



E-Infrastructures
H2020-INFRAEDI-2018-1

INFRAEDI-2-2018: Centres of Excellence on HPC

EoCoE-II

Energy oriented Center of Excellence :
toward exascale for energy

Grant Agreement Number: INFRAEDI-824158

D1.2

First scientific results and progress on exascale tools

D1.2 First scientific results and progress on exascale tools

Project and Deliverable Information Sheet

EoCoE-II	Project Ref:	INFRAEDI-824158
	Project Title:	Energy oriented Centre of Excellence: towards exascale for energy
	Project Web Site:	http://www.eocoe2.eu
	Deliverable ID:	D1.2
	Deliverable Nature:	Report
	Dissemination Level:	PU*
	Contractual Date of Delivery:	M18 30/06/2020
	Actual Date of Delivery:	30/06/2020
	EC Project Officer:	Evangelia Markidou

* - The dissemination level are indicated as follows: PU – Public, CO – Confidential, only for members of the consortium (including the Commission Services) CL – Classified, as referred to in Commission Decision 2991/844/EC.

Document Control Sheet

Document	Title :	First scientific results and progress on exascale tools
	ID :	D1.2
	Available at:	http://www.eocoe2.eu
	Software tool:	L ^A T _E X
Authorship	Written by:	Edoardo Di Napoli (FZJ)
	Contributors:	Sebastian Achilles (FZJ), Frédéric Blondel (IFPEN), Pauline Bozonnet (IFPEN), Johanna Bruckmann (RWTH), Francesco Buonocore (ENEA), Marie Cathelain (IFPEN), Matthias De Lozzo (IRT - Associated to CERFACS), Siham El Garroussi (CERFACS), Andrea Galletti (UNITN), Garrett Good (IEE), Yen Sen Lu (FZJ), Bruno Majone (UNITN), Bibi S. Naz (FZJ), Herbert Owen (BSC), Alessandro Pecchia (CNR), Andrea Piacentini (CERFACS), Kyle Reeves (CEA-MdIS), Sophie Ricci (CERFACS), Ulrich Rüde (FAU), Michele Ruggeri (CEA-MdIS), Mathieu Salanne (CEA-MdIS), Yanick Sarazin (CEA-IRFM), Helen Schottenhamml (FAU, IFPEN), Ani Anciaux-Sedrakian (IFPEN), Eric Sonnendruecker (MPG-IPP), Matthew Wolf (UBAH).
	Reviewed by:	Julien Bigot, Bruno Raffin

Document Keywords: Computational Fluid Dynamics, Large Eddy Simulation, Complex Terrain, Full Rotor, Density functional theory, Molecular dynamics, Quantum Transport, Photovoltaics, Quantum Monte-Carlo, Capacitance, Seebeck coefficient, Thermo-electrochemical devices, Multi-scale Modelling, Monte Carlo, TerrSysMp, Reservoir, Hydropower, Geothermal, ITER, Nuclear Fusion, Wind Turbine.

Contents

1	Executive summary	6
2	Acronyms	9
3	Wind for Energy (T1.1)	12
3.1	Task T1.1.1 - Scientific results	13
3.2	Task T1.1.2 - Scientific results	17
3.3	Code demonstrator	24
3.4	Roadmap revision	26
4	Meteorology for Energy (T1.2)	27
4.1	Task 1.2.1: Scientific results	28
4.2	Task 1.2.2: Scientific results	29
4.3	Task 1.2.3: Scientific results	32
4.4	Code demonstrator	34
4.5	Roadmap revision	36
5	Materials for Energy (T1.3)	37
5.1	Task 1.3.1: Scientific results	39
5.2	Task 1.3.2: Scientific results	50
5.3	Task 1.3.3: Scientific results	54
5.4	Code demonstrators	57
5.5	Roadmap revision	59
6	Water for Energy (T1.4)	61
6.1	Task 1.4.1: Scientific results	61
6.2	Task 1.4.2: Scientific results	67
6.3	Task 1.4.3: Scientific results	70
6.4	Tasks T1.4.4 & T1.4.5 (Geothermal Energy)	74
6.5	Code demonstrators	79
6.6	Roadmap revision	82
7	Fusion for Energy (T1.5)	85
7.1	T1.5.1: Scientific results	85
7.2	T1.5.2: Scientific results	87
7.3	T1.5.3: Scientific results	88
7.4	Code demonstrator	90
7.5	Roadmap revision	91
A	Annex A: Summary of delays and risk assessment	99
B	Annex B: Extra plots for subtasks T1.3.1-1 & T1.3.1-2	103

List of Figures

1	Word cloud of the Work Package 1.	12
2	Mean streamwise velocity.	14
3	velocity magnitude at two planes	15
4	ALEX17 elevation map with experimental layout and Z-shaped transect.	16

D1.2 First scientific results and progress on exascale tools

5	10-minute scan of the wind speed	16
6	Temporal evolution of forcing terms	17
7	Topography of the site modelled in Meso-NH	19
8	Instantaneous Q criterion isosurfaces for the NREL VI wind turbine	19
9	CAD model and shear webs	20
10	(a) Boundary conditions and modelled stations (b) Domain decomposition of the mesh.	20
11	Schematic of a wind turbine: view of a yawed turbine from the top (left) and side (right)	21
12	Schematic of a wind turbine blade element. The wind coming from the left.	22
13	Normal force distribution on the NewMexico wind turbine blade	23
14	Time-averaged velocity distribution upstream and downstream of the NewMexico wind turbine	23
15	Instantaneous velocity field on a 2D plane (NTNU test case)	24
16	Time-averaged velocity profiles in the wake of the second turbine	25
17	Exawind and Atmosphere to Electrons screenshot	27
18	Error of wind speed for PEARP and WRF	29
19	Calibration weight of a weather model to satellite irradiation data	31
20	Cloud motion	31
21	Ten-meter hourly wind speed score and cloud matching rate score	33
22	The of cloud fraction (0% (red) to 100% (blue) at 09:00 on 2015-08-23.	34
23	The schematic plot of grids analysis method that is use to proceed data analytics	35
24	Schematic plot of grids analysis, satellite data before and after re-sampling.	36
25	Snapshots of the a-Si:H/c-Si interfaces.	41
26	Projected density of states of the ab initio relaxed a-Si:H/c-Si interfaces	43
27	Projected density of states of c-Si for T=300 K.	43
28	Projected density of states of a-Si for T=300 K.	44
29	Local density of states of the intragap states of the a-Si:H/c-Si interfaces	45
30	Energy potential averaged over the xy-plane	46
31	The c-Si/a-Si:H interface and the difference of the charge density	47
32	Transmission plot for the c-Si/a-Si:H/c-Si interface at T=0	49
33	Transmission plot for the c-Si/a-Si:H/c-Si interface from T=300 to 900 K	49
34	DFT calculations for the adsorption of an acetonitrile molecule on graphite	51
35	Fitted forces and dipoles with respect to the DFT values.	52
36	Comparison of the simulated and experimental infrared spectra for acetonitrile.	53
37	vertical energy gaps machinelearning	54
38	KMC Simulations: Mobility	55
39	Mobilities in metal-halide perovskites	56
40	Parallel scaling of QMCPACK on GPUs and CPUs.	59
41	Model surface input data	62
42	Validation of the CLM simulated hydrological fluxes and states	63
43	Comparison of SM simulated by ParFlow, CLM with in-situ observations	64
44	Comparison of the spatially aggregated total water storage(TWS)	64
45	Schematic of CLM-PDAF workflow adopted to generate high resolution ESSMRA product. Adopted from [55].	65
46	Comparison of daily time series of soil water content	66
47	Schematic of the HYPERstreamHS modelling framework and workflow.	68
48	Conceptual model of the Adige river basin.	69
49	Sobol indices for 1D model MASCARET over the Garonne river.	73
50	Root mean square error and predictive coefficient for the POD+GP surrogate.	73
51	First and total Sobol' indices estimated with POD-GP surrogate	74
52	High-level EFCOSS architecture [78].	76
53	OED model	80

54	a) Single node performance comparison., b) Weak scaling comparison.	81
55	Embedded boundary, block-structured polar mesh, cartesian and polar grids.	86
56	POLOIDALLY-SYMMETRIC case	89
57	LIMITER case	89
58	Scanning of the GYSELA profiles	90
59	Strong scaling for GYSELAX on Irene-AMD supercomputer (TGCC).	91
60	Snapshots of the a-Si:H/c-Si interfaces at a) T=300 K; b) T=500 K; c) T=700 K; and d) T=900 K.	103
61	Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 500 K.	104
62	Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 700 K.	104
63	Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 900 K.	105
64	Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 500 K.	105
65	Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 700 K.	106
66	Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 900 K.	106
67	Local density of states of the intragap states of the a-Si:H/c-Si interfaces at T= 500 K.	107
68	Local density of states of the intragap states of the a-Si:H/c-Si interfaces at T= 700 K.	107
69	Local density of states of the intragap states of the a-Si:H/c-Si interfaces at T= 900 K.	108

List of Tables

1	Acronyms for the partners and institutes therein.	9
2	Acronyms of software packages	9
3	Acronyms for the Scientific Terms used in the report.	10
5	Wind turbine components with their respective degree of freedom	21
6	Employed Physical schemes setting of ESIAS-met	29
7	Common physical schemes setting of ESIAS-met	29
19	Current status of ongoing physics developments	90
20	Summary of delays and risk assesment listed by task.	99

1 Executive summary

This report describes the first scientific results of the Work Package 1 (WP1) and the progress of the corresponding flagship codes towards the goal of exascale computing. The scope of the EoCoE-II project is to build on its unique, established role at the crossroads of HPC and renewable energy to accelerate the adoption of production, storage and distribution of clean electricity. We realize such a target by developing a sustainable structure able to develop state-of-the-art numerical tools and promote the usage of HPC in the energy domain. At the core of this structure there are five Energy Science Challenges (SC) addressing the most important and diverse area in the energy domain. The scientific payload of these challenges is the target of the WP1. Each challenge is a task labeled by one letter (**T** for task) and two numbers, the first indicating the work package and the second specific to the challenge.

Label	Energy Science Challenge
T1.1	Wind for Energy
T1.2	Meteorology for Energy
T1.3	Materials for Energy
T1.4	Water for Energy
T1.5	Fusion for Energy

Each of these challenges focuses on a selected number of main tasks whose payload provides significant advances in its respective domain. These advances will trigger innovation in fields such as energy production by wind turbines, photovoltaic cells, and hydroelectric and geothermal stations; prediction of weather forecast and cloud cover for the better placement of renewable power stations; simulations and validation of water resources at the surface and subsurface level for better water management; up until future energy production from nuclear fusion. In addition, advances in these energy domains prove the benefit of exascale computing for the growing low-carbon energy community as well as pave the way to Exascale for other related energy sectors such as energy transportation and storage.

Wind. The scientific challenge “Wind for Energy” deals with two main objectives. First, we aim at improving the accurate simulation of large scale wind farms by including wall modeling and inflow boundary conditions. The second objective is to model and simulate accurately the entire wind turbine including the tower, full rotor and the rotation of the blades. Both goals require to expand the parallelism of the flagship code Alya, so to facilitate its scaling up to the largest computing cluster available. One of the main reasons behind this necessity is the large size of the wind farm domain we want to simulate (of the order of hundreds of square kilometers). Similarly, modeling the full turbine requires adaptive and refined meshes which transform the main equations describing the system into very large and sparse linear systems. In turns, the solution of such systems greatly benefit from scaling linear system solvers.

Meteo. In the “Meteorology for Energy” challenge, we address the issue of large-scale ensemble weather predictions aimed at improving the forecast of energy production due to renewable power plants. This challenge is divided into three main tasks. In the first, we develop ensemble data and additional software functionalities within the ESIAS hyper-ensemble system targeting solar irradiation and the wind speed at the rotor hub level. In the second task, we applied satellite modeling to the ESIAS ensemble data using the Solar Prediction (SPS) and Wind Power Management (WPMS) Systems. By including cloud-tracking, we are able to create continuous probabilistic predictions of the electric feed-in from the wind and solar ensembles. In the final task, we calibrate the ensembles to improve the wind and cloud forecast. The need for large scale computations in this challenge is a consequence of the sheer size of the ensemble we want to consider which is at least one order of magnitude larger than any other ensemble generated so far. By using such large ensembles, we plan to forecast with unprecedented accuracy extreme weather events.

D1.2 First scientific results and progress on exascale tools

Materials. The challenges of “Materials for Energy” focuses on researching novel materials for the production of electricity through photovoltaic devices and thermo-electrochemical devices applied to salinity gradients. The search of new photovoltaic materials probes into two different directions. In the first, we set out to understand the transport mechanisms underlying photovoltaic devices based on silicon hetero-junction technology. The challenge here is to understand the microscopic transport and recombination processes at the interface between passivated silicon and amorphous silicon. The second direction centers on optoelectronic devices based on two classes of emerging photovoltaic materials, namely organic molecular semiconductors and metal–halide perovskite semiconductors. Differences in the two materials necessitate different computational approaches, and therefore this task involves the development, implementation and application of simulation methodologies which are amenable to studying phenomena at a broad range of length and time scales.

Extracting electricity from salinity gradients aims at providing computational support for optimizing capacitive blue energy electrodes and thermo-electrochemical devices. Electric power production from salinity gradients harvests the free energy lost during the mixing of river with sea water in estuaries. The main technologies developed for this purpose to date exploit the electric potential differences applied across membranes. We analyze an alternative approach based on capacitive mixing. As part of this approach, we are investigating the electrode structure which optimizes a blue energy production and will enable optimal performances of thermo-electrochemical cells.

Water. The “Water for Energy” challenges tackles a number of diverse issues dealing with water management and energy production from surface water usage. This challenge is divided in five main tasks, each with its own objective both in terms of scientific payload and code scaling. The first task addresses the modeling of shallow subsurface water flow, which is of major importance in order to accurately simulate and predict the exchange of ground water with streams and rivers. The major challenge is the representation of the geological heterogeneity which drives ground water flow generation across a number of space scales ranging from centimeters to thousands of kilometers. In this context, the prediction of hydrologic states and fluxes requires a balanced combination of computational power to run climatic and hydrological models at the proper scale and enough information concerning the hydropower systems. This is extremely challenging and requires HPC technologies to enable hyper-resolution simulations at continental scales. In the second task, we aim at simulating hydropower generation in a meso-scale Alpine watershed. This is achieved by coupling highly detailed information on main hydropower systems that are present in the Italian Alpine region and runoff time series produced by hyper-resolved hydrological model as provided by the first task. Our hydrological model addresses the development of a large-scale model accounting explicitly for spatial and temporal variability of water usage and explicitly simulating water storage and diversions within a river basin. Similarly, the third task deals with hydrodynamics modeling for flood forecasting and water resources management. Uncertainties in river bed friction and inflow plays a key role in determining the uncertainties in simulated water level and discharge. These uncertainties are quantified with a non intrusive sensitivity analysis method based on statistics computed from an ensemble of model trajectories considering stochastic inputs sampled within prescribed probability density functions. We achieve this target by using a classical Monte Carlo approach adapted to massively parallel computational resources. To avoid that the cost of the ensemble integration gets prohibitive for large catchments, we use a surrogate model in place of the direct solver.

The last two tasks of the water challenge deal with the experimental design for predicting borehole locations in geothermal reservoirs in order to decrease the uncertainty in parameter estimation. Drilling boreholes during exploration and development of geothermal reservoirs not only involves high cost, but also bears significant risks of failure. In geothermal reservoir engineering, techniques of optimal experimental design (OED) have the potential to improve the decision making process, and minimize the uncertainty of the parameter estimation. In turns, this results in a more reliable parameterization of the reservoir simulation, improving the overall process in geothermal reservoir engineering. We tackle the problems arising from geothermal modeling by integrating the EFCOSS mathematical optimization software with the SHEMAT-

D1.2 First scientific results and progress on exascale tools

Suite, our geothermal simulation code for fluid flow and heat transport through porous media.

Fusion. The last scientific challenge, “Fusion for Energy”, addresses the future of renewable energy production. This challenge aims overall at building a first principle flagship code capable of addressing electromagnetic turbulence from the core to the far edge region in the magnetic geometries relevant to the ITER nuclear reactor. A breakthrough in this area will open the door to ground-breaking physics advances. To this end, we are developing the new release of the code GYSELAX , which capitalizes on and considerably extends the capabilities of the 5D flux-driven gyrokinetic Fortran parallel code GYSELA . The latter is capable of solving Vlasov and Poisson equations to simulate electrostatic plasma turbulence and transport in core Tokamak plasmas. In particular, we are actively working on improving the parallelization efficiency on the new code so as to be able to ultimately target exascale supercomputers.

Structure of the report and code demonstrators

Each scientific challenge has its own section. Scientific results for each of the main tasks of a scientific challenge are illustrated in separate subsections labeled with the name of the main task (e.g. T1.2.1). Each subsection is additionally split in parts corresponding to the subtask labeled by their relative subtask (e.g. T1.2.1-3). Last two subsections of a given challenge deal with a description of the code demonstrators and a revised roadmap and timeline.

Code Demonstrator. The description within the code demonstrator section refers to a snapshot of the current status of the flagship codes (and satellite codes, whenever possible) at the time of this report. This description includes an example of scientific simulation and relative results. For each code description, we have provided the relative software in homonymous folders within a protected Gitlab repository accessible only by request.

```
https://gitlab.maisondelasimulation.fr/eocoe-ii/code-demonstrators.git
```

The repository is not publicly available because each code demonstrator contains a snapshot of the code which may differ from the one available from a public repository (if one such repository exists) and the scientific data corresponding to the results described. In many instances, such data is not yet final or available to the public because not yet published and so needs to be protected until it can be made available. In the Gitlab repository, each code has its own folder with the source files and data available and a small Readme file with an explanatory text.

Roadmap revision. The roadmap revision describes changes in the timeline and roadmap with respect to the D1.1 deliverable report. Changes in the timeline are marked in pink color. Motivations, impact and mitigation are described in the text. A complementary table with all delays, corresponding risk assessments, and mitigation plans is provided for all tasks and partners in Annex A at the end of this report.

2 Acronyms

Table 1: Acronyms for the partners and institutes therein.

Acronym	Partner and institute
AMU:	Aix-Marseille University
BSC:	Barcelona Supercomputing Center
CEA:	Commissariat à l'énergie atomique et aux énergies alternatives
CERFACS:	Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique
CIEMAT:	Centro De Investigaciones Energeticas, Medioambientales Y Tecnologicas
CoE:	Center of Excellence
CNR:	Consiglio Nazionale delle Ricerche
EDF:	Électricité de France
ENEA:	Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile
FAU:	Friedrich-Alexander University of Erlangen-Nuremberg
FSU:	Friedrich Schiller University
FZJ:	Forschungszentrum Jülich GmbH
IBG-3:	Institute of Bio- and Geosciences Agrosphere
IEK-8:	Institute for Energy and Climate Research 8 (troposphere)
IEE:	Fraunhofer Institute for Energy Economics and Energy System Technology
IFPEN:	IFP Énergies nouvelles
INAC:	Institut nanosciences et cryogénie
INRIA:	Institut national de recherche en informatique et en automatique
IRFM:	Institute for Magnetic Fusion Research
ISMN:	Istituto per lo Studio dei Materiali Nanostrutturati
MdIS:	Maison de la Simulation
MF:	Meteo France
MPG:	Max-Planck-Gesellschaft
RWTH:	Rheinisch-Westfälische Technische Hochschule Aachen, Aachen University
UBAH:	University of Bath
UNITN:	University of Trento

Table 2: Acronyms of software packages

Acronym	Software and codes
CLM:	Community Land Model
EFCOSS:	Environment For Combining Optimization and Simulation Software
ESIAS:	Ensemble for Stochastic Interpolation of Atmospheric Simulations
EURAD-IM:	EUROpean Air pollution Dispersion-Inverse Model
HYPERstreamHS:	Dual-layer MPI large scale hydrological model including Human Systems
ICON:	Icosahedral Nonhydrostatic model
MDFT:	Molecular Density Functional Theory
MELISSA:	Modular External Library for In Situ Statistical Analysis
Meso-NH:	Mesoscale Non-Hydrostatic model
Nemo5:	NanoElectronics MOdeling Tools 5
neXGf:	non-equilibrium eXascale Green's functions library
OpenFOAM:	Open Source Field Operation and Manipulation
ParFlow:	PARallel Flow
PF-CLM:	Parflow-Community Land Model

D1.2 First scientific results and progress on exascale tools

PPMD:	Performance Portable Molecular Dynamics
ReaxFF:	Reactive Force Field
SHEMAT:	Simulator of HEat and MAss Transport
SOWFA:	Simulator fOr Wind Farm Application
SPS:	Solar Prediction System
TELEMAC:	TELEMAC-MASCARET system
TerrSysMP:	Terrestrial Systems Modeling Platform
WaLBerla:	A Widely Applicable Lattice-Boltzmann Solver
WanT:	Wannier Transport
WPMS:	Wind Power Management System
WRF:	Weather Research and Forecast model

Table 3: Acronyms for the Scientific Terms used in the report.

Acronym	Scientific Nomenclature
2D:	Two-Dimensional
3D:	Three-Dimensional
ABL:	Atmospheric Boundary Layer
AD:	Automatic Differentiation
ALEX17:	Alaiz Experiment 2017
AOT:	Aerosol Optical Thickness
PBE:	Perdew-Burke-Ernzerhof functional
BLYP:	Becke-Lee-Yang-Parr functional
COT:	Cloud Optical Thickness
COVID-19:	Corona Virus Disease 2019
COSMO-REA6:	Convective-scale Regional Reanalysis System 6 Km
CPU:	Central Processing Units
CSP:	Concentrated Solar Power
CUDA:	Compute Unified Device Architecture
DA:	Data Assimilation
DEM:	Digital Elevation Model
DFT:	Density Functional Theory
DMC:	Dynamic Monte Carlo
EnKF:	Ensemble Kalman Filter
E-RUN:	European Runoff
ESACCI:	European Space Agency Climate Change Initiative
ESSMRA:	European Surface Soil Moisture reanalysis
ET:	Evapotranspiration
FSI:	Fluid-Structure Interaction
FLUXNET:	Flux Network
GMTED2010:	Global Multi-resolution Terrain Elevation Data 2010
GPU:	Graphical Processing Unit
HLST:	High Level Support Team
HPC:	High Performance Computing
IEA-Wind:	International Energy Agency - Wind
ITER:	International Thermonuclear Experimental Reactor
ISMN:	International Soil Moisture Network
KMC:	Kinetic Monte Carlo
LES:	Large Eddy Simulations
LiDAR:	Light Detection and Ranging

D1.2 First scientific results and progress on exascale tools

MD:	Molecular Dynamics
MODIS:	Moderate Resolution Imaging Spectroradiometer
MPI:	Message Passing Interface
NEGF:	Non-Equilibrium Greens functions
NEWA:	New European Wind Atlas
NREL:	National Renewable Energy Laboratory
NWP:	Numerical Weather Prediction
OED:	Optimal Experimental Design
OpenMP:	Open Multi-Processing
PBC:	Periodic Boundary Conditions
PDAF:	Parallel Data Assimilation Framework
pdf:	probability density functions
PDOS:	Projected Density of States
PF-CLM:	Parflow-Community Land Model
PRUDENCE:	Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects
QMC:	Quantum Monte Carlo
QM:	Quantum Mechanics
SHJ:	Silicon HeteroJunction
SOL:	Scrape-Off Layer
WP:	Work Package



Figure 1: Word cloud of the Work Package 1.

3 Wind for Energy (T1.1)

T1.1 is divided into two main tasks:

T1.1.1	Development, Verification, and Validation for Complex terrain
T1.1.2	Development, Verification, and Validation for Full rotor

The goal of the first task is to develop wind resource assessment capabilities using the low dissipation Large Eddy Simulation (LES) formulation for Atmospheric Boundary Layer (ABL) flows in the code Alya. Improvements in wall modeling and inflow boundary conditions will be performed to enable accurate wind farm simulations. The model will include thermal coupling, Coriolis forces, canopy, and an actuator disc model. Benchmarking against the Meso-NH solver will be performed. The model will be tested on real wind farms as part of an ongoing collaboration with Iberdrola. Within the second task, a full rotor model where the actual geometry of the wind turbine blades and tower is modeled exactly is developed. A sliding mesh approach is used to incorporate the rotation of the blades. Their deformation is considered using a Fluid-Structure Interaction (FSI) approach. The model will be compared against the full rotor model available in the code FLOWer. A comparison between the actuator line model developed by IFPEN and full rotor simulations will also be performed.

It can be noted that Meso-NH substitutes SOWFA which was the code originally given in the proposal for Task 1.1 at IFPEN. There are several reasons that have motivated this choice. First, although Meso-NH is a LES code like SOWFA, it is more advanced from a meteorological point of view. Meso-NH can model more

D1.2 First scientific results and progress on exascale tools

thermo-dynamical phenomena such as radiation, deep and shallow convection. It embarks an advanced physical parameterizations for cloud and precipitation representation. It can be coupled with different modules for chemistry (aerosol...) or complex surface (vegetation, cities, ocean...) for instance. Then, Meso-NH is more advanced in term of HPC (good scalability, vectorization) and is actively supported. The last argument to use Meso-NH is the size of the benchmarks. Simulated domains for EoCoE have a size of 40 km by 40 km much higher than the size usually considered in usual SOWFA simulation at IFPEN (i.e. flat terrain over a domain of 5 km by 5 km).

3.1 Task T1.1.1 - Scientific results

The main task, T1.1.1, is subdivided into five subtasks.

T1.1.1-1	Improve wall modeling for Atmospheric Boundary Layer of Large Eddy Simulation in the code Alya developed at BSC.
T1.1.1-2	Improve the inflow boundary conditions for the Atmospheric Boundary Layer of Large Eddy Simulation for the Alya code.
T1.1.1-3	Including thermal coupling, Coriolis forces, canopy, and the actuator disc.
T1.1.1-4	Benchmarking against the Meso-NH code.
T1.1.1-5	Test the new model against a realistic set up.

We report on the first three tasks lead by the BSC partner and the T1.1.1-4 which is coordinated by the IFPEN team. These are the tasks that have started according to the timetable submitted in deliverable D1.1. In addition, we would like to mention that we have found an LES case of interest for Iberdrola (T1.1.1-5), and we have already started performing some simulations.

T.1.1.1-1 - Improve wall modeling for Atmospheric Boundary Layer of Large Eddy Simulation in the code Alya developed at BSC

The implementation of wall modeling usually used by the finite element community has been improved to increase its accuracy for high Reynolds numbers typical of ABL flows. We have presented a paper [60] that provides a detailed description of how the approach used by the finite volume and finite differences communities can be adapted to the finite element codes. The new implementation is as simple and easy to implement as the classical finite element one, but it provides vastly superior results. In the typical approach used in finite elements, the mesh does not extend all the way to the wall, and the wall stress is evaluated at the first grid point, based on the velocity at the same point. Instead, we adopt the approach commonly used in finite differences, where the mesh covers the whole domain, and the wall stress is obtained at the wall grid point, with the velocity evaluated at the first grid point off the wall. The method is tested in a turbulent channel flow at $Re_\tau = 2003$, a neutral atmospheric boundary layer flow, and flow over a wall-mounted hump, with significant improvement in the results compared to the standard finite element approach. Additionally, we examine the effect of evaluating the input velocity further away from the wall, as well as applying temporal filtering on the wall-model input.

Results for the mean streamwise velocity of the turbulent channel flow case are presented in Figure 2. U^+ and y^+ are the non dimensional mean steamwise velocity and the non dimensional distance to the wall respectively. EL stands for Exchange Location, the name given to the methodology proposed in [60]. It can be observed that the new methodology significantly outperforms the classical finite element approach. Moreover, a new methodology that provides increased robustness for complex geometries has been implemented. This methodology modifies how the wall law is applied to avoid problems that can appear at sharp edges. The difficulties are associated with particularities of the finite element method for the application of no penetration in the normal direction and slip conditions in the tangential direction. The new approach eliminates the problem by prescribing zero velocity at the wall and modifying the viscosity

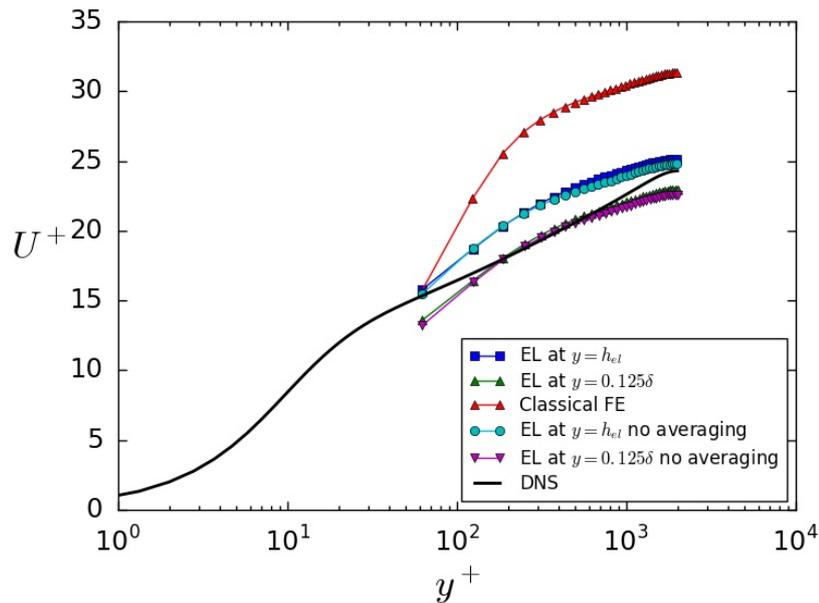


Figure 2: Mean streamwise velocity. DNS, reference direct numerical simulation; FE, finite element

in the first element so that the traction calculated from the wall law is obtained at the wall. Several tests have already been performed with this methodology with satisfactory results. When starting to write our latest paper manuscript, some subtleties have been discovered which are currently being corrected. Slight improvements in the results are being obtained, and we hope we can finish the publication shortly.

T.1.1.1-2 - Improve the inflow boundary conditions for the Atmospheric Boundary Layer of Large Eddy Simulation for the Alya code

Two different turbulent inlet approaches for the LES runs have been tested. In the first one, a turbulent inlet condition that matches the field data is synthesized using the digital filtering technique described in [5]. In the second one, a precursor type inlet is generated by coupling a periodic flat domain to the domain of interest. Both approaches have shown satisfactory results. Contrary to the wall modeling task, where significant improvements were obtained compared to those available in the literature for this task, the methodology we currently have in Alya is deemed satisfactory but similar to other approaches available elsewhere.

We have also performed Large Eddy Simulations for the Hornamossen test case [34] using neutral conditions. For this problem, periodic boundary conditions have been used. The terrain in the region far from the wind farm was modified to enable periodicity. The results were presented at the High-Performance Innovation Conference [59]. Figure 3 shows the velocity magnitude at two planes passing through the center of the region under analysis. We are now switching out attention towards the Alaiz Experiment [1] where better data is available.

T.1.1.1-3 - Including thermal coupling, Coriolis forces, canopy, and the actuator disc.

The Alya code now includes thermal coupling, Coriolis forces, canopy, and the actuator disc in the thermal and Navier Stokes equations for the LES simulations. Initial testing on the GABLS benchmark problem showed numerical instabilities close to the upper and lower boundaries. Initially, we thought that it was related to the Coriolis forces. A new stabilization method was developed for the Coriolis forces and tested in Alya. However, it did not solve the problem. Upon further numerical experiments, it was discovered that the problem was coming from the Boussinesq thermal coupling that needs to be included when Coriolis

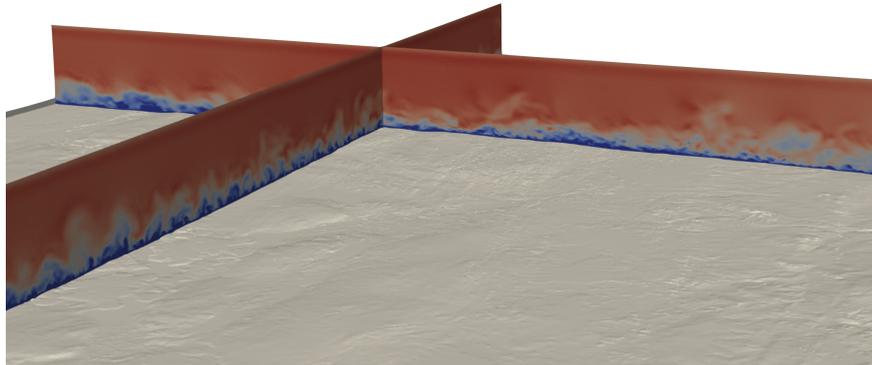


Figure 3: velocity magnitude at two planes passing through the center of the Hornamossen domain

forces are included for the problem to make sense. The problem is related to the interaction between the fractional step method we use to solve the incompressible Navier Stokes equations and the gravity forces introduced by the Boussinesq thermal coupling. We have found an initial solution that only works for flat terrain. It has been implemented in Alya and tested successfully. We believe that we have devised a more general solution that also works for complex terrain, but we still need to implement it and test it.

T.1.1.1-4 - Benchmarking against the Meso-NH code.

The ALEX17 benchmark, provided by NEWA and the IEA-Wind Task 31, will be used to compare Alya and Meso-NH in order to evaluate the ability to model complex flows over complex terrains.

The ALEX17 campaign is the last one of a series of experiments executed in the New European Wind Atlas (NEWA) project [4]. The main objective of NEWA is the development of a new European wind atlas and the improvement of advanced models towards a reduction of the cost of electricity generated by wind farms, and a better quantification of European wind energy potential [4]. To this end, a high-value data bank from a series of wind measurement campaigns has been developed and ALEX17 campaign is part of it. A benchmark has been launched by the IEA-Wind Task 31 Wakebench in order to characterize the wind conditions in the Alaiz mountains near Pamplona, Spain, and to compare the numerical modelling of participants to the field measurements. The IEA-Wind (International Energy Agency) is a collaborative program dedicated to the technological development and the global deployment of wind energy worldwide. The objective is to exchange information between member countries on the planning and execution of national large-scale wind system projects, and to undertake cooperative research and development projects called Tasks. The Task 31 Wakebench aims at developing verification, validation and uncertainty quantification frameworks to support a sustained improvement of wind farm models [3].

The ALEX17 benchmark is based on the Alaiz campaigns [74]. These experiments provide a vast range of atmospheric measurements within a mountain-valley area upstream from CENER wind turbine test site near Pamplona (Spain). The domain has dimensions of 20×20 km with an altitude of approximately 1100 m and elevation differences of up to 600 m. A large number of instruments were commissioned: five long-range WindScanner measuring the Z-shaped transect in Fig. 4 (WSx), six 80m-met masts (Mx), two 118m-met mast (MP1 and MP5) and a commercial Doppler wind LiDAR profiler (WLS70).

For this benchmark, four days have been selected (from September 30th 2018 to October 4th 2018) corresponding to inflow conditions from the north. The focus is on diurnal variability of mesoscale drivers and their impact on microscale wind conditions relevant for turbine siting. During this specific time period, particular events have been observed with recirculation zones and a windward flow blockage which can be observed in Fig. 5. The latter reports measurement on a vertical plane along the blue line in Fig. 4 (passing through WS3 and WS5) with 10-minute measurements of the velocity in the morning (top) and in the late afternoon (bottom). Most of the wind flow is directed to the south-southwest (positive values of the velocity) except downwind the first small mountain (recirculation zone) and upwind the larger one

D1.2 First scientific results and progress on exascale tools

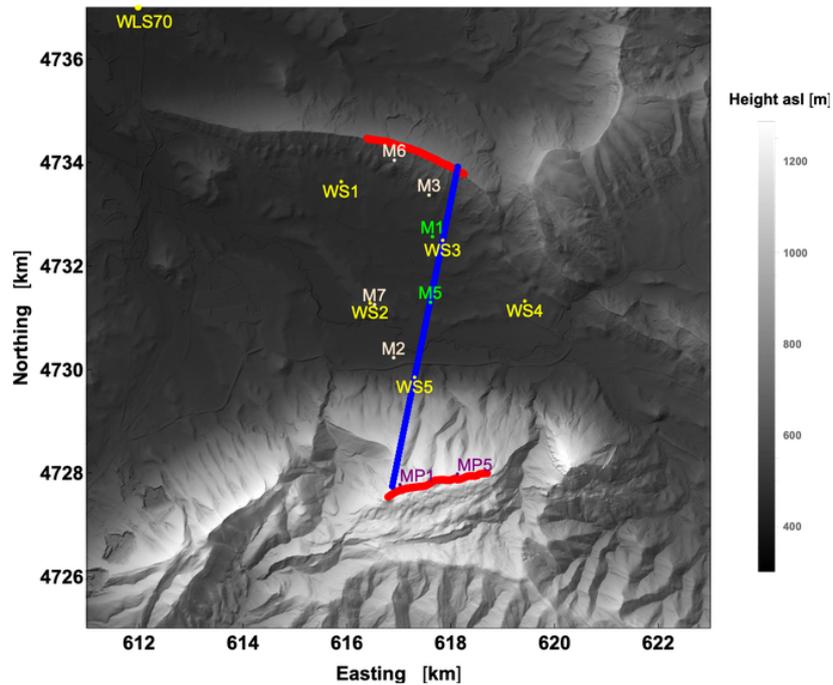


Figure 4: ALEX17 elevation map with experimental layout and Z-shaped transect.

(blockage zone).

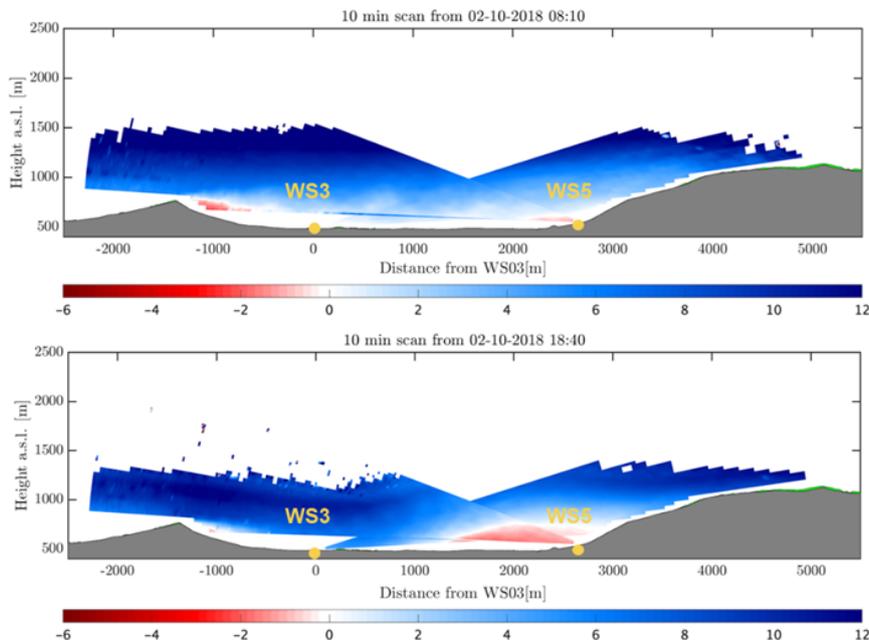


Figure 5: 10-minute scan (i.e. measurements) of the wind speed over the Z-shaped transect on October 2nd, 2018 at 8:10 AM and 6:40 PM.

This experiment has been modelled using the mesoscale numerical weather prediction model WRF on a regional domain with a resolution 27 km and two nested domains with resolutions of 9 and 3 km. The final area of interest is 45×45 km around the valley. The initialization and the forcing term of this large WRF simulation come from ERA5 reanalysis datasets. No recirculation and blockage zone have been observed

D1.2 First scientific results and progress on exascale tools

in this mesoscale WRF simulation hence the need to investigate meso-micro modelling.

A simulation with Meso-NH with a finer mesh resolution is set up with a geostrophic forcing in wind speed and potential temperature coming from the aforementioned WRF simulation. As the forcing terms from WRF are given at every 10 minutes during the 4-day period, it is necessary to proceed to an average of these terms. Fig. 6 shows the forcing terms such as the zonal and meridional geostrophic winds and the potential temperature at $z=100\text{m}$ above the ground level. A 2-hour average has been chosen as the red line matches quite well the 10-minute WRF data compared to the 4-hour average, especially the bump occurring before 02/10/18 (several hours before the wind recirculation observed in Fig. 5 top) which is totally missed with the 4-hour average (blue line). These 2-hour averaged forcing terms are thus implemented into the Meso-NH simulation.

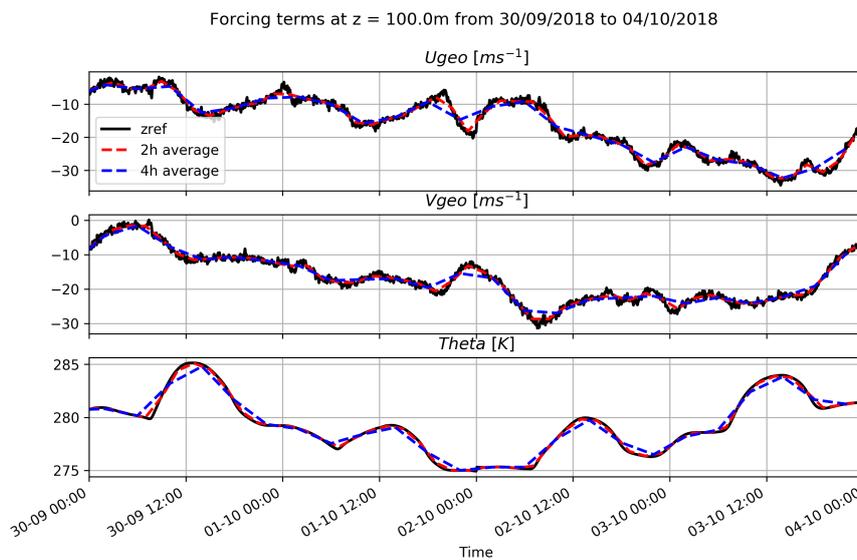


Figure 6: Temporal evolution of forcing terms (zonal and meridional geostrophic wind and potential temperature) from the WRF simulation (black line) with averages over 2 hours (red line) and 4 hours (blue line).

Concerning the Meso-NH setup, the physiographic data (describing the physical characteristics of the earth's surface) comes from the ECOCLIMAP-II database [31] and the terrain topography comes from the SRTM (Shuttle Radar Topography Mission) database [70]. Advanced meteorological features such as radiation or cloud microphysics are deactivated. The domain is 45 km by 45 km in order to have a distance sufficiently large between the northern boundary and the zone of interest to let the atmospheric boundary layer develop before reaching the zone of interest (Z-shaped transect in Fig. 7). Profilers have been included in the Meso-NH model (see black and red dots in Fig. 7). The study on the grid resolution sensitivity is still on-going.

3.2 Task T1.1.2 - Scientific results

The main task, T1.1.2, is subdivided into five subtasks.

T1.1.2-1	Develop and test the sliding mesh approach for rotating blades in Alya using rigid blades.
T1.1.2-2	Adapt shall elements to turbine blades.
T1.1.2-3	Merge the results of the previous two subtasks.
T1.1.2-4	Compare Alya with FLOWer.

T1.1.2-5 Comparing models.

We report on the first three subtasks that are the ones that have started according to the timetable submitted in deliverable D1.1.

T1.1.2-1 - Develop and test the sliding mesh approach for rotating blades in Alya using rigid blades

We have developed an implicit sliding mesh coupling for large scale industrial applications that allows us to simulate problems where one part of the mesh is fixed, and the other one is rotating. The wind turbine blades are enclosed in a cylinder that rotates inside the fixed mesh covering the whole wind farm. We have tested the methodology for an NREL Phase VI wind turbine, which is a relatively small, extensively tested turbine in the literature. Simulations have already been performed on a hybrid mesh formed by 50 million tetrahedral, pyramidal, and prismatic elements. We are currently running a case with twice smaller elements leading to a total of 400 million elements. In Figure 8 we show the Instantaneous Q criterion isosurfaces colored by velocity magnitude. For the moment we have not included the nacelle and the tower, mainly due to meshing difficulties which we expect to solve shortly. We have submitted a paper with the methodology to the PARCFD congress, where we will present a case with three wind turbines. The results presented at the congress will probably be later included in the Computers and Fluids Journal. Moreover, even though it was not included in the proposal, we would like to mention that we have started a collaboration with professor Riccardo Mereu from Politecnico di Milano, Department of Energy. His group has experience in the detailed simulation of the flow around wind turbine blades with the commercial code Fluent and the open-source code OpenFoam using Detached Eddy Simulation. We will compare his results with those obtained with Alya using Wall Modeled Large Eddy Simulation. We will start by repeating with Alya the simulations he has performed on the DU97-W-300 profile in clean configuration and equipped with vortex generators that improve the performance of the blade[54]. As part of this collaboration, Hugo Manuel Yauri Urbina, a master student from his group, has been awarded an HPC-Europa3 Programme travel grant to work with us at the Barcelona Supercomputing Center for three months.

T1.1.2-2 Adapt shell elements to turbine blades

The selected solid mechanics case is the WindPACT 1.5 MW horizontal-axis wind turbine blade (Malcolm and Hansen, 2002 [51]), which is a representative of megawatt-class horizontal-axis wind turbines. This is a 33m long wind turbine blade which includes information of the internal structural parts.

As the objective of this case is to perform a detailed structural simulation of the wind blade, the model is meshed using 3D linear brick elements. The main advantage of this element type is the real prediction of the stress state of the element (they take into account the out-of-plane stresses) while other element technologies such as conventional 4-node shells or beam elements do not. To the author's knowledge, the level of detail that we aim for this problem has not been performed by other people in the literature at a full-scale level either by the lack of computational resources or mesh complexity.

Due to the lack of geometry details and CAD models of real wind turbine blades in the literature, we have programmed a Python script to generate all the airfoils and internal parts using the available information in the literature. Fig. 9a and 9b show respectively the necessary points to define the surfaces of the wind turbine blade and also, the internal surfaces to define internal structural parts at each airfoil station. It is worth mentioning that this script serves as general tool to generate the CAD model of other blade designs including shear webs.

Before tackling the full blade model, we decided to first create a small model as a proof-of-concept model before meshing and working with the whole blade. The test model is shown in Fig 10a and the stations modelled are ranged from 9 to 11 (3.5m long). The material used is isotropic linear elastic with a Young Modulus of $E = 70GPa$ and a Poisson ratio of 0.33. The model is meshed with 8-node linear hexaedrons

D1.2 First scientific results and progress on exascale tools

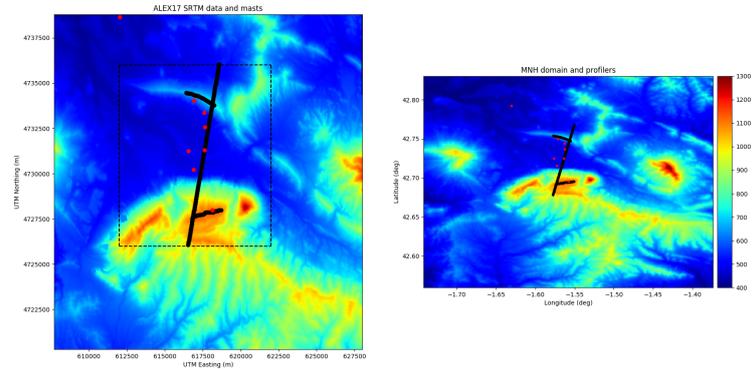


Figure 7: Topography of the site modelled in Meso-NH with implementation of met masts (red dots) and WindScanners (black dots).

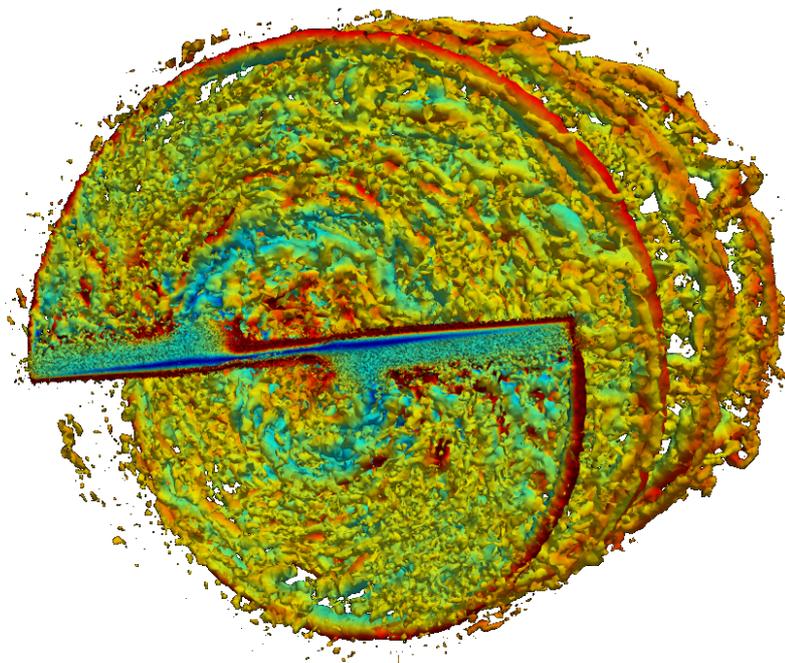


Figure 8: Instantaneous Q criterion isosurfaces coloured by velocity magnitude for the NREL VI wind turbine

and 6-node prisms resulting a total number of 6,768,684 DOF (degrees of freedom). For this test, a simple structural analysis is performed clamping the structure in one side and prescribing a displacement of 2.5cm in the other side. The simulation is conducted in Marenstrum IV [2] using 192 CPUs and the mesh domain decomposition is shown in Fig. 10b.

The model is solved using the Implicit Newmark-Beta scheme from [10] using $\beta = 0.65$, $\gamma = 0.9$ and $\alpha = 0$. In order to improve convergence of the solution a dynamic implicit analysis is used with a smooth step to reduce the inertial forces while applying the prescribed displacement. With respect to the algebraic solver, we use a GMRES solver with a RAS, BLOCK preconditioner, which is the one that performs better in terms of convergence of the solution among the available solvers in Alya. However, we have noticed that

D1.2 First scientific results and progress on exascale tools

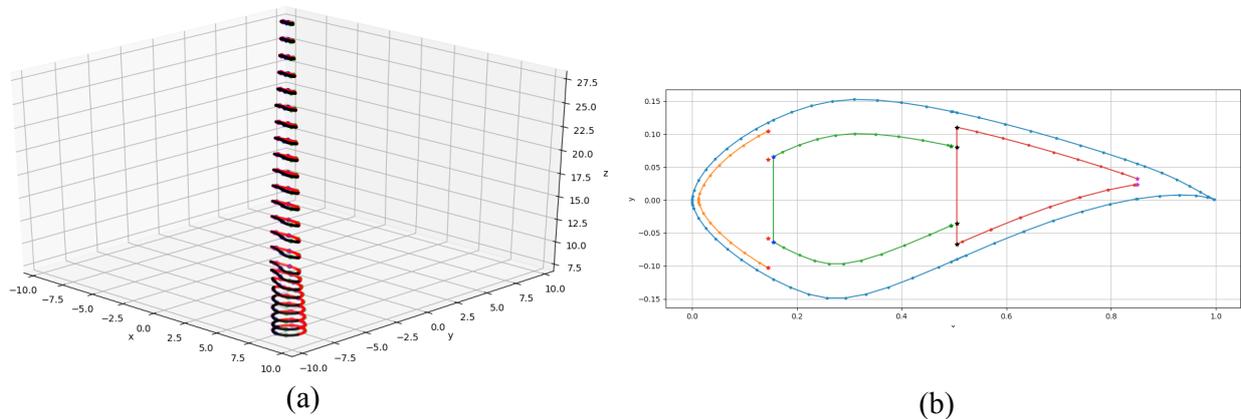


Figure 9: (a) Points for the generation of the CAD model. (b) Internal points for the definition of shear webs

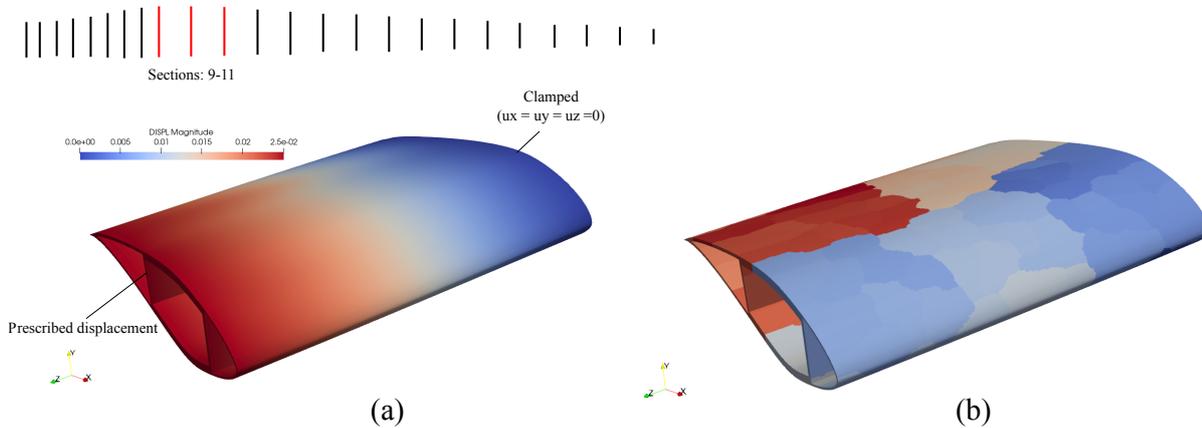


Figure 10: (a) Boundary conditions and modelled stations (b) Domain decomposition of the mesh.

the algebraic requires a huge number of iterations to solve the linear set of equations and it becomes the bottleneck in the total simulation time (99.49% of the time) in terms of computational performance. The average number of iterations for convergence of the solution is 800 for each N-R iteration. Thus, we are also exploring other solvers such as MUMPS [84] in order to figure out a more efficient and appropriate solver for this type of problems.

T1.1.2-3 Merge the results of the previous two subtasks

In this subtask, the outcome of the previous two subtasks will be used to simulate the rotating wind turbine blades and their deformation using the fluid-structure interaction (FSI) approach available in Alya. The fluid part will be treated with the sliding mesh approach from subtask T1.1.2-1, while the solid part will be treated using continuum shell elements. This task has started recently. The first step has been to perform a fluid-structure interaction problem using the portion of the blade analyzed in 1.1.2-2. The key bottleneck in this simulation has been that the solids part turned out to be much more costly than the fluid dynamics part. For the solids part, the most time-consuming part is the solution of a linear system for the displacements. We are actively collaborating with the linear algebra experts from Work Package 3 to improve this. We are currently working with the MUMPS Team. MUMPS has been interfaced, and preliminary tests have been performed. Results are promising and have been reported in D3.2. Moreover, other packages that have already been interfaced with Alya and tested in Computational Fluid Dynamics problems, such as

PSBlas/MLD2P4, could also be tested for Computational Solid Dynamics problems.

T1.1.2-5: Comparing models

WALBERLA : *Integration of the actuator line model*

Wind turbine model The first step towards the implementation of actuator line based wind turbine models within the WALBERLA solver consists of the physical representation of the wind turbine. This step is crucial since it has a direct impact on the solver versatility.

The present implementation follows roughly the work of JOULIN [44]. To allow for a large number of wind turbine configurations, a wind turbine is split into several components which have different degrees of freedom. The wind turbine components together with their respective degree of freedom are given in Table 5.

Component	Attachment point	Degrees of freedom
Tower	At the ground	Clamped to the ground
Nacelle	Tower top	Yaw around tower top
Hub	Nacelle end	Tilt around nacelle and azimuth
Blade elements	Hub center + offset (hub radius)	Precone, pitch and twist around hub

Table 5: Wind turbine components with their respective degree of freedom

Each of the components is discretized several nodes at which the positions, velocities, and orientation matrices are calculated at each time step. The usage of orientation matrices rather than direct velocity transformations makes the code, again, more versatile.

By this choice of components and nodes, multiple angles determining a wind turbine configuration can be defined see Figures 11 and 12:

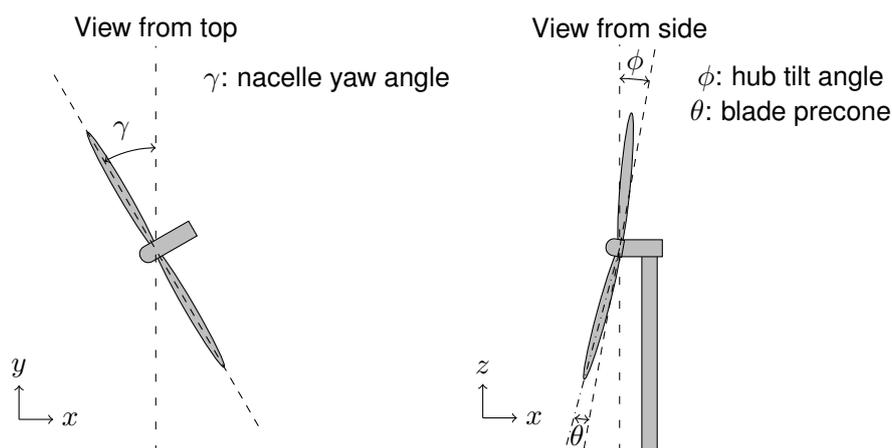


Figure 11: Schematic of a wind turbine: view of a yawed turbine from the top (left) and side (right)

Besides these angles and other quantities, the user needs to provide a distribution of chord, twist, and airfoil polars (i.e. lift and drag coefficients as a function of the angle of attack) for each airfoil used.

As a consequence of the details mentioned above, the representation of several wind turbine configuration is possible, amongst others horizontal axis wind turbines, vertical axis wind turbines, or even helicopter rotors. Note that the present implementation assumes a clamped tower: floating wind turbines cannot be simulated yet. This limitation will be alleviated in future work.

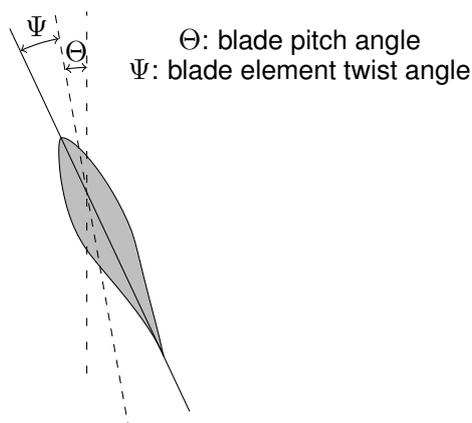


Figure 12: Schematic of a wind turbine blade element. The wind coming from the left.

Actuator line model It is yet to elaborate on how to couple the turbine information and the fluid flow. The actuator line model (ALM) has been implemented, for the blade and the tower modelling. The hub model is based on a rather simple disk model. The actuator line model follows three steps, the implementation is based on CHURCHFIELD ET AL. [18] and STEVENS ET AL. [82]:

1. Interpolation of velocity and density from the flow field to the actuator line nodes
2. Calculation of the blade element forces
3. Projection of the blade element forces back onto the flow

The interpolation and projection rely on the definition of kernels. To investigate the effect of the kernel choice on physics and performance, following kernels are available: the Discrete Dirac Delta kernel (ROMA ET AL. [71]), Gaussian kernels (e.g. [42]) and simple linear kernels with custom width.

After the interpolation step, velocities and density are known at each blade element. The angle of attack is computed using the relative velocity. Subsequently, the lift and drag coefficients are interpolated from the airfoil polars, and the forces are computed. These forces are projected back to the global reference frame and projected onto the flow field.

An issue often encountered with actuator line models is the over-prediction of blade forces near the blade tip, due to an under-resolution of the blade tip vortices. To circumvent this, blade forces are usually corrected using a model proposed by GLAUERT. This possibility has been implemented in the actuator line model. However, using a sufficiently fine mesh (*i.e.* at least 100 lattice per rotor diameter), realistic force distributions could be obtained, as will be shown later. The actuator line model for the tower is slightly altered by using only the drag coefficient as cylinders do not generate lift.

WALBERLA : *Preliminary validation of the actuator line model*

Two test cases have been considered for the preliminary validation of the actuator line model. The first case is based on the MexNext project experimental campaign, part of the IEA Task 29 [12]. The second one is based on measurements performed at the NTNU laboratory. Velocities in the wake of two aligned wind turbines have been measured [8].

Single rotor case: NewMexico First, let us consider the distribution of the time-average normal forces acting on one of the rotor blades. Pressure tap at five radial locations have been integrated from experimental measurement. Figure 13 shows the experimental data, WALBERLA simulation results for two mesh sizes (60 lattices per rotor diameter and 120 lattices per rotor diameter). Numerical results based on an inviscid vortex flow solver, developed at IFPEN (CASTOR) and on an actuator line model, implemented in

D1.2 First scientific results and progress on exascale tools

the MESONH meteorological solver, are also shown for comparison. The Glauert tip-loss correction is used in the MESONH simulations, while the vortex solver explicitly models the tip vortices using a non-diffusive Lagrangian method. No tip loss correction is used in WALBERLA, and the Roma et al. kernel is adopted.

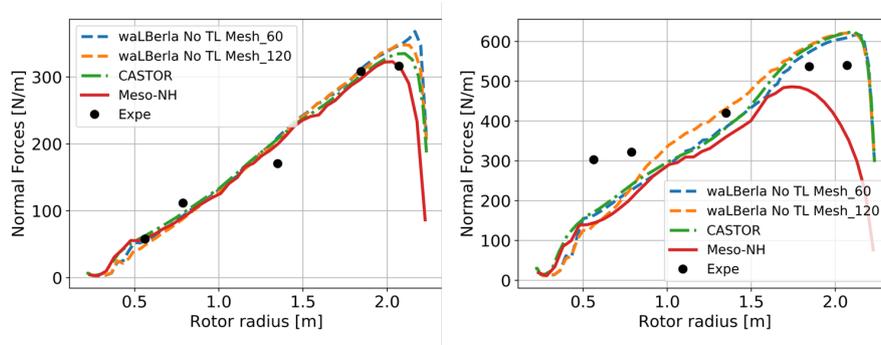


Figure 13: Normal force distribution on the NewMexico wind turbine blade

For the higher tip-speed-ratio (or lower wind speed, $u = 10\text{m/s}$), to the left, the agreement between simulations and measurements is very satisfactory. Some discrepancies are noticed near the tip of the blade. With 60 lattices per rotor diameter, the tip vortex is not sufficiently resolved, and a sharp decrease of the normal forces is noticed at the very tip of the blade. By increasing the mesh resolution, one gets a much smoother reduction of the forces near the tip of the blade, which is more physical. This indicates a better resolution of the tip vortex. WALBERLA results tend towards the vortex solver results, which is very encouraging. MESONH simulations, based on a lower mesh resolution, use the Glauert tip-loss correction for a realistic solution near the tip of the blade. For the lower tip-speed-ratio (or higher inflow velocity, $u = 24\text{m/s}$), results are less consistent with experimental data. In the first half of the blade, normal forces are strongly underestimated. This is due to the so-called rotational effects: at such operating conditions, large angles of attack are noticed. The blade is partially stalled, leading to the build-up of a span wise flow, that can not be represented using actuator line models. Near the tip of the blade, WALBERLA and CASTOR predict very similar trends, while MESONH underestimates the forces. This is due to the use of the Glauert tip-loss correction, that leads, in this case, to unphysical results. Comparison to fully-resolved RANS simulations can be found in [12].

Next figure shows the time-averaged velocities, given along a line starting from the hub and pointing toward the extremity of the blade in a 9 o'clock plane. Velocities are measured very close (30cm) in front and behind (in the wake) of the rotor. WALBERLA simulations, measurements and vortex simulations are compared in figure 14.

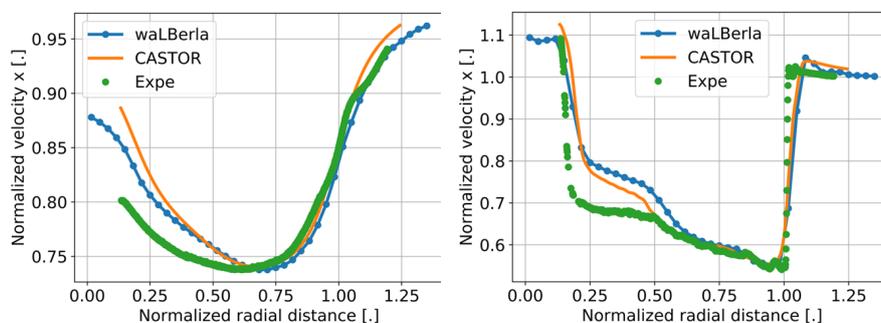


Figure 14: Time-averaged velocity distribution upstream and downstream of the NewMexico wind turbine

The second half of the blade shows good agreement between the simulations and the experimental data. Even the strong velocity decrease behind the rotor, due to the presence of a sharp and energetic tip vortex,

D1.2 First scientific results and progress on exascale tools

is rather well captured. In the first half, the hub and nacelle, that perturb the flow field, and are not modelled in either solver. This leads to some over-prediction of the velocities.

Tandem rotor case: NTNU The next case considered is a tandem rotor configuration. Two wind turbine are aligned, whereas the second one is located at a downstream distance of 3 diameters behind the first one. First, a snapshot of the instantaneous velocity field is given figure 15. One can notice the presence of the tip vortices right behind the first rotor. The flow is rather laminar behind the first turbine, with the impact of the blade passage visible in the wake. Afterwards, the flow becomes highly turbulent when it encounters the second wind turbine. At 8 diameters behind the first turbine, one observes almost homogeneous turbulence in the wake of the two turbines.

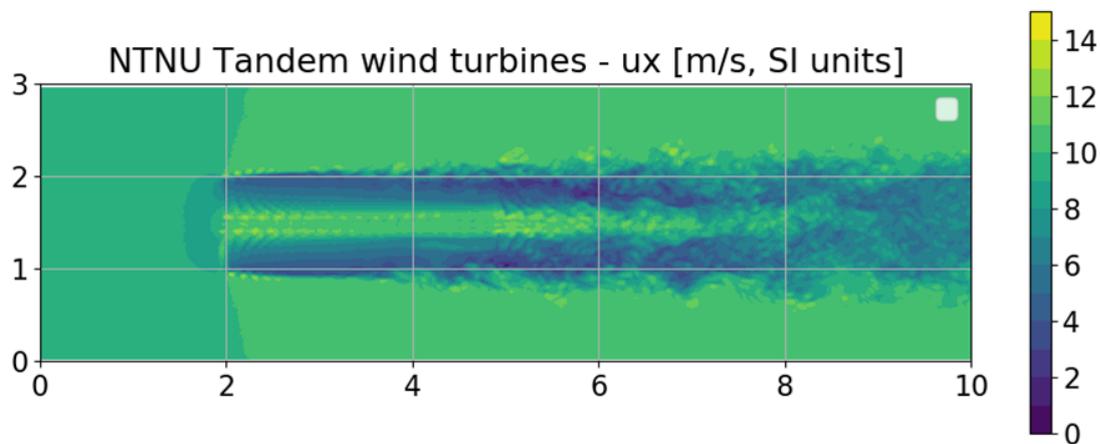


Figure 15: Instantaneous velocity field on a 2D plane (NTNU test case)

For a more quantitative comparison, velocity profiles at approximately 1, 3 and 4 diameters behind the second turbine are shown in figure 16. One can observe the transition from a nearly top-hat shape, very close to the second rotor plane, to a Gaussian shape, further away. This transition is rather well captured by WALBERLA, although the absence of hub and nacelle models in these simulations is visible, with a velocity equal to the inflow velocity at $1D$ behind the second rotor plane.

3.3 Code demonstrator

Alya

Alya [2] is a high-performance computational mechanics code to solve complex coupled multi-physics / multi-scale / multi-domain problems, which are mostly coming from the engineering realm. It is one of the two CFD codes of the Unified European Applications Benchmark Suite (UEBAS) as well as the Accelerator benchmark suite of PRACE. Among the different physics solved by Alya, we can mention incompressible/compressible flows, heat transfer, turbulence modeling, non-linear solid mechanics, chemistry, particle transport, electrical propagation, etc. For the wind problem, in particular, we use the first three physics to solve the fluid dynamics problem at the wind farm and wind turbine level. The solids mechanics module is used to analyze the wind turbine deformation. Moreover, we are starting to develop the fluid-structure interaction capabilities for the rotating wind turbine.

Since its conception, Alya was especially designed for massively parallel supercomputers, and the parallelization embraces four levels of the computer hierarchy. 1) A substructuring technique with MPI as the message passing library is used for distributed memory supercomputers. 2) At the node level, both loop and task parallelisms are considered using OpenMP as an alternative to MPI. Dynamic load balance techniques have been introduced as well to better exploit computational resources at the node level. 3) At the

D1.2 First scientific results and progress on exascale tools

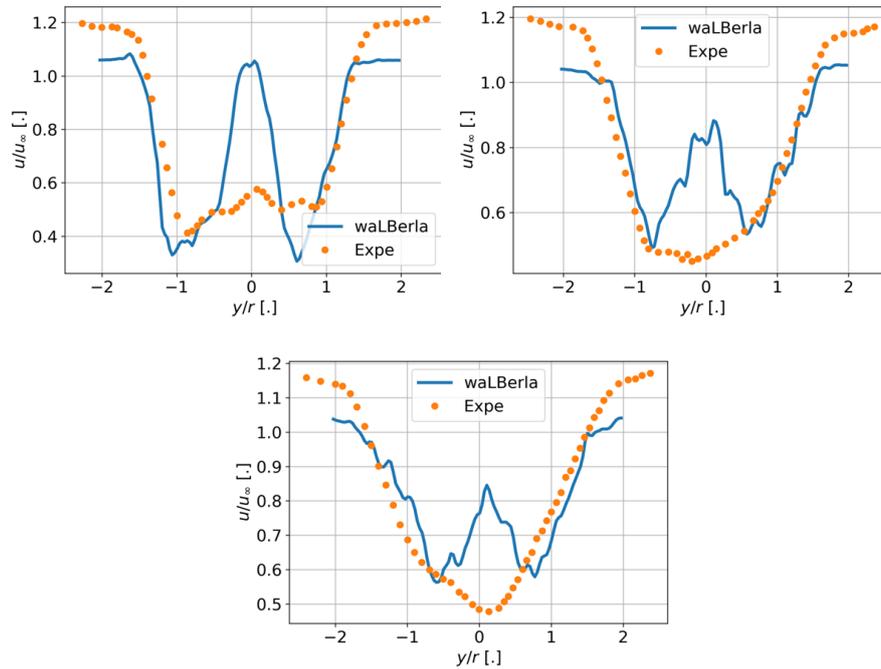


Figure 16: Time-averaged velocity profiles in the wake of the second turbine (NTNU test case) at 1 (top left), 2.5 (top right) and 4 (bottom) diameters behind the second turbine, with the second turbine at $x/D = 3$ away from the first turbine

core level, some kernels are also designed to enable vectorization. 4) Finally, accelerators like GPU are also exploited through OpenACC pragmas or with CUDA to further enhance the code's performance on heterogeneous computers. Within EoCoE-I, we introduced significant improvements at the CPU level and gave some first steps with the GPU version. Within EoCoE-II, we have a new team from Erlangen University (FAU) that are world-leading experts in node level optimization, and we expect to be able to make Alya even more efficient. Both Guillaume Houzeaux, Alya's main developer, and Herbert Owen, leader of the Wind scientific challenge, attended to the EoCoE Performance Evaluation Workshop at Erlangen in October 2019. An EoCoE Hackathon had been planned for the first week of April in Barcelona. Five developers from the Alya team had registered for the Hackathon taking advantage that it would be held at home. The objective was to create two teams to work on both the CPU and GPU implementations. Unfortunately the Hackathon was suspended due to Covid-19. Nevertheless, some of the work has been done remotely, and the FAU optimization experts are now working on the code.

Multiphysics coupling is achieved following a multi-code strategy, relating different instances of Alya. MPI is used to communicate between the different instances, where each instance solves a particular physics. This powerful technique enables the asynchronous execution of the various physics. Thanks to a careful programming strategy, coupled problems can be solved, retaining the scalability properties of the individual instances. This coupling that is being tested to allow for fluid-structured interaction problems. Moreover, for the case of the rotating wind turbine blades, a version of the coupling where just one instance of Alya is used for the fluid part allows the treatment of problems in which a part of the domain is rotating, and the other one is fixed. The coupling takes care of the interaction between both parts at the interface.

Finally, we would like to mention that most of the problems we deal with have two main parts: the assembly of a finite element matrix and the solution of a linear system. The assembly of the matrix is highly scalable, and the most demanding part for the scaling towards exascale is the linear system's solution. Before EoCoE, we relied only on in-house programmed code for the solution of the linear system. However, EoCoE-II includes partners that are leading figures of the European linear algebra community and their software packages. We have taken advantage of this opportunity to interface Alya with most of the packages:

D1.2 First scientific results and progress on exascale tools

Maphys, AGMG, and PSBlas/MLD2P4 iterative solvers; and Pastix and MUMPS direct solvers. We have performed successful weak scalability tests with both AGMG and PSBlas/MLD2P4 up to $2 * 10^9$ degrees of freedom (see deliverable D3.2). We believe that we could have applied one further level of refinement, leading to $1.6 * 10^{10}$ degrees of freedom and obtained positive weak scalability results. However, the main difficulty in performing such simulations has been the lack of computational resources. We spent all of the CPU time that was allocated in Marenostrum IV for half a year (200k CPU hours) with just six of $2 * 10^9$ degrees of freedom test. For the $1.6 * 10^{10}$ degrees of freedom test, we are not only short of CPU time, but we also find that the possibility of using a huge number of cores (around 100k) is only available once a year, at least in Marenostrum IV. This is typically in August, coinciding with the stopping of the machine for electrical maintenance. Further, details about the difficulties encountered can be found in D2.2 and D3.2.

3.4 Roadmap revision

There has been some delay in finishing T.1.1.1-1 (Improve wall modeling for Atmospheric Boundary Layer of Large Eddy Simulation in the code Alya developed at BSC). We could have finished this task with the submission of our first paper. Still, we came up with a new solution to improve the robustness of wall modelling in the finite context that deserved some further work. We have therefore put some additional resources into this task at the expense of T.1.1.1-2 (Improve the inflow boundary conditions for the Atmospheric Boundary Layer of Large Eddy Simulation for the Alya code), where we found that the solution we currently have works fine.

The overall work could have advanced much faster if it had not been for the COVID-19 crisis. For instance, the solution at BSC was to switch to home working which, for several researchers involved in this scientific challenge, has not been the right solution. Especially in case of research scientists with small children, switching to home office has resulted in a loss of concentration and productivity. This considerations together with the unrealistic hypothesis that a scientist can maintain the same level of efficiency during home office has lead to a lot of stress, which resulted in a degraded family environment. In turn, this degradation has further compromised productivity at home.

We have repeated the timeline from deliverable D1.1, where we have marked in a lighter color the periods we have some delay with respect to the original planing. For task T.1.1.1-3 (Including thermal coupling, Coriolis forces, canopy, and the actuator disc), we are advancing fine, and we hope we can finish within the expected dates.

WIND T1.1.1	Project Months											
		D1.1	MS1	D1.2				D1.3				
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.1.1-1												
T1.1.1-2												
T1.1.1-3												
T1.1.1-4												
T1.1.1-5												

We also repeat the timeline for T.1.1.2. As in the previous Table, we have marked in a lighter color the periods we have some delay with respect to the original planing. For subtask T.1.1.2-1 (Develop and test the sliding mesh approach for rotating blades in Alya using rigid blades), we have made significant advances. However, there are still some small points we plan to continue developing. The fact that the ParCFD congress has been postponed due to COVID-19 has given us some extra time to develop and test our sliding mesh implementation better. For subtask T.1.1.2-2 (Adapt shell elements to turbine blades), there has been some delay at the beginning, but now we are advancing at a good rhythm. Generating the meshes for the solids part has been significantly more time consuming than we had initially expected due to our lack of experience with this problem. For the Fluid-Structure interaction problem in T.1.1.2-3, our

D1.2 First scientific results and progress on exascale tools

initial steps have been quite promising. We now depend on the interaction with the linear algebra team in WP3 to make the solution more performant.

WIND T1.1.2	Project Months											
		D1.1	MS1	D1.2				D1.3				
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.1.2-1												
T1.1.2-2												
T1.1.2-3												
T1.1.2-4												
T1.1.2-5												

As part of increasing the visibility of the work carried out by the Alya team, Herbert Owen has contacted Michael Sprague, the leader of two US projects that are closely related to the Wind for Energy challenge; “Exawind” and “Atmosphere to Electrons”. In Figure 17, we show a screenshot from the interaction between Michael Sprague and Herbert Owen. Despite Europe has traditionally been a worldwide leader in wind energy, the resources nowadays being invested in the US to take wind energy to the Exascale are approximately one order of magnitude larger than those in Europe. With such resources, the US projects have been able to contract not only US researcher but also European leaders in the field. If we take into account the resources that have been invested in Wind for Energy, we believe that the Alya team is doing an excellent job . On the other hand, the resource disparity is such that it will be quite challenging to compete with the NALU code developed within the ExaWind Project.

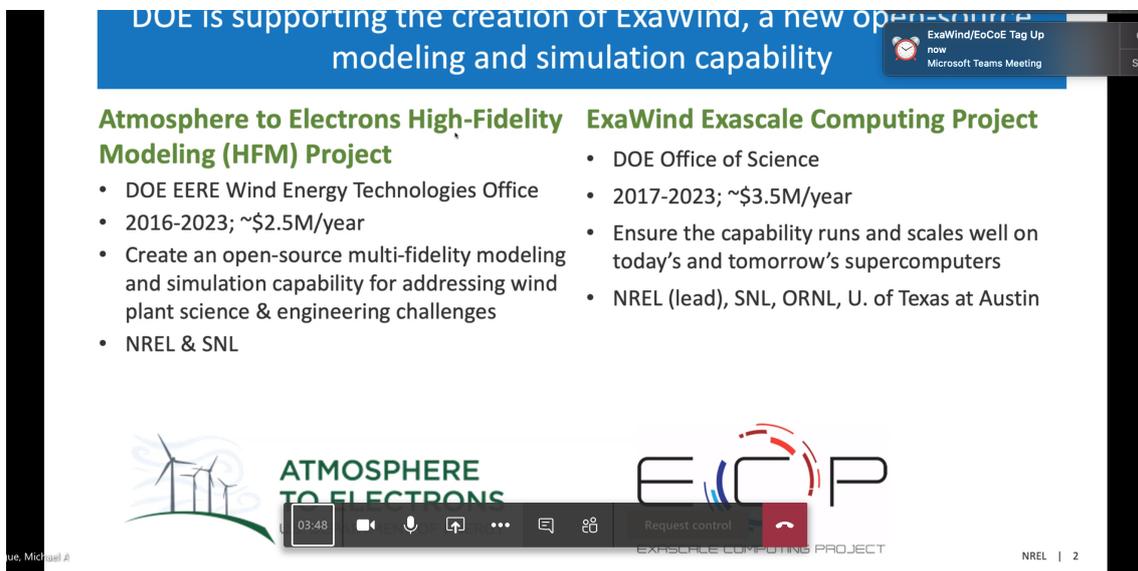


Figure 17: Exawind and Atmosphere to Electrons screenshot

4 Meteorology for Energy (T1.2)

T1.2 is divided in three main tasks.

T1.2.1	Continuous probabilistic short-term prediction of optical thickness and wind.
T1.2.2	Wind and solar power calculation for meteorological ensembles.
T1.2.3	Calibration of ensemble prediction of wind and cloud optical thickness.

D1.2 First scientific results and progress on exascale tools

Meteorology for Energy concerns large-scale ensemble weather predictions for renewables forecasting. The work is divided into three tasks. T.1.2.1 involves the development of the ensemble data at FZJ, including software developments to the ESIAS hyper-ensemble system and finally the simulation of forecasts for the other two tasks, including downward solar irradiation and the wind speed at turbine hub heights. T.1.2.2 applies satellite codes from Fraunhofer IEE to the ensemble data from FZJ. This includes the calculation of solar and wind power using the Solar Prediction (SPS) and Wind Power Management (WPMS) Systems, as well as the use of cloud-tracking to create continuous probabilistic predictions of the electric feed-in from the wind and solar ensembles. T.1.2.3 involves the calibration of the ensembles and the application of non-parametric model output statistics by CNRM to improve the wind and cloud predictions.

4.1 Task 1.2.1: Scientific results

Task 1.2.1 is divided in 3 subtasks:

T1.2.1-1	Probabilistic wind prediction. For the meteorological simulations the Weather Research and Forecast (WRF) model is adopted to predict winds typically at rotor hub heights, commonly taken at 100 m height.
T1.2.1-2	Probabilistic cloud optical thickness (COT) prediction. In addition, the radiative impact on solar energy by aerosol-induced turbidity (aerosol optical thickness, AOT) on solar energy production will be forecasted by the EUROpean Air pollution Dispersion-Inverse Model (EURAD-IM).
T1.2.1-3	Joint simulations. In both cases, non-Gaussian data assimilation by particle filter and smoother methods will be applied, with remote sensing data processing of satellite images as a prominent data source, combined with big data analytics, based on suitably selected metrics. A prominent objective includes a middleware based flexible and non-synchronous hyper-ensemble operation by MELISSA middleware. This model will also be integrated in stochastic mode by ESIAS-chem and operated in hyper-ensemble mode. The result will be evolving pdfs as an approximation to the corresponding Fokker-Planck equation.

T1.2.1-1 & T1.2.1-2: Probabilistic wind and COT prediction

While the results of EoCoE-I showed that the shape of the ensemble distribution contained useful information, e.g. about the type and likelihood of extreme events, the underlying Weather Research and Forecasting (WRF) model gave very biased wind speeds (see figure 18) and COT in Europe, where it was not developed or parameterized and is not commercially developed. This room for improvement in the WRF-based ensembles was seen both in the work comparing wind and solar power calculations to German meter data and in the model output statistical calibration based on wind measurements. While much attention was given to the correction of the boundary layer schemes for wind speeds in EoCoE-I, the calibration of the cloud optical thickness is a decided focus for EoCoE II.

A test of 12 micro-physics scheme, 7 cumulative physics schemes, and 8 boundary layer physics schemes has thus been conducted on ESIAS-met. The details of the selected physical schemes and common physical schemes used in all the cases are listed in Table 6 and Table 7, respectively. The test was performed on twelve days (288 total simulation hours) with 20 km spatial resolution. The results via-à-vis the wind and COT ensemble calibration are discussed in section 4.3.

Joint simulations with MELISSA The software structure is analysed and the information about how to integrate with the middle-ware based flexible for non-synchronous hyper-ensemble operation, MELISSA. This information is provided to WP5 for integrating MELISSA middle ware. Also a Python module, which is used to aid the ensemble simulation, is made for integrating into MELISSA.

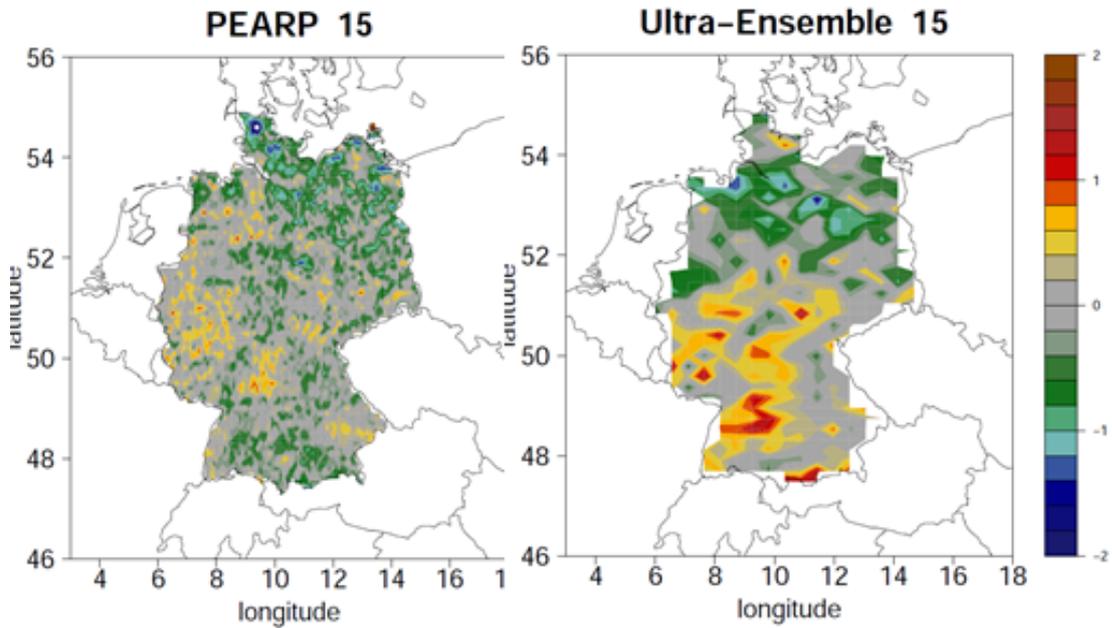


Figure 18: Error of ensemble ten-meter wind speed mean for the in-house PEARP ensemble (left) and WRF ensemble (right) groups from EoCoE-I.

Table 6: Employed Physical schemes setting of ESIAS-met

Micro-physics	Cumulative physics	planetary boundary physics
Kessler	Kain-Fritsch	YSU
Lin (Purdue)	Betts-Miller-Janjic	MYJ
WSM3	Grell-Freitas	GFS
WSM5	Simplified Arakawa-Schubert	QNSE
Eta (Ferrier)	Grell-3	MYNN2
WSM6	Tiedtke	MYNN3
Goddard	New SAS	ACM2
Thompson		BouLac
Milbrandt 2-mom		
Morrison 2-mom		
CAM 5.1		
SBU-YLin		

Table 7: Common physical schemes setting of ESIAS-met

Short-wave Radiation	Long-wave Radiation	Surface Layer	Land Surface
Duhia	RRTM	MM5 similarity	5-layer thermal diffusion

4.2 Task 1.2.2: Scientific results

Task 1.2.2 is divided in 2 subtasks:

T1.2.2-1	Power calculation from ensemble winds and COT¹. Training of the IEE wind power management system reference plants and ESIAS ensembles. Optimization of the IEE solar prediction system for varying modeling conditions in a two-step process: 1) Optimization of power model to German TSO electric meter data based on live satellite measurements of COT. 2) Spatially resolved calibration of a WRF input time series to the satellite COT grid. Solar and wind power calculations for WRF meteorological ensembles of COT and 100-m wind will be provided as probability density distributions.
T1.2.2-2	The confluence of ultra-large ensembles with IEE's satellite-based cloud-tracking system for short-term forecasting. Wind velocity gradients at cloud heights from a Taylor-based cloud-tracking system can identify flow structures and dynamically weight the intra-day wind power ensemble, as can satellite measurements of cloud cover for the solar power forecast. This ensemble member validation will provide feedback to the particle filter by weighting the ensemble member performance. The general output will be pdfs of power production, including inference of higher statistical moments.

T1.2.2-1: Power calculation from ensemble winds and COT

This first part of the subtask at Fraunhofer IEE regards the wind and solar power feed-in calculations based on the Task 1.2.1 ensemble data. While EoCoE I relied on raw ensemble data, the SPS and WPMS power models will be calibrated in EoCoE II based on meteorological training data expected in September, 2020 (M21), with the goal of having results comparable to actual meter data. For wind, the existing training framework of WPMS will be applied, which trains reference plant forecasts based on the meteorological model and reference plant power measurements.

SPS does not use reference plants to determine the local solar power. In fact, reference plants are not terribly representative for solar, due to the ubiquity of smaller, unmeasured residential systems. In the case of solar, however, satellite measurements of irradiation provide a real-time and spatially resolved target. In EoCoE II, a two-step trainings framework has been developed. In the first step, the weather model irradiation can be calibrated pixel-by-pixel to satellite data in Germany, as in figure 19. As each location is independently modeled, the smoothness of the calibration map is encouraging regarding its robustness. We note that the trends of the calibration mostly follows German topography, hopefully making reasonable corrections to shortcomings of the numerical weather prediction's parameterization or vertical resolution. In any case, for consistency, the satellite data is used as the meteorological input of the second step, the power model calibration to German meter data. This second step of the training framework ideally benefits from using real-time irradiation inputs instead of weather model data with day-ahead forecasting errors. Figure 19 represents data averaged from two operational models, ICON and IFS. With this approach,

¹Description edited from D1.1 for accuracy differentiating wind and solar training

D1.2 First scientific results and progress on exascale tools

the day-ahead forecast error of German solar power could be significantly reduced from 3.5% to 3.2% (r.m.s.e.). This framework will be applied to the training ensemble data from WRF expected in M21, such that the ultra-large ensemble power is comparable to real wind and solar power feed-in meter data.

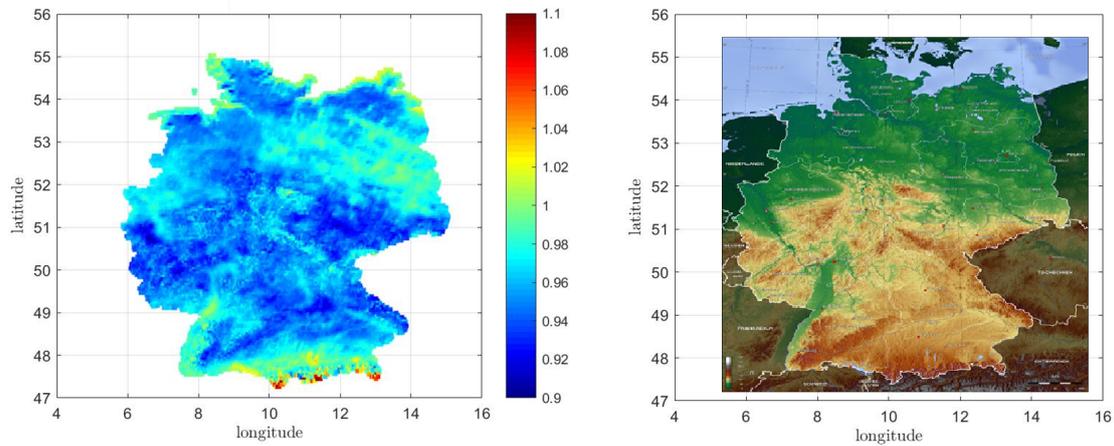


Figure 19: On the left, pixel-by-pixel linear calibration weight of a weather model to satellite irradiation data in SPS preproc. On the right, topographical map of Germany [13].

T1.2.2-2: Confluence of ultra-large ensembles with satellite-based cloud-tracking

The second part of the work at Fraunhofer IEE employs a new cloud-motion system that follows synoptic motion with the goal of scoring ensemble members. Scores are needed to dynamically weight the intra-day forecast and provide feedback to the particle filter. As the cloud-motion tracker is based on a Taylor expansion, there is both a continuous description of the flow u_i and its velocity gradient tensor $A_{ij} = \partial_i u_j$. Since all two-dimensional flow topologies can be linearly approximated by the invariants $P = -tr(A)$ and $Q = Det(A)$, these have been developed as constraints on the optimization to keep the result physical and ideally to serve as a basis to score the synoptic topologies of different ensemble members.

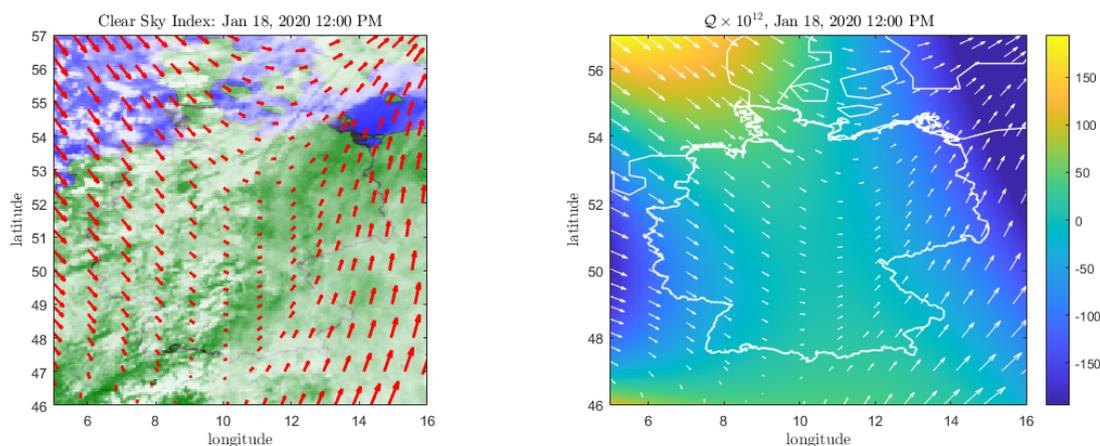


Figure 20: Cloud motion (left) tracked on January 18, 2020. The same streamlines are overlaid on the invariant Q (right), which characterizes the local flow.

An example is shown in figure 20, which shows a day that resulted in a large wind power feed-in forecasting failure, as day-ahead forecasts predicted that the center of the synoptic system and its light winds would pass over Germany, when it instead passed through southern Sweden, resulting in strong winds in Germany. The invariant Q is sensitive to the strength of the flow, but more generally is positive where

D1.2 First scientific results and progress on exascale tools

the flow is vortical and negative where straining motion is concerned. In the figure, Q helps differentiate regions of the flow over Germany. (P is the divergence of the synoptic flow and more an indication of its three-dimensionality; it appears to be less descriptive than Q of the streamlines). The classifications now possible in the cloud-motion system will be compared to the ensemble data from T1.2.1 when it is available. As the exact location of synoptic events is difficult to simulate and is often responsible for expensive forecast failures, it's hoped that this comparison with cloud-motion will be able to identify crucial ensemble members.

4.3 Task 1.2.3: Scientific results

T1.2.3-1	Calibration of improved probabilistic wind ensemble predictions. FZJ will calibrated the WRF physical schemes underlying the wind ensembles ² . CNRS-CNRM will exploit the ultra-large ensembles of forecasts in order to produce improved power predictions and wind velocity (for wind power), for lead times from 3 hours to one day. MF will work on the development of statistical non-parametric calibration. The method of non-parametric calibration will be adapted to 100 m winds at the site locations of wind farms.
T1.2.3-2	Calibration probabilistic downward solar radiation ensemble predictions. FZJ will calibrated the WRF physical schemes underlying the COT ensembles. CNRS-CNRM will exploit the ultra-large ensembles of forecasts in order to produce improved power predictions with focus on downward solar radiation, for photovoltaics, for lead times from 3 hours to one day. MF will work on the extension of the method of statistical non-parametric calibration for clouds. In addition, methods for the identification of cloud fields along their type of formation will be included.

T1.2.3-1: Calibration of probabilistic wind ensemble predictions

The first step for the ensemble calibration is to address the underlying physical schemes. After simulating an unprecedented 672 combination of schemes, the wind speed was scored by 1172 ground stations at the two-meter level using the index of agreement[49]. The Figure 21 heat map shows that the choice of PBL physics most affects the hourly wind speed. From the resulting score of wind speed, the Mellor-Yamada-Nakanishi-Niino scheme (MYNN2 and MYNN3) and the Asymmetric Convective Model (ACM2) perform best.

T1.2.3-2: Calibration of probabilistic downward solar radiation ensemble predictions

For solar, we found that using CAM5.1, WSM3/5/6 and Goddard can provide better detail of cloud coverage, as shown in Figure 22 . The calculation results is scored by comparing to the satellite data of cloud coverage from Satellite Application Facility on Climate Monitoring (CM SAF)[83]. A result of daily average score for cloud mask covering rate is shown in Figure 21 . The microphysics has the greatest effect on

²The T1.2.3 subtask descriptions from D1.1 has been complemented to include FZJ's precursor work on ESIA

D1.2 First scientific results and progress on exascale tools

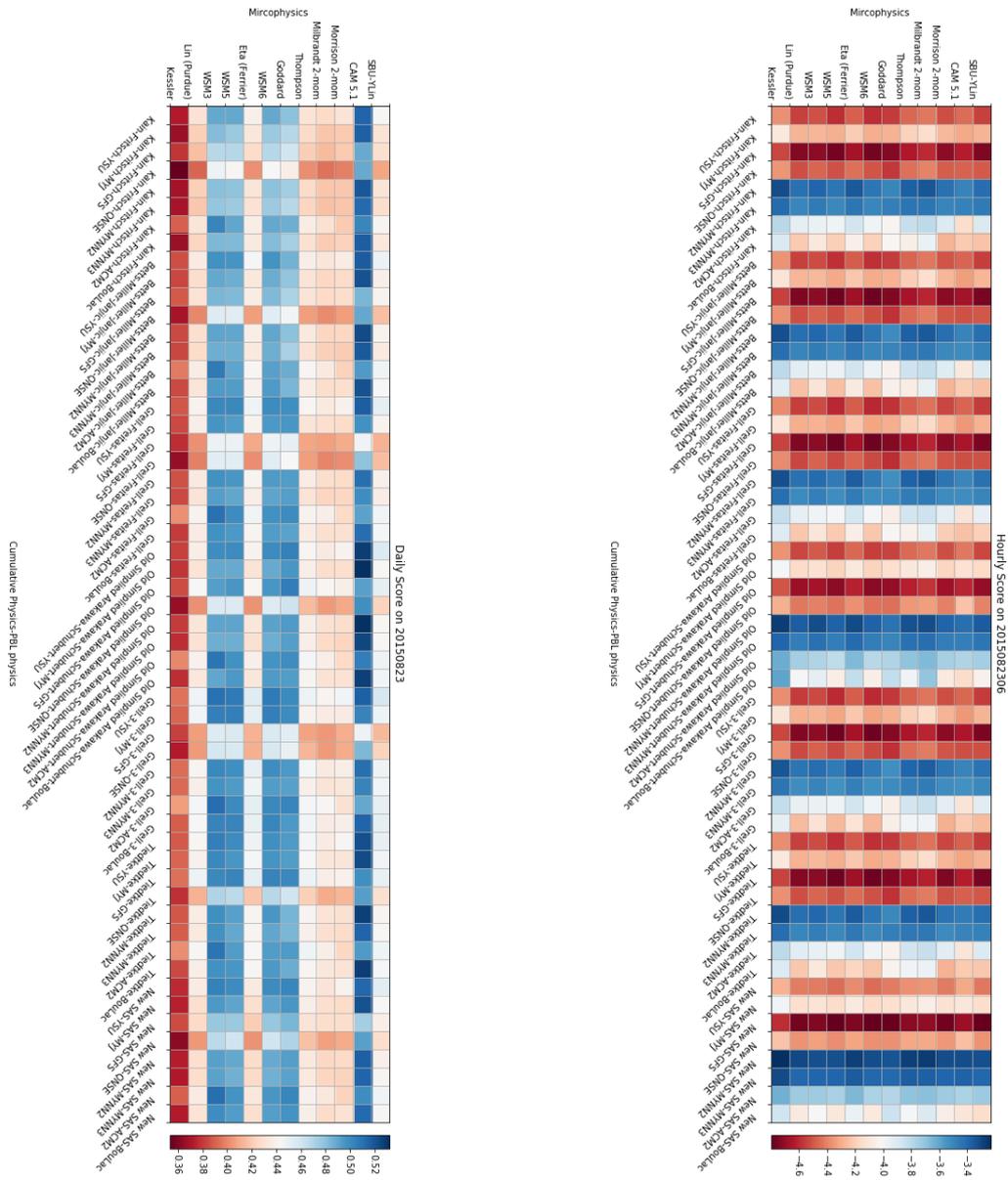


Figure 21: Ten-meter hourly wind speed score (top/right) on 2015-08-23 and cloud matching rate score (daily average, bottom/left).

D1.2 First scientific results and progress on exascale tools

the prediction of cloud coverage. Across all twelve day-long simulations, the most accurate microphysical scheme is the Community Atmospheric Model 5.1 (CAM5.1), though we note that this scheme takes longer to simulate than others. The Goddard scheme works better than WSM3/5/6 schemes during the late spring and early autumn, but produce similar results during the summer days. The cumulative schemes, Tiedtke, Grell-3, and Simplified Arakawa-Schubert, achieve good scores with above mentioned microphysics.

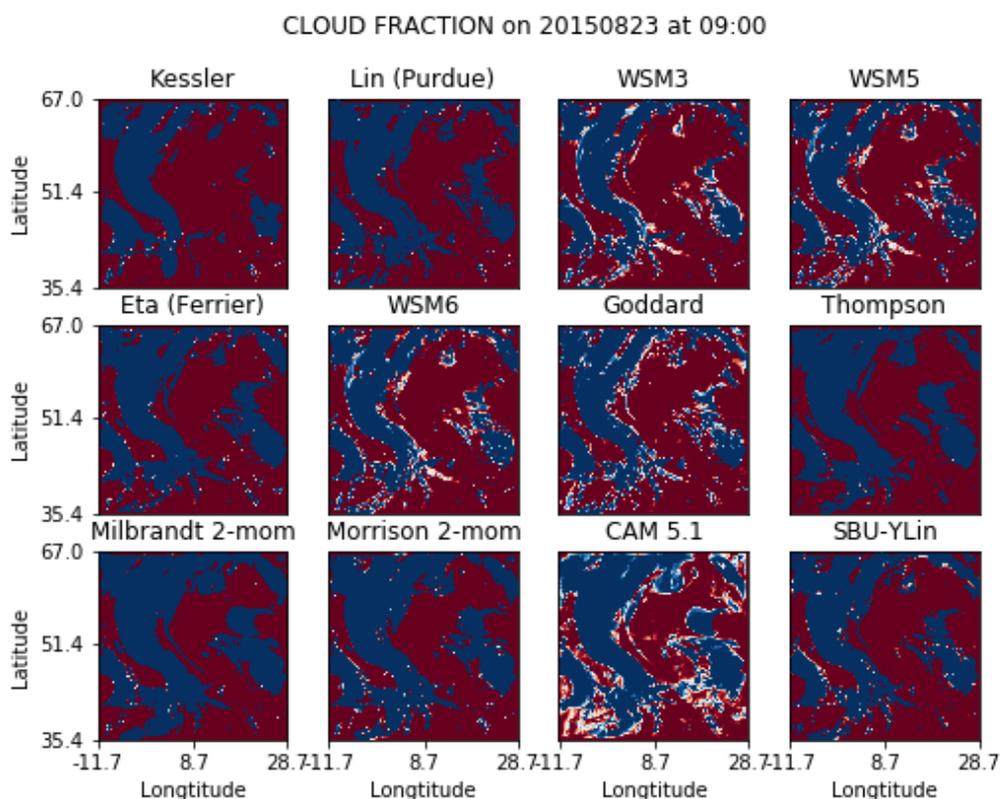


Figure 22: The of cloud fraction (0% (red) to 100% (blue) at 09:00 on 2015-08-23.

Summary and Outlook The 672 ensemble cases for sensitivity analysis on the parameterization combination of ESIAS-met indicate the most suitable choice of physics for the European domain. However, the investigation of choice of radiation parameterization and the choice of land surface model is not included yet. In the meanwhile, we are investigating on the effect of spatial resolution on the performance of weather forecasting. In the future study, we will continue this sensitivity analysis using the large ensemble capability of ESIAS-met to find the best choice of parameterization combination and the physics for performing better stochastic weather forecasting.

4.4 Code demonstrator

To create a sustainable development environment, a git repository is being hosted by JSC for documentation and a hub for deployment. The installation guide and pre- and post-processing tools are available on the git repository. The link to the git repository can be found at: jugit.fz-juelich.de/ye.lu/esias-met

T1.2.1-1/2: ESIAS

Though the framework of ESIAS-met is already ready, there is room to improve its operation. The framework script using to drive the ensemble simulation has been revised and re-written to provide better

D1.2 First scientific results and progress on exascale tools

usage for performing ensemble simulation. The instruction is well documented. The very user-friendly environment and common KSH code can be deployed very quickly for fast simulations. The WRF pre-processing system (WPS) is now driven without any modification, and therefore ESIAS can freely employ different version of WPS. Figure 23 shows the structure of the whole framework.

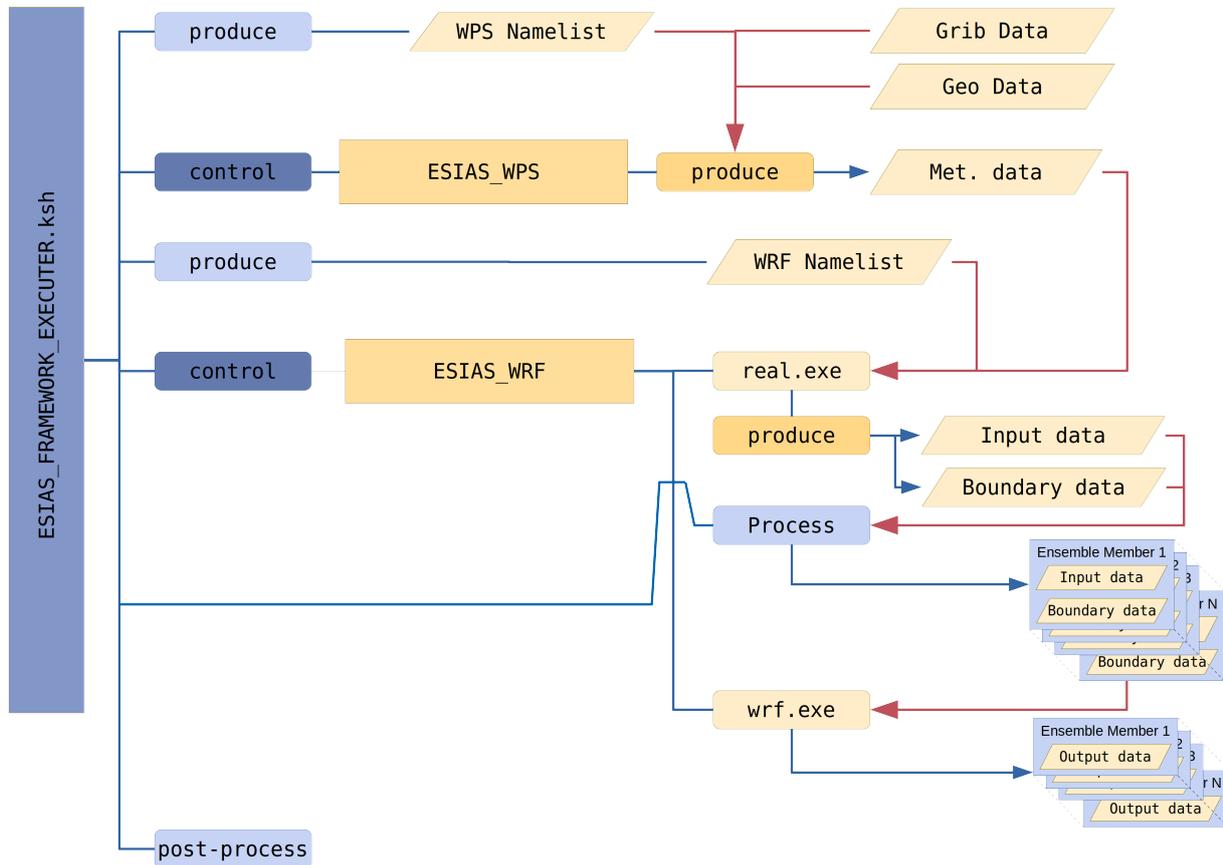


Figure 23: The schematic plot of grids analysis method that is use to proceed data analytics

T1.2.1-3: Data Analytics A base-algorithm for analyzing data with different spatial resolution is built to perform the processing of satellite images. This algorithm can also apply some re-sampling method like Gaussian filter or averaging filter. The concept of the resampling algorithm is shown in Figure 24

D1.2 First scientific results and progress on exascale tools

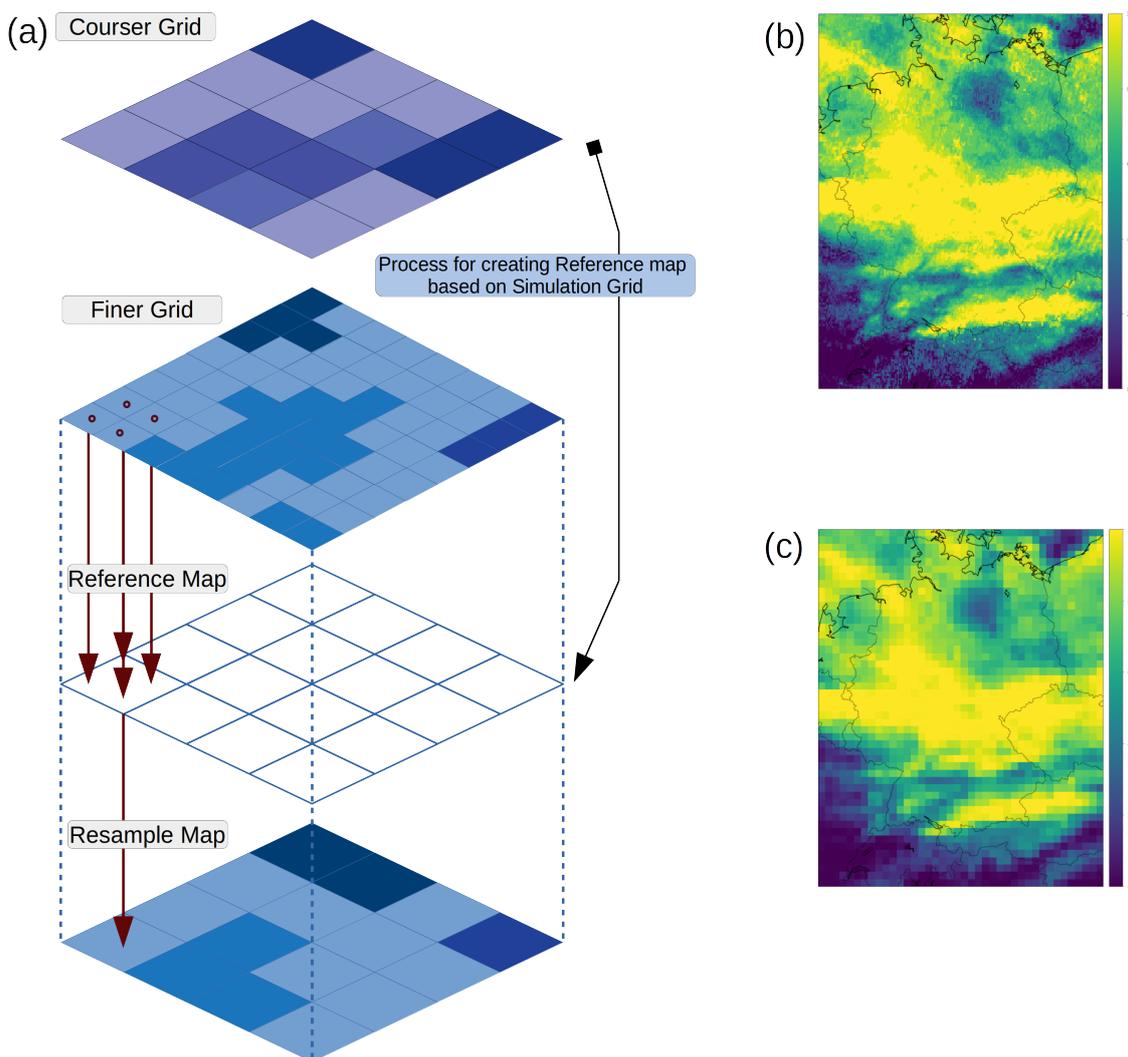


Figure 24: (a)The schematic plot of grids analysis method that is use to proceed data analytics. (b) The satellite data of cloud coverage before re-sampling and (c) after re-sampling by averaging.

4.5 Roadmap revision

Meteo T1.2	Project Months											
		D1.1	MS1	D1.2						D1.3		
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.2.1-1												
T1.2.1-2												
T1.2.1-3												
T1.2.2-1												
T1.2.2-2												
T1.2.3-1												
T1.2.3-2												

D1.2 First scientific results and progress on exascale tools

Status report Meteo for energy had a delayed start due to hiring processes following the departure of personnel at FZJ, including the original scientific challenge leader. Work on the first subtask was able to begin at M12, compressing the schedule for the rest of the project, including dependencies with the other subtasks. While not necessary for the final deliverable, a software analysis of ICON as an alternative to WRF was discussed in D1.1. This analysis or certainly any subsequent integration of a new weather model is now unlikely. We hope instead to achieve a better parameterization of WRF. T1.2.2 is currently on schedule. For the power calculations, the frameworks are ready for the delivery of data from T.1.2.1. Work on the methodology for T1.2.2-2 is ongoing and may proceed in the short term with in-house ensemble data. There are minor changes to the roadmap for T1.2.3, as a dedicated post-doc will be hired M21-M33 for EoCoE II work at CNRS-CNRM (non-parametric calibration). This extends to almost the end of the project for the final deliverable, but begins when the necessary data can be expected from the delayed T1.2.1. A candidate has already been selected for the position. In the meantime, the calibration of ESIAS-met has proceeded at FZJ.

Risk assessment Overall there remain minor to moderate risks due to hiring delays and the ongoing COVID-19 crisis affecting the efficiency of work, e.g. due to the availability of child care. This risk is less to the overall ability to deliver than to the quality of the final forecasts vis-à-vis calibrations, and the scope of the investigations, particularly regarding cloud-motion feed-back to the particle filter (T1.2.2-2), which is likely to require some iteration between partners to be successful. We will strive to mitigate these risks through careful coordination of the remaining work, as the compressed timeline will make it critical that work can be exchanged between partners without delay.

5 Materials for Energy (T1.3)

T1.3 is divided in three main tasks:

T1.3.1	Shedding light on carrier dynamics at hetero-interfaces in silicon solar cells
T1.3.2	Harvesting electricity from salinity or temperature gradients
T1.3.3	Organic and perovskite solar cells

T1.3.1 highlights the scientific objectives and roadmap for optimizing silicon solar cells to increase in performance and extension of lifetime. Passivation contacts is one of the most promising alternative techniques to crystalline silicon (c-Si) solar cells for a high performance-to-cost ratio. An ideal passivation contact is able to suppress the electrical losses at interfaces and, at the same time, to keep ultralow contact resistivity for the transport of one type of carriers (holes or electrons) while hindering the transport of the other type. In recent years, the silicon heterojunction (SHJ) solar cells reached the highest efficiency of 26.6% [86], mainly due to the passivation contacts. In these devices, intrinsic hydrogenated amorphous silicon (a-Si:H) was used to passivate the Si surface and the p/n-type doped hydrogenated amorphous silicon was employed to select the transport carriers. The application of SHJs offers several advantages: first, a-Si:H provides efficient passivation of Si dangling bonds at the interface; second, field-effect passivation can produce a significant inversion effect at the a-Si:H/c-Si interface that is able to increase the carrier lifetime. Moreover, fewer process steps are required to build-up the SC, and the low-temperature (< 200°C) processing allows the use of very thin wafers without any substrate damage. However the microscopic mechanisms of transport and recombination mechanisms at the interface are still poorly understood. The purpose of the present task is to understand the transport mechanisms underlying photovoltaic devices based on SHJ technology by simulating at atomistic resolution amorphous-crystalline heterointerfaces. We have used classic molecular dynamics (MD) simulations to build up realistic a-Si:H/c-Si interface structures at different temperatures. The ab initio characterization has been executed on selected configurations to monitor the electronic properties of the c-Si/a-Si:H interface.

D1.2 First scientific results and progress on exascale tools

T1.3.2 aims at providing computational support for optimizing capacitive blue energy electrodes and thermo-electrochemical devices. Electric power production from salinity gradients harvests the free energy lost during the mixing of river with sea water in estuaries. The main technologies developed for this purpose to date exploit the electric potential differences applied across membranes, but another approach based on capacitive mixing was recently proposed. Thermo-electrochemical devices employ the variation of the redox potential of an active species with temperature to convert a gradient into electricity. Ionic liquids were recently proposed as optimal media for performing such an energy harvesting. The purposes of this task are i) to ascertain the best electrode structure which optimizes a blue energy production and ii) to find compositions that will enable optimal performances of thermo-electrochemical cells. In both cases, a fundamental understanding of the cation and anion adsorption at the surface of the electrodes is essential. The challenge for simulation is that simulating the interfaces requires to rigorously account for the interactions between the electrode atoms and the adsorbed species. Due to the large size of the simulated systems for the final application, it is not possible to use electronic DFT for such calculations. We therefore aim at developing new force fields for classical molecular simulations. The parameterization can be made on a series of electronic DFT calculations. However, it was shown recently that the commonly used functionals may yield very different results for the adsorption energy of the molecules. During the period we have setup a series of systems in order to perform Quantum Monte-Carlo reference calculations to benchmark them. We have applied for computer time within several calls in order to run these calculations. In parallel, we have started to fine-tune force fields for classical molecular simulations for the bulk liquids, so that we will only have to develop the electrode-liquid potentials once the Quantum Monte Carlo calculations will have been completed. We have also developed a machine-learning approach to predict the redox potentials of active species in ionic liquids.

The scientific focus of T1.3.3 is optoelectronic devices based on two classes of emerging photovoltaic materials, namely organic molecular semiconductors and metal-halide perovskite semiconductors. The two classes of material have very different properties, which impact strongly on the electron dynamics within them. Organic molecular semiconductors are typically amorphous, possess low permittivities, and electrons are localised on individual molecules, so that transport proceeds via discrete hops between molecules. In contrast, metal-halide perovskites are (poly)crystalline, have high relative permittivities, and electron are delocalised so that their transport is "wave-like". These differences in the two materials necessitate different computational approaches, and therefore this task involves the development, implementation and application of simulation methodologies which are amenable to studying phenomena at a broad range of length and timescales. Transport phenomena in both classes of material remain ill understood, and are addressed in this task.

In the organics, low relative permittivities are expected to lead to strong interactions between carriers and ionised charged dopants, and there is experimental evidence that those interactions have electrical consequences which are qualitatively different from those in conventional inorganic semiconductors. Furthermore, the effects of injection and extraction of carriers from metallic contacts are poorly understood. Theoretical investigations into these effects have been severely limited due to the computational expense of calculating the long ranged electrostatic interactions in systems of sizes sufficient to capture the relevant phenomena. To address those issues, we will develop and utilise solvers based on bespoke algorithms for the fast and efficient computation of the electrostatic interactions in kinetic Monte Carlo (KMC) simulations. This allows the study of much larger systems with realistic levels of disorder and carrier densities, and the inclusion of metallic (Ohmic) contacts.

On the other hand, the soft nature of perovskites means that ionic distributions change under operating conditions, which strongly affects electron transport and device characteristics. Although continuum scale, drift-diffusion (DD) models have been developed and applied relatively extensively, such models do not allow for non-equilibrium electron distributions, which is crucial for understanding e.g. dependence of device characteristics on illumination wavelength and hot carrier effects. In order to address this shortcoming, we will develop a code which solves the semi-classical Boltzmann equation for electrons using the device Monte Carlo (DMC) method. This will be coupled to a drift-diffusion simulator for the ionic motion. We can

D1.2 First scientific results and progress on exascale tools

therefore identify how ion motion affects electron dynamics, in terms of scattering and hot carrier effects. All of the implementations developed as part of this task will be based on MPI parallelised, Python frameworks/libraries: (PPMD) [75] (KMC and DMC) and coulomb_kmc (KMC) [76]. Doing so allows us both to take advantage of the benefits of the separation of concerns approach to scientific programming, and allows for ready coupling between the different simulation techniques/codes to address the complex multi-scale problems in materials science.

5.1 Task 1.3.1: Scientific results

The main task T1.3.1 is subdivided in three subtasks:

T1.3.1-1	Classical Molecular Dynamics (MD) simulations of c-Si/a-Si:H interface. We have used ReaxFF (Reactive Force Field) molecular dynamics to efficiently simulate the thermalisation, quenching, and equilibration processes involving thousands of atoms forming realistic a-Si:H/c-Si interface structure (large interface model), during up to ten nano-seconds at T=300, 500, 700 and 900 K. Snapshots of the equilibrated c-Si/a-Si:H interface atom configurations have been generated to be characterized in T1.3.1-2.
T1.3.1-2	First-principles electronic properties of c-Si/a-Si:H interface. A workflow has been used for the first-principles characterization of snapshots of classical MD simulations (T1.3.1-1) to monitor the electronic properties of the c-Si/a-Si:H interface. We used the PWscf code of the Quantum Espresso suite to perform the first-principles calculations. The evolution of the intragap states at different temperatures is monitored by analyzing projected and local density of states, electrostatic potential, and charge density. This all will allow to select the atomic configurations that will be used in the electronic transport model, in order to design more efficient silicon solar cells belonging to the silicon heterojunction technology.
T1.3.1-3	Non-equilibrium Green's functions transport properties c-Si/a-Si:H interface. In order to analyze the effect of interfaces on the carrier transport and dynamics in silicon solar cells we use the results from T1.3.1-2 as input for a quantum transport code based on the NEGF formalism. To this end, we adopted a new code, libNEGF, in substitution to PVnegf. This includes the treatment of the contact self-energies, the coupling to a solver for the self-consistent Poisson equation, and the inclusion of electron-photon and electron-phonon scattering processes via the corresponding self-energies. The input for these calculations is a tight-binding Hamiltonian that reproduces the ab initio electronic structure of the heterostructure of T1.3.1-2. For the time being we used the relaxed and quenched structured provided by the T1.3.1-1 and generate a tight-binding Hamiltonian using DFTB+. At the same time, we are in the process of representing such an Hamiltonian in the more appropriate basis of localized Wannier functions that constitutes a very natural and very accurate basis for extended bulk states. The application of the NEGF formalism to the large interfaces from T1.3.1-2 requires a high-parallel and scalable code. Therefore, the optimization of the libNEGF code will need to be carried on in parallel to this task in the WP2 Task 2.4.

T1.3.1-1: Classical Molecular Dynamics (MD) simulations of c-Si/a-Si:H interface

Methods. Starting from the DFT optimized two a-Si:H/c-Si interface structures discussed in the next section, a subsequent classical MD analysis of the final ab initio configuration has been carried out by means of LAMMPS [63]) using the ReaxFF training set parametrization previously employed for the simulation of H bombardment of Si, Ge and SiGe (100) surfaces [64, 32]. This training set parametrization includes the dissociation of Si-Si bonds in the Si₂H₆ and Si₂H₄ molecules, therefore taking into account for single and double silicon bonds, as well as Si-H bond dissociation of the SiH₄ molecule. The entire dissociation energy landscape is obtained by means of adiabatic energy DFT calculations vs. bond length and covers from the equilibrium distance to the dissociation limit. The energy dependence on valence and torsion

D1.2 First scientific results and progress on exascale tools

angles in the Si-H force field is also incorporated. For example, for silicon bonds, the angular dependence is included in the form of the adiabatic DFT energy of the Si_3H_8 molecule as a function of the Si-Si-Si bond angle, and torsional terms are adapted by including energy differences between chair, boat, and planar conformations of $\text{c}-(\text{SiH}_2)_6$ six-member rings. Several other reactions are included, such as conversion of Si_2H_4 to $\text{H}_3\text{Si-SiH}$. Finally, the force field accurately takes into account also the cohesive energy and the equation-of-state DFT predictions for various silicon crystal phases, including the simple cubic Si, α -Si, and β -Si phases. The complete MD analysis starts with $T=0$ K minimization: firstly, fixing all cell dimensions: $L_x=L_y=23.22$ Å and $L_z=2\times L_x$ Å, then with $L_x=L_y$ still fixed but with L_z varying as an additional degree of freedom. The resulting geometry (see Figure 1b) is used as the initial condition for a subsequent quenching-thermalisation process. Initially, the system is heated up to 1100 K at zero pressure with a NPT (Nose-Hoover thermostat and barostat) [41] for 325 ps and next it is cooled down to the desired final temperatures of 300, 500, 700 and 900 K in 325 ps. In Figure 1c the final configuration of the quenching process is shown. A final thermalisation procedure is applied during 10 ns keeping a constant temperature with a csld [14] thermostat to avoid the flying ice cube artifact. A constant 1 fs integration time step is employed in the 300 and 500 K cases whereas at 700 and 900 K a smaller 0.5 fs time step was used to maintain the numerical stability of the simulations. The pressure is controlled along the z coordinate exclusively, keeping L_x and L_y box sizes fixed and allowing L_z to evolve isobarically ($P=0$).

Scientific results. In Figure 25 we report the snapshots of the c-Si/a-Si:H interfaces at different steps of the MD analysis. Silicon atoms and their bonds are in orange in the c-Si side and are in yellow in the a-Si:H side, hydrogen atoms and bonds with silicon atoms are in blue and bonds between c-Si and a-Si atoms are in red. It is applied a distance cut-off of 2.85 Å and 1.7 Å for Si-Si pairs and for Si-H pairs, respectively. In panel a) the ab initio relaxed system used as starting configuration of the MD simulations is shown. The following minimization at $T=0$ K (see panel b) produces a shrinking of the simulation cell along the z direction, indeed L_z changes from 46.44 Å to 40.67 Å. This gives rise to a narrowing of the distance between c-Si and a-Si:H slabs. Then, in the quenching process, in which the system is heated up to 1100 K and cooled down to 300 K, it is observed a displacement of the hydrogen atoms toward the interfaces. To quantify this effect, we inspect the hydrogen atoms near to the c-Si side setting a distance threshold $d_t=3.4$ Å (two times the distance cut-off for Si-H pairs). In Figure 25, the green atoms are hydrogen with distances to the c-Si sides less than d_t . In the ab initio relaxed system there are 5 hydrogen atoms that increase to 7 at the end of the MD minimization at $T=0$ K. During the quenching, the hydrogen atoms move toward the interfaces increasing the number of green atoms up to 22. This value is also maintained in the following thermalisation at $T=300$ K. This effect is also highlighted in the graphs above each configuration, in which the distributions of the hydrogen atoms along the z direction are reported. In the first two distributions there are two peaks at about 18 Å and 30 Å and a well-defined minimum in the middle. On the contrary, in the last two graphs, it is observed a more uniform distributions: the two peaks are reduced to fill the empty space in the middle and to form new peaks at the borders in correspondence with the interfaces. In Figure 60 in Annex B, we show the snapshots at $t=10$ ns of the c-Si/a-Si:H interfaces at $T=300, 500, 700$ and 900 K. The distribution of the hydrogen atoms along the z direction are reported for comparison. We found that the number of hydrogen atoms present at the interface is 22 at $T=300$ K and 23 at $T=500, 700$ and 900 K, so that it is almost constant at all the temperatures investigated.

In conclusion the ab initio relaxed system with a double a-Si:H/c-Si interface has been used as starting configuration for MD simulations. The minimization at $T=0$ K produces a shrinking of the system along the z direction of about 6 Å. The quenching process, in which the temperature raised up to 1100 K and cooled down to the thermalisation temperature, has been repeated for each of the thermalisation temperatures of $T=300, 500, 700$ and 900 K. We have observed a displacement of the hydrogen atoms towards the interfaces. This trend is maintained during each of the subsequent thermalisations of 10 ns. Indeed the concentration of hydrogen atoms near the interfaces remains almost constant for all temperatures.

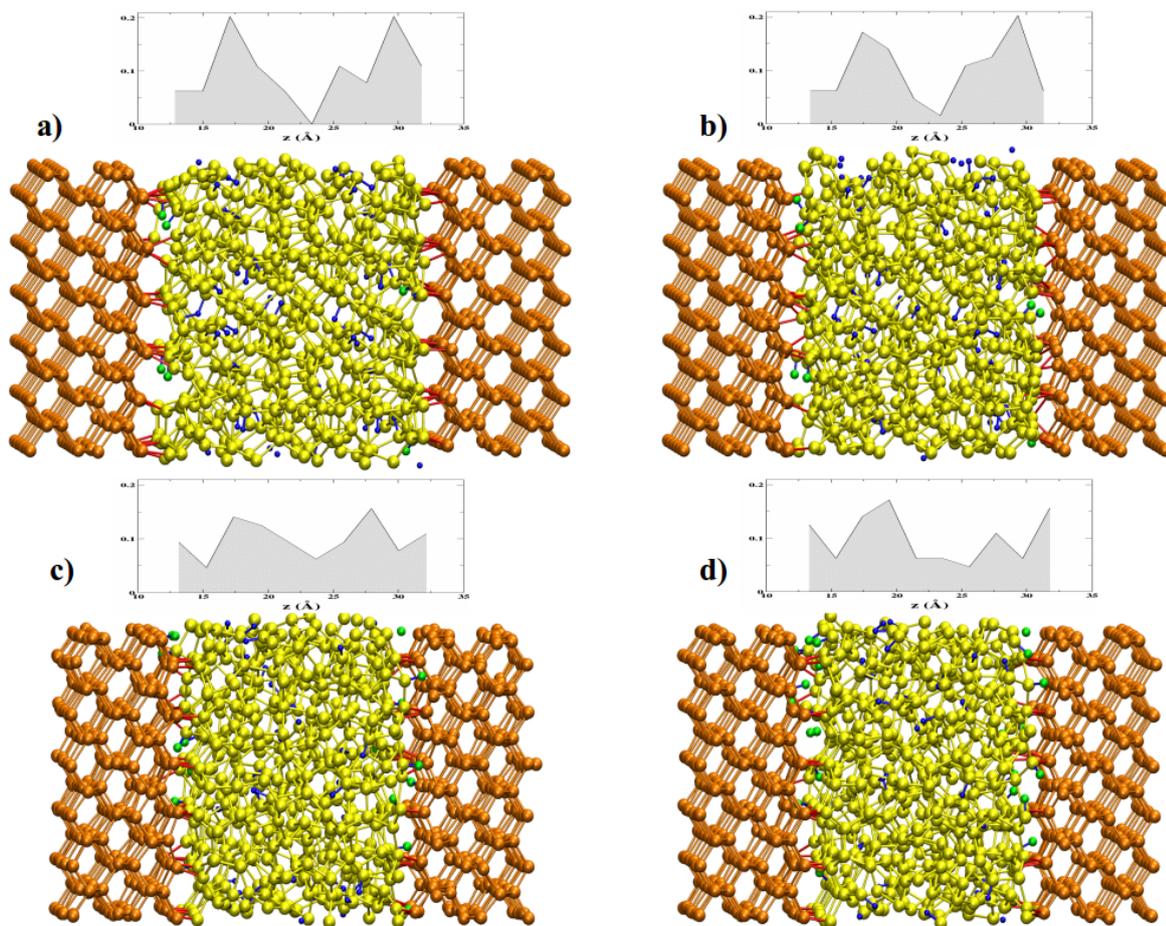


Figure 25: Snapshots of the a-Si:H/c-Si interfaces. a) The ab initio relaxed system; b) the system at the end of the minimization at $T = 0$ K; c) the system in the middle of the quenching process at $T = 1100$ K; d) the system at the end of the quenching process. The silicon atoms and their bonds are in orange in the c-Si side and are in yellow in the a-Si:H side, hydrogen atoms and bonds with silicon atoms are in blue. Bonds between c-Si and a-Si are in red. The hydrogen atoms close to the interfaces (with a distance less than 3.4 \AA to the c-Si sides) are in green. Above the configurations, the distribution of the hydrogen atoms along the z direction are reported.

T1.3.1-2: First-principles electronic properties of c-Si/a-Si:H interface

Methods. The two a-Si:H/c-Si interfaces are made of hydrogenated amorphous silicon (a-Si:H) between two crystalline silicon (c-Si) slabs. The relaxed $p(2 \times 1)$ symmetric reconstruction of the Si(001) surface constitutes the c-Si side of the interface. It is formed by 576 Si atoms, 16 layers of silicon with 36 atoms each. The a-Si:H side of the system is generated by cutting the a-Si:H structure, built as in reference [20], such that the surface area is equal to the c-Si side and the thickness is about 16 Å. It is composed of 512 Si atoms and 64 H atoms. The total length of the periodic cell is $L_z = 46.44$ Å, while in the x- and y-direction the system has $L_x = L_y = 23.22$ Å. Periodic boundary conditions are imposed in all directions. Therefore the a-Si:H/c-Si interface structures constituted by 1,152 atoms. A similar study has been performed in a full ab initio approach by some of us [19] for a smaller a-Si:H/c-Si interface formed by 336 atoms. The geometry of the ab initio relaxed structure is shown in Figure 1a: we use the PWscf (Plane-Wave Self-Consistent Field) code of the Quantum ESPRESSO suite [36, 35] to relax the a-Si:H/c-Si system. Si and H ultrasoft pseudopotentials with Perdew-Burke-Ernzerhof (PBE) [62] approximant GGA exchange-correlation potential, available in the Quantum ESPRESSO library. The electronic wave functions were expanded in a plane-wave basis set with a kinetic energy cut-off equal to 40 Ry (the charge density cut-off was 240 Ry). The Brillouin zone integration for the self-consistent calculation is restricted to the Γ -point, which is justified by the sufficiently large super cell. Gaussian smearing of 0.08 Ry is needed to reach convergence due to defect states at the Fermi level. All the parameters is chosen by checking the convergence of the total energy of the system. The energy minimization is performed by using conjugate gradient (CG) minimization energy method, with the convergence threshold for self-consistency equal to 10^{-6} Ry. Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-newton algorithm is used to perform geometry optimization. Ionic relaxation is stopped when both the following conditions is satisfied: energy changes less than 10^{-4} Ry between two consecutive self-consistent field (SCF) steps and all components of all forces were smaller than 5×10^{-4} Ry/Bohr. Then, this relaxed system is used as starting configuration for energy minimization and classical MD simulations. Snapshots of the thermodynamic properties of the system as well as per atom dynamic values (coordinates, velocities, forces, etc.) are taken at 1 ps intervals for DFT DOS post-processing. The non-self-consistent calculation of the electronic states is performed on a $2 \times 2 \times 1$ k-point grid, which was found to yield a sufficiently accurate representation of the relevant quantities (density of states, electron localization function, and charge density). In Figures 60 the configurations of the system at the end of the thermalisation process at 300, 500, 700 and 900 K are shown.

Scientific results. In Figure 26 we show the projected density of states (PDOS) of c-Si and a-Si:H ab initio relaxed, representing the PDOS of the a-Si:H/c-Si interfaces at $T = 0$ K that will be used as comparison. We see that c-Si has a gap of around 0.8 eV, below the experimental value of 1.1 eV [16]. It is well known that standard DFT [61] underestimates band gaps, due to the incomplete description of many-body effects. However, in this study the focus is on the formation of the intragap states related to the defects rather than on the evaluation of the band gap itself. From the PDOS in Figure 26b we see that broad peaks are induced from defects in a-Si:H bulk and at the interfaces.

We analyze the time evolution of the electronic properties of the a-Si:H/c-Si interfaces to monitor how intragap states change during the equilibration process after the quenching. We follow the time evolution of the projected density of states (PDOS) during the equilibration process starting from $t = 0$ ns, when the process begins until $t = 10$ ns. The PDOS during the equilibration process at 300, 500, 700 and 900 K are shown, respectively, in Figures 27, 61, 62 and 63 for c-Si and in Figures 28, 64, 65 and 66 for a-Si:H.

We observe at the start of the equilibration ($t = 0$ ns) that a dense concentration of peaks are found in the energy gap for the a-Si PDOS. Those peaks are related to defects at both the interface and in the a-Si:H bulk. In particular, intense peaks are found nearby the Fermi energy level at 0 eV. As the equilibration progresses, the energy of the PDOS peaks changes in energy and the corresponding intensity changes too, both in the intragap range and outside of it. Overall, after the ReaxFF MD annealing, quenching and equilibration we have a decrease of the density of the defects compared to that of the starting DFT relaxed a-Si:H/c-Si interfaces. In particular, the decrease of the PDOS related to intragap states is more evident

D1.2 First scientific results and progress on exascale tools

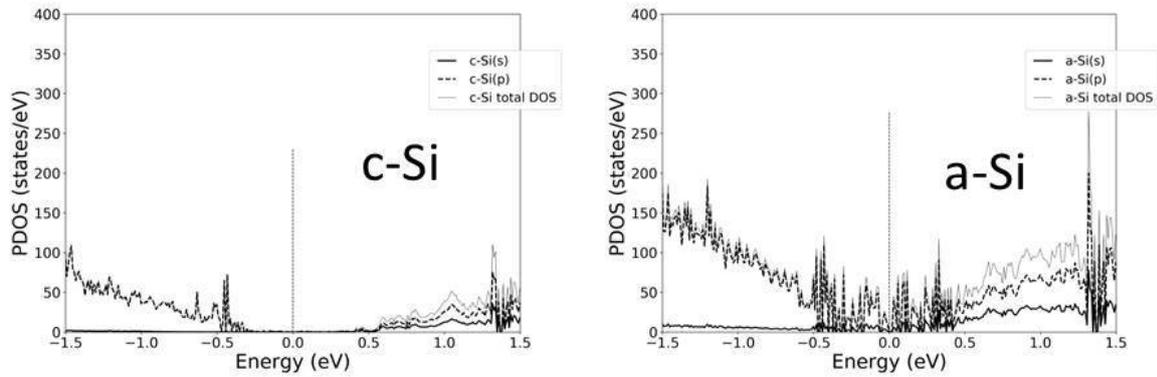


Figure 26: Projected density of states of crystalline and amorphous silicon of the ab initio relaxed a-Si:H/c-Si interfaces. The vertical dashed line at 0 eV evidences the Fermi energy.

c-Si, T = 300 K

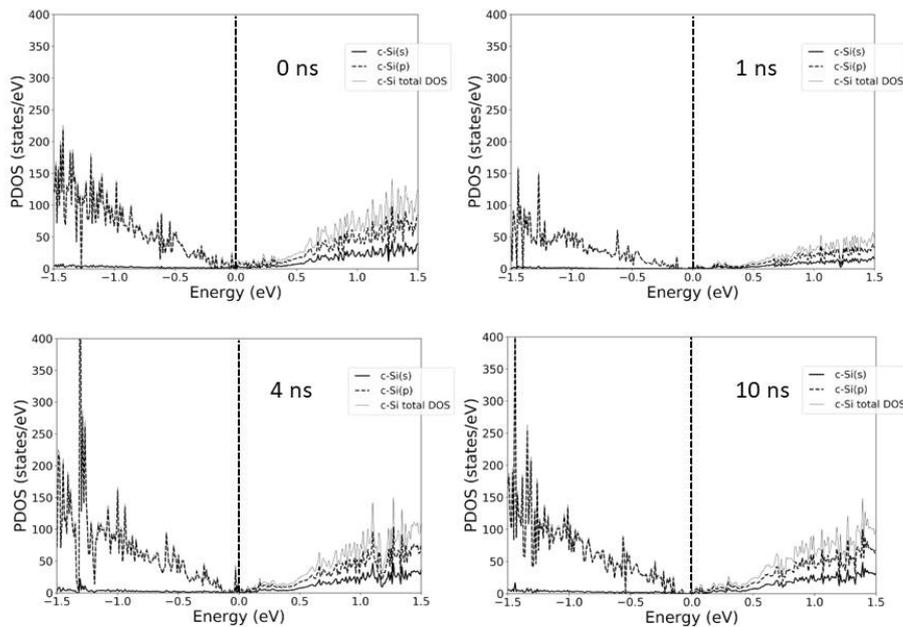


Figure 27: Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 300 K. The vertical dashed line at 0 eV evidences the Fermi energy.

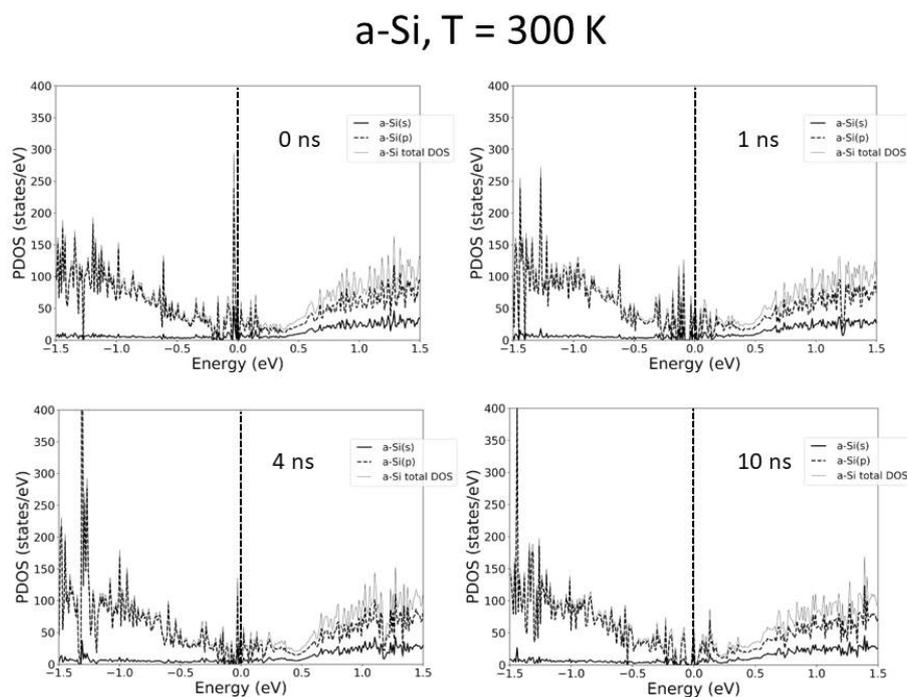


Figure 28: Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 300 K. The vertical dashed line at 0 eV evidences the Fermi energy.

for the thermalization at 900 K. To gain a deeper understanding of the structural properties, a coordination analysis of the Si atoms is performed. A geometrical criterion is used to identify the nearest neighbors in the coordination analysis, applying a distance cutoff of 2.85 Å and 1.7 Å for Si-Si pairs and for Si-H pairs, respectively. Concerning the $t= 10$ ns snapshot at T= 300 K, it is observed that the average number of neighbors of Si atoms is 4.01. In detail, 16 Si atoms have threefold coordination (1.4%), 1044 Si atoms have fourfold coordination (96.0%) and the remaining 28 Si atoms have fivefold coordination (2.6%). Just small variations have been found during the equilibration process taking into account that the average coordination number equals to 4.01 for the $t= 0$ ns snapshot too. Therefore, after the quenching process the system is quite ordered with a high percentage of fourfold coordinated Si atoms, and it keeps this condition during thermalisation at room temperature. By increasing the thermalisation temperature the number of fourfold coordinated Si has a monotonous increase until to 1054 atoms (96.9%) at 900 K. The fractions of threefold and fivefold coordinated Si are the same at the end of the thermalisation at 900 K with 1.55% of the total number of atoms each one, and the average coordination is 4.00.

In order to elucidate the spatial localization we calculated the local DOS (LDOS) of the intragap energy levels. This analysis allows to visualize both the localization and the DOS intensity of the energy levels in the band gap of c-Si. In Figures 29, 67, 68 and 69, we compare the LDOS at $t = 0$ and 10 ns for the different temperatures. The isosurfaces are taken for the same value of LDOS (20% of the maximum for the atomic configuration on the start of the thermalisation at room temperature) so that the intensity of localization can be compared. Following the above criteria for atomic distances, we investigated the coordination of the atoms nearby the LDOS isosurfaces and we found they are defects three- and five-fold coordinated. We found that intragap defect states are localized both in the bulk of a-Si and at the a-Si:H/c-Si interface. Intragap defect states can be formed in the few c-Si layers nearest to a-Si:H. However, we see a change in the distribution of intragap defect states in the thermalisation and for the different temperatures. Some intragap defect states present at room temperature disappear by increasing T and new ones can appear. For example, the LDOS isosurface at the left of a-Si:H visible at 300, 500 and 700 K in Figures 29b, 67b

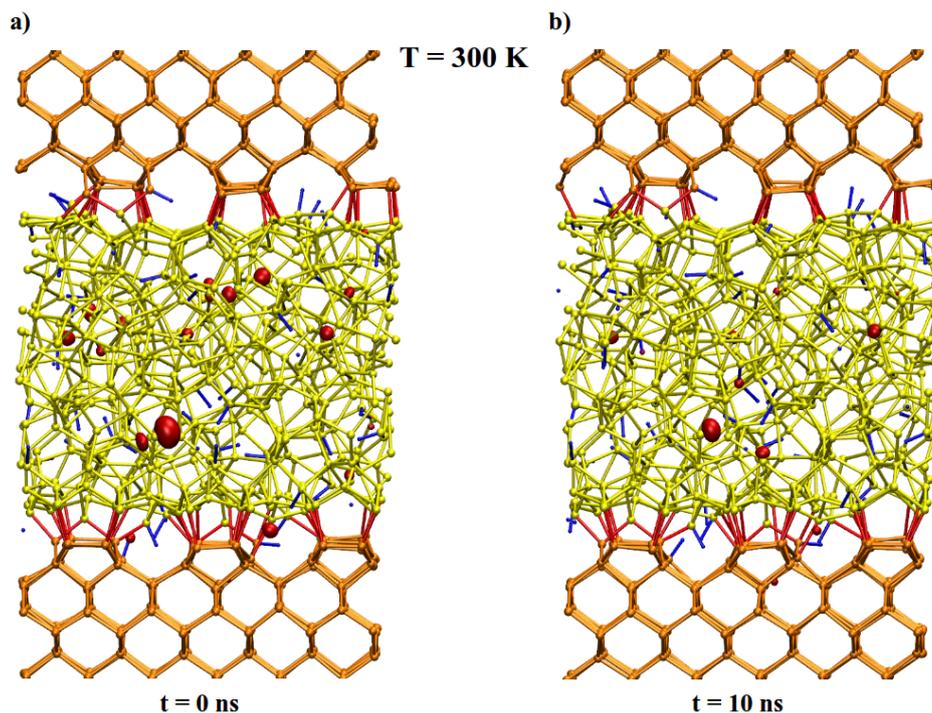


Figure 29: Local density of states of the intragap states of the a-Si:H/c-Si interfaces at the beginning a) and at the end b) of the thermalisation at $T = 300$ K.

and 68b disappears while a new LDOS isosurface appears on the top of c-Si at 900 K in Figure 69b. In general, the intensity of the LDOS decreases with the thermalisation. This finding is in agreement with the observation that the number and the intensity of the intragap states in the PDOS at the end of the equilibration process are lower than at the start.

We calculated the energy potential averaged in cross section along the direction perpendicular to the interfaces. We show the potential profiles after the different thermalisations in Figure 30. The averages are over small (1.0 \AA) and (5.0 \AA) large spatial interval, so that the barrier is calculated as the difference between the local maximum (for electrons) or minimum (for holes) of the small interval average at the interface, and the large interval average in the middle of c-Si and a-Si:H bulk regions. The profile of the average potential has small variations along all the times of the simulation for all the temperatures investigated. However, we found that the energy barrier at the interface decreases by increasing the temperature. Indeed, the barrier for electrons (holes) is about 6.0 (2.5) eV at $T = 300$ K, and 5.3 (2.0) eV at $T = 900$ K. We have investigated the charge transfer by calculating the difference of the charge density between the total system, and the c-Si and a-Si:H systems considered as isolated slabs, shown in Figure 31. We found that electron charge is accumulated along the c-Si/aSi:H interface and depleted from the nearby c-Si and a-Si:H surfaces. The two opposite pointing dipoles that are formed give rise to the potential profile showing the asymmetric barriers for electrons and holes.

In conclusion we combined ReaxFF MD simulations and ab initio calculations to investigate the time evolution of the intra-gap states of a large a-Si:H/c-Si interface system. Therefore, the electronic structure is calculated and analyzed with a focus on the identification and characterization of the intragap states, which have a crucial impact on the device performance due to their role as recombination centers. Throughout the annealing process, we monitor the evolution of the relevant structural and electronic properties. We have found that at the end of the equilibration process of 10 ns of the PDOS related to intragap states is decreased as well as the number of the electronic states into the gap. Nonetheless, the defects states are still localized both in the bulk of a-Si that at the interface with c-Si, until to be formed in the few c-Si

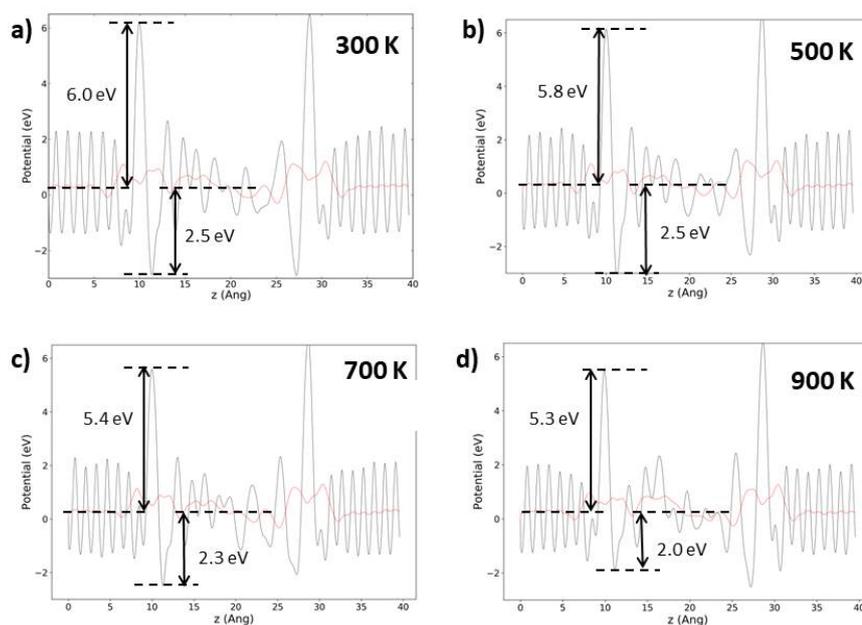


Figure 30: Energy potential averaged over the xy-plane at the end of the a) 300 K, b) 500 K, c) 700 K, d) 900 K equilibration. The energy barriers at the interface are shown. The black (red) curves are the average over a spatial interval of 1.0 (5.0) Å.

layers nearest to a-Si:H. However, the system is quite ordered after the quenching process with a high percentage of fourfold coordinated Si atoms, and it keeps this condition during the final thermalisation. By increasing the thermalisation temperature the number of fourfold coordinated Si has a monotonous increase, with a decrease of the defects. We found that electron charge is accumulated along the c-Si/aSi:H interface while it is depleted from the nearby c-Si and a-Si:H surfaces. The two opposite pointing dipoles that are formed give rise to different barriers to the each type of carriers favoring the hole transport while hindering the transport of electrons. However, the energy barrier at the interface decreases by increasing the thermalisation temperature.

This research will allow to select the atomic configurations that will be used in the electronic transport model, in order to design more efficient silicon solar cells based on the silicon heterojunction technology.

T1.3.1-3: Non-equilibrium Green's functions transport properties c-Si/a-Si:H interface

In order to study the role of defects at the disordered interfaces for energy applications, the tools developed in this subtask will be applied primarily to c-Si/a-Si/Si systems, as outlined in the roadmap report D1.1. The overall goal is to compute local Wannier functions for these systems and then run the transport calculations on the related Hamiltonians. As such the aim of this task can be separated in two parts: 1) Wannier Hamiltonian generation from the DFT configurations generated in T1.3.1-1 and T1.3.1-2 and 2) quantum transport simulations using these hamiltonian as inputs to libNEGF.

For the first part, a dedicated code for the generation of Wannier function Hamiltonian already exists as part of Quantum Espresso suite. Unfortunately, it became evident that even for the relatively small structures illustrated in T1.3.1-2, the Wannier functions generation is already quite an expensive task. A dedicated new code interfacing Wannier90 and Hamiltonian generation is necessary in order to accelerate the generation of the Wannier functions Hamiltonians. Since Wannier functions generation from DFT inputs is under the competence of the MaX-II CoE, a new agreement of collaboration between EoCoE-II and

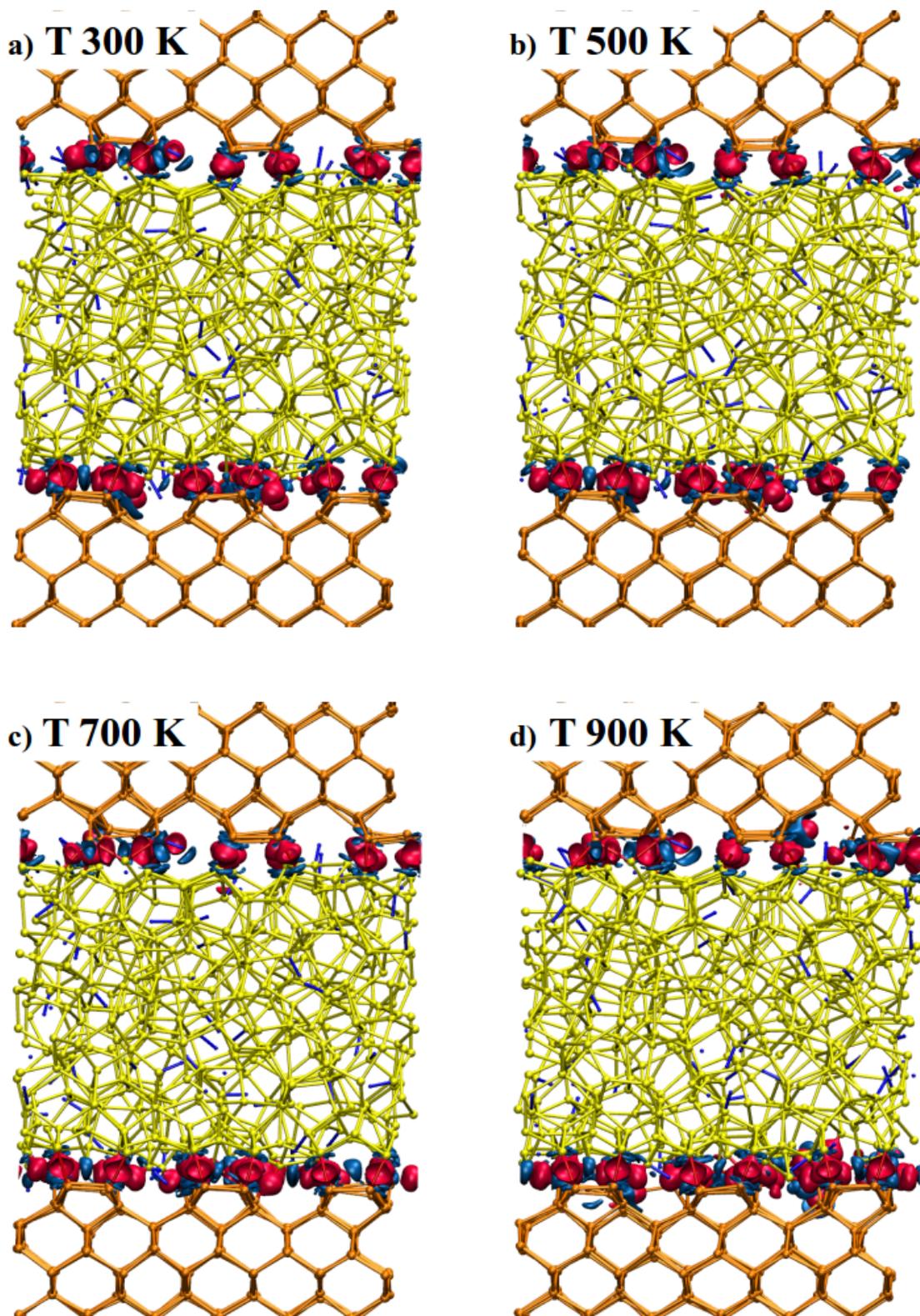


Figure 31: The c-Si/a-Si:H interface and the difference of the charge density between the total system and the c-Si and a-Si:H systems considered as isolated slabs at the end of the various thermalisations. Red (blue) isosurface is the positive (negative) difference of the charge density.

D1.2 First scientific results and progress on exascale tools

MaX-II has been recently drafted to carry on the development of such new optimized code. Under such an agreement, MaX-II will optimize and further parallelize the existing software, while EoCoE-II will provide test cases and validate the new code against such cases.

To obviate the lack of a performant routine for the generation of the Hamiltonians which are the input for the quantum transport simulation, we studied the same and larger structures using the Density Functional Tight Binding semi-empirical approach [27]. In DFTB all Hamiltonian matrix elements between atomic orbitals are computed from first principles DFT, however neglecting the three center integrals allows for a very simple compilation of look-up tables making the Hamiltonian generation as fast as in empirical TB. Yet the accuracy of the approach has been proven over several decades of developments. For the calculations shown below a specially tuned parameterization of Si has been developed that well reproduces the full bulk bandstructure [52]. Because the libNEGF library has been already interfaced to the dftb+ code (<https://www.github.com/dftbplus/dftbplus.git>), this approach has the clear advantage of producing preliminary results that we will be made more accurate when Wannier functions Hamiltonians will be available in the near future.

In Figure 32 we report the transmission function computed at temperature $T=0$ for the interface system. The calculation is obtained with a 4×4 k-point sampling mesh in the direction transverse to transport. A sharp drop of transmission can be observed within the Si bandgap and is due to the gap of the ideal Si crystal. The sharp peaks within the gap are connected to the local density of states in the a-Si layer. This plot should be compared to the temperature dependent transmissions shown in Figure 33.

For each line in the graph, we have considered the frozen geometry given by the last step obtained in the MD simulations at different temperatures, hence at $t=10$ ns. Thermal noise in the atomic positions considerably increases the gap density of states, leading to a clear increase of transmission, compared to the more ordered c-Si phase obtained at 0 K. This increase is likely to be connected to the disorder in the c-Si contacting leads, consistent with the results of Figures 27, 61, 62 and 63. On the other hand also the amorphous layer sees an increase of PDOS with temperatures (Figures 64, 65 and 66), but a decrease of PDOS at midgap. This well correlates with the decrease of transmission around the Fermi energy (-3.5 eV) that can be observed at 900 K and after 10 ns of annealing.

The disorder is also responsible for a sizeable increase of transmission both in the conduction and in the valence bands, albeit the latter is certainly less pronounced. This increase of transmission is directly related to an increase of percolation pathways in the a-Si that could be related to an increase of PDOS (see Figure 28), since disorder in the leads always reduces the transmission probability of the perfect crystal.

These studies are just a demonstration example of the capability of the combined MD/DFT/NEGF tools. The final goal is to include electron-photon and electron-phonon coupling via the related self-energies in the Green's functions. In the current status libNEGF can include elastic electron-phonon scattering and simple corrections to the transmission probability due to absorptions/emissions of photons. The implementation of inelastic electron/phonon scattering (e.g., optical phonons), on the other hand, requires important code improvements in the parallel communications in order to tackle problems of realistic size. This work will be carried out in the next months also within the WP2.

The resulting new tool (neXGf) will have several other applications beyond the scope of the project. The applications of the NEGF techniques have been extensively applied in recent years in quantum optics, quantum corrections to the Boltzmann transport equation, high field transport in bulk systems and electron transport through nano-scaled devices. For instance, the possibility to simulate structures with lateral dimensions of 5-10 nm will open the possibility to study non trivial interfaces, alloy disorder, stacking faults and in general defects that typically require large supercells in order to properly include the effects of charge localizations and charged defects. The problem is relevant both in modelling exact broadening of the PL spectrum as well as considering correctly injection and well to well charge transfer, including defect-mediated tunneling. In the emerging field of 2D materials, neXGf will be able to describe out of plane transport properties of stacked multilayers, where coherent transport might still play an important role.

In conclusion we apply transmission calculations to individual snapshots of ReaxFF MD simulations in order to investigate how intra-gap states of a large a-Si:H/c-Si interface couple to bulk Si and how coherent

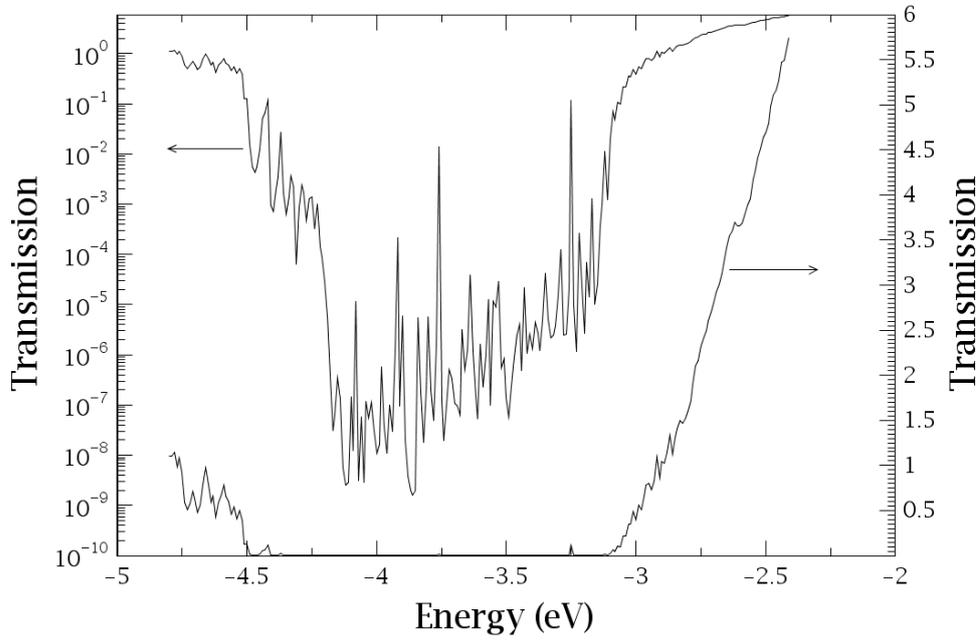


Figure 32: Transmission plot for the c-Si/a-Si:H/c-Si interface at T=0

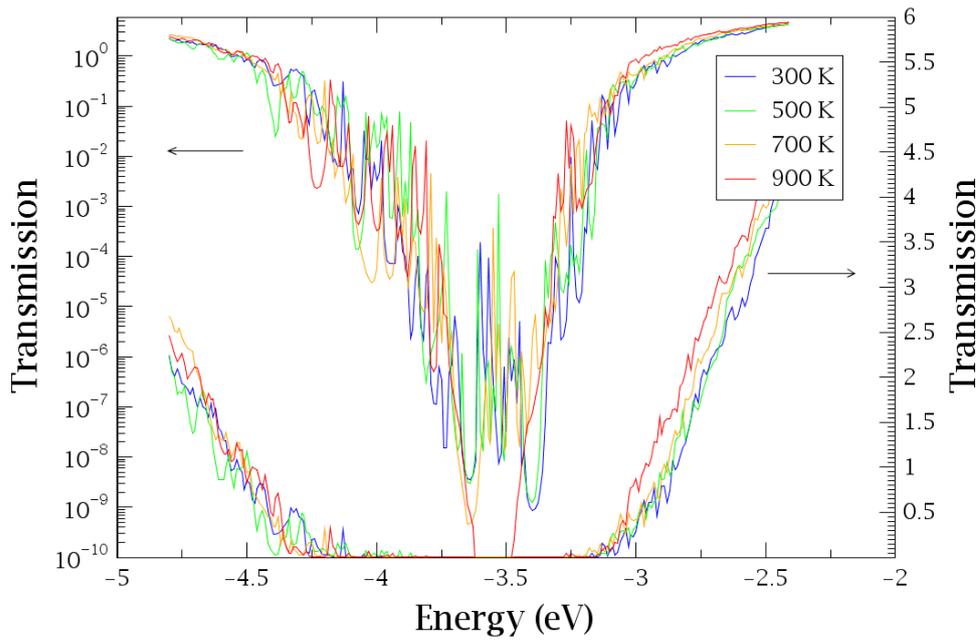


Figure 33: Transmission plot for the c-Si/a-Si:H/c-Si interface at increasing temperatures from T=300 to 900 K.

D1.2 First scientific results and progress on exascale tools

states travel across the amorphous layer. These states have a crucial impact on the device performance due to their role as recombination centers. Throughout the annealing process, we monitor the evolution of the relevant structural and electronic properties and found that during the annealing process the PDOS in the gap decreases.

5.2 Task 1.3.2: Scientific results

The main task T1.3.2 is subdivided in four subtasks:

T1.3.2-1	Electronic structure calculations. QMC reference calculations will be performed on a series of reference systems. A planar graphite electrode and/or a coronene molecule will be put in contact with several adsorbed molecules: Water, sodium, chloride, 1-ethyl-3-methyl-imidazolium and tetrafluoroborate. Then we will perform DFT calculations on the same systems in order to benchmark the various functionals available (either in the gradient-generalized approach, such as BLYP or PBE, or the more costly hybrid functionals such as HSE06). Once the best functional is chosen, we will perform a large series of DFT calculations on larger systems containing several layers of water or ionic liquids. These will produce a reference data set of forces, energies, and multipoles for T1.3.2-2.
T1.3.2-2	Force-field developments. The simulations of T1.3.2-3 and T1.3.2-4 require a realistic representation of the interatomic interactions, but on large systems, which puts them outside the range of what is currently feasible using brute-force DFT. We will therefore introduce physically-motivated model potentials for the interactions, in which additional degrees of freedom are introduced to account for the response of the electronic structure of the molecules and the electrode to their changing environments, namely induced dipoles for the former and atomic charge fluctuations for the latter. These potentials will be parametrized by fitting the reference data set gathered in T1.3.2-1 using a generalized force and multipole-fitting procedure that is now well-established.
T1.3.2-3	Capacitances of carbon materials for blue energy production. The force fields develop in T1.3.2-2 will be used in large-scale simulations. Molecular dynamics (Metalwalls) and molecular density functional theory (MDFT) techniques will be used in this task. MDFT will allow to screen a large amount of carbon electrode materials. The most promising ones will then be thoroughly studied using Metalwalls in order to determine precisely the capacitance of the device as well as the charging time, the structure of the adsorbed species, etc.
T1.3.2-4	Seebeck coefficients for redox active species in thermo-electrochemical cells. The force fields developed in T1.3.2.2 will be used to study redox active species, such as ferrocene, dissolved in ionic liquids. By using Metalwalls, we will determine the free energy profile for electron transfer in the bulk liquid and in the vicinity of electrodes. This will provide the redox potential; by varying the temperature we will be able to extract the Seebeck coefficient for ranking the systems in terms of performance for thermo-electrochemical devices.

T1.3.2-1: Electronic structure calculations

In principle, the quantities of interest in T1.3.2-3 and T.1.3.2-4 (adsorption energies, capacitances, structure of the adsorbed species, etc.) can be computed using either DFT or classical MD. Although DFT is exact

D1.2 First scientific results and progress on exascale tools

in theory, in practice it relies on approximations of the exchange-correlation (XC) functional: a plethora of these functionals are available in every DFT code. In the specific case of carbon surfaces, many DFT studies have been reported in the past. Nevertheless, the rigorousness of this approach has been questioned in two recent works involving a more fundamental correlated-electron method, i.e. Quantum Monte Carlo (QMC) [15]. Firstly, Al-Hamdani *et al.* have studied the adsorption of water molecules on carbon nanostructures using a wide range of XC functionals [6]. They have shown that the energy varies significantly from one functional to another, which strongly limits the predictive power of DFT for such systems. This is mostly due to the approximated treatment of the dispersion interaction, even when using the van der Waals functionals or the ad hoc dispersion correction (D3) proposed by Grimme and co-workers [37]. Shin *et al.* have reached a similar conclusion in their QMC study of O₂ adsorption on the graphene surface [80]. They obtain an adsorption energy in very good agreement with the experimental value and show that the adsorption site predicted differs from the one obtained in previous DFT calculations.

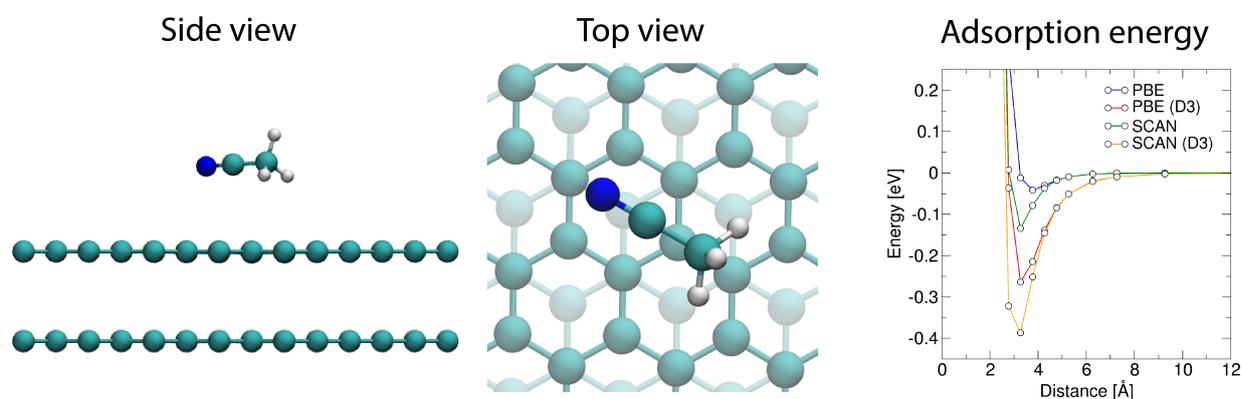


Figure 34: Preliminary DFT calculations performed for the adsorption of an acetonitrile molecule on graphite. The two first panels show a side and a top view of the system, and the third panel gives a comparison of the adsorption energy with various XC functionals.

In order to evaluate the importance of the XC choice for our specific case, we calculated the adsorption of acetonitrile on a surface made of two graphene layers with PBE and SCAN, which are two of the currently most employed functionals. We also did additional calculations adding the D3 correction, for which the results are summarized in Figure 34. It is clear that the difference between each choice is very large, and that dispersion interactions play an important role. For our systems of interest, which are made of the organic solvents and the various ions (Li⁺, Na⁺, but also organic species) involved in blue energy production and thermo-electrochemical cells, there is no experimental data to discriminate between the various XC functionals.

We therefore want to perform a large series of benchmark QMC calculations to study the interaction between energy storage electrolyte components and carbon surfaces. A proposal was submitted on the 20th PRACE Project Access Call to perform these calculations, but it was rejected. We have therefore submitted an INCITE proposal. In the meantime, preliminary QMC computations for the adsorption of an acetonitrile molecule on a carbon substrate were performed. The physical system studied is the same of the one as in the DFT calculations (2x5x5 carbon unit cell, with a single acetonitrile molecule on top), with the position of the atoms determined by DFT relaxation. Our QMC simulations were performed using the QMCPack software [30].

These computations included:

- a preliminary DFT calculation to determine the orbitals to be included in the trial wave function; for

D1.2 First scientific results and progress on exascale tools

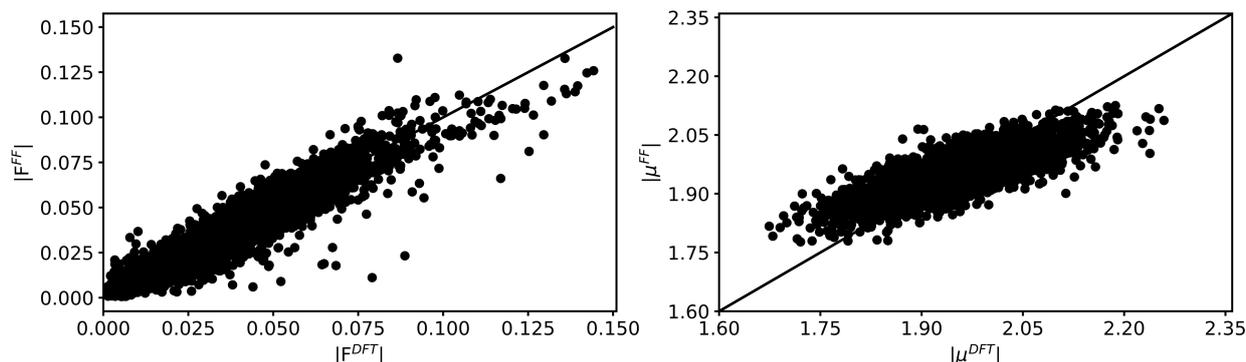


Figure 35: Fitted forces (left) and dipoles (right) with respect to the DFT values, for a series of configurations of anthraquinone and TEMPO redox-active molecules dissolved in acetonitrile. Each point corresponds to one atom of the configuration.

the DFT computations we used the Quantum Espresso software [29]

- the QMC optimization of the trial wave function;
- Variational and Fixed Node Diffusion Monte Carlo calculations of the total energy of the system.

The energy estimates were computed using twist averaged boundary conditions [50], using 8 different twists, to an accuracy of the mHa. These computations were done on the GPU-accelerated partition of the Idris Jean-Zay machine, using a total of about 30,000 core-hours. We plan to continue these computations by moving the acetonitrile molecule by changing its distance from the carbon substrate, to determine the full adsorption curve, using the additional CPU hours we have on Jean-Zay as well as the computer time we have on the Cineca Marconi100 machine. Additional QMC calculations will be performed on other molecules and ions if the INCITE project is awarded, otherwise we will have to rely on the results concerning the acetonitrile for the benchmark.

T1.3.2-2: Force-field developments

Although we need QMC results to start developing the force fields for the interactions between carbon surfaces and the liquid, we already started to parameterize them for the liquid species (for which the generic DFT functionals are usually accurate enough). For the moment, we have developed polarizable force fields for a series of redox species (anthraquinone and TEMPO) in acetonitrile. To do this we fit the forces and the dipoles obtained from DFT for several configurations, the results of such fits are shown on Figure 35.

We are now validating the force fields by extensive comparison with experimental data. An example of validation for the spectroscopic properties is shown on Figure 36, where we compare the simulated infrared spectrum to the experimental one. An excellent agreement is obtained. Preliminary results for the density, the viscosity, the diffusion coefficients are also good, but we need to gather more statistics in the next few weeks.

T1.3.2-3 Capacitances of carbon materials for blue energy production

This subtask was not started yet.

T1.3.2-4 Seebeck coefficients for redox active species in thermo-electrochemical cells

This subtask will involve the force fields developed in T1.3.2-2. Nevertheless the computation of the relevant quantities is generally made through a thermodynamic integrations schemes which was proven accurate in DFT-based MD. Nevertheless, in the case of classical MD, it is not clear that the method, which

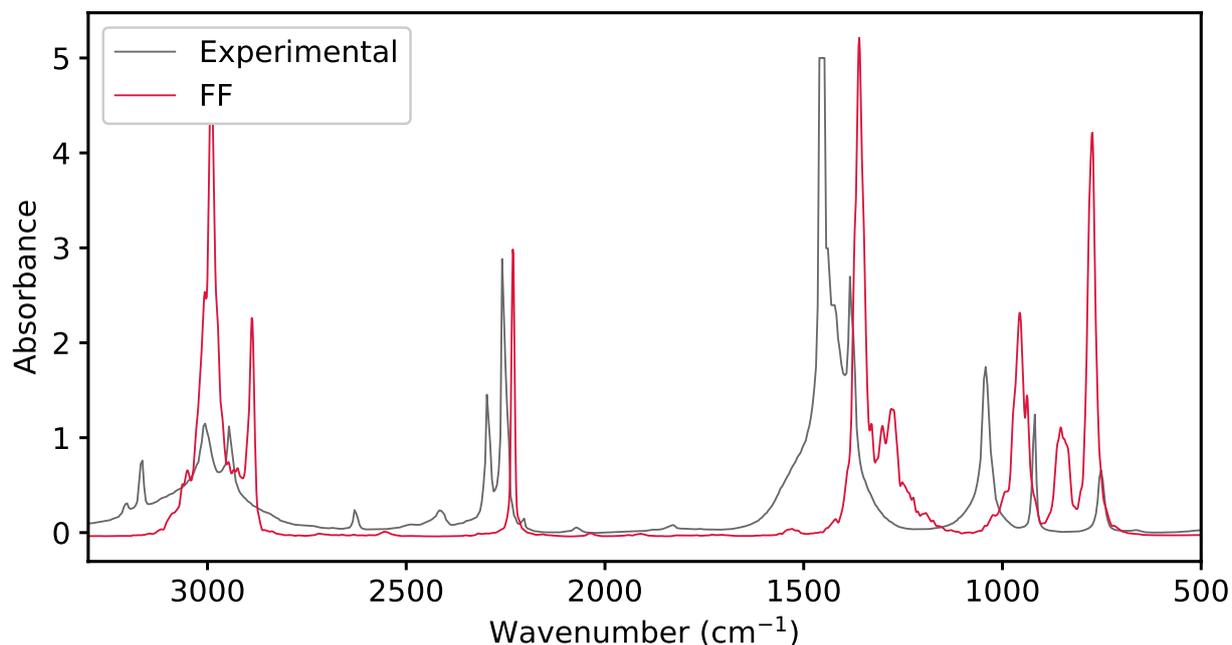


Figure 36: Comparison of the simulated and experimental infrared spectra for acetonitrile.

involves the calculation of the vertical energy gap (VEG) between the reduced and the oxidized species during the simulation, will be accurate enough to predict electrochemical properties (e.g. redox potentials). If not, an option would be to compute the VEG for hundreds of structures along the classical trajectory obtained from each molecular dynamics simulation using DFT.

Nevertheless, in a previous study performed in the laboratory, we have already obtained such informations using a full DFT-based approach [68]. We would therefore like to try whether it will be possible to use this data using Machine learning (ML) tools such as deep neural networks. The latter are quickly becoming integrated into research of complex chemical systems due to growing access to ever-improving computing resources and accessibility of codes. Deep neural networks are an ideal tool to capture complex, non-linear relationships between observable features and measurable outputs. Thus, it is worthwhile to explore the feasibility of integrating such tools into problems such as computing electrochemical properties to reduce computational costs in order to explore larger and more complicated systems.

Initial investigations have used the PyTorch library to generate a deep neural network (general architecture shown in Figure 37A), confirming that machine learning tools can indeed be extended to our investigation of redox potentials of biredox molecules. Through systematic modification of the hyperparameters associated with the neural network, five-hundred structures never seen before were passed through the model to predict their vertical energy gaps. Figure 37B shows in green a histogram associated with the ML-prediction of the VEGs, fit with a Gaussian (red). The resulting distribution computed via the deep neural network is encouragingly close to the distribution that would be expected using DFT calculations (blue).

With a tool such as the trained model we have developed, we are optimistic about the potential to integrate into our research of complex systems. Additional development will likely only improve the rigorous of the model. Future steps will investigate the transferability of the model to predict the vertical energy gaps based on new chemical environments (e.g. changes of temperature to probe the Seebeck coefficient). Additionally, we will explore the distributions generated by structures taken from classical molecular dynamics simulations. A schematic in Figure 37C suggest how with careful development, it can even be imagined to refine classical forcefields to improve predicted structures which ultimately may generate distributions of VEGs in greater agreement with DFT calculations.

D1.2 First scientific results and progress on exascale tools

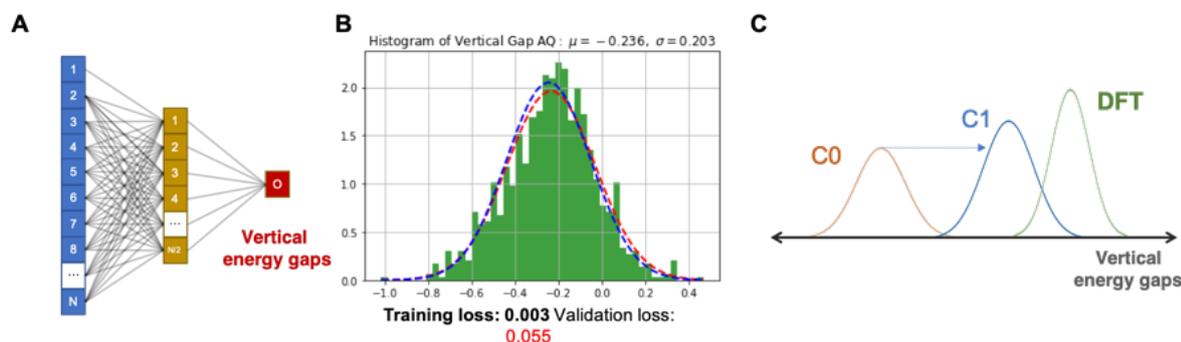


Figure 37: A) the architecture of the initial artificial neural network used to relate molecular dynamics simulation data to the vertical energy gap. B) A histogram of the vertical energy gaps predicted for five-hundred structures. A Gaussian fit of the machine learned predictions (red) is compared to a Gaussian fit of DFT-computed vertical energies for the same structures. C) Future direction where we propose to investigate whether classical dynamic simulations can generate structures coupled to the deep neural network to reproduce a distribution similar to that which would be generated using DFT. Different classical force fields will likely generate different distributions based on the phase space they sample.

5.3 Task 1.3.3: Scientific results

The main task T1.3.3 is subdivided in four subtasks:

T1.3.3-1	Development of KMC simulator. Algorithmic optimisation of the KMC simulator.
T1.3.3-2	Application of KMC simulator to electron transport in organic materials. Preliminary results applying the KMC simulator to the study of doped organic semiconductors have been published in a paper demonstrating the scaling of the electrostatics library [76]. A full scientific paper based on the study is currently in preparation; a summary of the study is presented separately below.
T1.3.3-3	Development of DMC simulator. Development of Monte Carlo simulator, which solves the Boltzmann transport equation, using the PPM Python framework.
T1.3.3-4	Application of DMC simulator to electron transport in in perovskites. We have a paper currently under review addressing polaronic effects in halide perovskites. Details of the results are given below.

T1.3.3-1: Development of KMC simulator

Algorithmic optimisation has been carried out on the KMC simulator such that the electrostatic potential due to static charges is calculated just once at the beginning of a simulation, rather than at every re-evaluation of the electrostatic energy over its course. This significantly decreases the time required for simulations involving ionised dopants, for example. We have also implemented Dirichlet boundary conditions in the electrostatic solvers so that the effects of metallic contacts may be included in the simulation. Furthermore, algorithmic developments have been implemented so as to enable injection and extraction events to take place from and to those contacts, and for creation and annihilation of electron-hole pairs due to photoexcitation and radiative recombination in the bulk of the material.

T1.3.3-2: Application of KMC simulator to electron transport in organic materials

The use of doping in organic semiconductors to provide additional charge carriers is a fundamental means of controlling organic device characteristics. The doping efficiency, i.e. the ratio of free charge carrier density to dopant density, deduced from current-voltage characteristics, is in general not equal to the probability of dopant ionisation, contrary to assumptions made on the basis of conventional semi-conductor theory. To understand the possible origins for this discrepancy, we studied the microscopic kinetics of charge carriers in the organic semiconductor α -N,N'-bis(1-naphthalene)-1,1'-biphenyl 4-4"-diamine (α NPD) doped by with varying densities of the molecule 2,2'-(perfluoronaphthalene- 2,6-diylidene)dimalononitrile (F6TNNQ) in which we assume 100% dopant ionisation. We used the KMC simulator developed as part of T1.3.3-1, to model dopant concentrations down to 0.01% in an off-lattice model that accounts for amorphous molecular packing, which allowed us to investigate up to $\sim 20,000$ charge carriers in a system containing 510300 molecular sites.

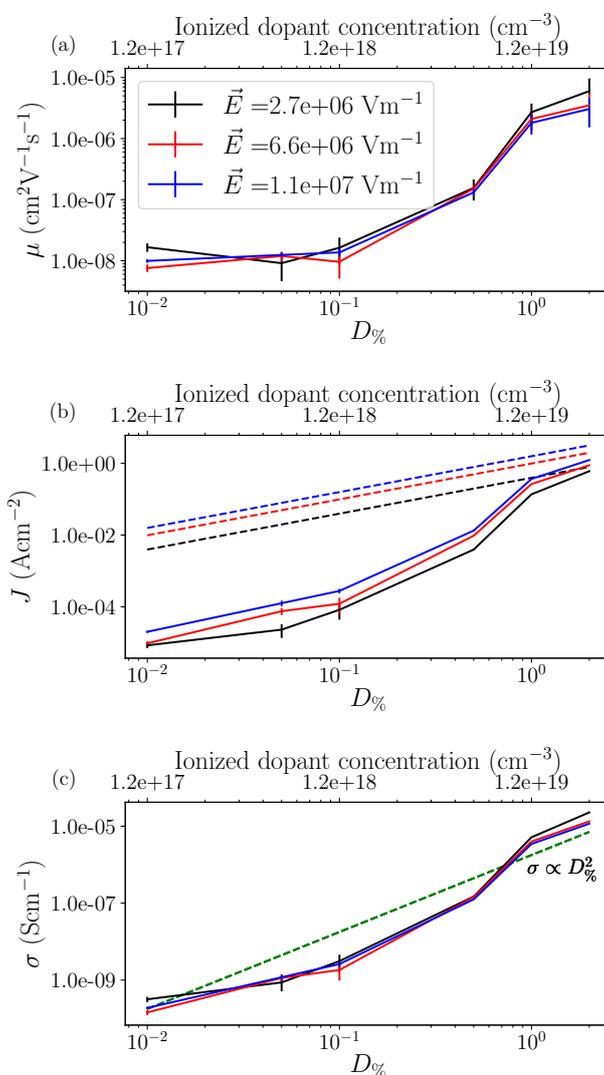


Figure 38: Quantities calculated during kMC simulations of the doped system against $D_{\%}$, the doping percentage, for different magnitudes of the applied electric field. (a) The hole mobility increases exponentially with $D_{\%}$. (b) The current density also increases exponentially with the doping percentage, the dashed lines represent a linear scaling due to ρ at a constant mobility of $7.8 \times 10^{-10} \text{ Acm}^{-2}$. (c) The conductivity collapses for different values of \vec{E} , showing that the field has a linear effect on current density. The dashed line is a square power law to guide the eye.

Our results show that the electrostatic energy minima created by ionized dopants have a large effect on the dynamics of charge carriers. More specifically, the energy scale set by the depth of the electrostatic basins induced by the ionised dopants is much greater than the applied field, and that the electronic current is therefore limited by the timescale on which the electrons pass from one basin to another. In turn, that timescale is determined by the dopants proximity to one another, which determines the barrier height of the basins, such that the conductivity exhibits a highly non-linear dependence on the dopant concentration, as shown in Fig. 38

T1.3.3-3: Development of DMC simulator

We have built a device Monte Carlo simulator, which solves the Boltzmann transport equation, using the PPM Python framework. It includes support for multiple band-structure minima (referred to as valleys in the literature); scattering from point defects (such as vacancies and impurities), and acoustic and optical phonons; and polaronic effects. It is built in a modular fashion such that additional stochastic events (due to additional scattering mechanisms, for example) can be added with minimal changes to the main code. Further developments will include the implementation of carrier–carrier scattering in order to study the timescale of thermalisation following photoexcitation, and the inclusion of a possible hot phonon distribution to allow for studies of the effects of a phonon bottleneck on carrier cooling. We also plan to implement ion motion in a drift–diffusion simulator, which will then be coupled with the DMC simulator in a 1D device model; a proof-of-concept 1D drift–diffusion model for electrons has also been implemented using the Firedrake Python framework [67]. However, this might not be possible due to delays caused by COVID-19.

T1.3.3-4: Application of DMC simulator to electron transport in in perovskites

Metal–halide perovskites are currently attracting intense interest for application in optoelectronic devices, due to their remarkable efficiencies in spite of low temperature synthesis methods. One hypothesis that has gained considerable traction in the community is that many of their favourable phenomenological parameters are due to the formation of large polarons. In order to test that hypothesis, we derived expressions for scattering of polarons by ionised impurities and acoustic phonons, and implemented them in the DMC simulator developed in T1.3.3-3, amongst other features, in order to compute and compare the temperature dependent mobilities of polarons and band electrons in three metal–halide perovskites.

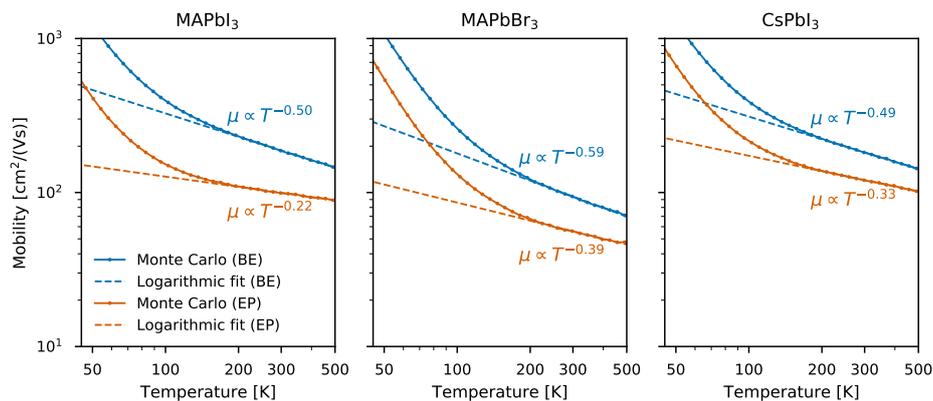


Figure 39: Computed temperature dependent mobilities of band electrons (BE) and electron polarons (PE) in three metal–halide perovskites. Logarithmic fits to the high temperature data are also shown.

The central results of our study are shown in Fig. 39. We observe that the main effect of polaron formation is to reduce the mobility of electrons, but the temperature dependence is not strongly altered. Our study therefore challenges the growing consensus that polaron formation plays a significant role in electron dynamics in metal–halide perovskites.

5.4 Code demonstrators

libNEGF → neXGf

The main motivation behind the joint effort within EoCoE is to fill the gap of available tools for quantum transport simulations on large supercomputing facilities, especially in the perspective of exascale computing facilities. The NEGF formalism is a highly computational intensive method that provides an excellent example of exa-scale applications. Scaling of the method up to 100,000 cores have been demonstrated, at least within OMEN [87], thanks to 3 levels of parallelism obtained by distributing \mathbf{k} -points, energy-points and a domain decomposition. Current peak performance reached in OMEN is 15 Pflops/s that has been recently pushed to 85 Pflops/s in a novel datacentric implementation [87]. The goal set on EoCoE is to increase the level of parallelization in order to approach and surpass the achievements reached by OMEN and pave the road towards an exascale-capable code.

The challenges to be solved are multiples, especially concerning communication bottlenecks that require smart strategies of data distribution. A notoriously problematic bottleneck of the NEGF method, both in terms of computation as well as memory consumption, is the scaling with lateral supercell size. Indeed the construction of the Green's functions when expressed on local orbitals (LCAO or Wannier functions) requires inverting dense matrices that scale with the size of the system cross-section.

The iterative (or recursive) algorithm used in libNEGF scales as $\sim \mathcal{O}(m^3M)$, where m is the size of the system cross-section and M is the number of partitions along the transport direction. Alternative strategies in order to reduce the rather heavy m^3 scaling to m^2 are in the process to be adopted, in particular the selected inversion algorithm or sub-partitioning of the layer. Important step forward have been already achieved in this project by improving parallel SMP cooperation efficiencies on the inversion step (See D2.2). Further strategies that will be considered include GPU acceleration and tensor products on GPU since it is clear that exascale machines will be heavily based on such devices.

The demonstrator developed so far is based on a multilevel modelling approach to describing the physical properties of the c-Si/a-Si interfaces. Structural properties are simulated using classical force-field MD using LAMMPS, a freely available package developed by Sandia Labs. Electronic properties are computed at Density Functional level of theory using Quantum Espresso, another freely available package (<https://www.quantum-espresso.org>), which is part of the software packages managed and optimized by the MaX-II CoE.

Local orthogonal Wannier functions are generated from the plane-wave calculations using the community code Wannier90 (ArXiv:1907.09788v1). Benchmarks have been performed on the JURECA cluster showing a relatively efficient strong scaling. On our systems, comprising 1152 atoms, this step is still rather expensive, limiting the applicability to few MD snapshots. The generated Hamiltonians can be given in input to a new code under development, WanTrans, that is interfacing libNEGF to solving the transport equations.

Currently, since WanTrans has not been completed, the transport calculations have been performed using the Density Functional Tight-Binding approach [27]. The current release of dftb+ is available on github (<https://www.github.com/dftbplus/dftbplus.git>) and already includes libNEGF interface. The reasons for using dftb+ are twofold. From the scientific point of view it provides a well-grounded way of building a transferable tight-binding parameterization of the c-Si/a-Si:H interface, allowing to scale up to systems beyond the size/time constraints of DFT/Wannier. In this way it provides a cheap way to get semi-quantitative trends of the electronic properties of such interfaces.

On the other hand, since the number of local TB orbitals can be mapped to the reduced Wannier representation (e.g., sp^3 localized states), we have a direct indication about the cost of libNEGF calculations, with a direct measure of its performance. In practice it allows us to develop and improve independently the library and the WanTrans code for the final demonstrator.

The systems included in the demonstrator are those reported in the calculations for Fig 33, including different MD snapshots obtained at increasing temperatures. The output geometries of the MD steps require some processing in order to be usable in the transport code. This step is needed since the NEGF

D1.2 First scientific results and progress on exascale tools

formalism is based on a system partitioning into central region and contacts. The latter provide the open boundary conditions for the transport calculation. This atomic reordering step is now easily accomplished with tools provided in the `dftbplus` package (`setupgeom`). The provided geometry files in the demonstrator have been already prepared for transport.

The batch file for the slurm queue system are included in the demonstrator folder. The crucial quantities to set are the resources requested in terms on number of nodes and threads per node. A mixed OpenMP/MPI parallelism is used. Energy and **k**-points are distributed over different MPI processes. Matrix algebra, including inversions and MM multiplications, exploits SMP parallelism implemented in the intel MKL library. The load balance is currently taken care by the user. In this case the calculations involve 10 independent **k**-points and 240 energy points. Hence the calculations are distributed over 25 nodes, each running 4 processes and 12 threads per process for a total 1200 cores of the JUWELS cluster in the Jülich Supercomputing Centre. This is a preliminary result on a small system with only ballistic transport. Once the new functionalities allowing for optical phonon interaction and increased parallelism will be in place we expect at least 10+ fold increase in parallel resources use.

QMCPack

QMCPack is one of the most efficient codes for Quantum Monte Carlo simulations, allowing to choose between a large variety of QMC methods with fast, highly scalable and parallelizable implementations. The computational cost of QMC simulations is proportional to (i) the size of the system to be simulated, i.e. the cube of the number of the valence electrons N^3 , and (ii) linear in the number of random walkers used in the sampling. Among the advantages of QMC methods with respect to other computation methods in quantum chemistry are a more favorable scaling with respect to system size, lesser memory requirements and a naturally parallel structure, which can be used to improve performance when a large number of parallel nodes are available. To this end QMCPack uses a fully hybrid (OpenMP,CUDA)/MPI parallelization scheme, to optimize memory usage and take advantage of the large number of cores on modern supercomputers. With the emergence and increasing relevance of GPU acceleration QMCPack is subject to constant updating to maximize its performance, with special attention to new equilibration schemes. The possibility of sharing the information concerning single particle orbitals over several nodes also allows to alleviate the problem of having only a limited amount of memory for each GPU. This code development is beyond the scope of our project, and performed for the most part by different research groups in several National Laboratories in the United States (Oak Ridge, Sandia, Argonne, Livermore).

The scaling of QMCPack has been object of extensive study; analysis performed on the Titan and Mira Blue Gene supercomputers have shown that QMCPack has an essentially ideal scaling even when executed on a large number (up to the order of 10^4) of either GPU or CPU nodes [28], as shown in figure 40. Such analyses were made performing DMC simulation of a system of crystalline NiO with 128 atoms, corresponding to 1536 valence electrons, using a total of $\simeq 5 \cdot 10^5$ walkers. All the code and data required to analyze the performance of the latest version QMCPack code is provided with the installation files, which can be downloaded from the QMCPack git repository <https://github.com/QMCPACK/qmcpack>.

At the current time we are waiting to have access to computer time to the Summit machine and to take advantage of a Grand Challenges project on the Idris Jean-Zay machine, which have a similar architecture to Titan and Mira Blue Gene and thus will ensure a similar ideal scaling.

We emphasize that several different physical systems can be simulated essentially independently at the same time on the same machine: this means that QMC studies have a modular nature, and it is easy to set up a workflow to make an efficient use of computational resources depending on their availability, i.e. it is possible to efficiently use all the nodes of a supercomputer or only a small fraction. To this end the Nexus Python library is distributed along with QMCPack; Nexus allows an easy management of all the phases of a QMC study, from the DFT generation of single particle orbitals to the wave function optimization to the DMC computation of physical quantities.

D1.2 First scientific results and progress on exascale tools

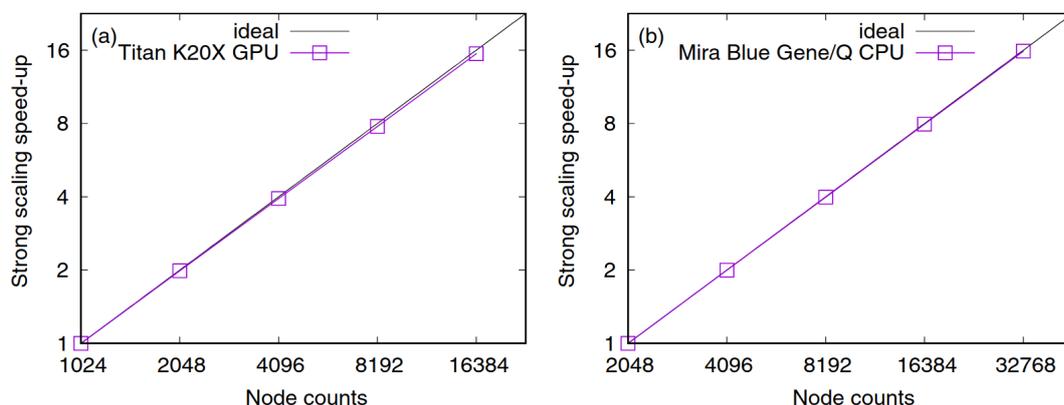


Figure 40: Parallel scaling of QMCPACK on GPUs and CPUs for DMC calculations of a NiO 128 atom cell with 1536 valence electrons. Titan nodes have a single GPU each, and these runs used 512 000 total walkers. Each Blue Gene Q node has a 16 core processor, and these runs used 458 752 total walkers. Figure taken from [28].

KMC/DMC

Our electrostatic solver for KMC enables access to systems that contain millions of charges and exhibits good parallel scaling. We demonstrated good strong and weak scaling on Intel and ARM platforms on 128 nodes. The largest test system used for scaling tests contained 1.28×10^8 charges. Our KMC code, that uses our electrostatic solver, was subsequently applied to a doped organic semi-conductor containing 2×10^4 charges, as described above. The previous state of the art code, using approximate electrostatics, could perform KMC simulations with systems containing less than 100 charges. Hence our developments enable further study into physically meaningful sized systems with accurate electrostatics in a manner previously not deemed possible.

5.5 Roadmap revision

T1.3.1

The timeline of this task has changed. Subtasks T1.3.1-1 and T1.3.1-2 are on time and did not experienced any large delay. T1.3.1-3 went through a radical change of code and experienced delays due to the pregnancy of one of the main developers at FZJ, Dr. Aguilera. Part of the work has been redirected and steered so as to keep making progress towards the goal of the main task. The Wannier function generation of the Hamiltonian is part of a starting collaboration between our CoE and MaX-II CoE.

The change of code from PVnegf to libNEGF needed a substantial replanning of the entire task. Targets that were part of PVnegf development had to be scrapped and substituted by alternative targets which are needed for libNEGF. As described in the D2.2 report, libNEGF is a full 3-dimensional (3D) multi-band code. This means that time planned to bring PVnegf to 3D and multi-band can now be devoted to add functionalities which were already present in PVnegf and are still in a development stage in libNEGF, such as non-ballistic scattering between charge carriers and phonons. Moreover, the search for the new flagship code introduced a substantial shift in the entire subtask. Despite the delay introduced, the higher level of development and maturity of the libNEGF code will in part compensate for the time lost. Moreover the team has now acquired a more dynamical and vibrant partnership. This is demonstrated by the quick progress made in the last 3 months of the project. We also expect Dr. Aguilera to come back from maternity leave at the end of 2020. This will give the project an additional boost which should eventually compensate for the time lost.

The new timeline of the project extends it as shown in the pink-colored areas in the timeline table for T1.3.1. The main target in this revised timeline are the optimization and parallelization of the Wannier90 kernels in

D1.2 First scientific results and progress on exascale tools

collaboration with MaX-II CoE and the further development of interaction terms which will greatly increase the computational load of the flagship code and pave the way for the use of large computational resources in the near-future largest clusters. In particular we plan to execute full quantum transport simulations using the entire JUWELS-Booster cluster (70 PFLOPS) which will be operative by the end of 2020. Such target will be reached by developing the libNEGF code in a new library ported on multi GPU devices. This target will be reached by a collaborative effort interl to the Jülich Supercomputing Centre. The final pre-exascale code will be renamed neXGf.

MATERIALS T1.3.1	Project Months											
		D1.1	MS1	D1.2				D1.3				
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.3.1-1												
T1.3.1-2												
T1.3.1-3												

T1.3.2

The first subtask has been delayed because our PRACE proposal was not accepted. Nevertheless, the second subtask could be started in advance by focusing on bulk systems first, for which the QMC calculations were not needed. Subtask 3 will start now as initially planned, and Subtask 4 was also started ahead of schedule.

MATERIALS T1.3.2	Project Months											
		D1.1	MS1	D1.2				D1.3				
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.3.2-1												
T1.3.2-2												
T1.3.2-3												
T1.3.2-4												

T1.3.3

The roadmap has undergone significant revision since D1.1. This was due to the unexpected continuation of the employment of Dr. Will Saunders in the group at Bath. Dr. Saunders is the main architect of the PPMD library, which underpins UBAH's contributions to the EoCoE-II project, and his continued participation has enabled, among other things, further development and optimisation of the KMC code base which significantly improves their performance and expands the scope of the scientific applications. Therefore, while the roadmap specified in D1.1 was focused mainly on development of the DMC simulator, the revised road map reflects a more equal division of effort between the development and applications of the KMC (T1.3.3-1 and T1.3.3-2) and DMC (T1.3.3-3 and T1.3.3-4) simulators. Since Dr. Saunders will leave the group at the end of August 2020, the remaining efforts on KMC will focus on applications to transport in organic devices. Development of the DMC simulator will continue along with applications to pervoskite devices.

D1.2 First scientific results and progress on exascale tools

MATERIALS T1.3.3	Project Months											
		D1.1	MS1			D1.2			D1.3			
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.3.3-1												
T1.3.3-2												
T1.3.3-3												
T1.3.3-4												

6 Water for Energy (T1.4)

T1.4 is divided in five main tasks:

T1.4.1	Exascale hyper-resolution hydrologic simulations
T1.4.2	Hydropower modelling in the Italian Alpine region
T1.4.3	Reduced model for high fidelity hydrodynamics simulation.
T1.4.4	Experimental design for geothermal modelling
T1.4.5	Modelling of geothermal reservoir system

6.1 Task 1.4.1: Scientific results

In hydrological applications, modeling of shallow subsurface flow is of major importance in order to accurately simulate and predict the exchange of groundwater with streams under low-flow conditions, and the transport of energy. The major challenge is the representation of topographically driven groundwater convergence and streamflow generation, and of the geological heterogeneity across a number of space scales ranging from centimeters to thousands of kilometers in case of continental river systems. Predicting hydrologic states and fluxes requires a balanced combination of computational power to run climatic and hydrological models at the proper scale and enough information concerning the hydropower systems. This is extremely challenging at the pan-European scale. For example, predicting water cycle processes for scientific and applied assessment of the terrestrial water cycle requires a high-resolution modeling framework on the order of 1-3km. Thus, HPC technologies are required to enable hyper-resolution simulations at continental scales. This was achieved by performing hydrological simulations using integrated Terrestrial Systems Modeling Platform (TerrSysMP; [81]), consisting of the three-dimensional surface-subsurface model ParFlow([47], and the Community Land Model, CLM3.5 (CLM;[58]). TerrSysMP is highly modular and scale-consistent and support highly scalable parallel applications. In EoCoE-I, ParFlow model has been setup at 3 km resolution over Europe. Using this setup, long-term (30 years) and high resolution hydrologic simulation using modernized version of ParFlow with GPU implementation will be performed in EoCoE-II. This would allow to accurately predict changes related to the climate events and its impact on the hydropower production.

The main task T1.4.1 is subdivided in three subtasks:

T1.4.1-1	Model Setup for hyper-resolution hydrologic simulations.
T1.4.1-2	Model Evaluation and validation.
T1.4.1-3	Uncertainty Quantification using Data Assimilation.

Task 1.4.1-1: Model Setup for hyper-resolution hydrologic simulations

The ParFlow and CLM models has been setup at a spatial resolution of 3 km over Europe. The meteorological forcing variables (variables such as temperature, precipitation, wind, vapor pressure, and downward longwave and shortwave radiation) were used as input to the model for the time period of the spinup and subsequent years simulated. In order to obtain an adequate representation of the natural variability, a higher spatial resolution is required in the vertical and lateral directions respectively. The inputs to the ParFlow model are the same as for the CLM model but with additional layers of bedrock geology going to a depth of 30 m is needed for the 3D groundwater simulations. The resolution of the 3km ParFlow-CLM model is 1592x1544x15. A spinup of nine years, by simulating the year 1997 nine times, was performed in order to obtain a stable and reasonable distribution of the initial state variables. The land surface static input data used in this work consist of topography, soil properties, plant functional types, and physiological vegetation parameters (Figure 41). Digital elevation model (DEM) data were acquired from the 1km Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) ([21]) as shown in Figure 41a. The land use data was based on the Moderate Resolution Imaging Spectroradiometer (MODIS) data set ([33]) (Figure 41c), where the land use types are transferred to Plant Functional Types (PFT). The properties of each of the sub-grid land fractions, such as the leaf area index, the stem area index, and the monthly heights of each PFT, were calculated based on the global CLM3.5 surface data set ([58]). To provide soil texture data in the model (Figure 41b and 41d), sand and clay percentages were prescribed based on pedotransfer functions from [77] for 19 soil classes derived from the FAO/UNESCO Digital Soil Map of the World ([9]). The high-resolution reanalysis COSMO-REA6 dataset ([11]) for the time period of 2000–2015 was used as the atmospheric forcing for CLM3.5 and ParFlow.

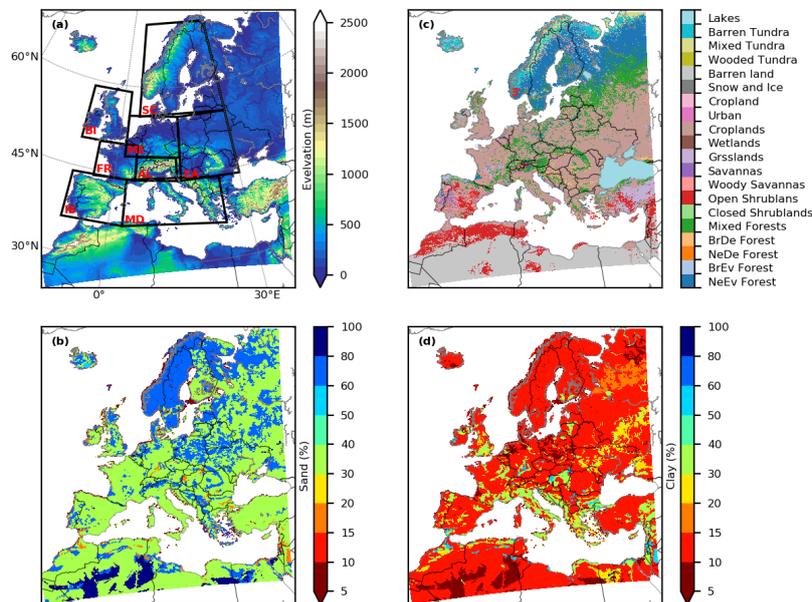


Figure 41: Model surface input data: a) GMTED2010 DEM, b) dominant land use type based on MODIS data, c) percent sand content, and d) percent clay content based on global FAO soil database. The inner boxes in (a) show the boundaries of the PRUDENCE regions (FR: France, ME: mid-Europe, SC: Scandinavia, EA: Eastern Europe, MD: Mediterranean, IP: Iberian Peninsula, BI: British Islands, AL: Alpine region; [17]. Adopted from [56].

D1.2 First scientific results and progress on exascale tools

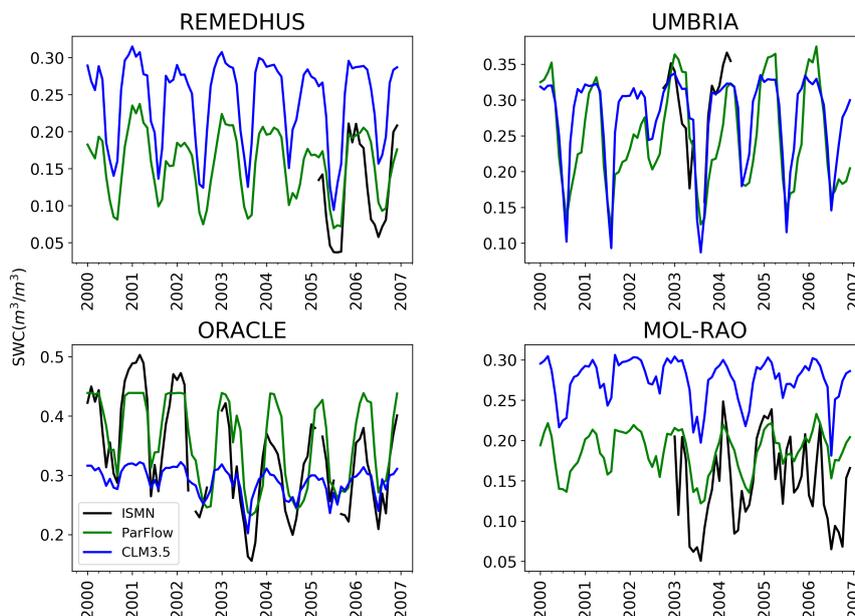


Figure 43: Comparison of SM simulated by ParFlow, CLM with in-situ observations in four ISMN soil moisture network for 2000–2006 for two networks REMEDHUS and SMOSMANIA located in Spain and France, respectively.

models captured well when compared with the GRACE data, However, both models show much stronger negative anomalies in summer over most regions (Figure 44).

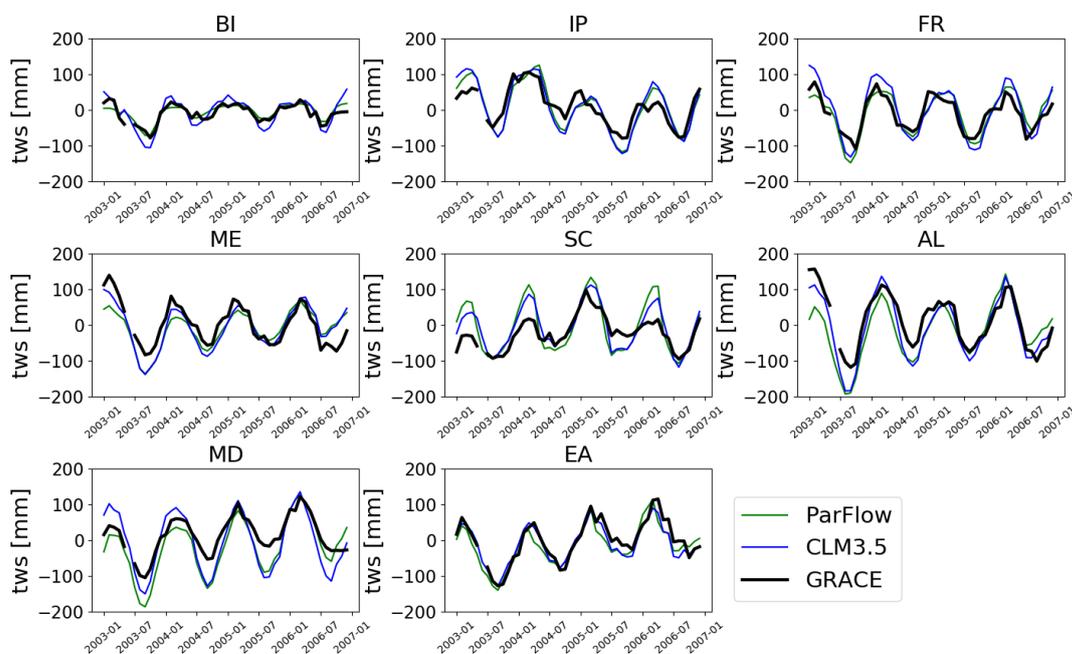


Figure 44: Comparison of the spatially aggregated total water storage (TWS) simulated by ParFlow, CLM with GRACE satellite dataset over PRUDENCE regions for 2003–2006) period.

Task 1.4.1-3: Data assimilation

To provide reliable hydrologic data and information for management of water resources, we used data assimilation (DA) techniques which combine observations and model to improve model estimates. In EoCoE-II, a series of data assimilation experiments were conducted using Parallel Data Assimilation Framework (PDAF; [57]) that has been coupled with CLM3.5 and ParFlow models ([48]). A key capability of CLM-PDAF is the support for data assimilation that combines land surface processes with satellite and in-situ observations for the estimation of optimal land surface states. The data assimilation structures in CLM-PDAF allows to directly ingest remotely sensed high resolution observations of land surface conditions to produce accurate, spatially and temporally consistent fields of land surface states, with reduced associated error. The CLM-PDAF is designed for high-performance computing infrastructures and can efficiently cope with the high computational burden of ensemble-based data assimilation.

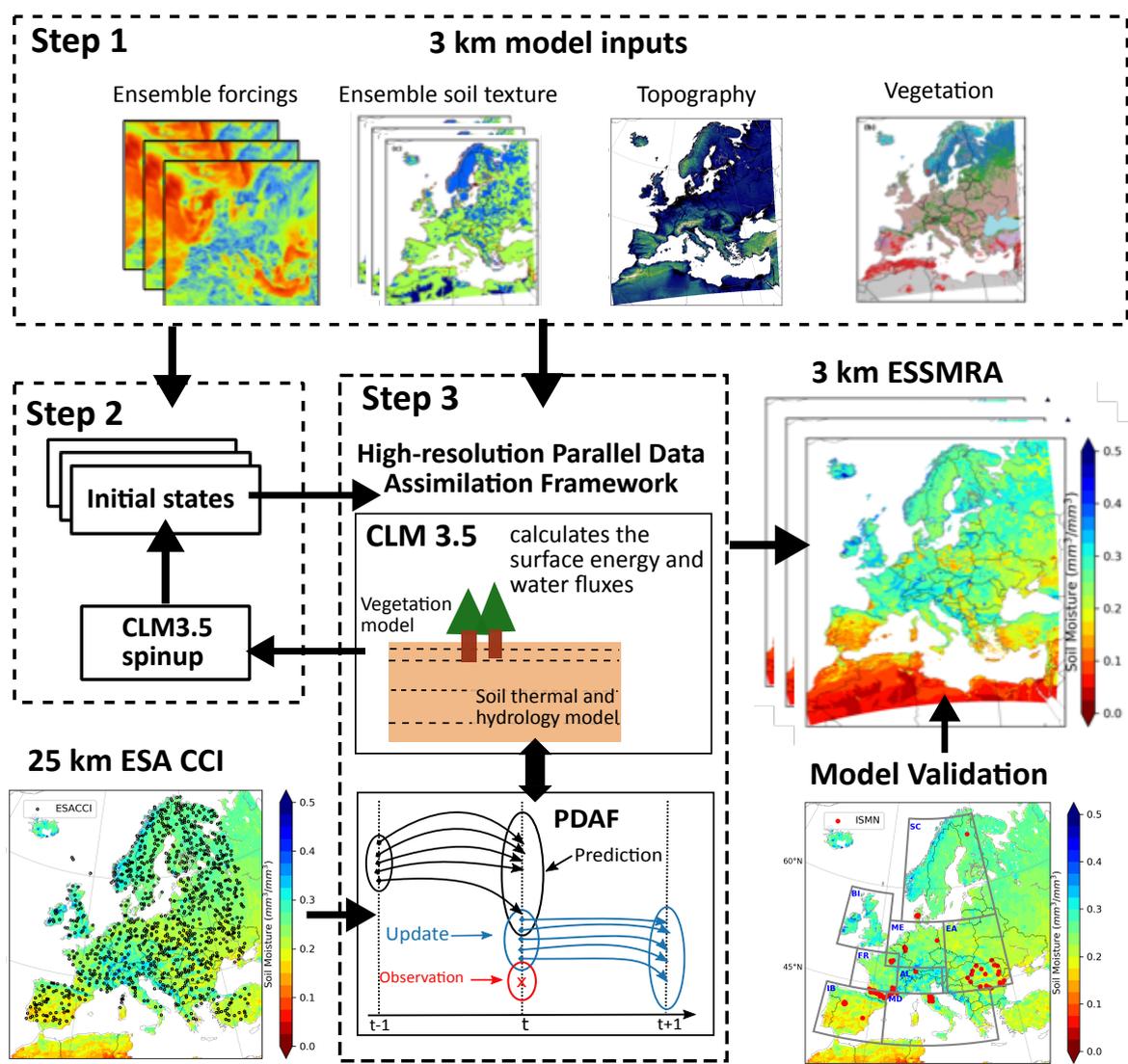


Figure 45: Schematic of CLM-PDAF workflow adopted to generate high resolution ESSMRA product. Adopted from [55].

We implemented CLM-PDAF to assimilate the satellite based soil moisture dataset ESACCI into CLM, producing a high-resolution European surface soil moisture reanalysis (called ESSMRA hereafter) dataset.

D1.2 First scientific results and progress on exascale tools

This product overcomes the shortcomings of sparse spatial and temporal distributions in observations and provides a better estimate of SM than obtained only by modeling or by sparse observations alone. The 3 km ESSMRA is generated by first implementing the regional land surface model setup coupled with data assimilation framework ([55];[56]) as shown in Figure 45. In the first step, the 3 km input data were prepared to implement CLM model for the EU-CORDEX domain. Second, CLM was initialized with the equilibrium initial state variables. In the third step, the model was run for 2000–2015 at hourly time step with the assimilation of ESACCI soil moisture data into the model once a day using the EnKF algorithm.

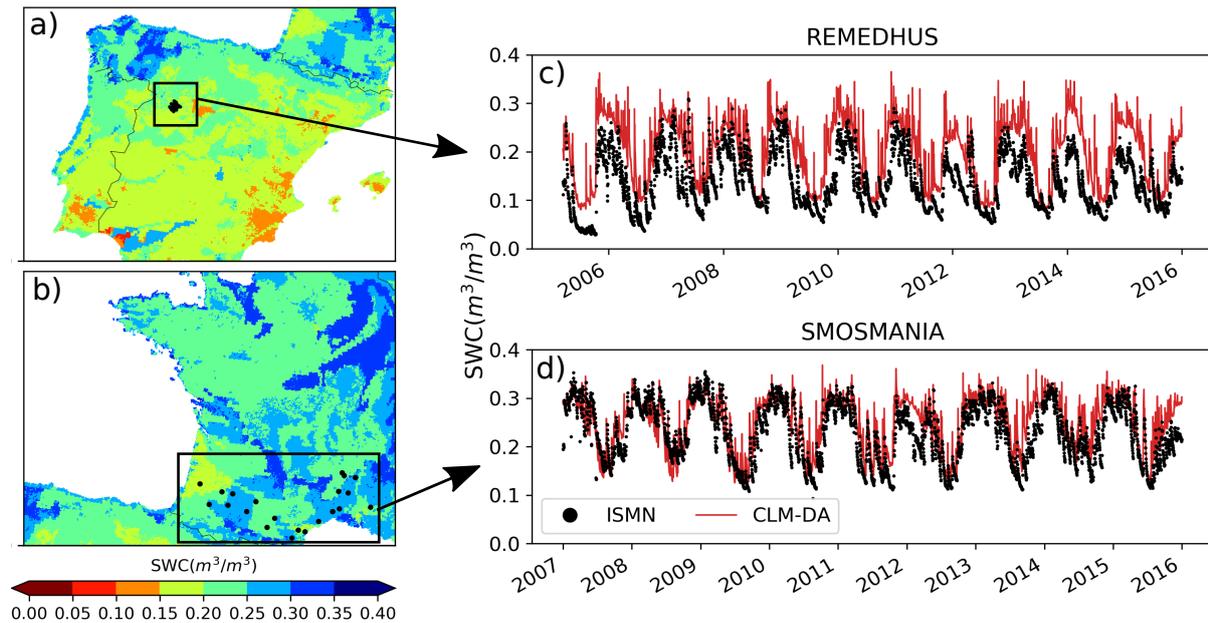


Figure 46: Comparison of daily time series of soil water content (m^3/m^3) from CLM-DA and in-situ observations for the REMEDHUS (a,c) and SMOSMANIA (b,d) soil moisture networks.

To evaluate the quality of ESSMRA, we compared the daily SM from ESSMRA against in-situ data using the SM measurement from 11 networks acquired from the International Soil Moisture Network (ISMN; [23] 2011) across Europe. The surface soil moisture data at selected 112 stations for top 5 cm surface layer were collected to evaluate the ESSMRA product in the top two CLM soil layers (about 3 cm). In-situ data were collected for 2000–2015 period, but their availability does not necessarily cover the whole period. For comparison with ESSMRA daily estimates, the measurements with hourly time scale were aggregated to daily time scale. If more stations are located within one 3 km grid cell, average of those stations were used for comparison. The average of the in-situ observations of all stations were compared with the averaged soil moisture of all grids within each ISMN network. ESSMRA product shows overall good agreement with in-situ observations at daily time scale as shown in Figure 46 for two networks REMEDHUS and SMOSMANIA located in Spain and France, respectively. This comparison shows that at daily scale the model is able to reproduce the daily variations in soil moisture fairly well, with overall correlation above 0.60 and root mean square error ranges 0.04 to 0.19.

Overall, multiple validations revealed that ESSMRA data were consistent with the in-situ soil moisture data and other existing global reanalysis products. The relatively long time series and fine spatial resolution of this new European gridded ESSMRA dataset could provide a valuable data source for many hydrological applications. For example, it can be used as an initial input data for climate change analysis and for numerical weather prediction models to improve the model forecast in terms of location and amount of extreme precipitation events. Because of the scarcity of the in-situ soil moisture observations over large areas, this dataset can also be used for validation of SM outputs in modelling studies. This dataset will be also useful to understand the development and persistence of extreme weather events such as droughts,

D1.2 First scientific results and progress on exascale tools

floods and heatwaves.

In addition to soil moisture assimilation, future work will focus on assimilating groundwater level observation in ParFlow model to characterize uncertainties in subsurface flow. However, implementing the data assimilation framework at 3 km resolution for ParFlow model is challenging because of the high computational cost associated with the EnKF, which relies on an ensemble of realizations to estimate model uncertainty. The computational burden can increase linearly with an increasing number of ensembles in the data assimilation system. To improve model computational efficiency, WP5 (Ensemble Runs; T5.1 and T5.4) develops a new version of the PDAF, which has been tested using ParFlow at watershed scale and shows good model performance in terms of speed. With these new technical development within WP5, the data assimilation framework will be also tested at the continental scale using modernized version of ParFlow.

6.2 Task 1.4.2: Scientific results

T1.4.2 aims at simulating hydropower generation in a meso-scale Alpine watershed. This is achieved by coupling highly detailed information on main hydropower systems that are present in the Italian Alpine region and runoff time series produced by hyper-resolved hydrological model (T1.4.1). Indeed, the development of a large-scale model accounting explicitly for spatial and temporal variability of water usage and explicitly simulating water storage and diversions within a river basin is a relevant topic in the hydrologic field, which we address here by means of HYPERstreamHS hydrological model. Activities conducted during this first reporting period concern with the following three subtasks:

T1.4.2-1	Refactoring of HYPERstreamHS model.
T1.4.2-2	Set-up of the model in the Adige river basin.
T1.4.2-3	Hydrological benchmarking exercise in the Adige river basin.

T1.4.2-1: Refactoring of HYPERstreamHS model

In Task 1.4.2-1 HYPERstreamHS hydrological model has undergone substantial refactoring in order to implement a dual-layer MPI framework for parallelization. Preliminary analyses of the code structure and of the computational workflow were conducted in order to assess how to implement parallel computing in the framework. Figure 47 depicts the overall structure and workflow of HYPERstreamHS. The simulation domain is partitioned into rectangular sub-areas of predefined size, called macrocells, which serve as computational units for the model. Each macrocell may contribute to runoff generation associated to multiple nodes of the river network where streamflow is simulated, thus identifying a given number of macrocell-node pairs. The first level of parallelization is designed to subdivide among a given number of CPUs the runoff simulation at the macrocell level and its routing along the river network to the closest downstream node. However, once runoff is routed and aggregated at the node level proper simulation of Human Systems related to hydropower (hereafter also referred to as HS, i.e. storage reservoirs, diversion channels, etc.) introduces dependencies between the calculated fluxes which require all computations to be solved sequentially. Indeed, node-to-node routing requires communication between all the CPUs before being performed, hence putting a limit to the scalability of the model.

The limitation of scalability during forward simulations as posed by node-to-node routing and by the presence of HS has been circumvented by implementing a second layer of parallelization in the code. In this second layer parallel computing is employed to subdivide the workload associated to the calibration and uncertainty analysis procedures of the hydrological model. In particular, two inverse modeling algorithms are embedded in the model for different purposes: i) the Particle Swarm Optimization (PSO) is used for exploring the parametric space and finding the optimal configuration of the parameters; and ii) the Latin

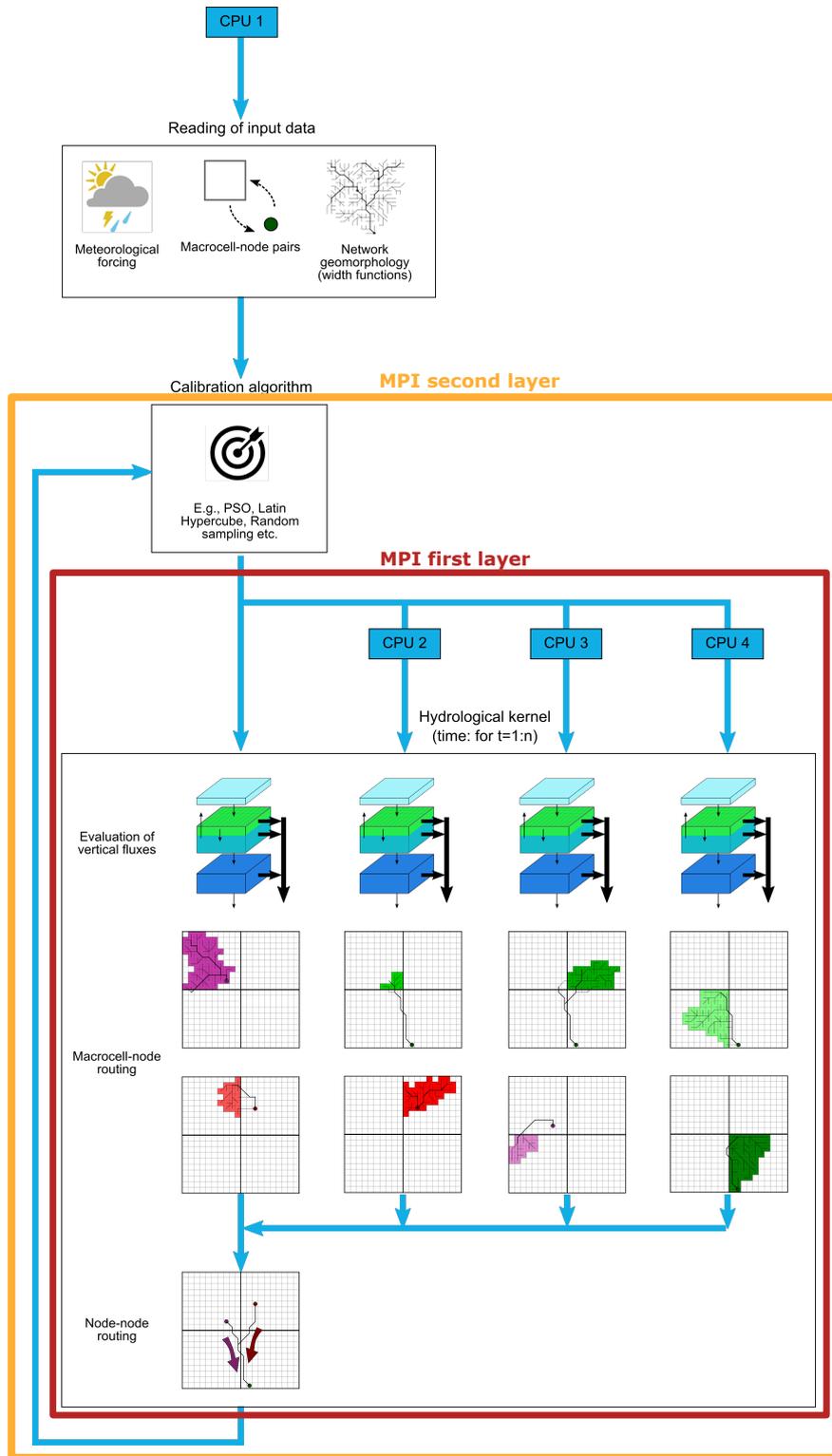


Figure 47: Schematic of the HYPERstreamHS modelling framework and workflow. Two nested MPI layers are embedded in the model, each one operating at a different level of the code.

D1.2 First scientific results and progress on exascale tools

Hypercube Sampler (LHS) is used to perform hydrological parametric uncertainty analysis. Dual-layer parallelization is organized in such a way that a given number of forward simulations is run in parallel (second layer), each of them using a different set of parameters selected by the adopted inverse modeling algorithm (i.e., PSO or LHS). Each forward simulation is performed in parallel as well by partitioning the hydrological kernel with a subset of processors (first layer). The inclusion of the second parallelization layer thus allows further exploitation of the available computational resources, hence improving model's scalability.

The scalability tests on HYPERstreamHS are ongoing and will be presented in the upcoming deliverables. Activities related to the improvement of model scalability were also accompanied by the development of suitable modules for simulating human infrastructures: storage reservoirs, water intakes, diversion channels, pumping storage reservoir, confluences between water tunnels, restitution points and hydropower plants. Each module is specifically characterized by a set of system-specific technical and operational information which introduce appropriate constraints to water mass balance equations.

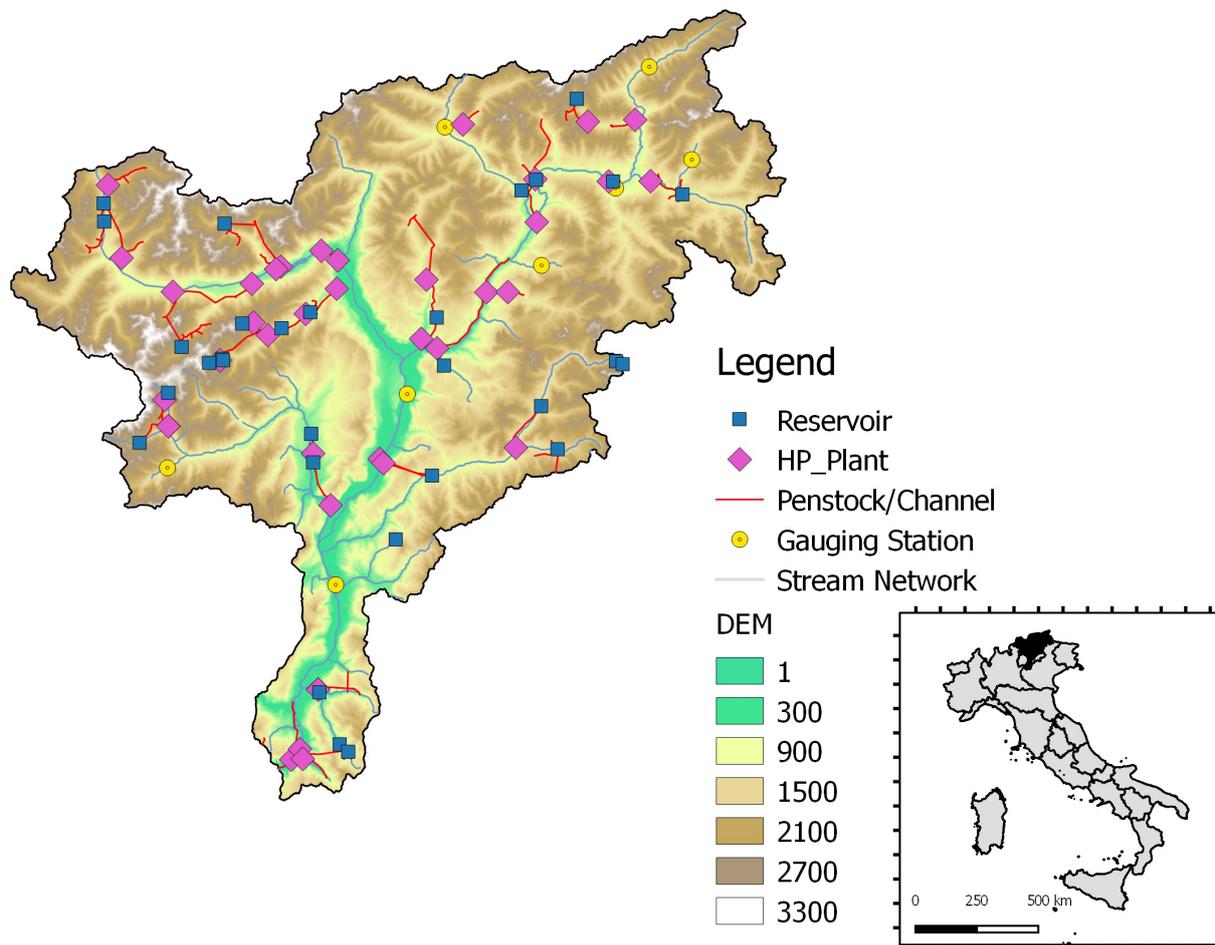


Figure 48: Conceptual model (network topology including Human System nodes) of the Adige river basin.

T1.4.2-2: Set-up of the model in the Adige river basin

In Task 1.4.2-2 detailed simulation of Human Systems related to hydropower was implemented into HYPERstreamHS model with reference to the Adige river basin case study. To this end, the development of the conceptual model (i.e. network topology including HS) of the Adige river basin was performed by identifying the location of all large hydropower systems (defined by the Italian law as those systems with installed power greater than 3 MW) and of their related infrastructures (i.e. dams, tunnels, penstock etc.).

D1.2 First scientific results and progress on exascale tools

Furthermore, extensive information was collected to characterize each HS node: a preliminary phase of model tuning allowed to reduce and optimize the set of information required for the implementation. The collected information allowed to develop reservoir-specific operation rules for the release of Environmental Flows (required by Italian law) and for hydropower production, as well as including an Equivalent Spillway Model in each reservoir that is only activated during flood events. The full conceptual model is depicted in Figure 48. HYPERstreamHS model was set up during the 1989-2013 time window using a daily time step. Meteorological input forcing (precipitation and temperature) was retrieved by the regional dataset ADIGE on a 5x5km grid.

We are currently investigating the model performance in order to assess the best setup to achieve optimal reproduction of historical observations of streamflows and hydropower production. Validation of HYPERstreamHS hydrological and hydropower production models against historical observations will be included in the next deliverable.

T1.4.2-3: Hydrological benchmarking exercise in the Adige river basin

In Task 1.4.2-3 we are evaluating suitable modifications of the Community Land Model v3.5 (CLM3.5) runoff outputs, to improve streamflow reproduction in the Adige river basin. In particular, we want to assess the effectiveness of introduced parameterizations of deep infiltration, return flow and streamflow routing. To this end, the performance of a calibrated run of HYPERstreamHS stand-alone model in reproducing observed streamflow has been evaluated to provide a hydrological benchmark. Note that all simulation runs will be informed by the same meteorological forcing. The multi-site calibration performed in the Trento and Bronzolo streamflow gauge stations, showed a satisfactorily NSE index of 0.65 for HYPERstreamHS model, while the uncalibrated CLM3.5 model run provided a NSE of -0.05.

In the ensuing reporting period, we will evaluate which additional parameterizations can provide a significant improvement of the performance of CLM3.5 in reproducing observed streamflow time series at the local scale.

6.3 Task 1.4.3: Scientific results

T1.4.3 deals with hydrodynamics modeling for flood forecasting and water resources management. Uncertainties in river bed friction coefficients and inflow that enters the hydraulic models translate into uncertainties in simulated water level and discharge. These uncertainties are quantified with a non intrusive sensitivity analysis method based on statistics computed from an ensemble of model trajectories considering stochastic inputs sampled within prescribed probability density functions. The classical Monte Carlo approach is simple and highly adapted to massively parallel computational resources. Yet, its convergence is slow as it scales inversely to the square root of the sample size and its cost gets prohibitive for expensive models. Indeed, full 2D simulations are computationally expensive for large catchments and the cost of the ensemble integration that allows to propagate uncertainty from the input space to the output space is often too expensive for use in a security or operational context. In order to limit the computational cost of the generation of the ensemble, a surrogate model is then used in place of the direct solver. T1.4.3-1 focusses on the construction of a surrogate model from a learning data base, resulting from a limited-size ensemble of integrations of the direct solver. T1.4.3-2 focusses on the use of the surrogate model for the global sensitivity analysis and specifically for the computation of Sobol sensitivity indices that allow to rank the sources of uncertainty by order of importance on the simulated water level and discharge.

The main task T1.4.3 is subdivided in three subtasks, progress are here reported on T1.4.3-1 and T1.4.3-2:

T1.4.3-1	Set up of the hydrodynamic surrogate model over Garonne river in France. The meta modeling exercise has been carried out over the Garonne river test case between Tonneins and La Réole, focusing on uncertainties related to friction and upstream forcing. The surrogate stands for replacement of the hydrodynamics solver TELEMAC (OpenSource software developed by EDF) using the OpenTURNS UQ dedicated library (Open Source software developed by EDF, AirbusGroup, Phimeca); a Galerkin projection (Polynomial Chaos Expansion) strategy has been favored. The construction of the learning data base relies on the development of Python classes for computationally efficient ensemble integration with TELEMAC.
T1.4.3-2	Sensitivity analysis using the surrogate model. The surrogate model implemented in the T1.4.3-1 has been used to carry out a global sensitivity analysis to classify sources of uncertainty to explain water level variance over the 2D domain. This work was carried out for stationary flow and led to the estimation of Sobol indices maps.

Task 1.4.3-1: Set up of hydrodynamic surrogate model over the Garonne river in France

The construction of a hydrodynamic surrogate model was investigated first at CERFACS (see general information in [69], in the context of Nabil El Moçayd Phd (El Mocayd,2016, [25]) and Pamphile Roy Phd (Roy, 2019, [72]) for the 1D Shallow Water model MASCARET that is part of the TELEMAC-MASCARET software (<http://www.opentelemac.org/>). Polynomial chaos Expansion and kriging strategies were investigated relying on algorithms implemented in OpenTURNS (<http://www.openturns.org/>) and ScikitLearn (<https://scikit-learn.org>) libraries dedicated to uncertainty quantification. In El Moçayd et al. (ENMO, 2017, [26]), as the water level varies over space, the PCE was achieved at each grid point using the same set of forward model evaluations, thus providing spatially dependent expansion coefficients. A convergence study focussed on truncation and sampling errors, for the construction of the surrogate with respect to 2 uncertain input parameters, showed that the L^2 truncation error between the surrogate and the direct solver remains below 10^{-6} for polynomial order greater than 6 and the sampling error remains below 10^{-6} as long as the learning data base gathers more than 49 samples. In Roy et al. (SERRA, 2017, [73]), the PCE strategy was compared with a kriging strategy based on Proper Orthogonal Decomposition to reduce the size of the output space and a Gaussian Process regression procedure. Both metamodeling strategies were applied over the Garonne river and succesfully led to efficient computation of water level mean and covariances over the 1D domain as well as Sobol sentitivity indices and probability density function. It was shown that in the upper part of the river, the flow is dominated by the upstream forcing while it is dominated by the friction coefficient in the lower part of the river as illustrated in Fig. 49. The expertise on surrogate modeling in 1D hydrodynamics was extented to 2D model in the context of S. El Garroussi Phd (funded by CERFACS and Région Occitanie) and EoCoE. In the 2D work, lie three major challenges, two of which that are tackled in Task 1.4.3-1:

- the large dimension of the 2D model output that should be reduced before formulating the surrogate,
- the computational cost of the learning data base that should be limited with efficient internal and external paralllism (since the cost of the 2D model is significantly larger than that of the 1D),
- the non linearity in the water level response with respect to friction and input discharge changes within the ensemble and formulation of an advanced surrogate modeling strategy (focus of S. El Garroussi Phd, work in progress and progress report planned for T1.4.3-3).

The previously cited surrogate modeling strategies are non intrusive to the solver. The efficient and modular implementation of these strategies was eased by the development of API (Application Program Interfaces)

D1.2 First scientific results and progress on exascale tools

that allow for elementary actions on the numerical model, its inputs and outputs such as initialization, run, stop, modification of input parameters, save of- and restart from- model state variables in a Python environment without need for recompilation of the direct solver. The development of MASCARET-TELEMAC Python-APIs was initiated by EDF, and further continued at CERFACS for user-dedicated needs related to *data driven* modeling with additional Python classes. Note that the terminology *data driven* is used here by extension of the perspective of ensemble generation to ensemble-based data assimilation. In concrete terms, it means that the model is integrated several times for varying inputs. These developments come in handing to change input parameters to the model, chain integration for the solver and manage outputs. This flexible environment is also fully compatible with the use of Python libraires in uncertainty quantification. As a consequence, the implementation of uncertainty quantification, surrogate modeling and sensitivity analysis algorithms remains fully non-intrusive to the hydrodynamic solver. Major efforts from S. El Garrousi at CERFACS (PhD, supervised by S. Ricci and M. De Lozzo from IRT) structured the classes in the user-dedicated APIs for TELEMAC2D as well as the connection between these classes and the UQ-dedicated functionalities, in a parrallel environment on a cluster at CERFACS. Additional work from A. Piacentini (funded by EoCoE) further improved the external and internal parallelism efficiency, so that the integration of an ensemble is now possible while making the most of available computational resources with MPI strategy through out available cores and nodes. Improvements were also made regarding the management of simulation outputs that are of large dimension, especially in the perspective of chaining the integration of the ensemble with data assimilation algorithm when observation operators require to retrieve the model outputs at observing points. On top of elementary TELEMAC Python-APIs, two main classes are dedicated to data-driven integration of the model. The main class allow to handle ensemble runs with perturbed inputs, working from a single instance of TELEMAC2D and using MPI through mpi4py to launch several members on available resources by batches. A second class is specific to the applicative case and defines which inputs to the model could be modified and how the perturbations are applied to these inputs that may vary in time and space. Additional classes are dedicated to the management for ensemble output fields or output values. These classes are available on the SVN server of the code TELEMAC, on a dedicated branch, together with a testing script for ensemble generation and an automatic validation procedure.

These classes are used to generate ensemble simulations with TELEMAC 2D over the Garonne river between Tonneins and La Réole, thus providing learning and validation data bases from which the surrogate model is respectively built and validated. The results are presented in the UNECECOMP conference paper by El Garroussi et al. (El Garroussi et al., 2019, [24]) for stationnary flow, meaning when the flow occupies the flood plain. Epistemic and aleatory uncertainties are taken into account considering two uncertainty sources on water level calculation for extreme flood event, respectively the roughness coefficient (4 areas for friction: 3 in the river bed and 1 in the flood plain) and the upstream discharge. The water level field discretized over an unstructured finite element mesh over the Garonne River (South-West France) between Tonneins and La Reole is simulated with TELEMAC-2D. A Gaussian process regression surrogate was favored here as its number of hyperparameters is limited to the small number of uncertain input parameters (1 inflow and 4 spatially distributed friction coefficients). Yet its formulation for large dimension model outputs (41 00 points in the present case) is challenging; thus the output space dimension is reduced using a Proper Orthogonal Decomposition (POD) strategy. The POD is achieved with the snapshot method, computing the Singular Value Decomposition of the covariance matrix of the snapshot matrix where a limited numer of simulation outputs are gathered. This leads to an approximate description of the spatial water level field, based on a limited number of the most significant components (4 modes are kept in the present study). The Gaussian process approximates the relation between the model inputs and these components coefficients, building from a squared exponential correlation function. The hyperparameters for the Gaussian process estimation result from the maximum likelyhood optimization on the learning data set with the L-BFGS algorithm. The POD+GP surrogate model is, in the end, formulated as the weighted sum of the most significant POD components where weights are the GP surrogate models depending on the sampled input values.

D1.2 First scientific results and progress on exascale tools

The quality of the POD+GP surrogate model is assessed computing the Root Mean Square Error between the surrogate outputs and the direct solver output on a validation data set (resulting from the integration for an ensemble of the direct solver) as well as by the predictive coefficient Q_2 criteria that measures the performance of the surrogate with respect to a raw average of the learning value, at each point of the spatial domain. These criteria are shown in Fig. 50. The POD+GP surrogate quality is very good with respect to global error metrics $RMSE = 0.8$ cm and $Q_2 = 0.99748$. The POD+GP surrogate is thus a reliable candidate to replace the direct solver, at a significantly reduced computational cost.

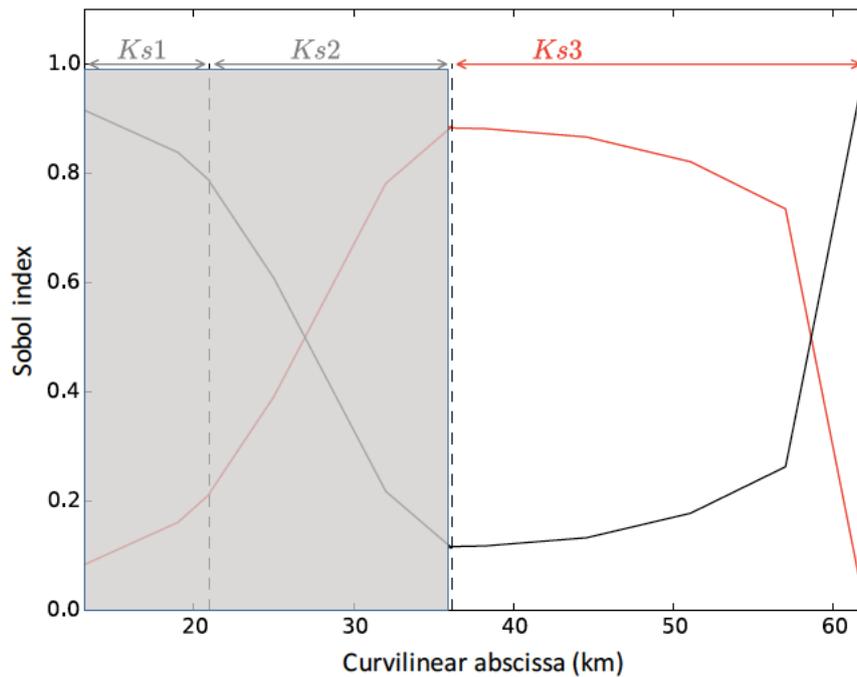


Figure 49: Sobol indices computed with the Polynomial Chaos Expansion for 1D model MASCARET over the Garonne river, with respect to friction coefficient $Ks3$ (red solid line) and upstream discharge Q (black solid line), El Moçayd et al. , ENMO, 2017 [26].

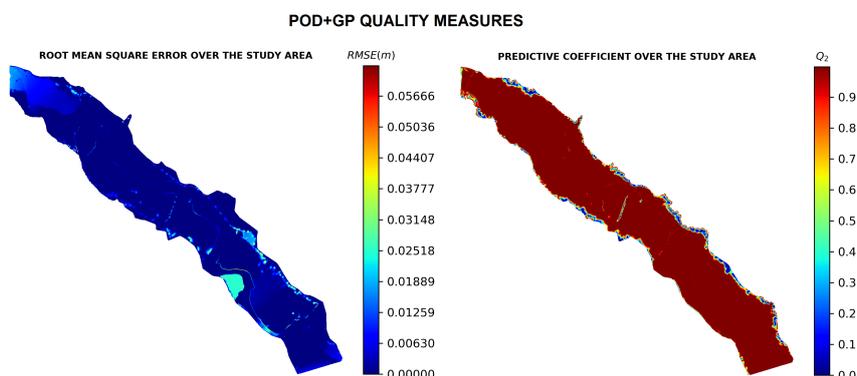


Figure 50: Root mean square error (left) and predictive coefficient (right) for the POD+GP surrogate El Garroussi et al., UNCECOMP, 2019, [24].

Task 1.4.3-2: Sensitivity analysis using the surrogate model

In this task the surrogate model formulated in Task 1.4.3-1 is exploited to carry out a global sensitivity analysis that leads to classifying sources of uncertainty at a reasonable computational cost. Sobol’ sensitivity indices are computed.

In El Garroussi et al. (2019) [24], over the Garonne river, for stationnary flow, the previously estimated POD+GP surrogate is used to carry out a variance-based sensitivity analysis over the entire simulated area, with a focus on the output location where the city of Marmande, prone to flooding, is located and where an in-situ observing station is operated. The sensitivie analysis at Marmande highlights that most of the water level variance is explained by the upstream discharge and to a lesser extend, by the Strickler friction coefficient Ks_4 prescribed between Marmande and La Reole as displayed in Fig. 51. It should also be noted that for each input variable, the first (S) and total Sobol’ (ST) indices at Marmande are equal, meaning that, at this location, the multivariate impact of the input on the water level is minimal. The map of Sobol indices was also computed with respect to each uncertain input parameter. It shows that, given the statistical assumptions that were made in the study, on the probability density functions for the friction and the upstream discharge, the flood plain friction has no impact on the water level, and major river bed friction coefficient has a local impact on water level in its own definition area mostly. Yet, over the entire domain, the most significant source of uncertainty is the upstream discharge.

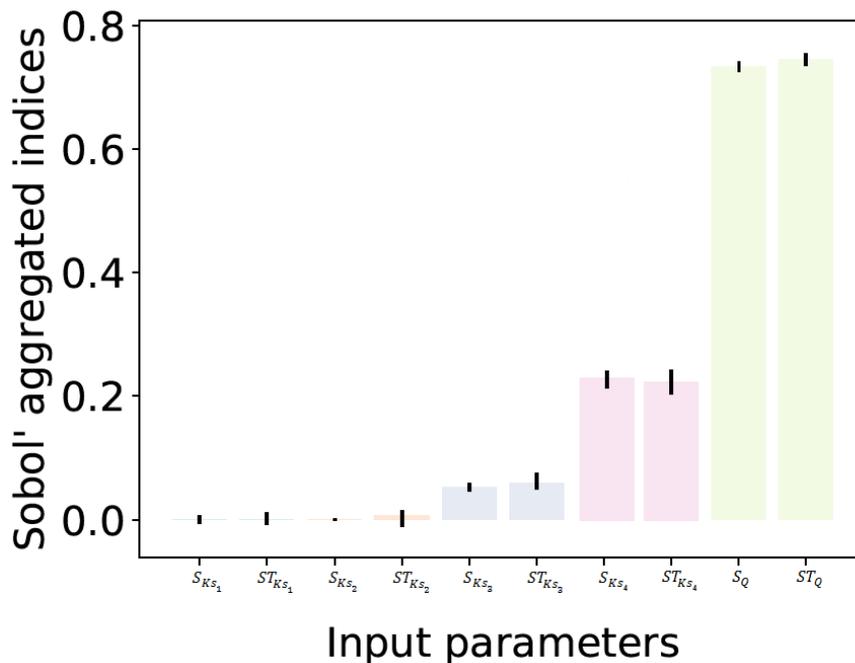


Figure 51: First (S) and total (St) Sobol’ indices estimated with POD-GP surrogate at Marmande, from El Garroussi et al., UNCECOMP, 2019.

6.4 Tasks T1.4.4 & T1.4.5 (Geothermal Energy)

T1.4.4 deals with experimental design for predicting borehole locations in geothermal reservoirs in order to decrease the uncertainty in parameter estimation, which, in turn, improves the quality of the reservoir simulation. Drilling boreholes during exploration and development of geothermal reservoirs not only involves high cost, but also bears significant risks of failure. In geothermal reservoir engineering, techniques of optimal experimental design (OED) have the potential to improve the decision making process. Earlier publications explained the formulation and implementation of this mathematical optimization problem and demonstrated its feasibility for finding borehole locations in two- and three-dimensional reservoir models that minimize the uncertainty of estimating hydraulic permeability of a model unit from temperature

D1.2 First scientific results and progress on exascale tools

measurements. Subsequently, minimizing the uncertainty of the parameter estimation results in a more reliable parametrization of the reservoir simulation, improving the overall process in geothermal reservoir engineering. Various OED techniques are implemented in the Environment for Combining Optimization and Simulation Software (EFCOSS). To address problems arising from geothermal modeling, this software framework links mathematical optimization software with SHEMAT-Suite, our geothermal simulation code for fluid flow and heat transport through porous media.

Within EoCoE-II we want to improve the computational performance of the existing OED workflow for geothermal modelling. This will enable us to apply it to more detailed numerical models of geothermal reservoir systems. Higher spatial resolution of numerical models, in turn, improves validity of OED results. Moreover, we aim at extending the OED workflow by additional functionalities, such as predicting optimal borehole depth or length of temperature logs or applying it for inverting parameters other than permeability (e.g., thermal conductivity). By simulating different model scenarios, we analyze the sensitivity of the OED approach to factors such as measurement errors, prior data density or quality. Our results improve the understanding of the presented workflow and show that the application of OED techniques to practical problems in geothermal reservoir engineering is feasible.

In Task 1.4.5 we carry out the application of optimal experimental design (OED) on a real world geothermal reservoir system model. The aim is to predict optimal locations for additional exploration wells for temperature measurements. We define the optimal location as the one, where the information content of the data (i.e temperature) is highest with respect to certain reservoir rock parameters (e.g. permeability, thermal conductivity); i.e, the rock parameters will be estimated with least uncertainty when inverting data from the optimal location. The quality of geothermal reservoir models and hydrothermal flow simulations highly depends on the quality of the subsurface parametrization. Reliable determination of physical rock properties such as porosity, permeability or thermal conductivity is crucial for the simulation results as they are directly linked to the observations (e.g., temperature, pressure, flow rate). This task will apply the OED workflow, which is improved and extended in Task 1.4.4. On the one hand, its results will demonstrate the feasible and successful applicability of the OED workflow to geothermal reservoir modeling. On the other hand, the results will provide new insights for the simulated geothermal reservoir system.

Task 1.4.4: Scientific results

The main task T1.4.4 is subdivided into seven subtasks that are described in detail in D1.1:

T1.4.4-1	Becoming acquainted with OED theory and the state of the art.
T1.4.4-2	Software access.
T1.4.4-3	Learning and understanding the EFCOSS software environment and its coupling with SHEMAT-Suite.
T1.4.4-4	Reproducing OED simulations and developing a post-processing workflow.
T1.4.4-5	Defining a research and development concept for OED for geothermal modelling.
T1.4.4-6	Setting-up a testmodel suite and extending OED functionality.
T1.4.4-7	Optimizing the OED workflow.

This deliverable reports on the progress made in these subtasks, of which the first four were completed and the others are work in progress.

The quality of geothermal reservoir models and hydrothermal flow simulations highly depends on the quality of the subsurface parametrization. Usually, direct and indirect measurements within deep formations are conducted within very few deep boreholes as drilling is expensive. Thus, parameters are often deter-

D1.2 First scientific results and progress on exascale tools

mined indirectly by inverse approaches that deduce reservoir properties from observation data. However, considering the effort, cost, and risk of borehole measurements, the goal is to determine the optimal position for such measurements in order to estimate rock properties with least uncertainty, i.e. optimizing the inversion quality. Seidler et al. [79] showed that this challenge of geothermal reservoir engineering can be addressed by methods from optimal experimental design (OED). Seidler et al. [79] extended the inversion approach presented in [66] and coupled OED techniques with the geothermal reservoir simulator SHEMAT-Suite. Starting from there, the main goal of this task is to optimize and extend the OED workflow for geothermal modeling.

Finding the optimal experimental conditions for constraining model parameters is a question of the sensitivity of the model with respect to the unknown parameters. This sensitivity, or the information content of the observations, is mathematically described by the Fisher matrix F :

$$F(p, X) = \left(\frac{\delta y}{\delta p}\right)^T \left(\frac{\delta y}{\delta p}\right)$$

F contains the first order derivative of the model output y (e.g. temperature) towards the parameters p (e.g. permeability); this term is the Jacobi matrix. Here, the experimental condition is the borehole location X at the surface. For evaluating the information contained in the Fisher matrix, OED criteria are formulated. One is the D-optimal design criterion which is based on the determinant of the Fisher matrix

$$\psi_D = -\log(\det(F))$$

The minimum of the D-optimal criterion contains the maximum sensitivity, which is the formulation of the OED problem:

$$\min_X \psi_D(p, X)$$

Other OED criteria are the A- and E-optimality that are based on the trace of the inverse Fisher matrix or the eigenvalue of the Fisher matrix, respectively. Various OED techniques are implemented in the software framework EFCOSS, which links mathematical optimization software with SHEMAT-Suite. The latter is our geothermal simulation code for fluid flow and heat transport through porous media, for addressing problems arising from geothermal modeling. Figure 52 shows the general high-level EFCOSS architecture. More details on EFCOSS and how the simulation code SHEMAT-Suite is linked are provided in section 6.5

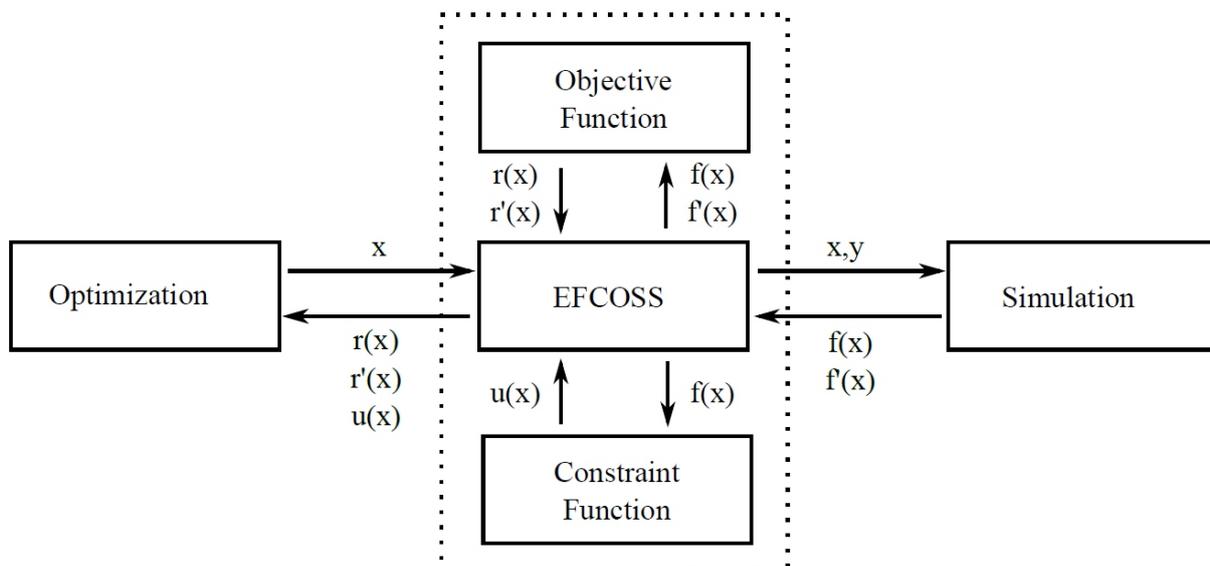


Figure 52: High-level EFCOSS architecture [78].

D1.2 First scientific results and progress on exascale tools

Access to the in-house software EFCOSS [65] including the existing OED workflow for geothermal simulations with SHEMAT-Suite [78] was kindly granted by the developers at the Chair for Advanced Computation, Friedrich Schiller University Jena. Furthermore, the group provided theoretical and practical advice during personal meetings in Aachen and Jena, as well as via email support.

During an OED simulation, EFCOSS executes SHEMAT-Suite for computing the model's Jacobi matrix using Automatic Differentiation (AD) for an exact calculation of the first order derivatives [66]. Subsequently, the Fisher information matrix is computed from which different optimal design criteria can be computed, e.g. the D-optimality. EFCOSS routines compute the D-optimality for each grid cell, which was specified in the steering script and write them to an output file. Recently, the option of storing the Jacobian as a Numpy array after the first SHEMAT run was implemented, so that EFCOSS can read from this array when computing the D-optimality for the next grid cell. In the previous implementation, SHEMAT-Suite was called and the complete inverse simulation was executed repeatedly for each new grid cell. Storing the Jacobi matrix greatly reduces the computing time for the OED simulation; for one of the 2D test models, for instance, it gets reduced from more than 11 hours to around 6 minutes.

For understanding the whole OED workflow, a deeper comprehension of the deterministic inverse simulation with SHEMAT-Suite was necessary. Activities included working through the source code, testing of the inverse mode, and finally documenting the source both at the developer- and user-level. This work and the reproduction of two existing OED studies revealed several bugs in the SHEMAT-Suite regarding the generation of the derivatives.

The code's derivatives are evaluated numerically by applying Automatic Differentiation (AD) in forward (tangent) mode, using the software Tapenade (<https://www-sop.inria.fr/tropics/tapenade.html>, [40]). AD is a set of techniques that transform a numerical model M , implemented as a program P , into a new program P_{dev} that computes derivatives of M . The techniques are based on the application of the chain rule. In forward mode, the differentiated program is a copy of the given program, where additional derivative instructions are inserted just before each instruction [39]. In contrast to approximation by divided differences, AD computes derivatives accurately to machine precision. The derivatives of the forward model, and thus the AD code generation, are essential for calculating the Jacobi matrix required for deterministic inversion and OED. Thus, some time was spent on bug fixing and consecutive testing in collaboration with developers in Jena (as forecasted as a risk in D1.1).

At the same time, the SHEMAT-Suite source code was refactored, restructured and cleaned-up during the last months, in order to make it publicly available as an open-source code. This makes the code easily accessible and usable for new and external users and developers and ensures reproducibility of simulations and scientific studies. Main features and results of this work are:

- Refactorization of the build management by replacing make with CMake.
- Bug fixes in ad-mode, so that AD code is available now for all property modules.
- Implementation of a semi-automatic testing environment, allowing for reproducible testing of all main code functionalities after new developments.
- Extensive documentation in doxygen and in the code's GitLab Wiki.
- Open-source code repository: <https://git.rwth-aachen.de/SHEMAT-Suite/SHEMAT-Suite-open>
- Code capsule at Code Ocean: <https://doi.org/10.24433/CO.9420115.v1>
- Publication in SoftwareX [45]

In order to evaluate OED simulation results fast, efficiently, and comparably, a post-processing workflow has been setup by using Jupyter Notebooks. Jupyter Notebooks allow for combining the processing and visualization of data with explanatory and descriptive text elements. The basic post-processing workflow is available for 2D models and will be extended for 3D models in the upcoming months.

D1.2 First scientific results and progress on exascale tools

The development of the OED workflow will concentrate on extending the OED approach in the following two ways: (i) We predict not only the borehole's position at the surface, but also its logging depth or logging interval needed for temperature measurements. (ii) We estimate not only the hydraulic permeability, but also the thermal conductivity. In summary, we determine different experimental conditions such that when estimating different parameters of model units from temperature measurements in the borehole the uncertainty is minimized. Numerical simulations of synthetic geothermal reservoir scenarios will demonstrate the applicability of OED for optimizing borehole or logging depth. Additionally, by simulating different model scenarios, we plan to analyze the sensitivity of the OED approach to factors such as measurement errors, prior data density or quality. Currently, this research is supported by a Bachelor student, who is studying the influence of the prior data location and the depth of the experiment on the OED result.

Until now, the testmodel suite comprises several synthetic 2D subsurface models. They cover different heat transport processes and geothermal reservoir characteristics; for instance free versus forced convection or fault-bound hydrothermal systems. The testmodel suite is smaller than originally planned, because of delay due to COVID-19. The testmodels need to be extended by a three dimensional model in the next step. Also, transient simulations are missing at the moment (e.g. geothermal doublet during production), and will be added.

For demonstrating the functionality of the OED workflow, we show first results of an OED study for a 2D synthetic model of a geothermal reservoir above a salt diapir, which is cross-cut by two high permeable faults (modified after [66]). Figure 53 - Top shows the geological reservoir units. Each model unit is characterized by constant thermal and hydraulic properties (e.g., permeability, porosity, thermal conductivity). High heat conductivity of salt results in higher temperatures at shallower depth. This makes the sedimentary units above and close to salt diapirs interesting for direct use of geothermal energy. In addition, the permeable faults in our model provide pathways for advective heat transport, resulting in a heat transport towards the surface through the western fault (unit 11, blue). There is a borehole at location $x_0 = 7475$ m with a certain depth $z_0 = 1475$ m providing a temperature log $Tlog_0$. The problem is to find the location x_1 for an additional borehole of the same depth for measuring another temperature log $Tlog_1$. These temperature data shall be used for estimating the fault permeability k_{10} and k_{11} with least uncertainty (i.e., optimally). The experimental condition is the horizontal position of the additional borehole. We assumed a prior fault permeability $k_{10} = k_{11} = 5 * 10^{-15} m^2$. The synthetic true reference fault permeability is $5 * 10^{-14} m^2$. For this OED problem, the numerical simulation provides two possible optimal horizontal ranges for placing the additional borehole (Figure 53, top): (i) from 1025 m to 1175 m and (ii) from 8225 m to 9975 m. The absolute D-optimality minimum is at $x = 8675m$. We will further use this model in the course of the project for simulating various OED problems and studying the impact of factors such as prior assumptions, location of prior data logs and their data quality.

The OED functionalities have not been extended yet. The next step is to work on the borehole depth optimization, according to the research and development concept. The option to store the Jacobi matrix from the SHERAT-Suite simulation, which is now available and was described above, is a main step for optimizing the performance of the OED workflow. Besides, we started to exploit the MPI parallelism within EFCOSS recently. It provides the possibility to run several independent OED problems in parallel, e.g. for applying several different prior parameters. This and potential other possibilities for optimizing the OED workflow will be evaluated within the upcoming project months.

Task 1.4.5: Scientific results

The main task T1.4.5 is subdivided in five subtasks that are described in detail in D1.1:

T1.4.5-1	Evaluation of the study area.
T1.4.5-2	Data collection.
T1.4.5-3	Setting-up a conceptual and numerical model.

T1.4.5-4	Model calibration.
T1.4.5-5	OED study.

This deliverable reports on task T1.4.5-1. The other subtasks are scheduled for the upcoming project months.

The originally proposed showcase of a geothermal reservoir in Tuscany, Italy, turned out to be not a feasible candidate for an OED study with SHEMAT-Suite within the framework of the project for several different reasons. Among them are practical issues such as no active collaboration and no cooperation agreement with Enel Green Power, the company that operates the reservoir and owns most of the relevant data. Besides, physical reservoir conditions in the very hot Tuscany reservoir are not favorable for our applications. It involves particular complications due to water/steam phase transition and possible two-phase flow, which burden the elaboration of the OED problem unnecessarily. In view of the available project duration, we therefore refrain from this application example. Currently, two alternative application examples, for which numerical reservoir models are available already, are under consideration. First, the Acocolco reservoir in Mexico that was recently scientifically studied within the Horizon2020 project GEMex [22]. Second, the reservoir model for the city of Geilenkirchen, which was setup and used during the first EoCoE phase.

6.5 Code demonstrators

ParFlow

ParFlow (v3.2) is a massively parallel, physics-based integrated watershed model. It simulates fully coupled dynamic 2D/3D hydrological, groundwater and land-surface processes for large scale problems. Saturated and variably saturated subsurface flow in heterogeneous porous media are simulated in three spatial dimensions using a Newton-Krylov nonlinear solver ([7]; [43]) and multigrid-preconditioners. ParFlow also features coupled surface-subsurface flow which allows for hillslope runoff and channel routing ([46]). ParFlow is a highly scalable code enabling it to be used for high resolution watershed simulations ranging from single river catchments to continents ([53]). ParFlow is successfully run on a large range of platforms ranging from single CPU notebooks and workstations, to distributed/shared memory clusters (e.g. JSC/JURECA) to massively parallel systems such as IBM BlueGene P and Q.

However, current version of ParFlow is time consuming when run over large scale and at high resolution (For example, one year of simulation takes about 1 month for the problem size of 365 x 1592 x 1544 x 15). This makes it difficult to perform long term simulations (i.e. 30 years) to accurately predict changes related to climate events. Therefore, it is needed to accelerate the run time efficiency of the model by implementing modern accelerator architectures, the graphics processing units (GPUs) within the ParFlow code. Within EoCoE-II, this new implementation for accelerator architectures in ParFlow hydrologic model demonstrates that a significant performance gain, high developer productivity, and minimally invasive implementation are all achievable at the same time while keeping the codebase well maintainable in the long-term. A representative benchmark problem was run on the JUWELS supercomputer where each utilized node is equipped with Dual Intel Xeon Gold 6148 processors (2 x 20 cores @ 2.4 GHz) and 4 NVIDIA V100 16GB GPUs. The nodes are connected through Dual EDR-Infiniband (Connect-X4) interconnect. The benchmark consist of a variably saturated infiltration problem into a homogeneous soil with a fixed water table at a depth of 6 m and a constant infiltration rate of m/hour. The vertical and lateral spatial discretization was 0.025 and 1 m, respectively. The time step size was 1 h. The profile was initialized with a hydrostatic profile based on a matric potential of -9 m at the top resulting in a considerable initial hydrodynamic disequilibrium with respect to the water table at the bottom boundary. The number of grid cells in the lateral directions was varied to change the total number of degrees of freedom in the performance testing. The reference results were obtained without accelerator devices by launching an MPI process for each CPU core. In case

D1.2 First scientific results and progress on exascale tools

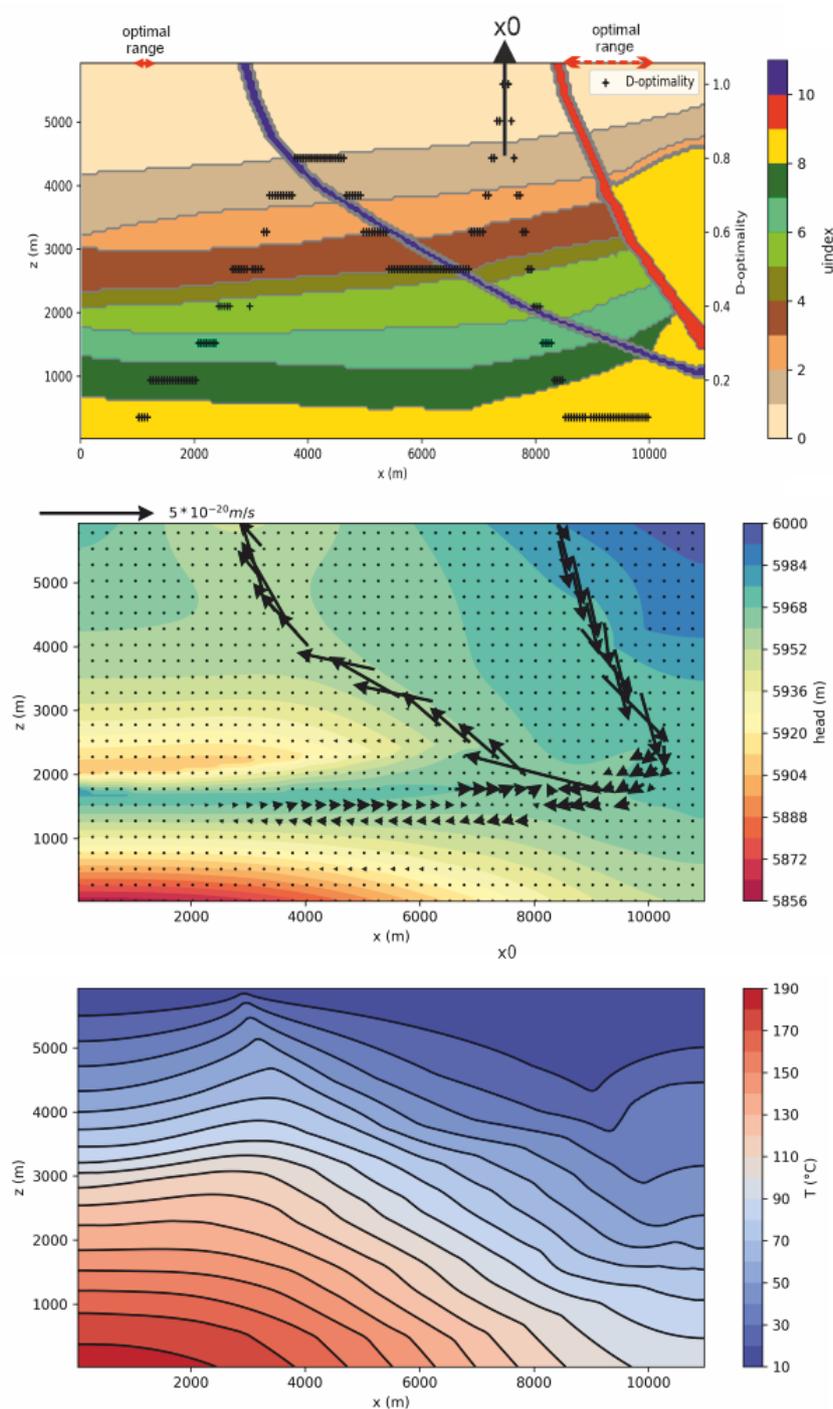


Figure 53: Numerical forward model of a synthetic geothermal reservoir above a salt diapir computed with SHEMAT-Suite. Top: Reservoir structure in terms of geological model units (uindex); unit 9 (yellow) is the salt and units 10 and 11 (red and blue) are permeable faults. The remaining units are various sedimentary layers. X0 marks the location of the existing borehole and the black crosses depict the OED result in terms of normalized and binned D-optimality. Red arrows display two optimal ranges for an additional borehole. Middle: Darcy flow in terms of hydraulic reference head and Darcy velocity (arrows) for the reference true reservoir properties. Bottom: Steady state temperature distribution for the reference true reservoir properties.

D1.2 First scientific results and progress on exascale tools

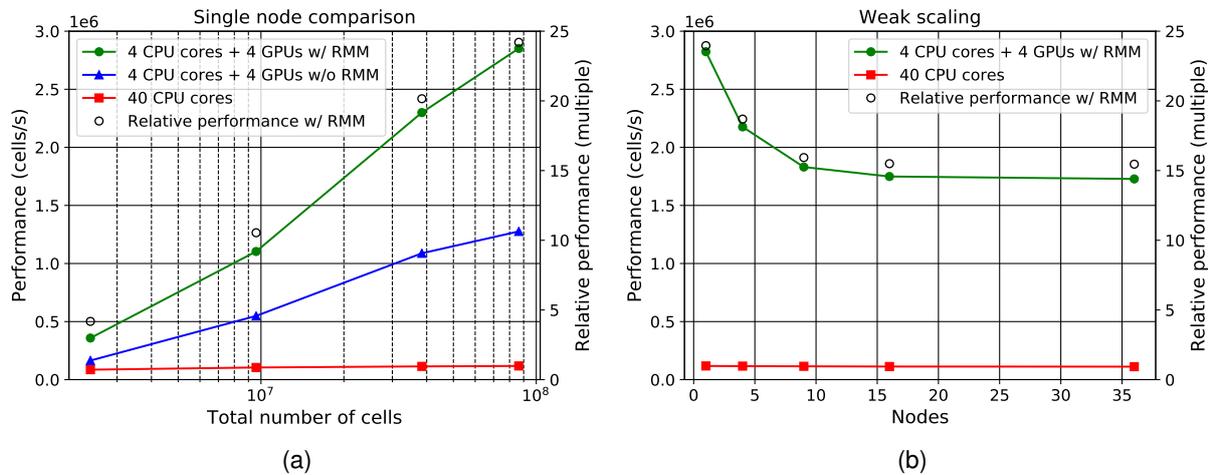


Figure 54: a) Single node performance comparison., b) Weak scaling comparison.

of the GPU accelerated runs, 4 MPI processes per node were launched such that each process uses one GPU and GPU-GPU data transfer is handled by a CUDA-aware MPI. The current, moderately optimized ParFlow GPU version runs a representative model up to 20 times faster on a node with 2 Intel Skylake processors and 4 NVIDIA V100 GPUs compared to the original version of ParFlow, where the GPUs are not used (Figure 54). These results demonstrate that the ParFlow GPU implementation may serve as a blueprint to tackle the challenges of heterogeneous HPC hardware architectures on the path to exascale.

TELEMAC-MASCARET

TELEMAC-MASCARET is an integrated suite of solvers for use in the field of free-surface flow. TELEMAC-MASCARET is managed by a consortium of core organisations: Artelia (formerly Sogreah, France), Bundesanstalt für Wasserbau (BAW, Germany), Centre d'Etudes et d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement (CEREMA, France), Daresbury Laboratory (United Kingdom), Electricité de France R&D (EDF, France), HR Wallingford (United Kingdom) and CERFACS (France). TELEMAC-MASCARET is used by most partners for dimensioning and impact studies, where safety is prevailing and, for this reason, reliability, validation and a worldwide recognition of our tools are of utmost importance. As a consequence and to improve the access to TELEMAC-MASCARET for the whole community of consultants and researchers, the choice of open source has been made. The source code, a list of various test cases and a automatic validation procedure can be found on the TELEMAC svn repository (<http://www.opentelemac.org/index.php/download>) together with a full documentation of the code. A git server should very soon replace the svn repository.

The data driven Python APIs previously mentioned in Sect. 6.3 are gathered in a dedicated svn branch, together with a testing script for ensemble generation and an automatic validation procedure.

Experimental design for geothermal modeling with SHEMAT-Suite and EFCOSS

SHEMAT-Suite is a finite difference code written in FORTRAN95 for simulating fluid flow, heat and species transport in porous media, which is commonly applied for simulating hydrothermal processes in geothermal reservoirs. It comprises forward simulation, deterministic inversion with automatic differentiation, stochastic inversion (Monte Carlo) and data assimilation (Ensemble Kalman filter). It features a two-level shared memory parallelization and a distributed memory parallelization for Monte Carlo simulation. An MPI parallel data assimilation framework is currently being integrated within WP2 and WP4. Detailed description of the code is available in [45]. This publication accompanies the recent open-source publication of the code: <https://git.rwth-aachen.de/SHEMAT-Suite/SHEMAT-Suite-open>.

D1.2 First scientific results and progress on exascale tools

The software framework EFCOSS consists of several Python classes. Its general software architecture is described in [78] (see also Fig. 52). For OED simulations, SHERAT-Suite is made available to EFCOSS as a static library from which the shared object `shemat.so` is generated. The user has to provide the EFCOSS problem definition in form of a Python class that steers the desired operations. In addition, the free parameters of the optimization need to be defined by the user.

There are several levels of parallelism implemented in the OED workflow [78]. On simulation code level, we use the OpenMP parallelism of SHERAT-Suite for executing the geothermal reservoir simulation in parallel. In addition, the inversion can run in parallel if more than one parameter is inverted, using as many processes as inversion parameters. Finally, several EFCOSS instances can be executed in parallel, e.g. when running simulations for a range of prior parameters.

The 2D model and OED results presented in subsection 6.4 demonstrate the functionality of the OED workflow (see also Figure 53). With 26180 grid cells and a steady state solution, this is a rather small test case. The job's runtime was 3 minutes using 24 threads. Two threads were used on the outer-level, because we have two active inversion parameters. Twelve threads were used on the inner-level for the reservoir simulation. However, before the storing of the Jacobi matrix was implemented, the same simulation took around 11 hours.

We will add a reservoir scale three-dimensional model to the testmodel suite, as well as transient simulations. Both will require longer runtimes and the usage of more compute resources. The problem size will increase on several levels: grid size, number of inversion parameters, range of prior parameter values. The goal is to simulate OED problems for models with 0.5 to 1 million grid cells, 5 to 10 inversion parameters, and around 10 prior parameter values. The independent OED problems for different prior parameter values can be spread over up to 10 compute nodes, while the computation of each OED problem will exploit all threads available per node, since the SHERAT simulation uses a shared-memory parallelization. The number of usable threads depends on the available hardware. For example, the compute cluster at RWTH Aachen University offers dedicated systems for shared-memory applications with up to 144 threads per node.

6.6 Roadmap revision

T1.4.1

For T4.1.1-2, the ParFlow model simulations are completed for the time period of 1997-2006. Since most of the observations and measurement of different hydrological variables such as soil moisture, evapotranspiration, total water storage discharge information at various gauge stations and measured ground water table depths across Europe are available for recent time period (i.e. after 2010), these simulation will be extended for the period of (1997 - 2018). The current version of ParFlow is time consuming (For example, one year of simulation takes about 1 month). The new implementation of GPU within ParFlow (See Section 6.2) showed 20 times faster run time for the synthetic experiment. The ParFlow-GPU version will be tested for the real test case over the EU-CORDEX domain, which will increase the run time efficiency of ParFlow. Similarly the T4.1.3, the simulation of ParFlow coupled with PDAF over Europe is highly dependent on the technical work package (i.e. WP5). While some progress has been done in making the ParFlow-PDAF code more computationally efficient using synthetic test cases, the model performance will be tested using the real test cases in the next half period of the project. In addition, because of the COVID-19 pandemic, the subtasks T4.1.1-2 and T4.1.1-3 are delayed by 3-4 months (as shown with pink color in table below) due to the reduced work caused by home office and school closures. However this will not effect the overall completion of the subtasks.

D1.2 First scientific results and progress on exascale tools

WATER T1.4.1	Project Months											
		D1.1	MS1			D1.2			D1.3			
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.4.1-1												
T1.4.1-2												
T1.4.1-3												

T1.4.2

Activities related to subtasks T1.4.2-1, T1.4.2-2 and T1.4.2-3 are in line with the project timeplan and did not suffer significant delays due to the recent COVID-19 emergency, nor to other causes. T1.4.2-4, however, requires the collection of technical information about main hydropower systems in the Italian Alpine Region from their respective owning companies and regional hydrological offices. Indeed the COVID-19 emergency reduced sensibly the responsiveness of such offices, which became harder to be contacted. The combination of this issue with the reduced work by project staff at UNITN due to home office and school closure is likely to impact the timing of T1.4.2-4, with a foreseen delay of the activities by 2-3 months. However this will not affect the overall completion of the subtask.

WATER T1.4.2	Project Months												
		D1.1	MS1			D1.2			D1.3				
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36	39
T1.4.2-1													
T1.4.2-2													
T1.4.2-3													
T1.4.2-4													

T1.4.3

The main lines of the code map has not significantly evolved for T1.4.3-1 and T1.4.3-2. Yet, some precisions are given here to highlight and justify some focus :

- The construction of the surrogate model for the Garonne river was carried out in T1.4.3-1 and T1.4.3-2 as development in the Python API dedicated to data driven were made to improve computational efficiency and allow for more tests for the following steps. Also a major upgrade of the TELEMAC code was pushed on the server and required quite some work for merging our previous developments.
- The extension of the previously described work is currently been extended to non stationnary flow. As local non linearities occur between friction, upstream discharge and simulated water level as the flow occupies the flood plain, a Mixture of Experts strategy is favored in the context of S. El Garroussi PhD, funded by CERFACS and Région Occitanie.
- The study is now totally focussed on the Garonne river. The application over the Gironde estuary will most likely not be achieved in the context of EoCoE. Indeed, there is some Intellectual properties limitations on this hydrodynamics model that could not be solved so far. Yet, the Garonne catchment offer sufficient opportunities to fully explore the methodologies in stationnary and non stationnary flow, with time varying upstream boudary conditions in the latter case. This modification is thus not considered as a penalty to the project, on the contrary, as the Garonne model is Open Source, it favors exchanges of the results between institutions that are active on this site.

D1.2 First scientific results and progress on exascale tools

Timeline for T1.4.3-3 will be updated soon as discussion between CERFACS and Juelich for the use of Parflow hydrology output on the Garonne river will take place in July 2020. We first need to move on to non stationary flow with observed upstream forcing at Tonneins, we secondly need to consolidate the Mixture of Expert surrogate strategy for non stationary flow, before considering using simulated or forecasted boundary conditions to the hydraulic code.

Water T1.4.3	Project Months											
		D1.1	MS1	D1.2						D1.3		
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.4.3-1												
T1.4.3-2												
T1.4.3-3												

T1.4.4 & T1.4.5

Some additional time had to be spent on subtasks T1.4.4-3 and T1.4.4-4 later than originally planned, because of bug fixes in the code. Therefore, and mainly because of the COVID-19 pandemic, subtasks T1.4.4-6 and T1.4.4-7 are delayed. The lockdown and the closure of childcare facilities resulted in reduced work efficiency and productivity. The subtasks' duration is extended accordingly in the revised timeline. They will be finalized approximately three months later than planned originally.

WATER T1.X.Y	Project Months											
		D1.1	MS1	D1.2						D1.3		
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.4.4-1												
T1.4.4-2												
T1.4.4-3												
T1.4.4-4												
T1.4.4-5												
T1.4.4-6												
T1.4.4-7												

Subtask T1.4.5-1 is completed. The other subtasks of T1.4.5 need to be rescheduled, because of delay in T1.4.4. T1.4.5-1 resulted in a revision of the study area for this task. Two new candidates were identified, which are more feasible for the project. Numerical reservoir models for SHEMAT-Suite exist already for both possible study areas. This reduces the time required for subtasks T1.4.5-2 and T1.4.5-3, what mitigates the risk of delay for this task.

WATER T1.X.Y	Project Months											
		D1.1	MS1	D1.2						D1.3		
Subtasks	3	6	9	12	15	18	21	24	27	30	33	36
T1.4.5-1												
T1.4.5-2												
T1.4.5-3												
T1.4.5-4												
T1.4.5-5												

7 Fusion for Energy (T1.5)

The scientific challenge aims at building a first principle flagship code capable of addressing (i) *electromagnetic turbulence* (ii) *from the core to the far edge region* in (iii) *ITER-relevant magnetic geometries*. The breakthrough of these simultaneous three major upgrades will open the door to groundbreaking physics advances.

The new code release GYSELAX is being developed to this end. It capitalizes on and considerably extends the capabilities of the 5D flux-driven gyrokinetic Fortran parallel code GYSELA , which solves Vlasov and Poisson equations to simulate electrostatic plasma turbulence and transport in core Tokamak plasmas. In particular, the parallelization efficiency – presently limited by network capability and memory performance – will be improved, targeting exascale supercomputers.

Task 1.5 aims at paving the way towards the preliminary exploitation of GYSELAX , and is divided in three main tasks:

T1.5.1	Prototype of GYSELA-X: arbitrary magnetic equilibrium in limiter configuration
T1.5.2	Advanced GYSELA-X: X-point configuration & alternative/complementary methods
T1.5.3	Core-edge-SOL physics: GYSELA-X & Tokam3X

7.1 T1.5.1: Scientific results

This task focuses on the basics with the help of prototypes, namely the derivation and implementation of the new set of equations and the identification of optimal choices for handling non-circular geometries and boundary conditions.

The main task T1.5.1 is subdivided in two subtasks:

T1.5.1-1	Identify minimal set of equations for ion turbulence.
T1.5.1-2	Comparative efficiencies of numerical strategies (flux-surface aligned and Cartesian mesh grids in poloidal plane, and of patches with regular or irregular meshes).

T1.5.1-1 Minimal set of equations

Core physics (closed magnetic field lines): GYSELA worked in the electrostatic limit in toroidal geometry with circular cross-section. The gyrokinetic equations including electromagnetic effects have been derived (modification of particle trajectories plus Ampère equation) and implemented in GYSELAX . The ongoing benchmarks – so far focusing on the development of a tearing mode instability – are encouraging. In particular, the so-called “magnetic cancellation” issue – related to the compensation of large terms in Ampère equation – has been adequately resolved within the semi-Lagrangian numerical scheme.

Scrape-Off Layer (SOL) physics: the periphery of the core confined plasma is made of open magnetic field lines (the SOL) which intercept solid elements, leading to plasma-wall interaction. To a large extent, this region governs tokamak edge plasma physics and contributes to the access to improved confinement regimes. In addition, it is the locus of heat and particle exhaust. The presence of a solid limiter has been successfully accounted for in GYSELAX by using immersed boundary conditions via the penalization technique, imported from the companion Tokam3X fluid code. So far, only ion-wall interactions are considered, electrons being treated adiabatically in the SOL.

The more complex case where both ions and electrons interact with a solid surface is being addressed with the VOICE kinetic code, which solves the coupled set of Vlasov-Poisson equations along the parallel (to the guiding magnetic field) direction in the SOL with the same penalization technique using ad-hoc mask functions. It is the prelude to the implementation in GYSELAX . In particular, the recovery of the theoretically expected physical characteristics of the plasma at the sheath entrance is currently being investigated when

D1.2 First scientific results and progress on exascale tools

scanning the ion to electron mass ratio, the parallel mesh grid and the mask parameters. The objective is to identify both physically satisfactory and numerically tractable conditions for GYSELAX .

The issue of axi-symmetrical X-point divertors (characterized by the vanishing of one component of the magnetic field at the X-point) will be addressed from a theoretical point of view by the end of the project.

T1.5.1-2 Comparative efficiencies of numerical strategies

In order to explore different reasonable possibilities for field solver for GYSELAX that is efficient for very large problems, we looked at Multigrid solvers on (i) a locally refined cartesian mesh on the one hand and on (ii) a flux surface aligned mesh on the other hand. Both options have advantages and drawbacks: a polar like grid would be aligned on the magnetic flux surfaces which makes it easier to handle the strong anisotropies (turbulent structures are much elongated in the direction parallel to the guiding magnetic field, and short-scale in the transverse plane). Moreover the refinement to handle the large gradients at the edge of the plasma would only need to be in the radial direction. On the other hand there is an issue with the singularity at the magnetic axis and also at the X-point in case the separatrix is included in the simulation domain. A specific multigrid solver is being developed for this configuration.

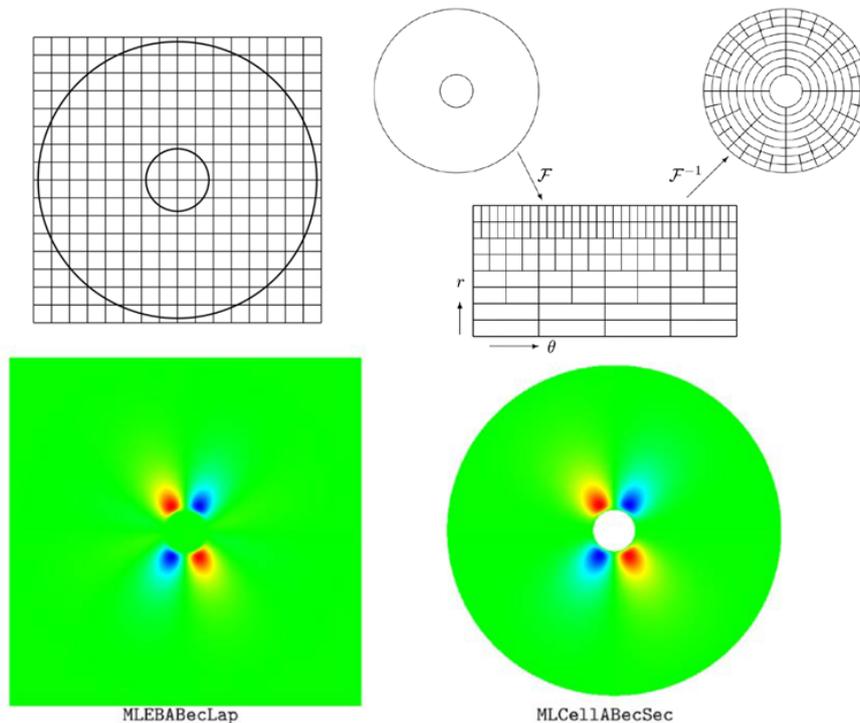


Figure 55: Embedded boundary (upper left), block-structured polar mesh (top right). Comparison of results on cartesian (bottom left) and polar (bottom right) grids.

The locally refined grid does not have the issue of the singularity at the magnetic axis and at the X-point but needs a strong 2D refinement in the edge region of the plasma. Moreover, there are very good open source libraries available for this method. We reviewed two algebraic multigrid libraries with adaptive mesh refinement (AMR), AMReX (from LBNL, USA) and waLBerla (from Universität Erlangen-Nürnberg, Germany). We decided to use the AMReX library, which seemed better suited for our project.

We tested two approaches based on the AMReX library to solve the gyrokinetic quasi-neutrality equation on a circular domain before going to the realistic shaped tokamak domain. This required the implementation of a new type of abstract equation and the extension of the multigrid solvers in AMReX. We also implemented a mapping from a block-structured logical grid refined at the edge to the circular physical domain. To avoid

D1.2 First scientific results and progress on exascale tools

the singularity at the axis we had to leave out a hole in the middle. See Figure 55 (top right). For the cartesian domain, there is no issue at the center but we had to implement an embedded boundary (Figure 55 (top left), provided by the library, to correctly describe the physical domain. We could obtain correct solutions for both test cases with a similar accuracy (Figure 55 (bottom)). The multigrid solver required considerably fewer iterations on the cartesian domain. We could document a good numerical performance for the multigrid solvers with OpenMP on a single node, but there is still an issue with the MPI parallelisation, which might be related to a bug or some wrong choice in the option. This has to be further investigated and improved.

On the other hand, in collaboration with CERFACS in WP3, a polar multigrid solver on a mesh that can be strongly refined in the radial direction has been developed from scratch. This solver is adapted specifically to our problem and should overcome the issues that we have adapting a polar multigrid solver from a library as we just saw. A discretization via symmetric finite difference stencils to yield a matrix free operator has been developed for the gyrokinetic quasi-neutrality equation and also a specific multigrid solver which converges at the optimal rate. The strong refinement at the edge of the domain does not seem to have any influence on the number of steps required for convergence. This solver is already available in circular and shaped geometry and will be coupled with the GYSELA code in the second part of the project. This work is described in more details in WP3.

7.2 T1.5.2: Scientific results

This task deploys in the advanced version the solutions retained in the prototypes, and extends the problem to critical large temperature variation and X-point issues. The main task T1.5.2 is subdivided in two subtasks:

T1.5.2-1	Handle the X-point of ITER-relevant magnetic configurations with multi-patch or flux coordinate independent schemes.
T1.5.2-2	Multi-patch treatment of the Vlasov equation to handle large variations of temperature.

T1.5.2-1 Handle the X-point

The numerical challenge to treat spatial regions of both closed and open magnetic field lines in a single simulation has been addressed by means of the gyrokinetic GENE code. To this aim, the so-called “flux coordinate independent (FCI)” approach – pioneered in fluid turbulence codes like GRILLIX (A. Stegmeir et al., IPP) – has been extended to GENE. Actually, both codes now use the same FCI routines (as an API). This technique has been tested thoroughly and successfully via the method of manufactured solutions. It allows for simulations of entire diverted tokamaks (from the magnetic axis to the boundary), even including, in principle, complicated structures like magnetic islands or stochastic regions. Further tests and applications are planned.

T1.5.2-2 Multi-patch treatment of Vlasov equation

As already mentioned, the CEA-IRFM team has lost one of the pillar developers of the GYSELA code, with an expertise on computational science and high-level parallelism which cannot be replaced by non-permanent staff. As a consequence, the rewriting of the code has been abandoned and replaced by its *complete refactoring* with enhanced modularity targeting exascale supercomputers.

Details of the first phase, which has led to a significant cleaning of the code (reduction by about one fifth of the number of lines), can be found in deliverable D2.2, section 9.3.

D1.2 First scientific results and progress on exascale tools

The refactoring has then entered its second phase, which has started 6 months ago and will continue until the end of the project. It aims at adding more modularity. We report here on the main actions which were undertaken and on preliminary results.

- The first step consists in isolating each kernel of calculation, the objective being then to be able to rewrite a certain number of these kernels in C or C ++ in order to take advantage of the most advanced technologies for optimization on new computing architectures (such as KOKKOS, etc.). In this context, we had to abandon the patch strategy initially proposed because it would have required an almost complete rewriting of the code.
- Yet, GYSELAX still needs to be able to address the large temperature variations from the very hot core to the far cool edge, as this represents a major issue in core-edge turbulence interplay in tokamak plasmas. To this end, we have decided to treat this intrinsic difficulty by using non-equidistant splines. This requires modularizing the existing equidistant splines and replacing this module with a new one. First coupling tests of a non-equidistant spline module developed in the SELALIB digital library (collaboration CEA-IRFM Cadarache and MPG-IPP Garching) have been performed on a prototype code and reveal conclusive. The coupling to the GYSELAX code is in progress and will require several weeks of development. More details can be found in deliverable D2.2, section 9.3.2.
- Testing the required resolution of the a non-equidistant mesh-grid to solve the Vlasov equation in the case of large temperature gradient was the aim of the 5-week study project proposed at CEMRACS. Indeed, the performance – both in terms of accuracy and of CPU time consumption – of the non-equidistant splines has to be tested against relevant physical problems. The targeted reduced model is the 2-dimensional (1D-1V) VOICE kinetic code, in view of preparing its implementation and use in GYSELAX . This work is key to the success of the ongoing developments regarding non-equidistant meshes, and seen as a must do to find the most suitable method to implement in the GYSELAX code. This meeting should have taken place this summer from July 20 to August 28, 2020. Unfortunately, it was canceled (actually postponed in 2 year) due to the Covid-19 sanitary crisis. This postponement is unfortunate and not compatible with the deadlines of the EoCoE project. We hope to be able to hire a post-doctoral researcher to work in collaboration with the French Universities of Aix-Marseille and Strasbourg on this critical issue. In any case, the deadlines will be longer than initially planned.

7.3 T1.5.3: Scientific results

This task takes advantage of all the developments to address first physics issues and is subdivided in two subtasks:

T1.5.3-1	Large scale equilibrium flows expected in the SOL.
T1.5.3-2	Core, edge and SOL interplay in turbulent regime.

T1.5.3-1 Large scale flows

An important aspect of the core–edge–SOL coupling lies with the development of large scale flows. Upon modification of the outer boundary condition and accounting for SOL physics in GYSELA, a drastic change occurs at the edge boundary and a steady large-scale shear flow develops and endures. The mechanism whereby the flow develops is an arduous question, under current investigation as several channels may be invoked and may interplay. Neoclassical or orbit loss physics are possibilities and are likely important and certainly so is turbulence as a strong and localised turbulent activity springs near the last closed flux surface (LCFS) in the presence of a limiter only. As an intermediate step to the global question of flow onset

D1.2 First scientific results and progress on exascale tools

mechanisms, we have thus thoroughly studied the linear nature of the instabilities that are responsible for near LCFS turbulence intensity, comparing nonlinear computations equal in every way but with respect to their outer boundary condition: with a poloidally-symmetric SOL (Fig.56) versus a limited configuration (Fig.57). This is done using the eigenvalue solver of the GKW local gyrokinetic framework applied to the nonlinear equilibrium profiles of the GYSELA computations.

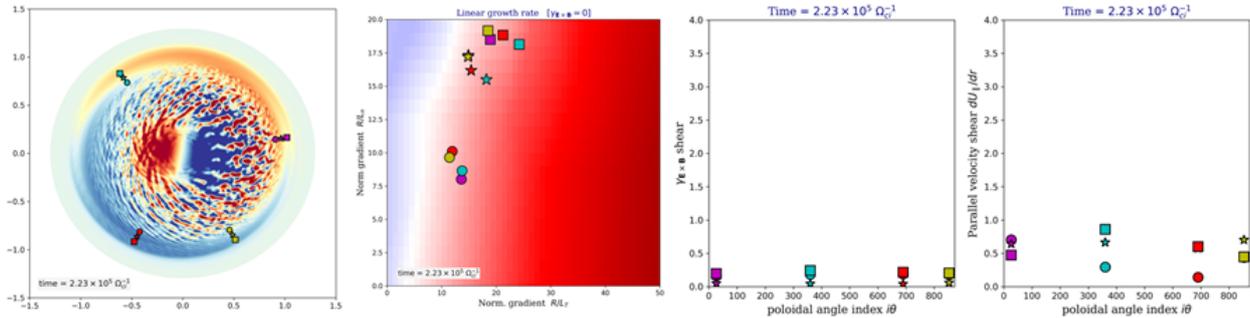


Figure 56: POLOIDALLY-SYMMETRIC case. (From left to right) 1. poloidal cross-section of the electric potential fluctuations with the 13 different radial-poloidal locations where linear stability analysis is performed; 2. the 13 poloidal locations are represented against 2-dimensional linear growth rate estimates at vanishing $\mathbf{E} \times \mathbf{B}$ shear; 3. Computation of the local $\mathbf{E} \times \mathbf{B}$ shear values with respect to poloidal location; 4. Computation of the local parallel velocity shear U' at the various poloidal locations. All data are radially averaged over 20 Larmor radii ρ_i

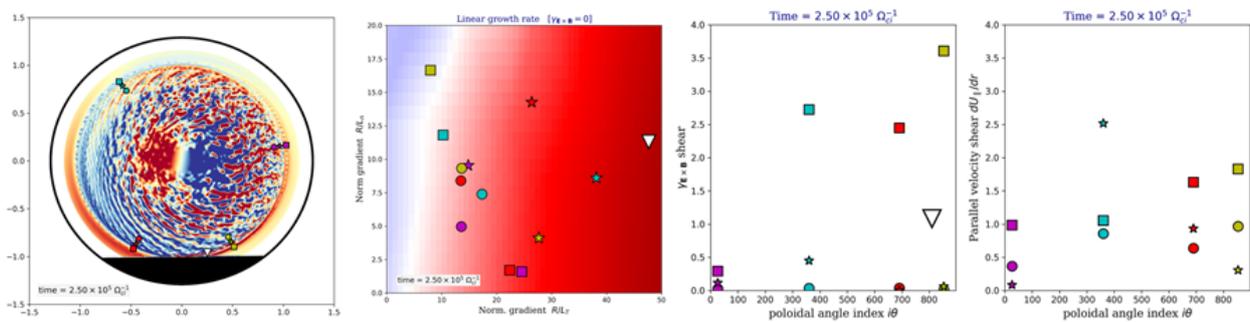


Figure 57: LIMITER case. Same captions as Fig.56.

We find that the dominant instability is of interchange character and dominantly driven by the ion temperature gradient (ITG). As clearly visible, there is a drastic increase in the $\mathbf{E} \times \mathbf{B}$ shear rate $\gamma_{\mathbf{E} \times \mathbf{B}}$ in the limited configuration in the immediate vicinity of the LCFS (squares) with respect to the poloidally-symmetric case, denoting the onset of a modest but persistent transport barrier at the plasma edge. The 2-dimensional map of instability growth rates as a function of the logarithmic gradients in Figs.56 and 57 is indeed estimated in GKW as an eigenvalue problem, on smoothed variations of the profiles and at vanishing $\mathbf{E} \times \mathbf{B}$ shear. The inclusion of $\mathbf{E} \times \mathbf{B}$ shear greatly changes the linear stability so that all 13 locations become stable or marginal stable except for the locations at normalised radius $r/a = 0.98$, just inside the LCFS, and especially in the low-field side quadrant and the immediate vicinity of the limiter. This points towards its key role for the nonlinear evolution in the upstream plasma (in the edge and core) as well as for the possible onset and sustainment of the edge transport barrier. This is under current investigation.

Also, despite significant velocity shear near the LCFS, with significant coarse-graining over a flux tube (i.e. whilst performing radial (over $20\rho_i$) and flux-surface average of the GYSELA profiles) Kelvin-Helmholtz-like instabilities (sometimes referred to as Parallel Velocity gradient-PVG) are found to contribute modestly to the global instability. Interestingly, local values however of the parallel velocity shear U' in the low-field

D1.2 First scientific results and progress on exascale tools

	PWI-AE	PWI-KE	Shaping	EM
VOICE	-	X	-	-
GYSELAX implementation	X	-	X	X
GYSELAX debugging	X	-	-	X
GYSELAX benchmark	X	-	-	X
GYSELAX production	X	-	-	-

Table 19: Current status of ongoing physics developments. Legend: PWI = Plasma Wall Interaction (open field line region or SOL); AE/KE = Adiabatic/Kinetic Electrons (ions are always kinetic); Shaping = non-circular poloidal cross-section; EM = Electromagnetic effects.

side quadrant near the limiter may locally reach up to 3 to 5 times the mean, which certainly opens the possibility for very localised yet significant instability sources of Kelvin-Helmholtz nature.

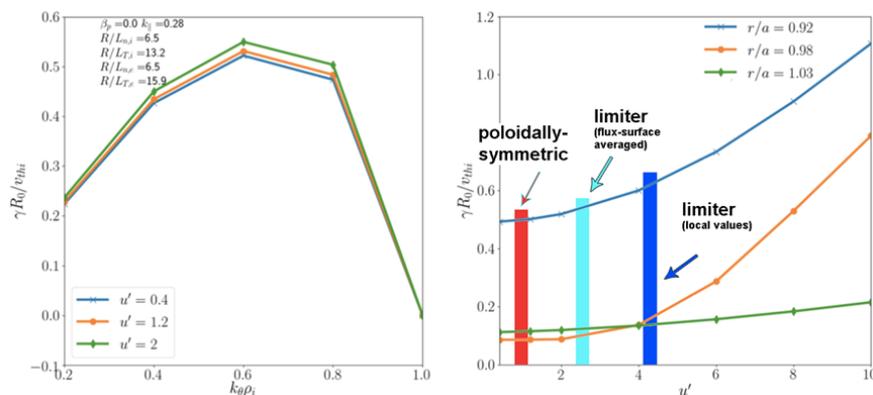


Figure 58: (left) scanning U' in the range 0 to 2, as a function of $k_{\theta}\rho_i$; (right) scanning U' at $k_{\theta}\rho_i = 0.6$ at 3 radial positions $r/a = 0.92, 0.98, 1.03$ for a radial (over $20\rho_i$) and flux-surface average of the GYSELA profiles with limiter. In red, cyan and dark blue are reported the maximum values for U' over the 13 locations for respectively the poloidally-symmetric, limited with radial averaging over $20\rho_i$ and limited with radial averaging over $10\rho_i$.

T1.5.3-2 Core-edge-SOL interplay in the turbulent regime

A few nonlinear runs have been performed with a penalized limiter in the SOL so as to identify the impact on plasma-wall interactions on core confinement. As already stated, electrons are still treated adiabatically so far, so that there is no turbulent transport of matter. Our preliminary results are twofold: (i) the level of turbulent fluctuations increases towards the edge in this case – as opposed to a drop of turbulence intensity at the edge in the absence of limiter – consistently with experimental observations, and (ii) this increase seems to be due to the combined effect of entrainment by asymmetric large scale poloidal flows and radial spreading.

7.4 Code demonstrator

The present status of the GYSELAX flagship code with respect to our final scientific goals is comprehensively detailed in table 19.

Besides, the completed refactoring has led to successful porting and scaling tests on the TGCC Irene-AMD machine, as illustrated on figure 59. The strong scaling leads to 80% (resp. 59%) relative efficiency on

D1.2 First scientific results and progress on exascale tools

49 152 (resp. 98 304, corresponding to the total number of cores of the machine) cores. So far, the code recovers on new machines previous scalability records. We expect that the ongoing work which aims at increasing the code modularity should help further improving the performance.

Yet, these optimizations have already allowed us to run GYSELAX with previously unaccessible – yet experimentally relevant – parameter ranges mimicking Tore Supra plasmas from the very core to the far edge (closed and open field lines), highlighting the complex interplay between core confinement and the physics of the SOL. The largest simulation has been performed within the PRACE project “ElectroPulse – Electric field and Pulse propagation in Magnetised Plasma Turbulence” on the AMD-Rome partition of the TGCC Irene machine. To this end, a total of about $2.06 \cdot 10^{11}$ grid points have revealed mandatory to mesh the 5-dimensional phase space ($r \times \theta \times \varphi \times v_{G\parallel} \times \mu = 512 \times 1024 \times 64 \times 128 \times 48$) and $2.5 \cdot 10^5$ time units (corresponding to ~ 1.4 ms of plasma time) to reach the statistically saturated turbulent state (corresponding to about 32 restarts of 24h on the supercomputer). Regarding parallelization, a total of 24 576 CPUs were used, corresponding to $4 \times 4 \times 48$ nodes with 34 threads each of an hybrid MPI-OpenMP parallelization.

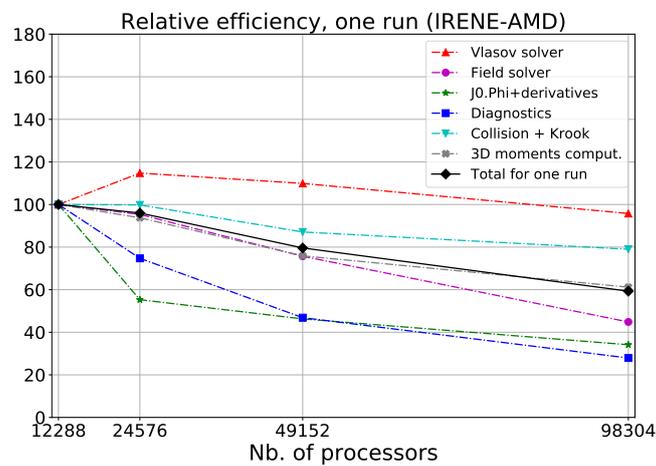


Figure 59: Strong scaling for GYSELAX on Irene-AMD supercomputer (TGCC).

7.5 Roadmap revision

The main scientific objectives are not modified, and the time line remains essentially the same as the one reported in D1.1. It is reproduced below for the different subtasks of the project. Light color cells indicate period of times where the work has been impacted by the sanitary crisis due to COVID-19. Details regarding risks and possible mitigations are given hereafter.

- Impact on T1.5.1-2: the Hackathon on code performance analysis and optimization, initially scheduled in April 2020, has been canceled. It has not been replaced by online lectures so far. We expect a possible moderate impact on HPC performance within the time frame of EoCoE-II. However, the code performance could still be improved after the end of the project if required.
- Impact on T1.5.2-1 and T1.5.3-2: the hiring of a post-doc has been delayed by 2 months at least. Slower activity is expected, hence delayed results. This may impact the achievable physics objectives of Task 1.5.3-2, although not dramatically. The only consequence one foresees is to get publications in journals with lower impact factor.
- Impact on T1.5.2-2: the CEMRACS of July-August 2020 is canceled, meaning that full-time 4 week collaborative work is abandoned. This is critical to GYSELAX . It may prevent to consider relevant

D1.2 First scientific results and progress on exascale tools

temperature profiles from the core to the far edge, where temperatures drops by 2 orders of magnitudes at least. Mitigating the impact require organizing interactive collaboration involving expected members of CEMRACS. The corresponding mission budget should typically cover 2 times 2 weeks for 2-4 people.

- Impact on T1.5.3-1: the Master-2 internship could not start on this task because of no access to numerical resources due to confinement regulations. The study of plasma-wall interaction with kinetic electrons using VOICE has been therefore delayed by 2-3 months.

FUSION Subtasks	Project Months											
		D1.1		MS1		D1.2					D1.3	
T1.5.1	3	6	9	12	15	18	21	24	27	30	33	36
T1.5.1-1												
T1.5.1-2												
T1.5.2-1												
T1.5.2-2												
T1.5.3-1												
T1.5.3-2												

References

- [1] The alaiz experiment - alex17. <https://thewindvaneblog.com/the-alaiz-experiment-alex17-revealing-mountain-valley-large-scale-flow-patterns-6176416dbf2>.
- [2] The alya system - large scale computational mechanics. <https://www.bsc.es/es/computer-applications/alya-system>.
- [3] lea wind task 31 wakebench. <https://community.ieawind.org/task31/home>.
- [4] New european wind atlas. <https://www.neweuropeanwindatlas.eu/mission/objectives>.
- [5] Kempf A., Klein M., and Janicka J. Efficient generation of initial- and inflow-conditions for transient turbulent flows in arbitrary geometries. *Flow Turbulence Combust*, 74:67 – 84, 2005.
- [6] Y. Al-Hamdani, D. Alfè, and A. Michaelides. How strongly do hydrogen and water molecules stick to carbon nanomaterials? *J. Chem. Phys.*, 146:094701, 2017.
- [7] Steven F. Ashby and Robert D. Falgout. A Parallel Multigrid Preconditioned Conjugate Gradient Algorithm for Groundwater Flow Simulations. *Nuclear Science and Engineering*, 124(1):145–159, September 1996.
- [8] Jan Bartl and Lars Sætran. Blind test comparison of the performance and wake flow between two in-line wind turbines exposed to different atmospheric inflow conditions. *Wind Energy Science Discussions*, pages 1–31, 09 2016.
- [9] N. H. Batjes. A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling. *Soil use and management*, 13(1):9–16, 1997.
- [10] T. Belytschko, W. Kam Liu, and B. Moran. *Numerical simulation of reactive flow in hot aquifers: SHEMAT and processing SHEMAT*. Springer Science & Business Media, 2000.
- [11] C. Bollmeyer, J. D. Keller, C. Ohlwein, S. Wahl, S. Crewell, P. Friederichs, A. Hense, J. Keune, S. Kneifel, and I. Pscheidt. Towards a high-resolution regional reanalysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society*, 141(686):1–15, 2015.
- [12] K. Boersma, J.G. Schepers, S. Gomez-Iradi, I. Herraez, T. Lutz, P. Weihsing, L. Oggiano, G. Pirrung, H.A. Madsen, W.Z. Shen, H. Rahimi, and P. Schaffarczyk. Final report of iea wind task 29 mexnext (phase 3). , Energy Research Centre of the Netherlands, 2018.
- [13] Botaurus-stellaris. Topographic map of germany, 2006. https://en.wikipedia.org/wiki/Geography_of_Germany#/media/File:Deutschland_topo.jpg.
- [14] G. Bussi and M. Parrinello. Accurate sampling using Langevin dynamics. *Physical Review E*, 75:056707, 2007.
- [15] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566, 1980.
- [16] T.C. Chiang and F.J. Himpsel. *Subvolume A 2.1.3 Si: Datasheet from Landolt-Börnstein—Group III Condensed Matter, vol. 23a: “subvolume a” in springermaterials*. Springer, Berlin, 1989.
- [17] Jens Hesselbjerg Christensen and Ole Bøssing Christensen. A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic change*, 81(1):7–30, 2007.

D1.2 First scientific results and progress on exascale tools

- [18] Matthew Churchfield, Sang Lee, Sven Schmitz, and Zhixiang Wang. Modeling wind turbine tower and nacelle effects within an actuator line model. In *33rd Wind Energy Symposium*, 01 2015.
- [19] P. Czaja, M. Celino, S. Giusepponi, M. Gusso, and U. Aeberhard. Ab initio study on localization and finite size effects in the structural, electronic, and optical properties of hydrogenated amorphous silicon. *Computational Materials Science*, 155:159–168, 2018.
- [20] P. Czaja, S. Giusepponi, M. Gusso, M. Celino, and U. Aeberhard. Computational characterization of a-Si:H/c-Si interfaces. *Journal of Computational Electronics*, 17:1457–1469, 2018.
- [21] Jeffrey J. Danielson and Dean B. Gesch. Global multi-resolution terrain elevation data 2010 (GMTED2010). Technical report, US Geological Survey, 2011.
- [22] P. Deb, D. Knapp, G. Marquart, C. Clauser, and E. Trumpy. Stochastic workflows for the evaluation of Enhanced Geothermal System (EGS) potential in geothermal greenfields with sparse data: the case study of Acoculco, Mexico. *Geothermics*, 88, 2020.
- [23] Wouter Dorigo, Wolfgang Wagner, Clement Albergel, Franziska Albrecht, Gianpaolo Balsamo, Luca Brocca, Daniel Chung, Martin Ertl, Matthias Forkel, and Alexander Gruber. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, 203:185–215, 2017.
- [24] S. El Garroussi, M. De Lozzo, S. Ricci, D. Lucor, N. Goutal, C. Goeury, and S. Boyaval. Uncertainty quantification in a two-dimensional river hydraulic model. In *UNCECOMP 2019, Crete, Greece, 24-26 June*, 2019.
- [25] N. ElMocayd. *La Décomposition en polynôme du chaos pour l'amélioration de l'assimilation de données ensembliste en hydraulique fluviale*. PhD thesis, Institut National Polytechnique de Toulouse, 2016.
- [26] N. ElMocayd, S. Ricci, M. Rochoux, N. Goutal, C. Goeury, S. Boyaval, D. Lucor, and O. Thual. Polynomial surrogates for open-channel flows in random steady state. *Env. Model. Assessment.*, pages 24–45, 2018.
- [27] B. Hourahine et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.*, 152:124101, 2020.
- [28] J. Kim et al. Qmcpack: an open source *ab initio* quantum monte carlo package for the electronic structure of atoms, molecules and solids. *J. Phys.: Condens. Matter*, 30:195901, 2018.
- [29] P. Giannozzi et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter*, 29:465901, 2017.
- [30] P. R. C. Kent et al. QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion Quantum Monte Carlo. *J. Chem. Phys.*, 152:174105, 2020.
- [31] S. Faroux, A. Kaptué Tchuenté, J.-L. Roujean, V. Masson, E. Martin, and P. Le Moigne. Ecoclimap-ii/europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. *Geosci. Model Dev*, 6:563–582, 2013.
- [32] J. C. Fogarty, H. M. Aktulga, A.Y. Grama, A.C.T. van Duin, and S.A. Pandit. A reactive molecular dynamics simulation of the silica-water interface. *Journal of Chemical Physics*, 132:17470, 2010.

D1.2 First scientific results and progress on exascale tools

- [33] Mark A. Friedl, Douglas K. Mclver, John CF Hodges, X. Y. Zhang, D. Muchoney, Alan H. Strahler, Curtis E. Woodcock, Sucharita Gopal, Annemarie Schneider, and Amanda Cooper. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83(1-2):287–302, 2002.
- [34] Pawel Gancarski. Hornamossen test case. <https://windbench.net/hornamossen>.
- [35] P. Giannozzi and et al. Quantum ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21:395502, 2009.
- [36] P. Giannozzi and et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *Journal of Physics: Condensed Matter*, 29:465901, 2017.
- [37] 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.*, 132:154104, 2010.
- [38] Lukas Gudmundsson and Sonia I. Seneviratne. Observational gridded runoff estimates for Europe (E-RUN version 1.0). *Earth System Science Data*, 2016:1–1, 2016.
- [39] L. Hascoët. Automatic differentiation by program transformation, 2007.
- [40] L. Hascoët and V. Pascual. The Tapenade Automatic Differentiation tool: Principles, Model, and Specification. *ACM Transactions On Mathematical Software*, 39(3), 2013.
- [41] W.G. Hoover and B.L. Holian. Kinetic moments method for the canonical ensemble distribution. *Physics Letters A*, 211:253–257, 1996.
- [42] Pankaj Jha, Matthew Churchfield, Patrick Moriarty, and Sven Schmitz. Guidelines for volume force distributions within actuator line modeling of wind turbines on large-eddy simulation-type grids. *Journal of Solar Energy Engineering*, 136:031003, 08 2014.
- [43] Jim E. Jones and Carol S. Woodward. Newton–Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, 24(7):763–774, 2001.
- [44] P.A. Joulin. *Modélisation à fine échelle des interactions entre parcs éoliens et météorologie locale*. PhD thesis, Université de Toulouse, 2019.
- [45] J. Keller, V. Rath, J. Bruckmann, D. Mottaghy, C. Clauser, A. Wolf, R. Seidler, H. M. Bücker, and N. Klitzsch. SHEMAT-Suite: An open-source code for simulating flow, heat and species transport in porous media. *SoftwareX*, 12, 2020.
- [46] Stefan J. Kollet and Reed M. Maxwell. Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7):945–958, 2006.
- [47] Stefan J. Kollet and Reed M. Maxwell. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, 44(2), 2008.
- [48] Wolfgang Kurtz, Guowei He, Stefan J. Kollet, Reed M. Maxwell, Harry Vereecken, and Harrie-Jan Hendricks Franssen. TerrSysMP–PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface–subsurface model. *Geoscientific Model Development*, 9(4):1341–1360, 2016.
- [49] David Legates and Gregory McCabe. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35:233–241, 01 1999.

D1.2 First scientific results and progress on exascale tools

- [50] C. Lin, F. H. Zong, and D. M. Ceperley. Twist-averaged boundary conditions in continuum quantum monte carlo algorithms. *Phys. Rev. E*, 64:016702, 2001.
- [51] D. Malcolm and A. Hansen. Windpact turbine rotor design study. *National Renewable Energy Laboratory, Golden, CO*, 2002.
- [52] S. Markov, G. Penazzi, Y. H. Kwok, A. Pecchia, B. Aradi, T. Frauenheim, and G.H. Chen. Permittivity of oxidized ultra-thin silicon films from atomistic simulations . *IEEE Elct. Dev. Lett.*, 36:1076–1078, 2015.
- [53] R. M. Maxwell, L. E. Condon, and S. J. Kollet. A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geoscientific Model Development*, 8(3):923, 2015.
- [54] Riccardo Mereu, Stefano Passoni, and Fabio Inzoli. Scale-resolving cfd modeling of a thick wind turbine airfoil with application of vortex generators: Validation and sensitivity analyses. *Energy*, 187:115969, 2019.
- [55] Bibi S. Naz, Stefan Kollet, Harrie-Jan Hendricks Franssen, Carsten Montzka, and Wolfgang Kurtz. A 3 km spatially and temporally consistent European daily soil moisture reanalysis from 2000 to 2015. *Sci Data*, 7(1):111, December 2020.
- [56] Bibi S. Naz, Wolfgang Kurtz, Carsten Montzka, Wendy Sharples, Klaus Goergen, Jessica Keune, Huilin Gao, Anne Springer, Harrie-Jan Hendricks Franssen, and Stefan Kollet. Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation. *Hydrology and Earth System Sciences*, 23(1):277–301, January 2019.
- [57] Lars Nerger and Wolfgang Hiller. Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, 55:110–118, June 2013.
- [58] K. W. Oleson, G.-Y. Niu, Z.-L. Yang, D. M. Lawrence, P. E. Thornton, P. J. Lawrence, Reto Stöckli, R. E. Dickinson, G. B. Bonan, and Samuel Levis. Improvements to the Community Land Model and their impact on the hydrological cycle. *Journal of Geophysical Research: Biogeosciences*, 113(G1), 2008.
- [59] Herbert Owen. Towards exascale for low carbon energy. In *High Performance Innovation Conference, Online*, 2020.
- [60] Herbert Owen, Georgios Chrysokentis, Matias Avila, Daniel Mira, Guillaume Houzeaux, Ricard Borrell, Juan Carlos Cajas, and Oriol Lehmkuhl. Wall-modeled large-eddy simulation in a finite element framework. *International Journal for Numerical Methods in Fluids*, 92(1):20–37, 2020.
- [61] J. P. Perdew. Density functional theory and the band gap problem. *Physical Review Letters*, 28:497, 1985.
- [62] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation made simple. *Physical Review Letters*, 77:3865, 1996.
- [63] S. Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117:1–19, 1995.
- [64] G. Psfogiannakis and A. C. T. van Duin. Development of a ReaxFF reactive force field for Si/Ge/H systems and application to atomic hydrogen bombardment of Si, Ge, and SiGe (100) surfaces. *Surface Science*, 646:253–260, 2016.

D1.2 First scientific results and progress on exascale tools

- [65] A. Rasch and H. M. Bückner. EFCOSS: An interactive environment facilitating optimal experimental design. *ACM Transactions on Mathematical Software*, 37(2):1–37, 2010.
- [66] V. Rath, A. Wolf, and H. M. Bückner. Joint three-dimensional inversion of coupled groundwater flow and heat transfer based on Automatic Differentiation: sensitivity calculation, verification, and synthetic examples. *Geophysical Journal International*, 167(1):453–466, 2006.
- [67] Florian Rathgeber, David A Ham, Lawrence Mitchell, Michael Lange, Fabio Luporini, Andrew TT McRae, Gheorghe-Teodor Bercea, Graham R Markall, and Paul HJ Kelly. Firedrake: automating the finite element method by composing abstractions. *ACM Transactions on Mathematical Software (TOMS)*, 43(3):1–27, 2016.
- [68] K. G. Reeves, A. Serva, G. Jeanmairet, and M. Salanne. A first-principles investigation of the structural and electrochemical properties of biredox ionic species in acetonitrile. *Phys. Chem. Chem. Phys.*, 22:10561–10568, 2020.
- [69] S. Ricci. *Quantification et réduction des incertitudes par assimilation de données en géosciences – Application à l’hydraulique à surface libre*. Hdr, CECI, Université de Toulouse, CNRS, CERFACS, 2019.
- [70] E. Rodriguez, C. Morris, J. Belz, E. Chapin, J. Martin, W. Daffer, and S. Hensley. An assessment of the srtm topographic products. *Technical Report JPL D-31639, Jet Propulsion Laboratory, Pasadena, California*, 2005.
- [71] Alexandre M Roma, Charles S Peskin, and Marsha J Berger. An adaptive version of the immersed boundary method. *Journal of Computational Physics*, 153(2):509 – 534, 1999.
- [72] P. Roy. *Uncertainty Quantification in High Dimensional Problem*. Phd thesis, Institut National Polytechnique de Toulouse - Ecole doctorale MEGeP, 5 2019.
- [73] P. Roy, N. ElMocayd, S. Ricci, M. DeLozzo, M. Rochoux, J-C. Jouhaud, and N. Goutal. Comparison of polynomial chaos and gaussian process surrogates for uncertainty quantification and correlation estimation of spatially-distributed open-channel steady flows. *Stochastic Environmental Research and Risk Assessment*, pages 24–45, 2017.
- [74] Pedros Santos. Alex17 diurnal cycles benchmark: a large domain in complex terrain, 2019. <https://thewindvaneblog.com/alex17-diurnal-cycles-benchmark-a-large-domain-in-complex-terrain-b5029e94485>.
- [75] William Robert Saunders, James Grant, and Eike Hermann Müller. A domain specific language for performance portable molecular dynamics algorithms. *Computer Physics Communications*, 224:119–135, 2018.
- [76] William Robert Saunders, James Grant, Eike Hermann Müller, and Ian Thompson. Fast electrostatic solvers for kinetic monte carlo simulations. *Journal of Computational Physics*, 410:109379, 2020.
- [77] Marcel G. Schaap and Feike J. Leij. Database-related accuracy and uncertainty of pedotransfer functions. *Soil Science*, 163(10):765–779, 1998.
- [78] R. Seidler, H. M. Bückner, M. A. Rostami, and D. Neuhäuser. On the design of the EFCOSS software architecture when using parallel and distributed computing. In *Proceedings of the 9th International Conference on Software Engineering and Applications*, pages 445–454. SCITEPRESS - Science and Technology Publications, 29.08.2014 - 31.08.2014.

D1.2 First scientific results and progress on exascale tools

- [79] R. Seidler, K. Padalkina, H. M. Bücke, A. Ebigbo, M. Herty, G. Marquart, and J. Niederau. Optimal experimental design for reservoir property estimates in geothermal exploration. *Computational Geosciences*, 20(2):375–383, 2016.
- [80] H. Shin, Y. Luo, A. Benali, and Y. Kwon. Diffusion monte carlo study of o_2 adsorption on single layer graphene. *Phys. Rev. B*, 100:075430, 2019.
- [81] P. Shrestha, M. Sulis, M. Masbou, S. Kollet, and C. Simmer. A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. *Monthly weather review*, 142(9):3466–3483, 2014.
- [82] Richard J.A.M. Stevens, Luis A. MartÁnez-Tossas, and Charles Meneveau. Comparison of wind farm large eddy simulations using actuator disk and actuator line models with wind tunnel experiments. *Renewable Energy*, 116:470 – 478, 2018.
- [83] Reto Stöckli, Anke Duguay–Tetzlaff, Bojanowski Jędrzej, Rainer Hollmann, Petra Fuchs, and Martin Werscheck. Cm saf cloud fractional cover dataset from meteosat first and second generation - edition 1 (comet ed. 1), 2017.
- [84] MUMPS team. Multifrontal massively parallel solver. <http://mumps-solver.org>.
- [85] W. Wagner, Wouter Dorigo, Richard de Jeu, Diego Fernandez, Jerome Benveniste, Eva Haas, and Martin Ertl. Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals)*, 7:315–321, 2012.
- [86] K. Yoshikawa, W. Yoshida, T. Irie, H. Kawasaki, K. Konishi, H. Ishibashi, T. Asatani, D. Adachi, M. Kanematsu, H. Uzu, and K. Yamamoto. Exceeding conversion efficiency of 26% by heterojunction interdigitated back contact solar cell with thin film Si technology. *Solar Energy Materials and Solar Cells*, 173:37, 2017.
- [87] A. Nikolaos Ziogas, T. Ben-Nun, G. Indalecio Fernández, T. Schneider, M. Luisier, and T. Hoeffler. data-centric approach to extreme-scale ab initio dissipative quantum transport simulations. *SC '19: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–13, 2019.

A Annex A: Summary of delays and risk assessment

The table in this annex show the overall description of the various delays incurred in the project across all five Scientific Challenges. The table describe such delays per subtask. It gives a brief description of the delays, their effect. It then gives a risk assessment using a typical criterion based on two parameters: impact and probability of such an impact. The risk is divided in 3 categories: high, medium and low. A final entry in the table describe the mitigation strategy adopted.

Table 20: Summary of delays and risk assesment listed by task.

Task #	Description	Partner	Reason or Cause	Effect	Prob.	Impact	Risk	Mitigation strategy
T1.1.1	Development, Verification and Validation for Complex terrain.	BSC	Sub-normal work conditions with a 6 year old kid at home (no school). It means reduction of available work time and concentration.	Difficulty to deliver all the work that had been planned.	75%	Moderate. Some of the promised deliverable may not be achieved by the end of the project.	High	Extend the duration of the project and PMs proportionately to the amount of time in which we work in sub-normal conditions.
T1.1.2	Development, Verification and Validation for Full rotor	BSC	Sub-normal work conditions with a 6 year old kid at home (no school). It means reduction of available work time and concentration.	Difficulty to deliver all the work that had been planned.	75%	Moderate. Some of the promised deliverable may not be achieved by the end of the project.	High	Extend the duration of the project and PMs proportionately to the amount of time in which we work in sub-normal conditions.
T1.2.1-3	Joint assimilation and the integration of MELISSA	FZJ	The no schooling for children during the restriction period will decrease the working efficiency	The tasks will be affected due to the reduction of working hours.	40%	Minor. Some of the promised deliverable may not be achieved by the end of the project.	Low	Extend the duration of the task of 2-3 PMs
T1.2.2	Wind and solar power calculation for meteorological ensembles	FZJ	The no schooling for children during the restriction period will decrease the working efficiency	The tasks will be affected due to the reduction of working hours.	40%	Minor. Some of the promised deliverable may not be achieved by the end of the project.	Low	Extend the duration of the task of 2-3 PMs
T1.2.2-2	Confluence of ultra-large ensembles with cloud-tracking for short-term forecasting.	IEE	Late start due to hiring process in T.1.2.1.	Compression of timeline exacerbates conflicts with other projects. Serial tasks with partners become very critical. Less to show in mid-term report. Small loss of resources for additional SC planning.	65%	Moderate. Some of the promised deliverable may not be achieved by the end of the project. (i.e. particle-filter feed-back).	Medium	Active planning of SC and contingencies to stay on schedule. (Stretching project would mitigate problems).

T1.2.3	Probabilistic calibration of ensemble predictions	MF	Non-permanent staff needs to be hired on one-year contract for task.	Hiring process brings uncertainty. Timeframe of task is compressed from original timeline, but not more so than from data delivery from T.1.2.1.	50%	Minor. Delays up to 10% of schedule and/or additional cost up to 10% of task	Low	September delivery of data from T1.2.1 to maximize use of resources.
T1.3.1-3	Non-equilibrium Green's functions transport properties c-Si/a-Si:H interface	FZJ	Home office and familial duties (schooling own child) have reduced working efficiency. Major change in the flagship code introduced additional challenges. Maternity leave of Irene Aguilera introduced one year delay on the final deliverables.	The simulation of medium to large interfaces may not be carried out before the mid term deliverable report. Final targets may not be reached.	60%	Moderate. Some of the promised deliverable may not be achieved by the end of the project.	Medium	Shift part of the workload to A. Pecchia group. Alternatively extend the duration of the task of 3-4 PMs
T1.3.3	Organic and perovskite solar cells	UBAH	Lockdown due to COVID-19 has been in place in the UK for the past two months, and is very likely to continue for at least another month. This has led to decreased productivity of individual researchers due to sub-optimal home working conditions, and also decreased team productivity due to the difficulty in performing tasks which require co-ordinated efforts.	Delay in deliverables and/or scaling back of features to be implemented in the flagship codes, or scientific applications of the codes.	70%	Minor. Some of the promised deliverable may not be achieved by the end of the project.	Low	Extend the duration of the project for 2-3 PMs
T1.4.1-2	Model Evaluation and validation	FZJ	Because of homeoffice and home schooling of children aged 6 and 11, the work efficiency is reduced	This task might be delayed by 2-3 months	20%	Insignificant. The promised deliverable can still be achieved by the end of the project.	Low	Shift part of the workload to later project months.
T1.4.1-3	Uncertainty Quantification using Data Assimilation	FZJ	Because of homeoffice and home schooling of children aged 6 and 11, the work efficiency is reduced.	The finalization of this task might be delayed by 2-3 months.	40%	Minor. Some of the promised deliverable may not be achieved by the end of the project.	Low	Extend the duration of the task of 2-3 PMs

T1.4.2-4	Set-up of the model in the Italian Alpine region	UNITN	Difficulties in contacting hydrological offices or water managers for requests of details concerning human infrastructures + family issues of task owner due to covid emergency	The simulation of hydropower production in the Italian Alpine region may be delayed of few months	30%	Minor. Some of the promised deliverable may not be achieved by the end of the project.	Low	Extend the duration of the task of 2-3 PMs
T1.4.4-6	Setting-up a test-model suite and extending OED functionality	RWTH	No childcare for my 3 and 5 year old children due to Covid-19 pandemic. Thus, reduced work time and efficiency.	Finalization of this task will be delayed compared to the original timeline presented in D1.1.	100%	Low. This deliverable can still be achieved by the end of the project.	Low	Shift part of the workload to later project months.
T1.4.4-7	Optimizing OED workflow	RWTH	No childcare for my 3 and 5 year old children due to Covid-19 pandemic. Thus, reduced work time and efficiency.	Finalization of this task will be delayed compared to the original timeline presented in D1.1.	100%	Low. This deliverable can still be achieved by the end of the project.	Low	Shift part of the workload to later project months.
T1.4.5	Modeling of geothermal reservoir system	RWTH	No childcare for my 3 and 5 year old children due to Covid-19 pandemic. Thus, reduced work time and efficiency.	This task might not be finalized by the end of the project.	80%	Moderate. Some of the promised deliverable may not be achieved by the end of the project.	High	Extend the duration of the project and PMs proportionately to the amount of time in which we work in sub-normal conditions.
T1.5.1-1	Identify minimal set of equations for ion turbulence	CEA-IRFM	Lack of black board style interactions (Covid-19)	Slower activity, delayed results	65%	Minor. Main results should be achieved by the end of the project	Low	None
T1.5.1-2	Comparative efficiencies of flux-surface aligned and Cartesian mesh grids in poloidal plane, and of patches with regular or irregular meshes	CEA-IRFM / MPG	Hackathon on code performance analysis and optimization cancelled (Covid-19). Not replaced by online courses so far	No local expertise on code optimization tools	65%	Moderate. Possible medium HPC performance within the time-frame of EoCOE-II	Medium	None. Code performance would be improved after the end of the project

T1.5.2-1	Handle the X-point of ITER-relevant magnetic configurations with multi-patch or flux coordinate independent schemes	CEA-IRFM / MPG	Delayed hiring of post-doc by 2 months at least (Covid-19)	Slower activity, delayed results.	50%	Moderate. May impact the achievable physics objectives of Task 1.5.3-2	Medium	Revise physics objectives of T1.5.3-2. Publications in journals with lower impact factor
T1.5.2-2	Multi-patch treatment of the Vlasov equation to handle large variations of temperature	CEA-IRFM	CEMRACS of July-August 2020 is cancelled (Covid-19)	Full-time 4 week collaborative work abandoned.	70%	Major. Critical to GYSELA-X. May prevent to consider relevant temperature profiles from core to far edge	High	Organize interactive collaboration involving expected members of CEMRACS
T1.5.3-1	Recover large scale equilibrium flows expected in the SOL	CEA-IRFM	M2 Internship could not start on this task because of no access to numerical resources (Covid-19).	Delay the study of plasma-wall interaction with kinetic electrons	65%	Minor delay by 2-3 months	Low	None
T1.5.3-2	Highlight the critical interplay between core, edge and SOL turbulence and transport in (some of) the ITER relevant parameter range	CEA-IRFM	Difficulty to get access to numerical resources (Covid-19)	Dealyed runs. Reduced expertise on the behaviour of new run types.	70%	Moderate	Medium	Revise objectives. Publications in journals with lower impact factor

B Annex B: Extra plots for subtasks T1.3.1-1 & T1.3.1-2

This annex displays a number of plots and figures in addition to the one already illustrated for the subtask T1.3.1-1 and T1.3.1-2 in Sec. 5.1.

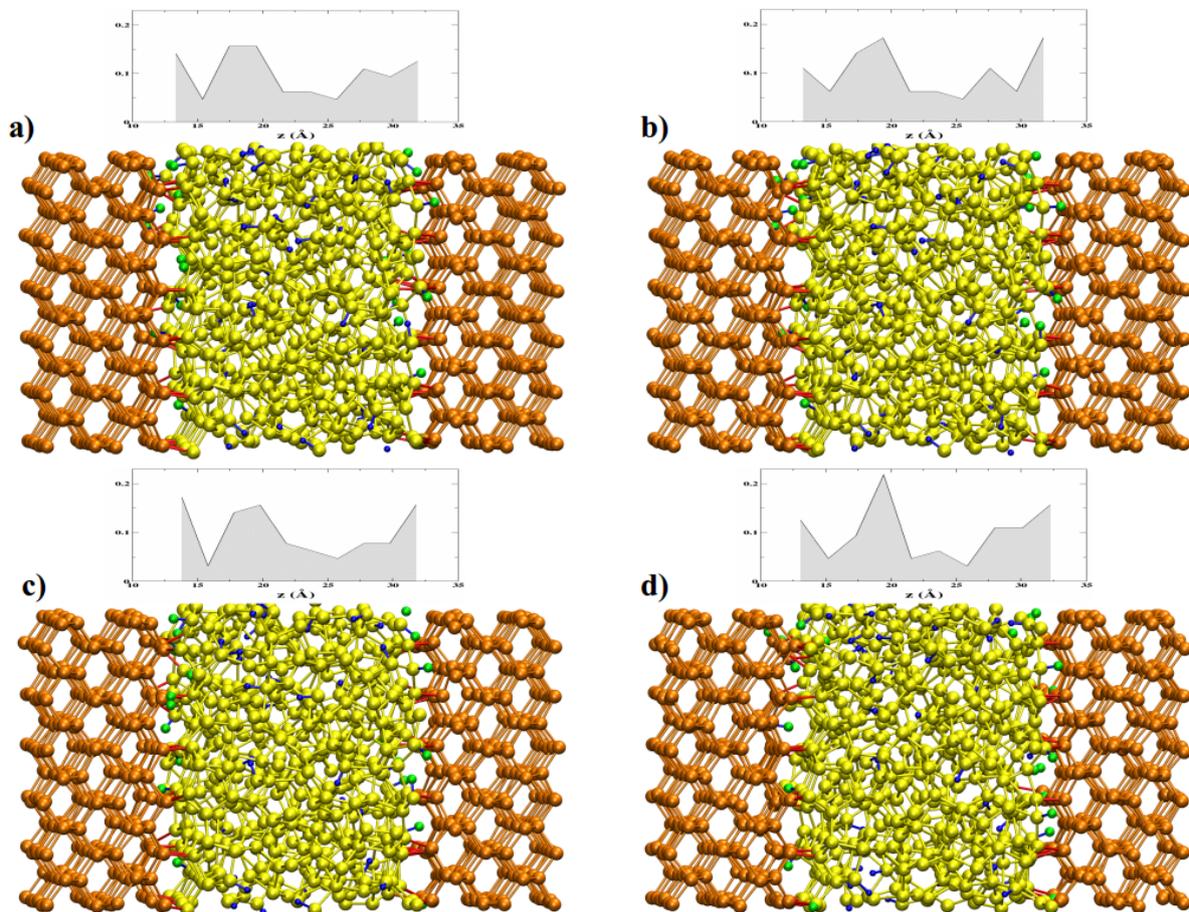


Figure 60: Snapshots of the a-Si:H/c-Si interfaces at a) T=300 K; b) T=500 K; c) T=700 K; and d) T=900 K. Above the configurations, the distribution of the hydrogen atoms along the z direction are reported.

c-Si, T = 500 K

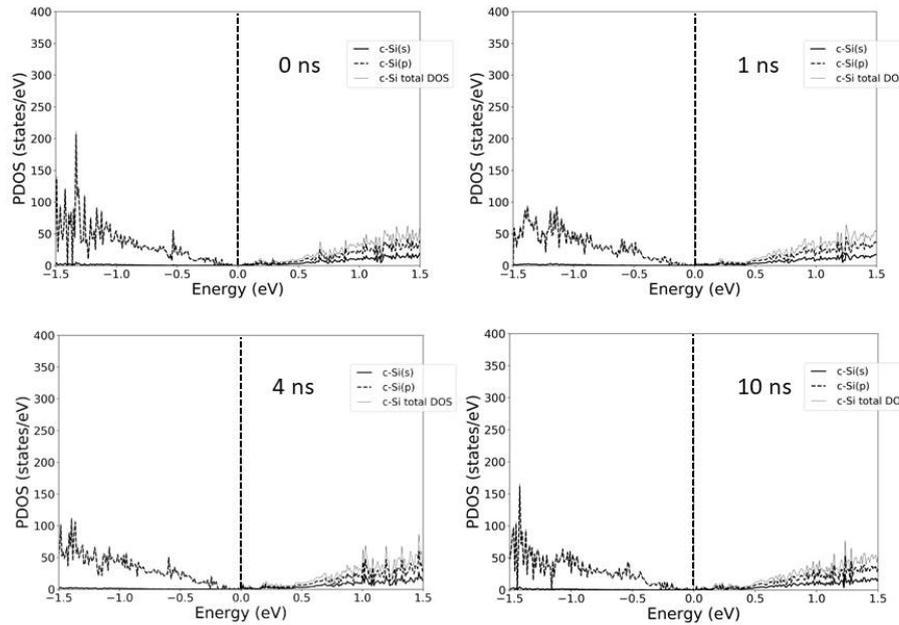


Figure 61: Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 500 K. The vertical dashed line at 0 eV evidences the Fermi energy.

c-Si, T = 700 K

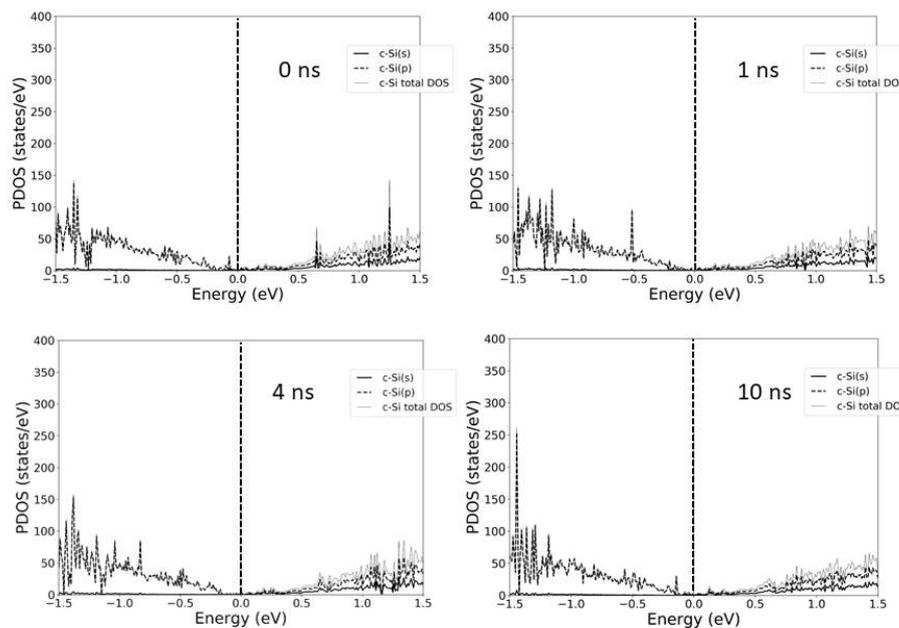


Figure 62: Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 700 K. The vertical dashed line at 0 eV evidences the Fermi energy.

c-Si, T = 900 K

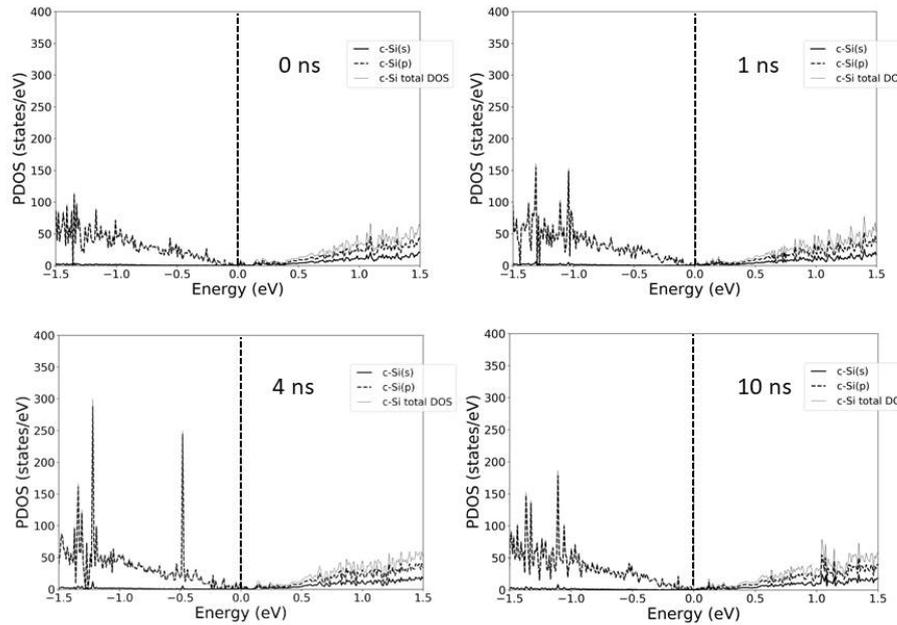


Figure 63: Projected density of states of crystalline silicon at 0, 1, 4 and 10 ns for T= 900 K. The vertical dashed line at 0 eV evidences the Fermi energy.

a-Si, T = 500 K

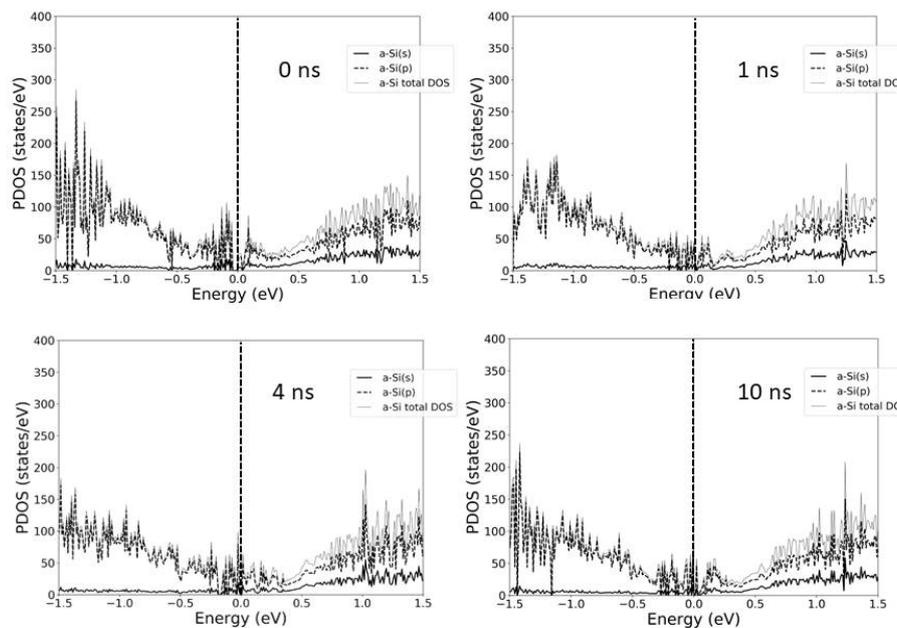


Figure 64: Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 500 K. The vertical dashed line at 0 eV evidences the Fermi energy.

a-Si, T = 700 K

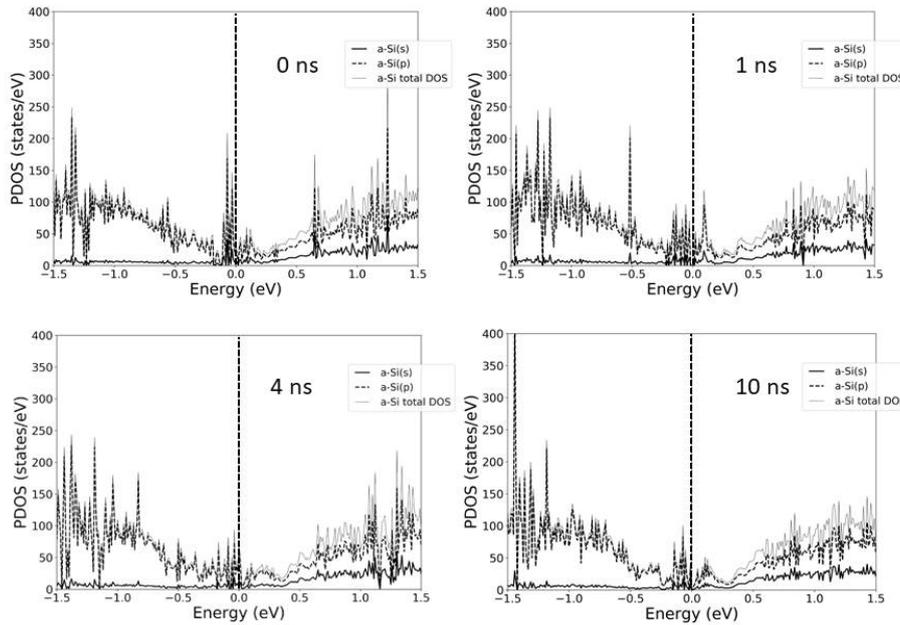


Figure 65: Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 700 K. The vertical dashed line at 0 eV evidences the Fermi energy.

a-Si, T = 900 K

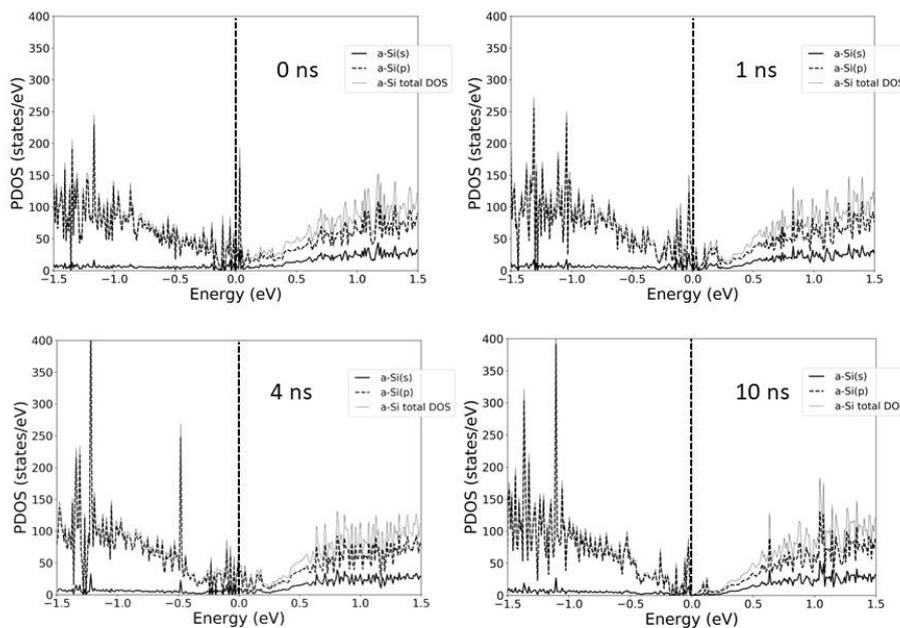


Figure 66: Projected density of states of amorphous silicon at 0, 1, 4 and 10 ns for T= 900 K. The vertical dashed line at 0 eV evidences the Fermi energy.

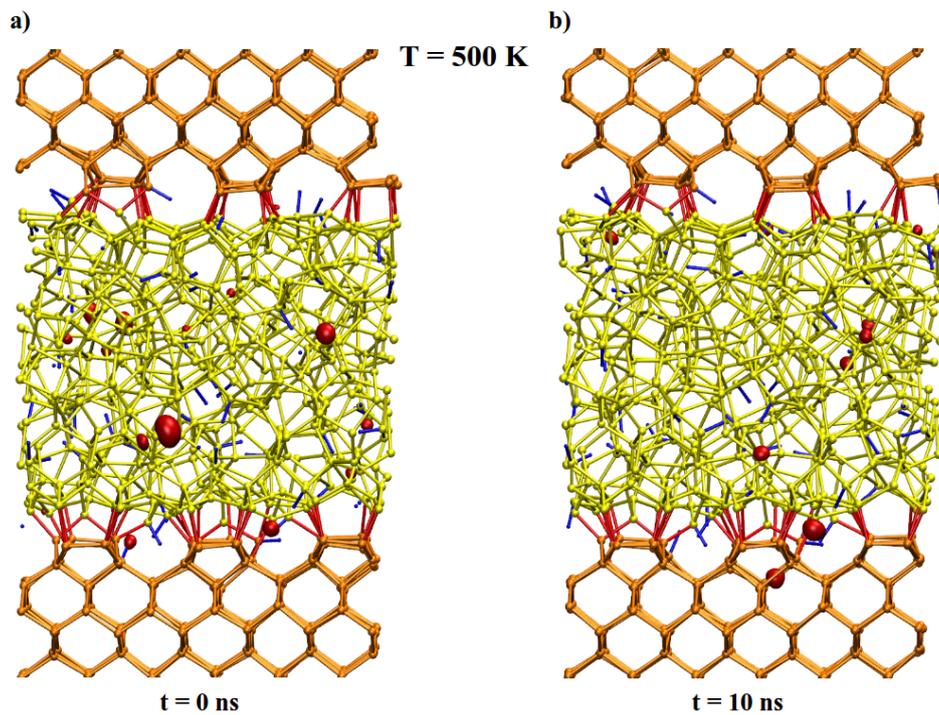


Figure 67: Local density of states of the intragap states of the a-Si:H/c-Si interfaces at the beginning a) and at the end b) of the thermalisation at $T= 500$ K.

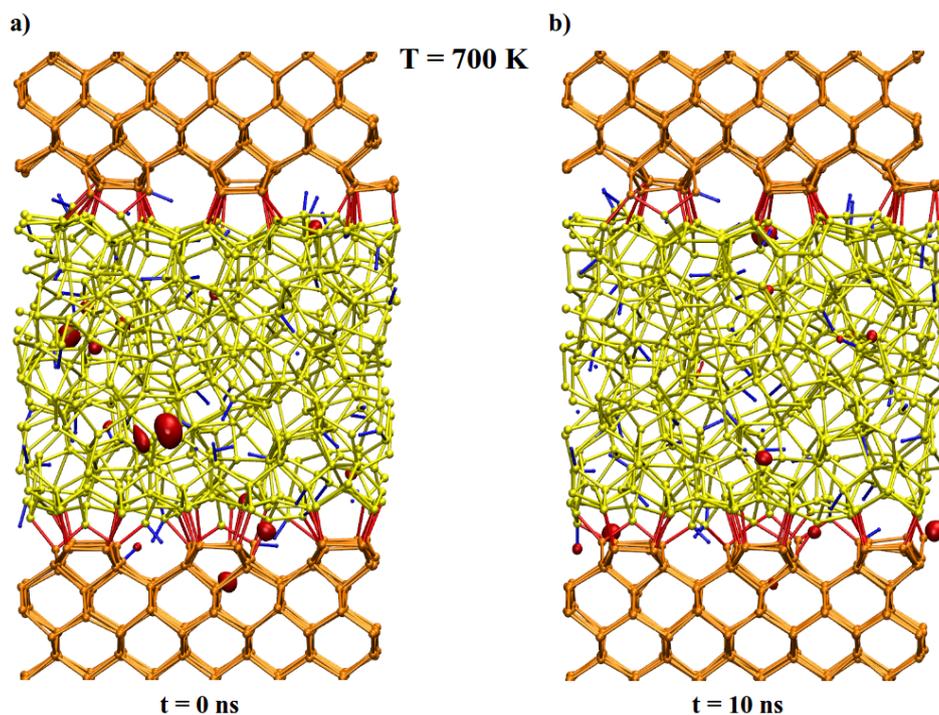


Figure 68: Local density of states of the intragap states of the a-Si:H/c-Si interfaces at the beginning a) and at the end b) of the thermalisation at $T= 700$ K.

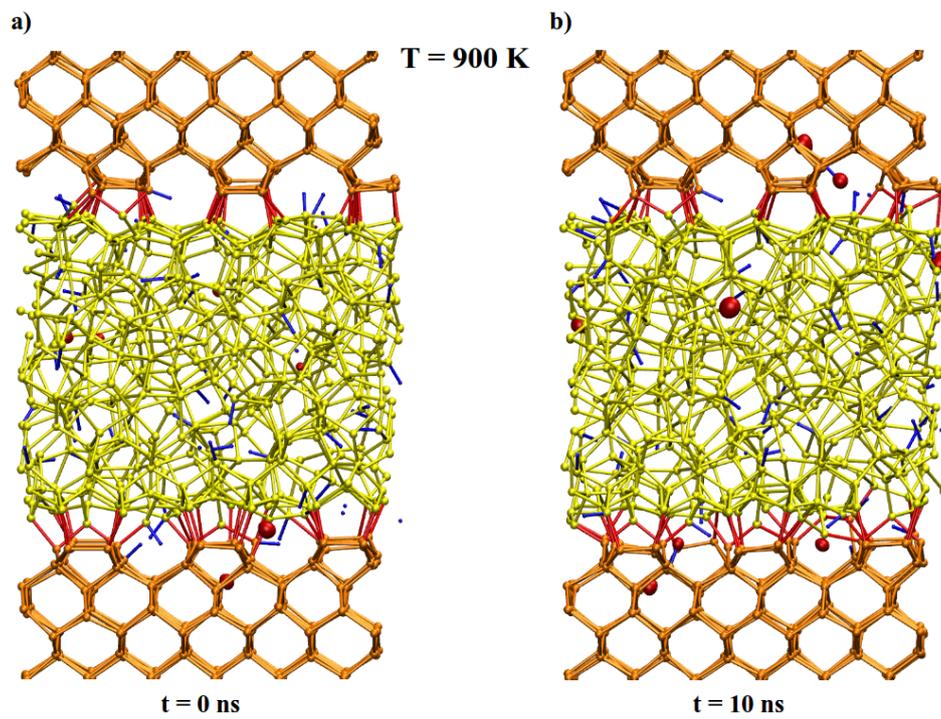


Figure 69: Local density of states of the intragap states of the a-Si:H/c-Si interfaces at the beginning a) and at the end b) of the thermalisation at T= 900 K.