



impress2017.aalto.fi

IMPRESS

Aalto University, Helsinki

7-9 June 2017



Interface Morphology Prediction with Robust and Efficient Structure Search





Opening Words

Determining or predicting the structure of organic ensembles on surfaces is a challenging problem that occupies basic science and engineering alike. Recently, novel machine-learning approaches have started to compete with more traditional, stochastic methods, such as basin hopping or simulated annealing. The objective of the present workshop is to bring the machine learning community and the traditional structure search community together, so that both can learn and benefit from each other's experiences.

We are very happy about the response and welcome all participants to Aalto University. We look forward to many stimulating talks and fruitful debates. We wish all of you a pleasant stay in Helsinki.

The Organizers: Milica, Oliver, and Patrick.



This workshop would not have been possible without the generous funding from various sources, including CECAM, Psi-k, the Academy of Finland, and NOMAD.

Conference Schedule

	Wednesday	Thursday	Friday
9:00	Registration/Coffee	Registration/Coffee	Reilly
9:30	Marom	Ghiringhelli	Discussion Session
–	Klapp	Laasonen	
11:00	Maurer	Sierka	
11:00-11:30	Coffee break		
11:30	Packwood	Karttunen	Schusteritsch
-	Scherbela	Jørgensen	Hirvonen
12:30	Todorović	Chen	Kloppenburg
12:30	Lunch	Lunch	Lunch
-			
14:00			
14:00	Discussion Session	Discussion Session	
-			
15:30			
15:30-16:00	Coffee break		
16:00	Blum	Goedecker	
-	Miao	Ma	
17:30	Oberhofer	Peterson	
17:30	Poster Session (catered)		
-			
19:30			
19:30	Welcome event (catered)	Conference dinner	

Effect of Crystal Packing on the Electronic Properties of Molecular Crystals

Noa Marom

Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh PA 15213

Molecular crystals have applications in nonlinear optics, organic electronics, and particularly in pharmaceuticals, as most drugs are marketed in the form of crystals of the pharmaceutically active ingredient. Molecular crystals are bound by dispersion (van der Waals) interactions, whose weak nature generates potential energy landscapes with many shallow minima that are close in energy. As a result, molecular crystals often exhibit polymorphism, the ability of the same molecule to crystallize in several structures. Crystal structure may profoundly influence the bioavailability, toxicity, manufacturability, and stability of drugs. In the context of technological applications, crystal structure affects the electronic and optical properties.

We perform large scale quantum mechanical simulations to predict the structure of molecular crystals and investigate the effect of crystal packing on their electronic and optical properties. The massively parallel genetic algorithm (GA) package, GAtor, relies on the evolutionary principle of survival of the fittest to find low-energy crystal structures of a given molecule. Dispersion-inclusive density functional theory (DFT) is used for structural relaxation and accurate energy evaluations¹. The structure generation package, Genarris, performs fast screening of randomly generated structures with a Harris approximation, whereby the molecular crystal density is constructed by replicating the single molecule density, which is calculated only once. Many-body perturbation theory, within the GW approximation and the Bethe-Salpeter equation (BSE), is then employed to describe properties derived from charged and neutral excitations.

For tricyano-1,4-dithiino[c]-isothiazole (TCS3), we propose a layered crystal structure, which is only 20 meV higher in energy than the observed cyclic dimer structure, and may exhibit a smaller gap, better transport properties, and broader optical absorption². For rubrene, we predict that a lesser known monoclinic crystal structure may exhibit higher singlet fission efficiency than the orthorhombic form, possibly rivaling that of pentacene³. Thus, the electronic and optical properties of organic semiconductors may be optimized by modifying the crystal packing.

¹N. Marom, R. A. DiStasio, Jr., V. Atalla, S. Levchenko, A.M. Reilly, J. R. Chelikowsky, L. Leiserowitz, A. Tkatchenko, *Angew. Chem. Int. Ed.* 125, 6761 (2013)

²F. S. Curtis, X. Wang, and N. Marom, *Acta Cryst. B* 72, 562 (2016)

³X. Wang, T. Garcia, S. Monaco, B. Schatschneider, N. Marom, *CrystEngComm* 18, 7353 (2016)

Coarse-grained modelling of the structure, phase behavior and non-equilibrium surface growth of anisotropic molecules

Sabine H. L. Klapp

Institute of Theoretical Physics, Technical University Berlin, Berlin, Germany

Anisotropic pair interactions of steric, chemical or electrostatic origin are known to strongly influence a molecular system's equilibrium phase behavior, self-assembly, and non-equilibrium surface growth. However, incorporating these interactions in an efficient manner in many-particle computer simulations at finite temperatures still poses a major challenge. In this talk I will present some recent results on coarse-graining and simulations of anisotropic systems, focusing on two aspects. First, I discuss the capability of coarse-grained potentials, which have been rigorously for a pair system (of coronene molecules), to predict the structure and phase behavior at high densities and strong coupling conditions¹. Second, I present (kinetic Monte Carlo) results for the sub-monolayer growth of a hybrid system of p-sexiphenyl (6P) molecules at a zinc oxide (1010) surface². These have been obtained on the basis of a heuristically motivated, anisotropic model which does contain information from DFT calculations. I will close with more general remarks on the impact of molecular size on surface growth³.

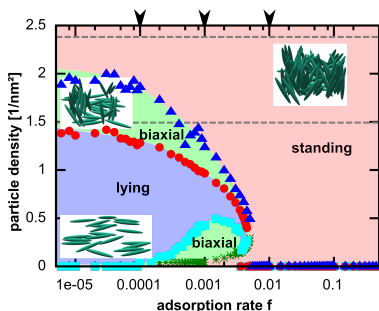


Fig. 1: Non-equilibrium state diagram for the growth of 6P molecules on a patterned substrate.

¹T. Heinemann, K. Palczynski, J. Dzubiella, and S. H. L. Klapp, J. Chem. Phys. **143**, 174110 (2015); J. Chem. Phys. **141**, 214110 (2014).

²N. Kleppmann and S. H. L. Klapp, Phys. Rev. B **94**, 241404(R) (2016).

³N. Kleppmann, F. Schreiber, and S. H. L. Klapp, Phys. Rev. E **95**, 020801(R) (2017).

Collective curvilinear coordinates in interface structure and function prediction

Reinhard Maurer

Department of Chemistry, Yale University, New Haven CT 06520, USA

Identification of relevant structures and reaction pathways in ever more complex hybrid organicinorganic interfaces and nanostructures poses a central challenge to computational materials science. Efficient, unbiased global structure prediction, tailored to identify chemically relevant domains of compound space, is an important tool to achieve rational materials design. In this talk, I will discuss a class of curvilinear, internal coordinate systems with a long history of application in chemistry in a different perspective, namely global materials structure search.¹ At hand of examples, I will show that curvilinear topologies provide the means to efficient structure determination of clusters, complex covalent molecules, and hybrid organic-inorganic interfaces,² while at the same time requiring only minor modification of existing stochastic search algorithms. By combining this smarter description of the potential energy landscape with more efficient, approximate electronic-structure methods,³ we are able to study interface structure and function at hitherto unfeasible length and time scales.⁴

¹Panosetti, Krautgasser, Palagin, Reuter, Maurer, Nano Letters 15, 8044-8048 (2015)

²Krautgasser, Panosetti, Palagin, Reuter, Maurer, J. Chem. Phys. 145, 084117 (2016)

³Stöhr, Michelitsch, Tully, Reuter, Maurer, J. Chem. Phys. 144, 151101 (2016)

⁴Lloyd et al., Nano Letters 16, 1884-1889 (2016)

Machine learning and dissimilarity analysis for surface-assisted molecular self-assembly

Daniel M. Packwood

*Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University, Kyoto, Japan
Japan Science and Technology Agency (PRESTO), Tokyo, Japan*

Bottom-up fabrication refers to the spontaneous formation of new nanomaterials via self-assembly of molecule precursors. In order to successfully assemble a specific nanomaterial via bottom-up fabrication, precursor molecules that interact and align correctly with each other during the self-assembly process must be first predicted. This presentation will introduce a new method for modeling the self-assembly of organic precursor molecules on metal surfaces. This method uses an Ising-like model, in which the energy function is constructed from first-principles data *via* machine learning. Markov chain Monte Carlo is then used to find the equilibrium states of the model ^{1 2}. Finally, the application of a dissimilarity analysis technique to the model will be described. This analysis quantifies how self-assembly depends upon the chemical properties of the precursor molecule, and can help identify appropriate precursor molecules for bottom-up fabrication of target nanomaterials.

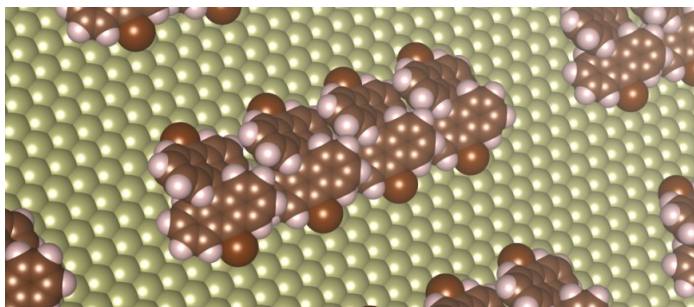


Fig. 1: Prediction of chain-shaped structures from self-assembly of dibromo-bianthracene precursors on copper(111)

¹D. M. Packwood, *et al.* Nat. Commun. 8, 14463 (2017)

²D. M. Packwood, *et al.* Roy. Soc. Open Sci. 3, 150681 (2016)

Structure Prediction at Organic/Inorganic Interfaces using Gaussian Process Regression

M. Scherbela, V. Obersteiner, L. Hörmann and O.T. Hofmann

Institute of Solid State Physics, Graz University of Technology, Austria

A major step towards designing new materials is understanding and predicting the crystal structures that form at interfaces. Finding thermodynamically stable phases could in principle be done by calculating the enthalpy of all possible structures. However, the rich polymorphism exhibited by these systems prohibits an exhaustive search for the global minimum using expensive ab-initio calculations. This configurational explosion is particularly problematic for crystal structures that contain many molecules per unit cell, due to the many degrees of freedom. In this work it is shown on the example of tetracyanoethylene (TCNE) on Ag(100) that this challenge can be tackled with a combination of coarse-graining and machine learning:

First the adsorption geometries that isolated molecules would adopt on the substrate are found. Supercells with multiple molecules per unit cell are built by combining these isolated adsorption geometries to generate a set of possible “guess polymorphs”. This discretizes the configurational space to a finite, but still huge size. Using optimal design methods, a small, representative subset is selected and its energies are calculated using Density Functional Theory (DFT). By training a Gaussian Process Regression model on these DFT calculations, effective interactions between adsorbates are obtained. These provide an efficient and accurate energy prediction for all remaining guess polymorphs which is exploited by sampling the energetically most promising structures and iterated relearning.

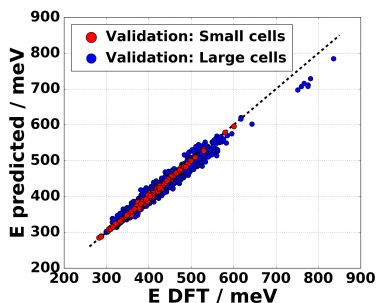


Fig. 1: Prediction for small and large unit cells after training only on small ones.

Using this method we explore the potential energy landscape of TCNE on Ag(100) and show that the machine learning model can achieve the accuracy of the expensive DFT calculations after having been trained only on a small fraction of the total dataset. Furthermore we are able to extrapolate energies of large unit cells by training the model on smaller, cheaper unit cells.

Efficient Bayesian Inference of Surface Adsorbate Structure

M. Todorović¹, M. U. Gutmann², J. Corander³ and P. Rinke¹

¹ *Department of Applied Physics, Aalto University, Espoo, Finland*

² *School of Informatics, University of Edinburgh, Edinburgh, UK*

³ *Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway*

The adsorption and self-organisation of molecules at inorganic surfaces is central to many industrial processes from catalysis and coatings, to organic electronics and solar cells. Computer simulations can help identify interface morphology and functionality, but sampling many atomic configurations over large length scales is prohibitively costly. We combined Bayesian optimisation¹ with accurate atomistic simulations in our efficient structure search tool BOSS, designed for intelligent probabilistic sampling of atomic configurations. The nearly parameter-free framework relies on Gaussian processes (GPs) to construct a probable potential energy surface (PES), which is then iteratively refined by input of energy data points from selected configurations.

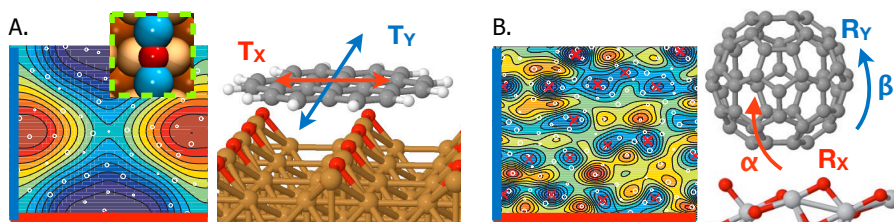


Fig. 1: BOSS application to A) adsorption registry of coronene/Cu(110)-O p(2x1), and B) molecular orientation of fullerene/TiO₂(101).

The BOSS framework was employed in up to six dimensions to identify the optimal adsorption structures of large organic molecules on functional oxide substrates (see Fig. 1). We report a dramatic speed-up in identifying optimal configurations, compared to the traditional chemical intuition technique, without significant loss of accuracy. Thanks to the clever Bayesian sampling scheme (balancing exploitation and exploration steps) and a streamlined "building block" approach to molecular structure, even complex interface problems can be solved for relevant global and local minima structures.

¹M. U. Gutmann and J. Corander, J. Mach. Learn. Res. 17, 1 (2016)

The Material that Wasn't There. Or, Would the Real Atomic Structure Causing Some Exciting Property Please Stand Up?

Volker Blum

Department of Mechanical Engineering and Materials Science and Department of Chemistry,
Duke University, Durham, NC 27708, USA

In 2009, a now-seminal paper¹ demonstrated that polymeric carbon nitride (PCN) materials facilitate the photocatalytic evolution of hydrogen. Ever since, there has been a run on these materials, particularly on a variety referred to as graphitic carbon nitride, $g\text{-C}_3\text{N}_4$ (rightmost structure in Figure 1). However, despite roughly thousands of publications assuming its existence, to date it remains unclear if $g\text{-C}_3\text{N}_4$ was ever actually made. In experiments, the nanocrystalline nature of PCN materials hinders their unambiguous atomic-scale assignment. Yet, a simple *ab initio* thermodynamic analysis² shows that the true PCN materials made in experiment likely contain hydrogen and have a different structure than the supposed $g\text{-C}_3\text{N}_4$. This – the identification of the nanoscale preferred atomic structure(s) governing key properties of materials and their interfaces – is where careful theory and modeling can truly make a difference to help understand what determines a materials' properties, and how we might improve them.

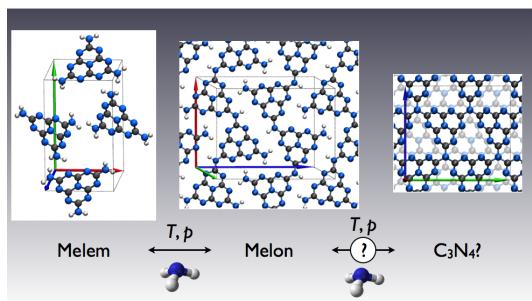


Fig. 1: Sketch of some phase equilibria in polymeric carbon nitrides.

This talk briefly summarizes the state of, and recent developments in, the FHI-aims electronic structure code,³ the primary tool used and developed in our group and by a global community of users and developers to address structural and electronic properties of molecules and materials across the periodic table, up to scales of thousands of atoms and with high numerical accuracy and reliability. The talk then focuses on the story of PCN materials, their *ab initio* thermodynamics, and of pathways to improve their performance. Finally, we address pathways for *ab initio* random structure searching of other *a priori* unknown molecular and interface systems with FHI-aims.

¹X.Wang *et al.*, Nature Materials 8, 76 (2009)

²T. Botari *et al.*, Chem. Mater. (2017) 10.1021/acs.chemmater.7b00965

³V. Blum *et al.*, Comp. Phys. Commun. 180, 2175 (2009)

Activating inert electrons and idle orbitals by high pressure

Maosheng Miao^{1,2}

¹ *Department of Chemistry and Biochemistry, California State University Northridge, CA, USA*

² *Beijing Computational Science Research Center, Beijing, China*

The chemistry at ambient condition has implicit boundaries rooted in the atomic shell structure: the inner-shell electrons and the unoccupied outer-shell orbitals do not involve as major component in chemical reactions and in chemical bonds. The chemical properties of atoms are determined by the electrons in the outermost shell, namely the valence electrons. These general rules govern our understanding of chemical structures and reactions.

Using first principles calculations and automatic crystal structure search method, we demonstrate that under high pressure, the above doctrines can be broken. We show that both the inert electrons and the idle orbitals can involve in chemical reactions. In the presence of fluorine and under pressure, the formation of CsFn ($n > 1$) compounds containing neutral or ionic molecules is predicted.¹ We also found that under high pressure Hg in Hg-F compounds transfers charge from the d orbitals to the F, thus behaving as a transition metal.² The Hg(IV) compound consists of HgF₄ planar molecules. HgF₃ is metallic and ferromagnetic, with a large gap between its partially occupied and unoccupied bands under high pressure.

Recently, it is found that, under high pressure, alkali metals such as Li and Na become semiconducting or insulating. As they do so, they adopt structures containing sites that accommodate electrons, leading to the formation of high-pressure electrides (HPE). Similar phenomena have also been predicted for Mg, Al and several other materials. We propose that the driving force for HPE formation can be attributed to the lack of core electrons in the interstitial sites, which causes the energies of the corresponding quantized orbitals to increase less significantly with pressure than normal atomic orbitals.³ These empty sites enclosed by surrounding atoms have been termed interstitial quasiatoms (ISQ); they may show some of the chemical features of atoms, including the potential of forming covalent bonds. Here we argue that some calculated ISQs in the high-pressure semiconducting Li phase (oC40, Aba2) actually form covalently bonded pairs.⁴ We suggest such quasimolecules may be found in other systems at high pressures as well.

¹M. S. Miao, *Nature Chemistry*, 5, 846 (2013).

²J. Botana, *et al.* *Angew. Chemie* 54, 9280-9283 (2015).

³M. S. Miao and R. Hoffmann, *Accounts of Chemical Research*, 47, 1311 (2014).

⁴M. S. Miao, *et al.*, *Angew. Chemie* 56, 972 (2017).

“Cheap” DFT for structure prediction, and what else it is good for

Harald Oberhofer

Chair for Theoretical Chemistry and Catalysis Research Center, Technische Universität München, Lichtenbergstr. 4, D-85747 Garching, Germany

Structure prediction of crystals of organic molecules rests—next to efficient sampling and optimisation of conformational spaces—to a large degree on the availability of cheap, yet accurate energy and force calculation protocols. The most accurate energy method is useless if it either cannot be applied to large enough systems or is so expensive that one can afford only a few calculations. Thereby, even the use of density functional theory (DFT) severely strains currently available computing resources. Moreover, standard DFT methods are of little use due to their failure to describe the dispersion forces that bind most organic crystals together. We therefore developed a simulation protocol where dispersion contributions are calculated with the recently presented many body dispersion (MBD) method of Tkatchenko and co-workers, yet based on an approximate single-molecule density akin to the well-known Harris functional.

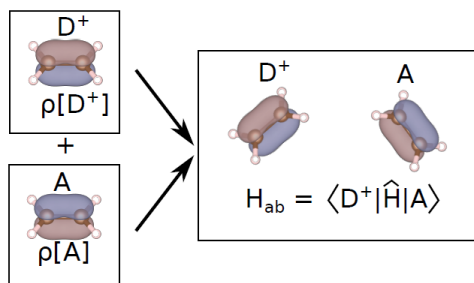


Fig. 1: Sketch of a Harris-functional like approach to determine electronic couplings H_{ab} between a donor and an acceptor molecule.

In our contribution we not only discuss this approach and the technical intricacies involved in its usage, but also other uses of the machinery developed for Harris-like DFT calculations. The main one of these, so called fragment orbital DFT (FO-DFT) yields electronic couplings between the molecules of the crystal which can then be employed to model e.g. charge transport properties of the sampled crystal structures.

Learning descriptors from materials-science (big) data

Luca Ghiringhelli

Fritz-Haber Institut of the Max-Planck Society, Faradayweg 4-6, 14195 Berlin, Germany

Scientific discoveries often proceed from the accumulation of consistent data to the identification of functional dependencies among the data, i.e., a model that is able to predict yet unseen phenomena. Ultimately, a theory may be constructed to explain the model with few simple principles. Classical examples are i) the three laws of Kepler, that were empirically found by observing the known data on the solar system, later justified by Newton's theory of gravitation, and ii) the periodic table of Mendeleev, empirically constructed from data on the chemistry of known elements, later justified by the atomic theory within quantum mechanics. In the last decades, statistical learning has been developed in order to find optimal and stable functional dependencies among data, in particular when some ancillary knowledge can be formalized and included in the search for optimal solutions.

We present a recently introduced compressed-sensing based methodology and its latest extension, for the identification of functional dependencies where the descriptor (the set of input variables of the functional dependence) is selected out of a dictionary of "well formed" candidate analytical expressions. Such candidates are constructed as non-linear functions of a set of basic "physically meaningful" features, called primary features. Furthermore, we present a complementary method, called subgroup discovery (SGD), designed for constructing statements, in the form of true/false boolean expressions, about an optimal subset of candidate functions of primary features.

Results from the application of both methods are presented for the crystal structure prediction of binary materials and (only for SGD) for the identification of relationships between electronic- and atomic-structure properties of metal nanoclusters.

Modeling of complex metal alloys and nanoparticles via multiscale first-principles methods

Kari Laasonen¹, and Juhani Teeriniemi¹

¹ *Department of Chemistry and Material Science, Aalto University, Finland*

This work is related to DFT based modelling of phase diagrams of complex metal alloys. The thermodynamic phase diagrams serve as road-maps and illustrate the state of a material as a function of temperature, concentrations, components, and pressure. In particular, the DFT can provide insight to microscopic features that are manifested as macroscopic properties of a material but the DFT calculations are very time consuming and usually not suitable for finite temperature simulations. We have developed a Cluster Expansion (CE) model with heuristic effective cluster interactions and Monte Carlo simulations have been used to determine the finite temperature configurational phase diagrams for several alloys^{1,2}. Machine learning methods have been used in the training of the CE parameters from the DFT data. The modelling of finite temperature properties provides substantially more information and knowledge than traditional CE studies, which are typically restricted to searching possible ordered ground-state structures at 0 K. We have recently apply the same DFT-CE method to nanosize FeNi clusters³.

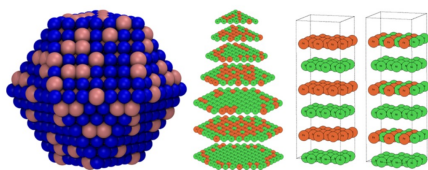


Fig. 1: Structure of a $\text{Fe}_{323}\text{Ni}_{600}$ cluster.

¹J. Teeriniemi, P. Taskinen, K. Laasonen, First-principles investigation of the Cu-Ni, Cu-Pd, and Ni-Pd binary alloy systems, *Intermetallics* 57 (2015) 41-50

²J. Teeriniemi, P. Taskinen and K. Laasonen, Modeling of complex ternary structures: Cu-Ni-Pd alloys via first-principles, *J. Comp. Mater. Sci.* 115, 2016, 202-207

³J. Teeriniemi, M. Melander, S. Lipasti, R. Hatz, and K. Laasonen, Fe-Ni Nanoparticles: A Multiscale First-Principles Study to Predict Geometry, Structure, and Catalytic Activity, *J. Phys. Chem. C* 2017, 121, 1667-1674

Atomic Structure of Oxide Materials: Synergy Between Theory and Experiment

M. Sierka

Otto Schott Institute of Materials Research, Friedrich Schiller University of Jena, Germany

Low-dimensional metal oxides in form of clusters and nanoparticles often exhibit unique properties very different from those of their atomic and bulk counterparts. This opens new possibilities for fabrication of highly functional materials with tailored properties. However, the prerequisite for the development of such materials is a detailed characterization of the low-dimensional oxides at the atomic level. Due to the limitations of both theory and experiments this task can very often only be achieved combining computer simulations and experimental investigations.¹

The examples presented in my talk include joined theoretical and experimental studies of new building blocks for ZnO and CdO based cluster-assembled materials,² structure evolution of nanoparticulate Fe₂O₃ starting from small clusters through to nanometer-sized particles³ as well as investigations of structure and phase transitions of novel negative thermal expansion materials. In all cases the full characterization of these systems could only be accomplished combining experiments and global structure optimizations at the density functional theory level or molecular dynamics simulations employing tailored, ab initio parameterized interatomic potential functions.

¹M. Sierka In: Computational Modelling of Inorganic Nanomaterials; S. T. Bromley, M. Zwijnenburg (Eds); CRC Press, 335-367 (2016)

²R. Lazarski, M. Sierka, J. Heinzelmann, A. Koop, R. Sedlak, S. Proch, G. F. Ganteför J. Phys. Chem. C 119, 6886 (2015); J. Heinzelmann, A. Koop, S. Proch, G. Ganteför, R. Lazarski, M. Sierka J. Phys. Chem. Lett. 5, 2642 (2014)

³A. Erlebach, H.-D. Kurland, J. Grabow, F. A. Müller, M. Sierka Nanoscale 7, 2960 (2015)

Crystal structure prediction of inorganic-organic coordination polymers

A. J. Karttunen, J. Heiska, M. Nisula, J. Linnerna, and M. Karppinen

Department of Chemistry and Materials Science, Aalto University, Helsinki, Finland

Atomic/molecular layer deposition (ALD/MLD) offers a highly controllable way to synthesize crystalline inorganic-organic coordination polymer (CP) thin films.¹ The fabricated materials could also be considered as metal-organic framework thin films and they show a layered structure with inorganic-organic interfaces (Fig. 1). We have recently investigated lithium aryloxide CP thin films fabricated by ALD/MLD (dilithium 1,4-benzenediolate, $\text{Li}_2[\text{p-C}_6\text{H}_4\text{O}_2]$).² The atomic-level structure of the CP thin films was solved by combining X-Ray diffraction and infrared spectroscopic studies with Density Functional Theory (DFT) calculations. In this case, a related bulk crystal structure facilitated the structure solution of the CP thin film.

Combining 2-aminoterephthalic acid (TPA-NH₂) and a suitable lithium precursor also results in a crystalline CP thin film (Fig. 1). However, in this case, there is no previous structural information from analogous bulk materials and the structure must be solved from scratch. We used the USPEX evolutionary crystal prediction approach together with DFT calculations to predict the atomic-level structure.³ When we define the inorganic-organic coordination polymer as a pseudo-molecular structure with rigid building blocks, the USPEX structure search works in a highly efficient manner. Our results for several materials recently synthesized by ALD/MLD show that USPEX is readily applicable to the structure prediction of inorganic-organic coordination polymers.

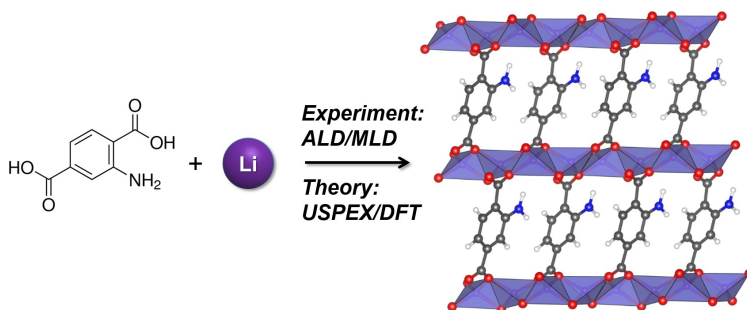


Fig. 1: The predicted structure of $\text{Li}_2\text{-TP-NH}_2$ coordination polymer (USPEX/DFT)

¹E. Ahvenniemi, M. Karppinen, *Chem. Mater.* **2016**, 28, 6260.

²M. Nisula, J. Linnerna, A. J. Karttunen, M. Karppinen, *Chem. Eur. J.* **2017**, 23, 2988.

³A. R. Oganov, A. O. Lyakhov, M. Valle *Acc. Chem. Res.* **2011**, 44, 227.

Combining Evolutionary Algorithms with Clustering toward Rational Global Structure Optimization at the Atomic Scale

Mathias S. Jørgensen, Michael N. Groves and Bjørk Hammer

Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark

Global structure optimization at the atomic scale can be a very costly task in terms of both computational resources and man-hours when done with only little *a priori* knowledge about the global minimum. Many current optimization techniques produce a vast amount of idle intermediate data before converging to a minimum. If these data could be analyzed during optimization, many new possibilities emerge for more rational search algorithms that require less human supervision and less computational resources to explore relevant regions of the search space. We combine an evolutionary algorithm (EA) with clustering, a machine-learning technique, to produce a rational algorithm where progression of the search is based on the entire configuration space of currently found structures¹ contrary to traditional EAs, where progression is blindly based on a few of the most stable structures (the population).

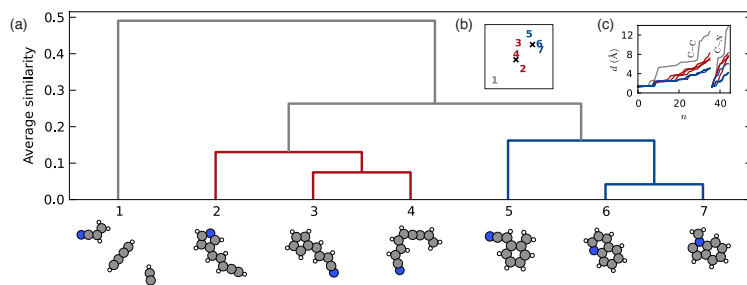


Fig. 1: Visualization of seven clustered C_9H_7N structures (a) in a dendrogram, (b) in a multidimensional scaling to two dimensions with cluster centroids marked as x, and (c) as feature plots with interatomic distances.

Clustering the configuration space of intermediate structures into regions of geometrically similar structures enables the EA to suppress certain regions and favor others. For two test systems, an organic molecule and an oxide surface, the global minimum search proves significantly faster using two clustering-enhanced algorithms. The first algorithm directly identifies and favors stable structures in unexplored regions, while the other algorithm penalizes structures in dense, explored regions. This clustering-enhanced EA is a step toward adaptive global optimization techniques that can act upon information in accumulated data.

¹M. S. Jørgensen, *et al.* J. Chem. Theory Comput. 13, 1486-1493 (2017)

Decompose local energy with artificial neural network

Xin Chen¹, Mathias S Jørgensen², Bjørk Hammer^{2,*} and Jun Li^{1,*}

¹ *Department of Chemistry and Key Laboratory of Organic Optoelectronics & Molecular Engineering of Ministry of Education, Tsinghua University, Beijing, China*

² *Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark*

In physical chemistry, it is desirable to develop interatomic potentials that allow for fast determination of dynamic and kinetic properties of molecules, clusters, and solids. We present an approach to this using an artificial neural network that is trained against a database of relaxed structures described in a density functional theory setup. The artificial neural network is designed to be atom-centered and scalable meaning that it can be trained on some stoichiometries and used on the same or other stoichiometries. Owing to its atom-centered nature, the artificial neural network yields local energy information which can be used to make rational decisions in global minimum searching algorithms like the basin-hopping (BH) algorithm¹ or the evolutionary algorithm (EA)². We explore this possibility and discuss on the basis of the success of the approach the usefulness of the locality of the energy as assigned by the network.

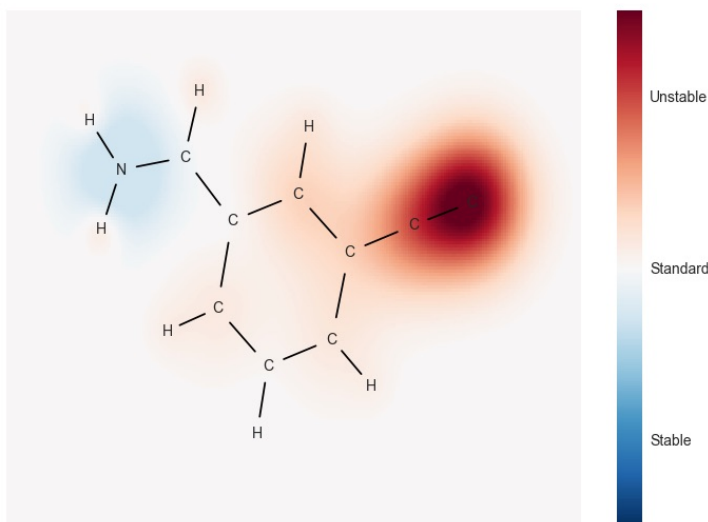


Fig. 1: The atomic heatmap of a C_9H_7N molecule. Atoms that tend to be unstable are colored red.

¹Ya-fan Zhao, *et al.* Nano. Res., Just Accepted, (2017)

²Mathias S. Jørgensen, *et al.* J. Chem. Theory Comput, 13, 1486-1493, (2017)

Structure prediction with the Minima Hopping Method

Stefan Goedecker

Department of Physics, University of Basel, Basel, Switzerland

I will present the Minima Hopping structure prediction method¹ as well as its extension that allows to find automatically a large number of saddle points² and to characterize in this way the entire potential energy surface. I will show that certain synthesizable systems share features of the potential energy surface. I will introduce a charge equilibration based neural network potential³ and present a new 2-dimensional titanium dioxide structure that was found with this potential and verified on the density functional level.⁴ I will present structure prediction results on a technologically important interface, namely the SiC-SiO₂ interface, for which we were able to determine the structure of the carbon clusters that are responsible for the deterioration of the electronic properties. I will comment on the "key challenges" that are posted on the workshop webpage.

¹S. Goedecker, JCP 120, 9911 (2004)

²B. Schaefer, S. Mohr, M. Amsler, Stefan Goedecker, JCP 140, 214102 (2014)

³S. A. Ghasemi, A. Hofstetter, S. Saha, S. Goedecker, PRB 92, 045131 (2015)

⁴H. A. Eivari et al., arXiv:1704.03983

Structure Prediction via Swarm Intelligence Algorithms

Yanming Ma

State Key Lab of Superhard Materials, College of Physics, Jilin University, Changchun 130012, China

Knowledge of atomistic structures is essential if the properties of materials are to be understood and exploited, particularly when establishing a correspondence between materials performance and their chemical compositions. Precise prediction of atomistic structures with the given information of chemical compositions is highly desirable, but it is extremely difficult as it basically involves in exploring a huge number of energy minima on the high-dimensional potential energy surface. There is an urge on development of global optimization algorithms on efficient exploration of energy surface to identify the global stable structure.

We have developed an efficient CALYPSO structure prediction approach^{1,2} via multi-objective swarm-intelligence optimization algorithms by taking the advantage of structures smart learning where both particle swarm optimization and artificial bee colony algorithms are employed (See more information at <http://www.calypso.cn>). The method has been widely used by more than 1600 users to design multi-dimensional structures ranging from 3D bulk crystals to 0D nanoclusters, 2D layers and surfaces, etc^{3,4,5}. Functionality-driven inverse design of electride, superhard, and optical materials are also feasible⁶. Recently, interface structure prediction module has been implemented in CALYPSO code. Benchmarks on known interface structures of graphene and TiO₂ have demonstrated the high efficiency of the method. Machine learning technique is equipped to construct the interatomic potentials for performing geometric optimizations and total energy calculations of sampled structures during CALYPSO structure searches. Computational costs are significantly reduced to allow a much-enhanced sampling of phase space.

¹Y. Wang, J. Lv, L. Zhu, and Y. Ma, Phys. Rev. B 82, 094116 (2010)

²Y. Wang, J. Lv, L. Zhu, and Y. Ma, Comput. Phys. Commun. 183, 2063 (2012)

³J. Lv, Y. Wang, L. Zhu, and Y. Ma, J. Chem. Phys. 137, 084104 (2012).

⁴S. Lu, Y. Wang, H. Liu, M. Miao, and Y. Ma, Nature Commun. 5, 3666 (2014)

⁵Y. Wang, et al., J. Chem. Phys. 137, 224108 (2012)

⁶X. Zhang, et al., J. Chem. Phys. 138, 114101 (2013)

Machine-learning algorithms to accelerating electronic structure calculations

Andrew Peterson

School of Engineering, Brown University, Providence, USA

Electronic-structure calculations are the workhorse of first-principles design of materials, chemicals, and reactions. However, a typical atomistic calculation involves much computational "waste": often, thousands of single-point electronic structure calculations are performed to produce each one or two configurations which are ultimately published. Recently, methods have been developed which can "learn" from the output of electronic structure calculations, allowing for the prediction of the energies and forces of atomic configurations that they have not been exposed to. In this talk, we will discuss new algorithms that can use such machine learning techniques to strategically accelerate and extend electronic structure calculations, through such routines as transition-state searches and hybrid quantum-mechanics / molecular-mechanics approaches. These methods will be introduced in terms of our open-source machine-learning package, "Amp".

Organic Crystal Structure Prediction: Blind Tests, Progress and Challenges

A.M. Reilly

School of Chemical Sciences, Dublin City University, Glasnevin, Dublin 9, Ireland

Organic molecules can crystallise in a wide range of potential crystal structures or solid forms. Each of these can have very different physical properties that impact the use of the molecule. For example, different solid forms of an active pharmaceutical ingredient can have different apparent solubilities, while different crystal morphologies can make some solid forms easier to process than others. The need to understand and predict such behaviour has led to organic crystal structure prediction (CSP) methods developing significantly over the past 20 years, with progress in the field being charted by series of blind tests organised by the Cambridge Crystallographic Data Centre. The most recent sixth blind test of organic CSP methods¹ tackled some of the most challenging molecules and systems to date (Fig. 1). This contribution will provide overview of the test and discuss a range of challenges that remain for organic CSP, particularly, accurate ranking and capturing conformational flexibility. With computational costs remaining prohibitively expensive for routine application of CSP, the potential for experimental information in the Cambridge Structural Database to speed up and direct CSP calculations will also be briefly discussed.

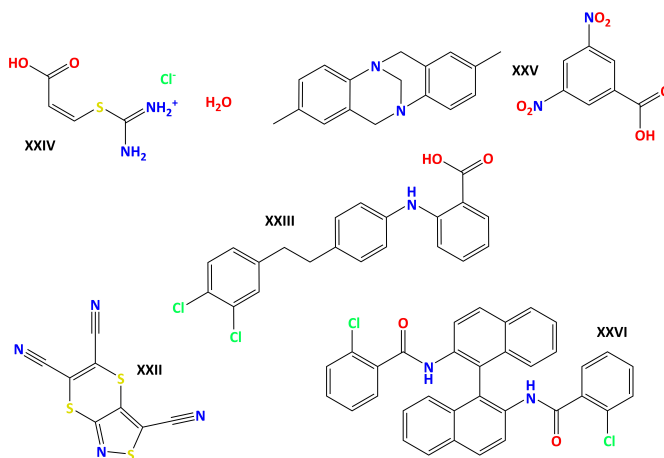


Fig. 1: The five target systems from the sixth blind test of organic CSP methods.

¹A. M. Reilly, *et al.* Acta Cryst. B 72, 493 (2016)

Predicting the structure of interfaces using the Ab Initio Random Structure Searching method

G. Schusteritsch¹ and C. J. Pickard^{1,2}

¹ *Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, U.K.*

² *Advanced Institute for Materials Research, Tohoku University, Sendai, Japan*

First-principles structure prediction of bulk materials is now routinely performed, however the field of predicting the atomic structure of interfaces from first-principles is still in its infancy. A detailed understanding of and ability to predict the atomic structure of interfaces is however of crucial importance for many technologies. Interfaces are very hard to predict due to the complicated geometries, crystal orientations and possible non-stoichiometric conditions involved and provide a major challenge to structure prediction. We present here the ab initio random structure searching (AIRSS)¹ method and how it can be used to predict the structure of interfaces². Our method relies on generating random structures in the vicinity of the interface and relaxing them within the framework of density functional theory. The method is simple, requiring only a small set of parameters, and can be efficiently run on modern parallel computer architectures. We focus here on the prediction of grain boundaries, but application to heterostructure interfaces and surfaces is equally possible. Examples of several grain boundary defects in technologically important materials will be presented, including grain boundaries in 2D materials (graphene), as well as much more complex 3D systems such as grain boundary interfaces in transition metal oxides (SrTiO_3 and TiO_2). Direct comparison to experimental results will be made.

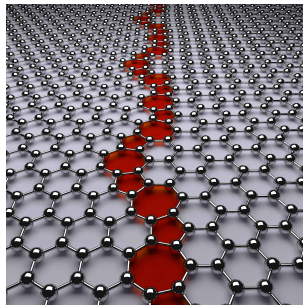


Fig. 1: Graphene grain boundary structure between armchair and zigzag graphene regions.

¹C. J. Pickard and R. J. Needs, Phys. Rev. Lett. 97, 045504 (2006)

²G. Schusteritsch and C. J. Pickard, Phys. Rev. B. 90, 035424 (2014)

Multiscale modeling of polycrystalline graphene

P. Hirvonen¹, Z. Fan¹, M.M. Ervasti¹, M. Jalalvand^{2,3}, K. Azizi^{1,2}, M. Seymour⁴, S.M. Vaez Allaei^{2,5}, N. Provatas⁴, A. Harju¹, K.R. Elder⁶, and T. Ala-Nissila^{1,7}

¹ *Department of Applied Physics, Aalto University School of Science, Espoo, Finland*

² *Department of Physics, University of Tehran, Tehran, Iran*

³ *Department of Physics, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran*

⁴ *Department of Physics, McGill University, Montreal, Canada*

⁵ *School of Physics, Institute for Research in Fundamental Sciences, Tehran, Iran*

⁶ *Department of Physics, Oakland University, Rochester, USA*

⁷ *Department of Mathematical Sciences and Department of Physics, Loughborough University, Loughborough, UK*

Defects and grain boundaries greatly influence the properties of graphene but modeling their formation is challenging due to the multiple length and time scales involved. We extend the Phase field crystal (PFC) approach¹ to quantitative modeling of defected graphene microstructures. We assess four PFC models by studying grain boundary structures and their formation energies. We compare PFC results to density functional theory (DFT) and molecular dynamics (MD). The one-mode PFC model is found to produce realistic defect topologies, whereas the three-mode model predicts quantitatively correct grain boundary energies.² We are currently studying grain boundary triple junctions, for which we find both positive and negative formation energies.

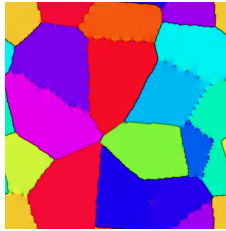


Fig. 1: Random polycrystalline system of graphene modeled with PFC. The local crystallographic orientations have been mapped to different hues.²

PFC models are able to capture the dynamics of large (poly)crystalline systems on diffusive time scales while retaining atomic resolution. We exploit these multiscale characteristics to construct large polycrystalline PFC graphene systems whose sizes and formation time scales are beyond the reach of DFT calculations and MD simulations, respectively. We use these systems as the starting point of MD simulations for investigating the thermal transport properties of polycrystalline graphene. Figure 1 demonstrates a random polycrystalline system of PFC graphene.

¹K.R. Elder *et al.*, *Phys. Rev. Lett.* **88**, 245701 (2002)

²P. Hirvonen *et al.*, *Phys. Rev. B* **94**, 035414 (2016)

The interfaces of 3×3 SiC-000 $\bar{1}$ with graphene and IrO₂ with water under applied potential

Jan Kloppenburg and Geoffroy Hautier

IMCN, Université catholique de Louvain, Louvain-la-Neuve, Belgium

3×3 SiC-000 $\bar{1}$ interface:

Silicon carbide (SiC) is a primary material candidate for large scale, high quality epitaxial graphene production. Graphene grown on the carbon face of SiC features high carrier mobilities. However, a controlled graphene growth is hindered by the precursor surface phases of graphene – the (3×3) -SiC- $(\bar{1}\bar{1}\bar{1})$ reconstruction. Despite more than two decades of research the precise atomic structure is still unclear. We solve the (3×3) -SiC- $(\bar{1}\bar{1}\bar{1})$ reconstruction by an *ab initio* random structure search (AIRSS) based on a van der Waals corrected density functional and identify its atomic structure in the Si-rich range. Simulated scanning tunneling microscopy (STM) images are in excellent agreement with previously reported experimental STM images.

IrO₂ in contact with water under applied potential:

The transition from fossil fuel based energy to renewable sources requires efficient energy conversion and storage solutions. Hydrogen and oxygen production from water is an attractive approach among the storages. The oxygen evolution reaction (OER) is sluggish and must be facilitated by a catalyst that is designed to catalyze the OER at low overpotential to maximize the energy conversion efficiency. Recently, the scientific interest has grown for oxide-based catalysts. The powerful yet-simple thermodynamic rate-limiting approach to optimize catalysts for OER using *ab initio* computed adsorption energies has shown predictive power on metals but still shows limits on oxides. A fundamental step back is required using model systems and comparing *ab initio* computed adsorption energies using density functional theory (DFT) to high quality single-crystal model catalysts. We present results of a theory-experiment comparison using cyclic voltammetry (CV) experiments on single-crystals of IrO₂ grown by molecular beam-epitaxy (MBE). We discuss the accuracy, strength and limits of DFT computations in modeling adsorption energies. We especially focus on water coverage effects under applied external potential.

Alignment of electronic energy levels with electrochemical redox potentials at semiconductor-water interfaces: from atomistic to continuum solvation models

L. Blumenthal¹, P. Tangey^{1,2} and J. Lischner²

¹ Department of Physics, Imperial College London, London, United Kingdom

² Department of Physics, Imperial College London, London, United Kingdom

Identifying efficient systems for storing energy constitutes a major challenge in the transition to sustainable and renewable energy sources. A potential solution is to store energy in a chemical form, for example, in hydrogen gas produced by photoelectrochemical (PEC) water splitting.¹ A detailed understanding of the electronic structure of photoelectrodes, in particular the alignment of the electrodes' electronic band edge positions with the relevant redox potentials of water, is required to guide experimental progress towards increased PEC device efficiencies. To address this problem, we introduce a new approach based on the combination of many-body perturbation theory within the GW method for the electronic structure of the photoelectrode and joint density functional theory (JDFT) for the description of solid-liquid interfaces.²

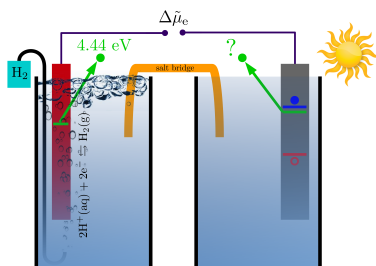


Fig. 1: Being able to predict the Fermi level of a semiconductor relative to that of the standard hydrogen electrode is essential for the *in silico* design of new photoelectrode materials.

We applied this approach to rutile (TiO_2) and determined the position of the valence and conduction band edges with respect to the redox potential of the hydrogen evolution reaction. Furthermore, we demonstrated that an explicit atomic description of the vicinity of the interface is crucial for obtaining an accurate energy level alignment when using JDFT. As the exact atomic structure of solid-liquid interfaces is often unknown, this currently limits the application of JDFT to high-throughput searches for novel photoelectrode materials. To overcome this drawback, future work will aim at identifying an efficient workflow that will enable us to predict the structure of the first monolayer of water next to a semiconductor surface.

¹H.-J. Lewerenz and L. Peter, Photoelectrochemical water splitting: materials, processes and architectures. RSC Publishing, 2013.

²S. A. Petrosyan et al. The Journal of Physical Chemistry B 109, 15436 (2005).

Structural and electronic properties of sexithiophene layers on gold

Laura Katharina Scarbath-Evers¹, Erik Schreck², René Hammer², Wolf Widdra² and Daniel Sebastiani¹

¹ *Institute of Chemistry, Martin Luther University Halle-Wittenberg, Germany*

² *Institute of Physics, Martin Luther University Halle-Wittenberg, Germany*

We investigate the layer-wise growth of α -sexithiophene on Au(100) using a hybrid QM/MM approach. In order to reduce the complexity of structure prediction we rely on experimental STM data that help us to find reasonable starting structures. α -sexithiophene (α -6T) is an organic semiconductor and serves as a model system for polythiophene which has already found application in electronic devices. An optimization of the device performance requires a fundamental understanding of the interplay between the packing structure of the molecules and their electronic properties, like HOMO-LUMO gaps. So far, research was focused on the adsorption structure of single molecules of α -6T up to one monolayer on different gold surfaces¹ and the bulk structure of α -6T² whereas the packing structure of several α -6T layers on gold has not been investigated. Our results reveal a flat adsorption for the first monolayer of α -6T on gold, whereas the molecules in the second monolayer are slightly tilted around their long molecular axis. This adopted double layer structure of α -sexithiophene significantly affects the electronic structure of the molecules which, in turn leads, to a remarkable decrease of the HOMO-LUMO gap.

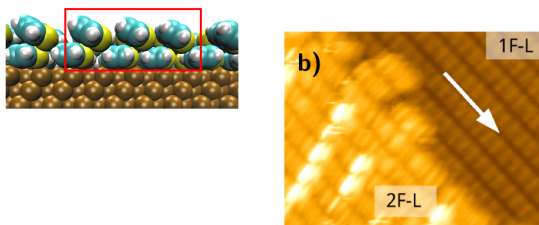


Fig. 1: **Growth of α -6T layers on Au(100).** a: Structure of the α -6T doublelayer from QM/MM calculations. b: STM image of a α -6T doublelayer on reconstructed Au(100) revealing two chiral adsorption configurations.

¹M. Kiel *et. al.* Phys. Rev. B, 75:195439, 2007

²Horowitz *et. al.*, Chemistry of Materials, 7(7):1337-1341, 1995

On the efficacy of an evolutionary algorithm

Simon Eskildsen and Bjørk Hammer

iNANO and Department of Physics and Astronomy, Aarhus University, Denmark

Evolutionary algorithms (EAs) have proven themselves successful in locating global minimum energy structures in conjunction with density functional energy expressions.^{1,2} In order to understand how well the evolutionary approach works, we have designed a model potential describing the interaction of two atomic species on a 2D grid. The 2D grid is subject to periodic boundary conditions with super cells of $N \times N$ grid points. As N is increased, the number of possible configurations of the two atomic species grow exponentially, cf. Fig. 1. The average number of EA iterations required to locate with 50% certainty the global minimum energy of the model system also grows exponentially with N . However, it does so with a smaller exponent than for the number of possible configurations. We discuss the origin of this favorable situation and look into possible measures to enhance the scaling of EAs.

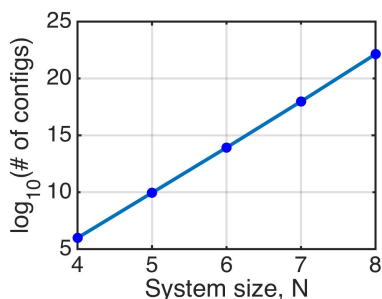


Fig. 1: Number of configurations as a function of the system size. N is the side length of the system.

¹L. B. Vilhelmsen and B. Hammer, J. Chem. Phys. 141, 044711 (2014)

²M. S. Jørgensen, et al., J. Chem. Theory Comput. 13, 1486-1493 (2017)

Local energies in density functional theory facilitated by machine learning

Thomas L. Jacobsen, Mathias S. Jørgensen and Bjørk Hammer

*Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy,
Aarhus University, Aarhus, Denmark*

When performing the search for global minimum energy structures of molecules, clusters, or solids with stochastic methods it is advantageous to have available information on the local energy of each atom. However, in first-principles approaches, where the energy expression is quantum mechanical, there is no local energy owing to the delocalized nature of the wave functions describing the constituents. Specifically, in density functional theory calculations, the electrons are described by delocalized one-electron wave functions whose (Kohn-Sham) energies enter the total energy expression but have no direct local origin. To facilitate the construction of local energies, we have introduced machine learning in the form of Kernel Ridge Regression (KRR) on an accumulated set of global-structure/global-energy relations. Using a simple finite difference approach, the KRR provides an estimate of the relative contribution to the total energy from each atom in a structure. To prove the validity of this local energy measure, we use it in conjunction with an evolutionary algorithm^{1,2} whose pairing and mutation steps may exploit the local energy information³.

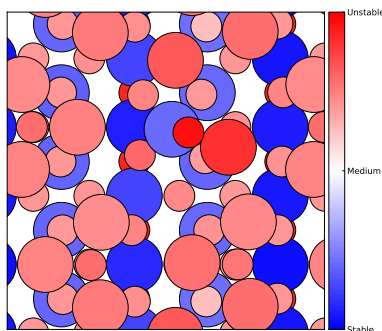


Fig. 1: Relative local energies for atoms on a SnO_2 surface.

¹L. B. Vilhelmsen and B. Hammer J. Chem. Phys. 141, 044711 (2014)

²M. S. Jørgensen, *et al.* J. Chem. Theory Comput. 13, 1486-1493 (2017)

³T. L. Jacobsen, *et al.* unpublished

Describing MoS₂ Nanoclusters with Classical Potentials

Marc Jäger*, Eiaki V. Morooka*, Adam Foster*

Department of Applied Physics, Aalto University, Espoo, Finland

MoS₂ has recently gained research interest in various fields. This inexpensive material is used currently mainly as a lubricant and for the hydrodesulfurisation reaction, but is gaining attention for other catalytic reactions, especially hydrogen evolution. Recent studies of single-layer MoS₂ have shown that its electronic band structure can be fine-tuned on the nanoscale¹. Remarkable metallic properties occur at edges of triangularly to hexagonally shaped nanoclusters which are likely to be the active site for the hydrogen evolution reaction¹²³.

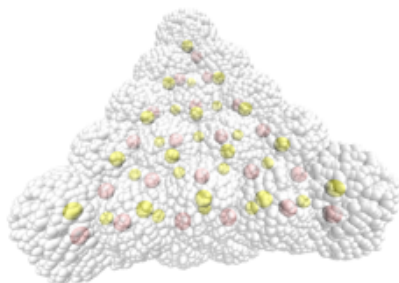


Fig. 1: Scanning hydrogen absorption sites over a MoS₂ nanocluster.

In this poster presentation, it will be shown how to simulate MoS₂ nanoclusters with a classical Stillinger-Weber potential. Trends on electronic energy and structure of MoS₂ with respect to different sizes and shapes will be presented.

¹Sorensen et al., ACS Nano 8, 6788 (2014).

²Bruix et al., ACS Nano 9, 9322 (2015).

³Walton et al., J. Catal. 308, 306 (2013).

Ab initio Prediction of Novel Electrode/Electrolyte Interfaces towards all Solid-State Batteries

B. Karasulu¹, J. Darby¹, A. J. Morris¹ and C. P. Grey²

¹ *Dept. of Physics, University of Cambridge, Cambridge, United Kingdom*

² *Dept. of Chemistry, University of Cambridge, Cambridge, United Kingdom*

Conventional rechargeable lithium-ion batteries (LIBs) utilize solid electrodes and liquid electrolytes. Combustible organic electrolytes pose potential safety risks, including volatilization, flammability and even explosion. Solid-state (SS) inorganic electrolytes have the potential to eliminate these safety concerns, while providing high-energy LIBs. Two major challenges, however, have been hindering the practical high-performance all-SS battery applications: (1) the rather low ionic conductivities of SS electrolytes compared to liquid counterparts and (2) high resistance at the electrolyte/electrode interfaces that further curtails the ion migration. Sulfide-based electrolytes are a possible solution for the former, displaying ionic conductivities comparable to the organic counterparts, with potential enhancements by dopants.¹ Besides, the interfacial resistances can be lowered by a rational design of the interfaces and the use of buffer layers.²

To tackle these issues, an automated computational procedure is adopted in this work for predicting novel SS electrode/electrolyte interfaces with lower interfacial resistances and high ionic conductivities. The procedure comprises several steps, first of which involves the generation of convex hulls by pre-screening the bulk structures of sulfide-based electrolytes with diverse compositions, phases, vacancies and doping. Subsequently, the promising electrolyte materials are screened this time for their ionic conductivity using molecular dynamics addressing Li-ion conduction. Stable surfaces of the selected bulk structures are then generated through random cuts and interfaced with the known cathode surfaces (e.g. LiCoO₂, Li-metal), while minimizing the lattice mismatch. The home-made ab initio random structure searching (PyAIRRS) code³ is used for the random search of the initial structures, and for introducing possible dopants and vacancies into the lattices. All electronic structure calculations are performed with the plane-wave density functional theory (DFT) using the CASTEP code. Further details of this procedure will be discussed in this contribution along with the most promising interfaces predicted by this approach.

¹C. George et al. *Chem. Mat.* 28, 7304-10 (2016); M. Butala et al. *Chem. Mat.*, 29,3070-82 (2017); K. See et al. *J. Am. Chem. Soc.* 136, 16368-16377 (2014)

²J. Haruyama, et al. *Chem. Mater.* 26, 4248-4255 (2014)

³C. J. Pickard, R. J. Needs, *Nat. Mater.* 7, 775-779 (2008)

Acceleration of NEB Calculations Using Gaussian Process Regression

O-P. Koistinen^{1,2}, E. Maras³, A. Vehtari¹ and H. Jónsson^{2,3}

¹ Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Finland

² Science Institute and Faculty of Physical Sciences, University of Iceland, Reykjavík, Iceland

³ Department of Applied Physics, Aalto University, Finland

The atomic ordering at interfaces and surfaces is often governed by kinetics rather than thermodynamic equilibrium. In such cases it is important to be able to identify the mechanism and estimate the rate of atomic scale structural changes and possibly simulate the interface evolution on a long time scale - much longer than what can be simulated by direct classical dynamics simulations. Stable states of the system are characterized by local minima on the energy surface and the transition paths of highest statistical weight are the minimum energy paths (MEPs) connecting the local minima. The nudged elastic band (NEB) method is commonly used to find MEPs for atomic transitions at interfaces and surfaces where each iteration involves computationally intensive DFT calculations. The Gaussian process regression (GPR) approach, using atomic forces as well as system energy, has been used to accelerate NEB calculations and greatly reduce the number of energy and force evaluations¹. The figure shows an example of how the GPR gradually builds up an accurate approximation of the energy surface in the vicinity of the MEP.

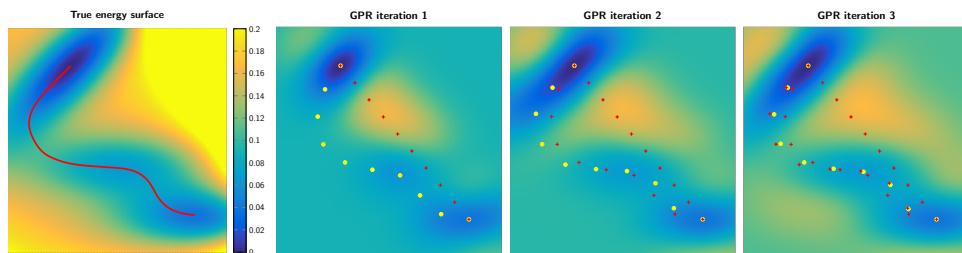


Fig. 1: The true and Gaussian process approximated energy surface and minimum energy path for a two-dimensional test problem. Far left: The true energy surface and minimum energy path (red line). Far right and intermediate figures: The approximate energy surface generated by the GPR after one, two and three iterations, points ('images') on the estimated minimum energy path (yellow dots) and points where the true energy and force have been calculated (red + signs) at each stage of the calculations.

¹O-P. Koistinen *et al.* Nanosystems: Phys., Chem., Math. 7, 925 (2016) arXiv:1703.10423

Estimating improvement potential in an evolutionary algorithm

U. F. Larsen¹, M. S. Jørgensen¹, K. W. Jacobsen², and B. Hammer¹

¹ *Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark*

² *Center for Atomic-Scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, Kgs. Lyngby, Denmark*

The objective of a global structure optimization is often to minimize the potential energy in order to find the thermodynamically most stable structures. A common challenge of such an algorithm is to balance the time spent optimizing structures that explore the entire potential energy surface and the time spent searching a local minimum. Using information about all found structures during a search to guide the progression of an evolutionary algorithm (EA) has been shown to increase performance¹. In this work, Bayesian optimization² is combined with an EA to construct a new fitness function that estimates the potential for improvement in current structures relative to the current state of the search.

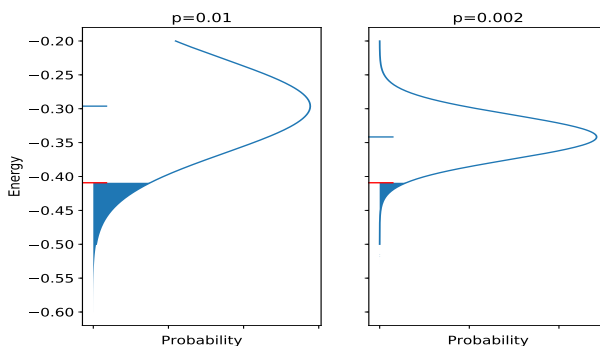


Fig. 1: Visualization of the normal distribution generated for each structure, with variance defined in the Gaussian process. The shaded area is the "probability" that the given structure is of lower energy than the best known structure, which is one choice of fitness function.

This improved fitness function favors structures in less dense regions of the search space, hence decreasing the number of duplicate structures produced, lowering stagnation in the population, while increasing the overall success rate of the EA as tested on a TiO_2 surface reconstruction. Our current work involves the implementation of kernel ridge regression to also predict this fitness of new offspring structures to avoid spending computational resources on candidate structures with little potential for improvement.

¹M. S. Jørgensen, *et al.* J. Chem. Theory Comput. 13, 1486-1493 (2017)

²J. Snoek, *et al.* NIPS 25, 2960-2968 (2012)

Machine learning structure search on alanine with gradient information

H. Paulamäki, M. Todorović and P. Rinke

Department of Applied Physics, Aalto University, Helsinki, Finland

In atomistic structure search it is important to simulate and minimize the system's multidimensional potential energy surface. Machine learning methods help to reduce the number of computationally costly evaluations needed for the task. We merged Gaussian process based Bayesian optimization with atomistic simulation codes into a tool for structure search. In tests on the alanine molecule, efficiency of finding the global minimum is considerably enhanced by adding information about the periodicity of the variables and gradients at the evaluation points. Evaluation locations are chosen iteratively using acquisition functions to make the most improvement on each step.

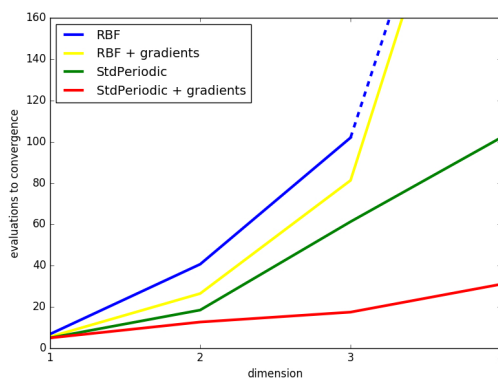


Fig. 1: Effect of kernel type and gradients on the efficiency of structure search as a function of search dimension. Used kernels are non-periodic (RBF) and periodic (StdPeriodic) both with and without gradient information.

Combination of machine learning and high-throughput DFT calculations for the prediction of thermodynamic stability

J. Schmidt¹, J. Shi² and M.A.L. Marques¹

¹ *Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, D-06099 Halle, Germany*

² *Institut Lumière Matière, UMR5306 Université Lyon 1-CNRS, Université de Lyon, F-69622 Villeurbanne Cedex, France*

We perform a large scale benchmark of machine learning methods for the prediction of the thermodynamic stability of solids. We start by constructing a data set that comprises density functional theory calculations of around 250 000 cubic perovskite systems. This includes all possible perovskite and anti-perovskite crystals that can be generated with elements from hydrogen to bismuth, and neglecting rare gases and lanthanides. Incidentally, these calculations already reveal a large number of systems (around 500) that are thermodynamically stable, but that are not present in crystal structure databases. Moreover, some of these phases have unconventional compositions and define completely new families of perovskites. This data set is then used to train and test a series of machine learning algorithms to predict the energy distance to the convex hull of stability. In particular, we study the performance of ridge regression, random forests, extremely randomized trees (including adaptive boosting), and neural networks. We find that extremely randomized trees give the smallest mean absolute error of the distance to the convex hull (121 meV/atom) in the test set of 230 000 perovskites, after being trained in 20 000 samples. Surprisingly, the machine already works if we give it as sole input features the group and row in the periodic table of the three elements composing the perovskite. Moreover, we find that the prediction accuracy is not uniform across the periodic table, being worse for first-row elements and elements forming magnetic compounds. Our results indicate that machine learning can be used to speed up considerably (by at least a factor of 5) high-throughput DFT calculations, by restricting the space of relevant chemical compositions without degradation of the accuracy.

Describing solid structure by local motifs obtained from clustering

Knud Haugaard Sørensen, Albert Bruix, Mathias S. Jørgensen and Bjørk Hammer

INANO and Department of Physics and Astronomy, Aarhus University, 8000 Aarhus C, Denmark

Representing the structure of a solid may be done uniquely by use of the unit cell - a 3×3 matrix - describing the 3D translational symmetries and a set of atomic coordinates of atoms in the basis - a $3 \times N$ matrix. In conjunction with regression methods, the use of the individual coordinates of atoms in solids is often abandoned in order to achieve translational and rotational invariance. Instead, features based on local bond-lengths, bond-angles, torsion angles and beyond are used, as formulated in various approaches, e.g. the fingerprint feature, the bag-of-bonds feature, or with feature vectors based on symmetry functions. We propose a new approach in which the information content held by such feature vectors is reduced via an unsupervised machine learning step - clustering - that maps the continuous feature vector space onto a discrete space of local atomic motifs.¹

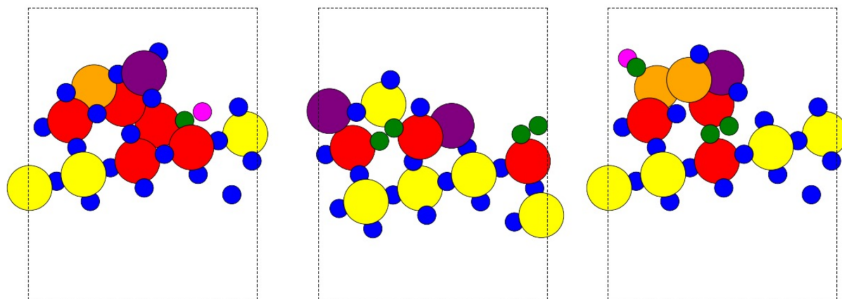


Fig. 1: Cross-section views of TiO₂ structures colored by atomic cluster assignment.

Given a large set of e.g. density functional theory total energies for a variety of solid structures, local energies can be assigned to the derived motifs using standard methods to solve over-determined linear equations. We use these motif energies to help the decision making in evolutionary algorithm searches for global minimum structures^{2,3} and look into the bottom-up construction of candidates for such structures by considering it a combinatorial problem in the space of the motifs.

¹K. H. Sørensen, *et al.*, unpublished.

²L. B. Vilhelmsen and B. Hammer, *J. Chem. Phys.* **141**, 044711 (2014).

³M. S. Jørgensen *et al.*, *J. Chem. Theory Comput.* **13**, 1486-1493 (2017).

Registered participants

Benjamin Volker	Aldritt Blum	Aalto University	Finland	benjamin.alldritt@aalto.fi
Lars Xin	Blumenthal Chen	Duke University	USA	volker.blum@duke.edu
Filippo Uffe	Federici Canova	Imperial College London	UK	lib11@ic.ac.uk
Srikanth Luca	Furlig Larsen	Aarhus University	Denmark	Bismarrok@me.com
Stefan Mario Rui	Gadicherla	Aalto University	Finland	filippo.federici@aalto.fi
Björk Ville	Gödecker	Aarhus University	Denmark	201205919@post.au.dk
Petri Alexander	Gonçalves Marques	Aalto University	Finland	srikanth.gadicherla@aalto.fi
Oliver Marc	Hammer	Fritz Haber Institute of the Max Planck Society	Germany	ghiringhelli@fhi-berlin.mpg.de
Farzaneh Nityasagar	Havu	University of Basel	Switzerland	Stefan.Goedecker@unibas.ch
Mathias Bora	Hirvonen	Martin-Luther-Universität Halle-Wittenberg	Germany	mario.goncalves-marques@physik.uni-halle.de
Sabine Jan	Hofmann	Aarhus University	Denmark	hammer@phys.au.dk
Jan Olli-Pekka	Hofmann	Aalto University	Finland	ville.havu@aalto.fi
Sanjay Kari	Hofmann	Aalto University	Finland	petri.hirvonen@aalto.fi
Thomas Jingrui Yanming	Jaeger	Umicore AG & Co. KG	Germany	alexander.hofmann@eu.umicore.com
George	Jahanbakhshi	Graz University of Technology	Austria	o.hofmann@tugraz.at
	Jena	Aalto University	Finland	marc.jager@aalto.fi
	Jørgensen	École Polytechnique Fédérale de Lausanne	Switzerland	farzaneh.jahanbakhshi@epfl.ch
	Karasulu	Institute of Nano Science and Technology Mohali	India	nityasagar.ph14216@inst.ac.in
	Karttunen	Aarhus University	Denmark	mj@inano.au.dk
	Klapp	University of Cambridge	UK	bk393@cam.ac.uk
	Kloppenburg	Aalto University	Finland	antti.karttunen@aalto.fi
	Koistinen	Technical University Berlin	Germany	klapp@physik.tu-berlin.de
	Kumar Singh	UCLouvain	Belgium	jan.kloppenburg@uclouvain.be
	Laasonen	Aalto University	Finland	olli-pekka.koistinen@aalto.fi
	Lemming Jacobsen	Aryabhata Group of Institutes	India	sanjayfizix@gmail.com
	Li	Aalto University	Finland	kari.laasonen@aalto.fi
	Ma	Aarhus University	Denmark	lemming@phys.au.dk
	Manyali	Aalto University	Finland	jingrui.li@aalto.fi
		Jilin University	China	mym@jlu.edu.cn
		Masinde Muliro University of Science and Technology	Kenya	gmanyali@mmust.ac.ke

Noa	Marom	Carnegie Mellon University	USA	nmarom@andrew.cmu.edu
Reinhard	Maurer	Yale University	USA	reinhard.maurer@yale.edu
Maosheng	Miao	California State University Northridge	USA	mmiao@csun.edu
Eiakhonroeda	Morooka	Aalto University	Finland	eiaki.morooka@aalto.fi
Harald	Oberhofer	Technical University of Munich	Germany	harald.oberhofer@tum.de
Daniel	Packwood	Kyoto University	Japan	dpackwood@icems.kyoto-u.ac.jp
Henri	Paulamäki	Aalto University	Finland	henri.paulamaki@aalto.fi
Andrew	Peterson	Brown University	USA	andrew_peterson@brown.edu
Patrick	Rinke	Aalto University	Finland	patrick.rinke@aalto.fi
Anthony	Reilly	Dublin City University	Ireland	anthony.reilly@dcu.ie
Mariana	Rossi	Fritz Haber Institute of the Max Planck Society	Germany	rossi@fhi-berlin.mpg.de
Giuseppe	Sansone	Università di Torino	Italy	giuseppe.sansone@unito.it
Laura	Scarbath-Evers	Martin-Luther-Universität Halle-Wittenberg	Germany	laura.scarbath-evers@chemie.uni-halle.de
Michael	Scherbela	Graz University of Technology	Austria	scherbela@student.tugraz.at
Jonathan	Schmidt	Martin-Luther-Universität Halle-Wittenberg	Germany	jonathan.schmidt@student.uni-halle.de
Georg	Schusteritsch	University of Cambridge	UK	gs550@cam.ac.uk
Mahdi	Shirazi	Eindhoven University of Technology	Netherlands	m.shirazi@tue.nl
Marek	Sierka	Friedrich Schiller University Jena	Germany	marek.sierka@uni-jena.de
Knud	Sørensen	Aarhus University	Denmark	khs@phys.au.dk
Olga	Syzgantseva	École Polytechnique Fédérale de Lausanne	Switzerland	olga.syzgantseva@epfl.ch
Milica	Todorović	Aalto University	Finland	milica.todorovic@aalto.fi

Aalto University Otaniemi Campus Map



- ①** Aalto Design Factory (ADF) - IMPRESS venue
- ②** Antell Restaurant - lunch
- ③** Otakaari 1 A-wing (door X) - Welcome Party
- ④** services (Alepa supermarket, pharmacy, ATM, R-kioski)
- ⑤** Aalto Inn residence
- ⑥** Otakaari 1 Y-wing: Aalto Dept. Applied Physics, COMP

