Technische Universität München
Fakultät für Chemie
Lehrstuhl für Theoretische Chemie

The Electrostatic Gap: Combining Electrostatic Models with Machine Learning Potentials

Carsten Gerald Staacke

Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

Technische Universität München
TUM School of Natural Sciences

The Electrostatic Gap: Combining Electrostatic Models with Machine Learning Potentials

Carsten Gerald Staacke

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Priv.-Doz. Dr. Aras Kartouzian

Prüfer der Dissertation:

1. Prof. Dr. Karsten Reuter
2. Prof. Dr. Jennifer Rupp
3. Prof. Dr. Volker Deringer

Die Dissertation wurde am 31.08.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 12.10.2022 angenommen.

Concordia res parvae crescunt, discordia maximae dilabuntur.

Preface

This dissertation is publication-based, meaning its scientific content is published in a series of related, but independent articles, all of which have undergone the scientific peer-review process in international scientific journals. The first chapters therefore mainly serve as an introduction to methods and relevant literature. Summaries for each article are then provided in chapter 5. The presented work has been carried out at the Chair of Theoretical Chemistry of the Technical University of Munich (TUM) between March 2019 and December 2020, under the supervision of Prof Dr. Karsten Reuter and has been completed between January 2021 and May 2022 at the Fritz Haber Institute of the Max Planck Society in Berlin. A research stay between January 2022 and Mai 2022 hosted by Prof Dr. Gábor Csányi at the University of Cambridge complemented this work.

Munich, August 2022

Abstract

Growths in the economy and population, a modern lifestyle, and a fully digitalized and connected world increase the global energetic demands each year. Currently, fossil fuels make up 80% of the global energy consumption, the combustion of which being the main driving force for the disastrous effects of climate change. Controlling and reducing CO_2 emissions are therefore key challenges of modern society. Renewable energy sources such as wind and solar panel would be ideal solutions to this problem, however both are a.) not necessarily predictable and b.) not evenly distributed geographicaly. To enable an even energy distribution, we require efficient energy storage. In the past, the combustion of coal and oil has been so successful as that is what carbon-based chemicals are: extraordinarily efficient forms of energy storage.

Many applications, such as laptops, mobile phones, and electric vehicles, utilize lithium-ion batteries as their primary energy storage. While lithium-ion batteries using liquid electrolytes entered the market in 1991, all-solid-state lithium-ion batteries (ASS-LIB), although investigated for decades, are still not widely applied. They promise several advantages in comparison to liquid electrolyte batteries: minimizing fire hazards, longer cycle lifetimes, more comprehensive temperature ranges, and enhanced energy density by potential usage of Li metal anodes. In particular, solid electrolytes of the Li_2S-P_2S_5 (LPS) material class have gained substantial attention due to their favorable properties. First, they possess high RT conductivities of up to 10^{-2} S/cm for crystalline LPS components, which ranks them among the most conductive solid electrolytes. Secondly, they are composed of the earth-abundant elements sulfur and phosphorous enabling applications at large scales. However, this material class' design of potent SSE is hampered by the poor understanding of structure-property relations. This manifests in massive deviations in reported Li-ion conductivity in different experimental setups and from theory and experiment.

Simulations based on Density Functional Theory (DFT) or classical force fields (FF) have enabled material comprehension e.g. new insights into material properties for decades. Insights at the atomistic level are irreplaceable for a mechanistic understanding of chemical processes. Unfortunately, due to high computational costs, DFT methods are limited to small systems while providing a highly accurate and complete description. At a much reduced computational cost, classical FFs allow to account for such effects. Yet, here the problem is an often reduced accuracy in the description of the potential energy surface (PES). To this end, emerging Machine Learning (ML) methods have shown to be increasingly able to bridge this gap, with good first-principles accuracy at a much reduced computational cost. However, the basic assumption of locality, implying the neglect of long-range interactions, is problematic in many cases.

To this end, the central topic of this thesis is threefold. First, we intended to systematically identify systems where this locality assumption does not hold. We especially tried to understand when and why the locality assumption holds for polar and ionic systems and when it fails. Second, as we realized that local ML models accurately predict isotropic bulk material properties, we developed a near-universal Gaussian Approximation Potential (GAP) model for the crystalline and amorphous compounds in Li_2S-P_2S_5. We then used the GAP model to systematically investigate the effect of the local anion composition in glassy Li_2S-P_2S_5 compounds.

At the same time we realized that a short-range model can accurately describe isotropic systems, we understood that we need an accurate description of non-local interactions for non-isotropic systems. To this end, we developed the kernel-based charge equilibration scheme called kQEq. The novel kQEq schemes enable the prediction of partial charges based on local environments by including the ability to predict non-local charge transfer.

Zusammenfassung

Wirtschafts- und Bevölkerungswachstum, ein moderner Lebensstil, in seiner Gesamtheit eine vollständig digitalisierte und vernetzte Welt, erhöhen den Ausstoß von Kohlenstoffdioxid jedes Jahr. Die Verbrennung fossiler Energieträger in Industriesektoren und im Individualverkehr sind dabei der Hauptgrund für katastrophale Auswirkungen sowohl auf den Planeten Erde als auch auf uns als Menschheit. Daher ist die Reduktion von CO_2-Emissionen die größte Herausforderung, vor der die Welt heute steht. Bereits heute stammen 20% des weltweiten Energieverbrauchs aus erneuerbaren Energien. Bei der Energiegewinnung aus nachwachsenden Rohstoffen stehen wir vor dem Problem, dass Wind und Sonne a.) nicht vorhersagbar und b.) nicht gleichmäßig über die Erde verteilt sind. Um eine gleichmäßige Energieverteilung zu ermöglichen, benötigen wir daher effiziente Energiespeicher. In der Vergangenheit war die Verbrennung von Kohle und Öl so allgegenwärtig, da kohlenstoffbasierte Chemikalien genau das sind: ein außerordentlich effizienter Energiespeicher.

Lithium-Ionen-Batterien werden häufig als Energiespeicher der Wahl eingesetzt. Laptops, Mobiltelefone und Elektroautos wären ohne Lithium-Ionen-Akkus nicht realisierbar gewesen. Während Lithium-Ionen-Batterien mit flüssigen Elektrolyten 1991 auf den Markt kamen, haben All-Solid-State-Lithium-Ionen-Batterien trotz jahrzehntelanger Erforschung noch immer keine breite Anwendung gefunden. Sie versprechen mehrere Vorteile im Vergleich zu Flüssigelektrolytbatterien: höhere Leistungsdichte, Minimierung von Sicherheits- und Brandgefahren, längere Zyklenlebensdauer, umfassendere Temperaturbereiche und Erhöhung der Energiedichte durch die potenzielle Verwendung von Li-Metall-Anoden. Festkörperelektrolyte der Materialklasse Li_2S-P_2S_5 haben hier aufgrund ihrer bemerkenswerten Eigenschaften große Aufmerksamkeit erlangt. Erstens besitzen sie hohe Leitfähigkeiten von bis zu 10^{-2} S/cm, womit sie zu den leitfähigsten SSEs zählen. Zweitens bestehen sie aus den auf der Erde reichlich vorkommenden Elementen Schwefel und Phosphor und ermöglichen daher Anwendungen in großem Maßstab. Allerdings wird das Design potenter SSEs dieser Materialklasse durch das schlechte Verständnis der Beziehung von Struktur und Materialeigenschaft behindert. Dies zeigt sich z.B. in großen Abweichungen von Li-Ionen Leitfähigkeit aus Theorie und Experiment.

Simulationen basierend auf der Dichtefunktionaltheorie (DFT) oder klassischen Kraftfeldern (force fields, FF) beschleunigen die Materialanalyse seit Jahrzehnten. Simulation und Modellierung auf atomarer Ebene sind für das mechanistische Verständnis chemischer Prozesse unersetzlich. DFT-Methoden sind auf kleine Systeme beschränkt, liefern aber trotz hoher Rechenkosten eine sehr genaue und vollständige Beschreibung. Obwohl der Rechenaufwand bei der Verwendung eines klassischen FF gering ist, liefern diese hingegen oft nur eine vereinfachte Beschreibung eines Materials. Hier haben sich neue Methoden des maschinellen Lernens (ML) als zunehmend in der Lage erwiesen, diese Lücke zu schließen, indem sie die Genauigkeit der Dichtefunktionaltheorie bei stark reduzierten Rechenkosten ermöglichen.

Die Grundannahme der Lokalität, die die Vernachlässigung langreichweitiger Wechselwirkungen impliziert, ist in vielen Fällen problematisch. Zu diesem Zweck ist das zentrale Thema dieser Arbeit in drei Fragestellungen aufgeteilt. Zunächst war es notwendig Systeme zu identifizieren, bei denen diese Lokalitätsannahme nicht ausreicht. Der Fokus lag insbesondere darauf, wann und warum die Lokalitätsannahme für polare und ionische Systeme wie SSEs gilt und wann sie versagt. Im zweiten Schritt haben wir, als wir erkannten, dass eine genaue Beschreibung isotroper Materialien durch ein lokales ML-Modell erreicht werden kann, ein nahezu universelles Gaussian Approximation Potential (GAP) für die kristallinen und amorphen LPS Materialien entwickelt.

Wir haben das GAP-Model verwendet, um systematisch den Einfluss der lokalen Anionenzusammensetzung auf die Li-Ionen-Leitfähigkeit in glasartigen Li_2S-P_2S_5-Verbindungen zu untersuchen. Genauso wie wir erkannten, dass isotrope Systeme durch ein lokales Modell nahezu exakt beschrieben werden können, haben wir verstanden, dass wir eine genaue Beschreibung nicht-lokaler Wechselwirkungen für nicht-isotrope Systeme benötigen. Hierfür haben wir das Kernel-basierte Ladungsmodel namens kQEq entwickelt. kQEq ermöglicht die Vorhersage von Partialladungen basierend auf lokalen atomaren Umgebungen.

Contents

1 Introduction 1

2 Li_2S-P_2S_5 Solid-State Electrolytes and their challenges 3
  2.1 Solid-State Electrolytes . . . 3
  2.2 The Li_2S-P_2S_5 Solid-State Electrolytes . . . 6
  2.3 Challenges from a Modeling Perspective . . . 10

3 Machine-learning interatomic potentials in materials science 13
  3.1 Machine Learning Potentials: A general Overview . . . 13
  3.2 The Gaussian Approximation Potential Framework . . . 14
  3.3 Descriptors of local atomic environments . . . 14
    3.3.1 Kernel Function . . . 15
    3.3.2 *n*-Body Descriptors . . . 16
    3.3.3 Smooth Overlap of Atomic Positions Descriptor . . . 17
  3.4 Regression Models . . . 20
    3.4.1 Gaussian Process Regression . . . 20
    3.4.2 Sparse Gaussian Process Regression . . . 21

4 Long-Range Electrostatics and Non-local Charge Transfer 25
  4.1 First steps towards QEq . . . 25
  4.2 Machine Learning Charge Prediction Schemes . . . 29
    4.2.1 Third generation NN potentials . . . 30
    4.2.2 The CENT approach . . . 31

5 Publications 33
  5.1 On the role of long-range electrostatics in machine-learned interatomic potentials for complex battery materials . . . 34
  5.2 Kernel Charge Equilibration: Efficient and Accurate Prediction of Molecular Dipole Moments with a Machine-Learning Enhanced Electron Density Model . . 35
  5.3 Tackling structural complexity in Li_2S-P_2S_5 solid-state electrolyte using Machine Learning Potentials . . . 36
  5.4 Additional Work on Machine Learning Potentials: Machine Learning Surface Complexions of Rutile IrO_2 and RuO_2 . . . 37

6 Summary, Conclusions and Outlook 39

Danksagung 41

Bibliography 43

Nomenclature

As this thesis combines the nomenclature of different fields I will use the following notation throughout this thesis. It is closely related to the original publications i.e. the publication where the underlying theory has been introduced for the first time.

χ^E	Electronegativity
$\chi^{E,0}$	Electronegativity of the isolated atom
χ^E_{CENT}	Environment dependent electronegativity used in the CENT approach
χ^d	Descriptor vector
J_{ii}	Hardness of atom i
J^0_{ii}	Hardness of an isolated atom i
r_{ij}	Distance between atoms i and j
r^A	Atomic radius
q_i	Charge of atom i
Q_{tot}	Total Charge
$\rho_{i,\alpha}$	Element specific neighbour density in SOAP
w	Weighting in the element-specific neighbor densities
$\mathbf{p}$	Power spectra vector
p	Elements of the power spectra vector $\mathbf{p}$
c	Regression weight
R_n	Radial basis function
Y_{lm}	Spherical Harmonics
f_{cut}	cutoff function
k^d	Kernel function using descriptor vectors d
$\mathbf{K}_{NN}$	Kernel matrix of dimension N x N
λ	Regularization parameter
Σ_{NN}	Diagonal matrix containing the regularization parameters λ
N	Number of input parameters
N_t	Number of training points
U_E	Electrostatic potential energy
k_C	Coulomb constant
$E(+1)$	Energy equivalent to removing an electron
E_{tot}	Total energy
$\epsilon^d(\chi^d)$	local energy corresponding to a descriptor d
$\delta^{(d)}$	Scaling parameter of descriptor $d \in$ [2b,3b,SOAP]

List of Abbreviations

ASS-LIB	All-solid-state Lithium ion battery
CENT	Charge equilibration via neural network technique
DFT	d Density Functional Theory
EA	Electron Affinity
EN	Electronegativity
FF	Force Field
GAP	Gaussian Approximation Potential
GPR	Gaussian Process Regression
IP	Ionisation potential
kQEq	Kernel QEq
KRR	Kernel ridge regression
LPS	LiPS material class
ML	Machine Learning
NN	Neural Network
PES	Potential energy surface
QEq	Charge equilibration scheme by Rappe and Goddard
RT	Room temperature
SSE	Solid-state electrolyte

1 Introduction

"Our dangerous reliance on carbon-based fuels is at the core of all these problems- economics, environmental, national security." [1]

In his Nobel prize lecture in 2007, Al Gore pinpointed the need for an immediate carbon-free economy. In times of inner European aggression, Al Gore seems right. More than ever.

Thirteen years later, clean energy technologies' worldwide market lies at approximately 130 billion USD only.[2] Nevertheless, there are signs of improvement. Whilst the global economy suffered from the impact of COVID-19,[3] renewable energies expanded by their fastest rate in two decades.[2] Record sales of electric vehicles are set daily.[4] Policy changes, such as a solar panel mandate for new non-residential buildings in Germany [5, 6], indicate that the energy economy of future generations will be unrecognizable to our current one.[7–9] Although sustainable energy in our economy is without alternative, it is unclear what the supply with solely regenerative energy will look like.

Batteries are by far the largest part of the clean energy technology market, a market expected to increase to 1.2 trillion USD by 2050.[2] The current majority of lithium-ion batteries (LIBs) utilize liquid electrolytes, which are in competition with All-solid-state (ASS)-LIBs.[8, 10–12] In theory, these are advantageous over liquid-electrolye LIBs, as they minimize safety and fire hazards, have longer cycle lifetimes and more comprehensive temperature ranges, and enhance energy density via the potential use of Li metal anodes.[11, 12]

A promising subgroup of ASS-LIBs is the Li_2S-P_2S_5 (LPS) material class.[13] First, they possess high conductivities of up to 10^{-2} S/cm, even higher than the well known electrolytes $Li_{10}GeP_2S_{12}$ or $Li_{1.3}Al_{0.3}Ti_{1.7}(PO_4)_3$.[14, 15] Secondly, they consist of earth-abundant elements sulfur and phosphorous.[16] A critical, yet often neglected factor, as abundance is required for sustainable large-scale global implementation. [8]

When first commercial energy storage devices entered the market, sustainable energy research targeted specific energy technologies and related materials.[17] While there is an ongoing hunt for the jack-of-all-trades material,[18] design of potent SSE is typically hampered by the poor understanding of structure-property relations.[19] Hence, efforts such as the *e*-conversion cluster or the BIG-MAP project target disorder and materials interfaces that underlie these changes of material functions.[20] Examples of successful material improvements by a mechanistic understanding range from atomic-scale complexions [21, 22], via interface amorphization processes that improve stability and capacity of batteries [23–25], to nanoscale disorder in solar cell materials.[26, 27]

Likewise, using LPS on a large scale is hindered by a poor understanding of Li-ion conductivity mechanisms, amorphization and degradation processes.[13] From a modeling perspective, investigating these observations realistically at the atomistic level strains the capabilities of state-of-the-art theoretical approaches.[28] On one hand, the system sizes and simulation time scales required are prohibitive for first-principles methods such as density functional theory (DFT).[29, 30] That simply means that the sheer number of calculations, requires simulation approaches multiple orders of magnitudes faster than standard DFT. On the other hand, parameterizations for empirical potentials are often not available, and these potentials may ultimately lack the desired predictive accuracy. Fortunately, modern machine learning (ML) potentials are increasingly able to bridge this gap, promising first-principles accuracy at a much reduced computational cost.[28] In the past ML potentials have increasingly been a standard tool for atomistic simulations. The most prominent examples are Neural Networks (NN) [31–36] and Gaussian Approximation Potentials (GAP)[37–40], which enable simulation with nearly DFT accuracy while accelerating simulations almost reaching the speed of classical empirical potentials in some cases.[28, 30] As indicated by their name, empirical potentials gain their speed from a simple fixed analytical form.[41] Flexibility and transferability are dominated by choice of the functional form of the potential energy surface (PES).[42] Opposed to that, ML potentials do not have a fixed functional form, but instead Machine Learning is used to learn an approximated PES as accurately as needed. [30]

When first introduced, ML mehods such as NN used a fixed structure. The NN was then optimized for a certain number of degrees of freedom, i.e., number of atoms.[32] These cannot be used to predict energies for a different system size/different number of atoms, since the optimized weights are valid only for a fixed number of input nodes of the NN. In order to overcome this scaling limitation, the locality assumption was introduced: The total energy E of the system can be defined as a sum of atomic contributions ϵ_i and hence the energy associated with a given atom depends on its immediate environment but not on atoms outside a given cutoff radius.[32, 43, 44]

However, this locality approximation implies that long-range contributions arising, e.g., from electrostatic interactions, are neglected beyond a certain cutoff.[37] Although the need to include long-range electrostatics in a similar fashion as in empirical potentials appears straightforward at first glance, the success of short-range ML potentials for the modeling of certain properties of ionic and polar materials appears to say something different.[45–49] Therefore, the first task was to identify a system and simulation tasks where the locality assumptions hold in some cases but fail in others. To study the role of long-range effects, GAP were constructed with and without an electrostatic correction term. As we studied the role of anisotropy, we realised that current electrostatic and charge equilibration models lack the required flexiblity. Hence, we extended the classical charge equilibration model QEq by an environment-dependent electronegativity and showed that our kernel QEq model (kQEq) can be used to generate accurate and highly data-efficient models for molecular dipole moments.

Chapter 2 will introduce the general concepts of SSEs specifically, materials of the Li_2S-P_2S_5 class, and discuss current challenges. Chapter 3 focuses on the two parts characteristic for ML potentials, namely the atomic descriptor and the concept of regression. I discuss the GAP framework and give a tutorial like introduction to the regression methods used in this work. The final chapter 4 focuses on the derivation of the charge equilibration and long-range interaction and the discussion of previously published machine learning charge prediction schemes.

2 Li_2S-P_2S_5 Solid-State Electrolytes and their challenges

2.1 Solid-State Electrolytes

Lithium-ion batteries (LIBs) are currently the most efficient electrochemical energy storage technology in terms of energy and power densities, reliability, and rechargability. While LIBs with liquid electrolytes reached commercialisation at the beginning of the 90s, all-solid-state Lithium-ion batteries (ASS-LIBs) are still not widely used. ASS-LIBs promise several advantages, regarding reduced flammability and operation safety in electrified mobility, longer lifetimes, and higher energy density. [11, 12]

A common drawback of liquid electrolytes are side reactions in which soluble products are generated at one electrode and consumed or further reacted at the other electrode, often called chemical cross-talk.[50] A well known example is the Li-S shuttle effect.[51, 52] Soluble electrode components can diffuse, being responsible for the progressive leakage of active material from the cathode.[53] In solid-state electrolytes (SSEs) only lithium ion transfer is exhibited as the SSE acts as a functional separators with only resulting in minor self-discharge. The negligible self-discharge is typically attributed to a low residual electronic conductivity.[21] In liquid electrolytes, most compounds (Li ions and most anions) are mobile. The mobility of all chemical species can cause concentration gradients of the conducting salts. This salt gradient leads to bulk polarization, limiting the cell current. As in SSBs only Li ions are mobile, higher current densities can be reached and hence lower charging times are enabled. [11, 14] A desired feature of SSEs is the use of lithium-metal anodes. The hope was that due to mechanical rigidity of SSEs dendrite formation caused by electrodeposition of lithium can be prevented.

SSEs can be divided into three groups. Organic solid polymers, inorganic solids, and solid like dispersion of nanoparticles in liquids.[54] The last group is often referred to as semi-solid electrolytes and form a group on their own. Although solid polymer electrolytes appear to be the preferred choice as they can compensate for volume changes of electrodes by elastic and plastic deformation, they lack the required high ionic conductivities for battery operation.[55] Batteries using inorganic solids — either crystalline, glass or glass-ceramic in nature — have demonstrated to improve battery performance at high currents.[56] This is in contrast to a common misconception that SSEs are inherently poor ionic conductors at ambient temperature.[56] In fact a number of ternary and quaternary sulfides and thiophosphates (Li_2S-P_2S_5 and $Li_{10}GeP_2S_{12}$) have been reported to exhibit room temperature conductivities equal or even higher than typical liquid electrolytes.[14, 57, 58]

So why is it that ASSBs are not fully applicable already? The major drawback of many inorganic SSEs is their low thermodynamic stability.[59] Just like liquid electrolytes, SEs are easily oxidized at intermediate potentials and reduced at low potentials.[60] Protecting interfaces are therefore required to stabilize the electrode/electrolyte contact.[61] While oxides often experience mechanical failure through cracking, thiophosphates such as the herein studied (Li_2S-P_2S_5) are ductile and easily form dense cathode composites.[62]

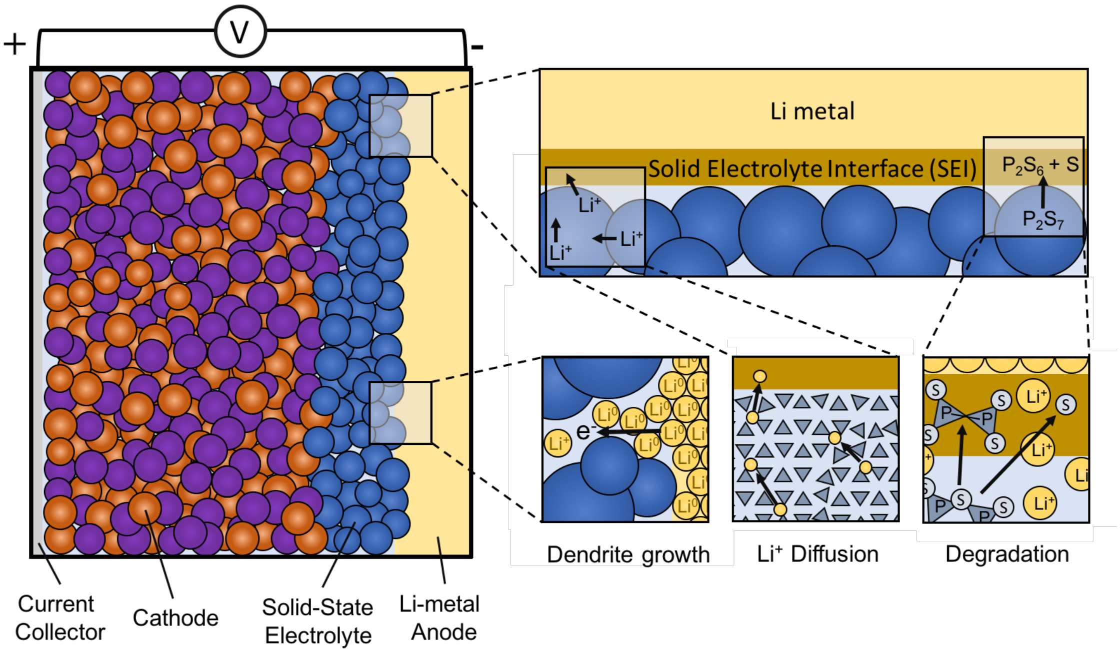

Fig. 2.1: *(Left): Schematic architecture of a solid-state battery adapted from Zeier and Janek.[11] Cathode and anode are separated by the solid-state electrolyte that allows for lithium-ion diffusion. Porous cathodes typically made of layered transition metals are the largest component of a battery. Anodes and cathodes both serve as the active storage component of the battery, but known cathode materials have a lower Li density. The cathode is typically coated on thin aluminium foils (current collector). In a SSB the classically used porous graphite anode could be replaced by Lithium metal. (Right top): Major challenges in modeling solid-state batteries: The use of a lithium metal anode can significantly increase the cell energy density resulting in a dramatically increased cell energy density. However, resistive solid electrolyte interfaces (SEI) may form between the lithium anode and the SSE (Right top). Inhomogeneous lithium metal deposition can form dendrites. Dendrite formation represents a simulation task that requires an accurate description of charge-transfer plus the ability to model SEI. Modeling lithium ion diffusion within a solid and over a solid-solid interface strains the capability of modern simulation techniques. Near ab inito accuracy for long and large simulation cells are required to determine a full picture of lithium ion motion. When modeling amorphization and the formation of interfaces, it is crucial to describe reactivity in order to predict degradation processes that happen during charging and discharging of a battery.*

In Fig. 2.1 I give a schematic architecture of a solid-state battery as proposed by Zeier and Janek and including current challenges in modeling solid-state batteries.[11] Already during synthesis but especially during operation, SSE form resistive solid electrolyte interfaces (SEI) between the lithium anode and the SSE. Although material interfaces can be engineered so they form a protective layer, realistic models are challenging in both cases.[21] Secondly, inhomogeneous lithium metal deposition in the anode as well as in the SSE itself can form dendrites.[63] For modelling dendrite formation we require an accurate description of charge-transfer plus the ability to model disorder in both electrodes and the SSE. The pure size of realistic simulation setups as well as the simulation time in order to realistically model lithium ion diffusion within a solid and over a solid-solid interface strains the capability of modern simulation techniques. Near *ab initio* accuracy, in long and large simulation cells are required to determine a full picture of lithium ion motion. Finally, when modeling disorder, defects, amorphization and the formation of interfaces, it is crucial that the underlying model is reactive and hence can predict degradation processes that happen during charging and discharging of a battery.

The key to further advance in the ASS-LIB field is to fully understand material disorder, interfacial properties and eventually feature a lithium metal electrode and outperform conventional lithium ion batteries.[11] In this light, it is evident that the development of Li ion batteries doesn't hinge on the development of appropriate solid bulk electrolyte materials, but a systematic understanding of material properties. In the following chapter I will summarize the structural and ionic conduction properties of the Li_2S-P_2S_5 material class, highlighting the challenges for an atomistic understanding of these materials.

2.2 The Li_2S-P_2S_5 Solid-State Electrolytes

Sulfur based electrolytes gained attention as SSEs due to an inherent high ionic conductivity (>10^{-3} Scm^{-1}), [64] and their good contact with electrode materials due to their mechanical soft nature. [65] The thiophosphate electrolytes in the "simple" Li_2S-P_2S_5 two component system (LPS family) are particularly interesting as they possess high conductivities without the addition of a transition metal (Si, Ge, Sn).[58] Accordingly, a variety of crystalline and amorphous materials in the LPS material class were reported over the past two decades.[13] In literature two nomenclatures for the LPS material class are found. Either the material is characterized by its chemical formula or by its mass percentage of Li_2S, i.e. Li_3PS_4 or 75 mol% Li_2S.

Although crystalline and amorphous materials are - in principle - of infinite structural diversity, in LPS five anionic species are commonly observed. As illustrated in Fig. 2.4, these species are characterized by central phosphorus atoms, each bonded to either four sulfur atoms or a neighbouring phosphorous atom and three sulfurs.

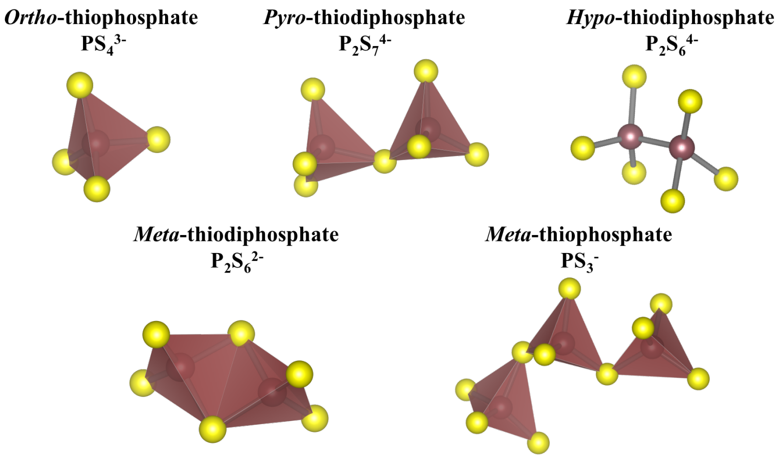

Fig. 2.2: *P-S microchemistry/ anionic species formed within Li_2S - P_2S_5.[13] Phosphorous is displayed in red, sulfur in yellow.*

These features were identified and characterized in detail for the first time by Dietrich *et al.* [58] in 2017 and can be summarized as follows:

- *Ortho*-thiophosphate moieties $\mathbf{PS_4^{3-}}$ (tetrahedra) are dominant for high Li_2S quantities (>75 mol%).

- *Pyro*-thiophosphate moieties $\mathbf{P_2S_7^{4-}}$ are formed by two corner sharing PS_4 tetrahedra. They are typically observed for Li_2S quantities <75 mol%.

- *Hypo*-thiodiphosphate moieties $\mathbf{P_2S_6^{4-}}$ are composed by two PS_3^- units with a direct P-P bond. It should be noted that phosphorus in $P_2S_6^{4-}$ has a formal oxidation state of +IV, whereas its formal charge in the rest of the LPS anions is +V.

- *Meta*-thiodiphosphate moieties $\mathbf{P_2S_6^{2-}}$ and *Meta*-thiophosphate $(\mathbf{PS_3^-})_n$ are both observed in the 60 mol% Li_2S crystalline composition. *Meta*-thiodiphosphate has two edge sharing PS_4^{3-} units, *Meta*-thiophosphate moieties are described as the polymeric corner-sharing chains of PS_3^-.

Crystalline Phases of Li_3PS_4, $Li_4P_2S_6$, and $Li_7P_3S_{11}$

In the ternary Li-P-S phase diagram, Li_3PS_4, $Li_4P_2S_6$, and $Li_7P_3S_{11}$ are probably the most commonly studied compounds.[13]

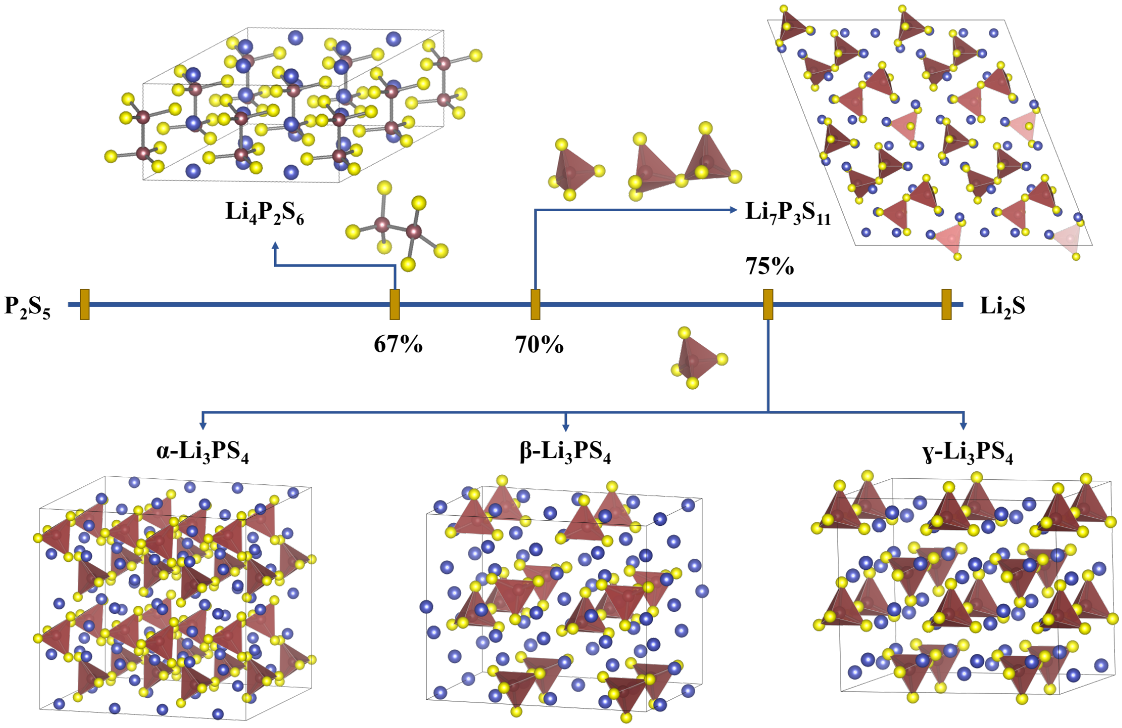

Fig. 2.3: *Crystal structures of the Li_2S–P_2S_5 composition line. The structures are grouped by their local P–S motifs. Note that $Li_4P_2S_6$ does not exactly lie on the Li_2S–P_2S_5 composition line, but is the crystallization product of glassy $Li_4P_2S_7$. (Li: blue; S: yellow; P: red)*

Li_3PS_4: Li_3PS_4 is the crystalline compound of Li_2S content of 75 mol% and can be found in three different phases: α, β, and γ -Li_3PS_4. All of them solely contain the simplest PS_4^{3-} anion. As seen in Fig. 2.4 conceptually, these three phases can be distinguished by their different arrangement of PS_4^{3-} anions. These are either all pointing in the same direction (γ), are arranged in a zig-zag fashion in one (α) or two directions (β) in space.[66] The room temperature stable phase is γ -Li_3PS_4, exhibiting a phase transition at 600K (γ to β) and at 800K (β to α).[67] As operating temperatures of ASSBs are usually smaller than 400K the α phase is less relevant for battery applications as it is not stable at RT. Apart from that, due to small stoichiometric changes, β-Li_3PS_4 has been shown to occur at RT. [68]

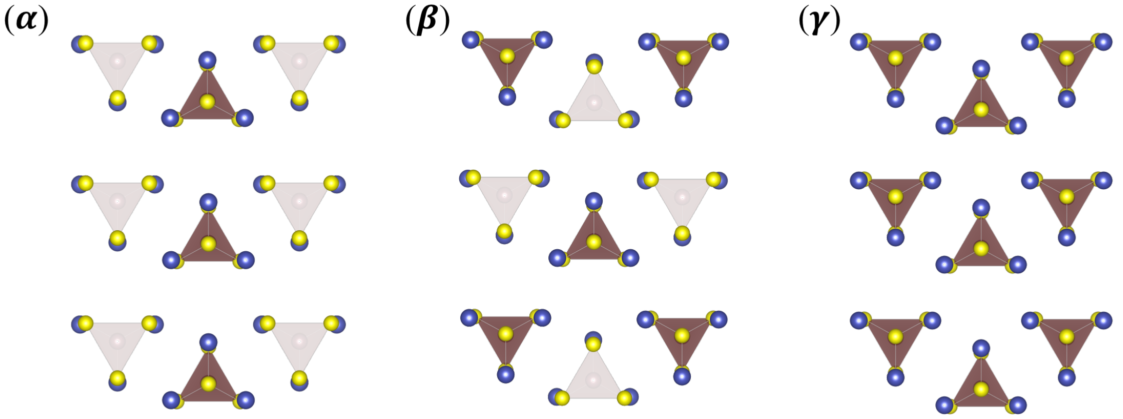

Fig. 2.4: Arrangements of PS_4 tetrahedra in the α, β, and γ in Li_3PS_4.[69] (Li: blue; S: yellow; P: red)

A fourth crystalline Li_3PS_4 called δ, was predicted by Iikubo *et al*. using an evolutionary algorithm under high-pressure of 5GPa.[70] To the best of my knowledge, δ-Li_3PS_4 has not been experimentally observed. Low RT-conductivities of 2.6 x 10^{-7} and 9.0 x 10^{-7} S cm^{-1} are experimentally reported for both relevant crystals (γ to β).[67]

$Li_7P_3S_{11}$: Crystalline LiP_3S_{11} is obtained for 70 mol% Li_2S. It is an extremely important member and well studied crystal of the LPS family due to its very high ionic conductivity (up to 1.7 x 10^{-2} Scm^{-1} at RT).[71] It has a triclinic P-1 space group, composed by a 1:1 ratio of PS_4^{3-} and $P_2S_7^{4-}$. Lithium ions are exclusively tetrahedrally coordinated.[72] Obtaining fully crystalline $Li_7P_3S_{11}$ is challenging due to its narrow stability window.

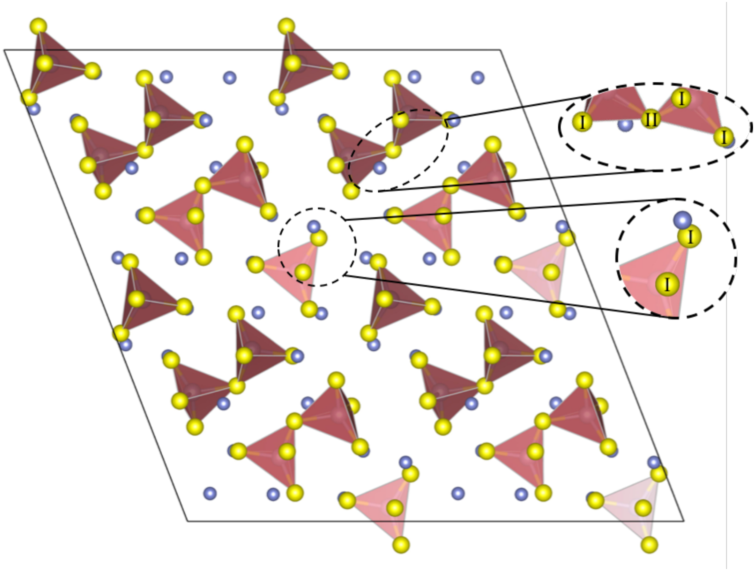

Fig. 2.5: Crystallographic structure of $Li_7P_3S_{11}$. It features terminal (I) as well bridging (II) sulfur atoms. (Li: blue; S: yellow; P: red)

As seen in Fig. 2.5 $Li_7P_3S_{11}$ features two distinct sulfur species, namely in bridging and terminal positions. Our evaluations using Hirshfeld population analysis [73] indicate that these species correspond to different charge states. It is generally obtained as a glass-ceramic. [74] While Li-ion conductivity in LPS is usually determined via a diffusion of defects (Li^+ vacancies), $Li_7P_3S_{11}$ exhibits a more collective Li^+ motion yielding superior conductivity compared to the other of its crystalline counterparts.[75, 76]

$Li_4P_2S_6$ Synthesizing $Li_4P_2S_6$ crystals is unique as its composition is not exactly in the stoichiometric line between Li_2S and P_2S_5. Li_2S contents of 67 mol% yields $Li_4P_2S_7$ glass. The corresponding reaction is:

$$2Li_2S + P_2S_5 \rightleftharpoons Li_4P_2S_7 \rightleftharpoons Li_4P_2S_6 + S$$

The crystal structure of $Li_4P_2S_6$ was initially reported as being P6 3/mcm when first synthesized in 1982,[77] recent studies predicted a stacking of $P_2S_6^{4-}$ yielding P-31 m.[78] Quenching the Li_2S contents of 67 mol% melt yields glassy $Li_4P_2S_7$. Subsequent annealing leads to the formation of crystalline $Li_4P_2S_6$ and sulphur.[77, 78] It is important to emphasize, that the local structures of glassy $Li_4P_2S_7$ and crystalline $Li_4P_2S_6$ differ significantly. The dominant anion of the $Li_4P_2S_7$ glass is $P_2S_7^{4-}$, while as shown in Fig. 2.6 in crystalline $Li_4P_2S_6$ the material solely contains $P_2S_6^{4-}$.[58, 78] $Li_4P_2S_6$ is a commonly observed byproduct of other members of the LPS material class.[78]

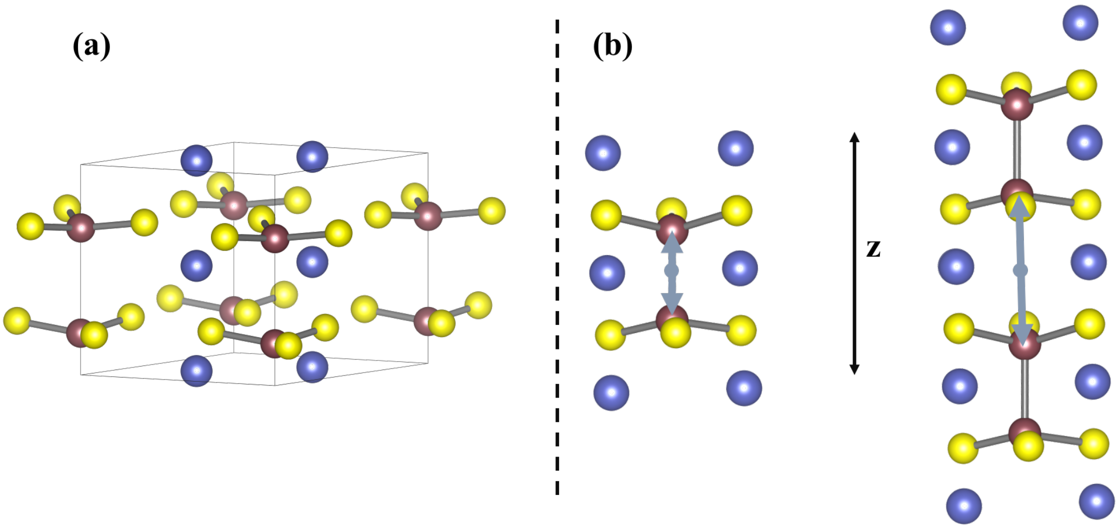

Fig. 2.6: *(a) rystallographic structure of $Li_4P_2S_6$ with a D_3d symmetry. The P–P bonds are along the crystallographic z axis. (b) Ball-and-stick diagram of the P_2S_6 ion units. The blue arrows indicate the two possible placements of the P.[79] The origin of the unit cell can either be at the center of the P-P bond or located between P_2S_6. (Li: blue; S: yellow; P: red)*

Especially, the metastable $Li_7P_3S_{11}$ crystal degrades at high temperatures above 1000K leading to the formation of $Li_4P_2S_6$.[80] Indeed, these reactions can be attributed to the evaporation of $Li_4P_2S_7$ and the high stability of the $Li_4P_2S_6$ phase. The ionic conductivity of $Li_4P_2S_6$ is very low (10^{-6} Scm^{-1}) but can be enhanced when synthesized as a glass ceramic structure.[58] In the glass-ceramic microstructure, the amorphous part is mainly composed of PS_4^{3-} units.[58]

Glasses and Thiophosphate Microchemistry

While LPS crystals have fixed stoichiometries and P-S anion matrices, high structural variability is found in LPS glasses.[72, 78] The plasticity arising from the amorphous nature of the glasses makes them more robust to mechanical stress, exerted by volume changes during cell cycling.[11]

Amorphous structures of $x\mathrm{Li_2S}$-(100-x)$\mathrm{P_2S_5}$ have been synthesized and systematically studied in the range $60 < x < 80$.[78] Similar to crystalline LPS, the choice of x influences the anion composition. Except for *Meta*-thiodiphosphate ($\mathrm{P_2S_6^{2-}}$) all thiophosphate anions were experimentally verified via ^{31}P magic angle spinning nuclear magnetic resonance (MAS-NMR) and Raman measurements.[78, 81, 82] In general though, compositions are dominated by $\mathrm{PS_4^{3-}}$ and $\mathrm{P_2S_7^{4-}}$ with the corresponding species ratio highly depending on the choice of x. At low $\mathrm{Li_2S}$ contents ($60 \leq x \leq 70$) $\mathrm{P_2S_7^{4-}}$ is the dominating anion, while higher $\mathrm{Li_2S}$ content favours $\mathrm{PS_4^{3-}}$ anions.

The $\mathrm{P_2S_6^{4-}}$ anion content is affected by the method of synthesis. A ball milling approach yields up to 10 % $\mathrm{P_2S_6^{4-}}$ anion independent of x.[78, 83] In contrast to that a microwave assisted synthesis route minimizes the formation of $\mathrm{P_2S_6^{4-}}$ anion at RT.[78, 81, 82] Thiophosphate-chains ($\mathrm{PS_3^{-}}$) have only been demonstrated for low $\mathrm{Li_2S}$ contents (x=60).[58, 78]

Essential differences of crystalline and glassy LPS, in addition to anion content, are $\mathrm{Li^+}$ conduction mechanisms. For glassy $\mathrm{Li_3PS_4}$, the so-called paddlewheel effect has been found to increase the RT conductivity by several orders of magnitude.[84] The paddlewheel effect describes quasi continuous $\mathrm{PS_4^{3-}}$ re-orientations during $\mathrm{Li^+}$ migration. Li ion conductivities are between 10^{-5} and 10^{-4} $\mathrm{Scm^{-1}}$. They are less conductive than the $\mathrm{Li_7P_3S_{11}}$ crystal, but significantly higher than the crystalline $\mathrm{Li_3PS_4}$ phases.[78–80, 85, 86] An interesting experimentally observed trend in LPS glasses is an increased Li-ion conductivity with increasing $\mathrm{Li_2S}$ content.[58]

2.3 Challenges from a Modeling Perspective

Realistic modelling of battery materials is an ongoing journey, challenging time and size scales of all available methods we currently have in our toolbox. Ranging from polaron assisted Li diffusion[87], studies on nano-scale complexions using empricial potentials[21] to kinetic Monte Carlo investigations of Lithium intercalated in graphite[88] our group always tried to push state of the art approaches to its limits. Developments in the field of operando spectroscopy enabled detailed structural insight into batteries during operation. A key finding of operando experiments, not just in battery research but also catalysis and other fields, is the awareness that a static picture of a material is not enough and sometimes even wrong.[89] To establish multiscale relations between atomistic features and a material's macroscopic behaviour and electrochemical properties we need to describe atomic interactions as accurate as possible on a size and time scale that is statistically robust in an ensemble like fashion. So far, the methods of choice are based on DFT for small/short but very accurate simulations and empirical potentials for large/long simulations that are limited in their accuracy. Empirical potentials gain their speed and robustness from fixed functional forms which comes at the cost of lacking reactivity or describing chemical variations in LPS material inaccurately. In that respect my personal experience was always that accuracy in a numerical sense often seemed to be taken too seriously and accuracy in a sense of consistency with experimental observations to be underrepresented.

The approach taken in this thesis, is to replace fixed functional forms of empirical potentials by flexible functional forms of machine learning interatomic potentials in order to combine the accuracy of first principle methods with the speed of empirical potentials. In materials modelling we are typically interested in the total energy E_{tot} of a system. We can define E_{tot} as a sum of local contributions E_{local} and long-range electrostatic E_{ES} yielding the following energy expression

$$E_{tot} = E_{local} + E_{ES} \tag{2.1}$$

The following two chapters will evolve around this expression of the total energy. While chapter 2 focuses on the local energy contribution E_{local} retrieved from the GAP framework, chapter 3 derives the E_{ES} expression in the context of the charge equilibration scheme QEq. Hence, the underlying structure of each chapter is always the same: Starting from the energy expression of either the local or electrostatic term each section introduces key steps how these energy expressions are obtained.

3 Machine-learning interatomic potentials in materials science

3.1 Machine Learning Potentials: A general Overview

In principle, the most accurate way to obtain total energies and atomic forces of a system is by solving the Schrodinger equation (SE).Unfortunately, an analytic solution is only feasible for very-simple systems such as the hydrogen atom. For larger chemical structures, the SE is typically solved approximately. However, even with approximations, an accurate numerical solution of the SE is a computationally extraordinary demanding task.

In the past, simple empirical functions are commonly used to model the relevant interactions. From these FF, energies and forces can be obtained with much reduced costs. However, while offering a qualitatively reasonable description of chemical interactions, the accuracy of the underlying FF and hence the quality of the simulations can be very limited.[90] ML methods could bridge this gap between accuracy of *ab initio* methods and efficiency of classical FFs. When using ML methods the user aims to train an algorithm to learn the functional relationship between inputs (chemical descriptors) and outputs (properties) from patterns or structure in the training data.

As we already introduced, in order to create a general ML potential that can be employed for systems of varying size and composition, just as with many empirical potentials (e.g. EAM, Tersoff) a locality assumption is typically made.[32, 38, 91, 92] The system's total energy is thus approximated purely as a sum of local (atomic) contributions:

$$E_{tot} \approx E_{local} = \sum_{i}^{N} \epsilon(Z_i, \chi_i) f_{cut} \tag{3.1}$$

where the sum runs over the N atoms in the system and each atom i contributes with an energy ϵ that only depends on its atomic number Z_i and its local chemical environment, represented by the descriptor χ_i. [32, 38, 93–95] This implies that electrostatic contributions outside a cutoff, in the above function defined by a cutoff function f_{cut}, are negligible.

Three components are needed to generate an ML potential for a given material/material class:

- A database of reference structures and associated quantum-mechanical data.
- A way to represent the atomic structure such that these can be used by the ML algorithm
- The regression or "learning" algorithm itself.

While I discussed the target materials in the previous chapter, I will discuss tasks two and three in the following chapter.

3.2 The Gaussian Approximation Potential Framework

In this work I used the Gaussian approximation potential (GAP) framework.[38, 39] The GAP software is implemented in the QUIP code.[96] As well as in other ML frameworks, the total energy of an atomistic system is a sum of atomic ("local") energies, from training data that consist of the system's cartesian coordinates, total energies and their derivatives. In GAP, the two components for modeling are the representation of atomic environments typically using n-body (with n=2,3) descriptors as well as the many-body descriptor smooth overlap of atomic positions (SOAP), and the regression task which is in this framework a Gaussian Process Regression (GPR).

The commonly used energy expression in GAP is

$$\begin{aligned} E_{total} = E_{local} &= (\delta^{(2b)})^2 \sum_{i \in pairs}^{N_t} \epsilon^{(2b)}(\chi_i^{(2b)}) \\ &+ (\delta^{(3b)})^2 \sum_{j \in triplets}^{N_t} \epsilon^{(3b)}(\chi_j^{(3b)}) \\ &+ (\delta^{(SOAP)})^2 \sum_{a \in atoms}^{N_t} \epsilon^{(SOAP)}(\chi_a^{(SOAP)}) \end{aligned} \tag{3.2}$$

"2b", "3b", and "SOAP" denote two-, three-, and many-body interactions each containing a scaling parameter $\delta^{(d)}$. The scaling parameter defines the energy contributions of a given interaction. To the best of my knowledge, Volker Deringer and Gabor Csanyi first introduced that energy expression for a GAP on amorphous Carbon in 2017.[37] From that energy expression, I will now introduce the ingredients of the GAP framework in a step-by-step manner starting with descriptors/representations of local atomic environments followed by the GPR framework.

3.3 Descriptors of local atomic environments

The set of descriptors $\chi^d = \{\chi_1^d, \chi_2^d, ..., \chi_N^d\}$ with $d \in \{2b, 3b, MB\}$ encode the local environment of every atom i.[97] One possible terminology defines a descriptor by being a mapping of an atomic configuration i, typically a molecule or a solid defined by the cartesian coordinates and chemical identity of its N atoms, into a suitable representation for the regression task.[30] The mapping associates i with points in feature space, which are then used to construct a machine-learning model to regress (fit) a structure-property relation.

One can define four desirable properties/requirements for a structural descriptor: First, the descriptor should obey fundamental physical symmetries. Second, the descriptor should be smooth i.e. continuous changes of a structure should yield a smooth change in the associated descriptor. Third, it should be complete hence inequivalent structures should yield distinguishable descriptors. Finally, in order to be able to ensure transferability to systems of varying molecular size, the descriptor should be additive e.g. structures should be decomposed in a sum of local environments.[98, 99]

The need to remove the dependency of the Cartesian coordinates on the origin and orientation of the reference system, is a key in chemical simulations. Already in classical FF different sets of internal coordinates (bonds, angles, and torsions) have been proposed, based on chemical intuition, as invariant descriptors of molecular geometry. In fact classical FF have been extremely effective in the modeling of biological systems. When ensuring fundamental physical symmetries this means that the descriptor $\chi^{(d)}$ has to be invariant [1] under symmetry operations and permutation of equivalent atoms in addition to translation and rotation of structures.

In practice, in addition to SOAP, a multitude of descriptors $\chi^{(d)}$ are available and different authors have their favorite descriptor.[32, 38, 93, 94]

3.3.1 Kernel Function

Before we can introduce n-Body Descriptors, we need to highlight the role of kernel functions. In GAP, the local energy corresponding to each type of descriptor $d \in \{2\text{b},3\text{b},\text{MB}\}$ can be given by a linear combination of kernel functions

$$\epsilon_i^{(d)}(\chi_i^{(d)}) = \sum_{t=1}^{N_t} w_t^{(d)} K^{(d)}(\chi_i^{(d)}, \chi_t^{(d)}) \tag{3.3}$$

where t denotes one of N_t training configurations $\chi_t^{(d)}$. Each training configuration has a weighting coefficient w_t. The weighting coefficient is attained during fitting. The covariance kernel K quantifies how similar the input configuration $\chi^{(d)}$ is to the training configuration $\chi_t^{(d)}$. In practice, one sparsifies the representation and only allows the sum to range over a number of "representative points" drawn from the full training database ($N_t \ll N_{full}$).

So what is now a kernel? In machine learning, a 'kernel' is usually used to refer to the kernel trick, a method of using a linear classifier to solve a non-linear problem. The kernel transform linearly inseparable data to linearly separable ones. As seen in Figure 3.1 the kernel function K is thus applied on each data point to map the original non-linear observations into a higher-dimensional space in which they become separable. In this work we thus use the following nomenclature: Cartesian coordinates are transformed by basis functions of the underlying descriptor, yielding a descriptor vector. Each element of that descriptor vector is referred to as a feature.

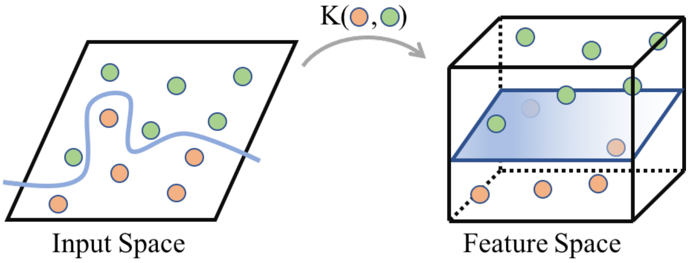

Fig. 3.1: *Graphical representation of the kernel trick: In 2-d, non-separable data (input space) become linearly separable data in 3-d (feature space) after applying the kernel transformation K.*

---

[1] It is important to note, that $\chi^{(d)}$ should be invariant to the things that the target property is invariant to. For example, energy is a translation and rotationally invariant property, but a molecular dipole isn't.

In the GAP-framework, for 2b and 3b contributions, a squared exponential kernel is

$$K^{(d)}(\chi^{(d)}, \chi_t^{(d)}) = exp\left(- \sum_{\varsigma} \frac{|\chi_{\varsigma}^{(d)} - \chi_{t,\varsigma}^{(d)}|^2}{2\sigma_{\varsigma}^2} \right) \tag{3.4}$$

with ς being an index running over the elements of the descriptor vector χ and σ being the width of the exponential kernel. The elements of the descriptor vector are often referred to as features. A polynomial kernel is used in order to compare many-body (SOAP) environments

$$K^{(SOAP)}(\chi^{(SOAP)}, \chi_t^{(SOAP)}) = \left(\frac{\chi^{(SOAP)} \cdot \chi_t^{(SOAP)}}{\sqrt{\chi^{(SOAP)} \cdot \chi_t^{(SOAP)} \chi^{(SOAP)} \cdot \chi_t^{(SOAP)}}} \right)^{\zeta} \tag{3.5}$$

3.3.2 n-Body Descriptors

In the case of two-body contributions we use the above introduced local energy expression

$$\epsilon_i^{(2b)}(\chi_i^{(2b)}) = \sum_{t=1}^{N_t} w_t^{(d)} K^{(2b)}(\chi_i^{(2b)}, \chi_t^{(2b)}) \tag{3.6}$$

using a squared exponential kernel yielding

$$\epsilon_i^{(2b)}(\chi_i^{(2b)}) = \sum_{t=1}^{N_t} w_t^{(d)} exp\left(- \sum_{\varsigma} \frac{|\chi_{\varsigma}^{(d)} - \chi_{t,\varsigma}^{(d)}|^2}{2\sigma_{\varsigma}^2} \right) \tag{3.7}$$

In this case, we simply use distances r_{ij} between atoms i and j, hence the descriptor has one single scalar component and the local energy expression for a two body contribution simplifies to

$$\epsilon_i^{(2b)}(\chi_i^{(2b)}) = \sum_{t=1}^{N_t} \sum_{j=1}^{N_t} w_t^{(d)} exp\left(- \frac{|r_{ij} - r_{jk}|^2}{2\sigma_{2b}^2} \right) \tag{3.8}$$

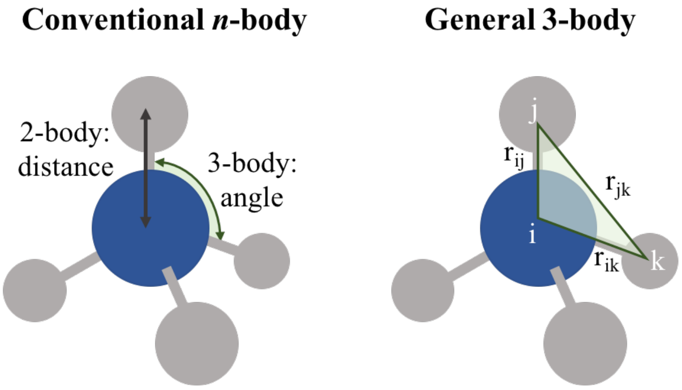

Fig. 3.2: *Graphical representation of n-body (here n=2,3) descriptors for a simple methane molecule. 2-body terms encode the chemical environments via atomic distances. 3-body terms add angular information between atoms. Instead of describing 3-body terms via the angle between three atoms, a general 3-body defines angular information by three atomic distances.*

The simplest way to guarantee rotational, translational and permutational invariance is to represent an atomic structure via a set of atomic pairwise distances $\mathbf{r}_{ij}$ accounting for all atom pairs i and j.

The two-body contribution $\mathrm{E}^{(2b)}$ can be obtained by comparing all eligible distances $r_{ij} = |r_i - r_j|$ in a molecule or crystal via a smooth and regular kernel function $k^{2b} : \mathbb{R}\mathrm{x}\mathbb{R} \rightarrow \mathbb{R}$

$$k^{(2b)}(r_{ij}, r_{ik}) = exp\left(-\frac{|r_{ij} - r_{jk}|^2}{2\sigma_{2b}^2} \right) \tag{3.9}$$

defining the similarity of distances via a Gaussian function of width σ_{2b}. The width σ_{2b} controls the smoothness of the kernel. The energy $E_i^{(2b)}(\mathbf{r}_{ij})$ of atom i is given by summing local energies, each defined by its kernel values $k^{(2b)}(r_{ij})$

$$E_i^{(2b)}(\mathbf{r}_{ij}) = \sum_j^{N_t} \epsilon_{ij} = \sum_j^{N_t} w_j k^{(2b)}(r_{ij}, r_{ik}) \tag{3.10}$$

with w being the regression weights. For a detailed discussion of the regression weights see sec. 3.4.

Three-body descriptors add angular information, i.e. energy contributions arising from the relative position of three atoms i, j, k. Equally to the 2-body descriptor, we can define a three-body kernel $k^{3b} : \mathbb{R}^2\mathrm{x}\mathbb{R}^2 \rightarrow \mathbb{R}$ defining the energy as

$$E_i^{(2b,3b)}(\mathbf{r}_{ijk}) = \sum_j^{N_t} w_j k^{(2b)}(r_{ij}, r_{ik}) + \sum_j^{N_t} w_{jk} k^{(3b)}(r_{ik}, r_{ij}, r_{jk}) \tag{3.11}$$

In the 3-body descriptor of GAP, distances r_{ij}, r_{ik}, and r_{jk} are not directly used, but a different form is used to enforce symmetry over permutation of the neighbor atoms j and k.

$$\mathbf{q}^{(3b)} = \begin{pmatrix} r_{ij} + r_{ik} \\ (r_{ij} - r_{ik})^2 \\ r_{jk} \end{pmatrix} \tag{3.12}$$

But as shown in Fig. 3.2, already a 3-body descriptor can be defined in multiple ways. Either by a set of three atomic distances (or a combination of these) or by an angle between three atoms. While it can be crucial to include higher body order terms to achieve a higher accuracy, when defining a many-body representation, the possibilities to define the descriptor space seems nearly infinite. [97] In the next section I will introduce the widely applied Smooth Overlap of Atomic Positions (SOAP) descriptor as my many-body descriptor of choice.

3.3.3 Smooth Overlap of Atomic Positions Descriptor

In the past, SOAP has been one of the most widely applied many-body representations.[39] In SOAP, neighboring atoms are represented by overlapping Gaussian functions yielding the neighbour density. While the neighbor density is by construction already invariant to permutation and translation, rotational invariance still has to be introduced. This is achieved by expanding the neighbor density in the basis of orthogonal radial distribution functions and spherical harmonics. We again start by the energy expression

$$\epsilon_i^{(d)}(\chi_i^{(d)}) = \sum_{t=1}^{N_t} w_t^{(d)} K^{(d)}(\chi_i^{(d)}, \chi_t^{(d)}) \tag{3.13}$$

using a polynomial kernel

$$\epsilon_i^{(d)}(\chi_i^{(d)}) = \sum_{t=1}^{N_t} w_t^{(d)} \Big(\frac{\chi^{(SOAP)} \cdot \chi_t^{(SOAP)}}{\sqrt{\chi^{(SOAP)} \cdot \chi_t^{(SOAP)} \chi^{(SOAP)} \cdot \chi_t^{(SOAP)}}} \Big)^{\zeta} \tag{3.14}$$

In a first step the atomic structure is transformed into atomic density fields ρ for each species α. We can define a set of element-specific neighbor densities $\rho_{i,\alpha}$ for each central atom i

$$\rho_{i,\alpha}(\mathbf{r}) = \sum_j f_{cut} \cdot (\mathbf{r}_{ij}) exp\Big(- \frac{|\mathbf{r} - \mathbf{r}_{ij}|^2}{2\sigma_\alpha^2} \Big) \tag{3.15}$$

with σ_α being an element-specific descriptor width and f_{cut} a cutoff function. As discussed the neighbour density in eq. 3.15 is already invariant to permutations between equivalent atoms and translation. Rotational invariance can then be introduced by expanding the neighbor density in a set of orthonormal radial basis functions g_n and spherical harmonics Y_{lm}

$$\rho_{i,\alpha}(\mathbf{r}) = \sum_{nlm} w_{nlm}^{i,\alpha} g_n(r) Y_{lm}(\hat{\mathbf{r}}) \tag{3.16}$$

In this definition of the neighbour density $\rho_{i,\alpha}$ $\mathbf{r}$ is the vector containing the cartesian coordinates of atom i, r in the radial basis functions $g_n(r)$ is the magnitude and $\hat{\mathbf{r}}$) in the spherical harmonics $Y_{lm}(\hat{\mathbf{r}})$) is the direction. The coefficients $w_{nlm}^{i,\alpha}$ can be obtained by projecting the density onto the basis functions via

$$w_{nlm}^{i,\alpha} = \iiint dV g_n(r) Y_{lm}(\theta, \phi) \rho_{i,\alpha}(\mathbf{r}). \tag{3.17}$$

One typically does not use the entire powerspectrum, but the elements of a finite truncation of the power spectrum (up to $n \leq n_{max}$ and $l \leq l_{max}$). Hence, these hyperparameters n_{max}, l_{max} have to be chosen according to the investigated system. The now rotationally invariant output is the partial power spectra vector $\mathbf{p}$ with the individual elements:

$$p_{nn'l}^{i\alpha\alpha'} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m w_{nlm}^{i,\alpha} \cdot w_{n'lm}^{i,\alpha'} \tag{3.18}$$

In an alternative definition of SOAP by Ceriotti and coworkers[97], the spherical harmonics can be defined by angular terms θ and ϕ as

$$\rho_{i,\alpha}(\mathbf{r}) = \sum_{nlm} w_{nlm}^{i,\alpha} g_n(r) Y_{lm}(\theta, \phi) \tag{3.19}$$

with the coefficients $w^{i,\alpha}_{nlm}$

$$w^{i,\alpha}_{nlm} = \iiint dV g_n(r) Y_{lm}(\theta, \phi) \rho_{i,\alpha}(\mathbf{r}). \tag{3.20}$$

and the elements of the powerspectrum

$$p^{i\alpha\alpha'}_{nn'l} = \sqrt{\frac{1}{2l+1}} \sum_{m} w^{i,\alpha}_{nlm} \cdot w^{i,\alpha'}_{n'lm} \tag{3.21}$$

The kernel function for SOAP is a dot product of the power spectrum elements. When then using a polynomial kernel (raising the dot product to a small integer ζ) we get the final polynomial kernel K as defined above

$$K(\chi_i, \chi_t) = \left(\frac{\chi_i \cdot \chi_t}{\sqrt{\chi_i \cdot \chi_t \chi_i \cdot \chi_t}} \right)^{\zeta} \tag{3.22}$$

In addition to the derivation of SOAP, Fig. 3.3 gives a graphical depiction of the above SOAP derivation. This figure focuses on the symmetry (permutation, translation, rotation) that is introduced in the individual steps.

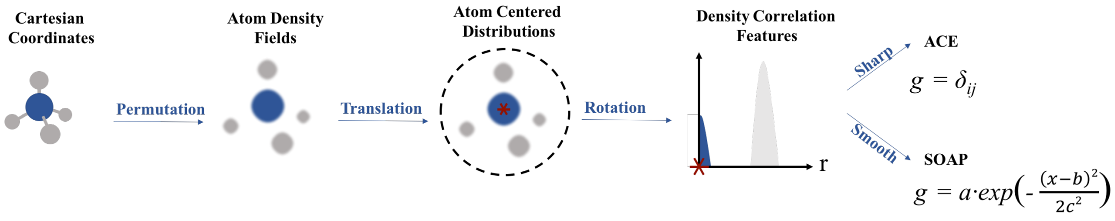
Cartesian Coordinates
Permutation
Atom Density Fields
Translation
Atom Centered Distributions
Rotation
Density Correlation Features
r
Sharp
ACE
$g = \delta_{ij}$
Smooth
SOAP
$g = a{\cdot}exp\left(-\frac{(x-b)^2}{2c^2}\right)$

Fig. 3.3: *Schematic summary of the steps in a symmetrized field construction. In contrast to cartesian coordinates, the atom density field is permutationally invariant. By summing over the continuous translation group we yield an atom centered distribution. Using orthonormal radial functions and spherical harmonics yields a discrete set of coefficients that transform as spherical harmonics. The atomic density functions can be either finite-width Gaussians, which leads to representations to SOAP features, or Dirac δ distributions, which recovers the third body order term of the atomic cluster expansion.[93]*

The energy expression can now be expressed in terms of kernel functions

$$\begin{aligned} E_{total} = E_{local} = (\delta^{(2b)})^2 \sum_{i}^{N_{full}} \sum_{t}^{N_t} K^{(2b)}(\chi^{(2b)}_i, \chi^{(2b)}_t) \\ + (\delta^{(3b)})^2 \sum_{j}^{N_{full}} \sum_{t}^{N_t} K^{(3b)}(\chi^{(3b)}_j, \chi^{(3b)}_t) \\ + (\delta^{(SOAP)})^2 \sum_{a}^{N_{full}} \sum_{t}^{N_t} K^{(SOAP)}(\chi^{(SOAP)}_a, \chi^{(SOAP)}_t) \end{aligned} \tag{3.23}$$

The coefficients w are determined during the fitting process. In the next section we turn to gaussian process regression (GPR).

3.4 Regression Models

The third component determining the use of ML in computational chemistry in addition to the database and a set of suitable descriptors is the regression of atomic properties. In that sense regression defines the functional dependence of a given quantity on the local structural environment. Regression models define y_i as a function of x_i and c_i, with c_i being the respective regression weights. The penalty term λ, often referred to as the regularization, is introduced in some cases

$$y = f(x, c) + \lambda \quad . \tag{3.24}$$

The aim is to find the function $f(x, c)$ that most closely fits the given data. In order to carry out regression tasks, the form of function f must be specified. First, I will introduce a generalized formalism of the underlying GPR approach including a simple tutorial-style example. The last section will discuss the formalism of a sparse regression approach and the role of sparsification for the prediction of atomic properties.

3.4.1 Gaussian Process Regression

In a recent review on "Gaussian Process Regression for Materials and Molecules" Deringer and coworkers defined two equivalent approaches deriving the GPR framework.[100] Both approaches highlight different aspects of the fitting process. While from my point of view the weight-space approach highlights the similarity of Kernel-Ridge Regression (KRR) and GPR based on the choice of regression weights, the function-space view discusses the fact that the estimator of the local energies only depends explicitly on the kernel function, and not on the basis functions. In this section I will discuss the weight-view derivation of GPR but will highlight key information taken from the function-space derivation.

In the weight-space view of GPR, a function $y(x)$ can be approximated by a function $f(x, c)$, defined as a linear combination of N data points in the training set (usually atoms)

$$f(x, c) = \sum_{n}^{N} c_n k(x, x_n) = \mathbf{c}^{\mathrm{T}} \mathbf{K}_{NN}, \tag{3.25}$$

with c being the regression weights, k the kernel function, and x_n the input data. I use the matrix notation for that regression problem $\mathbf{c}^{\mathrm{T}}\mathbf{K}_{\mathrm{NN}}$ in order to be consistent with Fig. 3.4. The fitting of the GPR model to the data is done by finding the coefficients c that minimize the loss function

$$\mathcal{L} = \sum_{i} (y_i - f(x_i, c_i))^2 + \lambda^2 \sum_{n,n'}^{N} c_n k(x_n, x_{n'}) c_{n'} = ||(\mathbf{y}_{\mathrm{ref}} - \mathbf{y})||^2 + \Sigma \mathbf{c}^T \mathbf{K}_{NN} \mathbf{c}. \tag{3.26}$$

Σ is a diagonal matrix of size N containing all values of λ. By setting $\frac{d\mathcal{L}}{dc} = 0$ to minimize $\mathcal{L}$ and solve for $\mathbf{c}$, to obtain:

$$\mathbf{c} = \left(\mathbf{K} - \Sigma\right)^{-1} \mathbf{y}_{\mathrm{ref}}. \tag{3.27}$$

In the GPR framework the Gaussian kernel, sometimes also called square exponential kernel,

$$k(x, x_n) = exp\left(\frac{(x - x_n)^2}{2\sigma_{\text{length}}^2}\right) \tag{3.28}$$

with σ_{length} being the spatial length scale/width of the Gaussian is used.

Let's now evaluate an example function: $y \colon \mathbb{R} \rightarrow \mathbb{R}$. These data points are reproduced from Ref. [101]. The exact function form of $y(x)$ is unknown so we collect a total of N observations $\mathbf{y}_{\text{ref}}$. In Fig. 3.4 the increasing similarity of the predicted function $f(x, c)$ and the true function $y(x)$ with an increasing number of data points is depicted. While this in principle shows the relevance of additional data for the accuracy of the prediction, additional data come with an increasing computation cost. In the last section of this chapter I discuss this and address the role of sparsification for such problems.

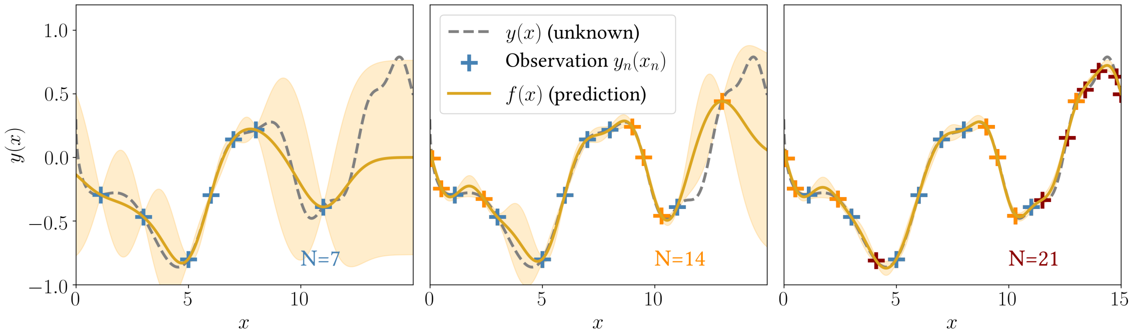

Fig. 3.4: *GPR prediction of a function $f(x, c)$ (solid yellow line) of an unknown one-dimensional function $y(x)$ (dashed black line). The orange shaded area is the standard deviation. GPR provides an uncertainty information that is not available with a kernel ridge. Using an increasing number of observations N of data points* $\mathbf{y}$ *the model hyperparameters are fixed.*

When updating the Gaussian process (hence adding a new observation n based on its predicted uncertainty), the updated Gaussian process is constrained to the possible functions that fit the observations N. Hence, the mean of $f(x, c)$ intercepts all observations N. Additionally, it is clear that the standard deviation is higher away from the observations, which reflects our lack of knowledge about these part of the function.

The kernel width σ_{length} in Gaussian kernel functions together with the regularization λ are both crucial for the smoothness and the accuracy of the predicted function. The role of σ_{length} and λ is depicted in Fig. 3.5. The choice is related to the often used terms 'underfitting' and 'overfitting' being the left and right panel in Fig. 3.5.

When decreasing regularization and/or kernel width ($\lambda \longrightarrow 0$, $\sigma \longrightarrow 0$) we obtain an improved alignment of training data N and the predicted function f. As illustrated in the right panel of Fig. 3.5 an increased precision on the prediction of the training data N when decreasing λ and or σ diminishes the predictive accuracy for input data unequal to the training data. It is therefore crucial to determine the accuracy/ predictive power on a set of independent data points.

3.4.2 *Sparse Gaussian Process Regression*

In practice when training an interatomic potential, full GPR is not applicable for large training sets, because the computational costs of training scale with $\mathcal{O}(N^3)$ i.e. cubical with the training

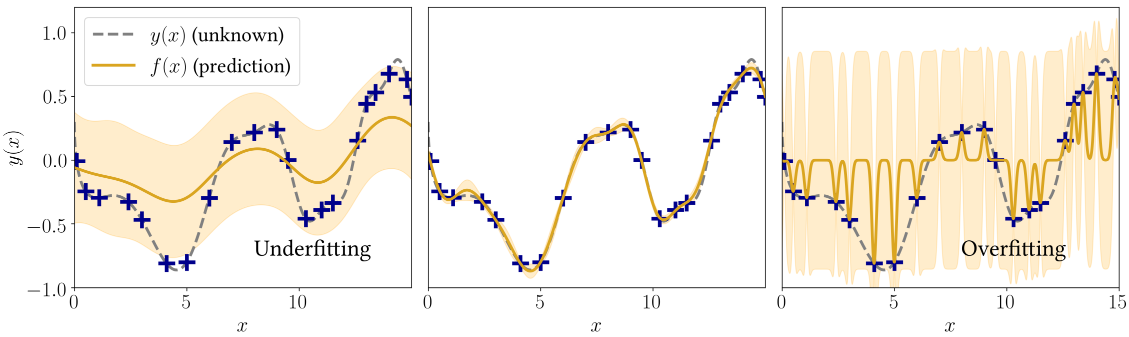

Fig. 3.5: *Effect of regularization λ and kernel width σ on the GPR prediction (solid yellow line). Left panel: For small penalties on the prediction (large λ, σ) $f(x, c)$ is centered around the average of the observations. Although the predicted function is smooth, it can not accurately predict the original function and has a high uncertainty. Right panel: In contrast, small λ, σ overfit the function y(x). Although high accuracy is achieved for the training data, high uncertainties are yielded for all other data points. Center panel: An optimal performance is acchieved when balancing smoothness and accuracy with appropriate λ and σ.*

set size.[100] In GAP, instead of using all data points N a smaller set of representative points M(i.e. $M \ll N$) defining the 'sparse' representation is used. We first recall the loss function

$$\mathcal{L} = \sum_{n}^{N} (y_n - f(x_n, c_n))^2 + R \tag{3.29}$$

where the relative importance of individual data points being controlled by R. Opposite to the general case discussed above in sparse GPR the regularization term R is now only depending on this representative set of M instead of N data points

$$R = \Sigma \sum_{m,m'}^{M} c_m k(x_m, x_{m'}) c_{m'}. \tag{3.30}$$

Adding eq. 3.29 to eq. 3.30 and rewriting the loss function in matrix form yields:

$$\mathcal{L} = \left(\mathbf{y}_{\text{ref}} - \mathbf{K}_{NM}\mathbf{c}\right)^T \Sigma^{-1} \left(\mathbf{y}_{\text{ref}} - \mathbf{K}_{NM}\mathbf{c}\right) + \mathbf{c}^T \mathbf{K}_{MM}\mathbf{c}. \tag{3.31}$$

The matrix elements are defined as $\mathbf{K}_{NM} = k(\mathrm{x}_n, \mathrm{x}_m)$ where N indicates the number of data points in the data set and M indicates the number of representative points, respectively. Minimizing $\mathcal{L}$ we obtain:

$$-\mathbf{K}_{MN}\Sigma^{-1}\mathbf{y} + \mathbf{K}_{MN}\Sigma^{-1}\mathbf{K}_{NM}\mathbf{c} + \mathbf{K}_{MM}\mathbf{c} = 0 \quad . \tag{3.32}$$

When solving for $\mathbf{c}$ we yield the following expression:

$$\mathbf{c} = \left(\mathbf{K}_{MM} + \mathbf{K}_{MN}\Sigma^{-1}\mathbf{K}_{NM}\right)^{-1} \mathbf{K}_{MN}\Sigma^{-1}\mathbf{y}_{\text{ref}} \tag{3.33}$$

Although, in the first glance eq. 3.27 appears to be 'simpler' compared to eq. 3.33, Fig. 3.7 clarifies that the coefficient vector is shorter. Fig 3.7 also sorts out the misconception of sparsification, that data points are left out. In sparse GPR, the full data vector $\mathbf{y}$ is used in training, yet now M ('sparse') locations are chosen to evaluate (unkown) input data. The coefficient vector is therefore of length M in the case of sparse GPR, while full GPR yields a coefficent vector of lenght N. Evaluating $f(x_{\text{new}}, c)=\tilde{y}$ for a new configuration $\mathbf{x}_{\text{new}}$ is done using eq. 3.25

$$f(x_{\text{new}}, \mathbf{c}) = \mathbf{c}^T \mathbf{k}(x_{\text{new}}) \tag{3.34}$$

A key benefit of sparse GPR becomes apparent. While the cost of prediction in full GPR scales with N, sparse GP is now independent of N. The notation of $\mathbf{k}(\mathbf{x}_{\text{new}})$ is used for the vector of kernel values at $\mathbf{x}_{\text{new}}$ and the set of representative points

$$\mathbf{k}(\mathbf{x})_m = k(x, x_m) \tag{3.35}$$

But what does that now mean in the case of the GAP framework? All GAP models are sparse kernel models, i.e. the basis functions for the linear expansion of the atomic energy do not directly correspond to the set of input data N. In the GAP framework individual atomic environments are chosen as the elements of the representative set M, and the corresponding kernel basis functions are used to expand the atomic energy. The crucial factor, beside the actual number of sparse points (which can be seen as convergence parameter), is the choice of representative environments for training. While for a two body-descriptor a homogeneously spaced grid is suitable due to low dimensionality, the CUR algorithm has been found to be a good algorithm to provide a decent set of representative data points within the SOAP-GAP model.[100]

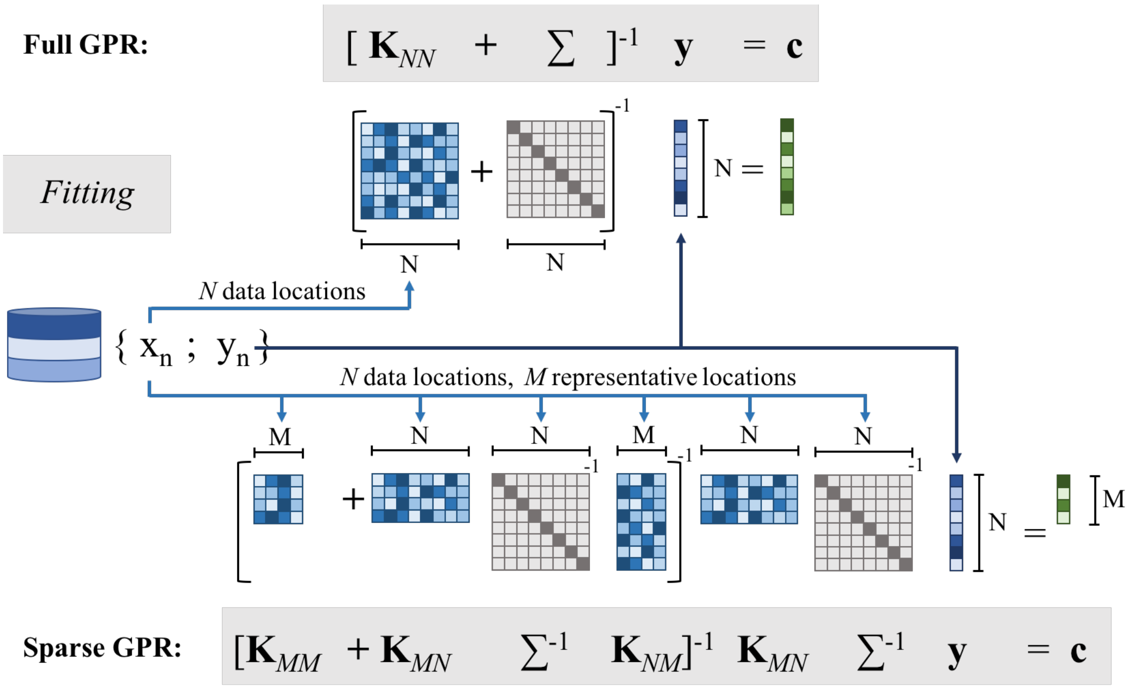
Full GPR:
$[\mathbf{K}_{NN} + \Sigma]^{-1}\ \mathbf{y} = \mathbf{c}$
Fitting
-1
N =
N
N
N data locations
$\{x_n ; y_n\}$
N data locations, M representative locations
M
N
N
M
N
N
-1
-1
-1
N
=
M
Sparse GPR:
$[\mathbf{K}_{MM} + \mathbf{K}_{MN}\ \Sigma^{-1}\ \mathbf{K}_{NM}]^{-1}\ \mathbf{K}_{MN}\ \Sigma^{-1}\ \mathbf{y} = \mathbf{c}$

Fig. 3.6: *Visualization of the full GPR (top) eq. 3.27 and sparse GPR (bottom) eq. 3.33. The training database consists of N entries. In full GPR all N entries are used to obain the coefficient vector c of length N. In sparse GPR, still all entries y are used, but M representative ("sparse") locations are chosen to obtain a coefficient vector c of length M. The figure is adapted from a review by Deringer and coworkers.[100]*

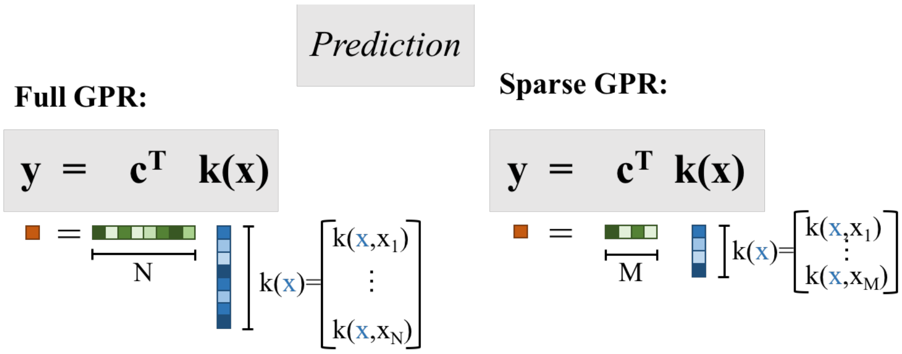
Prediction
Full GPR:
y = c^T k(x)
N
k(x)=
k(x,x_1)
⋮
k(x,x_N)
Sparse GPR:
y = c^T k(x)
M
k(x)=
k(x,x_1)
⋮
k(x,x_M)

Fig. 3.7: *The coefficient vector* c *(green) is computed, and can be used to make a prediction at a new location* $f(x_{new})$ *as given in eq. 3.34. The cost scales with the number of data locations, N. In sparse GPR, the full data vector y is used as well, but now M representative ('sparse') locations are chosen, with* $M \ll N$*. The coefficient vector is therefore of length M, and the cost of prediction is now independent of N. The figure is adapted from a review by Deringer and coworkers.[100]*

4 Long-Range Electrostatics and Non-local Charge Transfer

The locality assumption kick-started the fitting of ML potentials and their wide applicability spanning the range from molelucar to condensed systems.[37, 89] Even in water, where one would expect long-range electrostatics to be crucial for an accurate description, extraordinary accuracy of the structural and dynamic properties of bulk liquid water and different ice phases was achieved using a short-range potential.[102] This can be rationalized by the fact that these properties/materials are presumably highly isotropic, so that long-range interactions average out. The importance of long-range effects depends on the material and property of interest and thus demands a systematic analysis.

In this last chapter I will give a historical outline on long-range interactions. Starting with Coulomb's law from 18th century, to Sanderson's concept of electronegativity equalization introduced in the early 50s, and the first Charge Equilibration (QEq) scheme by Rappe and Goddard in 1991, section 4.1 will introduce the basic formalism of charge equilibration.

Section 4.2 covers machine-learning based charge equilibration approaches and recent developments in the field of machine learning based charge assignments.

4.1 First steps towards QEq

Coulomb's law, when first introduced in 1785 enabled the quantification of the force between two stationary, electrically charged particles. Charged particles' attraction or repulsion is directly proportional to the product of charges (q_i, q_j) and inversely proportional to the square of the distance between them:

$$F = k^c \frac{q_i q_j}{(r_i - r_j)^2} \quad . \tag{4.1}$$

Here, k^c is Coulomb's constant.

From Coulomb's law we can derive the electrostatic potential energy ($U_E(q_i)$) of one point charge q_i at position r_i in the presence of another point charge q_j as:

$$U_E(q_i) = k^c \frac{q_i q_j}{|r_i - r_j|} \quad , \tag{4.2}$$

as well as the electrostatic potential energy ($U_E(q_i)$) of one point charge q_i at position r_i in the presence of $N - 1$ other point charges q_j

$$U_E(q_i) = k^c q_i \sum_{j=1}^{N-1} \frac{q_j}{|r_i - r_j|} \quad , \tag{4.3}$$

In a molecule or condensed system we can then define the electrostatic potential energy of the entire system of N point charges as:

$$U_E = \frac{1}{2} k^c \sum_{i}^{N} q_i \sum_{j \neq i}^{N} \frac{q_j}{r_{ij}} \quad . \tag{4.4}$$

For small atomic distances r_{ij} e.g. in chemical bonds it is intuitive that the energy description can be beyond coulombic interactions. Linus Pauling defined electronegativity χ by the ability of an atom to attract shared electrons/electron density when forming a chemical bond.[103] He proposed an electronegativity scale which depends on bond energies, as a development of valence bond theory. By that we are able to understand that a covalent bond between two different atoms (i–j) is stronger than the average of the i–i and the j–j bonds. He proposed an electronegativity scale which depends on bond energies, as a development of valence bond theory. It is important to remember that electronegativity cannot be directly measured and strongly correlates with a number of other chemical properties. Two of these properties are the ionization potential (IP) and the electron affinity (EA).[104] The ionization potential is the energy needed to remove the outer valence electron. Electron affinity is the energy difference related to the injection of an extra electron. As such we can define the electronegativity of atom i as:

$$\frac{IP + EA}{2} = \chi_i^E \tag{4.5}$$

When two atoms i and j form a bond Sanderson postulated that the corresponding electronegativities equalize.[105] He first showed that concept based on bond lengths in alkali halide gas molecules, and proved that the same holds true for more than two atoms combining within a molecule. He defined the electronegativity as a stability ratio of the formed bond of atoms i and j

$$\chi^E = \frac{Z}{4.19 r_A^3 \rho_e} \tag{4.6}$$

where Z is the atomic number, r_A the atomic radius of each atom and ρ_e the electronic density of atom i. When comparing reported bond lengths to ionic or non-polar covalent radii, Sanderson identified electronegativities based on bond stability for different molecules. He then averaged χ^E to yield one electronegativity per element. But in principle he already put forth that the electronegativity depends on the formed bond and hence the atomic environment.

Rappe and Goddard extended the idea of the electronegativity equilibration of a bond and developed a framework to estimate the electrostatic potential energy of molecules and periodic systems by its charges. The QEq methods computes partial charges by using atomic coordinates and the two previous defined properties of isolated atoms $\chi_i^{E,0}$, J_{ii}^0. I use the notation $\chi_i^{E,0}$ instead of χ_i when we treat an isolated atom i instead of atom i in a molecule. This is one of the underlying assumptions: atomic ionization potential and electron affinity of isolated atoms are similar to the ones of the same element bonded inside a molecule or a crystal. The concepts involved in the QEq approach manages to combine earlier ideas of Pauling (chemical bond) [103], Mulliken (electron affinity)[104], Margrave (an extended definition of electronegativity)[106], Parr and Pearson (concept of hardness)[107], Mortier (Electronegativity-equalization for the prediction of partial charges)[108], and others. In that respect, I can highly recommend the original QEq publication of Rappe and Goddard. In their publication they clearly outline the relationship between QEq and some of these earlier ideas and methods.

Above, I already defined electronegativity based on IP and EA. Now let's consider how the energy of an isolated atom changes as a function of charge, e.g. how the energy changes by adding or removing an electron. Starting with the neutral reference, one can write the energy of atom i by a Taylor expansion

$$E_i(q_i) = E_{i0} + q_i\left(\frac{\partial E}{\partial q}\right)_{i0} + \frac{1}{2}q_i^2\left(\frac{\partial^2 E}{\partial q^2}\right)_{i0} + \dots \tag{4.7}$$

When including terms through second order and defining two examples of adding/removing an electron we yield:

$$E_i(+1) = IP = E_{i0} + \left(\frac{\partial E}{\partial q}\right)_{i0} + \frac{1}{2}\left(\frac{\partial^2 E}{\partial q^2}\right)_{i0} \tag{4.8}$$

$$E_i(-1) = EA = E_{i0} - \left(\frac{\partial E}{\partial q}\right)_{i0} + \frac{1}{2}\left(\frac{\partial^2 E}{\partial q^2}\right)_{i0} \tag{4.9}$$

These two cases represent the IP (the energy needed to remove the outer valence electron) and the EA (the energy difference related to the injection of an extra electron). When solving for $\left(\frac{\partial E}{\partial q}\right)_{i0}$ and $\left(\frac{\partial^2 E}{\partial q^2}\right)_{i0}$ we can define

$$\left(\frac{\partial E}{\partial q}\right)_{i0} = \frac{E_i(+1) + E_i(-1)}{2} = \frac{IP + EA}{2} \tag{4.10}$$

$$\left(\frac{\partial^2 E}{\partial q^2}\right)_{i0} = E_i(+1) - E_i(-1) = IP - EA \tag{4.11}$$

In eq. 4.5 we already identified the first term $\left(\frac{\partial E}{\partial q}\right)_i = \chi_i^{E,0}$ being equivalent to the electronegativity. For the second term let's consider a neutral atom with a singly occupied orbital. The orbital is empty for the positive ion and double occupied for the negative ion. The difference between the IP and EA for that orbital is:

$$IP - EA = J_{ii}^0 \tag{4.12}$$

with J_{ii}^0 being the Coulomb repulsion between two electrons in the orbital. This electronic repulsion quantity is the idempotential and is known as atomic hardness. This is an approximation as the optimum shape of the orbital changes upon adding an additional electron.
Using the definitions of electronegativity and hardness we get

$$E_i(q_i) = E_{i0} + \chi_i^0 q_i + \frac{1}{2}J_{ii}^0 q_i^2 \tag{4.13}$$

In order to yield an energy expression for molecules and crystals the sum of atomic values for $\chi_i^{E,0}$ and J_{ii}^0 are extended by pairwise interactions between the atoms:

$$E_{tot}(q_1, \dots, q_N) = \sum_i^N \left(E_{i0} + \chi_i^0 q_i + \frac{1}{2} J_{ii}^0 q_i^2\right) + \sum_{i<j}^N q_i q_j J_{ii} \tag{4.14}$$

with J_{ij} being the Coulomb interaction between atom i and j. This can be rewritten as

$$E_{tot}(q_1, \dots, q_N) = \sum_i^N \left(E_{i0} + \chi_i^{E,0} q_i\right) + \frac{1}{2} \sum_{i,j}^N q_i q_j J_{ij} \tag{4.15}$$

In order to obtain the partial charges we set up a system of partial differential equations of the energy with respect to the system charges as χ_i

$$\chi_i(q_1, \dots, q_N) = \frac{\partial E}{\partial q_i} = \chi_i^{E,0} + \sum_B^N q_B J_{ij} \tag{4.16}$$

χ_i is a function of the charges on all the atoms of the system. With the constraint on the total charge

$$Q_{tot} = \sum_{i=1}^N q_i \tag{4.17}$$

the minimun energy is found if

$$\chi_1 = \chi_2 = \dots = \chi_N \tag{4.18}$$

By that we have a set of N partial differential equations for the equilibrium self-consistent charges that are solved once for a given structure.

In order to solve the QEq scheme, we require the Coulomb potential J_{ij}, defined as the potential between the charge centers on atoms i and j, to be separated by a distance R. For distances R where the charge distributions of atoms overlap, the simple Coulomb law (eq. 4.3) is not valid. For $R \longrightarrow 0$, the Coulomb interaction $J_{ij} \longrightarrow \infty$, whereas it should lead to a finite value related to J$_{ii}$ and J$_{jj}$. To ensure that $J_{ij}(r)$ is physical meaningful, a shielding is needed. A variety of shielding approaches exist, Rappe and Goddard choose the shielding to be the Coulomb integral between atomic densities. For simplicity Slater-type densities of neighboring atoms are considered instead of atomic densities from *ab initio* calculations on atoms. Hence, QEq still considers neighbouring atoms like isolated atoms that are pushed close.

The electron densities of an atom with valence orbitals *n*s, *n*p, or *n*d can be written as normalized single *nS* Slater densities of the form

$$\rho_{n\zeta}^{Slater} = N_n r^{n-1} e^{-\zeta r} \tag{4.19}$$

where Nn is the normalization constant, n the valence shell and ζ the valence orbital exponent. The valence orbital simply represents the characteristic size of each atom by

$$\zeta_i = \frac{\lambda(2n+1)}{2r_i} \tag{4.20}$$

The scaling factor λ accounts for the difference between an average atom size and the covalent radius r_i. The Coulomb integral for short distances can now be expressed by atomic densities

$$J_{ij}(R) = \iint \rho_i(r_i)\frac{1}{r_{ij}}\rho_j(r_j)dV_i dV_j \tag{4.21}$$

In later work Rappe and coworker's extended the QEq for periodic systems by using the Ewald summation. By using the Ewald summation they ensure the convergence of the Coulomb term in an infinite periodic system. Since the first implementation of the original QEq scheme, a variety of extensions have been proposed in order to improve the quality of the computed charges. Ongari *et al.* compared systematic errors of different classical approaches for gas adsorption predictions in metalorganic frameworks (MOFs). [109] These variations can in general be distinguished by these four parameters:

- Choice of the atomic parameters
- Center and the order of the Taylor expansion of the energy
- Analytic form to compute the pairwise interaction between atoms with respect to it's geometry
- Inclusion of further parameters to characterize each bond type

In the next section, I want to discuss a few selected machine learning approaches for charge prediction and charge equilibration.

4.2 Machine Learning Charge Prediction Schemes

Including long-range electrostatic interactions in ML potentials as an electrostatic baseline is a great challenge. We not only require to take interactions beyond the cutoff radius into account, but also to include physically meaningful energy terms. Huge efforts have been made by various groups in the past, ranging from NN approaches like PhysNet[31] and HIPNN[110] predicting partial charges, electrostatic multipole coefficients for organic molecules from kernel-ridge regression [111, 112], partial charge prediction by random forest regression in drug like molecules [113], and many more.[114–116]

In this section I want to focus on two ideas that - from my perspective - are the key developments in the last decade that influenced the development of our kQEq model. Namely Behler's third generation NN in 2011[33] and Goedecker's charge equilibration neural network technique (CENT) in 2015[117]. Behler and coworkers were the first group to use a baseline neural network predicting environment-dependent charges from ab initio atomic charges. By that they could predict long-range interactions but lack the ability to predict non-local charge transfer. A first important step towards non local charge-transfer in ML potentials has then been done by Goedecker and co-workers. By predicting environment dependent electronegativities and determine charges by the QEq scheme they were able to include long-range charge transfer in a qualitatively correct way. In all subsections, I focus on the main idea of how to incorporate long-range electrostatics. Details of the underling NN frameworks can be found in the corresponding publications.

4.2.1 *Third generation NN potentials*

The key component for a successful application of such an electrostatic baseline is the accurate prediction of atomic charges. While for some cases, including our first attempt to use an electrostatic baseline[73], an element-wise fitting of charges can be sufficient, the work of Behler and others have shown the possibility to predict environment-dependent atomic charges.[33, 118]

In this approach a baseline NN (accounting for the long-range electrostatic) in addition to a NN potential (covering all remaining short-range interactions within the cutoff) is trained. This idea of using an electrostatic baseline is often referred to Δ-learning.

$$E_{total} = E_{short} + E_{elec} \tag{4.22}$$

In Δ-learning a double counting of electrostatic energy contributions is avoided by a simple subtraction of the electrostatic energy E_{elec} from the total energy E_{total}. As shown in Fig. 4.1, atomic charge NNs are trained using reference atomic charges obtained from electronic structure calculations and atomic positions.

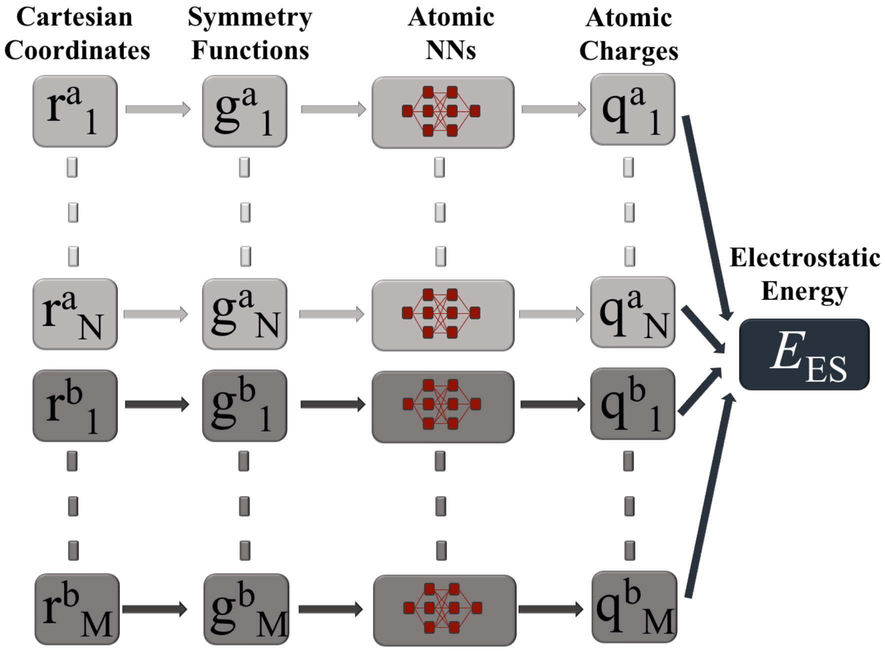

Fig. 4.1: *Schematic structure of a third generation NN potential by Behler. A set of atomic NNs (shown in red) is used to construct environment-dependent atomic charges. These predicted atomic charges can be used to calculate the long-range electrostatic energy. The total energy of the system is then given by the sum of the short-range* E_{short} *and the electrostatic energy* E_{elec}*.The figure is adapted from a review by Behler.[119]*

Using first-principle partial charges is a major drawback in that approach. Atomic partical charges are not physical observables and there is no unique/best choice. Although all of them are mathematically well-defined, benchmark studies have shown that different partitioning schemes yield very different results.[120] In addition, predicting partial charges directly from atomic environments, will not cover non-local charge transfer, e.g. charge redistribution outside the cutoff radius. As charges are predicted for each atomic environment individually without a constrain on Q_{tot}, the total charge is not conserved and molecules, which are formally charge neutral might be predicted to be charged by these type of NN.

4.2.2 The CENT approach

To account for non-local charge transfer Goedecker and co-workers define the total-energy expression similare to QEq by:

$$E_{tot}(q_1, \ldots, q_N) = \sum_{i}^{N} \left(E_i^0 + \chi_i^{CENT} q_i + \frac{1}{2} J_{ii} q_i^2 \right) + \frac{1}{2} \iint \rho_i(r_i) \frac{1}{r_{ij}} \rho_B(r_i) dV_i dV_j \tag{4.23}$$

with E_i^0 being a reference energy, q_i the atomic charges, J_{ii} the atomic hardness, and χ_i^{CENT} the environment-dependent atomic electronegativity of atom i. The charge density ρ in CENT is a superposition of normalized spherically symmetric Gaussian functions of width α_i centered at atomic positions r_i given by

$$\rho_i(\mathrm{r}) = \frac{q_i}{\alpha_i^3 \pi^{\frac{3}{2}}} exp\left(- \frac{|\mathrm{r} - \mathrm{r}_i|^2}{\alpha_i^2} \right) \tag{4.24}$$

The QEq approach as outlined above only requires three parameters, namely the electronegativity (χ_i^E), the non-classical contribution to the hardness (J_{ii}^0) and the atomic size (r_i) for each species in the system. As a flipside of this elegant simplicity, the accuracy and transferability of the QEq method is limited, however.

In the CENT approach this limitation is hurdled by allowing the electronegativity χ^{CENT} of an atom to change as a function of its chemical environment.[117] As shown in Fig. 4.2, for an ionic system, the cartesian coordinates (input) are transformed to atom-centered symmetry function vectors (descriptors for the NN). These inputs for the atomic NNs yield the environment-dependent electronegativities χ^{CENT}. Using the charge equilibarion framework similar to QEq, atomic charges q can be used to compute the total energy E_{tot} using eq. 4.23.

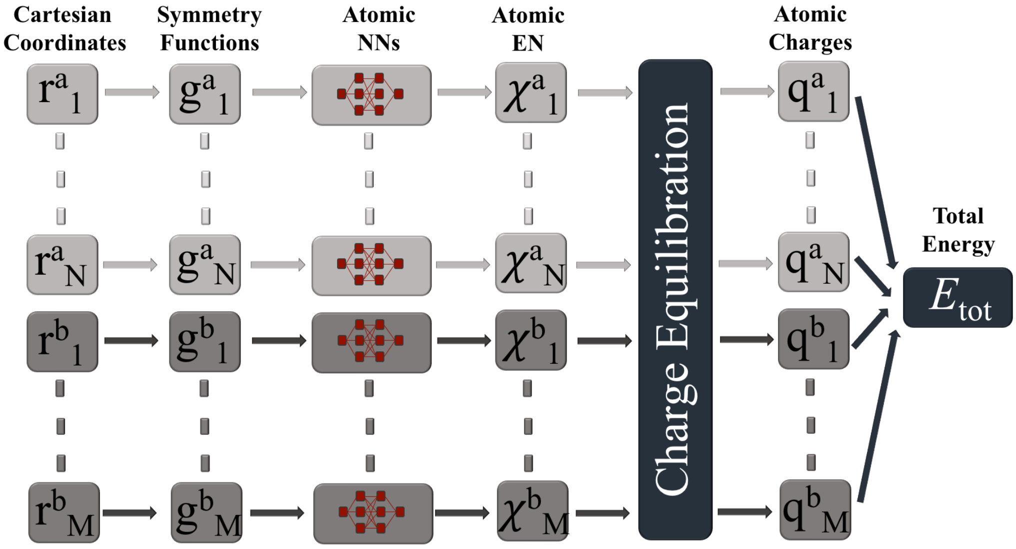
Cartesian Coordinates
Symmetry Functions
Atomic NNs
Atomic EN
Atomic Charges
Total Energy
r^a_1
g^a_1
χ^a_1
q^a_1
r^a_N
g^a_N
χ^a_N
q^a_N
r^b_1
g^b_1
χ^b_1
q^b_1
r^b_M
g^b_M
χ^b_M
q^b_M
Charge Equilibration
E_{tot}

Fig. 4.2: *Schematic structure of the charge equilibration neural network technique (CENT) for a system with two elements (a,b) with N atoms of element a and M elements for element b. Cartesian coordinates are transformed to atom-centered symmetry function vectors. These basis functions are the input for atomic NNs yielding environment-dependent electronegativities χ. Using a charge equilibration framework similar to QEq, atomic charges q can be used to compute the total energy E_{tot}. The figure is adapted from a review by Behler.[119]*

Goedecker, and co-workers applied the NN-based QEq model to ionic crystals.[117] In these ionic systems, the total energy can entirely be represented by the electrostatic contributions. In general there is no need for Q_{tot} to be zero such that the method is also applicable to charged systems. Notably in the CENT approach, Q_{tot} is conserved. For details on the setup on the NN infrastructure see [117].

5 Publications

As this thesis is publication based, in this chapter a summary of my two publications that resulted from my research during my PhD period is given. Each overview includes a brief summary of the genesis and its content and is followed by a more detailed elaboration on my personal contribution. The corresponding full articles together with the respective supporting information can be found in the appendix of this thesis.

5.1 On the role of long-range electrostatics in machine-learned interatomic potentials for complex battery materials

Carsten G. Staacke, Hendrik H. Heenen, Christoph Scheurer, Gábor Csányi, Karsten Reuter, Johannes T. Margraf
ACS Appl. Energy Mater. 2021, 4, 12562-12569
DOI: 10.1021/acsaem.1c02363

Summary: This first project aimed for the question, are long-range interactions needed when using an ML interatomic potential and if so when are they relevant? Modeling complex energy materials such as the herein investigated $Li_7P_3S_{11}$ SSE realistically at the atomistic level strains the capabilities of state-of-the-art theoretical approaches. Fortunately, modern ML potentials promise first-principles accuracy at a much reduced computational cost. However, the local nature of these ML potentials typically means that long-range contributions arising, e.g., from electrostatic interactions are neglected. Clearly, such interactions can be large in polar materials like electrolytes. In this work we investigated the effect that the locality assumption of ML potentials has on lithium mobility and defect formation energies in SSEs. We therefore developed a Δ learning protocol using a simple electrostatic baseline (ES-GAP). Comparing the classical GAP model with the newly developed ES-GAP, we found that neglecting long-range electrostatics is unproblematic for the description of lithium transport in isotropic bulk like environments. In contrast, simulating non-isotropic systems yielded the importance of ES contributions and provided new insights into interphase stability of $Li_7P_3S_{11}$.

Specifically, we studied Frenkel defects in an applied field mimicking the potential drop at a solid/solid interface. In this setup we found that a stabilization of the defects can occur already at moderate fields. This would favor the accumulation of defects towards the interphase, which could influence the kinetic stability of $Li_7P_3S_{11}$/electrode interfaces. Additionally, such stabilizations are anisotropic to crystallographic orientation making grain shape and orientation an additional parameter to be considered in battery engineering and beyond.

The findings of this work laid the foundation for the following two projects. First, now that we understood when ES contributions are relevant, we need to develop a suitable electrostatic model, that has the favourable properties of an ML approach, keeps the reactivity of ML interatomic potentials and can describe non-local charge transfer. Second, if we want to model SSE interphases realistically, we need to develop a near-universal ML potential that can likewise describe crystalline and amorphous LPS compounds.

Individual Contributions

The idea for this project was jointly conceived by Johannes Margraf, Gábor Csányi, Karsten Reuter, Christoph Scheurer and myself. Gábor Csányi introduced me to the GAP model while Johannes Margraf and Christoph Scheurer helped me to understand electrostatic modeling beyond first principles. While we have been able to identify cases of similar predictive power of the GAP and ES-GAP model, Karsten Reuter and Hendrik Heenen suggested interfacial like simulation tasks to strain the capability of a local GAP model. The manuscript was jointly written and edited by all authors.

5.2 Kernel Charge Equilibration: Efficient and Accurate Prediction of Molecular Dipole Moments with a Machine-Learning Enhanced Electron Density Model

Carsten G. Staacke, Simon Wengert, Christian Kunkel, Gábor Csányi, Karsten Reuter, Johannes T. Margraf
Mach. Learn.: Sci. Technol. 3, 2022, 015032
DOI: 10.1088/2632-2153/ac568d

Summary

In the previous project we identified the need for a charge model that can describe non-local charge transfer. Hence, in this project we developed a kernel based extension of the widely used charge equilibrition model (QEq) termed kernel Charge Equilibration (kQEq). In contrast to conventional QEq, a data-driven, environment-dependent description of atomic electronegativities is introduced. For this work we trained kQEq models on molecular dipoles and have been able to show an excellent performance, *en par* with or better than state-of-the-art kernel models, specifically tuned to predicting dipole moments.

The kQEq formalism presented in this work opens the door towards physics-based kernel ML models for predicting atomic charges, to be used in combination with reactive interatomic potentials such as the ES-GAP model. Most importantly, the presented approach was designed quite general and can be extended to other fit targets (*e.g* quadrupole moments and electrostatic potentials) and to more flexible density representations (*e.g* using atom centered dipoles in addition to partial charges).

While this work served as a proof of concept for molecular systems, we envision the extension for more complex finite systems, such as catalytic processes on nano-particles, and periodic systems, such as SSE interfaces.

Individual Contributions

Inspired by the work of Goedecker et al., Johannes Margraf suggested to replace fixed electronegativities by an atomic-environment based term and we quickly developed the first working implementation together with Simon Wengert. Thanks to Christian Kunkel, an online documentation is available. While we first aimed for partial charges as the fitting target, Gábor Csányi suggested to aim for atomic dipoles as a more realistic target. The manuscript was jointly written and edited by all authors.

5.3 Tackling structural complexity in Li_2S-P_2S_5 solid-state electrolyte using Machine Learning Potentials

Carsten G. Staacke*, Tabea Huss*, Johannes T. Margraf, Karsten Reuter, and Christoph Scheurer
* These authors contributed equally to the work.

Nanomaterials, 12, 2950, (2022)
DOI: 10.3390/nano12172950

Summary

For the final project we aimed for a near universal ML potential for the LPS material class that can likewise describe crystalline and amorphous LPS. So, how can we model amorphous LPS realistically at the atomic level and is there a way for a data-efficient description for glassy SSEs? As the lithium thiophosphate (LPS) material class provides promising candidates for solid-state electrolytes (SSEs) in lithium ion batteries due to high lithium ion conductivities, non-critical elements, and low material cost we want to be able to investigate this materials class as a whole. LPS materials are characterized by complex thiophosphate microchemistry and structural disorder influencing the material performance. To overcome the length and time scale restrictions of *ab initio* calculations in simulations of industrially applicable LPS materials, we develop a data-efficient training approach for SSEs with an emphasis on complex microchemistries. Our trained GAP model can likewise describe crystal and glassy materials and different P-S connectivities P_mS_n.

As we have been able to model this material class as a whole, we apply the GAP surrogate model to probe lithium ion conductivity and the influence of thiophosphate subunits on the latter. In our work we found that the paddle wheel effect, and hence, a constant reorientation of S in PS_4^{3-}, is happening as long as PS_4^{3-} is present, indepent of glassy or crystalline environments.

The general structure of the training protocol furthermore allows for a variety of extensions, such as dopants, other selection criteria, and including an electrostatic baseline. For future work we are currently aiming for a combination of the kQEq model with our training approach to realistically model internal SSE interfaces.

Individual Contributions

This project has been an ongoing effort of Tabea Huss and myself, that included a research internship and Tabea Huss masterthesis. While I developed the general fitting procedure, Tabea Huss refined the iterative training scheme. Christoph Scheurer and Johannes Margraf guided us when we were stuck in one of the many approaches we tried and helped us to analyse the data we produced. The manuscript was jointly written and edited by all authors.

5.4 Additional Work on Machine Learning Potentials: Machine Learning Surface Complexions of Rutile IrO_2 and RuO_2

IrO_2 Surface Complexions Identified through Machine Learning and Surface Investigations

Jakob Timmermann, Florian Kraushofer, Nikolaus Resch, Peigang Li, Yu Wang, Zhiqiang Mao, Michele Riva, Yonghyuk Lee, Carsten Staacke, Michael Schmid, Christoph Scheurer, Gareth S. Parkinson, Ulrike Diebold and Karsten Reuter
Phys. Rev. Lett. 125, 206101 (2020).
DOI: 10.1103/PhysRevLett.125.206101
Summary:
During an initial *ab initio* study of IrO_2 nanoparticles and surfaces, it quickly became obvious that relevant insights considering the catalytic surface demand a method multiple orders of magnitude faster than DFT. As classical force fields in this case are insufficient due to their lack of reactivity we turned to the Gaussian Approximation Potential (GAP) framework as a reactive, interatomic Machine Learning (ML) potential. Fruitful discussion and testing of hyperparameters with Jakob Timmermann led to a first GAP model. Already the first simulated annealing (SA) simulations based on the initial GAP revealed several unknown GAP minimum structures. In further iterative training and back then visual inspection of Jakob Timmermann revealed a variety of new, low energy surface complexions that have been revealed by surface reconstructions during the SA simulations. Together with the colleages from Prof Ulrike Diebold's group at the Technical University Vienna we have been able to confirm experimentally and theoretically that solely (101)-type surfaces confirm a (1 x 1) surface unit cell size and hence ruling out any reconstruction of higher symmetry.

Data-Efficient Iterative Training of Gaussian Approximation Potentials: Application to Surface Structure Determination of Rutile IrO_2 and RuO_2

Jakob Timmermann, Yonghyuk Lee, Carsten Staacke, Johannes T. Margraf, Christoph Scheurer and Karsten Reuter
J. Chem. Phys. 155, 244107 (2021)
DOI: 10.1063/5.0071249
Summary:
The instant success of the generation-based training protocol had an obvious flaw: the selection process via visual inspection is highly subjective, and can not be automated. Accordingly, we introduced a similarity measure as a systematic, quantifiable selection criterion into our workflow and developed an updated iterative and automated training protocol for the identification of global minimum structures of arbitrary metal oxide surfaces. I then helped with the hyperparameter selection procedure for rutile IrO_2 and RuO_2. Jakob Timmermann and Yonghyuk Lee then took this initial work and developed a systematic yet data-efficient scheme for bootstrapping the initial training set, detailed heuristics on how to test and select appropriate hyperparameters, and a straight-forward approach to determine the similarity threshold. This then fully-automatized, iterative training protocol was successfully applied to IrO_2 and RuO_2 and revealed additional low-energy complexions for both materials.

6 Summary, Conclusions and Outlook

Large scale energy storage is already indispensable in our todays life. In that respect batteries play a crucial role in modern mobility, transport and communication. While we are struggling with the complete picture of liquid electrolytes, we still barely scratch the surface when it comes to solid-state batteries. In the past, *ab initio* methods and empirical potentials have been powerful tools. Unfortunately, in the same way *ab initio* methods are limited given the size and time scales required for a detailed understanding of interfaces, while empirical potentials don't reach the required accuracy or lack flexibility. Here, emerging Machine Learning (ML) methods have shown to be increasingly able to combine the strength of both approaches, with good first-principles accuracy at a much reduced computational cost.

In this thesis, I developed a variety of new approaches and tackled these three challenges:

First, we identified simulation tasks which need an electrostatic baseline in order to achieve a correct description of the material. Using the same training data for crystalline $Li_7P_3S_{11}$, we found significant differences between a short range GAP model and the ES-GAP model when studying isotropic *vs.* non-isotropic systems. Here, for standard isotropic simulation tasks, such as determining Li diffusion barriers and ionic conductivities, both models yield similar results. In contrast, simulations on non-isotropic systems show the importance of ES contributions. More generally, our results confirm that short-ranged ML potentials can be surprisingly accurate for polar and ionic materials in the absence of non-isotropic chemical environments like interfaces or electric fields. In contrast we found important qualitative deviation between our GAP models in non-isotropic systems.

Second, we developed the kernel-based charge equilibration scheme called kQEq. The novel kQEq schemes enable the prediction of partial charges based on local environments by including the ability to predict non-local charge transfer. For a first application, kQEq models trained on molecular dipole moments display excellent performance, *en par* with or better than state-of-the-art dipole prediction schemes. The formalism of kQEq allows for physics-based kernel ML models for predicting atomic charges, to be used in combination with interatomic potentials such as GAP.

Third, we developed a near-universal GAP model for the crystalline and amorphous compounds in Li_2S-P_2S_5. We then used the GAP model to systematically investigate the effect of the local anion composition in glassy Li_2S-P_2S_5 compounds. The general structure of our training protocol allows for a variety of extensions, such additional selection criteria, doping with transition metals, and the future modeling of solid/solid interfaces.

Now in 2022, it seems we finally have all the tools to study solid-solid battery interfaces in an operando like fashion. Nevertheless, open challenges remain. First, we need to define a training procedure to train a kQEq model so that we can fit to energies, higher order moments, and periodic systems. Especially fitting to energies in a Δ-learning fashion is challenging. Second, material interfaces have shown to exhibit distinctively different stoichiometries, structure, and properties than either of the neighboring bulk phases. We therefore need a sensible way to to establish iterative training and exploration protocols that systematically extend the transferability of an electrostatic GAP model to the complex interphases that truly determine the performance of all solid-state batteries. The speed-up achieved by electrostatic GAPs as compared to direct *ab initio* calculations will then allow extensive searches and sampling that should provide a much clearer picture of the yet missing structure-performance relationships that will ultimately enable a rational design and advancement.

Danksagung

Mein erster Dank gilt meinem Doktorvater Prof. Dr. Karsten Reuter ohne den diese Arbeit nicht möglich gewesen wäre. Er war stets ein ehrlicher Berater und hat mir alle Türen geöffnet, die mich hierhergebracht haben.

Mein weiterer wissenschaftlicher Dank gilt meinem Gruppenleiter Dr. Christoph Scheurer und meinem Mentor Dr. Johannes Margraf. Beide haben wesentlich meine wissenschaftliche Ausbildung vorangetrieben und mir gezeigt wie ein Wissenschaftler in seiner Arbeitsweise, wie auch moralisch handeln sollte.

A special thanks belongs to Prof Gabor Csanyi. He supported me, as if I was one of his students, taught me that we should never stop asking and welcomed me in his group in Cambridge.

Forschung ist nur möglich, wenn man von einer Gruppe von Menschen bei den nicht wissenschaftlichen Themen unterstützt wird. Hier gilt mein Dank dem IT Support und Admin Team: Martin, Simon, Simeon, Xhristoph, Matthias, David und Steffen. Ein ganz besonderer Dank gilt Ruth und Julia, ohne die der administrative Dschungel uns alle mit Sicherheit schon verschlungen hätte.

Ich möchte einigen besonderen Kollegen bedanken: Zuerst meinem Mistery Room Mitstreitern Hanna und Frederic. Zu dritt haben wir uns nicht nur durch das Studium, sondern auch durch unsere ganze Diss durchgeboxt. Ein besonderer Dank gilt meinem Compadre Jakob u.a. dafür, das wir beide jetzt in einer der besten Bars Münchens als Pilsdoktoren bekannt und verewigt sind. Danke an die Studenten die ich über die Zeit betreuen durfte und an denen ich meine Fähigkeiten als Lehrer üben durfte. Hier gilt mein besonderer Dank Tabea, die ich zwei Jahre betreuen durfte.

Unendlicher Dank gilt meinen treuen Freunden die mich über die Jahre begleitet haben. Angefangen bei Nati und Konni meinen ältesten Schulfreunden. Danke an meine Freunde aus der Heckscherstrasse 15. Hier habe ich gelernt was es heißt einen Bund fürs Leben zu schließen. Mein Dank gilt meinen Freunden aus meinen ersten universitären Schritten in der Mathematik Fakultät. Mein vermutlich größter Dank gilt aber meinen Chemikern - Markus, Ala, Paddi, Michi, Sina, Maxi, Chrissi, Jonas und PG9.

Zuletzt möchte ich meiner Familie und Louisa danken. Danke für einfach alles.

Munich, August, 2022

C.

Bibliography

[1] A. Gore, *www.algore.com* (cit. on p. 1).

[2] International Energy Agency, *World energy outlook 2021, https://www.iea.org/reports/world-energy-outlook-2021* (cit. on p. 1).

[3] Organisation for Economic Co-operation and Development, *Oecd regional recovery platform, https://www.oecd.org/regional/recovery-platform.htm* (cit. on p. 1).

[4] International Energy Agency, *Https://www.iea.org/news/global-electric-car-sales-have-continued-their-strong-growth-in-2022-after-breaking-records-last-year* (cit. on p. 1).

[5] Senatsverwaltung für Wirtschaft, Energie und Betriebe Berlin, *Https://www.berlin.de/sen/energie/erneuerbare-energien/solargesetz-berlin/artikel.1053243.php* (cit. on p. 1).

[6] Bayerische Staatskanzlei, *Https://www.gesetze-bayern.de/content/document/bayklimag* (cit. on p. 1).

[7] J. B. Goodenough and M. H. Braga, Dalton Trans. **47**, 645 (2018) (cit. on p. 1).

[8] Z. Yang, J. Zhang, M. C. Kintner-Meyer, X. Lu, D. Choi, J. P. Lemmon, and J. Liu, Chem. Rev. **111**, 3577 (2011) (cit. on p. 1).

[9] C. Bauer, S. Burkhardt, N. P. Dasgupta, L. A.-W. Ellingsen, L. L. Gaines, H. Hao, R. Hischier, L. Hu, Y. Huang, J. Janek, et al., Nat. Sustain. **5**, 176 (2022) (cit. on p. 1).

[10] *Https://battery2030.eu/research/roadmap/* (cit. on p. 1).

[11] J. Janek and W. G. Zeier, Nat. Energy **1**, 1 (2016) (cit. on pp. 1, 3–5, 10).

[12] A. L. Robinson and J. Janek, *Solid-state batteries enter ev fray*, 2014 (cit. on pp. 1, 3).

[13] Ö. U. Kudu, T. Famprikis, B. Fleutot, M.-D. Braida, T. L. Mercier, M. S. Islam, and C. Masquelier, J. Power Sources **407**, 31 (2018) (cit. on pp. 1, 6, 7).

[14] N. Kamaya, K. Homma, Y. Yamakawa, M. Hirayama, R. Kanno, M. Yonemura, T. Kamiyama, Y. Kato, S. Hama, K. Kawamoto, et al., Nat. Mater. **10**, 682 (2011) (cit. on pp. 1, 3).

[15] H. Aono, E. Sugimoto, Y. Sadaoka, N. Imanaka, and G.-y. Adachi, J Electrochem Soc **137**, 1023 (1990) (cit. on p. 1).

[16] F. Mizuno, A. Hayashi, K. Tadanaga, and M. Tatsumisago, Solid State Ion. **177**, 2721 (2006) (cit. on p. 1).

[17] J. G. Kim, B. Son, S. Mukherjee, N. Schuppert, A. Bates, O. Kwon, M. J. Choi, H. Y. Chung, and S. Park, J. Power Sources. **282**, 299 (2015) (cit. on p. 1).

[18] A. D. Sendek, Q. Yang, E. D. Cubuk, K.-A. N. Duerloo, Y. Cui, and E. J. Reed, Energy Environ. Sci. **10**, 306 (2017) (cit. on p. 1).

[19] A. M. Abakumov, S. S. Fedotov, E. V. Antipov, and J.-M. Tarascon, Nat. Comm. **11**, 1 (2020) (cit. on p. 1).

[20] *Https://battery2030.eu/battery2030/projects/big-map/* (cit. on p. 1).

[21] S. Stegmaier, R. Schierholz, I. Povstugar, J. Barthel, S. P. Rittmeyer, S. Yu, S. Wengert, S. Rostami, K. Hans, K. Reuter, R. Eichel, and C. Scheurer, Adv. Energy Mater, 2100707 (2021) (cit. on pp. 1, 3, 4, 10).

[22] H. Türk, F.-P. Schmidt, T. Götsch, F. Girgsdies, A. Hammud, D. Ivanov, I. C. Vinke, L. de Haart, R.-A. Eichel, K. Reuter, et al., Adv. Mater. Interfaces **8**, 2100967 (2021) (cit. on p. 1).

[23] Y. Nagata, K. Nagao, M. Deguchi, A. Sakuda, A. Hayashi, H. Tsukasaki, S. Mori, and M. Tatsumisago, Chem. Mater. **30**, 6998 (2018) (cit. on p. 1).

[24] J. A. Lewis, F. J. Q. Cortes, M. G. Boebinger, J. Tippens, T. S. Marchese, N. Kondekar, X. Liu, M. Chi, and M. T. McDowell, ACS Energy Lett. **4**, 591 (2019) (cit. on p. 1).

[25] W. D. Richards, L. J. Miara, Y. Wang, J. C. Kim, and G. Ceder, Chem. Mater. **28**, 266 (2016) (cit. on p. 1).

[26] P. K. Nayak, G. Garcia-Belmonte, A. Kahn, J. Bisquert, and D. Cahen, Energy Environ. Sci. **5**, 6022 (2012) (cit. on p. 1).

[27] J. J. Choi, X. Yang, Z. M. Norman, S. J. Billinge, and J. S. Owen, Nano Lett. **14**, 127 (2014) (cit. on p. 1).

[28] H. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, et al., Electronic Structure **1**, 1 (2022) (cit. on pp. 1, 2).

[29] B. Huang and O. A. von Lilienfeld, Chem. Rev. **121**, 10001 (2021) (cit. on p. 1).

[30] M. Ceriotti, C. Clementi, and O. Anatole von Lilienfeld, Chem. Rev. **121**, 9719 (2021) (cit. on pp. 1, 2, 14).

[31] O. T. Unke and M. Meuwly, J. Chem. Theory Comput. **15**, 3678 (2019) (cit. on pp. 2, 29).

[32] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007) (cit. on pp. 2, 13, 15).

[33] N. Artrith, T. Morawietz, and J. Behler, Phy. Rev. B **83**, 153101 (2011) (cit. on pp. 2, 29, 30).

[34] N. Artrith and A. Urban, Comput. Mater. Sci. **114**, 135 (2016) (cit. on p. 2).

[35] N. Artrith, A. Urban, and G. Ceder, Phys. Rev. B **96**, 014112 (2017) (cit. on p. 2).

[36] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018) (cit. on p. 2).

[37] V. L. Deringer and G. Csányi, Phys. Rev. B **95**, 094203 (2017) (cit. on pp. 2, 14, 25).

[38] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010) (cit. on pp. 2, 13–15).

[39] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B **87**, 184115 (2013) (cit. on pp. 2, 14, 17).

[40] A. P. Bartók and G. Csányi, Int. J. Quantum Chem. **115**, 1051 (2015) (cit. on p. 2).

[41] F. Dommert, K. Wendler, R. Berger, L. Delle Site, and C. Holm, Chem. Phys. Chem. **13**, 1625 (2012) (cit. on p. 2).

[42] S. Riniker, J. Chem. Inf. Model **58**, 565 (2018) (cit. on p. 2).

[43] K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, J. Chem. Phys. **148**, 241727 (2018) (cit. on p. 2).

[44] F. Q. Nazar and C. Ortner, Arch. Ration. Mech. Anal **224**, 817 (2017) (cit. on p. 2).

[45] A. V. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016) (cit. on p. 2).

[46] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Müller, and A. Tkatchenko, J. Phys. Chem. Lett **6**, 2326 (2015) (cit. on p. 2).

[47] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, Nature **589**, 59 (2021) (cit. on p. 2).

[48] M. Gastegger, J. Behler, and P. Marquetand, Chem. Sci. **8**, 6924 (2017) (cit. on p. 2).

[49] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, Npj Comput. Mater. **6**, 1 (2020) (cit. on p. 2).

[50] O. C. Harris, S. E. Lee, C. Lees, and M. Tang, J. Phys. Energy **2**, 032002 (2020) (cit. on p. 3).

[51] J. Wang, Y.-S. He, and J. Yang, Adv. Mater. **27**, 569 (2015) (cit. on p. 3).

[52] Y. Jiang, F. Chen, Y. Gao, Y. Wang, S. Wang, Q. Gao, Z. Jiao, B. Zhao, and Z. Chen, J. Power Sources. **342**, 929 (2017) (cit. on p. 3).

[53] Y. Diao, K. Xie, S. Xiong, and X. Hong, J. Power Sources **235**, 181 (2013) (cit. on p. 3).

[54] A. Manthiram, X. Yu, and S. Wang, Nat. Rev. Mater. **2**, 1 (2017) (cit. on p. 3).

[55] A. M. Stephan and K. Nahm, Polymer **47**, 5952 (2006) (cit. on p. 3).

[56] Y. Kato, S. Hori, T. Saito, K. Suzuki, M. Hirayama, A. Mitsui, M. Yonemura, H. Iba, and R. Kanno, Nat. Energy **1**, 1 (2016) (cit. on p. 3).

[57] T. Minami, A. Hayashi, and M. Tatsumisago, Solid State Ion. **177**, 2715 (2006) (cit. on p. 3).

[58] C. Dietrich, D. A. Weber, S. J. Sedlmaier, S. Indris, S. P. Culver, D. Walter, J. Janek, and W. G. Zeier, J. Mater. Chem. A **5**, 18111 (2017) (cit. on pp. 3, 6, 9, 10).

[59] Y. Zhu, X. He, and Y. Mo, ACS Appl. Mater. Interfaces **7**, 23685 (2015) (cit. on p. 3).

[60] K. J. Kim, M. Balaish, M. Wadaguchi, L. Kong, and J. L. Rupp, Adv. Energy Mater. **11**, 2002689 (2021) (cit. on p. 3).

[61] A. C. Luntz, J. Voss, and K. Reuter, J. Phys. Chem. Lett. **6**, 4599 (2015) (cit. on p. 3).

[62] P. Li, Y. Zhao, Y. Shen, and S.-H. Bo, J. Phys. Energy **2**, 022002 (2020) (cit. on p. 4).

[63] F. Han, A. S. Westover, J. Yue, X. Fan, F. Wang, M. Chi, D. N. Leonard, N. J. Dudney, H. Wang, and C. Wang, Nature Energy **4**, 187 (2019) (cit. on p. 4).

[64] S. Chen, D. Xie, G. Liu, J. P. Mwizerwa, Q. Zhang, Y. Zhao, X. Xu, and X. Yao, Energy Storage Mater. **14**, 58 (2018) (cit. on p. 6).

[65] A. Sakuda, A. Hayashi, and M. Tatsumisago, Sci. Rep. **3**, 1 (2013) (cit. on p. 6).

[66] J.-S. Kim, W. D. Jung, S. Choi, J.-W. Son, B.-K. Kim, J.-H. Lee, and H. Kim, J. Phys. Chem. Lett. **9**, 5592 (2018) (cit. on p. 7).

[67] N. H. H. Phuc, M. Totani, K. Morikawa, H. Muto, and A. Matsuda, Solid State Ion. **288**, 240 (2016) (cit. on pp. 7, 8).

[68] M. Murayama, N. Sonoyama, A. Yamada, and R. Kanno, Solid State Ion. **170**, 173 (2004) (cit. on p. 7).

[69] K. Homma, M. Yonemura, T. Kobayashi, M. Nagao, M. Hirayama, and R. Kanno, Solid State Ion. **182**, 53 (2011) (cit. on p. 8).

[70] S. Iikubo, K. Shimoyama, S. Kawano, M. Fujii, K. Yamamoto, M. Matsushita, T. Shinmei, Y. Higo, and H. Ohtani, AIP Advances **8**, 015008 (2018) (cit. on p. 8).

[71] Y. Seino, T. Ota, K. Takada, A. Hayashi, and M. Tatsumisago, Energy Environ. Sci. **7**, 627 (2014) (cit. on p. 8).

[72] H. Yamane, M. Shibata, Y. Shimane, T. Junke, Y. Seino, S. Adams, K. Minami, A. Hayashi, and M. Tatsumisago, Solid State Ion. **178**, 1163 (2007) (cit. on pp. 8, 10).

[73] C. G. Staacke, H. H. Heenen, C. Scheurer, G. Csányi, K. Reuter, and J. T. Margraf, ACS Appl. Energy Mater. **4**, 12562 (2021) (cit. on pp. 9, 30).

[74] S. Wenzel, D. A. Weber, T. Leichtweiss, M. R. Busche, J. Sann, and J. Janek, Solid State Ion. **286**, 24 (2016) (cit. on p. 9).

[75] Y. Wang, W. D. Richards, S. P. Ong, L. J. Miara, J. C. Kim, Y. Mo, and G. Ceder, Nat. Mater. **14**, 1026 (2015) (cit. on p. 9).

[76] I.-H. Chu, H. Nguyen, S. Hy, Y.-C. Lin, Z. Wang, Z. Xu, Z. Deng, Y. S. Meng, and S. P. Ong, ACS Appl. Mater. Interfaces **8**, 7843 (2016) (cit. on p. 9).

[77] R. Mercier, J. Malugani, B. Fahys, J. Douglande, and G. Robert, J. Solid State Chem. **43**, 151 (1982) (cit. on p. 9).

[78] C. Dietrich, M. Sadowski, S. Sicolo, D. A. Weber, S. J. Sedlmaier, K. S. Weldert, S. Indris, K. Albe, J. Janek, and W. G. Zeier, Chem. Mat. **28**, 8764 (2016) (cit. on pp. 9, 10).

[79] Z. D. Hood, C. Kates, M. Kirkham, S. Adhikari, C. Liang, and N. Holzwarth, Solid State Ion. **284**, 61 (2016) (cit. on pp. 9, 10).

[80] Y. Seino, M. Nakagawa, M. Senga, H. Higuchi, K. Takada, and T. Sasaki, J. Mater. Chem. **3**, 2756 (2015) (cit. on pp. 9, 10).

[81] M. B. Preefer, J. H. Grebenkemper, C. E. Wilson, M. Everingham, J. A. Cooley, and R. Seshadri, ACS Appl. Mater. Interfaces **13**, 57567 (2021) (cit. on p. 10).

[82] K. Ohara, A. Mitsui, M. Mori, Y. Onodera, S. Shiotani, Y. Koyama, Y. Orikasa, M. Murakami, K. Shimoda, K. Mori, et al., Sci. Rep. **6**, 1 (2016) (cit. on p. 10).

[83] Y. Aoki, K. Ogawa, T. Nakagawa, Y. Hasegawa, Y. Sakiyama, T. Kojima, and M. Tabuchi, Solid State Ion. **310**, 50 (2017) (cit. on p. 10).

[84] J. G. Smith and D. J. Siegel, Nature Comm. **11**, 1 (2020) (cit. on p. 10).

[85] T. Baba and Y. Kawamura, Front. Energy Res. **4** (2016) `10.3389/fenrg.2016.00022` (cit. on p. 10).

[86] S. Shiotani, K. Ohara, H. Tsukasaki, S. Mori, and R. Kanno, Sci. Rep. **7** (2017) `10.1038/s41598-017-07086-y` (cit. on p. 10).

[87] M. Kick, C. Grosu, M. Schuderer, C. Scheurer, and H. Oberhofer, J. Phys. Chem. Lett. **11**, 2535 (2020) (cit. on p. 10).

[88] C. Grosu, S. Döpking, S. Merz, C. Panosetti, P. Jakes, S. Annies, S. Matera, J. Granwehr, and C. Scheurer, in Ecs meeting abstracts, 1 (IOP Publishing, 2020), p. 96 (cit. on p. 10).

[89] J. Timmermann, F. Kraushofer, N. Resch, P. Li, Y. Wang, Z. Mao, M. Riva, Y. Lee, C. Staacke, M. Schmid, et al., Phys. Rev. Lett. **125**, 206101 (2020) (cit. on pp. 10, 25).

[90] F. Vitalini, A. S. Mey, F. Noé, and B. G. Keller, J. Chem. Phys. **142**, 611 (2015) (cit. on p. 13).

[91] M. S. Daw and M. I. Baskes, Phys. Rev. B **29**, 6443 (1984) (cit. on p. 13).

[92] J. Tersoff, Phys. Rev. B **37**, 6991 (1988) (cit. on p. 13).

[93] R. Drautz, Phys. Rev. B **99**, 014104 (2019) (cit. on pp. 13, 15, 19).

[94] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, J. Chem. Phys. **148**, 241717 (2018) (cit. on pp. 13, 15).

[95] A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, J. Chem. Phys. **152**, 044107 (2020) (cit. on p. 13).

[96] G. Csányi, S. Winfield, J. R. Kermode, A. De Vita, A. Comisso, N. Bernstein, and M. C. Payne, IoP Comput. Phys. Newsletter, 1 (2007) (cit. on p. 14).

[97] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Chem. Rev. **121**, 9759 (2021) (cit. on pp. 14, 17, 18).

[98] S. N. Pozdnyakov, M. J. Willatt, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Phys. Rev. Lett **125**, 166001 (2020) (cit. on p. 14).

[99] B. Onat, C. Ortner, and J. R. Kermode, J. Chem. Phys. **153**, 144106 (2020) (cit. on p. 14).

[100] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, Chem. Rev. **121**, 10073 (2021) (cit. on pp. 20, 22–24).

[101] J. C. Timmermann, "Iridium oxide as catalyst in water electrolysis: identification of novel surface structures via machine learning," PhD thesis (Technische Universität München, 2022) (cit. on p. 21).

[102] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, Proc. Natl. Acad. Sci. **116**, 1110 (2019) (cit. on p. 25).

[103] L. Pauling, *The nature of the chemical bond, and the structure of molecules and crystals - an introduction to modern structural chemistry*, 3rd ed. (Cornell University Press, New York, 1960) (cit. on p. 26).

[104] R. S. Mulliken, J. Chem. Phys. **2**, 782 (1934) (cit. on p. 26).

[105] R. Sanderson, Science **114**, 670 (1951) (cit. on p. 26).

[106] R. P. Iczkowski and J. L. Margrave, J. Am. Chem. Soc. **83**, 3547 (1961) (cit. on p. 26).

[107] R. G. Parr and R. G. Pearson, J. Am. Chem. Soc. **105**, 7512 (1983) (cit. on p. 26).

[108] W. J. Mortier, S. K. Ghosh, and S. Shankar, J. Am. Chem. Soc. **108**, 4315 (1986) (cit. on p. 26).

[109] D. Ongari, P. G. Boyd, O. Kadioglu, A. K. Mace, S. Keskin, and B. Smit, J. Chem. Theory Comput. **15**, 382 (2018) (cit. on p. 29).

[110] B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, and S. Tretiak, J. Chem. Theory Comput. **14**, 4687 (2018) (cit. on p. 29).

[111] T. Bereau, D. Andrienko, and O. A. Von Lilienfeld, J. Chem. Theory Comput. **11**, 3225 (2015) (cit. on p. 29).

[112] T. Bereau, R. A. DiStasio Jr, A. Tkatchenko, and O. A. Von Lilienfeld, J. Chem. Phys. **148**, 241706 (2018) (cit. on p. 29).

[113] P. Bleiziffer, K. Schaller, and S. Riniker, J. Chem. Inf. Model. **58**, 579 (2018) (cit. on p. 29).

[114] B. K. Rai and G. A. Bakken, J. Comp. Chem. **34**, 1661 (2013) (cit. on p. 29).

[115] V. V. Korolev, A. Mitrofanov, E. I. Marchenko, N. N. Eremin, V. Tkachenko, and S. N. Kalmykov, Chem. Mater. **32**, 7822 (2020) (cit. on p. 29).

[116] R. Zubatyuk, J. S. Smith, B. T. Nebgen, S. Tretiak, and O. Isayev, Nat. Comm. **12**, 1 (2021) (cit. on p. 29).

[117] S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, Phys. Rev. B **92**, 045131 (2015) (cit. on pp. 29, 31, 32).

[118] T. Morawietz, V. Sharma, and J. Behler, J. Chem. Phys. **136**, 064103 (2012) (cit. on p. 30).

[119] J. Behler, Chem. Rev. **121**, 10037 (2021) (cit. on pp. 30, 32).

[120] A. E. Sifain, N. Lubbers, B. T. Nebgen, J. S. Smith, A. Y. Lokhov, O. Isayev, A. E. Roitberg, K. Barros, and S. Tretiak, J. Phys. Chem. Lett. **9**, 4495 (2018) (cit. on p. 30).