

Exploring chemical space using perturbation theory

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Kuang-Yu Samuel Chang

von Hsinchu, Taiwan

Basel, 2018

Originaldokument gespeichert auf dem Dokumentenserver
der Universität Basel edoc.unibas.ch



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). To view a copy of the license, visit creativecommons.org/licenses/by-nc-nd/4.0/

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. O. Anatole von Lilienfeld und Prof. Dr. Markus
Meuwly

Basel, 20. 6. 2017

Prof. Dr. Martin Spiess

“The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.”

Dirac, 1929

Abstract

We investigate the performance and limitations of perturbation theory applied to “alchemical” interpolations of nuclear charges and positions. The performance of first and second order estimates are examined for small molecules and crystalline systems within KS-DFT. Chemical accuracy can be achieved for some cases by first order estimate due to cancellations of higher order effects. In this case, the inclusion of second order correction gives worse results because the balance of error cancellation is disturbed. An empirical reference bond length is found for each bond type, which provides maximal error cancellation effects for first and second order estimates and is applicable across different chemical environments including σ - and π - bonding. We showed that if accurate density information is available, highly accurate first order estimates of potential energy surface is achievable. Results are presented for (i) covalent bonds to hydrogen in 12 molecules with 8 valence electrons (CH_4 , NH_3 , H_2O , HF , SiH_4 , PH_3 , H_2S , HCl , GeH_4 , AsH_3 , H_2Se , HBr); (ii) main-group single bonds in 9 molecules with 14 valence electrons (CH_3F , CH_3Cl , CH_3Br , SiH_3F , SiH_3Cl , SiH_3Br , GeH_3F , GeH_3Cl , GeH_3Br); (iii) main-group double bonds in 9 molecules with 12 valence electrons (CH_2O , CH_2S , CH_2Se , SiH_2O , SiH_2S , SiH_2Se , GeH_2O , GeH_2S , GeH_2Se); (iv) main-group triple bonds in 9 molecules with 10 valence electrons (HCN , HCP , HCAs , HSiN , HSiP , HSiAs , HGeN , HGeP , HGeAs). First order estimates are utilized to predict band structure of crystalline materials, including (i) III-V semiconductors AlP , AlAs , AlSb , GaP , GaAs , GaSb , InP , InAs , InSb ; (ii) IV-IV semiconductors Si , Ge , Sn , SiGe , GeSn , SnSi , where quantitative predictions is achieved with $\text{MAE} = 0.05$ eV for density changes less than 0.25 a.u./per atom. A hybrid gradient based genetic algorithm has been applied to design $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals with optimal band structure using first order alchemical estimates. Homogeneous $\text{Al}_{0.67}\text{Ga}_{0.33}\text{As}$ crystals are identified to have the largest direct band gap of 2.1 eV. Alchemical perturbation to geometry variations is also studied, which proves that it is necessary to take the response of electron-electron interaction energy into account. The behavior of alchemical estimates is investigated up to fourth order for one-electron H_2^+ within Hartree-Fock theory. A finite radius of convergence is observed due to swapped order of Hamiltonian eigenvalues.

Acknowledgements

It has been my great honor to be a student of my Ph.D. supervisor Prof. Dr. Anatole von Lilienfeld. He provided me with all the resources to freely explore this fascinating yet complex nature. I enjoyed all the freedom I imagined for independent academic research, and I was able to try out any thought. Among other great results, many ideas have been tested and failed, but his encouragements and warm support were always present. It has been his guidance and insight that steered our projects toward scientific relevance. I sincerely appreciate all the time we spent discussing, interpreting, attempting to understand many strange results I got during my study. I thank him for all the insightful discussions, valuable suggestions and comments, and most importantly, his worthful advice on every aspect.

Besides my advisor, I would like to express my gratitude to Prof. Dr. Markus Meuwly for his insightful comments and hard questions during the courses he gave and the presentations I did throughout my study in Basel. Those valuable comments and questions incited me to widen my research from various perspectives.

My sincere thanks also go to Dr. Raghunathan Ramakrishnan who provided tremendous help in the first two years of my study, when he was a post-doctoral researcher with us. As a student with mathematics/physics background, there is numerous chemistry knowledge to pickup. He patiently walked me through many programs line-by-line while explaining detailed chemical insight. The discussions on the dramatic behavior of electron density in H_2^+ is particularly helpful. He significantly deepened my understanding to the electronic response due to alchemical perturbation.

I appreciate the collaboration with Dr. Stijn Fias who introduced me to the world of linear response theory and conceptual density functional theory. He guided me through infinitely tedious Gaussian integrals and density functional evaluations. All the discussions we had were joyful and productive.

I thank all my labmates, especially my doctorate classmates Felix Faber and Diana Tahchieva for the stimulating discussions and our enjoyable time together. And I am also grateful for the opportunity and funding to join various conferences and tutorials throughout the course of my studies. I would also like to thank the friends in École Normale Supérieure during my stay there in the summer of 2016. A special thanks goes to Prof. Dr. Stéphane Mallat and Dr. Michael Eickenberg for introducing me to the world of artificial neural network. It was an unforgettable one-month stay in Paris filled with many insightful discussions.

Last but not least I would like to thank my family: my parents Chih-Han Chang and Feng-Shen Chen for their endless support and their confidence in me. Most importantly my wife, Chorpure Chang. I am deeply indebted to her for the sacrifices she made, allowing me to pursue my scientific dream without worrying about logistics at home.

During those sleepless nights of coding/writing, she always stood by me and supported me.

Contents

Abstract	iii
Acknowledgements	iv
1 Introduction	1
1.1 Rational compound design	1
1.2 Quantum chemistry	2
1.3 Alchemical derivatives in compound space	2
1.4 Overview	4
2 Theory	5
2.1 Quantum mechanics of N -electron system	5
2.1.1 Slater determinant	6
2.1.2 Hartree-Fock method	7
2.2 Density functional theory	8
2.2.1 Hohenberg-Kohn theorem	9
2.2.2 Kohn-Sham equation	9
2.2.3 PBE functional	10
2.3 Perturbation theory and alchemical derivatives	12
2.3.1 Linear alchemical path	13
2.3.2 First order energy derivative	14
2.3.3 Second order energy derivative	14
3 Overview of alchemy	15
3.1 Introduction	15
3.2 Common alchemical applications	16
3.3 Alchemy in Quantum Mechanics	18
3.3.1 Fictitious systems	18
3.3.2 Alchemical teleportation of an atom	19
3.4 Rational Compound Design	21
3.4.1 Motivation	21
3.4.2 Alchemical changes in geometry	22
3.4.3 Alchemical changes in stoichiometry	23
3.4.4 Linearizing chemical space	24
3.5 Conclusions	26

4	Alchemy on covalent bonds	27
4.1	Introduction	27
4.2	Method	28
4.2.1	Taylor expansion in CCS	28
4.2.2	Predicting changes in covalent bonds	29
4.2.3	Error measures	30
4.2.4	Computational details	31
4.3	Results and Discussions	32
4.3.1	Vertical iso-valence-electronic changes of X-H	32
4.3.1.1	Superior performance from better density	35
4.3.2	Vertical iso-valence-electronic changes involving single, double, and triple bonds	36
4.3.3	Empirical d_{opt}	40
4.4	Conclusions	40
5	Rational crystal design	42
5.1	Introduction	42
5.2	Results and discussion	43
5.2.1	Predictions and objective function	43
5.2.2	Alchemical band structure prediction of binary III-V and IV-IV semiconductors	44
5.2.3	Band structure engineering	47
5.3	Conclusion	54
5.4	Computational details	55
6	Varying geometry: Analytic derivative	57
6.1	Introduction	57
6.2	Methods	57
6.2.1	Second order derivative	58
6.2.2	Higher order derivatives	59
6.3	Alchemical stretching of H_2^+	59
6.4	Non-vertical iso-electronic changes	61
6.4.1	Computation details	65
6.5	Conclusion	65
7	Conclusions	67
A	Self-consistent field	69
B	Implementation	70
B.1	Gaussian basis set	70
B.2	Planewave basis set	71
B.2.1	Bloch wavefunction	71
B.2.2	Hamiltonian matrix element	72
B.3	Density functional evaluation	73
C	Pseudopotentials	74

Bibliography

76

To Joshua

Chapter 1

Introduction

1.1 Rational compound design

When the first battery was invented by Alessandro Volta in 1799, the physical mechanism of electrical current was not fully understood. His invention, or voltaic pile, used alternating layers of copper and zinc to create a flow of electrons when the cathode and anode are connected. Many ameliorations and redesigns have been made to improve the performance of batteries ever since. Due to the lack of understanding of the detailed role of the materials nature, those advances mainly came from trial and error. Such approach was also taken by Edison to develop the first light bulbs. “Edisonian” approach has been the principal method for materials innovations, in the realm of medicine, photovoltaics, energy storage, and many other industries.

Materials by design is the focus of much ongoing scientific research because many of the essential advancements of technologies depend on the performance of the employed materials. Modern techniques of synthesis, such as metal organic chemical vapor decomposition[1], molecular beam epitaxy[2], or additive nanomanufacturing[3–5] promise to deliver complex compounds with atomistic control. These techniques open the door to explore copious amounts of possible material candidates, all part of chemical compound space (CCS). However, the vast nature of CCS makes the quest to search for the optimal compound a daunting challenge. Due to the combinatorial scaling of CCS, Edisonian approaches are infeasible. Automatized algorithms with efficient evaluation mechanisms are necessary to systematically search through CCS.

Rational compound design (RCD) aims to alleviate this problem by predicting the optimal candidate compound in an efficient manner, before it is synthesized and experimentally evaluated. It holds many promises to deliver high-performance materials which could be the solutions to various problems the world is facing today. The goal of RCD corresponds to solving the inverse question, *i.e.* “which compounds exhibit a set of pre-defined desired properties?”, at a rate that is superior to mere screening.[6–10]

1.2 Quantum chemistry

It was a triumph of science when Schrödinger wrote down his famous equation in 1925[11]. The insight of his equation led to a fundamental understanding of the nature of chemistry. However complicated, all chemistry can be explained in principle. The “only” difficulty lies in the fact that the underlying equations are too complex to solve in general.[12] Due to many-body effects, the solution must be approximated. Depending on the required accuracy and the number of atoms and electrons in the chemical system, a versatile pool of approximations can be applied.

Among all the chemical approaches in existence, quantum chemistry, which applies quantum mechanics to chemical systems, is of particular interest since it requires no prior knowledge of the system. It exploits the descriptive power of quantum mechanics to atoms and electrons and approximates the solution to Schrödinger equation. Thanks to modern computers, it is nowadays possible to accurately calculate many of the properties, even before it is synthesized. The flexibility of quantum chemistry is necessary to RCD because it enables an automatized algorithm to explore the entire CCS, given sufficient computation resources. The theoretical breakthrough of quantum chemistry throughout the last century, provides us with a powerful tool to model atoms and electrons in molecules accurately. It is therefore safe to say that with the help of modern computer and the development of quantum mechanics, the cost of exploring CCS has been reducing dramatically.

The invention of computers, and the compelling improvements of processing power, permits the possibility of simulating the physical world with unprecedented accuracy and speed. Thanks to the rapid developments of silicon industry, the capacity of quantum chemistry prediction has been improving in step, *i.e.* exponentially. Unfortunately, the computational cost for solving Schrödinger equation is also growing enormously with respect to the system size. Due the immense amount of possible compound and atomic configurations, that naive enumeration is prohibitive. Efficient parameter-free methods that circumvent costly procedures of obtaining quantum solutions are therefore highly desirable for RCD applications. One promising approach of such a method correspond to the use of “alchemical derivatives”, which apply perturbation theory to chemical compositions.

1.3 Alchemical derivatives in compound space

Quantum mechanical alchemy with non-integer nuclear charges was introduced as a conceptual tool for density functional theory by E. B. Wilson as early as 1962.[13] More recently, conceptual density functional theory, pioneered by R. G. Parr and W. Yang,[14]

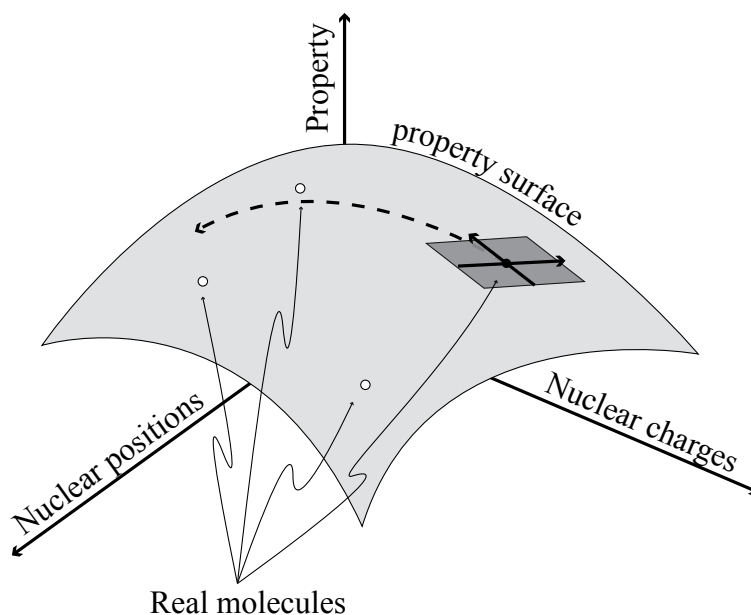


FIGURE 1.1: Cartoon illustration of a property surface parametrized by continuously defined CCS. A property is plotted as nuclear charges and nuclear positions change continuously in CCS. Real molecules with integer nuclear charges are indicated by empty circles. A reference molecule is highlighted by filled circle, where the local alchemical derivatives and the corresponding tangent plan are shown. The direction of steepest ascend is illustrated by dashed arrow.

further explores the implication of fractional electrons. These notations enable the establishment of a rigorous definition of CCS parametrized by a set of continuous variables, namely number of electrons, nuclear charges, and nuclear positions.[15] The concept of derivatives naturally arises within a continuous definition of CCS. This transforms the quest of finding optimal compounds into a mathematical optimization problem, where various gradient based optimization algorithm can be applied.

As shown schematically in Fig. 1.1, alchemical derivatives of the property of interest can be evaluated for any reference molecule. This information implies a response of the property due to infinitesimally small changes of nuclear charges and/or positions. By following the gradient to the next best molecule, a gradient base optimization algorithm can be constructed. However, not all the points in the CCS are “real”, since only molecules with integer electrons and nuclear charges are observable. In fact, real molecules form a sparse subset of all the theoretically interpolated compositions and configurations. Nonetheless, the alchemical derivatives evaluated for a given reference molecule contains important information about the local structure of the property surface in CCS.

1.4 Overview

The purpose of this thesis is to investigate the applicability of calculated alchemical derivatives. It examines the performance and limitations of utilizing alchemical derivatives in various circumstances. The thesis is organized as the following: Chapter 2 briefly summarizes the underlying theories and methods. Chapter 3 provides an overview on how to use alchemy in the context of quantum chemistry. Chapter 4 discusses first and second order derivatives, used to estimate changes in covalent binding in small molecules. Chapter 5 demonstrates the applicability of applying alchemical derivatives to crystalline systems, and discusses a RCD application for band structure engineering. Chapter 6 examines the feasibility of utilizing higher order alchemical derivatives and perturbations including changes in geometry. Chapter 7 concludes the thesis.

Chapter 2

Theory

2.1 Quantum mechanics of N -electron system

The underlying theory describing closed shell time-independent non-relativistic chemistry within the Born-Oppenheimer approximation for the N -electron and M -nucleus system is the Schrödinger equation[11]

$$H_{\{\mathbf{R}_I, Z_I\}} \Psi_{\{\mathbf{R}_I, Z_I\}}(\mathbf{r}_1, \dots, \mathbf{r}_N) = E_{\{\mathbf{R}_I, Z_I\}} \Psi_{\{\mathbf{R}_I, Z_I\}}(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (2.1)$$

where H , Ψ , and E are the Hamiltonian operator, electronic wavefunction, and eigenvalue respectively. \mathbf{r}_i denotes the coordinates and spin of the i^{th} electron and the nuclear locations and charges $\{\mathbf{R}_I, Z_I\}$ are treated as predefined parameters.

The potential energy of the system in its electronic ground-state $E_{\{\mathbf{R}_I, Z_I\}}$ is a scalar function of $\{\mathbf{R}_I, Z_I\}$, and can be defined as the potential energy hyper surface of the system, which is a function of 4^M degrees of freedom. Six out of 4^M can be subtracted for rotational and translational invariance. All the quantities are assumed to be dependent on nuclear locations and charges. The subscript $\{\mathbf{R}_I, Z_I\}$ will be omitted unless otherwise noted.

The N -electron Hamiltonian operator is

$$H = -\frac{1}{2} \sum_i^N \nabla_i^2 + \frac{1}{2} \sum_{i \neq j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i, I}^{N, M} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{I \neq J}^M \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \quad (2.2)$$

Once the wavefunction is found, the energy can be calculated by

$$\begin{aligned} E &= \langle \Psi | H | \Psi \rangle \\ &= \int d\mathbf{r}_1 \cdots d\mathbf{r}_N \Psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N) H \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \\ &= T + U + V + V_{nn}, \end{aligned} \quad (2.3)$$

where

$$\begin{aligned}
T &= -\frac{1}{2} \sum_i \int d\mathbf{r}_1 \cdots \mathbf{r}_N \Psi^*(\mathbf{r}_1, \cdots, \mathbf{r}_N) \nabla_i^2 \Psi(\mathbf{r}_1, \cdots, \mathbf{r}_N), \\
U &= \frac{1}{2} \sum_{i \neq j} \int d\mathbf{r}_1 \cdots \mathbf{r}_N \frac{\Psi^*(\mathbf{r}_1, \cdots, \mathbf{r}_N) \Psi(\mathbf{r}_1, \cdots, \mathbf{r}_N)}{|\mathbf{r}_i - \mathbf{r}_j|}, \\
V &= - \sum_{i,I} Z_I \int d\mathbf{r}_1 \cdots \mathbf{r}_N \frac{\Psi^*(\mathbf{r}_1, \cdots, \mathbf{r}_N) \Psi(\mathbf{r}_1, \cdots, \mathbf{r}_N)}{|\mathbf{r}_i - \mathbf{R}_I|}, \\
V_{nn} &= \frac{1}{2} \sum_{I \neq J}^M \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|},
\end{aligned} \tag{2.4}$$

stand for electronic kinetic energy, electron-electron repulsion energy, electron-nucleus attraction energy, and nucleus-nucleus repulsion energy respectively, the later being a constant independent of Ψ . The bracket notation used here is defined to integrate over all degree of freedom of the wavefunction. The electron density $\rho(\mathbf{r})$, or the possibility of finding an electron at location \mathbf{r} , can then be defined as^[16]

$$\rho(\mathbf{r}) = N \int d\mathbf{r}_2 \cdots \mathbf{r}_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \cdots, \mathbf{r}_N). \tag{2.5}$$

If Ψ was available, the only error in Eqs. (2.3) would come from neglect of relativistic effects and Born-Oppenheimer approximation. The total energy summed from these quantities in Eqs. (2.3) is also referred to as exact energy in the quantum chemistry community. Unfortunately, the information of $\Psi(\mathbf{r}_1, \cdots, \mathbf{r}_N)$ is not available due to $3N$ spacial degrees of freedom.

2.1.1 Slater determinant

Slater proposed to approximate Ψ by N one-particle spin orbitals, $\{\psi_i(\mathbf{r})\}$ where index i runs through each available electronic orbital, with 3 degrees of freedom each,¹ [17]

$$\Psi(\mathbf{r}_1, \cdots, \mathbf{r}_N) \approx \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{r}_1) & \cdots & \psi_1(\mathbf{r}_N) \\ \vdots & \ddots & \vdots \\ \psi_N(\mathbf{r}_1) & \cdots & \psi_N(\mathbf{r}_N) \end{vmatrix}. \tag{2.6}$$

The electron density becomes

$$\rho(\mathbf{r}) \approx \sum_i \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}). \tag{2.7}$$

¹Under this notation, the spin dependency is implicitly encoded in index i . For non-spin polarized ground state, or closed shell system, the identity $\psi_{2k} = \psi_{2k+1}$, $1 \leq k \leq N/2$ holds. This means the spin-up and spin-down electrons are indistinguishable.

Under Slater-Condon rules, the energy components in Eqs. (2.3) are given as

$$\begin{aligned}
T &= -\frac{1}{2} \sum_i \int d\mathbf{r} \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}), \\
U &= \sum_{i \neq j} \left(2 \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\psi_i^*(\mathbf{r}_1) \psi_i(\mathbf{r}_1) \psi_j^*(\mathbf{r}_2) \psi_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \right. \\
&\quad \left. - \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\psi_i^*(\mathbf{r}_1) \psi_j^*(\mathbf{r}_1) \psi_i(\mathbf{r}_2) \psi_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \right), \\
V &= -\sum_{i,I} Z_I \int d\mathbf{r} \frac{\psi_i(\mathbf{r})^2}{|\mathbf{r} - \mathbf{R}_I|}, \\
V_{nn} &= \frac{1}{2} \sum_{I \neq J}^M \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}.
\end{aligned} \tag{2.8}$$

And the energy is the sum of these terms. This approximation reduces the spacial degrees of freedom from $3N$ to 3 and producing a set of N self-consistent equations. And thus it makes the numerical approximation computationally feasible.

2.1.2 Hartree-Fock method

According to the variational principle,[18] the ground state energy can be obtained by varying orbitals until the minimum energy is achieved. The Slater determinate can be solved by setting the first variation of Eq. (2.3) to zero. With the constraints that $\{\psi_i\}$ are normalized, the variational principle yields

$$\begin{aligned}
0 &= \frac{\delta}{\delta \psi_i^*(\mathbf{r})} \left(E - \sum_i \varepsilon_i \int d\mathbf{r} \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}) \right) \\
&= \frac{-\nabla^2}{2} \psi_i(\mathbf{r}) + \psi_i(\mathbf{r}) \sum_j \int d\mathbf{r}' \frac{\psi_j^*(\mathbf{r}') \psi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\
&\quad - \sum_j \psi_j(\mathbf{r}) \int d\mathbf{r}' \frac{\psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \sum_I \frac{\psi_i(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} Z_I - \varepsilon_i \psi_i(\mathbf{r}) \\
&= H_{\text{core}}(\mathbf{r}) \psi_i(\mathbf{r}) + \sum_j \left(J_j(\mathbf{r}) - K_j(\mathbf{r}) \right) \psi_i(\mathbf{r}) - \varepsilon_i \psi_i(\mathbf{r}),
\end{aligned} \tag{2.9}$$

or

$$\left(H_{\text{core}}(\mathbf{r}) + \sum_j \left(J_j(\mathbf{r}) - K_j(\mathbf{r}) \right) \right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}), \tag{2.10}$$

where

$$\begin{aligned}
H_{\text{core}}(\mathbf{r}) &= -\frac{\nabla^2}{2} - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|}, \\
J_j(\mathbf{r}) &= \int d\mathbf{r}' \frac{\psi_j^*(\mathbf{r}')\psi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \\
K_j(\mathbf{r})\psi_i(\mathbf{r}) &= \psi_j(\mathbf{r}) \int d\mathbf{r}' \frac{\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}.
\end{aligned} \tag{2.11}$$

$\{\varepsilon_i\}$, $\{J_j(\mathbf{r})\}$, and $\{K_j(\mathbf{r})\}$ in Eq. (2.10) are Lagrange multipliers, Coulomb operator, and exchange operator respectively. Note that Coulomb operator and exchange operator depend explicitly on $\{\psi_i(\mathbf{r})\}$. This dependency implies Eq. (2.10) can only be solved iteratively until a the self-consistent solution is achieved. This is also called the self-consistent field (SCF) procedure.

The sum of Lagrange multipliers, referred to as orbital eigenvalues throughout the present thesis, is not the ground state energy due to double counting. Instead, the physical significance of ε_i is the energy required to add or remove an electron from ψ_i . The ground state energy must be computed by sum over Eqs. (2.8) within SCF. The difference of energy computed from Eq. (2.3) using Ψ and Eq. (2.10) using $\{\psi_i\}$ is referred to the correlation energy.[19]

2.2 Density functional theory

Various approaches have been proposed to approximate the correlation energy. Among them density functional theory (DFT) is one of the more successful ideas. Good accuracy can be achieved with modern generalized gradient approximation (GGA) or hybrid exchange-correlation (xc) functionals, at less or similar computational cost as Hartree-Fock. Within DFT, the energy is considered to be a functional of the electron density. Eq. (2.3) is rewritten as

$$E = E[\rho] = T[\rho] + U[\rho] + V[\rho] + V_{nn}, \tag{2.12}$$

where T is an unknown kinetic functional,

$$V[\rho] = \int d\mathbf{r} \sum_I \frac{-Z_I}{|\mathbf{r} - \mathbf{R}_I|} \rho(\mathbf{r}) \equiv \int d\mathbf{r} v(\mathbf{r}) \rho(\mathbf{r}), \tag{2.13}$$

is the external energy functional, and

$$U[\rho] = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[\rho] \equiv J[\rho] + E_{xc}[\rho], \tag{2.14}$$

is the electron-electron interaction energy functional consisting of Coulomb energy functional J and an unknown xc functional E_{xc} .

2.2.1 Hohenberg-Kohn theorem

Although DFT was proposed as early as 1927 by Thomas and Fermi,[20] an important keystone was not in place until 1964 when Hohenberg and Kohn proved that instead of $\{\psi_i\}$ also the electron density (Eq. (2.7)) can be treated as the fundamental variable. The Hohenberg-Kohn theorems state the following (for systems with non-degenerate ground state):[21]

1. $\rho(\mathbf{r})$ is as unique as $\{\psi_i\}$ uniquely determined by external potential $v(\mathbf{r})$.
2. The ground state energy can be obtained variationally, that is the density minimizes the energy is the ground state density.

The ground state energy and density obtained by solving

$$\frac{\delta E}{\delta \rho(\mathbf{r})} = 0. \quad (2.15)$$

Proof

The first part is easily shown by *reductio ad absurdum*: Assume two different external potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$ (with corresponding ground state wavefunction Ψ_1, Ψ_2 and energy E_1, E_2) would give the same ground state density $\rho(\mathbf{r})$. According to variational principle, using a trial function Ψ_2 on potential v_1 , and, a trial function Ψ_1 on potential v_2 give

$$\begin{aligned} \langle \Psi_2 | T + U + v_1 | \Psi_2 \rangle &> \langle \Psi_1 | T + U + v_1 | \Psi_1 \rangle = E_1, \\ \langle \Psi_1 | T + U + v_2 | \Psi_1 \rangle &> \langle \Psi_2 | T + U + v_2 | \Psi_2 \rangle = E_2. \end{aligned} \quad (2.16)$$

Because

$$\begin{aligned} \langle \Psi_2 | T + U + v_1 | \Psi_2 \rangle &= E_2 + \int d\mathbf{r} \rho(\mathbf{r}) (v_1(\mathbf{r}) - v_2(\mathbf{r})), \\ \langle \Psi_1 | T + U + v_2 | \Psi_1 \rangle &= E_1 + \int d\mathbf{r} \rho(\mathbf{r}) (v_2(\mathbf{r}) - v_1(\mathbf{r})). \end{aligned} \quad (2.17)$$

Insert Eqs. (2.17) into (2.16) and add up the two inequalities leads to contradiction that $E + E' > E + E'$, which implies that the assumption that $\rho(\mathbf{r})$ is not unique is wrong.

2.2.2 Kohn-Sham equation

Eq. (2.12) is exact in principle. But in reality T and E_{xc} are unknown and must be approximated. A great challenge is to approximate the kinetic energy accurately. Because the kinetic energy is in the same scale as the total energy, high accuracy of kinetic energy estimate is necessary. Kohn and Sham (KS) proposed a useful scheme in 1967 to approximate T . [22] They suggested to use a set of auxiliary normalized one-particle orbitals $\{\phi_i(\mathbf{r})\}$ such that their square sum corresponds to electron density $\rho(\mathbf{r}) = \sum_i^N \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r})$

and

$$T[\rho] \approx -\frac{1}{2} \sum_i^N \int d\mathbf{r} \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}). \quad (2.18)$$

Instead of minimizing Eq. (2.12) with respect to ρ , in KS approach it is solved with respect to $\{\phi_i\}$. Applying variational principle with normalization constraint, as in Eq. (2.9), one gets KS equations

$$\left(\frac{-\nabla^2}{2} + v(\mathbf{r}) + v_{xc}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \right) \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}), \quad (2.19)$$

where

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}}{\delta \phi_i^*(\mathbf{r})}, \quad (2.20)$$

and ε_i stand for Lagrange multipliers. Because ρ depends on ϕ , Eq. (2.19) also need to be solved iteratively. As such, all terms are known except for E_{xc} .

Exact exchange functional

Hartree-Fock exchange functional can be defined as

$$v_x^{\text{HF}}(\mathbf{r}) \phi_i(\mathbf{r}) = \left(\sum_j \frac{\delta E_x^{\text{HF}}}{\delta \phi_j^*(\mathbf{r})} \right) \phi_i(\mathbf{r}) = \sum_j \phi_j(\mathbf{r}) \int d\mathbf{r}' \frac{\phi_j^*(\mathbf{r}') \phi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}. \quad (2.21)$$

Since exchange energy is computed exactly without any approximation, Eq. (2.21) is also referred to as the exact exchange functional.

Eq. (2.10) and (2.19) are identical if v_{xc} in Eq. (2.19) is set to be v_x^{HF} . In other words, any other xc functional defines a different auxiliary Hamiltonian operator different from Eq. (2.2). Instead of adding missing pieces to Slater determinant, as all other correction schemes do, KS approach modifies the Hamiltonian operator to match up with Slater wavefunctions.

2.2.3 PBE functional

There are over 100 xc functionals proposed for various chemical environments today,[23] Perdew-Burke-Ernzerhof (PBE) functional[24] was exclusively used throughout this thesis. The PBE functional belongs to GGA family, which depends not only on density $\rho(\mathbf{r})$ but also on its gradient $\nabla\rho(\mathbf{r})$. It is an analytic functional with correct asymptotic limits. It is defined as

$$\begin{aligned} E_{xc}^{\text{PBE}}[\rho, \nabla\rho] &= E_x^{\text{PBE}}[\rho, \nabla\rho] + E_c^{\text{PBE}}[\rho, \nabla\rho] \\ &= \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_x^{\text{unif}}(r_s, \xi) F_x(s) + \int d\mathbf{r} \rho(\mathbf{r}) \left(\epsilon_c^{\text{unif}}(r_s, \xi) + H(r_s, \xi, s) \right). \end{aligned} \quad (2.22)$$

The components in Eq. (2.22) are explained below:

- r_s is the local Seitz radius defined as $r_s = \left(\frac{3}{4\pi\rho(\mathbf{r})}\right)^{\frac{1}{3}}$.
- ξ is the relative spin polarization defined as $\xi = \frac{\rho_{\uparrow}(\mathbf{r}) - \rho_{\downarrow}(\mathbf{r})}{\rho(\mathbf{r})}$, where ρ_{\uparrow} and ρ_{\downarrow} represent spin-up and spin down density respectively and $\rho_{\uparrow} + \rho_{\downarrow} = \rho$.
- s is the dimensionless density gradient defined as $s = \frac{|\nabla\rho(\mathbf{r})|}{2k_F(\mathbf{r})\rho(\mathbf{r})}$, where $k_F(\mathbf{r}) = (3\pi^2\rho(\mathbf{r}))^{\frac{1}{3}}$ is Fermi-wavevector.
- $\epsilon_x^{\text{unif}}(r_s, \xi), \epsilon_c^{\text{unif}}(r_s, \xi)$ are exchange and correlation energy density of spin polarized uniform electron gas, defined as

$$\begin{aligned}\epsilon_x^{\text{unif}}(r_s, \xi) &= -\frac{3}{4\pi r_s} \left(\frac{9\pi}{4}\right)^{\frac{1}{3}} \frac{(1+\xi)^{\frac{4}{3}} + (1-\xi)^{\frac{4}{3}}}{2}, \\ \epsilon_c^{\text{unif}}(r_s, \xi) &= -\frac{0.8851}{r_s} \left(1 - 0.5177 \left(\frac{(1+\xi)^{\frac{4}{3}} + (1-\xi)^{\frac{4}{3}}}{2}\right)\right).\end{aligned}$$

- $F_x(s)$ is the GGA exchange enhancement factor, defined as

$$F_x(s) = 1 + \kappa - \frac{\kappa}{1 + \frac{\mu s^2}{\kappa}},$$

where $\kappa \approx 0.804$, $\mu \approx 0.21951$ are derived quantities.

- $H(r_s, \xi, s)$ is the second order expansion of correlation energy, defined as

$$H(r_s, \xi, s) = \frac{e^2}{a_0} \gamma \phi^3 \ln \left(1 + \frac{\beta}{\gamma} t^2 \frac{1 + At^2}{1 + At^2 + A^2 t^4}\right),$$

where e and a_0 stand for electron charge and Bohr radius. $\beta \approx 0.066725, \gamma \approx 0.031091$ are derived quantities.

- $\phi = \frac{(1+\xi)^{2/3} + (1-\xi)^{2/3}}{2}$ is the spin scaling factor.
- Function $A = \frac{\beta}{\gamma} \left(\exp \left(\frac{-\epsilon_c^{\text{unif}}(r_s, \xi)}{\frac{e^2}{a_0} \gamma \phi^3} \right) - 1 \right)^{-1}$ in $H(r_s, \xi, s)$ is defined such that the second order expansion of correlation energy subjected to correct asymptotic limits.

Hybrid variant: PBE0

Becke pointed out in 1993 that the exact exchange functional, E_x^{HF} in Eq. (2.21), must play a role in highly accurate xc approximations.[25] Decent improvement to PBE functional is achieved by including a small fraction of exact exchange. It is called PBE0 functional,[26] and is defined as

$$E_{xc}^{\text{PBE}} \rightarrow E_{xc}^{\text{PBE0}} = \frac{1}{4} E_x^{\text{HF}} + \frac{3}{4} E_x^{\text{PBE}} + E_c^{\text{PBE}}. \quad (2.23)$$

For many circumstances, the PBE0 functional provides more accurate results. However, the four-centered integral of E_x^{HF} scales as N^4 , which makes it computationally demanding for large systems.

Range separation: HSE

Heyd, Scuseria, and Erzerhof proposed in 2003 that a screened Coulomb potential can be used to alleviate the computational demand for E_x^{HF} , where only the short range overlaps are evaluated.[27] In a recent parametrization it is defined as[28]

$$E_{xc}^{\text{HSE}} = 0.25E_x^{\text{HF,SR}}(\omega) + 0.75E_x^{\text{PBE,SR}}(\omega) + E_x^{\text{PBE,LR}}(\omega) + E_c^{\text{PBE}}, \quad (2.24)$$

where $\omega = 0.11 \text{ bohr}^{-1}$. The range separation of E_x^{HF} is done by replacing the Coulomb potential $\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{1}{r}$ kernel of Eq. 2.8 by a the screened Coulomb potential is defined as

$$\frac{1}{r} = \underbrace{\frac{1 - \text{erf}(\omega r)}{r}}_{\text{SR}} + \underbrace{\frac{\text{erf}(\omega r)}{r}}_{\text{LR}}, \quad (2.25)$$

where the superscripts SR and LR denote short range and long range while erf stands for error function. On the other hand, however, the range separation of E_x^{PBE} is more complex. It is done by scaling the exchange-correlation hole by a SR Coulomb screening factor. This yields a SR exchange enhancement factor $F_x^{\text{SR}}(s, \omega)$. The range separated E_x^{PBE} is therefore defined as²

$$\begin{aligned} E_x^{\text{PBE,SR}} &= \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_x^{\text{unif}}(r_s, \xi) F_x^{\text{SR}}(s, \omega), \\ E_x^{\text{PBE,LR}} &= E_x^{\text{PBE}} - E_x^{\text{PBE,SR}}. \end{aligned}$$

2.3 Perturbation theory and alchemical derivatives

Perturbation theory was proposed by Schrödinger in 1926 to estimate the responses of quantum mechanical systems under small perturbations. Later in 1943 Møller and Plesset applied this technique and considered electron-electron interaction as the perturbation, as an attempt to approximate the correlation energy.[29] In the present thesis, changes in external potential is considered as perturbation, or “alchemical perturbation”. Once a set of eigenvalues $\{\varepsilon_i\}$ and orbitals $\{\phi_i\}$ have been obtained for a reference system described by the Hamiltonian H_{R} , the solution of a similar target Hamiltonian H_{T} can be approximated by perturbation series, or alchemical derivatives. Throughout the thesis, alchemical perturbation is defined as an adiabatic infinitesimally small

²The detailed derivation of $F_x^{\text{SR}}(s, \omega)$ is outside the scope of the current thesis, and can be found in Ref. [27].

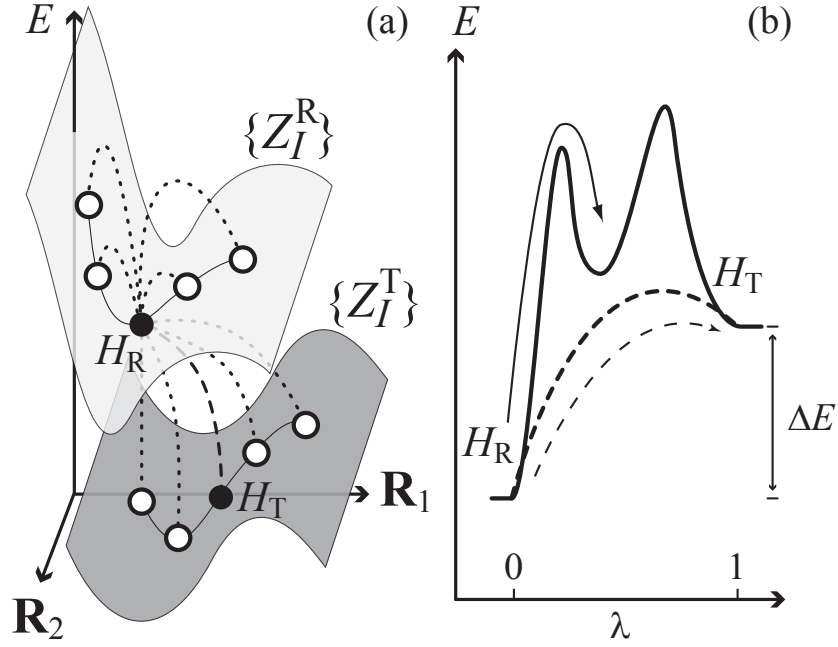


FIGURE 2.1: Cartoon illustration of alchemical paths. (a) Reference molecule is denoted by the filled circle on the upper PES, while one of the target molecules is denoted by the filled circle on the lower PES. While many targets can be connected to the reference molecule (dotted lines), a selected target is highlighted by dashed line. (b) Energy as a function of λ and the system goes from the reference molecule to the target molecule. Linear alchemical path and a realistic path is represented by dotted and solid lines respectively. The corresponding arrows indicates the reaction direction.

change along an interpolation between two distinct compounds with the same number of valence electrons.

2.3.1 Linear alchemical path

The reference molecule is defined by the Hamiltonian H_R with a collection of nuclear charges, nuclear positions, and number of electrons $\{\{Z_I^R\}, \{\mathbf{R}_I^R\}, N\}$. The target molecule is defined by H_T with $\{\{Z_I^T\}, \{\mathbf{R}_I^T\}, N\}$ as illustrated by the filled circles in Fig. 2.1(a). A linear alchemical path connecting H_R to H_T , illustrated by the dashed line in Fig. 2.1(a), is defined by the Hamiltonian

$$H_\lambda = H_R + \lambda(H_T - H_R), \quad (2.26)$$

where $0 \leq \lambda \leq 1$ is the coupling parameter.

The energy as a function of λ can be approximated by Taylor series

$$E_\lambda = E_R + \lambda \partial_\lambda E + \frac{\lambda^2}{2} \partial_\lambda^2 E + \dots \quad (2.27)$$

The derivatives with respect to λ are referred to as alchemical derivatives.

As λ goes from 0 to 1, the system gradually changes from the reference molecule to the target molecule, denoted by the dashed arrow in Fig. 2.1(b). The changes can include varying nuclear charges and positions. While a realistic path, denoted by the solid arrow in Fig. 2.1(b), incorporates such changes through complex chemical reactions, an alchemical path is merely a mathematical construct interconverting two realistic systems. As such, no reaction barriers should be expected when interpolating two molecules in their relaxed structure.

2.3.2 First order energy derivative

First order energy derivative was shown by Hellmann-Feynman theorem that[30, 31]

$$\partial_\lambda E_\lambda = \partial_\lambda \int d\mathbf{r} \Psi_\lambda^* H_\lambda \Psi_\lambda = \int d\mathbf{r} \Psi_\lambda^* (\partial_\lambda H_\lambda) \Psi_\lambda = \int d\mathbf{r} (\partial_\lambda v(\mathbf{r})) \rho_\lambda(\mathbf{r}), \quad (2.28)$$

where v is the external potential function defined in Eq. (2.13) and the derivative with respect to λ is the alchemical perturbation. Eq. (2.28) requires no more than a single integration of the electron density multiplied by the perturbation.³

2.3.3 Second order energy derivative

Second order energy derivative can be obtained by deriving Eq. (2.28) once more. Because the alchemical perturbation $\partial_\lambda v(\mathbf{r}) = H_R(\mathbf{r}) - H_T(\mathbf{r})$ does not depend on λ , $\partial_\lambda^2 E_\lambda$ becomes

$$\partial_\lambda (\partial_\lambda E_\lambda) = \int d\mathbf{r} (\partial_\lambda v(\mathbf{r})) \partial_\lambda \rho_\lambda(\mathbf{r}), \quad (2.29)$$

where the response of electron density needs to be taken into account.[32]

The electron response to a perturbation is complicated within DFT framework because T , U , and E_{xc} in Eq. (2.12) functionals implicitly depend on ρ . These effects can be computed analytically by susceptibility $\chi(\mathbf{r}, \mathbf{r}') = \frac{\delta \rho(\mathbf{r})}{\delta v(\mathbf{r}')}$, which measure how the electron density changes at location \mathbf{r} due to the perturbation at location \mathbf{r}' . This leads to

$$\partial_\lambda \rho_\lambda(\mathbf{r}) = \int d\mathbf{r}' \chi(\mathbf{r}, \mathbf{r}') \partial_\lambda v(\mathbf{r}'). \quad (2.30)$$

Detailed derivation of analytic expressions of $\chi(\mathbf{r}, \mathbf{r}')$ can be found in Ref. [33]. Finite difference is used to approximate $\partial_\lambda \rho(\mathbf{r})$ in chapter 4, while analytic expressions are used for higher order investigated in chapter 6.

³Eq. (2.28) can be proved without the notion of orbital but purely use electron density of variational principle. It is done by requiring first variation of Eq. (2.12), $\frac{\delta E}{\delta \rho_\lambda(\mathbf{r})} = 0$.

Chapter 3

Overview of alchemy

3.1 Introduction

Ever since the introduction of Hess’s law and Carnot’s cycle, chemists have known that some properties, called state functions, always change by the same amount when a system is moved reversibly from one state to another—regardless of how the change has been implemented. The freedom to choose *any* paths, even paths without any realistic correspondence except for the endpoints, is exploited within many applications. We generally refer to “alchemical” paths as paths that cannot be followed and verified through experimental observations. For example, Fig. 2.1(a) illustrates how, according to Hess’ law, the change of enthalpy of reaction can be calculated either by following the (realistic) reaction path, or, just as well, by following a more convenient yet non-realistic (alchemical) reaction path to product via dissembled elemental states as intermediates. Depending on the choice of state function, external conditions, system, and process, realistic reaction paths can be significantly more challenging because they can involve many intermediate and transition states which are difficult to identify and characterize. Even worse, they might even be experimentally impossible to probe, as it is the case for the chemistry of the earth’s core, some other planet’s bio-sphere, for distant historical or future events, or for very slow or very fast processes.

Within the atomistic theories of quantum and statistical mechanics, *any* path connecting the Hamiltonian of some reference molecule or material system, H_R , to some target system H_T , can be defined in a coupling order parameter λ as long as the end-points are met, [34–36] *i.e.*

$$H_\lambda = \begin{cases} H_R, & \lambda = 0, \\ H_T, & \lambda = 1, \end{cases} \quad (3.1)$$

where $0 \leq \lambda \leq 1$. H_λ in Eq. (3.1) denotes some intermediate state at λ , not necessarily differentiable. At boundaries of first order phase transitions, for example, the entropy

(state function) is not continuous in temperature (λ). Often, H_λ is (arbitrarily) chosen to be linear in λ , as in Eq. (2.26). As alluded to above, H_λ does not have to be realistic for all values of λ . Thermodynamics textbook examples of such changes include the calculation of the errors made when relying on the ideal gas equation. Introduced as “computational alchemy” [35, 36] in the realm of computational chemistry, this concept has successfully been used for the interpolation of forces and energies for molecular dynamics and Monte Carlo simulations. Also for the purpose of quantum mechanical observables, we can denote any such unrealistic path as “alchemical” [9, 15, 37]. We note however that Eq. (3.1) is also known as “mutation path” or “adiabatic connection” [38–40].

An even more intriguing possibility for exploiting the freedom of alchemical changes relates to the challenge of rational compound design. RCD attempts to circumvent (or at least reduce) the combinatorially scaling challenge of having to virtually enumerate and screen larger subsections of chemical or materials compound space using computationally demanding simulation methods. It has already been shown to yield promising results for the virtual atomistic control of material, nanoparticle, and even molecular structures [41, 42]. Because of the vastness of CCS, identification of novel compounds that meet desired property requirements still remains a challenge [9, 43]. Once an alchemical interpolating path, H_λ , is defined, property derivatives with respect to λ can be evaluated [31] (see Sec. 3.4). Similar to an iterative gradient descent-like algorithm, one can thus navigate gigantic combinatorial compound libraries at dramatically reduced computational costs by visiting the most promising compounds one after the other while avoiding the least promising candidates [44, 45].

The concept of connecting different systems via Eq. (3.1) has been in frequent use in various research fields, including computational engineering, physics, biophysics, and chemistry. Here, we first briefly summarize the most common application of Eq. (3.1) to calculate free energy changes, or alloy formation energies in Sec. 3.2. In Sec. 3.3 we review the quantum mechanical treatment of alchemical changes. To this end, we mainly rely on the use of DFT even though analogous arguments can be made using conventional wave-function based quantum chemistry methods. In Sec. 3.4 we present numerical results that demonstrate the use of alchemical derivatives for the screening of entire potential energy binding surfaces with semi-quantitative accuracy without additional self-consistent field calculations.

3.2 Common alchemical applications

Free energy is one of the most important state functions in chemistry. Since it is a statistical average, large numbers of configurations need to be taken into account to yield accurate predictions. [46] E.g., calculating a free energy of solvation following a

path that mimics the realistic complex process of reversible microscopic immersion of the solute into a condensed ensemble of a very large number of solvent molecules would imply a severe simulation effort that ensures that all relevant degrees of freedom have sufficiently been sampled. Furthermore, to account for hysteresis effects, this simulation should be repeated for various initial conditions and immersion rates. And one would have to start anew for any changes made to temperature, pressure, or solvent and solute species. Alternatively, one could also calculate the *change* in free energy with respect to some solute for which the free energy of solvation is already known. Thermodynamic integration, *i.e.* numerical integration of the statistical mechanical average of the “alchemical force” along the path converting known solute ($\lambda = 0$) into query solute ($\lambda = 1$), [36]

$$\Delta G = \int_0^1 d\lambda \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_\lambda. \quad (3.2)$$

Jorgensen and Ravimohan [38] proposed an even more efficient alternative: One can also estimate the change in free energy of solvation due to changing the solute using perturbation theory and MC simulation. Specifically, they considered the effect on the free energy of hydration due to an alchemical change of a methyl into hydroxy-group, $\Delta G = G_T - G_R = G_{\text{CH}_3\text{CH}_3} - G_{\text{CH}_3\text{OH}}$. One can show that if the sampling of the two states, H_R and H_T , yields sufficient overlap, the corresponding free energy difference can be accurately predicted using perturbation theory,

$$e^{-\beta \Delta G} \approx \left\langle e^{\beta(H_T - H_R)} \right\rangle_R. \quad (3.3)$$

Here, $1/\beta = k_B T$, and the right-hand-side refers to the average of the Hamiltonian difference Boltzmann’s weight over a trajectory generated using H_i . The authors used a linear interpolation of force field parameters for methanol and ethane, $H_\lambda = H_{\text{CH}_3\text{OH}} - \lambda(H_{\text{CH}_3\text{CH}_3} - H_{\text{CH}_3\text{OH}})$, from which the energy can be calculated for any λ .

As such, alchemical changes enable the prediction of changes in free energy differences without having to actually model the realistic process under investigation. Linear interpolation approaches have been applied to free energy calculations in various chemical and biological systems. [47–50] Smith and van Gunsteren found that non-linear alchemical coupling not necessarily leads to linear free energy changes. [51] Further applications of alchemical coupling to the estimation of free energy difference include the free energy of hydration of ions using *ab initio* molecular dynamics, [52] differences in free energy of binding between various host-guest complexes, [53] free energy differences at phase boundaries to predict melting points, [54, 55] the free energy of mixing to identify eutectics in ternary mixtures of molten alkali-nitrate salts, [56] kinetic isotope effects, [57] as well as constraints on the composition of the Earth’s core. [58]

But also from the solid state point of view the concept of alchemical coupling is used for the prediction of properties of disordered materials, such as co-crystals, solid solutions, or solid mixtures, as a function of mole-fraction.[59] It is computationally difficult to deal with such mixed disordered systems since the minimal self-repeating units can become very large. As a result it is nearly impossible to set up disordered systems within periodic boundary conditions. One alternative consists of using cluster-expansion methods [60], another alternative, akin to alchemical coupling, is the virtual crystal approximation (VCA)[61] which averages the system, rather than explicitly representing the full system. One of the simplest disordered class of materials are ternary semiconductors, $A_xB_{1-x}C$, where AC and BC are two different semiconductors while x is the mole-fraction between A and C. Consider, for example,[62] Eq. (3.1) applied to $Al_xGa_{1-x}As$: $H(x) = H_{GaAs} + x(H_{AlAs} - H_{GaAs})$. The linear interpolated alchemical path describes an averaged Hamiltonian between AlAs and GaAs for any mole-fraction of Al and Ga.

3.3 Alchemy in Quantum Mechanics

3.3.1 Fictitious systems

Within a first principles notion of CCS,[43] one can view every compound in any geometry as a state described by a unique Hamiltonian H . More specifically, the total potential energy's molecular Hamiltonian, H , is a function of a given set of nuclear charges, coordinates, and number of electrons, $\{\{Z_I\}, \{\mathbf{R}_I\}, N\}$, respectively. Without any loss of generality, we here rely on the Born-Oppenheimer approximation, neglecting all non-adiabatic electronic or nuclear quantum effects. Studies of alchemical paths have historically provided essential insight into the density functional theory (DFT) formulation of the many-electron problem in molecules. [16, 21] In 1974, Harris and Jones introduced an adiabatic connection,[63] coupling the system of interest to an fictitious but relevant system of non-interacting electrons,

$$H_\lambda = T + \lambda V_{ee} + V_{ext}, \quad (3.4)$$

where T , V_{ee} , and V_{ext} represent kinetic energy, electron-electron interaction energy, and external potential energy operator. By changing λ from 1 to 0, one can dial in the electron-electron interaction. For $\lambda = 0$, the electronic Schrödinger equation can thus be solved analytically, providing useful information on properties such as the exchange-correlation hole,[16, 64, 65] an important ingredient for current exchange-correlation potential development efforts.[24, 66–68] Another important study of electron-electron

interaction, carried out by Seidl, Perdew and Levy, introduces the limit of strictly correlated electrons.[69] Replacing the variable $\lambda = \frac{1}{\mu}$ for $0 < \mu \leq 1$ in Eq. (3.4), one obtains a coupled system where electron-electron interaction is dominant.

E. B. Wilson introduced the idea to alchemically couple any system to the uniform electron gas. Based on this path, he derived an expression for an exact four-dimensional density functional theory, integrating over three spatial and one λ -dimension.[13, 16] Subsequently, Politzer and Parr[70] showed that, by defining free-atom screening functions, Wilson’s functional can be decomposed into kinetic and potential energy of N electrons. These definitions of DFT related alchemical paths constitute the underlying framework for the results and discussions here within.

Within DFT,[21] we can explicitly calculate $E(\lambda)$ for *any* iso-electronic change of geometry and composition, *i.e.* under the constraint that $\int d\mathbf{r} \rho_\lambda(\mathbf{r}) = N \forall 0 \leq \lambda \leq 1$,

$$E[\rho_\lambda, \lambda] = T[n_\lambda] + V_{ee}[\rho_\lambda] + \int d\mathbf{r} \rho_\lambda(\mathbf{r}) v_{ext}(\mathbf{r}, \lambda). \quad (3.5)$$

Here, the coupling is introduced explicitly through the external potential. In practice, such coupling can be realized by scaling up or down the pseudopotentials or nuclear charges of reference and target molecules at their distinct clamped geometries. Note that kinetic and potential electron energy terms are only implicitly dependent on λ , namely through the electron density’s dependency on the λ -dependent external potential—which is imposed through application of the variational principle.

3.3.2 Alchemical teleportation of an atom

To illustrate the idea of alchemical changes within quantum chemistry, we now consider a process which is trivial when done through a realistic path, and non-trivial when done alchemically: The “teleportation” of an atom from one site to another with the constraint that the total number of electrons and protons is kept constant. Thus, instead of the trivial real space displacement of the atom, we continuously decrease the nuclear charge (annihilation) at one site while continuously increasing (creation) the nuclear charge at the other site by the same amount. For example, the external potentials of an atom at two sites can be linearly coupled through an alchemical path,

$$H_\lambda = T + V_{ee} + Z \sum_i^N \left(\frac{(1-\lambda)}{|\mathbf{r}_i|} + \frac{\lambda}{|\mathbf{r}_i - \mathbf{R}|} \right), \quad (3.6)$$

where the respective atomic sites are located at the origin and at \mathbf{R} . Considering only the endpoints ($\lambda = (0,1)$), the location of the atom obviously shifted from origin to \mathbf{R} . For any intermediate value of λ , however, the electrons will distribute among the

two competing poles of the external potential given in Eq. (3.6), forming an attractive chemical bond.

To numerically exemplify this process, we have chosen the highest occupied molecular orbital (HOMO) eigenvalue, ε , as property of interest, and an alchemical change corresponding to the linear teleportation of a $Z = 2$ and $N = 2$ system, *i.e.* effectively translating the He atom. The numerical calculation of ε for variable λ has been carried out using pseudopotential interpolation within plane-wave basis set PBE DFT calculations, in analogy to previous studies.[9, 31, 44]

In Fig. 3.1(a), the λ -dependence of ε is shown for various distances between the two atomic sites, $d = |\mathbf{R}|$. Clearly, while alchemical paths for small d yield simple parabolic shapes of ε , for teleportation involving larger interatomic distances ε develops into a double hill. ε versus d is plotted in Fig. 3.1(b) for various λ values. We note that for $\lambda = 0.5$ (magenta), the d dependency of ε corresponds to the case of stretching H_2 . ε increases monotonically at $\lambda = 0.1$ and $\lambda = 0.2$ as d increases. For these λ values, the buildup of integrated electron density at the \mathbf{R} , is still negligible, Fig. 3.1(c). Overall, the effect of nuclear potential in Eq. (3.6), $\frac{2\lambda}{|\mathbf{r}_i - \mathbf{R}|}$, amounts to a static electric field, which induces static Stark effect.[18, 71, 72] Because the electric field decreases according to Coulomb's law $\propto \frac{1}{d}$, ε rises as a result of decreasing electric field perturbation. Apart from the delocalization error of DFT,[73, 74] such nonlinear behavior could also be related to the instability of H_2^+ -like systems, which has been shown analytically.[75] Hogreve pointed out that strongly polarized electron density of asymmetric H_2^+ -like molecule severely destabilizes the system.[76] While the additional electron stabilizes the system, nonlinear behavior can be expected for ε in the case of strongly polarized density, *i.e.* for $\lambda > 0.3$ (in Fig. 3.1(c)). Fig. 3.1(c) displays integrated electron density slices, $P_\lambda(z) = \int dx dy \rho(x, y, z)$, for various λ values at interatomic distance, $d = 5\text{\AA}$. Note that for $\lambda = 0.5$, the electron density distribution corresponds to H_2 . The nonlinear dependency of electron density ρ on linearly changing growth of nuclear charge can be seen in Fig. 3.1(d) for the abrupt changes in electron density response induced by going from $\lambda \approx 0.2$ to $\lambda \approx 0.3$. To investigate the impact of parameterized exchange correlation potentials in DFT, Cohen and Mori-Sánchez calculated similar changes for $N = 1$ and $N = 2$ using the hydrogen atom plus one additional atomic site where a nuclear charge is grown, *i.e.* $Z(\lambda)$ with $Z(\lambda = 0) = 0$, $Z(\lambda = 0.5) = 1$ (H), and $Z(\lambda = 1) = 2$ (He). [77]

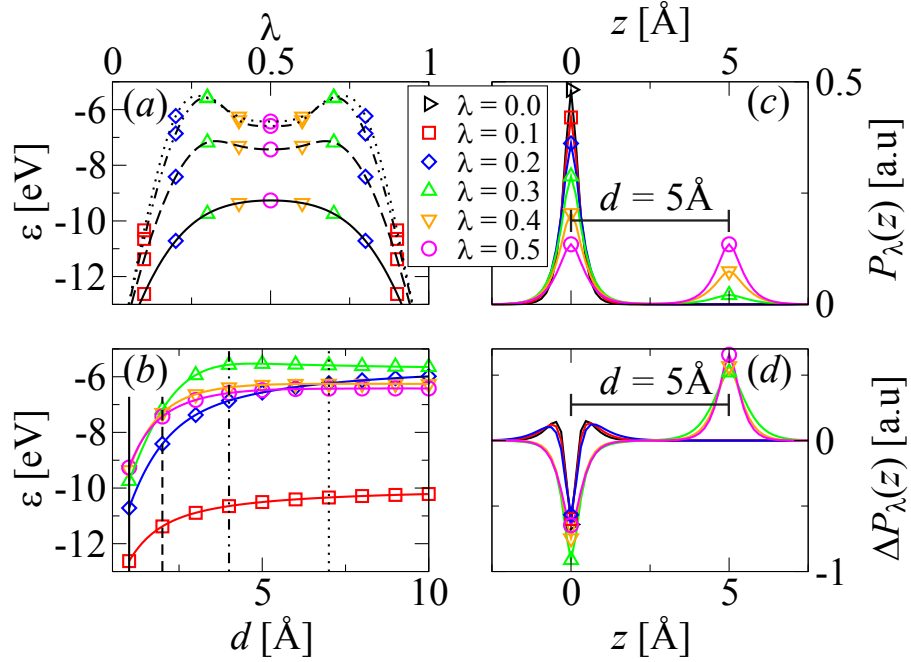


FIGURE 3.1: Alchemical transportation of a He atom. (a) ε as a function of λ for various distances $d \in \{1, 2, 4, 7\}$ Å denoted by solid, dashed, dash-dotted, dotted, respectively. (b) ε as functions of d . (c) Integrated electron density, $P_\lambda(z) = \int dx dy \rho_\lambda(x, y, z)$ for various λ at $d = 5$ Å. The electronic cusps at the nuclear sites have been highlighted by their corresponding λ symbols. (d) Integrated response of electron density due to changing λ , $\Delta P_\lambda(z) = \int dx dy \partial_\lambda \rho_\lambda(x, y, z)$ for various λ at $d = 5$ Å.

3.4 Rational Compound Design

3.4.1 Motivation

Various approaches tackle the challenges of RCD. These approaches include the inverse spectrum approach,[78] linear combination of atomic potentials,[79, 80] and many others.[8, 81–83] For the electronic potential energy, an alchemical path coupling $E(\lambda = 0)$ of one molecule to an unknown $E(\lambda = 1)$ of another compound makes explicit the compositional dependence of the energy. Understanding such a dependence holds promise to dramatically reduce the computational burden of having to stubbornly screen one compound after the other. More specifically, we can expand E in λ in terms of a Taylor series as in Eq. (2.27). In other words, if all derivatives of E_R were available one could simply follow a steepest descent procedure to screen a set of coupled “neighboring” molecules, *e.g.* with small differences in geometry or stoichiometry, to identify and proceed to more promising compound candidates. Fig. 2.1(b) illustrates the exploration of CCS following such alchemical predictions. Ideally, only a single calculation of the electronic ground-state E_R would be required (denoted by black circle). The energy of neighboring compounds can then be estimated via Eq. (3.7) (denoted by white circles).

As we discuss below, it is possible to make such scans through changes in geometry as well as composition.

In Ref. [31] we already discussed that for *any* iso-electronic alchemical change, the first order derivative is simply the Hellmann-Feynman derivative. [30] Consequently, differentiation of Eq. (3.5) yields,

$$\partial_\lambda E[\rho_\lambda, \lambda] = \langle \partial_\lambda H \rangle_\lambda = \int d\mathbf{r} \rho_\lambda(\mathbf{r}) \partial_\lambda v(\mathbf{r}, \lambda), \quad (3.7)$$

which is the same as the first order perturbation term.[18] Higher order derivatives can be evaluated or approximated by linear response theory,[84–86] and will be discussed below in the context of linearizing the energy in λ in Sec. (3.4.4).

3.4.2 Alchemical changes in geometry

We now consider alchemical changes that only involve teleportation. To demonstrate the versatility and transferability of the discussed approach, we have calculated alchemical predictions of changes in binding energy for two very different modes of binding: The covalent interatomic potential in hydrogen fluoride, as well as the hydrogen-bond-dominated van der Waals potential of the water dimer. In both cases the binding energy is given as the difference in potential energy of dimer (dim) and (relaxed) monomers m1 and m2,

$$E_b(d) = E_{\text{dim}}(d) - E_{\text{m1}} - E_{\text{m2}}. \quad (3.8)$$

Any approximate solution of the electronic Schrödinger equation at some initial distance d_i enables us to estimate the binding energy of any other d using the Hellmann-Feynman derivative and first order Taylor expansion in the alchemical teleportation path (Eqs. (3.7,2.27)),

$$E_b(d) \approx E_b^{(1)}(d) = E_b(d_i) + \partial_\lambda E_b(d_i), \quad (3.9)$$

where the super script (1) denotes the first order truncated Taylor estimate.

Considering now the case of d_i corresponding to the equilibrium distance, d_{eq} , the insets of the two top panels in Fig. 3.2 show the resulting scatter plots of $E_b^{(1)}(d)$ versus the actual $E_b(d)$ for various values of d in the case of HF and (H₂O)₂. While there is clear correlation, the scale differs dramatically for the two modes of binding. Most importantly, in the case of the dissociative tail $E^{(1)}$ correlates practically linearly with the actual binding energy. Consequently, if we now approximate the true $E_b \approx E_b^p = a_{l/r} E_b^{(1)} + b_{l/r}$, (l and r correspond to the left-hand repulsive wall and the right-hand attractive tail, respectively) one can solve for the coefficients if further constraints are known. Since this is a rather exploratory study, we here simply assume that (i) $E_b(d = d_{eq}) = E_b^{(1)}(d_{eq})$, and (ii) $E_b(d \rightarrow \infty) = 0$ in the dissociative region of the curve, and (iii) in the case of the repulsive region that $E_b(d = \frac{2}{3}d_{eq}) = 0$ for covalent binding, and

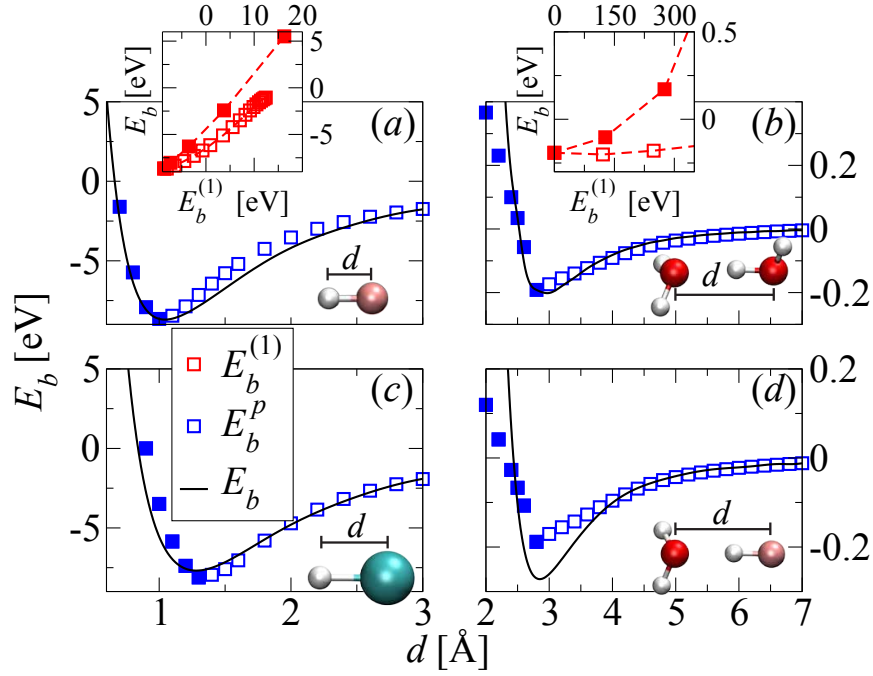


FIGURE 3.2: Actual (black lines) and alchemical (blue squares) binding energy E_b of repulsive (filled) and attractive (empty) regions of binding potentials for HF (a), $(\text{H}_2\text{O})_2$ (b), HCl (c), and H_2O -HF (d). Each screen corresponds to using only one self-consistent field (SCF) calculation at $d_i = d_{eq}$, together with the first order Taylor-expansion based model, $E_b^p = a_{l/r}E_b^{(1)} + b_{l/r}$ (Eq. (2.27)). Insets in (a) and (b) show E_b versus $E_b^{(1)}$. The screens in (c) and (d) are slightly less predictive because they are made using SCF results from HF and $(\text{H}_2\text{O})_2$, respectively. d_{eq} is set to 1, 2.8, 1.4 and 2.8 Å for (a), (b), (c), and (d) respectively.

$E_b(d = \frac{5}{6}d_{eq}) = 0$ for intermolecular binding. Assumption (iii) is based on experience using typical Morse and Lennard-Jones parameters. All resulting coefficients $\{a_{l/r}, b_{l/r}\}$ are specified in Ref. [87]. The predictions for scanning the entire binding potential agree reasonably well with the true binding potentials, and are shown together for both systems in the top panels in Fig. 3.2. Integrated deviations of these predictions are also shown in Table 3.1, yielding single digit percentage error for predicting the integral over the covalent bonding potential of hydrogen fluoride, and $\sim 14\%$ error for the integral over the van der Waals potential of the water dimer. We stress that the *entire* screen using this model only requires a single self-consistent field cycle to calculate energy and derivatives at $d = d_{eq}$.

3.4.3 Alchemical changes in stoichiometry

We now extend the use of Eq. (3.9) to also make predictions not only for teleportation changes in geometry but also for transmutational changes in stoichiometry. In particular, we have calculated predictions for changing hydrogen fluoride into hydrogen chloride at various distances, as well as changing the water dimer into the water-hydrogen fluoride

TABLE 3.1: Numerical integrals of reference energies E_b (REF) and of absolute deviation of alchemical predictions E_b^p from reference energies E_b (PRE-REF) over the binding region, *i.e.* for all d where $E_b < 0$, and percentage thereof (%) for the repulsive wall predictions as well as for the attractive tail. Columns correspond to (a) HF, (b) $(\text{H}_2\text{O})_2$, (c) HCl, and (d) H_2O -HF on display in Fig. (3.2),

Integral [$\text{eV} \times \text{\AA}$]	(a)	(b)	(c)	(d)
REF [$\text{eV} \times \text{\AA}$] (wall)	-2.542	-0.059	-3.054	-0.095
PRE-REF [$\text{eV} \times \text{\AA}$] (wall)	0.150	0.009	0.656	0.036
% (wall)	5.9	15.6	21.5	37.6
REF [$\text{eV} \times \text{\AA}$] (tail)	-8.199	-0.275	-6.017	-0.594
PRE-REF [$\text{eV} \times \text{\AA}$] (tail)	0.692	0.031	0.239	0.072
% (tail)	8.4	11.4	4.0	12.1

complex. Since we use pseudopotentials for both of these changes the total number of valence-electrons in our calculations does not change. To calculate $E_b^{(1)}$ according to Eq. (3.9) we have chosen d_{eq} to correspond to the equilibrium distance of the target system, *i.e.* HCl and H_2O -HF. Again, the same assumptions (i)-(iii) as above are used to calculate $a_{l/r}$ and $b_{l/r}$ to obtain a linear approximation of the actual $E_b(d)$ in $E_b^{(1)}$. Also for these changes, the resulting coefficients are specified in Ref. [87]. The predicted binding curves show reasonable agreement with the actual numbers, as shown for both systems in the bottom panels in Fig. 3.2. Again, integrated and relative errors are given in Table 3.1, and show a reasonable albeit slightly worse performance than in the case of predicting the water dimer or the hydrogen fluoride. We reiterate, however, that the entire screen results from only one self-consistent field cycle carried out to calculate energy and derivative of *another* molecular system—at the d_{eq} of the target system. While it is also possible to use other d to calculate energies and derivatives this typically leads to less accurate predictions. We do not think that this constitutes a problem since knowledge about equilibrium distances of target structures can easily be obtained from literature or through inexpensive force-field or semi-empirical quantum chemistry calculations which incur negligible computational overhead.

3.4.4 Linearizing chemical space

As we have seen above for the teleportation of the He atom, as well as in other studies,[31, 43] there are cases when the first order Taylor expansion of Eq. (3.9) does not provide satisfactory predictive power. This is not surprising since changes in composition correspond to large perturbations that typically lead to non-linear responses. We believe that the good performance obtained above for the binding curves is due to cancellation of higher order effects and due to the calibration of the linear model to the appropriate physical dissociation or repulsion limits. One way to systematically improve the predictive accuracy consists of including increasingly higher-order terms. Sebastiani and

coworkers[86, 88, 89] as well as Geerlings, De Proft and others[90–92] proposed promising approaches in this direction. For example, akin to our discussion above, Benoit, Sebastiani and Parrinello investigated the performance of second order linear response theory for screening the potential energy surface of the water dimer, and achieved very high predictive power. [93] How to efficiently calculate susceptibility accurately and in general, however, is still a matter of current research. Furthermore, typically one observes a (sometimes dramatic) increase in computational cost due to wave function-dependent susceptibilities, thereby defying the original motivation of RCD to navigate CCS *without* having to solve Schrödinger’s equation from scratch for each and every new geometry or molecule. As pointed out in Ref. [31], a promising alternative route towards improving the predictive power of the first order derivative consists of deviating from the assumption that the alchemical coupling must be linear in λ . In fact, as already mentioned above in the context of interpolating force-fields, [51] we are free to use *any* kind of coupling as long as we meet our endpoints, *i.e.* comply with Eq. (3.1). More specifically, if we knew the form of some coupling external potential $v_{ext}(\mathbf{r}, \lambda)$ that induces such changes in the electron density that $E(\lambda)$ becomes linear in λ , then Eq. (3.9) would result in perfect predictions. The quest for such a potential has been discussed in Ref. [43], in particular in connection to a 1-ounce-of-gold prize for anyone who provides a solution to this problem.

For a coupling path to generally fulfill the requirement that $E(\lambda)$ becomes linear in λ we note that the potential must have such a shape that the first order derivative, $\partial_\lambda E$ is a constant (as already pointed out and used in Ref. [31]), and that furthermore, all higher order energy derivatives must be zero. Consequently,

$$0 = \partial_\lambda^m E = \int d\mathbf{r} \partial_\lambda^{m-1}(\rho_\lambda(\mathbf{r})\partial_\lambda v_{ext}(\mathbf{r}, \lambda)), \quad (3.10)$$

$\forall m > 1$. This imposes certain constraints on the interpolating potential. For example, in the case of the second order derivative, equating the integrand to zero and solving for the electron density’s response results in

$$\partial_\lambda \rho(\mathbf{r}) = -\rho_\lambda(\mathbf{r}) \frac{\partial_\lambda^2 v_{ext}(\mathbf{r}, \lambda)}{\partial_\lambda v_{ext}(\mathbf{r}, \lambda)}. \quad (3.11)$$

Similar expressions can be obtained for higher order density response functions. Possibly, Eq. (3.10) could be transformed into a variational problem that yields an interpolating potential with the desired effect that the associated energies that are indeed linear in λ .

3.5 Conclusions

We discussed recent theoretical developments and approaches based on coupling states using unrealistic “alchemical” paths. Numerical evidence has been presented for the applicability and versatility of alchemical approaches applied to the inexpensive prediction of quantum mechanical observables of novel systems. The derivative based predictions certainly reflect the qualitative trend of the desired binding potentials, and are accurate within single, or low double, digit percentage accuracy. Results, discussions, and current state of the field indicate that the study of generalized coupling approaches still holds great promise for the predictive simulation of molecular and materials properties, as well as for rational compound design.

Chapter 4

Alchemy on covalent bonds

4.1 Introduction

Solving Schrödinger’s time independent equation for the unperturbed electronic ground-state within the Born-Oppenheimer approximation yields the potential energy surface (PES) of *any* molecule as a function of nuclear charges $\{Z_I\}$ (stoichiometry), nuclear positions $\{\mathbf{R}_I\}$ (geometry), and number of electrons N (molecular charge). [11, 12] The PES plays a fundamental role in chemistry and elsewhere because many properties can be derived from it. Efficient estimations of the PES for ensembles of molecules are crucial in the context of virtual compound design efforts.[47, 94, 95] These efforts typically attempt to search chemical compound space (CCS) spanned by $\{\{Z_I\}, \{\mathbf{R}_I\}, N\}$ [15, 43] for novel materials with desirable properties. As such, accurate yet efficient quantum mechanics (QM) based PES estimates hold the key for successful rational compound design applications.[6–8, 80, 96] While many inexpensive semi-empirical QM methods are available, for this study we restrict ourselves to first principles in the spirit of Refs. [9, 10, 31, 43, 44, 79, 97, 98] More specifically, we investigate the application of “alchemical” coupling to the problem of efficiently estimating the PES of new molecules using Taylor series expansions in CCS, rather than empiricism.

The alchemical coupling approach can be related to grand-canonical ensemble theory (Widom insertion) [99–101], and has been well established for empirical force-field based molecular dynamics studies. [34, 35, 38, 48, 49] Using QM, alchemical changes are less common despite E. B. Wilson’s early proposal of variable Z , back in 1962. [13] Within QM, any two iso-electronic molecules in CCS can be coupled “alchemically” through interpolation of their external potentials. Here, we have investigated if alchemical predictions can be used to model the PES of covalent bonds occurring in small closed-shell molecules made up from main group elements. We have limited ourselves to covalent hydrogen bonds, as well as single, double, and triple bonds in molecules with no more than 14 valence electrons. We present and discuss numerical evidence for the following

set of observations: First order Taylor-expansions of covalent bonding potentials can reach chemical accuracy (~ 1 kcal/mol) if two conditions are met. Firstly, the alchemical change has to be “vertical”, meaning that initial reference molecule as well as final target molecule have to possess the same number of atoms located at the exact same positions. Secondly, all elements involved in the alchemical change, *i.e.* all $\{Z_I\}$ destined to vary, have to occur late in the periodic table. Second order Taylor-expansion based predictions are less accurate than first order predictions if these conditions are met. If reference and target molecule have different geometries, the predictive power of the first order Taylor expansion substantially deteriorates, while second order estimates based on coupled perturbed Kohn-Sham equations offer some improvement, however, without reaching chemical accuracy. Second order estimates based on the independent particle approximation result in Taylor expansion estimates that are even worse than first order estimates.

In Sec. 4.2 we briefly summarize the framework of alchemical derivatives within Hartree-Fock and density functional theory (DFT) as well as our notations. Numerical estimations of covalent bond stretching of small molecules are presented and discussed in Sec. 4.3: Extending previous work on alchemical perturbation,[31, 98] we discuss alchemical energy derivatives with respect to vertical transmutation, interpolating only the identity of the atoms while keeping the geometry fixed. Estimates of single, double and triple bonds are included as an application. Finally, conclusions are drawn in Sec. 4.4.

4.2 Method

4.2.1 Taylor expansion in CCS

A Taylor expansion in CCS can be constructed with the exclusive knowledge acquired by solving Schrödinger’s equation for some reference molecule, with Hamiltonian H_R . And consequently, $\partial_\lambda^m E_\lambda = \partial_\lambda^m \langle H_\lambda \rangle$, with $\partial_\lambda H_\lambda = H_T - H_R$ being the alchemical perturbing Hamiltonian. If the alchemical derivatives can be computed, the energy of target system E_T can be estimated according to Eq. (2.27) by setting $\Delta\lambda = 1$. Note that we couple reference and target systems in a linear and global fashion. This is an arbitrary choice, non-linear and local interpolation functions could have been chosen just as well. In fact, in Ref. [31], an empirical quadratic interpolation function is found to yield superior results for first order predictions of highest occupied molecular orbital (HOMO) eigenvalues. In this study of alchemical changes of covalent bonding, we begin with linear and global interpolations, future work might deal with alternative functions.

Given a pair of isoelectronic reference/target systems, described by $\{\{Z_I^R\}, \{\mathbf{R}_I^R\}, N\}$ and $\{\{Z_I^T\}, \{\mathbf{R}_I^T\}, N\}$ respectively, one can couple the two systems such that certain Z_I^R

and Z_I^T are paired. Note that Z_I^R or Z_I^T can be scaled down to/up from zero if the number of atoms in one molecule is smaller. Under isoelectronic conditions, the λ -dependent terms in the coupling Hamiltonian (Eq. (2.26)) are the electron-nucleus and nucleus-nucleus interaction operators,

$$\begin{aligned} v_\lambda(\mathbf{r}) &= \sum_I^{N_I} \left(-\frac{(1-\lambda)Z_I^R}{|\mathbf{r} - \mathbf{R}_I^R|} - \frac{\lambda Z_I^T}{|\mathbf{r} - \mathbf{R}_I^T|} \right), \\ V_\lambda &= \sum_{I < J}^{N_I} \left(\frac{(1-\lambda)Z_I^R Z_J^R}{|\mathbf{R}_I^R - \mathbf{R}_J^R|} + \frac{\lambda Z_I^T Z_J^T}{|\mathbf{R}_I^T - \mathbf{R}_J^T|} \right). \end{aligned} \quad (4.1)$$

Since different pairing schemes result in different $v_\lambda(\mathbf{r})$ and V_λ , it is obvious that the alchemical perturbation is alignment dependent. To investigate the behaviour of higher order corrections and the effects of varying geometry/stoichiometry, we neglect all relaxation effects for vertical iso-valence-electronic changes (see Sec. 4.2.4).

4.2.2 Predicting changes in covalent bonds

For the study of covalent bonds we focus on the changes in binding potential due to alchemical coupling. We consider the difference in total potential energy between two bounded atoms at two arbitrary interatomic distances d and d_0 ,

$$\Delta E(d, d_0) = E(d) - E(d_0). \quad (4.2)$$

If, for example, d_0 is large and d is the geometry minimum, ΔE becomes the bond dissociation energy. We are interested in changes of $\Delta E(d, d_0)$ as a function of d due to alchemical changes for a fixed d_0 . More specifically, we couple a reference to target system via the corresponding Hamiltonians yielding expectation values as a function of λ ,

$$\begin{aligned} \Delta E_\lambda(d, d_0) &= E_\lambda(d) - E_\lambda(d_0) \\ &= \langle H_R(d) + \lambda(H_T(d) - H_R(d)) \rangle \\ &\quad - \langle H_R(d_0) + \lambda(H_T(d_0) - H_R(d_0)) \rangle. \end{aligned} \quad (4.3)$$

As λ goes from zero to one, the two components in Eq. (4.2) change from reference ($E_R(d), E_R(d_0)$) to target ($E_T(d), E_T(d_0)$) compound. The truncated Taylor expansion based estimate of the target compound's potential is then obtained via,

$$\begin{aligned} \Delta E_T(d, d_0) \approx \Delta E_T^{(m)}(d, d_0) &= \Delta E_R(d, d_0) \\ &\quad + \sum_{k=1}^m \frac{1}{k!} \partial_\lambda^k \Delta E_\lambda(d, d_0), \end{aligned} \quad (4.4)$$

where the superscript m stands for Taylor expansion with m terms, as a function of bond-length d for vertical alchemical changes. Since ΔE_T is the property of interest, the subscript T, λ , and the dependency of d_0 will be omitted for the rest of this work, unless otherwise noted. In this study we investigated orders up to $m = 4$ for the stretching of H_2^+ , and up to $m = 2$ for all other molecules. For a fixed d_0 , first and second order estimation are

$$\begin{aligned}\Delta E^{(1)}(d) &= (E_R(d) + \partial_\lambda E_\lambda(d)) \\ &\quad - (E_R(d_0) + \partial_\lambda E_\lambda(d_0)), \\ \Delta E^{(2)}(d) &= (E_R(d) + \partial_\lambda E_\lambda(d) + \frac{1}{2}\partial_\lambda^2 E_\lambda(d)) \\ &\quad - (E_R(d_0) + \partial_\lambda E_\lambda(d_0) + \frac{1}{2}\partial_\lambda^2 E_\lambda(d_0)).\end{aligned}\tag{4.5}$$

Since d and d_0 in Eq. (4.2) are arbitrary, one can infer the binding curve via scanning d for any fixed d_0 . The predictive power, however, happens to dependent on d_0 . For this reason, we optimize d_0 such that the integrated error in dissociation region is minimal. As shown in Fig. 4.5, an empirical linear relationship exists between equilibrium bond length of target molecule d_{eq}^T , and d_{opt}

$$d_{\text{opt}} \approx 0.76 d_{\text{eq}}^T + 0.97 \text{ \AA}.\tag{4.6}$$

d_0 is determined according to Eq. (4.6) for all vertical changes. If d_{eq}^T is not known it can easily be estimated with semi-empirical quantum chemistry methods.

4.2.3 Error measures

For bond lengths, we quantify the predictive power of the Taylor expansions by evaluating the deviation of prediction from the DFT bond length $\Delta d_{\text{eq}} = d_{\text{eq}}^{(m)} - d_{\text{eq}}$, where $d_{\text{eq}}^{(m)}$ stands for the predicted equilibrium distance of $\Delta E^{(m)}$. We calculate the deviation of the predicted energy at the minimum from the DFT energy at the DFT minimum, $\Delta E_{\text{eq}} = \Delta E^{(m)}(d_{\text{eq}}^{(m)}) - \Delta E(d_{\text{eq}})$. The deviation harmonic vibration frequency, $\Delta\omega = \omega_{\text{eq}}^{(m)} - \omega_{\text{eq}}$, of the stretching bond is also considered to quantify the predicted stiffness of the binding potential. The vibration frequency is computed via the curvature of cubic spline interpolated binding potential, $\omega_{\text{eq}} = \frac{1}{2\pi} \sqrt{\frac{k_{\text{eq}}}{\mu}}$ where $k_{\text{eq}} = \partial_d^2 \Delta E(d_{\text{eq}})$ and μ is the reduced mass. Finally, we measure the integrated error (IE) for the dissociative tail, defined as

$$\text{IE} = \frac{1}{|d_{\text{max}} - d_{\text{eq}}^{(m)}|} \int_{d_{\text{eq}}^{(m)}}^{d_{\text{max}}} dx |\Delta E^{(m)}(x) - \Delta E(x)|,\tag{4.7}$$

for vertical iso-valence-electronic changes. These four quantities provide a numerical indication of how good a prediction is. One would expect $(\Delta E_{\text{eq}}, \Delta d_{\text{eq}}, \Delta \omega_{\text{eq}}, \text{IE}) = (0, 0, 0, 0)$ for a perfect prediction. Note that we compare the predictions to DFT. This is an arbitrary choice, any other QM method could have been applied just as well.

4.2.4 Computational details

Alchemical interpolations of molecules containing elements from different rows in the periodic table can still be iso-electronic if effective core or pseudo potentials (PPs) are used, resulting in a constant number of valence electrons [43]. For example, one can couple carbon to silicon using just four valence electrons. Non-local PPs are widely used to mimic the presence of core electrons in atoms [102], and are amenable to the tuning of a wide range of properties including dispersion forces, band-gap, or vibrational frequencies [103–105]. The non-local external potential $v_\lambda(\mathbf{r})$ in Eq. (4.1) then becomes

$$v_\lambda(\mathbf{r}, \mathbf{r}') = \sum_I^{N_I} \left((1 - \lambda) v_I^{\text{R}}(\mathbf{r}, \mathbf{r}') + \lambda v_I^{\text{T}}(\mathbf{r}, \mathbf{r}') \right), \quad (4.8)$$

where v_I^{R} and v_I^{T} are PPs for Z_I^{R} and Z_I^{T} respectively. Note that $v_\lambda(\mathbf{r}, \mathbf{r}')$ in Eq. (4.8) and $v_\lambda(\mathbf{r})$ in Eq. (4.1) result in different coupling Hamiltonians, and therefore different derivatives.

All results have been obtained within the Born-Oppenheimer approximation, where nuclei are clamped, nuclear repulsion V_λ is decoupled from the electronic wavefunction, and is added as a geometry- and λ -dependent constant to the electronic energy. Nuclear-nuclear repulsion energy is computed automatically by most QM codes. However, it must be removed and recomputed independently for V_λ according to Eq. (4.1) to avoid self-repulsion between transmutating atoms. Throughout the present study, standard atomic and plane-wave basis functions, linearly interpolated PPs, as well as the PBE xc potential [24] within KS-DFT is used. The scanning of $0.5 \text{ \AA} \leq d \leq 3.0 \text{ \AA}$ is carried out with increments $\Delta d = 0.1 \text{ \AA}$. For each prediction order m , $\Delta E^{(m)}(d)$ are interpolated with cubic splines, from which the stiffness $\partial_d^2 \Delta E^{(m)}(d_{\text{eq}}^{(m)}) = k_{\text{eq}}$ is computed. All density volumetric data is printed into **Gaussian** CUBE files, from which integrated density slices are calculated.

Numerical results for vertical iso-valence-electronic alchemical changes (discussed in Sec. 4.3.1 and 4.3.2) have been obtained with **CPMD**[106], a plane wave basis with 100 Ry cutoff, and Goedecker PPs.[107–109] The periodic supercell size is $20 \times 15 \times 15 \text{ \AA}^3$, and one heavy atom is fixed at $(7.5 \text{ \AA}, 7.5 \text{ \AA}, 7.5 \text{ \AA})$ while the stretching atom shifts along $+x$ -axis. For each geometry, heavy atoms are mutated to other elements in the same column of the periodic table while all H are fixed at the same location as in the reference compound.

Since Eq. (4.8) is a non-local operator, Eqs. (2.28) and (2.29) need to be converted to wavefunction expressions. The first order derivative for the Hamiltonian $H_{R \rightarrow T}$ is evaluated using RESTART files in which the reference compound’s density and wavefunctions have been stored: $\partial_\lambda E = \langle \partial_\lambda H \rangle_R = E_T[\rho_R] - E_R[\rho_R]$. And the second order derivative is evaluated correspondingly relying on FD, $\partial_\lambda^2 E \approx \frac{\langle \partial_\lambda H \rangle_{\Delta\lambda} - \langle \partial_\lambda H \rangle_R}{\Delta\lambda}$, with $\Delta\lambda = 0.05$. Wavefunctions of reference compound are used for $\Delta E^{(1)}$, while $\Delta E_{\text{FD}}^{(2)}$ is evaluated by FD with linearly interpolated PPs parameter.

4.3 Results and Discussions

4.3.1 Vertical iso-valence-electronic changes of X-H

Using Taylor expansions binding potentials have been estimated for covalent bonds involving hydrogen (X-H) for the following 12 molecules with 8 valence electrons: (second period) CH₄, NH₃, H₂O, HF, (third period) SiH₄, PH₃, H₂S, HCl, (fourth period) GeH₄, AsH₃, H₂Se, HBr. Fig. 4.1 illustrates the first and second order truncated Taylor series estimations of the change in X-H binding energy as one goes from reference to target compound, *i.e.* $\Delta\Delta E$.

Semi-quantitative dissociative tail are obtained for all combinations of reference/-target molecules. The precise predictive power depends strongly on the choice of reference/target molecule pair. Overall, first order estimates among molecules with elements from the third or fourth row are highly accurate (Fig. 4.1). By contrast, predicting, or starting with, second row elements consistently yield worse results.

Surprisingly, inclusion of second order corrections does not necessarily lead to better performance. Second order truncated Taylor series only gives better predictions than first order when a molecule containing heavier elements is used as reference in order to predict target molecules with lighter element. Note that the performance of truncated Taylor series varies dramatically when choosing other d_0 values. Fig. 4.2 illustrates this for $\Delta E(2\text{\AA}, d_0)$ for HF \rightarrow HBr as a function of λ with $d_0 = d_{\text{opt}}$ as well as $d_0 = d_{\text{eq}}$ the equilibrium bond length of HF. While the general landscape of total energies is not changing much, small differences between $E(2\text{\AA})$, $E(d_{\text{eq}})$, and $E(d_{\text{opt}})$ give rise to great variations in ΔE as a function of λ . By choosing $d_0 = d_{\text{opt}}$, $\Delta E(d, d_{\text{opt}})$ in Eq. (4.3) becomes linear, while plotting $\Delta E(d, d_{\text{eq}})$ reveals substantial non-linearities. As a result, first order predictions of $\Delta E(d, d_{\text{opt}})$ are more predictive. This observation is very similar to the convergence behavior of the reference geometry Harris-Foulkes functional. [110, 111] Note that the Harris-Foulkes functional is often used for scanning the PES of a single compound, using the converged density at a reference geometry, while our alchemical approach offers a way to scan the PES across the CCS. The deterioration in predictive power of the second order estimate, when using molecules elements from

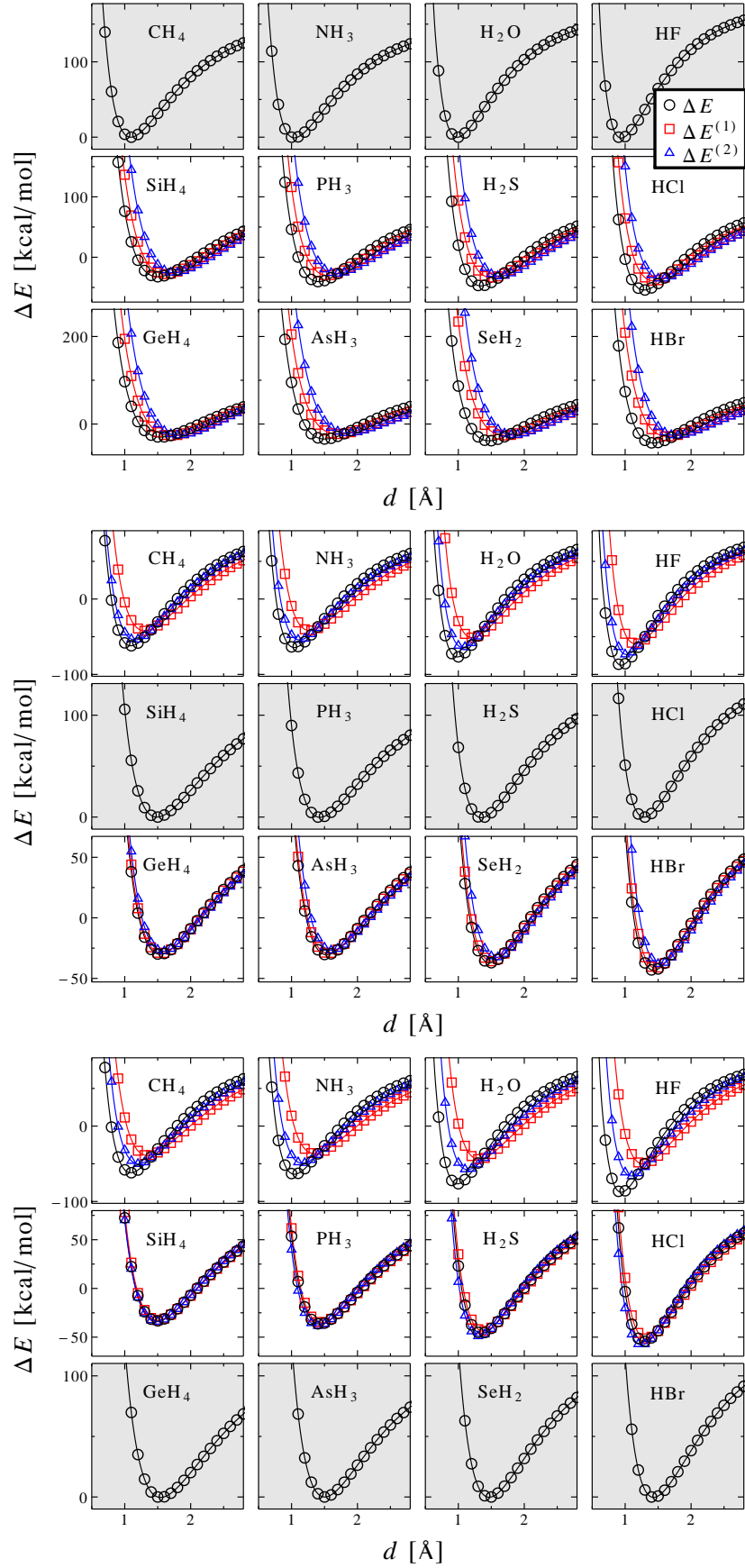


FIGURE 4.1: True (black circles), first (red squares) and second (blue triangles) order predictions of changes in the covalent bond of hydrogen due to vertical alchemical interpolations. The gray background panels correspond to the true potentials of reference compounds employed for the predictions. ΔE is shown as a function of d in Eq. (4.4).

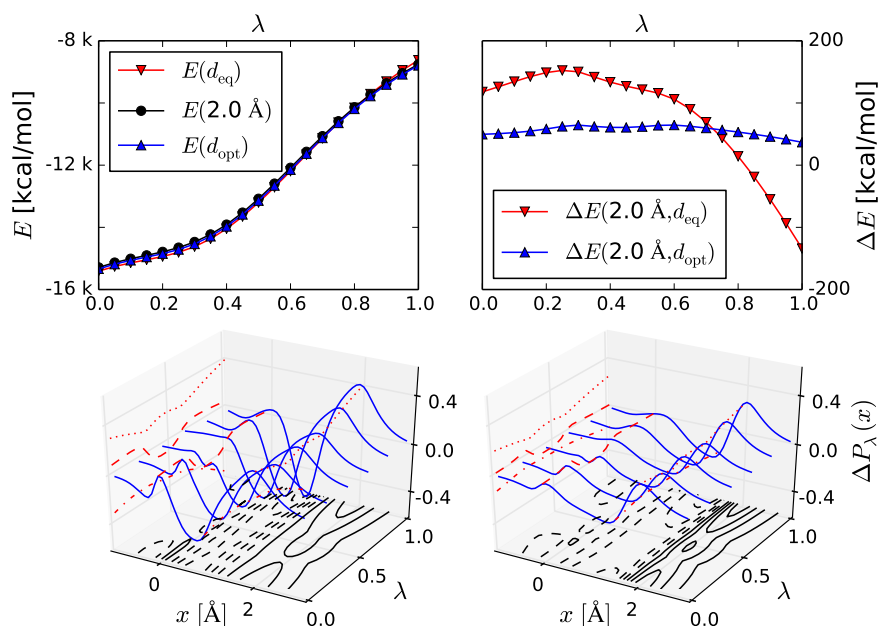


FIGURE 4.2: Alchemical coupling of HF ($\lambda = 0$) to HBr ($\lambda = 1$). TOP panels: E and ΔE where $d_{\text{eq}} = 0.94 \text{ \AA}$ and $d_{\text{opt}} = 1.57 \text{ \AA}$ denote the equilibrium bond length of reference molecule HF, and error minimizing distance, respectively. BOTTOM panels: y and z integrated valence electron density difference planes, where $\Delta P_{\lambda}(x) = \int dydz \rho_{\lambda}(\mathbf{r}, d) - \int dydz \rho_{\lambda}(\mathbf{r}, d_0)$ with $\mathbf{r} = (x, y, z)$, as a function of λ for vertical interpolation of $\Delta E(d, d_0)$ in Eq. (4.2) using $d_0 = 0.94 \text{ \AA}$ (left) or $d_0 = 1.57 \text{ \AA}$ (right), respectively. Location of heavy atom at $x = 0$ (red dashed lines), hydrogen atoms at $x = d$ (red dotted lines). Values for $x = d_{\text{eq}}$ or $x = d_{\text{opt}}$ (red dash-dotted line) are highlighted and projected onto $x = -5 \text{ \AA}$ plane.

second row, is due to the presence of an inflection point: The curvature at $\lambda = 0$ has the opposite sign. For coupling the other compound pairs in Fig. 4.1 analogous observations have been made.

Prediction errors for energy minima, equilibrium bond lengths, force constants, and integrated error in dissociation region (calculated as described in Sec. 4.2.3) have been obtained for all predictions in Fig. 4.1, and are listed in Table. 4.1. The results lend quantitative support for the observations made above: First order Taylor series based estimates are nearly chemically accurate alchemical interpolations among compounds with third and fourth row elements. Second order based predictions are *always* worse except when a molecule with heavier element is used as a reference to predict a molecule with lighter one, for example $\text{HBr} \rightarrow \text{HF}$.

Overall, the best prediction performance is found for first order based estimates using reference molecules containing third row elements ($n_{\text{R}} = 3$) in order to predict target molecules made up of fourth row elements ($n_{\text{T}} = 4$). The overall average deviation from reference bonding potential energies and integrated error are $\sim 2.5 \text{ kcal/mol}$. Corresponding predictions of equilibrium distances deviate at most by 0.03 \AA , and the

vibration frequencies deviate no more than 32 cm^{-1} . Second order estimates for the same third and fourth row combinations give slightly worse results. Not surprisingly, the worst predictions are made if the coupled molecules skip a row, *i.e.* involve elements from second and fourth row—for first as well as second order truncated Taylor expansions.

Moreover, second order overcorrections are also reported in Table. 4.1 whenever molecules containing fourth row elements are used to predict molecules containing third row elements. Both, predicted energy minimum and equilibrium bond length, show negative deviations.

We highlight the systematic asymmetry in the predictive power of first order based predictions: Consider any two iso-electronic compounds A and B, the respective alchemical derivatives will differ depending on the assignment of reference and target. In other words, the error in estimating A based on B will not be the same as the error in estimating B based on A. Results in Table. 4.1 suggest that predictions downward the columns in the periodic table are more accurate than upward. For example, predicting HBr ($n_T = 4$) using HF ($n_R = 2$) as a reference, a better estimate is obtained (error = $+25.7$ kcal/mol) than for predicting HF ($n_T = 2$) using HBr ($n_R = 4$) as a reference (error = $+61.1$ kcal/mol). Correspondingly, predicting HCl ($n_T = 3$) using HBr ($n_R = 4$) has an error = $+5.4$ kcal/mol, while the prediction of HBr ($n_T = 4$) using HCl ($n_R = 3$) has only 3.6 kcal/mol. Similar observations hold for bond lengths, and force constants. The asymmetry is also illustrated in Fig. 4.2. $\Delta E(d, d_0)$ is not necessarily symmetric with respect to $\lambda = 0.5$ for a given choice of (d, d_0) . Consequently, truncated Taylor series based predictions from either end will not be equally accurate.

Interestingly, the inflection point near $\lambda = 0$ occurs only when the reference molecule has lighter element. Conversely, no inflection point has been observed for atoms transmutating upward the column. The absence of inflection points gives a negative curvature, thus smaller error. We believe this behaviour is due to the specifics of the employed PPs. Future studies will show if similar trends hold for other PPs.

4.3.1.1 Superior performance from better density

Once the density is converged for reference molecule, inexpensive first order alchemical estimate comes with the cost of a single numerical integration as in Eq. (2.28). What if a more accurate density is used? In this section we demonstrate the applicability of Eq. (2.28) following one of the best performing alchemical coupling HBr \rightarrow HCl, using PBE0[26] as higher level of theory. Fig. 4.3 illustrates the first order predicted binding potential of HCl using converged HBr PBE0 density. It demonstrate similar behaviour as HBr \rightarrow HCl in Fig. 4.1. More importantly, when compared to PBE HCl binding potential, first alchemical prediction provides more accurate estimates to PBE0 HCl

TABLE 4.1: Errors for first (left tables) and second (right tables) order predictions of vertical iso-valence-electronic alchemical changes of covalent bond potentials in X-H \rightarrow A-H. The compound pairs are arranged in the same way as in Fig. 4.1. Error measures of each panel are collected in corresponding cell with unit [kcal/mol] for ΔE_{eq} and IE, [\AA] for Δd_{eq} , and $\Delta \omega_{\text{eq}}$ with [cm^{-1}]. Avg corresponds to averaged signed error for each target period with corresponding unit. Reference and target periods are denoted by the primary quantum number $n_{\text{R}}, n_{\text{T}} = \{2, 3, 4\}$. IV, V, VI, VII represent the column of X and A in the periodic table.

$\Delta E^{(1)}, n_{\text{R}} = 2$						$\Delta E^{(2)}, n_{\text{R}} = 2$							
n_{T}		IV.	V.	VI.	VII.	Avg	n_{T}		IV.	V.	VI.	VII.	Avg
$n_{\text{T}} = 3$	ΔE_{eq}	10.55	17.19	22.16	24.05	18.49	$n_{\text{T}} = 3$	ΔE_{eq}	18.92	26.34	31.39	33.94	27.65
	Δd_{eq}	0.12	0.14	0.15	0.15	0.14		Δd_{eq}	0.25	0.27	0.28	0.26	0.26
	$\Delta \omega_{\text{eq}}$	-46.9	-109.6	-188.5	-213.7	-139.7		$\Delta \omega_{\text{eq}}$	-53.6	-144.1	-233.7	-256.4	-172.0
	IE	13.01	20.00	27.43	30.18	22.65		IE	32.34	41.13	52.19	56.39	45.51
$n_{\text{T}} = 4$	ΔE_{eq}	10.76	19.97	23.77	25.74	20.06	$n_{\text{T}} = 4$	ΔE_{eq}	18.14	27.34	30.60	33.66	27.43
	Δd_{eq}	0.15	0.20	0.24	0.21	0.20		Δd_{eq}	0.28	0.34	0.36	0.34	0.33
	$\Delta \omega_{\text{eq}}$	-22.8	-137.8	-164.1	-196.8	-130.3		$\Delta \omega_{\text{eq}}$	-39.1	-136.4	-141.2	-182.0	-124.6
	IE	13.19	28.19	36.84	39.42	29.41		IE	32.30	55.27	67.82	68.92	56.08

$\Delta E^{(1)}, n_{\text{R}} = 3$						$\Delta E^{(2)}, n_{\text{R}} = 3$							
$n_{\text{T}} = 2$	ΔE_{eq}	32.90	36.17	41.62	44.65	38.84	$n_{\text{T}} = 2$	ΔE_{eq}	12.33	15.46	19.08	20.63	16.88
	Δd_{eq}	0.21	0.23	0.23	0.22	0.22		Δd_{eq}	0.07	0.08	0.09	0.08	0.08
	$\Delta \omega_{\text{eq}}$	-308.4	-373.8	-527.3	-575.0	-446.1		$\Delta \omega_{\text{eq}}$	-143.1	-202.4	-259.0	-312.3	-229.2
	IE	47.26	55.58	65.66	68.61	59.28		IE	12.95	16.79	21.26	22.16	18.29
$n_{\text{T}} = 4$	ΔE_{eq}	1.52	2.13	2.93	3.55	2.53	$n_{\text{T}} = 4$	ΔE_{eq}	4.66	7.24	9.33	10.90	8.03
	Δd_{eq}	0.01	0.02	0.03	0.03	0.02		Δd_{eq}	0.04	0.08	0.09	0.09	0.08
	$\Delta \omega_{\text{eq}}$	-5.8	-11.7	-31.9	-20.9	-17.6		$\Delta \omega_{\text{eq}}$	-25.1	-48.3	-56.5	-68.3	-49.6
	IE	1.45	2.26	2.83	3.25	2.45		IE	4.83	8.69	10.69	11.89	9.03

$\Delta E^{(1)}, n_{\text{R}} = 4$						$\Delta E^{(2)}, n_{\text{R}} = 4$							
$n_{\text{T}} = 2$	ΔE_{eq}	38.08	44.22	54.03	61.07	49.35	$n_{\text{T}} = 2$	ΔE_{eq}	19.91	22.19	30.44	33.75	26.57
	Δd_{eq}	0.26	0.31	0.33	0.33	0.31		Δd_{eq}	0.12	0.12	0.16	0.15	0.14
	$\Delta \omega_{\text{eq}}$	-331.4	-455.1	-636.6	-738.0	-540.3		$\Delta \omega_{\text{eq}}$	-211.2	-276.1	-418.1	-487.4	-348.2
	IE	60.39	78.48	106.16	121.95	91.75		IE	22.83	25.10	38.88	42.45	32.31
$n_{\text{T}} = 3$	ΔE_{eq}	1.26	2.43	3.81	5.39	3.22	$n_{\text{T}} = 3$	ΔE_{eq}	-1.30	-3.64	-5.32	-6.49	-4.19
	Δd_{eq}	0.01	0.02	0.02	0.04	0.02		Δd_{eq}	-0.01	-0.03	-0.03	-0.04	-0.03
	$\Delta \omega_{\text{eq}}$	-14.4	-15.0	-31.0	-60.4	-30.2		$\Delta \omega_{\text{eq}}$	12.3	12.4	24.0	51.1	24.9
	IE	1.16	2.31	3.56	5.13	3.04		IE	1.43	3.07	4.39	5.23	3.53

results. In other words, if accurate electron densities are available, one would expect better performance from first order alchemical prediction. However, we note that in the case of dissociation, PBE performs better than PBE0, as can be observed by the deviation from coupled-cluster results.

Coupled-cluster with single, double, and perturbative triple (CCSD(T)) results in this section is computed by Gaussian09[112] using basis set aug-cc-pVTZ-[113] with default setup.

4.3.2 Vertical iso-valence-electronic changes involving single, double, and triple bonds

Having discussed covalent bonds involving hydrogen, we now turn to single (XH₃-Y), double (XH₂=Y), and triple (HX#Y) bonds among *p*-block elements. Since third row elements can either be alchemically compressed to the corresponding second row element in the same column, or expanded to the fourth row element, we chose third row based

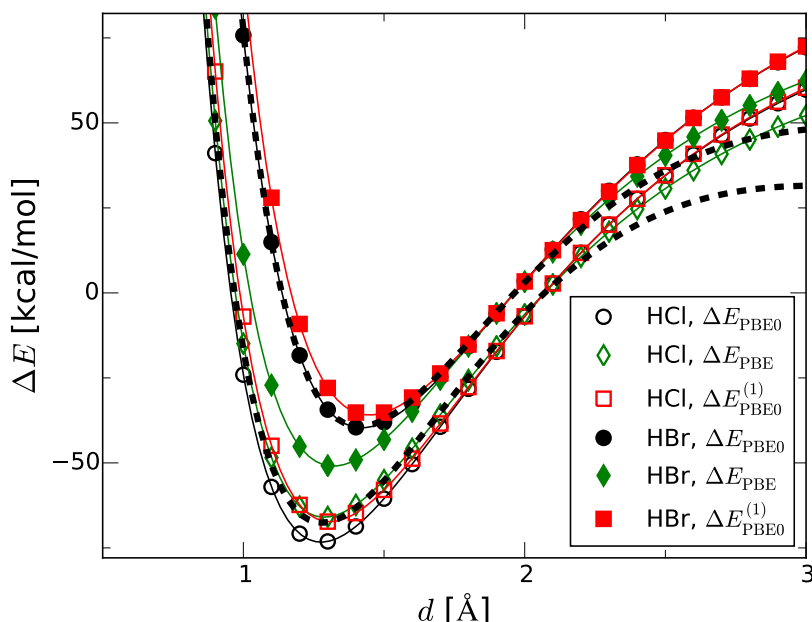


FIGURE 4.3: Binding potentials of HBr and HCl computed by PBE0 (black circles), PBE (green diamonds), first order alchemical prediction: HBr→HCl (red squares), and CCSD(T) (dashed line).

reference systems for single, double, and triple bonds, namely SiH_3Cl , SiH_2S , and HSiP . The resulting eight alchemical paths are combinations of changing the Si atom ($\text{Si} \rightarrow \text{C}$, $\text{Si} \rightarrow \text{Ge}$) or its binding partner ($\text{Cl} \rightarrow \text{F}$, $\text{Cl} \rightarrow \text{Br}$, $\text{S} \rightarrow \text{O}$, $\text{S} \rightarrow \text{Se}$, $\text{P} \rightarrow \text{N}$, $\text{P} \rightarrow \text{As}$). In Figs. 4.4 first and second order alchemical predictions are shown for the bonding potential using vertical transmutations from the three reference molecules.

More specifically, single bonds have been investigated in the nine following molecules with 14 valence electrons: ($n_X = 2$) CH_3F , CH_3Cl , CH_3Br , ($n_X = 3$) SiH_3F , SiH_3Cl , SiH_3Br , ($n_X = 4$) GeH_3F , GeH_3Cl , and GeH_3Br . For double bonds, we have considered the following nine unsaturated molecules with 12 valence electrons: ($n_X = 2$) CH_2O , CH_2S , CH_2Se , ($n_X = 3$) SiH_2O , SiH_2S , SiH_2Se , ($n_X = 4$) GeH_2O , GeH_2S , and GeH_2Se . And finally for triple bonds, we studied the following nine molecules with 10 valence electrons: ($n_X = 2$) HCN , HCP , HCAs , ($n_X = 3$) HSiN , HSiP , HSiAs , ($n_X = 4$) HGeN , HGeP , and HGeAs .

Prediction results in Fig. 4.4 suggest that the accuracy of the alchemical prediction of $\Delta E(d, d_0)$ exhibits very similar trends as the ones discussed above for vertical changes in the hydrogen containing single bond: First order predictions systematically achieve remarkable predictive power of chemical accuracy whenever a third row element is coupled to a fourth row element. Corresponding second order predictions deteriorate the accuracy due to inflection points near $\lambda = 0$. Second order Taylor series provides superior predictive power only in the case when at least one of the heavy elements in the

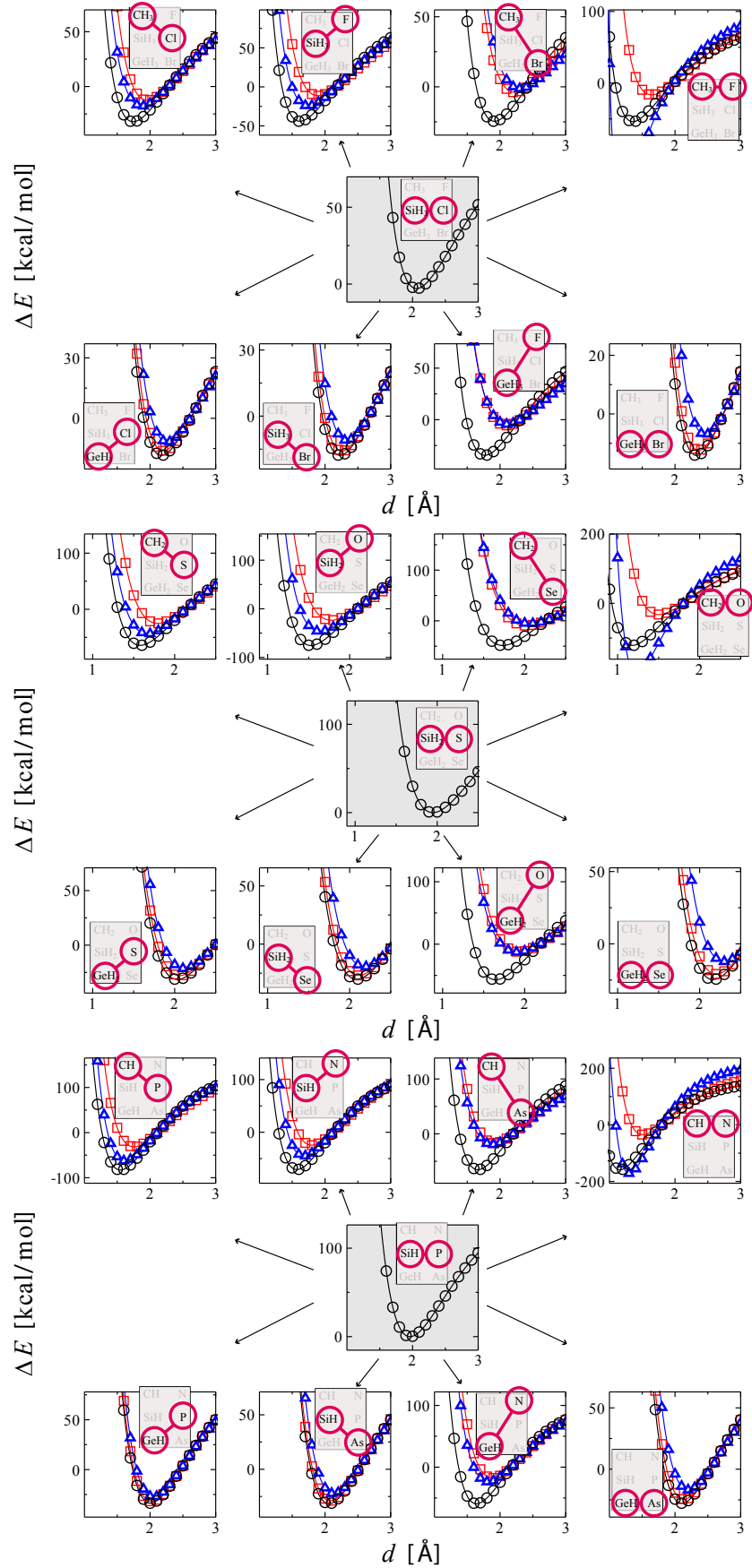


FIGURE 4.4: Alchemical predictions of single (top), double (middle), and triple (bottom) bond potentials. Curves are shown for eight target systems (specified as insets), iso-electronic with reference molecule SiH_3Cl (upper panel), SiH_2S (middle panel), and HSiP (bottom panel). True (black circles), first (red squares) and second (blue triangles) order vertical alchemical predictions of heavy atom bond dissociation curves.

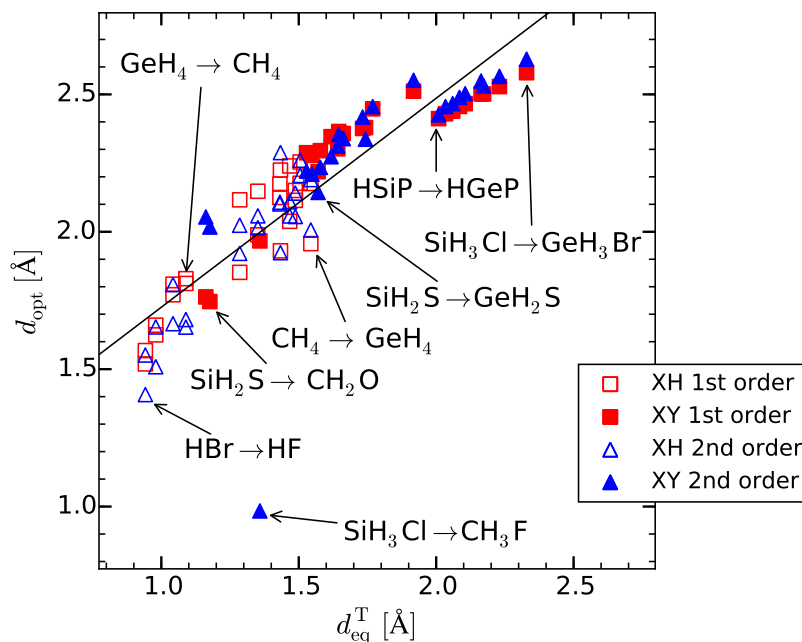


FIGURE 4.5: Scatter plot of optimized reference bond length d_{opt} versus equilibrium bond length of target molecule, denoted by d_{eq}^T . Linear regression gives $d_{\text{opt}} = 0.76d_{\text{eq}}^T + 0.97$ Å with MAE=0.11 Å and RMSE=0.15 Å. First order (red empty squares) and second order (blue empty triangles) d_{opt} of covalent X-H bond stretching, as well as first order (red filled squares) and second order (blue filled triangles) d_{opt} of X-Y, X=Y, X#Y stretching are shown, where -, =, # stand for single, double, and triple bonds. Some of the alchemical paths are highlighted by black arrows. All numbers are given in Table. 4.1 and Table. 4.2.

reference molecule couples to a lighter element. If both heavy atoms are simultaneously transmuted vertically up, *e.g.* $\text{SiH}_3\text{Cl} \rightarrow \text{CH}_3\text{F}$, second order estimates reduce the error of the first order prediction, a substantial over correction is observed.

For the case of one heavy element transmulating upward, the other vertically downward, second order corrections from both atoms nearly cancel each other. One atom couples to a lighter element resulting a negative curvature. The other couples to a heavier one giving a positive curvature, as observed in the asymmetry of ΔE in Fig. 4.2 and systematic behaviour of second order corrections in Fig. 4.1 and Table. 4.1. Summation of the two gives a near zero curvature behaviour at $\lambda = 0$.

The observations above are supported by quantitative prediction errors (described in Sec. 4.2.3), summarized in Table. 4.2. The best performance is achieved for the change $\text{SiH}_3\text{Cl} \rightarrow \text{GeH}_3\text{Br}$ ($n_X/n_Y = 4/4$) with energy error $\Delta E = 0.6$ kcal/mol and integrated IE = 0.9 kcal/mol. Corresponding predictions of equilibrium distance deviates 0.03 Å with vibration frequency deviate -1.1 cm^{-1} . However, the worst prediction occurs for the change $\text{SiH}_3\text{Cl} \rightarrow \text{CH}_3\text{F}$ ($n_X/n_Y = 2/2$), estimated by second order correction.

4.3.3 Empirical d_{opt}

Given these observations, it becomes clear that the choice of d_0 in Eq. (4.3) is crucial for linearizing the property in alchemical coupling parameter λ , and hence essential for the performance of the perturbation based predictions. All the error minimizing d_{opt} have an approximately linear dependence of target molecule’s d_{eq} , no matter if the reference is single, double, or triple bond. Also, the relationship is preserved, independent of the fact if predictions are made with first or second order estimates. This is shown in Fig. 4.5. The relationship obtained through linear regression is specified as well. The outlier in Fig. 4.5 at target $d_{\text{eq}} \approx 1.35\text{\AA}$ and $d_{\text{opt}} \approx 1.0\text{\AA}$ is due to the second order prediction of $\text{SiH}_3\text{Cl} \rightarrow \text{CH}_3\text{F}$, where a strong overcorrection is present.

4.4 Conclusions

The performance and predictive power of truncated Taylor series in iso-electronic chemical spaces through alchemical coupling has been investigated for changes in covalent bonding. For vertical linear transmutation interpolations (same geometry, same number of atoms) between reference and target compounds our results suggest that chemically accurate predictions are possible when transmutating p -block atoms from the third and fourth row of the periodic table and using first order, Hellmann-Feynman theory, based predictions. Since these estimates are analytical, our results suggest that it is possible to scan PES throughout CCS with unprecedented accuracy and speed.

Second order estimates are less accurate. This is not due to the fact that higher order terms cancel, but rather due to the fact that (a) changes in relative energies (bonding) are near-linear with respect to alchemical coupling, and (b) inflection points can occur which lead to worse predictions. An inflection point near $\lambda = 0$ is always observed when an main group element is coupled to heavier element. The absence of inflection points near $\lambda = 1$ gives rise to different predictive power of second order correction: As such, the asymmetry of $\Delta E(d, d_0)$ with respect to $\lambda = 0.5$ results in asymmetric predictive performance. Furthermore, the choice of the reference geometry has a dramatic impact on the predictive power of the alchemical estimates. A linear relationship has been identified, ($d_{\text{opt}} \approx 0.76 d_{\text{eq}}^{\text{T}} + 0.97 \text{\AA}$), that can be used to predict optimal d_0 requiring only rough estimates of the equilibrium bond-length in the target molecule.

In summary, our findings indicate that a careful choice of alchemical interpolation paths enables predictive power through alchemical perturbation.

TABLE 4.2: Summary of error measure in Fig. 4.4 for first order (upper table) and second order (lower table). The reference molecules are denoted by $H_R = \{\text{SiH}_3\text{Cl}, \text{SiH}_2\text{S}, \text{HSiP}\}$. The primary quantum number of heavy atom $X = \{\text{C}, \text{Si}, \text{Ge}\}$ and $Y = \{\text{N}, \text{P}, \text{As}, \text{O}, \text{S}, \text{Se}, \text{F}, \text{Cl}, \text{Br}\}$ in target molecule are denoted by n_X and n_Y respectively for each column. Error measures of each panel are collected in corresponding cell with unit [kcal/mol] for ΔE_{eq} and IE, [\AA] for Δd_{eq} , and $\Delta \omega_{\text{eq}}$ is in [cm^{-1}].

$\Delta E^{(1)}$									
H_R	2/3	3/2	2/4	2/2	4/3	3/4	4/2	4/4	
SiH ₃ Cl	ΔE_{eq}	9.04	18.25	10.37	20.37	0.95	0.80	15.83	0.60
	Δd_{eq}	0.22	0.27	0.31	0.29	0.03	0.03	0.33	0.03
	$\Delta \omega_{\text{eq}}$	-91.83	-169.02	-106.27	-237.85	-12.05	-5.02	-148.62	-1.09
	IE	10.14	16.82	12.98	14.85	1.18	0.98	16.74	0.88
SiH ₂ S	ΔE_{eq}	21.69	30.78	22.87	52.34	1.83	1.28	26.77	1.42
	Δd_{eq}	0.21	0.25	0.29	0.32	0.02	0.03	0.28	0.04
	$\Delta \omega_{\text{eq}}$	-191.24	-241.71	-201.69	-436.06	-9.88	-12.82	-218.43	-10.84
	IE	15.16	19.47	14.94	33.25	0.92	0.92	17.18	0.98
SiHP	ΔE_{eq}	31.35	26.34	30.42	76.29	0.79	1.32	23.91	1.28
	Δd_{eq}	0.21	0.23	0.25	0.35	0.01	0.02	0.24	0.03
	$\Delta \omega_{\text{eq}}$	-240.70	-229.35	-250.08	-568.76	-1.77	-9.15	-192.45	-2.34
	IE	22.28	19.42	23.73	41.45	0.93	1.22	18.35	1.56
$\Delta E^{(2)}$									
SiH ₃ Cl	ΔE_{eq}	7.79	11.84	13.29	-103.39	3.00	2.54	17.60	2.68
	Δd_{eq}	0.13	0.13	0.39	-0.18	0.09	0.10	0.34	0.13
	$\Delta \omega_{\text{eq}}$	-91.30	-111.67	-147.58	1912.55	-22.96	-18.27	-176.49	-13.20
	IE	6.48	8.47	16.19	32.26	3.36	3.61	17.63	3.83
SiH ₂ S	ΔE_{eq}	12.46	17.88	28.09	-101.05	5.18	4.59	29.36	5.36
	Δd_{eq}	0.09	0.11	0.33	0.06	0.08	0.09	0.26	0.13
	$\Delta \omega_{\text{eq}}$	-113.32	-128.03	-258.05	1290.29	-36.73	-30.13	-256.16	-27.49
	IE	6.84	9.58	15.34	33.66	2.47	2.75	14.82	2.42
SiHP	ΔE_{eq}	12.04	14.92	30.85	-50.92	3.05	4.22	21.95	4.93
	Δd_{eq}	0.07	0.11	0.22	0.14	0.04	0.07	0.18	0.10
	$\Delta \omega_{\text{eq}}$	-94.83	-118.62	-256.64	747.69	-11.84	-29.35	-182.63	-22.36
	IE	8.56	9.45	21.74	28.53	2.96	4.58	14.80	5.71

Chapter 5

Rational crystal design

5.1 Introduction

The essential advancements of technologies hinge upon the performance of underlying materials. The efficiency of photovoltaic materials heavily depends on the electronic properties of the solid-state or organic semiconductors, the performance of a drug molecule is usually affected by the free energy of binding between drug target compound, the stability of energy storage materials is governed by their energy landscapes and vibrational modes. Until the theoretical breakthrough of quantum mechanics (QM) at the beginning of last century, improvements in performance were achieved by trial and error. Nowadays, with the help of modern computers, properties of materials can be predicted *in silico* even before their assessments are carried out experimentally. [7, 10, 47–49, 114, 115] The costly experimental evaluations can focus only on those materials predicted to exhibit the most promising properties. Such computational materials design procedures therefore hold much promise to lift some of the chemical challenges that the world is facing today.

Finding materials with superior properties can be seen as an optimization problem in chemical compound space (CCS), the space consisting of all possible materials.[43, 116, 117] In practice, due to the combinatorial nature of CCS, an enormous number of possible materials makes a random search for optimal candidates a daunting challenge. Both, approximations to accelerate the evaluation of QM solution, as well as algorithms for efficient searching over CCS are desirable. Optimization algorithms can be applied iteratively by performing QM calculations for intermediate materials during each optimization step. [8, 79, 103, 118] Among these, gradient-based methods utilizing alchemical derivatives have been shown to yield promising results for many applications.[15, 31, 119, 120]

Once the QM solution of a reference material, represented by its Hamiltonian H_R , is found, any unknown target material, encoded by H_T , can be considered as the end point

of a coupling Hamiltonian $H(\lambda) = H_R - \lambda(H_T - H_R)$, where $0 \leq \lambda \leq 1$ is the coupling parameter. Electronic eigenvalues, of any target material can be approximated by a first order truncated Taylor series $E(\lambda) \approx E(0) + \partial_\lambda E(\lambda)|_{\lambda=0}$, where the “alchemical” derivative is readily available due to the Hellmann-Feynman theorem (Eq. (2.28)).[30] Note that the Hellmann-Feynman theorem, while typically used only for (relative) energies, including total energies, holds for any of the Hamiltonian eigenfunctions, as long as the eigenfunctions form an orthogonal set.

It has been demonstrated that chemical accuracy can be achieved by alchemical prediction,[120, 121] under the condition that the number of valence electrons and geometry remain unchanged, and that the interpolating elements come from the third or fourth period in Mendeleev’s table. In this chapter, we study the use of alchemical derivatives for the design of solid-state materials, and discuss some applications and limitations of exploring CCS.

Kohn-Sham density functional theory (KS-DFT)[22] and generalized gradient approximation (GGA) functional are used throughout. The prediction performance is measured in terms of deviation from KS-DFT/GGA results. Note that the choice of the level of theory is somewhat arbitrary: The approach pursued is valid for any level of theory if the perturbation theory can be applied. The predictive performance, however, possibly depends on the level of theory employed. We focus on the prediction of band structures of III-V semiconductors. More specifically, we perform a computational design application for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals, carried out by optimization in CCS based on a combination of alchemical derivatives with a genetic algorithm.

5.2 Results and discussion

5.2.1 Predictions and objective function

The band structure is one of the most important properties in semiconductors due to their many electronic applications.[104, 122–124] We investigate the predictive power of alchemical predictions for estimating band structures of III-V and IV-IV semiconductors drawn from 3^{rd} to 5^{th} period elements, *i.e.* AlP, GaP, InP, AlAs, GaAs, InAs, AlSb, GaSb, InSb for III-V semiconductors and Si, SiGe, SiSn, Ge, GeSn, Sn for IV-IV semiconductors. Alchemical prediction from the reference material (ref) of lattice constant a to any target material (tar) on the i^{th} band at a given crystal momentum \mathbf{k} is constructed as follows:

$$\varepsilon_{i,a}^{\text{pred}}(\mathbf{k}) = \varepsilon_{i,a}^{\text{ref}}(\mathbf{k}) + \partial_\lambda \varepsilon_{i,a}^{\text{ref} \rightarrow \text{tar}}(\mathbf{k}), \quad (5.1)$$

where the derivative is a 3D integral, $\partial_\lambda \varepsilon_{i,a}^{\text{ref} \rightarrow \text{tar}}(\mathbf{k}) = \langle \phi_{i,a} | H_{\text{T}} - H_{\text{R}} | \phi_{i,a} \rangle_{\mathbf{k}}$, calculated from the i^{th} Hamiltonian eigenfunction of the reference material of lattice constant a with crystal momentum \mathbf{k} , $|\phi_{i,a}\rangle_{\mathbf{k}}$. The prediction of band gap is therefore $E_{g,a}^{\text{pred}} = \min_{\mathbf{k}} (\varepsilon_{\text{LUMO},a}^{\text{pred}}(\mathbf{k}) - \varepsilon_{\text{HOMO},a}^{\text{pred}}(\Gamma))$, where the maximum of conduction band of materials investigated is located at Γ -point. A predicted band gap is direct if and only if the predicted minimum of $\varepsilon_{\text{LUMO}}$ is located at Γ -point. Lattice constants for the semiconductors are scanned from 5.4 Å to 6.5 Å in step of 0.05 Å including extra points correspond to the lattice constant reported in the literature (see Tabel 5.2). This range covers all equilibrium constants reported for all the semiconductors considered.

We quantify the performance of our predictions using the mean absolute error (MAE) of a prediction to the entire i^{th} band of a target material at lattice constant a , made from some reference material as

$$\text{MAE}_{i,a}(\text{ref}, \text{tar}) = \sum_{\mathbf{k}} |\varepsilon_{i,a}^{\text{pred}}(\mathbf{k}) - \varepsilon_{i,a}^{\text{true}}(\mathbf{k})| w(\mathbf{k}), \quad (5.2)$$

where $\varepsilon_{i,a}(\mathbf{k})$ is the i^{th} eigenvalue of Hamiltonian at lattice constant a and crystal momentum \mathbf{k} . $w(\mathbf{k})$ is the corresponding Monkhorst-Pack[125] weight for the sampled special k-points. Throughout this chapter, the error reported corresponds to average over all bands and lattice constants. The subscripts i and a in Eqs. (5.1) and (5.2) are omitted unless noted otherwise.

5.2.2 Alchemical band structure prediction of binary III-V and IV-IV semiconductors

The averaged MAE over all lattice constants and bands for all combinations of reference/target pair, or alchemical path, is on display in the upper panels in Fig. 5.1. Overall, decent agreement is found with most predictions being in agreement with target by less than 0.5 eV. Alchemical estimates with averaged MAE less than 0.2 eV are highlighted by the white circles. A general trend can be observed: Predictions using semiconductors containing 3rd-row elements in the reference give overall the largest averaged MAE, when the target material has elements from the 5th row, as can be seen by the red/orange corner of the upper panels (left for III-V semiconductors and right for IV-IV semiconductors) in Fig. 5.1. A similar target-reference pattern has been observed for alchemical predictions of covalent bonding, and is due to lack of similarity of electron densities (see Chapter 4)

A scatter plot between averaged MAE and integrated electron density difference, defined as $\Delta\rho(\mathbf{r}) = \rho_{\text{tar}}(\mathbf{r}) - \rho_{\text{ref}}(\mathbf{r})$, is shown in the lower left panel in Fig. 5.1. The results suggest that there is an upper error bound: Any alchemical path with small density changes will give a small predictive error to band structure, as all the points lie

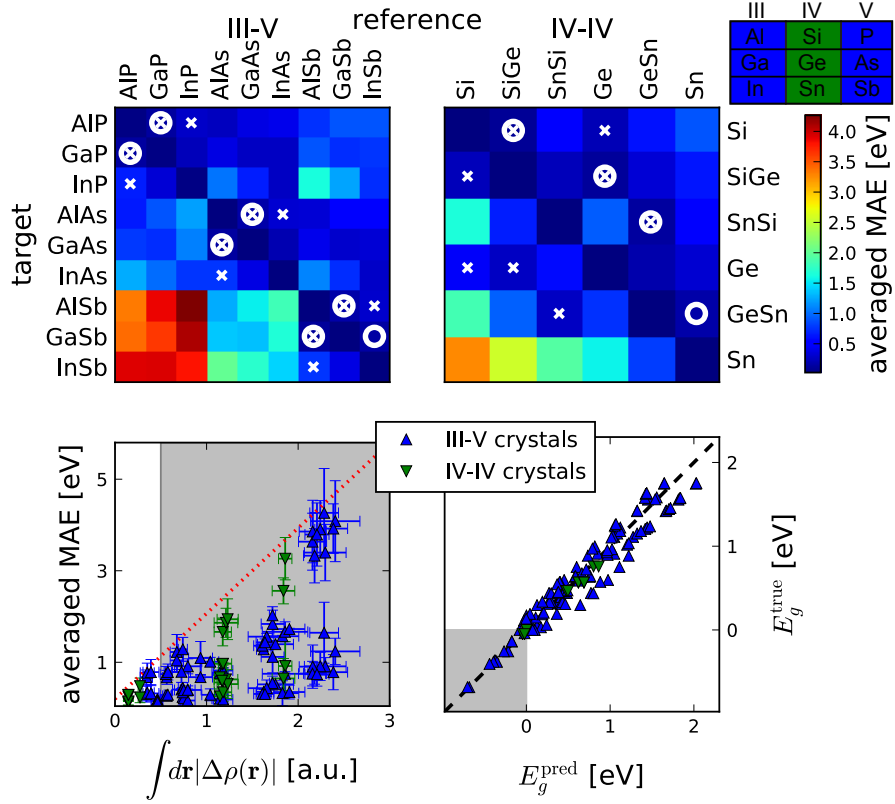


FIGURE 5.1: The upper panels: Prediction errors of band structure on alchemical paths for III-V (upper left) and IV-IV (upper right) semiconductors. Errors less than 0.2 eV are highlighted by white circles. Combinations with $\int d\mathbf{r}|\Delta\rho(\mathbf{r})| < 0.5$ a.u. are highlighted by white crosses. For guidance, the relevant block in the periodic table is shown in the top right corner. The lower left panel: Correlation of averaged MAE and integrated density difference for III-V and IV-IV semiconductors. The error bars represent the variations due to changes in lattice constant of the corresponding term. The lower right panel: Scatter plot of band gap prediction vs. true for alchemical paths with $\int d\mathbf{r}|\Delta\rho(\mathbf{r})| < 0.5$ a.u.. The region with negative E_g is highlighted by the gray background. A linear fit gives $E_g^{\text{true}} \approx 0.93E_g^{\text{pred}} + 0.019$ eV, with $R^2 = 0.93$, RMSE = 0.047 eV and MAE = 0.036 eV.

beneath the red dotted line. However, a small error does not necessarily imply small density changes, as many results with small predictive error correspond to large density changes. This is because the band structure is determined by the relative differences in Hamiltonian eigenvalues and the corresponding orbital structure. Cancellation of higher order (curvature) effects along the alchemical path can lead to a small predictive error for first order estimates yet large density change. Similar error cancellation has also been discussed in the context of covalent bonding in Chapter 4. The fact that small density changes imply accurate predictions can be used as a sufficient condition to detect good predictive power. This might be useful for future studies if decent approximations to inexpensively estimate electron density changes can be found. The region with $\int d\mathbf{r}|\Delta\rho(\mathbf{r})| < 0.5$ a.u. electron/primitive cell is highlighted by the white background in the lower left panel. The band gap predictions of the corresponding alchemical paths

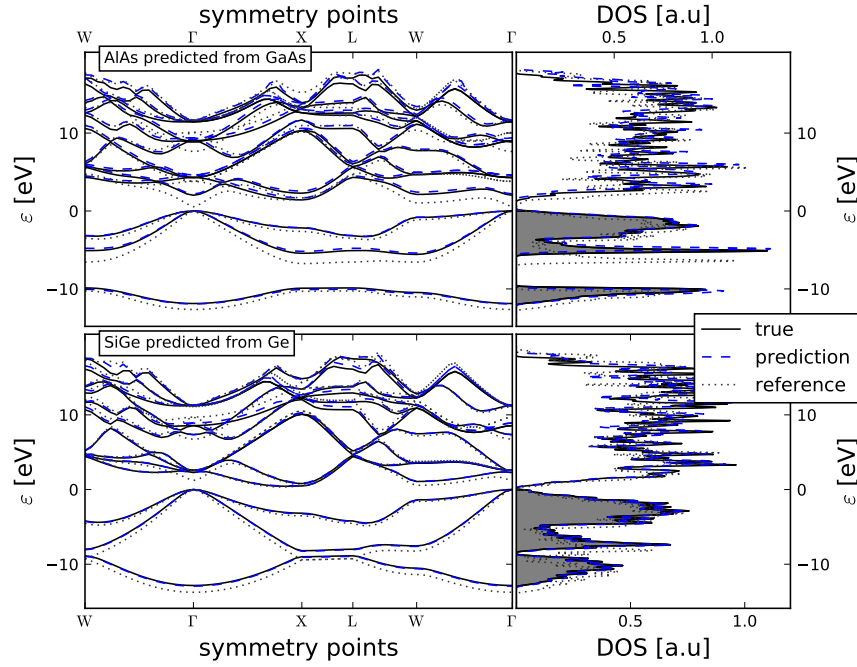


FIGURE 5.2: Band structure and DOS of alchemical predictions with minimal error. AlAs predicted from GaAs is plotted in the upper panels, while SiGe predicted from Ge is plotted in the lower panels. The true (predicted) band structure and DOS of the target crystals are represented by black solid (blue dashed) curves, while black dotted curves represent the reference crystals. Fermi-level is shifted to zero.

highlighted as white crosses in the upper panels of Fig. 5.1. A scatter plot versus the true band gap is shown in the lower right panel, remarkable agreement is found with MAE of less than 0.05 eV. The negative E_g in the lower right panel of Fig. 5.1 correspond to GaSb and InSb with small lattice constants. In such environment, the conduction band minimum is lower than the valence band maximum. Notice that decent predictive power can be achieved by alchemical predictions, satisfying the sufficient condition above, even at such extreme situation.

Band structures and density of states (DOS) corresponding to two alchemical paths exhibit minimal predictive errors: AlAs predicted with GaAs as a reference, and SiGe predicted with Ge as a reference. Corresponding results are plotted in Fig. 5.2. The prediction correctly captures the change from direct band gap at Γ in Brillouin zone for GaAs to indirect band gap at X for AlAs. Furthermore, most of the other details in the band structure and DOS are correctly predicted.

It is possible to use III-V semiconductor reference calculations to predict the band structure of other IV-IV semiconductors. However, such interpolations do not provide satisfactory predictive power when using only first order alchemical derivatives, due to the dissimilarity of the electronic density. Similar conclusions hold when attempting to predict band structures of II-VI semiconductors using III-V reference densities.

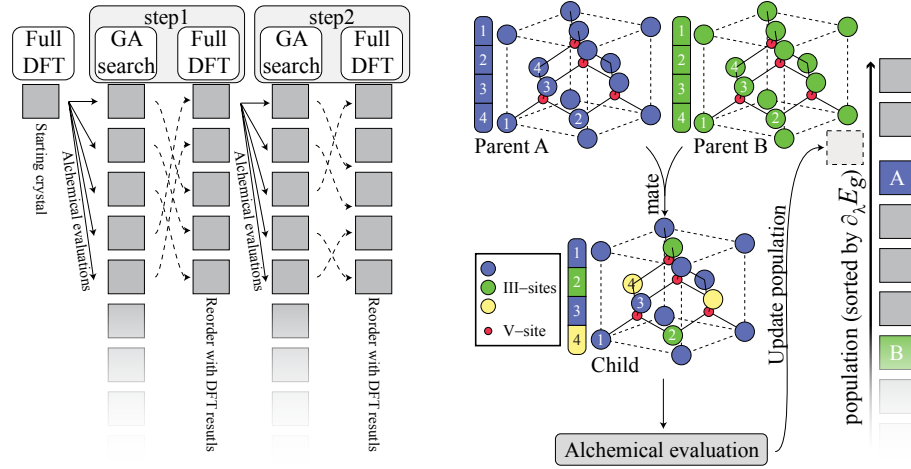


FIGURE 5.3: Schematic illustrations of the hybrid gradient based GA, described in Algorithm 1. Left panel presents the alternating procedures, full DFT and GA search, with two optimization steps. Right panel illustrates the mechanism of the mating procedure. For clarity, the cubical zinc blende unit cell of 8 atoms (4 Ga or Al and 4 As) are shown instead of fcc supercell. The 4 symmetrically unique atoms are highlighted. A child crystal is generated via from parent A (blue) and parent B (green). Atom site 1, 3, and 4 of the child crystal are inherited from parent A while atom site 2 is inherited from parent B. A mutation occurred at atom-site 4 (yellow). The resulting child crystal is evaluated alchemically and inserted to the population list.

5.2.3 Band structure engineering

Motivation

Now we turn to addressing the “inverse question”, utilizing alchemical derivatives.[8] Based on the sufficient condition discussed above and on display in Fig. 5.1, alchemical predictions provide good predictive power for III-V semiconductors when swapping Ga for Al or vice versa. We have investigated if this kind of predictive accuracy is sufficient to enable the design of crystals made up of Al, Ga, and As.

In this context, the inverse question can be formulated as follows: Assuming that $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals can be fabricated with atomistic control, *e.g.* using techniques such as additive nanomanufacturing[3–5], without vacancy or replacement defects, what is the precise spatial arrangement of Al and Ga in the crystal which will result in a maximal direct band gap? To search for crystals with large direct band gap is relevant for optoelectronic and quantum computing applications. However, from the alchemical perspective, the choice to maximize the direct band gap is arbitrary. In fact, any other property could have been targeted just as well, as long as alchemical estimates can be set-up.

In order to assess the usefulness of alchemy for design, we investigate disordered $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals with an explicit description of each III-site occupancy, using $3 \times 3 \times 3$ face-centered-cube (fcc) supercells with 54 atoms. The objective of the optimization is set to maximize the band gap while the bottom of the conduction band must be

located at Γ -point in the Brillouin zone. To this end, we vary the mixing ratio of Al and Ga, as well as their locations. Instead of performing a DFT evaluation for each candidate crystal, the band structure of a candidate crystal is estimated using alchemical derivatives as presented in Sec. 5.2.1. Each alchemical prediction takes less than 1% of the computational cost for a full DFT band structure calculation. This enables the rapid exploration of many configurations and mixture ratios with decent accuracy. The Brillouin zone of the supercell is unfolded such that it corresponds to the one from a primitive cell using spectral weights.[126, 127] There are 27 III-sites in a $3 \times 3 \times 3$ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystal, which can be either Al or Ga. This permits 2^{27} candidate crystals (including some redundant duplicates due to symmetry). As such, despite the location of atoms being clamped, the CCS for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is extremely high dimensional. Consequently, even with alchemical estimates, it is prohibitive to scan through the entire space.

Hybrid gradient based genetic algorithm

A genetic algorithm (GA) is used for stochastic search of the most promising directions for following steepest descent steps using alchemical derivatives. The pseudocode of the global optimization using such hybrid gradient based GA is shown in Algorithm 1

Each of the routines (denoted in bold) in Algorithm 1 is described below:

- **hybrid optimization:** The main routine consist of: A gradient sampling procedure **GA search**; and a wavefunction updating procedure **full DFT** following the steepest gradient, as schematically shown in the left panel of Fig. 5.3. The convergence criterion $|E_g^j - E_g^{(j-1)}| < 0.001$ eV at j^{th} optimization step.
- **full DFT:** It performs full DFT evaluations on the requested crystals and returns the best corresponding true band gap, E_g^{true} , and orbitals, $\{\phi_i\}$.
- **GA search:** The routine stochastically samples the alchemical estimates, using reference orbitals $\{\phi_i\}$. The initial population of twenty crystals is randomly generated. The best five crystals are used for **full DFT** evaluations. The sampling criterion is at least 1400 crystals with direct band gap are found. And the difference between the average of the top-20 population, \bar{E}_g^{pred} , and the best predicted crystal, E_g^{best} , is less than 0.02 eV.
- **get parents:** At each iteration of GA search, two parents, **ParentA** and **ParentB**, are randomly drawn from the top-20 of all searched crystals. In other words, the population size of GA is twenty.
- **mate:** The routine to generate a child from two parent crystals. The occupancy at each of the III-sites in the child crystal is randomly inherited from either parent

Algorithm 1 Hybrid gradient-based/genetic algorithm using alchemical derivative for crystal optimization. The pseudocode of the main routine, hybrid optimization, and two primary functions, full DFT and GA search, are explicitly stated. The variables are denoted by typewriter font.

```

procedure hybrid_optimization(starting_crystal)
   $E_g^{(0)} \leftarrow 0$ 
   $j \leftarrow 1$ 
   $(E_g^{(j)}, \{\phi_i^{(j)}\}) \leftarrow \text{full DFT}(\text{starting\_crystal})$ 
  while  $|E_g^{(j)} - E_g^{(j-1)}| > 0.001 \text{ eV}$  do
    top_5_predicted_crystals  $\leftarrow \text{GA search}(\{\phi_i^{(j)}\})$ 
     $(E_g^{(j+1)}, \{\phi_i^{(j+1)}\}) \leftarrow \text{full DFT}(\text{top\_5\_predicted\_crystals})$ 
     $j \leftarrow j + 1$ 
  return  $E_g^{(j)}$ 

```

```

function full_DFT(crystals)
  perform full DFT evaluations on every crystals
  return  $(E_g^{\text{true}}, \{\phi_i\})$  of the best crystal among crystals

```

```

function GA_search( $\{\phi_i\}$ )
  initialize population of size 20
  history  $\leftarrow$  population
   $E_g^{\text{pred}} \leftarrow \text{alchemical evaluation}(\{\phi_i\}, \text{population})$ 
  initialize  $\bar{E}_g^{\text{pred}}$  and  $E_g^{\text{best}}$ 
   $j \leftarrow 1$ 
  while  $j < 1400$  and  $|\bar{E}_g^{\text{pred}} - E_g^{\text{best}}| < 0.02 \text{ eV}$  do
    (ParentA, ParentB)  $\leftarrow \text{get parent}(\text{population})$ 
    Child  $\leftarrow \text{mate}(\text{ParentA}, \text{ParentB})$ 
    if Child is not in population then
       $\{E_g^{\text{pred}}\} \leftarrow \text{alchemical evaluation}(\{\phi_i\}, \text{Child})$ 
      update population
      if  $E_g^{\text{pred}}$  is direct then
         $j \leftarrow j + 1$ 
        update  $\bar{E}_g^{\text{pred}}$  and  $E_g^{\text{best}}$ 
      else
         $E_g^{\text{pred}} = 0$ 
      update history
  return top 5 crystals in the population and history

```

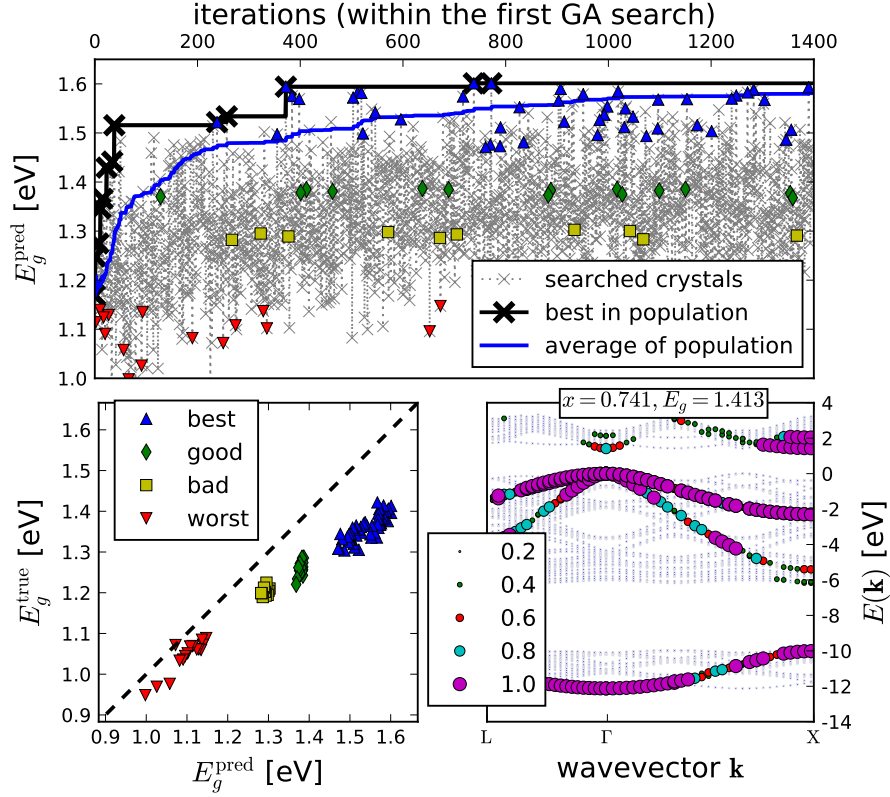


FIGURE 5.4: Optimization history of the first GA search is plotted in the upper panel, where the best crystal in the population at each iteration is highlighted by black crosses. The averaged band gap value over the population of the best 20 crystals is indicated by blue curve, while all searched crystals are plotted as gray crosses. Scatter plot of E_g^{true} vs. E_g^{pred} in the first GA search is shown in the lower left panel for the following four groups: The best, good, bad, and the worst. The same symbols are also used to highlight the corresponding iteration in the upper panel. The linear fit gives $E_g^{\text{true}} \approx 0.7E_g^{\text{pred}} + 0.28$ eV, where $R^2 = 0.85$, MAE = 0.014 eV, and RMSE = 0.018 eV. The true band structure of the $\text{Al}_{0.67}\text{Ga}_{0.33}\text{As}$ crystal with largest predicted band gap value (see Fig. 5.6 and table 5.1) is shown in the lower right panel.

with 50/50 possibility, as illustrated in the right panel of Fig. 5.3. A mutation rate of 0.05% is used at each of the III-sites, where the atom type would flip from Al to Ga or vice versa as indicated the yellow atom in the right panel of Fig. 5.3.

- **alchemical evaluation** estimate the band structure of the requested crystal using reference orbital $\{\phi_i\}$, as demonstrated in Sec. 5.2.2. If the prediction of the band gap is indirect, the value of the direct band gap is set to 0.

Analysis: Optimization history

Three independent hybrid gradient based GA optimizations have been carried out, start with three starting crystals: Pure GaAs and two random crystals, where either Ga or Al is randomly chosen for each of 27 III-sites. All three hybrid optimizations converged after two hybrid optimization steps, which give total six GA search histories. Only the

first GA search history is extensively investigated. 3137 crystals have been evaluated in total. Out of these, 1444 are predicted to be direct band gap crystal. Roughly half of the searched crystals are indirect band gap. E_g^{true} are calculated for the following four groups of crystals from the first GA search history, depending on the sorted direct band gap predictions: the best 50 E_g^{best} , the top 491 to 504 E_g^{good} , the top 1001 to 1010 E_g^{bad} , and the worst 15 E_g^{worst} . The analysis of the first optimization history is shown in Fig. 5.4, while the analysis of all three optimization histories is presented in Fig. 5.5.

In Fig. 5.4, the average trend of predicted direct band gap is moving upward as the iteration proceed. It implies that the mating procedure in GA search steers the population towards larger band gap. Since the band structure is determined by the structure of occupied and unoccupied orbitals, the fact that a simple mating mechanism (as in Fig. 5.3) works indicates that crystal truncation and catenation roughly preserve the local structure of orbitals around each atom. As a result, the algorithm identifies crystals with predicted band gap of around 1.6 eV after only several hundred iterations.

Among the best twenty crystals out of 1444 identified (see table 5.1) within the first GA search, a crystal, $\text{Al}_{0.74}\text{Ga}_{0.26}\text{As}$ (structure shown in Fig. 5.6), provides the largest direct band gap of 1.413 eV ($\#(\text{PBE})=1$ in Table 5.1). The unfolded true band structure is plotted in the lower right panel in Fig. 5.4, where the spectral weight at Γ -point is 61.4%. The prediction error of 0.18 eV, as the predicted band gap is 1.583 eV, is large but acceptable. The unexpectedly large error comes from the fact that 19 out of 27 Ga atoms in the pure GaAs crystal are being alchemically transformed into Al, while only one atom is transformed for the aforementioned case of the primitive cells used for Fig. 5.1. The larger error is expected due to the fact of larger perturbation. Accordingly, alchemical predictions have less prediction error when made from reference crystals with fewer alchemical transformations.

Despite the large prediction error, a decent correlation between the predicted band gap and the true band gap can be observed, as shown in the lower left panel of Fig. 5.4. Since the trends are preserved between E_g^{true} and E_g^{pred} , one GA search step is sufficient to identify the global optima. Following the steepest gradient of the band gap derivative, the consecutive GA searches confirm that the best crystals identified during the first GA search belong to the best global optima (see the lower left panel of Fig. 5.5). The qualitative identification of crystals with the largest band gap, therefore, indicates that one-step hybrid optimization is possible, as long as the steepest gradient is found during the first GA search.

Analysis: Optimized structure

The structure analysis on the history of the hybrid optimization is shown in Fig. 5.5. E_g^{pred} span from 0.9 eV to 1.6 eV in the history of the first GA search, as plotted as a

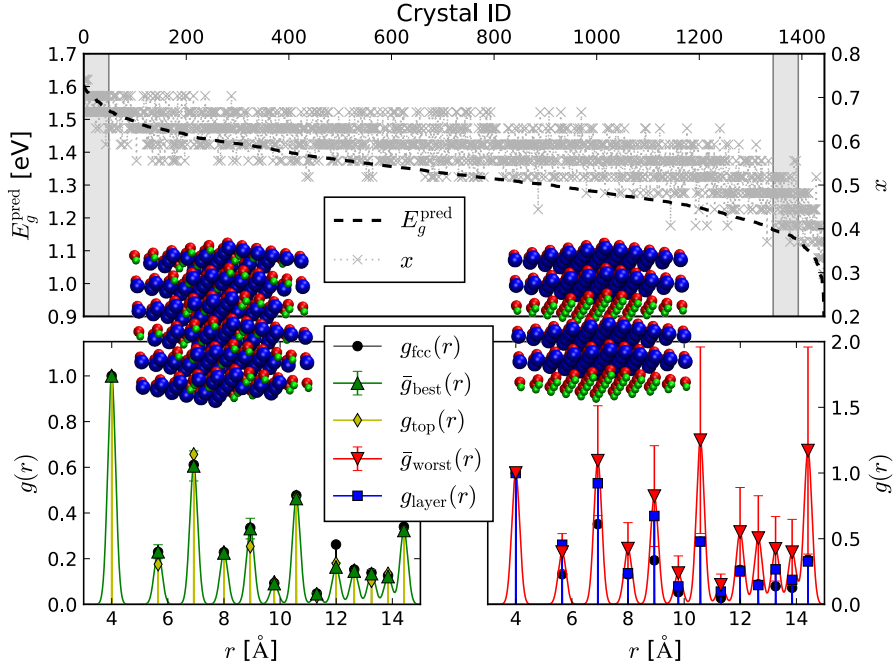


FIGURE 5.5: The upper panel shows the sorted history of the first GA search according E_g^{pred} in Fig. 5.4 in descending order, with corresponding mole fraction x . Radial distribution functions are shown in the lower panels. $g_{\text{fcc}}(r)$ is shown as the black circles in both panels. The crystals used to calculate $\bar{g}_{\text{best}}(r)$ and $\bar{g}_{\text{worst}}(r)$ are highlighted by the left and right gray regions in the upper panel respectively. Error bars for $\bar{g}_{\text{best}}(r)$ and $\bar{g}_{\text{worst}}(r)$ denote the standard deviation among three hybrid optimization within the corresponding best/worst regions. Gaussian smearing is applied for visualization. Corresponding crystals of $g_{\text{top}}(r)$ and $g_{\text{layer}}(r)$ are shown as the inset, where Al, Ga, and As atoms are denoted by blue, green, and red spheres respectively.

sorted sequence in the upper panel of Fig. 5.5. The mole fraction x fluctuates between 30% to 70%, roughly following the trend of E_g^{pred} . That is, larger x implies greater possibility of larger E_g^{pred} . However, the fluctuation of x implies the significance of explicit Ga/Al spatial configuration. On the contrary to implicit methods, such as virtual crystal approximation or special quasirandom structure[61, 128], a fixed mole fraction, say $x > 0.6$, provides crystals of a wide range of band gap values $1.12 < E_g < 1.36$ eV (according to the linear fit from the history of the first GA search in the lower right panel of Fig. 5.4). The band gap variation of 0.24 eV could be critical for applications where high precision is required, such as quantum computing.

The lower panels of Fig. 5.5 show the analysis of radial distribution functions between Ga and Al atoms, calculated from the history of all three hybrid optimization histories, or six GA searches (two hybrid optimization steps each). \bar{g}_{best} and \bar{g}_{worst} are calculated for the best and the worst crystals, as highlighted by the gray regions in the upper panel, and averaged over all six GA searches out of three independent hybrid optimization histories. Notice that the III-sites in the zinc blende crystal structure are

identical to fcc structure. Since the Al-Ga radial distribution function of a random alloy of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ will give $g_{\text{fcc}}(r)$. It is therefore meaningful to compare with $g_{\text{fcc}}(r)$, which is equivalent to Ga-Ga radial distribution function of a perfect GaAs crystal. The comparison to $g_{\text{fcc}}(r)$ quantifies how much the structure differs from random alloy.

$\bar{g}_{\text{best}}(r)$ follows closely $g_{\text{fcc}}(r)$ except for a significant decrease in the number of Al-Ga pairs separated by 12 Å, which indicates a structural motif in the best crystals. The small standard deviation over 300 best crystals (50 from each of the six GA search histories) suggests that this motif should be observable in most of the best crystals. Moreover, the feature similar to $g_{\text{fcc}}(r)$ implies that the Ga and Al atoms form more or less a homogeneous alloy, except for the structure motif which gives rise to the special feature at 12 Å. The overall small standard deviations imply that similar structures can be found within the best crystals. However, large standard deviations at $r = 5.6$ Å, $r = 6.9$ Å, and $r = 8.0$ Å imply that small structural variations should be expected among the best crystals. As an example, the best crystal $\text{Al}_{0.74}\text{Ga}_{0.26}\text{As}$ ($\#(\text{PBE})=1$ in Table. 5.1) from the first GA search is shown as $g_{\text{top}}(r)$, where all peaks coincide with \bar{g}_{best} except for peaks located at $r = 5.6$ Å, $r = 6.9$ Å, and $r = 8.0$ Å. Since multiple structures could reproduce any given $g(r)$, however, the spatial motif which gives rise to the decreased peak at 12 Å can not be easily visualized directly.

In contrast for the best crystals, there are more structural variations, as can be observed by the large standard deviations in the lower right panel of Fig. 5.5. A rough trend can be observed that $\bar{g}_{\text{worst}}(r)$ tend to increase with respect to r . This means the number of Al-Ga pairs as the nearest neighbor, $r \approx 4$ Å, is reduced. This can be exemplified by a layered $\text{Al}_{0.67}\text{Ga}_{0.33}\text{As}$ structure, which results in a relatively small band gap of $E_g^{\text{true}} = 1.19$ eV, (see Fig. 5.5). Because the Ga atoms form a layer, two out of four nearest neighbor sites for each Ga atom are occupied by two Ga atoms in the same layer. The increases of $g_{\text{layer}}(r)$ with larger distances can be observed in in Fig. 5.5. This suggests that Ga clusters are more likely to be found among the worst crystals, rather than among the best ones.

HSE validation

To validate our results, we also calculated the corresponding HSE band gap which is known to have better agreement with experiments than PBE.[129–131] The final population of the best 20 crystals predicted by the first GA search is listed in Tab. 5.1, where the crystals are sorted according to E_g^{HSE} ordering. The best four HSE crystals are shown in Fig. 5.6. The mole fraction $x \approx 0.7$ is identified as transition point from direct band gap to indirect band gap, which is in line with what implicit methods, such as virtual crystal approximation or special quasirandom structure, predicted.[61, 128, 132]

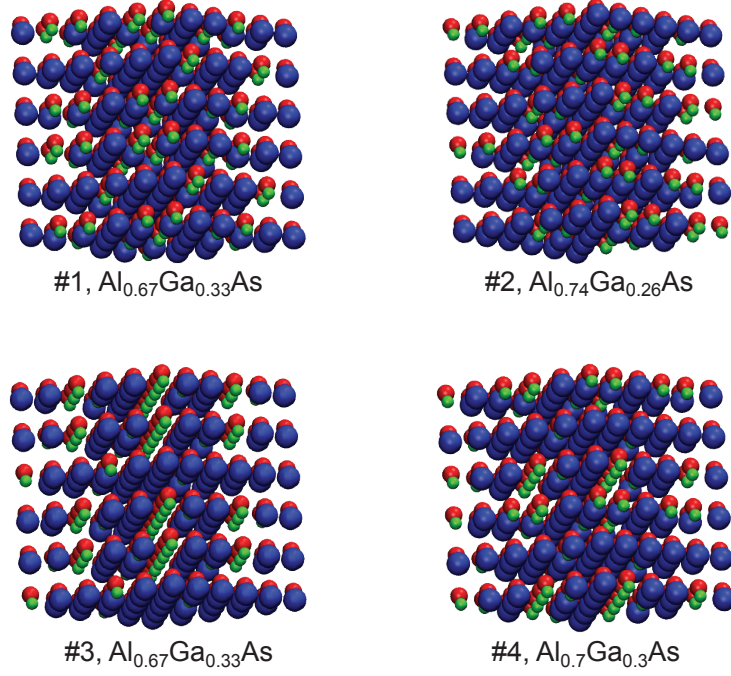


FIGURE 5.6: The best four crystals with $E_g^{\text{HSE}} \approx 2.1$ eV according to Table. 5.1. The HSE ordering rank and the crystal formula with mole fraction is shown for each crystal. The $3 \times 3 \times 3$ supercell is extended to $6 \times 6 \times 6$ to illustrate 2-fold symmetry, where Al, Ga, and As are represented by blue, green, and red spheres.

(Fig. (5.5) linear fit in Fig. 5.4). Yet the ordering of the band gap value changes significantly when using PBE or alchemical estimates.

Differences between the ordering of E_g^{PBE} and E_g^{HSE} are expected because it was shown that the exact exchange gives better accuracy.[133] Different ordering between E_g^{pred} and E_g^{PBE} is due to the higher order effects in CCS, where the first order gradient does not reproduce the correct value at the end point of the alchemical interpolation. Since the difference between E_g^{pred} and E_g^{PBE} is small, as shown in Fig. 5.1 and 5.4, significant differences between E_g^{pred} and E_g^{HSE} can be expected as well.

5.3 Conclusion

We have discussed the performance of first order alchemical derivatives to calculate band structures of III-V and IV-IV semiconductors. A sufficient condition for predictive estimate has been formed. It can be formulated as following: If the integrated density change is smaller than 0.5 a.u., errors of the predicted band structure are guaranteed to be smaller than 1 eV. This condition is used to select a RCD application, where the band structure of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is optimized.

Direct band gap $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals with $E_g > 2.1$ eV and mole fraction $x \approx 0.7$ were found. Inspection of radial distribution functions of crystals with largest direct band gap suggests that for heterogeneous $\text{Al}_x\text{Ga}_{1-x}\text{As}$, such as quantum wells, quantum wires,

TABLE 5.1: Sorted direct band gap of the best 20 $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals from the first GA search. The ordering of HSE, HSE band gap E_g^{HSE} , ordering of PBE, PBE band gap E_g^{PBE} , ordering of predicted band gap, predicted band gap E_g^{pred} , mole fraction x , and spectral weight of the bottom of conduction band at Γ -point w_Γ are listed. All band gap values are reported in eV.

#	E_g^{HSE}	#(PBE)	E_g^{PBE}	#(pred)	E_g^{pred}	x	w_Γ (%)
1	2.100	5	1.400	8	1.583	0.67	70.72
2	2.099	1	1.414	9	1.583	0.74	61.39
3	2.098	13	1.405	16	1.572	0.67	70.87
4	2.098	12	1.394	18	1.569	0.70	64.76
5	2.098	9	1.413	2	1.601	0.74	59.77
6	2.098	2	1.406	3	1.594	0.74	62.34
7	2.098	8	1.400	6	1.589	0.74	59.43
8	2.098	18	1.390	10	1.582	0.67	67.89
9	2.096	17	1.396	1	1.601	0.70	65.48
10	2.096	3	1.376	4	1.593	0.67	67.63
11	2.096	20	1.382	19	1.569	0.70	64.54
12	2.096	16	1.397	5	1.590	0.70	63.40
13	2.096	15	1.379	15	1.573	0.67	69.01
14	2.096	14	1.368	20	1.568	0.67	66.22
15	2.096	7	1.370	12	1.576	0.67	70.31
16	2.095	19	1.376	13	1.575	0.70	64.57
17	2.095	6	1.382	7	1.585	0.70	63.47
18	2.094	10	1.382	11	1.579	0.70	63.12
19	2.094	11	1.367	14	1.574	0.67	68.18
20	2.092	4	1.359	17	1.570	0.67	70.64

TABLE 5.2: Lattice constants of III-V and IV-IV semiconductors in Å. Most of the lattice constants are taken from Ref. [134], while SnSi from Ref. [135], GeSn from Ref. [136], and SiGe from Ref. [137]

AlP	5.4635	AlAs	5.6600	AlSb	6.1355	Si	5.4310	SiGe	5.4320
GaP	5.4510	GaAs	5.6535	GaSb	6.0900	Ge	5.6580	GeSn	6.0758
InP	5.8600	InAs	6.0500	InSb	6.4700	SnSi	5.9610	Sn	6.4892

and quantum dots, one should expect smaller band gaps than for more homogeneous crystals.

5.4 Computational details

Alchemy and PBE results were computed using the **ABINIT** package[138]. The PBE functional[24] with Goedecker norm-conserving pseudopotentials[108, 109] and planewave cutoff of 100 Ry was used for all single point GGA computations. Monkhorst-Park mesh[125] in Brillouin zone of $6 \times 6 \times 6$ for primitive cell band structure, and $3 \times 3 \times 3$ for band structure optimization (with $3 \times 3 \times 3$ fcc supercell) were used respectively. HSE[28] results for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ $3 \times 3 \times 3$ fcc supercell were obtained using the **VASP**[139]

package with PAW[140] pseudopotentials. VASP default energy cutoff is used for all HSE calculations. Experimental lattice constants (if available) were used for every crystal considered in this report. Otherwise calculated values were used from the literature (see table 5.2). Alchemical derivatives were printed by restarting with the reference wavefunction in ABINIT, and setting the SCF iteration step to 0 to evaluate Eq. (2.28).

The results of all searched crystals are synchronized at a centralized database, where the result are sorted and the top-20 population list are generated on the fly at each optimization step. This setup permits massively distributed threads to be executed at the same time. The 54-atom $3 \times 3 \times 3$ supercell crystal is extended to $18 \times 18 \times 18$ supercell with 11664 atoms to cover larger Ga-Al distances for radial distribution function analysis. Averaged radial distribution functions are calculated over the best 1-50 and the worst 51-100 crystals among six GA searches. Gaussian smearing with width 0.1 Å is applied to each of the peaks for visualization purpose. The choice of the worst 51-100 instead of the worst 1-50 crystals is to eliminate structural outliers.

Input generation and output analysis for band structure optimization were each distributed to compute node, across different networks on the fly to minimize data transfer. Only the final band structure was collected and maintained in a centralized database. All scripts to generate input and analyze output, rely on python quantum chemistry toolkit.[141] All of the data and most run scripts are publicly available on GitHub.[142]

Chapter 6

Varying geometry: Analytic derivative

6.1 Introduction

There is no mathematical limitation on how to construct alchemical paths under iso-electronic condition. Rather than using a previously obtained binding curve to explore PES of iso-electronic compounds with the same geometry as in the previous chapters, we have also investigated if one can use only *one* reference calculation in order to estimate the *entire* PES through “non-vertical” interpolations. We have further assessed the applicability of Taylor expansion of Eq. (2.27) to non-vertical changes in the following, *i.e.* varying geometry and/or number of atoms between reference and target molecule. We report numerical results for alchemical stretching of chemical bonds using non-vertical transmutations.

6.2 Methods

The same notations are used as introduced in Sec. 4.2.3 for binding energy to hydrogen atoms in small molecules. For non-vertical changes however, we fix the reference bond length $d_0 = d_{\text{eq}}$ to the equilibrium distance of reference molecule, resulting in $\Delta E^{(m)}(d_{\text{eq}}) = 0$. Eq. (4.5) becomes

$$\begin{aligned}\Delta E^{(1)}(d) &= (E_{\text{R}}(d) + \partial_{\lambda} E_{\lambda}(d)) \\ &\quad - (E_{\text{R}}(d_{\text{eq}}) + \partial_{\lambda} E_{\lambda}(d_{\text{eq}})), \\ \Delta E^{(2)}(d) &= (E_{\text{R}}(d) + \partial_{\lambda} E_{\lambda}(d) + \frac{1}{2} \partial_{\lambda}^2 E_{\lambda}(d)) \\ &\quad - (E_{\text{R}}(d_{\text{eq}}) + \partial_{\lambda} E_{\lambda}(d_{\text{eq}}) + \frac{1}{2} \partial_{\lambda}^2 E_{\lambda}(d_{\text{eq}})).\end{aligned}\tag{6.1}$$

6.2.1 Second order derivative

Differentiation of Eq. (2.28), based on linear interpolated Hamiltonian in Eq. (2.26), yields

$$\partial_\lambda^2 E_\lambda = \int d\mathbf{r} (\partial_\lambda \rho(\mathbf{r})) (\partial_\lambda v_\lambda(\mathbf{r})), \quad (6.2)$$

requiring the density response due to the alchemical perturbation. Again, at $\lambda = 0$ this amounts to the density response of the reference system. Evaluation of Eq. (6.2) implies a differing density response for each target system. We have considered three approximations to $\partial_\lambda \rho$ including second order perturbation theory with independent particle approximation[143] (IPA), coupled perturbed (CP) approaches,[33, 144] as well as finite difference approximation (FD). Note that Eq. (6.2) can be rewritten as $\partial_\lambda^2 E = \int d\mathbf{r} d\mathbf{r}' (\partial_\lambda v(\mathbf{r})) (\partial_\lambda v(\mathbf{r}')) \frac{\delta^2 E}{\delta v(\mathbf{r}) \delta v(\mathbf{r}')} = \chi(\mathbf{r}, \mathbf{r}')$ is the static linear response function or susceptibility, well established within conceptual DFT [16, 86, 92, 145–147].

Perturbation theory provides ways to estimate $\partial_\lambda \rho_\lambda(\mathbf{r})$. [85] Within IPA [143, 148, 149], the static density response for a close-shell system is approximated by

$$\begin{aligned} \partial_\lambda \rho_\lambda(\mathbf{r}) \approx & -4 \sum_{ia} \phi_i(\mathbf{r}) \phi_a(\mathbf{r}) \\ & \times \int d\mathbf{r}' \frac{\phi_i(\mathbf{r}') \phi_a(\mathbf{r}')}{\varepsilon_a - \varepsilon_i} \partial_\lambda v_\lambda(\mathbf{r}'), \end{aligned} \quad (6.3)$$

where $\{\phi_i, \varepsilon_i\}$ denote the i^{th} occupied molecular orbitals (MOs) and their eigenvalues, while $\{\phi_a, \varepsilon_a\}$ denote the a^{th} unoccupied counterparts. IPA neglects the influence of the alchemical perturbation on the Hartree and exchange-correlation (xc) potentials.[144, 146] Note that Eq. (6.3) becomes numerically exact for 1-electron system with converged basis set within Hartree-Fock approximation, because of the absence of Coulomb and xc interaction between electrons.

Recently, Yang, Cohen, De Proft and Geerlings derived an expression of the density response that also includes the dependence of Coulomb and xc potential,[33] the CP approach,[92]

$$\begin{aligned} \partial_\lambda \rho_\lambda(\mathbf{r}) = & -4 \sum_{ij} \sum_{ab} \phi_i(\mathbf{r}) \phi_a(\mathbf{r}) \\ & \times (\mathbf{M}^{-1})_{ia,jb} \int d\mathbf{r}' \phi_j(\mathbf{r}') \phi_b(\mathbf{r}') \partial_\lambda v_\lambda(\mathbf{r}'), \end{aligned} \quad (6.4)$$

where the matrix elements of \mathbf{M} are

$$\left\{ \begin{array}{lcl} \mathbf{M}_{ia,jb} & = & (\varepsilon_a - \varepsilon_i)\delta_{ij}\delta_{ab} + 4\mathbf{J}_{ia,jb} + 4\mathbf{X}_{ia,jb}, \\ \mathbf{J}_{ia,jb} & = & \int d\mathbf{r}d\mathbf{r}' \frac{\phi_i(\mathbf{r})\phi_a(\mathbf{r})\phi_j(\mathbf{r}')\phi_b(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \\ \mathbf{X}_{ia,jb} & = & \int d\mathbf{r}d\mathbf{r}' \phi_i(\mathbf{r})\phi_a(\mathbf{r})\phi_j(\mathbf{r}')\phi_b(\mathbf{r}') \\ & & \times \left(\frac{\delta^2 E_{xc}}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \right). \end{array} \right. \quad (6.5)$$

In the limit of $\mathbf{J}_{ia,jb} \rightarrow 0$ and $\mathbf{X}_{ia,jb} \rightarrow 0$, Eq. (6.3) and Eq. (6.4) are equivalent.

Alternatively, one can also introduce an explicit small perturbation and converge the new density at $\Delta\lambda \ll 1$. The density response can then be estimated via FD, $\partial_\lambda \rho(\mathbf{r}) \approx \frac{\rho_{\Delta\lambda}(\mathbf{r}) - \rho_R(\mathbf{r})}{\Delta\lambda}$. In practice, instead of starting the SCF for the perturbed system from atom based initial guesses, we restart with $\rho_R(\mathbf{r})$ resulting in convergence within few SCF steps.

6.2.2 Higher order derivatives

Møller-Plesset (MP) perturbation theory[29, 150] is used to estimate correlation energy corrections based on converged Hartree-Fock results. The derivation of higher order corrections in MP theory are equivalent to the m^{th} order alchemical derivative. Here, instead of the two-particle operator for electron-electron interaction as perturbation in MP theory, the alchemical perturbation operator $H_T - H_R$ can be used. Within IPA, the MP formula can be directly applied to obtain any m^{th} order derivative.

6.3 Alchemical stretching of H_2^+

Since Hartree-Fock approximation is numerically exact for one-electron systems, we employ atomic basis set and all-electron calculation with the following alignment scheme: One proton is centered at $\mathbf{R}_1 = (0, 0, 0)$, the other is aligned along the $+x$ -axis. Equilibrium bond length is used as the single reference system to estimate PES of H_2^+ . Instead of moving in real space, the stretching is done by simultaneous annihilation and creation of nuclear charges at $\mathbf{R}_2^R = (d_{\text{eq}}, 0, 0)$ and $\mathbf{R}_2^T = (d, 0, 0)$. Once a single SCF is done for d_{eq} , the entire PES of stretching is estimated up to $m = 4$ order, according to Eq. (4.4), by scanning through various d while setting $d_0 = d_{\text{eq}}$.

Results for alchemical predictions of H_2^+ bonding potential using only one SCF calculation at equilibrium geometry are shown in Fig. 6.1(a). Due to the variational principle for linearly coupled alchemical Hamiltonians,[43] $\Delta E^{(1)} > \Delta E$ for all interatomic distances. We note that this behaviour is absent for E in Fig. 4.2 due to the effect of PPs. Inclusion of second order term improves the $\Delta E^{(1)}$ prediction, yielding more reasonable

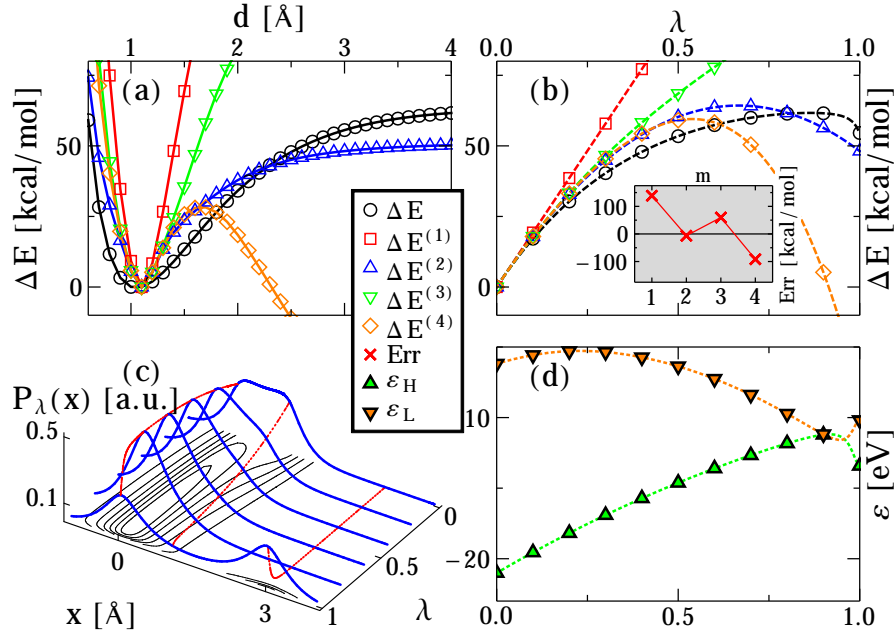


FIGURE 6.1: m^{th} order truncated Taylor series of H_2^+ are denoted by $\Delta E^{(m)}$ in (a) as a function of d at $\lambda = 1$ and in (b) as a function of λ at $d = 3$ Å. Inset shows the error of $\Delta E^{(m)}$ at $\lambda = 1$ (c) Integrated density $P_\lambda(x) = \int dydz \rho_\lambda(\mathbf{r})$ where $\mathbf{r} = (x, y, z)$, is presented as a function of both λ and x at $d = 3$ Å where the integrated density values at nuclei locations highlighted by red lines at $x = 0$ Å, $x = 1.1$ Å, and $x = 3$ Å, while contour lines are draw at the bottom. (d) HOMO/LUMO levels, denoted by ε_H and ε_L respectively, are plotted as a function of λ at $d = 3$ Å.

estimation. However, when including third and fourth order the performance deteriorates again with oscillating behaviour for varying order, as $\Delta E^{(3)}$ overshoot and $\Delta E^{(4)}$ over corrects. Overall $\Delta E^{(2)}$ gives the best prediction.

To explain the oscillating behaviour, it is necessary to understand how the system responds to alchemical perturbation. When λ increases gradually from 0 to 1, the nuclear charge decreases from 1 to 0 at \mathbf{R}_2^R , while increasing from 0 to 1 at \mathbf{R}_2^T . Using the alchemical derivatives at $\lambda = 0$, truncated Taylor series based estimates are plotted as a function of λ : $\Delta E^{(1)}$, $\Delta E^{(2)}$, $\Delta E^{(3)}$, and $\Delta E^{(4)}$ are linear, quadratic, third order, and fourth order polynomial, respectively. Fig. 6.1(b) shows truncated Taylor series as a function of λ , up to fourth order of ΔE at $d = 3$ Å. Clearly, the truncated Taylor series will fail to converge to ΔE at $\lambda = 1$ due to a sharp stabilization of ΔE at $\lambda \approx 0.9$. This implies a strong nonlinear electronic response occurring late in the alchemical coupling regime. This results in the oscillating behaviour of the predicted PES in Fig. 6.1(a). The inset of Fig. 6.1(b) demonstrates the oscillating behaviour of the prediction error at $\lambda = 1$ as a function of Taylor expansion order. While the sign of error changes, the magnitude of error also increases as one includes third or fourth order corrections. Similar behaviour can be observed for other values of d .

The energy stabilization at $\lambda \approx 0.9$ is due to a rapid rearrangement of electron density at $\lambda > 0.9$, as illustrated in Fig. 6.1(c) where integrated density $P_\lambda(x)$ is plotted

as a function of both λ and x at $d = 3 \text{ \AA}$. Cohen and Mori-Sánchez already pointed out that electronic structure changes dramatically with infinitesimal changes in infinitely separated nuclear charges.[77] One would expect this change to intensify as more basis functions are taken into account. This behaviour is also illustrated in Fig. 6.1(c), where integrated electron density planes are shown as a function of λ . The locations of the proton at origin \mathbf{R}_1 , as well as the location of the annihilated proton at \mathbf{R}_2^R , and created proton at \mathbf{R}_2^T , are indicated by red lines. For $\lambda > 0.5$, both ground state and first excited state orbitals are localized: The electronic ground state localized at \mathbf{R}_1 while the first excited state is localized at \mathbf{R}_2^T . At $\lambda \approx 0.9$, the eigenvalues become degenerate, resulting in a rapid change of the ground state density in order to meet the non polar symmetry requirement of H_2^+ , by taking a linear combination of both ground and first excited state. Note that there is no orbital node at midpoint indicating a true ground state for a dissociated H_2^+ molecule. Also, the degeneracy only happens for the system with fractional nuclear charges at $\lambda \approx 0.9$, the degenerate perturbation theory is not applicable in this case. The dramatic change in density stabilizes the system in λ , giving rise to the sharp decrease in energy in Fig. 6.1(b), as λ increases from 0.8 to 1. The degeneracy of the ground state and first excited state is illustrated by the eigenvalue crossing in Fig. 6.1(d), where the eigenvalue of the (highest) occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) are plotted as a function of λ . As the nuclear charge decreases at \mathbf{R}_2^R while increasing at \mathbf{R}_2^T , the HOMO-LUMO gap decreases from $\lambda > 0.3$. Degeneracy occurs at $\lambda \approx 0.9$ when HOMO-LUMO gap reduces to zero. The degeneracy breaks when ground state and first excited state switch order, which results in a delocalized ground state. By contrast, note that the eigenvalues will not cross each other if the stretching is carried out by moving \mathbf{R}_2^R in real space.

Crossing of eigenvalue surfaces limits the radius of convergence of alchemical Taylor expansion series. As a result, the Taylor expansion for this system is diverging at $\lambda = 1$, similar to many cases in Møller-Ploesset theory.[151–153] For asymmetric alchemical interpolations, as exemplified in other examples of this study as well as in previous studies,[9, 15, 31] the energy is typically smooth in all λ values, and perturbation based approaches are reasonable.

6.4 Non-vertical iso-electronic changes

In the final section of this paper, we discuss alchemical non-vertical changes of molecules with more than one electron. More specifically, we present numerical results of non-vertical iso-electronic changes of molecules with ten electrons, including $\{\text{CH}_4, \text{NH}_3, \text{H}_2\text{O}, \text{HF}\}$, using all electron DFT. The H_2^+ example has indicated that non-vertical changes can profit from second order estimates. Since exact analytical expressions are

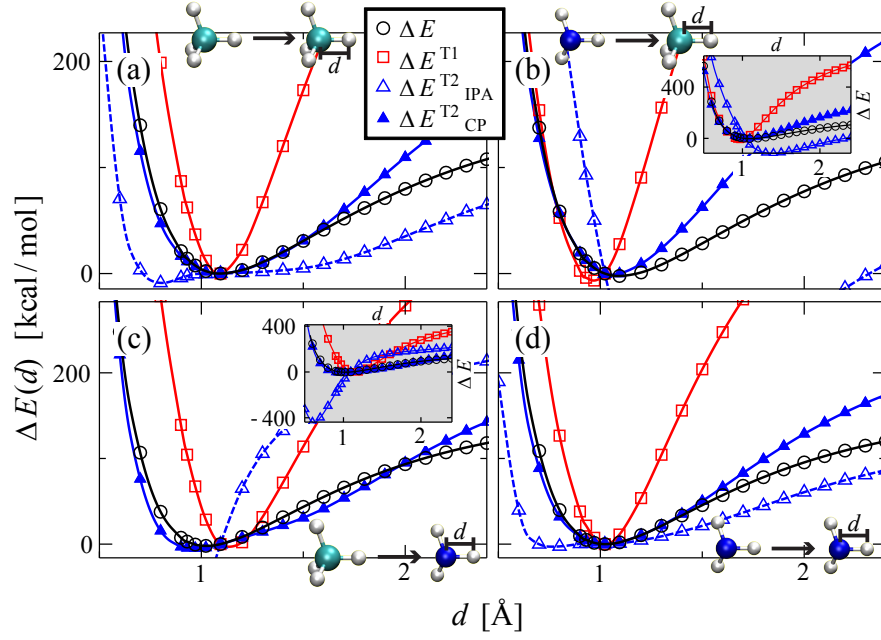


FIGURE 6.2: Energy difference ΔE , first order truncated Taylor series $\Delta E^{(1)}$, second order truncated Taylor series calculated by coupled perturbed $\Delta E^{(2)}_{\text{CP}}$, and second order truncated Taylor series calculated by independent particle approximation $\Delta E^{(2)}_{\text{IPA}}$ are plotted as black circles, red squares, blue open triangles, and blue filled triangles respectively. Coupling Hamiltonians are arranged as follow: (a) $\text{CH}_4 \rightarrow \text{CH}_4$, (b) $\text{NH}_3 \rightarrow \text{CH}_4$, (c) $\text{CH}_4 \rightarrow \text{NH}_3$, and (d) $\text{NH}_3 \rightarrow \text{NH}_3$. Insets of (b) and (c) show the zoom-out energy scale for overall landscape.

not available for systems with so many electrons, we have relied on approximative second order expressions IPA and CP (see above).

Fig. 6.2 presents estimated R-H covalent bond potentials of CH_4 and NH_3 predicted by a single SCF, using optimized CH_4 ((a) and (c)) or NH_3 ((b) and (d)) geometry as reference system. The second order derivative is calculated within IPA as well as within CP. When the chemical composition of H_R is the same as H_T , Fig. 6.2(a) and (d), $\Delta E^{(1)}$ is an upper bound, *i.e.* it always overshoots due to the concave behaviour of ΔE as a function of λ , demonstrated by Fig. 6.1(b). It also fails to capture the changes in equilibrium bond length when the chemical compositions of H_R differ from H_T , as can be seen in Fig. 6.2(b) and (c).

$\Delta E^{(2)}_{\text{IPA}}$ predicts a saddle point in Fig. 6.2(a) and (d), instead of a minimum at optimized geometry. When the chemical compositions of H_R and H_T are different, $\Delta E^{(2)}_{\text{IPA}}$ gives a large error as shown in the energy zoom out in the insets of Fig. 6.2(b) and (c). Considering $\mathbf{J}_{ia,jb}$ and $\mathbf{X}_{ia,jb}$ provides reasonable energy estimates, with $\Delta E^{(2)}_{\text{CP}}$ outperforming $\Delta E^{(2)}_{\text{IPA}}$. The superior performance of $\Delta E^{(2)}_{\text{CP}}$ over $\Delta E^{(2)}_{\text{IPA}}$ indicates that the contributions of Coulomb and xc energy due to density response are crucial. In other words, matrix elements $\mathbf{J}_{ia,jb}$ and $\mathbf{X}_{ia,jb}$ in Eq. (6.5) are not negligible for alchemical perturbations, especially for varying chemical composition.

TABLE 6.1: Prediction errors using first and second order based alchemical estimates for non-vertical changes in ten electron systems. Deviation of predicted energy minimum from actual ΔE_{eq} [kcal/mol], corresponding bond length deviation Δd_{eq} [Å], and vibration frequency $\Delta \omega_{\text{eq}}$. Numerical results from first order $\Delta E^{(1)}$, second order with independent particle approximation $\Delta E_{\text{IPA}}^{(2)}$, and second order with coupled perturbed $\Delta E_{\text{CP}}^{(2)}$ truncated Taylor series are presented. Average is calculated for each row in Avg column.

$$\Delta E^{(1)}:$$

H_{R}		CH ₄	NH ₃	H ₂ O	HF	Avg
CH ₄	ΔE_{eq}	-0.22	-2.30	-9.63	-21.99	-8.54
	Δd_{eq}	-0.01	0.12	0.22	0.30	0.15
	$\Delta \omega_{\text{eq}}$	6061.	3545.	2641.	1874.	3530.
NH ₃	ΔE_{eq}	-4.33	-0.04	-4.34	-14.01	-5.68
	Δd_{eq}	-0.13	0.002	0.10	0.19	0.04
	$\Delta \omega_{\text{eq}}$	4218.	4268.	4622.	2449.	3889.
H ₂ O	ΔE_{eq}	-32.78	-5.97	-0.0000	-4.88	-10.91
	Δd_{eq}	-0.27	-0.11	0.0005	0.09	-0.07
	$\Delta \omega_{\text{eq}}$	4360.	3933.	3886.	2568.	3687.
HF	ΔE_{eq}	-120.8	-38.63	-7.95	0.0003	-41.85
	Δd_{eq}	-0.44	-0.25	-0.10	-0.0008	-0.20
	$\Delta \omega_{\text{eq}}$	5772.	4047.	3982.	3560.	4340.

$$\Delta E_{\text{IPA}}^{(2)}:$$

CH ₄	ΔE_{eq}	-8.99	-437.5	-935.2	-1356.	-684.4
	Δd_{eq}	-0.31	-0.41	-0.40	-0.39	-0.38
	$\Delta \omega_{\text{eq}}$	2207.	9625.	12610.	14600.	9761.
NH ₃	ΔE_{eq}	-113.4	-4.10	-673.2	-1394.	-546.1
	Δd_{eq}	0.29	-0.27	-0.45	-0.43	-0.21
	$\Delta \omega_{\text{eq}}$	470.9	1471.	11200.	12660.	6452.
H ₂ O	ΔE_{eq}	-198.7	-90.18	-0.007	-809.4	-274.6
	Δd_{eq}	0.16	0.21	0.007	-0.43	-0.01
	$\Delta \omega_{\text{eq}}$	2949.	1124.	-1484.	9157.	2937.
HF	ΔE_{eq}	-212.6	-140.4	-64.95	0.002	-104.5
	Δd_{eq}	0.05	0.11	0.15	-0.001	0.08
	$\Delta \omega_{\text{eq}}$	6035.	4110.	1975.	-1177.	2736.

$$\Delta E_{\text{CP}}^{(2)}:$$

CH ₄	ΔE_{eq}	-0.39	-2.74	-19.70	-41.12	-15.99
	Δd_{eq}	-0.02	-0.07	-0.10	-0.10	-0.07
	$\Delta \omega_{\text{eq}}$	1663.	-495.7	2116.	3584.	1717.
NH ₃	ΔE_{eq}	0.90	-0.04	-1.84	-15.17	-4.04
	Δd_{eq}	-0.04	0.007	-0.06	-0.10	-0.05
	$\Delta \omega_{\text{eq}}$	2133.	140.4	-1480.	1690.	620.9
H ₂ O	ΔE_{eq}	5.78	0.77	0.0000	-1.51	1.26
	Δd_{eq}	-0.08	-0.02	0.002	-0.04	-0.03
	$\Delta \omega_{\text{eq}}$	2207.	-75.3	-194.2	-345.3	398.0
HF	ΔE_{eq}	13.12	4.25	0.65	0.001	4.50
	Δd_{eq}	-0.14	-0.07	-0.02	-0.0006	-0.06
	$\Delta \omega_{\text{eq}}$	3501.	2272.	1074.	-247.5	1650.

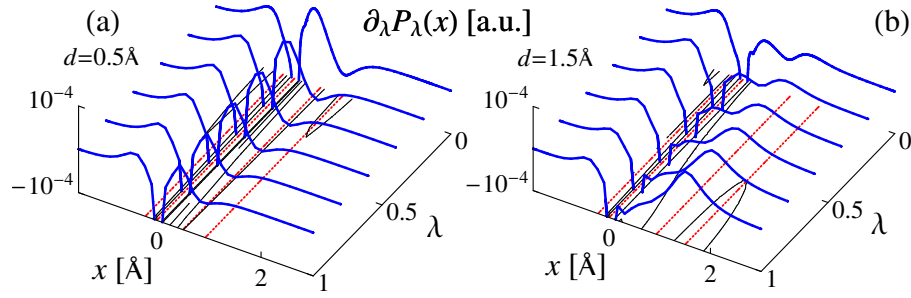


FIGURE 6.3: $\partial_\lambda P_\lambda(x)$ is calculated by finite difference $\partial_\lambda P_\lambda(x) \approx \frac{P_{\Delta\lambda}(x) - P_{\lambda=0}(x)}{\Delta\lambda}$. $\partial_\lambda P_\lambda(x)$, of $\text{HF} \rightarrow \text{H}_2\text{O}$ at (a) $d = 0.5 \text{ \AA}$ for compression and at (b) $d = 1.5 \text{ \AA}$ for extension are plotted as a function of x and λ . Nuclear positions are highlighted by red dashed lines, with $\text{F} \rightarrow \text{O}$ at $x = 0 \text{ \AA}$, $\text{H} \rightarrow \text{void}$ at $x = 0.93 \text{ \AA}$, two $\text{void} \rightarrow \text{H}$ at $x = -0.22 \text{ \AA}$ and $x = d$, where void denotes the nuclei with zero charge.

Different performances for compressing bonds $d < d_{\text{eq}}$, and stretching bonds $d > d_{\text{eq}}$, are found. $\Delta E_{\text{CP}}^{(2)}$ performs better in the region $0.5 \text{ \AA} \leq d \leq 1.5 \text{ \AA}$, except for Fig. 6.2(b) where $\Delta E_{\text{CP}}^{(2)}$ over shoot for $d > d_{\text{eq}}$. Another systematic asymmetric feature of $\Delta E^{(m)}$ is also observed *i.e.* $\Delta E_{\text{CP}}^{(2)}$ for $\text{H}_{\text{CH}_4} \rightarrow \text{H}_{\text{NH}_3}$ in Fig. 6.2(c) performs better than $\text{H}_{\text{NH}_3} \rightarrow \text{H}_{\text{CH}_4}$ in Fig. 6.2(b). When a growing nuclear charge is placed in a region of low density, it requires a larger nuclear charge to deform the density of reference system. By contrast the density easily adapts to a growing charge when it is placed in a region of high density. Lack of sensitivity results in nonlinear behaviour. Note that dramatic changes of electronic structure, as demonstrated for H_2^+ in Sec. 6.3, are not presented for many-electron systems.[74, 154] Since the performance of $\Delta E_{\text{CP}}^{(2)}$ is determined by how the density rearranges when λ is changing. One expects a near-constant $\partial_\lambda \rho$ through out the alchemical path for minimal higher order contributions. This is illustrated for the integrated density response of the alchemical path $\text{HF} \rightarrow \text{H}_2\text{O}$ in Fig. 6.3. Consistent behaviour is also observed for other alchemical paths of compressing vs stretching bond. $\partial_\lambda P_\lambda(x)$ varies less when λ changes from zero to one for $d = 0.5 \text{ \AA}$ Fig. 6.3(a), when compared with $d = 1.5 \text{ \AA}$ Fig. 6.3(b). A near constant $\partial_\lambda P_\lambda(x)$ at $d = 0.5 \text{ \AA}$ results in better performance to $\Delta E_{\text{CP}}^{(2)}$.

Table. 6.1 summarizes the results for all 4×4 combinations of $H_{\text{R}} \rightarrow H_{\text{T}}$, where PESs of H_{T} : $\{\text{CH}_4, \text{NH}_3, \text{H}_2\text{O}, \text{HF}\}$ are predicted by optimized geometry of H_{R} : $\{\text{CH}_4, \text{NH}_3, \text{H}_2\text{O}, \text{HF}\}$. Consistent trends of $\Delta E^{(1)}$, $\Delta E_{\text{IPA}}^{(2)}$, and $\Delta E_{\text{CP}}^{(2)}$ can be observed for other combinations of $H_{\text{R}} \rightarrow H_{\text{T}}$. Best performance of $\Delta E^{(1)}$, $\Delta E_{\text{IPT}}^{(2)}$, and $\Delta E_{\text{CP}}^{(2)}$ is observed when the chemical composition of H_{R} is the same as H_{T} , where the alchemical perturbation is minimal. When the chemical composition of H_{T} differs from H_{R} , a Coulomb potential is placed on the heavy atom to mutate it from H_{R} to H_{T} . Since the density peaks at the heavy atom, this Coulomb potential strongly perturbs the density. Worse performances are therefore expected for heavy atom perturbations. The importance of Coulomb and xc energy contribution to density response for alchemical

perturbation is further confirmed by the superior performance of CP over IPA, in terms of averaged signed error for ΔE_{eq} , Δd_{eq} , and $\Delta\omega$.

Interestingly, $\Delta E^{(1)}$ is competitively accurate with respect to $\Delta E_{\text{CP}}^{(2)}$ —at the energy minimum and when CH_4 is used as reference compound: $\Delta E^{(1)}$ gives Avg=-8.54 kcal/mol, while $\Delta E_{\text{CP}}^{(2)}$ gives -15.99 kcal/mol. This could be due to the more delocalized CH_4 density, when compared to NH_3 , H_2O , or HF . The delocalization results in smaller nonlinear effects in density response when alchemical perturbation is applied.

6.4.1 Computation details

Numerical results for non-vertical iso-electronic alchemical changes have been obtained using atom centered basis-sets. Restricted open-shell Hartree-Fock calculations have been carried out using Cartesian aug-cc-pVTZ basis set[113] for H_2^+ (Discussed in Sec. 6.3). Eq. (6.3) and higher order derivatives are evaluated analytically by Gaussian expansion of MOs. Reference geometry is first relaxed by **Gaussian09**[112] and the converged MO coefficients are extracted to evaluate orbital integrals. **NWChem**[155] is used to scan ΔE as a function of λ in Fig. 6.1(d) along alchemical path with discretization $\Delta\lambda = 0.01$. It is done by reassigning nuclear charges in the system.

Non-vertical alchemical changes in 10-electron molecules (discussed in Sec. 6.4) have been calculated using the uncontracted Cartesian Def2TZVP basis set[156]. Uncontracted neon basis is used for second row heavy atoms. Additional hydrogen basis functions are placed along the stretching pathway, from $d = 0.5 \text{ \AA}$ to $d = 3.0 \text{ \AA}$ in increments $\Delta d = 0.1 \text{ \AA}$. All systems with integer nuclear charges have been calculated using **Gaussian09**[112] while systems with fractional nuclear charges have been calculated using **NWChem**[155] with discretization $\Delta\lambda = 0.01$. For each $0 \leq \lambda \leq 1$, the atomic density for SCF initial guess iterates through {C, N, O, F, Ne} to ensure convergence. In all **Gaussian** and **NWChem** calculations we used Cartesian/Real spherical harmonic basis functions.

6.5 Conclusion

We have found oscillating behaviour in predictions of truncated Taylor series when varying the order in the non-vertical alchemical stretching of H_2^+ . The crossing of eigenvalue surfaces is due to the necessity to be symmetric at $\lambda = 0$ and $\lambda = 1$. This leads to a diverging series Taylor series, yet the second order correction could still provide fair predictions. The behavior of first and second order truncated alchemical Taylor series expansions in chemical space has been analyzed for molecules with many electrons. Numerical evidence of the superior performance of $\Delta E_{\text{CP}}^{(2)}$ over $\Delta E_{\text{IPA}}^{(2)}$ suggests that the response of Coulomb and xc energy to alchemical perturbation are crucial. Since the magnitude of higher order terms is determined by the variation of $\partial_\lambda \rho$ when changing

λ , a careful choice of target state could lead to better predictive power of truncated Taylor series. This indicates alchemical coupling via minimal density changes could lead to perturbative estimates with better predictive power.

Chapter 7

Conclusions

The applicability of alchemical derivatives to the exploration of CCS has been investigated in this thesis.

Some historical notes and recent efforts are discussed in chapter 3. Among other results, a sufficient condition for the linearizing alchemical path can be formulated as Eq. (3.11). It is practical because for alchemical perturbations it is possible to carefully design the coupling Hamiltonians such that they satisfy this condition.

Alchemical derivatives used to construct truncated Taylor series for estimating covalent bond energies in small molecules within fixed geometries are presented in chapter 4. Three significant conclusions can be drawn this study. Firstly, an universal correlation between error minimizing reference bond length and the target’s equilibrium bond length is found in Eq. (4.6). This relation holds across different elements. And it is also applicable to σ - as well as to π -bonding. Secondly, if electron density information from higher level of theory is numerically available, alchemical estimates can be made with chemical accuracy at a fraction of computational cost, as shown in Fig. 4.3. Thirdly, accurate first order predictions are possible if interpolating elements share the same position, and have similar core radii.

In chapter 5, alchemical estimates are applied to solid-state systems. Qualitative trends of alanates hydrogen absorption energy and vacancy energy ordering is achieved. Quantitative predictions are made for band structure as well as density of states of III-V and IV-IV semiconductors. It is found that if the integrated density differences between the reference and the target system is smaller than 0.25 a.u./per atom, high accuracy of the predicted band structure is guaranteed. With this insight, band structure optimization of explicit disordered $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals has been carried out using a hybrid gradient/genetic algorithm. The maximal direct band gap of 2.1 eV is achieved by $\text{Al}_{0.67}\text{Ga}_{0.33}\text{As}$, as shown in Fig. 5.6

Alchemical perturbation involving changes in geometry is explored in chapter 6. The significance of electron-electron interaction in second order alchemical estimates

has been studied. Our results imply that it is necessary to take the response of electron-electron interaction into consideration. If included, as within in coupled perturbed Kohn-Sham approach, a descent estimate can be achieved. Moreover, a finite radius of convergence, due to swapped eigenvalue ordering, is found for one-electron system of H_2^+ . This demonstrates that the choice of end points for alchemical paths can have a severe impact on the convergence behavior of Taylor series based estimates in CCS.

Appendix A

Self-consistent field

Solving Eq. (2.19) with the information of $\{\{Z_I\}, \{\mathbf{R}_I\}, N\}$ is in principle done by the following procedures:

1. At step $s = 0$, construct an starting density $\rho^{(s)}$ from an initial guess of $\{\phi_i^{(s)}\}$.
2. Solve the effective Schrödinger equation

$$\left(\frac{-\nabla^2}{2} + v_{eff}^{(s)}(\mathbf{r})\right)\phi_i^{(s+1)}(\mathbf{r}) = \varepsilon_i^{(s+1)}\phi_i^{(s+1)}(\mathbf{r}), \quad (\text{A.1})$$

with the effective potential

$$v_{eff}^{(s)}(\mathbf{r}) = v_{ext}(\mathbf{r}) + v_{xc}^{(s)}[\rho^{(s)}](\mathbf{r}) + \int d\mathbf{r}' \frac{\rho^{(s)}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \text{ where } \rho^{(s)} = \sum_{i=1}^{\text{occ}} |\phi_i^{(s)}(\mathbf{r})|^2.$$

3. Check consistency threshold, for example by $\sum_i \int d\mathbf{r} |\phi_i^{(s+1)}(\mathbf{r}) - \phi_i^{(s)}(\mathbf{r})|$ or by energy difference where energy is evaluated by Eq. (2.8).
4. If the new solution $\{\phi_i^{(s+1)}\}$ not yet consistent with old one $\{\phi_i^{(s)}\}$, update

$$\{\phi_i^{(s)}\} \mapsto \{\phi_i^{(s+1)}\}$$

and go back to procedure 2.

In practice, orbitals $\{\phi_i\}$ are expanded by a finite set of basis functions. These basis function can be a linear combination of 3D Gaussian functions, planewaves, a set of real space grid points, wavelet functions, or even some combinations of the basis sets mentioned above. Implementation of Gaussian basis set and planewaves are briefly summarized in Appendix B.

Appendix B

Implementation

B.1 Gaussian basis set

Gaussian basis sets are among the most intuitive ways to discretize the Schrödinger equation (Eq. A.1). They provide a set of predefined electron-independent “atomic orbitals” as a linear combination of some 3D Gaussian functions for each atom type

$$g_{I,\mathbf{l}}(x, y, z) = \sum_k c_{I,k} p_{I,l_x}(x) p_{I,l_y}(y) p_{I,l_z}(z) \exp^{-\alpha_{I,k}(x^2+y^2+z^2)}, \quad (\text{B.1})$$

where $p_{I,l}$ is the primitive polynomials $(1, x, x^2, x^3, \dots)$ of order l for atom I , $\mathbf{l} = \{l_x, l_y, l_z\}$ is the set of angular quantum number of each orbital, c_k is the contraction coefficient, α_k is the exponential factor of each Gaussian function, and $x^2 + y^2 + z^2 = |\mathbf{r}|^2$, \mathbf{r} being the distance to the atom’s position. Notice that the orthogonality is imposed by the linear combination factor c_k instead of the primitive polynomials for efficient Gaussian integral evaluations.

The wavefunction of the m^{th} molecular orbital can be expanded as

$$\phi_i(\mathbf{r}) = \sum_{I,\mathbf{l}} c_{I,\mathbf{l}}^{(m)} g_{I,\mathbf{l}}(\mathbf{r}), \quad (\text{B.2})$$

where $c_{I,\mathbf{l}}^{(m)}$ is the molecular orbital coefficient, which describes how the atomic orbital recombine to molecular orbital (or eigenfunction of the Hamiltonian). Since the Gaussian product can be integrated analytically,[19] each component of the Hamiltonian operator in Eq. (2.8) can be easily evaluated. The matrix elements in Eq. (2.19) is therefore easy to evaluate.

B.2 Planewave basis set

Planewaves are a convenient basis functions for periodic systems. The computation starts with a periodically repeating supercell, described by Bravais lattice vector $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ containing N electrons. These vectors determines the volume Ω , in which the wavefunction will be expanded by planewaves. A planewave basis function of this finite volume is

$$f_{\mathbf{G}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \exp(i\mathbf{G} \cdot \mathbf{r}) \equiv |\mathbf{G}\rangle, \quad (\text{B.3})$$

indexed by \mathbf{G} such that the orthonormality condition, $\langle \mathbf{G} | \mathbf{G}' \rangle = \delta_{\mathbf{G}\mathbf{G}'}$, is satisfied. Notice that $\{\mathbf{G}\}$ is a discrete set of vectors due to the periodic boundary condition on a finite volume boundary. The m^{th} molecular orbital m is expanded by

$$\phi_m(\mathbf{r}) = \sum_{\mathbf{G}} c_{m\mathbf{G}} |\mathbf{G}\rangle \quad (\text{B.4})$$

Due to the finite size of the Bravais lattice vector, not every \mathbf{G} satisfies the periodic condition

$$f_{\mathbf{G}}(\mathbf{r} + \mathbf{L}) = f_{\mathbf{G}}(\mathbf{r}), \quad (\text{B.5})$$

where $\mathbf{L} = l_1\mathbf{a}_1 + l_2\mathbf{a}_2 + l_3\mathbf{a}_3$, $\forall \{l_1, l_2, l_3\} \in \mathbb{Z}$ has the same periodicity as the lattice. In other words, only a discrete selection of \mathbf{G} is possible.

For convenience, the reciprocal lattice vector is defined as $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, such that $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$. The \mathbf{G} 's that satisfy Eq. (B.5) are integer multiples of reciprocal lattice vectors such that

$$\mathbf{G} \cdot \mathbf{L} = (g_1\mathbf{b}_1 + g_2\mathbf{b}_2 + g_3\mathbf{b}_3) \cdot \mathbf{L} = 2\pi(g_1l_1 + g_2l_2 + g_3l_3), \quad (\text{B.6})$$

where $\mathbf{G} = g_1\mathbf{b}_1 + g_2\mathbf{b}_2 + g_3\mathbf{b}_3$, $\forall \{g_1, g_2, g_3\} \in \mathbb{Z}$.

B.2.1 Bloch wavefunction

Under the notion of Slater determinant, the N -particle wavefunction can be approximated by a set of orbitals $\{\phi_m\}$ and eigenvalues $\{\varepsilon_m\}$, which solve the effective one-particle Schrödinger equation (Hartree-Fock or Kohn-Sham equation)

$$\left(-\frac{\nabla^2}{2} + v_{eff}(\mathbf{r}) \right) \phi_m(\mathbf{r}) = \varepsilon_m \phi_m(\mathbf{r}). \quad (\text{B.7})$$

If the effective potential satisfies the periodic condition *i.e.* $v_{eff}(\mathbf{r} + \mathbf{L}) = v_{eff}(\mathbf{r})$, Bloch theorem showed that the solution of one-particle Schrödinger equation must satisfy

$$\begin{aligned}\phi_{m\mathbf{k}}(\mathbf{r}) &= \exp(i\mathbf{k} \cdot \mathbf{r})u_{m\mathbf{k}}(\mathbf{r}) \\ &= \exp(i\mathbf{k} \cdot \mathbf{r})\frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{m\mathbf{G}}(\mathbf{k}) \exp(i\mathbf{G} \cdot \mathbf{r}) \\ &= \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{m\mathbf{G}}(\mathbf{k}) |\mathbf{G} + \mathbf{k}\rangle,\end{aligned}\tag{B.8}$$

where $c_{m\mathbf{G}}(\mathbf{k})$ is the \mathbf{k} -dependent expansion coefficient of $u_{m\mathbf{k}}(\mathbf{r})$ and \mathbf{k} is some given wave vector in the reciprocal space. With this condition, the effective Schrödinger equation can be rewritten in terms of $u_{i\mathbf{k}}$

$$\left(\frac{-1}{2}(\nabla + i\mathbf{k})^2 + v_{eff}(\mathbf{r})\right)u_{m\mathbf{k}}(\mathbf{r}) \equiv H_{\mathbf{k}}u_{m\mathbf{k}}(\mathbf{r}) = \varepsilon_{m\mathbf{k}}u_{m\mathbf{k}}(\mathbf{r}).\tag{B.9}$$

Notice that the momentum of electron in state $\phi_{m\mathbf{k}}$ is

$$\langle \phi_{m\mathbf{k}} | -i\nabla | \phi_{m\mathbf{k}} \rangle = \mathbf{k} + \langle u_{m\mathbf{k}} | -i\nabla | u_{m\mathbf{k}} \rangle.\tag{B.10}$$

Therefore the momentum of electron in state $\phi_{m\mathbf{k}}$ and $\phi_{m\mathbf{k}'}$ is different $\mathbf{k} \neq \mathbf{k}'$. An momentum transfer is required for the transition $\phi_{m\mathbf{k}} \rightarrow \phi_{m\mathbf{k}'}$.

The final results are the average over all possible \mathbf{k} values. In other words, instead of infinitely many electrons, it is possible to consider only N electrons in the supercell but averaging over infinitely many \mathbf{k} -points. Specifically, the electron density is expressed as

$$\begin{aligned}\rho(\mathbf{r}) &= \frac{1}{\Omega} \int_{\Omega} d\mathbf{k} \sum_m \phi_{m\mathbf{k}}^*(\mathbf{r}) \phi_{m\mathbf{k}}(\mathbf{r}) \\ &= \frac{1}{\Omega} \int_{\Omega} d\mathbf{k} \sum_m \sum_{\mathbf{G}, \mathbf{G}'} c_{m\mathbf{G}}(\mathbf{k})^* c_{m\mathbf{G}'}(\mathbf{k}) \exp(i(\mathbf{G} - \mathbf{G}') \cdot \mathbf{r}).\end{aligned}\tag{B.11}$$

B.2.2 Hamiltonian matrix element

If $v_{eff}(\mathbf{r})$ is known, the left hand side of Eq. (B.7) can be rewritten for each \mathbf{k} -vector as a matrix equation

$$\sum_{\mathbf{G}} \langle \mathbf{G}' + \mathbf{k} | (-\nabla^2 + v_{eff}(\mathbf{r})) | \mathbf{G} + \mathbf{k} \rangle c_{m\mathbf{G}}(\mathbf{k}) = \varepsilon_m(\mathbf{k}) c_{m\mathbf{G}'}(\mathbf{k}),\tag{B.12}$$

where subscript m loops over the electronic states in the system. Notice that $\varepsilon_m(\mathbf{k})$ gives rise to the electronic band structure for each electron at \mathbf{k} inside the first Brillouin

zone. The matrix element is

$$\langle \mathbf{G}' + \mathbf{k} | (-\nabla^2 + v_{eff}(\mathbf{r})) | \mathbf{G} + \mathbf{k} \rangle = \frac{1}{2} |\mathbf{G} + \mathbf{k}|^2 \delta_{\mathbf{G}'\mathbf{G}} + V_{eff}(\mathbf{G}) \delta_{\mathbf{G}\Delta\mathbf{G}}, \quad (\text{B.13})$$

where $V_{eff}(\mathbf{G})$ is the Fourier coefficient of $v_{eff}(\mathbf{r})$ at \mathbf{G} and $\Delta\mathbf{G} = \mathbf{G}' - \mathbf{G}$.

In practice, the large sum of reciprocal vector \mathbf{G} is truncated by a cutoff value, satisfying $\frac{1}{2}|\mathbf{G}|^2 < E_{cut}$. And the \mathbf{k} averaging is done with a finite set of \mathbf{k} vectors with corresponding weights $w_{\mathbf{k}}$. That is

$$\begin{aligned} \sum_{\mathbf{G}} &\mapsto \sum_{\mathbf{G}}^{E_{cut}} \\ \frac{1}{\Omega} \int_{\Omega} d\mathbf{k} &\mapsto \sum w_{\mathbf{k}}. \end{aligned} \quad (\text{B.14})$$

Moreover, Eq. (B.12) can be transformed into a set of predefined atomic wavefunction basis, which can dramatically reduce the matrix size.

Notice that the Coulomb integral in $v_{eff}(\mathbf{r})$ is computed via density instead of wavefunctions. This trick reduces the computational cost of the Coulomb integral from P^4 to P^2 , where P is the number of planewaves within a cutoff condition. For this reason, the exact exchange is very expensive under planewave implementation.

B.3 Density functional evaluation

One of the standard ways to calculate the density functional value $E_{xc}[\rho(\mathbf{r})]$ for a given density $\rho(\mathbf{r})$ at \mathbf{r} is to use LibXC[23] library, which provides standard C/Fortran API. Local functionals can be evaluated numerically for each \mathbf{r} . The density functional evaluation is therefore decoupled from the underlying implementation of the basis function.

Gaussian basis function Becke integration grid[157] can be used for numerical evaluations. It is essentially a spherical grid system centered at each atom with a weighting factor to guarantee correct integrated results. For Gaussian basis set, there is a open-source `Python` interface implemented in the `horton` code[158] for Becke integration grid generation.

Plane wave basis For planewave basis in periodic systems, on the contrary of using Becke integration grid, the real space grid points are determined by the energy cutoff value and the supercell lattice vector.

Appendix C

Pseudopotentials

Most of the chemical reactions can be accounted for by the changes in valence electron distribution. On the contrary, core electrons are typically inert to changes of the chemical environment. For this reason, it is common to replace core electrons by predefined core potentials, or pseudopotentials (PPs), to reduce computational cost. Moreover, the orbitals of valence electrons are more smooth at the core region which further reduces the number of required basis functions. The basic formulation is briefly discussed in this section.

Let $|\phi_i^{\text{AE}}\rangle$ and $|\phi_i^{\text{PP}}\rangle$ be an orbital of valence electron indexed by i with eigenvalue ε_i for all electron and PP Hamiltonian respectively. They are connected by

$$|\phi_i^{\text{AE}}\rangle = |\phi_i^{\text{PP}}\rangle + \sum_j^{\text{core}} a_j |\chi_j\rangle, \quad (\text{C.1})$$

where $|\chi_j\rangle$ is the j^{th} core state with eigenvalue ε_j . Notice that $|\phi_i^{\text{AE}}\rangle$ and $|\chi_j\rangle$ are both eigenfunctions of all electron Hamiltonian. Due to the orthogonality between eigenfunctions, we have

$$\langle \chi_k | \phi_i^{\text{AE}} \rangle = \langle \chi_k | \phi_i^{\text{PP}} \rangle + \sum_j^{\text{core}} a_j \langle \chi_k | \chi_j \rangle = 0. \quad (\text{C.2})$$

The all electron valence state can therefore be rewritten as

$$|\phi_i^{\text{AE}}\rangle = |\phi_i^{\text{PP}}\rangle - \sum_j^{\text{core}} \langle \chi_j | \phi_i^{\text{PP}} \rangle |\chi_j\rangle, \quad (\text{C.3})$$

which gives exactly the same orthogonality. Replace Eq. (C.3) in to all electron Schrödinger equation $H|\phi_i^{\text{AE}}\rangle = \varepsilon_i|\phi_i^{\text{AE}}\rangle$ we get

$$\begin{aligned} H\left(|\phi_i^{\text{PP}}\rangle - \sum_j^{\text{core}} \langle\chi_j|\phi_i^{\text{PP}}\rangle|\chi_j\rangle\right) &= H|\phi_i^{\text{PP}}\rangle - \sum_j^{\text{core}} \varepsilon_j|\chi_j\rangle\langle\chi_j|\phi_i^{\text{PP}}\rangle \\ &= \varepsilon_i\left(|\phi_i^{\text{PP}}\rangle - \sum_j^{\text{core}} \langle\chi_j|\phi_i^{\text{PP}}\rangle|\chi_j\rangle\right). \end{aligned} \quad (\text{C.4})$$

This implies the eigenvalue equation

$$H|\phi_i^{\text{PP}}\rangle + \left(\sum_j (\varepsilon_i - \varepsilon_j)|\chi_j\rangle\langle\chi_j|\right)|\phi_i^{\text{PP}}\rangle = \varepsilon_i|\phi_i^{\text{PP}}\rangle, \quad (\text{C.5})$$

where the PP operator can be defined as $V_{\text{PP}}(\varepsilon) = \sum_j (\varepsilon - \varepsilon_j)|\chi_j\rangle\langle\chi_j|$, which is also know as the projector.

In practice, PPs are parametrized for each atom type, combine with the shielded Coulomb attraction potential for valence electron. Core states are expanded by spherical harmonics to mimic the wavefunction of core electrons. The projector can be implemented as a nonlocal potential $\sum_j |\chi_j\rangle\langle\chi_j| = V_{\text{nl}}(\mathbf{r}, \mathbf{r}')$ where the integral gives

$$\langle\phi_i|\left(\sum_j |\chi_j\rangle\langle\chi_j|\right)|\phi_i\rangle = \int d\mathbf{r}d\mathbf{r}' \phi_i^*(\mathbf{r}) V_{\text{nl}}(\mathbf{r}, \mathbf{r}') \phi_i(\mathbf{r}'). \quad (\text{C.6})$$

Goedecker PPs are used throughout this thesis. It is in the form of

$$V_{\text{loc}}(r) + V_{\text{nl}}(\mathbf{r}, \mathbf{r}'), \quad (\text{C.7})$$

where V_{loc} is the local screened Coulomb potential for valence electrons and V_{nl} is the nonlocal potential consists of projector. The local part is of the form

$$V_{\text{loc}}(r) = -\frac{Z_{\text{eff}}}{r} \text{erf}\left(\frac{r}{\sqrt{2}r_{\text{loc}}}\right) + \sum_{i=1}^4 C_i \left(\frac{r}{r_{\text{loc}}}\right)^{2i-2} \exp\left(-\left(\frac{r}{\sqrt{2}r_{\text{loc}}}\right)\right), \quad (\text{C.8})$$

with parameters $\{r_{\text{loc}}, C_1, C_2, C_3, C_4\}$ defined for each atom type. The nonlocal part is of the form

$$V_{\text{nl}}(\mathbf{r}, \mathbf{r}') = \sum_{lm} \sum_{ij} \langle\mathbf{r}|p_i^{lm}\rangle h_{ij}^l \langle p_j^{lm}|\mathbf{r}'\rangle, \quad (\text{C.9})$$

where the projector is defined as $\langle\mathbf{r}|p_i^{lm}\rangle = N_i^l Y_{lm}\left(\frac{\mathbf{r}}{|\mathbf{r}|}\right) r^{l+2i-2} \exp\left(-\frac{r^2}{2r_i^2}\right)$ with parameters $\{h_{ij}^l, r_l\}$ defined for each atom type. Specific parametrizations can be found in Ref. [109].

Bibliography

- [1] S. Sakai, T. Saga, M. Takeyasu, and M. Umeno. Room-temperature laser operation of AlGaAs/GaAs double heterostructure fabricated on Si substrates by metalorganic chemical vapor deposition. *Appl. Phys. Lett.*, 48:413, 1986.
- [2] J. R. Gong, D. Dung, N. A. EI-Masry, and S. M. Bedair. Atomic layer epitaxy of AlGaAs. *Appl. Phys. Lett.*, 57:400, 1990.
- [3] A. P. Malshe, K. P. Rajurkar, K. R. Virvwani, C. R. Taylor, D. L. Bourell, G. Levy, M. M. Sundaram, J. A. McGeough, V. Kalyanasundaram, and A. N. Samant. Tip-based nanomanufacturing by electrical chemical mechanical and thermal processes. *CIRP ANN-MANUF TECHN*, 59:628, 2010.
- [4] I. G. Rau, S. Baumann, S. Rusponi, F. Donati, S. Stepanow, L. Gragnaniello, J. Dreiser, C. Piamonteze, F. Nolting, S. Gangopadhyay, O. R. Albertini, R. M. Macfarlane, C. P. Lutz, B. A. Jones, P. Gambardella, A. J. Heinrich, and H. Brune. Reaching the magnetic anisotropy limit of a 3D metal atom. *Science*, 344:6187, 2014.
- [5] D. S. Engstrom, B. Porter, M. Pacios, and H. Bhaskaran. Additive nanomanufacturing - a review. *J. Mater. Res.*, 29:1792, 2014.
- [6] C. Kuhn and D. N. Beratan. Inverse strategies for molecular design. *J. Phys. Chem.*, 100:10596, 1996.
- [7] S. R. Marder, D. N. Beratan, and L. T. Ceng. Approaches for optimizing the first electronic hyperpolarizability of conjugated organic molecules. *Science*, 252:103, 1991.
- [8] A. Franceschetti and A. Zunger. The inverse band-structure problem of finding an atomic configuration with given electronic properties. *Nature*, 402:60, 1999.
- [9] O. A. von Lilienfeld, R. D. Lins, and U. Rothlisberger. Variational particle number approach for rational compound design. *Phys. Rev. Lett.*, 95:153002, 2005.
- [10] T. Weymuth and M. Reiher. Inverse quantum chemistry: Concepts and strategies for rational compound design. *Int. J. Quantum Chem.*, 114:823, 2014.

- [11] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28:1049, 1926.
- [12] P. A. M. Dirac. Quantum mechanics of many-electron systems. *Proc. Math. Phys. Eng. Sci.*, 123:714, 1929.
- [13] E. B. Wilson Jr. Four-dimensional electron density function. *J. Phys. Chem.*, 36: 2232, 1962.
- [14] R. G. Parr and W. Yang. Density-functional theory of the electronic structure of molecules. 46:701, 1995.
- [15] O. A. von Lilienfeld and M. E. Tuckerman. Molecular grand-canonical ensemble density functional theory and exploration of chemical space. *J. Chem. Phys.*, 125: 154104, 2006.
- [16] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, USA, 1994.
- [17] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293, 1929.
- [18] David Griffiths. *Introduction to Quantum Mechanics*. 2nd edition edition, 2005.
- [19] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry*. Dover Publication, INC., 2nd edition, 1996.
- [20] M. Thomas. The calculation of atomic fields. *Proc. Camb. Phil. Soc.*, 23:542, 1927.
- [21] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.
- [22] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140, 1965.
- [23] M. A. L. Marques, M. J. T. Oliveira, and T. Burnus. Libxc: A library of exchange and correlation functionals for density functional theory. *Comput. Phys. Comm.*, 183:2272, 2012.
- [24] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.
- [25] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.
- [26] C. Adamo, M. Cossi, G. Scalmani, and V. Barone. Accurate static polarizabilities by density functional theory: Assessment of the PBE0 model. *Chem. Phys. Lett.*, 307:265, 1999.

- [27] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.*, 118:8207, 2003.
- [28] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.*, 125:224106, 2006.
- [29] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618, 1934.
- [30] R. P. Feynman. Forces in molecules. *Phys. Rev.*, 56:340, 1939.
- [31] O. A. von Lilienfeld. Accurate *ab initio* energy gradients in chemical compound space. *J. Chem. Phys.*, 131:164102, 2009.
- [32] K. Y. S. Chang and O. A. von Lilienfeld. Quantum mechanical treatment of variable molecular composition: From ‘alchemical’ changes of state function to rational compound design. *CHIMIA*, 68:602, 2014.
- [33] W. Yang, A. J. Cohen, F. De Proft, and P. Geerlings. Analytical evaluation of Fukui function and real space linear response function. *J. Chem. Phys.*, 136:144110, 2012.
- [34] W. G. van Gunsteren and H. J. C. Berendsen. Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput. Aided Mol. Des.*, 1:171, 1987.
- [35] T.P. Straatsma and J. A. McCammon. Computational alchemy. *Annu. Rev. Phys. Chem.*, 43:407, 1992.
- [36] P. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.*, 93:2395, 1993.
- [37] N. Marzari, S. de Gironcoli, and S. Baroni. Structure and phase stability of $\text{Ga}_x\text{In}_{1-x}\text{P}$ solid solutions from computational alchemy. *Phys. Rev. Lett.*, 72:4001, 1994.
- [38] W. L. Jorgensen and C. Ravimohan. Monte Carlo simulation of differences in free energies of hydration. *J. Chem. Phys.*, 83:3050, 1985.
- [39] D. C. Langreth and J. P. Perdew. Theory of nonuniform electronic systems. I. analysis of the gradient approximation and a generalization on that works. *Phys. Rev. B*, 21:5469, 1980.
- [40] J. Harris. Adiabatic-connection approach to Kohn-Sham theory. *Phys. Rev. A*, 29:1648, 1984.

- [41] J. V. Barth, G. Costantini, and K. Kern. Engineering atomic and molecular nanostructures at surfaces. *Nature*, 437:671, 2005.
- [42] Y. Xia, Y. Xiong, B. Lim, and S. E. Skrabalak. Shape-controlled synthesis of metal nanocrystals: Simple chemistry meets complex physics? *Angew. Chem. Int. Ed.*, 48:60, 2009.
- [43] O. A. von Lilienfeld. First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties. *Int. J. Quantum Chem.*, 113:1676, 2013.
- [44] D. Sheppard, G. Henkelman, and O. A. von Lilienfeld. Alchemical derivatives of reaction energetics. *J. Chem. Phys.*, 133:084104, 2010.
- [45] S. Goedecker. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Phys. Chem.*, 120:9911, 2004.
- [46] M. E. Tuckerman. *Statistical mechanics: Theory and molecular simulation*. Oxford University Press, 2010.
- [47] W. L. Jorgensen. The many roles of computation in drug discovery. *Science*, 303:1813, 2004.
- [48] C. Oostenbrink and W. F. van Gunsteren. Free energies of ligand binding for structurally diverse compounds. *Proc. Natl. Acad. Sci. USA*, 102:6750, 2005.
- [49] S. Riniker, C. D. Christ, H. S. Hansen, P. H. Hünenberger, C. Oostenbrink, D. Steiner, and W. F. van Gunsteren. Calculation of relative free energies for ligand-protein binding, solvation, and conformational transitions using the GRO-MOS software. *J. Phys. Chem. B*, 115:13570, 2011.
- [50] D. Jiao, K. Leung, S. B. Rempe, and T. M. Nenoff. First principles calculations of atomic Nickel redox potentials and dimerization free energies: A study of metal nanoparticle growth. *J. Chem. Theory Comput.*, 7:485, 2011.
- [51] P. E. Smith and W. G. van Gunsteren. When are free energy components meaningful? *J. Phys. Chem.*, 98:13735, 1994.
- [52] K. Leung, S. B. Rempe, and O. A. von Lilienfeld. Ab initio molecular dynamics calculations of ion hydration free energies. *J. Chem. Phys.*, 130:204507, 2009.
- [53] C. Oostenbrink. Efficient free energy calculations on small molecule host-guest systems—a combined linear interaction energy/one-step perturbation approach. *J. Comp. Chem.*, 30:212, 2009.

- [54] S. Jayaraman and E. J. Maginn. Computing the melting point and thermodynamic stability of the orthorhombic and monoclinic crystalline polymorphs of the ionic liquid 1-*n*-butyl-3-methylimidazolium chloride. *J. Chem. Phys.*, 127:214504, 2007.
- [55] S. Jayaraman, A. P. Thompson, O. A. von Lilienfeld, and E. J. Maginn. Molecular simulation of the thermal and transport properties of three alkali nitrate salts. 49:559, 2010.
- [56] S. Jayaraman, A. P. Thompson, and O. A. von Lilienfeld. Molten salt eutectics from atomistic simulations. *Phys. Rev. E*, 84:030201–1, 2011.
- [57] A. Pérez and O. A. von Lilienfeld. Path integral computation of quantum free energy differences due to alchemical transformations involving mass and potential. *J. Chem. Theory Comput.*, 7:2358, 2011.
- [58] D. Alfé, M. J. Gillan, and G. D. Price. Constraints on the composition of the Earth’s core from *ab initio* calculations. *Nature*, 405:172, 2000.
- [59] S. Adachi. GaAs, AlAs, and $\text{Al}_x\text{Ga}_{1-x}\text{As}$: Material parameters for use in research and device applications. *J. Appl. Phys.*, 58:R1, 1985.
- [60] T. Mueller and G. Ceder. Bayesian approach to cluster expansions. *Phys. Rev. B*, 80:024103, 2009.
- [61] L. Bellaiche and D. Vanderbilt. Virtual crystal approximation revisited: Application to dielectric and piezoelectric properties of perovskites. *Phys. Rev. B*, 61:7877, 2000.
- [62] D. G. Pettifor. Electron theory in materials modeling. *Acta Mater.*, 51:5649, 2003.
- [63] J. Harris and R. O. Jones. The surface energy of bounded electron gas. *J. Phys. F: Metal Phys.*, 4:1170, 1974.
- [64] O. Gunnarson and B. I. Lundqvist. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B*, 13:4274, 1976.
- [65] W. Koch and M. C. Holthausen. *A Chemist’s Guide to Density Functional Theory*. WILEY-VCH, Germany, 2nd edition, 2002.
- [66] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048, 1981.
- [67] M. Levy. Density-functional exchange correlation through coordinate scaling in adiabatic connection and correlation hole. *Phys. Rev. A*, 43:4637, 1991.

- [68] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671, 1992.
- [69] M. Seidl, J. P. Perdew, and M. Levy. Strictly correlated electrons in density-functional theory. *Phys. Rev. A*, 59:51, 1999.
- [70] P. Politzer and R. G. Parr. Some new energy formulas for atoms and molecules. *J. Phys. Chem.*, 61:4258, 1974.
- [71] M. Okrusch, R. Müller, and A. Hese. High-resolution ultraviolet laser spectroscopy on jet-cooled benzene molecules: Ground and excited electronic state polarizabilities determined from static Stark effect measurements. *J. Chem. Phys.*, 110:10393, 1999.
- [72] W. Xue and M. A. Ratner. Microscopic study of electrical transport through individual molecules with metallic contacts. I. Band lineup, voltage drop, and high-field transport. *Phys. Rev. B*, 68:115406–1, 2003.
- [73] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Insights into current limitations of density functional theory. *Science*, 321:792, 2008.
- [74] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Challenges for density functional theory. *Chem. Rev.*, 112:289, 2012.
- [75] W. G. Baber and H. R. Hassé. The two center problem in wave mechanics. *Proc. Camb. Phil. Soc.*, 31:564, 1935.
- [76] H. Høegreave. On the stability of the one electron bond. *J. Chem. Phys.*, 98:5579, 1993.
- [77] A. J. Cohen and P. Mori-Sánchez. Dramatic changes in electronic structure revealed by fractional charged nuclei. *J. Chem. Phys.*, 140:044110, 2014.
- [78] C. Hermann, J. Neugebauer, and M. Reiher. Finding a needle in a haystack: Direct determination of vibrational signatures in complex systems. *New J. Chem.*, 31: 818, 2007.
- [79] M. Wang, X. Hu, D. N. Beratan, and W. Yang. Designing molecules by optimizing potentials. *J. Am. Chem. Soc.*, 128:3228, 2006.
- [80] S. Keinan, X. Hu, D. N. Beratan, and W. Yang. Designing molecules with optimal properties using the linear combination of atomic potentials approach in an AM1 semiempirical framework. *J. Phys. Chem.*, 111:176, 2007.

- [81] J. H. Werner, S. Kolodinski, and H. J. Queisser. Novel optimization principles and efficiency limits for semiconductor solar cells. *Phys. Rev. Lett.*, 72:3851, 1994.
- [82] A. Silverman, Alex Zunger, R. Kalish, and Joan Adler. Atomic-scale structure of disordered $\text{Ga}_{1-x}\text{In}_x$. *Phys. Rev. B*, 51:10795, 1995.
- [83] M. Jansen. A concept for synthesis planning in solid-state chemistry. *Angew. Chem. Int. Ed.*, 41:3746, 2002.
- [84] W. Yang. Dynamic linear response of many-electron systems: An integral formulation of density-functional theory. 38:5512, 1988.
- [85] S. Baroni, S. de Gironcoli, A. D. Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515–562, 2001.
- [86] A. Putrino, D. Sebastiani, and M. Parrinello. Generalized variational density functional perturbation theory. *J. Chem. Phys.*, 113:7102, 2000.
- [87] a_l, a_r, b_l, b_r are linear coefficients [eV] for predicting E_b as $= aE_b^{T1} + b$. Subscripts l and r respectively denote the repulsive wall and attractive tail of the potential. For the four panels in Fig. (3.2) the parameters are (in same sequence): (a) 0.569, 0.361, -0.137, -0.203; (b) 1.058E-2, 1.473E-4, -7.036E-3, -7.043E-3; (c) 1.069, 0.326, 0.020, -0.197; (d) 5.805E-4, 1.375E-4, -6.920E-3, -0.00692256.
- [88] A. Scherrer, V. Verschinin, and D. Sebastiani. Eigensystem representation of the electronic susceptibility tensor for intermolecular interactions within density functional theory. *J. Chem. Theory Comput.*, 8:106–111, 2012.
- [89] A. C. Ihrig, A. Scherrer, and D. Sebastiani. Electronic density response to molecular geometric changes from explicit electronic susceptibility calculations. *J. Chem. Phys.*, 139:094102, 2013.
- [90] N. Sablon, F. de Proft, P. W. Ayers, and P. Geerlings. Computing second-order functional derivatives with respect to the external potential. *J. Chem. Theory Comput.*, 6:3671, 2000.
- [91] S. Fias, Z. Biosdenghien, T. Stuyver, M. Audiffred, G. Merino, P. Geerlings, and F. de Proft. Analysis of aromaticity in planar metal systems using the linear response kernel. *J. Phys. Chem. A*, 117:3556, 2013.
- [92] P. Geerlings, S. Fias, Z. Boisdenghen, and F. De Proft. Conceptual DFT: Chemistry from the linear response function. *Chem. Soc. Rev.*, 43:4989, 2014.

- [93] D. Benoit, D. Sebastiani, and M. Parrinello. Accurate total energies without self-consistency. *Phys. Rev. Lett.*, 87:226401, 2001.
- [94] Z. Cournia, L. Leng, S. Gandavadi, X. Du, R. Bucala, and W. L. Jorgensen. Discovery of human macrophage migration inhibitory factor (MIF)-CD74 antagonists via virtual screening. *J. Medic. Chem.*, 52:416, 2009.
- [95] W. L. Jorgensen. Efficient drug lead discovery and optimization. *Acc. Chem. Res.*, 42:724, 2009.
- [96] M. E. M. Noble, J. A. Endicott, and L. N. Johnson. Protein kinase inhibitors: Insights into drug design from structure. *Science*, 303:1800, 2004.
- [97] R. Car and M. Parrinello. Unified approach for molecular dynamics and density functional theory. *Phys. Rev. Lett.*, 55:2471, 1985.
- [98] V. Marcon, O. A. von Lilienfeld, and D. Andrienko. Tuning electronic eigenvalues of benzene via doping. *J. Chem. Phys.*, 127:064305, 2007.
- [99] B. Widom. Some topics in the theory of fluids. *J. Chem. Phys.*, 39(11):2808 – 2812, 1963.
- [100] M. D. Macedonia and E. J. Maginn. A biased grand canonical Monte Carlo method for simulating adsorption using all-atom and branched united atom models. *Mol. Phys.*, 96:1375, 1999.
- [101] M. D. Macedonia and E. J. Maginn. Pure and binary component sorption equilibria of light hydrocarbons in the zeolite silicalite from grand canonical Monte Carlo simulations. *Fluid Phase Equilibria*, 158-160:19, 1999.
- [102] L. Kleinman and D. M. Bylander. Efficient form for model pseudopotentials. *Phys. Rev. Lett.*, 48:1425, 1982.
- [103] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani. Optimization of effective atom centered potentials for London dispersion forces in density functional theory. *Phys. Rev. Lett.*, 93:153004, 2004.
- [104] O. A. von Lilienfeld and P. A. Schultz. Structure and band gaps of Ga-(V) semiconductors: The challenge of Ga pseudopotentials. *Phys. Rev. B*, 77:115202, 2008.
- [105] O. A. von Lilienfeld. Force correcting atom centred potentials for generalised gradient approximated density functional theory: Approaching hybrid functional accuracy for geometries and harmonic frequencies in small chlorofluorocarbons. *Molecular Physics*, 111(14-15):2147–2153, 2013.

- [106] Wanda Andreoni and Alessandro Curioni. New advances in chemistry and materials science with CPMD and parallel computing. *Parallel Computing*, 26:819–842, 2000.
- [107] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, 54:1703, 1996.
- [108] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B*, 58:3641, 1998.
- [109] M. Krack. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chim. Acta*, 114:145, 2005.
- [110] J. Harris. Simplified method for calculating the energy of weakly interacting fragments. *Phys. Rev. B*, 31:1770, 1985.
- [111] J. Harris. Tight-binding models and density-functional theory. *Phys. Rev. B*, 39:12520, 1989.
- [112] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- [113] T. M. Dunning Jr. Gaussian basis sets for use in correlated molecular calculations. I. the atoms Boron through Neon and Hydrogen. *J. Chem. Phys.*, 90:1007, 1989.
- [114] F. Capasso. Band-gap engineering: From physics and materials to new semiconductor devices. *Science*, 235:172, 1987.
- [115] M. C. Toroker and E. A. Carter. Transition metal oxide alloys as potential solar energy conversion materials. *J. Mater. Chem. A*, 1:2474, 2013.
- [116] P. Kirkpatrick and C. Ellis. Chemical space. *Nature*, 432:823, 2004.

- [117] L. Ruddigkeit, R. van Deursen, L. C. Blum, and J. L. Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.*, 52:2864, 2012.
- [118] F. De Vleeschouwer, P. Geerlings, and F. De Proft. Molecular property optimization with boundary conditions through the best first search scheme. *Comput. Phys. Comm.*, 17:1391, 2016.
- [119] R. Balawender, M. A. Welearegay, M. Lesiuk, F. De Proft, and P. Geerlings. Exploring chemical space with the alchemical derivatives. *J. Chem. Theory Comput.*, 9:5327, 2013.
- [120] A. Solovyeva and O. A. von Lilienfeld. Alchemical screening of ionic crystals. *Phys. Chem. Chem. Phys.*, 18:31078, 2016.
- [121] K. Y. S. Chang, S. Fias, R. Ramakrishnan, and O. A. von Lilienfeld. Fast and accurate predictions of covalent bond in chemical space. *J. Chem. Phys.*, 144:174100, 2016.
- [122] D. L. Smith and C. Mailhot. Theory of semiconductor superlattice electronic structure. *Rev. Mod. Phys.*, 62:173, 1990.
- [123] J. Liqiang, Q. Yichun, W. Baiqi, L. Shudan, J. Baojiang, Y. Libin, F. Wei, F. Hunggang, and S. Jiazhong. Review of photoluminescence performance of nanosized semiconductor materials and its relationships with photocatalytic activity. *Sol. Energ. Mat. Sol. Cells*, 90:1773, 2006.
- [124] N. López, L. A. Reichertz, K. M. Yu, and W. Walukiewicz. Engineering the electronic band structure for multiband solar cells. *Phys. Rev. Lett.*, 106:028701, 2011.
- [125] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:5188, 1976.
- [126] O. Rubel, A. Bokhanchuk, S. J. Ahmed, and E. Assmann. Unfolding the band structure of disordered solids: From bound states to high-mobility Kane-fermions. *Phys. Rev. B*, 90:115202, 2014.
- [127] V. Popescu and A. Zunger. Effective band structure of random alloys. *Phys. Rev. Lett.*, 104:236403, 2010.
- [128] S. H. Wei, L. G. Ferreira, J. E. Bernard, and A. Zunger. Electronic properties of random alloys: Special quasirandom structures. *Phys. Rev. B*, 42:9622, 1990.

- [129] J. Heyd and G. E. Scuseria. Efficient hybrid density functional calculations in solids: Assessment of the Heyd-Scuseria-Ernzerhof screened Coulomb hybrid functional. *J. Chem. Phys.*, 121:1187, 2004.
- [130] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *J. Chem. Phys.*, 123:174101, 2005.
- [131] E. R. Batista, J. Heyd, R. G. Hennig, B. P. Uberuaga, R. L. Martin, and G. E. Scuseria. Comparison of screened hybrid density functional theory to diffusion Monte Carlo in calculations of total energies of silicon phases and defects. *Phys. Rev. B*, 74:131102, 2006.
- [132] J. W. Nicklas and J. W. Wilkins. Accurate ab initio prediction of III-V direct-indirect band gap crossovers. *Appl. Phys. Lett.*, 97:091902, 2010.
- [133] O. A. Vydrov, J. Heyd, A. V. Krukau, and G. E. Scuseria. Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals. *J. Chem. Phys.*, 125:074106, 2006.
- [134] M. Levinshtein, S. Rumyantsev, and M. Shur. *Handbook series on Semiconductor Parameters*. World Scientific, 1st edition, 1999.
- [135] A. Tonkikh, A. Klavsyuk, N. Zakharov, A. Saletsky, and P. Werner. SnSi nanocrystals of zinc blende structure in a Si matrix. *Nano Res.*, 8:3905, 2015.
- [136] N. Bhargava, M. Coppinger, J. P. Gupta, L. Wielunski, and J. Kolodzey. Lattice constant and substitutional composition of GeSn alloy grown by molecular beam epitaxy. *Appl. Phys. Lett.*, 103:041908, 2013.
- [137] M. M. Rieger and P. Vogl. Electronic-band parameters in strained $\text{Si}_{1-x}\text{Ge}_x$ alloys on $\text{Si}_{1-y}\text{Ge}_y$ substrates. *Phys. Rev. B*, 48:14276, 1993.
- [138] X. Gonze, B. Amadon, P. M. Anglade, and J. M. Beuken. ABINIT: First-principle approach to material and nanosystem properties. *Comput. Phys. Comm.*, 180:2582, 2002.
- [139] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.
- [140] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758, 1999.
- [141] <https://github.com/SamKChang/qctoolkit.git>.
- [142] https://github.com/SamKChang/20170406_alchemy_optimization.git.

- [143] E. K. U. Gross and W. Kohn. Local density-functional theory of frequency dependent linear response. *Phys. Rev. Lett.*, 55:2850, 1986.
- [144] N. Sablon, F. De Proft, P. W. Ayers, and P. Geerlings. Computing second-order functional derivatives with respect to the external potential. *J. Chem. Theory Comput.*, 6:3671, 2010.
- [145] H. F. Wilson, F. Gygi, and G. Galli. Efficient iterative method for calculations of dielectric matrices. *Phys. Rev. B*, 78:113303, 2008.
- [146] N. Sablon, F. De Proft, and P. Geerlings. The linear response kernel: Inductive and resonance effects quantified. *J. Phys. Chem. Lett.*, 1:1227, 2010.
- [147] S. Fias, Z. Boisenghien, T. Stuyver, M. Audiffred, G. Merino, P. Geerlings, and F. De Proft. Analysis of aromaticity in planar metal systems using the linear response kernel. *J. Phys. Chem. A*, 117:3556, 2013.
- [148] S. L. Adler. Quantum theory of the dielectric constant in real solids. *Phys. Rev.*, 126(2):413, 1962.
- [149] N. Wiser. Dielectric constant with local field effects included. *Phys. Rev.*, 129(1):62, 1963.
- [150] D. Cremer. Møller-Plesset perturbation theory: From small molecule methods to methods for thousands of atoms. *Adv. Rev.*, 1:509, 2012.
- [151] A. V. Sergeev, D. Z. Goodson, S. E. Wheel, and W. D. Allen. On the nature of the Møller-Plesset critical point. *J. Chem. Phys.*, 123:064105, 2005.
- [152] M. L. Leininger, W. D. Allen, H. F. Schaefer III, and C. D. Sherrill. Is Møller-Plesset perturbation theory a convergent *ab initio* method? *J. Chem. Phys.*, 112:9213, 2000.
- [153] F. H. Stillinger. Møller-Plesset convergence issues in computational quantum chemistry. *J. Chem. Phys.*, 112:9711, 2000.
- [154] P. Mori-Sánchez, A. J. Cohen, and W. Yang. Localization and delocalization errors in density functional theory and implication for band-gap prediction. *Phys. Rev. Lett.*, 100:146401, 2008.
- [155] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, D. Wang, J. Nieplocha, T.L. Windus E. Apra, and W.A. de Jong. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Comm.*, 181:1477, 2010.

-
- [156] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297, 2005.
- [157] A. D. Becke. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.*, 88:2547, 1988.
- [158] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, K. Boguslawski, M. Chan, Y. Zhao, T. D. Kim, S. Vandenbrande, D. Yang, C. E. Gonzalez-Espinoza, S. Fias, P. A. Limacher, D. Berrocal, A. Malek, and P. W. Ayers, 2015. <http://theochem.github.com/horton/>.

Kuang-Yu Samuel Chang

AI/ML Software Engineer

Salmenweg 12

CH-4057, Basel

Switzerland

+41 (0)76 320 9932

✉ ky.sam.chang@gmail.com

GitHub: SamKChang

LinkedIn: SamKChang

Education

- 2013–2017 **Ph.D.**, *Computational Chemistry, University of Basel (unibas)*, Basel, Switzerland.
2008–2010 **M.Sc.**, *Material Physics, University of Linköping (LiU)*, Linköping, Sweden.
2005–2008 **B.Sc.**, *Mathematics and Physics, National Tsing Hua University (NTHU)*, Hsinchu, Taiwan.

Nanodegrees

- 2017 **Data analyst**, *Udacity*.
2017 **Deep learning foundation**, *Udacity*.
2017 **Artificial intelligence**, *Udacity*.

Working Experiences

- 2018–present **Software engineer**, *Google, Alphabet*, Zürich, Switzerland.
Focus on artificial intelligence and machine learning.
2017–2018 **Postdoctoral research fellow**, *Computational Chemistry, University of Basel*, Basel, Switzerland.
Apply deep learning/machine learning techniques to solve quantum mechanics.
2012–2013 **Research assistant**, *Institute of Molecular Biology, Academia Sinica*, Taipei, Taiwan.
Setup environment and perform rigid body protein docking theory prediction.
2010–2011 **Research assistant**, *Department of Biology, NTHU*, Hsinchu, Taiwan.
Setup experiment environment and carry out theory simulation for protein.

Projects

- 2015–present **qctoolkit**, *High throughput quantum chemistry framework*, C/Python.
Used for pipelining work flow on supercomputers (Mira, Piz Daint).
2016–present **CheML**, *Machine learning framework for Chemistry*, Python.

Languages

English **fluent**

Chinese **native**

Software Development Skills

- Languages C/C++, Fortran, Python, R, Perl, Javascript, UNIX scripting.
Involved in Horton (C++/Python), QuantumESPRESSO (FORTRAN90), CPMD (FORTRAN77).

Honors and Awards

- 2008–2009 Scholarship for Excellent Students to Study Overseas Internships, Ministry of education, Taiwan.

Bibliography

Original Publications

- 2018 **K. Y. S. Chang**, O. A. von Lilienfeld: $\text{Al}_x\text{Ga}_{1-x}\text{As}$ crystals with direct 2 eV band gaps from computational alchemy, *Phys. Rev. Mat.* 2, 073802.
2016 **K. Y. S. Chang**, S. Fias, R. Ramakrishnan, O. A. von Lilienfeld: Fast and accurate predictions of covalent bonds in chemical space, *J. Chem. Phys.* 144, 174119.
2014 **K. Y. S. Chang**, O. A. von Lilienfeld: Quantum mechanical treatment of variable molecular composition: From 'alchemical' changes of state functions to rational compound design, *CHEMIA*, 68, 602.

Oral Presentations

- 2017 American Physics Society (APS) March meeting, New Orleans, USA.
2016 Material Research Society (MRS) Fall meeting, Boston, USA.
2013 Institute for Pure and Applied Mathematics (IPAM) conference at UCLA on exploring chemical compound space, California, USA.