

International Workshop on  
**Machine-Learning  
for  
Materials Science**

**ML4MS 2018**

Organisers:

Milica Todorović, Filippo Federici Canova,  
Patrick Rinke, Adam Foster

# Oral contributions

# NEURAL NETWORK MOLECULAR DYNAMICS SIMULATIONS OF SOLID-LIQUID INTERFACES

V. Quaranta<sup>1</sup>, S. Kondati Natarajan<sup>1</sup>, M. Hellström<sup>2</sup> and J. Behler<sup>1,2</sup>

<sup>1</sup> Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, Germany

<sup>2</sup> Institut für Physikalische Chemie, Theoretische Chemie, Georg-August-Universität Göttingen, Germany

email: joerg.behler@uni-goettingen.de

Understanding the structural and dynamical properties of water at solid-liquid interfaces is essential for unravelling the atomistic details of key steps in electrochemistry, heterogeneous catalysis and corrosion. In recent years, in particular ab initio molecular dynamics (MD) simulations based on density-functional theory (DFT) have contributed significantly to the understanding of these processes. Still, due to their high computational costs, ab initio MD simulations of water interacting with solid surfaces are restricted to comparably small systems and short simulation times. This limitation can be overcome by employing high-dimensional neural network potentials (NNPs) [1], which are constructed from a set of electronic structure data and enable carrying out large-scale simulations with close to first-principles accuracy for a variety of systems.

In this talk I will first introduce the methodology of high-dimensional NNPs [2-4]. Using this method, NNPs for liquid water interacting with metal and oxide surfaces using copper [5] and zinc oxide [6] as benchmark systems will be presented. First, the convergence of various properties as a function of the supercell size and the thickness of the water layer in typical slab approaches will be investigated allowing to assess the reliability of conventional ab initio MD simulations. Then, the influence of copper and zinc oxide surfaces on various properties of water will be discussed addressing in particular the local water structure and the mechanisms of proton transfer events.

[1] J. Behler and M. Parrinello, *Phys. Rev. Lett.*, **98**, 146401 (2007).

[2] J. Behler, *J. Chem. Phys.*, **134**, 074106 (2011).

[3] J. Behler, *Int. J. Quant. Chem.*, **115**, 1032 (2015).

[4] J. Behler, *Angew. Chem. Int. Ed.*, **56**, 12828 (2017).

[5] S. Kondati Natarajan and J. Behler, *Phys. Chem. Chem. Phys.*, **18**, 28704 (2016).

[6] V. Quaranta, M. Hellström and J. Behler, *J. Phys. Chem. Lett.*, **8**, 1476 (2017).

# EFFICIENT N-BODY FORCE FIELDS FROM MACHINE LEARNING

C. Zeni<sup>1</sup>, A. Glielmo<sup>1</sup>, A. De Vita<sup>1,2</sup>

<sup>1</sup> Department of Physics, King’s College London, United Kingdom

<sup>2</sup> Department of Engineering and Architecture, University of Trieste, Italy  
email: claudio.zeni@kcl.ac.uk

In the recent years, several approaches to machine learning (ML) force fields (FF) for molecular dynamics have been proposed. Ultimately, despite their computational efficiency and promising level of accuracy on DFT reference calculations, these algorithms are not presently widespread as one would expect. A possible explanation for this can be found in the unintuitive functional form, and overall lower computational speed that these ML-FF offer when compared to parametric molecular dynamics FFs.

We present a method that improves our physical intuition of all classes of Gaussian Process (GP) ML FF while, most importantly, offering a orders-of-magnitude speed-up to GP predictions, thanks to a mapping procedure [1] We first introduce a set of GP kernels, offer physical interpretations of their formalism and systematically benchmark the accuracy of our GPs on a set of materials w.r.t. DFT reference force calculations. Afterwards, we showcase an example use of our in-development M-FF python package, looking at fast MD simulations of Ni nanoclusters with ab-initio accuracy on timescales of the order of 10 ns [2].

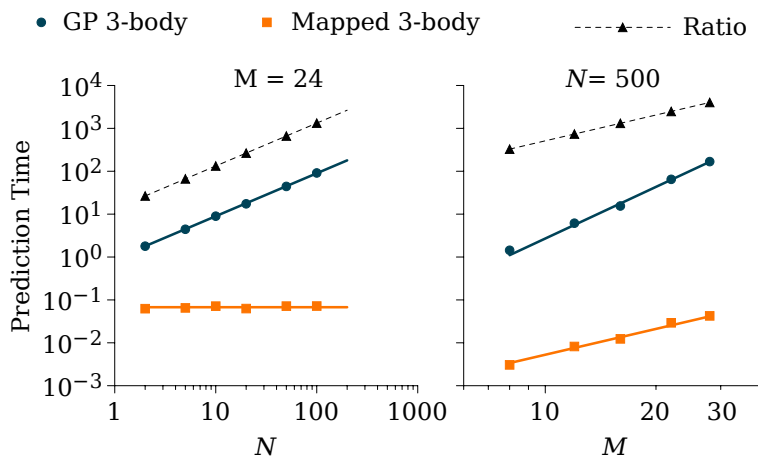


Figure 1: Computational cost of evaluating the 3-body energy as a function of the database size  $N$  and the number of atoms  $M$  located within the cutoff radius. Left: Prediction time (s) for the energy of a atomic triplet for the 3-body kernel using the GP (blue dots and solid line) and the remapped potential (orange dots and solid line), as a function of  $N$  for for a typical environment  $\rho$  including  $M = 24$  atoms.. Right: scaling of the same quantities as a function of  $M$  for  $N = 500$ .

[1] A. Glielmo, C. Zeni, A. De Vita *Submitted*, (2018).

[2] C. Zeni, K. Rossi, A. Glielmo, N. Gaston, F. Baletto, A. De Vita *Submitted*, (2018).

# Elucidating the growth mechanism in tetrahedral amorphous carbon with a machine-learning based interatomic potential

Miguel A. Caro<sup>1,2</sup>, Volker L. Deringer<sup>3,4</sup>, Jari Koskinen<sup>5</sup>, Tomi Laurila<sup>1</sup> and Gábor Csányi<sup>3</sup>

<sup>1</sup> Department of Electrical Engineering and Automation, Aalto University, Finland

<sup>2</sup> Department of Applied Physics, Aalto University, Finland

<sup>3</sup> Engineering Laboratory, University of Cambridge, United Kingdom

<sup>4</sup> Department of Chemistry, University of Cambridge, United Kingdom

<sup>5</sup> Department of Chemistry and Materials Science, Aalto University, Espoo, Finland  
email: mcaroba@gmail.com

Since the early days, there has been a debate in the community as to what physical processes are involved in tetrahedral amorphous carbon growth. Because of the impossibility to study the atomistic picture of film growth experimentally, many efforts have been made during the last 30 years to provide an explanation based on computational simulation. Unfortunately, due to the limitations of classical potentials and the cost of *ab initio* methods, these efforts have failed to provide definitive answers. Thus, the commonly accepted explanation that the “subplantation” mechanism is behind tetrahedral amorphous carbon growth and high  $sp^3$  fractions is based on indirect experimental observations and (to some degree) speculation.

In this presentation we solve the long-standing challenge of simulating the growth process which leads to the high degree of  $sp^3$  hybridization observed in dense amorphous carbon films. We do so by employing a state-of-the-art machine learning-based interatomic potential and carrying out large scale molecular dynamics simulations where the film deposition has been reproduced by studying the effect of atomic impacts one atom at a time. Thanks to this detailed and highly accurate simulation, we can elucidate the growth mechanism which takes place during deposition of energetic ions and leads to the formation of  $sp^3$ -rich films [1].

The GAP machine-learned potential that we use in this study bridges the gap between the speed of classical interatomic potentials and the accuracy of density functional theory [2, 3]. Based on the results we obtain for amorphous carbon with this new simulation framework, we challenge the subplantation theory and present compelling evidence that tetrahedral amorphous carbon grows predominantly via the “peening” process, whereby  $sp^3$  carbon is formed away from the impact site of the incident atoms due to pressure waves and as a result of a delicate balance between bond creation and annihilation.

While machine-learning approaches applied to physics, chemistry and materials science has been a trending topic for some years now, so far most studies have been restricted to proof-of-concept and validation work. In this presentation, we use machine learning as an extremely powerful tool to solve compelling physical problems and take a quantum leap in how atomistic simulations can be used to understand and predict the properties of complex materials.

[1] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, and G. Csányi, *under review*.

[2] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.* **104**, 136403 (2010).

[3] V. L. Deringer and G. Csányi, *Phys. Rev. B* **95**, 094203 (2017).

## MACHINE LEARNING IN THE CHEMICAL PHYSICS DOMAIN

Bjørk Hammer

Department of Physics and Astronomy, Aarhus University, Denmark  
email: hammer@phys.au.dk

A number of search methods based on basin hopping or evolutionary algorithms are routinely used to identify, e.g. in a density functional theory framework, the most optimal cluster and surface structures for various inorganic compounds. In this talk, I introduce simple machine learning models and show how such models, when adopted into the search methodologies, do accelerate the finding of optimal structures. The machine learning models include unsupervised and supervised models, such as clustering[1], kernel enabled regression methods[2,3], and artificial neural networks[4,5]. Common to the methods is a need for a proper representation of the compound structures and a discussion of different representations is hence taken, in particular with a view at the amount of data being available.

- [1] M. S. Jørgensen, M. N. Groves, and B. Hammer, *J. Chem. Theo. Comput.* 13, 1486 (2017).
- [2] T. L. Jacobsen, M. S. Jørgensen, and B. Hammer, *Phys. Rev. Lett.* 120, 026102 (2018).
- [3] M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, *J. Phys. Chem. A* 122, 1504 (2018).
- [4] E. L. Kolsbjerg, A. A. Peterson, and B. Hammer, *Phys. Rev. B*, *submitted*.
- [5] X. Chen, M. S. Jørgensen, J. Li, and B. Hammer, *J. Chem. Theo. Comput.*, *submitted*.

# STRUCTURE PREDICTION OF MAGNETIC TRANSITION METAL OXIDES BY USING EVOLUTIONARY ALGORITHM AND HYBRID DFT

M.S. Kuklin<sup>1</sup> and A. J. Karttunen<sup>1</sup>

<sup>1</sup> Department of Chemistry and Materials Science, Aalto University, Finland  
email: mikhail.kuklin@aalto.fi

Crystal structure prediction represents one of the major problems in physical sciences. Particularly, when some crystal structures can be successfully identified by using currently known approaches, prediction of transition metal oxides remains to be challenging. It is due to the complicated electronic structure of transition metal oxides which can be, however, reliably described by the hybrid DFT: the exact exchange part helps to solve the problem with overlocalization of metal d-orbital and, consequently, correctly treat the atomic magnetic moments.

Recently, we interfaced the USPEX code with the CRYSTAL17. It enables efficient use of the hybrid functionals for the structure predictions by using the Gaussian-type localized atomic basis sets. USPEX is an effective evolutionary algorithm-based method which utilizes ab initio codes for local relaxations of the candidates without requirements of experimental data.

For the first time, we show successful structure predictions of several transition metal oxides (NiO, CoO, and CuO) with correct atomic magnetic moments by using the USPEX in combination with the CRYSTAL17 code. The hybrid PBE0 functional with 25% Hartree-Fock and 75% PBE exchange was utilized with the optimized all-electron, Gaussian-type basis sets based on Karlsruhe def2 basis sets. For NiO and CoO oxides, minimal unit cell of 4 atoms requiring to describe antiferromagnetic structures was used. To obtain the correct magnetic ground state structure of CuO, a unit cell containing 16 atoms is needed. Otherwise, experimental data was not used in the crystal structure search by taking advantage of USPEX being a fully non-empirical approach. As a result, crystal structures and atomic magnetic moments of the studied oxides were found to be in line with experimental data.

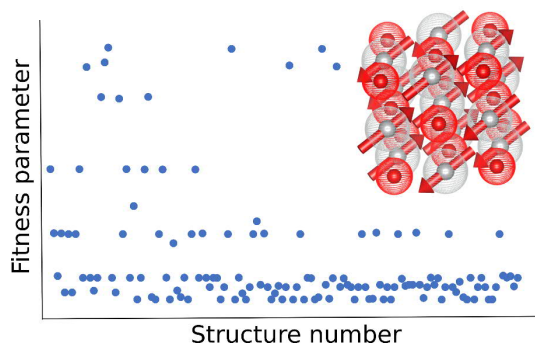


Figure 1: Example of predicted magnetic NiO structure

# EXPLORATION VERSUS EXPLOITATION IN GLOBAL ATOMISTIC STRUCTURE OPTIMIZATION

M. S. Jørgensen<sup>1</sup>, U. F. Larsen<sup>1</sup>, K. W. Jacobsen<sup>2</sup>, and B. Hammer<sup>1</sup>

<sup>1</sup> Interdisciplinary Nanoscience Center (iNANO) and Department of Physics and Astronomy, Aarhus University, Denmark

<sup>2</sup> Center for Atomic-Scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, Denmark  
email: mj@inano.au.dk

The ability to navigate vast energy landscapes of molecules, clusters, and solids is a necessity for discovering novel compounds in computational chemistry and materials science. For high-dimensional systems, it is only computationally feasible to search a small portion of the landscape, and hence, the search strategy is of critical importance. By introducing Bayesian optimization concepts in an evolutionary algorithm framework [1], I will show how one can quantify the concepts of exploration and exploitation in global minimum searches. This quantification enables control of the balance between probing unknown regions of the landscape (exploration) and investigating further regions of the landscape known to have low-energy structures (exploitation) via a single parameter. The search for global minima structures proves significantly faster with the optimal balance for three test systems (molecular compounds) and to a lesser extent also for a crystalline surface reconstruction. In addition, global search behaviors are analyzed to provide reasonable grounds for an optimal balance for different problems.

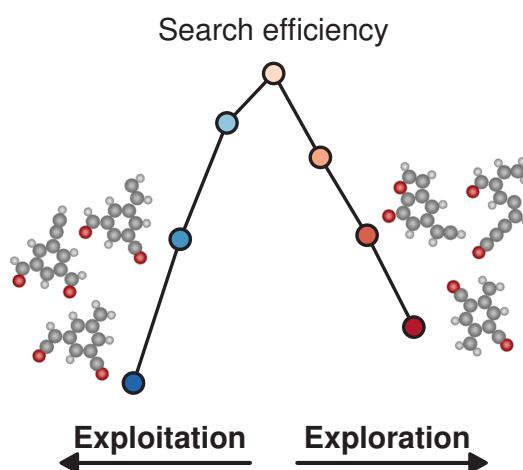


Figure 1: The success of an evolutionary global minimum search as a function of the exploitation/exploration balance.

[1] M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, *J. Phys. Chem. A*, 122, 1504-1509 (2018).



# MACHINE LEARNING FOR MATERIALS DESIGN

L. Ward<sup>1</sup>

<sup>1</sup> Computation Institute, University of Chicago, USA  
email: loganw@uchicago.edu

Computational methods offer the ability to drastically accelerate the design of materials, but there are many material properties which lack sufficient theoretical models or computational tools. One route to enable the computational design of more materials is to employ machine learning to automatically learn models from materials data. In this talk, we will present general-purpose approaches for creating machine learning models and demonstrate their application to the design of several materials. Specifically, we will show a machine learning technique that can link the composition of a material to a diverse range of properties. We then demonstrate the use of such models to optimize commercial bulk metallic glass alloys and discover new amorphous metal coatings.

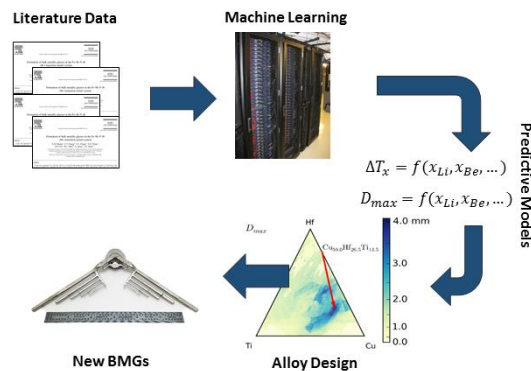


Figure 1: Approach used to design bulk metallic glass alloys with machine learning. We used data extracted from the materials science literature to build several machine learning models, which let us quickly identify candidate alloys.

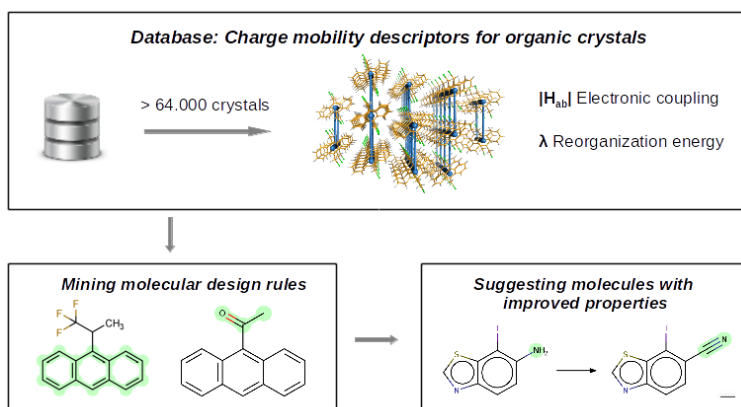
## How does molecular structure influence charge mobility? - Mining a database of organic semiconductors.

C. Kunkel, C. Schober, J. Margraf, K. Reuter and H. Oberhofer

Chair for Theoretical Chemistry, Technische Universität München, Lichtenbergstr. 4, D-85747 Garching, Germany  
email: christian.kunkel@tum.de

Improving charge carrier mobilities in organic semiconductors is a challenging task, usually tackled by structural tuning of a promising compound family, while relying on intuition or experience. Still, the vast chemical space is then only locally explored, while promising design strategies might also be uncovered from systematic analysis of large compound databases. We carry out such an analysis by applying charge carrier mobility simulations and data mining strategies to an inhouse database of > 64.000 organic crystals, obtained from the Cambridge Structural Database (CSD) and screened for charge transport properties using first principle derived descriptors[1].

The analysis shows, that our screening recovers many known and well performing materials, while also finding many promising candidates, not yet considered for organic electronics applications. To further derive design principles from the data, we evaluate the intrinsic suitability of  $\approx 200$  molecular scaffolds found to be contained as compound-clusters in our database. A similarity network analysis hints at already explored regions of chemical space, while statistical analysis of the clusters uncovers significant expectable performance differences. Likewise, an evaluation of attached sidegroups reveals significant tuning potential, demonstrated also in a simple molecular optimization approach. The discovered trends can therefore be a basis for further in-depth theoretical and experimental design of materials for organic electronics, highlighting the value of data-based approaches.



[1] C. Schober, K. Reuter and H. Oberhofer, *J. Phys. Chem. Lett.*, 7, 3973 (2016).

# High-throughput Screening of Transparent Conducting Oxides

Christopher Sutton

*Theory Department, Fritz Haber Institute of the Max Planck Society, Berlin, Germany*

Transparent conducting oxides (TCOs) are crucial for the operation of a variety of technological devices such as photovoltaic cells and light-emitting diodes; however, only a small number of compounds are currently known to display both transparency and conductivity suitable enough to be used as transparent conducting materials. We investigate ternary and quaternary alloys of group-III oxides with the formula  $(\text{In}_x\text{Ga}_y\text{Al}_z)_2\text{O}_3$  (where  $x+y+z=2$ ) as alternative TCOs by screening for stability and key computational parameters that define charge mobility. An important consideration in the identification of stable crystalline mixtures is that the lowest thermodynamic energy of approximately  $2^N$  possible configurations has to be examined (for a two-component mixture). However, an exhaustive search of a large configurational space becomes computationally infeasible at the density functional theory (DFT) level. Instead, numerically efficient methods for estimating the stability of new alloys are needed to efficiently search (meta)stable configurations.

To address the need for finding new materials with an ideal target functionality, the construction of cluster-expansion-based potentials for many lattice types using a sparse-regression machine-learning technique (compressed sensing) will be discussed. Combining this approach with the nested sampling algorithm, which is a Bayesian Markov chain Monte Carlo method, allows for an examination of the thermodynamics of ordering and phase separation in alloys. With a large computational database of group-III oxides, an open big-data competition was organized by Novel Materials Discovery Repository (NOMAD) and hosted by Kaggle. A summary of the best machine-learning models applied for the prediction of both the formation enthalpy (an indication of stability) and the bandgap energy (an indication of optical transparency) will be discussed.

## Towards an error-aware analysis of multiscale kinetic models

Sebastian Matera<sup>1</sup>

<sup>1</sup> Department of Mathematics and Computer Science, Freie Universität Berlin, Arnimallee 9, 14195 Berlin, Germany  
email: [matera@math.fu-berlin.de](mailto:matera@math.fu-berlin.de)

Multiscale modeling has become an important tool in materials research allowing to predict materials response on basis of microscopic models. A hitherto hardly addressed problem is the propagation of small scale errors to the targeted simulation result on the coarser scales. Such small scale errors might results from e.g. assumptions made for coarse graining or because we can only approximately solve the root model. The questions arise about what is the accuracy of the multiscale model predictions, which conclusions can still be drawn from the model and which small scale errors cause the large scale uncertainty.

In the talk, I will address these questions in the context of first principles based chemical kinetic models using Uncertainty Quantification (UQ) and Global Sensitivity Analysis (GSA)[1, 2]. It turns out that the errors in the Density Functional Theory derived energetics can lead to significant uncertainties in the predicted reactivities often spanning orders of magnitude. However, our analysis also reveals that it is still possible to draw conclusions on the key atomistic aspects controlling reactivity such as potential rate determining steps.

Large parts of my talk will focus on the utilization of surrogate modelling strategies, which we have developed during the last years. The first of these methods extend the well-known Modified Shepard interpolations by an automatic adjustment of the basis functions to the properties of original model[3], which allows to tackle models with rapid changes in the parameter response. The second is locally and dimension adaptive sparse grids (ASG), which overcome the *curse of dimensionality* to a large extend and which adaptively choose for which of the input parameters the original model needs to be evaluated[2]. Additionally, I will discuss the extension of the ASG to a multi-level strategy for stochastic models - such as kinetic Monte Carlo or Molecular Dynamics - allowing a reduction of the overall sampling effort by orders of magnitude.

- [1] S. Döpking, S. Matera *Chemical Physics Letters*, 674, 28-32 (2017).
- [2] S. Döpking, C.P. Plaisance, D. Strobusch, K. Reuter, Ch. Scheurer, S. Matera *The Journal of Chemical Physics*, 148, 034102 (2018)
- [3] J. M. Lorenzi, T. Stecher, K. Reuter, S. Matera *The Journal of Chemical Physics*, 147, 164106 (2017)

## Accelerating the Study on Nanocluster Catalysts with the Smooth Overlap of Atomic Positions(SOAP) Descriptor

Eiaki V. Morooka<sup>1</sup>

<sup>1</sup> Department of Applied Physics, Aalto University , Finland  
email: eiaki.morooka@aalto.fi

Platinum Group Metals (PGMs) are heavily used for catalysts that can withstand high acidic environments such as for fuel cells[1], however; there is a limited amount of supply [2]. We are studying characteristics of catalysts of nanoclusters using (ML) methods to automate the procedure of substituting expensive catalysts by rationally designing them with earth-abundant materials. We do so by eliminating redundant quantum mechanical calculations, classifying surface structures and interpolating adsorption energy using a state-of-the-art ML descriptor called Smooth Overlap of Atomic Positions (SOAP) [3].

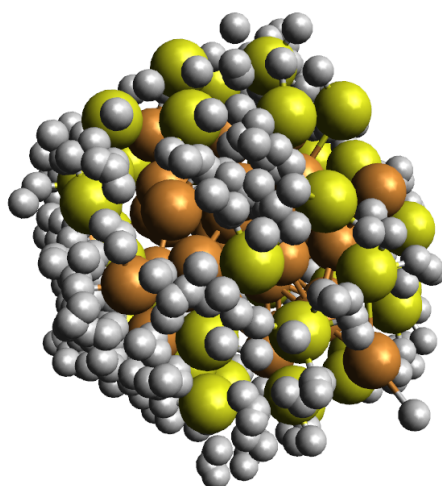


Figure 1: Relaxed Au<sub>40</sub>Cu<sub>40</sub> nanocluster with surface hydrogens that were selected uniquely with the SOAP measure differences.

- [1] European Commission. Report on Critical Raw Materials for the EU, Ad hoc Working Group on defining critical raw materials. *Technical report*, (2014).
- [2] <https://newscientist.com/data/images/archive/2605/26051202.jpg>
- [3] Albert P. Bartók, Risi Kondor, and G. Csányi. On representing chemical environments. *Physical Review B - Condensed Matter and Materials Physics*, 87(18):1-19(2013).

# MACHINE LEARNING OF MOLECULAR ORBITAL ENERGIES

A. Stuke<sup>1</sup>, L. Xie<sup>2</sup>, M. Todorović<sup>1</sup> and P. Rinke<sup>1</sup>

<sup>1</sup> Department of Applied Physics, Aalto University, Espoo, Finland

<sup>2</sup> Department of Computer Science, Hunter College and the City University of New York, New York City, USA

email: annika.stuke@aalto.fi

The capacity to efficiently design new and advanced optoelectronic materials is hampered by the lack of suitable methods to rapidly and accurately identify yet-to-be-synthesized materials that meet a desired application. To overcome such design challenges, we present a machine learning model based on kernel ridge regression (KRR) that can predict spectral properties of thousands of organic molecules using two different types of molecular representation: the Coulomb matrix [1] and the many-body tensor representation [2]. The model is trained on energies of the highest occupied molecular orbital (HOMO), pre-computed with density functional theory for three different datasets that consist of 44k, 64k and 134k molecules [3,4,5].

After training, the resulting KRR model successfully predicts HOMO energies of out-of-sample molecules at negligible computational cost. We find that the predictive accuracy of KRR greatly depends on the choice of molecular representation, with the many-body tensor representation performing superior to the Coulomb matrix for all three datasets in predicting HOMO energies. Moreover, we show that for large datasets, an equivalent prediction quality can be achieved by training on small, structurally diverse subsets that well represent the chemical space as covered by the original dataset while being only a fraction of its size.

[1] M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 108, 058301 (2012).

[2] H. Huo and M. Rupp, <https://arxiv.org/abs/1704.06439> (2017).

[3] M. Ropo, M. Schneider, C. Baldauf and V. Blum, *Sci. Data*, 3 (2016).

[4] C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Cryst.*, B72, 171-179 (2016).

[5] R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, *Sci. Data.*, 1 (2014).

# Using Machine-Learning to Create Predictive Material Property Models and Accelerate Combinatorial Searches

Chris Wolverton

*Dept. of Materials Science and Eng., Northwestern University, Evanston, IL (USA)*

Rational, data-driven materials discovery has the potential to make research and development efforts far faster and cheaper. In such a paradigm, computer models trained to find patterns in massive chemical datasets would rapidly scan compositions and systematically identify attractive candidates. Here, we present several examples of our work on developing machine learning (ML) methods capable of creating predictive models using a diverse range of materials data. As input training data, we demonstrate ML on both large computational datasets of DFT calculations, as implemented in the Open Quantum Materials Database (oqmd.org), and also experimental databases of materials properties. We construct ML models using a large and chemically diverse list of attributes, which we demonstrate can be used as an effective tool to automatically learn intuitive design rules, predict diverse properties of crystalline and amorphous materials, such as formation energy, specific volume, band gap energy, and glass-forming ability, and accelerate combinatorial searches.

## Machine Learning for Molecular Materials: Stability, Properties and Experimental Observables

Felix Musil<sup>1</sup>, Sandip De<sup>1</sup>, Federico M. Paruzzo<sup>2</sup>, Albert Hofstetter<sup>2</sup>, Yang Jack<sup>3</sup>, Joshua E. Campbell<sup>3</sup>, Graeme M. Day<sup>3</sup>, Lyndon Emsley<sup>2</sup>, Michele Ceriotti<sup>1</sup>.

1. Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

2. Institut des Sciences et Génie Matériaux, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

3. School of Chemistry, University of Southampton, Highfield, Southampton, UK.

Molecular crystals play an important role in several fields of science and technology. They frequently crystallize in different polymorphs with substantially different physical properties. To understand their structures, machine-learning (ML) methods have recently emerged as a way to overcome the need for expensive explicit high accuracy first-principle calculations. We show how a recently-developed ML framework [1] can be used to achieve sub-kJmol accuracy in the prediction of DFT lattice energies of pentacene and two azapentacene isomers [2], using only a few hundred reference configurations and provide a data-driven classification that is less biased and more flexible than typical heuristic rules.

At the same time, the elucidation of the structure of synthesized molecular crystals often requires theoretical inputs to interpret NMR spectra. We use a ML method based on local environments to accurately predict chemical shifts of different molecular solids and of different polymorphs within DFT accuracy [3], RMSE of 0.49 ppm (<sup>1</sup>H), 4.5ppm (<sup>13</sup>C), 13.3 ppm (<sup>14</sup>N), and 17.7 ppm (<sup>17</sup>O), and we demonstrate that the trained model is able to correctly determine structures of cocaine and the drug 4-[4-(2-adamantylcarbamoyl)-5-tert-butylpyrazol-1-yl]benzoic acid in a chemical shift based NMR crystallography approach.

[1] De, S., Bartok, A. P., Csanyi, G., & Ceriotti, M. (2016). *Phys. Chem. Chem. Phys.*, 18(20), 13754.

[2] Musil, F., De, S., Yang, J., Campbell, J. E., Day, G. M. & Ceriotti, M. (2016). *Chem. Sci.* 9, 1289–1300.

[3] Federico M. Paruzzo, Albert Hofstetter, Félix Musil, De Sandip, Michele Ceriotti, and Lyndon Emsley. *Chemical Shifts in Molecular Solids by Machine Learning* (submitted).



## A statistical learning approach to predict martensite start temperature and martensite fraction in steels

Moshiour Rahaman<sup>1</sup>, Wangzhong Mu<sup>1,2</sup>, Joakim Odqvist<sup>2,1</sup> and Peter Hedström<sup>2,1</sup>

<sup>1</sup>HiMat Engineering AB, SE-131 60 Nacka, Sweden

<sup>2</sup>Dept. Materials Science and Engineering, KTH Royal Institute of Technology,

SE-100 44 Stockholm, Sweden

email: moshiour.rahaman@himat.se

The martensite start temperature ( $M_s$ ) and the amount of athermal martensite formed at certain temperatures are important parameters when designing high-performance steels and their heat treatments. It has therefore attracted significant interest over the years with the application of numerous techniques to model the  $M_s$ . For example, thermodynamics-based, linear regression and artificial neural network (ANN)<sup>1,2,3</sup> models have been proposed to predict the  $M_s$  temperature. Recently, however, statistical learning (SL) techniques have made significant progress and the availability of open materials data motivates a new attempt at data-driven predictions of the  $M_s$  and the fraction of athermal martensite in steels. We report on the development of a SL software tool to predict the  $M_s$  temperature and the athermal martensite fraction in steels. Furthermore, we make a systematic comparison between ensemble learners<sup>4,5</sup> of SL techniques and ANN.

### References

1. Bhadeshia, H.K.D.H. Neural Networks in Materials Science, ISIJ International, (39)10: 966-979(1999).
2. T. Sourmail and C. Garcia-Mateo. Critical assessment of models for predicting the  $M_s$  temperature of steels. Computational Materials Science, 34, 323-334 (2005)
3. Yu, Z.W. Artificial Neural Networks in Material Science Application, Applied Mechanics and Materials, (20)23 1211-1216 (2010).
4. L. Breiman, Bagging Predictors, Machine Learning, 24, 123-140 (1996)
5. L. Breiman, Random Forests, Machine Learning, 45, 5-32 (2001)

# Insightful crystal-structure classification using deep learning

ANGELO ZILETTI<sup>1</sup>, MATTHIAS SCHEFFLER<sup>1</sup>, AND LUCA M. GHIRINGHELLI<sup>1</sup>

<sup>1</sup>*Fritz Haber Institute of the Max Planck Society, Berlin, Germany*  
ziletti@fhi-berlin.mpg.de

Big data is emerging as a new paradigm in materials science. A vast amount of three-dimensional structural data is provided by both computational repositories (e.g. <http://nomad-coe.eu>) and experiments (e.g. atom probe tomography). Computational methods that automatically and efficiently detect long-range order are of paramount importance for materials characterization and analytics. Current methods are either not stable with respect to defects, or base their representation on local atomic neighbourhoods, which in turn makes it difficult to detect “average” long-range order.

In the proposed approach, for a given crystal structure we first calculate its diffraction pattern, expand it on spherical harmonics, and then use a neural-network model to obtain a compact, low-dimensional representation. We apply this workflow to a subset of materials from the Novel Materials Discovery (NOMAD) Archive, and show that our deep-learning-based approach compactly encodes structural information, is robust to defects (e.g. point defects, and/or strain), and allows to build easily interpretable structural-similarity maps. This work received funding from the NOMAD Laboratory, a European Center of Excellence.

## **HOW MANY MATERIALS ARE LEFT TO DISCOVER? AN EXPLORATION OF QUATERNARY SPACE**

M. Sluydts<sup>1,2</sup>, M. Larmuseau<sup>1,2</sup>, T. Crepain<sup>1</sup>, K. Dumon<sup>1</sup>, K. Lejaeghere<sup>1,3</sup> and S. Cottenier<sup>1,2</sup>

<sup>1</sup> Center for Molecular Modeling, Ghent University, Belgium

<sup>2</sup> Department of Electrical Energy, Metals, Mechanical Constructions & Systems, Ghent University, Belgium

<sup>3</sup> Department of Applied Physics, Ghent University, Belgium

email: michael.sluydts@ugent.be

The frontier of materials science is shifting evermore towards the development of ‘exotic’ functional materials, which display an unfamiliar combination of properties. The underlying behavior giving rise to these materials’ properties is often too complex to predict purely from their crystal structure. New exotic materials are thus largely developed by mimicking existing materials, inevitably introducing bias.

While truly new exotic materials are likely to exist in unknown regions of materials space, it is unlikely we will find them through biased exploration. At the same time, random exploration is unsustainable given the time required to synthesize and characterize new materials. Several questions thus arise. How can we explore the vast materials space intelligently, yet without bias? And perhaps most importantly: how many materials are left to discover?

We investigate this fundamental question by creating a database of hypothetical crystals in quaternary space, where experimental exploration is limited. By employing high-throughput ab initio methods, we are able to predict various properties of these unknown materials, including their stability. Furthermore, applying machine learning during the screening procedure yields a ten-fold speedup over brute-force exploration. This yields a relatively unbiased, yet fast exploration method.

By comparing the discovery rate, composition and structure of the new materials with that of experimentally known quaternary phases, an estimation can be made of how many materials are yet to be discovered within this region of materials space.

## High-throughput structural classification of atomistic system

L. Himanen<sup>1</sup>, P. Rinke<sup>1</sup> and A. S. Foster<sup>1,2,3</sup>

<sup>1</sup> Department of Applied Physics, Aalto University, Finland

<sup>2</sup> Graduate School Materials Science in Mainz, Germany

<sup>3</sup> WPI Nano Life Science Institute, Kanazawa University, Japan

email: lauri.himannen@aalto.fi

Projects such as the Materials Genome Initiative[1] and the Novel Materials Discovery Laboratory [2] produce large *materials databases* containing information extracted from atomistic simulations, e.g., system geometries and methodology. Often database users would like to search these databases for specific structural classes such as crystals, molecules, surfaces or 2D-materials. To facilitate such searches, the database entries should be tagged according to a classification system. Unfortunately these structural classifications are not always provided. To cope with large heterogeneous datasets from atomistic calculations, automated and verifiable methods for classifying atomistic structures are becoming necessary.

We introduce a general structure classification concept that can be systematically used for all structural types. We also discuss how the materials genealogy can be intuitively mapped to produce a materials “tree of life”. We then discuss our implementation for the automatic and accurate classification of two-dimensional structures, including surfaces and 2D-materials, that requires no explicit search patterns.

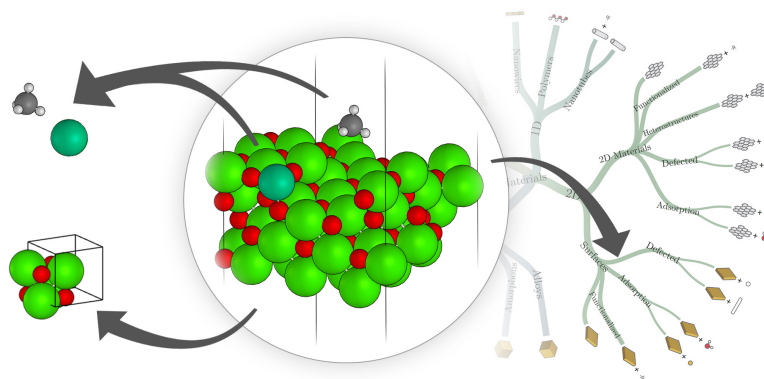


Figure 1: An illustration of the main features of our automatic structural classification scheme for two-dimensional materials. Given an atomic structure, it is placed on the materials genealogy and the bulk unit cell is detected together with outlier atoms, such as adsorbates.

[1] Materials Genome Initiative, <https://www.mgi.gov/>.

[2] The Novel Materials Discovery (NOMAD) Laboratory, <https://nomad-coe.eu/>.

# Poster contributions

## FROM MACHINE-LEARNING POTENTIALS TO MATERIALS CHEMISTRY

V. L. Deringer<sup>1,2</sup> and G. Csányi<sup>1</sup>

<sup>1</sup>Department of Engineering, University of Cambridge, UK

<sup>2</sup>Department of Chemistry, University of Cambridge, UK

email: vld24@cam.ac.uk

Machine-learning based interatomic potentials find increasing application in materials modelling due to their accuracy and attractive computational cost. On this poster, I will describe two recent developments that make use of the Gaussian Approximation Potential (GAP) framework [1] and specifically address questions in chemistry. First, having previously shown that GAPs can be used for crystal-structure searching [2], we now developed a more general protocol for fitting GAPs and concomitantly exploring structural space “from scratch”; this will be exemplified for the different allotropes of boron [3]. Second, we developed a strategy for fitting GAPs for guest atoms in host structures, exemplifying our approach for Li intercalation in carbon [4]. Our results hint towards the usefulness of GAPs and other machine-learning-based potentials in materials chemistry.

- [1] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Phys. Rev. Lett.*, 104, 136403 (2010).
- [2] V. L. Deringer, G. Csányi, and D. M. Proserpio, *ChemPhysChem*, 18, 873 (2017).
- [3] V. L. Deringer, C. J. Pickard, and G. Csányi, <https://arxiv.org/abs/1710.10475>.
- [4] S. Fujikake, V. L. Deringer, T. H. Lee, M. Krynski, S. R. Elliott, and G. Csányi, *J. Chem. Phys.*, in press (preprint at <https://arxiv.org/abs/1712.04472>).

## NEURAL NETWORKS TO SIMULATE DYNAMICS

S. Amabilino<sup>1</sup> and Dr. L. Bratholm<sup>1</sup>, Dr. S. Bennie<sup>1</sup>, Dr. D Glowacki<sup>1</sup>

<sup>1</sup> Department of Chemistry, University of Bristol, United Kingdom  
email: [silvia.amabilino@bristol.ac.uk](mailto:silvia.amabilino@bristol.ac.uk)

In our group we are looking at the reaction of a cyanide radical with squalane chains to study hydrogen abstractions. This reaction is interesting from an atmospheric chemistry point of view, because it represents a reaction at a gas-liquid interface. We want to simulate the dynamics of this reaction, but because of the electronic structure of the cyanide radical, high level quantum calculations are needed to obtain accurate forces and energies. This is impractical, as they require considerable computational resources. Consequently, we aim to use a neural network to fit the potential energy surface for this system and then use it to predict the forces in a molecular dynamics simulation. This should give forces and energies that have the accuracy of the high level quantum calculations at the speed of classical force fields.

## **Systematically identifying the best chemical representation for machine learning, with applications in crystal structure prediction**

D. McDonagh<sup>1</sup>, C. -K. Skylaris<sup>1</sup>, G. M. Day<sup>1</sup>

<sup>1</sup> Department of Chemistry, University of Southampton, United Kingdom  
email: dm1m15@soton.ac.uk

The ability of machine learning models to predict energy calculations, or energy corrections between two levels of theory, has been demonstrated for a number of systems in the literature. A key component of this is the representation or descriptor of the system, of which there are now many. Comparisons between descriptors are lacking, and are typically carried out by using different descriptors for a given model and observing the outcome. It would be useful to have an approach to comparing descriptors that is independent of the machine learning model used. Current rules of thumb for a good descriptor include that the descriptor must be invariant to permutation, unique, locally constant, and preferably small to reduce the effects of the curse dimensionality. The first of these can generally be shown a priori, but measuring the remaining properties can be difficult in practice.

Here we describe a simple metric for quantifying the effectiveness of a given descriptor, in terms of its uniqueness and local constancy, allowing for the best descriptor for a given dataset to be identified. Any hyperparameters of a descriptor can also then be optimised for these conditions, before any machine learning takes place. We demonstrate this in the field of crystal structure prediction, where thousands of lattice energy calculations at a high level of theory are often required. By systematically identifying the best descriptor and then selecting a training set that best represents the space, we find sub kJ/mol prediction errors in correcting force field lattice energies to density functional theory with a range of machine learning models, using training set fractions as low as 10%.



## Study of novel 2D structures with Siesta and AiiDA

Vladimir Dikan, Alberto García, Victor M. García-Suárez, Emanuele Bosoni, Pablo Ordejón

An interface between the DFT code Siesta [1] and the Automated Interactive Infrastructure and Database for Computational Science (AiiDA) [2] is presented. The interface allows to easily design Siesta calculations, submit them to local or remote computers and analyse the results. It includes the calculation of forces, stresses and band structures and also geometry relaxations. The Siesta-AiiDA infrastructure is designed to create workflows that can easily automate the design of simulations, the extraction of parameters and the generation of large amounts of data. An example of workflow that can be used to design new 2D structures is presented.

[1] J. M. Soler, E. Artacho, J. Gale, A. García, J. Junquera, P. Ordejón, and D. S. Portal, *J. Phys.: Condens. Matter* **14**, 2745 (2002).

[2] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, *Comp. Mat. Sci.* **111**, 218 (2016).

## Machine Learning Structural Descriptors on Nanocluster Catalysts

Marc Jäger<sup>1</sup>, Eiaki Morooka<sup>1</sup>, and Adam Foster<sup>1,2</sup>

— 1 Aalto University, Helsinki, Finland

— 2 Kanazawa University, Japan

Email: [marc.jager@aalto.fi](mailto:marc.jager@aalto.fi)

Scientists have advanced significantly in producing nanoparticles with defined composition, size and morphology in the last decade. Due to this and because of their remarkable properties, nanoclusters have gained attention in heterogeneous catalysis. Nanoclusters differ from bulk metal behaviour, their catalytic properties are sensitive to changes in size and morphology. Nanoparticles like molybdenum disulfide are known to catalyze the hydrogen evolution reaction (HER). The combinatorial and structural space of nanoclusters is vast, so extensive modelling is difficult. Structural descriptors are used to describe the geometry of an adsorption site and to predict properties which indicate a high catalytic activity, in particular the hydrogen adsorption free energy. We analysed the performance of state-of-the-art structural descriptors (SOAP, MBTR and ACSF). Simulations can provide energetic and kinetic analysis of HER using DFT. The vast amount of possible nanoclusters, all potential candidates for catalysing the HER, requires reduction and interpolation of DFT calculations. This is tackled by merging the combinatorial space with the chemical compound space and applying machine learning on diverse datasets.

## Global Search on a Machine Learned Energy Surface

Henrik Lund Mortensen<sup>1</sup>

<sup>1</sup> Department of Physics, Aarhus University, Denmark  
email: hlm@phys.au.dk

We search a machine learned energy surface to find the global minimum structure of various chemical systems. Although machine learning regression techniques are superb at interpolating between samples (obtained with DFT calculations) they often fail to be accurate when extrapolating beyond the sampled region. However, useful information can be extracted from features (such as local minima) in the extrapolation region. This information reveals interesting parts of the configuration space which are not yet sampled. The low cost of evaluating the machine learned potential allows us to greedily search the configuration space, and thereby quickly locating the regions of interest, which potentially contains the global minimum of the true, underlying potential energy surface.

## Machine learning enhanced Monte Carlo optimization of carbon structures

S. A. Meldgaard<sup>1</sup>, Esben L. Kolsbjerg<sup>1,2</sup>, and B. Hammer<sup>1,2</sup>

<sup>1</sup> Department of Physics and Astronomy, Aarhus University, Denmark

<sup>2</sup> Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Denmark  
email: sm@phys.au.dk

A geometric optimization of a 19 atom two-dimensional Lennard-Jones structure is performed using machine learning techniques to enhance the underlying Monte Carlo framework. By clustering local feature vectors a global feature vector is constructed and subsequently used for producing local energies by the use of ridge regression. The local energies are used to improve upon the traditionally stochastic choice of atoms to rattle in the Monte Carlo step. The method is then applied to a three-dimensional quantum system in the form of 24 carbon atoms. For both Lennard-Jones structures and the Carbon system an earlier optimization success is observed.

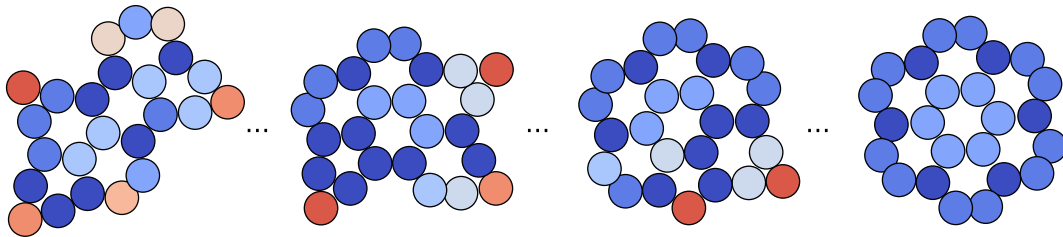


Figure 1: Sample of structures observed during a carbon optimization run, colored by predicted local energies.

[1] Meldgaard *et al.*, *in manuscript*

# CONSTRUCTION OF MACHINE LEARNING INTERATOMIC POTENTIAL FOR LIQUID RUBIDIUM

E. Oren and G. Makov

Department of Materials Engineering, Ben Gurion University of the Negev, Beer-Sheva  
84105, Israel  
email: oreneya@post.bgu.ac.il

In materials science, atomic scale problems are often investigated computationally by means of molecular dynamics simulations. These require the knowledge of interatomic interaction forces of the material in question. The classical models are constrained by physical approximations and thus have limited capabilities in terms of accuracy as compared with ab-initio calculations and experiments. Machine learning models, on the other hand, by their nature have the potential to fit any physical form that characterizing the relevant material system. In the present work we use density functional theory derived data of liquid rubidium to fit an interatomic potential model. In particular, we explore the space of hyperparameters that control fully connected neural networks and 3D convolutional neural networks.

# DEEP LEARNING REPRESENTATIONS OF MICROSTRUCTURES

M. Larmuseau<sup>1,2,3</sup>, M. Sluydts<sup>1,2</sup>, T. Dhaene<sup>3</sup> and S. Cottenier<sup>1,2</sup>

<sup>1</sup> Center for Molecular modeling, Ghent University, Belgium

<sup>2</sup> Electrical Energy, Metals, Mechanical Constructions and Systems, Ghent University, Belgium

<sup>3</sup> SUMO Lab, Ghent University, Belgium

email: michiel.larmuseau@ugent.be

For decades, metallurgists have relied on intuitive physical features of the microstructures such as the grain size to establish a quantitative link with the properties of the material. However, for complex, multiphase steels features such as the grain size cannot always be discerned and hence these traditional methods fall short. In order to analyse these types of microstructures, methods from computer vision have been successfully applied as an alternative. Different methods have been applied by different authors for the classification of distinct classes of steel, but a benchmark of these techniques on a common dataset is currently missing, to the best of our knowledge. We provide a benchmark of these features on a dataset of more than thousand optical microscopy images with more than 60 different classes of steel.

Inspired by the surge of deep learning in many scientific fields, we investigate the potential of deep learning in extract relevant features from microstructures. To this end, a supervised technique called "Siamese Networks" is used. This method allows to represent each microstructure in a low dimensional space where the distance between microstructure belonging to the same class is minimized.

# SUPPORT VECTOR MACHINE AND GAUSSIAN PROCESS AIDED AB INITIO PREDICTIONS: CLASSIFYING ULTRA-HIGH TEMPERATURE FRENKEL PAIRS<sup>A</sup> AND PREDICTING PHASE TRANSITION ENTROPY<sup>B</sup>

T.A. Mellan<sup>1,A,B</sup> and M.W. Finnis<sup>2,A</sup>

<sup>1</sup> Thomas Young Centre for Theory and Simulation of Materials, Department of Materials, Imperial College London, UK

<sup>2</sup> Department of Physics, and Thomas Young Centre for Theory and Simulation of Material, Department of Materials, Imperial College London, UK  
email: [t.mellan@imperial.ac.uk](mailto:t.mellan@imperial.ac.uk)

Density functional theory (DFT) has become ubiquitous in the prediction of solid state properties. Despite advances in high-performance computing (HPC) and code development, using DFT to compute anharmonic vibrations for transport or free energy predictions requires considerable computing resources. We illustrate two machine learning aided approaches that are useful in this regard:

- A) A support vector machine (SVM) is used to classify intrinsic defects in a DFT molecular dynamics simulation of an ultra-high temperature ceramic near the melting-point ( $T = 3800$  K). The concentration of different defect types identified by the classifier is useful for estimating the anharmonic free energy contributions associated with the intrinsic defects near the melting point.
- B) Gaussian process regression (GPR) is used to interpolate the Brillouin zone of a soft phonon mode based on limited experimental data points. Combined with harmonic DFT calculations, a good approximation to the temperature-renormalized phonon dispersion is computed, enabling the prediction of phase transition thermodynamics at minimal (harmonic level) computational cost.

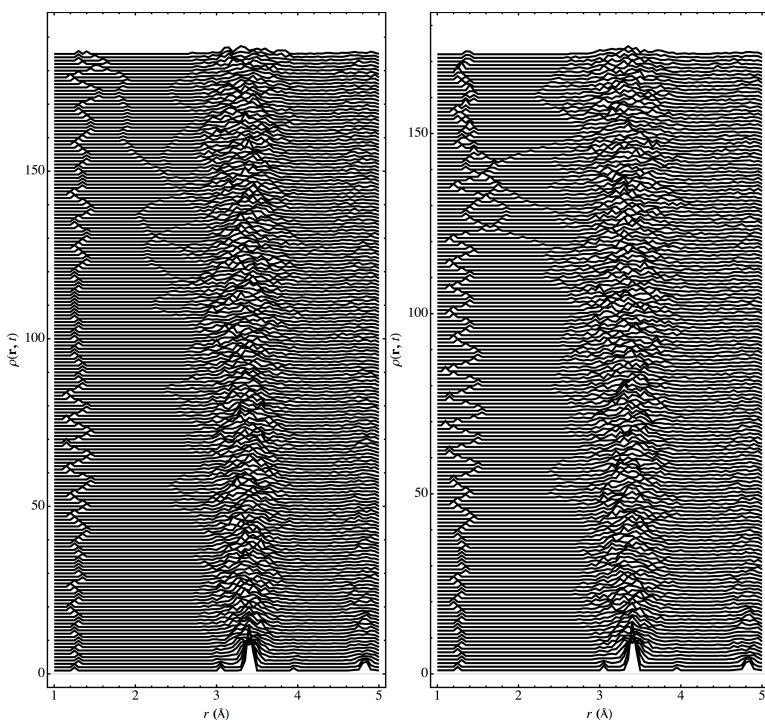


Figure 1: Evolution of the carbon-carbon pair correlation function over 200 fs of MD in a ZrC crystal at 3200 K. The data encodes Frenkel pair formation and is used to train a support vector machine (SVM).

# COMPETITIVE FORMATION OF HYDROGEN PEROXIDE AND WATER ON PALLADIUM(111): INSIGHTS FROM DENSITY FUNCTIONAL THEORY AND ARTIFICIAL NEURAL NETWORKS.

A. Sabadell,<sup>1</sup> M. Shahrokhi,<sup>1</sup> and N. López<sup>1</sup>

<sup>1</sup> Institute of Chemical Research of Catalonia, The Barcelona Institute of Science and Technology, Tarragona, 43003, Spain.

email: asabadell@icmq.es

Palladium-based catalysts are useful in many industrial processes, like the hydrogen peroxide synthesis. This process can be modeled by theoretical methods such as Density Functional Theory (DFT), which is a powerful tool to rationalize the development of new catalysts along with experimental procedures. However, DFT calculations comes with a high computational cost.[1]

Here, we implemented an Artificial Neural Network (ANN) that predicts the reaction barriers obtained through DFT. The chosen reaction network is the hydrogen peroxide synthesis, including the competing paths of the water formation. The active surface is Palladium (111), which is represented as an hexagonal lattice.

The ANN was trained with several parameters, including reactants on the surface, products markers, cell area, the energy of the clean surface, the number of electrons involved in each reaction, and the reactant coordinates on the cell.[2-4] The logistic function has been used as Neurons response, thus assuring smooth variations until convergence.[5-7] The reaction barriers obtained from the ANN are in good agreement to those computed with DFT, Figure 1. Since our database is expandable, it is possible to include results from other systems to improve the robustness and accuracy of this tool.

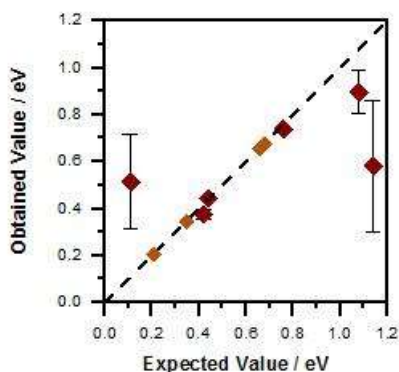


Figure 1. Reaction barriers obtained through ANN and DFT. Yellow (red) dots corresponds to the training (validation) set.

- [1] N. López *et al.*, *Catal. Sci. Technol.*, **2012**, 2, 2405–2417.
- [2] J. Behler and M. Parrinello, *Phys. Rev. Lett.*, **2007**, 98, 146401.
- [3] T. Derek, *Science*, **2018**, 36, 8075.
- [4] P. Raccuglia *et al.*, *Nature*, **2016**, 533, 73–76.
- [5] N. Castin and L. Malerba, *J. Chem. Phys.*, **2010**, 132, 074507.
- [6] J. S. Smith *et al.*, *Chem. Sci.*, **2017**, 8, 3192–3203.
- [7] F. Pascal *et al.*, *Sci. Rep.*, **2018**, 8, 2559.



## GENERALIZED MODEL FOR THE ADSORPTION OF INTERMEDIATES ON TRANSITION METALS

R. García-Muelas<sup>1</sup> and N. López<sup>1</sup>

<sup>1</sup> Institute of Chemical Research of Catalonia, The Barcelona Institute of Science and Technology, Tarragona, 43003, Spain.  
email: rgarcia@icmq.es

The rational design of new heterogeneous catalysts relies on the efficient survey of reaction networks by Density Functional Theory, DFT. However, massive reaction networks as those involved in the conversion of biomass cannot be sampled efficiently as they involve more than  $10^4$  intermediates even for a typical  $C_6$  sugar.[1–2] Here we present a statistical analysis applied on the thermochemical data of 71  $C_1$ – $C_2$  molecular fragments on 12 metal surfaces. The two main sources of thermochemical variability can be traced back to the  $d$ -band center and the oxidation potential. The robustness of the methodology was tested for a lower number of molecular descriptors. The full thermochemistry of a metal surface can be retrieved with mean absolute errors (MAE) between 0.08 and 0.12 eV.[3]

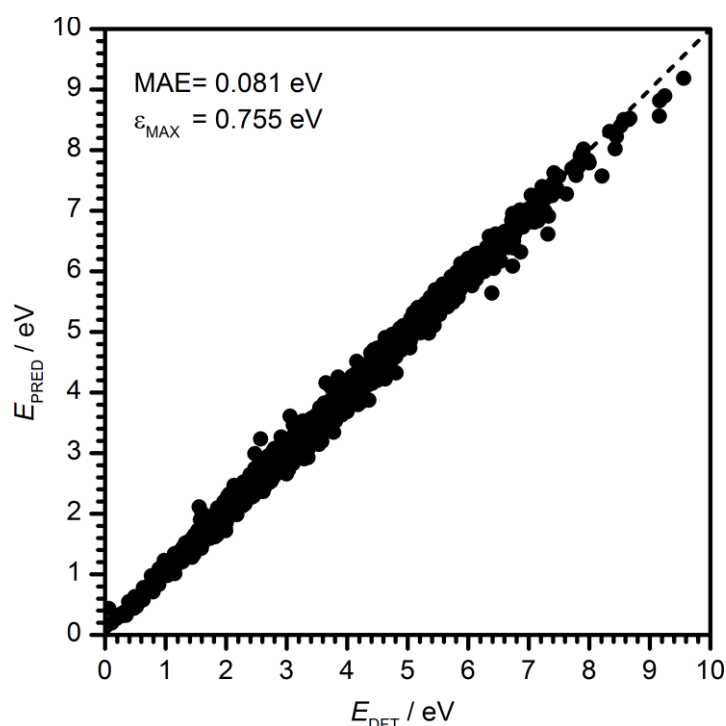


Figure 1. Parity plot of potential energies for  $C_1$ – $C_2$  intermediates on 12 metal surfaces.

[1] R. García-Muelas, Q. Li, and N. López. *ACS Catal.* **2015**, *5*, 1027–1036.

[2] Q. Li, R. García-Muelas, and N. López. *Nat. Comm.* **2018**, *9*, 526.

[3] R. García-Muelas and N. López. *Submitted*, **2018**.

# PREDICTING THE CURIE TEMPERATURE USING MACHINE LEARNING

J. Nelson and S. Sanvito

School of Physics and CRANN, Trinity College, Dublin 2, Ireland  
email: janelson@tcd.ie

When a ferromagnet is heated above the Curie Temperature ( $T_C$ ), its spontaneous magnetization abruptly falls to zero. For most practical applications involving ferromagnets a high  $T_C$  is required. However, at the moment there are a limited number of such magnets, thus any progress on speeding up the rate of discovery would be extremely valuable.

In this work we present our efforts at using Machine Learning to predict the  $T_C$  using solely the chemical formula of a material. The model was trained using 1500 experimental measurements of the  $T_C$  for different ferromagnets. On an independent test set consisting of 800 compounds the mean absolute prediction error was 65 K.

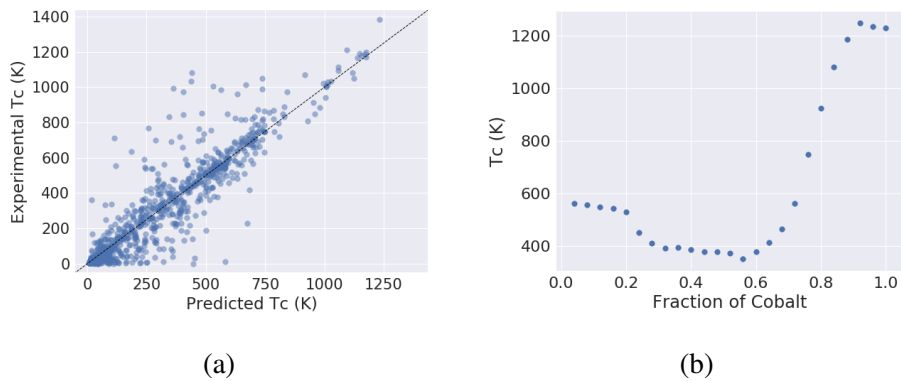


Figure 1: (a) The predictions on the  $T_C$  Test Set with  $R^2 = 0.85$ . (b) A plot showing how the predicted  $T_C$  varies with the fraction of Cobalt for the system Mn–Co.

## Accelerating atomistic structure search: gradient-assisted Bayesian optimization

H. Paulamäki<sup>1</sup>, M. Todorović<sup>1</sup>, Eero Siivola<sup>2</sup>, Aki Vehtari<sup>2</sup> and P. Rinke<sup>1</sup>

<sup>1</sup> Department of Applied Physics, Aalto University, Finland

<sup>2</sup> Department of Computer Science, Aalto University, Finland

email: henri.paulamaki@aalto.fi

In atomistic structure search it is often important to minimize the system's multidimensional potential energy surface (PES). Machine learning methods help to reduce the number of computationally costly static simulations needed for the task. We used Bayesian optimization on top of Gaussian process (GP) models [1] interfaced with atomistic simulation codes (BOSS method [2]) to find conformers of alanine dipeptide molecule. We studied the scaling of the efficiency of finding the global minimum structure when dimension of the PES is increased. We discovered an improvement in the scaling behaviour as gradient information of the PES function (forces) is added to the GP model [3].

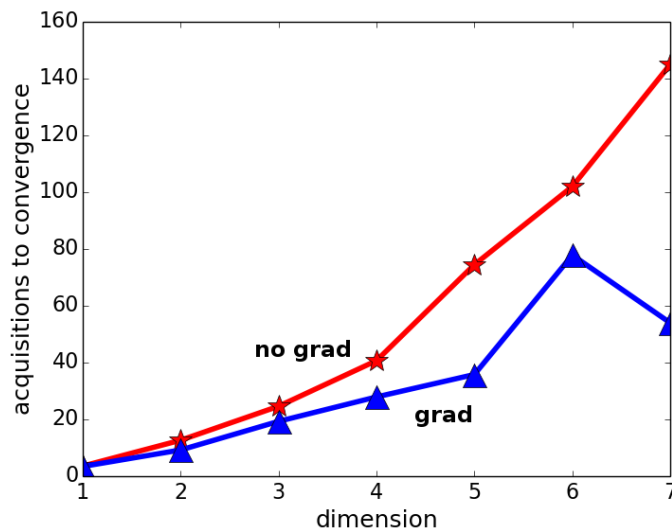


Figure 1: Average number of required acquisitions for global minimum convergence per dimension for non-gradient GP model and gradient-assisted model as found on dihedral angles of the alanine dipeptide molecule.

- [1] C. E. Rasmussen and C. Williams, Gaussian Processes for Machine Learning, *The MIT Press*, Cambridge, MA, USA (2006).
- [2] M. Todorović, M. U. Gutmann, J. Corander and P. Rinke, arXiv:1708.09274, [cond-mat.mtrl-sci] (2017).
- [3] E. Solak, R. Murray-Smith, W. E. Leithead, D. J. Leith and C. E. Rasmussen, Derivative observations in Gaussian process models of dynamic systems. *Advances in Neural Information Processing Systems 25*, pp. 1057-1064 (2003).



## Aalto University Otaniemi Campus Map

