

## Theory Department

### Poster List

Recent work of the Theory Department is displayed on 24 posters in building T.

The superscript <sup>OHG</sup> marks work by or done in collaboration with Mariana Rossi and her Otto Hahn group. And the superscript <sup>ERC</sup> marks the work of A. Tkatchenko's ERC group.

*Poster Location*      *Poster Title and Authors*  
*Index*

#### **New Concepts, Methods, and Techniques**

##### **TH 1      The Novel Materials Discovery Laboratory (NOMAD)**

Luca M. Ghiringhelli, Fawzi R. Mohamed, Angelo Ziletti, Ankit Kariryaa, Benjamin Regler, Alfonso Sastre, Christian Carbogno, Sergey V. Levchenko, Georg Huhs, Rubén Jesús García Hernández, Atte Sillanpää, Stefan Heinzl, Kimmo Koski, Dieter Kranzlmüller, Jose Maria Cela, Alessandro De Vita, Daan Frenkel, Francesc Illas, Risto Nieminen, Angel Rubio, Kristian Sommer Thygesen, Claudia Draxl, and Matthias Scheffler

##### **TH 2      The FHI-aims Project**

Volker Blum, Danilo S. Brambila, Björn Bieniek, Christian Carbogno, Luca M. Ghiringhelli, Dorothea Golze, Jan Hermann, Oliver T. Hofmann, Johannes Hoja, William P. Huhn, Yosuke Kanai, Matthias Kick, Jan Kloppenburg, Raul Laasner, Björn Lange, Marvin Lechner, Chi Liu, Andrew Logsdail, Andreas Marek, Florian Merz, Reinhard Maurer, Lydia Nemeč, Harald Oberhofer, Nathaniel Rimbault, Xinguo Ren, Karsten Reuter, Stefan Ringe, Patrick Rinke, Mariana Rossi, Christoph Schober, Honghui Shang, Markus Sinstein, Christopher Sutton, Alexandre Tkatchenko, Alvaro Vazquez-Mayagoitia, Mina Yoon, Victor W. Yu, Igor Y. Zhang, Tong Zhu, and Matthias Scheffler

##### **TH 3      Quantum-Chemistry Methods for Molecular and Periodic Systems in the Numeric Atom-Centered Orbital Framework**

Arvid C. Ihrig, Tong H. Shen, Matthias Scheffler, and Igor Y. Zhang

##### **TH 4      Towards Efficient Orbital-Dependent Density Functionals for Weak and Strong Correlation**

Igor Y. Zhang, Patrick Rinke, John P. Perdew, and Matthias Scheffler

- TH 5 Novel High-Entropy Carbides Discovered by Synthesizability Descriptors**  
Pranab Sarker, Tyler Harrington, Cormac Toher, Kenneth S. Vecchio, Jon-Paul Maria, Donald W. Brenner, and Stefano Curtarolo
- TH 6 Compressed-Sensing Approaches to Kohn-Sham Density-Functional Theory and Beyond**  
Adrian Steffens, Niklas Menzel, Ingo Roth, Christian Krumnow, Luca M. Ghiringhelli, Jens Eisert, and Matthias Scheffler
- TH 7<sup>ERC</sup> Bridging Scales from van der Waals to Casimir Interactions**  
Jan Hermann, Andrii Kleshchonok, Majid Mortazavi, and Alexandre Tkatchenko
- TH 8 Strongly Anharmonic Nuclear Dynamics in Solids: Accurate Computations and Rapid Estimates**  
Hagen-Henrik Kowalski, Florian Knoop, Maja-Olivia Lenz, Matthias Scheffler, and Christian Carbogno
- TH 9 Replica-Exchange Grand-Canonical Algorithm for Determining Phase Diagrams of Surfaces in Reactive Atmospheres**  
Yuanyuan Zhou, Matthias Scheffler, and Luca M. Ghiringhelli

### **Bulk Materials**

- TH 10<sup>OHG</sup> Adiabatic Vibronic Coupling in Molecules and Solids: Physical Insight from Density-Functional Perturbation Theory and Machine Learning**  
Honghui Shang, Nathaniel Raimbault, Carlos Mera Acosta, Patrick Rinke, Matthias Scheffler, Mariana Rossi, and Christian Carbogno
- TH 11 Massively Parallel Compressed-Sensing Optimizer for Identifying Materials-Property Descriptors: Application to the Classification of Materials as Metals**  
Runhai Ouyang, Emre Ahmetcik, Stefano Curtarolo, Matthias Scheffler, and Luca M. Ghiringhelli
- TH 12 Structure Dependence of Small Polarons in Transparent Oxides: *Ab Initio* Study of Polymorphs of TiO<sub>2</sub> and Ga<sub>2</sub>O<sub>3</sub>**  
Sebastian Kokott, Sergey V. Levchenko, Patrick Rinke, and Matthias Scheffler
- TH 13 Predicting and Understanding Quantum Spin-Hall Insulators with the Help of Machine Learning**  
Carlos Mera Acosta, Runhai Ouyang, Adalberto Fazio, Matthias Scheffler, Luca M. Ghiringhelli, and Christian Carbogno

- TH 14**     **The Face of Crystals: Insightful Classification Using Deep Learning**  
Angelo Ziletti, Devinder Kumar, Matthias Scheffler, and  
Luca M. Ghiringhelli
- TH 15**     **High-Throughput Screening of Transparent Conducting Oxides**  
Christopher Sutton, Luca M. Ghiringhelli, and Matthias Scheffler
- TH 16**     **Unified Representation of Molecules and Crystals for Machine  
Learning**  
Haoyan Huo and Matthias Rupp
- TH 17**     **Subgroup Discovery for Finding Interpretable Local Patterns in Data  
from Materials-Science**  
Bryan R. Goldsmith, Mario Boley, Jilles Vreeken, Matthias Scheffler, and  
Luca M. Ghiringhelli

**Surfaces, Adsorption, and Heterogeneous Catalysis**

- TH 18**     **Higher-Order Phase Transitions of a Hydrogenated ZnO (10-10)  
Surface**  
Maria E. Stournara, Sergey V. Levchenko, Santiago Rigamonti,  
Maria Troppenz, Oliver T. Hofmann, Patrick Rinke, Claudia Draxl, and  
Matthias Scheffler
- TH 19<sup>OHG</sup>**   **Impact of Nuclear Quantum Effects on Surface Mediated Water  
Dissociation at Finite Temperatures**  
Yair Litman and Mariana Rossi
- TH 20**     **New Insights into the Catalytic Activity of Solid Solutions: Carbon  
Dioxide Activation at Ni<sub>x</sub>Mg<sub>y</sub>O**  
Aliaksei Mazheika, Marie M. Millet, Sabine Wrabetz, Andrey Tarasov,  
Elias Frei, Robert Schlögl, Matthias Scheffler, and Sergey V. Levchenko
- TH 21**     **Compressed-Sensing Analysis of Data for Carbon-Dioxide Activation at  
Metal-Oxide and Metal-Carbide Surfaces**  
Aliaksei Mazheika, Yanggang Wang, Runhai Ouyang, Rosendo Valero,  
Christian Kunkel, Helena Muñoz Galan, Luca M. Ghiringhelli, Francesc Illas,  
Sergey V. Levchenko, and Matthias Scheffler
- TH 22**     **The Crucial Role of Many-Body Dispersion and Temperature Effects in  
the (Meta)Stability of Gold Clusters (Au<sub>n</sub>, n = 5–14)**  
Bryan R. Goldsmith, Philipp Gruene, Jonathan T. Lyon, David M. Rayner,  
André Fielicke, Matthias Scheffler, and Luca M. Ghiringhelli

## **Organic Materials, Interfaces, and Biophysics**

### **TH 23<sup>OHG</sup> Structure and Electronic Properties of Diverse (Bio)organic Molecules in Hybrid Inorganic/Organic Systems**

Carsten Baldauf, Volker Blum, Svenja Janke, Sergey V. Levchenko, Dmitrii Maksimov, Haiyuan Wang, Matthias Scheffler, and Mariana Rossi

### **TH 24 Peptide-Cation Systems: Observing Kinetically Trapped Liquid-State Conformers in the Gas Phase and How to Improve the Force Field Description with Machine Learning**

Markus Schneider and Carsten Baldauf

## The Novel Materials Discovery Laboratory (NOMAD)

Luca M. Ghiringhelli, Fawzi R. Mohamed, Angelo Ziletti, Ankit Kariryaa, Benjamin Regler, Alfonso Sastre, Christian Carbogno, Sergey V. Levchenko, Georg Huhs<sup>a,b</sup>, Rubén Jesús García Hernández<sup>c</sup>, Atte Sillanpää<sup>d</sup>, Stefan Heinzl<sup>e</sup>, Kimmo Koski<sup>d</sup>, Dieter Kranzlmüller<sup>c</sup>, Jose Maria Cela<sup>b</sup>, Alessandro De Vita<sup>f</sup>, Daan Frenkel<sup>g</sup>, Francesc Illas<sup>h</sup>, Risto Nieminen<sup>i</sup>, Angel Rubio<sup>j</sup>, Kristian Sommer Thygesen<sup>k</sup>, Claudia Draxl<sup>a</sup>, and Matthias Scheffler

Since its launch in November 2015, the *NOMAD European Centre of Excellence (CoE)*, <https://nomad-coe.eu>, has been working to unify and improve the usefulness of computational materials science data. Orthogonally to other databases in this field that typically focus on a single computer code and are restricted to a closed research group or consortium, the *NOMAD CoE* serves **all important computer codes** and develops open Big-Data tools and services that benefit the whole materials-science community.

*NOMAD* creates, collects, cleanses, refines, and stores data from computational materials science, and develops tools to mine this Big-Data to find patterns, structure, and novel information that could not be discovered from studying smaller data sets. The large volume of data and innovative tools already available from *NOMAD* will enable researchers to advance materials science, identify new physical phenomena, and help industry to improve existing and develop novel products and technologies.

A recent Nature Editorial: “Empty rhetoric over data sharing slows science (Not-so-open data)”<sup>1</sup>, notes that many fields are resistant to openly sharing data. In materials science, there has been a cultural shift in attitude towards open data, fostered by *NOMAD*.<sup>2</sup> The *NOMAD Repository* provides open access to input and output files of computational materials science from all<sup>3</sup> major data collections in the field, together with those of many individual researchers and groups. In this role, the *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 44 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), and it offers DOIs to make the data citable.

The *NOMAD Repository* serves all important computer codes used in computational materials science (more than 30 very different codes). As repository this is extremely useful. However, the files created by the many different codes are very heterogeneous. Thus, they are not directly suitable for Big-Data analytics and extensive comparisons. *NOMAD* therefore converts the code-dependent, heterogeneous data of the *Repository* into code-independent, homogeneous data that are stored in the *NOMAD Archive*. This is achieved via *NOMAD Meta Info*, a flexible and extensible metadata infrastructure.

Building on the *Repository & Archive*, the *NOMAD* team has developed various tools and services: *Encyclopedia*, *Advanced Graphics*, and *Big-Data Analytics Toolkit*.

The *NOMAD Encyclopedia* is a web-based public infrastructure that provides a materials-oriented view of the *Archive* data. Its purpose is to provide access to whatever property of a given material has been computed, through a user-friendly graphical user interface (GUI). Properties available span from structural features to mechanical

## TH 1

behavior, thermal properties, electronic structure, transport characteristics, and the materials response to light and other excitations.

*Advanced graphics* aims at offering a visual and interactive data analysis for a large and geographically dispersed user base in an accessible and interactive way. As a proof of concept 3D videos and virtual-reality applications have been created for the visualization of crystal structures, molecular dynamics, Fermi surfaces, and electron densities for a selection of systems.

With the *Analytics Toolkit*, we present a general, flexible framework to perform analyses on materials-science data. On the one hand, the *Analytics Toolkit* empowers users to perform complex analyses, even if they have only a basic understanding of data analytics techniques, and no prior knowledge on how the data are stored in the *NOMAD Archive*. On the other hand, the toolkit is also flexible, allowing more advanced users to explore different settings, add new functionalities, and even to further develop the example tools to answer their own scientific and engineering questions.

The central entity of the *Analytics Toolkit* is a *notebook*, i.e., an interactive web page for writing and executing code. From these notebooks, the user has immediate access to all the libraries needed for advanced data analyses and the whole *NOMAD Archive*. Users do not need to install any software on their local machines, as everything they need to perform analytics on *NOMAD Archive* data is available via their web browser.

In order to demonstrate the functionality of the *Toolkit*, we developed tools and associated tutorials for various specific, but complementary, materials-science topics, which cover different aspects of Big-Data analytics and showcase different relevant materials-science applications. Behind these tools, there are methodologies developed at FHI and presented separately in the following posters: TH 11, 13, 14, 16, 17, 18, 21, and 22.

## References

1. Nature **546**, 327 (2017): <https://www.nature.com/news/empty-rhetoric-over-data-sharing-slows-science-1.22133>.
2. C. Draxl, F. Illas, and Matthias Scheffler, Nature **548**, 523 (2017): <http://www.nature.com/nature/journal/v548/n7669/full/548523d.html>.
3. 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.

<sup>a</sup> Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

<sup>b</sup> Barcelona Supercomputing Center, Spain

<sup>c</sup> Leibniz Supercomputing Centre, Garching, Germany

<sup>d</sup> CSC - IT Center for Science Ltd., Espoo, Finland

<sup>e</sup> Max Planck Computing and Data Facility, Garching, Germany

<sup>f</sup> Department of Physics, King's College London, London, UK

<sup>g</sup> Department of Chemistry, University of Cambridge, Cambridge, UK

<sup>h</sup> Department of Materials Science and Physical Chemistry / IQTCUB, University of Barcelona, Spain

<sup>i</sup> Department of Applied Physics, Aalto University, Finland

<sup>j</sup> Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany

<sup>k</sup> Physics Department, Technical University of Denmark, Lyngby, Denmark

## The FHI-aims Project

Volker Blum<sup>a</sup>, Danilo S. Brambilla, Björn Bieniek, Christian Carbogno,  
 Luca M. Ghiringhelli, Dorothea Golze<sup>b</sup>, Jan Hermann, Oliver T. Hofmann<sup>c</sup>,  
 Johannes Hoja<sup>d</sup>, William P. Huhn<sup>a</sup>, Yosuke Kanai<sup>e</sup>, Matthias Kick<sup>f</sup>, Jan Kloppenburg<sup>g</sup>,  
 Raul Laasner<sup>a</sup>, Björn Lange<sup>a</sup>, Marvin Lechner<sup>f</sup>, Chi Liu<sup>a</sup>, Andrew Logsdail<sup>h</sup>,  
 Andreas Marek<sup>i</sup>, Florian Merz<sup>ij</sup>, Reinhard Maurer<sup>k</sup>, Lydia Nemeč<sup>f</sup>, Harald Oberhofer<sup>f</sup>,  
 Nathaniel Raimbault, Xinguo Ren<sup>l</sup>, Karsten Reuter<sup>f</sup>, Stefan Ringe<sup>f</sup>, Patrick Rinke<sup>b</sup>,  
 Mariana Rossi, Christoph Schober<sup>f</sup>, Honghui Shang, Markus Sinstein<sup>f</sup>,  
 Christopher Sutton, Alexandre Tkatchenko<sup>d</sup>, Alvaro Vazquez-Mayagoitia<sup>m</sup>,  
 Mina Yoon<sup>n</sup>, Victor W. Yu<sup>a</sup>, Igor Y. Zhang, Tong Zhu<sup>a</sup>, and Matthias Scheffler

The FHI-aims all-electron electronic structure code, which begun at FHI almost 13 years ago, continues to be a flagship project for computational molecular and materials science, with a global community of users and developers. The core strength of the code is high numerical accuracy of the underlying basis sets and other numerical algorithms [now validated in several high-profile, community-wide cross-comparisons for density-functional theory<sup>1,2</sup> (DFT) and many-body perturbation theory<sup>3</sup> (MBPT)] while enabling scalability to very large problem sizes with support for current and next-generation high-performance computers. This poster highlights key recent enhancements: (i) DFT-based ground state electronic structure (meta-generalized gradient approximations, DFT+*U*, advanced methods for van der Waals interactions, efficient inclusion and broad validation of spin-orbit coupling across the periodic table), (ii) MBPT (more general approaches to the *GW* self-energy, the Bethe-Salpeter equation, periodic implementations of the random-phase approximation, second-order Møller-Plesset theory, and *GW*), (iii) Access to efficient, massively parallel Kohn-Sham solvers scaling as  $O(N^3)$  or lower through the ELPA project and a new electronic structure infrastructure “ELSI”, (iv) A density-functional perturbation theory framework for phonons, electronic friction, and spectroscopic observables (hyperpolarizabilities, Raman, and magnetic resonance), (v) Embedding in effective environments (polarizable continuum solvation, a modified Poisson-Boltzmann environment for solvation effects including ionic effects, fragment-orbital density-functional theory to estimate matrix elements for molecular electronic transport), (vi) Enhancements for high-throughput screening and embedding into statistical-physics tools, dedicated to structure prediction, cluster expansion (including Wang-Landau and nested sampling), molecular dynamics and thermal conductivity calculations, (vii) First steps on GPUs and Intel many-core architectures. Beyond the scientific functionality, future directions will include increasing exchange with and creation of software for community libraries, e.g., with ELPA, ELSI, libxc, SPGLib (symmetry), the CECAM Electronic Structure Library, and other projects.

### References

1. K. Lejaeghere *et al.*, Science **351**, aad3000 (2016).
2. S.R. Jensen, S. Saha, J.A. Flores-Livas, W.P. Huhn, V. Blum, S. Goedecker, and L. Frediani, J. Phys. Chem. Lett. **8**, 1449 (2017).
3. M.J. van Setten *et al.*, J. Chem. Theor. Comput. **11**, 5665 (2015).

<sup>a</sup> Duke University, Durham, NC, USA. <sup>b</sup> Aalto University, Finland. <sup>c</sup> Graz University of Technology, Austria. <sup>d</sup> University of Luxembourg, Luxembourg. <sup>e</sup> University of North Carolina, Chapel Hill, NC, USA. <sup>f</sup> Technical University of Munich, Garching, Germany. <sup>g</sup> Université Catholique de Louvain, Belgium. <sup>h</sup> University Cardiff, Wales, UK. <sup>i</sup> Max Planck Computing and Data Facility, Garching, Germany. <sup>j</sup> Lenovo Germany. <sup>k</sup> Yale University, CT, USA. <sup>l</sup> University of Science and Technology of China, Hefei, China. <sup>m</sup> Argonne National Laboratory, IL, USA. <sup>n</sup> Oak Ridge National Laboratory, TN, USA.

## Quantum-Chemistry Methods for Molecular and Periodic Systems in the Numeric Atom-Centered Orbital Framework

Arvid C. Ihrig, Tong H. Shen, Matthias Scheffler, and Igor Y. Zhang

Accurate quantum-chemistry methods, in particular the second-order Møller-Plesset perturbation theory (MP2) and the coupled-cluster approach with singles, doubles, and perturbative triples (CCSD(T)), have drawn increasing attention in materials science<sup>1</sup>. In contrast to popular density-functional methods in mean-field approximations, quantum-chemical approaches treat correlation effects explicitly in a many-electron interaction picture and pave a systematic way towards the exact solution in theory. However, the price to pay for these potential benefits is a considerably increased numerical complexity, which manifests itself in a steep computational scaling with system size, tremendous memory consumption and often poor large-scale parallelization efficiency. This poster presents in-memory, massively parallel implementations for a lower-scaling MP2 and a canonical CCSD(T) with the numerical atom-centered orbital (NAO) framework used in the FHI-aims code<sup>2</sup>.

In this project, we propose a generalization of our previously developed local Resolution of Identity (RI-LVL)<sup>3</sup> for correlated methods. By construction, this generalized RI-LVL paves the way towards an ideal utilization of the localizability for both NAOs and electron correlations. We took the Laplace-transformed MP2 (LT-MP2) as an example to show that our generalized RI-LVL allows for an in-memory, accurate, and large-scale parallel MP2 implementation. Combined with a MPI/OpenMP hybrid parallelization, this implementation features a quadratic memory scaling, a high parallel efficiency, and the computational time scaling is at least two orders of magnitude lower compared to the canonical MP2. The accuracy and performance of our new LT-MP2 algorithm are demonstrated with water clusters containing up to 150 molecules and the adsorption of water on TiO<sub>2</sub> surfaces with up to 270 atoms.

CCSD(T) is frequently considered the ‘gold standard’ in quantum chemistry and is expected to provide reference data for the method development in material science<sup>1</sup>. In this poster, a dedicated data distribution strategy with matching MPI/OpenMP in-memory implementation was designed to efficiently manipulate the massive and multiform tensor contractions during the CCSD iterations, which is the main challenge preventing massively parallel CCSD(T) simulations for solids. We demonstrate the accuracy and efficiency of our CCSD(T) implementation with a group of molecular test sets, polymers and solids. Taking a set of polymers as an example, we compare the numerical convergence of finite cluster and periodic CCSD(T) calculations. We find that the deviations in total energies are less than 2 meV on average, highlighting that our implementation can handle molecules and solids on the same numerical footing.

### References

1. A. Michaelides *et al.*, J. Chem. Phys. **143**, 102601 (2015).
2. V. Blum *et al.*, Comput. Phys. Commun. **180**, 20175 (2009).
3. A. Ihrig *et al.*, New J. Phys. **17**, 093020 (2015).

## Towards Efficient Orbital-Dependent Density Functionals for Weak and Strong Correlation

Igor Y. Zhang, Patrick Rinke<sup>a</sup>, John P. Perdew<sup>b</sup>, and Matthias Scheffler

The popularity of density-functional theory (DFT) in physics, chemistry, and materials science stems from the favorable balance between accuracy and computational efficiency offered by semilocal or hybrid approximations to the exchange-correlation (XC) functional. However, certain well-documented failures, in particular the inability to correctly describe strongly interacting scenarios with pronounced multi-reference character, such as bond dissociation, limit the predictive power of these functionals in certain cases<sup>1</sup>. On this poster, we present a new concept for the design of exchange-correlation functionals in density-functional theory.

We started from the Bethe-Goldstone equation (BGE) to derive a simple orbital-dependent correlation functional, namely BGE2, which terminates the BGE expansion at the second-order, but retains the self-consistent coupling of electron-pair correlations<sup>2</sup>. This electron-pair coupling effect can be efficiently improved by a simple screened variant (sBGE2). We demonstrated that sBGE2 is size consistent and one-electron ‘self-correlation’ free. The screened electron-pair correlation coupling ensures the correct H<sub>2</sub> dissociation limit and gives a finite correlation energy for any system even if it has no energy gap. sBGE2 correctly dissociates H<sub>2</sub> and H<sub>2</sub><sup>+</sup> dimers, which is regarded as a big challenge in DFT.<sup>1</sup>

We then took the sBGE2 functional as a building block for an orbital-dependent functional, termed ZRPS, which is a natural extension of the PBE0 hybrid functional<sup>3</sup>. ZRPS delivers an overall mean absolute deviation of less than 70 meV in our collection of benchmarks that includes 55 atomization energies, 76 reaction barriers, 34 isomerization energies and 22 weak interactions. Taking N<sub>2</sub> and C<sub>2</sub> dissociations and 2 ozone-involved chemical reactions with pronounced many-body multi-reference character as examples, we demonstrate that ZRPS yields a remarkable and consistent improvement over other density functionals across various chemical environments from weak to strong correlation.

### References

1. A.J. Cohen, P. Mori-Sánchez, and W.T. Yang, *Chem. Rev.* **112**, 289 (2011).
2. I.Y. Zhang, P. Rinke, and M. Scheffler, *New J. Phys.* **18**, 073026 (2016).
3. I.Y. Zhang, P. Rinke, J.P. Perdew, and M. Scheffler, *Phys. Rev. Lett.* **117**, 133002 (2016).

<sup>a</sup>COMP/Department of Applied Physics, Aalto University, Finland

<sup>b</sup>Department of Physics, Temple University, Philadelphia, PA, USA

## Novel High-Entropy Carbides Discovered by Synthesizability Descriptors

Pranab Sarker<sup>a</sup>, Tyler Harrington<sup>b</sup>, Cormac Toher<sup>a</sup>, Kenneth S. Vecchio<sup>b,c</sup>,  
Jon-Paul Maria<sup>d</sup>, Donald W. Brenner<sup>d</sup>, and Stefano Curtarolo<sup>e,f</sup>

High-entropy materials having a highly disordered homogeneous single-phase (possibly stabilized entirely by entropic contributions) continue to attract a great deal of research interest due to their unique combination of superb properties and promising technological applications. Predicting their formation from previously known parameters remains the major hindrance to the discovery of new systems. Here, we propose a descriptor – Entropy Forming Ability – for predicting the synthesizability of such systems from first principles calculations. The formalism, based on the spectral distribution of the thermodynamic density of states, captures the accessibility of equally-sampled states near the ground state and quantifies configurational disorder potentially leading to high-entropy homogeneous single-phases. Appropriate extensivity considerations allow recursive subcomponents prescreening to further accelerate the search of novel systems.

The descriptor captures the propensity of a material at a given composition to form a high-entropy single-phase by measuring, in an extensive framework, the energy spectrum of the thermodynamic density of states: narrow distribution implies a low energy cost for accessing the metastable configurations for a given composition, hence promoting randomness of the cations (*i.e.* entropy) into the system at finite  $T$ . On the contrary, a wide distribution would suggest a composition with a high energy cost for introducing different configurations, and thus with a strong preference for ordered phases. The method is benchmarked by the carbides showcase: given a set of 8 refractory metals (Nb, Ta, Mo, W, Ti, Zr, Hf, and V) plus carbon, the formalism predicts the sets of synthesizable 5-metal high-entropy-carbides. Candidates are experimentally tested, leading to a new class of carbide materials. The successful outcome demonstrates the strength of the synergy between thermodynamics, high-throughput computation, and experimental synthesis. The method can be extended to systems with configurational disorder and has the potential to accelerate the search and development of disordered systems by rationally combining first principles, descriptors' analysis and experimental synthesis.

<sup>a</sup>Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

<sup>b</sup>Program of Materials Science and Engineering, University of California, San Diego, La Jolla, CA, USA

<sup>c</sup>Department of NanoEngineering, University of California, San Diego, La Jolla, CA, USA

<sup>d</sup>Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC, USA

<sup>e</sup>Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany

<sup>f</sup>Permanent address: Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

## Compressed-Sensing Approaches to Kohn-Sham Density-Functional Theory and Beyond

Adrian Steffens<sup>a</sup>, Niklas Menzel, Ingo Roth<sup>a</sup>, Christian Krumnow<sup>a</sup>,  
Luca M. Ghiringhelli, Jens Eisert<sup>a</sup>, and Matthias Scheffler

The choice of basis sets plays a crucial role in density-functional theory and quantum chemistry, determining the accuracy and computational cost of the calculation. Computing the solution of a Kohn-Sham (KS) eigenvalue problem is generally achieved by diagonalizing the Fock matrix in order to obtain the eigenspace that corresponds to the lowest eigenvalues. The required computational effort scales with  $O(n^3)$ , which becomes dominant for large basis size  $n$ . Making use of locality – equivalent to requiring the Fock matrix to be banded with limited band width – the computational effort can be substantially reduced<sup>1</sup>, but may restrict the set of allowed states too much. Moreover, choosing an optimal basis set is a nontrivial task for many systems of interest, and large basis sets to compensate for this do not only increase computational time, but may also lead to numerical instability.

In order to provide a more flexible approach, we propose a scheme that employs central notions of compressed sensing<sup>2</sup>, a signal processing paradigm that has revolutionized the recovery of signals by efficiently identifying their underlying sparsity patterns, which makes it possible to massively undersample, yet fully recover a signal. In this spirit, we have developed an algorithm based on multi-task regularized Stiefel-manifold optimization that enforces sparsity basis function coefficients, thus identifying those basis functions that carry the most relevant information of the system. Our method makes it possible to automatically identify subsets of higher tier basis sets that have the same size as lower tier ones but yield lower energies.

On the other hand, for advanced exchange-correlation functionals and correlated methods, standard basis sets were not sufficiently accurate which led to the development of ‘correlation-consistent’ basis sets<sup>3</sup>. In a complementary compressed sensing approach, we have developed a new LASSO<sup>4</sup>-based method to select basis functions: The signals of interest are converged atomic orbitals for the free atom that can be represented as a sparse linear combination of a given initial pool of basis functions. We calculate the total energies for the atoms H to Ar and achieve comparable total energies as the reference orbitals. Furthermore, as a proof of concept, we show that by using as input a set of converged KS eigenvectors, with a known basis set expansion, but represented numerically on a radial grid, we can identify the basis functions and expansion coefficients out of a largely overcomplete basis set. Our new approach enables us to determine accurate basis sets for heavier atoms including  $d$ - and  $f$ -elements.

### References

1. S. Goedecker, Rev. Mod. Phys. **71**, 4:1085-1123 (1999).
2. E.J. Candès, J. Romberg, and T. Tao, IEEE Trans. Inf. Theory **52**, 2:489–509 (2006).
3. T.H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
4. R. Tibshirani, J.R. Stat. Soc. Ser. B **58**, 267 (1996).

<sup>a</sup> Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany

## Bridging Scales from van der Waals to Casimir Interactions

Jan Hermann, Andrii Kleshchonok, Majid Mortazavi, and Alexandre Tkatchenko<sup>a</sup>

Quantum fluctuations of electrons in matter are manifested in observable phenomena ranging from van der Waals (vdW) interactions between neutral atoms and molecules (0.1–1 nm) to Casimir interactions between macroscopic metal plates (1–10  $\mu\text{m}$ ). Although vdW and Casimir effects share the same underlying mechanism, the respective quantitative theories are disconnected and often narrowly applicable. Increasingly, however, a single experimental process can probe the target systems at multiple length scales, exposing gaps between existing calculations. Here, we present several advances in bridging these gaps, all based on a common many-body dispersion (MBD) framework<sup>1,2</sup>, forming a coherent multiscale approach to vdW interactions.

First, we show that the MBD framework provides a straightforward access to the microscopic electronic fluctuations in nanoscale materials<sup>3</sup>, while its combination with DFT achieves accuracy in binding energies comparable to high-level *ab initio* methods. Our analysis demonstrates that  $\pi$ - $\pi$  stacking in supramolecular complexes is characterized by unusually delocalized, collective charge fluctuations.

Next, we extend the MBD model beyond the dipole approximation to include the effects of electric fields, and find that these can contribute up to 35% of the vdW stabilization of amino acid dimers in potassium channels. Although our model is efficient and applicable to mesoscopic environments, it is in a remarkable agreement with much more expensive electronic-structure methods such as RPA or the coupled-cluster method.

We further combine the MBD model with density-functional tight-binding (DFTB) calculations, extending its potential applicability to nanosecond-scale molecular dynamics, which is inaccessible to standard density-functional calculations. In particular, we demonstrate that DFTB+MBD predicts structures of molecular crystals that give a 5% accuracy of the subsequent lattice energy calculations.

At mesoscale, retarded vdW (Casimir) interactions are usually modelled by electrodynamic continuum techniques using bulk material properties, thus lacking any microscopic information. We demonstrate that this approximation is unnecessary by using the MBD response properties as an input for the continuum calculations. We find that the largest deviations from approximate models occur for elongated semi-metallic systems<sup>4</sup>.

Taken together, our developments unite modelling of vdW and Casimir interactions from atomistic to mesoscopic ranges into a single homogeneous framework. This will enable seamless treatment of multiscale systems, and could bring rapid development of more accurate and general models for complex composite materials.

### References

1. A. Tkatchenko *et al.*, Phys. Rev. Lett. **108**, 236402 (2012).
2. J. Hermann, R.A. DiStasio, Jr., and A. Tkatchenko, Chem. Rev. **117**, 4714 (2017).
3. J. Hermann, D. Alfe, and A. Tkatchenko, Nat. Comm. **8**, 14052 (2017).
4. P.S. Venkataram, J. Hermann, A. Tkatchenko, and A.W. Rodriguez, Phys. Rev. Lett. **118**, 266802 (2017).

<sup>a</sup> Physics and Materials Science Research Unit, University of Luxembourg, Luxembourg

## Strongly Anharmonic Nuclear Dynamics in Solids: Accurate Computations and Rapid Estimates

Hagen-Henrik Kowalski, Florian Knoop, Maja-Olivia Lenz,  
Matthias Scheffler, and Christian Carbogno

Many important material properties are influenced or even completely governed by the *anharmonicity* of the potential-energy surface (PES), i.e., by features of the PES that are not captured by a *harmonic* second-order Taylor expansion. Typically, such vibrational effects are only treated perturbatively in first-principles calculations by solely accounting for the leading contribution to anharmonicity, i.e., the third-order term in a Taylor expansion of the PES.<sup>1,2</sup> Despite the fact that such an approximated PES is known to be unphysical,<sup>3</sup> very little is yet known about the role of higher-order terms, which require a full *ab initio* molecular dynamics (AIMD) treatment.

We shed light on this question by systematically investigating hundreds of materials with the perturbative approach<sup>2</sup> first. For this purpose, we use our recently developed high-throughput framework that is specifically targeted at the automatized calculation of thermodynamic equilibrium (vibrational dispersion, Grüneisen parameters, thermal expansion, etc.) and non-equilibrium properties (lattice thermal conductivity). From this material set, we chose representative examples with increasing degree of anharmonicity (e.g. Si, CuCl, Ga<sub>2</sub>O<sub>3</sub>, CsCl, ZrO<sub>2</sub>). We investigate their thermodynamic equilibrium and non-equilibrium properties and discuss the limits of the perturbative approach by additionally performing AIMD simulations. The thermal conductivity is then obtained via Green-Kubo relations using a recently developed approach, which accounts for **all** anharmonic effects.<sup>4</sup> These investigations reveal that a correct treatment of higher-order anharmonicity plays a minor role in good thermal conductors, but is essential to achieve quantitative predictions and qualitative insights in highly anharmonic systems such as ZrO<sub>2</sub>. In the latter case, the higher-order anharmonicity also drastically alters the equilibrium properties such as the vibrational frequencies. Also, our investigations reveal that the exchange-correlation functional and accounting for thermal expansion can critically alter the results. Eventually, we critically discuss the computational cost of such AIMD calculations and strategies to reduce it, e.g., with temperature-dependent Taylor expansions of the PES.<sup>5</sup> The obtained potentials, in which higher-order anharmonicity is incorporated in lower-order terms via renormalization, can be constructed on-the-fly from relatively few AIMD steps. We show that they provide routes for a qualitative interpretation of strong anharmonic effects.

### References

1. D.A. Broido, *et al.*, Appl. Phys. Lett. **91**, 231922 (2007).
2. A. Togo, L. Chaput, and I. Tanaka, Phys. Rev. B **91**, 094306 (2015).
3. G. Baym, Phys. Rev. **117**, 886 (1960).
4. C. Carbogno, R. Ramprasad, and M. Scheffler, Phys. Rev. Lett. **118**, 175901 (2017).
5. K. Esfarjani and H.T. Stokes, Phys. Rev. B **77**, 144112 (2008).

## Replica-Exchange Grand-Canonical Algorithm for Determining Phase Diagrams of Surfaces in Reactive Atmospheres

Yuanyuan Zhou, Matthias Scheffler, and Luca M. Ghiringhelli

For the advancement and refinement of modern technology, a large range of time and length scales need to be described, from the electronic (microscopic/atomistic) to the macroscopic regime. This calls for multi-scale modeling. Moreover, catalytic processes take place at finite temperature and pressure. In this project, we aim at developing a multi-scale modeling approach, applicable to a wide range of realistic conditions.

A temperature-pressure phase diagram describes the composition and structure of a system at thermal equilibrium and is an essential tool for understanding material properties. The *ab initio* atomistic thermodynamics (aiAT)<sup>1,2</sup> approach has been very successful in predicting phase diagrams for surfaces and gas-phase clusters<sup>3</sup> at realistic  $T$ ,  $p$  conditions. However, its limitation is the lack of an unbiased sampling of the configurational space.

Here, we introduce a novel Replica-Exchange (RE) Grand-Canonical (GC) Monte-Carlo (MC)/Molecular-Dynamics (MD) algorithm that enables the efficient calculation of complete pressure-temperature phase diagrams. We specifically target open systems in the GC ensemble, aiming at describing (nano)structured surfaces in a reactive atmosphere at realistic  $T$ ,  $p$ . RE allows for an efficient sampling by shuttling configurations from regions of low  $T$  or high chemical potential ( $\mu$ ) to regions of high  $T$  or low  $\mu$ . This massively parallel algorithm requires no prior knowledge of the phase diagram and takes only the potential-energy function together with the desired  $\mu$  and  $T$  ranges as inputs. From the partition function of the system it is then straightforward to calculate all desired thermodynamic variables.

We demonstrate REGC-MC by exploring a two-component Lennard-Jones system (a model for purely dispersion-forces interacting systems) with a surface in contact with a gas phase. From the estimated partition function, we construct the phase diagram in a range of temperatures  $T$  and pressures  $p$  and compare the results with phase diagrams constructed via AT. The two phase diagrams are very similar at low coverage (up to the formation of the first adlayer), but at higher coverages our phase diagram contains more phases, including (off-lattice) disordered ones.

Furthermore, we implement the REGC algorithm in the framework of density-functional theory. *Ab initio* REGC-MD is applied to Si<sub>2</sub> and Si<sub>4</sub> clusters in contact with a H<sub>2</sub> reactive atmosphere. We identify the thermodynamically most stable phases, including dissociated and non-dissociated H<sub>2</sub> molecules, at realistic  $T$ ,  $p$ (H<sub>2</sub>) conditions, as a first step for the understanding of chemical vapor deposition of amorphous hydrogenated silicon.

### References

1. M. Scheffler and J. Dabrowski, Phil. Mag. A **58**, 107 (1988).
2. K. Reuter and M. Scheffler, Phys. Rev. B **65**, 035406 (2001).
3. S. Bhattacharya, S. Levchenko, L.M. Ghiringhelli, and M. Scheffler, Phys. Rev. Lett. **111**, 135501 (2013).

# Adiabatic Vibronic Coupling in Molecules and Solids: Physical Insight from Density-Functional Perturbation Theory and Machine Learning

Honghui Shang, Nathaniel Raimbault, Carlos Mera Acosta<sup>a</sup>,  
Patrick Rinke<sup>b</sup>, Matthias Scheffler, Mariana Rossi, and Christian Carbogno

First-principles calculations of the coupling of vibrational and electronic degrees of freedom have become possible only in recent years<sup>1</sup>, and still little is known about the underlying qualitative mechanisms, the limits of the current methodologies, and the trends across chemical and structural space. In this work, we combine our recent density-functional perturbation theory (DFPT) implementation<sup>2</sup> with machine-learning techniques to shed light on these issues for two examples: anharmonic vibrational Raman spectra and temperature-dependent electronic band structures.

Vibrational Raman spectroscopy is frequently used for characterizing different molecular-crystal polymorphs and their phase transitions. Using the DFPT framework for electric field responses, we obtain polarizability tensors for isolated and periodic systems, and calculate fully anharmonic Raman spectra from the Fourier transform of the polarizability autocorrelation functions. We focus on polymorphs of Paracetamol and Aspirin containing hundreds of atoms, and use the PBE functional augmented by many body van der Waals interactions<sup>2</sup>. Thermal lattice expansion is taken into account. Comparison with published experimental studies shows good quantitative agreement on low and mid-frequency ranges. We observe strong anharmonic effects on NH bending modes and on low-frequency modes corresponding to collective vibrations of the lattice. In order to move towards a high-throughput framework and to bypass the cost of these DFPT calculations, we employ Kernel Ridge Regression to obtain DFPT-level polarizability tensors by exploiting geometrical data and polarizabilities obtained from cheaper semi-empirical models.

We have also investigated the influence of nuclear motion on the electronic band structure of solids at 0K (zero-point renormalization: ZPR) and as a function of temperature for 82 octet binaries in both the zincblende and the rocksalt structure. After validating our approach and the employed exchange-correlation functional for polar and non-polar systems, we discuss the observed trends in chemical and structural space, especially the several exceptions to generally accepted rule-of-thumbs, e.g., to the so-called mass rule (ZPR should decrease with atomic mass)<sup>3</sup>. To explain these deviations, a machine-learning techniques (SISSO) is used to identify reliable, physically meaningful descriptors<sup>4</sup>. In turn, this analysis reveals (*i*) that the descriptors are crystal-structure dependent and (*ii*) that the investigated effects cannot be explained solely in terms of atomic descriptors, but require explicit information on the bonding, e.g., by including features of dimers in the descriptor identification.

## References

1. F. Giustino, *Rev. Mod. Phys.* **89**, 015003 (2017).
2. H. Shang *et al.*, *Comput. Phys. Commun.* **215**, 26 (2017).
3. M. Cardona and M.L.W. Thewalt, *Rev. Mod. Phys.* **77**, 1173 (2005).
4. R. Ouyang *et al.*, (in preparation), see TH 11.

<sup>a</sup> Present address: Instituto de Física, Universidade de São Paulo, Brazil

<sup>b</sup> COMP/Department of Applied Physics, Aalto University, Finland

## Massively Parallel Compressed-Sensing Optimizer for Identifying Materials-Property Descriptors: Application to the Classification of Materials as Metals

Runhai Ouyang, Emre Ahmetcik, Stefano Curtarolo<sup>a</sup>, Matthias Scheffler, and Luca M. Ghiringhelli

Discovering the relationship between a property of a material and its chemical composition and structure is of great importance for understanding the underlying mechanisms. In recent years, there has been a rapid growth of research interest in developing machine-learning approaches for finding the relationship between a set of descriptive parameters (termed descriptor) of a material and the property of interest. For this task, Ghiringhelli *et al.*<sup>1,2</sup> recently proposed the LASSO+ $\ell_0$  approach, a compressed-sensing (CS) method for descriptor identification from a pool of candidates (feature space) based on minimizing the  $\ell_1$  and  $\ell_0$  norms. The last two years have shown that this approach has severe limitations as it requires modest pool size and low mutual correlation between the offered descriptor candidates.

We have recently developed a new CS<sup>3</sup> scheme that combines Sure Independence Screening (SIS)<sup>4</sup> with a sparsifying operator (SO). SISSO overcomes the above-mentioned problems and enables us to find a proxy to the optimal descriptor in huge-sized feature spaces (billions of descriptor candidates or more). The approach starts with feature construction by building analytical functions of the input physical parameters (primary features), via the application of various algebraic operators. The SIS iteratively prunes the feature space by discarding features that have low correlation with the target property, and subsequent residuals. SIS is followed by the application of a SO method for finding the optimal solution inside the pruned feature space. Here, we apply the  $\ell_0$ -norm regularization that in practice finds the optimal solution by full enumeration.

The approach was applied to the metal/nonmetal classification of binary  $A_xB_y$  materials belonging to several structure prototypes. Several hundreds of materials as listed in the Springer Materials database<sup>5</sup> were used to train the model. A simple descriptor, which classifies all the training metal/nonmetal, was identified. It is a function of atomic radius, ionization energy, electronegativity, and cell volume. The found descriptor provides insights into the relationships between the metallic/nonmetallic character and the structure of a material.

### References

1. L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, and M. Scheffler, *Phys. Rev. Lett.* **114**, 105503 (2015).
2. L.M. Ghiringhelli *et al.*, *New J. Phys.* **19**, 023017 (2017).
3. E.J. Candès and M.B. Wakin, *IEEE Signal Process. Mag.* **25**, 21 (2008).
4. J. Fan and J. Lv, *J.R. Statist. Soc. B* **70**, 849 (2008).
5. <http://materials.springer.com/>.

<sup>a</sup> Permanent address: Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

## Structure Dependence of Small Polarons in Transparent Oxides: *Ab Initio* Study of Polymorphs of TiO<sub>2</sub> and Ga<sub>2</sub>O<sub>3</sub>

Sebastian Kokott, Sergey V. Levchenko, Patrick Rinke<sup>a</sup>, and Matthias Scheffler<sup>b</sup>

Thin films of transparent oxides are an important component for the fabrication of solar cells, photocatalysts, and more. Since the applications are extremely diverse, conductivity in the films is either required or should be suppressed for certain types of charge carriers. For both scenarios a detailed theoretical understanding is necessary in order to predict new functional materials. One important factor limiting the conductivity is the interaction of the charge carrier with phonon modes, i.e. the formation of polarons. The strength of the electron-phonon (el-ph) interaction determines the localization of the polaron, which in turn defines its characteristic temperature dependence for the charge-carrier mobility and ensuing properties. The present study focusses on materials with strong el-ph coupling, where the localization of the polaron has a length of the order of the interatomic distance (a so-called small polaron).

Though density-functional theory (DFT) has been often used for calculating properties of polarons, there are two challenges that were not addressed properly, so far: sensitivity to errors in the exchange-correlation approximation and finite-size effects in supercell calculations. In this work, we developed an approach that addresses these challenges. The polaron properties are obtained using a modified neutral potential-energy surface (PES)<sup>1</sup>. Using the hybrid HSE functional and considering the whole range  $0 \leq \alpha \leq 1$  of exact exchange, we show that the modified PES model significantly reduces the dependence of the polaron level and binding energy on the functional. Based on Pekar's model<sup>2</sup>, we derive the proper elastic long-range behavior of the polaron and a finite-size correction that allows to obtain the polaron properties in the dilute limit.

Specifically, we investigated the influence of the crystal structure on the polaron properties for rutile and anatase polymorphs of TiO<sub>2</sub> and for the monoclinic  $\beta$ - and orthorhombic  $\epsilon$ -polymorphs of Ga<sub>2</sub>O<sub>3</sub>. The existence of small polarons in TiO<sub>2</sub> is a matter of a continuing discussion<sup>3-5</sup>, and our study points out issues not considered thus far. We find that small-polaron formation in TiO<sub>2</sub> is sensitive to the crystal structure: While in rutile TiO<sub>2</sub> only small *electron* polarons are stable, only small *hole* polarons are found in anatase. On the contrary, small hole polarons exist in both Ga<sub>2</sub>O<sub>3</sub> polymorphs but have significantly different binding energies.

### References

1. B. Sadigh, P. Erhart, and D. Åberg, Phys. Rev. B **92**, 075202 (2015).
2. S.I. Pekar, Zh. Eksp. Teor. Fiz. **16**, 335 (1946).
3. C. Di Valentin and A. Selloni, J. Phys. Chem. Lett. **2**, 17 (2011).
4. C. Spreefico and J. VandeVondele, Phys. Chem. Chem. Phys. **16**, 47 (2014).
5. M. Setvin *et al.*, Phys Rev. Lett. **113**, 086402 (2014).

<sup>a</sup>COMP/Department of Applied Physics, Aalto University, Finland

<sup>b</sup> Also at: University of California at Santa Barbara, CA, USA

## Predicting and Understanding Quantum Spin-Hall Insulators with the Help of Machine Learning

Carlos Mera Acosta<sup>a</sup>, Runhai Ouyang, Adalberto Fazzio<sup>b</sup>, Matthias Scheffler, Luca M. Ghiringhelli, and Christian Carbogno

Quantum spin-hall insulators (QSHIs) are two-dimensional topological insulators (TIs) with a spin-orbit coupling (SOC) driven band inversion<sup>1</sup>. They have attracted considerable interest in recent years for spintronic applications<sup>2</sup>, since backscattering is forbidden in their edge states that are protected by time-reversal symmetry<sup>1</sup>. Accordingly, QSHIs can be characterized by a non-zero topological  $Z_2$  invariant<sup>2</sup>. So far, the computational search for new QSHIs has been a costly trial-and-error process guided by empirical rule-of-thumbs, e.g., focusing on heavy elements with a strong SOC.

In this work, we investigated the topological character of 220 two-dimensional,  $C$ -functionalized honeycomb lattice materials  $ABC_2$  ( $AB$  are IV-IV, III-V, II-VI materials, and  $C$  is from group VII). Starting from DFT-PBE calculations that account for van-der-Waals interactions<sup>3</sup> and spin-orbit coupling<sup>4</sup>, we computed the  $Z_2$  invariant using our recent local atomic orbitals based implementation of the Wannier centers of charge evolution formalism<sup>5</sup>. Besides confirming the QSHI character of known materials, e.g., functionalized stanene, our study found 15 other yet unreported QSHIs, as for example  $AlNBr_2$  and  $GaAsBr_2$ . Quite surprisingly, these TIs consist of relatively light elements and thus defy the widespread reasoning that heavy elements with strong SOC are required to induce a topological transition. To understand these trends, we use a recently developed machine-learning approach<sup>6</sup> to classify these compounds by means of a physically meaningful two-dimensional *descriptor* that only depends on the properties of the material's atoms, but not on the properties of the material itself. This analysis reveals that the actual topological transition is not only triggered by the strength of the SOC, but by the relative position and mixture of hybridized atomic orbitals in the valence and conduction band. As a consequence, strongly electronegative functionalization with fluorine or chlorine typically suppresses the QSHI state, whereas functionalization with bromine or iodine can trigger it. Besides correctly capturing this behavior, the found descriptors also correlate with the computed bond-lengths and lattice constants. Eventually, we show that the identified descriptors can predict over 50 additional QSHIs out of a different set of 140 2D honeycomb compounds that consist of elements not included in our original machine-learning procedure.

### References

1. V.A. Volkov and O.A. Pankratov, JEPT Lett. **42**, 178 (1985).
2. J.D. Jariwala *et al.*, ACS Nano **8**, 1102 (2014).
3. A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).
4. W.P. Huhn and V. Blum, Phys. Rev. Mater., in press (2017), preprint at arXiv:1705.0180.
5. R. Yu *et al.*, Phys. Rev. B **84**, 075119 (2011).
6. L.M Ghiringhelli *et al.*, New J. Phys. **19**, 023017 (2017).

<sup>a</sup> Present address: Instituto de Física, Universidade de São Paulo, Brazil

<sup>b</sup> Instituto de Física, Universidade de São Paulo, Brazil

## The Face of Crystals: Insightful Classification Using Deep Learning

Angelo Ziletti, Devinder Kumar<sup>a</sup>, Matthias Scheffler, and Luca M. Ghiringhelli

The crystal structure - how atoms are periodically arranged in space - is an essential ingredient for predicting properties from a material's chemical composition. It has also a large impact on real-world materials applications. For example, in iron, carbon solubility (important for steel formation) increases nearly forty times going from body-centered-cubic  $\alpha$ -ferrite to face-centered-cubic  $\gamma$ -austenite.

With the advent of high-throughput materials-science computations, millions of calculated crystals structures are now available to the scientific community<sup>1</sup>. A reliable identification of the lattice symmetry is a critical first step for their characterization and analysis. Current methods require a user-specified threshold, and are unable to detect “average” symmetries for defective structures. Clearly, there is no threshold which performs optimally (or even sub-optimally) for such a large amount of data, nor a well-defined procedure to test the soundness of a chosen threshold.

Here, we propose a new machine-learning based approach to automatically classify crystals based on their symmetry. Unlike other representations (e.g. Voronoi tessellation<sup>2</sup>, which is real-space based) we represent crystals in reciprocal space, and in particular by diffraction images. From these, we construct a deep neural network model<sup>3</sup> for classification. Using this procedure, we are able to correctly classify a dataset comprising more than 80,000 heavily defective cubic structures, without using any tolerance threshold. Unfolding the model's internal operations, we show that the neural network uses the same landmarks a materials scientist would use, although it was never explicitly instructed to do so.

Thanks to its robustness to defects, our method could also be used to determine local microstructures in atomic probe tomography experiments<sup>4</sup> (arguably the most important source of atomic structural information) where, however, more than 20% of the atoms escape detection.

### References

1. NOMAD Repository (<https://repository.nomad-coe.eu>), NOMAD Laboratory (<https://www.nomad-coe.eu/>) and references therein.
2. O. Isayev *et al.*, Nature Communications **8**, 15679 (2016).
3. Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
4. M.J. Duarte *et al.*, Science **341**, 372 (2013).

<sup>a</sup> Department of Systems Design Engineering, University of Waterloo, Ontario, Canada

## High-Throughput Screening of Transparent Conducting Oxides

Christopher Sutton, Luca M. Ghiringhelli, and Matthias Scheffler

Transparent conducting oxides (TCOs) are a well-developed and commercialized class of wide-bandgap semiconductors that are crucial for the operation of many electronic devices. Despite the widespread interest, the fundamental properties of the various binary polymorphs of group-III sesquioxides ( $M_2O_3$ , where  $M = Al, Ga, \text{ and } In$ ) are still poorly understood. This is in part because reliable experimental data on carefully characterized bulk crystals are largely missing. Therefore, an understanding of the key properties for these materials must instead be obtained through computations.

In this work, we first detail a new approach for the high-throughput screening of the oxides, which we define as `sp-HSE06@PBEsol`. This method utilizes a single-point HSE06 calculation at the PBEsol-optimized ground-state, which gives bandgap energies that are in good agreement with experimental values (13% error), but at a significantly reduced computational cost compared to a full structure relaxation at the HSE06 level. Using this approach, the trends in the bandgaps (for transparency) and effective masses (for conductivity) are examined for many polymorphs of  $Al_2O_3$ ,  $Ga_2O_3$ , and  $In_2O_3$ . We detail new polymorphs, some of which have a larger bandgap or smaller effective mass compared to the currently known thermodynamically stable phases.

A systematic examination of the thermodynamics of ordering and phase separation in ternary and quaternary mixtures with the formula  $(In_xGa_yAl_z)_2O_3$  (where  $x+y+z=2$ ) is also carried out in this work. A particular focus of this research is to investigate the stability and electronic properties of various phases for the alloy as a function of composition and temperature. This could allow for the design of new materials with an improved performance (*e.g.*, transparency over the visible range, while still maintaining *n*-type conductivity). To efficiently search (meta)stable configurations for many lattice types, a combination of DFT-based cluster expansion models and fast stochastic optimization techniques are used. The cluster expansion program has been developed in collaboration with the Draxl group at the Humboldt-Universität zu Berlin.<sup>1</sup> The fast stochastic optimization techniques that we employ are nested sampling<sup>2</sup> and our recently developed semi-grand extension,<sup>3</sup> which allows for a “one-shot” calculation over various compositions.

### References:

1. S. Rigamonti *et al.*, *In preparation*.
2. R.J.N. Baldock, L.B. Partay, A.P. Bartok, M.C. Payne, G. Csanyi, *Phys. Rev. B* **93**, 174108 (2016).
3. R.J.N. Baldock, C. Sutton, L.M. Ghiringhelli, M. Scheffler, N. Marzari, *In preparation*.

# Unified Representation of Molecules and Crystals for Machine Learning

Haoyan Huo<sup>a,b</sup> and Matthias Rupp

Accurate simulations of polyatomic systems require first-principles approaches whose high computational cost limits number and size of studied systems. In high-throughput settings, machine learning can reduce these costs significantly by interpolating between reference calculations<sup>1</sup>. This requires a numerical representation for arbitrary atomistic systems that supports interpolation. For this purpose we introduce<sup>2</sup> a *many-body tensor representation* (MBTR) enabling sub-kcal/mol prediction errors for molecular energies and single-digit meV/atom errors for materials (bulk crystal) in benchmark settings. For molecules, MBTR outperforms published results we are aware of; for materials, it is the first competitive approach not requiring locality assumptions.

The key idea of MBTR is to represent an atomistic system via distributions of  $k$ -body terms, such as (inverse) distances or angles, stratified by chemical elements (Fig. 1). Its value at  $x$  for a combination of elements  $\mathbf{Z} = Z_1, \dots, Z_k$  is

$$MBTR_k(x, \mathbf{Z}) = \sum_{i_1, \dots, i_k=1}^{N_a} w_k(i_1, \dots, i_k) D(x, g_k(i_1, \dots, i_k)) \prod_{j=1}^k c_{Z_j Z_{i_j}},$$

where  $x$  and  $\mathbf{Z}$  correspond to the  $x$ -axes and lines in Fig. 1, respectively,  $N_a$  is the number of atoms,  $g_k$  computes  $k$ -body terms,  $D$  is a normal distribution broadening  $g_k$ , and  $w_k$  is a weighting function decaying with distance between atoms. The product term allows for correlations between chemical elements.

We present atomization energy prediction errors below 1 kcal/mol for 7000 organic molecules and 6 meV/atom for 11000 elpasolite crystals. 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.

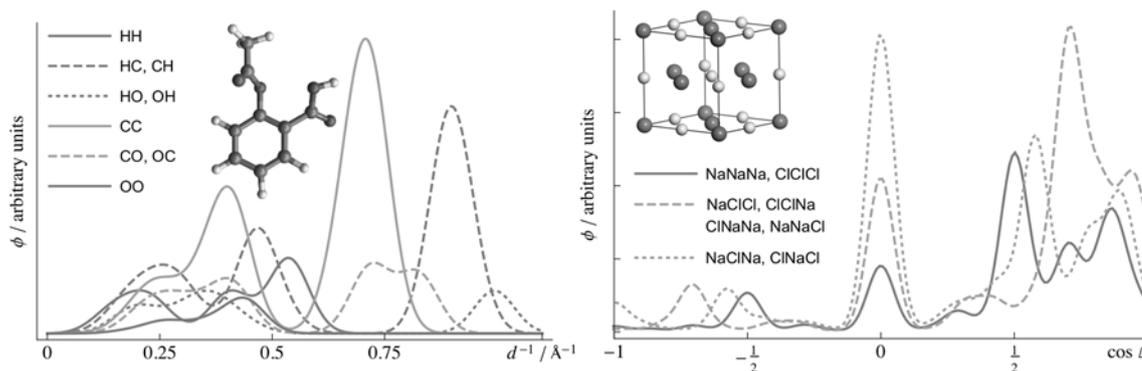


Figure 1: Visualization of many-body tensor representation for aspirin (left,  $k=2$ , inverse distances, quadratic weighting) and fcc salt (right,  $k=3$ , angles, exponential weighting).

## References

1. M. Rupp, A. Tkatchenko, K.-R. Müller, and O.A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012); J. Behler, Angew. Chem. Int. Ed., in press.
2. H. Huo and M. Rupp, arXiv, 1704.06439 (2017).

<sup>a</sup> School of Physics, Peking University, Beijing, China

<sup>b</sup> Present address: University of California at Berkeley, CA, USA

## Subgroup Discovery for Finding Interpretable Local Patterns in Data from Materials-Science

Bryan R. Goldsmith<sup>a</sup>, Mario Boley<sup>b</sup>, Jilles Vreeken<sup>b</sup>, Matthias Scheffler,  
and Luca M. Ghiringhelli

Big-data analytics tools (i.e., machine learning, compressed sensing, and data mining) in combination with the growth of materials-science repositories are creating opportunities for finding descriptors that help understand and screen novel materials. Data analytics tools applied to materials-science data often focus on the inference of a global prediction model for some property of interest for a given class of materials. However, the underlying mechanism for some target property could differ for different materials within a large pool of materials data. Consequently, a global model fitted to the entire dataset may be difficult to interpret and may well hide or incorrectly describe the actuating physical mechanisms. In these situations, local models describing subgroups would be advantageous to global models. As a more physical example, the transition metals of the periodic table are a subgroup, and the actinides and halogens are other subgroups. Therefore, identification of subgroups is useful to understand similarities and differences between materials.

In this poster, we demonstrate that subgroup discovery (SGD), a form of local pattern discovery for labeled data<sup>1</sup>, can help uncover physically meaningful descriptors from materials-science data obtained by first-principles calculations<sup>2</sup>. In contrast to global modelling algorithms, SGD finds interpretable descriptions of subpopulations in which, locally, the target property takes on an interesting distribution. At first, we formulate the SGD algorithm for materials applications. Subsequently, SGD is applied to 24 400 configurations of neutral gas-phase gold clusters (having 5 to 14 atoms) to discern general patterns between their geometrical and physicochemical properties<sup>2</sup>. For example, SGD helps find that van der Waals interactions within gold clusters are linearly correlated with their radius of gyration and are weaker for planar clusters than for nonplanar clusters. Additionally, SGD is demonstrated to identify subgroups that classify the crystal structures of the octet binary semiconductors as either rock salt or zincblende from only information of its chemical composition. SGD finds a simple two-dimensional model derived from only the atomic radii of valence  $s$  and  $p$  orbitals that classifies the crystal structures for 79 of the 82 octet binary semiconductors. An efficient optimal solver using branch-and-bound (an efficient algorithm for combinatorial searches) is also developed for objective functions that are corrected for the spread of the data distribution. This facilitates the discovery of meaningful subgroups<sup>3</sup>.

### References

1. W. Duivesteijn, A.J. Feelders, A. Knobbe, *Data Min. Knowl. Discov.* **30**, 47 (2016).
2. B.R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, and L.M. Ghiringhelli, *New J. Phys.* **19**, 013031 (2017).
3. M. Boley, B.R. Goldsmith, L.M. Ghiringhelli, and J. Vreeken, *Data Min. Knowl. Discov.* **31**, 1391 (2017).

<sup>a</sup> Present address: Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA

<sup>b</sup> Max Planck Institute for Informatics, Campus Mitte, Saarbrücken, Germany

## Higher-Order Phase Transitions of a Hydrogenated ZnO (10-10) Surface

Maria E. Stournara, Sergey V. Levchenko, Santiago Rigamonti<sup>a</sup>, Maria Troppenz<sup>a</sup>, Oliver T. Hofmann<sup>b</sup>, Patrick Rinke<sup>c</sup>, Claudia Draxl<sup>a</sup>, and Matthias Scheffler<sup>d</sup>

Zinc oxide (ZnO) has unique properties and a wide range of applications. In particular, it is a promising substrate for hybrid inorganic-organic photovoltaic and electronic devices. The performance of such devices is affected by the structure and composition of ZnO surfaces. Since hydrogen (H) is a common contaminant during and after ZnO fabrication, it is paramount to identify and predict surface phases of ZnO at realistic temperatures and hydrogen partial pressures.

Following the work of Borg *et al.*<sup>1</sup>, we applied *ab initio* atomistic thermodynamics on a cluster expansion model for atomic H adsorbed at the thermodynamically most stable ZnO (10-10) surface. Using the Wang-Landau algorithm<sup>2</sup>, we obtain all thermodynamic potentials, including the configurational entropy of the adsorbate. The model has been parameterized with density-functional theory (DFT) calculations (PBE functional). The effects of vibrations and higher-order methods (i.e., hybrid functionals) have been carefully tested and found to be qualitatively unimportant. The long-range many-body van der Waals (vdW) interactions are included using the Tkatchenko-Scheffler<sup>3</sup> scheme.

Our study reveals that at coverages below 6%, H atoms adsorb exclusively on surface O atoms. At higher coverages, H adsorbs also on Zn, but an excess (i.e., higher partial coverage) of O-H over Zn-H is retained at all coverages. Only at 50% and 100% coverage O-H equals the Zn-H coverage. Due to the interplay of long- and short-range interactions, neighboring O-H/Zn-H pairs form chains along surface -O-Zn- rows, with each chain anchored at an O-H. The interaction between the chains is dominated by the repulsion between the O-H ends, so that any possible relative ordering of the chains is destroyed at  $T > 60$  K. Interestingly, while no order-disorder (first-order) phase transitions are found at realistic conditions, there are phases of adsorbed H that are characterized by a constant O-H excess. This excess is a good descriptor for the work function. We use this descriptor to calculate the dependence of the work function on H coverage using large supercells, and find good agreement with recent experiments performed in the group of Julia Stähler at the FHI<sup>4</sup>. Our results offer a “road map” for H adsorption on the ZnO (10-10) surface for a wide range of different temperature and pressure conditions, consolidating previous experiments<sup>4-7</sup>.

### References

1. M. Borg *et al.*, ChemPhysChem **6**, 1923 (2005).
2. F.G. Wang and D.P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
3. A. Tkatchenko *et al.*, Phys. Rev. Lett. **108**, 236402 (2012).
4. J.C. Deinert *et al.*, Phys. Rev. B **91**, 235313 (2015).
5. Y. Wang *et al.*, Phys. Rev. Lett. **95**, 266104 (2005).
6. K. Ozawa and K. Mase, Phys. Stat. Sol. a **207**, 277 (2010).
7. K. Ozawa and K. Mase, Phys. Rev. B **83**, 125406 (2011).

<sup>a</sup> Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

<sup>b</sup> Institute of Solid State Physics, NAWI Graz, Graz University of Technology, Austria

<sup>c</sup> COMP/Department of Applied Physics, Aalto University, Finland

<sup>d</sup> Also at: University of California at Santa Barbara, CA, USA

## Impact of Nuclear Quantum Effects on Surface Mediated Water Dissociation at Finite Temperatures

Yair Litman and Mariana Rossi

The interface of molecules with inorganic substrates has received great attention in the scientific community in search of new technological devices. The presence of light atoms in these adsorbates can greatly enhance the importance of nuclear quantum effects in the structural stability and dynamical properties (e.g. reactions rates) of these systems, as highlighted by several studies that find non-trivial isotope effects in different properties of these systems<sup>1</sup>.

In this work, we explore the impact of nuclear quantum effects (NQE) on the dissociation of water wires on stepped Pt(221) surfaces, where NQE have been previously identified as necessary to stabilize dissociated structures<sup>2</sup>. By performing *ab initio* molecular dynamics simulations with density-functional theory and van der Waals dispersion<sup>3</sup>, we note that there are several competing minima for both the intact and dissociated structures, making the usually reported harmonic estimates of the contributions of NQE to their free energies a questionable approximation. We thus performed *ab initio* path integral molecular dynamics (PIMD) in order to calculate these contributions and their interplay with the electronic structure in an accurate fully anharmonic fashion. We propose a scheme based on a spatial partition of the system to accelerate PIMD simulations when adsorption is weak and NQE on the substrate are negligible, which we implement in the i-PI code<sup>4</sup>. We treat NQE on the adsorbate in full but in isolation, while the surface-adsorbate system is evaluated contracted to a few beads. This represents a considerable speed up of the simulations. This scheme allowed us to calculate the contribution of NQE to the free energies in an all-electron framework, summing up to an equivalent of 1.1 ns of simulations. We find that anharmonicities result in an increase of up to 20% per dissociated dimer of the quantum contribution to the dissociation free energy, compared to harmonic estimates.

We also show how both temperature and NQE indirectly impact dipoles and the redistribution of electron density, causing work function changes of up to 0.4 eV with respect to static estimates. This quantitative assessment provides a possible approach to determine experimentally the most stable configurations of water oligomers on the stepped surfaces. Finally, we find that zero point energy can increase the dissociation rate by a factor of ~1000 in the harmonic approximation. To go beyond this picture, we present an implementation of ring polymer instanton theory<sup>5</sup> that can be used to calculate tunneling contributions in these (and other) multidimensional atomistic simulations of proton transfer reactions.

### References

1. T. Koitaya and J. Yoshinobu, Chem. Rec. **14**, 848 (2014).
2. D. Donadio, L.M. Ghiringhelli, and L. Delle Site, J. Am. Chem. Soc. **134**, 19217 (2012).
3. V.G. Ruiz *et. al.*, Phys. Rev. Lett. **108**, 146103 (2012).
4. M. Ceriotti, J. More, and D.E. Manolopoulos, Com. Phys. Com. **165**, 1019 (2014).
5. A.N. Beyer *et. al.*, J. Phys. Chem. Lett. **7**, 4374 (2016).

## New Insights into the Catalytic Activity of Solid Solutions: Carbon Dioxide Activation at $\text{Ni}_x\text{Mg}_y\text{O}$

Aliaksei Mazheika, Marie M. Millet, Sabine Wrabetz, Andrey Tarasov, Elias Frei,  
Robert Schlögl, Matthias Scheffler<sup>a</sup>, and Sergey V. Levchenko

$\text{NiO/MgO}$  solid solutions are stable and active catalysts for dry reforming of  $\text{CH}_4$  and  $\text{CO}_2$  hydrogenation. Despite numerous experimental investigations, the structure of  $\text{Ni}_x\text{Mg}_y\text{O}$  surfaces as well as the nature of the catalytically active centers remained unclear. We present a combined hybrid DFT/microcalorimetry study of  $\text{CO}_2$  adsorption at  $\text{Ni}_x\text{Mg}_y\text{O}$  surfaces. For the first time, poly-crystalline samples with low concentrations (3-10%) of Ni in MgO were successfully synthesized and experimentally studied. The homogeneous distribution of Ni in MgO allowed a clear comparison of theory and experiment, and clarified the effects of Ni on MgO surface structure. Periodic DFT calculations with the hybrid HSE( $\alpha$ ) functional<sup>1</sup> augmented with a many-body dispersion correction<sup>2</sup> are performed, where the fraction of exact exchange  $\alpha$  is set to 0.3 based on a comparison of the HSE( $\alpha$ ) and CCSD(T) embedded-cluster calculations of energies of CO,  $\text{CO}_2$ , and  $\text{H}_2$  adsorption at  $\text{Ni}_x\text{Mg}_y\text{O}$  flat and stepped (100) surfaces, and  $\text{Ni}_{\text{Mg}}$  point defect formation energies in  $\text{MgO}$ <sup>3</sup>.

The calculations reveal that  $\text{Ni}_{\text{Mg}}$  defects prefer to occupy low-coordinated sites (corners and steps) at the MgO (100) surface. Thus, Ni doping promotes formation of such sites during synthesis of the solid solution. Indeed, diffuse reflectance infrared spectroscopy (DRIFTS) showed that the number of undercoordinated O sites at  $\text{NiMgO}$  surfaces is significantly increased compared to pure MgO samples. Calculated  $\text{CO}_2$  adsorption energy at the MgO (100) terrace (0.64 eV) at low coverage is much lower than the experimental one (1.15-1.25 eV for 3 at.% of Ni), while at the steps it is much higher, about 2.2 eV. Increasing coverage of  $\text{CO}_2$  at steps reduces the adsorption energy to 1.6 eV, and increasing the concentration of  $\text{Ni}_{\text{Mg}}$  at the steps decreases it further to values very close to the experimental ones. We also find that the calculated adsorption energy of  $\text{CO}_2$  at the MgO (110) and at the O-terminated octopolar reconstruction of the polar MgO (111) surface are close to the experimental results, strongly evidencing the presence of  $\text{Ni}_{\text{Mg}}$  at the surfaces. Thus, Ni doping and  $\text{CO}_2$  adsorption promote the formation of stepped (110) and reconstructed (111) MgO surfaces. After a dedicated sample pre-treatment, the adsorbed  $\text{CO}_2$  capacity measured by microcalorimetry strongly correlates with the amount of Ni in MgO. The experimental differential heat of adsorption values of these samples are perfectly explained by the occupation of  $\text{CO}_2$  vacancies in this surface carbonate as well as adsorption at the reconstructed (111) surface. The significance of carbonates as surface species and reaction intermediates are experimentally confirmed by *in situ* IR studies.

### References

1. J. Heyd, G.E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003); J. Chem. Phys. **124**, 219906 (2006).
2. A. Tkatchenko, R.A. DiStasio Jr., R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).
3. A. Mazheika and S.V. Levchenko, J. Phys. Chem. C **120**, 26934 (2017).

<sup>a</sup> Also at: University of California at Santa Barbara, CA, USA

## Compressed-Sensing Analysis of Data for Carbon-Dioxide Activation at Metal-Oxide and Metal-Carbide Surfaces

Aliaksei Mazheika, Yanggang Wang, Runhai Ouyang, Rosendo Valero<sup>a</sup>,  
Christian Kunkel<sup>a</sup>, Helena Muñoz Galan<sup>a</sup>, Luca M. Ghiringhelli, Francesc Illas<sup>a</sup>,  
Sergey V. Levchenko, and Matthias Scheffler<sup>b</sup>

CO<sub>2</sub> conversion to hydrocarbon fuels would help preventing uncontrolled climate changes, and provide a way to utilize energy from alternative sources. Existing catalytic systems for CO<sub>2</sub> activation contain rare or precious metals, so that their utilization at large scales is prohibited by high cost. Thus, there is an urgent need for finding new materials that do not contain expensive components, but have high stability, catalytic activity and selectivity for CO<sub>2</sub> conversion.

In this work, we employ a compressed-sensing approach to find *ab initio* descriptors for a very fast but reliable evaluation of the CO<sub>2</sub> adsorption geometry and energy on any surface of any metal carbide or oxide material. The O-C-O bending angle of an adsorbed CO<sub>2</sub> is an indicator of the activation<sup>1</sup>, while the adsorption energy is an indicator of the stability of a given surface to carbonation. Sure independence screening (SIS) combined with  $l_0$  minimization<sup>2</sup> is used to find physically meaningful descriptors for CO<sub>2</sub> activation. First, we calculate 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) using DFT. The accuracy of DFT functionals is tested by comparing measured and calculated adsorption energies and work function changes upon CO<sub>2</sub> adsorption. The primary features (properties of materials that are much easier to evaluate than the properties of the adsorbed CO<sub>2</sub>) include atom, bulk, and clean surface properties. The features are used to generate about 10<sup>13</sup> non-linear combinations (candidate descriptors), and SIS+ $l_0$  is employed to select the best 1D, 2D, 3D, etc., descriptors (i.e., single features, pairs, triples, etc.) from these combinations.

We find that DOS features (including width and second moment), atomic orbital radii, polarizabilities, and  $C_6$  coefficients enter the obtained 1D to 3D descriptors for the O-C-O angle. For oxides, another very important primary feature is the electrostatic potential above the O atom at which CO<sub>2</sub> is adsorbed. Adsorption energy descriptors in addition include  $p$ -band center, work function, and ionic (Hirshfeld) charges. Interestingly, ionic charges do not appear in the descriptors for the O-C-O angle. This is consistent with an observation that CO<sub>2</sub> adsorption leads to a localization of electronic density at the surface O atom, which decouples the final atomic structure from the charge of the surface O atom before adsorption. All data obtained in this study are available via the NOMAD repository<sup>3</sup>.

### References

1. H.-J. Freund and M.W. Roberts, Surf. Sci. Rep. **25**, 225 (1996).
2. R. Ouyang *et al.*, arXiv:1710.03319 [cond-mat.mtrl-sci].
3. <https://repository.nomad-coe.eu/>.

<sup>a</sup> Departament de Ciència de Materials i Química Física & IQTCUB, Universitat de Barcelona, Spain

<sup>b</sup> Also at: University of California at Santa Barbara, CA, USA

## The Crucial Role of Many-Body Dispersion and Temperature Effects in the (Meta)Stability of Gold Clusters ( $\text{Au}_n$ , $n = 5\text{--}14$ )

Bryan R. Goldsmith<sup>a</sup>, Philipp Gruene, Jonathan T. Lyon<sup>b</sup>, David M. Rayner<sup>c</sup>,  
André Fielicke, Matthias Scheffler, and Luca M. Ghiringhelli

Gold clusters in the gas phase exhibit a diversity of structures depending on cluster size and charge state at finite temperature<sup>1,2</sup>. Small gold clusters often adopt stable planar isomers, which may partially be a consequence of relativistic effects and 5d-6s orbital hybridization<sup>3</sup>. The critical size when gold clusters begin favoring nonplanar (3D) isomers over planar (2D) isomers has attracted sustained interest<sup>4</sup>, but the critical transition size has yet to be experimentally determined for neutral gold clusters, although it is now believed to be between 10 and 14 atoms at 0 kelvin based on *ab initio* predictions. However, the critical transition size likely depends on temperature, but this effect has not been closely examined. Additionally, van der Waals interactions have either been partially or fully neglected in most computational studies of gold clusters, but this interaction is important for correctly predicting gold cluster isomer stability due to its large polarizability.

In this poster, the (meta)stable structures of gas-phase gold clusters ( $\text{Au}_n$ ,  $n = 5\text{--}14$ ) at finite temperature are discerned based on free-energy estimates using replica-exchange *ab initio* molecular dynamics with the PBE exchange-correlation functional corrected for many-body dispersion (PBE+MBD). The many-body dispersion (MBD) is just the long-range correlation energy calculated from coupled atomic response functions. The PBE+MBD methodology for predicting cluster isomer energetics is shown to be accurate based on benchmark comparisons with HSE06+MBD or the random phase approximation (using PBE orbitals). Relative free energies of  $\text{Au}_n$  isomers, discriminated based on their distribution of Au-Au coordination numbers and radius of gyration, are accurately estimated using a replica-exchange implementation of the Multistate Bennett Acceptance Ratio<sup>5</sup>. Many-body dispersion interactions are found to be critical for predicting the correct isomer energetics of neutral gold clusters, and it stabilizes 3D clusters relative to 2D clusters (up to 0.5 eV). Far-infrared multiple photon dissociation spectroscopy and simulated IR spectra (including anharmonicity) demonstrate that  $\text{Au}_9$  and  $\text{Au}_{10}$  both adopt planar structures, whereas  $\text{Au}_{12}$  adopts a nonplanar structure at 100 K. Free energy predictions identify a 3D ground state structure of  $\text{Au}_{11}$ , which suggests that the critical transition size is 11 around 100 K. However, temperature effects give rise to significant populations of 3D isomers even below size 11 at 300 K, indicating that the 2D to 3D transition is gradual and temperature dependent.

### References

1. X. Xing, B. Yoon, U. Landman, and J.H. Parks, *Phys. Rev. B* **74**, 165423 (2006).
2. E.S. Kryachko and F. Remacle, *Int. J. Quant. Chem.* **107**, 2922 (2007).
3. H. Häkkinen, M. Moseler, and U. Landman, *Phys. Rev. Lett.* **89**, 033401 (2002).
4. M.P. Johansson, I. Warnke, A. Le, and F. Furche, *J. Phys. Chem. C* **118**, 29370 (2014).
5. M.R. Shirts and J.D. Chodera, *J. Chem. Phys.* **129**, 129105 (2008).

<sup>a</sup> Present address: University of Michigan, Department of Chemical Engineering, Ann Arbor, MI, USA

<sup>b</sup> Department of Natural Sciences, Clayton State University, Morrow, GA, USA

<sup>c</sup> National Research Council, Ottawa, Ontario, Canada

## Structure and Electronic Properties of Diverse (Bio)organic Molecules in Hybrid Inorganic/Organic Systems\*

Carsten Baldauf, Volker Blum<sup>a</sup>, Svenja Janke, Sergey V. Levchenko, Dmitrii Maksimov, Haiyuan Wang, Matthias Scheffler, and Mariana Rossi

Hybrid inorganic-organic materials are desirable for the use in technological and medicinal applications, in particular since organic and bio-inspired systems present a large range of functionalities, are abundant, cheap, and comparably environment friendly. Theoretical modeling of the electronic structure of these systems is challenging, because the results can be very sensitive to the choice of the exchange-correlation approximation and the high number of possible adsorbate geometries. We here present first-principles structure searches and electronic properties of organic systems ranging from rigid aromatic molecules to flexible bio-organic peptides adsorbed on both semiconductor [H-Si(111)] and metallic [Cu(111)] surfaces.

We consider F4-TCNQ, F6-TCNNQ, which are interesting for electronic doping, and tetracene (Tc) and pentacene (Pc) which are interesting for singlet exciton fission, adsorbed on H-Si(111). Internal degrees of freedom do not play a significant role in these systems and we explore adsorption under different coverage conditions with random structure searches. Generalized gradient approximation functionals yield the same qualitative level alignment compared to hybrid functionals, even if they differ quantitatively. F4-TCNQ and F6-TCNNQ prefer to lie flat on the surface for low coverages and the LUMO of the molecules lies slightly below valence band maximum (VBM) of H-Si(111) in our level of theory, causing a work function increase and a charge transfer from the surface to the molecule. In Tc and Pc the LUMO lies above the VBM, allowing, at best, little charge transfer to occur. We disentangle contributions to work-function changes by calculating core level shifts, the space charge layer, and charge transfer, also taking level-broadening due to temperature effects into account.

For the highly-flexible peptides HisProPheH<sup>+</sup> and ArgH<sup>+</sup>, which are known from experiment to self-assemble into dimers or hexamers on Cu(111), we tackle the structure-search problem by extending the Fafoom first-principles genetic algorithm package that works on internal degrees of freedom in order to include the position and orientation of molecules with respect to the surface. By comparing two protonation states (Arg and ArgH<sup>+</sup>) of an amino acid in the gas phase, we conclude that the charge reduces the size of the accessible conformational space: while Arg presents several isoenergetic alternative conformations, ArgH<sup>+</sup> presents a single energetically-privileged minimum structure. However, when ArgH<sup>+</sup> gets in contact with the Cu(111) surface, 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, even though it still retains one well defined minimum. This efficient screening of a peptide's charge is likely to play a role in the experimentally-observed self-assembly.

\* Work performed in collaboration with Michele Ceriotti (EPFL), Stephan Rauschenbach and Klaus Kern (MPI-FKF), as well as the group of Norbert Koch (HU Berlin)

<sup>a</sup> MEMS Department, Duke University, Durham, NC, USA

## Peptide-Cation Systems: Observing Kinetically Trapped Liquid-State Conformers in the Gas Phase and How to Improve the Force Field Description with Machine Learning\*

Markus Schneider and Carsten Baldauf

Metal cations are essential to life, *e.g.* by shaping the three-dimensional structure of proteins. One-third of the proteins in the human body require a metal cofactor for biological function<sup>1</sup>. Peptides in complex with cations undergo significant conformational changes that may alter important properties. As an example, we investigate the peptide AcPheAla<sub>5</sub>LysH<sup>+</sup>, a model system for studying helix formation in the gas phase. To address the question of whether the local fixation of the positive charge at the peptide's C-terminus is a prerequisite for forming helices, we replace the protonated C-terminal Lys residue by Ala and a sodium cation. The combination of molecular simulations based on DFT with experimentally observed conformer-selective vibrational spectra in the gas phase at 10 K allows for detailed structure elucidation<sup>2</sup>. For sodiated AcPheAla<sub>6</sub>, we find globular rather than helical structures, as the mobile positive charge strongly interacts with the peptide backbone and disrupts secondary-structure formation. Interestingly, the global minimum structure from simulation is not present in the experiment. Following a rigorous experiment-theory comparison, we interpret that this is due to high barriers involved in re-arranging the peptide-cation interaction in isolation that ultimately result in kinetically trapped structures being observed in the experiment.

In order to yield accurate conformational energies and vibrational spectra of peptide-cation systems, one needs to rely on computationally costly hybrid density-functional approximations. On the other hand, the accuracy of computationally cheaper empirical simulation frameworks, *i.e.* force fields (FFs), is anything but clear and, in fact, can be grossly misleading. An intent for a solution could be the machine learning approach presented here. In essence, torsional parameters and van-der-Waals parameters in the potential-energy function  $E_{\text{pot}}$  of a particular FF, here OPLS-AA<sup>3</sup>, are “tuned” by fitting  $E_{\text{pot}}$  using ridge regression<sup>4,5</sup> and/or LASSO regression<sup>6</sup> against higher-level energies from DFT calculations for a set of conformers. For the system of AcAla<sub>2</sub>NMe, this approach reduces the mean absolute error between FF and DFT energies from 2.5 to 0.9 kcal/mol, hence resulting in an energetic description within *chemical accuracy*.

### References

1. E.A. Permyakov, Metalloproteomics. John Wiley & Sons, 2009.
2. A. Svendsen *et al.*, Rev. Sci. Instrum. **81**, 073107 (2010).
3. W.L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).
4. A.N. Tikhonov, V.Y. Arsenin, Solution of Ill-posed Problems. Washington: Winston & Sons, 1977.
5. A.E. Hoerl and R.W. Kennard, Technometrics **12**, 55 (1970).
6. R. Tibshirani, J.R. Stat. Soc. B **58**, 267 (1996).

\*Work performed in collaboration with Thomas Rizzo and Chiara Masellis (EPFL)

