VOLUME 1 ◆ OCTOBER 2021

# NOMAD Center of Excellence

NOVEL MATERIALS DISCOVERY

*- NEWSLETTER -*

## WHAT ARE THE GOALS?

This first newsletter of the NOMAD Center of Excellence (CoE) starts with a brief description of its structure and bigger goals. The latter concern the development of novel materials that are urgently needed for solving the threatening environmental, economical, and energy challenges of the world society. Specifically, we will introduce concepts and methods that bring computational materials science to a new level by exploiting upcoming exascale technologies and advancing the handling and analyses of extreme-scale data by artificial intelligence (AI) tools.

The NOMAD CoE expands basic science and engineering and engages with scientists across many countries and across generations to promote innovations and further careers, and it reaches out to industry and society. For this purpose, we also invite students and experts to our tutorial series, interactive online seminars with introductory videos, hands-on exercises, and much more.

We already offer a working ***data infrastructure***, including the ***NOMAD Repository, Archive, Encyclopedia***, and ***AI Toolkit***: https://nomad-lab.eu/. This NOMAD ***Laboratory*** contains the biggest data store in computational materials science, and it is now being enhanced successively by including data from materials synthesis and experiments. The future CoE component of the Laboratory, https://www.nomad-coe.eu/, addresses the mentioned exascale computing and extreme-scale data.

By exploiting the capabilities of the upcoming high-performance computer generation, we will finally address predictions of novel materials for pressing energy, environmental, and societal challenges (two examples will be discussed below). In general, we will ***bring ab initio computational materials science to highly complex systems and realistic conditions.***

## THE COE'S GENERAL STRUCTURE

The NOMAD CoE is organized into three pillars and divided into 13 work packages (WPs), see Fig. 1.

- The goal of Pillar 1 with its three work packages (WPs 1-3) is ***to advance ab initio computational materials science not just for a few special software packages but for entire code families***. This enables tackling more complex problems with higher accuracy and precision than what is possible today.
- Efficient use of the exascale-ready libraries and codes developed in Pillar 1 requires ***sophisticated workflows*** that are capable of managing high-throughput computations and taking full advantage of exascale resources. This is the focus of WPs 4 and 5 of Pillar 2.
- The overall aim of Pillar 3 (WP 6) is to utilize exascale technology for ***advancing existing big-data AI tools and bringing them towards near-real-time performance***.
- WP 7 supports the three pillars, making the NOMAD data infrastructure exascale-ready.
- The task of WP 8 is the co-design of the elaborate *ab initio* software packages and high-performance computers (HPCs) to deliver new inputs to HPC architects. The aim is to ensure that hardware and software developments go hand in hand.

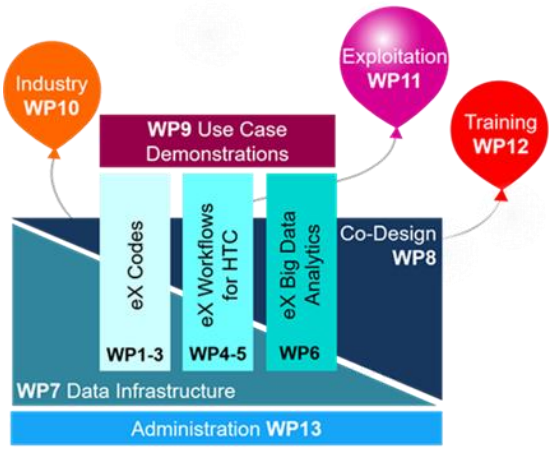


Fig. 1: The NOMAD CoE consists of 3 Pillars and 13 Workpackages.

- Our Use-Case Demonstrators (WP 9) will showcase the NOMAD CoE developments for urgent energy and environmental challenges with the examples of catalytic water splitting for hydrogen production and waste-heat recovery.
- WPs 10-13 comprise training, administration, and outreach to academia and industry.

## WHO WE ARE?

The NOMAD CoE represents an ***international, dynamic team*** organized in a 2-shell structure (see Fig. 2). The "first shell" represents the core scientists from ten academic and HPC computing centers across Europe. Renowned researchers from other materials-science groups, industry, and other stakeholders form the second shell.


Fig. 2: Participants in the NOMAD CoE Consortium: The first shell (inner circle) and the second shell (outer ring).

## CORE TEAM

- **Max Planck Society**:
  Matthias Scheffler (Coordinator, Fritz Haber Institute)
  Erwin Laure (Max Planck Computing and Data Facility)
- **Humboldt-Universität zu Berlin**:
  Claudia Draxl (Deputy Coordinator)
- **Aalto University**: Patrick Rinke
- **Barcelona Supercomputing Center**: José-Maria Cela
- **Commissariat à l'Énergie Atomique et aux Énergies Alternatives**: Christine Menache
- **CSC–IT Center for Science**: Kimmo Koski
- **Technical University of Denmark**:
  Kristian Sommer Thygesen
- **Technische Universität Wien**: Andreas Grüneis
- **Université Catholique de Louvain**: Xavier Gonze
- **University of Cambridge**: Gábor Csányi
- **University of Latvia:** Andris Gulans
- **University of Warwick**: James Kermode

## HOW TO FIND THE URGENTLY NEEDED MATERIALS?

To achieve the critical climate targets by primarily reducing the worldly $CO_2$ emission, we generally need new technologies that enable a $CO_2$-neutral energy production. Our contribution and responsibility on this challenge is to identify and develop new materials that serve as catalysts for a ***sustainable production of large quantities of hydrogen as a renewable and clean energy carrier***. This is one of our use-case demonstrators. The other one addresses the significant amount of energy that is wasted in the form of heat. Here we are searching for thermoelectric materials that can recover some of this lost energy by transforming heat into electricity.

### PHOTOCATALYTIC WATER SPLITTING

The sustainable production of hydrogen is of paramount importance for the production of chemicals and fuels. To this end, photocatalytic water splitting could become a key technology but has so far been unfeasible due to limitations of the employed materials. The key reaction: $2H_2O + \text{photons} \rightarrow 2H_2 + O_2$ requires (i) a stable semiconductor to harvest the solar photons and convert them into electron-hole pairs and (ii) a stable and efficient oxygen-evolution-reaction catalyst to split the water molecules and evolve $H_2$ and $O_2$ from the electron-hole pairs.
The discovery of new catalyst materials faces a number of challenges concerning the complexity of the materials space and the difficulties of describing, with sufficient accuracy, both the interactions of water with catalytic surfaces and the band gaps of the photo-absorbers. Standard density-functional theory (DFT) does not provide the

required accuracy, while beyond DFT methods are computationally too demanding with present algorithms and computers. To advance the field of photocatalytic water splitting, we will perform machine-learning-assisted high-throughput screening of novel materials with beyond-DFT methods, enabled by the developments within Pillars 1-3.

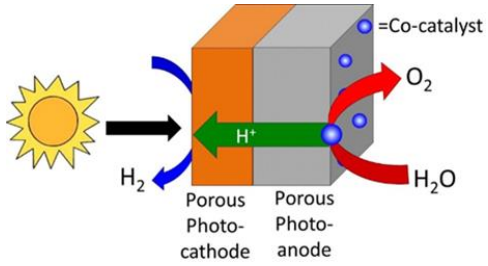


Fig. 3: Identification of a better catalyst for water splitting has enormous potential in carbon-dioxide management, in hydrogen-based energy, and more.

## WASTE-HEAT RECOVERY

Over 40% of our current energy consumption is wasted in the form of heat. Thermoelectric devices can convert heat into electricity, be it from industrial ovens, chemical plants, exhaust pipes, or computers. Thus, they have the potential to reduce the world's energy consumption and, hence, to facilitate the transition to a carbon-neutral economy.

More efficient and affordable ***thermoelectric materials*** are needed to make such devices economically viable and to boost their widespread deployment. Designing high-performance thermoelectric materials is a formidable task since the thermoelectric efficiency depends on subtle details of the charge and heat transport. Our developments in WP 2 will enable us to compute accurate, beyond-DFT electronic structures and to obtain electron mobilities fully from first principles. Similarly, the advances in WP 1 will allow an accurate and computationally efficient assessment of vibrational conductivities from *ab initio* molecular dynamics simulations.

Performing this kind of calculation in a high-throughput fashion by making use of the exascale workflows developed in WPs 4 and 5 will yield an exhaustive database of thermoelectric efficiencies with unmatched accuracy. Concurrently, these results will be data-mined using the high-performance AI techniques developed in WP 6, so to accelerate the exploration and identification of high-performance thermoelectrics.

## WHERE ARE WE ON OUR ROUTE?

First successful steps have already been taken in the last year. Importantly, newly developed methodologies and advances are already available to the public:

- Efficient ***eigensolvers*** are key to perform rapid large-scale DFT calculations. The 2021 release of the ***ELPA*** eigensolver allows harvesting the computational power provided by the most advanced (pre-)exascale technologies, including GPUs from Nvidia, Intel, and AMD as well as the ARM-based architecture used in Fugaku, the fastest supercomputer available to date.
- The newly developed ***Atomic Simulation Recipes (ASR)*** package allows for performing advanced computational tasks – often involving several concurrent or sequential first-principles calculations – in a reliable, flexible, and customizable high-throughput fashion. Examples are the assessment of Raman spectra, electronic effective masses, or energy barriers.
- The ***python package FHI-vibes*** provides the foundations for calculating and understanding the vibrational properties of solids from first principles, including heat transport. It seamlessly bridges between perturbative methods based on the harmonic approximation and fully anharmonic (Green-Kubo) molecular dynamics simulations.
- The ***SISSO++ package*** provides a massively parallel implementation of the sure-independence screening and sparsifying operator (SISSO) method. This AI technique can identify descriptors that reproduce or classify a target data set best. In the past, SISSO has proven successful in modeling and predicting, for example, the stability of phases, catalytic activity and reactivity, glass transition temperatures, and the topological character of materials.
- ***Coupled-cluster (CC)*** theory is considered the gold standard of quantum chemistry. It is *THE* benchmark

for approximate theories when experimental data are not available. The NOMAD team has developed an interface to the coupled-cluster-for-solids (CC4S) code. Thus, plane-wave pseudopotential, as well as all-electron local basis set codes, are covered already.

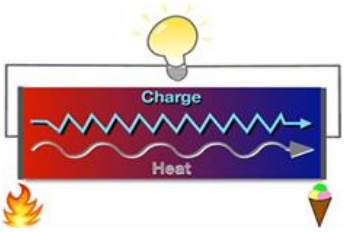


Fig. 4: In thermoelectric materials, a temperature gradient induces a charge flux that can be used to obtain a useful electrical current from otherwise wasted heat.

## COME AND MEET US!

Watch our series of tutorials and hands-on exercises, covering the various aspects of NOMAD! They were designed as interactive online seminars with introductory videos, hands-on exercises, and live Q&A zoom sessions. All materials are available on our website that invites you to learn more about the importance of data handling and state-of-the-art analysis methods. With more than 700 registrations and up to 120 participants in each event, the start of this tutorial series was a great success. We will continue and expand the program soon. Let us sketch some of our ***recently released tutorials:***

- **NOMAD Repository and Archive – Sharing, Publishing, and Managing Computational Materials Science Data with NOMAD**

The main focus of this tutorial is the FAIR (**F**indable, **A**ccessible, **I**nteroperable, and **R**eusable) sharing of materials-science data and how to do it with NOMAD. Markus Scheidgen covers the publication of new data and the exploration and download of NOMAD's existing data; both through our browser-based interface and API.

- **NOMAD Materials Encyclopedia**

The main focus of this second tutorial is on material exploration using NOMAD. In contrast to the NOMAD Archive that provides tools for exploring individual calculations, a complementary, materials-oriented data organization is provided by the NOMAD Materials Encyclopedia. Besides highlighting the key difference between these two approaches, Lauri Himanen discusses how this new layer of information brings additional value to the data.

- **Artificial Intelligence Toolkit**

This tutorial is focused on the analysis of data, introducing the functionalities of the NOMAD AI Toolkit. Luca Ghiringhelli introduces this web-based framework that allows for querying, filtering, and performing AI analysis on the data contained in the NOMAD Archive and explains the user interface that is based on the Jupyter notebook environment.

- **Workflow Management**

In this tutorial, the Atomic Simulation Recipes (ASR) is introduced by Kristian Thygesen. It is an open-source Python frame-work that is ideally suited for high-throughput studies. The FireWorks workflow manager, a free, open-source code for defining, managing, and executing workflows is explained by Geoffroy Hautier.

- **The Open Databases Integration for Materials Design (OPTIMADE)**

This tutorial, held by Gian-Marco Rignanese, is focussed on the usage of the OPTIMADE API that provides access to information from multiple databases. Like in the other tutorials, the participants can try out the capabilities of this tool, practice the different queries and analyze the responses.

## CONGRATULATIONS FAIRmat!

The FAIRmat project – FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids – is an off-spring of the NOMAD concept. It widens the computational materials science focus and includes materials synthesis and experiments. FAIRmat is being funded as a consortium of the German National Research-data infrastructure initiative (NFDI), starting October 1st, 2021. The initial runtime is 5 years (possible prolongation by another 5 years) and the funding is 3.3 Mio Euro/year. FAIRmat pursues a user-driven approach to develop easy-to-use tools and an infrastructure towards FAIR data processing, storage, curation, sharing, and AI readiness for future use of materials data. FAIRmat builds

on the extensive experience with the NOMAD Laboratory and thus the achievements of NOMAD CoE's first funding period and it will closely collaborate with the NOMAD CoE. Claudia Draxl with Matthias Scheffler as her deputy leads the FAIRmat consortium. For more information, visit its webpage: https://fair-di.eu/fairmat/.


Fig. 5: General Structure of the FAIRmat

## UPCOMING AND RECENT EVENTS

Our tutorials, hackathons, workshops and school are listed at the Events Web Page. Some upcoming events are sketched in the following.

- **Joint NOMAD - E-CAM Workshop: Modeling materials at realistic space and time scales via optimal exploitation of exascale computers and AI**
  DATE: shifted from November 2021 to spring 2022
  LOCATION: CECAM, HQ, Lausannne
  https://nomad-coe.eu/events/nomad-e-cam-workshop
- **High-throughput workflows for materials science with the Atomic Simulation Environment and Fireworks**
  DATE: **November 15-19, 2021**
  LOCATION: Technical University of Denmark, Lyngby, Denmark
- **NOMAD school: Towards exascale solutions in Green function methods and advanced DFT**
  DATE: **June 16-21, 2022**
  LOCATION: University of Latvia, Riga, Lativia

## SELECTED PRESENTATIONS OF NOMAD MEMBERS AND VIDEOS

- **Materials for Humanity**, International Conference, Singapore (virtual), July 6-9, 2021

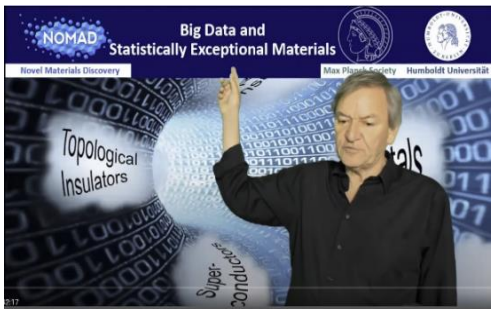


Fig. 6: Matthias Scheffler presented Big Data and Statistically Exceptional Materials at his plenary lecture in Singapore (virtual) 2021.

- **EUROMAT 2021,** Graz, Austria, (virtual), September 13-17, 2021


Fig. 7: Claudia Draxl presented a plenary talk about FAIR Data and Artificial Intelligence towards New Horizons in Materials.

- **FAIR-DI Workshop 2021 on FAIR data infrastructure for materials science**, Louvain-la-Neuve (virtual)**,** September 28 - October 01, 2021
  **This workshop was organized** by NOMAD PIs together with the Materials Genome Institute in Shanghai.

There were several other NOMAD PIs presented talks on one data sharing, analytics and exascale computing.

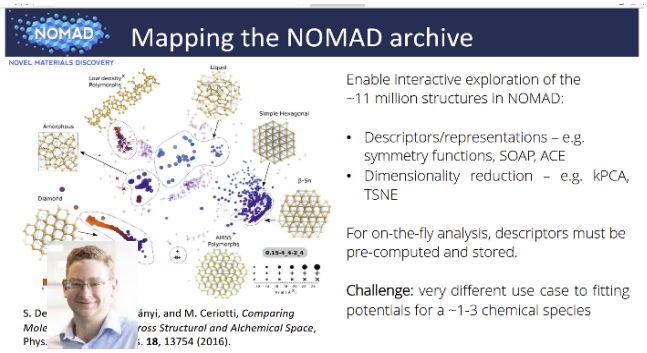


Fig. 8: James Kermode explaining "Upcoming Developments for Big Data Analytics in NOMAD".

## HONORS

NOMAD researcher, Professor Kristian Sommer Thygesen, has received a Villum Investigator Grant of 4 mio. Euro for research on data-driven discovery of functionalized 2D materials. Details can be found at the DTU web site.

The new Villum Investigator project is well aligned with the NOMAD CoE. In particular, the discovery of novel functionalized 2D materials will benefit from the exascale code libraries, workflow frameworks, and AI tools developed within the NOMAD CoE.

Congratulations Kristian!

Fig. 9: NOMAD PI Kristian Sommer Thygesen.

## GET IN TOUCH WITH US

- **General questions concerning the CoE**
  contact@nomad-coe.eu
- **NOMAD *ab initio* software**
  FHI-aims: sebastian.kokott@fhi.mpg.de
  exciting: claudia.draxl@physik.hu-berlin.de
  Abinit: xavier.gonze@uclouvain.be
  GPAW: thygesen@fysik.dtu.dk
- **Workflows**
  ASE/ASR: thygesen@fysik.dtu.dk,
  FireWorks: geoffroy.hautier@uclouvain.be
- **NOMAD data infrastructure**
  markus.scheidgen@physik.hu-berlin.de
- **Exascale HPC facilities**
  albert.farres@bsc.eu
- **Use-case demonstrators**
  christian.carbogno@fhi.mpg.de and thygesen@fysik.dtu.dk
- **Data analytics/ artificial intelligence:**
  luca.ghiringhelli@fhi.mpg.de

https://www.nomad-coe.eu/

@NoMaDCoE

The NOMAD Laboratory

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951786.