

# Ein systematischer, linear skalierender Fragmentansatz für das Elektronenstrukturproblem

**Dissertation**

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch–Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich–Wilhelms–Universität Bonn

vorgelegt von

Frederik Heber

aus

Freiburg im Breisgau

Bonn 2013

Angefertigt mit Genehmigung der Mathematisch–Naturwissenschaftlichen Fakultät der  
Rheinischen Friedrich–Wilhelms–Universität Bonn

1. Referent: Prof. Dr. Michael Griebel

2. Referent: Prof. Dr. Marc Alexander Schweitzer

Tag der Promotion: 12.02.2014

Erscheinungsjahr: 2014

## Zusammenfassung

Das Elektronenstrukturproblem bezeichnet eine Familie von Gleichungen, abgeleitet von der Schrödinger-Gleichung, welche das Verhalten von Elektronen im Potential der Atomkerne beschreiben. Zu ihr gehören die Hartree-Fock (HF)- oder auch die Kohn-Sham (KS)-Gleichung.

Wir werden in dieser Arbeit den Ansatz, der in der Chemie unter dem Begriff Fragmentierung oder auch Additivität bekannt ist, systematisch untersuchen. Dabei wird ein molekulares System mit chemischer Intuition in einzelne Teile, sogenannte Fragmente, zerlegt, um mittels bekannter Eigenschaften der Teile auf die Eigenschaften des Ganzen zu schließen. Größtenteils unverstanden ist, wieso dieser Ansatz funktioniert. In jüngster Zeit steht dieser Ansatz auch besonders im Fokus, da mit ihm leicht Näherungsverfahren möglich sind, deren Rechenkosten nur linear von der Anzahl der Atomkerne abhängen. Diese systematische Untersuchung liefert, ausgehend von der Schrödinger-Gleichung, das Verständnis, warum Fragmentansätze gute Näherungen liefern können.

Da die Schrödinger-Gleichung eine hochdimensionale Integro-Differential-Gleichung beschreibt, scheiterten außer bei sehr kleinen Systemen bisherige analytische und direkte numerische Lösungsversuche. Deswegen sind zwei Näherungsansätze geläufig: Entweder wird der zugehörige Operator modifiziert oder die gesuchte Lösung wird eingeschränkt. Wir leiten im Rahmen des zweiten Ansatzes mittels einer sogenannten Korrelationsentwicklung die elektronische Schrödinger-Gleichung und die Hartree-Fock-Gleichung her. Dabei schließt ein Dekorrelator pro Term der Entwicklung einen Anteil der Korrelation aus der gewünschten Lösung aus. Dieser fungiert als Gegenstück zum bekannten Anregungsoperator. Zum Zweck einer linear skalierenden Methode untersuchen wir die Ursachen der hohen Laufzeitkomplexität der Hartree-Fock-Methode. Dabei ist eine nicht-lineare Eigenwertgleichung über die Fockmatrix zu lösen, deren Aufstellung allgemein  $\mathcal{O}(n^4)$  Kosten mit der Anzahl an Basisfunktionen  $n$  verursacht. Wir definieren den Begriff der Lokalität mittels exponentiell abfallender Funktionen und zeigen ein Verfahren, um in linearer Zeit die Matrix zu negativen Eigenvektoren einer hermiteschen Matrix aus deren in gewissen Grenzen bekannter Dünnesetztheitsstruktur zu ermitteln. Dabei wird diese Strukturinformation am Ende heuristisch aus dem Bindungsgraphen des Systems gewonnen, der in alle verbundenen Teilgraphen bis zu einer vorgegebenen Menge an Kernen pro Teilgraph zerlegt wird, der Fragmentierung. Durch eine unter dem Namen ANOVA-Entwicklung bekannte Summation können wir elegant die Grundzustandsenergie des Ganzen aus den unabhängigen Systemen zu diesen Teilgraphen berechnen.

Wir haben die Methode im Rahmen dieser Arbeit vollständig und parallel implementiert, wobei kein chemisches Wissen seitens des Nutzers notwendig ist. Wir belegen ausführlich die sehr guten Approximations- und Skalierungseigenschaften. So liegt der Cross-over-Punkt bei 20 Kernen. Wir untersuchen die Berechnung der langreichweitigen Wechselwirkung mit einer hochskalierenden Mehrgittermethode und betrachten, wie fehlende Korrelationsenergie in Møller-Plesset 2. Ordnung ermittelt werden kann. Schließlich belegen wir die Möglichkeit der schnellen und genauen Berechnung großer Molekülsysteme mit bis zu 6000 Atomen.



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
1.1. Gliederung . . . . .	3
1.2. Notation . . . . .	5
<b>2. Tensorzerlegung</b>	<b>11</b>
2.1. Hochdimensionale Funktionen . . . . .	12
2.2. Tensoren . . . . .	16
2.3. Summen und Produkte von Funktionenräumen . . . . .	18
2.4. Bestapproximation . . . . .	22
2.5. Dekorrelatoren . . . . .	24
<b>3. Schrödinger-Gleichung</b>	<b>33</b>
3.1. Zeitabhängige Schrödinger-Gleichung . . . . .	33
3.2. Systeme von Kernen und Elektronen . . . . .	35
3.3. Klassische Moleküldynamik . . . . .	36
3.4. Elektronische Wellenfunktion . . . . .	39
<b>4. Hartree-Fock</b>	<b>45</b>
4.1. Closed-Shell Hartree-Fock . . . . .	46
4.2. Roothaan-Hall-Gleichung . . . . .	53
4.3. Komplexität . . . . .	58
<b>5. Lokalität</b>	<b>67</b>
5.1. Lokale Funktion . . . . .	67
5.2. Menge lokaler Funktionen . . . . .	69
5.3. Lokale Basis . . . . .	70
5.4. Schlussfolgerungen . . . . .	70
5.5. Lokalisierte räumlichen Orbitale . . . . .	74
<b>6. Dünnbesetzte, indefinite Matrizen</b>	<b>77</b>
6.1. Voraussetzungen . . . . .	77
6.2. Schwerpunkttheorem . . . . .	84
6.3. Störung . . . . .	94
6.4. Weiterführende Literatur . . . . .	96
<b>7. BOSSANOVA</b>	<b>99</b>
7.1. Dünnbesetzte Fockmatrix . . . . .	100

## Inhaltsverzeichnis

7.2. Dünnbesetzte Eigenvektoren . . . . .	103
7.3. Behandlung der langreichweitigen Wechselwirkung . . . . .	107
7.4. Gestörter Fockoperator . . . . .	110
7.5. Abhängigkeitsgraph und Fragmente . . . . .	116
7.6. Dichte-basierte HF-Methode . . . . .	118
<b>8. Korrelationsenergie</b>	<b>123</b>
8.1. Configuration Interaction . . . . .	123
8.2. Møller-Plesset Störtheorie . . . . .	126
8.3. Korrelationsentwicklung . . . . .	128
8.4. Verwandte Literatur . . . . .	129
<b>9. Experimente</b>	<b>133</b>
9.1. Vorüberlegungen . . . . .	134
9.2. Skalierungsverhalten der Implementierung . . . . .	139
9.3. Bindungsgraphbestimmung . . . . .	158
9.4. Approximationsverhalten . . . . .	163
9.5. Abschätzung der langreichweitigen Wechselwirkung . . . . .	185
9.6. Qualitative Untersuchungen . . . . .	198
<b>10. Schluss</b>	<b>209</b>
10.1. Zusammenfassung . . . . .	209
10.2. Ausblick . . . . .	212
<b>A. Appendix</b>	<b>215</b>
A.1. Bekannte Basissätze . . . . .	215
A.2. Bindungsarten . . . . .	217
A.3. Obere Schranke für Vier-Zentren-Integral bei lokaler Funktion . . . . .	220
A.4. Literatur . . . . .	227
A.5. Wasserstoff-Saturation . . . . .	233
<b>Abkürzungsverzeichnis</b>	<b>235</b>
<b>Index</b>	<b>236</b>
<b>Abbildungsverzeichnis</b>	<b>241</b>
<b>Literaturverzeichnis</b>	<b>244</b>

# 1. Einleitung

Das *Elektronenstrukturproblem* bezeichnet eine Familie von Gleichungen, abgeleitet von der Schrödingergleichung, welche das Verhalten von Elektronen im Potential der Atomkerne beschreiben. Zu ihr gehören die Hartree-Fock (HF)- oder auch die Kohn-Sham (KS)-Gleichung. Für unser alltägliches Leben sind Lösungen dieser Gleichungen von sehr grundsätzlichem Interesse. Egal ob es sich um stabilere, leichtere Materialien für Flugzeugrümpfe, um Reaktionsmechanismen in der technischen Synthese oder um das Bindungsverhalten eines Virus an die Wirtshülle handelt, es ist die Schrödingergleichung, die gelöst werden muss.

Die elektronische Schrödingergleichung,

$$\hat{H}^{(\text{el})}(\mathbf{R})\psi(\mathbf{r}; \mathbf{R}) = E^{(\text{el})}\psi(\mathbf{r}; \mathbf{R}),$$

ist eine  $3N$ -dimensionale Integro-Differential-Gleichung, die ein Eigenwertproblem beschreibt. Analytische Lösungen sind, neben einigen Modellsystemen, nur für das Wasserstoffatom bzw. für Wasserstoffartige Atome bekannt. Direkte numerische Behandlung dieser Gleichung, die also das Problem hochdimensionaler Integration überwinden muss, war bisher nur für sehr kleine Systeme erfolgreich, siehe Hamaekers [82]. Es existieren eine Reihe von Näherungen, die eine numerische Behandlung größerer Systeme möglich machen. Nach Defranceschi und Le Bris [43] lassen sich diese Näherungen in zwei Gruppen unterteilen. Bei der ersten Gruppe wird der Operator  $\hat{H}^{(\text{el})}$  modifiziert. Dazu stützt man sich auf die Existenzaussage des berühmten Theorems von Hohenberg und Kohn [95] für ein unbekanntes Funktional, das nur von der elektronischen Ladungsverteilung abhängt. Bei der zweiten Gruppe werden bestimmte Ansatzfunktionen benutzt, welche der Lösung  $\psi$  eine bestimmte Struktur vorgeben und die sich üblicherweise aus Produkten von Einteilchen-Wellenfunktionen zusammensetzen.

Aber auch die zentralen Repräsentanten dieser beiden Gruppen, die KS-Gleichung und die HF-Gleichung, lassen sich nicht für wirklich große Systeme in zeitlich erträglichem Aufwand lösen. Entwickelt man sie als Eigenwertprobleme in einer endlichen Basis, dann skaliert die benötigte Rechenzeit ca. mit  $n^3$  in der Anzahl an Basisfunktionen  $n$ .<sup>1</sup> Dabei steigt die Anzahl an Basisfunktionen proportional mit der Anzahl der Elektronen an.

*Linear skalierende Methoden*, bei denen die Rechenzeit linear mit der Anzahl der Elektronen anwächst, sind daher der Thema aktueller Forschung in der Quantenchemie und in der Physik kondensierter Materie. Dabei zielen die meisten Bemühungen in Richtung der

---

<sup>1</sup>Für eine genauere Untersuchung der Skalierungseigenschaften, zumindest von HF, sei auf Abschnitt 4.3.3 verwiesen.

## 1. Einleitung

einfacheren Beschreibung der Lösung mittels einer Ladungsverteilung, die auch im Rahmen der HF-Gleichung als dünnbesetzte Dichtematrix Eingang findet. Sie stützen sich auf die Aussage der Dünnbesetztheit in nicht-leitenden Materialien, siehe Benzi et al. [18], und suchen die Lösung über die Stationaritätsbedingung eines entsprechenden Funktionals. Diese Methoden sind eine Mischung aus den Ansätzen der beiden Gruppen, da die rein dichteabhängige Thomas-Fermi-Theorie bei der Bindungsvorhersage versagt und die nötige Diagonalisierung der HF-Gleichung linear skalierende Kosten verhindert.

Chemiker haben schon früh eine strukturelle Ähnlichkeit zwischen Molekül und Teilen des Moleküls erkannt. Daraus ergab sich die Idee der „Additivität“, dass also das Ganze aus der Summe seiner Teile folgert, siehe Benson und Buss [16]. Im Rahmen der numerischen Behandlung prägten u. a. Christoffersen und Maggiora [34] den Begriff des *Fragments* als Teil eines molekularen Systems. Diese große Methodenfamilie, welche den Ansatz verfolgt, ein molekulares System in Teile zu zerlegen und aus der Lösung der Teile die Lösung des Ganzen zu folgern, bezeichnen wir mit *Fragmentansätzen*. Uns ist dabei keine Arbeit bekannt, in der über eine Summe energetischer Beiträge hinaus erläutert wird, warum dieser Ansatz funktionieren sollte.

Wir wollen daher in dieser Arbeit den Ansatz der Fragmentierung des Systems aufgreifen und daraus systematisch einen allgemeinen Ansatz für linear skalierende Methoden für das Elektronenstrukturproblem entwickeln. Unser Ziel ist eine vollständige Herleitung dieser Fragmentansätze und die Klärung der Frage, warum diese seit 60 Jahren erfolgreich sind und dieser Erfolg bis in die empirischen Potentiale der Moleküldynamik ausstrahlt.

Wir entscheiden uns dabei für die oben besprochene zweite Gruppe, bei der zunächst die mögliche Lösung eingeschränkt, aber der Elektron-Hamiltonoperator in voller Form beibehalten wird. Dies hat folgenden Vorteil: Da die volle elektronische Schrödingergleichung direkter Behandlung unzugänglich ist, können wir so einen Umweg über eine zunächst eingeschränkte Lösung wählen, die aber anschließend dank des unveränderten Operators zur Lösung der vollen Gleichung erweitert werden kann. Bei der ersten Gruppe ist nicht ohne Weiteres klar, in welcher Beziehung die Lösung zum eingeschränkten Operator zu der Lösung des vollen Operators steht.

Die Lösung der hochdimensionalen elektronischen Schrödingergleichung ist selbst eine hochdimensionale Funktion. Über die Existenzaussage der Rang-1-Bestapproximation in Tensorprodukträumen definieren wir einen als Dekorrelator bezeichneten Projektor. Dieser schränkt die Lösung auf Terme aus niederdimensionalen Faktoren über Gruppen von unterscheidbaren oder ununterscheidbaren Eingangsvariablen ein. Effektiv schließt dieser einen Teil der Korrelation aus der zulässigen Lösung aus. Mittels einer Reihenentwicklung aus diesen Termen können wir anschließend zunächst die klassische Moleküldynamik aus der vollen Schrödingergleichung herleiten und weiter auch die HF-Gleichung aus der elektronischen Schrödingergleichung. Die Entwicklung in einer endlichen Basis überführt diese Gleichung in ein Eigenwertproblem mit der Fockmatrix. Über eine eingehende Komplexitätsanalyse dieses einfacheren, aber nicht-linearen Problems gelangen wir zum Konzept lokaler (Basis-)Funktionen. Außerdem beweisen wir eine schnelle Methode der



Diagonalisierung einer hermiteschen, indefiniten Matrix mit dünnbesetzten Eigenvektoren unter der Voraussetzung, dass uns die Struktur der Dünnbesetztheit in Form einer hierarchischen, überlappenden Zerlegung der Menge der lokalen Basisfunktionen vorliegt. Diese ergänzt das bekannte Interlacing-Theorem um Aussagen über den Schwerpunkt eines Teilspektrums der hermiteschen Matrix. Schließlich zeigen wir drei Varianten auf, wie die gesonderte Behandlung der langreichweitigen Wechselwirkung eingebracht werden kann.

Unser allgemeiner Zugang verhilft uns dabei, eine neue Methode vorzustellen und zu implementieren, die, im Gegensatz zu den uns bekannten Fragmentansätzen in der Literatur, gänzlich ohne chemische Intuition auskommt. Dazu fasst sie die Bindungsstruktur des molekularen Systems als Graph auf und fragmentiert diesen „blind“ in alle verbundenen Teilgraphen bis zu einer vorgegebenden Menge an Kernen pro Teilgraph. Dies generiert auch die für die schnelle Diagonalisierung nötige Zerlegung der Basisfunktionen. Wir belegen anhand vielfältiger numerischer Beispiele die lineare Komplexität der Methode und ihre sehr guten Parallelisierungs- und Approximationseigenschaften. Durch den Fragmentansatz erreichen wir eine um eine Größenordnung niedrigere Skalierungskonstante als andere linear skalierende Methoden. Das bedeutet, dass bereits ab 20 Kernen die von uns vorgestellte Methode bei ähnlicher Genauigkeit schneller rechnet als ein Standard-HF-Löser. Wir betrachten außerdem eine Erweiterung auf die Berechnung der Korrelationsenergie und beziehen den dispersiven Anteil der langreichweitigen Wechselwirkung ein. Abschließend betrachten wir große Moleküle mit bis zu  $10^4$  Kernen, deren Grundzustandsenergie bereits auf kleinen Commodity Clustern aktueller Leistungsfähigkeit in unter einer Stunde berechnet werden kann.

## 1.1. Gliederung

Die Idee, ein großes Problem in eine Menge kleinerer Probleme zu unterteilen, ist nicht neu: Divide et impera. Im Rahmen der Schrödingergleichung existiert ein Corpus an Literatur von mehr als 60 Jahren, die sich mit dieser Idee auseinandersetzen. Kaum eine Arbeit jedoch versucht zu erläutern, warum der Ansatz grundlegend überhaupt funktionieren kann. Meistens wird eine Formel präsentiert, die die Zerlegung des Problems motiviert und diese wird anschließend durch numerische Ergebnisse untermauert. Weiterhin existieren viele chemisch motivierte „Kniffe“.

Wir werden in dieser Arbeit den Ansatz, der in Chemie unter dem Begriff Fragmentierung oder auch Additivität bekannt ist, ausgehend von der Schrödingergleichung grundlegend motivieren und für den Fall einer Implementation mit einem HF-Löser vollständig darstellen. Unser Ziel ist dabei nicht, einen abstrakten Zugang zu der Idee darzustellen, sondern den gewählten Weg der Implementation mit seinen chemischen Eigenheiten detailliert zu erklären.

Diese Arbeit gliedert sich dann wie folgt: Die Schrödingergleichung beschreibt ein hochdimensionales Eigenwertproblem im Hilbertraum. Die gesuchte, hochdimensionale Lösung soll in einem endlich-dimensionalen Ansatzraum mit Hilfe von Tensoren dargestellt werden, worauf wir in Kapitel 2 eingehen. Wir beschreiben diese Lösung mittels

## 1. Einleitung

einer Reihenentwicklung, die auf Produkten von potentiell niederdimensionalen Funktionen über Gruppen von unterscheidbaren oder ununterscheidbaren Eingangsvariablen basiert. Die Projektion aus dem vollen Raum in die zugehörigen Unterräume erfolgt über einen formal definierten Dekorrelator. Durch Abschneiden der Korrelationsentwicklung wird die zulässige Lösung des Eigenwertproblems eingeschränkt.

Dies wird im folgenden Kapitel 3 benutzt, um verschiedene gängige Vereinfachungen der vollen Schrödingergleichung herzuleiten, darunter auch die BO-Approximation und die wichtige HF-Gleichung. Letztere stellt eine nicht-lineare, aber niedrig-dimensionale Eigenwertgleichung mit dem Fockoperator dar. Die Entwicklung in einer endlichen Basis überführt diese in ein algebraisches Eigenwertproblem mit der Fockmatrix. Mit der Lösung dieses Problems beschäftigen wir uns eingehend in Kapitel 4, wo wir die Ursachen für deren hohe Kosten von  $\mathcal{O}(n^4)$  in einer zunächst beliebigen Basis untersuchen.

Für die Lösungen der HF-Gleichung [105] sind bestimmte exponentielle Abfalleigenschaften bekannt. Wir definieren dazu in Kapitel 5 zunächst lokale Funktionen, wobei wir das Konzept des kompakten Trägers auf die endliche Maschinengenauigkeit des Rechners übertragen, indem wir exponentiellen Abfall um ein kompaktes Gebiet fordern. Die Menge dieser lokalen Funktionen ist dabei abgeschlossen unter Multiplikation, außerdem gehören zu ihnen die sogenannten Chemikerbasen. Eine Basis aus solchen lokalen Funktionen hat dann die notwendigen Eigenschaften, um das Aufstellen der Fockmatrix auf  $\mathcal{O}(n^2 \log n^2)$  zu vereinfachen.

Eine zweite Komplexitätshürde besteht in der notwendigen Diagonalisierung dieser hermiteschen Matrix, um negative Eigenwerte und -vektoren zu ermitteln. Unter der Voraussetzung, dass uns die Struktur der Dünnbesetztheit einer beliebigen hermiteschen Matrix in Form einer hierarchischen, überlappenden Zerlegung der Menge der Basisfunktionen vorliegt, zeigen wir in Kapitel 6, dass sich die Matrix der Eigenvektoren zu negativen Eigenwerten in linearer Zeit aus Lösungen zu Hauptuntermatrizen rekonstruieren lässt.

Dabei wurde die langreichweitige Wechselwirkung im Fockoperator zunächst vernachlässigt, damit auch das Aufstellen der Fockmatrix nur noch Kosten von  $\mathcal{O}(n)$  verursacht. Unter Auslassung einer eingehenden Betrachtung des nicht-linearen Zyklus erhalten wir also ein Verfahren mit einer abschließenden Komplexität von  $\mathcal{O}(n)$ , welches wir in Kapitel 7 zusammenfassend darlegen. Die langreichweitige Wechselwirkung kann jedoch bekanntermaßen mittels der Schnellen Multipol-Methode (FMM) oder, wie hier, mittels eines Mehrgitter-Ansatzes für das zugrundeliegende Poisson-Problem ebenfalls in linearer Zeit berechnet werden.

Dann schließen wir in Kapitel 8 den Kreis zur formalen Korrelationsentwicklung und legen im Rahmen der Møller-Plesset Störtheorie einen Ansatz dar, um die der HF-Gleichung fehlende Korrelationsenergie zu berechnen. Dabei identifizieren wir den in Kapitel 2 eingeführten Dekorrelator als Gegenstück zum bekannten Anregungsoperator.

Abschließend untersuchen wir in Kapitel 9 ausführlich die sehr guten Skalierungs- und Approximationseigenschaften der parallelen Implementation. Dabei belegen wir numerisch, dass ein heuristisch ermittelter Bindungsgraph sich sehr gut als notwendiger Abhängigkeitsgraph für das erwähnte Diagonalisierungsverfahren eignet. Im letzten Kapitel 10 geben wir eine Resümee und außerdem einen Ausblick auf zukünftig mögliche Entwicklungen.

Einige eher nebensächliche, aber der Vollständigkeit dienliche Teile sind in den Anhang [A](#) ausgelagert. So beschreiben wir die Notation der Chemikerbasen in [Abschnitt A.1](#), die chemischen Bindungen in [Abschnitt A.2](#), einige Nebenrechnungen in [A.3](#) und den (nicht vollständigen) Corpus der Literatur der Fragmentansätze in [Abschnitt A.4](#). Außerdem gehen wir auf die Wassersaturation im Wechselspiel mit der langreichweitigen Behandlung in [Abschnitt A.5](#) ein.

## 1.2. Notation

Wir haben uns dieser Arbeit bemüht, Symbole eindeutig zu machen. So sollte bspw.  $\hat{H}$  stets das Symbol für den Hamiltonoperator sein. Aufgrund der Endlichkeit der Alphabete bei gleichzeitiger Vielfältigkeit möglicher Symbole ist eine völlig strikte Trennung dabei aber nicht immer möglich. Dabei wird der jeweilige Buchstabe, wie geschehen, durch  $\tilde{\wedge}$  und andere Anhängsel modifiziert oder in einer bestimmten Schriftart gedruckt werden, um ihn in eine spezifischere Form zu bringen.

Im Allgemeinen schreiben wir diese spezifischen Formen wie folgt, hier exemplarisch an dem Symbol  $A$  aufgezeigt:

$a, A$  Skalar

$\hat{A}$  Operator

$\mathbf{A}, \mathbf{A}$  Vektor, Matrix (endlich-dimensional)

$\mathcal{A}$  Menge, Raum

Dabei bezeichnet  $\mathbf{a}_i$  einen, nämlich den  $i$ -ten, unter vielen möglichen Vektoren und  $a_i$  die  $i$ -te Komponente zu einem Vektor  $\mathbf{a}$ .

Alle anderen mathematischen Symbole wie bspw.  $\text{Tr}()$  für Spur, etc. folgen den üblichen Gepflogenheiten.

### Symbolliste

Im Folgenden als Referenz für den Leser geben wir eine Liste aller verwendeten Symbole mit fester Bedeutung an. Dies geschieht in alphabetischer Reihenfolge.

Alle Referenzen beziehen sich auf Gleichungen.

$\hat{A}$  Antisymmetrisierer. [2.3.7](#), [3.4.3](#), [3.4.4](#), [8.3.1](#), [A.4.1](#)

$\chi$  Kern-Wellenfunktion. [3.3.2](#), [3.3.4](#), [3.3.5](#)

$\mathbf{D}$  Dichtematrix. [4.2.7](#), [4.3.2](#), [4.3.3](#), [7.2.3](#), [7.2.4](#), [7.3.1](#), [7.3.5](#), [7.3.6](#), [7.4.1-7.4.6](#)

$d$  Freiheitsgrad. [2.0.1](#), [2.1.3-2.1.7](#), [2.2.1-2.2.7](#), [2.3.2-2.3.8](#), [2.4.3-2.4.5](#), [2.5.1](#), [2.5.3](#)

$\Delta$  Laplace-Operator. [9.5.1](#)

## 1. Einleitung

$\delta$  Delta-Distribution. [7.3.1](#), [7.3.5](#), [7.3.6](#)

$\delta$  Kronecker-Delta. [4.1.1](#), [4.1.11](#)

$E$  parallele Effizienz. [9.2.2](#)

$E$  Energie. [3.1.7](#), [7.6.1](#), [8.1.6](#), [8.2.1](#), [8.2.2](#), [8.2.6](#), [8.2.7](#), [9.4.1–9.4.3](#)

$E^{(0)}$  Grundzustandsenergie. [3.4.6](#)

$E^{(\text{ANOVA})}$  Bond-Order diSSection in an ANalysis Of VAriance-like fashion-Energie. [7.2.6](#),  
[9.4.1–9.4.3](#)

$E^{(\text{corr})}$  Korrelationsenergie. [4.1.18](#)

$E^{(c)}$  Coulombenergie. [7.4.15](#), [7.4.16](#)

$E^{(el)}$  Elektronischer Anteil der Gesamtenergie. [3.3.6](#)

$E^{(\text{SCF})}$  Self-Consistent Field-Energie. [4.1.6](#), [4.1.7](#), [4.1.17](#), [4.1.18](#), [7.2.3](#), [7.2.4](#), [7.2.6](#),  
[7.5.1](#), [9.4.1–9.4.3](#)

$E^{(\text{tot})}$  Gesamtenergie. [4.1.18](#)

$\epsilon$  Orbitalenergie. [4.1.16](#), [4.1.17](#), [4.2.1](#), [7.4.14](#), [7.6.1](#), [8.2.5](#), [8.2.6](#)

$\mathbf{F}$  Fockmatrix. [4.2.1](#), [4.2.2](#), [4.2.8](#), [4.2.9](#), [7.1.3–7.1.5](#), [7.2.3](#), [7.2.4](#), [7.4.5](#), [7.4.6](#)

$\hat{F}$  Fockoperator. [4.1.15](#), [4.1.16](#), [4.2.8](#), [7.2.1](#), [7.4.12](#), [7.4.14](#), [8.2.1](#)

$\hat{H}$  Hamiltonoperator. [3.1.3](#), [3.1.4](#), [3.1.7](#), [3.2.0](#), [3.2.2](#), [3.3.3](#), [3.4.6](#), [8.1.3](#), [8.1.6](#), [8.2.1](#),  
[A.4.2](#)

$h$  Einteilchenenergie. [4.1.6–4.1.8](#), [4.1.11](#), [4.1.17](#)

$\mathbf{h}$  Einteilchenmatrix. [7.2.3](#), [7.2.4](#), [7.4.1](#), [7.4.2](#), [7.4.5–7.4.7](#)

$\hat{h}$  Ein-Teilchen-Hamiltonoperator. [4.1.8](#), [4.1.13–4.1.15](#)

$\hat{H}^{(el)}$  Elektron-Hamiltonoperator. [3.3.6](#), [4.1.6](#)

$\mathbf{H}$  Hamiltonmatrix. [8.1.2](#), [8.1.3](#)

$l$  Vier-Zentren-Integral. [4.2.11](#), [4.3.2](#), [4.3.3](#), [5.4.6](#), [5.4.7](#), [7.1.2](#), [7.4.4](#), [8.2.5](#), [8.2.6](#)

$\mathcal{I}$  Irreguläre Kugelflächenfunktion. [7.4.8](#), [7.4.12](#)

$J$  Elektron-Elektron-Potentialenergie. [4.1.6](#), [4.1.7](#), [4.1.9](#), [4.1.11](#), [7.3.2](#)

$\hat{J}$  Coulomboperator. [4.1.13](#), [4.1.15](#)

- $\hat{K}$  Dekorrelator. 3.3.5, 3.4.2–3.4.4
- $K$  Austauschenergie. 4.1.6, 4.1.7, 4.1.10, 4.1.11
- $\hat{K}$  Austauschoperator. 4.1.13, 4.1.15, 7.4.12
- $\mathcal{L}_2$  Raum der quadratintegrierbaren Funktionen. 3.4.5
- $N$  Problemgröße. 9.2.3
- $n$  Anzahl an Basisfunktionen. 2.3.1, 2.3.4, 2.3.6, 2.3.7, 4.2.1, 4.2.6, 4.2.13, 4.2.16, 4.3.5, 4.3.6, 5.4.7–5.4.9, 6.1.1, 6.1.8, 6.3.1, 6.3.3, 6.3.4, 7.1.5, 9.5.7
- $n$  Anzahl. 3.1.1, 3.1.2, 3.1.4, 3.1.6, 3.1.7, 3.4.5
- $M$  Anzahl der Kerne. 3.2.0, 3.2.3, 3.3.3, 4.1.14, 7.3.1, 9.1.2, 9.2.1
- $N$  Anzahl der Elektronen. 3.2.0, 3.2.3, 3.3.3, 3.3.4, 3.3.8, 3.4.2–3.4.4, 4.1.2–4.1.4, 4.1.6, 4.1.7, 4.1.11, 4.1.13, 4.1.15, 4.1.17, 4.2.4, 4.2.5, 4.2.7, 8.1.1, 8.1.4, 8.3.1, A.4.1
- $P$  Prozesszahl. 9.2.2, 9.2.3
- $\mathbf{P}$  Legendre-Polynom. A.3.1, A.3.4, A.3.6, A.3.7
- $\mathbf{P}$  Projektion. 6.1.8, 6.1.10, 6.1.13, 6.2.10, 7.2.3, 7.4.2
- $q$  sphärisches Multipolmoment. 7.4.9, 7.4.10
- $\hat{\Phi}$  Fluktuationspotentialoperator. 8.2.1, 8.2.2, 8.2.6, 8.2.7
- $\Phi$  Coulomb-Potential. 7.3.4, 7.3.5, 7.3.8, 7.3.9, 7.4.15, 9.5.1
- $\hat{\phi}$  Fockabbildung. 7.4.1–7.4.4
- $\phi$  Einteilchen-Wellenfunktion. 4.1.1, 4.1.4, 4.1.8–4.1.13, 4.1.16, 4.2.1, 4.2.4–4.2.6, 4.2.13, 5.5.2, 7.2.1, 7.2.2, 7.4.14, A.4.2
- $\Psi$  Gesamtwellenfunktion. 3.1.1, 3.1.2, 3.1.4, 3.1.6, 3.2.3, 3.3.0, 3.3.2, 3.3.3, 3.3.5, 3.4.6
- $\psi$  elektronische Wellenfunktion. 3.3.2, 3.3.4–3.3.6, 4.1.2–4.1.4, 4.1.6, 8.1.1–8.1.4, 8.1.6, 8.2.1, 8.2.2, 8.2.4, 8.2.6, 8.2.7, 8.3.1, A.4.1
- $\psi$  Wellenfunktion. 3.1.4–3.1.7, 3.3.1
- $\psi^{(0)}$  Grundzustandselektronwellenfunktion. 3.3.7, 3.3.8, 3.4.2–3.4.4
- $\mathbf{R}$  Kernort. 3.2.0, 3.3.9, 4.1.14, 7.3.1, 7.3.5, 7.3.6, 10.2.2
- $\mathcal{R}$  Reguläre Kugelflächenfunktion. 7.4.8–7.4.10

## 1. Einleitung

$\mathbf{r}$  Elektronortskoordinate. 3.2.0, 3.3.4, 4.1.1, 4.1.4, 4.1.6, 4.1.8–4.1.11, 4.1.13, 4.1.14, 4.2.1, 4.2.4–4.2.6, 4.2.8, 4.2.10–4.2.13, 5.4.1–5.4.3, 5.5.2, 7.1.1, 7.3.1, 7.3.2, 7.3.5, 7.3.6, 7.4.8–7.4.10, 7.4.14, 7.4.15, 8.1.1, 8.1.3, 8.1.4, 8.2.1, 8.2.2, 8.2.6, 8.2.7, 8.3.1, 9.5.7, A.1.1, A.1.2, A.4.1, A.4.2

$\mathbb{R}^3$  Vektorraum der reellen Zahlen. 5.1.1

$\rho$  Ladungsdichte. 3.3.8, 3.3.9, 4.2.4, 4.2.5, 7.2.1, 7.2.6, 7.2.7, A.4.2, A.5.1

$\rho$  Ladungsverteilung. 7.3.1, 7.3.2, 7.3.4–7.3.6, 7.3.8, 7.4.9, 7.4.10, 7.4.15, 9.5.1, 9.5.7

$\mathbf{S}$  Überlappmatrix. 4.2.1, 4.2.2, 4.2.12–4.2.14, 4.2.16, 8.1.2, 8.1.4

$\mathcal{S}$  symmetrische Gruppe. 2.1.5

$\hat{S}$  Symmetrisierer. 2.1.5, 2.3.6

$S_c$  Scaleup. 9.2.3

$\sigma$  Permutation. 2.1.5

$\hat{T}$  Anregungsoperator. 8.1.4, 8.2.4, 8.2.7

$\hat{T}$  Kinetischer Operator. 3.1.3, 7.4.12

$T$  Laufzeit. 9.2.2, 9.2.3

$t$  Zeitpunkt. 3.1.1, 3.1.2, 3.1.4, 3.1.5, 3.3.0–3.3.2, 3.3.5, 3.3.6

$\hat{T}^{(\text{core})}$  Kinetischer Operator, Kerne. 3.2.2

$\hat{T}^{(\text{el})}$  Kinetischer Operator, Elektronen. 3.2.2

$V$  Potentialenergie. 7.3.2, 7.3.5, 7.4.9

$\hat{V}$  Potentialoperator. 3.1.3

$V$  Potentialfunktion. 3.3.7, 3.3.9

$\mathcal{V}$  Funktionenraum. 2.1.5, 2.3.5, 2.3.8, 2.4.1, 2.5.1, 2.5.3

$V^{(\text{BO})}$  Born-Oppenheimer-Oberfläche. 3.3.7

$\hat{V}^{(\text{core,core})}$  Kern-Kern-Wechselwirkungsoperator. 3.2.2, 8.2.1

$V^{(\text{core,el})}$  Kern-Elektron-Potentialenergie. 7.3.2

$\hat{V}^{(\text{core,el})}$  Kern-Elektron-Wechselwirkungsoperator. 3.2.2, 7.4.12

$\hat{V}^{(\text{el,el})}$  Elektron-Elektron-Wechselwirkungsoperator. 3.2.2, 7.4.12

- $\hat{V}^{(\text{tot})}$  Gesamtpotentialoperator. [3.3.4](#), [3.3.7](#)
- $\mathbf{X}$  verallgemeinerte Kernkoordinate. [3.2.0](#), [3.2.3](#), [3.3.0–3.3.7](#)
- $\mathbf{x}$  verallgemeinerte Elektronenkoordinate. [3.2.0](#), [3.2.3](#), [3.3.0–3.3.8](#), [3.4.2–3.4.4](#), [4.1.2–4.1.4](#)
- $\mathbf{x}$  verallgemeinerter Koordinate. [3.1.1](#), [3.1.2](#), [3.1.4–3.1.7](#), [3.4.5](#), [3.4.6](#)
- $\mathbf{x}$  Ort. [7.3.4](#), [7.3.5](#), [7.3.8](#), [7.4.15](#), [10.2.2](#)
- $\xi$  Basisfunktion. [4.2.1](#), [4.2.6](#), [4.2.8](#), [4.2.10–4.2.13](#), [5.4.1](#), [5.4.3](#), [7.3.2](#), [8.3.1](#), [A.1.1–A.1.3](#)
- $\mathbf{Y}$  Kugelflächenfunktion. [A.1.1](#)
- $Z$  Kernladung. [3.2.0](#), [4.1.14](#), [7.3.1](#), [7.3.5](#), [7.3.6](#)

## Danksagungen

Abschließend möchte ich die Gelegenheit ergreifen und vielen Leuten meinen Dank aussprechen.

Zunächst möchte ich einigen meiner studentischen und wissenschaftlichen Hilfskräfte danken. Saskia Metzler zeigte mir, dass auch das Programmieren eine Wissenschaft ist, welche sich durch Kontrolle und Wiederholbarkeit auszeichnet, in der Programmcode durch Komponententests wieder und wieder getestet sein muss, bevor er benutzt wird. Tillmann Crüger hat mir die Augen für Design Patterns und die Mächtigkeit der Programmiersprache C++ geöffnet, als er meinen anfänglichen Schritten darin einen mächtigen Schubs nach vorne gab. Michael Ankele hat viel Zeit, Mühe und Ideen in die graphische Benutzeroberfläche gesteckt, die ohne ihn all die Komplexität dieser Arbeit nicht so leicht verstecken könnte. Julian Iseringhausen stand mir stets mit Rat und Tat bei der Anbindung seines Mehrgitter-Coulombblösers zur Seite. Das Programm zur Erstellung bielementiger, saturierter Moleküle steuerte Christian Neuen bei.

Weiterhin möchte ich folgende Mitarbeiter am Institut für Numerische Simulation besonders hervorheben. Christian Rieger gilt mein Dank für das Aufzeigen von Theoremen zu gestörten hermiteschen Matrizen und die zahlreichen Gespräche. Weiterhin stand besonders Jan Hamaekers' Büro immer wieder offen, wenn es darum galt, die letzten Falten aus dem Theorie-Teil zu bügeln.

Meinem Doktorvater Michael Griebel danke ich für die geweckte Begeisterung, die Überlassung des Themas und die fortwährende Finanzierung.

Außerdem möchte ich all meinen Korrekturlesern meinen großen Dank aussprechen: Jutta Adelsberger, Björn Dickmann, Jens Külzer, Saskia Metzler, Christian Neuen, Alexander Rüttgers, Ralph Thesen, Stefan und Nadine Wich.

Schließlich und endlich danke ich allen Kollegen und Freunden am Institut für Numerische Simulation, die sich in vielen Seminaren immer wieder Fehlversuche anhörten und diskutierten: „Man lernt nur aus Fehlern, wenn überhaupt.“





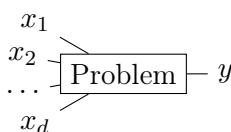
## 2. Tensorzerlegung

In diesem Kapitel beschäftigen wir uns mit der Approximation hochdimensionaler Funktionen. Es geht dabei aber nicht um die Approximation bekannter Funktionen, sondern um die Approximation von Funktionen, die unbekannte Lösungen zu einem bestimmten Problem darstellen. Wir wollen dabei die Funktion dergestalt ansetzen, dass sich das Problem, der Hochdimensionalität der Lösung angemessen, in einfachere Unterprobleme aufspaltet und dass uns gleichzeitig eine Vorschrift an die Hand gegeben ist, wie die Lösungen dieser Unterprobleme zusammengesetzt werden, um diese gesuchte hochdimensionale Lösung zu approximieren.

Ganz allgemein gesagt wollen wir uns zunächst vorstellen, dass diese hochdimensionale Funktion die Lösung zu einem Problem, bspw. ein physikalisches Modell, beschreibt, siehe auch Rabitz und Aliş [146]. Bezeichne diese Funktion  $f$  also den unbekanntem Zusammenhang

$$y = f(x_1, x_2, \dots, x_d) \quad (2.0.1)$$

zwischen den  $d$  Eingangsvariablen  $x_1, x_2, \dots, x_d$  mit  $x_i \in V$  und einem gesuchten komplexwertigen Ausgabewert  $y \in \mathbb{C}$ <sup>1</sup>, siehe Abbildung 2.1.  $V$  sei hierbei ein beliebiger Vektorraum über dem Körper der reellen Zahlen  $\mathbb{R}$ .



**Abbildung 2.1.:** Schema eines hochdimensionalen Problems, welches von  $d$  Eingangsvariablen  $x_i$  abhängt, die den Ausgabewert  $y$  bestimmen.

Weiterhin müssen wir das Problem angeben, welches festlegt, welches  $f$  wir genau approximieren wollen. Das Problem wird durch eine Gleichung angegeben, die  $f$  erfüllen muss, allgemein also<sup>2</sup>:

$$Lf = g \quad \text{auf } \Omega. \quad (2.0.2)$$

Hier werden verschiedene Gleichungen, die unsere Lösung  $f$  für  $(x_1, \dots, x_d) \in \Omega$  erfüllen muss, zu einem Operator  $L$  zusammengefasst, dessen Wirken auf  $f$  die Funktion  $g$  ergibt. Eine besondere Form ist, wenn die Funktion  $g$  von  $f$  abhängt,

$$Lf = \lambda f \quad \text{auf } \Omega, \quad (2.0.3)$$

<sup>1</sup>Der Einfachheit halber betrachten wir die Zielgröße als eindimensional, natürlich könnte dies auch einfach nur eine Komponente einer vektoriellen Zielgröße sein. Um Zusammenhänge zwischen Ausgabewerten geht es in dieser Arbeit jedoch nicht, sondern nur um Zusammenhänge zwischen Eingangsvariablen.

<sup>2</sup>Wir haben die Festlegung der Randwerte hier nicht explizit aufgeführt.

## 2. Tensorzerlegung

mit einem Skalar  $\lambda \in \mathbb{C}$ . Offensichtlich sollte  $f$  hier auf eine gewisse Art „normiert“ sein, denn würde man beide Seiten mit einem Faktor multiplizieren, würde sich an der Gleichung nichts ändern und damit wäre die skalierte Lösung gleichfalls eine Lösung. Probleme der Art (2.0.3) werden *Eigenwertprobleme* genannt,  $f$  heißt dann *Eigenfunktion* und  $\lambda$  *Eigenwert*. In dieser Arbeit werden wir uns mit einem besonderen dieser Eigenwertprobleme beschäftigen, der Schrödingergleichung. Dabei beschreiben wir jedoch in diesem Abschnitt einen Zugang, der grundsätzlich für alle Probleme dieser Art sinnvoll sein kann.

Wir wollen dafür im Laufe des Kapitels auf nachvollziehbare Art und Weise ein Konstrukt entwickeln. Dazu führen wir in Abschnitt 2.1 zunächst eine geeignete Funktionenklasse ein und teilen diese nach bestimmten Eigenschaften, insbesondere der Ununterscheidbarkeit ihrer Variablen, weiter auf. Abschließend bemühen wir eine Definition der Korrelation einer Funktion über die Eigenschaft, dass ihre Eingangsvariablen unkorreliert sind, wenn die Funktion eine Produktstruktur hat. Bei Produktstrukturen bietet es sich an, Tensorräume zu studieren. Diese behandeln wir in Abschnitt 2.2, wo allgemein Funktionen von einer Anzahl  $d$  Eingangsvariablen in vertensorierten Räumen über Funktionen mit nur einer Eingangsvariable dargestellt werden. Die Produktstruktur hängt dann mit der Rang-1-Bestapproximation einer Funktion in dem entsprechend vertensorierten Raum zusammen und wir beschäftigen uns mit deren Existenz in Abschnitt 2.4. Dieser Umstand liefert uns schließlich die Möglichkeit, formal Abbildungen einer Funktion auf ihre Rang-1-Bestapproximation zu definieren, die wir Dekorrektoren nennen. In Abschnitt 2.5 unterscheiden wir dabei zwischen unterscheidbaren und ununterscheidbaren Eingangsräumen bzw. Eingangsvariablen. Mittels geeignet gewählter Zerlegungen können wir schließlich verallgemeinerte Dekorrektoren definieren, die die Korrelation aus einer Funktion nur zwischen disjunkten Teilen ihrer Eingangsvariablen und nicht mehr zwischen allen Eingangsvariablen entfernen. Mittels diesen verallgemeinerten formalen Abbildungen schreiben wir dann hochdimensionale Funktionen in einer allgemeinen Reihenentwicklung, wobei mit jedem Summanden der Funktion weitere Korrelation hinzugefügt wird. Wir schließen mit einigen Bemerkungen in Abschnitt 2.5.4.

### 2.1. Hochdimensionale Funktionen

Die Komplexität einer Funktion bestimmt sich vor allem durch die Anzahl ihrer Eingangsfreiheitsgrade.

Wir wollen diese Eingangsfreiheitsgrade zunächst auftrennen in *Eingangsvariablen*, die veränderliche Freiheitsgrade darstellen, und *Eingangsparemeter*, die für eine Problemstellung feste Werte bezeichnen. Beispielhaft sei hier die Position eines Kerns als Variable, ihre Atomzahl aber als Parameter bezeichnet. Dies ist eine willkürliche Unterteilung, die derjenige zu treffen hat, der das komplexe Problem zu lösen hat und sich entscheiden muss, welche Eingangsfreiheitsgrade für ihn interessant und damit variabel sein sollten.

Wir betrachten also im Folgenden die Menge dieser Funktionen zu einem unbekanntem Zusammenhang (2.0.1) zu einer festen Zahl  $d$  an Eingangsvariablen.

**Definition 1** (Menge der hochdimensionalen Funktionen). Wir bezeichnen mit  $\mathcal{V}^{(d)}$  die Menge der komplex-wertigen Funktionen (2.0.1) zu einer festen Zahl  $d$  an Eingangsvariablen und mit einem reell-wertigen Eingangsvektorraum  $V$ .

Mit hochdimensionalen Funktionen sind also im Folgenden stets solche mit einer großen Zahl an Eingangsvariablen gemeint, also  $d \gg 1$ .

Diese Definition 1 ist jedoch noch viel zu allgemein für weitere Untersuchungen und wir wollen im Folgenden weitere Eigenschaften herausstellen. Gleichwohl bildet sie aber die Grundlage für unsere weiteren Überlegungen. Im Folgenden wird es immer um Funktionen wie in (2.0.1) gehen und wie sich die Komplexität bewältigen lässt, welche in der Suche nach dem richtigen  $f$  und auch in  $f$  selbst steckt.

### 2.1.1. (Un-)Unterscheidbarkeit

Unsere Eingangsvariablen beschreiben beliebige Informationen, welche die Grundlage des gesuchten Zusammenhangs  $f$  darstellen. Diese Informationen sind üblicherweise Orte in einem kontinuierlichen Raum, Orte eines diskreten Teilchens oder auch die Zeit.

Ergibt eine Funktion  $f$  den gleichen Zusammenhang, obwohl zwei ihrer Eingangsvariablen vertauscht wurden und gilt dies für jeden möglichen Wert aus  $V$ , dann sagt man, dass diese beiden Eingangsvariablen ununterscheidbar sind.

Die Ununterscheidbarkeit der Eingangsvariablen kann eine gewichtige Vereinfachung bedeuten. Dies ist genauso der Fall, wenn es sich nicht um Orte in einem kontinuierlichen Raum, sondern um die Aufenthaltsorte von diskreten Teilchen handelt. Dann sind in diesem Fall die Teilchen selbst ununterscheidbar. Das heißt, es ändert sich nichts, wenn zwei Teilchen willkürlich den Ort tauschen. Die Funktion ist invariant unter solch einer Permutation der beiden Eingangsvariablen.

Deswegen ist es lohnenswert, von den Funktionen der zuvor besprochenen Klasse jene Funktionen abzutrennen, deren Eingangsvariablen ununterscheidbar sind. Wir definieren dabei zunächst die Auswirkung der Permutation auf die Funktion  $f$  für zwei besondere Fälle.

**Definition 2** ((anti-)symmetrische Funktion). Eine Funktion  $f^S \in \mathcal{V}^{(d)}$  mit ununterscheidbaren Eingangsvariablen heißt *symmetrisch*, wenn für alle  $x_1, \dots, x_d \in V$  und für alle  $i \neq j$

$$f^S(\dots, x_i, \dots, x_j, \dots) = f^S(\dots, x_j, \dots, x_i, \dots) \quad (2.1.1)$$

gilt. Die Menge dieser Funktionen bezeichnen wir mit  $\mathcal{V}_S^{(d)}$ .

Die Funktion  $f^A$  heißt *antisymmetrisch*, wenn unter gleichen Voraussetzungen

$$f^A(\dots, x_i, \dots, x_j, \dots) = -f^A(\dots, x_j, \dots, x_i, \dots) \quad (2.1.2)$$

gilt. Analog sei die Klasse hier mit  $\mathcal{V}_A^{(d)}$  bezeichnet.

*Bemerkung 1* (gemischte Symmetrien). Natürlich kann eine Funktion auch gemischte Symmetrie haben. Dabei sind die Indexmengen  $I_s, I_a \subset \{1, \dots, d\}$  jedoch disjunkt, wie an folgendem Beispiel klar wird:

## 2. Tensorzerlegung

**Beispiel 1** Sei  $f \in \mathcal{V}^{(3)}$  und sei  $f$  in den ersten beiden Eingangsvariablen antisymmetrisch,  $I_a = \{1, 2\}$ , und in den letzten beiden wiederum symmetrisch,  $I_s = \{2, 3\}$ . Nun permutieren wir

$$\begin{aligned} f(x_1, x_2, x_3) &= -f(x_2, x_1, x_3) = -f(x_2, x_3, x_1) = f(x_3, x_2, x_1) \\ &= f(x_3, x_1, x_2) = -f(x_1, x_3, x_2) = -f(x_1, x_2, x_3) \end{aligned}$$

und damit muss  $f$  gleich 0 sein.  $\diamond$

Das Argument lässt sich für  $f$  mit  $d$  Eingangsvariablen analog führen: Wir benötigen nur ein  $x_i \in I_s$ ,  $x_k \in I_a$  und  $x_j \in I_s \cup I_a$ .

Aus diesem Grund beschränken wir uns auf (vollständig) symmetrische oder antisymmetrische Funktionen.

Aufgrund der Ununterscheidbarkeit der Eingangsvariablen schreiben wir die Funktion  $f$  auch so

$$f(\{x_1, \dots, x_d\}) =: f(\{x_i\}_I) \quad \text{mit } I := \{1, \dots, d\}, \quad (2.1.3)$$

wobei die Mengenklammern die Beliebigkeit in der Reihenfolge der Eingangsvariablen deutlich machen. Wir stellen damit die Menge der Eingangsvariablen  $\{x_1, x_2, \dots\}$  der Folge von Eingangsvariablen  $x_1, x_2, \dots$  mit fester Abfolge gegenüber, welche wir ähnlich verkürzt auch so schreiben

$$f(x_1, \dots, x_d) =: f(x_I) \quad \text{mit } I := 1, \dots, d. \quad (2.1.4)$$

Warum interessiert uns diese Eigenschaft der Unterscheidbarkeit? Natürlich ist sie durch wahrgenommene physikalische Realität motiviert<sup>3</sup>, jedoch steckt dahinter auch ein anderer Grund: Um die Komplexität des Problems zu durchbrechen, wollen wir später Funktionen aus  $\mathcal{V}^{(d)}$  durch Funktionen darstellen, die nur noch in einem Unterraum  $\mathcal{V}^{(k)}$  ihre Komplexität bewahren, wobei  $k$  kleiner und insbesondere in der Größe unabhängig von  $d$  ist. Zu diesem Zweck benötigen wir Abbildungen von  $\mathcal{V}^{(d)}$  nach  $\mathcal{V}^{(k)}$ . Offensichtlich müssen diese Abbildungen von einer bestimmte Untermenge der Eingangsvariablen auf die Eingangsvariablen des Zielraums wirken. Dabei ist natürlich entscheidend, ob die Eingangsvariablen unterscheidbar sind oder nicht.

Sind sie unterscheidbar, spielt es eine Rolle, welche Eingangsvariablen gewählt werden. Sind sie ununterscheidbar, spielt es keine Rolle welche wir wählen, sondern einzig wie viele.

Es scheint also so, als wäre das Problem mit ununterscheidbaren Eingangsvariablen einfacher.

**(Anti-)Symmetrisierer** Wir können jede Funktion aus  $\mathcal{V}_S^{(d)}$  und  $\mathcal{V}_A^{(d)}$  aus Funktionen aus  $\mathcal{V}^{(d)}$  durch zwei bestimmte Operatoren  $\hat{S}$  und  $\hat{A}$  erzeugen:

<sup>3</sup>So unterteilen sich in der Physik die Gesamtheit der Teilchen in genau zwei Klassen, Fermionen und Bosonen, deren beschreibende Funktionen entweder genau symmetrisch oder genau antisymmetrisch sind.

**Definition 3** ((Anti-)Symmetrisierer). Wir bezeichnen mit  $\hat{S} : \mathcal{V}^{(d)} \rightarrow \mathcal{V}_S^{(d)}$  den *Symmetrisierer* und mit  $\hat{A} : \mathcal{V}^{(d)} \rightarrow \mathcal{V}_A^{(d)}$  den *Antisymmetrisierer*. Sie definieren sich wie folgt in ihrer Wirkung auf eine Funktion  $f \in \mathcal{V}^{(d)}$

$$\hat{S}f(x_1, \dots, x_d) := \frac{1}{d!} \sum_{\sigma \in \mathcal{S}_d} f(x_{\sigma(1)}, \dots, x_{\sigma(d)}) \quad \text{mit } (\hat{S}f) \in \mathcal{V}_S^{(d)} \quad (2.1.5a)$$

$$\hat{A}f(x_1, \dots, x_d) := \frac{1}{d!} \sum_{\sigma \in \mathcal{S}_d} (-1)^{\text{sgn}(\sigma)} f(x_{\sigma(1)}, \dots, x_{\sigma(d)}) \quad \text{mit } (\hat{A}f) \in \mathcal{V}_A^{(d)}, \quad (2.1.5b)$$

wobei die Summe über alle Permutationen  $\sigma$  der symmetrischen Gruppe  $\mathcal{S}_d$  erfolgt und  $\text{sgn}(\sigma)$  das Vorzeichen der Permutation bezeichnet.

Die Wohldefiniertheit folgt offensichtlich aus der Definition der symmetrischen Gruppe. Symmetrisierer und Antisymmetrisierer erzeugen also schlicht alle möglichen Permutationen, summieren dies mit gewünschtem Vorzeichen auf und normalisieren, und sorgen so für das gewünschte Verhalten.

### 2.1.2. Korrelation und Produktfunktionen

Korrelation ist das Maß, mit dem wir die Komplexität einer hochdimensionalen Funktion ausdrücken möchten. Machen wir zunächst eine sehr allgemeine Definition.

**Definition 4** (Unkorreliert). In einer Funktion  $f(x_1, \dots, x_d) \in \mathcal{V}^{(d)}$  heißen die Variablen  $x_i$  und  $x_j$  mit  $i \neq j$  *unkorreliert*, wenn es zwei Funktionen  $f_1(x_1, \dots, x_{d-1}) \in \mathcal{V}^{(d-1)}$  und  $f_2(x_1, \dots, x_{d-1}) \in \mathcal{V}^{(d-1)}$  gibt, so dass

$$f(x_1, \dots, x_d) = f_1(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d) \cdot f_2(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d) \quad (2.1.6)$$

gilt.

Wir wollen also Korrelation bzw. deren Fehlen mit der Existenz einer Produktstruktur verknüpfen.

**Beispiel 2** Betrachten wir dazu eine Funktion mit zwei Eingangsvariablen  $f(x_1, x_2)$ . Gleichung (2.1.6) bedeutet, dass zwei Funktionen mit je einer Eingangsvariable  $f_1(x)$  und  $f_2(x)$  existieren, so dass  $f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$  gilt, sprich ein Separationsansatz gelingt. Wir können also eine Funktion mit zwei Eingangsvariablen in einem Produktansatz genau dann zerlegen, wenn ihre beiden Eingangsvariablen  $x_1$  und  $x_2$  unkorreliert sind.  $\diamond$

Verallgemeinern wir dieses Beispiel, nämlich dass alle Eingangsvariablen unkorreliert sind:

**Definition 5** (Unkorrelierte Eingangsvariablen). Wir bezeichnen die Eingangsvariablen  $x_1, \dots, x_d$  einer Funktion  $f \in \mathcal{V}^{(d)}$  als *unkorreliert*, wenn es  $d$  Funktionen  $f_i \in \mathcal{V}^{(1)}$  gibt, so dass für alle  $x_1, \dots, x_d$

$$f(x_1, \dots, x_d) = f_1(x_1) \cdot \dots \cdot f_d(x_d) \quad (2.1.7)$$

gilt.

## 2. Tensorzerlegung

Der gegenteilige Fall von korrelierten Variablen ist dann einfach wie folgt formulierbar:

**Definition 6** (Korreliert). In einer Funktion  $f(x_1, \dots, x_d) \in \mathcal{V}^{(d)}$  heißen die Variablen  $x_i$  und  $x_j$  mit  $i \neq j$  *korreliert*, wenn es *keine* zwei Funktionen nach Definition 4 gibt.

Leider ist diese Definition nicht sehr hilfreich, denn man würde sicherlich gerne die Korrelation zwischen Variablen messen können. Dazu möchte man bspw. als Maß den Betrag der Differenz  $|f(x_1, x_2) - f_1(x_1) \cdot f_2(x_2)|$  benutzen. Die Funktionen  $f_1$  und  $f_2$  sind aber nur für den unkorrelierten Fall wohldefiniert, genau dann wenn die Differenz gerade Null ist.

In der Wahrscheinlichkeitstheorie wird die Korrelation einer Funktion zwischen zwei Eingangsvariablen durch die Kovarianz bestimmt. Die Kovarianz ist genau dann Null, wenn der Erwartungswert der Funktion  $f(x_1, x_2)$  gleich dem Produkt der Erwartungswerte von  $f_1(x_1)$  und  $f_2(x_2)$  ist. Diese Erwartungswerte sind als Integrale mit entsprechendem Wahrscheinlichkeitsmaß über die Verteilung einer Zufallsvariable definiert, siehe [117, Kap. 5.1].

Natürlich schränkt das gewählte Maß die Funktionenklasse  $\mathcal{V}^{(d)}$  ein. Wählen wir das Integral zur Bestimmung von Erwartungswerten, so müssen die Funktionen integrierbar sein, d. h. das Integral muss einen endlichen Wert annehmen.

Allgemeiner gesagt sind Erwartungswerte Abbildungen vom Raum der Wahrscheinlichkeitsverteilungen auf reelle Zahlen. Eine solche Abbildung benötigen wir ebenfalls, wenn wir ein zur Kovarianz analoges Maß definieren wollen.

Um uns diesen Abbildungen zu nähern, welche quasi die Korrelation aus einer Funktion entfernen und die wir deswegen später *Dekorrelatoren* nennen werden, müssen wir uns zunächst mit Räumen beschäftigen, die eine Produktstruktur aufweisen, genau wie sie auch in (2.1.6) benutzt wird. Dazu betrachten wir in den folgenden Abschnitten zunächst Tensoren und dann die direkte Summe und das Tensorprodukt von Funktionenräumen.

## 2.2. Tensoren

In diesem Teil folgen wir Hackbusch [80, Abschn. 1.1] in seiner einführenden Darstellung von Tensoren als Erweiterung der Konzepte des Vektors und der Matrix aus der linearen Algebra zur multilinearen Algebra.

**Definition 7** (Tensor). Ein *Tensor*  $v$  ist ein Objekt im  $\mathbb{R}^{n_1 \times \dots \times n_d}$  der *Ordnung*  $d$ . Mit dem Multiindex

$$i = (i_1, \dots, i_d) \in \mathcal{I} := \mathcal{I}_1 \times \dots \times \mathcal{I}_d \quad (2.2.1)$$

mit  $\mathcal{I}_j = \{1, \dots, d_j\}$  können wir dann eine reellwertige Komponente

$$v_i = v[i] = v[i_1, \dots, i_d] \in \mathbb{R} \quad (2.2.2)$$

des Tensors  $v$  beschreiben<sup>4</sup>.

---

<sup>4</sup>Wir halten uns hier an die Notation aus [80] und bleiben im Körper  $\mathbb{R}$  anstelle von  $\mathbb{C}$ , da letzterer isomorph zu  $\mathbb{R}^2$  ist.

Damit ist ein Tensor der Ordnung 1 ein Vektor und ein Tensor der Ordnung 2 eine Matrix, siehe dazu auch [80, Abschn. 1.2.1].<sup>5</sup>

Ein Zusammenhang zwischen dem  $\mathbb{R}^{\mathcal{I}_j}$  und  $\mathbb{R}^{\mathcal{I}}$  ergibt sich über das Tensorprodukt.

**Definition 8** (Tensorprodukt). Für  $d$  gegebene Vektoren  $v^{(j)} \in \mathbb{R}^{\mathcal{I}_j}$  sei

$$v := v^{(1)} \otimes \dots \otimes v^{(d)} = \bigotimes_{j=1}^d v^{(j)} \quad (2.2.3)$$

das *Tensorprodukt* mit den Einträgen nach (2.2.2):

$$v_i = v[i_1, \dots, i_d] = v_{i_1}^{(1)} \cdot \dots \cdot v_{i_d}^{(d)} \quad (2.2.4)$$

Man kann sich Tensorprodukte als Verallgemeinerung des äußeren Produkts  $v_1 \cdot v_2^T$  zweier Vektoren  $v_1, v_2$  vorstellen. Dabei weist das hochgestellte  $T$ , die Transposition des zweiten Vektors, auf die besondere Art der Elementmultiplikation hin. Wir werden dazu das Symbol  $\otimes$  auch elementweise verwenden, siehe (2.2.3), um dieses verallgemeinerte äußere Produkt zwischen Vektoren kenntlich zu machen.

Diese Tensorprodukte, auch *zerlegbare Tensoren* genannt, erzeugen einen Raum.

**Definition 9** (Tensorraum). Der Tensorraum, geschrieben als

$$\bigotimes_{j=1}^d \mathbb{R}^{\mathcal{I}_j} = \mathbb{R}^{\mathcal{I}_1} \otimes \dots \otimes \mathbb{R}^{\mathcal{I}_d}, \quad (2.2.5)$$

ist definiert durch das Erzeugendensystem dieser zerlegbaren Tensoren

$$\bigotimes_{j=1}^d \mathbb{R}^{\mathcal{I}_j} = \text{span}\{v^{(1)} \otimes \dots \otimes v^{(d)} : v^{(j)} \in \mathbb{R}^{\mathcal{I}_j}\}. \quad (2.2.6)$$

*Bemerkung 2* (Beliebiger Tensor). Wir bemerken, dass ein beliebiger Tensor  $v \in \bigotimes_{j=1}^d \mathbb{R}^{\mathcal{I}_j}$  im Allgemeinen *nicht* als zerlegbarer Tensor dargestellt werden kann, sondern nur als *Linearkombination* solcher.

*Bemerkung 3* (Äquivalenz Funktionenraum und Tensorraum). Mit der Überlegung aus Bemerkung 2, dass sich jeder Tensor als Linearkombination zerlegbarer Tensoren darstellen lässt<sup>6</sup>, folgt auch sofort  $\bigotimes_{j=1}^d \mathbb{R}^{\mathcal{I}_j} = \mathbb{R}^{\mathcal{I}}$ , da nach Definition 9 schon  $\bigotimes_{j=1}^d \mathbb{R}^{\mathcal{I}_j} \subset \mathbb{R}^{\mathcal{I}}$  gelten muss.

<sup>5</sup>Wir schließen uns hier vor allem auch der Bemerkung an, dass kovariante und kontravariante Tensoren eine Vermischung von den dort dargestellten Tensoren und Einsteinscher Summenkonvention sind. Für eine derartige Darstellung in Bezug auf das elektronische Strukturproblem siehe Head-Gordon et al. [89].

<sup>6</sup>Anschaulich ist dies z. B. in Szabo und Ostlund [168, Abschn. 2.2.7] dargestellt.

## 2. Tensorzerlegung

**Lemma 1** (Tensorraumdimension). *Die Dimension des Tensorraums ist*

$$\dim\left(\bigotimes_{j=1}^d \mathbb{R}^{\mathcal{I}_j}\right) = \dim(\mathbb{R}^{\mathcal{I}}). \quad (2.2.7)$$

*Beweis.* Dies folgt mit  $|\mathcal{I}| = \prod_{j=1}^d |\mathcal{I}_j|$ . □

## 2.3. Summen und Produkte von Funktionenräumen

Im Folgenden werden wir das über Tensoren Gesagte auf Funktionenräume übertragen. Dazu benötigen wir zunächst eine Basis, damit die Räume den Charakter von (unendlich-dimensionalen) Vektorräumen erhalten. Wir wollen uns dabei aber mit den besonderen Schwierigkeiten<sup>7</sup> von unendlich-dimensionalen Vektorräumen nicht weiter beschäftigen, da unser Ziel die Behandlung komplexer Probleme in Rechnern mit endlichem Speicher ist.

Wir betrachten nun also Funktionenräume, die von einer (endlichen) Basis aufgespannt werden. Dabei widmen wir uns Summen und Produkten dieser Funktionenräume, aber auch deren Zerlegung für unser Ziel der Abbildung in Räume von Funktionen mit weniger Korrelation.

### 2.3.1. Von der Funktion zum Tensor

Wir wollen zunächst kurz auf den Zusammenhang zwischen einer Funktion  $f(x_1, \dots, x_d) \in \mathcal{V}^{(d)}$  und einem Tensor  $\mathbf{C}$  eingehen. Dieser ergibt sich, indem wir den Funktionenraum  $\mathcal{V}^{(d)}$  mittels einer (Produkt)Basis, elementweisen Addition und skalarer Multiplikation als einen Vektorraum auffassen. Sei dazu  $\mathcal{V}^{(1)}$  der Raum der Funktionen mit einer Eingangsvariable mit Basis  $\{b_i\}_{i=1}^n$ , dann können wir eine Funktion  $g(x) \in \mathcal{V}^{(1)}$  mit einer Eingangsvariable wie folgt schreiben:

$$g(x) = \sum_{i=1}^n c_i b_i(x). \quad (2.3.1)$$

$g(x)$  hat also diese Darstellung in der gegebenen Basis mit Koeffizientenvektor  $\mathbf{c}$  der Dimension  $n$ .

Sei nun  $\mathcal{V}^{(d)} := \bigotimes_{i=1}^d \mathcal{V}_i^{(1)}$ . Erinnern wir uns an Definition 7 und legen zunächst fest, dass die Indexmengen  $\mathcal{I}_i = \{1, \dots, n\}$  für alle  $i = 1, \dots, d$  gleich sind, dann können wir jedes  $f(x_1, \dots, x_d) \in \mathcal{V}^{(d)}$  mit der Tensorproduktbasis  $\{b_{j_1}(x_1) \cdot \dots \cdot b_{j_d}(x_d)\}_{j_i \in \mathcal{I}_i, i=1, \dots, d}$  darstellen:

$$f(x_1, \dots, x_d) = \sum_{j_i \in \mathcal{I}_i, i=1, \dots, d} \mathbf{C}[j_1 \dots j_d] b_{j_1}(x_1) \dots b_{j_d}(x_d). \quad (2.3.2)$$

---

<sup>7</sup>Beispielsweise muss bei der Darstellung eines Vektors als Linearkombination von Basisvektoren darauf geachtet werden, ob diese unendliche(!) Reihe überhaupt konvergiert.



Diesmal ist also  $\mathbf{C}[j_1 \dots j_d]$  in (2.3.2) der Koeffizient zu einer bestimmten Produktbasisfunktion  $b_{j_1}(x_1) \cdot \dots \cdot b_{j_d}(x_d)$ . Er ist eine Komponente eines Tensors  $\mathbf{C}$  der Ordnung  $d$ . Er ist das Pendant der multilinearen Algebra zum Koeffizientenvektor  $\mathbf{c}$  in (2.3.1) in der linearen Algebra. Mit der Frage nach der Darstellung bzw. Approximation von Funktionen, die nicht nur aus  $\mathcal{V}^{(d)}$  stammen, werden uns im nächsten Abschnitt 2.4 im Zusammenhang mit Bestapproximationen beschäftigen.

### 2.3.2. Summen

**Definition 10** (Direkte Summe von Funktionenräumen). Seien  $\mathcal{V}_1^{(d)}$  und  $\mathcal{V}_2^{(d)}$  Funktionenräume mit Basen  $\{a_i(x_1, \dots, x_d)\}_{i=1, \dots, n_A}$ ,  $\{b_i(x_1, \dots, x_d)\}_{i=1, \dots, n_B}$  mit gleicher Anzahl der Eingangsvariablen  $d$ . Dann ist die *direkte Summe*  $\mathcal{V}_1^{(d)} \oplus \mathcal{V}_2^{(d)}$  derjenige Raum  $\mathcal{V}^{(d)}$ , welcher von

$$\{a_i(x_1, \dots, x_d)\} \cup \{b_i(x_1, \dots, x_d)\} \quad (2.3.3)$$

aufgespannt wird. Ist die Vereinigung der Menge der Basisfunktionen  $\{a_i\}_i$  und  $\{b_i\}_i$  wiederum eine Basis – d.h. die Räume  $\mathcal{V}_1^{(d)}$  und  $\mathcal{V}_2^{(d)}$  sind orthogonal zueinander bzgl. eines gegebenen Skalarprodukts – so heißt die direkte Summe auch *orthogonal*, siehe [3, S. 11].

Man beachte, dass die Anzahl der Eingangsvariablen  $d$ , also die Anzahl der Argumente einer Basisfunktion bei der direkten Summe gleich sind, sowohl in den beiden Ausgangsräumen  $\mathcal{V}_1^{(d)}$  und  $\mathcal{V}_2^{(d)}$ , als auch in dem Zielraum  $\mathcal{V}^{(d)}$ .

**Zerlegung** Jeder Funktionenraum lässt sich in Unterräume zerlegen. Man kann sich die (orthogonale) Zerlegung der Basis eines Funktionenraums einfach als Auftrennen der Menge der Basisfunktionen in zwei (disjunkte) Mengen  $\{a_i\}_i$  und  $\{b_i\}_i$  vorstellen, so dass (2.3.3) gilt. Die Art der Auftrennung ist dabei zunächst nicht weiter beschränkt.

### 2.3.3. Produkte

**Definition 11** (Tensorprodukt von Funktionenräumen). Seien  $\mathcal{V}_1^{(d_1)}$  und  $\mathcal{V}_2^{(d_2)}$  Funktionenräume mit Basen  $\{a_i(x_1, \dots, x_{d_1})\}_{i=1, \dots, n_1}$ ,  $\{b_j(x_1, \dots, x_{d_2})\}_{j=1, \dots, n_2}$  mit nicht notwendigerweise gleicher Anzahl der Eingangsvariablen  $d_1$  bzw.  $d_2$ . Dann ist das *Tensorprodukt*  $\mathcal{V}_1^{(d_1)} \otimes \mathcal{V}_2^{(d_2)}$  derjenige Raum  $\mathcal{V}^{(d_1+d_2)}$ , welcher von

$$\{a_i(x_1, \dots, x_{d_1}) \cdot b_j(x_{d_1+1}, \dots, x_{d_1+d_2})\}_{i=1, \dots, n_1, j=1, \dots, n_2} \quad (2.3.4)$$

aufgespannt wird.

Die Basis von  $\mathcal{V}_1^{(d_1)} \otimes \mathcal{V}_2^{(d_2)}$  entsteht also durch Produktverknüpfung aller möglichen Paare einer Basisfunktion aus  $\{a_i\}_i$  und einer Basisfunktion aus  $\{b_j\}_j$ . Man beachte, dass sich die Anzahl der Eingangsvariablen  $d$  beim Tensorprodukt verändert; sie entspricht der Summe der Anzahlen der Eingangsvariablen  $d_1$  und  $d_2$  der „vertensorierten“ Räume.

## 2. Tensorzerlegung

**Zerlegung** Besitzt ein Funktionenraum eine Tensorproduktstruktur, dann kann er durch Aufspalten der Basisfunktionen in umgekehrter Weise zu Definition 8 in entsprechende Unterräume zerlegt werden. Aufgrund des schnell sehr großen Umfangs der Basis gestaltet sich im Allgemeinen diese Zerlegung jedoch sehr schwierig.

### 2.3.4. Spezielle Tensorprodukträume

Mit den zuvor gemachten Definitionen wollen wir nun einige Funktionenräume mit Tensorproduktstruktur näher betrachten. Dabei greifen wir auch wieder den Fall der ununterscheidbaren Eingangsvariablen aus Abschnitt 2.1.1 auf.

**Eindimensionale Unterräume** Wir widmen uns zunächst dem einfachen Fall, dass  $d$  Funktionenräume  $\mathcal{V}_i^{(1)}$  mit  $i = 1, \dots, d$  über Funktionen mit einer Eingangsvariable „vertensoriert“ werden sollen. Wir schreiben

$$\mathcal{V}^{(d)} = \bigotimes_{i=1}^d \mathcal{V}_i^{(1)}, \quad (2.3.5)$$

wobei die Gleichheit in Analogie zu Bemerkung 3 folgt. Nehmen wir an, dass die Räume  $\mathcal{V}_1^{(1)} = \dots = \mathcal{V}_d^{(1)} = \mathcal{V}^{(1)}$  alle gleich sind und von der Basis  $\{b_i(x)\}$  aufgespannt werden, dann ergibt sich die Basis für den aufgespannten Teilraum von  $\mathcal{V}^{(d)}$  durch Bildung aller möglichen  $d$ -Tupel aus der Menge der eindimensionalen Basisfunktionen,  $\{b_{i_1}(x_1) \cdot \dots \cdot b_{i_d}(x_d)\}$  mit  $i_1, \dots, i_d \in \{1, \dots, d\}$ .

**(Anti-)symmetrische Produkträume** Für den Fall, dass nur gleiche Räume vertensoriert werden, also  $\mathcal{V}_1^{(d_1)} = \mathcal{V}_2^{(d_2)} = \mathcal{V}^{(d)}$  in Definition 8, können wir noch zwei spezielle Tensorprodukte definieren, nämlich das symmetrische und das antisymmetrische Tensorprodukt. Diese benötigen wir, wenn die Eingangsvariablen ununterscheidbar sind. Dies bedingt offensichtlich gleiche Eingangsräume.

**Definition 12** (symmetrisches Tensorprodukt). Sei  $\mathcal{V}^{(d)}$  ein Funktionenraum mit Basis  $\{a_i(x_1, \dots, x_d)\}_{i=1, \dots, n}$  mit  $d$  Eingangsvariablen. Dann ist das *symmetrische Tensorprodukt*  $\mathcal{V}^{(d)} \overset{S}{\otimes} \mathcal{V}^{(d)}$  derjenige Raum  $\mathcal{V}_S^{(2d)}$ , welcher von

$$\{\hat{S}(a_i(x_1, \dots, x_d) \cdot a_j(x_{d+1}, \dots, x_{2d}))\}_{i,j=1, \dots, n}. \quad (2.3.6)$$

aufgespannt wird.

**Definition 13** (antisymmetrisches Tensorprodukt). Voraussetzungen wie in Definition 12. Dann ist das *antisymmetrische Tensorprodukt*  $\mathcal{V}^{(d)} \overset{A}{\otimes} \mathcal{V}^{(d)}$  derjenige Raum  $\mathcal{V}_A^{(2d)}$ , welcher von

$$\{\hat{A}(a_i(x_1, \dots, x_d) \cdot a_j(x_{d+1}, \dots, x_{2d}))\}_{1 \leq i < j \leq n} \quad (2.3.7)$$

aufgespannt wird. Das antisymmetrische Tensorprodukt ist auch als *Wedge Produkt* in bspw. Grassmann-Algebren bekannt, siehe [188, Kap. 8].

### 2.3. Summen und Produkte von Funktionenräumen

Bei diesen beiden Definitionen haben wir stets nur ein Erzeugendensystem angegeben und keine Basis. Der Grund liegt darin, dass die Produktbasisfunktionen zum Teil je beide Charaktere enthalten, symmetrisch und antisymmetrisch, wie an folgendem Beispiel klar wird:

**Beispiel 3** Betrachten wir die Darstellung einer Funktion  $f(x_1, x_2) \in \mathcal{V}^{(2)}$  mit zwei Eingangsvariablen in der endlichen Basis  $\{a_i(x)\}_{i=1,2}$  des Raums  $\mathcal{V}^{(1)}$ , welcher mit sich selbst vertensoriert zu  $\mathcal{V}^{(1)} \otimes \mathcal{V}^{(1)}$  wird,

$$\begin{aligned} f(x_1, x_2) &= \sum_{i,j=1}^2 \mathbf{C}[i, j] a_i(x_1) a_j(x_2), \\ &= \mathbf{C}[1, 1] a_1(x_1) a_1(x_2) + \mathbf{C}[1, 2] a_1(x_1) a_2(x_2) \\ &\quad + \mathbf{C}[2, 1] a_2(x_1) a_1(x_2) + \mathbf{C}[2, 2] a_2(x_1) a_2(x_2). \end{aligned}$$

Natürlich sind die Koeffizienten ein Tensor der Ordnung 2. Aber welche Bedingungen müssen diese Koeffizienten  $\mathbf{C}[i, j]$  erfüllen, wenn die entwickelte Funktion nun (anti-)symmetrisch ist? Betrachten wir also  $f(x_1, x_2) - f(x_2, x_1) = 0$  im symmetrischen und  $f(x_1, x_2) + f(x_2, x_1) = 0$  im antisymmetrischen Fall mit der obigen Entwicklung und fassen direkt beiden Fälle in einer Gleichung zusammen:

$$\begin{aligned} &(\mathbf{C}[1, 1] \mp \mathbf{C}[1, 1]) a_1(x_1) a_1(x_2) \\ &+ (\mathbf{C}[1, 2] \mp \mathbf{C}[2, 1]) a_1(x_1) a_2(x_2) \\ &+ (\mathbf{C}[2, 1] \mp \mathbf{C}[2, 1]) a_2(x_1) a_1(x_2) \\ &+ (\mathbf{C}[2, 2] \mp \mathbf{C}[2, 2]) a_2(x_1) a_2(x_2) = 0, \end{aligned}$$

wobei wir sehen dass das Vertauschen von  $x_1$  und  $x_2$  eine entsprechende Vertauschung der Tensorindizes bedingt.

Um dies im symmetrischen Falle ( $\mp \rightarrow -$ ) für beliebige Basisfunktionen zu erfüllen, muss  $\mathbf{C}[1, 2] = \mathbf{C}[2, 1]$  gelten. Der Raum, welchen wir über das symmetrische Tensorprodukt erhalten, wird also von drei Basisfunktionen aufgespannt, nämlich  $a_1(x_1) a_1(x_2)$ ,  $a_2(x_1) a_2(x_2)$  und  $\frac{1}{2} (a_1(x_1) a_2(x_2) + a_2(x_1) a_1(x_2))$ , wobei wir die Normiertheit der Basisfunktionen beibehalten.

Im antisymmetrischen Fall ( $\mp \rightarrow +$ ) wiederum muss gelten, dass  $\mathbf{C}[1, 1] = \mathbf{C}[2, 2] = 0$  und dass  $\mathbf{C}[1, 2] = -\mathbf{C}[2, 1]$ . Es verbleibt also nur eine Basisfunktion, nämlich  $\frac{1}{2} (a_1(x_1) a_2(x_2) - a_2(x_1) a_1(x_2))$ .  $\diamond$

Wir möchten festhalten, dass für die Vertensorierung *zweier* Räume stets gilt, dass

$$\mathcal{V}^{(d)} \otimes \mathcal{V}^{(d)} = \mathcal{V}^{(d)} \overset{S}{\otimes} \mathcal{V}^{(d)} \oplus \mathcal{V}^{(d)} \overset{A}{\otimes} \mathcal{V}^{(d)}, \quad (2.3.8)$$

wie aus der Analogie zum Aufsplitten einer Matrix in symmetrische und antisymmetrische Anteile sofort klar wird.

## 2. Tensorzerlegung

Dies gilt aber nicht im allgemeineren Fall. Dann können auch gemischte Produkträume auftreten<sup>8</sup>. Mittels der Projektion über den (Anti-)Symmetrisierer, siehe Definition 3, kann jedoch stets auf den entsprechenden Teilraum abgebildet werden. Auf diese Weise erhält man zumindest ein Erzeugendensystem des gesuchten symmetrischen oder antisymmetrischen Tensorraums. Rigoros ist dies in Hamaekers [82, Abschn. 2.5] definiert.

### 2.4. Bestapproximation

Sei nun  $\mathcal{V}^{(1)}, \dots, \mathcal{V}^{(d)}$  eine Folge von Funktionenräumen mit

$$\mathcal{V}^{(k)} = \bigotimes_{i=1}^k \mathcal{V}_i^{(1)}. \quad (2.4.1)$$

Wir nehmen dabei an, dass es sich stets um den gleichen Funktionenraum  $\mathcal{V}_1^{(1)} = \mathcal{V}_2^{(1)} = \dots = \mathcal{V}^{(1)}$  handelt. Für das  $k$ -fache Tensorprodukt des gleichen Raums  $\mathcal{V}^{(1)}$  schreiben wir im Folgenden kurz  $\bigotimes^k \mathcal{V}^{(1)}$ . Damit betrachten wir eine Funktion  $f \in \mathcal{V}^{(k)}$  mit  $k$  Eingangsvariablen. Ihre Darstellung in einer Basis des  $\mathcal{V}^{(1)}$  ergibt als Koeffizienten einen Tensor der Ordnung  $k$ .

Im Folgenden geht es uns um Bestapproximationen bestimmter Funktionen. Um die Qualität einer Approximation innerhalb eines bestimmten Raums  $\mathcal{V}$  messen zu können, benötigen wir eine Norm  $\|\cdot\|_{\mathcal{V}}$ , um Abstände  $\|f - \tilde{f}\|_{\mathcal{V}}$  zwischen einer Funktion  $f \in \mathcal{V}$  und ihrer Approximation  $\tilde{f} \in \mathcal{V}$  zu bestimmen.

Wir nehmen also im Folgenden an, dass unsere Funktionenräume  $\mathcal{V}$  mit einer Norm  $\|\cdot\|_{\mathcal{V}}$  ausgestattet sind und dass Approximations- und Funktionenraum der gleiche Raum sind.

Wir bemerken, dass innere Produkträume automatisch normierte Vektorräume sind, da deren inneres Produkt  $\langle \cdot, \cdot \rangle$  auf natürliche Weise eine Norm induziert, siehe [190, S. 23],

$$\|f\| = \sqrt{\langle f, f \rangle}. \quad (2.4.2)$$

#### 2.4.1. Tensorrang

Wir sind bereits vertraut damit, dass ein Koeffizient in einer Basisdarstellung eines Tensorraums die Komponente eines Tensors ist. Damit können wir uns nun dem Begriff des Tensorrangs zuwenden.

Wir betrachten wieder eine Funktion  $f$  mit Darstellung (2.3.2) in  $\mathcal{V}^{(d)}$ . Wir bemerken zunächst, dass diese Darstellung nicht eindeutig ist, denn sie hängt von den Basisfunktionen  $\{b_i\}$  ab.

---

<sup>8</sup>Diese haben dann aber disjunkte Indexmengen, die sich unter Permutation symmetrisch bzw. antisymmetrisch verhalten, siehe Bemerkung 1.

**Definition 14** (Zerlegbarkeit einer Funktion). Eine Funktion  $f \in \mathcal{V}^{(d)} = \bigotimes^d \mathcal{V}^{(1)}$  heißt *zerlegbar*, wenn es  $d$  Vektoren(!)  $\{\mathbf{c}_i\}_{i=1}^d$  der Dimension  $n$  gibt, so dass

$$f(x_1, \dots, x_d) = \left( \sum_{j_1} c_{1,j_1} b_{j_1}(x_1) \right) \cdot \dots \cdot \left( \sum_{j_d} c_{d,j_d} b_{j_d}(x_d) \right) \quad (2.4.3)$$

gilt.

Die zerlegbare Funktion  $f$  hat also eine Produktstruktur

$$\begin{aligned} f(x_1, \dots, x_d) &= f_1(x_1) \cdot \dots \cdot f_d(x_d) & (2.4.4) \\ &= \left( \sum_{j_1} c_{1,j_1} b_{j_1}(x_1) \right) \cdot \dots \cdot \left( \sum_{j_d} c_{d,j_d} b_{j_d}(x_d) \right) \\ &= \sum_{j_1, \dots, j_d} \underbrace{(\mathbf{c}_1 \otimes \dots \otimes \mathbf{c}_d)_{j_1 \dots j_d}}_{=: \mathbf{C}_{[j_1, \dots, j_d]}} b_{j_1}(x_1) \cdot \dots \cdot b_{j_d}(x_d), \end{aligned}$$

und jedes  $f_i(x_i)$  ist mittels Koeffizientenvektors  $\mathbf{c}_i$  in den Basisfunktionen  $\{b_j\}$  dargestellt, die zum Tensorprodukt herausgezogen wurden, vgl. auch mit (2.3.2).

In anderen Worten, nur auf den Koeffiziententensor  $\mathbf{C}$  bezogen und der Definition aus de Silva und Lim [41, Definition 2.2] folgend, deren Ergebnisse zur Existenz von Bestapproximationen wir später wiedergeben:

**Definition 15** (Zerlegbarkeit eines Tensors). Eine Tensor  $\mathbf{C}$  heißt *zerlegbar*, wenn es  $d$  Vektoren  $\{\mathbf{c}_i\}_{i=1}^d$  der Dimension  $d$  gibt, so dass

$$\mathbf{C} = \mathbf{c}_1 \otimes \dots \otimes \mathbf{c}_d \quad (2.4.5)$$

gilt.

Die Gegenüberstellung der Definitionen 14 und 15 soll die Analogie der Konzepte von Funktionen in Tensorprodukträumen und Tensoren deutlich machen. Im weiteren geben wir nur noch die auf Funktionen bezogenen Definitionen an.

Man sollte nicht der Versuchung erliegen, zu glauben, dass jeder Tensor  $\mathbf{C}$  der Ordnung  $d$  durch  $d$  Vektoren dargestellt werden kann. Das Produkt der Summen in (2.4.3) ergibt nicht die große Summe in (2.3.2).

**Beispiel 4** Dies ist offensichtlich bei Tensoren der Ordnung 2, welche analog zu Matrizen betrachten werden können. So ist eben im Allgemeinen *nicht* jede Matrix  $M$  durch Summen von zwei Vektoren  $\mathbf{c}_1$  und  $\mathbf{c}_2$  darstellbar,  $\left( (\sum_i c_{1,i} \mathbf{e}_i) \cdot (\sum_j c_{2,j} \mathbf{e}_j) \right)_{ij} = (\mathbf{c}_1 \cdot \mathbf{c}_2^T)_{ij}$ , wobei  $\mathbf{e}_j$  die Einheitsvektoren des  $\mathbb{R}^d$  sind. Dies trifft nur für solche mit Rang kleiner gleich eins zu.  $\diamond$

Wir legen nun fest, was wir unter dem Rang der Tensor Darstellung einer Funktion verstehen.

## 2. Tensorzerlegung

**Definition 16** (Tensorrang). Eine Funktion  $f \in \mathcal{V} = \bigotimes^d \mathcal{V}^{(1)}$  hat den *Tensorrang*  $r$ , wenn sie als eine Summe von  $r$  zerlegbaren Funktionen und nicht mit weniger geschrieben werden kann:

$$\text{rank}_{\otimes}(f) := \min \left\{ r \mid f(x_1, \dots, x_d) = \sum_{i=1}^r f_{i_1}(x_1) \cdots f_{i_d}(x_d) \right\},$$

wobei  $f_{i_1}(x_1) \cdots f_{i_d}(x_d)$  die Produktstruktur der  $r$  zerlegbaren Funktionen mit  $f_{j_i}(x_i) \in \mathcal{V}^{(1)}$  bezeichnet.

Wir bezeichnen die Menge der Funktionen in  $\mathcal{V}$  mit  $d$  Eingangsvariablen und Rang  $r$  mit  $\mathcal{S}(d, r)$ .

Offensichtlich ist eine Funktion genau dann zerlegbar, wenn sie Tensorrang 1 hat.

### 2.4.2. Bestapproximation

Wir suchen nun nach der nächsten Funktion  $\tilde{f}$  zur gegebenen Funktion  $f$ , welche einen bestimmten Rang  $r$  hat. Man spricht dabei von Rang- $r$ -Bestapproximation. Genauer werden wir uns später auf Rang-1-Bestapproximationen konzentrieren.

**Definition 17** (Rang- $r$ -Bestapproximation). Voraussetzungen wie in Definition 16. Dann heißt die Funktion  $f$  mit Rang  $r$  *Rang- $r$ -Bestapproximation*, wenn

$$\tilde{f} = \min_{g \in \mathcal{S}(d, r)} \|f - g\|_{\mathcal{V}}$$

gilt. Wenn also in einer entsprechenden Norm  $\tilde{f}$  der gegebenen Funktion  $f$  am nächsten kommt.

Entscheidend ist nun das folgende Lemma nach de Silva und Lim [41, Korollar 4.5], das wir hier nur wiedergeben wollen:

**Lemma 2** (Existenz der Rang-1-Bestapproximation). *Jeder Tensor  $A$  hat eine Rang-1-Bestapproximation.*

de Silva und Lim [41] zeigen dies zwar nur für endlichdimensionale Vektorräume, aber dies ist für unsere Zwecke ausreichend, denn damit hat also auch jede Funktion  $f \in \mathcal{V}$  eine Rang-1-Bestapproximation in der gegebenen (endlichen) Basis.

Wir bemerken noch, dass sich Foldy [59] mit der Frage beschäftigt hat, wann eine antisymmetrische Funktion sich durch eine Rang-1-Bestapproximation darstellen lässt. Im antisymmetrischen Funktionenraum heißt diese auch Slater-Determinante.

## 2.5. Dekorrelatoren

Sei  $\mathcal{V} = \bigotimes^d \mathcal{V}^{(1)}$  der  $d$ -fach vertensorierte Raum der Funktionen mit einer Eingangsvariable  $\mathcal{V}^{(1)}$ . Sei weiterhin  $\{b_i\}$  eine Basis des  $\mathcal{V}^{(1)}$ . Mit  $f \in \mathcal{V}$  existiert nach Lemma 2 eine Funktion  $\tilde{f} \in \mathcal{V}$  mit Rang 1, welche das Minimum

$$\min_{\tilde{f} \in \mathcal{S}(d, 1)} \|f(x_1, \dots, x_d) - \tilde{f}(x_1, \dots, x_d)\|_{\mathcal{V}}$$

annimmt, also die entsprechende Bestapproximation mit Rang 1. Diese Funktion  $\tilde{f}$  muss damit eine Produktstruktur haben. Und damit hat sie nach unserer anfangs gemachten Definition 5 unkorrelierte Eingangsvariablen.

Wir definieren nun formal folgende Abbildung:

**Definition 18** (Voller Dekorrelator). Die Abbildung  $\hat{K} : \mathcal{V} \rightarrow \mathcal{V}$ , welche  $f \in \mathcal{V}$  auf seine Rang-1-Bestapproximation  $\tilde{f}$  projiziert, heißt voller *Dekorrelator*.

Die genaue Gestalt von  $\hat{K}$  hat für uns hier keine Bedeutung, nach Lemma 2 wissen wir, dass sie für jedes  $f \in \mathcal{V}$  wohldefiniert ist.

Damit möchten wir nun festlegen, was mit Korrelation gemeint sein soll und die Namensgebung der Abbildung klar werden lassen:

**Definition 19** (Korrelation). Sei  $f \in \mathcal{V}^{(d)}$  eine Funktion mit  $d$  Eingangsvariablen. Sei weiterhin  $f_0 = \hat{K}f$  die Rang-1-Bestapproximation dieser Funktion  $f$  und sei  $f_1 := (1 - \hat{K})f$ . Dann sagen wir  $f_0$  ist der unkorrelierte Anteil von  $f$  und  $f_1$  beschreibt die *Korrelation* der Eingangsvariablen von  $f$ .

Die Funktion  $f_1$  ist nun definiert und damit auch messbar. Aber mit dem Werkzeug vertensorierter Funktionenräume gelingt noch viel mehr.

Wir können diese Abbildung  $\hat{K}$  nämlich verallgemeinern und genau herausarbeiten, welche Eingangsvariablen miteinander wie stark korreliert sind. Dazu müssen wir im Folgenden aber nun wieder zwischen unterscheidbaren und ununterscheidbaren Eingangsvariablen und entsprechenden Funktionen differenzieren. Wir werden zunächst allgemein Funktionen ohne solch spezielle Symmetrie betrachten, anschließend widmen wir uns antisymmetrischen Funktionen. Die symmetrischen folgen dabei ganz analog. Wir verweisen an dieser Stelle auf unsere Bemerkung am Ende von Abschnitt 2.1.1, die wir nun erfüllt sehen werden.

Für beide Fälle müssen wir jeweils die Indexmengen der Eingangsvariablen betrachten und diese geeignet zerlegen. Weiterhin werden wir eine Untermengenrelation definieren, die uns im letzten Unterabschnitt sehr nützlich sein wird.

### 2.5.1. Dekorrelatoren für unterscheidbare Eingangsräume

Betrachten wir  $d$  unterscheidbare Funktionenräume  $\mathcal{V}_i^{(1)}$ , so kommt es auf die Reihenfolge im Tensorprodukt

$$\mathcal{V} = \bigotimes_{i=1}^d \mathcal{V}_i^{(1)} \quad (2.5.1)$$

an, welches durch den unteren Index  $i$  deutlich gemacht ist.

Sei  $\mathcal{D} = \{1, \dots, d\}$  eine Indexmenge. Sei  $\mathcal{I}_j$  einer Zerlegung dieser Indexmenge  $\mathcal{D}$  in  $d$  disjunkte Untermengen  $\mathcal{I}_{j_i}$ . Und sei schließlich  $\mathcal{I} = \{\mathcal{I}_1, \dots\}$  die Menge aller Zerlegungen  $\mathcal{I}_j$ . Das heißt, es soll für alle Zerlegungen  $\mathcal{I}_j$  gelten,

$$\{\mathcal{I}_{j_1}, \dots, \mathcal{I}_{j_d}\} = \mathcal{I}_j, \quad \bigcup_{i=1}^d \mathcal{I}_{j_i} = \mathcal{D}, \quad \text{und} \quad \forall 1 \leq i < k \leq d : \mathcal{I}_{j_i} \cap \mathcal{I}_{j_k} = \emptyset.$$

## 2. Tensorzerlegung

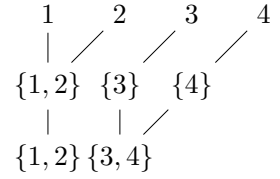
Wir möchten nun für diese Zerlegungen von  $\mathcal{D}$  eine Untermengenrelation definieren:

**Definition 20** (Untermengenrelation, Indexmengenzerlegung). Seien  $\mathcal{I}_j$  und  $\mathcal{I}_k$  aus  $\mathcal{I}$  zwei verschiedene Zerlegungen der Indexmenge  $\mathcal{D}$ . Dann gilt

$$\mathcal{I}_j \subset \mathcal{I}_k \quad (2.5.2)$$

genau dann, wenn für jede Untermenge  $\mathcal{I}_{j_i}$  aus  $\mathcal{I}_j$  eine Untermenge  $\mathcal{I}_{k_l}$  in  $\mathcal{I}_k$  existiert, so dass  $\mathcal{I}_{j_i} \subseteq \mathcal{I}_{k_l}$ .

**Beispiel 5** Betrachten wir die Zerlegungen  $\mathcal{I}_1 := \{\{1, 2\}, \{3, 4\}\}$  und  $\mathcal{I}_2 := \{\{1, 2\}, \{3\}, \{4\}\}$  von  $\mathcal{D} := \{1, 2, 3, 4\}$ , so gilt  $\mathcal{I}_2 \subset \mathcal{I}_1$ , denn  $\{1, 2\} \subset \{1, 2\}$  und  $\{3\}, \{4\} \subset \{3, 4\}$ , siehe auch die Abbildung rechts.  $\diamond$



Damit betrachten wir nun folgenden vertensorierten Raum zu einer Indexuntermenge  $\mathcal{I}_{j_i}$

$$\mathcal{V}_{\mathcal{I}_{j_i}}^{(|\mathcal{I}_{j_i}|)} := \bigotimes_{l \in \mathcal{I}_{j_i}} \mathcal{V}_l^{(1)},$$

wobei  $|\mathcal{I}_{j_i}|$  die Mächtigkeit der Indexmenge  $\mathcal{I}_{j_i}$  bezeichnet.

Ohne Beschränkung der Allgemeinheit stellen wir uns vor, dass die Indizes von  $\mathcal{I}_{j_i}$  die letzten  $|\mathcal{I}_{j_i}|$  Indizes von  $\mathcal{D}$  sind<sup>9</sup>, so dass wir

$$\bigotimes_{i=1}^d \mathcal{V}_i^{(1)} = \bigotimes_{l \notin \mathcal{I}_{j_i}} \mathcal{V}_l^{(1)} \otimes \mathcal{V}_{\mathcal{I}_{j_i}}^{(|\mathcal{I}_{j_i}|)} =: \mathcal{V}' \quad (2.5.3)$$

schreiben können. Soweit ist dies nur Notation: Der Raum  $\mathcal{V}'$  in (2.5.3) ist der gleiche wie der volle Tensorraum  $\mathcal{V}$  in (2.5.1).

Wir wissen jedoch, dass die Rang-1-Bestapproximation  $\hat{f}$  auch in  $\mathcal{V}'$  existiert. Sei also  $f \in \mathcal{V}$ , dann existiert deren Rang-1-Bestapproximation

$$\hat{f} = f_1(x_1) \cdot \dots \cdot f_{d-|\mathcal{I}_i|}(x_{d-|\mathcal{I}_i|}) \cdot g(x_{d-|\mathcal{I}_i|+1}, \dots, x_d).$$

Die Funktion  $g$  muss nun nicht weiter zerlegbar sein, insbesondere nicht in  $f_{d-|\mathcal{I}_i|+1}, \dots, f_d$ , die wir aus der Rang-1-Bestapproximation  $\tilde{f}$  in  $\mathcal{V}$  erhalten hätten. Damit sind die Variablen zur Indexmenge  $\mathcal{I}_i$  immer noch möglicherweise korreliert, aber nur untereinander und nicht mehr mit Variablen aus dem Komplement  $\mathcal{D} \setminus \mathcal{I}_i$ . Offensichtlich sind die restlichen Funktionen  $f_i \in \mathcal{V}^{(1)}$  mit  $i \notin \mathcal{I}_{j_i}$  von einer formal anderen Unterteilung unberührt und sind daher in beiden Bestapproximationen gleich, denn die Bestapproximation von  $g$  muss wieder zu  $f_{d-|\mathcal{I}_i|+1} \cdot f_d$  führen.

<sup>9</sup>Ist dies nicht der Fall, so müssen Eingangsräume und Eingangsvariablen entsprechend permutiert werden.



Übertragen wir diese Gedanken in Funktionenräumen auf Tensoren in folgendem Beispiel:

**Beispiel 6** Seien alle Eingangsräume  $\mathcal{V}_i^{(1)}$  gleich  $\mathcal{V}^{(1)}$  mit einer Basis  $\{b_i\}$ . Betrachten wir die Indexuntermenge  $\mathcal{J} = \{d-1, d\}$  und den Raum  $\mathcal{V}_{\mathcal{J}}^{(2)} = \mathcal{V}_{d-1}^{(1)} \otimes \mathcal{V}_d^{(1)}$  und entsprechend die Räume  $\mathcal{V}$  und  $\mathcal{V}'$  wie oben. Sei

$$\tilde{f}(x_1, \dots, x_d) = f_1(x_1) \cdot \dots \cdot f_d(x_d)$$

die Rang-1-Bestapproximation in  $\mathcal{V}$  mit  $f_i \in \mathcal{V}^{(1)}$  und

$$\hat{f}(x_1, \dots, x_d) = f_1(x_1) \cdot \dots \cdot f_{d-2}(x_{d-2}) \cdot \varphi(x_{d-1}, x_d)$$

die Rang-1-Bestapproximation in  $\mathcal{V}'$  mit  $\varphi \in \mathcal{V}_{\mathcal{J}}^{(2)}$ .

Was ist der genaue Unterschied zwischen  $\tilde{f}$  und  $\hat{f}$ ? Betrachten wir dazu die Darstellung von  $\varphi$  in  $\mathcal{V}_{\mathcal{J}}^{(2)}$ ,

$$\varphi(x_{d-1}, x_d) = \sum_{i_1, i_2} c[i_1, i_2] \cdot \xi_{i_1} \cdot \xi_{i_2}.$$

Analog existiert hier die Rang-1-Bestapproximation des Tensors  $c[i_1, i_2]$ , d. h. es existiert ein  $\tilde{\varphi} = (c_1 \otimes c_2)(\xi_1 \cdot \xi_2)$ . Im Unterraum  $\mathcal{V}_{\mathcal{J}}^{(2)}$  ist also die Einschränkung auf einen zerlegbaren Untertensor aufgehoben. Hier darf der Koeffizientenvektor also eine volle Matrix sein und ist nicht auf das Produkt zweier Vektoren eingeschränkt. Dementsprechend muss  $g$  eine mindestens genauso gute Approximation wie  $\tilde{f}$  ergeben, womöglich aber eine bessere durch diese größere Freiheit.  $\diamond$

Durch den formalen Trick, Untertensorprodukte zusammen zu fassen, heben wir also den Zwang zur Abbildung auf eine zerlegbare Funktion dort auf.

Nun können wir die Verallgemeinerung zum vollen Dekorrelator aus Definition 18 für unterscheidbare Eingangsräume definieren:

**Definition 21** (Teildekorrelator). Sei  $\mathcal{I}_j$  eine Zerlegung der Indexmenge  $\mathcal{D}$  in disjunkte Unterindexmengen. Sei  $\mathcal{I}_{j_i}$  eine Untermenge dieser Zerlegung und weiterhin  $\mathcal{V}_{\mathcal{I}_{j_i}}^{(|\mathcal{I}_{j_i}|)} = \bigotimes_{k \in \mathcal{I}_{j_i}} \mathcal{V}_k^{(1)}$  der  $|\mathcal{I}_{j_i}|$ -fach vertensorierte Raum über Funktionen mit einer Eingangsvariable. Entsprechend sei für eine Zerlegung  $\mathcal{I}_j$  die Vereinigung der vertensorierten Räume all ihrer Untermengen,

$$\mathcal{V}_{\mathcal{I}_j}^{(d)} = \bigotimes_{i=1}^d \bigotimes_{k \in \mathcal{I}_{j_i}} \mathcal{V}_k^{(1)},$$

wieder der Raum über Funktionen mit  $d$  Eingangsvariablen.

Dann heißt  $\hat{P}_{\mathcal{I}_j}^{-1} \hat{K}_{\mathcal{I}_j} \hat{P}_{\mathcal{I}_j}$  der *Teildekorrelator*, der eine Funktion  $f \in \mathcal{V}^{(d)}$  auf die Rang-1-Bestapproximation in  $\mathcal{V}_{\mathcal{I}_j}^{(d)}$  abbildet. Dabei bezeichnet  $\hat{P}_{\mathcal{I}_j}$  den Permutationsoperator, welche die Indexmengen in  $\mathcal{I}_j$  in eine aufeinanderfolgende, kanonische Reihenfolge  $\mathcal{I}_{j_1}, \mathcal{I}_{j_2}, \dots$  bringt, und  $\hat{P}_{\mathcal{I}_j}^{-1}$  den zugehörigen invers wirkenden Operator.

### 2.5.2. Dekorrelatoren für ununterscheidbare Eingangsräume

Nun wenden wir uns ununterscheidbaren Eingangsräumen zu. Konsequenterweise sei also  $\mathcal{V}^{(1)}$  der Raum der Funktionen mit einer Eingangsvariable, der wieder  $d$ -fach vertensoriert wird.

Wieder möchten wir uns entsprechende Zerlegungen einer Indexmenge  $\mathcal{D} = \{1, \dots, d\}$  definieren. Sind die Eingangsräume ununterscheidbar, dann spielt es keine Rolle, welche Indizes nun genau zusammengefasst werden, sondern nur wie viele.

Die Anzahl der möglichen Zerlegungen korrespondiert deswegen mit der Anzahl der Zerlegungen der Ganzzahl  $d$ , siehe [8] für eine Einführung. Zerlegungen der Ganzzahl  $d$  werden üblicherweise mit  $1^{d-2}2$  oder auch  $1^{d-2} + 2$  oder auch  $\underbrace{1 + \dots + 1}_{d-2\text{-mal}} + 2$  angegeben.

Deswegen sei also  $\mathcal{I}$  in diesem Fall die Menge aller Zerlegungen der Ganzzahl  $d$ , also  $\{1^d, 1^{d-2}2, 1^{d-3}3, \dots, 1^{d-4}2^2, \dots\}$ .

Aus der Kombinatorik ist einiges über diese Zerlegungen bekannt, da sie viele Gemeinsamkeiten mit der symmetrischen Gruppe und Permutationen haben, siehe Artin [10] und Herstein [92, Lemma 2.11.3]. Uns interessiert an dieser Stelle aber wiederum eine Untermengenrelation dieser Ganzzahlzerlegungen, wie wir sie auch schon für die Zerlegungen von Indexmengen definiert haben.

**Definition 22** (Untermengenrelation, Ganzzahlzerlegungen). Sei  $d$  eine Ganzzahl, seien  $\mathcal{I}_j, \mathcal{I}_k$  zwei Zerlegungen dieser Ganzzahl. Dann gilt

$$\mathcal{I}_j \subset \mathcal{I}_k \tag{2.5.4}$$

genau dann, wenn es für jede Zahl  $l \in \mathcal{I}_k$  wieder eine Ganzzahlzerlegung derart gibt, so dass die gemeinsame Menge dieser Zerlegungen genau  $\mathcal{I}_j$  ergibt.

**Beispiel 7** Beispielhaft betrachten wir  $n = 4$  und die beiden Ganzzahlzerlegungen  $2^2$  und  $1^22$ . Offensichtlich können wir  $2$  in  $1^2$  und  $2$  in  $2$  zerlegen und damit ist die zweite Ganzzahlzerlegung in der ersten enthalten.  $\diamond$

Wir erinnern an dieser Stelle an das Beispiel 6 mit der Rang-1-Bestapproximation der Darstellung einer Funktion im  $\mathcal{V}^{(2)}$ , dies enthält analog das Produkt der Rang-1-Bestapproximationen der vertensorierten Räume  $\mathcal{V}^{(1)}$ .

Analog zum Fall der unterscheidbaren Eingangsräume können wir nun die Verallgemeinerung zum vollen Dekorrelator aus Definition 18 für ununterscheidbare Eingangsräume definieren:

**Definition 23** ((Anti-)Symmetrischer Teildekorrelator). Sei  $\mathcal{I}_j$  eine Zerlegung der Ganzzahl  $d$ . Sei  $\mathcal{I}_{j_i}$  eine Untermenge dieser Zerlegung und weiterhin  $\mathcal{V}_{\mathcal{I}_{j_i}}^{(|\mathcal{I}_{j_i}|)} = \bigotimes_{k \in \mathcal{I}_{j_i}} \mathcal{V}_k^{(1)}$  der  $|\mathcal{I}_{j_i}|$ -fach vertensorierte Raum über Funktionen mit einer Eingangsvariable. Entsprechend sei für eine Zerlegung  $\mathcal{I}_j$  die Vereinigung der vertensorierten Räume all ihrer Untermengen,

$$\mathcal{V}_{\mathcal{I}_j}^{(d)} = \bigotimes_{i=1}^d \bigotimes_{k \in \mathcal{I}_{j_i}} \mathcal{V}_k^{(1)},$$

wieder der Raum über Funktionen mit  $d$  Eingangsvariablen.

Dann heißt  $\hat{K}_{\mathcal{I}_j}^S = \hat{S}\hat{K}_{\mathcal{I}_j}$  der *symmetrische Teildekorrelator* bzw.  $\hat{K}_{\mathcal{I}_j}^A = \hat{A}\hat{K}_{\mathcal{I}_j}$  der *antisymmetrische Teildekorrelator*, der eine Funktion  $f \in \mathcal{V}^{(d)}$  auf die (anti-)symmetrische Rang-1-Bestapproximation in  $\mathcal{V}_{\mathcal{I}_j}^{(d)}$  abbildet.

Wir schließen mit einem Beispiel zu den Dekorrelatoren ab.

**Beispiel 8** Betrachten wir  $f(x_1, x_2)$  und lassen den Dekorrelator zur Zerlegung  $\mathcal{I} = \{\{1\}, \{2\}\}$  darauf wirken, wobei die zugehörigen Permutationsoperatoren zur Identität werden,

$$\hat{K}_{\mathcal{I}}f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2).$$

Wir erhalten die Rang-1-Bestapproximation von  $f$  im Falle von unterscheidbaren Eingangsvariablen. Wären sie ununterscheidbar und die Funktion  $f$  antisymmetrisch, so ergäbe sich,

$$\hat{K}_{\mathcal{I}}^A f(x_1, x_2) = \hat{A}(f_1(x_1) \cdot f_2(x_2)) = f_1(x_1) \cdot f_2(x_2) - f_1(x_2) \cdot f_2(x_1).$$

◇

### 2.5.3. Korrelationsentwicklung

Ob nun eine Funktion  $f \in \mathcal{V}^{(d)}$   $d$  unterscheidbare oder ununterscheidbare Eingangsvariablen hat, mit den Untermengenrelationen aus Definition 20 und Definition 22 können wir nun eine elegante Zerlegung definieren

$$f(x_1, \dots, x_d) = \sum_{\mathcal{I}_j \in \mathcal{I}} \tilde{f}_{\mathcal{I}_j}(x_1, \dots, x_d) \quad (2.5.5a)$$

mit dem im Folgenden als *Korrelationsanteil* bezeichneten und mit dem Symbol  $\tilde{\phantom{x}}$  versehenen

$$\tilde{f}_{\mathcal{I}_j}(x_1, \dots, x_d) = \hat{K}_{\mathcal{I}_j}f(x_1, \dots, x_d) - \sum_{\mathcal{I}_k \subsetneq \mathcal{I}_j} \tilde{f}_{\mathcal{I}_k}(x_1, \dots, x_d), \quad (2.5.5b)$$

wobei wir den Dekorrelator allgemein mit  $\hat{K}$  schreiben. An dieser Stelle benötigen wir also die Untermengenrelation  $\subsetneq$  zwischen den Zerlegungen. Die Entwicklung gilt sowohl für unterscheidbare als auch für ununterscheidbare Eingangsräume, je nach entsprechender Zerlegung  $\mathcal{I}_j$  und Wahl des Dekorrelators nach Definition 21 bzw. 23.

Nun wird die Bezeichnung *Dekorrelatoren* für die Abbildungen  $\hat{K}$  offensichtlich: Für den Fall unterscheidbarer Eingangsvariablen ist der erste Summand  $f_{\{1, \dots, d\}}(x_1, \dots, x_d)$  in (2.5.5a) die Abbildung auf die Rang-1-Bestapproximation in  $\bigotimes_{j=1}^d \mathcal{V}_j^{(1)}$  und damit die Abbildung auf die unkorrelierte Produktfunktion. Wiederum für ununterscheidbare Eingangsvariablen ist der erste Summand  $f_{1^n}(x_1, \dots, x_d)$  die Abbildung auf die Rang-1-Bestapproximation in  $\bigotimes_{j=1}^{S,d} \mathcal{V}_j^{(1)}$  im symmetrischen Fall und in  $\bigotimes_{j=1}^{A,d} \mathcal{V}_j^{(1)}$  im antisymmetrischen Fall. Die resultierende Produktform ist dabei natürlich symmetrisch bzw. antisymmetrisch wie auch die ursprüngliche Funktion. Alle nachfolgenden Summanden tragen

## 2. Tensorzerlegung

immer mehr Korrelation hinein, bis die ganze Summe schließlich die vollständig korrelierte Funktion  $f$  liefert.

Ein Nutzen ergibt sich natürlich, wenn eine hochdimensionale Funktion  $f$  nur in Teilen korreliert ist und deswegen durch eine Summe von Funktionen mit teilweise Produktstruktur dargestellt werden kann. Diese Produktstruktur würde besonders bei der Integration und Differentiation vorteilhaft sein, denn bspw. bei einer unkorrelierten Funktion zerfällt ein Integral über den  $\mathbb{R}^n$  in  $d$  unabhängige Integralfaktoren über  $\mathbb{R}$ .

Die beiden Gleichungen (2.5.5a) und (2.5.5b) sind auch unter dem Begriff ANOVA-Zerlegung [76] bekannt, nach Hoeffding [94] für „ANalysis Of VAriance“. Aber der Trick der Zerlegung einer Funktion in eine Summe von Differenzanteilen kennt noch zahlreiche andere Namen. Beispielsweise nennt [169, Abschn. 5.4.1] diesen Trick „Moebius Inversion“ bekannt aus der Zahlentheorie. Wir werden sie im Folgenden als *Korrelationsentwicklung* bezeichnen, weil Korrelation ein sehr prägnanter Begriff in dem von uns betrachteten physikalisch-chemischen Themenkreis ist. Wir werden darauf in Kapitel 8 eingehen, wenn wir uns mit Methoden zur Bestimmung der Elektronenkorrelation beschäftigen.

Abschließend bemerken wir, dass Koordinatensystem und Basis großen Einfluss auf den Nutzen der Korrelationsentwicklung haben. Die Kenntnis der Struktur des Problems muss hierin einfließen.

### 2.5.4. Verwandte Literatur

Wir wollen abschließend noch kurz auf verwandte Ansätze in der Literatur eingehen. Eine tiefergehende Darstellung ist dabei aber aufgrund des Umfangs des Materials aus Statistischer Mechanik und Informationstheorie an dieser Stelle nicht möglich.

Kirkwood [113, 114] stellte bei der Betrachtung von Gasgemischen und flüssigen Lösungen die sogenannte Kirkwood-Approximation vor. Sie drückt die Wahrscheinlichkeitsdichtefunktion durch eine Integralgleichung über  $n$  Variablen aus, welche durch eine Abschlussgleichung zu einer Approximation bestimmter Güte trunziert wird. Dadurch findet er für die Dichtefunktion über drei Körper eine Näherung durch das Produkt von Dichtefunktionen von zwei Körpern.

Auch Bethe [20] beschäftigte sich mit einer ähnlichen Näherung bei der thermodynamischen Betrachtung der periodischen Zellen von Legierungen und stellte Ordnungsrelationen auf.

Kikuchi [112, 111] verallgemeinerte diese Näherung von Bethe und Kirkwood [21] bei der Betrachtung von Spin-Gitter-Modellen. Dabei nennt er Zwei-Körper-Terme Bindung, Drei-Körper-Terme Winkel und Vier-Körper-Terme Quadrate. Produkte dieser Terme ergeben Näherungen für die Wahrscheinlichkeit eines Spins an einem Gitterort.

Wir bemerken an dieser Stelle, dass aus den obigen Darstellung mittels Produkte Summen werden, und damit die Reihenentwicklung der uns bekannten Form, wenn anstelle der Dichtefunktion  $p$  die freie Energie  $W$  betrachtet wird,

$$p(\mathbf{R}_1, \dots, \mathbf{R}_n) = \exp(-W(\mathbf{R}_1, \dots, \mathbf{R}_n)/kT),$$

wobei  $k$  die Boltzmann-Konstante und  $T$  die Temperatur des thermodynamischen Systems ist, siehe [114, S. 395]. Im Sinne der statistischen Mechanik wird aus einer Energie in Bezug zur Temperatur des Systems mit der Exponentialfunktion eine Wahrscheinlichkeit.

Watanabe [182] analysierte multivariate Korrelation und stellte einige Theoreme zur Reihenentwicklung der hoch-dimensionalen Funktion auf. Diese stimmen mit der von uns gemachten Darstellung grundsätzlich überein.

Rabitz und Aliş [146] beschäftigten sich ebenfalls mit einer Reihenentwicklung der von uns dargestellten Gestalt. Ähnlich wie unsere Bemerkungen eingangs des Kapitels verstehen sie diese Entwicklung als eine allgemeine Möglichkeit der nieder-dimensionalen Darstellung des hoch-dimensionalen Zusammenhangs eines physikalischen Modells und gibt eine ausführliche Einleitung zu diesem Ansatz der linearen oder nicht-linearen Superposition von Funktionen.

In Somani et al. [160] findet sich eine Arbeit zur Wahrscheinlichkeitsdichtefunktion von Molekülen, welche sich auf eine Verallgemeinerung der Kirkwood-Approximation [114] bezieht. Diese Arbeit ist insbesondere in der Notation der Zerlegung von Indexmenge sehr nahe an unserer Darstellung zu unterscheidbaren Eingangsvariablen als die zuvor genannten frühen Arbeiten.

## Zusammenfassung

In diesem Kapitel beschäftigten wir uns mit Tensorprodukten von Funktionenräumen und der Darstellung von Funktionen in diesen Räumen, deren Koeffizienten Tensoren der entsprechenden Ordnung darstellen. Entscheidend ist dabei der Gedanke der Rang-1-Bestapproximation, in welchem Falle dieser Tensor als ein Produkt von Vektoren darstellbar ist. Für die mittels dieses Tensors dargestellte Funktion  $f$  bedeutete dies, dass sie eine Produktgestalt aufweist. In genau diesem Fall definierten wir ihre Eingangsvariablen als unkorreliert. Anschließend war es uns mittels geeigneter Zerlegungen und einer Untermengenrelation gelungen, eine Entwicklung einer allgemeinen Funktion  $f$  dergestalt zu definieren, dass deren Terme höherer Ordnung eine gleichfalls höhere Ordnung von Korrelation der Funktion hinzufügen. Diese Entwicklung kann im Anschluss die Grundlage bilden, um eine effiziente Darstellbarkeit der allgemeinen Funktion zu untersuchen. Wir bemerken, dass im Umfeld der theoretischen Chemie eine genau umgekehrte Entwicklung mittels Anregungsoperators bekannt ist. Wir werden darauf in Kapitel 8 eingehen.

Die in diesem Kapitel gemachten Ausführungen waren dabei von sehr grundlegender Natur. Wir sind nicht genauer auf die Natur des Problems und damit des Operators  $\hat{L}$ , siehe Abschnitt 2.1, eingegangen, deren Lösung die dargestellte Funktion  $f$  in den vertensorierten Funktionenräumen sein soll. Wir betrachteten nicht, ob der Operator separabel ist, siehe Hackbusch [80, S. 11], denn dies hängt von der speziellen Struktur des Operators selbst ab. Wir werden dazu im nächsten Kapitel konkret den Hamiltonoperator untersuchen und erläutern wie sich mit der trunkierten Korrelationsentwicklung Näherungslösungen definieren lassen. Wir betonen, dass eine trunkierte Entwicklung nur die Korrelation der möglichen Lösung und nicht den Operator selbst einschränkt.

## 2. Tensorzerlegung

Auch beschäftigten wir uns nicht damit, was genau eine hochdimensionale Funktion letztlich ausmacht oder berührten Themen wie effektive Dimension, siehe dazu auch Holtz [96]. Letztere Frage beschäftigt sich damit, ob eine Funktion in gegebenen Dimensionen überhaupt all diese Eingangsfreiheitsgrade braucht oder ob man sie dergestalt transformieren kann, dass sie weniger, wenn auch andere Eingangsvariablen benötigt, siehe Oettershagen [139] für eine mögliche Vorgehensweise.

Weitergehende Fragestellungen werden an anderer Stelle besprochen: Eine ausführlichere Darstellung findet sich natürlich bei Hackbusch [80] für die Tensorzerlegung, wo der Fokus auf die Tensorapproximation gelegt wird. Kolda und Bader [122] geben eine aktuelle Übersicht über dieses noch recht junge Thema. In diesem Zusammenhang ist auch [138] lesenswert. In [82] werden rigoros (antisymmetrische) Teilchenräume betrachtet, um geeignete Approximationen im Rahmen von Dünngitterdarstellungen herzuleiten.

Unsere Absicht ist zunächst die Benutzung dieser Korrelationsentwicklung (2.5.5a), welche wir trunkiert als formale Ansatzfunktion für eine gesuchte Lösung  $f$  im Folgenden benutzen werden. Der Ansatz wird uns dabei helfen, die Separabilität eines Problems zu erkennen bzw. ein lineares, aber hochdimensionales Problem in ein nicht-lineares, aber niederdimensionales Problem zu überführen.

## 3. Schrödinger-Gleichung

Im Folgenden betrachten wir die zeitabhängige Schrödingergleichung als grundlegende Modellgleichung eines quantenmechanischen Systems aus  $n$  Teilchen. Wir überblicken in Abschnitt 3.1 zunächst den zugrundeliegenden Konzepten von Zustand und Zustandsfunktion und betrachten allgemein den Hamiltonoperator, welcher die Dynamik bestimmt. In Abschnitt 3.2 stellen wir konkret den Hamiltonoperator für ein System aus Kernen und Elektronen auf. Anschließend widmen wir uns in Abschnitt 3.3 den beiden adiabatischen Ausprägungen, Born-Oppenheimer (BO) und Ehrenfest, der klassischen Moleküldynamik. Dabei wird der gemeinsamen Zustand von Kernen und Elektronen durch einen unkorrelierten Produktzustand angesetzt, den wir formal mit der trunkierten Korrelationsentwicklung des letzten Kapitels beschreiben. Für die als klassisch betrachteten Kerne verbleibt dann nur noch ein Poisson-Problem zu lösen, wenn die Lösung der elektronischen Schrödingergleichung bekannt ist. Zur Bestimmung dieser Lösung wiederum machen wir uns ebenso die zuvor eingeführten Dekorrektoren zunutze, welche die elektronische Wellenfunktion auf eine antisymmetrisierte Produktform überführen. Dies bildet den Anschluss zum folgenden Kapitel, wo diese Produktform in die bekannte Hartree-Fock (HF)-Gleichung mündet.

Wir werden in diesem Kapitel die Begriffe von der physikalischen Seite einführen. Anhand der Notation wird aber die Korrespondenz zu den zuvor gemachten allgemeinen Ausführungen der Zerlegung einer hochdimensionalen Funktion deutlich werden.

### 3.1. Zeitabhängige Schrödinger-Gleichung

Wir werden uns zunächst mit den allgemeinen Begriffen von Zustand, Zustandsraum und Zustandsfunktion befassen und anschließend die zeitabhängige Schrödingergleichung als jene Gleichung erfahren, welche die Dynamik dieses Systems beschreibt. Einen zentralen Platz nimmt dabei der Hamiltonoperator ein, welchen wir aber zunächst nur allgemein einführen. Die hochdimensionale Komplexität der Gleichung entsteht vor allem durch die Korrelation der Eingangsvariablen. Einfachere Modellsysteme entstehen, indem wir die als Ansatzfunktion eingesetzte Korrelationsentwicklung auf bestimmte Art ansetzen und abschneiden.

#### 3.1.1. Zustand und Zustandsfunktion

Wir folgen in diesem Abschnitt zunächst der Herleitung in [77, Abschn. 2.1]. Das Hauptwesen der Quantenmechanik verkörpert die sogenannte *Zustandsfunktion* oder Wellenfunktion:

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_n, t). \quad (3.1.1)$$

### 3. Schrödinger-Gleichung

Sie beschreibt als komplexwertige Funktion  $\Psi : \underbrace{\mathcal{V} \otimes \dots \otimes \mathcal{V}}_{n\text{-mal}} \otimes \mathbb{R} \rightarrow \mathbb{C}$  den *gleichzeitigen* Zustand von  $n$  Teilchen zum Zeitpunkt  $t$ . Jedwede Aussage über das System der Teilchen lässt sich aus ihr ableiten. Die Eingangsvariablen  $\mathbf{x}$  seien dabei aus einem später noch näher zu bestimmendem Vektorraum  $\mathcal{V}$ .

Mit Zustand ist dabei gemeint, dass der Ausdruck

$$\Psi^*(\mathbf{x}_1, \dots, \mathbf{x}_n, t) \Psi(\mathbf{x}_1, \dots, \mathbf{x}_n, t) d\mathbf{x}_1 \dots d\mathbf{x}_n \quad (3.1.2)$$

die Wahrscheinlichkeit beschreibt, die  $n$  Teilchen zum Zeitpunkt  $t$  in dem Volumenelement  $d\mathbf{x}_1 \dots d\mathbf{x}_n$  um die Koordinaten  $\mathbf{x}_1, \dots, \mathbf{x}_n$  des sogenannten Zustandsraumes vorzufinden. Dabei bezeichnet  $\Psi^*$  das komplex konjugierte der Gesamtwellenfunktion.

#### 3.1.2. Dynamik

Bezeichne  $\hat{H}$  den diesem System aus  $n$  Teilchen zugeordneten Hamiltonoperator,

$$\hat{H} = \hat{T} + \hat{V}, \quad (3.1.3)$$

welcher sich aus einem Kinetischen Operator  $\hat{T}$  und einem Potentialoperator  $\hat{V}$  zusammensetzt.

Dann beschreibt nach Schrödinger [150] die folgende zeitabhängige Schrödingergleichung die Dynamik des Systems:

$$i\hbar \frac{\partial \Psi(\mathbf{x}_1, \dots, \mathbf{x}_n, t)}{\partial t} = \hat{H} \Psi(\mathbf{x}_1, \dots, \mathbf{x}_n, t), \quad (3.1.4)$$

wobei  $i$  die imaginäre Einheit bezeichnet.

Wir wollen nun die Korrelationsentwicklung aus Kapitel 2 benutzen. Dazu definieren wir zunächst zwei Gruppen von Eingangsvariablen, nämlich  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  und  $\{t\}$ , also Orte und Zeit. Mit unseren Dekorrelatoren nach Definition 21 und der Reihenentwicklung aus 2.5.3 können wir dann folgenden Ansatz für die Zustandsfunktion formulieren:

$$\begin{aligned} \Psi(\mathbf{x}_1, \dots, \mathbf{x}_n, t) &= \sum_{\mathcal{I}_j} \tilde{\psi}_{\mathcal{I}_j}(\mathbf{x}, t) \\ &= \psi_1(\mathbf{x}) \cdot \psi_2(t) + \tilde{\psi}_{1,2}(\mathbf{x}, t), \end{aligned} \quad (3.1.5)$$

wobei hier  $d = 2$  ist und wir den Korrelationsanteil des ersten Terms ausgeschrieben haben. Damit erhalten wir nur folgende mögliche (unterscheidbare) Zerlegungen dieser Indexmenge  $\mathcal{D} = \{1, 2\}$ , nämlich  $\mathcal{I}_1 = \{\{1\}, \{2\}\}$  und  $\mathcal{I}_2 = \{1, 2\}$ .

Wir betonen, dass die eingesetzte Entwicklung die zugelassene Korrelation in der Lösung beschränkt, sie verändert nicht den Operator selbst.

Die Korrelation von Eingangsvariablen in der wahren Lösung ergibt sich stets daraus, wie diese im Operator verknüpft sind. Ein einfacher Zusammenhang besteht, wenn der Hamiltonoperator nicht explizit von der Zeit  $t$  abhängt. Denn dann sind die Orte



und die Zeit nicht miteinander korreliert und wir können die Entwicklung (3.1.5) ohne Einschränkung nach dem ersten Term abbrechen:

$$\Psi_{\mathcal{I}_1}(\mathbf{x}_1, \dots, \mathbf{x}_n, t) = \psi(\mathbf{x}_1, \dots, \mathbf{x}_n) \cdot f(t), \quad (3.1.6)$$

wobei die neue Wellenfunktion  $\psi$  zeitunabhängig ist. Es resultiert natürlich eine Produktgestalt und wir können deswegen die Gruppen von Eingangsvariablen in der Zustandsfunktion separieren, siehe Beispiel 2 in Abschnitt 2.1.2. Einsetzen in (3.1.4) und alle zeitabhängigen Terme auf die linke Seite bringen, gestattet dann die Separation der Variablen in zwei unabhängige Gleichungen bezüglich einer gemeinsamen Konstanten  $E$ : Wir erhalten für die linke Seite eine Differentialgleichung, welche die zeitliche Entwicklung der Wellenfunktion beschreibt. Ihre allgemeine Lösung ist bekanntlich  $f(t) = c \cdot \exp(iEt)$ . Die zweite Gleichung auf der rechten Seite ist die stationäre Schrödingergleichung:

$$\hat{H}\psi(\mathbf{x}_1, \dots, \mathbf{x}_n) = E\psi(\mathbf{x}_1, \dots, \mathbf{x}_n), \quad (3.1.7)$$

Dies ist die klassische Betrachtung stationärer Zustände der Schrödingergleichung für die Gesamtwellenfunktion  $\Psi$ , siehe Schwabl [151, Abschn. 2.8.1]. Aber völlig analog lassen sich auch andere Gruppen von Eingangsvariablen definieren, die nicht miteinander korreliert sein sollen, wie wir im Folgenden sehen werden.

## 3.2. Systeme von Kernen und Elektronen

Die obige Herleitung war soweit noch völlig allgemein, d. h. wir haben keine näheren Aussagen zu der Natur der Teilchen und damit auch nicht zum Raum  $\mathcal{V}$  der Eingangsvariablen gemacht. Wir befassen uns von nun an konkret mit Systemen aus  $M$  Kernen und  $N$  Elektronen unter Ausschluss von relativistischen Effekten. Der Raum der Kernteilchen ist damit nun  $\mathcal{V}^{(\text{core})} := \mathbb{R}^3 \otimes \mathbb{Z} \otimes \mathbb{R}^+$ , d. h. die verallgemeinerte Kernkoordinate fasst die Ortskoordinate  $\mathbf{R} \in \mathbb{R}^3$ , die Kernladung  $Z \in \mathbb{Z}$  und die Kernmasse  $m \in \mathbb{R}^+$  zusammen. Der Raum der Elektronen ist  $\mathcal{V}^{(\text{el})} = \mathbb{R}^3 \otimes \{\uparrow, \downarrow\}$ , d. h. er fasst die Ortskoordinate  $\mathbf{r} \in \mathbb{R}^3$  und den Spin des Elektrons  $\mathbf{s} \in \{\uparrow, \downarrow\}$  zusammen.

Der Hamiltonoperator aus (3.1.3) nimmt dann folgende konkrete Gestalt in atomaren Einheiten an, siehe dazu [168, Abschn. 2.1.1]:

$$\begin{aligned} \hat{H}(\mathbf{X}, \mathbf{x}) &= \sum_{k < j}^M \frac{Z_k Z_j}{\|\mathbf{R}_k - \mathbf{R}_j\|} - \frac{1}{2} \sum_{k=1}^M \Delta_{\mathbf{R}_k} \\ &\quad - \frac{1}{2} \sum_{k=1}^N \Delta_{\mathbf{r}_k} + \sum_{k < j}^N \frac{1}{\|\mathbf{r}_k - \mathbf{r}_j\|} - \sum_{k=1}^N \sum_{j=1}^M \frac{Z_j}{\|\mathbf{r}_k - \mathbf{R}_j\|}, \end{aligned} \quad (3.2.1)$$

wobei wir die parametrische Abhängigkeit des Hamiltonoperator von den Eingangsvariablen betont haben und dabei die abkürzende Schreibart von [77] übernehmen:  $\mathbf{X} := \mathbf{X}_1, \dots, \mathbf{X}_M$  und analog für  $\mathbf{x}$ . Jeden einzelnen Summanden in (3.2.1) abgekürzt können wir auch wie folgt schreiben:

$$\hat{H} = \hat{V}^{(\text{core,core})} + \hat{T}^{(\text{core})} + \hat{T}^{(\text{el})} + \hat{V}^{(\text{el,el})} + \hat{V}^{(\text{core,el})}. \quad (3.2.2)$$

### 3. Schrödinger-Gleichung

Wir bemerken, dass der Hamiltonoperator nicht explizit vom Spin der Elektronen abhängt, weil wir relativistische Effekte ausgeschlossen haben und keine äußeren Magnetfelder wirken. Speziell ist der Hamiltonoperator auch zeitunabhängig, was uns zu Gleichung (3.1.7) führt. Mit der zuvor beispielhaft ausgeführten Zerlegung der Zustandsfunktion  $\Psi$  in nicht-korrelierte und (teil-)korrelierte Anteile, können wir jedoch einen anderen Weg wählen, nämlich hin zur klassischen Moleküldynamik.

Dazu beginnen wir bei der zeitabhängigen Zustandsfunktionen des Systems aus  $M$  Kernen und  $N$  Elektronen,

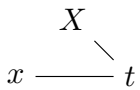
$$\Psi(\mathbf{X}_1, \dots, \mathbf{X}_M, \mathbf{x}_1, \dots, \mathbf{x}_N, t). \quad (3.2.3)$$

### 3.3. Klassische Moleküldynamik

Wir folgen nun Tully [173, Abschn. 6] in der Herleitung der Time-Dependent Self-Consistent Field (TDSCF)-Gleichungen. Zunächst spalten wir den Hamiltonoperator (3.2.2) auf in  $\hat{H} = \hat{T}^{(\text{core})} + \hat{H}^{(\text{el})} + \hat{V}^{(\text{core,core})}$  mit  $\hat{H}^{(\text{el})} = \hat{T}^{(\text{el})} + \hat{V}^{(\text{el,el})} + \hat{V}^{(\text{core,el})}$  und  $\hat{V}^{(\text{tot})} = \hat{V}^{(\text{el,el})} + \hat{V}^{(\text{core,el})} + \hat{V}^{(\text{core,core})}$ , da die Anteile  $\hat{T}^{(\text{core})}$  und  $\hat{V}^{(\text{core,core})}$  nicht auf die Elektronenfreiheitsgrade wirken.

Anschließend projizieren wir die Zustandsfunktion (3.2.3) wieder mittels der Dekorrektoren auf eine unkorrelierte Produktform, wobei wir aber diesmal die Zeit  $t$  als Eingangsparameter auffassen. Die Gruppen der Eingangsvariablen sind also diesmal,  $\mathbf{X}$  und  $\mathbf{x}$ . Wieder benutzen wir die Korrelationsentwicklung (3.1.5) mit  $d = 2$ , die wir hier ausführlich darstellen,

$$\begin{aligned} \Psi(\mathbf{X}, \mathbf{x}, t) &= \sum_{\mathcal{I}_j} \tilde{\Psi}_{\mathcal{I}_j}(\mathbf{X}, \mathbf{x}, t) \\ &= \psi_{\mathbf{X}}(\mathbf{X}) \cdot \psi_{\mathbf{x}}(\mathbf{x}) \cdot \psi_t(t) \\ &\quad + (\psi_{\mathbf{X}\mathbf{x}}(\mathbf{X}, \mathbf{x}) \cdot \psi_t - \psi_{\mathbf{X}}(\mathbf{X}) \cdot \psi_{\mathbf{x}}(\mathbf{x}) \cdot \psi_t(t)) + \dots \end{aligned} \quad (3.3.1)$$



**Abbildung 3.1.:** Angenommene Korrelation zwischen den Eingangsvariablen für die klassische Moleküldynamik.

Wir bemerken, dass einzelnen Gruppen von Eingangsvariablen voneinander unterscheidbar sind, weswegen wir uns um die Symmetrie bei dieser Projektion keine Gedanken machen müssen.

Nun benutzen wir die Näherung, dass die Korrelation zwischen Kernen  $\mathbf{X}$  und Elektronen  $\mathbf{x}$  vernachlässigbar ist. Die entsprechenden Terme, die gleichzeitig von beiden Eingangsvariablen  $\mathbf{X}, \mathbf{x}$  abhängen, werden zu Null gesetzt. Um die folgende Herleitungen zu erleichtern, bietet es sich an, die verbleibenden Terme weitestgehend wieder zusammen zu fassen:

$$\Psi(\mathbf{X}, \mathbf{x}, t) \approx \chi(\mathbf{X}, t) \cdot \psi(\mathbf{x}, t). \quad (3.3.2)$$

Diese Gleichung deckt sich mit unserer Definition 5 über unkorrelierte Eingangsvariablen.

Eine Betrachtung der Wellenfunktionen der Kerne und der Elektronen als unkorreliert resultiert in eine sogenannte *adiabatische* Dynamik. Je nachdem, wieviel Korrelation man zulässt<sup>1</sup>, berücksichtigt man zusätzliche diabatische Kopplungen, die andernfalls

<sup>1</sup>Hierzu müsste man die Eingangsvariablen jedoch genauer aufschlüsseln.

vernachlässigt würden. Wir verweisen an dieser Stelle aber auf Tully [173, Kap. 2] für Details zu diabatischer Dynamik, die wir nicht weiter betrachten werden.

Die Annahme, dass keine Korrelation zwischen Kernen und Elektronen besteht, ist eine Vereinfachung, denn im Hamiltonoperator taucht der Anteil  $\hat{V}^{(\text{tot})}$  auf, welcher beide Eingangsvariablengruppen koppelt. Eine Separation der Variablen ist somit eigentlich nicht möglich, weil unser Ansatz eben nicht die exakte Lösung darstellen kann. Eine Näherungsmöglichkeit besteht, den Einfluss der jeweils anderen Eingangsvariablen durch Ausintegration zu mitteln. Man erhält dann einen effektiven Operator, der nur noch Variablen innerhalb der Gruppe koppelt. Er hängt jedoch von der Lösung der anderen ab.

Wenn wir Energieerhaltung fordern,

$$\frac{d}{dt} \int \tilde{\Psi}^*(\mathbf{X}, \mathbf{x}, t) \hat{H} \tilde{\Psi}(\mathbf{X}, \mathbf{x}, t) d\mathbf{X}_1 \dots d\mathbf{X}_M d\mathbf{x}_1 \dots d\mathbf{x}_N = 0, \quad (3.3.3)$$

wird diese Mittelung des Einflusses der anderen Eingangsvariablen deutlich.

Zunächst führen wir aber noch eine komplexwertige Phase ein, die wir aufgrund der komplexwertigen Natur der Zustandsfunktion frei wählen dürfen, da nur das Betragsquadrat der Funktion  $\Psi$  physikalische Relevanz hat. Der Phasenfaktor  $P = \exp\left(i \int_{t_0}^t E(t') dt'\right)$  mit

$$E(t') = \int \psi^*(\mathbf{x}, t) \chi^*(\mathbf{X}, t) (\hat{H}^{(\text{el})} + \hat{V}^{(\text{core, core})}) \psi(\mathbf{x}, t) \chi(\mathbf{X}, t) d\mathbf{X}_1 \dots d\mathbf{X}_M d\mathbf{x}_1 \dots d\mathbf{x}_N$$

erleichtert die weiteren Herleitungen, siehe [173, Abschn. 6.1].

Damit erhalten wir schließlich ein gekoppeltes Gleichungssystem,

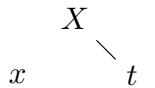
$$i \frac{\partial \psi}{\partial t} = - \sum_{k=1}^N \frac{1}{2} \Delta_{\mathbf{r}_k} \psi + \left( \int \chi^*(\mathbf{X}, t) \hat{V}^{(\text{tot})}(\mathbf{X}, \mathbf{x}) \chi(\mathbf{X}, t) d\mathbf{X} \right) \psi, \quad (3.3.4a)$$

$$i \frac{\partial \chi}{\partial t} = - \sum_{k=1}^N \frac{1}{2m_k} \Delta_{\mathbf{R}_k} \chi + \left( \int \psi^*(\mathbf{x}, t) \hat{V}^{(\text{tot})}(\mathbf{X}, \mathbf{x}) \psi(\mathbf{x}, t) d\mathbf{x} \right) \chi, \quad (3.3.4b)$$

bei der die eine Wellenfunktion jeweils das mittlere Potential der anderen wahrnimmt. Aus diesem Grund nennt man (3.3.1) auch Ansatz des mittleren Feldes. Diese Gleichungen bilden die Grundlage für den sogenannten TDSCF-Ansatz, indem auch die Betrachtung angeregter Elektronenzustände möglich ist. Dennoch bringt die Näherung durch die Projektion auf den Produktansatz Einschränkungen mit sich, die bspw. in [173, Abschn. 6.1] besprochen werden.

### 3.3.1. Born-Oppenheimer-Moleküldynamik

Wir möchten nun die Klassische Moleküldynamik im Rahmen der Born-Oppenheimer (BO)-Näherung herleiten. Die grundlegende Annahme ist, dass sich die Elektronen, aufgrund ihrer sehr viel kleineren Masse gegenüber den Kernen, adiabatisch auf Änderungen der Kernposition einstellen. Das soll bedeuten, dass sich der Elektronenzustand über die



**Abbildung 3.2.:** Angenommene Korrelation für die BO-Moleküldynamik.

### 3. Schrödinger-Gleichung

Zeit nicht ändert, bspw. verbleiben die Elektronen also immer in einem anfänglichen Grundzustand. Dies entspricht folgender Projektion:

$$\tilde{\Psi}(\mathbf{X}, \mathbf{x}, t) \approx \hat{K}_{\mathbf{x}t, \mathbf{x}} \Psi(\mathbf{X}, \mathbf{x}, t) = \chi(\mathbf{X}, t) \cdot \psi(\mathbf{x}) \quad (3.3.5)$$

Durch die fehlende Zeitabhängigkeit ist die bestimmende Gleichung für  $\psi$  die zeitunabhängige elektronische Schrödingergleichung, siehe auch die Herleitung in Abschnitt 3.2,

$$\hat{H}^{(\text{el})}(\mathbf{X}, \mathbf{x}, t) \psi(\mathbf{x}) = E^{(\text{el})} \psi(\mathbf{x}). \quad (3.3.6)$$

Zu beachten ist, dass  $\hat{H}^{(\text{el})}$  weiterhin von der Zeit  $t$  parametrisch über den effektiven Potentialoperator, siehe (3.3.4a), abhängt, denn dieser enthält die zeitabhängige Lösung der Kern-Wellenfunktion. Weiterhin haben wir die Kern-Kern-Wechselwirkungsoperator  $\hat{V}^{(\text{core, core})}$  nicht  $\hat{H}^{(\text{el})}$  zugeordnet, da er nicht von den Elektronenfreiheitsgraden abhängt und daher nur eine Konstante beitragen würde.

Sei  $\psi$  also zeitunabhängig und dieser durch den anfänglichen Zustand  $\psi(\mathbf{x}, t_0) = \psi^{(0)}(\mathbf{x})$  gegeben, dann schreiben wir das mittlere Potential durch die Elektronen und der Kerne zu einem bestimmten Zeitpunkt  $t$  als folgenden effektiven Operator

$$V^{(\text{BO})}(\mathbf{X}) = \int \psi^{(0)*}(\mathbf{x}) \hat{V}^{(\text{tot})}(\mathbf{X}, \mathbf{x}) \psi^{(0)}(\mathbf{x}) d\mathbf{x}, \quad (3.3.7)$$

wobei wir die wieder eigentlich vorhandene Kopplung der als unkorreliert angenommenen Eingangsvariablen ausintegrieren.

Dieser Ausdruck (3.3.7) wird auch die *BO-Potentialoberfläche* genannt.

Die weitere Herleitung der klassischen Dynamik nach Newton für die Kerne aus der quantenmechanischen Dynamik in Gleichung (3.3.4b) mit dem effektiven Operator  $V^{(\text{BO})}$  findet sich wieder in Tully [173, Abschn. 4.2].

#### 3.3.2. Poisson-Problem

Wir wollen die BO-Potentialoberfläche (3.3.7) zu einem bestimmten Zeitpunkt  $t$  genauer betrachten, dessen Gradient die Kräfte auf die Kerne in der BO-Moleküldynamik bestimmt:

$$V^{(\text{BO})}(\mathbf{X}) = \int \psi^{(0)*}(\mathbf{x}) \left( \hat{V}^{(\text{el, el})} + \hat{V}^{(\text{core, el})} + \hat{V}^{(\text{core, core})} \right) (\mathbf{X}, \mathbf{x}) \psi^{(0)}(\mathbf{x}) d\mathbf{x}.$$

Der erste Term wird offensichtlich zu einer Konstante  $C$ , welche aufgrund der reinen Auswertung der Potentialoberfläche als Gradient keine Auswirkungen hat.

Sei nun

$$\rho_j(\mathbf{x}_j) = \int \psi^{(0)*}(\mathbf{x}) \psi^{(0)}(\mathbf{x}) d\mathbf{x}_1 \dots d\mathbf{x}_{j-1} d\mathbf{x}_{j+1} \dots d\mathbf{x}_N \quad (3.3.8)$$

die sogenannte *reduzierte Dichtefunktion erster Ordnung*, siehe auch Szabo und Ostlund [168, Abschn. 4.4].<sup>2</sup>

<sup>2</sup>Zur Vereinfachung der Notation berücksichtigen wir hier nicht weiter den Spin der Elektronen. Ansonsten erhielte man zwei Dichtematrizen, eine für Spin Up und eine für Spin Down. Auf diesen gehen wir im nächsten Kapitel im Rahmen der HF-Methode genauer ein.

Betrachten wir den zweiten Term genauer und setzen (3.3.8) ein, dann erhalten wir mit (3.2.1)

$$\begin{aligned} \int \psi^{(0)*}(\mathbf{x}) \hat{V}^{(\text{core,el})}(\mathbf{X}, \mathbf{x}) \psi^{(0)}(\mathbf{x}) d\mathbf{x} &= - \sum_{j=1}^N \sum_{k=1}^M Z_k \int \frac{\rho_j(\mathbf{x}_j)}{\|\mathbf{r}_j - \mathbf{R}_k\|} d\mathbf{x}_j \\ &= - \sum_{k=1}^M Z_k \int \frac{\rho(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}_k\|} d\mathbf{r} \end{aligned}$$

wobei wir im zweiten Schritt die einzelnen reduzierten Dichtematrizen aufsummiert haben und die resultierende Dichtematrix nur noch von einer allgemeinen Ortskoordinate  $\mathbf{r} \in \mathbb{R}^3$  abhängt.

Im dritten Term ist der Operator unabhängig von den verallgemeinerte Elektronenkoordinaten  $\mathbf{x}$ . Da er die Form eines konstanten Gewichts hat, können wir ihn vor das Integral ziehen. Das Integral selbst wiederum wird zu 1, weil die elektronische Wellenfunktion normiert ist. Der Term lässt sich dann mit Hilfe von Delta-Distributionen auch wie folgt darstellen, siehe Griebel et al. [77, Abschn. 7.1.3]:

$$\sum_{k < j}^M \frac{Z_k Z_j}{\|\mathbf{R}_k - \mathbf{R}_j\|} = \sum_{k < j}^M Z_k Z_j \int \frac{\delta(y - \mathbf{R}_j)}{\|y - \mathbf{R}_k\|} dy,$$

wobei wir von offenen Randbedingungen ausgegangen sind.

Es verbleibt also

$$V^{(\text{BO})}(\mathbf{X}) = C - \sum_{k=1}^M Z_k \int \frac{\rho(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}_k\|} d\mathbf{x} + \sum_{k < j}^M Z_k Z_j \int \frac{\delta(y - \mathbf{R}_j)}{\|y - \mathbf{R}_k\|} dy.$$

Wir vernachlässigen die Konstante  $C$  und haben damit  $V^{(\text{BO})}(\mathbf{X})$  auf die Gestalt der Integralformulierung des Poisson-Problems

$$V(\mathbf{R}) = \int \rho(y) \frac{1}{\|y - \mathbf{R}\|} dy \quad (3.3.9)$$

zurückgeführt, siehe Griebel et al. [77, S. 240], wobei  $\rho(y)$  die Dichteanteile der elektronische Ladungsverteilung und der Delta-Distributionen der Kerne mal ihrer Ladung vereint. Das bedeutet, wenn die Lösung der elektronischen Wellenfunktion bekannt ist, dann muss für die Bestimmung der Dynamik der Kerne nur noch das Poisson-Problem im  $\mathbb{R}^3$  gelöst werden. Wir verweisen auf Abschnitt 9.5, wie das Problem effizient gelöst werden kann.

### 3.4. Elektronische Wellenfunktion

In der BO-Näherung verbleibt uns, die zeitunabhängige elektronische Schrödingergleichung (3.3.6) im Grundzustand zu lösen. Damit hat sich die Komplexität des Problems

### 3. Schrödinger-Gleichung

nur unwesentlich verändert. Wir sind immer noch vor eine sehr hochdimensionale Integro-Differentialgleichung gestellt.

Wir wollen an dieser Stelle die Korrelationsentwicklung (2.5.5a) auf die Grundzustandselektronwellenfunktion  $\psi^{(0)}(\mathbf{x})$  anwenden und sie ausführlich besprechen, da diese trun-kierte Entwicklung die Basis für die folgenden Kapitel über Hartree-Fock und Configuration Interaction bildet.

Sei also  $\mathcal{D} = \{1, \dots, N\}$  die Menge der Elektronindizes<sup>3</sup>. Wieder geht es um alle Partitionen  $\mathcal{I}_j$  dieser Indexmenge  $\mathcal{D}$ . Wir erhalten dann folgende Entwicklung:

$$\psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{\mathcal{I}_j \in \mathcal{I}} \widetilde{\psi}^{(0)}_{\mathcal{I}_j}(\mathbf{x}_1, \dots, \mathbf{x}_N) \quad (3.4.1a)$$

mit

$$\widetilde{\psi}^{(0)}_{\mathcal{I}_j}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \hat{K}_{\mathcal{I}_j} \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) - \sum_{\mathcal{I}_k \subsetneq \mathcal{I}_j} \widetilde{\psi}^{(0)}_{\mathcal{I}_k}(\mathbf{x}_1, \dots, \mathbf{x}_N). \quad (3.4.1b)$$

#### 3.4.1. Produktansatz

Beispielsweise hat der erste Term folgende Gestalt:

$$\widetilde{\psi}^{(0)}_{1N}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \hat{K}_{1N} \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N)$$

Wenn  $\hat{K}_{1N}$  der Dekorrelator auf unterscheidbare Eingangsräume nach Definition 21 ist, dann folgt:

$$\hat{K}_{1N} \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \psi^{(0)}_1(\mathbf{x}_1) \cdot \dots \cdot \psi^{(0)}_N(\mathbf{x}_N) \quad (3.4.2)$$

Dies entspricht dem Ansatz nach Hartree [85].

Wenn hingegen  $\hat{K}_{1N}^A$  der antisymmetrische Dekorrelator nach Definition 23 ist, dann folgt:

$$\hat{K}_{1N}^A \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \hat{A} \left( \psi^{(0)}_1(\mathbf{x}_1) \cdot \dots \cdot \psi^{(0)}_N(\mathbf{x}_N) \right) \quad (3.4.3)$$

Dieser Ansatz wird auch Slater-Determinante genannt, nach einer Bemerkung von Slater [158]. Fock [58] beschrieb einen solchen antisymmetrisierten Produktansatz, ebenso aufbauend auf der Arbeit von Hartree [85]. Dieser Ansatz führt die hochdimensionale, lineare Gleichung über in eine niederdimensionale, aber nicht-lineare Gleichung. Diese sogenannte HF-Methode werden wir im folgenden Kapitel näher beschreiben.

---

<sup>3</sup>Damit müssen nicht zwingend einzelne Elektronen gemeint sein, bspw. erlaubt diese Formulierung auch, je zwei Elektronen unterschiedlichen Spins zusammen zu fassen.

### 3.4.2. Configuration Interaction

Betrachten wir an dieser Stelle aber auch Terme höherer Ordnung, aber nur noch mit dem antisymmetrischen Dekorrelator. Der zweite Term in der Entwicklung lautet:

$$\begin{aligned}\widetilde{\psi^{(0)}}_{1^{N-2}2}(\mathbf{x}_1, \dots, \mathbf{x}_N) &= \hat{K}_{1^{N-2}2}^A \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) - \hat{K}_{1N}^A \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) \\ &= \hat{A} \left( \psi^{(0)}_1(\mathbf{x}_1) \cdot \dots \cdot \psi^{(0)}_{N-1N}(\mathbf{x}_{N-1}, \mathbf{x}_N) \right) \\ &\quad - \hat{A} \left( \psi^{(0)}_1(\mathbf{x}_1) \cdot \dots \cdot \psi^{(0)}_N(\mathbf{x}_N) \right).\end{aligned}$$

Entsprechend analoge Gestalt haben die weiteren Terme. Die Differenz betrachtet hier also stets den Unterschied in dem antisymmetrischen Raum über zwei Eingangsvariablen zur antisymmetrisierten Rang-1-Bestapproximation in den Einzelräumen.

Die CI-Methode beschreibt dabei diese Entwicklung mit Hilfe von Anregungsoperatoren. Hier wird die Rang-1-Bestapproximation durch Hinzunahme weiterer Basisfunktionen erweitert. Wir werden im Kapitel 8 darauf noch genauer eingehen.

### 3.4.3. Geminal-Entwicklung

Eine besondere Entwicklung ergibt sich, wenn  $N = \frac{d}{2}$ , denn dann lohnt die Betrachtung folgenden Terms:

$$\hat{K}_{\frac{N}{2}}^A \psi^{(0)}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \hat{A} \left( \psi^{(0)}_{1,2}(\mathbf{x}_1, \mathbf{x}_2) \cdot \dots \cdot \psi^{(0)}_{N-1,N}(\mathbf{x}_{N-1}, \mathbf{x}_N) \right). \quad (3.4.4)$$

Dies beschreibt die Entwicklung der elektronischen Wellenfunktion in sogenannten Geminal-Wellenfunktionen. Der Grund der Nützlichkeit liegt darin, dass in  $\hat{V}^{(\text{tot})}$  stets zwei Elektronortskordinaten enthalten sind. Die Geminal-Wellenfunktion stellt also die perfekte Ansatzfunktion für die durch diesen Operatorteil entstehende Korrelation. Im Rahmen dieser Arbeit werden wir auf diese Methode aber nicht näher eingehen, sondern verweisen auf die Arbeit von Hurley et al. [102], die dieser Methode den Anstoß gab, und außerdem auf die Einführung von Surján [167]. Weiterhin findet sich in [137] noch einmal eine schöne Darstellung des Grundgedankens.

### 3.4.4. Funktionalanalytische Bemerkungen

Wir werden im Folgenden nicht weiter darauf eingehen, in welchem Raum genau die Funktion  $\Psi$  lebt. Hier gibt es fundierte Aussagen, wie zuletzt von Yserentant und Kreuzler [187], welche sich mit Hilberträumen mit gemischten Ableitungen beschäftigen. Aussagen über die Regularität der Zustandsfunktion helfen den Raum, in dem sie existiert, genau einzuschränken. Je genauer der Raum beschrieben werden kann, desto wirksamer gelingt eine Approximation in diesem Raum. Für eine sparsame Approximation muss insbesondere eine hohe Regularität vorliegen. Eine solche Herangehensweise in Verbindung mit einer Tensorzerlegung findet sich z. B. in Hamaekers [82]. Wir sind aber zunächst nicht an einer qualitativ hochwertigen Approximation der exakten Lösung interessiert, sondern an der sparsamen Berechnung einer Näherungslösung in einem gewählten Unterraum.

### 3. Schrödinger-Gleichung

Uns genügt daher die Aussage, dass sie im (Hilbert-)Raum der quadratintegrierbaren Funktionen  $\psi^{(0)} \in \mathcal{L}_2(\Omega)$ ,  $\Omega \subset \mathbb{R}^{3n}$  mit dem inneren Produkt

$$\forall f, g \in \mathcal{L}_2(\Omega) : \quad \langle f, g \rangle = \int_{\Omega} f^*(\mathbf{x}_1, \dots, \mathbf{x}_n) g(\mathbf{x}_1, \dots, \mathbf{x}_n) d\mathbf{x}_1 \dots d\mathbf{x}_n \quad (3.4.5)$$

lebt. Dies muss gelten, damit die Aufenthaltswahrscheinlichkeit (3.1.2) über  $\Omega$  integriert auf die Wahrscheinlichkeit 1 normiert werden kann.

#### 3.4.5. Variationsprinzip

Allgemein machen wir uns ein mächtiges Prinzip zunutze, wenn wir eine Lösung der elektronischen Schrödingergleichung suchen wollen, die auf einen Unterraum eingeschränkt ist, siehe Szabo und Ostlund [168, Abschn. 1.3]:

**Theorem 1** (Variationsprinzip). *Nehmen wir an, dass die Eigenfunktionen  $\Psi_i$  des Hamiltonoperator vollständig bekannt sind. Weiterhin seien sie nach ihren Eigenwerten aufsteigend angeordnet:  $E^{(0)} = E^{(el)}_1 \leq E^{(el)}_2 \leq \dots$*

*Dann gilt für eine beliebige, normalisierte Funktion  $\tilde{\Psi}$ , die im Unendlichen verschwindet, dass*

$$\int \tilde{\Psi}^*(\mathbf{x}) \hat{H} \tilde{\Psi}(\mathbf{x}) d\mathbf{x} \geq E^{(0)} \quad (3.4.6)$$

*spricht, ihr Erwartungswert ist stets eine obere Schranke für die Energie des Grundzustands  $E^{(0)}$ .*

Der Beweis des Theorems macht klar, wo der Zusammenhang zum zuvor angesprochenen Unterraum liegt:



*Beweis.* Wir machen uns zunutze, dass die Eigenfunktionen  $\Psi_i$  eine vollständige Basis darstellen, damit ist  $\sum_i \Psi_i(\mathbf{x}')\Psi_i^*(\mathbf{x}'')$  der Identitätsoperator in dem Funktionenraum. Wir schreiben:

$$\begin{aligned}
 1 &= \int \tilde{\Psi}^*(\mathbf{x})\tilde{\Psi}(\mathbf{x})d\mathbf{x} \\
 &= \iiint \tilde{\Psi}^*(\mathbf{x}) \left( \sum_i \Psi_i(\mathbf{x})\Psi_i^*(\mathbf{x}') \right) \left( \sum_j \Psi_j(\mathbf{x}')\Psi_j^*(\mathbf{x}'') \right) \tilde{\Psi}(\mathbf{x}'')d\mathbf{x}d\mathbf{x}'d\mathbf{x}'' \\
 &= \sum_{i,j} \iiint \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})\Psi_i^*(\mathbf{x}')\Psi_j(\mathbf{x}')\Psi_j^*(\mathbf{x}'')\tilde{\Psi}(\mathbf{x}'')d\mathbf{x}d\mathbf{x}'d\mathbf{x}'' \\
 &= \sum_{i,j} \int \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})d\mathbf{x} \int \Psi_i^*(\mathbf{x}')\Psi_j(\mathbf{x}')d\mathbf{x}' \int \Psi_j^*(\mathbf{x}'')\tilde{\Psi}(\mathbf{x}'')d\mathbf{x}'' \\
 &= \sum_{i,j} \int \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})d\mathbf{x}\delta_{ij} \int \Psi_j^*(\mathbf{x}'')\tilde{\Psi}(\mathbf{x}'')d\mathbf{x}'' \\
 &= \sum_i \int \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})d\mathbf{x} \int \Psi_i^*(\mathbf{x}'')\tilde{\Psi}(\mathbf{x}'')d\mathbf{x}'' \\
 &= \sum_i \left| \int \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})d\mathbf{x} \right|^2.
 \end{aligned}$$

Dabei haben wir die Vertauschung der unendlichen Summen mit dem Lebesgue-Integral unter Voraussetzung der gleichmäßigen Konvergenz nach dem Satz von Levi zugelassen.

In völlig analoger Weise betrachten wir dann den Ausdruck:

$$\begin{aligned}
 \int \tilde{\Psi}^*(\mathbf{x})\hat{H}\tilde{\Psi}(\mathbf{x})d\mathbf{x} &= \sum_i E^{(\text{tot})}_i \left| \int \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})d\mathbf{x} \right|^2 \\
 &\leq E^{(0)} \sum_i \left| \int \tilde{\Psi}^*(\mathbf{x})\Psi_i(\mathbf{x})d\mathbf{x} \right|^2 = E^{(0)},
 \end{aligned}$$

was aufgrund von  $\int \Psi_i^*(\mathbf{x}')\hat{H}\Psi_j(\mathbf{x}')d\mathbf{x}' = E^{(\text{tot})}_i\delta_{ij}$  und der Anordnung der Eigenwerte folgt.  $\square$

Ein Unterraum wird nur durch einen Teil der Eigenfunktionen  $\Psi_i$  aufgespannt und muss deswegen automatisch eine Energie größer gleich der des Grundzustands im Ganzraum besitzen.

## Zusammenfassung

In diesem Kapitel befassten wir uns mit der zeitabhängigen Schrödingergleichung. Wir widmeten uns dabei genauer der Zustandsfunktion, welche das Verhalten *aller Teilchen gleichzeitig* beschreibt, und brachen, mit Hilfe von Dekorrelatoren, diese auf Produktformen herunter, welche aus Funktionen zusammen gesetzt sind, die jeweils nur *Teilmengen*

### 3. Schrödinger-Gleichung

aller Teilchen *gleichzeitig* beschreiben. Mit Hilfe dieser Produktformen gelangten wir zur klassischen Moleküldynamik für ein System aus Kernen und Elektronen, wobei die Produktform mit der adiabatischen Beschreibung der Wechselwirkungen zwischen Kernen und Elektronen korrespondiert. Die Kerne werden dabei klassisch durch eine Newtonsche Bewegungsgleichung beschrieben, welche vom mittleren Feld der Elektronen abhängt, was wir durch Ausintegration der unkorrelierten Eingangsvariablen nachvollzogen. Das dabei verbleibende Problem der elektronischen Schrödingergleichung muss aber immer noch gelöst werden. Der Energieerwartungswert jeder Produktform stellt dabei nach dem Variationsprinzip eine obere Schranke an die wahre Grundzustandsenergie dar. Der Term niedrigster Ordnung der in die elektronische Wellenfunktion eingesetzten Dekorrelator-entwicklung wird uns im nächsten Kapitel zur bekannten HF-Gleichung führen. Den Termen höherer Ordnung widmen wir uns dann in Kapitel 8.

## 4. Hartree-Fock

Im letzten Kapitel haben wir die volle Schrödingergleichung betrachtet und das Problem auf das Finden der immer noch hochdimensionale Lösung der elektronischen Schrödingergleichung reduziert, dem *Elektronenstrukturproblem*. Wir haben dabei die Gesamtwellenfunktion in eine Summe von Produktfunktionen entwickelt und bestimmte Terme unter der Annahme entfernt, dass ihre Eingangsvariablen unkorreliert sind. Nimmt man dies für eine ganze Gruppe von Eingangsvariable an, bspw. für alle Elektronenorte, obwohl der zugehörige Operator des Problems verschiedene Elektronenorte miteinander verknüpft, so wird dies der *Ansatz des mittleren Feldes* genannt. Der Grund liegt in der zur Lösung notwendigen Ausintegration, also Mittelung, der jeweils anderen Eingangsvariablen, welche im Operator eigentlich miteinander verknüpft sind. Dies führt zu einem gekoppelten Gleichungssystem, bei der die Gleichung jedes Teilchens nur den gemittelten Einfluss der anderen Teilchen beinhaltet. Wenn wir für die elektronische Wellenfunktion einen solchen Ansatz machen, dann entspricht dies der Trunkierung der Korrelationsentwicklung aus Kapitel 2 nach dem ersten Term. Anders formuliert, wir werden also nun die Lösung der elektronische Schrödingergleichung projiziert auf die Rang-1-Bestapproximation des antisymmetrisch vertensorierten Raumes der Funktionen mit einer Eingangsvariable betrachten, siehe (3.4.3).

Das Kapitel gliedert sich in die folgenden Abschnitte: Zunächst werden wir die nach dem ersten Term trunkierte Entwicklung in die elektronische Schrödingergleichung einsetzen. Dies wird in Abschnitt 4.1 zur bekannten Hartree-Fock (HF)-Gleichung mit dem Fockoperator führen. Zur Berechnung der Lösung im Computer wird diese Gleichung in einer Basis bekannter Funktionen entwickelt. Dies führt in Abschnitt 4.2 zur Roothaan-Hall-Gleichung, welche ein verallgemeinertes Eigenwertproblem darstellt. Dieses Problem wird in ein normales Eigenwertproblem überführt, um es Standardalgorithmen zugänglich zu machen. Anschließend wenden wir uns genauer den einzelnen Matrixelementen des nicht-linearen Fockoperators zu. Dabei spalten wir die Abhängigkeit des Operators von den anderen, als unkorreliert angenommenen Eingangsvariablen in die sogenannte Dichtematrix ab. Abschließend wenden wir uns in Abschnitt 4.3 der Komplexität des zu lösenden Matrixproblems zu, also dem Verhalten des Rechenaufwands gegen die Anzahl der Elektronen bzw. die Anzahl an Basisfunktionen.

Diese Untersuchungen erfolgen deshalb in so großer Detailtiefe, weil es im folgenden Kapitel darum gehen wird, die prinzipielle Komplexität von  $\mathcal{O}(N^4)$  bzw.  $\mathcal{O}(n^4)$  dieser Methode zu linearisieren.

## 4.1. Closed-Shell Hartree-Fock

Im Folgenden werden wir die HF-Methode für Systeme mit Gesamtspin gleich Null, dem sogenannten Closed-Shell Hartree-Fock-Fall, herleiten.

Die größte Quelle für Verwirrung für den nicht näher vorgebildeten Leser wird im Folgenden sicherlich die verschiedenen Basen bzw. Basisentwicklungen sein. Daher wollen wir sie schon einmal aufzählen: Zunächst treffen wir auf die in die elektronische Wellenfunktion eingesetzte Dekorrelatorentwicklung. Die darin enthaltenen unbekanntenen Einteilchen-Wellenfunktionen werden dann in einer bekannten Basis nach einer Idee von **Roothaan** und **Hall** entwickelt. Üblicherweise verwendet man dabei die Lösungen des Wasserstoffatoms, sogenannte Atomare Orbitale (AOs). Um schließlich Berechnungen zu vereinfachen, können diese Lösungen wiederum in einer weiteren Basis entwickelt werden, wie den wohlbekanntenen Gauss-Funktionen. Hier stecken also drei Entwicklungen wie die berühmten Matroschka-Puppen ineinander.

Aufgrund der Länge dieses Abschnitts wollen wir kurz auf seine Gliederung eingehen: Wir betrachten in Abschnitt 4.1.1 zunächst die elektronische Zustandsfunktion und welche Bedingungen sie erfüllen muss. Da Elektronen einen Spin besitzen, muss diesem und den Einteilchenfunktionen, die ihn beschreiben, besondere Aufmerksamkeit widerfahren. Wir gehen darauf in Abschnitt 4.1.2 näher ein. Wir werden dann in Abschnitt 4.1.3 den Ansatz von **Hartree** und die Verbesserung durch **Fock** betrachten, die diese zuvor aufgestellten Bedingungen zu erfüllen suchen, und sie im Rahmen unserer Korrelationsprojektoren verstehen. In Abschnitt 4.1.4 schränken wir den Ansatz auf den Fall ein, in dem der Gesamtspin des Systems gleich null ist, den sogenannten CLHF-Fall. Der Ansatz nach **Hartree** und **Fock** in die elektronische Schrödingergleichung führt uns zur HF-Gleichung in Abschnitt 4.1.5. Deren Eigenschaften besprechen wir in Abschnitt 4.1.6 und widmen uns besonders ihrem nicht-linearen Charakter.

In Abschnitt 4.1.7 fassen wir abschließend die gemachten Näherungen der HF-Approximation zusammen.

### 4.1.1. Bedingungen an die Wellenfunktion

Die Struktur des Systems, welche die elektronische Schrödingergleichung (3.3.6) festlegt, wird durch eine elektronische Wellenfunktion  $\psi$  beschrieben. Sie muss zur korrekten Modellierung der physikalischen Vorgänge drei Bedingungen erfüllen, siehe Staemmler [161]:

**Normierung** Da das Betragsquadrat der Zustandsfunktion einer Aufenthaltswahrscheinlichkeit entspricht, muss diese auf 1 normiert sein.

**Antisymmetrie** Elektronen gehorchen der Fermi-Dirac-Statistik. Die Wellenfunktion muss antisymmetrisch unter Vertauschung zwei beliebiger Elektronen sein.

**Eigenfunktion der Spinoperatoren** Eine vollständige Behandlung des Spins kann nur im Rahmen der Dirac-Gleichung erfolgen. Der Hamiltonian (3.2.1) enthält keine Spin-abhängigen Terme. Seine Eigenfunktionen kommutieren also mit den Spinoperatoren und es kann leicht gezeigt werden, dass sie daher auch Eigenfunktionen dieser Operatoren sein müssen.

### 4.1.2. Orbitale und Spin-Orbitale

Wir folgen hier Szabo und Ostlund [168, Abschn. 2.2.1].

Im Weiteren bezeichnen wir mit Orbital die elektronische Wellenfunktion für ein einzelnes Teilchen. Dabei ist das räumliche Orbital  $\phi(\mathbf{r}_1)$  der räumliche Anteil dieses Orbitals. Neben seiner Position verfügt das Elektron als weitere Eigenschaft auch über einen sogenannten Spin. Dieser wird durch zwei orthonormale Funktionen  $\alpha(\omega)$  und  $\beta(\omega)$ , dem *Up*- und dem *Down*-Spinor, beschrieben, die diesen Teilraum vollständig aufspannen. Die elektronische Wellenfunktion  $\psi(\mathbf{x}_1)$  die Ort und Spin beschreibt, heißt Spin-Orbital. Aus jedem räumlichen Orbital können dabei zwei Spin-Orbitale durch Produktbildung mit einem der Spinoren geformt werden,

$$\psi(\mathbf{x}_1) = \begin{cases} \phi(\mathbf{r}_1)\alpha(\omega) \\ \phi(\mathbf{r}_1)\beta(\omega) \end{cases} .$$

Die Orbitale bilden dabei eine orthonormale Menge. Da die beiden Spinoren sind per Definition orthonormal zueinander sind, müssen nur die räumlichen Orbitale orthonormal sein, d. h. sie sind jeweils auf 1 normiert und erfüllen

$$\langle \phi_i | \phi_j \rangle = \int \phi_i^* \phi_j d\mathbf{r} = \delta_{ij}, \quad (4.1.1)$$

wobei  $\delta$  das Kronecker-Delta bezeichnet.

### 4.1.3. Ansatzwellenfunktion und Slater-Determinanten

Betrachten wir die einzelnen Terme des Hamiltonoperators(3.2.1), so finden wir nur Terme, die entweder von der Position eines oder zweier Elektronen zugleich abhängen. Hartree [85] verfolgte den Ansatz, die gesuchte Wellenfunktion im gemittelten Feld der anderen Elektronen zu finden und untersuchten deswegen Lösungen der Schrödingergleichung in Nicht-Zentralfeldern. Slater [158] erkannte, dass Hartrees Ansatz einem Produktansatz für die Wellenfunktion entsprach,

$$\psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \psi_1(\mathbf{x}_1) \cdot \psi_2(\mathbf{x}_2) \cdot \dots \cdot \psi_N(\mathbf{x}_N). \quad (4.1.2)$$

Wir erkennen, dass dieser Ansatz dem ersten Term in unserer Korrelationsentwicklung (2.5.5a) entspricht, siehe auch (3.4.2), wobei Symmetrieeigenschaften nicht berücksichtigt werden.

Hierbei bezeichnen die Funktionen  $\psi_i(\mathbf{r}_j)$  jeweils das *i*te Spin-Orbital, welches mit Elektron *j* „besetzt“ ist.

Die Funktion (4.1.2) ist zwar leicht zu normieren, aber nicht antisymmetrisch, denn sie unterscheidet zwischen den verschiedenen Elektronen, siehe [158]. Mit dem Antisymmetrisierer, siehe (2.1.5b) und Fock [58, Abschn. 5], erhalten wir eine vollständig antisymmetrische Funktion. Diese Art von Funktionen wird auch Slater-Determinante nach Slater [158] genannt, da sie sich als Matrix schreiben lässt, in der über die Spalten

#### 4. Hartree-Fock

die Orbitale und über die Zeilen die Elektronen variieren, siehe Szabo und Ostlund [168, Abschn. 2.2.3]:

$$\psi_{\text{SCF}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = C \cdot |\psi_1(\mathbf{x}_1) \cdot \psi_2(\mathbf{x}_2) \cdot \dots \cdot \psi_N(\mathbf{x}_N)| \quad (4.1.3)$$

$$:= C \cdot \begin{vmatrix} \psi_1(\mathbf{x}_1) & \dots & \psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}, \quad (4.1.4)$$

wobei  $C = \sqrt{N!}^{-1}$  für die korrekte Normierung sorgt. Auch diesen Term erhalten wir aus der Korrelationsentwicklung mit dem antisymmetrischen Dekorrelator, siehe (3.4.3), jedoch mit abweichender Normierung aufgrund des normierten Betragsquadrats der Funktion.

Zu beachten ist, dass die obige Funktion, trotz ihrer einfachen Schreibweise, eine Linearkombination von  $N!$  Produktfunktionen darstellt. Das Anhängsel „SCF“ wird später klar werden. Es dient vorerst nur zur Unterscheidung der echten elektronischen Wellenfunktion  $\psi$  von dieser im Folgenden weiter betrachteten Ansatzfunktion  $\psi_{\text{SCF}}$ .

##### 4.1.4. Closed-Shell

Da Hamiltonoperator und z-Komponente des Spinoperators bzw. Gesamtspinoperator kommutieren, muss ein gemeinsames Eigensystem als Lösung der elektronischen Schrödingergleichung existieren.

Eine Slater-Determinante aus Spin-Orbitalen ergibt aber nur in speziellen Fällen eine Eigenfunktion der Spinoperatoren, nämlich für ein Zwei-Teilchen-System oder für ein System mit gerader Elektronenzahl und vollständig gefüllter äußerer Elektronenschale, siehe Staemmler [161, Abschn. 5,c)]. Diese speziellen Fälle nennt man Closed-Shell Hartree-Fock. Für alle anderen Fälle, als Open-Shell Hartree-Fock bezeichnet, muss eine Linearkombination von Slater-Determinanten verwendet werden, siehe Szabo und Ostlund [168, Abschn. 2.5].

Die Grundzustände der meisten Moleküle besitzen jedoch eine gefüllte äußere Schale, weswegen wir uns im Folgenden auf den Closed-Shell Hartree-Fock-Fall beschränken und unsere Wellenfunktion also stets als einzelne Slater-Determinante ansetzen. Für diese Zustände gilt, dass für jedes besetzte Orbital auch das Orbital aus dem gleichen räumlichen Orbital aber mit entgegengesetztem Spin besetzt sein muss. Das bedeutet, dass die Slater-Determinante stets sowohl beide möglichen Produkte von räumlichem Orbital und Spinor enthält,

$$\begin{aligned} \psi_{\text{SCF}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \\ C \cdot |\phi_1(\mathbf{r}_1)\alpha(\omega) \cdot \phi_1(\mathbf{r}_2)\beta(\omega) \cdot \dots \cdot \phi_{\frac{N}{2}}(\mathbf{r}_{N-1})\alpha(\omega) \cdot \phi_{\frac{N}{2}}(\mathbf{r}_N)\beta(\omega)|. \end{aligned} \quad (4.1.5)$$

### 4.1.5. Hartree-Fock-Gleichung

Mit einer Menge von Einteilchen-Wellenfunktionen  $\phi_i$  und mit der Ansatzfunktion (4.1.5) lässt sich nun aus der elektronischen Schrödingergleichung eine Einteilchengleichung ableiten. Wir erinnern kurz an die Form des Elektron-Hamiltonoperator,

$$\hat{H}^{(\text{el})} = \hat{T}^{(\text{el})} + \hat{V}^{(\text{el,el})} + \hat{V}^{(\text{core,el})} = -\frac{1}{2} \sum_{k=1}^N \Delta_{\mathbf{r}_k} + \sum_{k<j}^N \frac{1}{\|\mathbf{r}_k - \mathbf{r}_j\|} - \sum_{k=1}^N \sum_{j=1}^M \frac{Z_j}{\|\mathbf{r}_k - \mathbf{R}_j\|},$$

Nun bilden wir aus (3.3.6) den Rayleigh-Quotienten, wobei wir die Normiertheit der elektronischen Wellenfunktion direkt berücksichtigen, und setzen darin unsere Ansatzfunktion aus  $N$  Spin-Orbitalen ein, siehe [168, Abschn. 2.3.3],

$$E^{(\text{SCF})} = \int \psi_{\text{SCF}}^* \hat{H}^{(\text{el})} \psi_{\text{SCF}} d\mathbf{r} = \sum_{i=1}^N h_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N [J_{ij} - K_{ij}]. \quad (4.1.6)$$

*Bemerkung 4.* Wir bemerken, dass der Faktor  $\frac{1}{2}$  die doppelte Summation des äquivalenten Indexpaares  $ij$  und  $ji$  berücksichtigt. Im Fall  $i = j$  gilt  $J_{ii} = K_{ii}$  und der Term hebt sich weg.

Schränken wir die Ansatzfunktion auf doppelt besetzte räumlichen Orbitale<sup>1</sup> ein, um (4.1.5) zu erhalten, dann können wir die Spinoren ausintegrieren, siehe [168, Abschn. 2.3.5]. Dazu trennt man die Summen in (4.1.6) jeweils in zwei Summen über je  $\frac{N}{2}$  Terme mit gleichem Spin auf. Wir erhalten,

$$E^{(\text{SCF})} = 2 \sum_{i=1}^{\frac{N}{2}} h_i + \sum_{i,j=1}^{\frac{N}{2}} [2J_{ij} - K_{ij}], \quad (4.1.7)$$

weil bei der Austauschenergie Terme durch die Orthogonalität der Spinoren verschwinden.

Der wie folgt definierte kinetische und Kernpotential-Anteil hängt damit nur von jeweils einem („1“) Elektron ( $\mathbf{r}_1$ ) ab<sup>2</sup>,

$$h_i = \int \phi_i^*(\mathbf{r}_1) \hat{h} \phi_i(\mathbf{r}_1) d\mathbf{r}_1, \quad (4.1.8)$$

während der Coulomb-Term,

$$J_{ij} = \iint \phi_i^*(\mathbf{r}_1) \phi_i(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_j^*(\mathbf{r}_2) \phi_j(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \quad (4.1.9)$$

<sup>1</sup>Mit einem Elektron mit Up-Spinor und einem zweiten Elektron mit Down-Spinor.

<sup>2</sup>Aufgrund der Ununterscheidbarkeit der Elektronen brauchen wir die Gleichungen nur für das erste bzw. die ersten beiden Elektronen zu schreiben. Aufgrund der vollständig antisymmetrischen Gestalt der Ansatzfunktion sind damit alle Wechselwirkungen erfasst, wie man sich leicht klarmacht.

#### 4. Hartree-Fock

und der Austausch-Term,

$$K_{ij} = \iint \phi_i^*(\mathbf{r}_1)\phi_j(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_j^*(\mathbf{r}_2)\phi_i(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2, \quad (4.1.10)$$

von zwei („1“ und „2“) Elektronen ( $\mathbf{r}_1$ ) und ( $\mathbf{r}_2$ ) abhängen. Zu beachten ist, dass der letzte Term nur aufgrund der quantenmechanischen Natur, genauer wegen der geforderten Antisymmetrie, auftritt. Dieser Anteil wird auch als Austausch-Korrelation bezeichnet und deutet damit an, dass die Antisymmetrie eine Eigenschaft der Korrelation der Vielteilchenfunktion ist, siehe Szabo und Ostlund [168, S. 51].

Im Folgenden wollen wir nun als Unbekannte in obiger Gleichung die verbleibende Menge der räumlichen Orbitale betrachten und nach diesen lösen. Dazu wollen wir uns das Variationsprinzip, siehe Abschnitt 3.4.5, zunutze machen. Das heisst, wir suchen  $\frac{N}{2}$  dieser Funktionen, so dass der Eigenwert  $E^{(el)}$  am kleinsten wird, unter der Nebenbedingung dass die Menge orthonormal bleibt. Genauer suchen wir das Minimum des folgenden Funktionals, welches sich aus (4.1.7) ergibt,

$$\mathcal{J}[\phi] = 2 \sum_{i=1}^{\frac{N}{2}} h_i + \sum_{i,j=1}^{\frac{N}{2}} [2J_{ij} - K_{ij}] - \sum_{i,j=1}^{\frac{N}{2}} 2\lambda_{ij} \left[ \int \phi_i^* \phi_j d\mathbf{r} - \delta_{ij} \right], \quad (4.1.11)$$

wobei  $\lambda_{ij}$  die Lagrange-Multiplikatoren zur Erfüllung der Nebenbedingung der Orthonormalität sind.

Notwendig für die Existenz des Minimums ist, dass gilt,

$$\frac{\delta \mathcal{J}[\phi]}{\delta \phi_i} = 0 \quad (4.1.12)$$

für alle  $i$ . Setzen wir (4.1.11) in die Bedingung (4.1.12) ein und berechnen die Ableitung nach  $\phi_i$ , so erhalten wir die Einteilchengleichung

$$\left[ \hat{h} + \sum_j \left( 2\hat{J}_j - \hat{K}_j \right) \right] \phi_i(\mathbf{r}) = \sum_{j=1}^{\frac{N}{2}} \lambda_{ij} \phi_j(\mathbf{r}), \quad (4.1.13)$$

wobei wir besonders auf den im Vergleich zu (4.1.7) anderen Vorfaktor vor dem Einteilchen-Hamiltonoperator aufmerksam machen. Auf diesen kommen wir gleich noch zu sprechen, wenn wir den Zusammenhang zwischen der Gesamtenergie und der Summe der Eigenwerte darlegen. Der Ein-Teilchen-Hamiltonoperator  $\hat{h}$  und die beiden Operatoren  $\hat{J}_j$  und  $\hat{K}_j$  sind dabei wie folgt gegeben, wobei die beiden letzteren sich von  $J_{ij}$  und  $K_{ij}$  für eine beliebige Funktion  $\xi(\mathbf{r}_1)$  ableiten,

$$\hat{h} = -\frac{1}{2}\Delta_{\mathbf{r}} - \sum_{c=1}^M \frac{Z_c}{|\mathbf{r} - \mathbf{R}_c|}, \quad (4.1.14a)$$

$$\hat{J}_j(\mathbf{r}_1)\xi(\mathbf{r}_1) = \int \phi_j^*(\mathbf{r}_2)\phi_j(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \cdot \xi(\mathbf{r}_1), \quad (4.1.14b)$$

$$\hat{K}_j(\mathbf{r}_1)\xi(\mathbf{r}_1) = \int \phi_j^*(\mathbf{r}_2)\xi(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \cdot \phi_j(\mathbf{r}_1). \quad (4.1.14c)$$



Die Einteilchenwellenfunktionen  $\{\phi_i\}$  sind dabei invariant unter unitärer Transformation. Damit lässt sich die Matrix  $\lambda_{ij}$  auf Diagonalgestalt bringen. Fasst man außerdem alle weiteren Terme in (4.1.13) in dem sogenannten Fockoperator zusammen,

$$\hat{F} := \hat{h} + \sum_j^{\frac{N}{2}} (2\hat{J}_j - \hat{K}_j), \quad (4.1.15)$$

so erhält man die berühmte Hartree-Fock-Gleichung:

$$\hat{F}\phi_i = \epsilon_i\phi_i, \quad (4.1.16)$$

wobei  $\epsilon_i$  die Energie des Orbitals mit Index  $i$  bezeichnet, deren Reihenfolge wir im Folgenden als monoton ansteigend festlegen. Hätten wir in (4.1.11) die Summe nicht auf  $i = j$  erweitert, so wäre der Fockoperator für jedes räumliche Orbital unterschiedlich gewesen, siehe [168, Kap. 3.1.2].

#### 4.1.6. Selbst-konsistente Lösung

Die Gleichung (4.1.16) hat folgende Eigenschaften:

1. Sie ist eine Integro-Differential-Gleichung.
2. Sie hat unendlich viele Lösungen: Die  $\frac{N}{2}$  Lösungen mit den kleinsten Eigenwerte werden „besetzte“ Orbitale genannt, nach dem so genannten *Aufbau-Prinzip*<sup>3</sup>. Die verbleibenden Lösungen sind unbesetzte oder „virtuelle“ Orbitale. Sie sind nicht minimiert, da in der Summe in (4.1.13) nur über die ersten  $\frac{N}{2}$  summiert wird.
3. Sie hat  $N$  besetzte Lösungen  $\phi_i$ , wenn gilt  $\sum_{a=1}^M Z_a \geq N$ , siehe [128].
4. Die Eigenwerte  $\epsilon_i = \int \phi_i^*(\mathbf{r})\hat{F}\phi_i(\mathbf{r})d\mathbf{r}$  sind die Orbitalenergien. Für gebundene Zustände sind sie negativ, für ungebundene positiv. Sie stehen über Koopmans Theorem in Zusammenhang mit dem Ionisationspotential - hierbei wird jedoch die Änderung im Zustand der anderen Orbitale durch Entfernen eines einzelnen vernachlässigt.
5. Die Summe der Eigenwerte entspricht nicht der Gesamtenergie,

$$E^{(\text{SCF})} = \sum_i^{\frac{N}{2}} h_i + \epsilon_i. \quad (4.1.17)$$

wie durch Vergleich mit (4.1.7) deutlich wird.

<sup>3</sup>Das Prinzip ergibt sich aus der Definition des Grundzustands als derjenige mit der niedrigsten Energie und der Tatsache, dass jedes Orbital durch maximal zwei Elektronen besetzt sein kann.

## 4. Hartree-Fock

Sie ist eine Einteilchengleichung der elektronischen Schrödingergleichung, hängt jedoch über die Operatoren  $\hat{J}_j$  und  $\hat{K}_j$  von der Lösung selbst ab. Dies macht sie nicht-linear.

Hartree [86] schlug den „selbst-konsistenten“ Ansatz vor: Man beginnt mit einer beliebigen orthonormalen Funktionsmenge  $\{\phi_i\}$  für die räumlichen Orbitale. Mittels dieser konstruiert man die obigen Operatoren durch Integration mittels (4.1.14b) und (4.1.14c). Anschließend wird die Eigenwertgleichung (4.1.16) gelöst. Die daraus resultierenden Eigenfunktionen werden erneut zur Konstruktion der Operatoren verwendet, wieder wird (4.1.16) gelöst. Dieses Verfahren wird solange fortgeführt, bis die Orbitalenergien  $\epsilon_i$  konvergieren. Da die Elektronen sich dabei im gemittelten Feld aller bewegen, wie Slater [159] erkannte, spricht man vom Verfahren des „selbst-konsistenten Feldes“ (engl. Self-Consistent Field (SCF)).

Die Suche nach der anfänglichen Funktionsmenge  $\{\phi_i\}$  hat den Rang einer Suche nach einem Startwert in einer nicht-linearen Minimierung, der uns verlässlich in das globale Minimum führt. Es gibt zwar Aussagen über die Existenz dieses Minimums, aber unserem Kenntnisstand nach keine über dessen Eindeutigkeit, siehe Defranceschi und Le Bris [42]. Weiterhin gibt es Beweise [29], dass einer der bekannten nicht-linearen Minimierungsalgorithmen sicher zum Minimum konvergiert, wiederum aber nicht, ob dieses auch das gesuchte globale Minimum ist.

### 4.1.7. Näherungen und Einschränkungen

Wir wollen abschließend noch einmal die Einschränkungen auflisten, die die elektronische Schrödingergleichung aufgrund der eingesetzten Ansatzfunktion (4.1.3) erfährt.

**Dynamische Korrelation** Die HF-Gleichung ist eine Näherung an die exakte elektronische Schrödingergleichung. Sie kann nicht die *dynamische* Korrelation der Elektronen abbilden, durch die Elektronen die Coulombabstoßung untereinander verringern. Dies liegt in der Natur der Einteilchengleichung, die mittels Integration, also Mittelung, das Verhalten der anderen Elektronen berücksichtigt. So wird auch der Korrelationsfehler wie folgt definiert:

$$E^{(\text{corr})} := E^{(\text{tot})} - E^{(\text{SCF})}, \quad (4.1.18)$$

also als Differenz zwischen  $E^{(\text{tot})}$ , der exakten Energie der Lösung der elektronischen Schrödingergleichung, und  $E^{(\text{SCF})}$ , der Energie, welche aus der HF-Gleichung, siehe (4.1.17), resultiert.

**Endliche Basis** Ein weiterer Fehler entsteht durch die Entwicklung der unbekanntenen Einteilchenlösungen der HF-Gleichung in einer endlichen Basis. Das Variationsprinzip kann das Minimum nur im Rahmen des Unterraums finden, der von dieser Menge endlicher Basisfunktionen aufgespannt wird. Die gefundene Energie ist also stets nur eine obere Schranke zur echten HF-Energie, siehe Abschnitt 3.4.5.

**Einschränkungen** In den vorangegangenen Abschnitten haben wir nur den Closed-Shell Hartree-Fock (CLHF)-Fall betrachtet, also ein molekulares System bei dem stets beide

Spin-Zustände pro räumlichem Orbital besetzt sind und damit Gesamtspin null hat. Für einen Gesamtspin ungleich null, also Open-Shell Hartree-Fock (OHF), ist die gewählte Ansatzfunktion (4.1.3) aber keine Eigenfunktion des Spin-Operators mehr. Hier muss die Ansatzfunktion aus der Linearkombination mehrerer Slater-Determinanten konstruiert werden, dies geschieht im Rahmen der folgenden Erweiterungen von HF, die wir aber nur Stichpunktartig aufführen und in dieser Arbeit nicht weiter betrachten:

- Restricted Open Hartree-Fock (ROHF)
- Unrestricted Hartree-Fock (UHF)
- Multi-Configuration Self-Consistent Field (MCSCF)
- Complete Active Space Self-Consistent Field (CASSCF)

## 4.2. Roothaan-Hall-Gleichung

In Abschnitt 4.1 haben wir die Hartree-Fock (HF)-Gleichung als Operatorgleichung aufgeschrieben. Wir haben die einzelnen Operatoranteile genauer betrachtet und die grundlegenden Eigenschaften dieser Eigenwertgleichung beschrieben.

In diesem Abschnitt werden wir nun die Hartree-Fock-Gleichung zur Behandlung im Computer mittels Darstellung in einer geeigneten Basis in eine algebraische Eigenwertgleichung überführen. Wir folgen dabei grundsätzlich Szabo und Ostlund [168, Abschn. 3.4].

Ist die gewählte Basis nicht orthonormal, dann führt dies zu einer verallgemeinerten Eigenwertgleichung. Bekanntermaßen kann diese Gleichung mittels einer unitären Transformation, siehe Abschnitt 4.2.3, in eine einfache Eigenwertgleichung überführt, welche mit Standardalgorithmen behandelbar ist.

### 4.2.1. Allgemeines Eigenwertproblem

Gleichwohl das Variationsprinzip zu einer Eigenwertgleichung (4.1.16) führt, dessen Eigenfunktionen die gesuchten Lösungen sind, können wir diese nicht direkt lösen, handelt es sich schließlich doch nicht um reelle Zahlen, sondern um möglicherweise komplizierte Funktionen von unbekannter Gestalt. Um die unbekannteren räumlichen Orbitale  $\phi_i$  zu berechnen, müssen diese zunächst in einer Basis bekannter Funktionen entwickelt werden. Roothaan und Hall schlugen unabhängig voneinander die Menge der Atomaren Orbitale (AOs) vor. Wir werden speziellere Basen erst in Abschnitt 5.3 untersuchen. An dieser Stellen nehmen wir allgemein an, dass jedes räumliche Orbital in einer endlichen Menge von Basisfunktionen  $\{\xi_\nu\}_{\nu=1}^n$  entwickelt wird. Dann verbleiben als einzige Unbekannte die Entwicklungskoeffizienten  $c_{\nu i}$  dieser Entwicklung

$$\phi_i(\mathbf{r}) = \sum_{\nu=1}^n c_{\nu i} \xi_\nu(\mathbf{r}). \quad (4.2.1)$$

#### 4. Hartree-Fock

Wir bemerken, dass die Basis im Allgemeinen endlich sein wird. Damit spannt sie nicht mehr den vollen Ein-Teilchen-Hilbertraum sondern nur einen Unterraum auf. Wir finden also jetzt nicht mehr die echten Lösungen der HF-Gleichungen, sondern nur noch Näherungen innerhalb dieses aufgespannten Unterraums. Nach dem Variationsprinzip, siehe Abschnitt 3.4.5, erhalten wir also eine Energie zum Minimum von (4.1.11), die im Allgemeinen größer als die echte Grundzustandsenergie ist.

Betrachten wir nun die HF-Gleichung (4.1.16) für ein ausgewähltes räumliche Orbital mit dem Index  $i$ , setzen die Basisentwicklung (4.2.1) für diese Funktion  $\phi_i$  ein, multiplizieren die resultierende Gleichung mit  $\xi_\mu^*$  von der linken Seite und integrieren über den Ort  $\mathbf{r}_1$ , siehe [168, Abschn. 3.4.2], dann erhalten wir:

$$\begin{aligned}\hat{F}\phi_i(\mathbf{r}_1) &= \epsilon_i\phi_i(\mathbf{r}_1), \\ \sum_{\nu} \hat{F}c_{\nu i}\xi_{\nu}(\mathbf{r}_1) &= \sum_{\nu} \epsilon_i c_{\nu i}\xi_{\nu}(\mathbf{r}_1), \\ \sum_{\nu} \xi_{\mu}^*(\mathbf{r}_1)\hat{F}c_{\nu i}\xi_{\nu}(\mathbf{r}_1) &= \sum_{\nu} \xi_{\mu}^*(\mathbf{r}_1)\epsilon_i c_{\nu i}\xi_{\nu}(\mathbf{r}_1), \\ \sum_{\nu} c_{\nu i} \int \xi_{\mu}^*(\mathbf{r}_1)\hat{F}\xi_{\nu}(\mathbf{r}_1)d\mathbf{r}_1 &= \epsilon_i \sum_{\nu} c_{\nu i} \int \xi_{\mu}^*(\mathbf{r}_1)\xi_{\nu}(\mathbf{r}_1)d\mathbf{r}_1.\end{aligned}$$

Definieren wir die Integrale als Fockmatrix  $\mathbf{F}$  und als Überlappmatrix  $\mathbf{S}$  ein, so erhalten wir dann

$$\sum_{\nu} c_{\nu i}\mathbf{F}_{\mu\nu} = \epsilon_i \sum_{\nu} c_{\nu i}\mathbf{S}_{\mu\nu}. \quad (4.2.2)$$

Da die Gleichung für alle  $i = 1, \dots, \frac{N}{2}$  erfüllt sein muss, können wir sie auch kompakt als Matrix schreiben. Dabei müssen wir nun aber auf die richtige Reihenfolge der einzelnen, nicht notwendigerweise quadratischen Matrizen im Produkt achten,

$$\begin{aligned}\sum_{\nu} \mathbf{F}_{\mu\nu}\mathbf{C}_{\nu j} &= \sum_{\nu, i} \mathbf{S}_{\mu\nu}\mathbf{C}_{\nu i}\mathbf{E}_{ij} \\ \mathbf{FC} &= \mathbf{SCE}\end{aligned} \quad (4.2.3)$$

Dies ist die sogenannte Roothaan-Hall-Gleichung, eine allgemeine Eigenwertgleichung, mit Koeffizientenmatrix  $\mathbf{C} = [c_{\nu i}]$  und der Matrix  $\mathbf{E}$ , welche sich aus den Lagrange-Multiplikatoren über die Nebenbedingung der Orthonormiertheit ergibt, siehe auch (4.1.11).

Die Matrixeinträge von  $\mathbf{F}$  und  $\mathbf{S}$  werden dabei aus Integralen über ein und zwei Elektronenkoordinaten gebildet, auf deren genaue Gestalt wir im folgenden Abschnitt eingehen werden.

Ist die Fockmatrix  $\mathbf{F}$  aufgestellt, so muss sie diagonalisiert werden, um die Eigenwerte und Eigenfunktionen in der gewählten endlichen Basis zu finden. Im Allgemeinen lässt sich mit dem Variationsprinzip sagen: Je größer die Basis gewählt wird, desto genauer wird die Näherung der mittels (4.1.17) summierten Energie an  $E_{SCF}$ .

Die Erstellung guter Basissätze, die mit möglichst wenig Funktionen eine möglichst hohe Approximationsgüte liefern, hat viele Jahrzehnte gedauert. Mittlerweile existieren

eine ganze Reihe verschiedener Standardbasen. Abschnitt A.1 listet diese im Einzelnen auf. Leider existieren keine Aussagen, welche die Güte einer Basis auf Eigenschaften des Operators zurückführen. Hier ist nur die bekannte analytische Lösung des Wasserstoffproblems hilfreich, welche die Grundlage dieser Standardbasissätze bildet.

#### 4.2.2. Fockmatrixelemente

In diesem Abschnitt wollen wir die Matrixelemente der Fockmatrix und der Überlappmatrix, wie sie in der Roothaan-Hall-Gleichung (4.2.3) auftauchen, genauer betrachten. Dazu führen wir zuvor die Dichtematrix ein, da sich mit ihrer Hilfe der koeffizientenabhängige Teil aus der Fockmatrix herauslösen lässt.

#### Ladungsdichte und Dichtematrix

Die *Ladungsdichte* der elektronischen Wellenfunktion ist definiert als

$$\rho(\mathbf{r}) = 2 \sum_i^{\frac{N}{2}} |\phi_i(\mathbf{r})|^2, \quad (4.2.4)$$

wobei  $\rho(\mathbf{r})$  die Wahrscheinlichkeit angibt, die  $N$  Elektronen am Ort  $\mathbf{r}$  zu finden und wir von doppelt besetzten räumlichen Orbitalen ausgehen, siehe [168, Abschn. 3.4.3]

Mittels des Integrals über obige Ladungsdichte (4.2.4) erhalten wir die Gesamtzahl an Elektronen,

$$\int \rho(\mathbf{r}) d\mathbf{r} = 2 \sum_i^{\frac{N}{2}} \int |\phi_i(\mathbf{r})|^2 d\mathbf{r} = N. \quad (4.2.5)$$

Analog zur Prozedur für die Roothaan-Hall-Gleichung führen wir nun wieder die Entwicklung der räumlichen Orbitale in einer Basis  $\{\xi_\nu\}_1^n$  ein,

$$\phi_i(\mathbf{r}) = \sum_{\mu=1}^n c_{i\mu} \xi_\mu(\mathbf{r}), \quad \phi_i^*(\mathbf{r}) = \sum_{\mu=1}^n c_{i\mu}^\dagger \xi_\mu^*(\mathbf{r}). \quad (4.2.6)$$

Setzen wir diese Entwicklungen in (4.2.4) ein,

$$\rho(\mathbf{r}) = 2 \sum_i^{\frac{N}{2}} \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) = 2 \sum_i^{\frac{N}{2}} \sum_{\mu=1}^n c_{\mu i}^* \xi_\mu^* \sum_{\nu=1}^n c_{i\nu} \xi_\nu = \sum_{\nu,\mu=1}^n \left[ 2 \sum_i^{\frac{N}{2}} c_{\mu i}^* c_{i\nu} \right] \xi_\mu^* \xi_\nu,$$

dann wird der in eckige Klammern gesetzte Ausdruck als reellwertige *Dichtematrix* bezeichnet,

$$\mathbf{D}_{\mu\nu} := 2 \sum_i^{\frac{N}{2}} c_{\mu i}^* c_{i\nu} = 2 \sum_i^{\frac{N}{2}} c_{\mu i} c_{i\nu}^*. \quad (4.2.7)$$

Diese Matrix bestimmt offensichtlich die Ladungsdichte  $\rho$  vollständig zu einer gegebenen Menge von Basisfunktionen.

### Fockmatrix

Der Fockoperator  $\hat{F}$  besteht aus vier Teilen, einem kinetischen Teil  $\hat{T}$  und drei Potentialteilen: anziehende Kern-Elektron-Wechselwirkung  $\hat{V}^{\text{nucl}}$ , abstoßender Elektron-Elektron-Wechselwirkung  $\hat{J}$  und Austausch-Wechselwirkung  $\hat{K}$ .

Wir schreiben nun die Matrixelemente der Fockmatrix explizit auf, siehe auch [168, Abschn. 3.4.4]. Diese ergeben sich dabei aus dem Fockoperator mittels der gewählten Basisfunktionen  $\{\xi_\mu\}$  wie folgt als

$$\mathbf{F}_{\mu\nu} = \int \xi_\mu^*(\mathbf{r}_1) \hat{F} \xi_\nu(\mathbf{r}_1) d\mathbf{r}_1. \quad (4.2.8)$$

Damit ergibt sich das entsprechende Matrixelement für jeden einzelnen Operatorteil,

$$\mathbf{F}_{\mu\nu} = \mathbf{T}_{\mu\nu} + \mathbf{V}_{\mu\nu}^{\text{nucl}} + \mathbf{J}_{\mu\nu} - \mathbf{K}_{\mu\nu} \quad (4.2.9)$$

mit

$$\mathbf{T}_{\mu\nu} := \int d\mathbf{r} \xi_\mu^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2\right) \xi_\nu(\mathbf{r}) \quad (4.2.10a)$$

$$\mathbf{V}_{\mu\nu}^{\text{nucl}} := \sum_{c=1}^M Z_c \int \frac{1}{\|\mathbf{r} - \mathbf{R}_c\|} \xi_\mu^*(\mathbf{r}) \xi_\nu(\mathbf{r}) d\mathbf{r} \quad (4.2.10b)$$

$$\begin{aligned} \mathbf{J}_{\mu\nu} &:= 2 \sum_i^{\frac{N}{2}} \sum_{\lambda\sigma} c_{i\lambda}^\dagger c_{i\sigma} \left\{ \iint \xi_\mu^*(\mathbf{r}_1) \xi_\nu(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \xi_\sigma^*(\mathbf{r}_2) \xi_\lambda(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \right\} \\ &= \sum_{\lambda\sigma} \mathbf{D}_{\lambda\sigma} \left\{ \iint \xi_\mu^*(\mathbf{r}_1) \xi_\nu(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \xi_\sigma^*(\mathbf{r}_2) \xi_\lambda(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \right\} \end{aligned} \quad (4.2.10c)$$

$$\mathbf{K}_{\mu\nu} := \frac{1}{2} \sum_{\lambda\sigma} \mathbf{D}_{\lambda\sigma} \left\{ \iint \xi_\mu^*(\mathbf{r}_1) \xi_\lambda(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \xi_\sigma^*(\mathbf{r}_2) \xi_\nu(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \right\}, \quad (4.2.10d)$$

worin die zuvor schon erwähnte Dichtematrix  $\mathbf{D}_{\lambda\sigma}$  nun explizit auftritt. Dies ist wichtig, da wir somit die Fockmatrix in einen Teil, abhängig von den Koeffizienten  $c_{i\mu}$  der Einteilchen-Wellenfunktionen, und einen getrennten koeffizientenunabhängigen Teil aufspalten können. In der Dichtematrix steckt die Nicht-Linearität des Problems, denn sie ist von den räumlichen Orbitalen abhängig.

Offensichtlich können wir diesen vier Matrizen zwei Gruppen zuordnen, je nach dem über wieviele Elektronortskordinaten integriert wird. Üblicherweise spricht man jedoch nach der Anzahl an Basisfunktionen von Zwei- und Vier-Zentren-Integralen<sup>4</sup>.

Besonders für das Integral über zwei Elektronortskordinaten  $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{R}^3$  und über vier Basisfunktionen mit Indizes  $\mu, \nu, \sigma$  und  $\lambda$  – also ein sogenanntes Vier-Zentren-Integral, siehe auch [168, Abschn. 2.3.2] – bietet es sich an, es kompakt folgendermaßen zu schreiben,

$$\mathbf{I}_{\mu\nu\sigma\lambda} := \iint \xi_\mu^*(\mathbf{r}_1) \xi_\nu(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \xi_\sigma^*(\mathbf{r}_2) \xi_\lambda(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \quad (4.2.11)$$

<sup>4</sup>Der Name wird später verständlich, wenn wir lokale Basisfunktionen mit spezifischem Aufpunkt und exponentiellem Abfall einführen. Der Aufpunkt entspricht dann dem hier genannten Zentrum.

Wir bemerken, dass  $I_{\mu\nu\sigma\lambda}$  invariant unter gleichzeitiger Vertauschung  $\mu \leftrightarrow \sigma$  und  $\nu \leftrightarrow \lambda$  ist:  $I_{\mu\nu\sigma\lambda} = I_{\sigma\lambda\mu\nu}$ .

### Überlappmatrix

Es ist zu beachten, dass es zwei Arten von Überlappmatrizen gibt: Zum einen kann sie den Überlapp zwischen den Basisfunktionen beschreiben,

$$\mathbf{S}_{\mu\nu}^{\text{basis}} = \int \xi_{\mu}^*(\mathbf{r})\xi_{\nu}(\mathbf{r})d\mathbf{r}, \quad (4.2.12)$$

und hängt damit zusammen, dass (4.1.16) zu einer verallgemeinerten Eigenwertgleichung wird.

Zum anderen kann die Überlappmatrix den Überlapp zwischen den räumlichen Orbitalen beschreiben, siehe [168, Sec. 3.4.5],

$$\mathbf{S}_{ij}^{\text{orbital}} = \int \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})d\mathbf{r} = \sum_{\mu\nu=1}^n c_{\mu i}^*c_{j\nu} \int \xi_{\mu}^*(\mathbf{r})\xi_{\nu}(\mathbf{r})d\mathbf{r} = \sum_{\mu\nu=1}^n c_{\mu i}^*c_{j\nu}\mathbf{S}_{\mu\nu}^{\text{basis}}. \quad (4.2.13)$$

Wir halten aber fest, dass (4.2.13) zur Einheitsmatrix wird, sobald die Menge der  $\phi_i$  Eigenvektoren von  $\mathbf{F}$  unter entsprechender orthonormaler Transformation geworden sind bzw. unter entsprechenden Nebenbedingungen, siehe Abschnitt 4.1.5.

#### 4.2.3. Invarianz der Lösung

Auch wenn wir die Matrizen in (4.2.3) nun kennen, können wir sie aufgrund der Überlappmatrix  $\mathbf{S}$  nicht mit bekannten Standardmethoden für Eigenwertprobleme lösen. Es bietet sich an, sie zunächst in ein einfaches Eigenwertproblem zu überführen, um dann mit den bekannten Methoden fortzufahren, siehe [168, S. 146].

#### Unitäre Transformation

Die Rückführung auf ein einfaches Eigenwertproblem geschieht durch Transformation der nicht notwendigerweise orthonormalen Basis, und damit der Überlappmatrix  $\mathbf{S}$ , so dass die Matrix zur Identität wird. Da  $\mathbf{S}$  eine hermitesche, symmetrische Matrix ist, lässt sich dies wie folgt erreichen:

$$\mathbf{X}^{\dagger}\mathbf{S}\mathbf{X} = \mathbb{I}, \quad (4.2.14)$$

wobei  $\mathbf{X}$  eine komplexwertige, symmetrische Matrix ist. Dies folgt nach dem Spektralzerlegungstheorem, siehe Zhang [190, Theorem 3.4].

Wir erinnern, dass für unitäre Matrizen

$$\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{U}\mathbf{U}^{\dagger} = \mathbb{I}. \quad (4.2.15)$$

gilt und sie Eigenwerte vom Betrag eins besitzen.

#### 4. Hartree-Fock

Daher lässt sich die gesuchte allgemeine Matrix  $\mathbf{X}$  in ein Produkt aus zwei Komponenten zerlegen: eine unitäre Matrix  $\mathbf{U}$ , welche  $\mathbf{S}$  auf Diagonalgestalt bringt,

$$\mathbf{U}^\dagger \mathbf{S} \mathbf{U} = \text{diag}(s_1, \dots, s_n) \quad (4.2.16)$$

mit den Eigenwerten  $s_i$  von  $\mathbf{S}$ , und eine zweite Matrix  $\mathbf{s}^{-\frac{1}{2}}$ , welche schlicht die Diagonalmatrix der Wurzeln der reziproken Eigenwerte darstellt.

Für die Matrix  $\mathbf{X}$  gibt es üblicherweise zwei Arten, sie aus  $\mathbf{U}$  und  $\mathbf{s}$  zu bilden, siehe [168, Abschn. 3.4.5], wobei eine davon bei sehr kleinen Eigenwerten numerisch stabiler ist. Diese kann man auch als eine Cholesky-Zerlegung verstehen, siehe [18].

Mittels der Transformationsmatrix  $\mathbf{X}$  lässt sich die Koeffizientenmatrix  $\mathbf{C}$  auf zwei verschiedene Arten umwandeln:

**Zeilentransformation** Die Zeilentransformation mittels  $\mathbf{C} = \mathbf{X} \mathbf{C}'$  verändert Überlappmatrix  $\mathbf{S}$ , ebenso wie die Fockmatrix  $\mathbf{F}$ . Hiermit erreichen wir also ein einfaches Eigenwertproblem.

**Spaltentransformation** Eine Transformation der Spalten mittels  $\mathbf{C} = \mathbf{C}'' \mathbf{X}$  beeinflusst wiederum die Matrix  $\mathbf{E}$ , welche dann im Allgemeinen keine Diagonalgestalt mehr hat. Nur die sogenannten *kanonischen Orbitale* erfüllen die speziellere Diagonalgestalt, siehe [168, Abschn. 3.2.3]. Die Gesamtenergie bleibt davon jedoch unberührt, weil sie nur von der unter beiden Transformationen invarianten Dichtematrix abhängt. Diese Transformation lässt sich benutzen, um die räumlichen Orbitale in eine bestimmte Form zu überführen, bspw. um sie maximal zu lokalisieren, siehe bspw. Edmiston und Ruedenberg [50].

### 4.3. Komplexität

In diesem Abschnitt untersuchen wir nun die algorithmische Komplexität, welche zum Lösen der HF-Gleichung für eine beliebige Basis notwendig ist. Nachdem wir kurz die asymptotische Notation eingeführt haben, untersuchen wir zunächst die Kosten beim Aufstellen der Fockmatrix in Abhängigkeit zur Anzahl an Basisfunktionen. Anschließend betrachten wir die algorithmische Gesamtkomplexität. Abschließend untersuchen wir das prä-asymptotische Laufzeitverhalten des Algorithmus.

#### Asymptotische Notation

Die Komplexität einer Methode wird über die asymptotische Notation beschrieben. Dabei betrachtet man die Abhängigkeit der Laufzeit der Methode  $T(n)$  von der Anzahl der Freiheitsgrade  $n$ . Wir geben hier knapp die Ausführungen von Cormen et al. in [37, S.41ff] wieder, da wir diese Notation immer wieder benutzen werden.

Wir geben dabei nur die Definition einer oberen Schranke an.

**Definition 24** (Obere Schranke). Sei  $g(n)$  mit  $n \in \mathbb{N}$  eine Funktion. Dann definieren wir

$$\mathcal{O}(g(n)) = \{f(n) | \exists c \in \mathbb{R}^+ \text{ und } n_0 \in \mathbb{N} : 0 \leq f(n) \leq cg(n), \forall n \leq n_0\}.$$



Hier wird also eine Funktion  $f(n)$  durch eine Konstante  $c$  mal  $g(n)$  nur nach oben asymptotisch abgeschätzt.

Wächst die Laufzeit  $T(n)$  also prinzipiell linear mit der Anzahl der Freiheitsgrade  $n$ , dann sagt man,  $T(n)$  habe die asymptotische (obere) Schranke, oder auch knapper  $T(n) = \mathcal{O}(n)$ .<sup>5</sup>

Es ergeben sich noch folgende Rechenregeln.

**Lemma 3** (Rechenregeln). *Seien  $f(n)$ ,  $g(n)$  und  $h(n)$  mit  $n \in \mathbb{N}$  Funktionen. Dann gelten folgende Rechenregeln:*

- Transitivität:

$$f(n) = \mathcal{O}(g(n)) \text{ und } g(n) = \mathcal{O}(h(n)) \Rightarrow f(n) = \mathcal{O}(h(n)), \quad (4.3.1)$$

- Reflexivität:

$$f(n) = \mathcal{O}(f(n)), \quad (4.3.2)$$

#### 4.3.1. Komplexität der Fockmatrix

Wir haben bereits im Abschnitt 4.2.2 die Elemente der Fockmatrix aufgeschrieben. Dabei war ersichtlich, dass sich die beteiligten Matrizen in zwei Gruppen, abhängig von der Anzahl der involvierten Basisfunktionen, einordnen ließen. Nun untersuchen wir die Kosten, welche jede Matrix einer Gruppe beim Aufstellen verursacht. Die Kosten sind dabei eine Funktion in Abhängigkeit der Anzahl an Basisfunktionen  $n$ , welche wir mit der asymptotischen Notation beschreiben.

**Überlapp** Zur Transformation der Fockmatrix muss die Überlappmatrix diagonalisiert werden, um das verallgemeinerte Eigenwertproblem in ein normales zu überführen. Die Aufstellung der Überlappmatrix mittels (4.2.12) verursacht Kosten, die sich mit  $\mathcal{O}(n^2)$  zur Anzahl an Basisfunktionen verhalten.

**Kinetischer Term** Das Aufstellen des kinetischen Terms mit (4.2.10a) bedingt Kosten, die gleichfalls quadratisch mit der Anzahl an Basisfunktionen anwachsen.

**Anziehender Coulomb Term** Die Matrixelemente der Kern-Elektron-Wechselwirkung enthalten zusätzlich eine Summe über die Kerne, siehe (4.2.10b). Da die Anzahl der Kerne proportional zur Anzahl der Elektronen ist und diese wiederum proportional zur Anzahl an Basisfunktionen, hat diese Matrix also effektiv Kosten von der Größenordnung  $\mathcal{O}(n^3)$ .

<sup>5</sup>Diese "mißbräuchliche" Gleichsetzung hat besondere Vorteile, siehe [37, S.42]

#### 4. Hartree-Fock

**Austausch und abstoßender Coulomb Term** Mittels (4.2.11) geben wir die Austausch-Korrelation (4.2.10d) noch einmal kompakt wieder, siehe [168, Abschn. 3.4.4],

$$\mathbf{K}_{\mu\nu} = \frac{1}{2} \sum_{\sigma\lambda} \mathbf{D}_{\sigma\lambda} \mathbf{I}_{\mu\sigma\lambda\nu}. \quad (4.3.3)$$

Fast identisch schreiben wir ebenso die Coulomb-Wechselwirkung (4.2.10c) zwischen den Elektronen,

$$\mathbf{J}_{\mu\nu} = \sum_{\sigma\lambda} \mathbf{D}_{\sigma\lambda} \mathbf{I}_{\mu\nu\sigma\lambda}. \quad (4.3.4)$$

Beide Matrizen haben jeweils Kosten von der Ordnung  $\mathcal{O}(n^4)$ .

Insgesamt erwarten wir also, dass sich die Komplexität beim Aufstellen der Fockmatrix wie  $\mathcal{O}(n^4)$  verhält.

#### 4.3.2. Algorithmische Komplexität

Wir wollen nun die Gesamtkomplexität betrachten, dazu geben wir den algorithmischen Ablauf der klassischen HF-Methode wieder, wie er in Szabo und Ostlund [168, Abschn. 3.4.6] genauer beschrieben ist, und geben zu jedem einzelnen Punkt die Ordnung der anfallenden Kosten in Abhängigkeit zur Anzahl an Basisfunktionen  $n$  an. Wir gehen dabei davon aus, dass die Kosten der Diagonalisierung einer allgemeinen  $n \times n$ -Matrix sich wie  $\mathcal{O}(n^3)$  verhalten.

1. Mit gegebenem Molekül Aufstellen eines Basissatzes  $\{\xi_\mu\}$ :  $\mathcal{O}(n)$ .
2. Aufstellen des koeffizientenunabhängigen Teils der Fockmatrix: Zwei-Zentren-Integral, Vier-Zentren-Integral, Überlappmatrix:  $\mathcal{O}(n^4)$ .
3. Diagonalisieren der Überlappmatrix  $\mathbf{S}^{\text{basis}}$  und Aufstellen der Transformationsmatrix  $\mathbf{X}$ :  $\mathcal{O}(n^3)$ .
4. Aufstellen einer initialen Dichtematrix  $\mathbf{D}$ :  $\mathcal{O}(n^2)$ .
5. Aufstellen der vollständigen Fockmatrix aus der aktuellen Dichtematrix:  $\mathcal{O}(n^4)$ .
6. Transformation der Fockmatrix zum Zwecke eines normalen Eigenwertproblems:  $\mathcal{O}(n^3)$ .
7. Diagonalisieren der transformierten Fockmatrix:  $\mathcal{O}(n^3)$ .
8. Rücktransformation der Koeffizienten der Lösung und Erneuern der Dichtematrix:  $\mathcal{O}(n^3)$ .
9. Wenn noch nicht konvergiert, dann gehe zu Schritt 5. Andernfalls Ende.

Größtes Gewicht hat dann das einmalige Berechnen der koeffizientenunabhängigen Vier-Zentren-Integrale mit einer Komplexität von  $\mathcal{O}(n^4)$ , welches außerdem die Berechnungen auch sehr speicherintensiv macht. Bei der direkten SCF-Methode werden deswegen nur die mittels eines Ausschlusskriteriums, siehe Abschnitt 7.1.2, als notwendigen

gegebenen Vier-Zentren-Integrale in Schritt 5 immer wieder neu berechnet, siehe [2, 87]. Weiterhin großen Anteil an der Gesamtkomplexität pro Iterationsschritt haben die Matrixdiagonalisierungen und -multiplikationen mit einer Komplexität von  $\mathcal{O}(n^3)$ . Diese Diagonalisierung kann vermieden werden, indem mittels einer Gradientensuche nur die Dichtematrix variiert wird. Darauf gehen wir in Abschnitt 7.6 noch ein.

### Anfängliche Dichtematrix

Wir bemerken, dass wir den Schritt 9, nämlich die Anzahl der Iterationsschritte, nicht weiter aufgeschlüsselt haben. Diese Anzahl sollte durch geschickte Wahl der anfänglichen Koeffizienten bzw. der Dichtematrix unabhängig von  $n$  gemacht werden. Dabei wird meist ein noch einfacheres Modell als die HF-Gleichung gelöst.

Im Folgenden seien einige verschiedene Heuristiken stichwortartig aufgelistet, welche meist auf dem Lösen eines einfacheren Ersatzproblems beruhen:

1. Kernhamiltonian, siehe Szabo und Ostlund [168, S. 148] bzw. Van Lenthe et al. [176, HCORE],<sup>6</sup>
2. Minimalbasis, siehe Van Lenthe et al. [176, MINGUESS],
3. Superposition of Atomic Densities (SAD), siehe Van Lenthe et al. [176, ATOMS],
4. Generalized Wolfsberg-Helmholtz (GWH) [185], entspricht einem erweiterten Hückel-Verfahren.

Die Qualität des jeweiligen Verfahrens hängt dabei im Wesentlichen von zwei Faktoren ab: Ein Erreichen des Minimums und die Anzahl an benötigten Iterationsschritten bis zum Minimum bzw. die prinzipiellen Approximationsgüte der Gesamtenergie der anfänglichen Dichtematrix.

Van Lenthe et al. [176] vergleichen empirisch die Verfahren 1, 2 und 3, wobei das SAD-Verfahren als die geeignetste Methode in punkto Iterationsschritte und Approximationsgüte der Gesamtenergie herauskommt.

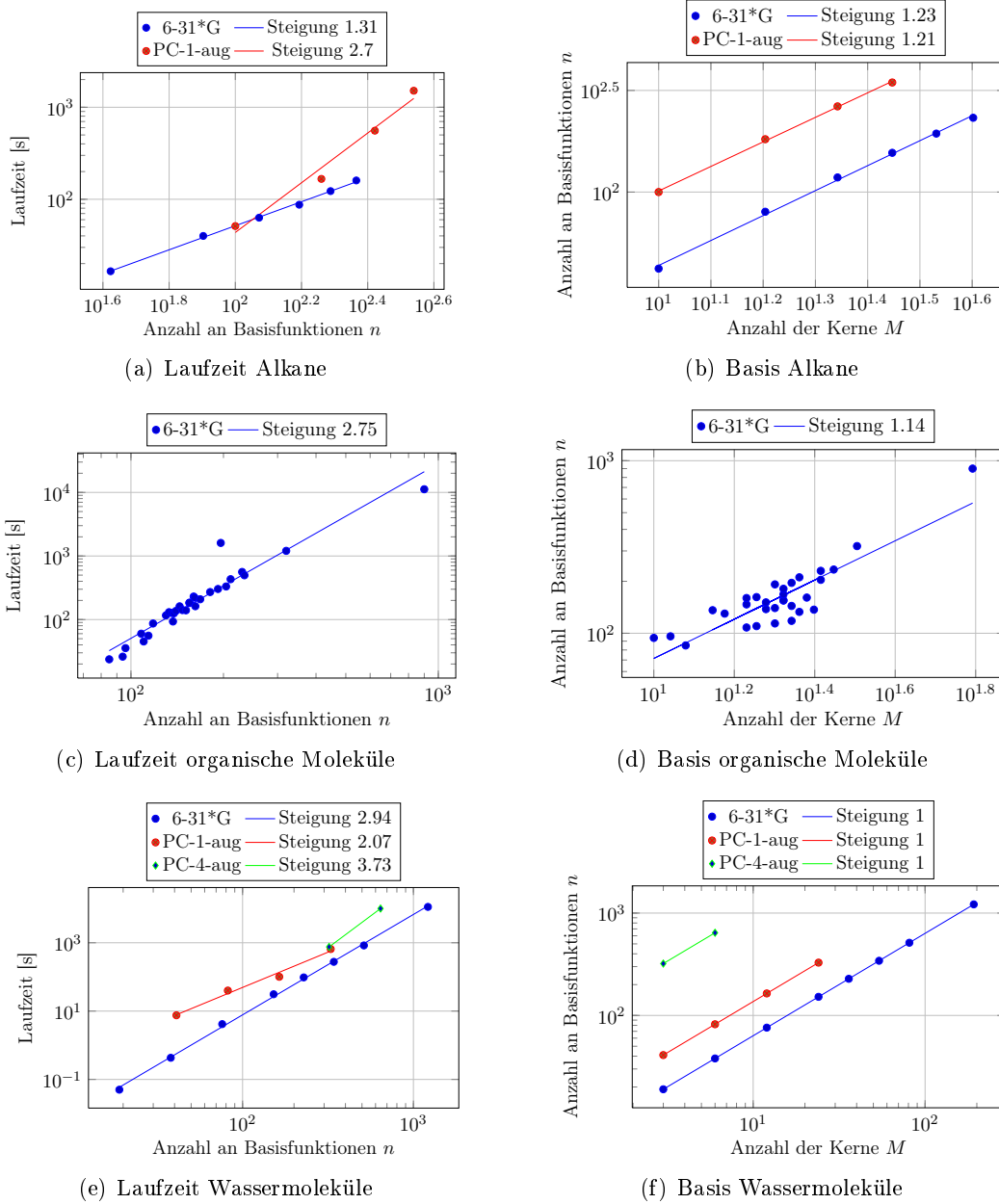
Wir gehen deswegen im Folgenden von der SAD-Methode aus, welche in vielen quantenchemischen Programmen wie [175], [152], [174], [45], [47] implementiert und die Standardeinstellung ist. Bei diesem Verfahren wird das molekulare System in einzelne neutrale Atome aufgeteilt und deren jeweilige Dichtematrix berechnet. Die resultierenden Blöcke werden zur anfänglichen Dichtematrix für das molekulare System durch Projektion in den größeren Ansatzraum und dortige Superposition zusammengeführt.

#### 4.3.3. Prä-asymptotisches Skalierungsverhalten

Wir wollen kurz die Komplexität beispielhaft anhand des HF-Programms MPQC [108] aufzeigen. Es gibt sehr viele Implementierungen der zuvor skizzierten Lösungsabfolge und wir geben hier keine repräsentative Schnittmenge wieder, sondern benutzen das Programm, welches uns auch in unseren späteren Experimenten als Basis dienen wird.

<sup>6</sup>Hierbei wird einfach die anfängliche Dichtematrix  $\mathbf{D}^{(0)} = 0$  gesetzt.

#### 4. Hartree-Fock



**Abbildung 4.1.:** *Links:* Darstellung der Komplexität der Anzahl an Basisfunktionen gegen die Rechenzeit einer typischen Hartree-Fock-Implementation. *Rechts:* Darstellung der Abhängigkeit der Anzahl an Basisfunktionen von der Anzahl der Kerne. Dabei wurden Alkane ((a), (b)), einige organische Moleküle ((c), (d)) und speziell Wassermoleküle ((e), (f)) betrachtet.

Zur Verdeutlichung der Alltagskomplexität haben wir drei Testsysteme näher untersucht: Alkane verschiedener Länge, unterschiedlich viele Wassermoleküle und eine Auswahl meist organischer kleinerer Molekülsysteme verschiedener Größe. Dabei haben wir verschiedene Standardbasen unterschiedlicher Güte benutzt. Die Anzahl der Basisfunktionen, mit denen die Elektronen eines Elements dargestellt werden, ist dabei nicht linear mit der Anzahl der Kerne, sondern schwankt, siehe Abbildung 4.1(d), da jedes Element eine unterschiedliche Elektronenzahl im neutralen Zustand besitzt und jedes Element daher auch eine unterschiedliche Anzahl an Basisfunktionen verursacht. Dies tritt bei dem organischen Testsatz hervor, bei dem die Elemente nicht proportional mit  $M$  zunehmen.

In Abbildung 4.1 (links) ist die Anzahl an Basisfunktionen gegen die Laufzeit aufgetragen. Dabei ist die Lösung bis auf auf eine relative Änderung von weniger als  $10^{-8}$  konvergiert.

Die Komplexität, d. h. die Laufzeit  $T(M)$ , zeigt das Verhalten von  $\mathcal{O}(M^4)$ . Für kleine Molekülsysteme und sehr gut auflösende Basen überwiegt also vor allem das Aufstellen der Vier-Zentren-Integrale. Dass die Komplexität im Einzelnen aber durchaus kleiner als  $n^4$  sein kann, liegt an noch nicht näher ausgeführten Abschätzungen, die durch die endliche Genauigkeit der numerischen Berechnung und anhand einer konkreten Basis möglich werden. Dies führen zu einem Ausschlusskriterium für die Vier-Zentren-Integrale und diese sind natürlich in MPQC integriert. Darauf werden wir im folgenden Kapitel genau eingehen, wobei unser Ziel sein wird, die dargestellte Methode so zu verändern, dass die Kosten schließlich nur noch linear mit der Anzahl an Basisfunktionen steigen.

Strout und Scuseria [165] untersuchten analog das (prä-asymptotische) Skalierungsverhalten  $n^\alpha$  der HF-Methode, wobei  $\alpha$  der Skalierungsexponent ist. Sie betrachteten einzelne Graphenlagen und Diamantstücke wachsender Größe und vernachlässigten dabei solche Vier-Zentren-Integrale, welche nach obig erwähntem Ausschlusskriterium keinen relevanten Anteil darstellen. Die größten untersuchten Systeme hatten dabei ca. 550 bzw. 650 Kerne. Dabei definierten sie den Skalierungsexponenten als

$$\alpha = \frac{\ln(I_2/I_1)}{\ln(n_2/n_1)}, \quad (4.3.5)$$

wobei  $I_1$  bzw.  $I_2$  die Anzahl der Vier-Zentren-Integrale und  $n_1$  bzw.  $n_2$  die Anzahl der Basisfunktionen für ein System 1 bzw. 2 bezeichnet. Sprich, die Skalierung wurde aus der Steigung zwischen zwei aufeinanderfolgenden Messpunkten in logarithmischer Skala bestimmt.

Allgemein fällt der Skalierungsexponent mit wachsender Anzahl Kerne eines Systems. Für Graphen fanden die Autoren einen (prä-asymptotischen) Grenzwert von  $\alpha = 2,1$ , für Diamant von  $\alpha = 2,4$ . Sehr kleine Systeme zeigten noch einen Exponenten  $\alpha > 3$ .

Wir bemerken, dass sich dies mit unseren Beispielrechnungen deckt. Insgesamt ist unter Ausschluss bestimmter Vier-Zentren-Integrale der realistische Skalierungsexponent  $2 < \alpha < 4$ . Die HF-Methode, auch mit sogenanntem Screening der Vier-Zentren-Integrale, siehe dazu auch Helgaker et al. [91, Abschn. 9.12], ist damit in jedem Fall *nicht* linear skalierend.

#### 4.3.4. Analyse des Laufzeitverhaltens

Wir wollen in diesem Abschnitt abschließend zur Betrachtung der Komplexität der HF-Methode untersuchen, inwieweit bestimmte Eigenschaften der Basisfunktionen es erlauben würden, die Vier-Zentren-Integrale in einer Laufzeit, die mit weniger als  $\mathcal{O}(n^4)$  skaliert, zu berechnen. Wir wollen hier jedoch keinen ernstzunehmenden Ansatz für die Basisfunktionen entwickeln, sondern dies soll als Vorbereitung auf das folgende Kapitel dienen.

##### Vier-Zentren-Integrale

Zur Berechnung der inneren Produkte über den Fockoperator, siehe (4.2.8), sind abstandsgewichtete Integrale über bis zu zwei Elektronortskoordinaten jeweils im Ganzraum zu lösen. Entwickelt man dabei den Operator in einer Basis, ist die paarweise Wechselwirkung bzw. die Wechselwirkung von bis zu vier Basisfunktionen im Integranden unvermeidlich.

Betrachten wir erneut die Definition des Vier-Zentren-Integrals, siehe (4.2.11),

$$I_{\mu\nu\sigma\lambda} := \iint \xi_\mu^*(\mathbf{r}_1)\xi_\nu(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \xi_\sigma^*(\mathbf{r}_2)\xi_\lambda(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2.$$

Da wir  $\mathcal{O}(n^4)$  dieser Terme berechnen müssen, können wir dieses Laufzeitverhalten nur durchbrechen, wenn wir im Vorneherein wissen, dass bestimmte Terme gleich null sind. Wir beobachten, dass jeweils zwei Basisfunktionen über eine gemeinsame Elektronortskoordinate integriert werden. Wenn wir keine konkrete Aussage über das Produkt zweier Basisfunktionen voraussetzen wollen, dann verbleibt nur der leere Schnitt der beiden kompakten Träger der an der Integration über eine Elektronortskoordinate beteiligten Basisfunktionen. Dazu definieren wir mit  $\mathcal{I}_\mu$  die folgende Menge von Indizes  $\nu$  zu einer Basisfunktion  $\xi_\mu$  mit einem Träger  $\mathcal{B}_\mu$ ,

$$\mathcal{I}_\mu := \{\nu \in \{1, \dots, n\} \mid \mathcal{B}_\mu \cap \mathcal{B}_\nu \neq \emptyset\}, \quad (4.3.6)$$

also diejenige Menge der Indizes von Basisfunktionen, deren kompakte Träger nicht-leeren Schnitt zur Basisfunktion  $\xi_\mu$  haben.

Konkretisieren wir obigen Gedankengang in folgendem Lemma:

**Lemma 4** (Quadratisches Laufzeitverhalten bei kompakter Basis). *Sei jede Basisfunktion  $\xi_\mu$  kompakt mit ihrem Träger eingebettet in die abgeschlossene Kugel  $\mathcal{B}_{\mathbf{r}_{\mu,0}}(R_{\mu,0})$  mit Radius  $R_{\mu,0}$  und Aufpunkt  $\mathbf{r}_{\mu,0}$ . Es gelte weiterhin, dass jede Indexmenge (4.3.6)  $|\mathcal{I}_\mu| = \mathcal{O}(1)$  für  $\mu \in \{1, \dots, n\}$  erfüllt.*

*Dann sind nur  $\mathcal{O}(n^2)$  Vier-Zentren-Integrale (4.2.11) ungleich null.*

*Beweis.* Sei die Funktion  $\xi_\mu(\mathbf{r}_1)$  gegeben, dann lässt sich aufgrund deren Kompaktheit das Integrationsgebiet in (4.2.11) von  $\mathbb{R}^3$  auf  $\mathcal{B}_{\mathbf{r}_{\mu,0}}(R_{\mu,0})$  einschränken. Damit ist das Integral über  $\mathbf{r}_1$  nur für Funktionen  $\xi_\nu$  mit  $\nu \in \mathcal{I}_\mu$  ungleich null. Es gilt aber  $|\mathcal{I}_\mu| = \mathcal{O}(1)$ . Analoges folgt für die Funktion  $\xi_\sigma$  und die zweite Elektronortskoordinate.  $\square$

Wenn also die Basisfunktionen kompakte Träger haben, die Dichte der Aufpunkte  $\mathbf{r}_{\mu,0}$  dünn bzw. die Radien  $R_{\mu,0}$  klein sind – woraus  $|\mathcal{I}_\mu| = \mathcal{O}(1)$  folgt – dann ergibt sich ein quadratisches Laufzeitverhalten.

### Problemfall: Coulomb-Wechselwirkung

Wir bemerken, dass wir dieses Laufzeitverhalten nicht mittels zusätzlicher Einschränkungen an die mögliche Basis weiter verbessern können. Der Coulomb-Teil des Fockoperators lässt zwei Elektronortskordinaten mit ihrem inversen Abstand  $r$  als Gewicht wechselwirken. Die Funktion  $\frac{1}{r}$  fällt aber im  $\mathbb{R}^3$  zu langsam ab, als dass wir das Integrationsgebiet ohne größeren Fehler analog einschränken könnten: Wechseln wir in ein sphärisches Koordinatensystem, dann wächst die Oberfläche über  $\phi$  und  $\theta$  der Kugel mit  $r^2$  und das Integral

$$\int_{\mathbb{R}} \int_0^{2\pi} \int_0^\pi \frac{1}{r} dr d\phi d\theta = 4\pi \int_{\mathbb{R}} r dr$$

konvergiert im Allgemeinen nicht. Im  $d$ -dimensionalen Fall gilt, dass die Funktion schneller als  $r^d$  abfallen muss, siehe [123, S. 275].

Nach Lemma 4 verbleiben zwei unabhängige Summationsindizes für das Coulomb-Matrixelement (4.3.4). Diese Doppelsumme lässt sich aber aufgrund des zu langsamen Abfalls von  $\frac{1}{r}$  zunächst nicht weiter einschränken.

### Matrixdiagonalisierung

Als zweite Hürde auf dem Weg zu einer linear skalierenden Laufzeit identifizieren wir die notwendigen Diagonalisierungen der Fockmatrix und der Überlappmatrix mit einer Komplexität von  $\mathcal{O}(n^3)$ .

Geringere Kosten werden erst möglich, wenn etwas über die Struktur der zu diagonalisierenden Matrix bekannt ist. Schnellere Verfahren benötigen die Eigenschaft der Dünnbesetztheit. Diese kann in speziellen Formen wie Blockdiagonal- oder Bandmatrizen vorliegen.

Wie wir noch sehen werden, überwinden wir diese zweite Hürde, wenn wir Eigenschaften festlegen, um die erste Hürde, die mit  $\mathcal{O}(n^4)$  anwachsende Berechnung aller Vier-Zentren-Integrale, zu überwinden.

## Zusammenfassung

In diesem Kapitel befassten wir uns mit der HF-Gleichung. Diese entstand, indem wir mit Hilfe der Korrelationsprojektoren aus Abschnitt 2.5.3 die elektronische Wellenfunktion zunächst auf die vollständig antisymmetrische, unkorrelierte Produktform aus Einteilchen-Wellenfunktionen projizierten und diese in die elektronische Schrödingergleichung einsetzten. Die resultierende Ein-Teilchen-Gleichung ist nicht-linear, weil sie von den untersten  $N$  Lösungen der Gleichung abhängt, den besetzten Zuständen des Systems. Bei diesen Einteilchen-Wellenfunktionen ist zusätzlich der Spin des Elektrons zu berücksichtigen. Wir zogen uns jedoch auf den CLHF-Fall zurück, indem der Gesamtspin des Systems

#### 4. Hartree-Fock

stets null ist und alle Einteilchen-Wellenfunktionen als doppelt besetzte, rein räumliche Orbitale aufgefasst werden können.

Setzt man eine Basis für diese räumlichen Orbitale an, so lässt sich die HF-Gleichung zu einer Eigenwertgleichung entwickeln und damit numerisch lösen. Diese sogenannte Roothaan-Hall-Gleichung stellt ein verallgemeinertes Eigenwertproblem dar, welches mittels einer unitären Transformationen in ein gewöhnliches Eigenwertproblem überführt werden kann. Weiter konnten wir mit Hilfe des Vier-Zentren-Integrals die einzelnen Teile der Fockmatrix kompakt schreiben. Dabei ließ sich der nicht-lineare Anteil der Fockmatrix getrennt als Dichtematrix schreiben.

Abschließend untersuchten wir die allgemeine Komplexität der HF-Methode. Wir sahen, dass diese prinzipiell  $\mathcal{O}(n^4)$  aufgrund der Produkte von vier Basisfunktionen beim Aufstellen der Fockmatrix beträgt. Wir bemerkten aber auch anhand numerischer Beispiele, dass die Komplexität im Einzelfall kleiner sein kann.

In den folgenden zwei Kapiteln befassen wir uns mit zwei getrennten Bausteinen, der Lokalität einer Funktion und dem Eigenwertproblem zu dünnbesetzten, indefiniten Matrizen. Mit diesen wird es uns anschließend gelingen, anstelle der obigen Komplexität eine linear mit der Anzahl an Basisfunktionen skalierende Methode vorzustellen.



## 5. Lokalität

Die im Abschnitt 4.3.4 über das Vier-Zentren-Integral betrachtete Kompaktheit des Trägers ist ein sehr strenges Kriterium, um die Komplexität von  $\mathcal{O}(n^4)$  zu überwinden. Auch ist sie im Allgemeinen nicht erfüllbar. Aber bei der approximativen Berechnung im Computer genügt bereits ein wesentliches schwächeres Kriterium, nämlich das der Lokalität (des Trägers). Dieses Kriterium wollen wir in diesem Abschnitt motivieren und zu einer allgemeinen Menge lokaler Funktionen erweitern.

Wir beschränken uns also ab jetzt auf eine Näherungslösung. Im Folgenden sei deswegen stets ein  $\epsilon > 0$  gegeben, welches im Rahmen der Maschinengenauigkeit  $\epsilon^{\text{masch}}$  liegt.

Weiterhin betrachten wir in diesem Abschnitt Funktionen aus dem Raum der quadratintegrierbaren Funktionen  $\mathcal{L}_2(\mathbb{R}^3, \mathbb{C})$ , kurz  $\mathcal{L}_2$  und verweisen auf die Definitionen in Abschnitt 3.4.4.

### 5.1. Lokale Funktion

Damit können wir nun zur Tat schreiten und die angekündigte Eigenschaft der Lokalität für eine Funktion aus diesem Raum definieren:

**Definition 25** (Lokale Funktion und lokaler Träger). Sei  $\epsilon > 0$  gegeben. Sei  $f(\mathbf{x}) \in \mathcal{L}_2(\mathbb{R}^3, \mathbb{C})$ . Dann ist diese Funktion *lokal* um einen Aufpunkt  $\mathbf{x}_0 \in \mathbb{R}^3$ , wenn ein  $\epsilon$ -abhängiger Radius  $R(\epsilon) > R_0 > 0$  existiert, so dass

$$0 \leq \int_{\mathbb{R}^3} |f(\mathbf{x})|^2 dx - \int_{\mathcal{B}_{\mathbf{x}_0}(R(\epsilon))} |f(\mathbf{x})|^2 dx = \int_{\mathbb{R}^3 \setminus \mathcal{B}_{\mathbf{x}_0}(R(\epsilon))} |f(\mathbf{x})|^2 dx < \epsilon \quad (5.1.1)$$

gilt, mit dem *lokalen Träger*  $\mathcal{B}_{\mathbf{x}_0}(R(\epsilon)) \subset \mathbb{R}^3$ , der Kugel mit  $\epsilon$ -abhängigem Radius  $R(\epsilon)$  um  $\mathbf{x}_0$ , und außerdem gilt,

$$R(\epsilon) \propto -\ln(\epsilon). \quad (5.1.2)$$

Dies ist aufgrund des Absolutbetrags im Integranden und der Annahme, dass  $f(\mathbf{x})$  quadratintegrierbar ist, wohldefiniert.

Wir weisen nochmal darauf hin, dass wir dabei nur fordern, dass Funktionen ab einem Abstand  $\|\mathbf{x} - \mathbf{x}_0\| > R_0$  exponentiell abfallen.

Wir wollen nun für einige Funktionen zeigen, dass diese lokal sind.

## 5. Lokalität

**Lemma 5** (Lokale Exponentialfunktionen). *Folgende Funktionen sind nach Definition 25 lokal um ein  $\mathbf{x}_0 \in \mathbb{R}^3$ .*

$$C \cdot \exp(-a\|\mathbf{x} - \mathbf{x}_0\|) \quad (5.1.3a)$$

$$p(\mathbf{x} - \mathbf{x}_0) \cdot \exp(-a\|\mathbf{x} - \mathbf{x}_0\|) \quad (5.1.3b)$$

$$p(\mathbf{x} - \mathbf{x}_0) \cdot \exp(q(\mathbf{x} - \mathbf{x}_0)), \quad (5.1.3c)$$

wobei  $C$  und  $a > 0$  Konstanten sind,  $\|\cdot\|$  die Norm im  $\mathbb{R}^3$  bezeichnet und  $p$  und  $q$  endliche Polynome mit mindestens Grad 1 über  $\mathbf{x} - \mathbf{x}_0$  sind, wobei die Konstante nullter Ordnung von  $q$  verschwindet und  $q(\mathbf{x}) < 0$  gilt.

*Beweis.* Wir müssen also zeigen, dass zu einem gegebenen  $\epsilon > 0$  für alle  $\mathbf{x} \in \mathbb{R}^3$  gilt:  $R(\epsilon) \propto \frac{1}{-\ln(\epsilon)}$ .

Wir besprechen zunächst (5.1.3a). Wir setzen den Term in (5.1.1) ein, verschieben den Aufpunkt  $\mathbf{x}_0$  in den Ursprung aufgrund des Ganzraumintegrals und transformieren in sphärische Koordinaten,

$$\int_{\mathbb{R}^3} |f(\mathbf{x})|^2 dx = C^2 \int_{\mathbb{R}^3} \exp(-a\|\mathbf{x} - \mathbf{x}_0\|) d\mathbf{x} = 4\pi C^2 \int_0^\infty r^2 \exp(-ar) dr.$$

Dann spalten wir das Integral bei  $R(\epsilon)$  auf,

$$= 4\pi C^2 \left\{ \int_0^{R(\epsilon)} r^2 \exp(-ar) dr + \int_{R(\epsilon)}^\infty r^2 \exp(-ar) dr \right\}.$$

Wir müssen nun also zeigen, dass der zweite Term durch Wahl von  $R(\epsilon)$  kleiner als ein vorgegebenes  $\epsilon$  gebracht werden kann. Dies gelingt mittels des bestimmten Integrals, siehe [26, Abschn. 21.6.2 (23.)], und mittels der Stammfunktion aus [26, Abschn. 21.5.4.2 (449.)], deren Konstante wir für  $x = 0$  auf null setzen,

$$\begin{aligned} & 4\pi C^2 \int_{R(\epsilon)}^\infty r^2 \exp(-ar) dr \\ &= 4\pi C^2 \left\{ \int_0^\infty r^2 \exp(-ar) dr - \int_0^{R(\epsilon)} r^2 \exp(-ar) dr \right\} \\ &= 4\pi C^2 \left\{ \frac{2}{a^3} - \left\{ \exp(-aR(\epsilon)) \left[ \frac{R(\epsilon)^2}{-a} - \frac{2R(\epsilon)}{a^2} + \frac{2}{-a^3} \right] - \frac{2}{-a^3} \right\} \right\} \\ &= 4\pi C^2 \exp(-aR(\epsilon)) \frac{1}{a^3} \left[ a^2 R(\epsilon)^2 + 2aR(\epsilon) + 2 \right] \end{aligned}$$

Wir müssen den polynomiellen Faktor zunächst nach oben abschätzen, bevor wir ihn gleich der Maschinengenauigkeit  $\epsilon$  setzen und dann invertieren können. Da der asymptotisch dominante Term  $R(\epsilon)^2$  ist, schätzen wir für  $R(\epsilon) > R_0$  wie folgt ab,

$$\begin{aligned} \alpha \cdot a^2 R(\epsilon)^2 &\geq a^2 R(\epsilon)^2 + 2aR(\epsilon) + 2 \\ \alpha &:= 1 + \frac{2}{aR_0} + \frac{2}{a^2 R_0^2}. \end{aligned}$$

Damit erhalten wir,

$$4\pi C^2 \exp(-aR(\epsilon)) \frac{1}{a^3} \left[ a^2 R(\epsilon)^2 + 2aR(\epsilon) + 2 \right] \leq 4\pi C^2 \exp(-aR(\epsilon)) \frac{\alpha}{a^3} a^2 R(\epsilon)^2$$

für  $R(\epsilon) > R_0$ .

Unsere Funktion schreiben wir nun also als

$$f(\epsilon) = \tilde{C} \exp(-aR(\epsilon)) (aR(\epsilon))^2,$$

wobei  $\tilde{C} = 4\pi C^2 \frac{\alpha}{a^3}$  alle Konstanten zusammenfasst.

Um diesen Ausdruck nun mit  $\epsilon$  gleichzusetzen und zu invertieren,

$$R(\epsilon) = -\frac{2W\left(\frac{\sqrt{\frac{\epsilon}{\tilde{C}}}}{2}\right)}{a},$$

benötigen wir die Lambert W-Funktion, welche die Inverse zu  $f(W) = W \exp(W)$  ist. Nach Hoofar und Hassani [97] verhält sich  $W(x) < \ln(x)$  für  $x > e$  und damit folgt das Gewünschte.

Da die Exponentialfunktion schneller abfällt als jedes Polynom und nur der höchste Polynomgrad von  $p(x)$  in (5.1.3b) zum zuvor abgeschätzten Term  $R(\epsilon)^2$  hinzu kommt, folgt das Gewünschte hierfür analog – nur mit entsprechend anderer Konstante.

Die letzte Funktion (5.1.3c) kann schließlich für  $\|x - x_0\| > 1$  durch eine entsprechende Funktion (5.1.3b) nach oben abgeschätzt werden.  $\square$

Anders ausgedrückt ist eine Funktion lokal, wenn sie um einen Aufpunkt ab einem bestimmten endlichen Radius exponentiell rasch abklingt. Lokalität ist also eine schwächere Form von Kompaktheit. Analog zum kompakten Träger einer Funktion nennen wir die Kugel  $\mathcal{B}_{r_0}(R(\epsilon))$ , die die Funktion bis auf  $\epsilon$  enthält, siehe (5.1.1), deswegen den *lokalen Träger* der Funktion.

## 5.2. Menge lokaler Funktionen

Schon am Beweis von Lemma 5 wird klar, dass es sich anbietet, diese Funktionen zu einer Menge lokaler Funktionen zu erweitern:

**Definition 26** (Menge lokaler Funktionen). Die Menge lokaler Funktionen  $\mathcal{K}$  sei die Untermenge der quadratintegrablen Funktionen  $\mathcal{L}_2(\mathbb{R}^d, \mathbb{R})$ , die sich durch folgende Gestalt (5.1.3c),

$$f(x) = p(x) \cdot \exp(q(x)), \tag{5.2.1}$$

ausdrücken lassen, wobei  $p(x)$  und  $q(x)$  endliche Polynome in  $x$  sind und  $q(x)$  mindestens Ordnung 1 hat, verschwindende Konstante für die nullte Ordnung besitzt und  $q(x) < 0$  gilt.

## 5. Lokalität

Nach Lemma 5 sind alle Funktionen in dieser Menge lokal.

Wir bemerken, dass wir uns auch schlicht an Definition 25 für eine allgemeinere Definition dieser Menge hätten halten können, anstatt eine konkrete Form wie (5.1.3c) festzulegen. Im Folgenden werden wir aber nur auf Funktionen der obigen Form (5.1.3c) treffen, weswegen für unsere Zwecke diese schwächere Definition der Menge genügt.

Schließlich merken wir an, dass Funktionen mit kompaktem Träger natürlich eine Untermenge dieser lokalen Funktionen sind.

### 5.3. Lokale Basis

Wir brauchen noch eine zweite Aussage, wenn eine Gruppe lokaler Funktionen als Basis quadratische Laufzeit bei der Berechnung der Vier-Zentren-Integrale bringen soll. Wir haben in Lemma 4 in Abschnitt 4.3.4 zu den Vier-Zentren-Integralen gesehen, dass sich für ein gegebenes  $\epsilon > 0$  nur konstant viele lokale Träger überlappen. Dies legen wir nun in der folgenden Definition fest.

**Definition 27** (Lokale Basis). Sei  $\epsilon > 0$  gegeben. Eine Basis  $\{\xi_\mu(\mathbf{r})\}_{\mu=1}^n$  heißt *lokal*, wenn jede ihrer Basisfunktionen in  $\mathcal{K}$ , also lokal, ist und darüber hinaus gilt, dass die Kugeln  $\mathcal{B}_{\mathbf{r}_\mu}(R_\mu(\epsilon))$  zu jeder Basisfunktion  $\xi_\mu$  mit ihrem spezifischen Radius  $R_\mu(\epsilon)$  und Aufpunkt  $\mathbf{r}_\mu \in \mathbb{R}^3$  einen nicht-leeren Schnitt nur mit einer höchstens konstanten Anzahl von lokalen Trägern anderer Basisfunktionen  $\xi_\nu$  hat, also dass gilt,  $|\mathcal{I}_\mu| = \mathcal{O}(1)$ , siehe (4.3.6).

Anders ausgedrückt: Stellen wir uns die Basen als Knotenmenge  $V$  eines Graphen  $G = (V, E)$  vor. Weiterhin enthalte dieser Graph genau dann eine Kante  $e \in E$  zwischen Knoten  $\nu, \mu \in V$ , wenn sich die lokalen Träger der Basisfunktionen  $\xi_\nu$  und  $\xi_\mu$  überlappen. Dann soll gelten, dass der maximale Bindungsgrad des Graphen nach oben durch  $\mathcal{O}(1)$  beschränkt ist. Wir erwähnen dies in Bezug zu Benzi et al. [18], die Funktionenapproximation und deren Abfall in Zusammenhang mit dünnbesetzten Matrizen untersuchen.

Wieder anders ausgedrückt, soll also die Dichte der Kugelaufpunkte und die Radien der Kugeln derart sein, dass sich nur benachbarte Kugeln überlappen und die Anzahl der Nachbarn konstant und insbesondere nicht allgemein  $\mathcal{O}(n)$ , sondern nur  $\mathcal{O}(1)$  ist.

### 5.4. Schlussfolgerungen

Wir wollen nun Schlussfolgerungen aus den obigen Definitionen untersuchen.

#### 5.4.1. Differentialoperator-Methode

Als Vorbereitung müssen wir zunächst Produkte verschiedener lokaler Funktionen aus  $\mathcal{K}$ , insbesondere mit verschiedenen Aufpunkten, betrachten.

Wir legen dabei als Erstes dar, warum die Betrachtung der Form (5.1.3a) für Funktionen im Integranden ausreichend ist, die wir kurz wiedergeben,

$$\xi_\mu(\mathbf{r}) = C \cdot \exp(-a \cdot \|\mathbf{r} - \mathbf{r}_\mu\|), \quad (5.4.1)$$

wobei  $C$  und  $a > 0$  Konstanten sind. Wir folgen hierbei Shavitt und Karplus [153, Abschn. II.], welche diese „Differentialoperator-Methode“ wiederum von Boys [23] übernehmen.

Da sich die Funktionen  $\exp(q(x))$ , siehe (5.1.3c), durch  $\exp(-a\|x - x_0\|)$  abschätzen lassen, brauchen wir sie hier nicht weiter zu betrachten.

Wir wollen zeigen, dass wir jeden polynomiellen Vorfaktor  $p(\mathbf{r})$  zu (5.4.1) vernachlässigen können, denn es gilt nach [153],

$$\|\mathbf{r} - \mathbf{r}_\mu\| \exp(-\alpha\|\mathbf{r} - \mathbf{r}_\mu\|) = \frac{\partial}{\partial \alpha} \exp(-\alpha\|\mathbf{r} - \mathbf{r}_\mu\|) \quad (5.4.2a)$$

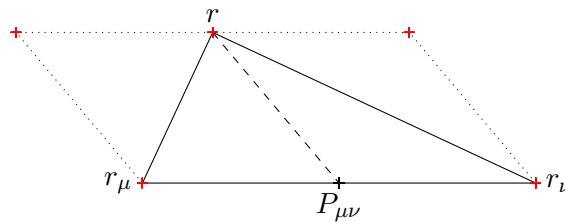
$$(\mathbf{r} - \mathbf{r}_\mu)_i \exp(-\alpha\|\mathbf{r} - \mathbf{r}_\mu\|) = -\frac{\partial}{\partial \alpha} \left[ \frac{1}{\alpha} \frac{\partial}{\partial \mathbf{r}_{\mu,i}} \right] \exp(-\alpha\|\mathbf{r} - \mathbf{r}_\mu\|). \quad (5.4.2b)$$

Da dies nur Ableitungen nach Variablen sind, über die nicht integriert wird, können wir uns Feynmans großen Trick [55, S. 86] zunutze machen: die *Differentiation unter dem Integralzeichen*, siehe auch [26, Abschn. 8.2.4.2]. Dabei dürfen wir Integration und Differentiation vertauschen.

Damit folgt, dass es zunächst genügt, Integrale über die einfacheren Funktionen (5.1.3a) zu betrachten, da die Differentiation nach obigem Muster und die Integration vertauschen. Weiterhin sind wir nur an einer Abschätzung nach oben und insbesondere an deren exponentiellem Abfall interessiert. Daher können wir einen beliebigen polynomiellen Vorfaktor in den asymptotischen Betrachtungen außen vor lassen, da die Exponentialfunktion stärker abfällt als jedes Polynom ansteigt.

### 5.4.2. Produkt lokaler Funktionen

Betrachten wir mit dieser Vereinfachung nun konkret das Produkt zweier Funktionen der Gestalt (5.1.3a), wie wir sie gleich häufig im Integranden auftauchen sehen werden.



**Abbildung 5.1.:** Darstellung der Längen und Positionen und des Parallelogramms im Beweis von Lemma 6. Die gepunkteten Linien geben dabei die Verschiebung der Grundseite  $\mathbf{r}_\mu - \mathbf{r}_\nu$  entlang  $\mathbf{r} - P_{\mu\nu}$  an, wodurch zwei gleiche Parallelogramme mit den Diagonalen  $\mathbf{r} - \mathbf{r}_\mu$  und  $\mathbf{r} - \mathbf{r}_\nu$  entstehen.

## 5. Lokalität

**Lemma 6** (Asymptotik des Produkts zweier lokaler Funktionen). *Gegeben sei ein  $\epsilon > 0$ . Seien  $\xi_\mu$  und  $\xi_\nu$  zwei (lokale) Funktionen aus  $\mathcal{K}$ .*

*Dann folgt, dass sich der Betrag des Produkts der beiden Funktionen wie folgt nach oben mit dem Mittelpunkt  $P_{\mu\nu} = \frac{1}{2}(\mathbf{r}_\mu + \mathbf{r}_\nu)$  abschätzen lässt,*

$$|\xi_\mu^*(\mathbf{r})\xi_\nu(\mathbf{r})| \leq C \cdot \exp(b\|\mathbf{r} - P_{\mu\nu}\|), \quad (5.4.3)$$

wobei  $\mathbf{r}_\mu$  bzw.  $\mathbf{r}_\nu$  die Aufpunkte der Träger,  $C \propto \exp(-\frac{a}{2}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|)$  und  $b$  Konstanten sind und  $a := \max\{a_\mu, a_\nu\}$  gilt.

*Beweis.* Wir setzen Definition 25 ein,

$$|\xi_\mu^*(\mathbf{r})\xi_\nu(\mathbf{r})| = \underbrace{|C_\mu C_\nu|}_{:=C} \cdot \exp(-a_\mu\|\mathbf{r} - \mathbf{r}_\mu\|) \exp(-a_\nu\|\mathbf{r} - \mathbf{r}_\nu\|).$$

Mit der Parallelogrammgleichung  $e^2 + f^2 = 2(g^2 + h^2)$  für die Seiten  $g, h$  und Diagonalen  $e, f$  eines Parallelogramms, siehe [26, Abschn. 3.1.4.1] und Abbildung 5.1 zur Erläuterung, und mit  $g^2 + h^2 \geq 2gh$  formen wir schrittweise um,

$$\begin{aligned} \|\mathbf{r} - \mathbf{r}_\mu\|^2 + \|\mathbf{r} - \mathbf{r}_\nu\|^2 &= 2(\|\mathbf{r} - P_{\mu\nu}\|^2 + \frac{1}{4}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|^2) \\ \|\mathbf{r} - \mathbf{r}_\mu\|^2 + 2\|\mathbf{r} - \mathbf{r}_\mu\|\|\mathbf{r} - \mathbf{r}_\nu\| + \|\mathbf{r} - \mathbf{r}_\nu\|^2 &\geq 2(\|\mathbf{r} - P_{\mu\nu}\|^2 + \frac{1}{4}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|^2) \\ (\|\mathbf{r} - \mathbf{r}_\mu\| + \|\mathbf{r} - \mathbf{r}_\nu\|)^2 &\geq \|\mathbf{r} - P_{\mu\nu}\|^2 + \frac{1}{4}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|^2 + \|\mathbf{r} - P_{\mu\nu}\|\|\mathbf{r}_\mu - \mathbf{r}_\nu\| \\ (\|\mathbf{r} - \mathbf{r}_\mu\| + \|\mathbf{r} - \mathbf{r}_\nu\|)^2 &\geq (\|\mathbf{r} - P_{\mu\nu}\| + \frac{1}{2}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|)^2. \end{aligned}$$

Wir erhalten schließlich als Abschätzung für die Summe der Argumente der beiden Exponentialfunktionen

$$\|\mathbf{r} - \mathbf{r}_\mu\| + \|\mathbf{r} - \mathbf{r}_\nu\| \geq \|\mathbf{r} - P_{\mu\nu}\| + \frac{1}{2}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|.$$

Damit folgt

$$\begin{aligned} &\leq C \cdot \exp(-a(\|\mathbf{r} - \mathbf{r}_\mu\| + \|\mathbf{r} - \mathbf{r}_\nu\|)) \\ &\leq C \cdot \exp(-a(\|\mathbf{r} - P_{\mu\nu}\| + \frac{1}{2}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|)) \\ &\leq C \cdot \exp(-a\|\mathbf{r} - P_{\mu\nu}\|) \exp(-\frac{a}{2}\|\mathbf{r}_\mu - \mathbf{r}_\nu\|). \end{aligned}$$

□

Wir weisen besonders darauf hin, dass die Konstante  $C$  in der obigen Abschätzung des Produktbetrags exponentiell mit dem Abstand der Aufpunkte abfällt.

Mit diesen Vorüberlegungen können wir jetzt die Auswirkungen der obigen Eigenschaften auf das Laufzeitverhalten betrachten.

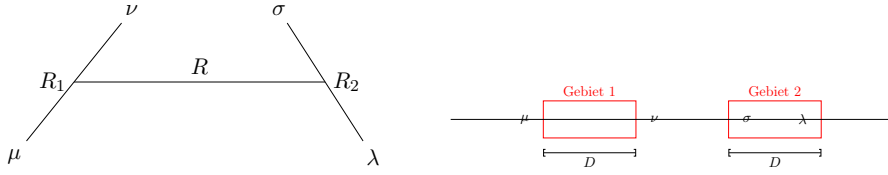
### 5.4.3. Quadratisches Laufzeitverhalten

Mit Lemma 6 und mit Definition 27 erhalten wir eine analoge Aussage zu Lemma 4:

**Lemma 7** (Quadratisches Laufzeitverhalten bei lokaler Basis). *Sei  $\epsilon > 0$  gegeben. Sei  $\{\xi_\mu(\mathbf{r})\}_{\mu=1}^n$  eine lokale Basis, siehe Definition 27, dann sind höchstens  $\mathcal{O}(n^2(\log n)^2)$  Vier-Zentren-Integrale  $I_{\mu\nu\lambda\sigma}$ , siehe (4.2.11), ungleich null.*

In Clementi [36]<sup>1</sup> wird zum ersten Mal die Idee geäußert, dass nicht alle Vier-Zentren-Integrale in gleichem Maße beitragen.

Wir folgen nun grundsätzlich der Darstellung durch Dyczmons [49], dort für Gauss-Funktionen, verallgemeinern aber den Beweis für beliebige lokale Basen, siehe Definition 27.



(a) Allgemeines Vier-Zentren-Integral  $I_{\mu\nu\sigma\lambda}$  mit Abständen  $R_1$  bzw.  $R_2$  der Zentren zu einer gemeinsamen Elektronenkoordinate und  $R$  als Abstand zwischen den Zentrenmittelpunkten.

(b) Vier Zentren auf einer Linie, wobei zwei Zentren zu einer Elektronenkoordinate einen Abstand größer  $D$  haben (Gebiet 1), das andere Paar einen kleineren (Gebiet 2).

**Abbildung 5.2.:** Illustrationen zu Vier-Zentren-Integral, reproduziert aus [49].

Da gilt,

$$I_{\mu\nu\sigma\lambda} \leq |I_{\mu\nu\sigma\lambda}| \leq \iint |\xi_\mu^*(\mathbf{r}_1)\xi_\nu(\mathbf{r}_1)| \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} |\xi_\sigma^*(\mathbf{r}_2)\xi_\lambda(\mathbf{r}_2)| d\mathbf{r}_1 d\mathbf{r}_2,$$

folgt mit Lemma 6,

$$\leq C_{\mu,\nu} \exp\left(-\frac{a_{\mu,\nu}}{2} \|\mathbf{r}_\mu - \mathbf{r}_\nu\|\right) C_{\sigma\lambda} \exp\left(-\frac{a_{\sigma\lambda}}{2} \|\mathbf{r}_\sigma - \mathbf{r}_\lambda\|\right) \cdot \dots \int \int \exp(-a_{\mu,\nu} \|\mathbf{r}_1 - \mathbf{r}_{\mu,\nu}\|) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \exp(-a_{\sigma\lambda} \|\mathbf{r}_2 - \mathbf{r}_{\sigma\lambda}\|) d\mathbf{r}_1 d\mathbf{r}_2. \quad (5.4.4)$$

Die längere Rechnung für die folgende Abschätzung des Integralterms nach oben haben wir in den Abschnitt A.3 im Anhang ausgegliedert. Wir geben hier nur das Ergebnis (A.3.20) an,

$$\leq C \exp(-a_1 R_1) \cdot \exp(-a_2 R_2) \frac{1}{R}, \quad (5.4.5)$$

wobei  $C$  alle weiteren Konstanten beinhalten soll und die Koeffizienten  $a_1$  und  $a_2$  sich jeweils aus den Koeffizienten des jeweiligen Basisfunktionspaares bestimmen. Zur Erläuterung der Variablen verweisen wir auf Abbildung 5.2(a).

<sup>1</sup>Der Originalartikel stand uns nicht zur Verfügung, wir gründen uns hier auf die Aussage in [35].

## 5. Lokalität

Der obige Ausdruck (5.4.5) wird offensichtlich dann klein, wenn die Abstände  $R_1$  bzw.  $R_2$  groß werden.

Dann betrachten wir vereinfachend den Fall, wo die vier Aufpunkte alle auf einer Linie liegen, siehe Abbildung 5.2(b). Sei dann eine Länge  $D$  gegeben und wir denken uns diese als Durchmesser des Gebiets, indem der Abstand  $R_2$  hinein passt und der Abstand  $R_1$  aber nicht. Die Aufpunkte von  $\mu$  und  $\nu$  zu  $R_1$  liegen außerhalb des gedachten Gebiets, also  $R_1 > D$ . Wir erhalten die folgende Abschätzung für (5.4.5), wenn  $R > 1$  in (5.4.5) gilt,

$$I_{\alpha\beta\gamma\delta} < C \exp(-c_1 D). \quad (5.4.6)$$

In der Bestimmung der Matrixelemente mit Vier-Zentren-Integralen wollen wir nun alle Terme vernachlässigen, so dass der Fehler  $\Delta I$  dabei kleiner als ein gegebenes  $\epsilon > 0$  ist. Maximal gibt es  $\mathcal{O}(n^4)$  zu vernachlässigende Terme, also ergibt sich

$$\Delta I < n^4 \cdot \exp(-c_M D) = \epsilon, \quad (5.4.7)$$

wobei  $c_M = \min(c_1, c_2)$  und  $D$  noch festzulegen ist. Lösen wir die Abhängigkeit jetzt nach der Länge  $D$  auf, so erhalten wir

$$D = \frac{1}{c_M} \left( 4 \ln(n) + \ln\left(\frac{1}{\epsilon}\right) \right) \propto \ln(n) \quad (5.4.8)$$

Insgesamt müssen wir zur Bestimmung aller Matrixelemente dann über  $\mathcal{O}(n^2)$  mögliche gedachte Gebiete für die beiden Aufpunktpaare summieren und für eine gegebene Genauigkeit  $\epsilon > 0$  verbleiben dann nur  $\mathcal{O}(\ln(n))$  mögliche Paarungen, so dass die Aufpunkte beide innerhalb des Gebiets liegen. Damit verhält sich die benötigte Rechenzeit  $T$  wie

$$T \propto n^2 \cdot (\ln(n))^2. \quad (5.4.9)$$

Im allgemeinen Fall folgt mit der Aufspaltung pro Koordinatenachse  $n = n_x \cdot n_y \cdot n_z$  mit gleicher Überlegung das analoge Ergebnis.

Damit schließen wir die Darstellung des Beweises ab und bemerken, dass in einer lokalen Basis und unter Vernachlässigung kleiner Terme die Komplexität nicht  $\mathcal{O}(n^4)$ , sondern nur  $\mathcal{O}(n^2 \cdot (\ln n)^2)$  beträgt. Wir haben also quadratische Komplexität erhalten.

## 5.5. Lokalisierte räumlichen Orbitale

Zu guter Letzt wollen wir die im Rahmen des Konzepts der Lokalität aufgestellten Definitionen auf die von uns gesuchten Lösungen der HF-Gleichung anwenden, den räumlichen Orbitalen.

Wir weisen an dieser Stelle aber nur auf das Resultat von Ishida und Ohno [105] und anderen hin, dass die Lösungen exponentiell abklingen,

$$r^\alpha \exp(-\beta r), \quad (5.5.1)$$

wobei  $r$  der radiale Abstand zum Zentrum,  $\beta = \sqrt{-2\epsilon_{\text{HOMO}}}$  mit dem größten Eigenwert  $\epsilon_{\text{HOMO}}$  aller besetzten Zustände und  $\alpha = (Z + 1)/\beta - 1$  ist. Ist die Gesamtladung des



Systems  $Z = \sum_{c=1}^M Z_C - N$  ausgeglichen, dann ist  $\alpha = \frac{1}{\beta} - 1$ . Damit sind die Lösungen also in der Menge der lokalen Funktionen nach Definition 26.

Die räumlichen Orbitale, welche als Lösung aus dem Algorithmus in Abschnitt 4.3.2 resultieren, sind im Allgemeinen nicht in unserem Sinne lokalisiert, sondern meist über das Gebiet „verschmiert“. Der Grund liegt in der Invarianz der Grundzustandsenergie bezüglich einer unitären Transformation der Koeffizientenmatrix, wie wir in Abschnitt 4.2.3 über Spaltentransformation erläutert haben.

Es gibt mehrere Verfahren [60, 50, 143], um eine solche unitäre Transformation zu finden, welche die Lösungen der HF-Gleichung *maximal lokalisiert*. Eines dieser Verfahren haben wir beispielsweise in [90] aufgegriffen und dort einen parallelen Algorithmus dafür implementiert. Allgemeines Ziel ist stets die gleichzeitige Minimierung der Varianz  $\sigma^2$  aller räumlichen Orbitale,

$$\sigma_k^2 = \int \phi_k^*(\mathbf{r}) \mathbf{r}^2 \phi_k(\mathbf{r}) d\mathbf{r} - \left( \int \phi_k(\mathbf{r}) \mathbf{r} \phi_k(\mathbf{r}) d\mathbf{r} \right)^2, \quad (5.5.2)$$

wobei  $\phi_k(\mathbf{r})$  die Lösungen bezeichnet.

## Zusammenfassung

In diesem Kapitel definierten wir zunächst, was wir unter lokalen Funktionen verstehen wollen. Anschließend zeigten wir, dass Produkte lokaler Funktionen wieder eine lokale Funktion ergeben. Damit konnten wir zeigen, dass im Rahmen der Maschinengenauigkeit sich nur quadratisch anwachsende Kosten für die Berechnung der Vier-Zentren-Integrale ergeben. Abschließend zitierten wir, dass die Lösungen der HF-Gleichung asymptotisch lokale Funktionen sind.

Lokalität spielt eine zentrale Rolle bei linear skalierenden Elektronstrukturmethoden, siehe Goedecker [67, Abschn. II.] für einige empirische Argumente. Dies wird nach Kohn [121] auch mit „Kurzsichtigkeit“ (der Wechselwirkung) bezeichnet. Diese zentrale Rolle spielt sie insbesondere auch bei den Fragmentansätzen, die wir in dieser Arbeit betrachten. Denn nur dadurch sind die Elektronenfreiheitsgrade zu einem gewissen Grade voneinander entkoppelt und ermöglichen es, das hochdimensionale Problem in Probleme kleinerer Dimensionalität zu zerlegen.



## 6. Negativer Eigenraum dünnbesetzter, indefiniter Matrizen

Im letzten Kapitel 5 haben wir uns mit der Lokalität der Lösungen der Hartree-Fock-Gleichungen beschäftigt und dabei bemerkt, dass deren Darstellung in einer entsprechenden Basis dünnbesetzt ist. Dies ist ein Baustein auf dem Weg zu einer linear skalierenden Methode. Ein weiterer ist der Umgang mit dünnbesetzten, indefiniten Matrizen, womit wir uns in diesem Kapitel auseinander setzen werden.

Dazu werden wir uns in diesem Kapitel allgemein mit dem Eigenwertproblem dünnbesetzter, hermitescher Matrizen beschäftigen.

### 6.1. Voraussetzungen

In diesem Kapitel betrachten wir eine hermitesche, indefinite  $n \times n$ -Matrix  $\mathbf{H}$  mit dünnbesetzten Eigenvektoren. Dabei nehmen wir der Einfachheit halber an, dass die Matrix vollen Rang habe und die Eigenwerte nicht entartet seien.

Wir interessieren uns für die Summe der negativen Eigenwerte dieser Matrix. Dazu müssen wir uns im weiteren Verlauf mit Hierarchien von Unterindexmengen und zugehörigen Hauptuntermatrizen auseinander setzen und verstehen wie dünnbesetzte Eigenvektoren damit zusammen hängen.

Zu diesem Zweck führen wir zunächst allgemein diese Konzepte der Dünnbesetztheit, der Unterindexmengen und der Hierarchien derselben, sowie der Hauptuntermatrizen ein.

Wir merken an, dass wir in diesem Kapitel keine entarteten Eigenwerte betrachten werden. So können wir konzeptionell im Bild der Eigenvektoren bleiben. Im Falle entarteter Eigenwerte müssen zugehörige Unterräume betrachtet werden, denn die zugehörigen Eigenvektoren zu diesen entarteten Eigenwerten sind nicht mehr eindeutig. Wir vermuten aber, dass sich die den Unterraum aufspannenden Vektoren für das untersuchte Problem stets so unitär transformieren lassen, dass eine dünnbesetzte Darstellung dieser resultiert. Dann kann das im Folgenden Gezeigte auf entartete Eigenwerte übertragen werden.

#### 6.1.1. Anteile der Eigenzerlegung

Da  $\mathbf{H}$  insbesondere normal ist, existiert nach dem Spektralzerlegungstheorem, siehe Horn und Johnson [98, S. 104].

$$\mathbf{H} = \mathbf{U} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{U}^\dagger \quad (6.1.1)$$

## 6. Dünnbesetzte, indefinite Matrizen

mit einer unitären Matrix  $\mathbf{U}$  aus den Eigenvektoren  $\mathbf{u}_i$  von  $\mathbf{H}$  als Spaltenvektoren und einer Diagonalmatrix aus den Eigenwerten  $\lambda_i$  von  $\mathbf{H}$ . Da  $\mathbf{H}$  hermitesch ist, sind die Eigenwerte reellwertig. Es gilt,

$$\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbb{I}. \quad (6.1.2)$$

Mit der Zerlegung (6.1.1) erhalten wir eine Darstellung der Matrix  $\mathbf{H}$  als eine Summe von Rang-1-Matrizen,

$$\mathbf{H}_{jk} = \sum_i \lambda_i \mathbf{U}_{ji} \mathbf{U}_{ki}^\dagger = \sum_i \lambda_i (\mathbf{u}_i)_j (\mathbf{u}_i^\dagger)_k. \quad (6.1.3)$$

Wir bemerken, dass die Rang-1-Matrix  $\mathbf{u}_i \mathbf{u}_i^\dagger$ , welche aus den Eigenvektoren  $\mathbf{u}_i$  der Matrix  $\mathbf{H}$  gebildet wird, stets positiv semidefinit ist. Wir nennen dabei Eigenvektoren  $\mathbf{u}_i$  zu negativen Eigenwerte  $\lambda_i$  auch *negative Eigenvektoren*. Bei ihnen wird die Rang-1-Matrix erst durch den Eigenwert als Vorfaktor negativ semidefinit.

Mit dieser auch als *Eigenzerlegung* bekannten Darstellung können wir die indefinite Matrix  $\mathbf{H}$  in einen negativ semidefiniten Anteil  $\mathbf{N}$  und einen positiv semidefiniten Anteil  $\mathbf{P}$  aufspalten,

$$\mathbf{N} = \sum_{\substack{i \\ \lambda_i < 0}} \lambda_i \mathbf{u}_i \mathbf{u}_i^\dagger \quad (6.1.4a)$$

$$\mathbf{P} = \sum_{\substack{i \\ \lambda_i > 0}} \lambda_i \mathbf{u}_i \mathbf{u}_i^\dagger. \quad (6.1.4b)$$

*Bemerkung 5.* Hat  $\mathbf{H}$  vollen Rang, so wird der Kern von  $\mathbf{N}$  durch das Bild von  $\mathbf{P}$  aufgespannt und umgekehrt.

### 6.1.2. Dünnbesetztheit

Während die Indefinitheit einer Matrix klar definiert ist, nämlich wenn sie weder positiv noch negativ semidefinit ist, so lässt sich die Dünnbesetztheit einer Matrix nicht so strikt fassen. Im Speziellen kann man sich eine Bandmatrix als eine prototypische dünnbesetzte Matrix vorstellen. Allgemeiner betrachtet man den Adjazenzgraphen zu einer Matrix, siehe Benzi et al. [18, S. 16].

**Definition 28** (Dünnbesetztheit). Wir betrachten die Spalten- und Zeilenindexmenge einer Matrix  $\mathbf{H}$  als Knoten des Adjazenzgraphen, bei dem eine Kante von Zeilenknoten  $i$  zum Spaltenknoten  $j$  führt, wenn  $|\mathbf{H}_{ij}|$  ungleich null ist. Eine Matrix  $\mathbf{H}$  heißt dann *dünnbesetzt*, wenn der maximale Bindungsgrad des Graphen durch eine Konstante, also unabhängig von  $n$ , beschränkt ist.

Ist die Matrix quadratisch, dann sind Zeilen- und Spaltenknotenmenge identisch. Ist die Matrix symmetrisch, so sind die Kanten des Graphen ungerichtet.

Weiter wollen wir die Dünnbesetztheit im Sinne der Maschinengenauigkeit verstehen, nämlich wie folgt:

**Definition 29** (Näherungsweise Dünnbesetztheit). Sei ein  $\epsilon > 0$  gegeben. Wir betrachten eine Matrix  $\mathbf{H}$  und deren Adjazenzgraphen. Eine Kante zwischen Knoten  $i$  und  $j$  in diesem Graphen bestehe nur dann, wenn  $|\mathbf{H}_{ij}| > \epsilon$  gilt. Ist der maximale Bindungsgrad wieder durch eine Konstante beschränkt, so heißt die Matrix *näherungsweise dünnbesetzt*.

Wir wollen im weiteren nicht gesondert zwischen diesen beiden Arten der Dünnbesetztheit unterscheiden, da letztere nur eine kleine Störung durch die Trunkierung der hermiteschen Matrix darstellt. Wie wir in Abschnitt 6.3 sehen werden, beeinflusst dies das Spektrum der Matrix nur geringfügig. Wählt man  $\epsilon \approx \epsilon^{\text{masch}}/n^2$ , so bleibt die Störung im Rahmen der Maschinengenauigkeit.

**Dünnbesetzter Eigenvektoren** In Gilbert [65] und in Mascarenhas [131] wird die Frage untersucht, inwieweit dünnbesetzte Matrizen und die Nicht-Null-Struktur ihrer Eigenvektoren zusammenhängen. Im Allgemeinen schließt man dabei von der Struktur der Matrix  $\mathbf{H}$ , also dem Adjazenzgraphen, auf die Struktur der Eigenvektormatrix  $\mathbf{U}$  über die möglichen Wege im Graphen. Diese sogenannte transitive Hülle ist dann die obere Schranke für die Struktur der Eigenvektormatrix. Ein allgemeiner Zusammenhang, dass dünnbesetzte Matrizen auch dünnbesetzte Eigenvektoren haben müssen, lässt sich daraus aber nicht folgern und gibt es auch nicht.

Wir werden uns im Folgenden deswegen auf die Annahme stützen, dass die Eigenvektoren der Matrix  $\mathbf{H}$  dünnbesetzt sind.

Sei  $\mathcal{N} := \{1, \dots, n\}$  die Menge aller möglichen Spalten- bzw. Zeilenindizes von  $\mathbf{H}$ . Die Menge hat dabei die Mächtigkeit  $|\mathcal{N}| = n$ .

Die Definitionen 28 und 29 sind so allgemein gewählt, dass sie auch sofort auf Vektoren anwendbar sind.

**Korollar 1** (Indexmenge dünnbesetzter Vektoren). *Sei ein dünnbesetzter Vektor  $\mathbf{u} \in \mathbb{C}^n$  gegeben, dann existiert eine Indexmenge  $\mathcal{I} \subset \mathcal{N}$  mit  $|\mathcal{I}| = \mathcal{O}(1)$ , so dass*

$$|\mathbf{u}_j| < \epsilon, \quad j \notin \mathcal{I} \quad (6.1.5)$$

*gilt.*

Dies ist schlicht die Menge der Nachbarn, welche mit dem einzigen Zeilenindex verknüpft sind, wenn  $\mathbf{u}$  ein Spaltenvektor ist.

**Orthogonalität dünnbesetzter Vektoren** Nun wollen wir uns den Beziehungen der dünnbesetzten Eigenvektoren untereinander widmen.

**Lemma 8** (Überlapp dünnbesetzter Vektoren). *Seien zwei dünnbesetzte Vektoren  $\mathbf{u}_i$  und  $\mathbf{u}_j$  mit Unterindexmengen  $\mathcal{I}_i$  und  $\mathcal{I}_j$  nach Korollar 1 gegeben. Dann muss zur Berechnung des Überlapps nur die Schnittmenge  $\mathcal{I}_k := \mathcal{I}_i \cap \mathcal{I}_j$  betrachtet werden, also*

$$S_{ij} = \sum_{l \in \mathcal{I}_k} (\mathbf{u}_i^\dagger)_l (\mathbf{u}_j)_l \in \mathbb{R}. \quad (6.1.6)$$

*Daraus folgt:*

## 6. Dünnbesetzte, indefinite Matrizen

1. Ist der Schnitt leer, sind die Vektoren orthogonal.
2. Sind die beiden Vektoren orthogonal,  $S_{ij} = 0$ , dann muss der Schnitt entweder leer sein oder mindestens zwei Indizes enthalten.

*Beweis.* Die erste Eigenschaft folgt sofort aus (6.1.6). Da die beiden Indexmengen nur die Komponenten ungleich null angeben, folgt die Eigenschaft 2, denn das Produkt zweier komplexer, von Null verschiedener Zahlen kann nicht null sein.  $\square$

### 6.1.3. Hierarchie von Unterindexmengen

Die Nicht-Null-Struktur von dünnbesetzten Vektoren lässt sich über die zugehörige Indexmenge beschreiben. Deswegen betrachten wir die Relationen zwischen diesen Indexmengen, die wir auch schon in Abschnitt 2.5 kennengelernt haben.

**Definition 30** (Hierarchische Zerlegung). Sei  $\mathcal{J} := \{\mathcal{I}_i\}$  eine (überlappende) Zerlegung von  $\mathcal{N}$ . Dann heißt diese Zerlegung *hierarchisch*, wenn diese abgeschlossen unter Schnittmengenbildung ist, d. h. für alle Paare  $\forall \mathcal{I}_i, \mathcal{I}_j \in \mathcal{J}$  gilt, dass die Schnittmenge  $\mathcal{I}_k := \mathcal{I}_i \cap \mathcal{I}_j$  ebenfalls in  $\mathcal{J}$  existiert,  $\mathcal{I}_k \in \mathcal{J}$ .

Im Rahmen von Hierarchischen oder kurz *H-Matrizen*, siehe bspw. Grasedyck [71, Abschn. 3.1], werden analoge Konzepte eingeführt. Dort spricht man von Hierarchischer Partitionierung. Es geht dabei jedoch nicht um Eigenwertprobleme, sondern um diskretisierte partielle Differentialgleichung. Da wir an dieser Stelle nicht den Anspruch verfolgen, die dort aufgeführten Konzepte auf die hier behandelten Eigenwertprobleme zu übertragen, verwenden wir anstelle von Partitionierung der Begriff Zerlegung. Damit möchten wir außerdem die rein surjektive Natur der Zuordnung betonen.

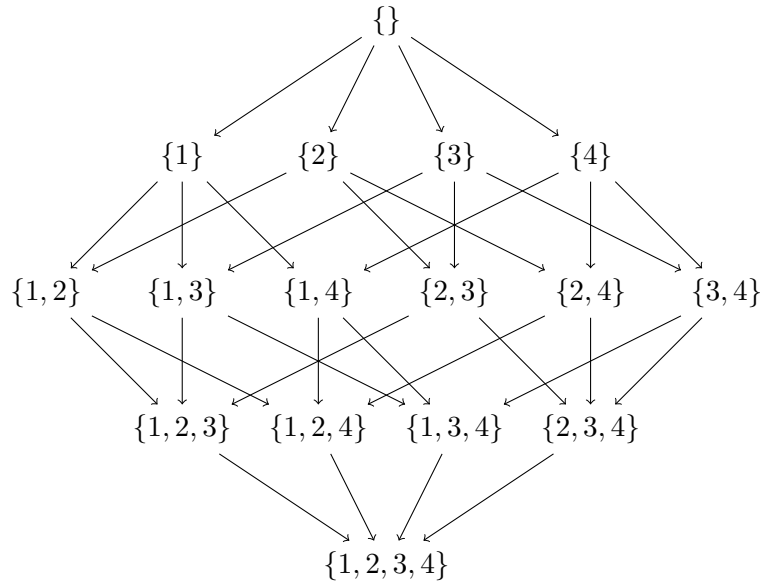
Die Hierarchie dieser Unterindexmengen von  $\mathcal{N}$  lässt sich dabei als ein Baumgraph darstellen, siehe Abbildung 6.1. Dabei stehen Unterindexmengen der gleichen Mächtigkeit auf der gleichen Ebene. Beziehungen zwischen den Indexmengen entstehen aufgrund ihrer Teilmengeneigenschaft. Diese sind im Graph als Wege gekennzeichnet, wobei gerichtete Kanten diese Eigenschaft zwischen den Ebenen angeben.

Gleichwohl in Abbildung 6.1 eine vollständige Zerlegung in alle möglichen Teilmengen angegeben ist, ist diese natürlich nicht die einzig mögliche unter den hierarchischen Zerlegungen von  $\mathcal{N}$ .

**Projektionsmatrizen** Der Zusammenhang zwischen den obigen Indexuntermengen und der Matrix  $\mathbf{H}$  und ihren Eigenvektoren ergibt sich über die folgenden kartesischen Projektionsmatrizen.

**Definition 31** (Kartesische Projektionsmatrix). Mit  $\pi_i = [\{\mathbf{e}_j\}_{j \in \mathcal{I}_i}] \in \mathbb{C}^{n \times |\mathcal{I}_i|}$  sei eine *kartesische Projektionsmatrix* zu einer Unterindexmenge  $\mathcal{I}_i$  bezeichnet.

Dies ist eine Matrix aus Spaltenvektoren, wobei  $\mathbf{e}_j$  den  $j$ -ten kartesischen Einheitsvektor bezeichnet.



**Abbildung 6.1.:** Beispiel einer vollständigen Unterindexmengenhierarchie für  $\mathcal{N} = \{1, 2, 3, 4\}$ . Die Existenz eines Wegs von einem Knoten  $i$  nach  $j$  im Graph entspricht dabei der Eigenschaft “ist Teilmenge von” für die zugehörigen Indexmengen.

Die Projektionsmatrix  $\pi_i^\dagger$  zu einer Menge  $\mathcal{I}_i$  projiziert einen Vektor  $\mathbf{u} \in \mathbb{C}^n$  in einen Unterraum  $\mathbb{C}^{|\mathcal{I}_i|}$  durch Weglassen aller Komponenten  $\mathbf{u}_i$  mit  $i \notin \mathcal{I}_i$ . Andersherum bettet die Projektionsmatrix  $\pi_i$  einen Vektor  $\mathbf{u} \in \mathbb{C}^{|\mathcal{I}_i|}$  in den  $\mathbb{C}^n$  ein.

Wir machen kurz deutlich, dass für diese Projektionsmatrizen eine analoge Hierarchie existiert wie für die Indexmengen.

**Korollar 2** (Hierarchie der kartesischen Projektionsmatrizen). *Seien  $\mathcal{I}_1$  und  $\mathcal{I}_2$  zwei Indexmengen mit  $\mathcal{I}_3 := \mathcal{I}_1 \cap \mathcal{I}_2$ , seien  $\pi_1, \pi_2$  und  $\pi_3$  die zugehörigen kartesischen Projektionsmatrizen. Dann folgt für einen beliebigen Vektor  $\mathbf{u} \in \mathbb{C}^n$  allgemein,*

$$\mathcal{I}_1 \cap \mathcal{I}_2 = \mathcal{I}_3 \quad \Rightarrow \quad \pi_1 \pi_2^\dagger \mathbf{u} = \pi_3 \pi_3^\dagger \mathbf{u}, \quad (6.1.7a)$$

und speziell,

$$\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset \quad \Rightarrow \quad \pi_1 \pi_2^\dagger \mathbf{u} = 0, \quad (6.1.7b)$$

$$\mathcal{I}_1 \subset \mathcal{I}_2 \quad \Rightarrow \quad \pi_1 \pi_2^\dagger \mathbf{u} = \pi_1 \pi_1^\dagger \mathbf{u}. \quad (6.1.7c)$$

*Beweis.* Diese Eigenschaften ergeben sich sofort aus der Definition 31 unter der Betrachtung, wo  $\pi_1$  und  $\pi_2$  jeweils eine eins stehen haben oder nicht.  $\square$

Mit Definition 31 erhalten wir aber noch keine orthogonale Projektion, siehe [190, S. 95], denn  $\pi$  hat unterschiedliche Ursprungs- und Zielräume.

## 6. Dünnbesetzte, indefinite Matrizen

Deswegen betrachten wir statt  $\pi_j$  die wie folgt definierte orthogonale Projektion

$$\mathbf{P}_j := \pi_j \pi_j^\dagger \in \mathbb{C}^{n \times n} \quad (6.1.8)$$

mit  $\text{rang}(\mathbf{P}_j) = |\mathcal{I}_j| \leq n_j$ . Dies projiziert nun vom  $\mathbb{C}^n$  nach  $\mathbb{C}^n$ , also in den gleichen Raum.

### 6.1.4. Hauptuntermatrizen

Die Bedeutung dünnbesetzter Eigenvektoren liegt in einem bekannten Zusammenhang zu Hauptuntermatrizen, auf den wir nun eingehen werden.

Zunächst definieren wir aber diese Hauptuntermatrizen mit den obigen Unterindexmengen von  $\mathcal{N}$ .

**Definition 32** (Hauptuntermatrix). Sei  $\mathcal{I}_i \subseteq \mathcal{N}$  eine Unterindexmenge von  $\mathcal{N}$ .

Dann ist die *Hauptuntermatrix*  $\mathbf{h}_i \in \mathbb{C}^{|\mathcal{I}_i| \times |\mathcal{I}_i|}$  diejenige Matrix, welche durch Streichung all jener Zeilen und Spalten von  $\mathbf{H}$  entsteht, deren Index nicht in der Menge  $\mathcal{I}_i$  enthalten ist.

*Bemerkung 6.* Wir verdeutlichen die Verknüpfung zwischen Unterindexmenge  $\mathcal{I}_i \subseteq \mathcal{N}$  und Hauptuntermatrix  $\mathbf{h}_i$  stets, indem wir beiden denselben Index geben.

Zur Vereinfachung betrachten wir im Weiteren nur noch die Hauptuntermatrizen  $\mathbf{H}_i$ , welche wir durch Rückeinbettung in den  $\mathbb{C}^{n \times n}$  erhalten. Beide Matrizen  $\mathbf{h}_i \in \mathbb{C}^{|\mathcal{I}_i| \times |\mathcal{I}_i|}$  bzw.  $\mathbf{H}_i \in \mathbb{C}^{n \times n}$  von  $\mathbf{H}$  lassen sich leicht mittels der Projektionen schreiben,

$$\mathbf{h}_i = \pi_i^\dagger \mathbf{H} \pi_i, \quad (6.1.9a)$$

$$\mathbf{H}_i = \mathbf{P}_i^\dagger \mathbf{H} \mathbf{P}_i = \mathbf{P}_i \mathbf{H} \mathbf{P}_i. \quad (6.1.9b)$$

Weiterhin halten wir einige Eigenschaften zu Hauptuntermatrizen fest.

**Lemma 9** (Eigenschaften, Hauptuntermatrizen hermitescher Matrizen). *Sei  $\mathbf{H}$  eine hermitesche  $n \times n$ -Matrix und sei  $\mathcal{I}_i \subseteq \mathcal{N}$  eine Unterindexmenge. Dann gilt:*

1.  $\mathbf{h}_i$  bzw.  $\mathbf{H}_i$  sind hermitesch.
2. Ist  $\mathbf{H}$  positiv bzw. negativ semidefinit, dann ist auch  $\mathbf{h}_i$  bzw.  $\mathbf{H}_i$  positiv bzw. negativ semidefinit.

*Beweis.* Eigenschaft 1 folgt sofort aus der Definition der Hermitizität, da wir eben nur Hauptuntermatrizen betrachten.

Mittels der Eigenzerlegung (6.1.3) können wir die Hauptuntermatrix  $\mathbf{H}_i$  auch wie folgt schreiben,

$$\mathbf{H}_i = \sum_j \lambda_j (\mathbf{P}_i \mathbf{u}_j) (\mathbf{P}_i \mathbf{u}_j)^\dagger. \quad (6.1.10)$$

Damit und mit der Tatsache, dass eine Summe positiv bzw. negativ semidefiniter Matrizen ebenso positiv bzw. negativ semidefinit ist, folgt die Eigenschaft 2, welche für positiv definite Matrizen als Sylvester-Kriterium bekannt ist. Für  $\mathbf{h}_i$  folgt die Aussage analog.  $\square$



Nun zeigen wir den bekannten Zusammenhang, der besagt, dass ein dünnbesetzter Eigenvektor einer Matrix auch gleichzeitig Eigenvektor einer ihrer Hauptuntermatrizen sein muss, siehe Johnson und Kroschel [109, Einleitung].

**Lemma 10** (Dünnbesetzte Eigenvektoren und Hauptuntermatrizen). *Sei  $\mathbf{u}$  ein dünnbesetzter Eigenvektor von  $\mathbf{H}$  mit Eigenwert  $\lambda$  und sei  $\mathcal{I}_i$  die zugehörige Indexmenge, siehe Korollar 1.*

*Dann ist der Vektor  $(\pi_i^\dagger \mathbf{u})$  Eigenvektor zur Hauptuntermatrix  $\mathbf{h}_i$  mit gleichem Eigenwert  $\lambda$  bzw. analog  $\mathbf{P}_i \mathbf{u} = \mathbf{u}$  ist Eigenvektor von  $\mathbf{H}_i$ .*

*Beweis.* Mittels  $\mathcal{N} \setminus \mathcal{I}_i =: \mathcal{I}_\setminus$  können wir die Eigenwertgleichung wie folgt partitionieren,

$$\begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}, \quad (6.1.11)$$

wobei  $\mathbf{H}_{11} = \mathbf{h}_i$ ,  $\mathbf{H}_{22} = \mathbf{H}_\setminus$ ,  $\mathbf{H}_{21} = \mathbf{H}_{12}^\dagger$  und  $\mathbf{u}_2 = \mathbf{0}$  ist. Damit folgt sofort, dass  $\mathbf{u}_1 = (\pi_i^\dagger \mathbf{u})$  einerseits Eigenvektor von  $\mathbf{h}_i$  ist und andererseits im Kern von  $\mathbf{H}_{21}$  liegen muss,

$$\mathbf{H}_{11} \mathbf{u}_1 + \underbrace{\mathbf{H}_{12} \mathbf{u}_2}_{=0} = \lambda \mathbf{u}_1 \quad \Rightarrow \quad \mathbf{H}_{11} \mathbf{u}_1 = \lambda \mathbf{u}_1, \quad (6.1.12a)$$

$$\mathbf{H}_{21} \mathbf{u}_1 + \underbrace{\mathbf{H}_{22} \mathbf{u}_2}_{=0} = \lambda \underbrace{\mathbf{u}_2}_{=0} \quad \Rightarrow \quad \mathbf{H}_{21} \mathbf{u}_1 = 0. \quad (6.1.12b)$$

Damit gilt das Gewünschte, siehe auch Johnson und Kroschel [109], wo die Frage untersucht wird, inwieweit Hauptuntermatrizen, die gleiche Eigenwerte mit der vollen Matrix haben, auch gleiche Eigenvektoren besitzen. Im Allgemeinen stimmt dies aber nicht.  $\square$

Wir merken an, dass jede Hauptuntermatrix  $\mathbf{H}_i$  einmal mit ihrer eigenen Eigenzerlegung (6.1.3), sowie durch die projizierte Darstellung (6.1.10) auch mit der vollen Matrix  $\mathbf{H}$  dargestellt werden kann. Diese beiden Darstellungen stellen jeweils Summen von positiv und negativ semidefiniten Matrizen dar, zwischen denen jeweils aber kein direkter Zusammenhang besteht.

Diese Hauptuntermatrix  $\mathbf{H}_i$  zu einem Eigenvektor  $\mathbf{u}$  von  $\mathbf{H}$  ist aber nicht eindeutig durch diese Eigenschaft des gleichen Eigenvektors gegeben. Es gibt weitere Hauptuntermatrizen  $\mathbf{H}_j$ , von denen  $\mathbf{u}$  Eigenvektor ist, nämlich all jene, deren assoziierte Indexmenge  $\mathcal{I}_j$  die Indexmenge  $\mathcal{I}_i$  vollständig enthält.

**Lemma 11** (Kleinste Hauptuntermatrix). *Sei  $\mathcal{I}_i$  die Indexmenge zu  $\mathbf{u}$ , siehe Korollar 1. Sei  $(\mathbf{u}, \lambda)$  dabei ein Eigenpaar von  $\mathbf{H}$ . Weiterhin existiere eine nicht-leere Menge*

$$\mathcal{J} = \left\{ \mathcal{I}_j \mid \mathbf{H}_j(\mathbf{P}_i \mathbf{u}) = \lambda(\mathbf{P}_i \mathbf{u}) \right\}. \quad (6.1.13)$$

*Dann ist  $\mathcal{I}_i \in \mathcal{J}$  die Menge kleinster Mächtigkeit, welche die gemeinsame Schnittmenge über alle Mengen  $\mathcal{I}_j$  ist, also  $\forall \mathcal{I}_j \in \mathcal{J} : \mathcal{I}_i \subseteq \mathcal{I}_j$  gilt.*

## 6. Dünnbesetzte, indefinite Matrizen

*Beweis.* Zunächst muss  $\mathcal{I}_i$  offensichtlich in  $\mathcal{J}$  sein, siehe Lemma 10. Damit ist  $\mathcal{J}$  nicht-leer und enthält sie nur eine Menge, gilt die Aussage trivialerweise.

Wegen  $\mathbf{H}_j(\mathbf{P}_i \mathbf{u}) = \mathbf{P}_j \mathbf{H} \mathbf{P}_j(\mathbf{P}_i \mathbf{u})$  muss gelten, dass  $\mathcal{I}_i \subseteq \mathcal{I}_j$  ist, also dass es gemeinsame Schnittmenge ist. Denn andernfalls erhielten wir nicht  $\mathbf{u}$ , sondern den trunkierten Vektor  $\mathbf{P}_j(\mathbf{P}_i \mathbf{u}) = (\mathbf{P}_k \mathbf{u})$  mit  $\mathcal{I}_k = \mathcal{I}_i \cap \mathcal{I}_j \neq \mathcal{I}_i$  als Eigenvektor zu  $\mathbf{H}_j$ .

Damit muss  $\mathcal{I}_i$  auch die Menge kleinster Mächtigkeit sein, denn jede andere verschiedene Menge  $\mathcal{I}_j$  muss sie als Teilmenge enthalten und mindestens einen unterschiedlichen Index enthalten und damit größer sein.  $\square$

Wir werden die Hauptuntermatrix  $\mathbf{H}_i$  zu dieser eindeutigen Schnittmenge  $\mathcal{I}_i$  im Weiteren als *kleinste Hauptuntermatrix* zu einem Eigenvektor  $\mathbf{u}$  von  $\mathbf{H}$  bezeichnen. Die Schnittmenge  $\mathcal{I}_i$  selbst heißt *kleinste Indexmenge* zu  $\mathbf{u}$ .

Mit diesen Mengen kleinster Mächtigkeit definieren wir nun eine besondere Zerlegung von  $\mathcal{N}$ .

**Definition 33** (Minimalüberlappende Zerlegung). Sei eine Matrix  $\mathbf{H}$  mit dünnbesetzten Eigenvektoren  $\mathbf{u}_i$  gegeben. Sei  $\mathcal{I}$  die Menge der assoziierten kleinsten Indexmengen  $\mathcal{I}_i$  zu diesen Eigenvektoren. Sei  $\mathcal{S} = \{\mathcal{I}_k \mid \mathcal{I}_k = \mathcal{I}_i \cap \mathcal{I}_j \mid \mathcal{I}_i, \mathcal{I}_j \in \mathcal{I}\}$  die Menge der paarweisen Schnitte. Für alle  $\mathcal{I}_i, \mathcal{I}_j \in \mathcal{S}$  gelte, dass sie paarweise disjunkt sind,  $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$ .

Dann heißt die Menge  $\mathcal{J} = \mathcal{S} \cup \mathcal{I}$  die *minimal überlappende Zerlegung* von  $\mathcal{N}$ .

Wir bemerken, dass eine minimal überlappende Zerlegung hierarchisch ist und außerdem dass diese Eigenschaft vor allem eine Einschränkung der Struktur der Eigenvektoren darstellt, da sich ihre Schnittmengen nicht mehr überlappen dürfen. Dies erfüllt beispielsweise eine Eigenvektormatrix  $\mathbf{U}$  aus sich überlappenden Blöcken auf der Diagonale.

## 6.2. Schwerpunkttheorem

Wir zeigen nun zunächst, dass wir die Matrix  $\mathbf{N}$  aus den negativen Eigenvektoren von Hauptuntermatrizen  $\mathbf{H}_i$  im Falle einer minimal überlappenden Zerlegung  $\mathcal{J}$  rekonstruieren können.

**Lemma 12** (Hauptuntermatrizenendarstellung). Sei  $\mathbf{H} \in \mathbb{C}^{n \times n}$  eine hermitesche Matrix mit dünnbesetzten Eigenvektoren mit einer minimal überlappenden Zerlegung. Sei  $\mathbf{H} = \mathbf{N} + \mathbf{P}$  wie in (6.1.4a), (6.1.4b) zerlegt. Sei weiterhin eine hierarchische Zerlegung  $\mathcal{J} = \{\mathcal{I}_i\}$  von  $\mathcal{N}$  gegeben, welche diese minimal überlappende Zerlegung enthält.

Dann gilt, dass

$$\tilde{\mathbf{N}} := \sum_{\mathcal{I}_i \in \mathcal{J}} \tilde{\mathbf{N}}_i \quad (6.2.1)$$

äquivalent zu  $\mathbf{N}$  ist, mit

$$\tilde{\mathbf{N}}_i := \mathbf{N}_i - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i} \tilde{\mathbf{N}}_j, \quad (6.2.2)$$

wobei  $\mathbf{H}_i = \mathbf{N}_i + \mathbf{P}_i$  in gleicher Weise wie in (6.1.4a), (6.1.4b) zerlegt wird.

Lemma 12 besagt also, dass die beiden Matrizen  $\tilde{\mathbf{N}}$  und  $\mathbf{N}$  die gleiche Spektralzerlegung und damit das gleiche Eigensystem haben sollen.

Wir betonen noch einmal, dass  $\mathbf{N}_i$  der Anteil der negativen Eigenvektoren von  $\mathbf{H}_i$  ist und *nicht*  $\mathbf{P}_i^\dagger \mathbf{N} \mathbf{P}_i$ . Dies ist der entscheidende Punkt, da wir nicht die Eigenvektoren von  $\mathbf{H}$  berechnen, sondern nur die der Hauptuntermatrizen  $\mathbf{H}_i$ .

Damit können wir nun das zentrale Theorem dieses Kapitels formulieren.

**Theorem 2** (Schwerpunkttheorem). *Gleiche Voraussetzungen wie in Lemma 12.*

Dann gilt,

$$\operatorname{Tr}(\mathbf{N}) = \sum_{\mathcal{I}_i \in \mathcal{J}} \operatorname{Tr}(\tilde{\mathbf{N}}_j) \quad (6.2.3)$$

mit

$$\operatorname{Tr}(\tilde{\mathbf{N}}_j) = \operatorname{Tr}(\mathbf{N}_j) - \sum_{\mathcal{I}_l \subsetneq \mathcal{I}_j} \operatorname{Tr}(\tilde{\mathbf{N}}_l).$$

*Beweis.* Mit Lemma 12 und der Additivität der Spur folgt,

$$\operatorname{Tr}(\mathbf{N}) = \operatorname{Tr}\left(\sum_{\mathcal{I}_i \in \mathcal{J}} \tilde{\mathbf{N}}_j\right) = \sum_{\mathcal{I}_i \in \mathcal{J}} \operatorname{Tr}(\tilde{\mathbf{N}}_j).$$

□

Im weiteren Verlauf dieses Abschnitts verbleibt nun Lemma 12 zu zeigen.

Wir merken zunächst an, dass für eine *vollständige* Zerlegung der Indexmenge  $\mathcal{N}$  zur Matrix  $\mathbf{N}$ , wie bspw. in Abbildung 6.1, die Aussage trivial gelten muss, da sie mit (6.2.2) letztlich nur einen Nulltrick darstellt.

### 6.2.1. Strukturelle Induktion

Die Schwierigkeit des Beweises hängt mit der rekursiven Natur der Formel (6.2.2) zusammen. Da sich die Indexmengen der Zerlegung  $\mathcal{J}$  beliebig überlappen können, müssen alle möglichen Arten von Schnittmengen berücksichtigt werden.

Die dazu nötige Beweistechnik nennt sich *strukturelle Induktion*, siehe Burstall [28], welche eine Verallgemeinerung der vollständigen Induktion auf Strukturen wie Listen oder Bäume darstellt. Dabei entscheidend ist, dass die Anordbarkeit der natürlichen Zahlen durch eine Ordnungsrelation zwischen den Objekten der Struktur ersetzt wird.

In unserem Fall ist die Struktur der Hierarchische Baum der Indexmengen, siehe Abbildung 6.1. Die Ordnungsrelation ist die Teilmengeneigenschaft. Gilt  $\mathcal{I}_j \subsetneq \mathcal{I}_i$ , so existiert ein Weg von  $\mathcal{I}_j$  nach  $\mathcal{I}_i$  im Baum.

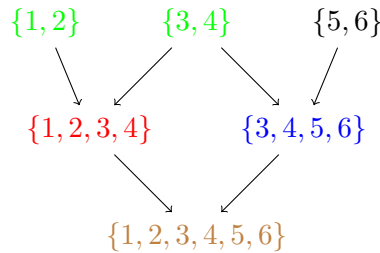
Bei der strukturellen Induktion ist zunächst die Induktionsvoraussetzung für alle Basisfälle zu zeigen. In unserem Fall bedeutet dies für alle Indexmengen, zu denen kein anderer Weg führt außer dem, der bei ihnen selbst beginnt, also bspw. alle Wurzelknoten des Baums. Anschließend wird mittels der Induktionsannahme, dass alle Teilmengen das Gewünschte erfüllen, im Induktionsschritt gefolgert, dass auch die Menge selbst die Aussage erfüllt.

### 6.2.2. Gleiches Eigensystem

Wir wollen nun zeigen, dass  $\tilde{\mathbf{N}}$  und  $\mathbf{N}$  die gleichen Eigenvektoren und -werte besitzen. Dazu betrachten wir jeweils einen beliebigen Eigenvektor von  $\mathbf{H}$  mit Eigenwert  $\lambda$ , also  $\mathbf{H}\mathbf{u} = \lambda\mathbf{u}$ , und untersuchen die Wirkung von  $\tilde{\mathbf{N}}$  auf diesen beliebigen Eigenvektor:  $\tilde{\mathbf{N}}\mathbf{u}$ .

Zunächst wissen wir mit Lemma 11, dass eine kleinste Hauptuntermatrix und damit eine kleinste Indexmenge  $\mathcal{I}_j$  zu  $\mathbf{u}$  existiert. Wir teilen nun die Mengen der Zerlegung  $\mathcal{J}$  je nach der Art auf, wie sich eine einzelne Indexmenge im Schnitt mit der kleinsten Indexmengen  $\mathcal{I}_j$  verhält:

1. Disjunkte Indexmengen:  $\mathcal{J}_\emptyset = \{\mathcal{I}_l \in \mathcal{J} \mid \mathcal{I}_l \cap \mathcal{I}_j = \emptyset\}$
2. Untermengen:  $\mathcal{J}_\subsetneq = \{\mathcal{I}_l \in \mathcal{J} \mid \mathcal{I}_l \subsetneq \mathcal{I}_j\}$
3. Obermengen:  $\mathcal{J}_\supsetneq = \{\mathcal{I}_l \in \mathcal{J} \mid \mathcal{I}_l \supsetneq \mathcal{I}_j\}$
4. Kleinste Indexmenge:  $\mathcal{J}_= = \{\mathcal{I}_j\}$
5. Schnittmengen:  $\mathcal{J}_\cap = \mathcal{J} - \mathcal{J}_\emptyset - \mathcal{J}_\subsetneq - \mathcal{J}_\supsetneq - \mathcal{J}_=$



**Abbildung 6.2.:** Beispiel einer Unterindexmengenhierarchie für  $\mathcal{N} = \{1, \dots, 6\}$ . Dabei sind die verschiedenen Untermengen wie folgt relativ zur Menge  $\mathcal{I}_j = \{1, 2, 3, 4\}$  markiert: disjunkte Mengen (schwarz), Untermengen (grün), Obermengen (braun), die kleinste Indexmenge selbst (rot) und restliche Schnittmengen (blau).

In Abbildung 6.2 ist beispielhaft eine solche Zerlegung der Menge  $\mathcal{N} = \{1, 2, 3, 4, 5, 6\}$  dargestellt und die verschiedenen Unterteilungen für ein bestimmtes  $\mathcal{I}_j$  sind farblich gekennzeichnet.

Wir bemerken, dass die zuvor definierten Mengen  $\mathcal{J}_\emptyset, \mathcal{J}_\subsetneq, \mathcal{J}_\supsetneq, \mathcal{J}_=, \mathcal{J}_\cap$  von Indexmengen paarweise disjunkt sind und ihre Vereinigung wieder  $\mathcal{J}$  ergibt. Deswegen lässt sich jede beliebige Menge  $\mathcal{I}_k \in \mathcal{J}$  – außer der leeren Menge selbst<sup>1</sup> – eindeutig einer der fünf Mengen zuordnen.

<sup>1</sup>Da die zugehörige Matrix aber die Nullmatrix ist, lassen wir sie in den weiteren Überlegungen außen vor.

Zunächst sind die Untermengen  $\mathcal{J}_\emptyset$  und  $\mathcal{J}_\subseteq$  abgeschlossen unter Schnittmengenbildung.

**Korollar 3** (Abgeschlossenheit unter Schnittmengenbildung). *Für zwei Mengen  $\mathcal{I}_m$  und  $\mathcal{I}_n$  aus  $\mathcal{J}_\emptyset$  bzw.  $\mathcal{J}_\subseteq$  einer hierarchischen Zerlegung  $\mathcal{J}$  ist die Schnittmenge  $\mathcal{I}_k := \mathcal{I}_m \cap \mathcal{I}_n$  aus  $\mathcal{J}_\emptyset$  bzw.  $\mathcal{J}_\subseteq$ .*

*Beweis.* Die Existenz der Schnittmenge folgt aus der Voraussetzung, dass  $\mathcal{J}$  hierarchisch ist. Mit  $\mathcal{I}_m, \mathcal{I}_n \in \mathcal{J}_\emptyset$  muss aber auch  $\mathcal{I}_k \cap \mathcal{I}_i = \emptyset$  gelten, also auch  $\mathcal{I}_k \in \mathcal{J}_\emptyset$ . Für  $\mathcal{I}_m, \mathcal{I}_n \in \mathcal{J}_\subseteq$  muss gleichfalls aber  $\mathcal{I}_k \subseteq \mathcal{I}_j$  gelten und damit  $\mathcal{I}_k \in \mathcal{J}_\subseteq$ .  $\square$

Anhand von Abbildung 6.2 ist leicht einzusehen, dass die verschiedenen Untermengentypen von  $\mathcal{J}$  sich in verschiedenen Bereichen des Baums befinden. Dies können wir benutzen, indem wir die Aussage zunächst für die Typen nahe den Wurzelknoten des Baums beweisen.

Dazu trennen wir die Summe in der Darstellung von  $\mathbf{N}$  in (6.2.1) auf,

$$\sum_{\mathcal{I}_i \in \mathcal{J}} \tilde{\mathbf{N}}_i \mathbf{u} = \sum_{\mathcal{I}_i \in \mathcal{J}_\emptyset} \tilde{\mathbf{N}}_i \mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_\subseteq} \tilde{\mathbf{N}}_i \mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_\supseteq} \tilde{\mathbf{N}}_i \mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_=} \tilde{\mathbf{N}}_i \mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_\cap} \tilde{\mathbf{N}}_i \mathbf{u},$$

und betrachten nun jeden einzelnen Anteil. Wir werden zeigen, dass die Anteile für die Mengen  $\mathcal{J}_\emptyset$ ,  $\mathcal{J}_\supseteq$  und  $\mathcal{J}_\cap$  verschwinden müssen.

Diese Auftrennung können wir auch in (6.2.2) nutzen,

$$\begin{aligned} \tilde{\mathbf{N}}_i &= \mathbf{N}_i - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i} \tilde{\mathbf{N}}_j \\ &= \mathbf{N}_i - \sum_{\substack{\mathcal{I}_j \in \mathcal{J}_\emptyset, \\ \mathcal{I}_j \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_j - \sum_{\substack{\mathcal{I}_j \in \mathcal{J}_\subseteq, \\ \mathcal{I}_j \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_j - \sum_{\substack{\mathcal{I}_j \in \mathcal{J}_\supseteq, \\ \mathcal{I}_j \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_j - \sum_{\substack{\mathcal{I}_j \in \mathcal{J}_=, \\ \mathcal{I}_j \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_j - \sum_{\substack{\mathcal{I}_j \in \mathcal{J}_\cap, \\ \mathcal{I}_j \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_j, \end{aligned} \quad (6.2.4)$$

wobei einige der Summen möglicherweise verschwinden, wenn die Mengen  $\mathcal{I}_j$  nicht beide Bedingungen gleichzeitig erfüllen.

**Korollar 4** (Verschwindende orthogonale Anteile aus  $\mathcal{J}_\emptyset$ ). *Für einen beliebigen Eigenvektor  $\mathbf{u}$  von  $\mathbf{H}$  mit  $\mathbf{H}\mathbf{u} = \lambda\mathbf{u}$  gilt für alle  $\mathcal{I}_i \in \mathcal{J}_\emptyset$ , dass  $\mathbf{N}_i \mathbf{u} = 0$ .*

*Beweis.* Da die Zerlegung hierarchisch ist, siehe Definition 30,  $\mathbf{u}$  dünnbesetzt, siehe Lemma 10, und mit (6.1.7b) folgt, dass für alle  $\mathcal{I}_i \in \mathcal{J}_\emptyset$  gilt,  $\mathbf{P}_i \mathbf{u} = 0$  und damit insbesondere  $\mathbf{N}_i \mathbf{u} = \mathbf{P}_i \mathbf{N}_i \mathbf{P}_i \mathbf{u} = 0$ . Damit folgt das Gewünschte mit (6.2.2) und der Abgeschlossenheit nach Korollar 3, sprich, alle Teilmengen einer Menge  $\mathcal{I}_i$  in  $\mathcal{J}$  sind auch in  $\mathcal{J}_\emptyset$ .  $\square$

Für diesen Beweis haben wir noch keine strukturelle Induktion benötigt, für die folgenden beiden aber brauchen wir sie und benutzen dabei die Darstellung (6.2.4).

**Korollar 5** (Anteil aus Schnittmengen). *Sei  $\mathbf{H}$  dergestalt, dass die Zerlegung  $\mathcal{J}$  minimal überlappend ist. Sei weiterhin  $\mathbf{u}$  ein dünnbesetzter Eigenvektor von  $\mathbf{H}$  mit zugehöriger Indexmenge  $\mathcal{I}_j \in \mathcal{J}$ .*

*Dann gilt für jede Indexmenge  $\mathcal{I}_i \in \mathcal{J}$  und mit  $\mathcal{I}_{i,\subseteq} = \mathcal{I}_i \cap \mathcal{I}_j$ ,*

$$\mathbf{N}_i \mathbf{u} = \mathbf{N}_{i,\subseteq} \mathbf{u}. \quad (6.2.5)$$

## 6. Dünnbesetzte, indefinite Matrizen

*Beweis.* Wir benutzen zunächst die Spektralzerlegung (6.1.10) von  $\mathbf{H}_i$ ,

$$\mathbf{H}_i \mathbf{u} = \sum_k \lambda_k (\mathbf{P}_i \mathbf{u}_k) (\mathbf{P}_i \mathbf{u}_k)^\dagger \mathbf{u},$$

und die Zuordnung  $I : \{1, \dots, n\} \rightarrow \{1, \dots, |\mathcal{J}|\}$ , welche zum Index eines Eigenvektors  $\mathbf{u}_k$  die zugehörige kleinste Indexmenge  $I(k) \in \mathcal{J}$  angibt, sowie  $\mathbf{P}_{I(k)} \mathbf{u}_k = \mathbf{u}_k$  und  $\mathbf{P}_j \mathbf{u} = \mathbf{u}$ ,

$$= \sum_k \lambda_k (\mathbf{P}_i \mathbf{P}_{I(k)} \mathbf{u}_k) (\mathbf{P}_i \mathbf{P}_{I(k)} \mathbf{u}_k)^\dagger (\mathbf{P}_j \mathbf{u}).$$

Wir betrachten isoliert das Skalarprodukt

$$(\mathbf{P}_i \mathbf{P}_{I(k)} \mathbf{u}_k)^\dagger \mathbf{P}_j \mathbf{u} = \mathbf{u}_k^\dagger (\mathbf{P}_{I(k)} \mathbf{P}_i \mathbf{P}_j) \mathbf{u}.$$

In einer minimal überlappenden Zerlegung gibt es keinen nicht-leeren Schnitt dreier paarweise verschiedener Indexmengen. Daher gilt,  $I(k) \in \{i, j\}$ . Weiterhin ist aber jeder Eigenvektor  $\mathbf{u}$  zur Indexmenge  $\mathcal{I}_j$  bereits auf dem Schnitt  $\mathcal{I}_i \cap \mathcal{I}_j$  orthogonal zu jedem Eigenvektor  $\mathbf{v}$  zur Menge  $\mathcal{I}_i$ , siehe Lemma 8,

$$\begin{aligned} &= \sum_{k, I(k)=j} \lambda_k (\mathbf{P}_i \mathbf{P}_{I(k)} \mathbf{u}_k) (\mathbf{P}_i \mathbf{P}_{I(k)} \mathbf{u}_k)^\dagger \mathbf{P}_j \mathbf{u} \\ &= \sum_{k, I(k)=j} \lambda_k (\mathbf{P}_i \mathbf{P}_j \mathbf{u}_k) (\mathbf{P}_i \mathbf{P}_j \mathbf{u}_k)^\dagger \mathbf{P}_j \mathbf{u} \\ &= \mathbf{H}_{i, \not\subseteq} \mathbf{u}, \end{aligned}$$

und mit der Aussage zu Projektoren zu überlappenden Indexmengen (6.1.7a) folgt damit die Aussage für  $\mathbf{H}_i \mathbf{u}$ , denn mit  $\mathcal{I}_{i, \not\subseteq}$  anstelle von  $\mathcal{I}_i$  ist die obige Argumentation gleich.

Die Aussage für  $\mathbf{N}_i \mathbf{u}$  folgt wegen der Aufspaltung (6.1.4a) sofort, da hier nur die Summation über  $k$  weiter eingeschränkt wäre. □

*Bemerkung 7.* Wir bemerken, dass Korollar 5 nicht ohne Weiteres für  $\mathcal{I}_i \in \mathcal{J}_\cap$  in allgemeinen Zerlegungen gilt.

**Korollar 6** (Verschwindende orthogonale Anteile aus  $\mathcal{J}_\cap$ ). *Sei  $\mathcal{J}$  eine minimal überlappende Zerlegung von  $\mathcal{N}$ .*

*Für einen beliebigen Eigenvektor  $\mathbf{u}$  von  $\mathbf{H}$  mit  $\mathbf{H}\mathbf{u} = \lambda\mathbf{u}$  gilt für alle  $\mathcal{I}_i \in \mathcal{J}_\cap$ , dass  $\tilde{\mathbf{N}}_i \mathbf{u} = 0$ .*

*Beweis.* Sei  $\mathcal{I}_i \in \mathcal{J}_\cap$ . Dann zerlegen wir zunächst  $\mathcal{I}_i$  in zwei disjunkte Anteile: eine Teilmenge  $\mathcal{I}_{i, \not\subseteq} = \mathcal{I}_i \cap \mathcal{I}_j$  von  $\mathcal{I}_j$  und deren Komplement  $\mathcal{I}_{i, \emptyset} = \mathcal{I}_i - \mathcal{I}_{i, \not\subseteq}$ .

Alle Teilmengen von  $\mathcal{I}_i$  gehören dann zu genau einer der folgenden Mengen: Teilmengen von  $\mathcal{I}_{i, \not\subseteq}$ , Teilmengen von  $\mathcal{I}_{i, \emptyset}$  oder  $\mathcal{J}_\cap$ .

Anschließend zerlegen wir die Summe mit (6.2.4), wobei  $\mathcal{I}_i$  offensichtlich keine Teilmengen in  $\mathcal{J}_{\supseteq}$  oder in  $\mathcal{J}_{=}$  besitzen kann.

$$\tilde{\mathbf{N}}_i \mathbf{u} = \mathbf{N}_i \mathbf{u} - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i} \tilde{\mathbf{N}}_j \mathbf{u} = \mathbf{N}_i \mathbf{u} - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i, \supseteq} \tilde{\mathbf{N}}_j \mathbf{u} - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i, \emptyset} \tilde{\mathbf{N}}_j \mathbf{u} - \sum_{\substack{\mathcal{I}_j \in \mathcal{J}_{\cap}, \\ \mathcal{I}_j \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_j \mathbf{u} \quad (6.2.6)$$

Nun benötigen wir die strukturelle Induktion: Zunächst müssen wir  $\tilde{\mathbf{N}}_i \mathbf{u} = 0$  für alle Basisfälle zeigen. Dies ist die Induktionsannahme. Alle möglichen Basisfälle  $\mathcal{I}_k \in \mathcal{J}_{\cap}$  sind aber genau jene, bei denen die letzte Summe verschwindet. Die vorletzte Summe über die disjunkten Teilmengen verschwindet nach Korollar 4. Mit (6.2.5) aus Korollar 5 sehen wir, dass sich der erste Term und die erste Summe gegenseitig aufheben.

Die Induktionsvoraussetzung lautet, dass für alle Teilmengen  $\mathcal{I}_k \in \{\mathcal{I}_l | \mathcal{I}_l \in \mathcal{J}_{\cap}, \mathcal{I}_l \subsetneq \mathcal{I}_i\}$  die Aussage  $\tilde{\mathbf{N}}_k \mathbf{u} = 0$  gilt. Wir müssen zeigen, dass diese damit auch für  $\tilde{\mathbf{N}}_i \mathbf{u} = 0$  gilt. Dies folgt aber mit (6.2.6) sofort, da nach Voraussetzung die Aussage für alle Teilmengen gilt und damit die letzte Summe gleich null ist. Die weitere Argumentation erfolgt analog wie bei der Annahme.  $\square$

**Korollar 7** (Verschwindende orthogonale Anteile aus  $\mathcal{J}_{\supseteq}$ ). *Für einen beliebigen Eigenvektor  $\mathbf{u}$  von  $\mathbf{H}$  mit  $\mathbf{H}\mathbf{u} = \lambda\mathbf{u}$  gilt für alle  $\mathcal{I}_i \in \mathcal{J}_{\supseteq}$ , dass  $\mathbf{N}_i \mathbf{u} = 0$ .*

*Beweis.* Sei  $\mathcal{I}_i \in \mathcal{J}_{\supseteq}$ . Es gilt,

$$\mathbf{H}_i \mathbf{u} = (\mathbf{P}_i \mathbf{H} \mathbf{P}_i) \mathbf{u} = \mathbf{P}_i \mathbf{H} \mathbf{P}_i \mathbf{P}_j \mathbf{u} = \mathbf{P}_i \mathbf{H} \mathbf{P}_j \mathbf{u} = \mathbf{P}_i \mathbf{H} \mathbf{u} = \lambda \mathbf{P}_i \mathbf{u} = \lambda \mathbf{u} = \mathbf{H}_j \mathbf{u}$$

wegen (6.1.7c). Und damit ist  $\mathbf{u}$  sowohl Eigenvektor von  $\mathbf{H}_i$  als auch von  $\mathbf{H}_j$  mit Eigenwert  $\lambda$ . Dann gilt mit (6.1.4a),

$$\mathbf{N}_i \mathbf{u} = \mathbf{N}_j \mathbf{u} = \begin{cases} \lambda \mathbf{u} & , \quad \lambda < 0 \\ 0 & , \quad \lambda > 0 \end{cases}$$

wegen Bemerkung 5. Dann trennen wir die Summe gemäß (6.2.2) auf und erhalten

$$\begin{aligned} \tilde{\mathbf{N}}_i \mathbf{u} &= \mathbf{N}_i \mathbf{u} - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i} \tilde{\mathbf{N}}_j \mathbf{u} \\ &= \underbrace{\mathbf{N}_j \mathbf{u} - \sum_{\mathcal{I}_k \subsetneq \mathcal{I}_j} \tilde{\mathbf{N}}_k \mathbf{u}}_{=0} - \sum_{\substack{\mathcal{I}_k \in \mathcal{J}_{\supseteq}, \\ \mathcal{I}_k \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_k \mathbf{u} - \sum_{\substack{\mathcal{I}_k \in \mathcal{J}_{\cap}, \\ \mathcal{I}_k \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_k \mathbf{u} - \sum_{\substack{\mathcal{I}_k \in \mathcal{J}_{\emptyset}, \\ \mathcal{I}_k \subsetneq \mathcal{I}_i}} \tilde{\mathbf{N}}_k \mathbf{u}. \end{aligned} \quad (6.2.7)$$

Nun benutzen wir wieder die strukturelle Induktion: Zunächst müssen wir  $\tilde{\mathbf{N}}_i \mathbf{u} = 0$  für alle Basisfälle zeigen. Dies ist die Induktionsannahme. Alle möglichen Basisfälle  $\mathcal{I}_k \in \mathcal{J}_{\supseteq}$  sind aber wiederum genau jene, bei denen es keine Teilmengen von  $\mathcal{I}_i$  in  $\mathcal{J}_{\supseteq}$  gibt, womit die zweite Summe wiederum verschwindet. Die vorletzte und letzte Summe verschwinden wegen Korollar 4 und Korollar 6.

Die Induktionsvoraussetzung lautet, dass für alle Teilmengen  $\mathcal{I}_k \in \{\mathcal{I}_l | \mathcal{I}_l \subsetneq \mathcal{I}_i, \mathcal{I}_j \subseteq \mathcal{I}_l\}$  die Aussage  $\tilde{\mathbf{N}}_k \mathbf{u} = 0$  gilt. Wieder müssen wir zeigen, dass diese auch für  $\tilde{\mathbf{N}}_i \mathbf{u} = 0$  gilt. Dies folgt aber mit (6.2.7) sofort, da dann analog die zweite Summe verschwindet.  $\square$

## 6. Dünnbesetzte, indefinite Matrizen

**Korollar 8** (Kern und Bild der  $\mathbf{N}_i$ ). *Für alle Mengen  $\mathcal{I}_i \in \mathcal{J}_{\subseteq} \cup \mathcal{J}_{=}$  gilt, dass für einen Eigenvektor  $\mathbf{u}$  von  $\mathbf{H}$  mit  $\mathbf{H}\mathbf{u} = \lambda\mathbf{u}$  der Beitrag  $\tilde{\mathbf{N}}_i\mathbf{u}$  im Allgemeinen ungleich Null ist. Für die Summe der Beiträge gilt,*

$$\begin{cases} \lambda < 0 & : \sum_{\mathcal{I}_i \in \mathcal{J}_{\subseteq} \cup \mathcal{J}_{=}} \tilde{\mathbf{N}}_i\mathbf{u} = \lambda\mathbf{u}, \\ \lambda > 0 & : \sum_{\mathcal{I}_i \in \mathcal{J}_{\subseteq} \cup \mathcal{J}_{=}} \tilde{\mathbf{N}}_i\mathbf{u} = 0. \end{cases} \quad (6.2.8)$$

*Beweis.* Der Fall für negative Eigenvektoren folgt sofort mit (6.2.2) und weil  $\mathbf{u}$  negativer Eigenvektor von  $\mathbf{H}_i$  und damit auch von  $\mathbf{N}_i$  ist.

Der Fall für positive Eigenvektoren folgt, weil  $\mathbf{u}$  wieder Eigenvektor von  $\mathbf{H}_i$  ist, welchen wir wie in (6.1.4a) und (6.1.4b) zerlegen können. Da  $\mathbf{u}$  aber dann positiver Eigenvektor in  $\mathbf{P}_i$  sein muss und daher im Kern von  $\mathbf{N}_i$  liegen muss, siehe Bemerkung 5, folgt das Gewünschte.  $\square$

Mit den Korollaren 4 bis 8 ist zu einer gegebenen minimal überlappenden Zerlegung  $\mathcal{J}$  gezeigt, dass für einen beliebigen negativen Eigenvektor  $\mathbf{u}$  mit zugehöriger Indexmenge  $\mathcal{I}_j$  mit  $\lambda_j < 0$  gilt,

$$\begin{aligned} \sum_{\mathcal{I}_i \in \mathcal{J}} \tilde{\mathbf{N}}_i\mathbf{u} &= \sum_{\mathcal{I}_i \in \mathcal{J}_0} \tilde{\mathbf{N}}_i\mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_{\subseteq}} \tilde{\mathbf{N}}_i\mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_{\supseteq}} \tilde{\mathbf{N}}_i\mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_{=}} \tilde{\mathbf{N}}_i\mathbf{u} + \sum_{\mathcal{I}_i \in \mathcal{J}_{\cap}} \tilde{\mathbf{N}}_i\mathbf{u} \\ &= \sum_{\mathcal{I}_i \subseteq \mathcal{I}_j} \tilde{\mathbf{N}}_i\mathbf{u} = \mathbf{N}_j\mathbf{u} = \lambda\mathbf{u}, \end{aligned}$$

und dass für einen beliebigen positiven Eigenvektor gilt,

$$\sum_{\mathcal{I}_i \in \mathcal{J}} \tilde{\mathbf{N}}_i\mathbf{u} = \sum_{\mathcal{I}_i \subseteq \mathcal{I}_j} \tilde{\mathbf{N}}_i\mathbf{u} = \mathbf{N}_j\mathbf{u} = 0. \quad (6.2.9)$$

Damit folgt die Behauptung von Lemma 12.

Die Forderung nach einer minimal überlappenden Zerlegung ist keine Einschränkung der Zerlegung selbst, sondern eine Einschränkung der möglichen Struktur der Eigenvektoren. Dies konnten wir jedoch in der aktuellen Arbeit nicht mehr zeigen, sondern führen es hier als Behauptung.

**Behauptung 1** (Größere Zerlegungen). *Die Gleichung (6.2.2) gilt für beliebige Zerlegungen  $\mathcal{J}$ , solange sie die minimal überlappende Zerlegung enthalten und hierarchisch sind.*

Zur Bestimmung der negativen Eigenwertsumme genügt also unter der geforderten Voraussetzungen die negative Eigenwertsumme der Hauptidealuntermatrizen zu einer Zerlegung  $\mathcal{J}$  geschickt aufzusummieren.

Da die Summe der Eigenwerte gewichtet mit ihrer Anzahl den Schwerpunkt dieser Eigenwerte beschreibt, haben wir das Theorem Schwerpunkttheorem genannt. Wir werden gleich noch näher ausführen, warum dieser Name auch bildlich sinnvoll ist.

Als weitere Behauptung führen wir auf, dass der Projektor in den negativen Eigenraum ebenfalls in analoger Weise zu Lemma 12 erstellt werden kann.



**Behauptung 2** (Projektor in negativen Eigenraum). *Unter gleichen Voraussetzungen wie in Lemma 12 lässt sich zeigen, dass*

$$\mathbf{P}^- = \sum_{i, \lambda_j < 0} \mathbf{u}\mathbf{u}^\dagger \quad (6.2.10)$$

einen Projektor in den negativen Eigenraum definiert, welcher nur aus Vektoren besteht, die aus kleinen Hauptuntermatrizen berechnet worden sind.

### 6.2.3. Ausführliches Beispiel

Wir wollen eine prototypische Beispielmatrix  $\mathbf{H}$  ausführlicher untersuchen. Dazu benötigen wir eine komplexwertige hermitesche oder reellwertige symmetrische Matrix der Dimension  $n$  mit einer Reihe Eigenvektoren zu negativen Eigenwerten, die jeweils nur  $\mathcal{O}(1)$  Komponenten ungleich Null besitzen.

Das folgende Beispiel soll zeigen, dass es eine dünnbesetzte Matrix mit einer Reihe dünnbesetzter Eigenvektoren gibt.

Wir setzen die Eigenvektoren wie folgt an,

$$\begin{aligned} & \left(-\frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0\right)^T, \\ & \left(0 \quad 0 \quad -\frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad 0 \quad 0 \quad 0 \quad \dots \quad 0\right)^T, \\ & \left(0 \quad 0 \quad 0 \quad 0 \quad -\frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad 0 \quad \dots \quad 0\right)^T, \dots \end{aligned} \quad (6.2.11)$$

welche sofort ersichtlich ein System orthonormaler Vektoren bilden. Im Raum der Dimension  $n$  können wir dabei  $\frac{1}{2}(n-2)$  Vektoren unterbringen.

Anschließend wollen wir für  $n=8$  eine Matrix  $\mathbf{H}$  erzeugen, welche also die ersten drei Vektoren der Gestalt (6.2.11) als Eigenvektoren  $\mathbf{u}_1$ ,  $\mathbf{u}_2$ ,  $\mathbf{u}_3$  zu Eigenwerten  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  haben soll. Wir stellen die Gleichungen auf,

$$\begin{aligned} \mathbf{H}\mathbf{v}_1 &= \lambda_1\mathbf{u}_1, \\ \mathbf{H}\mathbf{v}_2 &= \lambda_2\mathbf{u}_2, \\ \mathbf{H}\mathbf{v}_3 &= \lambda_3\mathbf{u}_3, \end{aligned}$$

und erhalten  $6 + 8 + 6 = 20$  nicht-triviale Gleichungen, welche die Matrixeinträge einer reellwertigen  $8 \times 8$ -Matrix<sup>2</sup> der folgenden Gestalt

$$\mathbf{H} = \begin{pmatrix} a & b & h_1 & h_1 & 0 & 0 & 0 & 0 \\ b & c & h_2 & h_2 & 0 & 0 & 0 & 0 \\ h_1 & h_2 & d & h_3 & h_4 & h_4 & 0 & 0 \\ h_1 & h_2 & h_3 & d & -h_4 & -h_4 & 0 & 0 \\ 0 & 0 & h_4 & -h_4 & f & h_5 & g & h_6 \\ 0 & 0 & h_4 & -h_4 & h_5 & f & -g & -h_6 \\ 0 & 0 & 0 & 0 & g & -g & h & h_7 \\ 0 & 0 & 0 & 0 & h_6 & -h_6 & h_7 & h_8 \end{pmatrix} \quad (6.2.12)$$

<sup>2</sup>Diese Matrix soll der Einfachheit halber eine Bandstruktur aufweisen und hat damit  $28 + 8 - 12 = 24$  Freiheitsgrade.

## 6. Dünnbesetzte, indefinite Matrizen

mit den Hilfsgrößen

$$\begin{aligned}
 h_1 &= \frac{1}{2}(-\lambda_1 + a - b), \\
 h_2 &= \frac{1}{2}(\lambda_1 + b - c), \\
 h_3 &= \frac{1}{2}a - b + \frac{1}{2}c - d, \\
 h_4 &= \frac{1}{2}(-\lambda_2 - \frac{1}{2}a + b - \frac{1}{2}c + 2d), \\
 h_5 &= -\frac{1}{2}a + b - \frac{1}{2}c + 2d - f, \\
 h_6 &= -\lambda_3 + \frac{1}{2}a - b + \frac{1}{2}c - 2d + 2f - g, \\
 h_7 &= \lambda_3 + 2g - h, \\
 h_8 &= -2\lambda_3 + a - 2b + c - 4d + 4f - 4g + h,
 \end{aligned}$$

bis auf 7 Freiheitsgrade  $a, \dots, g$  bestimmen. Wir bemerken dabei, dass das System von Gleichungen lösbar ist und einige Gleichungen äquivalent sind.

Natürlich müssen wir nun Werte für die 7 Freiheitsgrade finden, so dass keine anderen negativen Eigenwerte vorhanden sind. Solch einen Satz finden wir mit  $\{a, \dots, g\} = \{5; 2; 2, 5; 2; 2, 5; 2; 5\}$  und den gewählten negativen Eigenwerten  $-3, -2, -1$  und erhalten damit folgende Beispielmatrix,

$$\mathbf{H} = \begin{pmatrix} 5 & 2 & 3 & 3 & 0 & 0 & 0 & 0 \\ 2 & 2.5 & -1.75 & -1.75 & 0 & 0 & 0 & 0 \\ 3 & -1.75 & 2 & -0.25 & 2.125 & 2.125 & 0 & 0 \\ 3 & -1.75 & -0.25 & 2 & -2.125 & -2.125 & 0 & 0 \\ 0 & 0 & 2.125 & -2.125 & 2.5 & -0.25 & 2 & 1.75 \\ 0 & 0 & 2.125 & -2.125 & -0.25 & 2.5 & -2 & -1.75 \\ 0 & 0 & 0 & 0 & 2 & -2 & 5 & -2 \\ 0 & 0 & 0 & 0 & 1.75 & -1.75 & -2 & 4.5 \end{pmatrix},$$

mit den Eigenwerten  $-3, -2, -1, 4.3179, 6.25, 6.5, 7, 7.9321$ , wobei alle Eigenvektoren maximal 4 nicht-Null Komponenten haben.

Damit können wir nun die Summation in (6.2.3) nachvollziehen. Es gibt dabei nur drei Gruppen von Indexmengen zu Eigenvektoren, nämlich  $\mathcal{I}_1 = \{1, 2, 3, 4\}$ ,  $\mathcal{I}_2 = \{3, 4, 5, 6\}$  und  $\mathcal{I}_3 = \{5, 6, 7, 8\}$ . Wir benötigen die Schnittmengen  $\mathcal{I}_4 = \{3, 4\}$  und  $\mathcal{I}_5 = \{5, 6\}$  für die hierarchische, minimal überlappende Zerlegung  $\mathcal{J} = \{\mathcal{I}_i\}$ .

Betrachten wir

$$\begin{aligned}
 \sum_{\mathcal{I}_i \in \mathcal{J}} \text{Tr}(\tilde{\mathbf{N}}_j) &= \text{Tr}(\mathbf{N}_1) + \text{Tr}(\mathbf{N}_2) + \text{Tr}(\mathbf{N}_3) - \text{Tr}(\mathbf{N}_4) - \text{Tr}(\mathbf{N}_5) \\
 &= -3 + -2 + -1 - 0 - 0 \\
 &= -6 = \text{Tr}(\mathbf{N}),
 \end{aligned}$$

so erhalten wir die korrekte Eigenwertsumme. Es lassen sich leicht kompliziertere Beispiele finden.

### 6.2.4. Verwandtschaft mit Gebietszerlegungen

Es besteht eine gewisse Ähnlichkeit der dargestellten Zerlegung eines großen Eigenwertproblems in überlappende Teilprobleme zur Gebietszerlegung bei der Lösung linearer Gleichungssysteme, den Schwarz-Methoden: Die Kopplung der verschiedenen (überlappenden) Gebiete, Fragmente eines linearen Gleichungssystems, wird dabei über entsprechende gemeinsame Randbedingungen erreicht. Dabei müssen die Lösungen auf den verschiedenen Teilgebieten auf einer Untermenge, dem Rand, gleich sein. Bekannte Matrix-Methoden sind bspw. Block-Jacobi, welche der additiven Schwarz-Methode entspricht, siehe [149, Kap. 12.2 bzw. Kap. 14.3].

Bei Eigenwertproblemen gibt es auch einen solchen Rand: Er ist bestimmt durch Orthogonalität bestimmter Teilräume. Die Teilprobleme, welche sich aus den Eigenvektoren der Hauptuntermatrizen ergeben, spannen Räume auf, die jeweils zueinander orthogonal sein müssen, um ein gültiges volles Eigensystem zu bilden. Denn die Eigenvektoren des vollen Problems sind alle paarweise orthogonal. Der gemeinsame Rand zweier Unterprobleme bei den linearen Gleichungssystemen entspricht also genau dem Überlappunterraum  $\mathbb{C}^{\mathcal{I}_j}$  zweier Hauptuntermatrizen, indem deren Eigenvektoren zueinander Orthogonalität erfüllen müssen, der Schnittmenge  $\mathcal{I}_j$  der beiden Indexmengen.

Die Ähnlichkeit ergibt sich aber erst durch die geforderte Dünnbesetztheit der Eigenvektoren, da die Orthogonalität ansonsten ein globales Kriterium wäre. Nur durch die Dünnbesetztheit erhält sie einen lokalen Charakter, der nur auf einer kleinen Untermenge  $\mathbb{C}^{\mathcal{I}_j}$  ausgetragen wird. Analog bestünde andernfalls ein Rand zwischen jedem Paar von Fragmenten und nicht nur zwischen einem Fragment mit seinen wenigen Nachbarn. Bei den partiellen Differentialgleichungen ergibt sich diese Lokalität des Problems bspw. aus der Kompaktheit des Differenzensterns und der damit einhergehenden dünnbesetzten Matrix.

### 6.2.5. Beziehung zu Gelfand-Tsetlin-Mustern

Im Folgenden wollen wir einen schönen Zusammenhang herausarbeiten, der das Schwerpunkttheorem zusammen mit der Interlacingtheorem in Verbindung mit *Gelfand-Tsetlin-Mustern*, siehe [170, Remark 1.3.8], stellt und die Namensgebung des Theorems 2 noch einmal verdeutlicht.

Erinnern wir uns zunächst an das Interlacingtheorem.

**Theorem 3** (Interlacing Theorem). *Sei  $\mathbf{H}$  eine hermitesche  $n \times n$ -Matrix, die wie folgt unterteilt ist,*

$$\mathbf{H} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\dagger & \mathbf{C} \end{pmatrix},$$

wobei  $\mathbf{A}$  eine  $m \times m$ -Hauptuntermatrix von  $\mathbf{H}$  mit  $1 \leq m \leq n$  ist. Dann folgt,

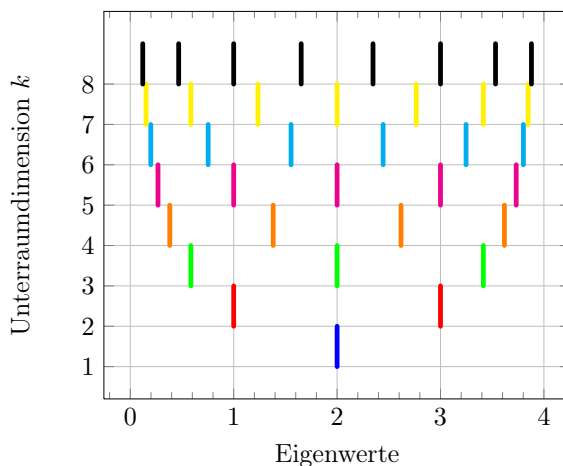
$$\lambda_{k+n-m}(\mathbf{H}) \leq \lambda_k(\mathbf{A}) \leq \lambda_k(\mathbf{H}), \quad k = 1, 2, \dots, m. \quad (6.2.13)$$

*Beweis.* Es finden sich in [190, S. 222ff.] drei verschiedene ausführliche Beweisversionen.  $\square$

## 6. Dünnbesetzte, indefinite Matrizen

Es existieren also jeweils Intervalle, gebildet aus den Eigenwerten von  $\mathbf{H}$ , welche die Eigenwerte einer Hauptuntermatrix  $\mathbf{A}$  beschränken.

*Bemerkung 8* (Intervallbreite). Wir halten fest, dass die (unterschiedlichen) Dimensionen von Matrix und Hauptuntermatrix entscheidend für die Breite der Intervalle  $\Delta := n - m$  sind.



**Abbildung 6.3.:** Spektrum der führenden Hauptuntermatrizen einer  $8 \times 8$ -Tridiagonalmatrix. Deutlich ist die Intervallierung der Eigenwerte zu erkennen und dass der Schwerpunkt, festgelegt durch die Spur, ebenfalls erhalten bleibt.

Wir betrachten nun eine  $8 \times 8$ -Tridiagonalmatrix  $\mathbf{H}$  mit den Einträgen 2 auf der Hauptdiagonalen und den Einträgen 1 auf der Nebendiagonalen. Die Eigenwerte für die führenden Hauptuntermatrizen  $\mathbf{H}_i$  zu  $\mathcal{I}_k = \{1, \dots, k\}$  mit  $1 \leq k \leq 8$  mit Mächtigkeit  $k$  sind in Abbildung 6.3 aufgetragen.

Zunächst wird die Gültigkeit des Interlacingtheorems deutlich sichtbar. Aber es fällt ebenso auf, dass der Schwerpunkt der Eigenwerte erhalten bleibt. Dies ist nicht verwunderlich, da er der Spur der Eigenwerte gewichtet mit ihrer Anzahl entspricht.

Würden wir nicht 1 sondern  $-1.5$  auf die Nebendiagonalen setzen, so ergäben sich auch negative Eigenwerte, aber keine zwei Schwerpunkte wie es das Schwerpunkttheorem nahegelegt, da die Eigenvektoren dieser Matrix nicht dünn-, sondern vollbesetzt sind. Wir werden im später folgenden Abschnitt 9.6.1 ein Beispiel einer indefiniten Matrix nachreichen, wo die negativen Eigenwerte einen eigenen Schwerpunkt nach Theorem 2 besitzen.

### 6.3. Störung

Es gibt viele Betrachtungen von gestörten Matrizen im Rahmen von Eigenwertproblemen. Hierbei muss zwischen dem Fokus auf Eigenwerte und dem auf Eigenvektoren bzw. besser

Eigenräumen<sup>3</sup> geachtet werden. Wir geben an dieser Stelle aber nur Ergebnisse in Bezug zur Störung der Eigenwerte wieder. Dabei beschränken wir uns auf die Ergebnisse für hermitesche Matrizen.

Wir verwenden dabei folgende Notation: Wir betrachten stets eine  $n$ -dimensionale, hermitesche Matrix  $\mathbf{A} \in \mathcal{M}_n$  mit Eigenwerten  $\{\lambda_i\}$  und Eigenvektoren  $\{\mathbf{u}_i\}$ . Es gilt also:

$$\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i \quad \forall i \in \{1, \dots, n\}. \quad (6.3.1)$$

Zu dieser Matrix wird eine zweite Matrix  $\mathbf{E}$  hinzuaddiert. Diese zweite Matrix betrachten wir dabei als eine Störung, d. h. die Norm von  $\mathbf{E}$  soll klein sein. Die resultierende, gestörte Matrix heißt hier stets  $\tilde{\mathbf{A}}$ .

Einige klassische Theoreme für die Eigenwerte gestörter Matrizen findet man schön zusammen gefasst in der Arbeit von Braun [24, Kap. 3.6] oder auch in [103, 52].

Das erste Theorem gibt uns eine Aussage darüber wie sehr sich ein bestimmter Eigenwert durch Addition von  $\mathbf{E}$  mindestens verändert:

**Theorem 4** (Bauer-Fike). *Sei  $\tilde{\lambda}$  ein bestimmter Eigenwert von  $\tilde{\mathbf{A}}$ . Dann gilt:*

$$\min_{\lambda \in \lambda(\mathbf{A})} |\lambda - \tilde{\lambda}| \leq \kappa(\mathbf{U})\|\mathbf{E}\|, \quad (6.3.2)$$

wobei  $\mathbf{U}$  die Matrix mit den Eigenvektoren von  $\mathbf{A}$  als Spaltenvektoren ist und  $\kappa(\mathbf{U})$  die Konditionszahl von  $\mathbf{U}$ .

Wir bemerken, dass wenn  $\mathbf{A}$  hermitesch ist, deren Eigenvektoren orthogonal und die Konditionszahl  $\kappa(\mathbf{U})$  damit natürlich 1 ist. Die Änderung durch  $\mathbf{E}$  ist also mindestens so groß wie dessen Norm.

Im folgenden Theorem betrachten wir die quadratische Abweichung über alle Eigenwertpaare  $(\lambda, \tilde{\lambda})$  und geben ein Theorem wieder, dass diese nach oben abschätzt:

**Theorem 5** (Hoffman-Wielandt). *Seien  $\mathbf{A}$  und  $\tilde{\mathbf{A}}$  hermitesche Matrizen. Seien weiterhin  $\lambda_1 \leq \dots \leq \lambda_n$  die Eigenwerte von  $\mathbf{A}$  und seien  $\tilde{\lambda}_1 \leq \dots \leq \tilde{\lambda}_n$  die Eigenwerte von  $\tilde{\mathbf{A}}$ . Dann gilt:*

$$\sum_{i=1}^n (\lambda_i - \tilde{\lambda}_i)^2 \leq \|\mathbf{E}\|_F^2, \quad (6.3.3)$$

wobei mit  $\|\cdot\|_F^2$  die Frobenius-Norm bezeichnet ist.

Dies lässt sich noch ein wenig genauer für einzelne Eigenwertpaare wie folgt ausdrücken:

**Theorem 6** (Weyl). *Seien  $\mathbf{A}$  und  $\mathbf{E}$  hermitesche  $n \times n$ -Matrizen. Dann gilt für  $1 \leq i \leq n$ , dass*

$$\lambda_i(\mathbf{A}) + \lambda_n(\mathbf{E}) \leq \lambda_i(\mathbf{A} + \mathbf{E}) \leq \lambda_i(\mathbf{A}) + \lambda_1(\mathbf{E}), \quad (6.3.4)$$

wobei mit  $\lambda_1(\cdot)$  wieder der größte Eigenwert und mit  $\lambda_n(\cdot)$  der kleinste Eigenwert einer Matrix bezeichnet ist. Damit gilt besonders:

$$|\lambda_i(\mathbf{A}) - \lambda_i(\mathbf{A} + \mathbf{E})| \leq \|\mathbf{E}\|. \quad (6.3.5)$$

<sup>3</sup>Da es zu einem bestimmten Eigenwert mehrere Eigenvektoren geben kann, sind die Eigenvektoren zu diesen *entarteten* Eigenwerten nicht wohldefiniert, die Eigenräume sind es aber schon.

## 6. Dünnbesetzte, indefinite Matrizen

Allgemein lässt sich sagen, dass das Spektrum einer hermiteschen Matrix stabil gegenüber kleinen Störungen ist, siehe Tao [170, Abschn. 1.3]. Die quadratische Abweichung über alle Eigenwertpaare bzw. auch die maximale Abweichung eines Eigenwertpaares korreliert mit der Norm der Störung.

Abschließend möchten wir noch bemerken, dass Theoreme zur Störung von Eigenräumen auf das so genannte  $\sin \theta$ -Theorem von Davis und M [40] zurückgehen. Darauf aufbauende, weitergehende Arbeiten finden sich unter anderen von Ipsen [103] und Eisenstat und Ipsen [52].

### 6.4. Weiterführende Literatur

Im Folgenden führen wir einige verwandte Arbeiten in der Literatur zusammen, welche ähnliche Aspekte wie das Schwerpunkttheorem untersuchen.

#### 6.4.1. Summen hermitescher Matrizen

Eine allgemeine Frage lautet: Seien zwei hermitesche Matrizen  $\mathbf{A}$  und  $\mathbf{B}$  gleicher Dimension gegeben und deren Spektren bekannt. Was kann man dann über das Spektrum von  $\mathbf{A} + \mathbf{B}$  aussagen?

Zunächst wurden eine Reihe von Ungleichungen gefunden, bspw. ergibt sich aus der Additivität der Spur leicht,  $\text{Tr}(\mathbf{A} + \mathbf{B}) = \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B})$ . Es wurden viele weitere solche Ungleichungen aufgestellt, die das mögliche Spektrum einschränken. Dies gipfelte schließlich in der Horn'schen Vermutung, welche kürzlich bewiesen werden konnte, siehe Fulton [61] für eine Übersicht und Knutson und Tao [119] für den Beweis der Vermutung im Zusammenhang mit sogenannten Honigwaben-Diagrammen.

Für das von uns gezeigte Theorem benötigten wir jedoch mehr als Ungleichungen, schließlich zeigen wir die Exaktheit der Eigenwertspektren. Dies gelingt jedoch nur, wenn die Eigenvektoren der Matrix eine spezielle, eben dünnbesetzte, Form besitzen.

#### 6.4.2. Dünnbesetzte Faktorisierungen

Eine Reihe von Arbeiten beschäftigen sich mit der Frage, warum für Matrizen, deren Nebendiagonalelemente einen raschen Abfall zeigen, auch die Faktoren von Cholesky-, LU- oder QR-Zerlegungen diesen Abfall behalten. Hare et al. [83] und Pothen [145] haben sich mit dieser Frage für die QR-Faktoren beschäftigt, Krishtal et al. [124] entwickeln eine vollständige Theorie für alle bekannten Faktorisierungen.

Dünnbesetzte Faktoren könnten helfen, die Annahme der Dünnbesetztheit der Matrix zu überwinden. Jedoch gibt es keinen direkten Zusammenhang zwischen der Faktorisierung einer Matrix und ihrer Eigenzerlegung.

#### 6.4.3. Dünnbesetzte hermitesche Matrizen

Benzi und Razouk [17] untersucht die glatten Funktionen  $f(\mathbf{A})$  für dünnbesetzte, diagonalisierbare Matrizen  $\mathbf{A}$  auf exponentiellen Abfall und zeigen, dass dünnbesetzte Ap-

proximationen von  $f(\mathbf{A})$  berechnet werden können. In [18] beschäftigen sie sich damit im Zusammenhang mit linear skalierenden Methoden in der Elektronenstrukturberechnungen. Sie gehen dabei von der dünnbesetzten Dichtematrix aus, die letztlich als ein spektraler Projektor betrachtet werden kann, da sie aus den negativen Eigenvektoren der Fockmatrix gebildet wird. Sie beweisen, dass Systeme mit Bandlücke zu einem solchen exponentiellen Abfallverhalten führen und damit lineare skalierende Algorithmen möglich machen und hatten zuvor gezeigt, dass die Matrix sich deswegen dünnbesetzt gut approximieren lässt. Die Betrachtung der Dichtematrix ist besonders im Rahmen der Dichte-basierten HF-Methode interessant, wo nicht die Fockmatrix diagonalisiert, sondern über ein direktes Minimierungsverfahren die richtige Dichtematrix gesucht wird, siehe auch den späteren Abschnitt 7.6. Letztlich geht die Aussage in die gleiche Richtung, wie auch in Abschnitt 5.5 zitiert, wo wir das asymptotische Verhalten der Einteilchen-Wellenfunktion als Lösungen der HF-Gleichung wiedergeben in Kombination mit der definierten lokalen Basis, siehe Abschnitt 5.3.

Wir haben uns in dieser Arbeit aber besonders auf das von uns vorgestellte Verfahren zur schnellen Diagonalisierung einer Matrix über die Hauptuntermatrizen konzentriert.

#### 6.4.4. Momente von Hauptuntermatrizen

Es gibt Arbeiten zur Verallgemeinerung der Poincaré-Ungleichung, welche die Anzahl positiver Eigenwerte einer Matrix mit der Anzahl positiver Eigenwerte von Hauptuntermatrizen vergleicht, siehe die frühe Arbeit von Haynsworth [88] oder später von Dancis [39].

Eine reine Abschätzung der Anzahl hilft zur Bestimmung der Dimensionen der positiven und negativen Eigenunterräume. Wir können aus diesen Anzahlen aber keinen Nutzen ziehen, sondern benötigen Aussagen über die Struktur der Unterräume.

#### 6.4.5. Blockdiagonalisierungsmethoden

Numerische Methoden, welche eine Matrix durch Betrachtung einzelner Blöcke diagonalisieren, haben eine lange Tradition, insbesondere seit parallele Algorithmen aufgrund zunehmende Präsenz von Vielprozessorarchitekturen immer wichtiger werden.

Sie sind besonders bekannt beim (iterativen) Lösen linearer Gleichungssysteme, wie die Block-Gauss-Elimination, die schon erwähnten alternierende Schwarz-Methoden oder auch Schur-Komplement-Methoden, siehe Saad [149, Kap. 14].

Auch für Eigenwertprobleme gibt es solche Ansätze. Beispielsweise ist der Cuppen-Algorithmus [38] eine effektive Methode zur Diagonalisierung tridiagonaler Matrizen. Außerdem haben wir bereits die Block-Jacobi- und Block Gauss-Seidel-Methoden erwähnt. Sie können als Vorkonditionierer für Eigenwertprobleme benutzt werden, bspw. beschäftigt sich Stathopoulos et al. [162] damit im Zusammenhang mit dem verallgemeinerten Davidson-Algorithmus. Eigenwertprobleme sind jedoch schwierig, da sie eben nicht-linear sind.

Die von uns beschriebene Methode ist aber eben nicht iterativ, sondern ein direkter Löser. Sie nutzt geschickt die Struktur der Lösung aus, so dass die "Gleichheit auf dem

Rand" automatisch erfüllt wird und die zwischenzeitlichen Lösung auf den einzelnen Teilgebieten nicht weiter auskonvergiert werden müssen, um diese Gleichheit bzw. die Orthogonalität zu erfüllen.

## **Zusammenfassung**

In diesem Kapitel befassten wir uns mit dünnbesetzten Matrizen mit dünnbesetzten Eigenvektoren im Rahmen von Eigenwertproblemen. Wir benutzten Konzepte der Dünnbesetztheit, Unterindexmengen, Hierarchien und Hauptuntermatrizen, um einen direkten Zusammenhang zwischen den negativen dünnbesetzten Eigenvektoren der vollen Matrix und den negativen Eigenvektoren der Hauptuntermatrizen zu finden, dem Schwerpunkttheorem. Ist also etwas über die Null/Nicht-Null-Struktur der Eigenvektoren bekannt, so lässt sich dies benutzen, um die Eigenwerte und Eigenvektoren der vollen Matrix in linearer Zeit aus kleinen Hauptuntermatrizen zu berechnen. Wir erläuterten, wie das Theorem mit dem sehr bekannten Interlacing-Theorem zusammenspielt, um Gelfand-Tsetlin-Muster im Eigenwertspektrum auftauchen zu lassen. Außerdem gaben wir verwandte Ansätze aus der Literatur in diesem Themenkreis wieder.



## 7. Linear skalierende Hartree-Fock-Methode

Im Kapitel 4 haben wir die HF-Gleichung aus dem Ansatz der antisymmetrischen, vollständig dekorrelierten Ansatzfunktion, besser bekannt als Slater-Determinante, hergeleitet. Der beschriebene Algorithmus skaliert dabei ohne basisspezifische Abschätzungen mit  $\mathcal{O}(n^4)$  mit der Anzahl an Basisfunktionen  $n$ , siehe Abschnitt 4.3.

Betrachten wir den in Abschnitt 4.3.2 dargestellten HF-Algorithmus, so erkennen wir dabei drei große Blöcke: Aufstellen von Matrizen, Diagonalisieren und Transformieren dieser und schließlich die notwendige Iteration.

In diesem Kapitel untersuchen wir mit den Analysen und Herleitungen der letzten beiden Kapitel die klassische HF-Methode bezüglich dieser drei Blöcke. Dabei wollen wir den Ansatz der Fragmentierungsmethoden benutzen. Diese gehen auf eine Idee von Hurley et al. [102] zurück, wo einzelne Elektronen zu Paaren zusammengefasst werden. McWeeny [133] beschäftigte sich ebenso mit solchen Gruppenfunktionen und Christoffersen und Maggiora [34] waren eine der ersten, die schließlich den Begriff *Fragment* für ein molekulares Teilsystem einführten und eine approximative Methode zur Lösung des HF-Problems vorstellten. Wir fassen die bekannteren und neueren Ansätze aufgrund ihrer großen Anzahl in einem getrennten Abschnitt A.4.3 im Anhang zusammen.

Bei diesen Methoden wird das Gesamtproblem nach dem bekannten Divide&Conquer-Muster in kleinere Teilprobleme aufgespalten, der Fragmentierung. Die gesuchte Eigenschaft wird nur für diese Fragmente berechnet und diese anschließend zu einer Näherung an die Eigenschaft des Gesamtsystems summiert. Skaliert die Anzahl der Fragmente linear mit der Anzahl an Basisfunktionen bzw. der Anzahl der Elektronen, so ergibt sich automatisch eine linear skalierende Methode. Wir wollen diese Gruppe von Methoden für die HF-Gleichung grundlegend motivieren. Dabei stellen wir auch eine neue approximative Methode namens BOSSANOVA vor, welche von dieser grundlegenden Betrachtung stark profitiert.

In Abschnitt 7.1 nutzen wir, dass die Dichtematrix in einer geeigneten lokalen Basis dünnbesetzt ist und untersuchen damit die Dünnbesetztheit der Fockmatrix. Anschließend erläutern wir in Abschnitt 7.2 wie dies die Anwendung des Schwerpunkttheorems aus Kapitel 6 erlaubt. Mit der langreichweitigen und langsam abfallenden Coulomb-Wechselwirkung, welche zunächst quadratische Kosten verursacht, befassen wir uns in Abschnitt 7.3, wo wir eine Aufspaltung in einen Anteil mit konstanten und einen Anteil mit linearen Kosten definieren. In Abschnitt 7.4 betrachten wir anschließend mittels eines Störansatzes, inwieweit wir die langreichweitige Wechselwirkung vernachlässigen können, wodurch wir die globale nicht-lineare Kopplung der HF-Gleichung in Kopplungen nur innerhalb der Fragmente auftrennen.

## 7. BOSSANOVA

In Abschnitt 7.5 erläutern wir schließlich, wie wir aus dem Bindungsgraphen des molekularen Systems die Information über die Dünnbesetztheit der Eigenvektoren gewinnen. Weiterhin stellen wir dar, wie dabei nur eine linear mit der Anzahl der Kerne skalierende Anzahl an überlappenden Fragmenten betrachtet wird.

Abschließend geben wir einen knappen Überblick im Abschnitt 7.6 über einen anderen Ansatz zur Erlangung linearer Komplexität, bei dem nicht iterativ die Einteilchen-Wellenfunktionen, sondern die richtige Dichtematrix über die Minimierung eines Funktionals gesucht wird.

### 7.1. Dünnbesetzte Fockmatrix

Zunächst widmen wir uns dem ersten der drei Blöcke, dem Aufstellen der Fockmatrix. Dazu befassen wir uns zunächst noch einmal mit den Elementen der Fockmatrix, siehe Abschnitt 4.2.2, angeregt durch die Konsequenzen aus dem Kapitel 5.

#### 7.1.1. Matricelemente in lokaler Basis

In Abschnitt 4.2.2 haben wir in Gleichung (4.2.9) beschrieben, dass sich die Matrix aus vier Anteilen zusammensetzt: Kinetischer Anteil, Kern-Elektron-Wechselwirkung, Austausch-Wechselwirkung und Elektron-Elektron-Wechselwirkung. In Abschnitt 4.3.1 haben wir anschließend die Komplexität ohne Einschränkung der Basis untersucht. Im Folgenden gehen wir nun von einer lokalen Basis aus, siehe Abschnitt 5.3, und untersuchen erneut jede dieser vier Matrizen einzeln.

Dabei nehmen wir nun an, dass die Basisfunktionen die Form (5.1.3c) haben,

$$p(\mathbf{r} - \mathbf{r}_0) \cdot \exp(q(\mathbf{r} - \mathbf{r}_0)),$$

sie also insbesondere in der Menge der lokalen Funktionen sind, siehe Abschnitt 5.2.

#### Kinetischer Anteil

Für den kinetischen Anteil müssen wir nur zeigen, dass die Ableitung einer beliebigen Funktion aus  $\mathcal{K}$  immer noch in der Menge der lokalen Funktionen ist. Damit folgt automatisch ihre asymptotische Abschätzung nach oben und mit Lemma 6 das Gewünschte.

Vorausgesetzt dass die zweite Ableitung der Funktion aus  $\mathcal{K}$  existiert, muss aber

$$\frac{\partial^2}{\partial \mathbf{r}_i^2} (p(\mathbf{r} - \mathbf{r}_0) \cdot \exp(q(\mathbf{r} - \mathbf{r}_0))) \quad (7.1.1)$$

wieder in der Menge  $\mathcal{K}$  liegen.

Damit folgt, dass die kinetische Matrix näherungsweise dünnbesetzt ist und ohnehin konstante Kosten pro Matricelement hat.

### Kern-Elektron-Wechselwirkung

Wir betrachten (4.2.10b), setzen dabei aber nur die asymptotisch relevante Form (5.1.3a) ein, schätzen also die allgemeine Funktion aus  $\mathcal{K}$  nach oben ab, und wählen ein gemeinsames Zentrum  $P_{\mu\nu}$ , siehe Lemma 6,

$$\sum_{c=1}^{d_{\text{cor}}} V_{c,\mu\nu}^{\text{nucl}} = \sum_{c=1}^{d_{\text{cor}}} Z_c \int C_\mu \cdot \exp(-a_\mu \|\mathbf{r} - \mathbf{r}_\mu\|) \frac{1}{\|\mathbf{r} - \mathbf{R}_c\|} C_\nu \cdot \exp(-a_\nu \|\mathbf{r} - \mathbf{r}_\nu\|) d\mathbf{r},$$

dabei summieren wir über alle Kerne  $c$  des molekularen Systems.

Analog zu unserer Nebenrechnung aus dem Anhang A.3.2 setzen wir  $\chi := \mathbf{r} - P_{\mu\nu}$ ,  $\xi_c := \mathbf{R}_c - P_{\mu\nu}$  und  $r_1 := \|\xi_c\|$ ,  $r_2 := \|\chi\|$  und benutzen die Translationsinvarianz des Integrals,

$$\leq C \cdot \exp(-\alpha \|\mathbf{r}_\mu - \mathbf{r}_\nu\|) \left( \sum_{c=1}^{d_{\text{cor}}} Z_c \int \frac{1}{\|\chi - \xi_c\|} \exp(-a \|\mathbf{r} - P_{\mu\nu}\|) d\mathbf{r} \right).$$

Mit (A.3.11) folgt dann schließlich,

$$= C \cdot \exp(-\alpha \|\mathbf{r}_\mu - \mathbf{r}_\nu\|) \left( \sum_{c=1}^{d_{\text{cor}}} Z_c \frac{4\pi}{a^2 r_1} \int_0^{r_1} (1 + ar_2) \exp(-ar_2) dr_2 \right).$$

Das Integral können wir schließlich wieder durch (A.3.19) nach oben abschätzen,

$$\leq C \cdot \exp(-\alpha \|\mathbf{r}_\mu - \mathbf{r}_\nu\|) \left( \sum_{c=1}^{d_{\text{cor}}} Z_c \frac{8\pi}{a^3 r_1} \right).$$

Wir erhalten also einen exponentiellen Faktor, der mit dem Abstand zwischen den beiden Aufpunkten der Basisfunktionen abfällt, aber auch eine Summe über alle Kernladungen gewichtet mit dem inversen Abstand des Kerns zum gemeinsamen Aufpunkt.

Auch hier ergibt sich also eine näherungsweise dünnbesetzte Matrix, aber die Kosten pro Matrixelement sind  $\mathcal{O}(n)$ .

### Dichtematrix

Anhand von (4.3.3) und (4.3.4) ist offensichtlich, dass die Dichtematrix entscheidenden Einfluss auf die beiden letzten verbleibenden Matrixanteile hat.

Da die Lösungen asymptotisch lokal sind, siehe Abschnitt 5.5, können diese durch Funktionen aus der Menge der lokalen Funktionen effektiv, nämlich dünnbesetzt, dargestellt werden.

Betrachten wir dazu nochmal die Gleichung (4.2.7),

$$\mathbf{D}_{\mu\nu} = 2 \sum_i^{\frac{N}{2}} c_{\mu i} c_{i\nu}^*,$$

dann ist aber offensichtlich, dass bei näherungsweise dünnbesetzter Darstellung der Lösung mittels der Koeffizienten  $c_{\mu i}$  in der Basisfunktion  $\xi_\mu$  auch die Dichtematrix näherungsweise dünnbesetzt sein muss, siehe auch die Arbeit von Benzi et al. [18].

**Austausch-Wechselwirkung**

Betrachten wir dann Gleichung (4.2.10d) bzw. (4.3.3), dann wird klar, dass durch Lemma 7 gezeigt ist, dass die Matrix zwar im Allgemeinen voll besetzt ist, aber pro Matrixelement nur konstante Kosten aufgrund der exponentiellen Unterdrückung mit dem Abstand der Basisfunktionspaare entstehen.

Betrachten wir jedoch noch einmal (4.3.3),

$$\hat{K}_{\mu\nu} = \frac{1}{2} \sum_{\sigma\lambda} \mathbf{D}_{\sigma\lambda} \mathbf{I}_{\mu\sigma\lambda\nu}.$$

so fällt auf, dass mit einer näherungsweise dünn besetzten Dichtematrix die Austauschmatrix ebenso dünnbesetzt sein muss, siehe auch [132, S. 2217].

Damit haben wir also eine näherungsweise dünnbesetzte Austauschmatrix und auch nur konstante Kosten pro Matrixelement.

**Elektron-Elektron-Wechselwirkung**

Für die Coulomb-Wechselwirkung zwischen Elektronen, siehe (4.3.4), ergibt sich ein ähnliches Bild wie zuvor mit der Austausch-Wechselwirkung,

$$\hat{J}_{\mu\nu} = \sum_{\sigma\lambda} \mathbf{D}_{\sigma\lambda} \mathbf{I}_{\mu\nu\sigma\lambda},$$

aber nicht das gleiche. Denn die Indizes der Vier-Zentren-Integrale haben eine leicht andere Struktur, weswegen eine dünnbesetzte Dichtematrix ihre Wirkung auf das Indexpärchen entfaltet, welches durch einen entsprechenden Exponentialfaktor aus (5.4.5) ohnehin schon eingeschränkt ist.

Die Coulombmatrix der Elektronen ist also zwar näherungsweise dünnbesetzt, hat aber Kosten von  $\mathcal{O}(n)$  pro Matrixelement, genau wie die Matrix der Kern-Elektron-Wechselwirkung.

**Gesamt**

Wir halten abschließend fest, dass bei einer (näherungsweise) dünnbesetzten Dichtematrix das Aufstellen der Fockmatrix immer noch Kosten von  $\mathcal{O}(n^2)$  aufgrund der Coulomb-Wechselwirkung verursacht.

**7.1.2. Schwarz-Ungleichung**

Dass die Coulomb-Wechselwirkung aufgrund der Exponentialfaktoren sich nur quadratisch und nicht quartisch verhält, lässt sich nur ausnutzen, wenn durch ein Kriterium ohne eigentliche Berechnung des Vier-Zentren-Integrals erkennbar ist, dass es vernachlässigbar klein ist.

Solch ein Kriterium ergibt sich mit der Schwarz-Ungleichung, siehe [91, Abschn. 9.12.4]. Dabei wird benutzt, dass im Falle reellwertiger räumlichen Orbitale die Coulomb-Matrixelemente positiv definit sind und damit als inneres Produkt mit der Metrik  $\mathbf{r}_{12}^{-1}$  benutzt

werden können, siehe [87]. Es gilt nämlich  $I_{\mu\nu\mu\nu} > 0$  und damit folgt mit Einsetzen in die Schwarz-Ungleichung,

$$|I_{\mu\nu\sigma\lambda}| \leq \sqrt{I_{\mu\nu\mu\nu}} \sqrt{I_{\sigma\lambda\sigma\lambda}}. \quad (7.1.2)$$

Berechnet man also zunächst alle Diagonalelemente  $I_{\mu\nu\mu\nu}$ , dann lassen sich mit diesen anschließend alle verbleibenden Matrixelemente ihrem Betrag nach nach oben abschätzen und, falls sie kleiner als ein gesetzter Schwellwert sind, von der Berechnung ausschließen.

Verbesserte Schranken wurden zum Beispiel in Gill et al. [66] entwickelt.

Mittels dieses Screenings ergibt sich das von Strout und Scuseria [165] gemessene Skalierungsverhalten, welches wir auch in Abschnitt 4.3.3 beobachten konnten. Trotzdem bleibt die Komplexität mindestens quadratisch, nicht linear.

### 7.1.3. Abschneiden kleiner Elemente der Fockmatrix

Wir betrachten schließlich noch die Störung der nur näherungsweise dünnbesetzten Matrix  $\mathbf{F}$ , wenn wir sehr kleine Matrixelemente vernachlässigen und sie so zu einer dünnbesetzten Matrix  $\tilde{\mathbf{F}}$  machen. Dazu betrachten wir die Matrix

$$\tilde{\mathbf{F}} = \mathbf{F} - \delta\mathbf{F}. \quad (7.1.3)$$

Nehmen wir an, dass die elementweise Störung kleiner als ein  $\epsilon > 0$  ist, also

$$|(\delta\mathbf{F})_{\mu\nu}| < \epsilon \quad (7.1.4)$$

gilt.

Wir haben uns bereits mit gestörten hermiteschen Matrizen in Abschnitt 6.3 befasst. Nun wenden wir diese Störtheoreme an. Das entscheidende Maß für die Abschätzung der Störung der Eigenwerte ist die Frobenius-Norm  $\|\delta\mathbf{F}\|_F$ . Mit der elementweisen Abschätzung nach (7.1.4) folgt also

$$\|\delta\mathbf{F}\|_F < n^2 \cdot \epsilon, \quad (7.1.5)$$

da die Matrix  $\delta\mathbf{F}$  voll besetzt sein muss, wenn  $\mathbf{F}$  dünnbesetzt werden soll.

Da die hierbei vernachlässigten Matrixelemente ausschließlich jene sind, welche stark exponentiell gedämpft sind, wie zuvor hergeleitet, stellen wir abschließend fest, dass bei entsprechender Wahl von  $\epsilon$  die Eigenwerte kaum gestört werden und wir damit im Folgenden mit einer gestörten, dünnbesetzten Fockmatrix weiter verfahren dürfen.

## 7.2. Dünnbesetzte Eigenvektoren

Wir widmen uns nun dem zweiten Block, der Diagonalisierung. Dazu wollen wir das Schwerpunkttheorem anwenden, welches uns eine Berechnung der Eigenwerte und -vektoren aus Hauptuntermatrizen ermöglicht, wenn diese Eigenvektoren dünn besetzt und ihre Struktur im Rahmen einer gegebenen Zerlegung bekannt sind.

### 7.2.1. Lokalisierte Eigenwertgleichung

Wir haben zuvor die Fockmatrix betrachtet. Die zugehörige Eigenwertgleichung liefert die räumlichen Orbitale als Eigenfunktionen zu negativen Eigenwerten.

Bevor wir jedoch die Erkenntnisse aus Kapitel 6 anwenden können, müssen wir die Voraussetzungen klären, ob die Eigenvektoren dünnbesetzt sind. Die kanonischen Orbitale sind im Allgemeinen nämlich gerade nicht lokalisiert und damit nicht effizient durch lokale Funktionen dargestellt. Erst durch Lokalisierung entstehen räumlichen Orbitale, die chemischer Intuition zugänglich sind, siehe [132]. Diese lokalisierten Orbitale wiederum sind aber keine Eigenfunktionen der Fockmatrix mehr, siehe Paragraph „Spaltentransformation“ in Abschnitt 4.2.3.

Ein Ausweg wurde durch Gilbert, siehe [50], aufgezeigt, indem man einen effektiven Operator wie folgt definiert,

$$\hat{F}^L \phi_i^L = [\hat{F} - \rho \hat{F} \rho + \rho \hat{L} \rho] = \lambda_i \phi_i^L, \quad (7.2.1)$$

wobei  $\hat{L}$  der Lokalisierungsoperator ist. Dabei hat die Dichte  $\rho$  die Eigenschaft eines Projektors in den Raum der besetzten Zustände bzw. der räumlichen Orbitale  $\phi_i$  mit  $\epsilon_i < 0$ . Die lokalisierten Orbitale  $\phi_i^L$  sind dabei Eigenfunktionen von sowohl  $\hat{F}^L$  als auch  $\hat{L}$ .

Die genaue Form des Lokalisierungsoperator ist dabei nicht bekannt, gleichwohl es in der Literatur verschiedene Vorgehensweisen gibt, siehe [60, 50, 143]. Er lässt sich aber über seine Spektraldarstellung teilweise angeben, denn mit

$$\phi_i^L = \hat{U}^\dagger \phi_i$$

schreiben wir

$$\hat{L} = \sum_i \lambda_i \phi_i^L \phi_i^{L*} = \sum_i \lambda_i \hat{U}^\dagger \phi_i \hat{U} \phi_i^*. \quad (7.2.2)$$

Dabei wird die unitäre Transformation  $\hat{U}$ , welche die Menge der besetzten Zustände  $\{\phi_i\}_{i=1}^N$  lokalisiert, mittels einer der bekannten Methoden bestimmt. Unbekannt verbleiben einzig die Eigenwerte  $\lambda_i$ .

Betrachtet man die Eigenfunktionen von  $\hat{F}^L$  in (7.2.1), in einer endlichen Basis der Größe  $n$  entwickelt, genauer, so wird klar, dass in der Menge der Eigenfunktionen von  $\hat{F}$ , nämlich  $\{\phi_1, \dots, \phi_n\}$ , die besetzten Zustände durch die lokalisierten ausgetauscht worden sind, nämlich  $\{\phi_1^L, \dots, \phi_N^L, \phi_{N+1}^L, \dots, \phi_n\}$ , wie man sich durch Einsetzen schnell klar macht. Dabei müssen wir beachten, dass die Dichte invariant gegenüber einer unitären Transformation ist,

$$\rho = \sum_{i=1}^N \phi_i \phi_i^* = \sum_{i=1}^N \phi_i^L \hat{U} \hat{U}^\dagger \phi_i^L = \sum_{i=1}^N \phi_i^L \phi_i^L.$$

Nun wollen wir endlich die Erkenntnisse aus Kapitel 6 anwenden. Stellen wir uns dazu (7.2.1) analog zur Roothaan-Hall-Gleichung in einer Basis entwickelt vor. Anschließend

betrachten wir die Zerlegung der zugehörigen Matrix  $\mathbf{F}^L = \mathbf{N}^L + \mathbf{P}^L$  in seinen negativ definiten Anteil  $\mathbf{N}^L = \mathbf{L}$  und positiv definiten Anteil  $\mathbf{P}^L = \mathbf{F} - \rho\mathbf{F}\rho$ .

Dabei wissen wir, dass die Eigenfunktionen von  $\mathbf{L}$  allesamt dünnbesetzt sind, weil es ja genau die lokalisierten Orbitale sind. Ihr exponentieller Abfall wird bspw. anhand von Alkanen in [132] untersucht. Anschließend können wir also das Schwerpunkttheorem anwenden und wissen, dass wir, eine geeignete Zerlegung der Indexmenge vorausgesetzt, die Matrix  $\mathbf{N}^L$  aus den Eigenvektoren der Hauptuntermatrizen darstellen können.

Anstelle jedoch die Diagonalisierung der Hauptuntermatrizen  $\mathbf{L}_i$  des unbekanntem Lokalisierungsoperators zu betreiben, diagonalisieren wir stattdessen die Hauptuntermatrizen der Fockmatrix  $\mathbf{F}_i$  und würden anschließend die dabei resultierenden Orbitale lokalisieren. Wir bemerken, dass wenn die Hauptuntermatrizen ausreichend klein sind, dann diese Lokalisierung entfallen kann, denn in der vollen Basis der Fockmatrix sind diese Lösungen automatisch dünnbesetzt dargestellt.

Für die Transferabilität dieser lokalisierten Orbitale vom einem Teilsystem in das volle System existieren bisher jedoch nur empirische Belege, siehe bspw. [156], wo ebenfalls ein Fragmentansatz für die Berechnung der Energie vorgestellt wird. Der direkte Zusammenhang nach Lemma 10 zwischen den Eigenvektoren der Hauptuntermatrizen  $\mathbf{F}_i$  und der Fockmatrix  $\mathbf{F}$  ist aufgrund der voll besetzten kanonischen Orbitale verwehrt. Aus diesem Grund müssen wir den Umweg über (7.2.1) nehmen.

Analog existieren Lokalisierungsmethoden für die virtuellen Orbitale, siehe [166], und eine weitere Anpassung ist gleichfalls möglich, so dass der effektive Fockoperator nur noch über lokalisierte Eigenfunktionen verfügt, egal ob zu negativen oder positiven Eigenwerten.

Die unbekanntem Eigenwerte in der Spektraldarstellung der beiden Lokalisierungsoperatoren spielt dabei keine Rolle, wie wir gleich sehen werden. Denn wir benötigen nur die Matrix, welche in den negativen Eigenraum projiziert. Die Eigenwerte erhalten wir dann über diese Matrix.

### 7.2.2. Berechnung der Gesamtenergie

Mittels der zuvor definierten effektiven Fockmatrix können wir das Schwerpunkttheorem 2 bzw. die Proposition 2 für den Projektor,

$$\mathbf{D} = \mathbf{P}^- = \sum_{\mathcal{I}_j \in \mathcal{J}} \tilde{\mathbf{D}}_i \quad \text{mit} \quad \tilde{\mathbf{D}}_i = \mathbf{D}_i - \sum_{\mathcal{I}_j \subseteq \mathcal{I}_i} \tilde{\mathbf{D}}_j, \quad (7.2.3)$$

anwenden. Damit beschreibt diese Dichtematrix also den Projektor in den Raum der besetzten Zustände aus den lokalisierten Zuständen der effektiven Fockmatrix.

Damit betrachten wir nochmal die Gesamtenergie des HF-Problems, siehe (4.1.17),

$$E^{(\text{SCF})} = \sum_i^{\frac{N}{2}} h_i + \epsilon_i.$$

## 7. BOSSANOVA

Die Summe über die Eigenwerte  $\epsilon_i$  erhalten wir dabei aus den Teilproblemen, aber wir müssen zusätzlich

$$h_i = \int \phi_i^*(\mathbf{r}_1) \hat{h} \phi_i(\mathbf{r}_1) d\mathbf{r}_1$$

auswerten.

Mittels des Projektors gelingt dies wie folgt,

$$\begin{aligned} E^{(\text{SCF})} &= \text{Tr} (\mathbf{D}(\mathbf{F} + \mathbf{h})) \\ &= \text{Tr} \left( \sum_{\mathcal{I}_j \in \mathcal{J}} \tilde{\mathbf{D}}_i(\mathbf{F} + \mathbf{h}) \right) \end{aligned}$$

Nach Definition der Differenzbeiträge  $\tilde{\mathbf{D}}_i$  zum Projektor betrachten wir dabei stets Terme der Art  $\mathbf{D}_j(\mathbf{F} + \mathbf{h})$ . Da sowohl  $\mathbf{F}$  als auch  $\mathbf{h}$  näherungsweise dünnbesetzte Matrizen von ähnlicher Struktur sind, siehe die numerische Untersuchung in [18] anhand von Alkanen, können wir uns auf die Form

$$E^{(\text{SCF})} \approx \text{Tr} \left( \sum_{\mathcal{I}_j \in \mathcal{J}} \tilde{\mathbf{D}}_i(\mathbf{F}_i + \mathbf{h}_i) \right) \quad (7.2.4)$$

zurückziehen. Danach können wir die Summe aus der Spur herausziehen und die Terme  $\mathbf{D}_i \mathbf{F}_i$  und  $\mathbf{D}_i \mathbf{h}_i$  werden bereits in den Hauptuntermatrizen ausgewertet. Sie stellen eigenständige HF-Probleme dar. Also können wir schreiben

$$E^{(\text{SCF})}_i = \text{Tr} (\mathbf{D}_i(\mathbf{F}_i + \mathbf{h}_i)) \quad (7.2.5)$$

die Energie zur Hauptuntermatrix  $\mathbf{F}_i$ .

Damit erhalten wir die Gesamtenergie  $E^{(\text{ANOVA})}$  aus der vorgestellten Methode über die ANOVA-artige Summe, siehe Abschnitt 2.5.3,

$$E^{(\text{SCF})} \approx E^{(\text{ANOVA})} = \sum_{\mathcal{I}_i \in \mathcal{J}} \widetilde{E^{(\text{ANOVA})}_i} \quad (7.2.6a)$$

$$\widetilde{E^{(\text{ANOVA})}_i} := E^{(\text{SCF})}_i - \sum_{\mathcal{I}_j \subsetneq \mathcal{I}_i} \widetilde{E^{(\text{ANOVA})}_j} \quad (7.2.6b)$$

aus den Energien der Teilprobleme  $E^{(\text{SCF})}_i$ .

Schließlich bemerken wir noch, dass sich die Grundzustandsdichte (4.2.4) aus den Dichten der Teilprobleme analog zu (7.2.3) ermitteln lässt.

$$\rho = \sum_i \tilde{\rho}_i, \quad (7.2.7)$$

$$\tilde{\rho}_i = \rho_i - \sum_{j, \mathcal{I}_j \subsetneq \mathcal{I}_i} \tilde{\rho}_j. \quad (7.2.8)$$



### 7.3. Behandlung der langreichweitigen Wechselwirkung

Wie im Abschnitt 7.1 festgestellt, hat das Berechnen jedes Matrixelements nicht ausschließlich konstante Kosten, sondern für die Coulombmatrizen ergeben sich Kosten *pro Matrixelement* der Ordnung  $\mathcal{O}(n)$  durch die Wechselwirkung jedes Basisfunktionspaars mit  $M$  Kernen bzw.  $N - 1$  anderen Elektronen, deren Darstellung der Ladungsdichte proportional zur Anzahl an Basisfunktionen ist.

Wir werden im Folgenden den Coulomb-Anteil pro Hauptuntermatrix der Fockmatrix so auftrennen, dass zwei Anteile entstehen. Ein Anteil soll dabei konstante Kosten beim Aufstellen der Hauptuntermatrix verursachen, der zweite linear mit der Anzahl an Basisfunktionen steigende Kosten pro Matrixelement.

Als ersten Schritt kombinieren wir allgemein das Coulomb-Potential von Elektronen und Kernen zu einem gemeinsamen Ausdruck. Die beiden Ladungsverteilungen sind mit unterschiedlichem Vorzeichen behaftet und können sich daher zu einem gewissen Grad gegenseitig aufheben. Dazu schreiben wir die Ladungsverteilung als eine Summe über die Punktladungen der Kerne und über die elektronische Ladungsverteilung, siehe auch die Abschnitte 3.3.2 über das Poisson-Problem und 4.2.2 über die Dichtematrix,

$$\rho(\mathbf{r}) := - \sum_{\sigma\lambda} \mathbf{D}_{\sigma\lambda} \Omega_{\sigma\lambda}(\mathbf{r}) + \sum_{a=1}^M Z_a \delta(\mathbf{r} - \mathbf{R}_a) \quad (7.3.1)$$

mit der verkürzenden Schreibweise für ein Basisfunktionspaar

$$\Omega_{\sigma\lambda}(\mathbf{r}) := \xi_{\sigma}^*(\mathbf{r}) \xi_{\lambda}(\mathbf{r}). \quad (7.3.2)$$

Damit können wir nun Matrixelemente definieren, welche beide Wechselwirkungen kombinieren, siehe auch (4.2.10b) und (4.2.10c),

$$\begin{aligned} V_{\mu\nu}^{\text{Coulomb}} &= V_{\mu\nu}^{(\text{core,el})} + J_{\mu\nu} \\ &= \iint \Omega_{\mu\nu}(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \rho(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2. \end{aligned} \quad (7.3.3)$$

Wir bemerken, dass die Delta-Distribution in (7.3.1) nur innerhalb des Integranden im Ausdruck (7.3.3) wohldefiniert ist.

#### Elektrostatistisches Potential

Im Folgenden halten wir uns an die Darstellung in Griebel et al. [77, Kap. 7].

Wir wissen, dass das Coulomb-Potential  $\Phi(\mathbf{x})$  an einem Ort  $\mathbf{x}$  einer Ladungsverteilung  $\rho$ , siehe Jackson [106, Abschn. 1.5],

$$\Phi(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int \rho(\mathbf{y}) \frac{1}{\|\mathbf{y} - \mathbf{x}\|} d\mathbf{y} \quad (7.3.4)$$

ist, wobei  $\epsilon_0$  die Dielektrizitätskonstante des Vakuums bezeichnet, siehe [106, S. 31].

## 7. BOSSANOVA

Die elektrostatische Energie  $V^{\text{Coulomb}}$  einer (beliebigen anderen) Ladungsverteilung  $\rho'(\mathbf{x})$  ergibt sich dann durch

$$V^{\text{Coulomb}} = \int \Phi(\mathbf{x})\rho'(\mathbf{x})d\mathbf{x}. \quad (7.3.5)$$

Dass die Ausdrücke (7.3.4) und (7.3.5) existieren, hängt dabei von den Eigenschaften der Ladungsverteilungen ab. An dieser Stelle werden wir die Existenz des Potentials, physikalischen Gegebenheiten folgend, schlicht voraussetzen.

Kennen wir nämlich den Term  $\Phi(\mathbf{r})$ , so gelingt die Berechnung des kombinierten Matrixelements (7.3.3) wie folgt,

$$\begin{aligned} V_{\mu\nu}^{\text{Coulomb}} &= \iint \Omega_{\mu\nu}(\mathbf{r}_1) \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \rho(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \int \Omega_{\mu\nu}(\mathbf{r}) \Phi(\mathbf{r}) d\mathbf{r}, \end{aligned}$$

also mit konstanten Kosten pro Matrixelement, da für das Basisfunktionspaar  $\Omega_{\mu\nu}$  selbst wieder ein exponentieller Faktor wirksam ist.

### Kurz- und langreichweitige Aufspaltung

Um das Potential  $\Phi$  zu berechnen, müssten wir jedoch die Gesamtladungsdichte kennen. Wie im letzten Kapitel erläutert, ist es jedoch unser Ziel, nur Hauptuntermatrizen  $\mathbf{F}_i$  der Fockmatrix zu einer Indexmenge  $\mathcal{I}_i \in \mathcal{J}$  zu diagonalisieren.

Wir setzen deswegen in die Ladungsverteilung (7.3.1) die ANOVA-artige Summe (7.2.3) der Dichtematrix ein und zerlegen diese Summe in einen kurzreichweitigen Anteil  $\rho_i^{(sr)}$  und einen langreichweitigen Anteil  $\rho_i^{(lr)}$  auf,

$$\begin{aligned} \rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}}(\mathbf{r}) &:= - \sum_{\mu\nu \in \mathcal{I}_i} \sum_{\mathcal{I}_j \subseteq \mathcal{I}_i} \left( \tilde{\mathbf{D}}_j \right)_{\mu\nu} \Omega_{\mu\nu}(\mathbf{r}) + \sum_{c \in \mathcal{U}_i} Z_c \delta(\mathbf{r} - \mathbf{R}_c), \\ &= - \sum_{\mu\nu \in \mathcal{I}_i} (\mathbf{D}_i)_{\mu\nu} \Omega_{\mu\nu}(\mathbf{r}) + \sum_{c \in \mathcal{U}_i} Z_c \delta(\mathbf{r} - \mathbf{R}_c), \end{aligned} \quad (7.3.6)$$

$$\rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{lr}}(\mathbf{r}) := - \sum_{\mu, \nu} \sum_{\mathcal{I}_j \not\subseteq \mathcal{I}_i} \left( \tilde{\mathbf{D}}_j \right)_{\mu\nu} \Omega_{\mu\nu}(\mathbf{r}) + \sum_{c \notin \mathcal{U}_i} Z_c \delta(\mathbf{r} - \mathbf{R}_c). \quad (7.3.7)$$

Hier tauchen hier nun zwei Mengen mit gemeinsamem Index  $i$  auf, die Mengen der Basisfunktionsindizes  $\mathcal{I}_i$  und die Menge der Kernindizes  $\mathcal{U}_i$ . Wir gehen hier aber nicht weiter auf diese zweite Indexmenge  $\mathcal{U}_i$  ein, sondern setzen sie, wie auch  $\mathcal{I}_i$ , als gegeben voraus.

Damit ergeben sich zwei entsprechende Anteile des Coulomb-Potentials (7.3.4),

$$\Phi_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}}(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int \rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}}(y) \frac{1}{\|y - \mathbf{x}\|} dy, \quad (7.3.8a)$$

$$\Phi_{\mathcal{U}_i, \mathcal{I}_i}^{\text{lr}}(\mathbf{x}) = \frac{1}{4\pi\epsilon_0} \int \rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{lr}}(y) \frac{1}{\|y - \mathbf{x}\|} dy. \quad (7.3.8b)$$

### 7.3. Behandlung der langreichweitigen Wechselwirkung

Wir bemerken, dass natürlich  $\rho = \rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}} + \rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{lr}}$  für jedes Indexmengenpaar  $\mathcal{I}_i, \mathcal{U}_i$  gilt und damit wegen der Linearität des Integrals auch

$$\Phi = \Phi_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}} + \Phi_{\mathcal{U}_i, \mathcal{I}_i}^{\text{lr}}. \quad (7.3.9)$$

Dabei können wir den langreichweitigen Anteil  $\Phi_{\mathcal{U}_i, \mathcal{I}_i}^{\text{lr}}$  nicht direkt berechnen, während wir die zugehörige Hauptuntermatrix  $\mathbf{F}_i$  diagonalisieren, da uns zunächst nur die Ladungsverteilung  $\rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}}$  zur Verfügung steht.

Mit allen kurzreichweitigen Anteilen  $\rho_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}}$  zusammen kennen wir aber die Ladungsverteilung  $\rho$  aus der Dichtematrix (7.2.7). Damit können wir das Coulomb-Potential  $\Phi$  über (7.3.4) berechnen. Ermitteln wir anschließend mit (7.3.8a) das kurzreichweitige Coulomb-Potential  $\Phi_{\mathcal{U}_i, \mathcal{I}_i}^{\text{sr}}$  der Ladungsverteilung zu den Indexmenge  $\mathcal{I}_i$  bzw.  $\mathcal{U}_i$ , so erhalten wir durch Differenz zum Gesamtpotential, siehe (7.3.9), schließlich den vernachlässigten langreichweitigen Matrixanteil  $\mathbf{F}_{\mathcal{U}_i, \mathcal{I}_i}^{(\text{lr})}$  mittels (7.3.3) für jede Hauptuntermatrix.

Bei dieser Vorgehensweise müssen die Coulomb-Potentiale in jedem Iterationsschritt vor dem Aufstellen der jeweiligen Hauptuntermatrix der Fockmatrix gelöst werden. Linear skalierende Verfahren für das Poisson-Problem sind bekannt, siehe [9]. An der nicht-linearen Iteration selbst ändert sich aber zunächst nichts.

Insgesamt erhalten wir also ein vor allem schnelles<sup>1</sup> und bis auf Näherungen im Rahmen der Maschinengenauigkeit exaktes Matrixaufstellungs- und Diagonalisierungsverfahren, welches linear mit der  $n$  skaliert, wenn mit der Zerlegung nur  $\mathcal{O}(n)$  viele Teilprobleme zusammenhängen. Andere sogenannte dichte-basierte HF-Verfahren sind ebenso linear skalierend, jedoch mit schlechterer Konstante, siehe auch den noch folgenden Abschnitt 7.6.

#### Kern-Kern-Wechselwirkung

In der Energie  $E^{(\text{SCF})}$ , welche wir aus der Lösung der HF-Gleichung erhalten, fehlt die Kern-Kern-Wechselwirkung, da sie nicht Teil des Elektron-Hamiltonoperator ist, siehe Abschnitt 3.3. Die Zerlegung der Kernindizes  $\{\mathcal{U}_i\}$  bedingt auch eine Aufspaltung dieser Wechselwirkung in einen kurzreichweitigen Anteil der Kernpaare zu Indizes in  $\mathcal{U}_i$  und restlichen langreichweitigen Wechselwirkung.

Da diese Wechselwirkung aber nicht von der Ladungsverteilung der Elektronen abhängt, betrachten wir sie als ein getrenntes Problem. Die naive paarweise Auswertung des zugehörigen Terms

$$\sum_{1 \leq c < d \leq M} \frac{Z_c Z_d}{\|\mathbf{R}_c - \mathbf{R}_d\|}$$

kann bei der klassischen HF-Methode durchgeführt werden, da mit ihr ja nur Kosten von  $\mathcal{O}(n^2)$  verbunden sind. Aber auch hier kann schlicht auf eine der bekannten Verfahren zu Berechnung in linearer Komplexität, siehe [9], zurück gegriffen werden. Aus diesem Grund gehen wir auf diesen Teil der Berechnung an dieser Stelle nicht weiter ein.

<sup>1</sup>Die obere Schranke der Komplexität  $\mathcal{O}(n)$  versteckt dabei die vorhandene Konstante, welche zu einem sogenannten Crossover-Punkt führt. Damit ist die Problemgröße  $M$  bezeichnet, aber der ein linear skalierendes Verfahren auch absolut schneller als das Standard-HF-Verfahren löst.

## 7.4. Gestörter Fockoperator

Wir haben uns soweit mit den beiden eingangs des Kapitels erwähnten Blöcken Aufstellung und Diagonalisierung der Fockmatrix beschäftigt. Dabei lösten wir die verbliebenen quadratischen Kosten in der Aufstellung der Matrix durch eine gesonderte Behandlung der langreichweitigen Coulomb-Wechselwirkung mittels linear skalierender Verfahren. Die Kosten der Diagonalisierung verringern wir durch Betrachtung von Hauptuntermatrizen im Rahmen einer gegebenen Zerlegung der Indexmenge der Basisfunktionen. Soweit beschreibt dies aber nur eine schnelle Lösung des Eigenwertproblems, anwendbar für die einzelnen Schritte der SCF-Iteration.

Wir wollen nun die Fockmatrix so stören, dass die nicht-linearen Kopplungen, welche durch die Abhängigkeit der Matrix von der Lösung in Form der Dichtematrix entsteht, nicht mehr global, sondern nur noch lokal wirken. Auf diese Weise trennen wir die Iterationsschleife in lauter Schleifen über unabhängige Teilprobleme auf, die in nullter Ordnung Störtheorie nicht weiter miteinander wechselwirken. Dazu werden wir formal eine Abbildung von der Dichtematrix auf die Fockmatrix definieren und diese Abbildung stören.

Wir haben die Coulomb-Wechselwirkung bereits formal in einen kurz- und einen langreichweitigen Anteil aufgetrennt. Seien dazu eine Zerlegung  $\mathcal{J}$  der Indexmenge  $\mathcal{N} := \{1, \dots, n\}$  der Basisfunktionsindizes, sowie eine Zerlegung  $\mathcal{U}$  der Menge der Kernindizes  $\mathcal{M} := \{1, \dots, M\}$  gegeben, welche zu den Indexuntermengen  $\mathcal{I}_i \in \mathcal{J}$  und  $\mathcal{U}_i \in \mathcal{U}$  jeweils die Fockmatrix in einen kurzreichweitigen Anteil  $\mathbf{F}_i^{(sr)}$  und einen langreichweitigen Anteil  $\mathbf{F}_i^{(lr)}$  aufspalten, so dass stets  $\mathbf{F}_i = \mathbf{F}_i^{(sr)} + \mathbf{F}_i^{(lr)}$  gilt, wobei  $\mathbf{F}_i$  die uns schon bekannte Hauptuntermatrix der Fockmatrix zu obigen Indexuntermengen darstellt.

Wir müssen anschließend sicherstellen, dass die Störung durch Weglassen des Anteils  $\|\mathbf{F}_i^{(lr)}\|$  klein ist.

### 7.4.1. Fockmatrix-Abbildung

Wir fassen die Fockmatrix formal als eine lineare Abbildung  $\hat{\varphi}$  vom Raum der hermiteschen Dichtematrizen  $\mathbb{C}^{n \times n}$  auf die hermiteschen Matrizen  $\mathbb{C}^{n \times n}$  auf,  $\hat{\varphi}[\mathbf{D}] = \mathbf{F}$ . Diese Fockabbildung  $\hat{\varphi}$  schreiben wir wie folgt auf: einen konstanten Anteil  $\mathbf{h}$  und einen mit der Dichtematrix variablen Anteil  $\hat{V}$ ,

$$\hat{\varphi}[\mathbf{D}] := \mathbf{h} + \hat{V}[\mathbf{D}], \quad (7.4.1)$$

wobei  $\mathbf{h}$  die Fockmatrixanteile der kinetischen Energie und der Kern-Elektron-Wechselwirkung repräsentiert und  $\hat{V}$  die der Austausch- und Elektron-Elektron-Wechselwirkung hinzufügt. Erinnern wir uns an die Gestalt der Ausdrücke (4.3.3) und (4.3.4), welche die Vier-Zentren-Integrale beinhalten und die von der Dichtematrix abhängigen Teile der Fockmatrix darstellen, so ist diese Abbildung linear.

Wir können diese Abbildung auch formal für die Hauptuntermatrizen der Fockmatrix aufschreiben. Dazu sei zu einem Indexmengenpaar  $\mathcal{I}_i \in \mathcal{J}, \mathcal{U}_i \in \mathcal{U}$

$$\hat{\varphi}_i[\mathbf{D}] := \mathbf{P}_i \mathbf{h} \mathbf{P}_i + \mathbf{P}_i \hat{V}[\mathbf{D}] \mathbf{P}_i. \quad (7.4.2)$$

Wir bemerken, dass hier nicht die Dichtematrix  $\mathbf{D}_i$  zur Hauptuntermatrix, sondern die volle Matrix  $\mathbf{D}$  eingeht. Dies stellt die eingangs bezeichnete globale Kopplung in der selbst-konsistenten Iteration noch einmal dar: das Teilproblem in Form der Hauptuntermatrix  $\mathbf{F}_i$ , welches wir lösen können, wie in dem vorigen Abschnitt beschrieben, ist stets abhängig von der kombinierten Lösung aller Hauptuntermatrizen des letzten Iterationsschritts in Form der Dichtematrix  $\mathbf{D}$ , siehe Gleichung (7.2.3).

Deswegen trennen wir nun diese Abbildung mittels der ANOVA-artigen (7.2.3) Summe der Dichtematrix in zwei Anteile analog zu (7.3.6), (7.3.7) auf, einen kurzreichweitigen und einen langreichweitigen Anteil,

$$\hat{\varphi}_i[\mathbf{D}] := \hat{\varphi}_i^{(sr)}[\mathbf{D}_i] + \hat{\varphi}_i^{(lr)}[\mathbf{D}], \quad (7.4.3)$$

wobei die kurzreichweitige Abbildung  $\hat{\varphi}_i[\mathbf{D}_i]^{(sr)}$  nur noch von der Dichtematrix  $\mathbf{D}_i$  der Hauptuntermatrix abhängt.

Schreiben wir schließlich die beiden Anteile explizit auf, wobei wir uns auf das Matrixelement zum Basisfunktionsindexpaar  $\mu, \nu \in \mathcal{I}_i$  beschränken, sprich nur auf den mit  $\mathcal{I}_i$  assoziierten Block der Fockmatrix,

$$\left(\hat{\varphi}_i^{(sr)}[\mathbf{D}_i]\right)_{\mu\nu} = T_{\mu\nu} + \sum_{c \in \mathcal{U}_i} V_{c,\mu\nu}^{\text{nucl}} + \sum_{\sigma, \lambda \in \mathcal{I}_i} (\mathbf{D}_i)_{\sigma\lambda} I_{\mu\nu\sigma\lambda} - \frac{1}{2} \sum_{\sigma, \lambda \in \mathcal{I}_i} (\mathbf{D}_i)_{\sigma\lambda} I_{\mu\sigma\lambda\nu}, \quad (7.4.4a)$$

$$\left(\hat{\varphi}_i^{(lr)}[\mathbf{D}]\right)_{\mu\nu} = \sum_{c \notin \mathcal{U}_i} V_{c,\mu\nu}^{\text{nucl}} + \sum_{\sigma, \lambda} \sum_{\mathcal{I}_j \not\subseteq \mathcal{I}_i} \left[ \left(\tilde{\mathbf{D}}_j\right)_{\sigma\lambda} I_{\mu\nu\sigma\lambda} - \frac{1}{2} \left(\tilde{\mathbf{D}}_j\right)_{\sigma\lambda} I_{\mu\sigma\lambda\nu} \right]. \quad (7.4.4b)$$

Dabei bezeichnet  $V_{c,\mu\nu}^{\text{nucl}}$  die Wechselwirkung des Basisfunktionspaars mit Kern  $c$ .

Der langreichweitige Anteil  $\hat{\varphi}_i^{(lr)}$  hängt immer noch von der vollen Dichtematrix  $\mathbf{D}$  ab. Der kurzreichweitige Anteil  $\hat{\varphi}_i^{(sr)}$  hingegen nicht mehr.

Im Anschluss widmen wir uns zunächst der Frage, wie groß der langreichweitige Anteil  $\hat{\varphi}_i^{(lr)}$  an der sich ergebenden Fockmatrix ist und ob er als Störung betrachtet werden kann.

### 7.4.2. Störung der Fockmatrix

Das Vernachlässigen von Anteilen in der Fockabbildung in nullter Ordnung Störtheorie verändert die Dichtematrix, welche die Methode pro Iterationsschritt liefert. Wir wollen nun abschätzen, wie stark diese dabei verändert wird, indem wir die Störung des Spektrums der aus der Abbildung resultierenden Matrix untersuchen. Dazu vergleichen wir  $\mathbf{F}_i^{(sr)} = \hat{\varphi}_i[\mathbf{D}_i]^{(sr)}$  aus (7.4.4a) mit  $\mathbf{F}_i$ , siehe (7.4.1).

Dabei müssen wir Störungen in zwei Formen unterscheiden: Da die Fockmatrix linear von der Dichtematrix  $\mathbf{D}$  abhängt, erfährt sie einerseits eine Störung, wenn die Dichtematrix gestört wird. Andererseits besteht die Fockmatrix aber aus mehreren Teilen. Vernachlässigen wir bspw. die Kern-Elektron-Wechselwirkung im konstanten Anteil  $\mathbf{h}$  zu einem gewissen Grad, dann stellt dies eine zweite Form der Störung dar. Wir hatten den Austauschanteil darin schon in Abschnitt 7.1.3 behandelt.

## 7. BOSSANOVA

Damit schreiben wir Störungen der Fockmatrix als Summe der beiden obigen Arten auf,

$$\mathbf{F} + \delta\mathbf{F} = \mathbf{h} + \delta\mathbf{h} + \hat{V}[\mathbf{D} + \delta\mathbf{D}] + \delta\hat{V}[\mathbf{D} + \delta\mathbf{D}], \quad (7.4.5)$$

und identifizieren die Störungsmatrix als

$$\delta\mathbf{F} := \delta\mathbf{h} + \hat{V}[\delta\mathbf{D}] + \delta\hat{V}[\mathbf{D} + \delta\mathbf{D}]. \quad (7.4.6)$$

Schreiben wir nun konkret die einzelnen Anteile der Störungen  $\delta\mathbf{F}_i$  zu einer Hauptuntermatrix für die obige Aufteilung in kurz- (7.4.4a) und langreichweitige Anteile (7.4.4b) auf,

$$\delta\mathbf{h}_i := \sum_{c \notin \mathcal{U}_i} V_{c,\mu\nu}^{\text{nucl}} \quad (7.4.7a)$$

$$\hat{V}_i[\delta\mathbf{D}] := \frac{1}{2} \sum_{\sigma,\lambda} (\delta\mathbf{D})_{\sigma\lambda} \mathbf{I}_{\mu\sigma\lambda\nu} + \sum_{\sigma,\lambda} (\delta\mathbf{D})_{\sigma\lambda} \mathbf{I}_{\mu\nu\sigma\lambda} \quad (7.4.7b)$$

mit

$$\delta\mathbf{D} = \sum_{\mathcal{I}_j \notin \mathcal{I}_i} (\tilde{\mathbf{D}}_j),$$

wobei alle anderen Störteile gleich null sind<sup>2</sup>.

Mit den Aussagen zur Veränderung des Spektrums einer hermiteschen Matrix unter additiver Störung, siehe Abschnitt 6.3, wissen wir, dass wir die Aufspaltung in kurz- und langreichweitige Wechselwirkung so zu wählen haben, dass sich die Matrixelemente möglichst wenig im Sinne der Frobenius-Norm ändern.

### 7.4.3. Multipolentwicklung

Um also weitere Aussagen über die Größe der Störung treffen zu können, müssen wir uns zunächst die Wirkung der Indexmenge  $\mathcal{I}_i$  im Zusammenspiel mit der Indexmenge  $\mathcal{U}_i$  verdeutlichen, um den verbleibenden Coulomb-Anteil zu verstehen.

Dazu führen wir eine Multipolentwicklung von (7.3.3) durch. Da die höheren Momente der Entwicklung sich wie  $\mathcal{O}(r^{-l})$  verhalten, wobei  $l$  die Ordnung bezeichnet und  $r$  den Abstand, konzentrieren wir uns auf die ersten Momente. Solch eine Entwicklung wurde bereits in einer Behandlung ohne Störansatz benutzt, siehe [184] und auch [91, Abschn. 9.14] für eine allgemeine Einführung.

Seien zwei Punkte  $P, Q \in \mathbb{R}^3$  gegeben. Wir schreiben die beiden Orte im Integranden in (7.3.3) relativ zu diesen auf,  $\mathbf{r}_1 = \mathbf{r}_{1P} + P$  und  $\mathbf{r}_2 = \mathbf{r}_{2Q} + Q$ , und definieren  $r_{12} = \|\mathbf{r}_{1P} - \mathbf{r}_{2Q}\|$  und  $r_{QP} = \|Q - P\|$ . Dann führen wir eine Multipolentwicklung des Coulomb-Kerns durch, siehe [91, Abschn. 9.13],

$$\frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} \mathcal{R}_{lm}(r_{12}) \mathcal{L}_{lm}^*(r_{QP}) \quad (7.4.8)$$

<sup>2</sup>Wir gehen von keiner Störung der Basisfunktionen selbst aus, welche zu  $\delta\hat{V} \neq 0$  führen würde

mit den sogenannten Regulären Kugelflächenfunktionen  $\mathcal{R}_{lm}$  und Irregulären Kugelflächenfunktionen  $\mathcal{I}_{lm}$  und setze diese in (7.3.3) ein,

$$V_{\alpha\beta}^{\text{Coulomb}} = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{+l'} q_{lm}^{\alpha\beta} T_{lm,l'm'}(r_{QP}) q_{l'm'}^{\rho}, \quad (7.4.9)$$

wobei wir die sphärischen Multipolmomente

$$q_{lm}^{\alpha\beta} = \int \Omega_{\alpha\beta}(\mathbf{r}) \mathcal{R}_{lm}(P) d\mathbf{r} \quad (7.4.10)$$

$$q_{l'm'}^{\rho} = \int \rho(\mathbf{r}) \mathcal{R}_{l'm'}(Q) d\mathbf{r} \quad (7.4.11)$$

und die sogenannte Interaktionsmatrix

$$T_{lm,l'm'}(r_{QP}) = (-1)^{l'} \mathcal{I}_{l+l',m+m'}^*(r_{QP}). \quad (7.4.12)$$

definieren. Da nun  $r_{QP}$  den Abstand zwischen den Wechselwirkungszentren bezeichnet, entscheiden daher die Multipolmomente über die Stärke der langreichweitigen Wechselwirkung.

Ist die Annahme der *lokalen Ladungsneutralität* gültig, dann verschwindet das Moment nullter Ordnung  $q_{00}^{\rho}$  und wir erhalten  $\propto \left(\frac{1}{\|\mathbf{r}\|}\right)^2$  als führende Ordnung des verbleibenden Entwicklung, siehe Gregory [74] für allgemeine Schranken zu den Kugelflächenfunktionen  $Y_{lm}$ . Wir bemerken, dass dann nur die Dipolmomente über längere Distanz wirksam sind, da wir in den Coulomb-Matrixelementen über den  $\mathbb{R}^3$  integrieren<sup>3</sup>. Wenn das zu betrachtende System prinzipiell nur ein- oder zwei-dimensional ist, sollte es allgemein keine starke langreichweitige Wechselwirkung geben.

Allgemein stellt dies entsprechende Bedingungen an die Menge der Kernindizes  $\mathcal{U}_i$  bei gegebener Menge der Basisindizes  $\mathcal{I}_i$  und umgekehrt. Sie sollten dergestalt gewählt sein, dass sich die beiden Ladungsverteilungen gegenseitig in einer Multipolentwicklung möglichst gut aufheben.

#### 7.4.4. Nicht-lineare Lösung aus Teilproblemen

Im Laufe der selbst-konsistenten Iteration berechnen wir immer wieder neue Eigenvektoren, mit denen anschließend die Dichtematrix und schließlich die Fockmatrix des nächsten Iterationsschritts aufgestellt wird.

Zunächst haben wir das Problem des Startwerts in Abschnitt 4.3.2 besprochen. Da sich bspw. aus den Dichten neutraler Atome ein guter Startwert  $\mathbf{D}^{(0)}$  zusammensetzen lässt und dieser in einer lokalen, atomzentrierten Basis sehr gut darstellbar ist, ist die für ein Teilprobleme aufgestellte Anfangsdichte  $\mathbf{D}_i^{(0)}$  zusammengesetzt exakt die gleiche wie die im vollen Problem und vollem Ansatzraum aufgestellte und projizierte Dichte

<sup>3</sup>Wir benötigen einen Abfall stärker als  $\left(\frac{1}{\|\mathbf{r}\|}\right)^3$  im Integranden, damit das Integral konvergent ist.

## 7. BOSSANOVA

$\mathbf{P}_i \mathbf{D}^{(0)} \mathbf{P}_i$ . Das bedeutet, dass vom ersten Iterationsschritt an  $\hat{\varphi}_i^{(sr)}[\mathbf{D}_i^{(0)}]$  einen im Teilproblem auswertbaren Ausdruck darstellt und insbesondere, dass mit diesem Startwert  $\hat{\varphi}_i^{(sr)}[\mathbf{D}^{(0)}] = \hat{\varphi}_i^{(sr)}[\mathbf{D}_i^{(0)}]$  gilt.

Wir bemerken anschließend, dass die Abbildung  $\hat{\varphi}_i^{(sr)}[\mathbf{D}_i]$  von einer Dichtematrix wieder zu einer Dichtematrix führen muss, welche bereits von Basisfunktionen aus der Untermenge  $\mathcal{I}_i$  dargestellt werden. Denn die Lösungen zu  $\hat{\varphi}_i^{(sr)}[\mathbf{D}_i]$  sind nur in den Basisfunktionen zu der Indexuntermenge  $\mathcal{I}_i$  dargestellt.

Daher kann der gesamte selbst-konsistente Zyklus ausschließlich in der Hauptuntermatrix  $\mathbf{F}_i$  betrachtet werden, wenn  $\hat{\varphi}_i^{(lr)}[\mathbf{D}]$  vernachlässigt wird. Dies führt zu einem Teilproblem des molekularen Systems, da sich die Anfangsdichte, siehe den Paragraphen in Abschnitt 4.3.2, ebenso auf den Unterraum der Hauptuntermatrix beschränkt. Dieses gestörte Teilproblem nennen wir auch *Fragment*.

Da wir keine notwendigen Bedingungen für die Eindeutigkeit, sondern nur für die Existenz kennen, beschränken wir uns darauf: Wenn die Teilprobleme ladungsneutral sind, dann wissen wir, dass die gefundenen Lösungen der HF-Gleichung eine entsprechende Anzahl negativer Eigenvektoren aufweist, siehe Abschnitt 4.1.6, Eigenschaft 3.

Dies hatten wir schon bei der Betrachtung der Größe der Störung mittels einer Multipolentwicklung festgestellt und die Bedingung aus dieser war noch stärker, denn die beiden Ladungsverteilungen mit unterschiedlichem Vorzeichen sollen sich am Ort  $\mathbf{x} \in \mathbb{R}$  möglichst aufheben. Diese Bedingung schränkt die Zerlegung der Kernindizes  $\mathcal{U}_i$  pro Fragment bei gegebener Basisindexuntermenge  $\mathcal{I}_i$  so stark ein, dass die eine von der anderen abhängig wird. Wir kommen darauf in Abschnitt 7.5 noch einmal zurück.

### 7.4.5. Störbehandlung der langreichweitigen Wechselwirkung

Wenn also die Kernladung und die elektronische Ladungsverteilung so kombiniert werden können, dass lokale Ladungsneutralität besteht oder sogar das Dipolmoment verschwindet, dann ist zunächst die Änderung der Matrixelemente bei Vernachlässigung des langreichweitigen Anteils klein, aber in Folge vor allem Störtheorie zulässig.

Dazu stellen wir uns vor, dass der Fockoperator einer Hintereinanderschaltung der Fockabbildungen entspricht, bei der die resultierende Dichtematrix eines Schritts in die Fockabbildung des nächsten Schritts eingeht. Wir erhalten die Dichtematrix dabei mittels Gleichung (7.2.3) aus den kurzreichweitigen Anteilen (7.4.4a). Bei jedem Schritt erfährt die Dichtematrix dadurch eine Störung. Wir nehmen an, dass die Störung insgesamt aufgrund der problemgrößenunabhängigen Anzahl der Iterationsschritte klein bleibt.

Wir spalten dann formal den Fockoperator in zwei Anteile auf,

$$\hat{F} = \hat{F}^{(0)} + \lambda \hat{F}^{(1)} = \hat{T} + \hat{K} + \left( \hat{V}^{(\text{core,el})} + \hat{V}^{(\text{el,el})} \right)^{\text{sr}} + \lambda \left( \hat{V}^{(\text{core,el})} + \hat{V}^{(\text{el,el})} \right)^{\text{lr}}. \quad (7.4.13)$$

Für  $\lambda = 1$  erhalten wir den vollen Operator  $\hat{F}$  und für  $\lambda = 0$  haben wir den gestörten Operator, bei dem nur bestimmte kurzreichweitige Anteile der Coulomb-Wechselwirkung berücksichtigt worden sind. Die trunkierten Austauschanteile werden dabei wirklich vernachlässigt.



Dann behandeln wir diesen Ansatz in erster Ordnung der Rayleigh-Schrödinger-Störtheorie. Zu diesem Zweck schreiben wir die Einteilchen-Wellenfunktionen als Lösungen zum Fockoperator als  $\phi_i = \phi_i^{(0)} + \lambda\phi_i^{(1)} + \dots$  und die resultierenden Energieeigenwerte als  $\epsilon_i = \epsilon_i^{(0)} + \lambda\epsilon_i^{(1)} + \dots$ . Wir benötigen weiterhin Orthogonalität zwischen ungestörten  $\phi_i^{(0)}$  und gestörten Einteilchen-Wellenfunktionen  $\phi_i^{(1)}$ .

Nachdem wir diese Definitionen in die Gleichung (4.1.16) eingesetzt und nach Ordnung von  $\lambda$  sortiert haben, erhalten wir,

$$\lambda^0 : \epsilon_i^{(0)} = \int \phi_i^{(0)*}(\mathbf{r}) \hat{F}^{(0)} \phi_i^{(0)}(\mathbf{r}) d\mathbf{r} \quad (7.4.14a)$$

$$\lambda^1 : \epsilon_i^{(1)} = \int \phi_i^{(0)*}(\mathbf{r}) \hat{F}^{(1)} \phi_i^{(0)}(\mathbf{r}) d\mathbf{r}. \quad (7.4.14b)$$

Erst höhere Ordnung würden berücksichtigen, dass sich durch die fehlende langreichweitige Wechselwirkung auch die Einteilchen-Wellenfunktionen verändern könnten und würden Korrekturen der Art  $\phi_i^{(1)}$  liefern.

### Fehlermaß

Lösen wir den vollen Operator mit  $\lambda = 1$  werden wir im Allgemeinen andere Lösungen  $\{\phi_i\}$  erhalten als mit dem gestörten Operator mit  $\lambda = 0$ . Das bedeutet, dass wir im Allgemeinen den Störanteil des Operators  $(\hat{V}^{(\text{core,el})} + \hat{V}^{(\text{el,el})})^{\text{lr}}$  nicht angeben können, denn er hängt ja von der uns unbekanntem Lösung des vollen Problems ab.

Wir gehen jedoch davon aus, dass der Störanteil, der von den ungestörten Lösungen  $\{\phi_i^{(0)}\}$  abhängt, eine betragsmäßig größere Coulomb-Wechselwirkungsenergie liefern wird. Denn die ungestörten Lösungen konnten ja nicht in diesem zusätzlichen Anteil minimiert werden.

Auch der kurzreichweitige Anteil  $(\hat{V}^{(\text{core,el})} + \hat{V}^{(\text{el,el})})^{\text{sr}}$ , welcher im ungestörten Anteil des Operators enthalten ist, hängt von den vollen Lösungen ab. Er wird im Allgemeinen jedoch unterschätzt werden, da über diesen Anteil in den Fragmenten minimiert wird und dabei gerade die langreichweitigen Wechselwirkungen fehlten.

Sei

$$E^{(c)}_i = \iint \rho_i(\mathbf{r}_1) \Phi(\mathbf{x}_1) \quad (7.4.15)$$

die Coulomb-Wechselwirkungsenergie zu einem Fragment mit Index  $i$ , siehe auch (7.3.5), mit dem Potential  $\Phi$  nach (7.3.9).

Wir betrachten die gesamte Coulomb-Wechselwirkungsenergie als Summe über entsprechende Beiträge der Teile (7.4.15) im Sinne von (7.2.6a),

$$\begin{aligned} E^{(c)} &= \sum_{I_j \in \mathcal{J}} \widetilde{E^{(c)}}_j^{(sr)} + \widetilde{E^{(c)}}_j^{(lr)} \\ &= \sum_{I_j \in \mathcal{J}} \widetilde{E^{(c)}}_j^{(sr)} + \sum_{I_j \in \mathcal{J}} \widetilde{E^{(c)}}_j^{(lr)} \end{aligned}$$

## 7. BOSSANOVA

Dann lösen wir nach der fehlenden Wechselwirkungsenergie auf,

$$\Delta E^{(c)} := \sum_{\mathcal{I}_j \in \mathcal{J}} \widetilde{E}^{(c)}_j^{(lr)} = E^{(c)} - \sum_{\mathcal{I}_j \in \mathcal{J}} \widetilde{E}^{(c)}_j^{(sr)}. \quad (7.4.16)$$

Nach unseren Argumenten liefert die Berechnung der gesamten Coulomb-Wechselwirkungsenergie  $E^{(c)}$  aus den ungestörten Lösungen eine Abschätzung nach oben. Weiterhin ist die Summe der Beiträge aus den Fragmenten  $\widetilde{E}^{(c)}_j^{(sr)}$  mit jeweils unabhängigen kurzreichweitigen Anteile gleichermaßen eine Abschätzung nach unten. Daher erwarten wir aus der Störtheorie eine Abschätzung des Fehlers nach oben. Dabei berechnen wir wie schon erwähnt, dass jeweils aus (7.2.7) mit den Fragmentanteilen  $\rho_i$ , welche nur über den kurzreichweitigen Anteil minimiert wurden, eine Näherung an die Ladungsdichte  $\rho$  der vollen Lösung. Damit ist das zugehörige Potential gleichfalls nur eine Näherung.

Wir werden dieses Maß in Kapitel 9 quantitativ untersuchen. Wie schon erwähnt, nehmen wir dabei an, dass sich die eigentliche Lösung nicht weiter verändert, wie sie nur Störtheorie höherer Ordnung erfassen könnte. Wir bemerken noch, dass diese Berechnung deswegen nicht den sogenannten Polarisationsanteil enthält, welcher sich aufgrund der wechselseitigen Änderungen der Einteilchen-Wellenfunktion dieser paarweisen Wechselwirkung ergibt, siehe [115], wo im PIMO-Ansatz eine äußere Iteration eingeführt wird, um diese vollständig zu erfassen.

### 7.5. Abhängigkeitsgraph und Fragmente

Die von uns vorgeschlagene Methode steht und fällt mit den Zerlegungen der Indexmenge der Elektronenfreiheitsgrade sowie der Kernfreiheitsgrade. Darüber wie wir diese Zerlegungen erhalten, haben wir bisher noch nicht gesprochen. Wir haben nur ausgeführt, dass die Bedingung zur Minimierung des Anteils der langreichweitigen Wechselwirkung eine der beiden Zerlegungen festlegt.

Grundsätzlich möchten wir die Zerlegung aus einem Graphen erzeugen, welcher die Abhängigkeiten der einzelnen Elektronenfreiheitsgrade voneinander angibt. Entfernungen in diesem Graphen sollen dabei die Stärke der wechselseitigen Abhängigkeit ausdrücken.

Auch im Rahmen von molekularen Systemen sind Graphen eine bekannte Darstellung. So stellen Gasteiger und Jochum [64] einen Algorithmus zur Findung elementarer Ringe innerhalb eines Molekülgraphen dar. Hu et al. [100] geben eine Übersicht über viele Methoden, die topologische Eigenschaften aus Molekülgraphen extrahieren.

Wir werden aus diesem Grund den (kovalenten) Bindungsgraphen des molekularen Systems als einen solchen Abhängigkeitsgraphen untersuchen.

Auf Details zur Bestimmung dieses Bindungsgraphen und auf dessen Eignung als Abhängigkeitsgraph gehen wir erst in Abschnitt 9.3 im Rahmen der numerischen Experimente ein.

Sei also ein solcher *Bindungsgraph*  $G = (\mathcal{V}, \mathcal{E})$  gegeben mit  $\mathcal{V}$  der Menge der Knoten im Graph, welcher der Anzahl der Kernindizes  $\mathcal{M} = \{1, \dots, M\}$  entspricht, und mit  $\mathcal{E}$  der Menge der Kanten im Graph bzw. der kovalenten Bindungen.

Dann legen wir die Zerlegung  $\mathcal{U}$  der Kernindizes wie folgt fest: Sei eine Zahl  $k$  gegeben, welche die maximale Anzahl an Knoten eines Teilgraphen bestimmt. Dann enthalte die Zerlegung  $\mathcal{U}$  alle Unterindexmengen von  $\mathcal{M}$ , so dass jedes  $\mathcal{U}_i$  die Knotenindizes eines *verbundenen Teilgraphen* von  $G$  enthalte und gleichzeitig die Mächtigkeit dieser Knotenmenge  $|\mathcal{U}_i|$  kleiner gleich  $k$  ist. Wir nennen dieses  $k$  die Trunkierungs- oder auch *Bindungsordnung*.

Wir wollen kurz an einem Beispiel erläutern, dass damit eine größenkonsistente Methode folgt. Wir betrachten zwei unabhängige Systeme  $A$  und  $B$ . Die Fockmatrix des gemeinsamen Systems  $AB$  besteht dann aus zwei getrennten Blöcken auf der Diagonalen. Der Bindungsgraph zerfällt in zwei verbundene Untergraphen. Wir betrachten die Energie des Gesamtsystems  $E^{(\text{SCF})}_{AB}$ . Diese Energie muss bspw. durch die HF-Methode bestimmt werden. Nehmen wir an, dass diese Methode größenkonsistent ist, also  $E^{(\text{SCF})}_{AB} = E^{(\text{SCF})}_A + E^{(\text{SCF})}_B$ , dann lässt sich zeigen, dass dies mit dem obig beschriebenen Auslassen nicht-verbundener Teilgraphen harmoniert, siehe auch Griebel et al. [78].

**Lemma 13** (Beiträge nicht-verbundener Untergraphen). *Sei  $G = (\mathcal{V}, \mathcal{E})$  gegeben und weiterhin zwei disjunkte Knotenteilmengen  $A \subset \mathcal{V}$ ,  $B \subset \mathcal{V}$  mit  $A \cap B = \emptyset$ . Sei  $\widetilde{E^{(\text{SCF})}}_{\emptyset} = 0$ . Wir nehmen an, dass  $E^{(\text{SCF})}_{AB} = E^{(\text{SCF})}_A + E^{(\text{SCF})}_B$  gilt. Dann folgt  $\widetilde{E^{(\text{SCF})}}_{AB} = 0$ .*

*Beweis.* Wir führen den Beweis per vollständiger Induktion über die Anzahl der Knoten in  $G$ . Die Induktionsannahme für  $|\mathcal{V}| \leq 2$  folgt sofort. Wir betrachten dann einen Graphen  $G' = (\mathcal{V}', \mathcal{E}')$  mit  $|\mathcal{V}'| = n + 1$  unter der Annahme, dass die Aussage für  $|\mathcal{V}'| \leq n$  gilt. Zunächst folgt aber aus (7.2.6b) sofort,

$$E^{(\text{SCF})}_i = \sum_{\mathcal{I}_j \subseteq \mathcal{I}_i} \widetilde{E^{(\text{SCF})}}_j. \quad (7.5.1)$$

Damit betrachten wir dann (7.2.6b),

$$\begin{aligned} \widetilde{E^{(\text{SCF})}}_{AB} &= E^{(\text{SCF})}_{AB} - \sum_{c \subsetneq AB} \widetilde{E^{(\text{SCF})}}_c \\ &= E^{(\text{SCF})}_{AB} - \sum_{a \subset A, b \subset B: |a|+|b| < |A|+|B|} \widetilde{E^{(\text{SCF})}}_{ab} \\ &= E^{(\text{SCF})}_{AB} - \sum_{a \subset A, b \subset B: 0 < |a|+|b| < |A|+|B|} \widetilde{E^{(\text{SCF})}}_{ab} - \widetilde{E^{(\text{SCF})}}_{\emptyset}. \end{aligned}$$

Da der leeren Menge ein Beitrag gleich null zugeordnet ist, fügen wir diesen nochmal hinzu,

$$\begin{aligned} &= E^{(\text{SCF})}_{AB} - \sum_{a \subset A, a \neq \emptyset} \widetilde{E^{(\text{SCF})}}_a - \sum_{b \subset B, b \neq \emptyset} \widetilde{E^{(\text{SCF})}}_b - 2\widetilde{E^{(\text{SCF})}}_{\emptyset} \\ &\quad - \sum_{a \subset A, b \subset B, a \neq \emptyset, b \neq \emptyset: |a|+|b| < |A|+|B|} \widetilde{E^{(\text{SCF})}}_{ab}, \end{aligned}$$

## 7. BOSSANOVA

und nutzen (7.5.1),

$$\begin{aligned}
 &= E^{(\text{SCF})}_{AB} - E^{(\text{SCF})}_A - E^{(\text{SCF})}_B - \sum_{a \subset A, b \subset B, a \neq \emptyset, a \neq \emptyset: |a|+|b| < |A|+|B|} \widetilde{E^{(\text{SCF})}}_{ab} \\
 &= - \sum_{a \subset A, b \subset B, a \neq \emptyset, a \neq \emptyset: |a|+|b| < |A|+|B|} \widetilde{E^{(\text{SCF})}}_{ab}, \\
 &= 0
 \end{aligned}$$

wo wir im letzten Schritt die Induktionsvoraussetzung für  $|V'| \leq n$  verwendet haben.  $\square$

Aus dem Bindungsgraphen wird durch das Auslassen aller Teilprobleme zu nicht-verbundenen Untergraphen eine Anzahl von Teilproblemen erzeugt, die nur linear mit der Anzahl der Elektronen ansteigt. Dies gilt, da der Bindungsgrad pro Kern bzw. Knoten des Graphen problemgrößenunabhängig beschränkt ist.

Mit den üblichen Atomaren Orbitale (AOs) als Basisfunktionen ergibt sich dadurch auch sofort die Zerlegung  $\mathcal{J}$ . So gehören zu einer Indexmenge  $\mathcal{I}_i$  alle Basisfunktionen, deren Zentren am Kernort zu Kernen aus dem zugehörigen  $\mathcal{U}_i$  liegen. Dies ergibt sich aus der Bedingung der Minimierung der langreichweitigen Störung.

Die lokale Ladungsneutralität hingegen fordert, dass die Anzahl der Elektronen im Teilproblem gleich der Summe der Kernladungen der in  $\mathcal{U}_i$  enthaltenen Kerne sein soll. Dadurch ergeben sich pro Teilproblem auch ausreichend viele negative Zustände.

Damit sind die Teilprobleme oder Fragmente vollständig spezifiziert. Die entscheidende Idee ist nun, dass die Hauptuntermatrizen aus der, wie weiter oben beschrieben, gestörten Fockmatrix zu diesen Fragmenten nichts anderes als eigenständige Fockmatrizen zu diesen Teilproblemen darstellen. Dies gilt, wenn wir die langreichweitige Wechselwirkung innerhalb der selbst-konsistenten Iteration vernachlässigen. Anstatt also Hauptuntermatrizen der Fockmatrix des vollen Molekülsystems zu betrachten, kann man direkt die Lösungen und damit die Energie nach (7.2.5) aus den HF-Gleichungen der Fragmente ermitteln. Wir bemerken noch, dass sich die aus diesem Verfahren berechnete Gesamtenergie nicht variationell verhalten wird, aber sie ist, wie oben verdeutlicht, größenkonsistent. Auf diese Punkt kommen wir im folgenden Kapitel 8 noch einmal zurück.

Damit schließt sich der Kreis zu den in der Einleitung des Kapitels besprochenen Fragmentansätzen. Bei diesen wird ein Molekülsystem in kleinere Teilsysteme aufgetrennt und die Modelle nur für diese Teile gelöst. Anschließend ergibt sich eine Näherung an die Eigenschaft des Gesamtsystems aus der Summe der Eigenschaften der Teile. Aus der Dünnbesetztheit der Dichtematrix und der Fockmatrix, aus der Existenz lokalisierter Zustände und aus einer später untersuchten Eignung des Bindungsgraphen als Generator der notwendigen Zerlegung können wir nun den Erfolg dieser Ansätze verstehen.

### 7.6. Dichte-basierte HF-Methode

Wir haben in Abschnitt 7.1 die Dünnbesetztheit der Fockmatrix ausgeführt, woraus quadratische Kosten für das Aufstellen der Matrix resultieren. Wir werden abschließend einen anderen Ansatz besprechen, um lineare Kosten zu erhalten.

Ein von dem in diesem Kapitel besprochenen Ansatz abweichende Methode – siehe [127] und Helgaker et al. [91, Kap. 10.7] für eine sehr schöne Darstellung – besteht darin, die in jedem selbstkonsistenten Iterationsschritt notwendige Diagonalisierung zu vermeiden und durch eine direkte Minimierung der Dichtematrix zu ersetzen. Da dabei nur Matrixmultiplikationen und -additionen dünnbesetzter Matrizen benötigt werden, verringert sich die Komplexität von andernfalls kubischen auf lineare Kosten. Diese Methoden-Familie fällt unter das Stichwort Dichtematrixsuche. Eine für diese Methoden wichtige theoretische Arbeit, welche die Dünnbesetztheit der Dichtematrix im Falle einer vorhandenen Lücke im Eigenwertspektrum zwischen positiven und negativen Eigenwerten beweist, findet sich bei Benzi et al. [18].

Im Rahmen dieser Vorgehensweise stehen jedoch die Orbitale als Eigenfunktionen des Fockoperator nicht mehr zur Verfügung, denn sie resultieren aus der Diagonalisierung der Fockmatrix. Stattdessen müssen die Matrixelemente, siehe Abschnitt 4.2.2, direkt mittels der AOs ausgedrückt werden, indem anstelle der Dichtematrix die AO Dichtematrix  $\mathbf{D}^{\text{AO}}$ ,

$$\mathbf{D}^{\text{AO}} = \mathbf{C}\mathbf{D}\mathbf{C}^T,$$

verwendet wird. Dabei muss aber darauf geachtet werden, dass die aus diesen Orbitalen resultierende Dichtematrix immer noch eine gültige Lösung darstellt. Dies gilt, wenn sie die drei folgende Bedingungen erfüllt, welche für die Dichtematrix (4.2.7) gilt:

**Symmetriebedingung**  $\mathbf{D} = \mathbf{D}^T$

**Spurbedingung**  $\text{Tr}(\mathbf{D}) = \frac{1}{2}N$

**Idempotenz**  $\mathbf{D}^2 = \mathbf{D}$

Alle Dichtematrizen aus Eigenlösungen des Fockoperators lassen sich mittels einer exponentiellen Parametrisierung durch eine antisymmetrische Drehmatrix  $\kappa$  aus einer Referenzmatrix  $\mathbf{D}$  darstellen,

$$\mathbf{D}(\kappa) = \exp(-\kappa)\mathbf{D}\exp(\kappa).$$

Analog gilt dies auch für Dichtematrizen aus AOs. Diese Darstellung sichert die obigen Eigenschaften zu, wenn sie für die ursprüngliche Dichtematrix galten. Anschließend wird die Exponentialfunktion mittels der Baker-Campbell-Hausdorff-Entwicklung dargestellt, aber nach einigen Ordnungen trunziert. Dieses Abschneiden der Entwicklung stört die Eigenschaften, wobei vor allem die Idempotenz nicht trivial wiederherzustellen ist. Dazu betrachtet man den Ansatz,

$$(\mathbf{D} + \delta\mathbf{D})^2 - (\mathbf{D} + \delta\mathbf{D}) = 0,$$

und wandelt dies in eine Fixpunktiteration um, die im Allgemeinen sehr schnell konvergiert. Diesen Schritt nennt man Bereinigung der Dichtematrix und er basiert auf McWeeny [134, Appendix 3].

Die HF-Energie wird dann bezüglich der Parametermatrix  $\kappa$  entwickelt und daraus Gradient und Hesse-Matrix berechnet, zur Minimierung mittels eines Standard-Verfahrens erster oder zweiter Ordnung. Alternativ kann die Lösung auch über einen selbstkonsistenten Zyklus erfolgen, da die Lösung der Roothaan-Hall-Gleichung gleichbedeutend

## 7. BOSSANOVA

ist mit dem Minimum über die Summe der besetzten Eigenwerte. Dabei wird in jedem selbstkonsistenten Iterationsschritt das Funktional  $E_\epsilon$  minimiert,

$$\frac{\partial E_\epsilon}{\partial \kappa} = 0 \quad (7.6.1)$$

welches direkt in der Dichtematrix der AOs ausgedrückt werden kann. Dies wird als Dichte-basierter SCF-Ansatz bezeichnet.

Die Coulomb-Wechselwirkung, welche eigentlich quadratische Kosten hat, wird in [184, continuous FMM] bzw. [164, Gaussian FMM] durch eine schnelle Multipol-Methode in linearer Zeit ausgewertet, siehe auch [91, Kap. 9.14]. Weiterhin wird in [91, Kap. 9.12] durch eine Betrachtung einer Kette von Wasserstoffatomen motiviert, dass große Systeme irgendwann nur noch lineare Kosten beim Aufstellen der restlichen Matrizen verursachen. Dies gilt auch für die AO Dichtematrix, wenn die AOs örtlich gut voneinander getrennt sind. Treffen obige Annahmen nicht zu, ergeben sich durch die vollbesetzten Matrizen natürlich wieder quadratische bzw. bei Matrix-Matrix-Multiplikationen kubische Kosten. Auch ist zu vermuten, dass der Crossover-Punkt deutlich höher liegt, ab dem diese linear skalierende Methode schneller als die traditionelle HF-Methode ist, siehe Abschnitt 9.2.8, wo wir das Skalierungsverhalten unserer Methode untersuchen und vergleichen.

Bei der von uns vorgeschlagenen Methode wird keine dünnbesetzte, große Dichtematrix variiert, sondern stattdessen kleine Blöcke der Fockmatrix diagonalisiert. In beiden Fällen ist dabei die Dünnbesetztheit der Dichtematrix Voraussetzung für lineare Kosten.

## Zusammenfassung

Wir untersuchten in diesem Kapitel Methoden, welche die HF-Gleichung im Rahmen von Fragmentansätzen mit linearen Kosten lösen.

Zunächst betrachteten wir die Fockmatrix, wobei wir die Entwicklung in einer lokalen Basis angenommen haben und von lokalen Lösungen bzw. einer dünnbesetzten Dichtematrix ausgegangen sind. Dabei stellten wir fest, dass wir mit dem Schwerpunkttheorem aus dem vorigen Kapitel bis auf die Coulomb-Wechselwirkung lineare Kosten erreichen. Für die Berücksichtigung dieser Wechselwirkung schauten wir anschließend drei verschiedene Vorgehensweise an. Dazu spalteten wir zunächst diese in einen kurz- und einen langreichweitigen Anteil auf. Dieser langreichweitige Anteil der Coulomb-Wechselwirkung kann dann entweder ganz vernachlässigt werden, wobei die Zulässigkeit dieser Trunkierung von den genauen Ladungsverteilungen des spezifischen Molekülsystems abhängt. Dies untersuchten wir anhand einer formalen Multipolentwicklung. Weiterhin kann die langreichweitige Wechselwirkung mittels Störtheorie erster Ordnung berechnet werden, wobei die Einteilchen-Wellenfunktion aber dann von diesem Wechselwirkungsanteil unbeeinflusst bleibt. Und schließlich kann sie vollständig berücksichtigt werden.

Bei den letzten beiden Varianten wird eine Methode zur Berechnung der Wechselwirkung mit linearen Kosten benötigt, wie sie jedoch in der Literatur existieren und auf die wir in Abschnitt 9.5 auch noch eingehen werden.

Die ersten beiden Varianten wiederum haben gemein, dass sich bei ihnen der selbstkonsistente Zyklus dergestalt aufspaltet, dass er über jedes Teilprobleme allein durch-

geführt werden kann. Dadurch ergeben sich voneinander unabhängige Fragmente, deren Berechnung sich besonders einfach parallelisieren lässt. Wir werden dies mit den im nächsten Kapitel folgenden numerischen Experimenten untersuchen.

Schließlich betrachteten wir abschließend noch einen anderen Ansatz, um lineare Skalierung des HF-Problems zu erhalten, die Dichte-basierten Methoden. Von diesen erwarten wir aber prinzipiell eine größere Skalierungskonstante als bei den von uns betrachteten ersten beiden Varianten der Fragmentansätze.





## 8. Korrelationsenergie

Defranceschi und Le Bris [42, Abschn. 1.1] bezeichnen die HF-Methode vom mathematischen Standpunkt aus als einen „rigorous energy/approximate wave function approach“ (Ansatz mit rigider Energie/genäherter Wellenfunktion). Hierbei wird der Hamiltonoperator unverändert von der zugrundeliegenden elektronischen Schrödingergleichung übernommen, aber der Raum, in dem die Lösung gesucht wird, erfährt durch den Ansatz einer einzigen Slater-Determinante starke Einschränkung. Aus diesem Grund ist die Energie der HF-Methode im Rahmen des Variationsprinzips, siehe Abschnitt 3.4.5, auch nur eine obere Schranke an die echte Energie. Dies ist unabhängig von der Größe der gewählten Basis, die zur Aufstellung der Fockmatrix verwendet wird. Das heißt, dass eine genauere Lösung nur in einem größeren Ansatzraum gefunden werden kann.

In diesem Abschnitt diskutieren wir die Möglichkeit, im Rahmen der von uns im letzten Kapitel vorgestellten BOSSANOVA-Methode eine Näherung auch an die Korrelationsenergie zu erhalten. Zunächst untersuchen wir die methodisch einfache Configuration Interaction (CI), welche aber nicht größenkonsistent ist und damit ungeeignet für unseren summativen Ansatz, siehe auch [142, Abschn. 2.2.3.4]. Die Møller-Plesset Störtheorie (MPPT) im Vergleich ist größenkonsistent, wenn auch nicht variationell wie CI. Sie passt damit aber sehr gut zu unserer Methode.

Unser Fokus in diesem Kapitel liegt darauf, notwendige Grundlagen darzulegen, um die Methodenauswahl zu motivieren und später folgende numerische Ergebnisse zu erklären. Eine vollständige Untersuchung der die HF-Theorie erweiternden Methoden wie CI, MPPT oder Coupled Cluster (CC) würde den Rahmen dieser Arbeit sprengen.

Schließlich weisen wir darauf hin, dass aufgrund der Notwendigkeit eines Bindungsgraphen all jene Systeme fern ihres Gleichgewichtszustands zunächst ausgeschlossen sind. Daher gehen wir im Folgenden davon aus, dass die einzelne Slater-Determinante als HF-Lösung eine grundsätzlich gute Beschreibung des Grundzustands liefert und damit den sogenannten *Referenzzustand* darstellt.

### 8.1. Configuration Interaction

Traditionell ist die sogenannte CI ein erster und sehr einfacher Ansatz zur Vergrößerung des Vielteilchenansatzraums, siehe Helgaker et al. [91, Kap. 5.6] oder einfach und knapp gehalten bei Staemmler [161]. Mit den Eigenfunktionen des Fockoperators existiert eine Basis des Einteilchenraums, die bekanntermaßen als Slater-Determinanten  $\psi_I$  auch den Vielteilchenraum vollständig aufspannen<sup>1</sup>, siehe [161, Abschn. 8]. Schränkt man den An-

---

<sup>1</sup>Anstelle der CSFs bleiben wir hier im einfacheren Bild der Determinanten, deren Verallgemeinerung sich unmittelbar aus (4.1.3) erschließt, siehe die Bemerkungen in [91, Abschn. 11.1.1].

## 8. Korrelationsenergie

satzraum der HF-Lösung auf eine endliche Basis ein, dann lässt sich die elektronische Wellenfunktion wie folgt entwickeln,

$$\psi = \sum_I c_I \psi_I(\mathbf{r}_1, \dots, \mathbf{r}_N), \quad (8.1.1)$$

wobei der Index  $I$  alle möglichen  $\binom{N}{n}$  Slaterdeterminanten aus je  $N$  der  $n$  Spin-Orbitale enumeriert. Dabei haben wir den Spin diesmal nicht ausintegriert.

Damit erhält man folgende Eigenwertgleichung aus der elektronischen Schrödingergleichung (3.3.6),

$$\sum_J (\mathbf{H}_{IJ} - E \mathbf{S}_{IJ}) \psi_J = 0, \quad (8.1.2)$$

die Indizes  $I$  und  $J$  gehen dabei über die erwähnten Slater-Determinanten, wobei wie folgt die Matrixelemente für den Hamiltonoperator, siehe (3.1.3),

$$\mathbf{H}_{IJ} = \int \psi_I^* \hat{H} \psi_J d\mathbf{r} \quad (8.1.3)$$

und für den Überlapp

$$\mathbf{S}_{IJ} = \int \psi_I^* \psi_J d\mathbf{r} \quad (8.1.4)$$

definiert sind.

Aufgrund des Binomialfaktors ist diese Full Configuration Interaction (FCI) nur für sehr kleine Systeme mit vertretbarem Aufwand berechenbar. Aus diesem Grund sind verschiedene auf Heuristik basierende Trunkierungen eingeführt worden, die Truncated CI, bei denen nur ein Teil aller möglichen Slater-Determinanten berücksichtigt wird.

Wir wollen noch kurz die übliche Notation für die verschiedenen Slater-Determinanten einführen. Wir verweisen auf Abschnitt 4.1.2 für die geläufigen Begriffe. Ausgehend von dem eingangs erwähnten Referenzzustand  $\psi^{(\text{SCF})}$ , welcher aus einer Slater-Determinante von besetzten Spin-Orbitalen besteht, werden jeweils besetzte gegen virtuelle Spin-Orbitale in der Determinante ausgetauscht. Man nennt dies *Anregungen*. Je nachdem wieviele Spin-Orbitale dabei getauscht werden, spricht man von Einfach-, Zweifach-, Dreifach-Anregungen, usw. . Mittels eines Anregungsoperators  $\hat{T}$ , welcher auf den Referenzzustand  $\psi^{(\text{SCF})}$  wirkt, lässt sich formal dieser Austausch von Spin-Orbitalen kompakt notieren,

$$\begin{aligned} \psi_{\text{FCI}} = & \psi^{(\text{SCF})} + \sum_{i_1, a_1} t_{i_1}^{a_1} \hat{T}_{i_1}^{a_1} \psi^{(\text{SCF})} + \sum_{i_1 i_2, a_1 a_2} t_{i_1 i_2}^{a_1 a_2} \hat{T}_{i_1 i_2}^{a_1 a_2} \psi^{(\text{SCF})} + \dots, \\ & \dots \sum_{i_1 \dots i_N, a_1 \dots a_N} t_{i_1 \dots i_N}^{a_1 \dots a_N} \hat{T}_{i_1 \dots i_N}^{a_1 \dots a_N} \psi^{(\text{SCF})} \end{aligned} \quad (8.1.5)$$

mit den Entwicklungskonstanten bzw. sogenannten Anregungsamplituden  $t_{\dots}$  des Zustands, wobei unten die besetzten und oben die virtuellen Spin-Orbitale stehen. Auf die weiteren Details der Trunkierung wollen wir hier nicht eingehen. Eine sehr geläufige ist die Trunkierung nach Einfach- und Zweifachanregungen, als CISD abgekürzt, also nach der ersten Zeile in (8.1.5).

### 8.1.1. Größenkonsistenz

Die FCI-Wellenfunktion erfüllt die Forderung der Größenkonsistenz. Sie ist aber aufgrund der großen Zahl zu berücksichtigender Determinanten  $\psi_I$  in (8.1.1) nur für sehr kleine Systeme berechenbar. Beim Truncated CI-Ansatz ist diese Zahl durch Trunkierung der Summe wesentlich kleiner. Dieser Ansatz ist aber wiederum nicht größenkonsistent und deswegen nicht für unser Summationsverfahren geeignet. Wir wollen dies kurz am Beispiel zweier nicht-wechselwirkender Systeme  $A$  und  $B$  ausführen, wobei wir [91, Abschn. 11.2] folgen.

#### Full Configuration Interaction

Sei mit  $\psi_A^{(\text{FCI})}$  die FCI-Wellenfunktion und mit  $\hat{H}_A$  der Hamiltonoperator des Systems  $A$  bezeichnet. Dann haben wir folgende Eigenwertgleichungen,

$$\begin{aligned}\hat{H}_A \psi_A^{(\text{FCI})} &= E_A \psi_A^{(\text{FCI})}, \\ \hat{H}_A \psi_A^{(\text{SCF})} &= E^{(\text{SCF})}_A \psi_A^{(\text{SCF})},\end{aligned}$$

wobei sich die Energie  $E_A = E^{(\text{SCF})}_A + E^{(\text{corr})}_A$  und die Wellenfunktion  $\psi_A^{(\text{FCI})} = \psi_A^{(\text{SCF})} + \psi_A^{(\text{corr})}$  aufspalten lassen, analog auch für System  $B$ . Wie eingangs des Kapitels erwähnt, sehen wir in obigen Gleichungen jeweils den vollen Hamiltonoperator, aber einmal die volle elektronische Wellenfunktion  $\psi_A^{(\text{FCI})}$  und einmal die approximative Wellenfunktion  $\psi_A^{(\text{SCF})}$ .

Betrachten wir nun die Kombination der beiden Systeme  $A$  und  $B$  im gegen unendlich gehenden Abstand. Mit dem Hamiltonoperator als Summe der einzelnen Operatoren und der Summe der Energie  $E_{AB} = E_A + E_B$  der beiden Systeme, können wir dann die Wellenfunktion des vollen Systems  $AB$  als Produkt der Einzelwellenfunktionen darstellen,  $\psi_{AB}^{(\text{FCI})} = \psi_A^{(\text{FCI})} \cdot \psi_B^{(\text{FCI})}$ , siehe [91, Abschn. 4.3.1] für eine detaillierte Darlegung in zweiter Quantisierung,

$$(\hat{H}_A + \hat{H}_B) \psi_A^{(\text{FCI})} \psi_B^{(\text{FCI})} = (E_A + E_B) \psi_A^{(\text{FCI})} \psi_B^{(\text{FCI})}. \quad (8.1.6)$$

Damit können wir die Wellenfunktion und Energie des Gesamtsystems  $AB$  mit den Anteilen der Einzelsysteme  $A$  und  $B$  schreiben,

$$\begin{aligned}\psi_{AB}^{(\text{FCI})} &= \psi_A^{(\text{SCF})} \psi_B^{(\text{SCF})} + \psi_A^{(\text{SCF})} \psi_B^{(\text{corr})} + \psi_A^{(\text{corr})} \psi_B^{(\text{SCF})} + \psi_A^{(\text{corr})} \psi_B^{(\text{corr})}, \\ E_{AB} &= E^{(\text{SCF})}_A + E^{(\text{SCF})}_B + E^{(\text{corr})}_A + E^{(\text{corr})}_B,\end{aligned}$$

wobei wir die Größenkonsistenz der HF-Methode hier nicht weiter ausführen.

#### Truncated Configuration Interaction

Gehen wir nun von einer trunkierten Wellenfunktion aus, wobei in (8.1.5) nur Zweifachanregungen zugelassen seien. Wenn wir dann die Wellenfunktionen der Einzelsysteme mit der Wellenfunktion des Gesamtsystems

$$\psi_{AB}^{(\text{FCI})} = \psi_A^{(\text{SCF})} \psi_B^{(\text{SCF})} + \psi_A^{(\text{SCF})} \psi_B^{(\text{corr})} + \psi_A^{(\text{corr})} \psi_B^{(\text{SCF})}$$

## 8. Korrelationsenergie

vergleichen, dann fehlt der Term  $\psi_A^{(\text{corr})}\psi_B^{(\text{corr})}$ , da er nur aus Vierfachanregungen bestehen würde, die aber bei unserem Ansatz nicht mehr berücksichtigt werden. Setzen wir diese trunkierte Wellenfunktion in (8.1.6) ein, so können wir  $E_{AB}$  nicht mehr in der gleichen additiven Form aufschreiben.

Für Truncated CI ergibt sich die Energie des Gesamtsystems nicht als Summe der Einzelsysteme und die Wellenfunktion lässt sich nicht als Produkt darstellen. Dies ist gemeint, wenn eine Methode nicht *größenkonsistent* ist. Wir möchten an dieser Stelle hervorheben, dass die Energie des Referenzzustands  $E^{(\text{SCF})}$  größenkonsistent bleibt, allein die Korrelationsenergie  $E^{(\text{corr})}$  nimmt mit zunehmender Größe des Systems ab, siehe [91, Abschn. 11.2.4] für numerische Beispiele anhand von Wassermolekülen.

Bei der BOSSANOVA-Methode wird aber insbesondere ein molekulares System in einzelne überlappende Fragmente aufgeteilt und diese aufsummiert, eine größenkonsistente Methode ist also unabdingbar.

## 8.2. Møller-Plesset Störtheorie

Møller und Plesset [136] schlugen bereits kurz nach der Einführung der HF-Methode einen störtheoretischen Ansatz vor, um die Genauigkeit der Näherung an die echte Energie zu verbessern. Wir folgen hier der knappen Darstellung in Helgaker et al. [91, Abschn. 5.8.1 und 14.4].

Zunächst wird der Hamiltonoperator in einen ungestörten und einen Störanteil aufgetrennt,

$$\hat{H} = \hat{F} + \hat{V}^{(\text{core,core})} + \hat{\Phi} \quad (8.2.1)$$

mit dem Fluktuationpotential  $\hat{\Phi}$ , welches den Unterschied zwischen dem echten Coulomboperator und dem effektiven Ein-Teilchen-Coulomboperator des Fockoperator beschreiben soll und dem Kern-Kern-Wechselwirkungsoperator, welcher kein Teil des Elektron-Hamiltonoperators ist.

Das ungestörte Problem ergibt sich dabei zu dem Referenzzustand  $\psi^{(\text{SCF})}$ ,

$$\hat{F}\psi^{(\text{SCF})} = \sum_i \epsilon_i \psi^{(\text{SCF})},$$

wobei die Summation über die besetzten Zustände in  $\psi^{(\text{SCF})}$  geht.

In nullter und erster Ordnung Störtheorie erhalten wir damit die Energien,

$$E^{(\text{MP0})} = \int \psi^{(\text{SCF})\dagger} \hat{F} \psi^{(\text{SCF})} d\mathbf{r} \quad (8.2.2)$$

$$E^{(\text{MP1})} = \int \psi^{(\text{SCF})\dagger} \hat{\Phi} \psi^{(\text{SCF})} d\mathbf{r}, \quad (8.2.3)$$

denn

$$\int \psi^{(\text{SCF})\dagger} \hat{\Phi} \psi^{(\text{SCF})} d\mathbf{r} = \int \psi^{(\text{SCF})\dagger} \left( \hat{H} - \hat{F} - \hat{V}^{(\text{core,core})} \right) \psi^{(\text{SCF})} d\mathbf{r}$$

und damit die HF-Energie erst als Summe  $E^{(\text{HF})} = E^{(\text{MP0})} + E^{(\text{MP1})} + V^{(\text{core,core})}$ , da  $E^{(\text{MP0})}$  nur die Summe der besetzten Eigenwerte darstellt, siehe auch (4.1.17) in Abschnitt 4.1.6.

In zweiter Ordnung Störtheorie wird auch eine Änderung des Zustands selbst berechnet. Dazu müssen wir den Ansatzraum vergrößern, um Korrelationen mit zu berücksichtigen. Analog werden dazu Einfach-, Zweifach- und höhere Anregungen benutzt, siehe (8.1.5). Für die Störung des Referenzzustands in erster Ordnung können aber nach dem Brillouin-Theorem nur die zweifach angeregten Determinanten beitragen. Deswegen setzen wir die Wellenfunktion als

$$\psi^{(\text{MP2})} = \hat{T}_2^{(1)} \psi^{(\text{SCF})} \quad \text{mit} \quad \hat{T}_2^{(1)} = \sum_{\substack{A>B, \\ I>J}} t_{IJ}^{AB} \hat{T}_{IJ}^{AB} \quad (8.2.4)$$

an, wobei sich die Anregungsamplituden  $t_{IJ}^{AB}$  aus der Rayleigh-Schrödinger-Störtheorie als

$$t_{IJ}^{AB} = \frac{1}{\epsilon_A + \epsilon_B - \epsilon_I - \epsilon_J} \int \psi^{(\text{SCF})\dagger} \hat{H} \hat{T}_{IJ}^{AB} \psi^{(\text{SCF})} d\mathbf{r}$$

ergeben, siehe [91, Abschn. 14.2.5]. Setzen wir den Teil des Hamiltonoperator ein, der auf zwei Elektronen gleichzeitig wirkt, also den Vier-Zentren-Integrale  $I_{AIBJ}$ , siehe Abschnitt 4.2.2, dann erhalten wir,

$$t_{IJ}^{AB} = -\frac{I_{AIBJ} - I_{AJBI}}{\epsilon_A + \epsilon_B - \epsilon_I - \epsilon_J}. \quad (8.2.5)$$

Mit der Wellenfunktion erster Ordnung ergibt sich dann die Korrektur zur Energie in zweiter Ordnung als

$$E^{(\text{MP2})} = \int \psi^{(\text{SCF})\dagger} \hat{\Phi} \psi^{(\text{MP1})} d\mathbf{r} = - \sum_{\substack{A>B, \\ I>J}} t_{IJ}^{AB} \frac{|I_{AIBJ} - I_{AJBI}|^2}{\epsilon_A + \epsilon_B - \epsilon_I - \epsilon_J}. \quad (8.2.6)$$

Diese Korrektur in zweiter Ordnung liefert dann einen gewissen Anteil der Korrelationsenergie der vollen Lösung.

Wir bemerken, dass die MPPT nicht konvergent ist, siehe [91, Abschn. 5.8.3], weswegen wir uns in den nachfolgenden Experimenten auf MP2 beschränken werden, welche insbesondere für Gleichgewichtszustände sehr gute und teils der CISD überlegene Ergebnisse liefert, siehe [91, Abschn. 5.8.2]. Auch ist MPPT nicht variationell. Dies trifft aber auch für BOSSANOVA zu.

### 8.2.1. Größenkonsistenz

Schließlich wollen wir am gleichen Beispiel wie oben erläutern, dass MPPT im Gegensatz zu Truncated CI *größenkonsistent* ist, siehe [91, Abschn. 14.2.6], und deswegen im Rahmen der orthogonalen Summation benutzt werden kann.

Wir betrachten also wieder die Lösung des kombinierten Systems,

$$\psi_{AB}^{(\text{MP2})} = \hat{T}_{2,AB}^{(1)} \psi_{AB}^{(\text{SCF})},$$

## 8. Korrelationsenergie

dabei separiert der Anregungsoperator  $\hat{T}_{2,AB}^{(1)} = \hat{T}_{2,A}^{(1)} + \hat{T}_{2,B}^{(1)}$ , da die Matrixelemente in erster Ordnung

$$\int \psi_{AB}^{(\text{SCF})\dagger} \hat{\Phi} \psi_{AB}^{(\text{MP1})} d\mathbf{r}$$

verschwinden, falls der Anregungsoperator auf beide Systeme gleichzeitig wirkt.

Die Energiekorrektur in erster Ordnung  $E^{(\text{MP1})}$  separiert additiv,

$$E_{AB}^{(\text{MP1})} = \int \psi_{AB}^{(\text{SCF})\dagger} \hat{\Phi}_{AB} \psi_{AB}^{(\text{SCF})} d\mathbf{r} = E_A^{(\text{MP1})} + E_B^{(\text{MP1})} \quad (8.2.7)$$

da das Fluktuationspotential  $\hat{\Phi}_{AB} = \hat{\Phi}_A + \hat{\Phi}_B$  additiv ist. Schließlich separiert auch die Energiekorrektur in zweiter Ordnung in die der Einzelsysteme,

$$\begin{aligned} E_{AB}^{(\text{MP2})} &= \int \psi_{AB}^{(\text{SCF})\dagger} \hat{\Phi} \psi_{AB}^{(\text{MP1})} d\mathbf{r} \\ &= \int \psi_{AB}^{(\text{SCF})\dagger} (\hat{\Phi}_A + \hat{\Phi}_B) (\hat{T}_A + \hat{T}_B) \psi_{AB}^{(\text{SCF})} d\mathbf{r} \\ &= E_A^{(\text{MP2})} + E_B^{(\text{MP2})} \end{aligned} \quad (8.2.8)$$

wegen der additiven Separabilität des Fluktuationspotentials und des Anregungsoperators und der Orthonormalität des Systems angeregter Determinanten.

## 8.3. Korrelationsentwicklung

Wir wollen nun eine Verbindung zwischen den obigen Entwicklung (8.1.5) mit dem Anregungsoperator  $\hat{T}$  und den zugehörigen neuen Entwicklungskoeffizienten, wie sie sich z. B. aus Truncated CI oder MPPT bestimmen lassen, und der Korrelationsentwicklung aus Abschnitt 2.5.3 mit dem Dekorrelator  $\hat{K}$  aufzeigen.

Stellen wir uns zunächst vor, dass der Ein-Teilchenraum  $\mathcal{V}^1$  von der Basis  $\{\xi_i\}_{i=1}^N$  aufgespannt würde. Als Basisfunktionen wählen wir dazu die  $N$  Lösungen des Fockoperator mit den niedrigsten Eigenwerten. Da wir anschließend im Rahmen der HF-Gleichung nur den antisymmetrisch vertensorierten Raum  $\mathcal{V}_A^{(N)} = \bigotimes_A^N \mathcal{V}^1$  betrachten, ist mit Beispiel 3 aus Abschnitt 2.3.4 klar, dass dieser Raum  $\mathcal{V}_A^{(N)}$  von nur einer einzigen Basisfunktion aufgespannt wird. Das bedeutet, dass alle Funktionen dieses Raums Produktstruktur besitzen bzw. unkorreliert sind. Der Referenzzustand schreibt sich schlicht als

$$\psi^{(\text{SCF})} = c \cdot \hat{A}(\xi_1(\mathbf{r}_1) \cdot \dots \cdot \xi_N(\mathbf{r}_N)) \quad (8.3.1)$$

mit dem festen Entwicklungskoeffizienten  $c = 1$ , wobei der Antisymmetrisierer mit der Normierung  $\frac{1}{\sqrt{N!}}$  anstelle von  $\frac{1}{d!}$  in (2.1.5b) wegen der Normierung des Betragsquadrats benutzt wird, siehe (4.1.3).

Der Anregungsoperator definiert also zunächst eine Erweiterung der Basis, in dem er in dem Produkt aus Basisfunktionen in (8.3.1) eine Funktion  $\{\xi_{N+1}\}$ ,

$$\hat{T}_1^{N+1} \psi^{(\text{SCF})} = \hat{A}(\xi_{N+1}(\mathbf{r}_1) \xi_2(\mathbf{r}_2) \cdot \dots \cdot \xi_N(\mathbf{r}_N)),$$

oder mehrere Funktionen  $\{\xi_{N+1}, \xi_{N+2}, \dots\}$  austauscht. Dadurch erzeugt er neben der einen Basisfunktion, welche den Referenzzustand beschreibt, neue, aufgrund der Orthogonalität der Lösungen des Fockoperators wohldefinierte Basisfunktionen.

Wir betonen, dass damit eben nicht die Basis des  $\mathcal{V}^{(1)}$  erweitert wird, sondern direkt die des  $\mathcal{V}_A^{(N)}$ .

Betrachten wir damit nun die Korrelationsentwicklung aus Abschnitt 2.5.3. Offensichtlich wirken dann Anregungsoperator und Dekorrelator genau konträr. Erstere projizieren den unkorrelierten Zustand in einen Raum  $\mathcal{V}_A^{(N)}$ , der auch Funktionen ohne völlige Produktstruktur enthält. Dazu benötigen sie zusätzlich die Anregungsamplituden, welche sie aus Truncated CI oder MPPT erhalten. Diese entsprechen schlicht Entwicklungskoeffizienten für Basisfunktionen außerhalb des Teilraums der einzelnen Slater-Determinante.

Letztere projizieren vom Raum  $\mathcal{V}_A^{(N)}$  in Teilräume  $\bigotimes_{A}^{N-k} \mathcal{V}^{(1)} \bigotimes_{A} \mathcal{V}^k$ , welche nur antisymmetrische Funktionen mit einer Produktstruktur beschreiben. Die Dekorrelatoren projizieren damit auf den Referenzzustand zurück und entfernen genau diese Anregungsamplituden aus der Darstellung.

Wir erhalten also die weiteren Funktionen in der Korrelationsentwicklung genau aus den Funktionen mit passendem Anregungsoperator und Anregungsamplitude. Die Güte der Approximation durch die trunkierte Korrelationsentwicklung misst sich dabei an dem Betrag der darüber hinaus weg gelassenen Anregungsamplituden.

Abschließend wollen wir die Berechnung der Korrelationsenergie im Rahmen der BOSSA-NOVA-Methode kurz erläutern. Durch die Zerlegung der Menge der Kernindizes in überlappende Unterindexmengen wird die elektronische Wellenfunktion in Teile mit bestimmter Produktstruktur im Rahmen der gewählten lokalen Basis unterteilt. Diese Unterteilung entspricht dann den zuvor im Rahmen der Größenkonsistenz betrachteten Systeme  $A$  und  $B$ . Anregungen würden also auf die Wirkung innerhalb eines Fragments beschränkt.

Da jedoch auch die Korrelation in Nicht- und Halbleitern kurzreichweitig ist, siehe [142], kann mit geschickter Wahl der Fragmente, die Korrelationsenergie mit einem Fragmentansatz berechnet werden. Da sich die Korrelation aber nicht notwendigerweise entlang kovalenter Bindungsstrukturen orientiert, muss hier ein anderer Abhängigkeitsgraph benutzt werden.

Wir gehen dazu in Abschnitt 9.4.6 auf eine Methode ein, die Energie auf Basis lokal beschränkter, aber dynamischer Korrelation der Elektronen zu berechnen. So besteht die Möglichkeit, bspw. van-der-Waals-Potentiale mit entsprechendem Trunkierungsradius zu ermitteln, wobei die Methode dennoch lineare Komplexität behält.

## 8.4. Verwandte Literatur

Der Fragmentierungsansatz ist auch innerhalb der Post-HF-Methoden angewandt worden. Wir wollen hier kurz einige Artikel zusammenfassen, aber können die Literatur nicht umfassend wiedergeben.

### 8.4.1. Fragmentansatz

Flocke und Bartlett resümieren in [57] die Bemühungen, linear skalierenden CC-Berechnung zu ermöglichen. Sie stellen fest, dass die Hauptschwierigkeit darin besteht, eine linear skalierende SCF-Rechnung als Ausgangsbasis für die CC-Methode zu haben. Sie lösen diese Schwierigkeit mit Hilfe des Fragmentansatzes und teilen dazu das molekulare System, und nicht die Basis, in kleine Untereinheiten auf. Hauptzweck dieser Art der Aufteilung ist der Erhalt lokalisierter Orbitale. Diese Orbitale werden anschließend auf das Gesamtsystem übertragen. Die Hauptschwierigkeit liegt aber in der Wahl gut übertragbarer Orbitale, wofür die Autoren eine spezielle lokalisierte Form vorschlagen, in der die Ein-Elektronen-Dichtematrix diagonal ist. Anschließend werden die Anregungsamplituden für den vollen Anregungsoperator aus den Anregungsamplituden approximiert, welche nur virtuelle Orbitale in Nachbarschaft zum gewählten lokalisierten Orbitale einbeziehen. Abschließend werden die Beiträge zur Korrelationsenergie aller lokalisierten Orbitale aufsummiert. Dabei kann jedoch die Wellenfunktion nicht direkt angegeben werden, da sich die Nachbarschaften dieser Orbitale überlappen. Abschließend betrachten die Autoren Kohlenwasserstoffketten und Polyglycin. Dabei werden die Untereinheiten mit chemischem Wissen gewählt und müssen bis zu 5 Bindungen einschließen, um Abfallverhalten in der Energie zu zeigen. Sie schlagen weiterhin vor, aromatische Ringe als Ganzes in den Untereinheiten mit einzubeziehen.

### 8.4.2. Inkrementenmethode

In [142, Abschn. 2.2.2] wird eine Übersicht über lokale Korrelationsmethoden für Festkörper gegeben. Diese Methoden basieren auf der Beobachtung, dass Elektronkorrelation in Nicht- und Halbleitern ein kurzreichweitiges Phänomen ist. Bei der diskutieren Inkrementenmethode werden lokalisierte Orbitale benutzt. Der volle Hamiltonoperator besteht dann neben den HF-Anteilen aus Anregungen von bis zu vier verschiedenen Orbitalzentren. Anschließend werden Näherung diskutiert, wobei diese Summation über bis zu vier Zentren schon vorher trunkiert wird. Es ergeben sich Gleichungen für die Korrelationsenergie in Übereinstimmung mit (7.2.6a) und (7.2.6b), aber nicht über Kernindizes, sondern über die Indizes zu lokalisierten Orbitalen. Auch Paulus verwendet einen Fragmentansatz, um Schwierigkeiten mit den periodischen Randbedingungen für Festkörper zu überwinden. Dazu werden kleine Fragmente aus der periodischen Zelle des Festkörpers extrahiert und im Falle von dabei durchtrennten kovalenten Bindungen mit Wasserstoffen saturiert. Je größer das Fragment, desto besser wird dabei die Korrelationsenergie approximiert. Bei ionischen Bindungen werden entsprechende Punktladungen anstelle von Wasserstoffen verwendet. Für metallische Systeme wird eine Minimalbasis verwendet, so dass es keine unbesetzten Orbitale gibt, jedoch hat das Fragment im Allgemeinen keinen metallischen Charakter mehr, siehe auch Riedel und Janiak [147, S. 188] zur notwendigen Größe von Metallclustern.



### 8.4.3. Dünne Gitter und Kombinationstechnik

Ein ganz anderes Vorgehen als die vorherigen Methoden beschreiben Garcke und Griebel [63], wo die volle elektronische Schrödingergleichung mittels einer Finite-Elemente-Methode in einem Galerkin-Ansatz gelöst wird. Dabei werden aber die Lösungen in nodalen, globalen Basen verschiedener Gitterdiskretisierungen geschickt miteinander kombiniert, so dass sie in der Approximationsqualität einem Dünngitter-Ansatz entsprechen, siehe [27]. Fasst man diese sogenannte *Kombinationstechnik* als eine Reihe von Projektoren auf, so ergibt sich, dass sich gleiche Approximationsgüte nur ergibt, wenn diese Projektoren miteinander kommutieren. Garcke beschreibt in [62], wie man durch Lösen eines Eigenwertproblems, den Fehler im nicht-kommutierenden Falle minimieren kann. Dabei besteht eine formale Gemeinsamkeit in der hierarchischen Kombinationsformel mit (2.5.5a) und (2.5.5b).

## Zusammenfassung

In diesem Kapitel untersuchten wir auf Basis einer Closed-Shell Hartree-Fock (CLHF)-Rechnung mit einer Slater-Determinante, wie die durch diesen Ansatz fehlende Korrelationsenergie berechnet werden kann.

Wir betrachteten dazu den Configuration Interaction (CI)-Ansatz und die Møller-Plesset Störtheorie (MPPT). Erstere schied für unsere in Kapitel 7 vorgestellte Methode aus, da sie nicht größenkonsistent ist. Die MPPT hat diese Eigenschaft und eignet sich damit sehr gut für unseren summativen Fragmentansatz. Die Grundidee bei beiden Methoden ist die Erweiterung der einzelnen Slater-Determinante, dem Referenzzustand, um weitere Vielteilchenfunktionen. In diesen kann dann eine Lösung der elektronischen Schrödingergleichung dargestellt werden, welche auch dynamische Korrelation der Elektronen beinhaltet. Die anderen Slater-Determinanten werden dabei formal mittels eines Anregungsoperators erzeugt, welcher auf den Referenzzustand wirkt. Durch Lösen der Gleichung müssen dabei die fehlenden Anregungsamplituden, die Entwicklungskoeffizienten der Slater-Determinanten für den Vielteilchenzustand, ermittelt werden.

Abschließend erörterten wir, wie sich diese Linearkombination von Slater-Determinanten, welche den Vielteilchenzustand darstellt, zu der von uns in Kapitel 2 beschriebenen Korrelationsentwicklung verhält. Wir erkannten, dass der Anregungsoperator genau das Gegenstück zum Dekorrelator in der von uns beschriebenen Korrelationsentwicklung ist. Mittels des Dekorrelators hatten wir in Kapitel 3 ein Ansatz beschrieben, um zunächst die volle elektronische Schrödingergleichung in das einfachere HF-Modell zu überführen. Mit dem Anregungsoperator lässt sich die Lösung dieses einfacheren Modells wieder in eine Lösung für das ursprüngliche Problem zurücktransferieren.

Schließlich gaben wir stellvertretend einige Fragmentansätze aus der Literatur wieder, welche ebenfalls die Korrelationsenergie berechnen und dabei die Lokalität dieser Wechselwirkung zugrunde legen. In Abschnitt 9.4.6 werden wir einige numerische Experimente mit der Methode aus dem letzten Kapitel betrachten.



## 9. Experimente

Wir hatten die Komplexität der Hartree-Fock (HF)-Methode in Abschnitt 4.3 und Kapitel 7 analysiert. Wir hatten festgestellt, wenn die Lösung bestimmte, physikalisch fundierte und asymptotisch geltende Annahmen erfüllt, dass wir eine Methode erhalten, deren Kosten linear mit der Anzahl der Kerne skalieren. Dabei wird das molekulare System in überlappende Teilsysteme aufgespalten und diese einzeln berechnet. Wir betrachteten dann drei Varianten: Bei der ersten wird die langreichweitige Wechselwirkung zwischen den Teilproblemen vernachlässigt. Bei der zweiten wird diese abschließend mittels einer Rechnung in erster Störtheorie abgeschätzt. Bei der dritten Variante wird diese Wechselwirkung voll mitberücksichtigt, dafür muss der selbst-konsistente Zyklus weiterhin über die globale Lösung erfolgen. Für Einzelheiten verweisen wir auf Kapitel 7.

Wir untersuchen im Folgenden eine von uns erstellte Implementation in C++ der ersten und der zweiten Variante. Bei diesen beiden kann ein beliebiger Standard-HF-Löser verwendet werden<sup>1</sup>. Für die nötige Störrechnung verwenden wir die in [104] implementierte Mehrgitter-Methode als Abschätzung der langreichweitigen Wechselwirkung.

Wir wollen dabei die folgenden Fragestellungen beantworten:

1. Skaliert die Implementation wirklich linear mit der Anzahl der Kerne?
2. Wie genau approximiert die Methode die Grundzustandsenergie?
3. Wie brauchbar ist die Methode bei alltäglichen Systemen?
4. Wie gut wird die langreichweitige Energie durch die Störrechnung abgeschätzt?

Dieses Kapitel gliedert sich dazu in die folgenden Abschnitte: Wir beginnen mit einigen Vorüberlegungen in Abschnitt 9.1 zu unseren Mess- und Implementationsgrundsätzen, sowie zur nötigen Saturation der Teilprobleme, um im Anschluss MPQC [108] als Closed-Shell Hartree-Fock (CLHF)-Löser benutzen zu können. Außerdem untersuchen wir die theoretische Komplexität der vorgestellten Methode anhand der verwendeten Algorithmen. Anschließend messen wir in Abschnitt 9.2 die Komplexität der einzelnen Schritte der Implementation. Diese bestehen aus: Graphenerstellung, Analyse des Graphen und dessen Zerlegung, Fragmentierung, Berechnung der Fragmentprobleme und abschließende Berechnung der Summe über die Teillösungen. Wir werden für jeden Schritt bestätigen, dass er lineare Komplexität hat und dabei auch die Effizienz der Parallelisierung der Implementation untersuchen. In Abschnitt 9.3 gehen wir auf implementierte Heuristiken zur Erstellung des Bindungsgraphen ein, welche nur auf der Kenntnis der Geometrie des molekularen Systems basieren. Im Anschluss widmen wir uns in Abschnitt 9.4 der

---

<sup>1</sup>Analog lässt sich auch ein DFT-Löser verwenden, erste Tests der Methode basierten noch auf dieser Theorie.

## 9. Experimente

Approximationsgüte der vorgestellten Methode in der ersten Variante und untersuchen eine Reihe von Systemen auf ihren Abfall im Fehler der Gesamtenergie bezüglich der Bindungsordnung  $k$ . Dabei untersuchen wir gezielt Stärken und Schwächen des Ansatzes. Außerdem belegen wir auch die Eignung zur Berechnung der Korrelationsenergie im Rahmen der MP2-Theorie. Um die zweite Variante der Methode und die darin zusätzlich enthaltene Berechnung der langreichweitigen Wechselwirkung in erster Ordnung Störtheorie geht es im Abschnitt 9.5. Dort untersuchen wir die Approximationseigenschaften dieses Teils relativ zu einer vollen HF-Rechnung. Abschließend betrachten wir noch einige besondere Aspekte der Methode in Abschnitt 9.6, wo wir die Spektren von Alkanen untersuchen und schließlich beispielhaft die Grundzustandsenergie einige sehr großer Biomolekülen berechnen.

Wir gehen dabei im Rahmen jeder Messung am Anfang eines Abschnitts, meist kurz, auf Details der Implementation ein, welche zum Verständnis der gemessenen Laufzeiten und Energien relevant sind.

### 9.1. Vorüberlegungen

Zunächst stellen wir einige Vorüberlegungen zu Messprinzipien und zur Skalierungskomplexität der verwendeten Algorithmen im Allgemeinen an. Anschließend gehen wir auf die Saturation der Fragmente mit Wasserstoff ein. Schließlich untersuchen wir die theoretische Skalierungskomplexität von der Seite der verwendeten Algorithmen.

#### 9.1.1. Messprinzipien

Um Reproduzierbarkeit sicherzustellen, werden bei physikalischen Messungen die benutzten Apparaturen und Prozeduren genau aufgelistet werden. Hersteller sind dabei für präzise Eichung und auch Wartung dieser Geräte verantwortlich.

Dieses Prinzip haben wir auch für die im Folgenden durchgeführten Experimente verfolgt:

- Vollständige Automatisierung jedes Experiments in Bezug auf Ablauf und Parameterdokumentation
- Durchführen der Experimente in stets drei getrennten Schritten: Vorbereitung, Durchführung und Analyse
- Jeder Schritt überprüft, ob notwendige Voraussetzungen erfüllt sind
- Festhalten des Experimentverlaufs in Protokollen und der Enddaten in Versionskontrollsystemen
- Strikte Trennung der Experimente auf dem Dateisystem und nach ihren Fragestellungen

Experimentieren bedeutet zunächst nichts anderes als Fehler machen. Experimente am Rechner erfolgen auf völlig analoge Weise. Hier ergeben sich vielerlei Fehlerquellen, die

jedoch nicht notwendigerweise der wissenschaftlichen Erkenntnis dienlich sind: Programmierfehler, falsche Parameterwahl oder auch besonders Tippfehler. Der wichtige Punkt dabei ist, diese Fehler zu dokumentieren und aus ihnen die richtigen Schlüsse zu ziehen. Die obigen Richtlinien sollen daher vor allem Wiederholbarkeit der Versuche und gleichzeitige Dokumentation der Fehler erleichtern und dabei vermeiden, dass Fehler, vor allem unbemerkt, wiederholt werden.

Bei der Programmierung haben wir uns folgende Leitlinien zugrunde gelegt:

- Weitest mögliche Absicherung gegen Fehler durch allgegenwärtige Komponenten- und Regressionstests
- Benutzung von vielgetesteten und erprobten Standardkomponenten wie die STL bzw. die Boost-Bibliotheken, die gleichzeitig bestimmte Komplexitäten garantieren
- Überprüfung der geforderten Komplexität durch Messung
- Festhalten der Softwarehistorie durch Versionskontrollsysteme und Versionierung

Dies sind die Prinzipien wie sie für das ingenieurhafte Schreiben von Software gelten. Nur durch solch einen Ansatz gelingen große Softwareprojekte in abschätzbarer Zeit. Numerische Simulationen und die dazugehörigen Softwarepakete haben vergleichbaren Umfang und Komplexität. Gleichzeitig folgen die obigen Prinzipien den gängigen Regeln wissenschaftlicher Praxis und maximieren die Vergleichbarkeit und Reproduzierbarkeit der erhaltenen Daten.

### 9.1.2. Wasserstoffsaturation

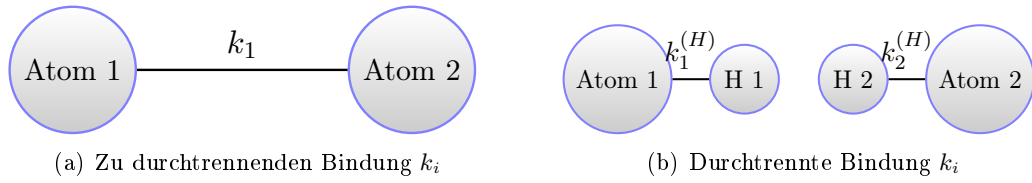
Wir wollen im Folgenden einen CLHF-Löser zur Bestimmung der Lösungen der Fragmente verwenden, siehe Abschnitt 7.5. Wenn wir jedoch verbundene Teilgraphen aus dem Bindungsgraphen extrahieren, dann müssen dabei im Allgemeinen Kanten, also Bindungen, durchtrennt werden. Ein solches Fragment hat damit aber nicht mehr den für diesen Lösertyp erforderlichen Gesamtspin null, denn nur solche Systeme sind mit einer einzigen Slater-Determinante darstellbar.

Die in der Literatur gängige Lösung, welche bereits in [130] vorgeschlagen wurde, ist, den jeweils ausgeschlossenen zweiten Knoten im Graph zu einer durchtrennten Kante durch einen Abschlussknoten zu ersetzen. In anderen Worten: Das abgeschnittene Atom wird durch Wasserstoff ersetzt. Darauf gehen wir Abschnitt 9.3 noch näher ein. Dieser Bindungsabschluss soll dabei möglichst wenig Einfluss auf das Fragment selbst entnehmen. Genau aus diesem Grund wählt man Wasserstoff als Saturelement, denn es kommt dieser Forderung am ehesten nahe<sup>2</sup>. Dies ist in Abbildung 9.1 veranschaulicht.

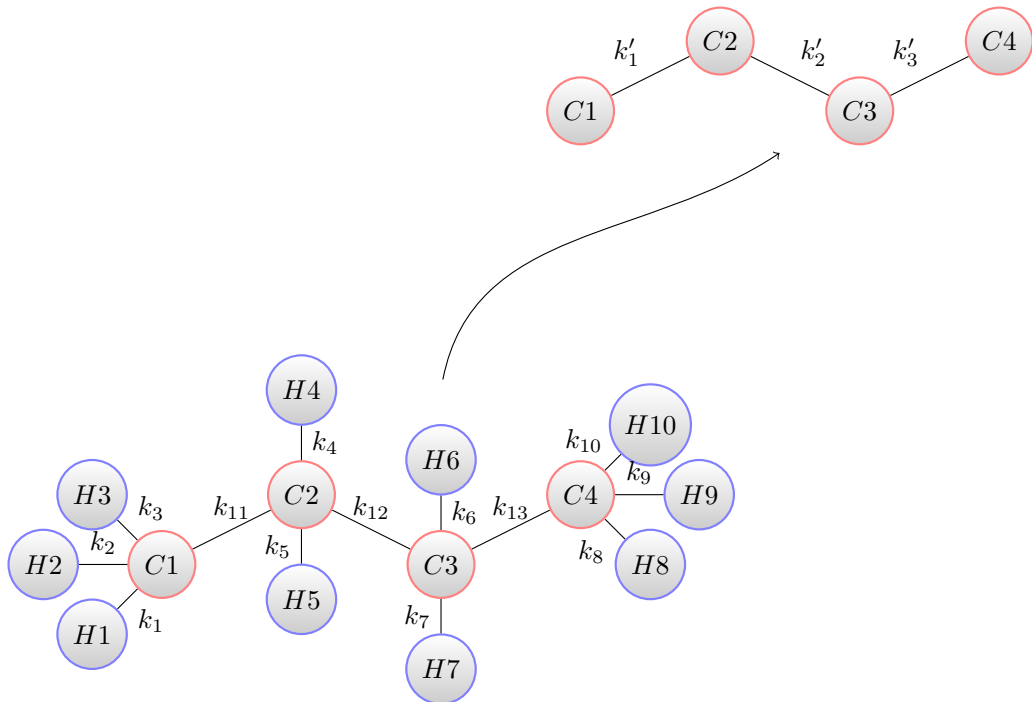
Dadurch erhalten schon vorhandene Wasserstoffatome eine Sonderstellung im Bindungsgraphen, denn Bindungen zu durchtrennen, an denen sie beteiligt sind, macht keinen Unterschied mehr, weil ein Saturelementwasserstoff an ihrer Stelle platziert würde.

<sup>2</sup>Diese Eigenschaft wird bspw. durch die relativ geringe chemische Verschiebung im Rahmen von NMR-Messungen an Molekülen belegt, siehe [90].

## 9. Experimente



**Abbildung 9.1.:** Eine Bindung  $k_i$  zwischen zwei Kernen wird durchtrennt und durch zwei Bindungen  $k_1^{(H)}$  und  $k_2^{(H)}$  zu hinzugefügten Wasserstoffen (H1 und H2) ersetzt.



**Abbildung 9.2.:** Wasserstoffe (blau umrandet) werden mit ihren Bindungsnachbarn (rot umrandet) zu neuen Knoten im Bindungsgraph zusammen gefasst. Die verbleibenden Kanten und Knoten werden dabei umbenannt.

Aus diesem Grund bietet es sich an, im Bindungsgraphen Wasserstoffe mit ihren Bindungsnachbarn zusammen zu fassen. Da sie im Allgemeinen nur eine kovalente Bindung eingehen, ist dies eindeutig. Ein Beispiel für einen solchen veränderten Bindungsgraphen geben wir in [Abbildung 9.2.](#)

### 9.1.3. Analyse der Skalierungskomplexität

In diesem Abschnitt geht es um die Analyse der Skalierungskomplexität der verwendeten Algorithmen auf Graphen. Wir wollen dabei zeigen, dass die implementierte Methode in der Theorie linear mit der Anzahl der Kerne<sup>3</sup> skalieren sollte.

Dazu wiederholen wir zunächst allgemein die Komplexität der verwendeten Graphenalgorithmen. Weiterhin gehen wir aber auch auf die Anzahl der resultierenden verbundenen Teilgraphen, der Fragmenten, ein, da wir sicherstellen müssen, dass die Anzahl dieser Teilprobleme ebenfalls linear mit der Anzahl der Kerne ansteigt.

#### Graphenalgorithmen

Die grundlegenden Algorithmen der Methode bearbeiten Graphen. Wir benötigen diese zur Darstellung der Bindungsabhängigkeit bzw. der Korrelation der verschiedenen elektronischen Orbitalfunktionen.

Die verwendeten Algorithmen sind dabei die Breitensuche [37, Abschn. 22.2] und die Tiefensuche [37, Abschn. 22.3] mit einer Komplexität  $\mathcal{O}(|\mathcal{V}| + |\mathcal{E}|)$  für die Suche von einem einzelnen Knoten  $s$  aus für eine Knotenmenge  $\mathcal{V}$  und Kantenmenge  $\mathcal{E}$ .

Wird der Erkundungshorizont durch  $k \sim \mathcal{O}(1)$  begrenzt, dann hat die Breitensuche auf einer Knoten-  $\mathcal{V}'$  und Kantenteilmenge  $\mathcal{E}'$  mit beschränktem Knotengrad  $d(v)$  leicht ersichtlich ein Laufzeitverhalten von  $\mathcal{O}((\max_{v \in \mathcal{V}} d(v))^k) = \mathcal{O}(1)$  pro Knoten  $s$ . Damit ergibt sich insgesamt eine Laufzeit, die linear mit der Anzahl der Kerne bei beschränkter maximaler Kantenzahl steigt.

Weiterhin verwenden wir aber auch den Maximum Matching-Algorithmus aus der Boost Graph Library [155], um die Bindungsgrade des Graphen robust zu bestimmen. Dies ist notwendig, um letztlich ladungsneutrale Teilprobleme zu erhalten. Dieser Algorithmus skaliert wie  $\mathcal{O}(|\mathcal{V}| \cdot |\mathcal{E}|) = \mathcal{O}(M^2)$ , also quadratisch mit der Anzahl der Knoten. Ein einfacherer, aber nicht robuster Algorithmus mit linearer Laufzeit in der Kantenzahl existiert ebenfalls. Wir werden später anhand einzelner Beispiele untersuchen, inwieweit sich die mögliche quadratische Laufzeit des Maximum Matching-Algorithmus auf die gesamte Laufzeit auswirkt.

#### Anzahl der Fragmente

Wir untersuchen nun die Anzahl der Teilprobleme, welche wir aus dem Graphen des molekularen Systems extrahieren.

Generell ist die Maximalzahl möglicher Fragmente eines Systems aus  $M$  Kernen durch die Potenzmenge der Mächtigkeit  $2^M$  bestimmt. Im Allgemeinen erhalten wir dann für die Teilmenge der Potenzmenge, welche nur Mengen bis  $k$  Kernen enthält eine Mächtigkeit von  $\sum_{l=0}^k \frac{M!}{l!(M-l)!} \approx M^k$  für kleines  $k$ . Dies würde bedeuten, dass mit  $k > 3$  die Anzahl der Teilprobleme so groß werden würde, dass die Methode langsamer als die Standard-HF-Methode ist.

<sup>3</sup>Wir merken an dieser Stelle noch einmal an, dass wir mit „Kern“ stets das Atom ohne Hülle und nicht den Prozessorkern meinen. Im Rahmen von Parallelisierung sprechen wir von mehreren „Prozessen“.

## 9. Experimente

Jedoch werden bei unserem Ansatz alle Fragmente vernachlässigt, die nicht einen verbundenen Untergraphen des molekularen Systems darstellen. Deswegen ist die wahre Zahl sehr viel kleiner, wie wir mit folgendem Lemma aufzeigen. Dies geben wir hier ausführlich wieder, da unsere Implementierung sehr nah dem konstruktiv gegebenem Beweis folgt und wir deswegen auf deren Details nicht weiter eingehen werden.

**Lemma 14** (Anzahl der verbundenen Untergraphen). *Sei ein verbundener Graph  $G = (\mathcal{V}, \mathcal{E})$  gegeben. Sei die Kantenzahl pro Knoten durch  $c > 0$  nach oben beschränkt. Dann kann die Anzahl der induzierten Untergraphen  $G' = (\mathcal{V}', \mathcal{E}')$ , die einen bestimmten Knoten  $s \in \mathcal{V}$  enthalten, die verbunden sind und deren Knotenzahl  $|\mathcal{V}'| \leq k$  durch eine Ordnung  $k$  beschränkt sind, durch*

$$\sum_{j=1}^{k-1} 2^{c \cdot (k-j)} = \sum_{j=1}^{k-1} (C)^{k-j} = \sum_{j=1}^{k-1} C^{k-j} < \sum_{j=0}^{k-1} C^j \stackrel{C \geq 2}{=} \frac{C^k - 1}{C - 1} \leq C^k, \quad (9.1.1)$$

mit  $C := 2^c$  nach oben abgeschätzt werden.

*Beweis.* Sei  $s \in \mathcal{V}$  gegeben. Wir teilen die Kanten um  $s$  in Ebenen gleicher Entfernung ein: Sei  $\mathcal{E}_s(j)$  die Endkanten der kürzesten Wege zu Knoten  $v \in \mathcal{V}$  in Entfernung  $d(s, v) = j$ .

Betrachten nun wir einen möglichen Untergraphen  $G'$  mit  $s \in \mathcal{V}'$ . Sei dann  $\mathcal{E}'_s(j)$  die reduzierte Kantenmenge zu  $\mathcal{E}_s(j)$ , für die nur einer der beiden zugehörigen Knoten in  $\mathcal{V}'$  liegt, siehe Abbildung 9.3 zur Illustration. Dies ist der Entdeckungshorizont von  $G'$  in Entfernung  $j$  zu  $s$ .

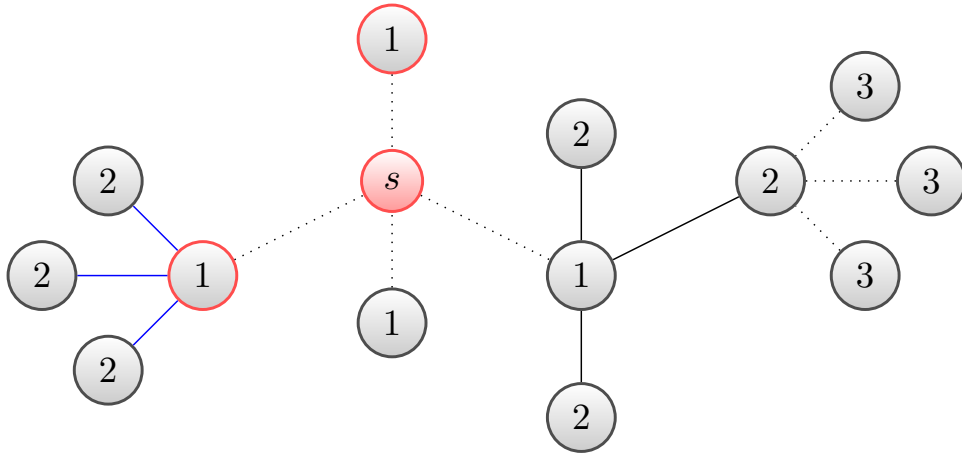
Die Kardinalität einer Potenzmenge der reduzierten Kantenmenge  $E'_s(j)$  ist  $2^{|\mathcal{E}'_s(j)|}$  für die Ebene  $j$ . Damit erhalten wir in der Summe über alle möglichen  $k - 1$  Ebenen insgesamt  $\sum_{j=1}^{k-1} 2^{|\mathcal{E}'_s(j)|}$  Möglichkeiten, wobei wir ignorieren, dass die Anzahlen pro Ebene nicht unabhängig voneinander sind. Mit der oberen Schranke des Bindungsgrades gilt, dass  $|\mathcal{E}'_s(j)|$  durch  $c \cdot |I_{j-1}|$  beschränkt ist, wobei  $I_{j-1}$  die Menge der Knoten, welche auf der Ebene  $j - 1$  hinzugefügt wurden, bezeichnet. In Abbildung 9.3 sind dies die mit „1“ beschrifteten, rot umrandeten Knoten. Weiterhin ist  $|I_j|$  beschränkt durch  $k - j$ , denn zumindest ein Knoten muss pro darunterliegender Ebene hinzugefügt worden sein und es gibt außerdem schon einen Wurzelknoten. Dann folgt das Gewünschte mit Kenntnis der Partialsumme der geometrischen Reihe.  $\square$

Die Summe der möglichen Untergraphen mit maximal  $k$  Knoten ist also nur abhängig von der Bindungsordnung  $k$  und dem höchsten Bindungsgrad  $c$  über alle Knoten  $v$  in  $G$ . Da wir über alle möglichen Knoten  $s \in V$  als Wurzelknoten summieren müssen, skaliert die Anzahl der Teilprobleme mit

$$\mathcal{O}(M \cdot C^k). \quad (9.1.2)$$

Durch das Kriterium nicht-verbundene Teilgraphen auszulassen, erhalten wir also eine nur noch linear mit der Anzahl der Kerne ansteigende Anzahl an Fragmente oder Teilproblemen.





**Abbildung 9.3.:** Darstellung der reduzierten Kantenmenge  $\mathcal{E}'_s(2)$  in blau zu einem gegebenen Untergraphen  $G' \subset G$  bestehend aus den rot umrandeten Knoten um  $s$ . Die Knoten des Graphen  $G$  mit  $c = 3$  wurden dabei mit ihrer Entfernung zu  $s$  beschriftet, alle Kanten außerhalb von Kantenmenge  $\mathcal{E}_s(2)$  sind nur gepunktet dargestellt.

## 9.2. Skalierungsverhalten der Implementierung

In diesem Abschnitt untersuchen wir das Skalierungsverhalten der Implementierung. Dabei muss zunächst das molekulare System geladen und falls nicht in der Datei gespeichert dessen Bindungsgraph erkannt und auf Zyklen analysiert werden. Der Graph wird in einzelne verbundene Untergraphen aufgetrennt. Anschließend werden alle verbundenen Teilgraphen bis zu einer festgelegten Knotenzahl aus diesem Bindungsgraph extrahiert und als Teilprobleme mittels MPQC [108] berechnet. Abschließend werden die Teillösungen zur Näherungslösung summiert.

Die Qualität dieser Näherung wird dabei Ziel der Untersuchung im nächsten Abschnitt sein. Hier geht es darum zu bestätigen, dass die Rechenzeit nur linear von der Anzahl der Kerne  $M$  abhängt, also dass die Implementierung unsere theoretischen Erwartungen an die Komplexität der Methode widerspiegelt.

**Erwartung an die Komplexität der Implementierung** Wir merken hierbei an, dass wir aufgrund verwendeter Datenstrukturen mit Sortierung wie der `std::map` der STL oder auch Algorithmen wie dem Linked-Cell-Verfahren ein Verhalten von  $\mathcal{O}(M \log(M))$  und nicht  $\mathcal{O}(M)$  erwarten. Es existieren Sortieralgorithmen mit einer Komplexität von  $\mathcal{O}(M)$ , siehe Cormen et al. [37, Kap. 8], da die Kernindizes aus einem beschränkten und bekannten Intervall  $[1, \dots, M]$  stammen. Den Mehraufwand bei der Implementierung würde der Laufzeitgewinn jedoch nicht wettmachen.

## 9. Experimente

**Allgemeiner Versuchsablauf** Unsere Messungen werden dabei immer nach der gleichen Prozedur ablaufen: Wir untersuchen die Laufzeit  $T$  eines Programmteils in Abhängigkeit der Anzahl der Kerne  $M$ . Wir messen ein Verhalten

$$c \cdot M^\alpha \tag{9.2.1}$$

mit einer Konstante  $c > 0$  und erwarten die aus dem Fit zu bestimmende Steigung als  $\alpha \approx 1$  für linear skalierend, vgl. (9.1.2). Zunächst führen wir Messungen an ausgetesteten Molekülsystemen durch, die jeweils auf bestimmte mögliche Schwachstellen der Implementierung abzielen. Die Erstellung dieser Systeme wird dabei jeweils kurz nach der ersten Erwähnung in einem eigenständigen, eingerückten Abschnitt erläutert. Anschließend werden diese Laufzeiten in doppelt logarithmischen Plots dargestellt. Aus den gemessenen Laufzeiten  $T(M)$  bestimmen wir eine Gerade mittels linearer Regression an den relevanten Teil der Kurve und notieren uns die resultierende Steigung  $\alpha$ . Abschließend vergleichen wir die gemessene Komplexität (9.2.1) mit der erwarteten Komplexität (9.1.2).

**Laufzeitmessung** Die Laufzeit eines Programms wird mittels des Linux-Programms *time* nach POSIX-Standard mit Hunderstelsekundengenauigkeit bestimmt. Einzelne Teile der Implementierung werden außerdem mit einem internen Zeiterfassungssystem auf Basis der Funktion „clock\_gettime()“ ausgemessen, welche eine Zeitauflösung in Prozessoraktsschritten, also genauer als Mikrosekunden, bietet. MPQC [108] erfasst ebenfalls die benötigte Zeit intern und gibt sie mit Hunderstelsekundengenauigkeit aus. Jede Messung wird jeweils dreimal durchgeführt und daraus Mittelwert und Abweichung bestimmt, um mögliche Laufzeitschwankungen zu erfassen.

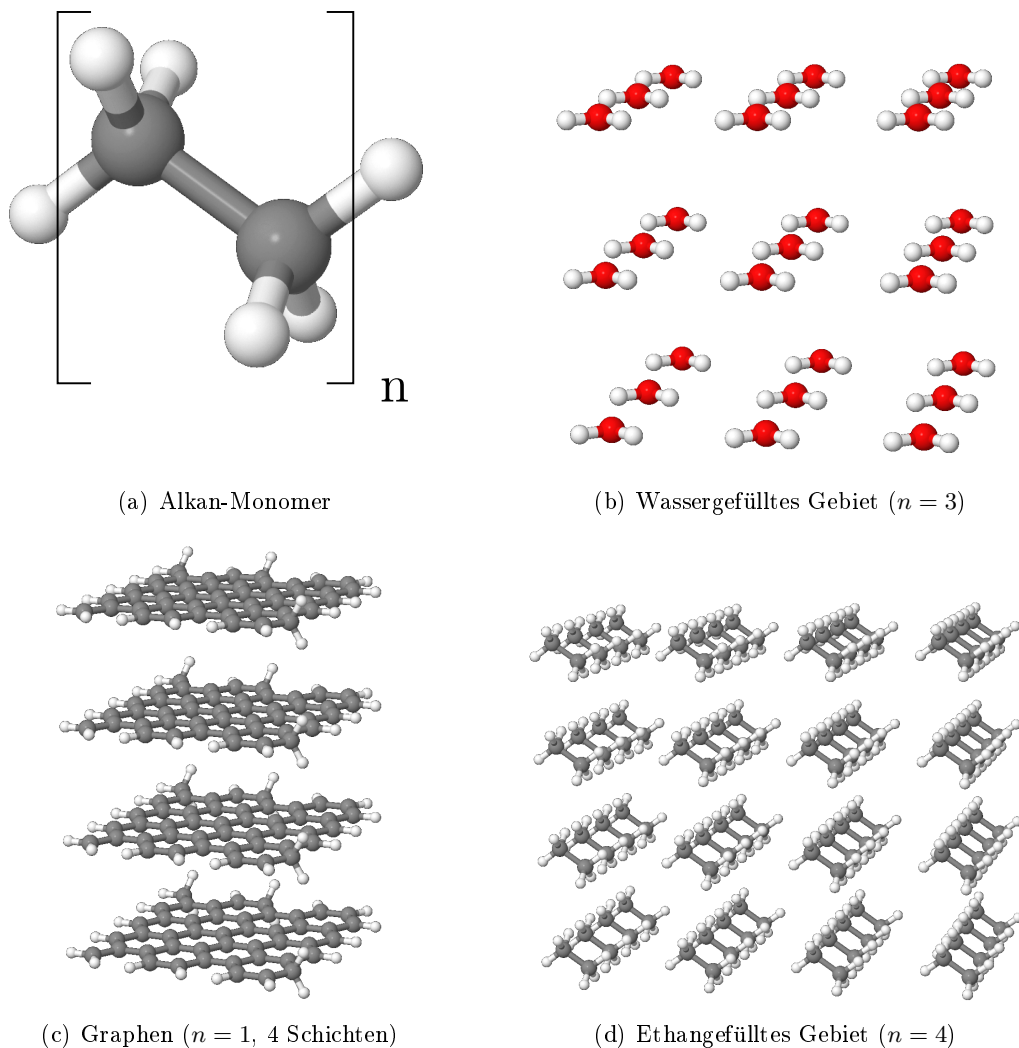
**Technische Details** Alle Berechnungen dieses Versuchsteils wurden auf einem AMD64-System (AMD FX-6100) mit 6 Kernen und 20 Gb Speicher unter Ubuntu 12.04 durchgeführt. Dabei wurden MoleCuilder in der Version 1.4.1, JobMarket in der Version 1.1.4, CodePatterns in der Version 1.2.7, Boost in der Version 1.47 und MPQC [108] in der Version 2.3.1 benutzt.

### 9.2.1. Laden

Wir gehen in dieser Analyse auf den stets notwendigen Ladeschritt nicht weiter ein, welcher Kernort und Kernladung, sowie eventuelle weitere Parameter für jeden Kern aus einer Datei in den Hauptspeicher des Rechners überträgt. Es sei nur festgehalten, dass dieser Teil nach Messung linear mit Steigungen von  $\alpha = 1 \pm 0.02$  für verschiedene Systeme skaliert.

### 9.2.2. Graphenerstellung

**Implementierung** Der Bindungsgraph der geladenen Ansammlung von Atomen wird über ein Abstandskriterium im Rahmen des Linked-Cell-Verfahrens, siehe [77, Abschn. 3.4], bestimmt.



**Abbildung 9.4.:** Kugel-Stab-Modelle der verwendeten molekularen Struktur zum Messen der Skalierungskomplexitäten.

Wir erwarten stets lineare Komplexität, unabhängig ob rein baumartige oder auch zyklische Anteile im Graphen gefunden werden. Wir erwarten aber auch aufgrund des Linked-Cell-Verfahrens eine leichte Abhängigkeit von der Dichte der Kerne, da die Kantenbreite dieser kubischen „Zellen“ nur von der größten typischen Bindungslänge abhängt.

Unsere Testfälle zielen deswegen auf verschiedene Arten des Bindungsgraphen ab und verschiedene Füllungen des Gebiets ab: Ein Gebiet gefüllt mit Wassermolekülen, eine lange Alkankette und gestapelte Graphenschichten. Das letzte System besteht aus viele

## 9. Experimente

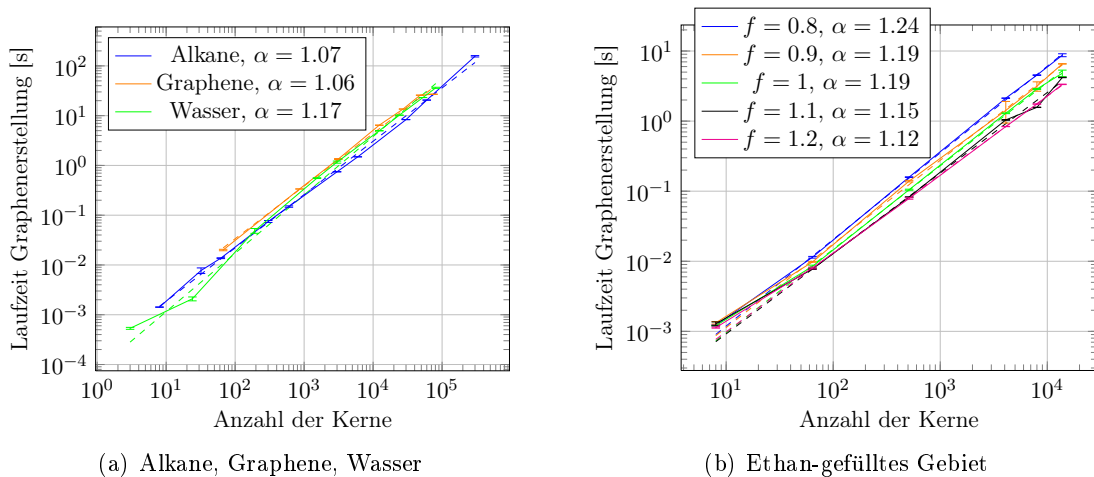
verschränkten Ringsystemen und unterscheidet sich daher durch seine höhere und zyklische Vernetztheit von Kohlenstoffatomen. Weiterhin haben wir mit diesen drei Testfällen Beispiele für 1d-, 2d- und 3d-Strukturen im Gebiet ausgewählt.

**Alkane** Wir benutzen Kohlenwasserstoffketten der Längen  $n$  von 1, 5, 10, 50, 100, 500, 1000, 5000, 10000, 50000. Dies resultiert jeweils in einem Alkan der Formel  $C_{2n}H_{2n+2}$ , siehe Abbildung 9.4(a).

**Wassergefülltes Gebiet** Die Wassermoleküle werden jeweils in ein kubisches Gebiet mit einem Abstand von  $3.1 \text{ \AA}$  zwischen den Molekülzentren eingefüllt. Das gefüllte Gebiet hat dann die Dichte von Wasser unter Standardbedingungen. Dabei werden 1, 2, 4, 8, 10, 16, 20, 26, 30 Moleküle pro Raumachse in einem regelmäßigen Gitter eingesetzt, also insgesamt  $3 \cdot n^3$  Atome, siehe Abbildung 9.4(b).

**Graphen** Die Graphenschichten werden aus periodisch replizierten Einheitszellen mit einer Chiralität von  $(6, 0)$  erstellt und aus dem resultierenden Gitter in verschiedenen quadratischen Größen ausgeschnitten. Die dabei gekappten Bindungen werden mit ein oder zwei Wasserstoffen saturiert, so dass sich ein insgesamt ladungsneutrales System ergibt. Die so erzeugten Graphenschichten haben eine Anzahl der Kerne von 67, 860, 3132, 12276, 27216, 48095 Atomen, siehe Abbildung 9.4(c).

In Abbildung 9.5(a) sind die Zeiten für die Erstellung des Bindungsgraphen in Abhängigkeit von der jeweiligen Anzahl der Kerne dargestellt. Die Steigung  $\alpha$  rührt jeweils aus der linearen Regression an (9.2.1) her.



**Abbildung 9.5.:** Laufzeiten für die Erstellung des Bindungsgraphen. In (a) werden verschiedene 1-, 2- und 3-dimensionale Systeme betrachtet. In (b) wird ein Ethan-gefülltes mit Dichte proportional zu  $1/f$  untersucht.

Der für die Graphenerstellung benutzte Linked-Cell-Ansatz hat eine Komplexität von  $\mathcal{O}(M \log(M))$  und hängt dabei insbesondere von der Struktur des Gebiets ab: In einem in alle Raumrichtungen gleichmäßig mit Wassermolekülen gefüllten Gebiet sind mehr Nachbarn zu überprüfen als entlang des 1-dimensionalen Strangs einer Alkankette, weswegen die Steigungen der gefitteten Geraden auch leicht ansteigen. Die Steigungen  $\alpha$ , welche leicht über der erwarteten linearen Komplexität von 1 liegt, hängen also vermutlich von der Dichte und Homogenität des gefüllten Gebiets ab.

Um diesen Umstand genauer zu untersuchen, untersuchen wir ein mit Ethan gefülltes Gebiet genauer, dessen Dichte wir variieren.

**Ethan-gefülltes Gebiet** Ethan ist das Alkan mit Länge  $n = 1$  und Formel  $C_2H_6$ , welches wir zuvor schon verwendet haben. Es wird wie das Wassermolekül entlang fester Gitterpunkte im Abstand  $(4f+1, 3f+1, 2f+1)$  mit Faktor  $f = 1$  zu seinen Nachbarn jeweils  $n = 1, 2, 4, 8, 10, 12, 16, 20$  pro Raumachse in das Gebiet eingefügt. Es ergeben sich  $8 \cdot n^3$  Atome, siehe Abbildung 9.4(d).

Wir haben die Laufzeiten für das System mit fünf verschiedenen Dichten durch unterschiedliche Faktoren  $f = 0.8, 0.9, 1, 1.1, 1.2$  gemessen. Wir sehen in Abbildung 9.5(b), dass sich mit abnehmender Dichte – also zunehmendem Faktor  $f$  – die Steigung  $\alpha$  dem linearen Fall nähert. Da die Kantenbreite der Zellen fest ist, sich aber die Dichte und damit die Anzahl der Kerne pro Zelle ändert, müssen mehr Nachbarn als mögliche Bindungspartner überprüft werden. Dies erklärt die zunehmende Abweichung von der 1 mit groß werdender Dichte. Der Anstieg kann nur am Linked-Cell-Verfahren liegen, da es die einzige Komponente ist, die von einer veränderten Dichte betroffen ist, da Atom- und Molekülzahl ansonsten konstant bleiben.

### Maximum Matching

Im Rahmen der Bindungsgraphenbestimmung wird bei der Bestimmung der einzelnen Bindungsordnungen auch ein „Maximum Matching“-Algorithmus benutzt, welcher eigentlich quadratische Laufzeit mit der Anzahl „ungesättigter“ Kanten hat. Diese Laufzeitabhängigkeit sollte besonders beim Graphen deutlich zu sehen sein, da hier stark vernetzte aromatische Netzwerke vorliegen, so dass eine große Kantenzahl in den Matching-Algorithmus eingeht. Wir bemerken jedoch keinen ungewöhnlichen Anstieg mit sehr großer Anzahl der Kerne. Wir vermuten also, dass die Skalierungskonstante des restlichen linearen Programnteils so viel größer ist als die für den Maximum Matching-Teil mit quadratischer Komplexität, dass letzter bei den untersuchten Systemgrößen nicht ins Gewicht fällt.

### 9.2.3. Analyse des Bindungsgraphen

**Implementierung** Im nächsten Schritt wird die Struktur des erstellten (oder zuvor geladenen) Bindungsgraphen auf Zyklen und Vernetztheit analysiert. Dabei kommen die aus der Graphentheorie bekannten Tiefen- und Breitensuchealgorithmen zum Einsatz, siehe Abschnitt 9.1.3.

## 9. Experimente

Wir verwenden dabei die selben drei Testsysteme wie zuvor, da die Komplexität wieder von den Eigenschaften des Graphen abhängt. Wir erwarten lineare Komplexität, da die obigen Algorithmen keine Abhängigkeit von der Struktur des Graphen haben.

In Abbildung 9.6(a) sind die Zeiten für die Analyse des Bindungsgraphen in Abhängigkeit von der jeweiligen Anzahl der Kerne festgehalten. Die Steigungen  $\alpha$  erhalten wir wieder aus linearer Regression an (9.2.1).

Wir haben bei dem Fit der Geraden Datenpunkte bei kleiner Anzahl der Kerne aufgrund überwiegender Initialisierungseffekte schwächer gewichtet. Diese Effekte bedingen auch den anfänglichen Knick.

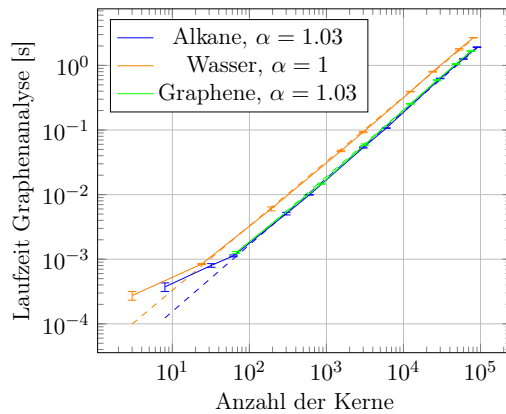
Ansonsten gibt es keinerlei wesentliche Unterschiede in den drei Geraden. Keine der Zahlen widerlegt die lineare Komplexität. Eine Abhängigkeit vom Vernetzungsgrad lässt sich anhand der Zahlen nicht belegen. Die Steigung ist  $\alpha \approx 1$ .

### 9.2.4. Zerlegung des Bindungsgraphen

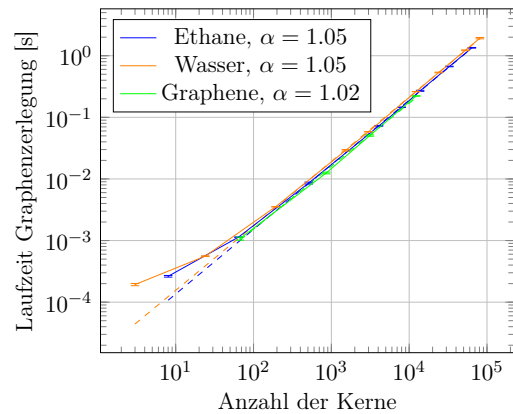
**Implementierung** Hier wird der Bindungsgraph in einzelne verbundene Teilgraphen zerlegt, welche zuvor bei der Analyse mittels der Tiefensuche erfasst wurden. Diese repräsentieren die einzelnen Moleküle des Systems, da wir in der gegenwärtigen Implementierung nur die kovalente Bindungsart als Bindung im Graphen erfassen.

Anstelle einer einzelnen Alkankette haben wir uns für ein mit Ethan gefülltes Gebiet entschieden, da bei einem einzelnen Molekül in diesem Schritt keine Arbeit zu erledigen wäre.

Die gemessenen Zeiten für die die Analyse und Zerlegung des Bindungsgraphen sind in Abbildung 9.6(b) abgebildet. Die Steigung  $\alpha$  erhalten wir wie zuvor bei der Analyse durch lineare Regression.



(a) Graphenanalyse



(b) Graphenzerlegung

**Abbildung 9.6.:** Laufzeiten für die Analyse des Bindungsgraphen und Zerlegung in verbundene Untergraphen für verschiedene molekulare Systeme.

Wir bemerken, dass die Steigungen allesamt leicht über  $\alpha = 1$  liegen, also dem linearen Fall. Dies liegt wohl aber an der Benutzung der sortierten Indexlisten. Da die Graphene nur aus einem Molekül bestehen, ist bei ihnen in diesem Schritt nichts zu tun.

### 9.2.5. Fragmentierung

**Implementierung** Das molekulare System liegt nach der Zerlegung in Form verbundener Teilgraphen vor. Nun wird jeder Teilgraph in alle möglichen verbundenen Untergraphen einer vorgegebenen maximalen Knotenzahl zerlegt. Der grundlegende Ansatz ist dabei die Betrachtung der Potenzmenge, aber mit eingeschränktem Horizont, siehe auch den Beweis zu Lemma 14.

Es gilt eine ganze Reihe verschiedener Fehlerquellen in der Implementierung zur überprüfen: Abhängigkeit von der Molekülzahl, nicht-lineare Abhängigkeit von der Maximalzahl an Atomen eines Moleküls, bestimmte Eigenschaften des Molekülgraphen.

Um diese Fehler auszuschließen, untersuchen wir die folgenden drei Bereiche: Robustheit, einzelnes Molekül und viele Moleküle.

#### Robustheit

Die folgende Untersuchung soll vor allem die Robustheit der Implementierung gegenüber verschiedensten Bindungsgraphen feststellen. Wir wollen also überprüfen, dass die Methode unabhängig von der Art des Molekülsystems funktioniert.

Dazu betrachten wir ein Testsystem verschiedener kleiner organischer Systeme. Diesen Testsatz werden wir besonders in Folgeabschnitten für die Analyse der Näherung der Gesamtenergie durch unsere Methode verwenden; eine Auflistung findet sich deswegen in Abschnitt 9.4.5 in Tabelle 9.3.<sup>4</sup>

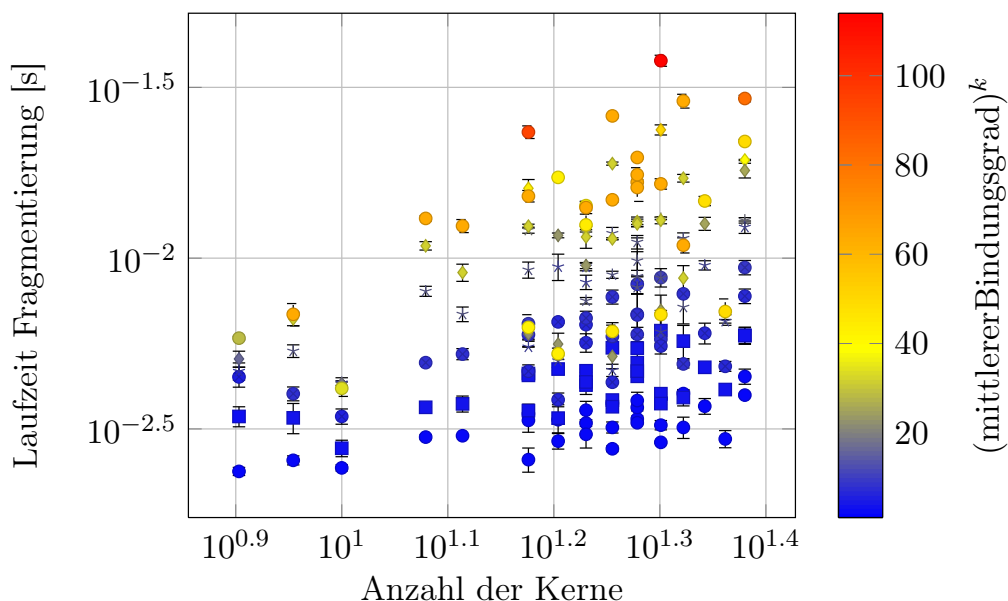
Da weiterhin der Bindungsgraph für diese Testsysteme bei der Auswahl relevant war und weniger ihre chemischen Eigenschaften, sind diese Moleküle dort in den Abbildungen 9.33 und 9.34 auch dargestellt.

In Abbildung 9.7 sind die Zeiten abgebildet. Aufgrund der kleinen Anzahlen von Atomen ist ein Fit an die Steigung pro Ordnung  $k$  nach (9.2.1) nicht sinnvoll. Wir haben die Datenpunkte mittels  $\hat{c}^k$  eingefärbt, wobei  $\hat{c}$  den mittleren Bindungsgrad pro Atom bezeichnet. Dieser gibt das allgemeine Verhalten sinnvoller wieder als die obere Schranke  $c$  aus Lemma 14.

Wir würden gerne erkennen, dass die Datenpunkte durch die Einfärbung in Gruppen gleicher Farbe und fester Steigung im Plot eingeteilt werden. Aber die untersuchten Systeme sind zu klein und zu unterschiedlich, als dass sich solch ein Verhalten ablesen ließe. Wir können nur schließen, dass Systeme mit größerem mittleren Bindungsgrad auch im Allgemeinen mehr Zeit benötigen.

<sup>4</sup>Im Folgenden benutzen wir in Abbildungen und Tabellen bei den Molekülnamen die englischen Bezeichnungen.

## 9. Experimente



**Abbildung 9.7.:** Laufzeiten für die Fragmentierung verschiedener kleiner organischer Moleküle für Ordnungen von  $1, \dots, 6$ . Dabei wurden die Datenpunkte nach  $\hat{c}^k$  mit der mittleren Bindungszahl  $\hat{c}$  pro Atom und der Bindungsordnung  $k$  eingefärbt.

### Einzelne Moleküle

Anschließend untersuchen wir Alkanketten und Graphene, also einzelne, möglicherweise große Moleküle mit 1- bzw. zwei-dimensionaler räumlicher Struktur.

In [Abbildung 9.8\(a\)](#) und [9.8\(b\)](#) sind die gemessenen Zeiten festgehalten. Wieder haben wir die Steigung  $\alpha$  durch lineare Regression an [\(9.2.1\)](#) erhalten.

Bei den Alkanen sind die Steigungen über die Ordnungen sehr gleichmäßig. Es ergibt sich nur ein minimaler Anstieg mit der Ordnung. Bei den verschiedenen großen Graphenschichten ergibt sich ebenso eine Steigung  $\alpha \approx 1$ . Ab  $k = 2$  liegt sie aufgrund der Benutzung von sortierten Listen etwas über 1, siehe unsere Bemerkungen eingangs [Abschnitt 9.2](#). Der höhere Bindungsgrad  $c$ , siehe [\(9.1.1\)](#), wirkt sich hier im Gegensatz zu den linearen Alkanen in [Abbildung 9.8\(a\)](#) sichtbar aus.

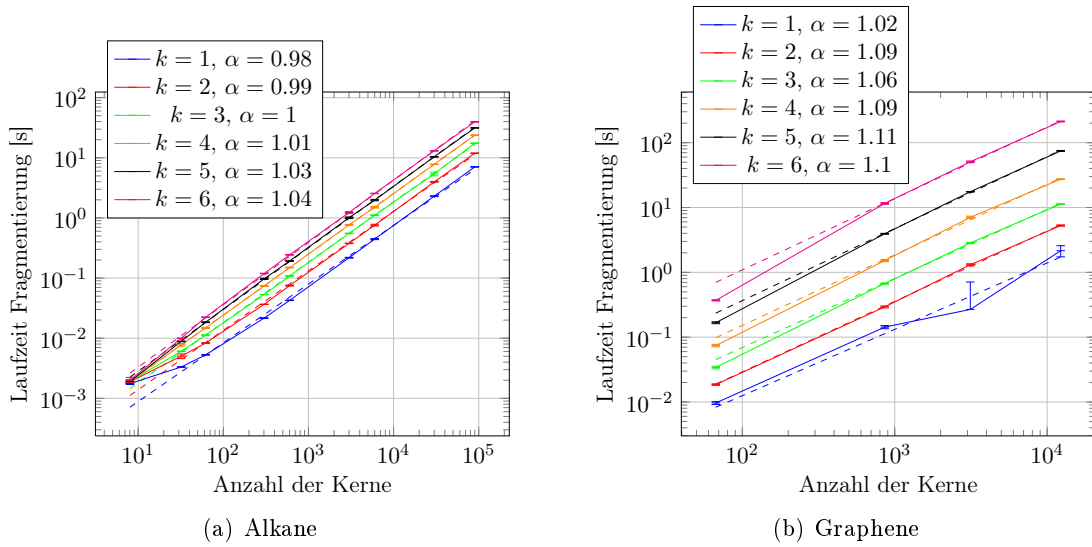
### Viele Moleküle

Schließlich werden wir zunächst Gebiete mit Wasser als ersten Test untersuchen, denn dabei ergeben sich nur Bindungsfragmente der Ordnung 1, und dann aber auch mit Alkanen verschiedener Längen und ebenso Schichten von Graphenen gefüllt beschäftigen. Es soll gezeigt werden, dass die Anzahl der Moleküle keinen Einfluss hat, sondern nur die Gesamtatomzahl eine Rolle spielt.

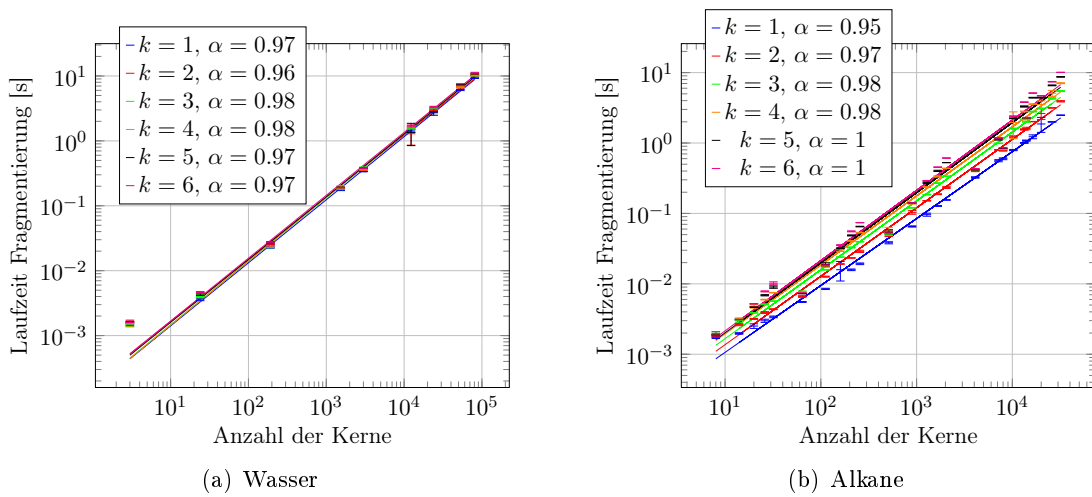
Die gemessenen Zeiten für Wasser sind in [Abbildung 9.9\(a\)](#) abgebildet. Wir sehen bestätigt, dass die Ordnung keine Rolle spielt, weil das Wasserstoffmolekül unter besonderer Behandlung der Wasserstoffe von der Bindungsordnung 1 ist. Es werden also immer



## 9.2. Skalierungsverhalten der Implementierung



**Abbildung 9.8.:** Laufzeiten für die Fragmentierung von Alkanketten unterschiedlicher Länge und verschieden großer Graphenschichten für Ordnungen von  $1, \dots, 6$ .



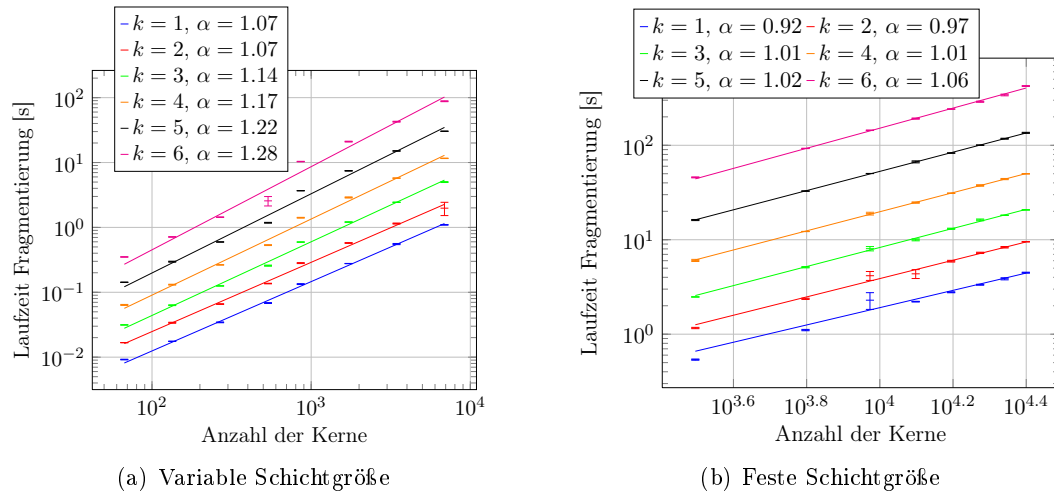
**Abbildung 9.9.:** Laufzeiten für die Fragmentierung eines mit Wasser und mit Alkanen fester Länge gefüllten Gebiets für Ordnungen von  $1, \dots, 6$ .

## 9. Experimente

gleich viele Fragmente erzeugt. Die Steigung liegt im Rahmen der Genauigkeit des Fits bei  $\alpha \approx 1$ . Das bedeutet, hier ergibt sich keine Abhängigkeit von der Molekülzahl.

Die Laufzeiten für Alkanketten finden sich in Abbildung 9.9(b). Abgesehen davon, dass die Ordnung hier nun eine Rolle spielt, da die Alkanmoleküle mehr Kerne als das Wassermolekül haben, zeigt sich ebenso eine Steigung von  $\alpha \approx 1$ .

Schließlich beschäftigen wir uns mit verschiedenen dicken Graphenschichten verschiedener Größe. Hier haben wir also einige wenige, aber sehr große Moleküle.



**Abbildung 9.10.:** Laufzeiten für die Fragmentierung verschieden großer ((a) bzw. mit fester Größe (b)) und verschieden häufig gestapelter Graphenschichten für Ordnungen von  $1, \dots, 6$ .

In Abbildung 9.10 sind diese gemessenen Zeiten des geschichteten Graphensystems dargestellt. Wieder wurde die Steigung  $\alpha$  durch lineare Regression an (9.2.1) bestimmt. Einmal wurden in (a) sowohl die Größe des Graphens als auch die Anzahl der Schichten variiert, daneben wurden in (b) nur die Anzahl der Schichten verändert.

Wir bemerken, dass sich mit wechselnder Schichtengröße des Graphens ein „Sprung“ im ersten Plot 9.10(a) ergibt. Es wurden jeweils Schichtzahlen von  $1, \dots, 4$  gewählt und zwei verschiedene Graphengrößen. Das bedeutet, dass die Steigung im ersten Plot nur scheinbar über  $\alpha = 1$  liegt, denn es muss der Rand des Graphens berücksichtigt werden. Diese spielt mit saturierten Wasserstoff vor allem beim Wechsel von dem kleineren Graphen zum größeren eine wesentliche Rolle. Denn der Rand nimmt Einfluss auf die mittlere Bindungszahl  $C$  und damit auf die Konstante. Der Sprung entspricht also einem Wechsel in dieser Konstante.

Beim zweiten Plot 9.10(b) haben wir ein noch größeres Graphenstück fest gewählt und Schichtzahlen von  $1, \dots, 8$  untersucht. Wir sehen, dass sich hier der erwartete lineare Zusammenhang sehr gut ergibt, gleichwohl wir auch einen kleinen, aber unwesentlichen Anstieg mit der Ordnung bemerken, analog zu Abbildung 9.8(b).

### 9.2.6. Fragmentberechnung

**Implementierung** Schließlich müssen die HF-Teilprobleme gelöst werden. Da sich unsere Methode besonders einfach aufgrund der voneinander unabhängigen Teilprobleme parallelisieren lässt, haben wir dies für diesen Programmteil realisiert. Da besonders sogenannte Commodity Cluster, bei denen Rechenknoten in einem von außen nicht erreichbaren internen Netzwerk liegen, heutzutage sehr beliebt sind, haben wir einen Client/Server-Ansatz gewählt. Die Teilprobleme werden dabei von einem Controller an den Server geschickt. Der Controller fragt in regelmäßigen Abständen nach fertigen Ergebnissen, er hat dabei keinen direkten Kontakt zu den Clients. Dieser Client/Server-Ansatz wurde im gesonderten Paket „JobMarket“ in C++ mit Hilfe der ASIO entwickelt. Dabei verbinden sich viele Clients an einen Server, der Arbeitspakete mit einem abstrakten Interface vorrätig hält und darüber wacht, dass die einzelnen Pakete von den Clients korrekt abgearbeitet wurden. Ein bekanntes Framework in weitaus mächtigerer Form ist BOINC [7].

Dabei interessieren wir uns für zwei Messwerte: Einerseits die Rechenzeit der Clients und zweitens die Zeit, welche für das Verschicken der Fragmente und das Erhalten und Summieren der Resultate benötigt wird. Letzteres stellt den Overhead dar, welcher nicht den Hauptteil der Rechenzeit ausmachen sollte.

Für die Rechnungen der Teilprobleme benutzen wir MPQC [108] in leicht modifizierter Form, damit das Programm als zum Server passender Client fungieren kann. Wir verwenden dabei stets *CLHF* als Theorie mit Berechnung des Gradienten und *6-311G* als Basis. In diesem Versuch interessiert uns aber die Genauigkeit der Ergebnisse noch nicht.

#### Skalierbarkeit des Client/Server-Ansatzes

Als erstes werden wir die Tauglichkeit dieses Client/Server-Ansatzes testen, indem wir als Arbeit „NoOp“ für den Client auswählen: Er tut nichts und meldet Erfolg zurück.

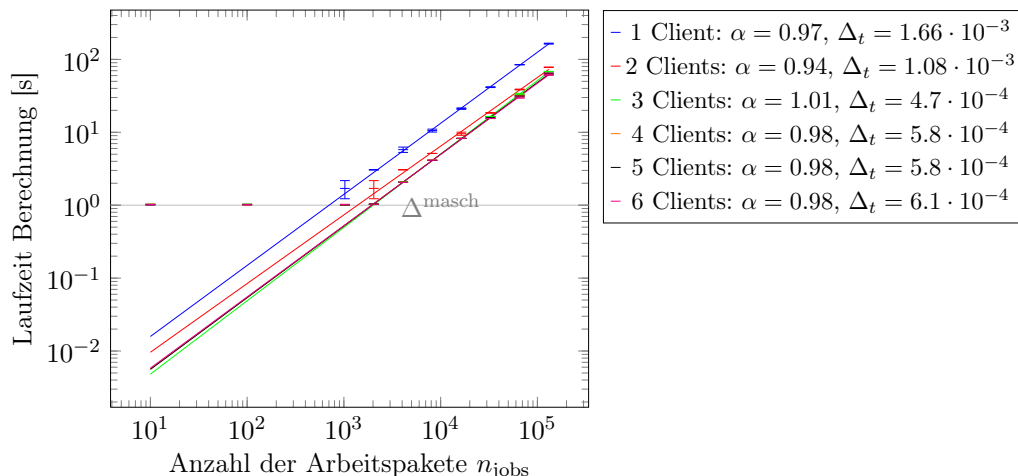
Wir erfassen damit in guter Näherung, wie schnell das Verteilen bzw. das Senden/Empfangen der Arbeitspakete von statten geht und damit wie groß der Overhead ist.

Da die Gesamtlaufzeit durch Polling seitens eines Controllers ermittelt werden muss, der nur mit einer Genauigkeit von  $\Delta^{\text{masch}} = 1\text{ s}$  nachfragt und mit höherer Frequenz die Messung signifikant beeinflussen würde, ist unsere Messgenauigkeit auf diesen Wert beschränkt. Dies kann aber durch eine entsprechende Gesamtlaufzeit kompensiert werden, weswegen wir eine Anzahl  $n = 2^p$  von Arbeitspaketen von  $p = 2; 3; 4; 6; 10; 11; \dots; 17$  betrachten. In Abbildung 9.11 sind die ermittelten Laufzeiten und die Grenze der Messgenauigkeit als horizontale Linie eingetragen.

Wir haben im Fit Datenpunkte mit  $T < 10\text{ s}$  aufgrund der endlichen Messgenauigkeit schwächer gewertet. Wir bemerken, dass wir bei mehr als 32.768 Arbeitspaketen auch an technische Grenzen stoßen<sup>5</sup>.

<sup>5</sup>Hierbei gehen insbesondere durch die häufigen Verbindungen die lokalen TCP-Ports des Systems aus. Man muss den Kernel anweisen, Ports direkt wiederzuverwenden.

## 9. Experimente



**Abbildung 9.11.:** Zeiten für das reine Senden/Empfangen von Arbeiten. Aus der Steigung ergibt sich eine gute Näherung an die Maximalzahl der Arbeitspakete, die ein einzelner Server pro Sekunden verschicken kann.

Die Steigung gibt klare Hinweise auf einen linearen Zusammenhang. Auch ergibt sich aus der doppelt logarithmischen Darstellung die Konstante, welche die Anzahl an Arbeitspaketen beschreibt, welche pro Sekunde lokal verteilt werden können,

$$\ln(\Delta_t \cdot n_{\text{jobs}}^\alpha) = \ln(\Delta_t) + \alpha \cdot \ln(n_{\text{jobs}}),$$

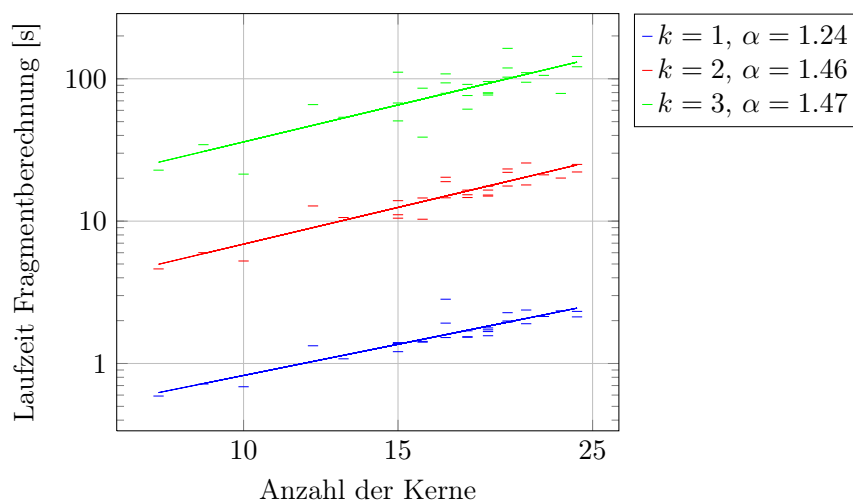
mit der Anzahl an Arbeitspaketen  $n_{\text{jobs}}$ . Wir ermitteln nebst  $\alpha$  auch  $\ln(\Delta_t)$  aus dem Fit, welcher in der Abbildung als Rate  $\Delta_t$  angegeben ist. Dabei kann  $\Delta_t \approx 6 \cdot 10^{-4}$  auch direkt als Overhead pro Paket interpretiert werden. Wir bemerken, dass wir ab 3 Clients in den Sättigungsbereich der Servers kommen. Er kann die Arbeitspakete nicht schneller verteilen, so dass die Gesamtzeit durch Verwendung weiterer Clients nicht weiter abfallen kann. Damit ergibt sich eine Leistung von ca. 1600 Paketen pro Sekunde auf dem Testsystem. Dabei läuft der Server mit nur einem Thread.

Wir bemerken, dass dieser Wert für unsere Zwecke völlig ausreichend ist, da die Bearbeitungszeit für das Arbeitspaket eines echten Fragmentproblems im Sekundenbereich liegt. Wir können also mit einem Server bis zu 1600 Clients mit Arbeitspaketen versorgen, wobei wir ein entsprechend schnelles Netzwerk voraussetzen, so dass die Latenz bei echten Netzwerkverbindungen keinen Einfluss hat.

### Robustheit

Nun analysieren wir die Laufzeiten für den Testsatz organischer Moleküle. Wir verweisen noch einmal auf Tabelle 9.3 für eine Auflistung des Testsatzes.

Wir interessieren uns hierbei weniger für das Skalierungsverhalten, da die Moleküle zu unterschiedliche Bindungsgrade haben, sondern vor allem für die allgemeine Robustheit des Verfahrens. Das bedeutet, dass alle Teilprobleme, die erzeugt werden, führen bei der



**Abbildung 9.12.:** Zeiten eines Clients für die Berechnung der Fragmentprobleme verschiedener einzelner Testmoleküle bei verschiedenen Ordnungen  $k$ .

Berechnung mit MPQC [108] zu im Rahmen der Abbruchbedingung auskonvergierten Lösungen. Die Zeiten sind in Abbildung 9.12 festgehalten.

Wir merken an, dass die Berechnung robust funktioniert hat. Es gibt keine Fehler oder nicht-konvergierte Fragmente über den Testsatz. Die Steigung  $\alpha$  wächst deutlich an und liegt über 1.

Im Allgemeinen wächst die benötigte Laufzeit für ein Fragment mit der Bindungsordnung  $k$  deutlich an. Entscheidend für die Gesamtlaufzeit ist also besonders die Anzahl der Fragmente zur höchsten Bindungsordnung. Gerade deren Zahl schwankt aber deutlich über den Testsatz trotz ansteigender Anzahl der Kerne. Dies liegt an der sehr unterschiedlichen Bindungsstruktur mit unterschiedlichem mittlerem Bindungsgrad, vgl. Abbildung 9.7. Deswegen lässt sich aus der bestimmten Steigung nicht ableiten, dass dieser Programmteil keine lineare Komplexität hat.

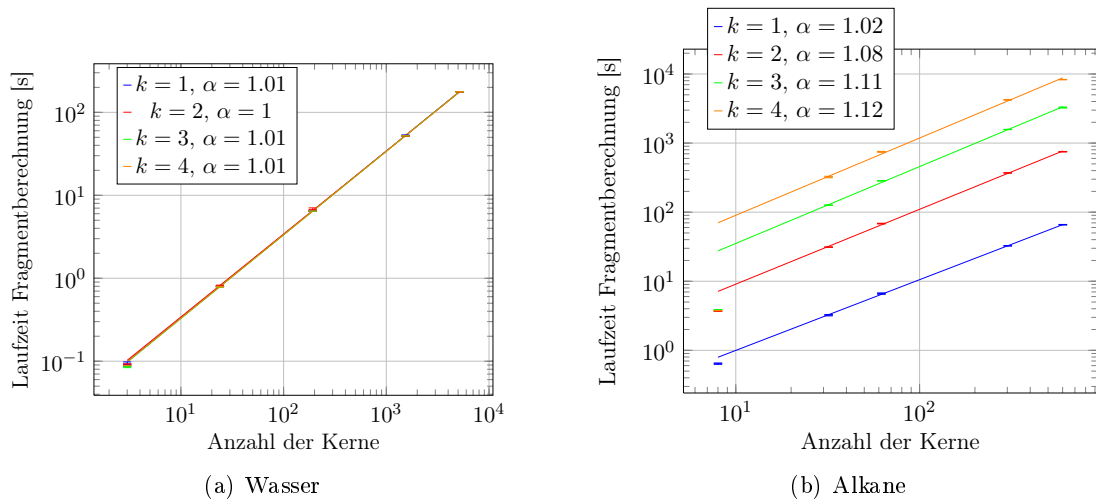
### Einzelne und viele Moleküle

Als nächstes betrachten wir das Laufzeitverhalten für das Wassergefüllte Gebiet und einzelne Alkanmoleküle bei verschiedenen Ordnungen.

Die ermittelten Laufzeiten für Wasser finden sich in Abbildung 9.13(a) für die Ordnungen 1, 2 und 3. Da die Ordnungen über 1 aufgrund der besonderen Behandlung von Wasserstoff nicht mehr Fragmente erzeugen können, erwarten wir die gleichen Zeiten. Wir bemerken, dass die Steigung perfekt linear ist.

Die Laufzeiten für Alkane sind in Abbildung 9.13(b) für Ordnungen 1 bis 4 und über verschieden lange Alkane aufgetragen. Auch hier ergibt sich eine hinreichend lineare Steigung und damit der Beleg des gewünschten Skalierungsverhaltens. Wir haben die ersten beiden Datenpunkte dabei für höhere Ordnungen schwächer gewichtet, da die Anzahl der Fragmente höchster Ordnung, welche die Hauptrechenlast beanspruchen, über die

## 9. Experimente



**Abbildung 9.13.:** Zeiten eines Clients für die Berechnung der Fragmentprobleme eines Wassergefüllten Gebiets und verschiedener Alkane bei verschiedenen Ordnungen.

ansteigende Anzahl der Kerne nicht in gleichem Maße zunimmt. Der Anstieg der Skalierungskonstante mit der Bindungsordnung  $k$  wird dabei in [Abbildung 9.13\(b\)](#) deutlich sichtbar, siehe auch [Lemma 14](#).

### Parallele Effizienz

Schließlich wollen wir anhand der Beispiele auch noch die parallele Leistungsfähigkeit der Implementierung messen. Wir interessieren uns also nun für parallele Effizienz und Scaleup und fassen dazu die Definitionen zunächst knapp aus [Griebel et al. \[77, Abschn. 4.4\]](#) zusammen.

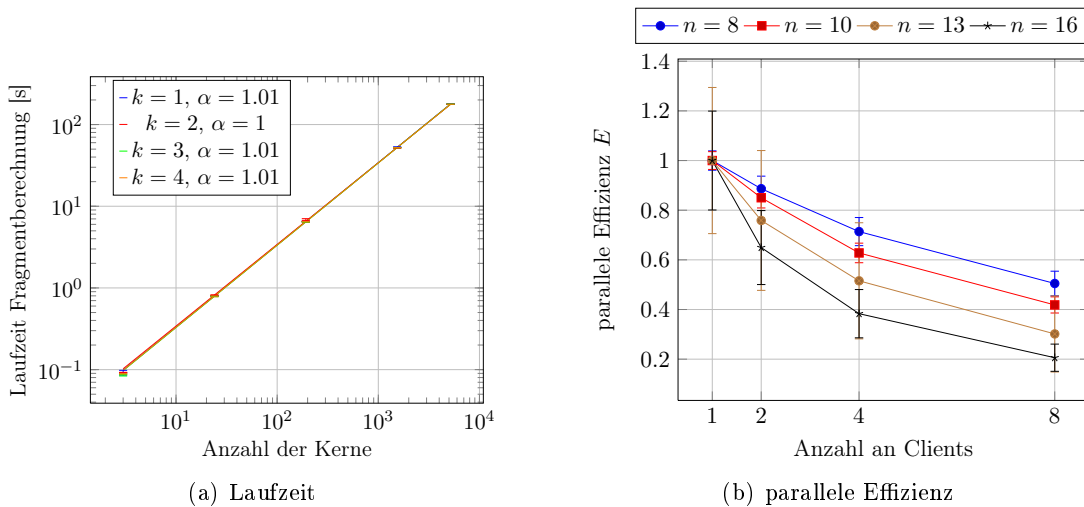
Für die nebenläufige Ausführung mit einer Prozesszahl  $P$  wird die Laufzeit mit  $T(P)$  bezeichnet. Als Systemunabhängige Größe dient die parallele Effizienz. Sie ist definiert als

$$E(P) := \frac{T}{PT(P)}. \quad (9.2.2)$$

Im optimalen Fall ergibt sich eine Effizienz von  $E(P) = 1$ . Sind nicht alle Teile eines Programms parallelisiert, so fällt diese Effizienz ab einer bestimmten Prozessorzahl und Problemgröße ab, weil der verbleibende serielle Anteil dominiert. Dies ist das Amdahlsche Gesetz. Deswegen misst man auch den Scaleup, wo Problemgröße und Prozessorzahl mit einem festen Faktor  $\kappa$  gemeinsam anwachsen,

$$S_c(P, N) = \frac{T(P, N)}{T(\kappa P, \kappa N)}. \quad (9.2.3)$$

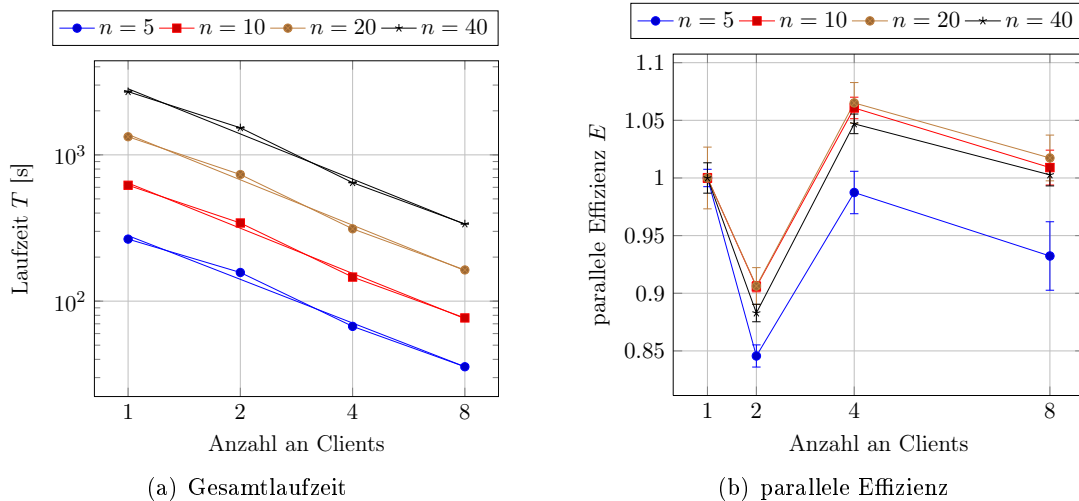
**Wassermoleküle** Wir messen zunächst für unterschiedlich große Wasser-gefüllte Gebiete mit  $n = 8, 10, 13, 16$  Molekülen pro Raumachse und verschiedene Anzahl an Clients  $w = 1, 2, 4, 8$  die Gesamtlaufzeit des Verfahrens. Wassermoleküle sind dabei aber eines der denkbar ungünstigsten Beispiele, da die Lösungen hier in  $T \approx 0.1$  s pro Fragment bestimmt werden. Das bedeutet, dass der parallele Programmteil im Vergleich zum seriellen nicht wesentlich mehr Zeit benötigt. Die Laufzeiten sind in Abbildung 9.14(a) festgehalten. Wir sehen, dass wir keine optimale parallele Effizienz erreichen, siehe Abbildung 9.14(b). Sie fällt von kleinen Systemen immer weiter ab. Bei größeren Systemen brauchen die nicht-parallelisierten Teile bei größer werdender Anzahl an Wassermolekülen einen immer größeren Teil der Gesamtlaufzeit, so dass eine noch größere Anzahl an Clients schließlich kaum noch einen weiteren Nutzen bringt. Dies ist das schon erwähnte Amdahlsche Gesetz.



**Abbildung 9.14.:** Laufzeit und parallele Effizienz des Verfahrens für die Berechnung eines Wasser-gefüllten Gebiets mit  $n = 8, 10, 13, 16$  Molekülen pro Raumachse über verschiedene Anzahl an Clients.

**Alkane** Weiterhin messen wir auch Alkanketten der Länge  $n = 5, 10, 20, 40$  mit gleicher Anzahl an Clients wie zuvor. Diese Zeiten finden sich in Abbildung 9.15. Wir sehen, dass wir diesmal im Rahmen der Messgenauigkeit optimale parallele Effizienz von ca. 1 erhalten, siehe Abbildung 9.15(b). Zusätzlich sind in Abbildung 9.16 der Scaleup und die Laufzeit des parallelen Anteils allein aufgetragen. Wir haben dabei die Tupel  $(n, w) = (5, 1), (10, 2), (20, 4), (40, 8)$  gewählt. Bei den Alkanketten macht ihre endliche Länge jedoch die Interpretation von Abbildung 9.16(b) schwieriger: Für kürzere Ketten entstehen nicht genau proportional weniger Fragmente höherer Ordnung als für längere Ketten, sondern nur  $2n - (k - 1)$ . Für Ordnung  $k = 4$  erhalten wir für die gewählten  $n$ : 7, 17, 37, 77. Es ist aber besonders die Berechnung dieser Fragmente höherer Ordnung, welche den Großteil der Laufzeit ausmachen und wie gesehen steigt deren Zahl nicht

## 9. Experimente



**Abbildung 9.15.:** Laufzeit und parallele Effizienz des Verfahrens für die Berechnung eines Alkans der Längen  $n = 5, 10, 20, 40$  über verschiedene Anzahl an Clients.

genau mit dem Faktor 2. Ein weiterer Effekt ist, ob die Anzahl der Clients zu der Anzahl Fragmente höchster Ordnung im Sinne eines Teilers passt. Dadurch kann die Rechenlast in den Einheiten der Arbeitspakete nicht perfekt gleichmäßig auf die Clients verteilt werden, woraus eine längere Laufzeit resultiert, als zu erwarten wäre. Dies ist in den Knicken der eigentlich flachen Kurven in Abbildung 9.15(a) zu sehen<sup>6</sup>, wo nur die Laufzeit der parallel laufenden Löser aufgetragen ist. Der gleiche Verlauf über die Anzahl der Clients zeigt sich auch im Scaleup. Unter Berücksichtigung dieser Einschränkungen verhält sich dieser also auch sehr gut.

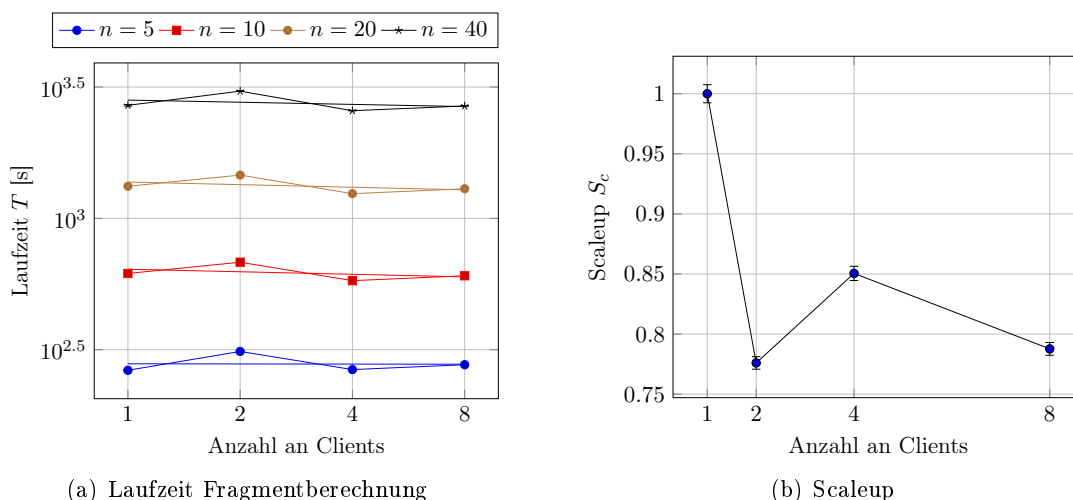
Abschließend fassen wir zusammen, dass die Parallelisierung bis zu kleiner Prozesszahl im zweistelligen Bereich ordentlich ist. Für die längste Alkankette beträgt die Zeit des seriellen Anteils ca. 3 s, während die kumulierte Löserzeit 2700 s ist. Bis also  $w = 100$  Clients ergibt sich noch ein sinnvoller Speedup. Zur effizienten Behandlung großer Systeme mit mehr Clients sollten auch die restlichen Teile der Implementierung parallelisiert werden. Die entsprechenden Strategien sind für alle Teile und Algorithmen in der Literatur bekannt, siehe [77, 73].

### 9.2.7. Aufsummierung der Lösungen der Teilprobleme

**Implementierung** Mit den gespeicherten Ergebnissen aus dem letzten Abschnitt werden nun die Teillösungen aufsummiert. Diesen Schritt haben wir mit Hilfe der Boost Fusion-Bibliothek unabhängig von einzelnen Datentyp implementiert. Dabei werden alle Elemente einer Datenstruktur einzeln nacheinander bearbeitet. Jedem Datum zu jeder Instanz, welche also bspw. die Energie eines Fragment repräsentiert, wird eine Indexmen-

<sup>6</sup>Dabei priorisiert die Implementierung die Arbeitspakete nach ihrer summierten Kernzahl, womit Fragmente höherer Ordnung zuerst berechnet werden und dieser Effekt schon gemäßig wird.





**Abbildung 9.16.:** Kumulative Laufzeit des rein parallelen Programmteils und Scaleup des Verfahrens für die Berechnung eines Alkans der Längen  $n = 5, 10, 20, 40$  über verschiedene Anzahl an Clients.

ge zugeordnet. Die Indexmengen und damit auch die Daten bilden anschließend einen Hierarchiebaum. Dazu muss zu jeder Indexmenge eine Anzahl von Referenzen zu Untermengen gespeichert werden. Diese erhält man aus der Potenzmenge der Indexmenge als mögliche Kandidaten und einer Suche in einer STL-Map in linearer Zeit, da die Mächtigkeit der Mengen durch die Bindungsordnung beschränkt ist.

Wir erwarten daher für diesen Programmteil erneut lineare Komplexität und keine Abhängigkeit von der genauen Gestalt des Bindungsgraphen. Für sehr große Indexmengen könnte die Komplexität von  $\mathcal{O}(M \log M)$  der sortierten Listen hervortreten. Die Skalierungskonstante hingegen sollte wieder deutlich vom der Bindungsordnung  $k$  abhängen, da sie die Größe der zu überprüfenden Potenzmenge bestimmt.

Wir untersuchen deswegen die zwei folgenden Systeme: Alkanketten der Länge  $n = 1, 5, 10, 50, 100$  und ein Wassergefülltes Gebiet mit  $n = 1, 2, 4, 8, 12$ . Das bedeutet, einmal mit dem Fokus auf einzelne große Moleküle und das zweite Mal mit dem Fokus auf viele verschiedene Moleküle, woraus jeweils große Indexmengen mit unterschiedlicher Hierarchie resultieren.

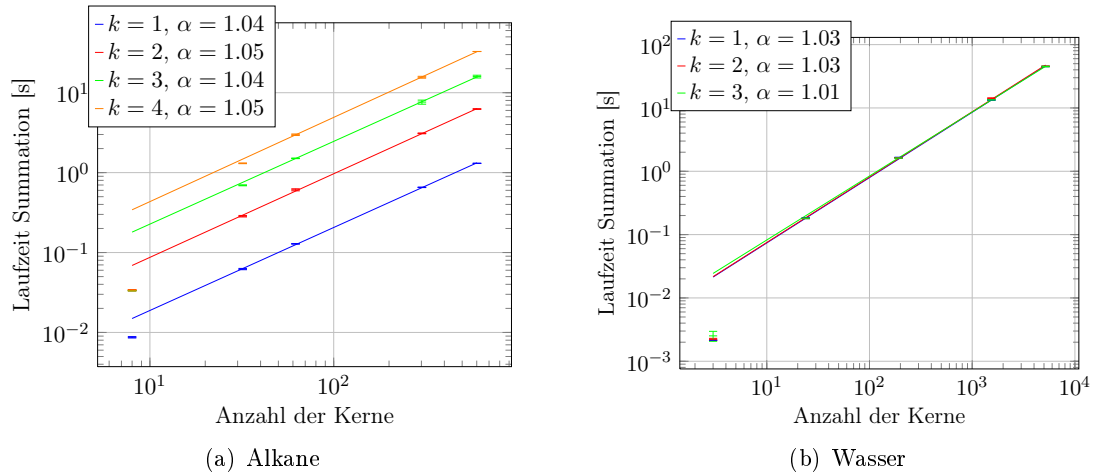
Die Zeiten für Alkane und Wasser sind in den Abbildungen 9.17(a) und 9.17(b) dargestellt. Auch hier haben wir die Steigung  $\alpha$  mittels linearer Regression an (9.2.1) bestimmt.

Wir sehen, dass sich die Laufzeiten beide linear mit der Anzahl der Kerne verhalten. Auch hier haben wir wieder die ersten Datenpunkte für höhere Ordnungen schwächer gewichtet.

### 9.2.8. Zusammenfassung

Wir haben jeden einzelnen Aspekt der Implementierung der BOSSANOVA-Methode in Kürze beschrieben und untersucht. Dabei haben wir festgestellt, dass wir das von den

## 9. Experimente



**Abbildung 9.17.:** Laufzeit für die Summation der Fragmentresultate für verschiedene Alkane und für Wassergefüllte Gebiet verschiedener Größe.

gewählten Datenstrukturen und Algorithmen zu erwartende Verhalten von  $\mathcal{O}(M \log(M))$  klar nachvollziehen konnten.

Es ergeben sich Einschränkungen in diesem Verhalten, wenn Atome sehr dicht beieinander liegen oder sehr stark vernetzt sind. Da wir aber von einer beschränkten Dichte und einem beschränkten maximalen Bindungsgrad ausgehen, ist auch die getestete Implementierung für die untersuchten Anzahl der Kerne insgesamt linear skalierend ist.

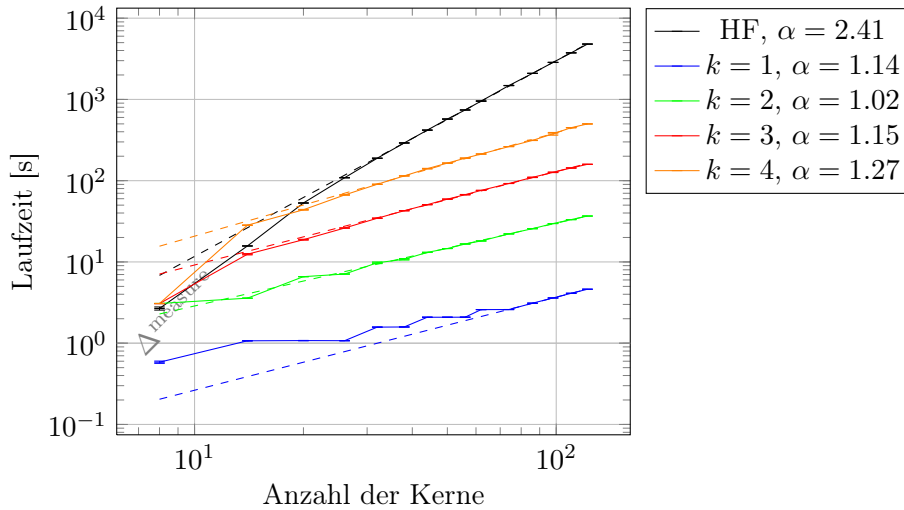
Abschließend wollen wir noch die Laufzeit unserer Methode gegen die Standard-HF-Methode vergleichen, wie sie in MPQC [108] implementiert ist. Damit ermitteln wir die typische Systemgröße, ab dem BOSSANOVA schneller ist, den sogenannten Crossover-Punkt.

Wieder untersuchen wir verschieden lange Alkanketten der Summenformel  $C_{2n}H_{4n+2}$  und berechnen deren Grundzustandsenergie. Dabei benutzen wir 4 Prozesse, also 4 Clients<sup>7</sup> für die parallele Berechnung der Teilprobleme mit BOSSANOVA und 4 Threads für MPQC bei der Berechnung des vollen Problems. Dabei wird die benötigte Gesamtlaufzeit gemessen und verglichen.

In Abbildung 9.18 sind die gemessenen Zeiten doppellogarithmisch dargestellt. Für jede Ordnung ergibt sich dabei ein eigener Kreuzungspunkt mit der Kurve für die Berechnung des vollen Problems. Für die Ordnungen  $k = 1, 2, 3, 4$  benötigen wir ab  $n = 1, 1, 2, 3$  jeweils effektiv weniger Zeit. Dies entspricht einer Anzahl der Kerne von ca. 20 für  $k = 4$ . Das bedeutet, dass ab einer Problemgröße von mehr als 20 Kernen die Berechnung der Grundzustandsenergie mit BOSSANOVA schneller ist als die Standard-HF-Rechnung bei ähnlicher Genauigkeit, wenn dabei die Endlichkeit der Basis berücksichtigt wird, siehe dazu den noch folgenden Abschnitt 9.4.

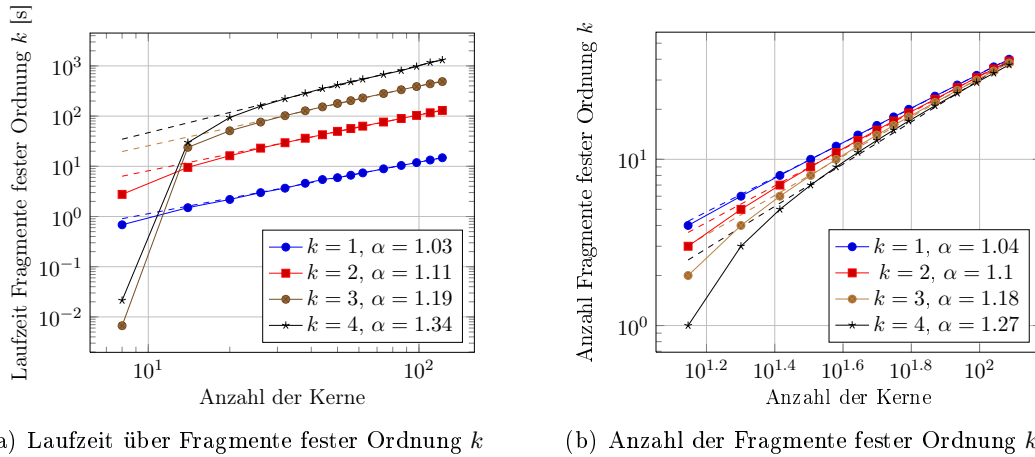
<sup>7</sup>Server- und Controller-Prozess sind völlig unterbeschäftigt, siehe Abschnitt 9.2.6, und werden deshalb nicht gezählt.

## 9.2. Skalierungsverhalten der Implementierung



**Abbildung 9.18.:** Gemessene Laufzeiten der Berechnung verschieden langer Alkane mittels Standard-HF und BOSSANOVA für  $k = 1, \dots, 4$ .

Dies ist um mehr als eine Größenordnung besser als bei anderen linear skalierenden Ansätzen: Goedecker [67] gibt in seiner Übersicht über linear skalierende Methoden hundert Atome als Crossover-Punkt an. Skylaris et al. [157] nennt einige hundert Atome. Weiterhin finden Fattbert und Gygi [53] den Punkt bei 500-600 Kernen, Horsfield et al. [99] nennen 200 Kerne für ihre Methode. Dixon und Merz Jr. [46] geben 173-193 Kerne, ein ebenfalls auf dem Divide&Conquer-Ansatz basierende Methode von Yang [186] ist ab 263 Kernen schneller als die Standard-HF-Methode.



**Abbildung 9.19.:** Vergleich des Verlaufs der kumulativen Laufzeiten der Löser für Fragmente fester Ordnung  $k$  und der Anzahl dieser Fragmente fester Ordnung  $k$  pro Ordnung.

## 9. Experimente

Abschließend wollen wir noch auf die scheinbar anwachsenden Steigungen  $\alpha$  mit der Bindungsordnung  $k$  eingehen. In Abbildung 9.19 vergleichen wir die Laufzeiten für Fragmente fester Ordnung mit der Anzahl der dabei berechneten Fragmente über den Längenfaktor der Alkane  $n$ . Dazu haben wir in Abbildung 9.19(a) die kumulierte Zeit aufgetragen, die alle vier Löser für die Berechnung bspw. für  $k = 3$  aller Molekülfragmente mit genau drei Nicht-Wasserstoff-Kernen benötigt haben. In Abbildung 9.19(b) hingegen ist die Anzahl dieser Fragmente festgehalten. Wir haben dabei jeweils die ersten drei Datenpunkte im Fit schwächer gewichtet. Wir sehen, dass die Steigung  $\alpha$  scheinbar mit der Ordnung  $k$  wächst und dies mit der Anzahl der Fragmente zu tun hat.

Die Formel  $2 \cdot n - (k - 1)$ , in Abbildung 9.19(b) dargestellt, gibt die Anzahl der Fragmente pro Ordnung  $k$  für die Alkane an. Sie wächst gerade für kleines  $n$  zunächst schneller als  $n$  an und geht erst für großes  $n$  in ein lineares Wachstum über. Da wir nur sehr kleine Systeme für die Bestimmung des Crossover-Punktes berechnen mussten, erfassen wir also eine scheinbar größere Steigung  $\alpha$ . Für große  $n$  würde die Steigung  $\alpha$  wieder gegen 1 gehen.

### 9.3. Bindungsgraphbestimmung

In diesem Abschnitt überprüfen wir anhand des bekannten Testsatzes von Molekülen, siehe Tabelle 9.3, wie sich die implementierten Methoden zur Bestimmung des Bindungsgraphen verhalten. Der Bindungsgraph hat dabei einen zentralen Aspekt im Rahmen der vorgeschlagenen BOSSANOVA-Methode, da er die Null/Nicht-Null-Struktur der zu erwartenden Eigenvektoren, also der Darstellung der lokalisierten Einteilchen-Wellenfunktion vorgibt. Dabei darf der Bindungsgraph mehr Kanten als nötig erhalten, aber die Ladungsneutralität muss für jedes Fragment gelten.

Es geht bei der Bestimmung des Graphen also um zwei Aspekte: Zunächst muss die Vernetztheit des Bindungsgraphen für die vorgegebene Knotenmenge bestimmt werden. Anschließend muss in einem zweiten Schritt pro Kante das richtige, ganzzahlige Gewicht, welches den Bindungsgrad repräsentiert, gefunden werden.

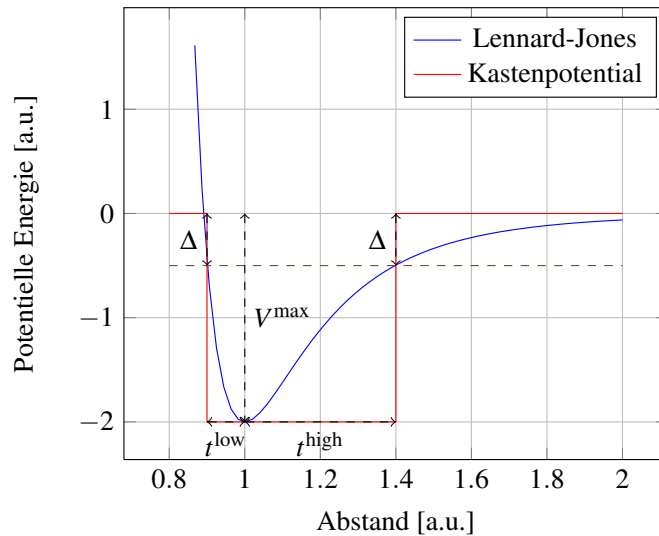
#### 9.3.1. Abstandskriterium

Allgemein fußen einfache, heuristische Methoden zur Bestimmung eines Bindungsgraphen auf einem Abstandskriterium, welches zwischen zwei Kernen  $a$  und  $b$  erfüllt sein muss, damit diese als gebunden gelten. Solch ein Kriterium erhält man bspw. mittels eines tabellarisierten Gleichgewichtsabstands  $R^{\text{eq}}(a, b)$  für zwei Elemente  $a$  und  $b$  und mittels zweier Schwellwerte  $t^{\text{low}}$  und  $t^{\text{high}}$ . Dadurch wird ein Intervall aufgespannt, in dem sich der Abstand zwischen zwei Kernen befinden muss,

$$R^{\text{eq}}(a, b) - t^{\text{low}} < \|\mathbf{R}_{ab}\| < R^{\text{eq}}(a, b) + t^{\text{high}}, \quad (9.3.1)$$

wobei  $\mathbf{R}_{ab} = \mathbf{R}_a - \mathbf{R}_b$  der Abstandsvektor den Kernorte ist.

Stellt man sich einen Potentialtopf zwischen Kernen  $a$  und  $b$  vor, in dem das Elektronenpaar innerhalb eines Bindungsorbitals lokalisiert ist, dann entspricht dies der Nähe-



**Abbildung 9.20.:** Der echte Potentialtopf – hier vereinfacht durch das Lennard-Jones-Modell – aufgrund der Coulomb-Wechselwirkung zwischen zwei Kernen  $a$  und  $b$  wird durch ein Rechteckpotential gleicher Tiefe ersetzt.

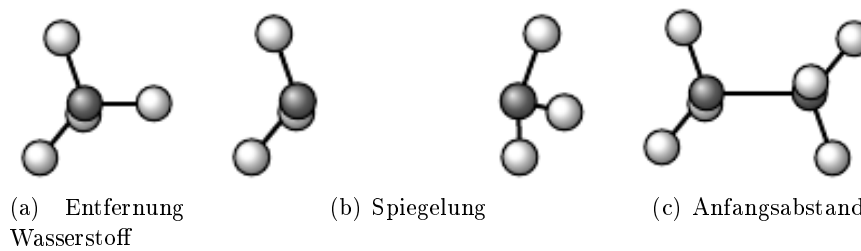
zung des Topfs durch ein Rechteckpotential, bei dem einzig dessen Breite für die Bindung Relevanz hat, siehe Abbildung 9.20.

Der tabellarisierte Abstand  $R^{\text{eq}}(a, b)$  soll dabei das Gleichgewicht bzw. das Potentialminimum beschreiben. Die Verwendung von zwei Schwellwerten  $t^{\text{low}}$  und  $t^{\text{high}}$  anstelle eines einzelnen dient der Berücksichtigung von Anharmonizitäten des Potentialtopfes. Sie sind dabei in Relation zur kinetischen Energie  $\Delta$  der Elektronen und der Tiefe  $V^{\text{max}}$  des Potentialtopfes zu setzen, siehe Abbildung 9.20. Üblicherweise werden aber beide Schwellwerte auch einfach auf den gleichen Wert gesetzt, z. B. von der Größe  $0.4 \text{ \AA}$ , siehe [1].

Wir halten noch fest, dass die Näherung über ein Rechteckpotential nur für ein molekulares System in der Nähe seiner *Gleichgewichtsgeometrie* sinnvolle Ergebnisse liefert, also wenn sich die jeweiligen Abstände gebundener Kerne deutlich innerhalb der Intervalle (9.3.1) befinden. Der Bruch einer Bindung kann darüber nicht sinnvoll erfasst werden und würde im Rahmen der Methode auch zu einer Unstetigkeit in der Näherung der Energie führen. Dieser *Bindungsbruch* könnte aber ohnehin im Allgemeinen nicht mittels der CLHF-Methode bestimmt werden, worauf wir uns aber in dieser Arbeit beschränkt haben.

Der Unterschied von dem von uns vorgestellten Verfahren zu von [1] besteht anschließend darin, wie der tabellarisierte Abstand  $R^{\text{eq}}(a, b)$  ermittelt wird, worauf wir nun eingehen.

## 9. Experimente



**Abbildung 9.21.:** Illustration zur Veranschaulichung des Algorithmus, um bivalente, gesättigte Systeme zu erzeugen. (a) Bei den gesättigten Atomen – hier nur das Element Kohlenstoff (grau) in der Ausgangsgeometrie – wird eine dem Bindungsgrad entsprechende Anzahl von Wasserstoffatomen (weiß) entfernt. (b) Eine der Geometrien wird gespiegelt. (c) Beide Geometrien werden schließlich in einem vorgegebenen Anfangsabstand zueinander zusammengeführt.

### Der kovalente Radius-Ansatz

Ein einfacher Ansatz für diesen Gleichgewichtsabstand ist nach [1] die Summe der kovalenten Radien der beiden beteiligten Elemente  $a$  und  $b$  zu benutzen,

$$R^{\text{eq}}(a, b) \approx R^{\text{cov}}(a) + R^{\text{cov}}(b). \quad (9.3.2)$$

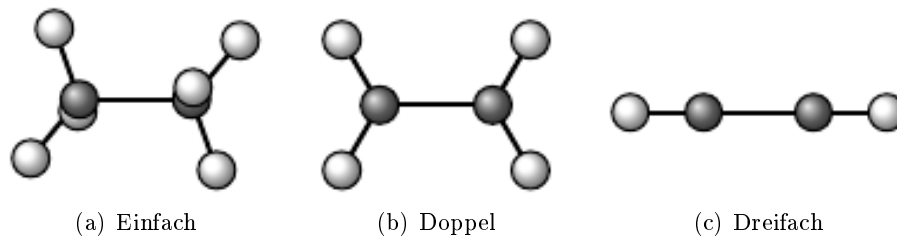
### Tabellarisierte Bindungsdistanzen

Ein anderer, in dieser Arbeit verfolgter Ansatz ist, die typischen Bindungsabstände zwischen zwei Elementen  $a$  und  $b$  aus Rechnungen zu tabellarisieren, welche Geometrieoptimierungen im Rahmen der elektronischen Schrödingergleichung für das bivalente, gesättigte<sup>8</sup> System liefern, siehe Abbildung 9.21 zur Erstellung der anfänglichen Geometrie. Die Sättigung des bivalenten Systems mit Wasserstoff erlaubt es, unterschiedliche Bindungsgrade einer Bindung zu berücksichtigen, siehe Abbildung 9.22. Diese resultieren in unterschiedlichen Gleichgewichtslängen und Bindungsstärken.

### 9.3.2. Bindungsgradbestimmung

Nun widmen wir uns der Bestimmung des Grades jeder Kante im Graphen, also jeder Bindung. Dies ist wichtig, da wir nur mit diesem Wissen die Annahme der Ladungsneutralität pro Fragment erfüllen können, indem wir jede Bindung mit der nötigen Anzahl an Wasserstoffen saturieren.

<sup>8</sup>Da im Allgemeinen ein Element über mehr Valenzelektronen verfügt als es für die Bindung zu einem anderen benötigt, werden diese verbleibenden Valenzelektronen üblicherweise durch Bindungen an eine dazu passende Anzahl an Wasserstoffatomen gesättigt. Damit wird sichergestellt, dass die Geometrie grundsätzlich stabil ist.



**Abbildung 9.22.:** Bivalentes System aus zwei Kohlenstoffatomen, welches mit einer solchen Anzahl von Wasserstoffatomen saturiert wurde, damit eine Einfach-, Zweifach- bzw. Dreifach-Bindung zwischen den Kohlenstoffatomen bei geschlossener äußerer Schale der beiden Atome möglich ist.

### Lokale Heuristik

Zunächst untersuchen wir die Variante, welche eine Komplexität von  $\mathcal{O}(n \log(n))$  hat. Sie verfährt nach folgenden lokalen Kriterien, basierend auf der Ausgangssituation, wo alle Bindungen auf Grad 1 gesetzt sind, indem sie über alle Atome iteriert und dabei jeweils pro Schritt maximal eine Erhöhung des Grades um 1 durchführt:

- Erhöhe, wenn für beide Bindungspartner gilt, dass ihre jeweilige Valenz größer als die jeweilige Bindungsgradgewichtete Summe über alle Bindungen ist.
- Erhöhe die Bindung zu dem Bindungspartner zuerst, der mehr Bindungen als das aktuelle Atom hat.
- Erhöhe die Bindung zu dem Bindungspartner zuerst, der näher an einer vollen äußeren Schale ist.

Mit dem oben beschriebenen Verfahren korrigieren wir die anfänglichen Bindungsgrade von 1 des Graphen der verschiedenen Moleküle des schon bei den Skalierungsuntersuchungen verwendeten Testsatzes, siehe Tabelle 9.3. Abschließend zählen wir für jedes Molekül die Anzahl der Kerne, deren Valenz nicht der Summe über die Bindungsgrade entspricht.

Wir geben aber hier nur das Ergebnis an: Mehrere Moleküle werden nicht korrekt bestimmt, darunter Adenine und Benzonitrile. Sie verfügen über aromatische Ringe, siehe Abbildung 9.34.

### Maximum Matching

Die rein lokalen Änderungen der einfacheren Methoden erfassen nicht die Mesomeriestruktur von aromatischen Ringen, welche gleichzeitige Änderungen bedingt, und laufen deswegen in Sackgassen.

Um mit solchen Ringsystemen umgehen zu können, betrachten wir ein maximales Matching aller Kanten bzw. Bindungen, deren zugehörige Ecken bzw. Atome nicht gesättigt sind. Ein Matching ist eine Kantenmenge, bei der keine zwei Kanten eine gemeinsame Ecke haben. Innerhalb eines solchen Matchings können wir also alle Kanten- bzw. Bindungsgrade erhöhen und verändern dabei nicht die gesuchte Sättigungsbedingungen für ein Atom über mehrere Bindungen gleichzeitig. Zu einem nicht-erweiterbaren Matching

## 9. Experimente

kann keine Kante hinzugefügt werden, so dass die Kantenmenge immer noch ein Matching ist. Ein maximales Matching ist ein nicht-erweiterbares Matching mit der größtmöglichen Kantenmenge. Es kann dabei mehrere solcher maximalen Matchings geben.

In der Boost Graph-Bibliothek [155] ist der Algorithmus nach Edmonds [51] für das Finden eines Maximalen Matchings, siehe Tarjan [171], implementiert. Leider ist bisher kein Algorithmus für dieses Problem mit einer linearen Laufzeit bekannt, der implementierte Maximum Matching hat eine Laufzeit von  $\mathcal{O}(|V| \cdot |E|)$ , wobei  $|V|$  die Anzahl der Knoten und  $|E|$  die Anzahl der Kanten bezeichnet. Da die Anzahl der Kanten und Knoten jeweils beide mit  $\mathcal{O}(M)$  ansteigen, erhalten wir also eine *quadratische Skalierung*. Wir merken aber an, dass dabei nur die Anzahl der Kanten zu ungesättigten Atomen eingeht. Hängt diese nur mit  $\mathcal{O}(1)$  von der Anzahl der Kerne ab, so bleibt die Gesamtkomplexität linear.

In gleicher Weise wie vorher wenden wir dieses Verfahren zur Korrektur der anfänglichen Bindungsgrade von 1 an und zählen die am Ende falsch erkannten Grade.

Wir geben wieder nur das Ergebnis an: Mit diesem Vorgehen sind die Bindungsgrade für alle Moleküle des Testsatzes korrekt bestimmt worden.

### Literatur

Gleichwohl wir uns für unsere obigen Methoden auf keine Literatur gestützt haben, gibt es dennoch einige Artikel über das Problem der Bindungsgraphenbestimmung. Wir haben aber keine vollständige Analyse der Methoden mehr durchgeführt, sondern möchte nun ein paar andere Ansätze zum Vergleich anführen, die wir aus Zhang et al. [191] entnehmen, die in der Einleitung knapp zehn Referenzen zu verschiedenen Methoden auflisten. Wir haben dabei solche Methoden von vorneherein ausgeschlossen, wenn sie auf Bindungslängen, Fragmenten/funktionellen Gruppen in Datenbanken oder reinen Wahrscheinlichkeiten basieren. Die in dem Artikel [191] selbst beschriebene Methode scheidet aus diesen Gründen ebenfalls aus.

Die von Labute [125] vorgeschlagene Methode basiert jedoch ebenfalls wie unsere auf Maximum Matching. Völlig richtig wird darauf hingewiesen, dass unser Algorithmus in seiner jetzigen Form schlicht die Anzahl der Doppelbindungen im System maximiert. Im Artikel wird eine Gewichtung der einzelnen Bindungen angegeben und das Matching im Rahmen dieser Gewichte maximiert.

Eine im Rahmen unserer Bedingungen weitere sinnvolle Methode ist im Artikel von Wang et al. [180] unter dem Stichwort „Bond Type Perception“ beschrieben: Sie benötigt als Eingabe nur die Bond Connectivity. Bei dieser Methode wird eine sehr spezifische Penalty-Tabelle pro Element verwendet, wobei berücksichtigt wird, zu welchen anderen Elementen Bindungen bestehen. Die Tabelle gibt dann Auskunft, welche Valenz für das Element die geeignetste ist. Anschließend wird ein Algorithmus, basierend auf vier Regeln angegeben, die ähnlich unserem lokalen Schätzer sind, mittels deren eine Struktur mit kleinstem Penalty gefunden wird. Diese Methode würde mit ihrer Verwendung einer Tabelle zu den tabellarisierten Bindungsabständen, siehe Abschnitt 9.3.1. Wir würden uns jedoch ein Verfahren wünschen, um diese Penalties aus einer ab initio-Rechnung zu



bestimmen, damit sie unabhängig von chemischer Intuition eingesetzt werden kann, auch wenn sie Parameter enthält.

Wir werden auf mögliche, noch bestehende Unzulänglichkeiten, unserer Maximum Matching-Methode noch im folgenden Abschnitt über die Güte der Approximation zu sprechen kommen.

## 9.4. Approximationsverhalten

Nachdem wir das Skalierungsverhalten der Implementation ausführlich getestet haben, wollen wir uns nun mit der Approximationsgüte des Verfahrens beschäftigen.

Die Genauigkeit der Näherung wird von einer Reihe verschiedener Faktoren beeinträchtigt:

- Die Fockmatrix und die lokalisierten Eigenfunktionen werden durch Nullsetzen numerisch kleiner Koeffizienten gestört
- Die Wasserstoffsaturation erweitert die Fockmatrizen der Teilprobleme und stört diese. Besonders durchtrennte Mehrfachbindungen stellen eine größere Störung dar
- Delokalisierte Elektronen, wie sie in aromatischen Ringsystemen vorkommen, werden bei zu niedriger Bindungsordnung künstlich lokalisiert
- Die langreichweitigen Coulomb-Wechselwirkung wird vernachlässigt

Diesen möglichen Problemfällen widmen wir uns wie folgt: Zunächst untersuchen wir Alkane, da sie nur Einfachbindungen aufweisen, keine delokalisierten Elektronen haben und durch ihre homonukleare und sehr symmetrische Struktur über kaum langreichweitige Wechselwirkung verfügen. Anschließend untersuchen wir mit den Alkanen verwandte Kohlenwasserstoffketten mit Zweifach- und Dreifachbindungen, um einen möglichen Einfluss der Saturation auf die Genauigkeit hervorzuheben. Dabei messen wir die Genauigkeit auch anhand von Bor-Stickstoff-Ketten, welche in ihrer Struktur zu den Alkanen sehr ähnlich, aber heteronuklear sind. Auch untersuchen wir qualitativ, anhand von Benzol, den Einfluss von aromatischen Ringen auf die Genauigkeit der Approximation. Dann messen wir die Genauigkeit anhand eines Testsatzes kleiner organischer Moleküle, die über Mehrfachbindungen und Ringe verfügen. Weiterhin betrachten wir einzelne Systeme mit metallischem Charakter, die delokalisierte Elektronen besitzen, wie etwa in vernetzten Ringsystemen. Abschließend untersuchen wir die Trunkierung der langreichweitigen Wechselwirkung und deren Abschätzung mittels des besprochenen Störansatzes anhand des S22-Testsatzes [110], welcher aus Dimeren besteht.

Wir werden dabei stets zwei Grundzustandsenergien messen, einmal die Energie aus der Standard-HF-Methode  $E^{(\text{SCF})}$  und einmal  $E^{(\text{ANOVA})}$ , welche wir mittels der BOSSANOVA-Methode erhalten. Wir betrachten dann mit der Bindungs- bzw. Bindungsordnung  $k$  einerseits die relative Abweichung

$$\delta E(k) = \frac{E^{(\text{SCF})} - E^{(\text{ANOVA})}(k)}{E^{(\text{SCF})}}, \quad (9.4.1)$$

## 9. Experimente

aber auch die absolute Änderung

$$\Delta E(k) = E^{(\text{SCF})} - E^{(\text{ANOVA})}(k). \quad (9.4.2)$$

Alle Berechnungen dieses Versuchsteils zur Genauigkeit wurden sowohl auf dem zuvor verwendeten AMD64-System, als auch auf Intel Xeon X5650-Systemen (12 Kerne) mit 24 Gb Speicher durchgeführt. Dabei wurden MoleCuider in der Version 1.4.4<sup>9</sup>, JobMarket in der Version 1.1.4, CodePatterns in der Version 1.2.8 und MPQC in der Version 2.3.1 benutzt.

### 9.4.1. Genauigkeit

Zunächst definieren wir den Fehler einer Näherungsmethode für die HF-Gleichung als den Unterschied zu der Gesamtenergie aus der vollen HF-Gleichung. Er kann dabei als relativer Fehler (9.4.1) oder absoluter Fehler (9.4.2) bestimmt werden.

HF ist selbst nur eine approximative Methode, um das Elektronenstrukturproblem zu lösen. Vorallem aber erfolgt eine Standard-HF-Rechnung durch Entwicklung in einer endlichen Basis, was nach dem Variationsprinzip wiederum in einem Fehler resultiert, siehe Abschnitt 4.1.7.

Wir wollen deswegen zuerst den relativen Fehler einiger der endlichen Standard-Basen abschätzen, indem wir die mit ihnen für ein Testsystem berechnete Gesamtenergie mit der niedrigsten Energie für das gleiche System über alle möglichen Basen vergleichen, die in MPQC implementiert sind. Für die Nomenklatur der Basen verweisen wir auf Abschnitt A.1 im Anhang. Der daraus resultierende Fehler ist eine untere Schranke an den echten Fehler im Grenzwert einer unendlichen Basis, da HF variationell ist. Diesen Fehler nennen wir die erreichbare Genauigkeit der Basis. Unterhalb dieser basisabhängigen Schranke  $\Delta^{\text{basis}}$  betrachten wir dann den relativen Fehler  $\delta E$  unserer Näherungslösung als sehr gut.

Als Testsystem verwenden wir Hexan, welches ein Alkan mit 6 Kohlenstoffatomen ist. Wir haben zuvor schon die sehr symmetrische, homonukleare Struktur des Moleküls erwähnt, welche nur Einfachbindungen und lokalisierte Elektronen aufweist.

Zunächst betrachten wir in Abbildung 9.23 die Gesamtenergie der vollen HF-Rechnung über die verschiedenen untersuchten Basen. Wir haben dabei die sehr kleinen und einfachen Basen wie „STO-...“ oder „3-21...“ und auch solche, die pro Kurve nur sehr wenige Punkte liefern würden, nicht weiter berücksichtigt, um die Darstellung klar zu halten. In Tabelle 9.1 sind dann alle betrachteten Basen aufgelistet.

Wir bemerken allgemein, dass sich für Basen mit gleicher Anzahl an Basisfunktionen durchaus recht signifikante Unterschiede ergeben. Dies ergibt sich, weil die Basissätze jeweils für unterschiedliche Einsatzzwecke optimiert sind und deswegen nicht notwendigerweise optimal für das untersuchte Hexan.

Für unseren Fall liefert die „pc-2“-Basis die niedrigste Gesamtenergie von  $-235.4459$  Ht bei 376 Basisfunktionen. Da wir mit der nächstbesten Basis „6-311++G(3df,3pd)“ eine

---

<sup>9</sup>Das Programm erfuhr nur Änderungen, die den langreichweitigen Schätzer betreffen und die die zuvor zur Skalierung gemachten Messungen nicht verändern.

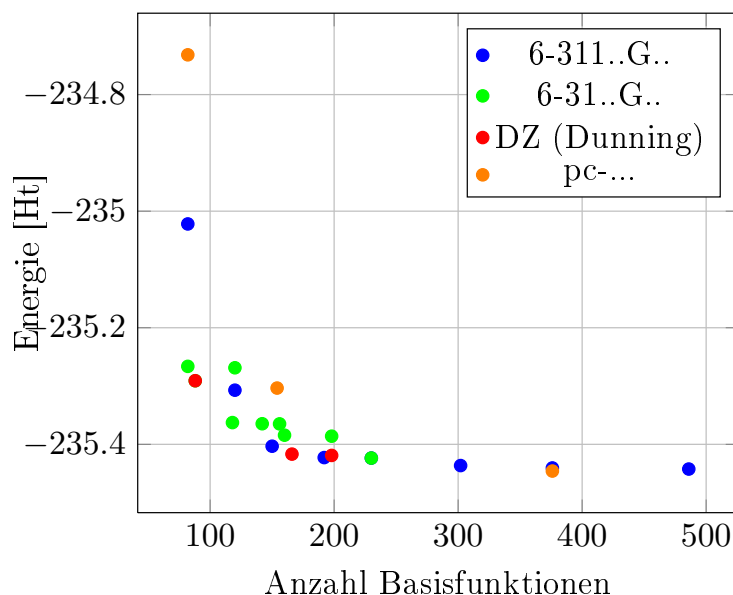


Abbildung 9.23.: Gesamtenergie von Hexan für verschiedene Standard-Basissätze.

Grundzustandsenergie von  $-235.4425$  Ht berechnen, erhalten wir bspw. einen relativen Unterschied zwischen diese beiden Werten von  $1.45 \cdot 10^{-5}$ . Alle weiteren relativen Fehler der gemessenen Basissätze sind in Tabelle 9.1 relativ zu „pc-2“ aufgelistet.

Eine zweite Fehlerschranke ergibt sich noch durch die endliche Ausgabegenauigkeit, welche in den dargestellten Plots zum Tragen kommt. Dies bezeichnen wir mit  $\Delta^{\text{float}} = 1.19 \cdot 10^{-7}$ . Eine höhere Ausgabegenauigkeit erscheint in Bezug auf die möglichen  $\Delta^{\text{basis}}$  wenig sinnvoll. Unterhalb dieser Schranke können also die dargestellten Fehler von numerischen Rundungsartefakten überlagert werden und daher nicht weiter sinnvoll interpretiert werden.

Als Kompromiss zwischen Genauigkeit und Geschwindigkeit werden die folgenden Berechnungen in der „6-311G\*“-Basis durchgeführt. Wir erwarten deswegen eine relative Genauigkeit von  $\Delta^{\text{basis}} \approx 10^{-4}$ .

### 9.4.2. Alkane

Mit diesen Vorüberlegungen zur Genauigkeit der endlichen Basis wollen wir nun verschiedene Kohlenwasserstoffketten als ersten Testfall für die implementierte, linear skalierende Methode untersuchen.

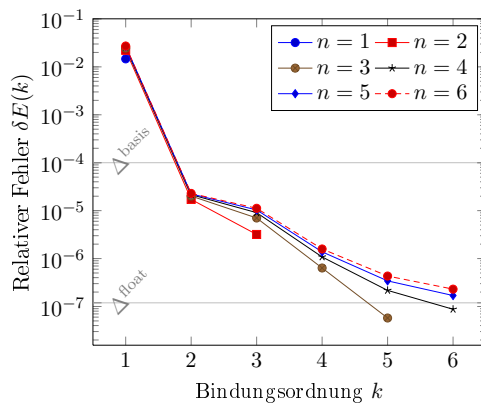
#### Idealfall

Zunächst betrachten wir die Alkane mit der Strukturformel  $C_{2n}H_{4n+2}$ , siehe [84, Abschn. 2]. Wir berechnen die Gesamtenergie, einmal des vollen Problems  $E^{(\text{SCF})}$ , siehe (4.1.17), und einmal als Teilprobleme per BOSSANOVA  $E^{(\text{ANOVA})}$ , siehe (7.2.6a). Wir ermitteln daraus den relativen Fehler (9.4.1) und den absoluten Fehler (9.4.2).

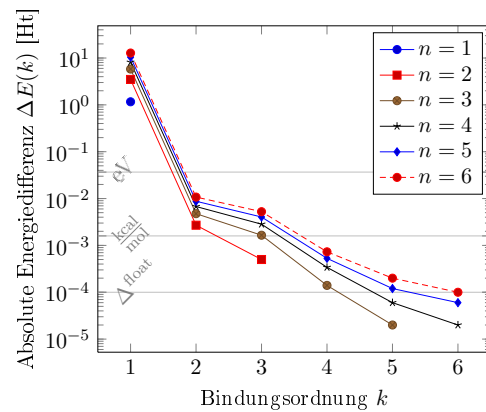
## 9. Experimente

**Tabelle 9.1.:** Liste der Basissätze mit  $n$  Basisfunktionen und ihrer gemessenen relativen Abweichung  $\Delta^{\text{basis}}$  der Gesamtenergie für Hexan von „pc-2“.

BASIS	n	$\Delta^{\text{basis}}$	BASIS	n	$\Delta^{\text{basis}}$
3-21G	82	$5.86 \cdot 10^{-3}$	6-31++G*	156	$3.44 \cdot 10^{-4}$
3-21G*	82	$5.86 \cdot 10^{-3}$	6-31++G**	198	$2.54 \cdot 10^{-4}$
3-21++G	120	$5.82 \cdot 10^{-3}$	DZ (Dunning)	88	$6.58 \cdot 10^{-4}$
3-21++G*	120	$5.82 \cdot 10^{-3}$	DZP (Dunning)	166	$1.23 \cdot 10^{-4}$
4-31G	82	$1.8 \cdot 10^{-3}$	DZP + Diffuse (Dunning)	198	$1.14 \cdot 10^{-4}$
6-311G	120	$5.89 \cdot 10^{-4}$	MIDI (Huzinaga)	82	$6.38 \cdot 10^{-3}$
6-311G(2df,2pd)	376	$2.24 \cdot 10^{-5}$	MINI (Huzinaga)	44	$8.81 \cdot 10^{-3}$
6-311G*	150	$1.81 \cdot 10^{-4}$	MINI (Scaled)	44	$6.91 \cdot 10^{-3}$
6-311G**	192	$9.81 \cdot 10^{-5}$	pc-0	82	$3.03 \cdot 10^{-3}$
6-311++G(2d,2p)	302	$3.99 \cdot 10^{-5}$	pc-1	154	$6.05 \cdot 10^{-4}$
6-311++G(3df,3pd)	486	$1.45 \cdot 10^{-5}$	pc-2	376	0
6-311++G**	230	$9.45 \cdot 10^{-5}$	pc-1-aug	264	$5.44 \cdot 10^{-4}$
6-31G	82	$7.63 \cdot 10^{-4}$	STO-2G	44	$4.05 \cdot 10^{-2}$
6-31G*	118	$3.53 \cdot 10^{-4}$	STO-3G	44	$1.2 \cdot 10^{-2}$
6-31G**	160	$2.61 \cdot 10^{-4}$	STO-3G*	44	$1.2 \cdot 10^{-2}$
6-31+G*	142	$3.45 \cdot 10^{-4}$	STO-6G	44	$2.38 \cdot 10^{-3}$
6-31++G	120	$7.52 \cdot 10^{-4}$	TZ (Dunning)	126	$5.51 \cdot 10^{-4}$



(a) Relativer Fehler zur Gesamtenergie



(b) Absoluter Fehler zur Gesamtenergie

**Abbildung 9.24.:** Approximation der Gesamtenergie für verschiedene Alkane pro Bindungsordnung  $k$ .

**Tabelle 9.2.:** Konvergenztabelle des relativen Gesamtenergiefehlers für Alkane verschiedener Länge  $n$  über die Bindungsordnung  $k$ .

$k$	$\delta E(n = 3)$	$\delta E(n = 4)$	$\delta E(n = 5)$	$\delta E(n = 6)$
1	$2.47 \cdot 10^{-2}$	$2.6 \cdot 10^{-2}$	$2.67 \cdot 10^{-2}$	$2.72 \cdot 10^{-2}$
2	$2.02 \cdot 10^{-5}$	$2.16 \cdot 10^{-5}$	$2.24 \cdot 10^{-5}$	$2.29 \cdot 10^{-5}$
3	$7.01 \cdot 10^{-6}$	$9.06 \cdot 10^{-6}$	$1.03 \cdot 10^{-5}$	$1.12 \cdot 10^{-5}$
4	$5.95 \cdot 10^{-7}$	$1.08 \cdot 10^{-6}$	$1.35 \cdot 10^{-6}$	$1.55 \cdot 10^{-6}$
5	$8.5 \cdot 10^{-8}$	$1.91 \cdot 10^{-7}$	$3.06 \cdot 10^{-7}$	$4.26 \cdot 10^{-7}$
6	0	$6.38 \cdot 10^{-8}$	$1.53 \cdot 10^{-7}$	$2.13 \cdot 10^{-7}$

In Abbildung 9.24 sind diese beiden Fehler inklusive der Fehlerschranken dargestellt. Für Alkane haben wir in Tabelle 9.2 den relativen Fehler festgehalten, für alle folgenden Untersuchungen halten wir jedoch die logarithmischen Darstellungen für ausreichend. Die Gesamtenergie für  $n = 3$ , also Hexan, entspricht für Ordnung  $k = 6$  exakt der vollen Energie: Der Fehler ist deswegen unter die Darstellungsgenauigkeit  $\Delta^{\text{float}}$  abgesunken.

Wir bemerken, dass der Fehler um ca. eine Größenordnung pro Bindungsordnung abfällt, dass wir also exponentielle Konvergenz erreichen. Natürlich wandern die Kurven im Plot der absoluten Fehler für längere Alkane nach oben, denn mit zunehmender Länge nimmt auch die Gesamtenergie zu. Um dann eine doppelt so lange Kette – mit ca. der doppelten Gesamtenergie – auf die gleiche absolute Genauigkeit zu berechnen, muss die Bindungsordnung um  $\log(2)$  erhöht werden. D. h. die Bindungsordnung für die gleiche Approximationsgüte für diese Kohlenwasserstoffketten wächst nur mit

$$k^{\text{fixed}} \propto \mathcal{O}(\log(n)).$$

### Gestörte Koordinaten

Wir wollen schließlich noch kontrollieren, dass das gute Abfallverhalten der Alkane nicht auf dessen perfekt symmetrischer Geometrie beruht.

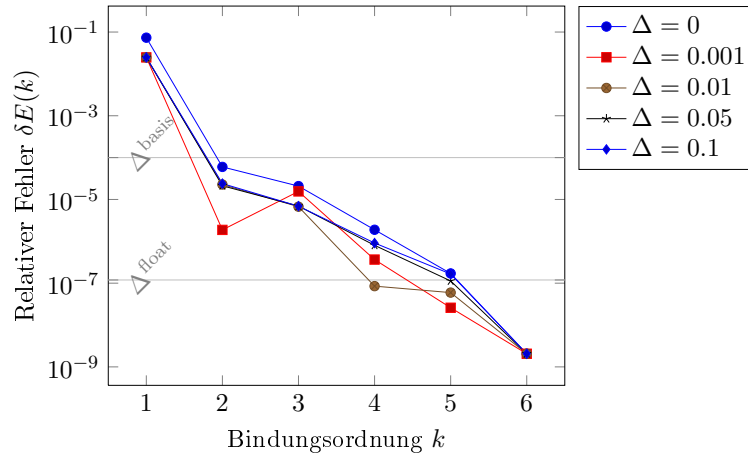
Dazu verändern wir die Kernorte von Hexan durch zufällige, uniform verteilte Störungen im Intervall  $[-\Delta, \Delta]$  mit  $\Delta = 0.001, 0.01, 0.05, 0.1$ .

Wir erkennen an Abbildung 9.25, dass die Konvergenz in allen Fällen gleich gut ausfällt. Damit ist diese Konvergenz also weitgehend unabhängig von der rein geometrischen Symmetrie.

### 9.4.3. Mehrfachbindungen

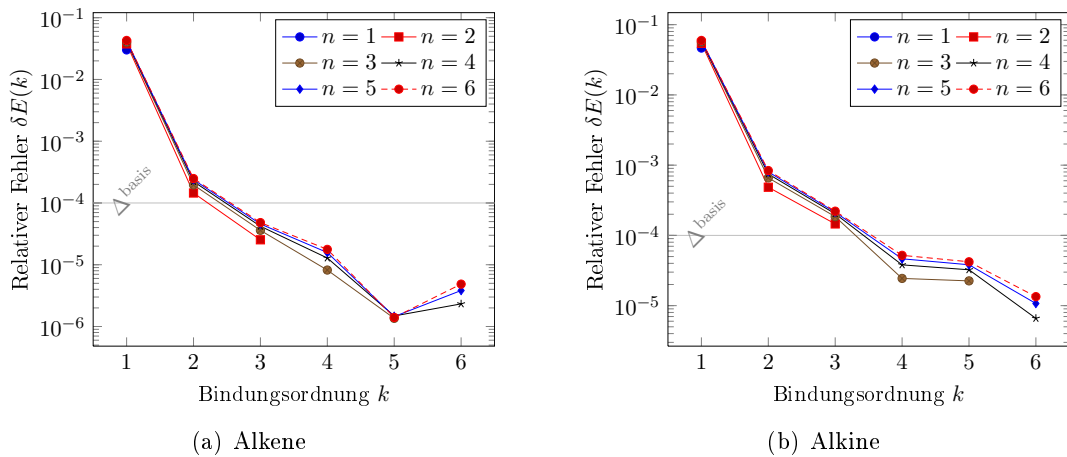
Anschließend betrachten wir die Auswirkung von Mehrfachbindungen im molekularen System. Um diese möglichst isoliert untersuchen zu können, berechnen wir die Gesamtenergie von Alkenen und Alkinen verschiedener Länge. Wir erwarten dabei eine geringere Genauigkeit aufgrund der Saturation mit Wasserstoffen, da dazu der genaue Ort der Mehrfachbindung festgelegt sein muss, siehe Abschnitt 9.3.2.

## 9. Experimente



**Abbildung 9.25.:** Approximation der Gesamtenergie für Hexan pro Bindungsordnung mit zufälligen Störungen der Geometrie in Intervallen der halben Breite  $\Delta$  um die Ausgangslage.

**Alkene und Alkine** Bei Alkenen wechseln sich entlang der Kette der Kohlenstoffatome eine Einfach- und eine Zweifachbindung ab, bei Alkinen sind es jeweils eine Einfach- und eine Dreifachbindung, siehe [84, Abschn. 3]. Sie verfügen über dementsprechend weniger Wasserstoffe, siehe auch Abbildung 9.22.



**Abbildung 9.26.:** Approximation der Gesamtenergie für verschiedene Alkene und Alkine pro Bindungsordnung  $k$ .

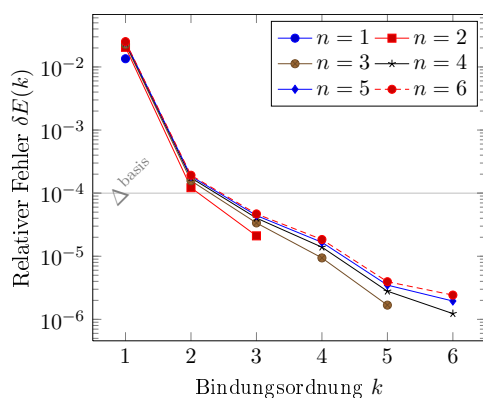
In den Abbildungen 9.26(a) und 9.26(b) sind die relativen Fehler für verschieden lange Alkene und Alkine dargestellt. Die Konvergenz ist diesmal etwas flacher als bei den Alkanen. In beiden Fällen bemerken wir ein unterschiedliches Verhalten bei  $k = 5$ : Bei den

Alkenen ergibt sich ein Knick, bei den Alkinen ein Plateau. Dieses Verhalten rührt vermutlich von den abwechselnden Einfach- und Mehrfachbindungen her, da es interessanterweise unabhängig von der Länge der Kette ist. Fragmente mit 5 Kohlenstoffatomen tauchen entweder mit zwei oder drei Mehrfachbindungen auf. Für Alkene scheint diese Mischung der beiden Fällen in der Summe die Ketten sehr gut zu beschreiben. Bei den Alkinen hingegen ergibt sich dadurch keine Verbesserung gegenüber  $k = 4$ , wo Fragmente nur genau zwei Dreifachbindungen besitzen. Der charakteristische Knick bei den Alkenen ist auch von der gewählten Basis abhängig: Bei „3-21G“ ist er nicht zu sehen. Er besteht in schwächerer Form auch bei der noch größeren, verwandten Basis „6-311G“.

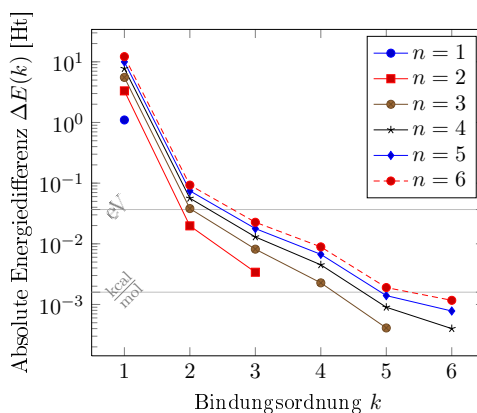
Wir halten fest, dass mit Mehrfachbindungen die Genauigkeit der Näherung bei gleicher Ordnung abnimmt, wie zu erwarten war. Dies kann neben der besprochenen Ursachen noch eine weitere haben: Eine Doppelbindung muss mit zwei Wasserstoffen und eine Dreifachbindung mit drei Wasserstoffen gesättigt werden, die sich aufgrund der räumlichen Nähe auch gegenseitig stören. Dieser Effekt sollte bei jedem Fragment in gleichem Maße bestehen. Insgesamt funktioniert die Näherung aber immer noch sehr gut und  $\Delta^{\text{basis}}$  wird um  $k = 3$  erreicht.

Im Unterschied dazu betrachten wir auch noch BN-Kettenmoleküle, welche die gleiche Struktur wie Alkene, aber nur Einfachbindungen aufweisen.

**BN-Kettenmoleküle** BN-Kettenmoleküle haben eine analoge Struktur zu Alkenen mit je zwei Wasserstoffen weniger, aber nur Einfachbindungen, da nur je drei statt vier Valenzelektronen zur Bindung frei sind. Ihre Strukturformel ist  $B_n N_n H_{2n+2}$ .



(a) Relativer Fehler zur Gesamtenergie



(b) Absoluter Fehler zur Gesamtenergie

**Abbildung 9.27.:** Approximation der Gesamtenergie für verschieden lange BN-Kettenmoleküle pro Bindungsordnung  $k$ .

In [Abbildung 9.27\(a\)](#) ist der verbleibende Fehler zur Gesamtenergie aufgetragen, in [Abbildung 9.27\(b\)](#) die relative Energieänderung pro Bindungsordnung. Wir sehen, dass der Fehler sich hier sehr ähnlich wie bei den Alkenen verhält. Der charakteristische Knick

## 9. Experimente

beim Übergang von  $k = 4$  zu  $k = 5$  bleibt aber aus, stattdessen geht die exponentielle Konvergenz weiter. Dies ist ein weiterer Hinweis darauf, dass der Knick mit der ungeraden Bindungsordnung der Kette zusammenhängt.

### 9.4.4. Konvergenz der einzelnen Energieanteile

Die im vorherigen Unterabschnitt betrachteten Kettenmoleküle konvergieren nicht alle gleich. Um die Ursache festzustellen, untersuchen wir nun die einzelnen Anteile an der Gesamtenergie getrennt. Dies sind die kinetische Energie, die Austauschenergie und die drei verschiedenen Coulomb-Wechselwirkungen: Elektron-Elektron, Kern-Elektron und Kern-Kern. Diese Energieanteile  $E_i$  sind in den Abschnitten 4.1.5 und 3.3.2 definiert, die wir hier kurz wiedergeben.

$$\begin{aligned}
 E_{\text{kin}} &= 2 \sum_{i=1}^{\frac{N}{2}} \int \phi_i^*(\mathbf{r}_1) \hat{T}^{(\text{el})} \phi_i(\mathbf{r}_1) d\mathbf{r}_1, \\
 E_{\text{XC}} &= \sum_{i,j=1}^{\frac{N}{2}} K_{ij}, \\
 E_{\text{EE}} &= 2 \sum_{i,j=1}^{\frac{N}{2}} J_{ij}, \\
 E_{\text{EN}} &= 2 \sum_{i=1}^{\frac{N}{2}} \int \phi_i^*(\mathbf{r}_1) \hat{V}^{(\text{core,el})} \phi_i(\mathbf{r}_1) d\mathbf{r}_1, \\
 E_{\text{NN}} &= \sum_{k < j}^M \frac{Z_k Z_j}{\|\mathbf{R}_k - \mathbf{R}_j\|}, \\
 E_{\text{coul}} &= E_{\text{EE}} + E_{\text{EN}} + E_{\text{NN}}.
 \end{aligned}$$

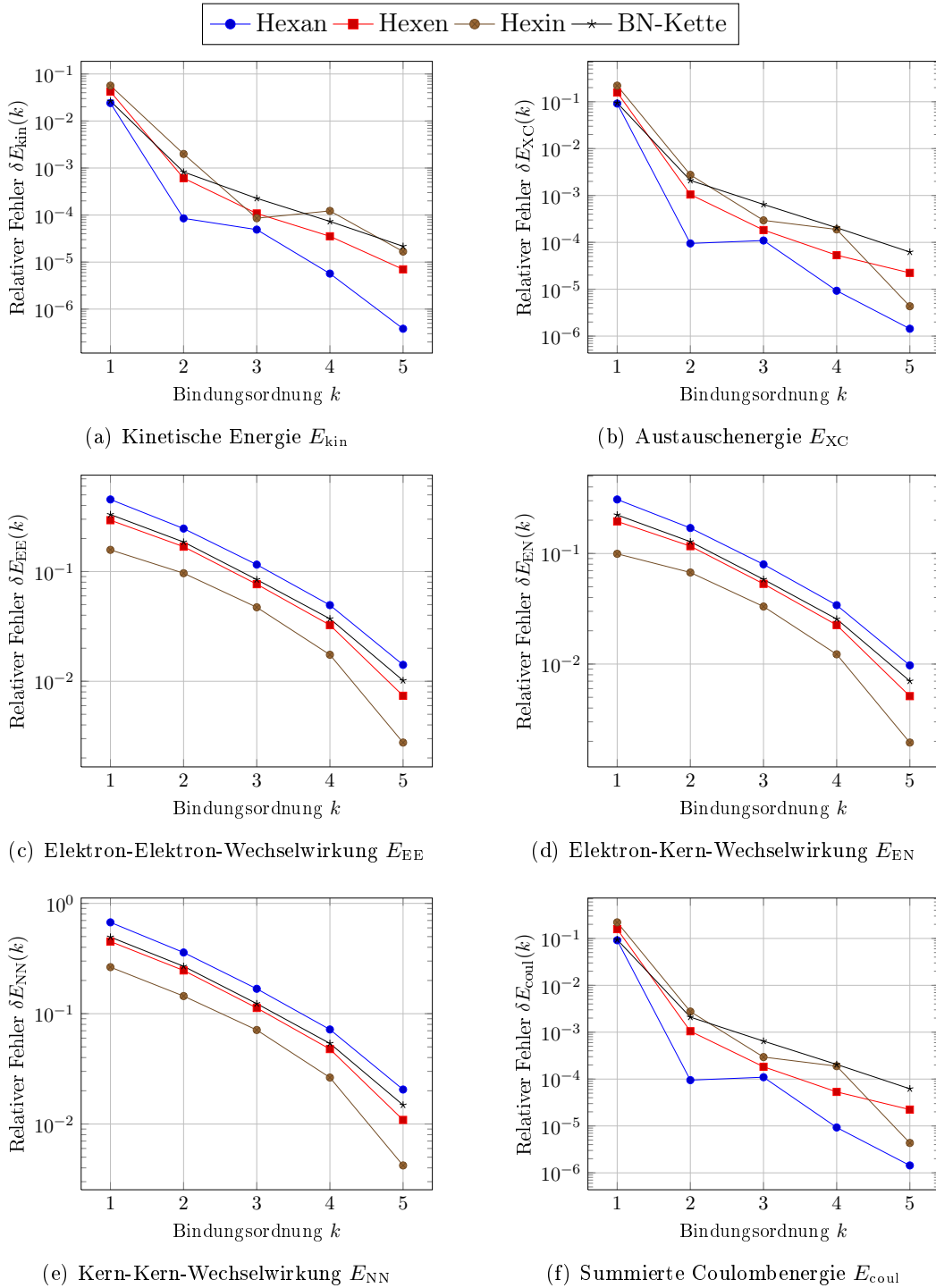
Wir untersuchen also den relativen Fehler dieser einzelnen Energieanteile

$$\delta E_i(k) = \frac{E^{(\text{SCF})}_i - E^{(\text{ANOVA})}_i(k)}{E^{(\text{SCF})}_i} \quad (9.4.3)$$

mit  $i = \{\text{kin, XC, EE, EN, NN, coul}\}$  und der Bindungsordnung  $k$ . Dabei bezeichnet  $E^{(\text{SCF})}_i$  den Energieanteil der vollen Lösung und  $E^{(\text{ANOVA})}_i(k)$  den mittels der BOSSANOVA-Methode genäherten Anteil  $i$  bei Ordnung  $k$ . Ziel ist zu ermitteln, welcher Anteil schlechter als andere durch BOSSANOVA erfasst wird.

In Abbildung 9.28 sind die relativen Fehler der einzelnen Anteile für die feste Länge  $n = 3$  dargestellt. Dabei wird mit Ordnung  $k = 6$  prinzipbedingt die Genauigkeit der vollen Lösung erreicht. Wir erkennen deutlich die exponentielle Konvergenz für die kinetische Energie 9.28(a) und für die Austauschenergie 9.28(b), wie wir es aufgrund des





**Abbildung 9.28.:** Relativer Fehler der einzelnen Anteile an der Gesamtenergie der betrachteten Kettenmoleküle der festen Länge  $n = 3$  über die Bindungsordnung  $k$ . Bei  $k = 6$  ist der Fehler für die betrachteten Moleküle exakt null.

## 9. Experimente

exponentiellen Faktors bei der Gauß-Basis auch erwarten. Die einzelnen Coulombenergien 9.28(c),9.28(d),9.28(e) hingegen konvergieren aufgrund des wirksamen Faktors  $1/r$  wesentlich langsamer.

Die summierte Coulomb-Wechselwirkung 9.28(f) hingegen fällt ähnlich rasch ab wie die ersten beiden Anteile. Da sich Ladungen nur entlang der Kette befinden, wirkt hier der Effekt der Abschirmung. Auch wenn die elektronischen Ladungsanteile von Alkanen, über Alkene bis zu Alkinen immer stärker in den wechselnden Bindungsgraden konzentriert werden, so ergibt sich dennoch keine große langreichweitige Wechselwirkungsenergie entlang der Kette. Auch bei den BN-Ketten mit ihren unterschiedlichen Kernladungen ergibt sich keine solche Wechselwirkungsenergie.

Stärker ist der Effekt der Lokalisierung. Dies wirkt in gleichem Maße auf alle Anteile – kinetisch, Austausch und Coulomb – denn Zweifach- bzw. Dreifachbindungen benötigen mehr Platz und sind daher schwächer lokalisiert als Einfachbindungen. Deswegen nimmt der relative Fehler von Alkanen, über Alkene zu den Alkinen zu.

In Abbildung 9.29 sind noch einmal die relativen Fehler der drei Anteile für die vier verschiedenen Molekülsysteme bei fester Länge  $n = 3$  geplottet.

Wir halten fest, dass die summierte Coulomb-Energie in keinem der bisher untersuchten Systeme gegenüber der kinetischen und der Austauschenergie eine große Rolle spielt. Ihr Fehler liegt um bis zu einer Größenordnung unterhalb des Fehlers der anderen Energieanteile. Wir beobachten stattdessen, dass alle drei Anteile in ähnlichem Maße zunehmen. Dies spricht für eine geringere Lokalisierung der Lösungen.

Natürlich kann auch die Wasserstoffsaturation Ursache der langsameren Konvergenz mit der Bindungsordnung sein. Aber die Alkene und die BN-Moleküle zeigen eine sehr ähnliche Konvergenz, gleichwohl einmal Doppelbindungen vorliegen und einmal nicht. Obwohl die Doppelbindung durch zwei Wasserstoffe saturiert werden muss, wirkt sich dies nicht auffällig auf.

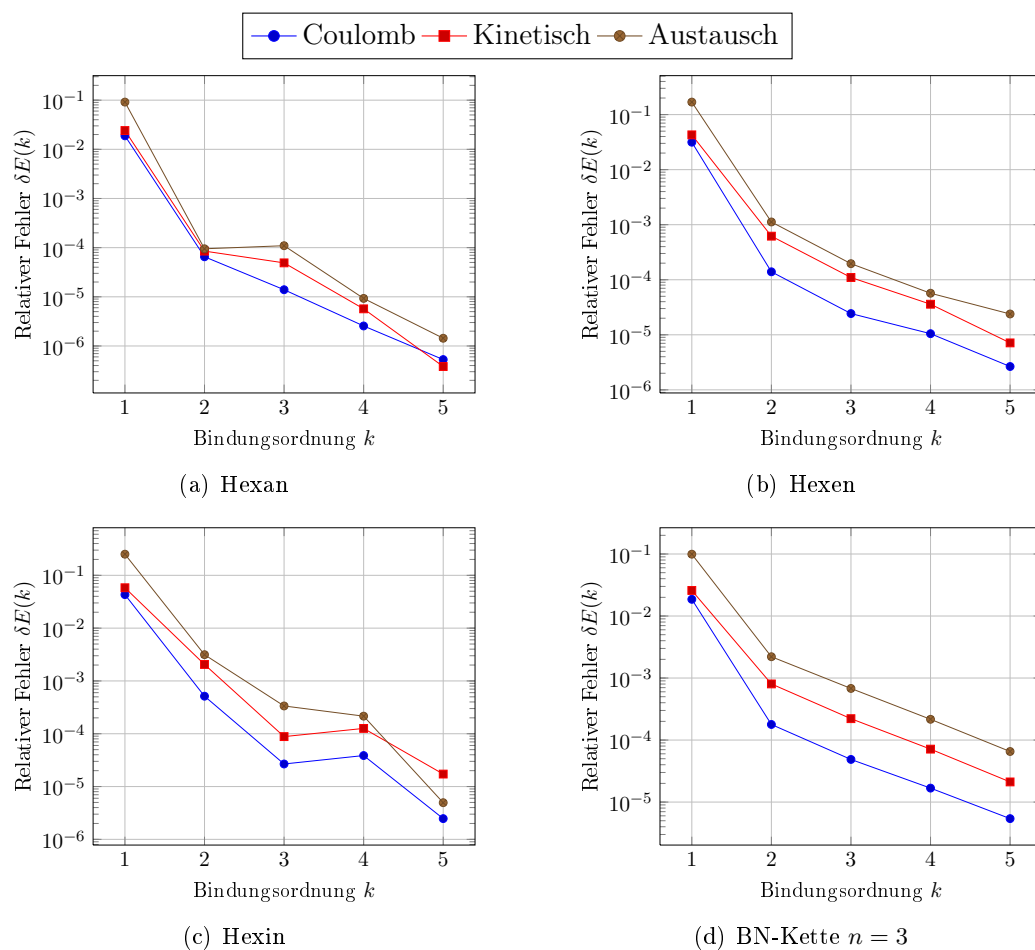
### 9.4.5. Delokalisierung

Wir betrachten nun Systeme mit delokalisierten Elektronen. Wir möchten zunächst betonen, dass dies der in Abschnitt 7.1.1 geforderten Annahme widerspricht. Wir erwarten also grundsätzlich kein konvergentes Verhalten wie bei den zuvor betrachteten Kettenmolekülen.

#### **Benzol**

Wir untersuchen zunächst Benzol als prototypisches System mit delokalisierten Elektronen. Die  $\pi$ -Orbitale bilden dort zwei Ringe oberhalb und unterhalb der Symmetrieebene des Moleküls, in denen sich die in diesen Orbitalen aufhaltenden Elektronen frei bewegen können, siehe [84, Abschn. 4.4]. Der daraus resultierende Energiegewinn ist leicht aus der Unschärferelation zu verstehen, da die Elektronen im Ring deutlich weniger örtlich eingeschränkt sind.

In Abbildung 9.30 ist der relative Gesamtfehler für den Benzolring mit steigender Bindungsordnung  $k$  aufgetragen.



**Abbildung 9.29.:** Relativer Fehler der einzelnen Anteile an der Gesamtenergie pro Kettenmolekül der festen Länge  $n = 3$  über die Bindungsordnung  $k$ . Bei  $k = 6$  ist der Fehler für die betrachteten Moleküle exakt null.

Bis Ordnung  $k = 3$  ergibt sich das gleiche Konvergenzverhalten, wie wir es auch bei den Kohlenwasserstoffketten beobachtet haben. Aber dann kehrt sich das Verhalten um und die Approximation wird bis  $k = 5$  immer schlechter. Bei  $k = 6$  erhalten wir prinzipbedingt die volle Lösung.

Der Grund liegt darin, dass die einzelnen Fragmente aus dem Benzolring, siehe Abbildung 9.31, immer stärker gekrümmt sind. Die Bindungen im vollen Benzolring sind dies auch. Der Benzolring erhält aber durch die Delokalisierung der  $\pi$ -Orbitale einen Bonus, welcher den Malus aus der zur Krümmung aufgewendeten Energie mehr als wett macht. Diesen Bonus kann die BOSSANOVA-Methode aber erst ab Ordnung  $k = 6$  einbeziehen.

Anders formuliert: Die Eigenvektoren, welche die delokalisierten Orbitale beschreiben, gehören zu Hauptuntermatrizen, deren Größe zur Basis des Fragments der Ordnung  $k = 6$  passt. In den Fragmenten kleinerer Ordnung werden sie künstlich trunziert und

## 9. Experimente

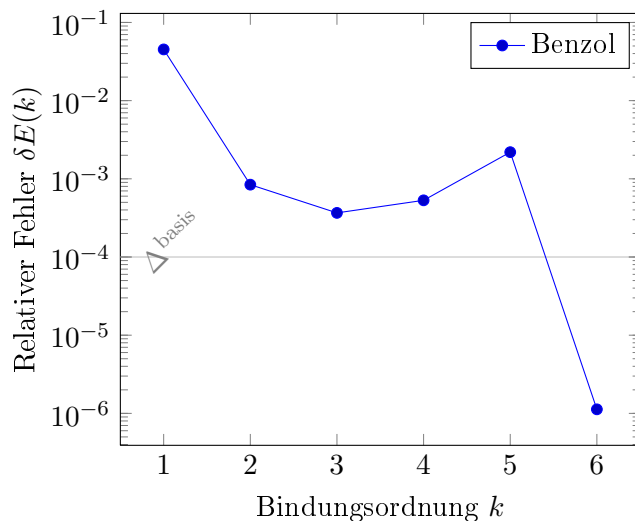


Abbildung 9.30.: Approximation der Gesamtenergie eines Benzolrings pro Bindungsordnung  $k$ .

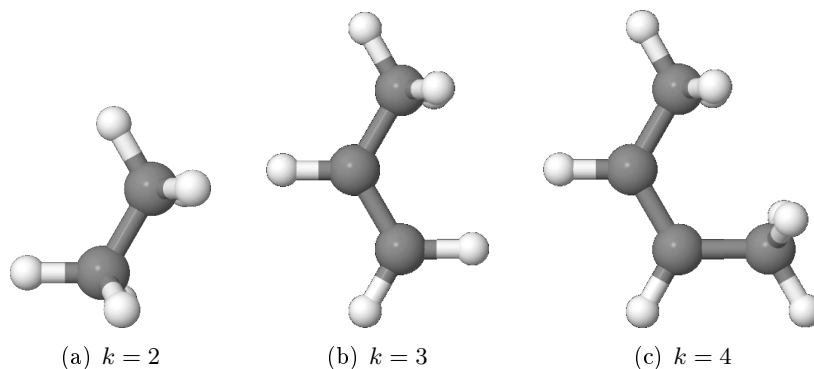


Abbildung 9.31.: Darstellung der zunehmenden Krümmung der Fragmente des Benzolrings mit ansteigender Ordnung  $k = 2, 3, 4$ .

beschreiben deswegen nicht die richtige Energie. Ein anderer Ansatz wäre hier, nicht nach Kernen aufzutrennen, sondern nach Elektronengruppen. So könnten die  $\pi$ -Elektronen des Rings als gemeinsame Gruppe beschrieben werden. Ein solcher Ansatz wurde in [118] verfolgt, siehe Abschnitt A.4, wo wir auf verwandte Ansätze in der Literatur näher eingehen.

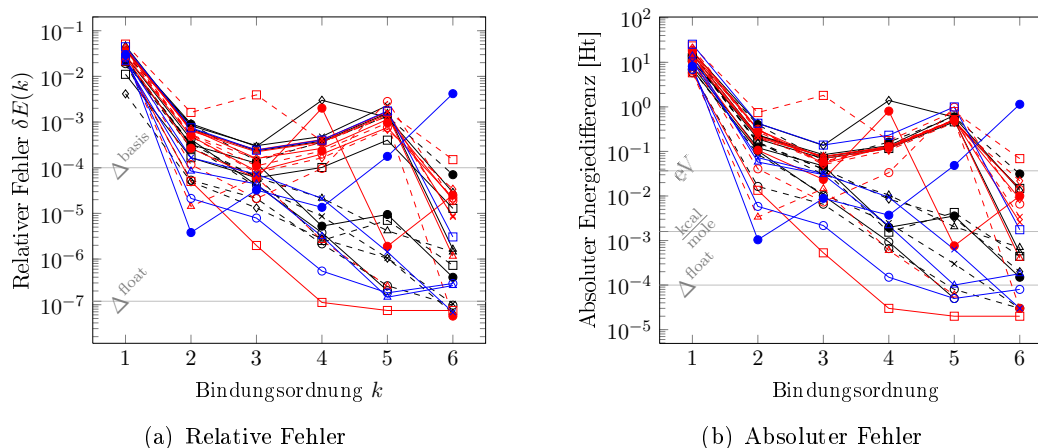
### Organischer Testsatz

Wir wollen im Weiteren den schon vorher im Rahmen der Skalierungsuntersuchungen verwendeten Testsatz von organischen Molekülen nun auf Konvergenz im Rahmen der BOSSANOVA-Methode untersuchen.

**Tabelle 9.3.:** Auflistung aller Moleküle des Testsatzes mit Legende.

NR.	LEGENDE	MOLEKÜL	NR.	LEGENDE	MOLEKÜL
1	●	oxalic acid	16	□	putrescine
2	●	1,3,5-triazine	17	△	2-phenylpropene
3	○	acrylamide	18	◇	acetanilide
4	□	1,4-dichlorobenzene	19	●	mandelic acid
5	△	benzonitrile	20	×	m-methylanisole
6	◇	adenine	21	○	2-methylcyclohexanone
7	●	melamine	22	□	biphenylene
8	×	N,N-dimethylacetamide	23	△	neohexane
9	○	1,2-dimethoxyethane	24	◇	aspirin
10	□	tartaric acid	25	◇	cycloheptane
11	△	asparagine	26	×	isoleucine
12	◇	dimethyl bromomalonate	27	○	heptane
13	●	proline	28	□	benzophenone
14	×	1,2-dimethylbenzene	29	△	glucose
15	○	gallic acid			

In Abbildung 9.32 sind die relativen und absoluten Fehler über den Testsatz dargestellt. Die Legende zu einzelnen Kurven findet sich in Tabelle 9.3, wo der Testsatz aufgelistet ist.



**Abbildung 9.32.:** Relativer und absoluter Fehler der Gesamtenergie über die Moleküle des Testsatzes mit der Bindungsordnung  $k$ . Die Legende zu den Kurven findet sich in Tabelle 9.3.

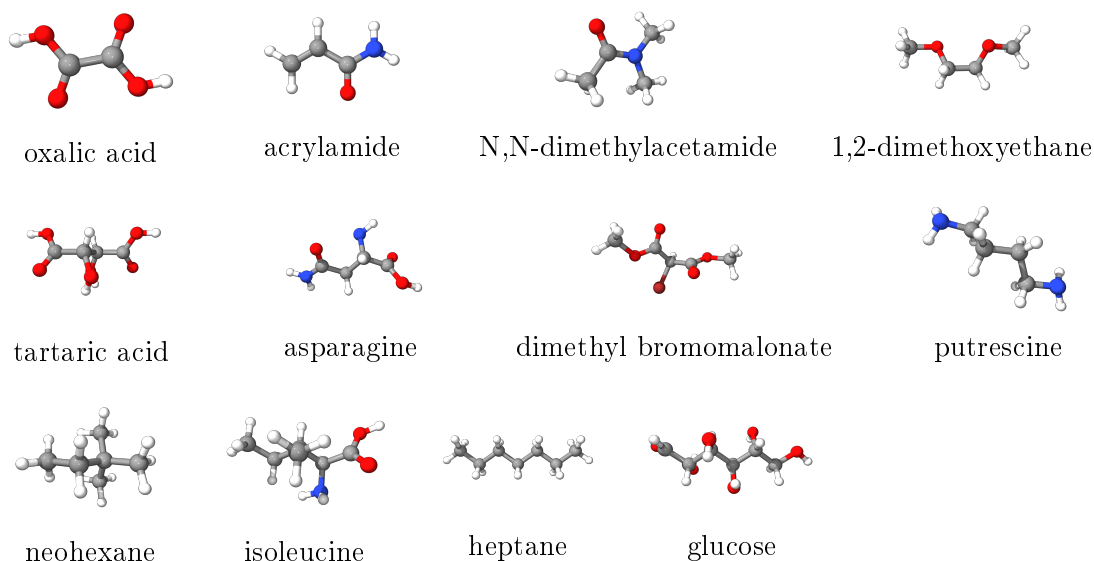
Wir erkennen grundsätzlich zwei Gruppen von Molekülen, insbesondere im Verlauf bis Ordnung  $k = 5$ .

## 9. Experimente

Eine Gruppe zeigt ein ähnliches Verhalten wie die zuvor betrachteten Kettenmoleküle, nämlich ein exponentieller Abfall des Gesamtenergiefehlers. Sie landen im Bereich unterhalb von  $10^{-4}$  bzw. annähernd im Bereich von 1 kcal/mol bei  $k = 5$ .

Die zweite Gruppe zeigt nach einem ähnlichen Abfall bis zur Ordnung  $k = 3$  anschließend wieder ein Anstieg, also eine Verschlechterung der Näherung mit höherer Ordnung.

Darüberhinaus sehen wir, dass Biphenylen, siehe -□-, eine Verschlechterung schon bei Ordnung  $k = 3$  und nicht erst danach zeigt. Weiter ist besonders in Abbildung 9.32(b) zu sehen, dass es eine ganze Reihe von übereinanderliegenden Kurven zu Molekülen gibt, die also den gleichen absoluten Fehler zur Gesamtenergie mit gleichem Verlauf von  $k = 2$  bis  $k = 5$  besitzen.

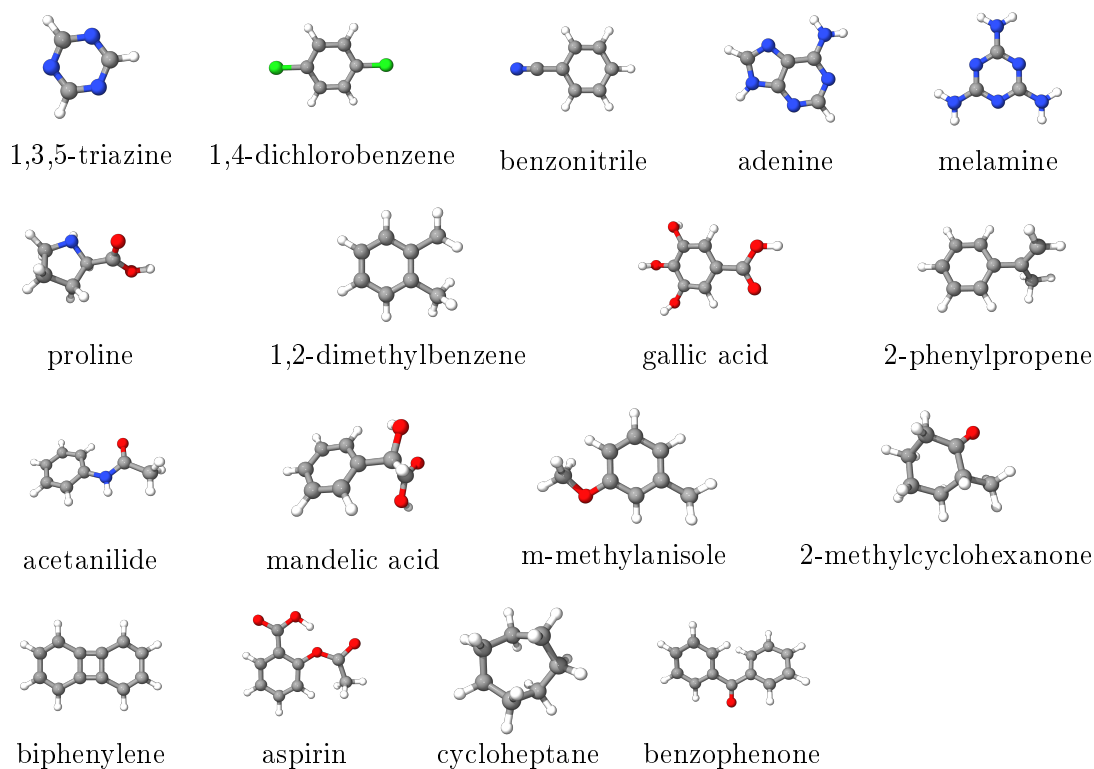


**Abbildung 9.33.:** Darstellung der Moleküle des Testsatzes ohne Ringstrukturen als Kugel-Stab-Modelle.

Mit unserer vorangegangenen Untersuchung von Benzol und der Abbildung der einzelnen Moleküle als Kugel-Stab-Modelle in den Abbildungen 9.33 und 9.34, getrennt nach solchen Systemen ohne und solchen mit Ringstrukturen, können wir dieses Verhalten leicht nachvollziehen: Diese Unterteilung trennt die Moleküle auch genau in die zuvor erläuterten beiden Gruppen mit Alkanähnlichem bzw. Benzolähnlichem Verhalten. In der zweiten Gruppe liegen dabei all jene Moleküle, welche über mindestens einen Ring verfügen.

Auf eine gewisse Weise kann man jedoch auch die in den aromatischen Ringen delokalisierten Elektronen als lokalisiert betrachten, indem man den Ring als Ganzes auffasst. Dies bedeutet, dass sobald wir den ganzen Ring als Fragment mitberücksichtigen, sollte sich wieder ein konvergentes Verhalten einstellen. Dies ist auch in [57] vorgeschlagen worden.

Im Rahmen der BOSSANOVA-Methode wird eine Analyse des Bindungsgraphen durchgeführt, welche auch die kleinsten Zyklen des Graphen erkennt, siehe Abschnitt 9.2.3.



**Abbildung 9.34.:** Darstellung der Moleküle des Testsatzes mit Ringstrukturen als Kugel-Stab-Modelle.

Aufgrund der Freiwählbarkeit der Unterindexmengen in der Methode, solange die Untermengeneigenschaft gewährt bleibt, siehe Abschnitt 6.1.3, können wir diese Zyklen einfach mit einbeziehen.

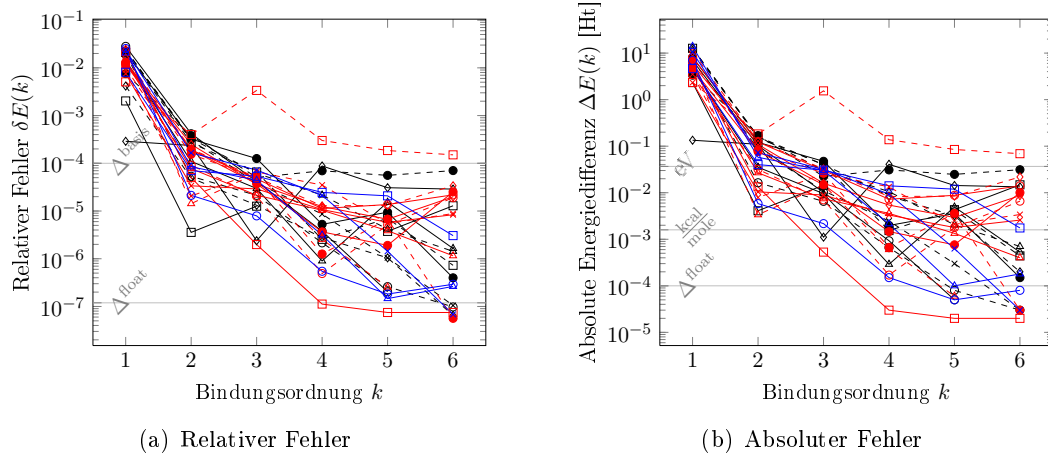
Die dadurch resultierende relative und absolute Gesamtenergie ist in Abbildung 9.35 festgehalten. Bei Ordnung  $k = 3$  setzt sich also die Näherung aus allen Fragmenten der Größen 1,2 und 3 und allen vollständigen Zyklen zusammen. 1,3,5-Triazin wird herbei also bereits bei Ordnung  $k = 1$  vollständig berechnet und die Näherung ist exakt, da es nur aus einem einzelnen Ring besteht, siehe Abbildung 9.34.

Wir vergleichen diese Abbildung 9.35 mit Abbildungen 9.32 und erkennen, dass die Aufspaltung in zwei Gruppen nahezu vollständig verschwunden ist. Über alle Moleküle des Testsatzes erreichen wir jetzt relative Genauigkeiten mindestens  $10^{-4}$  ab  $k = 4$ . Auf eine Ausnahme kommen wir gleich zu sprechen.

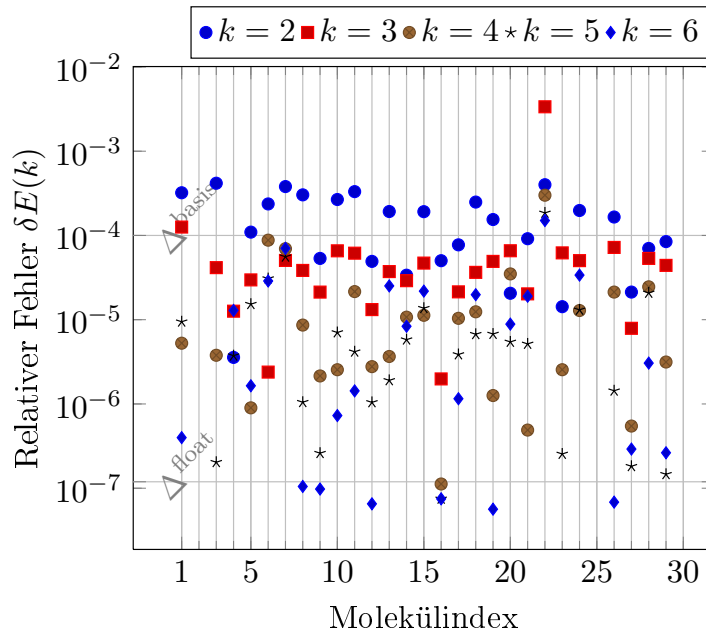
Diese Ausnahme ist noch einmal deutlicher in Abbildung 9.36 zu sehen, wo wir auf die durchgezogene Linie bei  $\Delta^{\text{basis}}$  und besonders auf die Datenpunkte der roten Kästchen zu  $k = 3$  verweisen, welche alle darunter liegen.

In Abbildung 9.37 haben wir die Energiebeiträge pro Fragment für jedes Molekül einzeln aufgeschlüsselt. Diese sind dabei nach Größe aufsteigend angeordnet, zuerst die Fragmente zu  $k = 1$ , dann die Fragmente zu  $k = 2$ , usw. . Wir haben dabei die nach der

## 9. Experimente



**Abbildung 9.35.:** Fehler der Gesamtenergie über die Moleküle des Testsatzes mit der Bindungsordnung  $k$  mit gesonderter Berücksichtigung aller Zyklen. Die Legende zu den Kurven findet sich in Tabelle 9.3.



**Abbildung 9.36.:** Auflistung des relativen Fehlers zur Gesamtenergie über die Moleküle des Testsatzes, siehe Tabelle 9.3 für die Ordnungen  $k=2, \dots, 6$ .



gesonderten Behandlung der Zyklen verbleibenden Problemfälle und zusätzlich Heptan dargestellt. Allgemein sollten wir gewisse Ebenen erkennen: Fragmente gleicher Größe tragen ungefähr auch in gleicher Größenordnung zur Gesamtenergie bei. Die Energiebeiträge zu diesen Ebenen wiederum sind durch schnellen Abfall gekennzeichnet. Dies ist der Hintergrund begründet in der gewünschten exponentiellen Konvergenz.

**Heptan** Die Konvergenz von Heptan ist auch am schnellen Abfall der Beiträge mit zunehmender Größe deutlich zu erkennen. Hier treten die Plateaus für Beiträge gleicher Größe aufgrund der Symmetrie des Systems sehr deutlich hervor.

**Prolin, Adenin** Prolin und Adenin fallen vor allem durch ihre Ausreißer bei  $k = 6$  auf. Deswegen ergeben sich die ansteigenden Flanken bei dieser Bindungsordnung.

**Biphenyl** Betrachten wir den Bindungsgraphen in Abbildung 9.38, welchen die Methode für Biphenyl erstellt, so fällt auf, dass der zentrale Vierer-Ring nur eine Bindung vom Grad 2, aber drei Bindungen vom Grad 1 hat. Wir vermuten, dass hier der Fehler zu suchen ist: Der Maximum-Matching-Algorithmus findet eines unter mehreren möglichen maximalen Matchings. Dabei existieren mehrere Varianten für die Bindungsgrade im zentralen Vierer-Ring: Einmal mit zwei Doppelbindungen, zweimal mit nur einer Doppelbindung und einmal nur mit Einfachbindungen. Das 4er-Ringfragment erhält aber die niedrigste Energie, wenn es über zwei Doppelbindungen verfügt. Nur in dieser Variante ergäbe sich eine sehr gute Näherung an die Grundzustandsenergie.

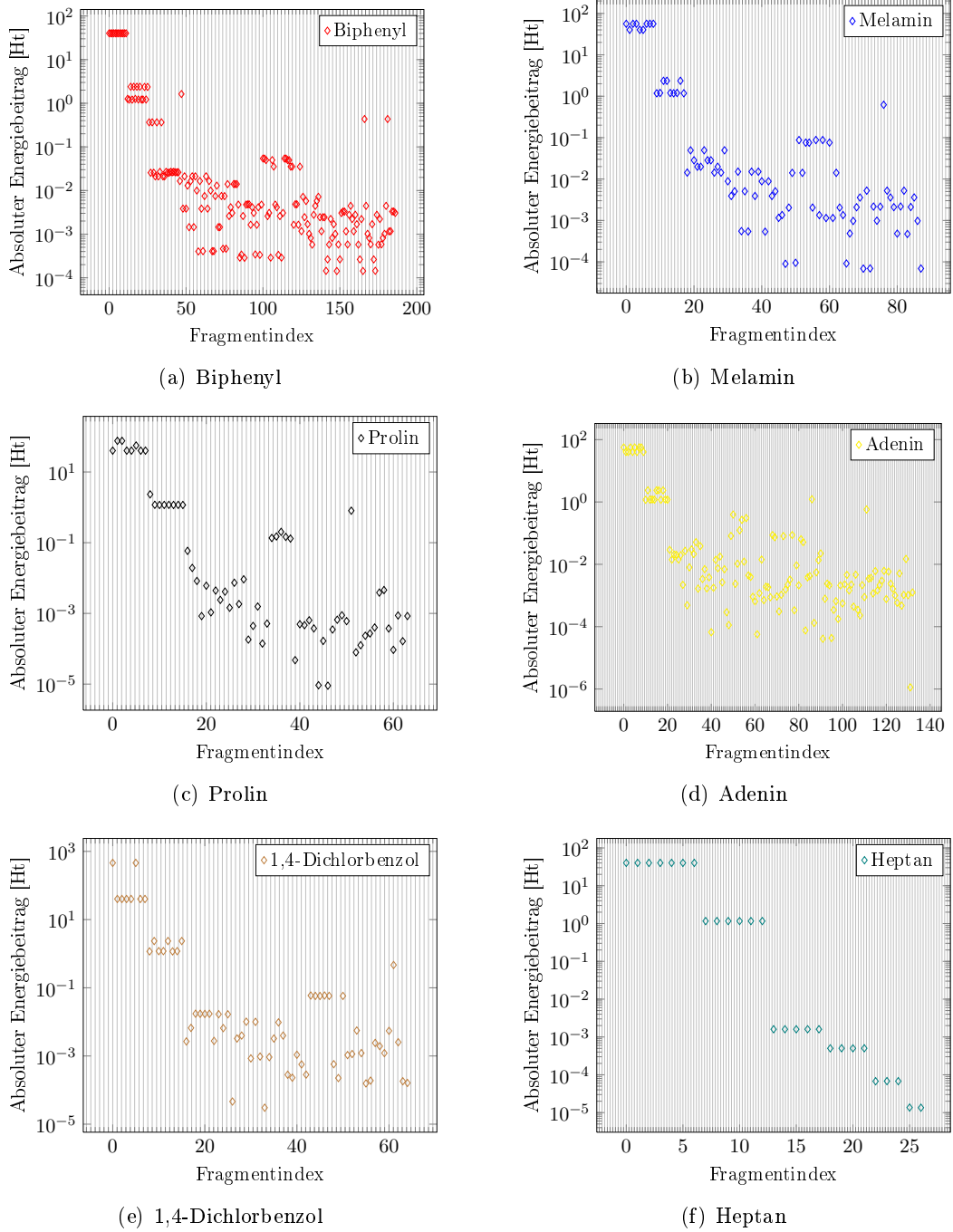
**Melamin, 1,4-Dichlorbenzol** Bei Melamin und 1,4-Dichlorbenzol greift dieses Argument aber nicht, da es bei ihnen diese Varianten nicht gibt und der Bindungsgradalgorithmus die richtige Grade bestimmt. Dafür ergeben sich jedoch besonders Beiträge von Fragmenten der Ordnung  $k = 5$ , die mit  $10^{-1}$  hervortreten und auch bei  $k = 6$  mit  $10^{-2}$ , bei denen der 6er-Ring fast vollständig erfasst wird, so dass sich die Saturationswasserstoffe von zwei entfernten Bindungen sehr nahe kommen.

Wir ziehen ein abschließende Resümee über die Ergebnisse zu diesem Testsatz: Obwohl geläufige organische Moleküle über aromatische Ringe und damit über delokalisierte Elektronen verfügen und damit die zentrale Annahme der Lokalität nur eingeschränkt gilt, lassen sie sich dennoch sehr gut berechnen. Der Fehler sinkt also in den Bereich von  $\Delta^{\text{basis}}$ , wenn die Ringe ganz berücksichtigt werden, oder auch zum größten Teil schon bei  $k = 3$  ohne gesonderte Berücksichtigung. Wir bemerken, dass die Hinzunahme der vollständigen Ringe nichts an dem linearen Skalierungsverhalten der Methode ändert, wenn die Größe der Ringe unabhängig von der Systemgröße ist und die Anzahl höchstens linear mit dieser steigt.

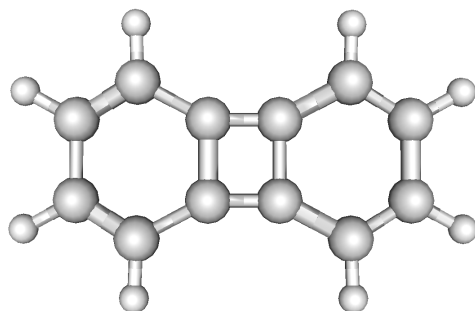
## Metallische Systeme

Makroskopische metallische Systeme lassen sich meist anhand von kristallinen Einheitszellen beschreiben. Die Berechnung dieser Einheitszellen gewinnt aber nicht durch unsere

## 9. Experimente



**Abbildung 9.37.:** Absoluter Beitrag zur Gesamtenergie pro Fragment über einige ausgesuchte Moleküle des Testsatzes mit der Bindungsordnung  $k$ . Dabei sind die Fragmentebeiträge sortiert nach der Ordnung  $k$ . Heptan zeigt das gewünschte Abfallverhalten, einige wenige Moleküle aus dem Testsatz in Tabelle 9.3 zeigen Aureißer.

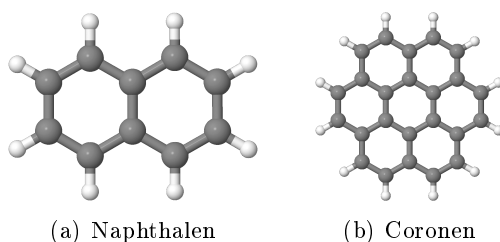


**Abbildung 9.38.:** Biphenyl als Kugel-Stab-Modell, wobei die Dicke der Stäbe den Bindungsgrad wiedergibt. Die linke Bindung im zentralen Vierer-Ring ist also eine Doppelbindung, die restlichen drei sind Einfachbindungen.

linear skalierende Methode. Da diese Zellen meist nur relativ wenige Atome mit hohem Bindungsgrad enthalten und unsere Methode erst ab 20 Kernen einen Laufzeitvorteil bringt, siehe Abschnitt 9.2.8, sehen wir der Berechnung solcher Systeme ab.

Alternativ könnten wir metallische Cluster an dieser Stelle betrachten. So hat beispielsweise ein Cluster aus Gold-Atomen ab  $M \approx 50$  Kernen metallische Eigenschaften, siehe Riedel und Janiak [147, S. 188]. Jedoch spielen hier aufgrund der Schwere des Kerns relativistische Effekte eine Rolle, siehe Wesendrup et al. [183] für Berechnungen auf HF/MP2-Niveau mit Pseudopotentialen in einer Artikelreihe über relativistische Effekte in Gold. Da wir den gleichen Löser wie in den vorherigen Untersuchungen verwenden wollen, sehen wir von diesen Systemen ebenfalls ab.

Stattdessen wollen wir noch einige Systeme aus der organischen Chemie diskutieren, die ausgesprochen metallischen Charakter durch vernetzte aromatische Ringe besitzen und damit auch Eigenvektoren der zugehörigen Fockmatrix, die sich nicht gut durch eine lokale Basis darstellen lassen. Als ausgedehntes prototypisches System dient hier das schon von uns im Rahmen der Skalierungsuntersuchungen besprochene Graphen. Wir wollen uns jedoch nur mit kleineren Systemen auseinandersetzen wie Naphthalen und Coronen, sogenannten polyaromatische Kohlenwasserstoffe. Die untersuchten Systeme sind in Abbildung 9.39 dargestellt.

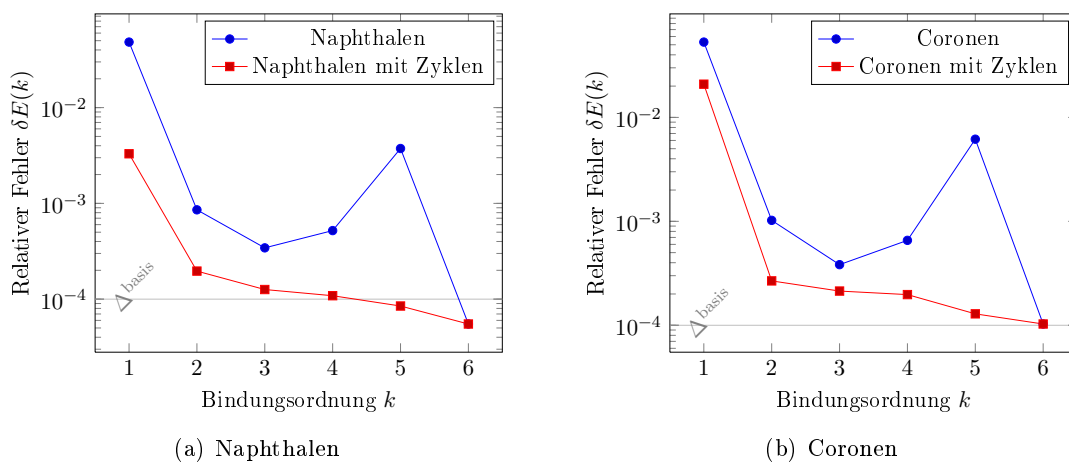


**Abbildung 9.39.:** Untersuchte Moleküle mit metallischem Bindungscharakter als Kugel-Stab-Modelle.

## 9. Experimente

**Naphthalen und andere polyaromatische Kohlenwasserstoffe** Generell merken wir an, dass schon die Rechnung in der Standard-HF-Methode schwieriger als zuvor wird. Innerhalb der gewählten Basis kommen wir nur mit Naphthalen und Coronen in unter 1000 Schritten unter die geforderte Änderungstoleranz von  $10^{-8}$ . Bei diesen Systemen reichen 10 – 20 Schritte.

In Abbildung 9.40 ist der relative Fehler der Gesamtenergie über die Bindungsordnung  $k$  aufgetragen. Dabei geben die Kurven einmal die approximierte Gesamtenergie ohne und einmal mit gesonderter Berücksichtigung aller Zyklen im Bindungsgraphen an.



**Abbildung 9.40.:** Relativer Fehler der Gesamtenergie für Moleküle mit delokalisierten Elektronen über die Bindungsordnung  $k$ . Einmal wurden dabei kleinste Zyklen im Graph unabhängig von der Ordnung  $k$  berücksichtigt, einmal nicht.

Wir erkennen das gleiche Bild wie auch schon bei dem organischen Testsatz. Ringsysteme werden ohne gesonderte Berücksichtigung der Zyklen mit zunehmender Bindungsordnung aufgrund der Krümmung der Fragmentsysteme zunehmend schlechter approximiert, bis der volle Ring als Fragment mit einbezogen wird. Weiterhin zeigt sich aber, dass sich ab Ordnung  $k = 3$  nicht mehr die gleiche starke Konvergenz einstellt wie bei den linearen Kohlenwasserstoffen, z. B. den Alkanen. Dies hängt mit den delokalisierten Elektronen zusammen, welche als Einteilchen-Wellenfunktion in einer lokalen Basis keine dünnbesetzte Darstellung haben und deswegen innerhalb der diagonalisierten Hauptuntermatrizen noch nicht vollständig erfasst werden.

Wir halten noch fest, dass die Ergebnisse von dem jeweilig gefundenen Bindungsgraphen und besonders den zugeordneten Bindungsgraden abhängen. Eine kleine, zukünftige Korrektur könnte darin liegen, dass wir nicht nur einen der möglichen Bindungsgraphen benutzen. In der Chemie spricht man von Mesomerie im Zusammenhang von aromatischen Ringen und der Resonanz zwischen mehreren möglichen Bindungszuständen, siehe [31, S. 131] oder auch [84, S. 26]. Eine bildliche Vorstellung ist dabei, dass sich in einem aromatischen Ring Doppel- und Einfachbindungen kontinuierlich abwechseln. Bei einem symmetrischen Benzolring haben beide Varianten des Graphen, siehe [84, Abschn. 4.2]

zu diesen Kekulé-Strukturen, die gleiche Energie, weswegen nur eine Variante berechnet werden muss. Bei den betrachteten polyaromatischen Kohlenwasserstoffen gilt dies jedoch nicht mehr zwangsläufig, siehe auch unsere Bemerkungen zu Biphenyl im letzten Abschnitt.

#### 9.4.6. Korrelation

In diesem letzten Unterabschnitt der Approximationsanalyse beschäftigen wir uns mit der Korrelationsenergie. Dazu benutzen wir den S22-Testsatzes, welcher von Jurecka et al. [110] für die hochgenaue Rechnung zur Eichung und zum Vergleich vorgeschlagen wurde.

Eine Auflistung der Molekülekomplexe des S22-Testsatzes [110] findet sich in den Tabellen in Abbildung 9.41 und 9.42. Jedes System besteht aus zwei Molekülen in bestimmter Ausrichtung zueinander.

##### Monomere

Wir bestimmen zunächst für jedes einzelne Monomer in jedem System die volle Korrelationsenergie innerhalb der MPPT-Theorie in der Basis „6-311G\*“. Außerdem bestimmen wir die Näherung an diese Energie im Rahmen der BOSSANOVA-Methode über wachsende Bindungsordnungen  $k = 1, 2, 3$  mit gleicher Theorie und gleicher Basis. Wir haben uns dabei auf eine maximale Bindungsordnung von  $k = 3$  beschränkt, um die bereits behandelten Schwierigkeiten mit Ringsystemen außen vor lassen zu können. Wir haben mögliche Auswirkungen aufgrund des sogenannten Basissatz-Superpositionsfehlers (BSSE) unberücksichtigt gelassen, siehe [14], der jedoch üblicherweise kleiner als 1 kcal/mol ist.

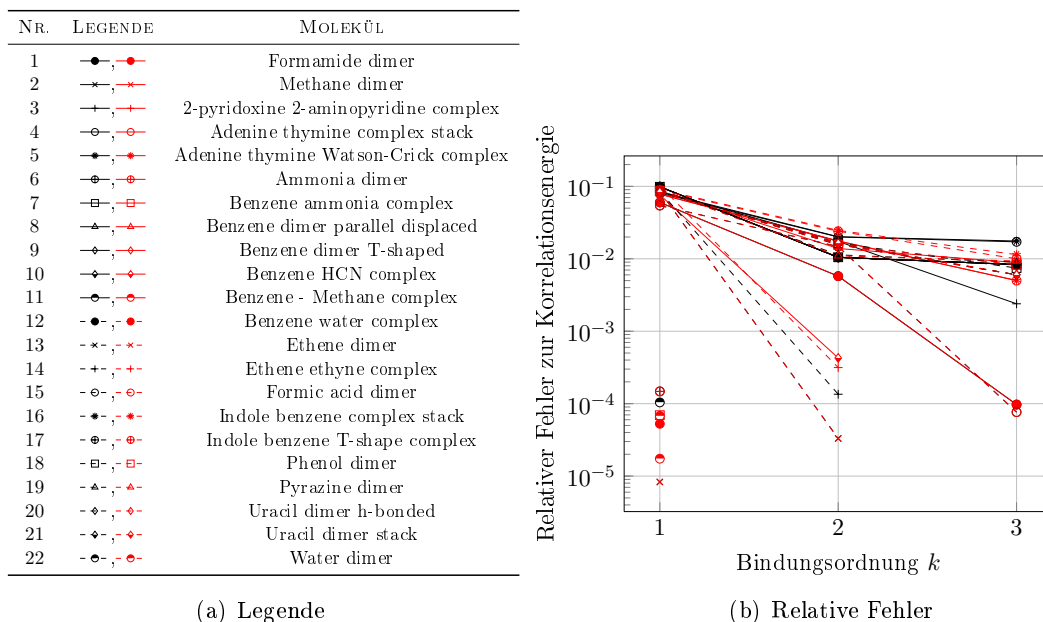
In [91, Abschn. 5.8.3] wird die Genauigkeit der MP2-Rechnung für ein Wassermolekül bestimmt. Dabei ist der absolute Unterschied zur FCI-Rechnung für die Gleichgewichtsgeometrie ca.  $10^{-2}$  Ht. Dies entspräche einem relativen Fehler in der Korrelationsenergie von  $10^{-2}$  bei den untersuchten Systemen. Deswegen werden wir uns hier mit einer solchen Genauigkeit im relativen Fehler der Korrelationsenergie begnügen.

Wir merken an, dass die MPPT-Theorie auf Basis einer einzelnen Slater-Determinante sich nur für Gleichgewichtsgeometrien gut eignet. Bei Geometrien fern dieses Zustands ist der von ihr erfasste Anteil der Korrelationsenergie wesentlich geringer, siehe [91, Abschn. 5.8.3]. Diese Eigenschaft und ihre Größenkonsistenz passen aber gut zu der von uns vorgestellten Methode, welche den Bindungsgraph benutzt, der nur in der Nähe der Gleichgewichtsgeometrie sinnvoll bestimmt werden kann. Deswegen haben wir uns für diese Methode entschieden.

In Abbildung 9.41 ist zunächst der relative Fehler der Korrelationsenergie für jedes Monomer aufgetragen.

Wie wir sehen, erreichen wir eine relative Genauigkeit von  $10^{-2}$ . Da der relative Anteil der Korrelationsenergie an der Gesamtenergie typischerweise nur im Promillebereich liegt, ist dies auch in Bezug zu unserer geforderten Basisgenauigkeit  $\Delta^{\text{basis}}$  ein akzeptabler Wert.

## 9. Experimente



**Abbildung 9.41.:** Relativer Fehler in der Korrelationsenergie jedes Monomers des Testsatzes über die Bindungsordnungen  $k$ . Fehlende Punkte resultieren aus der zu kleinen höchstmöglichen Ordnung des jeweiligen Monomers. In der Legende stehen dabei die Symbole der beiden Monomere des zugehörigen Molekülsystems.

### Dimere

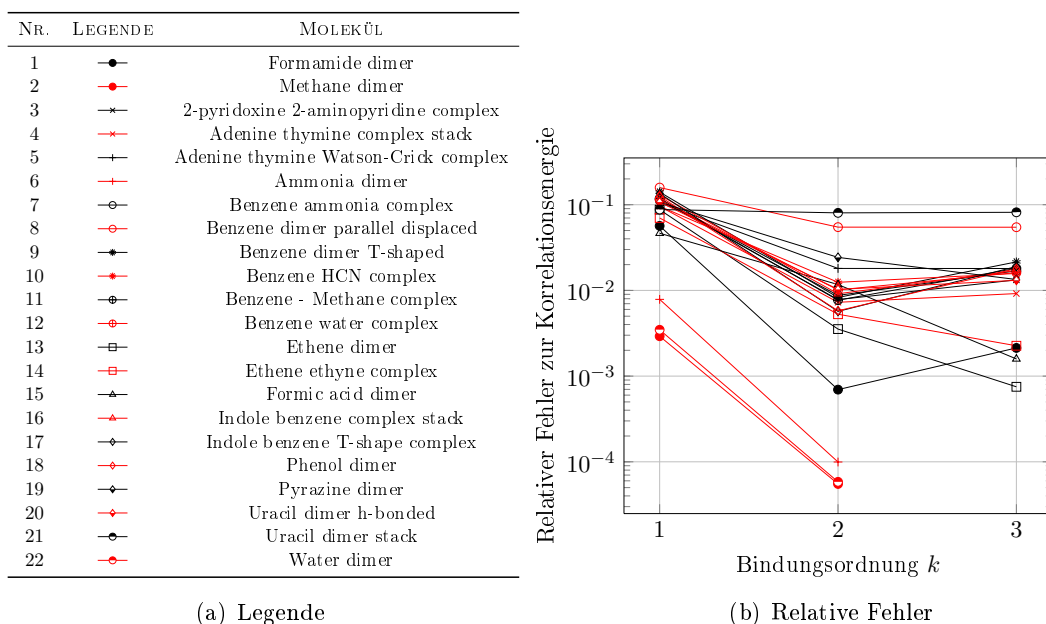
Da die Korrelationsenergie nicht an kovalente Bindungen gebunden ist, werden zusätzlich alle Fragmente bis Ordnung 1 in Paaren bis zu einer Entfernung von 5 Å voneinander mit berücksichtigt. Hier werden also Indexmengen der Größe  $|\mathcal{V}| = 2$  zusätzlich berücksichtigt, die keinen verbundenen Teilgraphen induzieren. Ohne diese zusätzlichen Indexmengen würden wir keine Korrelationsenergie zwischen den Monomeren berechnen.

In [Abbildung 9.42](#) haben wir dazu den relativen Fehler der Korrelationsenergie für jedes Molekül des Testsatzes aufgetragen.

Auch hier erreichen wir grundsätzlich eine Genauigkeit von  $10^{-2}$ . Zwei Konfigurationen, darunter bspw. „Uracil dimer stack“, erreichen jedoch nur  $10^{-1}$ .

Eine tiefere Untersuchung würde über den Rahmen dieser Arbeit hinausgehen. Es soll an dieser Stelle nur gezeigt werden, dass es mit akzeptabler Genauigkeit möglich ist, einen Teil der Korrelationsenergie zumindest im Rahmen der MP2-Theorie sinnvoll aus Fragmentlösungen zu berechnen. Dies spielt besonders für die Erfassung wie van-der-Waals-Bindungen eine Rolle, die schwächer als die kovalenten Bindungen sind, siehe [Abschnitt A.2](#) im Anhang für eine Zusammenfassung der verschiedenen Bindungstypen.

## 9.5. Abschätzung der langreichweitigen Wechselwirkung



**Abbildung 9.42.:** Relativer Fehler in der Korrelationsenergie der Dimere des Testsatzes über die Bindungsordnungen  $k$ . In der Legende steht dabei das Symbol des jeweiligen Molekülsystems. Fehlende Punkte resultieren aus der zu kleinen höchstmöglichen Ordnungen der einzelnen Monomere.

## 9.5. Abschätzung der langreichweitigen Wechselwirkung

In diesem Abschnitt wollen wir zunächst ausführlicher beschreiben, wie die Berechnung der langreichweitigen Wechselwirkungen mittels eines Störansatzes in *MoleCuilder* verwirklicht wurde.

Wenn wir, wie bisher, nur Hauptuntermatrizen der Fockmatrix behandeln, dann vernachlässigen wir den Teil der Coulomb-Wechselwirkung von Ladungen zwischen den Hauptuntermatrizen. Mit anderen Worten, es fehlt jede Wechselwirkung zwischen den Ladungen von Elektronen bzw. Kernen, die nicht in einem gemeinsamen Fragment vorkommen.

Die naheliegende Idee, um diese fehlende Wechselwirkung zu berechnen, ist die Vernachlässigung als eine Störung des Operators aufzufassen: Wir trennen den Fockoperator, und damit die Fockmatrix, für jedes einzelne Fragment in zwei Anteile auf. Die elektronischen Wellenfunktionen als Lösungen des Eigenwertproblems nehmen anschließend eine Gestalt ein, bei der sie nur innerhalb der ersten Anteils minimiert wurden. Wir können jedoch die Energie bezüglich des zweiten Anteils im Rahmen der Störtheorie erster Ordnung unter der Annahme wieder erlangen, dass die Störung durch die langreichweitige Wechselwirkung klein ist. Dies haben wir in Abschnitt 7.4 ausführlicher besprochen. Würden wir entsprechend höhere Ordnungen der Störtheorie benutzen, erhielten wir auch Korrekturen zu den ungestörten Einteilchen-Wellenfunktionen.

## 9. Experimente

Methoden zur Berechnung der langreichweitigen Coulomb-Wechselwirkung finden vor allem Einsatz in der Moleküldynamik, siehe [9] für eine aktuelle Übersicht, wo diese Methoden ergänzend zu Cluster- und anderen empirischen Potentialen [30] den langreichweitigen Anteil der Wechselwirkung berücksichtigen, wie er bspw. für Simulationen mit polarisierbaren Medien wie Wasser notwendig ist. Deswegen wurden diese Methoden vor allem für die Wechselwirkung der Kerne, also Punktladungen mit bestimmter effektiver Ladung, entwickelt.

Unsere Problemstellung ist nun aber leicht anders: Aus der Lösung der linearisierten HF-Methode des letzten Kapitels erhalten wir eine sehr gute Abschätzung der elektronischen Ladungsdichte. Deswegen brauchen wir keine effektiven Ladungen. Wir können die Elektronen direkt und (zumindest im klassischen Sinne) vollständig berücksichtigen.

### 9.5.1. Motivation Mehrgitter

Wir möchten kurz erläutern, warum wir uns im Folgenden für eine Mehrgittermethode anstelle bspw. der bereits gängigen Multipolmethoden entscheiden. Mehrgittermethoden lösen die diskretisierte Poisson-Gleichung zum Coulomb-Problem durch Iterationsverfahren für dieses lineare Gleichungssystem. Multipolmethoden benutzen eine Entwicklung des Coulomb-Kerns durch Legendre-Polynome, strukturieren Ladungen hierarchisch und drücken sie mittels Multipolen aus.

Das Hauptargument für den Mehrgitteransatz ist, dass für die Störungsrechnung eine begrenzte Genauigkeit ausreichend erscheint. Da wir aus dem kurzreichweitigen Anteil eine maximale Genauigkeit von  $\Delta_{\text{basis}} = 10^{-5}$  erwarten und die langreichweitige Coulomb-Wechselwirkung im Allgemeinen keinen relativen Anteil größer als  $10^{-3}$  an der Gesamtenergie hat, benötigen wir nur eine relative Genauigkeit von  $10^{-3}$ , um den Fehler abzuschätzen, den wir durch Trunkierung der langreichweitigen Wechselwirkung innerhalb der Fragmente machen.

Die Schwäche des Mehrgitters gegenüber der Multipolmethode ist insbesondere das notwendige Sampling der elektronischen Ladungsdichte und der punkartigen Kernladungen. Gerade weil wir erstere auch noch als lokalisiert voraussetzen, kann von einem gitterbasierten Ansatz zur Darstellung dieser Ladungsdichten keine hohe Genauigkeit erwartet werden. Aus diesem Grund können wir in den folgenden Messungen auch keine reine Konvergenz mit der Gittermaschenweite erwarten, da sich mit wechselnder Maschenweite auch das Sampling der Ladungsdichte ändert.

Die Schwäche der Schnelle Multipol-Methode (FMM) ist, dass diese Methode ursprünglich für Punktladungen gedacht war, siehe Greengard und Rokhlin [72]. White et al. [184] erweiterten sie auf kontinuierliche Ladungsverteilungen. In [164] wird diese Erweiterung speziell für Gauß-Funktionen vorgestellt. Aber dadurch ist die Methode konzeptionell für das vorliegende Problem deutlich schwieriger als der Mehrgitteransatz.

Wir betrachten deswegen die Mehrgittermethode als einen hervorragenden Kompromiss zwischen ausreichender Genauigkeit und einfacher Implementation im Rahmen dieser Arbeit.



### 9.5.2. Berechnung des Coulomb-Potentials

Im Rahmen dieser Arbeit benutzen wir die Mehrgitter-Implementation in Iseringhausen [104], siehe auch [22], für das Poisson-Problem unter offenen Randbedingungen.

Wir werden zunächst erläutern, wie der Mehrgitteralgorithmus prinzipiell funktioniert. Dabei müssen drei Dinge getan werden: Zunächst müssen wir die rechte Seite der Gleichung aufstellen, dazu muss die elektronische Ladungsdichte und auch die Punktladungen der Kerne auf das Gitter übertragen werden. Anschließend folgt das Lösen des Gleichungssystems. Schließlich muss das Potential vom Gitter rückinterpoliert werden.

#### Mehrgitter-Methoden

Man kann die Integralgleichung (7.3.4) auch als eine Differentialgleichung, die sogenannte Poisson-Gleichung in atomaren Einheiten,

$$-\Delta\Phi = 4\pi\rho \quad (9.5.1)$$

mit dem Laplace-Operator  $\Delta$  umformulieren, siehe [77, Kap. 7].

Wir folgen [172, Kap. 2.2], aber siehe auch das klassische Werk von Hackbusch [79]. Die Gleichung (9.5.1) ist eine prototypische elliptische partielle Differentialgleichung der folgenden Gestalt

$$\mathbf{A}^\Omega u(\mathbf{x}) = f^\Omega(\mathbf{x}), \quad (9.5.2)$$

wobei  $\mathbf{A}^\Omega = -\Delta$  auf einem kartesischen Gitter  $G_h$  mit Punkten

$$G_h = \{\mathbf{x} | \mathbf{x} = jh \text{ for } j \in \mathcal{Z}^3\} \quad (9.5.3)$$

durch einen formalen Diskretisierungsparameter  $h \in \mathbb{R}^3$  diskretisiert wird. Damit wird (9.5.2) zu

$$\mathbf{A}_h^\Omega u_h(\mathbf{x}) = f_h^\Omega(\mathbf{x}). \quad (9.5.4)$$

Die Lösung erhalten wir durch ein iteratives Relaxationsverfahren, welches den Fehler  $e_h^m = u_h - u_h^m$  der diskreten Lösung in jedem Iterationsschritt  $m$  minimiert:

$$u_h^{m+1} = u_h^m + \mathbf{C}_h \underbrace{(f_h - \mathbf{A}_h u_h^m)}_{d_h^m}, \quad (9.5.5)$$

wobei  $\mathbf{C}_h$  eine genäherte Inverse ist und  $d_h^m$  den Defekt im Iterationsschritt  $m$  bezeichnet. Wir trennen  $\mathbf{A}_h = \mathbf{L}_h + \mathbf{D}_h + \mathbf{U}_h$  in eine untere Dreiecksmatrix  $\mathbf{L}_h$ , eine Diagonalmatrix  $\mathbf{D}_h$  und eine obere Dreiecksmatrix  $\mathbf{U}_h$  auf. Mit  $\mathbf{C}_h := \mathbf{D}_h^{-1}$  erhalten wir Löser des Jacobi-Typs, mit  $\mathbf{C}_h := (\mathbf{L}_h + \mathbf{D}_h)^{-1}$  erhalten wir solche des Gauß-Seidel-Typs.

Relaxationsverfahren haben die Eigenschaft, die hochfrequenten Komponenten des Fehlers zu glätten, während die niederfrequenten Komponenten kaum berührt werden, siehe [172, Abschn. 1.5.1]. Was hoch- und was niederfrequente Anteile sind, hängt dabei mit der Gittermaschenweite zusammen. Daher konvergieren die Verfahren umso langsamer, je feiner das Gitter  $G_h$  aufgelöst ist.

## 9. Experimente

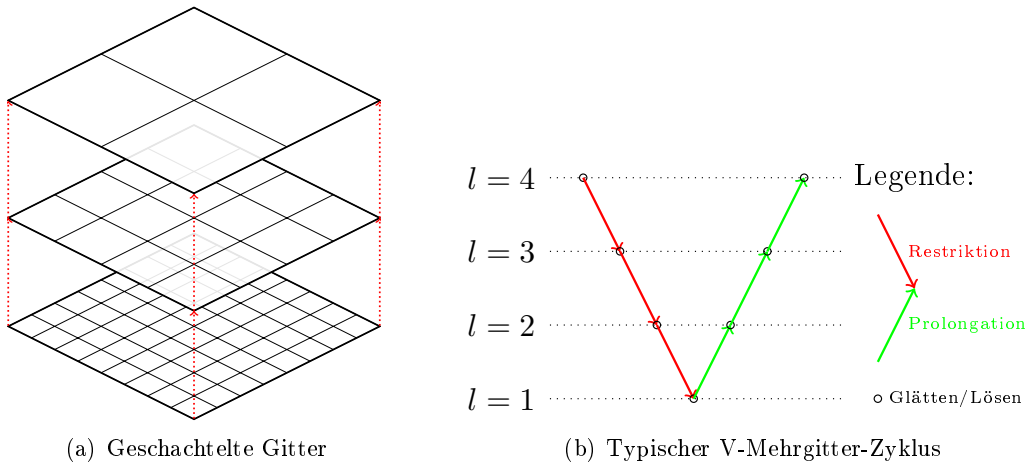
Wenn aber hochfrequente Komponenten des Fehlers bereits gedämpft wurden, kann der verbleibende Fehler auf einem größeren Gitter ähnlich gut dargestellt werden. Daher benötigen wir zunächst zwei Gitter, ein feineres  $G_h$  und ein gröberes Gitter  $G_H$ , im Verhältnis von z. B.  $h = \frac{H}{2}$ . Dann definieren wir einen Restriktionsoperator  $I_h^H : G_h \rightarrow G_H$ , sowie einen Prolongationsoperator  $I_H^h : G_H \rightarrow G_h$ , beispielsweise durch trilineare Interpolation. Damit haben wir alle Bestandteile eines Mehrgitteralgorithmus zusammen: Wir berechnen zunächst den Defekt  $d_h^m$  auf dem feinen Gitter, restringieren diesen dann auf das gröbere Gitter  $I_h^H d_h^m$  und lösen dort schließlich die sogenannte Defektgleichung

$$\mathbf{A}_H e_H^m = d_H^m, \quad (9.5.6)$$

wobei  $d_H^m = I_h^H d_h^m$  gilt. Die exakte Lösung von (9.5.6) ist dann der Fehler  $e_H^m$  auf dem gröberen Gitter, welche zurück auf das feinere Gitter als sogenannte Grobgitterkorrektur prolongiert wird,

$$u_h^{m+1} = u_h^m - I_H^h \mathbf{A}_H^{-1} I_h^H (\mathbf{A}_h u_h^m - f_h). \quad (9.5.7)$$

Gleichwohl die Defektgleichung (9.5.6) exakt gelöst werden muss, wird sie in der resultierenden Korrektur (9.5.7) doch nur anschließend auf das feinere Gitter zurückinterpoliert, welches auch eine Näherungslösung ausreichend macht. Daher können wir erneut den oben beschriebenen Zyklus von zwei Gittern benutzen: Diesmal ist das grobe Gitter das neue feine Gitter und wir lösen die Defektgleichung, indem wir zunächst hochfrequente Anteile dämpfen, diese entfernen, auf das neue grobe Gitter restringieren und dort wieder eine Grobgitterkorrektur berechnen. Dabei stellt es sich als besonders vorteilhaft heraus, wenn der Diskretisierungsparameter  $h$  als Potenzen von 2 gewählt wird. Denn dann kann diese Verschachtelung der Mehrgitter solange fortgesetzt werden, bis die echte Lösung auf dem größten Gitter effizient direkt oder gar trivial lösbar ist.



**Abbildung 9.43.:** Illustration zum Lösen des Poisson-Problems mittels eines Mehrgitteransatzes.

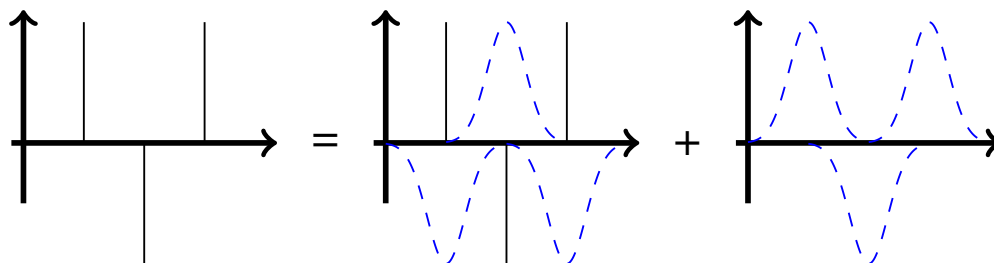
Ein Beispiel für diese geschachtelten Gitter ist Abbildung 9.43(a) dargestellt. Weiterhin findet sich in Abbildung 9.43(b) eine Darstellung des zuvor beschriebenen Zyklus über

$l = 4$  Ebenen, auch als V-Zyklus bekannt. Diese Zyklen werden solange wiederholt, bis die Norm des Fehlers  $e_H^m$  ausreichend klein geworden ist.

### Aufstellen der rechten Seite

Um die rechte Seite in (9.5.4) aufstellen zu können, müssen wir Gleichung (7.3.1) an den Gitterpunkten  $\mathbf{x} \in G_h$  auswerten.

Die Delta-Distribution in Gleichung (7.3.1) können aber nicht sinnvoll auf einem diskretisierten Gitter dargestellt werden, da es sich nicht um Funktionen handelt. Dazu benutzen wir die wohlbekannt Tatsache, dass getrennte, sphärisch symmetrische Ladungsverteilungen sich zueinander wie Punktladungen verhalten, siehe [123, S. 281]. Das bedeutet auch, dass wir die Delta-Distribution ohne Fehler durch eine andere Funktion mit diesen Eigenschaften ersetzen dürfen, solange sie nur in Wechselwirkungsintegralen mit räumlich getrennten Trägern vorkommt. Es würde sich dabei natürlich anbieten, die Basisfunktionen zur Darstellung der Ladungsverteilung der Kerne zu benutzen.



**Abbildung 9.44.:** Aufspaltung der punktierten Ladungsdichte (schwarz) mittels der Schildladungen (blau gestrichelt) in einen kurzreichweitigen Anteilen mit Singularitäten und einen langreichweitigen Anteil, der glatt ist. Reproduziert aus [77].

Diese Idee, anstelle der Punktladungen sogenannte Schildladungen zu benutzen, haben wir der Behandlung der klassischen Moleküldynamik bei der Trennung in kurz- und langreichweitige Anteile entnommen, siehe [77, Sec. 7.1.2]. Dies ist in Abbildung 9.44 für punktierte Ladungen veranschaulicht. Wir schreiben also Gleichung (7.3.1) wie folgt,

$$\rho(\mathbf{r}) = \rho(\mathbf{r}) + \sum_{\alpha,\beta=1}^n c_{\alpha\beta} \Omega_{\alpha\beta}(\mathbf{r}) - \sum_{\alpha,\beta=1}^n c_{\alpha\beta} \Omega_{\alpha\beta}(\mathbf{r}), \quad (9.5.8)$$

wobei wir die Schildladungen einfach durch eine entsprechende Entwicklung in unseren Basisfunktionen mit der Koeffizientenmatrix  $c_{\alpha\beta}$  angeben und den neuen Term einfach über einen Null-Trick in die Gleichung eingefügt haben. Die Darstellung der Schildladungen muss dabei folgende Bedingungen erfüllen [77, S. 246]:

1. normalisiert,
2. symmetrisch bezüglich des Ursprungs,

## 9. Experimente

3. kompakter Träger oder schneller Abfall,
4. glatt.

Glücklicherweise entsprechen drei dieser Forderungen jenen, die wir an die lokale Basis stellen. Wir erwarten außerdem, dass zumindest einige der Basisfunktionen symmetrisch sein werden. Damit können wir eine Darstellung der Schildladungen in der lokalen Basis erstellen, welche die Bedingungen erfüllt, die wiederum die Ladungsverteilung (9.5.8) repräsentiert.

### Extraktion des Potentials

Die berechnete Lösung auf dem Gitter entspricht schließlich dem Coulomb-Potential. Dieses wird über eine Newtoninterpolation ausreichend hoher Ordnung vom Gitter rückinterpoliert.

Wir bemerken, dass bei der Mehrgittermethode für Punktladungen, siehe [9], nur das langreichweitige Potential gewünscht ist. Aufgrund der Verschmierung der Ladungen existieren jedoch eine Reihe von Artefakten in der Lösung, um die das endgültige Potential bereinigt werden muss: die Selbst-Wechselwirkung und ein falsch berechneter kurzreichweitige Anteil, der mittels einer bekannten Stammfunktion der Schildladung abgezogen werden kann. Letzteres wäre für eine beliebige elektronische Ladungsverteilung nicht möglich.

Aufgrund der hierarchischen Summation der Energie (7.2.6a) haben wir jedoch solche Schwierigkeiten nicht. Die Selbstenergie eines Kerns  $a$ , also die Wechselwirkung von dessen Ladung mit dem eigenen Potentialanteil, steckt in exakt äquivalenter Weise auch in jedem Fragment  $\mathcal{I}_j$ , welches den Kern  $a$  enthält. Wenn wir also die Summe über die langreichweitige Wechselwirkungsenergie über jedes Fragment ermitteln und dies von derjenigen Energie abziehen, welche sich aus dem Potential der Gesamtladung (7.2.7) ergibt, so erhalten wir genau den gesuchten langreichweitigen Anteil, welcher in der Summe der Fragmente nicht erfasst wurde.

### Offene Randbedingungen

Offene Randbedingungen werden benötigt, wenn die langreichweitige Wechselwirkung im Vakuum anstelle eines periodisch fortgesetzten Mediums untersucht werden soll.

Ein erster Ansatz ist dabei eine systematische Vergrößerung des Gebiets  $\Omega$  nach außen, siehe Washio und Oosterlee [181], und auf dem größten Gebiet werden Dirichlet-Null-Randwerte gesetzt. Dieser Ansatz konvergiert jedoch recht langsam. Eine schnellere Lösung ist die Greens-Funktions-Entwicklung auf dem Rand der vergrößerten Gebiete explizit zu berechnen, siehe Bolten [22]. Dabei wird ein Finite-Volumen-Ansatz benutzt, da mit hängenden Knoten aufgrund der Gittervergrößerung umgegangen werden muss. Dieser Ansatz ist in Iseringhausen [104] implementiert, welchen wir hier benutzen. Für weitere Einzelheiten sei auf [104] bzw. auf [22] verwiesen.

### Implementationsbesonderheiten

Wir benutzen das Paket *vmg* [104] mit einem Differenzenstern nach Collatz, offenen Randbedingungen und kardinalen B-Splines als Schildladungen. Eine Verschmierung der Kernladung mittels der lokalen Basisfunktionen sollte die Genauigkeit der Methode verbessern, aber war im Rahmen dieser Arbeit nicht mehr möglich. Im V-Zyklus wird als Glätter Gauß-Seidel-Red-Black mit drei Vor- und drei Nachglättungsschritten benutzt. Der direkte Löser verwendet Givens-Rotationen. Die relative Abbruchtoleranz der Iteration lag bei  $10^{-10}$  oder maximal 15 Schritten. Die Rückinterpolation erfolgt über Newton-Interpolation.

Wir listen außerdem kurz die wichtigen Parameter auf:

- Gittermaschenweite  $h = L2^{-l}$ , bestimmt über die Anzahl der Gitterstufen  $l$  und die Kantenlänge  $L$  des Gebiets
- Halbe Breite des kompakten Trägers der B-Splines  $D$
- Grad der Newtoninterpolation  $d$

Dabei haben wir das Aufstellen der rechten Seite aus einer elektronischen Ladungsverteilung und deren Auswertung im Rahmen der Summation zusätzlich implementiert.

Eine Schwierigkeit ergibt sich durch die elektronische Ladungsverteilung. Werden die Kernladungen durch die Schildladungen verschmiert und daraus das Potential bestimmt, so erhalten wir eine falsche Wechselwirkungsenergie von verschmierten Kernladungen und elektronischer Ladungsverteilung. Da die elektrostatische Energie (7.3.5) aber symmetrisch in den beiden Ladungsverteilungen ist, können wir umgekehrt das Potential aus der elektronischen Ladungsverteilung bestimmen und es an den Kernorten mit den Punktladungen auswerten. Dadurch muss jedoch das Poisson-Problem zweimal gelöst werden: Einmal für das Elektron-Potential und ein zweites Mal für das summierte Elektron-Kern-Potential, um alle vier Wechselwirkungen zwischen Kernen (N) und Elektronen (E) zu erfassen: NN, EN, NE und EE.

#### 9.5.3. Analyse des Schätzers

Im Folgenden analysieren wir nun das Verhalten der Methode als Schätzer für die in den Fragmentrechnungen nicht erfasste langreichweitige Wechselwirkung.

#### Parameterbestimmung

Zunächst untersuchen wir das Verhalten des Löser, um den besten Parametersatz zu ermitteln.

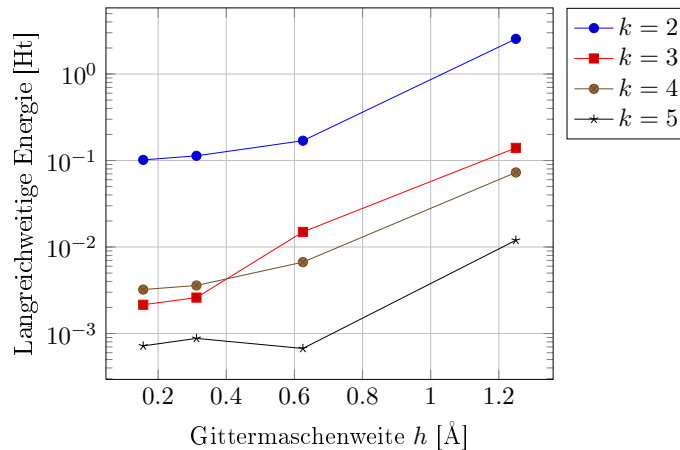
Die zu wählenden Parameter sind die Anzahl der Gitterebenen  $l$  – dies bedingt auch die Auflösung des feinsten Gitters mit  $2^l$  Punkten pro Achse – der Grad der Newton-Interpolation  $d$  und der Durchmesser des kompakten Trägers der B-Splines  $D$ .

Schließlich prüfen wir auch, inwieweit die Vernachlässigung von Kernelektronen gemeinsam mit entsprechender gleichgroßer Ladung der Kerne die Genauigkeit verbessert.

## 9. Experimente

Die Berechnungen führen wir jeweils anhand eines Hexan-Moleküls durch, da es bis zur sechsten Ordnung exakt berechnet wird. Wir ermitteln jeweils die absolute langreichweitige Coulomb-Energie, die innerhalb der berechneten Fragmente der BOSSANOVA-Methode nicht mehr erfasst wurde. Wir benutzen stets ein Gebiet der festen Größe von  $[0, L]^3$  mit  $L = 20 \text{ \AA}$ .<sup>10</sup>

**Gittermaschenweite** Wir variieren zunächst die Gittermaschenweite mit  $h = L2^{-l}$  für  $l = 4, 5, 6, 7$ , indem wir die Anzahl  $l$  der verschachtelten Gitter des Mehrgitteralgorithmus erhöhen. In Abbildung 9.45 ist die berechnete langreichweitige Coulombenergie für verschiedene Bindungsordnungen  $k$  dargestellt, wobei wir nicht die Gitterstufen, sondern die übliche Gittermaschenweite  $h$  angeben.



**Abbildung 9.45.:** Abhängigkeit der langreichweitigen Energie von der Gittermaschenweite  $h$  für verschiedene Bindungsordnungen  $k$ .

Wir erkennen Konvergenz der langreichweitigen Energie mit der Gittermaschenweite. In diesem Experiment erhalten wir ab  $l = 6$  verschachtelten Gittern ausreichende Genauigkeit.

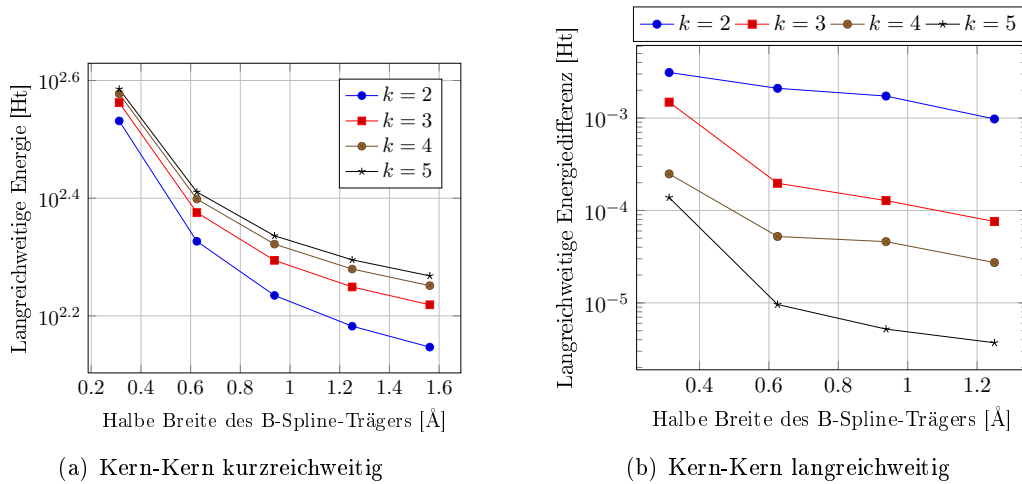
Dass die zu  $k = 3$  gehörende Kurve von  $l = 5$  zu  $l = 6$  springt, darf uns hier nicht verwundern, da die Orbitale eine Breite im Rahmen von 1 bis 2 Å haben, ist eine Maschenweite von  $h = 0,6 \text{ \AA}$  nicht ausreichend, um die Ladungsdichte sicher zu erfassen. Solche Sprünge sind also dem Sampling zu zuordnen, worauf wir später noch näher eingehen werden.

**Interpolationsgrad** Der Interpolationsgrad sollte passend zur Diskretisierungsordnung des Laplace-Operators gewählt werden. Ab einer Anzahl an Gitterebenen von  $l = 6$ , wie wir ihn zuvor auch als sinnvollen Mindestwert ermittelt haben, und ab einer damit zusammenhängenden kleinen Maschenweite spielt der Interpolationsgrad nur noch eine geringe Rolle. Somit ist  $d = 5$  ist ausreichend.

<sup>10</sup> *vmg* verwendet atomare Einheiten anstelle von Ångström für die Längen,  $L$  ist also 37,79 a. u.

**B-Spline-Breite** Wir untersuchen als letzten Parameter die Breite des kompakten Trägers der B-Splines. Diese wird als halbe Breite in *vmg* mit *near field cells* benannt und als Anzahl an Gitterpunkten  $D$  angegeben. Mittels der Kantenlänge  $L$  des Gebiets und der Anzahl an Gitterebenen  $l$  lässt sich über die Gittermaschenweite  $h = L \cdot 2^{-l}$  diese Angabe in eine Breite des Splines im Gebiet  $\Delta = hD$  in Å umrechnen.

Wir verändern die halbe Breite des B-Spline-Trägers in Schritten  $D = 2, 4, 6, 8, 10$ . Die berechneten Energien der Kern-Kern-Wechselwirkung, welche als einzige von der Splinebreite abhängen, sind in Abbildung 9.46 für den kurz- und langreichweitigen Anteil festgehalten.



**Abbildung 9.46.:** Abhängigkeit der Kern-Kern-Wechselwirkungsanteile von der Splinehalbbreite  $D$ . Um die Unterschiede deutlicher zu machen, ist der langreichweitige Anteil logarithmisch als relativer Unterschied zu  $D = 10$  dargestellt.

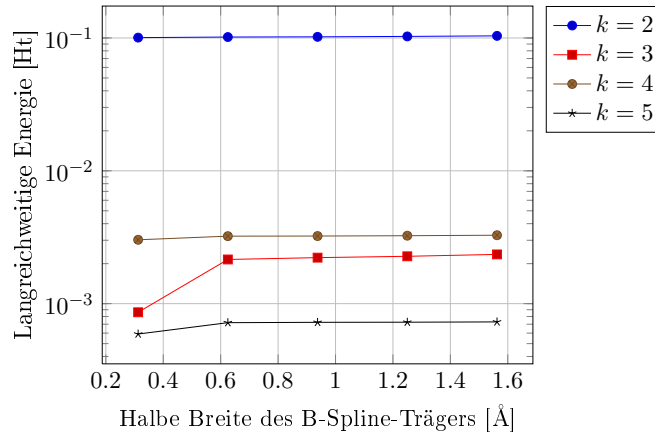
Für ausreichend feine Gittermaschenweiten muss die B-Splinenbreite nur ausreichend groß gewählt werden, damit die Kernladung glatt auf das Gitter übertragen werden. Anhand von Abbildung 9.47 sehen wir, dass die langreichweitige Coulombenergie ab einer halben Breite von 0.6 Å näherungsweise konstant bleibt, dies entspricht  $D = 4$ .

**Einfluss der Kernelektronen** Schließlich möchten wir noch den Einfluss der Kernelektronen untersuchen. In den zuvor dargestellten Zusammenhängen wurden jeweils bereits Kernelektronen vernachlässigt und gleichzeitig ein entsprechender Anteil der Ladung der zugehörigen Kerne.

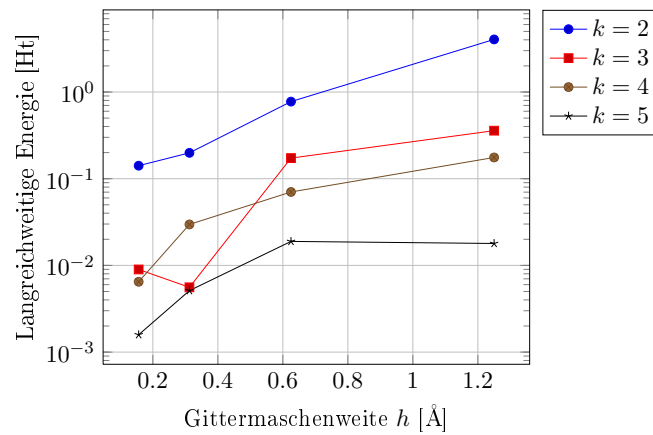
In Abbildung 9.48 ist erneut die Abhängigkeit der langreichweitigen Coulombenergie von der Gittermaschenweite dargestellt, dabei wurden Kernelektronen diesmal jedoch nicht vernachlässigt.

Wir erkennen deutlich, dass der konvergente Zusammenhang, der in Abbildung 9.45 noch sichtbar war, deutlich gestört ist, obwohl wir ähnliche Größenordnungen der berechneten Energien erhalten.

## 9. Experimente



**Abbildung 9.47.:** Abhängigkeit der langreichweitigen Coulombenergie von der Breite  $D$  des kompakten Trägers der B-Splines für verschiedene Bindungsordnungen  $k$  mit der Gittermaschenweite  $h = 2^{-7}$ .



**Abbildung 9.48.:** Abhängigkeit der langreichweitigen Energie von der Gittermaschenweite  $h$  für verschiedene Bindungsordnungen  $k$ , wobei Kernelektronen und zugehörige Kernladungsanteile *nicht* vernachlässigt wurden.

Aus diesem Grund werden wir im Weiteren nur die Valenzladungen von Kernen und Elektronen in der langreichweitigen Störungsrechnung betrachten.

### Einfluss der Gebietsgröße

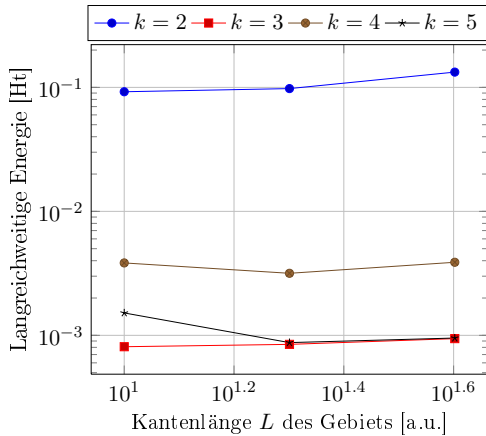
Mit dem festgelegten Parametersatz ermitteln wir nun, ob die Größe des Gebiets einen Einfluss auf die Energie hat, indem wir für ein einfaches Hexan-Molekül die Kantenlänge  $L$  des Gebiets in 2er-Potenzen anheben,  $L = 10, 20, 40$  Å, und dabei auch das Gitter  $h = 4, 5, 6$  und den Durchmesser des kompakten Trägers des B-Splines  $D = 2, 4, 8$  um einen Faktor 2 anpassen, da diese von der Gebietsgröße abhängen. Dadurch wird



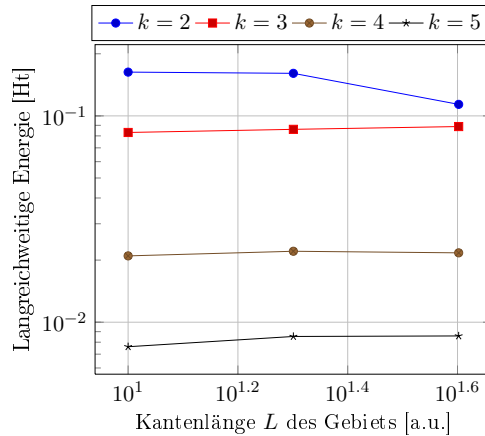
### 9.5. Abschätzung der langreichweitigen Wechselwirkung

der Effekt durch das Sampling unterdrückt und wir sehen einzig, ob eine anwachsende Gebietsgröße doch Einfluss auf die Gesamtenergie hat.

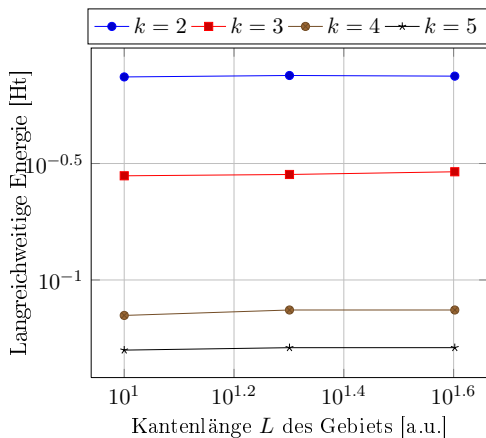
In Abbildung 9.49 ist die langreichweitige Gesamtenergie für verschiedene Bindungsordnungen für die untersuchten Kettenmoleküle jeweils mit dem Längenfaktor  $n = 3$  dargestellt. Diese Kettenlängen hatten wir auch schon in Abschnitt 9.4.4 untersucht, siehe Abbildung 9.29.



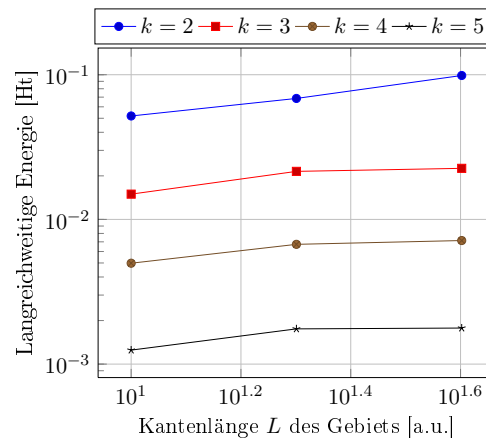
(a) Hexan



(b) Hexen



(c) Hexin



(d) BN-Kette  $n = 3$

**Abbildung 9.49.:** Abhängigkeit der langreichweitigen Energie vom in 2er-Potenzen vergrößerten Gebiet für die untersuchten Kettenmoleküle und für verschiedene Bindungsordnungen  $k$ .

Wir erkennen, dass die anwachsende Gebietsgröße im Rahmen der Genauigkeit der Methode keinen Einfluss auf die resultierende langreichweitige Gesamtenergie hat. Dies wird besonders deutlich, da von Hexan, über Hexen und Hexin die langreichweitige Gesamtenergie zunimmt und gleichzeitig auch die Schwankungen kleiner werden. Wir sehen also

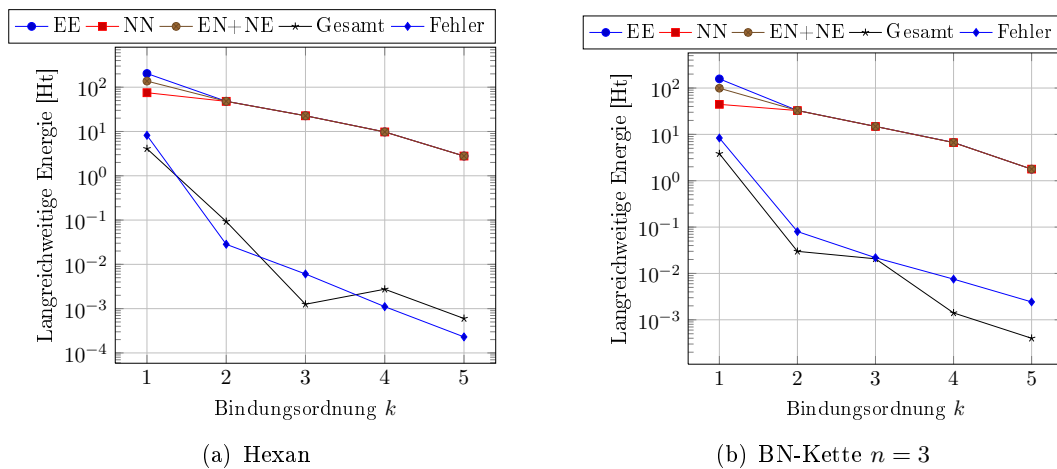
## 9. Experimente

jeweils den gleichen relativen Fehler. Dies belegt die sinnvolle Behandlung der offenen Randbedingungen.

### Kettenmoleküle

Wir wiederholen in diesem Versuch die Berechnungen aus den Abschnitten 9.4.2 und 9.4.3, berücksichtigen diesmal aber die langreichweitige Wechselwirkung unter offenen Randbedingungen zwischen den Fragmenten. Die Rechnungen werden dabei jeweils auf einem Gitter der Größe  $2^7$  Punkte pro Achse, mit einem Radius der Splines von  $D = 4$  und Interpolationsgrad  $d = 5$  berechnet.

**Alkane** Die Energie des langreichweitigen Potentials ist in Abbildung 9.50(a) für das Hexan, in die Beiträge für Elektronen, Kerne und insgesamt aufgeteilt, dargestellt.



**Abbildung 9.50.:** Approximation der langreichweitigen Energie für die Kettenmoleküle Hexan und  $B_3N_3H_{10}$  pro Bindungsordnung  $k$ .

Die Abbildung verdeutlicht, dass sich die Potentiale  $EE$  und  $NN$  mit  $EN + NE$  gegenseitig aufheben. Sie sind alle drei von gleicher Größenordnung. Für Ordnung  $k = 1$  ergeben sich Unterschiede aufgrund der Saturation der Fragmente mit Wasserstoff, was bei dieser Bindungsordnung noch in der gesampelten Ladungsdichte sichtbar wird, siehe Abschnitt A.5 im Anhang. Wir bemerken, dass die Ergebnisse für Hexan bei Bindungsordnung  $k = 6$  wieder exakt werden.

Mit einer Genauigkeit im Rahmen einer Größenordnung folgt die Abschätzung sehr gut dem Fehler aus der vollen Rechnung, gleichwohl kinetischer Anteil und Austauschanteil, siehe Abschnitt 9.4.4, ebenfalls in dieser Größenordnung liegen. Die Abschätzung behält diesen Abstand bis zu einer absoluten Genauigkeit von  $10^{-3}$  Ht, was unsere Erwartungen an die Genauigkeit mit diesem Mehrgitteransatz übertrifft.

**BN-Kettenmoleküle** Die Energie des langreichweitigen Potentials ist in Abbildung 9.50(b) für das Bor-Stickstoff-Kettenmolekül mit  $n = 3$  dargestellt.

Auch hier erhalten wir bis  $10^{-3}$  eine gute Abschätzung im Rahmen einer Größenordnung. Die Abschätzung der fehlenden langreichweitigen Wechselwirkung folgt dabei sehr gut dem Fehler in der summierten Coulomb-Energie der BOSSANOVA-Methode allein.

#### 9.5.4. Analyse als Bestimmer

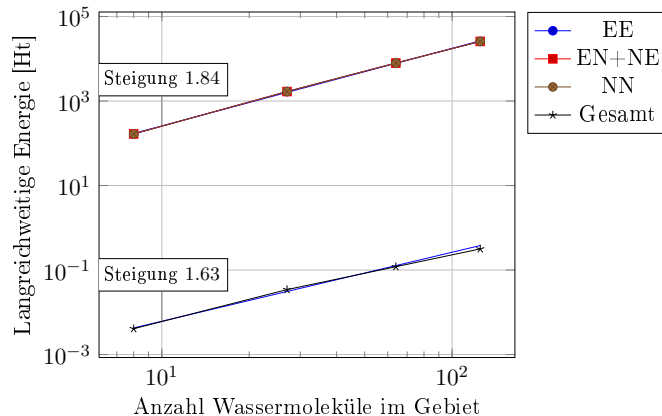
Für die langreichweitige Wechselwirkung innerhalb eines Moleküls gehen wir davon aus, dass die Lösung des Poisson-Problems nur eine Abschätzung der Größenordnung des zur vollen Lösung fehlenden Anteils liefert.

Im Folgenden untersuchen wir nun, inwieweit diese Art der Berechnung die Näherung an die Gesamtenergie der vollen HF-Rechnung insgesamt verbessert.

#### Wassergefülltes Gebiet

Zunächst betrachten wir die Energie eines Clusters von Wassermolekülen im Vakuum zunehmender Größe. Wir wollen dabei feststellen, ob die nicht erfasste langreichweitige Energie mit der Größe des Cluster gleichmäßig zunimmt.

Dazu haben wir jeweils ein Gitter mit regelmäßigen  $n = 2, 3, 4, 5$  Punkten in Abständen von  $3,1 \text{ \AA}$  mit Wassermolekülen gefüllt, wobei alle Moleküle gleich ausgerichtet sind, und anschließend die durch BOSSANOVA nicht erfasste langreichweitige Energie berechnet. Wir behalten dabei ein festes Gebiet von  $[0, L]^3$  mit  $L = 20 \text{ \AA}$ , da die Größe keinen Einfluss auf die langreichweitige Energie hat. Dadurch vermeiden wir Einflüsse durch wechselnde Gittermaschenweite und behalten einen festen Wert von  $h = L2^{-l}$  mit  $l = 7$ . Der Interpolationsgrad ist wieder  $d = 5$  und die Splinebreite  $D = 4$ .



**Abbildung 9.51.:** Approximation des nicht erfassten langreichweitigen Energieanteils für ein Wassergefülltes Gebiet.

Die berechnete langreichweitige Energie ist in Abbildung 9.51 dargestellt. Dabei nehmen die einzelnen Energiebeiträge der drei verschiedenen Coulomb-Wechselwirkungen schnell deutlich zu. Analog steigt aber auch die Summe der langreichweitigen mit fester Steigung von 1,63 in der doppelt logarithmischen Darstellung an.

## 9. Experimente

Abschließend bemerken wir also, dass sich die langreichweitige Energie sinnvoll mit anwachsender Systemgröße verhält.

### S22-Testsatz

Eine Auflistung der Molekülekomplexe des S22-Testsatzes [110] findet sich in Tabelle 9.4.

**Tabelle 9.4.:** Auflistung aller Moleküldimere und -komplexe des S22-Testsatzes.

NR.	LEGENDE	MOLEKÜL	NR.	LEGENDE	MOLEKÜL
1	●	Formamide dimer	12	⊖	Benzene water complex
2	●	Methane dimer	13	⊖	Ethene dimer
3	✱	2-pyridoxine 2-aminopyridine complex	14	⊖	Ethene ethyne complex
4	✱	Adenine thymine complex stack	15	⊖	Formic acid dimer
5	+	Adenine thymine Watson-Crick complex	16	⊖	Indole benzene complex stack
6	+	Ammonia dimer	17	⊖	Indole benzene T-shape complex
7	⊖	Benzene ammonia complex	18	⊖	Phenol dimer
8	⊖	Benzene dimer parallel displaced	19	⊖	Pyrazine dimer
9	⊖	Benzene dimer T-shaped	20	⊖	Uracil dimer h-bonded
10	⊖	Benzene HCN complex	21	⊖	Uracil dimer stack
11	⊖	Benzene - Methane complex	22	⊖	Water dimer

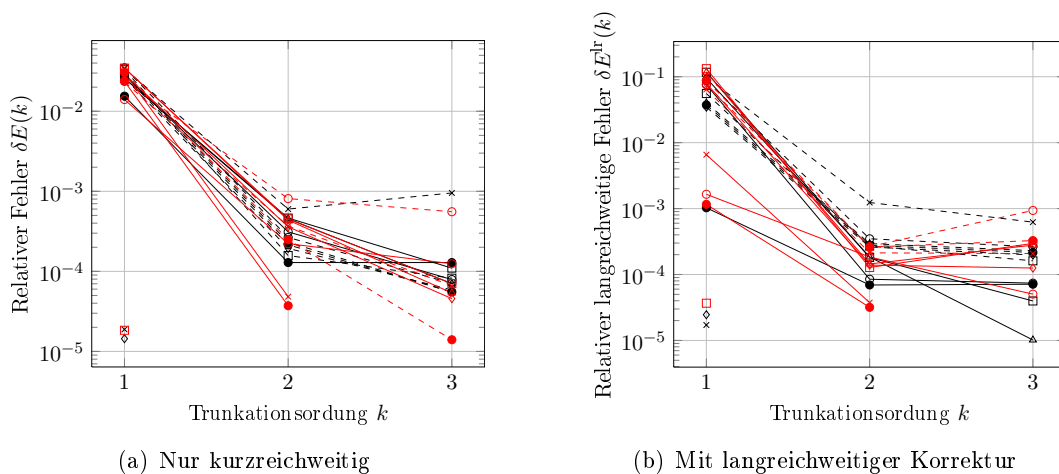
Dieser Testsatz ist eigentlich für die hochgenaue Berechnung im Rahmen von MPPT-, CI- und CC-Methoden gedacht, siehe auch Abschnitt 9.4.6. Er enthält aber insbesondere Dimere, die bei der BOSSANOVA-Methode als getrennte Monomere betrachtet werden, wenn die langreichweitige Wechselwirkung vernachlässigt wird. Deswegen stellt er auch für aktuelle Untersuchung einen guten Testsatz dar.

Wir vergleichen die summierte Coulomb-Wechselwirkung direkt, siehe Abschnitt 9.4.4. Dabei berechnen wir sie voll über eine HF-Referenzrechnung. Außerdem erhalten wir sie einmal als Summe eines kurzreichweitigen Anteils aus der BOSSANOVA-Rechnung und weiterhin mit einer zusätzlichen Abschätzung des dabei fehlenden langreichweitigen Anteils aus der Störungsrechnung über den Poisson-Löser *vmg*. Wir vergleichen diese beiden Energien wieder als relative Fehler zur Referenz (9.4.1) in Abbildung 9.52.

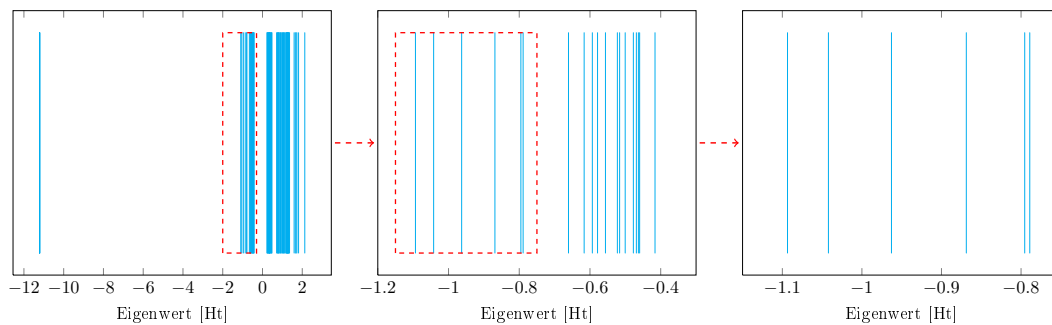
Bei Ordnung  $k = 2$  ist deutlich erkennbar, dass die Berechnung der Korrektur den Fehler um eine halbe Größenordnung verbessert, bis auf einige Systeme, die bereits sehr genau oder noch falsch erfasst wurden. Danach stagniert der Fehler, da die langreichweitige Berechnung keine höhere relative Genauigkeit als  $10^{-4}$  liefert. Insgesamt betrachten wir dieses Ergebnis für eine reine Störrechnung erster Ordnung mit den Sampling-Schwierigkeiten des Mehrgitters aber als sehr zufriedenstellend.

## 9.6. Qualitative Untersuchungen

Nach der quantitativen Untersuchung der Genauigkeit im vorherigen Kapitel wollen wir in diesem Abschnitt einige qualitative Aspekte näher besprechen.



**Abbildung 9.52.:** Summe der Anteile der langreichweitigen Coulombenergien: Einmal nur der kurzreichweitige Anteil und einmal mit zusätzlicher Korrektur aus der Poisson-Lösung als relativer Gesamtfehler relativ zur vollen Rechnung über den Testsatz über die Bindungsordnungen  $k$ .



**Abbildung 9.53.:** Eigenwertspektrum für Hexan. Die Bilderreihe fokussiert mit den rot gestrichelten Kästen schrittweise auf den Teil des Spektrums, welcher die Eigenwerte zu den  $\sigma$ -Bindungen enthält.

### 9.6.1. Eigenwertspektren

Im Folgenden interessieren wir uns für die resultierenden Eigenwertspektren der Näherung mittels BOSSANOVA. Wir wollen dabei an Überlegungen anknüpfen, die wir in Abschnitt 6.2.5 zu *Gelfand-Tsetlin-Mustern* angestellt haben.

Dabei achten wir nun auf das Verhalten der Eigenwertspektren über die wachsende Bindungsordnung  $k$ . Mit ihr wachsen auch die Hauptuntermatrizen der Fockmatrix an. Die Spektren haben wir für die Alkane mit  $n = 1, 2, 3$  in Abbildung 9.54 aufgetragen.

Zunächst werfen wir mit Abbildung 9.53 einen Blick auf das Eigenwertspektrum von Hexan insgesamt im rechten Bild.

## 9. Experimente

Dort finden sich insgesamt 82 Eigenwerte, denn so viele Basisfunktionen enthält die Basis für dieses  $C_6H_{14}$ -System. Wir bemerken in der ersten Abbildung dieser Reihe drei Gruppen von Eigenwerten: Die erste Gruppe konzentriert sich um  $-11,1$  Ht und wird von den Kernelektronen gebildet, welche sich im  $1s$ -Orbital der Kohlenstoffatome befinden. Die dritte Gruppe bilden die unbesetzten Orbitale jenseits der Energie von  $0$  Ht. Die zweite Gruppe schließlich, für die wir uns besonders interessieren, sind die Valenzorbitale, welche zwischen  $-1,2$  Ht und  $-0,3$  Ht liegen. Die erste und diese zweite Gruppe zusammen haben 50 Elektronen, also 25 doppelt besetzte Orbitale und damit 25 Eigenwerte kleiner  $0$  Ht.<sup>11</sup> Eine derartige Unterteilung findet sich bei jedem betrachteten Alkan.

Im mittleren Bild in Abbildung 9.53 haben wir auf diese zweite Gruppe fokussiert und stellen fest, dass sie in zwei Untergruppen zerfällt. Eine Untergruppe befindet sich im Intervall  $[-1,2; -0,75]$  Ht, eine zweite in  $[-0,7; -0,4]$  Ht. Die erste Untergruppe wird durch die Eigenwerte zu den  $\sigma$ -Bindungsorbitalen gebildet. Sie sind generell stärker gebunden als die  $\pi$ -Bindungen, weswegen ihre Energie niedriger ist.<sup>12</sup>

Auf diese erste Untergruppe konzentrieren wir uns schließlich im dritten Bild in Abbildung 9.53. Dort erkennen wir deutlich die sechs verschiedenen Eigenwerte.

Nachdem wir allgemein das Eigenwertspektrum von Hexan diskutiert haben, betrachten wir nun in Abbildung 9.54 das Eigenwertspektrum im Fokus der zuvor besprochenen zweiten Gruppe der Eigenwerte zu den Valenzorbitalen.

Dazu haben wir die orthogonalen Summen der Eigenwerte der verschiedenen Bindungsordnungen  $k$  übereinander in Abbildung 9.54(a) dargestellt. Ein Eigenwert entspricht dabei einem senkrechten Strich der Länge 1 und seine Farbe, sowie die Y-Ordinate kodieren die Ordnung  $k$ . Eine größere Bindungsordnung über die Anzahl der Nicht-Wasserstoffatome eines Systems ist nicht möglich, weswegen die Eigenwerte bei der Ordnung  $k = 6$  im vollen (Teil-)Spektrum von Hexan münden. Die oberste Reihe entspricht also dem mittleren Bild in Abbildung 9.53. Wir bemerken, dass hier aufgrund der Symmetrie des Systems teils mehrere Eigenwerte übereinander liegen.

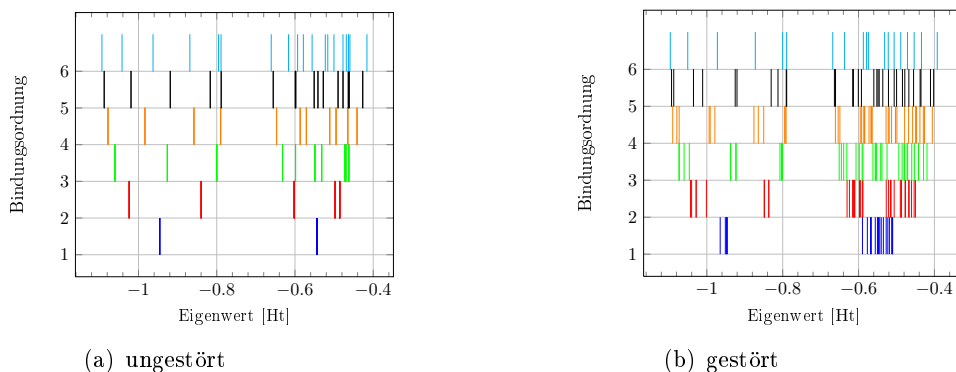
In Abbildung 9.54(b) ist der gleiche Teil des Spektrums für Hexan mit leicht gestörten Koordinaten wiedergegeben und man sieht die Eigenwerte durch die Störung aufgetrennt.

In Abbildung 9.54(a) fällt die verschachtelte Struktur der Eigenwerte, insbesondere in der linken Untergruppe, auf. Wir machen folgende Beobachtungen für diese Untergruppe:

- Die extremalen Eigenwerte  $\lambda_{\min}^{(k)}$  und  $\lambda_{\max}^{(k)}$  der Untergruppe bei einer Ordnung  $k$  begrenzen jeweils alle Eigenwerte kleinerer Ordnungen:  $\forall k' \leq k : \lambda^{(k')} \in [\lambda_{\min}^{(k)}, \lambda_{\max}^{(k)}]$ .
- Diese Struktur setzt sich auch innerhalb der Ordnungen fort: Generell liegt ein  $\lambda^{(k)}$  oft zwischen zwei Eigenwerten  $\lambda_i^{(k+1)}$ ,  $\lambda_{i+1}^{(k+1)}$  der nächsten Ordnung.
- Der Schwerpunkt der Gruppe  $\sum_i \lambda_i^{(k)}$  für die Ordnung  $k$  verändert sich kaum und zeigt dabei ein klares Konvergenzverhalten.

<sup>11</sup>Kohlenstoff hat 6 Elektronen, Wasserstoff hat ein Elektron, ergibt  $6 \cdot 6 + 14 \cdot 1 = 50$ .

<sup>12</sup>Wir möchten kurz festhalten, dass  $\sigma$ -Bindungen nicht automatisch CC-Bindungen entsprechen, auch sind nicht alle  $\pi$ -Bindungen CH-Bindungen.



**Abbildung 9.54.:** Eigenwertspektrum im Intervall  $[-2, 0]$  Ht für ein ungestörtes und gestörtes Hexan in Intervallen  $[-0, 1; 0, 1]$  Å für wachsende Bindungsordnung  $k$ .

Wir sehen hier also das gleiche Muster wie auch für die Eigenwerte von Hauptuntermatrizen, der in Abschnitt 6.2.5 betrachteten, sehr symmetrischen Matrix. Natürlich korrelieren obige Beobachtungen mit den Aussagen des Interlacing-Theorem, siehe Abschnitt 6.2.5, und des Schwerpunkttheorems, siehe Abschnitt 6.2.

Anhand von Abbildung 9.54(b) sehen wir auch, dass trotz Störung der Kernkoordinaten das grundsätzliche Intervallschachtelungsmuster erhalten bleibt und nur die einzelnen Zustände auseinander gedriftet sind. Wir konstatieren, dass das Muster also unabhängig von der perfekten symmetrischen Geometrie ist und stattdessen aus der Symmetrie der Elemente in der Konfiguration bedingt ist.

### 9.6.2. Sehr große Moleküle

Nachdem wir uns von der hinreichenden Genauigkeit und Leistungsfähigkeit der Methode überzeugt haben, wollen wir einige sehr große Moleküle qualitativ diskutieren, für die Standard-HF-Rechnungen in begrenzter Rechenzeit nicht möglich sind.

Insbesondere organische Moleküle bieten sich an, da hier vor allem kovalente Bindungen anzutreffen sind, deren Struktur mittels NMR-Spektroskopie und Röntgendiffraktion gemessen wurde, und Bindungsstrukturen bekannt sind. Wir beschränken uns auf Beispiele, die wir aus der Protein Data Bank (PDB) [19] erhalten.

Dabei bieten sich Kandidaten an, welche als „Molekül des Monats“ auf der Webseite der Datenbank erreichbar sind. Unsere Auswahl ist dabei wie folgt, wobei der vierstellige Code die Struktur in der PDB referenziert:

**1KDF** Anti-Frost-Protein

**1QM2** Menschliches Prion

**1ITF** Alpha-Interferon

**2KZ1** Interferon-Rezeptorkomplex

## 9. Experimente

Wir haben diese Strukturen aus der Datenbank entnommen, sie mittels des Programmpakets VMD [101] mit fehlenden Wasserstoffen ausgestattet und dies diente anschließend als Ausgangskonfiguration.

Unser Ziel ist vor allem aufzuzeigen, dass mit der vorgeschlagenen Methode die Berechnung großer Systeme in akzeptabler Zeit möglich ist. Da sich aus der Grundzustandsenergie viele wichtige Eigenschaften der Struktur bestimmen lassen, geben wir abschließend zu jedem Abschnitt die berechnete Näherung an die Gesamtenergie an, welche wir auf dem Niveau von MP2-Rechnungen ermittelt haben. Außerdem stellen wir die dafür benötigte kummulierte Laufzeit der parallel laufenden Löser, also die Summe der Laufzeiten für jedes einzelne Arbeitspaket, und die extrapolierte Rechenzeit einer vollen Rechnung auf gleichem Niveau einander gegenüber. Wir ermitteln diese Extrapolation dabei aus der Zeit eines kleinen Systems zusammen mit der von MP2 erwarteten Komplexität von  $\mathcal{O}(n^5)$ .

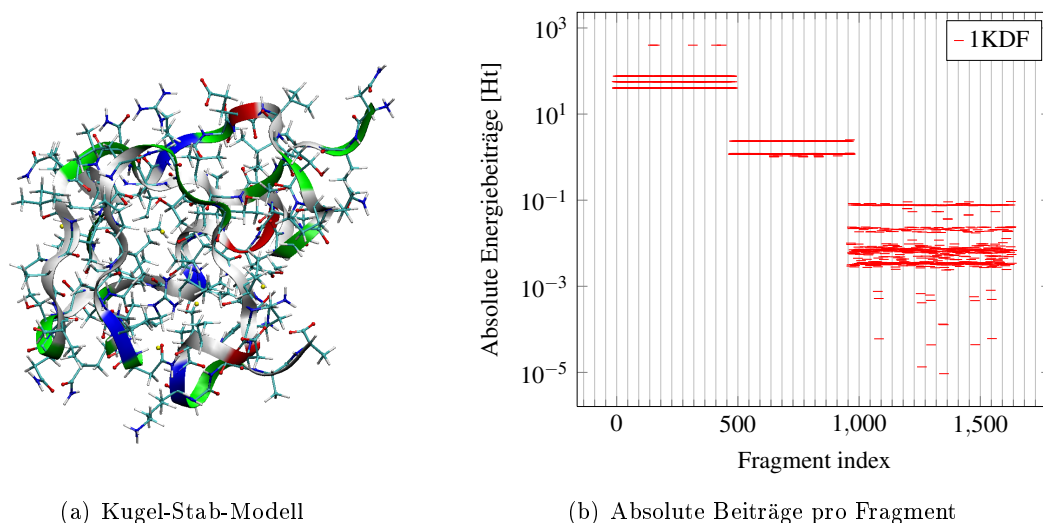
Aufgrund der Tatsache, dass es sich um organische Moleküle handelt, ergeben sich viele ähnliche Fragmente. Wir haben für jede Struktur den Absolutbetrag des Energiebeitrags pro Fragment mit einem kleinen Querstrich in den folgenden Abbildungen 9.55, 9.56, 9.57 und 9.58 dargestellt. Dort zeigen sich aus diesem Grund bestimmte Ebenen. Für das Alpha-Interferon haben wir die typischen Fragmente pro Ebene darüberhinaus durch kleine Bilder veranschaulicht. Bei den anderen drei Strukturen ergeben sich zum großen Teil die gleichen Ebenen, welche von den gleichen Fragmenten gebildet werden. Dies ist auch ein Grund für den Erfolg empirischer Potentialfunktionen im Umfeld der organischen Moleküle [129, 144].

### Anti-Frost-Protein

Pflanzen müssen sich im Winter um Schutz gegen die Kälte bemühen, denn diese kann durch Dichteänderung des Wassers in flüssiger und gefrorener Form dazu führen, dass Zellen platzen und sterben. Gewöhnlicherweise werden zum Schutz vor Frost Zucker oder Glycerol in den Zellkörper eingelagert, entweder direkt oder indem sie durch ein Protein gebunden werden, siehe [25, Abschn. 12.3.1]. Zellen in anderen Umgebungen müssen jedoch noch dramatischere Maßnahmen ergreifen, um ein Einfrieren zu verhindern. So enthalten Zellen bestimmter Dorsche, die im kalten Ozeanwasser leben, einen natürlichen Frostschutz. Wenn Wasser zu frieren beginnt, bilden sich zunächst viele kleine Eiskristalle. Das Anti-Frost-Protein verhindert, dass weitere Wassermoleküle sich an diese Kristallationsfront anlagern können, indem sie die Front krümmen und ihr weiteres Wachstum so thermodynamisch unattraktiver machen. Deswegen nennt man sie auch Eisrestrukturierungsproteine, da sie Eiskristalle in handhabbaren Größen halten. Wir haben uns hierbei an [69] gehalten.

Das Anti-Frost-Protein (PDB Schlüssel: 1KDF) enthält 992 Atome und wurde mit der BOSSANOVA-Methode mit Ordnung  $k = 3$  berechnet, siehe Abbildung 9.55(a). Die Gesamtenergie wurde dabei mit  $-24.996,98$  Ht approximiert. Diese Rechnung benötigte 2,6 Stunden an kumulierter Gesamtrechenzeit der Löser. Zum Vergleich würde die extrapolierte Laufzeit für eine HF-MP2-Rechnung 3 Jahre betragen.





**Abbildung 9.55.:** Kugel-Stab-Modell des Anti-Frost-Proteins kombiniert mit einer Bandansicht der Hauptkette, sowie die sich ergebenden absoluten Energiebeiträge pro Fragment bis Ordnung  $k = 3$ .

### Prionen

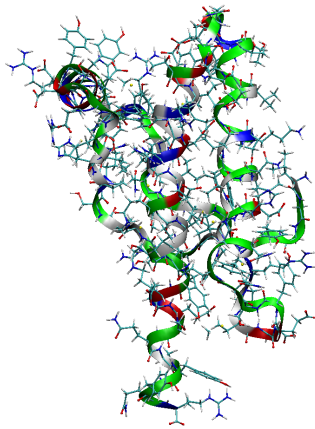
Prionen sind Proteine, die zwei Konformationen annehmen können: Eine normale und eine missgefaltete. Was sie jedoch so gefährlich macht ist, dass die missgefaltete Form andere normal gefaltete Proteine dazu anregen kann, ebenfalls in den missgefalteten Zustand zu wechseln. So kann also bereits ein einziges missgefaltetes Protein eine ganze Population normaler Proteine korrumpieren. Diese Proteine sind die Ursache des sogenannten *Rinderwahns* (Bovine Spongiforme Enzephalopathie). Wir halten uns hier an [68].

Das menschliche Prion (PDB Schlüssel: 1QM2) enthält 1694 Atome und wurde mit der BOSSANOVA-Methode mit Ordnung  $k = 3$  berechnet, siehe Abbildung 9.56(a). Die Gesamtenergie wurde dabei mit  $-45.918,01$  Ht approximiert. Diese Rechnung benötigte 4,4 Stunden an kummulierter Gesamtrechnenzeit der Löser. Zum Vergleich würde die extrapolierte Laufzeit für eine HF-MP2-Rechnung 30 Jahre betragen.

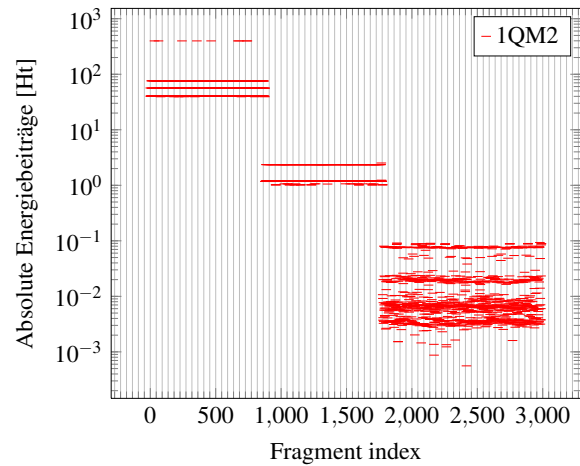
### Interferone

Zellen können sich gegen Virenangriffe auf vielfältige Weise verteidigen: Sie synthetisieren Enzyme, die die virale RNA zerschneiden, sie zeigen Teile der Viren auf ihrer Hülle, um Immunzellen zu alarmieren oder sie begehen sogar Apoptose, also Zelltod. Den Anstoß zu dieser Handlung gibt ein Übermittler. Interferone sind solche Nachrichtenüberbringer. Sie werden von infizierten Zellen abgesondert, um deren Nachbarn vor bevorstehenden Angriffen zu warnen. Sie sind dimerische oder monomerische Proteine, die sich an kleine Rezeptoren auf der Zelloberfläche binden. Das dimerische Protein ver-

## 9. Experimente



(a) Kugel-Stab-Modell



(b) Absolute Beiträge pro Fragment

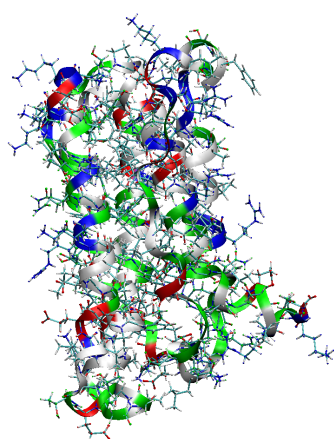
**Abbildung 9.56.:** Kugel-Stab-Modell des menschlichen Prions kombiniert mit einer Bandansicht der Hauptkette, sowie die sich ergebenden absoluten Energiebeiträge pro Fragment bis Ordnung  $k = 3$ .

bindet als Zwischenstück zwei unterschiedliche Rezeptoren, während das monomere zweifach an das gleiche Protein bindet. Wieder stammen diese Einzelheiten aus [70].

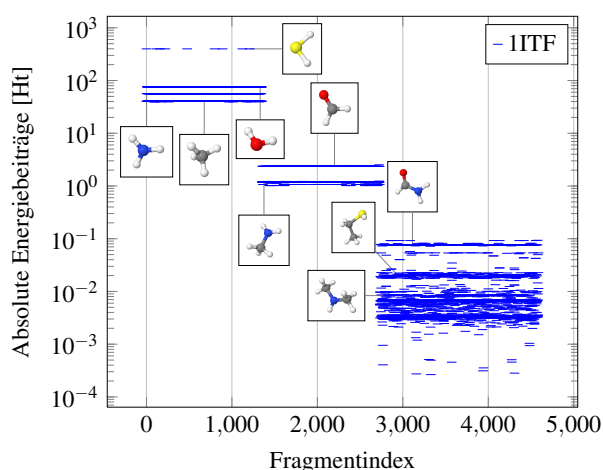
Interferone können in therapeutischen Maßnahmen zum Einsatz kommen, um virale Infektionen oder auch Krebs zu bekämpfen, da sie auch das Zellwachstum verlangsamen. Allerdings binden Interferone sehr spezifisch und nur solche von menschlichen Zellen oder Primatenzellen können effektiv zur Behandlung eingesetzt werden.

Das Alpha-Interferon (PDB Schlüssel: 1ITF) enthält 2698 Atome und wurde mit der BOSSANOVA-Methode mit Ordnung  $k = 3$  berechnet, siehe Abbildung 9.57. Die Gesamtenergie wurde dabei mit  $-68.467,41$  Ht approximiert. Die Absolutbeiträge pro Fragment sind in Abbildung 9.57(b) dargestellt. Dort zeigt ein kurzer Querstrich den Energiebeitrag je eines der sich ergebenden 4589 Fragmente an, wodurch besonders die Ebenen annähernd gleicher Beträge hervorgehoben werden. Zusätzlich haben wir exemplarisch einzelne typische Fragmente pro signifikanter Ebene abgebildet. Diese Rechnung benötigte 6,3 Stunden an kummulierter Gesamtrechnzeit der Löser. Zum Vergleich würde die extrapolierte Laufzeit für eine HF-MP2-Rechnung 186 Jahre betragen.

Außerdem betrachten wir einen 44 kDa-gewichtigen Teil des zugehörigen Rezeptorkomplexes (PDB Schlüssel: 2KZ1), welcher 5825 Atome enthält, siehe Abbildung 9.58(a), und haben diesen mit der BOSSANOVA-Methode mit Ordnung  $k = 3$  berechnet. Die Gesamtenergie wurde dabei mit  $-148.318,22$  Ht approximiert. Die Absolutbeiträge pro Fragment sind in wieder Abbildung 9.58(b) dargestellt. Diese Rechnung benötigte 62 Stunden



(a) Kugel-Stab-Modell



(b) Absolute Beiträge pro Fragment

**Abbildung 9.57.:** Kugel-Stab-Modell des Alpha-Interferons kombiniert mit einer Bandansicht der Hauptkette, sowie die sich ergebenden absoluten Energiebeiträge pro Fragment bis Ordnung  $k = 3$ . Zusätzlich sind einige sich ergebende, typische Fragmente abgebildet.

an kummulierter Gesamtrechnzeit der Löser.<sup>13</sup> Zum Vergleich würde die extrapolierte Laufzeit für eine HF-MP2-Rechnung 2600 Jahre betragen.

## Zusammenfassung

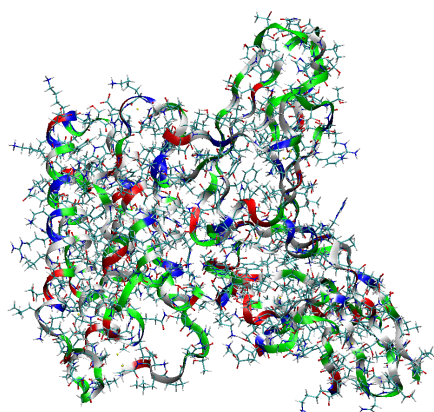
Wir untersuchten in diesem Kapitel ausführlich die Skalierungs- und Approximationseigenschaften der in Kapitel 7 vorgestellten Methode in den dort besprochenen Varianten eins und zwei. Wir legten eingangs den Fokus auf vier Fragestellungen, die wir unter folgenden Stichworten zusammenfassen: lineare Komplexität, Approximationsgüte, Robustheit und langreichweitige Wechselwirkung. Auf diese gehen wir in den folgenden Paragraphen nun kurz ein.

Wir betrachteten zunächst die Implementierung für die erste Variante und fanden über alle Programmteile hinweg klare Belege für lineare Komplexität anhand der gemessenen Laufzeiten über verschiedene molekulare Systeme. Diese Komplexität hatten wir auch aus unseren vorangegangenen theoretischen Überlegungen erwartet.

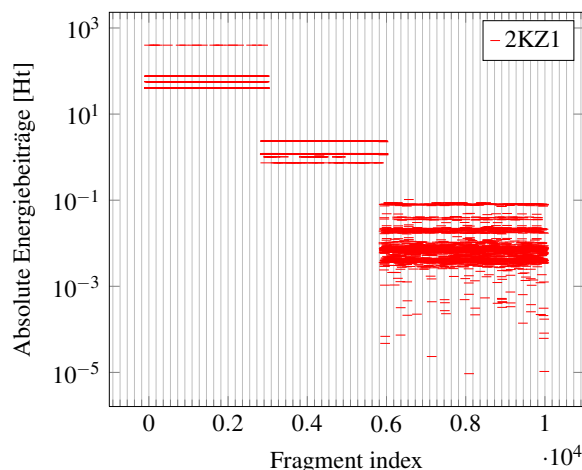
Anschließend analysierten wir die Approximationsgüte zunächst anhand verschiedener Kettenmoleküle. Der Fehler zur Grundzustandsenergie, welchen wir aus einer HF-Referenzrechnung ermittelten, nahm für diese Kettenmoleküle mit wachsender Bindungsordnung sehr rasch ab.

<sup>13</sup>Die Rechnung fand auf einem langsameren System als die vorherigen statt. Die Laufzeiten sind deswegen nicht direkt vergleichbar.

## 9. Experimente



(a) Kugel-Stab-Modell



(b) Absolute Beiträge pro Fragment

**Abbildung 9.58.:** Kugel-Stab-Modell des 44 kDa Interferon-Rezeptorkomplexes kombiniert mit einer Bandansicht der Hauptkette, sowie die sich ergebenden absoluten Energiebeiträge pro Fragment bis Ordnung  $k = 3$ .

Weiterhin behandelten wir aber auch einen Testsatz aus organischen Molekülen, deren Bindungsgraphen teils baumartige, teils durch aromatische Ringe zyklische Strukturen haben. Damit belegten wir die Robustheit der Methode. Dabei stellte sich für die Systeme mit aromatischen Ringen ein rasches Abfallverhalten des Fehlers erst ein, wenn die kleinsten Zyklen des zugehörigen Graphen vollständig mitberücksichtigt wurden. Dieses Verhalten hatten wir jedoch aufgrund der sonstigen Trunkierung der Eigenvektoren erwartet, welche die in den Ringen delokalisierten Elektronen beschreiben. Weiterhin untersuchten wir im Anschluss an die Überlegungen in Kapitel 8 auch die Näherung der Korrelationsenergie im Rahmen der Methode. Auch hier fanden wir hinsichtlich der erwarteten Genauigkeit eine gute Näherung. Dies belegte, dass kurzreichweitige Wechselwirkung auch über kovalente Bindungen hinaus erfasst werden können.

Die zweite Variante unserer Methode unterscheidet sich von der ersten durch eine nachgeschaltete Berechnung der langreichweitigen Coulomb-Wechselwirkung in erster Ordnung Störtheorie. Wir erläuterten, wie wir diese mittels eines Mehrgitter-Ansatzes für das assoziierte Poisson-Problem implementiert haben. Auch hier fanden wir im Rahmen der erwarteten Genauigkeit des Störtheorie- und des Mehrgitteransatzes eine gute Näherung, die sich sowohl als Abschätzung für den Fehler zur Grundzustandsenergie, als auch zur grundsätzlichen Berechnung der Wechselwirkung eignete. Ohne diese würden in unserem Ansatz getrennte Moleküle nicht miteinander wechselwirken, da nur kovalente Bindungen in den Bindungsgraphen eingehen.

Neben diesen vier Fragestellungen untersuchten wir außerdem heuristische Kriterien, die den Bindungsgraphen nur aus der Geometrie des molekularen Systems ermitteln. Wir sahen, dass für die Ermittlung eines Graphen mit korrekten Bindungsgraden eine quadratische Laufzeit aufgrund des Maximum Matching-Algorithmus notwendig ist. Diese

quadratische Abhängigkeit von der Anzahl der Kerne machte sich in der Gesamtkomplexität jedoch nicht bemerkbar, die sich weiterhin linear zu Anzahl der Kerne verhielt. Dem Bindungsgraphen wächst im Rahmen der Fragmentansätze und damit auch bei unserer Methode besondere Bedeutung zu. Die Qualität der Näherung bedarf seiner Korrektheit. Wir sahen anhand von Biphenyl, dass mehrere gleichberechtigte Zustände, sogenannte Mesomerie, die Berechnung mit mehreren zugehörigen Graphen nötig machen könnte, um die wahre Grundzustandsenergie wiederzugeben.

Schließlich stellten wir noch einige besondere Experimente heraus, die wir mit der vorgestellten Methode durchführten. So konnte in dem Spektrum von Alkanen ein Gelfand-Tsetlin-Muster nachgewiesen werden, wie wir es auch in Abschnitt 6.2.5 besprochen hatten. Außerdem berechneten wir einige sehr große Biomoleküle mit bis zu 6000 Atomen und konnten anhand der Beiträge pro Fragment einen starken Abfall des Fehlers zur hier aufgrund der Größe unbekanntem Grundzustandsenergie ablesen. Dies belegte einerseits die hohe Leistungsfähigkeit unserer Methode und andererseits Zweckhaftigkeit empirischer Potentialmethoden bei diesen Molekülen.



# 10. Schluss

In diesem Abschnitt fassen wir die Arbeit noch einmal zusammen. Außerdem geben wir einen Ausblick auf schon laufende und mögliche weitere Erweiterungen.

## 10.1. Zusammenfassung

In dieser Arbeit erläuterten wir einen systematischen Zugang, wie ausgehend von der vollen Schrödingergleichung eine linear skalierende Methode zur Lösung des Elektronenstrukturproblems gefunden werden kann.

Wir zeigten mit der eingeführten Korrelationsentwicklung eine Möglichkeit, welche die zulässige Lösung eines hochdimensionalen Problems gezielt einschränkt. Das Ziel dabei ist, Näherungslösungen zu erhalten, ohne dass der Operator, welcher die Problemstellung repräsentiert selbst eingeschränkt werden muss. Die Entwicklung basiert auf einer Rang-1-Bestapproximation in Unterräumen, die jeweils Funktionen über Gruppen von unterscheidbaren oder ununterscheidbaren Eingangsvariablen beinhalten. Die Projektion aus dem vollen Raum auf diese Unterräume erfolgt über einen formal definierten Dekorrektor.

Wir nutzten diese Entwicklung anschließend, um die Ansatzfunktion für die Schrödingergleichung derart einzuschränken, dass wir daraus die klassische und die Born-Oppenheimer (BO)-Moleküldynamik herleiten konnten. Für letztere verbleibt das Problem der zeitunabhängigen elektronischen Schrödingergleichung. Wir schnitten dabei die eingesetzte Korrelationsentwicklung für ununterscheidbare Eingangsvariablen nach dem ersten Term ab, welches zu der bekannten Slater-Determinante und nach kurzer Rechnung zur berühmten Hartree-Fock (HF)-Gleichung führte.

Diese HF-Gleichung muss zunächst in einer (endlichen) Basis entwickelt werden, damit sie numerischen Methoden zugänglich ist. Das resultierende Eigenwertproblem über die Fockmatrix untersuchten wir anschließend auf seine Komplexität, mit dem Ziel die ursprünglichen asymptotischen Kosten von  $\mathcal{O}(n^4)$  mit der Anzahl an Basisfunktionen  $n$  auf  $\mathcal{O}(n)$  zu reduzieren.

Eine Einschränkung der Basisfunktionen auf lokale Funktionen, solche die um ein kompaktes Gebiet betragsmäßig exponentiell abfallen, verringert die Komplexität für das Aufstellen der zu diagonalisierenden Matrix auf  $\mathcal{O}(n^2 \log(n^2))$ . Dazu definierten wir eine Menge von lokalen Funktionen, die abgeschlossen unter Produktbildung ist und die bekannten Chemikerbasen beinhaltet.

Um auch die Kosten der Diagonalisierung zu verringern, muss die Fockmatrix eine bekannte dünnbesetzte Struktur aufweisen. Diese ergibt sich aber aus der Einschränkung auf eine lokale Basis und aufgrund der exponentiellen Abfalleigenschaften der Lösungen

der HF-Gleichung. Wir nahmen an, dass uns die Struktur der Dünnbesetztheit einer beliebigen hermiteschen Matrix in Form einer hierarchischen, überlappenden Zerlegung der Menge der Basisfunktionen vorliegt. Damit konnten wir zeigen, dass sich die Matrix der Eigenvektoren zu negativen Eigenwerten in linearer Zeit aus Lösungen zu Hauptuntermatrizen rekonstruieren lässt. Das Spektrum der hermiteschen Hauptuntermatrizen ist stabil unter kleinen Störungen. Sie sind deshalb als eigenständige Teile, sogenannte Fragmente, des molekularen Systems interpretierbar. Unter der Forderung lokaler Ladungsneutralität ist außerdem die Existenz der lokalen Lösungen bekannt.

Die kanonischen Eigenfunktionen der HF-Gleichung sind aber im Allgemeinen nicht dünnbesetzt. Mittels eines modifizierten Fockoperators war es jedoch möglich, direkt lokalisierte Eigenfunktionen zu betrachten. Da diese auch aus einer nachträglichen Lokalisierung der kanonischen Eigenfunktionen erzeugt werden können, war diese Betrachtung aber rein formaler Natur.

Die letzte Hürde zu linearen Kosten fiel unter gesonderter Behandlung der langreichweitigen Coulomb-Wechselwirkung. Wir stellten dabei drei Varianten vor, diese zu behandeln. Bei der ersten wird sie nur über kurze Entfernungen behandelt. Bei der zweiten wird ihr langreichweitiger Anteil in erster Ordnung Störtheorie näherungsweise berechnet. Bei der dritten wird sie vollständig berücksichtigt. In jedem Fall wird die Wechselwirkung dabei über eines der bekannten linear skalierenden Verfahren ermittelt, siehe [9].

Die HF-Gleichung ist nicht-linear, denn der Fockoperator hängt selbst von den Lösungen der Gleichung ab. Ein gängiger Ansatz namens Self-Consistent Field (SCF) ist eine Iteration, bis sich die Lösungen nicht mehr weiter ändern. Bei den ersten beiden Varianten der Coulombbehandlung wird dabei diese nicht-lineare Kopplung der Lösungen aufgebrochen, so dass diese Abhängigkeit der Matrix von den Eigenvektoren nur noch innerhalb der Fragmente besteht.

Mittels dieser Bausteine ist dann zu verstehen, warum Fragmentansätze, deren mehr als 60-jährige Literatur wir im Anhang in Ausschnitten überblicken, erfolgreich sind. Mit chemischer Intuition wird die dünnbesetzte Struktur der Fockmatrix und ihrer Eigenvektoren geraten und aufgrund von Abschirmungseffekten kann die langreichweitige Wechselwirkung in guter Näherung vernachlässigt werden. Die Fragmente bilden dann eigenständige HF-Gleichungen, die völlig unabhängig von den anderen Fragmenten gelöst werden können. Im Rahmen unserer obigen Ausführungen verstehen wir, wie Eigenschaften eines molekularen Systems aus Teilen des Systems berechnet werden können. Dabei ist auch zu verstehen, warum der Ansatz mißlingen kann, wenn Elektronen zu stark delokalisiert sind, weil dann zumindest in einer lokalen Basis keine dünnbesetzte Struktur der Eigenfunktionen vorliegt.

Bei diesen Fragmentansätzen ist die Saturation mit Wasserstoff der bei der Fragmentierung durchtrennten Bindungen gängig, siehe Christoffersen und Maggiora [34]. Dies dient vor allem der Erleichterung der Berechnung der Fragmente, da sie dann über Gesamtspin null verfügen und mit der Closed-Shell Hartree-Fock (CLHF)-Methode berechnet werden können. Dabei wird die Einteilchen-Wellenfunktion nur mittels einer einzelnen



Slater-Determinante beschrieben, wie sie auch aus unserer abgeschnittenen Korrelationsentwicklung resultierte.

Wir schlossen am Ende des Kapitels den Kreis zur am Anfang der Arbeit besprochenen Korrelationsentwicklung, indem wir den darin enthaltenen Dekorrelator als Gegenstück zum sogenannten Anregungsoperator verstanden. Letzterer bricht die starke Einschränkung der Lösung auf eine einzelne Slater-Determinante und damit die Produktgestalt auf, indem er weitere Determinanten hinzufügt. Diese beinhalten dann mehr und mehr der herausprojizierten Korrelation. Wir betrachteten dazu die Møller-Plesset Störtheorie (MPPT) als Möglichkeit im Rahmen des Fragmentansatzes, eine genauere Lösung der elektronischen Schrödingergleichung auf Basis der Lösung der HF-Gleichung zu erhalten, welche die Korrelationsenergie in Teilen berechnet. Diese dient zur Bestimmung notwendiger Anregungsamplituden, den Entwicklungskoeffizienten dieser weiteren Slater-Determinanten.

Die vorgestellte Methode ist in den ersten beiden Varianten vollständig implementiert, dabei konnten wir aber auf MPQC [108] als CLHF-Löser zurückgreifen. Für die langreichweitige Wechselwirkung erweiterten wir die in [104] implementierte Mehrgittermethode, um die elektronische Ladungsverteilung auf das Gitter zu bringen. Wir belegten sowohl die lineare Komplexität der Implementierung, als auch die sehr guten Approximations- und Parallelisierungseigenschaften in vielen numerischen Beispielen. Dabei definierten wir eine Näherung als genau, wenn der relative Fehler zur Grundzustandsenergie kleiner als derjenige der verwendeten endlichen Basis in Relation zur unendlichen Basis wird. Dabei überprüften wir auch, dass mittels der Störrechnung die langreichweitige Wechselwirkung sinnvoll erfasst wird. Weiterhin konnten wir unter der Annahme, dass die dynamische Korrelation der Elektronen ein lokales Phänomen mit begrenzter Reichweite ist, auch an die Korrelationsenergie eine gute Näherung bestimmen.

Wir möchten abschließend noch die Vorteile unserer Methode herausstellen.

Im Unterschied zu den vorhandenen Fragmentansätzen wird kein chemisches Wissen seitens des Nutzers benötigt, da die Unterteilung anhand eines automatisch bestimmten Bindungsgraphen in alle möglichen verbundenen Teilgraphen erfolgt. Diese Teilgraphen überlappen sich dabei deutlich, ihre Größe wird durch eine vorgegebene Bindungsordnung beschränkt. Diese gibt die maximale Anzahl an Kernen pro Teilgraph vor.

Besonders hervorzuheben gegenüber anderen in der Literatur vorhandenen linear skalierenden Methoden ist die um eine Größenordnung geringere Skalierungskonstante. Wir fanden den Crossover-Punkt, ab dem unsere Methode schneller als Standard-HF ist, bei 20 Kernen. Gleichwohl die Methode approximativ arbeitet, erreicht oder übertrifft die Güte der damit berechneten Näherung die mögliche relative Genauigkeit aufgrund der endlichen Entwicklungsbasen der Referenz-HF-Methode.

Weiterhin ist besonders die sehr gute Parallelisierbarkeit interessant. Wir haben belegt, dass es mit einem einfachen Client/Server-Ansatz leicht möglich ist, eine sehr gute parallele Effizienz bei bis zu 100 Prozessen zu erreichen und vermutlich sogar deutlich darüber hinaus.

Schließlich möchten wir auf die Erweiterbarkeit des Ansatzes hinweisen. Leicht können periodischen Randbedingungen dank des flexiblen langreichweitigen Löser im Rahmen von HF-Rechnungen behandelt werden. Wir hatten gezeigt, dass die Berechnung der Korrelationsenergie mittels MP2 möglich ist. Dabei skalieren die Kosten in beiden Fällen linear mit der Anzahl der Kerne. Außerdem kann dank der Fragmente leicht ein Übergang zu den Simulationsmethoden der klassischen Moleküldynamik gefunden. Wir werden gleich noch darauf zu sprechen kommen.

## 10.2. Ausblick

Wir schließen im Ausblick auf gegenwärtige und mögliche zukünftige Entwicklungen.

### 10.2.1. Empirische Potentiale

Betrachten wir das Alpha-Interferon, welches wir in Abschnitt 9.6.2 gerechnet haben, dann ist mit der Darstellung typischer Fragmente in Abbildung 9.57(b) klar, dass sehr viele ähnliche Fragmente bei solch großen Rechnungen auftreten. Aufgrund der Systematik der Unterteilung des Bindungsgraphen bietet es sich an, die zahlreichen Lösungen der Teilprobleme gewissermaßen „abzuspeichern“. Dazu bedarf es einer geschickt gewählten Funktion mit Eingangsvariablen, die gerade die Freiheitsgrade des Fragments repräsentieren. Offensichtlich ist, dass Fragmente zu höheren Bindungsordnungen mehr Eingangsvariablen benötigen. Solche Funktionen sind in der Moleküldynamik als empirische Potentiale bekannt und es existieren eine Schar von Funktionen, die bspw. Schwingungsverhalten bielementiger Systeme sinnvoll repräsentieren, aber auch die vielfältigen Bindungsstrukturen der Hybridisierung von Kohlenstoff abbilden können. Mittels einer solchen Funktion wird dann der Beitrag eines Fragments zur Gesamtenergie dargestellt. Die Parameter der Funktion können dabei aus den Lösungen zu einer bekannten Geometrie des Fragments durch die Methode der kleinsten Quadrate gewählt werden.

### 10.2.2. Partialladungen

Im Rahmen der Berechnung langreichweitiger Wechselwirkung mittels einer Mehrgittermethode wird das elektrostatische Gesamtpotential des molekularen Systems auf einem Gitter  $G_h$  bestimmt. Das Gitter bestehe dabei aus  $n^3$  Punkten mit  $n = \frac{1}{h}$ , siehe (9.5.3),

$$G_h = \{\mathbf{x}_i | \mathbf{x} = jh \text{ for } j \in \{1, \dots, n\}\}_{i=1}^{n^3},$$

wobei  $h$  die Gittermaschenweite ist und wir von einem Gebiet  $[0, 1]^3$  ausgehen.

Damit lässt sich folgendes Problem formulieren: Wir möchten  $M$  Partialladungen an den Kernorten des Systems aufgehängt dergestalt bestimmen, so dass die Überlagerung der zugehörigen sphärischen Potentiale möglichst gut mit dem berechneten Gesamtpotential übereinstimmt. Dieses Problem lässt sich als überbestimmtes Gleichungssystem lösen,

$$\mathbf{Ax} = \mathbf{b}, \tag{10.2.1}$$

wobei  $\mathbf{x} \in \mathbb{R}^M$  der Vektor der Partialladungen,  $\mathbf{b} \in \mathbb{R}^{n^3}$  der Vektor der Potentialwerte pro Gitterpunkt und sich die Matrix des Problems  $\mathbf{A} \in \mathbb{R}^{n^3 \times M}$  wie folgt bestimmt,

$$\mathbf{A}_{ij} = \frac{1}{4\pi\epsilon_0} \frac{1}{\|\mathbf{x}_i - \mathbf{R}_j\|}. \quad (10.2.2)$$

Diese trägt also in den Spalteneinträgen jeweils das auf den Gitterpunkten  $\mathbf{x}_i \in G_h$  diskretisierte Potential der Punktladung des Kerns  $j$  am Ort  $\mathbf{R}_j$ .

Das System ist überbestimmt, da es im Allgemeinen wesentlich mehr Gitterpunkte als Kerne geben wird. Die Lösung kann mittels einer Singulärwertzerlegung bestimmt werden, so dass sie der Lösung der kleinsten Quadrate entspricht.

Partialladungen sind ebenfalls bei Moleküldynamiksimulationen wichtig. Kurzreichweitige Wechselwirkungen, die mittels der empirischen Potentiale bestimmt werden, enthalten einen Teil der Coulombwechselwirkung. Da bei diesen Simulationen die Elektronen im Allgemeinen nicht mehr explizit berücksichtigt werden, müssen solche effektive Ladungen für die Kerne ermittelt werden, welche die verbleibende langreichweitige Wechselwirkung widerspiegeln.

### 10.2.3. Mögliche weitere Entwicklungen

Es sind viele weitere Arbeiten auf Basis dieser Arbeit möglich. So ist die offensichtliche Beziehung zu den Bond Order-Potentiale, siehe Finnis [56], nicht geklärt. Weiterhin wurde in dieser Arbeit die Zerlegung nur anhand der Kerne und nicht anhand der Elektronen gesucht. Wenn rigoros etwas über den chemischen Charakter der Bindungen im Graphen bekannt wäre, so ließen sich diese automatisch in Gruppenfunktionen für  $\pi$ -Elektronen und andere nach McWeeny [135] einteilen. Auch konnten wir nicht auf die Verwicklungen eingehen, wenn der Ansatz einer einzelnen Slater-Determinante aufgegeben wird und damit auch Rechnungen mit Gesamtspin ungleich null möglich werden. Nach der Korrelationsentwicklung muss eine Linearkombination von Slater-Determinanten, wie im MCSCF-Ansatz, bereits über die reine Antisymmetrie hinausgehende, sogenannten statische [91, Abschn. 5.2.10] Korrelation beinhalten. Dies würde es eventuell ermöglichen, die aktuell notwendige Saturation mit Wasserstoff wegzulassen. Darüberhinaus sind Erweiterungen des Bindungsgraphen auf ionische Bindungen möglich oder solche, die die Berechnung äquivalenter Zustände im Rahmen von Mesomerie erlauben. Außerdem ist es mit unserer Methode im Rahmen der Energiesumme 7.2.6a möglich, die Bindungsenergie zumindest in Relation zu zwei Wasserstoffbindungen zu definieren, vgl. Abschnitt A.5. Schließlich erlaubt der Ansatz, HF-Rechnungen mit langreichweitiger Wechselwirkung [104] unter periodischen Randbedingungen aufzusetzen. Die periodische Zelle muss nur so groß sein (oder entsprechend vervielfältigt werden), dass sie Fragmente zur höchsten Bindungsordnung vollständig beinhaltet. Dies betrifft auch die Erweiterung zur Berechnung der Korrelationsenergie, analog zum Ansatz von Paulus [142].

### 10.2.4. Veröffentlichung der Implementierung

Abschließend möchten wir auf die Zugänglichkeit des Programmcodes innerhalb der GNU Public License unter der Adresse <http://www.moleculder.de/> hinweisen.



# A. Appendix

## A.1. Bekannte Basissätze

Über die Jahrzehnte wurden viele Basissätze entwickelt, die mittlerweile Standard sind. Wir folgen hier der Auflistung in [126, Abschn. 15.4].

### A.1.1. Slaterbasen

Zunächst wurden atomare Orbitale als Basisfunktionen verwendet, die Lösungen der elektronischen Schrödingergleichung für das Wasserstoffatom.

$$\xi_{\nu}^{STO}(\mathbf{r}_{\nu}) = N r_{\nu}^{n-1} \exp(-\zeta r_{\nu}) Y_l^m(\theta_{\nu}, \phi_{\nu}), \quad (\text{A.1.1})$$

wobei  $N$  die Normierungskonstante, der Orbitalexponent „Zeta“  $\zeta > 0$  und  $\mathbf{r}_{\nu} = \mathbf{r} - \mathbf{R}_{\nu}$  in sphärischen Koordinaten  $(r, \theta, \phi)$  relativ zum Kernort  $\mathbf{R}_{\nu}$  dargestellt sei. Barnett und Coulson [15] betrachten ausführlich wie die Matrixelemente dann berechnet werden können. Leider lassen sich mit diesen sogenannten STOs keine kompakteren, analytischen Ausdrücke für die Integrale über die Coulomboperatoren und Austauschoperatoren aufstellen, welche bis zu vier verschiedene Basisfunktionen einschließen. Die Berechnung der Matrixelemente in dieser Basis bleibt dadurch sehr aufwändig, siehe [126, S. 487].

### A.1.2. Gaussbasen

Boys [23] schlug alternativ vor, Gauss-Funktionen zu benutzen. Im Folgenden sei die *kartesische* Form dieser Basisfunktionen angegeben:

$$\xi_{\nu}^{GTO}(\mathbf{r}_{\nu}) = N \mathbf{r}_{\nu,1}^{l_1} \mathbf{r}_{\nu,2}^{l_2} \mathbf{r}_{\nu,3}^{l_3} \exp(-\alpha r_{\nu}^2), \quad (\text{A.1.2})$$

wobei wieder  $N$  die Normierungskonstante,  $\alpha$  der Orbitalexponent und  $\mathbf{r}_{\nu,i}^n$  die  $i$ te Komponente des Ortsvektors  $\mathbf{r}_{\nu} = \mathbf{r} - \mathbf{R}_{\nu}$  relativ zum Kernort  $\mathbf{R}_{\nu}$  mit Exponent  $l_i$  sei.

GTOs werden anhand der Summe  $n := \sum_{i=1}^3 l_i$  der Exponenten des Polynoms über die Komponenten des Ortsvektors klassifiziert, siehe (A.1.2). Für  $n = 0$  spricht man von einem s-Orbital, bei  $n = 1$  von p-Orbital,  $n = 2$  von d-Orbital usw. . Die Zahl  $n$  ist dann analog zum Exponenten  $n$  in (A.1.1).

Weiterhin müssen aufgrund der schlechten Wiedergabe der Einteilchen-Wellenfunktion am Kernort durch ein einzelnes GTO bereits für ein s-Orbital Linearkombinationen von GTOs verwendet werden. Da Boys [23] die GTOs vorschlug, um die Berechnung der Vier-Zentren-Integrale gegenüber den STOs zu vereinfachen, werden sogenannte CGTOs, also

## A. Appendix

Linearkombinationen von GTOs  $\xi_u^{GTO}$ , benutzt, die alle die gleichen Exponenten  $l_i$  und den gleichen Aufpunkt  $\mathbf{R}_\nu$  haben:

$$\xi_\nu = \sum_{u=1}^N d_{u\nu} \xi_u^{GTO} \quad (\text{A.1.3})$$

Es wird also einfach ein STO in einer begrenzten Menge bestimmter GTOs dargestellt. Die einfachen GTOs werden dabei auch *primitive* GTOs genannt. Die Koeffizienten  $d_{u\nu}$  werden pro  $\nu$  durch Fit an das entsprechende STO gewonnen. In der Entwicklung der Einteilchen-Wellenfunktion wird anschließend nur der Koeffizient für  $\xi_\nu$  variiert. Die Koeffizienten  $d_{u\nu}$  bleiben fest.

Es gibt verschiedene Arten, diese CGTOs zu bilden. Sie lassen sich zunächst grob anhand der Anzahl  $N$  der verwendeten primitiven GTOs einteilen.

### A.1.3. Klassifizierungen

**Slaterförmige Orbitale (STOs)** Wir verwenden hierbei die gebräuchlichen englischen Begriffe.

**Minimalbasis** Die Minimalbasis ergibt sich, wenn für jedes Elektron eines Atoms seiner inneren und äußeren Schale jeweils ein STO benutzt wird.

**DZ, TZ** Für diese Basis wird jedes STO der Minimalbasis durch zwei bzw. drei STOs mit leicht unterschiedlichem Orbitalexponenten  $\zeta$  ersetzt.

**SV** Hierbei wird für die Elektronen der inneren Schalen nur die Minimalbasis eingesetzt, während für die Elektronen der äußeren Schale zwei oder mehr STOs, also DZ, ..., angesetzt wird.

**P** Die Einteilchen-Wellenfunktion in Molekülen können durch Polarisierung verzerrt oder verschoben werden. Deswegen fügt man weitere STOs mit einer Quantenzahl  $l$  größer als die größtmögliche des Grundzustands hinzu.

Mittels der Abkürzungen wird dann eine Basis vollständig beschrieben. So bezeichnet VTZ eine SV mit TZ Basis. DZP wiederum ist eine DZ Basis mit Polarisierung. Für näherer Erläuterungen verweisen wir auf [126, S. 488f].

**Gaussförmige Orbitale (GTOs)** Die obige Klassifizierung der Standardbasen erfolgte also für STOs. Sie können, wie zuvor ausgeführt, jedoch einfach durch CGTOs ersetzt werden. Aufgrund der notwendigen Darstellung eines STO durch mehrere GTOs ergeben sich dabei weitere typische Namensgebungen, die wir hier noch kurz erläutern wollen. Wir unterteilen dabei nach den Gruppen, die sie entwickelt haben.

Die Basen von Pople und Mitarbeitern finden sich im bekannten Programm GAUSSIAN.

**STO-3G** Minimalbasis, wobei je ein STO durch je drei primitive GTOs dargestellt wird.

**3-21G,6-31G** SV-Basis, wobei das STO der inneren Schale durch 3 bzw. 6 primitive GTOs und die äußere Schale durch zwei CGTOs dargestellt wird, einmal aus 2 bzw. 3 primitiven GTOs und einmal aus einem einzigen (diffusen) GTO

**3-21G(\*),6-31G(\*)** P-Erweiterung der 3-21G bzw. 6-31G durch polarisierte GTOs.

**3-21++G,6-31++G(\*)** Erweiterung der 3-21G(\*) bzw. 6-31G(\*) durch weitere hoch diffuse GTOs mit sehr kleinem Orbitalexponenten. Dies dient der verbesserten Darstellung von Anionen und Wasserstoffgebundenen Dimern.

Zur Verwendung in Berechnungen, die die Korrelation zwischen den Elektronen einschließen soll, haben **Dunning** und Mitarbeiter folgende „correlation-consistent“ Basen entwickelt, siehe u. a. [48]:

**cc-VXZ** Das X bezeichnet dabei den die Anzahl von Zeta, bspw. cc-VDZ für DZ.

**cc-pVXZ** P-Erweiterung

**cc-pCVXZ** Zusätzliche primitive GTOs für Korrelationseffekte in der inneren Schale.

**aug-cc-pCVXZ** Zusätzliche diffuse polarisierte und nicht-polarisierte GTOs.

## A.2. Bindungsarten

Den berühmten „Bindungsstrich“, der von A. S. Couper eingeführt wurde, müssen wir wohl nicht erwähnen. Neben diesem unscheinbaren Strich in der chemischen Notation, der kovalente Bindungen bezeichnet, gibt es noch eine Reihe weiterer Bindungsarten, die wir im Folgenden näher beschreiben werden. Wieder halten wir uns an die Darstellung von H. R. Christen [32].

### A.2.1. Kovalente Bindung

Die MO-Methode eignet sich am besten zum Verständnis kovalenter Bindungen. Man spricht von solch einer Bindung, wenn sich ein doppelt besetztes Orbital zwischen zwei Atomrümpfen befindet. Ein Maß für die Stärke der Bindung ist der Überlapp der beiden an dem Orbital beteiligten Wellenfunktionen:  $\int \psi_1 \psi_2 d\nu$ .

Eine Bindung zwischen Atomrümpfen unterschiedlicher Elementen führt im Allgemeinen zu einer *polaren Bindung*. Hier halten sich die Elektronen in dem Orbital im zeitlichen Mittel mehr beim elektronegativeren Element auf. Hierbei ist die *Elektronegativität* aber ein qualitatives Konzept zur Einschätzung polarer Bindungen.

Ein Molekülorbital lässt sich daran charakterisieren, welche atomaren s-, p-, d-Orbitale genau zu diesem miteinander verknüpft werden mussten: So spricht von (rotationssymmetrischen)  *$\sigma$ -Bindungen*, wohin gegen  *$\pi$ -Bindungen* eine Knotenebene besitzen. Durch die gleichzeitige Linearkombination der atomaren Orbitale gelangt man zum Konzept der *Hybridisierung*, bei dem bspw. aus einem s- und 3 p-Orbitalen vier  *$sp^3$* -Orbitale entstehen und so die geometrische Form und die Tetraederwinkel, wie sie sich in  $\text{CH}_4$  finden, erklären. Andere Beispiele sind  $\text{BeCl}_2$  oder natürlich  $\text{H}_2\text{O}$ .

## A. Appendix

Ein weiteres wichtiges Konzept ist die *Delokalisierung*. Hierbei bilden mehrere Atomare Orbitale ein gemeinsames, ausgedehntes Molekülorbital. Man spricht von einer *polyzentrischen Bindung*, weil die atomaren Orbitale an verschiedenen Zentren aufgehängt sind. Aufgrund der Ausgedehtheit besitzen die Elektronen eine geringere kinetische Energie und das Gesamtsystem ist somit wesentlich stabiler. Prominentes Beispiel ist natürlich Benzol, bei dem sich je ein Orbital über und unter der Ringstruktur erstreckt. Der Stabilitätsgewinn mißt sich dabei an der *Delokalisierungsenergie*.

Allgemein gibt es folgende Charakteristika für kovalente Bindungen:

- Energie<sup>1</sup>
- Länge
- Polarität
- Bindungsordnung

Die *Kovalenzbindigkeit* eines Elements, also die allgemein maximal mögliche Anzahl an gleichzeitigen Bindungen, bestimmt sich durch die Anzahl einfach besetzter atomarer Orbitale.

### A.2.2. Ionische Bindung

Die ionische Bindung kann man sich als Extremfall einer polaren Bindung vorstellen, hierbei hat das Molekülorbital praktisch die Form eines der beteiligten atomaren Orbitale. Entscheidender Unterschied zur kovalenten Bindung ist, dass nicht der Überlapp die Stärke der Bindung ausmacht, sondern der Gewinn an Coulombenergie durch Nähe entgegen gesetzter Ladungen.

Eine grundlegende Besonderheit der ionischen Bindung wird erst klar, wenn man sich den Vorgang in Kristallen klar macht. Sie ist dort ein kollektives Phänomen, da die Ionisationsenergie der einzelnen Teilchen erst durch die gemeinsam gewonnene Gitterenergie wieder wettgemacht wird. Dies ist im Bildungsmodell von F. Haber und M. Born beschrieben. Sie geben einen Zyklus aus Sublimation – der Lösung vorhandener Bindungen –, Ionisation bzw. Affinisierung – der Abgabe und Aufnahme von Elektronen – und schließlich der Gitterbildung an, der es indirekt ermöglicht, die Gitterenergie näherungsweise zu messen.

Die Geometrie des Gitters hat entscheidenden Einfluss auf die Größe der Coulombenergie  $U$ , welche zusammen mit der Abstossungsenergie die Gitterenergie bildet. Die Gittergeometrie drückt sich in der Madelung-Konstante aus:  $U = -M \frac{e^2}{r} (Z_+ Z_-)$ , wobei  $r$  der Abstand der beiden Ladungen  $Z_+$  und  $Z_-$  im Gitter ist.

Die *maximale Ionisierung* ist zunächst beschränkt durch die Anzahl der Elektronen in der äußersten Schale. Elektronen aus tieferen Niveaus kosten überproportional mehr Ionisationsenergie, welches durch die hinzu gewonnene Gitterenergie nicht mehr kompensiert wird.

---

<sup>1</sup>Aus der Bindungsenergie leitete bspw. L. C. Pauling die Elektronegativität ab, indem er das (später) geometrische Mittel der Energien der homogenen Bindungen in Relation zur heterogenen Bindungsenergie betrachtete.



Ionen bilden sich nicht nur aus einzelnen Atomen sondern auch aus Molekülen, für welche S. M. Jørgensen und A. Werner den Begriff *Komplex-Ionen* prägten. Beispiele sind  $\text{CaSO}_4$ , welches aus  $\text{Ca}^{2+}$  und  $\text{SO}_4^{2-}$  besteht, oder  $\text{CaCO}_3$ , welches aus  $\text{Ca}^{2+}$  und  $\text{CO}_3^{2-}$  besteht.

Haben diese Komplex-Ionen mehrere, räumlich getrennte ionische Bindungen so spricht man von *Chelaten* oder “mehrzähniem” Komplex. Ein Beispiel hier ist Ethylendiamin.

### A.2.3. Metallbindung

Metallische Bindungen lassen sich durch ein Modell aus Elektronengas nach P. Drude und H. A. Lorentz und einem Gitter aus positiven Ionen beschreiben. Hierbei ist die Geometrie des Gitters meist eine dichteste Kugelpackung. Das Elektronengas entsteht durch Valenzelektronen im so genannten Leitungsband, welche nicht an ein Atom gebunden sind, sondern sich frei durch das Gitter bewegen. Die *metallische Bindung* ist also die elektrostatische Anziehung zwischen delokalisierten Elektronen, den Leitungselektronen. Besser spricht man nicht von Bindung, sondern vielleicht von einem Bund oder Verbund, was die kollektiven Natur noch deutlicher zum Ausdruck bringt.

#### Hintergrund: Metallische Bindung

Die grundsätzliche Vorstellung ist ein Bund positiv geladener Ionen, welcher durch einen See negativ geladener Elektronen zusammen gehalten wird. Dies ist das freie Elektronen-Modell nach Drude bzw. Sommerfeld, siehe Ashcroft und Mermin [11, Kap. 1 und 2]. Im Modell beinahe freier Elektronen wird zusätzlich berücksichtigt, dass aufgrund des Ionen-gitters diesem freien See kleine Potentialrumpfe aufgedrückt werden, siehe [11, Kap. 9].

Keines der Modelle vermag jedoch die Phänomene der metallischen Bindung vollständig zu beschreiben. So können s- und p-Elektronen, weil stark delokalisiert, richtig erfasst werden, aber d- und besonders f-Elektronen werden viel stärker von anderen Elektronen, also der lokalen Umgebung beeinflusst, was diese Modelle nicht wieder geben können.

Besser wird dies erfasst, wenn man ein Bild vom Übergang einzeln besetzter Orbitale zur umherwandernden Elektronen sich vor Augen führt.

Die Besonderheit metallischer Bindung ist aber nicht nur die Delokalisierung allein, sondern vor allem dass es viel mehr besetzbare Zustände als delokalisierte Elektronen gibt.

#### Delokalisierung

Delokalisierung in einer Bindung bedeutet, dass mehr als nur ein Paar von Atomen durch ein Paar von Elektronen zusammengehalten wird. Anders formuliert, ein Elektronenpaar trägt nicht nur zur Bindungsenergie zwischen zwei Atomen, sondern zwischen mehreren Atomen bei.

Das prototypische Beispiel für zwei-dimensionale Delokalisierung ist Benzol. Aneinanderreihung solcher 6er-Ringe führt zu noch größeren Beispielen dieser aromatischer Bindung: Naphtalen, Anthrazen, Ovalene, . . . .

### A.2.4. Dipolbindung

Als letzter Bindungstyp betrachten wir schließlich die Dipolbindungen. Sie sind nicht zu verwechseln mit den polaren Bindungen, welche "echt" kovalente Bindungen darstellen.

#### Wasserstoffbrückenbindung

Dipole entstehen jedoch unter anderem durch polare Bindungen und geben dadurch Anlass zu weiteren Dipolbindungen wie den Wasserstoffbrückenbindungen in Wasser. Hierbei wechselwirken statische Dipole dergestalt, um die gegenseitige Anziehung zu maximieren und die Abstoßung zu minimieren. Diese Art der Bindung ist wesentlich schwächer als die kovalente.

#### Van der Waals-Bindung

Es gibt jedoch auch dynamische Dipole, die aufgrund der Korrelation der Elektronen entstehen können. Diese lassen sich als induzierte Dipole auffassen und verursachen die so genannten Van der Waals-Wechselwirkung nach J. D. van der Waals. Die Van der Waals-Bindung kann Atombindungsstärke überschreiten.

## A.3. Obere Schranke für Vier-Zentren-Integral bei lokaler Funktion

Im Folgenden entwickeln wir eine obere Schranke für das Vier-Zentren-Integral über lokale Funktionen. Dabei folgen wir im Großen und Ganzen der Herleitung von Boys [23], der einen (beinahe) analytischen Ausdruck für das Vier-Zentren-Integral, aber für Gauss-Funktionen berechnete.

Die betrachteten Funktionen haben dabei die Gestalt:

$$f(x) = C \cdot \exp(-a\|x - A_x\|), \quad (\text{A.3.1})$$

wobei  $C$  und  $a > 0$  Konstanten sind und  $A_x \in \mathbb{R}^3$  ein Aufpunkt.

In unseren Überlegungen sind wir nur an einer oberen Schranke interessiert, weswegen wir die Konstante  $C$  in den folgenden Rechnungen zunächst vernachlässigen.

### A.3.1. Legendre-Polynomentwicklung

Mit Hilfe der Legendre-Polynome lässt sich der reziproke Abstand zwischen zwei Punkten  $x, y \in \mathbb{R}^3$ , welcher als Gewicht in den zu berechnenden Integralen häufig vorkommt, wie folgt entwickeln, siehe [26, S.508] oder auch [120, S.255]:

$$\begin{aligned} \frac{1}{\|x - y\|} &= \sqrt{\|x\|^2 + \|y\|^2 - 2\|x\| \cdot \|y\| \cos \theta}^{-1} \\ &= \sum_{l=0}^{\infty} \frac{r^l}{r^{l+1}} P_l(\cos \theta), \end{aligned} \quad (\text{A.3.2})$$

### A.3. Obere Schranke für Vier-Zentren-Integral bei lokaler Funktion

wobei  $\theta$  den Winkel zwischen  $x$  und  $y$ ,  $r$  bzw.  $r'$  die Abstände von  $x$  bzw.  $y$  zu einem gemeinsamen Referenzpunkt bezeichnet und besonders die Bedingung

$$r < r' \quad (\text{A.3.3})$$

gelten muss.

Die Legendre-Polynome, welche auf dem Intervall  $[-1, +1]$  definiert sind, bilden ein Orthogonalsystem mit folgender Relation, siehe [26, S.508]:

$$\int_{-1}^{+1} P_n(x)P_m(x)dx = \frac{2}{2m+1}\delta_{nm}. \quad (\text{A.3.4})$$

Mit einem kleinen Kniff folgt dann aus (A.3.4) sofort die Lösung für das Integral über Legendre-Polynome:

$$\int_{-1}^{+1} \sum_{l=0}^{\infty} \frac{r^l}{r'^{l+1}} P_l(x)dx = \sum_{l=0}^{\infty} \frac{r^l}{r'^{l+1}} \int_{-1}^{+1} \underbrace{P_0(x)}_{=1} P_l(x)dx = \sum_{l=0}^{\infty} \frac{r^l}{r'^{l+1}} \frac{2}{2l+1} \delta_{0l} = \frac{2}{r'} \quad (\text{A.3.5})$$

Analog lässt sich zeigen, dass aufgrund des Transformationsgesetzes mit  $x = \cos \theta$  und  $\frac{\partial \theta}{\partial \cos \theta} = \frac{-1}{\sin \theta}$  und der Vertauschung der Integrationsgrenzen gilt:

$$\int_0^{\pi} P_l(\cos \theta) \sin \theta d\theta = \int_{+1}^{-1} P_l(\cos \theta) \sin \theta \frac{-1}{\sin \theta} dx = \int_{-1}^{+1} P_l(x)dx. \quad (\text{A.3.6})$$

Dies werden wir später benötigen, wenn wir das dreidimensionale Integral über kartesische Koordinaten in sphärische Koordinaten transformieren.

#### A.3.2. Abstandsgewichtetes Gauss-Integral mit reziprotem Abstand

Für die Berechnung des Integral gewichtet mit dem reziproken Abstand zu einem zweiten Punkt  $z \in \mathbb{R}^3$  werden wir die Entwicklung nach Legendre-Polynomen aus Abschnitt A.3.1 benötigen.

$$\int_{\mathbb{R}^3} \frac{1}{\|x - y\|} \exp(-a\|x - A_x\|)dx. \quad (\text{A.3.7})$$

Wir transformieren in Koordinaten relativ zum Aufpunkt  $A_x$ , nämlich  $\chi = x - A_x$  und  $\xi = y - A_x$  mit  $r_1 := \|\xi\|$  und  $r_2 := \|\chi\|$ . Aufgrund der Translationsinvarianz des unbestimmten Integrals folgt,

$$= \int_{\mathbb{R}^3} \frac{1}{\|\chi - \xi\|} \exp(-a\|\chi\|)d\chi.$$

Wieder wenden wir nun den Satz von Fubini an und transformieren in ein sphärisches Koordinatensystem. Die Singularität am Ursprung durch den Term  $\frac{1}{\|\chi - \xi\|}$  ist dabei nur scheinbar, denn  $dV = r_2^2 \sin \theta dr_2 d\theta d\phi \propto r_2^2$ , während  $\frac{1}{\|\chi - \xi\|} \propto \frac{1}{r_2}$  ist. Wir erhalten also

$$= \int_0^{\infty} \int_0^{\pi} \int_0^{2\pi} \frac{1}{\|\chi - \xi\|} \exp(-ar_2)r_2^2 \sin \theta d\phi d\theta dr_2.$$

## A. Appendix

Wir stellen etwas um, um die Abhängigkeiten der Terme im Integranden von den Integrationsvariablen zu verdeutlichen,

$$= \int_0^\infty \exp(-ar_2) r_2^2 \int_0^\pi \int_0^{2\pi} \frac{1}{\|\chi - \xi\|} \sin \theta d\phi d\theta dr_2.$$

Um die Entwicklung (A.3.2) einsetzen zu dürfen, muss die Bedingung (A.3.3) erfüllt sein. Dazu trennen wir das Integral über  $dr_2$  in zwei Teile auf,

$$\begin{aligned} &= \int_0^{r_1} \exp(-ar_2) r_2^2 \int_0^\pi \int_0^{2\pi} \frac{1}{\|\chi - \xi\|} \sin \theta d\phi d\theta dr_2 \\ &+ \int_{r_1}^\infty \exp(-ar_2) r_2^2 \int_0^\pi \int_0^{2\pi} \frac{1}{\|\chi - \xi\|} \sin \theta d\phi d\theta dr_2. \end{aligned}$$

Nun setzen wir (A.3.2) ein,

$$\begin{aligned} &= \int_0^{r_1} \exp(-ar_2) r_2^2 \int_0^\pi \int_0^{2\pi} \sum_{l=0}^\infty \frac{r_2^l}{r_1^{l+1}} P_l(\cos \theta) \sin \theta d\theta d\phi dr_2 \\ &+ \int_{r_1}^\infty \exp(-ar_2) r_2^2 \int_0^\pi \int_0^{2\pi} \sum_{l=0}^\infty \frac{r_1^l}{r_2^{l+1}} P_l(\cos \theta) \sin \theta d\theta d\phi dr_2, \end{aligned}$$

und führen dann die triviale Integration über  $d\phi$  aus und benutzen (A.3.6), um das Integral über  $d\theta$  umzuformen,

$$\begin{aligned} &= 2\pi \left\{ \int_0^{r_1} \exp(-ar_2) r_2^2 \sum_{l=0}^\infty \frac{r_2^l}{r_1^{l+1}} \int_{-1}^{+1} P_l(z) dz dr_2 \right. \\ &\quad \left. + \int_{r_1}^\infty \exp(-ar_2) r_2^2 \sum_{l=0}^\infty \frac{r_1^l}{r_2^{l+1}} \int_{-1}^{+1} P_l(z) dz dr_2 \right\}. \end{aligned}$$

Schließlich integrieren wir mit (A.3.5) über  $dz$ :

$$\begin{aligned} &= 2\pi \int_0^{r_1} \exp(-ar_2) r_2^2 \frac{2}{r_1} dr_2 + 2\pi \int_{r_1}^\infty \exp(-ar_2) r_2^2 \frac{2}{r_2} dr_2. \\ &= \frac{4\pi}{r_1} \int_0^{r_1} r_2^2 \exp(-ar_2) dr_2 + 4\pi \int_{r_1}^\infty r_2 \exp(-ar_2) dr_2. \end{aligned} \tag{A.3.8}$$

Sei nun mit  $F(r_2) := \int r_2 \exp(-ar_2) dr_2$  die Stammfunktion bezeichnet, die wir [26, S. 1047] entnehmen,

$$F(r_2) = -(1 + ar_2) \frac{\exp(-ar_2)}{a^2}. \tag{A.3.9}$$

Wir integrieren dann partiell den ersten Summanden in (A.3.8), mit  $u = r_2$  und  $v' = r_2 \exp(-ar_2)$ , siehe [77, S.259] und erhalten

$$\int_0^{r_1} r_2 \cdot r_2 \exp(-ar_2) dr_2 = [r_2 \cdot F(r_2)]_0^{r_1} - \int_0^{r_1} F(r_2) dr_2. \tag{A.3.10}$$

### A.3. Obere Schranke für Vier-Zentren-Integral bei lokaler Funktion

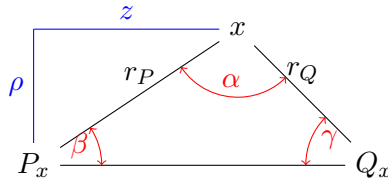
Setzen wir (A.3.10) in (A.3.8) ein, und wählen die Stammfunktion (A.3.9) – um die freie Konstante der Stammfunktion zu festzulegen – so dass also  $\lim_{x \rightarrow \infty} F(x) \rightarrow 0$  gilt, siehe ebenso [77, S.259], dann folgt

$$\begin{aligned} &= \frac{4\pi}{r_1} \left( [r_2 \cdot F(r_2)]_0^{r_1} - \int_0^{r_1} F(r_2) dr_2 \right) + 4\pi(0 - F(r_1)) \\ &= -\frac{4\pi}{r_1} \int_0^{r_1} F(r_2) dr_2. \end{aligned}$$

Wir setzen schließlich noch die Stammfunktion (A.3.9) ein und erhalten die Lösung

$$= \frac{4\pi}{a^2 r_1} \int_0^{r_1} (1 + ar_2) \exp(-ar_2) dr_2. \quad (\text{A.3.11})$$

#### A.3.3. Zwei-Zentren-Integral



**Abbildung A.1.:** Skizze zu den Abständen und Winkeln für die beiden Zentren  $P_x$  und  $Q_x$  im Zwei-Zentren-Gauss-Integral,  $x$  ist dabei der gegenwärtige Integrationspunkt.

In dieser Nebenrechnung betrachten wir das folgende doppelte Volumenintegral

$$\iint \frac{1}{\|x - y\|} \exp(-u\|x - P_x\| - v\|y - Q_x\|) dy dx, \quad (\text{A.3.12})$$

und wollen es vereinfachen. Dazu sind folgende Kniffe nötig: Zunächst werden wir die Lösung für die Ladung einer exponentiell abfallenden Funktion aus Abschnitt A.3.2 benutzen, welches unter anderem durch Transformation in sphärische Koordinaten das innere dreidimensionale Integral in ein eindimensionales Integral überführt. Anschließend werden wir auch das äußere Volumenintegral in ein spezielles Koordinatensystem transformieren, so dass insgesamt nur zwei eindimensionale Integrale verbleiben, siehe Unterabschnitt A.3.3. In A.3.3 werden wir schließlich die Nebenrechnungen aufgreifen und die Lösung präsentieren.

Wenden wir also zunächst (A.3.11) an,

$$\begin{aligned} &= \int \exp(-u\|x - P_x\|) \int \frac{1}{\|x - y\|} \exp(-v\|y - Q_x\|) dy dx \\ &= \int \exp(-u\|x - P_x\|) \frac{4\pi}{v^2 \|x - Q_x\|} \int_0^{\|x - Q_x\|} (1 + va) \exp(-va) da dx, \end{aligned}$$

## A. Appendix

und erhalten

$$= \frac{4\pi}{v^2} \int \exp(-u||x - P_x||) \frac{1}{||x - Q_x||} \int_0^{||x - Q_x||} (1 + va) \exp(-va) dadx. \quad (\text{A.3.13})$$

Es gilt nun auch das äußere Integral in ein geeignetes Koordinatensystem zu transformieren. Im Folgenden werden die das innere Integral verkürzt mit  $\int \dots da$  schreiben. Weiterhin soll  $PQ := ||P_x - Q_x||$  bezeichnen.

### Zwei-Zentren-Relativsystem

Aufgrund der Rotationssymmetrie des Systems um die Achse  $P_x Q_x$ , siehe Abbildung A.1 bietet es sich an, das Volumenintegral  $\int \dots dx$  in (A.3.13) mittels Satz von Fubini zunächst in zylindrischen Koordinaten zu schreiben:  $dx = \rho d\rho d\phi dz$ , siehe [26, S. 474]. Weiter gilt aufgrund des Kosinussatzes:

$$PQ^2 = r_p^2 + r_Q^2 - 2r_p r_Q \cos \alpha \quad (\text{A.3.14})$$

Dadurch folgt für die Zylinderkoordinaten:

$$z = r_p \cos \beta = \frac{PQ^2 + r_p^2 - r_Q^2}{2PQ} \quad (\text{A.3.15a})$$

$$PQ - z = \frac{PQ^2 - r_p^2 + r_Q^2}{2PQ} \quad (\text{A.3.15b})$$

$$\rho = \sqrt{r_p^2 - z^2} = \sqrt{r_p^2 - (PQ - z)^2} \quad (\text{A.3.15c})$$

Wir weisen darauf hin, dass wir  $\rho$  nur relativ zu  $z$  ausgedrückt haben, denn dies wird bei der folgenden Berechnung der Funktionaldeterminante, siehe [26, S. 474], eine Rolle spielen.

$$|D| = \begin{vmatrix} \frac{\partial \rho}{\partial r_P} & \frac{\partial \rho}{\partial r_Q} & 0 \\ 0 & 0 & \frac{\partial z}{\partial z} = 1 \\ \frac{\partial z}{\partial r_P} & \frac{\partial z}{\partial r_Q} & 0 \end{vmatrix} \quad \text{mit} \quad d\rho d\phi dz = |D| dr_P dr_Q dz. \quad (\text{A.3.16})$$

Wir erkennen, dass durch den Wechsel in das bekannte System der Zylinderkoordinaten die Funktionaldeterminante effektiv nur aus einer 2x2 Matrix zu bestimmen ist.

Wir berechnen die vier Terme,

$$\begin{aligned} \frac{\partial \rho}{\partial r_P} &= \frac{2r_P - 2z \frac{\partial z}{\partial r_P}}{2\rho} & \frac{\partial \rho}{\partial r_Q} &= \frac{2r_Q - 2(PQ - z) \left(-\frac{\partial z}{\partial r_P}\right)}{2\rho} \\ \frac{\partial z}{\partial r_P} &= \frac{r_P}{PQ} & \frac{\partial z}{\partial r_Q} &= \frac{-r_Q}{PQ}, \end{aligned}$$

### A.3. Obere Schranke für Vier-Zentren-Integral bei lokaler Funktion

und können dann mit diesen Termen die Determinante mit dem Laplace'schen Entwicklungssatz, siehe [26, Abschn. 4.2.1.1], bestimmen:

$$\begin{aligned} |D| &= \frac{\partial \rho}{\partial r_P} \left( 0 - 1 \cdot \frac{\partial z}{\partial r_Q} \right) - 0 + \frac{\partial z}{\partial r_P} \left( \frac{\partial \rho}{\partial r_Q} \cdot 1 - 0 \right) \\ &= \frac{r_P r_Q}{\rho P Q}. \end{aligned} \quad (\text{A.3.17})$$

Ingesamt erhalten wir damit durch Hintereinanderschaltung der beiden Funktionaldeterminanten, um direkt von  $dx$  nach  $dr_P dr_Q d\phi$  zu wechseln:

$$|D_{x \rightarrow r_P, r_Q, \phi}| = \frac{r_P r_Q}{P Q}. \quad (\text{A.3.18})$$

Schließlich müssen wir noch auf die Integrationsgrenzen im neuen Koordinatensystem eingehen: Da wir die Abbildung nicht explizit aufgeschrieben haben, veranschaulichen wir an dieser Stelle nur die Überlegungen:  $dx$  wurde über ganz  $\mathbb{R}^3$  integriert, in Zylinderkoordinaten ist das bekanntermaßen  $[0, \infty) \times [0, 2\pi) \times [0, \infty)$ . In dem Zwei-Zentren-System ergibt jedoch nicht jedes  $r_P, r_Q \in [0, \infty)$  ein gültiges  $x$ . Stellt man sich wie in Abbildung A.1 die beiden Abstände  $r_P, r_Q$  als Radien von Kreisen um  $P_x$  und  $Q_x$  vor, so müssen diese Kreise sich schneiden, um ein gültiges  $x$  zu erzeugen. Dies ist nicht mehr der Fall, wenn ein Kreis so groß wird, dass er den anderen komplett beinhaltet, z. B. bei  $r_P = PQ + 2r_Q$ . Damit folgen für  $r_P, r_Q, d\phi$  die Grenzen:  $[0, \infty) \times [PQ - r_Q, PQ + r_Q] \times [0, 2\pi)$ . Dieses Gebiet im  $\mathbb{R}^3$  wäre also das Bild der nicht explizit aufgeschriebenen Abbildung von  $\mathbb{R}^3$ .

#### Die kurze Lösung

Mit diesen Vorarbeiten können wir nun die Lösung für (A.3.12) recht schnell finden. Wir wenden also das Zwei-Zentren-Relativsystem (A.3.18) auf (A.3.13) an und führen die Integration über  $d\phi$  durch:

$$= \frac{8\pi^3}{v^2} \int_0^\infty \int_{|r_Q - PQ|}^{r_Q + PQ} \exp(-ur_P) \frac{1}{r_Q} \int_0^{r_Q} (1 + va) \exp(-va) da \frac{r_P r_Q}{P Q} dr_P dr_Q.$$

Die Stammfunktion von  $r_P \exp(-ur_P)$  ist durch (A.3.9) bekannt. Damit können wir die Integration über  $dr_P$  durchführen und wir trennen das Integral über die Differenz im Integranden auf,

$$\begin{aligned} &= \frac{8\pi^3}{u^2 v^2 P Q} \left\{ \int_0^\infty (1 + u|r_Q - PQ|) \exp(-u|r_Q - PQ|) \int_0^{r_Q} \dots dadr_Q \right. \\ &\quad \left. \dots - \int_0^\infty (1 + u(r_Q + PQ)) (\exp(-u(r_Q + PQ))) \int_0^{r_Q} \dots dadr_Q \right\}. \end{aligned}$$

Die Transformation des ersten Ausdruck  $r_Q \rightarrow -r_Q$  mit anschließender „Rücktransformation“  $a \rightarrow -a$ , um die Integranden und Integralgrenzen anzugleichen, wie es für Gauss-Funktionen möglich war, geht bei einfach exponentiell abfallenden Funktionen nicht.

## A. Appendix

Aber wir sind auch nur an einer Abschätzung nach oben interessiert. Wir können den zweiten Summanden fallen lassen, da seine Integranden nicht-negativ sind, genau wie auch seine Integralgrenzen,

$$\leq \frac{8\pi^3}{u^2v^2PQ} \int_0^\infty (1 + u|r_Q - PQ|) \exp(-u|r_Q - PQ|) \int_0^{r_Q} (1 + va) \exp(-va) da dr_Q$$

Führen wir die Integration des inneren Integrals aus, wobei wir die Stammfunktionen wieder [26, Abschn. 21.5.4.2 (447.-448.)] entnehmen, und schätzen anschließend ab, da  $v \leq 0$  und  $r_Q \leq 0$ ,

$$\begin{aligned} & \int_0^{r_Q} (1 + va) \exp(-va) da \\ &= \int_0^{r_Q} \exp(-va) da + v \int_0^{r_Q} a \exp(-va) da \\ &= \left[ \frac{1}{v} \exp(-va) \right]_0^{r_Q} + \left[ \frac{1}{v} (1 + va) \exp(-va) \right]_0^{r_Q} \\ &= \frac{2}{v} (1 - \exp(-vr_Q)) - \frac{1}{2} vr_Q \exp(-vr_Q) \\ &\leq \frac{2}{v}. \end{aligned} \tag{A.3.19}$$

Da die Abschätzung unabhängig von  $r_Q$  ist, können wir damit nun auch das äußere, verbleibende Integral lösen, siehe [26, Abschn. 21.6.2 (23.)],

$$\begin{aligned} & \int_0^\infty (1 + u|r_Q - PQ|) \exp(-u|r_Q - PQ|) dr_Q \\ &= \int_0^{PQ} (1 + u|r_Q - PQ|) \exp(-u|r_Q - PQ|) dr_Q \\ & \quad + \int_{PQ}^\infty (1 + u|r_Q - PQ|) \exp(-u|r_Q - PQ|) dr_Q \\ &= \int_{PQ}^0 (1 + ur_Q) \exp(-ur_Q) dr_Q + \int_0^\infty (1 + ur_Q) \exp(-ur_Q) dr_Q \\ &= - \int_0^{PQ} (1 + ur_Q) \exp(-ur_Q) dr_Q + \int_0^\infty (1 + ur_Q) \exp(-ur_Q) dr_Q \\ &\leq \int_0^\infty (1 + ur_Q) \exp(-ur_Q) dr_Q \end{aligned}$$

Die Lösung dieses Integrals und die Abschätzung kennen wir durch (A.3.19).

Insgesamt erhalten wir damit als obere Schranke,

$$\leq \frac{32\pi^3}{u^3v^3PQ} \propto \frac{1}{PQ}. \tag{A.3.20}$$



## A.4. Literatur

Der Corpus der Fragmentierungsansätze erstreckt sich über einen Zeitraum von über 60 Jahren. Wir wollen an dieser Stelle nicht den Ansatz versuchen, ihn umfassend und historisch korrekt wiederzugeben.

Dennoch müssen wir sowohl Arbeiten wiedergeben, die die in dieser Arbeit dargelegten Ideen schon früher gehabt haben, als auch jene Anzahl von Alternativmethoden kurz darlegen. Wir konzentrieren uns dabei vor allem auf Methoden im Rahmen der HF-Theorie und intramolekulare Wechselwirkungen.

### A.4.1. Vorarbeiten

Aufgrund der notwendigen Antisymmetrie der Gesamtwellenfunktion wurde anstelle der vollen  $N!$  Anzahl an Termen der Slater-Determinante, Unterteilungen gesucht, um eine Hierarchie einzuführen. Auf deren Basis können Teile dieses Hierarchiebaums vernachlässigt und damit die Gesamtzahl an Termen eingeschränkt werden.

Nach Kenntnis des Autors ist der Urvater der Fragmentierungs-idee in der Quantenmechanik die Arbeit von Hurley et al. [102], in der vorgeschlagen wird, Elektronen zu Paaren zu kombinieren.

McWeeny nahm schon früh diese Idee auf und schlug vor, die Einteilchen-Wellenfunktion in Gruppen zusammen zu fassen. Diese werden als verallgemeinerte Produktfunktionen bezeichnet. Mit Gruppen sind dabei eben nicht nur reine Paare, sogenannte Geminals, sondern auch Viel-Elektron-Gruppen wie konjugierte  $\pi$ -Elektronen oder die Kernelektronen gemeint und wie folgt angesetzt,

$$\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = M \hat{A} \left[ \psi_a(\mathbf{r}_1, \dots, \mathbf{r}_{N_a}) \cdot \psi_b(\mathbf{r}_1, \dots, \mathbf{r}_{N_b}) \cdot \dots \right], \quad (\text{A.4.1})$$

wobei  $\hat{A}$  den Antisymmetrisierer,  $M_k$  die Normierungskonstante und  $\psi_k$  die verschiedenen Gruppenfunktionen über  $N_k$  der Gruppe  $k$  bezeichnet, siehe [135, Kap. 14]. Mittels Linearkombination mehrerer dieser Funktionen, also MCSCF, wird dann die elektronische Wellenfunktion beschrieben. Entscheidend ist die starke Orthogonalität, welche zwischen Gruppenfunktionen gelten soll. Diese Aufspaltung betrachtet er dabei zur Berechnung intermolekularer Kräfte, zwischen denen solch eine starke Orthogonalität gelten kann. In [118] wird dieser Ansatz sehr früh anhand von Methan untersucht. Dort in den Schlussbemerkungen werden chemische Bindungen als mögliche Paare angesprochen, dass deren schwache Kopplung zu schneller SCF-Konvergenz führt und dass durch die Unterteilung in Gruppen die Teilprobleme sehr viel einfacher werden<sup>2</sup>. Dies beschreibt die grundlegende Motivation für alle Fragmentierungsansätze. Nachteilig ist, dass die Gruppen auf Basis chemischer Intuition gewählt werden müssen.

### A.4.2. Einschränkung der Basisgröße

Das grundsätzliche Problem bei der numerischen Lösung des Elektronenstrukturproblems in der Anfangszeit der numerischen Methoden mit damals knappen Rechenressourcen

<sup>2</sup>bspw. mussten dadurch nur  $3 \times 3$ -Matrizen gelöst werden.

## A. Appendix

war die Größe der Basis zur Beschreibung der Einteilchen-Wellenfunktionen. Entweder möchte man also das Problem in einem kleineren, aber möglichst besser geeigneten Ansatzraum aufstellen, oder man will den Ansatzraum so unterteilen, dass Lösungen schon in Teilräumen gefunden werden können.

### Fragmentansatz

In Christoffersen und Maggiora [34] werden nicht bestimmte Elektronengruppen zusammen gefasst, sondern das elektronische Problem auf Basis von Kerngruppen unterteilt, von sogenannten Molekülfragmenten. Nach Kenntnisstand des Autors taucht hier zum ersten Mal die Idee auf, die dabei durchtrennten Bindungen dieser Molekülfragmente mittels Wasserstoff zu saturieren.<sup>3</sup> Dabei werden FSGOs eingesetzt, welche jedoch nicht nur an den Kernorten, sondern auch zwischen diese gesetzt werden und deren Aufpunkt und andere Parameter zur Beschreibung des Fragments nicht-linear minimiert werden. Diese dann fest parametrisierten FSGOs werden als Basisfunktionen zur Entwicklung der gesuchten Einteilchen-Wellenfunktionen benutzt. Hier dient die Fragmentierung also der Bestimmung einer kleineren, geeigneteren Basis. In [154] wird dazu auf Basis der Fragmente das Aufstellen der anfänglichen Dichtematrix vorgeschlagen. Die Möglichkeit zur Beschreibung großer Moleküle entsteht dabei aufgrund der Einschränkung der Anzahl der FSGOs pro Fragment, wobei die Genauigkeit nicht im Vordergrund steht. In [33] werden eine ganze Reihe von Arbeiten zusammengefasst und mit Fokus auf Moleküleigenschaften, welche auf Basis dieser Methode berechnet wurden, mit Werten aus der Literatur verglichen.

### Molekülorbitalansätze

Mit der Idee der Lokalisierung der aus den SCF-Rechnung resultierenden kanonischen Orbitale nach Foster und Boys [60] oder auch Edmiston und Ruedenberg [50] wurde es möglich, diese MOe nach bekannter chemischen Intuition zu interpretieren. Dadurch entstand wohl auch die Idee der Transferabilität der MOe von einem Molekül auf ein ähnliches anderes.

Ein früher Ansatz ist die Arbeit von Degand et al. [44], der MOe von kleineren, verwandten Molekülen als Basisfunktionen benutzt, um die Fockmatrix in diesen aufzustellen und zu diagonalisieren. Hierbei wird die SCF-Iteration also nur in den kleineren Molekülen durchgeführt.

von Niessen [177] beschäftigt sich ebenfalls mit den Schwierigkeiten der Transferabilität. Neben der chemischen Interpretierbarkeit dieser räumlichen Orbitale stellt er dies als ihren Hauptnutzen heraus. In der vorgestellten Methode werden ebenso Orbitale aus kleineren Molekülfragmenten übernommen. Hierbei wird jedoch differenziert unterschieden, ob sich bestimmte Orbitale zu Kern- oder Valenzelektronen, nahe oder fern den Schnittstellen gegenüber dem Fragment verändern werden und deswegen erneut in einem

---

<sup>3</sup>In dieser noch jungen Zeit der computerbasierten Forschung machte dies die Berechnung sogar genauer, da dadurch die Basis des Gesamtmoleküls effektiv vergrößert wurde.

SCF-Zyklus minimiert werden sollten. Projektion sichert dabei die Orthogonalität des Orbital-Satzes.

### Zerlegung der Ladungsdichte

Von der Seite der Chemie ging von den „Additivity Rules“, siehe [16], die Idee zur Fragmentierung des Molekülsystems und besonders seiner Ladungsdichte aus. Dies wurde vor allem durch das wichtige Theorem von Hohenberg und Kohn [95] ermöglicht.

Bader [12] schlug, basierend auf einer Untersuchung des Virial-Theorems, vor, die Ladungs- und Stromdichte eines Moleküls in Virialfragmente zu unterteilen, um Eigenschaften des ganzen Moleküls aus einzelnen Teilen zu ermitteln. In [13] wird dies als „Atom im Molekül“ bezeichnet. Dabei wird das System in Untersysteme unterteilt, auf deren Rändern der Ladungsfluß verschwinden soll.

Auch Parr et al. [140] benutzte das Konzept vom „Atom im Molekül“ und untersuchte die Ladungsdichte im Rahmen der DFT. In Parr et al. [141] in informationstheoretischen Überlegungen mit Bezug auf Hirshfeld [93] bezeichnet er dieses Konzept aber als rein fiktives, wenn auch nützliches Konstrukt.

Später wird diese Idee der Dichtezerlegung aufgegriffen, um linear skalierende Fragmentierungsansätze zu beschreiben.

#### A.4.3. Fragmentmethoden

Mittels der Fragmentierungsidee kann auf sehr einfache Weise eine linear skalierende Methode erhalten werden, wenn die Anzahl der zu berechnenden Fragmente nur linear mit der Anzahl der Kerne ansteigt und diese selbst von konstanter Größe sind.

#### Molekülorbital-basiert

In Gresh et al. [75] wird eine zunächst entwickelte Methode namens SIBFA zur Berechnung der langreichweitigen elektrostatischen, sowie Polarisations- und Dispersionsenergien erweitert, indem die elektronische Wellenfunktion ebenfalls aus Untereinheiten, nämlich Molekülfragmenten, konstruiert wird. Auch hier werden Wasserstoffe zur Saturation benutzt, aber auch weitere fiktive Atome. Dabei wird die kurzreichweitige Abstoßung aus der Auswertung von Paar-Termen gewonnen. Der langreichweitige elektrostatische Anteil wird aus einer Multipolentwicklung ermittelt. Die Methode hat grundsätzlich quadratischen Aufwand  $\mathcal{O}(4M^2)$  aufgrund der Paar-Terme und benötigt chemisches Wissen, um geeignete Fragmente zu bestimmen.

Amovilli und McWeeny beschreiben ausführlich in [5] einen Ausdruck für die Wechselwirkungsenergie eines Dimers  $E_{\text{int}} = E - E_A - E_B$ , wobei  $A$  und  $B$  die beiden Monomere bezeichnet. Dabei wird die Wellenfunktion als Gruppenfunktion (A.4.1) angesetzt und der Ausdruck störtheoretisch nach zweiter Ordnung trunkiert. Sind die Monomere ausreichend entfernt voneinander, dann besteht die Grundzustandsenergie dabei nur aus der elektrostatischen Wechselwirkung. Zusätzliche Anregungen führen zu weiteren Polarisations- und Dispersionsanteilen in der Wechselwirkungsenergie. In kleineren Abständen zwischen den Monomeren, wo van-der-Waals-Wechselwirkungen wirksam sind,

tauchen zusätzlich Austausch- und Strafterme aufgrund des Überlapps der jetzt nicht mehr stark orthogonalen Gruppenfunktionen auf.

Walker und Mezey [179] beschreiben eine Methode, in der Ladungsdichte von Molekülfragmenten aus [130] bausteinartig zusammengesetzt werden. Dies können dabei einer Datenbank entnommen werden. Auch hier wird Wasserstoffsaturation benutzt. Der Fokus liegt vor allem in einer Analyse der Molekülform. Die Fragmente werden dabei mittels HF berechnet. Die Laufzeit skaliert linear mit der Anzahl der Kerne.

In einer späteren Arbeit verweist Amovilli et al. [6] auf die Ideen von Claverie, siehe [75], um die notwendigerweise große Einelektron-Basis für intermolekulare Energien durch paarweise Betrachtung von Molekülfragmenten zu verkleinern und damit die Berechnung größerer Systeme zu ermöglichen. Hierbei wird wieder das System mit chemischem Wissen aufgetrennt. Wird stattdessen das gesamte Molekülsystem in die Rechnung einbezogen, welches aus mehreren einzelnen Molekülen besteht, spricht man von einer *Supermolekül*-Rechnung. Stoll [163] hatte 1992 einen vergleichbaren Ansatz beschrieben, wobei er die Korrelationsenergie von Paaren, Tripeln, usw. von lokalisierten Bindungsorbitalen in Diamant betrachtete. Diese Inkremente stehen in direktem Zusammenhang zu (7.2.6b), wenn wir anstelle der Grundzustandsenergie die Korrelationsenergie betrachten würden, siehe auch Abschnitt 8.4.2. Hierbei stehen besonders die Wechselwirkungen im Fokus, welche aus der Korrelation der Elektronen resultieren.

Kitaura et al. [115] beschreiben den PIMO-Ansatz, bei dem anstelle einer Supermolekülrechnung Moleküle nur paarweise betrachtet werden. Dabei werden jeweils für Monomer  $I$  und Dimer  $IJ$  entsprechende Hamiltonoperatoren  $\hat{H}_I$  bzw.  $\hat{H}_{IJ}$  aufgestellt, wobei diese zusätzlich elektrostatische Wechselwirkungen zu allen anderen Monomeren bzw. Dimeren enthalten. Die zugehörigen elektronischen Schrödingergleichungen  $\hat{H}_I\Psi_I = E^{(\text{tot})}'_I\Psi_I$  bzw.  $\hat{H}_{IJ}\Psi_{IJ} = E^{(\text{tot})}'_{IJ}\Psi_{IJ}$  werden gelöst. Dabei werden die Energie jedes Monomers im elektrostatischen Feld der anderen Monomere minimiert. Mit der fehlenden Kernabstoßungsenergie erhalten wir dann die Gesamtenergie  $E^{(\text{tot})}$  des Molekülsystems,

$$E^{(\text{tot})} = \sum_{I < J} \left( E^{(\text{tot})}_{IJ} - E^{(\text{tot})}_I - E^{(\text{tot})}_J \right) + \sum_I E^{(\text{tot})}_I,$$

also die zu (7.2.6a) analoge orthogonale Summe. Hier werden aber ganze Moleküle und nicht einzelne Nicht-Wasserstoffatome als grundlegende Einheiten betrachtet. In [116] wird ein verwandter Ansatz namens FMO für die intramolekulare Energie großer Moleküle vorgeschlagen. Dabei wird das Molekül mit chemischem Wissen in Fragmente aufgeteilt und mit lokalisierten räumlichen Orbitalen beschrieben. Dabei enthält der Hamiltonoperator des Fragments wieder die elektrostatische Wechselwirkung mit der Ladungsdichte aller anderen Fragmente und diese wird iterativ berücksichtigt. Anschließend werden wieder diese Fragmente paarweise betrachtet, um die Gesamtenergie zu erhalten. Der Dimer-Hamiltonoperator folgt dann analog. Besonders werden dabei lokalisierte MOe nach [60] als Basisfunktionen benutzt, welche aus Referenzmolekülen, passend zu den Fragmenten, erzeugt wurden. In Fedorov und Kitaura [54] wird dieser Ansatz auf Trimere erweitert.

Vreven und Morokuma [178] kombinieren quantenmechanische und molekülmechanische Berechnung indem sie eine analoge orthogonale Summe wie im PIMO-/FMO-Ansatz

benutzen. Dabei wird das System wieder in verschiedene Anteile unterteilt, die mit unterschiedlich genauen Methoden berechnet werden. Die Unterteilung erfolgt wiederum mit chemischem Wissen.

Zhang et al. [189] stellen eine Methode namens MFCC vor, um die Wechselwirkungsenergie eines großen Moleküls zu einem kleinen Molekül zu berechnen. Dabei wird das große Moleküle in Fragmente mit chemischem Wissen unterteilt, wobei durchtrennte Bindungen mit spezifisch gewählten sogenannten Kappen saturiert werden. Eine Reihe weiterer Arbeiten wenden diesen Ansatz auf die Aufstellung der Gesamtdichte für DFT-Berechnungen der Gesamtenergie und für die Aufstellung von Dichtematrizen an.

In Jacobson und Herbert [107] findet sich schließlich eine umfassende, aber knappe Auflistung der verschiedenen Fragmentansätze.

Allgemein berichten diese Methoden von Fehlern im Bereich von 1 kcal/mol für untersuchte kleinere Systeme, also von relativen Genauigkeiten von ca.  $10^{-5}$ .

### Dichte-basiert

Die folgenden Ansätze zum Aufstellen linear skalierender Elektronenstrukturmethoden stützen sich alle auf die Idee der Zerlegung der Ladungsdichte.

Anstelle der von uns benutzten HF-Methode betrachtet [186] zunächst die KS-Gleichung im Rahmen der DFT und benutzt eine Zerlegung der Eins im  $\mathbb{R}^3$ , um die Ladungsdichte in fragmentarische Anteile aufzuteilen,

$$\rho(r) = \sum_a \rho^a(r) := 2 \sum_a p^a(r) \int \phi(\mathbf{r})^* \eta(\epsilon_F - \hat{H}^{(\text{KS})}) \phi(\mathbf{r}) d\mathbf{r} \quad (\text{A.4.2})$$

wobei  $\eta(x)$  die Heaviside-Funktion,  $\epsilon_F$  die Fermi-Energie und  $\hat{H}^{(\text{KS})}$  den KS-Hamiltonian bezeichnet. Mit der Zerlegung der Eins im  $\mathbb{R}^3$  geht eine Zerlegung des Ansatzraums für die Aufstellung der KS-Gleichung notwendigen Orbitale einher. Gleichzeitig wird auch der Hamiltonian durch lokale Funktionen auf einen Unterraum eingeschränkt. In der Informatik entspricht dies einem Divide&Conquer-Ansatz. Auch hier wird eine sinnvolle Unterteilung mit chemischem Wissen gewählt. Durch das notwendige Aufstellen der Matrix und die Berechnung der langreichweitigen Wechselwirkung bleibt sie aber zunächst quadratisch skalierend mit der Anzahl der Kerne.

Einige andere Ansätze zu linear skalierenden Methoden haben wir bereits in Abschnitt 7.6 zur Dichte-basierten HF-Methode besprochen.

#### A.4.4. Zusammenfassung

In dem gemachten Überblick wird die historische Entwicklung erkennbar. Anfangs stand noch die Suche nach einer geeigneten Basis und der anfänglichen Dichtematrix im Vordergrund. Dieses Problem verschwand mit den heute als Standard bekannten Chemikerbasen, siehe Abschnitt A.1 im Anhang, und robusten Methoden zur Startwertbestimmung des SCF-Zyklus, siehe den zugehörigen Paragraph unter Abschnitt 4.3.2. Später ergab sich auch stark ein phänomenologischer Zusammenhang in der Fragmentierung, wo die

## A. Appendix

Gesamtenergie in einzelne Anteile aufgespalten und diese auf verschiedene Arten berechnet wurden. Alternativ wurde die Ladungsdichte auf verschiedene Unterteilungen untersucht. Schließlich rückte der Fokus weg davon die Anzahl an Basisfunktionen klein zu halten und hin dazu den Exponenten in der Komplexität von  $\mathcal{O}(n^4)$  der Methode zu verringern. Besonderer Schub ergab sich durch die FMM [72], welche linear skalierende Komplexität der Berechnung der langreichweitigen Wechselwirkung ermöglichte. Dies sollte stets die Berechnung immer größerer Systeme ermöglichen. Parallel wird noch deutlich vor der Jahrtausendwende die Berechnung der Korrelationsenergie und die damit zusammenhängenden Post-HF-Methoden immer wichtiger. So wird in den Methoden der Charakter der intermolekularen Wechselwirkungen immer stärker.

Die vielen Methoden unterscheiden sich letztlich nur in wenigen Punkten:

- Aufteilung des Gesamtsystems
  - nach den Kernen (Fragmente)
  - nach der Dichte (Zerlegung der Eins)
- Selbst-konsistente oder störartige Berechnung intermolekularer Wechselwirkungen

Alle Methoden stützen sich auf die Annahme der Lokalität der Einteilchen-Wellenfunktion, gleich ob sie nun in einem Wellenfunktion-basierten Bild oder einem Ladungsdichtebild arbeiten. Dies bedeutet, dass Lösungen und deren darstellende Basisfunktionen einen exponentiellen Abfall im Betrag zeigen müssen und möglichst konzentriert um einen Aufpunkt liegen.

Die zweite Schwierigkeit liegt im Aufbrechen des SCF-Zykels, der sowohl bei HF- als auch bei DFT-basierten Methoden über das Gesamtsystem geht. Bei den Fragmentierungsmethoden wird dieser generell in zwei Zyklen hierarchisch aufgespalten: einen inneren, der nur über das Fragment selbst geht, und einen äußeren, der alle restlichen Wechselwirkungen berücksichtigt. Mit dem inneren Zyklus, dessen Lösung teils auch als Datenbankwert gespeichert werden kann, wird entweder ein guter Startwert für den äußeren Zyklus gesucht – wenn die MOs als Basisfunktionen oder zumindest als Anfangskoeffizienten benutzt werden – oder der äußere Zyklus wird ganz weggelassen, indem, bewußt oder unbewußt, der Hamiltonoperator zunächst um die Anteile trunkiert wird, welche Wechselwirkungen zwischen den Fragmenten darstellen. Diese fehlenden Anteile können anschließend in nachfolgenden Rechnungen in Störtheorie eingehen.

Schließlich kommt die Korrelationsenergie ins Spiel, welche HF-Methoden grundsätzlich nicht ermitteln können, für die intermolekulare Wechselwirkung aber eine große Rolle spielt. Hier kann entweder in einem reinen Störansatz die fehlende Energie in erster Ordnung berechnet, noch höhere Ordnungen Störtheorie benutzt oder gar innerhalb eines äußeren SCF-Zyklus minimiert werden.

Unsere Methode ordnet sich ebenso in diesen Kontext ein. In dieser Arbeit haben wir uns jedoch besonders bemüht, die Frage zu motivieren, warum Fragmentansätze überhaupt funktionieren und welche Voraussetzungen sie möglich machen. Wir haben in Kapitel 5 lokale Funktionen definiert und uns anschließend auf die asymptotische Lokalität der Einteilchen-Wellenfunktionen gestützt. Wir bauen direkt auf den lokalen Chemikerbasen auf, haben dazu die notwendigen Eigenschaften dieser Basisfunktionen untersucht,

um das durch ihre Verwendung resultierende algebraische Problem rasch lösen zu können. Wir haben uns außerdem mit einer linear skalierenden Methode für das eigentliche Eigenwertproblem beschäftigt, siehe Kapitel 6, was unseres Wissens nach in keinem der Artikel untersucht wird. Gleichzeitig legten wir Wert auf möglichst große Unabhängigkeit von chemischem Wissen durch Abstraktion auf einen (Bindungs-)Graphen und dessen automatisierte Zerlegung in alle verbundenen Teilgraphen, so dass die Methode robust und weitestgehend Molekülsystemunabhängig arbeitet. Auch auf diesen Aspekt konzentriert sich keine der dargestellten Methoden.

## A.5. Wasserstoff-Saturation

Die Fragmentlösungen weisen eine Besonderheit auf: Sie sind durch die Saturation durchtrennter Bindungen mit Wasserstoff verändert. Das bedeutet, dass die Ladungsdichte zusätzliche Anteile durch jedes zusätzliche Elektronen des Saturationswasserstoff enthält.

Wir wollen hier kurz darauf eingehen, warum sich ab einer Bindungsordnung von  $k = 2$  dieser Anteil an der Ladungsdichte in guter Näherung weghebt. Dazu machen wir nun ein illustratives Beispiel. Ein analoges Argument gilt für die Grundzustandsenergie.

Nehmen wir also an, wir hätten ein Stoffsystem bestehend aus drei nicht-Wasserstoff-Kernen  $a = 1, 2, 3$ , die jeweils mit Einfach-Bindungen zu einer Dreier-Kette zusammen geschlossen sind. In der üblichen Weise fragmentieren wir nun den Bindungsgraphen dieses Systems. Zunächst wählen wir als größte Ordnung  $k = 1$ . Dann erhalten wir drei verbundene Untergraphen  $I_1 = \{1\}$ ,  $I_2 = \{2\}$  und  $I_3 = \{3\}$ , die jeweils einen Index enthalten. Das dazugehörige Fragment besteht aus zwei Kernen, einmal dem Kern  $i$  des Fragments und im Falle von  $i = 1, 3$  einem und im Falle von  $i = 2$  sogar zwei weiteren Wasserstoffen, welche die durchtrennte Bindung jeweils saturieren. Anschließend ermitteln wir mittels der HF-Methode die Grundzustandselektronwellenfunktion und erhalten daraus auch die elektronische Ladungsdichten  $\rho_i$  zu jedem Fragment  $i$ , siehe (4.2.4). Wir gehen davon aus, dass aufgrund der Lokalität diese Dichten abseits der Kernorte des zugehörigen Fragments schnell abfallen. Nun enthält jede dieser Dichten zusätzliche Anteile  $\rho_i^s$  aufgrund der Saturation. Stellen wir uns ein anderes (unbekanntes) Verfahren vor, das uns die gleiche Dichte ohne Saturationsanteil  $\rho_i^0$  liefern würde und schreiben deswegen formal die Dichte wie folgt auf:

$$\rho_i = \rho_i^0 + \rho_i^s \quad (\text{A.5.1})$$

Setzen wir diese Dichte nun in Gleichung (7.2.7) sein – wobei wir noch nichts weiter abziehen müssen, da die Kernindexmengen ja (noch) disjunkt sind – so erhalten wir keine gute Näherung an die Gesamtdichte  $\rho$ :

$$\rho \approx \sum_{i=1}^3 \rho_i^0 + \underbrace{\sum_{i=1}^3 \rho_i^s}_{=:\Delta\rho^s},$$

denn sie enthält einen signifikanten Saturationsanteil  $\Delta\rho^s$ :  $\int \Delta\rho^s(\mathbf{x}) = 4$ , aufgrund der insgesamt vier Saturationswasserstoffe mit je einem Elektron.

Betrachten wir nun aber die Fragmentierung mit der Ordnung  $k = 2$ . Wir erhalten nun zwei weitere Fragmente  $I_4 = \{1, 2\}$  und  $I_5 = \{2, 3\}$ , da wir nur verbundene Teilgraphen zulassen. Diese ergeben wiederum entsprechende Ladungsdichten aus ihren Grundzustandselektronwellenfunktionen, welche wir uns wiederum wie in (A.5.1) aufgespalten denken. Die Fragmente  $I_4$  und  $I_5$  haben jedoch nur jeweils einen Sauerstoffwasserstoff. Setzen wir nun diese in die ANOVA-Formel (7.2.7) ein, so müssen wir nun zunächst von den neuen Anteilen wie in (7.2.8) die enthaltenen abziehen. Also zunächst am Fragment  $i = 4$ ,

$$\begin{aligned}
\tilde{\rho}_4 &= \rho_4 - \rho_1 - \rho_2 \\
&= \rho_4^0 + \rho_4^s - \rho_1^0 - \rho_1^s - \rho_2^0 - \rho_2^s - \rho_{2,a}^s - \rho_{2,b}^s \\
&= \rho_4^0 - \rho_1^0 - \rho_2^0 - \rho_1^s - \rho_{2,a}^s + \rho_4^s - \rho_{2,b}^s \\
&=: \tilde{\rho}_4^0 + \rho_4^s - \rho_{2,b}^s \\
&=: \tilde{\rho}_4^0 + \tilde{\rho}_4^s.
\end{aligned}$$

Dabei haben wir die zwei Sauerstoffwasserstoffe des Fragments  $I_2$  in zwei Anteile  $\rho_{2,a}^s$  und  $\rho_{2,b}^s$  aufgespalten. Da die Dichten  $\rho_1^s$  und  $\rho_{2,a}^s$  die Anzahl der Kerne mit  $\rho_4^0$  teilen, verschmelzen diese drei Anteile zum Beitrag  $\tilde{\rho}_4$ . Das ist schließlich genau der Zweck der Sauerstoffung: Sie nimmt bei Ordnung  $k = 1$  die sich auf höheren Ordnungen manifestierende Bindung voraus. Oder mit anderen Worten: Wir stellen die Änderung der Dichte aufgrund der Bindung zwischen den Nicht-Wasserstoffkernen 1 und 2 als Korrektur zu der überlagerten Dichte dar, wenn jeder Kerne stattdessen an Wasserstoff gebunden wäre.

$\rho_4^0$  ist nun also die Korrektur auf Ordnung  $k = 2$  zur Kernindexmenge  $\{1, 2\}$ . Für die verbleibende Sauerstoffdichte  $\tilde{\rho}_4^s$  gilt nun, dass sie näherungsweise verschwindet,

$$\int \rho_4^s(\mathbf{x}) - \rho_{2,b}^s(\mathbf{x}) d\mathbf{x} \approx 0,$$

wenn die Sauerstoffdichten nahezu unbeeinflusst von der chemischen Umgebung bleiben. Genau aus diesem Grund wählen wir Wasserstoff als Sauerstoffelement, denn es kommt dieser Forderung am nächsten<sup>4</sup>.

Für das verbleibende Fragment  $i = 5$  ergibt sich das Analoge.

Summieren wir nun erneut auf,

$$\rho = \sum_{i=1}^5 \tilde{\rho}_i^0 + \underbrace{\sum_{i=1}^5 \tilde{\rho}_i^s}_{=: \Delta\rho^s},$$

dann ergibt sich der Sauerstoffanteil zu:

$$\begin{aligned}
\Delta\rho^s &= \tilde{\rho}_1^s + \tilde{\rho}_2^s + \tilde{\rho}_3^s + \tilde{\rho}_4^s + \tilde{\rho}_5^s \\
&= \rho_4^s - \rho_{2,b}^s + \rho_5^s - \rho_{2,a}^s \approx 0.
\end{aligned}$$

<sup>4</sup>Diese Eigenschaft wird bspw. durch die relativ geringe chemische Verschiebung im Rahmen von NMR-Messungen an Molekülen belegt, siehe [90].



## Abkürzungsverzeichnis

ANOVA	. . . . .	ANalysis Of VAriance
AO	. . . . .	Atomares Orbital
BO	. . . . .	Born-Oppenheimer
Boost	. . . . .	Boost
ASIO	. . . . .	Boost ASynchronous Input/Output library
BOSSANOVA	. . .	Bond-Order diSSection in an ANalysis Of VAriance-like fashion
CASSCF	. . . . .	Complete Active Space Self-Consistent Field
CC	. . . . .	Coupled Cluster
CGTO	. . . . .	Zusammengefasste Gaussförmige Orbitale
CI	. . . . .	Configuration Interaction
CLHF	. . . . .	Closed-Shell Hartree-Fock
CSF	. . . . .	Configuration State Function
DFT	. . . . .	Dichtefunktionaltheorie
FMM	. . . . .	Schnelle Multipol-Methode
FMO	. . . . .	Fractional Molecular Orbit
FSGO	. . . . .	freies, sphärisches Gaussförmiges Orbital
GTO	. . . . .	Gaussförmiges Orbital
GWH	. . . . .	Generalized Wolfsberg-Helmholtz
HF	. . . . .	Hartree-Fock
KS	. . . . .	Kohn-Sham
LCAO	. . . . .	Linearkombination atomzentrierter Orbitale
MCSCF	. . . . .	Multi-Configuration Self-Consistent Field
MLWF	. . . . .	Maximal Lokalisierte Wannier-Funktion
MFCC	. . . . .	Molecular Fractionation with Conjugate Caps

MP2 . . . . .Møller-Plesset 2.Ordnung  
 MPPT . . . . .Møller-Plesset Störtheorie  
 MPQC . . . . .Massively Parallel Quantum Chemistry  
 OHF . . . . .Open-Shell Hartree-Fock  
 MO . . . . .Molekülorbital  
 PAH . . . . .polyaromatischer Kohlenwasserstoff  
 PIMO . . . . .Pair Interaction Molecular Orbital  
 PDB . . . . .Protein Data Bank  
 ROHF . . . . .Restricted Open Hartree-Fock  
 SAD . . . . .Superposition of Atomic Densities  
 SCF . . . . .Self-Consistent Field  
 STL . . . . .Standard Template Library  
 STO . . . . .Slaterförmiges Orbital  
 SIBFA . . . . .Sum of Interactions Between Fragments computed Ab initio  
 TB . . . . .Tight-Binding  
 TDSCF . . . . .Time Dependent Self-Consistent Field  
 TF . . . . .Thomas-Fermi-Theorie  
 UHF . . . . .Unrestricted Hartree-Fock  
 VMD . . . . .Visual Molecular Dynamics  
 XYZ . . . . .XYZ

# Index

- Adjazenzgraph, 78
- Alkan, 142
- Amdahlsches Gesetz, 152
- ANOVA, 30
- Anregung, 124
- Anregungsoperator, 124
- Anti-Frost-Protein, 202
- Antisymmetrisierer, 14–15
- Anzahl an Basisfunktionen, 55, 56, 60
- Anzahl der Elektronen, 35
- Anzahl der Kerne, 35
- Approximation
  - Kirkwood, 30
- Basis
  - lokal, 70, 100
- Basisfunktion, 53, 70
- Bestapproximation, 22, 24
  - Rang-1, 24
- Bindung
  - metallisch, 219
  - Van der Waals, 220
- Bindungsbruch, 159
- Bindungsgraph, 116
- Bindungsordnung, 117
- Born-Oppenheimer, 38
- Chelate, 219
- Coulomb-Potential, 107
- Dünnbesetztheit, 78–79
- Dekorrelator, 16, 25, 29–30
  - Teil-, 27–29
  - voll, 25
- Delokalisierung, 217–219
- Delokalisierungsenergie, 218
- Delta-Distribution, 39, 107
- Dichtematrix, 55
- Dirac-Gleichung, 46
- direkte Summe, 19
- Effizienz
  - parallele Effizienz, 152
- Eigenvektor
  - negativ, 78
- Eigenzerlegung, 78
- Elektronegativität, 217
- Elektronenstrukturproblem, 1, 45
- elektronische Wellenfunktion, 46
- Elektronortskordinate, 35
- Entwicklung
  - ANOVA, 30, 106
  - Korrelation, 29–30, 32
- Ethan, 143
- Fragment, 2, 99, 114, 118
- Fragmentansatz, 2, 130
- Funktion
  - antisymmetrisch, 13, 25
  - lokal, 67–69
    - Menge, 69–70
  - Produktstruktur, 23
  - symmetrisch, 13
  - zerlegbar, 22–24
- Funktionenraum, 17
- Gelfand=Tsetlin=Muster, 93, 199
- Gesamtelektronenanzahl, 55
- Gleichgewichtsgeometrie, 159
- Graphen, 142
- größenkonsistent, 117, 125–127
- Gruppe

symmetrische, 15  
 H=Matrizen, 80  
 Hamiltonoperator, 34–35  
   elektronischer, 49  
 Hartree-Fock, 51, 53  
 Hauptuntermatrix, 82  
 Indexmenge  
   Mächtigkeit, 79  
 Induktion  
   strukturell, 85, 88, 89  
 Innerer Produktraum, 22  
 Interferon, 203  
 Ionisierung  
   maximal, 218  
 Irreguläre Kugelflächenfunktion, 113  
 Kernladung, 35  
 Kernmasse, 35  
 Kernort, 35  
 Kombinationstechnik, 131  
 Komplex-Ionen, 218–219  
 Korrelation, 25  
 korreliert, 16  
 Kovalenzbindigkeit, 218  
 Kronecker-Delta, 47  
 Ladungsdichte, 55  
 Laufzeit, 152  
 lokalisiert  
   maximal, 75  
 Madelung-Konstante, 218  
 Maschinengenauigkeit, 67, 109  
 Methode  
   linear skalierend, 1  
 Modell  
   beinahe freie Elektronen, 219  
   freie Elektronen, 219  
 Moebius inversion, 30  
 Molekül, 144  
 Orbital  
   Gausstyp  
     kartesisch, 215  
     primitiv, 216  
     CGTO, 215–216  
     GTO, 215  
     STO, 215  
   Permutation, 13, 15  
   Potentialoperator, 34  
   Prion, 203  
   Produktfunktion  
     verallgemeinert, 227  
   Projektion, 82  
   Projektionsmatrix  
     kartesisch, 80  
   Prozesszahl, 152  
 Raum  
   Hilbert-, 41  
   quadratintegrierbare Funktionen, 42  
 Referenzzustand, 123, 124  
 Reguläre Kugelflächenfunktion, 113  
 Ring  
   aromatisch, 130  
 Scaleup, 152  
 Schildladung, 189  
 Schrödingergleichung, 34–35  
   elektronische, 38  
 Screening, 63, 103  
 Spinoperator, 46  
 Supermolekül, 230  
 Symmetrisierer, 14–15  
 Tensor, 16–17  
   zerlegbar, 17, 23  
 Tensorprodukt, 17, 19, 21  
   antisymmetrisch, 20–21  
   Funktionsraum, 22  
   symmetrisch, 20  
 Tensorrang, 23–24  
 Tensorraum, 17  
   Dimension, 17–18  
 Träger  
   kompakt, 64  
   lokal, 67, 70

unkorreliert, 15–16

verallgemeinerte Kernkoordinate, 35

Vier-Zentren-Integral, 56

Vorzeichen, 15

Wasser, 142

Wasserstoffbrückenbindung, 220

Wellenfunktion, 33

Zeitpunkt, 34

Zerlegung

    Ganzzahl, 28

    hierarchisch, 80

    Indexmenge, 25–26

    minimal überlappende, 84

Zustandsfunktion, 33–34



# Abbildungsverzeichnis

2.1. Schema eines hochdimensionalen Problems . . . . .	11
3.1. Korrelationsannahme, klassische Moleküldynamik . . . . .	36
3.2. Korrelationsannahme, BO-Moleküldynamik . . . . .	37
4.1. Skalierungsanalyse, Laufzeit und Basisgröße von HF . . . . .	62
5.1. Darstellungen zu Beweis von Lemma 6 . . . . .	71
5.2. Illustrationen zu Vier-Zentren-Integral . . . . .	73
6.1. Beispiel einer vollständigen Unterindexmengenhierarchie . . . . .	81
6.2. Unterindexmengenhierarchie, Unterteilung der Indexmengen . . . . .	86
6.3. Spektrum einer Tridiagonalmatrix, Gelfand-Tsetlin-Muster . . . . .	94
9.1. Illustration zur Wasserstoffsaturation . . . . .	136
9.2. Illustration, modifizierter Bindungsgraph bei Saturation . . . . .	136
9.3. Darstellung Graph zu Lemma 14 . . . . .	139
9.4. Skalierungsanalyse: Abbildung der Testsysteme . . . . .	141
(a). Alkan-Monomer . . . . .	141
(b). Wassergefülltes Gebiet ( $n = 3$ ) . . . . .	141
(c). Graphen ( $n = 1, 4$ Schichten) . . . . .	141
(d). Ethangefülltes Gebiet ( $n = 4$ ) . . . . .	141
9.5. Skalierungsanalyse, Graphenerstellung: Laufzeiten . . . . .	142
9.6. Skalierungsanalyse, Graphenanalyse&-zerlegung: Laufzeiten . . . . .	144
9.7. Skalierungsanalyse, Fragmentierung: Laufzeiten organische Moleküle . . . . .	146
9.8. Skalierungsanalyse, Fragmentierung: Laufzeiten einzelne Moleküle . . . . .	147
9.9. Skalierungsanalyse, Fragmentierung: Laufzeiten viele Moleküle . . . . .	147
9.10. Skalierungsanalyse, Fragmentierung: Laufzeiten Graphenschichten . . . . .	148
9.11. Skalierungsanalyse, Skalierbarkeit Server/Client: Laufzeiten . . . . .	150
9.12. Skalierungsanalyse, Fragmentberechnung: Laufzeiten organische Moleküle . . . . .	151
9.13. Skalierungsanalyse, Fragmentberechnung: Laufzeiten Wasser und Alkane . . . . .	152
9.14. Skalierungsanalyse, Parallelisierung: Wasser . . . . .	153
9.15. Skalierungsanalyse, Parallelisierung: Laufzeiten Alkane . . . . .	154
9.16. Skalierungsanalyse, Parallelisierung: Scaleup Alkane . . . . .	155
9.17. Skalierungsanalyse, Summation: Laufzeiten . . . . .	156
9.18. Skalierungsanalyse, Crossover-Punkt: Laufzeiten . . . . .	157
9.19. Skalierungsanalyse, Crossover-Punkt: Fragmentzahlen . . . . .	157

(a).	Laufzeit über Fragmente fester Ordnung $k$ . . . . .	157
(b).	Anzahl der Fragmente fester Ordnung $k$ . . . . .	157
9.20.	Bindungsgraph: Näherung Rechteckpotential . . . . .	159
9.21.	Bindungsgraph: Erzeugung bivalenter Systeme . . . . .	160
9.22.	Bindungsgraph: Bindungsgrad bivalenter Systeme . . . . .	161
9.23.	Approximationsanalyse: Genauigkeit endliche Basis . . . . .	165
9.24.	Approximationsanalyse: Alkane . . . . .	166
9.25.	Approximationsanalyse: Hexan, gestörte Kernkoordinaten . . . . .	168
9.26.	Approximationsanalyse: Alkene, Alkine . . . . .	168
9.27.	Approximationsanalyse: BN-Kette . . . . .	169
9.28.	Approximationsanalyse: Relativer Fehler der Energieanteile pro Anteil . . . . .	171
(a).	Kinetische Energie $E_{\text{kin}}$ . . . . .	171
(b).	Austauschenergie $E_{\text{XC}}$ . . . . .	171
(c).	Elektron-Elektron-Wechselwirkung $E_{\text{EE}}$ . . . . .	171
(d).	Elektron-Kern-Wechselwirkung $E_{\text{EN}}$ . . . . .	171
(e).	Kern-Kern-Wechselwirkung $E_{\text{NN}}$ . . . . .	171
(f).	Summierte Coulombenergie $E_{\text{Coul}}$ . . . . .	171
9.29.	Approximationsanalyse: Relativer Fehler der Energieanteile pro Molekül . . . . .	173
9.30.	Approximationsanalyse: Benzol . . . . .	174
9.31.	Approximationsanalyse: Krümmung Benzolfragmente . . . . .	174
9.32.	Approximationsanalyse: Organischer Testsatz . . . . .	175
9.33.	Darstellung des organischen Testsatzes, baumartige Systeme . . . . .	176
9.34.	Darstellung des organischen Testsatzes, Ringsysteme . . . . .	177
9.35.	Approximationsanalyse: Organischer Testsatz mit Zyklen . . . . .	178
9.36.	Approximationsanalyse: Organischer Testsatz Abfallverhalten . . . . .	178
9.37.	Approximationsanalyse: Organischer Testsatz, Energieanteile pro Fragment . . . . .	180
(a).	Biphenyl . . . . .	180
(b).	Melamin . . . . .	180
(c).	Prolin . . . . .	180
(d).	Adenin . . . . .	180
(e).	1,4-Dichlorbenzol . . . . .	180
(f).	Heptan . . . . .	180
9.38.	Bindungsgraph Biphenyl . . . . .	181
9.39.	Darstellung metallischer Testsysteme . . . . .	181
(a).	Naphthalen . . . . .	181
(b).	Coronen . . . . .	181
9.40.	Approximationsanalyse: Metallische Systeme . . . . .	182
9.41.	Approximationsanalyse: S22 Testsatz Monomere, Korrelationsenergie . . . . .	184
9.42.	Approximationsanalyse: S22 Testsatz Dimere, Korrelationsenergie . . . . .	185
9.43.	Illustrationen zu Mehrgitteransatz . . . . .	188
(a).	Geschachtelte Gitter . . . . .	188
(b).	Typischer V-Mehrgitter-Zyklus . . . . .	188
9.44.	Illustration, Aufspaltung der Ladungsdichte . . . . .	189
9.45.	Langreichweitiger Coulomb: Gittermaschenweite . . . . .	192



9.46. Langreichweitiger Coulomb: Splinehalbbreite, Kern-Kern-Anteil . . . . .	193
(a). Kern-Kern kurzreichweitig . . . . .	193
(b). Kern-Kern langreichweitig . . . . .	193
9.47. Langreichweitiger Coulomb: Splinehalbbreite . . . . .	194
9.48. Langreichweitiger Coulomb: Kernelektronen . . . . .	194
9.49. Langreichweitiger Coulomb: Gebietsgröße . . . . .	195
(d). BN-Kette $n = 3$ . . . . .	195
9.50. Langreichweitiger Coulomb: Hexan, BN-Kette . . . . .	196
9.51. Langreichweitiger Coulomb: Wasser, Skalierung mit Kernzahl . . . . .	197
9.52. Langreichweitiger Coulomb: S22 Testsatz, Coulombenergie . . . . .	199
9.53. Eigenwertspektrum: Hexan . . . . .	199
9.54. Eigenwertspektrum: Hexan, (un-)gestörte Kernkoordinaten . . . . .	201
9.55. Große Biomoleküle: Anti-Frost-Protein . . . . .	203
9.56. Große Biomoleküle: Prion . . . . .	204
9.57. Große Biomoleküle: Alpha-Interferon . . . . .	205
9.58. Große Biomoleküle: Interferon-Rezeptor . . . . .	206
A.1. Skizze, Zwei-Zentren-Gauss-Integral . . . . .	223



# Literaturverzeichnis

- [1] F. H. Allen. The cambridge structural database: a quarter of a million crystal structures and rising. *Acta Crystallographica*, B58:380–388, 2002.
- [2] J. Almlöf, K. Faegri Jr. und K. Korsell. Principles for a directSCF approach to LCAO-MO ab-initio calculations. *Journal of Computational Chemistry*, 3(3): 385–399, Januar 1982. doi: 10.1002/jcc.540030314.
- [3] Hans Wilhelm Alt. *Lineare Funktionalanalysis*. Springer-Verlag, Heidelberg, 6. Ausgabe, 2011. ISBN 978-3-642-22260-3. doi: 10.1007/978-3-642-22261-0.
- [4] Claudio Amovilli und R. McWeeny. Perturbation calculations of molecular interaction energies: an example, HF...HF. *Chemical Physics Letters*, 128(1):11–17, Juli 1986. doi: 10.1016/0009-2614(86)80137-3.
- [5] Claudio Amovilli und R. McWeeny. A matrix partitioning approach to the calculation of intermolecular potentials. General theory and some examples. *Chemical Physics*, 140(3):343–361, Februar 1990. doi: 10.1016/0301-0104(90)80002-F.
- [6] Claudio Amovilli, Ivo Cacelli, Silvio Campanile und Giacomo Prampolini. Calculation of the intermolecular energy of large molecules by a fragmentation scheme: Application to the 4-n-pentyl-4'-cyanobiphenyl (5CB) dimer. *Journal of Chemical Physics*, 117(7):3003–3012, 2002.
- [7] D. P. Anderson. BOINC: A System for Public-Resource Computing and Storage. In *Fifth IEEE/ACM International Workshop on Grid Computing*, Seiten 4–10. IEEE, 1999. ISBN 0-7695-2256-4. doi: 10.1109/GRID.2004.14.
- [8] George E. Andrews. *The Theory of Partitions*. Cambridge University Press, 1. Ausgabe, 1984. ISBN 0-521-63766-X.
- [9] Axel Arnold, Matthias Bolten, Holger Dachsels, Florian Fahrenberger, Franz Gähler, Rene Halver, Frederik Heber, Michael Hofmann, Christian Holm, Julian Iseringhausen, Ivo Kabadshow, Olaf Lenz, Michael Pippig, Daniel Potts und Godehard Sutmann. Comparison of scalable fast methods for long-range interactions. *Physical Review E*, 88(6):063308, Dezember 2013. doi: 10.1103/PhysRevE.88.063308.
- [10] Michael Artin. *Algebra*. Prentice Hall, Cambridge, USA, 1. Ausgabe, 1991.
- [11] Neil W. Ashcroft und David N. Mermin. *Festkörperphysik*. Oldenbourg Verlag, München Wien, 3. Ausgabe, 2007. ISBN 978-3-486-58273-4.

- [12] Richard F. W. Bader. Virial Field Relationship for Molecular Charge Distributions and the Spatial Partitioning of Molecular Properties. *The Journal of Chemical Physics*, 56(7):3320, 1972. doi: 10.1063/1.1677699.
- [13] Richard F. W. Bader. Molecular fragments or chemical bonds. *Accounts of Chemical Research*, 8(1):34–40, Januar 1975. doi: 10.1021/ar50085a005.
- [14] Roman M Balabin. Enthalpy difference between conformations of normal alkanes: Intramolecular basis set superposition error (BSSE) in the case of n-butane and n-hexane. *The Journal of chemical physics*, 129(16):164101, Oktober 2008. doi: 10.1063/1.2997349.
- [15] M. P. Barnett und C. A. Coulson. The Evaluation of Integrals Occurring in the Theory of Molecular Structure. Parts I & II. *Philosophical Transactions of the Royal Society of London. Series A*, 243(864):221–249, 1951.
- [16] Sidney W. Benson und Jerry H. Buss. Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. *The Journal of Chemical Physics*, 29(3):546, 1958. doi: 10.1063/1.1744539.
- [17] Michele Benzi und N. Razouk. Decay bounds and algorithms for approximating functions of sparse matrices. *Electronic Transactions on Numerical Analysis*, 28: 16–39, 2007.
- [18] Michele Benzi, Paola Boito und Nader Razouk. Decay properties of spectral projectors with applications to electronic structure. *SIAM Review*, 55:3–64, March 2013.
- [19] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov und P. E. Bourne. The protein data bank, 2000.
- [20] H. A. Bethe. Statistical Theory of Superlattices. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 150(871):552–575, Juli 1935. doi: 10.1098/rspa.1935.0122.
- [21] H. A. Bethe und J. G. Kirkwood. Critical Behavior of Solid Solutions in the Order-Disorder Transformation. *The Journal of Chemical Physics*, 7(8):578, 1939. doi: 10.1063/1.1750495.
- [22] Matthias Bolten. *Multigrid methods for structured grids and their application in particle simulation*. Dissertation, Bergische Universitaet Wuppertal, 2008.
- [23] S. F. Boys. Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 200(1063):542–554, Februar 1950. doi: 10.1098/rspa.1950.0036.

- [24] Mikio Ludwig Braun. *Spectral Properties of the Kernel Matrix and their Relation to Kernel Methods in Machine Learning*. Dissertation, Rheinische Friedrich-Wilhelms-Universität, 2005.
- [25] Andreas Bresinsky, Christian Körner, Joachim W. Kadereit, G. Neuhaus und Uwe Sonnewald. *Lehrbuch der Botanik*. Spektrum Akademischer Verlag, 36. Ausgabe, 2008. ISBN 978-3827414557.
- [26] I. N. Bronstein, K. A. Semendjajew, Gerhard Musiol und Heiner Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Frankfurt am Main, 7. Ausgabe, 2008.
- [27] Hans-Joachim Bungartz und Michael Griebel. Sparse grids. *Acta Numerica*, Seiten 1–123, 2004. doi: 10.1017/S0962492904000182.
- [28] R. M. Burstall. Proving properties of programs by structural induction. *The Computer Journal*, 12(1):41–48, Februar 1969. doi: 10.1093/comjnl/12.1.41.
- [29] Eric Cancès und Claude Le Bris. On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 34(4):749–774, April 2002. doi: 10.1051/m2an:2000102.
- [30] A. E. Carlsson. Beyond Pair Potentials in Elemental Transition Metals and Semiconductors. In H. Ehrenreich und D. Turnbull, Editoren, *Solid State Physics Vol. 43*, Seiten 1–91. Academic Press, San Diego, USA, 1. Ausgabe, 1990. ISBN 0-12-607743-6.
- [31] Hans Rudolf Christen. *Grundlagen der organischen Chemie*. Verlag Sauerlaender-Diesterweg-Salle, Aarau, Frankfurt am Main, 2. Ausgabe, 1972.
- [32] Hans Rudolf Christen. Chemische Bindung. In *Chemie - Atommodelle, Periodensystem, Chemische Bindung*, chapter 3, Seiten 65–137. Diesterweg - Salle - Sauerländer AG, Frankfurt am Main, Berlin, München, Aarau, 3. Ausgabe, 1972. ISBN 3-425-05381-7.
- [33] Ralph E. Christoffersen. Ab Initio Calculations on Large Molecules. In Per-Olov Löwdin, Editor, *Advances in Quantum Chemistry Vol. 6*, Seiten 333–393. Academic Press, 1972.
- [34] Ralph E. Christoffersen und Gerald M. Maggiora. Ab initio calculations on large molecules using molecular fragments. Preliminary investigations. *Chemical Physics Letters*, 3(6):419–423, Juni 1969. doi: 10.1016/0009-2614(69)80155-7.
- [35] E. Clementi und D. R. Davis. Electronic structure of large molecular systems. *Journal of Computational Physics*, 1(2):223–244, November 1966. doi: 10.1016/0021-9991(66)90004-0.
- [36] Enrico Clementi. Ab Initio Computations in Atoms and Molecules. *IBM Journal of Research and Development*, 9(1):2–19, Januar 1965. doi: 10.1147/rd.91.0002.

- [37] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest und Clifford Stein. *Introduction to Algorithms*. MIT Press, 2. Ausgabe, 2001. ISBN 0-262-03293-7.
- [38] J. J. M. Cuppen. A divide and conquer method for the symmetric tridiagonal eigenproblem. *Numerische Mathematik*, 36(2):177–195, 1981.
- [39] Jerome Dancis. On the Inertias of Symmetric Matrices and Bounded Self-Adjoint Operators. *Linear Algebra and its Applications*, 105:67–75, 1988.
- [40] Chandler Davis und Kahan W. M. Some new bounds on perturbation of subspaces. *Bull. Amer. Math. Soc*, 75(4):863–868, 1969.
- [41] Vin de Silva und Lek-heng Lim. Tensor Rank and the Ill-Posedness of the Best Low-Rank Approximation Problem. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1084–1127, Januar 2008. doi: 10.1137/06066518X.
- [42] M. Defranceschi und C. Le Bris. Computing a molecule: A mathematical viewpoint. *Journal of Mathematical Chemistry*, 21:1–30, 1997. doi: 10.1023/A:1019197613932.
- [43] Mireille Defranceschi und Claude Le Bris. Computing a molecule in its environment: A mathematical viewpoint. *International Journal of Quantum Chemistry*, 71(3):227–250, 1999. doi: 10.1002/(SICI)1097-461X(1999)71:3<227::AID-QUA2>3.0.CO;2-8.
- [44] P. Degand, G. Leroy und D. Peeters. L’utilisation d’orbitales localisées dans l’étude théorique des molécules. *Theoretica chimica acta*, 30(3):243–256, 1973.
- [45] B. Delley. Time dependent density functional theory with DMol3. *Journal of physics. Condensed matter : an Institute of Physics journal*, 22(38):384208, September 2010. doi: 10.1088/0953-8984/22/38/384208.
- [46] S. L. Dixon und K. M. Merz Jr. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *The Journal of Chemical Physics*, 107(3), 1997.
- [47] Roberto Dovesi, Roberto Orlando, Bartolomeo Civalleri, Carla Roetti, Victor R. Saunders und Claudio M. Zicovich-Wilson. CRYSTAL: a computational tool for the ab initio study of the electronic properties of crystals. *Zeitschrift für Kristallographie*, 220(5/6/2005), Januar 2005. doi: 10.1524/zkri.220.5.571.65065.
- [48] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007, 1989. doi: 10.1063/1.456153.
- [49] Volker Dyczmons. No  $N^4$ -dependence in the calculation of large molecules. *Theoretica Chimica Acta*, 28(3):307–310, 1973. doi: 10.1007/BF00533492.
- [50] Clyde Edmiston und Klaus Ruedenberg. Localized Atomic and Molecular Orbitals. *Reviews of Modern Physics*, 35(3):457–464, Juli 1963. doi: 10.1103/RevModPhys.35.457.

- [51] J. Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17: 449–467, 1965.
- [52] Stanley C. Eisenstat und Ilse C. F. Ipsen. Relative Perturbation Results for Eigenvalues and Eigenvectors of Diagonalizable Matrices. *BIT Numerical Mathematics*, 38(3):502–509, 1998. doi: 10.1007/BF02510256.
- [53] J.-L. Fattebert und F. Gygi. Linear-scaling first-principles molecular dynamics with plane-waves accuracy. *Physical Review B*, 73(11), March 2006. doi: 10.1103/PhysRevB.73.115124.
- [54] Dmitri G. Fedorov und Kazuo Kitaura. The importance of three-body terms in the fragment molecular orbital method. *The Journal of chemical physics*, 120(15): 6832–40, April 2004. doi: 10.1063/1.1687334.
- [55] Richard Feynman. *Surely you're joking, Mr. Feynman*. W. W. Norton, 1985. ISBN 0-393-01921-7.
- [56] M. Finnis. Bond-order potentials through the ages. *Progress in Materials Science*, 52(2-3):133–153, Februar 2007. doi: 10.1016/j.pmatsci.2006.10.003.
- [57] N. Flocke und Rodney J. Bartlett. A natural linear scaling coupled-cluster method. *The Journal of chemical physics*, 121(22):10935–44, Dezember 2004. doi: 10.1063/1.1811606.
- [58] V. Fock. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2):126–148, Januar 1930. doi: 10.1007/BF01340294.
- [59] Leslie L. Foldy. Antisymmetric Functions and Slater Determinants. *Journal of Mathematical Physics*, 3(3):531, 1962. doi: 10.1063/1.1724251.
- [60] J. Foster und S. F. Boys. Canonical Configurational Interaction Procedure. *Reviews of Modern Physics*, 32(2):300–302, April 1960. doi: 10.1103/RevModPhys.32.300.
- [61] William Fulton. Eigenvalues, invariant factors, highest weights, and Schubert calculus. *Bulletin of the American Mathematical Society*, Seiten 1–42, 2000.
- [62] Jochen Garcke. An optimised sparse grid combination technique for eigenproblems. *Proceedings in Applied Mathematics and Mechanics*, 7(1):1022301–1022302, Dezember 2007. doi: 10.1002/pamm.200700261.
- [63] Jochen Garcke und Michael Griebel. On the Computation of the Eigenproblems of Hydrogen and Helium in Strong Magnetic and Electric Fields with the Sparse Grid Combination Technique. *Journal of Computational Physics*, 165(2):694–716, Dezember 2000. doi: 10.1006/jcph.2000.6627.

- [64] J. Gasteiger und C. Jochum. An algorithm for the perception of synthetically important rings. *Journal of Chemical Information and Computer Sciences*, 19(1): 43–48, 1979.
- [65] J. R. Gilbert. Predicting structure in sparse matrix computations. *SIAM Journal on Matrix Analysis and Applications*, 1994.
- [66] Peter M. W. Gill, Benny G. Johnson und John A. Pople. A simple yet powerful upper bound for Coulomb integrals. *Chemical Physics Letters*, 217(1-2):65–68, Januar 1994. doi: 10.1016/0009-2614(93)E1340-M.
- [67] Stefan Goedecker. Linear scaling electronic structure methods. *Reviews of Modern Physics*, 71(4):1085–1123, Juli 1999. doi: 10.1103/RevModPhys.71.1085.
- [68] D. S. Goodsell. Prions. *RCSB Protein Data Bank*, Mai 2008. doi: 10.2210/rcsb\\_pdb/mom\\_2008\\_5.
- [69] D. S. Goodsell. Antifreeze Proteins. *RCSB Protein Data Bank*, Dezember 2009. doi: 10.2210/rcsb\\_pdb/mom\\_2009\\_12.
- [70] D. S. Goodsell. Interferons. *RCSB Protein Data Bank*, August 2010. doi: 10.2210/rcsb\\_pdb/mom\\_2010\\_8.
- [71] Lars Grasedyck. *Theorie und Anwendungen Hierarchischer Matrizen*. Doktorarbeit, Christian-Albrechts-Universität Kiel, 2001.
- [72] L. Greengard und V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73(2):325–348, 1987.
- [73] Douglas Gregor und Andrew Lumsdaine. The Parallel BGL: A generic library for distributed graph computations. In *Parallel Object-Oriented Scientific Computing (POOSC)*, Seiten 1–18, 2005.
- [74] R. Gregory. UNIFORM BOUNDS FOR  $P_{nm}(\cos\Theta)$  AND THE ABSOLUTE CONVERGENCE OF SERIES EXPANSIONS IN SPHERICAL SURFACE HARMONICS. *The Quarterly Journal of Mechanics and Applied Mathematics*, 50(3):467–479, August 1997. doi: 10.1093/qjmam/50.3.467.
- [75] Nohad Gresh, Pierre Claverie und Alberte Pullman. Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with ab initio SCF computations. *Theoretica Chimica Acta*, 66:1–20, 1984.
- [76] Michael Griebel. Sparse grids and related approximation schemes for higher dimensional problems. In J. Pardo, A. Pinkus, E. Suli und M. J. Todd, Editoren, *Foundations of Computational Mathematics (FoCM05)*, Santander, Seiten 106–161. Cambridge University Press, 2006.



- [77] Michael Griebel, Stephan Knapek und Gerhard Zumbusch. *Numerical Simulation in Molecular Dynamics – Numerics, Algorithms, Parallelization, Applications*. Springer-Verlag, Heidelberg, 1. Ausgabe, 2007.
- [78] Michael Griebel, Jan Hamaekers und Frederik Heber. A bond order dissection approach for efficient electronic structure calculations. In *Abschlussband SPP1324*, submitted, 2014. Springer-Verlag.
- [79] Wolfgang Hackbusch. *Multi-Grid Methods and Applications*. Springer Berlin Heidelberg, 1. Ausgabe, 1980. ISBN 3-540-12761-5.
- [80] Wolfgang Hackbusch. *Tensor Space and Numerical Tensor Calculus*. Springer-Verlag, Berlin, Heidelberg, 1. Ausgabe, 2012. ISBN 978-3-642-28026-9.
- [81] G. G. Hall. The Molecular Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 205(1083):541–552, March 1951. doi: 10.1098/rspa.1951.0048.
- [82] J. Hamaekers. *Sparse Grids for the Electronic Schroedinger Equation: Construction and Application of Sparse Tensor Product Multiscale Many-Particle Spaces with Finite-Order Weights for Schroedinger’s Equation*. Suedwestdeutscher Verlag fuer Hochschulschriften, Saarbruecken, 2010. ISBN 978-3-8381-1578-8.
- [83] Donovan R. Hare, Charles R. Johnson, D. D. Olesky und P. van den Driessche. Sparsity Analysis of the QR Factorization. 1991.
- [84] Harold Hart, Leslie E. Craine, David J. Hart und Christopher M. Hadad. *Organische Chemie*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 3. Ausgabe, 2007. ISBN 978-3-527-31801-8.
- [85] D. R. Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(01):89, Oktober 1928. doi: 10.1017/S0305004100011919.
- [86] D. R. Hartree. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(01):111, Oktober 1928. doi: 10.1017/S0305004100011920.
- [87] Marco Häser und Reinhart Ahlrichs. Improvements on the direct SCF method. *Journal of Computational Chemistry*, 10(1):104–111, Januar 1989. doi: 10.1002/jcc.540100111.
- [88] Emilie V. Haynsworth. Determination of the inertia of a partitioned Hermitian matrix. *Linear Algebra and its Applications*, 1(1):73–81, Januar 1968. doi: 10.1016/0024-3795(68)90050-5.

- [89] Martin Head-Gordon, Michael Lee, Paul Maslen, Troy van Voorhis und Steven Gwaltney. Tensors in Electronic Structure Theory: Basic Concepts and Applications to Electron Correlation Models. In Johannes Grotendorst, Editor, *Modern Methods and Algorithms in Quantum Chemistry*, Seiten 593–638. John von Neumann Institute for Computing, Jülich, 2000.
- [90] Frederik Heber. *Dichtefunktionaltheorie mit elektromagnetischen Effekten und ihre Anwendung*. Diplomarbeit, Universität Bonn, 2006.
- [91] Trygve Helgaker, Poul Jorgensen und Jeppe Olsen. *Molecular Electronic Structure-Theory*. John Wiley & Sons, West Sussex, England, 1. Ausgabe, 2000. ISBN 0-471-96755-6.
- [92] I. N. Herstein. *Topics in Algebra*. John Wiley & Sons, New York, 2. Ausgabe, 1975.
- [93] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta*, 44(2):129–138, 1977. doi: 10.1007/BF00549096.
- [94] Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948.
- [95] P. Hohenberg und Walter Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, 1964. doi: 10.1103/PhysRev.136.B864.
- [96] Markus Holtz. *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*. Dissertation, University Bonn, 2008.
- [97] Abdolhossein Hoorfar und Mehdi Hassani. Inequalities on the Lambert W Function and Hyperpower Function. *Journal of Inequalities in Pure and Applied Mathematics*, 9(2):1–5, 2008.
- [98] Roger A. Horn und Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1. Ausgabe, 1985.
- [99] A. P. Horsfield, A. M. Bratkovsky, D. G. Pettifor und M. Aoki. Bond-order potential and cluster recursion for the description of chemical bonds: Efficient real-space methods for tight-binding molecular dynamics. *Physical Review B*, 53(3):1656–1666, 1996.
- [100] Q. N. Hu, Y. Z. Liang und K. T. Fang. The matrix expression, topological index and atomic attribute of molecular topological structure. *Journal of Data Science*, 1:361–389, 2003.
- [101] William Humphrey, Andrew Dalke und Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, Februar 1996. doi: 10.1016/0263-7855(96)00018-5.

- [102] A. C. Hurley, John Lennard-Jones und J. A. Pople. The molecular orbital theory of chemical valency XVI . A theory of paired-electrons in polyatomic molecules. *Society*, 220(1143):446–455, 1953.
- [103] Ilse C. F. Ipsen. Relative perturbation results for matrix eigenvalues and singular values. *Acta Numerica*, 7:151, November 2008. doi: 10.1017/S0962492900002828.
- [104] Julian Iseringhausen. *Versatile MultiGrid - a highly scalable elliptic PDE solver for periodic, quasi-periodic, and open boundary conditions*. Diploma thesis, University Bonn, 2013.
- [105] Toshimasa Ishida und Koichi Ohno. On the asymptotic behavior of Hartree-Fock orbitals. *Theoretica Chimica Acta*, 81(6):355–364, 1992. doi: 10.1007/BF01134860.
- [106] John David Jackson. *Klassische Elektrodynamik*. Walter de Gruyter, Berlin, New York, 4. Ausgabe, 2006.
- [107] Leif D. Jacobson und John M. Herbert. An efficient, fragment-based electronic structure method for molecular systems: self-consistent polarization with perturbative two-body exchange and dispersion. *The Journal of chemical physics*, 134(9):094118, March 2011. doi: 10.1063/1.3560026.
- [108] Curtis L. Janssen, Ida B. Nielsen, Matt L. Leininger, Edward F. Valeev, Joseph P. Kenny und Edward T. Seidl. The Massively Parallel Quantum Chemistry Program (MPQC), Version 2.3.1, 2008.
- [109] C. R. Johnson und B. K. Kroschel. Principal submatrices, geometric multiplicities, and structured eigenvectors. *SIAM journal on matrix analysis and applications*, 16(3):1004–1012, 1995.
- [110] Petr Jurecka, Jirí Šponer, Jirí Cerný und Pavel Hobza. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Physical chemistry chemical physics : PCCP*, 8(17):1985–93, Mai 2006. doi: 10.1039/b600027d.
- [111] Ryoichi Kikuchi. A Theory of Cooperative Phenomena. II. Equation of States for Classical Statistics. *The Journal of Chemical Physics*, 19(10):1230, 1951. doi: 10.1063/1.1748002.
- [112] Ryoichi Kikuchi. A Theory of Cooperative Phenomena. *Physical Review*, 81(6):988–1003, March 1951. doi: 10.1103/PhysRev.81.988.
- [113] John G. Kirkwood. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics*, 3(5):300, 1935. doi: 10.1063/1.1749657.
- [114] John G. Kirkwood und Elizabeth Monroe Boggs. The Radial Distribution Function in Liquids. *The Journal of Chemical Physics*, 10(6):394–402, 1942. doi: 10.1063/1.1723737.

- [115] K. Kitaura, Takuya Sawai, Toshio Asada, Tatsuya Nakano und Masami Uebayasi. Pair interaction molecular orbital method: an approximate computational method for molecular interactions. *Chemical Physics Letters*, 312(2-4):319–324, Oktober 1999. doi: 10.1016/S0009-2614(99)00937-9.
- [116] Kazuo Kitaura, Eiji Ikeo, Toshio Asada, Tatsuya Nakano und Masami Uebayasi. Fragment molecular orbital method: an approximate computational method for large molecules. *Chemical Physics Letters*, 313:701–706, 1999.
- [117] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin Heidelberg, 3. Ausgabe, 2013. ISBN 978-3-642-36017-6. doi: 10.1007/978-3-642-36018-3.
- [118] M. Klessinger und R. McWeeny. Self-Consistent Group Calculations on Polyatomic Molecules. I. Basic Theory with an Application to Methane. *The Journal of Chemical Physics*, 42(10):3343, 1965. doi: 10.1063/1.1695733.
- [119] Allen Knutson und Terence Tao. Honeycombs and sums of Hermitian matrices. *Notices of the AMS*, 1:1–16, 2001.
- [120] Jorge Kohanoff. *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods*. Cambridge University Press, 2006.
- [121] W. Kohn. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Physical Review Letters*, 76(17):3168–3171, April 1996. doi: 10.1103/PhysRevLett.76.3168.
- [122] Tamara G. Kolda und Brett W. Bader. Tensor Decompositions and Applications. *SIAM Review*, 51(3):455, 2009. doi: 10.1137/07070111X.
- [123] Konrad Königsberger. *Analysis 2*. Springer Berlin Heidelberg, 2. Ausgabe, 1997.
- [124] Ilya Krishtal, Thomas Strohmer und Tim Wertz. Localization of Matrix Factorizations. Mai 2013.
- [125] Paul Labute. On the perception of molecules from 3D atomic coordinates. *Journal of chemical information and modeling*, 45(2):215–21, 2005. doi: 10.1021/ci049915d.
- [126] Ira N. Levine. *Quantum Chemistry*. Prentice Hall, New Jersey, 5. Ausgabe, 2000. ISBN 0-13-685512-1.
- [127] X. P. Li, R. W. Nunes und D. Vanderbilt. Density-matrix electronic-structure method with linear system-size scaling. *Physical Review B*, 47(16):10891–10894, 1993.
- [128] P. L. Lions. Solutions of Hartree-Fock equations for Coulomb systems. *Communications in Mathematical Physics*, 109(1):33–97, March 1987. doi: 10.1007/BF01205672.

- [129] Alexander D. MacKerell, D. Bashford, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin und M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, April 1998.
- [130] Gerald M. Maggiora, Donald W. Genson, Ralph E. Christoffersen und B. Vernon Cheney. Ab initio calculations on large molecules using molecular fragments First order electronic properties for hydrocarbons. *Theoretica Chimica Acta*, 22(4):337–352, 1971. doi: 10.1007/BF01036037.
- [131] Walter F. Mascarenhas. The structure of the eigenvectors of sparse matrices. *Linear Algebra and its Applications*, 207:1–20, August 1994. doi: 10.1016/0024-3795(94)90001-9.
- [132] P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee und M. Head-Gordon. Locality and Sparsity of Ab Initio One-Particle Density Matrices and Localized Orbitals. *The Journal of Physical Chemistry A*, 102(12):2215–2222, March 1998. doi: 10.1021/jp972919j.
- [133] R. McWeeny. The Density Matrix in Many-Electron Quantum Mechanics. I. Generalized Product Functions. Factorization and Physical Interpretation of the Density Matrices. *Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences*, 253(1273):242–259, 1959.
- [134] R. McWeeny. Some Recent Advances in Density Matrix Theory. *Reviews of Modern Physics*, 32(2):335–369, 1960.
- [135] R. McWeeny. *Methods of Molecular Quantum Mechanics*. Academic Press, 2. Ausgabe, 1992.
- [136] C. Møller und M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review*, 46(7):618–622, Oktober 1934. doi: 10.1103/PhysRev.46.618.
- [137] Martin J. Mohlenkamp. Capturing the Interelectron Cusp Using a Geminal Layer on an Unconstrained Sum of Slater Determinants. *SIAM Journal on Applied Mathematics*, 72(6):1742–1771, November 2012. doi: 10.1137/110823900.
- [138] Martin J. Mohlenkamp. Musings on multilinear fitting. *Linear Algebra and its Applications*, 438(2):834–852, Januar 2013. doi: 10.1016/j.laa.2011.04.019.
- [139] Jens Oettershagen. *Reduktion der effektiven Dimension und ihre Anwendung auf hochdimensionale Probleme*. Diploma, University Bonn, 2011.
- [140] Robert G. Parr, Robert A. Donnelly, Mel Levy und William E. Palke. Electronegativity: The density functional viewpoint. *The Journal of Chemical Physics*, 68(8):3801, 1978. doi: 10.1063/1.436185.

- [141] Robert G. Parr, Paul W Ayers und Roman F. Nalewajski. What is an atom in a molecule? *The journal of physical chemistry. A*, 109(17):3957–9, Mai 2005. doi: 10.1021/jp0404596.
- [142] B. Paulus. The method of increments—a wavefunction-based ab initio correlation method for solids. *Physics Reports*, 428(1):1–52, Mai 2006. doi: 10.1016/j.physrep.2006.01.003.
- [143] János Pipek und Paul G. Mezey. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *The Journal of Chemical Physics*, 90(9):4916, 1989. doi: 10.1063/1.456588.
- [144] Jay W. Ponder und David A. Case. Force Fields for Protein Simulation. *Advances in Protein Chemistry*, 66:27–85, 2003. doi: 10.1016/S0065-3233(03)66002-X.
- [145] Alex Pothen. Predicting the structure of sparse orthogonal factors. *Linear Algebra and its Applications*, 194:183–203, November 1993. doi: 10.1016/0024-3795(93)90121-4.
- [146] Herschel Rabitz und Omer F. Aliş. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25:197–233, 1999.
- [147] Erwin Riedel und Christoph Janiak. *Anorganische Chemie*. Walter de Gruyter GmbH & Co. KG, Berlin/New York, 8. Ausgabe, 2011. ISBN 978-3-11-022566-2.
- [148] C. C. J. Roothaan. New Developments in Molecular Orbital Theory. *Reviews of Modern Physics*, 23(2):69–89, April 1951. doi: 10.1103/RevModPhys.23.69.
- [149] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2. Ausgabe, 2003. ISBN 978-0-898715-34-7.
- [150] E. Schrödinger. 2. Quantisierung als Eigenwertproblem. *Annalen der Physik*, 384(4):361–376, 1926. doi: 10.1002/andp.19263840404.
- [151] Franz Schwabl. *Quantenmechanik*. Springer Berlin Heidelberg, 7. Ausgabe, 2007. ISBN 973-3-540-73674-5.
- [152] Yihan Shao, Laszlo Fusti Molnar, Yousung Jung, Jörg Kussmann, Christian Ochsenfeld, Shawn T. Brown, Andrew T. B. Gilbert, Lyudmila V. Slipchenko, Sergey V. Levchenko, Darragh P. O’Neill, Robert A. DiStasio, Rohini C. Lochan, Tao Wang, Gregory J. O. Beran, Nicholas A. Besley, John M. Herbert, Ching Yeh Lin, Troy Van Voorhis, Siu Hung Chien, Alex Sodt, Ryan P. Steele, Vitaly A. Rassolov, Paul E. Maslen, Prakashan P. Korambath, Ross D. Adamson, Brian Austin, Jon Baker, Edward F. C. Byrd, Holger Dachsel, Robert J. Doerksen, Andreas Dreuw, Barry D. Dunietz, Anthony D. Dutoi, Thomas R. Furlani, Steven R. Gwaltney, Andreas Heyden, So Hirata, Chao-Ping Hsu, Gary Kedziora, Rustam Z. Khallulin, Phil Klunzinger, Aaron M. Lee, Michael S. Lee, Wanzhen Liang, Itay Lotan,

- Nikhil Nair, Baron Peters, Emil I. Proynov, Piotr A. Pieniazek, Young Min Rhee, Jim Ritchie, Edina Rosta, C. David Sherrill, Andrew C. Simmonett, Joseph E. Subotnik, H. Lee Woodcock, Weimin Zhang, Alexis T. Bell, Arup K. Chakraborty, Daniel M. Chipman, Frerich J. Keil, Arieh Warshel, Warren J. Hehre, Henry F. Schaefer, Jing Kong, Anna I. Krylov, Peter M. W. Gill und Martin Head-Gordon. Advances in methods and algorithms in a modern quantum chemistry program package. *Physical chemistry chemical physics : PCCP*, 8(27):3172–91, Juli 2006. doi: 10.1039/b517914a.
- [153] I. Shavitt und M. Karplus. Gaussian-Transform Method for Molecular Integrals. I. Formulation for Energy Integrals. *The Journal of Chemical Physics*, 43(2):398, 1965. doi: 10.1063/1.1696757.
- [154] Lester L. Shipman und Ralph E. Christoffersen. Ab initio calculations on large molecules using molecular fragments. Initial P-matrices for scf calculations. *Chemical Physics Letters*, 15(4):469–474, September 1972. doi: 10.1016/0009-2614(72)80351-8.
- [155] Jeremy G Siek, Lie-Quan Lee und Andrew Lumsdaine. *The Boost Graph Library*. Addison-Wesley Professional, 1. Ausgabe, 2001. ISBN 978-0-201-72914-6.
- [156] M. Sironi, A. Famulari, M. Raimondi und S. Chiesa. The transferability of extremely localized molecular orbitals. *Journal of Molecular Structure: THEOCHEM*, 529(1-3):47–54, September 2000. doi: 10.1016/S0166-1280(00)00530-3.
- [157] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi und Mike C. Payne. Introducing ONETEP: linear-scaling density functional simulations on parallel computers. *The Journal of chemical physics*, 122(8):84119, 2005. doi: 10.1063/1.1839852.
- [158] J. C. Slater. Note on Hartree’s Method. *Physical Review*, 35(2):210–211, Januar 1930. doi: 10.1103/PhysRev.35.210.2.
- [159] J. C. Slater. A Simplification of the Hartree-Fock Method. *Physical Review*, 81(3):385–390, Februar 1951. doi: 10.1103/PhysRev.81.385.
- [160] Sandeep Somani, Benjamin J. Killian und Michael K. Gilson. Sampling conformations in high dimensions using low-dimensional distribution functions. *The Journal of chemical physics*, 130(13):134102, April 2009. doi: 10.1063/1.3088434.
- [161] Volker Staemmler. Introduction to Hartree-Fock and CI Methods. In J. Groten-dorst, S. Blügel und D. Marx, Editoren, *Computational Nanoscience: Do It Yourself!*, volume 31, Seiten 1–18. John von Neumann Institute for Computing, Jülich, 2006. ISBN 3-00-.017350-1.
- [162] Andreas Stathopoulos, Yousef Saad und Charlotte F. Fischer. Overlapping Domain Decomposition Preconditioners for the Generalized Davidson Method for the Eigenvalue Problem. 1994.

- [163] Hermann Stoll. Correlation energy of diamond. *Physical Review B*, 46(11):6700–6704, September 1992. doi: 10.1103/PhysRevB.46.6700.
- [164] Matthew C. Strain, Gustavo E. Scuseria und Michael J. Frisch. Achieving Linear Scaling for the Electronic Quantum Coulomb Problem. *Science*, 271(5245):51–53, Januar 1996. doi: 10.1126/science.271.5245.51.
- [165] D. L. Strout und G. E. Scuseria. A quantitative study of the scaling properties of the Hartree-Fock method. *The Journal of Chemical Physics*, 102(June):8448–8452, 1995.
- [166] Joseph E. Subotnik, Anthony D. Dutoi und Martin Head-Gordon. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *The Journal of Chemical Physics*, 123(11):114108, 2005. doi: 10.1063/1.2033687.
- [167] P. Surján. An Introduction to the Theory of Geminals. In Péter R. Surján, R. J. Bartlett, F. Bogár, D. L. Cooper, B. Kirtman, W. Klopper, W. Kutzelnigg, N. H. March, P. G. Mezey, H. Müller, J. Noga, J. Paldus, J. Pipek, M. Raimondi, I. Røeggen, J. Q. Sun, P. R. Surján, C. Valdemoro und S. Vogtner, Editoren, *Correlation and Localization*, volume 203 of *Topics in Current Chemistry*, Seiten 255–272. Springer, Berlin, Heidelberg, Juli 1999. ISBN 978-3-540-65754-5. doi: 10.1007/3-540-48972-X\\_4.
- [168] Attila Szabo und Neil S. Ostlund. *Modern Quantum Chemistry - Introduction to Advanced Electronic Structure Theory*. Dover Publications, Inc., New York, 1989. ISBN 978-0-486-69186-2.
- [169] Ellad B. Tadmor und Ronald E. Miller. *Modeling Materials - Continuum, Atomistic and Multiscale Techniques*. Cambridge University Press, Cambridge, UK, 1. Ausgabe, 2011. ISBN 978-0-521-85698-0.
- [170] Terence Tao. *Topics in random matrix theory*. American Mathematical Society, 2012. ISBN 978-0821874301.
- [171] Robert Endre Tarjan. *Data Structures and Network Algorithms*. Society for Industrial and Applied Mathematics, 1983. ISBN 978-0898711875.
- [172] U. Trottenberg, C. W. Oosterlee und A. Schuller. *Multigrid*. Academic Press, 2000.
- [173] John C. Tully. Nonadiabatic Dynamics. In Donald L Thompson, Editor, *Modern Methods for Multidimensional Dynamics Computations in Chemistry*, chapter 2, Seiten 34–72. World Scientific, 1. Ausgabe, 1998. ISBN 981-02-3342-6.
- [174] Justin M. Turney, Andrew C. Simmonett, Robert M. Parrish, Edward G. Hohenstein, Francesco A. Evangelista, Justin T. Fermann, Benjamin J. Mintz, Lori A. Burns, Jeremiah J. Wilke, Micah L. Abrams, Nicholas J. Russ, Matthew L. Leininger, Curtis L. Janssen, Edward T. Seidl, Wesley D. Allen, Henry F. Schaefer, Rollin A. King, Edward F. Valeev, C. David Sherrill und T. Daniel Crawford. Psi4: an



- open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(4):556–565, Juli 2012. doi: 10.1002/wcms.93.
- [175] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus und W. A. de Jong. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications*, 181(9):1477–1489, September 2010. doi: 10.1016/j.cpc.2010.04.018.
- [176] J. H. Van Lenthe, R. Zwaans, H. J. J. Van Dam und M. F. Guest. Starting SCF calculations by superposition of atomic densities. *Journal of computational chemistry*, 27(8):926–32, Juni 2006. doi: 10.1002/jcc.20393.
- [177] W. von Niessen. A theory of molecules in molecules. *Theoretica chimica acta*, 31(2):111–135, 1973.
- [178] Thom Vreven und Keiji Morokuma. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) Method. *Journal of Computational Chemistry*, 21(16):1419–1432, 2000.
- [179] P. Duane Walker und Paul G. Mezey. Molecular electron density Lego approach to molecule building. *Journal of the American Chemical Society*, 115(26):12423–12430, Dezember 1993. doi: 10.1021/ja00079a025.
- [180] Junmei Wang, Wei Wang, Peter A. Kollman und David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling*, 25(2):247–60, Oktober 2006. doi: 10.1016/j.jmgm.2005.12.005.
- [181] T Washio und C W Oosterlee. Error analysis for a potential problem on locally refined grids. *Numerische Mathematik*, 86:539–563, 2000. doi: 10.1007/s002110000189.
- [182] S. Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of research and development*, 4(1):66–82, 1960.
- [183] Ralf Wesendrup, Jon K. Laerdahl und Peter Schwerdtfeger. Relativistic effects in gold chemistry. VI. Coupled cluster calculations for the isoelectronic series AuPt[<sup>sup -</sup>], Au[<sub>sub 2</sub>], and AuHg[<sup>sup +</sup>]. *The Journal of Chemical Physics*, 110(19):9457, 1999. doi: 10.1063/1.478911.
- [184] Christopher A. White, Benny G. Johnson, Peter M. W. Gill und Martin Head-Gordon. The continuous fast multipole method. *Chemical Physics Letters*, 230(1-2):8–16, November 1994. doi: 10.1016/0009-2614(94)01128-1.
- [185] Max Wolfsberg und Lindsay Helmholz. The Spectra and Electronic Structure of the Tetrahedral Ions MnO<sub>4</sub><sup>-</sup>, CrO<sub>4</sub><sup>-</sup>, and ClO<sub>4</sub><sup>-</sup>. *The Journal of Chemical Physics*, 20(5):837, 1952. doi: 10.1063/1.1700580.

- [186] Weitao Yang. Direct calculation of electron density in density-functional theory. *Physical review letters*, 66(11):1438–1441, 1991.
- [187] Harry Yserentant und Hans-Christian Kreusler. The mixed regularity of electronic wave functions in fractional order and weighted Sobolev spaces. *Numerische Mathematik*, 121(4):781–802, 2012.
- [188] Arno Zaddach. *Grassmanns Algebra in der Geometrie*. BI-Wissenschaftsverlag, Mannheim, 1. Ausgabe, 1994. ISBN 3-411-17031-X.
- [189] Da W. Zhang, Yun Xiang und John Z H Zhang. New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin-Biotin Interaction Energy. *Journal of Physical Chemistry B*, 107:12039–12041, 2003.
- [190] Fuzhen Zhang. Hermitian Matrices. In S Axler, F W Gehring und K A Ribet, Editoren, *Matrix Theory - Basic Results and Techniques*, chapter 7, Seiten 208–239. Springer-Verlag, New York, 1. Ausgabe, 1999. ISBN 0-387-98696-0.
- [191] Qian Zhang, Wei Zhang, Youyong Li, Junmei Wang, Liling Zhang und Tingjun Hou. A rule-based algorithm for automatic bond type perception. *Journal of cheminformatics*, 4(1):26, Januar 2012. doi: 10.1186/1758-2946-4-26.