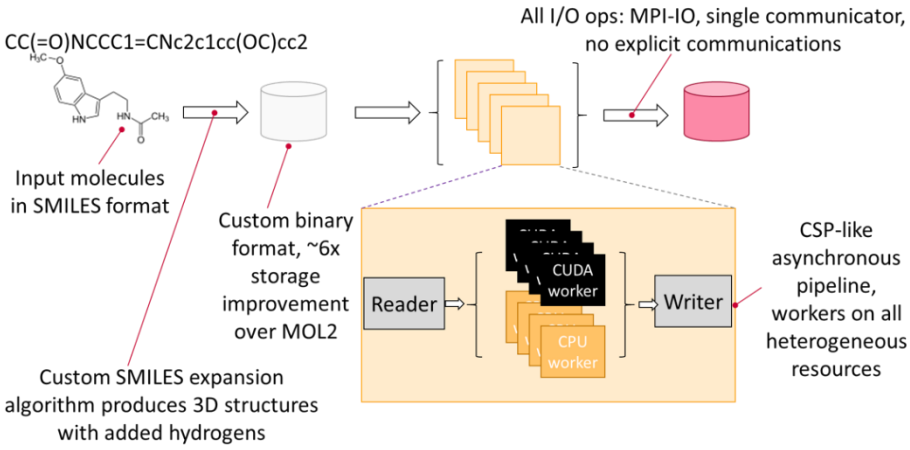


Fig. **5**: the E4C VS application architecture.

The input molecules are defined in the SMILES format. A custom SMILES expansion algorithm produces 3D structures with added hydrogens in binary format, with ~6 times storage improvement over reference method MOL2. All I/O works as MPI-IO, single communicator, with no explicit communications.
The resulting VS scale up (Figure 6) typically obtained a sustained single-node throughput of ~1600 ligands/s (~16000 poses/s, 10 best poses kept for scoring) on the full M100 system.

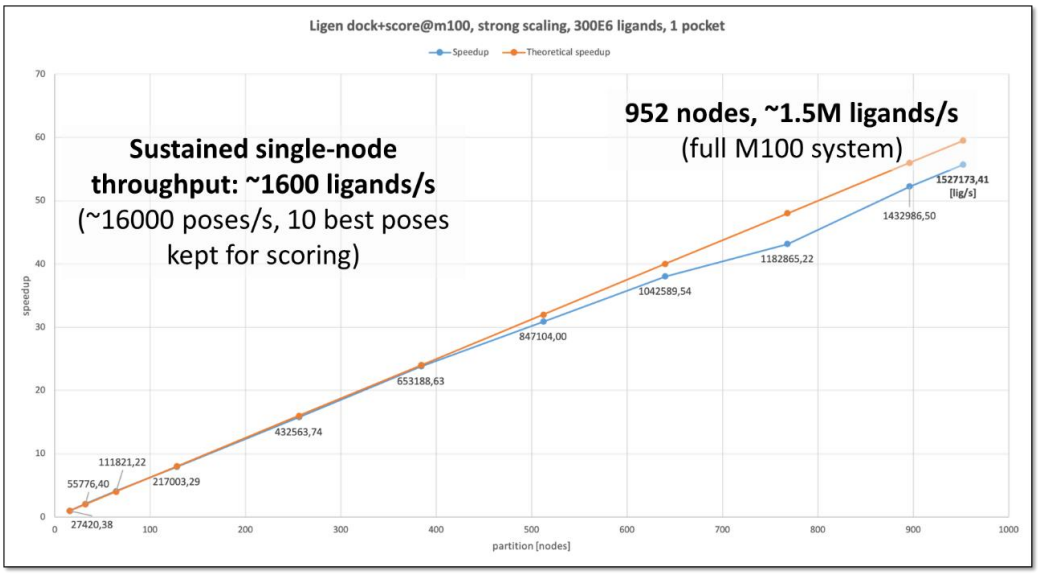


Fig. **6**: VS scale up on M100.

The first VS run of the E4C studied a database containing 400000 molecules. This in silico test selected ~7000 candidates while, throughout following in vitro tests, a total of ~100 structures were deemed interesting. At the end of this VS run ~40 candidates were experimentally found to be effective in limiting the SARS-CoV-2 virus replication.

A key achievement was then obtained during the 2020 spring: the identification of the osteoporosis drug Raloxifene (Figure 7) as possibly active against COVID-19.

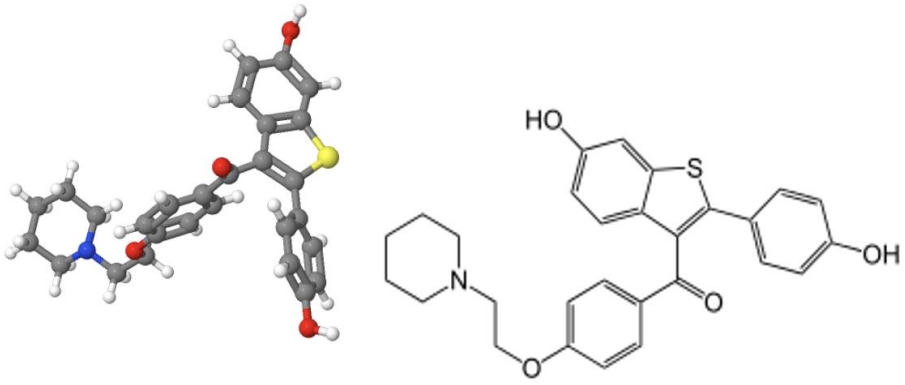


Fig. **7**: Two depictions of the chemical structure of Raloxifene.

This is a registered and generic drug with known safety characteristics (e.g. dosage and side effects). As a consequence, an application was made to the European Medicines Agency (EMA) and clinical trials have started.
The second run of the described simulation pipeline resulted in the largest VS experiment ever run. In November 2020 a library of 71 billion ligands was docked and scored in 15 active sites of 12 SARS-CoV-2 proteins: PLPRO, SPIKEACE, NS12thumb, NS12Palm, NSP12ortho, 3CL, NSP13allo, NSP13ortho, NSP3, NSP6, NSP9, NSP14, NSP15, NSP16, Nprot. Its workflow (Figure 8) is essentially the same as in the first run and it is composed of three steps: the input data preparation, the docking and scoring, the data post-processing.

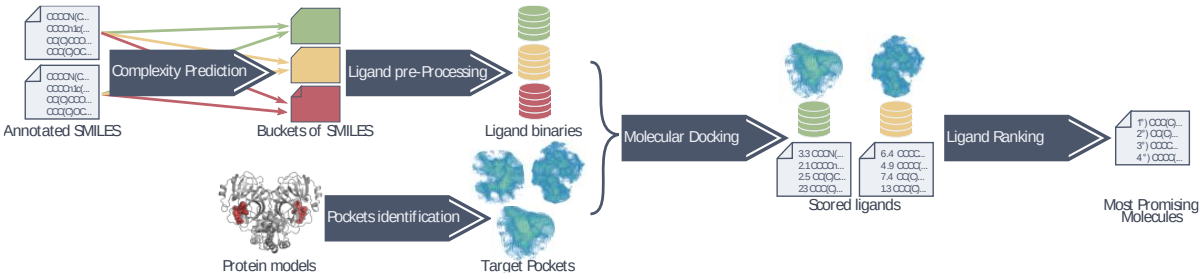


Fig. **8**: The VS workflow.

This meant that one trillion interactions were evaluated in 60 hours on the CINECA M100 (800 nodes) and the Eni HPC5 (1500 nodes). In details, that was accomplished by 2 K ligands/second/HPC-node and 5 M ligands/second overall. The post processing of the 65 TB of data generated by the big experiment is still ongoing.
A similar experiment run in June 2020 on the SUMMIT supercomputer at Oak Ridge National Lab evaluated less than 3 Billion interactions (1.4 B ligands in 2 sites). Therefore, the E4C experiment was 300 times larger and 500 times faster.

3 Conclusions

All scientific results are being published (https://www.exscalate4cov.eu/contribute.html#papers). Moreover, all produced data are being made public on mediate.exscalate4cov.eu (Figure 9), that is the Web portal of the MEDIATE initiative and was designed in collaboration with SAS. Arieh Warschel (Nobel prize winner for Chemistry in 2013), Rossen Apostolov, Igor Tetko and Yang Ye sit in its Scientific Committee. That is the starting point for any scientific and practical collaboration and it

declares the following purpose: “to collect the best possible chemical library of novel inhibitors of SARS-COV-2 targeting the most relevant viral proteins combining as much as possible VS simulations and Artificial Intelligence (AI) predictions generated by the best drug hunters worldwide”.

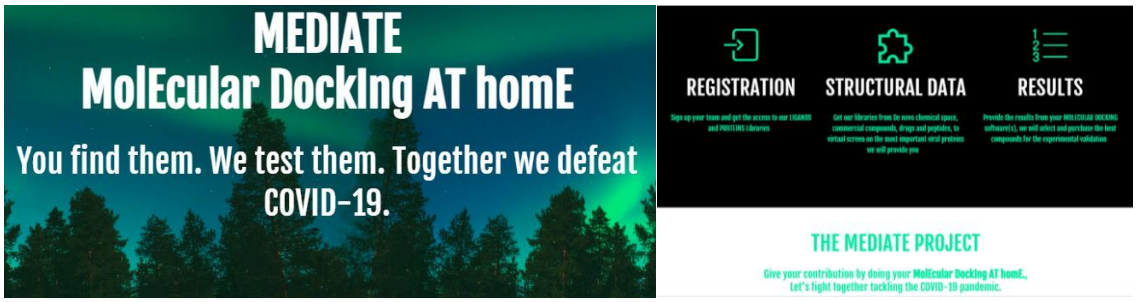
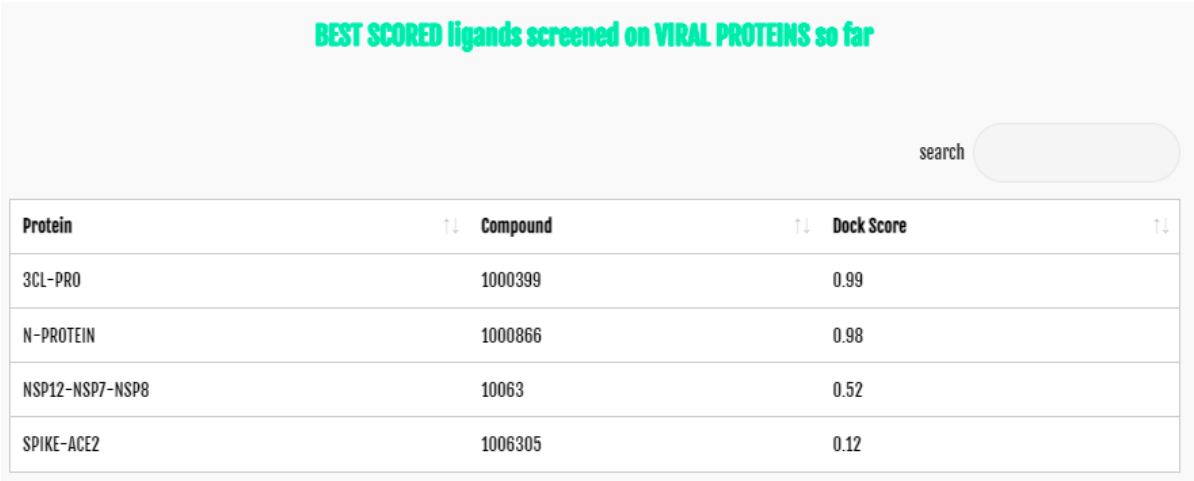


Fig. **9**: the MEDIATE initiative for collaboration on the E4C results.

After registration, it is possible to participate by taking at least one of five actions: submission and data collection of crowdsourced simulations, compound selection (generation of a single final ranking by using machine learning and AI technologies, thanks to the SAS platform), compounds acquisition (purchase of the best candidates selected in phase 2 of the VS), testing (the libraries will be tested in all E4C COVID-19 assays and the most promising could be crystallized), data sharing (results on the compounds will be published in public portals and made available).
The first VS phase of the E4C project selected Raloxifene as the best candidate for the hospital trials, that are currently ongoing. A new list of candidate drugs is going to come out after concluding the analysis of the second VS phase results. Some example can be found in the MEDIATE homepage (Figure 10).
Our acknowledgements go to the Project “EXaSCale smArt pLatform Against paThogEns for Corona Virus - Exscalate4CoV” funded by the EU’s H2020-SC1-PHE-CORONAVIRUS-2020 call under the grant N. 101003551. The two E4C computational teams are composed by:

- Carmine Talarico (Dompè), Alessandro Grottesi (CINECA), Andrew Emerson (CINECA), Neva Besker (CINECA), Giorgia Frumenzio (ABD) and Francesco Frigerio (Eni S.p.A.) for MD;
- Federico Ficarelli (CINECA), Chiara Latini (CINECA), Davide Gadioli (Politecnico di Milano), Emanuele Vitali (Politecnico di Milano), Gianluca Palermo (Politecnico di Milano), Carlo Cavazzoni (Leonardo S.p.A.) for VS.

Dedicated and efficient contributions to hardware operation and software installation and maintenance by Eni and CINECA support teams are kindly acknowledged.



Protein	Compound	Dock Score
3CL-PRO	1000399	0.99
N-PROTEIN	1000866	0.98
NSP12-NSP7-NSP8	10063	0.52
SPIKE-ACE2	1006305	0.12

Fig. **10**: MD and VS simulation results published within MEDIATE.

References

[1] K. Michalska, Y. Kim, R. Jedrzejczak, N. I. Maltseva, L. Stols, M. Endres and A. Joachimiak. Crystal structures of SARS-CoV-2 ADP-ribose phosphatase: from the apo form to ligand complexes. *International Union of Crystallography Journal* **7**, pp. 814-824, (2020).

[2] J. Wang, P. Cieplak and P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **21**, pp. 1049-1074, (2000).

[3] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2** pp. 19–25 (2015).

[4] A.R. Beccari, C. Cavazzoni, C. Beato and G. Costantino. LiGen: A High Performance Workflow for Chemistry Driven de Novo Design. *J. Chem. Inf. Model.* **53**, pp. 1518–1527 (2013)

THE ROLE OF THE NEUTRON TRANSPORT SIMULATION IN THE CALIBRATION OF THE MOSCAB DETECTOR

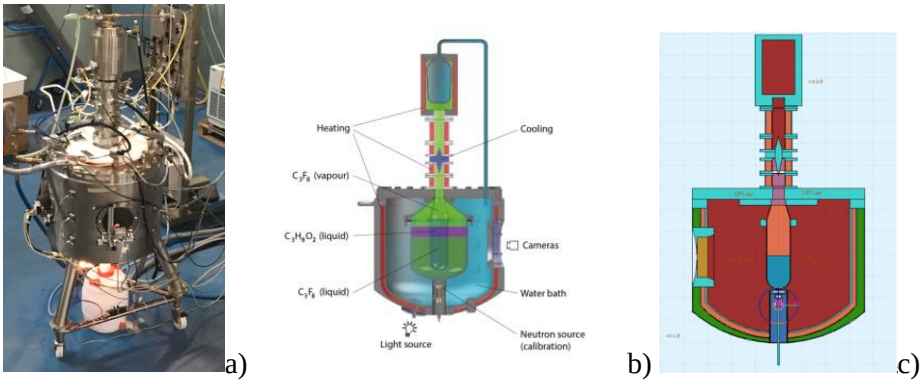
N. Burgio* and A. Santagata

ENEA, FSN-FIS-RNR, C.R. Casaccia, Roma, Italy

ABSTRACT. The Materia OScura Camera A Bolle (MOSCAB) equipment is a geyser-concept bubble-chamber thought to search dark matter in the form of Weakly Interactive Massive Particles (WIMPs). The detector calibration through a semi-empirical procedure is based on minimising the differences of simulated detector bubble rates obtained via MCNP Monte Carlo simulations and the homologues measurements with reference neutrons source. This work focuses on the Monte Carlo simulation of the nuclear stages of the MOSCAB detection chain modelled with the MCNP 6.2 code on the CRESCO computational facility.

1 Introduction

The MOSCAB geyser detector should reveal the WIMPs through a possible spin-dependent interaction with fluorine nuclei [1] bonded in a C_3F_8 molecule within liquid volume maintained in a supercooled (metastable) state. After such an interaction, the fluorine nucleus recoils away in the form of an ion, depositing the acquired kinetic energy into the bulk of the detecting liquid, promoting local vaporisation with subsequent growth of an optically detectable vapour bubble into the active volume. The liquid metastability, in which the vapour nucleation is kinetically favourite, is steadily maintained in MOSCAB for an extended period. The geyser technique, introduced initially to detect fission fragments [2], can be used for every nuclear particle's interaction, generating recoiling ions with sufficient kinetic energy. The neutron elastic scattering process is one of these interactions. Consequently, we used a low-intensity neutron source of Am-Be to calibrate the MOSCAB detector.



a) b) c)

Fig.1: The MOSCAB detector. (a) The real device. (b) Schematic View. (c) MCNP model.

As reported in Figure 1 consists of a closed vessel filled with a target liquid (e.g., C_3F_8) and its saturated vapour. The vessel's bottom contains the liquid kept at a constant temperature, T_L, by a thermal bath. The saturated vapour above it is kept at a constant temperature T_V, such that $T_V < T_L$. The superheat parameter can measure the degree of metastability SH [3] according to the following relation:

$$SH = \frac{T_L - T_V}{T_C - T_V} \quad (1)$$

T_c is the critical temperature (T_c = 71.87 °C for C_3F_8).

*Corresponding author. E-mail: nunzio.burgio@enea.it

A detailed description of the detector can be found in [4]. A schematic view is reported in Figure 1b, where the two configurations, 2 and 20 litres jars, are presented. The detector essentially has two thresholds: a thermodynamic critical energy threshold E_c [5], which represents the minimum energy required to create a stable vapour bubble, and the attitude of the recoiling ion to deposit E_c into a (critical) volume of the liquid that is quantified by critical stopping force threshold $(dE/dx)_c$. Therefore, the energy threshold, E_{th}, i.e., the minimum energy for an ion to generate bubble nucleation, is $E_{th} \geq E_c$ and depends on the stopping force function of the generated ion. Many authors have calculated E_c following different strategies. In the energy region of interest, we assume as valid the following relations

$$E_c = \frac{4}{3}\,\pi\,\rho_v \Delta H_{ev} R_C^3 + 4\pi(\gamma - T_L \frac{d\gamma}{dT}) R_c^2 \qquad (2)$$

$$R_c = \frac{2\gamma}{\rho_v \Delta H_{ev}(\frac{T_L - T_V}{T_V})} \qquad (3)$$

Where ρ_v is the vapour density, ΔH_{ev} is the evaporation enthalpy, γ is the surface energy of the vapour-liquid border, and R_c is the radius of the critical volume assumed as spherical. When the deposited energy density is higher than E_{th}, the nucleation of a vapour phase occurs. Then, the vapour germ could grow to a macroscopic, optically detectable bubble or be reabsorbed into the liquid phase. Tuning the liquid's metastability level, we can exclude the neutrons having kinetic energy below Eth from the detection. To accomplish this task, we perform a semi-empirical calibration procedure based on bubble rates count experiments and modelling the nuclear and thermodynamics detection stages. Once calibrated, the detector, which is entirely insensitive to minimum ionising particles and gammas, can be wholly muted of the residual source of background (such neutrons from natural background and alpha particles) selecting specific conditions. This work focuses on the neutron transport simulations performed with the MCNP 6.2 [6] on the CRESCO computational facility to model the nuclear detection stages of the MOSCAB detector. The simulations have been performed on the CRESCO Frascati cluster, compiling MCNP with the local INTEL compiler suite with the corresponding Intel MPI library. Each simulation took, on average, 7-14 hours using 192 cores and requesting an estimator total relative error less than 2%.

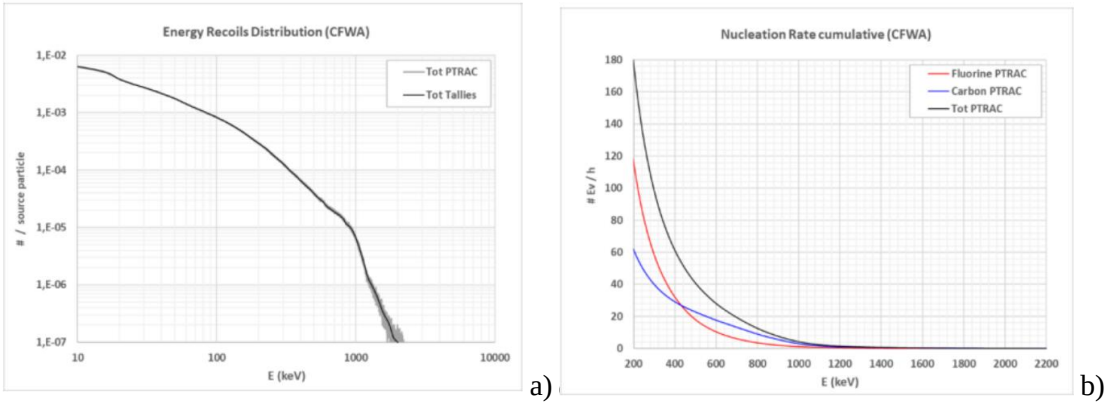
2 The detector Calibration

To explain the calibration procedure, we can split the MOSCAB detection mechanism into four stages:

Stage 1 - Neutron Transport: A given fraction of the neutrons emitted by the calibration source travel into the detector structure up to the active freon volume. In this stage, neutrons interact with the structural materials, insulating materials, thermostatics fluids modifying their kinetic energy.
Stage 2 - Neutron interaction in Freon: The neutron fraction that penetrates up to the freon volume interacts (mainly via scattering) with covalently bonded Carbon or Fluorine nuclei exchanging kinetic energy with them. Being the amount of transferred energy at least of the order of keV, the knocked nucleus easily breaks its molecular bond and start to travel into the bulk of the metastable liquids in the form of an ion. The recoiling nuclei have initial kinetic energy distributed according to the kinematic scattering lawn that governs the neutron-nuclei collision.
Stage 3 - Recoiling ion transport: Each recoiling ion, due to the interaction with the nuclei and electron clouds of the freon molecules, slow down releasing energy into the metastable liquid. When the ions' amount of locally deposited energy fulfils the previously mentioned critical, a vapour nucleation germ formation occurs.
Stage 4 – Growth of the nucleation germ to visible bubble: According to the thermophysical configuration of the nucleation germ and its interaction with the surrounding liquids, it can initiate a progressive growth to form a visible vapour bubble or collapse again into the liquid.

Stages 1 and 2 concern the neutron transport and interaction in freon using the MCNP 6.2 [9]. The complete 3D geometry of the detector, the composition of each material and the description of the experimental neutron source (Am-Be) are introduced in the MCNP input deck to estimates the neutron scattering in freon (Figure 1c). To estimate the recoiling ion's energy distribution, we use the MCNP's auxiliary file PTRAC (Particle TRACking) to store the relevant phase-space information (position,

energy, direction) on the neutron events taking place into the freon during the simulation. The PTRAC-analyser (a homemade program) reads the PTRAC file to calculate, via the scattering kinematic, the energy of the recoiling Fluorine and Carbon ions integrating the information in a frequency. We check this last result with the SPECTRA-PKA code [7] using as input the MCNP's estimate of the freon jar neutron flux obtained during the same simulation: figure 2a shows that the results from the two codes are in excellent agreement.



a) b)

Fig.2: a) Comparing of total ion's energy recoils distributions obtained from the PTRAC analysis and SPECTRA-PKA b) Total (black), Carbon (bleu), and Fluorine (red) cumulative expected number of nucleation germs generated by the recoiling ions vs E_c.

Stage 3 estimates the expected number of nucleation germs induced in the metastable liquid at an Ec value by the neutron irradiation. We obtain such estimate integrating the ion's recoil energy distribution for the considered Ec range and yielding in the cumulative nucleation rates distribution. Figure 3b reports the total and the partial (F and C) ions cumulative nucleation rates distributions in the E_c range from 200 to 2200 KeV. The cumulative ion recoils distribution is an upper boundary estimation of the experimental bubble rate. It does not consider that nucleation germs can be reabsorbed in the liquid bulk because some kinetic factors hinder their growth. Since the modelling of such factors goes beyond the scope of the present work, we use in **stage 4** an energetic approach that correlates the ion critical stopping force SF_c with E_c (equation 1) and R_c (equation 2) according to

$$SF_c = \frac{E_c}{2K_D R_c}; \quad \text{equation (4)}$$

Using equations 2 and 3 and applying the nucleation condition, we obtain for the K_D parameter the following expression

$$K_d = a \sqrt[6]{E_c} + b; \quad \text{equation (5)}$$

As shown in figure 3a, where the 1σ region is shaded grey, a satisfactory best fit of the data is obtained when the values of the two empirical constants, a and b, are assumed to be 0.42 and 1.70. Uncertainties of the source strength impact the values of K_D for less than 5%, with its distribution against E_c remaining substantially unchanged.

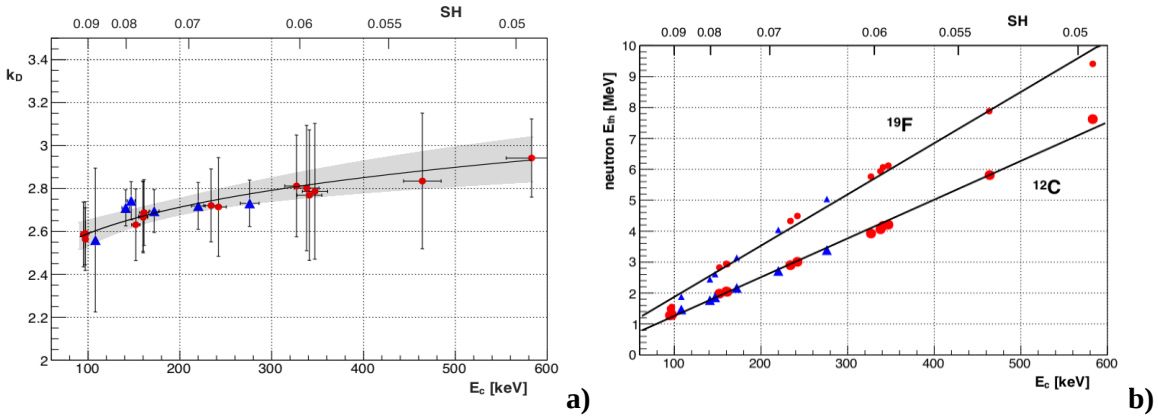


Fig.3: a) The nucleation parameter K_D vs the critical energy Ec and SH's superheat parameter. The red dots represent the 18 L vessel measurements, and the blue triangles are for the 18L vessel measurements. The grey shaded 1σ region represents the best fit curve; b) The neutron energy threshold as a function of Ec and SH for ^{19}F and ^{12}C.

Figure 3b reports the entire investigated neutron energy range in terms of the critical energy E_c and the superheat parameter SH whose relationship is well approximated by the following equations:

$$E_{th}^{n}(E_c) = \alpha\, E_c + \beta = (12.5 \pm 0.2)E_c + (7.5 \pm 70)\ ; \quad \text{equation (6)}$$

$$E_{th}^{n}(SH) = \frac{\delta}{SH^{\nu}} + \beta = \frac{(1.036 \pm 0.016)}{SH^{2.95}} + (7.5 \pm 70); \quad \text{equation (7)}$$

3 Conclusion

The calibration procedure shows that setting at a value of the reduced superheat parameter SH lower than 0.07, or at critical energy Ec higher than nearly 200 keV (2.5 MeV neutron energy threshold), the detector becomes sensitive only to neutrons. . In these conditions, the residual internal background rate of the detector is lower than five events $(y \cdot kg)^{-1}$, which corresponds to less than two events every ten days in the 18 L detector. The steadiness of operation of the detector for extended times of observation should be highly enhanced. These findings have been experimentally confirmed, as demonstrated by running the MOSCAB bubble chamber for more than 4000 hours in the LNGS underground laboratories.

References

[1] R. Bertoni et al., Nucl. Inst. Meth. A744 (2014) 61-68.
[2] B. Hahn, H.W. Reist, Proceedings of the Fifth International Conference on High Energy Physics and Nuclear Structure, Uppsala, 18-22 June 1973, edited by G. Tibell (Uppsala University, Sweden, 1973), p. 191.
[3] F. d'Errico, Rad. Prot. Dos. 84 (1999) 55Ð62.
[4] A. Antonicci et al.," MOSCAB: a geyser-concept bubble chamber to be used in a dark matter search", Eur. Phys. J. C (2017) 77:752
[5] F. Seitz, "On the Theory of the Bubble chamber", (Phys. Fluids) 1, (1958) 2.
[6] C.J.Werner Editor, MCNP6 User's Manual, October 2017, Manual Rev. 0. LA-UR-17-29981.
[7] M.R. Gilbert, J.Ch. Sublet," Differential DPA calculation with SPECTRA-PKA", Journal of Nuclear Materials 504 (2018) 101-108.

BLIND PREDICTIONS FOR THE SAMPL9 HOST-GUEST CHALLENGE

Piero Procacci*

University of Florence, Chemistry Department, Via Lastruccia n. 3, 50019, Sesto Fiorentino, Italy

ABSTRACT. We present our blind predictions for the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL), 9th challenge, focusing on binding of WP6 (carboxy-pillar[6]arene) with ammonium/diammonium cationic guests. Host-guest binding free energies have been calculated using the recently developed virtual double system single box approach, based on the enhanced sampling of the bound and unbound end-states followed by fast switching nonequilibrium alchemical simulations [M Macchiagodena, M Pagliai, M Karrenbrock, G Guarnieri, F Iannone, Procacci, J Chem Theory Comput, 16, 7260, 2020]. Experimental binding free energies for the thirteen host-guest complexes will be disclosed by November 2021.

1 Introduction

The Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) is a set of community-wide blind challenges aimed at advancing the computational techniques in drug design [1-3]. New experimental data, such as dissociation free energies, hydration free energies, acid-base dissociation constants or partition coefficients are withheld from participants until the prediction submission deadline, so that the true predictive power of methods can be assessed.

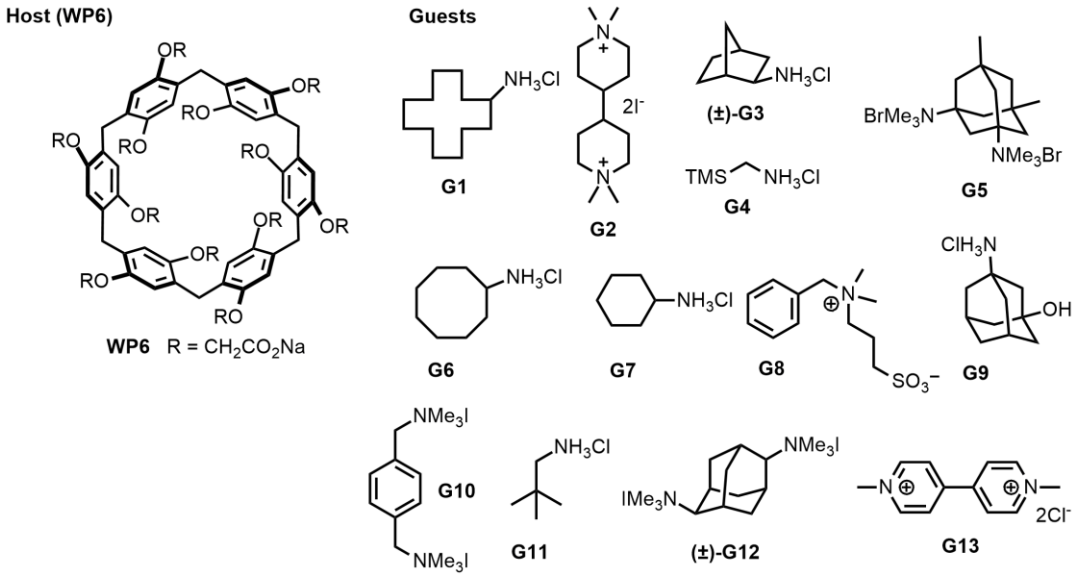


Fig. 1. Host-guest SAMPL9 challenge

In the latest 9th challenge, participants are required to predict the binding affinities of thirteen ammonium/diammonium cationic ligands (guests) *vs* WP6, a water-soluble toroidal macrocyclic host molecule (see Figure 1). Host-guest experimental binding affinities, measured by Isothermal Titration Calorimetry (ICT) at pH=7.4 will be disclosed by early November 2021.

*Corresponding author. E-mail: procacci@unifi.it

The host WP6, yet to be featured in previous SAMPL challenges, is structurally similar to cucurbit[n]urils (CBn) used in SAMPL6-SAMPL8[2,3] with important differences. Unlike the neutral CBn host, the quasi-D_{6h} WP6 bears six carboxylated moieties on the upper and lower rims that can be in part protonated at pH=7.4[4,5]. As the pKa's of WP6 are not available experimentally, such a feature represents an important challenge in the blind prediction of WP6-guest binding affinities. Participants will in fact need to deal with this complication.
In this report, we present the dissociation free energies of the fully anionic $WP6^{-12}$ for the thirteen cationic guests of Figure 1 plus five additional guest molecules with known dissociation constants, namely methylene blue (G14)[6], GNF-Pf-3194 (G15)[7], M2 (G16)[7], choline (G17)[8] and betaine (G18)[8]. The calculations have been performed on the CRESCO6-ENEA cluster[9] in Portici (Italy) using the so-called virtual Double System Single Box (vDSSB) method[10], based on a production of a swarm of concurrent nonequilibrium (NE) alchemical simulations that are started from end-state canonical configurations sampled using Hamiltonian Replica Exchange (HREM).

2 Methods

As stated previously, the twelve pKa's of WP6 host are not known. Strictly speaking, in presence of multiple host-guest complexes with various protonation states, the overall observed association constant for Gn-WP6 complexes is given by

$$K_a = \sum \; W_k K_a^{(k)} \qquad (1)$$

where $K_a^{(k)}$ is the association constant for the k-th protonated WP6 species and W_k is the corresponding normalized weight. The Gn-WP6 dissociation free energy ΔG_d should be hence calculated as

$$\Delta G_d = RTln\left(\sum \; W_k e^{\beta \Delta G_d^{(k)}}\right) \qquad (2)$$

where ΔG is the dissociation free energy with the host in the k-th protonation state. The predicted pKa of the WP6 template monobasic acid 2-(2,5-dimethylphenoxy) acetic acid is 3.23 [11]. Assuming a pKa distribution of the twelve equivalent protonation states modulated by the so-called statistical factor [12], we obtain that the prevalent species at pH=7.4 is the 12-anion with all deprotonated carboxylated groups ($W = 0.98$). We hence computed the dissociation constant for the $WP6^{12-}$ species only.

The vDSSB methodology is thoroughly described in Ref. [10]. In brief, the method consists of two massively parallel computational steps, the HREM stage and the nonequilibrium alchemical stage. The *HREM stage* relies on the enhanced sampling of the host-guest bound state in explicit water and of the isolated guest molecule. The initial configurations for the unbound state are obtained by combining the HREM-sampled (decoupled) ligand gas-phase snapshots with a pre-equilibrated box filled with explicit water. H-REM uses n=16 and n=8 replicas for the bound and unbound states respectively, with a maximum scaling factor of S=0.1 (corresponding to a temperature of 3000 K) involving only the intra-solute interactions. The scaling factors along the replica progression are computed according to the protocol $s_m = S^{\frac{(m-1)}{n}}$ with *m=1..n.* Only scaling factors are exchanged among neighboring replicas to minimize the communication overhead on the MPI layer. HREM simulations are run for 48 ns and 16 ns for the bound and gas-phase state, respectively.
In the *NE alchemical stage,* the bound state leg of the alchemical cycle is performed by rapidly decoupling the bound ligand in a swarm of 720 independent NE simulations, each lasting for 0.72 ns. The unbound leg of the cycle is done by growing (recoupling) the ghost ligand in the solvent in 480 NE alchemical simulations lasting 0.36 ns. For further details on the ligand coupling/decoupling protocols, we refer to Ref. [10]. The number of concurrent alchemical simulations in a parallel job corresponds to the number of requested MPI processes.
The final bound and unbound resulting alchemical work distribution are combined in the convolution

$$P(W) = P_b * P_u(W) = \int \; dw P_b(W) P_u(W - w) \qquad (3)$$

The dissociation free estimate can be performed by way of the Jarzynski identity[13]

$$\Delta G_{vDSSB} = -RTln\left(\int \quad P(W)e^{-\beta W}dW\right) \qquad (4)$$

or using the Crooks theorem in the assumption that the convolution $P(W)$can be described by a mixture of three normal distribution [10,14], i.e

$$\Delta G_{vDSSB} = -RTln\left(\sum \quad c_i e^{-\beta\left(\mu_i - \frac{\beta\sigma_i^2}{2}\right)}\right) \qquad (5)$$

where $c_i, \mu_i, \sigma_i^2$are the normalized weight, mean and variance of the i-th normal component, determined via the expectation-maximization algorithm [15]. Estimates based on eq. (4) or eq. (5) are used depending on the width and character of the work distribution as assessed by the Anderson-Darling normality test[16].
The free energy estimates are corrected for a volume and a charge term (for non-neutral ligand) given by

$$\Delta G_{Vol} = RTln\left(4\pi\frac{(2\sigma)^3}{(3V_0)}\right) \qquad (6)$$

$$\Delta G_q = \frac{\pi}{\alpha^2}\left[\frac{q_G^2+2q_Gq_H}{V_b} - \frac{q_G^2}{V_u}\right] \qquad (7)$$

where $\sigma$is the standard deviation of the host-guest COM-COM distance distribution in the bound state, $V_0$is the standard state volume, $q_G$and $q_H$are the net charges of the guest and the host, and $V_b, V_u$are the mean volume of the MD box for the bound and unbound state respectively, and α is the Ewald convergence parameter. The final blind prediction for the standard dissociation free energy is given by

$$\Delta G_d = \Delta G_{vDSSB} + \Delta G_{vol} + \Delta G_q \qquad (8)$$

The Force Field (FF) parameters and topology of the host and guests molecules were prepared using the PrimaDORAC interface [17] based on the GAFF2[18] parameter set. The initial bound state was prepared using the Autodock Vina code[19]. The bound complexes and the ghost ligands were solvated in about 1600 OPC3[20] and 512 water molecules, respectively. A background neutralizing plasma was assumed within the PBC Ewald method (PME[21]). All simulations, HREM or nonequilibrium, were performed in the NPT ensemble in standard conditions using an isotropic Parrinello-Rahman Langrangian[22] and a series of Nose' thermostats[23] for pressure and temperature control, respectively. All simulations have been performed using the hybrid OpenMP-MPI program ORAC[24] on the CRESCO6 cluster[9].

3 Results and Discussion

In Table 1 we show the results obtained for the dissociation free energy for the eighteen host-guest pairs. The correlation coefficient and the Kendall rank coefficient for the guests with known dissociation free energy are R=0.83 and τ=0.71, respectively with a mean *signed* error of -3.3 kcal/mol, indicating a systematic overestimate of the dissociation free energies.

Table 1: Blind predictions for the host-guest SAMPL9 set dissociation free energies. Free energies are given in kcal/mol.$\sigma_b$and $\sigma_u$are the standard deviation in kcal/mol of bound and unbound work distribution. "Type" refers to the estimate type (JR, eq (6); GM*n* eq (7) with *n* normal components). Experimental dissociation free energies, ΔG_{exp}, have been taken from the Refs. in the square bracket.

Host-guest	ΔG_d	$\Delta G_{vol} + \Delta G_q$	σ_b	σ_u	Type	$\Delta G_{exp.}$
G1-WP6	7.7∓1.9	-4.6	3.9	1.9	JR	-
G2-WP6	17.0∓0.4	-7.2	2.8	1.1	JR	-
G3-WP6	11.5∓0.6	-6.5	3.1	0.8	JR	-
G4-WP6	7.0∓0.9	-6.2	2.6	0.8	GM1	-
G5-WP6	9.5∓0.5	-8.8	3.0	1.3	JR	-
G6-WP6	14.7∓1.3	-6.3	3.1	0.9	GM1	-
G7-WP6	12.2∓1.7	-7.0	3.2	0.8	GM3	-
G8-WP6	12.2∓1.6	-3.3	3.9	1.6	GM3	-
G9-WP6	12.5∓0.6	-5.6	3.3	1.0	JR	-
G10-WP6	13.5∓1.1	-6.9	2.8	1.1	GM1	-
G11-WP6	8.7∓0.6	-6.2	2.6	0.8	GM3	-
G12-WP6	18.6∓1.2	-7.7	2.9	1.1	GM3	-
G13-WP6	8.3∓0.8	-7.3	3.0	0.9	GM3	8.5 [5]
G14-WP6	17.2∓0.4	-4.8	2.5	1.1	GM3	9.7 [6]
G15-WP6	11.2∓1.6	-4.8	2.8	0.9	GM3	8.4 [7]
G16-WP6	14.5∓1.5	-4.7	2.6	1.1	GM3	10.6 [7]
G17-WP6	7.9∓0.8	-6.3	2.1	0.7	GM3	6.5 [8]
G18-WP6	4.1∓1.3	-3.7	2.2	0.9	GM3	< 1 [8]

Such discrepancies might be due to systematic force field deficiencies in the ammonium/diammonium guests or in the host, or to missing contributions to the dissociation free energies due to species with protonation state differing from the fully anionic WP6. Predicted WP6-Gn dissociation free energies using vDSSB are quite high and comparable to those observed for CB8 in past SAMPL challenges with similar cationic guests. Nonetheless, due to the smaller cavity of the host, WP6 affinities are expected to be sensibly smaller than that of the structurally related (but larger) CB8 host with similar cationic guests. We may thus infer that our blind prediction of the standard Gn-WP6 dissociation free energies can be affected by a systematic *positive* bias, due to the presence of partially protonated WP6 species at pH=7.4. SAMPL9 organizers are working to obtain pKa values of the WP6 carboxy groups via pH-metric titration so it is possible this information may be available at some point, potentially before the challenge close. In that case, we may try to update our blind prediction by computing the dissociation free energies of the existing WP6 protonated species at the experimental pH using eq (2).

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High-Performance Computing infrastructure and its staff[9]. CRESCO/ENEAGRID High-Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes (see www.cresco.enea.it for information).

References

[1] Andrea Rizzi, Steven Murkli, John N. McNeill, Wei Yao, Matthew Sullivan, Michael K. Gilson, Michael W. Chiu, Lyle Isaacs, Bruce C. Gibb, David L. Mobley, and John D. Chodera. Overview of the sampl6 host–guest binding affinity prediction challenge. *J. Comput. Aided Mol. Des.*, **32** pp. 937-963, (2018).

[2] M. Amezcua, L. El Khoury, and David L. Mobley. Sampl7 host-guest challenge overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations. *J. Comput. Aided Mol. Des*., **35**, pp. 1-35 ,(2021)

[3] SAMPL8 analysis for CB8 and GDCC host. https://github.com/samplchallenges/SAMPL8, DOI: 10.5281/zenodo.4029560

[4] Guocan Yu 1, Xiangyan Zhou, Zibin Zhang, Chengyou Han, Zhengwei Mao, Changyou Gao, Feihe Huang, Pillar[6]arene/paraquat molecular recognition in water: high binding strength, pH-responsiveness, and application in controllable self-assembly, controlled release, and treatment of paraquat poisoning, *J. Am. Chem Soc.*, **134**, pp. 19489-97, (2012)

[5] Nicolas H, Yuan B, Xu J, Zhang X, Schönhoff M. pH-Responsive Host-Guest Complexation in Pillar[6]arene-Containing Polyelectrolyte Multilayer Films. *Polymers (Basel)*, **9**, 719, (2017)

[6] Yang K , Wen J , Chao S , Liu J , Yang K , Pei Y , Pei Z . A supramolecular photosensitizer system based on the host-guest complexation between water-soluble pillar[6]arene and methylene blue for durable photodynamic therapy. *Chem. Commun.*, **54**, pp. 5911-5914, (2018)

[7] Hessz D, Bádogos S, Bojtár M, Bitter I, Drahos L, Kubinyi M. Complexes of carboxylato pillar[6]arene with Brooker-type merocyanines: Spectral properties, pKa shifts and the design of a displacement assay for trimethyl lysine. *Spectrochim Acta A Mol Biomol Spectrosc.*, **252**:119455, (2021)

[8] Hua, B.; Zhang, Z.H.; Sun, J.F.; Yang, J. Pillar[6]arene/acridine orange host–guest complexes as colorimetric and fluorescence sensors for choline compounds and further application in monitoring enzymatic reactions. *Sens. Actuators B*, **255**, pp. 1430–1435, (2018)

[9] Iannone F.; Ambrosino F.; Bracco G.; De Rosa M.; Funel A.; Guarnieri G.; Migliori S.; Palombi F.; Ponti G.; Santomauro G.; Procacci P. CRESCO ENEA HPC clusters: a working example of a multifabric GPFS spectrum scale layout. *International Conference on High Performance Computing Simulation (HPCS)* 2019, pp 1051–1052.

[10] Macchiagodena M, Pagliai M, Karrenbrock M, Guarnieri G, Iannone F, Procacci P. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *J Chem Theory Comput.*, **16**, pp. 7160-7172. (2020)

[11] SciFinder; Chemical Abstracts Service: Columbus, OH; carbon-13 NMR spectrum; spectrum ID CC-03-C_SPC-3734; RN 50-52-2; https://scifinder.cas.org (accessed June 9, 2010).

[12]D. D. Perrin, Boyd Dempsey and E. P. Serjeant. Molecular Factors that Modify pKa Values. In pKa Prediction for Organic Acids and Bases. Netherlands: Springer (1981)

[13] Jarzynski, C., Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.*, **78** , pp. 2690, (1997)

[14] Procacci, P., Unbiased free energy estimates in fast nonequilibrium transformations using Gaussian mixtures, *J. Chem. Phys.* **142**, pp. 154117, (2015)

[15] Dempster, A.P.; Laird, N.M.; Rubin, D.B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. Royal Stat. Soc. B.*, **39**, pp. 1-38, (1977)

[16] Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. J. Am. Stat. Ass., **69**, pp. 730-737, (1974)

[17] Procacci, P., PrimaDORAC: A Free Web Interface for the Assignment of Partial Charges, Chemical Topology, and Bonded Parameters in Organic or Drug Molecules, *J. Chem. Inf. Model.*, **57**, pp. 1240-1245, (2017)

[18] Wang, J., Wang, W., Kollman P. A.; Case, D. A., Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graph. Model.* , **25**, pp. 247260, (2006)

[19] Trott, O., Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**, pp. 455-461, (2010)

[20] Izadi, Saeed; Onufriev, Alexey V., Accuracy limit of rigid 3-point water models *J. Chem. Phys.*, **145**, 074501, (2016)

[21] Essmann, Ulrich, Perera, Lalith, Berkowitz, Max L., Darden, Tom, Lee, Hsing, Pedersen, Lee G., A smooth particle mesh Ewald method, *J. Chem. Phys.*, **103**, pp.8577-8593, (1995)
[22] M Parrinello, A Rahman., *J. Appl. Phys.* **52**, pp. 7182-7190, (1981)
[23] S. Nose, A unified formulation of the constant temperature molecular dynamics methods *J. Chem. Phys.*, **81**, pp. 511, (1984).
[24] Procacci, P., Hybrid MPI/OpenMP Implementation of the ORAC Molecular Dynamics Program for Generalized Ensemble and Fast Switching Alchemical Simulations, *J. Chem. Inf. Model*.,**56**, pp. 1117-1121, (2016)

STRUCTURE AND STABILITY OF C-SI PHASES

Mariarosaria Tuccillo[1], Carla Albenga[2], Priscilla Reale[3], Antonino Santoni[3] and Sergio Brutti[2*]

[1]*University of Rome La Sapienza, Dept. Chemical Engineering Materials and Environment, Piazzale Aldo Moro 5, Rome, Italy*

[2]*University of Rome La Sapienza, Dept. Chemistry, Piazzale Aldo Moro 5, Rome, Italy*

[3]*ENEA, Via Enrico Fermi 45, Frascati, Italy*

ABSTRACT. Here we present a first principles study of the electronic and structural properties of the different Si-C polymorphs (SiC, Si_3C, SiC_3), and the calculation of the binding energy and core level shift for a theoretical prediction of the experimental XPS spectra, to understand the adequacy of the technique for detecting the formation of these interphase on the surface.

1 Introduction

Lithium-ion batteries have revolutionized our way of life, thanks to their large number of applications ranging from portable electronic devices to possible applications in electric-vehicles. The research focuses on the development and improvement of new electrodes and electrolytes to obtain high energy densities, high-power and at the same time long life-cycles. [1] In particular, considering the anodic materials, silicon represents the best alternative to graphite by virtue of its high theoretical capacity (4200 mAh/g) corresponding to the formation of a $Li_{4.4}Si$ alloy, ten times greater than that of carbonaceous materials (372 mAh/g for LiC_6) [2-3]. However, the use of this material as an anode is hindered by some problems: volumetric variations as a result of lithiation/de-lithiation processes greater than 300%; low electrical conductivity; instability of the solid electrolyte interface (SEI).

One way to improve its characteristics is to cover it with thin films of sufficiently elastic conductive materials such as graphene [4]. These thin films are made to increase the electronic conductivity of the silicon. The core-shell silicon nanotubes and nanowires have a higher reversible capacity than other nanostructures of this type (~ 2100 mAh/g), as carbon-coated 1D materials possess a better ability to carry electrons and accommodate volumetric expansions than nearly spherical nanostructures. [2,5] It has also been seen that the type of polymorph Si-C, which is formed during the covering, is a critical factor for increasing performance. In fact, the phase diagram of Si-C presents numerous polytypes in the various stoichiometries, the only thermodynamically stable ones are those with 1: 1 stoichiometry, but there are some dynamically stable structures in the 3: 1 and 1: 3 stoichiometries. [6]

In this scenario, our goal was a first principles study of the electronic and structural properties of the different Si-C polymorphs (SiC, Si_3C, SiC_3), and the calculation of the binding energy and core level shift for a theoretical prediction of the experimental XPS spectra, to understand the adequacy of the technique for detecting the formation of these interphase on the surface. A preliminary study was also conducted on a silicon surface (111) with a graphene layer deposited on top.

2 Methods

All calculations have been performed with the Vienna Ab-initio Simulation Package (VASP), [7] which performs periodic ab-initio quantum mechanical calculations within the Kohn-Sham density functional theory (DFT) framework, [8] with projector-augmented wave potentials and plane wave basis sets. We applied the generalized-gradient approximation (GGA) with the exchange–correlation density functional by Perdew, Burke and Ernzenhof (PBE). [9] In particular Van der Waals (vdW) force correction was also considered for our calculation using the semi-empirical force field method (DFT-

*Corresponding author. E-mail: sergio.brutti@uniroma1.it

D2). We have modelled three different polymorph of SiC based on various stacking patterns in the parent cubic and hexagonal diamond structures: 3C (ABC) F-43m structure, 2H (AB) and 4H (ABCB) $P6_3mc$ structure; Si_3C I-42d structure; SiC_3 Cm structure. [10] We optimized the structural parameters of supercells by relaxing iteratively the ion positions and the cell lattices without any symmetry constraints until the residual force on each atom was <0.01 eV $Å^{-1}$. We used a specific kinetic energy cut-off and a Γ-centered k-point mesh for each structure, based on convergence test.
We have modelled a silicon surface (111) from a cubic silicon cell, used a slab model with 2 x 2 of lateral periodicity. We have chosen this surface since (111) are the most stable surfaces in this structure. Our model consists of 4 layers and 13 Å of vacuum. For geometry optimizations, the positions of the two central layers have been frozen, while the atoms of the uppermost and the bottommost layers have been fully relaxed without symmetry constraints, like the graphene layer.
In order to calculate core-level binding energies (E_c), we worked within the initial state approximation, where in a selected core electron is removed but the remaining electrons are kept frozen. [11] Then E_c is simply given by the Kohn-Sham (KS) eigenvalue of the core electron (ε_{CL}), relative to the Fermi energy (ε_F), as shown in eq. 1:

$$E_c = \varepsilon_{CL} - \varepsilon_F \quad (1)$$

In the PAW formalism of VASP, E_c is computed in two steps: 1) the core electrons are frozen and the valence charge density is computed via the normal, self-consistent electronic relaxation; 2) the KS eigenvalues for the core electrons are solved inside the PAW spheres while keeping the valence charge density fixed. This method involves a non-negligible error in the absolute values of E_c as determined from the KS eigenvalues of DFT are typically 20–30 eV lower than the experimental values reported by XPS due to a breakdown of Koopman's theorem. [12] However, DFT does provide meaningful values of E_c, the shift of the core-level binding energy between two systems. Considering a reference atom, we can calculate core-level shift, as shown in eq. 2:

$$CLS = -(\varepsilon - \varepsilon^{rif}) \quad (2)$$

For this calculation we have built a supercell for each structure: 3x3x3, 3x3x2, 3x2x3 for SiC polymorph, Si_3C and SiC_3, respectively, and 3x3x1 for surface.

3 Results

In table 1 we reported the calculated lattice parameters with literature values for each polymorph. For metastable structures, it was not possible to make a comparison with the experimental data.
Our computational results are in good agreement with literature values. We have calculated the formation energies for each Si-C structure, as shown in fig.3 on the slide, considering as reference pure silicon (diamond structure) and pure carbon (graphite).
The stable polymorphs of SiC (3C, 2H and 4H) have very similar formation energy, lower than that of pure silicon diamond and graphite, as expected. The I-42d and Cm structures of Si_3C and SiC_3, respectively, are thermodynamically unstable with respect to pure silicon diamond and graphite. [6]

Table 1: Structural parameters of the optimized cells and parameters from the literature, in parentheses, for SiC (3C 2H 4H), Si_3C and SiC_3.

Si-C		Structure Parameters (Å)
SiC: 3C	a=b=c	**4.353** ($4.364^{calc.}$ $4.360^{exp.}$) [13]
SiC: 2H	a=b	**3.077** ($3.092^{calc.}$ $3.079^{exp.}$) [14]
	c	**5.048** ($5.074^{calc.}$ $5.053^{exp.}$) [14]
SiC: 4H	a=b	**3.078** ($3.094^{calc.}$ $3.073^{exp.}$) [14]
	c	**10.074** ($10.129^{calc.}$ $10.053^{exp.}$) [14]
Si_3C	a=b	**4.920** ($4.953^{calc.}$) [15]
	c	**9.537** ($9.599^{calc.}$) [15]
SiC_3	a	**10.077** ($10.123^{calc.}$) [15]
	b	**3.134** ($3.148^{calc.}$) [15]
	c	**4.939** ($4.962^{calc.}$) [15]

About surface optimization, we have calculated energy surface, according to eq. 3:

$$\gamma = \frac{E_{slab} - nE_{bulk}}{2A} \quad (3)$$

where E_{slab} is total energy of slab model, E_{bulk} is total energy of bulk system, n is the relationship between the number of total atoms and those present on the surface and A is the superficial area of slab. The value of γ is as small as 0.083 eV/ Å^2, which indicates a non-difficult formations of the surface. Following the relaxation of the structure, the graphene layer rearranges itself into a particular C_{18} ring. We have calculated the adsorption energy (E_{ads}) of graphene on the Si surface per unit of graphene and it assumes a positive value, indicating an un-favourable interaction between the carbon layer and the underlying silicon. This data is in agreement with the positive value of E_{ads} [6].
Overall the binding energies of 1s electrons of C and 2p electron of Si in its two components 1/2 3/2 were calculated for: graphite and silicon diamond (as a reference), polymorphs of SiC (3C, 2H, 4H), Si_3C and SiC_3. In the XPS spectra, the 2p peak of Si is a doublet due to spin-orbit coupling, a phenomenon that affects electrons with angular momentum l different from 0 (l=1 in case of Si). The coupling between l and s (spin moment) gives rise to two different values of J (total angular momentum), and therefore two different electronic states, which are detected in the XPS spectrum. In our computational approach we have not included an adequate representation of the spin-orbit coupling. This approximation inevitably reduces the doublet to a single value of E_c for the p orbitals. The energies of the two components can therefore be calculated using the formula as shown in eq. 4:

$$E_{l,s,j} = E_0 + \frac{A}{2}[j(j+1) - l(l+1) - s(s+1)] \quad (4)$$

where E_0 is the unperturbed energy that does not consider the spin-orbit coupling, A is the effective coupling constant -0.4 eV [16] obtained by knowing the energy difference of the two experimental peaks and s, l, j are, respectively, the quantum numbers of the spin, orbital and total angular momentum associated with the $2p^5$ electron configuration of the ionized Si.
In the case of E_c for C 1s of SiC_3 and Si $2p_{3/2}$ $2p_{1/2}$, we obtain two different values due to atoms with non-equivalent chemical neighborhoods. The calculated values are very far from the experimental values, but in calculating the core-level shift values, the error due to the use of PAW is partially understood and it is possible to obtain values in excellent agreement with the experimental ones as shown in table 2. As regards the surface of silicon (111) with a graphene layer, C 1s gave two BE values which give a negative and a positive CLS. These signals are related to carbon atoms with opposite partial charges produced as a result of the collapse of the regular structure of the C_6 rings. In fact, it has been seen from the structural data that the graphene layer is demolished and at the same time no strong bonds are formed between the Si and the carbon. This trend is consistent with the experimental observations [17]. The data obtained from the calculation of the BEs confirm what emerged from the structural analysis: there is no signal of the formation of Si-C bonds in either of the two spectral regions. The calculation of the BE of the Si 2p provides numerous BEs, each slightly different, without any component attributable to a Si-C bond. In general, the surfaces present complex situations from the point of view of the BEs since the structure undergoes reconstruction and relaxation and the atoms assume different positions on the surface with respect to the bulk. Even the first atomic layers below the surface will not have the same BE as the bulk because they are affected by the displacements of those above it [19].

Table 2: CLS of C 1s and Si 2p, calculated using the 1s electron of graphitic carbon and 2p electron of Si diamond as a reference, compared to literature values in parenthesis.

Core Level Shift [16]		
	C 1s (eV)	Si $2p_{3/2}$ (eV)
Graphite	**0**	
Si diamond		**0**
SiC: 3C	**-0.7** (-0.6) [16]	**1.7** (1.4) [16]
SiC: 2H	**-0.6** (-0.6) [16]	**1.7** (1.4) [16]
SiC: 4H	**-0.7** (-0.6) [16]	**1.7** (1.4)) [16]
Si_3C	**-0.5**	**0.9**
		-0.1
SiC_3	**-0.5**	**1.4**
	0.4	
$Si_{(111)}$ - C	**0.9** (1.1) [17]	**-0.2** (0.3) [17]
	-0.4 (0.1) [17]	

It was possible to reconstruct the XPS spectrum profiles in a semi-qualitative way for C 1s and Si 2p of each polymorph, assuming a Gaussian profile for the peaks with an amplitude value at half height equal to 1.5 eV (arbitrary value).
The structures of the SiC polymorphs (3C, 2H, 4H) would be indistinguishable with XPS spectroscopy alone as they show an almost identical CLS of the C 1s and Si 2p electrons. The metastable structures studied for Si_3C and SiC_3 provide CLS that are different from both stable polymorphs and between them, and distorted peaks due to the greater structural complexity they present.
The relaxation of $Si_{(111)} - C$ system led to a distortion of the hexagonal structure of the graphene layer, as can happen when the carbon layer is deposited on a silicon surface at temperatures below 700 ° C. [19] The theoretical prediction of the XPS spectra highlights the adequacy of this technique to detect the formation of these interphase.

Acknowledgments

This research has received funding by the Italian Ministry of Economic Development through ENEA in the frame of the "Ricerca Sistema Elettrico" PTR program 2019–2021 (Prot.Dip.Chimica Un. Roma La Sapienza 1381).

References

[1] Crabtree, G., Kócs, E. & Trahey, L. The energy-storage frontier: Lithium-ion batteries and beyond. MRS Bull. **40**, pp. 1067–1076, (2015).
[2] Su, L., Jing, Y. & Zhou, Z. Li ion battery materials with core-shell nanostructures. Nanoscale **3**, pp. 3967–3983, (2011).
[3] Ma, D., Cao, Z. & Hu, A. Si-based anode materials for li-ion batteries: A mini review. Nano-Micro Lett. **6**, pp. 347–358, (2014).
[4] Chan, C. K. et al. High-performance lithium battery anodes using silicon nanowires. Nat. Nanotechnol. **3**, pp. 31–35 (2008).
[5] Hu, Y. S. et al. Superior storage performance of a Si@SiOx/C nanocomposite as anode material for lithium-ion batteries. Angew. Chemie - Int. Ed. 47, pp. 1645–1649, (2008).
[6] Gao, G., Ashcroft, N. W. & Hoffmann, R. The unusual and the expected in the Si/C phase diagram. J. Am. Chem. Soc. **135**, pp. 11651–11656, (2013).
[7] Hafner, J.; Kresse, G. The Vienna AB-Initio Simulation Program VASP: An Efficient and Versatile Tool for Studying the Structural, Dynamic, and Electronic Properties of Materials. In *Properties of Complex Inorganic Solids*; Springer US, pp. 69–82, (1997)
[8] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters*, **18**, pp. 3865–3868, (1996).

[9] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *The Journal of Chemical Physics* **98**, pp. 5648–5652, (1993).
[10] V. A. Izhevskyi, L. A. Genova, J. C. Bressiani, A. H. A. Bressiani, Review article: Silicon Carbide. Structure, Properties and Processing, Cerâmica vol.46 n.297 São Paulo Jan./Feb./Mar. 2000.
[11] Köhler, L. & Kresse, G. Density functional study of CO on Rh(111) [78]. Phys. Rev. B - Condens. Matter Mater. Phys. **70**, pp. 1–9 (2004).
[12] García-Gil, S., García, A. & Ordejón, P. Calculation of core level shifts within DFT using pseudopotentials and localized basis sets. Eur. Phys. J. B **85**, (2012).
[13] Thakore, B. Y., Khambholja, S. G., Vahora, A. Y., Bhatt, N. K. & Jani, A. R. Thermodynamic properties of 3C - SiC. Chinese Phys. B 22, (2013).
[14] Sołtys, J., Piechota, J., ŁOpuszyńSki, M. & Krukowski, S. A comparative DFT study of electronic properties of 2H-, 4H- and 6H-SiC(0001) and SiC(000$\bar{1}$) clean surfaces: Significance of the surface stark effect. New J. Phys. **12**, (2010).
[15] Gao, G., Ashcroft, N. W. & Hoffmann, Supporting information: he Unusual and Expected in the Si/C Phase Diagram. *Journal of the American Chemical Society* **135**, pp. 11651-11656, (2013).
[16] NIST X-ray Photoelectron Spectroscopy Database, NIST Standard Reference Database Number 20, National Institute of Standards and Technology, Gaithersburg MD, 20899 (2000).
[17] Hackley, J., Ali, D., Dipasquale, J., Demaree, J. D. & Richardson, C. J. K. Graphitic carbon growth on Si(111) using solid source molecular beam epitaxy. *Applied Physical Lett*er **95**, pp. 133114, (2009).
[18] Terrones, M. et al. Graphene and graphite nanoribbons: Morphology, properties, synthesis, defects and applications. *Nano Today* **5**, pp. 351–372, (2010).
[19] Thanh Trung, P. et al. Direct growth of graphitic carbon on Si(111). *Applied Physical Letter* 102, pp. 2–6, (2013).

REACTIVITY EFFECTS OF COOLANT REGION MODELLING IN THE ALFRED CELL IN CASE OF THERMAL EXPANSION

Roberto Pergreffi*, Giacomo Grasso and Francesco Lodi

ENEA – Italian National Agency for New Technologies, Energy and Sustainable Economic Development Via Martiri di Monte Sole 4, 40129 Bologna - Italy

ABSTRACT. Effects of thermal expansion on the reactivity of a reactor core are an important feedback mechanism both in steady-state and during many postulated accident sequences. It is therefore necessary to model the expanded configuration in terms of shapes, densities and volumes as accurately as possible. Unfortunately, this is not self-evident for those regions that expand differently due to a temperature gradient. This is the case of the coolant region along the active height, where the change in temperatures determines a change in the flow area and consequently in both the mass and volume of the coolant itself. In this study, three alternative modelling approaches are applied to the elementary cell of the ALFRED core and their discrepancy on the reactivity of the system is assessed by detailed analysis with MCNP6.1. From the results obtained, it can be concluded that the MODEL 2 represents an appropriate trade-off between the correctness of the simulation and the affordability of the modelling effort and that, regardless of the model adopted, the expanded values of the pin radius and lattice pitch must refer to the midplane of the active region.

1 Introduction

In a nuclear reactor core, any thermal expansion causes a variation of the geometries and hence of the volumes, which corresponds to a variation in density for those materials where mass is conserved, such as solids, and in mass and density for liquids or gasses. Of particularly interest is the case of regions subject to a non-uniform temperature field, such as the coolant region corresponding to the active zone, where the progressive increase in temperatures determines a continuous change in the flow area. In fact, variation in temperatures of the lower and upper support grids at the contact points with the Fuel Assembly wrapper tube (if present) and gradients in the cladding determine different values of the flow area from the bottom to the top of the active height, which correspond to different axial values of mass and volume, therefore of density, of the coolant. This effect is larger in fast reactors, where the temperature increase by the coolant is larger than in light water reactors.

From a neutronic point of view, the effects of the coolant density on the reactivity in steady-state but also during many postulated accident sequences depend on the reactor type. In the Advanced Lead-cooled Fast Reactor European Demonstrator (ALFRED) [1], these effects are related to the neutron spectrum, and therefore to the average number of fission neutrons emitted per neutron absorbed in the fuel η and to the increase of fission rates for non-fissile isotopes where a threshold in the cross-section exists, but also to the leakage term due to elastic scattering reactions mainly impacting between a few keV and a few MeV [2].

To account for the effects of thermal expansion on the reactivity of the system, it is necessary to model the expanded configuration as accurately as possible. In order to investigate the neutronic effects due to different coolant modelling in the active region of the ALFRED core, three alternative models are tested in a subchannel of the ALFRED core by detailed analyses with MCNP6.1. This comparison, by assessing the discrepancy among the models on the reactivity of the system, is in fact necessary to obtain quantitative information for establishing a reference calculation route catching the aimed details while balancing the required efforts.

*Corresponding author. E-mail: roberto.pergreffi@enea.it.

2 Coolant region models

The three alternative approaches to modelling the coolant region are shown graphically in Fig. ***1*** where, for each of them, half of the axial section of ALFRED hexagonal lattice cell is drawn distinguishing between the fuel pin in grey and the coolant region in white.

In MODEL 1, the volume is calculated using averaged values of pitch and cladding radius whereas the density is calculated independently as a function of the average temperature of the coolant. As no axial changes in volume and density are taken into account, a unique mass value for the entire expanded subchannel along the active height is obtained.

In MODEL 2, the first approach is enhanced by an axial discretization of the active region in multiple sections. Although the volume is again calculated as in MODEL 1, the coolant density is calculated for each axial section using the average temperature of each region, so that the axial change in mass can be taken into account. Unfortunately, this approach does not conserve, with respect to the real case, neither the total coolant mass (equal to that calculated in MODEL 1) nor its distribution along the height, the latter being overestimated in the lower part of the active zone and underestimated in the upper one.

Finally, in MODEL 3, starting from the idea of explicitly preserving the reaction rates by preserving the coolant inventory compared to the real case, the mass is again calculated for each axial section in which the subchannel has been discretized, but applying the physical density to the real-case volume of each region. For this, a more accurate estimate of the volumes is achieved by using the pitch and cladding radius values relative to each section and not to the whole active region as done in the other models. The average pitch is calculated as the value of the function in the middle of each section, the function being the line passing through the points corresponding to the thermal expansions of the lower and upper support grids, while the outer radius is obtained by expanding the cladding at the actual average temperature of the region.

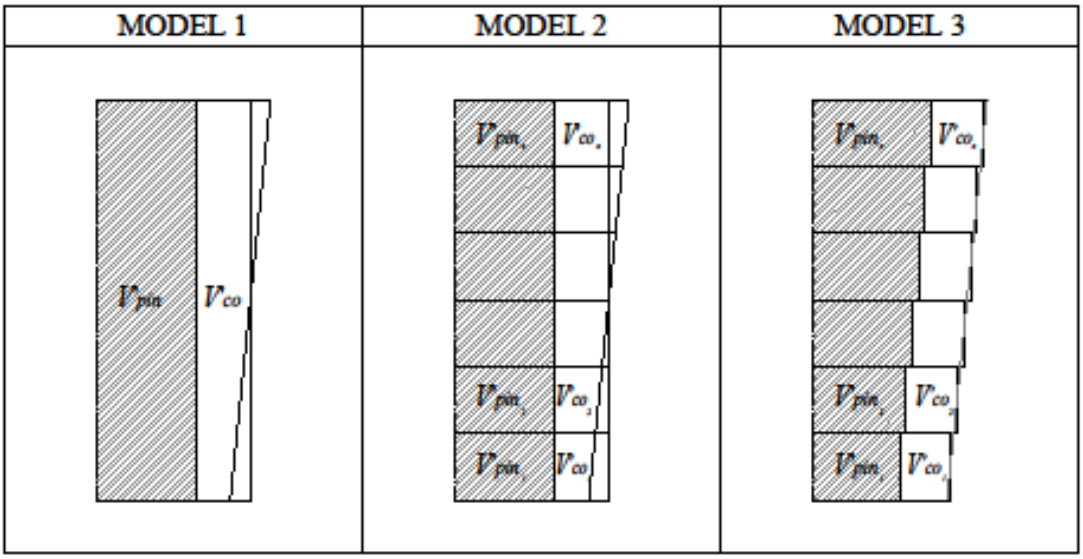


Fig. 1. Scheme of the three approaches to modelling the coolant in the active zone.

3 Coolant modelling in ALFRED

These three MODELs are compared to each other on the basis of their effects on the reactivity of the ALFRED fuel assembly hexagonal lattice. The “Cold Zero-Power” (CZP) state, with all temperatures at about 380 °C, and the “Hot Full-Power” (HFP) state, with temperatures referring to the hottest fuel pin in the hottest FA in the lower enrichment zone, are assumed respectively as initial state and expanded state. All criticality calculations are performed by means of the Monte Carlo code MCNP6.1. A schematic view of the geometry as modelled in MCNP is given in Fig. ***2***. As evaluated nuclear data library, the ENDF/B-VIII.0 distribution is chosen for all calculations. According to the available set in ENDF/B-VIII.0, the operating temperatures for Doppler broadening of the cross-sections are 1200 K for the fuel and 600 K and 900 K for the insulator, coolant and all structures respectively below and

above the active region. Calculations are done assuming axial and radial reflective boundary conditions. To minimize statistical variations, 100 k-eigenvalue source iteration cycles, of which the first 10 skipped, with 1.0E+07 histories per cycle are used. This corresponds to a standard deviation (1σ) of the criticality values never exceeding 3 pcm. All calculations are carried out in CRESCO6 over 60 nodes for a total of 2880 cores.

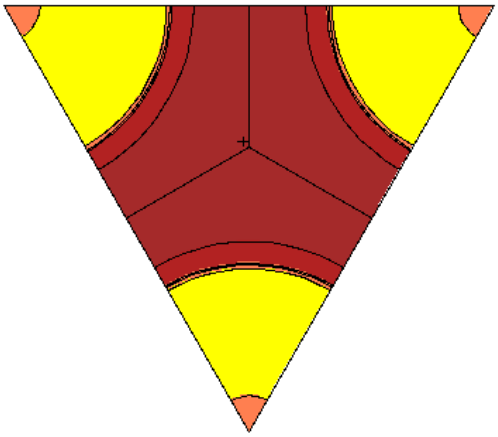


Fig. 2. Subchannel of the ALFRED cell as modelled in MCNP.

The effects on reactivity of the three models are of two types:

- that due to the choice of the reference expanded values of pitch and cladding outer radius,
- that due to the modelling approach per se.

Concerning the first effect, discrepancies are statistically appreciable, being about 20 pcm, between choosing the expanded values of the pitch and cladding outer radius with respect to the midplane of the active zone or to the midplane of the whole fuel pin. In this case, a clearer indication can be drawn in favour of referencing the expanded values of the pin diameter and lattice pitch to the midplane of the active region.
Concerning the second effect, the results indicate the tendency of MODEL 1 to underestimate reactivity. Moreover, if we add the fact that the implementation of MODEL 3 takes longer than that of MODEL 2 (and 1), it can be concluded that MODEL 2 could represent an appropriate trade-off between the correctness of the simulation and the affordability of the modelling effort.

References

[1] G. Grasso et al. An improved design for the ALFRED core, International Congress on Advances in Nuclear Power Plants ICAPP'2019 (Proc. Int. Conf., Juan-les-pins, 2019).
[2] M. Aufiero et al. Analysis of the coolant density reactivity coefficient in LFRs and SFRs via Monte Carlo perturbation/sensitivity, Physics of Reactors: Unifying Theory and Experiments in the 21st Century PHYSOR 2016 (Proc. Int. Conf., Sun Valley, 2016).
[3] J.T. Goorley et al. Initial MCNP6 release overview – MCNP6 version 1.0, internal report LA-UR-13-22934, LANL, Los Alamos, 2013.
[4] J.L. Conlin et al. Release of ENDF/B-VIII.0-based ACE data files, internal report LA-UR-18-24034, LANL, Los Alamos, 2018.
[5] F. Iannone et al. CRESCO ENEA HPC clusters: A working example of a multifabric GPFS spectrum scale layout, 2019 International Conference on High Performance Computing & Simulation HPCS 2019 (Proc. Int. Conf., Dublin, 2019).

WORK FUNCTION OF THE GRAPHENE/HYDROXYLATED GRAPHANE HETEROSTRUCTURE

Francesco Buonocore[1*], Massimo Celino[1], Andrea Capasso[2], Nicola Lisi[1] and Olivia Pulci[3]

[1]*ENEA, Casaccia Research Centre, I-00123 Rome, Italy*

[2]*International Iberian Nanotechnology Laboratory, 4715-330 Braga, Portugal*

[3]*Department of Physics, and INFN, University of Rome Tor Vergata, I-00133 Rome, Italy*

ABSTRACT. We investigate the electronic affinity modifications of graphane by progressive hydroxylation, i.e., by progressively substituting –H with –OH groups, within the density functional theory approach. We found that the electronic affinity of 0.8 eV for graphane can widely change in the 0.28-1.60 eV range depending on the geometric configuration. Hydroxylated graphane has two interfaces with vacuum, hence its electron affinity can be different on each interface, with the formation of an intrinsic dipole perpendicular to the monolayer. We demonstrate that the work function of graphene can be easily tuned by building heterostructures with hydroxylated graphane.

1 Introduction

Graphane is a 2D material with the same honeycomb structure of graphene, but the change of atomic configuration due to hydrogenation modifies the carbon atom hybridization from sp^2 into sp^3, giving rise to a band gap that GW calculations predicted to be in the range 5.4-6.1 eV [1-4]. Similarly to the effect of hydrogenation, simple arrangements of hydroxyl groups (OH) on graphene change the electronic and optical properties of the pristine material. The synthesis of hydroxylated graphane (HyGH), i.e., fully functionalized graphene derivatives with –H, and –OH groups bonded to the carbon atom network in full sp^3 hybridization, and of graphol (highly hydroxylated graphene), was achieved through hydroboration of graphene oxide (GO) followed by protonation [5]. These compounds showed catalytic properties towards the oxidation of biomarkers and even in hydrogen evolution reactions. Large scale (kilogram scale) synthesis of hydroxylated graphene has been reported in [6] through hydrolysis reaction cycles of halogenated graphene. The thermodynamic stability of HyGH structures has been recently theoretically investigated by us [7]. The HyGH structures were built up by progressive substitution of –H with –OH groups and the most stable configuration was selected for each substitution. The formation of HyGH from H_2 and O_2 is an exothermic reaction, and can be more favourable than the formation of water in presence of graphene. Moreover, the stability regions in the phase diagram of the most stable structures have been calculated, proving that HyGH structures with low contents of hydrogen are formed for high O partial pressure, while graphane and HyGH structures with high contents of hydrogen functionalization are formed for high H partial pressure [7].

In the present work we investigate the modification of the electronic affinity of the stable hydroxylated graphane structures. The electronic affinity (EA) is calculated for each geometry, as a function of the level of hydroxylation. Finally, we show that it is possible to tune the work function (WF) of graphene by forming heterostructures with HyGH.

2 Methods

The computational approach was based on a pseudo-potential plane-wave method using the PWSCF code as implemented in the QUANTUM-ESPRESSO package [8], compiled with Intel Fortran compiler, Math Kernel Library (MKL) and Message Passing Interface (MPI) parallelization on the Cresco6 cluster of the high-performance computer facility CRESCO [9], with a good scalability up to hundreds of cores. We used the generalized gradient approximation (GGA) with the Perdew, Burke and

*Corresponding author. E-mail: francesco.buonocore@enea.it

Ernzerhof (PBE) exchange-correlation functional [10]. The pseudo-potential plane-wave calculations were performed using Vanderbilt ultra-soft pseudo-potentials [11]. It is well known that calculations based on GGA functional perform well in the evaluation of structural properties, but underestimate the energy bandgap. Therefore, we also used the more computationally expensive screened hybrid functional of Heyd, Scuseria and Ernzerhof (HSE) [12] which mixes the Hartree-Fock (HF) exchange with the GGA exchange and correlation in the short-range portion of the potential. Indeed, the HSE hybrid functional has been demonstrated to successfully predict the electronic properties of graphene derivatives [13]. We have calculated the HSE total energy using the default value of 0.25 for the fraction of exact exchange (EXX), and then used a value of EXX equal to 0.50 such that the resulting electronic band gap of graphane was 5.4 eV, as predicted by more accurate GW calculations [4]. We indicate this latter functional with the name HSE_2.
All geometry optimizations were performed using the PBE exchange-correlation functional with an energy cut-off for the wave functions of 60 Ry, a cut-off for the charge density of 600 Ry and 12×12×1 k-points Monkhorst-Pack grid. The systems were fully relaxed with a convergence threshold of 0.001 Ry/Å on the inter-atomic forces, and imposing a final stress less than 0.04 GPa. We verified that the chosen cut-offs for the wave functions and charge density allow a convergence of the total energy better than 0.002 eV/atom. In all of the examined structures, we have added O and/or H atoms to the hexagonal 2×2 unit cell of the graphene, which contains 8 carbon atoms. The distance between each monolayer and its periodic image was set to 20 Å. Moreover, in the graphene/HyGH heterostructure calculations, we used the semiempirical Grimme's DFT-D3 correction [14] to take into account dispersion forces.
The HyGH structures here considered are indicated with the symbol GH(n_{OH}OH), where n_{OH} hydroxide groups (OH) substitute the same number of hydrogen atoms in the ideal graphane (GH) supercell. The HyGH structures differ for the percentage of hydroxylated C atoms, varying from 12.5% for GH(1OH) to 100%, for GH(8OH). The GH(n_{OH}OH) structures have all C atoms sp^3 hybridized as in graphane.

3 Results and Discussion

3.1 Electronic affinity

In a material, the EA is calculated as $EA = E_{vac} - E_{MCB}$, where the vacuum energy E_{vac} is defined as the electrostatic potential energy in the vacuum, far away from the system, and E_{MCB} is the energy of the bottom of the conduction band. The EA can be modified by the presence of different adsorbates on the surface. The EA of graphane calculated within the HSE_2 functional results to be 0.79 eV, to be compared with the more accurate GW value of 0.4 eV [15]. Discrepancies between GW and DFT calculations, even when using hybrid functional, are hence expected. PBE and HSE functionals give EA values of 1.35 and 1.02 eV respectively, further decreasing the agreement with the GW value. Therefore, in the following we will discuss the EA referring to HSE_2 results. When hydrogen is adsorbed on graphene, outward-pointing surface dipoles are produced because hydrogen is less electronegative with respect to C. On the other hand, since O is more electronegative than C, an inward-pointing surface dipole is produced that leads to the increase of the EA. The EA modulation of graphane due to functionalization is more complex than the EA behavior of the functionalized surface of a semi-infinite crystal. Indeed, the EA modification of the latter depends on the dipole orientation induced by atoms adsorption at just one interface with vacuum, while graphane has two such interfaces and the substitution of H atoms on one side influences the EA on both sides.

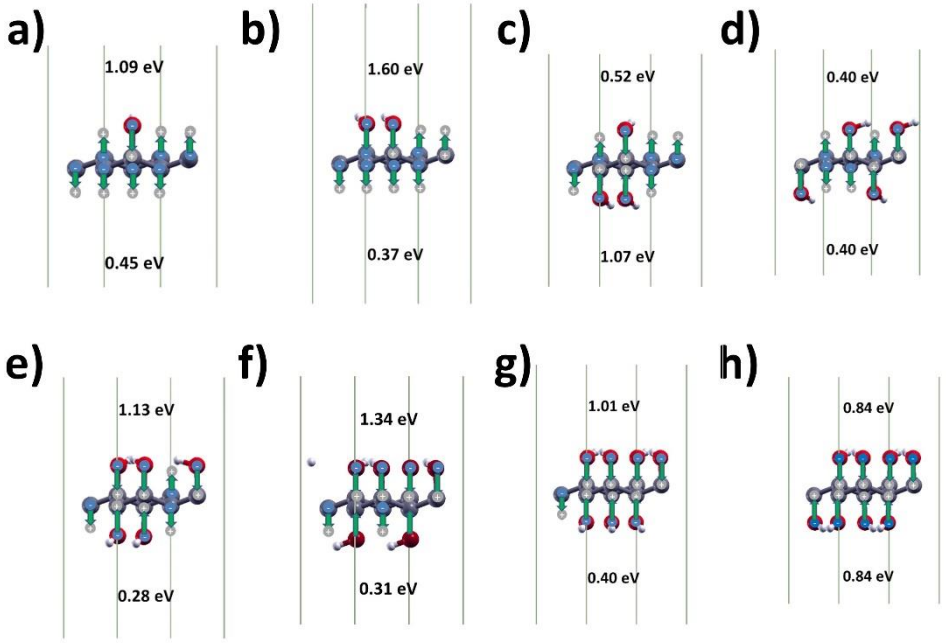


Fig.**1**: Top and bottom electronic affinity of a) GH(1OH), b) GH(2OH), c) GH(3OH), d) GH(4OH), e) GH(5OH), f) GH(6OH), g) GH(7OH) and h) GH(8OH) hydroxylated graphane calculated with HSE_2. The green arrows represent the microscopic electric dipoles along C–O and C–H bonds.

The EA of the up and down symmetric structures GH(4OH) and GH(8OH) is 0.40 and 0.84 eV as shown in Figures 1d and 1h, respectively. Instead, the EA of the remaining non-symmetric structures is different on the two sides of each structure. The lowest EA (0.28 eV) is found for the GH(5OH) on the bottom-side of the carbon lattice shown in Figure 1e. Indeed, for GH(5OH) the three inward-pointing microscopic electric dipoles along C–O on the top region add to the two outward-pointing microscopic electric dipoles along C–H on the bottom in such a way to lower the EA. On the other side, for GH(2OH) there are six microscopic electric dipoles on both sides hindering the escape of electrons on the top vacuum region so that the EA reaches the maximum value of 1.60 eV, as shown in Figure 1b. Therefore, the EA depends on the specificity of the surface geometry and on the presence of different adsorbates. In particular, EA variations due to the presence of adsorbates can be assessed very intuitively by comparing the electronegativity of the adsorbate atoms with respect to the substrate ones. When hydrogen is substituted by hydroxyl groups, inward-pointing surface dipoles are produced that lead to a EA increase on one side and a EA decrease on the opposite side. We found that HyGH EA can be reduced to 0.28 eV, allowing for low-EA 2D materials. However, the hydroxylation process induces asymmetric EA changes because graphane has two interfaces with vacuum, and thus the EA can be different on the two structure sides.

3.2 Work function of the graphene/GH(1OH) heterostructure

As already mentioned, the difference of the ‘upper’ and ‘lower’ EA in the HyGH structures with asymmetric distribution of top and bottom hydroxyl groups is related to the formation of an intrinsic dipole perpendicular to the xy plane. Such functionalized monolayers can then be used to shift the barrier height of any interface, depending on the orientation of the dipole layer. In this way the WF, given by $WF = E_{vac} - E_F$, where E_F is the Fermi energy, can be modified. For example, it is possible to tune the WF of graphene by forming heterostructure with HyGH. In Figure 2 the electrostatic potentials (averaged over planes parallel to the sheet) versus the distance z (parallel to the sheet normal) are shown for graphene and two graphene/GH(1OH) heterostructures. The heterostructures differ for the orientation of the HyGH intrinsic dipole. The flat potential regions on the right of Figure 2 give the vacuum levels for the different structures. For isolated graphene we find a WF of 4.5 eV in good agreement with experiments [16]. When the heterostructures are formed, the distance of the bottom

Hydrogen atoms of C–H from the graphene sheet is 2.9 (3.37) Å when the O–H is on the top (bottom) of GH(1OH). The vacuum level is shifted downwards of 0.6 eV when the OH group is at the bottom of graphane, resulting in a WF of 3.9 eV; in the other case, when the OH is on top of graphane, the vacuum level is shifted upwards of 0.7 eV resulting in a WF of 5.2 eV. The dipole potential of the isolated GH(1OH) (0.64 eV, see Figure 1a) is modified by the electron charge redistribution due to the interaction with graphene. Having a commensurate lattice structure to graphene and affording the WF tunability of different isomers, HyGH family materials could be ideal interlayers in graphene-based devices and 2D heterostructures. Transparent HyGH structures could be designed to serve as interfacial or electron/hole transport layer (ETL/HTL) in optoelectronic devices, such as OLED and solar cells, matching the interfacial energy levels of different materials. [17-18]

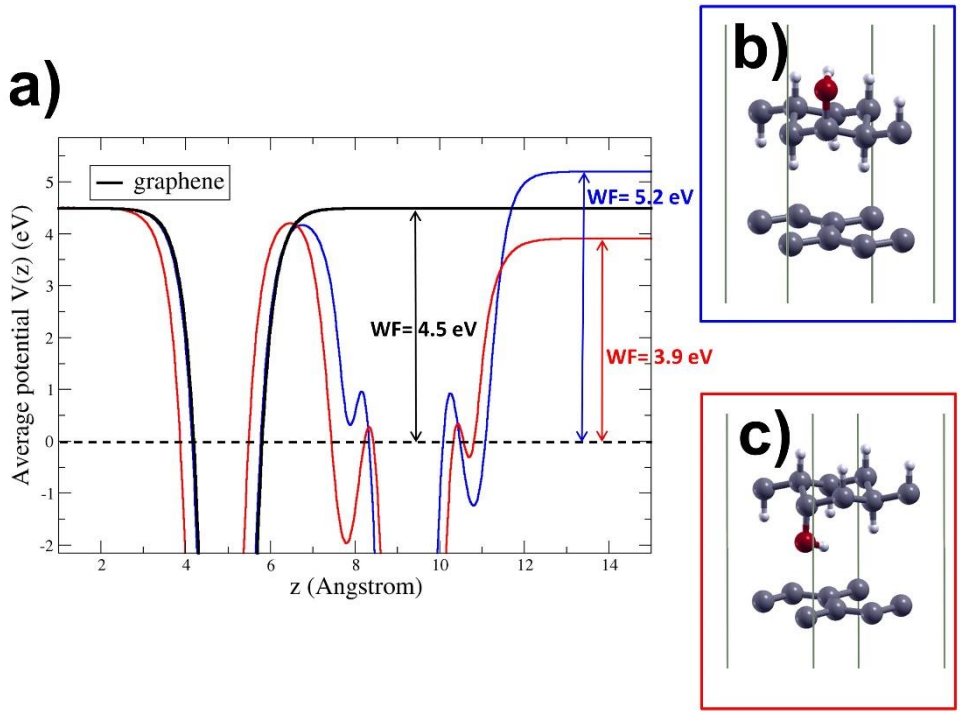


Fig.2: a) Average electrostatic potential V(z) for pristine graphene (black) and two graphene/GH(1OH) heterostructures (in b) and c) insets with blue and red curves, respectively). The horizontal dashed line indicates the Fermi energy.

4 Conclusions

In conclusion, we investigated the modifications of the EA of graphane as a consequence of progressive substitution of –H with –OH groups. The C–C sp^3 bonds determine the large energy gaps of HyGH. The EA is in the range 0.28-1.60 eV depending on the geometric configuration. Together with the already known possible applications of HyGH as novel biomaterial to be used in bone and skin regeneration, we envisage the possibility to use HyGH as 2D material with low EA. Moreover, by exploiting the intrinsic dipole we found in some of the structures, we demonstrate that the WF of graphene can be easily tuned by building GH(1OH) / graphene heterostructures. Finally, heterostructures fabricated with HyGH monolayers may play the role as electron transport layer or hole transport layer in solar cells, depending on the polarity of the 2D crystal.

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure [9] and its staff. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programs, see http://www.cresco.enea.it/english for information. This research

used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. O.P. acknowledges financial funding from the EU MSCA-RISE project DiSeTCom (GA 823728) and from INFN through the TIME2QUEST project. F.B. and M.C. acknowledges financial funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824158 (EoCoE-II). A.C. acknowledges the financial support of the project "GEMIS– Graphene-enhanced Electro-Magnetic Interference Shielding," with the reference POCI-01-0247- FEDER-045939, co-funded by COMPETE 2020–Operational Programme for Competitiveness and Internationalization and FCT–Science and Technology Foundation, under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

References

[1] S. Lebegue, M. Klintenberg, O. Eriksson and M.I. Katsnelson. Accurate electronic band gap of pure and functionalized graphane from GW calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **79**, No. 245117, (2009.

[2] P. Cudazzo, C. Attaccalite C., I.V. Tokatly and A. Rubio. Strong Charge-Transfer Excitonic Effects and the Bose-Einstein Exciton Condensate in Graphane. *Phys. Rev. Lett.* **104**, No. 226804, (2010).

[3] O: Pulci. P. Gori, M. Marsili, V. Garbuio, A.P. Seitsonen, F. Bechstedt, A. Cricenti and R. Del Sole. Electronic and optical properties of group IV two-dimensional materials. *Phys. Status Solidi A* **207**, pp. 291–299, (2010).

[4] O: Pulci. P. Gori, M. Marsili, V. Garbuio, R. Del Sole and F. Bechstedt. Strong excitons in novel two-dimensional crystals: Silicane and germanane. *EPL* **98**, p. 37004, (2012).

[5] H.L. Poh, Z. Sofer, P. Šimek, I. Tomandl and M. Pumera. Hydroboration of Graphene Oxide: Towards Stoichiometric Graphol and Hydroxygraphane. *Chem. Eur. J.* **21**, pp. 8130–8136, (2015).

[6] J. Sun, Y. Deng, J. Li, G. Wang, P. He, S. Tian, X. Bu, Z. Di, S. Yang, G. Ding and X. Xie. A New Graphene Derivative: Hydroxylated Graphene with Excellent Biocompatibility. *ACS Appl. Mater. Interfaces* **8**, p. 10226, (2016).

[7] F. Buonocore, A. Capasso and N. Lisi. An ab initio study of hydroxylated graphane. *J. Chem. Phys.* **147**, p. 104705, (2017).

[8] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, et al. QUANTUM ESPRESSO: A Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys. Condens. Matter* **21**, p. 395502, (2009).

[9] F. Iannone, *et al*. CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 *International Conference on High Performance Computing & Simulation (HPCS)*, Dublin, Ireland, pp. 1051-1052 (2019).

[10] J.P. Perdew, K. Burke and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, pp. 3865−3868, (1996).

[11] D. Vanderbilt. Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B: Condens. Matter Mater. Phys.* **41**, p. 7892, (1990).

[12] J. Heyd, G.E. Scuseria and M. Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, pp. 8207−8215, (2003)

[13] V. Barone, O. Hod, J.E. Peralta and G.E. Scuseria. Accurate Prediction of the Electronic Properties of Low-Dimensional Graphene Derivatives Using a Screened Hybrid Density Functional. *Acc. Chem. Res.* **44**, pp. 269−279, (2011).

[14] S. Grimme, J. Antony, S. Ehrlich and H. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, p. 154104, (2010).

[15] M. Marsili and O. Pulci. The fascinating physics of carbon surfaces: first-principles study of hydrogen on C(0 0 1), C(1 1 1) and graphene. *J. Phys. D: Appl. Phys.* **43**, p. 374016, (2010).

[16] Y.J. Yu, Y. Zhao, S. Ryu, L.E. Brus, K.S. Kim and P. Kim. Tuning the graphene work function by electric field effect. *Nano Lett.* **9**, pp. 3430– 3434, (2009).

[17] A.A.F. Husain, W.Z.W. Hasan, S. Shafie, M.N. Hamidon and S.S. Pandey. A review of transparent solar photovoltaic technologies. *Renew. Sustain. Energy Rev.* **94**, pp. 779−791, (2018).

[18] P. Mustonen, D.M.A. Mackenzie and H. Lipsanen. Review of fabrication methods of large-area transparent graphene electrodes for industry. *Front. Optoelectron.*, **13**, pp. 91–113, (2020).

CEPH FILESYSTEM IN THE ENEAGRID INFRASTRUCTURE

Marco Puccini[2*] and Angelo Mariano[1]

[1]*ENEA, TERIN-ICT, Apulian Office, via G. Petroni 15/F, Bari, Italy*

[2]*ENEA, TERIN-ICT-HPC, Headquarters, Lungotevere Thaon di Revel 76, Rome, Italy*

ABSTRACT. The growing interest in Machine Learning and Artificial Intelligence (AI) algorithms and applications covering multiple heterogeneous fields, entails the availability of modern, flexible and scalable solutions that will allow research teams to solve their tasks. The ENEAGRID infrastructure needs to be adapted to these needs not only in terms of computational resources and tools but also in the realm of data management and storage solutions. This work shows a short description of the Ceph storage cluster setup and benchmark carried out in Brindisi with the aim of finding a suitable solution for high-available, distributed and efficient data management in the contest of Big Data and AI.

1 Introduction

In the last few years, two areas of interest have mainly gained attention in computer science: Artificial Intelligence techniques [1] and data driven applications [2][3]. For both of them, new problems [4] gained attention with consequent development of new systems and technologies dedicated to face such new needs. Among others, data management issues conquered a primary role due to some characteristics, such as growing flows of information, dynamic models for structuring data, accessibility and replication concerns, interoperability [5]. An infrastructure that aims to manage projects in the large field of Machine and Deep Learning, neural networks and/or data lake tasks needs to be integrated with a set of data management tools. The ENEAGRID infrastructure, with its high-performance computing (HPC) CRESCO clusters, aims to offer not only computational resources but also efficient and scalable technologies to manage huge amount of data that needs to be automatically processed. Object storages are a valuable example of such modern technologies. They organize data in distributed systems capable of scaling, growing with the growth of data itself, allowing high I/O performances, managing high throughputs, modelling data as objects labelled with unambiguous keys to make them findable and portable. In this work, we are going to show the activities carried out on a test implementation of Ceph, a modern object storage, with the aim to be integrated in the ENEAGRID data workflows, as a storage support for the growing research activities in the above-mentioned fields.

2 Ceph highlights

Ceph is a modern technology (first stable release in 2012) for distributed storage, born to answer to the needs of managing data intensive scenarios, that is big flows of data not necessarily structured, heterogeneous, with high growth rate [6]. It is a completely open-source technology, which is therefore free from vendor lock-in. A first feature is the ability to use commodity hardware, thus allowing low investments from the start. Ceph has a growing community, with members from leading global IT organizations including Canonical, Cisco, Fujitsu, Intel, Red Hat, SanDisk, SUSE and CERN. Its strength is its scalability, which can be achieved by simply adding nodes to the cluster even with different types of storage drives, both in terms of type, size and speed. It can accept an arbitrary mix of files in terms of size and type. Another peculiar feature of Ceph is the possibility to expose data through 3 different interfaces on the same cluster:

- an object storage
- a block storage
- a POSIX compliant file system (CephFS)

[*]Corresponding author. E-mail: marco.puccini@enea.it

2.1 Ceph architecture overview

Ceph is a distributed and decentralized system so it has no single point of failure. Each node in the cluster has all the intelligence to self-manage and communicate with other nodes, thanks to some daemons that are in charge of performing different tasks. In particular these are:

- **mon** the monitor: it collects the node status, logs and metrics
- **mgr** the manager: it collects all the information from the monitors, keeps an archive of them and exposes them in a dashboard
- **msd** the metadata server: manages information about the data stored in the cluster
- **osd** the object storage device: manages data in the form of objects that can be fragmented and distributed throughout the cluster to ensure replication and accessibility
- **rgw** the RADOS (reliable autonomic distributed object store) gateway daemon: allow the object storage to be reach.

Basically Ceph, through the OSDs, manages data and files as an object storage thanks to the RADOS system [7]. It can be managed directly through different programming languages, both as a gateway compatible with S3 (AWS) and Swift (OpenStack) thanks to the RADOS gateway (RADOSGW), and as a block device through a Linux kernel client or a QEMU/KVM driver, and finally as a distributed file system thanks to a Linux kernel client and with FUSE support.

3 ENEA Ceph cluster architecture

The architecture of the Ceph cluster installed in the ENEA Brindisi research centre was obtained exclusively by recovering nodes of the *old* Galileo cluster, coming from CINECA, in order to configure a test environment with a proper commodity hardware. The configuration of the daemons for each node (with their hostnames) is represented in Table 1.

Table 1: ENEA Ceph cluster daemons map.

Daemon	# Of daemons	hosts
mon	3	ainsteinx09, ainsteinx10, ainsteinx12
mgr	2	ainsteinx09, ainsteinx12
mds	1	ainsteinx09
osd	4	ainsteinx09, ainsteinx10, ainsteinx11, ainsteinx12
rgw	1	ainsteinx12

Some benchmark tests were carried out on the cluster exposed as a file system using fio (Flexible I/O tester) [8], showing good performances both in terms of throughput and latency even compared with both a local SATA disk and another SSD one mounted via iSCSI. While this paper is being published some other tests are being conducted as an object storage. So far, it is proving to be extremely efficient in neural network training workflows. For example, see Fig. 1. As our benchmark shows, even if for small block size, CephFS throughput is comparable with that of a local SATA device, its performance increases a lot by increasing block size and outperforms a SATA device. Different line colors refer to the different queue depth for asynchronous jobs. In the following graphs, a red dashed line shows the average size of images downloaded from Instagram (3-4 MB) as a comparison term.

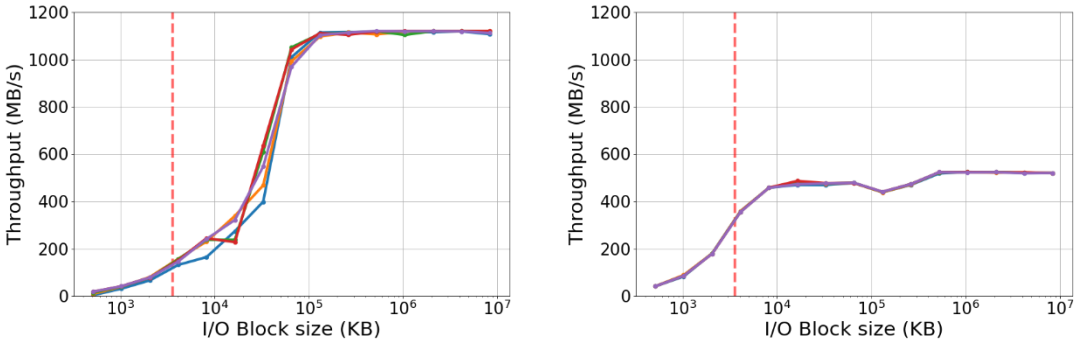


Fig.1: Throughput for different block sizes, at different number of queued jobs with 32 asynchronous read jobs. CephFS (left) vs. Local SATA disk (right).

A detailed benchmark will be the subject of a subsequent ENEA technical note.

3.1 ENEA Ceph cluster administration

Ceph can be installed and managed in multiple ways. The official Ceph recommended method is to use *cephadm*, a tool to deploy and manage clusters of any size with containers (docker or podman are required) and systemd services. It provides an isolated environment to deploy the needed daemons on every host, manage any kind of cluster and admin operation with a dedicated set of CLI administration tools. Among others it provides a dashboard GUI that allow cluster monitoring and management (restricted to a subset of operations). Fig. 2 shows the dashboard GUI installed in our site and how the health of the overall system can be managed.

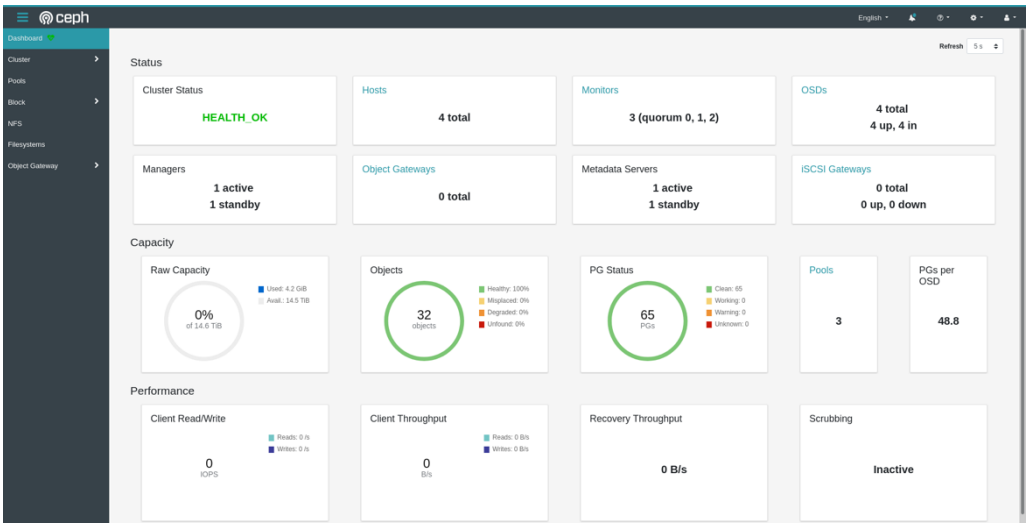


Fig.2: The Ceph dashboard currently in use in ENEA.

3.2 Ceph applications in ENEA

Currently, the Ceph test environment is implemented as a solution to store a set of images coming from the social media crawling connected to data exploitation and analysis required by the current AI projects in the fashion luxury sector being developed by the ENEA ICT Division. Another application is hosting data coming from Crystallographic Information Framework (CIF) in order to investigate molecular structures and tensors related to Crystal Graph Convolutional Neural Networks (CGCNN) and Geographic-enhanced CGCNNs, in the framework of applied research of Deep Learning in the materials for energy sector.

4 Conclusions and future improvements

Ceph has proven to be a viable choice that can efficiently handle large workloads. Thanks to its characteristic of being able to be managed by machines with commodity hardware, it is also a particularly sustainable option from an economic point of view and also in a green perspective. As it is an open-source (and completely free) software, it also guarantees a significant cost reduction against an initial training effort. Its extreme flexibility in configuring its architecture, makes it a technology valuable in different scenarios, and thus, of great interest in the research sector. The wide documentation together with the large community of developers and users finally constitutes a fundamental element in the evaluation compared to its competitors. Being it a mature technology, we are planning for the future to acquire a dedicated infrastructure for a production-ready system that will be seamlessly integrated in the computational facilities in ENEA.

Acknowledgments

We thank G. Cannataro and G. Caretto (ENEA TERIN-ICT-RETE) for their support in the network configuration in the Brindisi Research Center and C. Ronchetti for his involvement as a tester of the ENEA Ceph platform.

References

[1] S. Bianchini, M. Müller, and P. Pelletier. Deep Learning in Science. *ArXiv e-prints* (2020).

[2] K. M. Tolle, D. S. W. Tansley and A. J. G. Hey. The Fourth Paradigm: Data-Intensive Scientific Discovery [Point of View]. *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1334-1337, doi: 10.1109/JPROC.2011.2155130 (2011).

[3] D. Balazka, D. Rodighiero. Big Data and the Little Big Bang: An Epistemological (R)evolution. *Frontiers in Big Data* (2020) vol. 3, p.31, doi: 10.3389/fdata.2020.00031

[4] Daniel Ford, François Labelle, Florentina I. Popovici, et al. Availability in Globally Distributed Storage Systems. *9th USENIX Symposium on Operating Systems Design and Implementation* (OSDI) (2010).

[5] M. Kleppmann. Designing Data-intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. *O'Reilly Media* (2017).

[6] S. A. Weil, S. A. Brandt, E. L. Miller, D. D. E. Long, and C. Maltzahn. Ceph: A scalable, high-performance distributed file system. *Proceedings of the 7th Symposium on Operating Systems Design and Implementation* (2006).

[7] S. A. Weil, A. Leung, S. A. Brandt, Scott, C. Maltzahn. RADOS: A Fast, Scalable, and Reliable Storage Service for Petabyte-scale Storage Clusters. *Proceedings of the ACM Petascale Data Storage Workshop* (2007).

[8] https://github.com/axboe/fio

A STUDY OF TRAFFIC EMISSIONS BASED ON FLOATING CAR DATA FOR URBAN SCALE AIR QUALITY APPLICATIONS

Felicita Russo[1]*, Maria Gabriella Villani[2], Ilaria D'Elia[1], Massimo D'Isidoro[1], Carlo Liberto[3], Antonio Piersanti[1], Gianni Tinarelli[4], Gaetano Valenti[3], Luisella Ciancarella[1]

[1]*ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, SSPT-MET-INAT, 40129 Bologna,*

[2]*ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, SSPT-MET-INAT, 21027 Ispra (VA), Italy*

[3]*ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, TERIN-PSU-STMS, Roma, Italy*

[4]*ARIANET, Via Gian Giacomo Gilino, 9, 20128 Milan, Italy*

ABSTRACT. In this work, we summarise our study where we applied our microscale urban modelling system PMSS (Parallel Micro Swift Spray) at 3 m resolution over a 2x2 km2 domain in the city of Rome for the month of May 2013. By using a NOx emission dataset based on the estimation of real traffic emission, we calculated NOx concentrations, which we compared with referenced NOx concentrations. This work strongly relied on the ENEAGRID HPC facilities managed by ENEA, the cluster we used for the simulations.

1 Introduction

Urban air quality in cities is strongly influenced by road traffic emissions. Micro-scale models have been often used to evaluate the pollutant concentrations at the scale of the order of meters for estimating citizen exposure. Nonetheless, retrieving emissions information with the required spatial and temporal details is still not an easy task.

This paper aims at summarizing the work already published [1]. We conducted microscale simulation of NOx concentrations with the Parallel Micro Swift Spray (PMSS) model [2-4], which were fed by an emission dataset based on Floating Car Data (FCD), containing hourly data for a large number of road links within a 2x2 km^2 domain in the city of Rome for the month of May 2013. All the simulations were performed relying on the computational resources of CRESCO (Computational Centre for Research on Complex Systems) /ENEAGRID High Performance Computing infrastructure [5].

The main objective was to evaluate NOx concentrations simulated by PMSS using FCD-based emissions in comparison to NOx concentrations measured at the Magna Grecia urban air quality station. The possibility of using such detailed emissions, coupled with HPC performance represents a desirable goal for microscale modelling, which allow such modelling systems to be employed in quasi-real time and nowcasting applications.

2 Methodology

We relied on the ENEA's STMS (Systems and Technologies for Sustainable Mobility) laboratory FCD collected by Octo Telematics to get insights into travel patterns of private cars and to estimate traffic emissions in the case study (see [1] for details). The FCD used in this study represents about 5% of the circulating passenger cars in the study area. We estimated the emissions of the remaining part of the circulating fleet by considering the vehicle population in the city of Rome for the year 2013 and the public registers of vehicle licenses [6]. Scaling factor for each of the remaining vehicle categories not included in the FCD database were then calculated, in order to extrapolate their fluxes starting from the passenger cars hourly fluxes.

*Corresponding author. E-mail: felicita.russo@enea.it

The emission data were obtained by deploying TREFIC [7], which is the native traffic emissions pre-processor for the model PMSS, and it is based on the COPERT 4 [8] methodology for the calculation of the road vehicles emission factors. TREFIC takes as input vehicle type, fuel consumption, average travelling speed, road type, and time modulation factors of flow, speed and temperature to produce emission factors (EFs) for each road type. These pieces of information allow to generate a modulated input to the dispersion model and they support the user in generating time emission profiles for traffic flows, speed and temperature (see [1] for more details).
Micro-SWIFT-SPRAY (PMSS) is a modelling system which reproduces primary pollutant transport and dispersion at the microscale and calculates the dry and wet deposition of airborne chemical species. PMSS is the parallelized version of the MSS model suite. The system has two pre-processing phases for the meteorological and the emission data, respectively. These modules prepare the input for the main processing models, the meteorological driver PSWIFT, an analytically modified mass consistent interpolator over complex terrain, and the three-dimensional Lagrangian particle dispersion model, PSPRAY. For more details information please refer to [2], [3], [4]. The modelling system PMSS is a commercially available software developed by ARIANET [9]. The codes PSWIFT and PSPRAY used here, in their versions PSWIFT-2.1.1 and PSPRAY-3.7.3, were compiled with Intel16 compiler, using OPENMPI library.
We performed simulations for a period of 29 days from May 2nd to May 30th 2013, over 2 x 2 km2 horizontal domain that was centred on the urban traffic air quality (AQ) Station Magna Grecia. The chosen domain (see [1]) consisted of 667x667 gridpoints, and covered an adequate fraction of the emissions of the city with the AQ station location far enough from the domain border, where the model uncertainty is generally higher. The spatial resolution of 3 m was used to ensure the highest resolution possible while keeping run time within acceptable values. For the vertical grid, we chose the following 25 levels above the ground from 0 m to 700 m.
Meteorological data used to feed the diagnostic model PSWIFT were provided by the Weather Research and Forecasting [10] mesoscale model using ERA5 reanalyses [11] as boundary conditions at 28 km resolution and 3-hourly time-step. WRF simulations, performed using 3.9.1.1 model version, were based on two-way nesting over three grids, the coarser one covering whole Italy at 9 km horizontal resolution (over Italy), then an intermediate domain at 3 km resolution (over Lazio region) and finally the target domain at 1 km resolution over the city of Rome.
For the NOx emissions, we used the hourly fluxes of passenger cars provided by the ENEA's STMS database, where analyses were conducted to estimate the variability and the statistical significance of the modulation profiles computed from the emission database, in particular close to the monitoring station (see [1], for more details).
We conducted the simulations on the CRESCO/ENEAGRID High Performance Computing infrastructure funded by ENEA [5]. As the model system was structured, each 29-day-simulation consisted of 29 single model runs, each simulating 24 h and for each simulated day the values for the pollutant concentrations calculated for the last hour were saved and used as initial condition for the following run.
Our simulation duration depends mainly on the number of emitted particles, which depends on the concentration resolution required (i.e., the concentration contribution given by a single particle in a concentration cell). By using a concentration resolution of 0.5 mg/m3 for the NOx specie and using 528 cores for PSPRAY, which represents the most CPU demanding part of the system, the CPU time per core per simulated day was 8148s.

3 Results and discussion

We compared hourly NOx observed and modelled concentrations at Magna Grecia AQ station (Figure 1), where the absolute value of the daily fractional bias was considered as we are interested in discussing its magnitude rather that its sign. In general, NOx simulation shows a good agreement with the hourly NOx concentration variation with an average fractional bias of 0.2 across the period. This is confirmed by the daily fractional bias that shows values below 20% for half of the days and only for 7 days above 30%. Higher discrepancies are indicated by low values of R-squared. Generally, for almost half of the simulated days R-squared values larger than 0.4 are found.

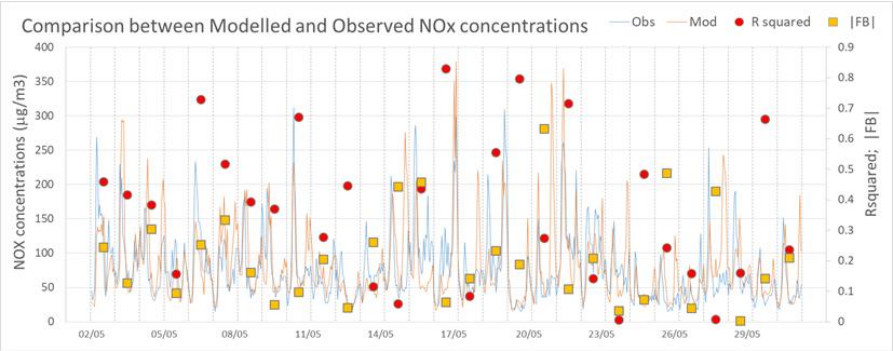


Fig. **1**: Hourly NOx concentration comparisons between observation (blue line) and simulation (orange line) with the linear correlation coefficients (red circles) and the absolute value of the daily fractional bias (yellow squares).

The reasons behind low correlations with low fractional bias can be ascribed to the three chosen modulations that were not always the most representative of the vehicle flow variability in all the 29 days period. These comparisons show nevertheless that the daily fractional bias is generally low, indicating that the background concentrations are fairly well transported by the model and that the reduced modulations used had little influence on most of the daily averages.
To evaluate our model performance, we used the acceptance criteria proposed by Ghermandi et al. 2020 [12], based on: Fractional bias (FB); Normalized Mean Squared Error (NMSE); Fraction of simulations within a factor of two of the measurements (FAQ2); Normalised Absolute Difference (NAD). We obtained FB = 0.049, NMSE = 0.37, FAC2 = 0.87, NAD = 0.20, which are values within the range set by the acceptance criteria (see Table 1).

Table 1: Definition of the acceptance criteria for dispersion model evaluation.

Type of measuring station	|FB|	NMSE	FAC2	NAD
Urban	|FB|≤0.67	≤6	≥0.3	≤0.5

4 Conclusions

The comparison between our model performances and the acceptance criteria indicates a good agreement of our simulation with observations. The innovative aspect of this work is the use of FCD computed emissions to feed our urban scale model. However, the use of massive FCD to extract traffic patterns and travel behaviours in urban areas is still at an early stage compared to other sectors, particularly due to the fragmented availability of transport/mobility data, institutional barriers and data privacy/security issues. The dataset used in this study was based on data collected by only 5% of the circulating passenger cars), and the extrapolation of this dataset to the entire fleet could represent an important approximation. Therefore, the 87% of the simulated values within a factor of two with respect to the measurements represents a remarkable success, even if the calculated linear correlations reported valued above 0.6 for few days only.

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [5]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programs, see http://www.cresco.enea.it/english for information.

References

[1] Russo,F.; Villani ,M.G.; D'Elia, I.; D'Isidoro, M.; Liberto, C.; Piersanti, A.; Tinarelli, G.; Valenti, G.; Ciancarella, L. A Study of Traffic Emissions Based on Floating Car Data for Urban Scale Air Quality Applications. Atmosphere 2021, 12, 1064. https://doi.org/10.3390/ atmos12081064

[2] Oldrini, O.; Armand, P.; Duchenne, C.; Olry, C.; Moussafir, J.; Tinarelli, G. Description and Preliminary Validation of the PMSS Fast Response Parallel Atmospheric Flow and Dispersion Solver in Complex Built-Up Areas, Environ. Fluid Mech. 2017, 17, 997–1014.

[3] Tinarelli, G.; Mortarini, L.; Trini Castelli, S.; Carlino, G.; Moussafir, J.; Olry, C.; Armand, P.; Anfossi, D. Review and Validation of Microspray, a Lagrangian Particle Model of Turbulent Dispersion. Geophys. Monogr. Ser. 2012, 200, 311–327.

[4] Tinarelli, G.; Gomez, F. PSPRAY General Description and User's Guide, 2017. Version Code 3.7.3.; ARIANET/ARIA Technologies: Milan, Italy, 2017. [10] Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. A Description of the Advanced Research WRF Version 3. NCAR Tech. Note 2008, NCAR/TN-475+STR; University Corporation for Atmospheric Research: Boulder, CO, USA, 2008; 113p.

[5] Iannone, F.; Ambrosino, F.; Bracco, G.; De Rosa, M.; Funel, A.; Guarnieri, G.; Migliori, S.; Palombi, F.; Ponti, G.; Santomauro, G.; et al. CRESCO ENEA HPC clusters: A working example of a multifabric GPFS Spectrum Scale layout. In Proceedings of the 2019 International Conference on High Performance Computing Simulation (HPCS), Dublin, Ireland, 15–19 July 2019; pp. 1051–1052.

[6] ACI. Available online: http://www.aci.it/laci/studi-e-ricerche/dati-e-statistiche.html (accessed on 29 March 2021).

[7] ARIANET. TREFIC—Traffic Emission Factors Improved Calculator. 2014. Available online: http://www.aria-net.it/front/ENG/codes/files/7.pdf (accessed on 19 April 2021).

[8] Emisia SA. COPERT—COmputer Programme to Calculate Emissions from Road Transport. 2018. Available online: https://www.emisia.com/utilities/copert/documentation (accessed on 11 June 2020).

[9] ARIANET. ARIA(NET). Available online: http://www.aria-net.it/ (accessed on 20 April 2020).

[10] Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. A Description of the Advanced Research WRF Version 3. NCAR Tech. Note 2008, NCAR/TN-475+STR; University Corporation for Atmospheric Research: Boulder, CO, USA, 2008; 113p.

[11] Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al.The ERA5 global reanalisys. Q. J. R. Met. Soc. 2020, 146, 1999–2049.

[12] Ghermandi, G.; Fabbri, S.; Veratti, G.; Bigi, A.; Teggi, S. Estimate of Secondary NO2 Levels at Two Urban Traffic Sites Using Observations and Modelling. Sustainability 2020, 12, 7897.

RADIOPHARMACEUTICALS MOLYBDENUM-99 AT ENEA TRIGA RC-1 REACTOR: EVALUATIONS FOR A CLOSED CYCLE PRODUCTION IN SHORT SUPPLY-CHAIN

Luigi Lepore[1*] and Nadia Cherubini[2]

[1]*ENEA, FSN-FISS-CRGR, Nuclear Material Characterization Laboratory and Nuclear Waste Management ENEA Casaccia Research Center Via Anguillarese, 301, 00123 Roma, RM, Italy*[1]

[2]*ENEA, FSN-FISS, Nuclear Fission Technology, Facilities and Materials Division ENEA Casaccia Research Center Via Anguillarese, 301, 00123 Roma, RM, Italy*

ABSTRACT. Molybdenum-99, with its daughter Technetium-99m, represents one of the most widely used radionuclide and radiopharmaceutical for medical diagnostics, respectively. Its availability all over the world is overcoming shortages because of production reactors decommissioning and difficulties related to transportation within the long supply chain among Countries. Moreover, the International Atomic Energy Agency is promoting variations in traditional production methods by reducing the usage of high enriched uranium as raw material (Mo-99 is extracted as a fission product of Uranium-235) and suggesting alternative technologies possibly using none of uranium at all. ENEA is studying activation methods to produce Mo-99, where raw material used is Molybdenum itself, at different neutron facilities. The key point is that Mo-99 produced by activation is requesting to develop technological solutions for the whole new supply chain: chemical processes, manipulation devices, transportation canisters, extraction methodologies, etc. Monte Carlo calculations have been extensively used to perform preliminary evaluation regarding system, structures and component to face the whole Mo-99 supply chain foreseen at ENEA facility, demonstrating a first-step feasibility of this new production method.

1 Introduction

The MOLY Project from ENEA has been developed in the framework of worldwide effort to keep and increase the availability of Tc-99m radiopharmaceutical.

Traditional production methods foresee the irradiation of high-enriched uranium targets (enrichment in U-235) in thermal reactors, from which Mo-99 (father of Tc-99m) is separated by means of chemical processing of the target itself [1].

Alternative ways to produce Mo-99 are foreseen and promoted to overcome two main issues: i) limitation in high-enriched uranium usage in civil applications; ii) diversification in production techniques and capillary diffusion of production centres near to users, to increase availability and a stable supply chain [2][3].

ENEA has investigated the possibility to produce Mo-99 at the TRIGA RC-1 reactor by means of neutron activation starting from Mo-98 in molybdenum samples, instead of uranium targets. The reactor is a thermal neutron pool-type water-cooled 1 MW facility. This alternative production method implies the revision of the whole process, regarding chemical processing of the sample and Mo-99 recovery, especially.

Mo-99 from activated samples is mixed with the Mo-98 matrix, and chemical purification of the sample is recovering both. The point here is a lower concentration of Mo-99 per unit mass of Molybdenum recovered with respect to the traditional method starting from uranium samples, in which Mo-99 is a fission product and it is directly separated during the chemical processing. Such a specific feature regarding this lower concentration of Mo-99 needs for a revision of the Tc-99m generator technology also [4][5][6].

[*]Corresponding author. E-mail: luigi.lepore@enea.it

ENEA has developed a feasibility study regarding Mo-99 production from neutron activation at the TRIGA RC-1 reactor. Particularly:

- reactor calculations to a weekly production of Mo-99 (2 TBq per week);
- procedures and machinery design to remote extraction and safe manipulation of irradiated samples;
- design of a transportation container to ship activated samples from the reactor hall to the facility exploiting the chemical processing;
- identification and development of a chemical process from the irradiated sample to the Tc-99m generator;
- chemical recovery of Mo-98 from exhausted generators to rebuild the sample to be newly irradiated in reactor to a new production batch, as to close the target cycle, to enhance recycling of materials, and to minimize the radioactive waste produced.

Research actions, theoretical and experimental, carried out at ENEA Research Nuclear Reactors Laboratory and ENEA Nuclear Material Characterization Laboratory, a MS thesis, a collaboration contract with SAPIENZA – University of Rome, have produced several internal reports covering all aspects before.
The paper herein is discussing all simulation efforts carried out in the framework of activities covered within the MOLY Project. Simulation carried out with the Monte Carlo N-Particles code, MCNP, have exploited the usage of multi-processing on CRESCO High Parallel Computing Resources. The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [7]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.
In particular, the irradiation target development has requested 172800 hour per core, using 720 threads per 240 hours; calculations regarding shielding purposes have requested 36000 hour per core, using 360 threads per 100 hours; dosimetry in chemical processing procedures has requested 18000 hour per core, using 360 threads per 50 hours.

2 Methodology

Currently, the most efficient method to simulate the radiation transport in media through complex (many volumes or “cells”, several radioactive sources, and different materials) and large (few centimetres to hundreds of meters) geometries is the Monte Carlo simulation.
By this approach, a lot of ‘histories’ of the particle of interest are run, in order to reproduce, into a virtual world, all the possible paths travelled by the radiation, by means of random samplings from all the probability density functions ruling physical phenomena. Very complex geometries can be built, and very accurate results can be produced if all the elements important to the transport of the analysed radiation are described into the model; moreover, an adequate computational power should be available for running a statistical significant number of histories. For this work, the MCNPX (Monte Carlo N-Particle eXtended) code has been used, in its version MCNPX 2.5.0 [8].

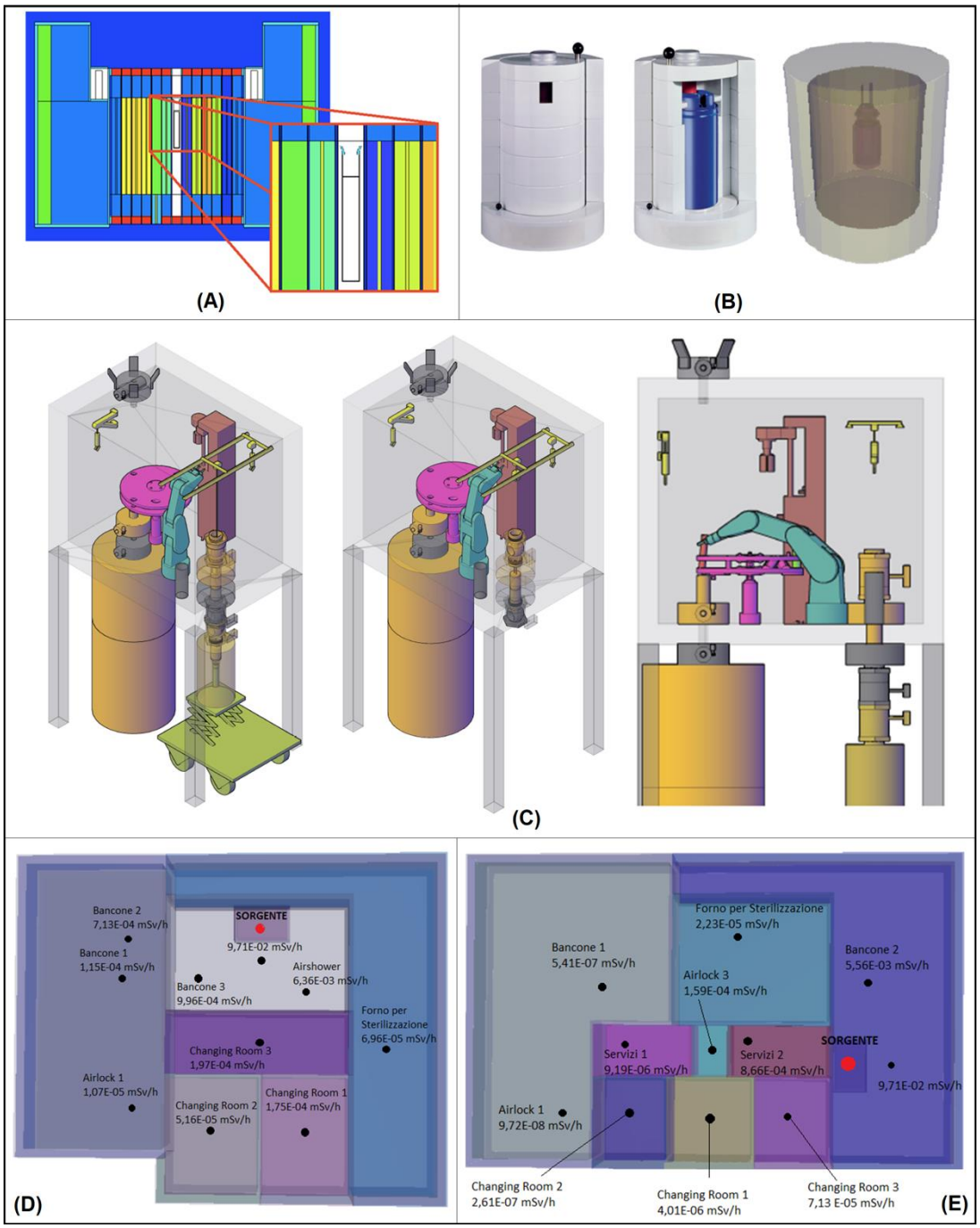


Fig.1: Sketches reporting some simulation activities: (A) reactor simulation for target activation developments; (B) Tc-99m generator design and shielding; (C) activated samples remote handling device development, with the shipping container; (D) and (E), some layouts of the facility designed for chemical processing of irradiated samples and Tc-99m generators production.

Monte Carlo calculations have been used extensively in reactor simulation first, as to obtain the nominal neutron flux for samples to be exposed for activation. Experimental characterizations have benchmarked simulated results as well as experimental activations of little quantities of Mo-98 to produce Mo-99.
For the current industry, Mo-99 production become significant when in the order of tens-hundreds of TBq. Even if TRIGA RC-1 production is lower (2 TBq per week), it could be of interest for a local supply of Mo-99/Tc-99m, and management of a 2 TBq sample require some remote handling device for radiation protection purposes. In the light of that, a fully remote machinery to handle the activated sample, to separate it from the container, and to get it into the shipping container has been conceived

and developed by means of Monte Carlo simulation. Beside shielding purposes, calculation about dose rates conferred to electronic components have been carried out.
Simulations have helped also in designing the layout of the facility for chemical processing of the activated sample, as well as all machinery devoted to target manipulation, dissolution, separation and extraction of the target nuclide Mo-99. Because of the low concentration of Mo-99 in Molybdenum as noted before, the revision of the Tc-99m generator technology has exploited Monte Carlo calculation also, and the same applies to the Mo-98 recovery process from exhausted generators.

3 Results obtained and future perspectives

The extensive work conduced and briefly described herein is constituting significant part of the Mo-99 production feasibility study carried out at the TRIGA RC-1 reactor facility at the ENEA Casaccia Research Centre.
Those results achieved are producing elements promoting further investigations, getting experience from the laboratory scale to a higher step, as to verify the effective capabilities of solutions now identified.
Monte Carlo simulation has played a pivotal role in such a study, representing one of the most powerful tools for scientist working with radioactive materials domain and all related issues.

References

[1] Technetium-99m pharmaceuticals, Zolle, Ilse. Vol. 1, Springer, 2007.
[2] Non-HEU Production Technologies for Molybdenum-99 and Technetium-99m. Number NF-T-5.4 in Nuclear Energy Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna, 2013.
[3] Feasibility of Producing Molybdenum-99 on a Small Scale Using Fission of Low Enriched Uranium or Neutron Activation of Natural Molybdenum, Number 478 in Technical Reports Series. INTERNATIONAL ATOMIC AGENCY, Vienna, 2015.
[4] Studio di fattibilità inerente il processo di attivazione di target per la produzione di radiofarmaci presso il reattore TRIGA RC-1 del C.R. ENEA Casaccia, Tesi magistrale, 2020
[5] Sankha Chattopadhyay, Sujata Saha Das, Malay K Das, and Naresh C Goomer. Recovery of 99mTc from Na2 [99Mo] MoO4 solution obtained from reactorproduced (n, γ) 99Mo using a tiny Dowex-1 column in tandem with a small alumina column. Applied Radiation and Isotopes, vol.66, no.12,pp.1814-1817,2008.
[6] Chattopadhyay, Sankha and Das, Sujata Saha and Barua, Luna and Pal, Asit Kumar and Kumar, Umesh and Alam, Md Nayer and Hudait, Arup Kumar and Banerjee, Sharmila and others A compact solvent extraction based 99Mo/99mTc generator for hospital radiopharmacy Journal of Applied Radiation and Isotopes, vol.143, pp. 41-46, 2019.
[7] F. Iannone et al., “CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout,” in 2019 International Conference on High Performance Computing Simulation (HPCS), 2019, pp. 1051–1052.
[8] J. F. Briesmeister, Ed., “MCNP: A General Monte Carlo N-Particle Transport Code,” 2000.

HYBRID SIMULATIONS ON ALFVÉNIC INSTABILITIES AND ENERGETIC PARTICLE TRANSPORT IN REVERSED SHEAR PLASMAS

Tao Wang[1,2*]

[1]*Institute for fusion theory and simulation, Zhejiang University, Hangzhou 310027, China*

[2]*Center for nonlinear plasma science and ENEA C.R. Frascati, Frascati 00044, Italy*

ABSTRACT. This report describes the research activities on CRESCO infrastructures during the calendar year 2020. The energetic particle (EP) transport by the synergetic effect of reversed shear and toroidal Alfvén eigenmodes is analysed by the hybrid MHD-kinetic code simulations. The results suggest that the relay of modes in different locations may create an efficient EP transport route.

1 Introduction

In the tokamak experiments aiming at controlled thermonuclear fusion, the auxiliary plasma heating and current drive (H&CD) largely relay on external power inputs, such as neutral beam and radio frequency wave injection. A small but finite population of energetic particle (EP) is routinely created in this process and is crucially responsible for efficient H&CD. Furthermore, Deuterium-Tritium fusion reaction generates 3.5 MeV alpha particles, which are needed to be well confined so as to sustain the thermal plasmas burning. However, various types of collective instabilities may be driven unstable by the inherent free energy of the nonuniform EP distribution function. In particular, it is well known that various branches of the shear Alfvén wave (SAW) eigenmode could be readily excited via wave-EP resonant interactions [1]. Once driven to a finite amplitude, the SAW instabilities could induce significant anomalous EP transport, which may lead to performance degradation and even device damage. Note that the interaction between SAW eigenmode and EPs is inherently non-perturbative and self-consistent, i.e., the mode structure and frequency could promptly follow the evolution of the EP distribution function. Thus, numerical simulation is a viable tool to study the complicated underlying physics.

In tokamaks, two of the most common SAW eigenmodes are the reversed shear Alfvén eigenmode (RSAE) and toroidal Alfvén eigenmode (TAE). The former is unique to reversed shear plasma, i.e., the radial profile of the safety factor q contains an off-axis minimum, denoted as q_{min}. The magnetic shear vanishes at q_{min}, and the RSAE could exist as a radially localized eigenmode due to minimized phase mixing dissipation. On the other hand, TAE is ubiquitous in toroidal plasmas, as coupled by two counter-propagating poloidal harmonics by toroidal effects. In typical conditions, RSAE resides deeply in the plasma core region, and TAE is usually closer to the plasma boundary, where the toroidal effect is more significant [2].

Due to the spatial separation, the RSAE and TAE are usually considered separately when studying their effect on the EP confinement property [3]. However, recent proof-of-principle simulations demonstrate that the inner-core localized RSAE could convectively pump the EPs outwards, while the TAE in the outer-core region may further induce EP diffusion via phase space stochasticity [3]. Thus, a synergistic transport route [4] is possible, where the EPs could be efficiently transported by the relay of RSAE and TAE. In this work, we attempt to use numerical simulations to demonstrate the possibility of this transport mechanism, and to delineate its general properties.

Section 2 describes the simulation setup, the main simulation results are presented in section 3, followed by a brief summary.

*Corresponding author. E-mail: t_wang@zju.edu.cn

2 Simulation setup

The simulations are performed by the hybrid code HMGC, which is originally developed in Frascati laboratories [5]. HMGC uses reduced ideal MHD equations for the electromagnetic wave fluctuations, and drift-kinetic description of the energetic particle pressure. The parameters are close to the typical ones of the HL-2A tokamak, with minor radius $a = 0.40$ m, $R_0 = 1.65$ m, on-axis magnetic field $B_0 = 1.0$ T. The equilibrium q profile is shown in figure 1(a), with $q_{min} = 1.92$. In the present work, we limit the analysis to $n = 3$ modes and neglect other perturbations. Here, n is the toroidal mode number.

The EPs are Deuterium beam ions and are described by an anisotropic slowing-down distribution function with birth energy $E_b = 45$ keV and critical energy $E_c = 28$ keV. The initial radial density profile is shown in figure 1(b), the on-axis EP density ratio is $n_{H0}/n_{i0} = 3\%$.

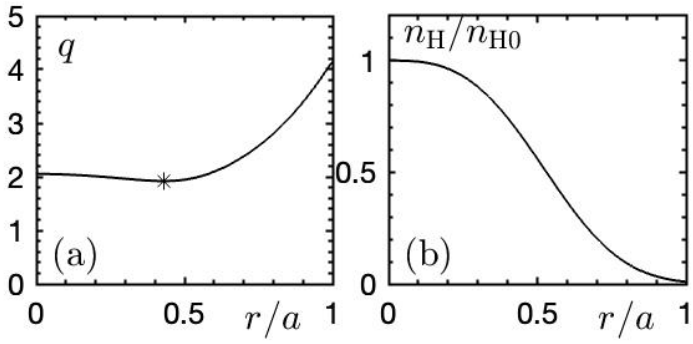


Fig.**1**: (a) equilibrium q radial profile, with the radial location of q_{min} denoted by an asterisk. (b) the initial EP density profile n_H/n_{H0}.

The MHD equations in HMGC are solved by semi-implicit, second-order finite difference method. The kinetic module, on the other hand, is implemented by the particle-in-cell method. Typical value of the computational particles in this work is about 12.6 million. The Calculation of the particle orbits consists most of the computational cost, the particles are decomposed by the number of assigned nodes and accelerated by OpenMP. Usually, a simulation case costs about one day using $4 * 48$ processors on CRESCO6.

3 Simulation analysis

We first give an overview of the simulation case in figure 2, which shows the time evolution of the total energy of all Fourier harmonics in the simulation. One sees that after the initial phase mixing process, an exponentially growing mode is formed. By inspecting the radial mode structure and frequency as shown in figure 3, the unstable mode is readily identified as a RSAE dominated by the $m = 6$ harmonic. Here, m is the poloidal mode number.

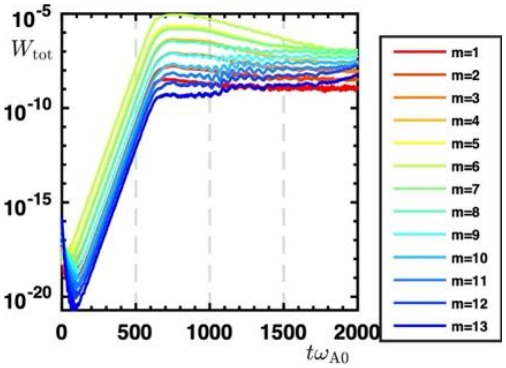


Fig.**2**: time evolution of the mode energy.

The RSAE saturates at about $t\omega_{A0} \cong 600$, with ω_{A0} being the on-axis Alfvén frequency. After the saturation, the mode amplitude gradually decreases, as the resonant EPs are pumped out in the core region and the mode drive decreases significantly. Starting from $t\omega_{A0} \cong 1100$, one sees that the amplitude of larger m modes slowly increase, and eventually becomes dominate at the later stages of the simulation ($t\omega_{A0} \cong 1800$). The larger m mode is a TAE located at the outer-core region, as shown in figure 4.

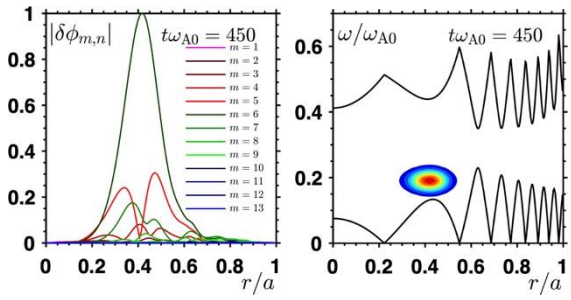


Fig.**3**: (left) radial structures of Fourier harmonics in the exponentially growing stage; (right) the corresponding integrated frequency spectrum, the solid curves are the SAW continuous spectra.

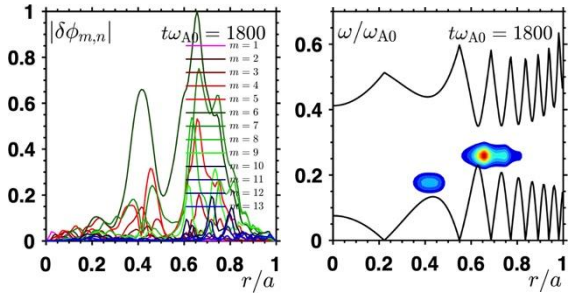


Fig.**4**: Same as figure 3 in the late nonlinear stage.

In order to further understand the excitation mechanism of the outer-core TAE, we first inspect the radial EP density profile. Figure 5 shows the results. We can see that at the time of the RSAE saturation, the EP density is only mildly perturbed. More significant transport appears only in the late nonlinear stage, consistent with the analysis above. From another perspective, we compare the phase space diagram of wave-EP resonant interactions at different stages in figure 6. Here, the red colours denote energy transfer from EPs to the wave, and blue vice versa. We can see that the dominant resonant interactions are significantly different for RSAE and TAE. The RSAE is mostly driven unstable by the relatively lower energy EPs, whilst the TAE absorbs energy from the high energy end of the distribution. Thus, we may conclude that the velocity space anisotropy also plays an important part of the relay mechanism. That is, since the RSAE only resonates with the lower energy EPs, its saturation leaves the high energy portion of the EP distribution almost intact, and the TAE could then extract energy from the velocity space anisotropy. In fact, in many other attempts to reproduce the similar effect, we find that the relay of different dominant modes is much less effective if the RSAE and TAE resonances are closely placed in phase space.

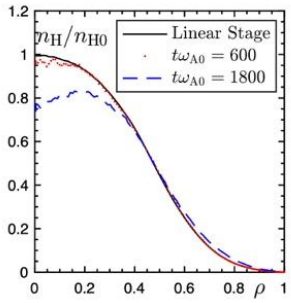


Fig.**5**: EP density profile in linear, initial saturation and late nonlinear stages.

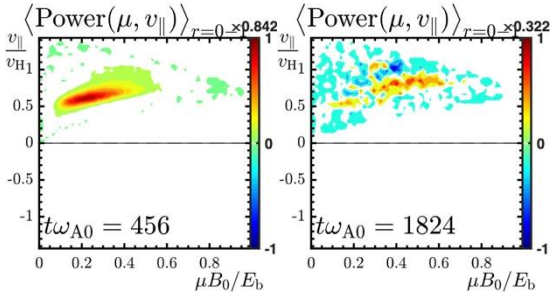


Fig.**6**: Comparison of the wave-EP resonance structures in phase space. The horizonal axis is the normalized magnetic moment, the vertical axis is the parallel velocity. The data is integrated in real space.

4 Summary

In this work, we use hybrid simulations to analyse the EP transport properties in reversed shear plasmas. The simulations suggest that a relay mechanism could take place, where the resonant EPs are firstly transported by core-localized RSAEs, and later by TAEs in the outer-core region. Clearly observing this mechanism requires that the resonances are sufficiently separated in phase space.

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [6].

References

[1] L. Chen and F. Zonca., Physics of Alfvén waves and energetic particles in burning plasmas. Rev. Mod. Phys. **88**, 015008 (2016).
[2] T. Wang *et al*., Shear Alfvén fluctuation spectrum in Divertor Tokamak Test facility plasmas. *Phys. Plasmas* **25**, 062509 (2018).
[3] T. Wang *et al*., Nonlinear dynamics of shear Alfvén fluctuations in Divertor Tokamak Test facility plasmas. *Phys. Plasmas* **26**, 012504 (2019).
[4] T. Wang *et al.*, Dynamics of reversed shear Alfvén eigenmode and energetic particles during current ramp-up. *Nucl. Fusion* **60**, 126032 (2020).
[5] S. Briguglio *et al*., Hybrid magnetohydrodynamic-gyrokinetic simulation of toroidal Alfvén modes. *Phys. Plasmas* **2**, 3711 (1995).
[6] F. Iannone *et al.*, CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, 2019, pp. 1051-1052, doi: 10.1109/HPCS48598.2019.9188135.

DETERMINISTIC ANALYSIS OF LOSS OF COOLANT ACCIDENT FOR DEMO DIVERTOR

Danilo Nicola Dongiovanni*, Tonio Pinna and Maria Teresa Porfiri

ENEA, Fusion Technical Unit, Nuclear Technologies Laboratory, Via Enrico Fermi 45, 00044 Frascati, Rome, Italy

ABSTRACT. DEMO Tokamak nuclear fusion reactor is currently under design and its compliancy with safety objectives shall be demonstrated since early design phases. Plant level safety performance is assessed by means of a systematic identification of representative accident events and their simulation by means of validated nuclear codes. Divertor is key in-vessel component in charge of exhausting the major part of the plasma ions thermal power in a region far from plasma core in order to control plasma pollution. A Loss of Coolant accident for the Divertor Plasma Facing Unit system was investigated by means of Melcor fusion adapted code to verify that accident consequences are bounded in terms of radionuclides confinement. The first containment pressurization and structures temperature evolution after accident initiating event occurrence were studied in the simulation.
Results sensitivity with respect to a set of parameters was also investigated exploiting CRESCO infrastructure to reduce computational time.

1 Introduction

The feasibility of nuclear fusion reactor energy exploitation as energy source relies on the actual possibility of achieving a safe and reliable reactor design. With this purpose DEMO Tokamak nuclear fusion reactor pre-conceptual design phase has just been concluded. As part of design activities, a set of analyses has been performed to assess its capability of safely operating under all circumstances.
Main systems in the DEMO plant [1] are analysed in the perspective of highlighting all related hazards (e.g. radionuclides) and possible failure mechanisms eventually resulting in such hazards to harm people or environment. For example, radionuclides inventories are confined within plant vacuum vessel (VV), but in-vessel actively cooled systems and associated coolant energies are potentially able to breach VV confinement barriers.
Divertor is an important in-vessel component in charge of exhausting the major part of the plasma ions thermal power in a region far from plasma core in order to control plasma pollution. It is actively cooled since subject to significant thermal and nuclear loads (order of 1 MW/m2). Main accidents associated to divertor system therefore originate from failures in heat transfer system as pointed out by Functional Failure Mode and Accident Analysis (FFMEA) methodology [2]. In particular the following accident have been identified as the most representative in terms of consequences

- In-vessel LOCA resulting from a large rupture of the divertor cassette/ Plasma Facing Units (PFU)
- Out-VV LOCA resulting from a large rupture in the divertor primary cooling loop in the water manifold feeder inside the PHTS Vault.
- Loss of flow (LOFA) in the water primary cooling loop of the divertor because of pump trip.
- Loss of heat sink in all FW, BZ and divertor primary cooling.

Here the simulation of an In-vessel LOCA event for the PFU divertor HTS is selected for a deterministic[1] simulation by means of MELCOR code. The input nodalization for simulation and main results are presented.

*Corresponding author. E-mail: danilo.dongiovanni@enea.it

[1] In nuclear safety analysis context, *probabilistic* safety analyses techniques (e.g. the mentioned FMEA) focus on the systematic identification of most likely and representative accident events to provide a comprehensive set of such events. Then the consequences of each of these accident events are deterministically assessed by means of validated codes able to cover and correctly simulate all relevant physical phenomena occurring during accident transients, so to provide accurate estimates

2 Materials and Methods

2.1 Divertor system

DEMO Divertor is a modular system. Its basic module is composed by a Divertor cassette and the related Plasma Facing Units (PFU). Divertor cassette (Figure.1-b) has the function of providing mechanical support for the PFUs and enabling an efficient pumping of neutral helium ashes and impurities out of plasma chamber. The PFUs (Figure.1-a) are implementing the main interfacing surface with plasma and related thermal and nuclear loads. Both PFU and cassette also implement an integrated shielding function of the underlying VV walls from plasma nuclear loads.
DEMO tokamak is assembled from 16 sectors, each including 3 cassette and PFU (Figure.1-c), for a total of 48 cassettes/PFUs around the VV. Every 8 sectors, Cassettes and PFU are actively cooled by means of two independent heat transfer loops (Figure.1-d). So a total of 4 loops provide heat sink functions for the divertor system as whole by means of a connection to DEMO Balance of Plant.

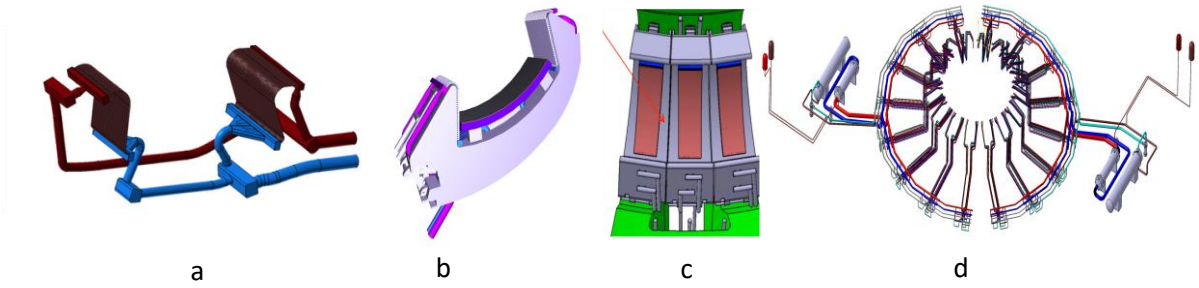


Fig. 1 Divertor cassette, PFU and related heat transfer systems.

2.2 System Design Analysis methodology

Methods for Estimation of Leakages and Consequences of Releases (MELCOR) is a fully integrated, engineering-level computer code initially developed to models the progression of severe accidents in light water reactor and more recently [4] adapted to model accident evolution in Fusion reactors. It is a system code for suitable for the analysis on integrated HTS containment as part of preliminary safety assessments in pre-conceptual designs. MELCOR can model phenomena relevant for fusion safety applications such as thermal-hydraulic response in the reactor coolant system with a variety of cooling fluids (helium, water, liquid metals) in the relevant pressure domain (including vacuum), containment and confinement buildings, hydrogen production, radionuclides transport (e.g., tritiated water) and combustion.
The nodalization shown in Figure.2 was adopted [5]. Ex-vessel divertor heat transfer system providing (5MPa/136°C) at PFU inlet includes a pressurizer (3.8 MPa), a heat exchanger providing 68 MW of heat sink power, hot leg, cold leg, pump supplying 1.82 MPa to the loop to compensate for pressure drops, hot ring, cold ring, 24 distributors and collectors serving each one PFU), 24 in-vessel PFU volumes resulting into a coolant inventory of about 57m^3. Vacuum Vessel (VV) volume (2898 m^3) and related upper and lower port volumes (1500 and 2000 m^3) are protected from pressurization by means of a pressure suppression system (VVPSS) connected by a rupture disk opening at 1.5bar. VV volume includes heat structures (breading blanket, divertor cassette, VV walls).
Parametric design cases were deterministically analyzed by means of multiple runs of Melcor code:
CASE A - Vacuum Vessel Walls at 40°C, with PFU HTS valves (FL417/420 in Fig.2)
CASE B - Vacuum Vessel Walls at 200°C, with PFU HTS valves (FL417/420)
CASE C - Vacuum Vessel Walls at 200°C, no valves.

concerning safety observables giving insight on accident consequences (i.e. containment pressure resulting from coolant vaporization within VV).

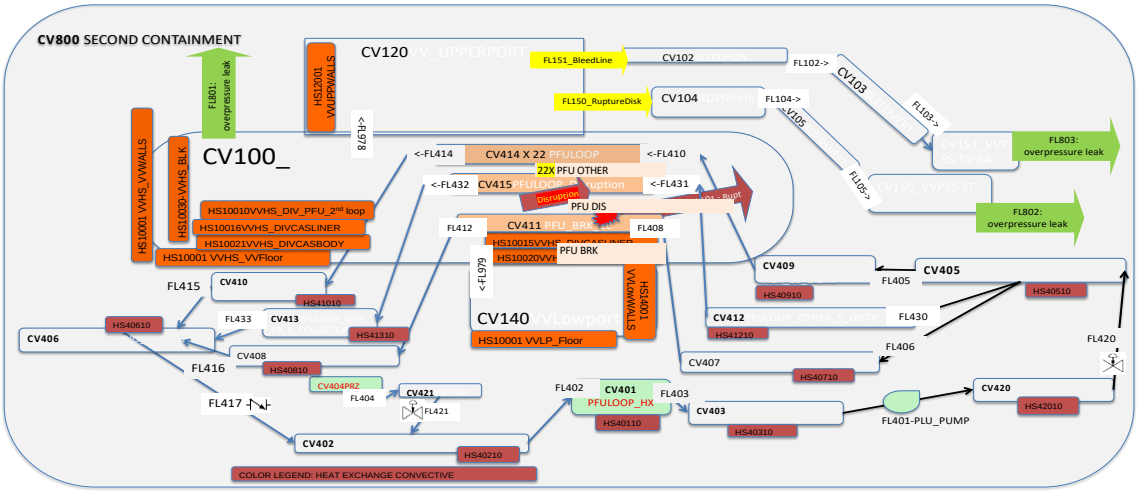


Fig. 2 Nodalization implemented for the simulation

3 Results

Of initial water PFU HTS inventory (58 m3 at 130°C/ 5MPa at PFU inlet), about 29 tons are released from Divertor HTS to VV in (Case A-B) while about 42 tons in (Case C) (Fig.3 left). Pressurization pattern is relatively slow (Fig.3 right). taking 2500s for VVPSS opening (in Case B/C) and depending on VV wall temperatures. No pressurization occurs in baseline case with VV wall at 40°C.

4 Conclusion

Divertor design foresees 4 independent loops (2 for PFU and 2 for cassettes) with direct distributor/collector serving each single cassette/PFU module. This choice results into a slow VV pressurization pattern in case of in-vessel LOCA because the low releasable coolant inventory and the low leak mass flow rate (low even in case the upper boundary is considered). In fact, it takes over 2500s to reach pressurization peak mainly as a result of vaporization of leaked water in contact with in-vessel components. Eventual pressurization peak is also effectively mitigated by safety devices.

Noticeably in the new baseline for VV wall temperature (40°C) promoting water condensation, no significant VV pressurization is observed. The isolation of HTS loop upstream rings has little impact on accident consequences, except protecting from under-pressurization effects on ex-vessel loop pipework. A cost-benefit assessment shall address the opportunity of such isolation valves studying the Loss of Flow Accident (possibly caused by spurious closure) [5].

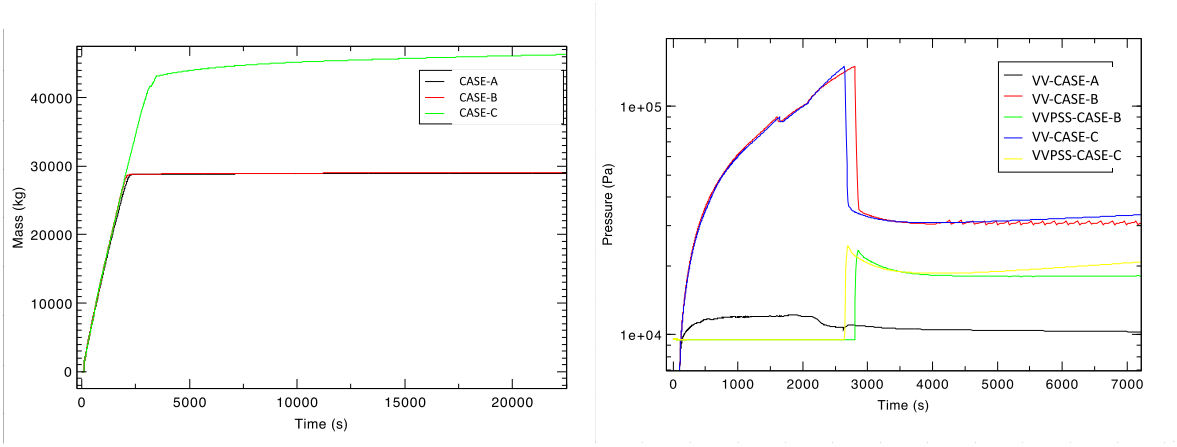


Figure 3 Left: Water Inventory released from FPU channels break (2.26E-3 m2) rupture to Vacuum Vessel (VV) for considered cases. Right: Pressure for Vacuum Vessel (VV) and related ports (total volume 6000m3) and VV Pressure Suppression System (VVPSS) for considered cases.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

[1] G. Federici *et al* 2017 European DEMO design strategy and consequences for materials, *Nucl. Fusion* 57 092002

[2] Tonio Pinna et al., Identification of accident sequences for the DEMO plant, Fusion Engineering and Design, Volume 124, 2017, Pages 1277-1280, ISSN 0920-3796, https://doi.org/10.1016/j.fusengdes.2017.02.026.

[3] G. Mazzone et al, Eurofusion-DEMO Divertor - Cassette Design and Integration, Fusion Engineering and Design, Volume 157, 2020, 111656, ISSN 0920-3796,https://doi.org/10.1016/j.fusengdes.2020.111656.

[4] Brad J. Merrill, et.al, "A recent version of MELCOR for fusion safety applications", Fusion Engineering and Design, 85 (2010), pp. 1479-1483

[5] D. Dongiovanni et al., DEMO Divertor preliminary safety assessment, Fusion Engineering and design, https://doi.org/10.1016/j.fusengdes.2021.112475, 2021.

APPLICATION OF INNOVATIVE MONTE CARLO RADIATION TRANSPORT ALGORITHMS TO VARIOUS PWR SAFETY PROBLEMS INCLUDING PRESSURE VESSEL AGEING AND ANALYSIS OF EXPERIMENTAL N_TOF FACILITY AT CERN

Patrizio Console Camprini[1*] and Kenneth W. Burn[2]

[1]*ENEA, Fusion and Technology for Nuclear Safety and Security Department, 40129, Bologna, Italy*

[2]*Former ENEA Researcher*

ABSTRACT. For a number of years, we have been involved in collaborations with the French Nuclear Safety Authority IRSN (Institut de Radioprotection et de Sûreté Nucléaire) on lengthening the lifetime of currently operating Pressurized Water Reactors (PWR's) and improving the safety of next-generation PWR's. As reported in previous CRESCO annual reports and documented in journal papers ([1,2] and citations therein), we have applied innovative Monte Carlo variance reduction (VR) algorithms, developed at ENEA, to calculating ex-core responses in PWR's (such as the pressure vessel (PV) damage, the signal in neutron detectors placed in the PV well and the activation of components outside the well. Our algorithms unify the two calculations of the "classical", decoupled approach – eigenvalue followed by fixed source with VR – in a single eigenvalue calculation, thus improving the accuracy by removing the approximations implicit in the decoupled approach. Our approach is based on the DSA (Direct Statistical Approach) [3]. There are two aspects of the DSA codes that are worth noting in the context of their execution on the machines of the CRESCO system. Within the collaborations with the INFN (Italian Institute for Nuclear Physics) Monte Carlo simulations are conducted to analyse the outcome of an experimental campaign performed at the n_TOF facility at CERN (European Center for Nuclear Research).

1 The handling of large variable arrays and integration at rendez-vous

The DSA generates large arrays to describe the 2nd moment function with the "enhanced point surface" approximation. This function is based on data accumulated at boundaries between cells in phase space (defined on real space and energy). These boundaries are not known a priori – they are only found during the Monte Carlo simulation. Furthermore, a binning is made at the boundaries that is a function both of the boundary phase space crossing point and what happens to the particle tracks having crossed the boundary. All this means that a fixed format cannot be used and pointers are employed to place the data in a hierarchical structure.

Thus, every time a boundary is encountered, a search must be made as to whether the particular boundary bin has been previously encountered or not (on this CPU). If it has not, space must be made in all the relevant arrays at the appropriate point to insert the new data. Clearly as the calculation proceeds and more data is found, more "accountancy" time is required.

Monte Carlo lends itself to multi-processing. In fixed source problems Monte Carlo is highly scalable with the number of processors – the master-slave and slave-master communication time is generally small compared with the tracking time on each CPU. (In eigenvalue problems that employ the source-iteration solution, scalability is not as good as in fixed source problems because after each fission generation the fission source must be reconstituted so as to be renormalized. A large fission source implies a high slave-master, master-slave communication time, although algorithms do now exist that tackle this problem.)

[*]Corresponding author. E-mail: patrizio.consolecamprini@enea.it

In eigenvalue calculations, the DSA employs superhistories that renormalize the fission source every n, n~10 fission generations, thus saving some inter-CPU communication time. However at rendez-vous points for the tally data where also the 2nd moment (and time) function data is called in from all the slaves, the same problem occurs as is met during the tracking on individual CPU's – how to merge the 2nd moment data.
Rather than sending the data from each slave in turn to the master, the data is firstly accumulated between slaves and then sent from the slaves to the master. It was intended thus to avoid too much "dead" time by having as many CPU's as possible working together. This works but it would be interesting to test it further.
The current version of the DSA uses as vehicle MCNP6 ver. 1.0 [4]. On the CRESCO6 system it is compiled with the default intel compiler (intel17) and with the default parallel processing option: openmpi_intel17-3.1.2. We have seen a marked improvement in the data passing performance compared with CRESCO4 and its analogous default options.

2 Minimizing the quality function using DUMING from IMSL

The quality function (product of the 2nd moment and time functions) needs to be minimized to give the optimum set of VR parameters. As the VR technique employed is splitting and Russian roulette and splitting may only be made in integer amounts, the 2nd moment function contains cusps. As a consequence, the quality function contains many local minima.
We employed the routine DUMING of the IMSL library (unconstrained minimization of a multi-variate function using a quasi-Newton method and a user-supplied gradient). Note as only positive VR parameters are acceptable, they are taken as the squares of the variables. Note also that the cusps are ignored in the gradient calculation.
We used the scalar version of DUMING, running on CRESCO4sm01/02 and CRESCO4fx002.frascati, latterly only on CRESCO4sm01. Although slow, presumably because of the nature of the quality function, we found the code to give excellent results for all problems so far treated. We have so far been unable to find a free software substitute that gives the same performance.

3 Some examples

We show a vertical (Fig. 1) and horizontal (Fig. 2) section of a typical PWR GEN II model with 14 positions indicating where the fast neutron flux (> 1 MeV and > 100 keV) that damages the PV is calculated. Fig. 3 shows the ratio of the results with an approximate approach, in this case assuming a homogeneous distribution of fission sites in the core and the default MCNP fission spectrum, to the results with the DSA single eigenvalue approach.
A vertical and horizontal section of a PWR GEN III model, with positions indicating where the various PV damage metrics are calculated, are shown in Figs. 4 and 5 respectively. Fig. 6 shows the ratio of the responses at position A4 in Fig. 5 with various decoupled approximations (all with the 235U thermal neutron-induced fission spectrum), to the single eigenvalue approach with the DSA.

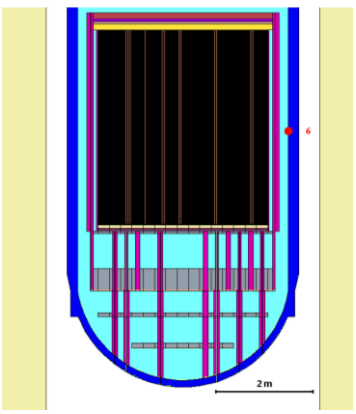


Fig.1: PWR GEN II: Vertical section

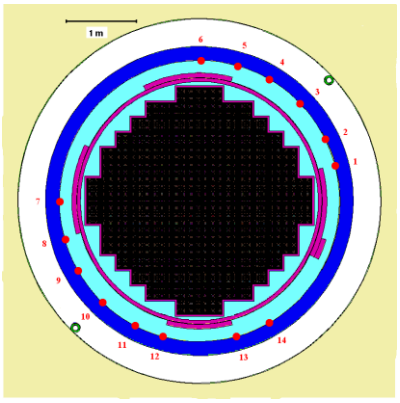


Fig.2: PWR GEN II: Horizontal core midplane

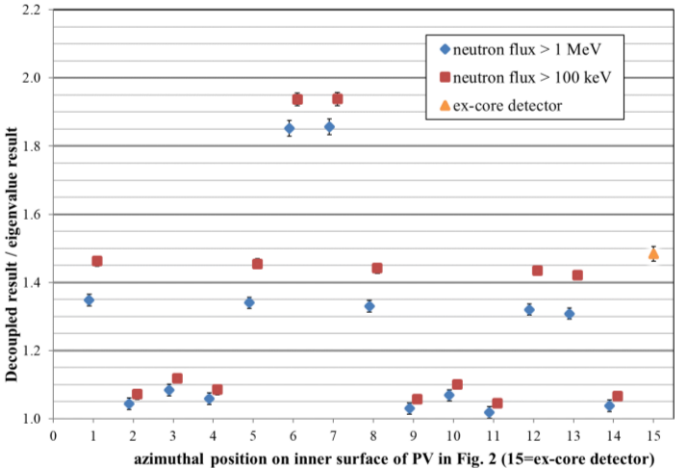


Fig.3: PWR GEN II: Ratio of decoupled to eigenvalue result for the MCNP default Watt fission spectrum and a homogeneous spatial distribution of fission sites

4 Simulation of Experiment for the n_TOF Facility at CERN

Within the framework of nuclear physics and specifically basic research, n_TOF facility at CERN (European Center for Nuclear Research) is an important facility utilized to study nuclear reaction cross sections at different energies, thanks to a particularly powerful neutron beam extracted from the large accelerator current, through spallation reactions on a dedicated target.

The present study has been carried out in collaboration with Italian National Institute for Nuclear Physics (INFN), with the Bologna Section.

Monte Carlo simulations have been oriented to a neutron beam impinging on a uranium target since ^{235}U(n, f) cross section is one of the most important cross section standards at thermal neutron energy between 0.15 MeV and 200 MeV, for measurements of neutron-induced reaction cross-sections [5]. It is used in a wide range of applications. In addition, as in most other neutron facilities, the ^{235}U(n,f) cross section is used as a reference for fission cross section measurements.

Despite its widespread use, however, the recommended ^{235}U(n,f) cross section data at energies above 20 MeV are based on a small set of measurements [6,7]. Hence, there is a clear and long-standing demand from the International Atomic Energy Agency (IAEA) to improve this situation. At energies above 200 MeV, the ^{235}U(n,f) cross section plays an important role for several applications, as well as for fundamental nuclear physics.

Despite the importance of the high-energy region, at present no data exist on neutron induced fission above 200 MeV, and it is necessary to rely on theoretical estimates.

However, new theoretical calculations [8] have indicated that the ^{235}U(n,f) cross section at high energy may be substantially different from what was previously thought. Furthermore, in a new evaluation [9], recently released by the International Atomic Energy Agency (IAEA) Nuclear Data Section, it is stated that new absolute measurements of the neutron induced fission cross sections (e. g. relative to n-p scattering) on uranium, bismuth, lead and plutonium have the highest priority in establishing neutron induced fission reaction standards above 200 MeV.

The n_TOF facility is then utilized for the measurement of neutron-induced reaction cross-section as a function of the energy.

At this stage of the analysis of the experiment, several simulations have been carried out in order to compare MCNP6.2 [10] and GEANT4 [11] Monte Carlo codes in order to find the best approach to simulate neutron-proton scattering reaction (n,p) taking place in many kind of inspected hydrogenated targets.

Neutron beam provided by n_TOF facility is utilised within the irradiation experiment and a good characterization of the proton recoil production is requested.

First, all experimental uncertainties related to geometrical setup have been verified and, ultimately, the point has been focused of nuclear data for scattering n-p reaction, in particular the Arndt cross-section implementation.

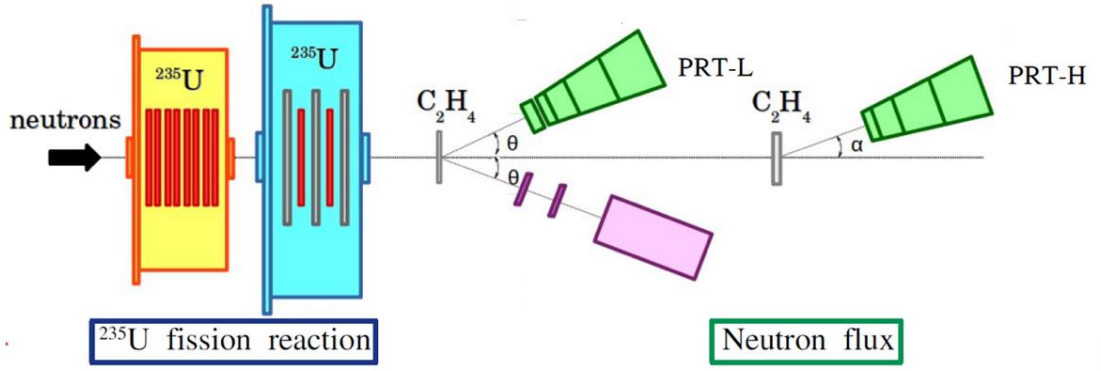


Fig.4: n_TOF experimental setup: neutron telescopes (PRT-L, H) and C_2H_4 intermediate targets.

Acknowledgement

The work on Gen-II PWR's was carried out in close collaboration with l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN). IRSN supplied the MCNP models and the requested responses: the authors would like to thank Mariya Brovchenko.

The author thanks the INFN (National Institute for Nuclear Physics) – Bologna Section – for the collaboration concerning n_TOF facility and all the information to prepare the calculations.

Simulations were performed through the CRESCO supercomputing facility since large HPC techniques and massive computing environment were necessary to achieve these calculations.

References

[1] M. Brovchenko, K.W. Burn, P. Console Camprini "On Integrating Monte-Carlo Calculations in and around Near-Critical Configurations – I. Methodology," to be submitted to *Ann. Nucl. Energy* (2021)

[2] M. Brovchenko, K.W. Burn and P. Console Camprini, "On Integrating Monte Carlo Calculations in and around Near-Critical Configurations – II. Pressurized Water Reactors," to be submitted to *Ann. Nucl. Energy* (2021)

[3] K.W. Burn and P. Console Camprini, "A Consistent Monte Carlo Treatment of Radiation Responses in and around Critical Configurations," *Proceedings of PHYSOR-2018*, Cancun, April 22 – 26, pp. 576–586 (2018)

[4] Denise B. Pelowitz (Ed.), "MCNP6™ User's Manual – Version 1.0", Los Alamos National Laboratory, LA-CP-13-00634, Rev. 0 (2013)

[5] A.D. Carlson, et al., Nuclear Data Sheets 110 (2009) 3215

[6] A.D. Carlson, O.A. Wasson, P.W. Lisowski, J.L. Ullmann, N.W. Hill, in Proc. of the Int. Conf. on Nuclear Data for Science and Technology, Juelich 1991, p.518 (1991), Germany

[7] R. Nolte, et al., Nuclear Science and Engineering, 156 (2007) 197

[8] S. Lo Meo, D. Mancusi, C. Massimi, G. Vannini, A. Ventura, , Nucl. Phys. A 933 (2015)

[9] INDC International Nuclear Data Committee, "209Bi(n, f) and natPb(n, f) Cross Sections as a New Reference and Extension of the 235U, 238U and 239Pu(n, f) Standards up to 1 GeV", INDC(NDS)-0681 Distr. ST/J/G/NM , IAEA 2015

[10] C.J. Werner, "MCNP User's Manual - Code Version 6.2", LA-UR-17-29981 (2017)

[11] GEANT4 Monte Carlo code: website reference https://geant4.web.cern.ch/node/1

FIGHTING SARS-COV-2 USING NATURAL COMPOUNDS: A VIRTUAL SCREENING ANALYSIS

Alice Romeo, Federico Iacovelli and Mattia Falconi*

Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy

ABSTRACT. The SARS-CoV-2 Spike glycoprotein is a promising target for early blocking virus infection, given its fundamental role in mediating cell recognition and virus fusion. Last year, we proposed a highly conserved internal cavity of the Spike as a novel therapeutic target for fusion inhibitors, and we performed a virtual screening using a set of FDA drugs to propose potential antiviral molecules. Considering the still limited access of developing countries to essential medicines and vaccines, we performed another virtual screening targeting the Spike cavity, evaluating a set of 77.500 natural compounds to suggest novel safe and low-cost promising antiviral molecules. The most promising compounds will be subjected to experimental tests.

1 Introduction

The Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first isolated in Wuhan (China) in December 2019, is the single-stranded RNA Betacoronavirus responsible for the ongoing COVID-19 pandemic [1]. The SARS-CoV-2 virion is composed by an outer membrane envelope, mainly coated with Spike glycoproteins [2]. The Spike N-terminal domain (NTD) is fundamental for virus entry and is divided into two functional subunits by protease cleavage. The S1 subunit (residues 1-685) is involved in receptor recognition through its receptor-binding domain (RBD) [1], while the S2 subunit (residues 686-1273) mediates viral fusion and it is composed by two trimeric α-helical regions called heptad repeats 1 (HR1) and 2 (HR2) [3]. After cell receptor binding, Spike activation by proteases cleavage triggers huge conformational rearrangements in the HRs regions, leading to the shedding of the S1 subunit and to the formation of a six-helical bundle (6-HB) structure [2]. This event marks the transition of the protein from a prefusion to postfusion conformation and allows upstream fusion peptides to insert within the cell membrane, providing the energy requirements needed to drive viral and cell membranes fusion and to allow virus entry [4,5]. Once entered, the virus hijacks the cellular machineries to ensure its own replication and propagation [2].

Targeting the Spike S2 subunit has attracted a great interest in therapeutic drug discovery since this region and its interaction modes are highly conserved among human coronaviruses [2]. Last year, we proposed a novel possible inhibition strategy involving the targeting of an internal cavity present only in the prefusion conformation of the Spike [6]. The idea was based on previous research carried out for the human respiratory syncytial virus (hRSV) F protein, which indicated that newly generated escape mutations for small-molecule fusion inhibitors were clustered inside the central cavity of the F protein in prefusion conformation [7–9]. Crystallographic studies confirmed that these inhibitors bind within the central cavity, and functional studies further demonstrated that these small-molecules act as antagonists that prevent the protein transition to the postfusion conformation [10,11]. To date, known small-molecule fusion inhibitors bind in the same pocket and have the same mechanism of action [12,13]. These data lead us to hypothesize that, also in the case of SARS-CoV-2, targeting the internal cavity of the Spike glycoprotein in prefusion conformation could be a valuable approach for hindering its prefusion to postfusion transition and consequently the virus entry into the host cells. Indeed, sequence alignments of different SARS-CoV-2 isolates, deposited in the GenBank database [14], showed that this internal pocket is conserved among different SARS-CoV-2 isolates and that it also shares a high sequence similarity with the SARS-CoV Spike, thus representing a possible target for a broad-spectrum treatment [6]. To characterize this novel target site, we performed a virtual screening (VS) using a set of 8755 FDA drugs, which allowed us to identify small compounds, in particular

*Corresponding author. E-mail: falconi@uniroma2.it

phthalocyanine and hypericin, which could adopt very stable binding poses inside the inner cavity of the prefusion Spike, potentially anchoring surrounding internal regions by establishing a thick network of interactions [6].
Given these promising results [6], we exploited again the internal cavity of the Spike as a VS target, this time using a larger set of 77.500 natural compounds, selected after a filtering procedure performed on a set of 224.205 molecules extracted from the ZINC15 database (https://zinc15.docking.org/) [15]. Indeed, natural products are commonly used in traditional medicine preparations and gathered considerable interest in recent years, particularly during the COVID-19 pandemic, as ecological and non-toxic drug alternatives [16]. Many plant-derived products are known for their intrinsic antiviral properties, their production costs are generally lower, and they are often widely available in developing countries, where access to vaccines or other expensive antiviral treatments is still very limited. VS results allowed us to proposed other possible fusion inhibitors compounds, that should be experimentally evaluated to assess their efficacy. As receptor to be used for molecular docking simulations, we extracted a representative structure of the Spike from a 10 μs classical molecular dynamics simulation released by the D.E. Shaw research team for public use (D. E. Shaw Research, "Molecular Dynamics Simulations Related to SARS-CoV-2," D.E. Shaw Research Technical Data, 2020; http://www.deshawresearch.com/resources_sarscov2.html). Simulation of the protein for this long timescale allowed to relax the cryo-EM structure and to better describe the protein structural dynamics, improving our subsequent evaluation of ligands binding.

2 Simulation Methods

2.1 Molecular modeling

Spike glycoprotein structure has been obtained from a cluster analysis performed on a 10 μs classical molecular dynamics simulation of the Spike glycoprotein (PDBID: 6VXX) [17], performed by the D.E. Shaw research team and released for public use (D. E. Shaw Research, "Molecular Dynamics Simulations Related to SARS-CoV-2," D.E. Shaw Research Technical Data, 2020) (http://www.deshawresearch.com/resources_sarscov2.html). Cluster analysis has been performed using the cluster module of the GROMACS 2019 software [18] and the *gromos* algorithm [19], setting a RMSD cut-off of 0.3 nm. The centroid of the first cluster, which represents about 70% of the trajectory, was chosen as receptor for the molecular docking simulations. Non-terminal small missing loops in the trimer structure have been modelled using the SWISS-MODEL web-server [20]. Modelling involved residues 70-79, 144-164, 173-185, 246-262, 445-446, 455-461, 469-488, 502, 621-640, 677-688, 828-853 of monomer A, B and C.

2.2 VS compounds database preparation

A library of 224.205 compounds structures was retrieved from the “natural products” subset of the ZINC 15 database (https://zinc15.docking.org/) [15]. Structures were filtered using a fingerprint-based similarity search implemented in the software Open Babel 2.4.1 [21], based on the calculation of the Tanimoto coefficient between each pair of structures. This coefficient is an index of similarity which ranges between 0 and 1, where 0 means that the structures are completely dissimilar and 1 that they are identical. When the similarity evaluated between two structures was higher than 0.95, the structure with the higher number of atoms was retained, while the other was removed, to discard all highly similar structures and substructures present in the library. Very small compounds were also removed, setting the threshold to at least 20 heavy atoms. Finally, 19 structures, containing Silicon (Si) atoms, were excluded due to the absence of AutoDock Vina [22] docking parameters for this atom type. The final library used to perform the VS procedure contained 77.500 molecules.

2.3 Virtual screening procedure

In order to obtain the input files for the VS procedure, Spike glycoprotein and compounds structures have been converted into the pdbqt format using the *prepare_ligand4.py* and the *prepare_receptor4.py* tools of AutoDockTools4 software [23]. The 77.500 molecular docking simulations have been performed through an in-house parallelized version of the Autodock Vina 1.1.2 program [38], implemented on the ENEA CRESCO6 HPC cluster [24] using the *mpi4py* Python3 library. One docking simulation, including ten docking runs, has been carried out for each compound, obtaining as output 10 binding poses representing the cluster centroids of all the different conformations generated. A box of

x=21.00, y=22.50, z=27.38 Å has been placed over the HR1 internal region (residues 897–920) of the Spike glycoprotein A monomer, and 14 receptor residues side-chains around the selected binding site have been considered as flexible, that is they were allowed to move during the procedure (N907, I909, V911, T912, Q913, N914, K1038, E1092, Q1106, R1107 and N1108 of the Spike monomer A and R1091, E1092 and F1121 of the Spike monomer C).

2.4 Filtering of the results set

The VS results set has been initially reduced by selecting only those molecules showing an interaction energy ≤ -10.0 kcal/mol with the Spike cavity, reducing the number of molecules to 7.734 (about 10% of the initial database). Then, subsequent filtering has been performed mainly based on ADMET parameters calculated for each compound.

Druglikeliness, physicochemical and pharmacokinetics properties of the molecules were predicted using the SwissADME web-server [25], to identify compounds predicted to be more suitable for a future experimental and clinical use. Key factors considered were gastrointestinal absorption rate, solubility, compliance to at least one druglikeliness rule-based filter, predicted bioavailability, absence of Pan Assay Interference Structures (PAINS) and Brenks fragments, and compliance to at least two of the three leadlikeliness rules. Compounds violating one or more of the ADME parameters evaluated have been discarded, reducing the number of molecules to 786.

Predicted interactions of the compounds with the P-glycoprotein efflux pump or with different cytochrome P450 isoforms were also considered, since the absence of interaction generally identifies more promising drug candidates. For each of these parameters, a positive (+1) or null (0) score was assigned to each compound according to the outcome of the prediction (binder or non-binder). An intermediate score, referred to as ADME score, was obtained by summing the partial scores of the six predictions.

General toxicity of compounds was predicted using the eToxPred approach [26] which employs an Extremely Randomized Tree (ET) classifier. All compounds in our results set showed a Tox-score < 0.58 which, according to the paper, discriminates between toxic and non-toxic molecules, identifying all our molecules as non-toxic. Compounds mutagenic potential, their potential hepatotoxic activity and potential inhibition of the hERG I and II K+ channels have been assessed using the pkCSM webserver [27]. Compounds having at least two out of three positive results have been discarded, reducing the set to 157 molecules. After applying the described filters, a score called ADMET score was computed by adding the ADME score to the probability of the compound to be non-toxic (evaluated as 1-Tox-score), normalizing the results.

Interaction energy of the 157 compounds passing all ADMET filters has been re-evaluated as an average of three repeated molecular docking simulations, using the same protocol described for the initial VS, ending up with a final set of 100 compounds showing an average energy ≤ -10.0 kcal/mol. The resulting average energy for each compound has been normalized and added to the ADMET score, obtaining a final score which has been used to sort the ranking, since it balances compounds affinity for the Spike with their pharmacological and toxicological properties.

3 Results

Results of the VS performed for 77.500 natural compounds within the Spike internal cavity have been filtered and ranked as described in the methods. For only 22 of the final 100 compounds passing all filters we could retrieve some information from literature, concerning their natural source, their common use, or any involvement in experimental studies. Top 10 molecules obtained are reported in Table 1.

Table 1: Top 10 molecules from the final filtered set of natural compounds.

PubChem ID	Name	Average binding energy (kcal/mol)	ADMET score	Final score
44559646	Subtrifloralactone K	-11.8	1.33	2.33
51136434	9(11)-dehydromanogenin	-11.4	1.38	216
11725801	Sauchinone	-11.2	1.38	2.05
188935	Mer-NF8054X	-11.0	1.41	1.97
442941	Argentine	-10.7	1.48	1.87
52931440	Neoallilogenin	-10.9	1.36	1.86
11385250	Asparacosin A	-10.7	1.46	1.85
9867531	Paspalinine	-10.7	1.36	1.75
76966159	Tryptoquivaline G	-10.1	1.67	1.73
102013420	Pierisformoside C	-10.4	1.38	1.60

Evaluating literature information, we excluded the compounds paspalinine and tryptoquivaline G for their potentially dangerous tremorgenic activities [28,29]. The withanolide subtrifloralactone K, the saponin 9(11)-dehydromanogenin and the lignan sauchinone are the compounds showing the highest final score and, also, the highest average interaction energy with the Spike cavity. Although from literature we retrieved no indication of a specific antiviral action exerted by these three molecules, several compounds belonging to their same chemical classes are known to have different antiviral or antimicrobial properties. For example, different saponins showed antiviral activity against enveloped viruses like the HCV and HIV, involving also the inhibition of virus entry into cells [30,31]. In particular, the glycosidic saponin glycyrrhizin was used in the treatment of SARS and was shown to neutralize the SARS-CoV-2 main protease (Mpro) *in vitro* [32,33]. Moreover, sauchinone possesses anti-inflammatory and antioxidant properties and in particular has been observed to upregulated the HO-1 cytoprotective pathway, which has been proposed as a possible therapeutic target to control the COVID-19 cytokine storm [34,35]. An *in silico* study also suggested the binding of sauchinone to the SARS-CoV-2 Mpro [36], proposing this molecule as a possible multi-target antiviral agent. Furthermore, extracts of *S. chiniensis*, the plant from which the compound derives, have shown antiviral effects on different viruses including the HIV-1 [37]. Oral administration of this molecules in mice yielded no side effects. Interestingly, the spirosteroid compound asparacosin A, the 7th ranked molecule, possessed an anti-inflammatory activity similar to sauchinone [38], and extracts of the plant *A. racemosus*, from which this compound derives, have been evaluated in a randomized clinical trial to treat COVID-19 patients [39].

Best binding poses obtained for these four promising compounds are shown in Fig. 1. The compounds bind to different regions of the pocket: subtrifloralactone K and 9(11)-dehydromanogenin bind to a region between the internal lobes of Spike monomers A and B, sauchinone locates below the internal lobe of monomer A, similarly to what we observed for hypericin and other FDA compounds in the previous VS [6], and asparacosin A binds almost at the center of the pocket, simultaneously contacting the internal lobes of the three monomers and also upper helical and loop regions of the pocket.

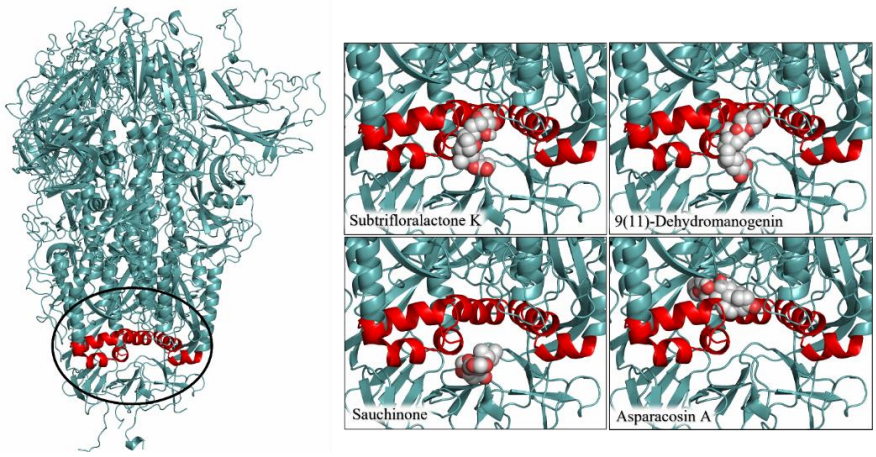


Fig.1: Structure of the Spike glycoprotein with the internal cavity, the site selected for the VS, highlighted in red. The region surrounded by the black circle is zoomed in the images on the right, showing the best binding poses obtained for the four most promising compounds, colored in white by atom type.

Interactions established by the compounds with surrounding regions of the pocket have been evaluated using the Ligand Interaction Diagram tool of the Schrödinger Maestro program (Schrödinger Release 2020-1: Maestro, Schrödinger, LLC, New York, NY, 2021). 2D ligand-interactions plots (Fig. 2) show that subtrifloralactone K establishes a total of 21 contacts within the pocket, mainly with residue from monomer A but also with 5 and 3 residues from monomer B and C, respectively, including one hydrogen bond with Lys1038 of monomer A. Similarly, 9(11)-dehydromanogenin interacts with 20 residues, including 13 residues from monomer A, 3 residues from monomer B and 4 residues from monomer C, also establishing 2 hydrogen bonds with residues Glu1092 of monomer A and Arg1091 of monomer C. Sauchinone and asparacosin A interact with 17 and 19 residues, respectively, contacting almost equally the three Spike monomers. Sauchinone can establish three hydrogen bonds with residues Gly1093 of monomer A, Asp1118 of monomer B and Arg1091 of monomer C, while asparacosin A establishes two hydrogen bonds with Gln1036 and Trp886 of monomer A (Fig. 2).
The obtained results open the path for a future evaluation, both through *in silico* and *in vitro* approaches, of the antiviral potential of these natural compounds, as novel SARS-CoV-2 fusion inhibitory molecules potentially targeting the internal cavity of the Spike glycoprotein.

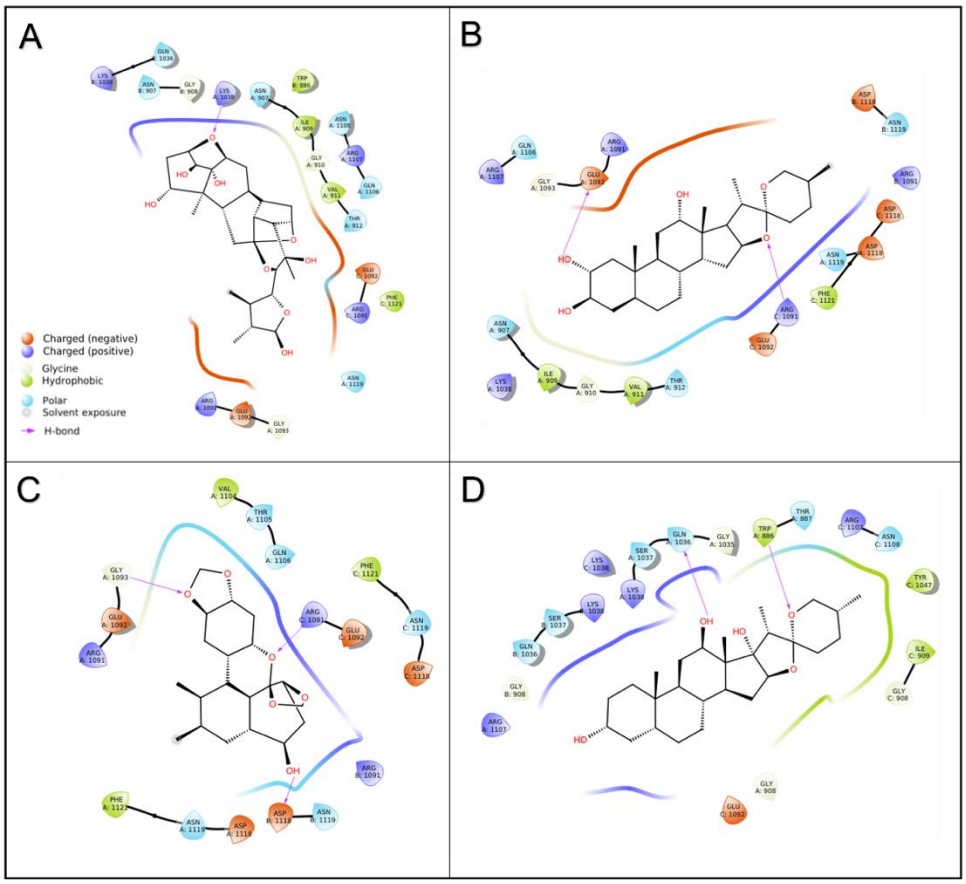


Fig.2: 2D representations of the interactions established within the Spike internal cavity by A) Subtrifloralactone K; B) 9(11)-dehydromanogenin; C) Sauchinone; D) Asparacosin A. Spike residues surrounding the ligands are colored according to their properties, as reported in the legend. Shape of the binding pocket is indicated by the line drawn around the ligands, colored as the nearest residue. This picture has been generated using Ligand Interaction Diagram Tool of the Schrödinger Maestro program (Schrödinger Release 2020-1: Maestro, Schrödinger, LLC, New York, NY, 2021).

Acknowledgment

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [24]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

References

[1] P. Zhou, X. Lou Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C.L. Huang, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. **579**, pp. 270–273, (2020).

[2] T. Tang, M. Bidon, J.A. Jaimes, G.R. Whittaker, S. Daniel. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res*. **178**, pp. 104792, (2020).

[3] B.J. Bosch, R. van der Zee, C.A.M. de Haan, P.J.M. Rottier. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol*. **77**, pp. 8801–11, (2003).

[4] A.C. Walls, M.A. Tortorici, J. Snijder, X. Xiong, B.J. Bosch, F.A. Rey, D. Veesler. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **114**, pp. 11157–11162, (2017).

[5] Y. Cai, J. Zhang, T. Xiao, H. Peng, S.M. Sterling, R.M. Walsh, S. Rawson, S. Rits-Volloch, B. Chen. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, pp. 1586-1592, (2020).
[6] A. Romeo, F. Iacovelli, M. Falconi. Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors. *Virus Res.* **286,** p. 198068, (2020).
[7] M.B. Battles, J.S. McLellan. Respiratory syncytial virus entry and how to block it. *Nat. Rev. Microbiol.* **17**, pp. 233–245, (2019).
[8] B.R.P. Lopes, M.F. da Costa, A. de Genova Ribeiro, C.S. Lima, I.P. Caruso, G.C. de Araújo, L.H. Kubo, F. Iacovelli, M. Falconi, A. Desideri, et al. Quercetin pentaacetate inhibits in vitro human respiratory syncytial virus adhesion. *Virus Res.* **276**, pp. 197805, (2020).
[9] D. Yan, S. Lee, V.D. Thakkar, M. Luo, M.L. Moore, R.K. Plemper. Cross-resistance mechanism of respiratory syncytial virus against structurally diverse entry inhibitors. *Proc. Natl. Acad. Sci.* **111**, pp. E3441–E3449, (2014).
[10] M.B. Battles, J.P. Langedijk, P. Furmanova-Hollenstein, S. Chaiwatpongsakorn, H.M. Costello, L. Kwanten, L. Vranckx, P. Vink, S. Jaensch, T.H.M. Jonckers et al. Molecular mechanism of respiratory syncytial virus fusion inhibitors. *Nat. Chem. Biol.* **12**, pp. 87–93, (2016).
[11] D. Samuel, W. Xing, A. Niedziela-Majka, J.S. Wong, M. Hung, K.M. Brendza, M. Perron, R. Jordan, D. Sperandio, X. Liu, R. Mackman, R. Sakowicz. GS-5806 Inhibits Pre- to Postfusion Conformational Changes of the Respiratory Syncytial Virus Fusion Protein. *Antimicrob. Agents Chemother.* **59**, pp. 7109–7112, (2015).
[12] J.P. DeVincenzo, R.J. Whitley, R.L. Mackman, C. Scaglioni-Weinlich, L. Harrison, E. Farrell, S. McBride, R. Lambkin-Williams, R. Jordan, Y. Xin, et al. Oral GS-5806 Activity in a Respiratory Syncytial Virus Challenge Study. *N. Engl. J. Med.* **371**, pp. 711–722, (2014).
[13] M. Stevens, S. Rusch, J. DeVincenzo, Y.-I. Kim, L. Harrison, E.A. Meals, A. Boyers, J. Fok-Seang, D. Huntjens, N. Lounis, et al. Antiviral Activity of Oral JNJ-53718678 in Healthy Adult Volunteers Challenged With Respiratory Syncytial Virus: A Placebo-Controlled Study. *J. Infect. Dis.* **218**, pp. 748–756, (2018).
[14] K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers. GenBank. *Nucleic Acids Res.* **44**, pp. D67–D72, (2016).
[15] T. Sterling, J.J. Irwin. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model*. **55**, pp. 2324–2337, (2015).
[16] J. Zhou, J. Huang. Current Findings Regarding Natural Components With Potential Anti-2019-nCoV Activity. *Front. Cell Dev. Biol*. **8**, pp. 1–13, (2020).
[17] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. **181**, pp. 281-292.e6, (2020).
[18] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*., **1–2**, pp. 19–25, (2015).
[19] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark. Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **38**, pp. 236-240, (1999).
[20] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, pp. W296–W303, (2018).
[21] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison. Open Babel: An open chemical toolbox. *J. Cheminform*. **3**, p. 33, (2011).
[22] O. Trott, A.J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, pp. 455-461, (2010).
[23] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30,** pp. 2785–2791, (2009).
[24] F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, F. Palombi, G. Ponti, G. Santomauro, P. Procacci. CRESCO ENEA HPC clusters: A working example of a multifabric GPFS Spectrum Scale layout. *2019 Int. Conf. High Perform. Comput. Simulation, HPCS 2019*. pp. 1051–1052, (2019).
[25] A. Daina, O. Michielin, V. Zoete. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **7**, pp. 42717, (2017).

[26] L. Pu, M. Naderi, T. Liu, H.-C. Wu, S. Mukhopadhyay, M. Brylinski. eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol. Toxicol*. **20**, p. 2, (2019).
[27] D.E. V. Pires, T.L. Blundell, D.B. Ascher. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem*. **58**, pp. 4066–4072, (2015).
[28] D.B. Gant, R.J. Cole, J.J. Valdes, M.E. Eldefrawi, A.T. Eldefrawi. Action of tremorgenic mycotoxins on GABAA receptor. *Life Sci*. **41**, 2207–2214, (1987).
[29] S. Dhingra, A.L. Lind, H.-C. Lin, Y. Tang, A. Rokas, A.M. Calvo. The Fumagillin Gene Cluster, an Example of Hundreds of Genes under veA Control in Aspergillus fumigatus. *PLoS One*. **8**, p. e77147, (2013).
[30] Y.J. Wang, K.L. Pan, T.C. Hsieh, T.Y. Chang, W.H. Lin, J.T.A. Hsu. Diosgenin, a Plant-Derived Sapogenin, Exhibits Antiviral Activity in Vitro against Hepatitis C Virus. *J. Nat. Prod*. **74**, pp. 580–584, (2011).
[31] C. Wang, L. Lu, H. Na, X. Li, Q. Wang, X. Jiang, X. Xu, F. Yu, T. Zhang, J. Li, Z. Zhang, B. Zheng, G. Liang, L. Cai, S. Jiang, K. Liu. Conjugation of a Nonspecific Antiviral Sapogenin with a Specific HIV Fusion Inhibitor: A Promising Strategy for Discovering New Antiviral Therapeutics. *J. Med. Chem.* **57**, pp. 7342–7354, (2014).
[32] G. Hoever, L. Baltina, M. Michaelis, R. Kondratenko, L. Baltina, G.A. Tolstikov, H.W. Doerr, J. Cinatl. Antiviral Activity of Glycyrrhizic Acid Derivatives against SARS−Coronavirus. *J. Med. Chem.* **48**, pp. 1256–1259, (2005).
[33] L. van de Sand, M. Bormann, M. Alt, L. Schipper, C.S. Heilingloh, E. Steinmann, D. Todt, U. Dittmer, C. Elsner, O. Witzke, A. Krawczyk. Glycyrrhizin Effectively Inhibits SARS-CoV-2 Replication by Inhibiting the Viral Main Protease. *Viruses.* **13**, p. 609, (2021).
[34] B. Li, D.S. Lee, H.G. Choi, K.S. Kim, D.G. Kang, H.S. Lee, G.S. Jeong, Y.C. Kim. Sauchinone Suppresses Pro-inflammatory Mediators by Inducing Heme Oxygenase-1 in RAW264.7 Macrophages. *Biol. Pharm. Bull.* **34,** pp. 1566–1571, (2011).
[35] M. Rossi, M. Piagnerelli, A. Van Meerhaeghe, K. Zouaoui Boudjeltia. Heme oxygenase-1 (HO-1) cytoprotective pathway: A potential treatment strategy against coronavirus disease 2019 (COVID-19)-induced cytokine storm syndrome. *Med. Hypotheses*. **144**, p. 110242, (2020).
[36] T. Joshi, T. Joshi, H. Pundir, P. Sharma, S. Mathpal, S. Chandra. Predictive modeling by deep learning, virtual screening and molecular dynamics study of natural compounds against SARS-CoV-2 main protease. *J. Biomol. Struct. Dyn.* **39**, pp. 6728–6746, (2021).
[37] J. Lee, M.S. Huh, Y.C. Kim, M. Hattori, T. Otake. Lignan, sesquilignans and dilignans, novel HIV-1 protease and cytopathic effect inhibitors purified from the rhizomes of Saururus chinensis. *Antiviral Res.* **85**, pp. 425-428, (2010).
[38] N. Karim, I. Khan, W. Khan, I. Khan, A. Khan, S.A. Halim, H. Khan, J. Hussain, A. Al-Harrasi. Anti-nociceptive and Anti-inflammatory Activities of Asparacosin A Involve Selective Cyclooxygenase 2 and Inflammatory Cytokines Inhibition: An in-vitro, in-vivo, and in-silico Approach. *Front. Immunol*. **10**, p. 581, (2019).
[39] H. Rangnekar, S. Patankar, K. Suryawanshi, P. Soni. Safety and efficacy of herbal extracts to restore respiratory health and improve innate immunity in COVID-19 positive patients with mild to moderate severity: A structured summary of a study protocol for a randomised controlled trial. *Trials.* **21**, p. 943, (2020).

EFFECTS OF MG DOPING ON THE ELECTRONIC PROPERTIES OF SNO2 AND ITS APPLICATION IN PEROVSKITE SOLAR CELLS AN AB-INITIO DFT STUDY

Gennaro Vincenzo Sannino[1], Adriana Pecoraro[2], Ana B. Muñoz García[2] and Michele Pavone[1*]

[1]*University of Naples "Federico II", Department of Chemical Sciences, Comp. Univ. Monte Sant'Angelo Via Cintia 21, 80126, Naples, Italy*

[2]*University of Naples "Federico II", Department of Physics "Ettore Pancini", Comp. Univ. Monte Sant'Angelo Via Cintia 21, 80126, Naples, Italy*

ABSTRACT. We provide a first-principles investigation on the structural and electronic effects of Mg doping in SnO_2. The substitution of Sn ions with Mg causes a reduction in the work function (WF) and, consequently, the conduction band minimum (CBM) of the material rises in energy. On the contrary, the insertion of Mg ions in interstitial sites increases the WF reversing the trend. Our results can explain experimental data extracted by perovskite solar cells (PSCs) involving Mg-doped SnO_2 as electron transport layer (ETL). In particular, an enhancement in open circuit potential (V_{OC}) is expected until Mg ions substitute Sn atoms and do not occupy interstitial positions.

1 Introduction

Perovskite solar cells (PSCs) are low cost devices able to convert sunlight into electricity with high efficiencies. SnO_2 is a promising candidate as electron transport layer (ETL), that is the component of the cell which has the task of extracting excited electrons from the conduction band (CB) of the perovskite film. The research carried out at the University of Naples "Federico II" focuses on the theoretical investigation of Mg doping as effective strategy to improve the performances of this material as ETL. Sn substitution with Mg can affect the work function (WF) and, consequently the power conversion efficiency (PCE) of the device. Considering the limited solubility of the Mg doping in SnO_2 lattice, besides the substitution in Sn sites, we also study the effect of Mg ions in interstitial sites. Here we report an ab initio DFT study on structural and electronic effects of Mg doping in SnO_2, explaining the enhancement in open circuit potential (V_{OC}) reported in literature [1].

2 Methods and Computational details

We perform periodic spin-polarized DFT calculations with the GGA/PBE [2, 3] and hybrid HSE06 [4, 5] functionals, projector-augmented wave (PAW) pseudopotentials and plane wave basis sets with the *Vienna Ab-initio Simulation Package* program (VASP, version 5.4.4) [6-9]. Dispersion is treated within the Grimme's D3 framework [10] with the damping scheme by Becke and Johnson (D3BJ) [11]. A kinetic energy cutoff of 600 eV and a 3x5x1 Γ-centered Monkhorst-Pack k-point mesh are used. We build up a 5L-slab model exposing the (110) surface containing 30 f.u. (90 atoms), in which 1, 2 and 3 inner Sn atoms are substituted with Mg in order to simulate 3, 6 and 10% doping content, respectively. Substituted inner tin atoms are selected as far as possible from each other. Furthermore, considering the 6% configuration, we moved a Mg atom in an interstitial inner site far from the other Mg atom.

*Corresponding author. E-mail: michele.pavone@unina.it

3 Results and Discussion

We first study Mg substitution in Sn sites with three different concentrations and then the insertitial case. The work function values and the corresponding variations are calculated with the HSE06 level of theory as follows:

$$WF = E_{vac} - E_{VB}$$

$$\Delta WF_{(x\%Mg-doped)} = WF_{(x\%Mg-doped)} - WF_{(0\%Mg-doped)}$$

Where E_{vac} is the value of the planar average electrostatic potential in the vacuum region of the the slab model and E_{VB} is the energy of the highest occupied level.

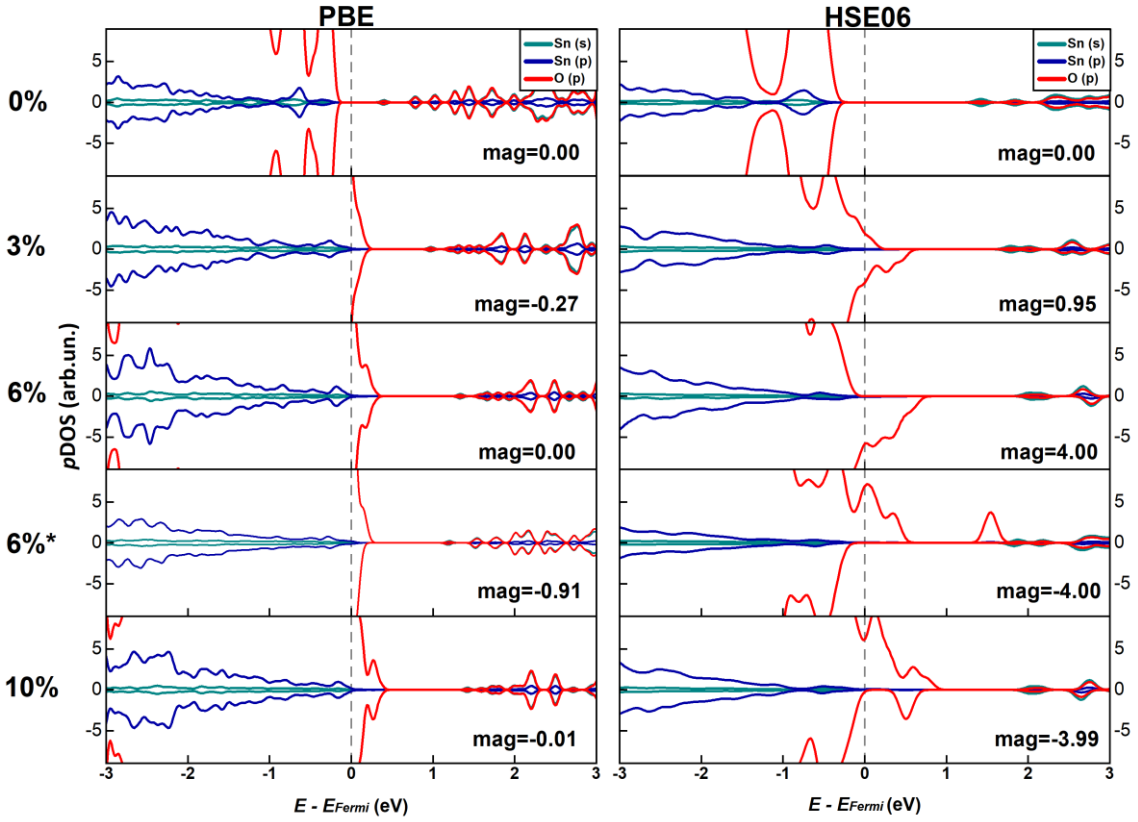


Fig.1: Spin resolved atom and orbital angular momentum-projected density of states (PDOS) calculated at PBE and HSE06 level for SnO_2 slab pristine and Mg-doped at 3, 6 and 10%; dashed line denotes Fermi level; 6%* is the configuration with Mg occupying the interstitial site.

PDOS in Figure 1 show a p-type conduction both in PBE and HSE06 doped slabs that confirms the acceptor behaviour of the doping ions. However, we note a difference for spin up and down states within the hybrid functional calculations of doped slabs and, consequently, the presence of ferromagnetism. The effects of Mg doping on magnetic properties of SnO_2 thin films is already well known [12]. When a Mg ion is placed in the interstitial site, in the 6%* doped configuration, we can observe O(p) states inside the band gap near the Sn (s) states of the CB. These intra-gap states, generated by oxygen atoms near the empty site shown in Figure 2, could promote charge recombination processes and, thus, reduce the PCE of the corresponding PSC.

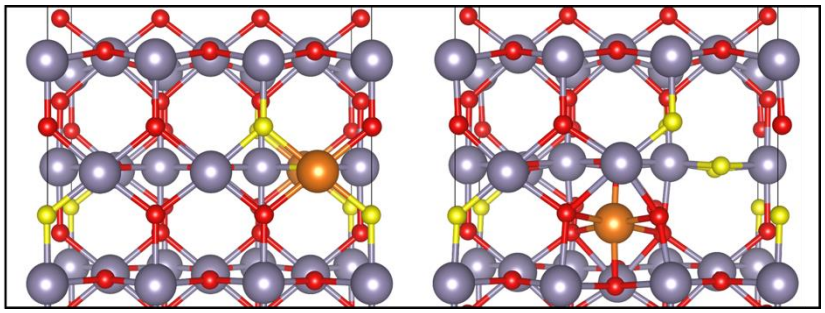


Fig.2: Upper part of the 6% doped slab model with and without Mg atom in the interstitial site, at right and left respectively; oxygen atoms responsible to generate intra-gap states are colored in yellow; legend: oxygen (red), tin (grey), magnesium (orange).

Substitution of Mg ions in Sn sites produces an evident reduction in WF (Figure 3): the higher the doping concentration, the lower the WF value. Due to the limited solubility of Mg doping in SnO_2 starting from ~8% [12], we also assess the influence of a Mg ion in the interstitial site considering a 6% doped model. The WF, shown in Figure 3 with a star symbol, is higher than the configurations with substitutional defects, but lower than the pristine SnO_2. Xiong et al. studied experimentally different percentages of Mg-doped SnO_2 as ETL in PSCs [1]. In Figure 3 are also shown the differences in experimental V_{OC} parameters obtained for the tested devices. This parameter increases with Mg doping and, then, starting from 7.5% it begins to decrease. We can explain this experimental trend considering our theoretical results. The decrease in WF suggests an upper conduction band minimum (CBM) for the ETL material and, consequently, the increase in V_{OC}.

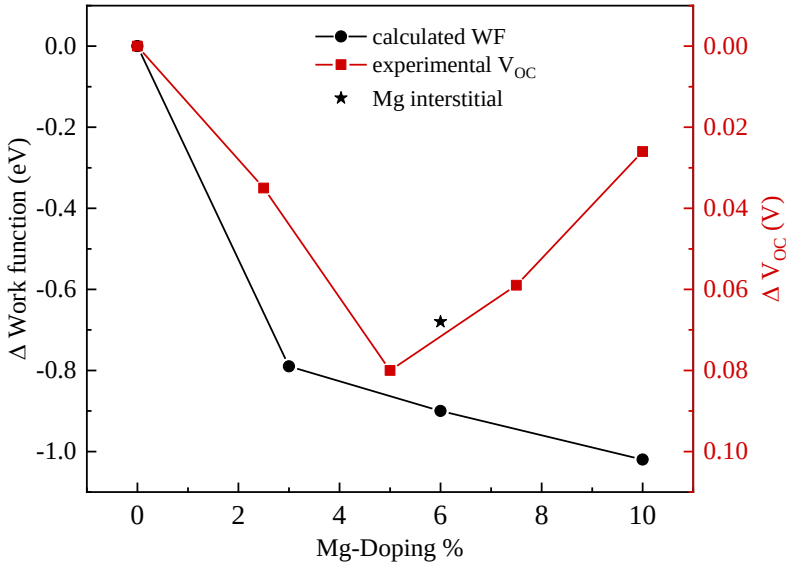


Fig.3: Differences in WF ($WF_{(x\%Mg)}$ - $WF_{(0\%Mg)}$) extracted from calculations at HSE06 level of the theory and experimental V_{OC} differences [1]; black star indicates the difference in workfunction calculated for the configuration with interstitial Mg.

For this reason, we see higher V_{OC} parameters at lower doping concentrations, and then this value starts to decrease again with the insertion of Mg ions in interstitial sites, which lowers the CBM.

4 Conclusions

Both PBE and HSE06 functionals simulate a p-type behaviour for Mg doped slab models, but the hybrid functional provides a more effective description of the electronic properties, pointing out the ferromagnetism already evinced in experimental studies. As experimentally established, Mg doping slightly improves the V_{OC} of SnO_2 based PSCs. We ascribe this effect to electronic reasons. In particular,

we expect higher CBM for doped slabs, which provide higher V_{OC}, considering the WF extracted from the calculations. On the contrary, high doping concentrations and, thus, Mg in interstitial sites are detrimental for the PCE of the devices. The oxygen atoms near the empty sites generated by interstitial Mg produce intra-gap states which are responsible for recombination processes. In addition, the WF increases and, consequently, we expect lower CBM which reduces the V_{OC} of the device. These results are particularly interesting because explain the reason why, regardless to structural interfacial modifications, Mg doping in SnO_2 can enhance the PCE of PSCs. Furthermore, our study paves the way for an alternative to the trial and error approach for the study of different elemental doping inside the most used ETL. The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff; CRESCO/ ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes. See: http://www.cresco.enea.it/english for information [13].

References

[1] Xiong, L., et al., Performance enhancement of high temperature SnO2-based planar perovskite solar cells: electrical characterization and understanding of the mechanism. Journal of Materials Chemistry A, 2016. 4(21): p. 8374-8383.
[2] Perdew, J.P., K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple. Physical Review Letters, 1996. 77(18): p. 3865-3868.
[3] Perdew, J.P., K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. Physical Review Letters, 1997. 78(7): p. 1396-1396.
[4] Heyd, J., G.E. Scuseria, and M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential. 2003. 118(18): p. 8207-8215.
[5] Heyd, J., G.E. Scuseria, and M. Ernzerhof, Erratum: "Hybrid functionals based on a screened Coulomb potential" [J. Chem. Phys. 118, 8207 (2003)]. 2006. 124(21): p. 219906.
[6] Kresse, G. and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. Physical Review B, 1996. 54(16): p. 11169-11186.
[7] Kresse, G. and J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Computational Materials Science, 1996. 6(1): p. 15-50.
[8] Kresse, G. and J. Hafner, Ab initio molecular dynamics for liquid metals. Physical Review B, 1993. 47(1): p. 558-561.
[9] Kresse, G. and J. Hafner, Ab initio molecular-dynamics simulation of the liquid-metal--amorphous-semiconductor transition in germanium. Physical Review B, 1994. 49(20): p. 14251-14269.
[10] Grimme, S., S. Ehrlich, and L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. 2011. 32(7): p. 1456-1465.
[11] Becke, A.D. and E.R. Johnson, A density-functional model of the dispersion interaction. 2005. 123(15): p. 154101.
[12] Wu, P., B. Zhou, and W. Zhou, Room-temperature ferromagnetism in epitaxial Mg-doped SnO2 thin films. 2012. 100(18): p. 182405.
[13] Ponti, G., et al., The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure. 2014: p. 1030-1033.

NUCLEAR ANALYSIS OF XD CONFIGURATION OF DEMO (CONTRIBUTION TO THE WP-DTT1-ADC)

N. Fonnesu*, D. Flammini, R. Villari, G. Mariano, A. Colangeli and F. Moro

ENEA, Department of Fusion and Nuclear Safety Technology, I-00044, Frascati, Rome, Italy

ABSTRACT. The investigation of alternative divertor configurations (ADC) for DEMO is dedicated to finding solutions to complex challenges through analysing the physics and engineering viability of various divertor configurations. Within WPDTT1-ADC the neutronic performances of three configurations have been considered, the single null (SN), SuperX (SX) and X divertor (XD). The present report describes the contribution ENEA to this, in particular on the MCNP model development and simulations performed for the XD configuration.

1 Introduction

EU DEMO will operate at power levels and for times scales approaching that of a commercial fusion plant. DEMO presents a range of complex challenges, some of which can be solved through our experiences operating ITER, but others completely novel, requiring breakthrough research and analysis. The divertor is one such example where a simple extrapolation of the ITER design is potentially not viable, and alternative solutions to the problem of plasma exhaust must be sought. The investigation of alternative divertor configurations (ADC) is dedicated to finding this solution through analyzing the physics and engineering viability of various divertor configurations for implementation in DEMO. For this, within the work package WPDTT1-ADC the neutronic analyses of three configurations have been considered, the single null (SN), SuperX (SX) and X divertor (XD). In this activity, coordinated by CCFE [1] and contributed by CCFE, ENEA, IPPLM, the role of ENEA was to develop the MCNP [2] model of XD configuration, implement tallies for nuclear responses studied, create weight windows to improve the efficiency of simulations and to run the final simulations to get the nuclear responses necessary for the comparison with the other configurations.

2 XD model preparation

The MCNP geometry model of DEMO X-Divertor configuration has been created starting from the two CAD files shown in Fig.1. These files, provided in STEP 3D format, consist of the 22.5° model of the tokamak and the generic model of the cryostat together with the bioshield.

These files were merged and some components adapted to the specific configuration (e.g., the ports have been extended up to the front face of bioshield, through the cryostat; angular opening of intercoil structure reduced to 22.5°). The merged CAD file has been then sorted into the main tokamak components and each of them divided into sub-components. Each sub-component is associated with a distinct material and this requires to define separated geometrical volumes. As examples of the breakdown, in Fig. 2 it is shown an exploded view of divertor (left side) and blanket (right) divided into sub-components.

Considering the limited level of detail of the CAD model, the simplification of small and not relevant geometrical parts as holes and bolts, was about negligible.

*Corresponding author. E-mail: nicola.fonnesu@enea.it

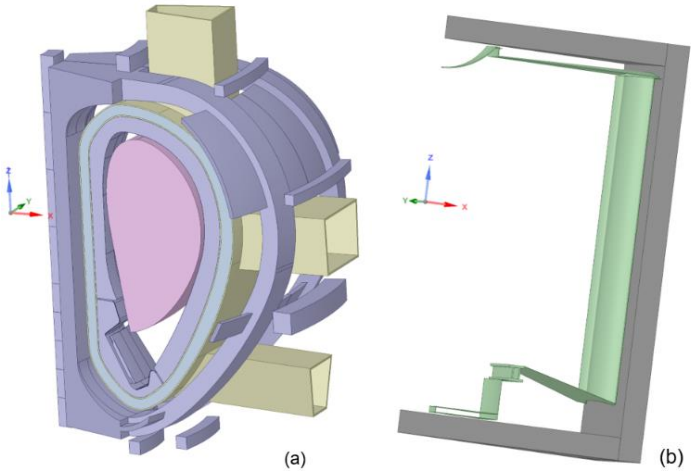


Fig. 1. CAD files in step format of tokamak sector (a) and cryostat with bioshield.

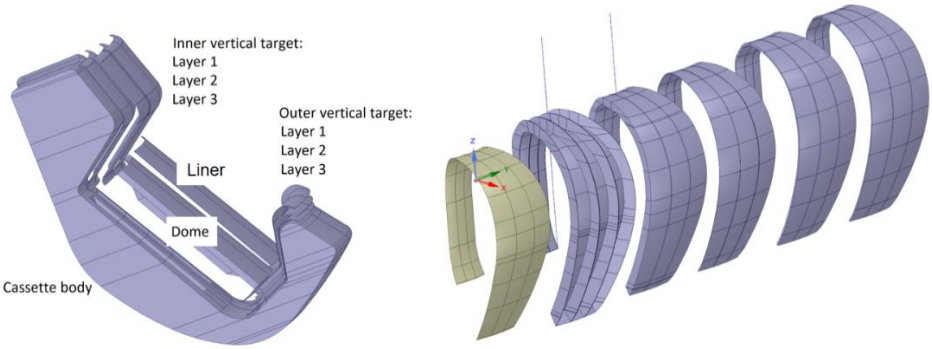


Fig. 2. Exploded views of divertor (left) and blanket (right) after the breakdown into sub-components.

The most relevant part of the geometry simplification process was devoted to the preparation of the CAD model to be converted into MCNP-readable format and this mainly consisted in some modifications applied to the components individually and then in some adjustments on the entire assembly, i.e., on each component in relation to the others.

As for the first type of actions, the correction of some geometry errors affecting the single components (as small faces, non-perfect definition of boundaries of solids, etc.) was performed and then some changes were applied to each component to make them compatible with the conversion tool (i.e., SuperMC [3]), in particular spline removal. MCNP is not capable of reading spline surfaces and SuperMC is not capable to convert them and this implies that all the splines have to be removed from the CAD before conversion.

Splines have been substituted mainly with toroidal, cylindrical and conical surfaces and this modification produced some volume variations in the components as the definition of their envelopes has been partially modified. The substitution of part of the body envelopes has to be precise enough not to alter significantly the volumes and for this, volume conservation is checked in the final model, as reported later.

As for the second type of actions, each component has been checked in relation to the others and the main modifications applied consisted in filling small gaps and removing interferences between bodies, as undefined and over-defined regions of space causes lost particles in the MCNP runs.

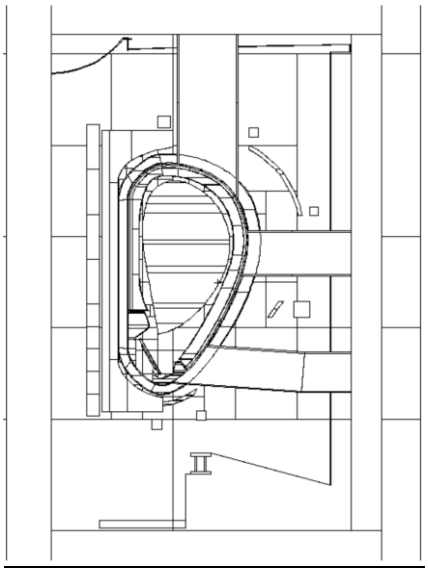


Fig. 3. Vertical cross-section of the MCNP geometry model created after correction.

Finally, before conversion, sub-components have been split into groups of geometry cells, as simple as possible to facilitate the conversion and avoiding some situations which are known to produce conversion errors (e.g., the intersection of a cylindrical and toroidal surfaces in a cell).
The MCNP geometry model created consists of 1460 geometry cells, of which 1171 are tokamak components and the rest are void space between the components and surrounding the tokamak. 1692 surfaces are employed to define the boundaries of these cells. A vertical cross-section of the model is shown in Fig. 3. To reduce the number of lost particles detected in some MCNP test-runs, the definition of the geometry cells involved has been reviewed.
The assumption that modifications introduced in the original CAD model do not alter significantly the volumes was checked by comparing the volume of each component of the CAD model with the corresponding one in the converted model. To calculate the volumes in the converted model, an MCNP run in void was launched with a spherical particle source pointing inward, 5 m of radius and surrounding the geometry domain. 5E9 particles were generated to sample the volumes and the statistical errors are below 0.12%. 83 particles got lost.
The maximum discrepancy is about 3.5% for the cryostat while the other components are below 0.2%. Cryostat shows the higher discrepancy and this is believed to be due to the large curved surfaces compared to the volume of the element. The conversion of this surfaces introduces some numerical approximations in the model and, considering the ratio surface/volume of the element, it has a higher impact on the volume discrepancy.
Lost particles were checked with some MCNP test-runs in void. Thousands of lost particles (out of 1E9) were observed in the model created by SuperMC thus requiring their reduction with manual correction of the geometry definition of some cells.
After the correction of these geometry errors, 84 lost particles out of 5E9 are tracked in MCNP runs for volume calculation and about 3000 out of 1E9 in the MCNP runs of the final model (i.e., tokamak sector with reflective surfaces, materials applied to each cell and weight windows).

3 XD weight window

To improve sampling in the regions of the phase space of importance for the evaluation of the nuclear responses selected and so to reduce statistical errors on these quantities, the method of weight windows was applied. Weight windows have been generated with ADVANTG code [4]. Even if ADVANTG can generate a weight window map both for neutrons and photons, to optimize the calculation time, such map was generated only for neutrons and the region chosen for variance reduction was the one defined for mesh tally #1 (i.e., neutron flux). The improvement of the statistics for photons relies on the secondary photon production.

For the deterministic transport simulation performed by DENOVO (i.e., the deterministic transport code used by ADVANTG for solving the adjoint transport problem) a mesh size of about 20x20x20 cm^3 covering the same region of mesh tally #1, was chosen and applied to the unfolded geometry model, i.e. the 360° model created by ADVANTG from the original 22.5° MCNP model. This results in about 7.4E6 voxels. Mesh size is about twice the one used for mesh tally #1 and a finer mesh was precluded due to memory limit reached in the workstation employed to generate weight windows.
Geometry reconstruction performed by ADVANTG with ray sampling, to reconstruct the geometry and to associate an equivalent material to each voxel, has produced the model shown in Fig. 4(a).
As a proof of the effectiveness of weight windows on the results, a comparison between statistical errors of neutron flux as scored in mesh tally #1 without (left-hand side) and with (right-hand side) the application of the weight windows is shown in Fig. 4(b). The overall sampling improvement is evident over the entire mesh cross-section and in particular in the inboard part of the TF coil, where statistics is extremely poor (deep blue in the mesh tally cross section on the left-side of Fig. 4(b) indicates that a very limited number of particles are scored in that region).

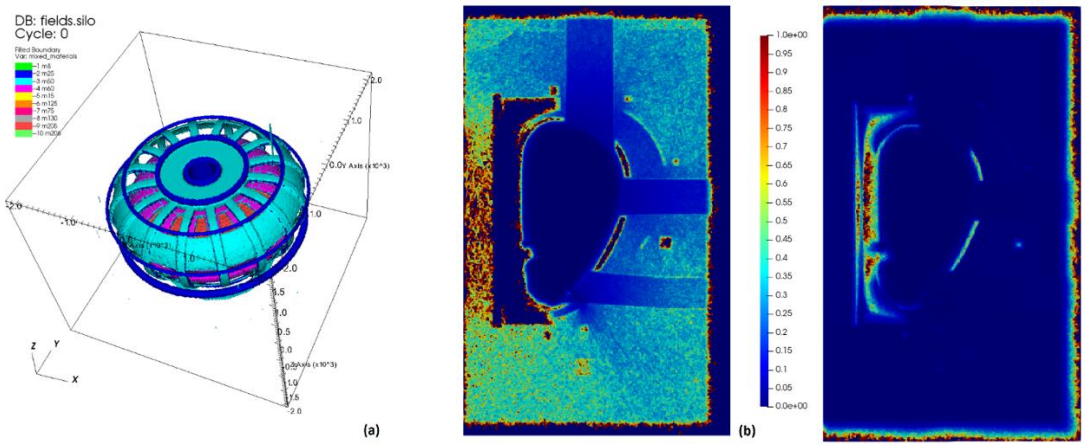


Fig. 4. (a) geometry model reconstructed by ADVANTG with ray sampling in voxels 20x20x20 cm^3; (b) statistical errors in mesh tally #1 without (left-hand side) and with (right-hand side) the application of weight windows as variance reduction technique.

4 Nuclear Responses

A range of different nuclear responses has been obtained as output of the MCNP runs, both specific the divertor and globally up to the bioshield, including neutron and photon flux, nuclear heating, power load, tritium production rate in the blanker breeder zone, DPA and helium production in divertor materials. Some of them are described in the following.
The neutron flux is a fundamental quantity underpinning all other responses. Studying the neutron flux distribution gives insight into the behavior observed in other responses. The neutron flux was calculated globally in the sector using a 10 cm voxel mesh tally as shown in Fig. 5(a). Around the divertor, the resolution of the mesh is increased using a 5 cm voxel size which is plotted in Fig. 5(b).
As significant to measure the damage in critical components, DPA per Full Power Year (FPY) is shown in Fig. 6(a) over the divertor region. The limit to the DPA/FPY in Eurofer is 6 DPA over 1.5 FPY, therefore we can assume a limit of 4 DPA/FPY for the mentioned plot. The highest DPA values are recorded on the dome and on the part of the two vertical targets which sit almost below the dome.
Helium production in Eurofer is in Fig. 6(b). The limit for rewelding in terms of helium production is equal to 1 appm/ 6 FPY. This gives 0.16 over 1 FPY, the contour for which is given in the figure. In theory, only beyond this contour can these operations be performed and this is deep within the cassette.
Another relevant aspect is the power deposited in the TF coils. To see more clearly the nuclear heat density with respect to the design limit, the calculated nuclear heat density is divided by 5 $x10^{-5}$ Wcm^{-3}. The result is shown in Figure 7(a). A minimum threshold is set to 1 so that only results exceeding the limit is shown.

The extracted nuclear heat density calculated through poloidal segmentation of the TF coils is plotted in Figure 7(b) for the casing (upper plot), and for the winding pack (lower plot). The limit is also plotted showing that for all poloidal segments, the configurations are below the limits. For all outboard segments the limit is exceeded. Numbering scheme of poloidal segments corresponding to the mentioned plots in given in the same figure, right-hand side.

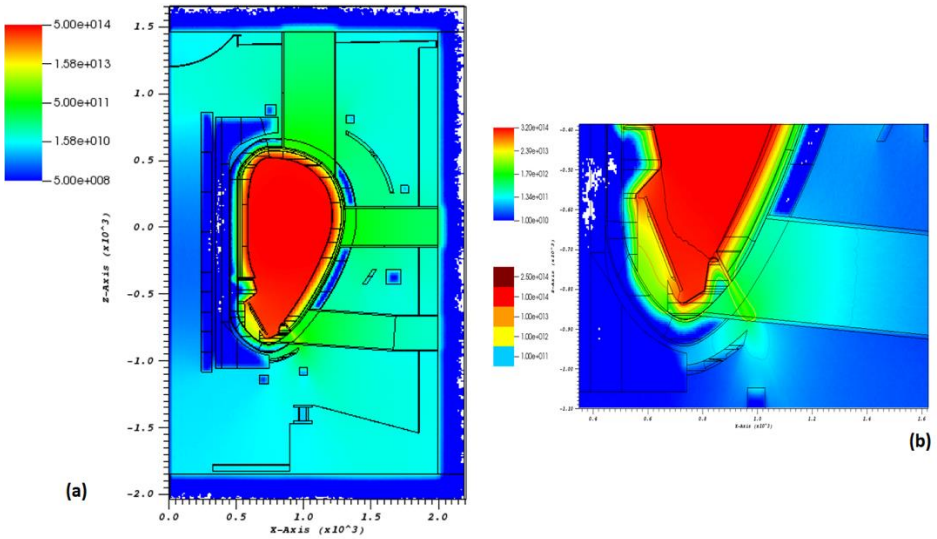


Fig. 5. Neutron flux ($cm^{-2}s^{-1}$) over the entire tokamak sector (a) and over the divertor area (b).

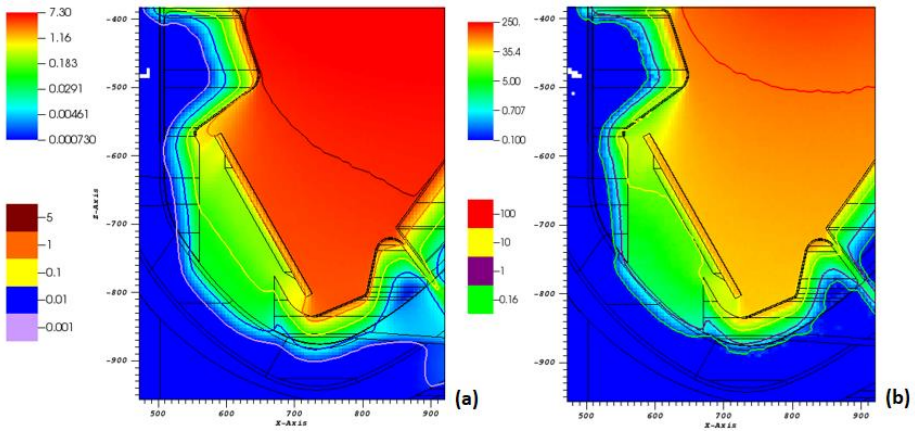


Fig. 6. (a) Damage (DPY/FPY) and (b) helium production (appm/FPY) in Eurofer over the divertor region; lower color bars refer to contour lines.

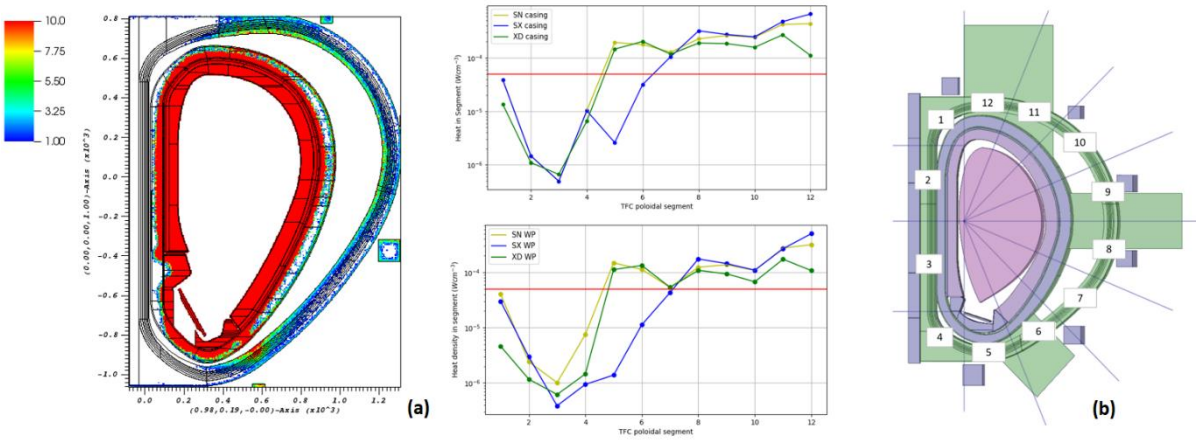


Fig. 7. (a) Ratio of the nuclear heat density to the limit of $5x10^{-5}$ W cm^{-3} in TF coil; (b) Poloidal profile of the nuclear heat density in the TF coil casing (upper plot) and in the winding pack (lower plot) according to the segmentation scheme shown in the same figure on the right. The limit of $5x10^{-5}$ W cm^{-3} is shown with a red line.

5 Conclusions

The present report presents the contribution of ENEA to the comparative analysis of the alternative divertor configurations that are being considered for implementation in EU DEMO. A simplified model with homogenized material descriptions has been produced for the XD model. A range of nuclear responses have been calculated, both local to the divertor but also spanning the entire sector to capture the different geometry of all components in each configuration considered in the comparative study.
The response of systems in the ex-vessel region is highly dependent on the size of the port openings.
A large neutron flux field in the lower port is observed in the XD model. This is predominantly the results of an 8 cm gap between the blanket and divertor coupled with bulk transport through the divertor itself. As for nuclear damage, peak values in the region 2-4 DPA/FPY were found for XD which should be reduced to ensure the lifetime of the cassette. Finally, the poloidal profile of nuclear heating in the TF coils was calculated and determined to exceed the limit of $5X10^{-5}$ W cm^{-3} on the outboard of the coils in the poloidal sections adjacent to the lower, equatorial and lower ports. In the XD model the limit is exceeded for all outboard segments, while for all inboard segments the limit is satisfied.

References

[1] A. Valentine at al., Neutronics assessment of EU DEMO alternative divertor configurations, Fusion Engineering and Design, Volume 169, August 2021, 112663
[2] C. J. Werner et al., “MCNP User’s Manual Code Version 6.2,” Los Alamos Natl. Lab., 2017.
[3] Y. Wu et al., “CAD-based Monte Carlo program for integrated simulation of nuclear system SuperMC,” Ann. Nucl. Energy, 2015.
[4] S. W. Mosher et al., ADVANTG: An Automated Variance Reduction Parameter Generator. 2015.

PRELIMINARY STUDIES FOR THE SORGENTINA-RF FUSION DEVICE RADIATION SHIELDING WITH CRESCO-4

P. Ferrari[1*], G. M. Contessa[2], G. Gadani[2], L. Lepore[2], F. Moro[2] and A. Pietropaolo[2]

[1]*ENEA, Radiation Protection Institute, Via Martiri di Monte Sole 4, 40129 Bologna (BO)*[1]

[2]*ENEA, Fusion and Technology for Nuclear Safety and Security Dep., Via E. Fermi 45, 00044 Frascati (Roma)*

ABSTRACT. In the framework of an agreement among ENEA, and the Regions Emilia-Romagna and Toscana it was established that the ENEA Reseach Center, located near Brasimone lake (Bologna), will host a series of installations for advanced technological research. SORGENTINA-RF is a new concept of plant designed to produce radionuclides of clinical relevance that relies on a neutron fusion source. This future high intensity neutron source will require a proper shielding in order to not exceed the annual limiting radiation doses for the workers. The CRESCO-4 grid has been employed with a multiprocessor version of the Monte Carlo code MCNP6 to perform a series of complex preliminary shielding evaluations in a reduced amount of time.

1 Introduction

In the framework of an agreement among ENEA, Emilia-Romagna and Toscana Regions the ENEA Reseach Center, located near Brasimone lake (Camugnano - Bologna), was selected to host a series of installations for advanced technological research, one of these, denominated SORGENTINA-RF, is a facility of new concept that will be used to test the feasibility of producing radionuclides of medical relevance employing a neutron fusion device [1-3]. In particular, SORGENTINA-RF is designed to produce ^{99}Mo [4], which is the precursor of ^{99m}Tc, that is employed in the 85% of the nuclear medicine diagnostic procedures in the world [5], and currently produced in a restricted number of research nuclear fission facilities [6].

The core of SORGENTINA-RF will be a 250 kW ion LINAC, generating a current of about 1 A, of 50% deuterium and 50% tritium ions, that will produce fusion reactions on a rotating Ti-coated aluminium target. The nuclear fusion neutrons, with energy of 14 MeV, will irradiate metallic ^{100}Mo spheres of few millimetres diameters, in turn inducing ^{100}Mo(n,2n)^{99}Mo reactions. The produced ^{99}Mo will be cumulated for the subsequent radiochemical processing.

The designed source intensity is established to be 7.10^{13} s^{-1}, then a proper radiation shielding is required to prevent undue workers exposure.

The aim of the present study is providing a preliminary evaluation of the material composition and thickness needed for this shielding. All the simulations have been done with MCNP6 Monte Carlo code but as the shielding calculations are really time consuming, the CRESCO-4 grid infrastructure [7] has been employed to assure the convergence of the results in a reduced amount of time. In this paper, a very concise description of the work done is presented.

2 Materials and methods

In these preliminary evaluations, a very simplified condition has been studied. The source term is defined as a disk source of 10 cm diameter, isotropically emitting monoenergetic 14 MeV neutrons. No additional element or structure has been considered within the shielding. The designed dose constraint, defined as the rate of the ambient dose equivalent, H*(10), to be fulfilled in contact with the external surface of the shielding, was set to 10 μSv/h.

All the simulations have been done with the Monte Carlo code MCNP6 [8] and CRESCO-4 grid infrastructure. A natural elemental and isotopic composition has been chosen for the simulated materials,

*Corresponding author. E-mail: paolo.ferrari@enea.it

when not supplied directly by the manufacturer, and S(α,β) factors, correcting neutron thermal elastic cross sections for the molecular and lattice binding energy, have been considered. For neutrons ENDFB-7.1 [9] libraries have been used. Photons are generated by the interactions of the neutrons with the shielding materials and the detailed treatment for their transport throughout the geometry has been followed.
The main problem associated to such a kind of calculations is related to the precision achievable for the results. As it is widely known, the uncertainty related to a Monte Carlo calculation is strictly dependent on the number of simulated events. The Law of Large Numbers implies that the average of a sample converges to the average of the population (so to the "real value") when the dimension of the sample goes to infinite. During a Monte Carlo simulation, each particle, generated in the source and transported through the geometry, corresponds to a track (history). The average of these histories, in a selected region of the phase space, corresponds to the value one is looking for. Only when the number of simulated histories is large enough, their average value corresponds to the "actual value" (the value the quantity should have in the "real" world).
In the present study, the Monte Carlo simulations aimed at estimating the value of the H*(10) rate outside the shielding, which is directly dependent on the number of the particles surviving the shielding. That means a contradicting situation: a good simulated shielding implies that only few particles can survive, but few particles outside the shielding implies a bad statistics, i.e. it is not possible to assure that the calculated value is representative of the "real" situation.
In order to solve this contradiction two methods were followed:

- variance reduction techniques (VRT);
- employing a MCNP6 multiprocessor version.

The VRT can be used to biasing the natural probability leading the radiation transport and to re-normalize the statistics at the end of the transport. In this way is possible to overcome the problem of lacking of histories in the geometry region that is more important for the results [10].
On the other hand, the multiprocessor version of MCNP6 allowed to drastically reduce the CPU time required for the calculations. In order to give an idea of the amount of time needed in this kind of simulations one has to consider that the employed inputs required the generation of 10^9 particles, moving in a volume of the order of 10^3 m^3 and producing about 0.9 10^{12} collisions events. Each event is treated in the detailed mode by the Monte Carlo, i.e., the event probability channel is selected on the basis of the energy of the particle, the products of the interactions are generated and followed and stored for the subsequent step. As a matter of fact, exploiting up to 256 CRESCO-4 grid nodes it was possible to reduce the uncertainties of the calculated values to 10-20% over 24 h simulation times.
The dosimetric evaluations have been performed simulating spheres, of proper dimension, filled by air, and placed in some selected points of the geometry. The rates of H*(10)~~,~~ have been calculated convolving the radiation flux spectra with the appropriate conversion coefficient taken from ICRU 57 publication [11].

3 Results

The project of the shielding followed three subsequent phases: from the very simple case, with a point source in a spherical geometry (an approach similar to that followed in [12]), then to a cylindrical model, with dimensions more similar to the expected final construction, arriving at the foreseen configuration for the project.
In figure 1, the proposed final configuration, as resulted from the calculations performed, is shown. The $4 \times 4 \times 4$ m^3 internal chamber will host the ion accelerator, the target and the necessary ancillary equipment. The walls, floor and ceiling will be constituted by 2 m thick ordinary concrete followed by one additional layer of 1 m of baritic concrete. The internal layer of the ordinary concrete can slow the 14 MeV neutrons through a series of elastic collision, produced, in particular, in the interaction of impinging neutrons with the low Z elements of the mixture. The higher density and different composition of outer layer of baritic concrete is employed to reduce the photon radiation [13] produced in the shielding itself by the absorption of the primary neutrons with the creation of deuterium, n(^{1}H,^{2}H)p. The access to the room is possible through a maze, as typically happens for this kind of structures, that will be closed by a properly shielded door. The whole structure of the shield consisted

of precast concrete slabs that will simplify the future decommissioning of the site, accordingly the Italian law implementing the 2013/59/Euratom Directive [14].

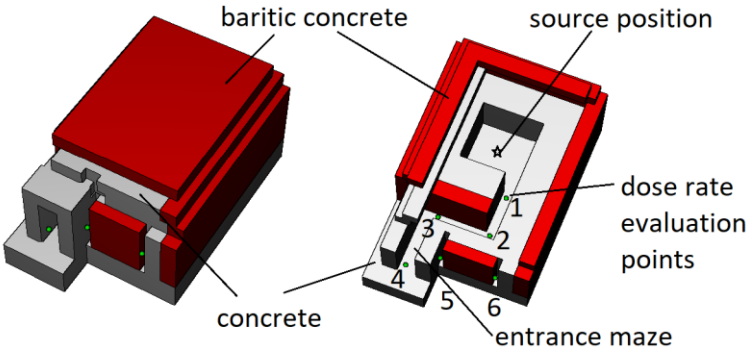


Fig.1: Shielding scheme of Sorgentina-RF facility.

In table 1, the rates of H*(10) in some selected points of the geometry are reported. These rates are the sum of the neutron and photon contribution to the dose equivalent. As can be seen, the proposed project allows to keep the rates below the dose constraint of 10 μSv/h, in the point 5 and 6, as required in the design phase.

Table 1: H*(10) rates in some selected points of the geometry. Numbers refer to figure 1. The asterisk (*) indicates uncertainties higher than 10%.

Position	Total H*(10) (mSv/h)	Mean neutron energy (keV)
1	1.09×10^5	4000
2	7.39×10^3	647
3	3.41×10^2	56.7
4	17.6	3,3
5	0.005 *	0.02 (*)
6	0.007 *	(not calculated)

4 Conclusions

The aim of this work was studying and optimizing the preliminary shielding of the SORGENTINA-RF facility. A multiprocessor version of the Monte Carlo code MCNP6 was employed. The majority of the simulations have been performed employing the CRESCO-4 grid infrastructure, by using up to 256 nodes, in combination with selected variance reduction techniques. That strategy allowed at handling the large computational workload needed for this investigation, fulfilling the requests of the SORGENTINA-RF project management.

References

[1] Capogni, M.; Pietropaolo, A.; Quintieri, L. 2016 99mTc Production via 100Mo(n,2n)99Mo Using 14 MeV Neutrons from a D-T Neutron Source: Discussion for a Scientific Case; RT/2016/32/ENEA;

[2] Capogni M., Pietropaolo A., Quintieri L., Fazio A., De Felice P., Pillon M., Pizzuto A. 2018b 99mTc by 99Mo produced at the ENEA-FNG facility of 14 MeV neutrons Appl. Rad. and Isotopes 134 105–107

[3] Pillon M., Angelone M., Pizzuto A., Pietropaolo A. Feasibility study of an intense D–T fusion source: "The New Sorgentina" Fusion Engineering and Design 89 (2014) 2141–2144

[4] A. Pietropaolo, G.M. Contessa, M. Farini, N. Fonnesu, R. Marinari, F. Moro, A. Rizzo, S. Scaglione, N. Terranova, M. Utili and The SRF-Collaboration. SORGENTINA-RF project: fusion neutrons for 99Mo medical radioisotope. EPJ-P, manuscript number EPJP-D-21-02529
[5] OECD-NEA 2019, The Supply of Medical Isotopes: An Economic Diagnosis and Possible Solutions, OECD Publishing, Paris, https://doi.org/10.1787/9b326195-en.
[6] NAP 2018, National Academies of Sciences, Engineering, and Medicine. 2018. Opportunities and Approaches for Supplying Molybdenum-99 and Associated Medical Isotopes to Global Markets: Proceedings of a Symposium. Washington,DC: The National Academies Press.
[7] Ambrosino et al. 2017 The CRESCO HPC environment The International Conference for High Performance Computing, Networking, Storage and Analysis November 12-17, 2017 Denver, (USA)
[8] D, B, Pelowitz (editor) 2013 MCNP6 USER'S MANUAL Version 1,0. LANL Los Alamos LA-CP-13-00634
[9] Chadwick et al. 2011 ENDF/B-VII.1 Nuclear Data for Science and Technology: Cross Sections, Covariances, Fission Product Yields and Decay Data Nuclear Data Sheets 112(12) 2887-2996
[10] Gualdrini G. and Ferrari P. Monte Carlo variance reduction techniques: an overview with some practical examples Radiation Protection Dosimetry (2011), Vol. 146, No. 4, pp. 425–433
[11] ICRU 1996 Report 57 Conversion Coefficients for use in Radiological Protection against External Radiation
[12] Yun S , Lee C W, Lee D W, Kim S, Jung B, Chang D, Jin H, Shin C.W. 2019 An optimization study for shielding design of D-D and D-T neutron generators Fusion Engineering and Design 146 (2019) 531–534
[13] Akkurt L. and El-Khayatt A.M. 2013 The effect of barite proportion on neutron and gamma-ray shielding Annals of Nuclear Energy 51 5-9
[14] Decreto Legislativo 31 luglio 2020, n. 101. Attuazione della direttiva 2013/59/Euratom, che stabilisce norme fondamentali di sicurezza relative alla protezione contro i pericoli derivanti dall'esposizione alle radiazioni ionizzanti, e che abroga le direttive 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom e 2003/122/Euratom e riordino della normativa di settore in attuazione dell'articolo 20, comma 1, lettera a), della legge 4 ottobre 2019, n. 117. GU Serie Generale n.201 del 12-08-2020 - Suppl. Ordinario n. 29.

AB INITIO STUDY OF THE ELECTRONIC PROPERTIES OF H- AND CL- FUNCTIONALIZED SILICON NANOWIRES

Barbara Ferrucci[1], Francesco Buonocore[2*], Simone Giusepponi[2] and Massimo Celino[2]

[1]*ENEA, Bologna Research Centre, I- 40129 Bologna, Italy*

[2]*ENEA, Casaccia Research Centre, I-00123 Rome, Italy*

ABSTRACT. High chemical reactivity and large surface-to-volume ratio have recently led to growing interest in the employment of silicon nanowires (Si NWs) in many optoelectronic applications. The working principle of Si NWs device resides in the possibility to induce modifications in their electronic properties via molecular interactions. However, the molecular interactions on confined structures (such as single Si NWs) have distinct properties than classical surface chemical reactions (such as on silicon surface). In the first phase of the SUPER project an ab initio approach has been developed to evaluate the stability and electronic properties of H- and Cl-functionalized Si NWs. Because the results will be used to generate different adsorption configurations of grafted molecules, the adsorption of a C8 alkane molecule on Cl- functionalized Si NW has been investigated. This research will provide tools to design efficient procedures to increase the overall efficiency of the devices based on Si NWs.

1 Introduction

In the first phase of the SUPER project, we modelled the electronic properties of SiNWs fully passivated with H and Cl by using the ab initio approach. Indeed, it is of considerable interest to develop a sound strategy to prevent extensive Si NW surface oxidation while preserving the low surface recombination velocities for longer time. Studies with two-dimensional (2D) substrates show that adsorption of organic molecules to the Si surface through Si-C bonds lead to lower rate of oxidation and electron-hole recombination than H-terminated Si surfaces. To this end, the H-terminated Si surfaces have been alkylated using chlorination/alkylation [1]. Methyl-terminated Si NW, which is a passive molecule, gives complete coverage that effectively inhibits oxidation and extend the carrier life time.
Genetic algorithm studies based on DFT and MD calculations have shown that the Si<112> NWs expose the (111) and (110) facets [2] in nice agreements with the HRTEM characterization [3]. In this study, DFT models have been established to find the optimized geometry and calculate the total energy of the chloride- and hydrogen-functionalized SiNWs. We investigated the electronic properties of Cl- and H-fully passivated SiNWs oriented along the <112> direction, focusing on two surface orientations, Si(111) and Si(110). Finally, the adsorption of a C8 alkane molecule on H- and Cl- functionalized Si NW has been investigated.

2 Methods

Calculations were performed within the periodic density functional theory (DFT) as implemented in the PWscf (plane-wave self-consistent field) code of the QUANTUM ESPRESSO package package [4], compiled with Intel Fortran compiler, Math Kernel Library (MKL) and Message Passing Interface (MPI) parallelization on the Cresco6 cluster of the high-performance computer facility CRESCO [5], with a good scalability up to hundreds of cores. The exchange-correlation energies were treated with the Perdew, Burke and Enzerholf (PBE) functional [6]. For all structures, the electron–ion interactions were described by the ultra-soft Vanderbilt pseudopotential [7] using an energy cut-off of 40 Ry, and a charge density of 400 Ry. The force and energy threshold have been imposed equal to 1.0E-4 Ry/Bhor and

*Corresponding author. E-mail: francesco.buonocore@enea.it

1.0E-5 Ry respectively. Self-consistent calculations to sample the Brillouin zone were performed to investigate the convergence of total energy with respect to different k-points grids. The geometries of all structure were fully optimized.
Using first-principles calculations we studied, (1) the preferential adsorption sites of H and Cl on the silicon (111) and (110) surfaces, (2) the effects of the cross-sectional area dimension on the electronic band structure of fully passivated H- and Cl- SiNWs, and (3) the formation energy and electronic band structure of SiNWs, fully passivated with H and Cl, and with an alkane chain absorbed on the (111) and (110) surfaces.
We modelled an eight-layers silicon slab with 8 atoms for each layer enclosed in a 2×2 supercell repeated in all three dimensions with a k-point mesh of 4×4×1 to investigate the preferential adsorption sites of H and Cl atoms on the silicon (111) and (110) surfaces which minimize the system's energy. The system consists of symmetric slabs, at the optimized bulk lattice constant of 5.469 Å, and with 30 Å of separation between periodically repeated cells along the z direction to prevent interactions between adjacent layers. As (geometry) optimizer was used the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm to minimize the slab energy.
First principles calculations based the density functional theory (DFT) were performed to investigate the formation energies and electronic properties of a series of fully passivated SiNWs with different cross sections. Using a supercell periodic along the <112> direction two sets of passivated systems were modelled: H-SiNWs (silicon nanowires fully passivated with hydrogen) and Cl-SiNWs (silicon nanowires fully passivated with chlorine). Each set consists of three SiNWs characterized by a different cross-sectional area (A) enclosed with the (111) and (110) facets. A (1×1×8) k-points grid has been used to sample the Brillouin zone which yields convergent results for the total energy.
We finally investigated the electronic properties of SiNWs, fully passivated with H and Cl, and with an alkane chain (C_8H_{17}) absorbed on the (111) and (110) surfaces.

3 Results and discussion

In this section, the results provided by the aforementioned simulations are reported. The total energy and the formation energy of Si slab systems and SiNWs systems were calculated to investigate the more stable configurations. The results shown that the preferential site of passivating atoms on the Si (111) and (110) surface are the top site positions. Based on this, the functionalization of the SiNWs was performed positioning the H or Cl atom directly above the Si atoms. We modelled two sets of hydrogenated and chlorinated SiNWs each consisting of three systems with different cross-sectional area, labelled as small, medium and large, and composed of 24, 96, and 216 atoms of Si, and 20, 40, and 60 atoms of Cl or H, respectively. For example, the optimized geometries of the fully Cl-passivated SiNWs are shown in Fig. **1** below.

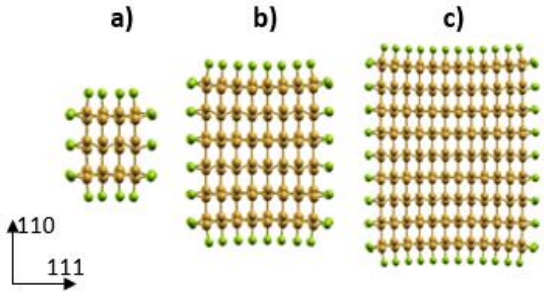


Fig. 1: Structures (green colour: Cl atoms, yellow colour: Si atoms) of fully passivated SiNWs, a) Si24Cl20; b) Si96H40; c) Si216H60.

For each system we calculated the formation energy, the energy band structure (Fig. 2) and the partial density-of-states (PDOS) (Fig. 4).From the formation energies calculations, we found that the stability trend is: $Si_{24}Cl_{20}$> $Si_{96}Cl_4$> $Si_{216}Cl_{60}$; and $Si_{216}H_{60}$> $Si_{96}H_{40}$ $Si_{24}H_{20}$, indicating that Cl-SiNWs are more stable than the H-SiNWs.

Due to the quantum confinement effect, the band-gaps reduce as the SiNWs cross-sectional areas increase, as shown in Figure 3, and the difference between the band gaps of the two passivated systems decreases from 0.35 to 0.03 eV going from the small to the large SiNWs.

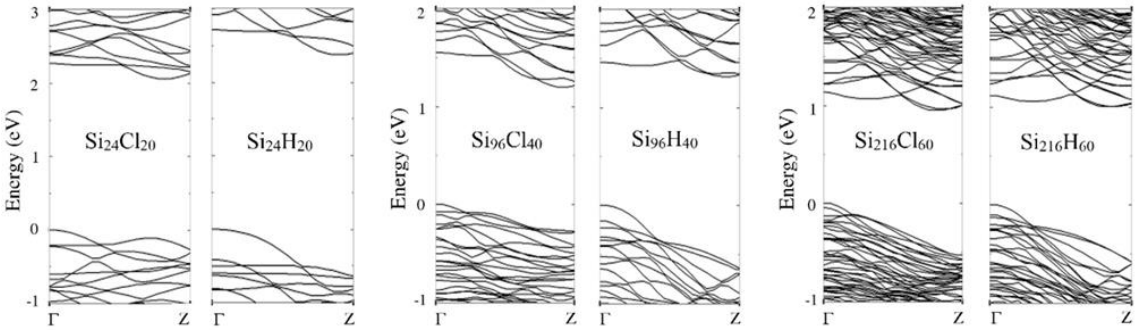


Fig. 2: Band structures of fully passivated <112> Cl- and H-SiNWs. The Fermi level is set to zero.

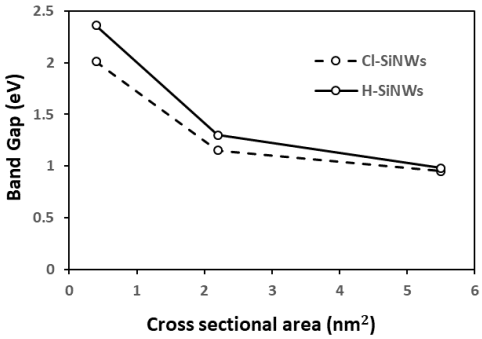


Fig. 3: Computed band-gap as a function of the cross-sectional area for <112> oriented fully passivated H-SiNWs and Cl-SiNWs.

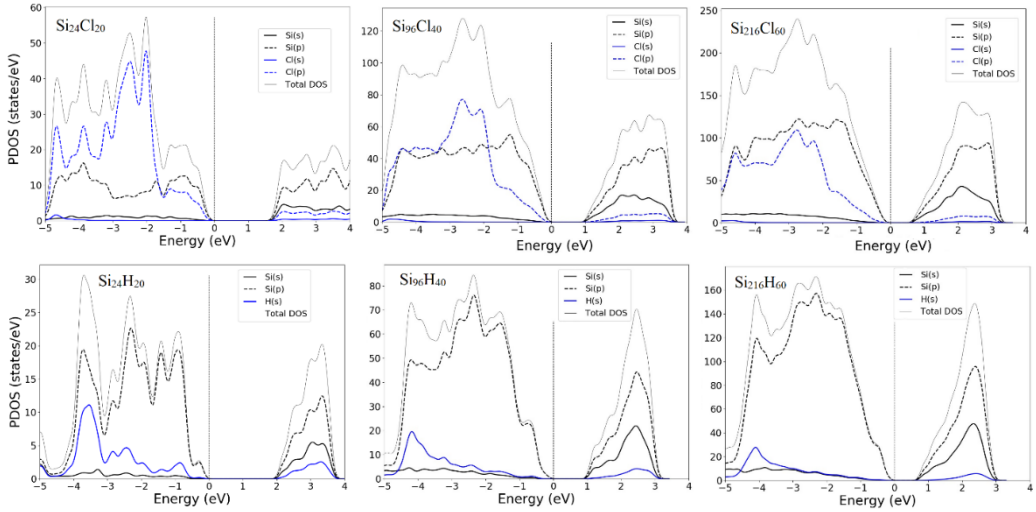


Fig. 4: Projected electron density of states (PDOS) of fully passivated <112> Cl-SiNWs and H-SiNWs.

4 Computer resources

We report the benchmark results for PWSCF module as implemented in Quantum Espresso suite. We inspected the CPU time to calculate the ground-state, and the total force of a Si system of 216 atoms with a 6x3x1 cell size, and 1x1x8 k-points grid. The number of cores used were 48, 96, 240, 480, and 720. Two cases were investigated: case1 with the 110 facets of the 216 SiNW parallel to the xy plane, (Lx= 43.201 Å, Ly= 48.216 Å, Lz = 6.668 Å), and the case2, obtained by rotating the system to make the 110 facets parallel to the yz plane (Lx = 6.668 Å, Ly= 43.201 Å, Lz= 48.216 Å). For each case two runs were done by setting the npools option to 1 and 5. In Fig. **6** the speedup and efficiency are normalized to the run on 2 nodes with 96 tasks for the case1 the, and to the run on 1 node with 48 tasks for the case2. In Table 4 are reported the CPU time value of the case1 and case2 simulations as a function of the core number and nodes, and the time gained passing from the case1 to the case2. The results show that for the case2, setting the option npools to 5 and scaling up from 96 to 480 cores, there is a gain of 5.4 times of the calculation time and a good performance with an efficiency of ~ 82%. This improvement in the performance is mainly due to (1) the rotation of the system and to (2) the npools option of the code. Since the code distributes the workload along the z dimension, making a rotation so that the side of the supercell along this direction is the largest, causes that the workload is better distributed among the cores. The npools option allows to dedicate for each k point a group of cores minimizing communication between them. This type of "parallelization" is almost linear with number of k points.

Table 4: CPU time calculations for the two simulation cases and the corresponding time (Δt) gained going from case1 to case2 and by using the options npools 1 and npools 5.

Nodes	Cores	CPU time (s) Case1 npools 1	CPU time (s) Case1 npools 5	CPU time (s) Case2 npools 1	CPU time (s) Case2 npools 5	Δt (s) npools 1	Δt (s) npools 5
1	48	-	-	8880	-		
2	96	5880	-	4200	-	1680	-
5	240	4860	2847.7	2216	2230.5	2644	617
10	480	5580	1619	1655.3	1083.1	3925	536
15	720	5460	1511	1553.9	1106.5	3906	404

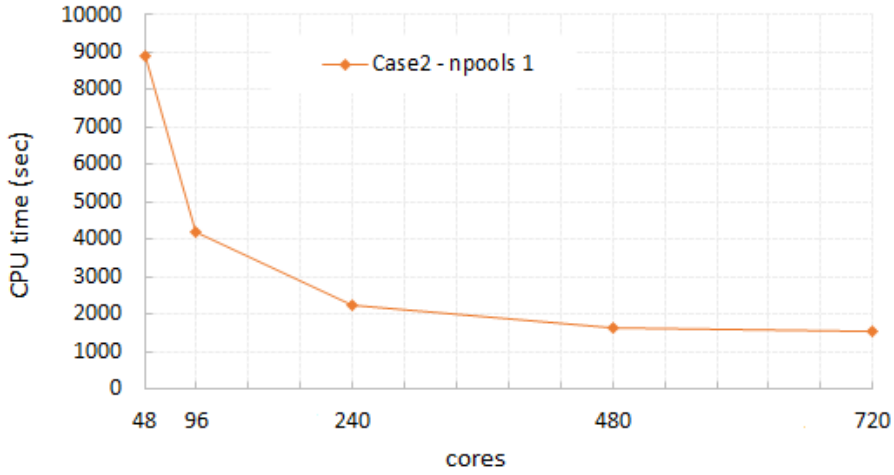


Fig. 5: CPU time as a function of the core number used in the Case2 simulations

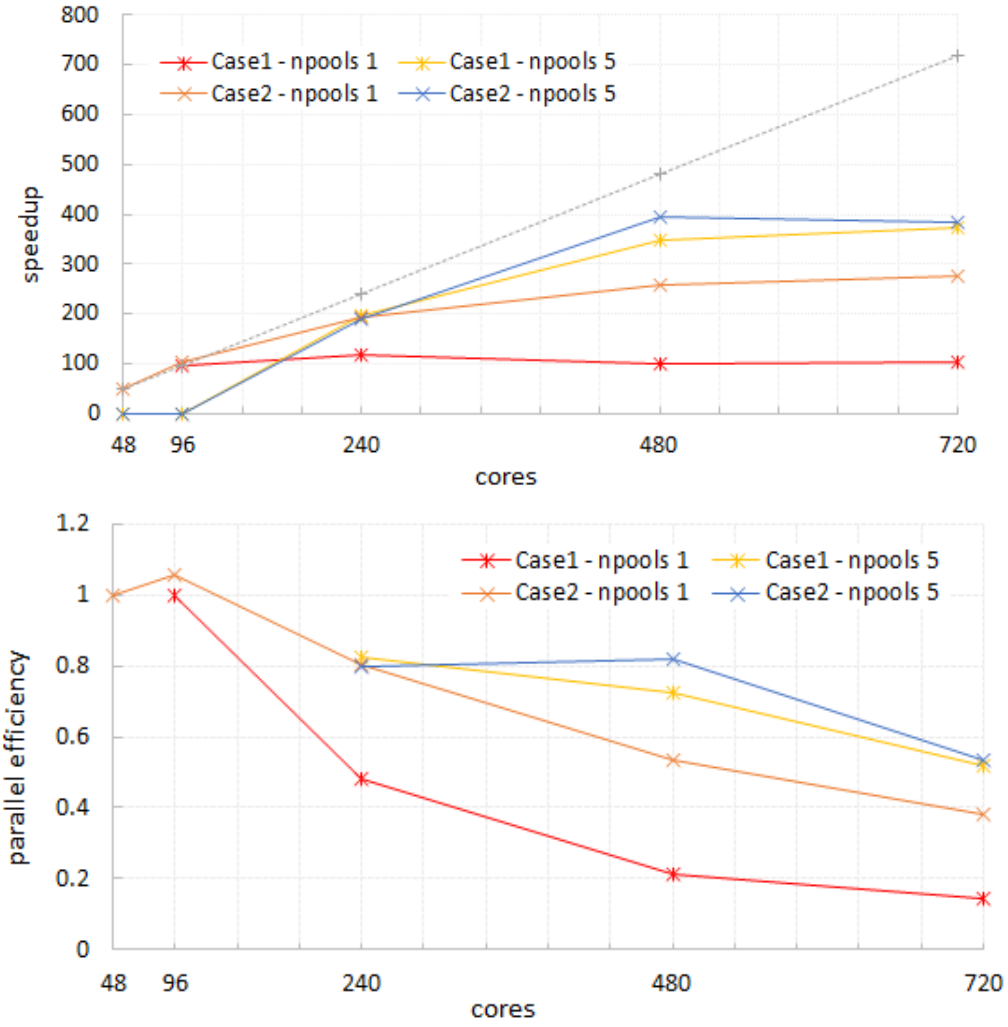


Fig. 6: Relative speedup and parallelization efficiency as a function of cores.

5 Conclusions

In conclusion, we assessed the most stable adsorption configuration of Hydrogen and Chlorine on (111) and (110) surfaces exposed by <112> Si NW. Hence, we investigated fully H- and Cl-passivated Si NWs of different diameter calculating the electronic structure and the formation energy. In the next phase of the project the H- and Cl-functionalized Si NWs will be used to investigate the adsorption of alkane and aliphatic chains. This research will allow to design efficient procedures to increase the overall efficiency of the devices based on Si NWs**.**

Acknowledgements

We acknowledge financial funding from the SUPER project in the Framework Programme Azione 1.5.1, POR-FESR 2014-2020 Emilia Romagna. SUPER (Supercomputing Unified Platform - Emilia-Romagna) is a regional project funded with the aim to extend the regional computing and storage infrastructure to support both scientific and industrial use cases.

References

[1] M. Bashouti et al. Covalent Attachment of Alkyl Functionality to 50 nm Silicon Nanowires through a Chlorination/Alkylation Process. *J. Phys. Chem. C* 113 (2009) pp. 14823–14828.

[2] Lu, N. et al. The structure of ultrathin H-passivated [112] silicon nanowires. *J. Phys. Chem. C* 111 (2007) pp. 7933–7937.

[3] A. Shalabny, F. Buonocore, M. Celino, G. Shalev, L. Zhang, W. Wu, P. Li, J. Arbiol, and M. Y. Bashouti. Semiconductivity Transition in Silicon Nanowires by Hole Transport Layer. *Nano Lett.* 20 (2020) pp. 8369–8374.
[4] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, et al. QUANTUM ESPRESSO: A Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys. Condens. Matter* **21**, p. 395502, (2009).
[5] F. Iannone, *et al*. CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 *International Conference on High Performance Computing & Simulation (HPCS)*, Dublin, Ireland, pp. 1051-1052 (2019).
[6] J.P. Perdew, K. Burke and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, pp. 3865−3868, (1996).
[7] D. Vanderbilt. Soft Self-Consistent Pseudopotentials in a Generalized Eigenvalue Formalism. *Phys. Rev. B: Condens. Matter Mater. Phys.* **41**, p. 7892, (1990).

AB INITIO STUDY OF NA INTERCALATION AT 2D-MOS2/GRAPHENE HETEROSTRUCTURE AS NEGATIVE ELECTRODE FOR NA BATTERIES

Arianna Massaro[1], Adriana Pecoraro[1], Ana B. Muñoz García[1] and Michele Pavone[2*]

[1]*University of Naples "Federico II", Department of Physics "Ettore Pancini", Comp. Univ. Monte Sant'Angelo Via Cintia 21, 80126, Naples, Italy*

[2]*University of Naples "Federico II", Department of Chemical Sciences, Comp. Univ. Monte Sant'Angelo Via Cintia 21, 80126, Naples, Italy*

ABSTRACT. We provide the first-principles investigation of Na intercalation and migration processes at 2D-MoS_2/Graphene heterostructures proposed as promising anode materials for Na-ion batteries (NIB). Structural and electronic features of 1T and 3R phases of MoS_2 monolayers interfaced to graphene are described at the DFT-D level of theory. Na-intercalated compounds act as equilibrium structures that define different migration pathways. We predict the 1T-MoS_2/Graphene material to be more active toward Na intercalation, while the 3R- counterpart allows easier diffusion thanks to peculiar structural features along the migration process.

1 Introduction

The research carried out at University of Naples "Federico II" focused on the ab initio simulation of Na intercalation and migration processes at the 2D-MoS_2/Graphene (MoS_2/G) heterostructures, where we consider both 1T and 3R phases as possible MoS_2 monolayers. Ion intercalation represents one of the main processes in a NIB device, thus understanding the detailed mechanism is crucial to boost the development of electrode materials that are expected to be good Na-hosts. MoS_2 has shown good performances as anode material in Na cells, and the combination with graphene layers has been proposed to improve both electric and mechanical properties [1]. The physico-chemical characterization of this hybrid interface is required in order to assess the role of the two materials when interacting with the Na ion. To this end, we report the DFT-D investigation of structural and electronic features occurring at the 2D interface and upon Na intercalation and migration, so as to shed light on the underlying mechanism and provide rational strategies for the design of efficient Na-hosts.

2 Methods and Computational details

We perform spin-polarized DFT calculations with projector- augmented wave (PAW) potentials and plane-wave (PW) basis set, as implemented in the Vienna Ab-initio Simulation Package (VASP) code (ver. 5.4.4). We use the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional and add the D3-BJ dispersion correction to account for van der Waals (vdW) interactions at the interface [2]. A kinetic energy cutoff of 600 eV and a 4×4×1 Γ-centered k-points sampling mesh are required to converge the PW basis set. Band structures are calculated self-consistently at the PBE-D3BJ level of theory along the Γ-K-M-Γ direction in reciprocal space by considering 10 k-points divisions along each segment [3]. Dipole corrections are applied to avoid long-range polarization from the periodic images along the c-direction. The CI-NEB method is used to determine minimum-energy paths (MEPs) of Na migration and the corresponding energy barrier heights [4].

3 Results and discussion

We obtain the 2D-MoS_2 layers by cleaving one single layer from the 1T- and 3R- bulk structures optimized at PBE-D3BJ level of theory. To combine the 1T- and 3R-MoS_2 monolayers with Graphene,

*Corresponding author. E-mail: michele.pavone@unina.it

we need to build up a 3x3 MoS_2 supercell and a 4x4 Graphene supercell in order to match the periodicity of both crystalline lattices. To quantify the strain between the two materials, we calculate the lattice mismatch according to Equation 1:

$$\varepsilon\,\% = \left(\frac{a^{hybrid}}{a^{pristine}} - 1\right) \cdot 100 \quad (1)$$

where $a^{pristine}$ is the lattice constant of the pristine material (*e.g.*, monolayered-MoS_2 and Graphene at their minima) and a^{hybrid} is the lattice constant of the heterostructure. We compute binding and adhesion energy variations to describe the thermodynamic stability of such hybrid materials:

$$\Delta E_{bind} = E_{MoS_2/G} - E_{MoS_2} - E_G \quad (2)$$

$$\Delta E_{adh} = E_{MoS_2/G} - E_{MoS_2}^{MoS_2/G} - E_G^{MoS_2/G} \quad (3)$$

where $E_{MoS_2/G}$ is the total energy of the hybrid material, E_{MoS_2} and E_G are the total energies of pristine MoS_2 and Graphene monolayers, $E_{MoS_2}^{MoS_2/G}$ and $E_G^{MoS_2/G}$ are the total energies of the pristine materials at the geometry of the heterostructure. Table 1 collects all these results.

Table 1: Lattice constant (a), interlayer distance (d), lattice mismatch (ε), binding energy (ΔE_{bind}) and adhesion energy (ΔE_{adh}) of both 1T-MoS_2/G and 3R-MoS_2/G heterostructures computed at PBE-D3BJ level of theory according to Equations 1-3.

phase	a (Å)	d (Å)	ε / MoS_2	ε / G	ΔE_{bind} (eV)	ΔE_{adh} (eV)
1T-MoS_2/G	9.810	3.110	2.17%	-0.71%	-0.665	-1.649
3R-MoS_2/G	9.758	3.350	3.26%	-1.25%	-0.026	-1.653

1T-MoS_2/G shows more negative binding energy compared to 3R-MoS_2/G, which can be related to the smaller lattice mismatch. Both phases can be combined to Graphene in thermodynamically stable hybrid materials without affecting Graphene structural properties. Electronic features of these heterostructures are investigated by means of band structures, projected density of states (PDOS) and charge density difference plots at the PBE-D3BJ level of theory, as depicted by Fig.1. Our calculations show that both heterostructures are electronically conductive and that a charge transfer occurs from Graphene to MoS_2, which is therefore expected to play the active role upon Na intercalation. We model Na intercalation at the MoS_2/G interface by considering a Na atom placed within the interlayer spacing in the 3x3MoS_2/4x4G supercell in several configurations. Sodium intercalation energy can be computed as:

$$\Delta E_{int} = E_{Na@host} - E_{host} - \frac{1}{2}E_{Na} \quad (4)$$

where $E_{Na@host}$ and E_{host} are the total energies of, respectively, Na-intercalated systems and the host materials, and E_{Na} is the total energy of Na metal. The results listed in Table 2 show that the overall intercalation energies are negative, suggesting that the formation of Na-intercalated compounds would be favourable for both phases. However, 1T- phase seems to be more active toward Na intercalation showing more negative ΔE_{int} values, which would explain the experimental evidence of phase stability variation upon ion intercalation.

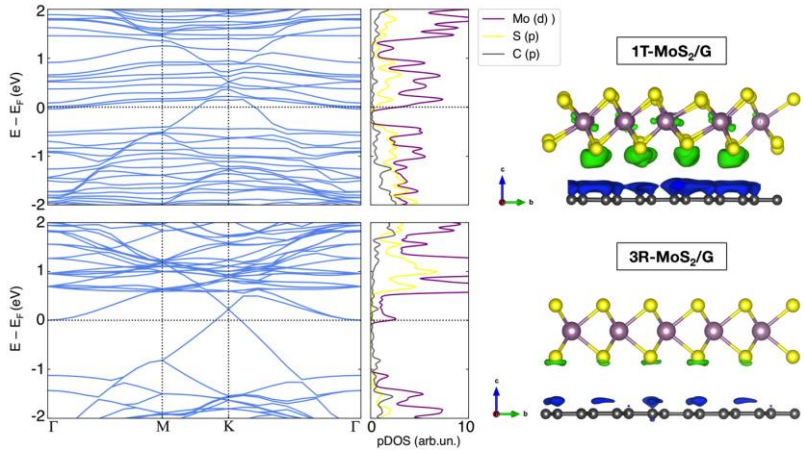


Fig. 1: Band structure, PDOS and charge density difference plot of 1T-MoS_2/G and 3R-MoS_2/G heterostructures (isosurface level: 2.7 meV/Å^3) computed at PBE-D3BJ level of theory. Colour code: Mo violet, S yellow, C black; electron gain (blue surface) and electron loss (green surface).

Table 2: Sodium intercalation energies, ΔE_{int}, at different intercalation sites computed at PBE-D3BJ level of theory according to Equation 4. Reported values are in eV.

phase	top-hollow	hollow-hollow	top-edge	hollow-edge
1T-MoS_2/G	-1.222	-0.914	-1.105	-1.150
3R-MoS_2/G	-0.162	-0.131	-0.089	-0.065

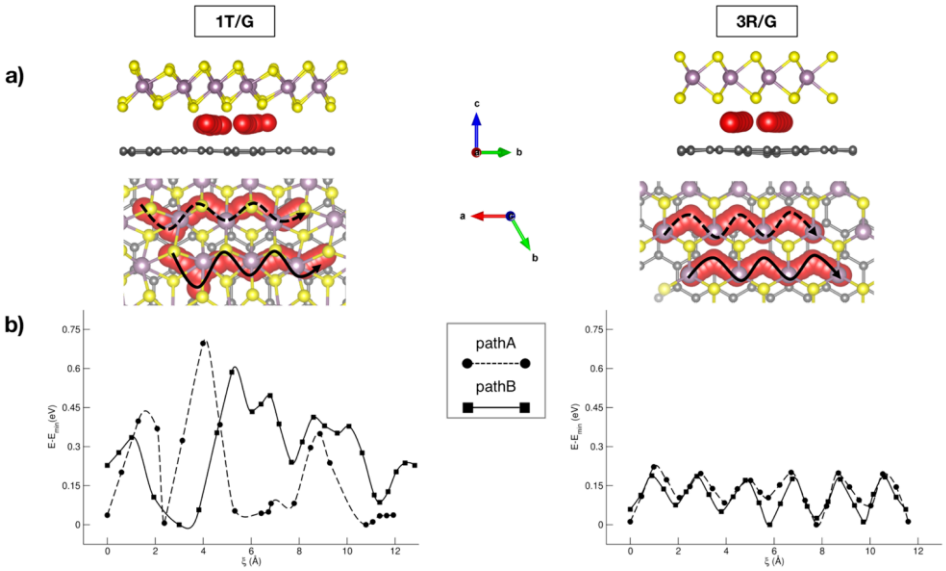


Fig. 2: a) Side and top views of Na migration pathways through **(left)** 1T-MoS_2/G and **(right)** 3R-MoS_2/G interfaces and **b)** corresponding energetics computed with CI-NEB method at PBE-D3BJ level of theory.

For the migration model, we identify two different pathways for each heterostructure, labelled A and B. In each path, Na atom can move along the *x* direction, where the S atoms pointing toward the interlayer spacing seem to create a proper diffusion channel. We consider all the intercalation sites lying on a given channel and applied the CI-NEB method to determine the minimum-energy path for Na migration (Fig. 2). In 3R-MoS_2/G, paths A and B (dashed and solid lines, respectively) result to be almost similar, with all the intermediate structures lying very close in energy. All the barriers associated to the Na jump are in the range 0.17-0.22 eV. Contrarily, migration through the 1T-MoS_2/G interface is much less regular

for both pathways, with larger energy differences among the explored sites. Path A (dashed lines) shows migration barriers going from ~0.08 to ~0.70 eV, while ~0.34 and ~0.59 eV values are found for path B (solid lines). So as to explain the different behavior among our explored pathways, we focus on the structural features of such intermediates along the migration coordinate. We look at the local coordination of the migrating Na cation and find out that the single jump is more likely to occur (*i.e.*, shows a lower energy barrier) when Na is furthest from S atoms (as a matter of fact, the hypothetic top-S intercalation site does not represent a stable configuration). Migration barriers generally arise from the shortening of two Na-S distances below the equilibrium value. In particular, this is more relevant in the case of 1T-MoS_2/G, where the Na-S distance becomes even shorter and leads to ΔE_{migr} equals to 0.696 eV and 0.586 eV for paths A and B, respectively. From these results, we would expect that 3R-MoS_2/G would allow an easier Na diffusion than its 1T- counterpart. However, considering the more favourable intercalation energies of the 1T-MoS_2/G heterostructure, we think that some strategies (*e.g.*, S vacancies, grain boundaries and/or defect engineering) aiming at lowering the migration barrier at this interface should be developed in order to boost its application as promising NIB electrode.

4 Conclusions

Single layers of 1T- and 3R-MoS_2 can be combined to Graphene into stable hybrid materials. Both heterostructures show a moderate charge transfer occurring from Graphene to MoS_2 layer and a metallic character in their band structures, thus proving that interfacial effects can completely modify the properties of pristine materials. Stable Na-intercalated compounds can be formed, with 1T- showing more favourable Na intercalation energies than 3R. MoS_2 turns out to be the active material upon Na intercalation since it is the one gaining the electronic charge from Na atom. Analysis of the diffusion mechanism shows that Na migration could be hindered by S atoms pointing towards the interlayer spacing, thus leading to higher migration barriers in 1T-MoS_2/G when Na-S distances along the path become too small. These results are subject of a scientific paper that has been recently published on Journal of Physical Chemistry C [5]. The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff; CRESCO/ ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes. See: http://www.cresco.enea.it/english for information [6].

References

[1] L. David, *et al.*, MoS_2/Graphene Composite Paper for Sodium-Ion Battery Electrodes. *ACS Nano*, **8** (2), pp 1759-1770 (2014).

[2] S. Grimme, *et al.*, A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.*, **132** (15), p 154104 (2010).

[3] A. Pecoraro, *et al.*, Structural and Electronic Properties of Defective 2D Transition Metal Dichalcogenide Heterostructures. *J. Comput. Chem.*, **41** (22), pp 1946-1955 (2020).

[4] G. Henkelman, *et al.*, A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.*, **22**, p 113 (2000).

[5] A. Massaro, *et al.*, First-Principles Study of Na Intercalation and Diffusion Mechanisms at 2D MoS_2/Graphene Interfaces. *J. Phys. Chem. C*, **125**, pp 2276-2286 (2021).

[6] G. Ponti, *et al.*, The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure, *Int. Conf. on High Perform. Comp. & Simul. (HPCS)* pp 1030-1033 (2014).

TRITON X-100 MICELLES IN AQUEOUS SOLUTION: SPHERE TO ROD LIKE MICELLAR SHAPE TRANSITION

Antonio De Nicola[1*], Antonio Pizzirusso[1], Stefano Caputo[1] and Yutaka Oya[2]

[1]*Dipartimento di Chimica e Biologia "A. Zambelli", Università di Salerno, Fisciano 84084, Italy*

[2]*Department of Organic Material Science, Yamagata University, Yonezawa 992-8510, Japan*

ABSTRACT. We present a computational investigation of the shape transitions mechanism of Triton X-100 micelles in an aqueous solution. Due to its peculiar structure, the non-ionic surfactant Triton X-100 forms aggregates that are not exclusively spherical. It has been observed, experimentally, that Triton X-100 has a strong tendency to form spherical aggregates at low surfactant concentration, while, increasing the concentration the shape changes from spherical to non-spherical, mainly prolate. We propose a simple theoretical model to calculate the interfacial free energy of Triton aggregates by using structural data derived from simulations.

1 Introduction

The Triton X-100 (TX-100) is a nonionic surfactant widely used in many biological applications such as purification of integral membrane protein,[1]–[5] and membrane solubilization.[6]–[10] The massive employ of TX-100 in many biological applications is due to its peculiar structure. Differently from most surfactants, the longest molecular fragment of TX-100 is characterized by a hydrophilic tail, while the shorter fragment is formed by a hydrophobic head. TX-100 consists of a hydrophobic core, 4-(1,1,3,3-tetramethylbutyl)-phenyl, bonded to a hydrophilic chain of polyethylene glycol (PEG), where the number of PEG units is between nine and ten. Various experimental techniques like Dynamic Light Scattering (DLS), X-Ray scattering, and NOESY Nuclear Magnetic Resonance (NMR) are used to gain more information about the size and shapes of TX-100 micellar aggregates. However, determining the shape of that aggregates is still challenging. Many computational studies investigating the shape transition of surfactant aggregates have been reported in the literature.[11]–[17] From all these studies it is possible to speculate about the behavior of TX-100 micelles, however, to the best of our knowledge, the shape transition of TX-100 micelles has been not investigated yet with a specific model. In our previous computational study we developed a CG model of TX-100 for the hybrid particle-field (hPF) approach, which is able to reproduce phase behavior in the whole TX-100/water phase diagram.[18] Our particular intent is to present a detailed investigation of the shape transition of TX-100 aggregates, with varying the aggregation number, looking at the interfacial free energy of the aggregates. To this aim, we developed a simple theoretical model to compute the interfacial free energy, based on structural data of the aggregates achievable from simulations. hPF MD simulations of TX-100 micelles in water are performed and analyzed to estimate parameters needed by theoretical model and to compute the interfacial energy profiles for the different micelle shapes.

2 TX-100 Model

In the present study we adopted the hPF MD technique, introduced by Milano and Kawakatsu in 2009.[19], [20] The hPF CG model of TX-100 has been taken from our previous papers.[21], [18], [22] Each CG bead represents, on average, four heavy atoms, while solvent bead W is representative of four atomistic water molecules. The intramolecular interactions such as bond and angle interactions are described by using harmonic potentials, while, according to the hPF treatment, the intermolecular interactions are calculated by using an external potential based on the density field.

*Corresponding author. E-mail: adenicola@unisa.it

3 Results

We present a simple theoretical model designed to predict the optimal micelle shape, in aqueous solution, formed by TX-100 molecules. The theoretical model proposed here contains elements used in previous approaches but is not directly comparable to any of them because its most important feature is the quantitative estimation of interfacial free energy as function of the N_{agg}, a problem which is partially treated by previous theoretical treatments.

To connect the geometry of micellar aggregate with its interfacial free energy the following assumptions are needed. Especially, we assume that: (i) TX-100 aggregates can adopt only defined geometries, i.e. spherical or spherocylindrical, which are the most probable shapes as emerge from experimental data[7], [23], [24] (see the scheme in Figure 1 left panel). (ii) The aggregate hydrophobic core is composed only of hydrophobic particle types (R, C) belonging the TX-100 molecule. (iii) The shape transition originates from the free energy contribution depending only on the interfacial region (I). (iv) Hydrophobic and hydrophilic (PEG) molecular segments are mixed with the water in I. According to our assumptions, we can define a general expression, valid for both shapes, to calculate the volume containing only the hydrophobic particles of an aggregate (V_H) with the expression:

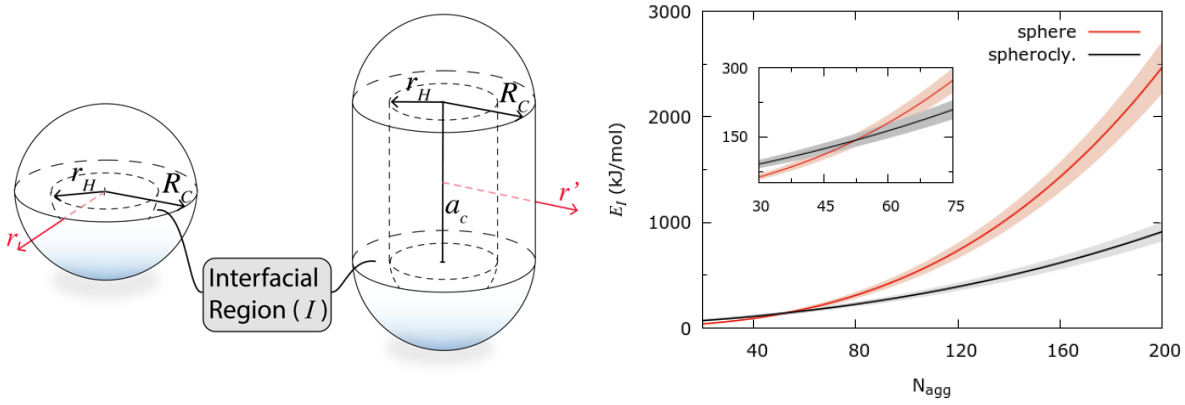


Fig. 1. (Left panel) Schematized 3D geometry for spherical and spherocylindrical (or capsule) TX-100 aggregate. The r_H and R_C are the radii of hydrophobic core and interfacial core, while a_c is the length of the cylindrical section in case of capsule shape. (Right panel) Interfacial energy E_I as function of N_{agg} calculated for both, spherical (red line), and spherocylindrical shape (black line). The N_{agg} range where the two E_I curves crosses each other is reported in the inset. The area around each curve represents the error on E_I due to the errors from the fitting procedure.

$$V_H = \int \phi_H(\boldsymbol{r}) d^3r = v\, N_{agg} \quad (1)$$

It can be demonstrated that the interfacial energy can be calculated with the equations 2 and 3[25].

$$I_{spherocylinder}\left(a_c, \phi_H(r), \phi_E(r)\right) = \int_{r_H}^{R_C} \phi_H(r)\phi_E(r) 4\pi r^2 dr + \int_{r_H}^{R_C} \phi_H(r')\phi_E(r') 2\phi r' a_c dr' \quad (2)$$

$$I_{sperhocylinder} = \frac{1}{15}\pi\left(R_C^3 - r_H^3 + R_C[R_C^2 - r_H^2]\right) + \frac{1}{30}\pi a_C(R_C^2 - r_H^2 + R_C[R_C - r_H]) \quad (3)$$

In the right panel of Figure 1, according to our theoretical model the behavior of E_I vs N_{agg} is reported. The trends of E_I for sphere and spherocylinder show a preference for the spherical shape at low N_{agg} ~20-40, while in the region N_{agg}~ 40 – 70 a coexistence between the two shapes is found (see the inset of Figure 1 right panel). Instead, for N_{agg} larger than ~70 the most favorable shape is the spherocylinder.

The emerging picture from the E_I profile agrees with the conjecture of Paradies[24] and Nemethy,[26] which claimed the non-sphericity of TX-100 micelles for large N_{agg}.

References

[1] D. A. Brown and E. London, "Functions of Lipid Rafts In Biological Membranes," Annual Review of Cell and Developmental Biology, vol. 14, no. 1, pp. 111–136, Nov. 1998, doi: 10.1146/annurev.cellbio.14.1.111.

[2] B. R. Casadei, C. C. Domingues, E. de Paula, and K. A. Riske, "Direct Visualization of the Action of Triton X-100 on Giant Vesicles of Erythrocyte Membrane Lipids," Biophysical Journal, vol. 106, no. 11, pp. 2417–2425, 2014, doi: 10.1016/j.bpj.2014.04.039.

[3] J. Lasch et al., "Interaction of Triton X-100 and octyl glucoside with liposomal membranes at sublytic and lytic concentrations. Spectroscopic studies," Biochimica et Biophysica Acta (BBA) - Biomembranes, vol. 1022, no. 2, pp. 171–180, 1990, doi: 10.1016/0005-2736(90)90111-Z.

[4] M. le Maire, P. Champeil, and J. V. Møller, "Interaction of membrane proteins and lipids with solubilizing detergents," Biochimica et Biophysica Acta (BBA) - Biomembranes, vol. 1508, no. 1, pp. 86–111, 2000, doi: 10.1016/S0304-4157(00)00010-1.

[5] D. Lichtenberg, H. Ahyayauch, and F. M. Goñi, "The Mechanism of Detergent Solubilization of Lipid Bilayers," Biophysical Journal, vol. 105, no. 2, pp. 289–299, 2013, doi: 10.1016/j.bpj.2013.06.007.

[6] M. N. Jones, "Surfactants in membrane solubilisation," International Journal of Pharmaceutics, vol. 177, no. 2, pp. 137–159, Jan. 1999, doi: 10.1016/S0378-5173(98)00345-7.

[7] E. A. Dennis, "Formation and characterization of mixed micelles of the nonionic surfactant Triton X-100 with egg, dipalmitoyl, and dimyristoyl phosphatidylcholines," Archives of Biochemistry and Biophysics, vol. 165, no. 2, pp. 764–773, 1974, doi: 10.1016/0003-9861(74)90305-1.

[8] C. Arnulphi, J. Sot, M. García-Pacios, J.-L. R. Arrondo, A. Alonso, and F. M. Goñi, "Triton X-100 Partitioning into Sphingomyelin Bilayers at Subsolubilizing Detergent Concentrations: Effect of Lipid Phase and a Comparison with Dipalmitoylphosphatidylcholine," Biophysical Journal, vol. 93, no. 10, pp. 3504–3514, 2007, doi: 10.1529/biophysj.107.104463.

[9] D. Lichtenberg, R. J. Robson, and E. A. Dennis, "Solubilization of phospholipids by detergents structural and kinetic aspects," Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes, vol. 737, no. 2, pp. 285–304, 1983, doi: 10.1016/0304-4157(83)90004-7.

[10] D. Lévy, A. Bluzat, M. Seigneuret, and J.-L. Rigaud, "A systematic study of liposome and proteoliposome reconstitution involving Bio-Bead-mediated Triton X-100 removal," Biochimica et Biophysica Acta (BBA) - Biomembranes, vol. 1025, no. 2, pp. 179–190, 1990, doi: 10.1016/0005-2736(90)90096-7.

[11] J. Gao, "Effect of concentration on surfactant micelle shapes—A molecular dynamics study," Sci China Ser B, vol. 48, no. 5, p. 470, 2005, doi: 10.1360/042004-71.

[12] S. Fujiwara, T. Itoh, M. Hashimoto, and R. Horiuchi, "Molecular dynamics simulation of amphiphilic molecules in solution: Micelle formation and dynamic coexistence," The Journal of Chemical Physics, vol. 130, no. 14, p. 144901, Apr. 2009, doi: 10.1063/1.3105341.

[13] M. Velinova, D. Sengupta, A. V. Tadjer, and S.-J. Marrink, "Sphere-to-Rod Transitions of Nonionic Surfactant Micelles in Aqueous Solution Modeled by Molecular Dynamics Simulations," Langmuir, vol. 27, no. 23, pp. 14071–14077, Dec. 2011, doi: 10.1021/la203055t.

[14] A. V. Sangwai and R. Sureshkumar, "Coarse-Grained Molecular Dynamics Simulations of the Sphere to Rod Transition in Surfactant Micelles," Langmuir, vol. 27, no. 11, pp. 6628–6638, Jun. 2011, doi: 10.1021/la2006315.

[15] A. G. Daful, J. B. Avalos, and A. D. Mackie, "Model shape transitions of micelles: Spheres to cylinders and disks," Langmuir, vol. 28, no. 8, pp. 3730–3743, 2012, doi: 10.1021/la204132c.

[16] S. Fujiwara, M. Hashimoto, T. Itoh, and R. Horiuchi, "Micellar Shape Change in Amphiphilic Solution: A Molecular Dynamics Study," Chem. Lett., vol. 41, no. 10, pp. 1038–1040, Oct. 2012, doi: 10.1246/cl.2012.1038.

[17]. K. Schäfer et al., "Supramolecular Packing Drives Morphological Transitions of Charged Surfactant Micelles," Angew. Chem. Int. Ed., vol. 59, no. 42, pp. 18591–18598, Oct. 2020, doi: 10.1002/anie.202004522.

[18] De Nicola, Antonio, T. Kawakatsu, C. Rosano, M. Celino, M. Rocco, and G. Milano, “Self-Assembly of Triton X-100 in Water Solutions: A Multiscale Simulation Study Linking Mesoscale to Atomistic Models,” J. Chem. Theory Comput., p. 13, 2015, doi: 10.1021/acs.jctc.5b00485.
[19] G. Milano and T. Kawakatsu, “Hybrid particle-field molecular dynamics simulations for dense polymer systems,” The Journal of Chemical Physics, vol. 130, no. 21, p. 214106, Jun. 2009, doi: 10.1063/1.3142103.
[20] G. Milano and T. Kawakatsu, “Pressure calculation in hybrid particle-field simulations,” The Journal of Chemical Physics, vol. 133, no. 21, p. 214102, Dec. 2010, doi: 10.1063/1.3506776.
[21] A. De Nicola, T. Kawakatsu, and G. Milano, “A Hybrid ParticleField CoarseGrained Molecular Model for Pluronics Water Mixtures,” Macromol. Chem. Phys., p. 11, 2013.
[22] A. Pizzirusso et al., “Biomembrane solubilization mechanism by Triton X-100: a computational study of the three stage model,” Phys. Chem. Chem. Phys., vol. 19, no. 44, pp. 29780–29794, 2017, doi: 10.1039/C7CP03871B.
[23] R. J. Robson and E. A. Dennis, “The size, shape, and hydration of nonionic surfactant micelles. Triton X-100,” Journal of Physical Chemistry, vol. 81, no. 11, pp. 1075–1078, 1977, doi: 10.1021/j100526a010.
[24] H. Hasko Paradies, “Shape and size of a nonionic surfactant micelle. Triton X-100 in aqueous solution,” The Journal of Physical Chemistry, vol. 84, no. 6, pp. 599–607, 1980, doi: 10.1021/j100443a008.
[25] W. Murakami et al., “Theoretical and Computational Study of the Sphere-to-Rod Transition of Triton X-100 Micellar Nanoscale Aggregates in Aqueous Solution: Implications for Membrane Protein Purification and Membrane Solubilization,” ACS Appl. Nano Mater., vol. 4, no. 5, pp. 4552–4561, May 2021, doi: 10.1021/acsanm.1c00171.
[26] F. Podo, A. Ray, and G. Nemethy, “Structure and Hydration of Nonionic Detergent Micelles. High Resolution Nuclear Magnetic Resonance Study,” Journal of the American Chemical Society, vol. 95, no. 19, pp. 6164–6171, 1973, doi: 10.1021/ja00800a003.

BRINGING AI PIPELINES ONTO CLOUD-HPC: SETTING A BASELINE FOR ACCURACY OF COVID-19 AI DIAGNOSIS

Iacopo Colonnelli[1*], Barbara Cantalupo[1], Concetto Spampinato[2], Matteo Pennisi[2] and Marco Aldinucci[1]

[1]*University of Torino, Computer Science Dept., Corso Svizzera 185, 10149, Torino, Italy*

[2]*University of Catania, Electrical Engineering Dept., Viale Andrea Doria 6, 95125, Catania, Italy*

ABSTRACT. HPC is an enabling platform for AI. The introduction of AI workloads in the HPC applications basket has non-trivial consequences both on the way of designing AI applications and on the way of providing HPC computing. This is the leitmotif of the convergence between HPC and AI. The formalized definition of AI pipelines is one of the milestones of HPC-AI convergence. If well conducted, it allows, on the one hand, to obtain portable and scalable applications. On the other hand, it is crucial for the reproducibility of scientific pipelines. In this work, we advocate the StreamFlow Workflow Management System as a crucial ingredient to define a parametric pipeline, called "CLAIRE COVID-19 Universal Pipeline", which is able to explore the optimization space of methods to classify COVID-19 lung lesions from CT scans, compare them for accuracy, and therefore set a performance baseline. The universal pipeline automatizes the training of many different Deep Neural Networks (DNNs) and many different hyperparameters. It, therefore, requires a massive computing power, which is found in traditional HPC infrastructure thanks to the portability-by-design of pipelines designed with StreamFlow. Using the universal pipeline, we identified a DNN reaching over 90% accuracy in detecting COVID-19 lesions in CT scans.

1 Introduction

The ability of AI-related techniques to transform raw data into valuable knowledge is growing at a breakneck pace. Among these techniques, Deep Learning (DL) has benefited from crucial results in Machine Learning (ML) theory and the large availability of data. The accuracy of the process is strictly related to the quality and size of available datasets. Usually, more data means more accurate predictions, but also more computing power needed to train the model.

In particular, in the last decade, Deep Neural Networks (DNNs) became larger and larger, and nowadays, a reasonably sized DL workload cannot prescind from the availability of heterogeneous (GPU-equipped) computing resources. For this reason, High-Performance Computing (HPC) is undoubtedly an enabling platform for AI, and, in turn, AI enables success in scientific challenges where traditional HPC techniques have failed. For their part, supercomputers are shifting more and more to heterogeneous hardware, both because of their better energy efficiency and to satisfy the ever-increasing need for GPU-enabled workloads pushed by DL.

Despite this potential, supercomputers are rarely used for standard AI workloads. This is mainly due to technical barriers, i.e., user-unfriendly SSH-based remote shells and queue-based job submission mechanisms, which prevent AI researchers without a strong computer science background from effectively unlocking their computing power. In practice, HPC centres are not designed for general purpose applications. Only scalable and computationally demanding programs can effectively benefit from the massive amount of processing elements and the low-latency network interconnections that characterize HPC facilities, justifying the high development cost of HPC-enabled applications. Moreover, some seemingly trivial features are not supported by HPC facilities, e.g., exposing a public web interface for data visualization in an air-gapped worker node.

[*]Corresponding author. E-mail: iacopo.colonnelli@unito.it

On the other hand, the technical barriers to Cloud-based infrastructures lowered substantially with the advent of the *-as-a-Service model. Alas, the multiple layers of virtualization that characterize modern cloud architectures introduce significant processing overheads and make it impossible to apply adaptive fine-tuning techniques based upon the underlying hardware technologies, making them incompatible with performance-critical HPC applications.
In this work, we advocate the combination of two distinct approaches as an effective way to lower the technical barriers of HPC facilities for AI researchers [1]:

- A *cluster-as-accelerator* design pattern, in which cluster nodes act as processing elements of user-defined tasks sent by a Cloud-based, general-purpose host executor. This paradigm can be used to offload computation to HPC facilities in a more intuitive way, as it mimics the GPGPU paradigm used by DL applications in a GPU-equipped machine;
- *Hybrid workflows*, i.e. workflows whose steps can be scheduled on independent and potentially not intercommunicating execution environments, as a programming paradigm to express this design pattern while ensuring portability and reproducibility of complex AI workloads.

In the evaluation part, we apply these two principles to a real AI application, i.e., a DNN training pipeline for COVID19 diagnosis from Computed Tomography (CT) scan images. The StreamFlow [2] Workflow Management System (WMS), which supports hybrid Cloud-HPC workflows as first-class citizens, is used as the underlying runtime system.

2 The StreamFlow toolkit

A workflow is commonly represented as an acyclic digraph $G = (N, E)$, where nodes refer to different portions of a complex program and edges encode *dependency relations* between nodes. In this representation, a direct edge connecting a node m to a node n means that n must wait for m to complete before starting its computation.
The workflow abstraction has already been explored for offloading computation to HPC facilities transparently [3]. Many of the existing WMSs come with a diverse set of *connectors*, some of them addressing Cloud environments and some others more HPC-oriented. Nevertheless, a far smaller percentage can deal with *hybrid workflows*, offering the possibility to seamlessly assign each portion of a complex application to the computing infrastructure that best suits its requirements.
Hybrid workflows can strongly reduce the necessary tradeoffs in relying on such high-level abstraction, both in terms of performance and costs. Indeed, complex applications usually alternate computation-intensive and highly parallelizable steps with sequential or non-compute-bound operations. When scheduling such applications to an HPC centre, only a subset of steps will effectively take advantage of all the available computing power, resulting in a low cost-benefit ratio.
Moreover, HPC facilities do not support some common operations that are instead trivial on Cloud-based infrastructures, such as exposing web interfaces for data visualization.
The StreamFlow toolkit[1] [2], whose logical stack is depicted in Fig.1, has been specifically developed to orchestrate hybrid workflows on top of heterogeneous and geographically distributed architectures.
Written in Python 3, it supports the CWL coordination standard [4] for expressing workflow models through a declarative JSON or YAML syntax. The translation of these declarative semantics into an executable workflow model is delegated to `cwltool`, the CWL reference implementation. The StreamFlow runtime layer is then able to efficiently execute such a model by translating it into a dataflow graph, identifying independent steps and running them in parallel whenever possible.
Alongside, one or more execution environments can be described in well-known external formats, e.g., Helm charts for Kubernetes deployments or Slurm files for HPC workloads. A `streamflow.yml` file constitutes the entry point of a StreamFlow run, relating each workflow step with the best suitable execution environment. This feature actually plugs the hybrid layer in the workflow design process.
Another distinctive feature of the StreamFlow WMS is the possibility to manage complex, multi-agent execution environments, ensuring the *co-allocation* of multiple heterogeneous processing elements to

[1] https://streamflow.di.unito.it/

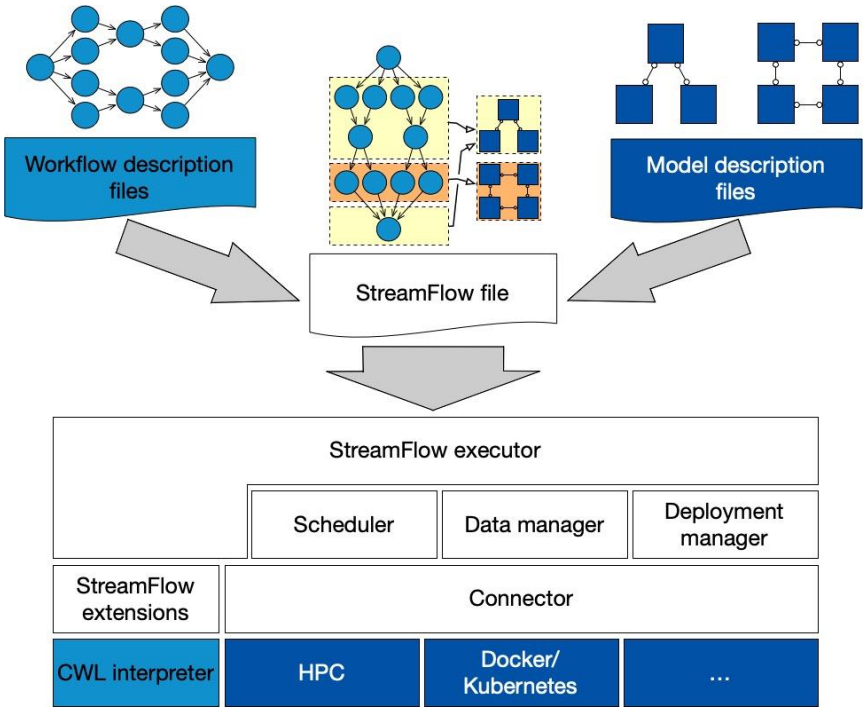


Fig.1: The StreamFlow toolkit's logical stack.

execute a single workflow step. The same interface can then be used to describe a diverse ecosystem of distributed applications, ranging from MPI clusters running on HPC facilities to microservices architectures deployed on Kubernetes. To provide enough flexibility, StreamFlow adopts a three-layered hierarchical representation of execution environments:

- A `model` is an entire multi-agent infrastructure and constitutes the *unit of deployment*, i.e., all its components are always co-allocated when executing a step;
- A `service` is a single agent in a model and constitutes the *unit of binding*, i.e., each step of a workflow can be offloaded to a single service for execution;
- A `resource` is a single instance of a potentially replicated service and constitutes the *unit of scheduling*, i.e., each step of a workflow is offloaded to a configurable number of service resources to be processed.

All communications and data transfer operations are started and managed by the StreamFlow controller, removing the need for bidirectional channels between the management infrastructure and the target resources and allowing tasks to be offloaded to HPC infrastructures with air-gapped worker nodes. Moreover, StreamFlow does not need any specific package or library to be installed on the target resources, other than the software dependencies required by the host application. As a consequence, virtually any target infrastructure reachable by a practitioner can serve as a target model, as long as a compatible connector implementation is available.

3 The CLAIRE COVID-19 universal pipeline

To demonstrate how StreamFlow can help bridge HPC and AI workloads, enabling reproducibility and portability across different platforms, we present the COVID-19 universal pipeline, developed by the Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE)[2] task force on AI & COVID-19 during the first COVID-19 outbreak. The group, composed of fifteen researchers in complementary disciplines (Radiomics, AI, and HPC) and led by Prof. Marco Aldinucci [5], investigated the diagnosis of COVID-19 pneumonia assisted by Artificial Intelligence (AI).

At the start of the pandemic, several studies outlined the effectiveness of chest radiology imaging for COVID-19 diagnosis. Even if X-Ray scans represent a cheaper and most effective solution for large

[2]https://claire-ai.org/

scale screening, their low resolution led AI models to show lower accuracy than those obtained with CT scans. Therefore, the latter has become the gold standard for the investigation of lung diseases.
Several research groups worldwide began to develop DL models for the diagnosis of COVID-19, mainly in the form of deep Convolutional Neural Networks (CNN). As is especially the case in the medical field, reproducibility of the results was an important issue to address. Providing AI pipelines for COVID diagnosis with reproducible steps should not be an option to ensure the goodness of the results. This is particularly important when dealing with DL models, which are obscure by definition.
Furthermore, such a significant number of proposals was not accompanied by any baseline of the accuracy expectation. A comprehensive study of the proposed solutions highlighted that it would not be easy to evaluate the most promising approaches due to the adoption of different architectures, pipelines and datasets. Therefore, instead of proposing yet another hopefully better solution, the task force commitment was to organize the knowledge so far to consolidate and formalize all or the most state-of-the-art deep learning models to diagnose COVID-19.
The result of such commitment was the distillation of a reproducible workflow, the *CLAIRE COVID-19 universal pipeline* represented in Fig.2, capable of automating the comparison of the proposed DL models and supporting the definition of a baseline for any further evaluation.
The pipeline is basically designed by composing the main steps in a standard AI workflow. TC scan images are pre-processed to insulate the region of interest, the lungs in this case, and then used to train a classifier to recognize the typical lesions of interstitial pneumonia caused by COVID-19, specifically consolidation, crazy paving and grown glass.

The pipeline is composed of two main parts:

- A *data preparation* phase (yellow elements), comprising *pre-processing*, where standard techniques for cleaning the training images are applied, and *segmentation*, for extracting and selecting the region of interest for the next training. This step is performed just once for each dataset;
- The *core training* phase (blue elements), composed of standard AI steps such as *data augmentation*, to generate image variants, model *pre-training*, to generate an initial set of weights for initialization, and eventually *classification*, which labels each image with a class identified with a kind of lesion that is typical of the disease. The final steps are *cross-validation*, which increases the pipeline robustness by applying the training on different portions of the dataset, and *performance metrics*, obtained by collecting and comparing all the measures from the different pipelines.

The effectiveness of the DL approach depends on many parameters, e.g., the input dataset, the pre-processing steps, the chosen DL model, and the hyperparameters of the training algorithm, such as learning rate, weight decay, learning rate decay. It is worth noting that, in the universal pipeline, the DNN itself, which is the model used for classification, is just one of the variables that can be set for training. Different variants of existing networks (such as AlexNet, ResNet, DenseNet) can be plugged into the pipeline, but any future network could be included in principle.
Finally, the pipeline is modelled as a workflow where every single step is a kind of container without any dependency on external libraries or vendor-specific technology. This choice enables portability on different platforms allowing to run the same pipeline in different platforms or even across different ones.
The first set of experiments has been already completed comparing about 1% of the variants (11 of 990) and applying different segmentation types. Results show that the pipeline can generate models with excellent accuracy in classifying typical interstitial pneumonia lesions due to COVID-19, with sensitivity and specificity metrics over 90% in the best cases. More detailed information on the experiment's results are available in [6] and demonstrate that the proposed approach is able to carry out the same task with an accuracy that is at least on par with, or even higher than, human experts, thus showing the potential impact that these techniques may have in supporting physicians in decision making.

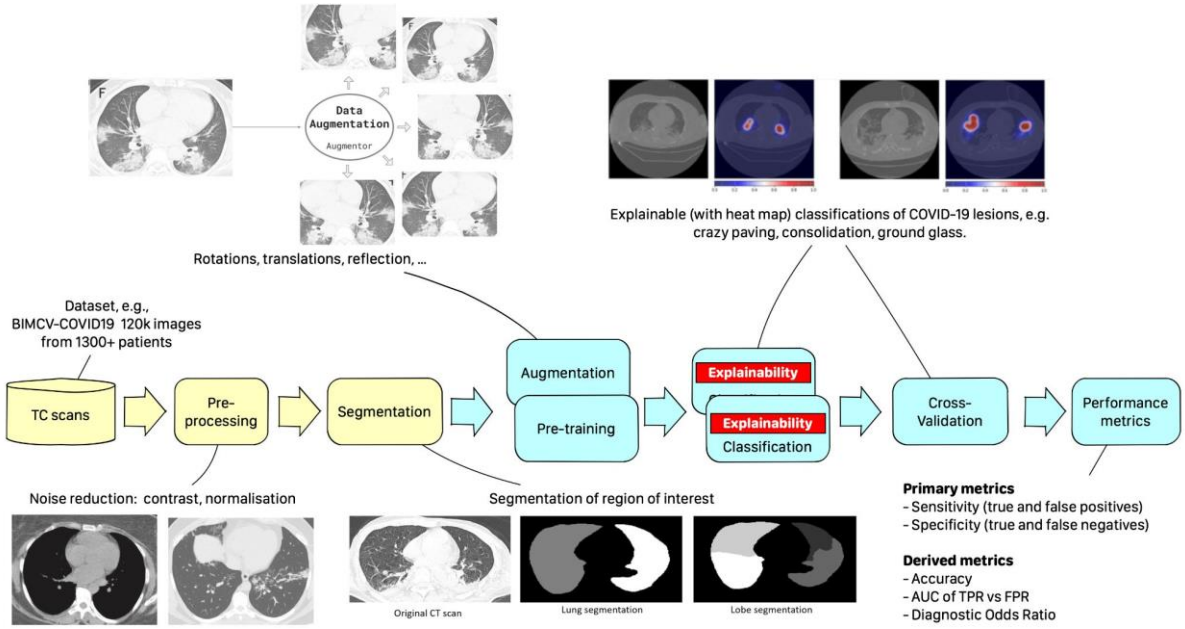


Fig. **2**: The CLAIRE COVID-19 universal pipeline.

4 Experimental evaluation

On the move from the design to the implementation, the universal pipeline takes advantage of the StreamFlow technology. The workflow is defined using CWL, using traditional parallel computing operators (such as scatter, broadcast, gather, reduce) to explicitly annotate the parallelizable portions of the pipeline (Fig. 3). The search space is composed of all the combinations of network models, training hyperparameters, and one or more datasets. Each point in such space is a tuple, which constitutes the input for a single pipeline instance. As each pipeline instance is independent of each other, the overall execution is a typical embarrassingly parallel problem, whose parallelism can be exploited by distributing the input tuples to the work units through the "scatter" operator.

In turn, every single pipeline instance can further exploit parallelism, distributing different partitions of the dataset (generated by the cross-validation algorithm) to as many instances of the classification step through another scatter operator. In this setting, the initial weights of the DNN are broadcasted to all the classifiers of a specific pipeline instance. Performance metrics are then collected through a combination of reduce operators, first reducing internally in the single pipeline instance and then globally collecting results from all the pipelines. The baseline performance of the analyzed pipelines for COVID-19 diagnosis is finally obtained.

For evaluation purposes, we ran the pipeline on the BIMCV-COVID19 dataset, with more than 120k images from 1300 patients. Assuming to train each pre-trained model for 20 epochs on such dataset, a single variant of the pipeline takes over 15 hours on a single NVidia V100 GPU, one of the most powerful accelerators in the market. Therefore, exploring all the 990 pipeline variants we have selected would take over two years. Fortunately, as we already pointed out, the universal pipeline has an embarrassingly parallel structure, and therefore using a supercomputer could reduce the execution time down to 15 hours in the best case (i.e., when 990 GPUs are available at the same time).

Post-training steps, as performance metrics extraction and comparison, are better suited for a Cloud infrastructure. Indeed, they do not require much computing power and can significantly benefit from web-based visualization tools. Given that, we used StreamFlow to model the pipeline as a hybrid workflow, offloading the training portions to HPC nodes and collecting the resulting networks on the host execution flow for visualization purposes. In particular, different portions of the training spectrum have been offloaded to three different heterogeneous architectures:

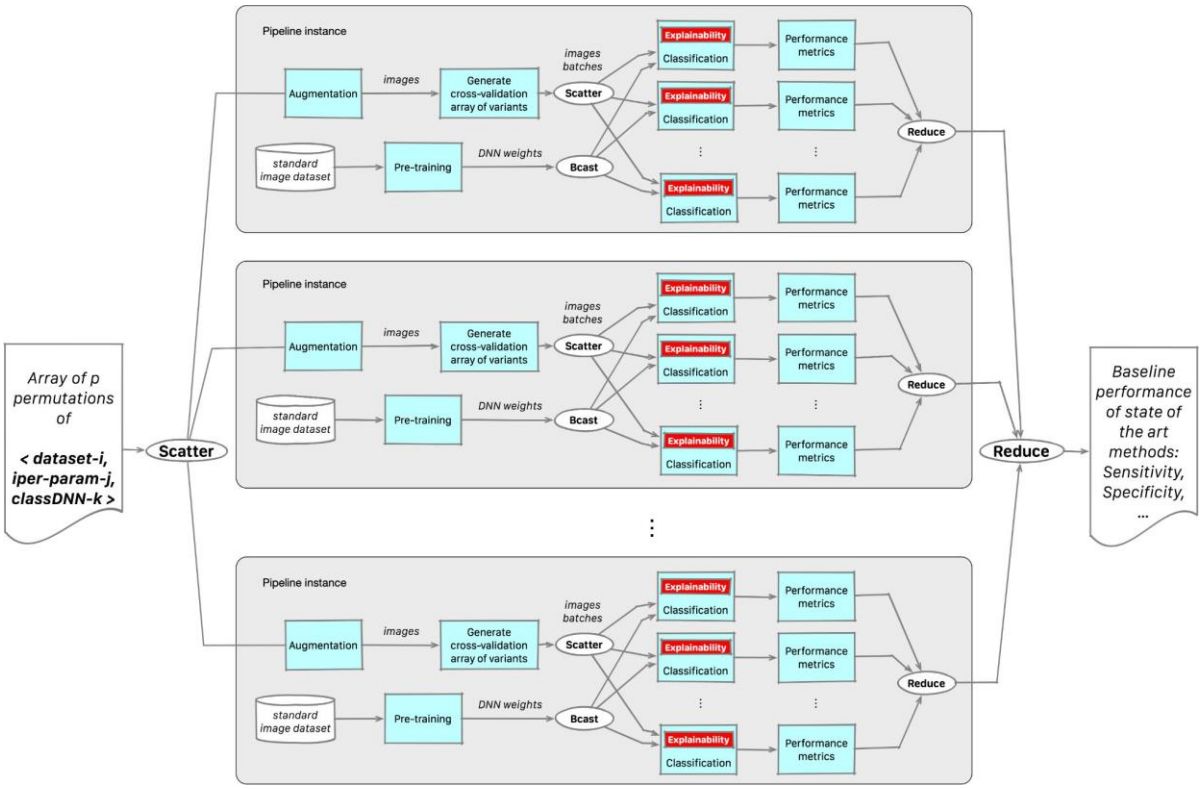


Fig. **3**: Unfolded implementation of the pipeline training components

- The ENEA CRESCO D.A.V.I.D.E. cluster, composed of 45 nodes with 2 IBM POWER8 sockets, 256 GB of RAM and 4 NVidia P100-SMX2 GPUs each, that can be used on-demand through bare SSH connections;
- The CINECA MARCONI100 cluster, a SLURM-managed HPC facility with 32 IBM POWER9 cores, 256 GB of RAM and 4 NVidia V100 GPUs per node;
- The High-Performance Computing for Artificial Intelligence (HPC4AI) infrastructure at the University of Torino, a multi-tenant hybrid Cloud-HPC system with 80 cores and 4 GPUs per node (T4 or V100-SMX2) managed by OpenStack [7].

As an interface towards Cloud-HPC infrastructures, StreamFlow seamlessly manages data movements and remote step execution with each of these infrastructures, automatically transferring back the training results to the Cloud-based host node to perform post-training steps.

5 Conclusion and future work

Presenting the work on the universal COVID-19 pipeline, we demonstrated that AI can be an effective support for human activity, that HPC is crucial to perform complex tasks in a useful time, and last but not least, that the convergence of different platforms is the next big thing. In this scenario, the general adoption of hybrid infrastructures from the scientific communities can only be obtained by leveraging advanced software tools like StreamFlow to enable portability and reproducibility.

The computing power required by the largest AI training runs has been increasing exponentially with a 3.4-month doubling time in the last 10 years [8]. A need that today can be only matched by way of accelerated computing provided by specialized processors such as GPUs and TPUs. With this background, the next era of HPC will inevitably see a further increase of heterogeneous hardware, with general-purpose CPUs flanked by highly parallel co-processors as GPUs and special-purpose hardware as neuromorphic chips and quantum annealing. Even if each architecture comes with its peculiar programming paradigm for local computations, the accelerator pattern is becoming a de-facto standard for moving computation away from CPUs.

We believe that a sound and stable system software part is crucial for the mainstream industrial adoption of HPC, enabling technology to transform applications into easily usable services hence into innovation.

While in scientific computing the modernization of HPC applications is a scientific desideratum required to boost industrial adoption, the shift toward the Cloud model of services is a must in AI. AI applications are already modern, and they will not step back.
We advocate StreamFlow as an intuitive programming paradigm to foster the design of portable and scalable AI pipelines and reduce technical barriers to HPC facilities for domain experts without a strong computer science background.
Such paradigms can be further extended and improved. From one side, support for specific hardware (e.g., quantum processors) can be added to the list of connectors offered by hybrid WMSs. Moreover, more intuitive and user-friendly technologies (e.g., Jupyter Notebook) can be augmented with hybrid workflow semantics to evolve them from prototyping technologies to production-ready toolchains. Both these challenges are essential parts of the StreamFlow roadmap, together with further applications in the domains of deep learning, bioinformatics and molecular dynamics simulations.

Funding & Acknowledgement

We gratefully acknowledge the support of Francesco Iannone from ENEA and the CRESCO/ENEAGRID High Performance Computing infrastructure and its staff. This work has been partially supported by the DeepHealth project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825111, by the HPC4AI project funded by Regione Piemonte (POR FESR 2014-20 - INFRA-P). Access to CINECA resources has been possible thanks to the CINI-CLAIRE-ABD MoU 2020 supporting research on COVID-19.
We want to thank Emanuela Girardi and Gianluca Bontempi, who are coordinating the CLAIRE task force on COVID-19, for their support, and the group of volunteer researchers who contributed to the development of the CLAIRE COVID-19 universal pipeline, they are: Marco Calandri and Piero Fariselli (Radiomics & medical science, University of Torino, Italy); Marco Grangetto, Enzo Tartaglione (Digital image processing Lab, University of Torino, Italy); Simone Palazzo, Isaak Kavasidis (PeRCeiVe Lab, University of Catania, Italy); Bogdan Ionescu, Gabriel Constantin (Multimedia Lab @ CAMPUS Research Institute, University Politechnica of Bucharest, Romania); Miquel Perello Nieto (Computer Science, University of Bristol, UK); Inês Domingues (School of Sciences University of Porto, Portugal).

References

[1] I. Colonnelli, B. Cantalupo, R. Esposito, M. Pennisi, C. Spampinato, and M. Aldinucci. HPC Application Cloudification: The StreamFlow Toolkit. *12th Workshop on Parallel Programming and Run-Time Management Techniques for Many-core Architectures and 10th Workshop on Design Tools and Architectures for Multicore Embedded Computing Platforms*, Dagstuhl, Germany, pp. 5:1–5:13, (2021). doi:10.4230/OASIcs.PARMA-DITAM.2021.5
[2] I. Colonnelli, B. Cantalupo, I. Merelli, and M. Aldinucci. StreamFlow: cross-breeding Cloud with HPC. *IEEE Transactions on Emerging Topics in Computing*, (2020). doi:10.1109/TETC.2020.3019202
[3] R.F. da Silva, R. Filgueira, I. Pietri, M. Jiang, R. Sakellariou, and E. Deelman. A characterization of workflow management systems for extreme-scale applications. *Future Generation Computer Systems*, Vol. 75, pages 228–238, October 2017. doi:10.1016/j.future.2017.02.026
[4] P. Amstutz, M. R. Crusoe, N. Tijanić, B. Chapman, J. Chilton, M. Heuer, A. Kartashov, J. Kern, D. Leehr, H. Ménager, M. Nedeljkovich, M. Scales, S. Soiland-Reyes, and L. Stojanovic. Common workflow language, v1.0, (2016). doi:10.6084/m9.figshare.3115156.v2
[5] High-performance computing and AI team up for COVID-19 diagnostic imaging. https://aihub.org/2021/01/12/high-performance-computing-and-ai-team-up-for-covid-19-diagnostic-imaging/, (2021). Accessed: 2021-06-30
[6] M. Pennisi, I. Kavasidis, C. Spampinato, V. Schinina, S. Palazzo, F.P. Salanitri, G. Bellitto, F. Rundo, M. Aldinucci, M. Cristofaro, P. Campioni, E. Pianura, F. Di Stefano, A. Petrone, F. Albarello, G. Ippolito, S. Cuzzocrea, S. Conoci. An Explainable AI System for Automated COVID-19 Assessment and Lesion Categorization from CT-scans. *Artificial Intelligence in Medicine*, (2021). doi:10.1016/j.artmed.2021.102114
[7] M. Aldinucci, S. Rabellino, M. Pironti, F. Spiga, P. Viviani, M. Drocco, M. Guerzoni, G. Boella, M. Mellia, P. Margara, I. Drago, R. Marturano, G. Marchetto, E. Piccolo, S. Bagnasco, S. Lusso, S. Vallero,

G. Attardi, A. Barchiesi, A. Colla, and F. Galeazzi. HPC4AI, an AI-on-demand federated platform endeavour. *ACM Computing Frontiers*, Ischia, Italy, (2018). doi:10.1145/3203217.3205340
[8] OpenAI. AI and Compute. https://openai.com/blog/ai-and-compute/, (2018). Accessed: 2021-06-30

THE ENEA-REG REGIONAL EARTH SYSTEM MODEL: DESIGNING SIMULATIONS FOR PRESENT AND FUTURE CLIMATE

Alessandro Anav*, Adriana Carillo, Massimiliano Palma, Maria Vittoria Struglia and Gianmaria Sannino

ENEA SSPT-MET-CLIM, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, CR Casaccia, Rome, Italy

ABSTRACT. An effort is made to implement a regional earth system model (ENEA-REG) to simulate present and future climate over the Med-CORDEX region. In this configuration, the ENEA-REG system uses the regional climate model WRF as atmospheric component, a river model (HD), and an ocean model (MITgcm). This multi-model system has been developed by the SSPT-MET-CLIM group in ENEA over the CRESCO6 HPC. Here we briefly presents the modeling system, the input data used and some results as an example.

1 Introduction

The Mediterranean basin is a complex region, characterized by the presence of pronounced topography and a complex land-sea distribution including a considerable number of islands and several straits. These features generate strong local atmosphere–sea interactions leading to the formation of intense local winds, like Mistral, Etesian and Bora which, in turn, dramatically affect the Mediterranean ocean circulation. Given the relatively fine spatial scales at which these processes take place, the Mediterranean basin provides a good opportunity to study regional climate, with a special focus on the air-sea coupling. For these reasons, regional coupled models are extensively developed and used to study both present and future Mediterranean climate system.

2 Description of the regional Earth system model

In the configuration presented here, the ENEA-REG system uses the Weather Research and Forecasting (WRF v4.2.2) model to simulate the atmospheric dynamic. WRF is a limited-area, non-hydrostatic, terrain-following eta-coordinate mesoscale model developed by the NCAR/MMM (National Center for Atmospheric Research, Mesoscale and Microscale Meteorology division). We use a one way nesting to simulate (offline) the atmospheric dynamics at coarse spatial resolution over the whole Europe: in this case, the model domain is projected on a Lambert conformal grid with a horizontal resolution of 36 km and with 51 vertical levels extending from land surface up to 10 hPa (Fig.1). Then, a finer domain, with resolution of 12 km, is used coupled to the ocean model to simulate present and future climate over the Mediterranean region.

*Corresponding author. E-mail: alessandro.anav@enea.it

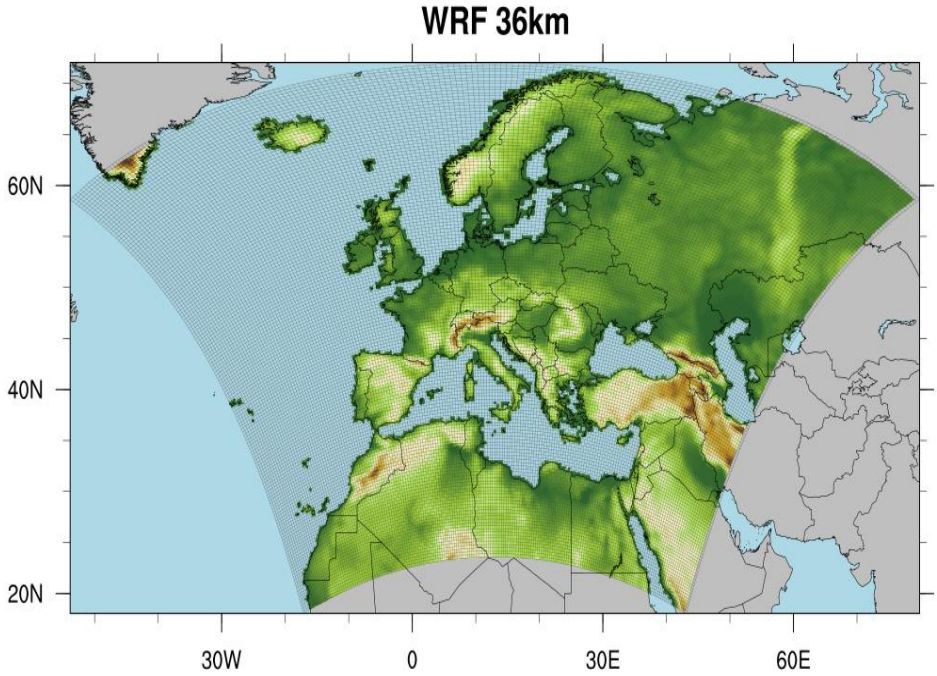


Fig. 3: Corse resolution domain of the ENEA-REG system, with green shading representing the topography of the atmospheric model and solid gray lines indicating the computational grid.

The ocean component of the ENEA-REG system is the Massachusetts Institute of Technology General Circulation Model (MITgcm, version z69). The MITgcm solves both the hydrostatic and nonhydrostatic Navier-Stokes equations under the Boussinesq approximation for an incompressible fluid with a spatial finite-volume discretization on a curvilinear computational grid using the z* rescaled height vertical coordinate. The model domain has a horizontal resolution of 1/12°, corresponding to 570x264 grid points, and covers the entire Mediterranean Sea with the boundary conditions in the Atlantic Ocean. In the vertical the model is discretized using 75 unevenly spaced Z-levels going from 1 m at the surface to about 300 m in the deepest part of the basin.
In the coupled configuration, the atmospheric model retrieves sea surface temperature (SST) from the ocean model (where grids are overlapped), while the ocean model collects surface pressure, wind components, freshwater (evaporation-precipitation, i.e. E-P) and heat fluxes from the atmospheric component. Similarly, the hydrological model uses surface and sub-surface runoff simulated by the atmospheric component to compute the river drainage and exchanges this field with the ocean component to close the water cycle (figure 1). The model coupling time step between ocean and atmosphere is set to 3-hours, while the coupling with the hydrological model is defined as 1-day. All the ENEA-REG components has been compiled with the intel compiler (mpiifort) and ran over CRESCO6 using 418 CPUs; the domain decomposition has been achieved with MPI.

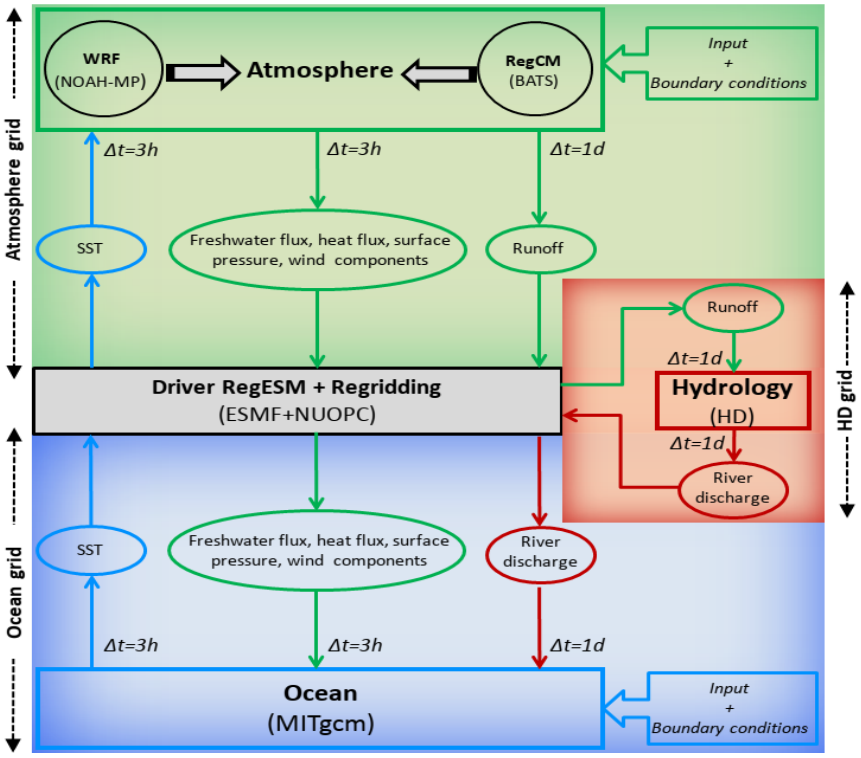


Fig.4: Schematic description of the ENEA-REG regional coupled model. The green block represents the atmosphere with the two components that can be selected and used (i.e. WRF and RegCM), the blue block is the ocean component (i.e. MITgcm), the red block represents the river routing component while the grey block is the ESMF/NUOPC coupler which collects, regrids and exchanges variables between the different components of the system.

Meteorological input data for hind-cast simulations are provided by the ERA-Interim and ERA5 reanalysis; in addition, data from different CMIP6 ESMs (i.e. CESM2 and MPI-ESM1-2-HR) are used for historical and future simulations under different emission scenarios (ssp126, ssp245, ssp585). Because of the large local to regional biases of CMIP6 models, a bias correction system has been implemented to remove the mean bias from input forcings keeping their interannual variability and trends.

3 First results and discussion of future work

For present climate we planned to run our system forced with different reanalysis (ERA-Interim and ERA5) for the period 1980-2014, while for future simulations covering the time period 2015-2100 we use CMIP6 models. In addition, to check the ability of the coupled model to reproduce realistic ocean and atmospheric dynamics for the present climate, we compare results of the models forced with a reanalysis and CMIP6 data over the temporal period 1980-2014. Overall, a model that provides realistic results for the present climate, likely will be reliable for future projections.

Figure 3 shows the time evolution of the mean annual surface air temperature averaged over the whole domain of Figure 1. The black curve corresponds to the values from the reanalysis (used as reference data), whereas dashed lines represent different WRF simulations using ERA5 as forcing (blue line), MPI-ESM1-2-HR (green line) and MPI-ESM1-2-HR bias corrected (red line). Results indicate that when WRF is forced with ERA5, the model and reference values are quite close, with differences smaller than 0.5 °C, while when the same simulation is performed using a CMIP6 ESM (i.e. MPI-ESM1-2-HR) a large cold bias occur. Besides, results highlight how the bias correction performed on the input forcing works well, being the red line in good agreement with the reference data.

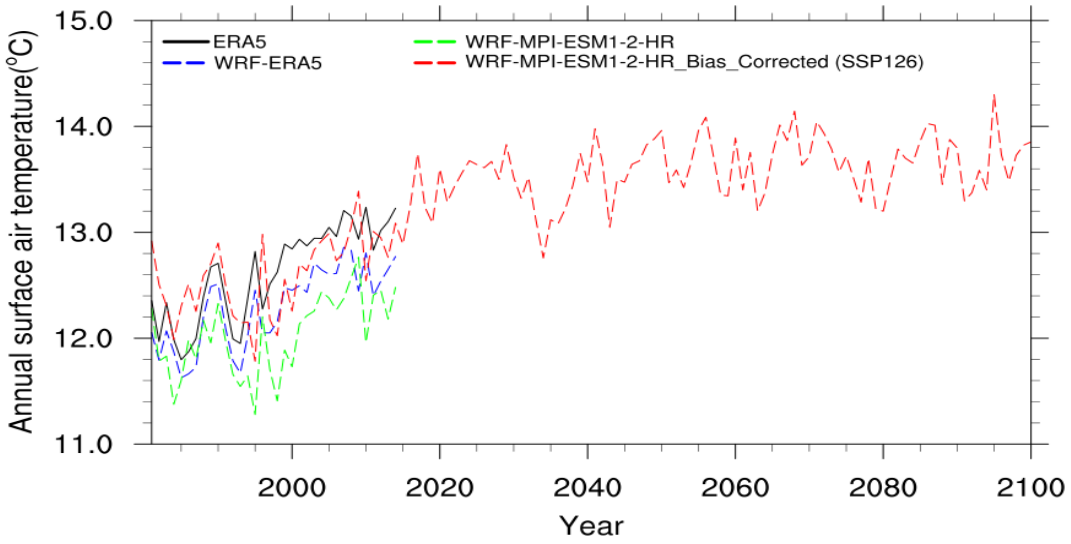


Fig.5: Comparison of the evolution of mean annual surface air temperature as simulated by the atmospheric model using different forcing data and ERA5 reanalysis.

Considering the temporal performance of the models, the entire regional Earth system is composed by serial and parallel processes that use up to 418 cores. To simulate one single year with the climate model only (i.e. 36 km over Europe), the procedure takes about 11 hours on CRESCO6 with 168 cores, while the full coupled system requires about 40 hours to perform one single year over the finer domain (12 km) centred over the Mediterranean region.

References

[1] Anav, A., Carillo, A., Palma, M., Struglia, M. V., Turuncoglu, U. U., and Sannino, G.: The ENEA-REG system (v1.0), a multi-component regional Earth system model: sensitivity to different atmospheric components over the Med-CORDEX (Coordinated Regional Climate Downscaling Experiment) region, Geosci. Model Dev., 14, 4159–4185, https://doi.org/10.5194/gmd-14-4159-2021, 2021.

ANALYSIS OF NONLINEAR DYNAMICS OF CHIRPING ALFVÉN MODES

Sergio Briguglio[1*], Xin Wang[2], Claudio Di Troia[1], Giuliana Fogaccia[1], Valeria Fusco[1] and Gregorio Vlad[1]

[1]*ENEA, Fusion and Nuclear Safety Department, C. R. Frascati, Via E. Fermi 45, 00044 Frascati (Roma), Italy*

[2]*Max-Planck-Institut für Plasmaphysik, Boltzmannstraße 2, Garching D-85748, Germany*

ABSTRACT. Nonlinear dynamics of chirping Alfvén modes is analysed on the basis of particle-in-cell simulations performed by the XMHGC particle-in-cell code.

1 Chirping Alfvén modes

Alfvén modes can be driven unstable, in Tokamak plasmas, by the resonant interaction with alpha particles produced by fusion reactions and/or energetic ions produced by auxiliary heating methods, characterised by speeds of the order of the Alfv\'{e}n velocity. Mode-particle interaction can in turn deteriorate the confinement of these particles, preventing their thermalisation in the core plasma and eventually damaging the first wall. Some of the Alfvén modes, like the Toroidal Alfvén Eigenmodes (TAEs) have a MHD counterpart; that is, they exist as marginally stable or quasi marginally stable modes even in the absence of energetic-particle drive. Other modes, like the Energetic Particle Modes (EPMs) [1] have no such counterpart: they are oscillations of the Alfvén continuum driven unstable only when the energetic-particle drive exceeds the continuum damping; in the absence of this drive, they are strongly damped. While the former modes are typically characterised by an approximately constant frequency, even in the nonlinear stage, essentially constrained to keep within the frequency gaps opened in the Alfvén continuum by the coupling between different poloidal harmonics, the latter are able to vary their frequency (they have been named chirping modes), during the nonlinear evolution, in order to extract as much power as possible from the energetic particles. Both because of this feature and the strong dependence on the energetic-particle drive they usually present above the instability threshold, such chirping modes are considered as potentially harmful for the confinement of energetic particles in burning Tokamak plasmas and have attracted much interest.

2 Nonlinear dynamics of chirping Alfvén modes

We have investigated, by numerical particle simulation performed with the XHMGC code [2], the nonlinear dynamics of chirping modes. The case of an Alfvén spectrum characterised by a single toroidal mode number n has been considered. Once the relevant resonances have been identified, the corresponding phase-space region is sampled by a large number of test particles (moving in the fields yielded by the self-consistent full-population simulation) in order to get detailed information on the nonlinear evolution of the mode-particle power exchange.

We have shown that an island-like structure forms around the resonance radius in the plane (Θ,r), enclosing the bounded orbits of particles instantaneously trapped in the potential well of the wave; here, Θ corresponds to the wave-phase seen by the particle, and r is the radial coordinate of the particle. The formation of the island creates a flattened density region delimited by large negative density gradients moving inward and outward as the island width grows with the increasing mode amplitude. If the mode frequency were fixed, saturation would be reached when the density flattening completely covers the resonant interaction region [3]. The width of such region essentially coincides with the smallest between resonance and mode widths. As the former scales with the linear growth rate of the mode, for relatively weak modes, like that considered in the present case, the resonant interaction region would be limited

*Corresponding author. E-mail: sergio.briguglio@enea.it

by the resonance width, and the saturation condition would occur when the large negative density gradients reach the boundaries of this region. If the frequency is allowed to decrease, however, resonance radius and resonance region move inward, and the inner large gradient is able to further drive the mode. In turn, the island reconstitutes around the new resonance radius, drifting inward too. As a consequence, the large negative density gradient further moves inward. This process goes on as long as the frequency can decrease and the the resonance region can move inward in such a way to accommodate the large gradient in the resonance region. When this is no longer possible, either because a further change in frequency is disadvantageous in terms of drive/damping balance, or because such a change does not result in a significant inward shift of the resonance radius and resonance region, the gradient ceases to be effective in driving the mode. To further grow, the mode has to tap a different resonant structure, possibly making use of additional frequency variations.

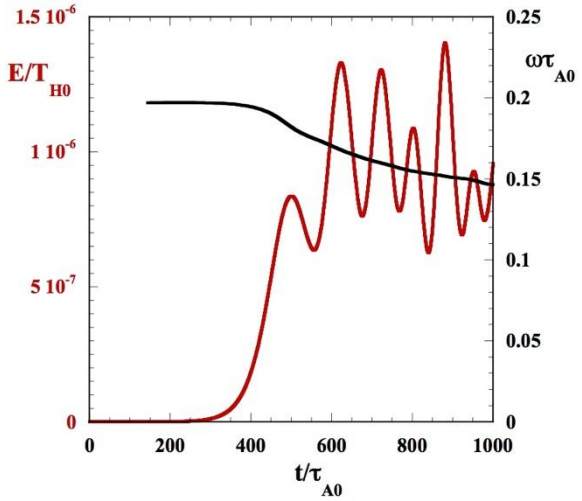


Fig. 1: Time evolution of mode frequency and amplitude for the Energetic Particle Mode investigated in our simulations.

Figure 1 shows the time evolution, for the case analyzed in our simulations, of mode frequency and amplitude. Figure 2 shows the island structure, in the plane (Θ,r), at two different times in the nonlinear stage. Here, Θ, is the wave-phase seen by the particle, r is its radial coordinate. The inward radial drift of the island, due to the downward frequency chirping, is clearly observed.

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [4]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

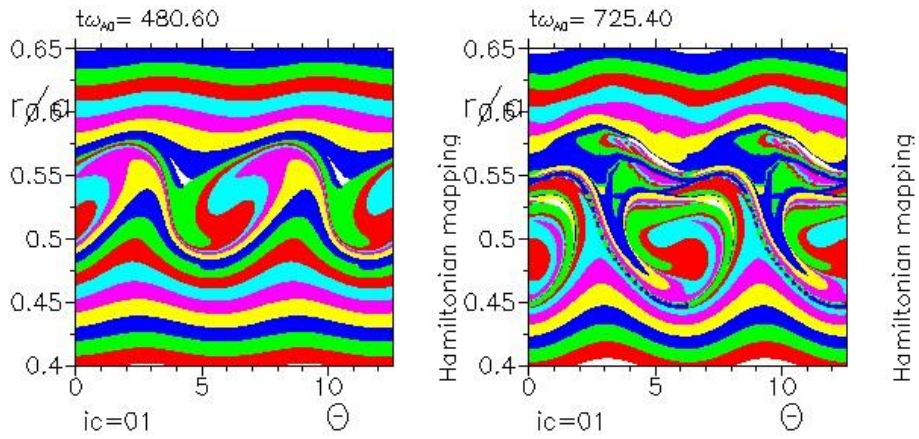


Fig. 2: Island structure in the plane (at two different times in the nonlinear stage. The inward radial drift of the island, due to the downward frequency chirping, is clearly observed

References

[1] L. Chen. Theory of magnetohydrodynamic instabilities excited by energetic particles in tokamaks. *Phys. Plasmas* **1**, p. 1519, (1994)

[2] S. Briguglio, G. Vlad, F. Zonca, C. Kar. Hybrid magnetohydrodynamic-gyrokinetic simulation of toroidal Alfvén modes. *Phys. Plasmas* **2**, pp. 3711-3723, (1995)

[3] S. Briguglio, M. Schneller, X. Wang, C. Di Troia, T. Hayward, V. Fusco, G. Vlad and G. Fogaccia. Saturation of Alfvén modes in Tokamak plasmas investigated by Hamiltonian mapping techniques. *Nucl. Fusion* 57, p. 072001, (2017)

[4] F. Iannone et al., "CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout," 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland, 2019, pp. 1051-1052, doi: 10.1109/HPCS48598.2019.9188135

3D DIRECT NUMERICAL SIMULATION OF CH4/O2/CO2 SUPERCRITICAL DIFFUSION FLAME

Donato Cecere*, Eugenio Giacomazzi and Nunzio Arcidiacono

ENEA, Laboratory of Processes and Systems Engineering for Energy Decarbonisation, Department of Energy Technologies and Renewable Sources, Rome, Italy

ABSTRACT. Turbulent non-premixed methane oxy-combustion in an atmosphererich in CO2 at supercritical conditions (300 bar) is investigated by means of 3DDirect Numerical Simulation. A simple shear-layer configuration typical of slotburners is considered: the fuel flows through the central slot, the oxidant O2/CO2 (90% O2 by mass) mixture flows on both sides of the slot. Compressible N-Sequations coupled to the Peng-Robinson cubic equation of state (PR-EoS) in itstranslated volume formulation (VTPR) are solved with a reduced 23-species kineticscheme derived by ARAMCO 2.0 full mechanism. Local dilution due to the sCO2 coflow decreases the local CO production; the very high pressure results inturbulent scales smaller than at lower pressure with an inner cut-off larger than theKolmogorov scale, likely related to the Prandtl number larger than one; the flow exhibits uphill (counter-gradient) diffusion, with the main contribution due to the Hirshfelder and Curtiss' mass flux. Temperature iso-surfaces fractal dimension is calculated with values in the range 2.43-2.59 , decreasing as temperature increases indicating that turbulent structures are more dissipated.

1 Introduction

High-pressure combustion of reactants exhibiting real gas behaviour is becoming an important research topic for the growing interest in supercritical CO_2 and organic Rankine gas turbine cycles, diesel engines with higher and higher pressures, liquid oxygen rocket engines, cooling systems. In such applications, the flows can be far away from ideal thermodynamics, and the different fluid behaviour has to be accounted for by means of real gas equations of state and specific models for molecular transport properties. Experimental work at such high-pressure conditions may be prohibitive: the use of advanced laser diagnostics is not an easy task to achieve, and very few examples exist in literature; besides, facilities themselves may be very expensive. Hence, most of research and design in this area is based on numerical simulation. However, also the numerical approach is not an easy task, suffering problems of both modelling accuracy, computational efficiency and lack of robustness by numerical schemes. Furthermore, numerical schemes are stressed hardly: the high-density gradients typical of some applications (as those involving liquid injection) and the multi- species transport enhance wiggle formation in fully compressible solvers; spurious pressure oscillations are generated when a fully conservative scheme is adopted, due to the high non-linearity of the real-fluid equation of state.

2 The Numerical Experiment and its Set-Up

The numerical experiment here simulated has a nominal pressure around 300 bar and consists in asimple shear-layer configuration that can be encountered in slot burners; the flow is confined by means of two no-slip adiabatic walls (at left and right of the domain). At the bottom of the computational domain there is the inlet, while the outlet is located at the top. Simulations are performed in a threedimensional framework and the domain X x Y x Z is 0.5 x 8 x 10mm, discretised by means of 384x2880x5000 points. The computational nodes are not uniformly distributed along the streamwise direction z, with a minimum Δz =1.05 μm close to the inlet, and stretched from the center towards the sides along the transversal direction y, with y_{min} = 1.08 μm and y_{max} = 21 μm (in the center of the coflow CO_2 streams). The CH_4 is injected centrally at 28 m/s and 388K; its width is 4 x 10^{-4} m. The oxidant O_2/CO_2 (Y_{O2} = 0.9

*Corresponding author. E-mail: donato.cecere@enea.it.

by O_2 mass fraction) mixture flows on both sides of the methane. Each oxidant jet flows at 84 m/s and 450K; their width is 1.98 x 10^{-4} m. A pilot flame (10^{-4} m widen) is imposed at the inlet between the two streams of reactants to force ignition. Adjacent to each of the oxidant streams, there is a co-flowing stream of CO_2 at 60 m/s and 973K. For this combustor with only two inlets the global equivalence ratio is 0.7. The reduced temperatures for the fuel, oxidant and CO_2 streams are nearly 2, 2.9 and 3.2; their reduced pressures are 6.5, 5.9 and 4, respectively. The (isotropic) turbulence characteristics of the jets are specified in terms of velocity fluctuations and auto-correlation length-scales in table 1.

Table 1: Injection data for the present DNS simulation. Y_{O2} = 0.9 by O_2 mass fraction in the oxidant flow. Pilot flame at 1600 K are imposed.

Pressure 300 bar	T_{inj} K	ρ_{inj} kg/m^3	μ_{inj} kg/(m · s)	U_{inj} m/s	u'_{inj} m/s	ℓ_t mm	Re_t [−]	η μm
O_2/CO_2	450	240.36	$3.79 \cdot 10^{-5}$	66	3.5	0.22	646.21	1.78
CH_4	388	150.66	$4.54 \cdot 10^{-5}$	22	3.0	0.2	499.10	1.89
CO_4	950	150.87	$4.26 \cdot 10^{-5}$	60	6.0	3.5	637.47	27.58

3 Physical and Numerical Models

In this work the compressible Navier-Stokes equations are solved for a reacting real gas flow. The Peng-Robinson cubic equation of state in its improved translated volume formulation is assumed [1, 2]: it requires only the acentric factor on top of the critical properties and is thus easy to implement for a wide range of species. It provides slightly better accuracy than the Redlich-Kwong EoS because of a more complex temperature dependence. As a consequence, energies and specific heats predictions are more accurate, an important feature for reacting flows. The mathematical models adopted are derived for a Newtonian and Stokesian fluid of N_s chemical species. The mass diffusion flux has three contributions: the first due to concentration gradients (modelled through the Hirschfelder and Curtiss' law for multicomponent mixtures), the second due to pressure gradients (the baro-diffusion mechanism), and the third one due to temperature gradients (the thermo-diffusion or Soret effect). The heat diffusion has also three contributions: the first due to temperature gradients (the Fourier diffusion), the second due to mass diffusion fluxes, and the third one is the Dufour effect (reciprocal of the Soret effect). The thermodiffusion, or Soret effect, is the mass diffusion due to temperature gradients, driving light species towards hot regions of the flow. This effect, often neglected, is neverthless important, in particular for hydrogen combustion, and in general when very light species play an important role. Usually the Dufour effect is negligible even when thermo-diffusion is not, and hence it is neglected. Molecular viscosity and thermal conductivity are accurately modelled through NIST models in REFPROP with an Extended Corresponding States method and fluid-specific correlations. The diffusion coefficient D_i of the i-th species into the rest of mixture is modelled according to the Hirschfelder and Curtiss expression, where the required binary diffusion coefficient is calculated by means of kinetic theory. The thermodiffusion coefficient $D_{T\,i}$ is estimated by means of the EGLIB routines. Just as an example, Fig. 1 shows the ratio between the high pressure corrected i-th species mixture diffusion (the Takahashi's correction factor (DP_R) is adopted in the i-th species binary diffusivity) and the classical species mixture diffusion from an instantaneous flowfield as a function of temperature: at low temperature (T < 600K) the mixture species diffusivity is nearly reduced by 20% (i.e. CO_2;OH), while at higher temperature the mixture diffusion is increased up to 7% with the H_2O having the widest correction factor distribution.

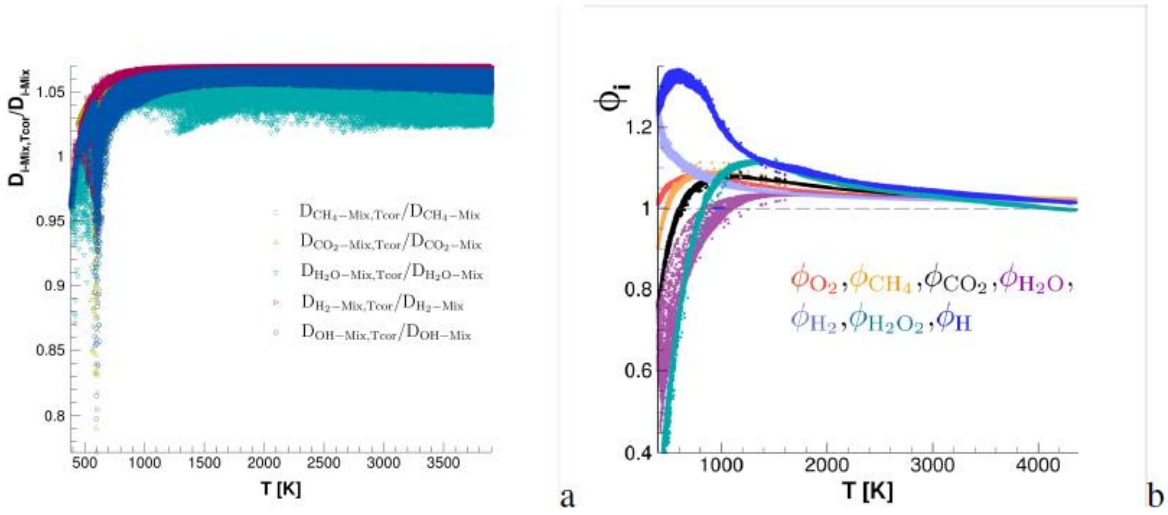


Fig.1: a) Instantaneous ratio between corrected i-th species mixture diffusion and standard i-th species mixture diffusion; b) Instantaneous species fugacity coefficients as function of temperature.

4 Chemical Mechanism

For the present study, a 142-step 23-species finite-rate methane-oxygen kinetic mechanism developed for high pressure oxy-combustion and derived from ARAMCO2:0 is employed. This scheme is able to perform very well when compared to the predictions of the detailed ARAMCO2.0 mechanism in terms of ignition delay time, perfectly stirred reactor estimation under various CO_2 dilution and equivalence ratios at high pressure [4]. A maximum difference of 3% is found in temperature peak near the stoichiometric mixture fraction line and a very good agreement is shown also in CO, CO_2, H_2, OH and H_2O mass fraction comparisons for all the scalar dissipation rates investigated (4% of maximum mass fraction difference in the range of _ between [0.001 - 1.5 x10^6]). Due to the high-pressure conditions of the present study, individual species fugacities cannot be neglected when calculating backward reaction constants. The fugacities reach the ideal value only for very high temperatures, while the deviation also reaches 60% in low temperature flow areas (T < 1200K) where real gas thermodynamics and kinetics is important for correct ignition delay times calculations and so for the correct flame position [5].

5 Results

As shown in Fig.2a the fuel core penetrates the computational domain within its radial dimension for 1.5mm, then mixes with the oxidant by means of corrugated finger-like structures associated to the high-momentum difference between the inner and outer streams (4200 vs 22000 kgm^{-2} s^{-1}, respectively).

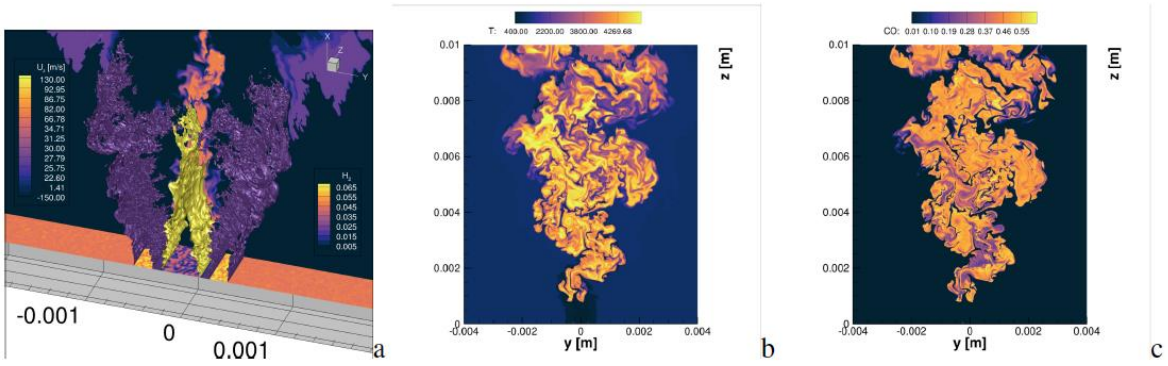


Fig. 2: a) Instantaneous flame structure, CH_4 mass fraction at Y=0.5 an;d O_2 mass fraction at Y=0.9; b) Instantaneous flame temperature; c) Instantaneous distribution of CO mass fraction.

The high momentum gradient of the shear layer promotes turbulent mixing with vortices that force the internal gas towards the external stream. The extruded structures are elongated by the velocity difference to form the ligaments since the diffusion processes have not time to smear out the momentum and species gradients between the jet core and the surroundings. These very thin structures (50 μm), protruding from the inner core, contribute to reduce the external velocity stream and consequently to form new vortices. Figure 2c shows an instantaneous temperature profile. The flame is stably anchored thanks to the pilot flame and at stoichiometric isolines the temperature reaches very soon a maximum value of 3850K. Anyway, the CH_4 interacts very soon with the pilot flame (1600K) and the O_2 flux, and it ignites only approximately at 0.8mm , producing H_2, C_2H_4 and CO (see Fig. 2c) in rich conditions. This suggests that some further methods must be adopted in order to increase mixing and to rich also locally the global lean equivalence ratio of the flame.

References

[1] K. Harstad, R. Miller, J. Bellan, Efficient high pressure state equations. Chemical Engineers Journal 43 (6) (1997) 1605-1610.

[2] D. Wong, S. Sandler, A theoretically correct mixing rule for cubic equations of state. American Institute of Chemical Engineers Journal 38 (1992) 671-680.

[3] S. Takahashi, Preparation of a generalized chart for the diffusion coefficients of gases at high pressures, J. Chemical Engineering of Japan

[4] Manikantachari K.R.V., Vesely L., Martin S., Bobren-Diaz J.O., Vasu S., Reduced Chemical Kinetic Mechanisms for Oxy/Methane Supercritical CO_2 Combustor Simulations, J. Energy Resour. Technol., 140, (2018) 092202.

[5] G. Kogekar, C. Karakaya, G.J. Liskovich, M.A. Oehlschlaeger, S.C. deCaluwe, R.J. Jee, Impact of non-ideal behavior on ignition delay and chemical kinetics in high-pressure shock tube reactors, Combustion and Flame 189 (2018) 1-11.

POLYMER PHYSICS MODELING AND COMPUTER SIMULATIONS RECAPITULATE GENOME 3D ORGANIZATION AFTER COHESIN DEPLETION

Mattia Conte, Andrea Esposito, Alex Abraham, Simona Bianco and Andrea Maria Chiariello*

Dipartimento di Fisica, Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant'Angelo, 80126, Naples, Italy

ABSTRACT. Genome three-dimensional (3D) spatial organization is crucial to cell activity. Indeed, recent technologies revealed that it is non-randomly folded in 3D space so to bring in spatial proximity genes and regulators. Nevertheless, the mechanisms explaining how architectural specificity is achieved are still poorly understood. In this context, polymer physics models turn out to be a powerful tool to quantitatively investigate such aspect of Molecular Biology. Here, we report some recent advances in the field related to the predictive power of such models compared against independent microscopy experimental data. On the same line, we stress that an efficient algorithmic implementation and computational resources are fundamental to produce increasingly accurate models able to improve our knowledge of genomic structure.

1 Introduction

Different, recent technologies in molecular biology, e.g. Hi-C (1), GAM (2) or SPRITE (3) methods, have shown that, within the cell nucleus, chromosomes are organized in complex non-random three-dimensional (3D) architecture (**Fig. 1a**) (4). For instance, it has been discovered that chromosomes are organized in megabase sized domains of self-interaction (named TADs) (5,6), which in turn have an inner structure (7) and, on higher scales, interact in hierarchical structures (metaTADs) up to cover entire chromosomes (**Fig. 1b**) (8). Such organization is essential to the correct activity of the cell, by establishing a precise network of contacts between genes and their regulatory elements (enhancers) (9,10) so to control gene transcription. In this scenario, quantitative models based on polymer physics laws combined with massive numerical simulations allowed to systematically interpret the enormous amount of data generated by these new experimental methods and to better understand the physical mechanism shaping genome 3D organization (11,12). In this report, we review some recent advances in the field based on the Strings and Binders Switch (SBS) model, in which molecular binders mediate the interactions between distant DNA binding sites (13,14) so to shape chromatin structure. The model can be trained on experimental data and specialized to describe real genomic regions by means of a machine learning based method. Afterwards, massive parallel molecular Dynamics simulations are performed to generate an ensemble of 3D structures that quantitatively describe the considered region (15). As this is very computationally demanding step, efficient calculus resources and parallelization strategies are of essence, especially for the systems of high complexity. As case of study, we show that such model is able to explain the 3D chromatin structure at the single-cell level (16) in WT and in perturbed conditions, i.e. cohesion depletion, for a 2.5 Mb long locus on chr21, human HCT116 cell line, and successfully validated against publicly available microscopy data (17).

2 A polymer physics-based algorithm to investigate 3D genome folding

In the SBS model of chromatin there are two main ingredients: (1) a self-avoiding chain of beads, a fraction of which acts as binding sites for the interaction with (2) diffusive molecular factors, i.e. the binders. Binding sites have a specific interaction type (or color) and can experience an attractive potential only with binders of the same color (**Fig. 1c**). In the 3D nuclear space, the complex network

*Corresponding author. E-mail: chiariello@na.infn.it.

of interactions among beads and binders drives the folding of the system via a mechanism of microphase separation. As dictated by polymer physics (18), the system transits from a randomly open conformational state (coil) to a more compact, globular phase as soon as the binder concentration (or affinity strength) exceed a threshold value (**Fig. 1d**). Dynamical and equilibrium proprieties of such a system are studied by Molecular Dynamics (MD) simulations with classical interaction potentials (15,19) and Langevin equations, solved by a standard velocity Verlet integrator using, e.g., the LAMMPS package (20).

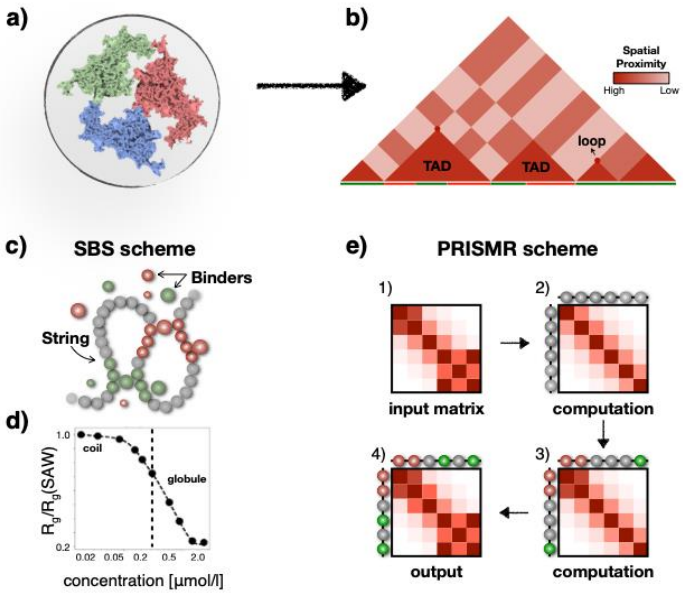


Fig.1: Polymer physics-based model and algorithm to investigate the principles of 3D genome folding. a) Novel experimental methods, such as Hi-C or multiplexed FISH, can measure the spatial proximity of genomic regions across cell populations and in single cells. **b)** Output data can be arranged in a symmetric pairwise matrix whose heatmap shows strikingly non-random pattern of interactions. **c)** Theoretical models can help understanding the genome folding principles. In the Strings and Binders Switch (SBS) model, the process of folding is guided by homotypic interaction between binding sites and cognate diffusing binders. **d)** The folding of the polymer in the 3D space can be monitored, for instance, by the gyration radius R_g of the polymer itself. As the binder concentration grows above a certain threshold value, R_g undergoes a sharp decrease which marks the microphase transition of the system from the coil to the globular state. **e)** A scheme of PRISMR (12). The algorithm aims to find the optimal number of binding site types and their positioning along the polymer that best describes a given input matrix. Items adapted from (21).

A complete simulation of the described system from the coil to the globular state can require many hours of running time depending on the size of the studied polymer model. In this sense, High-Performance-Computing (HPC) resources play a fundamental role in this approach.
The state of an SBS polymer is fully characterized by the location of the different binding sites along the polymeric chain. In order to apply this model to real genomic regions, we have developed a machine learning based algorithm, named PRISMR (12). Precisely, starting from the pairwise contact matrix of the region of interest, PRISMR aims to find the optimal number of binding site types and their positioning along the polymer that best describes the input within a given accuracy (**Fig. 1e**). This is done by using a standard Simulated Annealing approach to find the minimum of a specifically defined cost function consisting of two terms: the mean squared error function between the input and model-derived contact matrix and a Bayesian term which penalizes the addition of new binding sites, so to avoid overfitting.

3 Validation of model predictions upon cohesin depletion against independent single-molecule 3D distance data

The SBS model has been shown to recapitulate large scale features of chromosome 3D organization (14,19), as well as the folding of different genomic loci across different cell types (22,23). The predictive power of the model has been tested against data from different independent experimental technologies (16,24) and, importantly, it provides a powerful tool to predict in-silico the effects on genome 3D structure of disease-linked chromatin rearrangements, such as deletions or inversions (12). As the output of our approach is a thermodynamic ensemble of steady-state 3D conformations, the SBS model also allows a quantitative investigation of the properties underlying chromatin folding at the level of single DNA molecules, i.e., well beyond population-averaged data. Model predictions can then be validated with available experimental (e.g., microscopy-based) data, thus providing insights on the machinery guiding folding in single-cells. Based on such a rationale, in the paper (16) we systematically tested the single-molecule 3D structures predicted by the model against single-cell multiplexed FISH data (17) available for different genomic loci in human cells (e.g., in IMR90 and HCT116 cells). As discussed in (16), the consistent agreement between model and experiments provides evidence that a thermodynamic physical mechanism of polymer phase-separation, as envisaged by the SBS theory, controls chromosome architecture in single-cells. We review here as case study the model of a 2.5 Mb DNA region (chr21:34.6–37.1 Mb) in human HCT116 cells upon acute cohesin depletion (hereafter referred to as the HCT116+Auxin locus), which is fully detailed in the original work (16).

Based on the pipeline of our approach discussed in the previous section, the putative binding sites of the SBS model of the HCT116+Auxin locus are optimally inferred from bulk Hi-C data (25) by using PRISMR. The algorithm returned three distinct types of specific binding sites, each correlated with a specific combination of know chromatin organizing factors and epigenetic marks (16). Next, we ran massive, parallel Molecular Dynamics (MD) simulations of the optimal polymer model to derive an ensemble of equilibrium 3D conformations in the steady state of the system. As stated in classical polymer physics theory (18), the model structures fold in two main classes that correspond to the thermodynamic phases of the system: the coil (i.e., randomly folded) and the globule phase-separated state (see section 2). Hence, to investigate whether the SBS-predicted single molecules are a bona-fide representation of the real imaged single-cell conformations, we performed a quantitative all-against-all structural comparison between model and microscopy 3D structures. To this aim, we employed the minimum RMSD computational method (16,26) whereby each conformation from imaging data is univocally and significantly associated to a corresponding best-matching structure from the model (in the coil or globule state) (**Fig. 2a**). Technically, the efficiency of the method is optimized by the QCP (Quaternion Characteristic Polynomial) algorithm, which is the fasted method for assessing the least RMSD between two 3D coordinate sets (e.g., experimental vs model) (27). By using this method in HCT116+Auxin cells, we found that 80% experimental structures statistically map onto SBS model 3D conformations in the coil and 20% in the phase-separated globule state (**Fig. 2a**), suggesting that the removal of the cohesin complex tends to reverse the globule phase-separation into the coil state in most single-cells. As a first validation of the model, we computed the median distance matrix from the mixture ensemble (80% coil, 20% globule) of SBS conformations and we compared it against independent microscopy data (17) (**Fig. 2b**). As known from previous studies (25), cohesin depletion removes patterns at the bulk-averaged level, hence resulting in featureless contact (or distance) maps. Consistently, we found that the model-derived median distance matrix has no structure and well matches the experiment, as quantified by its high correlation values (Pearson r=0.96, genomic-distance corrected Pearson r'=0.57). Thus, a mixture of polymer structures mostly folded in the coil state is consistent with bulk distance data. Additionally, to further test the model, we checked that the experiment-model best-matches identified via the RMSD criterion indeed correspond to similar distance matrices and 3D structures (**Fig. 2c**). To investigate whether the SBS structures recapitulate the finer folding details of the studied locus, we computed the boundary probability function, i.e., the probability to find in a single-molecule a TAD-like boundary at a given genomic position (17)(**Fig. 2d**). In both model and experiment, we found that such a function is flat and uniformly spread across the locus, again consistent with a scenario where chromatin tends to form more fleeting and random interactions (as in the coil phase) rather stable contacts (as in the globule state). Finally, we measured the level of single-cell structural variability in both model and experimental conformations (**Fig. 2e**). To this aim, we computed the all-against-all r' correlation between pairs of experimental (blue), model (red), and experimental-model

(grey) single-cell distance matrices. We found all three distributions to be statistically similar to each other with an average correlation value r'=0.0, which reflects the high degree of chromatin cell-to-cell variability upon cohesin depletion.
Overall, model predictions on chromatin structure after cohesin depletion are consistently validated against independent single-cell data. In the emerging picture, consistent with the known role of cohesin as a key architecture organizing factor, cohesin depletion reverses chromatin globule phase separation to the coil state in most single cells, whose highly variable contact patterns originate mainly from random chromatin collisions rather than from phase-separated domains (16).

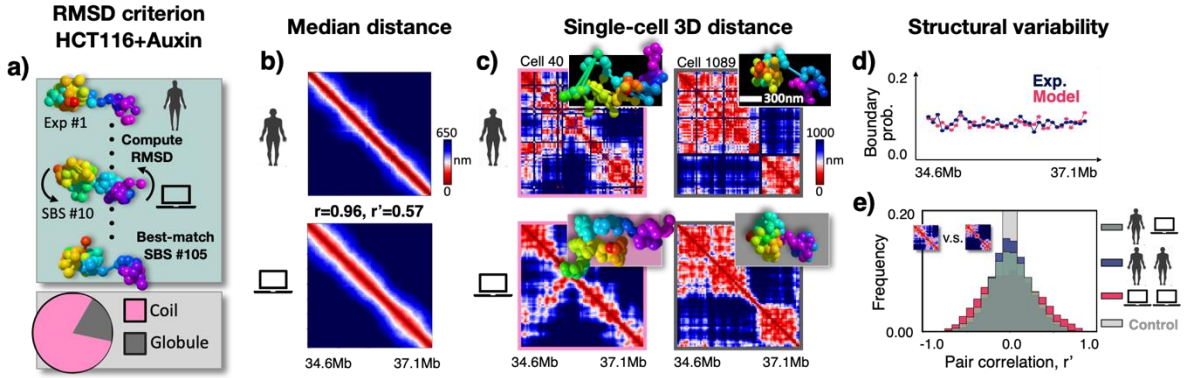


Fig.2: Single-molecule predictions of the SBS model upon cohesin depletion are validated against independent microscopy data. **a)** Scheme of the RMSD all-against-all structural criterion between SBS model-predicted (16) and microscopy (17) 3D structures. 80% experimental structures are optimally mapped into model conformations belonging to the coil state. **b)** Comparison between the imaged [Bintu18] (top) and model derived (bottom) median distance matrices of the studied 2.5 Mb DNA region (chr21:34.6–37.1 Mb) in human HCT116+auxin cells. **c)** Examples of imaged single-cell distance matrices and relative 3D structures are shown along with their best-matching pairs from the model. **d)** In both model and experiment, the boundary probability function is uniformly spread along the locus, as a reflex of the high cell-to-cell variability. **e)** Quantification of the single-molecule structural variability by measuring the all-against-all r' pair correlation between distance matrices. The distributions shown have an average value r'=0.0, which is consistent with the theoretical prediction whereby chromatin mostly folds in coil states after cohesin depletion. All items are adapted from (16).

4 Conclusions

In this report we discussed a general computational approach based on polymer physics that is able to reproduce the 3D structure of genomes in real organisms. As example, we considered the model of a 2.5 Mb genomic region located on chr21 in human HCT116 cell line in which structural single cell variability in WT and perturbed (cohesion depletion) conditions has been predicted with high accuracy. It is important to stress that the accuracy of the description is strictly linked to the complexity of the implemented model, which depends on system parameters as the polymer length, molecule number and equilibration time. Those aspects in turn impact the computational effort required to perform these numerical simulations. In this regard, High Parallel Computing resources from ENEA (28) allow to efficiently implement the appropriate Molecular Dynamics paradigms and represent a valuable tool to this kind of research.

References

[1] Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (80-). 2009;326(5950):289–93.
[2] Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, Xie SQ, et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature. 2017;543(7646):519–24.

[3] Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell [Internet]. 2018;174(3):744-757.e24. Available from: https://doi.org/10.1016/j.cell.2018.05.024
[4] Kempfer R, Pombo A. Methods for mapping 3D chromosome architecture [Internet]. Vol. 21, Nature Reviews Genetics. Nature Research; 2020 [cited 2021 Mar 16]. p. 207–26. Available from: www.nature.com/nrg
[5] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature [Internet]. 2012;485(7398):376–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22495300%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3356448
[6] Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485(7398):381–5.
[7] Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013;
[8] Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. Mol Syst Biol [Internet]. 2015;11(12):852–852. Available from: http://msb.embopress.org/cgi/doi/10.15252/msb.20156492
[9] Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 2016;164(6):1110–21.
[10] Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep [Internet]. 2016;15(9):2038–49. Available from: http://dx.doi.org/10.1016/j.celrep.2016.04.085
[11] Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc Natl Acad Sci. 2016;113(31):E4504–12.
[12] Bianco S, Lupiáñez DG, Chiariello AM, Annunziatella C, Kraft K, Schöpflin R, et al. Polymer physics predicts the effects of structural variants on chromatin architecture. Nat Genet. 2018;50(5):662–7.
[13] Nicodemi M, Prisco A. Thermodynamic pathways to genome spatial organization in the cell nucleus. Biophys J. 2009;96(6):2168–77.
[14] Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A, et al. Complexity of chromatin folding is captured by the strings and binders switch model. Proc Natl Acad Sci [Internet]. 2012;109(40):16173–8. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1204799109
[15] Annunziatella C, Chiariello AM, Esposito A, Bianco S, Fiorillo L, Nicodemi M. Molecular Dynamics simulations of the Strings and Binders Switch model of chromatin. Vol. 142, Methods. 2018. p. 81–8.
[16] Conte M, Fiorillo L, Bianco S, Chiariello AM, Esposito A, Nicodemi M. Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. Nat Commun [Internet]. 2020 Dec 3;11(1):3289. Available from: http://www.nature.com/articles/s41467-020-17141-4
[17] Bintu B, Mateo LJ, Su J-HH, Sinnott-Armstrong NA, Parker M, Kinrot S, et al. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science (80-). 2018;362(6413):eaau1783.
[18] De Gennes PG. Scaling concepts in polymer physics. Cornell university press. Ithaca NY,. 1979;
[19] Chiariello AMAM, Annunziatella C, Bianco S, Esposito A, Nicodemi M. Polymer physics of chromosome large-scale 3D organisation. Sci Rep. 2016;6.
[20] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995;117(1):1–19.
[21] Esposito A, Bianco S, Fiorillo L, Conte M, Abraham A, Musella F, et al. Polymer models are a versatile tool to study chromatin 3D organization. Biochem Soc Trans [Internet]. 2021 Aug 27 [cited 2021 Oct 18];49(4):1675–84. Available from: /biochemsoctrans/article/49/4/1675/229278/Polymer-models-are-a-versatile-tool-to-study
[22] Esposito A, Chiariello AMAM, Conte M, Fiorillo L, Musella F, Sciarretta R, et al. Higher-order

Chromosome Structures Investigated by Polymer Physics in Cellular Morphogenesis and Differentiation. J Mol Biol [Internet]. 2019;432(3). Available from: http://www.sciencedirect.com/science/article/pii/S002228361930717X
[23] Bianco S, Chiariello AMAM, Annunziatella C, Esposito A, Nicodemi M. Predicting chromatin architecture from models of polymer physics. Chromosom Res. 2017;25(1):25–34.
[24] Fiorillo L, Musella F, Conte M, Kempfer R, Chiariello AM, Bianco S, et al. Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. Nat Methods 2021 185 [Internet]. 2021 May 7 [cited 2021 Oct 18];18(5):482–90. Available from: https://www.nature.com/articles/s41592-021-01135-1
[25] Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, et al. Cohesin Loss Eliminates All Loop Domains. Cell [Internet]. 2017;171(2):305-320.e24. Available from: https://doi.org/10.1016/j.cell.2017.09.026
[26] Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature [Internet]. 2017 Apr 6 [cited 2021 Mar 29];544(7648):59–64. Available from: https://pubmed.ncbi.nlm.nih.gov/28289288/
[27] Theobald DL. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. urn:issn:0108-7673 [Internet]. 2005 Jun 23 [cited 2021 Oct 18];61(4):478–80. Available from: http://scripts.iucr.org/cgi-bin/paper?sh5029
[28] Iannone F, Ambrosino F, Bracco G, De Rosa M, Funel A, Guarnieri G, et al. CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. In: CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019.

A FUNCTIONAL DATA ANALYSIS APPROACH TO SHIP CO_2 EMISSION PROFILE MONITORING

Antonio Lepore*, Biagio Palumbo, Christian Capezza and Fabio Centofanti

University of Naples Federico II, Department of Industrial Engineering, Piazzale Tecchio 80, 80136, Naples, Italy

ABSTRACT. In this report, we show the implementation of advanced statistical methods to face the industrial problem of monitoring the profiles acquired at each voyage of the CO_2 emission due to navigation of a Ro-pax ship. The methods are based on functional data analysis, which deal with data observed as functions defined over a continuum, such as curves or images. The infinite dimensionality of the data requires appropriate methods that are computationally intensive. We present a real-case study to show the applicability of such methods.

1 Introduction

The massive amount of data available in many applications thanks to modern sensing networks allows to provide information about complex data, such as curves, surfaces, which vary over a continuum and thus can be modelled as functional data. Functional data analysis (FDA) refers to the set of statistical methods where the observation units are functional data, which are known to be mathematical objects in an infinite dimensional space. Thorough overviews of FDA techniques are provided by Ramsay [1], Horváth and Kokoszka [2], Kokoszka and Reimherr [3]. More specific theoretical insight can be found in [4,5,6]. Although a finite basis representation is often used, we are still in a high-dimensional setting that requires appropriate statistical techniques able to face the computational burden required to analyse this type of data. In this report, we show how to perform statistical process monitoring of functional data, which is also known as profile monitoring. Motivated by an industrial application, the objective is to provide the reader with a very transparent set of steps for the statistical process monitoring of functional data in real-world case studies. We propose a procedure that is applied to a real-case study from the maritime field in monitoring CO_2 emissions from real navigation data of a roll-on/roll-off passenger cruise ship, i.e., a ship designed to carry both passengers and wheeled vehicles that are driven on and off the ship on their own wheels.

In particular, the proposed method can be divided into three main steps. Firstly, the functional data are obtained from the raw data through a smoothing technique based on spline functions. Then, a functional principal component analysis is performed to extract the relevant principal component scores. Lastly, the retained principal component scores are used in a monitoring procedure that is based on the simultaneous application of the Hotelling's T^2 and the squared prediction error (SPE) control charts in a nonparametric framework. In the following, these steps are illustrated.

2 Methods

2.1 Data Smoothing

Data are collected by devices in a discrete fashion, that is as n discrete observed curves $\{Y_i(t_j), j = 1, \dots, p\}_{i=1,\dots,n}$, where $\{t_j\}_{j=1,\dots,p}$ are the observation points in a given closed interval $\mathcal{T} \subseteq \mathbb{R}$. Hence, appropriate methods are required to convert discrete raw data $\{Y_i(t_j)\}$ into functional data $\{X_i(t)\}$ computable for any $t \in \mathcal{T}$, which are random realizations of a functional quality characteristic. A common approach consists of representing each functional datum $\{X_i(t)\}$ by introducing a basis function system, i.e., a set of K known, linearly independent functions $\Phi = (\phi_1, \dots, \phi_K)^T$, usually B-spline basis functions, that have the property that we can approximate arbitrarily well any function by

*Corresponding author. E-mail: antonio.lepore@unina.it.

taking a weighted sum or linear combination of a sufficiently large number of these functions [1]. Then, we have

$$X_i(t) = \sum_{l=1}^{K} c_{il}\phi_l(t) = c_i^T \Phi(t), \quad t \in \mathcal{T} \tag{2}$$

where c is the coefficient vector for each curve. Then, the problem of recovering the functional data $\{X_i(t)\}$ reduces to the estimation of the unknown coefficient vectors c_i for every $i = 1, \dots, n$. In particular, the coefficient vector c_i is estimated as $\hat{c}_i$ by penalized least squares:

$$\widehat{c_i} = \underset{c \in \mathbb{R}^K}{\operatorname{argmin}} \left\{ \sum_{j=1}^{p} \left(Y_i(t_j) - \boldsymbol{c}^T \boldsymbol{\Phi}(t_j) \right)^2 + \lambda \boldsymbol{c}^T \boldsymbol{R} \boldsymbol{c} \right\}, \tag{3}$$

where $\lambda > 0$ is a smoothing parameter and $\boldsymbol{R}$ is a matrix whose (i,j)-th entry is $\int_{\mathcal{T}} \phi_i^{(m)}(t)\phi_j^{(m)}(t)dt$, with $\phi^{(m)}$ the m-th derivative of ϕ. Finally, the functional data we are interested in are as follows

$$\hat{X}_i(t) = \hat{c}_i^T \Phi(t), \quad t \in \mathcal{T}. \tag{4}$$

2.2 Functional Principal Component Analysis

FPCA aims at reducing the infinite dimensionality of the functional data, by retaining a finite number L of scores $\{\xi_{il}\}_{l=1,\dots,L}$, which explain the largest part of the sample variability, for each functional observation $\{X_i(t)\}$. By assuming, without loss of generality, that $\{X_i(t)\}$ have zero mean, scores are defined as

$$\xi_{il} = \int_{\mathcal{T}} \psi_l(t) X_i(t) dt \tag{5}$$

where $\{\psi_l\}_{l=1,\dots,L}$ are weight functions referred to as functional principal components (FPCs), which are such that $\int_{\mathcal{T}} \psi_l(t)^2 dt = 1$ and $\int_{\mathcal{T}} \psi_i(t)\psi_j(t)dt = 0$, for $i \neq j$. FPCs are calculated by an iterative algorithm which at each step finds the eigenfunctions of the covariance operator of the functional data, they maximize the following mean square of the scores, i.e., their sample variance,

$$\psi_l = \underset{\psi}{\operatorname{argmax}} \sum_{i=1}^{n} \xi_{il}^2 = \underset{\psi}{\operatorname{argmax}} \sum_{i=1}^{n} \left(\int_{\mathcal{T}} \psi(t) X_i(t) dt \right)^2, \quad l = 1 \dots, L, \tag{6}$$

under the normalization and orthogonality constraints.

2.3 Monitoring Procedure

Two functional control charts are introduced based on the following T^2 and the SPE statistics. The T^2 statistic is as follows

$$T_i^2 = \sum_{l=1}^{L} \frac{\xi_{il}^2}{\lambda_l} \tag{8}$$

where $\lambda_1, \dots, \lambda_L$ are the variances of $\xi_{i1}, \dots, \xi_{iL}$ and correspond to the eigenvalues of the covariance function of X. The SPE_i statistic is defined for each i as

$$SPE_i = \int_{\mathcal{T}} \left(X_i(t) - \hat{X}_i^{PC}(t) \right)^2 dt \tag{9}$$

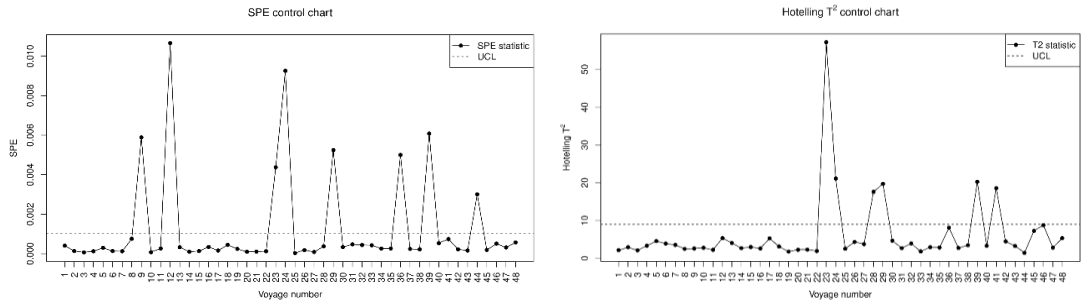
where $\hat{X}_i^{PC}$ are defined in Equation (4).

The control limits for both the T^2 and the SPE control charts are obtained as the $(1-\alpha)$-quantiles of the empirical distribution of the two statistics, based on the estimated $\{T_i^2\}$ and $\{SPE_i\}$.

3 A real-case study in the shipping industry

We illustrate the proposed monitoring procedure by means of a real-case study from the maritime field in monitoring CO_2 emissions during the navigation phase of a roll-on/roll-off passenger (Ro-Pax) cruise ship. The data analysed in this paper are a courtesy of the owner Grimaldi Group. Information about ports, name of the ship and CO_2 emissions are omitted for confidentiality reasons. Two years of data are

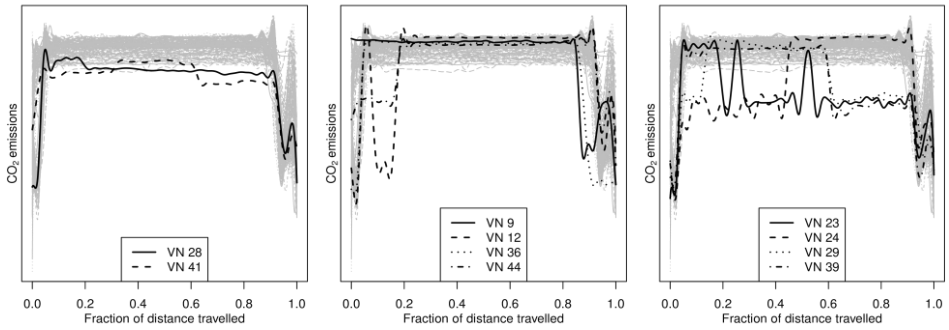
available with five-minute frequency. In the proposed application, we focus only on one route sailed by the ship to link two ports. The available data set contains the discrete values for 194 voyages of CO_2 emissions due to propulsion, which is the functional quality characteristic to be monitored at the end of each voyage. The functional domain for each voyage is the fraction of total distance travelled from the beginning of the voyage, which is a dimensionless quantity between zero and one. Figure 1 shows the T^2 and SPE control charts used for Phase II monitoring.



(a) (b)

Fig.1: (a) T^2 and (b) SPE phase II control charts. In each control chart, points joint by solid line indicate monitoring statistic values at each voyage, while dashed lines indicate upper control limit (UCL), at $\alpha = 0.05$.

Several scenarios are possible, and it is interesting to notice that the use of both control charts supports the interpretation of the type of anomalies encountered. In Figure 2 we report OC profiles against those of the training data set, which are plotted for ease of comparison as grey lines.
Note that VN 28 and 41 are OC in the T^2 control chart only (Figure 2a), VN 9, 12, 36, and 44 are OC in the SPE control chart only (Figure 2b), whereas VN 23, 24, 29, and 39 are OC in both control charts (Figure 2c). In Figure 2a, profiles of VN 28 and 41 show a clear deviation in magnitude only, that is the CO_2 emissions plot below the average. Strictly speaking, it is worth noting that lower CO_2 emissions, which are in fact desired, often are trivially associated to voyages sailed at lower-than-usual speed over ground that, in turn, imply other types of undesired costs for the shipping company due to arrival delay. Therefore, it is crucial that the proposed control charting procedure can signal these profiles. In Figure 2b, VN 9, 12, 36, and 44 show that, during most of the voyage, the CO_2 emissions were not particularly different from the reference profiles in the training data set. However, some non-negligible slowdowns are highlighted in brief parts of these voyages.



(a) (b) (c)

Fig.2: OC CO_2 emission profiles (black lines) are superimposed on phase I reference ones (grey lines) and grouped by OC in (a) in T^2 control chart only, (b) SPE control chart only, (c) both T^2 and SPE control charts.

The most important deviation from reference behaviour occurs in fact at the beginning of VN 12, which shows the largest SPE in Figure 2b. The other voyages seem to postpone the acceleration phase at the beginning of the voyage, or to anticipate the slowdown at the end of the voyage, then they show lower amounts of CO_2 emissions. More generally, these voyages show a different shape from standard Phase I profiles. Finally, In Figure 2c, with respect to the other voyages we discussed above, the CO_2 emissions for VN 23, 24, 29, and 39 show much larger deviations from the reference profiles in terms of both magnitude and shape.

4 Conclusion

In this paper, we showed benefits and practical applicability of a functional data analysis approach in real-world case studies, with a very transparent set of steps. The proposed functional control charting scheme is shown to be able to monitor CO_2 emissions in practice from real navigation data and to support the detection and interpretation of anomalous voyage profiles. Different scenarios have validated the capability of distinguishing, with respect to the reference profiles characterizing the standard operating conditions, the type of deviation based on which control charts has issued the out-of-control signal.

References

[1] J. O. Ramsay. Functional data analysis. Wiley Online Library (2005).
[2] L. Horváth and P. Kokoszka. Inference for functional data with applications. Springer Science & Business Media (2012).
[3] P. Kokoszka and M. Reimherr. Introduction to functional data analysis. CRC Press (2017).
[4] T. Hsing and R. Eubank. Theoretical foundations of functional data analysis, with an introduction to linear operators. John Wiley & Sons (2015).
[5] R. L. Eubank. Nonparametric regression and spline smoothing. CRC press (1999).
[6] D. Bosq. Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics. Springer New York (2012).

CLAR'S AROMATICITY IN MIXED SP2/SP3 STRUCTURES

Emiliano Burresi* and Maria Lucia Protopapa

ENEA, Brindisi Research Center, S.S. APPIA Km 706, Brindisi (Br)

Abstract. We report a brief discussion on the possibility to apply the Clar's aromaticity rules in hybrid sp^2/sp^3 structures, currently containing carbon and hydrogen atoms only. The goal is to verify if the stability conditions of an aromatic network of sp^2 carbons, which contain one or more benzenoid rings embedded within an sp^3 matrix are governed by Clar's aromaticity rules, commonly employed to study the stability of polycyclic aromatic molecules.

1 Introduction

The aromaticity of polycyclic systems was formulated by Erich Clar and extensively applied as Clar's π-sextet rule, also used recently to study aromaticity in graphite, graphene and carbon nanoribbons [1-4]. Briefly Clar's rule states that a polycyclic system feels a greater stabilization when it is present a large number of disjoint aromatic π-sextet, which are defined as six π-electrons localized in a single-benzene ring separated from others rings by C-C single bonds [3]. In Clar's theory four types of rings are defined: aromatic sextets, migrating sextets, empty rings and rings with localized double bonds. Molecules with 6n π electrons (n integer) and all carbon atoms in sextet rings are so-called "fully benzenoid" and present a great stability. Nevertheless, in structures not fully benzenoid (like for example Coronene) the existence of aromatic rings condensed into rings with two C-C double bonds permit a sextet migration which creates an extra current increasing the aromaticity of the structure. A single Clar's representation contains all the possible Kekulè structures of an aromatic polycyclic hydrocarbon. Here, we propose the Clar's approach for analyzing the aromaticity in the sp^2 structures embedded inside the sp^3 matrix. Clar's aromaticity has been also extensively applied to study aromaticity in Polycyclic Aromatic Hydrocarbons (PAHs). The analogy among the HOMOs orbitals of PAHs molecules and our structures act as a guide in this direction.

2 Computational Details

The geometry of the all structures analysed in this paper were fully optimized with the PWSCF (Plane Wave Self Consistent Field) code of the QUANTUM ESPRESSO (QE) [5] suite to investigate structural parameters, total energies, band gaps and indices of aromaticity. Plane-wave basis functions, exchange-correlation functional PBE generalized-gradient approximation and norm-conserving pseudopotentials were used for PWSCF calculations. The Brillouin zone is sampled using the Gamma-point only and the cutoff energy for the wave function was set at 80 Ry. A cubic box of 40 Å in side was used to apply the periodic boundary conditions and the system was placed in the center of the box.

3 Results and Discussion

In the starting structure GR_0 carbon atoms are completely saturated in a sp^3 hybridization; each carbon is bonded to other three neighbor carbons and one hydrogen whereas the carbon atoms on the edge are bonded to two carbon and two hydrogen atoms. This structure corresponds to a reduced version in size of the graphane, a novel polymer of infinite $(CH)_n$ units where all the carbon atoms are in sp^3 hybridization and the hydrogen atoms are alternatively placed up and down the carbon plane in a typical graphane chair-like structure [6]. Two notable cases of mixed sp^2/sp^3 structure are obtained by removing some hydrogen atoms in specific positions as reported in figure 1 and named GR_1 and GR_2.

*Corresponding author. E-mail: emiliano.burresi@enea.it.

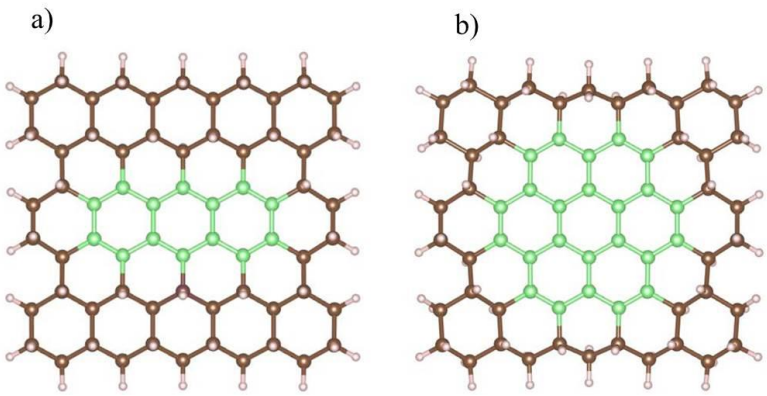


Fig. 1. GR_1 (a) and GR_2 (b) structures obtained by removing specific hydrogen atoms from carbons colored in green.

The thus obtained network of sp^2 hybridized carbon atoms form an sp^2 structure embedded inside an sp^3 matrix, corresponding to two typical aromatic structures as Anthracene (GR_1) and Coronene (GR_2). While Anthracene is almost trivial for a Clar's approach to the aromaticity, Coronene has been considered as the prototype of nanographene (NGR) or as a starting graphene model and it is the smaller specie which belongs to the PAHs class. Obviously, in our case we do not have a real structure of Coronene because the benzenoid rings are obtained inside an sp^3 carbon matrix, hence the carbons on the boundary of the aromatic rings are bonded to another carbon and not with hydrogens. Moreover, it is also possible that a sensible bond stress can derive for effect of sp^2 carbons (in green) inserted in a planar structure bonded with sp^3 carbon (in brown) which are in a no planar configuration. Kekulé structure of Coronene was reported elsewhere and each benzenoid ring was connected with other rings which have two carbon double bond which permitted a π-sextet migrations, enhancing the aromaticity, according to Clar's rule. By following Clar's rules, while for GR_1 the Clar's structure can be seen as made only by superposition of three single equivalent structure where each have a Clar's sextet, in GR_2 the Clar's structure is better represented by six sextet rings and an empty central ring which leads a greater stability. Typical approach for studying PAH molecules regards the calculations of quantitative aromatic indices which should support the Clar's rules of the aromaticity. Some of these one largely used were based on bond criteria as Pauling bond order in conjugative molecules, bond-based p-electron content ECb(R) reported in [3], and HOMA index [Harmonic Oscillator Model of Aromaticity]. One of the index widely used based on magnetic criteria was expressed by the nucleus independent chemical shifts (NICS) and ^{3}He chemical shifts, anisotropy and exaltation of magnetic susceptibility. NICS is probably the index more accurate which it has been used to construct the aromaticity patterns in PAH. HOMA index is an empirical index which was used in several studies about the chemical of the PAH and was in good agreement with NICS index. The HOMA index is defined as

$$HOMA = 1 - \frac{K}{n}\sum_{i=1}^{n}(R_i - R_{opt})^2$$

where K is a normalization constant for C-C bond fixed at 257.7, n is the number of C-C bonds involved into the summation. In the summation R_i stands for running bond length and R_{opt} is a C-C reference bond that in the primordial use of this index was taken as average value of the R_i distance; in case of C-C length was set at 1.388 Å. In table 1 we summarized the main results on geometrical parameters and HOMO-LUMO gap while in figure 2 we reported the HOMA index for both cases GR_1 and GR_2.

Table 1: Geometrical parameters (Å), HOMO-LUMO (H_L) gap (eV) and total HOMA (T_HOMA) index were reported for GR_1 and GR_2 structures.

Structure	Min	Max	Aver	T_HOMA	H-L
GR_1	1.382	1.445	1.421	0.564	2.09
GR_2	1.384	1.433	1.420	0.651	2.63

In table 1 T_HOMA index was calculated as sum on all C-C bond lengths. More the HOMA index is closer to 1 and greater is the stability for aromaticity (HOMA is set equal to 1 as reference for benzene). Moreover also the greater value of H-L gap for GR_2 with respect GR_1 ensures a great stability for GR_2. It is interesting to observe as in GR_2 the average bond length is 1.420 Å (and 1.421 Å in GR_1) which is the bond length of the graphene. This correspondence could be due to the sp^2 network in GR_2 which does not present atoms on the boundary of the structure; so it is made of graphitic carbon bonded with another carbon atoms and no bonds with hydrogen occur.

The single values of HOMA for each aromatic rings were reported in figure 2. For GR_1 all the three rings assume similar values. By doing a parallelism with Anthracene, Clar's sextet is localized on single aromatic ring obtaining three separated structure whose aromaticity is similar. We can argue that similar value for three ring is due to the extra current or migrating of sextet which contributes to the superposition of three similar Kekulé formula. In GR_2 (fig. 2) we found HOMA value very close for peripheral rings; the number of π electrons are 24 (n=4) hence only three rings are occupied by π-sextet. Also in this case the migration of π-sextets generates an extra current of stabilization and similar value of HOMA index could indicate this behavior.

We presented a simple discussion, based on HOMA index, H-L gap, and geometrical considerations, for analyzing domains made up of carbon atoms characterized by sp^2 hybridizations inside a matrix of sp^3 atoms. The existence of these domains is expected for example in reduced graphene oxide, obtained by reduction processes of graphene oxide or in carbon dots (CDs). The latest are a new class of prominent eco-friendly low cost materials with exceptional luminescent properties. Recent works show that CDs are a mixture of sp^2 and sp^3 carbons whose structure depends by synthesis procedure, including in their structures PAH and organic dyes as most likely chromophore centers of the CDs.

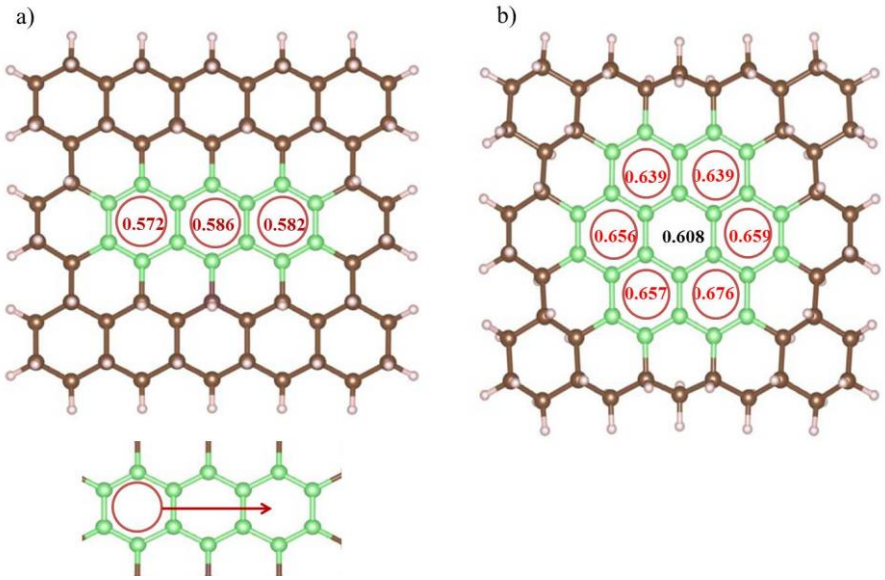


Fig. 2. HOMA index for single benzenoid rings in GR_1 (a) and GR_2(b).

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [7]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

References

[1] T. Wassmann, A.P. Seitsonen, A. M. Saitta, M. Lazzeri and F. Mauri. Clar's Theory, π-Electron Distribution, and Geometry of Graphene Nanoribbons. *Journal American Chemical Society* **132**, pp. 3440-3451 (2010).

[2] A. D. Zdetsis. Bridging the Physics and Chemistry of Graphene(s): From Huckel's Aromaticity to Dirac's Cones and Topological Insulators. *The journal of physical chemistry A* **124**, pp. 976-986 (2020).

[3] I. Gutman, Z. Tomovic, K. Mullen, J.P. Rabe. On the distribution of π-electrons in large polycyclic aromatic hydrocarbons. *Chemical Physics Letters* **397**, pp. 412-416 (2004).

[4] D. Moran, F. Stahl, H.F. Bettinger, H.F. Schaefer III, and v. R. Schleyer. Towards Graphite: Magnetic Properties of Large Polybenzenoid Hydrocarbons. *Journal American Chemical Society* **125**, pp. 6746-6752 (2003).

[5] P. Giannozzi *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J.Phys.:Condens. Matter* **21**, 395502, (2009).

[6] S. Lebegue, M. Klintenberg, O. Eriksson, and M.I. Katsnelson, Accurate electronic band gap of pure and functionalized graphane from GW calculations. arXiv:0903.0310v1 (2009).

[7] F. Iannone et al., "CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout," *2019 International Conference on High Performance Computing & Simulation (HPCS)*, Dublin, Ireland, pp. 1051-1052, (2019)
doi: 10.1109/HPCS48598.2019.9188135.

TOWARD ATOMISTIC RESOLVED PEPTIDE SEQUENCING VIA TUNNELING CURRENT IN GRAPHENE NANORIBBONS DEVICES

Giuseppe Zollo* and Tommaso Civitarese

Dipartimento di Scienze di Base e Applicate per l'Ingegneria- Sapienza Università di Roma via A. Scarpa 14-16; 00161 Roma- ITALY

ABSTRACT. Nanogaps in graphene nanoribbons are employed for atomistic re- solved amino acids recognition in polypeptides during their translocation. The transversal tunneling current is calculated using the non equilibrium Green func- tion scheme, based on the density functional theory. A periodical signal from the backbone of small neutral and polar peptides is related to the peptide bonds for both X (Asn, Ala, Asp, Ser) homo peptides and Gly-X etero peptides. The X side chain is only partially involved in the electron injection and no role is played by the polar ends. Most of the current is injected through the CαH group and the partially resonant peptide bonds. These results encourage the search of unique triggering signals related to the passage of each residue during translocation by atomic resolved tunneling currents.

1 Introduction

Efficient and cheap protein sequencing techniques are needed to support proteomics and to identify mutations that affect the protein 3D structure and behavior within living organisms. Current methods require long sequencing times that represent a serious drawback and so new sequencing techniques are searched in the context of nanoscience. Recently new nanodevices and methods have been proposed for amino acids recognition and protein sequencing, especially employing nanopores and nanogaps[1-3]. Among the various strategies, the measurement of the transversal tunneling current flowing between two nano-electrodes during the protein translocation across a nanogap is particularly appealing due to the direct atomistic scale insight of amino acids (AAs) obtained using 2D electrodes, such as graphene [4, 5]. A device made of sub-nanometer gaps between graphene nanoribbons (GNR) has been demonstrated to attain an atomistic resolution sensing of peptide bonds (PB) for Gly peptides [6]: the idea was to devote the various gaps of the array to sense different parts of the peptide chain, with the central one devoted to the backbone. This signal involves the pseudo-π and pseudo-$\pi*$ orbitals of the GNRs electrodes and the PB Molecular Orbitals (MOs). Here we report on the case of small polar AAs, with side chain size similar to Ala, to evidence the eventual role played by the side chain polarity and how it affects the PB signal.

2 Theoretical methods

The ideal device is schematically drawn in Fig. 1. Two semi-infinite narrow zig-zag hydrogenated graphene nanoribbons 2-ZGNR (H2-ZGNR) are employed as electrodes. Although their ground state is semi-conducting ferromagnetic [7], there are several examples of metallic or half-metallic stable phases in particular situations [8-10]. Then the unpolarized metallic phase of H2-ZGNR is considered

*Corresponding author. E-mail: giuseppe.zollo@uniroma1.it

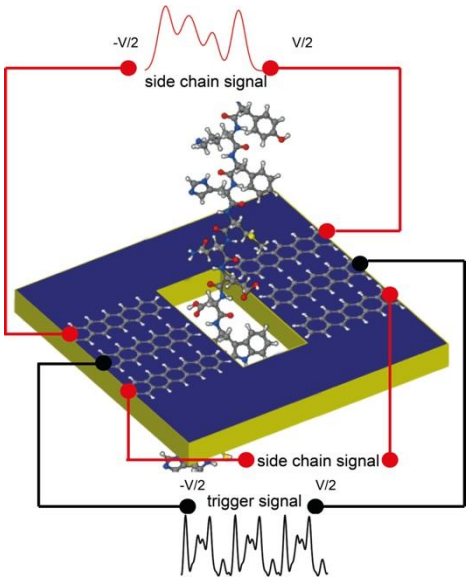


Fig. 1: Ideal device made of an array of nanogaps and GNRs onto a nano-pore. Signal related to the peptide backbone is collected from the central nanogap.

as a paradigmatic case of metallic or half-metallic H2-ZGNR electrodes. The gap devices have been fully relaxed by DFT using the Quantum ESPRESSO package [11] as detailed elsewhere [12]. Before translocation, the whole system (device, peptide and water molecules) has been equilibrated using a sequence of NVT/NPT runs. Then the peptide has been translocated via non-equilibrium steered classical molecular dynamics (SMD) in water at T = 300 K and constant velocity. Residual strain has been treated using a selection protocol [12]. Finally we have calculated the tunneling current, according to the DFT-NEGF scheme [13] as implemented in the TRANSIESTA code [14], for selected atomic configurations having chemically meaningful groups in middle of the gap, namely the carboxyl *CO* and the amino *NH* groups of the PB, the side chain (*SC*) and the Center of the Bond between the carbon bonded to the Side chain and the N/Cα atom, namely *CBSN*/*CBSC*. According to DFT-NEGF [13], the tunneling current is obtained from the transmission function $T(\varepsilon)$ through the Landauer-Buttiker formula for an external bias voltage V applied along the z direction

$$I(V) = \frac{2e}{h}\int_{-\infty}^{+\infty} T(\varepsilon)[f(\varepsilon-\mu_L) - f(\varepsilon-\mu_R)]\,d\varepsilon \text{ with}$$

$T(\varepsilon) = Tr\left[G(\varepsilon)\Gamma_L(\varepsilon)G^{\dagger}(\varepsilon)\Gamma_R(\varepsilon)\right]$ where $f(\varepsilon)$ is the Fermi-Dirac distribution function and $\mu L(R)$ is the electrochemical potential of the left(right) electrode, $G(\varepsilon) = \lim_{\eta\to 0+}(\varepsilon+i\eta-H)^{-1}$ is the Green's function of the system and $\Gamma_{L(R)}(\varepsilon) = i[\Sigma_{L(R)}(\varepsilon) - \Sigma^{\dagger}_{L(R)}(\varepsilon)]$ is the left(right) coupling function. The computational parameters employed can be found elsewhere [12] The atomic current between have also been calculated according to: $I_{n,n'} = e\int_{-\infty}^{+\infty} d\varepsilon[f_L(\varepsilon) - f_R(\varepsilon)]Tr\left[G(\varepsilon)\Gamma_L(\varepsilon)G^{\dagger}(\varepsilon)J_{n,n'}\right]$ with $J_{n,n'} = \frac{2\pi}{ih}(P_nHP_{n'} - P_{n'}HP_n)$ and the projection operator on the n site $P_n = \sum_{\gamma}|\varphi_{n,\gamma}\rangle\langle\varphi_{n,\gamma}|$ expressed in terms of a complete localized basis set $\{|\varphi n,\gamma\rangle\}$ in a subspace of the full device region where $|\varphi_{n,\gamma}\rangle$ is an atomic orbital of type γ, centered at the atomic site n [15].

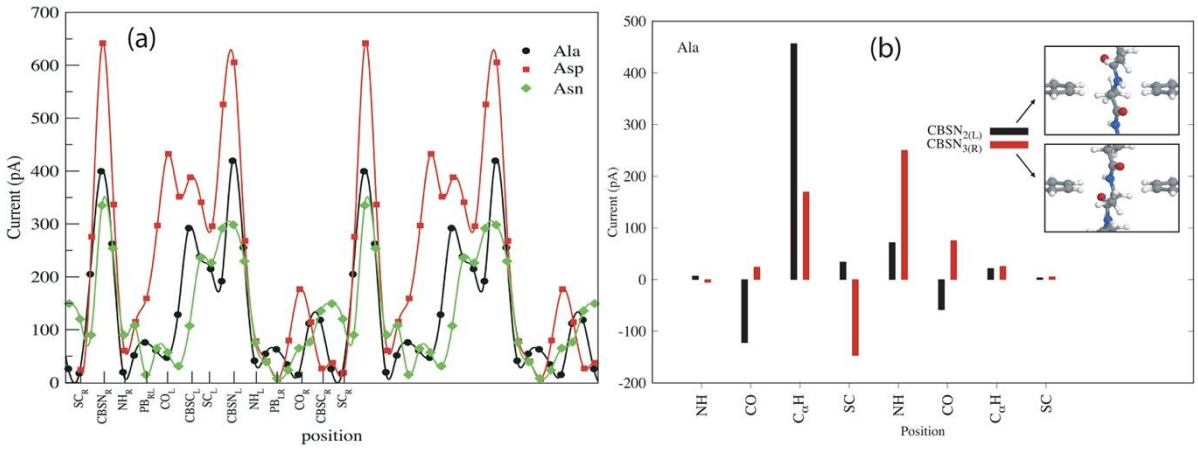


Fig. 2: Current signal calculated for Ala, Asp and Asn homo-peptides (a). Bond currents injected from the left lead into the peptide for the CBSN configurations with different orientation of the Cα-H bond either toward the left lead electron source (L) or toward the right lead electron drain (R) (b).

3 Results

In Fig. 2 we report the tunneling current measured during the translocation for Ala, Asn, and Asp homopeptides. The subscripts *L*/*R* to indicate whether the *CH* bond containing the α-carbon of the amino acid backbone protrudes towards the left (L) or the right (R) lead (see fig 2). The signal for the Ala case is dominated by one main peak per amino acid at the mid-bond between the α-carbon and the *NH* group involved in the PB. To understand the origin of the current signal observed, we have calculated the cumulative atomic bond currents injected from the left lead (electron source) into the various peptide chemical groups specified in Sec. 2 (see Fig 2(b)). This analysis evidences that the peaks at the $CBSN_{L/R}$ configurations are due to the interplay between the injection through the *NH* groups of the closest PB and of the closest *CH* groups containing the α-carbons: if the configuration favors the overlap between the pseudo-π orbital of the left lead and the partially empty p_z orbital of the N atom (involved in the resonant PB), then the injection occurs though both the $C_\alpha H$ and the *NH* group. If, on the contrary, the *NH* groups protrudes towards the right lead (and is closer to it), then the dominant injection occurs mostly through the *CαH* group because the C-H bond is oriented toward the left lead. A second minor peak occurs at the $CBSC_j$ (j=L,R) configurations, i.e. at the mid-bond between the *CO* group and the α-carbons showing the same interplay of the injection through the closest *NH* and *CH* groups. Basically the injection still proceeds through the groups protruding towards the injection lead: the electrons flow mainly through the $C_\alpha H$ group in one case or through $C_\alpha H$ and PB (via the *NH* group) in the other case. This phenomenon explains also why the currents at SC_L configurations are greater than the ones measured for SC_R [12]. All the homopeptides have the same behavior and signal periodicity except for just minor, specific peculiarities concerning the secondary peaks and satellites. For Asp and Asn we still see a structured secondary peak on the left side of $CBSC_L$ and a simple small peak on the left of $CBSC_R$ but with a certain shift that depends on the homo-peptide. All the attempts to relate the amino acid polarity to this shift have failed and we can state that there is no clear signature of the amino acid polarity on the tunneling current, at least according to the present level of theory. This is confirmed by the fact that the polar ends do not affect the peptide backbone, neither concerning the electron density not the total electrostatic potential [12]. We have also considered the cases of the Gly-X peptides with alternating Gly and X (X=Ala, Asn, Asp) AAs. As in the previous case, the neutral apolar case of the Gly-Ala peptide is taken as a reference. The tunneling currents are reported in Fig. 3. Because the relevant chemical groups here belong either to Gly (G) or to X, X=Ala (A), Asn (N), Asp (D) of the alternating hetero peptide, then we have labelled the groups

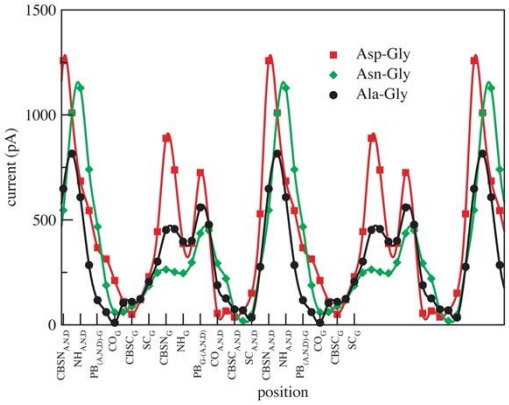


Fig. 3: Current signals calculated for the alternating Gly-X (Asp, Asn) peptides. For better clarity these signals have been superimposed to the GLy-Ala case.

accordingly instead of using the ”left/right” labels. Independently on the polarity, we see that the signal has a doubled periodicity with respect to the Ala homo-peptide, with main peaks at approximately the $CBSN_X$ (X=A, N, D) configurations and secondary double peaks, resembling the typical ones for Gly homo-peptides, occurring when the Gly residue is in the gap, namely at the $CBSN_G$ and $PBG{-}X$ configurations (the last one is the mid-bond of the PB between Gly and Ala/Asn/Asp). As evidenced in the previous literature [6], this feature is related to the current flow across the Gly CH2 group made of the Gly side chain (a simple H atom) and the *C*αH group. Concerning the main peak, instead, we can see that the electron injection occurs mainly through part of the X side chain, the *NH* and the *C*αH groups [12]. Thermal fluctuations explain the major main difference regarding the secondary double peak, and the minor difference in the main peak as well. [12]. The hetero-peptides main peaks are much higher than the corresponding homo-peptides ones with a relative increase up to nearly 300% for Asn: this is due to to a marked contribution of the side chain to the tunneling current caused by the alignment of a second *CH* bond along the ZGNR direction [12] and is a consequence of the gap constraint on the alternating sequence of Gly and the X residues. Thus peptides with polar AAs that include at least one methylene bridge seem to have reproducible behaviors and signals that are basically lead by the backbone geometry during the translocation and that appear independent on the polar nature of the side chain. Among the smaller polar peptides, we have analyzed Serine (Ser, S) as a paradigmatic case of this category as reported in Fig. 4 for the Ser homo-peptide (a) and the Gly-Ser hetero-peptide (b) The main difference between the Ser homo-peptide and the reference Ala case is in an extra peak at the mid position of the PB between the ”left” and the ”right” oriented residues. The extra peak is due to the additional coupling between the electrodes through the Ser side chain. However the atomic current branches injected from the left lead (not shown) show that the main contribution comes from hydrogen atom oriented towards the left source electrode [12]. Then no role is played by the side chain polar hydroxyl. The second main difference is the poor current signal at $CBSN_R$ that, contrarily to the Ala reference case, is much smaller than the *CBSNL* one. The described interplay of the current injected through the $C_\alpha H$ and *NH* groups still holds, with an enhanced reflection at the bottom carboxyl group that lowers the injected current. The Gly-Ser case follows nicely the reference Gly-Ala case confirming that the polar nature of the side chain affects negligibly the main features of the signal thus showing again a double residue periodicity (see Fig.4(b)).

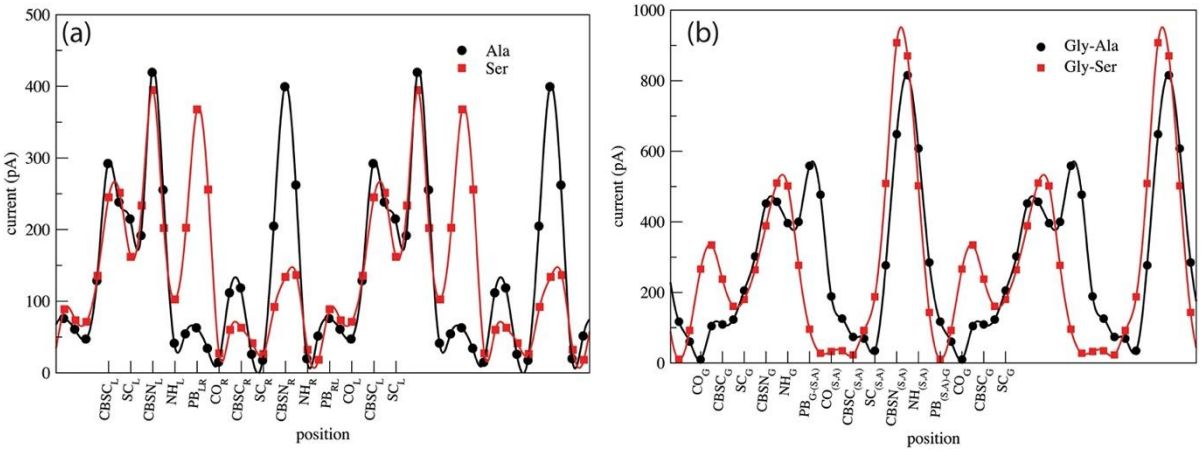


Fig. 4: Tunneling current calculated during the translocation of Ser homo-peptides (a) and Gly-Ser hetero-peptides (b).

4 Conclusions

As concluding remarks the signal from the peptide backbone of small neutral polar (Asp, Asn, Ser) and apolar (Ala) AAs is characterized by single/double residue periodicity with the main peaks when the mid-bond of the backbone *NC* bond is in the ZGNR gap. The current is injected into the peptide through the closest $C_{\alpha}H$ and *NH* groups that are out of the ZGNR plane, the relative extent depending on the orientation of the C_{α}-H bond. The signal not show any marked dependence on the side chain polarity and thus the same trigger signal id expected also for larger AAs. Smaller AAs show different tunneling current features (due to the stronger coupling) that, however, are not affected by the side chain polarity. The peculiarities of the backbone signal coming from the neutral and polar peptides enforce the idea that peptide/protein primary structure sequencing by tunneling current across nanogaps in graphene nano ribbons could be triggered with atomistic resolution by looking at the tunneling signal coming from the peptide backbone.

Acknowledgments

Computing resources and the related technical support used for this work have been provided also by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [16]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programmes, see http://www.cresco.enea.it/english for information.

References

[1] Eamonn Kennedy, Zhuxin Dong, Clare Tennant, and Gregory Timp. Reading the primary struc- ture of a protein with 0.07 nm3 resolution using a subnanometre-diameter pore. *Nat Nano*, 11(11):968–976, 11 2016.

[2] Takahito Ohshiro, Makusu Tsutsui, Kazumichi Yokota, Masayuki Furuhashi, Masateru Taniguchi, and Tomoji Kawai. Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nature nanotechnology*, 9(10):835–840, 2014.

[3] Michael Zwolak and Massimiliano Di Ventra. Electronic signature of dna nucleotides via trans- verse transport. *Nano Letters*, 5(3):421–424, 03 2005.

[4] Stephanie J. Heerema and Cees Dekker. Graphene nanodevices for dna sequencing. *Nature Nanotechnology*, 11:127 EP –, 02 2016.

[5] James Wilson, Leila Sloman, Zhiren He, and Aleksei Aksimentiev. Graphene nanopores for protein sequencing. *Advanced Functional Materials*, 26(27):4830–4838, 2016.

[6] Aldo Eugenio Rossini, Fabrizio Gala, Mauro Chinappi, and Giuseppe Zollo. Peptide bond detec- tion via graphene nanogaps: a proof of principle study. *Nanoscale*, 10(13):5928–5937, 2018.

[7] Young-Woo Son, Marvin L. Cohen, and Steven G. Louie. Energy gaps in graphene nanoribbons. *Phys. Rev. Lett.*, 97:216803, Nov 2006.
[8] Daniel Gunlycke, Denis A. Areshkin, Junwen Li, John W. Mintmire, and Carter T. White. Graphene nanostrip digital memory device. *Nano Letters*, 7(12):3608–3611, 12 2007.
[9] Young-Woo Son, Marvin L. Cohen, and Steven G. Louie. Half-metallic graphene nanoribbons. *Nature*, 444(7117):347–349, 2006.
[10] Daniel J. Rizzo, Gregory Veber, Jingwei Jiang, Ryan McCurdy, Ting Cao, Christopher Bronner, Ting Chen, Steven G. Louie, Felix R. Fischer, and Michael F. Crommie. Inducing metallicity in graphene nanoribbons via zero-mode superlattices. *Science*, 369(6511):1597–1603, 2020.
[11] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(19):395502–1–19, 2009.
[12] Tommaso Civitarese and Giuseppe Zollo. Triggering amino acid detection by atomistic resolved tunneling current signals in graphene nanoribbon devices for peptide sequencing. *ACS Applied Nano Materials*, 4(1):363–371, 01 2021.
[13] Supriyo Datta. *Electronic Transport in Mesoscopic Systems*. Cambridge Studies in Semiconductor Physics and Microelectronic Engineering. Cambridge University Press, 1995.
[14] K. Stokbro, J. Taylor, M. Brandbyge, and P. Ordejo´n. Transiesta - a spice for molecular electronics. *Annals of the New York Academy of Sciences*, 1006(1):212–226, 2003.
[15] Nick Papior, Nicols Lorente, Thomas Frederiksen, Alberto Garca, and Mads Brandbyge. Improvements on non-equilibrium and transport green function techniques: The next-generation transiesta. *Computer Physics Communications*, 212(Supplement C):8 – 24, 2017.
[16] F.Iannone,F.Ambrosino,G.Bracco,M.DeRosa,A.Funel,G.Guarnieri,S.Migliori,F.Palombi, G. Ponti, G. Santomauro, and P. Procacci. Cresco enea hpc clusters: a working example of a multifabric gpfs spectrum scale layout. In *2019 International Conference on High Performance Computing Simulation (HPCS)*, pages 1051–1052, 2019.

HEAT CAPACITY OF PUO2 AT HIGH TEMPERATURE: RESULTS OF MD CALCULATIONS

Rolando Calabrese

ENEA, Safety and Sustainability of Nuclear Energy Division, I-40129, Bologna, Italy

ABSTRACT. A new generation of fast breeder reactors (FBRs) is under development with the objective of making nuclear energy more sustainable. Most promising reactor designs are loaded, at least during their early phase of deployment, with UO_2-PuO_2 mixed oxide fuel (MOX). Values of PuO_2 concentrations that are considered in designs span up to 30 mol%. Heat capacity is important for the evaluation of thermal conductivity and performance under transient conditions. A lack of experimental measurements especially at high temperatures makes drawing final conclusions on PuO_2 complex. This paper presents results based on MD calculations aimed at discussing the behaviour of PuO_2 heat capacity from 2000 K up to the melting temperature. For the purpose, we have applied a theoretical model that considers three contributions to the heat capacity.

1 Introduction

Next generation FBRs aim at improving nuclear energy sustainability by overcoming, among others, shortages in uranium resources. New reactors' designs will require the use of MOX fuel with a concentration of plutonium dioxide up to 30 mol% [1]. In their study Kato and Matsumoto confirm that the thermophysical properties of MOX fuel are relevant for safety assessment under the demanding conditions typical of FBRs (e.g., high fuel temperature and high burn-up) [2]. If, on the one hand, the heat capacity of UO_2 has been deeply studied, on the other hand, toxicity, high radiation level, and behaviour at high temperature are all factors that make measurements of PuO_2 thermophysical properties complex [3]. Several authors have discussed the heat capacity of PuO_2 and models are available in the literature [4-7]. A lack of experimental measurements above 2370 K has raised questions on the behaviour of PuO_2 heat capacity in the high temperature domain extending from 2400 K up to the melting temperature. A wide consensus has been reached on the assumption that the formation of oxygen Frenkel pairs (OFPs) plays a relevant role for plutonium dioxide heat capacity [2,7-8]. A theoretical model based on statistical mechanics predicts a pronounced increase in heat capacity at high temperatures [8] and recommendations are given in this sense [7]. This paper presents MD calculations on the heat capacity of PuO_2. The interatomic potential used in this analysis has been developed by Uchida et al. [9]. MD calculations have been carried out by means of the LAMMPS code (v. 2019) [10]. Code runs have been carried out on the CRESCO6 cluster [11]. This machine supports the MPI message-passing library giving the opportunity to take advantage of the built-in parallel structure of the code. LAMMPS uses spatial-decomposition techniques to partition the simulation domain into small sub-domains of equal computational cost, one of which is assigned to each processor.

2 Details of calculations

The interatomic potential employed in calculations is based on the model that was presented in [9]. Authors have developed a Born-Mayer-Huggins pair potential (BMH) that is consistent with their measurements on the thermal expansion of PuO_2. The analytical expression of BMH potential partially ionic is presented in Eq. 1. Each of the terms in Eq. 1 accounts for different types of interactions between ions: the first for long-range Coulombic; the second and the third for short-range interactions due to Pauli's repulsion principle and van der Waals forces, respectively. The level of iconicity fixes the effective electronic charges z_i and z_j; r_{ij} stands for the distance between ion i and ion j. A Morse

*Corresponding author. E-mail: rolando.calabrese@enea.it

potential accounts for the covalent bond between anion and cation; see Eq. 2. In this equation r_{ij}^{*} is the covalent bond length. D_{ij} and β_{ij} determine the depth and shape of potential.

$$U_{ij}(r_{ij}) = \frac{z_i z_j e^2}{4\pi\epsilon_0 r_{ij}} + A_{ij} exp\left(-\frac{r_{ij}}{\rho_{ij}}\right) - \frac{C_{ij}}{r_{ij}^6} \qquad (1)$$

$$U_{ij}(r_{ij}) = D_{ij}\{exp[-2\beta_{ij}(r_{ij} - r_{ij}^{*})] - 2exp[-\beta_{ij}(r_{ij} - r_{ij}^{*})]\} \qquad (2)$$

Coefficients used in calculations are based on the potential published by Uchida et al. [9]. They have been slightly modified to comply with a value of lattice constant at room temperature (300 K) of 5.3946 Å. Concerning iconicity, its value (0.565) sets the effective charge of plutonium to +2.26 and the corresponding quantity of oxygen to -1.13. Coefficients of the model are listed in Tab. 1.

Table 1: Coefficients of the interatomic potential.

Pair (BMH)	A_{ij} (eV)	ρ_{ij} (Å)	C_{ij} (eV·Å ^{6})
O-O	483029.685	0.178	96.0
Pu-Pu	0.0	0.200	0.0
Pu-O	5700.0	0.252	0.0
Pair (Morse)	D_{ij} (eV)	β_{ij} (Å^{-1})	r_{ij}^{*} (Å)
Pu-O	0.165	2.0	2.37

The supercell used in calculations is composed of 12000 atoms arranged in a face-centered cubic lattice (*fcc*). The elementary cell, formed by 12 atoms, has been replicated 10x10x10 along the x, y, z orthogonal axes. Periodic boundary conditions have been applied. Energy minimization was performed by means of a steepest descent algorithm. The Nose-Hoover thermostat/barostat has been used to reach equilibrium (30 ps) [10]. Enthalpy and lattice parameters have been recorded during a period of 30 ps. Calculations have been performed in the NPT and NVT ensembles at different temperatures. In the low temperature domain (300-1500 K) calculations have been performed with an interval of 100 K. From 1500 K up to 2800 K the interval between measurements has been reduced to 50 K.
The theoretical model applied in our analysis employs three contributions to the heat capacity of PuO_2: phonon C_{ph}, dilation C_d, and formation of defects C_{df}; see Eq. 3 [12]. According to the indications of the literature, it is assumed that the third term is strongly correlated with the creation of OFPs at high temperature.

$$C_p = C_{ph} + C_d + C_{df} \qquad (3)$$

The dilation term is presented in Eq. 4. In this equation $\alpha(T)$ is the coefficient of volumetric thermal expansion and γ the Grüneisen constant; T is the temperature.

$$C_d = \alpha(T)\gamma C_{ph}(T)T \qquad (4)$$

C_p and $\alpha(T)$ have been evaluated from enthalpy and lattice constant that have been determined in NPT calculations. The phonon term C_{ph} has been evaluated by means of NVT calculations as the derivative of total energy. As shown in Eq. 4, $\alpha(T)$ and C_{ph} together with the Grüneisen constant are used in the evaluation of C_d. Values of the Grüneisen constant of plutonium dioxide are reported in [13] (2.30), [14] (1.62), and [15] (2.60). In our results, the Grüneisen constant was set to 1.62. After that, the contribution due to the formation of defects can be derived from Eq. 3.

3 Results

Values of C_p, C_{ph}, and C_{ph} plus C_d in the temperature interval 800-2800 K are presented in Fig. 1 (left). In this figure are also included for comparison the curves recommended for UO_2 and PuO_2 [7]. Fig. 1 (right) presents the contribution due to the formation of defects resulting from MD recalculations. In this figure are also shown three curves derived from the model proposed by Konings and Beneš [8]. Their model is presented in Eq. 5.

$$C_{p,exc} = \frac{\Delta H_{OFP}^2}{\sqrt{2}RT^2} exp\left(\frac{-\Delta H_{OFP}}{2RT}\right) exp\left(\frac{\Delta S_{OFP}}{2R}\right) \qquad (5)$$

In Eq. 5 ΔH_{OFP} and ΔS_{OFP} are the enthalpy and entropy of oxygen Frenkel pair formation, R is the universal gas constant. The formation enthalpy and entropy of oxygen Frenkel pair of plutonium dioxide are 279 kJ mol^{-1} and 66.7 J mol^{-1} K^{-1}, respectively [8]. The resulting curve is included in Fig. 1 (right). Predictions of the model assuming a formation enthalpy of OFP of 295 and 380 kJ mol^{-1} have been calculated. These curves are shown in Fig. 1 (right).

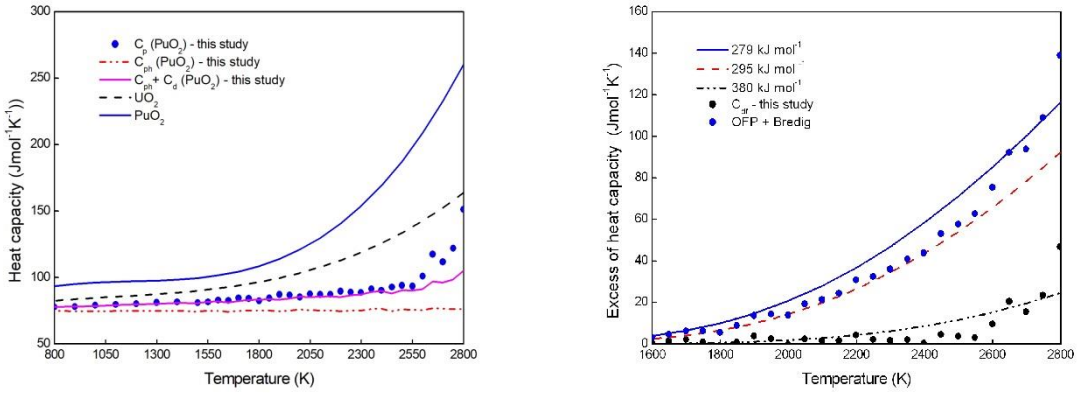


Fig. **1**: Comparison of recommended curves [7] with MD calculations (left); excess of heat capacity due to defects formation: comparison of MD calculations with the model in Eq. 5 having the formation enthalpy of OFP as parameter (recommended: 279 kJ mol^{-1} [8]).

4 Discussion and conclusion

Results presented in Fig. 1 (left) indicate that MD calculations show a nearly constant deviation from the recommended curve up to 1500 K. This behaviour is consistent with the occurrence of the Schottky effect. The literature has confirmed the existence of additional energy levels available for electrons with a consequent increase in heat capacity [2,16]. At higher temperatures the deviation is larger. A curve complying with Eq. 5 and obtained by assuming a value of OFP formation enthalpy of 380 kJ mol^{-1} has been calculated. It provides predictions consistent with our calculations; see Fig. 1 (right). However, MD calculations suggest a threshold behaviour and a roughly linear dependence on temperature that do not agree with the model [8]. The existence of the Bredig transition at temperatures around 85% of the melting temperature has been reported for actinides oxides [4,6]. Therefore, stated that the recommended curve has been obtained by fitting available experimental data, these calculations could suggest that in this temperature domain OFP formation and pre-melting transition could co-exist. In Fig. 1 (right) it is proposed a combination of these two processes that maintains a reasonable agreement with the recommended curve suggesting an increase in the formation enthalpy and a linear dependence of heat capacity on temperature approaching to melting. In conclusion, our MD calculations confirm the need for considering additional terms to predict the heat capacity of $PuO_2$with good accuracy (Schottky effect, formation of OFPs). The inclusion of a third additional contribution due to the pre-melting transition suggests that the formation enthalpy of OFP is higher than reported in [8].

References

[1] A.K. Sengupta, K.B. Khan, J. Panakkal, H.S. Kamath, S. Banerjee, Evaluation of high plutonia (44% PuO_2) MOX as a fuel for fast breeder test reactor, *Journal of Nuclear Materials* **385**, pp. 173-177, (2009).

[2] M. Kato, T. Matsumoto, Thermal and Mechanical Properties of UO_2 and PuO_2, NEA/NSC/ R(2015)2, pp. 172-177, OECD/Nuclear Energy Agency, Paris, (2015).

[3] H. Balboa, L. Van Brutzel, A. Chartier, Y. Le Bouar, Assessment of empirical potential for MOX nuclear fuels and thermomechanical properties, *Journal of Nuclear Materials* **495**, pp. 67- 77, (2017).

[4] J.K. Fink, Enthalpy and Heat Capacity of the Actinide Oxides, *International Journal of Thermophysics* **3**(2), pp. 165-200 (1982).

[5] J.-M. Bonnerot, Proprietes Thermiques des Oxydes Mixtes d'Uranium et de Plutonium, Ph.D. thesis, CEA-R-5450, (1988).

[6] J.H. Harding, D.G. Martin, P.E. Potter, Thermophysical and thermochemical properties of fast reactor materials, EUR 12402 EN, Commission of the European Communities, Luxemburg, (1989).

[7] R.J.M. Konings, O. Beneš, A. Kovăcs, D. Manara, D. Sedmidubsky, L. Gorokhov, V.S. Iorish, V. Yungman, E. Shenyavskaya, E. Osima, The Thermodynamic Properties of the f-Elements and their Compounds. Part 2. The Lanthanide and Actinide Oxides, *Journal of Physical and Chemical Reference Data* **43**, p. 013101, (2014).

[8] R.J.M. Konings, O. Beneš, The heat capacity of NpO_2 at high temperatures: The effect of oxygen Frenkel pair formation, *Journal of Physics and Chemistry of Solids* **74**(5), pp. 653-655, (2013).

[9] T. Uchida, T. Sunaoshi, K. Konashi, M. Kato, Thermal expansion of PuO_2, *Journal of Nuclear Materials* **452**, pp. 281-284, (2014).

[10] S. Plimpton, Fast Parallel Algorithms for Short-Range Molecular Dynamics, *Journal of Computational Physics* **117**, pp. 1-19, (1995).

[11] High Performance Computing on CRESCO infrastructure: research activities and results 2019, Italian National Agency for New Technologies Energy and Sustainable Economic Development, Rome, (2020).

[12] K. Kurosaki, K. Yamada, M. Uno, S. Yamanaka, K. Yamamoto, T. Namekawa, Molecular dynamics study of mixed oxide fuel, *Journal of Nuclear Materials* **294**, pp. 160-167, (2001).

[13] S. Hirooka, M. Kato, Sound speeds in and mechanical properties of $(U,Pu)O_{2-x}$, *Journal of Nuclear Science and Technology* **55**(3), pp. 356-362, (2017).

[14] H. Serizawa, Y. Arai, An examination of the estimation method for the specific heat of TRU dioxides: evaluation with PuO_2, *Journal of Alloys and Compounds* **312**, pp. 257–264, (2000).

[15] R.B. Jr Roof, An experimental determination of the characteristic temperature for PuO_2, *Journal of Nuclear Materials* **2**(1), pp. 39–42, (1960).

[16] L. Manes, in: 4th Int. Conf. Plutonium and Other Actinides, Vol. 1, T.M.S.O.A.W.N. Miner, Santa Fe, NM, USA, p. 254, (1970).

NEUTRONIC ANALYSES OF THE PFC OF DEMO DIVERTOR 2020

Simone Noce[1*], Davide Flammini[2], Giuseppe Mazzone[2], Fabio Moro[2], Francesco Romanelli[1], Rosaria Villari[2] and Jeong-Ha You[3]

[1]*University of Rome "Tor Vergata", Industrial Engineering Department, I-00133, Rome, Italy*

[2]*ENEA, Department of Fusion and Nuclear Safety Technology, I-00044 Frascati (Rome), Italy*

[3]*Max-Planck-Institut für Plasmaphysik, Boltzmannstrasse 2, 85748, Garching, Germany*

ABSTRACT. Detailed nuclear analyses have been carried-out in 2020 on the Plasma Facing Components (PFCs) of DEMO divertor, using MCNP5 Monte Carlo code on HPC CRESCO cluster, in order to assess their performance response under the expected nuclear loads and harsh operational conditions. The latest design configuration of the ITER-like PFCs has been studied comparing the effects of two different blankets (WCLL and HCPB) on the PFCs nuclear loads. Two DEMO MCNP models have been reproduced, consisting of a semi-heterogenous representation of the blanket (one WCLL and the other one HCPB), both with a fully heterogenous description of the divertor and PFCs. Nuclear heating, cumulated damage and helium production have been calculated in all divertor PFCs targets.

1 Introduction

In the framework of the work package 'Divertor' (WPDIV) within the EUROfusion Power Plant Physics and Technology (PPPT) program, detailed neutronics analyses have been performed aimed at evaluating the divertor ITER-Like PFCs nuclear responses to provide inputs for thermo-hydraulic and thermo-mechanical analyses and to assess its shielding and structural performances. A heterogeneous neutronics model of DEMO divertor (including both the inner and outer target) has been developed and integrated in the DEMO MCNP model with a Water Cooled Lithium Lead (WCLL) and Helium Cooled Pebble Bed (HCPB) blankets.

Nuclear heating density, cumulated damage and helium production have been calculated in all divertor PFCs, comparing the effect of the different blanket type and composition, using MCNP5v1.60 [1] code and JEFF 3.3 [2] nuclear data library on HPC CRESCO-6 cluster.

2 DEMO MCNP models description

The 2020 analyses have been performed using WCLL and HCPB DEMO MCNP geometrical models with semi-heterogeneous blanket representations. The WCLL BB design is based, as for on single-module-segment (SMS) concept where a basic breeding unit (BU), is replicated in Inboard and Outboard along the poloidal direction. The single BU includes the FW and side walls, top and bottom caps, internal stiffening and baffle plates, BSS, cooling pipes, LiPb manifolds and water manifolds for the FW and breeding zone cooling [3]. Taking into account the complexity of the BU cooling pipes assembly, the WCLL BB MCNP modelling approach has been based on the segmentation of the BB into radial layers with specifically defined material mixtures [4]. A similar modelling approach has been used for the HCPB, representing the BU (consisting of FW and side walls, top and bottom caps, internal BZ, ribs, BSS, purge gas manifold, fuel breeder pin) [5], into radial layers with specifically defined material mixtures [4]. The detailed MCNP model of the entire ITER-Like PFCs (design in Fig.2) targets [6,7] have been reproduced and integrated in the divertor model both with WCLL and HCPB blanket (above-mentioned). These MCNP models are shown in Fig.1.

[*]Corresponding author. E-mail: simone.noce@uniroma2.it.

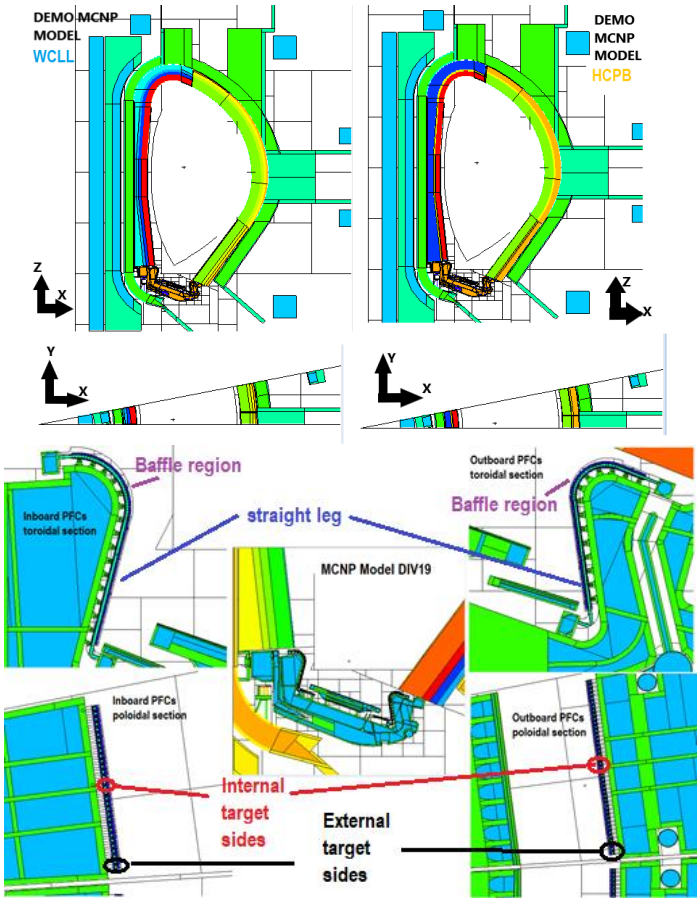


Fig. 1: WCLL (upper left) and HCPB (upper right) DEMO SMS semi-heterogeneous MCNP models with heterogeneous divertor 2019 CB model (central bottom), detailed IL PFCs inboard (bottom left) and outboard (bottom right) targets representation.

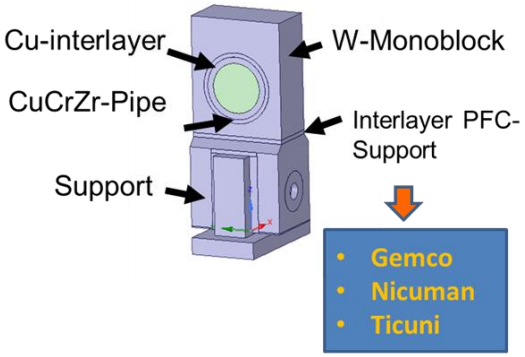


Fig. 2: ITER-Like PFCs reference design [6,7].

3 Nuclear analyses results

Detailed assessment of the spatial distribution of the nuclear loads on the DEMO divertor PFCs targets for both the BB cases during operations has been performed with MCNP5. The results are provided along the poloidal profiles as function of a curvilinear abscissa which starts from the upper part of the targets (baffle region, see Fig.1) and directed towards the lower part (Fig.3); in particular the peak values of each nuclear load are summarized in the table shown in Fig.4.

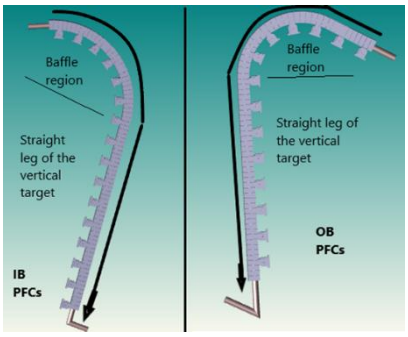


Fig. 3: Inboard and outboard poloidal profiles of the PFCs, in which is highlighted the curvilinear abscissa used in the graphs, which originates in the upper part of the target and it is directed downwards, along the profiles.

Absolute maximum nuclear loads	Helium production (appm/FPY)			Nuclear damage (dpa/FPY)			Nuclear heating density (W/cm^3)			Peak position target # (external sides)			Peak position target # (external sides)		
	WCLL	HCPB	WCLL/ HCPB	WCLL	HCPB	WCLL/H CPB	WCLL	HCPB	WCLL/H CPB	WCLL			HCPB		
W Monoblock	1.88	1.90	0.99	2.10	1.85	1.14	21.40	19.00	1.13	18th OB	13th IB	24th OB	18th OB	18th OB	21th IB
Cu/CuCrZr interlayer/ pipe	55.6	56.3	0.99	8.60	7.18	1.20	8.70	7.84	1.11	19th OB	13th IB	19th IB	17th OB	13th IB	19th IB
Eurofer PFC-CB support	58.03	55.8	1.04	5.07	4.41	1.15	7.29	6.45	1.13	4th IB	3rd IB	4th IB	4th OB	4th OB	4th IB

Fig. 4: Comparison between the absolute maximum nuclear loads (helium production, nuclear damage, nuclear heating density) considering both the WCLL and HCPB case, together with the indication of the target points where they are achieved.

Every graph shows the poloidal profiles comparing both the external and internal (see Fig.1) sides of the targets loads, both for WCLL and HCPB blanket. Each point is representative of the nuclear load value on each component (W, Cu/CuCrZr and Eurofer) along the profile. The number of components is 78 (IB) and 70 (OB) W-Monoblocks/Cu-Interlayers/CuCrZr pipe, 16 (IB) and 14 (OB) Eurofer supports. By way of example, only the graph related to the nuclear damage on the inboard Cu/CuCrZr components (which is of great interest for the design of the cooling pipes) is shown, among the many cases available. The poloidal profiles of nuclear damage on inboard Cu/CuCrZr components are shown in Fig.5. Table shown in Fig.4, highlights the absolute maximum nuclear loads values on the DEMO divertor targets, comparing them between the WCLL and HCPB case. These peak values are in correspondence of the target position (the numbering follows that of the graphs) reported in the last columns, which are located on the external sides of the target (see Fig.1). The absolute minimum values are reached on the lower zone of the target straight leg.

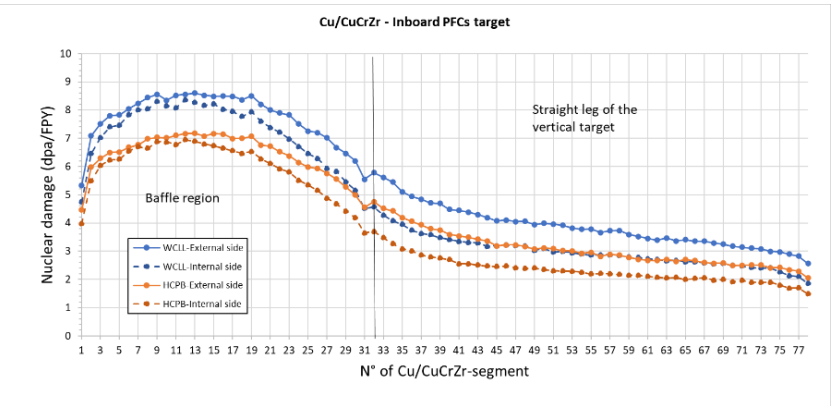


Fig. 5: Inboard poloidal profiles of the nuclear damage (dpa/FPY) on Cu/CuCrZr, comparing the external and internal target sides values for both WCLL and HCPB blanket cases.

These analyses lead to the following main conclusions: the symmetrical behaviour with respect to the central part of the target is also confirmed. The differences between internal and external are more pronounced on the straight leg (e.g., about +50/60% as peak difference), and more attenuated on the baffle region (e.g., around +5/10%), especially on the outboard targets. WCLL vs. HCPB: the different type of blanket has an impact on the neutron spectrum (water vs coolant) impinging on the targets. He-production appears to be not significantly affected by the blanket configuration. This is not true for Eurofer, which shows higher He-production levels with WCLL (+3/5% on baffle region, +5/10% on straight leg) than HCPB. Nuclear damage seems to be higher with the WCLL blanket than HCPB. These differences amount to about +15%/10% on W, +20% on Cu/CuCrZr, +15% on Eurofer (baffle region) and around +20% on W, +20/30% on Cu/CuCrZr and +20% on Eurofer (straight leg). Nuclear heating density shows a trend similar to the nuclear damage. This is due to the greater gamma generation, whose production on the divertor targets is favoured by the water-cooled blanket. These differences are more reduced in the baffle region (+10% on W, +10% on Cu/CuCrZr, +10% on Eurofer) and mitigated on the straight leg zone (+20% on W, +15/20% on Cu/CuCrZr, and Eurofer).

References

[1] X-5 Monte Carlo Team: MCNP - A General Monte Carlo N-Particle Transport Code, Version 5, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, April 2003.
[2] JEFF-3.3 evaluated data library, https://www.oecd-nea.org/dbdata/jeff/jeff33/index.html.
[3] Internal Deliverable BB-3.2.1-T007-D001: WCLL BB design and Integration studies 2020 activities https://idm.euro-fusion.org/?uid=2LF79B.
[4] Internal Deliverable DIV-1-T007-D014: Neutronic analyses of the PFC of DEMO divertor 2020 https://idm.euro-fusion.org/?uid=2P5PL4.
[5] Internal Deliverable BB-1.2.1-T006-D001: HCPB Design and Integration Studies 2019 (EFDA_D_2NKC7G v1.0) https://idm.euro-fusion.org/?uid=2NKC7G.
[6] Report on DIV-JUS-2-CD1 Target Analysis Guidelines & Structural Integrity Assessment https://idm.euro-fusion.org/?uid=2PAV56&action=get_document.
[7] Report on DIV.TAR.JUS-2-CD1 Target concepts detailed design description document https://idm.euro-fusion.org/?uid=2NVYGP&action=get_document.

CFD HYDRODYNAMICS STUDY OF INTERNALLY CIRCULATING BUBBLING FLUIDIZED BED GASIFIER

Emanuele Fanelli*

Laboratory thermochemical process for biomass, residues and wastes valorization, Division of Bioenergy, Bio-refinery and Green Chemistry, ENEA Trisaia Centre, 75026 Rotondella (MT), Italy.

ABSTRACT. Biomass gasification is nowadays considered a viable option for clean energy production. Furthermore, still more efforts need to be spent to make this technology fully available at commercial scale. Drawbacks that greatly limit the full-time plant availability—and so its economically feasibility—mainly concerns syngas purification by contaminants such as tars. Among technologies, fluidized beds are surely the most promising architectures for power production at thermal scale above 1 MW_{th}. Gasifier can be surely considered the key component of the whole power plant and its proper design, the main engineering effort. This process involves different engineering aspects: thermo-structural, heat, and mass transfer, and chemical and fluid-dynamic concerns being the most important. In this study, with the aim to reach an optimal reaction chamber design, the hydrodynamics of a bubbling fluidized bed reactor was investigated by using a CFD approach. A Eulerian–Eulerian multiphase model, supported by experimental data, was implemented to describe the interactions between the solid and fluid phases inside the reactor while a discrete dense phase model (DDPM) was considered to investigate momentum exchange among continuous phases and solid particles simulating char. Different process parameters, such as the bed recirculation rate and the particles circulation time inside the bed, were at least analyzed to characterize the hydrodynamics of the reactor. Results indicate that the recirculation time of bed material is in the order of 6–7 s at bench scale and, respectively, of 15–20 s at full scale. Information about solid particles inside the bed that should be used to avoid elutriation and agglomeration phenomenon, suggest that the dimension of the mother fuel particles should not exceed the value of 5–10 mm.

1 Introduction

Amongst all energy sources for power generation, renewables had in the last decade the highest rate of growth. IEA (International Energy Agency) projections in its SDS (Sustainable Development Scenario), project bioenergy share by 2030 from by the actual 590 TWh (2019) to about 1170 TWh (2030), with an annual increase of 6% (electricity generation from bioenergy) [1]. Moreover, it is expected that several policies and market initiatives will encourage this trend. In some countries such as China, India and Brazil are implementing similar policies to encourage the development of bioenergy as a way to support in a sustainably manner the to meet their ever-increasing energy demand.

Amongst different technologies available today for biomass conversion, gasification is considered to have the greatest potential due its higher thermochemical conversion efficiency [2]. Gasification is a thermochemical process that occurs inside reactors known as gasifiers where solid feedstocks are converted in gaseous form at high temperature (typically in the range of 700–900 °C). Via gasification, a fraction of the initial chemical energy is stored into the producer gas: in such processes

*Corresponding author. E-mail: emanuele.fanelli@enea.it

(**a**) (**b**)

Fig. **1**: ENEA's 1 MW_{th} Internally Circulating Bubbling Fluidized Bed (ICBFB) pilot plant: (a) Gasifier details; (b) Whole pilot plant layout: it is visible the biomass feeding system, the gasifier, the heat exchanger and the syngas cleaning system.

about 80% (cold efficiency) of the original energy is converted. Gaseous form is most suitable for different kinds of application for power production such as internal combustion engines, fuel cells, and gas turbines. Furthermore, syngas can be converted to fuel catalytically via FT (Fischer Tropsch) synthesis [3]. Depending on the end-use of syngas on power scale plant, though different gasifier architectures can be considered but mainly two main groups, namely, fixed beds and fluidized beds, are available. Common issue of both fixed and fluidized bed reactors is the large quantities of tar in the producer gas that compromises the plant availability. Other than the selection of the right operational conditions and the injection of additives (to prevent also the formation of ashes agglomerates) or catalysts during gasification, the approach used in the gasifier design plays a key role in tar formation mechanisms. In the last two decades efforts have been made to investigate the hydrodynamic of fluidized bed reactors, both experimentally and numerically. The main aim is to well understand phenomenon involved during gas–solid interactions as well as to know how these can influence the reactor behavior. Recent developments in numerical techniques and the availability of computational infrastructures, make the application of CFD affordable to investigate design of fluidized bed reactors and their evaluation. In this field two types of methods can mainly be applied: the Eulerian–Eulerian (E-E) approach and the Eulerian–Lagrange (E-L) approach [4]. The E-E approach considers both fluid and particles in the continuous interpenetrating phases. In such a manner granular solid phase is treated as a pseudo-fluid. The E-L approach still considers fluid phase (i.e., gas phase) as continuous but the granular one, is treated by tracking a large number of particles through the computational domain. The E-E approach is less computational expensive compared to E-L approach and is surely the most widely applied in the study of the hydrodynamic of fluidized beds. A great number of numerical studies are today available in the literature but most of these are referred to the simple cylindrical geometry of the reactor at bench scale, and mainly focused only on the evaluation of the main phenomenon involved in the fluidizing process. This work proposes the use of CFD techniques as a new approach to the real rector design by using information achievable by the scaled cold model reactor. In this study the hydrodynamic of the new concept of the ENEA's ICBFB (Internally Circulating Bubbling Fluidized Bed), as revision of the 1 MW_{th} pilot plant [5], [6], is under investigation. This configuration merges both advantages of bubbling and internally circulating fluidized beds by keeping all reaction zones in only one vessel.

2 Materials and Methods

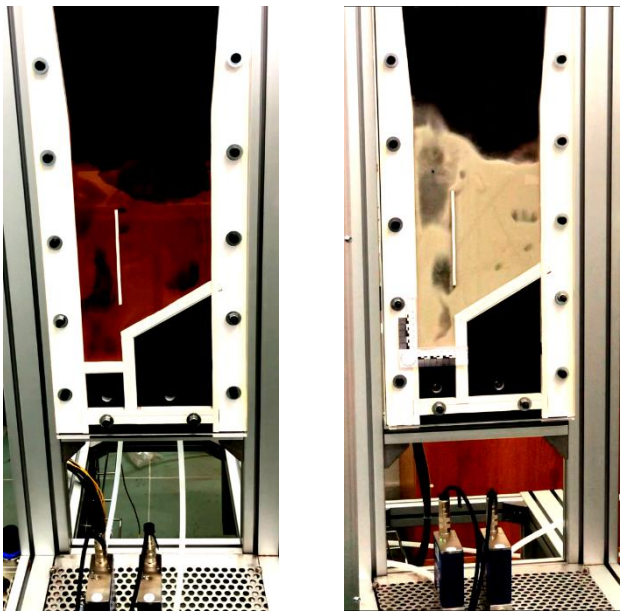
This work deals about the hydrodynamic optimization of the new configuration of the ENEA's ICBFB pilot scale reactor, Fig. 1. Within only one vessel, two bubbling fluidized beds of different cross-sectional area are delimitated by a vertical plate at the lower edge of which an interconnection opening between them (fluidized beds) is realized. Different flow rate of fluidizing agents (air) are introduced in the two chambers so that different regimes of fluidization are induced. These conditions create, inside the reactor, two areas at different bed density, which promote the circulation of the inert bed material through the opening from the denser chamber (down-flowing bed, DFB) to the lesser one (up-flowing bed, UFB). By varying the difference between the two superficial velocities inside the two reaction chambers, a different circulation rate of bed material is induced.

2.1 Cold model definition

Numerical investigations were carried out on bench scale cold model reactor in fluid dynamic similar to the real ICBFB pilot plant. For this purpose, scale transposition laws were assumed to be in agreement with some dimensionless parameters (nine dimensional groups) as proposed by Glicksman [7]. In such a manner phenomenon involved and the hydrodynamic behavior observed on bench scaled cold model can directly be transposed to real reactor. In this study, the methodology proposed by Foscolo et al. [8], where only four of these parameters must be satisfied, was considered. With respect to bench scale tests, with known fluidizing flow (i.e., air), well defined temperature, pressure (i.e., reference conditions), and geometry of the real reactor, the geometry of the cold model is fully determined, as well. In addition to such parameters, also the Reynolds and Froude numbers (even if their combination defines the Archimedes number) were verified.

2.2 Mathematical model and computational methodology

Eulerian–Eulerian multiphase approach was considered to describe gaseous flow (fluidizing flow as primary phase) and granular solid material (bed material as secondary phase) behavior. As defined earlier in the introduction, all interacting phase are considered, by E-E approach, i.e., as continuous and interpenetrating. Ansys Fluent® code was used to formulate the governing equations of the problem adopting a finite volume approach. PC-SIMPLE (phase-coupled semi-implicit method for pressure linked equation) algorithm with a first order formulation of convective terms, was selected to carry out transient simulations by setting 10^{-3} s as time step. Overall simulation time of 20 s was imposed at each run for a total number of 20.000-time steps (each one with 100 iterations). Two UDFs (user defined functions) were implemented in the original code to introduce a modified form of both the interphase exchange coefficient $\boldsymbol{K_{sg}}$, and the drag factor $\boldsymbol{f_p}$. The first one parameter, was determined by implementing the numerical procedure proposed by Syamlal et al. [9], when minimum experimental fluidization velocity was known. The second term was instead defined as proposed by Papadikis et al. [10]. Both 2D and 3D numerical grids with different refinement grade were considered during calculations. Initial solid volume fraction as well as initial field velocity, were imposed both for freeboard and granular bed region. Constant velocity profiles were imposed at inlet sections (velocity inlet), while atmospheric pressure was imposed at outlet section (pressure outlet) of the numerical grid. Gaseous phase velocity at walls was imposed to be equal to zero (no-slip boundary conditions) while no tangential stress conditions (free-slip boundary conditions) was imposed for the solid phase.



(**a**) (**b**)

Fig. **2**: ENEA's ICBFB pseudo-2D cold model test ring implementing LaVision PIV system for velocity field evaluations. Images of test conducted with: (a) 0.14 mm copper powder particles; (b) 0.55 mm glass beads.

3 Results

3.1 Numerical Bechmark

Validation of mathematical model is of crucial importance in order to achieve good accuracy in results with respect to physical observations. The importance is even greater if obtained results are to be used as guideline in the designing of equipment. The effective hydrodynamic conditions inside a fluidized bed reactor, is very difficult to capture. In this study, mathematical model was experimentally validated by using data available by Deza et al. [11] which conducted tests on bench scale cold model by the implementation of X-ray methodology. Experiments were carried out by using an acrylic cylindric shaped reactor whose diameter is equal to 9.5 cm. Glass beads were used as bed material (with an initial bed height of 10 cm) and air as fluidizing agent by imposing a superficial gas velocity of $U_g = 1.3 \cdot U_{mf}$. Results obtained show as the maximum percentage error is less than 3.5%. The independence of results toward grid refinement was evaluated following ASME (American Society of Mechanical Engineers) procedure for estimation of the uncertainty due to discretization in CFD applications [12]. GCI procedure (grid convergence method), based on the Richardson extrapolation method, was used to evaluate the selected variables (for the purpose: pressure drop through the bed and mean void fraction at z = 4 cm and z = 8 cm) do not exhibit a monotone dependence toward grid refinement. It must be highlighted that bed material used by Deza et al. in their study was different from that identified for the cold model considered here, where copper powder was selected. This methodological choice is justifiable in view of the lack of available experimental data for copper powder. Furthermore, it must be stated that both glass and copper particles are classifiable as Geldart Group B, so that the phenomenological nature of the observations should be the same. This means that mathematical model, parametrized with respect to the different particles used and describing the two processes, should have the same response. Preliminary experiments were conducted on the ENEA's ICBFB pseudo-2D cold model; PIV (Particle Image Velocimetry) methodology that was adopted to investigate the velocity field structure, seems to confirm these assumptions. A schematic representation of the ICBFB pseudo-2D cold model is shown in Fig. 2.

A reactor, 10 mm depth in perpendicular direction to the plane, was equipped with a front panel realized in polycarbonate in order to investigate solid motion inside the two chambers. A black back panel was instead used in order to create the right light contrast for the acquisition. The acquisition area was lighted

by using two light led sources while a high speed camera, provided by La Vision, was used for frame acquisitions. Cross-correlation method was used to correlate frames at different time in order to solve the solid velocity field. A comparison of two different images of tests conducted, under the same fluidizing conditions, with 0.14 mm copper powder and 0.55 mm glass beads, are shown in Fig. 2. The comparison should be regarded as qualitative. Of course, the two experiments were carried out in the same fluidizing conditions but with different density number, in two cases. Nevertheless, the behavior of the two systems appears to be the same. Dimension of the bubbles is in the same order of magnitude. Moreover, bed expansion and up-flowing chamber is visibly less dense than down-flowing one. Furthermore, circulation of bed material through the opening between the two chambers is well visible in the two cases.

3.2 ENEA's ICBFB Cold Model Analysis

The validated numerical code was then applied to study the behavior of the cold model describing the ENEA's ICBFB. The geometry of the ICBFB cold model reactor and the fluidizing conditions were fully defined by applying scaling laws to preserve fluid dynamic similarities between the real pilot reactor at full scale (1 MWth), operating at high temperature, and that one at bench scale operating at ambient conditions. Behavior of the inert bed material was investigated under the design fluidizing conditions and critical conditions (e.g., elutriation and agglomeration phenomenon) were identified. Recirculation rate of bed material between the two reaction chambers was also evaluated and compared with theoretical one calculated by using semi-empirical correlations by different authors [13]. The numerical results are related to the best grid refinement at solution convergence. This latter was evaluated by following the ASME procedure [12] based on the Richardson extrapolation method. By analyzing numerical results, it is clearly evident as the two regions inside the reactor (UFB and DFB), are characterized by a different behavior. This is well noticeable by evaluating for each of these the bed expansions ratio, the bubbles fraction and the mean void fraction terms. DFB chamber is denser than UFB one, that also exhibits a greater bed expansion respect to the initial condition. Again, the flux vector evaluated on the opening plane, exhibits as its main component that one in perpendicular direction to the plane along the positive x axes. All these conditions clearly are explained by a motion of the bed material between the two regions inside the reactor from the denser to less dense chamber across the opening at the bottom side of the vertical plate.

All the above considerations are well visible in Fig. 3, where a comparison between results obtained with finest mesh is proposed for different simulation time. Images show, on the background, the solid phase volume fraction and, in overlapping, the solid phase velocity field. Each of these, clearly shows the bed motion between the two reactor regions. The behavior of solid particles modelling char was also investigated by injecting solid particles of different size inside the well fluidized bed and by keeping for each one all information (e.g., position, velocity, and momentum) at different simulation time. 3D numerical simulations, Fig. 4, were carried out by setting up a DDPM model. Five equivalent particles were injected at simulation time t = 10 s when inert bed material was well fluidized. Simulations were terminated at simulation time t = 20 s and data about particles (e.g., position, velocity, and momentum) were collected every 0.01 s. Results showed that particles whose diameter is lesser than 0.22 mm (equivalent to 0.80 mm char particles), segregate on bed surface in the upper left region of the bed, near the reactor walls, where stagnant fluid dynamic conditions are encountered. Instead, particles whose diameter is in the range 0.22 mm to 0.27 mm (equivalent to 1.00 mm char particles) well follow bed motion, circulating between the two reaction chambers. The calculated recirculation time of average particles is in the order of 5 s for 0.22 mm particle, and 6–7 s for 0.27 mm particle, well comparable with the average bed recirculation time estimated to be in the order of 6–7 s at bench scale and, respectively, of 15–20 s at full scale. All these information suggest that the dimension of the mother fuel particles, to feed to the real pilot reactor, should not exceed the evaluated value of 5–10 mm.

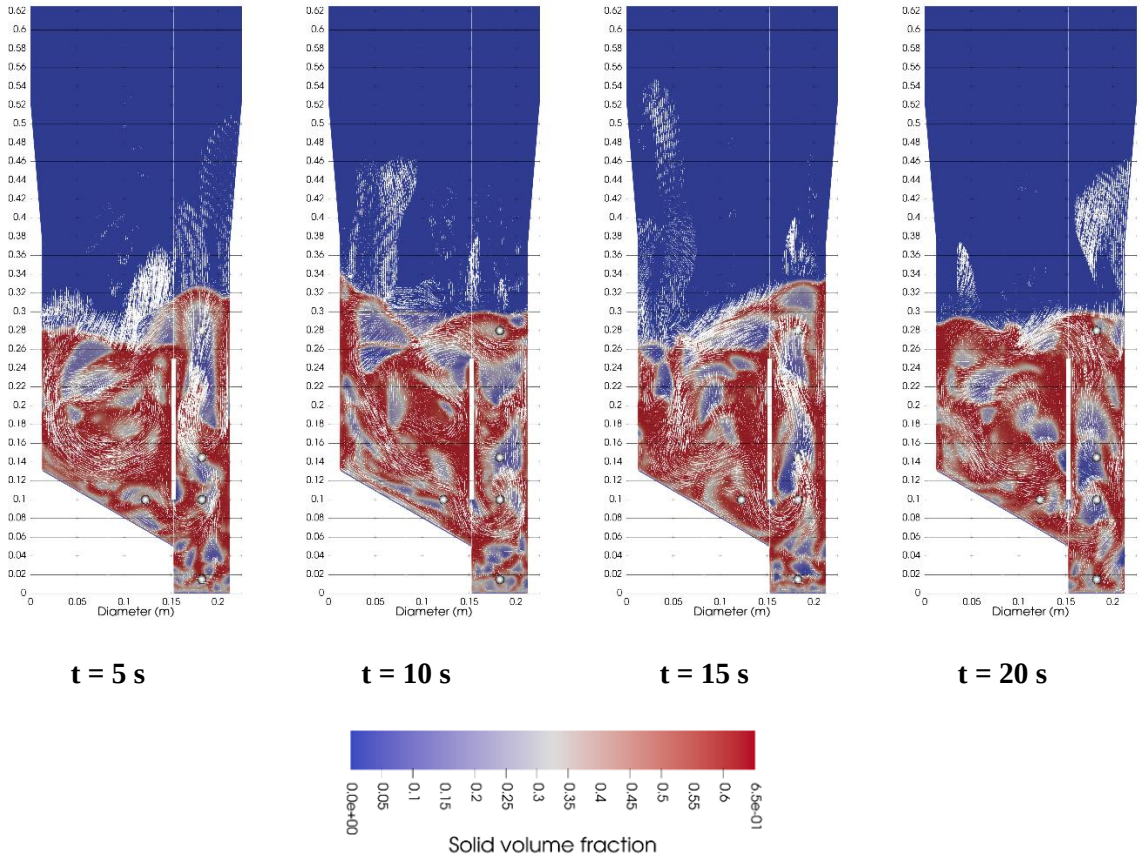
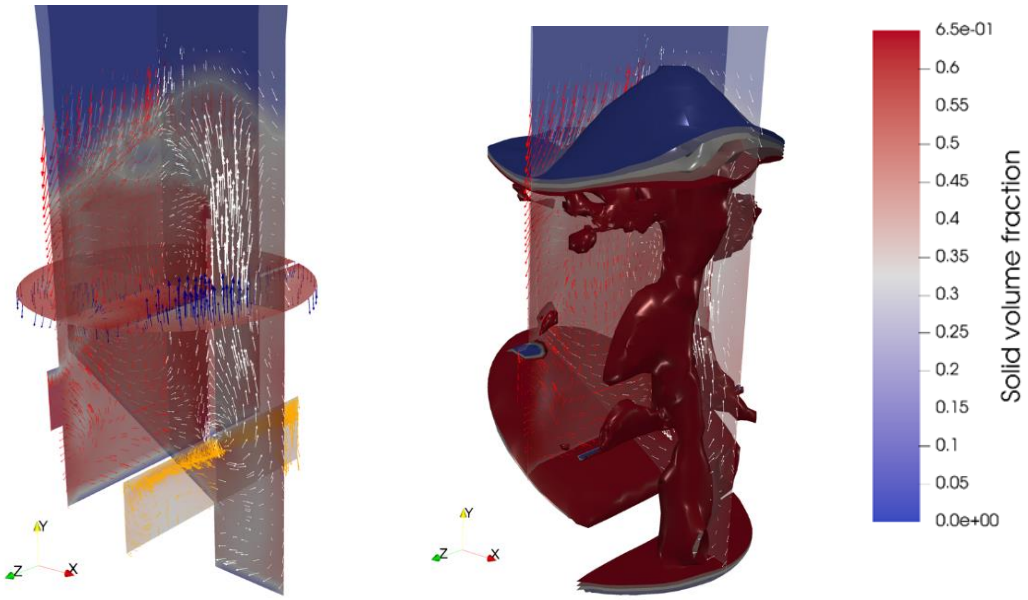


Fig. **3**: Solid phase volume fraction and field velocity comparison between CFD solutions at various simulation time.

Acknowledgments

The computing resources and the related technical support used for this work have been provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [14]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development and by Italian and European research programs, see http://www.cresco.enea.it/english for information.
Research activities were funded by H2020 Spring G2E project supported by the Italian ministry of economic.

3D CFD Solution at simulation time t = 10 s



(**a**) (**b**)

Fig.4: 3D-CFD visual representation at simulation time t = 10 s: (**a**) solid phase volume fraction and solid velocity field visualization on main section planes (colors of arrows are only delimiting the cutting plane where vector field is displayed); (**b**) granular solid phase iso-surfaces (red $\varepsilon_s = 0.35$, blue $\varepsilon_s = 0.15$) delimiting bubbles formation inside the two chambers.

Patents

E. Fanelli, F. Nanna, D. Barisano, G. Cornacchia, G. Braccio. "Reattore per la gassificazione di biomasse" IT Patent n° 202019000004530, issue December 13, 2019.

References

[1] IEA, "Bioenergy Power Generation," Paris, 2020. [Online]. Available: https://www.iea.org/reports/bioenergy-power-generation.

[2] L. Devi, K. J. Ptasinski, and F. J. J. G. Janssen, "A review of the primary measures for tar elimination in biomass gasification processes," Biomass and Bioenergy, vol. 24, no. 2, pp. 125–140, 2003, doi: https://doi.org/10.1016/S0961-9534(02)00102-2.

[3] R. G. dos Santos and A. C. Alencar, "Biomass-derived syngas production via gasification process and its catalytic conversion into fuels by Fischer Tropsch synthesis: A review," Int. J. Hydrogen Energy, 2019, doi: https://doi.org/10.1016/j.ijhydene.2019.07.133.

[4] W. K. H. Ariyaratne, E. V. P. J. Manjula, C. Ratnayake, and M. C. Melaaen, "CFD Approaches for Modeling Gas-Solids Multiphase Flows - A Review," Proc. 9th EUROSIM Congr. Model. Simulation, EUROSIM 2016, 57th SIMS Conf. Simul. Model. SIMS 2016, vol. 142, pp. 680–686, 2018, doi: 10.3384/ecp17142680.

[5] D. Barisano et al., "Steam/oxygen biomass gasification at pilot scale in an internally circulating bubbling fluidized bed reactor," Fuel Process. Technol., vol. 141, pp. 74–81, 2016, doi: https://doi.org/10.1016/j.fuproc.2015.06.008.

[6] D. Barisano et al., "Production of Gaseous Carriers Via Biomass Gasification for Energy Purposes," Energy Procedia, vol. 45, pp. 2–11, 2014, doi: https://doi.org/10.1016/j.egypro.2014.01.002.

[7] L. R. Glicksman, M. Hyre, and K. Woloshun, "Simplified scaling relationships for fluidized beds," Powder Technol., vol. 77, no. 2, pp. 177–199, 1993, doi: 10.1016/0032-5910(93)80055-F.

[8] R. D. I. Felice, S. Rapagna, and P. U. Foscolo, “Cold modelling studies of fluidized bed reactors,” pp. 2233–2238, 1992.
[9] M. Syamlal, N. Energy, T. O. Brien, N. Energy, and U. States, “The Derivation of a Drag Coefficient Formula from Velocity-Voidage Correlations,” no. January, 1994.
[10] K. Papadikis, A. V. Bridgwater, and S. Gu, “CFD modelling of the fast pyrolysis of biomass in fluidised bed reactors, Part A: Eulerian computation of momentum transport in bubbling fluidised beds,” Chem. Eng. Sci., vol. 63, no. 16, pp. 4218–4227, 2008, doi: 10.1016/j.ces.2008.05.045.
[11] M. Deza, N. P. Franka, T. J. Heindel, and F. Battaglia, “CFD modeling and x-ray imaging of biomass in a fluidized bed,” J. Fluids Eng. Trans. ASME, vol. 131, no. 11, pp. 1113031–11130311, 2009, doi: 10.1115/1.4000257.
[12] F. E. Division et al., “Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Applications,” vol. 130, no. July, pp. 2008–2011, 2008, doi: 10.1115/1.2960953.
[13] M. Kuramoto, D. Kunii, and T. Furusawa, “Flow of dense fluidized particles through an opening in a circulation system,” Powder Technol., vol. 47, no. 2, pp. 141–149, 1986, doi: 10.1016/0032-5910(86)80110-3.
[14] G. et al. Ponti, “The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure,” in Proceedings of the 2014 International Conference on High Performance Computing and Simulation, 2014, pp. 1030–1033.

COMPUTATIONAL METHODS APPLIED TO THE DETECTION OF SARS-COV-2 INHIBITORS TARGETING THE SPIKE GLYCOPROTEIN

Alice Romeo, Federico Iacovelli and Mattia Falconi*

Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy

ABSTRACT. Given the pivotal role of surface glycoproteins in mediating recognition and fusion of enveloped viruses to host cells, targeting these proteins is a promising strategy for blocking the viral life cycle already at the early entry phase. Considering previous data reported for the respiratory syncytial virus, we identified a highly conserved internal cavity of the SARS-CoV-2 Spike as a new possible therapeutic target due to its key structural role in the viral membrane fusion process. To evaluate this hypothesis, a virtual screening was carried out on a set of FDA drugs to identify possible fusion inhibitors strongly binding to this region, and molecular dynamics simulations were performed for the two top-ranking complexes to characterize compounds interactions with the cavity using the MM/GBSA method. In parallel, considering the well-known antiviral properties of Lactoferrin, a glycoprotein present in all human secretions, molecular docking and molecular dynamics simulations have been carried out to evaluate its potential to interfere with host cells recognition acting as a competitive inhibitor of Spike's binding to ACE2, and the promising results obtained were also supported by *in vitro* evidences.

1 Introduction

Despite the recent release of highly effective vaccines against the novel coronavirus SARS-CoV-2, the causative agent of the COVID-19 pandemic, more than one year after its outbreak there is still a need for potent, safe, and broad-spectrum antiviral drugs to treat infected patients and, possibly, to control any future outbreak of similar viruses [1]. The SARS-CoV-2 Spike glycoprotein represents one of the most promising therapeutic targets and is composed by a large extracellular domain, which is divided into two functional subunits and mediates the key steps of the virus entry process: the S1 subunit (residues 1-685) is involved in the recognition of the ACE2 cell receptor, while the S2 (residues 686-1273) mediates viral and cell membranes fusion and the subsequent release of viral genome inside the cell [2]. Fusion is mediated by trimeric α-helical regions called heptad repeats 1 (HR1) and 2 (HR2) that, upon receptor binding and protease cleavage, undergo huge conformational changes determining the transition of the Spike from a metastable prefusion conformation to a postfusion conformation, providing the energy requirements to drive viral and cell membranes fusion [3]. The S2 domain and its interaction modes are highly conserved among human coronaviruses and enveloped viruses in general [2]. Recently, structural and experimental data reported for the human respiratory syncytial virus (RSV) fusion (F) glycoprotein showed that specific small molecules can bind within a central cavity of this protein and act as fusion inhibitors, preventing the protein transition to the postfusion state [4]. Given the structural similarity of the RSV F protein with the Spike, we hypothesized that the internal cavity of the Spike in prefusion conformation could represent a possible new broad-spectrum therapeutic target that would allow to interfere with coronavirus infection at an early stage. To evaluate this hypothesis and to suggest possible fusion inhibitor compounds, we performed a virtual screening (VS) within this internal pocket using a library of thousands of FDA drugs [5].

In parallel, we also characterized the antiviral potential of the glycoprotein Lactoferrin (Lf) against SARS-CoV-2. Lf is a member of the transferrin family, is present in all human secretions and is part of the innate immunity [6]. Several studies showed that this protein can interfere with receptor recognition and prevent viral entry into cells either by obscuring cellular receptors or by directly binding to surface viral particles, like the RSV F, the HIV gp120 and the HCV E1 and E2 fusion glycoproteins [6]. To

*Corresponding author. E-mail: falconi@uniroma2.it

evaluate if this protein could also interfere with Spike-ACE2 recognition, we performed protein-protein molecular docking and molecular dynamics (MD) simulations to determine the presence of binding sites for the bovine (bLf) or human (hLf) form of Lf, sharing about 70% of sequence identity, on the Spike surface [7].

2 Methods

2.1 Virtual screening of FDA drugs

A cryo-EM structure of the SARS-CoV-2 Spike glycoprotein in prefusion conformation (PDB ID: 6VSB) [8] was used as receptor for the VS procedure (Fig. 1, left) after modelling several non-terminal missing loops using the SWISS-MODEL webserver [9]. A drug library of 8755 FDA-approved, experimental, or investigational drugs, was obtained from the DrugBank database [10]. Molecular docking simulations have been performed using an in-house parallelized version of the Autodock Vina program [11] implemented on the ENEA HPC cluster CRESCO6. Parallelization was obtained through in-house written scripts employing the *mpi4py* Python3 library, which allowed to run 4 different Vina processes at the same time on each node of the CRESCO6 cluster. The use of 12 nodes of the cluster (576 CPUs) allowed to perform about 1200 molecular docking simulations per day, extremely reducing the computational time. The simulation box was placed over one of the HR1 internal lobes (residues 897–920) of the Spike, selecting 15 receptor sidechains inside the box as flexible. Binding energies of the 10 top-ranking compounds were calculated as an average of the best poses obtained in three repeated molecular docking simulations. The top-10 ranking was then reweighted using an additional scoring procedure, based on: binding energies, known side effects, physiological effects, and antiviral properties of the compounds, which resulted in a final score (named S-final) for each drug [5]. Due to the reciprocal size of cavity and ligand and hypothesizing the simultaneous binding of three molecules within the cavity, sequential molecular docking simulations have been performed for the two top-ranking drugs placing the simulation boxes over the other two internal lobes of the cavity and, again, selecting 15 receptor side chains as flexible. Each docking simulation was repeated three times and binding energies were calculated as an average of the interaction energies obtained for the best poses. Contacts between the ligands and the Spike were analysed using the LigPlot+ program [12]. Pictures were generated using the PyMOL 2.1.0 [13] and the UCSF Chimera programs [14].

2.2 Molecular dynamics simulations and trajectory analyses

Complexes obtained after sequential molecular docking simulations of phthalocyanine and hypericin were simulated using classical MD. Topologies and coordinates files of the input structures were generated using the tLeap module of the AmberTools 19 package [15], parametrizing the Spike with the ff19SB force field [16] and the ligands with the antechamber module of AmberTools 19 [15] and the general Amber force field [17]. Each complex was solvated using a box filled with TIP3P water molecules and 0.15 mol/L of NaCl, setting a minimum distance of 12.0 Å from the box sides. Four minimization cycles were performed for both systems, each composed by 500 steps of steepest descent followed by 1500 steps of conjugate gradient method, slowly decreasing the constraints applied on protein and ligand atoms, starting from 20.0 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2}. Systems were gradually heated from 0 to 300 K in a NVT ensemble over a period of 2.0 ns using the Langevin thermostat [18], imposing a starting restraint of 0.5 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2} on protein and ligand atoms, which was decreased every 500 ps. Simulations were then carried out in an isobaric-isothermal (NPT) ensemble for 2.0 ns, imposing a pressure of 1.0 atm using the Langevin barostat [19] and fixing the temperature at 300 K. A production run of 30 ns was finally performed for both systems with a timestep of 2.0 fs, using the pmemd.cuda module of the AMBER16 software [18]. Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm [19], the PME method [20] was used for calculating long-range interactions and a cut-off of 9.0 Å was set for short-range interactions. Hydrogen bonds (Hbonds) have been evaluated using the Hbonds plugin of VMD 1.9.3 [21]. Interaction analysis has been performed using the molecular mechanics energies combined with generalized Born and surface area continuum solvation (MM/GBSA) method [22], coupled to per-residue decomposition, using the MMPBSA.py.MPI program of the AMBER16 software [18] on 3 nodes of the CRESCO6 HPC cluster.

2.3 Protein-protein molecular docking and molecular dynamics simulations of Spike-Lactoferrin complexes

Protein-protein molecular docking simulations were carried out between the Spike protein and the lactoferrin structures using the FRODOCK webserver [23]. Structure of the Spike in prefusion conformation was extracted from a previously performed trajectory [5], while that of the bLf and hLf were obtained from the PDB Database (PDB IDs: 1BLF [24] and 1B0L [25]). Topology and coordinate files for MD simulations were generated using the tLeap module of the AmberTools 19 package [15] and the ff19SB force field [16], inserting the proteins into a box of TIP3P water molecules and 0.15 mol/L of NaCl ions and setting a minimum distance of 12.0 Å from the box sides. The same simulation protocol described in the previous paragraph was used and production runs of 30 ns were performed for both systems on the CRESCO6 HPC cluster, using the NAMD 2.13 MD package [26]. Hbonds were analysed using the hbonds module of the GROMACS 2019 analysis tools [27], while hydrophobic contacts were identified using the mdtraj Python library [28]. MM/GBSA analysis [22] was performed as previously described.

3 Results

3.1 Virtual screening of FDA drugs and evaluation of top-ranking compounds

The VS procedure allowed to rapidly evaluate the affinity of 8755 drugs against one of the internal lobes (residues 897-920) of the Spike cavity (Fig.1, left). The ten top-ranking compounds, obtained after performing a literature-based ranking reweighting [5], are reported in Table 1.

Considering their high interaction energies with the Spike, subsequent analyses were focused only on phthalocyanine (PHT) and hypericin (HYP). HYP and different PHT derivatives have been previously evaluated as antiviral and, in particular, anti-HIV compounds for their capability to interfere with viral gp120 glycoprotein binding and fusion to the host cell, and both drugs also possess other targets on different viruses, including the avian coronavirus infection bronchitis virus (IBV) [5]. Recently, the use of a PHT-containing mouthwash has also helped reduce symptoms and hospitalization time for COVID-19 patients [29]. Considering the trimeric structure of the Spike, we performed sequential molecular docking simulations, proposing an inhibitory mechanism by which at least three molecules could arrange inside the inner cavity, binding the three HR1 internal lobes (Fig. 1A, B).

Table 1: Final reweighted ranking of the VS.

Compound	Binding energy (kcal/mol)	Final score S-final
phthalocyanine	-16.3	62.6
hypericin	-15.1	55.3
TMC-647055	-12.5	49.5
quarfloxin	-12.6	35.2
tepotinib	-12.0	24.1
laniquidar	-12.8	23.0
tadalafil	-12.4	21.4
ergotamine	-13.2	20.1
JNJ-10311795	-13.1	19.1
TZ2PA6	-12.7	12.7

Results showed that PHTs and HYPs arrange in a shifted and tilted orientation within the pocket. PHTs mainly establish hydrophobic contacts with surrounding residues and within themselves, while HYPs establish both hydrophobic contacts and Hbonds within themselves and with other residues.

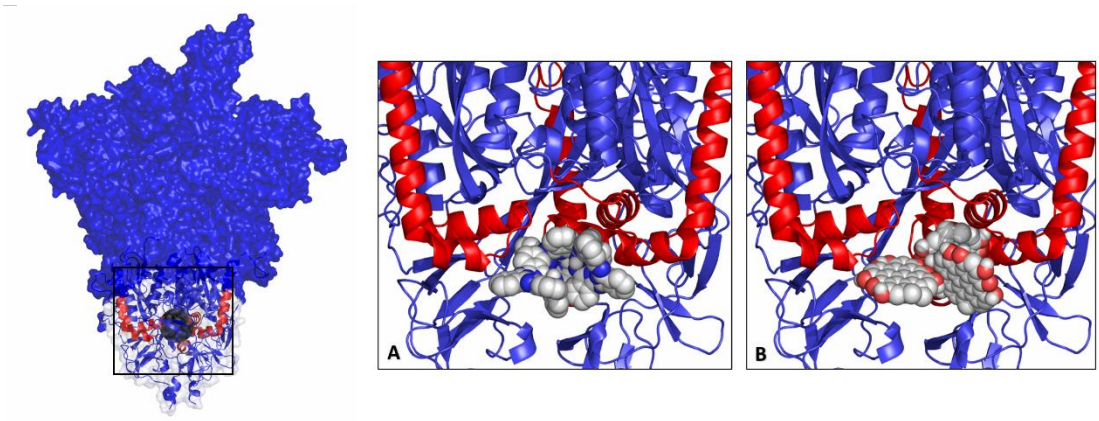


Fig.1: On the left, the Spike glycoprotein modelled structure. The black sphere placed within the structure highlights the internal cavity, selected for the VS, shown in cartoon representation surrounded by a transparent surface. The region surrounded by the black square is magnified in the images on the right, showing the trimeric Spike-PHT (**A**) and Spike-HYP (**B**) complexes after three sequential molecular docking simulations. HR1 regions of the Spike are highlighted in red.

To further assess the stability of the obtained binding poses, we also performed 30 ns MD simulations of the trimeric Spike-PHT and Spike-HYP complexes. Free energies of binding, calculated using the MM/GBSA method [22], confirmed the high interaction energies observed through molecular docking simulations and highlighted the presence of a predominant hydrophobic contribution within the pocket during both simulations, indicated by the presence of highly favourable Van der Waals energies for all six compounds (Table 2).
Furthermore, HYP interactions are also characterized by a negative electrostatic contribution (Table 2) and Hbonds analysis showed that the three HYPs can establish from of 1 to 4 Hbonds during the simulation time, while no valuable Hbond was observed for PHTs.

Table 2: MM/GBSA results for the three PHT and HYP molecules.

Compound	VdW (kcal/mol)	Electrostatic (kcal/mol)	Nonpolar solvation (kcal/mol)	Polar solvation (kcal/mol)	$\Delta G_{binding}$ (kcal/mol)
PHT #1	-84.8 ± 4.7	-5.6 ± 3.0	-6.8 ± 0.2	30.6 ± 2.2	-66.6 ± 4.7
PHT #2	-76.5 ± 3.0	-0.7 ± 2.3	-6.7 ± 0.2	24.7 ± 2.0	-59.1 ± 3.0
PHT #3	-54.3 ± 2.7	3.0 ± 2.6	-5.0 ± 0.3	20.7 ± 2.9	-35.6 ± 2.6
HYP #1	-55.3 ± 3.6	-18.4 ± 6.4	-4.7 ± 0.3	34.7 ± 5.5	-43.6 ± 4.4
HYP #2	-53.3 ± 3.0	-19.4 ± 5.2	-5.2 ± 0.2	33.8 ± 4.0	-44.1 ± 3.5
HYP #3	-37.4 ± 2.4	-14.4 ± 5.5	-4.3 ± 0.2	28.8 ± 4.7	-27.3 ± 3.3

Per-residue decomposition analyses [22] showed that each drug can establish from 9 to 15 contacts, with interaction energies ranging from -0.5 to -5.1 kcal/mol. During the simulation time, the three PHTs and HYPs created a clustered arrangement within the pocket, contacting residues from the HR1 internal lobes and from upward and downward regions of the Spike. This is expected to generate a hard impairment to the motion of the internal regions and supports our hypothesis that the strong network of interactions established, and the steric hindrance generated should be sufficient for the compounds to block the Spike transition to the postfusion state.

3.2 Molecular docking and molecular dynamics simulations of Spike-Lactoferrin complexes

Protein-protein molecular docking simulations of bLf and hLf targeting the Spike glycoprotein, indicated that both proteins mainly interact with the Spike receptor binding domain (RBD) in “up”

conformation (Fig. 2 B-C). The stability of the best obtained complexes was further evaluated through 30 ns of classical MD simulations. Simulations showed that close and stable contacts are established between the proteins' interfaces, and MM/GBSA [22] interaction analyses also showed that both proteins interact with the Spike RBD domain with high free energies of binding. In particular, bLf shows an interaction energy of -28.0 ± 9.0 kcal/mol, while the hLf reaches an almost two-fold higher energy of -48.3 ± 17.0 kcal/mol. This is explained by the higher number of interactions established by the human protein in comparison with the bovine form (45 and 28, respectively). A detailed analysis of the interaction types showed that the bLf can arrange 20 hydrophobic contacts, 3 salt bridges and 5 Hbonds with the Spike RBD, while the hLf sets up 23 hydrophobic contacts, 12 salt bridges and 10 Hbonds. Furthermore, only two Spike residues (Gly502 and Tyr505) are shared between the ACE2 [30] and lactoferrins binding interfaces, despite the three proteins bind to close sites on the RBD surface (Fig. 2). The obtained results allowed us to hypothesize that one of the many beneficial effects of Lf could involve the competitive binding of the Spike RBD, preventing Spike attachment to the human ACE2 receptor and, in this way, blocking host cell recognition and virus entry. Indeed, results obtained *in vitro* support our computational results showing that pre-incubation of bLf with SARS-CoV-2 reduced the infection of two different cell lines [7].

4 Conclusions

The VS carried out using the ENEA HPC cluster CRESCO6, and the following MD simulations, allowed us to characterize a novel SARS-CoV-2 therapeutic target, setting the path for future research that could be useful also against other threatening enveloped viruses sharing a similar entry mechanism. Furthermore, MD simulations of two Spike-Lf complexes, performed on the CRESCO6 cluster, allowed us to suggest a further and unknown protective role exerted by Lf against SARS-CoV-2, that binding the Spike RBD domain *de facto* prevents virus attachment and entry into cells.

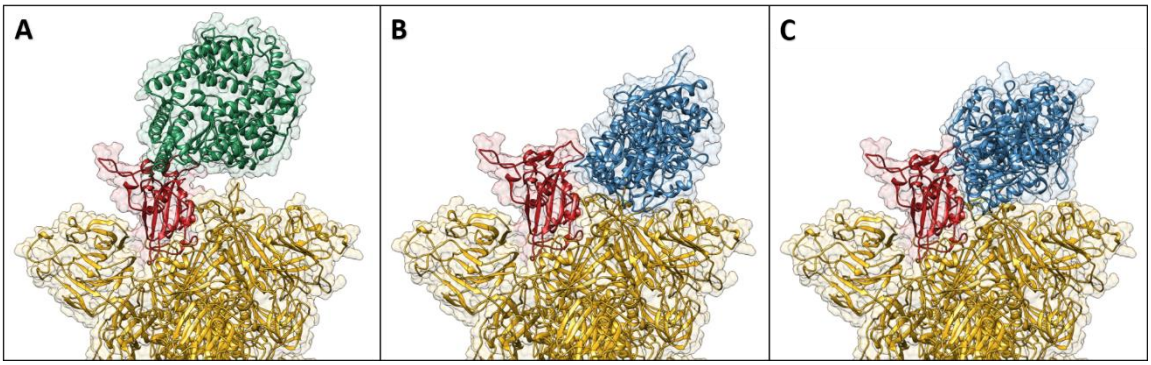


Fig. 2: Spike-ACE2 complex (PDB ID: 6LZG) (**A**), compared with the best FRODOCK binding poses obtained between the Spike and the bLf (**B**) or hLf (**C**), represented as ribbons surrounded by a transparent surface. The Spike is represented in yellow and the RBD domain is highlighted in red. ACE2 is shown in green, while both lactoferrins are shown in blue.

References

[1] M. Mei and X. Tan. Current Strategies of Antiviral Drug Discovery for COVID-19. *Front. Mol. Biosci.* **8**, p. 671263, (2021).

[2] T. Tang, M. Bidon, J.A. Jaimes, et al. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res.* **178**, p. 104792, (2020).

[3] A.C. Walls, M.A. Tortorici, J. Snijder, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **114**, pp. 11157–11162, (2017).

[4] M.B. Battles and J.S. McLellan. Respiratory syncytial virus entry and how to block it. *Nat. Rev. Microbiol.* **17**, pp. 233–245, (2019).

[5] A. Romeo, F. Iacovelli and M. Falconi. Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors. *Virus Res*. **286**, p. 198068, (2020).
[6] E. Campione, T. Cosio, L. Rosa, et al. Lactoferrin as protective natural barrier of respiratory and intestinal mucosa against coronavirus infection and inflammation. *Int. J. Mol. Sci*. **21**, pp. 1–14, (2020).
[7] E. Campione, C. Lanna, T. Cosio, et al. Lactoferrin against SARS-CoV-2: in vitro and in silico evidences. *Front. Pharmacol.* **12**, p.1524,(2021).
[8] D. Wrapp, N. Wang, K.S. Corbett, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, pp. 1260-1263, (2020).
[9] A. Waterhouse, M. Bertoni, S. Bienert, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, pp. W296–W303, (2018).
[10] D.S. Wishart, Y.D. Feunang, A.C. Guo, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, pp. D1074–D1082, (2018).
[11] O. Trott and A.J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, pp. 455–461, (2010).
[12] R.A. Laskowski and M.B. Swindells. LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, pp. 2778–2786, (2011).
[13] The PyMOL Molecular Graphics System, Version 2.1.0 Schrödinger, LLC.
[14] E.F. Pettersen, T.D. Goddard, C.C. Huang, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, pp. 1605–1612, (2004).
[15] R. Salomon-Ferrer, D.A. Case and R.C. Walker. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **3**, pp. 198–210, (2013).
[16] C. Tian, K. Kasavajhala, K.A.A. Belfon, et al. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **16**, pp. 528–552, (2020).
[17] J. Wang, R.M. Wolf, J.W. Caldwell, et al. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, pp. 1157–1174, (2004).
[18] N. Goga, A.J. Rzepiela, A.H. de Vries, et al. Efficient algorithms for Langevin and DPD dynamics. *J. Chem. Theory Comput.* **8**, pp. 3637–3649, (2012).
[19] K.M. Aoki, M. Yoneya and H. Yokoyama. Constant pressure Md simulation method. *Mol. Cryst. Liq. Cryst.* **413**, pp. 109–116, (2004).
[18] D.A. Case, R.M. Betz, D.S. Cerutti, et al. AMBER 2016. University of California, San Francisco (2016).
[19] J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, pp. 327–341, (1977).
[20] T. Darden, D. York and L. Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, pp. 10089–10092, (1993).
[21] W. Humphrey, A. Dalke and K. Schulten. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, pp. 33–38, (1996).
[22] S. Genheden and U. Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, pp. 449–461, (2015).
[23] E. Ramírez-Aportela, J.R. López-Blanco and P. Chacón. FRODOCK 2.0: fast protein–protein docking server. *Bioinformatics* **32**, pp. 2386–2388, (2016).
[24] S.A. Moore, B.F. Anderson, C.R. Groom, et al. Three-dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. *J. Mol. Biol.* **274**, pp. 222–236, (1997).
[25] X.L. Sun, H.M. Baker, S.C. Shewry, et al. Structure of recombinant human lactoferrin expressed in Aspergillus awamori. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, pp. 403–407, (1999).
[26] J.C. Phillips, R. Braun, W. Wang, et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, pp. 1781–1802, (2005).
[27] M.J. Abraham, T. Murtola, R. Schulz, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, pp. 19–25, (2015).
[28] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, pp. 1528–1532, (2015).

[29] P.S.S. Santos, B.F. Orcina, R.R.G. Machado, et al. Beneficial effects of a mouthwash containing an antiviral phthalocyanine derivative on the length of hospital stay for COVID-19: Randomised trial. *Research Square*, (2021)
[30] Q. Wang, Y. Zhang, L. Wu, et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, pp. 894-904, (2020).

HPC-DRIVEN HIT-TO-LEAD PROCESS FOR SARS-COV-2 MAIN PROTEASE INHIBITION

Piero Procacci[1*], Marina Macchiagodena[1], Maurice Karrenbrock[2], Marco Pagliai[1], Guido Guarnieri[3] and Francesco Iannone[3]

[1]*University of Florence, Chemistry Department, Via Lastruccia 3, Sesto F.no 50019 (Italy)*

[2]*University of Geneva, Pharmaceutical Sciences, 30, quai Ernest-Ansermet CH-1211 Genève (Switzerland)*

[3]*ENEA, Energy Technologies & Renewable Sources Department - Information Communication Technologies Division, Lungotevere Thaon di Revel, 76, 00196 Rome Italy*

ABSTRACT. In this contribution, we present a molecular dynamics-based technique for the calculation of the absolute binding free energies (ABFE) in drug-receptor systems. The technique, called virtual double system single box (vDSSB), is a versatile nonequilibrium variant of the so-called alchemical approach for ABFE calculation, specifically tailored for homogenous and heterogenous high-performing computing platforms. The technique has been applied to the calculation of potential non-covalent inhibitors of the main protease of the SARS-CoV-2 virus. We report here the results obtained in the summer 2020 on the CRESCO6 facilities using the program ORAC [M. Macchiagodena et al., J. Chem. Theory Comput. 16, 7160 (2020), P. Procacci et al., Chem. Comm., 56, 8854 (2020)] on a series of ligands selected as docking hits from a previous virtual screening study. The algorithm has been further adapted to the GROMACS code and tested successfully on the Marconi100 heterogeneous architecture at CINECA.

1 Introduction

Ligand-receptor binding free energy (BFE) prediction is one of the main research topics in computational chemistry today due to the potential impact on drug discovery and public health. Powered by the exponential growth in computer speed of the last two decades, modern supercomputing architectures are now affording high-throughput screening (HTS) with an efficiency largely outperforming experimental HTS, at a much lower cost and accessing a much larger and unrestricted chemical space domain. According to the Scopus database, more than 6% of all peer-reviewed Covid-19-related scientific output in 2020 involved computational approaches, mostly based on molecular docking. A contemporary high-end HPC platform is capable of screening via docking many millions of compounds per day on a given biological target [1]. Docking techniques have benefited in the last years of knowledge-based and machine learning (ML) methods. Scoring functions (SF) have gradually evolved towards the use of empirically weighted simplified physical descriptors heavily trained on ever-extending databases of ligand-receptor binding free energy. For example, in the widely popular Vina program [2], a recent evolution of the Autodock code, electrostatic interactions based on atomic charges have been replaced by piecewise linear functions for H-bonded atoms, with weighted steric atom-atom functions accounting for hydrophobic effects. The Vina code was found to significantly improve the binding mode prediction accuracy on the well-established virtual screening benchmark called the Directory of Useful Decoys [3] with respect to its precursor Autodock relying on a complex, more physically-grounded SF. The performance of docking SFs in terms of BFE predictions is generally measured using ranking-based binary metrics such as the area under the Receiver Operating Characteristic (ROC) curve (AUC) or the Enrichment Factor (EF). These methods assume that ligands can be classified in two groups, namely active and inactive compounds according to some affinity threshold. Performances are measured with respect to a random decision (e.g., the random-hit rate in case of the EF or the random guess in case of the ROC-AUC).

[*]Corresponding author. E-mail: piero.procacci@unifi.it

While powerful in screening large datasets of compounds, docking is unable to provide secure affinity predictions. This is especially true when the target/ligand pair is deflecting significantly from the chemical and structural traits of the ML training sets [4]. Vina, exhibiting ROC-AUC from 0.70-0.90 in the DUD-E derived test sets [5]. When tested on a simple host-guest database, it yields a prediction efficiency only slightly higher that based on the flipping of a coin with ROC-AUC of only 0.55 [6].
The main problem in docking technology lies in the high number of false positives produced in screening campaigns, a limit that is shared with experimental HTS. False negatives are unavoidable in HTS processes, whether virtual or experimental. False positives, on the other hand, are one of the key factors that currently restricts the discovery potential of HTS techniques, as they require time, energy, and high cost to be identified in wet-lab low-throughput protocols by medicinal chemists [7]. Docking hits urgently need, prior to wet-lab confirmation, a further computational assessment using advanced techniques, based on atomistic MD simulations with explicit solvent, using increasingly accurate all-atom force fields [8]. In this regard, in the last two decades, the so-called alchemical method has emerged as one of the most powerful and promising approaches to the calculation of binding free energies in ligand-receptor systems. The alchemical protocol evaluates the BFE in a thermodynamic cycle as a difference (see Fig. 1) of the solvation energy of the ligand in the bound state and in the bulk solvent [9].

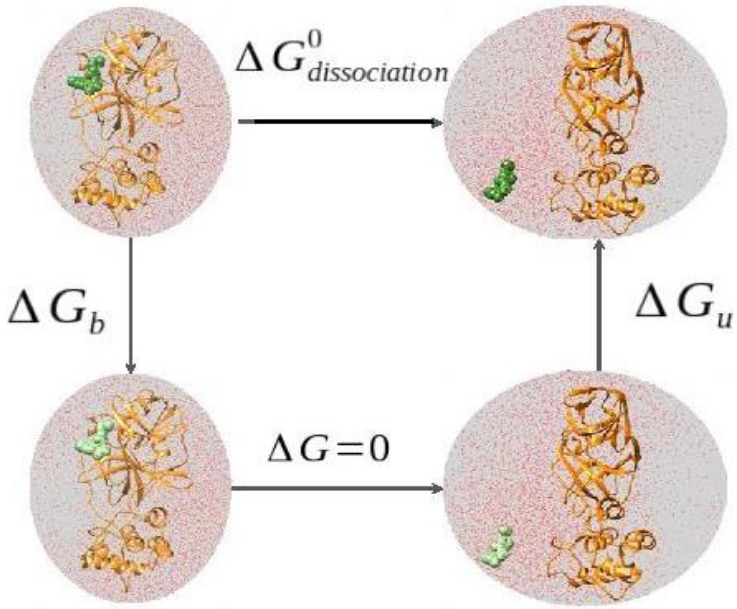


Fig 1: Thermodynamic cycle in alchemical transformations. Left and right legs of the cycle represent the solvation free energy of the ligand in the binding site and in the bulk solvent, respectively. The ligand is shown using van der Waals spheres; dark and light green: fully interacting and decoupled (gas-phase) ligand.

These solvation energies are computed independently by connecting the ligand end-states through a stratification of discrete intermediate states where the ligand-environment interactions are progressively turned off via a λ coupling parameter. The solvation free energies (namely the reversible work to bring the ligand from the gas phase to the condensed phase) are given by the sum of the free energy contributions along the stratification, computed using the free energy perturbation (FEP) approach [9], requiring equilibrium MD simulations for each of the intermediate λ-states. Alchemical techniques are costly and generally limited to the calculations of relative binding free energies (RBFE) of strictly congeneric series of compounds. This is a severe limitation since hits form docking campaigns can be chemically distant and not easily amenable to RBFE calculations through, e.g., intricate "perturbation graphs" gradually connecting the chemically distant interesting molecules, thereby spending computational resources in determining RBFE's between uninteresting intermediate decoys [10]. Quoting Yutong Zhao [11], lead engineer at Relay Therapeutics, FEP-based absolute binding free

energy (ABFE) calculations may easily (absolutely) end up in tears due to daunting sampling challenges emerging at low coupling, when the ligand in the binding site experiences an order-disorder transition from low enthalpy to high entropy states [12]. Paradoxically, the advent of GPUs in scientific calculations, that allows simulating a typical drug-receptor system for up to hundreds ns/day, has strengthened the illusion that a single sufficiently long MD trajectory can achieve correct sampling in FEP applications of complex biomolecular systems. Conformational transitions are in fact sudden events that occur rarely in the time scale that can be attained in a simulation of a single typical ligand-receptor system (microseconds at most).

An important tightly connected limiting factor in ABFE calculations using FEP is the lack of a reliable confidence interval (CI) determination and the sensibility to the initial conditions of the ABFE prediction [13]. Credible CIs can be only assessed by repeating the FEP calculation several times, thereby expanding considerably the computational cost. From an implementation point of view, while the calculations on the windows of the stratification can be run independently, the convergence rate of these concurrent simulations is strongly λ-dependent [9,13], posing severe challenges to an efficient parallelization of the algorithm on HPC systems.

In the last years, in the context of the alchemical approach we have been developing a massively parallel computational variant based on enhanced sampling methodologies of the end states only and on the production of nonequilibrium (NE) alchemical trajectories rapidly connecting the end-states. The ABFE is recovered by computing the work done in these driven NE trajectories exploiting the Crooks [14] and Jarzynski [15] NE theorems on the resulting work distribution. At variance with FEP, these NE-based techniques strictly require that the concurrent NE trajectories are run according to a common alchemical time schedule, hence automatically satisfying a perfect load balance in the embarrassingly parallel implementation. Besides, in the bound leg of the alchemical thermodynamic cycle, enhanced sampling is applied only to the low enthalpy ordered fully coupled ligand states, according to a powerful torsional tempering scheme involving the so-called “hot zone” comprising the ligand and residues of the binding site. During 2020, in the midst of the Covid-19 crisis, we set up an automated workflow for the calculations of MD-based accurate ABFEs starting from the knowledge of the docking pose. The method, termed virtual Double System Single Box (vDSSB) allows to screen on a dedicated HPC platform such as CRESCO6 tens of compounds per day, hence potentially providing the missing link for chemically distant docking hits assessment (usually few tens or hundreds over millions of tested compounds) and false positive filtering. Compound refinement on surviving hits can be effectively and rapidly achieved using unidirectional NE technique applied to RBFE calculations of congeneric series.

This contribution is organized as follows. In the method section we briefly describe the theoretical background of the vDSSB approach discussing its parallel implementation on the CRESCO6 and M100/CINECA HPC platforms. In Section 3 we report on some recent results obtained with this technique on the inhibition of the $3CL^{pro}$ protein, the main protease of the SARS-CoV-2 virus. Concluding remarks and perspectives are drawn in the last section.

2 Non-equilibrium Virtual Double System Single Box: theoretical background and methodological workflow

vDSSB is an inherently parallel approach composed of two distinct computational stages for HPC execution with ideal parallel efficiency applied to both legs of the alchemical thermodynamic cycle (see Fig. 1), namely the HREX stage and, in sequence, the NE stage, followed by the fast post-processing of the resulting work distribution.

2.1 HREX stage: equilibrium sampling of the fully coupled bound end-state and of the decoupled unbound state

This stage involves the canonical sampling of the bound end-state (top left panel in Fig. 1) and of the ligand in the gas-phase (bottom right panel in Fig. 1) embedded as a ghost molecule in equilibrated solvent. These two equilibrium simulations are performed by launching multiple batteries of Hamiltonian Replica Exchange (HREX) simulation with a torsional tempering scheme [16]. Parallel tempering or Replica exchange schemes are paradigmatic low-communication weak scaling parallel algorithms in the MD simulations of complex systems. The HREX torsional variant, by selectively scaling the crucial degrees of freedom for binding, allows to keep the replica number (and hence the computational cost) to a minimum while affording an effective enhanced sampling of the ligand-receptor

rugged conformational free energy landscape. In each HREX battery, the torsional potential of a selected subset of degrees of freedom of the system is scaled along a progression of n replicas up to a maximum scaling factor 1. The lowest *s* factor sets the temperature of the system subset in the highest replica to$T_t = \frac{T_0}{s}$ (where T_0 is the target temperature, 300 K), while the rest of the degrees of freedom of the system remains cold at T_0. In the bound state, the system subset includes the ligand and the residues of the binding site, implemented in a progression of 12 replica states with torsional scaling decreasing from 1 to 0.2 corresponding to a torsional temperature of 1500 K. In the gas-phase state, the starting end-states of the decoupled ligand are efficiently obtained by combining HREX sampling of a single molecule (*s*=0.1 and 8 replicas) in vacuo with pre-equilibrated samples of pure explicit solvent. The HREX sampling of the unbound leg end-state can be performed on a local workstation in a matter of minutes. Higher torsional and nonbonded scaling is needed in the unbound leg due to the strong and unscreened electrostatic interactions that may occur between polar moieties of flexible ligands.

The HREX computational stage starts from a docking-generated ligand-receptor structure, and includes a series of complex operations consisting in i) the generation of the MD engine topology files using appropriate software tools (e.g. PrimaDORAC [17], LigParGen [18], Paramchem [19]) for the ligand force field parameterization; ii) a preliminary minimization of the complex with the user-selected force field; iii) the NPT equilibration in standard conditions of the resulting solvated receptor-ligand structure in a MD box of optimal size; iv) the setting-up of the input files with the definition of the hot-zone in the bound state and specification of the scaling protocols to perform HREM simulations for bound and unbound ligand state on the HPC platforms. These error-prone steps, which are central in the vDSSB approach, have been fully automatized by the middleware [20], HPC_Drug, an effective tool for the guided submission of vDSSB jobs on a HPC system. HPC_Drug is a python application distributed under the GPL that can be cloned from the GitHub repository https://github.com/MauriceKarrenbrock/HPC_Drug. All prerequisite software (GPL) for using HPC_Drug and installation instructions are detailed at the GitHub site. The generation of ready-to-use HREX input files for HPC batch submission is achieved by HPC_Drug in a matter of minutes on the HPC front-end or on a local workstation. Currently, HPC_Drug supports the ORAC MD code [12] and the GROMACS MD suite [20]. ORAC has been designed to run on homogeneous CPU-based architecture such as the CRESCO6 platforms [21]. In ORAC, HREX is implemented using an hybrid OpenMP-MPI approach, whereby scaling factors rather than system configurations are exchanged on the MPI layer during the simulation with a minimal impact on the communication overhead. For GROMACS, HPC_Drug generates the HREX input files exploiting the versatile open-source Plumed library [22].

For a typical drug-receptor systems (50K atoms), the HPC execution of the HREX stage of the bound state using ORAC may typically involve from 32 to 64 Skylake 48-cores nodes on CRESCO6 with 8 to 16 HREX batteries producing a total simulation time of the order of 1-2 microseconds simulation time with 30/60 ns on the target state in one wall clock day. GROMACS/Plumed may produce on the M100/CINECA platform equipped with 4 Volta100 GPU per node up to 200 ns a day on the target state engaging 36 nodes. The HREX output consists in uncorrelated end-state snapshots, sampled at regular time intervals on the target state, providing a high-quality starting set of equilibrium system configurations for the subsequent NE stage.

2.2 NE stage: production of the NE alchemical trajectories

The NE stage is launched, for each leg of the cycle, in a single parallel job starting from the configurations generated by the preceding HREX stage. In the annihilation job (bond-state leg) and in the growth job (unbound-state lef), each MPI instance read its own equilibrium starting configuration and then proceeds independently (no communication) completing the NE alchemical simulation in the *common* prescribed time schedule[1] with a perfect load balance by design. The output of these independent runs consists in files where the work done on the driven system by the switching off (bound leg) or on (unbound leg) of the ligand-environment alchemical coupling parameter is saved at regular intervals. Examples of work time recorded during the NE trajectories are reported in Fig. 2. The main output of NE stage growth and annihilation jobs is hence a N_{MPI}-sample of *final* growth and annihilation work values. The final growth work values refer to processes where we start with the equilibrated ligand

1 For the annihilation job, the atomic charges of the ligand are first turned off in about 100-200 ps, followed by ligand-environment Lennard-Jones interaction in, typically, 300-500 ps.

in the gas-phase and we end in a non-equilibrium state with a fully coupled ligand in the solvent. The final annihilation work values refer to processes where an equilibrated ligand is brought into the gas-phase ending up in a NE conformation. Typically, on the CRESCO6 cluster, for Covid-19 related drug-receptors pairs, we launched $N_{MPI} = 640$NE trajectories in each leg engaging 80 Skylake nodes. Using the ORAC code, the annihilation job (bound state) for a typical system (50K atoms) is completed on CRESCO6 in few wall clock hours, while the growth job (unbound state) for a drug-size ligand (5K atoms) takes few tens of wall clock minutes. The total simulation times produced by the jobs is of the order of 0.3 microseconds for each leg of the cycle. Using GROMACS on M100/Cineca, wall-clock times for the NE stage are cut by a factor of 4 to 5.

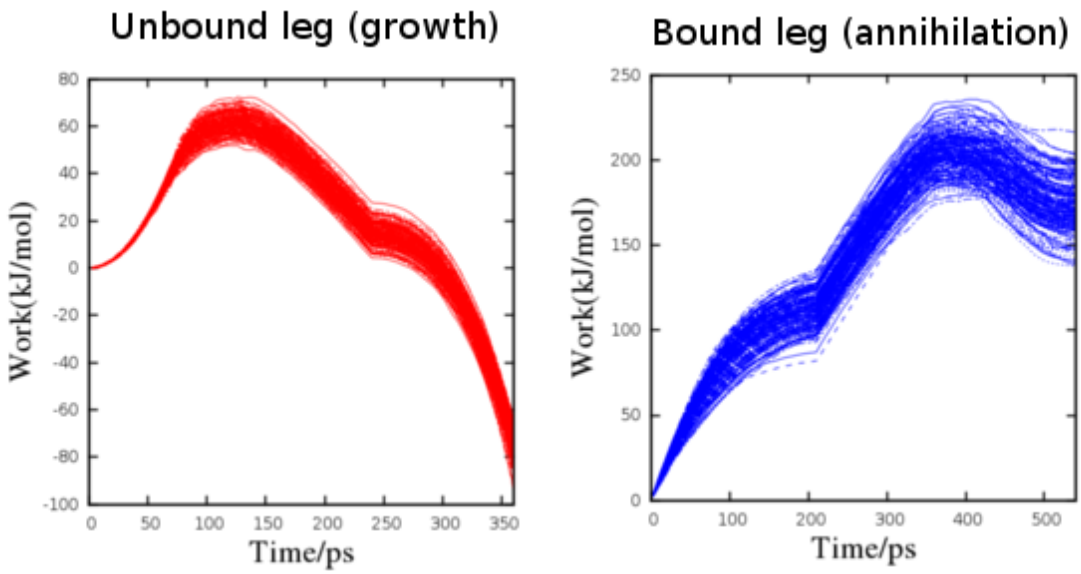


Fig. 2: Alchemical work trajectories for a typical ligand of the SARS-CoV-2 main protease.

2.3 Post processing stage: production of ligand decoupling equilibrium sampling of the fully coupled bound end-state and of the decoupled unbound state

The bound and unbound work values are two sets of independent random variables. We can hence emulate a "double-system-single-box" by combining each annihilation work with each growth work hence producing a sample total of N_{MPI}^2work values for a vDSSB NE process where the ligand is annihilated on the protein and is simultaneously grown in the bulk. In terms of work distribution, this operation corresponds to the *convolution* of the annihilation and growth work distributions, as it is shown in equation 1:

$$P_b(W) * P_u(W) = \int P_b(W) P_u(W - w) dw \qquad (1)$$

In Fig.3 we show the work distributions obtained for the final work values reported in Fig.2. The estimate of the dissociation free energy ΔG_{b+u}can be readily calculated from the combined growth and annihilation work values by exploiting the Jarzynski identity [15] or, equivalently, by assuming that the convolution distribution can be represented as a combination of Gaussian mixture and using the Crooks theorem [13]. The number and weights of the normal components can be evaluated using the Expectation-Maximization algorithm [23]. The so-computed dissociation free energy must be corrected[9] by adding a standard state dependent term of the form $\Delta G_{vol} = RTln\left(\frac{V_{site}}{V_0}\right)$, where $V_0 = 1661$ Å^{3}is the standard state volume (corresponding to 1M concentration) and $V_{site} = 4\pi\frac{(2\sigma)^3}{3}$is the binding site volume estimated from the variance $\sigma^2$of the distance between the centers of mass of the ligand and the receptor in the HREX simulation of the bound state. The ligand-receptor dissociation constant (in M units) is finally given by equation 2:

$$K_d = exp\left[\frac{-\left(\Delta G_{b+u} + RTln\left(\frac{V_{site}}{V_0}\right)\right)}{RT}\right] \quad (2)$$

The confidence interval on ΔG_{b+u} can be reliably estimated by bootstrapping the N_{MPI} uncorrelated growth and annihilation work values *before* performing the convolution.

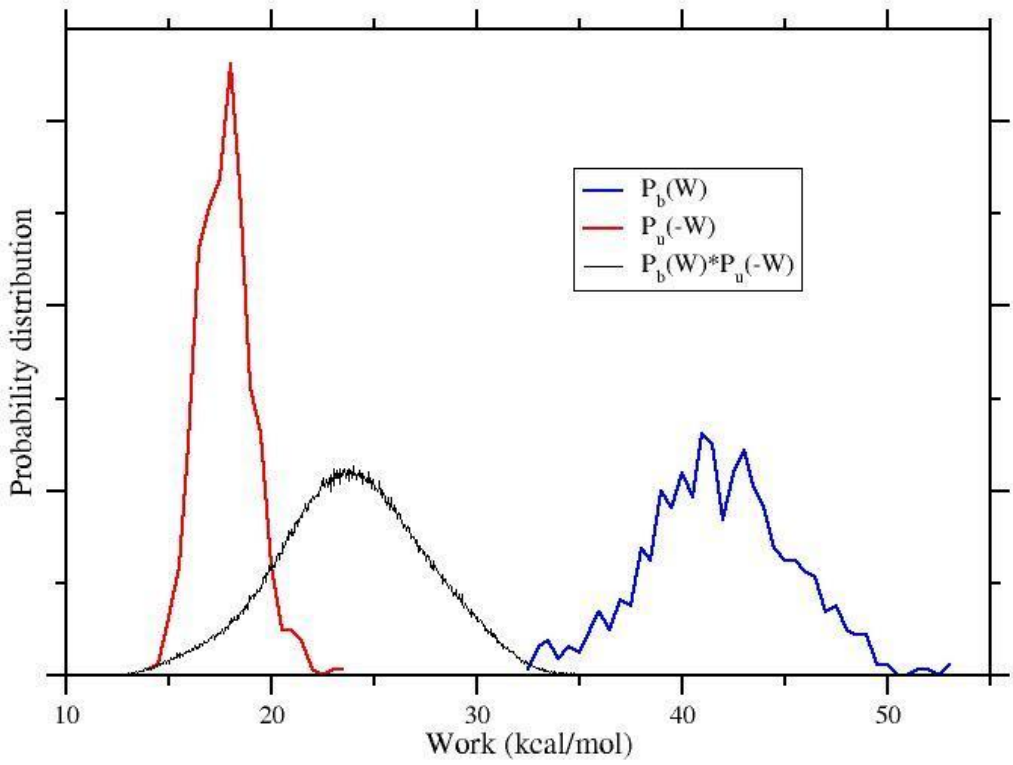


Fig. 3: Growth and annihilation final work distribution for a typical Covid-19 related ligand of the main protease.

The post-processing stage has been fully automated by simple application scripts to be executed on the front-end once the HREX stage and the subsequent NE stage have been completed. Further technical details of the vDSSB methodology are provided in Refs. 23 and 24.

3 Assessing docking hits for the inhibition of the SARS-CoV-2 main protease using vDSSB on CRESCO6.

The SARS-CoV-2 main protease ($3CL^{pro}$) is a non-structural protein that cleaves the pp1a and pp1ab polyproteins expressed by the viral m-RNA upon host cell entry. $3CL^{pro}$ is generated by self-excision from the pp1a polyprotein following dimerization. The catalytic activity of $3CL^{pro}$ is expressed by the $(3CL^{pro})2$ dimer. The $3CL^{pro}$ monomer is in turn composed of two loosely coupled units, the chymotrypsin-like domains I+II (residues 1-197), harbouring the catalytic site, and the cluster of helices domain III (residues 198-304), regulating dimerization via two intertwined salt bridges involving ARG4(A)-GLU290(B) and GLU290(A)-ARG4(B) of the A and B protomers. The dimer is characterized by two symmetric extended clefts for pp1a, pp1ab adhesion. Each dimer cleft ends at the solvent exposed catalytic site with the CYS145-HIS41 proteolytic dyad. The two catalytic dyads, on opposite sides of the dimer and far from the monomer-monomer adhesion surface, very likely act

independently for maximizing the catalytic efficiency. To identify an effective $3CL^{pro}$ inhibitor, only the I+II chymotrypsin catalytic domain needs to be considered in the ligand receptor simulation of the bound state. In Fig. 4, we show, as an example, the chymotrypsin domain of $3CL^{pro}$ bound to hydroxychloroquine [26].

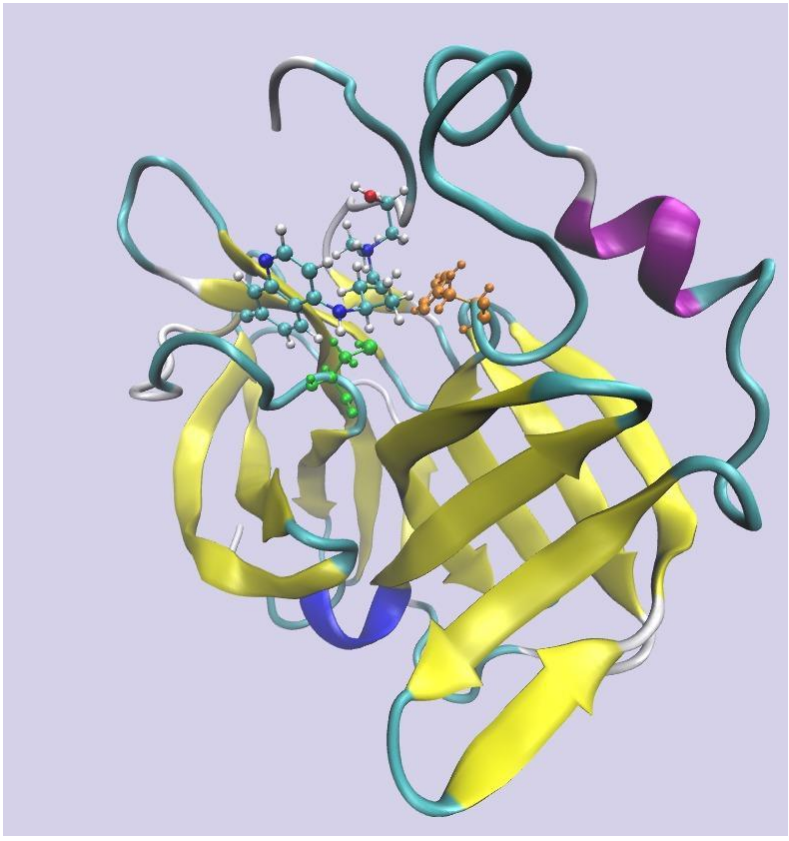
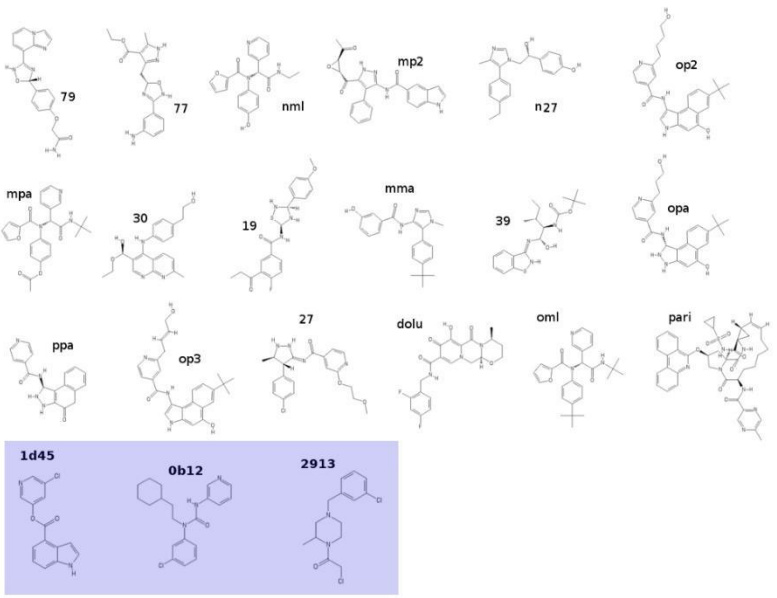


Fig. 4: Chymotrypsin domain of the SARS-CoV-2 main protease in complex with hydroxychloroquine. The catalytic dyad CYS145-HIS41 is in orange color.

In Ref. 25, we have identified via multimodal structure-based design and molecular docking several possible inhibitors of the SARS-CoV-2 $3CL^{pro}$ protease. Most of these putative chemically distant inhibitors are not commercially available. The experimental validation hence requires the synthesis of these non-congeneric compounds. In Fig. 5, we present the results obtained on CRESCO6 using the vDSSB approach for the dissociation free energies of these docking hits. The compounds within the light blue background in the left panel have been identified in the context of the Covid-19 Moonshot initiative as possible $3CL^{pro}$ inhibitors. Their activities were experimentally determined using fluorescence techniques. vDSSB-predicted dissociation free energies for these compounds are in fair agreement with the experimental data (within 1:1.5 kcal/mol), except for the 1d45 compound where the predicted value for the dissociation free energy is significantly smaller than the experimental counterpart. However, while 1d45 is labelled as a non-covalent $3CL^{pro}$-binder according to the Covid-19 Moonshot activity data [27], the same compound was found to be a potent covalent inhibitor of the highly homologous SARS-CoV-1 main protease with approximately the same dissociation free energy ($\Delta G_{exp.}$ = 10.3 kcal/mol) [28]. Covalent binding (that is not accounted for in vDSSB of FEP-based techniques) may explain the observed difference between experimental and calculated dissociation free energy for 1d45-$3CL^{pro}$ interaction.
Assuming as a threshold for activity $K_d = 1\mu M$, only compound 27 among the docking hits identified in Ref. 25 survives after vDSSB assessment. The compounds marked in red color in the right panel of Fig. 5 are hence false positives according to our calculations. With the availability of the entire CRESCO6 cluster (~500 nodes), these false positives could have been identified in less than 2 wall-clock days with an estimated cost of less than 100K Euros (assuming a cost of 0.1 euro cent per core

hours), a time and a cost incomparably smaller than that needed in low-throughput medicinal chemistry involving in all cases the synthesis of the compound.



Ligand	ΔG^0_J(kcal/mol)	$\Delta G_{exp.}$(kcal/mol)
pari	10.7 ± 0.6	-
opa	9.4 ± 0.5	-
ml	9.0 ± 0.7	$7.9^{(a)}$
op3	8.8 ± 0.8	-
27	8.3 ± 0.5	-
39	7.6 ± 1.5	-
mma	7.4 ± 0.8	-
ppa	7.3 ± 0.8	-
op2	7.1 ± 2.0	-
dolu	6.7 ± 1.6	-
19	6.5 ± 0.7	-
30	6.1 ± 0.9	-
mp2	5.8 ± 0.6	-
mpa	4.3 ± 1.2	-
77	4.0 ± 0.4	-
n27	3.9 ± 0.6	-
79	3.7 ± 0.9	-
nml	3.1 ± 0.4	-
1d45	5.4 ± 0.8	10.0
0b12	9.3 ± 0.9	7.46
2913	5.8 ± 0.8	7.0

(a) BFE for 3CLpro of SARS-CoV-1 [Jakobs, J. Med. Chem 56, 534-546, 2013]

Fig. 5: Dissociation free energies computed on CRESCO6 with the vDSSB approach.

4 Conclusion

Most of the (successful) coordinated efforts by governments and big-pharma industry in Covid-19 therapy has been directed toward the SARS-CoV-2 viral structural proteins (spike (S), nucleocapsid (N), membrane (M), and envelope (E)). Pfizer and Moderna mRNA vaccine, based on the translation of the prefusion Spike protein, are the chief achievement of this extraordinary effort. Despite the success, we must nonetheless remain vigilant as SARS-CoV-2 can evolve to wane vaccine efficacy or new zoonotic pandemics can arise in the future. On the other hand, research for antiviral therapeutic agents have been erratic and poorly coordinated. During the pandemic, drug discovery has disorderly dispersed in plethora of biological targets including human proteins such as ACE2, bromodomain, sigma receptors, immunophilins, kinases, while a well-coordinated collaborative effort in targeting the highly conserved nonstructural proteins involved in the viral life cycle (such as the precursor SARS-CoV-2 $3CL^{pro}$ and PL^{pro} proteases) could have maybe delivered an effective and specific antiviral agent.

In this context, HPC-driven vHTS can provide an effective tool for a rationalization of drug-discovery process with the potential to substantially increase the productivity of pharmaceutical research, especially in the field of antiviral design. We do believe that the paradigm based on non-equilibrium thermodynamics, relying on rigorous and general enhanced sampling approaches for the end-states and on the equal-time condition for the NE swarm of alchemical trajectories, is effective in by-passing the well-known sampling-related pitfalls and entanglements of conventional equilibrium based MD-based technologies such as FEP+ [29] that have prevented so far its widespread use in industrial setting. NE-based vDSSB has the potential to provide, through fully automated procedures, a firm framework for an HPC based measurement of K_d in realistic thermodynamic conditions with minimal end-user intervention/tweaking. Accuracy of the prediction via SF-vDSSB can further benefit from the progress in force field parameterization (e.g., polarizable force fields, QM/MM schemes, Car-Parrinello MD schemes) from better algorithm for force kernels and from increased capacity of the hardware technology. While the competence for the setting up of the Docking/NE-vDSSB hit-to-lead framework on a HPC system are high and with a strong interdisciplinary character, the required skills for its bare use are limited. vDSSB may hence provide the missing link for the setting up of an automated HPC-based instrument for hit-to-lead in drug discovery.

References

[1] A. Acharya *et al.* Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19, J. Chem. Inf. Model. **60**, pp. 5832-5852, (2020).

[2] O. Trott and A. J. Olson. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. J. Comput. Chem., **31,** pp. 455-461, (2010).

[3] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet. Directory of useful decoys, enhanced (dud-e): Better ligands and decoys for better benchmarking. J. Med. Chem., **55**, pp. 6582-6594, (2012).

[4] M. Wójcikowski, P. Ballester, and P. Siedlecki. Performance of machine-learning scoring functions in structure-based virtual screening. Scientific Reports, **7**, pp. 1-10 (2017).

[5] L. Masters, S. Eagon, and M. Heying. Evaluation of consensus scoring methods for AutoDock Vina, smina and idock. J. Mol. Graph. Model., **96,** 107532, (2020).

[6] L. Casbarra, and P. Procacci, J. Comput. Aided Mol. Des., *in press* DOI:10.1007/s10822-021-00388-4 (2021).

[7] J. Bibette. Gaining confidence in high-throughput screening. Proc. Natl. Acad. Sci., **109**, pp. 649-650 (2012).

[8] F. Palazzesi, M. K. Prakash, M. Bonomi, and A. Barducci. Accuracy of current all-atom force-fields in modeling protein disordered states, J. Chem. Theory Comput., **11**, pp. 2-7 (2015).

[9] A. Pohorille, C. Jarzynski, and C. Chipot. Good practices in free-energy calculations. J. Phys. Chem. B, **114**, pp.10235-10253 (2010).

[10] L. F. Song, T.-S. Lee, C. Zhu, D. M. York, and K. M. Merz. Using amber18 for relative free energy calculations. J. Chem. Inf. Model., **59**, pp. 3128-3135, (2019).

[11] https://twitter.com/proteneer/status/1376314932732571658

[12] R. K. Pal and E. Gallicchio. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. J. Chem. Phys., **151**, pp. 124116 (2019).

[13] P. Procacci. Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics. Curr. Op. Struct. Biol. Biology **67**, pp. 127-134 (2021).

[14] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. J. Stat. Phys. **90**, pp. 1481–1487 (1998).

[15] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. Phys. Rev. Lett., **78**, 2690-2693 (1997).

[16] S. Marsili, G. F. Signorini, R. Chelli, M. Marchi, and P. Procacci. Orac: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. J. Comput. Chem., **31**, pp. 1106-1116 (2010).

[17] P. Procacci. PrimaDORAC: A Free Web Interface for the Assignment of Partial Charges, Chemical Topology, and Bonded Parameters in Organic or Drug Molecules J. Chem. Inf. Model. **57**, 1240-1245 (2017).

[18] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W. L. Jorgensen LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands. Nucleic Acids Research, Volume 45, Issue W1, 3 July 2017, Pages W331-W336.

[17] K. Vanommeslaeghe, E. Prabhu Raman, and A. D. MacKerell, Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. J. Chem. Inf. Model. **52**, pp. 3155-3168 (2012).

[19] M. Karrenbrock, HPC_DRUG: A middleware python tool for computational drug discovery on HPC architectures, https://github.com/MauriceKarrenbrock/HPC_Drug (2021).

[20] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. Bioinformatics, **29**, 845-854 (2013).

[21] F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, F. Palombi, G. Ponti, G. Santomauro, P. Procacci. CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 International Conference on High Performance Computing & Simulation (HPCS), pp. 1051-1052 IEEE Dublin (2019).

[22] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi. PLUMED2: New feathers for an old bird, Comp. Phys. Comm. **185**, 604-613 (2014).

[23] J. A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Tech Rep. International Computer Science Institute, Berkeley CA (1998).
[23] M. Macchiagodena, M. Karrenbrock, M. Pagliai, G. Guarnieri, F. Iannone, P. Procacci. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. J. Chem. Theory Comput. **16**, pp. 7160-7172 (2020).
[24] M. Macchiagodena, M. Karrenbrock, M. Pagliai, G. Guarnieri, F. Iannone, P. Procacci. Chapter: Nonequilibrium Alchemical Simulations for the Development of Drugs Against Covid-19, in Methods in Pharmacology and Toxicology, Springer, New York, NY (2021).
[25] M. Macchiagodena, M. Pagliai, P. Procacci, Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. Chem. Phys. Lett. **750**, pp. 137489 (2020).
[26] P. Procacci, M. Macchiagodena, M. Pagliai, G. Guarnieri, F. Iannone. Interaction of hydroxychloroquine with SARS-CoV2 functional proteins using all-atoms non-equilibrium alchemical simulations. Chem. Comm. **56**, pp. 8854-8856 (2020).
[27] J. Chodera, A. A. Lee, N. London, F. von Delft. Crowdsourcing drug discovery for pandemics. Nature Chemistry, **12**, pp. 581-583 (2020).
[28] A. K. Ghosh, G. Gong, V. Grum-Tokars, D. C. Mulhearn, S. C. Baker, M. Coughlin, V. S. Prabhakar, K. Sleeman, M. E. Johnson, A. D. Mesecar. Design, synthesis and antiviral efficacy of a series of potent chloropyridyl ester-derived SARS-CoV 3CLpro inhibitors. Bioorg. Med. Chem. Lett. **18**, pp. 5684–5688 (2008).
[29] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner, R. Abel. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. J. Am. Chem. Soc. **137**, pp. 2695–2703 (2015).

MULTISCALE MODELING OF THE WILD-TYPE AND ALPHA VARIANT SARS-COV-2 SPIKE PROTEIN

Marco Lauricella[1*], Letizia Chiodo[2], Fabio Bonaccorso[3,4], Mihir Durve[3], Andrea Montessori[1], Adriano Tiribocchi[1], Alessandro Loppini[2,3], Simonetta Filippi[2] and Sauro Succi[3,1]

[1]*Consiglio Nazionale delle Ricerche, Istituto per le Applicazioni del Calcolo IAC-CNR, 00185, Rome Italy*

[2] *Campus Bio-Medico University, Engineering Department, 00128, Rome, Italy*

[3] *Istituto Italiano di Tecnologia, Center for Life Nano- & Neuro-Science@Sapienza – IIT, 00161, Rome, Italy*

[4] *Department of Physics and INFN, University of Rome "Tor Vergata", 00133 Rome, Italy*

ABSTRACT. Physiological solvent flows act in such a way as to promote collective motions of biological structures. In virus/host-cell interactions, the solvent flow may facilitate the virus adhesion on the target receptors and drive the hierarchy of multivalent adhesion mechanisms.
To elucidate functional interactions between flows and molecules, we couple the all-atom atomistic molecular dynamics (for proteins) with the computational Lattice Boltzmann fluid dynamics (for solvent flows).
Preliminary results are presented for SARS-CoV-2 viral spike glycoprotein S in implicit solvent, as well as for its Alpha variant. Our multiscale simulations are performed with the LAMMPS classical molecular dynamics code, used within a customized installation on Cresco6 at ENEA.
The mesoscopic solvent description is critically compared to the all-atom solvent model, to quantify advantages and limitations of the multiscale description.

1 Introduction

In the last decades, computational tools have undergone spectacular methodological and technological progress, which can play a decisive role in fighting contagious diseases, by providing *in silico* simulations of biological molecules and drugs design. Concurrently, computing infrastructures endure great technological progress, largely fueled by the computing power of the graphics processing units (GPUs). The confluence of such major advances spawns opportunities to develop a new multiscale modelling tool by coupling a mesoscale solvent representation to the molecular dynamics to obtain an efficient computational biomedicine tool able to boost the simulation power and the understanding of the biological mechanisms at the atomistic level.
Nowadays, the molecular dynamics (MD) method has shown its massive potential in describing at the atomistic level the biological mechanisms underlying the activities of several proteins. Remarkable examples in computational biomedicine are the recent simulations of an entire cell organelle, a photosynthetic chromatophore vesicle from a purple bacterium [1] or the study of the N-Methyl-D-Aspartate (NMDA) neuroreceptor by the DE Shaw research group [2]. As of today, an enormous scientific effort has been spent to investigate in-silico the molecular behaviour of SARS-CoV-2 proteins, both for drug repurposing [3] and for antibody design. Standard MD simulations have been used, for example, to estimate binding free energies of spike in interaction with the human angiotensin-converting enzyme 2 (ACE2) receptor [4,5,6,7] alongside with their interaction scores [8].

*Corresponding authors. E-mail: marco.lauricella@cnr.it; l.chiodo@unicampus.it

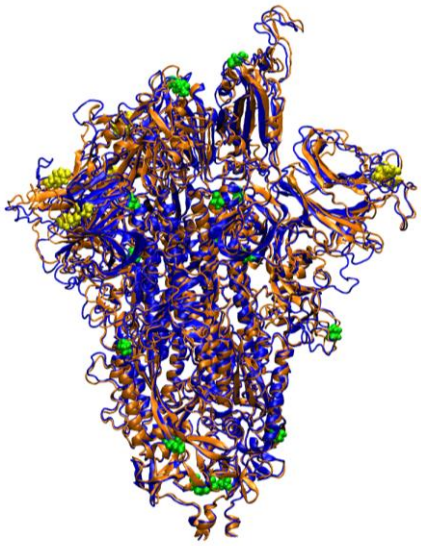


Fig. **1**: Secondary structure of the spike protein, wild type (blue) and Alpha variant (orange) as obtained from experimental pdb (6vsb.pdb) and homology modelling, respectively. In yellow and green, deletions and mutations of the wild type giving the Alpha variant are highlighted.

Nonetheless, the long time scales of the movements associated with the allosteric and functional response of biological mechanisms normally lie in several microseconds, beyond the actual high-performance computational limits to obtain a statistically meaningful description [9,10], even by exploiting optimised codes for GPUs clusters [11]. Thus, in the last three decades, the development of coarse-grained models has shown great scope in overcoming these limits. The coarse-grained strategy usually aims at reducing the details of the protein structures alongside their aqueous solvent. Such a reduction shall be made with particular care to preserve the detailed description, where necessary to appropriately describe the protein structure and function.

The coexistence of different time- and length- scales necessarily requires the use of a multiscale approach in new computational biomedicine tools, that shall be pursued with prompted concern in the actual pandemic scenario [12**,**13].

The present work exploits a multiscale description based on the Lattice Boltzmann (LB) method for the solvent fluid of the aqueous combined with all-atom molecular dynamics (MD) description for the protein structures. Several multiscale approaches to coupling LB/MD are already available in the literature. Here, the coupling of the LB velocity field with MD objects is implemented via a Stokes friction term in the overlay region of the two descriptions to realise the multiscale description [14**]**.

As a test case, we use the SARS-CoV-2 spike protein S, a heavily glycosylated protein anchored in the viral membrane. It is constituted by three chains, each one made of an identical primary sequence of more than 1200 amino acids of which 1146 form the extracellular domain. Each chain of the trimer is composed of two fragments: the receptor-binding fragment S1, containing the receptor-binding domain (RBD), interacting with ACE2, and the fusion fragment S2 [15]. The S protein is cleaved by a furin-like protease at residue 686 into the S1 and S2 fragments [16], initiating the membrane fusion process. We also study the Variant of Concern 202012/01 (lineage B.1.1.7), also commonly referred to as Alpha variant (α-Spike).

Equilibrium properties, such as rigidity and elasticity of specific sites, are pivotal for binding and other functional activities of the viral protein [15,17], therefore they must be properly described within the multiscale model. We compare equilibrium state properties from LB/MD data obtained by all-atom molecular dynamics, to assess the quality and efficiency of the multiscale description. The key role of water molecules for protein structure and function is highlighted [18].

In perspective, our multiscale approach could permit us to overcome the statistical sampling limitation affecting current explicit solvent atomistic description, in particular providing a unique tool to investigate large conformational changes and dynamics solvent flow effects.

2 Methodology

2.1 Simulated systems

Based on available cryo-EM data [19,20], we built the two models, hereafter called Spike and α-Spike. By relying on the 6vsb.pdb structure [10], we added the missing loops in the receptor-binding domain (RBD) as obtained from the 6m17.pdb structure [20]. The glycans from the 6vsb.pdb structure are used in these simulations. The α-Spike has been modelled via I-Tasser [21], using the wild-type as template, by including in the sequence [22, 23] three deletions: ΔH69/ΔV70 and ΔY144, and six mutations; N501Y, A570D, P681H, T716I, S982A, and D1118H (Fig. 1).

The protonation states have been calculated via Playmolecule webserver [24], based on PROPKA 3.1 [25] to determine protein pKa values, and on PDBTOPQR 2.1 [26] to optimize the protein for favourable hydrogen bonding.

The cell of the LB/MD systems is a cubic box of 19.2 nm side length, surrounding the all-atom glycosylated-protein (53k atoms). The same box size is used for the all-atom simulations (676k atoms). The cell is built and neutralized (100 nM solution) via the Solvate and Ionize plugins included in Visual Molecular Dynamics (VMD) [27].

2.2 Coupled Lattice-Boltzmann and Molecular Dynamics (LB/MD)

LB/MD simulations have been performed with LAMMPS (stable release 3 March 2020) [28] on the CRESCO-6 cluster based on Intel(R) Xeon(R) Platinum 8160 with 24 cores and two CPUs per computing node.

The two systems (wild-type and mutant) endured an initial 10000 steps of conjugate gradient minimization and were afterwards equilibrated in an NVT ensemble, with the temperature increased to 310 K over 2.0 nanoseconds of MD simulations. Hence, both systems were evolved in time for 500 nanoseconds (NVT, 310 K). The direct summation method was exploited to assess the Coulomb interactions. In particular, the additional screening of the solvent was modelled by a Coulomb correction for implicit solvent interactions which exploits a distance-dependent dielectric permittivity, scaling with an additional 1/r term included in the Coulomb formula. The cut-off of the intermolecular interactions was set to 7 Å corresponding to the Bjerrum length in water [29]. The CHARMM36 force field [30] has been used to model inter- and intra-molecular interactions, including the glycan and N-linked glycan bond descriptions.

The aqueous solvent is described by a specific mesoscale technique, known as the Lattice Boltzmann (LB) method, namely a minimal lattice version of the Boltzmann equation. The LB approach allows modelling the dynamic behaviour of fluid flows without directly solving the Navier-Stokes equations of continuum fluid mechanics. In this framework, the solvent is treated via a fictitious ensemble of particles, whose motion and interactions are confined to a regular space-time lattice. The dramatic reduction of the degrees of freedom associated with the velocity space is the main advantage of the LB approach. Thus, the solvent is described in terms of probability to find a certain quantity of solvent particle at position $\vec{r}$ and time t moving with velocity $\vec{c_i}$ along a possible grid direction. In the LB approach, the particle collisions are represented through a relaxation to the local equilibrium. Here, we rely on the simplest form of the collision operator that is the celebrated Bhatnagar-Gross-Krook operator, where the operator is a simple single-time relaxation term [31].

The standard LB scheme in single-relaxation time (BGK) form reads as follows:

$$f_i(\vec{r} + \vec{c_i}\Delta t, t + \Delta t) = f_i(\vec{r}, t) - \omega\left[f_i(\vec{r}, t) - {f^{eq}}_i(\vec{r}, t)\right] + S(\vec{r}, t) \quad (1)$$

where f is the discrete Boltzmann distribution associated with the discrete velocity $\vec{c_i}$, i = 0, b running over the discrete lattice, in our case 19-speed lattices, commonly denoted D3Q19. In Eq. 1, the relaxation frequency ω is used to set the kinematic viscosity ν of the fluid by the relation $\omega = 2/(6\nu + 1)$, while ${f^{eq}}_i$ is the lattice local equilibrium, basically the local Maxwell-Boltzmann distribution truncated to the second order in the Mach number. mass density and mass flow are obtained in terms of moments of the distribution functions:

$$\rho = m\sum_{i=0}^{b} f_i(\vec{r}, t) \quad (2)$$

$$\rho\vec{u} = m\sum_{i=0}^{b} f_i(\vec{r}, t)\vec{c_i} \quad (3)$$

where m denotes a scaling mass factor set to obtain the correct water density of 993.4 kg/m^3 at 1 atm and 310 K. The fluid and particles are coupled as follows. The effect of the particle on the surrounding fluid is modelled via a friction force term, $\overrightarrow{F_{nj}} = \gamma(\vec{v}_n - \vec{u}_j)$, where $\vec{v}_n$ denotes the particle velocity and $\vec{u}_j$ the fluid velocity at the particle position obtained by a linear interpolation over the nearest eight lattice points. The coupling force is then added to Eq. 1 by the extra force term $S(\vec{r}, t)$. Hence, an equal and opposite force is applied to the particle to model the counterpart of the coupling term (from the fluid to the particle). Following the literature [32], the friction coefficient γ is taken equal to $0.1\ fs^{-1}$, while the kinematic viscosity was set equal to 0.07 Å^2 /*fs* corresponding to the water kinematic viscosity at 310 K. The LB scheme is evolved in time step by step with the MD integration scheme, with a timestep equal to 2.0 femtoseconds.

2.3 All-atom molecular dynamics (AA-MD)

AA-MD simulations have been performed with GROMACS-2020.6 [33] on the Cineca Marconi100 cluster, based on IBM Power9 architecture and Volta NVIDIA GPUs. Both Spike and α-Spike endured an initial 10000 steps of conjugate gradient minimization. Hence, the systems were equilibrated up to 310 K over 2.0 nanoseconds, as in the LB/MD simulations. The MD simulations last 400 nanoseconds (NPT, 310 K, 1 atm). The simulation timestep is 1.6 femtoseconds. Periodic boundary conditions were used, with particle-mesh Ewald long-range electrostatics, using a grid spacing of 1.5 Å along with a fourth-order B-spline charge interpolation scheme. Both the Coulomb and Lennard-Jones interactions use a cut-off of 12 Å with a force switching function acting from 10 Å [34]. The CHARMM36 force field [30] has been used to model inter- and intra-molecular interactions, including the glycan and N-linked glycan bond descriptions. The water is described via the TIP3P water model as implemented in CHARMM [35] which specifies a 3-site rigid water molecule with Lennard-Jones parameters and charges assigned to each of the 3 atoms.

3 Results and Discussion

Using the AA-MD and LB-MD trajectories for Spike and α-Spike, we calculated the Root Mean Square Deviation (RMSD) and the Root Mean Square Fluctuations (RMSFs) (for chain A, that has the RBD in open conformation). We also performed the principal component analysis (PCA) of the internal motion involving the linked S1 and S2 fragments and evaluated the cross-correlation matrix (CCM) between pairs of Cα. Comparison with experimental or MD data, when available, is reported.

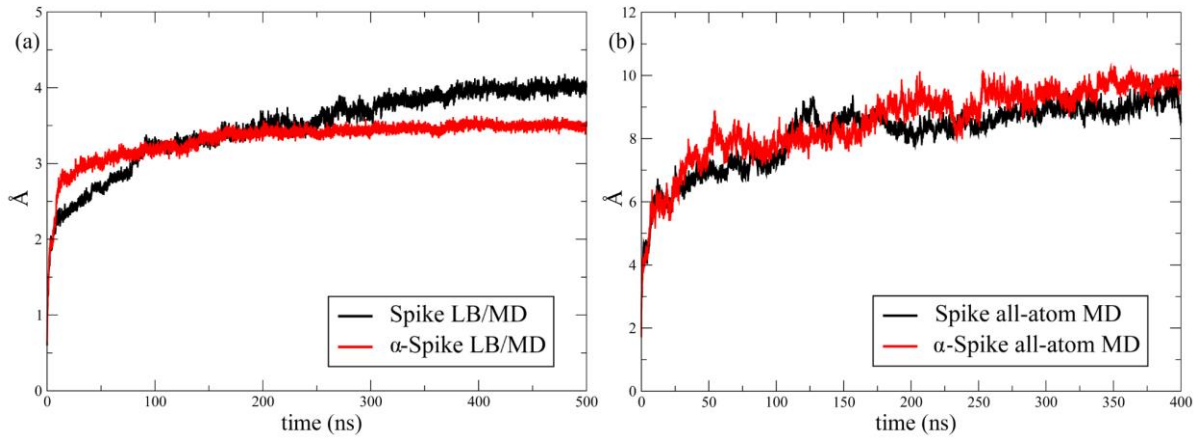


Fig. 2: Root Mean Square Deviation of the Cα atoms, for LB/MD (left panel) and all-atom MD (right panel) simulations, for the 400 ns long trajectories. Slightly differences are observed between wild and mutated. The intensity decrease for the LB/MD case is due to the low mobility of the protein.

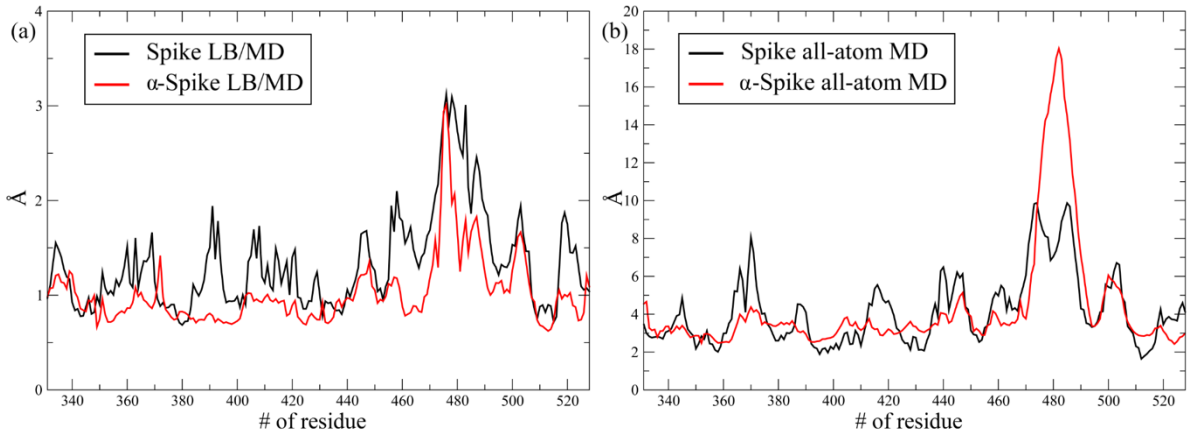


Fig. **3**: Root Mean Square Fluctuations (Cα atoms, RBD) for LB/MD (left panel) and AA/MD (right panel) simulations [15]. Slightly differences are observed between wild and mutated. The significant intensity decrease in the LB/MD is due to the low protein mobility. This result, expected on qualitative grounds, to the best of our knowledge was never inspected on quantitative grounds before.

The RMSD (Fig.2) shows that the equilibration of the AA/MD is quite slow, with large changes from the initial models as obtained from Cryo-EM data and homology modelling. The relaxation is much faster in LB/MD, and changes are smaller with respect to starting structure. Also, RMSD fluctuations are larger in AA/MD. Part of the observed differences are ascribable to the intrinsic rigidity of the LB/MD model, but we cannot exclude a role of the starting Cryo-EM structure.
The protein rigidity in the LB/MD description is confirmed by the RMSFs (Fig.3). However, despite the intensity differences, the LB/MD is able to obtain a qualitative agreement with AA/MD, with fluctuations observed in the same regions. We highlight that the RBD (res 331-528) is less flexible in the α-Spike than in the wt-Spike, both in LB/MD and AA/MD simulations, apart from the recognition loop L3 (at residue 480). L3 is a loop of interest because its rigidity has been related to a stable interaction with ACE2 in an MD study [17].
The PCA (Figs. 4-5) is also quite indicative, for the sake of methods' comparison. Focusing on Spike/α-Spike comparison, their description is quite similar in the framework of the same method. Obviously, due to the lower flexibility of the LB/MD protein, the size of the modes is almost 3-4 times larger in AA/MD. Moreover, also the weight of the various modes is different, meaning that different regions of the protein show different flexibility changes when the solvent is described all-atom or at the mesoscale.

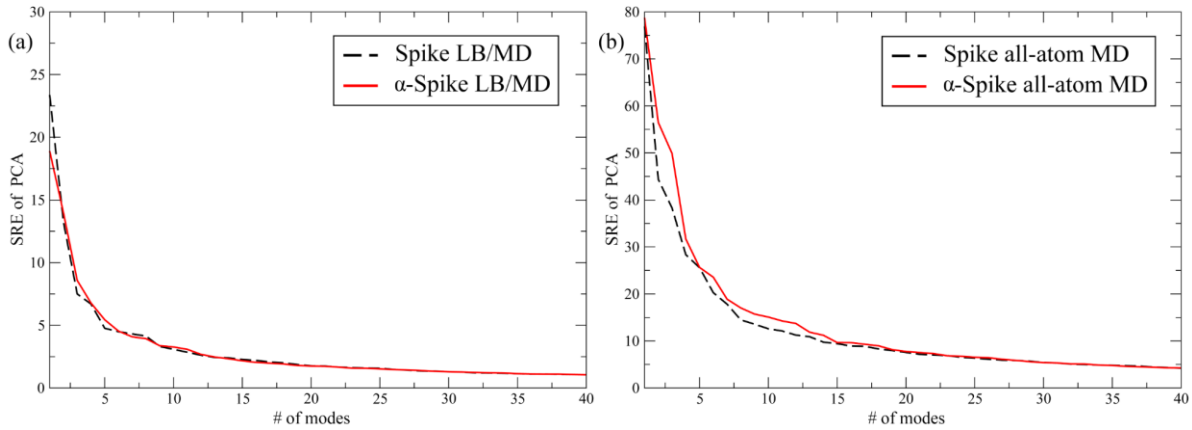


Fig. **4**: Square root of the eigenvalues (SRE) of the largest 40 modes from the PCA, for chain A of LB/MD (left panel) and AA/MD (right panel) simulations. Slightly differences are observed between wild and mutated. A strong intensity decrease is observed for the LB/MD cases, corresponding to the reduced mobility of the protein.

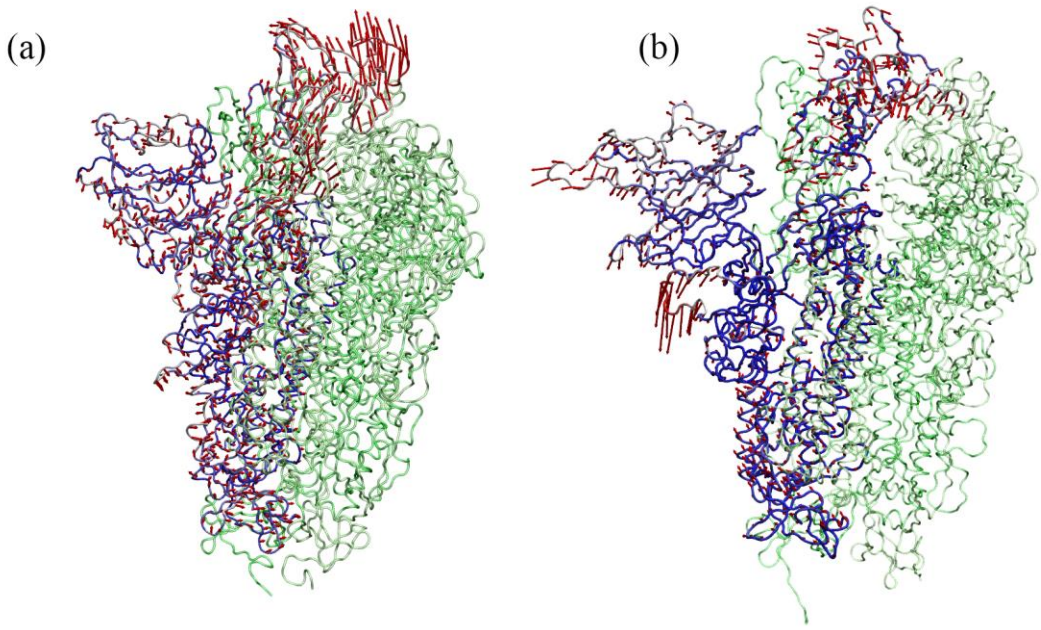


Fig. **5**: The main variation mode in principal component analysis (PCA) for chain A of the wild type Spike protein in the LB/MD simulation (a) and in the AA/MD simulation (b).

Finally, the map of the CC (Fig. 6) shows something comforting, because, overall, correlations between residues are reasonably described by LB/MD with respect to AA/MD. This result is due to the fact that intramolecular local interactions, not mediated by solvent, are kept and well described in the LB/MD method. To note that, in both methods, the α-Spike presents higher correlation values with respect to the wild type, denoting that mutations indeed affect the protein structure and interaction propensity, and that these fundamental features are kept in the mesoscale description.
Overall, from this preliminary investigation, two results emerge a) the *in nuce* potential of the multiscale approach to describe large systems, with the possibility of including flow motion effects in the protein dynamics; b) the needing for a more refined coupling scheme at the protein/solvent interface, possibly a scheme where the solvent is explicitly described only close to the protein interface, providing a reliable description of the interaction among water molecules and amino acids, a pivotal interaction in determining both equilibrium and dynamic properties of biological molecules.

4 Conclusions

To conclude, on one side our simulations demonstrate and quantitatively estimate the pivotal role of the explicit water molecules treatment to obtain a statistically reliable characterization of biological molecules. In this respect the use of the LB solvent, while computationally advantageous, does not deliver quantitatively accurate information. On the other hand, our study highlights that many structural features, important in biological activities, are preserved in the LB/MD mesoscale solvent description, as shown for instance by the CC map and the RMSFs. The RMSFs, despite being quantitatively different, show similar qualitative behaviour when comparing Spike and α-Spike. In particular, we observe an increased rigidity of the RBD in the α-Spike compared to the wild type Spike. The main correlations inside the sub-domains are preserved, and the large correlation of the RBD block in the CC map for the α-Spike reinforces the information on the stability of the mutated structure. Indeed, the capability of the α-Spike RBD to maintain a rigid structure can be related, from the statistical point of view, with its remarkable human-to-human transmissibility [17], since the smaller fluctuations are limiting the sampling of possible RBD configuration basins to the subregion where the structure more efficiently binds to ACE2. This behaviour is somehow similar to what is observed for RBD rigidity of SARS-CoV-2 with respect to the Sars virus [17]. Moreover, the same correlation trends over Spike and α-Spike structures are found in the LB/MD simulations, showing the capability of the multiscale approach to preserve essential information even if at a lower description level of the solvent.

Future work should be directed to the development of a new class of LB models capable of supporting larger fluctuations than presently possible. A promising direction along this line is the resort to higher-order sets of discrete velocities.

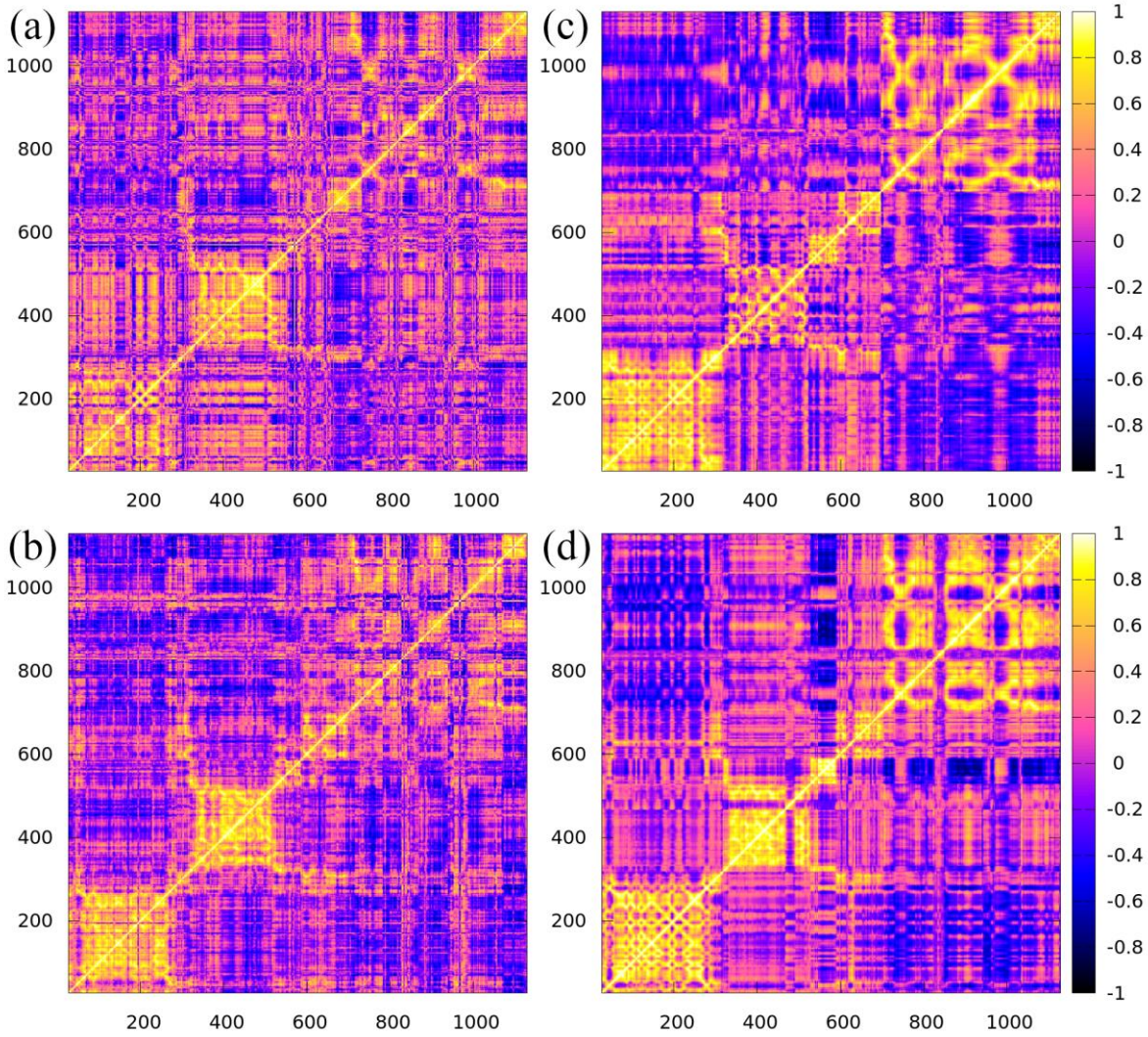


Fig. **6**: Cross-correlation map of the cross-correlation matrix for chain A of Spike and α-Spike in LB/MD simulations (panels a, b), and in AA/MD simulation (panels c, d). All cases show a significant correlation block associated with residues from the RBD (residues 331-528). The higher mobility of the protein in the all-atom simulations is reflected by the richer structure of the CCM (panels c,d), yet the qualitative structure of the patterns is preserved in LB/MD.

Acknowledgements

M. L., F. B., M. D., A. M., A. T. and S. S. acknowledge funding from the European Research Council under the European Union's Horizon 2020 Framework Programme (No. FP/2014-2020) ERC Grant Agreement No.739964 (COPMAT). L. C., A. L. and S. F. acknowledge the support of the International Center for Relativistic Astrophysics Network (ICRANet) and of the Italian National Group for Mathematical Physics (GNFM-INdAM). All the Authors gratefully acknowledge ENEA for the availability of high-performance computing resources and support on the HPC CRESCO facility used in the LB/MD simulations. Also, we acknowledge CINECA Project ABRISP under the ISCRA initiative, for the availability of high-performance computing resources and support used in the all-atom MD simulations.

References

[1] J. D. Rochaix. (2019). Dynamic Modeling of a 100-Million-Atom Organelle at the Source of Life. Cell, 179 (5), 1012-1014.

[2] X. Song, M. Ø. Jensen, V. Jogini, et al. (2018). Mechanism of NMDA receptor channel block by MK-801 and memantine. Nature, 556(7702), 515-519.

[3] https://www.exscalate4cov.eu/eu

[4] E. Taka, S. Z. Yilmaz, M. Golcuk, C. Kilinc, U. Aktas, A. Yildiz and M. Gur. (2020). Critical Interactions Between the SARS-CoV-2 Spike Glycoprotein and the Human ACE2 Receptor. bioRxiv.

[5] J. He, H. Tao, Y. Yan, S.-Y. Huang and Y. Xiao. (2020). Molecular Mechanism of Evolution and Human Infection with SARS-CoV-2. Viruses, 12, 428.

[6] V. Armijos-Jaramillo, J. Yeager, C. Muslin and Y. Perez-Castillo. (2020). SARS-CoV-2, an evolutionary perspective of interaction with human ACE2 reveals undiscovered amino acids necessary for complex stability. Evol. Appl., 13, 2168– 2178.

[7] J. Zou, J. Yin, L. Fang, M. Yang, T. Wang, W. Wu, M. A. Bellucci and P. Zhang. (2020). Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Relative Free Energy Calculations. J. Chem. Inf. Model. 60, 5794– 5802.

[8] E. S. Brielle, D. Schneidman-Duhovny and M. Linial. (2020). The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. Viruses, 12, 497.

[9] A. Ali and R. Vijayan. (2020). Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. Scientific reports, 10(1), 1-12.

[10] W. Zheng, H. Wen, G. J. Iacobucci and G. K. Popescu. (2017). Probing the structural dynamics of the NMDA receptor activation by coarse-grained modeling. Biophys. J., 112(12), 2589-2601.

[11] B. Lev, S. Murail, F. Poitevin, B. A. Cromer, M. Baaden, M. Delarue and T. W. Allen. (2017). String method solution of the gating pathways for a pentameric ligand-gated ion channel. PNAS, 114(21), pp. E4158-E4167.

[12] P. V. Coveney, A. Hoekstra, B. Rodriguez and M. Viceconti. (2021). Computational biomedicine. Part II: organs and systems.

[13] W. Edeling, H. Arabnejad, R. Sinclair, et al. (2021). The impact of uncertainty on predictions of the CovidSim epidemiological code. Nature Computational Science, 1(2), pp.128-135.

[14] M. Bernaschi, S. Melchionna and S. Succi. (2019). Mesoscopic simulations at the physics-chemistry-biology interface. Reviews of Modern Physics, 91(2), p.025004.

[15] Y. Cai, J. Zhang, T. Xiao, et al. (2020). Distinct conformational states of SARS-CoV-2 spike protein. Science, 369(6511), pp.1586-1592.

[16] B. J. Bosch, R. Van der Zee, C. A. De Haan and P. J. Rottier. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. Journal of virology, 77(16), pp.8801-8811.

[17] A. Spinello, A. Saltalamacchia and A. Magistrato. (2020). Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations. J. Phys. Chem. Lett. 2020, 11, 12, 4785–4790

[18] M. C., Bellissent-Funel, A. Hassanali, M. Havenith, et al. (2016). Water determines the structure and dynamics of proteins. Chemical reviews, 116(13), 7673-7697.

[19] D. Wrapp, N. Wang, K. S. Corbett, et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science, 367(6483), 1260-1263.

[20] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, & Q. Zhou, (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science, 367(6485), 1444-1448.

[21] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, (2015). The I-TASSER Suite: protein structure and function prediction. Nature methods, 12(1), pp.7-8.

[22] A. Rambaut, N. Loman, O. Pybus, et al., (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations Virological. org

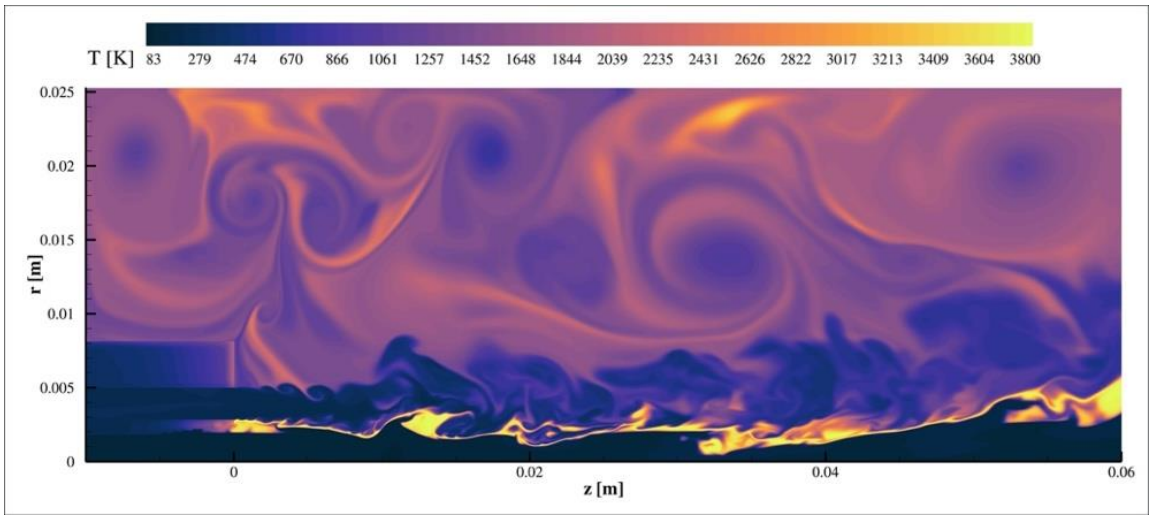
[23] V. Kovacova, K. Boršová, E.D. Paul, et al., (2021), A novel, room temperature-stable, multiplexed RT-qPCR assay to distinguish lineage B.1.1.7 from the remaining SARS-CoV-2 lineages, medRxiv, doi: https://doi.org/10.1101/2021.02.09.21251168

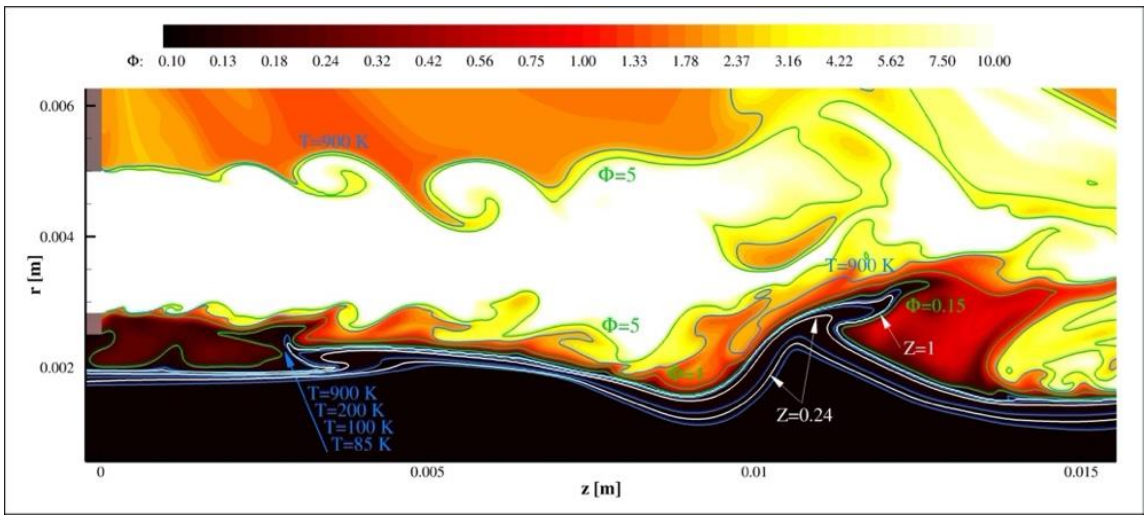
[24] G. Martínez-Rosell, T. Giorgino and G. De Fabritiis, (2017). PlayMolecule ProteinPrepare: A web application for protein preparation for molecular dynamics simulations. Journal of chemical information and modeling, 57(7), pp.1511-1516.

[25] M.H. Olsson, C.R. Søndergaard, M. Rostkowski and J.H., Jensen, (2011). PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. Journal of chemical theory and computation, 7(2), pp.525-537.
[26] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon and N.A. Baker, (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic acids research, 32(suppl_2), pp.W665-W667.
[27] W. Humphrey, A. Dalke and K. Schulten, (1996). VMD: visual molecular dynamics. Journal of molecular graphics, 14(1), pp.33-38.
[28] S. Plimpton, (1995). Fast parallel algorithms for short-range molecular dynamics. Journal of computational physics, 117(1), pp.1-19.
[29] U. Micka, C. Holm, and K. Kremer, (1999). Strongly charged, flexible polyelectrolytes in poor solvents: molecular dynamics simulations. Langmuir, 15(12), pp.4033-4044.
[30] J. Huang, and A.D. MacKerell Jr, (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. Journal of computational chemistry, 34(25), pp.2135-2145.
[31] S. Succi, (2018). The lattice Boltzmann equation: for complex states of flowing matter. Oxford University Press.
[32] F. Sterpone, P. Derreumaux, and S. Melchionna, (2015). Protein simulations in fluids: Coupling the OPEP coarse-grained force field with hydrodynamics. Journal of chemical theory and computation, 11(4), pp.1843-1853.
[33] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl, (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1, pp.19-25.
[34] P. J. Steinbach and B.R. Brooks, (1994). New spherical-cutoff methods for long-range forces in macromolecular simulation. Journal of computational chemistry, 15(7), pp.667-683.
[35] A.D. MacKerell Jr, D. Bashford, M.L.D.R. Bellott, et al., (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. The journal of physical chemistry B, 102(18), pp.3586-3616.

AUTHOR INDEX

ENEA
Promotion and Communication Service
www.enea.it
Printed at the ENEA Technographic Laboratory – Frascati
December 2021


T [K]
83 279 474 670 866 1061 1257 1452 1648 1844 2039 2235 2431 2626 2822 3017 3213 3409 3604 3800
0.025
0.02
0.015
0.01
0.005
0
r [m]
0
0.02
0.04
0.06
z [m]


Φ: 0.10 0.13 0.18 0.24 0.32 0.42 0.56 0.75 1.00 1.33 1.78 2.37 3.16 4.22 5.62 7.50 10.00
0.006
0.004
0.002
r [m]
T=900 K
Φ=5
T=900 K
Φ=0.15
Φ=5
Z=1
Φ=1
T=900 K
T=200 K
T=100 K
T=85 K
Z=0.24
0
0.005
0.01
0.015
z [m]