Recent Developments in the *Theory Department*

Director: Matthias Scheffler

1. General Remarks

Materials science is at present developing a new research paradigm: Parallel to (*i*) experimental or empirical studies, (*ii*) theoretical science (deriving laws and models) and (*iii*) computational materials science (performing simulations or computer experiments), we are now adding (*iv*) big-data-driven materials science. Here, one starts from a large body of experimental or theoretical data, and by means of "big-data analytics", which utilize methods from the signal-processing, image-recognition and machine-learning fields, one identifies patterns or structure in the data that would not be visible in a single study or even in several studies. In this way, we build maps that identify regions in the huge parameter space of compound materials, where different regions correspond to materials with different properties, such as, for example, electrical conductors, efficient catalysts for a certain chemical reaction, quantum-hall insulators, etc.

The challenge in building such a map is to find the appropriate descriptive parameters (called descriptors) that define these regions of interest. Importantly, inside these regions, interpolation between already known data enables an efficient identification of potentially interesting novel materials that have not been studied so far.

Up to now, scientific research is often characterized by a built-in bias deriving from the application of established criteria. For example, when looking for better catalytic materials, one studies substances from the same class of materials as those that have previously shown good catalytic properties. Obviously, this will hardly lead to revolutionary results. Moreover, the selection of materials that were studied in greater detail in the past is subjective, and this may imply that patterns in data just reflect the bias behind the selection, a phenomenon called the "Berkson paradox". In general, it appears that from the about 240,000 inorganic materials that have been synthesized so far[1] (many more are possible), the properties mentioned in the first paragraph have been studied for less than 1% of them.

Developing this fourth research paradigm, but with the necessary caution, is becoming more and more the focus of the work of the *Theory Department*. For example, it was the motivation behind a paper that was published in March 2015 (L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Phys. Rev. Lett. (2015): *Big data of materials science: Critical role of the descriptor*). That this was a rather new research field in 2014 is reflected in the initial editorial rejection of this paper, in which it was argued: "PRL has to decide how

[1] http://materials.springer.com/

narrow it should be in the topics that it covers". Interestingly, by September 1, 2017 this "narrow paper" had been cited 90 times!

Big-data-driven materials science was also the main idea behind our proposal for a "*Centre of Excellence (CoE)* for Computing Applications of the EC Horizon 2020 program: *The Novel Materials Discovery Laboratory (NOMAD)*, https://nomad-coe.eu/. The center, officially launched in November 2015, connects key scientists in many different institutions. At the FHI there are Claudia Draxl (Humboldt University Berlin, Max Planck Fellow at the FHI), Luca M. Ghiringhelli, Angelo Ziletti, and Christian Carbogno. More and more members are becoming involved (see section 4.2 below) also through international collaborators (e.g. Prof. Dr. Stefano Curtarolo, Duke University, now at FHI as AvH Bessel Awardee). A concise 3-minute introduction of the *NOMAD CoE* can be watched on YouTube:
https://youtu.be/yawM2ThVlGw.

While the last two years have witnessed substantial progress in the *NOMAD CoE* activities, other topics have also been highly successful. Four outstanding achievements of the *Departmen*t are briefly described in the following:

- From *LASSO (least absolute shrinkage and selection operator)* to *SISSO (sure independence screening plus sparsifying operator)*. Our initial work in data analytics started with the LASSO mathematical approach, originally developed for signal processing applications. LASSO can also be considered a forerunner of the field of *compressed sensing*. Unfortunately severe instabilities appear when the candidate descriptors are both highly correlated and in great number (e.g. more than 5,000 elements meant to describe 100 observables). These problems are solved by our SISSO approach, recently developed in the *Department*. SISSO does not suffer from the correlation problem. Instead, it can efficiently identify the best 2-3 descriptors out of the immensity of trillions candidates.
- *An efficient, non-perturbative approach for determining the thermal conductivity of solids* has been developed by Christian Carbogno. It was published in Physical Review Letters in April 2017. The preliminary work was already presented at the *Fachbeirat* meeting in 2015. The two most important aspects of this theory are: a) a first-principles formulation of the Green-Kubo method that allows the accurate assessment of the phonon-driven thermal conductivity of solid semiconductors and insulators by accounting for anharmonic effects to all orders via *ab initio* molecular dynamics calculations, and b) the achievement of size- and time convergence by a robust,

asymptotically exact extrapolation scheme. The technique was demonstrated for extremely-high and low heat conductors, namely Si and ZrO_2. A systematic high-throughput study of many materials has just been started.

- *Van der Waals physics on the nano and mesoscopic scales.* Alexandre Tkatchenko's group has developed new methods for studying van der Waals (vdW) interactions in nano and mesoscopic systems, also in the presence of static and dynamic fields. This has led to the discovery of "dipole waves". In addition, a multiscale method was developed to compute the fully retarded vdW energy for an arbitrary arrangement of microscopic and macroscopic subsystems, leading to a unified treatment of vdW and Casimir interactions. Taken together, these new developments achieve control of mesoscale structures by tailoring matter and light waves.
- Last but not least, we mention the further development of our all-electron electronic-structure code FHI-aims. The code was released in 2009 and its functionality has been impressively extended due to contributions from an active worldwide community of FHI-aims users/developers. In brief, the code has been characterized to be "as accurate as WIEN2k and as fast as or faster than VASP". The key developer and coordinator is Volker Blum, now Associate Professor at Duke University. The "standard" publication [V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler: *Ab initio molecular simulations with numeric atom-centered orbitals.* Comp. Phys. Comm. (2009)] has been cited more than 800 times to date! Several specific aspects are described below and, of course, on an extensive poster.

The following Section describes organizational developments. Section 3 addresses personnel and related matters. Subsequently, a more complete discussion of the work of the *Department* can be found in Sections 4 and 5, as well as in the special sections on the Otto Hahn Group of Mariana Rossi (established in October 2016) and on the remaining FHI activities of Alexandre Tkatchenko. He holds a professorship at the University of Luxembourg since November 2015. A more in-depth presentation is provided by the posters displayed in the *Department* and the poster abstracts.

The account given here and the poster presentations focus on the developments in the last two years (since September 2015).

2. Organizational

As of September 1, 2017, the FHI *Theory Department* is structured into 9 research groups. Nevertheless, most of the work in the *Department* involves more than one of these groups. The groups and their leaders are:

- *Advanced First-Principle Methods for Materials Science and Engineering*, headed by **Igor Ying Zhang**
- *Unifying Concepts in Catalysis*, headed by **Sergey V. Levchenko**
- *Heat and Charge Transport*, headed by **Christian Carbogno**
- *Big-Data Analytics for Materials Science*, headed by **Luca M. Ghiringhelli**
- *Data Science for Similarity and Crystal-Structure Prediction in Materials*, headed by **Matthias Scheffler** (deputy: **Angelo Ziletti**)
- *Machine Learning for Materials*, headed by **Matthias Rupp**
- *Otto Hahn Group***:** *Simulations from Ab Initio Approaches: Structure and Dynamics from Quantum Mechanics* headed by **Mariana Rossi**
- *Organic Functional Materials and Molecular Interactions* – FHI branch of **Alexandre Tkatchenko**
- *Ab Initio Biomolecular Simulations*, headed by **Carsten Baldauf**

In addition, we proudly host the *Max Planck Fellow Group* of **Claudia Draxl** and we support the *Max Planck Partner Group for Advanced Electronic-Structure Methods* of **Xinguo Ren** in Hefei/China. Emeritus **Alexander M. Bradshaw** 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.

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 6 additional institutions. It focuses on networking, backup, virtualization, and security services. The *GNZ* is headed by Gerd Schnapka.

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, since October 2014.

- 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, since February 2016.
- DFG – German Research Foundation, Cluster of Excellence 314: *Unifying Concepts in Catalysis (UniCat)*, chair: M. Driess; project leader: M. Scheffler (S. V. Levchenko), first funding period: November 2007 - October 2012, new funding period: November 2012 - October 2017.
- DFG – German Research Foundation, SFB 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.
- DFG – German Research Foundation, FOR 1543: *Shear-flow Regulation of Hemostasis (SHENC)*, spokesperson: R. Schneppenheim; project leaders: F. Gräter, C. Baldauf (C1), second funding period: September 2014 - August 2017.
- Einstein Foundation Berlin – Einstein Research Project *ETERNAL: Exploring Thermoelectric Properties of Novel Materials*, K.-R. Müller, C. Draxl, M. Scheffler, April 2013 - March 2016.
- ERC – European Research Council, Advanced Grant: *Big-Data Analytics for the Thermal and Electrical Conductivity of Materials from First Principles*; M. Scheffler – starting October 2017.
- EU – European Commission, E-Infrastructures: *NOMAD (Novel Materials Discovery) Center of Excellence*, coordinator: M. Scheffler; work package leaders: L. M. Ghiringhelli and C. Draxl, since November 2015.
- Leibniz-Gemeinschaft – Leibniz-WissenschaftsCampus (LWC): *Growth and fundamentals of oxides for electronic applications (GraFOx)*, spokesperson: H. Riechert; project leaders: M. Scheffler, S. Levchenko, and C. Sutton, 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; board members: B. Deveaud, A. Fontcuberta i Morral, A. Wodtke, M. Scheffler, since 2013. The *Theory Department* is currently involved in two projects: *Biomolecules on their Way to Solvation*, C. Baldauf, G. von Helden, T. R. Rizzo (EPFL) and *Peptides at Surfaces: Simulating Organic-Inorganic Interface Systems*, C. Baldauf, M. Ceriotti (EPFL), M. Rossi.

- Max Planck Society - Central Funds – *MP Fellow Group* of C. Draxl, since July 2014.
- Max Planck Society - Central Funds – *MP Partner Group for Advanced Electronic-Structure Methods* by X. Ren at the University of Science and Technology of China (USTC) in Hefei, since December 2015.
- Max Planck Society - Central Funds – *BigMax, MaxNet on Big-Data-Driven Materials Science*, coordinator: P. Benner, M. Scheffler, since March 2017.
- 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.
- Vetenskapsrådet – Swedish Research Council: *Catalysis on the Atomic Scale*, organizer: E. Lundgren; project leaders: S. V. Levchenko and M. Scheffler, 2012 - 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 board of directors of the *MM1P.de node* consists of Thomas Frauenheim (University of Bremen), Eberhard K. U. Gross (MPI Halle), Matthias Scheffler, and Björn Winkler (Goethe University Frankfurt). The node will continue to exist at least until 2018.

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.

3. Personnel and Related Matters

Since the last visit of the *Fachbeirat* in November 2015 the following noteworthy developments have taken place:

- In October 2016, **Mariana Rossi** returned to the *Theory Department* as independent group leader of the Otto Hahn Group on *Simulations from Ab Initio Approaches: Structure and Dynamics from Quantum Mechanics* (see Section 6).
- **Christian Carbogno** was promoted to group leader in November 2015. His group is called *Heat and Charge Transport.*
- The Max Planck Partner Group on *Advanced Electronic-Structure Methods* at the University of Science and Technology of China (USTC) led by **Xinguo Ren** began its work in September 2015. The objective of the group is to develop advanced and numerically tractable electronic-structure methods to satisfy the increasing needs for higher accuracy and reliability of first-principles calculations in physics and materials science. The inauguration workshop on the *Frontiers of Advanced Electronic-Structure Methods* took place in Hefei in June 2016. The lineup of speakers was outstanding: V. Blum, K. Burke, Z. Fang, X.G. Gong, X. Gonze, E.K.U. Gross, L.X. He, W.X. Li, X.Z. Li, W. Liu, H. Jiang, S. Meng, I. Mertig, J. P. Perdew, P. Rinke, A. Ruzsinszky, M. Scheffler, G.E. Scuseria, E.G. Wang, H. Wu, X. Xu, Y. Xu, J.L. Yang, W.T. Yang, I.Y. Zhang, Z.Y. Zhang. The meeting was judged as highly successful.
- **Carsten Baldauf** completed his habilitation at the FU Berlin in 2017.
- In November 2015, **Alexandre Tkatchenko** became the Chair of Theoretical Condensed Matter Physics at the University of Luxemburg.
- **Bryan Goldsmith** has accepted a position as Assistant Professor of Chemical Engineering at the University of Michigan. He will start there in September 2017.
- **John P. Perdew** and **Gustavo Scuseria** were bestowed with Humboldt Research Awards and spent some time in the *Department*. **Andrew Rappe** received this award recently and will visit us next year.
- **Stefano Curtarolo** received the Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation for his pioneering work in computational materials science, in particular the development of data-mining and high-throughput algorithms for computationally-assisted discovery of new materials. He is the initiator and director of the biggest repository for computational materials science data in the US, AFLOWlib. He is at present in Berlin and already involved in several joint research

projects. It is also planned to establish close links between the AFLOW and *NOMAD* projects and coordinate future developments.

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 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, now 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, alternating between Berlin and an international location. For example, this year we had more than 200 applicants but were only able to accept 70 participants (from 28 countries!). The workshop in Iran in 2016 was generously funded by the Alexander von Humboldt Foundation, and also the DAAD (German academic exchange program) contributed noticeably. The lack of official support from the Max-Planck Society was disappointing. The workshop 2016 in Esfahan and that in 2017 in Berlin-Adlershof (funded by Volkswagen Foundation) were great successes, and Carsten Baldauf, the main organizer, did a most impressive job. As always, the whole *Department* helped greatly.

Further activities in the last two years included:

- Workshop *Big Data of Materials Science - Critical Next Steps*, Lausanne, Switzerland, November 30 - December 4, 2015, organizers: L. M. Ghiringhelli, M. Scheffler, S. V. Levchenko
- Workshop *Towards a Common Format for Computational Materials Science Data*, Lausanne, Switzerland, January 25 - 27, 2016 (first part: discussion) and January 27 - February 4, 2016 (second part: implementation), organizers: D. Caliste (Alternative Energies and Atomic Energy Commission, France), C. Carbogno, L. M. Ghiringhelli, G. Huhs (Barcelona Supercomputing Center, Spain), M. Lueders (Daresbury Laboratory, UK), M. Oliveira (University of Liege, Belgium), R. Ramakrishnan, M. Scheffler

- *NOMAD*: first industrial meeting, Berlin, Germany, March 1 - 2, 2016, organizers: A. De Vita (King's College London, UK), A. Rubio (MPI Hamburg), M. Scheffler
- Symposium on *Frontiers of Electronic-Structure Theory: Focus on Topology and Transport* at the DPG Spring Meeting in Regensburg, Germany, March 7 - 11, 2016, organizers: I. Mertig (Martin Luther University Halle), G. Vignale (University of Missouri-Columbia, USA), M. Scheffler
- Workshop *Probing Potential-Energy Surfaces PPES IV 2016*, Zermatt, Switzerland, April 10 - 15, 2016, organizers: C. Baldauf, J. Pach, M. Scheffler
- Hands-on workshop and Humboldt-Kolleg: *Density-Functional Theory and Beyond - Basic Principles and Modern Insights*, Isfahan, Iran, May 2 - 13, 2016, organizers: H. Akbarzadeh (Isfahan University of Technology, Iran), C. Baldauf, S. Javad Hashemifar (Isfahan University of Technology, Iran), M. Scheffler
- Sino-German Workshop on *Biomolecular Simulations across Scales*, Shanghai, China, May 26 - 30, 2016, organizers: S. Xiao (Norwegian University of Science and Technology, Trondheim), X. Fei (East China Normal University) I. Y. Zhang, C. Baldauf
- The 1st USTC-FHI Workshop on the *Frontiers of Advanced Electronic-Structure Methods*, Hefei, China, June 14 - June 18, 2016, organizers: X. Ren (USTC, China), M. Scheffler
- FHI-aims Developers' and Users' Meeting on *Electronic-Structure Theory with Numeric Atom-Centered Basis Functions*, München, Germany, July 20 - 22, 2016, organizers: K. Reuter (TU München), C. Baldauf, V. Blum (Duke University, USA), M. Scheffler
- Workshop on *Big-Data-Driven Materials Science*, Ringberg, Germany, July 27 - 30, 2016, organizer: M. Scheffler
- *NOMAD* Analytics Hackathon, Berlin, Germany, January 18 - 20, 2017, organizers: L. M. Ghiringhelli, F. Mohamed, A. Ziletti, M. Scheffler
- *NOMAD*: second industrial meeting, Hamburg, Germany, February 6 - 7, 2017, organizers: A. De Vita (King's College London, UK), Angel Rubio (MPI Hamburg), M. Scheffler
- Symposium on *Frontiers of Electronic Structure Theory: New Concepts and Developments in Density Functional Theory and Beyond* at the DPG Spring Meeting in Dresden, Germany, March 19 - 24, 2017, organizers: J. P. Perdew (Temple University, USA), A. Rubio (MPI Hamburg), M. Scheffler
- *NOMAD* Analytics Hackathon, Berlin, Germany, July 3 - 5, 2017, organizers: L. M. Ghiringhelli, F. Mohamed, A. Ziletti, M. Scheffler

- *NSF/CECAM* School on *Computational Materials Science: From Basics to Applications*, Lausanne, Switzerland, July 17 - 27, 2017, organizers: K. Fichthorn (PennState, USA), J. Sofo (PennState, USA), M. Scheffler
- Hands-on workshop on *Density-Functional Theory and Beyond - Accuracy, Efficiency and Reproducibility in Computational Materials Science*, Berlin, Germany, July 31 - August 11, 2017, organizers: C. Baldauf, I. Y. Zhang, P. Pavone (HU Berlin), V. Blum (Duke University, USA), M. Scheffler
- Workshop on *Big-Data Driven Materials Science*, Lausanne, Switzerland, September 11 - 13, 2017, organizers: L. M. Ghiringhelli, M. Scheffler
- *NOMAD* SUMMER: A Hands-on Course on *Tools for Novel-Materials Discovery*, Berlin, Germany, September 25 - 29, 2017, C. Draxl (HU Berlin), M. Rampp (MPCDF, Garching), S. Rigamonti (HU Berlin), A. Ziletti, G. Huhs (HU Berlin; Barcelona Supercomputing Center, Spain)
- Workshop on *Quantum-Chemistry Methods for Materials Science*, Lausanne, Switzerland, November 8 - 10, 2017, organizers: I. Y. Zhang, M. Scheffler

Alex Bradshaw has served on several committees of the National Academy of Sciences (Leopoldina), National Academy of Engineering (acatech), and the Royal Society in the UK. He is also a member of the energy steering panel of the European Academies Science Advisory Council (EASAC).

4. New Concepts, Methods, and Techniques

Since the last meeting of the *Fachbeirat*, two wider themes have dominated the work of the *Department*, namely, the continuing development of the *FHI-aims* code and the *NOMAD CoE* project. These themes will be discussed first, followed by 9 additional topis. ***As mentioned above, the work by Mariana Rossi and Alexandre Tkatchenko will not be described here, but in separate chapters.***

4.1 *FHI-aims* Code

The FHI-aims code implements a unique set of features that distinguish it from other codes used in computational materials science. It combines (*i)* high-accuracy basis set (numerical atom-centered functions), suitable for efficient large-scale calculations, (*ii)* support for finite systems

(molecules and clusters) as well as extended, periodic systems on equal footing and (*iii) a* broad spectrum of methods from ground-state DFT via several quantum-chemistry methods to excited states. Recent advances cover new exchange-correlation functionals (e.g. self-consistent SCAN), van der Waals interactions, spin-orbit coupling, *GW*, BSE, and more. A density-functional perturbation theory framework addresses phonons, electronic friction, and spectroscopic observables. Important further enhancements target high-throughput screening and embedding into statistical-mechanics tools that are dedicated to structure prediction as well as to cluster expansion (including Wang-Landau and nested-sampling approaches for the evaluation of the density of configurational states), molecular dynamics, and thermal conductivity calculations. (This work was performed by many colleagues and typically coordinated by Volker Blum and Matthias Scheffler jointly; see the poster abstract for the full list of authors.)

4.2 The *NOMAD (Novel Materials Discovery) CoE (Center of Excellence)*

The *NOMAD CoE* maintains the largest *Repository* for input and output files of all important computational materials science codes worldwide (more than 32 codes). Using its open-access data *NOMAD* forms the basis for several *Big-Data Services* which in turn help to advance materials science and engineering.

Before 2015, it was very unusual in materials science to provide access to, and share fully research data (other disciplines were not any different!). This was also noted in a recent *nature editorial*[2], entitled: Empty rhetoric over data sharing slows science -- not-so-open data. Since end of 2014, this issue has been successfully addressed by the *NOMAD Repository*. Our work emphasizes that ***data is a crucial raw material of the 21st century***, from which valuable knowledge can be extracted using novel data-mining tools.[3]

In fact, the *NOMAD*[4] project has fostered a cultural shift in attitudes towards open data in the materials science community. The *NOMAD Repository* now provides open access to input and output files of computational materials science from all[4] major data collections in the field, together with those of many individual researchers and research groups. In this role, the

[2] Nature **546**, 327 (2017), https://www.nature.com/news/empty-rhetoric-over-data-sharing-slows-science-1.22133

[3] The *NOMAD* Center of Excellence https://nomad-coe.eu

[4] The biggest "single" repository is AFLOWlib (http://aflowlib.org) which provides numerous results as well as the input and output files of the underlying calculations. OQMD (http://oqmd.org) and Materials Project (https://materialsproject.org) offer numerous results but the underlying input and output files can be only obtained from *NOMAD*.

NOMAD Repository is also listed among the recommended repositories of *Nature Scientific Data*, where it is the only one for materials science. The *Repository* currently contains more than 40 million total-energy calculations, corresponding to billions of CPU-core hours used at high-performance computers worldwide.

NOMAD hosts the data for at least 10 years (for free), it offers DOIs (digital object identifiers) to make data citable, and it provides several services. For example, it supports all important electronic-structure codes used in the community (more than 32 codes). Moreover, it converts this very heterogeneous data and stores it in a ***code-independent archive***. We offer an *encyclopedia* to access all this data, numerous ***big-data analytics tools,*** and ***advanced graphics***. (The PIs of the *NOMAD CoE* are: Jose Maria Cela (Barcelona Supercomputing Centre (BSC)), Ciaran Clissmann (Pintail Ltd), Alessandro De Vita (King's College London), Claudia Draxl (Humboldt-Universität zu Berlin), Daan Frenkel (University of Cambridge), Stefan Heinzel (Max Planck Computing and Data Facility, Garching), Francesc Illas (University of Barcelona), Kimmo Koski (CSC – IT Center for Science, Espoo), Dieter Kranzlmüller (Leibniz Supercomputing Centre (LRZ), Garching), Risto Nieminen (Aaalto University), Angel Rubio (Max Planck Institute for the Structure of Dynamics of Matter, Hamburg), Kristian S. Thygesen (Technical University of Denmark, Lyngby), and Matthias Scheffler (coordinator).)

4.3 Equality of the DFT Band Gaps When Read from the Kohn-Sham Band Structure or Calculated by Total-Energy Differences

During a visit of John P. Perdew to the *Theory Department* we realized that an elementary misunderstanding persists in many publications with respect to the discussion of *band gaps of solids in generalized Kohn-Sham theory* (GKS). This was addressed by giving a simple proof of a theorem[5]: In GKS, the band gap of an extended system equals the fundamental gap for the approximate functional, if the GKS potential operator is continuous and the density change is delocalized when an electron or hole is added. This work explains how GKS band gaps from meta-generalized gradient approximations (meta-GGAs) and hybrid functionals can be more realistic than those from GGAs or even from the exact KS potential. The band edges in the GKS one-electron spectrum are also related to measurable energies. (J.P. Perdew, W. Yang,

[5] J.P. Perdew, W. Yang, K. Burke, Z. Yang, E.K.U. Gross, M. Scheffler, G.E. Scuseria, T.M. Henderson, I.Y. Zhang, A. Ruzsinszky, H. Peng, J. Sun, E. Trushin, and A. Görling, *Understanding Band Gaps of Solids in Generalized Kohn-Sham Theory*. Published in PNAS (March 6, 2017).

K. Burke, Z. Yang, E.K.U. Gross, G.E. Scuseria, T.M. Henderson, I.Y. Zhang, A. Ruzsinszky, H. Peng, J. Sun, E. Trushin, and A. Görling)

4.4 Advanced DFT and Quantum-Chemistry Methods for Molecules, Clusters, and Extended Solids

While the connection from the ground-state electron density to the total energy of the many-electron system is clear in principle[6], it is also plausible that any exchange-correlation (xc) functional, that can be written in a closed mathematical form, will be an approximation.

A significant amount of work in the *Theory Department* has therefore been devoted to implementing quantum-chemistry methods into the FHI-aims code for molecules and clusters as well as for extended, periodic systems. This concerns, in particular, coupled-cluster theory with single, double and perturbative triple excitations (CCSD(T)) as well as standard and reduced-scaling MP2. Obviously, CCSD(T) is restricted to closed shell systems or insulators, and it is expensive. However, it is very important for benchmark studies. (Tonghao Shen, Arvid C. Ihrig, and Igor Y. Zhang)

We have also developed and analyzed an improved orbital-dependent correlation functional that explicitly includes two-electron correlations. The functional is called sBGE2 (screened Bethe-Goldstone equation to second order). We then took the sBGE2 functional as a building block for an orbital-dependent xc functional, termed ZRPS (the authors' initials), which is a natural extension of the PBE0 hybrid functional. ZRPS delivers an overall mean absolute deviation of less than 70 meV in an extensive list of benchmarks. (Igor Y. Zhang, Patrick Rinke, and John P. Perdew)

4.5 Compressed-Sensing Approaches for Materials Science

In various (big-) data analytics studies we have developed and employed tools that build on compressed sensing (CS), a recent signal-processing concept that has revolutionized the recovery of signals by efficiently identifying their underlying sparsity patterns. One application direction aims to improve the quality and size of basis sets, as they crucially control the accuracy and CPU and memory requirements. Two complementary schemes are compared, where each

[6] The exact route from the density to the total energy goes as follows: The ground-state electron density determines the many-body hamiltonian. Subsequently, solving the many-electron Schrödinger equation gives the total energy. In this sense $n(\boldsymbol{r})$ also determines not just the ground state but also all excited states and any observable.

identifies the basis functions that carry the most relevant information of the system. (Ingo Roth, Adrian Steffens, Christian Krumnow, Jens Eisert, Niklas Menzel, and Luca M. Ghiringhelli)

A very different application of CS addresses materials properties and building "a map of materials". Our previous "LASSO plus ℓ_0 sparsification" approach is limited to modest pool sizes and descriptor candidates with low mutual correlation. The newly developed CS scheme replaces LASSO by Sure Independence Screening (SIS). It overcomes the above-mentioned problems. In most cases we find that the ℓ_0-norm regularization provides the best solution. The method has be carefully tested and applied to various timely systems, some will be discussed in Section 5 below. (Runhai Ouyang and Luca M. Ghiringhelli)

4.6 Phase Diagrams of Surfaces in Reactive Atmospheres: A Replica-Exchange Grand-Canonical Algorithm

Our *ab initio* atomistic thermodynamics approach has been very successful in predicting phase diagrams for surfaces and gas-phase clusters in contact with a realistic environment at realistic (T, p) conditions. The surface structures considered in such studies may be plausible, but, in general, the selection process is subjective, and it may miss the relevant structure. Indeed, in several situations, relevant to catalysis and corrosion, the surface composition and structure may be hard to guess. We have therefore developed a replica-exchange grand-canonical Monte-Carlo/molecular-dynamics algorithm that enables the unbiased, efficient calculation of these phase diagrams. We specifically target open systems (surfaces and free clusters in a reactive atmosphere) in the grand-canonical ensemble. Test-case applications demonstrate the potential of the approach. (Yuanyuan Zhou and Luca M. Ghiringhelli)

4.7 Efficient High-Throughput Calculations of Lattice Anharmonicities and Thermal Conductivity

Anharmonic effects of solids are typically treated perturbatively by solely accounting for the third-order term in a Taylor expansion of the potential-energy surface. To overcome this approximation, we are presently developing a high-throughput framework that is specifically targeted at the automatized calculation of fully anharmonic thermodynamic equilibrium (phonons, Grüneisen parameters, thermal expansion, specific heat, etc.) and non-equilibrium properties (thermal conductivity). Planning calculations for 1,000 specific systems, we began by looking at representative examples with increasing anharmonicity (e.g. Si, Mg_2Si, Ga_2O_3,

CuCl, ZrO_2) and calculating their properties both perturbatively and with our recently developed *ab initio* Green-Kubo molecular-dynamics approach. For strongly anharmonic systems such as CuCl and ZrO_2 the higher-order anharmonicity drastically alters the equilibrium properties, e.g., the phonon band structure. We are presently analyzing strategies to reduce the numerical cost by using higher-order, temperature-dependent Taylor expansions of the PES. (Hagen-Henrik Kowalski, Florian Knoop, Maja-Olivia Lenz, and Christian Carbogno)

4.8 Insightful Classification of Crystal Structures by Deep Learning

Computational methods that automatically extract knowledge from data are critical for enabling data-driven materials science. Identification of lattice symmetry is crucial for materials characterization and analytics. Current methods require a user-specified threshold, and are unable to detect "average symmetries" for defective structures. Here, we propose a machine learning-based approach to automatically classify structures by crystal symmetry. First, we represent crystals by a diffraction image, and then construct a deep-learning model for classification. Our approach is able to correctly classify a dataset comprising more than 80,000 structures, including heavily defective ones, belonging to the cubic crystal system. The internal operations of the neural network are unraveled through attentive response maps, demonstrating that it uses the same landmarks a materials scientist would use, although never explicitly instructed to do so. (Angelo Ziletti, Devinder Kumara, and Luca M. Ghiringhelli)

4.9 Unified Representation of Molecules and Crystals for Machine Learning

Machine learning can accurately interpolate between reference calculations given a sufficiently high volume of data. For this, an invariant, unique, differentiable, fast numerical representation that supports interpolation is essential. We developed and tested a many-body tensor representation (MBTR). Numerical results are presented for 7,000 organic molecules as well as 7 *ab initio* molecular dynamics trajectories (energy prediction errors are below 1 kcal/mol) and for 11,000 elpasolite crystals (energy prediction errors are below 3 meV/atom). Applicability is demonstrated for phase diagrams of Pt-group/transition-metal binary systems, where convex hulls are correctly identified for AgPt, IrZn, CdPt, AgPt, and IrPt. (Haoyan Huo and Matthias Rupp)

4.10 Finding Interpretable Local Patterns in Big-Data of Materials by "Subgroup Discovery"

Data-analytics tools applied to materials-science data often focus on the inference of a global prediction model for some property of interest. However, the underlying mechanism for some target property could differ even qualitatively for different materials. As a physical example, the transition metals of the periodic table are a subgroup, and the actinides and halogens are other subgroups.

Thus, identification of subgroups helps to understand similarities and differences between materials. In this spirit we developed subgroup discovery (SGD) for materials-science data by characterizing physical properties in terms of Boolean conditions. In contrast to global modelling algorithms, SGD finds interpretable descriptions of subpopulations in which, locally, the target property takes on an interesting distribution. The approach is demonstrated for several examples. (Bryan R. Goldsmith, Mario Boley, Jilles Vreeken, and Luca M. Ghiringhelli)

4.11 Error Bars, Extrapolation Schemes, and Test Sets

Significant work in the *Department* is devoted to determine numerical error bars associated to DFT calculations in diverse Kohn-Sham implementations, specifically, plane-wave pseudopotentials (and PAWs), LAPWs, and numerical, atom-centered basis sets. With all these implementations, extended convergence studies were performed for elemental and binary solids with respect to the numerical parameters up to full numerical convergence. On the basis of the error bars identified in this way, schemes are proposed to estimate reliably the numerical error bars associated with the incompleteness of the basis set and not fully converged k-meshes for arbitrary systems: https://analytics-toolkit.nomad-coe.eu. These schemes are validated for randomly chosen ternary systems obtained from the *NOMAD Repository*. (Christian Carbogno, Oliver Hofmann, Claudia Draxl, Kristian S. Thygesen, *et al.*)

We have also been strongly involved in benchmarking different codes for PBE calculations of perfect solids[7] and for *GW* calculations for 100 molecules[8]. This work has had significant

[7] K. Lejaeghere, *et al.*, Reproducibility in density functional theory calculations of solids. Science **351**, aad3000 (2016).

[8] M.J. van Setten, *et al.*, *GW 100: Benchmarking G_0W_0 for molecular systems.* J. Chem. Theory Comput. **11**, 5665 (2015).

impact, and is presently being continued for more advanced functionals as well as for solids with surfaces and defects. (Xinguo Ren, Christian Carbogno, Claudia Draxl, *et al.*)

The work for creating a *test set for materials science and engineering (MSE)* continues. This project considers cohesive energies, lattice constants, and bulk moduli for representative materials. Specifically, we consider (as the first step) first- and second-row elements and their binaries with cubic crystal structures and various bonding characters. So far we studied several different density functionals as well as MP2 and RPA. Over the next few months it is planned to include also CCSD(T). (Igor Y. Zhang, Andrew J. Logsdail, Norina A. Richter, Toktam Morshedloo, Xinguo Ren, Sergey V. Levchenko, and Luca M. Ghiringhelli)

5. Applications to Timely Questions in Materials Science

5.1 Classification of Materials Properties with SISSO

As a first application of our SISSO approach, we addressed the challenge to represent the vast amount of materials in terms of a map that assigns different regions to electrical conductors and semiconductors/insulators. Because of the size of available experimental data we restricted our study to A_xB_y binary compounds and certain structure prototypes. As primary features for our machine-learning strategy we used properties of neutral atoms, interatomic distances and the unit-cell volume of the considered materials. From these primary features more than a billion candidate descriptors were created; SISSO then determines the best two-dimensional descriptor. This defines the wanted two-dimensional map as well as the equation of the line dividing the two regions, which contains the key physical factors that determine the metallic vs. nonmetallic character of the material. The model achieves perfect (100%) classification of the training data. Interestingly, though the learning did not include information on the pressure dependence, the resulting equation also describes pressure-induced insulator → metal transitions. (Runhai Ouyang, Emre Ahmetcik, Stefano Curtarolo, and Luca M. Ghiringhelli)

5.2 Predicting and Understanding Quantum Spin-Hall Insulators with SISSO

Quantum spin-hall insulators (QSHIs) are two-dimensional topological insulators (TIs). They have attracted considerable interest in recent years for spintronic applications, since backscattering is forbidden in their edge states due to the topological band inversion. We have computed the topological Z_2 invariant of 220 two-dimensional, functionalized honeycomb-

lattice materials ABC_2 (*AB* are IV-IV, III-V, II-VI materials, and *C* is from group VII) using FHI-aims (DFT-PBE). Their classification into metals, insulators, and QSHIs can be described and understood using a two-dimensional descriptor identified via SISSO. Since this descriptor is physically meaningful and only depends on the properties of the isolated constituent atoms, it is highly transferable. This was demonstrated by using it to predict and subsequently validate over 50 additional 2D honeycomb QSHIs that consist of elements not included in our original machine-learning study. (Carlos Mera Acosta, Runhai Ouyang, Adalberto Fazzio, Luca M. Ghiringhelli, and Christian Carbogno)

5.3 Stabilities, Band Gaps, and Effective Masses of Polymorphs of Group-III Oxides

Despite the strong interest in transparent, conducting oxides, the fundamental properties of the various polymorphs of group-III sesquioxides ($Ga_xAl_yIn_zO_3$, with $x+y+z=2$) are still poorly understood. In fact, reliable experimental data are rare. We have therefore performed a high-throughput screening of these materials, focusing on the stability and electronic properties of various phases for the alloy as a function of composition and temperature. The results are obtained by sp-HSE06@PBEsol, i.e. single-point HSE06 calculations at the PBEsol-optimized ground-state and wavefunctions. This gives bandgap energies that are in good agreement with experimental values (13% error). To efficiently search (meta)stable configurations for many lattice types, we use a combination of DFT-based cluster-expansion models and fast stochastic optimization techniques. (Christopher Sutton and Luca M. Ghiringhelli)

5.4 Vibrational Effects on the Electronic Properties of Molecules and Solids

Our recent real-space implementation of density-functional perturbation theory (DFPT) was applied to compute the phonon-induced renormalization of the electronic band structure of 82 octet binaries (zincblende and rock salt structure) for both zero-point vibrations and finite temperatures. The SISSO approach is then used to identify physically meaningful descriptors for these effects, which allows the several exceptions observed with respect to established scaling rules, e.g., for the mass, to be rationalized.

Furthermore, we computed fully anharmonic Raman spectra for both molecular and crystalline polymorphs of paracetamol and aspirin. These spectra are obtained by evaluating the DFPT polarizability tensors along equilibrium *ab initio* molecular-dynamics trajectories. Besides being in quantitative agreement with experimental data, these results reveal strong anharmonic

effects on some localized (NH bending) and some delocalized (lattice phonon) modes. (Honghui Shang, Nathaniel Raimbault, Carlos M. Acosta, Patrick Rinke, Mariana Rossi, and Christian Carbogno)

5.5 Polarons in Different Polymorphs of TiO_2 and Ga_2O_3

An important factor limiting the electrical conductivity in solids is the interaction of the charge carriers with phonon modes and polaron formation. We studied oxides of Ti and Ga where it is expected that the electron-phonon coupling is large and polarons are strongly localized.

In order to address uncertainties caused by the exchange-correlation approximation, we used the hybrid HSE functional and took into account the whole range $0 \leq \alpha \leq 1$ of exact exchange. Furthermore, in order to obtain the polaron properties in the dilute limit, we consider elastic long-range distortions and finite-size corrections in a supercell approach. The approach is applied to rutile and anatase TiO_2, and monoclinic β- and orthorhombic ε-Ga_2O_3. Unexpectedly, we find that in rutile TiO_2 only small, i.e., strongly localized, *electron* polarons are stable, whereas only small *hole* polarons are found in anatase. On the other hand, for Ga_2O_3 we predict the existence of small hole polarons for both polymorphs. Their binding energies are, however, significantly different. (Sebastian Kokott, Sergey V. Levchenko, and Patrick Rinke)

5.6 Adsorption and Disorder-Disorder Phase Transitions of H at ZnO (10-10) at Realistic (*T*, *p*) conditions

Following earlier work in the *Department*[9], we combined a DFT-based cluster-expansion model with the Wang-Landau algorithm to describe the thermodynamics of atomic H adsorbed at ZnO (10-10), including the configurational entropy of the adsorbate. Our study reveals that at coverages below 6%, H atoms adsorb exclusively on surface O atoms. At higher coverages, H adsorbs also on Zn. Neighboring O-H/Zn-H pairs form chains along surface -O-Zn- rows, always ending by O-H. The interaction between chains is dominated by the repulsion between the O-H ends, and ordering of the chains is destroyed for $T > 60$ K. Interestingly, there are different disordered phases of adsorbed H that are characterized by their quantitative O-H

[9] M. Borg, C. Stampfl, A. Mikkelsen, J. Gustafson, E. Lundgren, M. Scheffler, and J.N. Andersen, "*Density of configurational states from first-principles: The phase diagram of Al-Na surface alloys*". ChemPhysChem **6**, 1923 (2005).

excess over Zn-H. The calculated work function (as function of coverage) is in good agreement with recent experimental results by Julia Stähler *et al*. (PC Department). (Maria E. Stournara, Sergey V. Levchenko, Santiago Rigamonti, Maria Troppenz, Oliver T. Hofmann, Patrick Rinke, and Claudia Draxl)

5.7 Catalytic Conversion of CO_2 at Metal-Oxide and Metal-Carbide Surfaces

The catalytic conversion of the greenhouse gas CO_2 to fuels or to other useful chemicals is highly desired, and thus there is a need for finding new, more efficient catalysts. The O-C-O bending angle of an adsorbed CO_2 is an indicator of the activation of the molecule, while the adsorption energy is an indicator of the stability of the system with respect to carbonate formation. Using DFT with a carefully validated xc approximation, we calculated the properties of the molecule adsorbed on various cuts of about 14 binary carbides and 72 binary and ternary oxides (over 200 different surfaces overall). SISSO is then used to identify physically meaningful descriptors for the CO_2 activation. (Aliaksei Mazheika, Yanggang Wang, Runhai Ouyang, Rosendo Valero, Christian Kunkel, Helena Muñoz Galan, Luca M. Ghiringhelli, Francesc Illas, and Sergey V. Levchenko)

In a co-operation with colleagues from the AC Department we have also studied Ni_xMg_yO for dry reforming of CH_4 and for CO_2 hydrogenation. Poly-crystalline samples with low concentrations (3-10%) of Ni in MgO were successfully synthesized for the first time,. The homogeneous distribution of Ni in MgO allowed a clear comparison of theory and experiment. The calculations reveal that Ni doping and CO_2 adsorption stabilize low-coordinated sites (corners and steps) and promote the formation of stepped (110) and reconstructed (111) MgO surfaces. The experimental differential heat of adsorption values of these samples are perfectly explained by the occupation of CO_2 vacancies in the surface carbonate at steps, as well as adsorption at the reconstructed (111) surface with Ni_{Mg} defects. (Aliaksei Mazheika, Marie M. Millet, Sabine Wrabetz, Andrey Tarasov, Elias Frei, Sergey V. Levchenko, and Robert Schlögl)

5.8 *Ab Initio* Biomolecular Simulations

The *Ab Initio Biomolecular Simulations* group aims at laying a solid foundation for the description of the energetics and dynamics of biomolecules by simulation. They develop structure search techniques and test across-the-scale energy functions. The group works closely

with Mariana Rossi. One topic concerns hybrid inorganic-organic materials. Modeling their electronic structure is challenging, since xc approximations are often not working reliably. We have studied organic systems ranging from rigid aromatic molecules to flexible biomolecules adsorbed on semiconductor [H-Si(111)] and metallic [Cu(111)] surfaces. The peptide $HisProPheH^+$ and the amino acid $ArgH^+$, for example, are known from experiment to form distinct structures on metal surfaces like Cu(111). To predict such structures, our first-principles structure search package *Fafoom* was extended to include the position and orientation of molecules with respect to the surface. (Carsten Baldauf, Volker Blum, Svenja Janke, Sergey V. Levchenko, Dmitrii Maksimov, Haiyuan Wang, and Mariana Rossi)

Furthermore, peptide-cation systems were studied to understand how metal cations shape the structure of proteins. Specifically, we investigated a model system for helix formation in the gas phase by a combination of first-principles molecular simulations with cold conformer-selective vibrational spectroscopy experiments in the gas phase that allow for detailed structure elucidation. (Markus Schneider, Chiara Masellis, Tom Rizzo, and Carsten Baldauf)

6. Otto Hahn Group Simulations from *Ab Initio* Approaches: Structure and Dynamics from Quantum Mechanics — Head: Mariana Rossi

Guest scientists, staying for at least six months, paid by FHI:

Nathaniel Raimbault	*Otto Hahn*
Haiyuan Wang	*Otto Hahn*

Guest scientists, staying for at least six months, paid from external funds:

Svenja Janke	*DFG (together with M. Scheffler)*

Graduate students: 2 (*1 from FHI, 1 from MP-EPFL center*)

General Remarks

The Otto Hahn Group led by Mariana Rossi was established in October 2016. In the Max Planck Society, a very small number of recipients of the Otto Hahn Medal are selected for the Otto Hahn Award. This distinction gives them the opportunity to choose a Max Planck Institute of their choice in order to head a small group, independently funded by the MPS, for a period of 3 years (plus 2, after evaluation), after doing a post-doc abroad. In Mariana Rossi's group H-bonded systems composed of biomolecules, organic molecules, and their interfaces with inorganic systems are studied. In particular, the focus lies on the influence of temperature, anharmonic effects, and nuclear quantum phenomena, studied using first-principles quantum mechanical simulations. The two members of the group funded directly from the Otto Hahn Award started in November 2016, and, between February and March 2017, three more members have joined the group. There are active collaborations with experimental groups in the Fritz Haber Institute and other Max Planck Institutes, as well as co-operations with other group leaders in the *Theory Department* and theoretical scientists in other institutions, such as Prof. Michele Ceriotti (EPFL), Prof. Jeremy Richardson (ETH-Zürich), Prof. Davide Donadio (UC Davis), and Prof. Volker Blum (Duke University).

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. Most of our research involves bio-organic materials and bio-organic/inorganic interfaces. In general, matter is not static, and for non-covalently bonded systems the fluctuations induced by temperature can completely change the structure, the thermodynamic equilibrium, and also the

properties of the electronic structure. If the materials involved are organic or biological, the ubiquitous presence of hydrogen atoms makes approximations like the treatment of nuclei as classical particles unsuitable. The interplay of temperature and nuclear quantum effects can give rise to several important effects such as the displacement of thermodynamic equilibria in reactions and the changing of the pH and pKa of substances. Moreover it is responsible for band gap renormalization in water and other materials; it can induce changes in charge transfer behavior, and also induce work function changes. In particular, we are focusing on bio-organic/inorganic interfaces, where new technological developments can stem from the combination of the robust properties of inorganic materials and the flexibility and modularity of bio-organic components. We develop structure searches tailored for flexible molecules, acceleration techniques and new methods to treat dynamical properties including nuclear quantum effects, and frameworks to calculate anharmonic vibrational properties of materials. Our group can study these effects with speed and accuracy, thus bringing the electronic-structure calculations closer to actual experiments.

Research Highlights and Recent Developments

Hydrogen Transfer Reactions on Surfaces Down the Deep Tunneling Regime

Surface-mediated hydrogen transfer reactions can be particularly impacted by nuclear quantum effects, but these can either be dominated by zero-point energy (ZPE) or tunneling. For example, we have studied the impact of nuclear quantum effects (NQE) on the dissociation of water wires on stepped Pt(221) surfaces at finite temperatures. Some interesting findings of our study are that (i) several competing minima co-exist at finite temperatures, (ii) temperature and anharmonicities make NQE contributions to the free energies 20% larger per dissociated dimer, (iii) the magnitude of NQE contributions is comparable to the free energy differences themselves at finite temperatures, and (iv) temperature and NQE impact particularly work-function changes on these stepped surfaces, changing their average value by 0.2 to 0.4 eV. We find that the dissociation process is dominated by ZPE, which can enhance the rate of dissociation by three orders of magnitude. In collaboration with Takashi Kumagai in the PC department, we are also studying hydrogen transfer reactions in molecular switches like porphycenes adsorbed on metallic surfaces. In this case the reactions are dominated by tunneling at low temperatures, and the surface changes the potential energy landscape of the reaction. In order to perform our simulations we have proposed and implemented in the i-PI code an acceleration scheme that makes a spatial partition in the system where a ring polymer contraction can be performed, resulting in considerable computational savings for this class of

systems. In collaboration with Jeremy Richardson (ETH Zürich), we have also implemented the ring polymer instanton method in i-PI, thus giving access to calculations of reaction rates in the deep tunneling regime and tunneling splittings, which, through the interface with FHI-aims, can now be used also for *ab initio* calculations on surfaces. (Yair Litman)

Anharmonic Vibrational Spectra: IR, Raman, and Sum Frequency Generation

Vibrational Raman spectroscopy is important for the structural characterization of many systems. Through the optimization and extension of the DFPT framework for electric field responses in the FHI-aims code, we could obtain polarizability tensors for isolated and periodic systems. Joining this framework with molecular dynamics gives access to different vibrational spectra where temperature and anharmonicities are automatically accounted for. We take as an example molecular crystals, where we calculate anharmonic Raman spectra from the Fourier transform of the polarizability autocorrelation functions. We also treat lattice expansion in a quasi-harmonic framework that gives direct access to the anisotropic expansion of different lattice components. By studying paradigmatic molecular crystals like aspirin and paracetamol, and performing simulations including many body dispersion interactions, we find an extremely good agreement with published data for the far infrared region, which is known to be quite anharmonic and very sensitive to the crystal structure. Our implementation can also give access to vibrational sum frequency generation spectra, which is sensitive to interfaces. It can also be coupled to recently developed methods like thermostatted ring polymer molecular dynamics to account for nuclear quantum effects in vibrational spectra. We are currently exploring ways, based on kernel ridge regression techniques, to make the calculation of the polarizability tensors cheaper in these simulations. (Nathaniel Raimbault)

Organic molecules adsorbed on H-Si(111)

Hybrid inorganic-organic systems (HIOS) are promising materials especially for optoelectronic devices such as light-emitting diodes, solar cells, or sensors. When the interface between the organic layer and the semiconductor is formed, measurable work function changes take place that one would ideally like to control. In our group we study highly efficient dopants (F4-TCNQ and F6-TCNNQ) as well as renowned singlet fission materials (tetracene and pentacene) adsorbed on hydrogen-passivated Si(111) [H-Si(111)]. A key difficulty in describing quantitatively the electronic structure of organic molecules at surfaces is to determine the adsorption geometry of the organic layer at the surface. To tackle this problem, we develop structure search strategies that allow us to determine the adsorption geometry for different

coverages on H-Si(111). Through the calculation of core-level shifts, consideration of the space-charge layer, and comparison to X-ray photoelectron spectroscopy data from the group of Prof. Norbert Koch (HU-Berlin), we are able to disentangle different contributions to the work function change (band bending and surface dipole) and to provide physical insight into this process. (Svenja Janke, Haiyuan Wang, Sergey V. Levchenko, Volker Blum, and Matthias Scheffler)

Conformational Search of Biomolecules on Inorganic Surfaces

For the structure search of flexible bio-organic molecules in contact with inorganic surfaces, where properly accounting for the internal degrees of freedom is a challenge, we have extended the *Fafoom* code (a genetic algorithm code that works on internal degrees of freedom) to include orientations and translations of flexible molecules with respect to one another and to fixed frames, such as surfaces. This work is performed jointly with the group of Carsten Baldauf in the *Theory Department*. We obtained results from density-functional theory for the case of the arginine amino-acid in different protonation states. From our data we can show how the conformational space of the molecule is changed upon adsorption and can consider how to obtain simpler energy functions and entropy estimations for an initial structural pre-screening of these systems. When $ArgH^+$ adsorbs on metallic surfaces, we find that the positive charge is efficiently screened by concentration of electrons from the surface close to the charged group and the conformational space is expanded with respect to the gas-phase picture. This efficient screening of the charge on a peptide is likely to play a role in the 2D self-assembly observed for these same systems in the group of Klaus Kern and Stephan Rauschenbach (MPI FKF Stuttgart), with whom we actively collaborate in this project. (Dmitrii Maksimov and Carsten Baldauf)

External funding

- CSCS - Swiss National Supercomputing Centre, 450’000 node hours over 1 year on the Cray XC30, Piz Daint, since October 1st 2016.
- MP-EPFL Center for Molecular Nanoscience and Technology, Project *Peptides at Surfaces: Simulating Organic-Inorganic Interface Systems*, 1 Ph.D. student.

7. ERC Group on Organic Functional Materials and Intermolecular Interactions — Head: Alexandre Tkatchenko

Guest scientists, staying for at least six months, paid from external funds:

Mausumi Chattopadhyaya	*until 05/2017*	*AvH Fellow*
Fairoja C. Kabeer	*until 07/2017*	*DFG*
Andrii Kleshchonok		*DFG*
Gionni Marchetti	*until 08/2016*	*DFG*
Majid Mortazavi		*DFG*
Huziel E. Sauceda	*until 12/2017*	*DFG*
Limin Zheng	*until 03/2016*	*China Scholarship Council*

Graduate students: 3 *(2 from DFG, 1 FHI)*

General Remarks

The ERC (European Research Council) Research Group *Organic Functional Materials and Intermolecular Interactions*, led by Alexandre Tkatchenko, was established in May 2010 and is hosted by the *Theory Department*. In November 2015, A. Tkatchenko became a full professor at the University of Luxembourg, where he holds a chair in theoretical chemical physics and also has an ERC Consolidator grant. There is a strong connection between his group in Luxembourg and FHI; part of his group is hosted by the FHI. To strengthen the collaboration between the two groups, A. Tkatchenko visits Berlin once every two months. In addition, the FHI members of the group visit Luxembourg at least twice a year.

The ERC Group has recently been responsible for fundamental developments in efficient first-principles methods for the accurate modeling of non-covalent interactions in molecules, solids, and interfaces. The methods developed are applied to a wide range of fundamental complex materials, with the aim of understanding their qualitative and quantitative structural, cohesive, and electronic properties at the atomic scale. The group at FHI currently consists of three postdoctoral researchers, and two PhD students, together with the Principal Investigator (PI). Since the last meeting of the *Fachbeirat*, several postdoctoral researchers and PhD students have left for academic and industry positions elsewhere. Since February 2017, Anthony Reilly is an assistant professor in computational chemistry in the School of Chemical Sciences at Dublin City University. In total, three PhD students have defended during the last two years (Vivekanand V. Gobre, Victor G. Ruiz, and Nicola Ferri). All of them are pursuing academic careers: Vivekanand Gobre has become a research associate at the Interdisciplinary School of Scientific Computing in Pune, India; Victor Ruiz is a postdoctoral fellow at the Helmholtz-

Zentrum in Berlin; and Nicola Ferri is considering several postdoctoral opportunities after his contract at FHI expires. These successful career moves of former and current ERC Group members highlight the high quality of the research and scientific training in the group.

Scientific Scope

Non-covalent van der Waals (vdW) interactions are ubiquitous in nature. They make possible the existence of molecular liquids and solids, control protein-protein and drug-protein binding inside our cells, and give geckos the ability to "defy gravity" by climbing walls and ceilings. An accurate description of vdW interactions is extremely challenging, since the vdW dispersion energy arises from collective motion of electrons and must be described by many-electron quantum mechanical methods. The lack of accurate and efficient methods for capturing vdW interactions in large and complex systems hinders truly quantitative predictions of properties and functions of technologically relevant materials. The ERC Group has successfully addressed this challenging age-old problem by developing a hierarchy of efficient and advanced quantum-mechanical methods with accuracy and capacity to predict new phenomena in complex materials of fundamental and technological interest. These developments are based on a combination of techniques from many-body physics, statistical mechanics, chemistry, and computer science. The methods developed by the group are now widely used worldwide to enable molecular simulations with predictive power for large and complex materials, chemicals, and biological systems.

In addition to non-covalent interactions, the group has recently expanded its interests towards developing efficient methods for modeling the coupling between electronic correlations and nuclear dynamics, and employing data-driven machine learning techniques to construct accurate representations of molecular properties in chemical space and next-generation interatomic potentials.

Research Highlights

New van der Waals Physics at the Nano and Mesoscopic Scale

With a strong focus on the accurate description of interactions that define the functionality of materials at the nano and mesoscopic scales, it becomes crucial to develop increasingly reliable methods for the description of vdW interactions in materials of increasing complexity and size (systems under confinement, subject to heterogeneous electromagnetic and optical fields, as well as nano/bio heterostructures and engineering materials). In addition, new physical

phenomena often emerge when coupling external fields to electronic fluctuations inherent to a material of interest. In this context, several major advances have been accomplished by the ERC Group during the last two years, including:

(1) We have demonstrated that in large molecular complexes and low-dimensional nanostructures, vdW correlations are more naturally described in terms of coupling between fluctuating dipole waves. This leads to several nontrivial effects in vdW interactions that cannot be explained by dispersion corrections and non-local functionals based on electron density, but that are correctly captured by our developed many-body dispersion (MBD) model. In linear chain structures and layered 2D materials, the interaction energy shows complex power laws that can be systematically tuned from insulating to metallic behavior. The associated wave fluctuations are largely delocalized over the whole system, continuously changing the interaction power law as a function of the separation between nanostructures. For molecules interacting through nanostructures, we observe a regime in which their interaction energy increases, instead of decreasing as naively expected from dielectric screening arguments. Further application to carbon-based supramolecular complexes demonstrates that omitting wave-like vdW fluctuations leads to differences in association constants of up to five orders of magnitude where there should be degeneracy according to high-level quantum Monte Carlo calculations. Taken together, our findings indicate that the wavelike nature of vdW interactions provides a hitherto unexplored avenue that could be used for tailoring the assembly of complex polarizable systems at the nanoscale (*Science* 2016; *Chem. Mater.* 2016; *Nature Commun.* 2017; *Chem. Rev.* 2017; *J. Chem. Phys.* 2017).

(2) In collaboration with P. Venkataram and A. W. Rodriguez (Princeton University), we have developed a method that unifies calculation of van der Waals interactions with Maxwell-based solvers for Casimir interactions. Specifically, we developed an approach for computing long-range van der Waals (vdW) interactions between complex molecular systems and arbitrarily shaped macroscopic bodies, melding atomistic treatments of electronic fluctuations based on density functional theory in the former with continuum descriptions of strongly shape dependent electromagnetic fields in the latter, thus capturing many-body and multiple scattering effects to all orders. Such a theory is especially important when considering vdW interactions at mesoscopic scales, i.e., between molecules and structured surfaces with features on the scale of molecular sizes, in which case the finite sizes, complex shapes, and resulting non-local electronic excitations of molecules are strongly influenced by electromagnetic retardation and

wave effects that depend crucially on the shapes of surrounding macroscopic bodies. We show that these effects together can modify vdW interaction energies and forces, as well as molecular shapes deformed by vdW interactions, by orders of magnitude compared to previous treatments based on Casimir-Polder, non-retarded, or pairwise approximations. The latter are valid only at macroscopically large or atomic-scale separations or in dilute insulating media (*Phys. Rev. Lett.* 2017; *Rev. Mod. Phys.* 2016).

Quantitative Prediction of Molecular Crystal Polymorphs

Molecular crystals are versatile materials that find applications as pharmaceuticals, explosives, organic semiconductors, in solid-state reactions, and plastic materials. A molecular crystal composed of the same moieties can have several different crystal-packing motifs or polymorphs. This has far-reaching consequences, since polymorphs can exhibit completely different physical and chemical properties. Therefore, accurately predicting the relative energies of all thermally accessible polymorphs of a given molecule is a difficult fundamental problem, which also plays a crucial role in the pharmaceutical industry.

In order to demonstrate the capabilities of our methods in the study of polymorphism, our group participated in the 6th blind test of organic crystal structure prediction, organized by the Cambridge Crystallographic Data Centre (CCDC). Remarkably, our submission achieved the best energetic ranking for eight molecular crystal polymorphs among the 25 research groups and companies who participated in the blind test (*Acta Crys B.* 2016).

Following on this success, we have established collaborations with several leading research groups and pharmaceutical companies to further develop and apply our methods to increasingly challenging polymorphic systems (*Chem. Sci.* 2017; *WIRES: Comput. Mol. Sci.* 2017; *Phys. Rev. Lett.* 2015, 2017).

Towards Machine Learning of Chemical Space and Interatomic Potentials

Learning from data has led to paradigm shifts in a multitude of disciplines, including web, text and image search, speech recognition, as well as bioinformatics. Can machine learning enable similar breakthroughs in understanding quantum chemistry? The main challenge is the disproportionately large size of chemical space, estimated to contain 10^{60} molecules even when only counting small organic drug-like candidates.

With the aim of applying machine learning to quantum chemical space, we have developed an efficient deep learning approach that enables spatially and chemically resolved insights into

quantum-mechanical observables of molecular systems. This is done by unifying concepts from many-body Hamiltonians with purpose-designed deep tensor neural networks, which in turn leads to size-extensive and uniformly accurate (1 kcal/mol) predictions in compositional and configurational chemical space for molecules of intermediate size. As an example of chemical relevance, the machine learning model reveals a classification of aromatic rings with respect to their stability. Further applications for predicting atomic energies and local chemical potentials in molecules, reliable isomer energies, and molecules with peculiar electronic structure demonstrate the potential of machine learning for revealing insights into complex quantum-chemical systems (*Nature Commun.* 2017).

For constructing next-generation machine-learned force fields, we have used conservation of energy to develop an efficient gradient-domain machine learning (GDML) approach to construct accurate molecular force fields using a restricted number of samples from *ab initio* molecular dynamics (aiMD) trajectories. The GDML implementation is able to reproduce global potential energy surfaces of intermediate-sized molecules with an accuracy of 0.3 kcal/mol for energies and 1 kcal/mol/Å for atomic forces using only 1000 conformational geometries for training. We demonstrate this accuracy for aiMD trajectories of molecules, including benzene, toluene, naphthalene, ethanol, uracil, and aspirin. The challenge of constructing conservative force fields is accomplished in our work by learning in a Hilbert space of vector-valued functions that obey the law of energy conservation. The GDML approach enables quantitative molecular dynamics simulations for molecules at a fraction of cost of explicit aiMD calculations, thereby allowing the construction of efficient force fields with the accuracy and transferability of high-level *ab initio* methods (*Science Adv.* 2017).

Third-Party Funded Projects

- ERC – European Research Council, Starting Grant: Van der Waals Interactions in Complex Materials; A. Tkatchenko – Sept. 2011 – Aug. 2016.
- DFG – German Research Foundation, Basis module, project: Exploring Chemical Compound Space with Machine Learning; A. Tkatchenko, K.-R. Müller – since 2014.
- DOE – US Department of Energy, EFRC (Energy Frontier Research Centers): Integrated Mesoscale Architectures for Sustainable Catalysis (IMASC), coordinator: C. Friend; A. Tkatchenko – August 2014 – Oct. 2016.
- DFG – German Research Foundation, SFB 951, Hybrid Inorganic/Organic Systems (HIOS), spokesperson: Norbert Koch; A. Tkatchenko, M. Scheffler – since 2015.