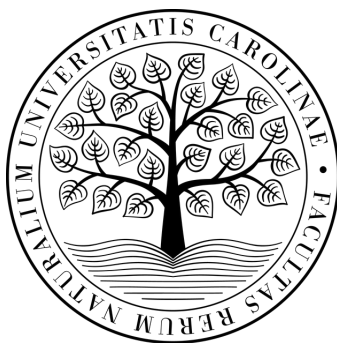


TOWARDS RELIABLE SIMULATIONS OF
NANOPOROUS MATERIALS UNDER OPERANDO
CONDITIONS

(Habilitation Thesis)



Lukáš Grajciar

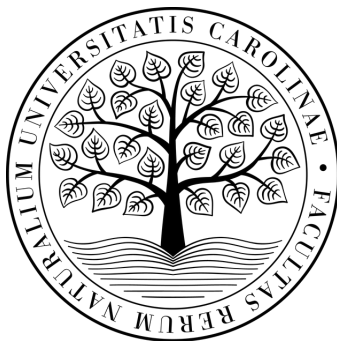
DEPARTMENT OF PHYSICAL AND MACROMOLECULAR CHEMISTRY,
FACULTY OF SCIENCE, CHARLES UNIVERSITY

Faculty of Science
Charles University

– January 2023 –

VÝVOJ A APLIKACE SPOLEHLIVÝCH SIMULACÍ PRO
POPIS NANOPORÉZNÍCH MATERIÁLŮ ZA
REALISTICKÝCH PROVOZNÍCH PODMÍNEK

(Habilitační práce)



Lukáš Grajciar

KATEDRA FYZIKÁLNÍ A MAKROMOLEKULÁRNÍ CHEMIE,
PŘÍRODOVĚDECKÁ FAKULTA, UNIVERZITA KARLOVA

Přírodovědecká fakulta
Univerzita Karlova

– leden 2023 –

I would like to express my gratitude to my teacher, close colleague and a friend Petr Nachtigall, who formed my way towards this habilitation and who passed away much too early.

I am also in debt to my family, especially my wife and kids, whose time with me was sacrificed.

Contents

1	Introduction	7
2	Methodological advances	10
2.1	"Better" - Dispersion corrections to DFT	10
2.2	"Stronger/Larger" - Large scale DFT calculations	13
2.3	"Harder" - Global optimization in confined spaces	17
2.4	"Faster" - Machine Learning Potentials and More	19
3	Applications	26
3.1	Accurate description of adsorption in nanoporous materials: a static approach	26
3.1.1	Benchmarking adsorption on coordinatively unsaturated metal sites	26
3.1.2	Coverage effects in MOF models	29
3.1.3	Embedding environment effects in zeolitic models	31
3.2	Towards reactivity and dynamical models	33
3.3	High-throughput calculations, material design and beyond	38
4	Conclusions and outlook	46
5	References	49
6	Attached papers	66
6.1	Low-memory iterative density fitting [Grajciar, 2015]	68
6.2	Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole method: Analytical gradients [Lazarski et al., 2016]	69
6.3	PbS clusters embedded in sodalite zeolite cavities of different compositions: Unraveling the structural evolution and optical properties using ab initio calculations [Grajciar, 2016]	70

6.4	Towards operando computational modeling in heterogeneous catalysis [Grajciar et al., 2018]	71
6.5	Accurate large-scale simulations of siliceous zeolites by neural network potentials [Erlebach et al., 2022]	72
6.6	Constructing collective variables using invariant learned representations [Sípka et al., 2023]	73
6.7	The role of water loading and germanium content in germanosilicate hydrolysis [Jin et al., 2021b]	74
6.8	Identification of the most active sites for tetrahydropyranylation in zeolites: MFI as a test case [Liu et al., 2020]	75
6.9	Fast room temperature lability of aluminosilicate zeolites [Heard et al., 2019b]	76
6.10	Mechanism of zeolite hydrolysis under basic conditions [Jin et al., 2021a]	77
6.11	Zeolite (in)stability under aqueous or steaming conditions [Heard et al., 2020]	78
6.12	The effect of water on the validity of Löwenstein’s rule [Heard et al., 2019a]	79
6.13	The Lewis acidity of three- and two-dimensional zeolites: The effect of framework topology [Ho et al., 2018]	80
6.14	The Brønsted acidity of three- and two-dimensional zeolites [Thang et al., 2019]	81

1 Introduction

Nanoporous materials play a crucial role in a wide range of applications in chemical engineering, chemistry, and material science. The prime example of such materials are zeolites, the crystalline aluminosilicates composed of covalently bound tetrahedral networks, acting as heterogeneous catalysts with confinement-induced reactivity and selectivity, which are one of the most important industrial catalysts and are used nowadays, in particular, in crude oil processing and petrochemistry. In addition, zeolites have already found or are expected to find utilization in sustainable fields including production of chemicals and fuels using abundant and renewable feedstocks (biomass conversion), thermal energy storage, water purification and desalination, NO_x emission control from exhaust fumes and CO₂ capture and conversion [Li et al., 2017]. The advantages of zeolites are that they are environmentally friendly, reasonably cheap, and thermochemically stable. The space of zeolite framework topologies is large theoretically, however, only about 250 different zeolite frameworks have been synthesized so far, mostly in a trial-and-error fashion. This limited flexibility is partially offset by the possibility to tune the character of the extra-framework cations, compensating the negative charge of the framework, as well as to engineer the aluminum content and distribution. In addition, the silicon or aluminum atoms can be isomorphously substituted for other elements (such as B, Ge, Sn, Ti, *etc.*), which further extends the versatility of zeolites. Lately, there is a growing demand for new zeolites for much more diverse applications than before. Even though zeolites have been studied for decades, understanding of their (trans)formation and decomposition mechanisms at the molecular level is still lacking [Heard et al., 2020].

Another very important and recently extremely popular class of nanoporous materials are the metal-organic frameworks (MOFs), joining the family of crystalline nanoporous materials rather recently. The MOFs are a subclass of coordination polymers made of metal-containing (metal or metal oxide clusters) nodes coordinated to the bridging organic linkers (such as carboxylic acids or azols). Their strengths are the ease of their synthesis and the huge tunability of their physiochemical properties (surface area, sorption capac-

ity, catalytic activity, *etc.*) thanks to a large variety of organic and metal-containing building blocks that could be coordinated together [mof, 2012]. However, MOFs typically suffer from lower thermochemical stability and high production costs which translates into a fact that MOFs have not been yet commercialized in larger scale [Kokcam-Demir et al., 2020]. Similarly to zeolites, a lot of effort has focused on synthesizing and designing MOFs (ideally *in silico*) with targeted properties.

This habilitation thesis maps my efforts in developing and applying methods that fill the gap in our current understanding of processes taking place in the nanoporous materials (in particular zeolites and MOFs) at the atomistic level, with the ultimate aim to enable the rational design of these nanoporous materials. This journey could be characterized as continuous attempt to strike a balance between the complexity of the model system (*i.e.*, typically more complex means more realistic) and the complexity of the method (*i.e.*, typically more complex means more accurate), so that our computational results could be readily experimentally verifiable/falsifiable. In our earlier studies the pendulum swung more towards the side of the method accuracy, as we tried to describe as accurately as possible the physisorption (or weak chemisorption) of various small gas molecules in few reference nanoporous materials (3.1), developing the dispersion-corrected density functional theory (DFT) methods (2.1) and carrying out a lot of so-called static calculations (typically just local geometry optimization) on a set of reasonably representative structures. Later, the pendulum swung more towards our aim to describe more complex (realistic) systems, *e.g.*, zeolites interacting with aqueous solutions or modelling defective zeolites (3.2). To reliably model such systems, a dynamical description, *e.g.*, using (biased) *ab initio* molecular dynamics, is often necessary, enabling a more thorough sampling of the configuration space that accounts for temperature effects and which allows for more open-ended search for accessible reaction/transformation paths. However, in both of the above-mentioned directions, one was still limited, due to large computational costs, to a few assumingly representative systems. We tried to assess this problem initially by extensive high-throughput-like DFT-based calculations for broader range of systems (3.3). However, to properly reconcile this conflict between the complexity of model and the complexity of the method, we have recently started to leverage the power of machine learning for our purposes (via developing neural network potentials and ML-driven collective variables - see Chapter 2.4), which are expected to allow us to carry out

truly comprehensive studies of nanoporous materials such as zeolites under realistic conditions maintaining an *ab initio* level of accuracy (3.3).

The habilitation thesis is composed of two parts - the first part focuses on my efforts towards developing novel methods for simulation of (not only) nanoporous materials (chapter 2), while the second part highlights some of my application results in the field of nanoporous materials (Chapter 3). At the end of each (sub)chapter, I provide, in a two short paragraphs, i) a specification of my own contribution to the particular topic including a list of related articles co-authored by me, and ii) the importance, the broader context and the future outlook of the particular methodological development or the particular application finding. The articles on which this habilitation thesis is based are attached at the end of the thesis in the Chapter 6.

The text of habilitation thesis discusses my more recent work in the broad context of my career including my PhD years, which are summarized in subchapters 2.1 and 3.1. The publications from my PhD years are not *per se* a subject of thesis and none of them is included in the list of the attached papers (Chapter 6).

2 Methodological advances

*“Work it harder, make it better
Do it faster, makes us stronger.”*

— Daft Punk. “Harder, Better, Faster, Stronger.” 2007

2.1 "Better" - Dispersion corrections to DFT

The attractive Van der Waals (vdW) dispersion interactions, arising from electrostatic interactions between fluctuating electron densities, are omnipresent in molecules and materials. The dispersion interactions are essential not only for description of binding in molecules and materials but has been shown to influence wide range of seemingly unrelated properties (spectroscopic, mechanical, electronic or kinetic) [Hermann et al., 2017, Xu et al., 2020]. Hence, the need to describe them accurately straightforwardly extends to nanoporous materials, be it in the context of adsorption, reactivity/catalysis, or phase transformation.

Unfortunately, proper description of dispersion interaction, which is a many-body correlation effect, is difficult and necessitates deployment of high-level correlated calculations, which can be, due to exorbitant computational costs, applied only to rather small molecules (typically tens and up to very low hundreds only if extensive supercomputer resources are available). Use in condensed systems, such as nanoporous materials, is thus mostly precluded and approximate treatments, typically within the context of density functional theory (DFT) methods, are necessary. This is still an active field of research despite it maturing significantly in the recent years [Caldeweyher et al., 2019, Hermann and Tkatchenko, 2020, Kim et al., 2020]. A useful classification of a "zoo" of various approaches put forward in ref. [Hermann et al., 2017] is the following: i) the exact treatment via the adiabatic-connection fluctuation-dissipation theorem within DFT framework (with a necessity to approximately remove double-counting of correlation energy in the short-range), ii) coarse-grained (*e.g.*, atom-based) many-body vdW methods, iii) fine-grained (infinitesimal) pairwise density functional approximations, and iv) coarse-grained (*e.g.*, atom-based) pairwise vdW methods. The last class of atom-based pairwise vdW methods, represented by a family of *DFT-D* methods of Grimme et al. [Caldeweyher et al., 2019], is cost- and implementation-wise

least demanding, and despite a typically significant degree of semi-empiricism involved, has been shown to provide acceptable accuracy in most applications (excluding mostly extended low-dimensional systems or materials with non-cubic symmetry) and is (unfortunately) still mostly a method of choice for the majority of users in the (nanoporous) materials field. Our own approach, the DFT/CC method [Bludský et al., 2008, Grajciar et al., 2010], falls also within this last class of dispersion correction methods.

The DFT/CC method, introduced first by Bludský et al. [Bludský et al., 2008], is a coarse-grained pairwise vdW-*like* method that expresses the vdW-*like* correction ΔE to interaction energy between molecules (or material and molecule) derived from a pure semi-local DFT functional (typically a PBE functional was used) as a sum of atomtype-atomtype distance-dependent pairwise corrections (generalization to angle dependence or different atom types is straightforward):

$$\Delta E = \sum_{i \in \mathbf{A}} \sum_{j \in \mathbf{B}} \epsilon_{At(i), At(j)}, \quad (2.1)$$

which is similar to the approach taken by the original DFT-D1 [Grimme, 2004] and DFT-D2 [Grimme, 2006] works by Grimme. In contrast to those Grimme’s approaches, the goal of the DFT/CC method was not universality but a very high accuracy for specific systems of interest with an ability to correct for other deficiencies of semi-local DFT functionals beyond missing dispersion and focusing on consistently accurate description of larger portions of the potential energy surface (PES) beyond the equilibrium distances only. To obtain the atomtype-atomtype correction functions $\epsilon_{At(i), At(j)}$ a gridded reference data (at highly-accurate correlated CCSD(T) level) were generated and the correction functions were fitted on this reference data using a robust reciprocal power Reproducing Kernel Hilbert Space (RKHS) interpolation [Ho and Rabitz, 1996]:

$$\epsilon_{ij}(r_{ij}) = \sum_{k=1} \alpha_k^{ij} q(r_{ij}^2, r_k^2), \quad (2.2)$$

where k runs over the gridded reference points, α_k^{ij} are coefficients to-be-fitted and the kernel function is defined as:

$$q(r_{ij}^2, r_k^2) = \frac{1}{3r_{>}^6} - \frac{r_{<}^2}{5r_{>}^8}, \quad (2.3)$$

with $r_{<} = \min(r, r')$ and $r_{>} = \max(r, r')$. This form of kernel function guarantees correct r^{-6} asymptotics of the correction, as it constructed as a vdW-*like* correction. From today's perspective, our DFT/CC approach could be considered as an early attempt at delta-machine learning approach [Ramakrishnan et al., 2015] (which constructs a correction surface between high- and low-level theories) that utilizes nowadays well-established and very popular machine learning method (RKHS), which is today better known under different names, *i.e.*, as kernel ridge regression (KRR) or as a Gaussian process regression (GPR) [Unke et al., 2021b]. Hence, it seems we were a bit ahead of time with the DFT/CC approach.

The DFT/CC method has been successfully applied to study interactions of hydrocarbons [Bludský et al., 2008] and adsorption of small gas molecules on (hydro)carbons [Rubeš et al., 2010] and in extended materials (zeolites and metal-organic frameworks) [Grajciar et al., 2012, 2015a] exhibiting unprecedented accuracy. It has been also later extended and coupled with vdW non-local functionals [Hermann and Bludský, 2013] to improve originally rather cumbersome way of creating atomtype-atomtype corrections functions for systems with multiple atomic types. However, it appears that with the advent of machine learning, much more streamlined and automatized big data extraction procedures are available and with the focus on universality rather than specificity (at least in the vdW-corrected DFT field) the DFT/CC method is now superseded with more clearly data-oriented approaches such as machine learning thermodynamic perturbation theory [Herzog et al., 2022] and by the universal vdW correction schemes [Hermann and Tkatchenko, 2020, Kim et al., 2020, Caldeweyher et al., 2019].

My contribution to the DFT/CC was in its extension: i) for description of adsorption in complex materials (metal-organic frameworks) with multiple atomic types, and ii) for an explicit use of DFT/CC as general correction scheme to standard DFT moving beyond the vdW-correction only. Thus, *e.g.*, I was able to achieve unprecedented level of accuracy for description of adsorption of small molecules in computationally very challenging metal-organic frameworks with open-metal sites as illustrated in Figure 1 (see Chapter 3.1 for more details). Also, it represented my first encounter with the machine learning approaches that I built upon later down the line (see Chapter 2.4).

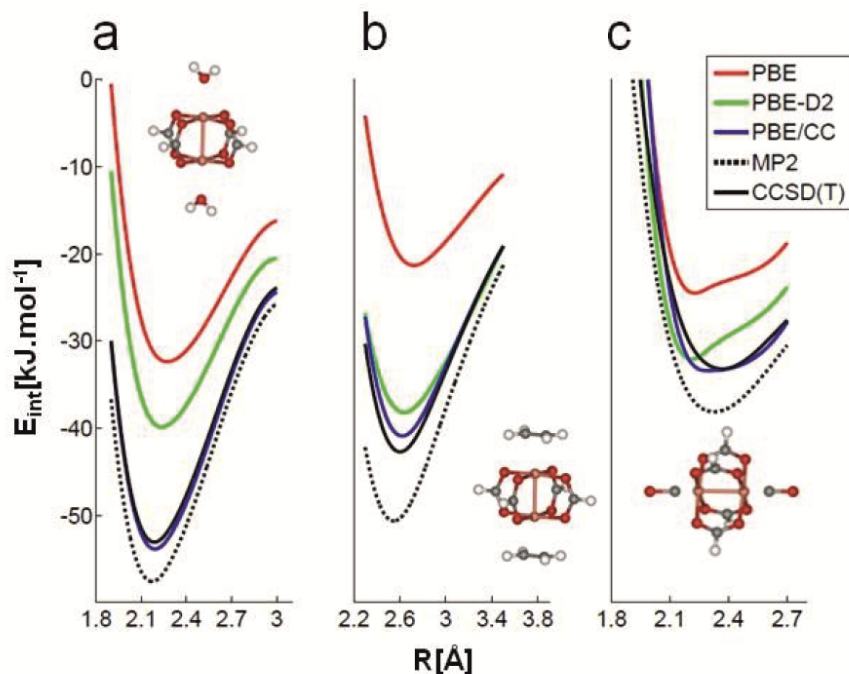


Figure 1: The performance of the DFT/CC method tested on the 1-D dissociation curves of various adsorption complexes of small molecules with the paddlewheel model of the metal cluster in a reference MOF named CuBTC: (a) water, (b) ethane and (c) carbon monoxide. Figure adapted from [Grajciar, 2013].

2.2 "Stronger/Larger" - Large scale DFT calculations

The ability to routinely treat large nanoporous systems such as zeolites, metal-organic frameworks, or covalent-organic frameworks at the standard (semi-local) DFT level is still a challenge as many of those systems have unit cells containing high hundreds to low thousands of atoms. In addition, these systems are sparsely packed making the use of standard plane-wave based periodic DFT implementations inefficient [Kudin and Scuseria, 2000]. In this context the use of compact localized atom-centered basis sets such as Gaussian-type orbitals (GTOs) is more "natural" and allows for consistent treatment of both molecular and periodic systems of any dimensionality (*i.e.*, of 1D nanowires such as nanotubes or of 2D surfaces) without a need to use

artificial 3D periodic models.

However, the use of GTOs comes with a caveat in a need to evaluate computationally expensive four-center electron repulsion integrals (ERIs) constituting the Coulomb (and exchange) contributions. One of the most successful ways how to reduce the cost of ERI evaluations is to use a density fitting (DF) procedure [Whitten, 1973] which approximates electron density $\rho(\mathbf{r})$ (a two-center entity) as linear combination of atom-centered GTOs from the "so-called" auxiliary basis set $\xi_\alpha(\mathbf{r})$:

$$\rho(\mathbf{r}) \approx \tilde{\rho}(\mathbf{r}) = \sum_{\alpha} c_{\alpha} \xi_{\alpha}(\mathbf{r}), \quad (2.4)$$

where linear expansion coefficients c_{α} are obtained by minimizing the error defined as $(\rho - \tilde{\rho}|w|\rho - \tilde{\rho})$ with respect to metric w . All-in-all this leads to system of linear equations:

$$\mathbf{V}^{\mathbf{w}} \mathbf{c} = \gamma^{\mathbf{w}}, \quad (2.5)$$

where $\gamma_{\mathbf{w}}$ is a vector with elements $\gamma_{\alpha}^w = (\xi_{\alpha}|w|\rho)$ and $\mathbf{V}^{\mathbf{w}}$ is a matrix with elements $V_{\alpha\beta}^w = (\xi_{\alpha}|w|\xi_{\beta})$. Thus, DF turns the four-center ERIs into a sum of two- and three-center ERIs that can be evaluated order of magnitude faster [Eichkorn et al., 1995] for small and mid-size systems (tens to low hundreds of atoms) as the cost of solving Equation 2.5 is small for those systems. However, the cost increases for very large systems as: i) inversion of Equation 2.5 scales as $O(N_{aux}^3)$ with the size of auxiliary basis set N_{aux} , albeit with very small prefactor, and ii) the memory demand scale as $O(N_{aux}^2)$ as the matrix $\mathbf{V}^{\mathbf{w}}$ is dense. The problem is compounded by the fact that auxiliary basis sets are three to five times larger than orbital basis sets. In addition, in periodic calculations even the need to evaluate large number of three-center ERIs becomes a computational bottleneck very fast. Multiple workarounds have been proposed in the literature based mostly on localizing the auxiliary basis set expansion such as using a metric w different than the optimal Coulomb metric [Jung et al., 2005, Luenser et al., 2017] or explicitly partitioning fitted densities into localized subsystems and fitting them separately [Sodt et al., 2006, Pisani et al., 2008]. However, these approaches have their own caveats such as increased error of the fit (using non-Coulomb metric) or introduction of the structure-dependent error that can lead to discontinuities in the potential energy surface.

Our approach for DF of large molecular and periodic systems is to use the global Coulomb metric with a key ingredient being the continuous fast multi-

pole method (CFMM) [Kudin and Scuseria, 2000]. The CFMM can evaluate a large portion of Coulomb long-range interactions very efficiently using multipole expansions rather than by calculating costly ERIs and can achieve near-linear scaling for evaluation of the Coulomb term. Our combination of DF and CFMM for periodic systems was coined as DF-CFMM [Lazarski et al., 2015, 2016] and the approach for large molecular systems was termed as low-memory iterative density fitting (LMIDF) [Grajciar, 2015] as it included also an implementation of the iterative solution of Equation 2.5 using preconditioned conjugate gradient (CG) method. In the LMIDF, by evading direct inversion of Equation 2.5 we alleviate a problem with $O(N_{aux}^3)$ scaling of the direct solution and by using the CFMM to evaluate the majority of the $\mathbf{V}^w\mathbf{c}$ products *on-the-fly*, making the remaining part of the \mathbf{V}^w matrix sparse, we mitigate the $O(N_{aux}^2)$ memory demand. For the best conjugate-gradient preconditioner (*SP block*), providing optimal trade-off between its storage/calculations costs and the ability to decrease the number of CG iterations (which necessitate repeated evaluation of the $\mathbf{V}^w\mathbf{c}$ products), we achieve at least 7-fold reduction of memory demands compared to standard DF with at least 13-times smaller prefactor for the $O(N_{aux}^3)$ step. The cost of the iterative solution of Equation 2.5 is a modest increase (by a few tens of percent) of total computation time due to repeated evaluation of the $\mathbf{V}^w\mathbf{c}$ products, which, however, decreases with the system size. The upshot is that using the LMIDF one can routinely carry out calculation of very large molecules (or molecular fragments) containing few thousands of atoms described by over 100,000 auxiliary basis functions on a single computational node equipped with a memory of a standard size (high ones and low tens of GB per CPU core). We demonstrated this by performing a DFT single point calculation for a large chabazite zeolite fragment containing 2592 atoms with the electron density expanded using 121,248 auxiliary basis functions and all this done on a single 12-core computational node with 128 GB of memory only (see Figure 2). The ability to run such calculations on a single node is of particular importance for studies that need to perform many runs in parallel such as global structure optimization (see Chapter 2.3) or high-throughput screenings (see Chapter 3.3).

Since their implementations DF-CFMM and LMIDF approaches have found applications mostly in description of covalent organic frameworks [Gottschling et al., 2020], molecular crystals [Stein and Heimsaat, 2019] and zeolites [Grajciar, 2016]. In addition, I expect their potential to lie also in description of

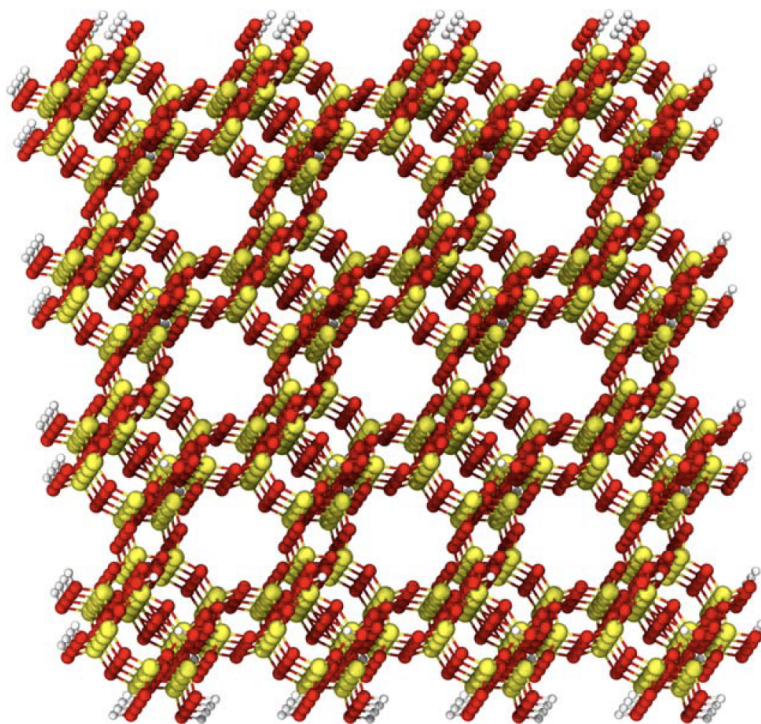


Figure 2: The example of an extremely large cluster (of the zeolite chabazite) containing 2592 atoms and 121,248 auxiliary basis functions, for which a single point DFT calculation could be performed, using the low-memory iterative density fitting developed by me [Grajciar, 2015], on a single 12-core CPU node equipped with 128 GB memory only. Figure adapted from [Grajciar, 2015].

supramolecular systems possibly within the QM/MM or QM/QM framework either as a one-off calculation or carried out within a high-throughput study. Lastly, the DF-CFMM implementation has become a bedrock on which new capabilities were built including calculation of stress tensors [Becker and Sierka, 2019], wave-function-in-DFT embedding [Sharma and Sierka, 2022] and implementation of real-time time-dependent DFT [Müller et al., 2020].

The development of the LMIDF was solely my own undertaking [Grajciar, 2015] and I implemented the most-complex and the most important part of the analytical gradients, the Coulomb term gradients, for the DF-CFMM

method [Lazarski et al., 2016]. The analytical gradients enable calculation of forces acting on atoms and are needed for structure optimizations or the molecular dynamics runs. Importantly, the implementation was done within one of the leading quantum chemistry packages on the market, the TURBOMOLE [TUR], which: i) enabled me to gain experience in developing methods in the high-performance environment, ii) provided me with the contacts in the method-development field, and iii) allowed my implementations to gain broader audiences. These developments were funded by internal TURBOMOLE project and the Deutsche Forschungsgemeinschaft Grant [No. 269386423](#), in both of which I was a principal investigator.

2.3 "Harder" - Global optimization in confined spaces

Reliable characterization of the atomic structures of nanoporous materials and in particular of the structures of molecules and clusters embedded inside the nanoporous materials is a key prerequisite for understanding their physical and chemical properties. One way how to solve this non-trivial problem is to perform a global structure optimization (GSO) search [Wales, 2004], in which a structure-dependent cost function $E(\mathbf{x})$, typically a potential energy of a system, is globally minimized over a breadth of the structures \mathbf{x} providing the putative global minimum structure as well as a library of low-lying local minima obtained during the optimization run. Since the cost function $E(\mathbf{x})$ over the structure space is typically highly non-linear and non-convex, GSO algorithms are highly heuristic with many system-specific tunable hyperparameters (convergence and acceptance criteria, moveclass types, local minimization algorithm, *etc.*). The two main classes of these heuristic GSO methods are Monte Carlo based methods such as basin-hopping [Wales and Doye, 1997] or minima-hopping [Goedecker, 2004] and nature-inspired GSO methods represented mainly by the genetic/evolutionary algorithms [Heiles and Johnston, 2013]. The main application fields for GSO searches are in structure determination of [Grajciar et al., 2018]: i) free-standing clusters/nanoparticles, ii) external surfaces, and iii) surface-deposited/surface-adsorbed clusters/molecules with a focus on globally optimizing the structure of the clusters/molecules rather than allowing for a surface-reconstruction. However, the application of GSO methods to structure determination of clusters/molecules confined in the embedding matrix (*e.g.*, nanoporous material such as zeolites) has been largely nonexistent, with a single exception [Vil-

helmsen et al., 2012] that basically just re-used the older GSO implementation for surface-deposited nanoclusters.

Therefore, the aim was to develop a tool for global structure optimization in confined spaces, GOCONS, which allows for locating of the most stable structures of molecular and nanoparticulate systems confined in host materials in a computationally efficient way. The GOCONS tool has been implemented as an extension of the DoDo program [Sierka, 2010] for global structure optimizations using genetic algorithm and being a part of the TURBOMOLE program package [TUR]. The extensions over standard GSO methods for surface-deposited clusters were: i) modification of the cut-and-splice crossover operator that would reflect the fact that the stability of confined clusters is not only a function of their structure but also of their location and orientation within the host (in particular in case of the strong interaction with the environment), ii) robust and automatic definition of the void space in the host used for generation of the candidate structures based on alpha shape theory [Edelsbrunner et al., 1998] and weighted Voronoi decomposition, and iii) employment of the LMIDF [Grajciar, 2015] to enable running DFT calculations within GOCONS in parallel even for such large cluster-host systems (see Chapter 2.2). The applicability of the GOCONS tool was later demonstrated for global structure determination of lead sulfide quantum dots confined in the zeolite host with sodalite topology [Grajciar, 2016] (see Chapter 3.1 for more details).

The GSO techniques are most relevant for systems at low temperatures, in low pressure environments and without large solvent effects - for such systems and conditions a putative global minimum along with few low-lying minima do have outsize Boltzmann weights in the ensemble and thus represent the system in question reasonably well. However, at increased temperatures and pressures, with strong solvent effects or for systems undergoing reconstruction at such reactive *operando* conditions, the applicability of GSO is limited mostly to qualitative screening of candidate structures then needs to be refined by methods able to describe the free energy surface quantitatively such as molecular dynamics or Monte Carlo [Grajciar et al., 2018]. Since I have been lately interested in complex systems characterized by strong solvent effects at increased temperatures (see Chapter 3.2) my interest in GSO has faded and when needed I opted out [Hou et al., 2020] for simpler and more robust, albeit less fine-tuned and more brute-force, approaches such as basin-hopping [Wales and Doye, 1997]. However, there have been lately

few interesting GSO developments related to my current focus on using machine learning methods in chemistry (see Chapter 2.4) that may rekindle my interest in GSO such as: i) a GOFEE method leveraging cheap machine learning surrogate models to speed up local optimizations while mostly retaining the accuracy of the high-level (DFT) method [Bisbo and Hammer, 2020], or ii) use of robust non-heuristic global optimizers based on covariance matrix adaptation evolution strategy algorithm [Arrigoni and Madsen, 2021].

The development and the test applications of the GOCONS tool was solely my own undertaking [Grajciar, 2016] done within the Deutsche Forschungsgemeinschaft Grant No. 269386423, in which I was the principal investigator. In addition, thanks to this project I gained a broad overview of the GSO field, of its strengths and weaknesses and it thus put me in good position to write a review article that included GSO as a one of its main topics [Grajciar et al., 2018].

2.4 "Faster" - Machine Learning Potentials and More

With the increasing availability of big data in chemistry, the application of machine learning (ML) in chemistry has become a huge trend in the last decade and despite a significant hype associated with it, its usefulness is now largely undeniable. This is illustrated in multiple application domains of ML in chemistry ranging from material design and retrosynthesis predictions to autonomous experimentation [Moosavi et al., 2020, Strieth-Kalthoff et al., 2020]. The most of ML algorithms can be thought as universal regularized interpolators that are really good at inferring implicit knowledge from (big) data and are very efficient in removing redundancies, *i.e.*, in reducing dimensionality of the data [Keith et al., 2021]. However, as interpolators, ML algorithms can be prone to artifacts, still need significant number of often expensive data (from quantum chemistry calculations), are difficult to interpret in order to gain some high-level conceptual insight and have limited extrapolation capabilities. Hence, the ML is not a "silver bullet" to our experimental and computational limitations and problems but, if handled cautiously and with a regard for its limitations, can be used, *e.g.*: i) to derive statistically significant predictive models from large data sets that could allow for refinement and easier understanding of complex problems such as zeolite synthesis design [Schwalbe-Koda et al., 2021, Jablonka et al., 2020], or ii) to significantly accelerate computational simulations of atomistic sys-

tems using *machine learning potentials* [Unke et al., 2021b] allowing us to adopt more realistic models or to carry out high-throughput calculations on large databases such as on a database of existing and hypothetical zeolites containing more than three hundred thousand structures [Erlebach et al., 2022].

The machine learning potentials (MLPs), also known as ML-based force fields, represent the mapping $E(\mathbf{x})$ between the (chemical) structures and their potential energy and belong to a class of supervised ML as the reference (typically quantum chemical data) are needed for their training. There are two big classes of MLPs, the *kernel-based* MLPs and the *neural network based* (NN) MLPs [Unke et al., 2021b, Keith et al., 2021]. In both cases one tries to find a general function approximator linking structure descriptors to energy $\tilde{E}(\mathbf{x})$. For the kernel-based methods such approximator is expressed as a linear approximation:

$$E(\mathbf{x}) \approx \tilde{E}(\mathbf{x}) = y(\mathbf{x}) = \sum_i c_i K(\mathbf{x}, \mathbf{x}_i), \quad (2.6)$$

where c_i are coefficients, and $K(\mathbf{x}, \mathbf{x}_i)$ is a non-linear kernel that evaluates similarity between reference data points (structure descriptors) \mathbf{x}_i and the previously unseen input (structure descriptor) \mathbf{X} in some high-dimensional "feature space" (see example of such kernel in Equation 2.3). This then leads to a linear system for coefficients c_i :

$$\mathbf{c} = (\mathbf{K} - \lambda \mathbf{I})^{-1} \mathbf{y}, \quad (2.7)$$

where strength of regularization (noise) is controlled by hyper-parameter λ . Equation 2.7 is solved by matrix factorization which, however, scales unfavorably with number of reference data points ($\mathcal{O}(N^3)$ for standard Cholesky decomposition or as $\mathcal{O}(N^2)$ for iterative solvers). For neural network based MLPs the general function approximator can be expressed as:

$$\begin{aligned} \mathbf{y} &= \mathbf{W}_n \mathbf{h}_{n-1} + \mathbf{b}_n \\ \mathbf{h}_i &= \sigma(\mathbf{W}_i \mathbf{h}_{i-1} + \mathbf{b}_i) \\ \mathbf{h}_0 &= \sigma(\mathbf{W}_0 \mathbf{x} + \mathbf{b}_0) \end{aligned} \quad (2.8)$$

where \mathbf{h}_i denotes i -th "hidden" layer of the network, weights \mathbf{W}_i and biases \mathbf{b}_i are parameters of i -th layer and σ is a non-linear activation function. The optimal values of weights \mathbf{W}_i and biases \mathbf{b}_i are found, *i.e.*, the *NN-based*

MLPs are trained, by minimizing the loss function that evaluates the error between the reference data and the MLPs’ predictions. This is typically done iteratively using (stochastic) gradient descent.

Both *kernel-based* and *NN-based* MLPs have their strengths and weaknesses and it is not possible to recommend any of those two as a general purpose approach. The *kernel-based* MLPs are typically more data-efficient than *NN-based* MLPs, *i.e.*, they need less reference data to reach required accuracy. One of the reasons is the fact that it is much more straightforward to incorporate known symmetries, conservation laws, invariances, or asymptotic behaviors into *kernel-based* MLPs, *i.e.*, into their kernels (*e.g.*, see Eq. 2.3), which all constrain the function space searched. However, *NN-based* MLPs can handle much larger data sets and are more flexible, which becomes particularly important if a good similarity metric, *i.e.*, a good kernel, is not known *a priori*.

The MLPs are becoming increasingly popular in multiple areas of material science as they enable rapid [Unke et al., 2021b, Artrith, 2019, Deringer et al., 2021]: i) structural and spectroscopic (IR or Raman spectra) characterization of the material including evaluation of the complete phase diagrams, ii) quantification of the chemical reaction dynamics (*e.g.*, via free energy profiles) for realistic and complex systems (*e.g.*, reactions at the liquid-solid interface), iii) routine evaluation of the nuclear quantum effects, or iv) evaluation of transport processes (*e.g.*, diffusion) at the interfaces. The reason for the success of MLPs is their ability to reach *ab initio* accuracy at the cost of the standard analytical reactive force fields (*i.e.*, being few orders of magnitude faster than DFT) using established and robust training and data curation procedures. Also, the MLPs are inherently reactive as they are able to learn the structure-energy map (see, *e.g.*, Eq 2.6) without a need for an *a priori* expert knowledge about particular chemical bonding patterns. However, many challenges in MLP development still remain [Poltavsky and Tkatchenko, 2021] including the problems of accurately describing the long-range interactions (*i.e.*, combating the multiscale nature of real interactions), generating balanced reference data sets that would deliver good performance across the potential energy surface (PES), interpreting/explaining the MLPs, and most importantly, the problem of limited MLP transferability and generality across the configurational and chemical space. Hence, the vast majority of the MLP studies still typically focuses on systems with rather low-dimensional configurational and chemical space such as simple metals, binary alloys, a single

chemical reaction, non-reactive interaction of a solvent with a simple surface (metallic, alloy or simple oxide) or some more complex systems but staying close to a specific thermodynamic state point [Artrith, 2019, Schran et al., 2021]. This apparent curse of dimensionality could be behind the fact that MLPs for nanoporous solids such as zeolites are, to our best knowledge, basically non-existent with one exception [Eckhoff and Behler, 2019].

Therefore, in our work, we embarked on the development of the linear scaling reactive MLPs for various classes of zeolites. Our MLPs are *NN-based* using the message-passing graph convolutional SchNet architecture [Schütt et al., 2018] with a end-to-end trainable environment representation that can seamlessly incorporate multiple atom types and that has been shown to provide high accuracy and high data efficiency. We generated a large, curated library (few hundred thousand structures) of density functional theory (DFT) energetic and force data for siliceous materials [Erlebach et al., 2022], germanosilicates, aluminosilicates zeolites with water [Saha et al., 2022] and silicious materials with platinum clusters. The library was created by a strategy [Erlebach et al., 2022] that attempts to optimally cover the important parts of the configurational space including not only structures close to equilibria both also non-equilibrium structures necessary for description of bond-breaking events, making our NN-based MLPs truly reactive. Figure 3 showcases and briefly explains the strategy for a case of siliceous zeolitic materials. The resulting NN-based MLPs are able to retain the accuracy of DFT calculations across the complex zeolitic classes considered, *e.g.*, outperforming specialized ReaxFF force fields by order(s) of magnitude in accuracy, while speeding up the calculations in comparison to DFT by at least three orders of magnitude. In addition, our tests show that our NNPs are qualitatively correct even in the parts of the configurational space sparsely covered by the database, *i.e.*, they seem to exhibit transferability and generality across rather broad configurational and chemical space, defying thus the curse of dimensionality to the extent that we can start working with realistic zeolitic models under realistic conditions (such as zeolites undergoing decomposition under hot liquid water conditions).

Using our newly developed NNPs we have been already witnessing intriguing results such as: i) revision of the Deem zeolite database (330 thousand hypothetical zeolites) revealing more than 20 thousand additional hypothetical frameworks in the thermodynamically accessible range for zeolite synthesis [Erlebach et al., 2022], ii) reactive diffusion and sintering of sub-nanometer

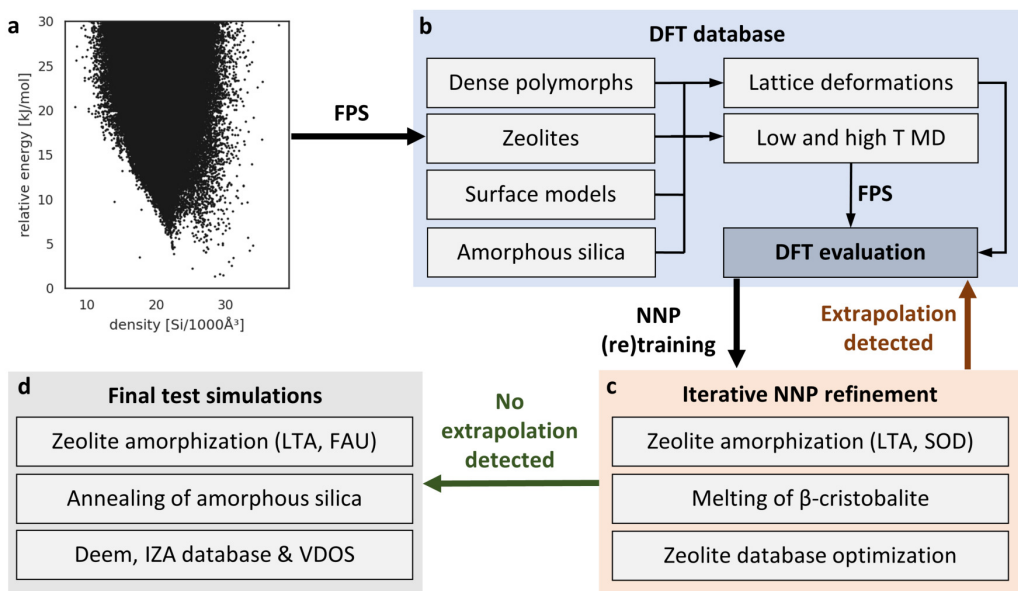


Figure 3: Illustration, for siliceous zeolites, of a strategy to create a library of DFT energetic and force data in a way to optimally cover the important part of configurational space including both close to equilibrium and activated structures close to transition state of bond-breaking events. One starts by (a) generating structurally diverse subset of zeolites selected by farthest point sampling (FPS) (b), followed by DFT single-point calculations on an FPS sparsified set of configurations. The MLP trained on the initial database are (c) iteratively refined by active learning. (d) The final MLP is then used for the production calculations. Figure taken from [Erlebach et al., 2022].

Pt clusters in the siliceous zeolite associated with intermittent breaking of the windows connecting the zeolite cages obtained from multi-nanosecond long molecular dynamics (MD) simulations, iii) effects of minor topology variations on germanium distribution for particular Si/Ge ratios in germanosilicate zeolites with profound effects on their delamination propensity obtained from comprehensive Monte Carlo simulations, and iv) effect of heteroatom concentration on proton solvation, water diffusion and stability of acidic zeolites [Saha et al., 2022]. Hence, our reactive zeolite NNPs seem to facilitate accurate simulations of even the largest zeolitic frameworks under realistic conditions (including temperature, pressure, heteroatom concentrations, or solvent effects), making them a new standard in the field. We expect that

future NNP simulations with our potentials will enable not only high throughput calculations for the targeted catalyst design but also provide atomic-level insights into zeolite synthesis and stability. This assumption is supported by the impact that the dissemination of our NNP-related work recently had on the rate of establishing of not only new academic collaborations but in particular new partnerships with industry partners, including the partnerships with the world-leading (petro)-chemical corporations.

However, even the acceleration of 3-4 orders of magnitude compared to DFT is not sufficient to directly probe reactive events such as those taking place during the zeolite synthesis. One of the most common ways how to increase the chance of observing such events is to accelerate sampling using enhanced sampling methods that bias the simulations along a low-dimensional representation of the reactive event [Chipot and Pohorille, 2007], *i.e.*, along the collective variable (CV). But guessing a "good" collective variable *a priori* is often a formidable task that is, however, amenable to acceleration/automatization by machine-learning, specifically by applying a dimension-reduction method to extract the essential information from the system [Sidky et al., 2020]. Despite the sustained effort in the recent years multiple questions remain open such as: i) the way how to efficiently featurize the atomic structures retaining inherent invariances, a property essential for the efficiency of the whole dimensionality reduction method, or ii) the ability to deploy these data-driven methods for reactions with high barriers. Lastly, most of the development is oriented towards applications in biomolecular systems with their applicability in the material science context being unclear. Building on our experience with end-to-end trained representations and message-passing neural networks, we managed to develop new neural network potential collective variables (NNP-CVs) coupling learned SchNet-based representations of the structures with variational autoencoder [Sípka et al., 2023]. Our approach is unique because it is using robust, well-scalable, inherently invariant representation of the structure that are generated automatically (no need to hand-pick the important descriptors as in previous attempts) and that are expected to be "aware" of the underlying potential energy surface. We tested it for enhanced sampling of reactive events in large nanoporous systems, *i.e.*, in zeolites. This combination of our NNPs with our automatically determined NNP-CVs is expected to enable us to simulate (catalytic) materials under *in operando* conditions, *e.g.*, making it possible to realistically simulate synthesis of zeolites, a process of high importance

and equivalently high lack of atomistic insight. And all that done without much of an *a priori* bias.

The above described development of machine learning potentials (so far published here [Erlebach et al., 2022, Saha et al., 2022]) and associated machine learned collective variables (so far published here [Sípka et al., 2023]) was conceived, conceptualized and supervised by me as a primary investigator heading a small group of very skillful postdocs and doctoral students. These developments were funded by my junior group leader research project by Charles University ([Primus/20/SCI/004](#)).

3 Applications

“Fall in love with some activity, and do it! Nobody ever figures out what life is all about, and it doesn’t matter. Explore the world. Nearly everything is really interesting if you go into it deeply enough.”

— Richard Feynmann

3.1 Accurate description of adsorption in nanoporous materials: a static approach

3.1.1 Benchmarking adsorption on coordinatively unsaturated metal sites

One of the most important classes of MOFs are the MOFs which contain in their metal-containing building units some metals ions that are not coordinatively saturated by organic linkers (see Figure 4). These coordinatively unsaturated sites (CUS) are in as-synthesized samples typically occupied by a solvent molecule, which is removed by heating. Activated CUS, which are regularly distributed throughout the MOF sample, show enhanced adsorption energies and are catalytically active - thus they are considered as very promising materials for catalytic, sensing, gas separation and gas sorption applications [Kokcam-Demir et al., 2020].

However, from the computational perspective, CUS-containing MOFs turned out to be particularly challenging to describe accurately for multiple reasons [Odoh et al., 2015]: i) a large unit cell sizes basically precluding application of other than DFT level of theory, ii) a need to include long-range van der Waals (vdW) interactions missing in the standard semi-local DFT functionals (see also Chapter 2.1), iii) description of transition metal cations by standard semi-local DFT functionals is often just semi-quantitative, suffering, *e.g.*, from problems with incomplete cancellation of electron self-interaction, and iv) the presence of multiple metal ions in the metal (or metal oxide) node with unpaired electrons that can couple, and which therefore have to be described by multireference calculations, or at least the validity of the standard single-reference approach should be tested. These compounded challenges necessitated a general approach for correcting standard (and computationally convenient) semi-local DFT functionals, which we tried to provide within the

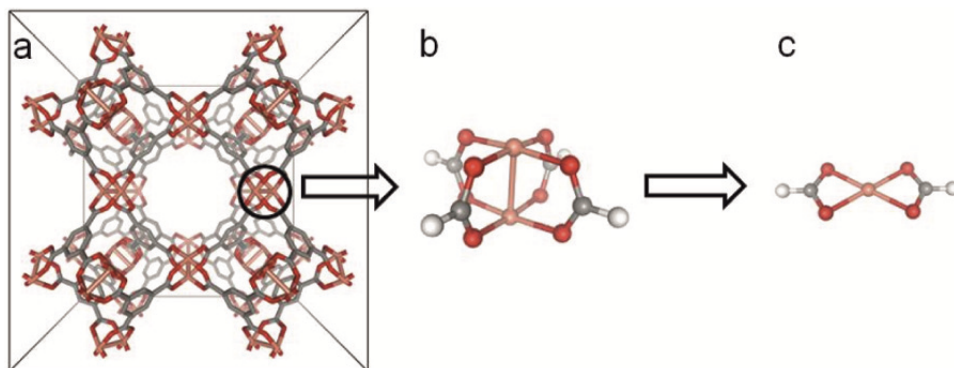


Figure 4: An example of the metal-organic framework with the coordinatively unsaturated metal sites - namely the CuBTC formed by benzene-1,3,5-tricarboxylate linkers and copper metal ions: (a) CuBTC unit cell, (b) paddlewheel (PDWL) unit, and (c) copper(II) formate model. Figure adapted from [Grajciar, 2013].

context of the DFT/CC scheme (see Chapter 2.1 for more details). Amongst the reported CUS-containing MOFs, the CuBTC MOF, $\text{Cu}_3(1,3,5\text{-benzenetricarboxylate})_2$, also known as HKUST-1, is considered in the literature as a reference CUS-containing MOF [Kokcam-Demir et al., 2020], which we also chose for our studies, in which we investigated how small gas molecules adsorb in CuBTC MOF. The metal-containing node of CuBTC MOF, also known as a paddlewheel (PDWL) unit, contains two Cu^{II} cations that are coordinated to benzenetricarboxylate ligands (see Figure 4) and contain two unpaired electrons that can couple either ferromagnetically or antiferromagnetically leading to single-reference triplet and multi-reference singlet ground state, respectively. We also note, that a half of our studies were carried out in collaboration with experimental investigators [Grajciar et al., 2011, Rubeš et al., 2012, 2013], highlighting our aim to overlap with experiment where possible and providing an extremely important "reality check" to our simulation work.

First part of our studies focused on benchmarking the accuracy of DFT functionals of different level of sophistication (LDA, GGA, meta-GGA and hybrid exchange-correlations DFT functionals) for description of interaction of water [Grajciar et al., 2010] and carbon monoxide [Rubeš et al., 2012] with the

CUS site in CuBTC at the zero-coverage limit, *i.e.*, with a single adsorbate bound to the CUS site. For the benchmarking, we employed a copper formate cluster models (see Figure 4), which were trackable by the "golden standard" method of the quantum chemistry, namely the CCSD(T) method. Luckily, employing multi-reference calculation (at the CASPT2-level), we found out that the interaction energies are mostly independent of the PWDL spin state, which allowed us to employ single-reference methods (describing PDWL in a triplet state) throughout our following works on the CuBTC described in this Chapter 3.1. The extensive benchmarking studies of various exchange-correlations DFT (XC-DFT) functionals showed that failures of XC-DFT functionals in description of small molecules interaction with CUS sites can be related not only to lacking ability to account for the vdW interactions but also to unrealistic charge and spin distribution that is expected to stem from the incomplete cancellation of the self-interaction, which was supported by an improved performance of XC-DFT functionals with increased admixture of the exact exchange (such as B3LYP). Such multifactorial nature of the standard XC-DFT functional failure represented a good case for our pragmatic but accurate DFT/CC approach (Chapter 2.1), which can not only effectively correct for these deficiencies but can be also straightforwardly extended beyond the limited cluster models to full periodic MOF models. Importantly, comparison with the available experimental data for zero-coverage limit confirmed that using DFT/CC approach for description of adsorption on CUS sites, we can truly achieve "chemical" accuracy of description (*i.e.*, errors within 1 kcal/mol).

In the follow-up study [Grajciar et al., 2015a], we extended our benchmarking to other transition metals (iron, in particular), larger set of small gas molecules (CH_4 , H_2 , N_2 , CO_2 , CO , NH_3 , H_2O) and other adsorption sites (mostly vdW-interaction dominated) in the CUS-containing CuBTC MOF. On average, we saw some improvement in adsorption description accuracy with the quality of the XC-DFT functional (going from GGA and meta-GGA to hybrid or even double-hybrid). However, none of the XC-DFT functionals could be recommended for general use, rather a metal- and MOF-topology-dependent benchmarking of XC-DFT functionals is suggested considering all the relevant adsorption sites present in the MOF material. Alternatively, the DFT/CC method could be leveraged for this purpose as we exemplified in our investigations of coverage effects described below (Chapter 3.1.2).

3.1.2 Coverage effects in MOF models

After a thorough benchmarking described above we focused on employing the DFT/CC scheme to more realistic and more challenging models, namely to studying absorption behavior in CuBTC MOF across a much broader coverage range (*i.e.*, across broader partial pressure range of the adsorbing gas). The challenge lies in a need to describe adsorption in a consistently accurate fashion across various types of adsorbate-adsorbent interactions (*e.g.*, interaction with CUS and vdW-dominated interactions with organic linkers) and in addition including lateral interactions, *i.e.*, adsorbate-adsorbate interactions. The coverage effects were studied for an important example of hydrocarbon sorption separation (propane vs. propene [Rubeš et al., 2013]) and for sorption of important greenhouse gases - methane [Chen et al., 2011] and carbon dioxide [Grajciar et al., 2011]. In CO₂ and propane/propene cases, we carried out an extensive enumeration of various representative structures of adsorbates at various loadings manually, however, for the methane case, we took a more thorough and comprehensive approach. We joined forces with the experts in classical force field simulation of adsorption from University in Edinburgh (group of prof. Tina Düren) collaborating within [MACADEMIA FP7 program](#), with whom we generated a simplified potential energy surface of methane at the DFT/CC level, that was used within grand-canonical Monte Carlo simulations to provide comprehensive information on adsorption behavior of methane in CuBTC covering range of temperatures and pressures (*e.g.*, calculating adsorption isotherms).

For all adsorbates considered, we achieved a quantitative agreement between our theoretical predictions and the experimental measurements (microcalorimetry, adsorption isotherms and neutron diffraction). Based on this agreement, we could reliably explain the adsorption mechanism for each of the adsorbates almost up to saturation, highlighting the remarkable role of lateral interactions and framework topology (*e.g.*, presence of pores of different sizes or CUS site orientation and separation allowing for cooperative adsorption), which defied simple explanations based on standard adsorption models (such as Langmuir, Freundlich or BET models). In addition, proposed mechanisms differed qualitatively from previous (incorrect) predictions from classical force fields, despite the disturbing fact that some of the classical force field macroscopic predictions, such as adsorption isotherms, were reasonable. Lastly, we confirmed our previous observations (Chapter 3.1.1) that the *ab initio* method employed for description of adsorption matters and that

the accuracy provided by DFT/CC scheme is essential to provide reliable description in line with the experiment.

Our work on adsorption in CUS-containing MOFs: i) provided a first benchmarking of performance of XC-DFT functionals for description of this challenging systems including the role of multi-reference effects and dispersion interactions, ii) generated an effective and pragmatic corrections (within the DFT/CC scheme) that provided, up to that point, unparalleled accuracy in description of adsorption of small gas molecules in these systems, and iii) stimulated a big push towards development of more complex classical force field that could effectively describe interaction with the CUS sites [Odoh et al., 2015]. The importance of our contribution to the field is also illustrated by the fact that our six articles on the topic accrued almost five hundred citations so far. However, this very competitive field of research started to focus lately rather on the ability of simulations to rapidly screen MOFs for particular properties, including adsorption-related ones, developing, *e.g.*, very pragmatic ways on how to describe adsorbate-CUS interaction using rather simple empirical analytical forms "embeddable" within the classical force fields [Chen et al., 2012, Vanduyfhuys et al., 2015, Campbell et al., 2018]. Such *ab-initio* derived force fields typically fit rather simple analytical formulas using the reference data from the scans of the potential energy surface, an approach reminiscent of the DFT/CC scheme, however, implemented on top or within the classical force field framework rather than being considered as a DFT correction. Despite a loss of accuracy compared to the DFT/CC scheme, such approaches definitely provide a marked improvement over previous classical force fields, are orders of magnitude faster and are more straightforwardly extendable/transferable to other topologies and metal centers (however, the transferability of the accuracy can be questionable). All this makes these approaches better suited for rapid high-throughput screening studies, in which the correct description of trends rather than the absolute accuracy is of a priority. On the other hand, for particularly challenging problems such as description of MOFs with metal nodes containing multiple spin-coupled transition metals [Odoh et al., 2015], I would personally see the wave-function-in-DFT embedding methods [Graham et al., 2020, Sharma and Sierka, 2022] as the most elegant and robust framework that would be up to the task. Lastly and rather surprisingly, this field has yet to be "discovered" by the machine learning potentials developers/users, with the only harbinger of this trend being now a rather old study by Eckhoff and

Behler [Eckhoff and Behler, 2019].

The work presented in these two subchapters (3.1.1 and 3.1.2) represented a new topic in my PhD group (within [MACADEMIA FP7 program](#)) that was driven mostly by me, which is illustrated by me being a first author on half of the referenced articles [Grajciar et al., 2010, 2011, 2015a]. Also the collaborative effort with prof. Düren group [Chen et al., 2011] was conceived, initiated and pushed forward by me. On the remaining two articles I assumed a supporting role [Rubeš et al., 2012, 2013]. Our efforts were capped by a chapter in the book on MOF modeling [Grajciar et al., 2015b].

3.1.3 Embedding environment effects in zeolitic models

Zeolites, similarly, to MOFs, also have handles that can modulate their adsorption properties - there are multitude of framework topologies available, and the concentration and the type of charge-compensating cations may be also varied. There are multiple potential and realized applications of zeolites, in which ability to tune the zeolitic embedding environment is useful, ranging from separation and energy storage application to catalysis or even electronic and optical applications [Li et al., 2017]. Since zeolites are already commercialized, cheap and reasonably thermochemically stable, one of the key application the zeolites are nowadays considered for is carbon separation and carbon capture, which is illustrated by the fact that zeolite 13X (sodium exchanged zeolite with FAU topology) serves as a benchmark material in the development of new materials for carbon capture and storage [Fu and Davis, 2022]. Hence, understanding the role of the embedding zeolitic environment (topology, cation type and concentration) in CO₂ adsorption in zeolites was both intriguing and very important.

In particular, we investigated how cation type (Li vs. Na) effects the low-coverage adsorption heats of CO₂ in FAU zeolite [Thang et al., 2014] and how different cation concentrations and distributions (assuming similar cation concentrations) give rise to very different adsorption heat profiles as a function of CO₂ coverage in Na-FER zeolite [Nachtigall et al., 2012]. Again, this work was carried out it in close collaboration with the experimental colleagues, whose data confirmed the ability of the DFT/CC scheme to deliver quantitative accuracy for the description of the weakly bound molecules in nanoporous materials and which allowed us to provide mechanistic interpretation of the adsorption including the emergence of the so-called dual cation

sites (see Figure 5), that were able to explain the observation of surprisingly high absorption heats of CO₂ in the experiment at low CO₂ coverages. These observations, along with the related earlier investigations within the group [Nachtigallová et al., 2006, Zukal et al., 2010], allowed us to propose a general model of absorption of CO₂ in zeolites [Grajciar et al., 2012] that consisted of: i) the effect from the bottom - interaction of CO₂ molecule with the primary extra-framework cation that is modulated by cation size, charge and its coordination to the framework, ii) the effect from the top - a secondary interaction of CO₂ with other nearby cations creating a bridging dual cationic sites or even multiple cation sites at extremely high cation concentrations, and iii) dispersion interactions mostly modulated by the framework density and the size of the channels or cages, in which the CO₂ molecule interacts with the cations. This absorption model not only allowed to interpret the available experimental data spanning the zeolite topologies, cation types and concentrations on an equal atomistic footing [Grajciar et al., 2012] but also allows to devise a guiding principle for designing zeolitic materials for particular separation, sequestration or purification application (*e.g.*, homogenous absorption heats for a broad range of CO₂ partial pressures vs. high zero-coverage heats).

I will also briefly mention a related work of mine [Grajciar, 2016], which investigated how extra-framework cation type influences the structure and the optical properties of a small lead sulfide quantum dots embedded in the zeolitic framework - such embedded quantum dots were reported to exhibit extremely high nonlinear optical properties [Kim and Yoon, 2012]. After extensive global structure optimizations with a newly developed tool (see Chapter 2.3), followed by the expensive relativistic calculations at the hybrid DFT level, I managed to reliably compare with the available experimental spectra establishing a structure-property relationship for these materials highlighting a pronounced role of the cation type in both the structure and optical properties of the QD.

Similarly, as in our studies of adsorption in CUS-containing MOFs, the DFT/CC scheme provided an unparalleled accuracy for description of CO₂ adsorption in zeolites. This was recognized in the field, with our reference data being used for fitting classical force fields [Fang et al., 2013] of improved quality and applying these force fields across zeolite topologies and cation concentrations [Fang et al., 2013, Yang et al., 2021]. This again followed a general trend in the field mentioned above (Chapter 3.1.1) that preferred the

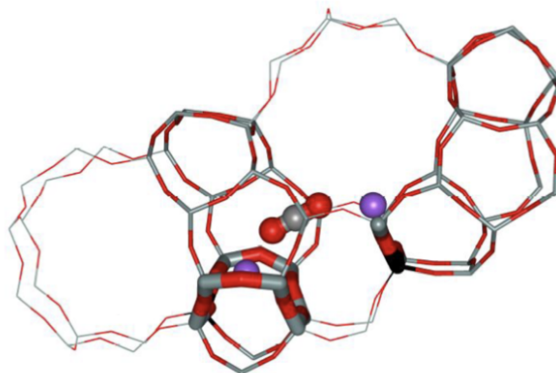


Figure 5: An example of a dual CO₂ adsorption complex in a sodium-exchanged MFI zeolite. Figure adapted from [Grajciar, 2013].

ability to rapidly screen adsorption properties of various zeolites, that can be fed into multi-scale screening protocols [Farmahini et al., 2021], at the expense of absolute accuracy. However, one of the outstanding issues for carbon capture is the CO₂ adsorption in humid conditions [Fu and Davis, 2022] as water has significant detrimental effects on adsorption performance both due to adsorption site blocking or due to hydrolytic degradation of zeolite. Investigation of such zeolite-water interactions is currently one of my key research interests (see Chapters 3.2 and 3.3 for details).

In the work on the embedding environment effects on the CO₂ adsorption in zeolites I mostly followed up on body of earlier related work in the group [Nachtigallová et al., 2006, Zukal et al., 2010] and contributed to it both from the supporting [Nachtigall et al., 2012, Thang et al., 2014] and the main author role [Grajciar et al., 2012]. The study on the structure and optical properties of PbS QDs embedded in zeolite cavities of different cation composition was my own single-author work [Grajciar, 2016].

3.2 Towards reactivity and dynamical models

An increased sophistication of experimental investigations, able to probe nanoporous systems under realistic *operando* conditions, have been highlighting for some time now the very dynamical nature of the nanoporous systems undergoing significant changes under operation conditions, *e.g.*, with (cat-

alytically) active species being often created only transiently. In addition, complex effects of solvent and reactive species concentration as well as temperature or pressure on material stability and reaction rates have been observed [Kalz et al., 2017, Reuter et al., 2017]. These reports suggest, that for many systems, a move beyond simple descriptive models, based on probing the potential energy surface at 0 K and low reactant or solvent concentration is a must [Grajciar et al., 2018] (see Figure 6) since under realistic conditions: i) the structure of active site may change, ii) new energetically feasible reaction pathways may become available, or iii) the relative preference of competing reaction pathways may be significantly altered. However, employing a computational model that could account for realistic operation conditions, *e.g.*, global structure optimization or (biased) *ab initio* molecular dynamics (AIMD), significantly increases computational demands, which significantly limits the breadth of the systems and problems that could be investigated. Nevertheless, even employing realistic models within such limited scope may be sufficient to provide significant new insights and help novel concepts to materialize, as we exemplified in our investigations into zeolite interaction with (liquid) water [Heard et al., 2019b].

The major challenge in the zeolite science is the durability of zeolite-based catalysts in technological processes. Numerous chemical processes, including biomass conversion and crude oil processing, involve aqueous media at elevated temperatures and pressures, rendering zeolite hydrolysis the critical factor in catalyst durability [Heard et al., 2020]. In addition, the (partial) zeolite hydrolysis could be also reigned in and adopted for synthesis of novel zeolite structures (using so-called ADOR method [Roth et al., 2013]) or for introduction of mesoporosity, which diminishes the diffusion limitations and thus improves the catalytic performance of zeolites [Verboekend and Pérez-Ramírez, 2011]. Despite a large body of available experimental work on the topic [Heard et al., 2020], a deeper mechanistic understanding of the hydrothermal zeolite (in)stability and the role of various factors (*e.g.*, concentration of defects, framework composition and topology, partial water pressure, temperature, *etc.*) is still mostly missing, impairing the ability to rationally design zeolitic materials with desired level of water tolerance. Our work tried to contribute to providing answers to these unresolved questions, employing realistic zeolitic models in aqueous media treated typically dynamically (using (biased) AIMD), allowing for more open-ended exploration of reactive events taking place at the zeolite/water interface.

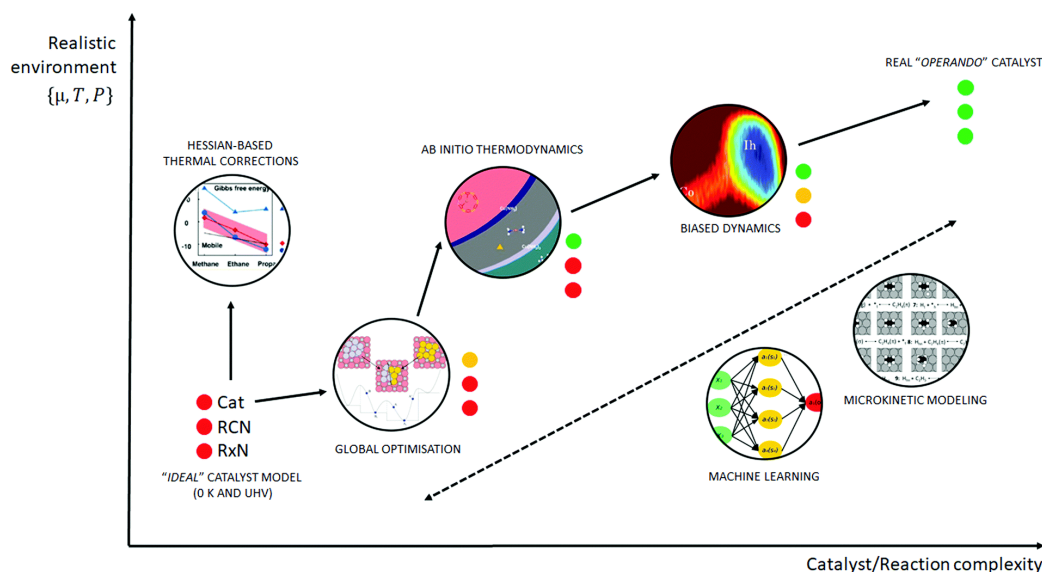


Figure 6: Schematic of the various computational methods applied to heterogeneous catalysis, which lie between an idealised UHV model and a realistic, *operando* model. Figure taken from [Grajciar et al., 2018].

Our investigations fall roughly into two broad classes: i) quantifying the effect of temperature and water concentration in zeolitic pores on hydrolytic mechanisms [Heard et al., 2019b, Jin et al., 2021b,a], and ii) determining the character of active species under realistic conditions including the presence of liquid water [Heard et al., 2019a, Liu et al., 2020, Jin et al., 2022, Jin, 2022]. In the first class of studies, we managed either to discover novel energetically favorable hydrolytic mechanisms (see Figure 7) or at least to observe significant modulation of the "known" mechanisms under increased water concentration. A common thread in these studies was the "emergent" collective/cooperative nature of those mechanisms that is difficult to predict *a priori*, and which justifies the use of open-ended dynamical description of the zeolite/water interface. The collectiveness was exhibited not only by the water molecules (*e.g.*, shuffling the proton via Grotthuss mechanism towards to-be-broken Al-O, Si-O or Ge-O bonds) but also, *e.g.*, by the germanium heteroatoms in the zeolitic framework, which upon mild clustering, opened hitherto unreported low-energy hydrolytic pathways [Jin et al., 2021b]. Alternatively, the collectivity may also manifest in a more subtle ways such as modification of activation barriers and reaction energies for the "known"

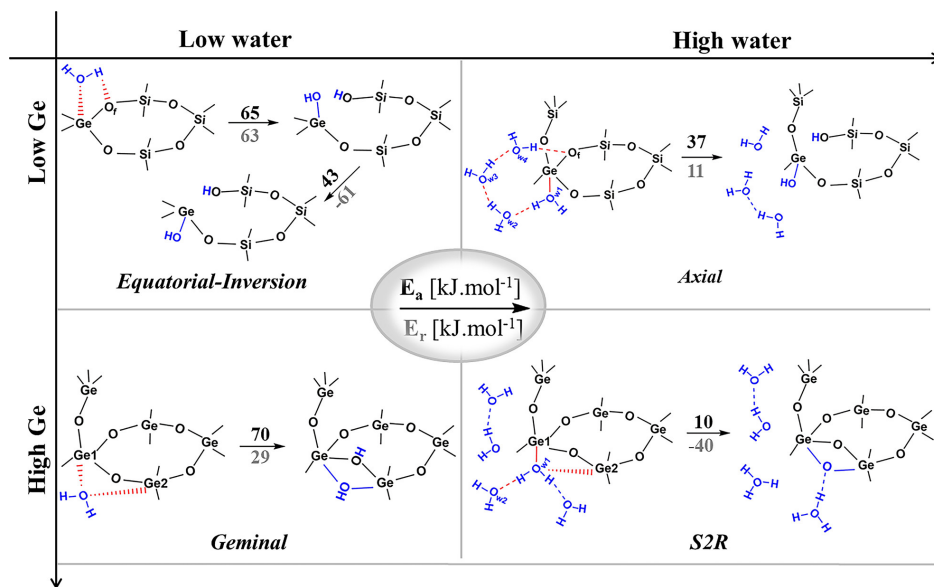


Figure 7: Scheme summarizing how by varying water loading and germanium content one may open up very distinct mechanistic pathways for zeolite decomposition. Figure taken from [Jin et al., 2021b].

competing reactive pathways as we witnessed in our study of zeolite hydrolysis in alkaline aqueous solution [Jin et al., 2021a], in which increased water concentration in nanopores starts to favor hydrolytic mechanism in which sodium hydroxide acts as a catalyst rather than as a reactant. Importantly, such observations are not only theoretical curiosities but have been supported by water adsorption measurements [Jin et al., 2021b] and NMR spectroscopy [Heard et al., 2019b] of our experimental collaborators and have important practical consequences on how one thinks about zeolites and their interaction with water. In particular, our work on hydrolytic lability of aluminosilicate zeolites under ambient conditions [Heard et al., 2019b] changed the tradition view of zeolites as stable and rigid frameworks towards a much more dynamical picture of fast (Si-O and Al-O) bond-forming and bond-breaking system creating transient defects that can affect many of the defining properties of zeolitic frameworks such as their catalytic activity or molecular sieving. Partial zeolite hydrolysis leads to creation of defective sites, many of which may become either the initial sites for further decomposition or active sites for a sought-for reactive transformation [Heard et al., 2020]. Hence, in the second

class of studies we tried to determine the character of active species, *i.e.*, of the Brønsted acid site pair [Heard et al., 2019a] and of the elusive framework-associated octahedral aluminum species [Jin et al., 2022, Jin, 2022], in the presence of liquid water. We also complemented this class with a study of the effect of various active site species on the zeolite reactivity using a nucleophilic double-bond-attack test reaction [Liu et al., 2020], limiting ourselves to a static 0 K description. In the first character-of-active-species investigation [Jin et al., 2022, Jin, 2022] we were, in collaboration with our experimental colleagues, able to determine the plausible structure of the long sought-for but elusive framework-associated octahedral aluminum species, Al(Oh), and explained why it can be formed only under sufficient water solvation reverting back to the tetrahedral coordination upon drying. The structure of Al(Oh) species also explained why it can be formed only in zeolites with sufficiently high concentration of aluminum. The essential role of water in determination of the active species was also confirmed in our investigation [Heard et al., 2019a] on how increased water concentration stabilizes canonical aluminum distributions (so-called Loewenstein distributions) over the non-canonical Al-O-Al non-Loewenstein pairs of aluminums. This observation has an important consequence as a mechanistical justification for the experimentally observed validity of the so-called Loewenstein rule that "forbids" formation of the Al-O-Al aluminum pairs, *i.e.*, the formation of the non-Loewenstein pairs. Lastly, we tried to screen multiple conceivable acidic active sites in a zeolite for their activity in a particular test reaction (tetrahydropyranylation), comparing it with the experimental catalytic tests [Liu et al., 2020]. To reliably explain the experimental observations for the most of the samples, we either had to consider the existence of defective active sites or the existence of significant diffusion barriers experienced by the bulky products. These results highlight the complex nature of the processes taking place under *operando* conditions, in which a simplified model of the nanoporous material with an idealized active site, on which the reaction proceeds at 0 K under ultra-high vacuum, is insufficient not only for quantitative but also for qualitative understanding of reactivity of and in the nanoporous materials.

Our work focused on simulation of zeolitic materials in interaction with water at *operando* conditions [Heard et al., 2020]. It highlighted the potential of advanced simulation techniques (such as biased AIMD) to reveal unexpected mechanistic routes at realistic conditions and dynamical character of zeolitic

surface with defects formed often transiently and in low concentrations hardly accessible to experimental observations. I expect that these observations of dynamical nature of zeolites and the complex character of the zeolite-water interactions will be of great importance to our understanding of zeolite behavior not only in established fields such as in petroleum refinement but also in evolving fields such as bio-refinery catalysis or in synthetic manipulations enabling access to new zeolitic materials. However, due to significant computational costs of these advanced simulation techniques, our work had to be limited only to a very small subset of conditions and systems. A way how to move beyond this "looking through a peephole" approach, leveraging the power of machine learning, was discussed from a methodological point of view in Chapter 2.4 with some relevant applications mentioned in the following Chapter 3.3.

The push towards adopting advanced simulation techniques able to describe zeolites under realistic conditions in our group has been conceived and initiated by me. This is also reflected in me being the first author on the review article summarizing use of these techniques and highlighting their best practices in heterogeneous catalysis [Grajciar et al., 2018]. I am a corresponding author on the two articles referenced in this section [Jin et al., 2021b, Liu et al., 2020] and I made sizable contributions (*e.g.*, discovery and quantification of the new transformation routes, guidance of the PhD students, co-drafting manuscript, *etc.*) to other three standard articles [Heard et al., 2019b, Jin et al., 2021a, 2022] and one review article on zeolite-water interactions [Heard et al., 2020]. In the remaining article [Heard et al., 2019a] I assumed mostly a consulting role.

3.3 High-throughput calculations, material design and beyond

The rational design of novel materials, including novel nanoporous materials, is a holy grail of the material science. Despite a significant progress in recent years, thanks to, *e.g.*, the development of automatized experimentation or ML-accelerated computational screenings [Schmidt et al., 2019, Burger et al., 2020], this goal is still mostly elusive.

Zeolitic nanoporous materials are unfortunately no exception to this, despite being particularly rewarding targets for targeted property (reactivity, selec-

tivity, adsorption capacity, *etc.*) design with multiple parameters that could be tuned such as framework topology and dimensionality, heteroatom type and concentration or character and concentration of the charge-compensating cations. The topology parameter appears to be particularly underleveraged as only about 250 zeolite topologies have been synthesized so far despite tens to hundreds of thousands hypothetical zeolitic topologies being proposed as thermodynamically accessible [Deem et al., 2009]. Moreover, most of the newly obtained zeolitic topologies are synthesized in a trial-and-error fashion. There are few exceptions to this rule such as the ADOR protocol [Roth et al., 2013] (including my very minor contribution), which allows for the generation of new topologies with targeted pore sizes by chemo-selective partial hydrolysis of germanosilicate zeolites followed by targeted recondensation of non-hydrolyzed parts of the parent zeolitic structure.

A way how to accelerate the material design is to transfer this endeavor from lab to the computer, *i.e.*, to leverage the computer-aided *in silico* design. This is typically realized by screening large number of candidate structures in a high-throughput fashion for a desired property/behavior. The candidate structures are mostly taken either from the existing databases or generated using computationally-cheap analytical force fields, but more comprehensive first principles-based investigations have also emerged with the increased performance of and accessibility to high performance computing resources [Li et al., 2018]. With the advent of machine learning, the generation of the structural databases can be accelerated, and high-throughput screening and analysis can be made more robust, efficient, and statistically significant.

In the zeolite science, the ML application has been centered so far around agglomeration and classification/interpretation of the existing data [Muraoka et al., 2019, Schwalbe-Koda et al., 2021] (*e.g.*, the zeolite synthesis parameters) or around identification of the most relevant structural descriptors for various properties, *e.g.*, mechanical [Evans and Coudert, 2017] or catalytic [Zhu et al., 2022]. However, the ML has not been yet utilized in the zeolite field (or even in other nanoporous material field such as metal-organic frameworks besides one exception [Eckhoff and Behler, 2019]) to accelerate atomistic simulations, *i.e.*, to generate new structural and energetic data. Hence, most of the existing screening studies in zeolites still rely heavily on specialized analytical force fields. We have benchmarked some of those analytical force fields in number of occasions, such as in the study of zeolite 2D-layer interaction and (re)ordering in the context of the ADOR protocol [Grajciar et al., 2013] or when developing our neural network potentials for

(alumino)silicate zeolites [Erlebach et al., 2022, Saha et al., 2022], and the general outcome is that they are unreliable, providing only very qualitative description at best, but mostly fail to do even that. And as I have illustrated in previous chapters (chapters 3.2 and 3.1), it seems clear that achieving reasonable level of accuracy accompanied by adoption of realistic models is essential for reliable description of nanoporous materials. Such reliable description, if significantly accelerated, could then allow for reliable predictions in the context of the computer-aided material design.

We followed two different routes towards achieving the goal of realizing the comprehensive high-throughput-like but reliable simulations of zeolitic materials. Initially, we carried out extensive first-principles based studies evaluating the Lewis and Brønsted acidity of selected zeolites as well as their two-dimensional analogues [Ho et al., 2018, Thang et al., 2019], with the aim to quantify how much and if the 3D-to-2D transformation affects the intrinsic zeolite acidity strength. In both studies we considered two to three zeolite topologies and for each topology we considered all the symmetrically inequivalent acid sites. This amounted to hundreds of acid sites that were screened with DFT calculations. Multiple effects of the zeolite transformation to the layered form, such as change in cation coordination and cation distribution around particular aluminum site or emergence of new acid sites on the external surface, were encountered and their aggregate result on macroscopically observable acidity descriptors (*e.g.*, OH stretching vibration and IR spectrum of the adsorbed CO probe) was quantified and compared favorably with experimental observations. Another important conclusion of these studies was that all sites (a few hundreds of them) needed to be considered in order to get not only statistically relevant but also (and most importantly) experimentally observed results, which are that the Brønsted acidity of three- and two-dimensional zeolites is basically the same. Focusing only on a subset of seemingly representative sites, as is often done in first-principles calculations to save the costs, could lead to qualitatively different conclusions, as many of such sites do exhibit pronounced changes upon 3D-to-2D transformation - these effects, however, average out if all aluminum sites are considered. This stresses the need for comprehensive screening. However, such screening even for only slightly more complex systems than considered in these studies (such as water-zeolite systems discussed in Chapter 3.2) is prohibitively costly. Therefore, we adopted a novel machine learning potential (MLP) methodology that has been shown to be able to accelerate the calculations by few

orders of magnitude while being able to retain first principles accuracy (see Chapter 2.4 for more details). However, the majority of existing MLP applications focus on acceleration of simulations: i) close to the specific thermodynamic state point (*e.g.*, equilibrium dynamics at a well-defined non-reactive solid-liquid interface), or ii) for a particular reaction in question (*e.g.*, quantifying non-classical quantum effects), or iii) for simple, mostly elemental, systems (*e.g.*, such as evaluating elemental phase diagram). In contrast, our aim was to generate a more general (so-called *global*) MLP that would cover larger fractions of the configuration space and chemical space at once, *i.e.*, covering not only close-to-equilibrium structures but also higher-energy transition states (including bond-breaking and bond-forming events) and describing multielemental systems in multitude of reactive scenarios. Up to now we have managed to generate MLPs for multiple systems including (see Figure 8) : i) siliceous zeolites (including silica glass) [Erlebach et al., 2022], ii) germanosilicate zeolites, iii) siliceous zeolites with hydroxylated defects (*e.g.*, OH-group-terminated silica surface), iv) platinum and silver clusters embedded in (defective) siliceous zeolites, v) aluminosilicate zeolites in acidic form (with hydrogen cation as a charge-compensating cation) interacting with neutral/acidic/basic aqueous solution [Saha et al., 2022], and vi) aluminosilicate zeolite in sodium form interacting with neutral/acidic/basic aqueous solution. With the exception of the germanosilicate MLPs, which are non-reactive by construction, all potentials were constructed and successfully tested against first principles calculations (see, *e.g.*, Figure 9) for treatment of reactive events, such as zeolite melting, reactive platinum clusters agglomeration with intermittent breaking of the framework or proton solvation and partial zeolite hydrolysis. Moreover, the MLPs for more complex systems such as the MLP for aluminosilicate zeolites in sodium form in interaction with water is generated in a way that it covers also (defective) siliceous systems or aluminosilicate systems in acidid form, *i.e.*, our MLPs truly tend towards *global* MLPs.

Our first testing grounds was the development of the MLP for general potential surface (PES) modelling of silica [Erlebach et al., 2022] covering a broad range of silica densities from siliceous zeolites with large pores through denser zeolitic frameworks and silica glass towards very dense silica polymorphs such as quartz or even octahedrally coordinated stishovite. We thoroughly tested the accuracy of the potential in multiple out-of-domain generalization tests, with the DFT calculations serving as a reference, such as cristobalite and zeolite melting, Stone-Wales defect formation [Klemm et al., 2020], evaluation

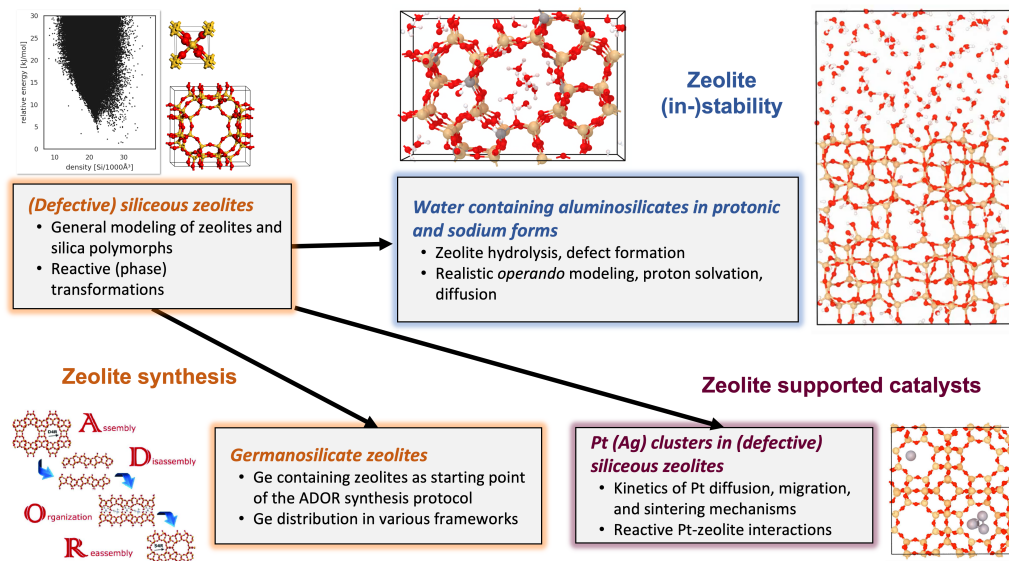


Figure 8: Overview of the application fields for which we have generated our neural network potentials with the examples of the particular application goals.

of the vibrational density of states for glass and cristobalite or evaluation of the single-point energies for close-to-equilibrium structures randomly chosen from the Deem zeolite database [Deem et al., 2009]. In all instances, our MLPs retained the DFT-level of accuracy. In contrast, other popular and for silica specialized reactive (ReaxFF) or non-reactive (SLC) analytical force fields exhibited at least on order of magnitude larger errors with respect to the DFT reference than our MLP (see Figure 9). Interestingly, even in some cases of extrapolation or sparse interpolation (*i.e.*, in parts of the PES sparsely covered by the training data), our MLP showed qualitative agreement with the reference calculations demonstrating unexpected robustness of the underlying NNP architecture, which could help in explaining the experienced ability to generate more *global* MLP and which also allows for straightforward improvement and extension by active learning. Having the reliable silica MLP trained, we applied it for high-throughput reoptimization of the IZA (International Zeolite Association) and Deem zeolite databases which contain roughly 330 thousand real and hypothetical zeolite structures. These databases are used as an input to evaluate the correlation between

the zeolite density and energy, a central quantity used by experimentalists to estimate the synthetic feasibility of new candidate zeolitic structures [Henson et al., 1994]. The database reoptimization at the MLP level revealed more than twenty thousand additional hypothetical zeolites within the synthetic feasibility range as compared to the results obtained with the state-of-the-art analytical force field. In addition, the revised Deem database provides essential input for the future screening studies on structure-property correlations in zeolites inching us closer towards the ultimate goal of computer-aided design and discovery of zeolites.

We made a significant step towards more complex and truly realistic systems with the development of the NNP for all aluminosilicate zeolites in the protonic form in reactive interaction with the water solution [Saha et al., 2022]. We showcased its potential by determining the effect of aluminum content, water loading and temperature on water dynamics and proton solvation in one of the industrially most important acidic zeolites, the faujasite (FAU). This computational screening across multiple variables showed, *e.g.*, that diffusivity of confined water can change by as much as an order of magnitude as a function of the aluminum content or that FAU samples with very high concentration of Brønsted acid sites appear to be on average less acidic than the samples with lower Brønsted acid site concentration (*i.e.*, they solvate less protons). In addition, we managed to observe unexpected "new" chemistry in the MLP simulations such as formation of the peculiar "pinned" hydroxonium species residing over 6-membered rings abundant with aluminum or witness the formation of the 3-membered ring defects containing non-Loewenstein Al-O-Al aluminum pairs in samples with very high aluminum content. These observations seem to be supported by the preliminary DFT tests and could be related to some of the experimental observations from neutron diffraction and stability tests of acidic forms of FAU zeolite with very high aluminum content [Lee et al., 2013, Czjzek et al., 1992]. Hence, such results indicate that we may be able to generate more *global* MLPs that go beyond a straightforward acceleration of simulations nearby known thermodynamic state points but could be used also for exploration of new effects and transformations such as those needed to explore zeolite synthesis pathways.

Development of our new reasonably *global* MLPs supplemented with our ML-based collective variables (see Chapter 2.4 for more details) indeed appears to open up a possibility to simulate (nanoporous) materials under realistic condition retaining the first principles accuracy and also being able to carry

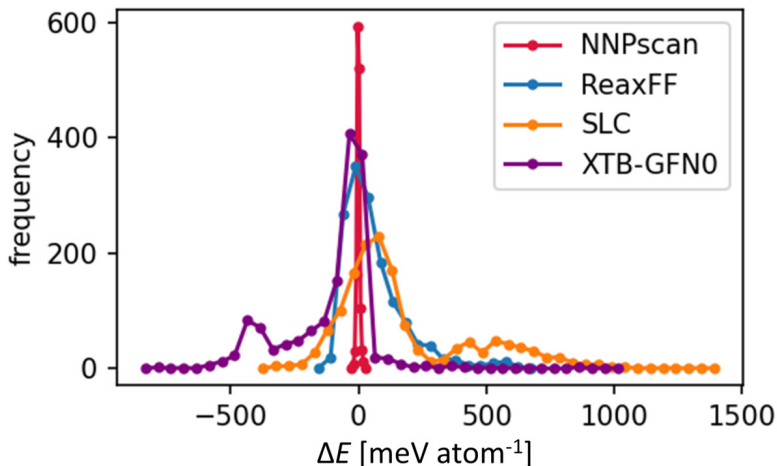


Figure 9: Example of energy errors with respect to DFT reference for the out-of-domain generalization tests for our MLPs for siliceous systems with comparison to the other specialized analytical force fields and a semi-empirical method XTB-GFN0. Figure taken from [Erlebach et al., 2022].

out large-scale screenings of the conceivable material and reaction system targets, with a bonus of providing the atomic-level resolution and insight. The MLPs accelerate the simulations while ML-based collective variables accelerate sampling of the rare events (*e.g.*, chemical reactions) along relevant degrees of freedom, determination of which is made basically automatically, solving to some degree the well-known "chicken-and-egg" problem [Rohrdanz et al., 2013] of determining a good collective variables without *an priori* knowledge of a transformation/reaction path. Hence, the targeted zeolite design might be actually within our reach now. In addition, our approach of generating more *global* MLPs is expected to be extensible beyond the zeolite field, opening up the option of the *on-demand* training of the NNPs for a particular application - such assumption is demonstrated by our newly established partnerships and *on-demand* NNP development projects (for the time being within the zeolite field) with the world-leading (petro)chemical companies. Next, multiple new advanced neural network architectures have been proposed recently such as equivariant NNs [Schütt et al., 2021, Batzner et al., 2022], including also easily parallelizable implementations able to treat hundreds of thousand atoms [Musaelian et al., 2022], or physically-motivated NNs including long-range interactions [Unke et al., 2021a, Frank et al., 2022].

The adoption of these architectures is expected to further improve the performance of our NNPs in particular for tensorial properties (*e.g.*, such as forces, polarizabilities or NMR tensors) and also it should decrease the amount of the reference data needed to cover the desired parts of the chemical and configurational space thanks to the improved data-efficiency of these new architectures. Lastly, improved data-efficiency should also allow for more efficient application of the Δ -learning strategies [Ramakrishnan et al., 2015], *i.e.*, composite strategies able to achieve the accuracy of the computationally much more expensive first principles methods such as hybrid DFT or post-HF methods, if that turns out to be necessary for a particular application. These strategies work by constructing ML-based correction surfaces to computationally less expensive baseline methods, such as our NNPs, with the corrections being evaluated using a much smaller and/or sparser grid of the expensive higher-level data points, since one assumes that: i) the correction surface is smoother than the baseline model, or that ii) the improvement in description is desired only for a specific subset of the configuration space, *e.g.*, a specific chemical reaction. Alternatively, the free energy perturbation approach could be applied in the latter case as well.

The above described application of the machine learning potentials and ML-driven collective variables (so far published here [Erlebach et al., 2022, Saha et al., 2022, Sípka et al., 2023]) was conceived, conceptualized and supervised by me as a primary investigator heading a small group of very skillful postdocs and doctoral students¹ and funded by my junior research group project by Charles University ([Primus/20/SCI/004](#)). In addition, the high-throughput-like DFT based studies on Lewis and Brønsted acidity of selected three- and two-dimensional zeolites [Ho et al., 2018, Thang et al., 2019] were also driven and supervised by me as indicated by me being a corresponding author on both publications. These high-throughput-like DFT based studies were generated within the standard Czech Science Foundation project ([No. 17-01440S](#)), on which I was a co-investigator at the Faculty of Science, Charles University.

¹The ML-based development and publications were also mentioned in Chapter 2.4

4 Conclusions and outlook

This habilitation thesis recapitulates my more than a decade long effort to understand various processes taking place in nanoporous materials, in particular zeolites and metal-organic frameworks, at the mechanistic level using simulations with the atomistic resolution and *first principles* accuracy. This effort involved multiple diverse methodological advances (Chapter 2) that were typically leveraged for diverse problems of interest (Chapter 3), albeit all within the field of nanoporous materials. The diversity in the methods developed and the applications pursued stems, in my opinion, from my inherent novelty-seeking character, due to which I am eager to try out new ways of solving scientific questions or look at existing problems from the unorthodox viewpoints - all this provides me with a broader overview and a broader toolkit of methods, but it clearly comes at some efficiency costs compared to somebody specialized in a very narrow field. The common thread in my scientific endeavors was the striving for a reasonable mix of generality, elegance and practicality in the methods developed (Chapter 2) and for the deployment of those methods for applications, in which our simulations could be readily experimentally verifiable/falsifiable. The latter "imperative" represents my effort to try to simulate something "real", *i.e.*, adopting a reasonably realistic model for a problem in question. This "imperative" to adopt realistic models, with a tacit assumption of using methods able to deliver at least qualitatively correct description, has guided my scientific path starting with simple models for a particular nanoporous material (see Chapter 3.1 describing our work on (physi)sorption of small molecules in MOFs and zeolites), continuing with more complex models but still being limited to a particular system (see Chapter 3.2 describing our work on various aspects of water-zeolite interaction) and recently aiming for a more comprehensive description of nanoporous materials under realistic *in operando* conditions covering a much broader structure/topology space (see Chapter 3.3 describing our attempts towards accelerating simulations using machine learning (ML)).

Some of the main results of from the methodological side (Chapter 3) can be summarized as follows:

- Development of molecular and periodic DFT implementations able to efficiently (both CPU-time- and memory-wise) treat large nanoporous or even mesoporous materials.

- Development of global structure optimization tool for determination of structure of molecules/clusters embedded in nanoporous materials.
- Development of ML-based tools for accelerated simulations of nanoporous materials, namely the development of machine learning potentials (MLPs) to speed up the sampling of the potential energy surface and the development of approaches to accelerate and automatize sampling of the rare events such as the generation of the collective variables based on variational autoencoders.

Many of these methodological advances were leveraged later for numerous application studies, in many of which the use of advanced simulation techniques accompanied with an adoption of reasonably complex model system was rewarded with novel unexpected insights into: i) adsorption in zeolites and metal-organic frameworks, ii) hydrolytic stability of zeolites and iii) speciation of the catalytically active site in zeolites under *operando* conditions. In particular, the herein presented work:

- Unraveled the dynamical nature of zeolites, the emergence of transient active sites under *operando* conditions and the complex collective character of the zeolite-water interactions, which are difficult to predict *a priori* using "chemical" intuition. This justified the resolve to use the advanced open-ended dynamical methods for description of these interfaces.
- Generated state-of-the-art *reactive* MLPs for multiple systems spanning increasingly complex configuration and chemical space starting with silicious zeolites in vacuum, continuing with platinum clusters embedded in silicious zeolites and going (for the time being) all the way to aluminosilicate zeolites in both acid or sodium form interacting with neutral/acidic/basic aqueous solution. Using these MLPs, we were, *e.g.*, able to reveal more than twenty thousand additional zeolite framework candidates within the synthetic feasibility range.

Hence, the focus of my application work shifted from specific simple systems that we could treat with quantitative accuracy (close to golden-standard of CCSD(T) post-HF method) to a comprehensive set of complex systems that we could treat at least qualitatively correct with MLPs (with a semi-local DFT level accuracy). However, I expect, supported by some of our own recent

developments and related works in the field [Ang et al., 2021, Chmiela et al., 2018], that in the very near future, one might not need to compromise much between the complexity of the model and the complexity of the methods, *i.e.*, for a particular (complex) system of interest one could obtain results of almost arbitrary accuracy or at least able to reach the chemical accuracy (1 kcal/mol). The solution, in my opinion, lies in composite strategies (*e.g.*, similar to Δ -learning or free energy perturbation mentioned in Chapter 3.3), in which: 1) a *global* MLP, such as our own, would be used for pre-screening of broad configuration and chemical spaces, searching for a limited subset of relevant reactions, and/or close-to-optimal compositions and conditions, and 2) a local "correction"-MLP for these relevant reactions and conditions would be trained using a very limited set of expensive higher-level data points, which could be generated, *e.g.*, using various embedding strategies such as mechanical embedding [Sauer, 2019] or rather density-in-wavefunction [Sharma and Sierka, 2022] embedding. In addition, it is not unreasonable to expect breakthroughs in the development of ML-accelerated quantum chemical methods such as ML-based exchange-correlation DFT functionals [Pederson et al., 2022] in the near future. Therefore, being a bit techno-optimistic, I think that machine learning has truly a transformative potential for our ability to realistically simulate and eventually design (nanoporous) materials and that in the rather near future most of the consequential (nanoporous) material studies will be ML-accelerated in one way or the other.

References

- TURBOMOLE V7.6 2021, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from <http://www.turbomole.com>.
- Introduction to metal–organic frameworks. *Chemical Reviews*, 112(2): 673–674, 2012. doi: 10.1021/cr300014x. URL <https://doi.org/10.1021/cr300014x>. PMID: 22280456.
- S. J. Ang, W. Wang, D. Schwalbe-Koda, S. Axelrod, and R. Gómez-Bombarelli. Active learning accelerates ab initio molecular dynamics on reactive energy surfaces. *Chem*, 7(3):738–751, 2021. ISSN 2451-9294. doi: <https://doi.org/10.1016/j.chempr.2020.12.009>. URL <https://www.sciencedirect.com/science/article/pii/S2451929420306410>.
- M. Arrigoni and G. K. H. Madsen. Evolutionary computing and machine learning for discovering of low-energy defect configurations. *npj Computational Materials*, 7(1):71, May 2021. ISSN 2057-3960. doi: 10.1038/s41524-021-00537-1. URL <https://doi.org/10.1038/s41524-021-00537-1>.
- N. Artrith. Machine learning for the modeling of interfaces in energy storage and conversion materials. *Journal of Physics: Energy*, 1(3):032002, jul 2019. doi: 10.1088/2515-7655/ab2060. URL <https://doi.org/10.1088/2515-7655/ab2060>.
- S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. URL <https://doi.org/10.1038/s41467-022-29939-5>.
- M. Becker and M. Sierka. Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole method: Stress tensor. *Journal of Computational Chemistry*, 40(29):2563–2570, 2019. doi: <https://doi.org/10.1002/jcc.26033>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.26033>.
- M. K. Bisbo and B. Hammer. Efficient global structure optimization with a machine-learned surrogate model. *Physical Review Letters*, 124:086102,

- Feb 2020. doi: 10.1103/PhysRevLett.124.086102. URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.086102>.
- O. Bludský, M. Rubeš, P. Soldán, and P. Nachtigall. Investigation of the benzene-dimer potential energy surface: Dft/ccsd(t) correction scheme. *The Journal of Chemical Physics*, 128(11):114102, 2008. doi: 10.1063/1.2890968. URL <https://doi.org/10.1063/1.2890968>.
- B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, and A. I. Cooper. A mobile robotic chemist. *Nature*, 583(7815):237–241, July 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2442-2. URL <https://doi.org/10.1038/s41586-020-2442-2>.
- E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Banwarth, and S. Grimme. A generally applicable atomic-charge dependent london dispersion correction. *The Journal of Chemical Physics*, 150(15):154122, 2019. doi: 10.1063/1.5090222. URL <https://doi.org/10.1063/1.5090222>.
- C. Campbell, J. R. B. Gomes, M. Fischer, and M. Jorge. New model for predicting adsorption of polar molecules in metal–organic frameworks with unsaturated metal sites. *The Journal of Physical Chemistry Letters*, 9(12):3544–3553, 2018. doi: 10.1021/acs.jpcclett.8b00967. URL <https://doi.org/10.1021/acs.jpcclett.8b00967>. PMID: 29886744.
- L. Chen, L. Grajciar, P. Nachtigall, and T. Düren. Accurate prediction of methane adsorption in a metal–organic framework with unsaturated metal sites by direct implementation of an ab initio derived potential energy surface in gcmc simulation. *The Journal of Physical Chemistry C*, 115(46):23074–23080, 2011. doi: 10.1021/jp2090878. URL <https://doi.org/10.1021/jp2090878>.
- L. Chen, C. A. Morrison, and T. Düren. Improving predictions of gas adsorption in metal–organic frameworks with coordinatively unsaturated metal sites: Model potentials, ab initio parameterization, and gcmc simulations. *The Journal of Physical Chemistry C*, 116(35):18899–18909, 2012. doi: 10.1021/jp3062527. URL <https://doi.org/10.1021/jp3062527>.
- C. Chipot and A. Pohorille. *Free energy calculations*, volume 86. Springer, 2007.

- S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1):3887, Sept. 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06169-2. URL <https://doi.org/10.1038/s41467-018-06169-2>.
- M. Czjzek, H. Jobic, A. N. Fitch, and T. Vogt. Direct determination of proton positions in d-y and h-y zeolite samples by neutron powder diffraction. *The Journal of Physical Chemistry*, 96(4):1535–1540, 1992. doi: 10.1021/j100183a009. URL <https://doi.org/10.1021/j100183a009>.
- M. W. Deem, R. Pophale, P. A. Cheeseman, and D. J. Earl. Computational discovery of new zeolite-like materials. *The Journal of Physical Chemistry C*, 113(51):21353–21360, 2009. doi: 10.1021/jp906984z. URL <https://doi.org/10.1021/jp906984z>.
- V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi. Gaussian process regression for materials and molecules. *Chemical Reviews*, 121(16):10073–10141, 2021. doi: 10.1021/acs.chemrev.1c00022. URL <https://doi.org/10.1021/acs.chemrev.1c00022>. PMID: 34398616.
- M. Eckhoff and J. Behler. From molecular fragments to the bulk: Development of a neural network potential for mof-5. *Journal of Chemical Theory and Computation*, 15(6):3793–3809, 2019. doi: 10.1021/acs.jctc.8b01288. URL <https://doi.org/10.1021/acs.jctc.8b01288>. PMID: 31091097.
- H. Edelsbrunner, M. Facello, and J. Liang. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, 88(1):83–102, 1998. ISSN 0166-218X. doi: [https://doi.org/10.1016/S0166-218X\(98\)00067-5](https://doi.org/10.1016/S0166-218X(98)00067-5). URL <https://www.sciencedirect.com/science/article/pii/S0166218X98000675>. Computational Molecular Biology DAM - CMB Series.
- K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. Auxiliary basis sets to approximate coulomb potentials. *Chemical Physics Letters*, 240(4):283–290, 1995. ISSN 0009-2614. doi: [https://doi.org/10.1016/0009-2614\(95\)00621-A](https://doi.org/10.1016/0009-2614(95)00621-A). URL <https://www.sciencedirect.com/science/article/pii/000926149500621A>.

- A. Erlebach, P. Nachtigall, and L. Grajciar. Accurate large-scale simulations of siliceous zeolites by neural network potentials. *npj Computational Materials*, 8(1):174, Aug. 2022. ISSN 2057-3960. doi: 10.1038/s41524-022-00865-w. URL <https://doi.org/10.1038/s41524-022-00865-w>.
- J. D. Evans and F.-X. Coudert. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chemistry of Materials*, 29(18):7833–7839, 2017. doi: 10.1021/acs.chemmater.7b02532. URL <https://doi.org/10.1021/acs.chemmater.7b02532>.
- H. Fang, P. Kamakoti, P. I. Ravikovitch, M. Aronson, C. Paur, and D. S. Sholl. First principles derived, transferable force fields for co₂ adsorption in na-exchanged cationic zeolites. *Physical Chemistry Chemical Physics*, 15:12882–12894, 2013. doi: 10.1039/C3CP52246F. URL <http://dx.doi.org/10.1039/C3CP52246F>.
- A. H. Farmahini, S. Krishnamurthy, D. Friedrich, S. Brandani, and L. Sarkisov. Performance-based screening of porous materials for carbon capture. *Chemical Reviews*, 121(17):10666–10741, 2021. doi: 10.1021/acs.chemrev.0c01266. URL <https://doi.org/10.1021/acs.chemrev.0c01266>. PMID: 34374527.
- J. T. Frank, O. T. Unke, and K.-R. Müller. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems, 2022. URL <https://arxiv.org/abs/2205.14276>.
- D. Fu and M. E. Davis. Carbon dioxide capture with zeotype materials. *Chemical Society Reviews*, pages –, 2022. doi: 10.1039/D2CS00508E. URL <http://dx.doi.org/10.1039/D2CS00508E>.
- S. Goedecker. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *The Journal of Chemical Physics*, 120(21):9911–9917, 2004. doi: 10.1063/1.1724816. URL <https://doi.org/10.1063/1.1724816>.
- K. Gottschling, G. Savasci, H. Vignolo-González, S. Schmidt, P. Mauker, T. Banerjee, P. Rovó, C. Ochsenfeld, and B. V. Lotsch. Rational design of covalent cobaloxime–covalent organic framework hybrids for enhanced photocatalytic hydrogen evolution. *Journal of the American Chemical Society*, 142(28):12146–12156, 2020. doi: 10.1021/jacs.0c02155. URL <https://doi.org/10.1021/jacs.0c02155>. PMID: 32564604.

- D. S. Graham, X. Wen, D. V. Chulhai, and J. D. Goodpaster. Robust, accurate, and efficient: Quantum embedding using the huzinaga level-shift projection operator for complex systems. *Journal of Chemical Theory and Computation*, 16(4):2284–2295, 2020. doi: 10.1021/acs.jctc.9b01185. URL <https://doi.org/10.1021/acs.jctc.9b01185>. PMID: 32105469.
- L. Grajciar. *Quantum-chemical study of adsorption in microporous materials*. PhD thesis, Univerzita Karlova, 2013.
- L. Grajciar. Low-memory iterative density fitting. *Journal of Computational Chemistry*, 36(20):1521–1535, 2015. doi: <https://doi.org/10.1002/jcc.23961>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23961>.
- L. Grajciar. Pbs clusters embedded in sodalite zeolite cavities of different compositions: Unraveling the structural evolution and optical properties using ab initio calculations. *The Journal of Physical Chemistry C*, 120(47):27050–27065, 2016. doi: 10.1021/acs.jpcc.6b09423. URL <https://doi.org/10.1021/acs.jpcc.6b09423>.
- L. Grajciar, O. Bludský, and P. Nachtigall. Water adsorption on coordinatively unsaturated sites in cubtc mof. *The Journal of Physical Chemistry Letters*, 1(23):3354–3359, 2010. doi: 10.1021/jz101378z. URL <https://doi.org/10.1021/jz101378z>.
- L. Grajciar, A. D. Wiersum, P. L. Llewellyn, J.-S. Chang, and P. Nachtigall. Understanding co2 adsorption in cubtc mof: Comparing combined dft–ab initio calculations with microcalorimetry experiments. *The Journal of Physical Chemistry C*, 115(36):17925–17933, 2011. doi: 10.1021/jp206002d. URL <https://doi.org/10.1021/jp206002d>.
- L. Grajciar, J. Cejka, A. Zukal, C. Otero Arean, G. Turnes Palomino, and P. Nachtigall. Controlling the adsorption enthalpy of co2 in zeolites by framework topology and composition. *ChemSusChem*, 5(10):2011–2022, 2012. doi: <https://doi.org/10.1002/cssc.201200270>. URL <https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cssc.201200270>.
- L. Grajciar, O. Bludský, W. J. Roth, and P. Nachtigall. Theoretical investigation of layered zeolite frameworks: Interaction between ipc-1p layers derived from zeolite utl. *Catalysis Today*, 204:15–21, 2013. ISSN

- 0920-5861. doi: <https://doi.org/10.1016/j.cattod.2012.07.018>. URL <https://www.sciencedirect.com/science/article/pii/S0920586112005184>. Challenges in Nanoporous and Layered Materials for Catalysis.
- L. Grajciar, P. Nachtigall, O. Bludský, and M. Rubeš. Accurate ab initio description of adsorption on coordinatively unsaturated Cu^{2+} and Fe^{3+} sites in MOFs. *Journal of Chemical Theory and Computation*, 11(1):230–238, 2015a. doi: [10.1021/ct500711z](https://doi.org/10.1021/ct500711z). URL <https://doi.org/10.1021/ct500711z>. PMID: 26574221.
- L. Grajciar, M. Rubeš, O. Bludský, and P. Nachtigall. Accurate ab initio Description of Adsorption on Coordinatively Unsaturated Sites in Metal–Organic Frameworks. In *Metal–Organic Frameworks: Materials Modeling towards Engineering Applications*, page 578. CRC Press, 2015b. ISBN 978-981-4613-45-3. URL <https://www.routledge.com/Metal-Organic-Frameworks-Materials-Modeling-towards-Engineering-Applications/author/p/book/9789814613460>.
- L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko, and P. Nachtigall. Towards operando computational modeling in heterogeneous catalysis. *Chemical Society Reviews*, 47:8307–8348, 2018. doi: [10.1039/C8CS00398J](https://doi.org/10.1039/C8CS00398J). URL <http://dx.doi.org/10.1039/C8CS00398J>.
- S. Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, 2004. doi: <https://doi.org/10.1002/jcc.20078>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20078>.
- S. Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006. doi: <https://doi.org/10.1002/jcc.20495>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20495>.
- C. J. Heard, L. Grajciar, and P. Nachtigall. The effect of water on the validity of Löwenstein’s rule. *Chemical Science*, 10:5705–5711, 2019a. doi: [10.1039/C9SC00725C](https://doi.org/10.1039/C9SC00725C). URL <http://dx.doi.org/10.1039/C9SC00725C>.
- C. J. Heard, L. Grajciar, C. M. Rice, S. M. Pugh, P. Nachtigall, S. E. Ashbrook, and R. E. Morris. Fast room temperature lability of aluminosilicate

- zeolites. *Nature Communications*, 10(1):4690, Oct. 2019b. ISSN 2041-1723. doi: 10.1038/s41467-019-12752-y. URL <https://doi.org/10.1038/s41467-019-12752-y>.
- C. J. Heard, L. Grajciar, F. Uhlík, M. Shamzhy, M. Opanasenko, J. Čejka, and P. Nachtigall. Zeolite (in)stability under aqueous or steaming conditions. *Advanced Materials*, 32(44):2003264, 2020. doi: <https://doi.org/10.1002/adma.202003264>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202003264>.
- S. Heiles and R. L. Johnston. Global optimization of clusters using electronic structure methods. *International Journal of Quantum Chemistry*, 113(18):2091–2109, 2013. doi: <https://doi.org/10.1002/qua.24462>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24462>.
- N. J. Henson, A. K. Cheetham, and J. D. Gale. Theoretical calculations on silica frameworks and their correlation with experiment. *Chemistry of Materials*, 6(10):1647–1650, 1994. doi: 10.1021/cm00046a015. URL <https://doi.org/10.1021/cm00046a015>.
- J. Hermann and O. Bludský. A novel correction scheme for dft: A combined vdw-df/ccsd(t) approach. *The Journal of Chemical Physics*, 139(3):034115, 2013. doi: 10.1063/1.4813826. URL <https://doi.org/10.1063/1.4813826>.
- J. Hermann and A. Tkatchenko. Density functional model for van der waals interactions: Unifying many-body atomic approaches with nonlocal functionals. *Physical Review Letters*, 124:146401, Apr 2020. doi: 10.1103/PhysRevLett.124.146401. URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.146401>.
- J. Hermann, R. A. DiStasio, and A. Tkatchenko. First-principles models for van der waals interactions in molecules and materials: Concepts, theory, and applications. *Chemical Reviews*, 117(6):4714–4758, 2017. doi: 10.1021/acs.chemrev.6b00446. URL <https://doi.org/10.1021/acs.chemrev.6b00446>. PMID: 28272886.
- B. Herzog, M. Chagas da Silva, B. Casier, M. Badawi, F. Pascale, T. Bučko, S. Lebègue, and D. Rocca. Assessing the accuracy of machine learning thermodynamic perturbation theory: Density functional theory and beyond. *Journal of Chemical Theory and Computation*, 18(3):1382–1394, 2022. doi:

- 10.1021/acs.jctc.1c01034. URL <https://doi.org/10.1021/acs.jctc.1c01034>. PMID: 35191699.
- T. Ho and H. Rabitz. A general method for constructing multidimensional molecular potential energy surfaces from ab initio calculations. *The Journal of Chemical Physics*, 104(7):2584–2597, 1996. doi: 10.1063/1.470984. URL <https://doi.org/10.1063/1.470984>.
- T. V. Ho, P. Nachtigall, and L. Grajciar. The lewis acidity of three- and two-dimensional zeolites: The effect of framework topology. *Catalysis Today*, 304:12–21, 2018. ISSN 0920-5861. doi: <https://doi.org/10.1016/j.cattod.2017.06.025>. URL <https://www.sciencedirect.com/science/article/pii/S0920586117304649>. 7th Czech-Italian-Spanish Symposium on Zeolites and Catalysis.
- D. Hou, L. Grajciar, P. Nachtigall, and C. J. Heard. Origin of the unusual stability of zeolite-encapsulated sub-nanometer platinum. *ACS Catalysis*, 10(19):11057–11068, 2020. doi: 10.1021/acscatal.0c01344. URL <https://doi.org/10.1021/acscatal.0c01344>.
- K. M. Jablonka, D. Ongari, S. M. Moosavi, and B. Smit. Big-data science in porous materials: Materials genomics and machine learning. *Chemical Reviews*, 120(16):8066–8129, 2020. doi: 10.1021/acs.chemrev.0c00004. URL <https://doi.org/10.1021/acs.chemrev.0c00004>. PMID: 32520531.
- M. Jin. *Theoretical Investigation of the Zeolite Hydrolysis under Realistic Conditions*. PhD thesis, Univerzita Karlova, 2022.
- M. Jin, M. Liu, P. Nachtigall, L. Grajciar, and C. J. Heard. Mechanism of zeolite hydrolysis under basic conditions. *Chemistry of Materials*, 33(23):9202–9212, 2021a. doi: 10.1021/acs.chemmater.1c02799. URL <https://doi.org/10.1021/acs.chemmater.1c02799>.
- M. Jin, O. Veselý, C. J. Heard, M. Kubů, P. Nachtigall, J. Čejka, and L. Grajciar. The role of water loading and germanium content in germanosilicate hydrolysis. *The Journal of Physical Chemistry C*, 125(43):23744–23757, 2021b. doi: 10.1021/acs.jpcc.1c06873. URL <https://doi.org/10.1021/acs.jpcc.1c06873>.
- M. Jin, R. Manoj, L. Grajciar, C. J. Heard, F. Brivio, C. Lei, Z. Tošner, J. A. van Bokhoven, and P. Nachtigall. Dynamical equilibrium between brønsted

and lewis sites in zeolites: Framework-associated octahedral aluminum. *submitted*, 2022.

- Y. Jung, A. Sodt, P. M. W. Gill, and M. Head-Gordon. Auxiliary basis expansions for large-scale electronic structure calculations. *Proceedings of the National Academy of Sciences*, 102(19):6692–6697, 2005. doi: 10.1073/pnas.0408475102. URL <https://www.pnas.org/doi/abs/10.1073/pnas.0408475102>.
- K. F. Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Gläser, U. Krewer, K. Reuter, and J.-D. Grunwaldt. Future challenges in heterogeneous catalysis: Understanding catalysts under dynamic reaction conditions. *ChemCatChem*, 9(1):17–29, 2017. doi: <https://doi.org/10.1002/cctc.201600996>. URL <https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cctc.201600996>.
- J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chemical Reviews*, 121(16):9816–9872, 2021. doi: 10.1021/acs.chemrev.1c00107. URL <https://doi.org/10.1021/acs.chemrev.1c00107>. PMID: 34232033.
- H. S. Kim and K. B. Yoon. Increase of third-order nonlinear optical activity of pbs quantum dots in zeolite y by increasing cation size. *Journal of the American Chemical Society*, 134(5):2539–2542, 2012. doi: 10.1021/ja211547s. URL <https://doi.org/10.1021/ja211547s>. PMID: 22279958.
- M. Kim, W. J. Kim, T. Gould, E. K. Lee, S. Lebègue, and H. Kim. umbd: A materials-ready dispersion correction that uniformly treats metallic, ionic, and van der waals bonding. *Journal of the American Chemical Society*, 142(5):2346–2354, 2020. doi: 10.1021/jacs.9b11589. URL <https://doi.org/10.1021/jacs.9b11589>. PMID: 31920085.
- H. W. Klemm, M. J. Prieto, F. Xiong, G. B. Hassine, M. Heyde, D. Menzel, M. Sierka, T. Schmidt, and H.-J. Freund. A silica bilayer supported on ru(0001): Following the crystalline-to vitreous transformation in real time with spectro-microscopy. *Angewandte Chemie International Edition*, 59(26):10587–10593, 2020. doi: <https://doi.org/10.1002/anie.202002514>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202002514>.

- U. Kokcam-Demir, A. Goldman, L. Esrafil, M. Gharib, A. Morsali, O. Weingart, and C. Janiak. Coordinatively unsaturated metal sites (open metal sites) in metal–organic frameworks: design and applications. *Chemical Society Reviews*, 49:2751–2798, 2020. doi: 10.1039/C9CS00609E. URL <http://dx.doi.org/10.1039/C9CS00609E>.
- K. N. Kudin and G. E. Scuseria. Linear-scaling density-functional theory with gaussian orbitals and periodic boundary conditions: Efficient evaluation of energy and forces via the fast multipole method. *Physical Review B*, 61:16440–16453, Jun 2000. doi: 10.1103/PhysRevB.61.16440. URL <https://link.aps.org/doi/10.1103/PhysRevB.61.16440>.
- R. Lazarski, A. M. Burow, and M. Sierka. Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole methods. *Journal of Chemical Theory and Computation*, 11(7):3029–3041, 2015. doi: 10.1021/acs.jctc.5b00252. URL <https://doi.org/10.1021/acs.jctc.5b00252>. PMID: 26575740.
- R. Lazarski, A. M. Burow, L. Grajciar, and M. Sierka. Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole method: Analytical gradients. *Journal of Computational Chemistry*, 37(28):2518–2526, 2016. doi: <https://doi.org/10.1002/jcc.24477>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.24477>.
- S. Lee, H. Kim, and M. Choi. Controlled decationization of x zeolite: mesopore generation within zeolite crystallites for bulky molecular adsorption and transformation. *Journal of Materials Chemistry A*, 1:12096–12102, 2013. doi: 10.1039/C3TA11256J. URL <http://dx.doi.org/10.1039/C3TA11256J>.
- S. Li, H. Li, R. Gounder, A. Debellis, I. B. Müller, S. Prasad, A. Moini, and W. F. Schneider. First-principles comparison of proton and divalent copper cation exchange energy landscapes in ssz-13 zeolite. *The Journal of Physical Chemistry C*, 122(41):23564–23573, 2018. doi: 10.1021/acs.jpcc.8b07213. URL <https://doi.org/10.1021/acs.jpcc.8b07213>.
- Y. Li, L. Li, and J. Yu. Applications of zeolites in sustainable chemistry. *Chem*, 3(6):928–949, 2017. ISSN 2451-9294. doi: <https://doi.org/10.1016/j>.

chempr.2017.10.009. URL <https://www.sciencedirect.com/science/article/pii/S2451929417304370>.

- M. Liu, O. Veselý, P. Eliášová, M. Shamzhy, P. Lyu, and L. Grajciar. Identification of the most active sites for tetrahydropyranlation in zeolites: Mfi as a test case. *Catalysis Today*, 345:165–174, 2020. ISSN 0920-5861. doi: <https://doi.org/10.1016/j.cattod.2019.08.017>. URL <https://www.sciencedirect.com/science/article/pii/S0920586119304389>. 8th Czech-Italian-Spanish Symposium on Zeolites and Catalysis.
- A. Luenser, H. F. Schurkus, and C. Ochsenfeld. Vanishing-overhead linear-scaling random phase approximation by cholesky decomposition and an attenuated coulomb-metric. *Journal of Chemical Theory and Computation*, 13(4):1647–1655, 2017. doi: 10.1021/acs.jctc.6b01235. URL <https://doi.org/10.1021/acs.jctc.6b01235>. PMID: 28263577.
- S. M. Moosavi, K. M. Jablonka, and B. Smit. The role of machine learning in the understanding and design of materials. *Journal of the American Chemical Society*, 142(48):20273–20287, 2020. doi: 10.1021/jacs.0c09105. URL <https://doi.org/10.1021/jacs.0c09105>. PMID: 33170678.
- K. Muraoka, Y. Sada, D. Miyazaki, W. Chaikittisilp, and T. Okubo. Linking synthesis and structure descriptors from a large collection of synthetic records of zeolite materials. *Nature Communications*, 10(1):4459, Oct. 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-12394-0. URL <https://doi.org/10.1038/s41467-019-12394-0>.
- A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics, 2022. URL <https://arxiv.org/abs/2204.05249>.
- C. Müller, M. Sharma, and M. Sierka. Real-time time-dependent density functional theory using density fitting and the continuous fast multipole method. *Journal of Computational Chemistry*, 41(30):2573–2582, 2020. doi: <https://doi.org/10.1002/jcc.26412>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.26412>.
- P. Nachtigall, L. Grajciar, J. Pérez-Pariente, A. B. Pinar, A. Zukal, and J. Čejka. Control of co2 adsorption heats by the al distribution in fer

- zeolites. *Physical Chemistry Chemical Physics*, 14:1117–1120, 2012. doi: 10.1039/C1CP22816A. URL <http://dx.doi.org/10.1039/C1CP22816A>.
- D. Nachtigallová, O. Bludský, C. Otero Areán, R. Bulánek, and P. Nachtigall. The vibrational dynamics of carbon monoxide in a confined space—co in zeolites. *Physical Chemistry Chemical Physics*, 8:4849–4852, 2006. doi: 10.1039/B612238H. URL <http://dx.doi.org/10.1039/B612238H>.
- S. O. Odoh, C. J. Cramer, D. G. Truhlar, and L. Gagliardi. Quantum-chemical characterization of the properties and reactivities of metal–organic frameworks. *Chemical Reviews*, 115(12):6051–6111, 2015. doi: 10.1021/cr500551h. URL <https://doi.org/10.1021/cr500551h>. PMID: 25872949.
- R. Pederson, B. Kalita, and K. Burke. Machine learning and density functional theory. *Nature Reviews Physics*, 4(6):357–358, June 2022. ISSN 2522-5820. doi: 10.1038/s42254-022-00470-2. URL <https://doi.org/10.1038/s42254-022-00470-2>.
- C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat. Periodic local mp2 method for the study of electronic correlation in crystals: Theory and preliminary applications. *Journal of Computational Chemistry*, 29(13):2113–2124, 2008. doi: <https://doi.org/10.1002/jcc.20975>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20975>.
- I. Poltavsky and A. Tkatchenko. Machine learning force fields: Recent advances and remaining challenges. *The Journal of Physical Chemistry Letters*, 12(28):6551–6564, 2021. doi: 10.1021/acs.jpcllett.1c01204. URL <https://doi.org/10.1021/acs.jpcllett.1c01204>. PMID: 34242032.
- R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld. Big data meets quantum chemistry approximations: The delta-machine learning approach. *Journal of Chemical Theory and Computation*, 11(5):2087–2096, 2015. doi: 10.1021/acs.jctc.5b00099. URL <https://doi.org/10.1021/acs.jctc.5b00099>. PMID: 26574412.
- K. Reuter, C. P. Plaisance, H. Oberhofer, and M. Andersen. Perspective: On the active site model in computational catalyst screening. *The Journal of Chemical Physics*, 146(4):040901, 2017. doi: 10.1063/1.4974931. URL <https://doi.org/10.1063/1.4974931>.

- M. A. Rohrdanz, W. Zheng, and C. Clementi. Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Annual Review of Physical Chemistry*, 64(1):295–316, 2013. doi: 10.1146/annurev-physchem-040412-110006. URL <https://doi.org/10.1146/annurev-physchem-040412-110006>. PMID: 23298245.
- W. J. Roth, P. Nachtigall, R. E. Morris, P. S. Wheatley, V. R. Seymour, S. E. Ashbrook, P. Chlubná, L. Grajciar, M. Položij, A. Zukal, O. Shvets, and J. Čejka. A family of zeolites with controlled pore size prepared using a top-down method. *Nature Chemistry*, 5(7):628–633, July 2013. ISSN 1755-4349. doi: 10.1038/nchem.1662. URL <https://doi.org/10.1038/nchem.1662>.
- M. Rubeš, J. Kysilka, P. Nachtigall, and O. Bludský. Dft/cc investigation of physical adsorption on a graphite (0001) surface. *Physical Chemistry Chemical Physics*, 12:6438–6444, 2010. doi: 10.1039/C001155J. URL <http://dx.doi.org/10.1039/C001155J>.
- M. Rubeš, L. Grajciar, O. Bludský, A. D. Wiersum, P. L. Llewellyn, and P. Nachtigall. Combined theoretical and experimental investigation of co adsorption on coordinatively unsaturated sites in cubtc mof. *ChemPhysChem*, 13(2):488–495, 2012. doi: <https://doi.org/10.1002/cphc.201100602>. URL <https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cphc.201100602>.
- M. Rubeš, A. D. Wiersum, P. L. Llewellyn, L. Grajciar, O. Bludský, and P. Nachtigall. Adsorption of propane and propylene on cubtc metal–organic framework: Combined theoretical and experimental investigation. *The Journal of Physical Chemistry C*, 117(21):11159–11167, 2013. doi: 10.1021/jp401600v. URL <https://doi.org/10.1021/jp401600v>.
- I. Saha, A. Erlebach, P. Nachtigall, C. J. Heard, and L. Grajciar. Reactive neural network potential for aluminosilicate zeolites and water: Quantifying the effect of si/al ratio on proton solvation and water diffusion in h-fau. *ChemRxiv*, 2022. doi: 10.26434/chemrxiv-2022-d1sj9.
- J. Sauer. Ab initio calculations for molecule–surface interactions with chemical accuracy. *Accounts of Chemical Research*, 52(12):3502–3510,

2019. doi: 10.1021/acs.accounts.9b00506. URL <https://doi.org/10.1021/acs.accounts.9b00506>. PMID: 31765121.
- J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5(1):83, Aug. 2019. ISSN 2057-3960. doi: 10.1038/s41524-019-0221-0. URL <https://doi.org/10.1038/s41524-019-0221-0>.
- C. Schran, F. L. Thiemann, P. Rowe, E. A. Müller, O. Marsalek, and A. Michaelides. Machine learning potentials for complex aqueous systems made simple. *Proceedings of the National Academy of Sciences*, 118(38):e2110077118, 2021. doi: 10.1073/pnas.2110077118. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2110077118>.
- D. Schwalbe-Koda, S. Kwon, C. Paris, E. Bello-Jurado, Z. Jensen, E. Olivetti, T. Willhammar, A. Corma, Y. Román-Leshkov, M. Moliner, and R. Gómez-Bombarelli. A priori control of zeolite phase competition and intergrowth with high-throughput simulations. *Science*, 374(6565):308–315, 2021. doi: 10.1126/science.abh3350. URL <https://www.science.org/doi/abs/10.1126/science.abh3350>.
- K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. Schnet – a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018. doi: 10.1063/1.5019779. URL <https://doi.org/10.1063/1.5019779>.
- K. T. Schütt, O. T. Unke, and M. Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra, 2021. URL <https://arxiv.org/abs/2102.03150>.
- M. Sharma and M. Sierka. Efficient implementation of density functional theory based embedding for molecular and periodic systems using gaussian basis functions. *Journal of Chemical Theory and Computation*, 18(11):6892–6904, 2022. doi: 10.1021/acs.jctc.2c00380. URL <https://doi.org/10.1021/acs.jctc.2c00380>. PMID: 36223886.
- H. Sidky, W. Chen, and A. L. Ferguson. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Molecular Physics*, 118(5):e1737742, 2020. doi: 10.1080/00268976.2020.1737742. URL <https://doi.org/10.1080/00268976.2020.1737742>.

- M. Sierka. Synergy between theory and experiment in structure resolution of low-dimensional oxides. *Progress in Surface Science*, 85(9):398–434, 2010. ISSN 0079-6816. doi: <https://doi.org/10.1016/j.progsurf.2010.07.004>. URL <https://www.sciencedirect.com/science/article/pii/S0079681610000237>.
- A. Sodt, J. E. Subotnik, and M. Head-Gordon. Linear scaling density fitting. *The Journal of Chemical Physics*, 125(19):194109, 2006. doi: 10.1063/1.2370949. URL <https://doi.org/10.1063/1.2370949>.
- M. Stein and M. Heimsaat. Intermolecular interactions in molecular organic crystals upon relaxation of lattice parameters. *Crystals*, 9(12), 2019. ISSN 2073-4352. doi: 10.3390/cryst9120665. URL <https://www.mdpi.com/2073-4352/9/12/665>.
- F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler, and F. Glorius. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chemical Society Reviews*, 49:6154–6168, 2020. doi: 10.1039/C9CS00786E. URL <http://dx.doi.org/10.1039/C9CS00786E>.
- M. Sípka, A. Erlebach, and L. Grajciar. Constructing collective variables using invariant learned representations. *Journal of Chemical Theory and Computation*, 2023. doi: 10.1021/acs.jctc.2c00729. URL <https://doi.org/10.1021/acs.jctc.2c00729>. PMID: 36696574.
- H. V. Thang, L. Grajciar, P. Nachtigall, O. Bludský, C. O. Areán, E. Frýdová, and R. Bulánek. Adsorption of co2 in fau zeolites: Effect of zeolite composition. *Catalysis Today*, 227:50–56, 2014. ISSN 0920-5861. doi: <https://doi.org/10.1016/j.cattod.2013.10.036>. URL <https://www.sciencedirect.com/science/article/pii/S0920586113005117>. Molecular Sieves and Catalysis: Selected papers from CIS-5 Conference.
- H. V. Thang, J. Vaculík, J. Přeč, M. Kubů, J. Čejka, P. Nachtigall, R. Bulánek, and L. Grajciar. The brønsted acidity of three- and two-dimensional zeolites. *Microporous and Mesoporous Materials*, 282:121–132, 2019. ISSN 1387-1811. doi: <https://doi.org/10.1016/j.micromeso.2019.03.033>. URL <https://www.sciencedirect.com/science/article/pii/S1387181119301817>.
- O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, and K.-R. Müller. SpookyNet: Learning force fields with electronic degrees of

- freedom and nonlocal effects. *Nature Communications*, 12(1):7273, Dec. 2021a. ISSN 2041-1723. doi: 10.1038/s41467-021-27504-0. URL <https://doi.org/10.1038/s41467-021-27504-0>.
- O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021b. doi: 10.1021/acs.chemrev.0c01111. URL <https://doi.org/10.1021/acs.chemrev.0c01111>. PMID: 33705118.
- L. Vanduyffhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. Van Speybroeck. Quickff: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *Journal of Computational Chemistry*, 36(13):1015–1027, 2015. doi: <https://doi.org/10.1002/jcc.23877>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23877>.
- D. Verboekend and J. Pérez-Ramírez. Design of hierarchical zeolite catalysts by desilication. *Catalysis Science & Technology*, 1:879–890, 2011. doi: 10.1039/C1CY00150G. URL <http://dx.doi.org/10.1039/C1CY00150G>.
- L. B. Vilhelmsen, K. S. Walton, and D. S. Sholl. Structure and mobility of metal clusters in mofs: Au, pd, and aupd clusters in mof-74. *Journal of the American Chemical Society*, 134(30):12807–12816, 2012. doi: 10.1021/ja305004a. URL <https://doi.org/10.1021/ja305004a>. PMID: 22734664.
- D. Wales. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge Molecular Science. Cambridge University Press, Cambridge, UK, 2004. ISBN 9780521814157.
- D. J. Wales and J. P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997. doi: 10.1021/jp970984n. URL <https://doi.org/10.1021/jp970984n>.
- J. L. Whitten. Coulombic potential energy integrals and approximations. *The Journal of Chemical Physics*, 58(10):4496–4501, 1973. doi: 10.1063/1.1679012. URL <https://doi.org/10.1063/1.1679012>.

- P. Xu, M. Alkan, and M. S. Gordon. Many-body dispersion. *Chemical Reviews*, 120(22):12343–12356, 2020. doi: 10.1021/acs.chemrev.0c00216. URL <https://doi.org/10.1021/acs.chemrev.0c00216>. PMID: 33166127.
- K. Yang, G. Yang, and J. Wu. Quantitatively understanding the insights into co₂ adsorption on faujasite from the heterogeneity and occupancy sequence of adsorption sites. *The Journal of Physical Chemistry C*, 125(28):15676–15686, 2021. doi: 10.1021/acs.jpcc.1c04254. URL <https://doi.org/10.1021/acs.jpcc.1c04254>.
- Q. Zhu, Y. Gu, X. Liang, X. Wang, and J. Ma. A machine learning model to predict co₂ reduction reactivity and products transferred from metal-zeolites. *ACS Catalysis*, 12(19):12336–12348, 2022. doi: 10.1021/acscatal.2c03250. URL <https://doi.org/10.1021/acscatal.2c03250>.
- A. Zukal, A. Pulido, B. Gil, P. Nachtigall, O. Bludský, M. Rubeš, and J. Čejka. Experimental and theoretical determination of adsorption heats of co₂ over alkali metal exchanged ferrierites with different si/al ratio. *Physical Chemistry Chemical Physics*, 12:6413–6422, 2010. doi: 10.1039/C001950J. URL <http://dx.doi.org/10.1039/C001950J>.

6 Attached papers

6.1 Lukáš Grajciar. Low-memory iterative density fitting. *Journal of Computational Chemistry*, 36(20), 1521-1535, 2015.

- 6.2 Roman Lazarski, Asbjorn M. Burow, Lukáš Grajciar, and Marek Sierka. Density functional theory for molecular and periodic systems using density fitting and continuous fast multipole method: Analytical gradients. *Journal of Computational Chemistry*, 37(28), 2518-2526, 2016.

- 6.3 Lukáš Grajciar. PbS clusters embedded in sodalite zeolite cavities of different compositions: Unraveling the structural evolution and optical properties using ab initio calculations. *The Journal of Physical Chemistry C*, 120(47), 27050–27065, 2016.

- 6.4 Lukáš Grajciar, Christopher J. Heard, Anton A. Bondarenko, Mikhail V. Polynski, Jittima Meep-rasert, Evgeny A. Pidko and Petr Nachtigall. Towards operando computational modeling in heterogeneous catalysis. *Chemical Society Reviews*, 47, 8307–8348, 2018.

- 6.5 Andreas Erlebach, Petr Nachtigall, and Lukáš Grajciar. Accurate large-scale simulations of siliceous zeolites by neural network potentials. *npj Computational Materials*, 8, 174, 2022.

- 6.6 Martin Šípka, Andreas Erlebach, and Lukáš Grajciar. Constructing collective variables using invariant learned representations. *Journal of Chemical Theory and Computation*, 10.1021/acs.jctc.2c00729, 2023.

- 6.7 Mengting Jin, Ondřej Veselý, Christopher J. Heard, Martin Kubů, Petr Nachtigall, Jiří Čejka, and Lukáš Grajciar. The role of water loading and germanium content in germanosilicate hydrolysis. *The Journal of Physical Chemistry C*, 125(43), 23744–23757, 2021.

- 6.8 Mingxiu Liu, Ondřej Veselý, Pavla Eliášová, Mariya Shamzhy, Pengbo Lyu, and Lukáš Grajciar. Identification of the most active sites for tetrahydropyranylation in zeolites: MFI as a test case. *Catalysis Today*, 345, 165–174, 2020.

- 6.9 Christopher J. Heard, Lukáš Grajciar, Cameron M. Rice, Suzi M. Pugh, Petr Nachtigall, Sharon E. Ashbrook, and Russell E. Morris. Fast room temperature lability of aluminosilicate zeolites. *Nature Communications*, 10, 4690, 2019.

- 6.10 Mengting Jin, Mingxiu Liu, Petr Nachtigall, Lukáš Grajciar, and Christopher J. Heard. Mechanism of zeolite hydrolysis under basic conditions. *Chemistry of Materials*, 33(23), 9202–9212, 2021.

- 6.11 Christopher J. Heard, Lukáš Grajciar, Filip Uhlík, Mariya Shamzhy, Maksym Opanasenko, Jiří Čejka, and Petr Nachtigall. Zeolite (in)stability under aqueous or steaming conditions. *Advanced Materials*, 32(44), 2003264, 2020.

- 6.12 Christopher J. Heard, Lukáš Grajciar, and Petr Nachtigall. The effect of water on the validity of Löwenstein's rule. *Chemical Science*, 10, 5705–5711, 2019.

- 6.13 Ho Viet Thang, Petr Nachtigall, and Lukáš Grajciar. The Lewis acidity of three- and two-dimensional zeolites: The effect of framework topology. *Catalysis Today*, 304, 12-21, 2018.

- 6.14 Ho Viet Thang, Jan Vaculík, Jan Přech, Martin Kubů, Jiří Čejka, Petr Nachtigall, Roman Bulánek, and Lukáš Grajciar. The Brønsted acidity of three- and two-dimensional zeolites. *Microporous and Mesoporous Materials*, 282, 121-132, 2019.