

**Director of the Theory Department
(in January 2020 renamed in: The NOMAD Laboratory)**

Prof. Dr. Matthias Scheffler

Director at the FHI and Scientific Member of the Max Planck Society (since 1988)
Honorary Professor at the Freie Universität Berlin, Humboldt-Universität zu Berlin,
and Technische Universität Berlin
Distinguished Visiting Professor at the University of California, Santa Barbara
Habilitation, Technische Universität Berlin (1984)
Dr. rer. nat., Technische Universität Berlin (1978)

Recent Developments in the Theory Department

On July 1, 2019, the FHI Theory Department turned 31 years old; its founding director has since become an Emeritus. He will head the department until the new director assumes office, which will probably be on January 1, 2020. Karsten Reuter from the Technical University of Munich has been identified for this position and was recently elected by the MPG Section for Chemistry, Physics, and Technology (CPTS). We are very happy with this development and hope that by the time of the meeting of the Fachbeirat, he may already have signed his contract with the Max Planck Society. Currently, it is planned that the upper floor of the T building will be cleared by January 1, 2020; the “Scheffler Group” will be located in the Richard Willstätter House (RWH) and, together with those parts of its projects oriented towards research-data infrastructure (see the NOMAD/FAIR-DI activities below), also at the Humboldt-Universität zu Berlin.

From the very beginning, research in the department has been concerned with fundamental aspects of the chemical and physical properties of surfaces, interfaces, clusters, and nanostructures. Obviously, this theme is sufficiently general and covers a wide range of topics, so that there has been little reason for change. Having said this, we note, however, that a group working in biophysics was added a few years ago. Moreover, since 2013, activities on FAIR¹ data infrastructure and on (big) data analytics based on artificial-intelligence concepts have been systematically added. These activities have received considerable attention worldwide and are now affecting many of our new developments, as will be sketched below.

As of September 1, 2019, the FHI Theory Department is structured in 9 research groups, although most of the work in the department is performed as cross-cutting activities within this group structure. The groups and their leaders are:

1. *Unifying Concepts in Catalysis*, headed by **Sergey V. Levchenko**
2. *Heat and Charge Transport*, headed by **Christian Carbogno**

¹ FAIR stands for Findable, Accessible, Interoperable, and Re-usable. The FAIR Data Principles: <https://www.force11.org/group/fairgroup/fairprinciples>. See also: C. Draxl and M. Scheffler, Big Data-Driven Materials Science and its FAIR Data Infrastructure. Plenary Chapter in Handbook of Materials Modeling (eds. S. Yip and W. Andreoni), Springer (2019).
<https://arxiv.org/ftp/arxiv/papers/1904/1904.05859.pdf>

3. *Ab Initio Biomolecular Simulations*, headed by **Carsten Baldauf**
4. *Simulations from Ab Initio Approaches: Structure and Dynamics from Quantum Mechanics*, **Otto Hahn Group of Mariana Rossi**
5. *Big-Data Analytics for Materials Science*, headed by **Luca M. Ghiringhelli**
6. *Crystal-Structure Prediction and Heterogeneous Catalysis*, headed by **Matthias Scheffler**
7. *Novel Materials Discovery and a FAIR Research Data Infrastructure*, **Max Planck Fellow Group of Claudia Draxl** of the Humboldt-Universität zu Berlin
8. *Advanced Electronic-Structure Methods*, **Max Planck Partner Group of Xinguo Ren** at USTC in Hefei, China (in collaboration with Matthias Scheffler)
9. *Atomistic Simulations of Electrochemistry*, **Max Planck Tandem Group of Luana Pedroza** at UFABC, Santo André, Brazil (in collaboration with Mariana Rossi).

Furthermore, the Theory Department hosts Emeritus Alexander M. Bradshaw who has been a valued, stimulating member of the department since January 2009. He works on photoionization in molecules and clusters as well as on questions of energy supply and resources in the context of the sustainability debate. He plans to retire finally within the next three years on completion of a project with the University of Augsburg.

Ongoing organizational developments are summarized in the following:

Part of group 1 is moving to group 6, but the main part is being transferred to the renowned Skolkovo Institute of Science and Technology (Skoltech). This is a private graduate research institute in Moscow, Russia (established in 2011 in collaboration with the MIT), and Sergey V. Levchenko took up an Assistant Professorship there in November 2018. Group 2 will continue at the FHI, largely within Scheffler's **ERC project TEC1p: *Big-Data Analytics for the Thermal and Electrical Conductivity of Materials from First Principles***. The TEC1p funding ends in September 2022. Group 3 is being transferred partly to group 4 and partly to the MP Department of the FHI. Mariana Rossi and her group 4 will move to the Max Planck Institute for the Structure and Dynamics of Matter in Hamburg, where she will set up her distinguished Lise Meitner Group. A small subgroup connected to the DFG-funded Collaborative Research Centre (CRC) 951 (HIOS)¹¹ will remain at the FHI, hosted by the "Scheffler Group" in the RWI. Groups 5, 6, 7 are key elements of the NOMAD (Novel Materials Discovery) Center of Excellence¹², the FAIR Data Infrastructure association (FAIR-DI e.V.)¹³, and the activities of the Max Planck Network on Big-Data-Driven Materials Science (BiGmax)¹⁴. Some aspects are part of the Max Planck Graduate Center for Quantum Materials¹⁵. The materials science data infrastructure activities in NOMAD/FAIR-DI are strongly related to the FAIRmat consortium¹⁶ which may

¹¹ <https://www.physik.hu-berlin.de/de/sfb951>

¹² <https://nomad-coe.eu>

¹³ <https://fairdi.eu>

¹⁴ <https://bigmax.iwww.mpg.de/scope>

¹⁵ <https://www.quantummaterials.mpg.de>

¹⁶ <https://fairdi.eu/fairmat>

possibly become part of the NFDI (National Research Data Infrastructure)¹⁷; more information is given below. These groups (5-7) will remain, and their funding is planned at present for the period until fall 2025. Group 8 was favorably evaluated in December of last year, and its funding will continue until November 2020. Moreover, the funding of group 9 in Brazil will continue until July 2023. Concluding, by January 1, 2020, four groups will remain in Berlin: *NOMAD/FAIR-DI* (Claudia Draxl), *DFT and Beyond* (Matthias Scheffler), *Big-Data Analytics for Materials Science* (Luca M. Ghiringhelli), and *Heat and Charge Transport* (Christian Carbogno). They will occupy offices in the RWH and at the HU Berlin.

Together with Martin Vingron (director at the MPI for Molecular Genetics), Matthias Scheffler is responsible for the Joint Network Center (Gemeinsames Netzwerkzentrum, GNZ) for the Max Planck activities in Berlin and Brandenburg. This is a regional IT competence center, serving 8 MPIs and 7 additional institutions. It focuses on networking, backup, virtualization, and security services. The GNZ is headed by Gerd Schnapka, and more information is given in the report of the GNZ in this booklet.

Groups 1, 2, 4, 5, 7, 8 are described by their respective group leaders in special chapters, following this general overview. At this point it is perhaps appropriate to mention a few highlights, in particular in those areas that are likely to comprise the future activities of the “Scheffler Group”.

NOMAD and FAIR-DI

These activities started with the NOMAD (Novel Materials Discovery) Repository in 2013 (online since 2014), together with Claudia Draxl and the MPCDF (Max Planck Computing and Data Facility). They have developed into the highly successful NOMAD Center of Excellence (established in October 2015). In brief, NOMAD has five key components (see Fig. 1).

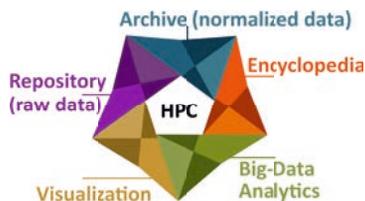


Figure 1: Structure of the NOMAD (Novel Materials Discovery) Center of Excellence which now is the biggest data base in computational materials science. For details see the report by Claudia Draxl and the respective poster.

With its combination of the Repository and Archive, it represents a FAIR¹ data infrastructure. FAIRness also implies the underlying metadata and ontology infrastructure, the NOMAD Metainfo. Many people have contributed here, but four of them should be mentioned explicitly: *Fawzi Mohamed*, *Luca M. Ghiringhelli*, *Claudia Draxl*, and *Markus Scheidgen*. The metadata and ontology

¹⁷ https://www.dfg.de/en/research_funding/programmes/nfdi/index.html

development will be advanced in the coming years, supported by the BMBF project STREAM (Semantische Repräsentation, Vernetzung und Kuratierung von qualitätsgesicherten Materialdaten).

Figure 2 shows a transparency that summarizes the worldwide leading data infrastructures in materials science as shown by James Warren, the director of the Materials Genome Program of NIST (USA) in his welcome speech at a workshop in Washington D.C. in August 2019. It not only includes NOMAD but also mentions the OPTiMaDe (Open Databases Integration for Materials Design)¹⁸ project. OPTiMaDe aims to make various materials databases interoperational by developing a common REST API. This effort is led by Gian-Marco Rignanesi from Louvain La Neuve, and FHI / NOMAD scientists form a significant part of the consortium. Further aspects of NOMAD are described in the report by Claudia Draxl and in the next paragraph on artificial intelligence for materials science.



Figure 2: Slide listing the main materials science data infrastructures. This was shown in the welcome talk of the director of the Materials Genome Program of NIST in August 2019.

NOMAD also plays a significant role for the Max Planck Network on Big-Data-Driven Materials Science (BiGmax¹⁴; Fig. 3 left) in which several institutes of the Max Planck Society combine their know-how to further data-driven materials science. The aim is to develop concepts and methods for handling and analyzing large amounts of materials data. In its present funding period, BiGmax will run until February 2022. Together with Peter Benner (MPI for Dynamics of Complex Technical Systems, Magdeburg), Matthias Scheffler is coordinating the BiGmax network and the Theory Department is part of several cooperation projects.

In September 2018, institutions from the Netherlands and Germany founded the not-for-profit association FAIR-DI e.V.¹³ (Fig. 3 right; "gemeinnütziger, eingetragener Verein" according to German law). It is headed by Matthias Scheffler (supported by two executive secretaries Carsten Baldauf and Alexander Brink (Radboud University)). Its mission statement reads: "Scientific

¹⁸ <http://www.optimade.org>

data are a significant raw material of the 21st century. To exploit its value, a proper infrastructure that makes it Findable, Accessible, Interoperable, and Re-purposable – FAIR – is a must. For the fields of computational and experimental materials science, chemistry, and astronomy, FAIR-DI e.V. sets out to make this happen. This enabling of extensive data sharing and collaborations in data-driven sciences (including artificial-intelligence tools) will advance basic science and engineering, reaching out to industry and society.”



Figure 3: Two important facilities co-lead by the FHI.

The FAIR-DI concept is largely identical with the goals of the planned large-scale German Data Infrastructure NFDI (the corresponding DFG call for proposals was published in June this year and successful submissions may expect funding from fall 2020 onwards)¹⁷. FAIR-DI's materials-science component will submit a proposal called FAIRmat¹⁶ in October 2019. FAIRmat, coordinated by Claudia Draxl, is supported by the CPTS of the Max Planck Society, the Condensed Matter Section of the German Physical Society, and numerous research groups in Germany.

Artificial Intelligence for Materials Science

We realized early on that the direct impact of high-throughput screening of materials may remain limited as the amount of possible materials is practically infinite. Thus, two goals were defined: First, to construct “property maps of materials” analogous to the periodic table of the elements (see Fig. 4). Second, to note that the number of high-performance materials that may be suitable for a certain purpose, may be very small, e.g. just 10 or 100 out of the practically infinite number of all possible materials. Thus, the goal is not a statistical-learning process (or a “fit”) of the big data but the identification of “statistically exceptional regions” in the data space (“needles in a haystack”). Two important methods have been developed for these purposes: a) SISSO (Sure Independence Screening and Sparsifying Operator) and b) Subgroup Discovery. In simple words, SISSO follows the spirit of LASSO (Least Absolute Shrinkage and Selection Operator), a popular method in compressed sensing (signal processing). However, there are several significant advancements. In LASSO, the selection is approximated by replacing the ℓ_0 by the ℓ_1 norm, since there are problems with ℓ_0 in actual calculations because of its NP hardness. Furthermore, LASSO becomes unstable when the number of features offered are correlated and/or exceed about 5,000, with about 100 data points, which is a typical number of data points in materials discovery. All these issues were resolved with SISSO: SISSO considers the ℓ_0 norm and the number of features can be huge, e.g. billions or trillions. For the one-dimensional feature selection, the result is even exact. And for higher dimensions we implemented an iterative approach. The key people behind this work are *Runhai Ouyang* and *Luca M.*

Ghiringhelli. More details are described in the report by Luca M. Ghiringhelli and on the posters.

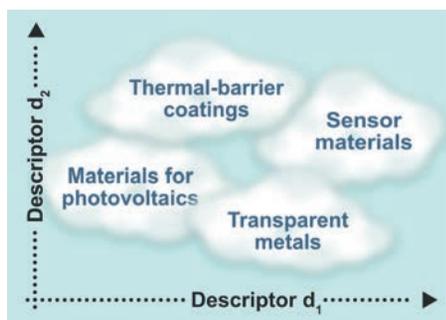


Figure 4: Big data of materials contain correlations and structure that are invisible in small data sets. Finding descriptors that determine a specific property or function of a material is a crucial challenge. Once this is in place, it will be possible to machine learn the data and eventually draw maps of materials.

The second highly important development is “subgroup discovery”. As mentioned above, fitting, or machine learning, all available data with a single, global model means fitting the hay, where one may accidentally average away the special features of the interesting minority, namely, the needles. Using Boolean conjunctions, the method of subgroup discovery identifies statistically-exceptional subgroups in a large amount of data. The key developers of this approach were *Mario Boley*, *Jilles Vreeken (Saarland University)*, *Luca M. Ghiringhelli*, and *Bryan Goldsmith*. A recent application of the approach was the identification of materials that may be good for catalytic CO₂ activation (turning a greenhouse gas into fuels and useful chemicals). This work was done with *Sergey V. Levchenko* and *Aliaksei Mazheika* (see also the report of *Sergey V. Levchenko*).

Thermal and Electrical Conductivity from First Principles

Thermal conductivity (TC) is a key characteristic of many materials, e.g. thermoelectrics, thermal-barrier coatings, and catalysts. However, TC is largely unknown – of the 225,000 identified inorganic semiconductor and insulator crystals, only 100 have any experimental TC data available. We started a high-throughput study of TC by first developing a rapid-decision algorithm on what materials actually exhibit strong anharmonic effects and thus low TCs. This revealed that for many of the technologically interesting materials anharmonic effects are large, so that the standard description in terms of perturbation theory and Boltzmann equation does not provide a reliable description. To investigate accurately these strongly anharmonic materials, we have developed and coded a full *ab initio* molecular dynamics description with a numerically efficient algorithm that enables convergence of the statistical mechanics in system size and time. Together with the decision algorithm mentioned above, this has allowed us to screen thousands of materials and to study the TC of several complex material classes, including half-Heuslers, inorganic halide perovskites, and ultra-insulating layered chalcogenides like SnSe.

Analogously, we have developed an *ab initio* theory of electrical transport for systems with strong electron-phonon coupling and/or strong anharmonic effects. At this point, our approach is the only one that does not involve perturbation theory – neither for the electron-phonon coupling nor for the nuclear dynamics. Rather, all anharmonic effects in the nuclear motion are accounted for via *ab initio* molecular dynamics, while electron-phonon coupling is treated to all orders by analyzing the evolution of the self-consistent electronic-structure along the anharmonic trajectory. The approach is being implemented and tested right now. The key developer behind these TEC1p methods is *Christian Carbogno* (see corresponding report).

Related to these studies we are also working on advanced exchange-correlation treatments, e.g. a self-interaction corrected SCAN (Strongly Constrained and Appropriately Normed) functional, RPA (Random-Phase Approximation) and coupled-cluster theory in order to test the accuracy of the computed properties of strongly anharmonic systems. The key developers are *Igor Y. Zhang* and *Xinguo Ren* (see corresponding report).

Let us finish with a note on the recent scientific impact of the Theory Department, specifically on our NOMAD/FAIR-DI/AI activities. This may be measured by the number of invited, keynote, and plenary talks held by members of the Theory Department on related topics. For the last 2-year reporting period (from September 1, 2017 to August 31, 2019) the number was 76. And there were numerous contributed talks by our students and postdocs.

Members of the Theory Department take part in various national, European, and international research programs and initiatives. The following list of these activities is ordered alphabetically according to the funding agency:

- BMBF – Federal Ministry of Education and Research (Germany), Berlin Big Data Center (BBDC) of Competence, coordinator: V. Markl; project leader: M. Scheffler, October 2014 - September 2018.
- BMBF – Federal Ministry of Education and Research (Germany), Eigenwert-Löser für PetaFlop-Anwendungen: Algorithmische Erweiterungen und Optimierungen (ELPA - AEO), coordinator: H. Lederer; project leaders: M. Scheffler, C. Carbogno, February 2016 - Januar 2019.
- BMBF – Federal Ministry of Education and Research (Germany), Semantische Repräsentation, Vernetzung und Kuratierung von qualitätsgesicherten Materialdaten (STREAM), coordination: Fraunhofer-Institut für Werkstoffmechanik IWM; project leaders: M. Scheffler, C. Baldauf, August 2019 - July 2022.
- BMWi – Federal Ministry for Economic Affairs and Energy (Germany), Cooperation Project within “Zentrales Innovationsprogramm Mittelstand”: Entwicklung einer neuartigen Softwarekombination; FHI project leaders: M. Scheffler, S. Kokott, May 2019 - April 2021.
- DFG – German Research Foundation, CRC 951: Hybrid Inorganic/Organic Systems for Opto-Electronics (HIOS), spokesperson: N. Koch; project leaders: P. Rinke, M. Scheffler (B4) and A. Tkatchenko, M. Scheffler (A10), second funding period: July 2015 - June 2019. A third funding period has been approved: July 2019 - June 2023, Current project: Temperature effects

on the atomic and electronic structure of organic/TMDC interfaces (A13), project leader: M. Rossi.

- EPSRC - Engineering and Physical Sciences Research Council (UK): International Centre-to-Centre Collaboration: New trimetallic nanoparticles as catalysts for the conversion of carbon dioxide to renewable fuels, coordinator: G. Hutchings; project leader: M. Scheffler, starting October 2019.
- ERC – European Research Council, Advanced Grant: Big-Data Analytics for the Thermal and Electrical Conductivity of Materials from First Principles (TEC1p), M. Scheffler, October 2017 - September 2022.
- EU – European Commission, E-Infrastructures: NOMAD (Novel Materials Discovery) Center of Excellence, coordinator: M. Scheffler; work package leaders: L. M. Ghiringhelli, C. Draxl, November 2015 - October 2018.
- FAPESP – São Paulo Research Foundation, Max Planck Tandem Group for Atomistic Simulations of Electrochemistry lead by L. Pedroza at UFABC in Santo André, Brazil, August 2018 - July 2023.
- Leibniz-Gemeinschaft – Leibniz-WissenschaftsCampi (LWC): Growth and fundamentals of oxides for electronic applications (GraFOx), spokesperson: H. Riechert; principal investigators: M. Scheffler, C. Draxl, since July 2016.
- Max Planck Society and École Polytechnique Fédérale de Lausanne – MP-EPFL Center for Molecular Nanoscience and Technology, directors: K. Kern, T. Rizzo; since 2013. Current project: Machine learning of experimental observables in surface and interface science, project leaders: M. Ceriotti (EPFL), M. Rossi.
- Max Planck Society - Central Funds – MP Fellow Group of C. Draxl, since July 2014. In July 2019, C. Draxl was appointed as MP Fellow of the Max Planck Graduate Center for Quantum Materials for 5 years, with her MP Fellow Group based at the FHI.
- Max Planck Society - Central Funds – MP Partner Group for Advanced Electronic-Structure Methods lead by X. Ren at the University of Science and Technology of China (USTC) in Hefei, December 2015 - November 2020.
- Max Planck Society - Central Funds – BiGmax, Max Planck Network on Big-Data-Driven Materials Science, coordinators: P. Benner, M. Scheffler; project leaders: C. Draxl, L. M. Ghiringhelli, M. Rossi, M. Scheidgen, M. Scheffler, March 2017 - February 2022.
- Max Planck Society - Central Funds - Max Planck Graduate Center for Quantum Materials, speaker: B. Keimer; FHI coordinators: M. Rossi, M. Scheffler, since January 2019.
- Max-Planck-Cardiff Centre on the Fundamentals of Heterogeneous Catalysis 'FUNCAT', coordinators: G. Hutchings, F. Schüth, project leader: M. Scheffler, starting 2020.
- UCSB – University of California, Santa Barbara, College of Engineering and College of Mathematical, Life & Physical Sciences; M. Scheffler (distinguished visiting professor for materials science and engineering), since 2005.

Additionally, Matthias Scheffler is Visiting Professor at the Institute for Catalysis, Hokkaido University, Japan, since 2016.

We continue our strong involvement and support of CECAM and Psi-k, e.g. by organizing workshops, tutorials, and schools and working in the CECAM Council and the Psi-k Board of Trustees (the governance institutions of these two organizations).

CECAM is a high-level European organization devoted to the promotion of fundamental research on advanced computational methods and their application to important problems in frontier areas of science and engineering. It is supported by various European research organizations including MPS and DFG. The CECAM headquarters is at the EPFL Lausanne, and CECAM nodes have been established in several member states. We are part of the joint node, cecam-MM1P.de which focusses on multi-scale modeling from first principles (MM1P) with emphasis on methods and applications to materials and biophysics. The consortium of the MM1P.de node consists of the University of Bremen (represented by the Node Director T. Frauenheim and T. Wehling), the MPI Hamburg (represented by A. Rubio), Goethe University Frankfurt (represented by B. Winkler and R. Valenti), and the FHI (represented by M. Scheffler). The node will continue to exist at least until 2022.

Psi-k is a Europe-based worldwide network of researchers working on the advancement of first-principles computational materials science. Its mission is to develop fundamental theory, algorithms, and codes in order to understand, predict, and design materials properties and functions. Theoretical condensed matter physics, quantum chemistry, thermodynamics, and statistical mechanics form the scientific core. Applications encompass inorganic, organic, and bio-materials and cover a whole range of diverse scientific, engineering, and industrial endeavors. Key activities of Psi-k are the organization of conferences, workshops, tutorials, and training schools and the dissemination of scientific thinking in society.

In the context of NOMAD and the FAIR-DI activities, the FHI has also established a collaboration with Shanghai University, specifically with its Materials Genome Institute on "Data Infrastructure and Materials Informatics".

Personnel and Related Matters

Since the last visit of the Fachbeirat in November 2017 the following noteworthy developments have taken place:

- **Mariana Rossi** was selected as leader for a highly prestigious Lise Meitner Group which she will establish in January 2020 at the Max Planck Institute for the Structure and Dynamics of Matter in Hamburg. Additionally, she was accepted into the Elisabeth Schiemann Kolleg in June 2018. Within the Elisabeth Schiemann Kolleg scientific members of the Max Planck Society foster the careers of excellent female scientists after their postdoc phase, helping them to succeed on their way to an appointment as a tenured professor or as a director of a research institution.

- Since September 2018, **Carsten Baldauf** has been Visiting Professor (Professurvertretung Theoretische Chemie) at the Wilhelm-Ostwald-Institut, Fakultät für Chemie und Mineralogie, Universität Leipzig. Starting October 1, 2019, he assumes a permanent position in the administration of the FHI as representative of the board of directors.
- **Igor Ying Zhang** became Professor at Fudan University in March 2018.
- **Yang-Gang Wang** became Associate Professor at the Southern University of Science and Technology of China in September 2018.
- **Sergey V. Levchenko** assumed an Assistant Professorship at the Skolkovo Institute of Science and Technology in November 2018.
- **Honghui Shang** became Associate Professor at the Institute of Computing Technology of the Chinese Academy of Sciences in November 2018.
- **Runhai Ouyang** assumed a position as Associate Professor of Materials at the Genome Institute at Shanghai University in February 2019.
- **Jungho Shin** assumed a Senior Researcher Position at the Korea Research Institute of Chemical Technology in March 2019.
- In February 2019, **Matthias Rupp** became an Applied Scientist at Citrine Informatics, a leading company for data-driven developments of chemicals and materials.
- **Angelo Ziletti** accepted a position as Senior Data Scientist in the Artificial-Intelligence Lab of Bayer in April 2019.
- **Stefano Curtarolo** received the Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation in 2016 and spent a sabbatical year in Berlin from July 2017 - July 2018.

Scientists of the Theory Department were involved in various services and activities supporting the surface-science and electronic-structure communities worldwide. For example, they lecture at the Free University Berlin, the Technical University Berlin, and the Humboldt-Universität zu Berlin as well as in the International Max Planck Research School. They organized or co-organized conferences, workshops, tutorials, and summer schools on topics in electronic-structure theory, multi-scale modeling, surface science, biophysics, and big-data science which have impacted upon the careers of very many students of theoretical materials science. Let us emphasize here in particular the 10-day hands-on schools which are organized by the Theory Department since 1994. They cover the whole field of DFT and *GW* and all important methodologies (e.g. pseudo-potential plane wave, LAPW, numeric atomic orbital, Quantum Monte Carlo methods, etc.), and we invite top experts in these fields. The practical sessions take place in the afternoons and evenings, using the FHI-aims code. Originally these schools were run every second year. However, the demand is significant and in the last few years we have organized these events yearly, in Berlin and different international locations. This year we had more than 100 applicants but were only able to accept 74 participants (from 16 countries). The workshop in Beijing in 2018 was generously funded by the Sino-German Center for Research Promotion and the Max Planck Society. This workshop as well as the one in Barcelona in 2019 (funded by CECAM and Psi-k) were great successes, and Carsten Baldauf, the main organizer, did a very impressive job. As always, the whole Theory Department helped greatly.

Further activities in the last two years included:

- CECAM workshop Quantum-Chemistry Methods for Materials Science, Lausanne, Switzerland, November 8–10, 2017, organizers: I. Y. Zhang, M. Scheffler.
- NOMAD: Third Industrial Meeting, London, United Kingdom, February 5–6, 2018, organizers: A. De Vita (King's College London), A. Rubio (MPI for the Structure and Dynamics of Matter, Hamburg), M. Scheffler.
- Symposium on Managing and Exploiting the Raw Material of the 21st Century at DPG Spring Meeting in Berlin, Germany, March 11–16, 2018, organizers: C. Draxl, P. Fratzl (MPI of Colloids and Interfaces, Potsdam).
- Symposium on Frontiers of Electronic Structure Theory: Correlated Electron Materials at DPG Spring Meeting in Berlin, Germany, March 11–16, 2018, organizers: S. Biermann (École Polytechnique, Palaiseau), P. Kent (Oak Ridge National Laboratory), M. Scheffler.
- BiGmax Workshop 2018 on Big-Data-Driven Materials Science, Kloster Irsee, Germany, April 10–13, 2018, organizers: J. Neugebauer (MPI für Eisenforschung GmbH, Düsseldorf), P. Benner (MPI for Dynamics of Complex Technical Systems, Magdeburg).
- CECAM workshop Path Integral Quantum Mechanics: From the Basics to the Latest Developments, Lausanne, Switzerland, June 25–29, 2018, organizers: M. Rossi, J. Richardson (ETH Zürich), M. Ceriotti (EPF Lausanne), T. Markland (Stanford).
- Electronic Structure Theory with Numeric Atom-Centered Basis Functions - FHI-aims Developers' and Users' Meeting, Munich, Germany, July 9–11, 2018, organizers: V. Blum (Duke University, USA), W. Huhn (Duke University), H. Oberhofer (TU Munich), X. Ren (University of Science and Technology of China), M. Rossi.
- Hands-On DFT and Beyond: Frontiers of Advanced Electronic Structure and Molecular Dynamics Methods, Peking University, Beijing, China, July 30–August 10, 2018, organizers: X. Li (Peking University), C. Baldauf, X. Ren (University of Science and Technology of China), H. Jiang (Peking University), V. Blum (Duke University), M. Scheffler.
- HoW exciting! 2018, Humboldt-Universität zu Berlin, Berlin, Germany, July 31–August 9, 2018, organizers: C. Cocchi (HU Berlin), A. Gulans (HU Berlin), P. Pavone (HU Berlin), C. Draxl.
- Theory Department Workshop 2018, Ringberg, Germany, September 12–14, 2018, organizers: C. Carbogno, M. Rossi, M. Scheffler.
- Long Program on Science at Extreme Scales: Where Big Data Meets Large-Scale Computing at the Institute for Pure & Applied Mathematics, Los Angeles, USA, September 12–December 14, 2018, organizers: J. Buhmann (ETH Zürich), H.-J. Bungartz (TU Munich), E. Candes (Stanford University), C. Draxl, J. Hittinger (Lawrence Livermore National Laboratory), F. Jenko (MPI for Plasma Physics and UCLA), D. Keyes (KAUST), A. Lee (AMD), T. Warnow (University of Illinois).
- NOMAD Summer: A Hands-On Course on Tools for Novel-Materials Discovery, Lausanne, Switzerland, September 24–27, 2018, organizers: L. M. Ghiringhelli, M. Rampp (MPCDF, Garching), A. Ziletti, M. Scheffler.

- Exciting Shanghai School, Shanghai, China, November 17–22, 2018, organizers: W. Ren (Shanghai University), A. Stroppa (CNR-SPIN), C. Draxl, P. Pavone (HU Berlin).
- FAIRmat Meeting, FHI, Berlin, Germany, March 22, 2019, organizer: C. Baldauf, C. Draxl, M. Scheffler.
- Symposium on Frontiers of Electronic-Structure Theory: Focus on The Interface Challenge at DPG Spring Meeting in Regensburg, Germany, March 31–April 5, 2019, organizers: J. Nørskov (Technical University of Denmark), K. Reuter (TU Munich), M. Scheffler.
- NOMAD-FAIR-DI workshop: Shared metadata and data formats for Big-Data Driven Materials Science, Berlin-Adlershof, Germany, July 8–12, 2019, organizers: C. Draxl, L. M. Ghiringhelli, G. Hutchings (Cardiff University), J. Kermode (University of Warwick), C. Koch (HU Berlin), C. Liebscher (MPI für Eisenforschung GmbH, Düsseldorf), M. Tourdot de Oliveira (MPI for the Structure and Dynamics of Matter, Hamburg), G.-M. Rignanese (UCLouvain), M. Scheffler, C. Wöll (KIT).
- FAIRmat Meeting, Berlin-Adlershof, July 15–16, 2019, organizers: C. Baldauf, C. Draxl, M. Scheffler.
- Hands-On DFT and Beyond: High-throughput screening and big-data analytics, towards exascale computational materials science, Barcelona, Spain, August 26–September 6, 2019, organizers: F. Illas Riera (University de Barcelona), E. Ruiz Sabin (University de Barcelona), C. Baldauf, S. Kokott, V. Blum (Duke University), M. Scheffler.
- BiGmax summer school Big Data Summer, Platja d'Aro, Spain, September 9–13, 2019, organizers: G. Dehm (MPI für Eisenforschung GmbH, Düsseldorf), C. Draxl, M. Scheffler, J. Vreeken (CISPA – Helmholtz Center for Information Security).



Prof. Dr. Sergey V. Levchenko

Assistant Professor at the Skolkovo Institute of Science and Technology, Moscow, Russia (since 2018)
Group Leader at the FHI (2008-2019)
Postdoc at the University of Pennsylvania, PA, USA (2005-2008)
Ph.D. (Chemistry), University of Southern California, CA, USA (2005)

Activities of the “Unifying Concepts in Catalysis” Group

Our group studies surfaces and interfaces under realistic conditions of temperature (T), pressure (p), and doping using first-principles modelling combined with statistical mechanics. In particular, we investigate the surface stoichiometry and atomic structure of heterogeneous catalysts. Below, we describe the projects pursued in the last two years.

Accurate Single-Particle Energies and Band Gaps of Solids from Second-Order Møller-Plesset Perturbation Theory and Beyond

With: Maria Dragoumi, Igor Y. Zhang (Fudan University), and Matthias Scheffler

Previously, we have unveiled the important role that charge-carrier doping plays in defect formation at surfaces under realistic (T , p) conditions [1]. This has prompted us to develop a methodology for describing polarons in materials using *ab initio* methods [2]. As a continuation of this work, we develop a many-body perturbation theory method for accurate evaluation of polaron properties in solids. While local density/generalized gradient approximations to density-functional-theory (DFT) exchange-correlation (XC) functional suffer from the self-interaction error that influences charge-carrier localization energetics (an important property of a polaron), hybrid functionals contain adjustable parameters the values of which can be set by comparing with more accurate parameter-free methods. Therefore, we start from the Møller-Plesset second-order perturbation theory for quasiparticle excitations [3], but go beyond by solving the Dyson equation. This way, an infinite summation of diagrams is considered, including second-order exchange that cancels the one-electron self-interaction error. Calculations of the band gap for semiconductors and insulators with the new approach show that the Dyson equation is crucial for a proper treatment of single-particle energies. The new approach shows a competitive or even superior performance in comparison to hybrid functionals and the *GW* approximation, where the second-order exchange diagram is missing.

Oxidation of Elemental and Alloyed Transition-Metal Surfaces at Realistic Temperature and Pressure

With: Zhong-Kang Han, Santiago Rigamonti (HU Berlin), Maria Troppenz (HU Berlin), Claudia Draxl, and Matthias Scheffler

In this project, we study thermodynamics of elemental and alloyed metal surfaces under an oxygen atmosphere. The thermodynamics of oxidation and its dependence on the metal type and surface orientation are still not fully understood. We have performed a DFT study (comparing various XC approximations) for the adsorption of atomic and molecular oxygen on the (111) and (110) surfaces of Ag, Au, Pt, Rh, Pd, and Cu elemental metals, and on

the Pd- or Pt-alloyed Cu(111) and Cu(110) surfaces. The effects of configurational entropy were evaluated by combining a cluster expansion (CE) model with Monte Carlo sampling [4]. To the best of our knowledge, this is the first study that takes into account configurational entropy for both alloy components and adsorbates simultaneously. It leads to a very large number of possible compact clusters in CE, so that machine-assisted selection of the clusters is required to find a predictive model. We showed that it is essential to take into account the Coulomb repulsion between oxygen anions to explain the stability of ordered structures of adsorbed O atoms on the (110) and (111) surfaces of elemental metals. The structures of adsorbed O at the Pd or Pt-alloyed Cu surfaces exhibit a larger variety of patterns depending on coverage; Pd/Pt segregate to the surface at low oxygen coverages, while at high oxygen coverages Cu segregates.

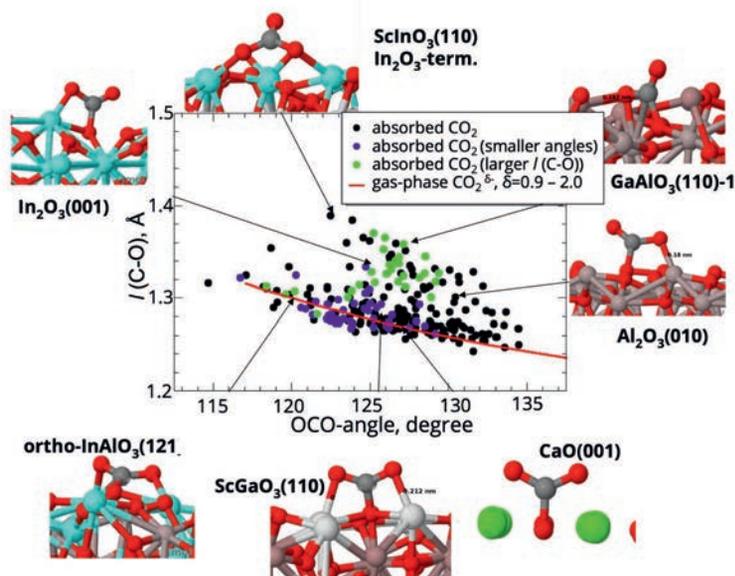


Figure 1: The correlation between the larger of the two C-O bond lengths (in case the two bond lengths are different) and the OCO-angle in charged gas-phase (red line) and adsorbed CO_2 (dots, blue – subgroup with small OCO-angles, green – subgroup with large $I(\text{C-O})$, black – the remaining samples).

A Route Towards Discovering Better Catalysts Using Artificial-Intelligence Analysis of *Ab Initio* Data Together with Experimental Data

With: Aliaksei Mazheika, Yang-Gang Wang, Rosendo Valero (University of Barcelona), Luca M. Ghiringhelli, Francesc Illas Riera (University of Barcelona), and Matthias Scheffler

In this project we have applied data-analytics techniques to systems relevant for heterogeneous catalysis [5,6]. In particular, we have studied the activation of carbon dioxide at oxide surfaces [6]. Uncontrolled emission of CO_2 into the atmosphere is one of today's major environmental threats. The most promising way to recycle CO_2 is to convert it to fuels and other useful chemicals using

heterogeneous catalysis. For a rational design of catalytic materials for CO₂ activation, a reliable prediction of catalytic performance (e.g., reaction turn-over frequency, TOF) requires understanding and modelling the full catalytic cycle, which is currently not feasible. A way out is to find *indicators* for activation, i.e., common features of good CO₂-conversion catalysts. In order to be useful, these features should be much easier to evaluate than the TOF. We consider several physically motivated indicators for CO₂ activation, including the OCO bending angle and the C-O bond length $l(\text{C-O})$ in the adsorbed CO₂ molecule. Overall, 71 different binary and ternary oxide materials, 141 surfaces, and 270 unique adsorption sites were considered. We then used the subgroup-discovery (SGD) [7] approach to identify subgroups of adsorption sites that provide a small OCO angle or a large C-O bond length. Interestingly, for the $l(\text{C-O})$ subgroup, *all* experimentally tested materials are good CO₂-conversion catalysts. On the other hand, several materials that have been identified as unsuitable for CO₂ activation belonged to the “small OCO angle” subgroup. This is explained by the fact that the OCO angle is reduced on surfaces that also bind CO₂ too strongly. The longer $l(\text{C-O})$ is achieved due to the binding of an O atom in the adsorbed molecule to a nearby surface cation, which provides an additional mechanism for weakening C-O bonds (Fig. 1). This suggests that C-O bond elongation might be used as an indicator for CO₂ activation. Based on these findings, we propose a set of promising new catalysts for CO₂ conversion, and a recipe to find more. In future, we plan to extend this work to other materials classes, including metal carbides.

Atomic and Electronic Structure of β -Ga₂O₃ Surfaces at Realistic Conditions

With: Konstantin Lion (HU Berlin), Matthias Scheffler, and Claudia Draxl

The transparent conducting oxide Ga₂O₃, which exhibits a band gap of about 4.9 eV, is a promising candidate for a number of applications, such as semiconducting lasers and transparent electrodes for UV optoelectronic devices and solar cells. As such, the bulk properties of its thermodynamically stable β phase have been extensively studied in the last two decades. The surface properties that play a vital role in epitaxial growth, electrical contacts, and gas sensors are, however, still not well understood. We performed DFT calculations for several β -Ga₂O₃ surfaces with the hybrid exchange-correlation functional HSE06, which mixes a portion of exact exchange from Hartree-Fock theory with exchange energy from a generalized gradient calculation. The mixing parameter was chosen as 26% exact exchange, which reproduces the experimental band gap. Surface structures were modelled as slabs in the supercell approach, including up to 140 atoms per supercell. Surface thermodynamic stability over a wide range of temperature and pressure conditions was characterized by the surface free energy. While the total energy is the dominating contribution to the free energies, vibrational contributions can potentially affect the relative stability of different surface orientations. Therefore, we also included vibrational contributions for the two most stable surfaces.

We found that the (100) surface is the most stable one, in agreement with previous reports, but, quite surprisingly, it is followed closely by the (201) surface. Here, the topmost surface layers are significantly flattened upon

relaxation, reducing the surface energy to 40% of the value of the unrelaxed surface. Vibrational contributions further slightly stabilize both surfaces but do not change their relative stability. This result explains recent experimental findings where $(\bar{2} 01)$ facets were found to form during homoepitaxial growth on off-oriented $\beta\text{-Ga}_2\text{O}_3$ (100) substrates [8].

References

1. Richter, N.A., S. Sicolo, S.V. Levchenko, J. Sauer, and M. Scheffler: Concentration of vacancies at metal-oxide surfaces: Case study of MgO (100). *Physical Review Letters* **111**(4), 045502 (2013)
2. Kokott, S., S.V. Levchenko, P. Rinke, and M. Scheffler: First-principles supercell calculations of small polarons with proper account for long-range polarization effects. *New Journal of Physics*, **20**(3), 033023 (2018)
3. Sun, J.Q., and R.J. Bartlett: Second-order many-body perturbation-theory calculations in extended systems. *The Journal of Chemical Physics*, **104**(21), 8553-8565 (1996)
4. Stampfl, C., H.J. Kreuzer, S.H. Payne, H. Pfnür, and M. Scheffler: First-principles theory of surface thermodynamics and kinetics. *Physical Review Letters*, **83**(15), 2993 (1999)
5. Andersen, M., S.V. Levchenko, M. Scheffler, and K. Reuter: Beyond scaling relations for the description of catalytic materials. *ACS Catalysis*, **9**(4), 2752-2759 (2019)
6. Mazheika, A., S.V. Levchenko, L.M. Ghiringhelli, and M. Scheffler: *Ab initio* data-analytics study of carbon-dioxide activation on semiconductor oxide surfaces, in preparation
7. Goldsmith, B.R., M. Boley, J. Vreeken, M. Scheffler, and L.M. Ghiringhelli: Uncovering structure-property relationships of materials by subgroup discovery. *New Journal of Physics*, **19**, 013031 (2017)
8. Schewski, R., K. Lion, A. Fiedler, C. Wouters, A. Popp, S.V. Levchenko, T. Schulz, M. Schmidbauer, S. Bin Anooz, R. Grüneberg, Z. Galazka, G. Wagner, K. Irmscher, M. Scheffler, C. Draxl, and M. Albrecht: Step-flow growth in homoepitaxy of $\beta\text{-Ga}_2\text{O}_3$ (100)—The influence of the miscut direction and faceting. *APL Materials* **7**, 022515 (2019)



Dr. Christian Carbogno

Group Leader at the FHI (since 2015)

Deputy Group Leader at the FHI (2013-2015)

Post-doctoral Researcher at the University of California Santa Barbara, USA (2010–2012)

Dr. rer. nat., Universität Ulm (2009)

Activities of the “Heat and Charge Transport” Group

Heat and charge transport processes play a key role in materials science, e.g., in the discovery of novel materials for thermoelectric devices. Nonetheless, little is known about the actual mechanisms that drive or hinder heat and charge transport in complex materials under actual operational temperatures $T > 300$ K. To date, perturbative formalisms based on an idealized, semi-harmonic description of the nuclear motion and its coupling to the electrons are still the most popular tool in materials science [1] – despite the fact that these assumptions are prone to failure in real materials and/or at elevated temperatures. The main focus of the *Heat and Charge Transport* group is to overcome these approximations by developing novel first-principles methodologies based on *ab initio* molecular dynamics (aiMD), so as to accurately account for **all anharmonic** effects in the nuclear motion and in its coupling to the electrons. These developments massively increase the range of materials, temperatures, and pressures, for which highly-accurate first-principles predictions are possible. We demonstrate this by applying the developed techniques – mostly in a *high-throughput* fashion – to fundamental, yet still unsolved, materials-science problems such as the design of improved thermoelectric materials. Accordingly, this research covers the fundamental challenges within the ERC-funded TEC1p project of Matthias Scheffler. To facilitate this, the group also addresses practical aspects of density-functional theory (DFT), e.g., decreasing its computational cost and estimating its numerical errors, in collaborative efforts.

Vibrational Heat Transport in Strongly Anharmonic Solids

With: Florian Knoop, Thomas A.R. Purcell, Maja-Olivia Lenz, Stefano Curtarolo, and Matthias Scheffler

In the last two years, the group has implemented a *high-throughput* framework for the efficient and high-precision prediction of vibrational thermal conductivities $\kappa(T)$ using the *ab initio* Green-Kubo (aiGK) approach [2]. This *ab initio* MD-based formalism developed earlier in the group is able to capture anharmonic effects to **all** orders and thus to accurately predict $\kappa(T)$ in solids, even at elevated temperatures and/or for strongly anharmonic materials, e.g., for materials close to a phase transition. Accordingly, this formalism is uniquely suited to aid and guide the search for novel thermoelectric materials, which require ultra-low $\kappa(T)$ and thus strongly anharmonic dynamics, well beyond the regime in which perturbative approaches are applicable. Clearly, an exhaustive *high-throughput* search for such thermal insulators is only possible if good conductors are disregarded early in the search process, even before computational time is invested in calculating their $\kappa(T)$. For this purpose, we

have developed an analytical metric $r^2(T) \sim 1 - \langle (\mathbf{F}^{\text{DFT}} - \mathbf{F}^{\text{ha}})^2 \rangle$ that quantitatively captures the degree of anharmonicity in the nuclear dynamics. It relies on the statistical comparison of the fully first-principles anharmonic forces \mathbf{F}^{DFT} experienced by the nuclei under specific thermodynamic conditions $\langle \cdot \rangle$ with those forces \mathbf{F}^{ha} that would act in the *harmonic* approximation. Accordingly, an essentially harmonic system features $r^2(T)$ values close to unity, whereas strongly anharmonic systems feature $r^2(T)$ values below 0.8. Since $r^2(T)$ can be obtained in a robust fashion within few force evaluations – several orders of magnitude faster than $\kappa(T)$ – we are able to scan over the materials space in a *high-throughput* fashion, covering thousands of materials including several complex thermoelectric material classes (half-Heuslers, inorganic halide perovskites, layered chalcogenides, etc.). This reveals that strongly anharmonic effects are much more common than hitherto believed. As shown in Fig. 1 for a representative test set of known materials obtained via the *AFLOW Library of Crystallographic Prototypes* [3], many compounds are well described by a perturbative quasi-harmonic description at 300 K ($r^2 > 0.8$). For 900K, however, this only holds for 55% of the materials. Over 20% of them, especially complex functional materials like perovskites, feature $r^2(T)$ values lower than 0.8. Accordingly, harmonic and anharmonic effects are of the same order and anharmonic effects can thus no longer be treated in a perturbed harmonic theory. To obtain the thermal conductivity $\kappa(T)$ of such materials, we perform *high-throughput aiGK* calculations, revealing that these compounds indeed feature low and ultra-low conductivities $\kappa(T) \ll 10$ W/mK. We demonstrate that accounting for all orders of anharmonicity is critical in this regime to obtain correct insights, i.e., quantitative predictions and a qualitative understanding of the atomistic mechanisms.

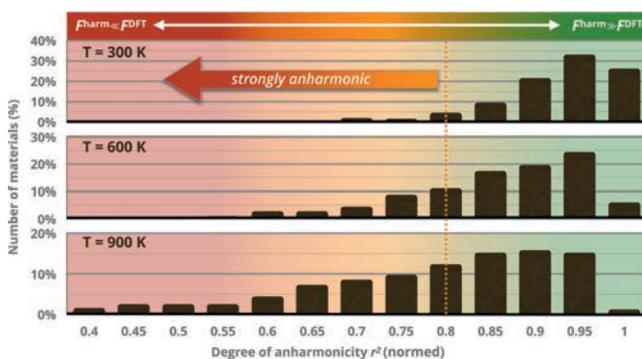


Figure 2: Degree of anharmonicity $r^2(T)$ obtained from first-principles with the PBE functional at 300, 600, and 900K for a representative set of over 300 materials, including over 100 perovskites structures. For compounds with $r^2(T) < 0.8$ (orange line) harmonic and anharmonic effects are comparable in magnitude.

Electronic Transport beyond Perturbation Theory

With: Marios Zacharias, Zhenkun Yuan, and Matthias Scheffler

The frequent emergence of strongly anharmonic effects in complex materials discussed above raises the question, to what extent these effects also determine electronic transport in solids. Existing perturbative formalisms [1] are unsuitable to resolve this doubt, since they intrinsically rely on a harmonic

description of the nuclear motion. Furthermore, they rely on an additional, potentially problematic approximation [4], i.e., a first-order model for the electron-phonon coupling. To overcome **both** these approximations and to account for **all** orders of phonon-phonon and electron-phonon coupling, we have developed a first-principles theory of electronic transport, in which the electronic flux, i.e., the time derivative of the electronic polarization $\mathbf{P}(t)$, is computed along aiMD trajectories. Since $\mathbf{P}(t)$ requires computing the *Berry phase* on a dense \mathbf{k} -grid [5], this approach seems at first incompatible with aiMD calculations in large supercells. We circumvent this via a newly developed electronic-structure unfolding scheme: The time-dependent, single particle wave-functions $\Psi_{\mathbf{k},n}^{\text{aiMD}}(t) = \sum_{j,\mathbf{G}} C_{\mathbf{k},\mathbf{k}+\mathbf{G}}^{n,j}(t) \psi_{\mathbf{k}+\mathbf{G},j}^{\text{ref}}$ in the aiMD supercell are expressed in terms of a static reference $\psi_{\mathbf{k}+\mathbf{G},j}^{\text{ref}}$, i.e., the wave-functions of the primitive unit cell. Here, \mathbf{G} denotes the reciprocal vectors connecting the respective Brillouin zones. By this means, the Berry phase can be evaluated in the primitive unit cell; the time-dependent expansion coefficients $C_{\mathbf{k},\mathbf{k}+\mathbf{G}}^{n,j}(t)$ then modulate the individual contributions, thus capturing the anharmonic electron-phonon couplings. Eventually, the spectral function $A_{\mathbf{k},j}(\epsilon, T) \sim \langle \sum_n |C_{\mathbf{k},\mathbf{k}+\mathbf{G}}^{n,j}(t)|^2 \rangle$, i.e., the T -dependent electronic band-structure in the primitive unit cell, is computed from the thermodynamic average of $C_{\mathbf{k},\mathbf{k}+\mathbf{G}}^{n,j}(t)$ obtained via aiMD in supercells. Analogously, the lifetimes $\tau_{\mathbf{k},j}(T)$ are computed from the correlation functions of $C_{\mathbf{k},\mathbf{k}+\mathbf{G}}^{n,j}(t)$, thus providing a direct route to evaluate electronic transport. As an example, Fig. 2 shows the spectral function $A_{\mathbf{k},j}(\epsilon, T)$ obtained for silicon at 0K as well as the T -dependence of the indirect band-gap, as extracted from a series of $A_{\mathbf{k},j}(\epsilon, T)$ calculations for different T . For this validation calculation, a quantum-mechanical harmonic model is used for the nuclear motion, so that our approach is in excellent agreement with perturbative theoretical results. Disagreement with experiment becomes noticeable at higher temperatures. Since our approach is also fully compatible with anharmonic aiMD, ongoing investigations on simple as well as on complex functional materials aim at understanding the role of anharmonicity for electronic self-energies and for the resulting transport coefficients.

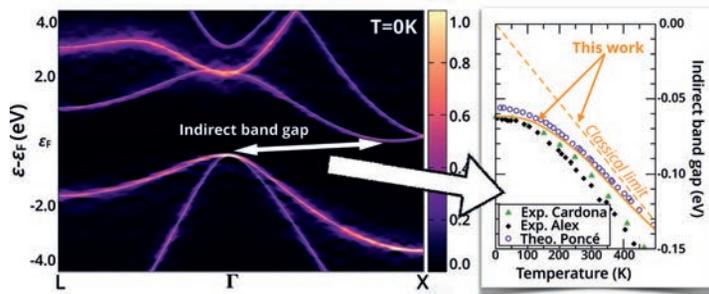


Figure 3: Spectral function $A_{\mathbf{k},j}(\epsilon, 0\text{K})$ (left) and temperature-dependence of the indirect band gap in silicon (right). All results obtained within the harmonic approximation for the nuclear dynamics to allow a direct comparison with the also shown perturbative literature results.

ELPA-AEO: Towards Eigensolvers for Materials Science at the Exascale

With: Hagen-Henrik Kowalski, Danilo S. Brambila, and Matthias Scheffler

The above described DFT calculations in extended supercells come at the cost of considerable computational effort; the bottleneck of which is solving

(generalized) eigenproblems. To exploit existing and upcoming exascale high-performance architectures, e.g., hybrid CPU-GPU installations, the group participated in the BMBF-funded ELPA-AEO project¹⁹, coordinated by Hermann Lederer (MPCDF), aimed at improving and optimizing the eigensolver library *ELPA*. In particular, the group was involved in guiding its development and evaluating the achievements. Due to this effort, *ELPA* nowadays “outperforms *ScalLAPACK* on all architectures tested and all matrix sizes and concurrencies”, as recently found by NERSC and Cray [6]. Furthermore, these advantages are accessible in an unprecedented user-friendly fashion, since performance-critical parameters are chosen in an automated way.

Numerical Quality Control for DFT-based Materials Databases

With: Björn Bieniek, Daniel Speckhard, Claudia Draxl, and Matthias Scheffler

Despite the widespread use of DFT calculations in *high-throughput* studies, little is yet known about code- and method-specific numerical errors that arise when the basis sets and k -grids commonly chosen in such workflows are used. Hence, potentially uncontrollable uncertainties prevent re-using data from different sources or created for different target properties, in spite of the fact that the data may be publicly available. To overcome this, we have studied the convergence of different key properties (band gaps, total and relative energies) in four conceptually-different DFT codes (*exciting*, *FHI-aims*, *GPAW*, and *VASP*) for typical settings used in production calculations, but also for extremely precise settings that serve as fully converged reference to evaluate the numerical errors. On the basis of data obtained for 71 elemental solids, we have proposed an analytical model that allows total-energy errors for any compound to be estimated, as demonstrated for 73 binary and ternary solids. We are investigating the extensibility of our approach towards complex materials properties such as band gaps.

References

1. Giustino, F.: Electron-phonon interactions from first principles. *Review of Modern Physics* **89**, 015003 (2017)
2. Carbogno, C., R. Ramprasad and M. Scheffler: *Ab Initio* Green-Kubo Approach for the Thermal Conductivity of Solids. *Physical Review Letters* **118**, 175901 (2017)
3. Mehl, M. J., *et al.*: The AFLOW Library of Crystallographic Prototypes. *Computational Material Science* **136**, S1 (2017) and **161**, S1 (2019)
4. Saidi, W. A., S. Poncé, and B. Monserrat: Temperature Dependence of the Energy Levels of Methylammonium Lead Iodide Perovskite from First-Principles. *Journal of Physical Chemistry Letters* **7**, 5247 (2016)
5. King-Smith, R.D. and D. Vanderbilt: Theory of polarization of crystalline solids. *Physical Review B* **47**, 1651 (1993)
6. Cook, B., *et al.*: Eigensolver performance comparison on Cray XC systems. *Concurrency and Computation: Practice and Experience* **180**, e4997 (2018)

¹⁹ <http://elpa-aeo.mpcdf.mpg.de>



Otto Hahn Group Leader
Dr. Mariana Rossi

Independent Otto Hahn Group Leader at the FHI (Oct. 2016–Dec. 2019)
Visiting Researcher, EPF Lausanne, Switzerland (2016)
Junior Fellow in St. Edmund Hall, Oxford, UK (2013–2015)
Dr. rer. nat., Technische Universität Berlin (2012)

Activities of the Otto Hahn Group “Simulations from Ab Initio Approaches: Structure and Dynamics from Quantum Mechanics”

General Remarks

The Otto Hahn Group, led by Mariana Rossi, is completing its third year in October 2019. In this period, the group has been pushing back the boundaries of atomistic simulations of weakly bonded systems at finite temperatures. The group's four PhD students and two postdocs are funded through the budget allocated to the Otto Hahn Group and through Dr. Rossi's third party-funding. This consists of a sub-project within the Collaborative Research Centre 951 (Hybrid Inorganic/Organic Systems for Opto-Electronics - HIOS), the Max Planck Research Network on Big-Data-Driven Materials Science (BiGmax) and the Max Planck-EPFL Center for Molecular Nanoscience and Technology. One student is funded through the International Max Planck Research School (IMPRS) for functionalized interfaces. Since August 2018, Mariana Rossi is also the main partner of the Max Planck Tandem Group on “Atomistic Simulations for Electrochemistry”, led by Luana Pedroza at the UFABC, São Paulo, Brazil. The group has long-term collaborations with the experimental groups of Takashi Kumagai (PC Department, FHI) and Norbert Koch (Humboldt-Universität zu Berlin). New collaborations in the framework of the CRC-951 (third funding period, since June 2019) have been established with Katharina Franke (Freie Universität Berlin), Ralph Ernstorfer (FHI), and Christoph Koch (Humboldt-Universität zu Berlin). In the past two years, the group has also collaborated with other theory groups, for example those led by Jörg Behler (Georg-August-Universität Göttingen), Jeremy Richardson (ETH Zürich), Michele Ceriotti (EPF Lausanne), Patrick Rinke (Aalto University), and Volker Blum (Duke University).

In January 2020, the group will move to the Max Planck Institute for Structure and Dynamics of Matter in Hamburg, where Mariana Rossi will be taking up a tenure track position within the scope of a Lise Meitner Group. She will continue with a 20% appointment at the FHI for the duration of the third funding period of the CRC-951, which will end in June 2023.

Scientific Scope

The research in our group focuses on including temperature and nuclear quantum effects in electronic-structure calculations, for systems where these effects are particularly important. Especially in weakly bonded systems, the nuclear fluctuations induced by temperature can completely change the structure, the thermodynamic equilibrium, and also the electronic structure of materials. In addition, given the ubiquitous presence of light nuclei in organic materials, not only the electrons but also the nuclear fluctuations must be

treated within the first principles of quantum mechanics. We achieve an all-atom, all-electron quantum description of these systems by developing methodology and tools that allow joining density-functional-theory calculations to path-integral methods. We belong to the core developers of the FHI-aims electronic-structure package [1] and the i-PI software [2], which performs nuclear dynamics. As described in detail below, this combination has allowed us to solve questions related to level broadening in semiconductor interfaces, isotope effects in electron-phonon coupling and apparently contradictory observations in hydrogen-transfer reactions. None of this would be possible if the atomic structure of the systems was unknown or incorrect. This is why we also conduct exhaustive *ab initio* structure searches of single molecules and of self-assembled monolayers as well as augment genetic-algorithm program packages in order to make these searches more efficient (see Fig. 1).

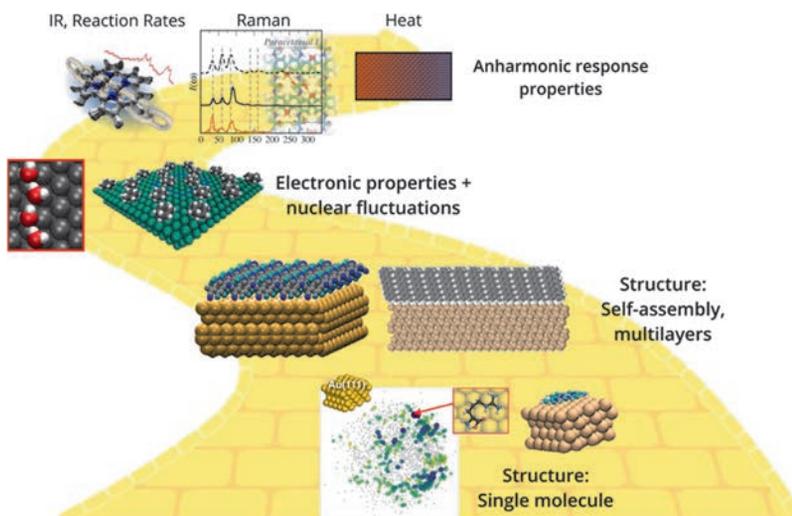


Figure 1: Research topics explored in the group. For isolated molecules on surfaces, we develop strategies for conformational space exploration that can handle the flexibility of the adsorbates. We address self-assembled organic structures at interfaces including the impact of atomic structure on the electronic-level alignments. We also include anharmonic nuclear quantum effects in electron-phonon coupling at weakly bonded interfaces. Finally, we develop novel techniques for the calculation of anharmonic nuclear response properties including nuclear quantum effects.

Research highlights

Hydrogen Transfer Reactions in High-Dimensional Anharmonic Systems

With: Yair Litman, Jeremy Richardson (ETH), and Takashi Kumagai (PC Department)

Intra- and inter-molecular hydrogen transfer is strongly impacted by zero-point energy and tunneling. The inclusion of these effects in high-dimensional anharmonic systems represents a considerable challenge in atomistic simulations. We implemented and derived optimizations for the ring-polymer instanton method [3] and thermostatted ring-polymer molecular dynamics [4] in the i-PI code. The former method gives access to reaction rates and the latter to vibrational spectra, both including zero-point energy and tunnelling effects,

and are applicable to high-dimensional anharmonic systems. For the paradigmatic example of porphycene, a prototype molecular switch, experimental intramolecular hydrogen transfer reaction rates show a controversial coexistence of Arrhenius (classical) behaviour and a strong isotope effect. Our simulations are able to explain fully the experimental data (see Fig. 2a) – a feat that had not been previously achieved [5]. In particular, the failure of the classical-nuclei description and apparent controversies for reaction rates are fully explained by the vibrational coupling between quasi-classical thermally activated modes and high-frequency quantized modes in this system. The framework we developed is easily generalizable to even more complex situations.

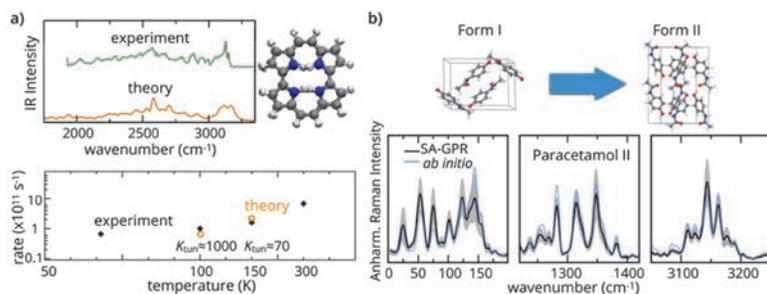


Figure 2: a) Top: Simulated anharmonic *ab initio* IR spectrum from approximate quantum dynamics and experimental spectrum at room temperature for the porphycene molecule. Bottom: Comparison between experimental tautomerization rates and simulated ones from approximate quantum dynamics, including the tunneling enhancement factors K_{tun} . See Ref. 3 and references therein. b) Gaussian process regression (GPR) prediction of the vibrational Raman spectrum of the paracetamol crystal. With the symmetry-adapted GPR (SA-GPR) framework, we obtained a model that is transferable between different polymorphic forms, see Ref. 4 and references therein.

Molecular Crystals at Finite Temperatures

With: Nathaniel Raimbault, Marcin Krynski, Andrea Grisafi (EPFL), and Michele Ceriotti (EPFL)

Structure prediction in molecular crystals is incomplete if the thermodynamic conditions (temperature, pressure) are not considered. Anharmonic contributions represent the largest contribution to free-energy differences of molecular crystal polymorphs composed of flexible molecular units. We have computed lattice changes at different conditions, confirming the essential role of van der Waals interactions for the stability and anisotropic lattice expansion and contraction of aspirin and paracetamol. Further, we have studied the accuracy of approximate methods to include anharmonic free-energy contributions (which were implemented in i-PI) for these systems, and found that methods based on normal-mode expansions fail when librational and rotational motion are present. Thermal free energies are especially sensitive to the low-frequency vibrational modes, which are also an important region to differentiate different crystal polymorphs. With the efficient density-functional-perturbation-theory implementation in FHI-aims, we were able to simulate anharmonic Raman spectra based on *ab initio* molecular dynamics trajectories. In order to cope with the added cost of calculating Raman intensities, we

devised a machine-learning model based on a symmetry-adapted Gaussian process regression algorithm (SA-GPR) that makes use of models trained for isolated molecules to learn dielectric responses of crystals [6]. The models are very accurate with only a few hundred points and, importantly, are transferable between different polymorphs (see Fig. 2b).

Hybrid Organic-Inorganic Interfaces and Electron-Phonon Coupling

With: Haiyuan Wang, Sergey V. Levchenko, Karen Fidanyan, Svenja M. Janke (Duke University), Volker Blum (Duke University), Thorsten Schultz (HU Berlin), Norbert Koch (HU Berlin) and Matthias Scheffler

A major challenge in describing organic-inorganic interfaces is predicting and controlling the structure-function relationship. The weakly-bonded organic components show flexibility and polymorphism, which are exacerbated by the typical temperatures at which these systems operate. We have studied the impact of strain on interfaces of silicon and pentacene films, finding that even low strain can strongly change the frontier-orbital alignment. When silicon is interfaced with strong organic electron acceptors, we found that the level alignment is more robust but the structure itself is strongly dependent on the strength of van der Waals interactions. Moreover, we predict a broadening of frontier orbitals due to anharmonic nuclear fluctuations at room temperature that far surpasses $k_B T$ [7]. Anharmonic electron-phonon coupling also explains the isotope effect in work-function changes that has been measured in metal-organic interfaces, showing that finite temperature nuclear quantum effects are relevant also for the electronic structure

This work was performed partly within the framework of the DFG CRC-951 (HIOS) project.

Other Research Fronts

A major challenge in predicting structure of flexible organic films and solids is navigating the conformational space of not only the self-assembled structure but also of the individual molecular constituents. In collaboration with Carsten Baldauf and the PhD student Dmitrii Maksimov, we improved a genetic-algorithm-based structure search that can handle flexible molecules in order to work at interfaces. In addition, we are investigating how to bias these searches to favour regions of lower free-energy minima under specific thermodynamic conditions.

In collaboration with the Tandem Group led by Luana Pedroza (UFABC, São Paulo, Brazil) we are implementing a framework to capture nuclear quantum effects at water interfaces relevant for electrochemistry. In particular, we have a setup based on the i-PI code and a density-functional non-equilibrium Green's function code that performs path-integral molecular dynamics with (self-consistent) forces from simulations where the potential bias is included explicitly. This bias has a direct impact on the structure of water at the interface, while nuclear quantum effects can strengthen (or weaken) intramolecular H-bonds and metal-hydrogen bonds. These effects play a direct role on interface degradation at different voltage biases and on the efficiency of the cells at different temperatures, but simulations have so far not been able to combine these physical interactions.

Regarding heat transport through materials, we wish to perform simulations for regimes where strong anharmonicity and quantum effects play a significant role. These regimes are of particular importance for organic electronics due to the anharmonic (weak) interactions and the ubiquitous presence of light atoms. We collaborate with Christian Carbogno and David Manolopoulos (Oxford University), as well as with the students Hagen-Henrik Kowalski and Benjamin Sutherland to achieve novel frameworks based on calculations of thermal diffusivities from non-equilibrium simulations and thermal conductivities directly from the Green-Kubo formalism. With these developments, we hope to achieve a robust approximate method that is efficient in combination with *ab initio* simulations.

References

1. Blum, V., R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler: *Ab initio* molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications* **180**, 2175-2196 (2009)
2. Kapil, V., M. Rossi *et al.*: i-PI 2.0: A universal force engine for advanced molecular simulations. *Computer Physics Communications* **236**, 214-223 (2019)
3. Richardson, J.O., and S.C. Althorpe: Ring-polymer instanton method for calculating tunneling splittings. *Journal of Chemical Physics* **134**, 054109 (2011)
4. Rossi, M., M. Ceriotti, and D.E. Manolopoulos: How to remove the spurious resonances from ring polymer molecular dynamics. *Journal of Chemical Physics* **140**, 234116 (2014)
5. Litman, Y., J. Richardson, T. Kumagai, and M. Rossi: Elucidating the Nuclear Quantum Dynamics of Intramolecular Double Hydrogen Transfer in Porphycene. *Journal of the American Chemical Society* **141**, 2526-2534 (2019)
6. Raimbault, N., A. Grisafi, M. Ceriotti, and M. Rossi: Using Gaussian Process Regression to Simulate the Vibrational Raman Spectra of Molecular Crystals. *New Journal of Physics*, in press (2019)
7. Wang, H., S.V. Levchenko, T. Schultz, N. Koch, M. Scheffler, and M. Rossi: Modulation of the Work Function by the Atomic Structure of Strong Organic Electron Acceptors on H-Si(111). *Advanced Electronic Materials* **5**, 1800891 (2018)



Dr. Luca M. Ghiringhelli

Group Leader at the FHI (since 2011)

Postdoc at the FHI (2008–2010)

Postdoc at the Max-Planck-Institut für Polymerforschung in Mainz (2005–2008)

Dr. rer. nat., Universiteit van Amsterdam (2006)

Activities of the “Big-Data Analytics for Materials Science” Group

The group “Big-Data Analytics for Materials Science” was established in 2015 with the research focus on developing and applying methods for finding patterns, trends, and analytical relationships in materials-science data, in order to establish what has now become the field of “data-driven materials science”. Primarily, the methods developed belong to the wide area of artificial intelligence (AI). AI is an umbrella term describing any technique that mimics (and potentially even surpasses) human intelligence, in particular its ability to infer regularities from empirical data. It embraces disparate areas from signal analysis, data mining, and statistical (or machine) learning. The developed AI algorithms are applied to relevant problems in materials science, in particular the identification of descriptors enabling the construction of simple *maps* (or *charts*) of materials properties (see Fig. 1 for an example). Below is a short account of the major advances within the group in the last two years.

SISSO, an Efficient, Massively Parallel Algorithm for the Compressed-Sensing- Based Identification of Materials-Property Maps

Method development with: Runhai Ouyang, Emre Ahmetcik, and Matthias Scheffler

Applications also with: Stefano Curtarolo, Christopher J. Bartel, Christopher Sutton, Bryan Goldsmith, and Charles B. Musgrave (University of Colorado Boulder)

Building on the breakthrough introduction in 2015 of LASSO+ ℓ_0 , a method that combines symbolic regression²⁰ and compressed sensing²¹ for identifying descriptors and accurate models from a set of candidate features, we developed SISSO [1]. This is a massively parallel algorithm that overcomes the limitation of LASSO+ ℓ_0 , i.e., it can cope with hugely and highly correlated dimensional feature spaces. SISSO stands for “Sure Independence Screening and Sparsifying Operator” and deals with feature spaces with billions or more elements. Together with the algorithm for identifying the best descriptor in the set of candidates, we also constructed a “self-diagnostic” technique that determines the maximum complexity of the descriptor by means of so-called cross validation²². Perhaps the most notable application was the identification

²⁰ Symbolic regression is a technique that searches the space of mathematical expressions in order to find the model (equation) that best fits a given data set. This is in contrast to parametric regression, e.g. the usual linear (least-square) regression, where only the parameters of fixed functional form are fit to the data.

²¹ Compressed sensing is a mathematical technique developed in signal analysis for solving massively underdetermined (fewer equations than unknowns) linear systems.

²² Cross validation is a family of techniques having in common the iterated splitting of the data set into training and test. The model is fitted to the training data, but the performance is

of a tolerance factor for predicting whether the ground-state crystal structure of a material with a given chemical formula within the prototype ABO_3 is perovskite [2]. Having been trained using about 600 experimental data points retrieved from literature, the descriptor found by SISO utilizes the same input features as the hundred-year-old, famous Goldschmidt tolerance factor t , confirming that the sizes of the atoms (namely, their Shannon radii) determine their spacing and their stability as perovskites, but proposes a slightly more complex equation. The newly found descriptor, dubbed τ , outperforms Goldschmidt's t in terms of accuracy and crucially allows a monotonic probability to be assigned as to whether a material is a perovskite or not. In practice, for very small (large) values of τ , the material is predicted as certainly (not) a perovskite, while for intermediate values, the probability ranges from 100% to 0%. We have exploited the model found by SISO by scanning all possible values of atomic radii and identified several perovskites candidates that were previously unknown. The model can be extended without loss of accuracy to double perovskites, by just inputting the average Shannon radii of the atomic species occupying the AA' , (or BB') positions. The latter circumstance highlights one of the big merits of the SISO-like descriptor and model learning: The relative simplicity of the descriptor learned and the very small number of fitting parameters (e.g., only one fitting parameter for τ) gives rise to a remarkable stability and extrapolation ability well beyond the training data set.

Learning Several Materials Properties Simultaneously from Incomplete Databases with Multi-Task SISO

With: Runhai Ouyang, Emre Ahmetcik, Christian Carbogno, and Matthias Scheffler

The SISO algorithm was generalized to so-called multi-task learning, a framework that allows for the simultaneous learning of several target properties by means of a family of models sharing certain characteristics. In the case of the newly developed multi-task (MT) SISO, all predicting models (one per target property) have in common the same descriptor (as usual identified from a huge set of candidates), while they differ in the fitting coefficients. The remarkable characteristic of MT-SISO is that it is not necessary to have information about all considered properties for all training data points. In fact, the training database can be rather incomplete and the enhanced robustness introduced by the simultaneous learning promotes the identification of an accurate model. MT-SISO was applied, among other things, to *crystal-structure prediction* (within Scheffler's ERC project TEC1p) of the relative formation energy of several (5) crystal structures for octet binary materials, see Fig. 1. MT-SISO promises to be a versatile approach, coping with incomplete databases and circumventing the present difficulty of single-task SISO, where it is problematic to characterize materials found in different crystal structures. Instead of finding an explicit descriptor encoding the crystal structures, the problem is shifted to the fitting coefficients, which are implicit functions of the crystal structures.

evaluated over the test data. The minimization of the test error allows for choosing the so-called hyperparameters, that in the case of SISO are those determining the complexity (how many input features and how many mathematical operators acting on them) of the learned formula.

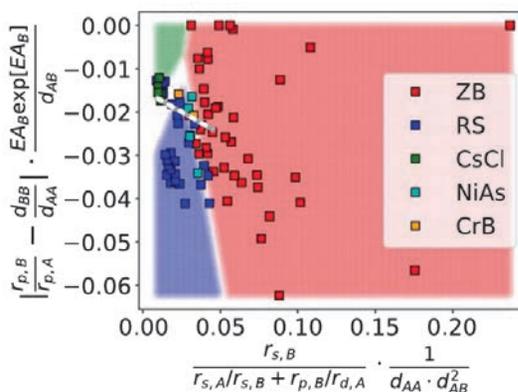


Figure 1: Map (phase diagram) of the ground-state crystal structure for octet binary materials. Five prototypes are considered: Zinc blende (ZB), rock salt (RS), CsCl, NiAs, and CrB. The extended colored regions correspond to the stability of the corresponding crystal phase, while the colored squares are the training points. The descriptor depends on s, p, d valence-orbital radii for the A or B component of the binary and the equilibrium distances of the isolated AA , BB , and AB dimers, and the electron affinity.

NOMAD Activity: The Kaggle Competition

With: Christopher Sutton, Angelo Ziletti, and Matthias Scheffler

In the context of the NOMAD (Novel Materials Discovery) Laboratory (see report of Claudia Draxl), a machine-learning competition was organized. By means of high-throughput density-functional-theory (DFT) calculations, the formation energy and band gap of 3,000 transparent conducting oxides (prototype formula, $\text{Al}_x\text{Ga}_y\text{In}_z\text{O}_3$, $x+y+z=2$) were calculated. The input data (chemical formula, atomic coordinates, and unit cell lattice vectors) were then provided to the participants of a competition hosted on the Kaggle platform for machine-learning contests²³. The aim of this challenge was to identify the best machine learning (ML) model for the prediction of two target properties.

As many as 883 teams participated. The first three best-performing models are described in Ref. [4]. The first place solution is a novel approach, based on both crystal-graph representation and chain of words (n -grams) borrowed from natural-language processing.

This competition is unprecedented in materials science and the possibility to analyze AI models designed and trained by independently working teams revealed valuable information, since the bias introduced by one team selecting AI models due to personal preferences is removed by the crowd-sourcing approach.

²³ <https://www.kaggle.com/c/nomad2018-predict-transparent-conductors>

Subgroup Discovery for the Analysis of the Performance of Machine-Learning Models.

With: Mario Boley (Monash University), Christopher Sutton, Matthias Rupp (Citrine Informatics), Jilles Vreeken (Helmholtz Center for Information Security), and Matthias Scheffler

In 2017, we adapted subgroup discovery (SGD) to materials-science problems. SGD is a data-mining technique that identifies subsets of the data with outstanding values of a given target property; the subgroups are described by means of descriptors identified within a set of candidates, in a spirit similar to SISSO. However, rather than fitting the target property, the algorithm looks for statistically exceptional subsets, where the target property exhibits, e.g., a narrow distribution of values and its average is much smaller than for all data points. In a recent project (paper to be submitted), we trained 3 ML models on the same dataset as for the Kaggle competition and analyzed the distribution of *test (prediction) errors*. In practice, we identified what we define *domains of applicability (DA)* of the three models, described in terms of the input parameters (e.g., the lattice vectors). These are regions of the data space where the model performs better, in terms of reduced prediction root mean-square error. This (meta)analysis of AI models by means of another AI algorithm can be used to guide the development and improvement of ML representations through the identification of their systematic strengths and weaknesses.

NOMAD Activity: Analytics Toolkit

With: Emre Ahmetcik, Angelo Ziletti, Benjamin Regler, Ádám Fekete (King's College London), Fawzi Mohamed, and Matthias Scheffler

In the context of the NOMAD Laboratory, the Analytics Toolkit was first introduced in 2017. This is an online library of python notebooks illustrating step-by-step AI projects developed within the NOMAD team. The projects described above, in particular SISSO, the winners of the Kaggle competition, and more are available at <https://analytics-toolkit.nomad-coe.eu/>. Users can reproduce the results published in the related papers, starting from the raw training data; furthermore, they can interactively modify the training parameters to build an understanding of their effect on the trained models.

Robust Crystal-Structure Recognition via Deep Learning

With: Angelo Ziletti, Andreas Leitherer, Devinder Kumar (University of Waterloo and Vector Institute of AI Toronto), and Matthias Scheffler

Automatically recognizing local arrangements of atoms in order to assign crystal-structure prototypes by similarity and localizing boundaries among crystalline regions is a challenging, as well as important, application of AI. Existent and widely applied methods are not robust, since they typically rely on human-defined tolerance and/or are limited to few prototypes. Furthermore, it is desirable to recognize the parent structure also in case of heavily defected (e.g., by the presence of vacancies or atomic displacements). By designing a suitable descriptor, based on the Fourier transform of the atomic structure, and by training a state-of-the-art deep convolutional neural network (ConvNet) developed in the image-recognition field, we have obtained a parameter-free algorithm able to classify several prototypes in possibly heavily defected structures [5]. Furthermore, by using a cutting-edge technique [6], we were able

to interpret the reasoning of the trained ConvNet, which turns out to use bright spots in the input image as an experienced human crystallographer would do.

In a recent development, we have introduced an alternative local descriptor, based on the expansion of the local environment over spherical harmonics and then averaged over the observed sample. With this as input for a newly trained ConvNet, we achieved the contrast to classify more than a hundred prototypes, including lower dimensional ones. We are able now to investigate structural transitions such as the Bain path between fcc and bcc structures, where the intermediate structures along the path are smoothly assigned to the expected crystal-structure prototypes even though the network is trained only on pristine structures.

Ongoing Projects

- *(with Benjamin Regler)* Development of an information-theory based score for assessing the content of information in a set of features with respect to a target property. This approach does not assume or search for any functional form for the feature set \rightarrow property relationship, but rather quantifies how much a set of features “knows about” a target property. This algorithm would run preliminary to a symbolic regression or in general a ML step where the actual model for the selected features is trained. For this project, the theoretical development is essentially fulfilled, and few proof-of-concept applications are being carried out.
- *(with Emre Ahmetcik)* Development of a ML empirical potential for crystal-structure prediction (within Scheffler’s ERC project TEC1p) across chemical space, such that the fitting parameters depend explicitly on the chemical species involved. These are characterized, for example, by their row and column in the periodic table, electronegativity, etc. The purpose is to be able to transfer the interacting potential from species to species with a sparse database, where only (relatively) few data points for each species and configuration are available. Preliminary tests show very promising results.
- *(with Yuanyuan Zhou)* Development of a potential-energy-surface efficient sampling technique in the grand-canonical ensemble. Based on the extension of the replica-exchange framework to the grand-canonical ensemble, this method is developed having in mind the unbiased and beyond-harmonic study of reactive surfaces in reactive atmospheres at finite temperature and pressure. An extensive application to the Si(111) surface in contact with H_2 , modeled at the gradient-corrected DFT level, is under development. This work is going to be part of the Max-Planck-Cardiff Centre on the Fundamentals of Heterogeneous Catalysis ‘FUNCAT’.

References

1. Ouyang, R., S. Curtarolo, E. Ahmetcik, M. Scheffler, and L. M. Ghiringhelli: SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* 2, 083802 (2018)
2. Bartel, C.J., C. Sutton, B.R. Goldsmith, R. Ouyang, C.B. Musgrave, L. M. Ghiringhelli, and M. Scheffler: New Tolerance Factor to Predict the Stability of Perovskite Oxides and Halides. *Science Advances* 5, no. 2, eaav0693 (2019)
3. Ouyang, R., E. Ahmetcik, C. Carbogno, M. Scheffler, and L. M. Ghiringhelli: Simultaneous Learning of Several Materials Properties from Incomplete Databases with Multi-Task SISSO. *Journal of Physics: Materials* 2, 024002 (2019)
4. Sutton, C., L. M. Ghiringhelli, T. Yamamoto, Y. Lysogorskiy, L. Blumenthal, T. Hammerschmidt, J. Golebiowski, X. Liu, A. Ziletti, and M. Scheffler: NOMAD 2018 Kaggle Competition: Solving Materials Science Challenges Through Crowd Sourcing. *NPJ Computational Materials*, in press. <https://arxiv.org/abs/1812.00085> (2019)
5. Ziletti, A., D. Kumar, M. Scheffler, and L. M. Ghiringhelli: Insightful classification of crystal structures using deep learning. *Nat. Commun.* 9, 2775 (2018)
6. Kumar, D., V. Menkovski, G.W. Taylor, and A. Wong: Understanding anatomy classification through attentive response maps, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, 1130-1133 (2018)



Max Planck Fellow **Prof. Dr. Claudia Draxl**

Max Planck Fellow of the MP Graduate Center for Quantum Materials (since July 2019)

Max Planck Fellow at the FHI (July 2014–June 2019)

Einstein Professor at the Humboldt-Universität zu Berlin (since 2011)

Habilitation Universität Graz (1996)

Dr. rer. nat., Universität Graz (1988)

Activities of the Max Planck Fellow Group

The main focus of the Max Planck Fellow group centers on data-driven research, covering all important aspects, from the creation of high-quality benchmark data to the design of FAIR [1] data infrastructure and the development and application of artificial-intelligence (AI) tools. Through the close interaction with the FHI, the NOMAD Laboratory has developed into a world-leading platform for materials discovery (see also Poster *From NOMAD to FAIR-DI*).

The NOMAD Laboratory

The Novel Materials Discovery (NOMAD) Laboratory is a unique user-driven platform for sharing and exploiting computational materials science data for speeding up research in materials and engineering. With the NOMAD Repository²⁴ and the NOMAD Archive, its code-independent and normalized form, the NOMAD Laboratory represents the worldwide largest data collection in this field. Based on its FAIR (findable, accessible, interoperable, reusable) data infrastructure, various services are offered, comprising the NOMAD Encyclopedia²⁵, advanced visualization, and artificial-intelligence tools as realized in the NOMAD Analytics Toolkit²⁶. Prerequisite for all this is the NOMAD metadata, a hierarchical description of the data produced by all important computer codes in the community. Users can request a DOI to make data citable.

NOMAD is orthogonal to, but interacts synergistically with, other data collections worldwide. A new pipeline for automatic ingest from the *Materials Project* (MP)²⁷ input and output files is currently installed, which can be downloaded by interested users from NOMAD. Likewise, data from the *Automatic FLOW for Materials Discovery* (AFLOW)²⁸ and *The Open Quantum Materials Database* (OQMD)²⁹ have all been ingested into NOMAD.

²⁴ <https://repository.nomad-coe.eu>

²⁵ <https://encyclopedia.nomad-coe.eu>

²⁶ <https://analytics-toolkit.nomad-coe.eu>

²⁷ <https://materialsproject.org>

²⁸ <http://aflow.org>

²⁹ <http://oqmd.org>

The NOMAD Laboratory – in the meantime well established – has recently been incorporated into a non-profit association, FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V. (FAIR-DI e.V.)³⁰. Although an association under German law, it is international in nature and has close connections to the European Open Science Cloud (EOSC). The latter initiative is dedicated to providing a stable and sustainable data infrastructure that hosts and further develops the NOMAD Repository, Archive, and Encyclopedia, so as to guarantee long-term viability, independent of the individual research projects currently driving it. It also forms an ideal basis for advancing data-driven science, for example, also addressing experimental studies. First steps in this direction have already been taken and also involve other departments at the FHI: Together with Ralph Ernstorfer and Martin Wolf (PC Department) we are tackling the FAIRness of time-resolved angle-resolved photoemission spectroscopy (tr-ARPES) data, while with Annette Trunschke and Robert Schlögl (Inorganic Chemistry) we focus on catalysis data.

All the above are components of FAIRmat³¹ a consortium, led by Claudia Draxl and co-led by Matthias Scheffler, with significant involvement of Carsten Baldauf, which is running for the National Research Data Infrastructure (NFDI). This is a large-scale project of the BMBF (Federal Ministry of Education and Research, Germany) with a proposal deadline on October 15, 2019. FAIRmat brings together individual researchers, research institutions, joint research programs, and data and computing centers from all over Germany with the aim of establishing a FAIR data infrastructure for the wide field of materials science. The latter comprises all the sub-communities in condensed-matter science and reaches out to solid-state chemistry, catalysis, and soft matter. It covers synthesis, as well as experimental and computational materials science and will demonstrate its potential for selected applications such as batteries, heterogeneous catalysis, optoelectronics, spintronics, and AI.

We emphasize that such an infrastructure is the crucial basis for data science. After demonstrating the potential of compressed sensing for learning descriptors for materials properties, [2] i.e., parameters that capture the physical mechanisms underlying certain material properties, the methodology has been further refined [3] (see also the report by Luca M. Ghiringhelli). In the PhD thesis of Benedikt Hoock (co-supervised with Matthias Scheffler), this approach has been applied in order to predict lattice parameters and energetics of ternary alloys. In this context, various cross-validation-based model-selection strategies have been introduced and compared, at the same time probing various ways for choosing the feature space [4]. Other activities are dedicated to exploring similarities of materials in terms of properties and structures.

Critical for operating in a huge data space are validation and verification of the data. The most severe restrictions that may hamper interoperability concern the computational parameters that are used in different calculations, thus

³⁰ <https://fairdi.eu>

³¹ <https://fairdi.eu/fairmat>

affecting their numerical precision. Data are typically created for different purposes which may require different levels of convergence in terms of basis-set size and alike. We have to ask as well whether different codes aiming at the solution of one and the same problem give the same results. For the latter, the community effort led by Stefaan Cottenier showed for the equations of state for 71 elemental solids, that over a time span of a few years, many different *ab initio* electronic-structure packages finally led to basically the same answer [5]. While such investigations have set the stage towards a culture of benchmarking in the field of computational solid-state science, clearly, other properties like energy barriers, band gaps, spectra, etc., and systems like surfaces, interfaces and inorganic/organic hybrid materials, etc. will be much less forgiving than total energies of simple bulk solids and will make discrepancies more obvious. Also, the comparison of codes could only be made with parameter sets that represent full convergence, while calculations performed on a regular basis are often far from this optimal case. So, the question is how to compare and operate on data that have been obtained with different settings, e.g. in terms of basis sets, meshes for Brillouin-zone integrations, and other parameters. That this is, in fact, possible, i.e., that fully converged results of complex materials can be estimated by learning from errors of calculations of simple materials has been recently demonstrated by employing four different codes, two of them being developed in Berlin (*exciting* in Draxl's group, FHI-aims in Scheffler's group). For details, we refer to the report by Christian Carbogno.

Finally, we note that Claudia Draxl has been an organizer of the Long Program *Science at Extreme Scales: Where Big Data Meets Large-Scale Computing* (Sep.12–Dec. 14, 2018) at the Institute for Pure & Applied Mathematics (IPAM) at the University of California, Los Angeles (UCLA) to bring together innovators and pioneers from applied mathematics, computer science, and various applications areas to catalyze the new, emerging field linking HPC with Big Data³².

The Cluster-Expansion Package CELL

With: Santiago Rigamonti (HU Berlin), Maria Troppenz (HU Berlin), Martin Kuban (HU Berlin), Axel Hübner (HU Berlin), Christopher Sutton, Luca M. Ghiringhelli, and Matthias Scheffler

Predicting stability and properties of alloys is a challenging task due to the need to evaluate relative energies of many atomic configurations accessible at realistic temperatures. A direct *ab initio* approach is typically hampered by the significant amount of time required to evaluate these energies. By mapping the configuration-dependent properties of the material to an Ising-like Hamiltonian parametrized with *ab initio* data, the cluster expansion (CE) approach reduces this time by several orders of magnitude. Still, for complex alloys with large parent cells, traditional CE approaches are not feasible.

The python package CELL³³ has been devised to deal with such situations by allowing for an iterative solution of the CE and, as a result, can handle materials with complex primitive cells comprising 30 atoms and more. An impressive

³² <http://www.ipam.ucla.edu/programs/long-programs>

³³ <https://sol.physik.hu-berlin.de/cell>

example for this capability has been demonstrated by predicting ground-state configurations and electronic properties of thermoelectric clathrates [6]. CELL can be applied to a broad variety of cases, covering one-, two- and three-dimensional materials, as well as multi-component and multi-lattice systems. A joint ongoing project between Draxl and Scheffler is on the *ab initio* atomistic thermodynamics of adsorbed atomic hydrogen at the (10-10) ZnO surface (in preparation).

Other joint activities

Claudia Draxl and Matthias Scheffler co-supervise three PhD students: Benedikt Hooch, working on data analytics (see work on compressed sensing above); Daniel Speckhard, working on verification and validation aspects of computational materials data (going beyond what is described above); Konstantin Lion, within the LeibnizCampus GraFOx on surfaces and growth of group-III oxides (see report by Sergey V. Levchenko).

References

1. Wilkinson, M.D., *et al.*: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016)
2. Ghiringhelli, L.M., J. Vybiral, S.V. Levchenko, C. Draxl, and M. Scheffler: Big Data of Materials Science: Critical Role of the Descriptor. *Physical Review Letters* **114**, 105503 (2015)
3. Ghiringhelli, L.M., J. Vybiral, E. Ahmetcik, R. Ouyang, S.V. Levchenko, C. Draxl, and M. Scheffler: Learning physical descriptors for materials science by compressed sensing. *New Journal of Physics* **19**, 023017 (2017)
4. Hooch, B., S. Rigamonti, L.M. Ghiringhelli, M. Scheffler, and C. Draxl, in preparation
5. Lejaeghere K., *et al.*: Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000 (2016)
6. Troppenz, M., S. Rigamonti, and C. Draxl: Predicting ground-state configurations and electronic properties of the thermoelectric clathrates $\text{Ba}_8\text{Al}_x\text{Si}_{46-x}$ and $\text{Sr}_8\text{Al}_x\text{Si}_{46-x}$. *Chemistry of Materials* **29**, 2414 (2017)



Max Planck Partner Group Leader Prof. Dr. Xinguo Ren

Max Planck Partner Group Leader (since 2015)
Professor at the University of Science and Technology of China (since 2013)
Postdoc at the FHI (2006–2012)
Dr. rer. nat., Universität Augsburg (2006)

Activities of the Max Planck Partner Group on “Advanced Electronic-Structure Methods”

Introduction

The Max Planck Partner Group on Advanced Electronic-Structure Methods was set up in December 2015 at the University of Science and Technology of China (USTC). Xinguo Ren was appointed as the Head of this Partner Group. The group is hosted by the FHI Theory Department, led by Matthias Scheffler. The basic task of the Partner Group is to develop cutting-edge electronic-structure methods and the corresponding computer code to deal with challenging problems in computational materials science. During the past few years, the Partner Group has collaborated closely with the Theory Department of the FHI on multiple common projects, facilitated amongst other things by several mutual visits.

Below, we describe the research efforts of the Partner Group, on both implementing existing electronic-structure methods in a new numerical framework, and developing new concepts and methodologies.

Implementation of Periodic G_0W_0 within an All-Electron, Numerical Atomic Orbital Basis-set Framework

With: Florian Merz (Lenovo HPC Innovation Center), Hong Jiang (Peking University), Hermann Lederer (MPCDF), Volker Blum (Duke University), and Matthias Scheffler

We have implemented the periodic G_0W_0 method for quasiparticle energy calculations within an all-electron, numerical atomic orbital (NAO) basis-set framework. A straightforward implementation of such correlated methods within a NAO framework results in a computationally expensive code. Here we employ a localized variant of the resolution of identity (RI) approximation, enabling a significant reduction of the required CPU time and memory consumption. We demonstrate that the error arising from the localized RI approximation is controllable and can be made negligibly small by enhancing the set of auxiliary basis functions (ABFs). A particularly effective algorithm, suitable for the NAO framework, has been developed to deal with the Coulomb singularity in the Brillouin zone sampling.

Our implementation is carried out within the FHI-aims software package. In Fig. 1, we present the computed G_0W_0 band gaps, on top of the PBE starting point, versus the experimental values for a selected set of semiconductors and insulators. The three sets of all-electron G_0W_0 results presented here are obtained by the FHI-aims code with a standard NAO *tier 2* basis set, and by the FHI-gap code with the linearized augmented plane wave (LAPW) basis as well as

with the LAPW complemented by high-energy localized orbitals (HLOs), respectively. For most of the materials, our implementation yields band gap values in close agreement (within 0.2 eV) with those obtained by the LAPW+HLOs framework. However, for systems with a strong ionic character, such as LiF and ZnO, previous investigations indicate that the standard LAPW basis set is not sufficient to converge the G_0W_0 gap. Adding HLOs significantly enlarges the band gap up to 0.7 eV for ZnO and 0.9 eV for LiF. Our standard NAO basis set, though surpassing the standard LAPW basis set in performance, suffers from a similar under-convergence issue. We are currently working to resolve this issue by complementing our standard *tier 2* basis set with highly localized Slater-type orbitals (STOs). Preliminary results for ZnO shows that this strategy works rather well.

Implementation of Periodic RPA within a NAO Basis-Set Framework

With: Florian Merz (Lenovo HPC Innovation Center), Igor Ying Zhang (Fudan University), Hermann Lederer (MPCDF), Volker Blum (Duke University), and Matthias Scheffler

Our periodic RPA (Random-Phase Approximation) implementation follows a similar numerical algorithm as G_0W_0 , and has been used to compute the binding energy curves of bulk materials and layered materials. Benchmark calculations for a set of 14 solid materials have been reported previously [1]. We would like to point out our periodic RPA implementation also includes a correction arising from renormalized single excitation contributions (rSE) [2], which can further increase the accuracy of binding-energy calculations. In Fig. 2 we present benchmark RPA and RPA+rSE calculations for bilayer boron nitride (BN), compared with results from other functionals.

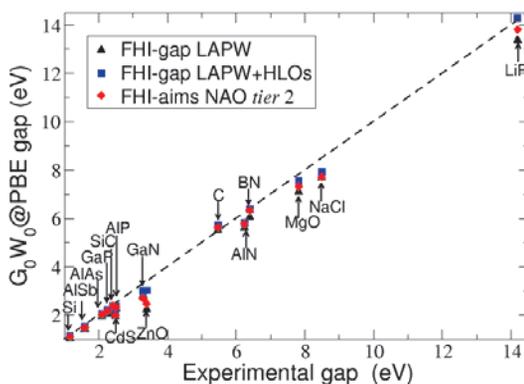


Figure 1: The calculated G_0W_0 @PBE band gaps versus the experimental gap for a sequence of semiconductors and insulators. Presented theoretical results are obtained by the FHI-aims code using the NAO *tier 2* basis set (red diamonds, this work), and by the FHI-gap code with the LAPW (black triangles, and LAPW+HLOs (blue squares) basis sets

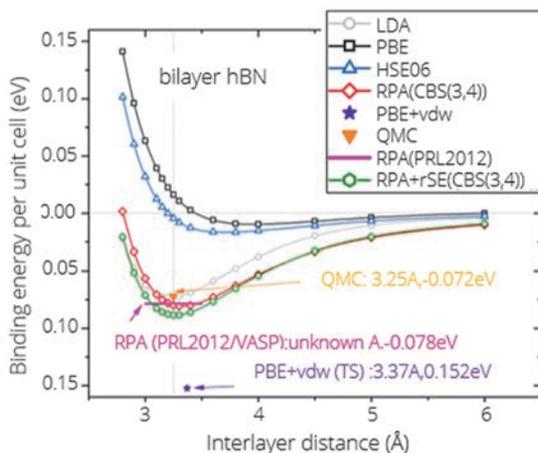


Figure 2: Binding energy curves of a double-layer BN obtained by Local Density Approximation (LDA), Perdew-Burke-Ernzerhof (PBE) Generalized Gradient Approximation, Heyd-Scuseria-Ernzerhof (HSE) hybrid functional, RPA, and RPA+rSE [2].

Particle-Particle RPA and Its Combination with Standard Particle-Hole RPA

With: *Muhammad N. Tahir (USTC)*

In nuclear physics, the usual RPA used in condensed matter physics and quantum chemistry is called particle-hole RPA (ph-RPA). In addition to the ph-RPA, a particle-particle channel of RPA (pp-RPA), corresponding to the ladder diagrams, has also been discussed. Weitao Yang's group from Duke University has played a pioneering role in bringing the pp-RPA to the attention of the electronic-structure community [3]. It seems that pp-RPA does not suffer from the charge delocalization error as the ph-RPA does. We implemented pp-RPA in FHI-aims, with the aim to develop a computational scheme that combines pp-RPA and ph-RPA. We analyzed the critical role of the exchange-type interactions, which are included in pp-RPA and usually not in ph-RPA. Unfortunately, although pp-RPA and ph-RPA are diagrammatically distinct, a straightforward combination of the two RPA channels does not work well. We carried out an analysis which helps us to understand why this is the case [4]. At the moment, we are trying to combine the two channels of RPA in a range-separation framework, which has been shown to work for a homogeneous electron gas.

Force Calculations within the RPA method

With: *Muhammad N. Tahir (USTC), Tong Zhu (Duke University), and Volker Blum (Duke University)*

The RPA force calculation is an important step in developing RPA into a viable method in computational materials science. In complex systems where different bonding mechanisms compete with each other, the geometry often cannot be reliably determined. Geometry relaxations with the RPA method should be very helpful in such situations. In the literature, RPA force implementations based on the Gaussian orbital basis sets and the projector augmented wave (PAW) method have been reported. In the present work, we have derived the RPA force formalism within the NAO framework and implemented it in the FHI-aims code.

Benchmark results for small molecules indicate that our derivation and implementation are correct. At the moment we are working to improve the implementation efficiency in order to treat more complex and interesting systems. Next, we will extend the implementation to periodic systems.

References

1. Zhang, I.Y., A.J. Logsdail, X. Ren, S.V. Levchenko, L.M. Ghiringhelli, and M. Scheffler: Main-group test set for materials science and engineering with user-friendly graphical tools for error analysis: systematic benchmark of the numerical and intrinsic errors in state-of-the-art electronic-structure approximations. *New Journal of Physics* **21**, 013025 (2015)
2. Ren, X., P. Rinke, G.E. Scuseria, and M. Scheffler: Renormalized second-order perturbation theory for the electron correlation energy: Concept, implementation, and benchmarks. *Physical Review B* **88**, 035120 (2013)
3. van Aggelen, H., Y. Yang, and W. Yang: Exchange-correlation energy from the pairing matrix fluctuation and the particle-particle random-phase approximation. *Physical Review A* **88**, 030501 (R) (2013)
4. Tahir, M.N., and X. Ren: Comparing particle-particle and particle-hole channels of the random phase approximation. *Physical Review B* **99**, 195149 (2019)