

**Berechnung von Stufenkanten
auf der As-bedeckten Si(111)-Oberfläche
mit einem parallelisierten *ab initio* Programm**

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von
Diplom-Physiker
Ralf Berger
aus Warstein

Berichter: apl.Prof. Dr.rer.nat. K. Schroeder
Univ.-Prof. Dr.rer.nat. P.H. Dederichs

Tag der mündlichen Prüfung: 23. September 2002

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online
verfügbar.

Zusammenfassung

Stufenkanten auf der As-bedeckten Si(111)-Oberfläche

Im ersten Teil der Arbeit wird die Parallelisierung des **ES**t**Co**MPP¹-Simulationsprogrammes vorgestellt. Damit kann man im Rahmen der Dichtefunktionaltheorie (DFT) Grundzustandseigenschaften von Vielelektronensystemen berechnen, wie sie z.B. in Molekülen und Festkörpern auftreten, wobei inhomogene Systeme durch eine periodische Anordnung von Superzellen approximiert werden. Im Rahmen der DFT wird die Wechselwirkung der Elektronen durch ein dichteabhängiges Einteilchen-Potential berücksichtigt, das selbstkonsistent bestimmt wird. Die Wechselwirkung der Valenzelektronen mit den Ionenrümpfen wird durch normerhaltende Pseudopotentiale beschrieben. Die Elektronenzustände ergeben sich numerisch aus der Lösung eines Eigenwertproblems, welches für einen Satz von \mathbf{k} -Punkten gelöst wird, um jeweils die $\nu = 1 \dots M$ energetisch niedrigsten Eigenzustände zu finden. Die Eigenzustände werden durch Linearkombination ebener Wellen zu den Vektoren \mathbf{G} des reziproken Gitters dargestellt. Entsprechend den drei Indizes \mathbf{k} , ν und \mathbf{G} der Koeffizienten der Wellenfunktionen werden Parallelisierungsstrategien betrachtet, die jeweils Teilbereiche des entsprechenden Index einzelnen Prozessoren zuweisen. Die \mathbf{k} -Punkt-Parallelisierung ist sehr effizient, da die Eigenwertprobleme zu verschiedenen \mathbf{k} -Punkten unabhängig voneinander lösbar sind, so daß nur 0.2% der Rechenzeit auf nicht parallele Programmteile entfallen. Bei der zusätzlich implementierten \mathbf{G} -Vektor-Parallelisierung läßt sich ein günstiges Skalierungsverhalten bei großen Systemen beobachten, das den Einsatz vieler Prozessoren erlaubt. Bei einer Rechnung mit 110 Atomen werden z.B. weniger als 0.4% der Rechenzeit für nicht parallelisierte Programmteile benötigt.

Im zweiten Teil der Arbeit werden Stufenkanten auf der As-bedeckten Si(111)-Oberfläche mit dem **ES**t**Co**MPP-Programm untersucht. Gesamtenergien und Strukturen von verschiedenen metastabilen Zuständen werden durch Relaxation vorgegebener Startkonfigurationen gefunden. Die je nach Struktur der Stufenkante unterschiedliche As-Bedeckung der Oberfläche führt zu einer Abhängigkeit der Kantenenergien vom chemischen Potential für As, das als Referenzenergie in die Rechnung eingeht. Sowohl für As-Bulk, als auch für As_4 als Referenz ergibt sich die $(11\bar{2})$ -Stufenkante, bei der alle Si-Kantenatome durch As-Atome ersetzt sind, als stabilste Stufenkante.

Mit einem zweidimensionalen Wulff-Plot kann mittels der gefundenen Kantenenergien die energetisch günstigste Inselform für Si-Homoepitaxie vorhergesagt werden. Für eine Referenzenergie von As_4 ergeben sich dreieckige, nur von $(11\bar{2})$ -Kanten begrenzte Inseln. Mit As-Bulk als Referenz finden sich hingegen sechseckige Inseln, bei denen sich kurze $(\bar{1}\bar{1}2)$ -Kanten mit langen $(11\bar{2})$ -Kanten abwechseln. Die vorliegenden STM-Bilder wurden nicht an Gleichgewichtsinseln gemacht, so daß ein Vergleich mit den Vorhersagen noch nicht möglich ist.

¹**ES**t**Co**MPP steht für “**E**lectronic **S**tructure **C**ode for **M**aterials **P**roperties and **P**rocesses”.

Abstract

Investigation of step-edges on As-covered Si(111) using a parallelized *ab initio* code

The first part of this thesis presents the parallelisation of the **ES**t**Co**MPP²-code. Using Density-Functional-Theory **ES**t**Co**MPP simulates many electron systems, found in molecules and solids and computes characteristics of their ground-state. The program employs periodic supercells to approximate inhomogeneous systems. The DFT accounts for the electron-electron interaction by means of a one-electron-potential that depends on the electron-density which is determined selfconsistently. The interaction of valence-electrons with ionic cores is described by norm-conserving pseudo potentials. The electronic states are numerically calculated by solving an eigenproblem for a set of \mathbf{k} -points to find the $\nu = 1..M$ eigenstates with the lowest energies. These eigenstates are represented in a basis of plane-waves labeled by \mathbf{G} -vectors of the reciprocal lattice.

Corresponding to the indices \mathbf{k} , ν and \mathbf{G} of the coefficients of the wavefunctions three parallelisation strategies are considered, each allocating partial ranges of the respective index to individual processors. The \mathbf{k} -point-parallelisation is highly efficient since eigenproblems assigned to different \mathbf{k} -points can be solved independently, thus only 0.2% of the computing time account for non-parallelized tasks. In combination with the \mathbf{G} -vector-parallelisation performance scales well with the number of processors for large systems. For a supercell containing with 110 Atoms less than 0.4% of computing time is spent in unparallelized routines.

The second part of this thesis utilizes the **ES**t**Co**MPP-code to investigate step-edges on As-covered Si(111)-surfaces. The total energies and structures of several metastable states are calculated by relaxing given starting configurations. Since the amount of As contained in the step-edges varies for different structures, the formation energies of the respective step-edges depend on the reference-energy that is used as chemical potential for As. For As-bulk as well as As_4 as reference the most stable structure found was the $(11\bar{2})$ -step-edge with all exposed Si-atoms at the step-edge replaced by As.

Based on the total energies of the various step-edges the form of As-covered Si-islands with the lowest overall energy is predicted by two-dimensional Wulff-plots. Using As_4 -molecules as reference we find triangular islands terminated by $(11\bar{2})$ -step-edges. For As-bulk as reference the result are six-sided islands terminated alternately by short $(\bar{1}\bar{1}2)$ -edges and long $(11\bar{2})$ -edges. The available STM-pictures for Si-homoepitaxy don't show islands in equilibrium so that the predictions of these calculations could not be compared to experimental data.

²**ES**t**Co**MPP is an acronym for “**E**lectronic **S**tructure **C**ode for **M**aterials **P**roperties and **P**rocesses”.

Inhaltsverzeichnis

1	Einleitung und Motivation	5
2	DFT und Pseudopotentiale	9
2.1	Die Dichtefunktionaltheorie	9
2.1.1	Die Born-Oppenheimer-Näherung	9
2.1.2	Grundlagen der DFT	10
2.1.3	Die Kohn-Sham-Gleichungen	12
2.1.4	Die Lokale Dichte Näherung	14
2.1.5	Darstellung des Eigenwertproblems	15
2.1.6	Die Pseudopotentialmethode	17
2.2	Darstellung der Operatoren	23
2.3	Das EStCoMPP-Programm	28
2.4	Skalierungsverhalten	30
3	Parallelisierung	35
3.1	Die \mathbf{k} -Punkt-Parallelisierung	37
3.2	Die \mathbf{G} -Vektor-Parallelisierung	39
3.2.1	Aufteilung der \mathbf{G} -Vektoren	42
3.2.2	Die Fourier-Transformation	44
3.3	Die Band-Parallelisierung	45
3.3.1	Ein paralleles Gram-Schmidt-Verfahren	46
3.4	Zusammenfassung	50
4	Stufenkanten	51
4.1	Motivation	51
4.2	Aufbau der Superzelle und Bezeichnungen	55
4.2.1	Bezeichnungen	55
4.2.2	Gestaltung der Superzelle	59
4.3	Struktur der Stufenkanten	62
4.3.1	Die $(11\bar{2})$ -Stufe mit Silizium an der Kante	65
4.3.2	Die $(11\bar{2})$ -Stufe mit Arsen an der Kante	67
4.3.3	Die $(11\bar{2})$ -Stufe mit überhängendem Arsen	70
4.3.4	Die $(\bar{1}\bar{1}2)$ -Stufe mit Silizium an der Kante	73

4.3.5	Die $(\overline{112})$ -Stufe mit ersetzttem Arsen	75
4.3.6	Die $(\overline{1\overline{1}2})$ -Stufe mit überhängendem Arsen	77
4.4	Energiebetrachtungen	80
4.4.1	Vereinfachte Wulff-Konstruktion	82
4.4.2	STM-Bilder	89
5	Zusammenfassung und Ausblick	93
	Anhang	94
A	Struktur der Stufenkanten	95
A.1	Die $(11\overline{2})$ -Stufe mit Silizium an der Kante	96
A.2	Die $(11\overline{2})$ -Stufe mit Arsen an der Kante	98
A.3	Die $(11\overline{2})$ -Stufe mit überhängendem Arsen	99
A.4	Die $(\overline{112})$ -Stufe mit Silizium an der Kante	100
A.5	Die $(\overline{112})$ -Stufe mit ersetzttem Arsen	102
A.6	Die $(\overline{112})$ -Stufe mit überhängendem Arsen	104
B	Davidson-Kosugi-Algorithmus	105
	Literaturverzeichnis	110

Kapitel 1

Einleitung und Motivation

Der Einsatz moderner Techniken zur Herstellung künstlicher, vertikal und lateral strukturierter Schichtsysteme mit nahezu atomar scharfen Grenzflächen aus bekannten oder neuen Materialien hat der Festkörperphysik einzigartige Möglichkeiten an die Hand gegeben, Systeme mit ganz speziellen elektronischen Eigenschaften zu entwickeln. Bekannte Beispiele sind MOSFETs (**M**etal-**O**xide **S**emiconductor **F**ield **E**ffect **T**ransistor) oder die HEMT-Systeme (**H**igh **E**lectron **M**obility **T**ransistor), an denen dann erstmalig der 1-Elektron-Quanten-Hall-Effekt und der fraktionale Quanten-Hall-Effekt entdeckt wurden.

Heute steht die Herstellung kleinerer Strukturen in der Größenordnung von ca. 1 bis 100 Nanometern in Form von Quanten-Punkten, "Hut-Clustern" oder Nanodrähten im Vordergrund. Bei der Herstellung dieser Strukturen versucht man, das selbstorganisierte Wachstum auszunutzen. Eine Kontrolle der Wachstumsprozesse ist unabdingbar, um solche Strukturen herstellen zu können. Derartige Kontrolle setzt ein detailliertes Verständnis des Wachstums und entsprechende Analysetechniken voraus.

Obwohl die Wachstumsvorgänge oftmals auf makroskopischen Längen- und Zeitskalen stattfinden, liegen diesen mikroskopische, atomistische Prozesse und Mechanismen zugrunde. Erforderlich ist ein Verständnis des Ablaufs und der Relevanz der einzelnen Wachstumsprozesse, wie z.B. Diffusion oder Austausch von Atomen auf der Oberfläche, aber auch der Einfluß atomistischer Inhomogenitäten wie z.B. Oberflächendefekten, Verunreinigungen, Stufenkanten oder Oberflächenrekonstruktionen.

Dabei steht man bei der theoretischen Behandlung der Wachstumsprozesse vor mehreren Problemen: Zum einen wird das gezielte Absuchen eines großen Konfigurationsraumes großer, inhomogener Systeme verlangt. Dazu sind molekulardynamische Verfahren ideal geeignet. Zweitens bedingen die mikroskopischen Prozesse das Brechen und Schließen von chemischen Bindungen. Dabei verlangt die Modifikation der chemischen Bindungen, die bei ständig neuen atomaren Konfigurationen stattfindet, ein verlässliches Vielkörperpotential, das direkt aus der elektronischen Struktur abgeleitet wird. Um diese Struktur zu berechnen, sind

daher “Erste-Prinzipien”- (*ab-initio*-) Methoden nötig, deren Ergebnis die Grundlage der Molekulardynamik bilden.

Die Zeitskala elektronischer Prozesse liegt typischerweise im Femtosekundenbereich, während die Zeitskala der Diffusion im Bereich von Pikosekunden und die des Wachstums im Bereich von Sekunden liegt. Diese verschiedenen Zeitskalen lassen sich nicht in einer einzigen Rechnung erfassen, daher werden sie tatsächlich getrennt voneinander betrachtet. Bei der Relaxation der elektronischen Struktur wird gemäß der Born-Oppenheimer-Näherung davon ausgegangen, daß die Atomkerne statisch sind bzw. daß sich die Elektronen beliebig schnell auf eine Veränderung der atomaren Koordinaten einstellen, so daß atomare Bewegungen und die Zeitskala, auf der diese ablaufen, nicht berücksichtigt werden müssen. Aufgrund der so berechneten elektronischen Struktur und den daraus resultierenden Kräften wird die Molekulardynamik durchgeführt, die wiederum totale Energien für Gleichgewichtszustände und für Übergänge zwischen diesen die Anregungsenergien der entsprechenden atomaren Prozesse liefert. Aufgrund der daraus resultierenden Wahrscheinlichkeiten läßt sich dann schließlich eine kinetische Monte-Carlo-Simulation der Oberfläche durchführen, bei der die einzelnen atomaren Prozesse entsprechend ihrer Wahrscheinlichkeit auftreten.

Die Vielfalt der dabei zu berücksichtigenden atomaren Prozesse und die Schwierigkeit, optimale Reaktionspfade in hochdimensionalen Räumen zu finden, wirft Probleme auf, die selbst unter Ausnutzung modernster Algorithmen den Einsatz von Supercomputern zur Bewältigung der Rechnungen erfordern. Erst die heute zur Verfügung stehenden Leistungen von Rechnern und Großrechnern mit paralleler Architektur ermöglichen die *ab initio* Berechnung der atomaren und elektronischen Struktur von realen Festkörpern mit der notwendigen Zuverlässigkeit. Da bei diesen Simulationsrechnungen mikroskopische Vorgänge detailliert berechnet werden, unterstützen sie die Interpretation von experimentellen Ergebnissen. Sie gestatten dabei ein aussergewöhnliches Maß an Einsicht in die Vorgänge auf atomarer Längenskala. Dies liegt vor allem daran, daß die Simulation es gestattet, einzelne Atome und Elektronenorbitale sowohl zu beobachten als auch durch Zwangsbedingungen ‘festzuhalten’.

Mit Hilfe der Dichtefunktionaltheorie (DFT) [HK64] gelingt es derzeit, die elektronischen Grundzustandseigenschaften von Systemen mit ca. 500 Atomen zu berechnen. Mit dem **ESTCoMPP**-Simulationsprogramm¹ können die Grundzustandseigenschaften beliebiger Kristallstrukturen untersucht werden, für deren konstituierende Atome sich geeignete Pseudopotentiale [BHS82] finden lassen. Durch geeignete Zwangsbedingungen und Betrachtung der Gesamtenergie in Ausschnitten des Konfigurationsraumes lassen sich auch dynamische Prozesse analysieren. Das Programm beschreibt die elektronischen Zustände mit Hilfe von ebenen Wellen. Aufgrund dieser Basiswahl werden grundsätzlich periodische Strukturen berechnet, die sich periodisch wiederholenden Zellen werden dabei als Super-

¹**ESTCoMPP** steht für “**E**lectronic **S**tructure **C**ode for **M**aterials **P**roperties and **P**rocesses”.

zellen bezeichnet. Zur Betrachtung der Bulkigenschaften eines Kristalles wird z.B. die Elementarzelle des Kristalles als Superzelle vorgegeben und mit Atomen entsprechend der Basis des Kristalles besetzt. Die Wahl einer ausreichend großen Superzelle gestattet auch die Untersuchung nichtperiodischer Strukturen, wie kristalline Oberflächen und Überstrukturen auf diesen. Aber es gibt auch periodische Strukturen, zu deren Untersuchung sehr große Superzellen nötig sind. Auf der reinen Si(111)-Oberfläche kommt es zu einer 7×7 -Rekonstruktion für die von Takayanagi 1985 eine DAS (Dimer-Adatom-Stackingfault) Struktur vorgeschlagen wurde [TTTT85]. Erst durch den Einsatz massiv paralleler Rechner gelang 1991 zwei Gruppen die *ab-initio*-Berechnung der dazu nötigen Superzellen mit 400 bzw. 700 Atomen [SPKSL92, BNLJ92, Nee93]. Auch das Verhalten von Adatomen auf kristallinen Oberflächen läßt sich untersuchen. Dabei ist darauf zu achten, daß Strukturen benachbarter Superzellen sich nicht gegenseitig beeinflussen². Die Superzellenmethode eignet sich ideal zur Untersuchung von Überstrukturen, wie sie unter anderem auf rekonstruierten Oberflächen, beispielsweise bei Adsorption von Fremdatomen auf Oberflächen entstehen.

In unserer Arbeitsgruppe sind vor allem Halbleitersysteme untersucht worden, z.B. die Diffusion von Adatomen auf As- und Sb-bedeckten Oberflächen [SERB98, SAB⁺00, SABB02] und die ersten Stadien der Inselbildung bei Homo- und Heteroepitaxie auf diesen Oberflächen [Ant01]. Der theoretische Beitrag besteht in der Berechnung von atomaren Defekt-Strukturen mit minimaler Gesamtenergie. Daraus kann man Schlüsse auf realisierte Strukturen ziehen und Vorschläge für Experimente machen. Die Berechnungen mit dem **ES**t**Co**MPP-Programm gehen für die Festkörperelektronen von der DFT [HK64] aus, deren Grundlage in Kapitel 2.1 vorgestellt wird. Die Wechselwirkung der für die Bindung verantwortlichen Valenzelektronen mit den Ionen wird durch Pseudopotentiale beschrieben. Als Basis für die Darstellung der Elektronendichte wird eine Ebene-Wellen-Basis gewählt, die für Halbleitersysteme mit vollem Valenz- und leerem Leitungsband besonders geeignet ist [PTA⁺92]. Da der Basiswechsel in den Ortsraum durch eine Fourier-Transformation vollzogen wird, ist die Entwicklung der Fast-Fourier-Transformation für die numerische Formulierung der DFT in einer Ebenen-Wellen-Basis von besonderer Bedeutung. Durch diese Möglichkeit eines schnellen Wechsels zwischen reziprokem Raum und Ortsraum können die einzelnen Bestandteile des Hamiltonoperators in der jeweils günstigsten Basis angewandt werden, was zur expliziten Darstellung des Hamiltonoperators in Kapitel 2.2 führt. Mit dieser läßt sich der durch die DFT vorgegebene Ansatz zur Lösung des Vielelektronenproblems als numerischer Algorithmus formulieren. Kapitel 2.3 stellt die im **ES**t**Co**MPP-Programm gewählte Implementation vor, die mit Rücksicht auf Speicherplatz- und Rechenzeitbedarf optimiert wurde, um die Berechnung möglichst großer Systeme zu ermöglichen. Beim Übergang

²Dies kann z.B. durch Vergleich der Untersuchungen für Superzellen unterschiedlicher Form oder Größe gezeigt werden.

zu immer größeren Systemen bestimmt vor allem das in Kapitel 2.4 angegebene Skalierungsverhalten von Rechenzeit und Speicherplatz in Abhängigkeit von der Systemgröße, welche Berechnungen mit einer gegebenen Rechenleistung und Speicherausstattung möglich sind.

Um die Leistung moderner, massiv paralleler Rechner nutzen zu können, wurde das **ESTCoMPP**-Programm parallelisiert [BBA⁺00]. Kapitel 3 beschreibt die notwendigen Umstellungen zur Verteilung der Rechenschritte und Daten auf viele ($\mathcal{O}(100)$) Rechenknoten (**P**rocessing **E**lements, PEs) und geht auf die verschiedenen Ansätze zur Parallelisierung und deren Effizienz ein. Es werden drei mögliche Parallelisierungsstrategien vorgestellt, die im wesentlichen auf der Aufteilung der Wellenfunktionen nach **k**-Punkten (Kap. 3.1), nach **G**-Vektoren (Kap. 3.2) und nach Energiebändern (Kap. 3.3) beruhen. Dabei zeigt sich, daß die **G**-Vektor-Parallelisierung von besonderem Interesse ist, da sie bei der Berechnung großer Systeme den Einsatz vieler Prozessoren erlaubt und daher das **ESTCoMPP**-Programm ideal an den Einsatz auf massiv parallelen Rechnern anpasst. Erst die Parallelisierung ermöglicht die Berechnung der oben erwähnten großen Systeme, die z.T. 500 und mehr Atome enthalten.

In solch großen Systemen lassen sich z.B. atomare Strukturen und Prozesse auf fremdatombedeckten Oberflächen beobachten. Die Untersuchung der Mechanismen, durch die Surfactants (“oberflächenaktive Fremdatome”) das Lagenwachstum bei der Heteroepitaxie beeinflussen, ist von besonderem Interesse bei der Herstellung neuer Halbleitermaterialien. Durch verschiedene Surfactants läßt sich ein verbessertes heteroepitaktisches Wachstum erreichen. Z.B. beeinflußt eine Monolage von As oder Sb das Lagenwachstum von Ge auf der Si(111)-Oberfläche günstig [MVA92, VZ96, HCT⁺94]. Bei diesen Surfactants beobachteten Voigtländer *et al.* im homoepitaktischen Fall eine Verkürzung des Diffusionsweges. Sie schlossen daraus, daß kürzere Diffusionswege zu Verbesserung des heteroepitaktischen Lagenwachstums führen. Eine wichtige Rolle beim Wachstum spielen die Stufenkanten, die Terrassenflächen und Inseln, die sich auf den Terrassen bilden, begrenzen. In Kapitel 4 werden solche Stufenkanten auf der Asbedeckten Si(111)-Oberfläche mit Hilfe des **ESTCoMPP**-Programmes untersucht, um so deren energetisch günstigste Struktur zu finden und Einblicke in die Wirkungsweise von Arsen als Surfactant auf dieser Oberfläche zu erhalten.

Kapitel 2

Dichtefunktionaltheorie und Pseudopotential-Methode

2.1 Die Dichtefunktionaltheorie

Die Dichtefunktionaltheorie ist ein wichtiges Werkzeug bei der Berechnung und Simulation elektronischer und struktureller Eigenschaften von Atomen, Molekülen und Festkörpern, da sie es ermöglicht, das quantenmechanische Viel-Elektronenproblem auf Einteilchenprobleme abzubilden und so zu lösen. Daher kommt die DFT, anders als z.B. Näherungen unter Benutzung phänomenologischer Potentiale, ohne Anpassung von Parametern an experimentell gewonnene Daten aus, Atome werden allein durch ihre Kernladungszahl charakterisiert. In solche *ab initio* Rechnungen gehen als zusätzliche Parameter nur noch Vorgaben für die Qualität notwendiger Näherungen ein.

Die Vereinfachung durch die DFT erschließt den Zugang zu realen Problemen mit vielen Atomen. Sie ermöglicht auch, experimentell beobachtbare Grundzustandseigenschaften größerer Systeme zu berechnen. Mit dem in dieser Arbeit beschriebenen parallelisierten Programm haben wir Systeme mit bis zu 300 Atomen auf einer Cray-T3E behandelt. So helfen *ab initio* Rechnungen experimentelle Beobachtungen zu erklären und gestatten, Eigenschaften betrachteter Systeme bis ins Detail zu untersuchen. Von Interesse sind dabei beispielsweise die Verknüpfung von elektronischen Eigenschaften mit der atomaren Struktur und die theoretische Simulation von experimentellen Ergebnissen.

2.1.1 Die Born-Oppenheimer-Näherung

Betrachtet man das aus Atomkernen und Elektronen bestehende Vielteilchensystem, so läßt sich der zugehörige Hamiltonoperator aufgrund der stark unterschiedlichen Massen von Elektronen und Atomkernen in zwei Anteile zerlegen, da die Bewegungen auf sehr unterschiedlichen Zeitskalen ablaufen: die leichten und

schnellen Elektronen passen sich dabei adiabatisch jeder Veränderung der Ionenkonfiguration an und befinden sich bezüglich des durch die Atomkerne vorgegebenen äußeren Potentials im Grundzustand. Die Atomkerne sehen dann das durch die Elektronendichte verursachte Potential und die daraus resultierenden Kräfte. Dabei werden Anregungen der Elektronen durch Bewegungen der Atomkerne, also z.B. Elektron-Phonon-Wechselwirkungen, vernachlässigt. Die Bewegung der Atomkerne kann, außer im Falle der leichtesten Elemente, klassisch behandelt werden.

Die chemischen und strukturellen Eigenschaften von Festkörpern und Molekülen werden im wesentlichen von den äußeren Valenzelektronen bestimmt. Die Elektronen in den gefüllten inneren Schalen werden kaum von den Potentialen benachbarter Atome beeinflusst, da der Überlapp ihrer Wellenfunktionen mit denen anderer Atome gering ist. Zudem läßt sich schon in Störungstheorie sehen, daß die Wellenfunktionen dieser sogenannten Rumpf- oder Coreelektronen aufgrund ihres großen energetischen Abstandes zu denen der äußeren Valenzelektronen kaum zu den Eigenfunktionen des Festkörpers im Bereich des Valenz- oder Leitungsbandes beitragen. D.h. die Wellenfunktionen der Cores können bei der Berechnung von Festkörpereigenschaften i.A. festgehalten werden. Damit vereinfacht sich das Elektronenproblem, da nur die Valenzelektronen als quantenmechanisches Vielteilchensystem im äußeren Feld der Ionenrümpfe zu berechnen sind.

2.1.2 Grundlagen der Dichtefunktionaltheorie

Der Hamiltonoperator eines Systems N wechselwirkender Elektronen mit den Orts- und Spinkoordinaten $X_\nu = (\mathbf{r}_\nu, \sigma_\nu)$, $\nu = 1 \dots N$ in einem durch Atomkerne oder Ionenrümpfe erzeugten äußeren Potential $V_{ext}(X_\nu)$ lautet:

$$\mathcal{H} = \mathcal{T} + V_{ee} + \sum_{\nu=1}^N V_{ext}(X_\nu) \quad (2.1)$$

mit der kinetischen Energie

$$\mathcal{T} = - \sum_{\nu=1}^N \partial_{\mathbf{r}_\nu}^2, \quad (2.2)$$

und der Elektron-Elektron Wechselwirkung

$$V_{ee} = \sum_{\nu \neq \mu}^N \frac{1}{|\mathbf{r}_\nu - \mathbf{r}_\mu|}. \quad (2.3)$$

Dabei wird der der Zustand des Systems durch eine Vielteilchenwellenfunktion

$$\Psi(X_1, X_2, \dots, X_N) \quad X_\nu = (\mathbf{r}_\nu, \sigma_\nu) \quad (2.4)$$

angegeben. Aufgrund des Pauliprinzips muß jeder Eigenzustand antisymmetrisch in bezug auf Vertauschung zweier Teilchenkoordinaten $(X_\nu, X_{\nu'})$ sein.

Die Bestimmung der Eigenfunktion bleibt, auch nach obigen Vereinfachungen, ein quantenmechanisches Vielteilchensystem, das analytisch nur in Spezialfällen zu lösen ist. Die vollständige numerische Behandlung (beispielsweise mit Quanten-Monte-Carlo-Methoden) ist so aufwendig, daß sich derzeit nur Systeme in der Größenordnung von zehn Ionenrümpfen untersuchen lassen. In der Dichtefunktionaltheorie¹, deren Grundlagen in den sechziger Jahren von Hohenberg und Kohn [HK64] sowie Kohn und Sham [KS65] formuliert wurden, wird anstelle der Vielteilchenwellenfunktion (2.4) die Elektronendichte

$$\begin{aligned} n(\mathbf{r}) &= \langle \Psi | \sum_{\nu=1}^N \delta(\mathbf{r} - \mathbf{r}_\nu) | \Psi \rangle \\ &= \int \prod_{\nu=1}^N dX_\nu \sum_{\nu=1}^N \delta(\mathbf{r} - \mathbf{r}_\nu) |\Psi(X_1, \dots, X_N)|^2 \end{aligned} \quad (2.5)$$

benutzt, um den Zustand des Systems zu beschreiben. Hohenberg und Kohn [HK64] zeigten, daß ein Energiefunktional $E[n]$ existiert, das durch die Grundzustandsdichte $n_0(\mathbf{r})$ minimiert wird², so daß diese durch ein Variationsprinzip gefunden werden kann, das die Ladungserhaltung berücksichtigt. Eine mögliche Konstruktion eines solchen Funktionals fand Levy [Lev79], indem er $E[n]$ als minimale Energie aller möglichen antisymmetrischen Zustände Ψ des Systems definierte, die nach (2.5) die Ladungsdichte $n(\mathbf{r})$ ergeben:

$$E[n] = \min_{\{\Psi\}} \langle \Psi | \mathcal{H} | \Psi \rangle \quad (2.6)$$

In diesem Fall ist $E[n_0]$ minimal, man erhält also

$$\delta E[n] = 0 \quad |_{n=n_0} \quad (2.7)$$

unter der Nebenbedingung der Teilchenzahlerhaltung

$$\int d^3r n(\mathbf{r}) = N. \quad (2.8)$$

Wie üblich gelangt man unter Einführung eines Lagrange Parameters μ zu einem neuen Energiefunktional,

$$\tilde{E}[n] = E[n] - \mu \left(\int d^3r n(\mathbf{r}) - N \right). \quad (2.9)$$

¹Ein ausführlicher Übersichtsartikel findet sich in [JG89].

²Der Dichtefunktionalformalismus ist, wie in [Koh85] und [DG90] gezeigt, auch auf entartete Grundzustände erweiterbar.

Die Minimalbedingung führt auf die Gleichung

$$\left. \frac{\delta \tilde{E}[n]}{\delta n(\mathbf{r})} \right|_{n=n_0} = \mu , \quad (2.10)$$

die gemeinsam mit (2.8) eine Bestimmung der Grundzustandsdichte $n_0(\mathbf{r})$ erlaubt.

Allerdings ist das Energiefunktional (2.6) nicht explizit darstellbar. Setzt man aber die volle Vielteilchenwellenfunktion (2.4) in (2.6) ein, so führt dies aufgrund des Wechselwirkungsterms V_{ee} wieder zur Schrödingergleichung für ein wechselwirkendes Vielteilchensystem zurück.

2.1.3 Die Kohn-Sham-Gleichungen

Es muß also eine geeignete Näherung für das Energiefunktional $E[n]$ gefunden werden, um das elektronische Vielteilchenproblem numerisch behandeln zu können. Dazu haben Kohn-Sham [KS65] dieses Funktional wie folgt aufgespalten:

$$E[n] = T_0[n] + E_{ext}[n] + E_H[n] + E_{xc}[n] . \quad (2.11)$$

Zur Parametrisierung der Dichte schlugen sie vor, ein System nicht wechselwirkender Hilfselektronen mit den Einteilchenwellenfunktionen $\psi_\nu(\mathbf{r})$ zu benutzen. Mit den Besetzungszahlen f_ν ergibt sich eine Dichte

$$n(\mathbf{r}) = 2 \sum_{\nu} f_{\nu} |\psi_{\nu}(\mathbf{r})|^2 , \quad (2.12)$$

wobei der Faktor zwei durch Spinartung zustande kommt³ und aufgrund der Wahl normierter Wellenfunktionen $\psi_\nu(\mathbf{r})$ gilt:

$$N = 2 \sum_{\nu} f_{\nu} . \quad (2.13)$$

Mit diesem Ansatz kann die kinetische Energie $T_0[n]$ der wechselwirkungsfreien Elektronen exakt berechnet werden zu

$$T_0[n] = -2 \sum_{\nu} f_{\nu} \int d^3r \psi_{\nu}^*(\mathbf{r}) \partial_{\mathbf{r}\nu}^2 \psi_{\nu}(\mathbf{r}) . \quad (2.14)$$

Die potentielle Energie $E_{ext}[n]$ beschreibt die Wechselwirkung der Elektronen mit einem äußeren Potential der Atomkerne bzw. Ionenrümpfe:

$$E_{ext}[n] = \int d^3r n(\mathbf{r}) V_{ext}(\mathbf{r}) . \quad (2.15)$$

³Der Formalismus läßt sich analog auch für den allgemeinen Fall nicht spinentarteter Systeme anwenden, im Folgenden beschränkt sich diese Arbeit jedoch der Vereinfachung halber auf den Fall spinentarteter Systeme.

Die Elektron-Elektron Wechselwirkung ist unterteilt in den Hartree Anteil

$$E_H[n] = \int d^3r d^3r' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.16)$$

und das Austausch-Korrelations-Funktional $E_{xc}[n]$, das die Austausch- und Korrelationsbeiträge sowie Korrekturen der kinetischen Energie und der in $E_H[n]$ enthaltenen Selbstwechselwirkung enthält. Außer dem Anteil $E_{xc}[n]$ können also alle Energien exakt behandelt werden. Somit muß nur für das Austausch-Korrelations-Funktional eine geeignete Näherung gefunden werden. Der Vorteil beispielsweise gegenüber Thomas-Fermi und ähnlichen Methoden, bei denen die kinetische Energie genähert wird, ist, daß $E_{xc}[n]$ den kleinsten Anteil an $E[n]$ hat und somit eine Näherung von $E_{xc}[n]$ die Gesamtenergie am wenigsten verfälscht [JG89].

Aufgrund der geforderten Wechselwirkungsfreiheit der Hilfselektronen zerfällt nun die Lagrangegleichung (2.10) in unabhängige Einzelprobleme, wie man durch Variation nach den Wellenfunktionen $\psi_\nu^*(\mathbf{r})$ feststellen kann:

$$\frac{\delta E[n(\mathbf{r})]}{\delta \psi_\nu^*(\mathbf{r})} = \frac{\delta E[n(\mathbf{r})]}{\delta n(\mathbf{r})} \frac{\delta n[\dots, \psi_\nu, \dots]}{\delta \psi_\nu^*(\mathbf{r})}. \quad (2.17)$$

Die Nebenbedingung der Ladungserhaltung wird mittels Normierung der Wellenfunktionen

$$\int d^3r |\psi_\nu(\mathbf{r})|^2 \quad (2.18)$$

durch Einführung des Lagrangeparameters ε_ν im Energiefunktional berücksichtigt. Man erhält so:

$$(-\Delta_{\mathbf{r}}^2 + V_{ext}(\mathbf{r}) + V_H[n(\mathbf{r})] + V_{xc}[n(\mathbf{r})]) \psi_\nu(\mathbf{r}) = \varepsilon_\nu \psi_\nu(\mathbf{r}). \quad (2.19)$$

Diese Form ist analog zur stationären Schrödingergleichung im wechselwirkungsfreien Fall. Die wechselwirkungsfreien Hilfselektronen sind dabei einem Potential $V_{eff} = V_{ext} + V_H + V_{xc}$ ausgesetzt, mit dem externen Potential V_{ext} der Ionenrümpfe, dem Hartree-Potential

$$V_H[n(\mathbf{r})] = \frac{\delta E_H[n(\mathbf{r})]}{\delta n(\mathbf{r})} = 2 \int d^3r' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (2.20)$$

und dem Austausch-Korrelations-Potential

$$V_{xc}[n(\mathbf{r})] = \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \quad (2.21)$$

Die Wellenfunktionen ψ_ν lassen sich also nun aus dem Eigenwertproblem zu (2.19) bestimmen. Allerdings ist das effektive Potential in diesem Eigenwertproblem von der Dichte $n(\mathbf{r})$ abhängig, die wiederum nach (2.12) von den Wellenfunktionen

selbst abhängt. Man gelangt so zu einem *Selbstkonsistenzproblem*, welches sich iterativ lösen läßt. Als Testfunktion für das Erreichen der Selbstkonsistenz kann entweder die Dichte $n(\mathbf{r})$ oder das Potential $V_{eff}(\mathbf{r})$ gewählt werden.

Die Wellenfunktionen ψ_ν und die Lagrange-Parameter ε_ν der Kohn-Sham-Gleichungen beschreiben formal die Hilfselektronen, einzig die Dichte $n(\mathbf{r})$ und aus dem Variationsprinzip (2.10) ableitbare Größen sind also tatsächliche physikalische Größen. In der Praxis zeigt sich jedoch, daß, besonders für delokalisierte Zustände, Differenzen von ε_ν sehr gut mit den Ionisierungsenergien übereinstimmen und beispielsweise zur Darstellung der Bandstruktur genutzt werden können. Durch Aufsummieren der Aufenthaltswahrscheinlichkeiten $|\psi_\nu|^2$ der Hilfselektronen über einen Energiebereich $\varepsilon_l < \varepsilon_\nu < \varepsilon_u$ lassen sich sogar mit STM-Bildern vergleichbare Dichteplots generieren.

2.1.4 Die Lokale Dichte Näherung

Da das Austausch-Korrelations-Potential $V_{xc}[n(\mathbf{r})]$ nicht in einer expliziten Form angegeben werden kann, ist es notwendig, für diesen Anteil eine geeignete Näherung zu finden. Während alle anderen Anteile im Rahmen des Kohn-Sham-Formalismus exakt behandelt werden, ist es nötig, diesen Term zu approximieren, um von einem System wechselwirkender Elektronen zu einem System nicht wechselwirkender Hilfselektronen zu gelangen. Die wesentliche Näherung betrifft die Darstellung der Paarkorrelationsfunktion des wechselwirkenden Systems. Gunnarsson und Lundqvist [GL76] haben gezeigt, daß die Austausch-Korrelationsenergie nicht empfindlich von Details dieser Paarkorrelation abhängt.

Der Ansatz der Lokalen Dichte Näherung (LDA) setzt für das Austausch-Korrelations-Funktional $E_{xc}[n]$ eine lokale Energiedichte pro Teilchen $\epsilon_{xc}(n(\mathbf{r}))$ an, die nur von der lokalen Dichte abhängt, also kein Funktional der Dichte ist. Damit wird

$$E_{xc}^{LDA}[n] = \int d^3r n(\mathbf{r}) \epsilon_{xc}(n(\mathbf{r})) . \quad (2.22)$$

Das Austausch-Korrelations-Potential erhält man, indem (2.22) nach der Dichte variiert wird:

$$V_{xc}^{LDA}(n(\mathbf{r})) = \epsilon_{xc}(n(\mathbf{r})) + n(\mathbf{r}) \frac{\partial \epsilon_{xc}}{\partial n}(n(\mathbf{r})) \quad (2.23)$$

Für die Energiedichte $\epsilon_{xc}(n)$ wird dazu der Austausch-Korrelations-Anteil der Energiedichte des homogenen Elektronengases benutzt. Diese ist analytisch für die Grenzfälle hoher und niedriger Elektronendichte bekannt, unser Programm verwendet eine Parametrisierung von Vosko, Wilk, Nusair [VWN80] nach numerischen Ergebnissen einer Monte-Carlo-Simulation von Ceperley und Alder [CA80] mit relativistischen Korrekturen nach MacDonald und Vosko [MV79]. Wahlweise lassen sich auch andere Parametrisierungen verwenden, wobei die Ergebnisse nur

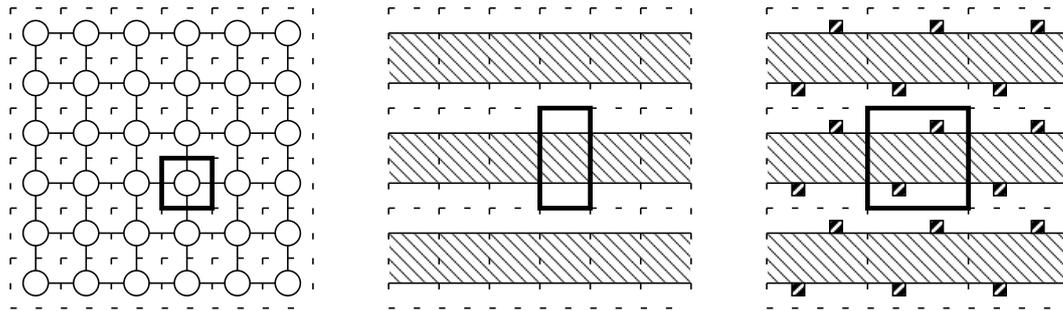


Abbildung 2.1: Superszellen zur Untersuchung von Oberflächen und Strukturen auf Oberflächen (Seitenansicht). Links: Untersuchung der Bulkeigenschaften von Kristallen; die Superszelle entspricht der Einheitszelle des Kristalles. Mitte: Untersuchung von Oberflächen; Materialschichten werden durch Vakuumschichten getrennt, die Zelle muß in vertikaler Richtung ausgedehnt werden um ausreichend Bulkmaterial und Vakuum einzuschließen. Rechts: Untersuchung von Strukturen auf Oberflächen; tatsächlich werden periodisch angeordnete Strukturen auf der Oberfläche berechnet. Die Superszelle muß auch in lateraler Richtung ausgedehnt werden, damit sich die Strukturen verschiedener Superszellen nicht beeinflussen.

wenig voneinander abweichen. Allerdings lassen sich nur Gesamtenergien miteinander vergleichen, die mit derselben Parametrisierung berechnet wurden.

Obwohl der Ansatz der LDA von Systemen mit möglichst homogener Elektronendichte ausgeht, zeigt sich in der Praxis, daß sie auch für Systeme mit starken Dichteschwankungen, wie z.B. Oberflächen, noch gute Ergebnisse liefert. Dies liegt daran, daß zu einer exakten Berechnung der Austausch-Korrelations-Energie nur das Mittel über das Austausch-Korrelations-Loch korrekt wiedergegeben werden muß [JG89, GL76]. Allerdings erzielt man für stark inhomogene Systeme, wie z.B. Moleküle, mit Hilfe der Generalized Gradient Approximation (GGA) bessere Resultate [PW86, Per86, PCV⁺92, OEK99].

2.1.5 Mathematische Darstellung des Eigenwertproblems

Wir wollen die mathematischen Hilfsmittel der etablierten Bandstrukturmethoden benutzen. Dazu ist es notwendig, ein periodisches System zu definieren. Perfekte Kristallgitter haben diese Periodizität. Für inhomogene Systeme, wie z.B. Oberflächen oder Defekte, definiert man eine sogenannte Superszelle, die so viele Atome enthält, daß unerwünschte Wechselwirkungen zwischen den Oberflächen und den Replikationen der Defekte minimiert werden, wie in Abbildung 2.1 gezeigt. Die Superszelle wird periodisch in allen drei Raumrichtungen wiederholt. Die elektronischen Wellenfunktionen sind in einem entsprechenden periodischen Potential gemäß dem Bloch-Theorem periodisch bis auf einen Phasenfaktor $\exp\{i\mathbf{k}\mathbf{r}\}$. Daher kann durch Berechnung für genügend \mathbf{k} -Werte das elektronische Problem für ein unendlich ausgedehntes Bravais-Gitter periodisch aneinandergelagert werden.

setzter Superzellen gelöst werden⁴.

Weiter muß nun eine geeignete Basis für die Wellenfunktionen gewählt werden. Eine Betrachtung des Hamiltonoperators ergibt, daß dieser sich in zwei Teile zerlegen läßt, für die jeweils eine optimale Basis existiert: zum einen der Ortsraum, in dem Multiplikationsoperatoren, wie z.B. der Ausdruck $V_{ps}(\mathbf{r})\psi_\nu(\mathbf{r})$, eine einfache Form annehmen, zum anderen der Impulsraum, in welchem Differentialoperatoren, wie z.B. die Anwendung der kinetischen Energie $T_0\psi_\nu$, besonders einfach werden. Auch zur Gewinnung des Hartree Potentials aus der lokalen Dichte gemäß (2.16) ist die Impulsraumdarstellung geeignet. Diese beiden Darstellungen lassen sich durch die Fourier-Transformation ineinander überführen, allerdings wird dieser Lösungsweg numerisch erst durch die Fast-Fourier-Transformation (FFT) interessant, da die herkömmliche Fourier-Transformation durch das quadratische Anwachsen des Aufwandes mit der Vektorlänge bei großen Problemen zu rechenintensiv wird. Die Berechnung von Operatoren zu verschiedenen \mathbf{k} -Punkten wirkt sich im Ortsraum als Anmultiplikation eines Phasenfaktors aus, während sie im Impulsraum nur eine Verschiebung um \mathbf{k} bedeutet.

Die Wellenfunktionen werden in der Ebenen-Wellen-Basis durch einen Satz Koeffizienten $c_{\mathbf{k}}(\mathbf{G}_n)$ zu den Wellenvektoren \mathbf{G}_n dargestellt:

$$\psi(\mathbf{r}) = \sum_{n=1}^{N_G} c_{\mathbf{k}}(\mathbf{G}_n) e^{-i(\mathbf{k}+\mathbf{G})\mathbf{r}} . \quad (2.24)$$

Dabei wird die Randbedingung der Periodizität der Bloch-Funktionen auf dem Bravais-Gitter der Superzellen erzwungen, indem nur \mathbf{G} -Vektoren des entsprechenden reziproken Gitters zugelassen werden. Somit bestimmt die Größe der Superzelle die Dichte der \mathbf{G} -Vektoren. Um die Basisfunktionen auf eine endliche Anzahl zu beschränken, wird ein Abschneideradius \mathbf{G}_{max} eingeführt, und es werden nur Basisfunktionen mit $|\mathbf{G}| < \mathbf{G}_{max}$ berücksichtigt. Diese Einschränkung gestattet nur eine endliche Auflösung der Wellenfunktionen. Mathematisch wird dies klar, wenn man die Fourier-Transformation, also die Rücktransformation, aus dem Impulsraum in den Ortsraum betrachtet. Dazu kann man sich auf eine Box beschränken, die die Kugel mit dem Radius \mathbf{G}_{max} einhüllt⁵. Sie wird durch

⁴Da tatsächlich nur endlich viele, meist äquidistant gewählte, \mathbf{k} -Punkte berücksichtigt werden, erzwingt man tatsächlich eine Periodizität der Wellenfunktionen nach endlich vielen Superzellen.

⁵Tatsächlich muß man zur Darstellung der Dichte $n(\mathbf{r})$ nach (2.12) eine in jeder Raumrichtung doppelt so große Box nehmen, da das Quadrat $|\psi_\nu(\mathbf{r})|^2$ in der Ebenen-Wellen-Darstellung einer Faltung der Koeffizienten $c_{\mathbf{k}}(\mathbf{G}_n)$ mit sich selbst entspricht, und es sonst zu aliasing-Effekten kommen würde. Entsprechend besteht die Basis zur Darstellung der Dichte $n(\mathbf{r})$, des davon abhängigen Hartree-Potentials $V_H[n(\mathbf{r})]$, sowie des lokalen Anteils des Potentials $V_{ext}(\mathbf{r})$ in der Ebenen-Wellen-Darstellung aus \mathbf{G} Vektoren innerhalb einer Kugel mit Radius $2\mathbf{G}_{max}$. Die Wellenfunktionen $\psi_\nu(\mathbf{r})$ werden, nachdem das Potential $V_{ps}(\mathbf{r})$ auf sie angewandt wurde, wieder auf ihren ursprünglichen Basissatz projiziert. So werden auch bei dieser Operation aliasing-Effekte vermieden.

$\mathbf{G}_{x,max}$, $\mathbf{G}_{y,max}$ und $\mathbf{G}_{z,max}$ aufgespannt. Dies entspricht einer diskreten Fourier-Transformation auf dieser Box, so daß $\psi(\mathbf{r})$ auch nur auf einem diskreten Gitter von Stützpunkten $\{\mathbf{r}\}$ dargestellt wird, mit $\mathbf{r} \cdot \mathbf{G}_{[xyz],max}$ ganzzahlig. Die Dichte der Stützpunkte ist also von der Größe der Box und somit von \mathbf{G}_{max} abhängig. Physikalisch entspricht das einer Beschränkung auf kinetische Energien mit

$$E_{max} = \frac{\hbar^2}{2m_e} \mathbf{G}_{max}^2 \quad (2.25)$$

und ist daher nur im Falle eines flachen Potentials sinnvoll, da ein Valenzelektron in einem tiefen Potentialtopf eine stark variierende Wellenfunktion und eine hohe kinetische Energie hätte, also durch Basisfunktionen mit entsprechend großen \mathbf{G} -Vektoren dargestellt werden müßte

Die oben genannten Methoden erlauben es, den Hamiltonoperator auf eine Wellenfunktion ψ_ν anzuwenden. Allerdings verbietet sich eine Darstellung als Matrix aufgrund des Speicherbedarfs. Daher werden, wie von Car und Parrinello vorgeschlagen [CP85a], iterative Methoden benutzt, um das Eigenwertproblem zu lösen. Dabei kann jeder Anteil des Hamiltonoperators in der optimalen Darstellung, also Orts- oder Impulsraum, angewandt werden. In Kapitel 2.2 wird die explizite Darstellung der Operatoren angegeben und diskutiert.

2.1.6 Die Pseudopotentialmethode

Die Eigenschaften von Festkörpern und Molekülen sind im wesentlichen durch die chemischen Bindungen bestimmt, an denen vor allem die äußeren Valenzelektronen beteiligt sind. Dies liegt daran, daß die Orbitale energetisch tiefer liegender core-Elektronen einander nicht überlappen, so daß die zugehörigen Energieniveaus sich nicht zu Bändern verbreitern. Der zugehörige Term aus erster Ordnung Störungstheorie ist durch den Faktor $1/(\varepsilon_{core} - \varepsilon_{valenz})$ also klein genug, daß die Wellenfunktionen der core-Elektronen als ungestört angenommen werden können. Daher läßt sich ein Atom in einem Festkörper hinreichend genau beschreiben, indem man die Wechselwirkung des Ionenrumpfes mit den Valenzelektronen durch ein "Pseudopotential" ersetzt, das die Streuung der Valenzelektronen korrekt reproduziert. Dadurch ergibt sich eine erste Vereinfachung, da nur noch Valenzelektronen tatsächlich berechnet werden müssen. Bei diesem Pseudopotential kann man zunächst zwischen einem äußeren Bereich, weit entfernt vom Atomkern, und einem inneren Bereich, nahe dem Atomkern, unterscheiden. Im äußeren Bereich, in dem die Einflüsse von verschiedenen Atomen überlappen können, muß das Potential korrekt wiedergegeben werden. Im inneren Bereich muß das Potential so gewählt werden, daß die relevanten Streueigenschaften des Einzelatoms im Energiebereich der Valenzelektronen korrekt reproduziert werden.

Da die Wellenfunktionen mittels ebener Wellen dargestellt werden [IZC79], ist es sinnvoll, ein möglichst "weiches" Pseudopotential zu wählen, um so die

Anzahl benötigter ebener Wellen zu reduzieren. Das Potential des Ionenrumpfes wird also durch ein Pseudopotential mit folgenden Eigenschaften ersetzt:

- i) Das Potential hat gleiche Streueigenschaften wie das ursprüngliche atomare Potential im Energiebereich der Valenzelektronen.
- ii) Die Grundzustände dieses Pseudopotentials entsprechen den Zuständen der Valenzelektronen.
- iii) Das Potential selbst und die Grundzustände sind mit möglichst wenigen ebenen Wellen darstellbar.

Dies bringt zwei weitere Vereinfachungen mit sich: Da die Wellenfunktionen der Valenzelektronen den Grundzuständen des Pseudopotentials entsprechen, ist es nicht nötig, bei der Lösung des elektronischen Problems auf Orthogonalität zu den Wellenfunktionen der core-Elektronen zu achten. Außerdem können die Wellenfunktionen der Valenzelektronen und somit das Eigenwertproblem mit wenigen ebenen Wellen beschrieben werden.

Konstruktion des Pseudopotentials

Zur Konstruktion eines Pseudopotentials berechnet man zunächst ein vollständiges Referenzatom⁶. Weiter wird das Potential durch einen Parameter R_c in einen Rumpfbereich und einen Valenzbereich unterteilt. Im inneren Rumpfbereich ($r < R_c$) halten sich die stark gebundenen Elektronen der inneren Schalen auf. Da diese, wie oben gesagt, wenig Einfluss auf die chemischen Eigenschaften haben und auch wenig von der äußeren Umgebung beeinflusst werden, werden ihre Wellenfunktionen und Dichten in der “frozen core”-Näherung als unveränderlich angenommen. Im äußeren Valenzbereich ($r > R_c$) ist das Potential flach, und durch die core-Elektronen abgeschirmt. Allerdings überlappen hier die Wellenfunktionen der Valenzelektronen mit denen anderer Atome und bilden chemische Bindungen oder ausgedehnte Zustände. Deshalb müssen die Eigenschaften der Elektronen im Außenbereich möglichst genau reproduziert und für jedes behandelte Festkörperproblem neu berechnet werden.

Ausgehend von den berechneten Wellenfunktionen für das komplette Atom “friert” man also nun die Rumpfelektronen ein und erhält ein effektives Potential $V_{core}^{eff}(\mathbf{r})$ des Ionenrumpfes in dem sich die Valenzelektronen bewegen. Im nächsten Schritt ersetzt man die echten Wellenfunktionen $\psi_\nu(\mathbf{r})$ (mit Eigenwert ε_ν) der Valenzelektronen durch knotenlose Pseudowellenfunktionen $\psi_\nu^{ps}(\mathbf{r})$. Diese müssen außerhalb des Abschneideradius’ identisch mit den ursprünglichen $\psi_\nu(\mathbf{r})$ sein. Sie werden stetig und möglichst glatt ins Innere fortgesetzt. Zu diesen Wellenfunktionen wird ein Pseudopotential $V^{ps}(\mathbf{r})$ bestimmt, das im Außenbereich mit dem

⁶Oder Referenzion, falls das Atom in der späteren Rechnung erwartungsgemäß in einer ionischen Form vorkommt. Das ermöglicht es, den Energiebereich, in dem die Streuung akkurat beschrieben wird, leicht zu verschieben.

Rumpfpotential $V_{core}^{eff}(\mathbf{r})$ übereinstimmt und im gesamten Raum die Pseudowellenfunktion $\psi_\nu^{ps}(\mathbf{r})$ als Lösung der Schrödingergleichung mit dem vorgegebenen Eigenwert ε_ν liefert. Dabei wird gefordert, daß die Aufenthaltswahrscheinlichkeit außerhalb des Abschneideradius zwischen ursprünglicher und Pseudowellenfunktion übereinstimmt. D.h. für den Innenbereich gilt wegen der Normierung auch

$$\int_{|\mathbf{r}| < R_c} d^3r |\psi_\nu^{ps}(\mathbf{r})|^2 = \int_{|\mathbf{r}| < R_c} d^3r |\psi_\nu(\mathbf{r})|^2, \quad (2.26)$$

da sonst die elektronische Dichte im Außenbereich verändert würde. Mit dieser wichtigen Forderung der Normerhaltung gelang Bachelet, Chiang, Hamann und Schlüter [HSC79, BHS82] der Durchbruch bei der Verwendung von *ab initio* Pseudopotentialen in der DFT. So fallen betragsmäßig große energetische Anteile in der numerischen Behandlung weg, wie die Energie der Rumpfelektronen (die ja konstant gehalten wird) und die kinetische Energie der Valenzelektronen in Kernnähe, die durch eine entsprechende potentielle Energie aufgehoben wird. Dies senkt den numerischen Fehler erheblich, da selbst ein verhältnismäßig kleiner relativer Fehler der großen Energiebeiträge die energetischen Unterschiede im Valenzbereich erheblich verfälschen würde.

Zur Lösung des atomaren Problems werden die Wellenfunktionen mit Hilfe der Kugelflächenfunktionen $Y_{lm}(\Omega_{\mathbf{r}})$ in der Form

$$\psi_{lm}(\mathbf{r}) = \frac{g_l(r)}{r} Y_{lm}(\Omega_{\mathbf{r}}) \quad (2.27)$$

dargestellt. Dies führt zur radialen Schrödingergleichung

$$\left(-\partial_r^2 + \frac{l(l+1)}{r^2} + V(r) \right) g_l(r) = \varepsilon g_l(r). \quad (2.28)$$

In einer Dichtefunktionalrechnung berechnet man so für ein Referenzatom oder -ion sowohl Valenz- als auch Rumpfelektronen. Zu jeder exakten (all electron, ae) radialen Wellenfunktion g_l^{ae} zum Drehimpuls l eines Valenzelektrons des Referenzatoms wird nun eine Pseudo-(ps)-Wellenfunktion g_l^{ps} konstruiert, die stetig differenzierbar ist, außerhalb von R_c mit g_l^{ae} übereinstimmt und folgende Normerhaltung erfüllt:

$$\int_0^{R_c} d^3r |g_l^{ps}|^2 = \int_0^{R_c} d^3r |g_l^{ae}|^2. \quad (2.29)$$

Durch geeignete Parametrisierung von g_l^{ps} im Corebereich $r < R_c$ [Ker80, Van85, RRKJ90, TM90, TM91, VS92] wird dabei eine möglichst glatte, mit wenigen ebenen Wellen darstellbare Pseudowellenfunktion gesucht.

Durch Invertieren der radialen Schrödingergleichung wird nun das abgeschirmte Potential $V_{scr,l}^{ps}(r)$ errechnet, das g_l^{ps} erzeugt:

$$V_{scr,l}^{ps}(r) = \varepsilon_l - \frac{l(l+1)}{r^2} + \frac{\partial_r^2 g_l^{ps}}{g_l^{ps}} \quad (2.30)$$

Da dieses Potential noch den Hartree- und den Austauschkorrelationsanteil für die Ladungsdichte der Pseudovalenzelektronen enthält, die in einer späteren Dichtefunktionalrechnung für den Festkörper neu berechnet und wieder hinzugerechnet werden, werden diese Teile für das Atom abgezogen:

$$V_{ion,l}^{ps}(r) = V_{scr,l}^{ps}(r) - (V_H[n^{ps}(r)] + V_{xc}[n^{ps}(r)]) . \quad (2.31)$$

Damit ist die Kohn-Sham-Gleichung des Pseudoatoms mit der Pseudoladungsdichte $n^{ps}(r)$, die nur Pseudovalenzelektronen berücksichtigt, analog zu der des Referenzatoms mit der vollen Ladungsdichte $n^{ae}(r)$ aller Elektronen:

$$\begin{aligned} \left(-\partial_r^2 + \frac{l(l+1)}{r^2} - \frac{e^2 Z}{r} + V_H[n^{ae}(r)] + V_{xc}[n^{ae}(r)] - \varepsilon_l \right) g_l^{ae} &= 0 \\ \left(-\partial_r^2 + \frac{l(l+1)}{r^2} + V_{ion,l}^{ps}(r) + V_H[n^{ps}(r)] + V_{xc}[n^{ps}(r)] - \varepsilon_l \right) g_l^{ps} &= 0 \end{aligned} \quad (2.32)$$

So erhält man für jeden Drehimpuls l ein Potential $V_{ion,l}^{ps}$, aus dem sich das Gesamtpotential durch Summation über all diese Potentiale ergibt:

$$\mathcal{V}^{ps} = \sum_l V_{ion,l}^{ps}(r) \mathcal{P}_l . \quad (2.33)$$

Dabei ist \mathcal{P}_l der Projektor auf den Unterraum zum Drehimpuls l . Da dieses Potential durch die Drehimpulsoperatoren im Winkelanteil nicht mehr lokal ist, wird es als *semilokales* Potential bezeichnet. Aufgrund der Konstruktion hat das Pseudopotential folgende Eigenschaften:

- es reproduziert die Eigenwerte ε_l der Valenzelektronen des Referenzatoms,
- es reproduziert außerhalb von R_c die Wellenfunktionen ψ_ν der Valenzelektronen,
- die knotenlosen Pseudozustände ψ_ν der Valenzelektronen sind die energetisch niedrigsten,
- sowohl Potential als auch Wellenfunktionen sind innerhalb von R_c möglichst glatt und lassen sich mit wenigen ebenen Wellen darstellen.

Die partielle Core-Korrektur

Man beachte, daß beim Abziehen des Potentials der Pseudovalenzelektronen vom abgeschirmten Potential in (2.31), dem sogenannten unscreening, eine lineare Näherung für das Austausch-Korrelations-Potential angesetzt wird

$$V_{xc}[n_{core}^{ae}(r) + n^{ps}(r)] \approx V_{xc}[n_{core}^{ae}(r)] + V_{xc}[n^{ps}(r)] . \quad (2.34)$$

Diese Näherung ist exakt, wenn sich Rumpfladungsdichte $n_{core}^{ae}(r)$ und Pseudovalenzladungsdichte $n^{ps}(r)$ nicht überlappen. Diese Näherung läßt sich mit Hilfe der

“partiellen core Korrektur” [LFC82] vermeiden. Dabei wird eine partielle Rumpfladungsdichte $n_{pc}(r)$ erstellt, die möglichst glatt ist und im Bereich des Überlapps von $n^{ps}(r)$ und $n_{core}^{ae}(r)$ mit der Rumpfladungsdichte $n_{core}^{ae}(r)$ übereinstimmt. Statt mit der reinen Pseudovalenzladungsdichte $n^{ps}(r)$ wird das “unscreening” in (2.31) dann für V_{xc} mit $n^{ps}(r) + n_{pc}(r)$ durchgeführt:

$$V_{ion,l}^{ps}(r) = V_{scr,l}^{ps}(r) - (V_H[n^{ps}(r)] + V_{xc}[n^{ps}(r) + n_{pc}(r)]) . \quad (2.35)$$

Bei der späteren Verwendung dieses Potentials muß dann die Austauschwechselwirkung für die “korrigierte” Ladungsdichte $n^{ps}(r) + n_{pc}(r)$ berechnet werden, in LDA-Näherung also $V_{xc}(n^{ps}(r) + n_{pc}(r))$.

Die Zusammensetzung des Potentials aus Anteilen $V_{ion,l}^{ps}(r)\mathcal{P}_l$, die in zueinander orthogonalen Unterräumen wirken, erlaubt es, einen rein lokalen Anteil $V_{loc}(r)$ abzuspalten:

$$\mathcal{V}^{ps} = V_{loc}(r) + \sum_l^{l_{max}} \Delta V_l^{SL}(r)\mathcal{P}_l$$

mit $\Delta V_l^{SL}(r) = V_{ion,l}^{ps}(r) - V_{loc}(r)$. (2.36)

Die Entwicklung nach Drehimpulsen wird bei einem gewählten l_{max} abgebrochen. Für die zu $l > l_{max}$ zugehörigen Unterräume wirkt nur das lokale Potential. Da außerhalb von R_c die Potentiale $V_{ion,l}^{ps}(r)$ per Konstruktion übereinstimmen, speziell in der Beschreibung des langreichweitigen Anteils $\sim -e^2 Z_v / r$, wird für $r > R_c$ das Potential $V_{loc}(r) = V_{core}^{eff}(r)$ gewählt, so daß die Anteile $\Delta V_{ion,l}^{ps}(r)\mathcal{P}_l$ nur innerhalb von R_c von Null verschieden sind. Da in den Unterräumen zu $l > l_{max}$ das Potential $V_{loc}(r)$ wirkt (die semilokalen Anteile sind durch die orthogonalen Projektoren $\mathcal{P}_{l \leq l_{max}}$ ausgeblendet), ist es sinnvoll, $V_{loc}(r) = V_{l_{max}+1}$ zu wählen, um so effektiv die Drehimpulsentwicklung bis $l_{max} + 1$ zu berücksichtigen.

Pseudopotentiale in Kleinman-Bylander Form

Die Darstellung des Projektors \mathcal{P}_l ist sehr aufwendig, da man letztlich im Impulsraum die Elemente der Matrix berechnen müßte. Die Anzahl der Rechenschritte zur Anwendung des Operators auf eine Wellenfunktion wäre zudem proportional zur Anzahl $N_{\mathbf{G}}$ der ebenen Wellen zum Quadrat, eine klassischen Eigenwertbestimmung erforderte gar $\mathcal{O}(N_{\mathbf{G}}^3)$ Rechenschritte.

Durch die von Kleinman und Bylander [KB82] vorgeschlagene Form des Pseudopotentials als Projektionsoperator kann dies mit wenigen Dyaden dargestellt werden, wodurch die Berechnung großer Superzellen praktikabel wird. Anstelle der Anteile $\Delta V_l^{SL}(r)$ des semilokalen Potentials treten beim Pseudopotential in Kleinman-Bylander Form die Anteile

$$\Delta \mathcal{V}_l^{KB} = \sum_{l=-m}^m \frac{|\Delta V_l^{SL} \psi_{lm}^{ps}\rangle \langle \psi_{lm}^{ps} \Delta V_l^{SL}|}{\langle \psi_{lm}^{ps} | \Delta V_l^{SL} | \psi_{lm}^{ps} \rangle} , \quad (2.37)$$

mit den Pseudowellenfunktionen des Referenzatoms:

$$\langle \mathbf{r} | \psi_{lm}^{ps} \rangle = \frac{g_l^{ps}(r)}{r} Y_{lm}(\Omega_{\mathbf{r}}) . \quad (2.38)$$

Aufgrund der Orthogonalität der Kugelflächenfunktionen wirkt auch der Anteil $\Delta \mathcal{V}_l^{KB}$ nur auf den Unterraum zum Drehimpuls l . Insbesondere ergibt die Anwendung auf die Referenzwellenfunktion zum gleichen Drehimpuls

$$\Delta \mathcal{V}_l^{KB} \psi_{lm}^{ps}(\mathbf{r}) = \Delta V_l^{SL}(r) \psi_{lm}^{ps}(\mathbf{r}) . \quad (2.39)$$

In dem von den Wellenfunktionen ψ_{lm}^{ps} aufgespannten Unterraum wirken also die semilokale und die Kleinman-Bylander Form gleich. Allerdings ergibt sich bei Einsetzen dieses veränderten Potentials in die Schrödingergleichung natürlich ein neues Eigenwertspektrum und eventuell neue, unphysikalische Lösungen, sogenannte Ghoststates [BK90, GKS90, GSS91, SSO92]. Da das Pseudopotential die Streueigenschaften des Referenzatoms aber nur im wichtigen Energiebereich der Valenzelektronen⁷ richtig beschreiben soll, ist es ausreichend, zu verlangen, daß die Eigenwerte dieser Zustände weit genug über denen des betrachteten Energiebereichs liegen. Projektionspotentiale, bei denen Ghoststates mit Eigenenergien unterhalb denen der Valenzelektronen auftreten, sind nicht brauchbar, denn bei der Suche nach dem Grundzustand werden diese besetzt, oder sie müßten extra ausprojiziert werden⁸. Es gibt jedoch verschiedene Methoden, diese Ghoststates anhand der logarithmischen Ableitung der Wellenfunktionen zu erkennen und durch geeignete Wahl der Parametrisierung des Pseudopotentials zu vermeiden [BK90, GKS90, GSS91, SSO92].

Die Anteile $\Delta \mathcal{V}_l^{KB}$ des Potentials lassen sich nun zerlegen in der Form:

$$\Delta \mathcal{V}_l^{KB} = E_l^{KB} \sum_{m=-l}^l |t_{lm}\rangle \langle t_{lm}| , \quad (2.40)$$

wobei die t_{lm} die normierten Wellenfunktionen ψ_{lm}^{ps} des Referenzatoms sind und die Vanderbilt-Gewichtsfaktoren E_l^{KB} sich errechnen als

$$E_l^{KB} = \frac{\langle \psi_{lm}^{ps} \Delta V_l^{SL} | \Delta V_l^{SL} \psi_{lm}^{ps} \rangle}{\langle \psi_{lm}^{ps} | \Delta V_l^{SL} | \psi_{lm}^{ps} \rangle} = \frac{\int dr (g_l^{ps}(r))^2 (\Delta V_l^{SL}(r))^2}{\int dr (g_l^{ps}(r))^2 \Delta V_l^{SL}(r)} . \quad (2.41)$$

Die Koeffizienten der Wellenfunktionen t_{lm} lassen sich im Impulsraum berechnen. Für ein Atom am Ursprung der Superzelle ergibt sich dann mit Hilfe der sphärischen Besselfunktionen j_l

$$\langle \mathbf{k} + \mathbf{G} | t_{lm} \rangle = \frac{4\pi i^l}{\sqrt{\Omega}} Y_{lm}(\Omega_{\mathbf{k}+\mathbf{G}}) \int_{\infty}^0 j_l(|\mathbf{k} + \mathbf{G}| r) t_l(r) r dr . \quad (2.42)$$

⁷Wie Hamman zeigte [Ham89] müssen als Energie der Referenzzustände nicht notwendigerweise Eigenenergien gewählt werden.

⁸Im Dichtefunktionalprogramm wird man letztlich den Grundzustand der Pseudoelektronen suchen, bzw. die niedrigsten Eigenwerte. Findet man einen Ghoststate in diesem Teil des Eigenwertspektrums, muß dieser identifiziert und bei der Berechnung der Dichte künstlich unbesetzt gehalten werden.

Durch Multiplikation mit $\exp\{i(\mathbf{k} + \mathbf{G})\tau\}$ läßt sich auch das Potential eines um τ verschobenen Atoms darstellen. Durch diese einfache Darstellung wird jede l -Komponente des Pseudopotentials durch eine Funktion $t_l(r)$ und die Vanderbilt-Gewichtsfaktoren E_l^{KB} charakterisiert. Zur Darstellung in ebenen Wellen muß dann pro Atomsorte, l -Komponente und \mathbf{k} -Punkt nur ein Vektor der Länge $N_{\mathbf{G}}$ im Speicher behalten werden. Durch Anmultiplikation der Kugelflächenfunktionen $Y_{lm}(\Omega_{\mathbf{k}+\mathbf{G}})$ erhält man dann die Komponenten der t_{lm} , und durch Anmultiplikation des Faktors $\exp\{i(\mathbf{k} + \mathbf{G})\tau\}$ das verschobene Potential für ein Atom am Ort τ .

2.2 Explizite Darstellung der Energieanteile und Operatoren

Zur Berechnung der totalen Energie ist es nötig, deren Anteile explizit aufzustellen. Die schnelle Fourier-Transformation (FFT) bietet einen einfachen und effizienten Weg, zwischen Impulsraum- und Realraumdarstellung zu wechseln. Daher werden die Anteile der totalen Energie jeweils im günstigsten Raum dargestellt, so daß sie in einer möglichst einfachen Darstellung vorliegen, die eine entsprechend einfache Berechnung erlaubt. Zur Energie des Elektronischen Systems aus (2.11) kommt noch die Wechselwirkung der Ionenrümpfe untereinander hinzu. Die Gesamtenergie setzt sich zusammen:

$$E_{tot} = E_{ion} + T_0 + E_H + E_{xc} + E_{loc} + E_{nl}. \quad (2.43)$$

Bis auf die reine Ionen-Ionen Wechselwirkung E_{ion} sind dies Funktionale der Dichte, die sich gemäß (2.5) aus den Wellenfunktionen $\psi_{\mathbf{k}\nu}$ der (Hilfs-) Valenzelektronen zusammensetzt. Dabei sind die Anteile:

E_{ion} : die Wechselwirkung (WW) der Ionenrümpfe untereinander,

T_0 : die kinetische Energie der Valenzelektronen,

E_H : die Hartree-Energie der WW der Valenzelektronen untereinander,

E_{xc} : die Austausch-Korrelations-Energie der Valenzelektronen,

E_{loc} : der lokale Anteil der WW der Ionenrümpfe mit den Valenzelektronen und

E_{nl} : der nichtlokale Teil der WW zwischen Ionenrümpfen und Valenzelektronen.

Die **Coulomb-Wechselwirkung der Ionen** miteinander wird als Ewald-Energie bezeichnet. Da die Rumpfladungen einander nicht überlappen können die Ionenrümpfe als Punktladungen beschrieben werden, die sich gegenseitig abstoßen. Jedes Atom μ hat die Atomsorte $\alpha(\mu)$, $Z_{\alpha(\mu)}$ Valenzelektronen und die

Atomkoordinaten τ_μ innerhalb der Superzelle. Nach dem Coulomb-Gesetz erhält man für die Energie der Ionenrümpfe einer Superzelle im Feld der Ionen des gesamten Kristalls durch Summierung über alle Vektoren \mathbf{R} des Bravaisgitters und alle Ionenpaare⁹:

$$E_{ion} = \frac{1}{2} \sum_{\mathbf{R}} \sum_{\mu\mu'} \frac{Z_{\alpha(\mu)} Z_{\alpha(\mu')}}{|\tau_\mu - (\mathbf{R} + \tau_{\mu'})|} \left(1 - \delta_{\mu, \mathbf{R} + \tau_{\mu'}}\right) \quad (2.44)$$

$$= \frac{1}{2} \sum_{\mathbf{G} \neq 0} \sum_{\mu\mu'} \frac{4\pi}{\Omega \mathbf{G}^2} Z_{\alpha(\mu)} Z_{\alpha(\mu')} e^{-i\mathbf{G}(\tau_\mu - \tau_{\mu'})} \quad (2.45)$$

Diese Summen konvergieren in dieser Form beide nicht. Für Gleichung (2.44) im realen Raum liegt das an der Langreichweitigkeit der Coulomb-Wechselwirkung, eine Zelle des Bravais-Gitters hätte im Feld der unendlich ausgedehnten Ladung des Gitters unendliche Energie. Dies läßt sich beheben, indem man die Ladung der Ionenrümpfe mit einer homogen über die Superzelle (und somit das gesamte Gitter) verteilten Ladung neutralisiert. Tatsächlich würden die Ionen ja durch die Valenzelektronen abgeschirmt, dieser Beitrag ist also unphysikalisch. Mathematisch hebt er sich mit analogen Beiträgen bei der Berechnung der Hartree-Energie der Valenzelektronen und dem Beitrag durch das lokale Pseudopotential auf [Ric93]. Die langreichweitige Wechselwirkung entspricht in Gleichung (2.45) in der Darstellung im reziproken Raum dem Summanden für $\mathbf{G} = 0$, daher wird dieser explizit zu Null gesetzt. In Gleichung (2.45) wird allerdings nicht der Fall $\delta_{\mu, \mathbf{R} + \tau_{\mu'}} = 1$ ausgeschlossen, daher divergieren dort die Beiträge für $\mu = \mu'$.

Bei der Ewald-Summation [Ewa21] wird daher ein konvergenzerzeugender Faktor η eingeführt, der es ermöglicht, die Summation teils im Ortsraum und teils im Impulsraum auszuführen. Dies erlaubt nicht nur die Vermeidung der Divergenzen, sondern führt auch zu einer schnellen Konvergenz der Summe, da dabei die Summanden exponentiell verschwinden:

$$E_{ion} = \frac{1}{2} \sum_{\mu\mu'} Z_{\alpha(\mu)} Z_{\alpha(\mu')} X_{\tau_{\mu\mu'}} \quad \text{mit} \quad \tau_{\mu\mu'} = \tau_\mu - \tau_{\mu'} \quad \text{und} \quad (2.46)$$

$$X_{\tau_{\mu\mu'}} = \frac{4\pi}{\Omega} \sum_{\mathbf{G} \neq 0} \frac{e^{-\mathbf{G}^2/4\eta^2}}{\mathbf{G}^2} \cos(\mathbf{G}\tau_{\mu\mu'}) - \frac{2\eta}{\sqrt{\pi}} \delta_{\tau_{\mu\mu'}} \quad (2.46)$$

$$+ \sum_{\mathbf{R}} \frac{\text{erfc}(\eta |\tau_{\mu\mu'} - \mathbf{R}|)}{|\tau_{\mu\mu'} - \mathbf{R}|} \left(1 - \delta_{\tau_{\mu\mu'}, \mathbf{R}}\right) - \frac{\pi}{\eta^2 \Omega} \quad (2.47)$$

⁹Mit $\sum_{\mu\mu'}$ werden die Ionenpaare tatsächlich doppelt gezählt, daher der Faktor 1/2. Um die Wechselwirkung des Ions mit sich selbst auszuschließen muß in der Summation der Fall $\delta_{\mu, \mathbf{R} + \tau_{\mu'}} = 1$ ausgeschlossen werden.

Die Ewald-Energie ist dabei unabhängig von η , daher wird ein η so gewählt, daß die Summen in (2.46) und (2.47) etwa gleich schnell konvergieren.

Die **kinetische Energie** läßt sich mit Hilfe des Laplace-Operators im Ortsraum als $-\partial_{\mathbf{r}}^2$ (angewandt auf dem gesamten Volumen Ω der Superzelle) schreiben. Wendet man dies auf die Wellenfunktion in der Darstellung (2.24) an, so erhält man im reziproken Raum:

$$\begin{aligned} T_0[n] &= \langle \psi | -\partial_{\mathbf{r}}^2 | \psi \rangle_{\Omega} \\ &= \sum_{\mathbf{k}, \nu} f_{\mathbf{k}\nu} \langle \psi_{\mathbf{k}\nu} | -\partial_{\mathbf{r}}^2 | \psi_{\mathbf{k}\nu} \rangle_{\Omega} \\ &= \sum_{\mathbf{k}, \nu} \sum_n f_{\mathbf{k}\nu} |\mathbf{k} + \mathbf{G}_n|^2 c_{\mathbf{k}\nu}^n * c_{\mathbf{k}\nu}^n \end{aligned} \quad (2.48)$$

Im reziproken Raum kann man daher die kinetische Energie in Diagonalform darstellen:

$$\begin{aligned} \langle \mathbf{k} + \mathbf{G} | T_0 | \psi_{\mathbf{k}\nu} \rangle &= |\mathbf{k} + \mathbf{G}_n|^2 c_{\mathbf{k}\nu}^n \\ \langle \mathbf{k} + \mathbf{G}_n | T_0 | \mathbf{k}' + \mathbf{G}_{n'} \rangle &= |\mathbf{k} + \mathbf{G}_n|^2 \delta_{\mathbf{k}, \mathbf{k}'} \delta_{n, n'} \end{aligned} \quad (2.49)$$

Die Hartree-Energie, der lokale Anteil des Pseudopotentials und die Austausch-Korrelations-Energie werden im Ortsraum beschrieben. Das **Hartree-Potential** V_H wird durch die Poisson-Gleichung bestimmt:

$$\partial_{\mathbf{r}}^2 V_H(\mathbf{r}) = -8\pi n(\mathbf{r}). \quad (2.50)$$

Da der Laplaceoperator diagonal im reziproken Raum ist, vereinfacht sich dort die Darstellung zu

$$\begin{aligned} V_H(\mathbf{G}) &= \frac{1}{\Omega} \frac{8\pi}{\mathbf{G}^2} n(\mathbf{G}) \\ \text{mit} \quad V_H(\mathbf{G}) \frac{1}{\Omega} \int d^3\mathbf{r} V_H(\mathbf{r}) e^{i\mathbf{G}\mathbf{r}}, \quad n(\mathbf{G}) \frac{1}{\Omega} \int d^3\mathbf{r} n(\mathbf{r}) e^{i\mathbf{G}\mathbf{r}}. \end{aligned} \quad (2.51)$$

Dabei tritt auch in (2.51) wieder, wie schon bei der Ewald-Energie, ein divergenter Term für $\mathbf{G} = \mathbf{0}$ auf. Auch hier kommt dieser Term wieder aufgrund der langreichweitigen Coulombwechselwirkung zustande und wird zu Null gesetzt. Die Anwendung des Hartree-Potentials im reziproken Raum führt zu der Konvolution

$$\langle \mathbf{k} + \mathbf{G} | V_H | \psi_{\mathbf{k}\nu} \rangle = \sum_{\mathbf{G}'} V_H(\mathbf{G} - \mathbf{G}') c_{\mathbf{k}\nu}(\mathbf{G}'). \quad (2.52)$$

Die Matrixdarstellung des Hartree-Potentials im reziproken Raum ist also eine voll besetzte Matrix:

$$\langle \mathbf{k} + \mathbf{G} | V_H | \mathbf{k} + \mathbf{G}' \rangle = V_H(\mathbf{G} - \mathbf{G}'). \quad (2.53)$$

Um die Konvolutionssumme in (2.52) zu vermeiden, wird das Hartree-Potential Fourier-transformiert und im Ortsraum angewandt, dort nimmt es Diagonalform an:

$$\langle \mathbf{r} | V_H | \psi_{\mathbf{k}\nu} \rangle = V_H(\mathbf{r}) \psi_{\mathbf{k}\nu}(\mathbf{r}). \quad (2.54)$$

Die **Austauschwechselwirkung** in Lokaler-Dichte-Näherung ist lokal, und damit auch diagonal im Ortsraum. Nachdem das Potential $V_{xc}(\mathbf{r}) = V_{xc}(n(\mathbf{r}))$ gemäß (2.23) aus der Dichte $n(\mathbf{r})$ bestimmt wurde, kann es einfach anmultipliziert werden:

$$\langle \mathbf{r} | V_{xc} | \psi_{\mathbf{k}\nu} \rangle = V_{xc}(\mathbf{r}) \psi_{\mathbf{k}\nu}(\mathbf{r}). \quad (2.55)$$

Der **lokale Anteil des Pseudopotentials** eines Ionenrumpfes der Atomsorte $\alpha(\mu)$ am Ort τ_μ ist durch das Potential $V_{loc}^{\alpha(\mu)}(|\mathbf{r} - \tau_\mu|)$ gegeben. Die Translation \mathcal{T}_{τ_μ} um τ_μ läßt sich im reziproken Raum einfach durch die Anmultiplikation der Strukturfaktoren $T_{\tau_\mu} = \exp\{i(\mathbf{k} + \mathbf{G})\tau_\mu\}$ bewerkstelligen. Bei der Darstellung des lokalen Potentials muß berücksichtigt werden, daß die Atomrümpfe periodisch an den Orten $\mathbf{R} + \tau$ im Bravaisgitter angeordnet sind. Deren periodisch angeordneten, radialsymmetrischen Potentiale müssen aufsummiert werden. Im reziproken Raum erhält man:

$$V_{loc}(\mathbf{r}) = \sum_{\mathbf{G}} V_{loc}(\mathbf{G}) e^{-i\mathbf{G}\mathbf{r}} \quad (2.56)$$

$$V_{loc}(\mathbf{G}) = \sum_{\mu} V_{loc}^{\mu}(\mathbf{G}) \quad (2.57)$$

$$\begin{aligned} V_{loc}^{\mu}(\mathbf{G}) &= \frac{1}{\Omega} \int_{\Omega} d^3r \left\{ \sum_{\mathbf{R}} V_{loc}^{\alpha(\mu)}(|\mathbf{r} - (\mathbf{R} + \tau_\mu)|) \right\} e^{i\mathbf{G}\mathbf{r}} \\ &= \frac{1}{\Omega} \int_{\Omega} d^3r \sum_{\mathbf{R}} \mathcal{T}_{\mathbf{R}} V_{loc}^{\alpha(\mu)}(|\mathbf{r}|) e^{i\mathbf{G}(\mathbf{r} + \tau_\mu)} \\ &= \frac{1}{\Omega} \int_{\sum_{\mathbf{R}} \mathcal{T}_{\mathbf{R}}\Omega} d^3r V_{loc}^{\alpha(\mu)}(|\mathbf{r}|) e^{i\mathbf{G}\mathbf{r}} e^{i\mathbf{G}\tau_\mu} \\ &= \left\{ \frac{1}{\Omega} \int_{\infty} d^3r V_{loc}^{\alpha(\mu)}(|\mathbf{r}|) e^{i\mathbf{G}\mathbf{r}} \right\} S^{\tau_\mu}(\mathbf{G}) \\ &= V_{loc}^{\alpha(\mu)}(\mathbf{G}) S^{\tau_\mu}(\mathbf{G}) \end{aligned} \quad (2.58)$$

$$\text{mit } V_{loc}^{\alpha(\mu)}(\mathbf{G}) = \frac{1}{\Omega} \int_{\infty} d^3r V_{loc}^{\alpha(\mu)}(|\mathbf{r}|) e^{i\mathbf{G}\mathbf{r}}$$

$$\text{und } S^{\tau_\mu}(\mathbf{G}) = e^{i\mathbf{G}\tau_\mu}.$$

Der lokale Anteil des Pseudopotentials läßt sich also im reziproken Raum gemäß (2.58) in die Formfaktoren $V_{loc}^{\alpha(\mu)}(\mathbf{G})$ und die Strukturfaktoren $S^{\tau_\mu}(\mathbf{G})$ zerlegen. Dabei hängen die Formfaktoren nur von der Atomsorte ab, $V_{loc}^{\alpha(\mu)}(\mathbf{G})$ entspricht

dem Potential, das entsteht, wenn man Atome der Sorte $\alpha(\mu)$ an den Punkten \mathbf{R} des Bravaisgitters anordnet. Da dieser Anteil nur von $\alpha(\mu)$ abhängt, muß er nur einmal bestimmt werden. Da $V_{loc}^\alpha(r)$ radialsymmetrisch ist, findet man durch Entwicklung der ebenen Wellen $e^{i\mathbf{G}\mathbf{r}}$ nach sphärischen Besselfunktionen $j_l(Gr)$ mit $G = |\mathbf{G}|$:

$$\begin{aligned} V_{loc}^\alpha(\mathbf{G}) &= \frac{4\pi}{\Omega} \begin{cases} \int_\infty^0 r^2 dr V_{loc}^\alpha(r)(Gr) j_0(Gr) & \text{für } G > 0 \\ \int_\infty^0 r^2 dr V_{loc}^\alpha(r)(Gr) & \text{für } G = 0 \end{cases} \quad (2.59) \\ j_0(Gr) &= \frac{\sin(Gr)}{Gr} \end{aligned}$$

Das Pseudopotential $V_{loc}(\mathbf{G})$ ist wieder ein lokales Potential, das bedeutet, dass es im Ortsraum diagonal ist, die Anwendung im reziproken Raum aber wieder zu einer Konvolution wie in (2.52) führt. Es wird daher mittels (2.56) in den Ortsraum transformiert und dort durch einfaches Anmultiplizieren angewandt:

$$\langle \mathbf{r} | V_{loc} | \psi_{\mathbf{k}\nu} \rangle = V_{loc}(\mathbf{r}) \psi_{\mathbf{k}\nu}(\mathbf{r}). \quad (2.60)$$

Wie bei der Hartree- und der Ewald-Energie tritt auch in (2.59) für $\mathbf{G} = \mathbf{0}$ eine Divergenz aufgrund der langreichweitigen Coulombwechselwirkung auf. Diese divergenten Terme heben sich gegenseitig auf. Übrig bleibt eine Konstante, die nur von den verwendeten Pseudopotentialen abhängt [Ric93]. Dies ist auch physikalisch einsichtig, da die Ladungsdichte der Ionenrümpfe ja durch die im Mittel gleich große Ladungsdichte der Valenzelektronen abgeschirmt wird.

Der **nichtlokale Anteil des Pseudopotentials** wird entsprechend (2.40) als Summe dyadischer Produkte dargestellt, diese ließen sich sowohl im Ortsraum wie im reziproken Raum aufstellen. Wie beim lokalen Anteil des Pseudopotentials läßt sich auch bei den nichtlokalen Anteilen ähnlich wie in (2.58) die Abhängigkeit von der Verschiebung τ_μ als Strukturfaktor abspalten. Daher ist es günstiger, diese Dyaden im reziproken Raum darzustellen.

Um die Verschiebung τ zu berücksichtigen, wird diese auf die dyadischen Summanden in (2.40) angewandt:

$$\mathcal{T}_\tau |t_{lm}\rangle \langle t_{lm}| = |\mathcal{T}_\tau t_{lm}\rangle \langle t_{lm} \mathcal{T}_\tau|. \quad (2.61)$$

In reziproker Darstellung läßt sich dann der Translationsoperator als Strukturfaktor darstellen

$$\begin{aligned} \langle \mathbf{k} + \mathbf{G} | \mathcal{T}_\tau t_{lm} \rangle &= \langle \mathbf{k} + \mathbf{G} \mathcal{T}_\tau | t_{lm} \rangle \\ &= e^{i(\mathbf{k} + \mathbf{G})\tau} \langle \mathbf{k} + \mathbf{G} | t_{lm} \rangle \\ &= e^{i\mathbf{k}\tau} S^\tau(\mathbf{G}) \langle \mathbf{k} + \mathbf{G} | t_{lm} \rangle. \end{aligned} \quad (2.62)$$

Da der Faktor $\exp\{i\mathbf{k}\tau\}$ in (2.61) immer paarweise mit seinem konjugiert komplexen auftritt, kann man ihn herauskürzen und das nichtlokale Potential unter

Verwendung von (2.42) schreiben:

$$\begin{aligned}
\langle \mathbf{k} + \mathbf{G} | \mathcal{V}_{KB} | \mathbf{k} + \mathbf{G}' \rangle &= \sum_{\mu} \sum_{lm} E_l^{\alpha(\mu)} T_{lm}^{\mu}(\mathbf{k}, \mathbf{G}) T_{lm}^{\mu *}(\mathbf{k}, \mathbf{G}') \\
T_{lm}^{\mu}(\mathbf{k}, \mathbf{G}) &= S^{\tau_{\mu}}(\mathbf{G}) Y_{lm}(\Omega_{\mathbf{k}+\mathbf{G}}) t_l^{\alpha(\mu)}(\mathbf{k} + \mathbf{G}) \\
t_l^{\alpha(\mu)}(\mathbf{k} + \mathbf{G}) &= \frac{4\pi}{\sqrt{\Omega}} \int_{\infty}^0 j_l(|\mathbf{k} + \mathbf{G}|r) t_l^{\alpha(\mu)}(r) r dr . \quad (2.63)
\end{aligned}$$

2.3 Das EStCoMPP-Programm

Um nun tatsächlich einen Festkörper zu simulieren, ist es notwendig, die vorgehend dargelegten Methoden in Form eines Programmes zu implementieren. Im Folgenden soll nun die Struktur des EStCoMPP-Programmes vorgestellt werden, das eine solche Implementation darstellt. Das EStCoMPP-Programm ermöglicht unter anderem die Berechnung einer atomaren Struktur mit minimaler Energie (lokales Minimum), die sich aus einer vorgegebenen Konfiguration von Atomen in einer Superzelle durch Relaxation ergibt. Dazu werden zu einer gegebenen Konfiguration von Atomen jeweils die Kräfte auf alle Atomrümpfe berechnet und diese dann entsprechend der Kräfte verschoben, bis das System kräftefrei ist. Um im hochdimensionalen Raum der Atomkoordinaten τ_{μ} aller Atome μ schneller zum energetischen Minimum zu gelangen, werden Quasi-Newton-Methoden angewandt [Eng95]. Diese Molekulardynamikschritte werden so oft iteriert, bis die Kräfte hinreichend klein (entsprechend einem vorgegebenen Kriterium) werden.

Zur Berechnung des Anteils der Kräfte, der durch die Valenzelektronen verursacht wird, ist es zunächst nötig, zu den vorgegebenen Atomkoordinaten τ_{μ} den Grundzustand der Valenzelektronen zu finden. Dazu muß im Formalismus der Dichtefunktionaltheorie das Selbstkonsistenzproblem gelöst werden. Auch dies wird iterativ gelöst: Ausgehend von einer Startladungsdichte¹⁰ $n^{ps,0}$ errechnet man das zugehörige Startpotential, zu dem der Grundzustand der Pseudoelektronen gefunden wird. Daraus ergibt sich wiederum eine neue Ladungsdichte $n^{ps,1}$. Dies wird iteriert, bis schließlich alte und neue Dichte übereinstimmen: $n^{ps,i} = n^{ps,i+1}$. Auch hier bedient man sich wieder Pseudo-Newton-Methoden, um schnellere Konvergenz zu erzielen. Die Anzahl der Iterationen bis zur Selbstkonsistenz hängt dabei stark vom System ab. Während für Systeme mit Bandlücken, also z.B. Halbleiter, typischerweise ca. 20 Schritte genügen, werden für metallische Systeme eher 100 Schritte benötigt¹¹.

Zur Berechnung der Elektronendichte müssen in jedem Schritt der elektronischen Selbstkonsistenz die Kohn-Sham-Gleichungen gelöst werden. Aus den

¹⁰z.B. einer gaußförmigen Verteilung der Valenzelektronendichte um die Atomrümpfe.

¹¹Fällt die Fermienergie in eine Bandlücke, so werden sich die Besetzungszahlen (fast) nicht verändern. In Metallen hängen diese Besetzungszahlen, für Zustände nahe der Fermienergie, jedoch stark von ε_{ν} ab. Dies kann auch so interpretiert werden, daß durch die Fermioberfläche in Metallen zusätzliche Freiheitsgrade berücksichtigt werden müssen.

Abbildung 2.2: EStCoMPP — *ab initio* Molekulardynamik mit DFT, in Klammern sind die Namen der jeweiligen Subroutine angegeben.

Dimensionierung und Initalisierung entsprechend den Parametern des Problems (dim, setup)	
Berechnung der Ewald Energie und Kraft (mfewld)	
Initialisiere die Wellenfunktionen $ \psi_\nu\rangle$ für $\nu = 1 \dots M$, falls nötig (trdhmd)	
Berechne das lokale Potential und die Strukturfaktoren (prpvlc)	
Für jeden k -Punkt:	
Berechne Diagonalmatrix der kinetischen Energie (prpham)	
Berechne den Diagonaleil $H_{diag} = H_{ii}$ für $i = 1 \dots N$ (genhdg)	
Bilde $ \psi'_\nu\rangle = H \psi_\nu\rangle$ für $\nu = 1 \dots M$ (accwvf)	
Schätze Eigenwerte $\varepsilon_{k\nu} = \langle \psi'_\nu \psi_\nu \rangle$ für $\nu = 1 \dots M$	
Drei Iterationen des Davidson Eigenwertlösers:	
Bestimme den Rest $ \psi'_\nu\rangle = \psi'_\nu\rangle - \varepsilon_{k\nu} \psi_\nu\rangle$ für $\nu = 1 \dots M$	
Bestimme den Fehler $ \psi'_\nu ^2$ für $\nu = 1 \dots M$	
Berechne Korrekturvektor $ \psi_{\nu+M}\rangle_i = \frac{1}{H_{ii} - \varepsilon_{k\nu}} \psi'_\nu\rangle_i$ für $\nu = 1 \dots M; i = 1 \dots N$	
Orthonormiere die $ \psi_\nu\rangle$ für $\nu = 1 \dots 2M$ (clgrsh)	
$ \psi'_\nu\rangle = H \psi_\nu\rangle$ für $\nu = 1 \dots 2M$ (accwvf)	
Projiziere H auf den Unterraum: $H'_{ij} = \langle \psi_i \psi'_j \rangle$ für $i, j = 1 \dots 2M$	
Löse das Unterraumproblem: $H'_{ij} \Phi_\nu\rangle_j = \varepsilon_{k\nu} \Phi_\nu\rangle_i$ für $\nu = 1 \dots M; i, j = 1 \dots 2M$	
Rücktransformation: $ \psi_\nu^{(l)}\rangle = \sum_i \psi_i^{(l)}\rangle \Phi_\nu\rangle_i$ für $\nu = 1 \dots M$ und $i = 1 \dots 2M$	
Besetze die Zustände nach der Fermiverteilung (efrghi)	
Bestimme die neue Ladungsdichte und das von ihr erzeugte Potential (chdpot)	
Mische das alte und das neue Potential bzw. Ladungsdichte (... mix)	
Bis eine selbstkonsistente Ladungsdichte erreicht ist (eldgfr)	
Berechne die Kräfte auf Atome aufgrund der Elektronenverteilung (force)	
Bewege die Atome im Quasi-Newtonverfahren (mdyns)	
Molekulardynamik, bis die Kräfte auf Ionenrümpfe vernachlässigbar klein sind	

Lösungen wird die Grundzustandsdichte der Pseudovalenzelektronen zum Hamiltonoperator $\mathcal{H}[n^{ps,i+1}]$ konstruiert. Wie schon gesagt, skaliert die vollständige Diagonalisierung von \mathcal{H} ungünstig mit der Anzahl $N_{\mathbf{G}}$ der \mathbf{G} -Vektoren. Zudem werden zur Berechnung der Dichte und der elektronischen Gesamtenergie nur die besetzten Zustände (also die Eigenzustände zu den niedrigsten Energieeigenwerten) benötigt. Zu deren Berechnung wird das in Anhang B beschriebene iterative Davidson Verfahren angewandt [Ric93], bei dem das Eigenwertproblem nur in einem sehr viel kleineren Unterraum der Dimension M_{dvd} gelöst wird. Durch Anwendung des Hamiltonoperators auf die Wellenfunktionen ψ_{ν} werden neue Wellenfunktionen $\mathcal{H}\psi_{\nu}$ gefunden. Dann wird der Hamiltonoperator auf den von ψ_{ν} und $\mathcal{H}\psi_{\nu}$ aufgespannten Unterraum projiziert und dort diagonalisiert. Die so gefundenen Eigenvektoren des Unterraumproblems bilden dann den Ausgangspunkt der nächsten Iteration.

In jedem Schritt der Molekulardynamik muß also die elektronische Selbstkonsistenz gelöst werden. Dazu müssen wiederum in jedem Schritt der Selbstkonsistenz die Kohn-Sham-Gleichungen gelöst werden. Auch diese werden dann wiederum in einem iterativen Verfahren gelöst. Damit ergibt sich eine dreifach verschachtelte Schleifenstruktur, in deren innerster Schleife letztlich die einzelnen Davidson-Schritte ausgeführt werden. Das Verfahren ist in Abbildung 2.2 veranschaulicht.

Aufgrund der verschachtelten Programmstruktur entfällt der Hauptanteil der Berechnungen auf die innerste Schleife. Da das Eigenwertproblem wegen der Orthogonalität der ebenen Wellen separiert, wird sie für jeden \mathbf{k} -Punkt pro Molekulardynamik-Schritt für Halbleitersysteme ca. 60 mal ausgeführt. Indem in jedem Selbstkonsistenzschritt nur drei Davidson-Iterationen ausgeführt werden, konvergieren Eigenvektoren und Selbstkonsistenz gemeinsam. Dies vermeidet eine unnötig genaue Berechnung der Eigenvektoren bei schlecht konvergierter Selbstkonsistenz. Um den Grundzustand der Atompositionen zu erreichen, bedarf es ca. 20-40 Molekulardynamikiterationen. Dies ergibt $\mathcal{O}(1000)$ Iterationen pro \mathbf{k} -Punkt. Allerdings benötigt man für größere Probleme und entsprechend höher dimensionalen Lösungsraum auch oft mehr Iterationen bis zum Erreichen der Konvergenz.

2.4 Skalierungsverhalten

Für ein konkretes Problem muß man zunächst abschätzen, wie groß der Rechenaufwand und die Speicheranforderungen sind. So läßt sich entscheiden, ob das Problem lösbar ist, auf welcher Hardware es berechnet werden kann und wieviel Rechenzeit benötigt wird. Letztlich sind für Rechenaufwand und Speicheranforderungen einige wenige Programmparameter wesentlich, die sich grob in zwei Gruppen unterteilen lassen. Dies sind zum einen Parameter, die sich direkt auf die physikalischen Eigenschaften des Problems beziehen, im wesentlichen die Größe

Ω der Superzelle und die Anzahl betrachteter Atome. Zum anderen benötigt man Parameter, um die Genauigkeit der Berechnung vorgeben zu können, wie z.B. die Dichte $n_{\mathbf{k}}$ der \mathbf{k} -Punkte in der Brillouinzone oder die durch \mathbf{G}_{max} vorgegebene Ortsauflösung der Wellenfunktionen. Letztlich sind auch diese Parameter wieder von der Physik des Systems abhängig, so benötigt man zur Berechnung metallischer Systeme eine höhere \mathbf{k} -Punktdichte, um die Fermifläche mit hinreichender Genauigkeit festlegen zu können. Der Parameter \mathbf{G}_{max} ist meist durch die verwendeten Pseudopotentiale vorgegeben, die ihrerseits von den Atomsorten abhängen. Aus diesen Größen ergeben sich folgende Parameter des **EStCoMPP**-Programms, die im wesentlichen die Zeit- und Speicheranforderungen bestimmen:

$$\begin{aligned}
& \text{die Anzahl der } \mathbf{k}\text{-Punkte:} & N_{\mathbf{k}} \sim n_{\mathbf{k}}/\Omega; , \\
& \text{die Zahl der Wellenfunktionen pro } \mathbf{k}\text{-Punkt:} & M \sim N_{at} , \\
& \text{die Dimension des Davidson-Unterraumes:} & M_{dvd} \sim N_{at} , \\
& \text{die Anzahl der } \mathbf{G}\text{-Vektoren:} & N \sim \mathbf{G}_{max}^3 \Omega , \\
& \text{und die Anzahl der Atome:} & N_{at} .
\end{aligned}$$

Dabei gibt M die Anzahl berechneter Zustände an¹². Als Dimension des Davidson-Unterraums M_{dvd} wird ein Wert zwischen $2M$ und $3M$ gewählt. Ein größerer Wert für M_{dvd} verbessert die Konvergenz des Eigenproblems, führt aber auch zu längeren Rechenzeiten pro Davidson-Schritt und höheren Anforderungen an den Speicher. Bei einem Halbleiter, z.B. Silizium, mit vier Valenzelektronen pro Atom, kommt man wegen Spinartung auf¹³ $M \geq 2.1N_{at}$ und $M_{dvd} \geq 2M \geq 4.2N_{at}$. Mit $M_{dvd} = 2.5M \approx 5.5N_{at}$ läßt sich allerdings bei akzeptablem Mehraufwand ein verbessertes Konvergenzverhalten erzielen.

In jeder der oben angesprochenen Davidson-Iterationen¹⁴ werden zunächst neue Korrekturvektoren berechnet, welche gemeinsam mit den Ausgangsvektoren orthonormiert werden. Indem der Hamiltonoperator auf jeden dieser Vektoren angewandt wird, läßt sich \mathcal{H} auf den von diesen Vektoren aufgespannten Unterraum projizieren. Dort wird der projizierte Hamiltonoperator \mathcal{H}' diagonalisiert, und die Eigenvektoren werden auf den gesamten Raum zurücktransformiert. Die Anwendung des Hamiltonians auf eine Wellenfunktion ψ läßt sich noch in drei Anteile

¹²Mindestens die Zahl aller (auch nur teilweise) besetzten Zustände. M wird typischerweise ca. 5-25% größer gewählt als die halbe Zahl (wegen Spinartung) der Valenzelektronen. So werden alle, auch nur teilweise besetzten Zustände, bzw. bei Halbleitern auch einige Zustände oberhalb der Bandlücke erfasst. Die Wahl eines etwas größeren M begünstigt die Konvergenz des Eigenvektorproblems (siehe Anhang B), da so in späteren Davidson-Schritten mehr Startvektoren zur Verfügung stehen.

¹³4 Atome pro Si-Atom, $\div 2$ wegen Spinartung, $+5\%$ um einige Zustände oberhalb der Bandlücke zu berücksichtigen.

¹⁴Das im **EStCoMPP**-Programm verwendete Davidson-Kosugi-Verfahren wird in Anhang B genauer erläutert.

unterteilen, entsprechend dem lokalen Potential $V_{loc}(\mathbf{r})$, dem kinetischen Anteil $T_0(\mathbf{G} + \mathbf{k})$ und dem nichtlokalen Potential \mathcal{V}_{nl} , das in Kleinman-Bylander-Form durch Dyaden dargestellt wird. Da die Zustände in der Form der Koeffizienten $c_{\mathbf{k}\nu}^n$ im Impulsraum vorliegen, ist die Anwendung des kinetischen Anteils trivial. Zur Anwendung des lokalen Potentials bedient man sich der schnellen Fourier-Transformation, die für Vektoren der Länge N mit $N \log(N)$ skaliert¹⁵. Zur Anwendung des nichtlokalen Potentials müssen (für $l_{max} = 1$) pro Atom vier Dyaden anmultipliziert werden. Betrachten wir nun den Rechenaufwand aller Operationen, die pro Davidson-Schritt ausgeführt werden und ungünstiger als $N_{\mathbf{k}}NM_{dvd}$ skalieren:

Orthonormierung der Wellenfunktionen:	$N_{\mathbf{k}}NM_{dvd}^2$
Nichtlokaler Anteil des Potentials:	$N_{\mathbf{k}}N \log(N)M_{dvd}$
Lokaler Anteil des Potentials:	$N_{\mathbf{k}}NN_{at}M_{dvd}$
Projektion auf Unterraum:	$N_{\mathbf{k}}NM_{dvd}^2$
Lösung des Unterraumproblems:	$N_{\mathbf{k}}M_{dvd}^3$
Rücktransformation aus Unterraum:	$N_{\mathbf{k}}NM_{dvd}^2$

Als speicherintensivste Größen des Programms stellen sich die Koeffizienten der Wellenfunktionen $c_{\mathbf{k}\nu}^n$ und die Dyaden des nichtlokalen Potentials heraus. Einmaliges Ausrechnen und Vorhalten der Dyaden würde (bei $l_{max} = 1$) das Abspeichern von $4N_{\mathbf{k}}N_{at}N$ komplexen Zahlen verlangen. Tatsächlich werden die Dyaden aber bei Bedarf erzeugt, so daß pro Atomsorte und Drehimpuls nur eine Dyade gespeichert werden muß. Damit bleiben als speicherintensivste Größe die Koeffizienten der Wellenfunktionen, die den Davidson-Unterraum aufspannen. Dazu werden $N_{\mathbf{k}}NM_{dvd}$ komplexe Zahlen benötigt. Da nur die Daten für den jeweils berechneten \mathbf{k} -Punkt im Hauptspeicher gehalten werden, indem die Wellenfunktionen der anderen \mathbf{k} -Punkte auf einem externen Speichermedium gehalten werden, beanspruchen die Wellenfunktionen nur für NM_{dvd} komplexe Zahlen Speicherplatz. Allerdings ist es zur Effizienzsteigerung sinnvoll, nicht nur die Wellenfunktionen $\psi_{\mathbf{k}\nu}$ abzuspeichern, sondern zusätzlich auch $\mathcal{H}\psi_{\mathbf{k}\nu}$, um diese Funktionen nicht jedesmal erneut berechnen zu müssen. Da die Dimension M_{dvd} des Davidson-Unterraums für Halbleiter ca. $5N_{at}$ ist, kommt man unter Berücksichtigung einiger Hilfsfelder auf ca. $12N_{at}N$ abzuspeichernde Koeffizienten. Bei einem Beispiel mit 190 Atomen (Arsen-bedeckte Siliziumoberfläche mit $\mathbf{G}_{max} = 3.7$ und 20% Vakuum) erhält man $N = 30000$ Koeffizienten pro Wellenfunktion. Mit 16 Byte pro komplexer Zahl ergibt $12N_{at}N$ etwa ein GByte Speicherbedarf allein für die Koeffizienten $c_{\mathbf{k}\nu}^n$. Hinzu kommen Größen wie z.B. die Formfaktoren des nichtlokalen Potentials, die jeweils N Koeffizienten, also etwa 0.5MByte, benötigen. Weiter werden bei der Anwendung des lokalen Potentials Hilfsfelder zur Real-

¹⁵Im Gegensatz zur Anmultiplikation an die $N \times N$ -Matrixdarstellung des lokalen Potentials im Impulsraum, die mit N^2 skaliert.

raumdarstellung gebraucht, die grob $15N$ (entsprechend 7MByte) Koeffizienten benötigen¹⁶.

¹⁶Da bei der Fourier-Transformation eine Box benutzt wird, deren Seitenlänge das Vierfache des Abscheideradiuses \mathbf{G}_{max} beträgt, werden $4^3/(4\pi/3) \times N \approx 15 \times N$ Koeffizienten in Realraumdarstellung benötigt.

Kapitel 3

Parallelisierung

Die Ausnutzung modernster Supercomputer gestattet nicht nur eine quantitative, sondern auch eine qualitative Erweiterung der mittels **ES**tCoMPP berechenbaren Systeme, da es möglich wird, Superzellen mit so großer lateraler Ausdehnung zu untersuchen, daß die vorgegebene Periodizität vernachlässigbar wird. Dies gestattet zum Beispiel die Betrachtung von Adatomen auf Oberflächen und an Stufenkanten, da die periodisch angeordneten Adatome weit genug auseinanderliegen, so daß die gegenseitige Beeinflussung verschwindend gering ist.

Supercomputing bedeutet heute den Einsatz massiv paralleler Systeme. Aus Gründen der Skalierbarkeit handelt es sich dabei um Systeme mit verteiltem Speicher, bei denen jeder Rechenknoten (PE) mit lokalem Speicher versehen ist. Diese Rechenknoten¹ kommunizieren über ein Verbindungsnetzwerk. Deren Benutzung macht allerdings die Anpassung (“Parallelisierung”) des jeweiligen Programmes an diese Rechner nötig, da die notwendige Kommunikation durch Aufruf spezieller Prozeduren gestaltet werden muß. Abgesehen von der Verteilung der Rechenzeit muß bei Parallelrechnern mit verteiltem Speicher auch auf die Verteilung der Daten eingegangen werden, da der zur Verfügung stehende Speicher zwar in der Gesamtheit groß ist, aber pro PE nur wenig Speicher verfügbar ist.

Da die Koeffizienten der Wellenvektoren $c_{\mathbf{k}\nu}^n$, die größten Datenfelder im Programm beanspruchen und die sie betreffenden Operationen den wesentlichen Anteil der Programmlaufzeit ausmachen, bietet eine Verteilung der $c_{\mathbf{k}\nu}^n$ und der entsprechenden Berechnungen auf verschiedene PEs die Möglichkeit, sowohl Speicher als auch Rechenleistung gleichermaßen effizient zu verteilen. Gleichzeitig wird, da die von Berechnungen auf einem PE betroffenen Koeffizienten in dessen lokalem Speicher liegen, der Transfer der $c_{\mathbf{k}\nu}^n$ zwischen den PEs vermieden.

Der Großteil der Rechenzeit wird auf die Berechnung des Eigenwertproblems verwandt, die ja im wesentlichen aus der Anwendung des Hamiltonoperators auf die Wellenfunktionen, Orthonormierung und Lösen des Unterraumproblems besteht. Die Verteilung der $c_{\mathbf{k}\nu}^n$ auf die Prozessoren sollte also im Hinblick auf die

¹Im Folgenden werden die Begriffe Prozessor, Rechenknoten und PE synonym verwandt.

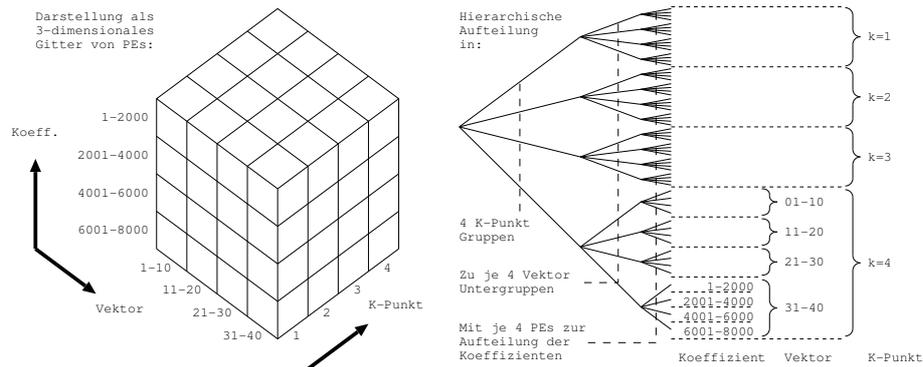


Abbildung 3.1: Aufteilung der Koeffizienten $c_{\mathbf{k}\nu}^n$ auf Prozessoren: Die Datenverteilung kann unter verschiedenen Gesichtspunkten betrachtet werden. Zum einen als dreidimensionales Prozessorgitter (links), wobei jeder Dimension einer der Indices \mathbf{k} , ν und \mathbf{G} zugeordnet ist und in Teilbereiche unterteilt wird. Zum anderen (rechts) lässt sich die Zuordnung auch als Aufteilung in \mathbf{k} -Punkt Gruppen auffassen, die wiederum in Band-Untergruppen unterteilt sind, welche letztlich die Koeffizienten der entsprechenden Vektoren auf einzelne PEs verteilen.

optimale Parallelisierung dieser Teile geschehen.

Entsprechend den drei Indices n , \mathbf{k} und ν ergeben sich nun drei Ansätze zur Parallelisierung, wobei jeweils der Wertebereich des entsprechenden Index in disjunkte Teilbereiche unterteilt wird, die dann entsprechenden Prozessoren (oder Prozessorgruppen) zugeordnet werden. Ein Prozessor bearbeitet dann nur noch $c_{\mathbf{k}\nu}^n$, deren Indices in den ihm zugewiesenen Teilbereichen liegen. Die drei sich so ergebenden Parallelisierungen sind:

- (i) Die \mathbf{k} -Punkt-Parallelisierung:

\mathbf{k} -Punkte werden Prozessorgruppen zugeteilt und \mathbf{k} -Punkt abhängige Daten entsprechend verteilt. Da die Eigenwertprobleme für die einzelnen \mathbf{k} -Punkte unabhängig voneinander sind, ist zu ihrer Lösung keine Kommunikation zwischen PEs verschiedener \mathbf{k} -Punkte nötig.

- (ii) Die Zustands-Parallelisierung:

Prozessorgruppen werden verschiedene Zustände ν zugeordnet. Zur Orthonormierung und zur Lösung des Unterraumproblems müssen Skalarprodukte verschiedener Vektoren gebildet werden, was Kommunikation in erheblichem Maße erforderlich macht. Insbesondere die Orthonormierung ist wegen der Abhängigkeiten der Rechenschritte nicht trivial.

- (iii) Die \mathbf{G} -Vektor-Parallelisierung:

Die \mathbf{G}^n -Vektoren werden Prozessoren zugeordnet, entsprechend werden die Koeffizienten $c_{\mathbf{k}\nu}^n$, der Wellenvektoren und anderer Objekte (Teilchendichte, Dyaden des nichtlokalen Potentials, Operator der kinetischen Energie) über diese Prozessoren verteilt. In diesem Fall ist sowohl für Skalarprodukte wie auch für Fourier-Transformationen Kommunikation notwendig.

Da die einzelnen Parallelisierungsstrategien weitgehend unabhängig voneinander sind, lassen sich alle diese Konzepte in einem Programm implementieren. Dabei wird das Feld der $c_{\mathbf{k}\nu}^n$ mit den drei Dimensionen n , \mathbf{k} und ν auf ein dreidimensionales Feld von Prozessoren abgebildet, wie in Abb. 3.1 zu sehen ist. Man kann sich auch eine hierarchische Struktur vorstellen: Die \mathbf{k} -Punkte werden auf Prozessorgruppen verteilt, die wiederum in Untergruppen aufgeteilt werden, die jeweils einen Teil der Zustände berechnen. Jeder Prozessor dieser Untergruppen beherbergt schließlich nur einen Teil der Koeffizienten $c_{\mathbf{k}\nu}^n$, der ihm zugeordneten Vektoren. Diese Hierarchie orientiert sich an der Struktur des ESTCoMPP-Programms und der Verschachtelung von \mathbf{k} -, ν - und n -Schleifen ineinander: dabei sind \mathbf{k} -Punkt-Schleifen die am weitesten außen liegenden und Schleifen über die Koeffizienten eines Wellenvektors die innersten.

Bei dieser Parallelisierung werden die \mathbf{k} -Punkte auf verschiedene Prozessoren oder Prozessorgruppen verteilt. Diese Parallelisierung ist mit vergleichbar geringem Aufwand zu implementieren, da nur wenige \mathbf{k} -Punkt-Schleifen und Datenfelder betroffen sind. Zudem liegen diese \mathbf{k} -Punkt-Schleifen weit außen. Da die Kommunikation auf Anfang und Ende der parallel abgearbeiteten Programmteile beschränkt ist, läßt sich der dadurch entstehende Overhead weitgehend vermeiden. Zudem wird der wesentliche Rechenaufwand auf die Lösung des Unterraumproblems verwandt, diese wird für jeden \mathbf{k} -Punkt unabhängig durchgeführt und somit parallelisiert. Dies hat auch den weiteren Vorteil, daß an den Routinen des Eigenwertlösers nur wenige Änderungen vorgenommen werden müssen. Allerdings muß schon bei dieser Parallelisierung besonderer Augenmerk auf die Behandlung von Dateien gerichtet werden, um zu gewährleisten, daß jeder Prozessor auf aktuelle Daten zugreift und Datenkorruption durch Caching oder Überschreiben vermieden wird.

3.1 Die \mathbf{k} -Punkt-Parallelisierung

Um die Effizienz weiter zu verbessern und Kommunikation der $c_{\mathbf{k}\nu}^n$ zu vermeiden, wurden alle \mathbf{k} -Punkt-Schleifen parallelisiert, in denen diese Koeffizienten betroffen sind, also z.B. die Generierung der neuen Ladungsdichte $n(\mathbf{r})$ nach einer Iteration der elektronischen Selbstkonsistenz oder die Kraftberechnung. Eine parallelisierte \mathbf{k} -Punktschleife hat dann folgende Form:

- Vor der Schleife werden alle globalen (\mathbf{k} -Punkt unabhängigen) Daten an alle PEs verteilt.
- Auf jedem PE (bzw. jeder Gruppe von PEs) werden die \mathbf{k} -Punkt abhängigen Berechnungen analog zur nichtparallelisierten Version durchgeführt.
- Nach der Schleife werden die Resultate eingesammelt bzw. globale Reduktionen (z.B. Summen) angewandt.

Abbildung 3.2: Elektronik-Schritt — Schritt der elektronischen Selbstkonsistenz

Y	Ist dieser PE in \mathbf{k} -Punkt-Gruppe 0?		N
Berechne V_{loc} (prpvlc)			
Vervielfältige V_{loc} auf alle Prozessoren		Empfange V_{loc} von PE 0	
Verteile $\varepsilon_{\nu,\mathbf{k}}$ nach \mathbf{k} -Punkten an PEs		Empfange $\varepsilon_{\nu,\mathbf{k}}$ für diesen PE	
Für alle diesem PE zugeordneten \mathbf{k} -Punkte			
Berechne kinetischen Anteil für diesen \mathbf{k} -Punkt (prpham)			
Lade Eigenvektoren $c_{\mathbf{k}\nu}^n$ aus Datei			
Führe einige Iterationen des Eigenwertlösers aus (dvdhm)			
Speichere Eigenvektoren $c_{\mathbf{k}\nu}^n$ in Datei			
Y	Ist dieser PE in \mathbf{k} -Punkt-Gruppe 0?		N
Sammele $\varepsilon_{\nu,\mathbf{k}}$ PEs von anderen PEs ein		Schicke $\varepsilon_{\nu,\mathbf{k}}$ an PE 0	

In Abb. 3.2 ist dies am Beispiel der elektronischen Selbstkonsistenz dargestellt. Da die \mathbf{k} -Punkt-Parallelisierung aufgrund der wenigen, weit außen liegenden \mathbf{k} -Punkt-Schleifen nur wenig Kommunikation erfordert, erreicht sie, wie in Abb. 3.3 ersichtlich, eine hohe Effizienz.

Wichtige Voraussetzung für diese Parallelisierung ist, daß die parallel abgearbeiteten Programmteile, also besonders die Ausführung des Eigenwertlösers `dvdham`, etwa gleich viel Rechenzeit beanspruchen, da es sonst bei der Synchronisation am Ende solcher Teile zu Wartezeiten bei weniger ausgelasteten Prozessoren kommt. Da die Zahl N der \mathbf{G} -Vektoren für verschiedene \mathbf{k} -Punkte leicht unterschiedlich ist, kommt es zu kleinen Unterschieden in der Laufzeit. Theoretisch wäre es sogar denkbar, daß bei einigen \mathbf{k} -Punkten die Lösung des Eigenwertproblems schneller konvergiert als bei anderen. Wie am Skalierungsverhalten in Bild 3.3 zu erkennen ist, spielen derartige Effekte jedoch kaum eine Rolle.

Von der Parallelisierung der Daten sind besonders die Eigenvektoren betroffen, allerdings wurden schon im seriellen Programm die Eigenvektoren immer nur für den zu berechnenden \mathbf{k} -Punkt im Speicher gehalten, so daß sich hier keine Einsparung des Speicherplatzes ergibt. Dies ist auch schon einer der wesentlichen Nachteile dieser Parallelisierung: Jeder einzelne Prozessor benötigt fast ebensoviel Speicherplatz wie das serielle Programm.

Der zweite Nachteil dieser Parallelisierung besteht darin, daß gerade für die von uns berechneten Halbleitersysteme nur wenige \mathbf{k} -Punkte benötigt werden, so-

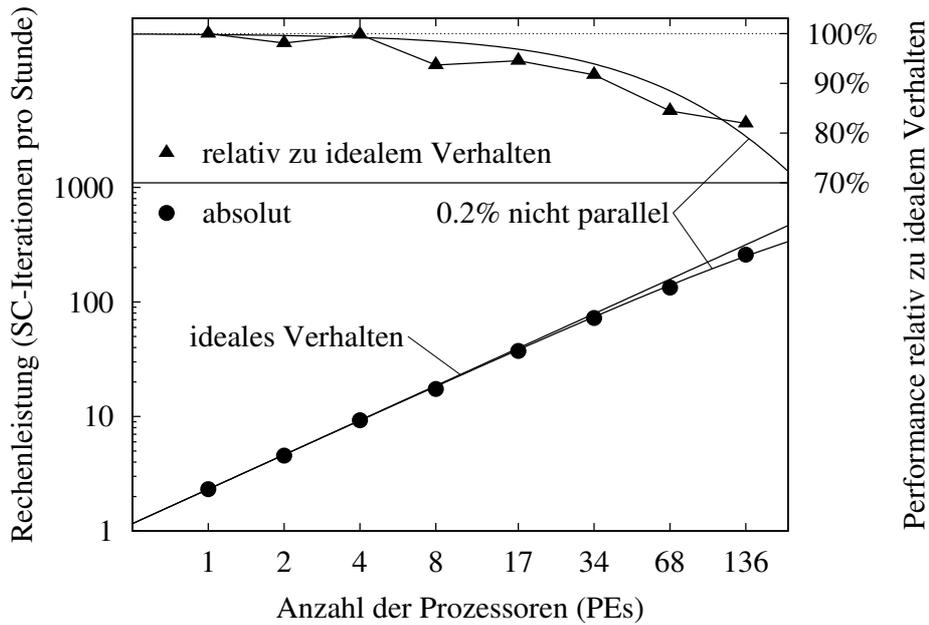


Abbildung 3.3: Effizienz der \mathbf{k} -Punkt-Parallelisierung aufgetragen gegen Anzahl der PEs am Beispiel von Bulk- β -Zinn (2 Atome pro Superzelle; 272 \mathbf{k} -Punkte), berechnet auf einer T3E. Im oberen Teil die Effizienz (Relation tatsächlicher zur idealer Rechenleistung), unten die absolute Rechenleistung: Die Gerade zeigt die hypothetische Rechenleistung bei idealer Parallelisierung, der durchgezogene Fit die theoretische Rechenleistung eines zu 0.2% nicht parallelisierten Programms.

gar Rechnungen mit nur einem \mathbf{k} -Punkt sind denkbar. Da die Anzahl der von dieser Parallelisierung nutzbaren Prozessorgruppen durch die Anzahl der \mathbf{k} -Punkte beschränkt ist, entsteht die paradoxe Situation, daß dann für größere Superzellen immer weniger Prozessoren nutzbar werden.

Allerdings ist diese Parallelisierung besonders geeignet für Metalle und kleine Systeme mit vielen \mathbf{k} -Punkten, sie ist aber auch für große Systeme in Verbindung mit weiteren Parallelisierungen sinnvoll.

Berechnungen großer Halbleitersysteme mit vielen Atomen pro Superzelle benötigen typischerweise wenige \mathbf{k} -Punkte, haben allerdings einen hohen Speicherplatzbedarf pro PE. Daher reicht die \mathbf{k} -Punkt-Parallelisierung zur Berechnung solcher Systeme nicht aus, was die Implementation weiterer Parallelisierungsstrategien nötig macht. Entsprechend Abb. 3.1 teilt man dann jedem \mathbf{k} -Punkt eine Prozessorgruppe zu, zwischen deren Prozessoren die weiteren, feinkörnigeren Parallelisierungsstrategien realisiert werden.

3.2 Die G-Vektor-Parallelisierung

Von den verbleibenden Parallelisierungsstrategien wurde zunächst die \mathbf{G} -Vektor-Parallelisierung implementiert. Bei dieser Parallelisierung wird jeder \mathbf{G} -Vektor einem Prozessor zugeordnet und die Koeffizienten aller Größen, die im Impuls-

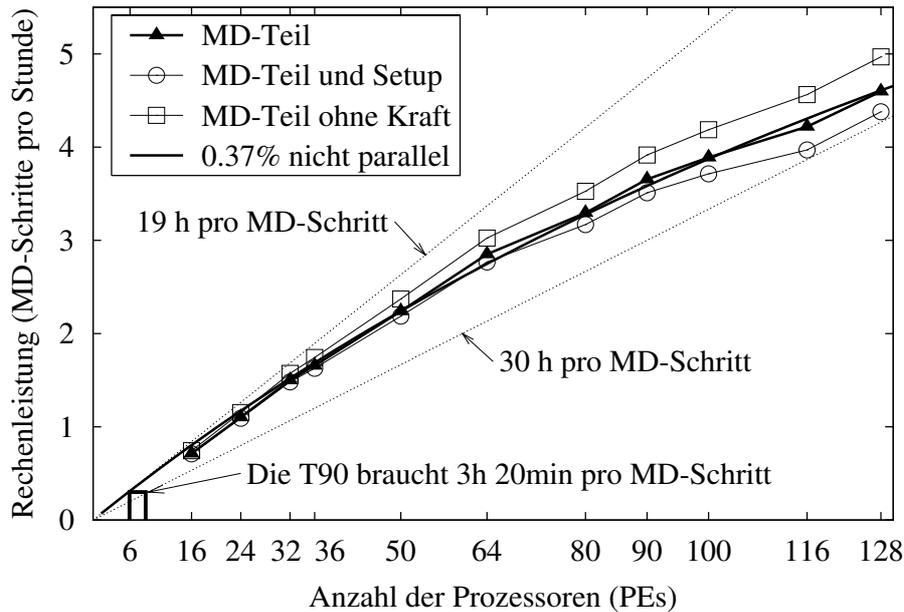


Abbildung 3.4: Rechenleistung aufgetragen gegen Anzahl der PEs für die \mathbf{G} -Vektor-Parallelisierung (Beispiel As auf Si(111); 110 Atome pro Superzelle; zwei \mathbf{k} -Punkte, entsprechend zwei \mathbf{k} -Punkt Gruppen), durchgeführt auf einer Cray-T3E (DEC Alpha EV5 (21164, 450MHz) Prozessoren und 128MB RAM pro PE). Die gepunkteten Geraden zeigen die hypothetische Rechenleistung eines ideal parallelisierten Programmes das 19 bzw. 30 CPU-Stunden pro Molekulardynamikschritt benötigt. Der durchgezogene Fit repräsentiert ein Programm das zu 0.37% nicht parallelisiert ist. Zum Vergleich: ein Cray-T90 Prozessor benötigt 3h 20min pro MD-Schritt, entspricht also (je nach Effizienz) sechs bis neun T3E Prozessoren. Mit 128 PEs wurden insgesamt etwa 15 GFLOPS erreicht.

raum dargestellt werden, und durch die \mathbf{G}^n -Vektoren indiziert werden, entsprechend verteilt:

$$\begin{aligned} \{c_{\mathbf{k}\nu}^n\} &= \{c_{\mathbf{k}\nu}^{(1)}, \dots, c_{\mathbf{k}\nu}^{(N)}\} \\ &\longrightarrow \{(c_{\mathbf{k}\nu}^{(1)}, \dots, c_{\mathbf{k}\nu}^{(N_{PE_1})})_{PE_1}, \dots, (c_{\mathbf{k}\nu}^{(N_{PE_{N-1}}+1)}, \dots, c_{\mathbf{k}\nu}^{(N_{PE_N})})_{PE_N}\}. \end{aligned}$$

Dies sind z.B. die Wellenfunktionen $c_{\mathbf{k}\nu}^n$, die Potentiale $V_{loc}(\mathbf{G}^n)$ und $T_{lm}^{\mu,n}$ und die Ladungsdichten $n(\mathbf{G}^n)$. Da die Anzahl N der Koeffizienten groß ist, erlaubt diese Parallelisierung eine sehr effiziente und flexible Nutzung vieler Prozessoren. Zudem skaliert die Anzahl der \mathbf{G} -Vektoren und somit die Anzahl nutzbarer Prozessoren mit der Größe der simulierten Superzelle.

Allerdings sind von der Implementation sämtliche Loops über \mathbf{G} -Vektoren und somit sämtliche Berechnungen mit den oben genannten Größen betroffen. Der entsprechend hohe Kommunikationsanteil setzt ein schnelles Kommunikationsnetzwerk, wie z.B. das auf der T3E, voraus. Wie in Abb. 3.4 zu sehen, lassen sich so für größere Beispiele 99.63% der Rechenleistung parallelisieren obwohl die Nutzung von nur zwei \mathbf{k} -Punkt Gruppen häufige Kommunikation zwischen großen Clustern von Prozessoren nötig macht. Der Gesamtverlust von 0.37% läßt sich

auf Kommunikation, unparallelisierte Teile des Programms und Synchronisationseffekte zurückführen. Gemäß Amdahls Gesetz führt dies für 128 PEs zu einer Effizienz von 68% für das gesamte Programm, da in diesem Fall die unparallelisierten 0.37% von jedem der 128 PEs bearbeitet werden.

Die Verteilung der **G**-Vektoren auf die PEs bestimmt, wie Operationen auf die entsprechend verteilten c_{kv}^n implementiert werden müssen. Daher ist sie wesentlich für die Effizienz der **G**-Vektor-Parallelisierung, und es gibt verschiedene Ansätze, durch eine geeignete Verteilung der **G**-Vektoren eine optimale Parallelisierung zu erreichen [CvP92, NPS93, WJ95]. Es lassen sich im Hinblick auf die Parallelisierbarkeit grundsätzlich drei Klassen von Operationen unterscheiden:

- a) Multiplikation eines Vektors mit einem Skalar, Summen von Vektoren, Applikation einer diagonalen Matrix:

$$\mathbf{a} + \mathbf{b} = \left(\begin{array}{c} \left[\begin{array}{cc} a_1 & + b_1 \\ & \dots \\ a_{PE_1} & + b_{PE_1} \end{array} \right]_{PE_1} \\ \cdot \\ \cdot \\ \cdot \\ \left[\begin{array}{cc} a_{PE_{N-1}+1} & + b_{PE_{N-1}+1} \\ & \dots \\ a_{PE_N} & + b_{PE_N} \end{array} \right]_{PE_N} \end{array} \right)$$

- b) Skalarprodukte zweier Vektoren:

$$\mathbf{a} \cdot \mathbf{b} = \left[\sum_{i=1}^{N_{PE_1}} a_i b_i \right]_{PE_1} + \dots + \left[\sum_{i=N_{PE_{N-1}}+1}^{N_{PE_N}} a_i b_i \right]_{PE_N}$$

- c) Fourier-Transformationen der Vektoren von Impuls- zu Ortsraum und umgekehrt.

Die unter a) genannten Operationen sind lokal ausführbar. Dazu müssen im Falle der Vektoraddition die Koeffizienten beider zu verknüpfenden Vektoren gleich über die Prozessoren verteilt sein. Analoges gilt für das Anmultiplizieren einer Diagonalmatrix. Bei Skalarprodukten b) werden zunächst die lokalen Teilsommen der einzelnen PEs berechnet, die dann mittels einer Reduktionsoperation über alle PEs zum eigentlichen Skalarprodukt aufsummiert werden. Um eine optimale Verteilung von Rechenzeit und Speicherplatz zu gewährleisten, sollten die Koeffizienten der Vektoren möglichst gleichmäßig über die PEs verteilt werden, da der Speicherplatzbedarf, und für Operationen nach a) und b) die Rechenzeit, mit der maximalen Anzahl der N_{PE_i} Koeffizienten pro PE skaliert. Im Falle der

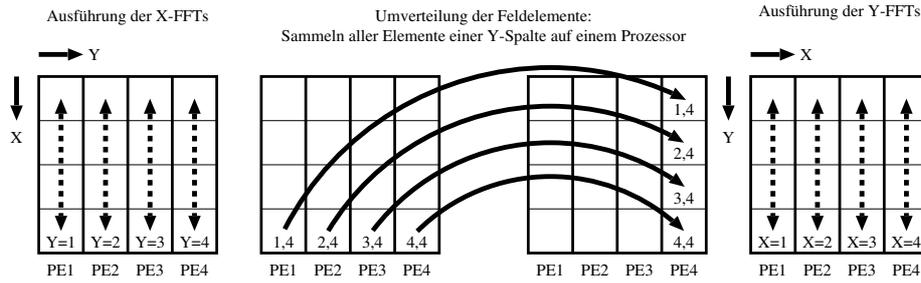


Abbildung 3.5: Parallelisierung einer 2-dimensionalen FFT. Am einfachsten lassen sich mehrdimensionale Fourier-Transformationen implementieren, indem einzelne Fourier-Transformationen immer auf einem PE ausgeführt werden. Im Beispiel sind zunächst die 4 y-Spalten der Matrix den 4 PEs zugeordnet, so daß alle x-Elemente einer y-Spalte auf einem PE liegen. Daher können die Fourier-Transformationen in y-Richtung (y-FFTs) lokal auf den PEs ausgeführt werden. Dann werden die Daten umverteilt, so daß nun die 4 x-Spalten den 4 PEs zugeordnet werden. Daraufhin können die x-FFTs lokal ausgeführt werden.

Skalarprodukte b) sollten außerdem möglichst viele Skalarprodukte zusammengefaßt werden, um durch Zusammenfassen der Kommunikation Latenzzeiten zu verringern.

Im Fall c) der Fourier-Transformationen ist der Vektor als dreidimensionales Feld (die FFT-box) an den Gitterpunkten $\mathbf{G}^n = \mathbf{G}^{n_x, n_y, n_z} = (G_x^{n_x}, G_y^{n_y}, G_z^{n_z})$ zu betrachten. Der dreidimensionale Vektor $\mathbf{G}^{n_x, n_y, n_z}$ (bzw. $\mathbf{r}^{n_x, n_y, n_z}$) entspricht dabei den drei Raumdimensionen des reziproken (oder realen) Raumes. Die dreidimensionale FFT entspricht eindimensionalen FFTs entlang den drei Raumrichtungen. Aufgrund der Verteilung der Koeffizienten liegen nun für mindestens eine der Raumrichtungen die an einer FFT beteiligten Koeffizienten auf verschiedenen Prozessoren. Daher müssen die Koeffizienten vor diesen Transformationen umverteilt werden, wie in Abb. 3.5 am Beispiel einer zweidimensionalen FFT dargestellt. Die Fourier-Transformationen werden also nacheinander für die x-, y- und z-Richtung ausgeführt, beginnend mit der Richtung mit der größten Anzahl von Elementen $n_i; i = x, y, z$. In unserem Programm wird die Zelle so gedreht, daß dies die z-Richtung ist. Um die Durchführung der FFT ohne vorherige Umverteilung der Koeffizienten zu ermöglichen, sollten daher alle Koeffizienten einer z-Spalte zusammen auf einem PE liegen.

3.2.1 Aufteilung der G-Vektoren

Die Optimierung oben genannter Operationen stellt einige Bedingungen an die Aufteilung der \mathbf{G} -Vektoren auf die Prozessoren: Um einen Teil der Fourier-Transformationen ausführen zu können, ohne vorher Koeffizienten zwischen den PEs umzuverteilen, müssen Koeffizienten, die an der gleichen Fourier-Transformation beteiligt sind, auf dem gleichen PE liegen. Da zunächst die längsten (z-) FFTs ausgeführt werden bedeutet das, daß die z-Spalten den Prozessoren

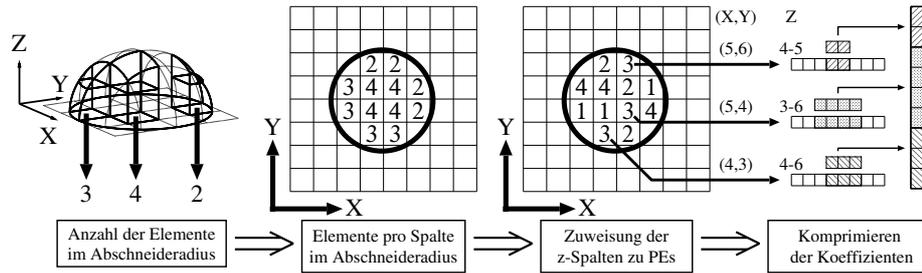


Abbildung 3.6: Spaltenweise Zuweisung der \mathbf{G} -Vektoren zu PEs (am Beispiel von PE 3). Das linke Diagramm zeigt welche \mathbf{G} -Vektoren innerhalb des Abschneideradius \mathbf{G}_{max} um \mathbf{k} liegen. Bei der Fourier-Transformation in den Realraum werden alle Elemente außerhalb dieses Radius zu Null gesetzt bzw. bei der Rücktransformation vernachlässigt. Daraus ergibt sich für jedes (x, y) Tupel eine z-Spalte mit verschiedener Anzahl berücksichtigter Elemente, wie im zweiten Diagramm dargestellt. Im nächsten Diagramm wird jede Spalte einem Prozessor zugeordnet, wobei auf möglichst gleichmäßige Verteilung der Koeffizienten geachtet wird. Alle z-Spalten eines Prozessors werden dann zu einem langen Vektor komprimiert, der nur noch $c_{kv}(\mathbf{G}_n)$ zu \mathbf{G}_n innerhalb des Abschneideradius enthält.

zugewiesen werden. Dabei sollten auf jeden PE nach Möglichkeit etwa gleichviele Koeffizienten und etwa gleichviele z-Spalten entfallen, um eine optimale Verteilung der Rechenleistung sowohl für Vektoroperationen als auch für Fourier-Transformationen zu erreichen.

Wie in Abb. 3.6 zu sehen, liegen die im reziproken Raum berücksichtigten \mathbf{G} -Vektoren innerhalb eines durch \mathbf{G}_{max} festgelegten Abschneideradius, also in einer Kugel. Dies bedeutet, daß der 3-tupel (n_x, n_y, n_z) in einem Ellipsoid liegt, da die Koordinaten im allgemeinen nicht kartesisch sind. Daher enthalten die jeweiligen z-Spalten eine unterschiedliche Anzahl berücksichtigter Koeffizienten. Zur Verteilung der Spalten werden diese nach Anzahl der Koeffizienten sortiert. Dann werden die Spalten nacheinander, beginnend mit der ‘längsten’, immer dem PE zugeordnet, der gerade die wenigsten Spalten und bei mehreren PEs mit gleicher Spaltenzahl die wenigsten Koeffizienten hat. Danach werden Spalten paarweise zwischen den PEs vertauscht, bis sich auf diese Weise keine gleichmäßigere Verteilung mehr erzielen läßt. Dieses Design hat den Vorteil, daß die Zahl der PEs für die \mathbf{G} -Vektor Parallelisierung nicht durch Bedingungen auf wenige Werte eingeschränkt wird. Der hohe Rechenaufwand für ein derartiges Verfahren ist dadurch gerechtfertigt, daß es pro Programmablauf nur einmal (im setup) ausgeführt werden muß, während die durch suboptimale Verteilung der Koeffizienten entstehende ungleiche Verteilung der Rechenlast sich während des gesamten Programmablaufs auswirkt.

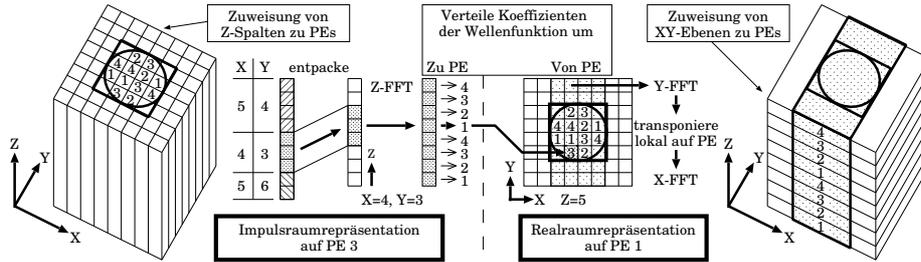


Abbildung 3.7: Parallelisierte 3-dimensionale FFT: Links ist die Verteilung der z-Spalten über die PEs für die Impulsraumdarstellung zu sehen. Zunächst werden die Vektorstücke entpackt (hier $(x, y) = 4, 3$ auf PE 3) und mit Nullen aufgefüllt. Daraufhin wird für alle Spalten, die von Null verschiedene Elemente enthalten, die z-FFT ausgeführt. Danach werden die Elemente umverteilt, wobei nun den PEs xy-Ebenen reihum zugeordnet werden (siehe rechts). Im Beispiel werden alle Elemente der xy-Ebenen mit $z = 1$ und $z = 5$ Prozessor 1 zugewiesen. Die Umverteilung ist in der Mitte dargestellt: z.B. wird das fünfte Element der $(4, 3)$ z-Spalte von PE 3 an PE 1 weitergegeben, der es den Koordinaten $(4, 3)$ der ihm zugewiesenen xy-Ebene 5 zuordnet. Nun werden alle y-Reihen mit von Null verschiedenen Elementen (schattierte Region) y-transformiert, gefolgt von einer lokalen Transposition und den x-FFTs. Rechts ist die Verteilung der xy-Ebenen im Realraum dargestellt. Zur Rücktransformation wird der Algorithmus rückwärts ausgeführt.

3.2.2 Die Fourier-Transformation

Die am häufigsten anfallenden Fourier-Transformationen betreffen die Wellenfunktionen. Diese sind im Impulsraum nur innerhalb des Abschneideradius \mathbf{G}_{max} von Null verschieden und werden außerhalb dieses Radius numerisch als Null betrachtet. Da die Anmultiplikation des Potentials im Realraum im Impulsraum einer Faltung entspricht:

$$\begin{aligned} c'_{\mathbf{k},\nu}(\mathbf{r}) &= V(\mathbf{r})c_{\mathbf{k},\nu}(\mathbf{r}) \\ \xrightarrow{FFT} c'_{\mathbf{k},\nu}(\mathbf{G}) &= \sum_{\mathbf{G}'} V(\mathbf{G} - \mathbf{G}')c_{\mathbf{k},\nu}(\mathbf{G}') , \end{aligned} \quad (3.1)$$

müssen Koeffizienten $V(\mathbf{G})$ mit $|\mathbf{G}| \leq 2\mathbf{G}_{max}$ berücksichtigt werden. Das lokale Potential und daher auch die lokale Dichte müssen also innerhalb des doppelten Abschneideradius \mathbf{G}_{max} berücksichtigt werden. Die FFT-box ist so zu wählen, daß sie diese Kugel beinhaltet. Sie hat also eine Kantenlänge von $4\mathbf{G}_{max}$.

Dies bedeutet, daß bei den Transformationen der Wellenfunktionen etwa die Hälfte der eindimensionalen FFTs erspart werden kann, da diese ausschließlich Nullen enthalten. Dies kann von üblichen Bibliotheksroutinen nicht berücksichtigt werden, gestattet aber eine wesentliche Optimierung, indem bei Transformationen vom Impuls- in den Realraum nur solche Fourier-Transformationen ausgeführt werden, die von Null verschiedene Elemente enthalten, und bei der Rücktransformation nur solche, die zur Berechnung der am Ende berücksichtigten Elemente (innerhalb \mathbf{G}_{max}) wichtig sind. Bei Transformation von Impuls- zu

Realraum müssen also zunächst nur die z-FFTs ausgeführt werden, deren Elemente zum Teil im Abschneideradius liegen, dann alle y-FFTs in allen yz-Ebenen, die solche Elemente enthalten, und schließlich alle x-FFTs.

Die dreidimensionale FFT ist, wie in Abb. 3.7 veranschaulicht², aus eindimensionalen FFTs zusammengesetzt, die auf den PEs lokal ausgeführt werden. Indem die Koeffizienten, wie in Abb. 3.6 gezeigt, z-Spaltenweise an die PEs verteilt werden, können die z-FFTs ohne vorherige Kommunikation durchgeführt werden. Um dabei eine möglichst gleichmäßige Arbeitsaufteilung zu gewährleisten, sollten die z-Spalten möglichst gleichmäßig auf die PEs verteilt werden.

Nun befinden sich auf jedem PE alle (bereits z-transformierten) Elemente jeweils der z-Spalten, die diesem PE zugewiesen sind. Für die x- und y-FFTs ist es nun notwendig, die Koeffizienten so umzuverteilen, daß alle Elemente einer xy-Ebene auf einem PE liegen. Daher werden die xy-Ebenen reihum an die PEs verteilt, jeder PE sammelt nun die Elemente aller z-Spalten auf allen PEs, die zu den ihm zugewiesenen xy-Ebenen gehören. Daraufhin können y- und x-Fourier-Transformationen wieder lokal ausgeführt werden. In der Realraumdarstellung sind die Elemente dann ebenenweise auf die Prozessoren verteilt. Auch hier sollten die xy-Ebenen wieder möglichst gleichmäßig auf die PEs verteilt werden.

3.3 Die Band-Parallelisierung

Der Vorteil der Band-Parallelisierung liegt darin, daß die Anzahl M bzw. M_{dvd} der Wellenfunktionen groß ist und mit der Anzahl N_{at} der Atome, also mit der Systemgröße skaliert. Dies bedeutet, daß auch die Anzahl nutzbarer Prozessoren mit der Systemgröße skaliert.

Da die in unserem Programm gewählte Implementation des Davidson-Verfahrens in jedem Schritt sowohl eine Orthonormierung als auch eine Projektion auf den Unterraum durchführt, macht es diese Parallelisierung nötig, entweder den Eigenwertlöser zu modifizieren oder eine parallele Orthonormierung für verteilte Vektoren zu implementieren. Die Schwierigkeit liegt dabei darin, daß einerseits Skalarprodukte zwischen Vektoren ausgeführt werden müssen, die auf verschiedenen Prozessoren residieren, andererseits diese Skalarprodukte im Gram-Schmidt-Verfahren nicht in beliebiger Reihenfolge ausgeführt werden können (wie das z.B. für die Projektion auf den Unterraum der Fall ist). Eine Möglichkeit ist, ein alternatives Orthonormierungsverfahren zu benutzen [WJ94], um die Abhängigkeiten des Gram-Schmidt-Verfahrens zu vermeiden.

Ein anderer Ansatz ist, eine geeignete Verteilung der Eigenvektoren $\psi_{\mathbf{k}\nu}$ zu wählen, um dennoch eine effiziente Implementation des Gram-Schmidt-Verfahrens zu ermöglichen. Die in Abb. 3.8 dargestellten Abhängigkeiten führen

²Zur Vereinfachung der Darstellung wurde hier der Ursprung der FFT, und somit des Abschneideradius, im Zentrum der FFT-Box dargestellt, tatsächlich muß der Abschneideradius jedoch um das $(0, 0, 0)$ -Element in einer Ecke der Box zentriert werden.

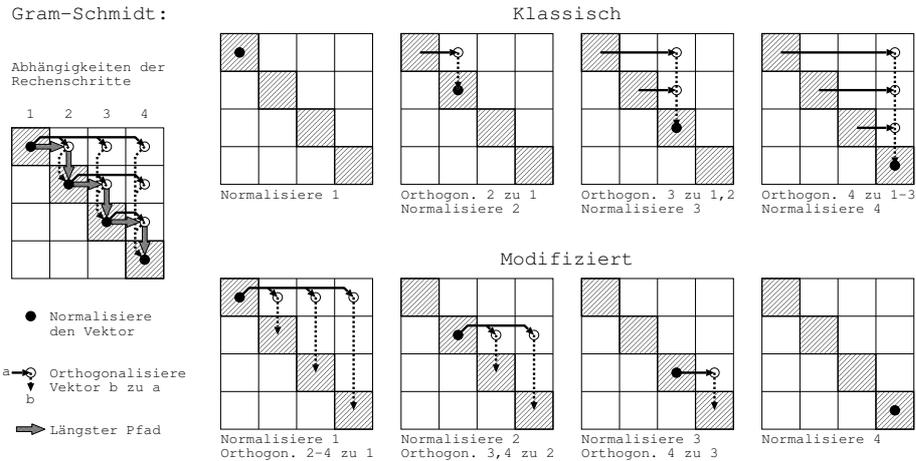


Abbildung 3.8: Gram-Schmidt-Verfahren. Im linken Teil des Diagramms sind die Abhängigkeiten der Rechenschritte dargestellt: Jeder Vektor muß zunächst zu allen vorhergehenden orthogonalisiert werden, bevor er normiert werden kann, erst daraufhin können nachfolgende Vektoren zu diesem Vektor normiert werden. Daraus ergibt sich ein längster Pfad nacheinander auszuführender und voneinander abhängiger Rechenoperationen, da ein Vektor (unter anderem) jeweils zum vorhergehenden orthogonalisiert werden muß, bevor er normiert werden kann, erst daraufhin kann der nächste Vektor zu diesem orthogonalisiert werden. Rechts oben ist eine Implementation des klassischen Gram-Schmidt-Verfahrens dargestellt: In jedem Schritt wird ein Vektor zu allen vorhergehenden orthogonalisiert (dabei werden die Skalarprodukte mit dem ursprünglichen Vektor gebildet, woraus ein Korrekturvektor entsteht der erst dann zum ursprünglichen Vektor addiert wird) und dann normiert. Rechts unten hingegen das modifizierte Gram-Schmidt-Verfahren: Jeder Vektor wird zunächst normiert, daraufhin werden alle folgenden Vektoren zu diesem orthogonalisiert (die entstehenden Korrekturvektoren werden sofort angewandt).

zu Synchronisationseffekten, indem sie dazu führen, daß ein PE auf die Ergebnisse eines anderen PE warten muß. Solche Wartezeiten müssen möglichst vermieden werden, um zu gewährleisten, daß die Prozessoren möglichst gleichmäßig ausgelastet sind. Bisher wurde diese Parallelisierung noch nicht implementiert, trotzdem sollen hier einige Vorüberlegungen dargelegt werden.

3.3.1 Parallelisierung des Gram-Schmidt-Verfahrens

Eine mögliche Parallelisierungsstrategie für die Orthonormierung ist in Abb. 3.9 gezeigt. Wie man sieht, wurden einzelne Schritte der in Abb. 3.8 dargestellten Gram-Schmidt-Verfahren parallelisiert. Ein Schritt des klassischen Verfahrens besteht in der Berechnung des Gesamtkorrekturvektors aus der Summe der einzelnen Projektionen des betrachteten Vektors auf alle vorangegangenen. Dazu muß der Vektor zunächst an alle PEs verteilt werden, dort werden jeweils die Projektionen auf die dort liegenden Vektoren berechnet und die Teilsummen pro PE

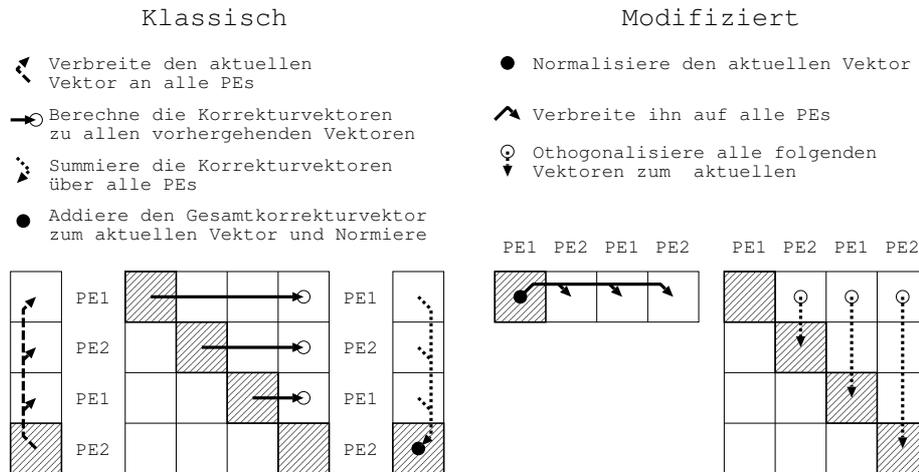
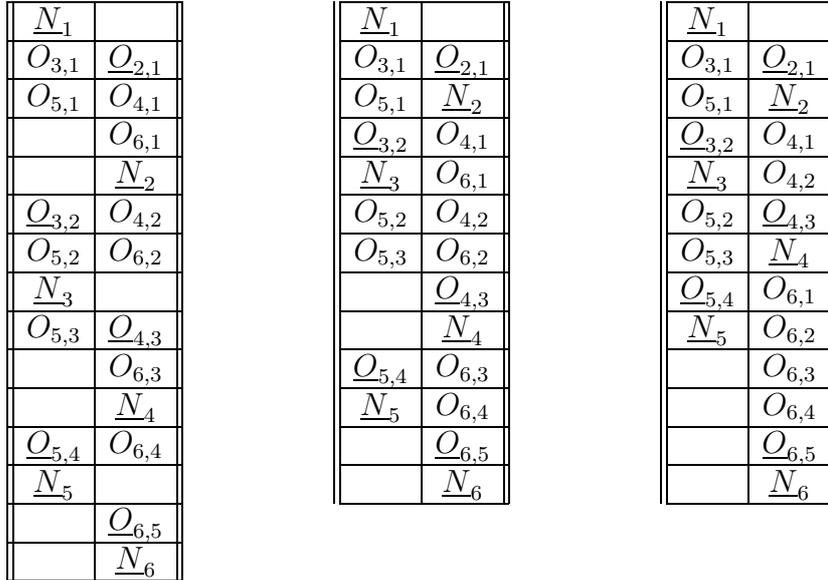


Abbildung 3.9: Parallelisierung eines Gram-Schmidt-Schrittes am Beispiel von vier Vektoren und zwei Prozessoren (PE1 und PE2): Links: Orthogonalisierung von Vektor 4 zu seinen (bereits orthonormierten) Vorgängern bei klassischem Gram-Schmidt. Zunächst wird der Vektor an alle PEs verteilt, die dort berechneten Korrekturvektoren zurückgesandt, und angewandt. Rechts: Orthogonalisierung der Vektoren 2-4 zu Vektor 1 bei modifiziertem Gram-Schmidt. Der normierte Vektor wird an alle PEs verteilt, wo die nachfolgenden Vektoren zu diesem orthogonalisiert werden.

gebildet. Diese werden dann an den ursprünglichen Prozessor zurückgeschickt, dort aufsummiert und vom zu orthogonalisierenden Vektor abgezogen. Nach dessen Normierung wird mit dem folgenden Vektor fortgefahren.

Beim modifizierten Verfahren wird der aktuelle Vektor normiert und dann an alle PEs verschickt. Dort werden alle nachfolgenden Vektoren zu diesem orthogonalisiert, woraufhin der nächste Vektor bearbeitet wird. Das klassische Gram-Schmidt-Verfahren läßt sich z.B. auch analog zum modifizierten parallelisieren, indem die Korrekturvektoren zunächst aufsummiert und erst der Gesamtvektor angewandt wird. Wie im Diagramm zu sehen, sollten die Vektoren reihum auf die Prozessoren verteilt werden. Liegen nämlich z.B. die zehn letzten zu orthogonalisierenden Vektoren auf einem PE, so wird sich (für den Fall des modifizierten Gram-Schmidt-Verfahrens) am Schluß nur noch dieser mit deren Orthonormierung befassen, während die anderen Prozessoren vermutlich an einer anderen Stelle des Programms warten.

Wird nun das modifizierte Gram-Schmidt-Verfahren wie oben beschrieben implementiert, so ergibt sich durch die Abhängigkeiten der Rechenschritte, daß Prozessoren auf Ergebnisse anderer Prozessoren warten müssen. Dies ist in Abb. 3.8 unten links zu beobachten: bevor ein Prozessor seine Vektoren zum nächsten Vektor orthogonalisieren kann, muß dieser erst berechnet und verteilt werden, es kommt für jeden Schritt des Verfahrens (also pro Vektor) zu Leerlaufzeiten auf allen Prozessoren. Also ist die Gesamtleeplaufzeit $t_{idle} \propto N_{PE} \times M_{dvd}$, und da die Zahl der PEs mit M_{dvd} skalieren soll, ist $t_{idle} \propto M_{dvd}^2$ proportional zur Zahl der nötigen Operationen und somit zur Rechenzeit.



$O_{a,b}$ Orthogonalisiere Vektor a zu Vektor b
 N_a Normalisiere Vektor a

Abbildung 3.10: Verteilung der Rechenschritte zur Orthogonalisierung von 6 Vektoren auf zwei Prozessoren beim modifizierten Gram-Schmidt-Verfahren. Hier wurde zur Vereinfachung der Darstellung angenommen, daß alle Orthogonalisierungen und Normierungen gleichlange dauern und normierte Vektoren sofort an alle Prozessoren verteilt werden. Durch Vorziehen der Operationen des kritischen Pfades aus Abb. 3.8 (hier durch Unterstreichen hervorgehoben) lassen sich Wartezeiten zum Großteil vermeiden. Im mittleren Diagramm werden nur Normierungen vorgezogen, während im letzten Diagramm zusätzlich die Reihenfolge der Orthogonalisierungen optimiert wird. Dazu ist allerdings einiger Buchhaltungsaufwand nötig, und auch im Falle von mehr Prozessoren und Vektoren sind die Einsparungen nur gering.

Werden die Operationen des kritischen Pfades aus Abb. 3.8 bevorzugt ausgeführt, so lassen sich diese Wartezeiten zum Großteil vermeiden, wie man in Abb. 3.10 sehen kann. Durch zusätzliche Umstellung der Orthogonalisierungen läßt sich in Fällen mit mehr Vektoren und mehr Prozessoren noch etwas Leerlaufzeit einsparen. Dies ist aber mit großem Buchhaltungsaufwand verbunden, da dazu Vektoren, die noch für spätere Orthonormierungsschritte gebraucht werden, vorgehalten werden müssen. Es zeigt sich, daß das Vorziehen der Normierungen ausreicht, da in diesem Fall nur für die letzten 3-4 Vektoren pro Prozessor Leerlaufzeiten entstehen, somit $t_{idle} \propto N_{PE}$ ist und für genügend viele Vektoren pro PE gegenüber der Rechenzeit nicht mehr ins Gewicht fällt. In der Implementation läßt sich dazu die asynchrone Kommunikation vorteilhaft ausnutzen, für das modifizierte Gram-Schmidt-Verfahren gestaltet sich das wie in Abb. 3.11 zu

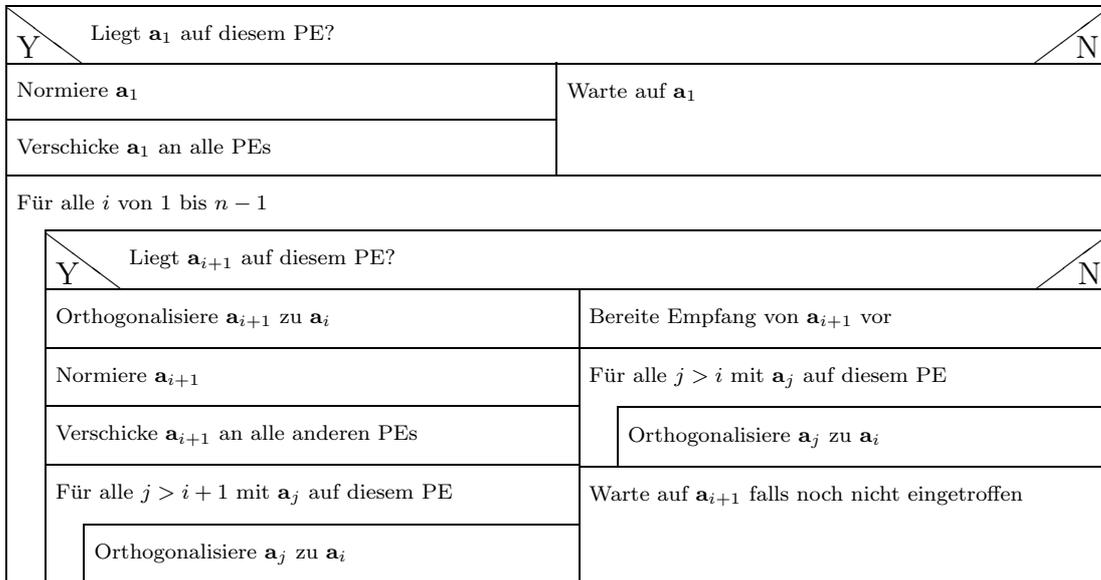
Gram-Schmidt — Orthonormiere Vektoren \mathbf{a}_i für $i = 1 \dots n$ 

Abbildung 3.11: Nassi-Schneiderman Diagramm für ein parallelisiertes modifiziertes Gram-Schmidt-Verfahren, bei dem Normierungsschritte vorgezogen werden.

sehen.

Die asynchrone Kommunikation vermeidet hier weitgehend Synchronisationseffekte, unter anderem indem sie ermöglicht, daß jeder Prozessor individuell mit dem nächsten Schritt beginnen kann, wenn er seine Operationen des aktuellen Schrittes abgearbeitet hat, sofern der nächste orthonormierte Vektor bereits eingetroffen ist.

Während das Verteilen der Vektoren reihum eine sehr gleichmäßige Arbeitsaufteilung gestattet, ist einzelnes Versenden von Vektoren nicht wünschenswert, da jede Kommunikation zu Latenzzeiten führt. Als Kompromiß können die Vektoren statt einzeln auch blockweise reihum verteilt werden. Ersetzt man einen Vektor durch einen Block von Vektoren, Normierung eines Vektors durch Orthonormierung aller Vektoren des Blockes auf sich und Orthogonalisierung zweier Vektoren durch Orthogonalisierung jedes Vektors eines Blockes zu allen Vektoren eines anderen Blockes, so lassen sich die Verfahren zur Behandlung von Vektoren auf Blöcke von Vektoren übertragen³.

In einem automatisierten Ablauf ist es auch wichtig, eine Fehlerabschätzung, bzw. eine Abschätzung der Qualität der Orthonormierung zu implementieren, um dem aufrufenden Programm z.B. die Möglichkeit zu geben, auf eine zu schlechte

³Gestattet man beim modifizierten Verfahren, daß die Korrekturvektoren nicht einzeln sondern blockweise angewandt werden, läßt sich die Orthogonalisierung zweier Blöcke zueinander sogar mit BLAS 3 Routinen realisieren.

Orthonormierung mit Abbruch, Wiederholung der Orthonormierung oder Warnmeldungen zu reagieren. Eine Möglichkeit dazu wäre, nach Durchführung der Orthonormierung die Skalarprodukte der Vektoren untereinander zu überprüfen. Der Aufwand für diese Art der Fehlerabschätzung ist aber sehr hoch: Es werden etwa die Hälfte der Rechenoperationen eines vollständigen Orthonormierungsschrittes benötigt. Einfacher ist es, die Länge eines Vektors vor und nach der Anwendung des Korrekturvektors miteinander zu vergleichen: ist der Ergebnisvektor nach Subtraktion des Korrekturvektors um Größenordnungen kleiner als der ursprüngliche Vektor, so wurden zwei etwa gleiche Vektoren voneinander abgezogen. Dies führt üblicherweise zu Rundungsfehlern, die sich durch die Rechengenauigkeit (Anzahl der berücksichtigten Dezimalstellen) der verwendeten Architektur abschätzen lassen. Indem z.B. alle Vektoren vor der Orthonormierung einmal normiert werden, läßt sich bei der abschliessenden Normierung aus der Länge des orthogonalisierten Vektors auch eine Abschätzung des Rundungsfehlers ableiten.

3.4 Zusammenfassung

Von den besprochenen Parallelisierungsansätzen wurden bisher die \mathbf{k} -Punkt- und die \mathbf{G} -Vektor-Parallelisierung im EStCoMPP-Programm implementiert. Für Systeme mit grossen Superzellen und entsprechend wenigen \mathbf{k} -Punkten ist besonders die \mathbf{G} -Vektor-Parallelisierung wichtig, da erst mit dieser das Eigenwert-Problem auf vielen Prozessoren verteilt gelöst werden kann. Bei der Aufteilung der \mathbf{G} -Vektoren, bzw. der entsprechenden Koeffizienten der Wellenfunktionen, auf verschiedene Prozessoren wurde darauf geachtet, daß die häufig auszuführenden Fouriertransformationen mit einem Minimum an Kommunikation verbunden sind. Zur Band-Parallelisierung wurde bisher nur der in Kapitel 3.3.1 beschriebene parallele Gram-Schmidt-Algorithmus implementiert.

Für die Parallelisierung auf der Cray-T3E wurden soweit möglich MPI-Aufrufe benutzt. Nur an sehr kommunikationsintensiven Stellen, wie z.B. der Fouriertransformation und der Unterraumdiagonalisierung, wurde auf "shmem"-Routinen von Cray zurückgegriffen, um das Kommunikationsnetzwerk der T3E optimal auszunutzen. Diese Programmteile müssen bei einer Portierung auf andere MPP-Maschinen überarbeitet und z.T neu parallelisiert werden.

Kapitel 4

Stufenkanten

4.1 Motivation

Der Trend zur Miniaturisierung elektronischer und opto-elektronischer Bauelemente ist ungebrochen. Zur Herstellung schnellerer Bauelemente mit geringerem Leistungsverbrauch ist es essentiell, ebene Heterostrukturen aus wenigen Atomlagen mit atomar scharfen Grenzflächen zu bauen. Für zwei Materialien mit unterschiedlichen Gitterkonstanten ist Wachstum von ebenen epitaktischen Schichten nicht die Regel. So wächst z.B. Germanium auf Silizium im sogenannten Stranski-Krastanov-Modus, d.h. nach einer ca. drei Monolagen dicken benetzenden Ge-Schicht bilden sich große dreidimensionale Ge-Inseln.

Wie experimentell gezeigt wurde [MVA92, VZ96, HCT⁺94], kann man mit einer monoatomaren Schicht aus Arsen-, Antimon- oder Wismut- Fremdatomen auf der Si-Oberfläche das Wachstum von Ge auf Si so verändern, daß ebene Ge-Schichten aufwachsen. Man nennt solche Atome wegen der Oberflächenaktivität "Surfactant"-Atome. Die Surfactantatome verändern dabei die Oberflächenstruktur, z.B. zeigt Si(111):As eine 1×1 -Struktur ohne Überstruktur, statt der komplizierten 7×7 Überstruktur der reinen Si(111)-Oberfläche. Die Surfactantatome nehmen Plätze auf der Oberfläche ein, an denen nur drei nächste Nachbarn zur Verfügung stehen. Da sie fünfwertig sind, sind die Surfactantatome ideal koordiniert, da sie durch die Ausbildung dreier kovalenter Bindungen Edelgaskonfiguration erreichen. Die Ausbildung doppelt besetzter dangling bonds senkt die Oberflächenenergie. Deshalb segregieren die Surfactantatome an die Oberfläche.

Die Surfactant-Schicht 'schwimmt' sozusagen auf der Oberfläche und die hinzukommenden Adatome werden unter der Surfactantlage eingebaut. Da die Atome des Surfactants zunächst die Gitterplätze besetzen, die die aufgedampften Atome dann beim Wachstum einnehmen, werden dabei einzelne Surfactantatome durch hinzukommende Adatome ersetzt. Für auf die Oberfläche auftreffende Adatome gibt es also zwei mögliche Prozesse: entweder diffundieren sie auf der Oberfläche oberhalb der Surfactant-Schicht oder sie werden eingebaut, indem

sie mit einem Surfactantatom den Platz tauschen. Allerdings ist dieses "Aufschwimmen" des Surfactants nicht ausreichend, um Lagenwachstum zu erzwingen. Zusätzlich zur Absenkung der Oberflächenenergie bewirkt das Surfactant auch eine Veränderung der Kinetik der an der Oberfläche stattfindenden Prozesse. Um zu untersuchen, welche Elemente als Surfactant geeignet sind, muß untersucht werden, wie die Kinetik des epitaktischen Wachstums verändert wird.

Es lassen sich nun zwei unterschiedliche Mechanismen vorstellen, die auf einer durch ein Surfactant passivierten Oberfläche ein Lagenwachstum begünstigen: [VZ93]: Zum einen wäre vorstellbar, daß Adatome auf der passivierten Oberfläche durch Diffusion sehr weite Strecken zurücklegen. Da alle Bindungen an der Oberfläche abgesättigt sind, können sich auf der Oberfläche auftreffende Atome schnell bewegen und erreichen so Stufenkanten, an denen sie eingebaut werden. Da so Material bevorzugt an den Stufenkanten eingebaut wird, kommt es zu einem "Stufen-Fluss"-Wachstum (step-flow). Andererseits ist ein Modell denkbar, bei dem die Diffusion eher gebremst wird. Ein ankommendes Atom wird nicht weit diffundieren, sondern nach geringer Distanz eingebaut und bleibt dann im energetisch günstigeren (eingebauten) Zustand. Dieser Mechanismus führt zu einer erhöhten Inseldichte. Die Inseldichte ist ein Indikator, welcher dieser Mechanismen tatsächlich stattfindet. Bei langen Diffusionswegen bis zum Einbau an einer Stufenkante entstehen wenige Inseln. Werden Adatome nach wenigen Diffusionsschritten in die Oberfläche eingebaut, so entstehen viele Inseln.

Bei Heteroepitaxie kommt es zu Effekten, wie Verspannung der Oberflächlagen durch die verschiedenen Gitterkonstanten von Germanium und Silizium. Um diese Effekte zu vermeiden und nur die Kinetik der Adatome auf der Oberfläche zu untersuchen, bietet es sich daher an, den Einfluss verschiedener Surfactants bei Homoepitaxie von Silizium auf der Si-(111) Oberfläche zu beobachten. Durch Auswertung von STM-Bildern untersuchten Voigtländer *et al.* den Einfluß verschiedener monoatomarer Fremdatombedeckungen auf die Länge des Diffusionswegs von Si-Adatomen bis zum Einbau [VZ93, VZWB95] und fanden, daß eine Atomlage In auf Si(111) die Diffusion der Si-Adatome erleichtert, während As und Sb zu höheren Inseldichten führen, als auf der reinen Si-Oberfläche. Dies läßt sich so interpretieren, daß diese Surfactants zu einer Verkürzung der Diffusionswege führen, die die Si-Adatome auf der Oberfläche zurücklegen, bevor sie eingebaut werden. As und Sb sind, im Gegensatz zu In, gute Surfactants für die Ge-Heteroepitaxie. Daher schlossen Voigtländer *et al.*, daß die Verkürzung der Diffusionswege die Entstehung dreidimensionaler Inseln reduziert.

Inzwischen ist es möglich, solche mikroskopischen Prozesse, wie den Austausch eines Surfactantatoms durch ein Siliziumatom auf der Si(111)-Oberfläche in *ab-initio* Berechnungen zu simulieren. Die Nutzung moderner Supercomputer gestattet die Berechnung ausreichend großer Superzellen, um so auch nichtperiodische Systeme, wie z.B. einzelne Atome auf Oberflächen, zu untersuchen. Dabei wird zwar weiterhin tatsächlich ein periodisches System berechnet, bei einer ausreichend großen Superzelle liegen die periodisch angeordneten Wiederholungen der

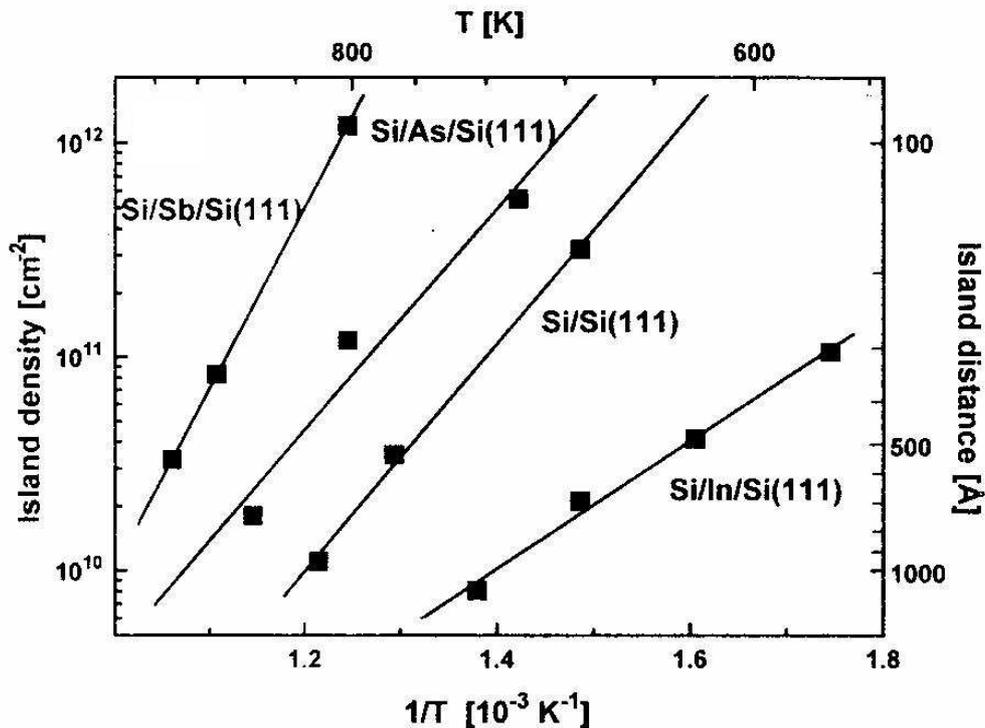


Abbildung 4.1: Inselfichte gegen Temperatur. An der Inselfichte läßt sich der Einfluss des Surfactants auf die Kinetik an der Oberfläche und speziell die Diffusionswege der Adatome ersehen: Hohe Inselfichten entsprechen kurzen, niedrige Inselfichten langen Diffusionswegen.

untersuchten Struktur allerdings weit genug auseinander, so daß sie sich nicht gegenseitig beeinflussen.

So untersuchte K. Schroeder [SERB98, SAB⁺00, SABB02] das Verhalten von Silizium- und Germanium-Adatomen auf Arsen-terminierten Si(111)-Oberflächen. Dazu betrachtete er sowohl Diffusionspfade einzelner Atome auf der Arsen-terminierten Si(111)-Oberfläche, als auch den Austausch eines Siliziumatoms mit einem Arsenatom der Oberfläche. Wie oben beschrieben muß dieser Austauschprozess stattfinden, damit die Adatomlage auf der Oberfläche ‘schwimmt’. Erst die Simulation ermöglichte, die genauen Pfade und die Energie des Systems entlang dieser Pfade zu berechnen. Die Betrachtung der Energie entlang der Reaktionspfade lieferte dabei Aktivierungsenergie und Energiegewinn der betrachteten Reaktionen und somit die grundlegenden Parameter, um die Kinetik der Adatome auf solchen Oberflächen zu verstehen. Somit konnte die durchschnittliche Länge des Diffusionspfades eines Adatoms bis zum Einbau berechnet werden. Diese Ergebnisse brachten erhebliche Fortschritte beim Verständnis des Lagenwachstums auf der As-bedeckten Si(111)-Oberfläche [SABB02]:

- Siliziumatome werden sehr schnell in die Arsenschicht eingebaut, da die

Aktivierungsenergie für den Si-As-Austausch etwa gleich groß ist wie die Barriere für die Diffusion auf der Arsen-terminierten Oberfläche (0.25 eV)

- Für Germaniumatome ist die Austauschbarriere viel höher (0.7 eV). Sie werden daher längere Diffusionswege auf der Oberfläche zurücklegen und können so vorhandene Stufenkanten erreichen, bevor sie in die As-Schicht eingebaut werden.

Um das Lagenwachstum vollständig zu verstehen und es letztlich simulieren zu können, ist es nötig, alle in Frage kommenden Reaktionsprozesse zu betrachten. Die so berechneten Energien entlang der Reaktionspfade können dann als Grundlage einer Monte-Carlo-Simulation des Oberflächenwachstums dienen. Eine grobe Vorstellung des epitaktischen Wachstums und der daran beteiligten Prozesse auf der Arsen-bedeckten Si(111)-Oberfläche liefert folgendes Bild [Ant01]:

- Es wird von einer Oberfläche ausgegangen, auf der sich die verschiedenen Atomsorten durch Diffusionsprozesse bewegen können.
- Die deponierten Adatome werden einzeln in die Oberfläche eingebaut.
- In benachbarten Positionen werden weitere Atome eingebaut, es bilden sich Cluster.
- Durch Anlagerung weiterer Atome wachsen die Cluster und bilden As-bedeckte Inseln, die durch Stufenkanten begrenzt sind.
- Es finden auch Diffusionsprozesse über und entlang Stufenkanten statt.
- Durch Zusammenwachsen der Inseln bildet sich wiederum eine neue Lage.

Tatsächlich laufen natürlich all diese Prozesse auf verschiedenen Lagen gleichzeitig ab. Ob und mit welcher Wahrscheinlichkeit in einer gegebenen Situation einer dieser Prozesse stattfindet, hängt vor allem von den Aktivierungsenergien für die jeweiligen Reaktionspfade ab. Wenn die Inseln groß genug sind, ist ihr Rand äquivalent zu einer Stufenkante. Somit ist die Untersuchung von Stufenkanten ein wichtiger Schritt für ein umfassendes Verständnis des Lagenwachstums von Oberflächen. Experimentell findet man auf As-bedecktem Si(111) nur Inseln, die durch Stufen mit der Höhe einer Doppellage begrenzt sind [VZWB95]. Dicht gepackte Stufenkanten kommen in zwei Orientierungen vor: die $(11\bar{2})$ -Kante und die gegenüberliegende $(\bar{1}\bar{1}2)$ -Kante.

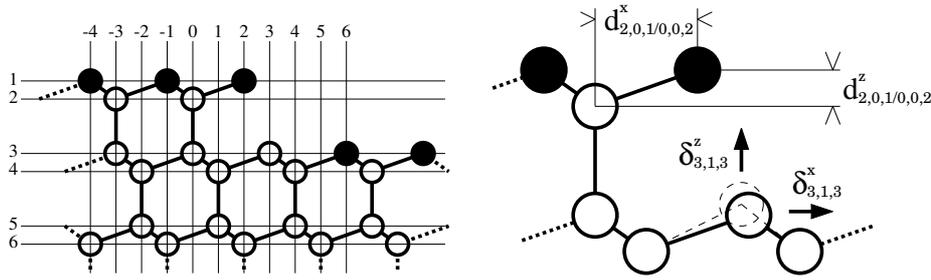


Abbildung 4.2: Indizierung der Atome für die $(11\bar{2})$ -Stufenkante, Seitenansicht; hier am Beispiel einer Kante mit überhängendem Arsen. Gefüllte Kreise stehen für As, offene für Silizium. In z -Richtung werden Ebenen, beginnend mit der obersten, durchgezählt. Abstände zwischen Atomen werden mit $d_{atom_1,atom_2}$ bezeichnet, Verschiebungen gegenüber der idealen Gitterposition mit δ_{atom} . Positive Verschiebungen weisen in z -Richtung aus der Ebene heraus und in x -Richtung entsprechend der Kantennormalen (also von der oberen Ebene weg, bzw. zur unteren Ebene hin).

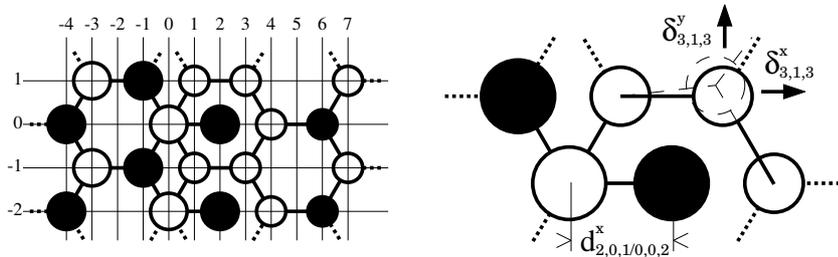


Abbildung 4.3: Indizierung der Atome für die $(11\bar{2})$ -Stufenkante, Aufsicht. Kleinere Kreise stehen für Atome der unteren Terrasse. In x -Richtung wird so indiziert, daß positive Indices auf der unteren, negative auf der oberen Terrasse liegen. Der Nullpunkt wurde dabei auf den Kantenatomen in der zweiten Ebene (hier die letzte Siliziumreihe der oberen Terrasse). Die Ausrichtung der y Koordinate ist schließlich so gewählt, daß ein rechtshändiges System entsteht. Da die Koordinaten äquidistant gewählt wurden, sind nicht alle Punkte des Koordinatennetzes besetzt.

4.2 Aufbau der Superzelle und Bezeichnungen

4.2.1 Bezeichnungen an den Stufenkanten

Da die Richtungen auf der (111) -Ebene nicht mit den aufspannenden Vektoren der kubischen Einheitszelle zusammenfallen, wird im weiteren Text folgendes Koordinatensystem gebraucht:

- Die z -Richtung ist senkrecht zur Oberfläche, fällt also mit der (111) -Richtung zusammen und zeigt von der Oberfläche in Richtung Vakuum.
- Die x -Richtung wird senkrecht zur Kante in der (111) -Ebene gewählt und zeigt von der höheren zur niedrigeren Terrasse. Je nach Stufenkante ist dies die $(11\bar{2})$ bzw. die $(\bar{1}\bar{1}2)$ -Richtung.

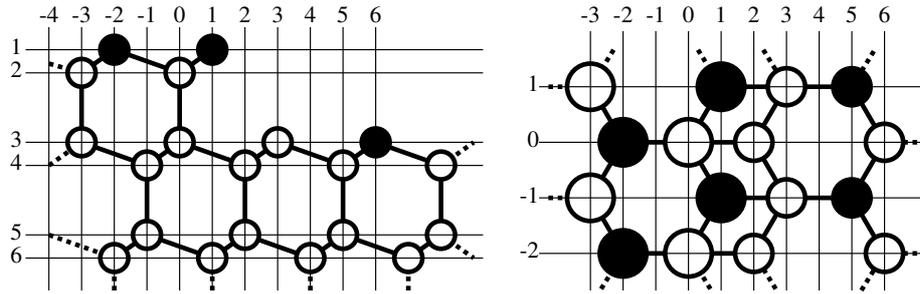


Abbildung 4.4: Indizierung der Atome für die $(\bar{1}\bar{1}2)$ -Stufenkante, Seitenansicht und Aufsicht, auch hier wieder für die Kante mit überhängendem Arsen.

- Die y -Richtung zeigt entlang der Stufenkante und wird so gewählt, daß x , y und z ein rechtshändiges System bilden.

Dies legt ein rechtshändiges orthonormales System fest. Um die Atome zu bezeichnen, wurde ein orthogonales Gitter entlang dieser Koordinaten so gelegt, daß alle Atompositionen des Siliziumgitters auf einen Punkt des orthogonalen Gitters fielen, wie in den Abbildungen 4.2 bis 4.4 gezeigt. Die Nummerierung wurde dabei in x - und y -Richtung aufsteigend gewählt¹, bei den Ebenen wurde jedoch, beginnend bei der oberen Lage der oberen Terrasse, in den Bulk hinein gezählt², so daß die erste und zweite Lage die obere Doppellage ausmachen, während die Doppellage der unteren Terrasse aus der dritten und vierten Ebene besteht.

Weiter wird von einer Perspektive ausgegangen, die ein von der Stufenkante auf die untere Terrasse blickender Beobachter hätte, wenn die Begriffe ‘vor’ (in x -Richtung), ‘hinter’ (entgegengesetzt), ‘neben’ (positive oder negative y -Richtung) und ‘unter’ (negative z -Richtung) verwandt werden. Diese sind als grobe Richtungsangaben zu verstehen, sie dienen nur zur zusätzlichen Orientierung und Unterstützung der Vorstellung. Zusätzlich werden in y -Richtung, also parallel zur Stufenkante, nebeneinanderliegende Atome als ‘Reihe’ bezeichnet. Da die Ausgangskonfigurationen der Stufen durch ‘Abschneiden’ der obersten Doppellage nach einer Atomreihe entsteht, bilden diese vordersten Atome einer Doppellage eine Reihe. Diese Atome werden als Kantenatome bezeichnet.

Bei Materialberechnungen im **ESTCoMPP**-Programm wird durch Energieminimierung in einer Bulkrechnung zunächst eine theoretische Gitterkonstante berechnet. Diese hängt von der Wahl des Basissatzes, also der \mathbf{k} -Punkt-Dichte und des cut-offs \mathbf{G}_{max} der ebenen Wellen, ab und weicht leicht von der physikalischen Gitterkonstante 5.4304\AA ab, wie man Tabelle 4.1 entnehmen kann.

Sämtliche Rechnungen wurden bei $\mathbf{G}_{max} = 3.0$ durchgeführt. Da bei den Berechnungen der Oberflächen und Stufenkanten die z -Richtung der Superzelle

¹In x -Richtung wird dabei als Nullpunkt das letzte Atom der zweiten Lage (untere Lage der oberen Terrasse) gewählt.

²Dabei wurden Ebenen des orthogonalen Gitters, die nicht mit Ebenen des Siliziumgitters zusammenfallen, übersprungen

Si	$\mathbf{G}_{max}=3.0$	$\mathbf{G}_{max}=3.5$	$\mathbf{G}_{max}=4.0$
$5 \times 5 \times 5$	5.4259	5.4110	5.3970
$6 \times 6 \times 6$	5.4247	5.4077	5.3941
$9 \times 9 \times 9$	5.4138	5.4062	5.3928
$12 \times 12 \times 12$	5.4234	5.4063	5.3928

Tabelle 4.1: Abhängigkeit der theoretischen Gitterkonstanten (in Å) von \mathbf{G}_{max} und der \mathbf{k} -Punkt-Dichte.

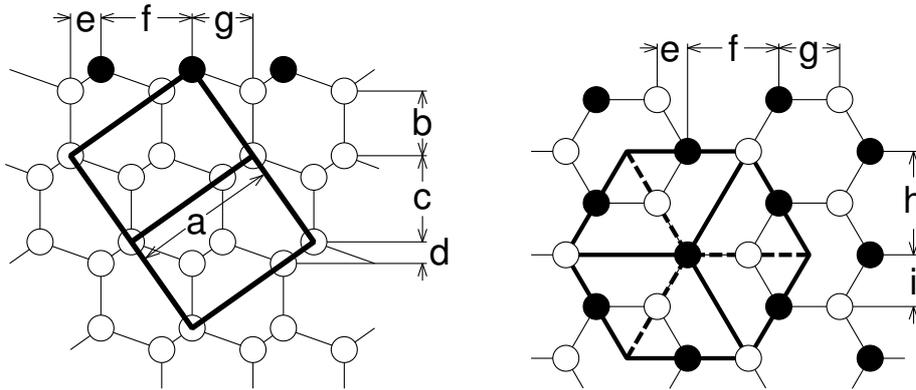
entlang der (111)-Richtung des Gitters ausgerichtet ist, wurde zur Berechnung der theoretischen Gitterkonstanten und der Referenzenergie für Si-Bulk auch eine hexagonale Zelle mit einer Basis aus 6 Atomen gewählt, bei der die z -Achse entlang der (111)-Richtung des Siliziums lag. Die Berechnung des Si-Bulk wurde mit $8 \times 8 \times 3$ \mathbf{k} -Punkten durchgeführt. Dies entspricht einer \mathbf{k} -Punktdichte zwischen $8 \times 8 \times 8$ und $9 \times 9 \times 9$ \mathbf{k} -Punkten in der fcc-Einheitszelle. Die so berechnete theoretische Gitterkonstante ist 5.4239 \AA und die Referenzenergie -7.93113 Ry/Atom .

Bei den Oberflächen- und Stufenrechnungen ist nur die laterale \mathbf{k} -Punktdichte wichtig, da vorausgesetzt wird, daß sich die Oberflächen nicht gegenseitig beeinflussen und die Wellenfunktionen im Vakuum verschwinden. Daher hängen sie nicht von der z -Koordinate des \mathbf{k} -Punktes (im reziproken Raum) ab, und in z -Richtung ist keine Dispersion zu beobachten. Die Referenz für die As-bedeckte Oberfläche wurde mit 4×4 hexagonalen Zellen und $2 \times 2 \times 1$ \mathbf{k} -Punkten berechnet. Damit entspricht ihre laterale \mathbf{k} -Punktdichte genau der Si-Bulk Referenz.

Um Verzerrungseffekte zu vermeiden, werden Strukturrechnungen für jedes Material mit der theoretischen Gitterkonstanten durchgeführt. Da sich in der Ausgabe des Programms sämtliche Angaben absoluter Längen auf die theoretische Gitterkonstante beziehen, wurden diese Werte in relativen Maßeinheiten, also bezogen auf die benutzte Gitterkonstante, angegeben. Maßangaben wie zum Beispiel Verschiebungen der Atome gegenüber ihren Idealpositionen sind in Prozent der Gitterkonstanten a angegeben. In diesen Einheiten sind die meisten Verschiebungen von der Größenordnung $\mathcal{O}(1)$. Abb. 4.5 enthält einige Längenangaben in dieser Form für das ideale hexagonale Gitter.

Si	$\mathbf{G}_{max}=3.0$	$\mathbf{G}_{max}=3.5$	$\mathbf{G}_{max}=4.0$
$5 \times 5 \times 5$	-7.92906	-7.94451	-7.95043
$6 \times 6 \times 6$	-7.93018	-7.94611	-7.95198
$9 \times 9 \times 9$	-7.93122	-7.94698	-7.95285
$12 \times 12 \times 12$	-7.93110	-7.94703	-7.95291

Tabelle 4.2: Abhängigkeit der Energie pro Atom (in Rydberg) von \mathbf{G}_{max} und der \mathbf{k} -Punkt-Dichte.



a	a	100.00	Gitterkonstante, Seitenlänge der kubischen Zelle
b	$\sqrt{3}a/4$	43.30	Nächst-Nachbar-Abstand
c	$\sqrt{3}a/3$	57.74	Höhenunterschied zwischen zwei Doppелеbenen
d	$\sqrt{3}a/12$	14.43	Höhenunterschied der Atome einer Doppелеbene
e	$\sqrt{6}a/12$	20.41	minimale y-Komponente des Abstandes zweier Atome
f	$\sqrt{6}a/4$	61.24	Abstand zweier 'Reihen' auf der (111)-Oberfläche
g	$\sqrt{6}a/6$	40.82	Seitenlänge der Hexagonalen Zelle in der Aufsicht
h	$\sqrt{2}a/2$	70.71	Höhe einer hexagonalen Zelle auf der (111)-Oberfläche
i	$\sqrt{2}a/4$	35.36	Halbe Höhe der Hexagonalen Zelle

Abbildung 4.5: Längen im idealen hexagonalen Gitter relativ zur Gitterkonstanten a . Die Buchstaben $a-i$ in der Tabelle bezeichnen die entsprechenden Abstände im obigen Diagramm. Im weiteren Text werden nur die Gitterkonstante a und der Nächstnachbarabstand b benutzt. Die Längen sind in Prozent der Gitterkonstanten a angegeben, um einen direkten Vergleich mit den Verschiebungen δ zu ermöglichen, welche auch in '% a ' angegeben sind.

Im Fall von Bindungslängen sind diese in Prozent der Bindungslänge im Bulk angegeben, um einen leichteren Vergleich mit dem Abstand benachbarter Gitterpositionen im idealen Gitter zu erlauben. In physikalischen Systemen entspricht somit $100\% \simeq 2.35\text{\AA}$ der Si-Si-Bindungslänge im Silizium-Kristall. Für die As-Si und As-As Bindungen, die in den betrachteten Strukturen auftreten, sind größere Bindungslängen zu erwarten. Die As-As-Bindungslänge in kristallinem Arsen beträgt 2.51\AA . Dies entspricht 106.7% der Bindungslänge in Silizium-Bulk, was in den meisten Fällen recht gut mit den As-As Bindungslängen in den relaxierten Kanten-Strukturen übereinstimmt. Für As-Si Bindungslängen ist etwa der Mittelwert zwischen As-As- und Si-Si-Bindungslängen zu erwarten, also etwa 103% - 104% . In diesem Bereich liegt mit 103.79% auch die Bindungslänge zwischen As und Si auf der As-terminierten Si(111)-Oberfläche.

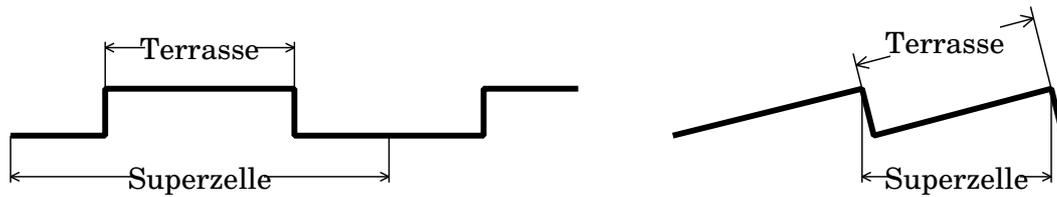


Abbildung 4.6: Superzellenkonstruktionen zur Untersuchung von Stufenkanten. Links: eine auf- und eine absteigende Stufenkante schließen eine Terrasse ein. Rechts: Indem die Superzelle nicht parallel zur Terrasse ausgerichtet ist, wird eine hoch indizierte Oberfläche berechnet, die auch als endlose Treppe betrachtet werden kann.

4.2.2 Gestaltung der Superzelle

Wie schon in Kapitel 2.1.5 beschrieben, berechnet das **ES**tCoMPP-Programm periodische Strukturen, somit ist die Berechnung einer einzelnen Stufenkante nicht möglich. Daher wird tatsächlich eine sich periodisch wiederholende Anordnung von Stufenkanten untersucht. Wie auch bei der Untersuchung von einzelnen Defekten wird dabei davon ausgegangen, daß die periodisch angeordneten Stufenkanten durch ausreichend breite Terrassen voneinander getrennt sind, so daß sie sich nicht gegenseitig beeinflussen. Da es an einer Stufenkante naturgemäß zu einem Höhenunterschied in der Oberfläche kommt, bieten sich dazu, wie in Diagramm 4.6 gezeigt, grundsätzlich zwei Möglichkeiten, damit eine Superzelle bündig (ohne unterschiedlich hohe Oberflächen) an die ihre nächste Repräsentation anschließt:

- Es werden zwei Stufenkanten berechnet, eine auf- und eine absteigende, so daß insgesamt kein Höhenunterschied entsteht.
- Es wird eine hoch indizierte Oberfläche berechnet, also eine leicht geneigte Oberfläche, auf der die parallelen Stufenkanten eine unendliche Treppe bilden.

Im ersten Fall enthält jede Superzelle zwei Stufenkanten, was folgende gravierenden Nachteile hat: zum einen muß die Superzelle in der Raumrichtung parallel zur Oberfläche und senkrecht zur Stufenkante doppelt so lang gewählt werden, um zwei Stufenkanten und den dazwischenliegenden Terrassen Platz zu bieten. Zum anderen sind ‘gegenüberliegende’ Stufenkanten verschiedenartig, was die unabhängige Energiebetrachtung für verschiedene mögliche Stufenkanten erheblich erschwert. Im zweiten Fall gestaltet sich hingegen der Aufbau der Superzelle schwieriger, da die Oberflächennormale der Si(111)-Oberfläche nicht mit der Oberflächennormalen der gewählten Superzelle zusammenfällt und die Gitterebenen somit ‘schräg’ in der Superzelle liegen. Um die Superzelle überhaupt auf eine berechenbare Größe zu beschränken und auch um die Auswertung der so berechneten Energien zu vereinfachen, wurde der zweite Weg gewählt.

Die Superzelle wird durch die Vektoren \mathbf{b}_1 , \mathbf{b}_2 und \mathbf{b}_3 aufgespannt. Dabei wird der \mathbf{b}_3 -Vektor stets senkrecht zur Terrassenfläche und der \mathbf{b}_2 -Vektor entlang der

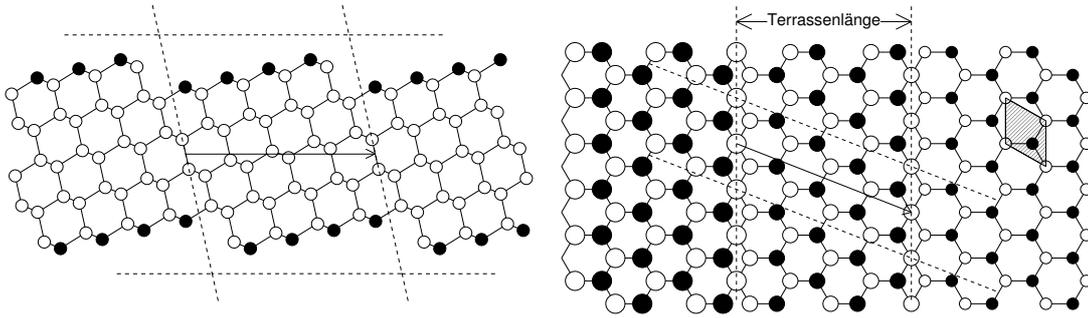


Abbildung 4.7: Seitenansicht und Aufsicht der $(11\bar{2})$ -Stufenkante, hier abgeschnitten, mit Siliziumatom an der Kante. Die Superzelle ist gestrichelt eingezeichnet. Die schraffierte Fläche stellt die Elementarzelle einer idealen Si(111)-Oberfläche dar. Jede Lage ist um ein Drittel gegenüber der darüberliegenden verschoben, so daß zwischen zwei Stufenkanten keine ganze Anzahl Oberflächenelementarzellen passt. Der Stufenabstand wird durch den \mathbf{y} -Vektor vorgegeben.

Stufenkante ausgerichtet. Der \mathbf{b}_1 -Vektor ist parallel zur tatsächlichen hoch indizierten Oberfläche ausgerichtet, z.B. von einem Atom an einer Stufenkante zum Repräsentanten dieses Atoms an der dazu parallelen benachbarten Stufe³. Die Länge des \mathbf{b}_2 -Vektors ist als Vielfaches des Gitterabstandes zweier Atome entlang der Stufenkante, also in y -Richtung, zu wählen. Durch den \mathbf{b}_1 -Vektor wird die Terrassenlänge festgelegt. Er zeigt von einem Atom einer Terrasse zu seinem Repräsentanten in der nächsten. Es ist möglich, diesen Vektor senkrecht zur Kante zu wählen, dies schränkt jedoch die Wahl möglicher Terrassenlängen ein⁴. Um in der \mathbf{b}_3 -Richtung Periodizität zu erzielen, werden, wie bei Oberflächenrechnungen mit periodischen Randbedingungen üblich, stets zwei gegenüberliegende Oberflächen berechnet, so daß sich periodisch Materialschichten, getrennt durch Vakuum, wiederholen, wie in Abb. 2.1 in der Mitte dargestellt. Zur Erleichterung der späteren Auswertungen, z.B. der Dichteinformationen in Bezug zur Terrassenoberfläche, wurde dieser Vektor, wie bereits erwähnt, senkrecht zur Terrassenebene gewählt. Die Länge dieses Vektors ist dadurch bestimmt, wieviele Materialschichten in die Tiefe berechnet werden und wieviel Vakuum über den Oberflächen nötig ist, um auszuschließen, daß sich die Oberflächen, sowohl über das Vakuum als auch durch die Materialschicht, gegenseitig beeinflussen. Indem man die Zelle inversionssymmetrisch aufbaut, wird die Berechnung der zwei Oberflächen wiederum vereinfacht, da so die Zahl der Freiheitsgrade der beweglichen Atome halbiert wird.

Es ist wünschenswert, mit möglichst kleinen Superzellen zu rechnen, da sich die Größe der Zelle und die Anzahl der zu berechnenden Atome stark auf die Rechenzeit auswirken. Gleichzeitig ist es aber nötig, die Zelle groß genug zu

³Somit fällt \mathbf{b}_2 mit der y -Richtung, und \mathbf{b}_3 mit der z -Richtung zusammen.

⁴Je nach Ausrichtung der Kante kommen dann nur gerade bzw. nur ungerade Terrassenlängen in Frage.

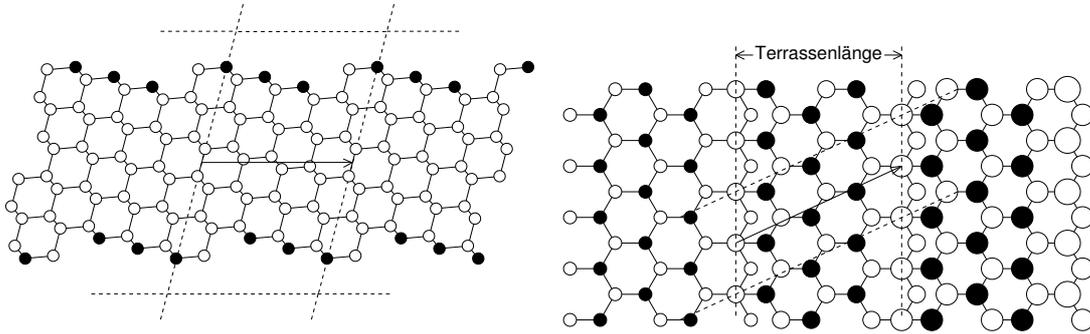


Abbildung 4.8: Seitenansicht und Aufsicht der $(\bar{1}\bar{1}2)$ -Stufenkante, wieder mit Siliziumatom an der Kante.

wählen, um qualitativ brauchbare Resultate zu erhalten. Dabei lassen sich zur Abschätzung der ‘Höhe’ der Superzelle, also der Länge des \mathbf{b}_3 -Vektors, und zum nötigen Abstand zwischen zwei Stufenkanten Erfahrungswerte aus anderen Rechnungen, beispielsweise Berechnungen von Adatomen, heranziehen. Als Terrassenlänge wurden dabei vier Doppelreihen Atome gewählt, die Oberflächen wurden durch 6 Lagen Bulkmaterial voneinander getrennt. Um die Zelle nicht unberechenbar groß werden zu lassen, war es nach Vorgabe dieser Parameter nötig, die Gitterkonstante der Zelle in \mathbf{b}_2 -Richtung möglichst klein zu wählen. Als praktikables Maximum erwies sich dabei eine Länge von drei Kantenatomen. Dies würde bedeuten, daß jedes dritte Atom der Kante das gleiche Atom repräsentiert, der Kante würde also eine künstliche Dreier-Periodizität aufgezwungen.

Bei ersten Berechnungen zeigte sich dann, daß an der Kante ‘Dimerisierungen’ auftraten, d.h. zwei benachbarte Atome bewegten sich aufeinander zu, um so ihre freien Bindungen möglichst abzusättigen. Durch die künstlich erzwungene Periodizität ergab sich, daß das dritte Kantenatom keinen Partner fand, da sich die beiden anderen Atome, indem sie sich aufeinander zubewegten, von diesem wegbewegt hatten. Dies führte zu sichtlich unphysikalischen Ergebnissen, weshalb in \mathbf{b}_2 -Richtung eine andere Länge für die Superzelle gewählt wurde. Um die bereits beobachteten Dimerisierungen weiterhin zuzulassen, wurde daher für diese Gitterkonstante der doppelte Abstand zweier Kantenatome gewählt.

Bei genauerer Betrachtung der Terrassenlängen ergibt sich, daß sich diese nicht als ganzzahliges Vielfaches des Abstandes zweier Atomreihen in der Ebene ausdrücken lassen. Das liegt daran, daß die obere Doppellage gegenüber der darunterliegenden um ein Drittel dieses Abstandes verschoben ist. Somit sind die Terrassenlängen entsprechend verkürzt bzw. verlängert. Dies muß bei den Berechnungen der Oberflächenenergien berücksichtigt werden, da bei diesen gleich große Oberflächen verglichen werden müssen.

Aufgrund der Konstruktion der Zelle ist deren Länge in \mathbf{b}_1 -Richtung kein ganzzahliges Vielfaches der Länge der Einheitszelle des Gitters, was zu einer inkommensurablen Zahl von \mathbf{k} -Punkten pro Einheitszelle führt. Zur Berechnung der

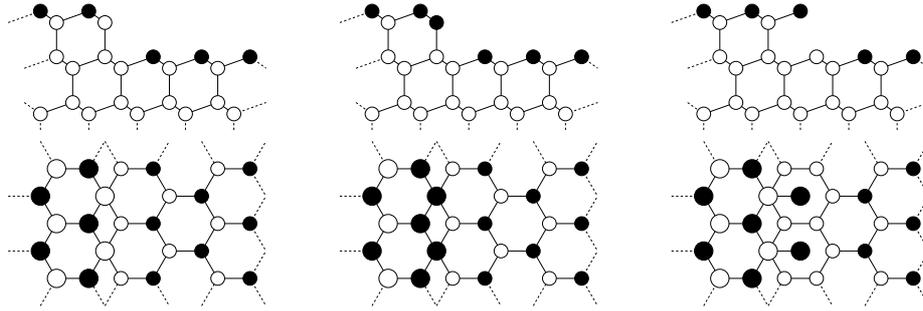


Abbildung 4.9: Verschiedene Ausgangspunkte für Strukturen an der $(11\bar{2})$ -Stufenkante, von links nach rechts: Abgeschnitten mit Siliziumatom an der Kante, Silizium an der Kante ersetzt durch Arsen, überhängendes Arsenatom.

Bulkeigenschaften, also der theoretischen Gitterkonstanten und der Referenzenergie für Bulk-Si, wird daher die \mathbf{k} -Punktdichte angenähert, indem für die Zahl der \mathbf{k} -Punkte pro Einheitszelle die nächstliegende ganze Zahl gewählt wird. Wie den Abbildungen 4.7 und 4.8 zu entnehmen ist, entspricht die laterale Ausdehnung der Superzelle annähernd 4×2 Einheitszellen der Oberfläche. Daher entsprechen $2 \times 4 \times 1$ \mathbf{k} -Punkte sehr gut der \mathbf{k} -Punktdichte der Referenzrechnung. Die lateralen Ausdehnungen der Superzellen, und somit die lateralen \mathbf{k} -Punktdichten der Stufenrechnungen, weichen von der einer 4×2 -Zelle um $\pm 1/12$, also ca. 8% ab. Wie man Tabelle 4.1 entnehmen kann, liegt die Differenz der theoretischen Gitterkonstanten für verschiedene \mathbf{k} -Punktdichten im Promillebereich. Die Kräfte, die sich aus der Abweichung der theoretischen Gitterkonstanten aufgrund der leicht veränderten \mathbf{k} -Punktdichte ergeben, können daher vernachlässigt werden.

4.3 Struktur der Stufenkanten

Bei der Struktur der Stufenkanten ist es nötig, sinnvoll erscheinende Strukturen vorzugeben und diese relaxieren zu lassen. Ausgehend von der vorgegebenen Konfiguration findet das **ESCoMPP**-Programm das im Konfigurationsraum nächstliegende lokale Minimum. Erst durch Vergleich der Energien dieser Konfigurationen läßt sich beurteilen, wie stabil diese Kantenstruktur tatsächlich ist. Die Ergebnisse der Rechnung sind also wesentlich von der Wahl sinnvoller Ausgangskonfigurationen abhängig. Wie bereits erwähnt, wird dazu die oberste Doppellage entlang einer Linie einfach abgeschnitten und die oberen Atome der darunterliegenden freiwerdenden Doppellage durch Arsenatome ersetzt. Je nachdem, an welcher Stelle man die obere Doppellage dabei abschneidet, erhält man entweder eine Reihe Siliziumatome (in der zweiten Lage) oder Arsenatome (überhängend, in der ersten Lage) an der Kante.

Die Atome an der Kante haben im Vergleich zu anderen Oberflächenatomen der gleichen Lage ein oder zwei nächste Nachbarn weniger. Es hat sich schon in anderen Rechnungen gezeigt, daß energetisch günstige Konfigurationen oft dar-

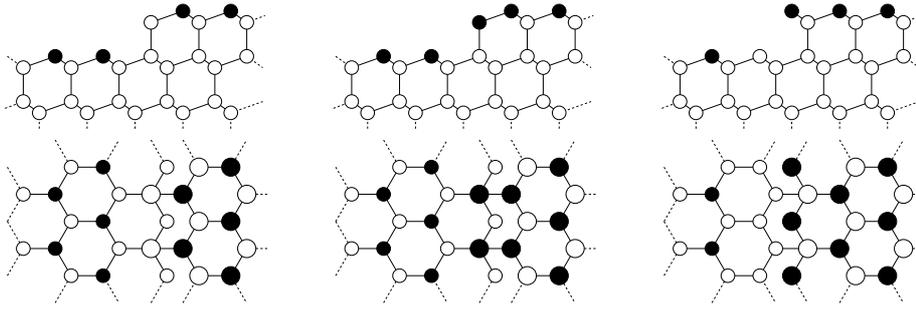


Abbildung 4.10: Verschiedene Ausgangspunkte für Strukturen an der $(\bar{1}\bar{1}2)$ -Stufenkante, von links nach rechts: Abgeschnitten mit Siliziumatom an der Kante, Silizium an der Kante ersetzt durch Arsen, überhängendes Arsenatom.

an erkennbar sind, daß sämtliche Valenzbindungen abgesättigt sind, jedes Atom also so viele freie Nachbarn findet, daß es Edelgaskonfiguration erreicht. Bei den beteiligten Atomsorten Silizium und Arsen ist dabei von kovalenten Bindungen auszugehen. Das heißt, Siliziumatome werden idealerweise vier nächste Nachbarn haben und Arsenatome drei, wie dies auch bei der Arsen-terminierten Si(111)-Oberfläche der Fall ist.

Da die an den Kanten gelegenen Siliziumatome zu wenige nächste Nachbarn haben, wurden auch Kantenstrukturen berechnet, bei denen diese Atome durch Arsenatome ersetzt wurden. Während Arsenatome in der unteren Schicht einer Doppellage üblicherweise nicht vorkommen, erscheint ihre Position an der Kante energetisch günstig, da sie dort ihre optimale Koordinationszahl leichter erreichen können als Siliziumatome in der gleichen Position.

Wir haben also drei verschiedene Konfigurationen der Stufenkanten gerechnet, und zwar für beide Kantennormalen $(11\bar{2})$ und $(\bar{1}\bar{1}2)$:

- (1) Kante mit exponiertem Si-Atom in der 2. Lage.
- (2) Kante mit überhängendem As-Atom.
- (3) Kante mit As-Atom in der 2. Lage, an Stelle des exponierten Si in (1).

Die verschiedenen Kantenstrukturen wurden jeweils relaxiert, bis Konvergenz der Kräfte bis auf 10^{-4} Ry/a.u. eintrat. Die relaxierte Struktur ist in den folgenden Diagrammen und Tabellen angegeben. Um die Identifizierung der Atome zu erleichtern, wurden sie zusätzlich nummeriert. Im unteren Teil jedes Bildes sind die Elektronendichten durch Iso-Flächen dargestellt, um Bindungen zwischen den Atomen zu verdeutlichen. Bei den angedeuteten Bindungen in den oberen Bildteilen ist tatsächlich eine erhöhte Elektronendichte zwischen den jeweiligen Atomen vorhanden, so daß dort von kovalenten Bindungen ausgegangen werden kann.

Wie oben bereits erwähnt, beziehen sich die Auslenkungen δ auf die Verschiebung gegenüber einer Position im idealen Gitter. Sie sind in Prozent der Gitterkonstanten ('%a') angegeben. Nächstnachbarabstände sind in Prozent des

Nächstnachbarabstandes ('%d') im idealen Gitter angegeben, um den Vergleich von Bindungslängen untereinander zu erleichtern. Bei diesen Angaben ist zu berücksichtigen, daß schon bei der Arsen-terminierten glatten Si(111)-Oberfläche die Atome der obersten Lagen gegenüber der idealen Gitterposition verschoben sind. Aus Symmetriegründen treten dabei jedoch nur Verschiebungen in z -Richtung auf, die zudem für alle Atome einer Lage gleich sind. Dabei ist die Arsenlage um 3.96%a nach oben verschoben ($\delta_{*,*,1}^z = 3.96$), die darunterliegende Siliziumlage (in der gleichen Doppellage) um 0.4%a nach unten ($\delta_{*,*,2}^z = -0.4$), in der nächsten Siliziumlage ist die Verschiebung nur noch 0.11%a nach unten ($\delta_{*,*,3}^z = -0.11$), die vierte Lage wurde bei der Berechnung der reinen Arsen-terminierten Siliziumoberfläche festgehalten. Dies führt zu einer Bindungslänge von $d_{*,*,1/*,*,2} = 103.79\%$ d zwischen Arsenatomen der obersten und Siliziumatomen der darunterliegenden Lage. Diese Bindungen schließen miteinander, sowohl am Arsen als auch am Siliziumatom, einen Winkel von 103.74° ein. Die Bindungen der Siliziumatome der zweiten mit denen der dritten Lage haben eine Länge $d_{*,*,2/*,*,3} = 99.31\%$ d und schließen am oberen Siliziumatom einen Winkel von 114.79° mit den Si-As Bindungen in der ersten Doppellage ein.

4.3.1 Die $(11\bar{2})$ Stufe mit Siliziumatom an der Kante

Diese Stufenkante ergibt sich durch ‘Abschneiden’ der Oberfläche nach einer Reihe (in y -Richtung nebeneinanderliegender) Siliziumatome aus der zweiten Lage. Auf der tiefer liegenden Terrasse werden die Oberflächenatome (in der dritten Lage) durch Arsenatome ersetzt. Im Fall der $(11\bar{2})$ -Kante sind die an der Kante gelegenen Siliziumatome dreifach koordiniert: sie besitzen je zwei Bindungen zu Arsenatomen aus der ersten Lage und eine zu den Siliziumatomen unter ihnen. Sie erreichen somit aufgrund ihrer Koordination keine Edelgaskonfiguration.

Abb. 4.11 zeigt die relaxierte Konfiguration, Tabelle 4.3 zeigt einen Auszug der Koordinaten und Bindungswinkel aus Anhang A. Nach Relaxation aus der Ausgangslage und Erreichen des energetischen Minimums zeigt sich, daß an der Kante Symmetriebrechung auftritt: jedes zweite Kantenatom bewegt sich sichtbar in die Kante hinein, bis seine nächsten Nachbarn nahezu in einer Ebene angeordnet sind, die drei Bindungen an diesem Atom liegen dabei nahezu im 120° Winkel. In dieser Bindungsanordnung ist das Atom nahezu sp^2 -hybridisiert, wobei die hybridisierten Orbitale in Richtung der Bindungen ausgerichtet sind. Das nicht hybridisierte p-Orbital ist dabei senkrecht zur Bindungsebene ausgerichtet und energetisch ungünstiger als die sp^2 -Orbitale. Auch die verkürzten Bindungslängen an diesem Siliziumatom stehen mit der sp^2 -Hybridisierung in Einklang, da sp^2 -Orbitale aufgrund des höheren s-Anteils zu einer kürzeren Bindungslänge im Vergleich zu sp^3 -Orbitalen führen. Während der Abstand zum Siliziumatom mit 98.2% des Nächstnachbar-Abstandes im Bulk nur geringfügig von der typischen Si-Si-Bindungslänge abweicht, ist die Verkürzung der Si-As-Bindung auf 98.6% gegenüber sonst beobachteten Werten um 104% deutlich.

Wenn nun diese Konfiguration günstiger ist als eine Position näher der ursprünglichen Gitterposition, ist zu klären, warum sie nur von jedem zweiten Kantenatom eingenommen wird. Bei genauerer Betrachtung der Elektronendichte an diesen Atomen fällt auf, daß das nicht hybridisierte p-Orbital nicht besetzt ist. Dieses ragt senkrecht aus der Ebene der Bindungen heraus und eine entsprechende Ausprägung in der Elektronendichte ist nicht zu beobachten. Das vierte Elektron steht also den anderen Siliziumatomen an der Kante zur Verfügung, so daß diese mit fünf Elektronen und drei kovalenten Bindungen Edelgaszustand erreichen können. Dies zeigt sich auch am Backbond des Siliziumatoms, der von der Ausprägung her eher den Backbonds der Arsenatome in der Oberfläche entspricht. Auch entspricht die Anordnung der nächsten Nachbarn, und somit der Bindungen an diesen Atomen, noch am ehesten einer sp^3 -Hybridisierung und der entsprechenden Tetraederanordnung, wie sich an den Bindungswinkeln ablesen läßt: Die Winkel zwischen der Bindung zum daruntergelegenen Siliziumatom und der zu einem der schräg dahinterliegenden Arsenatome aus der ersten Lage betragen ca. 108.1° , liegen also nahe am Tetraederwinkel. Der Winkel zwischen den Bindungen zu den Arsenatomen beträgt nur 100.5° . Dieser kleine Winkel kommt im wesentlichen durch die größere Bindungslänge der Silizium-Arsen-Bindung

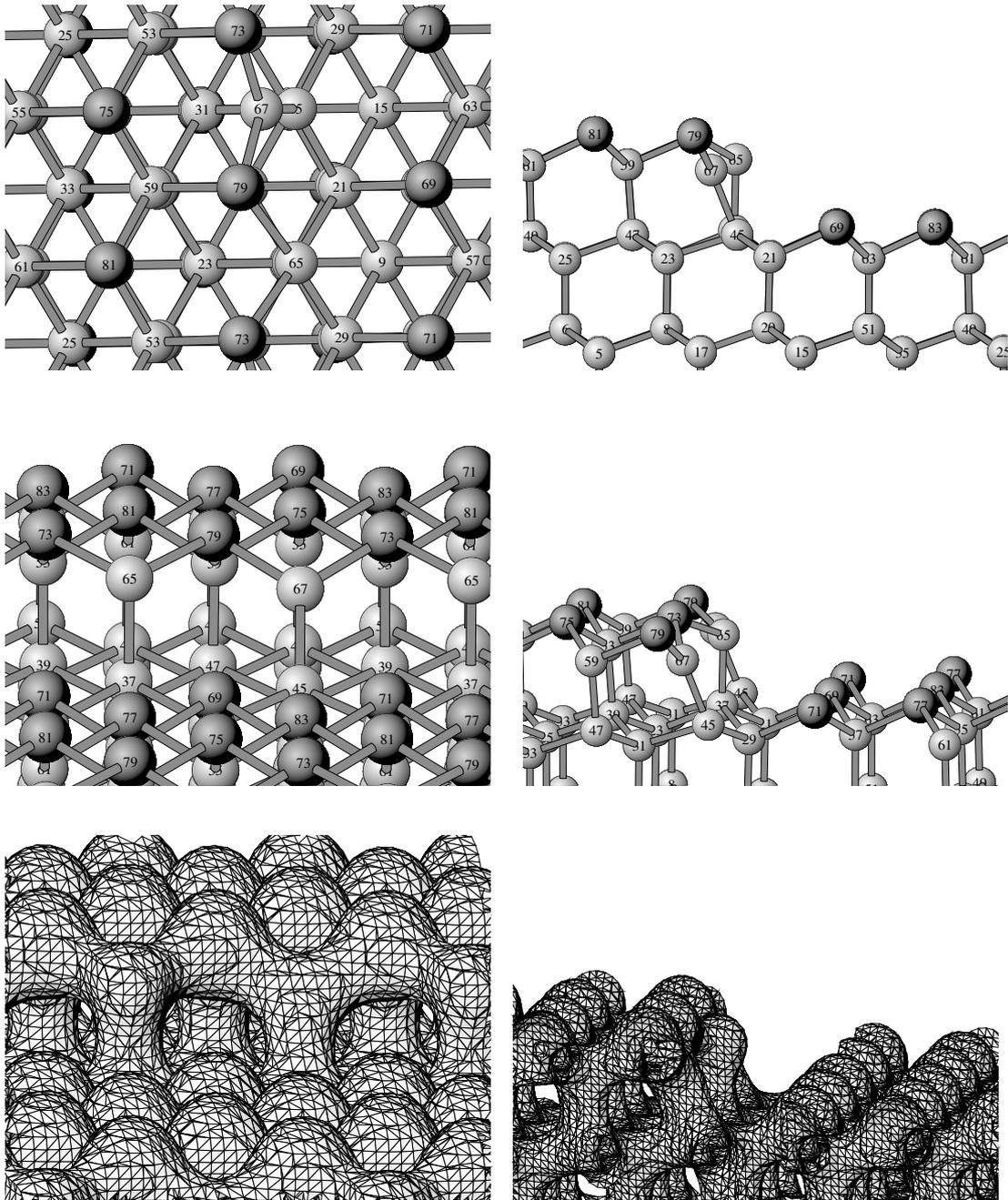


Abbildung 4.11: Seitenansicht und Aufsicht der $(11\bar{2})$ -Stufenkante mit Siliziumatom an der Kante. Arsenatome sind dunkel dargestellt. Zur besseren Identifizierung sind die Atome zusätzlich nummeriert. In den unteren Bildern sind Isosflächen der Elektronendichte dargestellt.

zustande. Zusätzlich bewegen sich die Arsenatome leicht vom sp^2 -hybridisierten Siliziumatom weg, was den Winkel am sp^3 -hybridisierten Siliziumatom zusätzlich verkleinert.

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge			
79	As	-3.09	-0.46	2.28	3.87	79-59	79-67	79-65	
-1, 1, 1						102.15	98.54	104.66	
73	As	-3.19	0.58	2.31	3.98	73-53	73-67	73-65	
-1, -1, 1						102.08	98.64	104.65	
67	Si	-13.48	0.06	-4.10	14.09	67-79	67-73	67-45	
0, 2, 2						98.54	98.64	98.23	
65	Si	1.43	0.02	2.04	2.49	65-79	65-73	65-37	
0, 0, 2						104.66	104.65	100.60	
45	Si	2.13	0.02	-0.37	2.16	45-67	45-31	45-21	45-29
0, 2, 3						98.23	102.75	100.26	100.22
37	Si	0.39	0.01	1.79	1.83	37-65	37-23	37-21	37-29
0, 0, 3						100.60	100.33	100.60	100.58
Atom		Winkel							
79	As	59-79-67	59-79-65	67-79-65					
-1, 1, 1		90.78	112.95	111.08					
73	As	53-73-67	53-73-65	67-73-65					
-1, -1, 1		90.72	113.06	110.84					
67	Si	79-67-73	79-67-45	73-67-45					
0, 2, 2		114.37	122.63	122.80					
65	Si	79-65-73	79-65-37	73-65-37					
0, 0, 2		100.47	108.10	108.13					
45	Si	67-45-31	67-45-21	67-45-29	31-45-21	31-45-29	21-45-29		
0, 2, 3		86.08	117.19	117.11	109.85	109.89	113.08		
37	Si	65-37-23	65-37-21	65-37-29	23-37-21	23-37-29	21-37-29		
0, 0, 3		113.20	110.31	110.35	108.96	108.90	104.75		

Tabelle 4.3: Koordinaten und Bindungswinkel der $(11\bar{2})$ -Stufe mit exponiertem Si.

Wesentlich direkter wirkt sich die Verschiebung des sp^2 -hybridisierten Siliziumatoms (Nummer 67 in Abb. 4.11) aber auf die Bindungswinkel an dessen direkten Nachbarn aus. Da das Atom mit den Arsenatomen der ersten Lage (73 und 79) und dem unter ihm liegenden Siliziumatom (45) in einer Ebene liegt, verkleinert sich der Winkel zu den jeweils dahinter gelegenen Atomen (53 und 59, bzw. 31) auf 90.75° an den Arsenatomen und sogar auf 86.1° am Siliziumatom.

4.3.2 Die $(11\bar{2})$ -Stufe mit Arsenatom an der Kante

Ausgehend von der Überlegung, daß die Atome kovalente Bindungen mit ihren nächsten Nachbarn ausbilden, ist offensichtlich, daß Siliziumatome, die idealerweise vier nächste Nachbarn haben, an der Kante in einer ungünstigeren Position

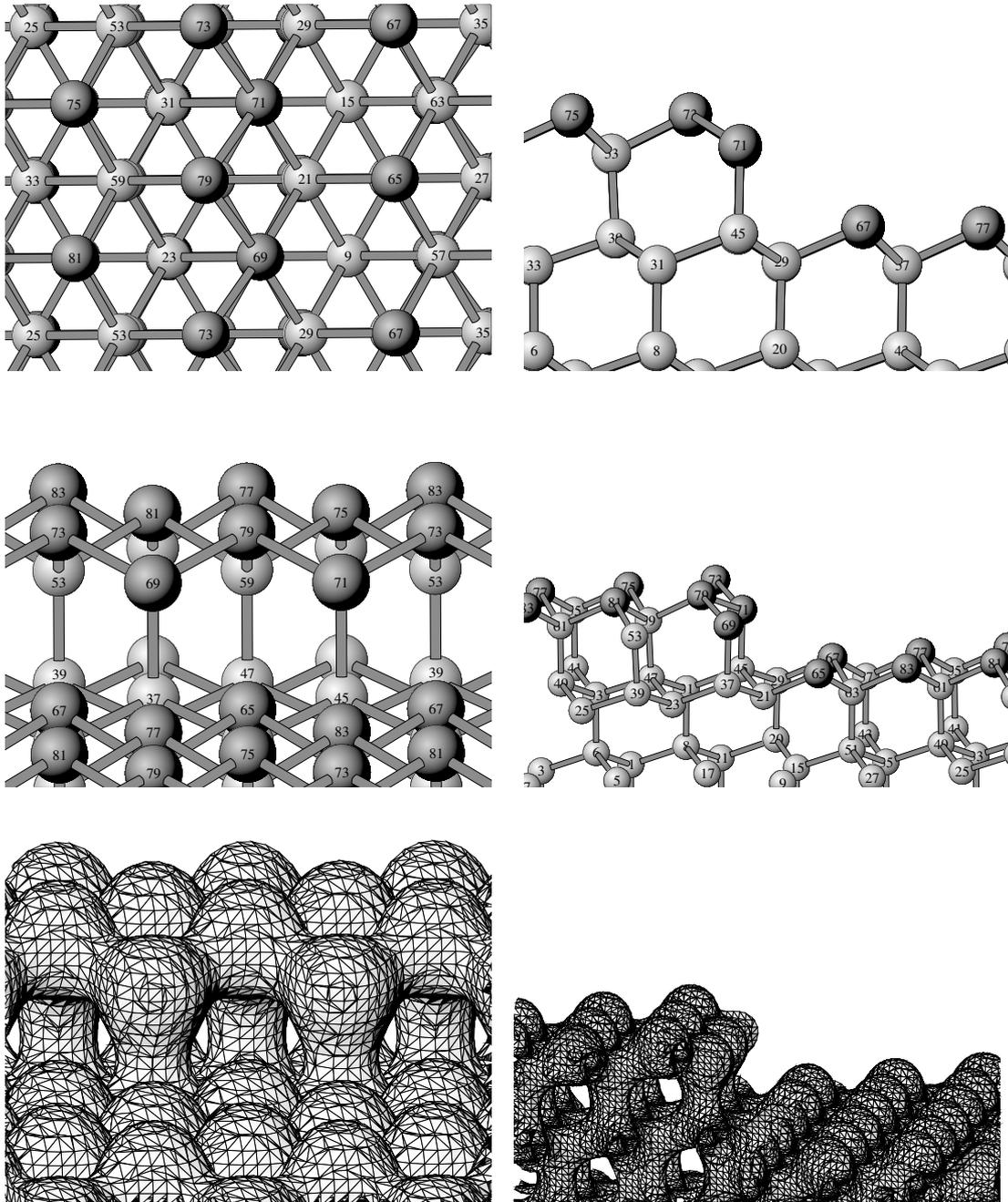


Abbildung 4.12: Seitenansicht und Aufsicht der $(11\bar{2})$ -Stufe mit Arsen an der Kante.

sind als Arsenatome, die nur drei nächste Nachbarn benötigen, um Edelgaskonfiguration zu erreichen. Somit sind Arsenatome nicht nur auf der Oberfläche, sondern auch an der $(11\bar{2})$ -Kante ideal koordiniert. Da in dieser Struktur sämtliche Atome ideal koordiniert sind, sind auch die Verschiebungen der Atome gegenüber den idealen Gitterpositionen vergleichsweise klein, wie sich Tabelle 4.4 entnehmen läßt. Die relaxierte Struktur zeigt, wie in Abb. 4.12 zu sehen, keine Symmetrie-

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge			
79(73)	As	-3.16	0.01	3.66	4.83	79-59	79-71	79-69	
-1, 1, 1						101.74	106.61	106.61	
59(53)	Si	-1.79	0.01	-1.49	2.33	59-75	59-81	59-79	59-47
-3, 1, 2						103.33	103.33	101.74	98.54
71(69)	As	1.33	0.02	1.93	2.34	71-79	71-73	71-45	
0, 2, 2						106.61	106.60	101.97	
45(37)	Si	0.18	0.01	1.09	1.11	45-71	45-31	45-21	45-29
0, 2, 3						101.97	100.34	101.17	101.17
65(67)	As	1.38	0.01	4.62	4.82	65-21	65-63	65-57	
3, 1, 3						102.90	103.42	103.43	

Atom		Winkel					
79(73)	As	59-79-71	59-79-69	71-79-69			
-1, 1, 1		109.11	109.10	99.97			
59(53)	Si	75-59-81	75-59-79	75-59-47	81-59-79	81-59-47	79-59-47
-3, 1, 2		104.41	101.71	116.47	101.72	116.46	114.01
71(69)	As	79-71-73	79-71-45	73-71-45			
0, 2, 2		99.97	109.63	109.65			
45(37)	Si	71-45-31	71-45-21	71-45-29	31-45-21	31-45-29	21-45-29
0, 2, 3		112.67	108.23	108.21	109.97	109.98	107.62
65(67)	As	21-65-63	21-65-57	63-65-57			
3, 1, 3		103.12	103.11	104.27			

Tabelle 4.4: Koordinaten und Bindungswinkel der $(11\bar{2})$ -Stufe mit As-Kantenatom in der 2. Lage.

brechung: die Verschiebungen der einzelnen Atome waren mit den in y -Richtung neben ihnen liegenden im Rahmen der Rechengenauigkeit identisch.

Die Verschiebungen der Atome gegenüber der Ausgangsposition sind vor allem auf die längeren As-As Bindungen der Kantenatome mit den Arsenatomen der ersten Ebene in der dahintergelegenen Reihe zurückzuführen. Diese Länge ist, wie zu erwarten, mit 106.6% deutlich länger als As-Si-Bindungen. Aus geometrischen Gründen führt eine Vergrößerung des As-As Abstandes zu einer Verkleinerung der Winkel dieser Bindungen zueinander, sonst würden die Arsenatome vermutlich einen noch größeren Abstand voneinander einnehmen. Eine weitere Auswärtsrelaxation des As-Atoms an der Kante zerstört auch die Tetraederkonfiguration des darunterliegenden Si-Atoms (Atom 45 bzw. 37) und würde damit zu einer Energieerhöhung führen. Aufgrund der Geometrie ist der Winkel der As-As Bindungen an den Kantenatomen (69, 71) und den dahinterliegenden Arsenatomen (73, 79) gleich und beträgt 99.97° . Diese As-terminierte Stufenkante ist die stabilste $(11\bar{2})$ -Stufenkante, siehe Kap. 4.4.

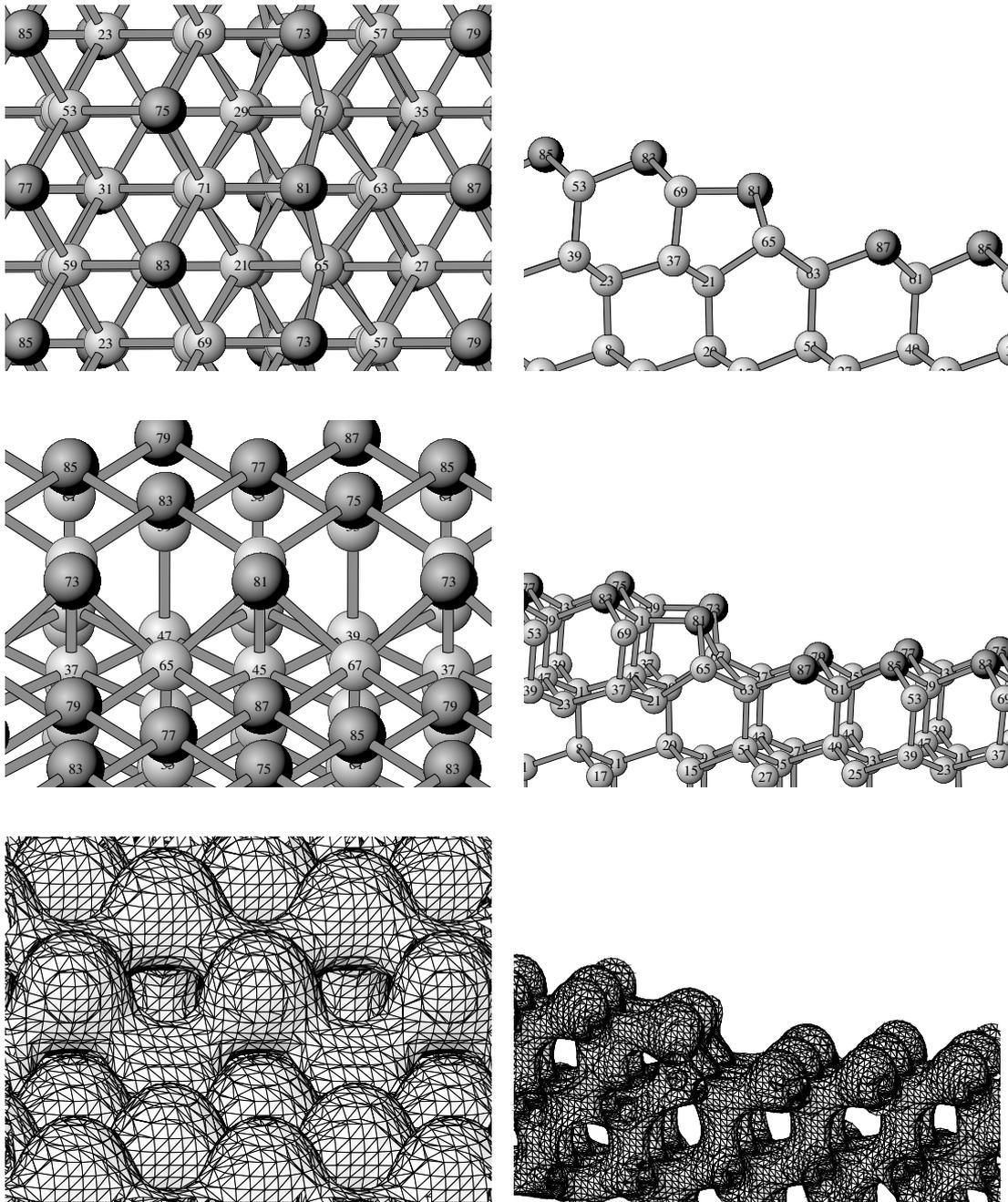


Abbildung 4.13: Seitenansicht und Aufsicht der $(11\bar{2})$ -Stufe mit überhängendem Arsen an der Kante.

4.3.3 Die $(11\bar{2})$ Stufenkante mit überhängendem Arsen

Wie in Abbildung 4.13 zu sehen ist, wird bei dieser Konfiguration die obere Terrasse, entgegen den vorhergehenden Strukturen, nach einer Atomreihe in der ersten Ebene abgeschnitten. In der unteren Terrasse wird die Reihe der vor den

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge			
83(75)	As	4.23	0.01	1.66	4.54	83-59	83-71	83-69	
-1, 1, 1						104.85	104.10	104.10	
81(73)	As	7.06	-0.01	-18.42	19.73	81-71	81-65	81-67	
2, 2, 1						103.80	109.97	110.00	
71(69)	Si	2.94	0.01	-4.30	5.21	71-83	71-75	71-81	71-45
0, 2, 2						104.10	104.10	103.80	97.70
45(37)	Si	-1.31	0.00	-3.10	3.36	45-71	45-31	45-21	45-29
0, 2, 3						97.70	97.85	99.47	99.47
65(67)	Si	-5.08	-0.00	8.50	9.90	65-81	65-73	65-21	65-63
3, 1, 3						109.97	110.00	101.04	112.07
65(67)	Si	-5.08	-0.00	8.50	9.90	65-57			
3, 1, 3						112.02			
Atom		Winkel							
83(75)	As	59-83-71		59-83-69		71-83-69			
-1, 1, 1		103.04	103.05	103.32					
81(73)	As	71-81-65		71-81-67		65-81-67			
2, 2, 1		99.77	99.73	95.87					
71(69)	Si	83-71-75		83-71-81		83-71-45		75-71-81	75-71-45
0, 2, 2		103.32	114.88	114.06	114.92	114.07	81-71-45	96.16	
45(37)	Si	71-45-31		71-45-21		71-45-29		31-45-21	31-45-29
0, 2, 3		112.27	104.13	104.12	112.67	112.66	21-45-29	110.34	
65(67)	Si	81-65-73		81-65-21		81-65-63		81-65-57	73-65-21
3, 1, 3		95.87	102.38	79.31	153.69	102.39	73-65-63	153.70	
65(67)	Si	73-65-57		21-65-63		21-65-57		63-65-57	
3, 1, 3		79.31	103.89	103.91	93.56				

Tabelle 4.5: Koordinaten der $(11\bar{2})$ -Stufe mit überhängendem As.

Kantenatomen liegende Siliziumatome nicht durch Arsenatome ersetzt, obwohl die darüberliegenden Siliziumatome entfernt wurden. Die freien Bindungen dieser Siliziumatome stehen den Arsenatomen an der Kante daher zur Verfügung. Diese Konfiguration läßt sich auch so auffassen, daß in einer der vorher diskutierten Strukturen die Kantenatome entfernt wurden.

Aufgrund der fehlenden Atomreihe müssen die Kantenatome sich sehr weit nach vorne bewegen, um mit den Siliziumatomen der unteren Terrasse eine Bindung einzugehen. Dabei 'ziehen' sie so stark an den hinter ihnen gelegenen Atomen der oberen Terrasse, daß sich diese Verzerrung deutlich sichtbar in die Terrasse hinein fortsetzt. Dieser Effekt wird sogar noch verstärkt, da die Kantenatome sich zudem nach unten auf die Si-Atome der 3. Lage zubewegen, bis sie sogar unterhalb der 2. Atomlage sind, wie der Tabelle 4.5 zu entnehmen ist. Auch um

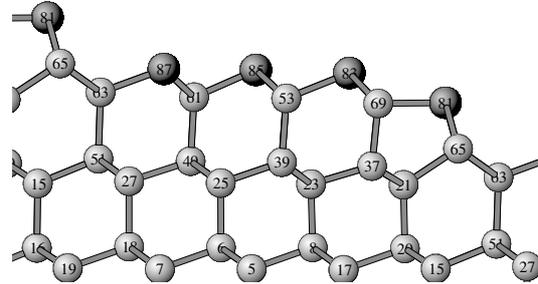


Abbildung 4.14: Seitenansicht einer ganzen Terrasse der $(11\bar{2})$ -Stufenkante mit überhängendem Arsenatom an der Kante. Wie deutlich zu sehen ist, setzt sich die strukturelle Verzerrung über die gesamte Oberfläche fort.

den Bindungswinkel zwischen dem Kantenatom und dem unter ihm gelegenen Siliziumatom (81-71-45) nicht zu klein werden zu lassen, bewegt sich das hinter dem Kantenatom gelegene Si-Atom nach vorne. Von dieser Verzerrung ist daher die gesamte Terrassenbreite betroffen, wie in Abb. 4.14 zu beobachten ist. Es ist aufgrund dessen nicht auszuschließen, daß sich in dieser Rechnung die Stufen gegenseitig beeinflussen, da unklar ist, inwieweit die Verschiebung der Atome 57 und 63 auf die Verzerrung der Terrasse zurückzuführen ist, also auf die über die Bindungen 63-87 bzw. 57-79 vermittelten Kräfte, und wie stark sich das wiederum auf die Atome 65 bzw. 67 und somit auf die nächste Stufe auswirkt.

Außerdem führen die fehlenden Atome auch zu einer veränderten Topologie: die Kantenatome benötigen je zwei Bindungspartner, die Siliziumatome der unteren Terrasse (65, 67) bieten jedoch nur einem der beiden nächstgelegenen Kantenatome (73, 81) eine freie Bindung an. In der Endkonfiguration sind die Siliziumatome 65 und 67 schließlich fünffach koordiniert. Atom 65 hat Bindungen zu den Kantenatomen 73 und 81 sowie zu den Atomen 21, 57 und 63 aus der vierten Ebene. Anstatt zu nur einem, baut es also zu zwei Kantenatomen eine Bindung auf. Die resultierende fünffache Koordination dieser Atome sowie die Geometrie ihrer Anordnung, die am ehesten einer vierseitigen Pyramide entspricht, ist nur unter Zuhilfenahme ihrer unbesetzten d-Orbitale zu erklären. Dies erklärt auch die vergleichsweise großen Bindungslängen dieser Atome, die zu den Kantenatomen 110% und zu den Siliziumatomen 57 und 63 sogar 112% der Bindungslänge im Bulk betragen. Die Anteile der energetisch ungünstigen d-Orbitale heben allerdings auch die Energie der elektronischen Wellenfunktionen, an denen sie beteiligt sind, an. Dies läßt sich z.B. anhand der gefundenen Bandlücke beobachten, die für diese Rechnung sehr klein war.

Aufgrund der ungünstig koordinierten Siliziumatome ist es denkbar, daß hier nur ein lokales Minimum gefunden wurde, und, ausgehend von einer anderen Ausgangsposition, eine energetisch günstigere Struktur gefunden werden kann.

4.3.4 Die $(\bar{1}\bar{1}2)$ Stufenkante mit Siliziumatom an der Kante

Die $(\bar{1}\bar{1}2)$ -Kanten schließen mit den $(11\bar{2})$ -Kanten einen Winkel von 120° ein, bzw. liegen ihnen gegenüber. Wie zuvor wurden auch hier drei Fälle betrachtet: Abschneiden nach einer Reihe Siliziumatome in der zweiten Ebene, Ersetzen der Kantenatome durch Arsenatome und schließlich Abschneiden nach einer Reihe Arsenatome aus der ersten Ebene.

In Abbildung 4.15 ist die relaxierte Struktur zu sehen. Einige der zugehörigen Koordinaten und Bindungswinkel sind in Tabelle 4.6 enthalten. Vollständige Tabellen der Koordinaten, Bindungswinkel und Nächstnachbarabstände finden sich in Anhang A. In diesem Fall haben die Atome der zweiten Ebene im Gegensatz zur $(11\bar{2})$ -Kante nur eine Bindung zu hinter ihnen liegenden Atomen und zwei zu Atomen der vor ihnen liegenden Reihe in der obersten Ebene. Wird die Ebene also nach einer Reihe Siliziumatome abgeschnitten, ist es dazu nötig, zwei Bindungen pro Kantenatom zu trennen, im Gegensatz zu nur einer im Fall der $(11\bar{2})$ -Kante. Ein Kantenatom ist also in der Ausgangslage zunächst nur zweifach koordiniert. Ein Siliziumatom in dieser Position benötigt also zwei zusätzliche Elektronen, z.B. durch kovalente Bindungen zu Nachbaratomen, um Edelgaskonfiguration zu erreichen. In der Ausgangsposition sind die neben ihm liegenden Kantenatome zu weit entfernt, um Bindungen zu bilden. Allerdings sind in dieser Position nur zwei Bindungen, die zu den hinter ihnen liegenden Arsenatomen der oberen Terrasse und die zu den unter ihnen liegenden Siliziumatomen, vorgegeben, was besonders in y -Richtung eine größere Flexibilität gestattet.

Diese Flexibilität nutzen die Siliziumatome, indem sich je zwei Kantenatome aufeinander zu bewegen und dimerisieren. Somit ist jedes dieser Atome dreifach koordiniert, allerdings bleibt eine Bindung frei. Es läßt sich weiter beobachten, daß diese Atome keine symmetrischen Positionen einnehmen. Während einige Kantenatome (67) bei dieser Dimerisierung die Bindungslängen zu den hinter ihnen liegenden Arsenatomen (79) und den unter ihnen liegenden Siliziumatomen (63) im wesentlichen beibehalten und somit um die durch diese Atome verlaufende Achse kippen, bewegt sich jedes zweite der Kantenatome (65) zusätzlich weiter in die Ebene seiner Bindungspartner hinein. Es läßt sich hier ein ähnlicher Effekt wie schon bei den Siliziumatomen an der $(11\bar{2})$ -Kante beobachten: Jedes zweite Si-Atom bildet sp^2 -hybridisierte Orbitale aus. Das verbleibende p-Orbital wird nicht besetzt, das entsprechende Elektron wandert stattdessen zum benachbarten Kantenatom, das so Edelgaskonfiguration erreichen kann. Auch hier läßt sich wieder beobachten, daß das sp^2 -hybridisierte Siliziumatom kürzere Bindungen zu seinen Nachbarn ausbildet und nahezu in einer Ebene mit seinen Bindungspartnern liegt, die etwa im 120° -Winkel zueinander angeordnet sind.

Durch die Dimerisierung und die damit verbundene Verschiebung der Kantenatome werden die Bindungswinkel an den benachbarten Atomen verzerrt, so daß diese keine Tetraederkonfiguration mehr einnehmen können. Bei den zusätzlich

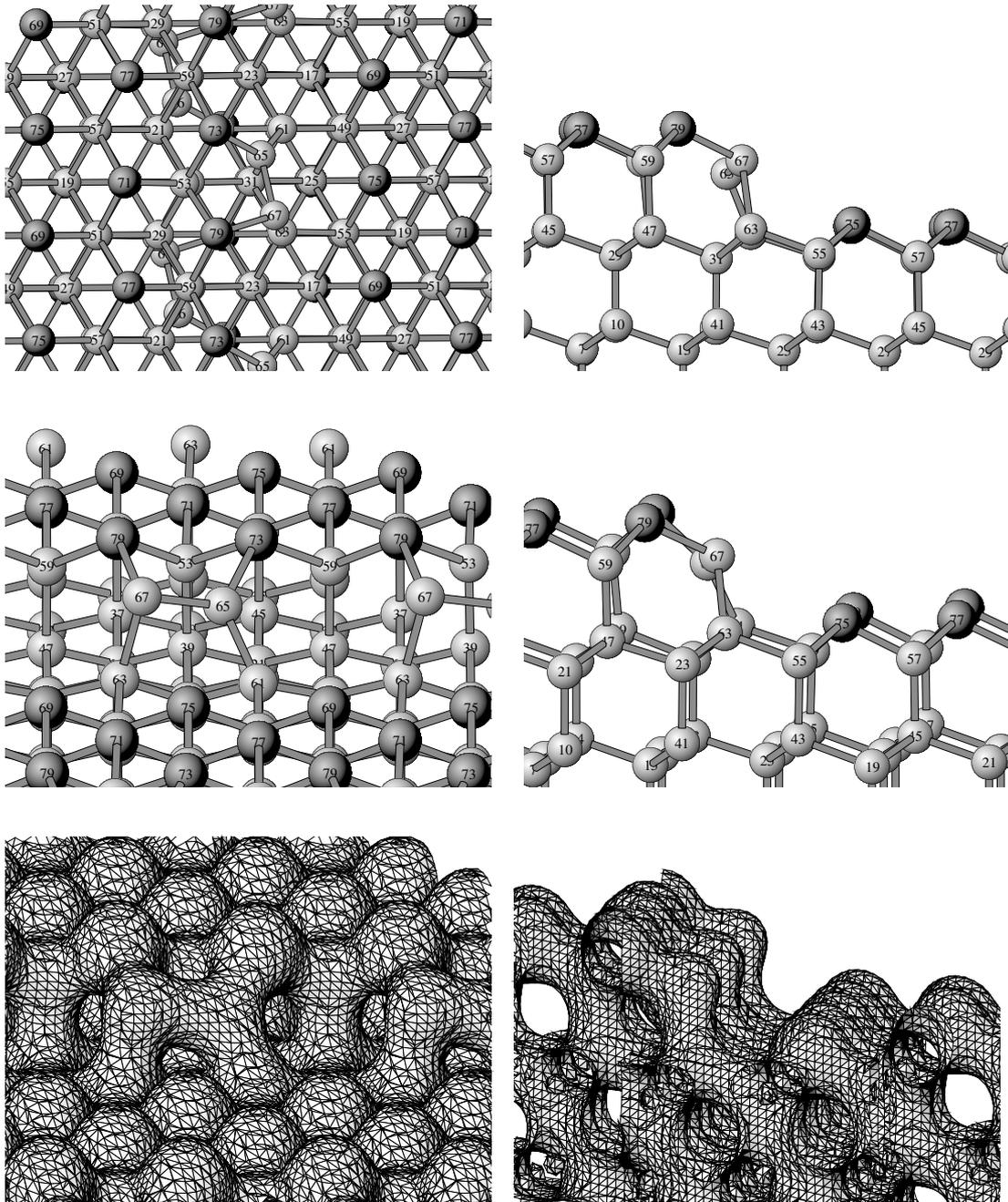


Abbildung 4.15: Seitenansicht und Aufsicht der $(\bar{1}\bar{1}2)$ -Stufenkante mit Siliziumatom an der Kante, analog zu Abb. 4.11.

einwärts verschobenen Kantenatomen (65) bzw. ihren Nachbarn (73) und (61) tritt eine besonders starke Verzerrung aus der Tetraederkonfiguration auf, die an den Bindungswinkeln in Tabelle 4.6 abzulesen ist. Aufgrund der Verzerrung der Bindungswinkel und da nur jedes zweite Kantenatom Edelgaskonfiguration erreicht ist diese Stufenkante energetisch ungünstig.

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge			
79	As	-2.99	0.21	4.78	5.64	79-67	79-59	79-53	
-2, 2, 1						103.55	103.39	102.25	
73	As	-3.49	-1.24	3.95	5.42	73-65	73-59	73-53	
-2, 0, 1						99.90	104.17	103.57	
67	Si	-5.09	12.20	0.05	13.22	67-79	67-65	67-63	
0, -2, 2						103.55	97.05	101.02	
65	Si	-13.87	-18.07	-7.32	23.93	65-73	65-67	65-61	
0, 0, 2						99.90	97.05	99.23	
63	Si	0.24	1.12	1.38	1.79	63-67	63-55	63-31	63-23
0, -2, 3						101.02	101.63	100.11	100.94
61	Si	0.72	-0.44	-0.38	0.93	61-65	61-49	61-31	61-23
0, 0, 3						99.23	100.59	101.31	100.46

Atom		Winkel						
79	As	67-79-59	67-79-53	59-79-53				
-2, 2, 1		112.24	91.40	102.55				
73	As	65-73-59	65-73-53	59-73-53				
-2, 0, 1		110.15	77.05	106.49				
67	Si	79-67-65	79-67-63	65-67-63				
0, -2, 2		98.73	116.57	95.80				
65	Si	73-65-67	73-65-61	67-65-61				
0, 0, 2		114.66	125.61	118.19				
63	Si	67-63-55	67-63-31	67-63-23	55-63-31	55-63-23	31-63-23	
0, -2, 3		116.66	97.42	118.28	109.17	107.05	107.24	
61	Si	65-61-49	65-61-31	65-61-23	49-61-31	49-61-23	31-61-23	
0, 0, 3		126.07	79.06	115.58	110.15	110.58	109.73	

Tabelle 4.6: Koordinaten und Bindungswinkel der ($\bar{1}\bar{1}2$)-Stufe mit exponiertem Si.

4.3.5 Die ($\bar{1}\bar{1}2$) Stufenkante mit ersetzttem Arsenatom

Wie schon bei der ($11\bar{2}$)-Kante ist es auch für die ($\bar{1}\bar{1}2$)-Kante günstig, die Kantenatome durch Arsenatome zu ersetzen. Diese benötigen nur drei nächste Nachbarn. Dies wird, wie schon bei den Siliziumatomen, durch Dimerisierung der Kantenatome erreicht (Abb. 4.16). Die errechnete Endkonfiguration ist im Rahmen der Rechengenauigkeit bezüglich der Ebenen zwischen den Kantenatomen symmetrisch. Im Gegensatz zu den Siliziumatomen erreichen die Arsenatome an der Kante (65 und 67) bereits durch die Dimerisierung und die so erzielte dreifache Koordination Edelgaskonfiguration. Die Arsen-Arsen Bindungen sind dabei, wie zu erwarten, deutlich länger als As-Si Bindungen (siehe Tabelle 4.7). Zu den hinter ihnen gelegenen Arsenatomen der oberen Terrasse (73 bzw. 79)

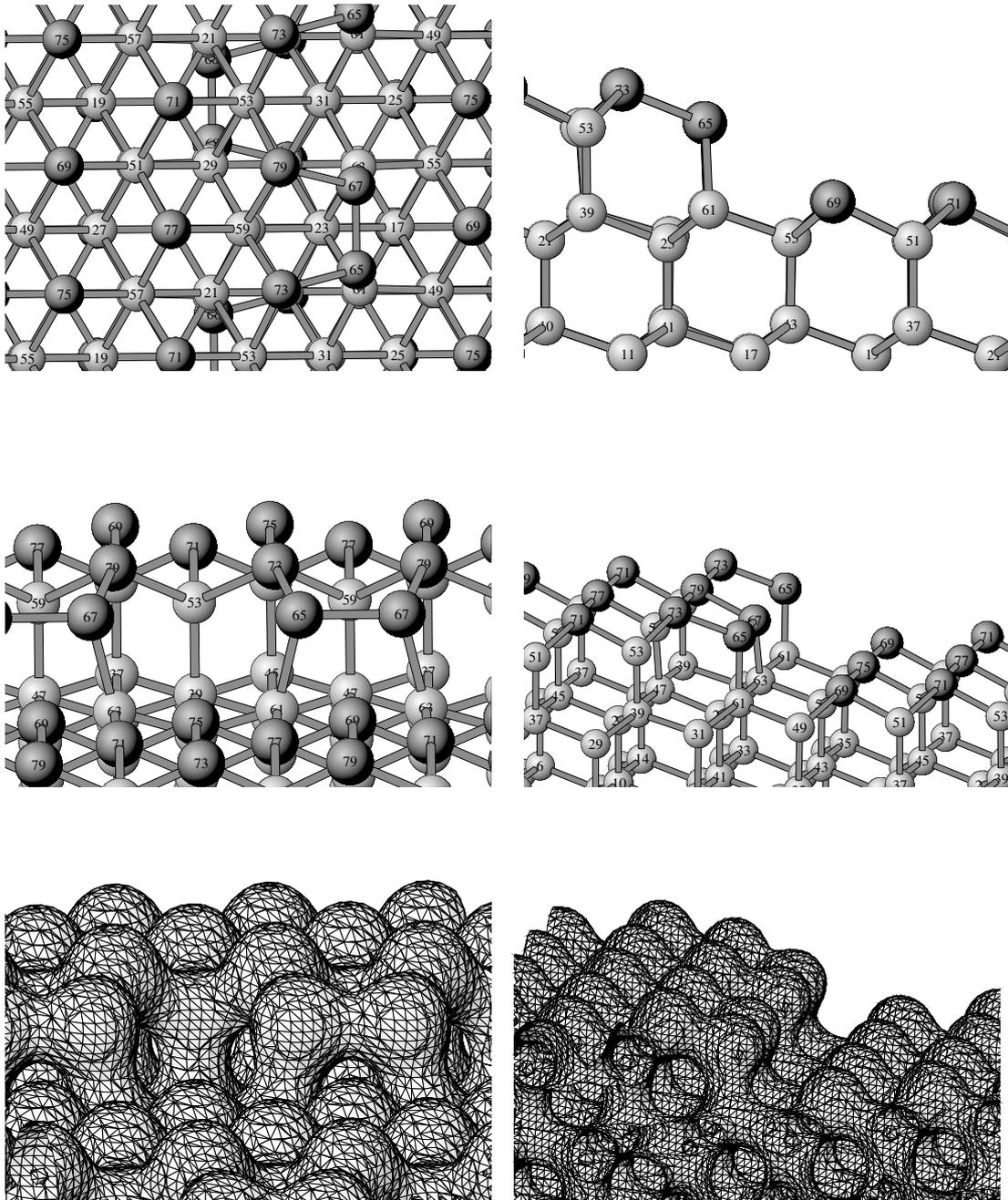


Abbildung 4.16: Seitenansicht und Aufsicht der $(\bar{1}\bar{1}2)$ -Stufenkante mit Arsenatom an der Kante, analog zu Abb. 4.11.

beträgt die Bindungslänge 106% des Nächstnachbarabstandes im Bulk, die Bindungslänge zwischen zwei Kantenatomen beträgt sogar 107.5%. Die Bindung der Kantenatome untereinander bildet einen Winkel von 103.3° mit der Bindung zum Arsenatom der oberen Terrasse bzw. 103.6° mit der zum darunterliegenden Siliziumatom. Der verbleibende Winkel ist 109.6° . Die Bindungspartner befinden

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge			
79						79-67	79-59	79-53	
-2, 2, 1	As	-2.66	-1.56	3.12	4.39	106.05	101.78	104.72	
73						73-65	73-59	73-53	
-2, 0, 1	As	-2.57	1.54	3.11	4.32	106.05	101.74	104.67	
67						67-79	67-65	67-63	
0, 2, 2	As	-2.00	-12.04	0.89	12.24	106.05	107.53	103.20	
65						65-73	65-67	65-61	
0, 0, 2	As	-1.92	12.11	0.90	12.29	106.05	107.53	103.22	
63						63-67	63-55	63-31	63-23
0, 2, 3	Si	-0.27	-1.56	0.78	1.77	103.20	101.82	101.22	99.21
61						61-65	61-49	61-31	61-23
0, 0, 3	Si	-0.04	1.58	0.80	1.77	103.22	102.11	101.33	99.30
Atom		Winkel							
79		67-79-59	67-79-53	59-79-53					
-2, 2, 1	As	95.59	113.28	104.76					
73		65-73-59	65-73-53	59-73-53					
-2, 0, 1	As	95.65	113.50	104.55					
67		79-67-65	79-67-63	65-67-63					
0, 2, 2	As	103.27	109.51	103.58					
65		73-65-67	73-65-61	67-65-61					
0, 0, 2	As	103.23	109.62	103.61					
63		67-63-55	67-63-31	67-63-23	55-63-31	55-63-23	31-63-23		
0, 2, 3	Si	111.21	118.85	100.42	106.67	109.92	109.52		
61		65-61-49	65-61-31	65-61-23	49-61-31	49-61-23	31-61-23		
0, 0, 3	Si	111.28	118.74	100.31	107.03	110.28	108.95		

Tabelle 4.7: Koordinaten und Bindungswinkel der $(\bar{1}\bar{1}2)$ -Stufe mit As an der Kante.

sich also annähernd in einer Tetraederkonfiguration. Allerdings führt die Dimerisierung und das entsprechende ‘Kippen’ der Kantenatome zueinander bei deren Bindungspartnern zu stärkeren Abweichungen von einer tetraedrischen Anordnung der Bindungspartner, wie sich Tabelle 4.7 entnehmen läßt.

4.3.6 Die $(\bar{1}\bar{1}2)$ Stufenkante mit überhängendem Arsenatom

Im Falle des überhängenden As-Atoms bei der $(\bar{1}\bar{1}2)$ -Kante (Abb. 4.17 bzw. Tabelle 4.8) hat dieses zwei Bindungen zu Si-Atomen der Reihe hinter den Kantenatomen. Durch das Abschneiden der oberen Terrasse nach der As-Reihe wird also, im Gegensatz zum überhängenden As-Atom an der $(11\bar{2})$ -Kante, nur eine Bin-

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge			
75(69)	As	13.18	-0.00	-12.66	18.28	75-63	75-59	75-53	
-2, 2, 1						111.67	105.24	105.22	
67(73)	As	6.33	-0.01	2.27	6.73	67-57	67-53	67-51	
-5, -1, 1						104.82	107.13	104.81	
63(61)	Si	-3.45	0.00	3.21	4.71	63-75	63-55	63-31	63-23
0, 2, 3						111.67	109.27	100.23	100.23
55(49)	Si	1.07	0.01	4.11	4.25	55-71	55-65	55-63	55-43
2, 2, 4						104.66	104.67	109.27	103.08
53(59)	Si	5.93	-0.01	-6.04	8.47	53-75	53-69	53-67	53-39
-3, -1, 2						105.22	105.23	107.13	97.12
39(47)	Si	0.15	-0.01	-4.40	4.40	39-53	39-31	39-29	39-21
-3, -1, 3						97.12	99.18	98.96	98.96
31(23)	Si	0.68	0.00	-1.55	1.69	31-63	31-61	31-39	31-33
-1, -1, 4						100.23	100.23	99.18	98.85
Atom		Winkel							
75(69)	As	63-75-59	63-75-53	59-75-53					
-2, 2, 1		98.94	98.92	101.78					
67(73)	As	57-67-53	57-67-51	53-67-51					
-5, -1, 1		104.14	102.33	104.14					
63(61)	Si	75-63-55	75-63-31	75-63-23	55-63-31	55-63-23	31-63-23		
0, 2, 3		136.64	101.26	101.25	103.48	103.49	109.10		
55(49)	Si	71-55-65	71-55-63	71-55-43	65-55-63	65-55-43	63-55-43		
2, 2, 4		102.55	113.10	111.74	113.10	111.70	104.89		
53(59)	Si	75-53-69	75-53-67	75-53-39	69-53-67	69-53-39	67-53-39		
-3, -1, 2		101.78	116.42	104.68	116.42	104.68	111.46		
39(47)	Si	53-39-31	53-39-29	53-39-21	31-39-29	31-39-21	29-39-21		
-3, -1, 3		97.75	110.34	110.35	113.25	113.22	111.19		
31(23)	Si	63-31-61	63-31-39	63-31-33	61-31-39	61-31-33	39-31-33		
-1, -1, 4		109.10	104.00	116.06	104.02	116.06	106.08		

Tabelle 4.8: Koordinaten und Bindungswinkel der ($\bar{1}\bar{1}2$)-Stufe mit As an der Kante.

allerdings sehr weit nach vorn bewegen, was auch hier wieder zu Verzerrungen über die gesamte Terrassenlänge führt. Daher gilt auch hier, daß eine Beeinflussung der Stufen untereinander vermutet werden muß, deren quantitative Auswirkungen, z.B. auf die Bindungslänge zwischen den Si-Atomen 61 und 49, schwer abzuschätzen sind. Auch hier zeigen sich die Auswirkungen der ‘fehlenden’ Reihe von Si-Atomen wieder durch vergleichsweise lange Bindungen an den Kantenatomen, wie sich Tabelle 4.8 entnehmen läßt. Besonders die Bindung zum Si-Atom der dritten Gitterebene ist mit 111.7% besonders lang.

4.4 Energiebetrachtungen

Um zu vergleichen, welche der Stufenkanten energetisch am günstigsten ist und wie günstig eine glatte Oberfläche gegenüber einer mit Inseln ist, werden die Energien der berechneten Systeme miteinander verglichen. Dabei ist wichtig, daß die Systeme auch wirklich vergleichbar sind. Damit die Berechnungen nicht durch Oberflächenenergien verfälscht werden, können nur Zellen mit gleicher (auf die (111)-Ebene projizierter) Oberfläche berücksichtigt werden. Zudem müssen vergleichbare Zellen auch vergleichbare Anzahlen der jeweiligen Atomsorten enthalten. Dies ist bei den berechneten Zellen nicht der Fall, daher werden diese Unterschiede durch Referenzenergien ausgeglichen. So wird, um den Einfluss unterschiedlich großer Oberflächen zu kompensieren, die Oberflächenenergie der Differenz beider Oberflächen mitberücksichtigt. Dazu wird die Energie einer zuvor berechneten idealen Arsen-terminierten Si(111)-Oberfläche benutzt.

Um die Auswirkung unterschiedlicher Anzahlen von Atomen zu berücksichtigen, müssen auch die Energien entsprechender Referenzsysteme einbezogen werden. Dazu ist zunächst zu bedenken, mit welchem konkreten System verglichen werden muß, woher also ein zusätzliches Silizium- oder Arsenatom kommt bzw. wohin ein überschüssiges Atom abwandert. Dabei wird von einer Oberfläche ausgegangen, die bis auf die von uns betrachtete Struktur eine ideale Arsen-bedeckte Si(111)-Oberfläche ist. Die Energie einer beliebig strukturierten Oberfläche ist dann die Summe der Energien der einzelnen Strukturen, solange angenommen werden kann, daß sich die Strukturen nicht gegenseitig beeinflussen.

Damit der Austausch der Atome mit der Atmosphäre über der Oberfläche korrekt berücksichtigt wird, müssen auch über diese Annahmen gemacht werden. Um den experimentellen Bedingungen möglichst gerecht zu werden, wird dazu angenommen, daß die Arsenatome auf der Oberfläche mit As_4 -Molekülen in der Atmosphäre im Gleichgewicht stehen. Somit erhält man als Referenzenergie für Arsen die Energie eines Arsenatoms des As_4 -Moleküls. Dabei ist allerdings zu berücksichtigen, daß sich je nach den experimentellen Bedingungen das chemische Potential für Arsen verschieben läßt. Um die As_4 -gesättigte Atmosphäre zu erzeugen, wird festes Arsen abgedampft. Stünde die gesamte Atmosphäre mit festem Arsen im Gleichgewicht, so wäre auch das Arsen auf der Siliziumoberfläche über die Atmosphäre mit festem Arsen im Gleichgewicht. In diesem Fall wäre festes Arsen zur Berechnung der Referenzenergie heranzuziehen. Die geeignete

As_4 :	-12.579728	Ry/Atom
As-bulk:	-12.615445	Ry/Atom
Si-bulk:	-7.931134	Ry/Atom
As-terminiertes Si(111):	-0.032312	Ry pro lateraler Einheitszelle

Tabelle 4.9: Referenzenergien bei $\mathbf{G}_{max} = 3.0$.

	As ersetzt	Si an der Kante	As überhängend
Energie pro Zelle (Ry)	-759.737	-740.760	-772.345
Si-Atome	64	68	72
As-Atome	20	16	16
Fläche	$17\frac{1}{3}$ Oberflächeneinheitenzellen		
zusätzliches As	$2\frac{2}{3}$	$-1\frac{1}{3}$	$-1\frac{1}{3}$
Referenzenergie As ₄ : -12.579728 Ry/Atom			
Energiedifferenz in mRy	10.305	392.434	531.984
Energie pro Kantenlänge	2.576	98.108	132.996
Referenzenergie As-bulk: -12.615445 Ry/Atom			
Energiedifferenz in mRy	105.550	344.811	484.361
Energie pro Kantenlänge	26.388	86.203	121.090

Tabelle 4.10: Energien der $(11\bar{2})$ -Stufenkanten

Referenzenergie und somit das chemische Potential für Arsenatome hängt daher von den experimentellen Bedingungen ab. Es ist über den Temperaturunterschied zwischen dem Arsenmaterial und der betrachteten Siliziumoberfläche einstellbar und wird sich zwischen den Referenzenergien für As₄ und Arsen-Bulkmaterial bewegen.

Hingegen enthält die Atmosphäre nur vereinzelte Siliziumatome, die bei Auftreffen auf der Oberfläche nach kurzem Diffusionsweg eingebaut werden. Da sich beim Abdampfen des Siliziums die Größe der Oberflächen und deren Struktur im Mittel nicht verändern sollen, stammen die zusätzlichen Siliziumatome also aus dem Bulk. Als Referenzenergie für Silizium wird daher die Energie von Siliziumbulkatomen genommen.

	As ersetzt	Si an der Kante	As überhängend
Energie pro Zelle (Ry)	-709.265	-690.295	-658.565
Silizium-Atome	64	68	64
Arsen-Atome	16	12	12
Fläche	$14\frac{2}{3}$ Oberflächeneinheitenzellen		
zusätzliches As	$1\frac{1}{3}$	$-2\frac{2}{3}$	$-2\frac{2}{3}$
Referenzenergie As ₄ : -12.579728 Ry/Atom			
Energiedifferenz in mRy	77.100	452.777	457.852
Energie pro Kantenlänge	19.275	113.194	114.463
Referenzenergie As-bulk: -12.615445 Ry/Atom			
Energiedifferenz in mRy	124.723	357.532	362.607
Energie pro Kantenlänge	31.181	89.383	90.652

Tabelle 4.11: Energien der $(\bar{1}\bar{1}2)$ -Stufenkanten

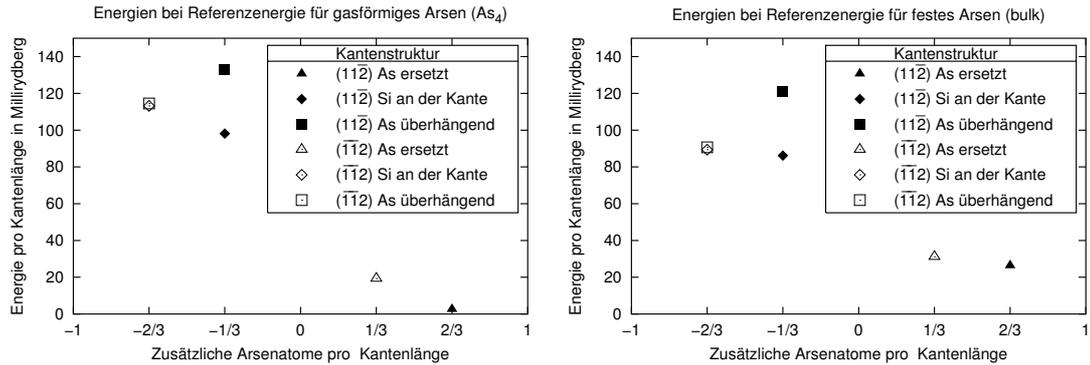


Abbildung 4.18: Energie aufgetragen gegenüber zusätzlich benötigtem Arsen pro Kantenlänge. Als Einheit für die Kantenlänge wurde die Anzahl der Kantenatome genommen. Im linken Diagramm wurde gasförmiges Arsen (As_4) als Referenz benutzt, für das rechte festes (bulk) Arsen.

Um die Auswirkungen verschiedener Referenzenergien für Arsen zu ermessen, wurden die Ergebnisse für die verschiedenen Referenzenergien in Abbildung 4.18 graphisch gegenübergestellt. In den Tabellen wurden die Ergebnisse für beide Fälle angegeben. Es ist zu beobachten, daß sich die Energien zwar verschieben, allerdings nicht soweit, daß sich die Reihenfolge der Energien verändern würde.

Wie zu erwarten, ist bei der $(11\bar{2})$ -Stufenkante die Konfiguration, bei der die Siliziumatome an der Kante durch Arsen ersetzt werden, am günstigsten. Wie anhand der Struktur im Grundzustand zu sehen ist, haben hier schon in der Ausgangsposition alle Atome die jeweils bevorzugte Anzahl nächster Nachbarn. Zudem sind die Verschiebungen der Atome aus ihren Ausgangslagen nur sehr klein. Am ungünstigsten ist bei der $(11\bar{2})$ -Kante offensichtlich die Anordnung mit überhängenden Arsenatomen, was vermutlich auf die ungünstige elektronische Konfiguration der fünffach koordinierten Siliziumatome zurückzuführen ist.

Auch bei der $(\bar{1}\bar{1}2)$ -Kante ist wieder die Struktur mit Arsenatomen auf Siliziumplätzen an der Kante am günstigsten. In diesem Fall erreichen die Kantenatome ideale Koordination, indem sie an der Kante Dimere bilden, so daß je zwei Kantenatome eine Bindung zueinander ausbilden. Allerdings sind dazu größere Verzerrungen nötig als bei der $(11\bar{2})$ -Stufenkante, was erklärt, warum die Stufenkante mit ersetzten Siliziumatomen in $(11\bar{2})$ -Richtung energetisch günstiger ist als die in $(\bar{1}\bar{1}2)$ -Richtung. Es fällt auf, daß die energetisch günstigsten Stufenkanten auch das meiste zusätzliche Arsen benötigen, da hier die Siliziumatome an der Kante durch Arsenatome ersetzt werden, die dort leichter ihre ideale dreifache Koordination erreichen.

4.4.1 Vereinfachte Wulff-Konstruktion

Auch wenn die $(11\bar{2})$ -Kante energetisch günstiger ist als die $(\bar{1}\bar{1}2)$ -Kante, werden die energetisch günstigsten Inseln nicht unbedingt nur von $(11\bar{2})$ -Kanten begrenzt.

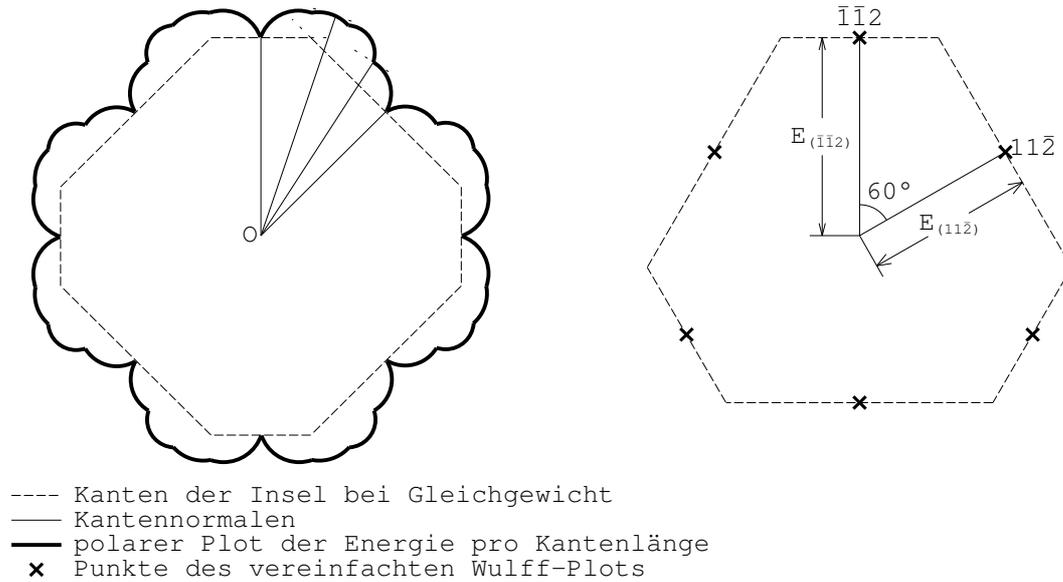


Abbildung 4.19: Links ist ein zweidimensionaler Wulffplot dargestellt. Die durchgezogene Kurve entsteht durch einen Polarplot der Energie pro Kantenlänge gegen den Normalenvektor der Kante in der Ebene. Die innersten der zugehörigen Kanten begrenzen die stabile Form einer Insel. Rechts ist der stark vereinfachte Wulff-Plot dargestellt. Die polar aufgetragene Energiekurve beschränkt sich auf sechs Punkte, die sich aus den jeweils günstigsten Energien für die $(11\bar{2})$ und die $(\bar{1}\bar{1}2)$ -Kante ergeben, hier mit Bulk-Arsen als Referenz.

Untersucht man, welche Inselform bei einer vorgegebenen Anzahl von Siliziumatomen am günstigsten ist, so ist die Gesamtenergie ausschlaggebend, die für alle Kanten der Insel aufzuwenden ist. Dabei ist die Fläche der Insel durch die Zahl der Siliziumatome in den zweiten und dritten Lagen der Insel vorgegeben. Damit ergibt sich das Optimierungsproblem, bei vorgegebener Flächengröße A die für Kanten aufgewandte Energie E_K zu minimieren bzw. bei vorgegebener Energie der Kanten die Fläche der Insel zu maximieren. Bezeichnet man die Kantenstücke mit S_i , und die Energie pro Kantenlänge mit k_i , so ist der Beitrag der Kanten zur Energie:

$$E_K = \sum_i k_i S_i \quad (4.1)$$

Zur Gesamtenergie der Insel trägt natürlich auch die Bindungsenergie der Atome bei, da aber die Anzahl der Atome durch die Flächengröße vorgegeben sei, bleibt dieser Beitrag konstant. Zur Minimierung der Energie wird die Insel daher eine optimale Form anstreben, die durch Minimierung der Energie über alle Kanten nach (4.1) gegeben ist. Für den Fall von Kristallen im dreidimensionalen Raum⁵ wurde das schon 1878 von J.W. Gibbs [Gib77, Gib92] und 1885 von

⁵Im dreidimensionalen Fall treten Oberflächen, Oberflächenspannungen und das Volumen des Kristalls an die Stelle von Kanten, Kantenenergien und Oberfläche der Insel.

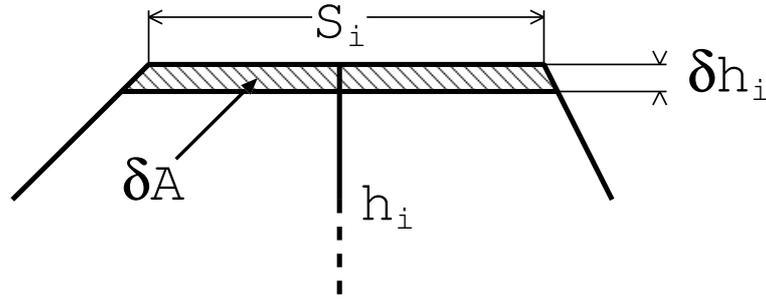


Abbildung 4.20: Ändert sich h_i um δh_i so ändert sich die Fläche A der Insel um $\delta A = S_i \cdot \delta h_i$.

P. Curie [Cur85] abgeleitet. Die geometrische Form einer solchen Insel läßt sich konstruieren, indem man die von G.Wulff 1901 [Wul01] beschriebene Methode auf den zweidimensionalen Fall anwendet. In diesem Fall bedeutet das: es existiert innerhalb der Inselfläche ein Punkt, der Wulff'sche Punkt, so, daß die Abstände h_i der Kanten von diesem Punkt sich verhalten wie die k_i :

$$\frac{h_1}{k_1} = \frac{h_2}{k_2} = \dots = \frac{h_N}{k_N} \quad (4.2)$$

Dies wurde von Wulff für einige Spezialfälle und für den allgemeinen Fall von H. Liebmann 1914 [Lie14] bewiesen. M.v. Laue stellt in [vL43] einige Beweise für das Wulff'sche Theorem zusammen. Der Vollständigkeit halber möchte ich daraus den zweiten Beweis für den zweidimensionalen Fall anführen. Sei zunächst vorausgesetzt, daß sich die Kanten S_i für $i = 1 \dots N$ in der günstigsten Inselform tatsächlich ausbilden (die Konstruktion wird später erlauben festzustellen, welche Kanten tatsächlich berücksichtigt werden müssen). Ist die Ausrichtung der Kanten gegeben, dann läßt sich die Inselform durch die Abstände h_i der Kanten von einem vorgegebenen Punkt festlegen. Da A aus Dreiecken mit Grundlinie S_i und Höhe h_i , deren Spitzen sich im Wulff'schen Punkt treffen, zusammengesetzt ist, läßt sich die Fläche A bestimmen als:

$$A = \frac{1}{2} \sum_i h_i S_i, \quad (4.3)$$

wobei $S_i = S_i\{h_i\}$. Die Wahl der h_i ist dabei nicht völlig frei. Zum einen ist die Fläche A vorgegeben. Zum anderen sollen Freiheitsgrade aufgrund reiner Verschiebungen der Insel gegenüber dem gewählten Punkt nicht berücksichtigt werden. Diese weitere Einschränkung um zwei Freiheitsgrade geschieht durch zwei Gleichungen, mittels derer die h_i für zwei nichtparallele Kanten festgelegt werden:

$$\frac{h_1}{k_1} = \mu, \quad \frac{h_2}{k_2} = \mu \quad . \quad (4.4)$$

Die Konstante μ wird später festgelegt. Da die Flächenänderung dA bei einer Verschiebung dh_i einer Kante $S_i dh_i$ beträgt (Abb. 4.20), gilt:

$$\frac{\partial A}{\partial h_i} = S_i, \quad \frac{\partial^2 A}{\partial h_i \partial h_k} = \frac{\partial S_k}{\partial h_i} = \frac{\partial S_i}{\partial h_k}. \quad (4.5)$$

Aus (4.3) und wegen Abb. 4.20 folgt für die Flächenänderung dA :

$$\begin{aligned} \Rightarrow \quad dA &= \frac{1}{2} \sum_i (S_i dh_i + h_i dS_i) = \sum_k S_k dh_k \\ \Rightarrow \quad 2dA - dA &= \sum_i (S_i dh_i + h_i dS_i) - \sum_k S_k dh_k \\ \Rightarrow \quad dA &= \sum_i h_i dS_i. \end{aligned} \quad (4.6)$$

Bildet man dA/h_k in der ersten sowie der letzten Zeile von (4.6), erhält man⁶:

$$S_k = \sum_i h_i \frac{\partial S_i}{\partial h_k}. \quad (4.7)$$

Um nun die Energie E_k bei gegebener Fläche A zu minimieren wird der Lagrangemultiplikator λ eingeführt und man erhält:

$$0 = \frac{\partial}{\partial h_k} \left(\sum_i k_i S_i - \lambda A \right) = \sum_i k_i \frac{\partial S_i}{\partial h_k} - \lambda S_k \quad (4.8)$$

daraus wird einerseits nach Einsetzen von (4.7):

$$0 = \sum_i (k_i - \lambda h_i) \frac{\partial S_i}{\partial h_k}. \quad (4.9)$$

Andererseits läßt sich durch Multiplikation der Gleichungen (4.8) mit h_k und (4.6) λ bestimmen:

$$\sum_{i,k} k_i h_k \frac{\partial S_i}{\partial h_k} = \lambda \sum_k h_k S_k \quad (4.10)$$

$$\sum_i k_i S_i = \lambda 2A \quad (4.11)$$

$$\frac{E_{K,min}}{2A} = \lambda \quad (4.12)$$

wobei $E_{K,min}$ das Minimum für E_K sei. Damit ergibt (4.9) ein homogenes Gleichungssystem für die Größen $(k_i - \lambda h_i)$ mit N Gleichungen. Allerdings würde eine freie Wahl aller h_i bei konstanter Fläche A auch Translationen gegenüber

⁶Da $S\{\lambda h\} = \lambda S\{h\}$ findet man dies auch gemäß dem Eulerschen Satz über homogene Funktionen.

dem Wulff'schen Punkt mit einschließen. Mathematisch bedeutet das, daß die Matrix zu (4.9) höchstens vom Grad $N - 2$ ist. Daher werden zwei der h_i mittels (4.4) festgelegt, S_1 und S_2 dürfen dabei offensichtlich nicht parallel sein, da durch die vorgegebenen Abstände h_1, h_2 sonst kein Punkt gegeben wäre. Man legt nun den frei wählbaren Parameter μ in der Randbedingung (4.4) als $\mu = \lambda^{-1}$ fest und eliminiert die Freiheitsgrade der Translation so, daß die Terme $(k_1 - \lambda h_1)$ und $(k_2 - \lambda h_2)$ in (4.9) verschwinden. Es ist ersichtlich, daß die triviale Lösung von (4.9), bei der alle Terme $(k_i - \lambda h_i)$ verschwinden, stationär ist, es ist also:

$$k_i - \lambda h_i = 0 \quad \Leftrightarrow \quad k_i = \lambda h_i \quad \Leftrightarrow \quad h_i = \frac{1}{\lambda} k_i \quad (4.13)$$

Das entspricht genau der Wulff-Konstruktion. Allerdings ist so nicht gezeigt, daß diese auch stabil ist oder daß derart das globale Minimum gefunden wird. Leitet man wie Dinghas in [Din43] das Wulff-Theorem aus der Brunn-Minkowskischen Ungleichung ab, findet man, daß durch die Wulff-Konstruktion die energetisch günstigste Form gefunden wird. Wie zu erwarten ist die so bestimmte Form der Insel von deren Fläche unabhängig, sofern man die Diskretisierung durch das Kristallgitter sowie besondere Effekte an den Ecken außer acht läßt. Dies ist bei ausreichend großen Inseln, deren Kantenlängen groß gegenüber der Gitterkonstanten sind, gerechtfertigt.

Der Faktor $1/\lambda$ kann als Skalierungsfaktor betrachtet werden, durch den die Größe der Inseln nach (4.13) vorgegeben werden kann. Da die Strecken h_i proportional zu $1/\lambda$ sind, gilt aufgrund der Ähnlichkeit der Inseln, daß auch die Seitenlängen S_i und damit wegen (4.1) auch E_K proportional zu $1/\lambda$ sein müssen. Wegen (4.3) ist die Fläche A dann proportional zu $1/\lambda^2$. Daher ist E_K^2/A von der Größe einander ähnlicher Inseln unabhängig und hängt nur von der Form der Inseln ab, ist also für alle Inseln, die nach (4.13) konstruiert sind, gleich. Will man die günstigste Form einer Insel unabhängig von ihrer Größe bestimmen, muß man eine Insel mit minimalem E_K^2/A finden, wobei dies dann allein durch die k_i und die Kantennormalen bestimmt ist. Außerdem ist für alle Inseln mit einer solchen Form unabhängig von ihrer Fläche:

$$E_K \sim \sqrt{A} \quad \Leftrightarrow \quad \frac{E_K}{A} \sim \frac{1}{\sqrt{A}} \quad . \quad (4.14)$$

Die Kantenenergie pro Fläche nimmt also ab, je größer die Insel ist. Betrachtet man verschieden große Inseln auf einer Fläche, die durch Diffusion untereinander Atome austauschen können, so sind die größten Inseln die energetisch günstigsten. Kleine Inseln werden sich also auflösen und ihre Atome an größere Inseln abgeben. Dabei können Stufen, die Terrassen begrenzen, als unendlich große Insel gewertet werden: Werden dort Atome angelagert, so verschiebt sich die Stufe einfach, ohne dass sich die Kantenlänge verändert, es muß also keine zusätzliche Energie zur Bildung von Kanten aufgewendet werden. Betrachtet man Prozesse,

bei denen frei auf der Oberfläche diffundierende Atome sich an Inseln anlagern, so muß die Bindungsenergie der auf der Oberfläche diffundierenden Atome betrachtet werden. Deponiert man Atome aus der Gasphase, sind entsprechend Bindungsenergien von Molekülen in der Gasphase zu berücksichtigen. Diese sind mit der Bindungsenergie der eingebauten Atome zu vergleichen. Letztere ist die Energie von Bulkatomen, sobald einmal eine weitere Atomlage vervollständigt wurde, da in diesem Fall letztlich der Bulk um eine Atomlage gewachsen ist. Werden Atome zu Inseln hinzugefügt, so muß die Energie zur Schaffung zusätzlicher Kanten entsprechend mitgerechnet werden. Da alle anderen Energien und Energiedifferenzen für jedes Atom gleich und unabhängig von der Insel, an der ein Atom sich anlagert, sind, hängt die Bindungsenergie eines Atoms nur über E_K von dem Ort ab, an dem das Atom eingebaut wird, sobald es einmal eine Insel oder Terrassenkante erreicht hat. Also ist wegen (4.14) die Anlagerung an größere Inseln energetisch günstiger.

In den experimentellen Untersuchungen zur Form der Inseln auf As-bedecktem Si(111) [Voi] treten nur zwei Kantenorientierungen auf: $(11\bar{2})$ - bzw. $(\bar{1}\bar{1}2)$ -Kanten. Für diese Kanten haben wir die energetisch günstigsten Konfigurationen und deren Energien bestimmt. Wenn man diese Kanten mit jeweils drei Ausrichtungen berücksichtigt, kann man die Gleichgewichtsform der Inseln berechnen, wenn man davon ausgeht, daß die Energie für andere Kantenorientierungen so hoch ist, daß diese nicht auftreten. Die Annahme, daß nur $(11\bar{2})$ - bzw. $(\bar{1}\bar{1}2)$ -Kanten berücksichtigt werden müssen, scheint auch aus theoretischer Sicht plausibel, da Kantengeometrien für andere Facetten, bei denen sämtliche Kantenatome ihre Bindungen mit mindestens drei nächsten Nachbarn absättigen können, schwer vorstellbar sind. Außerdem ist die Kantenenergie k_i für alle Kanten eines Typs gleich, für die drei möglichen $(\bar{1}\bar{1}2)$ -Kanten ist dies $k_{(\bar{1}\bar{1}2)}$, für die $(11\bar{2})$ entsprechend $k_{(11\bar{2})}$. Daher wird die entsprechende Wulff-Konstruktion auch eine dreizählige Symmetrie haben. Alle Kantenlängen $S_{(\bar{1}\bar{1}2),i}$ werden also gleich sein und auch deren Abstände $h_{(\bar{1}\bar{1}2),i}$ vom Wulff-Punkt. Gleiches gilt für $S_{(11\bar{2}),i}$ bzw. $S_{(\bar{1}\bar{1}2),i}$. Aus der Konstruktion können sich so drei- oder sechseckige Inseln ergeben. Dabei wechseln sich bei sechseckigen Inseln $(\bar{1}\bar{1}2)$ und $(11\bar{2})$ Kanten ab, während dreieckige Inseln nur von einer Kantensorte begrenzt werden. Zu dreieckigen Inseln kommt es dann, wenn eine der Kantenenergien mindestens doppelt so groß ist wie die andere. In diesem Fall läge die energetisch ungünstigere Kante bei der Wulff-Konstruktion außerhalb des von drei energetisch günstigeren Kanten gebildeten Dreiecks und würde somit nicht zu den innersten Kanten beitragen.

Der Faktor $1/2$, der den Grenzfall des Energieverhältnisses markiert, ab dem dies auftritt, ergibt sich aus geometrischen Überlegungen. Zum einen läßt sich dieser Faktor direkt aus der Wulff-Konstruktion ableiten. Zum anderen kann man sich überlegen, daß es erst dann energetisch günstiger sein kann, von einer dreieckigen zu einer sechseckigen Inselform überzugehen, wenn durch die zusätzlichen Kanten die Kantenenergie vermindert wird. Da die zusätzlichen Kanten aber gleichseitige Dreiecke von der Insel abschneiden, erspart man sich also durch Ein-

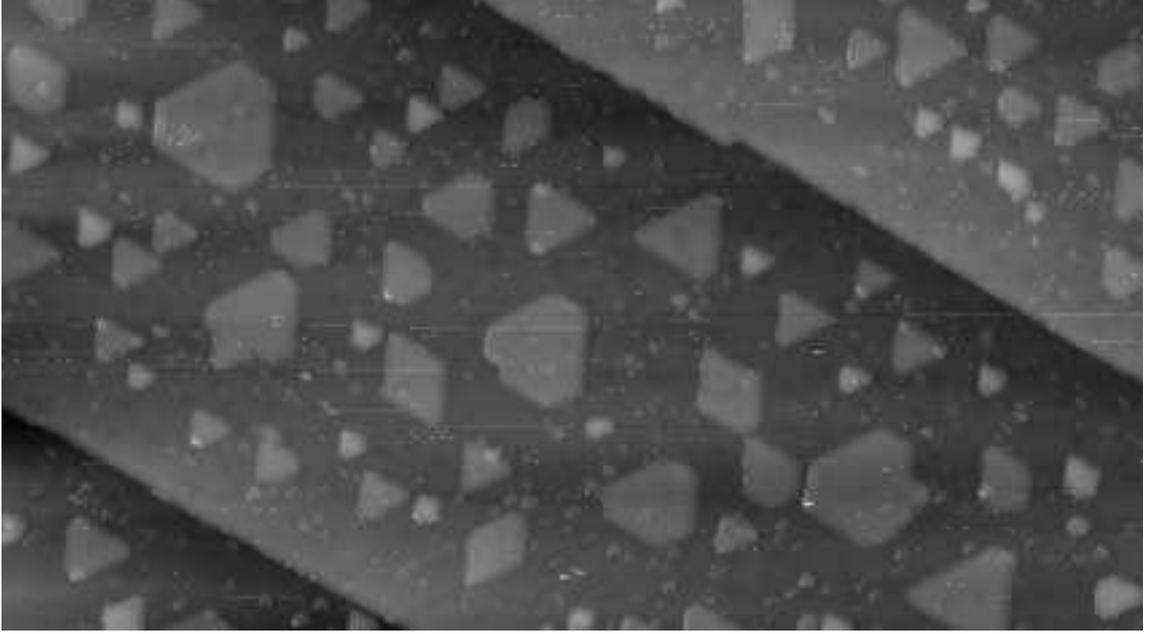


Abbildung 4.21: Inselbildung auf einer Arsen-bedeckten Si(111)-Oberfläche im STM-Bild. Die lange, sich durchs Bild ziehende Stufenkante ist eine $(\bar{1}\bar{1}2)$ -Kante. Während die meist kleineren nach rechts zeigenden Dreiecke $(11\bar{2})$ -Kanten haben, überwiegen bei größeren Inseln (z.B. oben links) eher $(\bar{1}\bar{1}2)$ -Kanten.

setzen einer solchen Kante S_1 mit Kantenenergie k_1 die doppelte Kantenlänge $\Delta S_2 + \Delta S'_2 = 2S_1$ der benachbarten Kanten, die in unserem Fall vom gleichen Typ mit Kantenenergie k_2 sind. Dabei wird die Fläche um das abgeschnittene Dreieck verkleinert, wobei dieses ΔA aber durch Abschneiden entsprechend kleiner Dreiecke beliebig klein gehalten werden kann. Damit dabei E_K^2/A verringert werden kann, muß offenbar E_K kleiner werden, also $\Delta E_K = S_1 k_1 - 2S_1 k_2 < 0$ sein. Sechseckige Inseln ergeben sich daher für:

$$\frac{1}{2}k_{(11\bar{2})} < k_{(\bar{1}\bar{1}2)} < 2k_{(11\bar{2})}. \quad (4.15)$$

Aus der Geometrie der Inselnform ergibt sich in unserem Fall:

$$S_{(11\bar{2})} = \frac{2}{\sqrt{3}}(2h_{(\bar{1}\bar{1}2)} - h_{(11\bar{2})}), \quad S_{(\bar{1}\bar{1}2)} = \frac{2}{\sqrt{3}}(2h_{(11\bar{2})} - h_{(\bar{1}\bar{1}2)}). \quad (4.16)$$

Durch Einsetzen in (4.1) und (4.3) erhält man dann:

$$\frac{E_K^2}{2A} = 2\sqrt{3} \times \begin{cases} 3k_{(11\bar{2})}^2 \\ 3k_{(\bar{1}\bar{1}2)}^2 \\ 4k_{(11\bar{2})}k_{(\bar{1}\bar{1}2)} - k_{(11\bar{2})}^2 - k_{(\bar{1}\bar{1}2)}^2 \end{cases} \quad \text{für } \begin{cases} k_{(\bar{1}\bar{1}2)} > 2k_{(11\bar{2})} \\ k_{(11\bar{2})} > 2k_{(\bar{1}\bar{1}2)} \\ \text{sonst.} \end{cases} \quad (4.17)$$

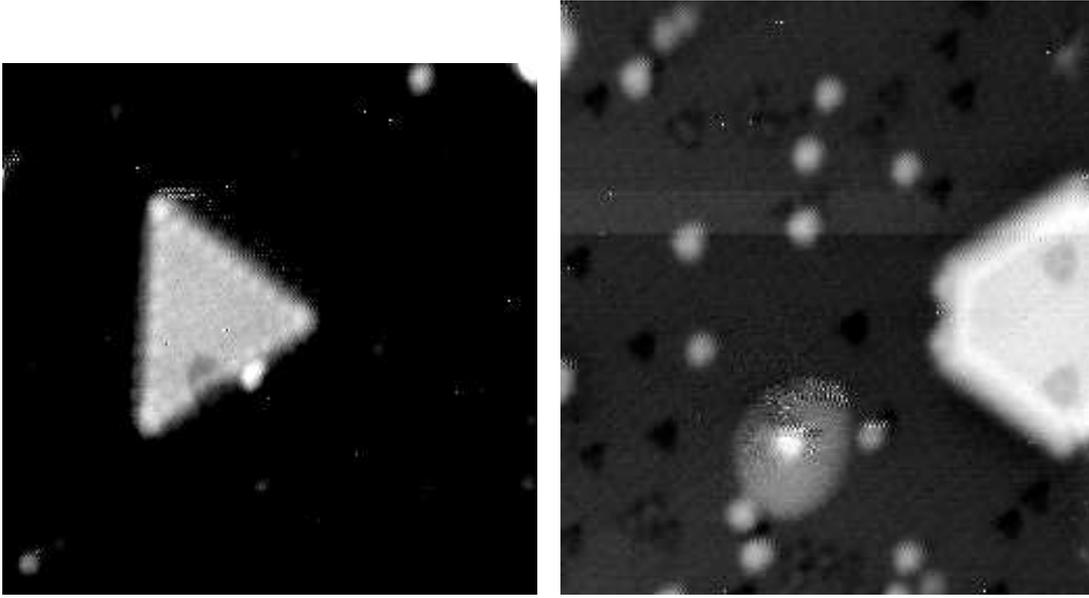


Abbildung 4.22: Auf der Oberfläche beobachtete Inseln. Inseln mit der links zu beobachteten Ausrichtung, und $(11\bar{2})$ -Kanten sind am häufigsten zu beobachten und meist kleiner. Die Inselnform des rechten Teilbildes mit längeren $(\bar{1}\bar{1}2)$ -Kanten ist eher in der Terrassenmitte zu beobachten und meist größer.

Indem man (4.13) in (4.16) einsetzt erhält man für den Anteil einer Kantenart an der Gesamtseitenlänge (wenn (4.15) erfüllt ist):

$$\frac{S_{(11\bar{2})}}{S_{(11\bar{2})} + S_{(\bar{1}\bar{1}2)}} = \frac{2k_{(\bar{1}\bar{1}2)} - k_{(11\bar{2})}}{k_{(11\bar{2})} + k_{(\bar{1}\bar{1}2)}}, \quad \frac{S_{(\bar{1}\bar{1}2)}}{S_{(11\bar{2})} + S_{(\bar{1}\bar{1}2)}} = \frac{2k_{(11\bar{2})} - k_{(\bar{1}\bar{1}2)}}{k_{(11\bar{2})} + k_{(\bar{1}\bar{1}2)}}. \quad (4.18)$$

Da $k_{(\bar{1}\bar{1}2)}$ und $k_{(11\bar{2})}$ von der Arsen-Referenzenergie abhängen, wirkt sich diese auch auf die Kantenlängen $S_{(\bar{1}\bar{1}2)}$ und $S_{(11\bar{2})}$ und so auf die Inselnform aus. Im rechten Bild des Diagramms 4.19 wurden die Größen für Arsen-Bulk als Referenz benutzt. Mit der As_4 -Referenzenergie erhält man nur von $(11\bar{2})$ -Kanten begrenzte dreieckige Inseln, da in diesem Fall $k_{(\bar{1}\bar{1}2)} > 2k_{(11\bar{2})}$ wird.

4.4.2 STM-Bilder

Von B. Voigtländer [Voi] wurden mir Rastertunnelmikroskopie (STM)-Bilder der Arsen-bedeckten Si(111)-Oberfläche zur Verfügung gestellt. In Abbildung 4.21 lassen sich drei Inselnformen unterscheiden; nach ‘rechts’ zeigende Dreiecke, nach ‘links’ zeigende Dreiecke mit abgestumpften Spitzen und unregelmäßige Inseln. Besonders interessant sind dabei die dreieckigen Formen. Die nach rechts zeigenden Dreiecke haben $(11\bar{2})$ Kanten⁷. Diese Inselnform ist am häufigsten, allerdings

⁷Die Ausrichtung des Siliziumkristalls wurde vor der Beschichtung mit Arsen festgestellt. Dazu wurde die 7×7 -Rekonstruktion auf der reinen Si(111)-Oberfläche beobachtet. Bei negativer

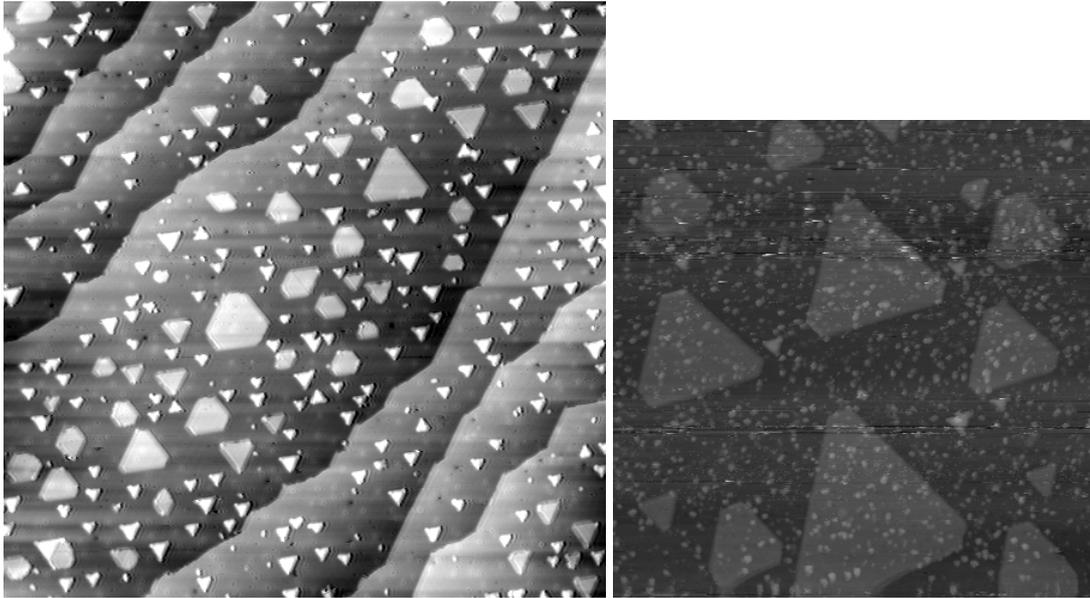


Abbildung 4.23: Auswirkung längeren Heizens. Auf der linken Oberfläche (aus Abb. 4.21) wurde 80 Minuten lang bei 700°C Arsen deponiert. Rechts wurde zunächst zwei Minuten bei 700°C Arsen deponiert, daraufhin wurde die Oberfläche drei Stunden bei 750°C geheizt. Die Seitenlänge der Ausschnitte sind: links 1707\AA , rechts 1365\AA , sie sind hier im gleichen Maßstab wiedergegeben.

sind diese Inseln meist kleiner als die anderen Formen. Sie sind auch eher an Terrassenrändern zu finden. Bei den Sechsecken in Form eines nach links weisenden abgestumpften Dreiecks ist die längere Kante eine $(\bar{1}\bar{1}2)$ -Kante, die kürzere eine $(11\bar{2})$ Kante. Diese sechseckigen Formen sind meist größere Inseln und finden sich eher in der Terrassenmitte. Es fällt weiter auf, daß für viele dieser Sechsecke die Verhältnisse der Kantenlängen zueinander sehr ähnlich sind.

In dem den STM-Bildern zugrundeliegendem Experiment wurde allerdings nicht von einer Arsen-terminierten Siliziumoberfläche ausgegangen, auf die dann weiteres Silizium aufgedampft wurde. Stattdessen wurde von einer reinen $\text{Si}(111)$ -Oberfläche ausgegangen, auf die Arsen aufgedampft wurde. Dabei ersetzen die Arsenatome dann die oberste Lage der Siliziumatome, die somit für die Bildung von Inseln auf der Arsen-terminierten Oberfläche zur Verfügung stehen. Dies ist ein dynamischer Vorgang, bei dem die Kantenbildung mit anderen Einzelprozessen in Konkurrenz steht, die auch Arsenatome benötigen, beispielsweise zur Bildung der obersten Atomlage dort, wo die Arsenterminierung noch nicht vollständig abgeschlossen ist. Dabei kann weder von einem einheitlichen chemischen Potential für Arsen über der gesamten Oberfläche noch von einer relaxierten Inselnform ausgegangen werden. Da die Inseln, wie zu sehen, auch sehr nah beieinander liegen,

Probenspannung erscheint dabei eine Hälfte der Einheitszelle (die "faulted half") heller als die andere ("unfaulted half"). Anhand dessen kann bei der Ausrichtung des Kristalls die $(11\bar{2})$ -Richtung von der $(\bar{1}\bar{1}2)$ -Richtung unterschieden werden.

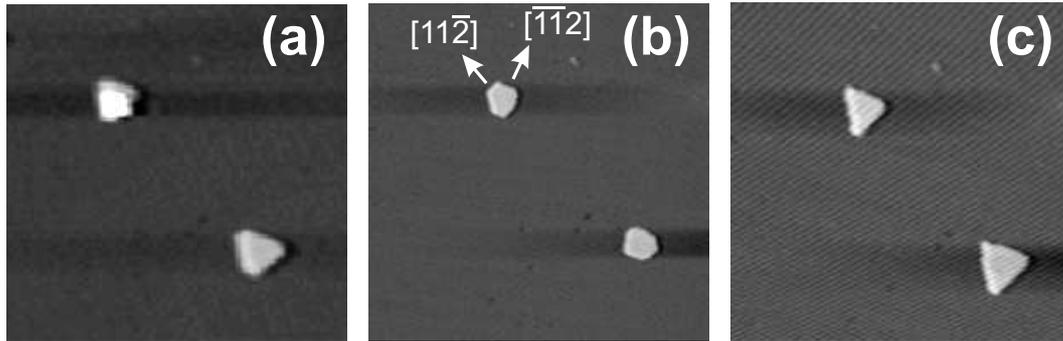


Abbildung 4.24: Inseln während des Wachstums, und nach Unterbrechung des Wachstums auf reinem Si(111) (Temperatur des Substats 725 K). Während des Wachstums haben die Inseln eine dreieckige Form (a). Nach 18 Minuten Heizen bei 775 K haben die Inseln hexagonale Form angenommen (b). Bei Fortsetzen des Wachstums nehmen die Inseln wieder eine dreieckige Form an. Die Kantenlängen entsprechen jeweils 550Å. (Aus [Voi01] p166.)

muß zudem davon ausgegangen werden, daß die Inseln auch Atome miteinander austauschen. Das STM-Bild stellt also eine Momentaufnahme eines dynamischen Prozesses dar, bei der die Inseln noch nicht ihre Idealform ausgebildet haben, sondern entweder im Wachsen oder im Schrumpfen begriffen sind. Da größere Inseln energetisch günstiger sind als kleine, ist anzunehmen, daß die kleineren eher an den Terrassenrändern auftretenden Inseln schrumpfen und dabei Material an die Terrassenkanten und größere Inseln in der Terrassenmitte abgeben. Den Prozess, bei dem größere Inseln auf Kosten kleinerer wachsen, nennt man Ostwald-Reifung. Daß hier Ostwald-Reifung beobachtet wird, ist daran zu erkennen, daß auf dem STM-Bild 4.21 verschiedene Inseln gleichzeitig auftreten. Die Häufung größerer Inseln in der Terrassenmitte ist so zu erklären, daß an den Terrassenstufen die energetisch günstigste Möglichkeit ist, Material anzulagern. Daher steht in der Nähe der Stufenkanten, also an den Rändern der Terrasse, weniger Material zum Wachstum der Inseln zur Verfügung.

Die berechneten Daten sind offenbar nicht ausreichend, um das vorliegende STM-Bild zu erklären, da in den STM-Bildern dynamische Prozesse eine Rolle spielen und nicht von einer idealen Inselnform gemäß dem Wulff-Theorem ausgegangen werden kann (diese wäre für alle Inseln gleich). Die Form einer wachsenden Insel ist vor allem durch die Geschwindigkeit bestimmt, mit der an einer Kante neues Material angelagert werden und die Kante so weiter nach außen verschoben werden kann. Wie schon Wulff experimentell beobachtete [Wul01] werden Kristalle beim Wachstum vor allem von den Flächen begrenzt, an die sich am langsamsten Material anlagert. Dies ist unmittelbar plausibel, denn diese Flächen werden am beständigsten sein, und durch Anlagerung von Material an benachbarte Flächen werden sie weiter vergrößert. Analoges gilt für wachsende Inseln und deren Kanten: In Abb 4.24 kann man dies auf reinem Si(111) deutlich beobachten: während des Wachstums in (a) bzw. (c) nehmen die Inseln eine andere Form an als nach ei-

ner Unterbrechung des Wachstums und längerem Heizen (b). Dabei spielen aber vor allem die Energiebarrieren eine Rolle, die zur Anlagerung neuen Materials an einer Kante zu überwinden sind. Umgekehrt sind schrumpfende Inseln vor allem von Kanten begrenzt, von denen schnell Material abgegeben wird. Durch die Einwärtsbewegung dieser Kanten werden benachbarte Kanten bis hin zum Verschwinden verkürzt. Aus Abb. 4.21 läßt sich daher ablesen, daß an den $(\bar{1}\bar{1}2)$ -Kanten schneller Material abgegeben oder angelagert werden kann.

In Abbildung 4.23 sind die Auswirkungen längeren Heizens zu beobachten. Das rechte Bild zeigt eine Oberfläche, die nach Dotierung mit Arsen drei Stunden lang bei 750 K geheizt wurde. Dabei wurde ein Ausschnitt der Oberfläche betrachtet, der weit von Terrassenkanten entfernt ist, so daß frei auf der Oberfläche diffundierende Atome eher auf eine Insel als auf die Terrassenkante stoßen. Man sieht, daß sich bei längerer Reifung größere Inseln ausbilden, aber es treten immer noch verschiedene Inseln gleichzeitig auf: kleine, dreieckige, von $(11\bar{2})$ -Kanten begrenzte Inseln und große Inseln, die von langen $(\bar{1}\bar{1}2)$ -Kanten und kurzen $(11\bar{2})$ -Kanten begrenzt werden. Also ist auch hier noch Ostwald-Reifung zu beobachten.

Kapitel 5

Zusammenfassung und Ausblick

Wie in Kapitel 1 dargelegt, ermöglicht die *ab-initio* Simulation von Festkörpern mit Hilfe der Dichtefunktionaltheorie detaillierte Einblicke in atomare Strukturen und Prozesse. Diese können dazu dienen, experimentelle Ergebnisse zu erklären und Anregungen für neue Experimente zu geben. Die Implementation der DFT im EStCoMPP-Programm ist in Kapitel 2 dargelegt. Die Berechnung größerer Zellen mit mehr Atomen erweitert die Auswahl von Strukturen und Prozessen, die simuliert werden können. Für die dazu nötige Rechenleistung werden moderne Supercomputer, wie z.B. die Cray-T3E eingesetzt. Derzeit sind die leistungsstärksten Computersysteme massiv parallele Rechner. Um deren Rechenleistung optimal auszunutzen, müssen Programme entsprechend angepasst, "parallelisiert" werden.

Dieser Parallelisierung widmet sich Kapitel 3. Die in Kapitel 3.1 beschriebene **k**-Punkt-Parallelisierung bietet sich dabei zunächst an, da dabei große Teile der Berechnungen unabhängig voneinander auf verschiedenen Prozessoren ablaufen können. Dies führt auch zu einem guten Skalierungsverhalten der Rechenzeit mit der Anzahl der Prozessoren. Leider ist die Parallelisierbarkeit des EStCoMPP-Programms in dieser Weise durch die Anzahl der **k**-Punkte beschränkt, gerade Berechnungen sehr großer Systeme kommen mit wenigen **k**-Punkten aus. Aus diesem Grund wurde die Parallelisierung mit der in Kapitel 3.2 beschriebenen **G**-Vektor-Parallelisierung weiter getrieben. Dabei muß, z.B. durch Auswahl einer geeigneten Datenstruktur, darauf geachtet werden, daß Rechenlast und Speicherbedarf möglichst gleichmäßig auf die einzelnen Prozessoren verteilt werden, gleichzeitig aber Overhead und Latenzzeiten durch Kommunikation nach Möglichkeit vermieden wird. Wie in Abb. 3.4 zu sehen, gelingt es so, mit bis zu ca. 60 Prozessoren pro **k**-Punkt ein sehr gutes Skalierungsverhalten der Rechenleistung mit der Zahl der Prozessoren zu erzielen. In Kapitel 3.3 wird diskutiert, wie durch Aufteilung der Zustandsvektoren auf Prozessorgruppen weiter parallelisiert werden kann, um noch mehr Prozessoren nutzen zu können. Die durch die Parallelisierung nutzbare Rechenleistung ermöglicht es, Strukturen auf Oberflächen zu simulieren.

In Kapitel 4 wird auf die Untersuchung von Stufen auf Arsen-bedeckten Si-

(111)-Oberflächen mit dem *ESCoMPP*-Programm eingegangen. Zur Untersuchung der Stufen muß, wie in Kap. 4.2.2 beschrieben, eine speziell konstruierte Superzelle verwendet werden. Die relaxierten Strukturen der berechneten Stufenkanten werden in Kapitel 4.3 und Anhang A beschrieben und diskutiert. Dabei ist die energetisch günstigste Struktur die $(11\bar{2})$ -Kante mit einem Arsenatom auf einem Siliziumplatz in der zweiten Lage, wie anhand der Tabellen in Kapitel 4.4 abzulesen ist.

Die Energie der Kanten in Abhängigkeit von deren Ausrichtung bestimmt die energetisch günstigste Form einer Insel. Diese läßt sich durch den in Kapitel 4.4.1 beschriebenen Wulff-Plot finden. Da bei der Bildung von Kanten zusätzliches Arsen gebraucht wird oder Arsen abgegeben wird, hängt die Energie der Kanten und damit auch die energetisch günstigste Inselform von der Referenzenergie für Arsen ab. Diese hängt wiederum von den experimentellen Bedingungen ab, die man zu simulieren wünscht. In den in Kapitel 4.4.2 zu sehenden STM-Bildern lassen sich von $(11\bar{2})$ - und $(\bar{1}\bar{1}2)$ -Kanten begrenzte Inseln beobachten. Da auf den Bildern der Arsen-bedeckten Si-(111)-Oberfläche jedoch verschieden geformte Inseln beobachtbar sind, ist davon auszugehen, daß die Bilder eine Momentaufnahme wachsender bzw. schrumpfender Inseln erfassen. Bei diesen ist die Inselform nicht allein durch die Kantenenergie bestimmt, sondern auch durch die Geschwindigkeit, mit der an einer Kante neues Material angelagert werden kann, was wiederum von den dynamischen Prozessen bei der Anlagerung von Atomen an der Kante und den dabei auftretenden Aktivierungsenergien abhängt.

Um einen vollständigen Überblick über die Vorgänge beim Oberflächenwachstum zu gewinnen und schließlich das Wachstum mit Monte-Carlo-Methoden zu simulieren ist es nötig, die Prozesse, die bei der Wachstumstemperatur energetisch möglich sind, zu erfassen. Ausgehend von der berechneten Struktur der Stufenkanten können nun Prozesse an den Kanten, wie z.B. Diffusion über und entlang der Stufenkanten und Austausch eines Atomes an der Stufenkante, in weiteren Rechnungen untersucht werden. Die dabei zu untersuchenden Superzellen sind jedoch sehr groß, da die betrachteten Strukturen immer ausgedehnter werden und entsprechend mehr Atome in der Oberflächenebene berücksichtigt werden müssen. In einem weiteren Schritt ist zur Berechnung der Energie einer gegebenen strukturierten Oberfläche auch die Betrachtung strukturierter Kanten (Übergänge zwischen verschiedenen Kantenorientierungen, wie z.B. versetzte Kanten und Knicke) nötig.

Anhang A

Struktur der Stufenkanten

In den folgenden Tabellen sind die berechneten Strukturen der Stufenkanten angegeben. Die Atome sind dabei mit Nummern und zusätzlich mit den in Kapitel 4.2.1 erläuterten Koordinaten bezeichnet, um die relative Lage der Atome zueinander zu verdeutlichen. Die Verschiebungen δ sind in Prozent relativ zur Gitterkonstanten a des Siliziumgitters angegeben, in physikalischen Systemen entspräche $\delta = 100$ einer Verschiebung um 5.4304\AA . Die Bindungslängen sind in Prozent relativ zum Nächstnachbarabstand in Si-Bulk angegeben, in physikalischen Systemen entspricht daher eine Bindungslänge von $100 \simeq 2.35\text{\AA}$. Die Tabellen sind wie folgt aufgebaut:

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge	Winkel
Nummer (Sym.)						Bindungspartner	Bindungen
Koordinaten	Typ	δ_* in %a				Längen in %d	Winkel in °

In Systemen, in denen an der Kante keine Symmetriebrechung in y-Richtung auftritt, so daß die Verschiebungen für alle Atome in einer Reihe gleich sind, ist neben der Nummer des Atoms in Klammern die Nummer des jeweils anderen Atoms mit gleichen x- und z- Koordinaten angegeben.

A.1 Die (112) Stufe mit Silizium an der Kante

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge				Winkel						
75	As	-2.37	0.03	3.63	4.34	75-59	75-53	75-55		59-75-53	59-75-55	53-75-55				
-4, 2, 1						102.88	102.84	101.72		104.56	102.84	102.80				
81	As	-2.15	-0.00	3.96	4.51	81-59	81-53	81-61		59-81-53	59-81-61	53-81-61				
-4, 0, 1						103.44	103.39	101.95		104.80	102.09	102.06				
79	As	-3.09	-0.46	2.28	3.87	79-59	79-67	79-65		59-79-67	59-79-65	67-79-65				
-1, 1, 1						102.15	98.54	104.66		90.78	112.95	111.08				
73	As	-3.19	0.58	2.31	3.98	73-53	73-67	73-65		53-73-67	53-73-65	67-73-65				
-1, -1, 1						102.08	98.64	104.65		90.72	113.06	110.84				
59	Si	-2.96	0.14	-0.61	3.03	59-75	59-81	59-79	59-47	75-59-81	75-59-79	75-59-47	81-59-79	81-59-47	79-59-47	
-3, 1, 2						102.88	103.44	102.15	99.18	104.73	104.82	117.15	103.04	116.53	109.09	
53	Si	-2.98	-0.11	-0.67	3.06	53-75	53-81	53-73	53-39	75-53-81	75-53-73	75-53-39	81-53-73	81-53-39	73-53-39	
-3, -1, 2						102.84	103.39	102.08	99.26	104.64	104.78	117.17	102.83	116.67	109.22	
67	Si	-13.48	0.06	-4.10	14.09	67-79	67-73	67-45		79-67-73	79-67-45	73-67-45				
0, 2, 2						98.54	98.64	98.23		114.37	122.63	122.80				
65	Si	1.43	0.02	2.04	2.49	65-79	65-73	65-37		79-65-73	79-65-37	73-65-37				
0, 0, 2						104.66	104.65	100.60		100.47	108.10	108.13				
47	Si	0.01	-0.25	-0.15	0.29	47-59	47-31	47-23		59-47-31	59-47-23	31-47-23				
-3, 1, 3						99.18	100.44	100.33		109.84	111.57	108.80				
39	Si	-0.01	0.23	-0.24	0.33	39-53	39-31	39-23		53-39-31	53-39-23	31-39-23				
-3, -1, 3						99.26	100.44	100.35		109.78	111.39	108.91				
45	Si	2.13	0.02	-0.37	2.16	45-67	45-31	45-21	45-29	67-45-31	67-45-21	67-45-29	31-45-21	31-45-29	21-45-29	
0, 2, 3						98.23	102.75	100.26	100.22	86.08	117.19	117.11	109.85	109.89	113.08	
37	Si	0.39	0.01	1.79	1.83	37-65	37-23	37-21	37-29	65-37-23	65-37-21	65-37-29	23-37-21	23-37-29	21-37-29	
0, 0, 3						100.60	100.33	100.60	100.58	113.20	110.31	110.35	108.96	108.90	104.75	
69	As	1.17	-0.06	4.20	4.36	69-21	69-63	69-57		21-69-63	21-69-57	63-69-57				
3, 1, 3						102.42	103.33	103.26		103.63	102.25	104.61				
71	As	1.09	0.09	4.19	4.33	71-29	71-63	71-57		29-71-63	29-71-57	63-71-57				
3, -1, 3						102.35	103.28	103.19		103.71	102.33	104.39				
83	As	-0.57	0.07	3.71	3.75	83-63	83-55	83-61		63-83-55	63-83-61	55-83-61				
6, 2, 3						102.15	103.21	103.29		102.65	102.99	104.48				
77	As	-0.45	-0.03	3.75	3.78	77-57	77-55	77-61		57-77-55	57-77-61	55-77-61				
6, 0, 3						102.00	103.17	103.26		102.45	102.76	104.61				
31	Si	0.54	-0.02	0.61	0.82	31-47	31-39	31-45	31- 8	47-31-39	47-31-45	47-31- 8	39-31-45	39-31- 8	45-31- 8	
-2, 2, 4						100.44	100.44	102.75	101.43	109.85	111.39	107.97	111.37	107.81	108.31	

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge				Winkel					
23	Si	0.89	0.01	0.06	0.89	23-47	23-39	23-37	23- 4	47-23-39	47-23-37	47-23- 4	39-23-37	39-23- 4	37-23- 4
-2, 0, 4						100.33	100.35	100.33	100.17	107.85	109.45	108.50	109.54	108.36	113.01
21	Si	1.49	-0.84	0.58	1.81	21-45	21-37	21-69	21-14	45-21-37	45-21-69	45-21-14	37-21-69	37-21-14	69-21-14
1, 1, 4						100.26	100.60	102.42	100.26	108.91	105.95	106.42	108.59	111.10	115.57
29	Si	1.46	0.87	0.56	1.79	29-45	29-37	29-71	29-20	45-29-37	45-29-71	45-29-20	37-29-71	37-29-20	71-29-20
1, -1, 4						100.22	100.58	102.35	100.23	108.95	105.89	106.46	108.52	111.19	115.53
63	Si	0.16	0.05	-0.54	0.57	63-69	63-71	63-83	63-51	69-63-71	69-63-83	69-63-51	71-63-83	71-63-51	83-63-51
4, 2, 4						103.33	103.28	102.15	99.38	104.76	102.28	115.19	102.34	115.33	115.22
57	Si	0.37	-0.02	-0.53	0.65	57-69	57-71	57-77	57-43	69-57-71	69-57-77	69-57-43	71-57-77	71-57-43	77-57-43
4, 0, 4						103.26	103.19	102.00	99.53	104.24	102.46	115.12	102.55	115.21	115.52
55	Si	-1.34	0.03	-0.81	1.56	55-75	55-83	55-77	55-41	75-55-83	75-55-77	75-55-41	83-55-77	83-55-41	77-55-41
7, 1, 4						101.72	103.21	103.17	99.26	102.43	102.29	113.86	104.80	115.87	115.82
61	Si	-1.08	0.01	-0.80	1.34	61-81	61-83	61-77	61-49	81-61-83	81-61-77	81-61-49	83-61-77	83-61-49	77-61-49
7, -1, 4						101.95	103.29	103.26	99.28	102.49	102.36	114.60	104.29	115.65	115.67
8	Si	-0.00	0.00	-0.00	0.00	8-31									
-2, 2, 5						101.43									
4	Si	-0.00	0.00	-0.00	0.00	4-23									
-2, 0, 5						100.17									
14	Si	0.30	-0.04	0.48	0.57	14-21									
1, 1, 5						100.26									
20	Si	0.32	0.05	0.48	0.58	20-29									
1, -1, 5						100.23									
51	Si	-0.01	-0.00	-0.27	0.27	51-63									
4, 2, 5						99.38									
43	Si	0.03	0.01	-0.33	0.33	43-57									
4, 0, 5						99.53									
41	Si	-0.21	-0.03	-0.47	0.51	41-55									
7, 1, 5						99.26									
49	Si	-0.20	0.03	-0.48	0.52	49-61									
7, -1, 5						99.28									

A.4 Die $(\bar{1}\bar{1}2)$ Stufe mit Silizium an der Kante

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge				Winkel					
79	As	-2.99	0.21	4.78	5.64	79-67	79-59	79-53		67-79-59	67-79-53	59-79-53			
-2, 2, 1						103.55	103.39	102.25		112.24	91.40	102.55			
77	As	-0.48	-0.14	3.89	3.92	77-59	77-57	77-51		59-77-57	59-77-51	57-77-51			
-5, 1, 1						102.83	103.46	103.67		103.60	103.46	103.17			
75	As	0.80	0.00	4.65	4.72	75-57	75-55	75-49		57-75-55	57-75-49	55-75-49			
3, -1, 3						101.86	103.41	102.57		103.24	102.36	104.18			
73	As	-3.49	-1.24	3.95	5.42	73-65	73-59	73-53		65-73-59	65-73-53	59-73-53			
-2, 0, 1						99.90	104.17	103.57		110.15	77.05	106.49			
71	As	-2.71	-0.18	4.20	5.00	71-57	71-53	71-51		57-71-53	57-71-51	53-71-51			
-5, -1, 1						102.45	101.33	102.40		102.13	106.66	100.11			
69	As	0.51	-0.38	4.60	4.64	69-55	69-51	69-49		55-69-51	55-69-49	51-69-49			
3, 1, 3						103.67	101.92	102.88		102.87	105.21	102.23			
67	Si	-5.09	12.20	0.05	13.22	67-79	67-65	67-63		79-67-65	79-67-63	65-67-63			
0, -2, 2						103.55	97.05	101.02		98.73	116.57	95.80			
65	Si	-13.87	-18.07	-7.32	23.93	65-73	65-67	65-61		73-65-67	73-65-61	67-65-61			
0, 0, 2						99.90	97.05	99.23		114.66	125.61	118.19			
63	Si	0.24	1.12	1.38	1.79	63-67	63-55	63-31	63-23	67-63-55	67-63-31	67-63-23	55-63-31	55-63-23	31-63-23
0, -2, 3						101.02	101.63	100.11	100.94	116.66	97.42	118.28	109.17	107.05	107.24
61	Si	0.72	-0.44	-0.38	0.93	61-65	61-49	61-31	61-23	65-61-49	65-61-31	65-61-23	49-61-31	49-61-23	31-61-23
0, 0, 3						99.23	100.59	101.31	100.46	126.07	79.06	115.58	110.15	110.58	109.73
59	Si	-1.08	0.04	-0.79	1.34	59-79	59-77	59-73	59-47	79-59-77	79-59-73	79-59-47	77-59-73	77-59-47	73-59-47
-3, 1, 2						103.39	102.83	104.17	99.05	100.66	106.84	116.14	100.07	116.97	114.11
57	Si	-0.52	0.09	-0.51	0.73	57-77	57-75	57-71	57-45	77-57-75	77-57-71	77-57-45	75-57-71	75-57-45	71-57-45
-6, 0, 2						103.46	101.86	102.45	99.29	102.93	105.11	114.65	100.06	117.17	115.01
55	Si	0.76	-0.02	0.78	1.09	55-75	55-69	55-63	55-43	75-55-69	75-55-63	75-55-43	69-55-63	69-55-43	63-55-43
2, 2, 4						103.41	103.67	101.63	100.36	104.90	105.60	114.53	107.68	114.50	109.05
53	Si	-3.75	-1.26	0.15	3.96	53-79	53-73	53-71	53-39	79-53-73	79-53-71	79-53-39	73-53-71	73-53-39	71-53-39
-3, -1, 2						102.25	103.57	101.33	100.57	102.00	105.73	113.46	103.13	110.50	120.09
51	Si	-0.72	-0.34	-0.47	0.93	51-77	51-71	51-69	51-37	77-51-71	77-51-69	77-51-37	71-51-69	71-51-37	69-51-37
-6, 2, 2						103.67	102.40	101.92	99.33	104.81	103.14	114.49	100.54	114.92	117.09
49	Si	1.24	-0.36	0.38	1.35	49-75	49-69	49-61	49-35	75-49-69	75-49-61	75-49-35	69-49-61	69-49-35	61-49-35
2, 0, 4						102.57	102.88	100.59	100.06	104.48	107.04	116.01	106.82	115.20	106.71
47	Si	0.09	-0.25	-0.36	0.45	47-59	47-29	47-23	47-21	59-47-29	59-47-23	59-47-21	29-47-23	29-47-21	23-47-21
-3, 1, 3						99.05	100.08	100.23	100.02	108.17	108.78	108.60	110.01	109.89	111.31

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge				Winkel						
45	-6, 0, 3	Si	0.08	-0.08	-0.20	0.23	45-57 99.29	45-27 99.95	45-21 100.16	45-19 99.93	57-45-27 108.66	57-45-21 110.04	57-45-19 108.90	27-45-21 110.29	27-45-19 109.42	21-45-19 109.50
43	2, 2, 5	Si	0.06	0.03	0.63	0.63	43-55 100.36	43-19 100.17			55-43-19 109.29					
41	-1, 1, 5	Si	0.33	0.19	0.64	0.75	41-23 101.36									
39	-3, -1, 3	Si	0.19	-0.54	0.08	0.58	39-53 100.57	39-31 101.19	39-29 99.91	39-21 100.27	53-39-31 116.79	53-39-29 106.45	53-39-21 107.78	31-39-29 108.83	31-39-21 107.96	29-39-21 108.83
37	-6, 2, 3	Si	-0.07	-0.18	-0.17	0.26	37-51 99.33	37-29 100.15	37-27 99.81	37-19 100.02	51-37-29 110.41	51-37-27 108.69	51-37-19 108.88	29-37-27 109.48	29-37-19 109.60	27-37-19 109.76
35	2, 0, 5	Si	0.04	-0.09	0.37	0.38	35-49 100.06	35-27 100.06			49-35-27 108.44					
33	-1, -1, 5	Si	-0.20	0.28	-1.15	1.20	33-31 98.96									
31	-1, -1, 4	Si	0.11	-0.27	-1.60	1.63	31-63 100.11	31-61 101.31	31-39 101.19	31-33 98.96	63-31-61 104.95	63-31-39 106.74	63-31-33 114.52	61-31-39 108.62	61-31-33 110.49	39-31-33 111.18
29	-4, 2, 4	Si	-0.03	-0.23	-0.25	0.34	29-47 100.08	29-39 99.91	29-37 100.15	29-14 99.89	47-29-39 108.82	47-29-37 109.68	47-29-14 109.48	39-29-37 109.53	39-29-14 109.68	37-29-14 109.64
27	-7, 1, 4	Si	-0.07	-0.13	-0.04	0.15	27-45 99.95	27-37 99.81	27-35 100.06	27- 6 99.91	45-27-37 109.45	45-27-35 109.53	45-27- 6 109.36	37-27-35 109.25	37-27- 6 109.12	35-27- 6 110.12
23	-1, 1, 4	Si	0.72	0.41	1.23	1.48	23-63 100.94	23-61 100.46	23-47 100.23	23-41 101.36	63-23-61 111.99	63-23-47 110.45	63-23-41 109.99	61-23-47 110.38	61-23-41 107.13	47-23-41 106.72
21	-4, 0, 4	Si	0.20	-0.45	-0.06	0.49	21-47 100.02	21-45 100.16	21-39 100.27	21-10 100.07	47-21-45 109.15	47-21-39 109.90	47-21-10 108.68	45-21-39 109.86	45-21-10 109.05	39-21-10 110.18
19	-7, -1, 4	Si	-0.07	-0.06	0.07	0.12	19-45 99.93	19-43 100.17	19-37 100.02	19- 2 100.16	45-19-43 109.33	45-19-37 109.74	45-19- 2 108.99	43-19-37 109.31	43-19- 2 110.30	37-19- 2 109.16
14	-4, 2, 5	Si	0.05	-0.02	-0.20	0.20	14-29 99.89									
10	-4, 0, 5	Si	0.06	-0.04	-0.09	0.11	10-21 100.07									
6	-7, 1, 5	Si	-0.00	0.00	0.00	0.00	6-27 99.91									
2	-7, -1, 5	Si	-0.00	0.00	0.00	0.00	2-19 100.16									

A.5 Die $(\bar{1}\bar{1}2)$ Stufe mit ersetzttem Arsen

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge				Winkel						
79	-2, 2, 1	As	-2.66	-1.56	3.12	4.39	79-67	79-59	79-53		67-79-59	67-79-53	59-79-53			
							106.05	101.78	104.72		95.59	113.28	104.76			
77	-5, 1, 1	As	-1.58	-0.05	3.71	4.04	77-59	77-57	77-51		59-77-57	59-77-51	57-77-51			
							101.24	102.83	102.71		101.48	101.09	106.32			
75	3, -1, 3	As	0.60	-0.24	4.63	4.67	75-57	75-55	75-49		57-75-55	57-75-49	55-75-49			
							102.53	103.36	103.80		103.30	102.88	105.05			
73	-2, 0, 1	As	-2.57	1.54	3.11	4.32	73-65	73-59	73-53		65-73-59	65-73-53	59-73-53			
							106.05	101.74	104.67		95.65	113.50	104.55			
71	-5, -1, 1	As	0.22	0.04	3.75	3.76	71-57	71-53	71-51		57-71-53	57-71-51	53-71-51			
							103.80	103.35	103.62		103.73	102.80	103.31			
69	3, 1, 3	As	1.09	0.27	4.59	4.73	69-55	69-51	69-49		55-69-51	55-69-49	51-69-49			
							102.97	102.22	103.40		103.92	103.61	103.43			
67	0, 2, 2	As	-2.00	-12.04	0.89	12.24	67-79	67-65	67-63		79-67-65	79-67-63	65-67-63			
							106.05	107.53	103.20		103.27	109.51	103.58			
65	0, 0, 2	As	-1.92	12.11	0.90	12.29	65-73	65-67	65-61		73-65-67	73-65-61	67-65-61			
							106.05	107.53	103.22		103.23	109.62	103.61			
63	0, 2, 3	Si	-0.27	-1.56	0.78	1.77	63-67	63-55	63-31	63-23	67-63-55	67-63-31	67-63-23	55-63-31	55-63-23	31-63-23
							103.20	101.82	101.22	99.21	111.21	118.85	100.42	106.67	109.92	109.52
61	0, 0, 3	Si	-0.04	1.58	0.80	1.77	61-65	61-49	61-31	61-23	65-61-49	65-61-31	65-61-23	49-61-31	49-61-23	31-61-23
							103.22	102.11	101.33	99.30	111.28	118.74	100.31	107.03	110.28	108.95
59	-3, 1, 2	Si	-3.17	-0.05	-1.41	3.47	59-79	59-77	59-73	59-47	79-59-77	79-59-73	79-59-47	77-59-73	77-59-47	73-59-47
							101.78	101.24	101.74	99.11	103.47	100.20	113.30	103.59	120.54	113.30
57	-6, 0, 2	Si	-0.38	-0.26	-0.52	0.69	57-77	57-75	57-71	57-45	77-57-75	77-57-71	77-57-45	75-57-71	75-57-45	71-57-45
							102.83	102.53	103.80	99.32	101.65	104.30	114.41	103.67	116.60	114.53
55	2, 2, 4	Si	0.53	-0.23	0.77	0.97	55-75	55-69	55-63	55-43	75-55-69	75-55-63	75-55-43	69-55-63	69-55-43	63-55-43
							103.36	102.97	101.82	100.28	103.58	108.79	114.28	106.63	114.79	108.37
53	-3, -1, 2	Si	-0.40	0.02	-1.48	1.53	53-79	53-73	53-71	53-39	79-53-73	79-53-71	79-53-39	73-53-71	73-53-39	71-53-39
							104.72	104.67	103.35	98.49	108.99	100.11	114.52	100.19	114.50	116.75
51	-6, 2, 2	Si	-0.05	0.26	-0.56	0.62	51-77	51-71	51-69	51-37	77-51-71	77-51-69	77-51-37	71-51-69	71-51-37	69-51-37
							102.71	103.62	102.22	99.33	104.86	101.23	114.65	103.19	114.80	116.38
49	2, 0, 4	Si	0.93	0.25	0.85	1.28	49-75	49-69	49-61	49-35	75-49-69	75-49-61	75-49-35	69-49-61	69-49-35	61-49-35
							103.80	103.40	102.11	100.39	105.08	108.33	114.19	106.17	114.73	107.91
47	-3, 1, 3	Si	-0.14	-0.04	-0.92	0.93	47-59	47-29	47-23	47-21	59-47-29	59-47-23	59-47-21	29-47-23	29-47-21	23-47-21
							99.11	99.69	100.60	99.64	106.96	114.58	106.90	109.02	110.15	109.16

Atom		δ_x	δ_y	δ_z	δ	Bindungslänge				Winkel					
45	Si	-0.05	-0.12	-0.22	0.26	45-57	45-27	45-21	45-19	57-45-27	57-45-21	57-45-19	27-45-21	27-45-19	21-45-19
-6, 0, 3						99.32	99.99	100.11	99.82	108.98	110.27	108.79	109.33	109.69	109.75
43	Si	-0.05	-0.03	0.65	0.66	43-55	43-19			55-43-19					
2, 2, 5						100.28	100.11			109.58					
41	Si	-0.05	-0.02	-1.08	1.08	41-23									
-1, 1, 5						98.43									
39	Si	0.11	0.03	-0.82	0.83	39-53	39-31	39-29	39-21	53-39-31	53-39-29	53-39-21	31-39-29	31-39-21	29-39-21
-3, -1, 3						98.49	100.05	99.93	99.88	106.78	108.72	108.70	111.46	111.57	109.50
37	Si	0.04	0.11	-0.27	0.29	37-51	37-29	37-27	37-19	51-37-29	51-37-27	51-37-19	29-37-27	29-37-19	27-37-19
-6, 2, 3						99.33	100.09	100.00	99.84	109.94	109.07	108.88	109.47	109.92	109.54
35	Si	0.10	0.05	0.69	0.70	35-49	35-27			49-35-27					
2, 0, 5						100.39	100.16			109.19					
33	Si	0.51	0.02	0.84	0.98	33-31									
-1, -1, 5						101.77									
31	Si	0.92	0.05	1.61	1.85	31-63	31-61	31-39	31-33	63-31-61	63-31-39	63-31-33	61-31-39	61-31-33	39-31-33
-1, -1, 4						101.22	101.33	100.05	101.77	114.71	109.57	108.31	109.91	108.37	105.55
29	Si	-0.01	0.02	-0.53	0.53	29-47	29-39	29-37	29-14	47-29-39	47-29-37	47-29-14	39-29-37	39-29-14	37-29-14
-4, 2, 4						99.69	99.93	100.09	99.60	109.94	109.48	108.96	109.51	109.02	109.92
27	Si	-0.04	0.01	0.10	0.11	27-45	27-37	27-35	27- 6	45-27-37	45-27-35	45-27- 6	37-27-35	37-27- 6	35-27- 6
-7, 1, 4						99.99	100.00	100.16	100.22	110.00	109.25	108.99	109.31	108.95	110.33
23	Si	-0.17	-0.04	-1.76	1.77	23-63	23-61	23-47	23-41	63-23-61	63-23-47	63-23-41	61-23-47	61-23-41	47-23-41
-1, 1, 4						99.21	99.30	100.60	98.43	103.63	107.70	113.17	107.99	113.23	110.69
21	Si	-0.11	-0.03	-0.50	0.52	21-47	21-45	21-39	21-10	47-21-45	47-21-39	47-21-10	45-21-39	45-21-10	39-21-10
-4, 0, 4						99.64	100.11	99.88	99.62	109.60	109.71	108.87	109.66	110.02	108.96
19	Si	-0.23	-0.02	0.02	0.23	19-45	19-43	19-37	19- 2	45-19-43	45-19-37	45-19- 2	43-19-37	43-19- 2	37-19- 2
-7, -1, 4						99.82	100.11	99.84	100.04	109.43	109.23	109.01	109.53	110.65	108.97
14	Si	0.06	0.03	-0.36	0.37	14-29									
-4, 2, 5						99.60									
10	Si	0.03	-0.03	-0.34	0.34	10-21									
-4, 0, 5						99.62									
6	Si	-0.00	0.00	0.00	0.00	6-27									
-7, 1, 5						100.22									
2	Si	-0.00	0.00	0.00	0.00	2-19									
-7, -1, 5						100.04									

Anhang B

Davidson-Kosugi-Algorithmus

Iterative Eigenwert-/ Eigenvekturlöser gehen von einem Satz Startvektoren $|\psi_\nu^{(0)}\rangle$ ($\nu = 1..M$) aus, aus denen im ersten Schritt ein verbesserter Satz Vektoren $|\psi_\nu^{(1)}\rangle$ gewonnen wird, die eine bessere Näherung der Eigenvektoren darstellen. Dies wird so oft iteriert, bis schließlich ein Satz von Vektoren $|\psi_\nu^{(k)}\rangle$ ein vorgegebenes Qualitätskriterium erfüllt. Iterative Eigenwertlöser haben verschiedene Vorteile, die sich das **EStCoMPP**-Programm zunutze macht:

- Die zu diagonalisierende Matrix muß nicht explizit aufgestellt werden, es genügt, den Operator, dessen Eigenvektoren gesucht sind, auf Vektoren anwenden zu können. Eine explizite Darstellung von \mathcal{H} würde schon für $N = 30000$ ebene Wellen ca. 13.5GB Speicher benötigen. Dies hätte schon 128 PEs mit je 128MB Speicher¹ allein mit der Speicherung von \mathcal{H} ausgelastet. Auch wenn massiv parallele Maschinen inzwischen derartigen Speicheranforderungen genügen, wächst der benötigte Speicher quadratisch mit N . Ab $N = 130000$ würden selbst der Speicher von 512 mit je 512MB ausgestatteten PEs² nicht mehr ausreichen.
- Es ist möglich, nur für einen Teil des Eigenwertspektrums die Eigenvektoren zu berechnen, nicht benötigte Eigenvektoren werden nicht berechnet. Daher skaliert die Rechenzeit mit NM^2 , und nicht wie bei direkten Eigenwertlösern mit N^3 . Da nur die untersten 3-5% der Eigenvektoren und -werte tatsächlich benötigt werden, bietet diese Einsparung der Rechenzeit Zugang zu erheblich größeren Systemen.
- Die Eigenvektoren konvergieren mit der Zahl der Iterationen. Da auch die Dichte $n(\mathbf{r})$ sich erst mit der Zahl der elektronischen Selbstkonsistenzschritt-

¹Ursprüngliche Ausstattung der CRAY-T3E im ZAM des Forschungszentrums Jülich.

²Derzeitige Ausstattung der CRAY-T3E-1200 am John von Neumann-Institut für Computing (NIC)

te verbessert, ist es nicht nötig, die Eigenvektoren schon im ersten elektronischen Selbstkonsistenzschritt zur Konvergenz zu bringen. Stattdessen wird schon nach wenigen Davidson-Schritten ein Schritt der elektronischen Selbstkonsistenz berechnet und die letzten berechneten Eigenvektoren als Startvektoren im nächsten Selbstkonsistenzschritt verwendet. Da sich mit der Konvergenz der selbstkonsistenten Dichte $n(\mathbf{r})$ der Hamiltonian \mathcal{H} immer weniger ändert und die Eigenvektoren des vorigen Selbstkonsistenzschrittes immer bessere Startvektoren bilden, können so die Dichte $n(\mathbf{r})$ und die Eigenvektoren $|\psi_\nu\rangle$ gleichzeitig zur Konvergenz gebracht werden.

Die Qualität der Startvektoren ist wesentlich für die Zahl der benötigten Iterationen. Ab dem zweiten Schritt der elektronischen Selbstkonsistenz stehen dafür die im vorigen Schritt berechneten $|\psi_\nu\rangle$ zur Verfügung, im ersten Schritt muß jedoch einmalig ein Satz Startvektoren bestimmt werden. Dazu wird \mathcal{H} einfach auf einen Unterraum projiziert, der klein genug ist, um mit konventionellen Eigenwertlösern eine Lösung zu finden³. Die ersten M Lösungsvektoren bilden die Startvektoren für das Davidson-Kosugi-Verfahren im ersten Selbstkonsistenzschritt.

Die Auswahl des Unterraums wird dadurch getroffen, daß man die Startvektoren als ebene Wellen darstellt oder zusätzlich als Linearkombination von atomaren Orbitalen zu den gegebenen Pseudopotentialen, die dann ebenfalls nach ebenen Wellen entwickelt werden.

Bei der iterativen Bestimmung von Eigenvektoren werden zu jedem Startvektor $|\psi\rangle$ jeweils Korrekturvektoren $|\psi^{korr}\rangle$ berechnet und hinzuaddiert. Bei der Konstruktion von Korrekturvektoren spielt der Rayleigh-Quotient eine entscheidende Rolle:

$$\rho[\psi] = \frac{\langle\psi|\mathcal{H}|\psi\rangle}{\langle\psi|\psi\rangle} \quad (\text{B.1})$$

Ist $|\psi_{i_0}\rangle$ ein Eigenvektor, so ist der entsprechende Rayleigh-Quotient $\rho[\psi_{i_0}]$ identisch mit dem Eigenwert ϵ_{i_0} . Daher wird der Rayleigh-Quotient für Eigenlösungen $|\psi_\nu\rangle$ des Hamiltonoperators \mathcal{H} stationär. Damit läßt sich das hermitesche Eigenwertproblem auf ein Optimierungsproblem abbilden, dessen Zielfunktion der Rayleigh-Quotient ist. Im allgemeinen werden Gradientenverfahren zur Optimierung eingesetzt, wenn der Gradient einer Zielfunktion leicht zu bestimmen ist. Der Gradient des Rayleigh-Quotienten ergibt sich zu:

$$|\mathcal{R}[\psi]\rangle = (\mathcal{H} - \rho[\psi])|\psi\rangle \quad (\text{B.2})$$

$$\text{mit: } \left\langle \frac{\delta\rho[\psi]}{\delta\psi} \right\rangle = \frac{|\mathcal{R}[\psi]\rangle}{\langle\psi|\psi\rangle} \quad (\text{B.3})$$

³Der Unterraum zur Bestimmung der Startvektoren kann z.B. von den ebenen Wellen mit den niedrigsten $|\mathbf{G}|$ aufgespannt werden. Es können auch zusätzlich Linearkombinationen der atomaren Orbitale zu den gegebenen Pseudopotentialen hinzu genommen werden, die ebenfalls nach ebenen Wellen entwickelt werden.

Das Residuum $|\mathcal{R}[\psi]\rangle$ legt die Richtung des Gradienten und somit die Korrekturrichtung fest. Des weiteren liefert der Absolutbetrag $||\mathcal{R}[\psi]\rangle|$ ein Maß für die Güte des approximierten Eigenwertes und Eigenvektors. Für einen Eigenvektor $|\psi_\nu\rangle$ verschwindet er:

$$|\mathcal{R}[\psi_{i_0}]\rangle = (\mathcal{H} - \epsilon_{i_0}) |\psi_{i_0}\rangle = \mathbf{0} \quad (\text{B.4})$$

Unterschreitet $||\mathcal{R}[\psi]\rangle|$ ein vorgegebenes Kriterium wird der zugehörige Vektor im **ESStCoMPP**-Programm als auskonvergiert betrachtet.

Folgt man dem Gradienten mit einem Steepest-Descent-Algorithmus, so ist das Verfahren für kleine Schrittweiten zwar sehr stabil, es konvergiert jedoch im allgemeinen nur langsam. Steepest-Descent-Algorithmen konvergieren ideal für isotrope Minima, werden jedoch für anisotrope Minima zunehmend schlechter. Anschaulich bedeutet diese Anisotropie im zweidimensionalen Fall ein langgestrecktes schmales Tal, welches der Steepest-Descent-Algorithmus im Zick-Zack zu seinem tiefsten Punkt verfolgt. Durch eine "Präkonditionierung" werden die anisotropen Minima auf möglichst isotrope Minima abgebildet, die gut mit einem Steepest-Descent-Algorithmus auffindbar sind. Dazu ist es nötig, für die Zielformation, bzw. ihren Gradienten eine geeignete Transformation zu finden. Entwickelt man (B.3) um (B.4), so findet man:

$$|\mathcal{R}[\psi_{i_0} + \delta\psi]\rangle = (\mathcal{H} - \epsilon_{i_0}) |\delta\psi\rangle + \mathcal{O}((\delta\psi)^2) \quad (\text{B.5})$$

$$= (\mathcal{H} - \rho[\psi]) |\delta\psi\rangle + \mathcal{O}((\delta\psi)^2) . \quad (\text{B.6})$$

Indem (B.6) mit einer Approximation der Matrix $(\mathcal{H} - \rho[\psi])^{-1}$ multipliziert wird erhält man einen guten Korrekturvektor $|\delta\psi\rangle$. Der Hamiltonian ist im reziproken Raum diagonal dominant, da die Diagonalelemente aufgrund der kinetischen Energie $\sim (\mathbf{k} + \mathbf{G})^2$ quadratisch ansteigen. Andererseits konvergieren die lokalen Pseudopotentiale $V_{loc}(\mathbf{G} - \mathbf{G}')$ schnell im Impulsraum. Als Näherung für den Präkonditionierer wird daher nur die Diagonale der Matrix $(\mathcal{H} - \rho[\psi])$ invertiert. Dies hat den Vorteil, daß der Präkonditionierer einfach zu konstruieren ist und kaum Rechenzeit oder Speicherplatz beansprucht. Insgesamt ergibt sich damit ein Korrekturvektor:

$$c_{\mathbf{k}\nu}^{korr}(\mathbf{G}) = \frac{-1}{\rho[\psi_{\mathbf{k}\nu}] - \mathcal{H}_{\mathbf{k},\mathbf{G}\mathbf{G}}} \left\{ \rho[\psi_{\mathbf{k}\nu}] c_{\mathbf{k}\nu}(\mathbf{G}) - \sum_{\mathbf{G}'} \mathcal{H}_{\mathbf{k},\mathbf{G}\mathbf{G}'} c_{\mathbf{k}\nu}(\mathbf{G}') \right\} . \quad (\text{B.7})$$

Dieser Korrekturvektor muß nun auf den Eigenvektor angewandt werden, um zu einer neuen verbesserten Approximation des Eigenvektors zu gelangen. Das einfachste Vorgehen ist es, den Korrekturvektor, gewichtet mit einer Schrittweite α , zum Startvektor zu addieren:

$$|\psi_\nu^{(j+1)}\rangle = |\psi_\nu^{(j)}\rangle + \alpha |\psi_\nu^{(j),korr}\rangle \quad (\text{B.8})$$

Dieses Verfahren wurde von Car und Parrinello [CP85b] ursprünglich gewählt. Es hat den Nachteil, daß kleine Schrittweiten α zu sehr langsamer Konvergenz führen, zu große α das Verfahren hingegen divergieren lassen.

Daher ist es besser, den Hamiltonian in einem Unterraum zu diagonalisieren, der sowohl die Start-, als auch die Korrekturvektoren enthält. Dazu wird aus den Start- und Korrekturvektoren, z.B. mittels Gram-Schmidt, eine Orthonormalbasis erstellt, der Hamiltonian auf diesen Unterraum projiziert, und dort diagonalisiert. Man kann auf die Orthonormierung verzichten, indem man die Überlappmatrix der Basisvektoren in einem verallgemeinerten Eigenwertproblem berücksichtigt. Der im **ESStCoMPP**-Programm implementierte Algorithmus [Ric93] geht vom Davidson-Algorithmus [Dav75] aus, der jedoch ursprünglich nur für die Bestimmung eines einzelnen Eigenvektors $|\psi\rangle$ konzipiert war. Der Davidson-Algorithmus geht von einem Startvektor zum gesuchten Eigenvektor aus. In jeder Iteration wird ein neuer Korrekturvektor hinzugefügt und die Matrix im jeweils aufgespannten Unterraum diagonalisiert. Spätestens nach $N - 1$ Iterationen entspricht der aufgespannte Unterraum dem gesamten Raum und das Eigenwertproblem ist konvergiert. Die Erfahrung zeigt jedoch, daß die Eigenwerte deutlich schneller konvergieren.

Zur Berechnung mehrerer Eigenvektoren ist es ineffektiv, den Davidson-Algorithmus für jeden Eigenvektor einzeln durchzuführen. Stattdessen wird der Davidson-Algorithmus als Block-Algorithmus ausgeführt [Liu78], bei der von einem Block $|\psi_{\mathbf{k},1..M}\rangle$ von Startvektoren ausgegangen wird, zu dem entsprechend viele Korrekturvektoren $|\delta\psi_{\mathbf{k},1..M}\rangle$ berechnet werden. Dabei kommen in jedem Schritt M Vektoren hinzu. Dies führt rasch zu einem Aufblähen des Unterraumes, welches Rechenkapazitäten und Speicher schnell erschöpfen würde. Nach einer Idee von Kosugi [Kos84] wird die Dimension des Unterraumes daher auf M_{dvd} beschränkt und alte Korrekturen verworfen. Durch Wahl eines größeren Wertes für M_{dvd} läßt sich das Konvergenzverhalten eines Davidson-Schrittes auf Kosten des Speicherbedarfs und der Rechenzeit pro Davidson-Schritt verbessern. Im **ESStCoMPP**-Programm haben sich Werte von $2 \times M$ bis $3 \times M$ für M_{dvd} bewährt, um ein gutes Konvergenzverhalten bei vertretbarer Rechenzeit zu erhalten. Der Davidson Unterraum wird dabei von folgenden Vektoren aufgespannt:

- Den M Lösungen des letzten Davidsonschrittes.
- Den neuen Korrekturvektoren, die sich aus Anwendung des Präkonditionierers auf die Residuen der noch nicht konvergierten Vektoren ergeben.
- Den Lösungsvektoren die in vorherigen Iterationen noch nicht konvergiert waren.

Um die Dimension des Unterraums nicht über M_{dvd} anwachsen zu lassen, werden die ältesten Lösungsvektoren vorheriger Iterationen weggelassen. Am Anfang des Davidsonverfahrens werden die M Lösungsvektoren des letzten elektronischen Selbstkonsistenzschrittes übernommen und bilden die Startvektoren. Die Korrekturvektoren für $\nu > M$ werden mit Einheitsvektoren $\psi_\nu(\mathbf{G}) = \delta_{\mathbf{G}\mathbf{G}_\nu}$ initialisiert. Um im **ESStCoMPP**-Programm elektronische Selbstkonsistenz und das elektronische

Davidson-Kosugi — Iterative Matrixdiagonalisierung in EStCoMPP

Berechne die Startwellenfunktionen $ \psi_\nu^{(0)}\rangle$ und Starteigenwerte $\epsilon_\nu^{(0)}$ für den ersten elektronischen Selbstkonsistenzschritt, indem \mathcal{H} auf einen niedrigdimensionalen Unterraum projiziert und das Eigenwertproblem dort gelöst wird.
Elektronische Selbstkonsistenzschritte
...
Berechne die Diagonalmatrix $(\mathcal{D}_\mathcal{H})_{\mathbf{G}\mathbf{G}'} = \mathcal{H}_{\mathbf{k},\mathbf{G}\mathbf{G}} \delta_{\mathbf{G}\mathbf{G}'}$ aus der Diagonalen des Hamiltonoperators.
Übernehme die M Wellenfunktionen $ \psi_\nu^{(0)}\rangle$ des letzten Selbstkonsistenzschrittes bzw. der Initialisierung.
Berechne $\mathcal{H} \psi_\nu^{(0)}\rangle$ für $\nu = 1 \dots M$.
Initialisiere die Korrekturvektoren $ \psi_\nu^{(0)}\rangle$ mit Einheitsvektoren: $\psi_\nu(\mathbf{G}) = \delta_{\mathbf{G}\mathbf{G}_\nu}$ für $M_{dvd} \geq \nu > M$.
Bis zur Erfüllung des Davidson-Konvergenzkriteriums:
Berechne die Residuen $ \mathcal{R}[\psi_\nu^{(j)}]\rangle = (\mathcal{H} - \epsilon_\nu^{(j)}) \psi_\nu^{(j)}\rangle$.
Unterschreitet die Norm der Residuen $\ \mathcal{R}[\psi_\nu^{(j)}]\ $ eine vorgegebene Schranke, so markiere die entsprechenden Wellenfunktionen $ \psi_\nu^{(j)}\rangle$ als konvergiert.
Berechne zu nicht konvergierten Zuständen die neuen Korrekturvektoren $(\mathcal{D}_\mathcal{H} - \epsilon_\nu^{(j)})^{-1} \mathcal{R}[\psi_\nu^{(j)}]\rangle$ und ersetze damit die ältesten Korrekturvektoren.
Bilde eine Orthonormalbasis $ e_m\rangle$ aus den Vektoren $ \psi_\nu^{(j)}\rangle$ für $\nu = 1 \dots M_{dvd}$. Dazu müssen neue Korrekturvektoren zur alten Eigenvektoren und alten Korrekturvektoren orthonormiert werden.
Berechne $\mathcal{H} \psi_\nu^{(j)}\rangle$ für $\nu = (M + 1) \dots M_{dvd}$.
Projiziere \mathcal{H} auf den Davidson Unterraum durch Bildung von $\mathcal{H}'_{\nu\mu} = \langle \psi_\nu^{(j)} \mathcal{H} \psi_\mu^{(j)} \rangle$.
Löse das M_{dvd} -dimensionale Eigenwertproblem für \mathcal{H}' .
Ersetze die ältesten Korrekturvektoren durch die $ \psi_\nu^{(j)}\rangle$ für $\nu = 1 \dots M$, die nicht als konvergiert markiert wurden.
Übertrage die Lösungsvektoren zu den M niedrigsten Eigenwerten aus dem Unterraum zurück in die Basis der \mathbf{G} -Vektoren und gewinne so $ \psi_\nu^{(j+1)}\rangle$ und $\mathcal{H} \psi_\nu^{(j+1)}\rangle$ für $\nu = 1 \dots M$.
...

Abbildung B.1: Der in EStCoMPP implementierte iterative Eigenwertlöser.

Eigenwertproblem gleichzeitig zur Konvergenz zu bringen, werden pro Selbstkonsistenzschritt nur drei Davidson-Iterationen ausgeführt, indem ein entsprechendes Konvergenzkriterium angegeben wird.

Literaturverzeichnis

- [Ant01] A. Antons. First-Principles Investigation of Initial Stages of Surfactant Mediated Growth on the Si(111) Substrate. Master's thesis, RWTH Aachen, 2001.
- [BBA⁺00] R. Berger, S. Blügel, A. Antons, Wi. Kromen, and K. Schroeder. A Parallelized ab initio Molecular Dynamics Code for the Investigation of Atomistic Growth Processes. In R. Esser, P. Grassberger, J. Grotdorst, and M. Lewerenz, editors, *Molecular Dynamics on parallel Computers, Workshop am John von Neumann Institut für Computing, Jülich, 08-10. Feb. 1999*, pages 185–198, Singapur, Feb. 2000. World Scientific Publishing Co.
- [BHS82] G.B. Bachelet, D.R. Hamann, and M. Schlüter. Pseudopotentials that work: From H to Pu. *Phys. Rev. B*, 26(8):4199, 1982.
- [BK90] D.M. Bylander and L. Kleinman. 4f resonances with norm-conserving pseudopotentials. *Phys. Rev. B*, 41(2):907, 1990.
- [BNLJ92] K. D. Brommer, M. Needels, B.E. Larson, and J. D. Joannopoulos. *Ab Initio* Theory of the Si(111)-(7×7) Surface Reconstruction: A Challenge for Massively Parallel Computation. *Phys. Rev. Lett.*, 68:1355, 1992.
- [CA80] D.M. Ceperley and B.J. Alder. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.*, 45:566, 1980.
- [CP85a] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theorie. *Phys. Rev. Lett.*, 55:2471, 1985.
- [CP85b] R. Car and M. Parrinello. Unified Approach for Molecular Dynamics and Density-Functional Theorie. *Phys. Rev. Lett.*, 55:2471, 1985.
- [Cur85] P. Curie. Sur la formation des cristaux et sur les constantes capillaires de leurs différentes faces. *Bull. soc. min. fr.*, 8:145, 1885.

- [CvP92] L. J. Clarke, I. Štich, and M. C. Payne. Large-scale ab initio total energy calculations on parallel computers. *Comp. Phys. Comm.*, 72:14–28, 1992.
- [Dav75] W.E. Davidson. The iterative calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *J. Comput. Phys.*, 17:87, 1975.
- [DG90] R.M. Dreizler and E.K.U. Gross. *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Springer-Verlag, Berlin, 1990.
- [Din43] A. Dinghas. Über einen geometrischen Satz von Wulff für die Gleichgewichtsform von Kristallen. *Z. f. Kristallogr.*, 105:304, 1943.
- [Eng95] B. Engels. *Ab initio Berechnung der (110) Oberfläche von III-V Halbleitern: Simulation von Rastertunnelmikroskopie- Aufnahmen*. PhD thesis, RWTH Aachen, 1995.
- [Ewa21] P.P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 64:253, 1921.
- [Gib77] J. W. Gibbs. On the equilibrium of heterogeneous substances. *Transactions of the Connecticut Academy*, 3:108–248 and 343–524, 1877.
- [Gib92] J. W. Gibbs. *Thermodynamische Studien*. Engelmann, Leipzig, 1892.
- [GKS90] X. Gonze, P. Käckell, and M. Scheffler. Ghost states for separable, norm-conserving, *ab initio* pseudopotentials. *Phys. Rev. B*, 41(17):12264, 1990.
- [GL76] O. Gunnarsson and B.I. Lundqvist. Local Density Approximation. *Phys. Rev. B*, 13:4274, 1976.
- [GSS91] X. Gonze, R. Stumpf, and M. Scheffler. Analysis of separable potentials. *Phys. Rev. B*, 44(16):8503, 1991.
- [Ham89] D.R. Hamann. Generalized norm-conserving pseudopotentials. *Phys. Rev. B*, 40(5):2980, 1989.
- [HCT+94] M. Horn-von Hoegen, M. Copel, J.C. Tsang, M.C. Reuter, and R.M. Tromp. Surfactant-mediated growth of Ge on Si(111). *Phys. Rev. B*, 50(15):10811–10822, October 1994.
- [HK64] P. Hohenberg and W. Kohn. Inhomogenous Electron Gas. *Phys. Rev. B*, 136(3):864–871, 1964.

- [HSC79] D.R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving Pseudopotentials. *Phys. Rev. Lett.*, 43(20):1494, 1979.
- [IZC79] J. Ihm, A. Zunger, and M.L. Cohen. Momentum-space formalism for the total energy of solids. *J. Phys. C.: Solid State Phys.*, 12:4409, 1979.
- [JG89] R.O. Jones and O. Gunnarson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61(3):689–746, 1989.
- [KB82] L. Kleinman and D.M. Bylander. Efficacious Forms for Model Pseudopotentials. *Phys. Rev. Lett.*, 48(20):1425, 1982.
- [Ker80] G.P. Kerker. Non-singular atomic pseudopotentials for solid state applications. *J. Phys. C.: Solid State Phys.*, 13:L189–94, 1980.
- [Koh85] W. Kohn. Density Functional Theory: Fundamentals and Applications. In M.P. Tosi F. Bassani, F. Fumi, editor, *Highlights of Condensed Matter Theory*. North Holland, Amsterdam, 1985.
- [Kos84] N. Kosugi. Modification of the Liu-Davidson method for obtaining one or simultaneously several eigensolutions of a large real symmetric the preconditioned gradient-type iterative methods in a subspace for partial generalized symmetric eigenvalue problem. *J. Comput. Phys.*, 55(3):426–436, 1984.
- [KS65] W. Kohn and L.J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140(4A):A1133, 1965.
- [Lev79] M. Levy. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solutions of the v -representability problem. *Proc. Natl. Acad. Sci. USA*, 76:6062, 1979.
- [LFC82] S.G. Louie, S. Froyen, and M.L. Cohen. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B*, 26(4):1738, 1982.
- [Lie14] H. Liebmann. Der Curie-Wulff'sche Satz über Combinationsformen von Krystallen. *Z. f. Krystallogr.*, 53:171, 1914.
- [Liu78] B. Liu. Numerical Algorithms in Chemistry: Algebraic methods. *Report on Workshop*, 1978.
- [MV79] A.H. MacDonald and S.H. Vosko. A relativistic density functional formalism. *J. Phys. C.*, 12:2977, 1979.

- [MVA92] G. Meyer, B. Voigtländer, and N. M. Amer. Scanning tunneling microscopy of surfactant-mediated epitaxy of Ge on Si(111): strain relief mechanisms and growth kinetics. *Surf. Sci. Lett.*, 274:L541–L545, 1992.
- [Nee93] M. Needels. Comment on “*Ab Initio* Theory of the si(111)-(7×7) Surface Reconstruction”. *Phys. Rev. Lett.*, 71:3612, 1993.
- [NPS93] J. S. Nelson, S. J. Plimpton, and M. P. Sears. Plane-wave electronic-structure calculations on a parallel supercomputer. *Phys. Rev. B*, 47(4):1765–1774, 1993.
- [OEK99] K. Ohno, K. Esfarjani, and Y. Kawazoe. *Computational Material Science*. Springer Series in Solid State Sciences. Springer, 1999.
- [PCV⁺92] John P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671, 1992.
- [Per86] P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B*, 33:8822, 1986.
- [PTA⁺92] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, and J.D. Joannopoulos. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Computer Physics Reports*, 64(4):1045–1097, 1992.
- [PW86] P. Perdew and Y. Wang. Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation. *Phys. Rev. B*, 33:8800, 1986.
- [Ric93] P. Richard. Iterative Lösung der Schrödingergleichung für große Systeme. Master’s thesis, RWTH Aachen, 1993.
- [RRKJ90] A.M. Rappe, K.M. Rabe, E. Kaxiras, and J.D. Joannopoulos. Optimized pseudopotentials. *Phys. Rev. B*, 41(2):1227, 1990.
- [SAB⁺00] K. Schroeder, A. Antons, R. Berger, W. Kromen, and S. Blügel. Surface Diffusion and Models for the Kinetics of Epitaxial Growth. *Proc. Int. Sym. Str. Dyn. Het. Sys.*, pages 71–94, February 2000.
- [SABB02] K. Schroeder, A. Antons, R. Berger, and S. Blügel. Surfactant Mediated Heteroepitaxy versus Homoepitaxy: Kinetics for Group-IV Adatoms on As-Passivated Si(111) and Ge(111). *Phys. Rev. Lett.*, 88:46101, Jan. 2002.

- [SERB98] K. Schroeder, B. Engels, P. Richard, and S. Blügel. Reexchanged Controlled Diffusion in Surfactant-Mediated Epitaxial Growth: Si on As-Terminated Si(111). *Phys. Rev. Lett.*, 80(13):2873–2876, March 1998.
- [SPKSL92] I. Stich, M.C. Payne, R.D. King-Smith, and J-S. Lin. *Ab initio* Total-Energy Calculations for Extremely Large Systems: Application to the Takayanagi Reconstruction of Si(111). *Phys. Rev. Lett.*, 68(9):1351–1358, March 1992.
- [SSO92] M. Saito, O. Sugino, and A. Oshiyama. Real-space-partitioned separable pseudopotential. *Phys. Rev. B*, 46(4):2606, 1992.
- [TM90] N. Troullier and J.L. Martins. A straightforward method for generating soft transferable pseudopotentials. *Solid State Commun.*, 74(7):613, 1990.
- [TM91] N. Troullier and J.L. Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43(3):1993, 1991.
- [TTTT85] K. Takayanagi, Y. Tanishiro, S. Takahashi, and M. Takahashi. Structure Analysis of Si(111)- 7×7 reconstructed Surface by Transmission Electron Diffraction. *Surf. Sci.*, 164(367), 1985.
- [Van85] D. Vanderbilt. Optimally smooth norm-conserving pseudopotentials. *Phys. Rev. B*, 32(12):8412, 1985.
- [vL43] M. v. Laue. Der Wulffsche Satz für die Gleichgewichtsform von Kristallen. *Z. f. Kristallogr.*, 105:124, 1943.
- [Voi] B. Voigtländer. Private Mitteilung.
- [Voi01] B. Voigtländer. Fundamental processes in Si/Si and Ge/Si epitaxy studies by scanning tunneling microscopy during growth. *Surf. Sci. Rep.*, 43(5-8):127–254, 2001.
- [VS92] J. Vackar and A. Simunek. Phase-Shift Technique for Generating Generalized Norm-Conserving Semilocal Pseudopotentials. *Solid State Commun.*, 81:837, 1992.
- [VWN80] S.H. Vosko, L. Wilk, and N. Nussair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58:1200, 1980.
- [VZ93] B. Voigtländer and A. Zinner. Influence of surfactants on the growth-kinetics of Si on Si(111). *Surf. Sci. Lett.*, 292:775–780, 1993.

- [VZ96] B. Voigtländer and A. Zinner. Structure of the Stranski-Krastanov layer in surfactant-mediated Sb/Ge/Si(111) epitaxy. *Surf. Sci.*, 351:L233–L238, 1996.
- [VZWB95] B. Voigtländer, A. Zinner, T. Weber, and H.P. Bonzel. Modification of growth kinetics in surfactant-mediated epitaxy. *Phys. Rev. B*, 51(12):7583–7591, March 1995.
- [WJ94] J. Wiggs and H. Jónsson. A parallel implementation of the Car-Parinello method by orbital decomposition. *Comp. Phys. Comm.*, 81:1–18, 1994.
- [WJ95] J. Wiggs and H. Jónsson. A hybrid decomposition parallel implementation of the Car-Parinello method. *Comp. Phys. Comm.*, 87:319–340, 1995.
- [Wul01] G. Wulff. Zur Frage der Geschwindigkeit des Wachstums und der Auflösung von Krystallflächen. *Z. f. Krystallogr.*, 34:449, 1901.

Danksagung

An dieser Stelle möchte ich all denen danken, die zum Gelingen dieser Arbeit beigetragen haben:

Vor allem möchte ich meinem Betreuer Prof. Dr. Kurt Schroeder für die maßgebliche Unterstützung dieser Arbeit danken. Besonders in den schwierigen Phasen dieser Arbeit habe ich von der Geduld, mit der er sich meinen Problemen widmete und der motivierenden Arbeitsatmosphäre, die er verbreitete, profitiert.

Auch Prof. Dr. Stefan Blügel möchte ich besonderen Dank aussprechen für die interessante Aufgabe der Parallelisierung des Programmes, die Einsichten, die er mir während meiner Arbeit daran vermittelte und für viele erhellende und offene Dialoge.

Prof. Dr. Stefan Blügel und Prof. Dr. Kurt Schroeder danke ich beiden sehr für die Freundlichkeit, mit der sie mich in ihrer Arbeitsgruppe aufnahmen und durch diese Arbeit begleiteten und für das Engagement, mit dem sie meine Ausbildung förderten.

Herrn Prof. Dr. H. Müller-Krumbhaar danke ich vielmals dafür, daß er mir die Möglichkeit gab, diese Doktorarbeit an seinem Institut zu erstellen und für die gemeinschaftliche Arbeitsatmosphäre dort, für die er stets sorgt.

Dr. Armin Antons möchte ich sehr für seine enge Kooperation und Hilfsbereitschaft während des letzten Teils meiner Arbeit danken.

Dr. Bert Voigtländer danke ich für die mir zur Verfügung gestellten STM-Bilder und sein Zuvorkommen bei der Erörterung meiner diesbezüglichen Fragen.

Dr. Reiner Vogelsang danke ich für die schnelle und zuverlässige Hilfe bei der Klärung der Probleme, die bei der Parallelisierung auftraten.

Dr. Winfried Kromen und Dipl. Phys. Diedrich Frielinghaus möchte ich für die informativen Gespräche und die angenehme Zusammenarbeit, besonders während des ersten Teils meiner Arbeit, danken.

Dr. Peter Richard und Dr. Benedikt Engels danke ich für die Mühe, die sie sich gaben, mich mit dem Programm vertraut zu machen und all meine Fragen dazu zu beantworten.

Nicolae Atodiresei möchte ich für seine Unterstützung bei der Berechnung einiger Referenzdaten danken.

Annemarie Grabus und Michael Krohm möchte ich an dieser Stelle für das Korrekturlesen dieser Arbeit danken.

Bei allen Mitarbeitern des Instituts Theorie III, auch denen die hier nicht genannt wurden, möchte ich mich herzlich für die freundschaftliche Arbeitsatmosphäre bedanken.

Von ganzem Herzen danke ich meiner Mutter, die mir meine Ausbildung und letztlich mein Studium ermöglichte, mich in all den Jahren unterstützte und immer für mich da war, wenn ich sie brauchte.

Lebenslauf

persönliche Daten

Name: Ralf Berger
Geburtsdatum: 4. April 1968
Geburtsort: Warstein
Staatsangehörigkeit: deutsch
Familienstand: ledig

Schulbildung

Aug. '74 - Juni '78 Martin-Luther-Schule, Düren
Aug. '78 - Mai '87 Gymnasium am Wirteltor, Düren
Mai '87 Abitur

Wehrdienst

Juli '87 - Sep. '88 Fernmelder, Nörvenich

Studium

Okt. '88 Immatrikulation an der RWTH-Aachen
Okt. '90 Vordiplom in Physik
Apr. '93 - Nov. '94 Diplomarbeit bei Prof. Dr. Vollhardt, Thema:
*“Das Phasendiagramm des Hubbard-Modells
in einer dynamischen Mean-Field Theorie”*
Okt. '92 - Feb. '93 Übungsbetreuung Quantenmechanik
Okt. '93 - Feb. '94 Übungsbetreuung Quantenmechanik
Mai '94 Benutzerbetreuung im Rechner-Pool
Apr. '95 - Juni '95 Übungsbetreuung Elektrodynamik
Dez. '95 Diplom in Physik
Jan. '96 Beginn der Promotion bei Prof. Dr. Schroeder
am Institut für Festkörperforschung
im Forschungszentrum Jülich
Sep. '02 Promotionsprüfung