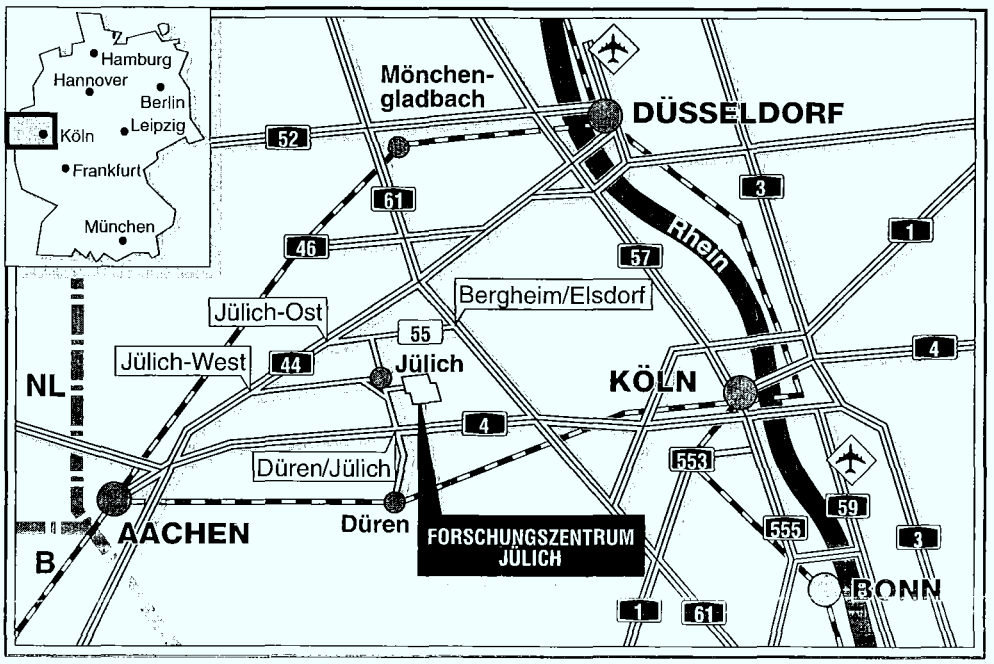
FORSCHUNGSZENTRUM JÜLICH GmbH



Institut für Festkörperforschung

Untersuchungen der strukturellen Eigenschaften von InGaAs/GaAs-Übergittern und zur Rekonstruktion des projizierten Kristallpotentials in der hochauflösenden Elektronenmikroskopie

Markus Lentzen



Berichte des Forschungszentrums Jülich ; 3324
ISSN 0944-2952
Institut für Festkörperforschung Jül-3324
D82 (Diss. RWTH Aachen)

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek
D-52425 Jülich · Bundesrepublik Deutschland
Telefon: 02461/61-6102 · Telefax: 02461/61-6103 · Telex: 833556-70 kfa d

Untersuchungen der strukturellen Eigenschaften von InGaAs/GaAs-Übergittern und zur Rekonstruktion des projizierten Kristallpotentials in der hochauflösenden Elektronenmikroskopie

Markus Lentzen

Inhaltsverzeichnis

I Struktureigenschaften von InGaAs/GaAs-Übergittern 9

Zusammenfassung Teil I 11

1 Einleitung 14

2 Relaxationsmodelle 19
2.1 Grundlagen der Elastiziätstheorie 19
2.2 Ableitung von Relaxationsmodellen 21
2.3 Homogen deformierte Einzelschicht 22
2.4 Verzerrungsfeld einer Versetzung 22
2.4.1 Versetzung im unendlichen Kristall 26
2.4.2 Versetzung nahe einer Kristalloberfläche 27
2.5 Gleichgewichtszustand einer Einzelschicht 28
2.6 Gleichgewichtszustand eines Übergitters 33
2.6.1 Kohärentes Übergitter 33
2.6.2 Relaxation freier Übergitter 35
2.6.3 Endliche Übergitter 38
2.7 Kinetik der Relaxation 41
2.7.1 Versetzungserzeugung 41
2.7.2 Versetzungsmultiplikation 43
2.7.3 Bewegung von Versetzungen 47
2.7.4 Orientierung von Versetzungen 48
2.7.5 Ergänzungen der Gleichgewichtsmodelle 48
2.8 Äquivalenz der Energiemethode und des Kräftegleichgewichts 50

3 Experimentelle Methoden 52
3.1 Probenherstellung 52
3.2 Transmissionselektronenmikroskopie 54
3.2.1 Probenpräparation 54
3.2.2 Mikroskopische Abbildung von Versetzungen 55

4 Charakterisierung der Übergitter 58
4.1 Orthogonale Versetzungsnetzwerke 58
4.1.1 Experimentelle Resultate 58
4.1.2 Diskussion 61

4.2 Gleitfähige 60°-Versetzungen 61
4.2.1 Experimentelle Resultate 61
4.2.2 Diskussion 63
4.3 Versetzungsdipole 65
4.3.1 Experimentelle Resultate 65
4.3.2 Diskussion 69
4.4 Relaxation durch Versetzungsdipole 71
4.4.1 Experimentelle Resultate 71
4.4.2 Diskussion 71
4.5 Wechsel von Gleitebenen 73
4.5.1 Experimentelle Resultate 73
4.5.2 Diskussion 77
4.6 Superkinken 78
4.6.1 Experimentelles Resultat 78
4.6.2 Diskussion 81
4.7 Versetzungen im Substrat 81
4.7.1 Experimentelle Resultate 81
4.7.2 Diskussion 88
4.8 Plan-View-Streifenkontrast der Übergitter 89
4.8.1 Experimentelles Resultat 89
4.8.2 Diskussion 91
4.9 Versetzungsdichten 93
4.9.1 Experimentelle Resultate 93
4.10 Zusammenfassung der experimentellen Beobachtungen 100

5 Partielle Relaxation von Übergittern 102
5.1 Bildkräfte bei der Versetzungserzeugung 102
5.2 Bildkräfte bei der Multiplikation 104
5.3 Elastische Verspannung des Substrats 104
5.4 Wechselwirkung Versetzung–Netzwerk 105
5.5 Einzelschicht-Barrieren 107
5.5.1 Einsetzen der Relaxation 108
5.5.2 Gleiten bei partieller Relaxation 108
5.5.3 Kinetische Energie einer Versetzung 111
5.5.4 Der Doppelkinkenmechanismus 114
5.6 Der Einfluß der Wachstumssequenz 115
5.6.1 Unterkritische GaAs-Schicht 116
5.6.2 Überkritische GaAs-Schicht 117
5.6.3 Kinetische Barrieren 119
5.6.4 Wachstumssequenz bei kleiner GaAs-Einzelschichtdicke 119
5.6.5 Partielle Relaxation während der Wachstumssequenz 120
5.6.6 Experimenteller Beleg 121
5.7 Zusammenfassung der Diskussion 122

II Rekonstruktion des projizierten Potentials 125

6 Zusammenfassung und Einleitung Teil II 127

7 Verschlüsselung der Strukturinformation 130
7.1 Die Blochwellenmethode 130
7.1.1 Nichtlinearität der dynamischen Beugung 135
7.1.2 Interpretierbarkeit des schwachen Phasenobjekts 135
7.2 Der Multislice-Algorithmus 136
7.3 Modell der Kontrastübertragung 138
7.3.1 Idealer Kontrasttransfer 139
7.3.2 Realer Kontrasttransfer 140
7.3.3 Die Raumfrequenzabhängigkeit der Aberrationsfunktion . . . 140
7.3.4 Die Gestalt der Dämpfungsterme 143
7.3.5 Verschlüsselung der Austrittswellenfunktion 147
7.4 Zusammenfassung 149

8 Existierende Strukturermittlungsverfahren 150
8.1 Konventioneller Bildvergleich 150
8.2 Quantitativer Bildvergleich 151
8.2.1 Kleinstquadrat-Kriterium 151
8.2.2 Kreuzkorrelationskoeffizient 152
8.2.3 Vergleichskriterien beim schwachen Objekt 153
8.3 Automatisierter Bildvergleich 154
8.3.1 Bestimmung der Beugungs- und Abbildungsparameter 154
8.3.2 Strukturermittlung 155
8.4 Rekonstruktionsmethoden 156
8.4.1 Phasenrekonstruktionsverfahren 156
8.4.2 Potentialrekonstruktionsverfahren 158

9 Theorie der stochastischen Rekonstruktionsverfahren 161
9.1 Vergleich von Austrittswellenfunktionen 163
9.2 Simulated Annealing 164
9.3 Genetische Algorithmen 167

10 Anwendung stochastischer Phasenrekonstruktionsverfahren 169
10.1 Anwendung auf eine simulierte Defokusserie 169
10.1.1 Rekonstruktion mittels Simulated Annealing 171
10.1.2 Rekonstruktion mittels eines genetischen Algorithmus 176
10.2 Anwendung auf eine experimentelle Defokusserie 179
10.2.1 Aufbereiten der experimentellen Defokusserie 179
10.2.2 Rekonstruktion der Austrittswellenfunktion 180
10.3 Diskussion 183

11 Anwendung der stochastischen Potentialrekonstruktion **189**
11.1 Simulation der Austrittswellenfunktionen von GaAs[110] 189
11.2 Rekonstruktion des projizierten Kristallpotentials 192
11.3 Diskussion 197

A Theorie der Zweistrahlbeugung **203**

Abbildungsverzeichnis

1.1 Relaxationsmoden der Übergitterrelaxation 16
1.2 Wachstumsmodi des Schichtwachstums 17

2.1 Burgersumlauf einer Stufenversetzung 24
2.2 Burgersumlauf einer Schraubenversetzung 25
2.3 Dehnungen im kohärenten Übergitter 34
2.4 Eingeschobene Ebene einer relaxierten Einzelschicht 35
2.5 Abnahme der Restdehnung eines Übergitters während der Relaxation 40
2.6 Wachstum einer Versetzungshalbschleife 44
2.7 Gleiten von Fehlpassungssegmenten 45
2.8 Hagen-Strunk-Multiplikation von Fehlpassungsversetzungen 46

4.1 Projektion der Übergitternetzwerke 59
4.2 Projektion der Übergitternetzwerke und des Substratnetzwerks 60
4.3 Schnittlinie der Substratgrenzfläche und der Probenunterseite 62
4.4 Kontrastauslöschung von Versetzungslinien 64
4.5 „Dunkel-Hell-Dunkel"-Kontrast an Versetzungsdipolen 66
4.6 Stereopaar von Hellfeldabbildungen eines Dipols zweier Fehlpassungsversetzungen 67
4.7 Stereopaar von Hellfeldabbildungen zweier benachbarter Dipole von Fehlpassungsversetzungen 68
4.8 Haarnadelkonfiguration zweier paralleler Fehlpassungssegmente 70
4.9 Abgleiten der Fadensegmente von Haarnadelkonfigurationen 72
4.10 Rechtwinkliges Abknicken einer Fehlpassungsversetzung 74
4.11 Mehrfaches rechtwinkliges Abknicken einer Fehlpassungsversetzung . 75
4.12 Doppel-U-Konfiguration von abknickenden Fehlpassungsversetzungen 76
4.13 Superkinke zwischen zwei Fehlpassungssegmenten 79
4.14 Superkinke zwischen zwei Fehlpassungssegmenten 80
4.15 Versetzungsschleifen unterhalb der Substratgrenzfläche 82
4.16 Versetzungsschleifen unterhalb der Substratgrenzfläche 82
4.17 Versetzungsschleifen unterhalb der Substratgrenzfläche 83
4.18 Versetzungsschleifen unterhalb der Substratgrenzfläche 84
4.19 Fehlpassungssegment mit langem Fadensegment 85
4.20 Lange, geradlinige Versetzung im Substrat 86
4.21 Stapelfehler an der Substratgrenzfläche 87
4.22 Übergitter-Streifenkontrast in Plan-View-Abbildungen 90

4.23 Übergitter-Streifenkontrast in Plan-View-Abbildungen 92
4.24 Seitenansicht von Fehlpassungsnetzwerken an den Substratgrenzflächen 94
4.25 Seitenansicht eines (18/6)-Übergitters 95
4.26 Seitenansicht eines (18/2)-Übergitters 96
4.27 Abhängigkeit des relaxierten Anteils der Fehlpassung von der GaAs-Einzelschichtdicke 98
4.28 Abhängigkeit der homogenen Restdehnung von der GaAs-Einzelschichtdicke 99

5.1 Peach-Koehler-Kraft eines Netzwerks auf eine 60°-Versetzung 106
5.2 Fehlpassung und relaxierter Anteil der Fehlpassung eines (18/2)-Übergitters 117
5.3 Fehlpassung und relaxierter Anteil der Fehlpassung eines (18/10)-Übergitters 118
5.4 Fehlpassung und relaxierter Anteil der Fehlpassung eines (18/25)-Übergitters 118

9.1 Kompakte Darstellung des Simulated-Annealing-Algorithmus 166
9.2 Kompakte Darstellung des genetischen Suchalgorithmus 168

10.1 Projektion der $D1_a$-Struktur von Ni_4Mo 170
10.2 Simulierte Eingabebilder von Ni_4Mo 172
10.3 Austrittswellenfunktionen von Ni_4Mo für eine Probendicke von 25 nm 173
10.4 Konvergenzplots der Simulated-Annealing-Rekonstruktion 175
10.5 Konvergenzplots des genetischen Rekonstruktionsalgorithmus 178
10.6 Aufbereitung zweier Bilder einer Defokusserie von Ni_4Mo 181
10.7 Rekonstruktion aus einer experimentellen Defokusserie 182

11.1 Projektion der Kristallstruktur von GaAs 190
11.2 Simulierte Austrittswellenfuktionen von GaAs 191
11.3 Phase der Transmissionsfunktion von GaAs 193
11.4 Rekonstruierte Potentiale von GaAs 194
11.5 Konvergenzplot des Simulated-Annealing-Algorithmus 198

Teil I

Strukturelle Eigenschaften von InGaAs/GaAs-Übergittern

Zusammenfassung Teil I

Elektronische und optoelektronische Bauelementstrukturen auf der Basis von Gruppe-IV-Halbleitern, III-V- und II-VI-Verbindungshalbleitern bestehen häufig aus epitaktischen Schichtsystemen mit zueinander kompatibler Diamant- oder Zinkblendestruktur. Die mechanischen Eigenschaften der verwendeten Materialien, insbesondere die Gitterkonstanten, können sich so drastisch unterscheiden, daß einzelne Kristallbereiche zum Teil erheblich mechanisch verzerrt sind.

Die Verzerrung ist mit einer inneren Spannung verknüpft, die die freie Energie des Systems um den Betrag der mechanischen Energie des Verzerrungsfeldes erhöht. Oberhalb eines bestimmten Schwellenwertes wird die Verspannung in Abhängigkeit von der Geometrie einer Bauelementstruktur durch Defekterzeugung relaxiert, und die freie Energie so erniedrigt. Der Einbau einer zu hohen Defektdichte ist in praktisch allen Bauelementanwendungen unerwünscht, weil sie die elektronischen und optoelektronischen Eigenschaften verschlechtert.

Epitaxieexperimente, mikrostrukturelle Charakterisierung von epitaktisch gewachsenen Halbleiterschichten und die Entwicklung und Verfeinerung von Relaxationsmodellen verfolgen das Ziel, eine Reduktion der Defektdichte oder eine Defektvermeidung bei der Bauelementherstellung zu erreichen.

In dieser Arbeit werden die strukturellen Eigenschaften eines epitaktischen Modellsystems untersucht, das aus einem $In_xGa_{1-x}As/GaAs$-Übergitter besteht, welches mittels Molekularstrahlepitaxie auf ein GaAs(001)-Substrat gewachsen wurde. In fehlgepaßten Übergittern, die aus einer periodischen Abfolge von Einzelschichten bestehen, gibt es zwei verschiedene Relaxationsmoden [123, 126]. Einerseits können die Einzelschichten einer Übergitterperiode durch den Einbau von Fehlpassungsversetzungen gegeneinander relaxieren, andererseits kann ein Übergitter durch den Einbau von Fehlpassungsversetzungen an der Substratgrenzfläche als ganzes relativ zum Substrat relaxieren.

Das Einsetzen und das Ausmaß beider Relaxationsprozesse hängt von der Fehlpassung zwischen InGaAs- und GaAs-Schichten, den Einzelschichtdicken, der Gesamtdicke des Übergitters und der Indiumkonzentration x einer $In_xGa_{1-x}As$-Schicht ab.

Die beiden Relaxationsmoden wurden durch die Variation der GaAs-Einzelschichtdicke und Konstanthalten der InGaAs-Schichtdicke untersucht. Die Indiumkomposition betrug $x = 20$ %, so daß die Fehlpassung der InGaAs-Einzelschichten den Wert 1.4 % annahm. Bei dieser Komposition herrscht der zweidimensionale Wachstumsmodus vor, der die Voraussetzung für die Herstellung strukturell relativ perfekter Übergitter ist. Die Gesamtdicke der Übergitter wurde in der Probense-

rie konstant gehalten, so daß die mittlere Fehlpassung der Übergitter wegen des unterschiedlichen GaAs/InGaAs-Schichtdickenverhältnisses ebenfalls variierte.

Die mikrostrukturelle Charakterisierung der Übergitter erfolgte durch konventionelle und hochauflösende Transmissionselektronenmikroskopie. Die anschließende Bestimmung der Grenzflächendefektdichten und der Verspannungszustände wurde mit den Voraussagen bestehender Relaxationsmodelle [123, 126] verglichen, die für die Schichtdickenabfolgen der untersuchten Übergitter eine vollständige relative Verspannung der Einzelschichten der Übergitterperioden, und eine vollständige Relaxation der gesamten Übergitter relativ zum Substrat voraussagen.

Es wird experimentell beobachtet, daß alle Übergitter nur partiell relativ zum Substrat relaxiert sind. Die Grenzflächen der Übergitterperioden sind nicht defektfrei und sie enthalten ebenfalls Fehlpassungsversetzungen. Die Fehlpassungsnetzwerke bestehen aus zwei zueinander orthogonalen Scharen geradliniger Segmente, mit Linienrichtungen $\vec{u} = [110], [\bar{1}10]$, und sie sind vom gleitfähigen 60°-Versetzungstyp. Der Substratbereich direkt unterhalb der Substratgrenzfläche, der nach den Modellvorstellungen defektfrei sein sollte, enthält Versetzungshalbschleifen.

Dieser deutliche Widerspruch zwischen Beobachtungen und Relaxationsmodellen wurde auf den Einfluß der parallel zum epitaktischen Schichtwachstum ablaufenden kinetischen Teilprozesse der Relaxation, der Erzeugung, der Multiplikation und des Gleitens von Versetzungen zurückgeführt.

Es wurde anhand einfacher Modellerweiterungen untersucht, in welchem Ausmaß jeder dieser möglicherweise nur unvollständig ablaufenden Teilprozesse zu einer nur partiellen Relaxation eines Übergitters beitragen kann. Es konnte insbesondere mit Hilfe der Analyse einzelner charakteristischer Versetzungskonfigurationen wie Halbschleifen, Versetzungsdipolen, Haarnadelkonfigurationen, Superkinken, Fadenversetzungen und rechtwinklig abknickenden Fehlpassungssegmenten für jeden plausiblen Mechanismus, der während der einzelnen Teilprozesse ablaufen kann, entweder ein experimenteller Beleg oder eine Widerlegung gefunden werden.

Eine wichtige Ursache für die partielle Relaxation ist eine mögliche Hemmung der Versetzungserzeugung und der Versetzungsmultiplikation, wenn die treibende Kraft, die von der Restdehnung der Einzelschichten abhängt, klein ist. Diese Schlußfolgerung stimmt mit analogen Betrachtungen überein, die schon in der Literatur über fehlgepaßte Einzelschichten durchgeführt wurden [129, 127, 121, 83, 141, 142, 69, 31, 32].

Die wichtigste Ursache der partiellen Relaxation hat kein Analogon bei fehlgepaßten Einzelschichten, und sie ist einzigartig für die Übergitterrelaxation: Nachdem die Relaxation oberhalb einer kritischen Gesamtdicke eines Übergitters eingesetzt hat, ist eine bestimmte Zahl von Versetzungen zur Substratgrenzfläche geglitten, und diese haben dadurch die InGaAs-Schichten relaxiert, jedoch die vorher unverzerrten GaAs-Schichten gleichzeitig verspannt. Die GaAs-Einzelschichten der Übergitterperioden wirken als Barrieren für das weitere Gleiten von Fehlpassungssegmenten, und ab einem bestimmten Relaxationsgrad wird die gesamte Relaxation effektiv gehemmt.

Die Existenz dieses Mechanismus konnte experimentell belegt werden, denn der resultierende maximale Relaxationsgrad eines Übergitters hat nach der Modellvor-

stellung, die auf einem Doppelkinkenmechanismus an der GaAs-Barriere beruht, eine charakteristische $1/h_{GaAs}$-Abhängigkeit von der GaAs-Einzelschichtdicke, die im Rahmen der Meßfehler mit den Meßwerten der Versetzungsdichten übereinstimmt. Das Gleiten einer Versetzung über die GaAs-Einzelschichtbarrieren ist jedoch mit einer hohen Aktivierungsenergie verknüpft.

Andere Einflüsse, wie die Wechselwirkung einer gleitenden Versetzung mit einem schon existierenden Fehlpassungsnetzwerk, oder einer möglichen lokalen Substratverspannung können in der Größenordnung gegenüber einer Behinderung der Versetzungserzeugung und der Multiplikation und gegenüber der Größe der GaAs-Einzelschichtbarriere vernachlässigt werden.

Kapitel 1

Einleitung

In den letzten Jahrzehnten haben Halbleiterbauelemente, deren Herstellung auf Kombinationen von Gruppe-IV-Elementen, III-V-Verbindungshalbleitern und II-VI-Verbindungshalbleitern, sowie ihren ternären und quaternären Verbindungen beruht, neuartige technologische Perspektiven eröffnet und eine Reihe materialwissenschaftlicher Fragestellungen aufgeworfen. Das technologische Potential beruht darauf, daß die elektrischen und optoelektronischen Eigenschaften vieler Verbindungshalbleiter durch die Existenz einer direkten Bandlücke bestimmt werden, und einige Verbindungshalbleiter eine sehr hohe Ladungsträgerbeweglichkeit aufweisen [117]. Das Maßschneidern elektrischer Eigenschaften, sowie das Erreichen einer hohen Integrationsdichte von Bauelementen erfordert fast ausnahmslos eine strukturelle Perfektion auf sehr kleinen Längenskalen, bis hinunter auf die atomare Ebene.

Ein wichtiger Elementarprozeß bei der Bauelementherstellung ist das Aufwachsen eines kristallinen Materials auf ein anderes kristallines Material, wobei die aufwachsende Schicht die kristallographische Orientierung des Substrats übernimmt. Diese *Heteroepitaxie* zweier verschiedener Materialien ist technologisch in einer Reihe von Verfahren realisiert, wie der Molekularstrahlepitaxie (*molecular beam epitaxy*, MBE), der Flüssigphasenepitaxie (*liquid phase epitaxy*, LPE), der Deposition aus der Gasphase (*chemical vapour deposition*, CVD), sowie Hybridverfahren, die nach der Zusammensetzung der Quellen oder nach der Art der Prozeßführung benannt sind: Molekularstrahlepitaxie mit metallorganischen Verbindungen (*metal organic molecular beam epitaxy*, MOMBE), Niedrigdruck-Gasphasendeposition (*low pressure chemical vapour deposition*, LPCVD), und so fort.

Die grundlegende Problematik der Halbleiter-Heteroepitaxie liegt in den oft sehr unterschiedlichen strukturellen und mechanischen Eigenschaften der beteiligten Materialien, wie zum Beispiel der Kristallstruktur, der Gitterkonstante, des Schubmoduls und des thermischen Ausdehnungskoeffizienten. In einer Vielzahl von Fällen ist die Kristallstruktur der beteiligten Materialien jedoch gleich, nämlich vom Diamant- oder Zinkblende-Typ. In diesem Fall wächst die Schicht in der kristallographischen Orientierung auf, die durch die Substratoberfläche vorgegeben wird.

Häufig unterscheiden sich jedoch die Gitterkonstanten der beiden Materialien voneinander, so daß ein defektfreies, gitterangepaßtes Wachstum nur dann möglich ist, wenn das aufwachsende Material die Gitterkonstante des Substratmaterials an-

nimmt. Die dabei auftretende elastische Verzerrung der aufwachsenden Schicht ist mit einer inneren Spannung verknüpft. Dadurch ist die freie Energie des Kristalls um den Betrag der Energie des Verzerrungsfeldes erhöht [109].

Der Gitterkonstantenunterschied wird oberhalb einer bestimmten Schichtdicke durch einen *Relaxationsprozeß* reduziert, so daß die freie Energie erniedrigt wird. Die Relaxation erfolgt durch den Einbau von Kristalldefekten, deren Energie niedriger als die des ursprünglichen Verzerrungsfeldes ist. Der Defekttyp und die Defektdichte sind sehr stark abhängig vom ursprünglichen Gitterkonstantenunterschied und von den gegebenen geometrischen Abmessungen, wie zum Beispiel der Schichtdicke des aufwachsenden Materials oder der lateralen Ausdehnung einer Bauelementstruktur.

In praktisch allen Bauelementanwendungen sind zu große Defektdichten unerwünscht, da sie die elektrischen Eigenschaften empfindlich beeinträchtigen können, bis hin zum Totalausfall eines Bauelementes [154, 132, 169]. Daraus erwächst erstens die Notwendigkeit, epitaktisch gewachsene Halbleiterschichten mikrostrukturell zu charakterisieren, zweitens, den Defekteinbau durch geeignete Relaxationsmodelle zu beschreiben, und drittens, durch Epitaxieexperimente und Modellverfeinerungen Kriterien zu finden, die eine Defektreduzierung oder eine Defektvermeidung ermöglichen.

In dieser Arbeit werden die strukturellen Eigenschaften eines Modellsystems untersucht, das aus einem InGaAs/GaAs-Übergitter besteht, welches mittels Molekularstrahlepitaxie (Kapitel 3) auf ein GaAs(001)-Substrat gewachsen wurde. In fehlgepaßten Übergittern, die aus einer periodisch fortgesetzten Abfolge von Schichten bestehen, können zwei verschiedene Relaxationsmoden unterschieden werden [123, 126]. Erstens kann der relative Gitterkonstantenunterschied, die sogenannte *Fehlpassung*, zwischen den einzelnen fehlgepaßten Schichten durch Defekterzeugung abgebaut werden. Zweitens kann das Übergitter insgesamt relativ zum Substrat relaxieren. Das Einsetzen und das Ausmaß beider Relaxationsmoden ist empfindlich abhängig von der ursprünglichen Fehlpassung zwischen den Einzelschichten, der mittleren Fehlpassung des gesamten Übergitters relativ zum Substrat, den jeweiligen GaAs- und InGaAs-Einzelschichtdicken, sowie der Gesamtdicke des Übergitters. Abbildung 1.1 (A) zeigt schematisch das Resultat des ersten Relaxationsmodus, der durch Defekteinbau an den Übergittergrenzflächen gekennzeichnet ist, und Abbildung 1.1 (B) das Resultat des zweiten Relaxationsmodus, der durch Defekteinbau an der Substratgrenzfläche charakterisiert ist.

Die beiden Relaxationsmoden wurden durch Variation der Einzelschichtdicken untersucht, wobei die Fehlpassung der $In_xGa_{1-x}As$-Einzelschichten durch die Wahl der Indiumkonzentration von $x = 20$ % in allen Proben konstant auf 1.4 % gehalten wurde. In diesem Kompositions- und Fehlpassungsbereich ist das Wachstum durch den für die Herstellung strukturell relativ perfekter Übergitter notwendigen *zweidimensionalen Wachstumsmodus* gekennzeichnet, der oberhalb einer kritischen Fehlpassung von etwa 2.0 % durch dreidimensionales Inselwachstum bzw. durch eine Mischform aus beiden, das Stranski-Krastanow-Wachstum [166], abgelöst wird (Abbildung 1.2). Das Wachstum bei der in diesem System maximalen Fehlpassung von 7.1 %, im Falle reiner InAs-Schichten, ist im Anfangsstadium durch Inselwachstum und sehr hohe Defektdichten gekennzeichnet. Das gleichzeitige Konstanthalten

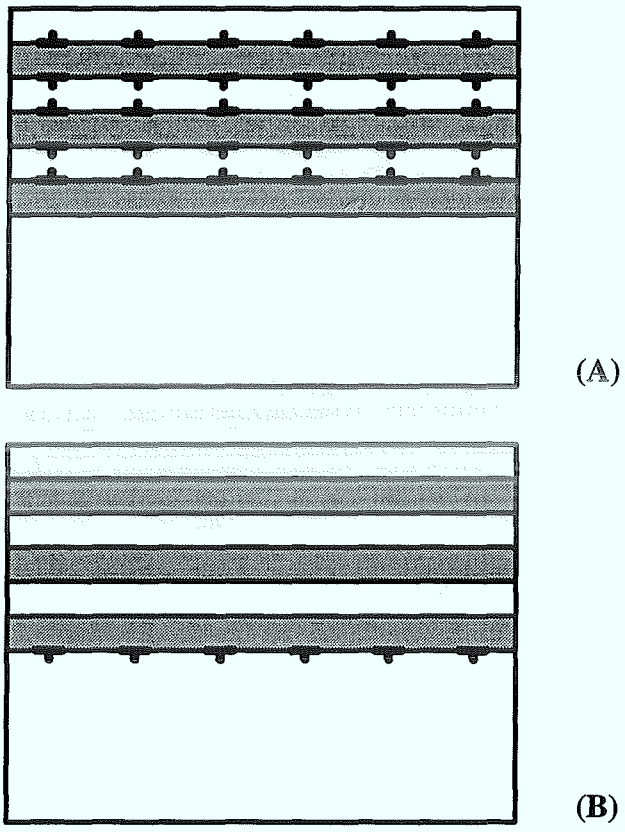


Abbildung 1.1: Relaxationsmoden der Übergitterrelaxation. (A) Relaxation der Einzelschichten durch Versetzungseinbau an den Übergittergrenzflächen. (B) Relaxation des gesamten Übergitters durch Versetzungseinbau an der Substratgrenzfläche.

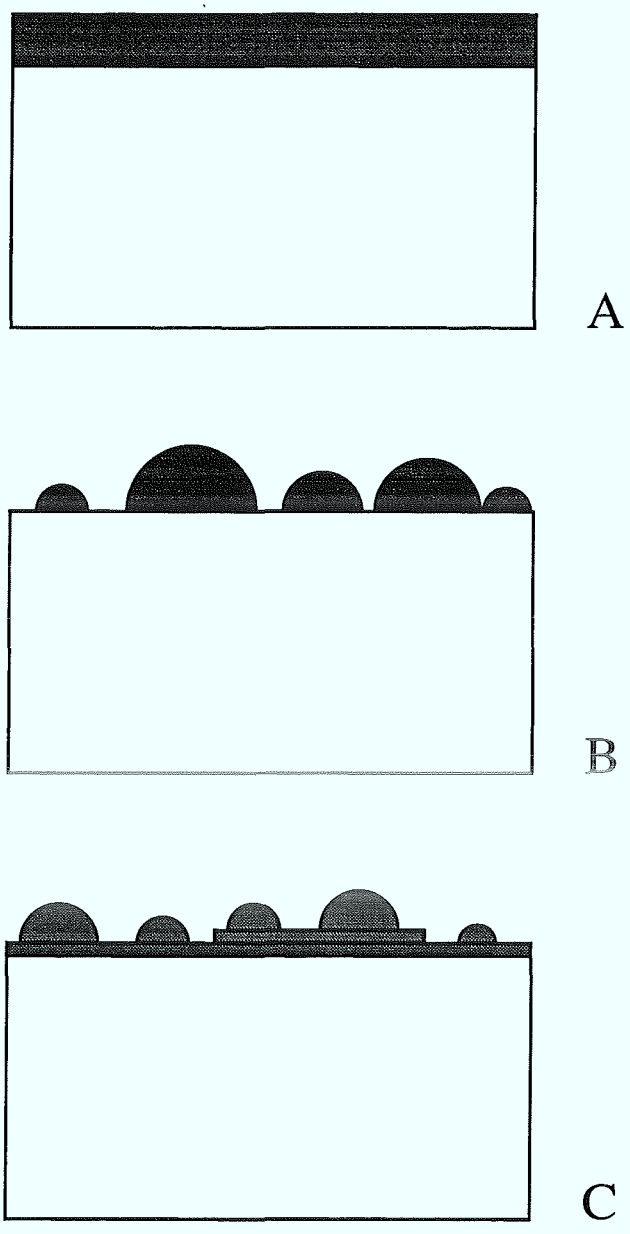


Abbildung 1.2: Wachstumsmodi des Schichtwachstums. (A) zweidimensionales Volmer-Weber-Wachstum, (B) dreidimensionales Inselwachstum, (C) Stranski-Krastanow-Wachstum.

der Gesamtdicke der Übergitter in allen Proben ergibt unter den vorher genannten Bedingungen eine Variation der mittleren Fehlpassung der Übergitter relativ zum Substrat.

Bei der vorliegenden Zinkblende-Struktur und beim zweidimensionalen Wachstum auf (001)-orientierten Substraten werden bei der Relaxation vorwiegend Grenzflächendefekte eines bestimmten charakteristischen Typs beobachtet. Die plastische Verformung während der Relaxation wird durch 60°-*Versetzungen* getragen, die durch Aktivieren der ⟨110⟩{111}-Gleitsysteme und die (001)-Orientierung der Grenzflächen regelmäßige Netzwerke von Fehlpassungsversetzungen (*misfit dislocations*) ergeben [95, 97, 3, 2].

Die mikrostrukturelle Charakterisierung der Übergitter erfolgte durch konventionelle und hochauflösende Transmissionselektronenmikroskopie (Kapitel 4). Die anschließende quantitative Bestimmung der Grenzflächendefektdichten und der Verspannungszustände ermöglichte die Einordnung und den Vergleich mit bestehenden Relaxationsmodellen (Kapitel 2), die entweder den Fall einer fehlgepaßten Einzelschicht auf einem Substrat beschreiben [129, 127, 121, 83, 141, 142, 69, 31, 32], oder die auf spezielle Geometrien fehlgepaßter Übergitter erweitert sind [123, 126].

Die klassischen Relaxationsmodelle gelten jedoch nur für eine Reihe spezieller Schichtanordnungen und Annahmen, die im gegebenen Fall kritisch überprüft wurden (Kapitel 5), und die insbesondere für den Fall des fehlgepaßten Übergitters mit unterschiedlichen Einzelschichtdicken erweitert werden mußten. Auf diese Weise wurde eine Relaxationsbarriere identifiziert, die eine vollständige Relaxation eines Übergitters verhindert.

Kapitel 2

Relaxationsmodelle

In diesem Kapitel wird ein ausführlicher Überblick über die bestehenden Relaxationsmodelle gegeben, einschließlich einer Darstellung der Grundlagen, die auf der Anwendung der linearen Elastizitätstheorie basieren. Die reichhaltige Literatur über Relaxationsmodelle, sowie ihrer Anwendungen in der Materialwissenschaft, ist von einer bemerkenswert hohen Zahl von Ungenauigkeiten und numerischen Fehlern durchsetzt (einen Überblick gibt [51, 40]), deren Herkunft oft an den Endformeln, die charakteristische Größen wie die kritische Schichtdicke oder die homogene Restdehnung beschreiben, schwer nachzuvollziehen ist.

Aus diesem Grund wurde im Folgenden eine Darstellung des Zusammenhangs versucht, der bei den Grundlagen der Elastizitätstheorie beginnt [109], zum einfachen Modell der Einzelschichtrelaxation fortschreitet, und bei dem relativ komplexen Modell der Übergitterrelaxation endet. Die Struktur des Aufbaus vom Einfachen zum Komplexen ist durch die Gültigkeit des Superpositionsprinzips der linearen Elastizitätstheorie gegeben. Insbesondere wurde Wert auf die klare Darstellung von Modellerweiterungen gelegt, die durch das Einarbeiten von Vorfaktoren in die vorher genannten Endformeln möglich sind.

Das Kapitel ist durch einen Abschnitt ergänzt, der sich in Hinblick auf die spätere Diskussion mit den Teilprozessen befaßt, die bei der Betrachtung der Relaxationskinetik zu berücksichtigen sind.

2.1 Grundlagen der Elastiziätstheorie

Im allgemeinen beinhalten *Realkristalle* gegenüber dem mathematischen Modell des unendlich ausgedehnten, idealen Kristallgitters, das durch eine langreichweitige Translationssymmetrie gekennzeichnet ist, Gitterstörungen, wie zum Beispiel lokale Verzerrungen, Punktfehler, Versetzungen, Stapelfehler, Korngrenzen, Zwillingsgrenzen, Risse und so fort. Jeder dieser Defekte kann durch ein Feld aus Verschiebungsvektoren $u_i(\vec{r})$, $i = 1, 2, 3$ beschrieben werden, das die Verrückung eines jeden Atoms des Realkristalls aus seiner Position im Idealkristall angibt.

Die Ursache für die Verschiebung einzelner Atome ist das Wirken äußerer oder innerer Kräfte, deren örtliche Verteilung durch den *Spannungstensor* $\sigma_{ik}(\vec{r})$, $i, k = 1, 2, 3$ beschrieben wird. Die Komponente σ_{ik} ist die i-Komponente der Kraft pro

Fläche auf die Seitenfläche eines Volumenelements $\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$ mit der Oberflächennormalen parallel zur k-Koordinate. Bezüglich des gleichen Volumenelements wird der *Verzerrungstensor* $u_{ik}(\vec{r})$, $i,k = 1,2,3$, definiert,

$$u_{ik} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right), \tag{2.1}$$

der durch die räumliche Ableitung des Verschiebungsfeldes gegeben ist.

In vielen praktisch realisierten Fällen sind die Verzerrungen des Realkristalls klein im Vergleich zu den Gitterkonstanten des Idealkristalls, und es besteht ein linearer Zusammenhang zwischen Spannungs- und Verzerrungstensor. In einem solchen Fall kann der Realkristall in guter Näherung im Rahmen der *linearen Elastizitätstheorie* beschrieben werden [109].

Oft wird außer der Linearität noch eine weitere Vereinfachung gemacht, die die Anisotropie eines Festkörpers betrifft. Im allgemeinen Fall wird die Linearität zwischen Spannungs- und Verzerrungstensor durch einen Satz von 81 unabhängigen elastischen Konstanten beschrieben, eine Mannigfaltigkeit, die sich jedoch bei Vorhandensein von Kristallsymmetrien drastisch reduziert, bis hinunter auf einen Satz von drei unabhängigen Konstanten im einfachsten Fall der kubischen Symmetrie.

Eine Anwendung der linearen Elastizitätstheorie schon auf einfache Kristalldefekte zeigt jedoch, daß diese mit Verzerrungsfeldern verknüpft sein können, für die es selbst bei korrekter Berücksichtigung kubischer Symmetrie eine analytische Darstellung nur noch in den Spezialfällen gibt, in denen Defekte selbst schon eine bestimmte Symmetrie und Orientierung besitzen [45, 47, 167]. Aus diesem Grund behilft man sich mit der Reduktion auf den Fall des *isotropen Körpers*, dessen Symmetrie so weit erhöht ist, daß er durch nur zwei elastische Konstanten beschrieben wird. In der Literatur werden jedoch meist nicht die elastischen Konstanten selbst, sondern die äquivalenten Größen des *Schubmoduls* G und der *Poissonzahl* ν angegeben.

Im thermodynamischen Gleichgewicht nimmt der Kristall ein Minimum der freien Energie F ein, deren elastischer Anteil im Rahmen der linearen Elastizitätstheorie durch

$$F = \frac{1}{2}\int_V \sigma_{ik}u_{ik}\,\mathrm{d}V \tag{2.2}$$

gegeben ist. In diesem Zustand befinden sich auch die internen Kräfte zwischen den einzelnen Volumenelementen $\mathrm{d}V$ und die äußeren Kräfte auf die Kristalloberfläche im mechanischen Gleichgewicht. Das Kräftegleichgewicht bei Abwesenheit äußerer Kräfte resultiert in den beiden *Grundgleichungen* der linearen Elastizitätstheorie

$$\frac{\partial \sigma_{ik}}{\partial x_k} = 0 \tag{2.3}$$

$$\sigma_{ik}n_k = 0 \tag{2.4}$$

Die erste Gleichung ist im Volumen des Kristalls erfüllt, während die zweite auf der *freien Oberfläche* des Kristalls gilt. Die n_k sind die Komponenten des Normalenvektors der Oberfläche.

Aus der Linearität der Grundgleichungen folgt außerdem die Gültigkeit des Superpositionsprinzips zweier Lösungen. Dieses für die Herleitung von Relaxationsmodellen wichtige Prinzip erlaubt eine wesentliche Vereinfachung in der Berechnung der Verzerrungsfelder einer unter Umständen sehr komplexen Defektstruktur. In einem ersten Schritt wird das Gleichgewichtsverzerrungsfeld für jeden einzelnen Defekt getrennt berechnet, anschließend werden die einzelnen Lösungen linear zur Gesamtlösung kombiniert.

In der Regel ist die freie Energie der komplexen Defektstruktur verschieden von der Summe der freien Energien der einzelnen Defekte, was für den Fall der Überlagerung zweier Verzerrungsfelder (L) und (D) unmittelbar einsichtig ist:

$$\begin{aligned} F &= \frac{1}{2}\int_V (\sigma_{ik}^{(\mathrm{L})}(\vec{r}) + \sigma_{ik}^{(\mathrm{D})}(\vec{r}))(u_{ik}^{(\mathrm{L})}(\vec{r}) + u_{ik}^{(\mathrm{D})}(\vec{r}))\,\mathrm{d}\vec{r} \qquad (2.5)\\ &= \underbrace{\frac{1}{2}\int_V \sigma_{ik}^{(\mathrm{L})}(\vec{r})u_{ik}^{(\mathrm{L})}(\vec{r})\,\mathrm{d}\vec{r}}_{F_\mathrm{L}} + \underbrace{\frac{1}{2}\int_V \sigma_{ik}^{(\mathrm{D})}(\vec{r})u_{ik}^{(\mathrm{D})}(\vec{r})\,\mathrm{d}\vec{r}}_{F_\mathrm{D}} + \underbrace{\int_V \sigma_{ik}^{(\mathrm{L})}(\vec{r})u_{ik}^{(\mathrm{D})}(\vec{r})\,\mathrm{d}\vec{r}}_{F_\mathrm{WW}}. \end{aligned}$$

Hierbei bezeichnen F_L und F_D die freien Energien der einzelnen Defekte und F_WW den Differenzbetrag zur freien Energie des Gesamtsystems, der als Wechselwirkungsenergie zwischen den einzelnen Defekten interpretiert werden kann.

Daraus folgt die Regel, daß ein „ungestörtes“ System (L) labil gegen eine Störung (D) ist, wenn

$$F_\mathrm{D} + F_\mathrm{WW} < 0, \qquad (2.6)$$

das heißt, daß insgesamt eine Energieabsenkung des Systems stattfindet, wenn der zusätzliche Energieaufwand für die Einführung eines Defekts durch die Wechselwirkungsenergie überkompensiert wird.

2.2 Ableitung von Relaxationsmodellen

Grundsätzlich erfolgt der Entwurf eines Relaxationsmodells einer beliebigen deformierten Kristallstruktur in vier Schritten. Zuerst wird festgestellt, welche geometrische Form die Oberfläche des gegebenen Systems hat, und welche Gitterstörungen experimentell beobachtet werden.

Anschließend wird das Gleichgewichtsverzerrungsfeld für jede Defektklasse getrennt aus den Grundgleichungen der Elastizitätstheorie berechnet. Im allgemeinen enthalten die analytischen Lösungen noch eine Zahl freier experimenteller Parameter, wie zum Beispiel die Schichtdicke eines verspannten Films auf einem Substrat, den Abstand eines Defektes von der Kristalloberfläche oder den Abstand zwischen zwei benachbarten Defekten.

Im dritten Schritt werden die Verzerrungsfelder der einzelnen Defekte überlagert und sind somit nach dem Superpositionsprinzip Lösungen der Grundgleichungen des Gesamtsystems. Aus dem Verzerrungsfeld und dem dazu korrespondierenden Verspannungsfeld wird die freie Energie des Gesamtsystems berechnet. Sie enthält immer noch den Satz freier experimenteller Parameter.

Zuletzt wird das Minimum der freien Energie bestimmt, indem die noch freien geometrischen Parameter variiert werden.

Im vorliegenden Fall eines fehlgepaßten Übergitters auf einem Substrat werden experimentell eine planare Kristalloberfläche sowie eine Anzahl von *Versetzungen* beobachtet, deren Versetzungslinien parallel zur Kristalloberfläche verlaufen. In den nächsten Abschnitten wird die Ableitung eines Relaxationsmodells für diesen konkreten Fall dargelegt.

Nach der Bestimmung des Verzerrungszustandes einer homogen verspannten, jedoch versetzungsfreien Einzelschicht und des Verzerrungszustandes einer parallel zu einer Kristalloberfläche verlaufenden Versetzung wird zunächst ein Modell für eine einzelne relaxierte Schicht auf einem Substrat abgeleitet. Die dabei gewonnenen Erkenntnisse vereinfachen die Betrachtungen, die zu einem Relaxationsmodell eines relaxierten Übergitters führen, das ja aus einer Sequenz verspannter Einzelschichten besteht.

2.3 Homogen deformierte Einzelschicht

Es wird eine unendlich ausgedehnte, elastisch isotrope Schicht konstanter Dicke h betrachtet, die nur in der Ebene der Schicht homogen verspannt ist, wobei die Scherspannungen jeweils Null sind. Die Koordinatenwahl erfolgt so, daß die x- und y-Achse in der Schichtebene liegen, und die z-Achse parallel zur Schichtnormalen. Dann sind außer σ_{xx} und σ_{yy} alle anderen Komponenten des Spannungstensors Null und die zugehörigen Verzerrungen haben die Werte [109]

$$u_{xx} = \frac{1}{E}\left(\sigma_{xx} - \nu\sigma_{yy}\right), \tag{2.7}$$

$$u_{yy} = \frac{1}{E}\left(\sigma_{yy} - \nu\sigma_{xx}\right), \tag{2.8}$$

mit dem *Elastizitätsmodul* $E = 2\ (1+\nu)\ G$. Die Grundgleichungen der linearen Elastizitätstheorie ergeben außerdem die Beziehung

$$u_{zz} = -\frac{\nu}{1-\nu}\ \left(u_{xx} + u_{yy}\right), \tag{2.9}$$

das heißt, die Schicht ist entlang der z-Achse verformt, ohne daß eine Spannung σ_{zz} in dieser Richtung wirkt. Die freie Energie pro Flächeneinheit der Schicht hat den Wert

$$F = \frac{Gh}{1-\nu}\left(u_{xx}^2 + u_{yy}^2 + 2\nu u_{xx}u_{yy}\right). \tag{2.10}$$

2.4 Verzerrungsfeld einer Versetzung

Versetzungen sind Liniendefekte, die die Eigenschaft haben, daß die Kristallstruktur im gesamten Kristall, bis auf einen kleinen Bereich an der Linie selbst, lokal erhalten bleibt. Global ist die Kristallstruktur jedoch in einer Weise gestört, die sich erst bemerkbar macht, wenn man die Umgebung der Versetzungslinie mit Hilfe eines

Burgersumlaufs analysiert [13, 59]. Dieser besteht aus einem geschlossenen Umlauf L im ungestörten Kristall, bei dem eine beliebige Anzahl von Schritten in jede Kristallrichtung ausgeführt wird, so daß zum Schluß der Ausgangspunkt wieder erreicht ist. Anschließend wird der Umlauf mit der gleichen Anzahl von Schritten im gestörten Kristallbereich durchgeführt, so, daß der Umlauf die Versetzungslinie einschließt. Im allgemeinen endet der zweite Umlauf nicht am Ausgangspunkt, denn jeder Elementarschritt des zweiten Umlaufs enthält ja zusatzlich die lokale Komponente des Verschiebungsfeldes $u_i(\vec{r})$.

Es werden zwei grundlegende Typen von Versetzungen, nämlich *Stufenversetzungen* und *Schraubenversetzungen*, unterschieden.

Stufenversetzungen sind gekennzeichnet durch eine einzelne Kristallgitterebene, die an der Versetzungslinie endet. Das Erzeugen einer Stufenversetzung wird am einfachsten durch einen sogenannten *Volterra-Prozeß* veranschaulicht [182], bei dem zuerst der Kristall bis zur späteren Versetzungslinie aufgeschnitten, und die Hälften auf beiden Seiten des Schnitts *senkrecht* zur Linie auseinandergezogen werden. Dann wird die einzelne Kristallgitterebene bis zur Versetzungslinie eingeschoben und schließlich werden die Kristallhälften wieder zusammengefügt (Abbildung 2.1). Schon ohne eine genaue Berechnung des Verzerrungsfeldes sieht man, daß die lokal überschüssige Halbebene in einem unendlich ausgedehnten Kristall zu einer Stauchung einer Hälfte des Kristalls und zu einer Dehnung der anderen Hälfte führt.

Eine Schraubenversetzung ensteht ebenfalls durch einen Schnitt durch den Kristall bis zur späteren Versetzungslinie. Die beiden Kristallhälften auf beiden Seiten des Schnitts werden diesmal jedoch *parallel* zur Versetzungslinie verschoben und anschließend wieder zusammengefügt. Die zugehörige Verformung des Kristallgitters führt dazu, daß die vorher ungestörten Gitterebenen senkrecht zur Versetzungslinie anschließend die Form schraubenförmiger Flächen annehmen (Abbildung 2.2). Hier sieht man ohne genaue Berechnung des Verzerrungsfeldes, daß das Kristallgitter um die Versetzungslinie herum einer gleichsinnigen Scherung unterworfen ist.

Aus den beiden Volterraprozessen ist direkt abzulesen, daß bei einer Stufenversetzung das zugehörige Verschiebungsfeld $u_i(\vec{r})$ nur Komponenten ungleich Null enthält, die senkrecht zur Versetzungslinie verlaufen, während bei einer Schraubenversetzung nur Komponenten des Verschiebungsfeldes parallel zur Versetzungslinie auftreten. Aus den Prozessen folgt weiter, daß die Verschiebungsfelder einen jeweils resultierenden Betrag b_i haben müssen, der dem Auseinanderziehen der Kristallhälften im einen Fall und dem Verschieben der Hälften im anderen Fall entspricht. Dieser resultierende Betrag heißt *Burgersvektor* und er ergibt sich aus dem Verschiebungsfeld durch

$$b_i = -\oint_L \mathrm{d}u_i = -\oint_L \frac{\partial u_i}{\partial x_k}\mathrm{d}x_k, \tag{2.11}$$

wenn der Integrationsweg die Versetzungslinie einschließt [109]. Die Berechnung des Ringintegrals ist die mathematische Fassung des oben dargestellten Burgersumlaufs, der ja nicht wie hier mit infinitesimalen Elementarschritten $\mathrm{d}x_k$, sondern in Schritten einzelner Gitterkonstanten in einer hochaufgelösten Gitterabbildung der projizierten Struktur durchgeführt wird. Für Stufenversetzungen liegt der Burgersvektor senkrecht zur Versetzungslinie, während er für Schraubenversetzungen parallel zur

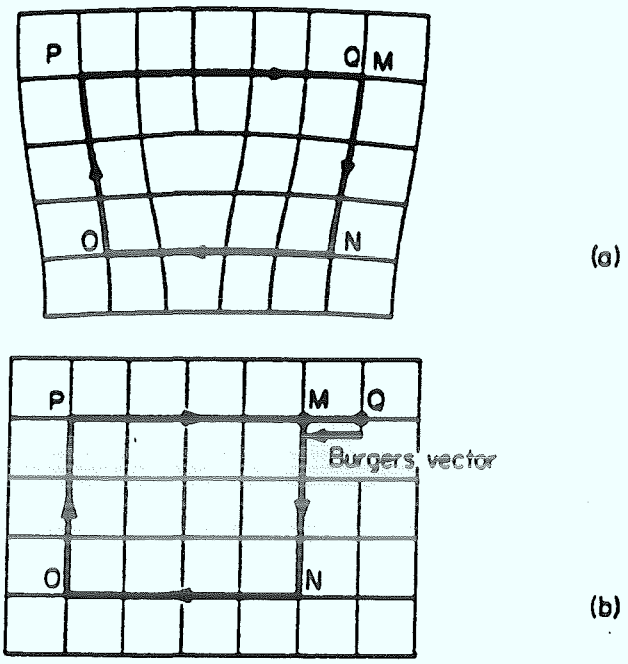


Abbildung 2.1: Burgersumlauf einer Stufenversetzung [88]. (A) Umlauf MNOPQ im gestörten Kristallbereich. Zwischen den Punkten P und Q ist die von oben eingeschobene Halbebene zu erkennen. (B) Der gleiche Umlauf MNOPQ im perfekten Kristall. Der Schließungsvektor $\vec{QM}$ ist der Burgersvektor der Versetzung.

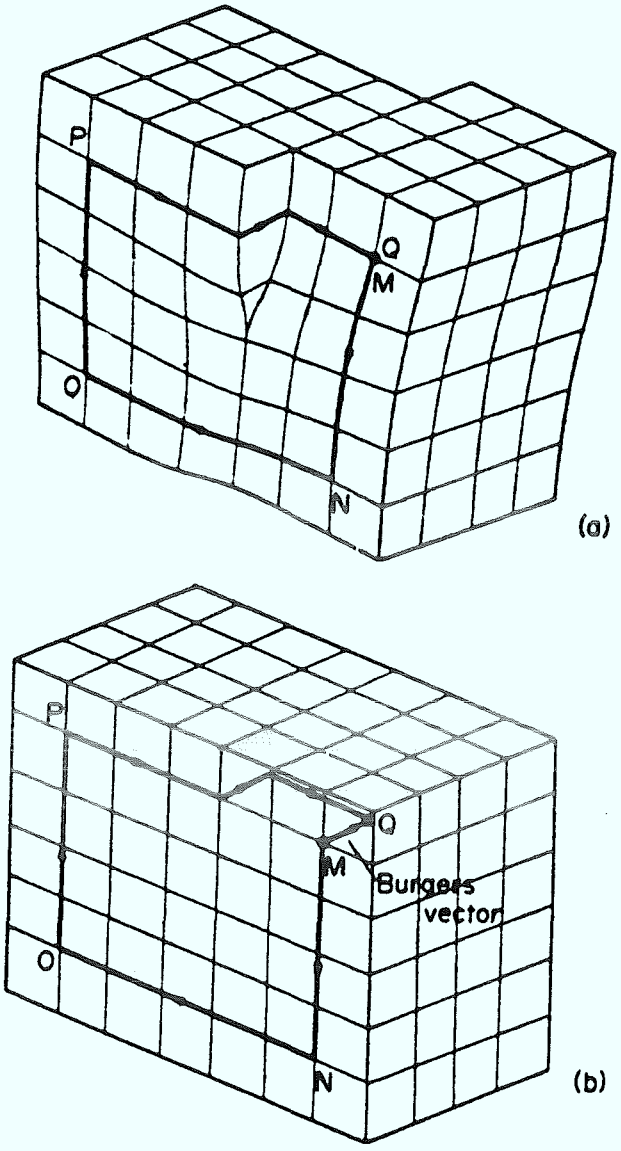


Abbildung 2.2: Burgersumlauf einer Schraubenversetzung [88]. (A) Umlauf MNOPQ im gestörten Kristallbereich. Zwischen den Punkten P und Q ist die Verschiebung der Kristallhälften zu erkennen. (B) Der gleiche Umlauf MNOPQ im perfekten Kristall. Der Schließungsvektor $\overrightarrow{QM}$ ist der Burgersvektor der Versetzung.

Versetzungslinie verläuft. In der Realität treten außerdem „gemischte" Versetzungstypen auf (*mixed dislocations*), deren Burgersvektoren sowohl eine Stufen- als auch eine Schraubenkomponente haben.

2.4.1 Versetzung im unendlichen Kristall

Die Berechnung der Deformation eines isotropen Kristalls in der Umgebung einer Stufenversetzung liefert im Rahmen der linearen Elastizitätstheorie eine analytisch darstellbare Lösung, deren Herleitung auf den oben angegeben Grundgleichungen beruht, unter der zusätzlichen Bedingung, daß ein Burgersumlauf einen nichtverschwindenden Burgersvektor senkrecht zur Versetzungslinie ergibt [106, 114].

Die Koordinatenwahl erfolgt so, daß x- und y-Achse senkrecht zur Versetzungslinie liegen, und die z-Achse parallel zur Versetzungslinie verläuft. Dann lautet die Lösung für das Verschiebungsfeld einer Stufenversetzung [83]

$$u_x = \frac{b}{2\pi}\left(\arctan\frac{y}{x} + \frac{1}{2(1-\nu)}\frac{xy}{x^2+y^2}\right), \tag{2.12}$$

$$u_y = -\frac{b}{2\pi}\left(\frac{1-2\nu}{2(1-\nu)}\ln\sqrt{x^2+y^2} + \frac{1}{4(1-\nu)}\frac{x^2-y^2}{x^2+y^2}\right). \tag{2.13}$$

Der sich daraus ergebende Spannungstensor hat seine einfachste Gestalt in Zylinderkoordinaten

$$\sigma_{rr} = \sigma_{\varphi\varphi} = -\frac{Gb}{2\pi(1-\nu)}\frac{\sin\varphi}{r}, \tag{2.14}$$

$$\sigma_{r\varphi} = \frac{Gb}{2\pi(1-\nu)}\frac{\cos\varphi}{r}. \tag{2.15}$$

Der elastische Anteil der freien Energie einer Stufenversetzung pro Längeneinheit ist

$$F_{\mathrm{el}} = \frac{Gb^2}{4\pi(1-\nu)}\ln\frac{R}{r_0}, \tag{2.16}$$

wobei die Integration in Zylinderkoordinaten von einem inneren Abschneideradius r_0 nahe der Versetzungslinie bis zu einem äußeren Abschneideradius R ausgeführt wurde. Das Aussparen des Bereichs um die Versetzungslinie selbst, des *Versetzungskerns*, ist nötig, weil Spannungs- und Verzerrungstensor an dieser Stelle eine Singularität haben. In der Regel wird dem Versetzungskern selbst ein empirischer Wert der freien Energie zugeordnet, der den Betrag

$$F_{\mathrm{Kern}} = \frac{Gb^2}{4\pi(1-\nu)} \tag{2.17}$$

hat [51], sodaß die Gesamtenergie pro Längeneinheit

$$F = \frac{Gb^2}{4\pi(1-\nu)}\left(\ln\frac{R}{b} + 1\right) \tag{2.18}$$

ist. Hierbei wurde eine zweite empirische Näherung gemacht, nämlich daß der Versetzungskern einen Radius von $r_0 = b$ haben soll.

Die Berechnung der Deformation eines isotropen Kristalls in der Umgebung einer Schraubenversetzung liefert mit der gleichen Koordinatenwahl eine analytisch darstellbare Lösung, jedoch jetzt unter der Bedingung, daß ein Burgersumlauf einen nichtverschwindenden Burgersvektor parallel zur Versetzungslinie ergibt [106, 114].

Die Lösung für das Verschiebungsfeld lautet in Zylinderkoordinaten

$$u_z = \frac{b}{2\pi}\varphi, \tag{2.19}$$

sowie die Lösung für den Spannungstensor

$$\sigma_{z\varphi} = \frac{Gb}{2\pi r}. \tag{2.20}$$

Die freie Energie einer Schraubenversetzung pro Längeneinheit ist dann, mit einem empirischen Betrag

$$F_{\text{Kern}} = \frac{Gb^2}{4\pi} \tag{2.21}$$

für den Versetzungskern [51] und den Abschneideradien R, $r_0 = b$, gleich:

$$F = \frac{Gb^2}{4\pi}\left(\ln\frac{R}{b} + 1\right). \tag{2.22}$$

2.4.2 Versetzung nahe einer Kristalloberfläche

Verläuft eine Versetzung nahe einer Kristalloberfläche, so muß bei der Berechnung des Verzerrungsfeldes im Gegensatz zum Fall des unendlich ausgedehnten Kristalls zusätzlich die Randbedingung $\sigma_{ik}n_k = 0$ erfüllt sein.

Die Lösungen sind für eine Stufen- oder Schraubenversetzung parallel zur Oberfläche noch analytisch zu berechnen, jedoch wird, insbesondere für die Stufenversetzung, die analytische Darstellung unhandlich, und sie ist nicht mehr in kompakter Form darzustellen [46]. Für das Verständnis und die weitere Verwendung des Ergebnisses für die freie Energie in Relaxationsmodellen ist es hilfreich, den Weg des Rechengangs zu skizzieren.

Die Grundidee besteht darin, daß, analog zur Methode der Spiegelladungen in der Elektrostatik, das Verspannungsfeld einer zweiten, virtuellen Versetzung im Vakuum über dem Kristall das Verspannungsfeld der ursprünglichen Versetzung auf der Oberfläche gerade so kompensieren soll, daß die zusätzliche Randbedingung erfüllt ist. Zu diesem Zweck muß die Versetzungslinie dieser *Bildversetzung* parallel zur ursprünglichen Versetzung im gleichen Abstand zur Kristalloberfläche verlaufen, und sie muß einen gleich großen Burgersvektor entgegengesetzter Richtung besitzen.

Es erweist sich, daß diese Konstruktion der Lösung im Fall der Schraubenversetzung exakt ist [46]. Für eine Stufenversetzung ist die Kompensation durch eine Bildversetzung jedoch nicht ausreichend, und es müssen weitere Terme hinzukonstruiert werden, die aber keine direkt anschauliche Bedeutung mehr haben.

Interessanterweise reduzieren sich trotz der komplizierteren resultierenden Verspannungsfelder die analytischen Ausdrücke für die freie Energie einer Stufen- oder Schraubenversetzung auf Ausdrücke, die analog zu denen im Fall des unendlich ausgedehnten Kristalls aufgebaut sind. Es sind lediglich statt der äußeren Abschneideradien R die geometrischen Abstände h der Versetzungen zur Kristalloberfläche einzusetzen [46]:

$$F_{\text{Stufe}} = \frac{Gb^2}{4\pi(1-\nu)}\left(\ln\frac{h}{b}+1\right), \tag{2.23}$$

$$F_{\text{Schraube}} = \frac{Gb^2}{4\pi}\left(\ln\frac{h}{b}+1\right), \tag{2.24}$$

wobei die Ausdrücke wieder einen empirischen Beitrag „+1“ für den Versetzungskern enthalten.

2.5 Gleichgewichtszustand einer Einzelschicht

Experimentell wird beobachtet, daß eine epitaktische, fehlgepaßte Schicht, die auf ein Substrat aufgewachsen wurde, ab einer bestimmten Schichtdicke Fehlpassungsversetzungen in der Grenzfläche zwischen Schicht und Substrat enthält [95, 97, 3, 2]. Die Fehlpassung ist definiert durch den relativen Gitterkonstantenunterschied

$$f = \frac{a_{\text{S}} - a_{\text{L}}}{a_{\text{L}}}, \tag{2.25}$$

mit der Gitterkonstante der Schicht a_{L} und der Gitterkonstante des Substrats a_{S}. Sie ist gleich den Werten u_{xx} und u_{yy} des Verzerrungstensors, wenn die Schicht defektfrei mit der Gitterkonstante des Substrats kohärent aufwächst.

Dieses *kohärente* Wachstum ist jedoch nur bis zu einer *kritischen Schichtdicke* h_{c} energetisch günstig. Bei größerer Schichtdicke h wird die freie Energie des Systems durch den Einbau von Fehlpassungsversetzungen verringert. Sowohl in der Diamant- als auch in der Zinkblendestruktur sind die $\langle 110\rangle\{111\}$-Gleitsysteme die energetisch günstigsten [83]. Die Schnittgeraden zwischen den $\{111\}$-Gleitebenen und den im hier untersuchten Material vorliegenden (001)-Oberflächen ergeben $\langle 110\rangle$-Kristallrichtungen, die somit die Linienrichtungen der Fehlpassungsversetzungen zwischen den GaAs- und den InGaAs-Schichten vorgeben. Aus diesem Grund verlaufen in der natürlichsten Koordinatenwahl die x- und die y-Achse parallel zu den [110]- und $[1\bar{1}0]$-Linienrichtungen der Fehlpassungsversetzungen und die z-Achse parallel zur (001)-Schichtnormalen.

Das Relaxationsproblem ist symmetrisch bezüglich der x- und der y-Achse, so daß es für die weiteren Betrachtungen zunächst ausreicht, Relaxation nur entlang der x-Richtung anzunehmen. Die daraus folgenden Ergebnisse sind dann auf einfache Weise auf die Relaxation entlang beider Richtungen zu verallgemeinern.

Die Gesamtenergie des Systems bei Einsetzen der Relaxation, das heißt beim Einführen der ersten Versetzung, ergibt sich mit Hilfe des Superpositionsprinzips

und der Gleichung (2.5), aus den Ausdrücken der Energie pro Längeneinheit einer Versetzung F_{D} (Gleichung (2.23)) und der Energie einer homogen deformierten Einzelschicht F_{L} (Gleichung (2.10)).

Für $u_{yy} = 0$ ist der Wert der Wechselwirkungsenergie mit der Spannungskomponente

$$\sigma_{xx}^{(\mathrm{L})}(\vec{r}) = \frac{2G}{1-\nu} u_{xx}^{(\mathrm{L})} = \text{const.} \tag{2.26}$$

der Schicht und

$$-\int_{-\infty}^{\infty} u_{xx}^{(\mathrm{D})}(x)\,\mathrm{d}x = u_x^{(\mathrm{D})}(x=\infty) - u_x^{(\mathrm{D})}(x=-\infty) = b \tag{2.27}$$

gleich

$$F_{\mathrm{WW}} = -\frac{2G}{1-\nu} b h u_{xx}^{(\mathrm{L})}. \tag{2.28}$$

Aus dem Stabilitätskriterium (2.6) folgt dann für das Einsetzen der Relaxation der Fehlpassung $f = u_{xx}$ durch Stufenversetzungen bei einer kritischen Schichtdicke h_{c} die wichtige Beziehung

$$f = \frac{b}{8\pi h_{\mathrm{c}}}\left(\ln\frac{h_{\mathrm{c}}}{b} + 1\right). \tag{2.29}$$

Beim weiteren Anwachsen der Schichtdicke werden sukzessive immer mehr Versetzungen eingebaut, und es ist interessant zu betrachten, wie die Versetzungsdichte mit der Schichtdicke h zunimmt und gleichzeitig die Deformation $\epsilon = u_{xx}$ der aufgewachsenen Schicht abnimmt.

Zu diesem Zweck wird ein Relaxationszustand betrachtet, bei dem schon eine Anzahl von Versetzungen eingebaut ist, deren Versetzungslinien parallel verlaufen und die einen äquidistanten Abstand S haben. Durch eine rein geometrische Betrachtung sieht man, daß entlang des Abstandes S eine Gitterverschiebung vom Betrag b des Burgersvektors abgebaut ist, und deshalb ist die ursprüngliche Deformation f der aufgewachsenen Schicht um den Wert b/S reduziert. Die *mittlere* Deformation ϵ beträgt dann

$$\epsilon = f - \frac{b}{S}, \tag{2.30}$$

und die Restdeformation und die Versetzungsdichte $1/S$ sind direkt miteinander verknüpft. Grundsätzlich ist die mittlere Deformation ortsabhängig mit einer Periodizität des Versetzungsabstandes S, denn sie ist ja eine Superposition der homogenen Fehlpassungsdeformation und der Deformation des Versetzungsnetzwerkes. Wenn das Netzwerk jedoch genügend dicht ist, dann ist die periodische Modulation schwach, und ϵ ist annähernd konstant. Abweichungen von dieser Näherung werden gewöhnlich in der materialwissenschaftlichen Literatur, in der die klassischen Relaxationsmodelle angewendet werden, ignoriert.

Den Zusammenhang zwischen Restdeformation ϵ und Schichtdicke h erhält man auf die gleiche Weise, wie bei der vorangehenden Bestimmung der kritischen Schichtdicke. Das System aus homogen deformierter Schicht und Versetzungsnetzwerk wird

als der ungestörte Zustand im Sinne von Gleichung (2.5), und eine neu einzuführende Versetzung als Störung aufgefaßt. Das System befindet sich im Gleichgewicht, wenn die Energie der neu eingeführten Versetzung F_{D} gerade durch die Wechselwirkungsenergie F_{WW} kompensiert wird. Die Berechnung der Wechselwirkungsenergie verläuft genau gleich, jedoch wird diesmal statt der vollen Fehlpassung f der reduzierte Wert ϵ eingesetzt. Damit ergibt sich die Restdeformation ϵ zu

$$\epsilon = \frac{b}{8\pi h}\left(\ln\frac{h}{b}+1\right). \tag{2.31}$$

Diese Berechnung gilt jedoch nur für den Fall, daß die restliche Wechselwirkung zwischen neuer Versetzung und dem schon existierenden Versetzungsnetzwerk klein ist. Wegen des äußeren Abschneideradius $R = h$ ist diese Bedingung erfüllt, wenn die Versetzungen um mindestens einen Abstand $S > 2R$ voneinander entfernt sind [121].

Ist das Netzwerk dichter, das heißt $S < 2h$, dann sind noch die Wechselwirkung der neuen Versetzung mit dem Netzwerk, und die Wechselwirkungen der Versetzungen des Netzwerkes untereinander zu berücksichtigen. Gleichung (2.5) spaltet dann in eine Vielzahl von Termen auf. Es ist in diesem Fall einfacher, zuerst das Spannungsfeld des Netzwerkes zu berechnen, und dieses mit der homogenen Fehlpassungsspannung der Schicht linear zu kombinieren. Im Verlauf der immer noch elementar durchzuführenden, jedoch etwas länglichen Rechnung zeigt sich, daß das Verspannungsfeld des Netzwerks einen ortsabhängigen, periodischen Anteil und einen konstanten Anteil hat, der der relaxierten Deformation b/S entspricht [112, 113]. Dieser konstante Anteil wird genau wie oben (Gleichung (2.31)) mit der homogenen Fehlpassungsdeformation f der Schicht zusammengefaßt, der ortsabhängige, periodische Anteil wird in den klassischen Relaxationsmodellen vernachlässigt, weil er exponentiell mit dem Abstand vom Netzwerk auf Null abfällt [112, 113]. Das Ergebnis der Rechnung sind schließlich zwei Terme, die die Energie der homogenen Deformation und die Eigenenergie pro Flächeneinheit des Versetzungsnetzwerks beschreiben:

$$F_{\mathrm{L}} = \frac{Gh}{1-\nu}\epsilon^2, \tag{2.32}$$

$$F_{\mathrm{D}} = \frac{Gb\,(f-\epsilon)}{4\pi\,(1-\nu)}\left(\ln\frac{1}{2(f-\epsilon)}+1\right), \tag{2.33}$$

mit $F_{\mathrm{D}} = F_{\mathrm{Stufe}}/S$ und $R = S/2 = b/2(f-\epsilon)$. Die Minimierung der Gesamtenergie nach dem noch freien Parameter ϵ ergibt

$$\epsilon = \frac{b}{8\pi h}\ln\frac{1}{2(f-\epsilon)}, \tag{2.34}$$

den Zusammenhang zwischen der Verspannung der Einzelschicht und der Schichtdicke.

In der bisherigen Analyse wurde angenommen, daß die beiden Materialien des heteroepitaktischen Systems den gleichen Schubmodul G besitzen. Eine Verallgemeinerung auf den Fall unterschiedlicher Schubmoduln ist auf relativ einfache Weise möglich, wobei nur die Ausdrücke für die Eigenenergie einer Versetzung modifiziert

werden müssen. Das Spannungs- und Verzerrungsfeld einer Grenzflächenversetzung reicht sowohl in das Gebiet der aufgewachsenen Schicht, als auch in das Substrat hinein. Die zugehörige Eigenenergie ist nun nicht einfach nur die arithmetische Summe der Teilenergien in beiden Halbräumen, sondern an der Grenzfläche muß im Gleichgewicht wieder die Bedingung (2.3) erfüllt sein. Das zugehörige anisotrope Problem der linearen Elastizitätstheorie ist analytisch lösbar und das Ergebnis unterscheidet sich von den oben gefundenen Ausdrücken nur dadurch, daß ein modifizierter Schubmodul eingesetzt wird [115],

$$G \longrightarrow \frac{2G_{\mathrm{L}}G_{\mathrm{S}}}{G_{\mathrm{L}}+G_{\mathrm{S}}}, \tag{2.35}$$

mit G_{L} dem Schubmodul der Schicht und G_{S} dem Schubmodul des Substrats. In den Ausdrücken für die kritische Schichtdicke (Gleichung (2.29)) und die Restdeformation (Gleichungen (2.31), (2.34)) müssen dann die Faktoren vor dem Logarithmus jeweils mit dem dimensionslosen Parameter

$$\frac{2G_{\mathrm{S}}}{G_{\mathrm{L}}+G_{\mathrm{S}}} \tag{2.36}$$

multipliziert werden.

Bis zu diesem Punkt war die Analyse auf den Fall der Relaxation nur entlang der x-Richtung erstreckt worden. Im vorliegenden Kristallgitter ist eine Fehlpassung entlang der x-Richtung jedoch immer gleichbedeutend mit einer gleich großen Fehlpassung entlang der y-Richtung. Im Rahmen der isotropen Elastizitätstheorie verläuft dann auch der Relaxationsprozeß entlang beider Richtungen gleich, und aus dem Superpositionsprinzip folgt, daß sich die Spannungen der beiden zugehörigen Versetzungsnetzwerke addieren, und daß der Versetzungsabstand S für beide Richtungen die gleiche Größe hat.

Da die Burgersvektoren der beiden Netzwerke für Stufenversetzungen rechtwinklig zueinander verlaufen, hängen die einzelnen Komponenten des resultierenden Spannungstensors nur entweder von den Koordinaten x und z oder von den Koordinaten y und z ab, und es gibt keine Wechselwirkung zwischen beiden Netzwerken [83]. Die Energien pro Längeneinheit der Versetzungen addieren sich, und es muß keine zusätzliche Wechselwirkungsenergie bei der Entstehung der Netzwerke aufgewendet werden.

Die Energie der homogen deformierten Schicht (Gleichung (2.10)) enthält jedoch eine Kopplung zwischen den Verzerrungen entlang der x- und der y-Richtung. Die Werte u_{xx} und u_{yy} sind wegen der Symmetrie gleich, so daß die Energie der zweiachsig verspannten Schicht einen um den Faktor

$$2\,(1+\nu) \tag{2.37}$$

größeren Wert gegenüber dem Fall der einachsigen Verspannung hat.

Zur Berechnung des Energiegleichgewichts wird eine Volumenzelle betrachtet, deren quadratische Grundfläche mit der Seitenlänge S von Versetzungslinien beider Netzwerke berandet wird. Aus der Berandung folgt, daß eine totale Länge der Versetzungslinien von $2S$ zur Gesamtenergie der Zelle beiträgt.

Bei der Minimierung der Gesamtenergie bezüglich der Grundfläche $S \times S$ nach dem freien Parameter $\epsilon = u_{xx} = u_{yy}$ bleibt dann der funktionale Zusammenhang zwischen der Verspannung ϵ und der Schichtdicke h (Gleichungen (2.31), (2.34)) erhalten, wenn die Formeln um einen dimensionslosen Korrekturfaktor

$$\frac{1}{1+\nu} \tag{2.38}$$

vor dem Logarithmus ergänzt werden. Der Grund liegt darin, daß zwar die Verspannungsenergie der Schicht um einen Faktor 2 $(1+\nu)$ größer ist (Term (2.37)), aber andererseits bei der Relaxation auch die doppelte Energie zur Versetzungserzeugung aufgewendet werden muß [127].

Eine dritte Erweiterung der Ausdrücke für die kritische Schichtdicke und die Verspannung im Gleichgewicht betrifft die Komponente des Burgersvektors einer Fehlpassungsversetzung, die zur Relaxation beiträgt. Im vorliegenden System niedriger Fehlpassung sind die Gleitebenen zur Grenzfläche geneigt, und die Burgersvektoren besitzen neben einer Stufenkomponente, die in der Grenzfläche liegt, auch noch eine Stufenkomponente senkrecht zur Grenzfläche, sowie eine Schraubenkomponente.

Zur Relaxation trägt nach Gleichung (2.30) nur die Stufenkomponente in der Grenzfläche bei, während in die Energie der Versetzungen die volle Burgersvektorlänge eingeht. Gegenüber der obigen Darstellung, die unter der Annahme erfolgte, daß der gesamte Burgersvektor in der Grenzfläche liegt, und die Versetzung vom reinen Stufentyp ist, sind für das vorliegende Materialsystem die entsprechenden Formeln um einen Faktor

$$b/b_{\text{eff}} \tag{2.39}$$

für die Projektion in die Grenzfläche, und um einen Faktor

$$1 - \nu \cos^2 \alpha \tag{2.40}$$

für die Kombination eines Stufen- und eines Schraubenanteils, vor dem Logarithmus zu ergänzen [127]. b_{eff} bezeichnet die in die Grenzfläche projizierte Stufenkomponente des Burgersvektors und α den Winkel zwischen Versetzungslinie und Burgersvektor.

Einschließlich aller Erweiterungen lauten dann die Zusammenhänge für die kritische Schichtdicke

$$f = \frac{2G_{\text{S}}}{G_{\text{L}} + G_{\text{S}}} \frac{b^2(1-\nu\cos^2\alpha)}{8\pi b_{\text{eff}} h_{\text{c}}(1+\nu)} \left(\ln \frac{h_{\text{c}}}{b} + 1 \right), \tag{2.41}$$

für die Restdeformation bei $S > 2h$

$$\epsilon = \frac{2G_{\text{S}}}{G_{\text{L}} + G_{\text{S}}} \frac{b^2(1-\nu\cos^2\alpha)}{8\pi b_{\text{eff}} h(1+\nu)} \left(\ln \frac{h}{b} + 1 \right), \tag{2.42}$$

und für die Restdeformation bei $S < 2h$

$$\epsilon = \frac{2G_{\text{S}}}{G_{\text{L}} + G_{\text{S}}} \frac{b^2(1-\nu\cos^2\alpha)}{8\pi b_{\text{eff}} h(1+\nu)} \ln \frac{b_{\text{eff}}}{2b(f-\epsilon)}, \tag{2.43}$$

mit $S = b_{\text{eff}}/(f - \epsilon)$. Im vorliegenden heteroepitaktischen System sind wegen der niedrigen Fehlpassungen nur Gleitsysteme mit 60°-Versetzungen angeregt [51], so daß in allen in dieser Arbeit betrachteten Fällen $b/b_{\text{eff}} = 2$ und $\alpha = 60°$ ist.

An dieser Stelle wird ausdrücklich darauf hingewiesen, daß die gesamte Literatur zur Theorie der Relaxation implizit der Konvention folgt, daß die Fehlpassung f und die Restdeformation ϵ positive Werte haben sollen. Dadurch erhalten geometrische Größen wie der Versetzungsabstand S automatisch positive Werte, die Argumente der Logarithmen bleiben positiv, und die Rechnungen bleiben frei von Fallunterscheidungen und Vorzeichenbetrachtungen. Der physikalische Sachverhalt bleibt natürlich für negative Fehlpassungen und Restdeformationen der gleiche, so daß die Formeln gültig bleiben, wenn man den entsprechenden Größen das entgegengesetzte Vorzeichen gibt.

2.6 Gleichgewichtszustand eines Übergitters

Im folgenden wird ein Schichtpaket betrachtet, das aus einer Abfolge von Einzelschichten A, B mit Schichtdicken h_{A}, h_{B} zweier verschiedener Materialien mit Gitterkonstanten a_{A}, a_{B}, Schubmoduln G_{A}, G_{B} und Poissonzahlen ν_{A}, ν_{B} besteht. Zur Ableitung der speziellen elastischen Eigenschaften und des Relaxationsverhaltens solcher heteroepitaktischer *Übergitter* ist es zunächst ausreichend, den Fall des unendlich ausgedehnten, *freien* Übergitters zu betrachten [124]. Hierbei ist die aus den zwei Einzelschichten A, B bestehende Schichtperiode entlang der Schichtnormalen unendlich oft fortgesetzt, und es ist kein Substrat vorhanden, das dem Übergitter eine zusätzliche Fehlpassung aufprägt. Die Erweiterung auf endliche Übergitter, sowie der Einfluß eines Substrats werden weiter unten diskutiert.

2.6.1 Kohärentes Übergitter

In Analogie zum Fall der fehlgepaßten Einzelschicht existieren kritische Kombinationen der Schichtdicken $h_{\text{A,c}}$, $h_{\text{B,c}}$, im Verhältnis zu den Materialkonstanten, oberhalb derer Fehlpassungsversetzungen an den Grenzflächen eingebaut werden. Zur Bestimmung des kohärenten Verspannungszustandes wird jetzt angenommen, daß die Einzelschichtdicken unterhalb der jeweiligen kritischen Dicke liegen, so daß die relative Fehlpassung durch eine rein tetragonale Verzerrung angepaßt wird. In vielen Fällen, so auch im vorliegenden System niedriger Fehlpassung, ist der Unterschied der beiden Poissonzahlen vernachlässigbar, so daß nur die jeweiligen Gitterkonstanten und Schubmoduln unterschiedliche Werte annehmen.

Durch Anwendung der Gleichung (2.10) auf die Einzelschichten und unter der zusätzlichen Randbedingung der Kohärenz erhält man für die freie Energie des Übergitters pro Schichtperiode und Fläche den Wert [124]:

$$F = 2G_{\text{A}}h_{\text{A}}\frac{1+\nu}{1-\nu}\epsilon_{\text{A}}^2 + 2G_{\text{B}}h_{\text{B}}\frac{1+\nu}{1-\nu}\epsilon_{\text{B}}^2. \tag{2.44}$$

Wegen der Kohärenz wird die Schicht mit der kleineren Gitterkonstanten gedehnt, während die Schicht mit der größeren Gitterkonstanten komprimiert wird. Die zu-

gehörigen Verspannungen ϵ_A, ϵ_B haben dann *entgegengesetztes* Vorzeichen, und durch Vergleich zweier Einheitszellen der Materialien A, B erhält man unter Beachtung der Vorzeichen die Beziehung

$$\epsilon_A - \epsilon_B = f, \tag{2.45}$$

wobei

$$f = \frac{a_B - a_A}{\frac{1}{2}(a_A + a_B)} \tag{2.46}$$

die auf die mittlere Gitterkonstante bezogene Fehlpassung ist.

Einsetzen von Gleichung (2.45) in Gleichung (2.44) und Minimieren nach dem freien Parameter ϵ_A ergibt die Beziehung [124]

$$G_A h_A \epsilon_A = -G_B h_B \epsilon_B. \tag{2.47}$$

Sie gibt drei Eigenschaften des kohärenten Zustandes wieder, die auch schon der direkten Anschauung entspringen (Abbildung 2.3): Die Einzelschichten sind mit ent-

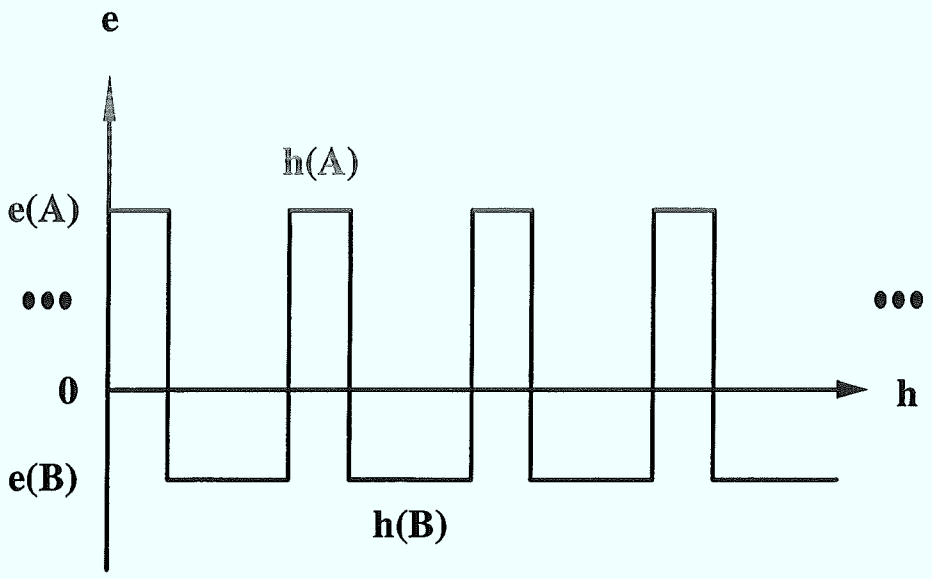


Abbildung 2.3: Dehnungen der Einzelschichten im kohärenten Übergitter. Bei gleichen Schubmoduln sind die Dehnungen umgekehrt proportional zu den Einzelschichtdicken.

gegengesetztem Vorzeichen verspannt, unter sonst gleichen Bedingungen wird das härtere Material weniger verzerrt als das weichere, und die dickere Schicht wird weniger verzerrt als die dünnere.

2.6.2 Relaxation freier Übergitter

Der Gleichgewichtszustand des freien kohärenten Übergitters, mit Verzerrungen ϵ_A^0, ϵ_B^0 der Einzelschichten A, B, wird im folgenden als Ausgangspunkt genommen für den Fall, daß das System durch Vergrößern der Schichtdicken h_A, h_B überkritisch wird, und deshalb Fehlpassungsversetzungen an den Grenzflächen eingebaut werden.

Schon ohne eine genaue Analyse des Relaxationsgrades lassen sich drei wichtige Eigenschaften des Gleichgewichtszustandes nach der Relaxation ableiten. Die erste betrifft das Vorzeichen der Stufenkomponenten der Burgersvektoren, die in den A–B- und den B–A-Grenzflächen liegen, wobei die Schichtabfolge gemäß der Wachstumssequenz gezählt ist. Für eine positive relative Fehlpassung enthält die Schicht B nach einer teilweisen Relaxation mehr Gitterebenen pro Längeneinheit als die Schicht A (Abbildung 2.4). Ein Burgersumlauf um eine Versetzung in der A–

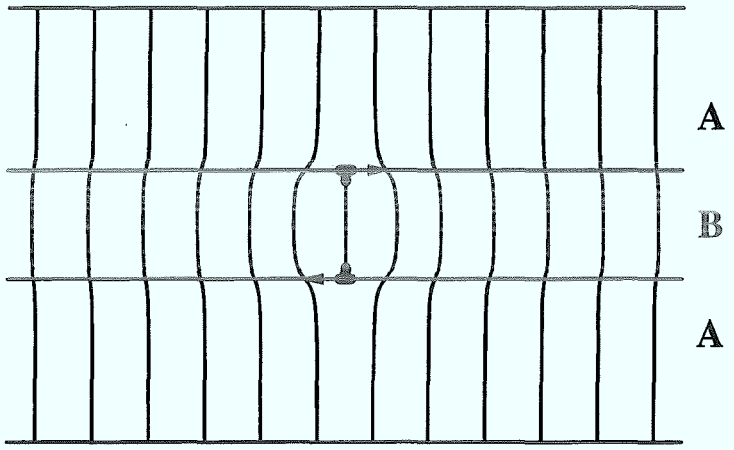


Abbildung 2.4: Eingeschobene Ebene einer relaxierten Einzelschicht. Die Fehlpassungsversetzungen an den A–B- und B–A-Grenzflächen haben entgegengesetzte projizierte Stufenkomponenten.

B-Grenzfläche ergibt also einen Schließungsvektor der das genau entgegengesetzte Vorzeichen des Schließungsvektors eines Burgersumlaufs um eine Versetzung in der B–A-Grenzfläche hat. Durch Fortsetzen der Übergitterperiode entlang der Wachstumsrichtung sieht man, daß sich stets „positive“ und „negative“ Versetzungsnetzwerke abwechseln. Das gleiche Ergebnis folgt natürlich auch für eine negative relative Fehlpassung.

Die zweite Eigenschaft betrifft die Dichte der Fehlpassungsversetzungen in den

A–B- und den B–A-Grenzflächen. Wegen der Periodizität der Schichtabfolge des Übergitters entlang der Wachstumsrichtung müssen das Spannungs- und das Verzerrungsfeld im Gleichgewicht die gleiche Periodizität besitzen. Insbesondere sind dann auch die homogenen Restdeformationen in benachbarten Perioden $i, i+1$ gleich:

$$\epsilon_{A,i} = \epsilon_{A,i+1}, \tag{2.48}$$

$$\epsilon_{B,i} = \epsilon_{B,i+1}. \tag{2.49}$$

Da andererseits die Differenzen $\epsilon_{A,i} - \epsilon_{B,i}$ und $\epsilon_{A,i+1} - \epsilon_{B,i}$, jeweils abzüglich der Fehlpassung f, proportional den Versetzungsdichten an den Grenzflächen sind, ist

$$S_i = S_{i+1}. \tag{2.50}$$

Die Versetzungsabstände in den benachbarten „positiven" und „negativen" Versetzungsnetzwerken sind also gleich groß.

Die dritte Eigenschaft betrifft die relative Lage der Versetzungslinien zweier benachbarter Fehlpassungsnetzwerke. Nach dem Superpositionsprinzip der linearen Elastizitätstheorie ist die totale freie Energie einer Konfiguration aus zwei Versetzungen gegeben durch die Summe der Selbstenergien der beiden Versetzungen (Gleichungen (2.18), (2.22)), die ja unabhängig von der relativen Lage der Linien sind, plus der Wechselwirkungsenergie (Gleichung (2.5)), die durch das Verzerrungsfeld der einen und das Verspannungsfeld der anderen Versetzung gegeben ist, und die noch von der Position der Linien abhängt.

Die Versetzungen der jeweiligen Netzwerke haben unterschiedliches Vorzeichen, so daß die Wechselwirkungsenergie um so kleiner wird, je kleiner der Abstand zwischen den Linien ist. Im Grenzfall des Abstands Null annihilieren die Felder und die Wechselwirkungsenergie wird Null. In Übergittern sind die Versetzungen jedoch an den Ort der Grenzfläche gebunden, so daß die energetisch günstigste Position einen Minimalabstand von h_A oder h_B hat, bei der die Versetzungslinien, in Wachstumsrichtung gesehen, übereinander liegen. Eine solche Versetzungskonfiguration wird analog zur Elektrostatik „Dipol" genannt [83].

Um den quantitativen Unterschied zum Fall der fehlgepaßten Einzelschicht deutlich zu machen, wird die kritische Schichtdicke für ein einachsig verspanntes Übergitter berechnet, dessen Einzelschichten gleiche Dicken h und gleiche Materialkonstanten haben sollen. Aus Gleichung (2.47) folgt, daß das kohärente Übergitter symmetrisch verspannt ist, mit $\epsilon_A = -\epsilon_B$. Die freie Energie pro Übergitterperiode besteht aus dem Beitrag

$$F_L = \frac{Gh}{1-\nu}\epsilon_A^2 + \frac{Gh}{1-\nu}\epsilon_B^2 \tag{2.51}$$

der homogenen Verspannung der Einzelschichten, und aus dem Beitrag

$$F_D = \frac{Gb(f-(\epsilon_A-\epsilon_B))}{2\pi(1-\nu)}\left(\ln\frac{h}{2b}+1\right) \tag{2.52}$$

der beiden, aus Stufenversetzungen bestehenden Fehlpassungsnetzwerke, mit $F_D = 2F_{Stufe}/S$, $S > h$, und

$$S = \frac{b}{f-(\epsilon_A-\epsilon_B)}. \tag{2.53}$$

Für den oberen Abschneideradius R der Versetzungen ist der Wert $h/2$ einzusetzen, da die Verzerrungsfelder jetzt nicht durch die Kristalloberfläche, beziehungsweise durch eine Bildversetzung im Abstand $2h$ abgeschnitten werden, sondern durch die darüber- und darunterliegenden Versetzungen der benachbarten Netzwerke.

Das Minimieren der Ausdrücke nach den freien Parametern ϵ_A, ϵ_B ergibt wieder eine symmetrische Restdehnung des Übergitters mit

$$\epsilon = \epsilon_A = -\epsilon_B = \frac{b}{4\pi h}\left(\ln\frac{h}{2b} + 1\right). \tag{2.54}$$

Es ergibt sich der gleiche Zusammenhang wie bei der Relaxation der fehlgepaßten Einzelschicht (Gleichung (2.31)), bis auf einen Faktor „2" vor dem Logarithmus und einen halb so großen Abschneideradius. Wenn man also vom geänderten Abschneideradius absieht, so hat das freie Übergitter eine etwa *viermal* höhere kritische Fehlpassung der einzelnen Schichten, als eine Einzelschicht auf einem dicken Substrat, bei gleicher Einzelschichtdicke h [123]. Dazu trägt nicht nur der zusätzliche Faktor „2" vor dem Logarithmus bei, sondern auch die Tatsache, daß schon im kohärenten freien Übergitter beide Schichten A und B durch tetragonale Dehnung relaxieren, während im Fall der Einzelschicht das starre Substrat nicht dazu beiträgt.

Die Erweiterung auf den Fall der Relaxation eines *zweiachsig* verspannten Übergitters, auf unterschiedliche Schubmoduln G_A, G_B, und auf nicht zur Relaxation beitragende Burgersvektorkomponenten führt zu den exakt gleichen Korrekturfaktoren (2.36), (2.38), (2.39), (2.40) wie im Fall der fehlgepaßten Einzelschicht.

Haben die Schichten A und B unterschiedliche Dicken, dann überlagern sich die Verzerrungsfelder der Fehlpassungsversetzungen in den A–B- und den B–A-Grenzflächen so, daß sich jeweils die Felder der nächstliegenden „positiven" und „negativen" Versetzungen möglichst kompensieren. In diesem Fall ist für den äußeren Abschneideradius der Wert $\bar{h}/2$ einzusetzen, mit $\bar{h} = \min(h_A, h_B)$. Wegen der unterschiedlichen Schichtdicken sind dann auch die Restdehnungen unterschiedlich, die im Gleichgewicht, analog zu Gleichung (2.54) die Werte

$$\epsilon_A = \frac{b}{4\pi h_A}\left(\ln\frac{\bar{h}}{2b} + 1\right) \tag{2.55}$$

und

$$\epsilon_B = -\frac{b}{4\pi h_B}\left(\ln\frac{\bar{h}}{2b} + 1\right) \tag{2.56}$$

annehmen. Die beiden Gleichungen folgen aus der Minimierung der Gesamtenergie nach ϵ_A und ϵ_B, wobei die Schicht mit der positiven Verzerrung mit „A" und die mit der negativen Verzerrung mit „B" bezeichnet wurde. Selbstverständlich kann man in einem solchen Übergitter nicht mehr von einer einzigen „kritischen Schichtdicke" sprechen, sondern für einen gegebenen Satz kritischer Schichtdicken $(h_{A,c}, h_{B,c})$ muß mit den Gleichungen (2.55), (2.56) die Kohärenzbedingung (2.45)

$$f = \frac{b}{4\pi h_{A,c}}\left(\ln\frac{\bar{h}_c}{2b} + 1\right) + \frac{b}{4\pi h_{B,c}}\left(\ln\frac{\bar{h}_c}{2b} + 1\right) \tag{2.57}$$

erfüllt sein. Wenn die aus dieser Gleichung errechnete Fehlpassung größer als die tatsächlich im System vorliegende Fehlpassung ist, dann ist das System unterkritisch und die Einzelschichten bleiben kohärent, ansonsten relaxiert es. Dieses Kriterium ist analog zum Kriterium im Fall der relaxierten Einzelschicht, wo das System kohärent bleibt, solange die aus Gleichung (2.31) ermittelte homogene Restdehnung größer als die tatsächlich vorliegende Fehlpassung ist. Mit der Gleichung (2.57) ist ein Satz kritischer Schichtdicken dadurch gekennzeichnet, daß f genau gleich der tatsächlichen Fehlpassung ist.

2.6.3 Endliche Übergitter

Praktisch realisierte Übergitter bestehen aus einer endlichen Anzahl von Übergitterperioden, die auf ein Substrat aufgewachsen werden. Dadurch wird die periodische Randbedingung ersetzt durch zwei neue Randbedingungen, nämlich die Einspannung des Übergitters durch ein starres Substrat an der ersten Periode und die *freie Oberfläche* der letzten Periode.

Relaxierte Übergitter haben im Vergleich zu relaxierten Einzelschichten ein weniger dichtes Netzwerk von Fehlpassungsversetzungen (Gleichung (2.54)), und die Versetzungen der alternierenden „positiven" und „negativen" Netzwerke bilden *Dipolkonfigurationen*, deren Verspannungs- und Verzerrungsfelder Ausdehnungen von der Größenordnung des Abschneideradius $\bar{h}/2$ haben.

Die Bildkräfte auf die Versetzungsdipole der letzten Übergitterperiode sind im allgemeinen zu vernachlässigen, weil die Dipolfelder proportional zu $1/r^2$ mit dem Abstand r vom Dipol abfallen [83], und deshalb die Bedingung (2.4) der freien Oberfläche praktisch erfüllt ist, wenn der Abstand des Dipols von der Oberfläche mindestens von der Größenordnung $\bar{h}$ ist. Wenn der Abstand kleiner ist, oder das Übergitter statt mit einer vollständigen Zweischichtperiode nur mit einer Einzelschicht endet, dann wird das oberste Versetzungsnetzwerk eine gegenüber den „inneren" Netzwerken verringerte Dichte haben. In Analogie zum Fall der relaxierten Einzelschicht ist für den Abschneideradius des obersten Netzwerkes statt $\bar{h}/2$ die Dicke der terminierenden Schicht einzusetzen.

Die Gesamtenergie des Übergitters wird jedoch im überwiegenden Maße durch die Energie der inneren Perioden bestimmt, so daß die obige Änderung der Versetzungsdichte der terminierenden Schicht nur einen kleinen Einfluß auf die Lage des Gleichgewichts hat. Dies ist auch anschaulich klar, da die Dipolfelder sämtlicher innerer Perioden wegen des noch größeren Abstandes zur Oberfläche auf Null abgefallen sind, und dementsprechend auch keine Korrektur der Abschneideradien vorgenommen werden muß.

Das epitaktische Wachstum des gesamten Übergitters auf ein dickes Substrat mit der Gitterkonstanten a_S und dem Schubmodul G_S hat demgegenüber einen viel massiveren Einfluß auf die Gesamtenergie. Ähnlich zum Fall der fehlgepaßten Einzelschicht hängt die Größe der Energieänderung vom Gitterkonstantenunterschied zwischen Übergitter und Substrat ab. Da das Übergitter selbst aus Einzelschichten mit unter Umständen verschiedenen Gitterkonstanten a_A, a_B besteht, und diese Einzelschichten nach dem Modell des freien Übergitters schon gegeneinander relaxiert

sein können, ist es sinnvoll, zuerst eine „mittlere" Gitterkonstante des Übergitters zu definieren, und diese dann zur Definition einer Fehlpassung des Übergitters relativ zum Substrat, der *Substratfehlpassung*, zu benutzen.

Die mittlere Gitterkonstante ist das mit dem jeweiligen Schubmodul und der jeweiligen Schichtdicke gewichtete arithmetische Mittel der einzelnen Ruhegitterkonstanten a_A, a_B:

$$a_{SL} = \frac{G_A h_A a_A + G_B h_B a_B}{G_A h_A + G_B h_B}. \tag{2.58}$$

Diese Definition berücksichtigt, daß die Gitterkonstante des Übergitters im mechanischen Gleichgewicht stärker von einer härteren oder dickeren Einzelschicht bestimmt wird, als von einer weicheren oder dünneren Einzelschicht. Sie ist außerdem sowohl für den Fall des kohärenten, tetragonal verspannten Übergitters, als auch für ein relaxiertes Übergitter gültig, was durch Ersetzen von a_A durch $(1 + \epsilon_A)\, a_A$ und a_B durch $(1 + \epsilon_B)\, a_B$ unter Berücksichtigung der Gleichungen (2.47), (2.55) und (2.56) folgt.

Mit dieser Definition erhält die Substratfehlpassung den Wert

$$f_{SL} = \frac{a_{SL} - a_S}{a_S}. \tag{2.59}$$

Es ist nun wichtig zu bemerken, daß der Einfluß des Substrats keinen grundsätzlich neuen Verspannungszustand erzeugt. Aus dem Superpositionsprinzip (2.5) folgt nämlich, daß die Substratfehlpassung f_{SL} analog zum Fall der fehlgepaßten Einzelschicht durch eine *homogene* tetragonale Verzerrung des Übergitters und, oberhalb der kritischen Dicke $h_{SL,c}$ des Übergitters, durch ein Netzwerk von Fehlpassungsversetzungen an der Substratgrenzfläche angepaßt werden kann. Die relativen Verspannungszustände der Einzelschichten bleiben dabei gleich.

Zwei Ursachen sind für diese, auf den ersten Blick überraschend einfache Situation verantwortlich. Erstens ist die Versetzungsdichte an den A–B- und den B–A-Grenzflächen nur von der Differenz der Dehnungen ϵ_A, ϵ_B abhängig (Gleichung (2.53)). Wenn also einem relaxierten, freien Übergitter zusätzlich eine homogene Dehnung ϵ_{SL} aufgeprägt wird, so werden die Dehnungen der Einzelschichten transformiert nach [124]:

$$\epsilon_A \rightarrow \epsilon_A + \epsilon_{SL}, \tag{2.60}$$

$$\epsilon_B \rightarrow \epsilon_B + \epsilon_{SL}. \tag{2.61}$$

Damit bleibt auch die Differenz gleich, die Versetzungsdichten ändern sich nicht, und der Energiebeitrag der Versetzungsnetzwerke bleibt bei der Transformation konstant.

Zweitens hat die Wechselwirkungsenergie zwischen der periodisch alternierenden, homogenen Verspannung der Einzelschichten A und B (Gleichungen (2.47), (2.55) und (2.56)) und der homogenen Verzerrung ϵ_{SL} exakt den Wert Null.

Nachdem mit Hilfe des Superpositionsprinzips gezeigt wurde, daß der Energiebeitrag der homogenen Dehnung einfach zur Energie des freien Übergitters addiert werden kann, entkoppelt das Problem der Relaxation eines fehlgepaßten, auf ein

Substrat aufgewachsenen Übergitters einfach in die Relaxation eines freien Übergitters und in die Relaxation einer einzelnen fehlgepaßten Schicht, mit einer Dicke gleich der Gesamtdicke h_{SL} des Übergitters, und einer Fehlpassung f_{SL}.

Die für die Relaxation des gesamten Übergitters relativ zum Substrat wichtigen Größen der kritischen Dicke $h_{SL,c}$ und der Restdehnung $\epsilon_{SL,c}$ ergeben sich dann aus den schon abgeleiteten Gleichungen für die fehlgepaßte Einzelschicht, mit den entsprechenden Ersetzungen für f, ϵ und h und den mittleren Schubmodul des Übergitters. Abbildung 2.5 zeigt schematisch die Abnahme der Restdehnung ϵ_{SL} vom voll

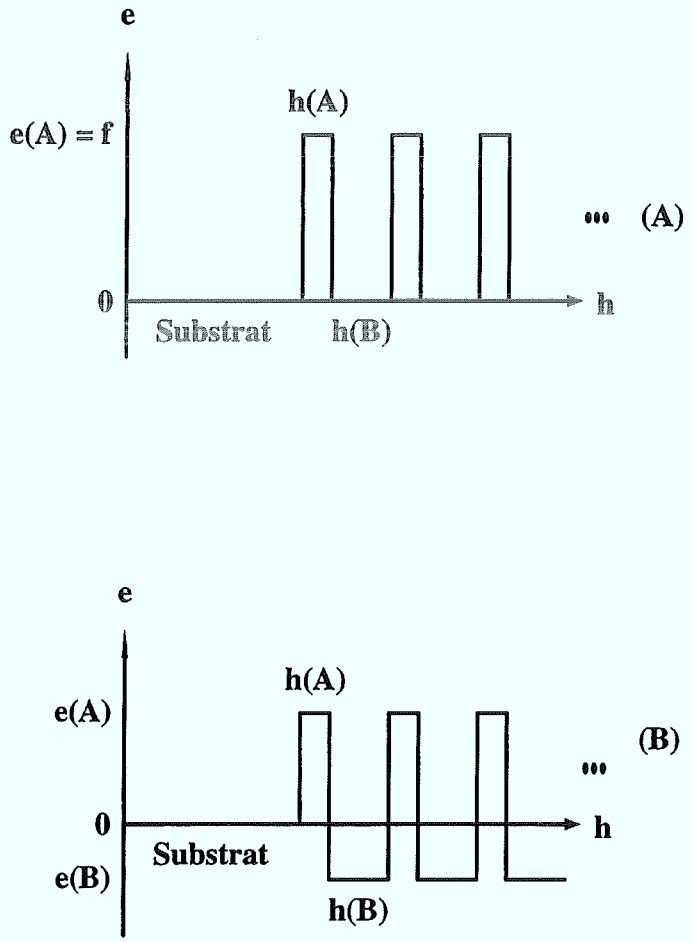


Abbildung 2.5: Abnahme der Restdehnung eines Übergitters während der Relaxation. (A) Zustand vollständiger Verspannung, (B) teilweise Relaxation des Übergitters.

verspannten Zustand (A) zu einem teilweise relaxierten Zustand (B) des Übergitters relativ zum Substrat. Zusätzlich sind noch die Dehnungen der Einzelschichten ϵ_A, ϵ_B eingetragen.

Im allgemeinen enthalten durch epitaktisches Wachstum hergestellte Übergitter sehr viele Perioden, und wenn die mittlere Gitterkonstante des Übergitters nicht sehr exakt mit der des Substrats übereinstimmt, dann ist die Gesamtdicke meist sehr viel größer als die entsprechende kritische Dicke des Übergitters. Das zugehörige Versetzungsnetzwerk an der Substratgrenzfläche ist relativ dicht, mit $S < h_{SL}$, und die Restdehnung ϵ_{SL} nimmt Werte nahe Null an.

2.7 Kinetik der Relaxation

Die bis zu diesem Punkt dargestellten Relaxationsmodelle beschreiben ausschließlich Gleichgewichtszustände, die durch das Minimum der freien Energie der Schichtsysteme gekennzeichnet sind. Dabei wurde erstens vorausgesetzt, daß zu jedem Zeitpunkt des Relaxationsprozesses genügend Versetzungen vorhanden sind, die zusammen Fehlpassungsnetzwerke mit einer Dichte bilden, die dem Gleichgewicht entspricht. Zweitens wurde angenommen, daß die Versetzungen durchweg Burgersvektoren und Linienrichtungen besitzen, die mit der Relaxation der durch die Fehlpassungen induzierten Verspannungen verträglich sind. Drittens sollten die Fehlpassungssegmente nach ihrer Entstehung auch am Ort der entsprechenden Grenzfläche verfügbar sein, das heißt, sie müssen sich gegebenenfalls vom Ort ihrer Entstehung zur Grenzfläche bewegen.

Das bisher betrachtete Gleichgewicht ist also der Endpunkt in einer Kette, die in Hinblick auf Versetzungen aus folgenden Teilprozessen besteht:

Erzeugung
→Multiplikation
→Bewegung
→Orientierung
→Gleichgewichtsdichte.

2.7.1 Versetzungserzeugung

Durch eine jahrzehntelange Optimierung der Wachstumsprozesse der verschiedenen Kristallzüchtungsverfahren ist es gelungen, technisch relevante Substratmaterialien wie Silizium und Germanium mit sehr niedriger Versetzungsdichte oder praktisch versetzungsfrei zu wachsen. Substrate auf der Basis von Verbindungshalbleitern, wie Galliumarsenid, Indiumarsenid oder Indiumphosphid weisen eingewachsene Versetzungsdichten in der Größenordnung von 10^4 cm^{-2} auf [51].

Schichtsysteme, die auf solchen Substraten gewachsen werden, haben jedoch oft geometrische Abmessungen und Fehlpassungen durch nichtangepaßte Gitterkonstanten der verwendeten Materialien, die teilweise stark überkritisch bezüglich der Relaxation durch Defekteinbau sind. In solchen Systemen, wie auch im hier untersuchten

InGaAs/GaAs-Schichtsystem, treten schon bei niedrigen Fehlpassungen relativ hohe Versetzungsdichten auf, die nicht mehr allein durch eine Multiplikation der im Substratmaterial vorhandenen Versetzungen erklärt werden können.

In der Literatur über epitaktisch gewachsene Schichtsysteme wird als einzige echte Versetzungsquelle die *Erzeugung von Halbschleifen* an Kristalloberflächen diskutiert [17, 189]. Andere relevante Versetzungs-„quellen", die im nächsten Abschnitt beschrieben werden, erzeugen nämlich nicht neue Versetzungssegmente, sondern multiplizieren lediglich die Segmente einer schon vorhandenen Versetzungskonfiguration.

Im Fall der *homogenen Nukleation* ist der Mechanismus der Halbschleifenerzeugung ein thermisch aktivierter Prozeß. Die Wahrscheinlichkeit pro Zeiteinheit für die Erzeugung einer Versetzungshalbschleife bei einer Temperatur T wird durch die Boltzmannverteilung beschrieben und sie ist proportional zu $\exp(-E_c/kT)$. E_c ist dabei eine für den Defekt typische Aktivierungsenergie.

An der perfekten Oberfläche des Kristalls, auf den eine Spannung τ wirken soll, werde die Energie zur Erzeugung einer kleinen Versetzungshalbschleife mit dem Radius r verwendet, deren Burgersvektor in der Schleifenebene liegt. Dazu muß einerseits der Energiebetrag

$$\frac{2-\nu}{8(1-\nu)} Gb^2 r \left(\ln \frac{4r}{b} - 2\right) \tag{2.62}$$

aufgewendet werden [83], der der Energie des Verzerrungsfeldes der Versetzung entspricht. Andererseits hat die Versetzung durch Überstreichen der Fläche $\frac{1}{2}\pi r^2$ einen Teil der homogenen Verspannung τ abgebaut, wodurch ein Energiebetrag von

$$\frac{1}{2}\pi r^2 \tau b \tag{2.63}$$

frei wird.

Die spontan erzeugte Halbschleife ist natürlich nur dann stabil, wenn die Energiebilanz negativ ist. Durch Minimieren der Gesamtenergie nach dem Radius r, wird ein kritischer Radius r_c gefunden, der die Gleichung

$$\tau = \frac{2-\nu}{8(1-\nu)} \frac{Gb}{\pi r_c} \left(\ln \frac{4r_c}{b} - 1\right) \tag{2.64}$$

erfüllt [127, 121]. Wenn der Radius einer spontan gebildeten Halbschleife größer als r_c ist, dann ist das weitere Wachstum energetisch günstig, und die Versetzungslinie wird weiter vergrößert.

Einsetzen von r_c in (2.62), (2.63) ergibt jedoch für Verspannungen fehlgepaßter Schichtsysteme von der Größenordnung $\tau \approx G/100$ Werte der Aktivierungsenergie E_c, die viel größer als kT sind. Mit anderen Worten, die homogene Nukleation ist bei Wachstumstemperaturen von einigen hundert Grad Celsius ein extrem seltener Prozeß.

In der Literatur [127] werden zwei Ursachen diskutiert, die dennoch eine spontane Nukleation begünstigen können. Die erste ist der Einfluß weiterer Energieterme, die die Energiebilanz zugunsten der Nukleation verschieben können. Während

des epitaktischen Wachstums, insbesondere im zweidimensionalen Wachstumsmodus, ist ständig eine bestimmte Dichte von Oberflächenstufen vorhanden, die mit ihrer zusätzlichen Oberflächenenergie zur Gesamtenergie beitragen. Bei der spontanen Erzeugung der Versetzungshalbschleife kann durch den Stufenanteil der Versetzung eine Oberflächenstufe passender Orientierung entfernt werden, und die dabei freiwerdende Oberflächenenergie trägt zur Stabilisierung der Schleife bei [127].

Schon eine grobe Abschätzung der Oberflächenenergie für ein gegebenes Material zeigt jedoch, daß dieser kleine zusätzliche Beitrag den kritischen Radius r_c und damit auch die Aktivierungsenergie E_c bei realistischen Spannungen τ noch nicht weit genug verschiebt.

Die zweite Möglichkeit, den Energiegewinn (2.63) zu vergrößern, sind lokale Spannungskonzentrationen, die während des epitaktischen Wachstums durch die Anwesenheit von Verunreinigungen, Ausscheidungen oder massiven Schädigungen der Kristalloberfläche, wie zum Beispiel Mikrorissen entstehen [12].

2.7.2 Versetzungsmultiplikation

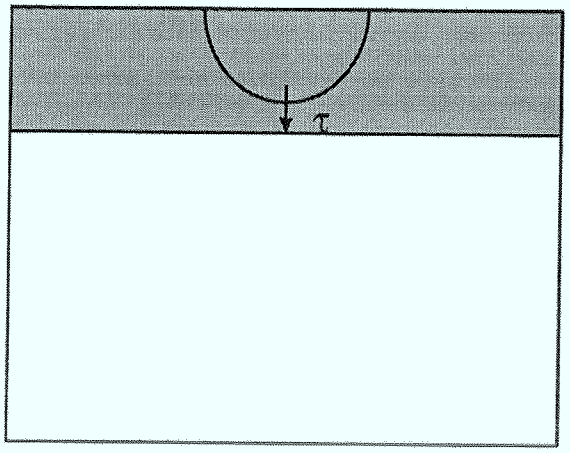
Die Multiplikation von neu erzeugten, oder schon im Kristall vorhandenen Versetzungen beruht grundsätzlich auf der Ausdehnung von Versetzungssegmenten spezieller Versetzungskonfigurationen unter Einwirkung einer äußeren Spannung.

Im vorliegenden heteroepitaktischen Schichtsystem niedriger Fehlpassung dominieren drei Multiplikationsmechanismen. Der erste ist das Wachstum von durch spontane Nukleation erzeugten Versetzungsschleifen oberhalb des kritischen Radius r_c. Aus der Geometrie der Versetzungsschleife folgt, daß die Versetzungslinie auf eine totale Länge von πr wächst, wenn die Halbschleife eine Fläche von $1/2\ \pi r^2$ überstreicht. Das Wachstum der Schleife wird aufrecht erhalten, solange die Versetzungslinie Gebiete des Kristalls überstreicht, die unter der Spannung τ stehen.

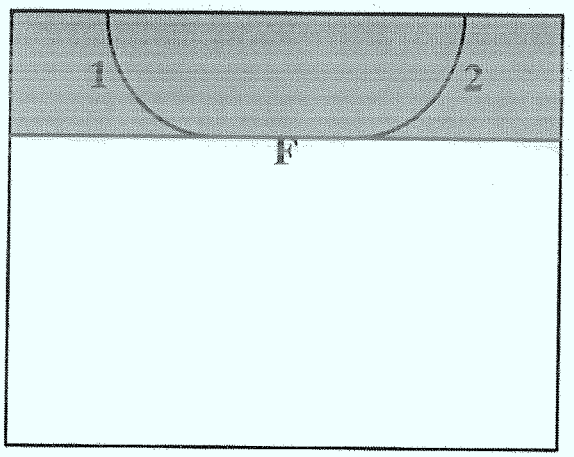
Bei der Relaxation einer fehlgepaßten Einzelschicht führt diese Bedingung zum Abbruch des Wachstums, wenn ein Punkt der Versetzungslinie die Substratgrenzfläche erreicht hat, an der die äußere Spannung, die durch die Fehlpassung verursacht wird, auf Null abfällt. Das Schleifenwachstum hinterläßt somit zwei sogenannte *Fadensegmente*, die die aufgewachsene Schicht von der Substratgrenzfläche bis zur Oberfläche der Schicht durchdringen [129] (Abbildung 2.6).

Der zweite Multiplikationsmechanismus ist die Erzeugung von Fehlpassungssegmenten an Grenzflächen durch das Gleiten von Fadensegmenten in der aufwachsenden Schicht durch die Wirkung einer äußeren Spannung τ (Abbildung 2.7). Bei diesem Prozeß wird ein Energiebetrag von $E_{\text{disl}} \times L$ aufgewendet, wenn ein Fadensegment durch die Spannung τ um eine Strecke L abgleitet, mit E_{disl} der Energie pro Längeneinheit einer Versetzung (Gleichungen (2.23), (2.24)). Andererseits wird durch das Überstreichen der Fläche $h \times L$, mit h der Schichtdicke, durch das Gleiten des Fadensegments eine Energie τbhL frei.

Die Berechnung der Energiebilanz führt auf einen quantitativen Zusammenhang zwischen Spannung τ und Schichtdicke h, der exakt mit Gleichung (2.31) übereinstimmt, wenn der lineare Zusammenhang zwischen der Spannung τ und der Dehnung ϵ eingesetzt wird.

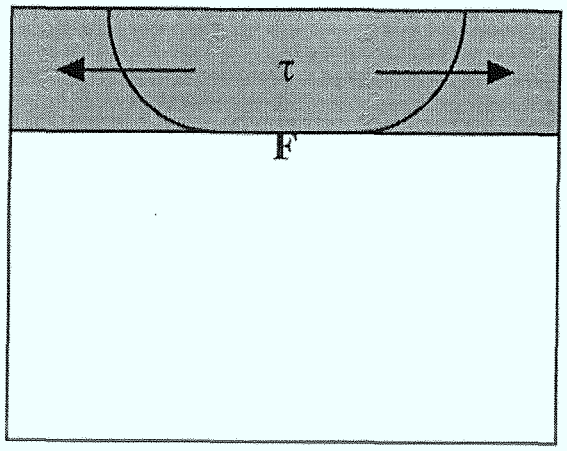


(A)

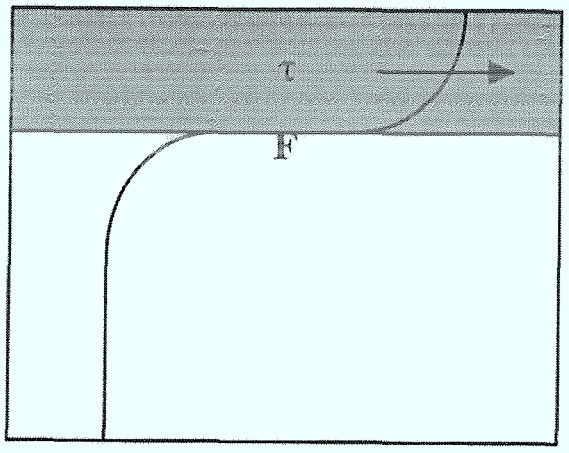


(B)

Abbildung 2.6: (A) Wachstum einer Versetzungshalbschleife unter der Fehlpassungsspannung τ_L. (B) Das Wachstum hinterläßt ein kurzes Fehlpassungssegment (F) und zwei gleitfähige Fadensegmente (1), (2).



(A)



(B)

Abbildung 2.7: Gleiten von Fehlpassungssegmenten unter der Fehlpassungsspannung τ_L. (A) Halbschleifen-Quelle, (B) eingewachsene Versetzung. In beiden Fällen wird das Fehlpassungssegment (F) durch das Abgleiten verlängert.

Ein dritter Mechanismus, die *Hagen-Strunk-Multiplikation* [80, 168], beruht auf der Wirkung dreier Spannungen auf Versetzungslinien an der Grenzfläche zwischen einer dünnen epitaktischen Schicht und einem Substrat. Die erste Spannung ist die durch die Fehlpassung induzierte Spannung τ_f, die zweite ist die Spannung τ_B auf das Fehlpassungssegment, die von der Anwesenheit einer virtuellen Bildversetzung entgegengesetzten Vorzeichens oberhalb der freien Kristalloberfläche herrührt (Abschnitt 2.4.2).

Ist das Fehlpassungssegment weit genug von anderen Versetzungen entfernt, dann ist die Lage der Versetzung stabil, wenn $\tau_f = \tau_B$. Diese Bedingung ist im Energiegleichgewicht eines Netzwerkes von Fehlpassungsnetzwerken erfüllt.

Falls das Netzwerk jedoch nicht genügend dicht ist, $S \gg h$, und die Versetzungslinie eine andere Versetzung des zu ihr orthogonalen Netwerkes kreuzt, dann wirkt bei parallelen Burgersvektoren am Versetzungsknoten eine zusätzliche repulsive Spannung τ_{WW} zwischen den Linien, so daß $\tau_f < \tau_B + \tau_{WW}$.

Unter der so entstandenen Spannungsdifferenz wird eine Hälfte der Überkreuzung am Versetzungsknoten abgelöst und zur nahen Kristalloberfläche getrieben, wo die Versetzungslinie unterbrochen wird und in zwei Fadensegmente zerfällt (Abbildung 2.8). Das Ergebnis ist eine rechtwinklig abknickende Fehlpassungsversetzung

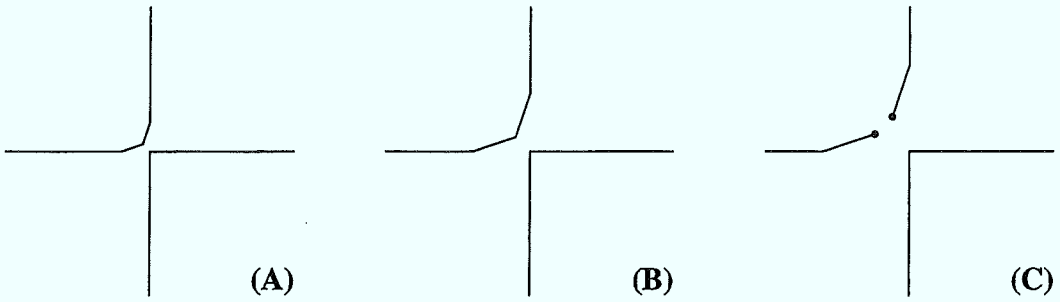


Abbildung 2.8: Hagen-Strunk-Multiplikation an einer Überkreuzung zweier Fehlpassungsversetzungen. (A) Ablösen eines rechtwinkligen Segmentes vom Versetzungsknoten, (B) Abgleiten des Segmentes zur Kristalloberfläche, (C) Bildung zweier Fadensegmente.

nahe der Grenzfläche, sowie die beiden neu erzeugten Fadensegmente, die unter der

Wirkung des obigen Multiplikationsmechanismus weitere Fehlpassungssegmente erzeugen können.

An dieser Stelle wird zur Unterstützung der späteren Diskussion hervorgehoben, daß die ersten beiden Multiplikationsmechanismen im Prinzip sowohl in dünnen als auch in dicken Schichten aktiv sein können, während die Hagen-Strunk-Multiplikation wegen des starken Abfalls der Bildspannung τ_B mit der Schichtdicke nur bei kleinen Schichtdicken h arbeitet.

2.7.3 Bewegung von Versetzungen

Die Erzeugung und die Multiplikation von Versetzungen enthält immer auch die Bewegung eines Teils der Versetzungslinie. Bei der Ableitung der Relaxationsmodelle im Rahmen der linearen isotropen Elastizitätstheorie wurde bisher an jeder Stelle vorausgesetzt, daß eine für einen Relaxationsmechanismus notwendige Bewegung auch jederzeit möglich ist.

Reale Kristalle sind praktisch nie isotrop, und es ist aus energetischen Gründen günstig, die Translationssymmetrie eines Kristalls soweit wie möglich zu erhalten, so daß bei einer Deformation nur eine diskrete Anzahl von Gleitsystemen angeregt sind, die eine beschränkte Anzahl von Burgersvektoren und Gleitebenen enthalten. Diese Beschränkung wurde in den Modellen durch die Korrekturfaktoren (2.36), (2.39), (2.40) berücksichtigt.

Bei der Bewegung der Versetzung über die Gleitebene wirken zusätzliche Spannungen auf die Versetzungslinie, die ihre Ursache in der Kristallperiodizität haben. Im Rahmen der isotropen Elastizitätstheorie ist es ja bei der Bildung einer Versetzung durch einen Volterra-Prozeß (Abschnitt 2.4) gleichgültig, an welcher Stelle des Kontinuums eine zusätzliche Halbebene der Dicke b, im Falle einer Stufenversetzung, eingeschoben wird. Gerade bei kovalenten Bindungen in Halbleitern sind jedoch nicht alle Orte äquivalent, so daß es einen mikrostrukturellen Unterschied macht, ob der Versetzungskern zum Beispiel „auf" einer mit Atomen besetzten Ebene liegt, oder „dazwischen".

Diesem strukturellen Unterschied entspricht auch ein Unterschied der Energie des Versetzungskerns, dessen Ursache die in den genannten Fällen unterschiedlichen Bindungsverhältnisse zwischen Atomen „oberhalb" und „unterhalb" der Gleitebene sind. Maxima und Minima des zusätzlichen Energiebeitrags folgen alternierend aufeinander und dieser hat somit eine Periodizität, die gleich dem Atomebenenabstand in der Gleitebene parallel zum Burgersvektor ist.

Der periodische Energiebeitrag und die damit verbundene periodische Spannung wurde von verschiedenen Autoren abgeschätzt [66, 136, 130, 78]. Eine exakte quantitative Bestimmung hat sich als schwierig erwiesen, da sie von vielen Einflußfaktoren wie zum Beispiel der genauen Linienrichtung der Versetzung, der Temperatur oder von Dotierstoffkonzentrationen zum Teil empfindlich abhängt. Eine verbreitete und für Diskussionen ausreichende Darstellung ist die Größe der Amplitude der periodischen Spannung in der Form von Peierls und Nabarro [136, 130]:

$$\tau_P = \frac{2G}{1-\nu} \exp\left(-2\pi \frac{w}{b}\right). \tag{2.65}$$

Der Parameter w bezeichnet die Breite des Verzerrungsfeldes der Versetzung, die in Halbleitermaterialien gut mit einem Wert von wenigen Burgersvektorlängen abgeschätzt wird. Der Vorfaktor enthält im Prinzip noch eine zusätzliche Temperaturabhängigkeit, die für kubische Kristalle von Dietze [29] und Kuhlmann-Wilsdorf [108] abgeschätzt wurde.

Bei allen Unsicherheiten der zugrundeliegenden Annahmen dieser Arbeiten ist für die spätere Diskussion nur wichtig, daß die Spannung τ_P in den kovalent gebundenen Halbleitermaterialien einen experimentell gefundenen Wert von bis zu $10^{-2}G$ haben kann. Das Versetzungsgleiten ist ein thermisch aktivierter Prozeß, und die Versetzungsgeschwindigkeit v läßt sich durch die empirische Gleichung

$$v = v_0 \left(\frac{\tau_{\text{eff}}}{\tau_0}\right)^m \exp\left(-\frac{E_v}{kT}\right) \tag{2.66}$$

beschreiben [88], mit den materialabhängigen Parametern v_0 und τ_0, dem Spannungsexponenten m, der Aktivierungsenergie E_v des Versetzungsgleitens, in die auch die Peierls-Energie eingeht, der Boltzmann-Konstanten k und der absoluten Temperatur T. Der Spannungsexponent, v_0 und die Aktivierungsenergie haben für $\tau_0 = 1$ MPa in Galliumarsenid die Werte $m = 1.6$, $v_0 = 3 \times 10^3$ ms^{-1}, $E_v = 1.4$ eV und in Indiumarsenid die Werte $m = 1.4$, $v_0 = 4 \times 10^4$ ms^{-1}, $E_v = 1.4$ eV [194].

2.7.4 Orientierung von Versetzungen

Die im vorigen Abschnitt genannten Spannungen, die die Bewegung von Versetzungen hemmen können, sind im allgemeinen selbst wieder abhängig von der Bewegungsrichtung der Versetzung, das heißt sowohl von der Linienrichtung als auch von der Richtung des Burgersvektors.

Solange die Versetzung sich in einem Kristallvolumen bewegt, das einer homogenen Spannung τ ausgesetzt ist, wird die Linienrichtung während des Gleitens die energetisch günstigste Lage mit der kleinsten Peierls-Nabarro-Spannung einnehmen. Im vorliegenden Fall trifft dies in erster Linie auf Fadensegmente zu, die in den homogen gedehnten Einzelschichten gleiten, während die [110]- und [$\bar{1}$10]-Linienrichtungen der Fehlpassungssegmente ja schon geometrisch durch die Schnittgeraden zwischen den {111}-Gleitebenen und den (001)-Grenzflächen festgelegt sind.

2.7.5 Ergänzungen der Gleichgewichtsmodelle

Aus den Erläuterungen zu Erzeugung, Multiplikation und Bewegung von Versetzungen geht hervor, daß jeder der Teilprozesse die Relaxation behindern kann, so daß die Gleichgewichtsdichte nicht in jedem Fall nach Abschluß des Wachstumsprozesses erreicht werden muß. Auch wenn grundsätzlich für große Zeiten t jeder Teilprozeß zum Gleichgewicht führt, so kann die durch die Schichtherstellung limitierte Wachstumszeit t_W zu einem Nichtgleichgewicht des epitaktischen Systems führen.

Parallel zur Entwicklung der Gleichgewichtsmodelle wurden deshalb eine Anzahl empirischer Ergänzungen gemacht, die die zeitliche Entwicklung der Relaxation durch Modellieren der mittleren Erzeugungs- und Multiplikationsraten, sowie einer

limitierten mittleren Versetzungsgeschwindigkeit berücksichtigen. Das Ergebnis der einzelnen Ansätze ist eine Reihe physikalischer Größen, die den Endzustand des relaxierten Systems zusätzlich bestimmen. Der Vorteil einer solchen Vorgehensweise ist die Möglichkeit, wichtige Größen, wie Versetzungsdichten oder Restdehnungen, vollständig quantitativ zu erklären. Der Nachteil ist die in vielen Anwendungsfällen vorliegende Unterbestimmtheit des Anpaßproblems, weil zum Beispiel in Bauelementanwendungen nur wenige verschiedene oder sogar nur ein einziges konkretes Schichtsystem gewachsen wird, und unter Umständen ein einziger Meßpunkt zur Ableitung mehrerer Einflußgrößen herangezogen wird.

Die erste Erweiterung betrifft die Annahme einer maximalen Versetzungsgeschwindigkeit, deren Ursache eine material- und temperaturabhängige Reibungskraft ist. Sie resultiert unter anderem aus dem Peierls-Potential, dessen Größe durch die Beweglichkeit eines Versetzungskerns bestimmt ist. In der Untersuchung von Matthews [129] wird ein linearer Zusammenhang zwischen Reibungskraft und Geschwindigkeit angenommen. Das Resultat der Analyse ist eine lineare Differentialgleichung für den durch Versetzungseinbau relaxierten Anteil der Fehlpassung, $\delta = f - \epsilon$, deren Lösung ein asymptotisches Wachstum der Versetzungsdichte bis auf den Gleichgewichtswert δ_∞ ist:

$$\delta = \delta_\infty \left(1 - \mathrm{e}^{-\beta t}\right). \tag{2.67}$$

β enthält die schon aus den Gleichgewichtsmodellen bekannten Materialkonstanten, die Wachstumstemperatur T, und eine konstante Dichte ρ_a von Fadensegmenten, die zur Multiplikation notwendig sind. Mit Hilfe dieser empirischen Gleichung kann eine vom Gleichgewicht abweichende Dichte von Fehlpassungsversetzungen bei festem β durch eine endliche Wachstumszeit erklärt werden. Gleichzeitig verdeutlicht sie, daß, wegen der Materialabhängigkeit von β, unterschiedliche Schichtsysteme, bei gleichen Wachstumszeiten, verschieden weit vom Gleichgewicht entfernt sein können.

Eine zweite Erweiterung nimmt die obige Berücksichtigung einer geschwindigkeitsabhängigen Reibungskraft zum Ausgangspunkt. In der Arbeit von Dodson und Tsao [31, 32] wird jedoch keine konstante, sondern eine spannungsabhängige Versetzungsgeschwindigkeit (Gleichung (2.66)) angenommen, wobei die in die Geschwindigkeit eingehende „wahre" Spannung τ_{eff} proportional zum Unterschied zwischen der aktuellen Versetzungsdichte und dem Gleichgewichtswert δ_∞ ist.

Zusätzlich wird eine Multiplikationsrate angenommen, die proportional zur wahren Spannung, zur Versetzungsgeschwindigkeit und zur schon vorhandenen Versetzungsdichte ist. Der Multiplikationsterm enthält außerdem noch eine „Anfangsdichte" δ_0 von Versetzungen, die sowohl die Dichte der schon vor dem Schichtwachstum im Substrat vorhandenen Versetzungen, als auch die Dichte der während des Wachstums aktiven Versetzungsquellen berücksichtigt.

Die resultierende Differentialgleichung für ϵ ist nichtlinear, und ihre Lösung ist für kleine Fehlpassungen von bis zu $f = 1\%$ mit der Lösung von Matthews identisch, während für größere Fehlpassungen wesentlich niedrigere Werte der Restdehnung ϵ gefunden werden. Der Grund dafür ist die Abhängigkeit der Multiplikationsrate von der Versetzungsdichte und der wahren Spannung, die ja beide für kleine Fehlpassungen klein sind.

2.8 Äquivalenz der Energiemethode und des Kräftegleichgewichts

Die Entwicklung von Relaxationsmodellen ist in der Vergangenheit zwei unterschiedlichen, jedoch im Rahmen der klassischen Mechanik äquivalenten Wegen gefolgt [68, 93]. Die erste Gruppe von Modellen beruht auf der Berechnung der freien Energie eines Systems und der anschließenden Minimierung nach freien Parametern, die in erster Linie mit der Geometrie eines gegebenen Schichtsystems verknüpft sind. Diese Energiemethode, die auch in den ersten Abschnitten dieses Kapitels zur Ableitung von kritischen Schichtdicken und Restdehnungen von fehlgepaßten Einzelschichten und Übergittern benutzt wurde, gründet direkt auf den Grundgleichungen der linearen Elastizitätstheorie.

Bei einer zweiten Gruppe von Modellen werden verschiedene, auf eine Versetzungslinie wirkende mechanische Kräfte betrachtet. Die Kräfte hängen von freien Parametern ab, die wieder mit der Geometrie eines gegebenen Schichtsystems verknüpft sind, und die Berechnung des Kräftegleichgewichts ergibt dann Ausdrücke für kritische Schichtdicken und Restdehnungen.

Die Ausdrücke für die einzelnen Kräfte auf eine Versetzungslinie werden bei dieser *Methode des Kräftegleichgewichts* jedoch vorher mittels des mechanischen *Prinzips kleiner Verrückungen* aus den zugehörigen Termen für die freie Energie berechnet. Zu diesem Zweck wird der Energieunterschied $\mathrm{d}E$ des Systems für zwei Positionen $\vec{x}$ und $\vec{x}+\mathrm{d}\vec{x}$ des Versetzungskerns berechnet, und die Größe $f_i = -\frac{\mathrm{d}E}{\mathrm{d}x_i}$ als Komponente i, $i = 1,2,3$, der Kraft pro Längeneinheit der Linie interpretiert.

Diese Vorgehensweise erspart gegenüber der Energiemethode natürlich keine Rechenarbeit, da ja auch hier zuerst die Ausdrücke für die freie Energie berechnet werden müssen.

Die Methode des Kräftegleichgewichts wird erst gleich effizient wie die Energiemethode, wenn die Kräfte nicht aus den freien Energien, sondern aus den ihnen zugrundeliegenden Spannungsfeldern errechnet werden. Das Prinzip der kleinen Verrückungen kann nämlich statt auf ein konkretes Spannungsfeld eines gegebenen Schichtsystems auf eine Versetzung angewendet werden, die sich in einem beliebigen äußeren Spannungsfeld σ_{ik} befindet. Das Resultat ist die Peach-Koehler-Gleichung der linearen Elastizitätstheorie [109]

$$f_i = e_{ikl} b_m \xi_k \sigma_{lm}, \tag{2.68}$$

die nur noch von der ortsabhängigen äußeren Spannung abhängt. Die bei der Verrükkung konstanten Größen e_{ikl}, b_m, ξ_k sind die Komponenten des antisymmetrischen Einheitstensors, des Burgersvektors und der Linienrichtung der Versetzung.

Eine Erweiterung der Peach-Koehler-Gleichung, die in der Form (2.68) nur für Gleitprozesse gilt, wurde von Weertman für das Versetzungsklettern eingeführt [183]. Hierbei muß der Wert der äußeren Spannung in (2.68) um den Betrag $-\frac{1}{3}\delta_{lm}\sigma_{nn}$ korrigiert werden. Das Versetzungsklettern ist mit einer Diffusion von Punktdefekten verbunden, die erst bei höheren Wachstumstemperaturen eine Rolle spielt, und es

ist im vorliegenden InGaAs/GaAs-Schichtsystem, das in dieser Arbeit untersucht wurde, nicht beobachtet worden.

Mit Benutzung der Peach-Koehler-Gleichung wird die Methode des Kräftegleichgewichts besonders nützlich in den Fällen, wo neben den Kräften der Verspannungsfelder noch zusätzliche Kräfte, wie zum Beispiel „Reibungskräfte", eingearbeitet werden, die ein empirisches Modellieren der Relaxationskinetik ermöglichen.

Kapitel 3

Experimentelle Methoden

3.1 Probenherstellung

Das in dieser Arbeit untersuchte epitaktische System von InGaAs- und GaAs-Schichten auf GaAs(001)-Substraten hat eine vom Indiumanteil x der $In_xGa_{1-x}As$-Legierung abhängige Fehlpassung. Die Fehlpassung einer Einzelschicht gegenüber dem Substrat kann linear mit zunehmendem Indiumanteil zwischen Null und 7.1 % variiert werden. Für eine Indiumkomposition von $x = 0.6$ liegt die der Bandlückenenergie von InGaAs zuzuordnende Wellenlänge in einem Bereich relativ niedriger Absorption von Glasfasermaterial. Für Bauelementanwendungen ist jedoch in diesem System häufig der Bereich kleiner Indiumanteile von $x < 30$ % interessant, weil erstens die relativ kleine Fehlpassung niedrige Defektdichten zur Folge hat, und weil zweitens in diesem Kompositionsbereich der zweidimensionale Wachstumsmodus vorherrscht, der die Voraussetzung für die Planarität von Bauelementstrukturen und relativ niedrige Grenzflächenrauhigkeiten ist.

Andererseits werden trotz der angestrebten Defektfreiheit in vielen Bauelementanwendungen Schichtdicken benötigt, die oberhalb einer entsprechenden kritischen Schichtdicke liegen. Es ist zwar durch Wahl geeigneter Materialkombinationen möglich, die Fehlpassung zu beseitigen, wie zum Beispiel im System GaAs/AlAs, oder es ist möglich, durch das Wachsen von gitterangepaßten Puffern zwischen Substrat und aufwachsenden Schichten zumindest die Substratfehlpassung zu verkleinern. Im Hinblick auf die elektronischen und optischen Eigenschaften stehen jedoch dadurch interessante Materialkombinationen nicht zur Verfügung, oder das zusätzliche Pufferwachstum erzeugt selbst eine Reihe technologischer Schwierigkeiten.

Bei Übergittern liegt oft die komplexe Situation vor, daß die Einzelschichtdicken zwar überkritisch relativ zum Substrat sind, eine einzelne Übergitterperiode wegen des kohärenten Gleichgewichtszustandes (2.47), (2.57) unterkritisch ist, während das gesamte Übergitter wieder überkritisch relativ zum Substrat ist.

Daraus wird ersichtlich, daß eine Untersuchung der Relaxation von Übergittern den Einfluß einer Vielzahl von Parametern, wie den Einzelschichtdicken, der Gesamtdicke des Übergitters, den Schichtkompositionen, der Wachstumstemperatur, sowie der Wachstumsgeschwindigkeit berücksichtigen muß. Um eine maximale Kontrolle der Randbedingungen zu haben, ist es bei einer systematischen Vorgehensweise rat-

sam, zunächst einen der genannten Parameter zu variieren, und die übrigen konstant zu halten.

Beim in der vorliegenden Arbeit untersuchten Modellsystem wurde aus zwei Gründen nur die Einzelschichtdicke der GaAs-Schichten variiert, bei Konstanthalten der InGaAs-Schichtdicke und der Gesamtdicke des Übergitters. Der erste Grund ist die Realisierung des oben genannten, praktisch wichtigen Falls eines überkritischen Übergitters und überkritischer Einzelschichten, sowie unterkritischer Übergitterperioden.

Der zweite Grund ist die Konzentration auf die Frage nach dem Relaxationsverhalten von Übergittern, während andere, als die weiter unten dargestellten Bedingungen nach Abschätzungen mit Hilfe der in Kapitel 2 dargestellten Modelle auch im Fall des Übergitters auf die Untersuchung des in der Literatur schon ausgiebig diskutierten Relaxationsverhalten einer Einzelschicht hinauslaufen [129, 127, 121, 83, 141, 142, 69, 31, 32].

Im vorliegenden Fall wurde eine Serie von Übergittern hergestellt, mit Einzelschichtdicken der GaAs-Schichten von $h_{\mathrm{GaAs}} = 2$, 6, 10, 16, 20, 25, 30 nm, einer konstanten Einzelschichtdicke der InGaAs-Schichten von $h_{\mathrm{InGaAs}} = 18$ nm, und einer Gesamtdicke des Übergitters von $h_{\mathrm{SL}} = 800$ nm. Der Indiumgehalt der Einzelschichten beträgt in allen Fällen $x = 20$ %. Die Probenserie wurde mittels Molekularstrahlepitaxie bei einer Wachstumstemperatur von 460°C und bei einer Wachstumsgeschwindigkeit von 0.8 μm pro Stunde auf GaAs(001)-Substraten gewachsen.

Für die weitere Diskussion wird eine abkürzende Nomenklatur der Probenbenennung eingeführt: Ein Übergitter mit der Gesamtdicke $h_{\mathrm{SL}} = 800$ nm, den Einzelschichtdicken h_{GaAs} und h_{InGaAs} einer Übergitterperiode und einem Indiumgehalt $x = 20$ % wird im folgenden als „$(h_{\mathrm{GaAs}}/h_{\mathrm{InGaAs}})$-Übergitter“ bezeichnet. Für h_{GaAs}, h_{InGaAs} werden die oben genannten Werte eingesetzt, so daß die Probenserie aus (2/18)-, (6/18)-, (10/18)-, (16/18)-, (20/18)-, (25/18)- und (30/18)-Übergittern besteht.

Die Fehlpassung zwischen den GaAs-Schichten und den InGaAs-Schichten beträgt 1.4 % und die Variation der GaAs-Einzelschichten bewirkt entsprechende Fehlpassungen der mittleren Gitterkonstante der Übergitter von $f_{\mathrm{SL}} = 1.26$, 1.05, 0.90, 0.74, 0.66, 0.59, 0.53 % relativ zum Substrat.

Die kritische Schichtdicke bei einer Fehlpassung von $f = 1.4$ % beträgt $h_{\mathrm{c}} = 5.7$ nm, so daß die InGaAs-Schichten deutlich überkritisch relativ zum Substrat sind. Die kritische Fehlpassung des kohärenten freien Übergitters (2.57) ist für eine (2/18)-Periode 4.7 %, die einer (30/18)-Periode ist 1.6 %, das heißt in allen Fällen sind die Übergitterperioden wegen $f = 1.4$ % unterkritisch. Die kritische Dicke $h_{\mathrm{SL,c}}$ für die Relaxation des gesamten Übergitters relativ zum Substrat liegt für Fehlpassungen $1.26\ \% \geq f_{\mathrm{SL}} \geq 0.53\ \%$ im Intervall $6.6\ \mathrm{nm} \leq h_{\mathrm{SL,c}} \leq 20.4\ \mathrm{nm}$. Die Übergitter sind also wegen $h_{\mathrm{SL}} = 800$ nm stark überkritisch.

Das Minimum der freien Energie der Übergitter ist mit den Modellannahmen, die in Kapitel 2 dargestellt sind, in diesem Modellsystem durch Konfigurationen gegeben, die aus kohärenten, versetzungsfreien Grenzflächen der Übergitterperioden und einem Netzwerk von Fehlpassungsversetzungen an der Übergitter–Substrat-Grenzfläche bestehen.

3.2 Transmissionselektronenmikroskopie

Mit der Messung der Grenzflächenversetzungsdichten in den real hergestellten Übergittern sind präzise Aussagen über den wirklichen Verspannungszustand der Übergitter und insbesondere über mögliche Abweichungen von den Gleichgewichtskonfigurationen zu gewinnen.

Aus drei Gründen ist die Methode der Transmissionselektronenmikroskopie besonders zur Untersuchung des vorliegenden Relaxationsproblems geeignet. Erstens haben die durch die möglichen Relaxationsprozesse erzeugten Versetzungen im vorliegenden System zu erwartende Dichten, die mit typischen Versetzungsabständen in der Größenordnung von einigen Nanometern bis zu einigen Mikrometern korrespondieren. Diese Längenskala wird durch die Transmissionselektronenmikroskopie einerseits an der unteren Grenze durch das erreichbare Auflösungsvermögen, und andererseits an der oberen Grenze durch ein ausreichend großes Gesichtsfeld gut abgedeckt. Zweitens haben die einzelnen Übergitter eine Gesamtdicke, die es bei typischen Beschleunigungsspannungen von 200 kV bis 400 kV erlaubt, elektronenmikroskopisch transparente Proben herzustellen, die ein jeweils vollständiges Übergitter und Teile des Substrats enthalten. Auf diese Weise sind Projektionsansichten von Grenzflächenversetzungen herstellbar, die eine quantitative Auswertung der Grenzflächenversetzungsdichten ermöglichen. Drittens ist wegen der Beobachtung diskreter Versetzungen eine direkte Information über konkret vorliegende Versetzungskonfigurationen zugänglich. Diese Information ist besonders wertvoll, weil sie Rückschlüsse auf die verschiedenen Prozesse der Versetzungskinetik zuläßt, die im vorangehenden Kapitel beschrieben wurden.

3.2.1 Probenpräparation

Zur Herstellung elektronentransparenter Proben wurden zwei Methoden angewandt, die in der Mikroskopie halbleitender Materialien als Standardpräparationsmethoden etabliert sind, nämlich das mechanische Dünnen durch eine Reihe von Spalt-, Säge-, Bohr- und Schleifprozeduren und die Ionendünnung mittels Argonionen bei gleichzeitiger Probenkühlung auf Stickstofftemperatur.

Es wurden sowohl *Rückdünnungsproben* präpariert, die eine Aufsicht der Übergitterstruktur und des Substrats entlang der [001]-Wachstumsrichtung ermöglichen, als auch *Querschnittproben*, die eine Seitenansicht eines Übergitters und des Substrats entlang der ⟨110⟩-Richtungen gestatten.

Die Präparation von Rückdünnungen beginnt in einem ersten Schritt durch Spalten der Wafer entlang der {110}-Spaltflächen der Zinkblendestruktur in kleine rechteckige Stücke und das anschließende Herausbohren von Scheiben, die einen Durchmesser von 3 mm haben, mittels eines Ultraschallbohrers.

In einem zweiten Schritt wird durch Schleifen ein großer Teil des Substratvolumens enfernt, bis hinunter zu einer Scheibendicke von etwa 150 μm. Dann wird das restliche Substratmaterial mittels Muldenschleifen soweit entfernt, bis die Dicke in der tellerartigen Vertiefung noch etwa 30 μm beträgt. Das Muldenschleifen geschieht durch das Auflegen eines rotierenden Schleifrads auf die Scheibe, die selbst

auf einem rotierenden Probenteller befestigt ist.

Im letzten Schritt des *Ionendünnens* wird die Scheibe auf einem rotierenden Halter befestigt, der sich in einem Vakuum befindet und der auf Stickstofftemperatur gekühlt wird. Die Mitte der rotierenden Scheibe wird von der Ober- und von der Unterseite mit Argonionen beschossen, die unter einem flachen Winkel von 18° einfallen und die auf eine Energie von 4 keV beschleunigt werden. Nachdem in der Mitte der Scheibe so viel Material abgetragen ist, daß ein kleines Loch entstanden ist, wird die Ionendünnung zur Reduzierung amorpher Deckschichten für eine kurze Zeit bei etwa halber Energie und halbem Einfallswinkel fortgesetzt.

Die so präparierten Lochränder sind sehr dünn, mit Dicken bis hinunter zu wenigen Nanometern. Entlang des Lochrandes gibt es Flächen von bis zu einigen Hundert Quadratmikrometern, die bei Beschleunigungsspannungen von 200 kV bis 400 kV elektronentransparent sind.

Die Präparation von Querschnittproben beginnt mit dem Herstellen eines Pakets, das aus mehreren gespaltenen Waferstücken besteht, die mit den (001)-Flächen gegeneinander geklebt werden, so daß die Seitenflächen eine {110}-Orientierung haben. Anschließend werden von den Seitenflächen der 3 mm dicken Pakete Scheiben gesägt, die selbst eine Dicke von etwa 0.5 mm haben.

Diese Scheiben werden analog zur Präparation von Rückdünnungen durch Schleifen und Ionendünnen weiterverarbeitet, so daß Flächen nahe des Lochrandes und nahe der aufgewachsenen Schichtstrukturen elektronentransparent sind.

3.2.2 Mikroskopische Abbildung von Versetzungen

In der konventionellen und in der hochauflösenden Transmissionselektronenmikroskopie wird eine Strukturinformation durch zwei physikalische Prozesse verschlüsselt. Der erste ist die dynamische Beugung hochenergetischer Elektronen am Kristallpotential, der zweite ist die nachfolgende elektronenoptische Abbildung der an der Probenunterseite austretenden Elektronenwelle in die Bildebene des Mikroskops. Eine vollständige Darstellung der Theorie der dynamischen Beugung und der elektronenoptischen Abbildung wird im zweiten Teil dieser Arbeit gegeben.

Die für die Untersuchung der Relaxationsprozesse in Übergittern wichtige Strukturinformation ist in erster Linie die Dichte von Versetzungen, sowie die Lage und die Größe ihrer Verzerrungsfelder.

Für die benötigte Charakterisierung von Versetzungsnetzwerken und für die Bestimmung von Versetzungsdichten ist in besonderer Weise die konventionelle Transmissionselektronenmikroskopie geeignet, mit welcher unter Benutzung der *Zweistrahlabbildungstechnik* eine Ortsauflösung von wenigen Nanometern erreichbar ist [214].

Bei Zweistrahlabbildungen wird die zu untersuchende Probe mechanisch so gekippt, daß bei der dynamischen Beugung hochenergetischer Elektronen am Kristallpotential außer dem transmittierten Strahl nur ein abgebeugter Strahl angeregt ist. Diese an der Probenunterseite austretenden Strahlen enthalten in ihren Amplituden und Phasen eine Information über das Kristallpotential, sowie über Verzerrungsfelder, die beide Strahlen während der Beugung durchlaufen haben. Ein Abriß der

Theorie der Zweistrahlbeugung [214], die aus der im zweiten Teil dieser Arbeit dargestellten Theorie der dynamischen Vielstrahlbeugung hervorgeht, ist im Anhang A zusammengefaßt.

Es existieren drei Arten von Zweistrahlabbildungen, die sich durch die Art der Anregung des abgebeugten Strahls, sowie durch die nachfolgende elektronenmikroskopische Abbildung unterscheiden.

Bei der *Hellfeldabbildung* wird der abgebeugte Strahl durch eine Blende von der nachfolgenden Abbildung ausgeschlossen. Ein Bildkontrast entsteht in erster Linie dadurch, daß die transmittierte Elektronenwelle durch ein ortsabhängiges Verzerrungsfeld entlang von Kristallsäulen parallel zum einfallenden Elektronenstrahl verschieden stark gebeugt wird. Im Gegensatz dazu wird bei einer *Dunkelfeldabbildung* der transmittierte Strahl durch eine Blende von der nachfolgenden Abbildung ausgeschlossen.

Eine dritte Abbildungsmethode ist die *Weak-Beam-Dunkelfeldtechnik*, bei der wie bei einer einfachen Dunkelfeldabbildung der transmittierte Strahl ausgeblendet wird. Im Gegensatz zu den beiden vorher genannten Methoden wird hier die Probe zusätzlich so gekippt, daß der Anregungsfehler, das heißt die nach Betrag und Richtung festgelegte Abweichung von der exakten Bragglage im reziproken Raum, möglichst groß ist. Der Anregungsfehler kann jedoch nicht beliebig vergrößert werden, da die zur Abbildung benötigte Intensität mit wachsendem Anregungsfehler abnimmt [214].

Für die Interpretation der konventionellen elektronenmikroskopischen Abbildungen von Defekten sind eine Reihe von Eigenschaften der genannten Zweistrahlabbildungen wichtig:

- Wenn das von den Strahlen durchlaufene Verzerrungsfeld an allen Orten Null ist, dann haben der transmittierte Strahl ϕ_0 und der abgebeugte Strahl ϕ_g eine Dickenperiodizität mit der *Extinktionslänge* ξ_g. Die wechselseitige Auslöschung von transmittiertem und abgebeugtem Strahl wird als *Pendellösung* der Zweistrahlbeugung bezeichnet und äußert sich zum Beispiel in der Abbildung keilförmiger Proben durch Kontraststreifen – sogenannte Dickenkonturen – entlang Orten gleicher Probendicke.

- Je größer der Anregungsfehler s_g ist, desto kleiner ist die Extinktionslänge

$$\xi_g = \frac{\xi_{g,0}}{\sqrt{1 + s_g^2\, \xi_{g,0}^2}}, \tag{3.1}$$

mit

$$\xi_{g,0} = \frac{h^2}{2me} \frac{\cos\theta_B}{\lambda V_g} \tag{3.2}$$

und der relativistisch korrigierten Elektronenmasse m, der Elektronenladung e, der Planckschen Konstante h, dem Bragg-Winkel θ_B und der relativistisch korrigierten Elektronenwellenlänge λ. V_g ist der Betrag der zugehörigen Fourierkomponente des relativistisch korrigierten Kristallpotentials, und es wichtig

zu bemerken, daß die Größe der Extinktionslänge umgekehrt proportional zu V_g ist. Der Bragg-Winkel erfüllt die Bragg-Bedingung der elastischen Streuung:

$$2d_{hkl} \sin \theta_B = n\lambda, \tag{3.3}$$

mit dem Ebenenabstand d_{hkl} der beugenden Kristallebenen und einer ganzen Zahl n.

- Die Information über das Verzerrungsfeld ist im Howie-Whelan-Differentialgleichungssystem (Anhang A), das die Zweistrahl-Elektronenbeugung beschreibt, mit einem Term $\vec{g} \cdot \vec{R}$ enthalten, worin $\vec{R} = \vec{u}$ das zugehörige Verschiebungsfeld und $\vec{g}$ der Beugungsvektor im reziproken Raum ist. Für Versetzungen sind hier konkret die Verschiebungsfelder (2.12), (2.13), (2.19) einzusetzen.

- Wenn das Verschiebungsfeld nur Komponenten rechtwinklig zum Beugungsvektor besitzt, das heißt $\vec{g} \cdot \vec{R} = 0$, dann verläuft die dynamische Beugung effektiv wie in einem perfekten Kristall, und das Verschiebungsfeld wird nicht abgebildet. Daraus folgt für die Verschiebungsfelder von Versetzungen die wichtige Auslöschregel [214]

$$\vec{g} \cdot \vec{b} = 0, \tag{3.4}$$

für Beugungsvektoren, die rechtwinklig zum Burgersvektor $\vec{b}$ einer Versetzung verlaufen. Sie gestattet die eindeutige Bestimmung von Burgersvektoren durch die systematische Abbildung einer Versetzung mittels verschiedener Beugungsvektoren, wenn zwei Auslöschbedingungen gefunden werden.

- Wenn eine Versetzung nicht ausgelöscht ist, und ihre Linie parallel zur Kristalloberfläche verläuft, dann wird sie als Linienkontrast abgebildet, dessen Halbwertsbreite vom Wert $n = \vec{g} \cdot \vec{b}$ abhängt. Er liegt zum Beispiel für $n = 1$ in der Größenordnung von 0.2 ξ_g bis 0.4 ξ_g, abhängig vom Versetzungstyp [214]. Die Beobachtung des Linienkontrastes unter verschiedenen Einfallsrichtungen des transmittierten Strahls ermöglicht die eindeutige Bestimmung der Linienrichtung einer Versetzung.

- Aus den Punkten 2 und 5 folgt die selbständige Bedeutung der Weak-Beam-Dunkelfeldtechnik: Durch das extreme Vergrößern des Anregungsfehlers wird die Extinktionslänge stark reduziert und damit kann die Kontrastbreite einer Versetzung bis auf Werte von 1.5 nm für $\vec{g} \cdot \vec{b} = 1$ herabgesetzt werden. Eine solch kleine Kontrastbreite ist nützlich, wenn zwei in einer Projektion eng benachbarte Versetzungen getrennt abgebildet werden müssen, oder wenn sich langreichweitige Kontraste von Versetzungen in einer Zweistrahlabbildung mit niedrigerem Anregungsfehler überlappen.

Kapitel 4

Charakterisierung der Übergitter

In diesem Kapitel wird eine Darstellung der experimentellen Beobachtungen gegeben, die den Verspannungszustand, den Relaxationsmechanismus und die damit verbundene charakteristische Defektstruktur der untersuchten Übergitter betreffen. Die experimentellen Resultate beleuchten zum Teil sehr verschiedene Aspekte des epitaktischen Wachstums und der Spannungsrelaxation. Deshalb wird zu jedem Resultat eine Diskussion angefügt, die die jeweilige Bedeutung erläutert. Eine zusammenfassende Diskussion der Beobachtungen wird im letzten Abschnitt des Kapitels gegeben.

4.1 Orthogonale Versetzungsnetzwerke

4.1.1 Experimentelle Resultate

In allen Übergittern wurden, unabhängig von der unterschiedlichen Einzelschichtdicke der GaAs-Schichten, Versetzungsnetzwerke gefunden. Die Untersuchung von Rückdünnungsproben der gesamten Probenserie ermöglichte die getrennte Beobachtung von Netzwerken, die sich an den Grenzflächen zwischen den GaAs- und InGaAs-Schichten innerhalb der Übergitter befinden, oder die sich an der Substratgrenzfläche zwischen den Übergittern und dem GaAs-Substrat befinden.

Wegen des präparationsbedingten, keilförmigen Dickenprofils der elektronenmikroskopischen Proben sind insbesondere zwei verschiedene Projektionen entlang der [001]-Richtung zu beobachten. Bei großer Probendicke, $t > h_{SL}$, enthält die Projektion alle Netzwerke innerhalb des Übergitters, das Netzwerk an der Substratgrenzfläche und Teile der Defektstruktur des Substrates. Bei genügend kleiner Probendicke, $t < h_{SL}$, ist das Substrat einschließlich der Substratgrenzfläche durch die Präparation entfernt worden, und in der Projektion sind allein die Netzwerke innerhalb des Übergitters enthalten.

Die Abbildung 4.1 zeigt eine Projektion der Netzwerke innerhalb eines (18/16)-Übergitters. Bei (D) befindet sich eine Haarnadelkonfiguration zweier Fehlpassungsversetzungen in benachbarten Übergittergrenzflächen (siehe Abschnitt 4.3), an den Stellen (a), (b) und (c) sind rechtwinklig abknickende Fehlpassungssegmente zu erkennen (siehe Abschnitt 4.5). Die Abbildung 4.2 zeigt eine Projektion eines voll-

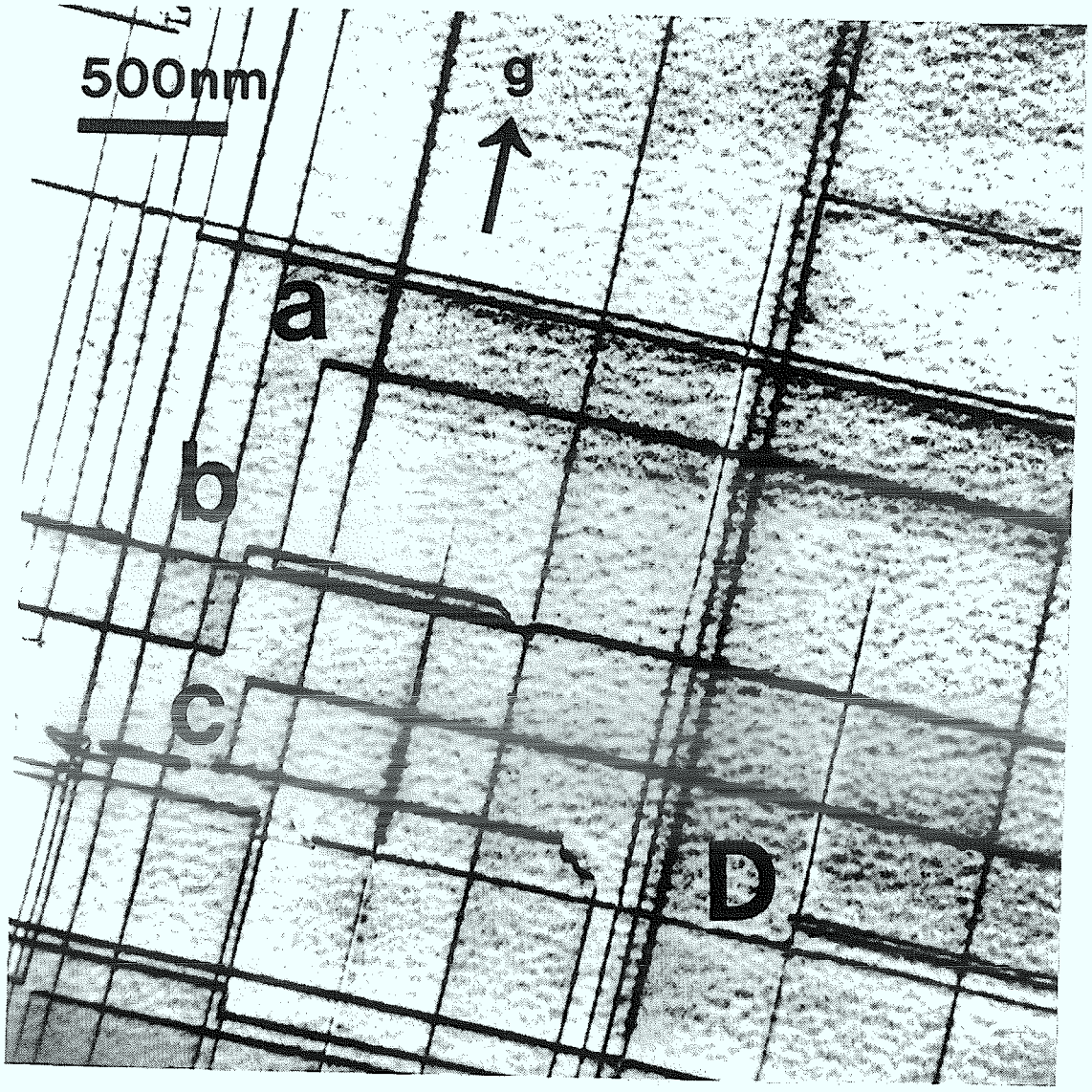


Abbildung 4.1: Projektion der Fehlpassungsnetzwerke an den Grenzflächen der Perioden eines (18/16)-Übergitters. In der Hellfeldabbildung, $\vec{g} = (220)$, sind entlang der Linienrichtungen $\vec{u} = [110]$ und $\vec{u} = [\bar{1}10]$ geradlinig verlaufende Fehlpassungsversetzungen zu erkennen. Bei (D) befindet sich eine Haarnadelkonfiguration, an den Stellen (a), (b) und (c) knicken Fehlpassungssegmente rechtwinklig ab.

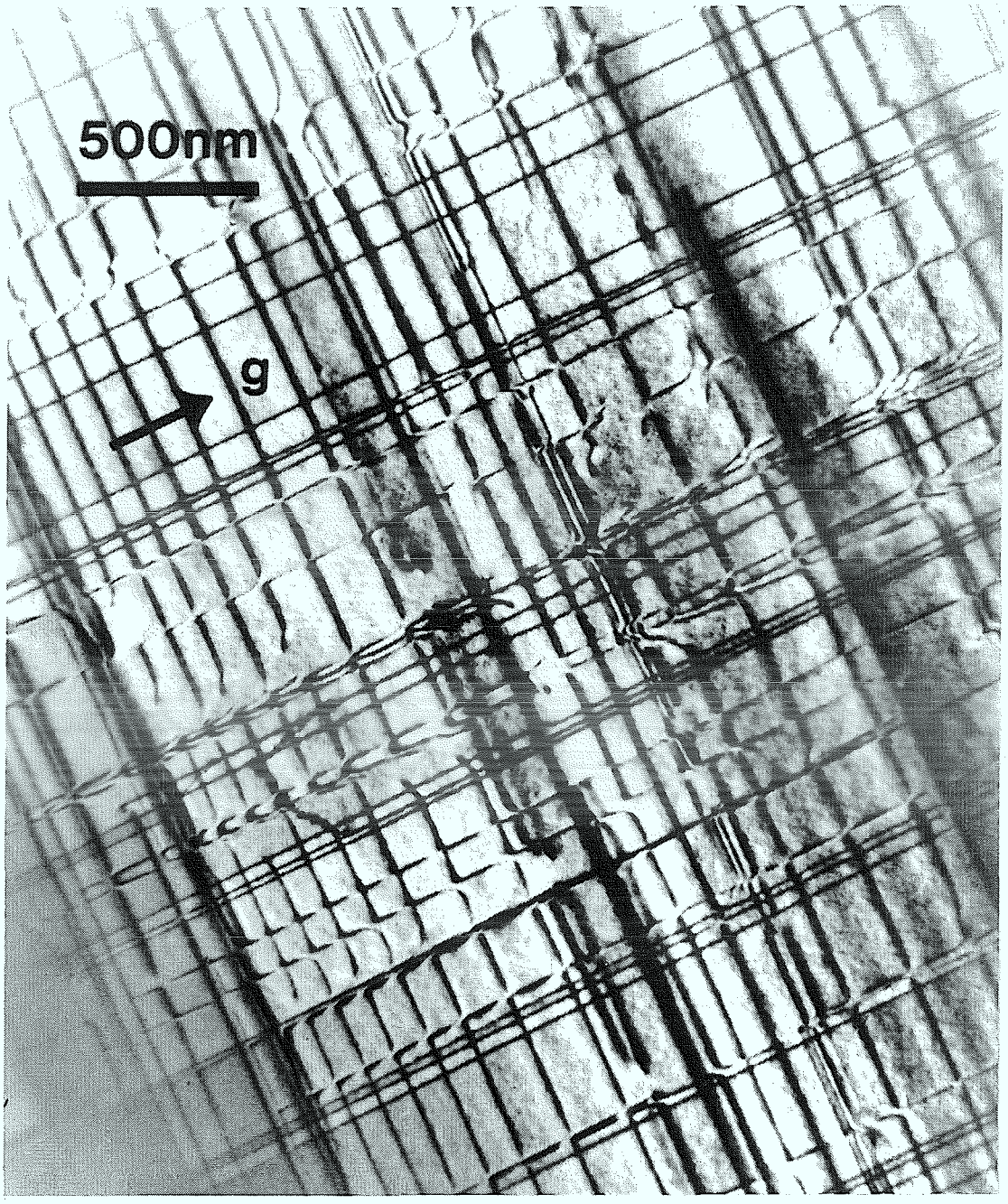


Abbildung 4.2: Projektion der Fehlpassungsnetzwerke an den Grenzflächen der Perioden eines (18/2)-Übergitters und an der Substratgrenzfläche. In der Hellfeldabbildung, $\vec{g} = (220)$, sind entlang der Linienrichtungen $\vec{u} = [110]$ und $\vec{u} = [\bar{1}10]$ geradlinig verlaufende Fehlpassungsversetzungen zu erkennen.

ständigen (18/2)-Übergitters, einschließlich eines dichten Versetzungsnetzwerkes an der Substratgrenzfläche. Durch den Vergleich von Abbildung 4.1 und Abbildung 4.2 erkennt man, daß die Versetzungsdichte an der Substratgrenzfläche deutlich höher ist, als die gesamte Versetzungsdichte im Übergitter.

In Abbildung 4.3 ist die Linie zu erkennen, an der die Dicke der keilförmigen Probe gerade gleich der Gesamtdicke eines (18/16)-Übergitters ist. An dieser Stelle treten die Versetzungslinien des Netzwerks an der Substratgrenzfläche an der Unterseite der Probe aus, einige Austrittspunkte sind mit Pfeilen markiert. Rechtwinklig abknickende Fehlpassungsversetzungen an den Übergittergrenzflächen sind mit (a), (b) und (c) gekennzeichnet (siehe Abschnitt 4.5). Bei (D) ist eine Haarnadelkonfiguration zweier Fehlpassungsversetzungen zu erkennen (siehe Abschnitt 4.3).

Alle Netzwerke bestehen aus zueinander rechtwinkligen Versetzungssegmenten, deren Linien in der Mehrzahl der Fälle über viele Mikrometer hinweg geradlinig verlaufen. Die Linienrichtungen dieser geraden Segmente sind für die eine Schar paralleler Versetzungen $\vec{u} = [110]$ und für die dazu orthogonale Schar $\vec{u} = [\bar{1}10]$. Querschnittaufnahmen zeigen, daß sich die geradlinig verlaufenden Segmente durchweg an den entsprechenden GaAs/InGaAs-Grenzflächen befinden.

4.1.2 Diskussion

Die Beobachtung orthogonaler Netzwerke an den heteroepitaktischen Grenzflächen, mit Linienrichtungen $\vec{u} = [110]$ und $\vec{u} = [\bar{1}10]$, ist in Übereinstimmung mit Beobachtungen früherer Arbeiten im System InGaAs/GaAs, als auch mit Beobachtungen anderer fehlgepaßter, epitaktischer Systeme mit Diamant- oder Zinkblendestruktur, wie zum Beispiel Ge/Si oder GaAs/Si [33, 35, 36, 118].

Die Ausbildung solcher Netzwerke ist typisch für Systeme niedriger Fehlpassung, wie im hier untersuchten Fall, aber auch für Systeme mit höheren Fehlpassungen. Es treten jedoch, abhängig von der Fehlpassung, unterschiedliche Gruppen von Burgersvektoren auf (Abschnitt 4.2).

Da der weitaus überwiegende Teil der Segmente, bezogen auf die totale Linienlänge, vom Fehlpassungstyp mit einer ⟨110⟩-Linienrichtung ist, und die Dichte von Fadenversetzungen dagegen vernachlässigt werden kann, ist auch die Modellannahme, die in Kapitel 2 gemacht wurde, zutreffend: Die Linienrichtungen der Fehlpassungssegmente sind in der Tat durch die Schnittgerade zwischen der (001)-Substratoberfläche und den {111}-Gleitebenen festgelegt.

4.2 Gleitfähige 60°-Versetzungen

4.2.1 Experimentelle Resultate

Es wurden drei experimentelle Hinweise dafür gefunden, daß in praktisch allen Fällen die Fehlpassungssegmente der orthogonalen Netzwerke vom gleitfähigen 60°-Typ sind.

Erstens wurde bei einer Anzahl von Versetzungssegmenten eine Auslöschung des Linienkontrastes bei Beugungsvektoren $\vec{g} = (400), (040)$ beobachtet, die mit

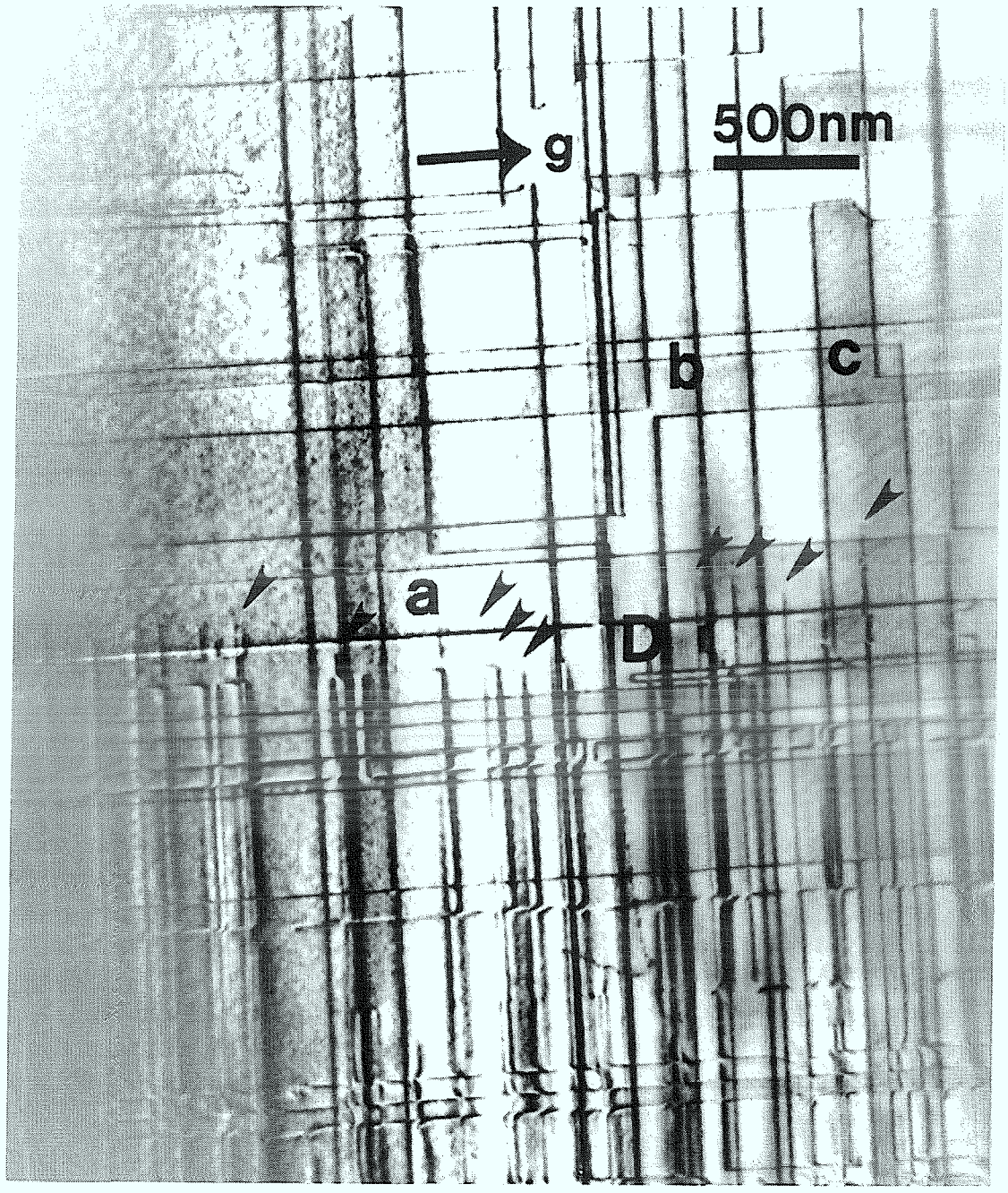


Abbildung 4.3: Schnittlinie zwischen der Substratgrenzfläche eines (18/16)-Übergitters und der Probenunterseite. In der Hellfeldabbildung, $\vec{g} = (220)$, sind Austrittspunkte der Versetzungslinien des Fehlpassungsnetzwerkes an der Substratgrenzfläche mit Pfeilen markiert. Bei (D) befindet sich eine Haarnadelkonfiguration zweier Fehlpassungsversetzungen, an den Stellen (a), (b) und (c) knicken Fehlpassungssegmente rechtwinklig ab.

Benutzung des Auslöschkriteriums (3.4) zu Burgersvektoren kompatibel ist, deren Richtungen jeweils unter einem Winkel von 60° zur Versetzungslinie verlaufen, und die auf {111}-Gleitebenen liegen.

Abbildung 4.4 zeigt ein Paar von Hellfeldabbildungen einer Schar paralleler Fehlpassungsversetzungen. Bei der Abbildung mit dem Beugungsvektor $\vec{g} = (220)$ (oben) sind alle Versetzungen, die mit (1), (2), (3), (4), (5) gekennzeichnet sind, in Kontrast, während die Versetzungen (2) und (3) bei der Abbildung mit $\vec{g} = (400)$ (unten) so weit ausgelöscht sind, daß nur ein schwacher Restkontrast sichtbar bleibt. Mit dieser Auslöschung sind Burgersvektoren von Typ $a/2\langle 011\rangle$ kompatibel, die auf zur Grenzfläche geneigten {111}-Gleitebenen liegen.

Zweitens konnte der ebenfalls in fehlgepaßten Schichtsystemen der Diamant- und Zinkblendestruktur beobachtete Typ der Lomer-Versetzung für praktisch alle Segmente der Fehlpassungsnetzwerke ausgeschlossen werden, weil in keinem der untersuchten Fälle die damit verbundene Auslöschung von Linienkontrasten bei der Abbildung ein und derselben Stelle eines Netzwerks mit (220)- und $(\bar{2}20)$-Beugungsvektoren beobachtet wurde.

Drittens konnte die Atomsäulenstruktur einiger Versetzungskerne entlang der [110]-Linienrichtung mit der hochauflösenden Transmissionselektronenmikroskopie direkt und atomar aufgelöst abgebildet werden. Der 60°-Typ der Fehlpassungsversetzungen konnte durch die Ausführung von Burgersumläufen (Gleichung (2.11) und dort folgender Absatz) in hochaufgelösten Gitterabbildungen verifiziert werden.

4.2.2 Diskussion

Das experimentelle Ergebnis liefert zwei direkte und einen indirekten Hinweis darauf, daß praktisch alle Fehlpassungsversetzungen vom gleitfähigen 60°-Typ sind. Dabei hat besonders der obige indirekte Nachweis große praktische Bedeutung, weil das systematische Testen von Auslöschbedingungen unter anderem eine Reihe von Großwinkelkippungen der Proben erfordert, die wegen des mechanisch begrenzten Kippwinkels in einem Mikroskop und wegen einer oft starken Probenverbiegung nicht an allen gewünschten Probenstellen ausgeführt werden können.

Die Beobachtung des gleitfähigen 60°-Versetzungstyps in diesem System niedriger Fehlpassung ist in Übereinstimmung mit den Resultaten früherer Arbeiten im InGaAs/GaAs-System, in denen ebenfalls keine Lomer-Versetzungen bei niedrigen Fehlpassungen gefunden wurden [33, 118].

Lomer-Versetzungen entstehen in höher fehlgepaßten Systemen durch die Versetzungsreaktion zweier gleitfähiger 60°-Versetzungen [83]. Die möglichen resultierenden Burgersvektoren $\pm a/2\,[\bar{1}\bar{1}0]$ und $\pm a/2\,[\bar{1}10]$ sind zwar vom perfekten $a/2\,\langle 110\rangle$-Typ, sie spannen jedoch mit den entsprechenden Linienrichtungen keine {111}-Gleitebenen auf, und sie sind somit nur noch bei höheren Temperaturen oder in extrem hohen Spannungsfeldern beweglich. Die Bildung von Lomer-Versetzungen ist energetisch günstig, weil ihre Burgersvektoren vollständig in der Grenzfläche liegen, und sie dadurch eine zweimal höhere Relaxationseffizienz als 60°-Versetzungen haben [51]. Ihre Bildungsrate steigt allerdings erst bei höheren Fehlpassungen an, weil für die Versetzungsreaktion ein schon relativ dichtes Netz von 60°-Versetzungen

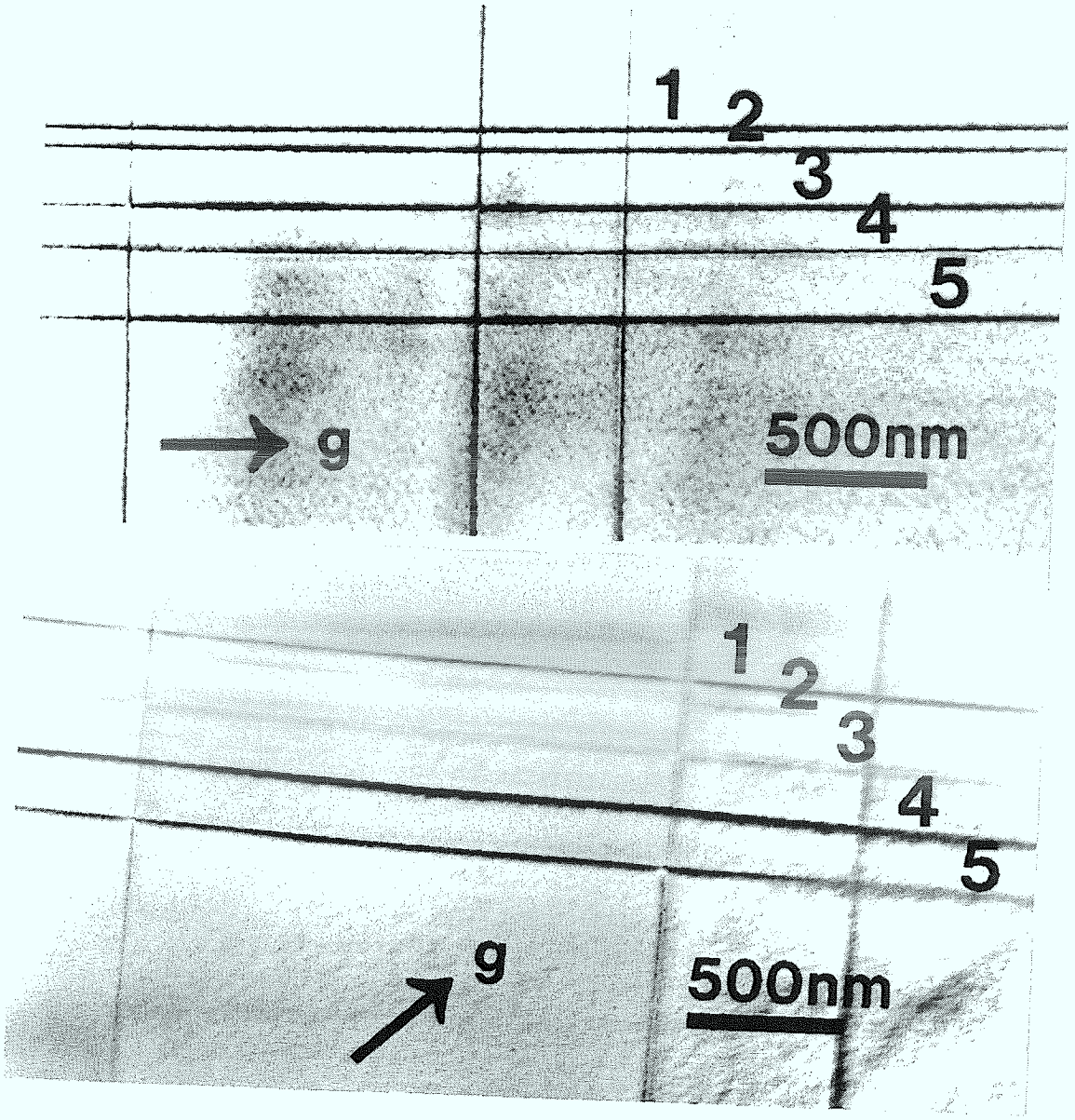


Abbildung 4.4: Auslöschung des Versetzungskontrastes von Fehlpassungssegmenten. In der Hellfeldabbildung mit dem Beugungsvektor $\vec{g} = (220)$ sind die parallelen Fehlpassungsversetzungen (1), (2), (3), (4), (5) in Kontrast (oben). In der Hellfeldabbildung mit dem Beugungsvektor $\vec{g} = (400)$ sind die Versetzungen (2) und (3) bis auf einen schwachen Restkontrast ausgelöscht (unten).

vorhanden sein muß.

Der indirekte Nachweis des 60°-Typs über den Ausschluß von Lomer-Versetzungen basiert auf der Tatsache, daß in der Diamant- und der Zinkblendestruktur außer Lomer- und 60°-Versetzungen bei ⟨110⟩-Linienrichtungen nur noch reine Schraubenversetzungen mit perfektem $a/2\langle 110\rangle$-Burgersvektor existieren. Die Scherkomponenten des Spannungstensors der epitaktisch aufgewachsenen, fehlgepaßten Schichten sind jedoch Null, so daß der zusätzliche Einbau von Schraubenversetzungen energetisch nicht begünstigt ist. Experimentell konnte die Existenz von reinen Schraubenversetzungen ausgeschlossen werden, da die mit Schraubenversetzungen verbundene Auslöschung von ⟨110⟩-Versetzungslinien mit (220)- und $(2\bar{2}0)$-Abbildungsvektoren nicht beobachtet wurde.

4.3 Versetzungsdipole

4.3.1 Experimentelle Resultate

In einer Reihe von Fällen wurde beobachtet, daß parallele Fehlpassungsversetzungen an benachbarten Übergittergrenzflächen in der (001)-Projektion einen kleinen lateralen Abstand $S_{(001)}$ haben, der ungefähr der Größe der Schichtdicke h der dazwischenliegenden Einzelschicht entspricht. Eine quantitative Auswertung des Abstands unter Berücksichtigung der Probenkippung aus der [001]-Zonenachse, die für das Einstellen einer Zweistrahlbedingung erforderlich ist, ergab für das Verhältnis von Schichtdicke zu Abstand einen Wert von $h/S_{(001)} = 1.4$. Dieser Wert stimmt gut mit dem Tangens des Winkels 54.7° zwischen einer $\{111\}$-Gleitebene und der (001)-Grenzfläche überein. Die beiden Segmente befinden sich wegen dieser Geometrie der Konfiguration auf derselben Gleitebene.

Zwei Beobachtungen belegen, daß die beiden Segmente einen entgegengesetzten Burgersvektor besitzen. Zweistrahlabbildungen eines solchen Versetzungspaares an einer dünnen Probenstelle mit einem Beugungsvektor $\vec{g}$ senkrecht zu den parallelen Versetzungslinien ergaben einen deutlichen „Hell-Dunkel-Hell"- beziehungsweise einen „Dunkel-Hell-Dunkel"-Kontrast des benachbarten ungestörten Kristalls.

Abbildung 4.5 zeigt die Hellfeldaufnahme zweier Versetzungsdipolkonfigurationen (A), (B), bei welchen die projizierte Fläche zwischen zwei parallelen Fehlpassungssegmenten heller erscheint, als die Hintergrundintensität. Abbildung 4.6 zeigt ein Stereopaar von Hellfeldabbildungen eines Dipols zweier Fehlpassungsversetzungen. Durch die Variation der Probenkippung ändert sich in einem Stereopaar der projizierte Abstand der Versetzungslinien, die sich an benachbarten Übergittergrenzflächen befinden.

In Abbildung 4.7 ist ein Stereopaar von Hellfeldaufnahmen einer Konfiguration von vier parallelen Fehlpassungsversetzungen dargestellt. Diese besteht aus zwei parallelen Versetzungsdipolen (A), (B), die sich jedoch in ihrem Kontrastverhalten unterscheiden. Die von Dipol (A) eingeschlossene projizierte Fläche erscheint heller als der Hintergrund, die von Dipol (B) eingeschlossene Fläche dunkler. Aus der Änderung der projizierten Versetzungsabstände läßt sich ableiten, daß in beiden Konfigurationen die linke Fehlpassungsversetzung oberhalb der rechten Fehl-

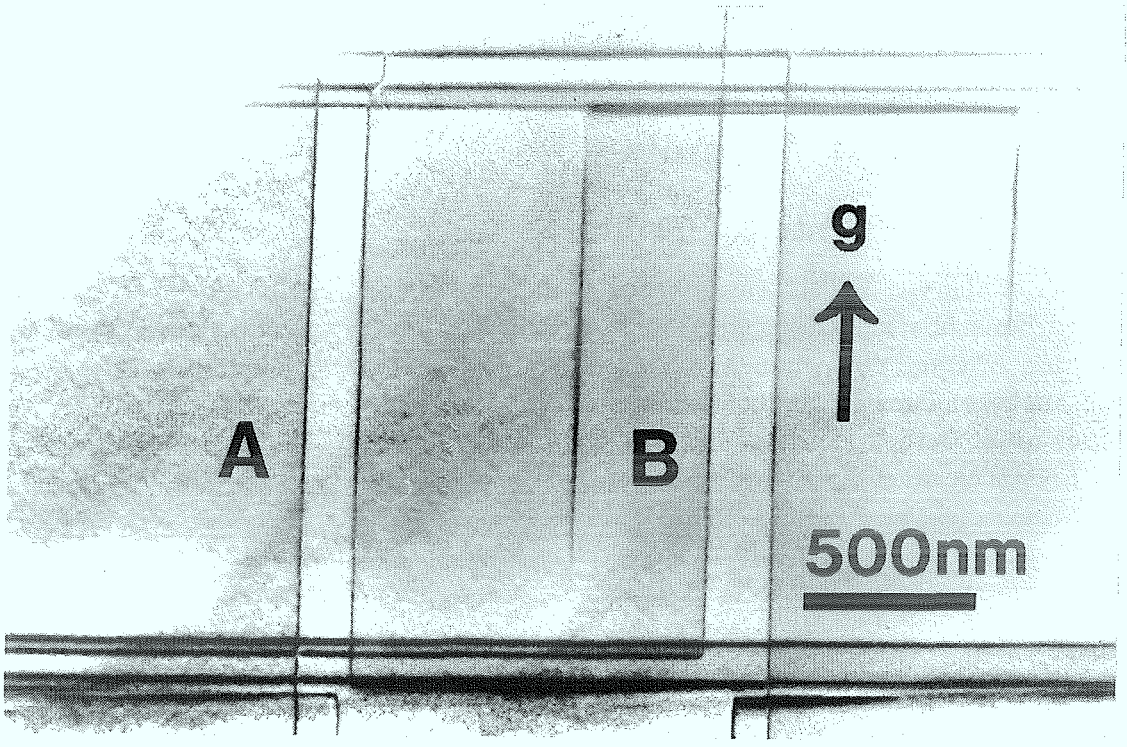


Abbildung 4.5: „Dunkel-Hell-Dunkel"-Kontrast an Versetzungsdipolen. In der Hellfeldabbildung, $\vec{g} = (220)$, zweier Dipolkonfigurationen von Fehlpassungsversetzungen (A) und (B) erscheint die projizierte Fläche zwischen den parallelen Fehlpassungssegmenten aufgrund einer lokalen Probenkippung heller.

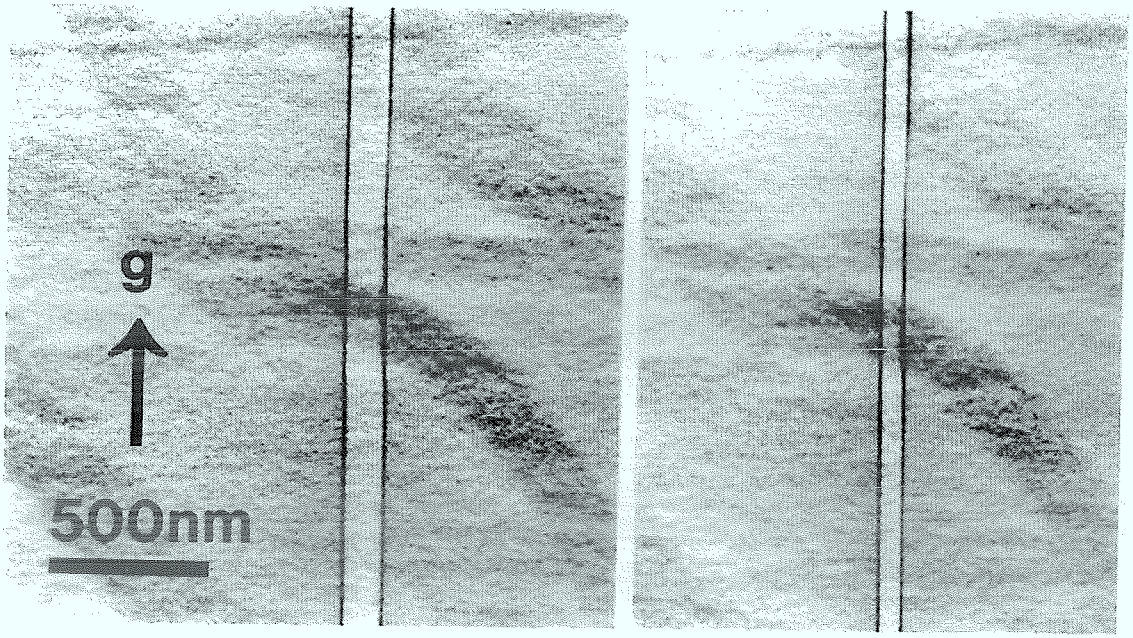


Abbildung 4.6: Stereopaar von Hellfeldabbildungen eines Dipols zweier Fehlpassungsversetzungen. Durch eine Variation der Probenkippung in beiden Hellfeldabbildungen, $\vec{g} = (220)$, ändert sich der projizierte Abstand der beiden Fehlpassungssegmente, die sich an unterschiedlichen Grenzflächen des (18/25)-Übergitters befinden. Die projizierte Fläche zwischen den Versetzungslinien erscheint wegen einer lokalen Probenkippung heller.

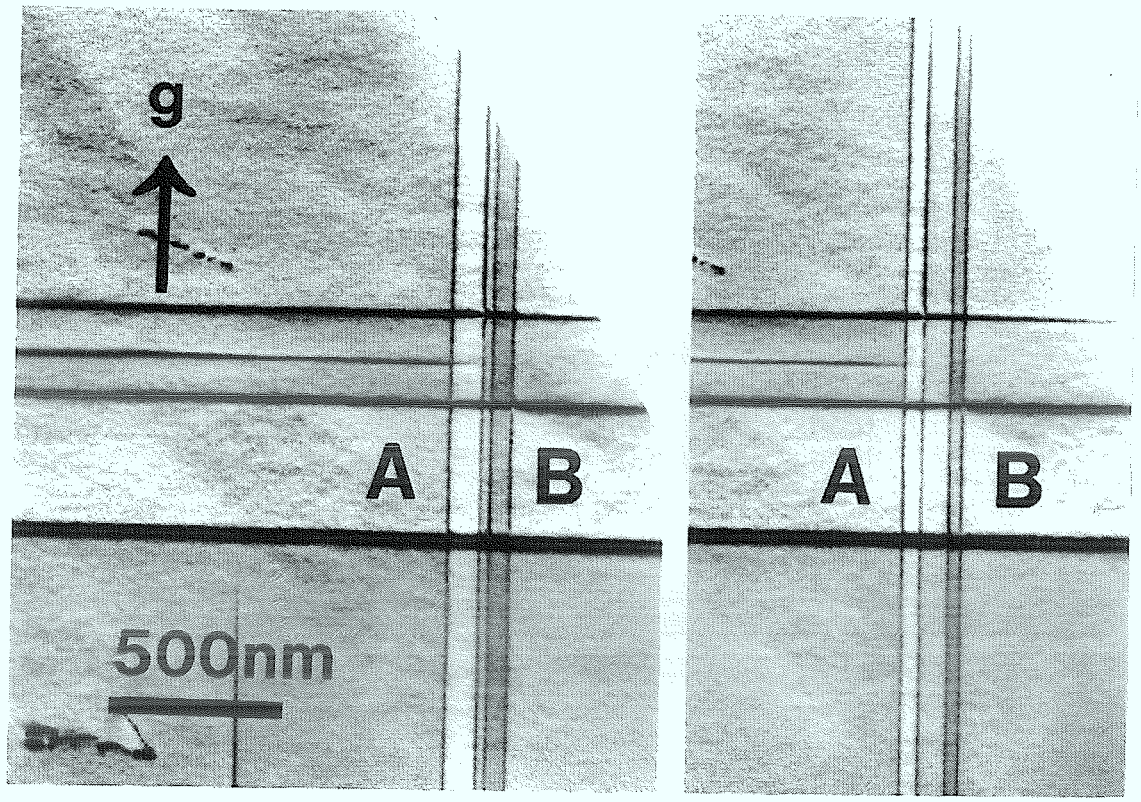


Abbildung 4.7: Stereopaar von Hellfeldabbildungen zweier benachbarter Dipole (A) und (B) von Fehlpassungsversetzungen. Durch eine Variation der Probenkippung in beiden Hellfeldabbildungen, $\vec{g} = (220)$, ändert sich der projizierte Abstand der beiden Fehlpassungssegmente in (A) und (B), die sich an unterschiedlichen Grenzflächen des (18/25)-Übergitters befinden, sowie der projizierte Abstand zwischen beiden Dipolen. Die projizierte Fläche zwischen den Versetzungslinien erscheint wegen einer lokalen Probenkippung in (A) heller und in (B) dunkler.

passungsversetzung liegt. Die projizierten Abstände der Segmente in (A) und (B) verhalten sich wie die Einzelschichtdicken des (18/25)-Übergitters, deshalb schließt Dipol (A) eine GaAs-Einzelschicht und Dipol (B) eine InGaAs-Einzelschicht ein.

Im Rahmen der Theorie der Zweistrahlbeugung ist dieser Kontrast an dünnen Probenstellen mit konstanter Dicke als entsprechende Variation des Anregungsfehlers zu deuten, der seine Ursache in einer kleinen Probenkippung hat [214]. Der mathematische Zusammenhang zwischen Probenkippung und Bildintensität ist durch Gleichung (A.9) des Anhangs gegeben.

Eine Stufenversetzung, die parallel zur Oberfläche in der Mitte einer dünnen Probe verläuft, bewirkt eine lokale Verbiegung, die die beiden ungestörten Kristallhälften um einen Winkel $\theta = 3b/2t$ gegeneinander verdreht [214]. Dabei liegt der Burgersvektor mit dem Betrag b senkrecht zur Probennormalen und t ist die Probendicke.

Der Kontrast „rechts" und „links" des Versetzungspaares war jeweils gleich, so daß die lokale Verbiegung der ersten Versetzung durch die lokale Verbiegung der zweiten wieder aufgehoben wird. Deshalb haben die beiden Stufenkomponenten senkrecht zur Probennormalen ein entgegengesetztes Vorzeichen und die Versetzungskonfiguration ist ein *Dipol* zweier Fehlpassungsversetzungen.

Mit Hilfe dieses Kontrastverhaltens kann der unterschiedliche Charakter der beiden Dipole, die in Abbildung 4.7 dargestellt sind, direkt erschlossen werden: Die unterschiedliche Probenkippung auf den von beiden Dipolen eingeschlossenen Flächen resultiert aus der vertauschten Reihenfolge der Vorzeichen der Stufenkomponenten beider Burgersvektoren in (A) und (B). Deshalb schließen die Dipole jeweils unterschiedliche Einzelschichten ein. Diese Kontrastinterpretation ist in Übereinstimmung mit dem oben dargestellten Resultat der Stereo-Analyse der Versetzungskonfiguration.

Die zweite Beobachtung ist ein direkter Beweis des Dipolcharakters solcher benachbarter Fehlpassungsversetzungen, denn in einigen Fällen wurden Fadensegmente gefunden, die zwei Fehlpassungssegmente verbinden und so einen Dipol terminieren. Ein charakteristisches Beispiel einer solchen „Haarnadel"-Konfiguration ist in Abbildung 4.8 dargestellt. Die Funktion der Fadensegmente wird im nächsten Abschnitt erläutert.

4.3.2 Diskussion

Die Existenz von Versetzungsdipolen in benachbarten Grenzflächen der Übergitterperioden ist im Rahmen der in Kapitel 2 dargestellten Relaxationsmodelle ein charakteristisches Ergebnis der Relaxation eines freien Übergitters [124] oberhalb eines Satzes kritischer Einzelschichtdicken $h_{A,c}$, $h_{B,c}$ (Gleichung (2.57)).

Diese Relaxation der Einzelschichten sollte jedoch in den hier untersuchten Übergittern *nicht* stattfinden, weil die Einzelschichtdicken jeweils unterkritisch sind. Daß sie trotzdem stattfindet, und daß auch Versetzungsnetzwerke an den Grenzflächen der Übergitterperioden existieren, die *keine* Dipolkonfiguration mit benachbarten Netzwerken bilden, wird bei der späteren Diskussion der Versetzungskinetik wichtig sein (Abschnitt 5.5.4).

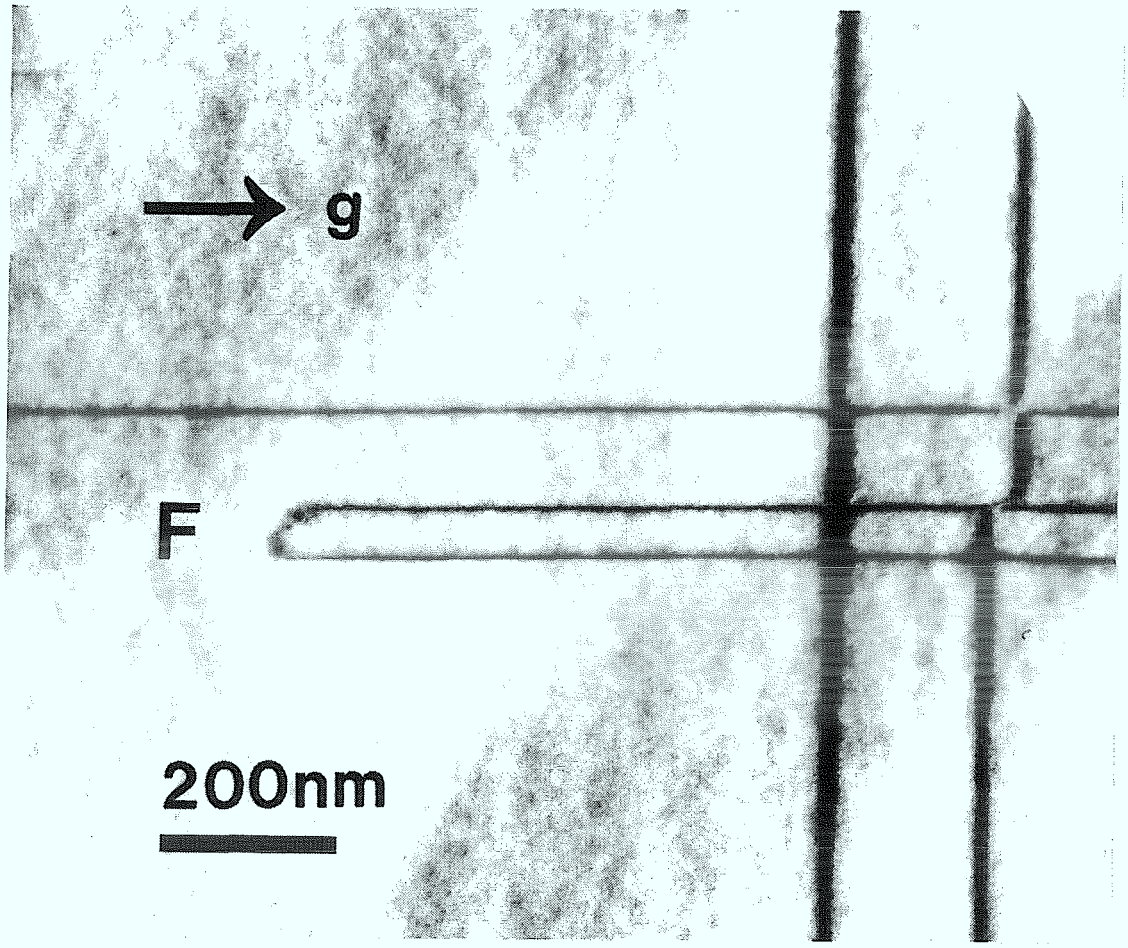


Abbildung 4.8: Haarnadelkonfiguration zweier paralleler Fehlpassungssegmente. In der Hellfeldabbildung, $\vec{g} = (220)$, ist ein Fadensegment (F) zu erkennen, das zwei parallele Fehlpassungssegmente an benachbarten Übergittergrenzflächen eines (18/25)-Übergitters verbindet und so die Dipolkonfiguration terminiert.

Eine weitere Eigenschaft der Versetzungsdipole macht eine Korrektur des Relaxationsmodells eines freien Übergitters (Abschnitt 2.6.2) nötig, die aus der Geometrie der parallelen Fehlpassungssegmente folgt. In diesem Modell, wie auch in der Originalarbeit von Matthews wurde angenommen, daß die Versetzungslinien in einer (001)-Projektion direkt übereinander liegen, weil dies die Konfiguration minimaler freier Energie ist. Da die beiden Segmente auf einer Gleitebene liegen, gibt es noch die zusätzliche geometrische Randbedingung [124]

$$h = S_{(001)} \tan \alpha \tag{4.1}$$

für die Schichtdicke h, den Abstand $S_{(001)}$ in der (001)-Projektion, und $\tan \alpha = \sqrt{2}$ für den Winkel zwischen der Gleitebene und der Grenzfläche. Wegen dieser Bedingung ist der im Relaxationsmodell benutzte Abschneideradius $S = \bar{h}/2$ zu ändern, und zwar auf den Wert

$$S = \sqrt{3/8}\,\bar{h}. \tag{4.2}$$

4.4 Relaxation durch Versetzungsdipole

4.4.1 Experimentelle Resultate

Das Fadensegment eines Versetzungsdipols kann durch Abgleiten in der Einzelschicht einen Teil der Fehlpassungsspannung abbauen, indem es das obere und das untere Fehlpassungssegment verlängert, und so als Multiplikator des Relaxationsprozesses wirkt [124] (Abschnitt 2.7.2).

Dieser Multiplikationsmechanismus konnte direkt an einzelnen Dipolen beobachtet werden. Zu diesem Zweck wurde jeweils eine Abbildung einer „Haarnadel"-Konfiguration aus zwei Fehlpassungssegmenten und einem Fadensegment aufgenommen, und anschließend der Elektronenstrahl stark auf das Fadensegment fokussiert. Durch das damit verbundene Aufheizen des Kristalls dehnte sich die elektronenmikroskopische Probe lokal aus, und an der Stelle des Elektronenstrahls wirkte eine zusätzliche mechanische Spannung. Unter dieser Spannung und der Spannung der fehlgepaßten Schicht glitt das Fadensegment ein Stück weit in den kühleren Teil der Einzelschicht ab. Nach dem Abgleiten wurde eine zweite Abbildung der Haarnadelkonfiguration aufgenommen. Dieser Vorgang konnte an ein und derselben Fadenversetzung mehrfach wiederholt werden. In Abbildung 4.9 sind fünf solcher Bilderserien zusammengefaßt.

4.4.2 Diskussion

Der Dipolmechanismus wurde ursprünglich von Matthews [124] vorgeschlagen, um die Relaxation von Einzelschichten eines Übergitters zu erklären. Es wurde angenommen, daß es oberhalb einer kritischen Dicke einer Einzelschicht energetisch günstig ist, daß ein Fadensegment unter der Spannung der Gitterfehlpassung in die Einzelschicht abgleitet und dabei zwei Fehlpassungssegmente erzeugt.

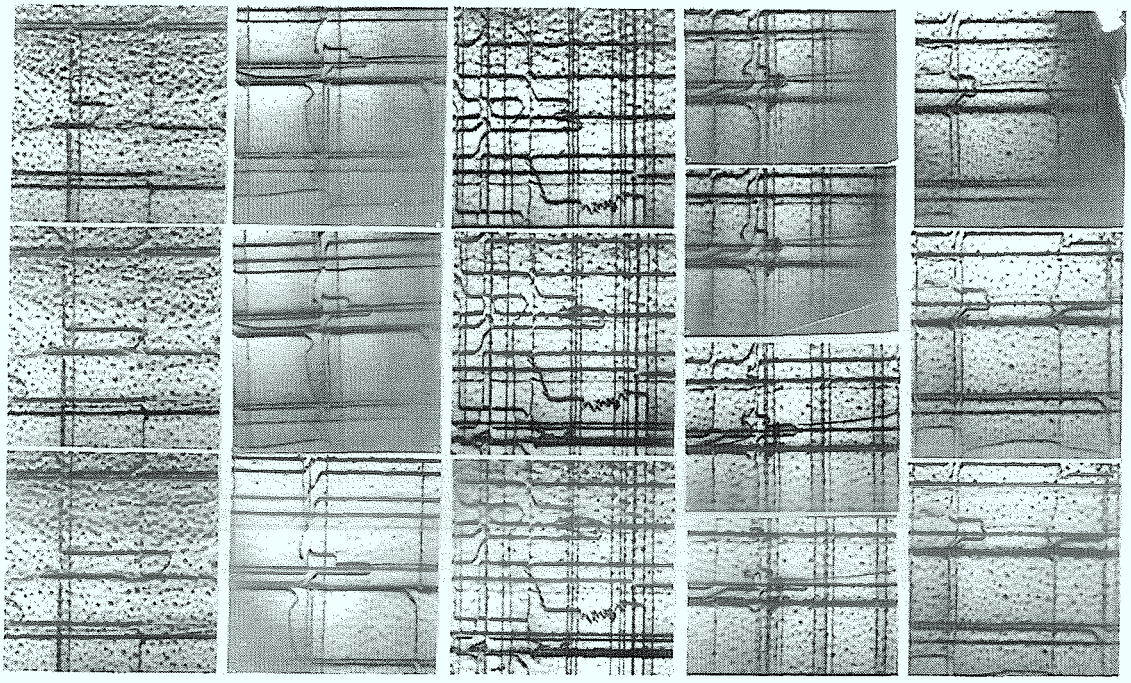


Abbildung 4.9: Abgleiten der Fadensegmente von Haarnadelkonfigurationen unter Elektronenbestrahlung. Die fünf Spalten enthalten Bilderserien von Hellfeldabbildungen, $\vec{g} = (220)$, die von oben nach unten den Fortschritt des Abgleitens mit der Zeit dokumentieren. Jede Serie umfaßt ein Zeitintervall von etwa einer Minute. Das entsprechende Fadensegment befindet sich jeweils im Zentrum eines Teilbildes.

Eine Analyse der kritischen Schichtdicke für diesen Prozeß ergibt jedoch, daß diese generell etwa zweimal größer als die kritische Schichtdicke einer fehlgepaßten Einzelschicht ist, wenn die Schichten oberhalb und unterhalb nicht verspannt sind [51] („single quantum well"). Die zu Gleichung (2.31) analoge Beziehung lautet nämlich für den Fall des Single-Quantum-Well:

$$\epsilon = \frac{b}{4\pi h}\left(\ln\frac{h}{b}+1\right). \tag{4.3}$$

Die kritische Schichtdicke ist etwa viermal größer, wenn die Relaxation eines Übergitters betrachtet wird (Abschnitt 2.6.2).

Der Mechanismus sollte also in dem hier untersuchten, verspannten Schichtsystem nicht zu beobachten sein, weil wegen der kleineren kritischen Dicke die fehlgepaßte Einzelschicht in der Wachstumssequenz schon relaxiert, bevor die Deckschicht oder weitere Übergitterperioden gewachsen sind.

Daß trotzdem Dipole beobachtet werden, ist mit hoher Wahrscheinlichkeit durch eine kinetische Behinderung (Abschnitt 2.7) der Versetzungsbewegung während der Wachstumssequenz verursacht. Der Einfluß der Wachstumssequenz wird ausführlich in einem späteren Abschnitt (5.6) diskutiert. Während des Schichtwachstums sind zum Beispiel die InGaAs-Schichten wegen $h_{\mathrm{InGaAs}} = 18$ nm überkritisch bezüglich der Einzelschichtrelaxation, so daß Fadensegmente in der InGaAs-Schicht abgleiten (Abschnitt 2.7.2). Beim Wachstum der nächsten GaAs-Schicht wird die darunterliegende InGaAs-Schicht wieder unterkritisch, weil die kohärente GaAs-Schicht die Fehlpassungsspannung der InGaAs-Schicht reduziert, so daß die entsprechenden Fadensegmente wieder zurückgleiten müssen. Kann ein Fadensegment wegen eines Hindernisses, zum Beispiel wegen einer quer zu ihrer Gleitrichtung verlaufenden Fehlpassungsversetuzng mit parallelem Burgersvektor diesen Gleitprozessen nicht folgen, so hat dies eine Erhöhung der kritischen Schichtdicke zur Folge, denn die zusätzliche lokale Spannung des Hindernisses addiert sich zur Spannung der Gitterfehlpassung. Das Fadensegment gleitet dann beim Wachstum der Einzelschicht oberhalb der höheren kritischen Dicke ab, und zwar so lange, bis es durch ein neues Hindernis gestoppt wird.

Die Möglichkeit, das Gleiten eines Fadensegments durch Anlegen einer zusätzlichen lokalen Spannung zu aktivieren, stützt die These einer solchen kinetischen Behinderung.

4.5 Wechsel von Gleitebenen

4.5.1 Experimentelle Resultate

In den Fehlpassungsnetzwerken der Übergitterperioden wurden an manchen Stellen Versetzungslinien beobachtet, die aus zwei Segmenten mit Linienrichtungen entlang der [110]- und der [$\bar{1}$10]-Richtung bestehen. Dieses rechtwinklige „Abknicken" einer Versetzungslinie ist bemerkenswert, weil beim gewöhnlichen Relaxationsmechanismus die Spannung durch Gleiten eines Fadensegments entlang einer festen Richtung

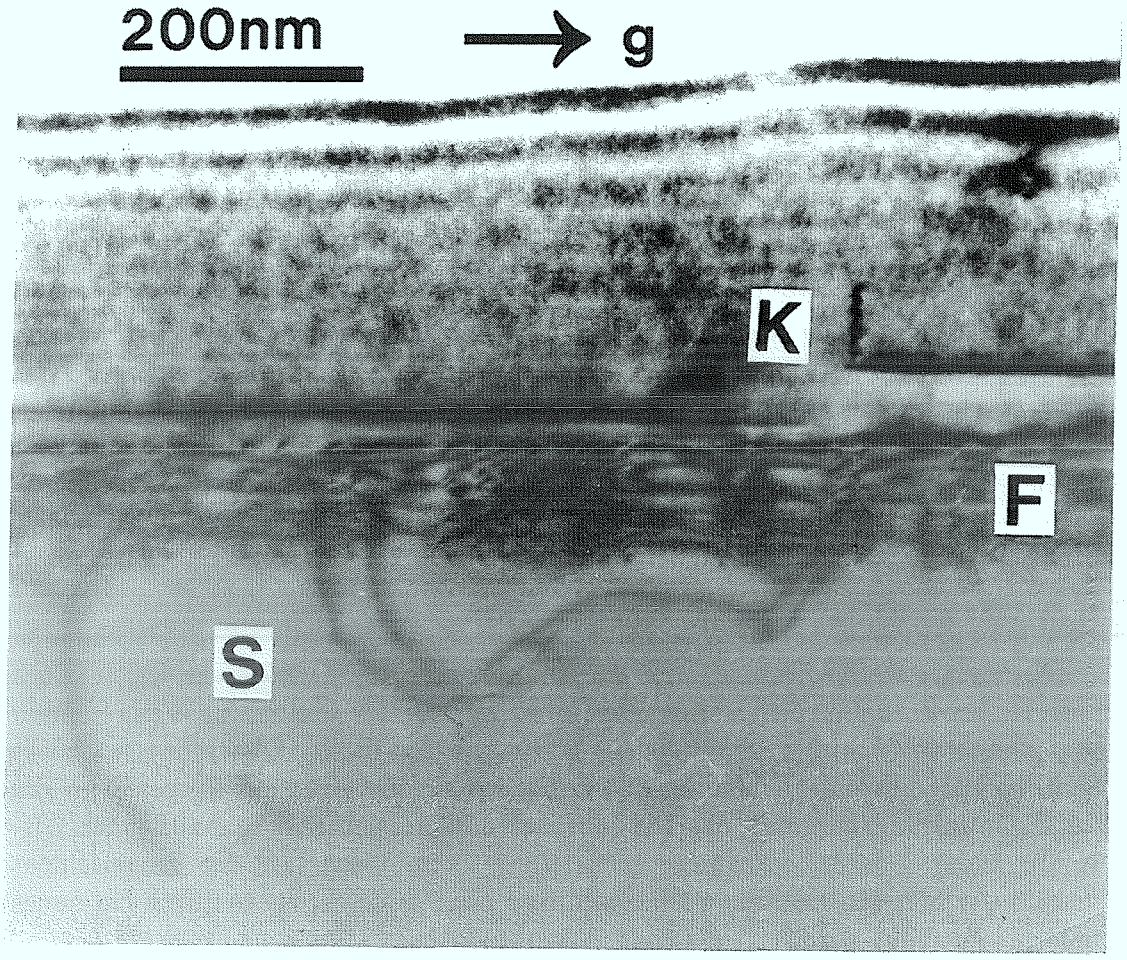


Abbildung 4.10: Rechtwinkliges Abknicken einer Fehlpassungsversetzung an einer Übergittergrenzfläche bei (K). In der Hellfeldabbildung, $\vec{g} = (220)$, sind außerdem das dichte Fehlpassungsnetzwerk an der Substratgrenzfläche eines (18/6)-Übergitters (F) und Versetzungsschleifen, die von der Substratgrenzfläche in das Substrat reichen (S), zu erkennen.

abgebaut wird. Abbildung 4.10 zeigt eine Querschnittaufnahme eines (18/6)-Übergitters mit einem rechtwinklig abknickenden Fehlpassungssegment an einer Übergittergrenzfläche bei (K), dem dichten Fehlpassungsnetzwerk an der Substratgrenzfläche (F) und Versetzungsschleifen, die von der Substratgrenzfläche in das Substrat reichen (S) (siehe Abschnitt 4.7). Ein zweites Beispiel ist in Abbildung 4.11 dargestellt, wo eine Fehlpassungsversetzung an einer Übergittergrenzfläche an zwei

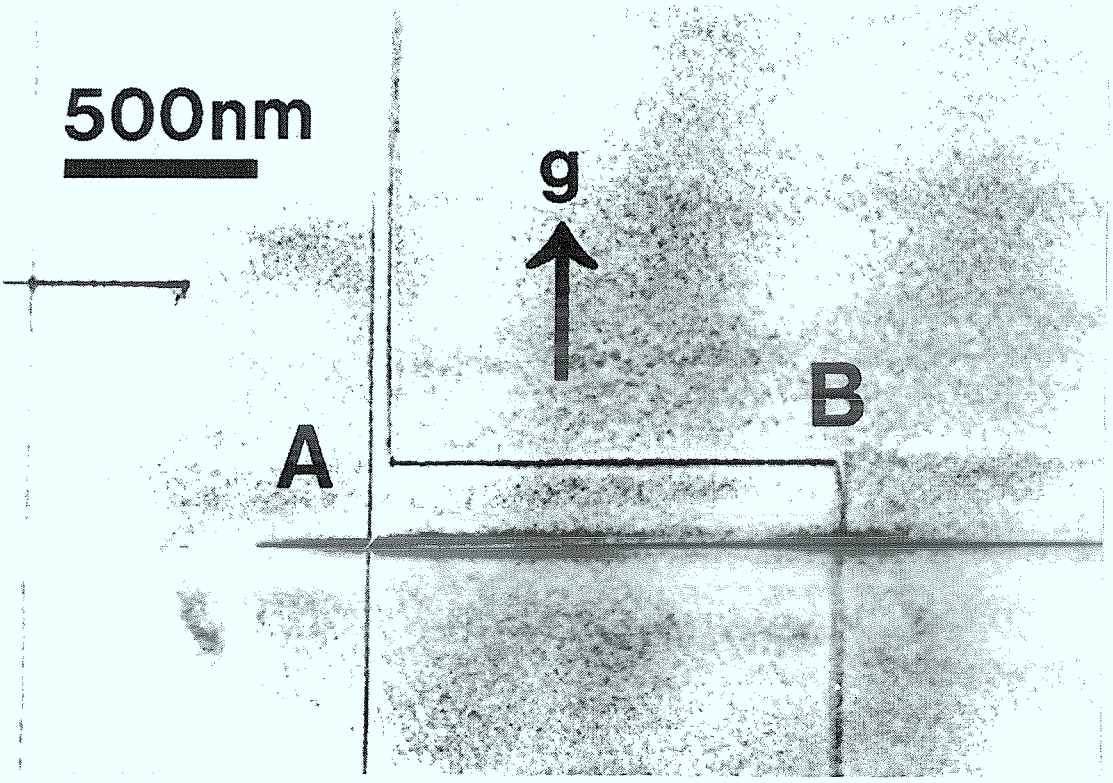


Abbildung 4.11: Hellfeldabbildung, $\vec{g} = (220)$, einer mehrfach abknickenden Fehlpassungsversetzung an einer Übergittergrenzfläche. Das Fehlpassungssegment wechselt die Linienrichtung bei (A) von einer auf die andere ⟨110⟩-Richtung, und bei (B) wieder zurück auf die ursprüngliche Richtung.

hintereinanderliegenden Stellen (A) und (B) abknickt.

Es wurden außerdem Konfigurationen beobachtet, die aus einer Aufeinanderfolge solcher abknickender Fehlpassungssegmente bestehen. In einigen Fällen bilden zwei abknickende Fehlpassungssegmente einer Versetzungslinie eine U-förmige Anordnung, die sich in einer auf der gleichen Grenzfläche liegenden, benachbarten U-förmigen Anordnung so wiederholt, daß die Verbindungssegmente, die jeweils den unteren Schenkel einer U-Konfiguration bilden, auf einer gemeinsamen Verbindungslinie liegen. Eine solche Konfiguration ist in Abbildung 4.12 gezeigt, mit den Eckpunkten (A), (B) der ersten U-Konfiguration und (C), (D) den Eckpunkten der zweiten U-Konfiguration. Die Punkte (A), (B), (C) und (D) liegen auf einer gemein-

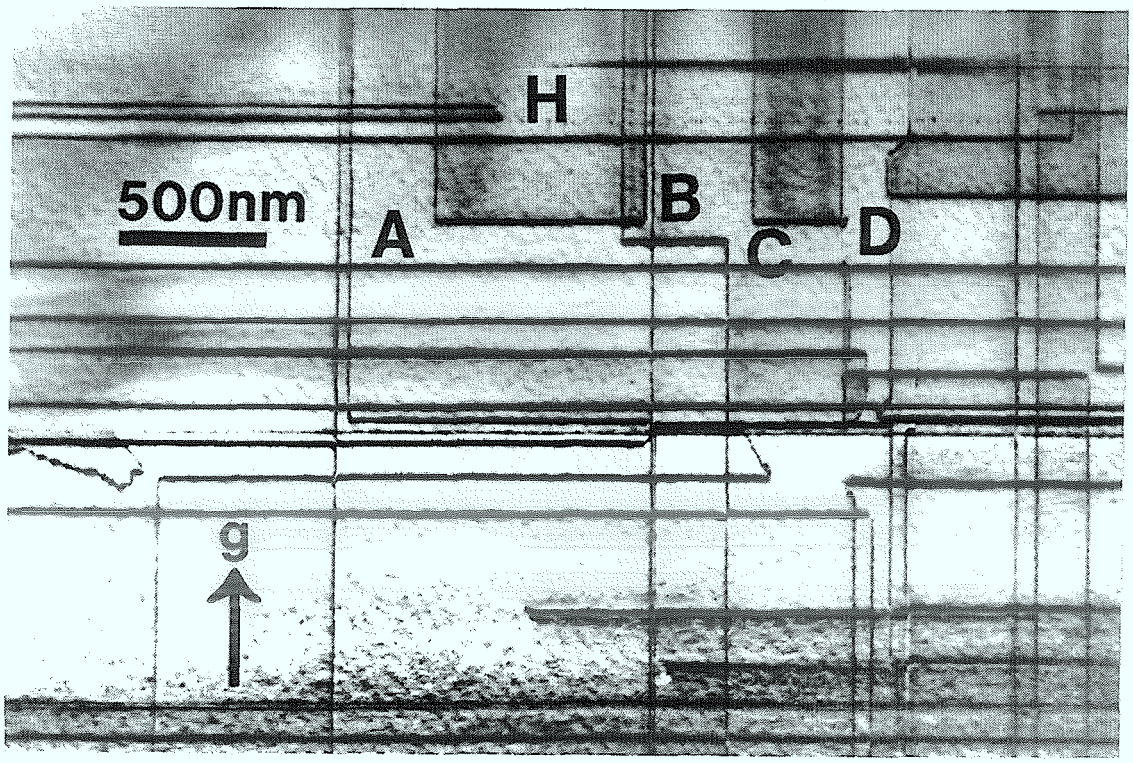


Abbildung 4.12: Hellfeldabbildung, $\vec{g} = (220)$, einer Doppel-U-Konfiguration zweier rechtwinklig abknickender Fehlpassungsversetzungen. Die Eckpunkte (A), (B) und (C), (D) der beiden U-Konfigurationen liegen auf einer gemeinsamen Verbindungsgeraden. Bei (H) ist eine Haarnadelkonfiguration eines Dipols zweier Fehlpassungsversetzungen zu erkennen.

samen Geraden. Der dunklere Kontrast innerhalb der beiden U-Konfigurationen ist unter Berücksichtigung der Richtung des Abbildungsvektors auf eine Probenverbiegung aufgrund der Stufenkomponente der Segmente zwischen den Punkten (A), (B) und (C), (D) zurückzuführen (Abschnitt 4.3). Die Abbildung enthält außerdem noch weitere Konfigurationen abknickender Fehlpassungssegmente, sowie eine Haarnadelkonfiguration bei (H) (siehe Abschnitte 4.3, 4.4).

Zwei Gründe sind dafür verantwortlich, daß eine Versetzung während der Relaxation die Möglichkeit hat, die Gleitebene zu wechseln. Erstens ist die fehlgepaßte Schicht zweiachsig verspannt, das heißt es wirkt sowohl eine Spannung entlang der [110]- als auch entlang der [$\bar{1}$10]-Richtung. Zweitens sind die Versetzungen im vorliegenden Fall vom 60°-Typ, dessen Burgersvektor in drei Komponenten aufgespalten werden kann: eine Schraubenkomponente entlang der Linienrichtung, eine Stufenkomponente entlang der Schichtnormalen und eine Stufenkomponente in der Grenzflächenebene. Letztere ist die Komponente, die zur Relaxation beiträgt, die andere Stufenkomponente und die Schraubenkomponente haben den gleichen Betrag. Der Burgersvektor einer 60°-Versetzung hat außerdem die Eigenschaft, zwei verschiedenen {111}-Ebenen anzugehören.

Das rechtwinklige Abknicken ist also möglich, indem die Versetzung während des Gleitens des Fadensegments von der einen auf die andere Gleitebene wechselt. Dabei vertauschen die Stufenkomponente in der Grenzfläche und die Schraubenkomponente ihre Rollen, während die Stufenkomponente entlang der Schichtnormalen erhalten bleibt.

Durch diesen Mechanismus wird die Relaxation entlang der ersten Spannungsachse durch die Versetzung beendet, während sie die Relaxation entlang der zweiten Achse unterstützt.

4.5.2 Diskussion

Zwei Ursachen sind grundsätzlich für einen Wechsel der Gleitebene während der Relaxation denkbar.

Die erste ist die Tendenz, den entlang der x- und y-Achsen symmetrischen Verspannungszustand auf möglichst symmetrische Weise abzubauen. Das heißt, wenn an irgendeiner Stelle des fehlgepaßten Schichtsystems zum Beispiel die xx-Komponente des Spannungstensors stärker relaxiert sein sollte, als die yy-Komponente, dann wird wegen der höheren Restspannung das y-Gleitsystem mit höherer Wahrscheinlichkeit angeregt, als das x-Gleitsystem.

Die zweite Ursache ist eine mögliche Hemmung des Gleitprozesses entlang einer Richtung durch ein Hindernis, zum Beispiel eine quer zur ursprünglichen Gleitrichtung verlaufende Fehlpassungsversetzung mit gleichem Burgersvektor. Dabei wird das Fadensegment durch die zusätzliche Spannung des Hindernisses blockiert, und es besteht nur noch eine Beweglichkeit entlang der zweiten Richtung [83]. Das Fadensegment muß dann zunächst eine passende ⟨110⟩-Linienrichtung annehmen, bevor es die Gleitebene wechseln kann.

Es gibt einen dritten Mechanismus, der jedoch nur aktiviert werden kann, wenn eine fehlgepaßte Schicht zuerst relaxiert, ein Fehlpassungsnetzwerk ausbildet, und

dann in der Wachstumssequenz eine Deckschicht mit einer entgegengesetzten Spannung aufgewachsen wird, die unterhalb einer kritischen Schichtdicke mit der ersten eine kohärente Grenzfläche bildet. Nach den Prinzipen der Minimierung der freien Energie wird die Spannung der ersten Schicht durch die kohärente zweite Schicht reduziert, so daß das bis dahin gebildete Fehlpassungsnetzwerk eine zu große Dichte hat. Diese Situation ist nicht nur für die erste InGaAs-Schicht relevant, die direkt auf das Substrat aufwächst, sondern auch für alle folgenden Einzelschichten, deren Schichtdicke überkritisch bezüglich der Einzelschichtrelaxation ist.

Das Netzwerk muß folglich zum Erreichen des neuen Gleichgewichts Versetzungslinien eliminieren. Bei genügend dünnen Schichten ist dies durch die Bildkraft (Abschnitt 2.4.2) auf einzelne Versetzungssegmente möglich, die die reduzierte Fehlpassungsspannung dominiert. In Analogie zum Hagen-Strunk-Mechanismus [80, 168] (Abschnitt 2.7.2) können sich an Überkreuzungen zweier Versetzungen mit gleichem Burgersvektor zwei zueinander rechtwinklige Segmente vom Versetzungsknoten ablösen und zur Oberfläche gleiten. Die beiden resultierenden Fadensegmente können dann unter Wirkung der Bildkraft vom Knoten weggleiten und dadurch die mit ihnen verbundenen Fehlpassungssegmente verkürzen und schließlich eliminieren. Die beiden anderen zurückbleibenden Segmente bilden dann die beobachtete rechtwinklige Konfiguration.

Mit großer Wahrscheinlichkeit ist die beobachtete „Doppel-U"-Konfiguration von Fehlpassungssegmenten (Abbildung 4.12) ein Ergebnis dieses Eliminationsmechanismus. Eine unabhängige Entstehung der beiden U-Konfigurationen hat nämlich nicht notwendigerweise eine exakte Ausrichtung der beiden Verbindungssegmente entlang einer Linie zur Folge, während sie mit dem obigen Mechanismus zwanglos durch das Eliminieren des dazwischenliegenden Segmentes (B)–(C) aus dem ursprünglich vollständigen rechtwinkligen Netzwerk erklärbar ist.

4.6 Superkinken

4.6.1 Experimentelles Resultat

In Querschnittaufnahmen der Übergitter wurde eine Versetzungskonfiguration aus zwei Fehlpassungssegmenten gefunden, die sich an zwei verschiedenen Grenzflächen zwischen den Einzelschichten der Übergitterperioden befinden. Sie haben einen Abstand, der jeweils genau einer vollständigen Periode aus einer GaAs- und einer InGaAs-Schicht entspricht, und sie sind durch ein Fadensegment verbunden, das die beiden dazwischenliegenden Schichten durchdringt. Die gesamte Versetzungskonfiguration ist vom gleitfähigen 60°-Typ.

Abbildung 4.13 zeigt eine solche *Superkinke*, die mit (S) gekennzeichnet ist, zwischen zwei Fehlpassungssegmenten (F) an den Übergittergrenzflächen eines (18/6)-Übergitters. Ein zweites Beispiel ist in Abbildung 4.14 dargestellt.

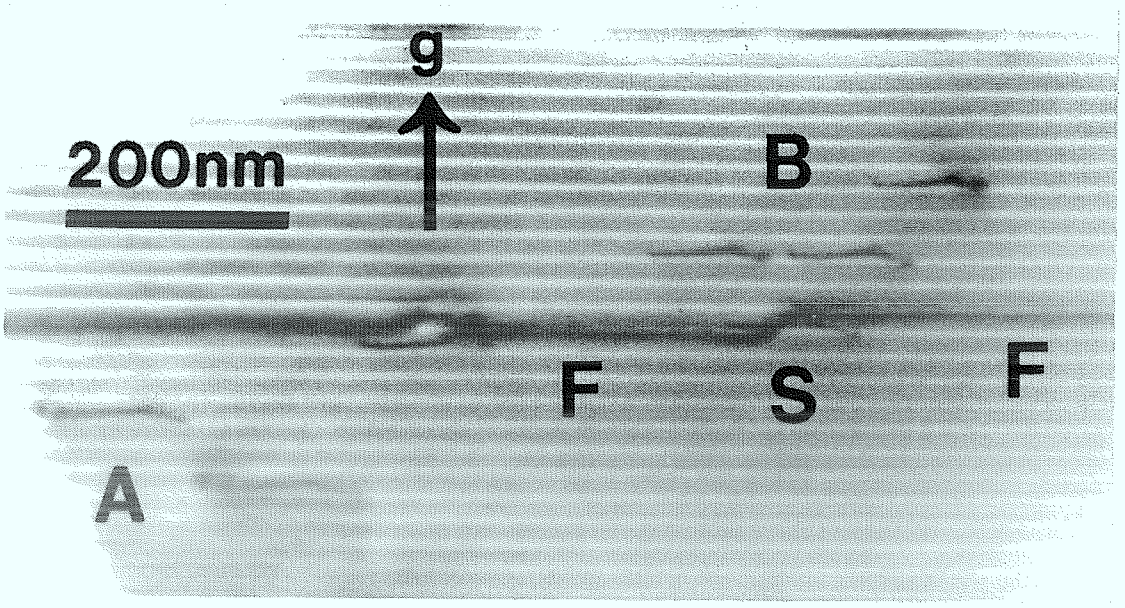


Abbildung 4.13: Hellfeldabbildung, $\vec{g} = (004)$, einer Superkinke (S) zwischen zwei Fehlpassungsversetzungen (F) an den Übergittergrenzflächen eines (18/6)-Übergitters. Die Kinkenhöhe ist gleich der Dicke einer vollständigen Periode aus einer GaAs- und einer InGaAs-Einzelschicht. Bei (A) und (B) befinden sich Fehlpassungssegmente des orthogonalen Versetzungsnetzwerks.

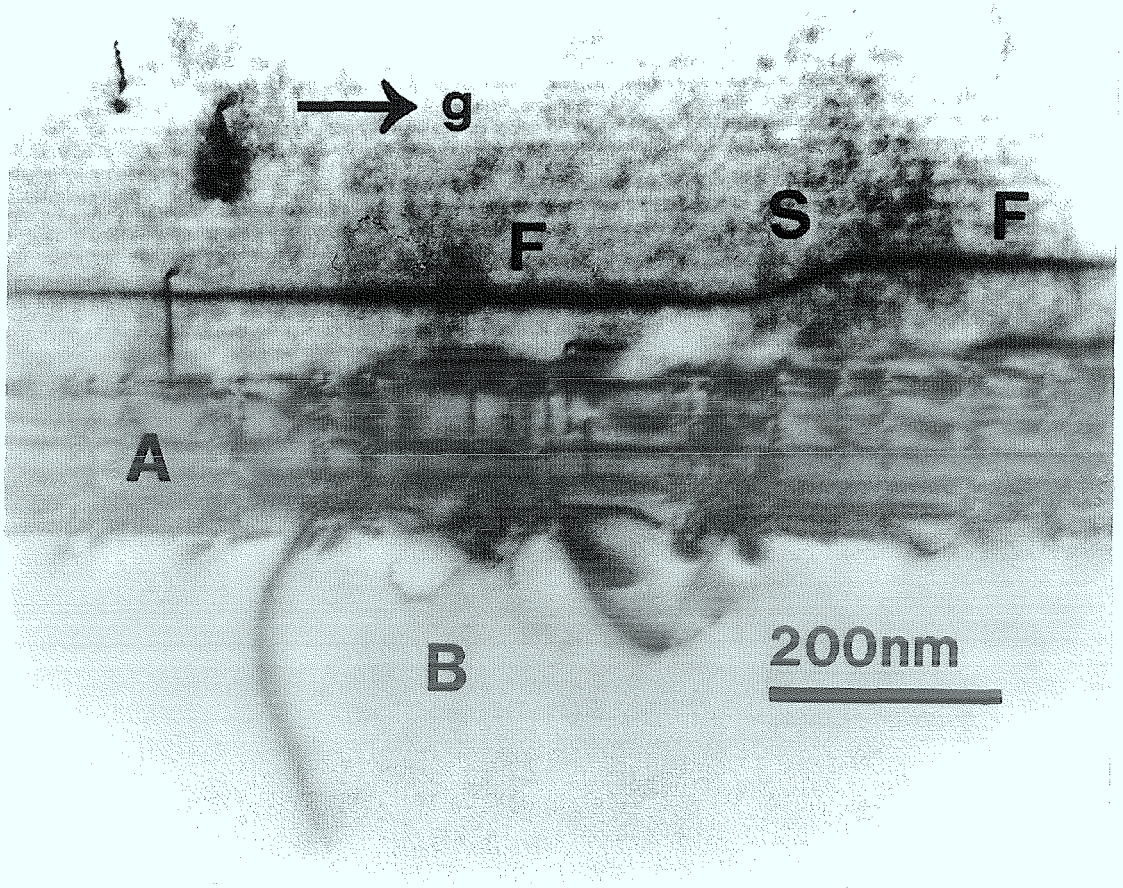


Abbildung 4.14: Hellfeldabbildung, $\vec{g} = (220)$, einer Superkinke (S) zwischen zwei Fehlpassungsversetzungen (F) an den Übergittergrenzflächen eines (18/6)-Übergitters. Die Kinkenhöhe ist gleich der Dicke einer vollständigen Periode aus einer GaAs- und einer InGaAs-Einzelschicht. Bei (A) befindet sich das dichte Fehlpassungsnetzwerk an der Substratgrenzfläche, bei (B) sind Versetzungsschleifen unterhalb der Substratgrenzfläche zu erkennen.

4.6.2 Diskussion

Die beobachtete Superkinke zwischen den beiden Fehlpassungssegmenten ist das Produkt des Gleitprozesses einer Versetzung, die ihren Ursprung an der Kristalloberfläche hat, und die unter der Wirkung der mittleren Fehlpassungsspannung des Übergitters zur tiefer liegenden Grenzfläche zwischen Übergitter und Substrat geglitten ist.

Kinken enstehen während des Gleitens im allgemeinen an ausgedehnten Hindernissen, die eine Versetzung aus energetischen Gründen nicht mit der gesamten Versetzungslinie gleichzeitig überwinden kann. Stattdessen gleitet nur ein kleines Segment über das Hindernis, wobei an beiden Enden des Segmentes unter Aufwenden einer bestimmten Bildungsenergie je eine Kinke erzeugt wird [83]. Anschließend wandern die beiden Kinken durch die äußere Spannung, die die Ursache des Versetzungsgleitens ist, in entgegengesetzte Richtungen auseinander [124]. Bei diesem letzten Prozeß muß eine relativ geringe Aktivierungsenergie aufgebracht werden, die in etwa der Höhe der Peierls-Nabarro-Barriere (Abschnitt 2.7.3) beim gewöhnlichen Versetzungsgleiten entspricht [83].

Die Zahl der beobachteten Superkinken ist jedoch sehr klein, weil unter normalen Umständen die Kinkenwanderungsenergie so klein ist, daß sie bei den im untersuchten Schichtsystem vorliegenden Fehlpassungsspannungen mobil sind, und praktisch so weit gleiten, bis das obere Fehlpassungssegment vollständig auf die untere Grenzfläche transferiert ist. In den beiden abgebildeten Fällen sind die beiden Kinken einer Doppelkinke sehr weit voneinander entfernt, so daß jeweils nur eine der beiden beobachtet wird. Die beiden einzelnen Kinken sind relativ weit von anderen Versetzungen entfernt, und sie sind vermutlich durch kleine Hindernisse oder lokale Spannungsfelder gestoppt worden.

Die Existenz von Superkinken in den Übergittern ist ein Beleg dafür, daß die Übergitterperioden selbst als Hindernisse für das Versetzungsgleiten wirken, und daß während des Relaxationsprozesses neue Versetzungen an der Kristalloberfläche gebildet werden. Die Auswirkung dieser kinetischen Behinderung des Relaxationsprozesses in Übergittern wird ausführlich in Kapitel 5 diskutiert.

4.7 Versetzungen im Substrat

4.7.1 Experimentelle Resultate

Das Fehlpassungsnetzwerk an der Grenzfläche zwischen Übergitter und Substrat wies in einigen Fällen eine Abweichung von der perfekten Orientierung der Versetzungslinien entlang der [110]- und der [$\bar{1}$10]-Richtungen auf. Teile der Fehlpassungssegmente hatten Halbschleifen gebildet, die bis zu einer Tiefe von ungefähr 100 nm in das Substratvolumen unterhalb der Grenzfläche reichten.

Die Abbildungen 4.15, 4.16 und 4.17 zeigen in Querschnittaufnahmen charakteristische Beispiele von Halbschleifen, deren Linien vom dichten Substratversetzungsnetzwerk ausgehen, in das Substrat führen und von dort zur Substratgrenzfläche zurück verlaufen. Bei der in Abbildung 4.18 dargestellten Querschnittaufnahme rei-

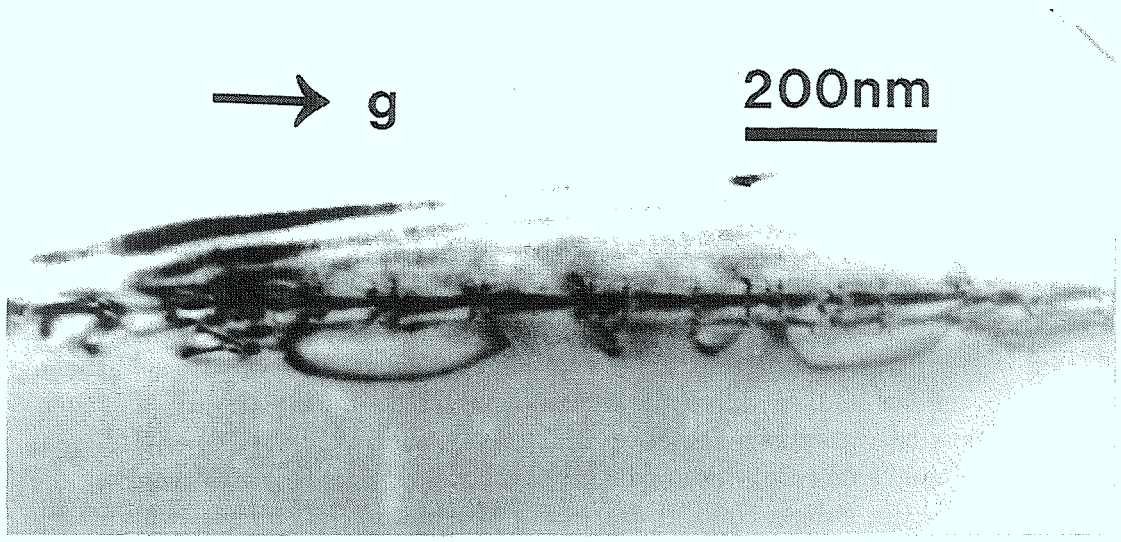


Abbildung 4.15: Hellfeldabbildung, $\vec{g} = (220)$, eines Querschnitts durch ein (18/6)-Übergitter. Unterhalb des dichten Substratversetzungsnetzwerks verlaufen Versetzungsschleifen in das Substrat hinein.

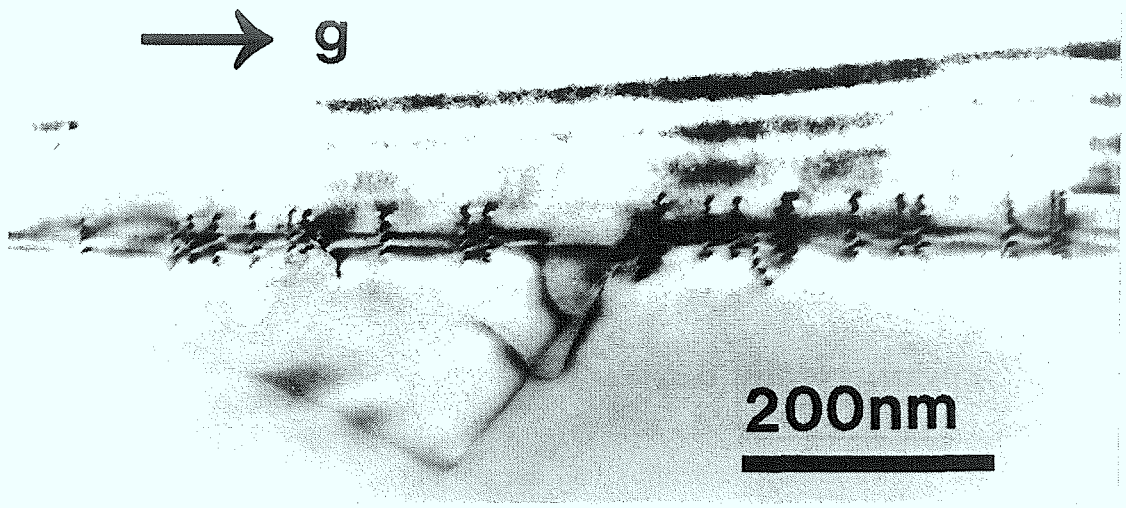


Abbildung 4.16: Hellfeldabbildung, $\vec{g} = (220)$, eines Querschnitts durch ein (18/6)-Übergitter. Unterhalb des dichten Substratversetzungsnetzwerks verlaufen Versetzungsschleifen in das Substrat hinein.

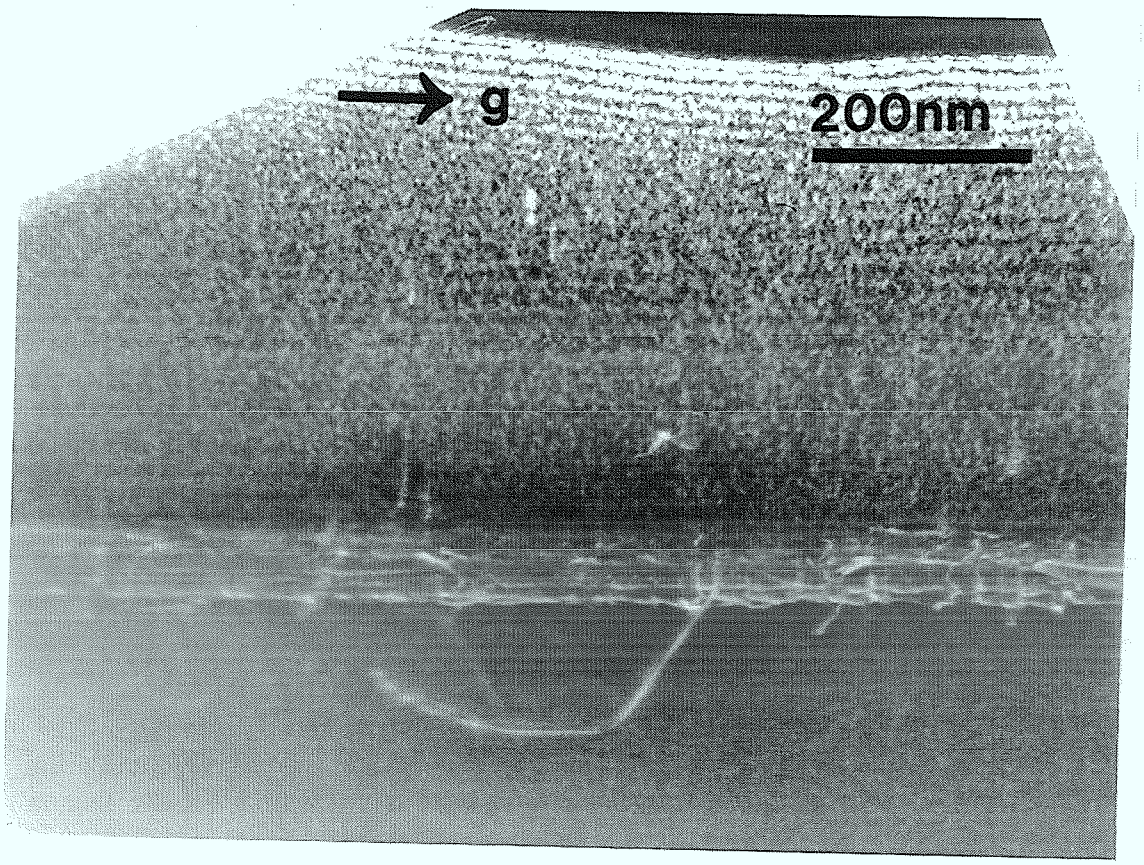


Abbildung 4.17: Weak-Beam-Dunkelfeldabbildung, $\vec{g} = (220)$, $(\vec{g}/4\vec{g})$ eines Querschnitts durch ein (18/2)-Übergitter. Unterhalb des dichten Substratversetzungsnetzwerks verläuft eine Versetzungsschleife in das Substrat hinein.

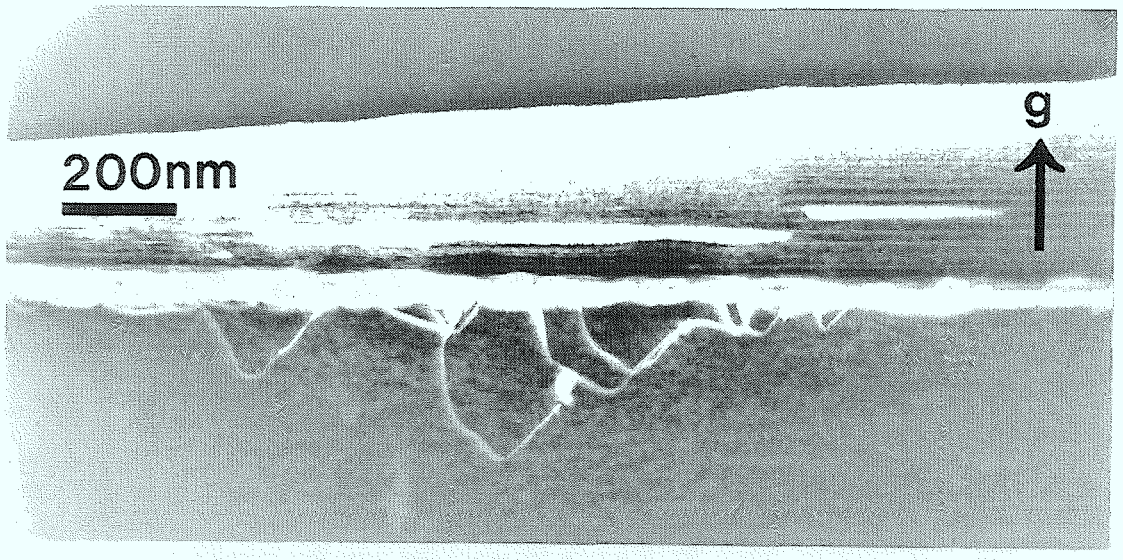


Abbildung 4.18: Weak-Beam-Dunkelfeldabbildung, $\vec{g} = (004)$, $(\vec{g}/4\vec{g})$ eines Querschnitts durch ein (18/6)-Übergitter. Unterhalb des dichten Substratversetzungsnetzwerks verläuft eine Versetzungsschleife etwa 200 nm in das Substrat hinein. Es sind außerdem zwei geradlinige Fehlpassungssegmente an den Übergittergrenzflächen in Kontrast.

chen die Schleifen bis in eine Tiefe von 200 nm in das Substrat, es sind außerdem zwei Fehlpassungssegmente an den Übergittergrenzflächen zu erkennen. In der Querschnittaufnahme von Abbildung 4.19 sind große Teile der Versetzungsschleifen präparationsbedingt abgeschnitten, zum Beispiel bei (S). Im Übergitter ist bei (A) ein gerades Fehlpassungssegment zu erkennen, das mit einem relativ langen Fadensegment (B) verbunden ist. Es ist wegen der geringen Dichte der Substratversetzungsschleifen kein signifikanter Unterschied der Eindringtiefen in den hier untersuchten Übergittern zu beobachten.

In einem Fall konnte eine Versetzung beobachtet werden, Abbildung 4.20, die sich von den anderen Halbschleifensegmenten dadurch unterscheidet, daß sie sehr schwach gekrümmt ist und sehr weit in das Substrat hineinreicht, bevor ihre Versetzungslinie die Oberfläche der dünnen elektronenmikroskopischen Probe durchstößt. Sie ist sehr wahrscheinlich eine der schon von der GaAs-Kristallzüchtung her vorhandenen Versetzungen, die Multiplikationsquellen für Fehlpassungssegmente während des ersten Relaxationsstadiums sind, bei dem noch keine neuen Versetzungssegmente erzeugt werden (Abschnitt 2.7.2).

Ein dritter Defekt, der jedoch nur sehr selten beobachtet wurde, ist das Aufspalten eines Fehlpassungssegments in zwei Partialversetzungen mit einem dazwischenliegenden Stapelfehler. Dabei verbleibt die eine Partiale an der Substratgrenzfläche und die andere gleitet ein Stück weit in das Substrat hinein. In der in Abbildung 4.21 dargestellten Querschnittaufnahme sind die Partialen (P1), (P2) in Kontrast, und

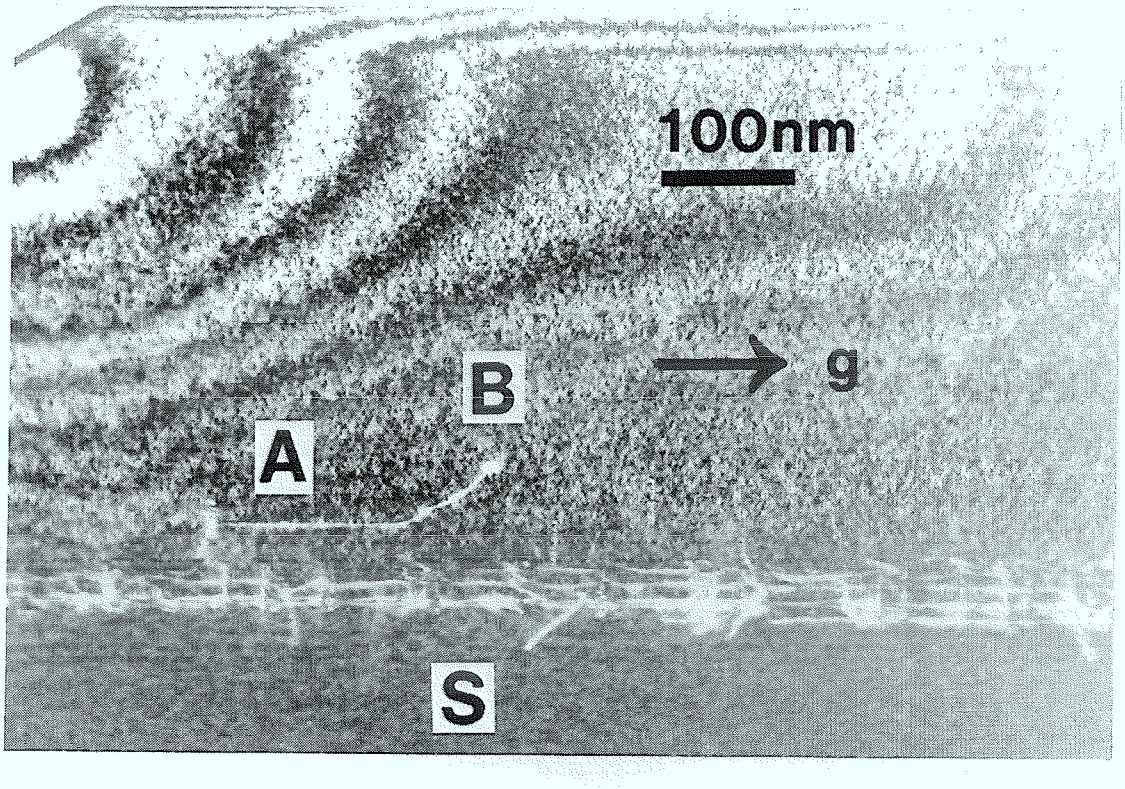


Abbildung 4.19: Weak-Beam-Dunkelfeldabbildung, $\vec{g} = (220)$, $(\vec{g}/4\vec{g})$ eines Querschnitts durch ein (18/2)-Übergitter. Große Teile der Versetzungsschleifen sind präparationsbedingt abgeschnitten (S). Im Übergitter ist ein geradliniges Fehlpassungssegment (A) mit einem relativ langen Fadensegment (B) verbunden.

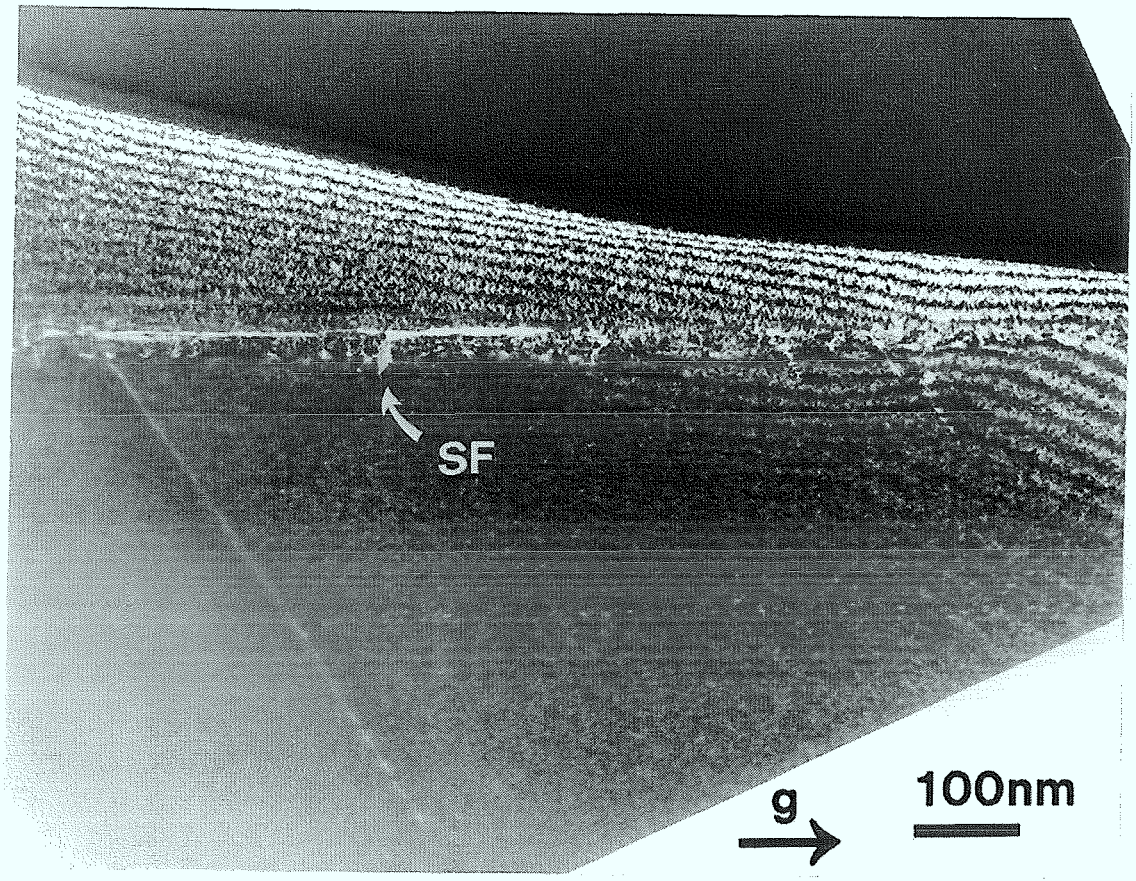


Abbildung 4.20: Weak-Beam-Dunkelfeldabbildung, $\vec{g} = (220)$, $(\vec{g}/4\vec{g})$ eines Querschnitts durch ein (18/2)-Übergitter. Von der Substratgrenzfläche reicht eine geradlinige Versetzungslinie sehr weit in das Substratvolumen hinein. An der Substratgrenzfläche ist bei (SF) ein Stapelfehlerkontrast zu erkennen.

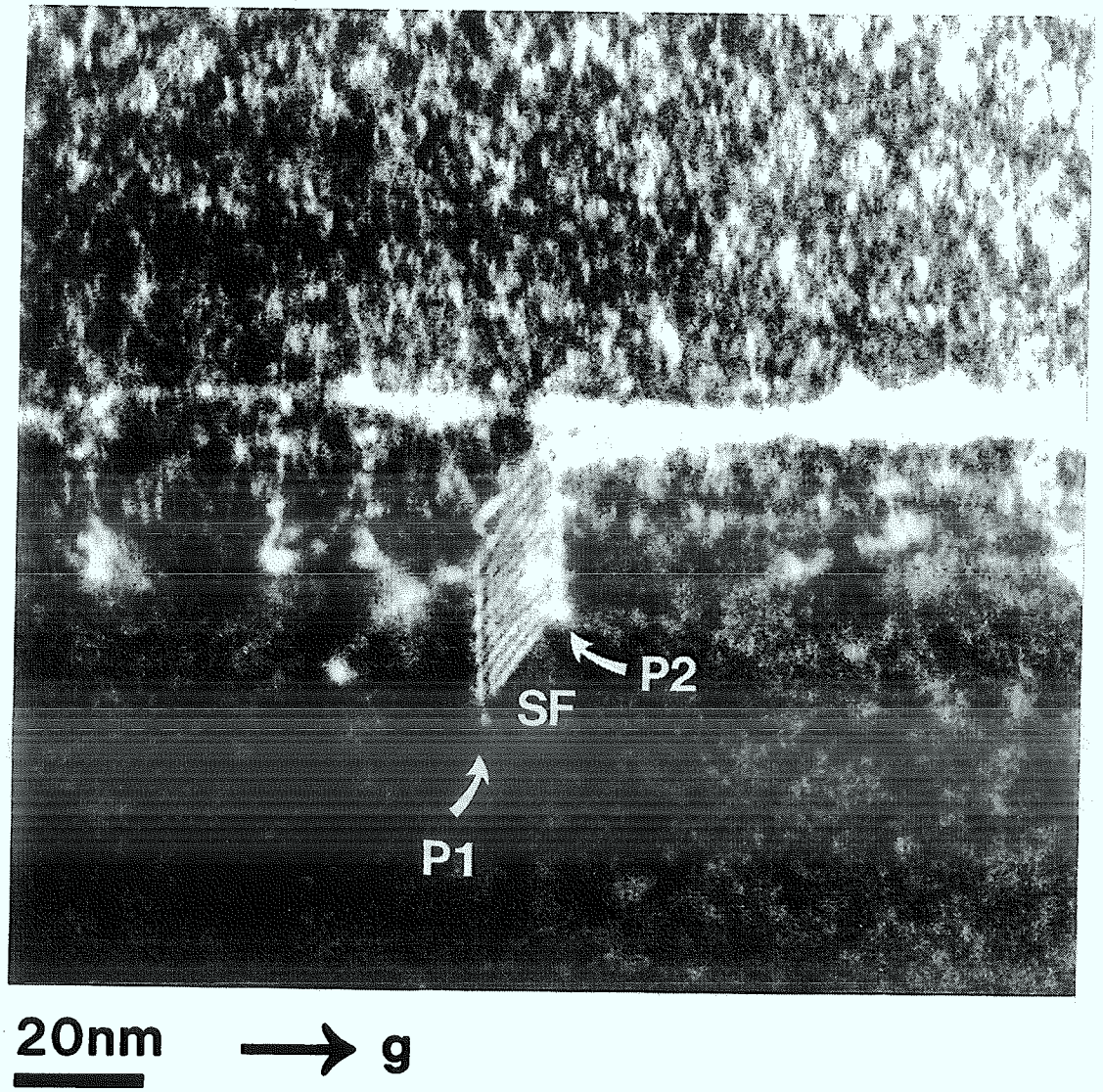


Abbildung 4.21: Weak-Beam-Dunkelfeldabbildung, $\vec{g} = (220)$, $(\vec{g}/4\vec{g})$ eines Querschnitts durch ein (18/2)-Übergitter. Eine Fehlpassungsversetzung an der Substratgrenzfläche ist in zwei Partialversetzungen (P1), (P2) aufgespalten. (P1) befindet sich im Substrat, (P2) liegt im Netzwerk an der Substratgrenzfläche. Der dazwischenliegende Stapelfehler erzeugt einen typischen Streifenkontrast.

es ist der für Stapelfehler typische Streifenkontrast zu erkennen. Die beobachtete Aufspaltungsweite beträgt etwa 10 nm.

Es ist noch wichtig zu bemerken, daß die Halbschleifenbildung in keinem einzigen Fall an den Grenzflächen der einzelnen Übergitterperioden beobachtet wurde.

4.7.2 Diskussion

Vom Standpunkt der Gleichgewichtsmodelle ist es grundsätzlich energetisch ungünstig, daß eine Versetzungsschleife im Substrat gebildet wird, die das gleiche Vorzeichen des Burgersvektors wie ein Fehlpassungssegment hat. Die zur Erzeugung der Halbschleife benötigte Energie kann nämlich nicht durch eine Reduzierung der Substratverspannung kompensiert werden, weil das Substrat viel dicker als ein aufgewachsenes Übergitter ist, und deshalb praktisch nicht verspannt ist.

Die Existenz von Halbschleifen zeigt jedoch, daß zumindest eine lokale Substratverspannung vorhanden ist. Die Dichte der Halbschleifensegmente ist sehr viel kleiner als die Dichte eines Fehlpassungsnetzwerkes an der Substratgrenzfläche, so daß die Modellannahme eines starren Substrats für die Berechnung der Gleichgewichtsdichte von Fehlpassungsversetzungen praktisch weiter gültig bleibt.

Das Aufspalten von Fehlpassungssegmenten in zwei Partialversetzungen und einen dazwischenliegenden Stapelfehler ist hier ebenfalls ein Hinweis auf die Existenz einer lokalen Substratverspannung. Im hier untersuchten System haben die Grenzflächenversetuzngen, wenn überhaupt, eine wesentlich kleinere Aufspaltungsweite, die durch die Energiebilanz aus der Energie der Partialen, ihrer Wechselwirkung und der Stapelfehlerenergie bestimmt ist. Das Vergrößern der Stapelfehlerfläche ist offensichtlich nur dann möglich, wenn das Abgleiten einer der Partialen unter der Wirkung des äußeren Spannungsfeldes der lokalen Substratverzerrung energetisch günstig ist. Mit Hilfe der konventionellen Elektronenmikroskopie kann eine Aufspaltung grundsätzlich nur auf einer Längenskala beobachtet werden, die mindestens so groß wie die erreichbare Kontrastbreite ist (Abschnitt 3.2.2). Die hochauflösende Abbildung der atomaren Säulenstruktur einer 60°-Grenzflächenversetzung zeigte jedoch eindeutig, daß die Fehlpassungssegmente nicht aufgespalten sind.

Ein zweiter Grund für die Existenz von Halbschleifen ist eine Eigenart der Schichtsysteme mit niedriger Fehlpassung. Eine Berechnung der Gesamtenergie in Abhängigkeit von der Eindringtiefe einer zur Kristalloberfläche parallelen Versetzungslinie bei konstanter Schichtdicke ergibt, daß für die Lage direkt an der Grenzfläche ein Energieminimum existiert [145]. An dieser Stelle hat die Versetzung einerseits den maximal möglichen Energiebetrag durch Relaxieren der Schichtverspannung freigesetzt, andererseits muß für ein weiteres Gleiten in das Substrat hinein Energie aufgewendet werden.

Das Minimum ist für Schichten mit niedriger Fehlpassung flacher als für höhere Fehlpassungen, so daß bei gleicher Eindringtiefe in das Substrat eine kleinere Aktivierungsenergie pro Versetzung eines Netzwerks aufgewendet werden muß.

Die Abwesenheit von Halbschleifen an den Grenzflächen der einzelnen Übergitterperioden ergibt sich in diesem Zusammenhang zwanglos, indem man berücksichtigt, daß zur Erzeugung der Halbschleife nicht nur Energie zur Verlängerung der Ver-

setzungslinie aufgewendet werden muß, wie bei einer Substratschleife, sondern daß zusätzlich noch die mechanische Arbeit gegen die Spannung der darunterliegenden Schicht geleistet werden muß. Diese zusätzliche Aktivierungsenergie verhindert effektiv das Abgleiten von Fehlpassungssegmenten in die darunterliegende Einzelschicht und stabilisiert so die Versetzungsnetzwerke, obwohl auch hier die relative Fehlpassung klein ist. Der hier diskutierte Fall der Halbschleife, die in die darunterliegende Schicht eindringt, ist vom später besprochenen Fall der *Doppelkinke* zu unterscheiden, die die darunterliegende Schicht vollständig durchdringt (Abschnitt 5.5.4).

Eine dritte treibende Kraft für das Versetzungsgleiten in das Substrat resultiert aus dem Typ der Fehlpassungsversetzungen. In der Diamant- und Zinkblendestruktur gibt es Gleitsysteme, die Verspannungen entlang der beiden ⟨110⟩-Richtungen relaxieren, und die trotz verschiedener Gleitebenen gleiche Burgersvektoren haben (siehe auch Abschnitt 4.5). Das heißt, innerhalb der zueinander orthogonalen Fehlpassungsnetzwerke gibt es Versetzungsknoten, an denen sich zwei Linien mit parallelen Burgersvektoren überkreuzen.

In diesem speziellen Fall existiert eine repulsive Wechselwirkung zwischen beiden Versetzungslinien [83], die eines der beiden Segmente in das Substrat drücken kann. Der umgekehrte Prozeß, nämlich daß das Segment nach oben in die aufgewachsene fehlgepaßte Schicht gleitet, ist weniger wahrscheinlich, denn es muß dabei zusätzlich die Wechselwirkungsenergie zwischen Versetzung und verspannter Schicht aufgewendet werden, die ja bei der Relaxation der Schicht durch das Gleiten der Versetzung zur Grenzfläche freigeworden war. Dieser Mechanismus ist bei den in höher fehlgepaßten Schichten auftretenden Lomer-Versetzungen nicht möglich, weil sie nicht gleitfähig sind.

4.8 Plan-View-Streifenkontrast der Übergitter

4.8.1 Experimentelles Resultat

In den keilförmig rückgedünnten elektronenmikroskopischen Proben wurde ein charakteristisches System aus abwechselnd hellen und dunklen Streifen beobachtet. Sie befinden sich an Probenstellen, die nur eine Projektion durch die Übergitter enthalten, und an denen das Substrat präparationsbedingt entfernt ist. Die Streifenbreiten stehen im zahlenmäßig gleichen Verhältnis wie die Einzelschichtdicken der entsprechenden Übergitter. Häufig enden die Linien von Fehlpassungssegmenten exakt an der Grenze zwischen einem hellen und einem dunklen Streifen. Insbesondere enden die Streifensysteme exakt an der Grenzfläche zwischen Übergitter und Substrat, und die Berandung des letzten Streifens an der Substratgrenzfläche enthält die Durchstoßpunkte des dichten Substratversetzungsnetzwerks.

Abbildung 4.22 zeigt eine Projektion durch die Übergitterperioden eines (18/25)-Übergitters. Die dunklen Streifen korrespondieren mit den Austrittsflächen der InGaAs-Einzelschichten an der zur [001]-Richtung geneigten Probenunterseite, helle Streifen korrespondieren mit den Austrittsflächen der GaAs-Einzelschichten. An den Stellen (A), (B), (C) und (D) sind Durchstoßpunkte von Fehlpassungsversetzungen

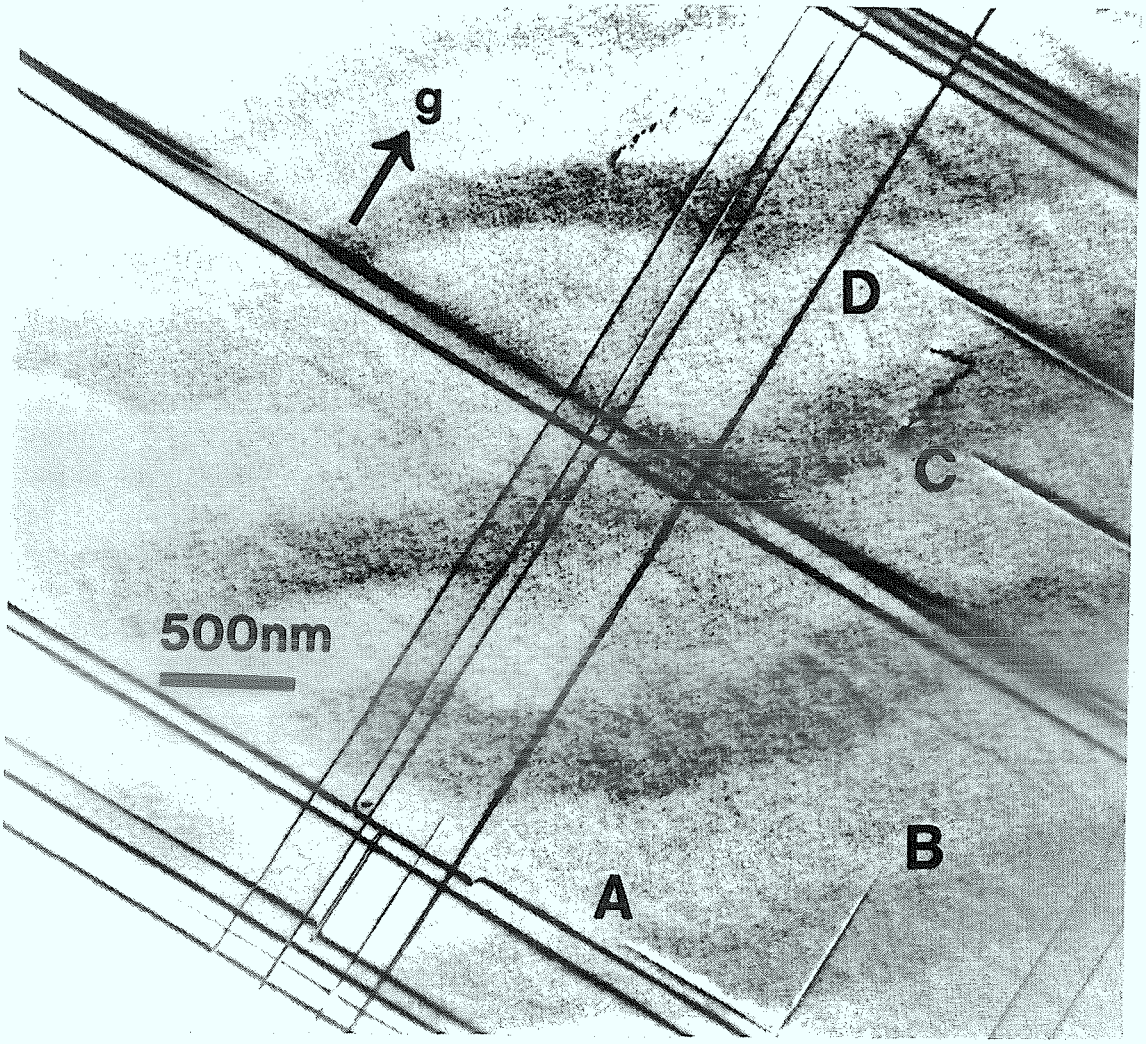


Abbildung 4.22: Hellfeldabbildung, $\vec{g} = (220)$, eines (18/25)-Übergitters. Die Aufsicht enthält ein System aus breiten, hellen und dunklen Sreifen, die mit den Austrittsflächen von GaAs- (hell) und InGaAs-Einzelschichten (dunkel) korrespondieren. An den Orten (A), (B), (C) und (D) befinden sich Austrittspunkte von Fehlpassungsversetzungen an den Übergittergrenzflächen, die exakt auf einer Streifengrenze enden.

an den Übergittergrenzflächen zu erkennen, die sich jeweils exakt an der Grenze zwischen einem hellen und einem dunklen Streifen befinden.

Abbildung 4.23 zeigt einen Probenabschnitt, der sowohl eine Projektion nur durch das Übergitter, als auch eine Projektion durch Übergitter und Substrat enthält. An der Grenzlinie zwischen beiden Abschnitten schneidet die Substratgrenzfläche die Probenunterseite, und deshalb ist dort eine relativ dichtbesetzte Linie von Durchstoßpunkten des dichten Fehlpassungsnetzwerks zu erkennen. Nahe der Grenzlinie ist ein abknickendes Fehlpassungssegment an einer Übergittergrenzfläche zu erkennen. Dunkle Streifen korrespondieren wieder mit den InGaAs-Einzelschichten, helle Streifen mit den GaAs-Einzelschichten.

4.8.2 Diskussion

Die Streifensysteme sind ein direktes Abbild der GaAs- und InGaAs-Einzelschichten, und die Begrenzungslinien zwischen den Streifen sind die Durchstoßlinien der Übergittergrenzflächen an der Probenunterseite. Die Position und die Breite der Streifen folgen aus der Probengeometrie, die in erster Linie durch den Einfallswinkel der Argonionen während des Ionendünnens gegeben ist (Abschnitt 3.2.1).

In der Literatur [144] werden zwei mögliche Ursachen für das beobachtete Kontrastphänomen diskutiert. Die erste resultiert aus der Pendellösung der dynamischen Beugung, die eine Intensitätsvariation in Abhängigkeit von der Probendicke zur Folge hat (Abschnitt 3.2.2). Da die Streifen exakt mit der Projektion der Einzelschichten korrespondieren, wären sie das Ergebnis einer Dickenvariation, die zum Beispiel durch vorzugsweises Abtragen eines der beiden Materialien während der Ionendünnung entstehen kann.

Die zweite mögliche Ursache ist eine lokale Verkippung eines Kristallbereichs nahe der unteren Grenzfläche der keilförmigen Probe, die bei der Pendellösung der dynamischen Beugung ebenfalls zu einer Intensitätsänderung führt. Im hier abgebildeten Fall gibt es keine Überlagerung mit Kontrasten gleicher Probendicke, weil die lokale Probendicke bei etwa 800 nm liegt, wo die Dickenoszillation der Intensität des transmittierten und des abgebeugten Strahls wegen Absorption schon abgeklungen ist.

Die GaAs- und die InGaAs-Einzelschichten stehen in einem Übergitter unter Verspannungen mit entgegengesetztem Vorzeichen (Abschnitt 2.6), und die entsprechenden Komponenten des Spannungstensors erfüllen an der unteren Probenfläche, die wegen des flachen Dünnungswinkels zu den (001)-Grenzflächen geneigt ist, nicht mehr die Randbedingung der freien Oberfläche (Gleichung (2.4)). Die resultierende Relaxation nahe der Oberfläche, die das elastische Gleichgewicht wiederherstellt, bewirkt dann eine lokale Kristallkippung, die für die GaAs- und InGaAs-Schichten ein unterschiedliches Vorzeichen hat.

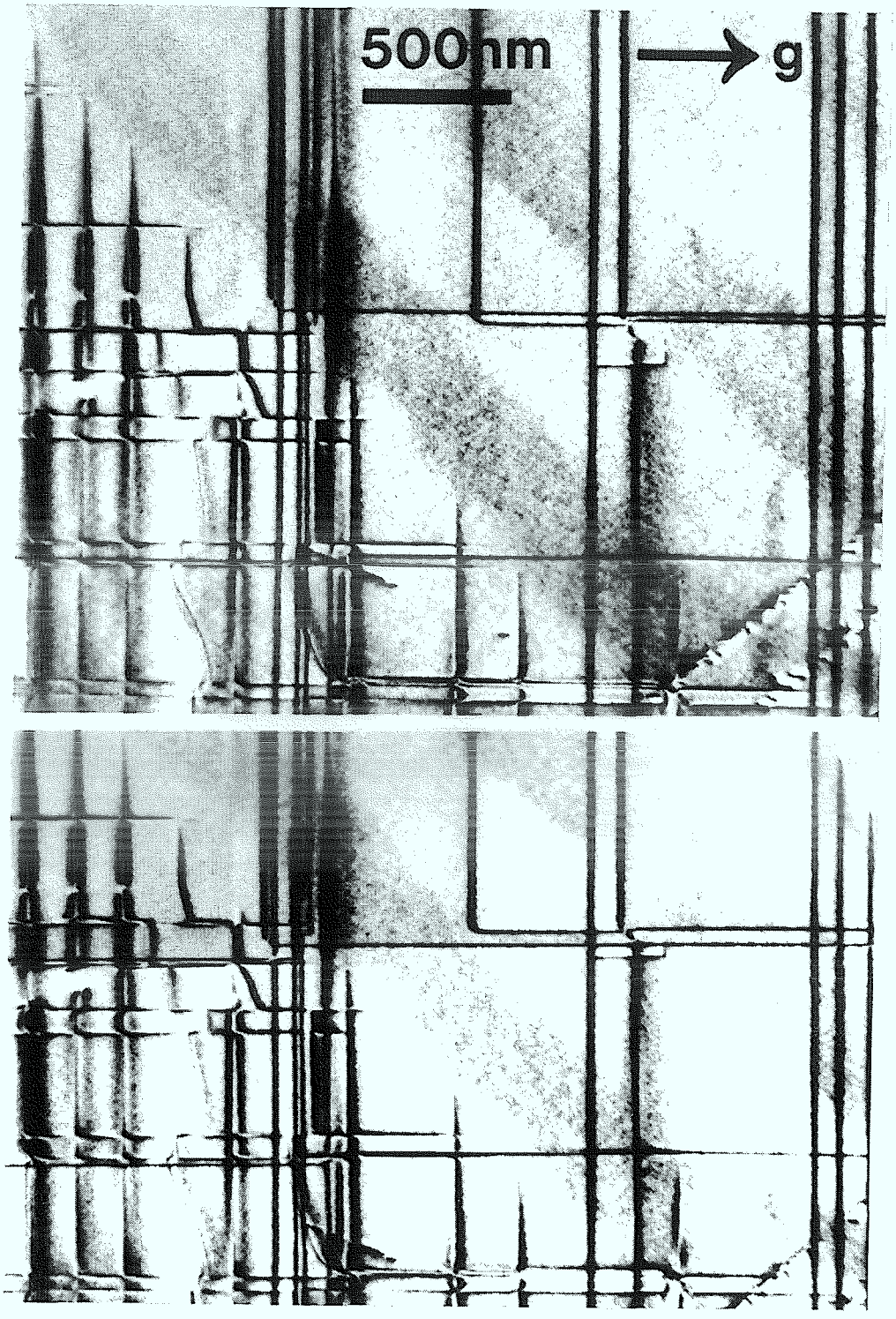


Abbildung 4.23: Stereopaar zweier Hellfeldabbildungen, $\vec{g} = (220)$, eines (18/25)-Übergitters. Der Projektion durch das Übergitter ist ein System aus breiten, hellen und dunklen Sreifen überlagert, die mit den Austrittsflächen von GaAs- (hell) und InGaAs-Einzelschichten (dunkel) korrespondieren. Der linke untere Teil enthält zusätzlich eine Projektion der Substratgrenzfläche, an der Grenzlinie zwischen beiden Projektionen endet das dichte Substratversetzungsnetzwerk.

4.9 Versetzungsdichten

4.9.1 Experimentelle Resultate

Neben einer Voraussage der kritischen Dicke einer fehlgepaßten Schicht, oder eines Satzes kritischer Schichtdicken im Fall eines Übergitters, ist die zweite wichtige Aufgabe eines Relaxationsmodells die Voraussage der Grenzflächenversetzungsdichte in Abhängigkeit von der Schichtdicke.

Die Messung der experimentell vorliegenden Versetzungsdichten ist einer der empfindlichsten Tests eines Relaxationsmodells, weil deren Dichte das Produkt des Relaxationsprozesses ist, der selbst aus den vier Teilprozessen der Erzeugung, Multiplikation, Bewegung und Orientierung von Versetzungen besteht (Abschnitt 2.7).

Eine mögliche Abweichung vom vorausgesagten Wert der Versetzungsdichte läßt unter Berücksichtigung der beobachteten Versetzungskonfigurationen, die in den vorigen Abschnitten dargestellt wurden, Rückschlüsse auf die einzelnen Teilprozesse der Relaxation zu.

Die Messung der Versetzungsdichten erfolgte jeweils durch Abzählen der diskreten Versetzungslinien über Längenskalen von einigen Mikrometern und anschließendes Teilen durch die Meßdistanz. Die Meßwerte für die verschiedenen Übergitter der untersuchten Probenserie wurden an verschiedenen Probenstellen, entlang der beiden ⟨110⟩-Richtungen, sowohl aus Abbildungen von Rückdünnungs- als auch von Querschnittproben ermittelt.

Netzwerke der Übergitterperioden

Aus einer Übersicht der Abbildungen von Rückdünnungs- und Querschnittproben ist schon auf den ersten Blick zu sehen, daß die Dichten der Fehlpassungssegmente an den Grenzflächen zwischen den Einzelschichten der Übergitterperioden im Vergleich zu den Dichten der Segmente zwischen Übergittern und Substrat klein sind. Abbildung 4.24 zeigt die dichten Fehlpassungsnetzwerke an der Substratgrenzfläche eines (18/6)-Übergitters und eines (18/2)-Übergitters. Aus den Querschnittaufnahmen der Abbildungen 4.25 und 4.26 ist zu sehen, daß Fehlpassungsversetzungen an den Übergittergrenzflächen vergleichsweise selten auftreten.

Berücksichtigt man die relativ große Zahl von Einzelschichten, die die Übergitter bei einer Gesamtdicke von $h_{SL} = 800$ nm enthalten, so kommt man zum Ergebnis, daß die Verstzungsdichte an einer einzelnen Grenzfläche einer Periode verschwindend gering im Vergleich zu der an der Substratgrenzfläche ist. Es müßten zur Gewinnung eines verläßlichen Meßwertes dieser kleinen Dichte Meßdistanzen ausgewertet werden, die das Gesichtsfeld auch beim Aneinanderreihen benachbarter Abbildungen überschreiten. Es läßt sich weiterhin keine signifikant unterschiedliche Versetzungsdichte pro Grenzfläche in den Übergittern der hier untersuchten Proben feststellen.

Die Versetzungsdichten pro Grenzfläche sind so klein, daß sie für die quantitativen Betrachtungen der Relaxationsmodelle zu vernachlässigen sind, das heißt die Annahme von vollständig gegeneinander verspannten Einzelschichten ist für jeweils benachbarte Schichten eine vernünftige Vereinfachung.

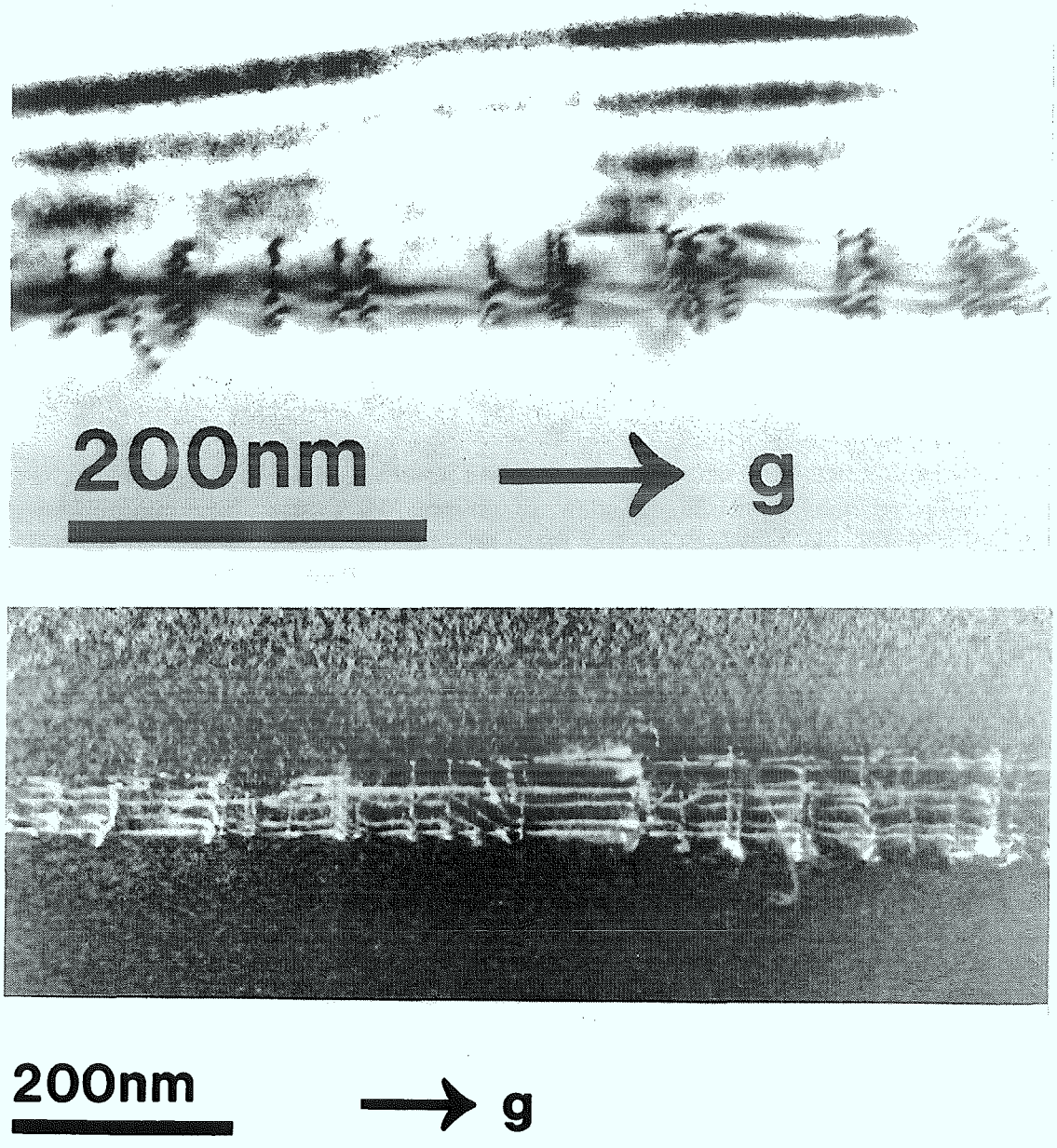


Abbildung 4.24: Seitenansicht von Fehlpassungsnetzwerken an den Substratgrenzflächen. Hellfeldabbildung, $\vec{g} = (220)$, eines (18/6)-Übergitters (oben), Weak-Beam-Dunkelfeldabbildung, $\vec{g} = (220)$, $(\vec{g}/4\vec{g})$, eines (18/2)-Übergitters (unten).

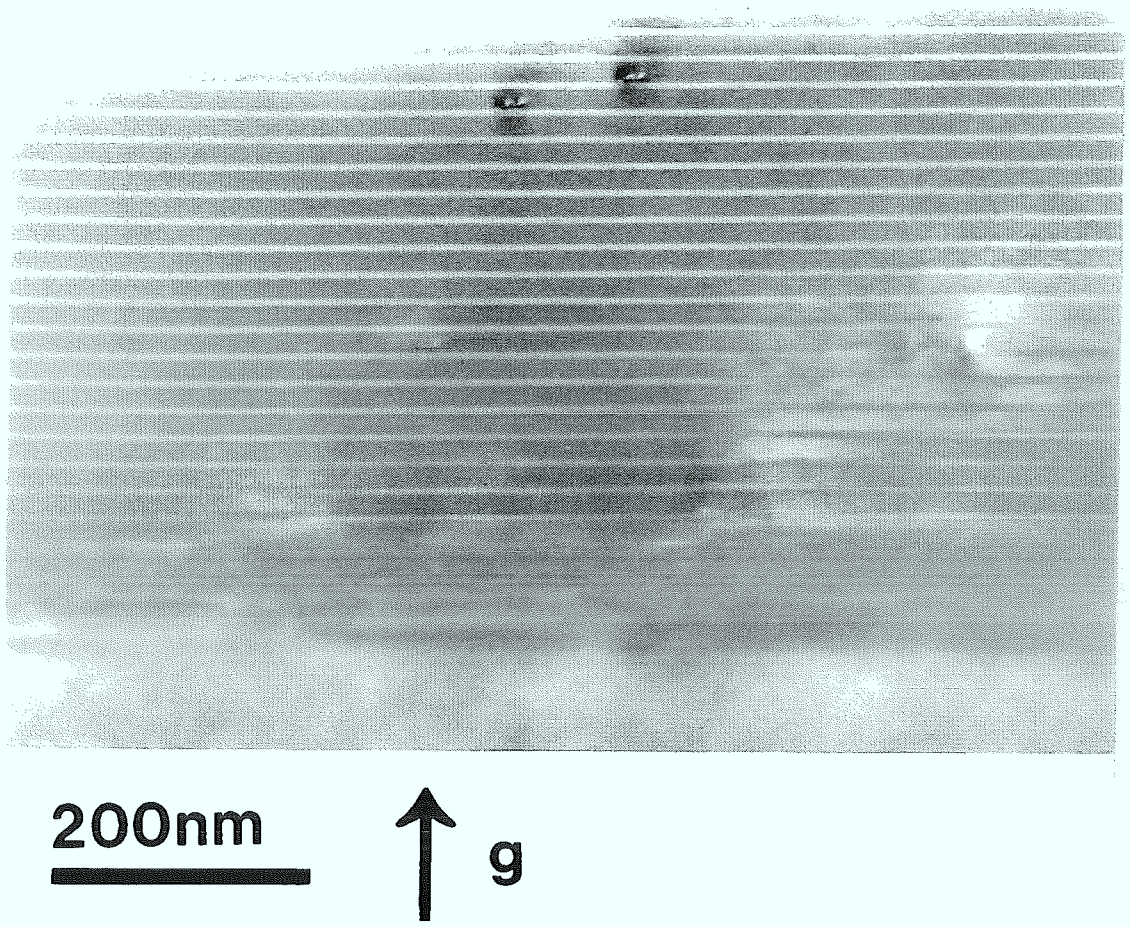


Abbildung 4.25: Seitenansicht eines (18/6)-Übergitters, $\vec{g} = (220)$. Die Übergitterperioden enthalten nur sehr wenige Fehlpassungsversetzungen, und sie sind über weite Distanzen praktisch defektfrei.

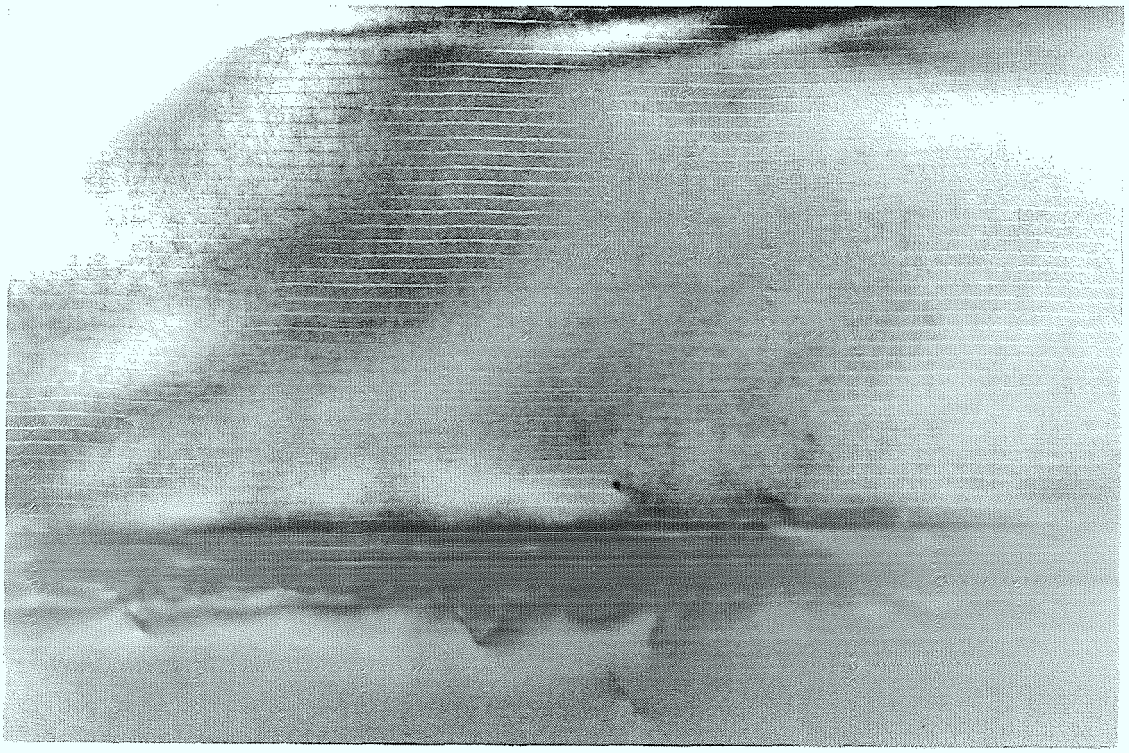


Abbildung 4.26: Seitenansicht eines (18/2)-Übergitters nahe einer [110]-Zonenachse. Die Übergitterperioden enthalten nur sehr wenige Fehlpassungsversetzungen, und sie sind über weite Distanzen praktisch defektfrei.

Es ist jedoch zu beachten, daß jede einzelne dieser Versetzungen zur Relaxation der Schichten beiträgt, die zwischen ihr und der Kristalloberfläche liegen. Je nachdem, ob sie in einer der unteren Übergitterperioden oder in einer der obersten Übergitterperioden liegt, trägt sie mehr oder weniger zur Relaxation des gesamten Übergitters bei.

Es hängt von der Verteilung der Versetzungen über alle Grenzflächen ab, wie weit die Summe aller Versetzungen zur Relaxation des gesamten Übergitters beiträgt. Wenn man in erster Näherung eine gleichmäßige Verteilung annimmt, so sind dann etwa die Hälfte der Versetzungen bei der Bestimmung der Versetzungsdichte der Substratgrenzfläche hinzuzuzählen.

Diese Korrektur kann jedoch effektiv entfallen, denn die Übergitter enthalten ja nicht nur Versetzungen, die einen zur Relaxation gegenüber dem Substrat kompatiblen Burgersvektor haben, sondern auch eine ganze Anzahl von Dipolkonfigurationen (Abschnitt 4.3), die aus Versetzungssegmenten mit entgegengesetzten Burgersvektoren bestehen. Sie relaxieren nur die betreffende Einzelschicht und tragen wegen ihres schnell abfallenden Verspannungsfeldes [83] nicht zur Relaxation des gesamten Übergitters bei.

Praktisch kann aus den genannten Gründen bei einem Verhältnis der Summe aller Versetzungen in den einzelnen Perioden zu der Summe aller Versetzungen in der Substratgrenzfläche von < 10 % der Einfluß der Versetzungen der einzelnen Perioden auf die Relaxation des gesamten Übergitters vernachlässigt werden, weil der effektiv anzurechnende Anteil, der dann bei etwa einem Viertel von 10 % läge, schon der Ungenauigkeit der Längenmessung bei der Bestimmung der Versetzungsdichte entspricht.

Netzwerke an den Substratgrenzflächen

Mit Hilfe der oben dargestellten Abzählmethode wurden für die Übergitter der untersuchten Probenserie jeweils die mittleren Versetzungsabstände S an den Grenzflächen zwischen den Übergittern und dem Substrat ausgewertet. Die Meßwerte sind zusammen mit den für die Interpretation der Daten durch Relaxationsmodelle wichtigen Größen der Fehlpassung f_{SL} der Übergitter, des relaxierten Anteils der Gitterfehlpassung b_{eff}/S, der homogenen Restdehnung ϵ_{SL}, und des Relaxationsgrades

$$\frac{b_{eff}/S}{f_{SL}}, \tag{4.4}$$

in der Tabelle 4.1 zusammengefaßt. Die bisher noch nicht eingeführte Größe des Relaxationsgrades ist ein auf die Fehlpassung normiertes Maß, das bei vollständiger Verspannung eines Schichtsystems den Wert Null hat, und das bei vollständiger Relaxation den Wert Eins annimmt. In der letzten Spalte ist für jedes Übergitter das Produkt aus der GaAs-Schichtdicke und dem relaxierten Anteil der Gitterfehlpassung eingetragen, das in der Diskussion zur Versetzungskinetik in Übergittern (Kapitel 5) eine wichtige Rolle spielen wird.

Die Meßpunkte des relaxierten Anteils der Gitterfehlpassung und die der homogenen Restdehnung sind jeweils zusammen mit den Werten der mittleren Fehlpassung

GaAs h [nm]	InGaAs h [nm]	f_{SL} [%]	S [nm]	$\frac{b_{eff}}{S}$ [%]	ϵ_{SL} [%]	$\frac{b_{eff}/S}{f_{SL}}$	$h_{GaAs}\frac{b_{eff}}{S}$ [nm]
2	18	1.260	32.5	0.615	0.645	0.49	1.23×10^{-2}
6	18	1.050	45.6	0.439	0.611	0.42	2.63×10^{-2}
10	18	0.900	93.9	0.213	0.687	0.24	2.13×10^{-2}
16	18	0.741	144.9	0.138	0.603	0.19	2.21×10^{-2}
25	18	0.586	210.5	0.095	0.491	0.16	2.38×10^{-2}

Tabelle 4.1: Abhängigkeit des mittleren Versetzungsabstands S von der Schichtdicke h_{GaAs} der GaAs-Einzelschichten. Es sind außerdem die Schichtdicke der InGaAs-Einzelschichten, die Fehlpassung f_{SL}, der relaxierte Anteil der Fehlpassung b_{eff}/S, die homogene Restdehnung ϵ_{SL}, der Relaxationsgrad $\frac{b_{eff}/S}{f_{SL}}$ und das Produkt aus der GaAs-Schichtdicke und dem relaxierten Anteil der Fehlpassung angegeben.

der Übergitter in Abhängigkeit von der GaAs-Schichtdicke in den Diagrammen der Abbildungen 4.27 und 4.28 dargestellt.

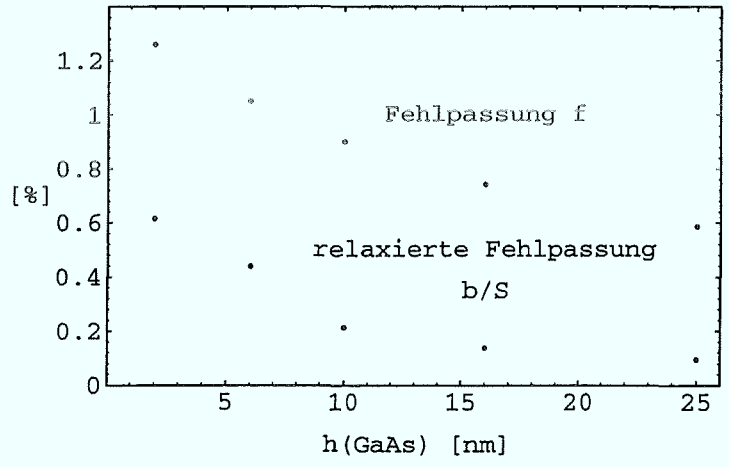


Abbildung 4.27: Abhängigkeit des relaxierten Anteils der Fehlpassung b_{eff}/S der Übergitter von der GaAs-Einzelschichtdicke h_{GaAs}. Zusätzlich ist die Fehlpassung f_{SL} der entsprechenden Übergitter eingetragen.

Die quantitative Analyse der Versetzungsdichten und der daraus abgeleiteten Größen kann zu folgenden Aussagen zusammengefaßt werden:

- Der mittlere Versetzungsabstand S nimmt mit steigender GaAs-Schichtdicke h_{GaAs} stetig zu. Generell ist eine Zunahme des Versetzungsabstandes nach den Gleichgewichtsrelaxationsmodellen zu erwarten, weil die mittlere Fehlpassung der Übergitter relativ zum Substrat mit steigender GaAs-Schichtdicke im un-

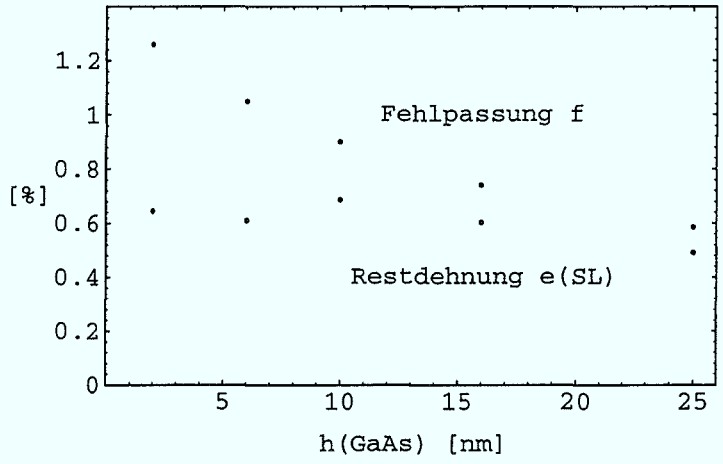


Abbildung 4.28: Abhängigkeit der homogenen Restdehnung ϵ_{SL} der Übergitter von der GaAs-Einzelschichtdicke h_{GaAs}. Zusätzlich ist die Fehlpassung f_{SL} der entsprechenden Übergitter eingetragen.

tersuchten Schichtsystem nach der Relation

$$f_{SL} = \frac{h_{InGaAs}}{h_{GaAs} + h_{InGaAs}} f_L \tag{4.5}$$

abnimmt, mit der Gitterfehlpassung $f_L = 1.4$ % zwischen den Einzelschichten.

- Der relaxierte Anteil der Gitterfehlpassung b_{eff}/S nimmt vom Wert 0.615 % für die dünnste GaAs-Schicht mit $h_{GaAs} = 2$ nm auf den Wert 0.095 % für die dickere Schicht mit $h_{GaAs} = 25$ nm ab. Dies steht in Widerspruch zu den Voraussagen der Gleichgewichtsmodelle, weil die Übergitter stark überkritisch sind, und deshalb bei fast vollständiger Relaxation

$$\frac{b_{eff}}{S} \approx f_{SL} \tag{4.6}$$

gelten sollte.

- Die nur partielle Relaxation der Übergitter wird durch den Wert des Relaxationsgrades belegt, der anzeigt, daß ein Übergitter bei der dünnsten GaAs-Schicht mit $h_{GaAs} = 2$ nm nur etwa zur Hälfte relaxiert ist, während ein Übergitter bei einer dickeren GaAs-Schicht mit $h_{GaAs} = 25$ nm nur zu etwa einem Achtel relaxiert ist.

- Die Werte der homogenen Restdehnung ϵ_{SL} schwanken für kleine GaAs-Schichtdicken, $h_{GaAs} \leq 10$ nm, zwischen 0.611 % und 0.687 %. Mit zunehmender GaAs-Schichtdicke nimmt die Restdehnung stetig bis auf den Wert 0.491 %

für $h_{GaAs} = 25$ nm ab. Alle gemessenen Werte der Restdehnung stehen in Widerspruch zu den Voraussagen der Gleichgewichtsrelaxationsmodelle, weil die stark überkritischen Übergitter eine Restdehnung von $\epsilon_{SL} \approx 0$ haben sollten.

- Mit Ausnahme der kleinsten GaAs-Schichtdicke $h_{GaAs} = 2$ nm ist in allen Übergittern das Produkt aus der GaAs-Schichtdicke und aus dem relaxierten Anteil der Gitterfehlpassung fast konstant, mit einem Mittelwert von $2.34 \pm 0.22 \times 10^{-2}$ nm.

- Die Dichten paralleler Fehlpassungsversetzungen sind in einem Netzwerk entlang der [110]- und der [$\bar{1}$10]-Richtungen im Rahmen der Meßgenauigkeit gleich.

Eine weitergehende Interpretation der Meßergebnisse nimmt einen relativ breiten Raum ein, deshalb ist der Diskussion der partiellen Relaxation der untersuchten Übergitter das anschließende Kapitel gewidmet.

4.10 Zusammenfassung der experimentellen Beobachtungen

Alle in dieser Arbeit untersuchten fehlgepaßten InGaAs/GaAs-Übergitter sind gegenüber dem GaAs-Substrat relaxiert. Die dabei an der Substratgrenzfläche erzeugten Versetzungsnetzwerke bilden zwei orthogonale Scharen von Fehlpassungssegmenten aus, mit Linienrichtungen $\vec{u} = [110], [\bar{1}10]$, und sie bestehen fast ausschließlich aus dem bei Systemen niedriger Fehlpassung vorherrschenden, gleitfähigen 60°-Versetzungstyp [33, 35, 36, 118].

Es wird die Bildung von Versetzungshalbschleifen durch Abgleiten einzelner Fehlpassungssegmente in das Substrat beobachtet, deren Ursache eine kleinere Stabilität von Versetzungsnetzwerken bei niedrigen Fehlpassungen ist. Sie entsteht durch die mögliche repulsive Wechselwirkung zweier, sich überkreuzender Versetzungslinien mit parallelen Burgersvektoren, und durch mögliche lokale Verspannungen des Substrates.

Die GaAs- und InGaAs-Einzelschichten der Übergitterperioden sind praktisch nicht gegeneinander relaxiert. Es wird eine sehr geringe Versetzungsdichte an den einzelnen Grenzflächen beobachtet, die jedoch einen vernachlässigbaren Beitrag zur Relaxation des gesamten Übergitters liefert.

Die in den Übergitterperioden auftretenden einzelnen Versetzungskonfigurationen, wie Versetzungsdipole, Haarnadelkonfigurationen, Superkinken, Fadenversetzungen und rechtwinklig abknickende Fehlpassungssegmente sind Produkte der während der Relaxation ablaufenden kinetischen Prozesse.

Insbesondere die Existenz von Superkinken ist aufschlußreich, weil sie erstens zeigt, daß während der Relaxation neue Versetzungen erzeugt werden, die von der Kristalloberfläche zur Substratgrenzfläche gleiten, und weil sie zweitens einen Hinweis darauf gibt, daß einzelne Übergitterperioden als Hindernis während des Versetzungsgleitens wirken können.

Das beobachtete rechtwinklige Abknicken von Fehlpassungssegmenten bewirkt einen Ausgleich einer möglichen asymmetrischen Relaxation entlang der [110]- und der [$\bar{1}$10]-Richtung. Das Auftreten mehrerer solcher Konfigurationen, die exakt entlang einer gemeinsamen Verbindungslinie angeordnet sind, legt nahe, daß während der Relaxation ein zum Hagen-Strunk-Mechanismus analoger Prozeß abläuft, der Fehlpassungssegmente eliminiert.

Das wichtigste experimentelle Resultat ist die Feststellung einer nur partiellen Relaxation der Übergitter: Obwohl die Gesamtdicken der Übergitter in allen Fällen weit über die kritischen Schichtdicken hinausgehen, sind die Übergitter bei der größten hier untersuchten mittleren Gitterfehlpassung von 1.26 % etwa nur zur Hälfte, bei der kleinsten untersuchten Fehlpassung von 0.59 % nur zu etwa einem Achtel relaxiert.

Die experimentellen Beobachtungen legen den Schluß nahe, daß mindestens einer der Teilprozesse während der Übergitterrelaxation im Hinblick auf die Gleichgewichtskonfiguration eines vollständig relaxierten Übergitters nur unvollständig abläuft. Das nächste Kapitel ist der Frage gewidmet, in welchem Ausmaß die Erzeugung, die Multiplikation oder das Gleiten von Versetzungen zur Substratgrenzfläche, unter Berücksichtigung der in diesem Kapitel dargestellten Versetzungskonfigurationen behindert sein können.

Kapitel 5

Partielle Relaxation von Übergittern

Die wichtigste der experimentellen Beobachtungen, die im vorangegangenen Kapitel beschrieben wurden, ist die Feststellung, daß alle untersuchten Übergitter *partiell* relaxiert sind, obwohl sie abhängig vom Verhältnis der Einzelschichtdicken um ein bis zwei Größenordnungen dicker als die zugehörige kritische Schichtdicke sind (Abschnitt 3.1).

Das folgende Kapitel wird den Zusammenhang der Beobachtungen in Hinblick auf die mögliche Hemmung der Relaxation während der Erzeugung, der Multiplikation und des Gleitens von Versetzungen zur Substratgrenzfläche herstellen. Dabei werden teilweise die Ergebnisse der kurzen Diskussionen benutzt, die im vorangehenden Kapitel jeder experimentellen Beobachtung gewidmet waren, und die die grundsätzliche Funktion und Bedeutung zum Beispiel einer bestimmten Versetzungskonfiguration erläutert haben.

Darüber hinaus werden zwei zusätzliche Sachverhalte diskutiert (Abschnitte 5.5, 5.6), die einzigartig für den Fall der Übergitterrelaxation sind, und die bisher in keinem einzigen kinetischen Relaxationsmodell berücksichtigt sind.

5.1 Bildkräfte bei der Versetzungserzeugung

Aus der Analyse des kritischen Schleifenradius r_c bei der homogenen und heterogenen Halbschleifenerzeugung und der dazu korrespondierenden Aktivierungsenergie E_c ist bekannt, daß die Versetzungserzeugung behindert sein kann, wenn die Temperatur während des epitaktischen Wachstums nicht hoch genug ist, um die Aktivierungsenergie durch thermische Fluktuationen der freien Energie aufzubringen [127, 121]. Beispielsweise beträgt der kritische Radius einer $In_{0.2}Ga_{0.8}As$-Schicht, die um $f = 1.4$ % komprimiert ist, 5.1 nm, mit einer zugehörigen Aktivierungsenergie von 22 eV.

Für die in dieser Arbeit untersuchten Übergitter wurde untersucht, wie weit ein Versetzungssegment in eine verspannte Schicht eindringen muß, damit die Summe aller Kräfte pro Längeneinheit auf die Versetzungslinie ein Resultierende bildet, die

das Gleiten zur Grenzfläche gerade noch ermöglicht.

Zu diesem Zweck wurde angenommen, daß sich die Fadensegmente einer Versetzungsschleife schon so weit voneinander entfernt haben, daß die gegenseitige Wechselwirkung zwischen beiden vernachlässigt werden kann. Dann ist die einzige repulsive Kraft auf das praktisch parallel zur Kristalloberfläche verlaufende Fehlpassungssegment die Bildkraft ihres Spiegelbildes, das die Randbedingung der freien Oberfläche aufrecht erhält (Abschnitt 2.4.2).

Im Rahmen der linearen isotropen Elastizitätstheorie ist der auf die Gleitebene projizierte Betrag der Bildkraft durch

$$F = \frac{G \sin\alpha}{4\pi d\,(1-\nu)} \left(b_e^2 + (1-\nu)\, b_s^2\right) \tag{5.1}$$

gegeben [83], mit b_e und b_s der Stufen- und Schraubenkomponente des Burgersvektors, α dem Winkel zwischen der Kristalloberfläche und der Gleitebene und d dem Abstand der Versetzungslinie von der Kristalloberfläche.

Die minimale Eindringtiefe $d_{\min}$ der Versetzungslinie, die überwunden werden muß, damit die Versetzung unter der Wirkung der äußeren Spannung der aufgewachsenen Schicht zur Substratgrenzfläche weitergleitet, erhält man dann durch Gleichsetzen der Bildkraft und der Peach-Koehler-Kraft, die proportional zur äußeren Spannung τ der Einzelschicht ist (Abschnitt 2.68).

Die Übergitter besitzen nach der partiellen Relaxation eine Restdehnung ϵ_{SL}, die in den untersuchten Fällen im Intervall von 0.5 % bis 0.7 % liegt. Die Umrechnung der Restdehnung in die Restspannung ergibt mit Gleichung (5.1) eine minimale Eindringtiefe von 3.5 nm bis 4.0 nm. Die zugehörige Spannung in der Gleitebene beträgt 240 MPa.

Die berechnete Eindringtiefe hat ungefähr die gleiche Größe, wie der oben berechnete kritische Schleifenradius r_c, so daß die Abschätzung mittels der Restverspannung und der Bildkraft auf das Fehlpassungssegment gültig ist. Der Wert der Eindringtiefe hat eine plausible Größenordnung, die unterhalb der Einzelschichtdicke der fehlgepaßten Einzelschicht liegt. Die hohe Aktivierungsenergie der Halbschleifenerzeugung macht es jedoch unwahrscheinlich, daß die stabile Konfiguration bei der oben berechneten Eindringtiefe auch tatsächlich erreicht wird. In Abschnitt 2.7.1 wurden bereits experimentelle Bedingungen diskutiert, die dennoch eine Halbschleifenerzeugung und das Erreichen einer genügend großen Eindringtiefe ermöglichen.

Es gibt jedoch keinen experimentellen Befund, der belegt oder widerlegt, daß die berechnete Eindringtiefe während der Halbschleifenerzeugung auch tatsächlich erreicht wird. Das zum kritischen Schleifenradius gehörige Gleichgewicht ist offensichtlich so instabil, daß im Gegensatz zu anderen Versetzungskonfigurationen, die mit der Relaxationskinetik zusammenhängen, keine einzige kritische Halbschleife nach Abschluß des epitaktischen Schichtwachstums „eingefroren“ wurde. Die Halbschleifenerzeugung kann also im vorliegenden System als limitierender Faktor des Relaxationsprozesses nicht mit Sicherheit ausgeschlossen werden.

5.2 Bildkräfte bei der Multiplikation

Nach dem im vorangehenden Abschnitt die Erzeugung von Halbschleifen und deren Wachstum, das ja strenggenommen auch eine Multiplikation der total verfügbaren Versetzungslänge ist, abgehandelt wurde, ist in diesem Abschnitt lediglich noch die Hagen-Strunk-Multiplikation zu diskutieren.

Die Teilprozesse des Mechanismus sind in Abschnitt 2.7.2 dargestellt. Es ist an dieser Stelle mit Bezug auf die in dieser Arbeit untersuchten Übergitter festzustellen, daß die neben der Wechselwirkungskraft zweier sich überkreuzender Fehlpassungssegmente wichtige treibende Kraft die Bildkraft auf die sich vom Versetzungsknoten ablösenden Segmente ist.

Diese Bildkraft fällt mit dem reziproken Abstand von der Kristalloberfläche ab [83], und bei einer Gesamtdicke der Übergitter von 800 nm ist sie an den Orten der Substratgrenzflächen der Übergitter zu vernachlässigen. Konkret ist sie an der Substratgrenzfläche um einen Faktor 4 nm/800 nm $= 5 \times 10^{-3}$ kleiner, als bei der im vorangehenden Abschnitt diskutierten Halbschleifenerzeugung.

Der Hagen-Strunk-Mechanismus ist bei großen Schichtdicken nicht aktiv, so daß hier einzig die Halbschleifenerzeugung als Versetzungsquelle und Multiplikationsmechanismus übrig bleibt.

5.3 Elastische Verspannung des Substrats

Es ist bekannt, daß epitaktische Systeme, die in einem dreidimensionalen Wachstumsmodus aufwachsen, ihre Fehlpassungsspannung nicht allein durch den Einbau von Versetzungen an der Substratgrenzfläche relaxieren [39, 21]. Das in solchen Systemen auftretende Inselwachstum ist zusätzlich durch verspannte Substratbereiche gekennzeichnet, die sich direkt unterhalb der Inseln befinden.

Weil das Vorzeichen der Spannungskomponenten in den Substratbereichen genau entgegengesetzt zur Fehlpassungsspannung des aufgewachsenen Materials ist, wird dadurch die Fehlpassung effektiv verringert und das entsprechende Netzwerk von Fehlpassungsversetzungen ist im mechanischen Gleichgewicht weniger dicht. Das Verzerrungsfeld fällt auf einer Längenskala, die in etwa einem Inseldurchmesser entspricht, auf Null ab.

In perfekt zweidimensionalen Systemen ist eine solche Substratverspannung jedoch grundsätzlich vernachlässigbar, weil die entsprechende Längenskala von zum Beispiel einem Waferdurchmesser sehr viel größer als ein typischer Inseldurchmesser ist, und eine solche makroskopische Verspannung des Substrates energetisch ungünstig ist.

Die im Substrat unterhalb der Grenzfläche der ersten Übergitterperiode beobachteten Versetzungshalbschleifen und aufgespaltenen Fehlpassungssegmente können grundsätzlich ihre Ursache jedoch in einer *lokalen* Substratverspannung haben. Dazu muß man allerdings annehmen, daß die Relaxation an unterschiedlichen Stellen der Substratgrenzfläche verschieden dichte Netzwerke erzeugt. Experimentell wird jedoch keine ausreichend starke Inhomogenität der Versetzungsdichten in der Nähe

der Versetzungshalbschleifen beobachtet.

Im Vergleich zu der im vorigen Kapitel diskutierten Instabilität von Fehlpassungsnetzwerken in niedrig fehlgepaßten Systemen mit gleitfähigen 60°-Versetzungen (Abschnitt 4.7.2) ist das Vorliegen einer globalen Substratverspannung in einer Größenordnung der beobachteten Restdehnung der Übergitter somit wenig wahrscheinlich.

5.4 Wechselwirkung Versetzung–Netzwerk

In Abschnitt 2.5 wurde der Gleichgewichtszustand einer fehlgepaßten Einzelschicht abgeleitet, wobei für den Fall eines dichten Netzwerkes mit einem Versetzungsabstand $S < 2h$ die Gesamtenergie in die Energie der homogenen Schichtdeformation und in die Energie des Fehlpassungsnetzwerkes aufgespalten wurde. Bei dieser Aufspaltung wird eine schwache periodische Modulation des Spannungsfeldes vernachlässigt, weil deren Amplitude stark mit dem Abstand von der Substratgrenzfläche abfällt.

Im Rahmen dieser Arbeit wurde überprüft, ob diese Näherung auch für die Schichtdicken und die Fehlpassungen der untersuchten Übergitter gültig ist, und wie die Spannungsverteilung in der Nähe eines schon existierenden Netzwerkes auf eine neu erzeugte Versetzung wirkt, die von der Kristalloberfläche zur Substratgrenzfläche gleitet.

Zu diesem Zweck wurde die Spannungsverteilung eines regelmäßigen Fehlpassungsnetzwerks aus Stufenversetzungen berechnet, deren Burgersvektoren in der Grenzfläche zwischen Substrat und aufgewachsener Schicht liegen. Die Näherung eines Fehlpassungsnetzwerks aus 60°-Versetzungen durch das genannte Netzwerk ist gerechtfertigt, weil in realen Netzwerken die Summe aller Schraubenkomponenten und Stufenkomponenten senkrecht zur Grenzfläche im Mittel über das ganze Netzwerk verschwindet. Andernfalls würden nämlich die aufwachsenden Einzelschichten einer energetisch ungünstigen makroskopischen Kristallkippung oder einer energetisch ungünstigen Scherung unterworfen. Für jeden Punkt einer Zelle mit der Breite des Versetzungsabstandes S und mit der Höhe gleich der Schichtdicke h wurden die Spannungskomponenten der Grenzflächenversetzungen numerisch aufaddiert. Das Ergebnis für die Gesamtspannungen σ_{xx}, σ_{yy} wurde auf drei gültige Dezimalen genau berechnet, so daß die Auswertung der zugehörigen unendlichen Reihen nach einer endlichen Zahl von Summanden abgebrochen werden konnte.

Die Gleitebene der neu erzeugten 60°-Versetzung, deren Linie parallel zur Grenzfläche angenommen wurde, ist zur Grenzfläche geneigt, so daß nach der Peach-Koehler-Gleichung die auf die Versetzungslinie wirkende Kraft von beiden Spannungskomponenten σ_{xx}, σ_{yy} abhängt (Abschnitt 2.68). Der resultierende Betrag der Kraft wurde in eine zweidimensionale Karte eingetragen, die in Abbildung 5.1 gezeigt ist.

Die berechnete Kraftverteilung kann in drei Anteile aufgespalten werden. Der erste ist ein konstanter Anteil, der proportional zur relaxierten Dehnung b_{eff}/S ist, und der zusammen mit der Gitterfehlpassung die homogene Restdehnung ϵ der auf-

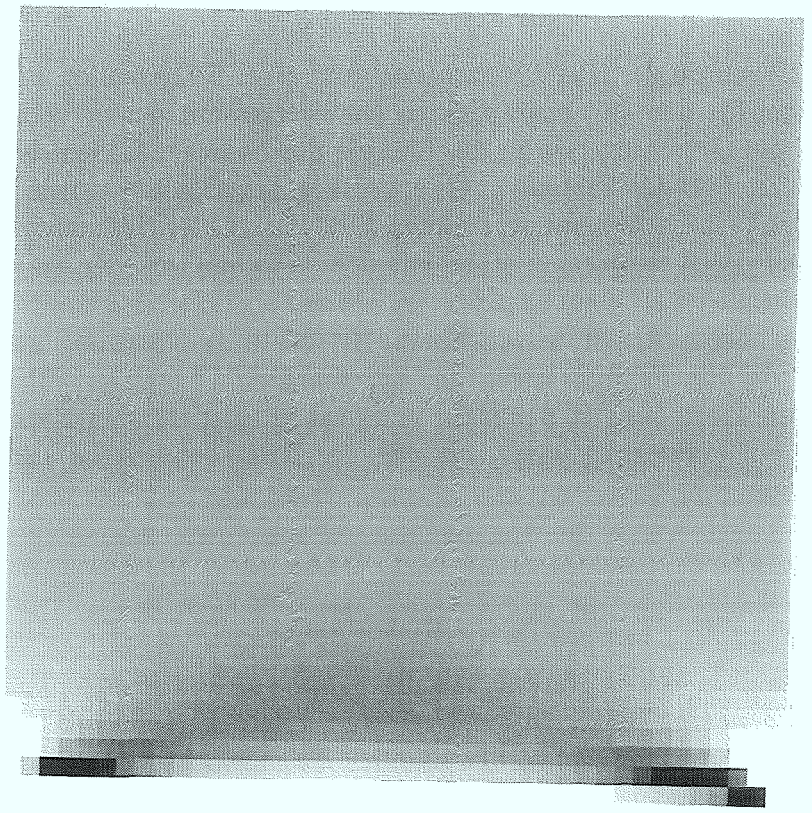


Abbildung 5.1: Peach-Koehler-Kraft auf eine gleitende 60°-Versetzung in Abhängigkeit von der Position relativ zu einem dichten Fehlpassungsnetzwerk aus Stufenversetzungen. Unten rechts und unten links befindet sich jeweils eine Fehlpassungsversetzung. Der Ausschnitt ist periodisch nach rechts und links fortzusetzen, das Verhältnis von Schichtdicke zu Versetzungsabstand beträgt 3.

gewachsenen Schicht repräsentiert. Der zweite Anteil ist das $1/r$-Nahfeld einer einzelnen Grenzflächenversetzung, das zusammen mit den Nahfeldern der übrigen Versetzungen das Spannungsfeld des Netzwerkes ergibt, welches in der in Kapitel 2 dargestellten Form in Relaxationsmodellen enthalten ist.

Der dritte Anteil, der in Relaxationsmodellen gewöhnlich vernachlässigt wird, hat die Gestalt eines Bandes, das sich von einer Versetzung zur nächsten erstreckt, und das sich in der Mitte zwischen zwei Fehlpassungsversetzungen von der Grenzfläche weg zur Kristalloberfläche hin auswölbt.

Wenn man annimmt, daß das Versetzungsnetzwerk zusammen mit der neu erzeugten Versetzung ein möglichst regelmäßiges Netzwerk bilden soll, dann ist es plausibel, daß die entsprechende Gleitebene die Grenzfläche etwa in der Mitte zwischen zwei Versetzungen schneidet.

Durch Auswerten der Karte der Kraftverteilung sieht man, daß der homogene Anteil der Kraft vor und hinter dem eben beschriebenen Band fast gleich ist, und daß das Band selbst eine Barriere für das Versetzungsgleiten bildet. Es gibt jedoch zwei Gründe, weshalb diese Barriere im Fall der hier untersuchten Übergitter nicht der begrenzende Mechanismus der Relaxation ist. Erstens beträgt die Barrierenhöhe nur etwa 10 % des Anteils der homogenen Kraft auf die Versetzungslinie, das heißt die Barriere ist nicht hoch genug, um die Übergitterrelaxation bis auf die Hälfte oder bis auf ein Achtel zu reduzieren. Zweitens würden gleitende Versetzungen etwa am Ort der Barriere gestoppt, so daß sich eine Reihe von Versetzungen in einem charakteristischen Abstand von der Grenzfläche ansammeln würden. Dies wird jedoch experimentell nicht beobachtet.

Die Analyse der Barrierenhöhe kann natürlich nur für Koordinaten sinnvoll ausgewertet werden, die nicht zu nahe an den aus der Karte abzulesenden Singularitäten der Spannungsfelder liegen. An diesen Orten ist die oben geschilderte Mittelung über die Schraubenkomponenten und die Stufenkomponenten senkrecht zur Grenzfläche nicht mehr gültig, weil dort das Spannungsfeld der einzelnen Versetzungen dominiert und die Peach-Koehler-Kraft konkret von der relativen Orientierung der Burgersvektoren abhängt. Das oben diskutierte „Band" in der Spannungsverteilung entsteht jedoch als Überlagerung *aller* Versetzungen des Netzwerks, und die zugeordnete relative Barrierenhöhe wird deshalb nicht drastisch von einer einzelnen Versetzungskonfiguration beeinflußt.

5.5 Einzelschicht-Barrieren

Aus den Überlegungen des letzten Abschnitt über die Höhe einer möglichen Barriere ist ein wichtiger Schluß zu ziehen: Eine Barriere, die das Gleiten von Versetzungen so massiv behindert, daß der Relaxationsgrad bis auf Achtel reduziert sein kann, muß ein lokales Spannnungsfeld besitzen, dessen Größe ein *bedeutender* Bruchteil der Fehlpassungsspannung ist. Unter diesem Gesichtspunkt wurde untersucht, welchen lokalen Kräften eine Versetzung ausgesetzt ist, die an der Kristalloberfläche erzeugt worden ist, und die durch das Übergitter zur Substratgrenzfläche gleitet. Hierbei wurde das Problem der Bildkräfte an der Kristalloberfläche ausgeklammert, das

schon in den Abschnitten 5.1, 5.2 diskutiert wurde.

Es wurden zwei Fälle untersucht, nämlich erstens das Einsetzen der Relaxation eines defektfreien Übergitters, dessen Gesamtdicke gerade größer als die kritische Gesamtdicke ist, und zweitens das Gleiten einer Versetzung in einem schon partiell relaxierten Übergitter.

5.5.1 Einsetzen der Relaxation

Der Verspannungszustand eines kohärenten Übergitters ist im vorliegenden InGaAs-GaAs-Materialsystem durch verspannte InGaAs-Schichten und unverspannte GaAs-Schichten gekennzeichnet. Die Grenzflächen der Übergitterperioden und die Substratgrenzfläche sind versetzungsfrei, die Dehnungen der InGaAs-Schichten betragen $\epsilon_x = \epsilon_y = f = -1.4$ %. Die GaAs-Schichten und das GaAs-Substrat haben die Dehnung Null.

Die erste Versetzung, die durch das Übergitter gleitet, erfährt eine Kraft pro Linienlänge, die durch die Peach-Koehler-Gleichung gegeben ist (Gleichung (2.68)). Die in dieser Gleichung auftretende Spannungskomponente σ_{lm} ist proportional zu den Dehnungskomponenten ϵ_x, ϵ_y.

Das Einsetzen der Burgersvektoren und Linienrichtungen der in diesem System vorliegenden Gleitsysteme ergibt eine in Richtung auf die Substratgrenzfläche wirkende Kraft, deren Betrag proportional zu b_{eff}, der zur Relaxation beitragenden Burgersvektorkomponente, und zu ϵ_x ist. Die Kraft ist nur ungleich Null in den InGaAs-Schichten, während sie in den nicht gedehnten GaAs-Schichten verschwindet.

Das Gleiten einer Versetzung ist also beim Einsetzen der Relaxation nicht behindert, und es wird durch die Verspannung der InGaAs-Schichten in Richtung auf die Substratgrenzfläche vorangetrieben.

5.5.2 Gleiten bei partieller Relaxation

Das partiell relaxierte Übergitter besitzt ein Netzwerk von Fehlpassungsversetzungen an der Substratgrenzfläche, mit einem mittleren Versetzungsabstand S, der bei symmetrischer Relaxation in x- und y-Richtung gleich groß ist. Wenn die Einzelschichten unterkritisch sind, dann sind die Übergitterperioden weiter versetzungsfrei, so daß die Relaxation des gesamten Übergitters zu einer Relaxation der InGaAs-Schichten um den Betrag b_{eff}/S geführt hat. Gleichzeitig sind jedoch auch alle Versetzungen des Netzwerks durch die GaAs-Schichten geglitten und haben diese um den gleichen Betrag b_{eff}/S verspannt.

Hier wird jetzt zur Vermeidung von möglicherweise verwirrenden Vorzeichenbetrachtungen wieder der Konvention gefolgt, daß die Fehlpassung der InGaAs-Schichten als positiv angenommen wird, mit entsprechenden Konsequenzen für die Vorzeichen der Dehnungen und der Größe von b_{eff}/S in den InGaAs- und GaAs-Schichten (Bemerkung in Abschnitt 2.5).

Die Dehnungen nehmen in den Einzelschichten die Werte

$$\epsilon_{\mathrm{InGaAs}} = f - \frac{b_{\mathrm{eff}}}{S}, \tag{5.2}$$

$$\epsilon_{\mathrm{GaAs}} = -\frac{b_{\mathrm{eff}}}{S} \tag{5.3}$$

an. Mit der Proportionalität von Spannungen und Dehnungen folgt aus der Peach-Koehler-Gleichung, analog zur Betrachtung des vorigen Abschnitts, daß auf eine durch das Übergitter gleitende Versetzung in den InGaAs-Schichten eine Kraft in Richtung der Substratgrenzfläche wirkt, während sie in den GaAs-Schichten in die entgegengesetzte Richtung, zur Kristalloberfläche hin wirkt.

Das Gleiten von Versetzungen wird also in den InGaAs-Schichten durch die Restdehnung $\epsilon_{\mathrm{InGaAs}}$ unterstützt, und in den GaAs-Schichten gehemmt. Natürlich ist die Hemmung im Frühstadium der Relaxation nicht so groß, daß diese vollkommen unterbunden wird, denn sonst wäre ja der zugehörige Relaxationsgrad nahe Null, was experimentell nicht beobachtet wird.

Deshalb sind jetzt zwei Fragen zu untersuchen, nämlich erstens, ab welchem Relaxationsgrad die Hemmung so groß ist, daß die weitere Relaxation effektiv unterbunden wird, und zweitens durch welchen Mechanismus die Barriere, die die GaAs-Schicht schon bei einem kleinen Relaxationsgrad bildet, überwunden werden kann.

Die Untersuchung der ersten Frage beinhaltet eine Berechnung der Aktivierungsenergie, die aufgewendet werden muß, um ein Segment der Fehlpassungsversetzung über die Barriere gleiten zu lassen. Der Vorgang ist analog zum Gleiten einer Versetzung über die Peierls-Nabarro-Barriere eines anisotropen Kristalls (Abschnitt 2.7.3).

Die zu verrichtende mechanische Arbeit ist gegeben durch die Peach-Koehler-Kraft auf die Versetzungslinie multipliziert mit der Länge des Weges, die die Versetzung während des Gleitprozesses entlang der z-Achse zurücklegt:

$$\Delta E = L \cdot \int_0^{h_{\mathrm{GaAs}}} f_z \, \mathrm{d}z, \tag{5.4}$$

mit L der Länge des Fehlpassungssegmentes und h_{GaAs} der Dicke der GaAs-Schicht. Die Komponenten f_x und f_y der Peach-Koehler-Kraft sind Null, weil der Burgersvektor und die Linienrichtung des Segmentes so orientiert sind, daß das entsprechende Produkt mit dem Spannungstensor verschwindet.

Für die folgende Diskussion ist es an dieser Stelle ausreichend, den funktionalen Zusammenhang zwischen ΔE und den Parametern h_{GaAs} und ϵ_{GaAs} abzuleiten, die die Barriere charakterisieren.

Dazu sind folgende Feststellungen zu machen:

- Die Peach-Koehler-Kraft ist proportional zur Spannung der GaAs-Schicht,
- die Spannung ist konstant über dem Volumen der GaAs-Schicht, und
- die Spannung ist proportional zur Dehnung ϵ_{GaAs}.

Mit Gleichung (5.4) folgt, daß die aufzuwendende mechanische Arbeit direkt proportional zur Länge des Fehlpassungssegmentes, zur GaAs-Schichtdicke und zur Dehnung ist:

$$\Delta E \propto L\, h_{\mathrm{GaAs}}\, \epsilon_{\mathrm{GaAs}}. \tag{5.5}$$

Eine Abschätzung der im Experiment auftretenden Aktivierungsenergie pro Längeneinheit kann mit

$$\frac{\Delta E}{L} = \tau b_{\mathrm{eff}} h_{\mathrm{GaAs}}, \tag{5.6}$$

$$\tau = 2G\frac{1+\nu}{1-\nu}\epsilon_{\mathrm{GaAs}} \tag{5.7}$$

und Einsetzen einer beobachteten Restdehnung von $\epsilon_{\mathrm{GaAs}} = 0.6$ % und den Materialkonstanten von GaAs gewonnen werden. Das Ergebnis ist 1.9 eV nm^{-1} für $h_{\mathrm{GaAs}} = 2$ nm und 24 eV nm^{-1} für $h_{\mathrm{GaAs}} = 25$ nm. Die folgenden Betrachtungen werden zunächst unter dem Vorbehalt durchgeführt, daß die von der GaAs-Einzelschichtdicke abhängenden, hohen Aktivierungsenergien thermisch aufgebracht werden müssen. Im nächsten Abschnitt wird ein möglicher mechanischer Energiebeitrag diskutiert, der es ermöglicht, einen Teil der Aktivierungsenergie pro Längeneinheit aufzubringen.

Die Aktivierungsenergie kann bei Abwesenheit anderer äußerer Spannungen auf die Versetzung nur durch thermische Fluktuationen der freien Energie aufgebracht werden, und die Wahrscheinlichkeit pro Zeiteinheit für das Überwinden der Barriere ist dann proportional zu [79]

$$\exp\left(-\frac{\Delta E}{kT}\right), \tag{5.8}$$

mit der Wachstumstemperatur T.

Während der Relaxation nimmt die Aktivierungsenergie mit dem Relaxationsgrad, der proportional zu $\epsilon_{\mathrm{GaAs}} = -b_{\mathrm{eff}}/S$ ist, wegen der Beziehung (5.5) linear zu, und das Gleiten von Versetzungen wird zu dem Zeitpunkt effektiv gehemmt, bei dem die Aktivierungsenergie nicht mehr durch thermische Fluktuationen der freien Energie bei der Temperatur T aufgebracht werden kann.

In allen untersuchten Übergitterproben sind die gleichen Gleitsysteme aktiviert, in allen Fällen bestehen die als Barriere wirkenden Einzelschichten aus dem gleichen Material, und alle Proben wurden bei der gleichen Wachstumstemperatur hergestellt. Dann ist es vernünftig anzunehmen, daß die kritische Aktivierungsenergie pro Längeneinheit der Versetzungslinie, bei der das Gleiten von Versetzungen effektiv gehemmt wird, in allen untersuchten Übergittern gleich groß ist, so daß mit Gleichung (5.5) folgt:

$$h_{\mathrm{GaAs}}\, \epsilon_{\mathrm{GaAs}} = \mathrm{const.} \tag{5.9}$$

Nach Ersetzen der Dehnung (Gleichung (5.3)) ergibt sich der Zusammenhang

$$\boxed{\frac{b_{\mathrm{eff}}}{S} \propto \frac{1}{h_{\mathrm{GaAs}}}.} \tag{5.10}$$

Wenn also das Gleiten über die GaAs-Einzelschicht-Barrieren der Prozeß ist, der die Relaxationskinetik bestimmt, dann sollte der Relaxationsgrad eine $1/h_{\mathrm{GaAs}}$-Abhängigkeit haben, und das Produkt aus dem relaxierten Anteil der Fehlpassung, b_{eff}/S, und der GaAs-Schichtdicke sollte für alle Übergitter konstant sein.

Dies ist genau das Verhalten, das im Rahmen der Meßgenauigkeit mit einer Ausnahme für alle GaAs-Schichtdicken experimentell beobachtet wird (Tabelle 4.1). Bei der kleinsten GaAs-Schichtdicke $h_{\mathrm{GaAs}} = 2$ nm ist offensichtlich ein anderer Mechanismus der limitierende Faktor für die Relaxation. Diese Ausnahme wird in einem anderen Zusammenhang in Abschnitt 5.6 diskutiert.

Der Wert der Proportionalitätskonstanten der Relation (5.10) wurde durch einen Kleinstquadrat-Fit an die Meßwerte bestimmt, unter Auslassung des Meßpunktes für $h_{\mathrm{GaAs}} = 2$ nm. Das Resultat ist:

$$\frac{b_{\mathrm{eff}}}{S} = 2.47 \times 10^{-2}\ \mathrm{nm}\ \frac{1}{h_{\mathrm{GaAs}}}. \tag{5.11}$$

Die Konstante enthält neben Materialkonstanten auch noch die Wachstumstemperatur und einen dimensionslosen Faktor, der durch den Mechanismus bestimmt ist, der das Gleiten über die Barriere ermöglicht. Deshalb ist es an dieser Stelle vom Standpunkt der Verformungsmechanik und vom Standpunkt der Thermodynamik aus schwierig zu beurteilen, ob die Konstante in einer realistischen Größenordnung liegt. Aus der obigen Abschätzung der Aktivierungsenergie pro Längeneinheit (Gleichung (5.6)) ist jedoch deutlich abzulesen, daß ein beträchtlicher Teil der Aktivierungsenergie durch andere, als thermische Energiebeiträge aufgebracht werden muß.

5.5.3 Kinetische Energie einer Versetzung

Die im letzten Abschnitt diskutierte Barriere, die die entgegengesetzt verspannten GaAs-Einzelschichten bezüglich des Versetzungsgleitens bilden, hat eine von der GaAs-Einzelschichtdicke abhängige Aktivierungsenergie pro Längeneinheit der Versetzungsline von bis zu 24 eV nm^{-1}, die bei den in dieser Arbeit untersuchten Übergittern nicht allein durch thermische Fluktuationen der freien Energie aufgebracht werden kann. Das heißt, eine Versetzung, die vor der Barriere an einer InGaAs/GaAs-Grenzfläche gestoppt wurde, kann die Barriere bei der Wachstumstemperatur von 460°C nicht überwinden.

Wie in Abschnitt 5.6 ausführlich dargestellt wird, hätte dies zur Folge, daß ein Übergitter nach dem Aufwachsen der ersten GaAs-Einzelschicht nicht weiter relaxieren kann, als es dem Relaxationsgrad der untersten InGaAs-Einzelschicht entspricht. Schon die unterste GaAs-Schicht würde, praktisch unabhängig von der Einzelschichtdicke, eine so effektive Barriere für das weitere Versetzungsgleiten darstellen, daß der Relaxationsgrad wegen der konstanten InGaAs-Schichtdicke für alle Übergitter der untersuchten Probenserie gleich sein müßte. Die experimentellen Meßwerte zeigen jedoch, daß der Relaxationsgrad sehr wohl von der GaAs-Einzelschichtdicke abhängt (Tabelle 4.1).

Diese auf den ersten Blick paradoxe Situation ist unter anderem das Ergebnis der *quasistatischen* Betrachtungsweise der klassischen Mechanik, die auch die Grundlage

des Teils der linearen Elastizitätstheorie ist, der die gesamte Versetzungsmechanik betrifft [109], also die Gestalt der Verzerrungsfelder von Versetzungen und die Wechselwirkung dieser Verzerrungsfelder mit äußeren Spannungsfeldern. Im Rahmen dieses Formalismus werden grundsätzlich dynamische Vorgänge, die mit der Bewegung von Versetzungen zusammenhängen, vernachlässigt.

Eine vollständige Behandlung des Problems erfordert das Einbeziehen der Dynamik des Verzerrungsfeldes einer sich bewegenden Versetzung. Im Rahmen der Versetzungsdynamik [83] ist einer geradlinigen Versetzung, die sich mit konstanter Geschwindigkeit v durch einen Festkörper bewegt, eine kinetische Energie pro Linienlänge zuzuordnen, die, analog zur klassischen Mechanik, eine effektive Masse m^* pro Längeneinheit der Versetzungslinie enthält:

$$E_{\text{kin}} = \frac{1}{2} m^* v^2. \tag{5.12}$$

Diese Beziehung gilt nur, solange die Versetzungsgeschwindigkeit klein gegenüber den Schallgeschwindigkeiten c_l, c_t der elastischen Kompressions- und Scherwellen ist:

$$c_\text{l} = \sqrt{\frac{E(1-\nu)}{\rho(1+\nu)(1-2\nu)}}, \tag{5.13}$$

$$c_\text{t} = \sqrt{\frac{E}{2\rho(1+\nu)}}, \tag{5.14}$$

mit dem Elastizitätsmodul E, der Poisson-Zahl ν und der Massendichte ρ des Festkörpers. Die effektive Masse pro Längeneinheit hat für eine Schraubenversetzung den Wert [83]

$$m^*_{\text{Schraube}} = \frac{\rho b^2}{4\pi} \ln \frac{R}{r_0}, \tag{5.15}$$

und für eine Stufenversetzung den Wert [83]

$$m^*_{\text{Stufe}} = \frac{\rho b^2}{4\pi} \left(1 + \left(\frac{c_\text{t}}{c_\text{l}}\right)^4\right) \ln \frac{R}{r_0}. \tag{5.16}$$

R und r_0 bezeichnen den oberen und unteren Abschneideradius des Verzerrungsfelds.

Die Gesamtenergie einer gemischten Versetzung mit Stufen- und Schraubenanteil ist dann gegeben durch die Summe ihrer „Ruheenergie“ und der kinetischen Energien der Schrauben- und Stufenkomponente:

$$E = E_0 + \frac{1}{2}(m^*_{\text{Schraube}} + m^*_{\text{Stufe}})v^2. \tag{5.17}$$

Die Ruheenergie E_0 ist die in Abschnitt 2.4 hergeleitete Größe der Energie pro Linienlänge einer ruhenden Versetzung.

Die Energiebetrachtung kann jetzt auf den Fall einer Versetzung angewendet werden, die sich, von der Fehlpassungspannung angetrieben, gleichmäßig durch eine InGaAs-Schicht bewegt, und die dann auf eine GaAs-Schicht trifft, die mit ihrer

entgegengesetzten Spannung die Bewegung hemmt. Die Rechnung des vorigen Abschnitts zeigte, daß die Aktivierungsenergie der GaAs-Einzelschichtbarriere zu groß ist, um alleine durch eine thermische Aktivierung des Gleitprozesses aufgebracht zu werden. Da innerhalb eines perfekten Übergitters keine Spannungskonzentration vorliegen kann, die es einer Versetzung lokal ermöglichen könnte, die Barriere zu überwinden, ist jetzt im Rahmen der Versetzungsdynamik die Frage zu beantworten, ob die kinetische Energie der Versetzung, die diese durch das Gleiten in der InGaAs-Schicht erhalten hat, ausreichend groß ist, um die Aktivierungsenergie aufzubringen.

Dazu ist es hilfreich, zunächst eine obere Grenze $\bar{E}_{\mathrm{kin}}$ der kinetischen Energie zu berechnen, die sich aus der Energiebilanz des Gleitpozesses in der InGaAs-Schicht ergibt. Wenn die InGaAs-Schicht unter einer Spannung τ_{InGaAs} steht, dann wird durch das Abgleiten der Versetzung eine Energie pro Längeneinheit von

$$\Delta E = \tau_{\mathrm{InGaAs}} b h_{\mathrm{InGaAs}} \tag{5.18}$$

freigesetzt [83]. Dieser Energiebetrag ΔE ist die maximal erreichbare kinetische Energie, wenn die Versetzung vorher in Ruhe war: $\bar{E}_{\mathrm{kin}} = \Delta E$. Mit

$$\tau_{\mathrm{InGaAs}} = 2G\frac{1+\nu}{1-\nu}\epsilon_{\mathrm{InGaAs}}, \tag{5.19}$$

Gleichung (5.2) und den experimentell gemessenen Werten der Restdehnungen $\epsilon_{\mathrm{InGaAs}}$ erhält man für $h_{\mathrm{GaAs}} = 2$ nm eine obere Schranke der kinetischen Energie pro Linienlänge von $\bar{E}_{\mathrm{kin}} = 21$ eV nm^{-1}, für $h_{\mathrm{GaAs}} = 25$ nm eine obere Schranke von $\bar{E}_{\mathrm{kin}} = 35$ eV nm^{-1}. Beide Werte der maximalen kinetischen Energie sind jeweils größer als die in Abschnitt 5.5.2 abgeschätzte Aktivierungsenergie. Daraus folgt, daß die maximale kinetische Energie einer in einer InGaAs-Einzelschicht gleitenden Versetzung groß genug ist, um einen nennenswerten Beitrag zur Aktivierungsenergie aufzubringen. Die maximale kinetische Energie ist sogar größer als die Aktivierungsenergie, so daß die Barriere in jedem Fall überwunden werden könnte, was wiederum der experimentellen Beobachtung widerspricht.

Der Grund dafür ist die Tatsache, daß die maximale kinetische Energie beim Versetzungsgleiten nicht erreicht wird. Dies wird einsichtig, wenn man aus der Energiebilanz auf die mit dem Gleiten verbundene Versetzungsgeschwindigkeit v zurückschließt. Bei hohen Versetzungsgeschwindigkeiten, die einen wesentlichen Bruchteil der Schallgeschwindigkeit ausmachen, gilt die Beziehung

$$E = \frac{E_0}{\sqrt{1-\frac{v^2}{c^2}}} \tag{5.20}$$

zwischen der Ruheenergie E_0 und der Gesamtenergie E einer sich gleichmäßig bewegenden Versetzung [83]. Für eine Stufenversetzung ist $c = c_{\mathrm{l}}$ zu setzen, für eine Schraubenversetzung $c = c_{\mathrm{t}}$.

Die kinetische Energie ist dann die Differenz zwischen Gesamtenergie und Ruheenergie:

$$E_{\mathrm{kin}} = E - E_0. \tag{5.21}$$

Nach Einsetzen aller Materialkonstanten und der InGaAs-Einzelschichtdicke h_{InGaAs} erhält man mit den Gleichungen (2.23) und (2.24) für die Ruheenergie pro Linienlänge einer 60°-Versetzung den Wert $E_0 = 15.7$ eV nm^{-1}.

Jetzt kann mit den Gleichungen (5.20) und (5.21) berechnet werden, wie hoch die Versetzungsgeschwindigkeit sein müßte, um die im vorigen Abschnitt abgeschätzte Aktivierungsenergie jeweils aufzubringen. Für die GaAs-Barriere mit $h_{\mathrm{GaAs}} = 2$ nm betrug die Aktivierungsenergie 1.9 eV nm^{-1}. Daraus folgt eine Versetzungsgeschwindigkeit von $v = 0.45c$. Für die GaAs-Barriere mit $h_{\mathrm{GaAs}} = 25$ nm betrug die Aktivierungsenergie 24 eV nm^{-1}. Daraus folgt eine Versetzungsgeschwindigkeit von $v = 0.92c$. Die Schallgeschwindigkeit von Scherwellen, c_{t}, beträgt 2.4 km/s, die Schallgeschwindigkeit von Kompressionswellen, c_{l}, beträgt 4.7 km/s. Die Versetzungsgeschwindigkeit müßte also in beiden Fällen sehr große Werte annehmen, um die GaAs-Einzelschichtbarriere durch das Aufwenden kinetischer Energie zu überwinden.

Diese hohen Versetzungsgeschwindigkeiten werden während des Relaxationsprozesses mit Sicherheit nicht erreicht [89, 173], so daß auch die kinetische Energie einer Versetzung nicht ausreicht, die Aktivierungsenergie aufzubringen.

5.5.4 Der Doppelkinkenmechanismus

Vom thermodynamischen Gesichtspunkt aus ist es sehr unwahrscheinlich, daß eine Fehlpassungsversetzung mit einer sehr großen Linienlänge gleichzeitig über die Einzelschichtbarriere gleitet. Die total aufzuwendende Energie wird abgesenkt, wenn man den mikroskopischen Gleitprozeß einer Versetzungslinie über die Peierls-Nabarro-Barriere auf die größere Längenskala der GaAs-Einzelschichtdicken überträgt.

Aus Gleichung (5.4) folgt nämlich, daß die Energieänderung kleiner wird, wenn nur ein kurzes Stück der Linie über die Barriere gleitet. Allerdings muß dazu die Bildungsenergie zur Erzeugung zweier Kinken aufgewendet werden [83].

Das Segment, das die Barriere überwunden hat, gleitet dann ungehindert unter der Fehlpassungsspannung der hinter der GaAs-Barriere liegenden InGaAs-Schicht zur nächsten InGaAs/GaAs-Grenzfläche. Die beiden Kinken gleiten schließlich unter der Spannung der InGaAs-Schicht in entgegengesetzte Richtungen auseinander und die Fehlpassungsversetzung wird kontinuierlich über die Barriere transportiert.

Der experimentelle Beleg dafür, daß der Doppelkinkenmechanismus im vorliegenden Übergittersystem aktiv ist, ist die Beobachtung einzelner Superkinken (Abschnitt 4.6), die eine in die [001]-Richtung projizierte Höhe von der Dicke genau einer Übergitterperiode haben. Die beobachteten Einzelkinken sind das Produkt des letzten Schrittes, bei dem die Doppelkinken schon weit voneinander entfernt sind.

Vom theoretischen Standpunkt aus existiert bei der Bildung von Doppelkinken das gleiche Problem einer sehr hohen Aktivierungsenergie des Barrierenmodells. Eine realistische Kinkenbreite L_{K} muß nämlich aus geometrischen Gründen etwa so groß sein, wie die entsprechenden GaAs-Einzelschichtdicken [83], so daß die total aufzuwendende Energie des Gleitprozesses bei der Doppelkinkenbildung gleich dem in Abschnitt 5.5.2, Gleichung (5.6), berechneten Wert der Aktivierungsenergie $\Delta E(h_{\mathrm{GaAs}})$ mal der Kinkenbreite L_{K} ist.

Zum Schluß dieses Abschnitts ist es wichtig zu bemerken, daß die Existenz von Fehlpassungsversetzungen an den Grenzflächen der Übergitterperioden, die ja nach den Gleichgewichtsmodellen eigentlich versetzungsfrei sein sollten, grundsätzlich mit in das Szenario einer durch die GaAs-Einzelschichten gehemmten Relaxation paßt. Diese können nämlich als Segmente aufgefaßt werden, die bei Erreichen des kritischen Relaxationgrades an einer InGaAs/GaAs-Grenzfläche gestoppt wurden, weil die entsprechende thermische oder mechanische Energie gerade nicht mehr ausreichte, um die Aktivierungsenergie aufzubringen.

5.6 Der Einfluß der Wachstumssequenz

Neben der Versetzungskinetik gibt es im Fall der Übergitter noch einen zweiten wichtigen Aspekt, der in Gleichgewichtsmodellen üblicherweise vernachlässigt wird, nämlich den Einfluß der zeitlichen Abfolge des Schichtwachstums auf die Gleichgewichtsversetzungsdichte.

In Modellen der Einzelschichtrelaxation ist die Berücksichtigung der zeitlichen Abfolge unwichtig, weil vom Standpunkt der Elastizitätstheorie aus die beiden Teilprozesse, Schichtwachstum und Relaxation, zeitlich entkoppelt werden können. Das heißt, die zeitliche Verlauf der Restdehnung $\epsilon(t)$ und des Versetzungsabstandes $S(t)$ sind mit Hilfe der zeitabhängigen Schichtdicke $h(t)$ aus den in Kapitel 2 abgeleiteten Gleichungen zu ermitteln.

Im Fall des Übergitterwachstums wird jedoch häufig schon durch das Wachsen weniger Übergitterperioden die kritische Gesamtdicke überschritten. Wenn schon einige Perioden gewachsen sind, dann kann man die Kinetik der Relaxation und das zu erreichende Gleichgewicht verfolgen, so wie es in der Diskussion der experimentellen Ergenbnisse bis zu diesem Abschnitt geschehen ist.

In dieser Situation ist es auch gerechtfertigt, analog zur Einzelschichtrelaxation das Übergitter zunächst versetzungsfrei anzunehmen, und dann oberhalb der kritischen Gesamtdicke den Verlauf der Relaxation, die sich hier wegen der Versetzungserzeugung an der Kristalloberfläche *entgegen* der Wachstumsrichtung entwickelt, bis zu ihrem Stillstand zu betrachten. Dieses Vorgehen führte im vorhergehenden Abschnitt zur Identifikation des Hemmungsmechanismus der Einzelschichtbarrieren.

In vielen fehlgepaßten Systemen ist jedoch oft schon die erste Einzelschicht der ersten Übergitterperiode überkritisch. In dieser Situation verläuft der Relaxationsprozeß nicht wie vorher von der Kristalloberfläche zur Substratgrenzfläche, sondern er entwickelt sich parallel zum Wachstum der Einzelschichten schon von der Substratgrenzfläche ab, bis hin zur Kristalloberfläche.

Das Minimum der freien Energie bleibt im Rahmen der konventionellen Relaxationsmodelle, zum Beispiel bezüglich der Versetzungsdichte an der Substratgrenzfläche, für den Grenzfall eines genügend dicken Übergitters derselbe, wie im ersten Fall, wo das Übergitter erst nach dem Wachstum einiger Perioden die kritische Gesamtdicke überschreitet. Der Relaxationsprozeß verläuft jetzt jedoch *in Wachstumsrichtung*, so daß auch die kinetischen Teilprozesse der Versetzungserzeugung und des Versetzungsgleitens anders verlaufen.

Zwei Beispiele sollen verdeutlichen, welche Konsequenz dieser zweite, „invertierte" Relaxationsmodus für die Versetzungsdichte an der Substratgrenzfläche hat.

5.6.1 Unterkritische GaAs-Schicht

In diesem Beispiel wird das Wachstum der ersten beiden InGaAs/GaAs-Übergitterperioden, also der ersten vier Einzelschichten betrachtet. Es wird angenommen, daß die InGaAs-Schichten überkritisch relativ zum GaAs-Substrat und die GaAs-Schichten unterkritisch relativ zur Gitterkonstante von InGaAs sind. Es wird wieder der Konvention gefolgt, daß f und ϵ positive Werte haben sollen. Die Betrachtung erfolgt zunächst unter der Voraussetzung, daß die Gleichgewichtsversetzungsdichte zu jedem Zeitpunkt des Schichtwachstums erreicht wird.

Nachdem die erste InGaAs-Schicht aufgewachsen ist, nimmt die Restdehnung dieser Schicht den Wert ϵ_1 an, der wegen der Relaxation kleiner als die Fehlpassung $f_1 = f$ ist. Der relaxierte Anteil der Fehlpassung, $\delta_1 = f_1 - \epsilon_1$, nimmt den Gleichgewichtswert an, der aus dem in Kapitel 2 dargestellten Modell hervorgeht.

Nach dem Wachstum der ersten GaAs-Schicht auf die InGaAs-Schicht ist die Fehlpassung des Schichtpakets auf den Wert

$$f_2 = \frac{h_{\mathrm{InGaAs}}}{h_{\mathrm{GaAs}} + h_{\mathrm{InGaAs}}} f \tag{5.22}$$

reduziert worden, $f_2 < f_1$, weil die GaAs-Schicht kohärent aufwächst, und weil sie mit ihrer kleineren Gitterkonstante zur Verminderung der Gesamtspannung beiträgt.

Wegen der größeren Gesamtdicke $h_1 = h_{\mathrm{GaAs}} + h_{\mathrm{InGaAs}}$ ist die mittlere Restdehnung ebenfalls kleiner geworden, $\epsilon_2 < \epsilon_1$. Da die GaAs-Schichtdicke unterkritisch ist, ist auch der relaxierte Anteil der Fehlpassung kleiner geworden, $\delta_2 = f_2 - \epsilon_2 < \delta_1$.

Daraus folgt, daß nach dem Wachsen der GaAs-Schicht ein Teil der Fehlpassungssegmente an der Substratgrenzfläche *eliminiert* werden muß. Dieser Sachverhalt ist der grundlegende Unterschied zwischen der gewöhnlichen Relaxation einer Einzelschicht oder der eines überkritischen Übergitters und des hier diskutierten „invertierten" Relaxationsmodus.

Der gesamte Prozeß wiederholt sich beim Aufwachsen der nächsten InGaAs-Schicht, denn die Fehlpassung des Schichtpakets steigt auf den Wert

$$f_3 = \frac{2h_{\mathrm{InGaAs}}}{h_{\mathrm{GaAs}} + 2h_{\mathrm{InGaAs}}} f > f_2. \tag{5.23}$$

Die Gesamtschichtdicke nimmt um den Betrag h_{InGaAs} zu, so daß die mittlere Restdehnung um einen kleinen Betrag abnimmt, $\epsilon_3 < \epsilon_2$. Der relaxierte Anteil der Fehlpassung nimmt somit zu, $\delta_3 = f_3 - \epsilon_3 > \delta_2$. Zum Erreichen des neuen Gleichgewichts müssen also wieder Versetzungen zur Substratgrenzfläche gleiten.

Das Aufwachsen der nächsten GaAs-Schicht resultiert analog zum Wachstum der ersten GaAs-Schicht wieder in einer Reduktion der Fehlpassung, einer kleinen Abnahme der mittleren Restdehnung, und in einer kleinen Abnahme des relaxierten Anteils der Restdehnung.

Durch Fortsetzen der Wachstumssequenz sieht man, daß sowohl die Fehlpassung, als auch der relaxierte Anteil der Fehlpassung, und damit die Versetzungsdichte an der Substratgrenzfläche, nach jedem Aufwachsen einer InGaAs-Schicht ein Maximum, und nach jedem Aufwachsen einer GaAs-Schicht ein Minimum annimmt. Eine quantitative Auswertung der Fehlpassungen f_i zeigt, daß die Amplitude der Zu- und Abnahme der Versetzungsdichte mit größer werdender Gesamtdicke abnimmt, und dem Gleichgewichtswert zustrebt, der auch schon durch die Gleichgewichtsrelaxationsmodelle vorhergesagt wird. Das Ergebnis der Berechnungen ist für ein (18/2)-Übergitter, ein (18/10)-Übergitter und ein (18/25)-Übergitter in den Abbildungen 5.2, 5.3 und 5.4 dargestellt. Die obere Kurve zeigt jeweils den Verlauf

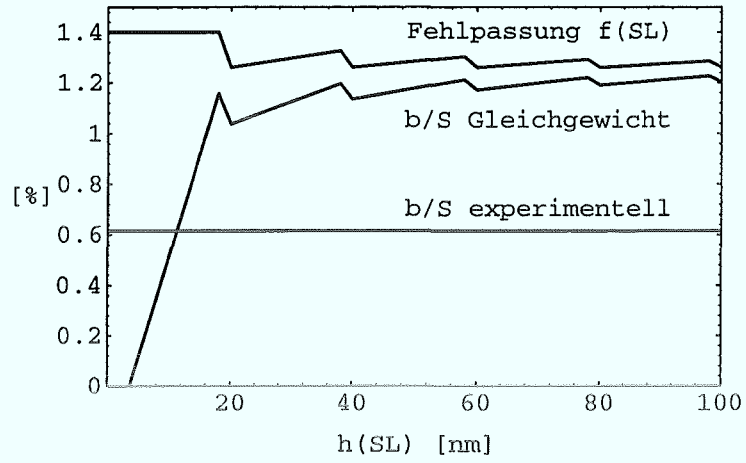


Abbildung 5.2: Fehlpassung und relaxierter Anteil der Fehlpassung eines (18/2)-Übergitters in Abhängigkeit von der Gesamtdicke des Übergitters. Die obere Kurve zeigt die Fehlpassung, die untere Kurve den relaxierten Anteil der Fehlpassung. Die untere, waagrechte Linie gibt den experimentell beobachteten, relaxierten Anteil der Fehlpassung für eine Gesamtdicke von 800 nm an.

der Fehlpassung mit zunehmender Gesamtdicke des Übergitters, die untere Kurve den relaxierten Anteil der mittleren Fehlpassung nach dem Gleichgewichtsmodell. In die Diagramme sind zusätzlich noch Linien eingezeichnet, die den experimentell gemessenen, relaxierten Anteil der Fehlpassung repräsentieren.

5.6.2 Überkritische GaAs-Schicht

Die Wachstumssequenz der Einzelschichten verläuft für eine GaAs-Schicht, deren Dicke überkritisch bezüglich der InGaAs-Gitterkonstante ist, grundsätzlich analog. Es ist, wie im vorigen Abschnitt dargestellt, eine Zu- und Abnahme der Fehlpassung des Schichtpakets und damit auch der Versetzungsdichte an der Substratgrenzfläche zu erwarten.

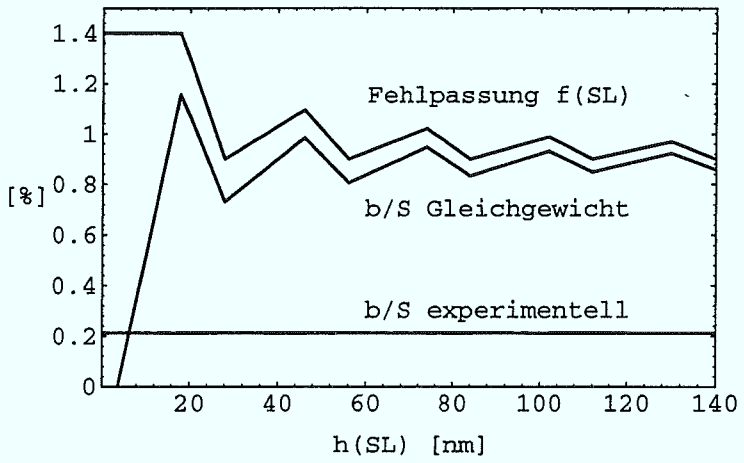


Abbildung 5.3: Fehlpassung und relaxierter Anteil der Fehlpassung eines (18/10)-Übergitters in Abhängigkeit von der Gesamtdicke des Übergitters. Die obere Kurve zeigt die Fehlpassung, die untere Kurve den relaxierten Anteil der Fehlpassung. Die untere, waagrechte Linie gibt den experimentell beobachteten, relaxierten Anteil der Fehlpassung für eine Gesamtdicke von 800 nm an.

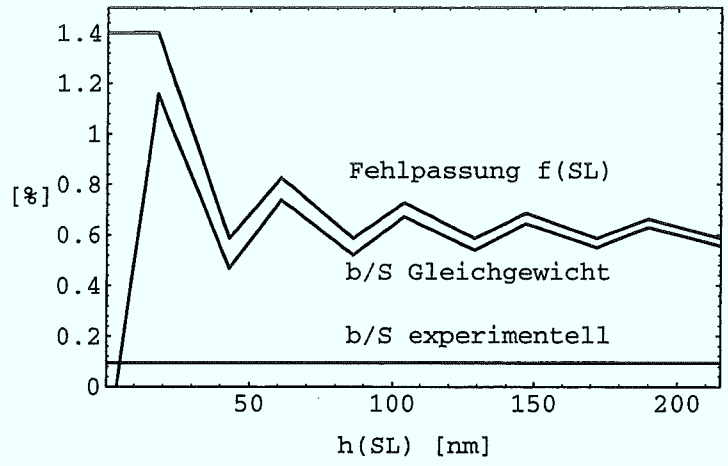


Abbildung 5.4: Fehlpassung und relaxierter Anteil der Fehlpassung eines (18/25)-Übergitters in Abhängigkeit von der Gesamtdicke des Übergitters. Die obere Kurve zeigt die Fehlpassung, die untere Kurve den relaxierten Anteil der Fehlpassung. Die untere, waagrechte Linie gibt den experimentell beobachteten, relaxierten Anteil der Fehlpassung für eine Gesamtdicke von 800 nm an.

Hier tritt jedoch zusätzlich die Komplikation auf, daß die jeweils oberste GaAs-Schicht relaxiert und somit ein Fehlpassungsnetzwerk an der obersten InGaAs-GaAs-Grenzfläche existiert. Wenn die GaAs-Schichtdicke zusammen mit der InGaAs-Schichtdicke ein unterkritisches Paar von Schichtdicken bezüglich der Relaxation eines Übergitters bildet (Gleichung (2.57)), dann muß das Netzwerk der obersten Grenzfläche beim weiteren Aufwachsen von Übergitterperioden wieder entfernt werden. Dieser zweite kinetische Teilprozeß ist der Zu- und Abnahme der Versetzungsdichte an der Substratgrenzfläche überlagert.

5.6.3 Kinetische Barrieren

Die beiden vorangehenden Beispiele zeigen, daß das Erreichen der Gleichgewichtsdichte von Versetzungen an der Substratgrenzfläche während der Wachstumssequenz, unabhängig von der Dicke der GaAs-Einzelschichten, sowohl mit der Erzeugung als auch mit dem Entfernen von Fehlpassungssegmenten verbunden ist.

Dieser komplexe Relaxationsmodus ist mit dem Gleiten von Versetzungen verbunden, wobei die einzelnen Versetzungen die zu einem bestimmten Zeitpunkt existierenden Einzelschichten durchdringen müssen. Dabei wirken die verspannten Einzelschichten analog zu der in Abschnitt 5.5 diskutierten Situation als Barrieren des Versetzungsgleitens.

Es ist zu erwarten, daß die InGaAs-Barrieren im Vergleich zu den GaAs-Barrieren eine untergeordnete Rolle spielen, weil im ersten Fall eine Hemmung auf eine schon existierende Versetzung wirkt, und weil im zweiten Fall die GaAs-Barriere nicht nur beim Gleiten wirksam ist, sondern auch schon hemmend auf die Versetzungserzeugung an der Kristalloberfläche wirken kann.

Aus dieser qualitativen Betrachtung und mit den gleichen Argumenten, die im Abschnitt 5.5 zur Abschätzung des relaxierten Anteils der Fehlpassung benutzt wurden, ist auch bei der hier diskutierten Berücksichtigung der Wachstumssequenz eine $1/h_{\mathrm{GaAs}}$-Abhängigkeit der Versetzungsdichte zu erwarten.

5.6.4 Wachstumssequenz bei kleiner GaAs-Einzelschichtdikke

Eine Berechnung der sukzessiven Fehlpassungen f_i (Gleichungen (5.22), (5.23)) für sehr kleine GaAs-Einzelschichtdicken, $h_{\mathrm{GaAs}} \ll h_{\mathrm{InGaAs}}$, zeigt, daß die Zu- und Abnahme der Fehlpassungen nur schwach ausgeprägt ist. Gleichzeitig variieren die sukzessiven homogenen Restdehnungen ϵ_i beim Wachsen einer GaAs-Schicht noch viel schwächer mit der Gesamtdicke des Übergitters, weil die GaAs-Einzelschichtdicke sehr klein ist (Abbildung 5.2).

Daraus folgt, daß die Abnahme des relaxierten Anteils der Restdehnung, $\delta_i = f_i - \epsilon_i$, die in erster Linie durch die Abhnahme der Fehlpassung f_i bestimmt ist, ebenfalls schwach ausgeprägt ist.

In diesem Fall ist die Entwicklung der Versetzungsdichte an der Substratgrenzfläche durch den Anstieg der mittleren Fehlpassung während des InGaAs-Schichtwachstums bestimmt, so daß das wachsende Übergitter sich fast wie eine dicke,

relaxierende Einzelschicht verhält. Das Versetzungsgleiten findet dann vorwiegend von der Kristalloberfläche zur Grenzfläche statt, während die Elimination von Fehlpassungsversetzungen in der umgekehrten Richtung vernachlässigt werden kann.

In Übergittern mit sehr dünnen GaAs-Einzelschichten ist auch bei der Berücksichtigung der GaAs-Barrieren der höchste Relaxationsgrad zu erwarten, und die damit verbundenen Erzeugungs- und Multiplikationsraten von Fehlpassungsversetzungen sind hier größer als bei dickeren GaAs-Einzelschichten.

Bei dem in dieser Arbeit untersuchten Übergitter mit der kleinsten GaAs-Schichtdicke, $h_{\mathrm{GaAs}} = 2$ nm, ist durch Extrapolation der Relation (5.11) ein relaxierter Anteil der Fehlpassung von 1.24 % zu erwarten. Die Fehlpassung selbst beträgt 1.26 %, so daß für diese kleine Schichtdicke eine praktisch vollständige Relaxation des Übergitters zu erwarten wäre.

Experimentell wird jedoch eine Restdehnung von 0.65 % beobachtet, was ein starker Hinweis darauf ist, daß die Hemmung der Versetzungserzeugung und der Versetzungsmultiplikation im vorliegenden InGaAs/GaAs-System dort eine Rolle spielt, wo die durch die Restdehnung verursachte treibende Kraft relativ klein ist.

5.6.5 Partielle Relaxation während der Wachstumssequenz

Das Resultat des vorigen Abschnitts war die Feststellung, daß auch bei kleinen GaAs-Einzelschichtdicken experimentell keine vollständige Relaxation beobachtet wird. Im Prinzip wäre jetzt die gesamte Betrachtung der Wachstumssequenz, die ja in Abschnitt 5.6.1 unter der Voraussetzung durchgeführt wurde, daß zu jedem Zeitpunkt die Gleichgewichtsdichte erreicht werden kann, mit der Voraussetzung einer partiellen Relaxation der gewachsenen Einzelschichten zu wiederholen.

Aus den Diagrammen für die Fehlpassung und den relaxierten Anteil der Fehlpassung, die in den Abbildungen 5.2, 5.3 und 5.4 gezeigt sind, ist jedoch abzulesen, daß der experimentell gemessene, relaxierte Anteil der Restdehnung $\delta_{\mathrm{exp}}(h_{\mathrm{SL}})$, der in den Diagrammen mit einer waagrechten Linie eingetragen ist, in allen Übergittern niedriger liegt, als das Minimum des Gleichgewichtswerts $\delta_{\mathrm{min}}(h_{\mathrm{SL}})$. Bei der experimentell beobachteten, partiellen Relaxation strebt das System also beim Wachstum jeder Einzelschicht dem Gleichgewichtswert $\delta(h_{\mathrm{SL}})$ zu, ohne ihn tatsächlich zu erreichen.

Im Hinblick auf die in den vorangegangenen Abschnitten diskutierte GaAs-Einzelschichtbarriere ist festzustellen, daß die während der Wachstumssequenz erzeugten Versetzungssegmente, auch bei der hier diskutierten partiellen Relaxation, von ihren Erzeugungsort an der Kristalloberfläche zu Substratgrenzfläche gleiten müssen. Die aus der Modellvorstellung der Einzelschichtbarriere abgeleitete Abschätzung der benötigten Aktivierungsenergie pro Linienlänge einer Fehlpassungsversetzung ist zwar auch hier so groß, daß das Überwinden der einzelnen GaAs-Schichten allein durch thermische Anregung unmöglich sein sollte. Doch werden jetzt zwei Extremfälle der Einzelschichtbarriere diskutiert, aus denen abgeleitet werden kann, daß das Gleiten über die Barrieren dennoch während der Wachstumssequenz stattfindet.

Im ersten Fall wird angenommen, daß die Barrieren so hoch sind, das das Ver-

setzungsgleiten schon bei geringen Spannungen τ_{GaAs} vollständig unterbunden wird. Dann würde bei allen Übergittern die erste aufwachsende InGaAs-Schicht, die ja in allen Fällen eine Dicke von $h_{InGaAs} = 18$ nm hat, und die überkritisch bezüglich der Einzelschichtrelaxation ist, bis zu einem bestimmten Grad δ_1 relaxieren. Die folgende GaAs-Schicht, die eine unterkritische Schichtdicke hat, würde nach dem Aufwachsen eine Dehnung von $\epsilon_{GaAs} = \delta_1$ besitzen, und nach der Voraussetzung jedes weitere Versetzungsgleiten, und somit auch eine Änderung von δ_1 unterbinden. Die Konsequenz daraus wäre ein konstanter Relaxationsgrad der untersten InGaAs-Schicht und damit aller Übergitter, *unabhängig* von der GaAs-Einzelschichtdicke. Dies steht in Widerspruch zu den experimentellen Meßwerten (Tabelle 4.1).

Im zweiten Fall wird angenommen, daß die Barrieren so niedrig sind, daß das Versetzungsgleiten zu keinem Zeitpunkt während der Wachstumssequenz behindert ist. In diesem Fall wäre die Versetzungsdichte an der Substratgrenzfläche durch die Kinetik der Versetzungserzeugung und Versetzungsmultiplikation bestimmt. Da alle Übergitter eine gleiche Gesamtdicke von 800 nm haben, wäre der Relaxationsgrad nur eine Funktion der Restdehnung ϵ_{SL} eines Übergitters, die ja die treibende Kraft für Erzeugung und Multplikation ist. Dies steht ebenfalls in Widerspruch zu den experimentellen Meßwerten, denn wäre die Restdehnung ϵ_{SL} der einzig bestimmende Faktor, dann müßte die Relaxation ja in allen Übergittern beim exakt gleichen Wert der Restdehnung zum Stillstand kommen.

Aus den beiden Grenzfällen folgt unmittelbar, daß auch im Fall einer kinetischen Behinderung der Versetzungserzeugung und der Versetzungsmultiplikation die GaAs-Einzelschichtbarrieren eine Rolle spielen. Denn wenn die Barrieren das Versetzungsgleiten weder völlig behindern, noch völlig unbeeinflußt lassen, dann bleibt als einzige Konsequenz die schon in Abschnitt 5.6.1 diskutierte Abhängigkeit der Barrierenhöhe von der GaAs-Schichtverspannung, die die schon beschriebene $1/h_{GaAs}$-Abhängigkeit des relaxierten Anteils der Restdehnung zur Folge hat.

Natürlich hat der überhaupt niedrige Relaxationsgrad in allen Übergittern seine Ursache mit Sicherheit, genauso wie im Fall einer relaxierenden Einzelschicht, in einer kinetischen Behinderung der Versetzungserzeugung und der Versetzungsmultiplikation.

5.6.6 Experimenteller Beleg

Nach dieser Analyse der Wachstumssequenz kann eine Erklärung für zwei experimentelle Beobachtungen gegeben werden, die mit den bisherigen Gleichgewichtsmodellen und auch mit Hilfe des Einzelschichtbarrieren-Mechanismus nicht eingeordnet werden konnten, und die ein Beleg für den Einfluß der Wachstumssequenz sind.

Es wurde nämlich beobachtet, daß Einzelschichten mit Hilfe des Dipolmechanismus relaxiert sind (Abschnitte 4.3, 4.4), obwohl sie bezüglich der Relaxation von Übergitterperioden unterkritisch sind.

Mit Berücksichtigung der Wachstumssequenz können die unteren Versetzungssegmente einer Dipol- und einer Haarnadelkonfiguration als Fehlpassungsversetzungen gedeutet werden, die bei der Relaxation einer gerade gewachsenen überkritischen obersten Einzelschicht entstanden sind (Abschnitt 5.6.2). Beim Wachsen der

nächsten Einzelschicht wird die neue Übergitterperiode unterkritisch, und die noch von der Relaxation her existierenden Fadensegmente sollten so durch die Einzelschichten gleiten, daß die Fehlpassungsversetzungen wieder eliminiert werden.

Wenn jedoch dabei ein Fadensegment durch ein Hindernis blockiert wird, dann kann das zugehörige Fehlpassungssegment nicht mehr vollständig entfernt werden. Das Fadensegment durchdringt die weiter aufwachsende Einzelschicht, und erfährt dort eine zusätzliche Kraft, weil diese Schicht entgegengesetzt verspannt ist. Deshalb wird dort das Zurückgleiten des Fadensegments vorangetrieben, wobei ein zweites paralleles Fehlpassungssegment erzeugt wird.

Der so erzeugte Dipol hat die gleiche Gestalt wie ein Matthewsscher Dipol [124], der ja durch das Abgleiten eines Fadensegmentes in einem schon existierenden Schichtpaket entsteht (Abschnitt 4.4.2), unterscheidet sich aber von diesem dadurch, daß er hier durch „Vor- und Zurückgleiten" eines Fadensegmentes schon während des Schichtwachstums entsteht.

Der zweite experimentelle Beleg dafür, daß während der Wachstumssequenz tatsächlich Fehlpassungssegmente eliminiert werden, ist die Beobachtung von „Doppel-U"-Konfigurationen, die mit großer Wahrscheinlichkeit durch Entfernen des zwischen den U-Konfigurationen liegenden Segments entstanden sind (Abschnitt 4.5).

5.7 Zusammenfassung der Diskussion

In den vorangegangenen Abschnitten dieses Kapitels wurden eine Reihe von Mechanismen diskutiert, die mögliche Ursachen für die experimentell beobachtete partielle Relaxation der InGaAs/GaAs-Übergitter sind.

Eine Abschätzung der einzelnen Wirkungen auf die Versetzungserzeugung und das Versetzungsgleiten während der Relaxation zeigt, daß folgende Einflüsse eine untergeordnete Bedeutung haben: Die elastische Verspannung des Substrats direkt unterhalb der Substratgrenzfläche hat eine nur lokale Wirkung auf die Versetzungskonfiguration der Fehlpassungssegmente (Abschnitt 5.3). Die Wechselwirkung einer gleitenden Versetzung mit den schon an der Substratgrenzfläche vorhandenen Versetzungen eines Fehlpassungsnetzwerks ist ebenfalls lokal. Die durch das inhomogene Verzerrungsfeld des Netzwerks verursachte Barriere ist nur schwach ausgeprägt, und sie kann deshalb nur wenig zur Reduktion des Relaxationsgrades beitragen (Abschnitt 5.4).

Die wichtigste Ursache für die nur partielle Relaxation eines stark überkritischen Übergitters ist die kinetische Behinderung der Versetzungserzeugung und der Versetzungsmultiplikation [35]. Damit verhält sich ein dickes, relaxierendes Übergitter nicht anders, als eine dicke relaxierende Einzelschicht mit homogener Indiumkomposition.

Die experimentell beobachtete Abhängigkeit des Relaxationsgrades von der GaAs-Einzelschichtdicke kann auf eine Behinderung des Versetzungsgleitens durch eine Verspannung der GaAs-Einzelschichten, die der ursprünglichen Fehlpassungsspannung entgegengesetzt ist, zurückgeführt werden (Abschnitt 5.5). Diese Einzelschichtbarrieren behindern das Gleiten von Versetzungen von der Kristalloberfläche zur

Substratgrenzfläche (Abschnitt 5.6). Eine Abschätzung der Aktivierungsenergie der GaAs-Einzelschichtbarriere durch die mechanische Arbeit pro Linienlänge, die eine Versetzung beim Gleiten durch die verspannte GaAs-Schicht leisten muß, führt jedoch zu so großen Werten, daß Versetzungen bei der Wachstumstemperatur eigentlich nicht über die GaAs-Barrieren gleiten können.

Andererseits geben die experimentell gemessenen Versetzungsdichten an den Substratgrenzflächen und die Mikrostruktur innerhalb der Übergitter, insbesondere die Beobachtung einzelner Superkinken einen starken Hinweis darauf, daß das Versetzungsgleiten über die GaAs-Barrieren während der Wachstumssequenz doch stattfindet. Weiterhin läßt sich ein Widerspruch zu den experimentellen Meßwerten ableiten, wenn man annimmt, daß die GaAs-Barrieren wegen der hohen Aktivierungsenergien das Versetzungsgleiten völlig behindern würden (Abschnitt 5.6.5).

Daß die kinetische Behinderung der Versetzungserzeugung und die Versetzungsmultiplikation mit großer Wahrscheinlichkeit für die überhaupt niedrigen Relaxationsgrade verantwortlich ist, zeigt die Restdehnung bei kleinen GaAs-Einzelschichten, $h_{\mathrm{GaAs}} < 6$ nm (Abschnitt 5.6.4), wo zu erwarten ist, daß die GaAs-Barrieren im Vergleich zu den anderen Fällen, $h_{\mathrm{GaAs}} > 6$ nm, eine geringere Rolle spielen. Tatsächlich weicht der relaxierte Anteil der Restdehnung bei $h_{\mathrm{GaAs}} = 2$ nm auch von der empirisch gefundenen $1/h_{\mathrm{GaAs}}$-Abhängigkeit ab (Gleichung (5.11)).

Teil II

Rekonstruktion des projizierten Potentials in der hochauflösenden Elektronenmikroskopie

Kapitel 6

Zusammenfassung und Einleitung Teil II

In modernen Anwendungen der Festkörpertechnologie, der Mikroelektronik und der Mikrotechnik ist die Materialwissenschaft zunehmend vor die Aufgabe gestellt, strukturelle Charakterisierung auf kleinsten Längenskalen, bis hinunter auf die atomare Ebene zu leisten. Der Grund dafür liegt einerseits in der ursächlichen Bestimmung makroskopischer Festkörpereigenschaften durch eine gegebene Mikrostruktur, andererseits in dem Bestreben, Materialeigenschaften durch geeignete Kontrolle der Mikrostruktur maßzuschneidern.

Die hochauflösende Transmissionselektronenmikroskopie bietet beim heutigen Stand der Technik ein Auflösungsvermögen im Scherzerfokus von 0.17 nm, die absolute Auflösungsgrenze liegt bei 0.13 nm. Unter erheblichem technischem und finanziellem Aufwand ist es in Einzelfällen – gemessen an der Zahl der verfügbaren Mikroskope – möglich, die absolute Auflösungsgrenze auf 0.10 nm zu reduzieren. Beobachtungen auf dieser Längenskala ermöglichen die Untersuchung einer Vielzahl interessanter materialwissenschaftlicher Problemstellungen.

Ein Beispiel einer solchen materialwissenschaftlichen Fragestellung ist die Aufklärung der Kernstruktur einer 60°-Fehlpassungsversetzung, die der bei weitem häufigste Defekt bei der Relaxation fehlgepaßter epitaktischer Halbleiterschichten mit Diamant- oder Zinkblendestruktur bei niedriger Fehlpassung ist.

Die technisch verfügbare, maximale Auflösungsgrenze elektronenmikroskopischer Aufnahmen ist jedoch nicht das einzige Kriterium, das für eine erfolgreiche Strukturaufklärung maßgebend ist. Die zweite wichtige Eigenschaft einer elektronenmikroskopischen Abbildung ist ihre *direkte Interpretierbarkeit*, die danach beurteilt wird, in welchem Grad zum Beispiel die atomare Struktur eines Kristalldefektes der subjektiven Betrachtungsweise eines Beobachters zugänglich ist.

Im Gegensatz zur Lichtmikroskopie, die durch eine jahrundertelange technische Verfeinerung dem menschlichen Sehen so angepaßt werden konnte, daß dem Beobachter eine direkte Strukturinterpretation ohne den Gebrauch zusätzlicher Hilfsmittel möglich ist, kann die Elektronenmikroskopie beim derzeitigen technischen Stand nur ein zweifach verschlüsseltes Abbild der Strukturinformation liefern.

Der erste Verschlüsselungsprozeß ist die Beugung hochenergetischer Elektronen

an der Objektstruktur, die selbst bei den derzeit präparationsbedingt erreichbaren, kleinsten Probendicken eine nichtlineare Abbildung der Strukturinformation in die an der Probenunterseite austretende Elektronenwelle ist. Der zweite Verschlüsselungsprozeß ist die elektronenoptische Abbildung der Austrittswellenfunktion in die Bildebene des Mikroskops, der wegen der Abbildung einer komplexen Wellenfunktion in reelle Bildintensitäten ebenfalls ein grundsätzlich nichtlinearer Vorgang ist. Die elektronenoptische Abbildung ist zusätzlich durch eine Reihe von Aberrationen gekennzeichnet, deren wichtigste die Defokussierung der Objektivlinse, die sphärische und chromatische Aberration, der zweizählige und dreizählige Astigmatismus, sowie eine Strahlverkippung gegenüber der optischen Achse sind.

Wegen dieser komplexen Abbildung der Strukturinformation beschränkt sich die konventionelle Strukturbestimmung gegenwärtig auf die Vorgabe eines Strukturmodells, der nachfolgenden Simulation von dynamischer Beugung und elektronenmikroskopischer Abbildung und dem anschließenden Vergleich der simulierten Bilder mit den experimentellen Bildern per Augenschein [243]. Diese Sequenz wird solange mit unterschiedlichen Strukturmodellen wiederholt, bis eine befriedigende Übereinstimmung zwischen Experiment und Simulation erreicht ist. Die Methode ist eingeschränkt einerseits durch die beschränkte Anzahl von Strukturmodellen, die in vertretbarer Zeit durchprobiert werden können, und andererseits in der erreichbaren Präzision der Strukturdaten aufgrund der subjektiven Beurteilung durch den Beobachter.

Prinzipiell bestehen zwei Möglichkeiten, die genannten Beschränkungen zu umgehen. Die erste Möglichkeit basiert auf der Idee, die konventionelle Strukturbestimmung durch Anwendung eines objektiven, quantitativen numerischen Bildvergleichs, einer Bildmustererkennung und Bildmusterinterpretation zu automatisieren. Sie ist besonders effektiv in Fällen, in denen ein Vorwissen über die zu untersuchende Struktur besteht [215, 233]. Die zweite Möglichkeit besteht in der zweistufigen Rekonstruktion der Objektstruktur durch Invertieren des mikroskopischen Abbildungsprozesses, der von den aufgezeichneten Bildern zur Austrittswellenfunktion führt [221, 223, 212, 201, 249, 250, 247], und der anschließenden Lösung des inversen Streuproblems, das von der zuvor bestimmten Austrittswellenfuntion zur projizierten Struktur führt [251, 235, 211, 197, 241]. Sie ist einzusetzen in Fällen, in denen kein Vorwissen über die zu untersuchende Struktur besteht, oder in denen die vorgenannte Methode scheitert.

In den letzten Jahren haben technische Fortschritte in der Elektronenmikroskopie, wie die Verwendung hochkohärenter Feldemissions-Elektronenquellen und die Verringerung der wichtigsten Primäraberration, der sphärischen Aberration der Objektivlinse, sowie der Einsatz von digitalen Bildaufnahmetechniken auf der Basis von CCDs *(charge coupled devices)* einerseits, und der stetige Leistungszuwachs moderner Digitalrechner andererseits, den Einsatz von Phasenrekonstruktionsalgorithmen und Strukturrekonstruktionsalgorithmen möglich gemacht. Unter *Phasenrekonstruktion* werden im folgenden alle Rekonstruktionstechniken zusammengefaßt, die die Ermittlung der Austrittswellenfunktion aus elektronenmikroskopischen Bildern zum Ziel haben, während unter *Strukturrekonstruktion* oder *Potentialrekonstruktion* der zweite Rekonstruktionsschritt verstanden wird, der von der Austrittswellenfunktion

zur zweidimensional projizierten atomaren Struktur führt.

In der vorliegenden Arbeit werden zwei alternative Verfahren vorgestellt, die die Phasenrekonstruktion und die Potentialrekonstruktion zum Gegenstand haben. Die Invertierung der beiden Prozesse, die die Strukturinformation verschlüsseln, wird dabei durch die Verwendung stochastischer Suchalgorithmen ermöglicht, die besonders effizient sind, wenn die zu ermittelnde Objektstruktur periodisch ist. Die Phasenrekonstruktion auf der Basis einer stochastischen „Simulated-Annealing"-Suche und auf der Basis eines „genetischen" Suchalgorithmus wird sowohl mittels einer simulierten Defokusserie elektronenmikroskopischer Bilder als Eingabe, als auch anhand einer experimentell aufgenommenen Defokusserie der intermetallischen Verbindung Ni_4Mo demonstriert. Die Potentialrekonstruktion wird mittels einer stochastischen „Simulated-Annealing"-Suche unter Benutzung von simulierten Austrittswellenfunktionen von Galliumarsenid demonstriert.

Kapitel 7

Verschlüsselung der Strukturinformation

Den in dieser Arbeit untersuchten Phasen- und Potentialrekonstruktionsverfahren liegen Modelle zugrunde, die die beiden Verschlüsselungsprozesse der hochauflösenden Transmissionselektronenmikroskopie, die dynamische Beugung hochenergetischer Elektronen am atomaren Kristallpotential der Objektstruktur und die elektronenmikroskopische Abbildung beschreiben. Diese Modelle resultieren in Berechnungsverfahren, die zum Beispiel im Rahmen der quantenmechanisch-optischen Bildsimulation Bestandteil der konventionellen Strukturbestimmung sind.

Alle fortgeschrittenen Rekonstruktionsverfahren, einschießlich der hier untersuchten stochastischen Rekonstruktionsalgorithmen, enthalten intern jeweils eine „Vorwärts"-Berechnung von Wellenfunktionen, beziehungsweise eine „Vorwärts"-Berechnung elektronenmikroskopischer Bilder, obwohl das Ergebnis der Rekonstruktionsalgorithmen eine „Rückwärts"-Berechnung der Austrittswellenfunktion, beziehungsweise eine „Rückwärts"-Berechnung des projizierten Kristallpotentials darstellt. Dieser Zusammenhang wird im späteren Kapitel 9 dargestellt.

Dieses Kapitel enthält eine Darstellung der beiden Standardsimulationsverfahren zur Berechnung einer Austrittswellenfunktion aus dem atomaren Kristallpotential, der *Blochwellenmethode* [198] und des *Multislice-Algorithmus* [202], und eine kurze Darstellung des Standardmodells der elektronenmikroskopischen Abbildung, das wiederum auf dem Modell der *Transmissionskreuzkoeffizienten* von Zweistrahlinterferenzen unter partiell kohärenter Beleuchtung beruht [206, 219].

7.1 Die Blochwellenmethode

Die Blochwellenmethode zur Berechnung einer Austrittswellenfunktion $\psi(\vec{r})$ geht aus von der Lösung der Schrödingergleichung für ein gegebenes atomares Kristallpotential $V(\vec{r})$ und ein die Elektronen beschleunigendes Potential E:

$$\nabla^2\psi(\vec{r}) + \frac{8\pi^2 me}{h^2}\left(E + V(\vec{r})\right)\psi(\vec{r}) = 0, \tag{7.1}$$

mit der Elektronenmasse m, der Elementarladung e und der Planckschen Konstante h. Bei typischen, in Elektronenmikroskopen verwendeten Beschleunigungsspannungen von 100 kV bis 1 MV ist grundsätzlich eine relativistische Rechnung auf der Grundlage der Dirac-Gleichung erforderlich, es wurde jedoch gezeigt [207, 208], daß die nicht-relativistische Schrödingergleichung in guter Näherung verwendet werden darf, wenn für die De-Broglie-Wellenlänge der Elektronen λ, die Potentiale E und $V(\vec{r})$, und die Elektronenmasse m relativistisch korrigierte Werte eingesetzt werden.

Für die Bewegung im Vakuum, bei $V(\vec{r}) = 0$, lautet die Lösung

$$\psi(\vec{r}) = \exp(2\pi i \vec{\chi} \cdot \vec{r}), \tag{7.2}$$

wobei der Betrag des Wellenvektors $\vec{\chi}$ die Gleichung

$$\frac{h^2\chi^2}{2m} = eE \tag{7.3}$$

erfüllen muß. Die Vakuumwellenlänge der Elektronen hat dann den Wert $\lambda = 1/\chi$.

Für die Bewegung eines Elektrons in einem Kristallpotential $V(\vec{r})$ ist es naheliegend, nachdem die Lösung im Fall des Vakuums eine ebene Welle ist, eine Lösung $\psi(\vec{r}) = b(\vec{k}, \vec{r})$ zu suchen, die aus einer ortsabhängigen Funktion $C(\vec{r})$ mit Kristallperiodizität und einem Phasenfaktor besteht:

$$b(\vec{k}, \vec{r}) = C(\vec{r}) \exp(2\pi i \vec{k} \cdot \vec{r}). \tag{7.4}$$

$\vec{k}$ ist der Wellenvektor der Blochwelle $b(\vec{k}, \vec{r})$.

Geeignete Lösungen für $\vec{k}$ und $C(\vec{r})$ müssen die Schrödingergleichung erfüllen. Sie werden gewonnen, indem man die Fourierdarstellung der Blochfunktion

$$C(\vec{r}) = \sum_{\vec{g}} C_{\vec{g}}(\vec{k}) \exp(2\pi i \vec{g} \cdot \vec{r}) \tag{7.5}$$

in Gleichung (7.4) einsetzt und dann $b(\vec{k}, \vec{r})$ und die Fourierdarstellung des Kristallpotentials

$$V(\vec{r}) = \sum_{\vec{g}} V_{\vec{g}} \exp(2\pi i \vec{g} \cdot \vec{r}) = \frac{h^2}{2me} \sum_{\vec{g}} U_{\vec{g}} \exp(2\pi i \vec{g} \cdot \vec{r}) \tag{7.6}$$

in die Schrödingergleichung (7.1) einsetzt. Die Fourierentwicklungen erstrecken sich über rezpiroke Gittervektoren $\vec{g}$ des periodischen Kristalls.

Das Resultat ist eine Eigenwertgleichung, die die Blochwellenkoeffizienten $C_{\vec{g}}(\vec{k})$ und die zugehörigen Wellenvektoren $\vec{k}$ enthält:

$$\left(K^2 - \left(\vec{k} + \vec{g}\right)^2\right) C_{\vec{g}}(\vec{k}) + \sum_{\vec{h} \neq 0} U_{\vec{h}} C_{\vec{g}-\vec{h}}(\vec{k}) = 0. \tag{7.7}$$

Die Konstante K ist der Betrag des Wellenvektors im Kristall, er ist mit dem Betrag des Wellenvektors im Vakuum, χ, über das mittlere Potential U_0 des Kristalls verknüpft:

$$K^2 = \frac{2meE}{h^2} + U_0 = \chi^2 + U_0. \tag{7.8}$$

Dieser Sachverhalt ist die zur Lichtoptik analoge Situation der Änderung der Wellenlänge aufgrund einer Änderung des Brechungsindex.

Aus rechentechnischen Gründen müssen die Fourierentwicklungen des Kristallpotentials und der Blochfunktionen bei einer endlichen Zahl N von ebenen Wellen abgebrochen werden. In diesem Fall hat das System linearer Eigenwertgleichungen (7.7) die Dimension N, so daß bei Vernachlässigung der Rückwärtsstreuung grundsätzlich N verschiedene Lösungen $C_{\vec{g}}(\vec{k})$ zu N verschiedenen Eigenwerten $\vec{k}$ existieren. Die volle Lösung der Schrödingergleichung ist dann eine beliebige Linearkombination der N Blochwellen $b(\vec{k},\vec{r})$:

$$\psi(\vec{r}) = \sum_{\vec{k}} \alpha(\vec{k}) b(\vec{k},\vec{r}), \tag{7.9}$$

mit der Anregungsstärke $\alpha(\vec{k})$ der Blochwelle mit dem Wellenvektor $\vec{k}$.

Im Fall der dynamischen Beugung einer aus dem Vakuum auf einen Kristall einfallenden ebenen Elektronenwelle (Gleichung (7.2)) ist zusätzlich noch die Kontinuitätsbedingung an der Kristalloberfläche für $\psi(\vec{r})$ und $\nabla\psi(\vec{r})$ zu erfüllen, die diese Mannigfaltigleit auf eine eindeutige Lösung reduziert. Im symmetrischen Laue-Fall ist die Anregungsstärke einer Blochwelle gleich dem konjugiert komplexen Wert ihrer Fourierkomponente der Raumfrequenz $\vec{g} = 0$:

$$\alpha(\vec{k}) = C_0^*(\vec{k}). \tag{7.10}$$

Für praktisch alle in der Transmissionselektronenmikroskopie verwendeten Beschleunigungsspannungen gilt die sogenannte Hochenergienäherung, das heißt das Kristallpotential ist viel kleiner als das die Elektronen beschleunigende Potential, $V(\vec{r}) \ll E$. Dann sind die Größen K, k und $|\vec{k}+\vec{g}|$ sehr groß gegenüber ihren relativen Differenzen, so daß der Betrag des Anregungsfehlers

$$\vec{s}_{\vec{g}} = \vec{k} - \vec{K} - \vec{g} \tag{7.11}$$

und die Größe

$$\gamma(\vec{k}) = k_z - K_z \tag{7.12}$$

klein sind. Die Koordinate z verläuft hierbei senkrecht zur Laue-Ebene und damit senkrecht zu $\vec{g}$. Dann gilt für den Koeffizienten in Gleichung (7.7):

$$(K^2 - (\vec{k}+\vec{g})^2) \approx K^2 - (K + (\gamma(\vec{k}) - s_{\vec{g}})\cos\theta_{\vec{g}})^2 \approx 2K(s_{\vec{g}} - \gamma(\vec{k}))\cos\theta_{\vec{g}}, \tag{7.13}$$

und der Winkel $\theta_{\vec{g}}$ liegt in der Größenordnung des sehr kleinen Bragg-Winkels, so daß $\cos\theta_{\vec{g}} \approx 1$.

Mit den genannten Näherungen lautet die Eigenwertgleichung (7.7) in Matrixschreibweise:

$$\boldsymbol{AC} = \boldsymbol{C}\{\gamma(\vec{k})\}, \tag{7.14}$$

mit

$$A_{\vec{g},\vec{h}} = \begin{cases} s_{\vec{g}}, & \vec{g} = \vec{h} \\ U_{\vec{g}-\vec{h}}/2K, & \vec{g} \neq \vec{h} \end{cases}, \tag{7.15}$$

$\boldsymbol{C}$ der Matrix, die spaltenweise die Blochwellenkoeffizienten $C_{\vec{g}}(\vec{k})$ enthält, und der Diagonalmatrix $\{\gamma(\vec{k})\}$ mit den zugehörigen Größen $\gamma(\vec{k})$.

In einer Berechnung ist man in einer Vielzahl von Fällen eher an einer Darstellung der Austrittswellenfunktion interessiert, die Amplituden und Phasen einzelner „Strahlen", also ebener Wellen der Raumfrequenz $\vec{g}$, enthält, als an einer Realraumdarstellung $\psi(\vec{r})$. Zu diesem Zweck ordnet man die Strahlen, die ja die Fourierkomponenten $\Psi_{\vec{g}}$ von $\psi(\vec{r})$ sind, in einem Spaltenvektor $\boldsymbol{\Psi}$ an.

Mit dem Spaltenvektor $\boldsymbol{\alpha}$ der Anregungsstärken $\alpha(\vec{k})$ der Blochwellen lautet das Ergebnis:

$$\boldsymbol{\Psi} = \boldsymbol{C}\{\exp(2\pi i\gamma(\vec{k})z)\}\boldsymbol{\alpha}. \tag{7.16}$$

Für die Berechnung der Austrittswellenfunktion bei einer bestimmten Probendicke muß dann nur noch die Koordinate z durch die Probendicke t ersetzt werden.

Mit dem Matrixformalismus kann auch der Zusammenhang zwischen der Darstellung einer Wellenfunktion in der Basis der ebenen Wellen, die mit den Beugungsvektoren $\vec{g}$ abgezählt werden, und der Darstellung in der Basis der Blochwellen, die nach ihren Wellenvektoren $\vec{k}$ abgezählt werden, kompakt dargestellt werden. Die Basiswechselmatrix ist die Spaltenmatrix der Blochwellenkoeffizienten $C_{\vec{g}}(\vec{k})$ und die Transformation lautet:

$$\boldsymbol{C}\boldsymbol{\alpha} = \boldsymbol{\Psi}. \tag{7.17}$$

Die Rücktransformation kann mittels $\boldsymbol{C}^{-1} = \tilde{\boldsymbol{C}}$ umgeformt werden, weil die Spaltenmatrix der Blochwellenkoeffizienten eine orthogonale Matrix der Eigenvektoren ist:

$$\tilde{\boldsymbol{C}}\boldsymbol{\Psi} = \boldsymbol{\alpha}. \tag{7.18}$$

Die Tilde bezeichnet eine komplex konjugierte und transponierte Matrix.

Dieses Ergebnis kann mit Gleichung (7.16) benutzt werden, um die Abhängigkeit einer Austrittswellenfunktion von einer beliebigen „Eintritts"-Wellenfunktion $\boldsymbol{\Psi}'$, die nicht notwendigerweise eine ebene Welle sein muß, darzustellen:

$$\boldsymbol{\Psi} = \underbrace{\boldsymbol{C}\{\exp(2\pi i\gamma(\vec{k})z)\}\tilde{\boldsymbol{C}}}_{\boldsymbol{P}}\boldsymbol{\Psi}'. \tag{7.19}$$

$\boldsymbol{P}$ ist die *Streumatrix* der dynamischen Beugung.

Ein wichtiger praktischer Aspekt bei der Benutzung der Blochwellenmethode ist die Wahl einer genügend großen Zahl N ebener Wellen in den Fourierentwicklungen. Es werden nämlich sehr oft Einstrahlrichtungen betrachtet, die nahe einer hochsymmetrischen Kristallrichtung liegen, so daß bei der dynamischen Beugung eine große Zahl von Strahlen angeregt ist. Dann kann die Qualität einer Berechnung dadurch überprüft werden, daß eine zweite Berechnung mit einer höheren Strahlzahl durchgeführt wird, und anschließend die Differenz zur ersten Rechnung kontrolliert wird.

Bei der Beugung hochenergetischer Elektronen an realistisch dicken Proben tritt neben der elastischen Streuung auch inelastische Streuung auf, deren wichtigste Ursachen Plasmonenanregung, Phononenanregung und Elektron-Elektron-Streuung sind. Inelastisch gestreute Elektronen tragen nicht mehr zum Interferenzkontrast

der elastisch gestreuten Elektronen bei, weil sie entweder so stark gestreut wurden, daß sie die Objektivblende nicht mehr passieren können, oder weil ihre Wellen nicht mehr kohärent sind.

Das einfachste Modell, welches die Reduktion der elastisch gestreuten Elektronen beschreibt, benutzt den Übergang von reellen zu komplexen Wellenvektoren der gebeugten Elektronenwellen, wobei der zusätzlich eingeführte Imaginärteil $\vec{q}(\vec{k})$ des Wellenvektors $\vec{k}$ wegen der Beziehung (7.16) als Dämpfungskonstante wirkt. Dies ist mathematisch äquivalent mit dem Übergang von einem reellen Kristallpotential $V(\vec{r})$ zu einem komplexen Potential. In der einfachsten Näherung wird dabei angenommen, daß der zusätzliche Imaginärteil proportional zum Realteil ist:

$$V(\vec{r}) \longrightarrow V(\vec{r}) + \mathrm{i}\kappa V(\vec{r}). \tag{7.20}$$

In der Fourierdarstellung entspricht dies der Ersetzung:

$$V_{\vec{g}} \longrightarrow V_{\vec{g}} + \mathrm{i}\kappa V_{\vec{g}}, \tag{7.21}$$

$$U_{\vec{g}} \longrightarrow U_{\vec{g}} + \mathrm{i}\kappa U_{\vec{g}}. \tag{7.22}$$

Für κ werden Werte eingesetzt, die von der Ordnungszahl der streuenden Atome abhängen, und die etwa 0.05 bis 0.15 betragen.

Die Absorptionskonstante κ ist deutlich kleiner als 1, und die für eine numerische Rechnung benötigten Größen $q_z(\vec{k})$ können deshalb mittels einer quantenmechanischen Störungsrechung aus der obigen, ungestörten Lösung für die Absorption Null ermittelt werden. Das Ergebnis ist

$$q_z(\vec{k}) = \frac{\kappa m e}{h^2 k_z} \int b^*(\vec{k},\vec{r}) V(\vec{r}) b(\vec{k},\vec{r})\, \mathrm{d}A, \tag{7.23}$$

mit Flächenelementen $\mathrm{d}A$ senkrecht zur z-Achse. In der Matrixschreibweise lautet das gleiche Ergebnis:

$$q_z(\vec{k}) = \kappa \sum_{\vec{g},\vec{h}} C^*_{\vec{g}}(\vec{k})\, \Delta A_{\vec{g},\vec{h}}\, C_{\vec{h}}(\vec{k}), \tag{7.24}$$

mit

$$\Delta A_{\vec{g},\vec{h}} = \left(1 - \delta_{\vec{g},\vec{h}}\right) \frac{U_{\vec{g}-\vec{h}}}{2K}. \tag{7.25}$$

Die Dämpfungskonstanten $q_z(\vec{k})$ werden mit den Anregungsstärken $\alpha_{\vec{k}}$ zusammengefaßt, so daß die Darstellung (7.16) gültig bleibt, wenn die Ersetzung

$$\alpha_{\vec{k}} \longrightarrow \alpha_{\vec{k}} \exp(-2\pi q_z(\vec{k}) z) \tag{7.26}$$

gemacht wird.

7.1.1 Nichtlinearität der dynamischen Beugung

Mit Hilfe des Blochwellenformalismus kann ein direkter Zusammenhang zwischen dem atomaren Kristallpotential und der Austrittswellenfunktion hergleitet werden. Zu diesem Zweck wird die Diagonalmatrix in Gleichung (7.19) so umgeformt, daß die Exponentialfunktion nach „außen" vorgezogen werden kann:

$$\{\exp(2\pi i\gamma(\vec{k})z)\} = \exp\{(2\pi i\gamma(\vec{k})z)\}. \tag{7.27}$$

Die Identität ist durch Taylorentwicklung beider Seiten einzusehen.

Die zweite Umformung betrifft das Matrixprodukt in Gleichung (7.19), das die Streumatrix $\boldsymbol{P}$ bildet. Auch hier kann der Exponentialoperator vor das Produkt gezogen werden, weil die Spaltenmatrix der Blochwellenkoeffizienten, $\boldsymbol{C}$, orthogonal ist:

$$\boldsymbol{C}\exp\{(2\pi i\gamma(\vec{k})z)\}\tilde{\boldsymbol{C}} = \exp(\boldsymbol{C}\{(2\pi i\gamma(\vec{k})z)\}\tilde{\boldsymbol{C}}). \tag{7.28}$$

Die Identität ist ebenfalls durch Taylorentwicklung der Exponentialoperatoren auf beiden Seiten einzusehen.

Das Produkt auf der rechten Seite in Gleichung (7.28) kann mit der Eigenwertgleichung (7.14) und Ausnutzen der Orthogonalität von $\boldsymbol{C}$ weiter zusammengefaßt werden:

$$\exp(\boldsymbol{C}\{(2\pi i\gamma(\vec{k})z)\}\tilde{\boldsymbol{C}}) = \exp(2\pi i\boldsymbol{A}z). \tag{7.29}$$

Die Gleichung (7.19) lautet dann nach allen Umformungen:

$$\Psi = \exp(2\pi i\boldsymbol{A}z)\Psi'. \tag{7.30}$$

Sie stellt die mathematische Abbildung des Kristallpotentials bei fester Eintrittswellenfunktion Ψ' in die Austrittswellenfunktion Ψ dar.

Die Einträge der Matrix $\boldsymbol{A}$ sind, von den Diagonalelementen abgesehen, proportional zu den Fourierkoeffizienten des Kristallpotentials (Gleichung (7.15)), jedoch geht diese Linearität in der Austrittswellenfunktion verloren, weil der Exponentialoperator nichtlinear ist.

Die in Gleichung (7.30) sichtbare Nichtlinearität ist der Grund dafür, daß eine Strukturinformation, die im atomaren Kristallpotential enthalten ist, auf eine komplexe Weise in der Austrittswellenfunktion verschlüsselt ist.

7.1.2 Interpretierbarkeit des schwachen Phasenobjekts

Die Verschlüsselung der Strukturinformation in der Austrittswellenfunktion ist nur schwach, wenn die Probendicke t sehr klein ist, oder wenn das atomare Kristallpotential schwach ist. In solchen Fällen ist das Argument des Exponentialoperators in Gleichung (7.30), die Matrix $2\pi i\boldsymbol{A}z$, „klein" gegenüber der Einheitsmatrix $\boldsymbol{I}$, das heißt jeder ihrer Einträge ist viel kleiner als 1.

Dann ist die lineare Entwicklung nach der Matrix $2\pi i\boldsymbol{A}z$ in Gleichung (7.30) eine gute Näherung für die Streumatrix $\boldsymbol{P}$:

$$\Psi \approx (\boldsymbol{I} + 2\pi i\boldsymbol{A}z)\Psi'. \tag{7.31}$$

Bei einer kleinen Probendicke t und Einstrahlung einer ebenen Welle, $\Psi'_{\vec{g}} = \delta_{\vec{g},0}$, hat der Nullstrahl, Ψ_0, den Wert 1 und die abgebeugten Strahlen erhalten die Werte

$$\Psi_{\vec{g}} \approx 2\pi \mathrm{i} \frac{U_{\vec{g}}}{2K} t. \tag{7.32}$$

Unter den genannten Bedingungen ist die Intensität der abgebeugten Strahlen viel kleiner als 1, $|\Psi_{\vec{g}}|^2 \ll 1$, und die Strukturinformation ist direkt und linear im Imaginärteil der Austrittswellenfunktion enthalten. Weil der Imaginärteil für kleine Argumente gleich dem Phasenwinkel in der komplexen Ebene ist, wird die gesamte Situation als *Beugung am schwachen Phasenobjekt* bezeichnet.

Durch Transformation der Gleichung (7.32) in den Realraum sieht man, daß die Phase der Austrittswellenfunktion ein direkt interpretierbares Abbild des Kristallpotentials, und damit der projizierten atomaren Struktur darstellt.

Diese Situation trifft aber praktisch nur für Objektstrukturen zu, die nur Atome relativ niedriger Ordnungszahlen enthalten, und die eine nur sehr kleine Probendicke aufweisen. Reale Objektstrukturen, die zum Beispiel Gegenstand einer materialwissenschaftlichen Untersuchung sind, enthalten jedoch erstens in sehr vielen Fällen Atome mittlerer bis großer Ordnungszahl, und die Präparationstechnik setzt zweitens oft eine untere Grenze für die minimal herstellbare Probendicke.

Deshalb ist in materialwissenschaftlichen Anwendungen die Voraussetzung des schwachen Phasenobjekts nicht gegeben, so daß eine direkte Strukturinterpretation der Austrittswellenfunktion in dem hier dargestellten Zusammenhang nicht möglich ist.

Eine wichtige Anwendung der schwachen Phasenobjektnäherung ist jedoch bei der Kontrastanalyse spezieller Testpräparate gegeben, die aus dünnem amorphem Material, zum Beispiel Kohlenstoff oder Silizium, bestehen und die zur Bestimmung der Kontrasttransfereigenschaften eines Elektronenmikroskops dienen (Abschnitt 7.3).

7.2 Der Multislice-Algorithmus

Die Herleitung der Blochwellenmethode stützte sich auf die Fourierentwicklung eines periodischen Kristallpotentials und die entsprechende Entwicklung von Wellenfunktionen nach einem Satz N ebener Wellen. Die Verwendung der Blochwellenmethode auf den in der hochauflösenden Transmissionselektronenmikroskopie relevanten Fall der Einstrahlung entlang einer niedrigindizierten, hochsymmetrischen Kristallrichtung erfordert in der Regel das Einbeziehen mehrerer hundert Strahlen in die Fourierentwicklung, um eine Konvergenz der Lösung zu erreichen. Die Zahl der zu berücksichtigenden Strahlen wächst sehr stark an, wenn die Beugung an nichtperiodischen Strukturen berechnet werden soll, ohne daß es zu gravierenden Randeffekten an den Grenzen der periodisch fortgesetzt gedachten Austrittswellenfunktion einer Defektstruktur kommt.

Der benötigte Speicherplatz bei der Blochwellenmethode ist proportional zu N^2, und der Rechenaufwand bei der Lösung des Eigenwertproblems ist bei vollbesetzter

Matrix im ungünstigsten Fall proportional zu N^2. Diese letzte Eigenschaft macht das Blochwellenverfahren für die Simulation der dynamischen Beugung an Defektstrukturen wenig geeignet, und es wurden Methoden entwickelt, die bei gleicher Strahlzahl N effizienter arbeiten.

Das gegenwärtig populärste, und in allen Standard-Simulationspaketen enthaltene Verfahren ist der Multislice-Algorithmus. Er ist hergeleitet aus dem Prinzip der Beugung am schwachen Phasenobjekt (Abschnitt 7.1.2) und der Wellenausbreitung nach dem Huygens-Fresnelschen Prinzip.

Zur Simulation der dynamischen Beugung wird der Modellkristall in N Schichten (engl.: *slices*) der Dicke Δz entlang der z-Achse aufgeteilt. Nach dem Huygens-Fresnelschen Prinzip ist die Austrittswellenfunktion an der Unterseite der Schicht $n = 1 \ldots N$, $\psi(\vec{r})$, gegeben durch:

$$\psi_n(\vec{r}) = [\psi_{n-1}(\vec{r}) \otimes p_n(\vec{r})] \cdot q_n(\vec{r}). \tag{7.33}$$

$p_n(\vec{r})$ ist die *Propagationsfunktion*, $q_n(\vec{r})$ die *Transmissionsfunktion* und $\otimes$ bezeichnet das Faltungsprodukt. Indem man die Iterationsvorschrift (7.33) für $n = 1 \ldots N$ ausführt, erhält man die Austrittswellenfunktion $\psi_N(\vec{r})$ nach einer Probendicke von $t = N\ \Delta z$. An den Anfang der Iteration wird die eingestrahlte ebene Welle, $\psi_0(\vec{r}) = 1$ gestellt.

Die Iterationsvorschrift hat ein Analogon im Reziprokraum, sie lautet nach der Fouriertransformation der Gleichung (7.33)

$$\Psi_n(\vec{k}) = \left[\Psi_{n-1}(\vec{k}) \cdot P_n(\vec{k})\right] \otimes Q_n(\vec{k}), \tag{7.34}$$

wobei $\Psi_n(\vec{k})$, $P_n(\vec{k})$ und $Q_n(\vec{k})$ die Fouriertransformierten der Funktionen $\psi_n(\vec{r})$, $p_n(\vec{r})$ und $q_n(\vec{r})$ sind.

Gewöhnlich wird die sehr zeitaufwendige direkte Berechnung des Faltungsprodukts in Gleichung (7.34) unter Ausnutzen des Faltungstheorems durch eine Fouriertransformation der Faktoren in den Realraum, einer Multiplikation im Realraum und die anschließende Rücktransformation des Produkts in den Reziprokraum ersetzt. Die resultierende Iterationsvorschrift für den Multislice-Zyklus lautet dann im Reziprokraum:

$$\Psi_n(\vec{k}) = \mathcal{F}\left(\mathcal{F}^{-1}\left(\Psi_{n-1}(\vec{k}) \cdot P_n(\vec{k})\right) \cdot q_n(\vec{r})\right). \tag{7.35}$$

Die Realraumdarstellung der Multislice-Iteration (Gleichung (7.33)) kann in einer völlig äquivalenten Weise umformuliert werden.

In vielen Anwendungen sind die Propagationsfunktion und die Transmissionsfunktion für alle Schichten n gleich, und sie haben die Werte

$$P(\vec{k}) = \exp(2\pi i s_{\vec{k}} \Delta z) \tag{7.36}$$

und

$$q(\vec{r}) = \exp(i\sigma \Phi_P(\vec{r}) \Delta z). \tag{7.37}$$

$s_{\vec{k}}$ ist der Anregungsfehler, σ die relativistische Wechselwirkungskonstante und

$$\Phi_P(\vec{r}) = \frac{1}{\Delta z} \int_0^{\Delta z} V(\vec{r})\, \mathrm{d}z \tag{7.38}$$

das projizierte atomare Kristallpotential pro Längeneinheit.

Der rechenaufwendigste Schritt im Multislice-Verfahren ist die Ausführung einer Fouriertransformation. Sie muß pro Schicht zweimal ausgeführt werden. Der numerische Rechenaufwand steigt mit der Zahl der einbezogenen Strahlen N proportional zu $N \log_2 N$, was immer noch erheblich günstiger ist, als die Lösung der Eigenwertaufgabe im Blochwellenverfahren, auch wenn der Modellkristall in eine größere Anzahl von Schichten unterteilt werden muß.

Die Schichtdicke Δz muß in der Praxis für ein gegebenes Kristallpotential so dünn gewählt werden, daß bei der Beugung der einfallenden Welle $\psi_{n-1}(\vec{r})$ durch das projizierte Potential die Näherung des schwachen Phasenobjekts gültig ist. Das heißt, die Phase der Transmissionsfunktion $q_n(\vec{r})$ muß stets viel kleiner als 1 sein. Diese Beschränkung führt immer noch zu einer Zahl von Einzelschichten, die klein genug ist, um den Rechenaufwand nicht übermäßig anwachsen zu lassen. In einem konkreten Anwendungsfall ist es ausreichend, die gesamte Multislice-Iteration für die doppelte Strahlzahl N oder eine halbierte Schichtdicke Δz zu wiederholen, um die numerische Korrektheit eines Ergebnisses zu überprüfen.

Das außerordentlich große Potential des Multislice-Verfahrens liegt darin, daß bei gleichem Rechenaufwand gegenüber dem Blochwellenverfahren eine sehr viel größere Zahl von Strahlen einbezogen werden kann, so daß auch die Simulation der dynamischen Beugung an Defekten ermöglicht wird, die nichtperiodisch senkrecht zum Elektronenstrahl verlaufen. Obwohl die Herleitungen von Multislice- und Blochwellenverfahren verschieden sind, liefern sie im Limes großer Strahlzahlen und kleiner Schichtdicken gleiche Resultate.

7.3 Modell der Kontrastübertragung

Die in dieser Arbeit beschriebenen Rekonstruktionsalgorithmen zur Phasenrekonstruktion, also der Bestimmung der Austrittswellenfunktion aus einer Serie von elektronenmikroskopischen Bildern, enthalten Berechnungsprozeduren, die es gestatten, aus einer gegebenen Austrittswellenfunktion unter Berücksichtigung der Kontrasttransfereigenschaften eines Elektronenmikroskops Bilderserien zu simulieren.

Diese Bildberechnung basiert auf einem Standardmodell der Kontrastübertragung, das schon seit rund zwei Jahrzehnten etabliert ist [206, 219], und das in allen relevanten Simulationspaketen implementiert ist. Die vollständige mathematische Herleitung des Modells ist sehr umfangreich und insbesondere die quantitative Diskussion aller Eigenschaften würde den Rahmen der vorliegenden Darstellung sprengen.

Stattdessen geht die folgende Darstellung einen Weg, der von der Beschreibung der Eigenschaften eines „idealen" Kontrasttransfers ausgeht, und dann zu einer Diskussion der real vorliegenden Kontrasttransferbedingungen fortschreitet. Die Darstellung schließt dann mit einem Spezialfall des realen Kontrasttransfers, der linearen Kontrastübertragung, ab. Parallel dazu wird der Rechengang skizziert, der der ausführlichen mathematischen Herleitung zugrundeliegt, und dessen Ergebnisse werden abschließend in einer kompakten Form dargestellt.

7.3.1 Idealer Kontrasttransfer

Die nach der dynamischen Beugung an der Probenunterseite austretende Wellenfunktion trägt die Strukturinformation des beugenden Objektes grundsätzlich mit einer Auflösung, die nach den Gesetzen der Optik nur durch die verwendete Wellenlänge begrenzt ist. Bei gebräuchlichen Beschleunigungsspannungen liegt diese Auflösung in der Größenordnung weniger Pikometer, das ist eine um etwa zwei Größenordnungen kleinere Längenskala, als zur Auflösung von projizierten Atomabständen kristalliner Materialien entlang niedrigindizierter Kristallrichtungen notwendig ist.

Ein perfektes Kontrasttransfersystem, das die Austrittswellenfunktion in beobachtbare Bildintensitäten übersetzen würde, bräuchte also zunächst keine andere Eigenschaft, als einen Ausschnitt dieser hochaufgelösten Austrittswellenfunktion stark zu vergrößern und auf einem Schirm als Bildintensität darzustellen. Der ideale Kontrasttransfer lautet also:

$$I(\vec{r}) = |\psi(\vec{r})|^2, \tag{7.39}$$

wobei in der gesamten Theorie des Kontrasttransfers immer mit der Vergrößerung $M = 1$ gerechnet wird.

Es sind fünf Eigenschaften des idealen Kontrasttransfers abzulesen:

- Wenn die Austrittswellenfunktion rein phasenmoduliert ist, zum Beispiel nach der Beugung an einem schwachen Phasenobjekt (Abschnitt 7.1.2), dann ist die Bildintensität an allen Orten 1, und die Strukturinformation ist vollständig verloren.

- Eine Amplitudenmodulation der Austrittswellenfunktion wird vollständig übertragen.

- Wenn die Austrittswellenfunktion sowohl amplituden- als auch phasenmoduliert ist, dann geht die Phaseninformation durch die Betragsquadratbildung der Austrittswellenfunktion verloren.

- Selbst wenn die Austrittswellenfunktion eine direkte und lineare Strukturinformation trägt (Abschnitt 7.1.2), dann ist die Bildinformation wegen der Betragsquadratbildung nichtlinear.

- Die Bildintensität hat grundsätzlich die gleiche hohe Ortsauflösung wie die Austrittswellenfunktion.

Das bedeutendste Defizit ist somit der Verlust der Phaseninformation, egal ob die Austrittswellenfunktion rein phasenmoduliert oder gemischt amplituden- und phasenmoduliert ist. Das ideale Kontrasttransfersystem sollte also zusätzlich noch eine Einrichtung haben, mit der bei Bedarf zuerst die Phasenmodulation in eine Amplitudenmodulation umgewandelt, und dann erst die Betragsquadratbildung durchgeführt wird. In der komplexen Zahlenebene bedeutet es, daß die abgebeugten Strahlen in der Austrittswellenfunktion mit einem zusätzlichen Phasenfaktor

$\exp(2\pi i\chi)$, mit $\chi = \pm 1/4$ multipliziert werden. Das Resultat wird nach dem Vorzeichen der zusätzlichen Phasendrehung „positiver" oder „negativer" Phasenkontrast genannt.

7.3.2 Realer Kontrasttransfer

Experimentell wird beobachtet, daß der Kontrasttransfer für hohe Raumfrequenzen der Bildintensität, also bei der Auflösung kleiner Objektdetails, effektiv zusammenbricht. Eine Ortsauflösung in der Größenordnung der Wellenlänge wird nicht erreicht, sondern sie ist mit Werten größer 0.10 nm etwa zwei Zehnerpotenzen kleiner.

Mathematisch kann diese Beobachtung durch einen Satz von Dämpfungskoeffizienten $T(\vec{k}, \vec{k}')$ berücksichtigt werden, der am einfachsten in einer Fourierraumdarstellung der Gleichung (7.39) eingearbeitet wird [206]:

$$I(\vec{g}) = \int \Psi(\vec{g}+\vec{h})\, \Psi^*(\vec{h})\, T(\vec{g}+\vec{h}, \vec{h})\, \mathrm{d}\vec{h}. \tag{7.40}$$

Die Dämpfungskoeffizienten der Zweistrahlinterferenz haben zwei Argumente, weil sie gleichzeitig den im vorigen Abschnitt diskutierten zusätzlichen Phasenfaktor $\exp(2\pi i\chi(\vec{k}))$ aufnehmen können, der jetzt schon in Hinblick auf die folgende Diskussion raumfrequenzabhängig angenommen wurde.

Die Funktion $\chi(\vec{k})$ wird *Aberrationsfunktion* genannt. $T(\vec{k}, \vec{k}')$ wird durch Einbeziehen der Phasenfaktoren komplex und der Betrag soll zwischen Null (totale Dämpfung) und 1 (vollständige Transmission) liegen. Die Größe T wird *Transmissionskreuzkoeffizient* genannt [219]. In den folgenden Abschnitten werden die experimentellen Ursachen für das Auftreten der zusätzlichen Phasenfaktoren und Dämpfungsterme genannt, und es wird die Gestalt der einzelnen Faktoren diskutiert, soweit sie die in dieser Arbeit untersuchten Phasenrekonstruktionsverfahren betreffen.

7.3.3 Die Raumfrequenzabhängigkeit der Aberrationsfunktion

Das elektronenoptische Abbildungssystem eines Transmissionselektronenmikroskops hat, analog zu lichtoptischen Linsensystemen, eine Reihe von Aberrationen, die derzeit technisch noch nicht vollständig korrigiert werden können. Die wichtigste nichtkorrigierte Aberration ist die sphärische Aberration der Objektivlinse, die die von einem Gegenstandspunkt ausgehenden perfekten Kugelwellen so „deformiert", daß die bildseitigen Wellenfronten nicht wieder in einem Punkt auf der optischen Achse konvergieren.

Die sphärische Aberration hat eine Raumfrequenzabhängigkeit, die proportional zu k^4 ist, das heißt, die Abschnitte der einfallenden Kugelwellen, die unter den größten Winkeln auf die Ebene der Objektivlinse treffen, werden am stärksten mit einer zusätzlichen Phase $\chi(\vec{k})$ beaufschlagt. Diese Phase hängt zusätzlich noch von der Wellenlänge λ der einfallenden Elektronen ab:

$$\chi_s(\vec{k}) = \frac{1}{4} C_s \lambda^3 k^4, \tag{7.41}$$

mit der Konstante der sphärischen Aberration C_s. Diese gerätespezifische Konstante hat in Transmissionselektronenmikroskopen, die speziell für hochauflösende Abbildung optimiert sind, einen Wert von zum Beispiel 1.0 mm (Herstellerangabe für ein JEOL 4000EX).

Die zweite bedeutende Aberration des Kontrasttransfersystems ist eine mögliche Defokussierung der Objektivlinse. Die Defokussierung setzt sich zusammen aus dem geometrischen Abstand Z_G des abzubildenden Objektes von der „idealen" Gegenstandsweite der Objektivlinse, und einem Anteil Z_O, der von einer Abweichung des „idealen" Objetivlinsenstroms herrührt. Der zweite Anteil ist eine charakteristische Besonderheit eines elektronenoptischen Kontrasttransfersystems, denn durch eine Variation des Objektivlinsenstroms kann die Brechkraft der Linse kontinuierlich geändert werden.

Die Aberration aufgrund der Defokussierung $Z = Z_G + Z_O$ hat eine Raumfrequenzabhängigkeit, die proportional zu k^2 ist, sie ist ebenfalls abhängig von der Wellenlänge λ:

$$\chi_F(\vec{k}) = \frac{1}{2} Z \lambda k^2. \tag{7.42}$$

Diese beiden wichtigsten Beiträge zur Aberrationsfunktion

$$\chi(\vec{k}) = \chi_s(\vec{k}) + \chi_F(\vec{k}) = \frac{1}{4} C_s \lambda^3 k^4 + \frac{1}{2} Z \lambda k^2 \tag{7.43}$$

sind Funktionen des Betrags von $\vec{k}$, sie sind deshalb rotationssymmetrisch um die optische Achse.

Der Vollständigkeit halber wird an dieser Stelle ergänzt, daß es noch weitere Aberrationsterme gibt, die nicht nur vom Betrag, sondern auch von der Richtung von $\vec{k}$ abhängen. Die beiden wichtigsten zusätzlichen Aberrationen sind der zweizählige und der dreizählige Astigmatismus, deren Ursache zum Beispiel eine nicht perfekt rotationssymmetrische Linsenanordnung sein kann.

Ein weiterer wichtiger Abbildungsfehler des Kontrasttransfersystems ist eine Strahlverkippung des Nullstrahls gegenüber der optischen Achse. Die Richtung und die Stärke der Srahlverkippung kann, analog zu den Beugungsvektoren $\vec{k}$, durch einen Vektor $\vec{k}_{tilt}$ im Reziprokraum dargestellt werden. Die Kippung des Strahls korrespondiert zu einer Nullpunktsverschiebung im Reziprokraum, so daß alle in der Aberrationsfunktion vorkommenden Vektoren $\vec{k}$ durch $\vec{k} + \vec{k}_{tilt}$ ersetzt werden müssen. Dadurch kann die Aberrationsfunktion eine relativ komplexe Gestalt erhalten, insbesondere, wenn zweizähliger und dreizähliger Astigmatismus berücksichtigt werden.

Die Strahlverkippung, der zweizählige und der dreizählige Astigmatismus sind nicht in den untersuchten Phasenrekonstruktionsverfahren berücksichtigt, so daß hier bezüglich einer vollständigen Darstellung auf die Literatur verwiesen wird.

Phasenkontrastabbildung im Scherzer-Paßband

Anhand der Eigenschaften eines idealen Kontrasttransfersystems wurde gezeigt, daß zur Abbildung eines reinen Phasenobjekts eine zusätzliche Phasendrehung von $\pm\pi/2$

notwendig ist, um eine Transformation in einen Amplitudenkontrast zu erreichen (Abschnitt 7.3.1).

Die Aberrationsfunktion $\chi(\vec{k})$ (Gleichung (7.43)) hat mit C_s und Z zwei Parameter, die das Verhältnis der k^4-Abhängigkeit und der k^2-Abhängigkeit, und damit den Phasenwinkel bestimmen. C_s ist eine unveränderliche Gerätekonstante, während die Defokussierung Z durch eine Änderung der Objektivlinsenstroms variiert werden kann. Dieser variable Parameter erlaubt es, für einzelne Raumfrequenzen $\vec{k}$ die Bedingung $\chi(\vec{k}) = \pm 1/4$ zu erfüllen. Aus der Periodizität des Phasenfaktors $\exp(2\pi i\chi(\vec{k}))$ folgt, daß auch noch ganzzahlige Werte n zur Aberrationsfunktion addiert werden können, ohne daß sich eine Kontraständerung ergibt.

In einem Experiment ist man nicht daran interessiert, nur eine einzige Raumfrequenz, oder einen diskreten Satz mit der „richtigen" Phasendrehung zu übertragen, sondern ein möglichst breites Band von Raumfrequenzen, das zusätzlich noch von einer möglichst kleinen Raumfrequenz $k_{\min}$ nahe Null bis zu einer möglichst großen Raumfrequenz $k_{\max}$ verlaufen soll. Der Grund dafür liegt in einer wichtigen Eigenschaft des menschlichen Sehens, dessen Mustererkennung optimiert ist, tiefpaßgefilterte Bildinformation in direkt interpretierbare Sinneseindrücke zu verwandeln.

Aus der zweifachen Forderung ein „möglichst breites" Raumfrequenzband, das nahe $k = 0$ beginnen soll, und das eine Aberration „möglichst nahe" bei $\pm(1/4 + n)$ haben soll, resultiert ein Optimierungsproblem bezüglich des Defokusparameters Z und der ganzen Zahl n, dessen Lösung

$$Z_{\text{Scherzer}} = -\sqrt{\frac{4}{3} C_s \lambda} \tag{7.44}$$

für die ganze Zahl $n = 1$ ist [238]. Bei diesem optimalen Fokus (*Scherzerfokus*) nimmt die Aberrationsfunktion an den Stellen $k = 0$ und

$$k_{\max} = (3/4\ C_s \lambda^3)^{-1/4} \tag{7.45}$$

den Wert Null an. Dazwischen befindet sich ein flaches Minimum, und die Aberrationsfunktion nimmt über einen weiten Bereich innerhalb des Intervalls $[0, k_{\max}]$ Werte nahe $\chi = 1/3$ an. Die Abweichung vom idealen Wert $\chi = 1/4$ wurde dabei in Kauf genommen, um $k_{\max}$ zu einer etwas höheren Raumfreqenz zu verschieben.

Oberhalb der Raumfrequenz $k_{\max}$ folgen wegen der k^4-Abhängigkeit der sphärischen Aberration in immer dichteren Abständen Maxima, Nullstellen und Minima der Aberrationsfunktion aufeinander. Diese Oszillation führt zu Kontrastabbildungsbedingungen, die nicht mehr mit der Forderung nach der Herstellung eines direkt interpretierbaren Sinneseindrucks vereinbar sind. Deshalb wird dieser Raumfrequenzbereich experimentell durch Blenden abgeschnitten, wenn nicht schon die in $T(\vec{k}, \vec{k}')$ enthaltenen Dämpfungsterme den Kontrasttransfer bei hohen Raumfrequenzen effektiv unterdrücken.

Zum Schluß wird noch einmal ausdrücklich betont, daß die Benutzung des Scherzer-Paßbands nur dann sinnvoll ist, wenn ein reines Phasenobjekt abgebildet werden soll. Gemischt amplituden- und phasenmodulierte Austrittswellenfunktionen erfüllen nämlich nicht mehr die Grundvoraussetzung, daß ihre ursprüngliche Phasendrehung, die ja durch $\chi(\vec{k})$ teilweise kompensiert wird, $\pi/2$ beträgt. Zweitens

ist die Bemühung um einen möglichst direkt interpretierbaren Sinneseindruck durch Ausnutzen der Tiefpaßcharakteristik des Kontrasttransfers an die Voraussetzung geknüpft, daß schon die Austrittswellenfunktion ein direkt interpretierbares Abbild der Strukturinformation ist. Beide Voraussetzungen sind in der Regel bei Proben, die Gegenstand einer materialwissenschaftlichen Untersuchung sind, selbst bei dünnen elektronenmikroskopischen Objekten, nicht gegeben.

7.3.4 Die Gestalt der Dämpfungsterme

In der Beschreibung der dynamischen Beugung hochenergetischer Elektronen wurden zwei grundlegende Annahmen gemacht, die an dieser Stelle kritisch beleuchtet werden müssen. Sie betreffen die Natur der einfallenden Welle, die bisher als perfekt eben und monochromatisch angenommen wurde.

Die in der Elektronenmikroskopie verwendeten Elektronenquellen, die auf dem Mechanismus der Glühemission, beziehungsweise der Feldemission basieren, senden ein Spektrum von Wellenlängen aus, dessen Verteilung durch die thermische Energieverteilung der austretenden Elektronen und durch die nicht perfekte Spannungsstabilität des Elektronenbeschleunigers gegeben ist.

Das Erreichen vernünftig kurzer Belichtungszeiten bei elektronenmikroskopischen Aufnahmen erfordert das Erzeugen einer bestimmten Elektronenintensität durch ein mehr oder weniger starkes Konvergieren des auf eine Probe treffenden Elektronenstrahls. Im Teilchenbild bedeutet es, daß die einzelnen Elektronenbahnen nicht mehr parallel verlaufen, wie es in der Modellvorstellung angenommen wurde, sondern sie bilden einen Kegel, dessen Spitze auf die Probe zeigt. Im Wellenbild bedeutet es die Überlagerung der Elektronenwellen, die nicht mehr wie in der Modellvorstellung eine gemeinsame Ausbreitungsrichtung haben, aus den vielen unterschiedlichen Richtungen, die im Konvergenzkegel enthalten sind. Bei modernen Feldemissionsquellen ist der resultierende Konvergenzwinkel des Kegels relativ klein, so daß die einfallenden Teilwellen insgesamt einer ebenen Welle sehr nahe kommen, jedoch bei Glühemissionsquellen kann der Konvergenzwinkel Werte annehmen, die keine Näherung durch eine ebene Welle zulassen.

Wenn die Verteilung der Wellenlängen nicht zu breit ist, und der Konvergenzwinkel der einfallenden Elektronen nicht zu groß ist, dann bleibt das Modell der dynamischen Beugung weiter in der bisher dargestellten Form gültig, so daß beide Einflüsse auf das Modell des Kontrasttransfers übergewälzt werden können.

Partielle zeitliche Kohärenz

Durch die Wellenlängenverteilung aller auf die Probe eingestrahlten Elektronenwellen, sind einzelne Wellen untereinander nicht mehr kohärent. Eine Ursache für die Wellenlängenverteilung ist, wie oben dargestellt, die zeitliche Spannungsschwankung des Elektronenbeschleunigers. Deshalb wird der Zustand der gesamten Elektronenwelle als *partiell zeitlich kohärent* bezeichnet. Zu dieser Bennung trägt bei, daß formal eine weitere zeitliche Instabilität eines Elektronenmikroskops einer Wellenlängenänderung zugeordnet werden kann. Der Objetivlinsenstrom ist nämlich ge-

nau wie die Beschleunigungsspannung nicht konstant, sondern ebenfalls eine zeitlich veränderliche Größe. Dadurch wird die Brechkraft der Linse geringfügig geändert, so daß sich auch die bildseitige Fokusebene verschiebt. In gewissen Grenzen ist diese Verschiebung einer chromatischen Aberration gleichzusetzen, denn auch Elektronen unterschiedlicher Wellenlängen werden in verschiedenen Bildebenen fokussiert.

Die drei Einflüsse der thermischen Energieverbreiterung der Elektronen, die Instabilität der Beschleunigungsspannung und die Linsenstromschwankungen können alle auf eine Defokusschwankung umgerechnet werden. In einer Näherung werden die Maxwellverteilung der Elektronenenergien E, und die Zeitabhängigkeiten der Beschleunigungsspannung V und des Linsenstroms I durch eine Gaußverteilung modelliert, und die drei unabhängigen Ereignisse mittels des Gaußschen Fehlerfortpflanzungsgesetzes zu einer Größe

$$\Delta = C_c \sqrt{\left(\frac{\Delta E}{E}\right)^2 + \left(\frac{\Delta V}{V}\right)^2 + \left(\frac{2\Delta I}{I}\right)^2} \tag{7.46}$$

zusammengefaßt. ΔE, ΔV und ΔI sind die Verteilungsbreiten der entsprechenden Größen, und C_c ist die Konstante der chromatischen Aberration.

Im Rahmen der Theorie der zeitlich partiellen Kohärenz ist die Defokusstreuung durch eine auf 1 normierte Verteilungsfunktion des Defokus Z,

$$f(Z) = \frac{1}{\sqrt{\pi}\Delta} \exp\left(-\frac{Z^2}{\Delta^2}\right), \tag{7.47}$$

zu berücksichtigen. Die grundlegende Idee ist dabei das inkohärente Aufsummieren von Teilbildern, die bei verschiedenen Defokuswerten entstanden sind.

Der Transmissionskreuzkoeffizient $T(\vec{k}, \vec{k}', Z)$, der wegen der Defokusabhängigkeit der Aberrationsfunktion $\chi(\vec{k}, Z)$ ebenfalls defokusabhängig ist, ist somit das Integral über die den beiden Raumfrequenzen $\vec{k}$ und $\vec{k}'$ zugeordenten Phasenfakoren und die Verteilungsbreite $f(Z)$:

$$T(\vec{k}, \vec{k}', Z) = \int \exp\left(-2\pi i \chi(\vec{k}, Z + Z')\right) \exp\left(2\pi i \chi(\vec{k}', Z + Z')\right) f(Z)\, \mathrm{d}Z'. \tag{7.48}$$

Partielle räumliche Kohärenz

Im einleitenden Abschnitt wurde dargelegt, daß der auf eine Probe eingestrahlte Elektronenstrahl eine kegelförmige Gestalt hat, und daß dieser Kegel im Wellenbild als eine kontinuierliche Überlagerung ebener Teilwellen angesehen werden kann. Die zugehörigen Wellenvektoren $\vec{K}$ haben eine Komponente $\vec{k}_0$ entlang der optischen Achse und wegen der gekippten Einstrahlung zusätzlich noch eine kleine Komponente $\vec{q}$ senkrecht zur optischen Achse. Die dynamische Beugung verläuft wegen $q \ll k_0$ in erster Näherung so, als wäre der Wellenvektor parallel zur optischen Achse. Deshalb muß zur inkohärenten Überlagerung der Teilbilder nicht die gesamte Berechnung der Austrittswellenfunktion für viele verschiedene Wellenvektoren $\vec{K} = \vec{k}_0 + \vec{q}$ der eingestrahlten ebenen Welle durchgeführt werden, sondern stattdessen werden Bilder der Austrittswellenfunktionen mit den verschiedenen Abbildungsvektoren $\vec{q}$

aufsummiert. Weil die inkohärente Überlagerung der Teilwellen mit der räumlichen Anordnung der Wellenvektoren zusammenhängt, wird der Zustand des gesamten Strahlenbündels als *partiell räumlich kohärent* bezeichnet.

Analog zum vorigen Abschnitt ist zuvor noch die Verteilungsdichte der Einstrahlrichtungen $\vec{q}$ zu bestimmen. In der Literatur werden zwei Ansätze diskutiert: Der erste besteht aus dem Modell einer scheibenförmigen, gleichmäßig hellen Quelle, denn in einer Blickrichtung von der Probe durch das Kondensorsystem eines Elektronenmikroskops zur Elektronenquelle erscheint diese wie das Bild der gleichmäßig hell ausgeleuchteten Kondensorblende.

Die resultierenden mathematischen Ausdrücke eines solchen Zylinderprofils erhalten wegen der steilen Kanten der Quellverteilung am Rand der Blende eine unbequeme Gestalt. Deshalb wird im Standardmodell der partiellen räumlichen Kohärenz eine mathematisch bequemere Quellverteilung verwendet, die aus einem gaußförmigen Strahlprofil besteht:

$$s(\vec{q}) = \frac{1}{\pi q_0^2} \exp\left(-\frac{q^2}{q_0^2}\right). \tag{7.49}$$

Die halbe 1/e-Breite des Strahlprofils, q_0, heißt *Semikonvergenzwinkel.*

Durch die inkohärente Überlagerung der Teilbilder wird die Integration in Gleichung (7.48) um eine Integration über die Einstrahlrichtungen ergänzt. Der Transmissionskreuzkoeffizient erhält nach Ersetzen der ursprünglichen Einstrahlrichtungen $\vec{k}$, $\vec{k}'$ durch $\vec{k}+\vec{q}$, $\vec{k}'+\vec{q}$ folgende Gestalt:

$$\begin{aligned} T(\vec{k},\vec{k}',Z) &= \int\int \exp\left(-2\pi i\chi(\vec{k}+\vec{q},Z+Z')\right) \exp\left(2\pi i\chi(\vec{k}'+\vec{q},Z+Z')\right) \times \\ &\quad \times s(\vec{q}) f(Z)\, \mathrm{d}\vec{q}\, \mathrm{d}Z'. \end{aligned} \tag{7.50}$$

Dies ist die vollständige Integraldarstellung des Transmissionskreuzkoeffizienten der Zweistahlinterferenz unter partieller räumlicher und zeitlicher Kohärenz.

Eine letzte Ergänzung muß noch gemacht werden für den Fall, daß bei einem Experiment eine Objektivblende benutzt wird, die einen Teil der hochfrequenten Raumfrequenzen abschneidet. Dies ist zum Beispiel sinnvoll, wenn die bei hohen Raumfrequenzen stark oszillierende Aberrationsfunktion die Aufnahme eines direkt interpretierbaren Abbildes einer Strukturinformation behindern würde (Abschnitt 7.3.3).

Die Aperturfunktion einer Kreisblende hat im Reziprokraum die Gestalt eines Zylinders mit dem Radius k_B und der Höhe 1, und wenn die Blende rotationssymmetrisch zur optischen Achse zentriert ist, lautet die Funktion:

$$a(\vec{k}) = \begin{cases} 1 & k < k_\mathrm{B} \\ 0 & k \geq k_\mathrm{B} \end{cases}. \tag{7.51}$$

In das Doppelintegral gehen die zu den Wellenvektoren $\vec{k}$ und $\vec{k}'$ gehörenden Aperturfunktionen als Produkte ein:

$$\begin{aligned} T(\vec{k},\vec{k}',Z) &= \int\int \exp\left(-2\pi i\chi(\vec{k}+\vec{q},Z+Z')\right) \exp\left(2\pi i\chi(\vec{k}'+\vec{q},Z+Z')\right) \times \\ &\quad \times a(\vec{k}+\vec{q}) a(\vec{k}'+\vec{q}) s(\vec{q}) f(Z)\, \mathrm{d}\vec{q}\, \mathrm{d}Z'. \end{aligned} \tag{7.52}$$

Produktdarstellung des Transmissionskreuzkoeffizienten

Es ist vom Standpunkt der numerischen Bildberechnung aus unbefriedigend, bei jeder Bildberechnung erneut das Doppelintegral (7.50) auszurechnen zu müssen. Aus diesem Grund wurden Näherungsformeln entwickelt, die auf der Taylorreihenentwicklung der in Gleichung (7.52) enthaltenen Aberrationsfunktion beruhen.

Der Rechengang beginnt mit der Entwicklung von $\chi(\vec{k}+\vec{q}, Z+Z')$ um die Stelle $(\vec{k}, Z)$. Das Integral kann für die dann auftretenden Summanden der Ableitungen niedrigster Ordnung nach der Koordinate $\vec{k}$ und der Koordinate Z und der gemischten Ableitung niedrigster Ordnung nach $\vec{k}$ und Z explizit unter Verwendung der Verteilungsfunktionen $f(Z)$ und $s(\vec{q})$ gelöst werden. Die beiden Phasenfaktoren der Strahlen $\vec{k}$ und $\vec{k}'$ hängen nicht mehr von den Integrationsvariablen $\vec{q}$ und Z' ab, und können vor das Integral gezogen werden.

Die zugehörige Rechnung ist langwierig, aber noch analytisch durchzuführen. Das Ergebnis ist die Produktdarstellung des Transmissionskreuzkoeffizienten:

$$\begin{aligned} T(\vec{k},\vec{k}',Z) &= \underbrace{\exp\left(-2\pi\mathrm{i}\chi(\vec{k},Z)\right)\exp\left(2\pi\mathrm{i}\chi(\vec{k}',Z)\right)a(\vec{k})a(\vec{k}')}_{\text{kohärent}} \times \\ &\quad \times \underbrace{E_{\mathrm{S}}(\vec{k},\vec{k}',Z)E_{\mathrm{F}}(\vec{k},\vec{k}')E_{\mathrm{X}}(\vec{k},\vec{k}')P(\vec{k},\vec{k}',Z)}_{\text{partiell kohärent}}. \end{aligned} \tag{7.53}$$

Der mit „kohärent" gekennzeichnete Anteil besteht aus den schon oben eingeführten Phasenfaktoren aufgrund der Aberration der Strahlen $\vec{k}$ und $\vec{k}'$, und den zugehörigen Aperturfunktionen $a(\vec{k})$ und $a(\vec{k}')$.

Die beiden Faktoren $E_{\mathrm{S}}(\vec{k},\vec{k}',Z)$ und $E_{\mathrm{F}}(\vec{k},\vec{k}')$ beinhalten den Einfluß der partiellen räumlichen (S) und zeitlichen (F) Kohärenz. Die Faktoren $E_{\mathrm{X}}(\vec{k},\vec{k}')$ und $P(\vec{k},\vec{k}',Z)$ sind sowohl von der Breite der Quellverteilung als auch von der Defokusstreuung abhängig, und sie repräsentieren eine Dämpfung und eine Phasenverschiebung durch den kombinierten Einfluß von partieller räumlicher und zeitlicher Kohärenz. Mit den bis hierhin eingeführten Konstanten lauten die Faktoren:

$$E_{\mathrm{S}}(\vec{k},\vec{k}',Z) = \exp\left(-\frac{(\pi q_0)^2}{u}(\chi_{\vec{k}}-\chi'_{\vec{k}})^2\right), \tag{7.54}$$

$$E_{\mathrm{F}}(\vec{k},\vec{k}') = \exp\left(-\frac{(\pi\Delta)^2}{u}(\chi_Z-\chi'_Z)^2\right), \tag{7.55}$$

$$E_{\mathrm{X}}(\vec{k},\vec{k}') = \frac{1}{\sqrt{u}}\exp\left(-\frac{(\pi^2 q_0^2\Delta)^2}{u}\left|(\chi_{\vec{k}}-\chi'_{\vec{k}})\times(\chi_{\vec{k}Z}-\chi'_{\vec{k}Z})\right|^2\right), \tag{7.56}$$

$$P(\vec{k},\vec{k}',Z) = \exp\left(-2\pi\mathrm{i}\frac{(\pi q_0\Delta)^2}{u}(\chi_Z-\chi'_Z)(\chi_{\vec{k}}-\chi'_{\vec{k}})\cdot(\chi_{\vec{k}Z}-\chi'_{\vec{k}Z})\right). \tag{7.57}$$

Die Abkürzungen bedeuten:

$$\chi_{\vec{k}} = \left.\frac{\partial\chi(\vec{k},Z)}{\partial\vec{k}}\right|_{\vec{k}} = \lambda\vec{k}(C_{\mathrm{s}}\lambda^2k^2+Z), \tag{7.58}$$

$$\chi_Z = \left.\frac{\partial\chi(\vec{k},Z)}{\partial Z}\right|_Z = \frac{\lambda}{2}k^2, \tag{7.59}$$

$$\chi_{\vec{k}Z} = \left.\frac{\partial^2\chi(\vec{k},Z)}{\partial\vec{k}\partial Z}\right|_{\vec{k},Z} = \lambda\vec{k} \tag{7.60}$$

und

$$u = 1 + (\pi q_0\Delta)^2(\chi_{\vec{k}Z} - \chi'_{\vec{k}Z})^2. \tag{7.61}$$

Es ist gezeigt worden, daß der Faktor $E_X(\vec{k},\vec{k}')$, der unter anderem von der gemischten Ableitung nach $\vec{k}$ und Z in der Taylorreihenentwicklung abhängt, gegenüber dem Einfluß von $E_S(\vec{k},\vec{k}',Z)$ und $E_F(\vec{k},\vec{k}')$ vernachlässigt werden kann [219]. Die Phasendrehung $P(\vec{k},\vec{k}',Z)$ ist ebenfalls zu vernachlässigen.

7.3.5 Verschlüsselung der Austrittswellenfunktion

In einer Rückschau auf Gleichung (7.40), die der Ausgangspunkt für die Entwicklung der Theorie der Kontrastübertragung unter partieller räumlicher und zeitlicher Kohärenz war, sind in Hinblick auf den Transfer einer Strukturinformation von der Austrittswellenfunktion in die Bildintensitäten folgende Eigenschaften des Kontrasttransfers festzuhalten:

- Die Kontrastübertragung ist nichtlinear, weil die Intensität einer Bildfrequenz vom Produkt zweier Fourierkomponenten der Austrittswellenfunktion abhängt.
- Die Integration über alle Fourierkomponenten der Austrittswellenfunktion verknüpft grundsätzlich alle Raumfrequenzbeiträge untereinander.
- Die Phasendrehung und die Dämpfung der einzelnen Beiträge führt zu einer hochkomplexen Verschlüsselung der Austrittswellenfunktion, weil der Transmissionskreuzkoeffizient von fünf Parametern, k_x, k_y, k'_x, k'_y und Z, abhängt, wenn man C_s, λ, Δ und q_0 als Abbildungskonstanten ansieht.

Aus dieser Aufzählung folgt, daß eine Bildinformation im allgemeinen nicht direkt interpretierbar bezüglich einer in ihr enthaltenen Strukturinformation ist.

Im folgenden wird ein Spezialfall diskutiert, bei welchem dennoch eine lineare Kontrastübertragung stattfindet, die unter einer weiteren Einschränkung sogar ein direkt interpretierbares Strukturabbild liefert.

Schwaches Objekt

Aus der komplizierten Gestalt des Transmissionskreuzkoeffizienten ist abzulesen, daß es trotz des experimentell beeinflußbaren Defokusparameters Z keine Möglichkeit gibt, in der Vielstrahlinterferenz einer Wellenfunktion mit mehreren Fourierkomponenten mittels einer geschickten Wahl von Z einzelne Kontrastbeiträge so auszuschalten, daß Gleichung (7.40) linear in $\Psi(\vec{g})$ wird. Daraus folgt, daß die Austrittswellenfunktion selbst eine spezielle Gestalt haben muß.

Dies ist der Fall, wenn die dynamische Beugung der hochenergetischen Elektronen durch das atomare Kristallpotential so schwach gewesen ist, daß die Amplitude des Nullstrahls $\Psi(\vec{k}=0)$ groß gegenüber den Amplituden der abgebeugten Strahlen $\Psi(\vec{k})$ ist. Dann können in Gleichung (7.40) alle Interferenzen zwischen abgebeugten Strahlen vernachlässigt werden, weil sie von zweiter Ordnung der kleinen Amplituden sind, so daß nur die Interferenzterme zwischen Nullstrahl und abgebeugten Strahlen übrigbleiben:

$$I(0) \approx |\Psi(0)|^2 \tag{7.62}$$

und

$$I(\vec{g}) \approx \Psi(0)\,\Psi^*(-\vec{g})\,T(0,-\vec{g},Z) + \Psi(\vec{g})\,\Psi(0)\,T(\vec{g},0,Z) \qquad \vec{g}\neq 0, \tag{7.63}$$

mit der Normierung der Nullstrahlphase auf den Wert Null. Die Beugung an einem *schwachen Objekt* resultiert in einer Bildintensität $I(\vec{g})$, die linear in $\Psi(\vec{g})$ und $\Psi^*(-\vec{g})$ ist. Die Funktion $T(\vec{g},0,Z)$ wird deshalb *lineare Kontrasttransferfunktion* genannt.

Schwaches Phasenobjekt

In Abschnitt 7.1.2 wurde gezeigt, daß bei der Beugung an einem schwachen Phasenobjekt die Strukturinformation direkt und linear in den Fourierkomponenten der Austrittswellenfunktion $\Psi(\vec{g})$ enthalten ist. Die abgebeugten Strahlen haben dabei eine Phasendrehung von $\pi/2$ gegenüber der korrespondierenden Fourierkomponente des atomaren Kristallpotentials, $U_{\vec{g}}$, erhalten (Gleichung (7.32)). Aus Gleichung (7.63) ist abzulesen, daß im Fall des schwachen Phasenobjekts allein der Imaginärteil der linearen Kontrasttransferfunktion für die zusätzliche Phasendrehung verantwortlich ist, die zur Transformation des Phasenkontrastes in einen Amplitudenkontrast notwendig ist (Abschnitt 7.3.3). Der Imaginärteil ist proportional zu $\sin(\chi(\vec{g},Z))$ und die optimale Defokuseinstellung ist ebenfalls schon ausführlich in Abschnitt 7.3.3 dargestellt.

An dieser Stelle ist mit dem zusätzlichen Wissen über die Form der Dämpfungsfaktoren eine Aussage über die Hauptursache der Dämpfung sehr hoher Raumfrequenzen beim linearen Kontrasttransfer möglich. Durch Einsetzen von $\vec{k}=\vec{g}$ und $\vec{k}'=0$ in die Ausdrücke der dominierenden Dämpfungsfaktoren $E_{\mathrm{S}}(\vec{k},\vec{k}',Z)$ und $E_{\mathrm{F}}(\vec{k},\vec{k}')$ sieht man, daß in den Vorfaktor der höchsten vorkommenden Potenz von $\vec{k}$ die Konstante der sphärischen Aberration C_{s} eingeht.

Eine Verkleinerung von C_{s} hat also den Effekt, daß die Dämpfung hoher Raumfrequenzen reduziert wird. Gleichzeitig verschiebt sich auch die obere Grenze des Scherzer-Paßbandes zu größeren Raumfrequenzen hin (Gleichung (7.45)), so daß auch die Grenzauflösung des Phasenkontrastes einer direkt interpretierbaren Strukturabbildung verbessert wird. Dieser Zusammenhang ist der Anlaß verstärkter Bemühungen um eine Reduktion der sphärischen Aberration durch die Konzeption und Verwirklichung eines fortschrittlichen Linsendesigns im Rahmen eines von der Volkswagenstiftung geförderten Projekts.

7.4 Zusammenfassung

Die Verschlüsselung einer Strukurinformation durch die dynamische Beugung hochenergetischer Elektronen am atomaren Kristallpotential und die anschließende Abbildung der Austrittswellenfunktion durch den nichtlinearen Kontrasttransfer unter partiell kohärenter Beleuchtung resultiert im allgemeinen in einem komplexen Interferogramm, das eine für einen menschlichen Betrachter nicht direkt interpretierbare Intensitätsverteilung enthält.

Die Nichtlinearität beider Verschlüsselungsprozesse ist in den beiden Grundgleichungen der mathematischen Formulierung der dynamischen Beugung (Gleichung (7.30)) und des Kontrasttransfers (Gleichung (7.40)) einzusehen. Es existieren in der Gestalt des schwachen Objekts, beziehungsweise des schwachen Phasenobjekts Beugungsbedingungen, die einen anschließenden nichtlinearen Kontrasttransfer zum Beispiel durch Ausnutzen des Scherzer-Paßbandes ermöglichen. In einer Vielzahl von Fällen, besonders bei der Beugung an Materialien, die Gegenstand einer materialwissenschaftlichen Untersuchung sind, gelten die grundlegenden Näherungen nur für unrealistisch kleine Probendicken.

Solange die durch das Elektronenmikroskop gelieferte Primärinformation in der Form beobachtbarer Intensitätsverteilungen nicht direkt interpretierbar ist, müssen mehr oder weniger aufwendige Bildinterpretations- und Strukturermittlungsverfahren nachgeschaltet werden. Das folgende Kapitel beschreibt eine Reihe solcher Verfahren, die entweder auf der quantenmechanisch-optischen Bildsimulation der dynamischen Beugung und der elektronenoptischen Abbildung beruhen, oder die in einem zweistufigen Prozeß zuerst die Austrittswellenfunktion aus elektronenmikroskopischen Bildern rekonstruieren und dann aus der Austrittswellenfunktion die projizierte Struktur rekonstruieren.

Kapitel 8

Existierende Strukturermittlungsverfahren

8.1 Konventioneller Bildvergleich

Die naheliegendste und immer noch gebräuchlichste Methode der Strukturermittlung ist die Simulation der dynamischen Beugung und der elektronenoptischen Abbildung auf der Grundlage der im vorangegangenen Kapitel dargestellten Berechnungsverfahren zur Bestimmung der Austrittswellenfunktion aus dem atomaren Kristallpotential und zur Bestimmung der Bildintensität [243, 248, 234, 220]. Dazu muß erstens ein Strukturmodell zugrundegelegt werden, aus dem das atomare Kristallpotential berechnet wird, und zweitens muß ein Satz von Beugungs- und Abbildungsparametern vorgegeben werden, der unter anderem aus der Wellenlänge der Elektronen, der Probendicke, dem Objektivlinsendefokus, der Konstanten der sphärischen Aberration, der Semikonvergenz des einfallenden Strahlenbündels und der Defokusstreuung besteht. Die Simulation hat das Ziel, sowohl ein Strukturmodell als auch einen Satz von Beugungs- und Abbildungsparametern zu finden, so daß die experimentell gemessene Intensitätsverteilung möglichst genau durch die simulierte Intensitätsverteilung wiedergegeben wird.

Einige der Beugungs- und Abbildungsparameter können mit ausreichender Genauigkeit gemessen werden, wie zum Beispiel die Konstante der sphärischen Aberration, andere Parameter, wie zum Beispiel die Probendicke, müssen jedoch schlicht „durchprobiert" werden. Das Bestehen dieser Unsicherheit deutet bereits an, daß ein gewisser „Suchaufwand" getrieben werden muß, um einen „passenden" Satz von Parametern zu finden.

Bedeutend größer ist der Suchaufwand für ein passendes Strukturmodell, denn wenn nicht schon eine konkrete Vermutung zum Beispiel über eine vorliegende Defektstruktur existiert, dann hat man wegen der Vielzahl der Atome, aus denen ein Streuobjekt bestehen kann, einschließlich der jeweiligen Ortskoordinaten eine sehr große Anzahl von durchzuprobierenden Freiheitsgraden. Daraus ist schon ersichtlich, daß eine Strukturbestimmung mittels Durchprobierens einer Anzahl von Strukturmodellen abhängig von der Größe des Problems an eine Grenze stoßen muß.

Eine zweite Unzulänglichkeit dieser konventionellen Strukturermittlung besteht

in der Methode, wie die Änhlichkeit zwischen simulierten und experimentell gemessenen Bildintensitäten beurteilt wird. Es ist nämlich gängige Praxis, den Bildvergleich „per Augenschein" durchzuführen, also die Bewertung einem subjektiven Beobachter zu überlassen.

Erst in den letzten Jahren gibt es vermehrt Anstrengungen, den Bildvergleich objektiv auf der Basis numerischer Kriterien durchzuführen. Diese Entwicklung wurde nicht zuletzt durch den Preisverfall von Off-line-Digitalisierungsgeräten gestützt, die das ausschnittweise Digitalisieren von Photoplatten ermöglichen, sondern auch durch den vermehrten On-line-Einsatz von CCD-Kameras *(charge coupled devices)* anstelle von Photoplatten. Die digitalisierte Bildinformation wird zusammen mit simulierten Intensitätsverteilungen in einen Digitalrechner geladen, und dort per Programm verglichen.

8.2 Quantitativer Bildvergleich

Der quantitative Bildvergleich beruht auf dem Prinzip, die zu vergleichende Bildinformation, die aus zwei zweidimensionalen Datensätzen besteht, auf einen einzigen, aussagekräftigen Zahlenwert zu reduzieren. Während der jahrtzehntelang andauernden und immer noch fortschreitenden Entwicklung der numerischen Bildverarbeitung wurde eine sehr große Anzahl von Vergleichskriterien erfunden, deren mathematische Gestalt in der Regel auf die zu erwartende Form einer Bildinformation zugeschnitten sind. Ein erschöpfender Überblick ist an dieser Stelle weder sinnvoll noch möglich, stattdessen werden im folgenden zwei grundlegende und verbreitete Vergleichskriterien diskutiert, die in den in dieser Arbeit untersuchten Rekonstruktionsalgorithmen verwendet wurden.

8.2.1 Kleinstquadrat-Kriterium

Ein sehr altes, auf den Prinzipien der mathematischen Statistik gründendes Kriterium ist das Gaußsche Kleinstquadrat-Kriterium, welches dem Ausgleichungsproblem zwischen simulierten und experimentell gemessenen Daten gerecht wird, wenn die gemessenen Daten statistische Meßfehler aufweisen, und wenn es eine Anzahl „überschüssiger" Messungen gibt. Damit ist gemeint, daß die Zahl der statistisch unabhängigen Meßwerte die Zahl der zu bestimmenden Parameter überschreitet, von denen die simulierten Daten abhängen.

Der zentral wichtige Begriff in der Methode der kleinsten Quadrate ist das *Residuum*, welches auf den numerischen Bildvergleich übertragen die Differenz zwischen einer simulierten und einem experimentell gemessenen Bildintensität an der Stelle $\vec{r}_i$ ist:

$$R_i(\vec{r}) = I_{\text{sim}}(\vec{r}_i) - I_{\text{exp}}(\vec{r}_i) \tag{8.1}$$

$\vec{r}_i$ ist der zweidimensionale Vektor der Bildkoordinaten.

Das aus den Residuen abgeleitete Vergleichskriterium ist die Summe aller Qua-

drate der Residuen:

$$R = \sum_i R_i^2 = \sum_i (I_{\text{sim}}(\vec{r}_i) - I_{\text{exp}}(\vec{r}_i))^2. \tag{8.2}$$

Der Bildvergleich auf der Basis dieses Kriteriums ist natürlich nur sinnvoll, wenn die Bildintensität des simulierten und des experimentell gemessenen Bildes jeweils auf 1 normiert ist:

$$\sum_i I_{\text{sim}}^2(\vec{r}_i) = 1 \tag{8.3}$$

und

$$\sum_i I_{\text{exp}}^2(\vec{r}_i) = 1. \tag{8.4}$$

Das in Kapitel 7 dargestellte Modell der Kontrastübertragung unter partiell kohärenter Beleuchtung ist durchgehend im Reziprokraum formuliert (Gleichung (7.40)), und die Bildinformation eines simulierten Bildes liegt zunächst in der Form der Fourierkomponenten $I_{\text{sim}}(\vec{g})$ im Reziprokraum vor. Die experimentell gemessene Bildinformation ist jedoch ein Satz von Realraumintensitäten $I_{\text{exp}}(\vec{r})$. Vor der Anwendung des Kleinstquadrat-Kriteriums (8.2) ist deshalb das simulierte Reziprokraumbild in den Realraum zu transformieren.

Genausogut kann man aber auch das experimentell gemessene Bild in den Reziprokraum transformieren und den numerischen Bildvergleich an Hand der Fourierkomponenten $I_{\text{sim}}(\vec{g})$ und $I_{\text{exp}}(\vec{g})$ durchführen. Das zum Realraumkriterium (8.2) analoge Kriterium lautet im Reziprokraum:

$$R = \sum_{\vec{g}} R_{\vec{g}}^2 = \sum_{\vec{g}} |I_{\text{sim}}(\vec{g}) - I_{\text{exp}}(\vec{g})|^2, \tag{8.5}$$

mit den Normierungen

$$\sum_{\vec{g}} |I_{\text{sim}}(\vec{g})|^2 = 1 \tag{8.6}$$

und

$$\sum_{\vec{g}} |I_{\text{exp}}(\vec{g})|^2 = 1. \tag{8.7}$$

Für beide Arten des Kleinstquadrat-Kriteriums gelten die folgenden qualitativen Bewertungen: Je kleiner der Wert R ist, desto besser stimmen die beiden verglichenen Bilder überein. Der Wert $R = 0$ bedeutet eine perfekte Übereinstimmung beider Bilder.

8.2.2 Kreuzkorrelationskoeffizient

Die Korrelation ist ein ebenfalls aus der mathematischen Statistik stammender Begriff, der die stochastische Abhängigkeit zweier Zufallsgrößen beschreibt. Die mit der Korrelation verbundene Bewertungsgröße, der Kreuzkorrelationskoeffizient, hat die mathematische Gestalt eines normierten Skalarprodukts der beiden korrelierten Größen.

Für den Bildvergleich zweier Bildintensitäten $I_{\text{sim}}(\vec{r}_i)$ und $I_{\text{exp}}(\vec{r}_i)$ an den Orten $\vec{r}_i$ lautet der Kreuzkorrelationskoeffizient:

$$R = \frac{\sum_i I_{\text{sim}}(\vec{r}_i) I_{\text{exp}}(\vec{r}_i)}{\sqrt{\sum_i I_{\text{sim}}^2(\vec{r}_i)} \sqrt{\sum_i I_{\text{exp}}^2(\vec{r}_i)}}. \tag{8.8}$$

Eine anschauliche Interpretation dieser Größe wird möglich, wenn man die Intensitätsverteilung $I(\vec{r}_i)$ nicht mehr als zweidimensionales Bild ansieht, sondern die einzelnen Zahlenwerte in einem eindimensionalen Vektor anordnet, dessen i-te Komponente den Wert $I(\vec{r}_i)$ hat. Dann ist der Zähler in Gleichung (8.8) das Skalarprodukt der betreffenden „Intensitätsvektoren" und der Nenner enthält die Normierung der Vektoren auf den Wert 1.

Mit dieser Interpretation ist der Wert der Kreuzkorrelation, R, als Kosinus des Winkels zwischen diesen beiden abstrakten Vektoren anzusehen. Deshalb wird häufig statt des Wertes R der Wert

$$\alpha = \arccos R \tag{8.9}$$

benutzt [236]. Die Verwendung des Winkels hat neben der Anschaulichkeit zusätzlich den Vorteil, bei einer großen Korrelation, wo R Werte nahe 1 annimmt, und sich Korrelationsunterschiede nur durch die Angabe vieler Nachkommastellen ausdrücken lassen, eine besser differenzierende „Größenskala" zu liefern. Der Grund dafür ist die Steigung Null der Kosinusfunktion für den Wert $R = 1$.

Analog zu der Situation beim Kleinstquadrat-Kriterium (Abschnitt 8.2.1) ist eine vollständig äquivalente Formulierung im Reziprokraum möglich:

$$R = \frac{\sum_{\vec{g}} I_{\text{sim}}(\vec{g}) I_{\text{exp}}^*(\vec{g})}{\sqrt{\sum_{\vec{g}} |I_{\text{sim}}(\vec{g})|^2} \sqrt{\sum_{\vec{g}} |I_{\text{exp}}(\vec{g})|^2}}. \tag{8.10}$$

Für beide Arten des Kreuzkorrelationskriteriums gelten die folgenden qualitativen Bewertungen: Je größer der Wert R und je kleiner der Winkel α ist, desto besser stimmen die beiden verglichenen Bilder überein. Die Werte $R = 1$ und $\alpha = 0$ bedeuten eine perfekte Übereinstimmung beider Bilder.

8.2.3 Vergleichskriterien beim schwachen Objekt

Neben den in den vorhergehenden Abschnitten 8.2.1 und 8.2.2 dargestellten Versionen des Kleinstquadrat-Kriteriums und des Kreuzkorrelationskoeffizienten gibt es noch zwei wichtige abgeleitete Bewertungsgrößen, die besonders vorteilhaft eingesetzt werden können, wenn die experimentell gemessene Bildintensität ihren Ursprung in der Beugung an einem schwachen Objekt oder an einem schwachen Phasenobjekt hat (Abschnitte 7.3.5 und 7.3.5).

In einem solchen Fall ist wegen der praktisch linearen Kontrastübertragung nicht nur die Amplitude des transmittierten Strahls der Austrittswellenfunktion viel größer als die Amplituden der abgebeugten Strahlen, sondern dann dominiert auch die Nullstrahlintensität $I(\vec{g} = 0)$ des Bildes die Intensität der anderen Bildfrequenzen. Die Werte R in den Reziprokraumversionen (8.5) und (8.5) des Kleinstquadrat-

und des Kreuzkorrelationskriteriums werden dann in erster Linie durch die Nullstrahlintensitäten $I_{\text{sim}}(0)$ und $I_{\text{exp}}(0)$ der Bilder bestimmt, und die Werte R in den Realraumversionen (8.2) und (8.8) des Kleinstquadrat- und des Kreuzkorrelationskriteriums werden in erster Linie durch die Intensitätsmittelwerte $\bar{I}_{\text{sim}}$ und $\bar{I}_{\text{exp}}$ der Bilder bestimmt.

Mit anderen Worten, die genannten Kriterien sind dann weniger sensitiv auf die Werte der Bildfrequenzen $I(\vec{g} \neq 0)$, also den Bildkontrast. Es ist dann aus numerischen Gründen günstig, in den Realraumkriterien statt der Intensitäten $I_{\text{sim}}(\vec{r}_i)$, $I_{\text{exp}}(\vec{r}_i)$ die Werte $I_{\text{sim}}(\vec{r}_i) - \bar{I}_{\text{sim}}$, $I_{\text{exp}}(\vec{r}_i) - \bar{I}_{\text{exp}}$ einzusetzen, also die Mittelwerte vorher abzuziehen. In den Reziprokraumkriterien wird dann einfach nur über die Vektoren $\vec{g} \neq 0$ summiert.

Das Ausklammern der Intensitätsmittelwerte führt nicht zu einer Mehrdeutigkeit der Vergleichskriterien, in dem Sinne, daß jetzt zwei Bilder als gleich angesehen werden, die zwar in ihren Bildfrequenzen übereinstimmen, jedoch unterschiedliche Mittelwerte haben. Denn die Mittelwerte sind wegen der Normierungen (8.3), (8.4), (8.6) und (8.7) keine statistisch unabhängigen Größen.

Die Kreuzkorrelation nach Abzug der Intensitätsmittelwerte, beziehungsweise nach dem Ausklammern der Nullstrahlintensität, hat in der mathematischen Statistik eine eigenständige Bedeutung, sie heißt dann *Kreuzkovarianz.*

8.3 Automatisierter Bildvergleich

8.3.1 Bestimmung der Beugungs- und Abbildungsparameter

Ein wesentlicher Bestandteil einer erfolgreichen Strukturermittlung ist die relativ genaue Bestimmung der Beugungs- und Abbildungsparameter (Abschnitt 8.1). Besonders wichtig ist die Bestimmung der Parameter Probendicke und Defokus, die empfindliche Einflußgrößen in den Modellen der dynamischen Beugung und der elektronenoptischen Abbildung sind.

Im allgemeinen ist es äußerst schwierig, Probendicke und Defokus aus Abbildungen einer noch unbekannten Defektstruktur zu bestimmen. Viele Defekte kristalliner Materialien haben jedoch im direkt benachbarten Kristallvolumen genügend große Bereiche, die als perfekt kristallin angesehen werden können. Weil die perfekte Kristallstruktur der meisten Materialien bekannt ist, lassen sich Probendicke t und Defokus Δf im Prinzip einfach mittels Durchprobieren verschiedener Paare $(t, \Delta f)$ bei der Simulation der dynamischen Beugung und der elektronenoptischen Abbildung und mittels eines anschließenden Bildvergleichs ermitteln. Die „passende" Probendicke und der „passende" Defokus sind dann durch das Bild bestimmt, welches die beste Übereinstimmung mit einem experimentellen Bild zeigt.

Kleine Veränderungen der Probendicke schon im Bereich weniger Nanometer können die zugehörige Austrittswellenfunktion erheblich beeinflussen, und Defokusänderungen in der Größenordnung von 10 nm können die Abbildung einer Austrittswellenfunktion drastisch ändern. Daraus folgt, daß für eine erfolgreiche Suche

das zu durchsuchende Dickenintervall $[t_{\min}, t_{\max}]$ und das zu durchsuchende Defokusintervall $[\Delta f_{\min}, \Delta f_{\max}]$ in genügend dichten Schritten Δt, $\Delta(\Delta f)$ abgetastet werden müssen. Die daraus resultierende Zahl von Defokusschritten

$$N_{\mathrm{F}} = \frac{\Delta f_{\max} - \Delta f_{\min}}{\Delta(\Delta f)} \tag{8.11}$$

und Dickenschritten

$$N_{\mathrm{T}} = \frac{t_{\max} - t_{\min}}{\Delta t} \tag{8.12}$$

ergibt eine Zahl von auszuführenden Bildberechnungen von

$$N_{\mathrm{B}} = N_{\mathrm{F}} \times N_{\mathrm{T}}. \tag{8.13}$$

Die Zahl der Bildberechnungen kann in Hinblick auf eine erfolgreiche Suche so groß werden, daß ein Durchprobieren mittels manueller Eingabe der Parameterpaare $(t, \Delta f)$ in Standardsimulationspaketen zu zeitraubend wird. Aus diesem Grund wurden automatisierte Durchmusterungs- und Bildvergleichsprogramme entwickelt, die auf quantitative numerische Bildvergleichskriterien zurückgreifen [246].

Es ist zu bemerken, daß in der automatisierten Defokus- und Probendickensuche nicht etwa N_{B}-mal die dynamische Beugung *und* die elektronenoptische Abbildung simuliert werden müssen, denn für einen konstanten Defokus ist nur die Probendicke durchzuprobieren, und für eine konstante Probendicke der Defokus. Zweckmäßigerweise wird zuerst die dynamische Beugung mit Hilfe des Multislice-Algorithmus für die erforderlichen Probendicken simuliert (Abschnitt 7.2), und die resultierenden Austrittswellenfunktionen abgespeichert. Für einen konstanten Defokus braucht man anschließend nur einmal den Satz der erforderlichen Transmissionskreuzkoeffizienten für die relevanten Beugungsvektoren zu berechnen, wenn man dann sukzessive auf die abgespeicherten Austrittswellenfunktionen zurückgreift, und dann die Bildberechnung (Gleichung (7.40)) und den Bildvergleich durchführt (Gleichung (8.10)).

8.3.2 Strukturermittlung

Durch den Einsatz der automatisierten Suche nach den Beugungs- und Abbildungsparametern kann die Strukturermittlung der ursprünglichen Defektstruktur mit reduziertem Aufwand durchgeführt werden.

Im ersten Abschnitt dieses Kapitels wurde geschildert, daß die konventionelle Strukturermittlung, wie vorher die Bestimmug der Beugungs- und Abbildungsparameter, auf das Durchprobieren einer möglichst großen Anzahl atomarer Strukturmodelle hinausläuft. In jüngster Zeit wurde diese Strukturermittlung ebenfalls automatisiert [215, 233], wobei die entsprechenden Suchalgorithmen wegen des großen Berechnungsaufwands auf das Unterproblem der *Strukturverfeinerung* spezialisiert wurden. Eine Strukturverfeinerung betrifft nicht die Suche nach einer völlig unbekannten Objektstruktur, sondern sie geht von einer Schätzung der Objektstruktur aus, die zum Beispiel aus theoretischen Vorstellungen oder aus anderen experimentellen Untersuchungen gewonnen wurde.

In einem Verfahren auf der Basis eines Multislice-Algorithmus besteht der Rechenaufwand pro zu bewertender „Test"-Struktur aus der sukzessiven Berechnung des atomaren Kristallpotentials und der Transmissionsfunktion (Gleichung (7.37)), einer vollen Multislice-Iteration (Gleichung (7.35)), einer Bildberechnung (Gleichung (7.40)) und einem Bildvergleich (Abschnitt 8.2). Die Optimierung der vorgegebenen Objektstruktur wird danach beurteilt, ob die Teststruktur, die eine leicht veränderte Variante der Ausgangsstruktur ist, in einer Bildintensität resultiert, die auf der Basis des benutzten Bildvergleichskriteriums eine bessere Übereinstimmung mit einem experimentellen Bild zeigt. Die Optimierung der Teststruktur wird durch Wiederholen des gesamten Zyklus solange fortgesetzt, bis kein weiterer Fortschritt in der Übereinstimmung zwischen experimentellem und simuliertem Bild mehr möglich ist.

8.4 Rekonstruktionsmethoden

Der konventionelle Bildvergleich und die daraus abgeleitete automatisierte, quantitative Strukturermittlung basieren auf der numerischen Nachbildung der die Strukturinformation verschlüsselnden Prozesse, der dynamischen Beugung und der elektronenoptischen Abbildung. Parallel zur Entwicklung hochauflösenden Transmissionselektronenmikroskopie wurden zunehmend Anstregungen unternommen, diese relativ zeitraubende Prozedur durch sogenannte *Rekonstruktionsmethoden* zu umgehen.

Rekonstruktionsmethoden haben das Ziel, in einem ersten Schritt durch eine Inversion des Abbildungsprozesses die Austrittswellenfunktion direkt aus der Bildinformation zurückzugewinnen *(Phasenrekonstruktion)*, und dann in einem zweiten Schritt das atomare Kristallpotential durch Lösen des inversen Streuproblems direkt aus der zuvor rekonstruierten Austrittswellenfunktion zu ermitteln *(Potentialrekonstruktion)*. Aus den Grundgleichungen der dynamischen Beugung (Gleichung (7.30)) und des nichtlinearen Kontrasttransfers (Gleichung (7.40)) ist abzulesen, daß ein Invertieren beider Gleichungen im allgemeinen Fall eines stark streuenden Objekts und einer nichtlinearen Abbildung unter partiell kohärenter Beleuchtung nicht direkt möglich ist.

8.4.1 Phasenrekonstruktionsverfahren

Lineare Rekonstruktion

Eine wesentliche Schwierigkeit der direkten Inversion des Abbildungsprozesses kann unter der Voraussetzung umgangen werden, daß die Austrittswellenfunktion das Ergebnis der Beugung an einem schwachen Objekt oder einem schwachen Phasenobjekt ist. Dann können die nichtlinearen gegenüber den linearen Kontrastbeiträgen vernachlässigt werden, und die Bildintensität repräsentiert ein direktes Abbild des projizierten Potentials der Objektstruktur, gefaltet mit der Realraum-Transferfunktion. Eine entsprechende Entfaltung, die günstigerweise im Reziprokraum durch Division der fouriertransformierten Bildintensität durch die Transferfunktion ausgeführt wird, ergibt dann wieder ein Bild des projizierten Potentials der Objektstruktur.

Lücken in der Transferfunktion können durch die Auswertung zusätzlicher Bilder bei anderen Defokussierungen aufgefüllt werden. Lineare Rekonstruktionsmethoden [239, 229, 217] sind in der Praxis selten anzuwenden, weil für physikalisch interessante und für die Festkörpertechnologie wichtige Materialien in der weit überwiegenden Zahl die Näherung durch ein schwaches Phasenobjekt nicht gültig ist.

Nichtlineare Rekonstruktion

Im Fall stark streuender Objekte ist die Strukturinformation nicht alleine in der lokalen Phase der Austrittswellenfunktion, sondern auch in ihrer lokalen Amplitude enthalten. Zusätzlich kann die Amplitude der abgebeugten Strahlen vergleichbar groß zur Amplitude des transmittierten Strahls werden. Dann sind bei der elektronenmikroskopischen Abbildung nichtlineare Kontrastbeiträge – die Interferenzen zwischen abgebeugten Strahlen – nicht mehr zu vernachlässigen. Nichtlineare Rekonstruktionsmethoden berechnen die Austrittswellenfunktion aus mehreren Bildern [205, 231, 210, 218, 221, 223, 212, 201, 249, 250, 247] einer *Defokusserie*.

Das Aufnehmen einer solchen Bilderserie ist grundsätzlich nötig, weil es wegen der Komplexität der beiden Verschlüsselungsprozesse, der dynamischen Beugung und der elektronenoptischen Abbildung, a priori nicht klar ist, welche Bildkontrastdetails durch den nichtlinearen Kontrasttransfer verursacht sind. Durch systematisches Variieren des Defokus Z in einer Bilderserie wird die Aberrationsfunktion (7.43) geändert, so daß bei unveränderter Austrittswellenfunktion $\psi(\vec{r})$ in den Bildern neben einer Information über die Fourierkomponenten $\Psi(\vec{g})$ auch noch ein genügend großer Informationsgehalt über die Transmissionskreuzkoeffizienten $T(\vec{k}, \vec{k}', Z)$ vorliegt.

Eine nichtlineare Rekonstruktionsmethode hat also zwei Aufgaben zu lösen: Erstens, das „Entfernen" der nichtlinearen Kontrasttransferfunktion, die in der Gestalt der Transmissionskreuzkoeffizienten in den Produkten (7.40) enthalten ist, und zweitens die „Entflechtung" der übrigbleibenden Interferenzterme $\Psi(\vec{g}+\vec{h})\,\Psi^*(\vec{h})$.

In frühen Arbeiten wurde gezeigt, daß mit nichtlinearen Rekonstruktionsalgorithmen, die Information aus zwei Bildern mit verschiedenem Defokus [205, 231] oder aus einem Bild und einem Beugungsbild [210, 218] entnehmen, eine Berechnung der Austrittswellenfunktion eines stark streuenden Objektes möglich ist. Jedoch blieb zu diesem Zeitpunkt der Einfluß der partiell kohärenten Beleuchtung unberücksichtigt, der der wichtigste auflösungsbegrenzende Faktor in der hochauflösenden Mikroskopie ist. In den vergangenen Jahren wurden aufgrund der immens gewachsenen Leistungsfähigkeit von Digitalrechnern die Realisierung anspruchsvollerer Rekonstruktionstechniken möglich, die alle den Einfluß der partiell kohärenten Beleuchtung einschließen.

In einer von Kirkland [221, 223, 222] eingeführten rekursiven Methode wird eine Serie von Bildern ausgewertet, die bei systematisch variiertem Defokus der Objektivlinse aufgenommen wurden. Nach einer ersten Schätzung der Austrittswellenfunktion, zum Beispiel aus der Näherung des schwachen Phasenobjektes, werden die aus der geschätzten Austrittswellenfunktion berechneten Bilder mit den experimentellen Bildern verglichen. Anschließend wird unter Berücksichtigung des Unterschiedes

zwischen den experimentellen und den simulierten Bildern eine Korrektur zur ersten Schätzung der Wellenfunktion berechnet. Die Sequenz wird solange wiederholt, bis eine Übereinstimmung zwischen simulierten und experimentellen Bildern erreicht ist. Später wurde gezeigt [212], daß die Rekonstruktionsmethode versagt, wenn die erste Schätzung der Austrittswellenfunktion zu weit von der wahren Lösung entfernt ist, das heißt in dem Fall, wo die Näherung des schwachen Phasenobjektes nicht mehr gültig ist. In der jüngsten Vergangenheit wurde eine Modifikation des Rekonstruktionsalgorithmus für den Fall hochkohärenter Beleuchtung mit Feldemissionsquellen veröffentlicht [201], der eine sehr schnelle Bildberechnung auf der Basis einer quasi-kohärenten Näherung der Bildentstehung enthält.

In einer von Van Dyck und Mitarbeitern [249, 250] entwickelten Rekonstruktionsmethode wird eine Defokusserie mit engen Defokusabständen ausgewertet mit der Idee, die linearen von den nichtlinearen Bildkontrastbeiträgen zu separieren. Dazu werden die Bilder der Defokusserie bezüglich der Bildkoordinaten $\vec{r}$ und der Fokuskoordinate Z fouriertransformiert. Im dreidimensionalen reziproken Raum, der durch die konjugierten Variablen $\vec{g}$ und ξ aufgespannt wird, liegen dann die linearen Kontrastbeiträge im Idealfall auf einem Paraboloid, während die nichtlinearen Kontrastbeiträge weiter über den reziproken Raum verstreut sind. Die linearen Kontrastbeiträge können dann separiert und zu einer relativ einfachen Bestimmung der Austrittswellenfunktion verwendet werden. Mittels aus dieser Austrittswellenfunktion berechneter Bilder wird anschließend ein Vergleich zwischen simulierten und experimentellen Bildern durchgeführt. Die gesamte Sequenz wird wiederholt, wobei in jeder Iteration eine Korrektur zur berechneten Wellenfunktion ermittelt wird, die aus dem Unterschied zwischen den experimentellen und den simulierten Bildern hervorgeht. Der Erfolg des Verfahrens hängt davon ab, wie groß der Anteil linearer Bildinformation ist, und wie weit nichtlineare Anteile während der Iteration unterdrückt werden können. Dabei werden die nichtlinearen Kontrastbeiträge als nichtkorreliertes „Rauschen" angesehen, und die Verwendung einer Defokusserie aus N Bildern führt dann zu einer Verbesserung des Signal-zu-Rausch-Verhältnisses um den Faktor $\sqrt{N}$.

Es bleibt abschließend festzustellen, daß die beschriebenen Phasenrekonstruktionsverfahren auf dem Ansatz der linearen Bildkontrastentstehung beruhen, oder daß sie eine erste Schätzung der Austrittswellenfunktion benötigen, die zum Beispiel aus der Näherung des schwachen Phasenobjekts abgeleitet wird.

8.4.2 Potentialrekonstruktionsverfahren

Die logische Reihenfolge einer Strukturrekonstruktion erfordert, daß zuerst die Austrittswellenfunktion aus den experimentellen Bildern bestimmt wird, und anschließend die Rekonstruktion des projizierten Potentials der Objektstruktur aus der Austrittswellenfunktion erfolgt. So wurden auch – historisch gesehen – mehrere Algorithmen zur Rekonstruktion der Austrittswellenfunktion entwickelt, während bis jetzt nur sehr wenige Verfahren zur hochaufgelösten Potentialrekonstruktion existieren. Der erste Grund dafür ist, daß eine allgemeine Potentialrekonstruktion die Lösung des zugehörigen inversen quantenmechanischen Streuproblems erforderlich macht,

was vom Standpunkt der Theorie aus ohne ein bestimmtes Maß an Vorwissen über die Struktur äußerst schwierig ist [242]. Der zweite Grund ist die experimentelle Situation: Wegen des Informationslimits des Mikroskops kann die Austrittswellenfunktion nur mit einer bestimmten Grenzauflösung gemessen werden, das heißt es gibt in den Bildern eine experimentell begrenzte maximale Raumfrequenz. Für die dynamische Beugung und damit für die Potentialrekonstruktion sind aber im allgemeinen auch Potentialkoeffizienten höherer Raumfrequenzen wichtig und bestimmend. Der dritte Grund ist der enorme numerische Berechnungsaufwand für eine hochauflösende Potentialrekonstruktion. Die stetig wachsende Leistungsfähigkeit moderner Digitalrechner hat eine algorithmische Umsetzung erst sinnvoll gemacht.

In den vergangenen Jahren wurde von Van Dyck und Mitarbeitern [251, 235] ein Potentialrekonstruktionsverfahren entwickelt, das auf dem Bild des Elektronen-Channeling entlang niedrigindizierter Kristallrichtungen im Wechselwirkungspotential zwischen hochenergetischen Elektronen und Atomsäulen beruht. Es wurde gezeigt, daß für nicht zu enge Atomsäulenabstände und nicht zu hohe Masse pro Längeneinheit einer Atomsäule die Wechselwirkung zwischen den schnellen Elektronen und dem projizierten Kristallpotential dazu führt, daß die Wellenfunktion während der dynamischen Beugung an den Atomsäulen lokalisiert bleibt. Dies führt zu einem relativ einfachen Zusammenhang zwischen der Streustärke einer Atomsäule und der Wellenfunktion, wobei die totale Austrittswellenfunktion durch die Superposition der Wellenfunktionen der einzelnen Atomsäulen gegeben ist. Die Methode ist jedoch nur anwendbar, wenn zuvor die Austrittswellenfunktion mit einer Auflösung gemessen oder rekonstruiert wurde, die die Trennung einzelner Atomsäulen gestattet. Außerdem existiert eine obere Grenze für die Probendicke, ab der das einfache Bild des Channeling nicht mehr gültig ist. In seiner einfachsten Form ist das Verfahren nur dann durchzuführen, wenn die dynamische Beugung im Rahmen einer Zwei-Blochwellen-Näherung beschrieben werden kann.

Eine zweites Rekonstruktionsverfahren wurde von Gribelyuk [211] veröffentlicht, das auf der Verfeinerung einer anfänglichen Schätzung des projizierten Streupotentials mit Hilfe des Multislice-Verfahrens [202] beruht. In einem ersten Schritt wird eine Schätzung des projizierten Streupotentials aus der experimentell ermittelten Austrittswellenfunktion unter der Annahme von Einfachstreuung und Propagation der gestreuten Welle über die Distanz der Probendicke gewonnen. Im zweiten Schritt wird diese Näherung des Streupotentials verbessert, indem der Unterschied zwischen der experimentell bestimmten und einer simulierten Austrittswellenfunktion berücksichtigt wird, die mittels des Multislice-Verfahrens aus dem genäherten Streupotential berechnet wird. Der letzte Schritt wird so lange wiederholt, bis Konvergenz gegen eine stabile Lösung eintritt. Es wurde gezeigt, daß diese Rekonstruktionsmethode nicht gegen die richtige Lösung konvergiert, wenn die Annahmen, die zur ersten Schätzung des projizierten Streupotentials führen, nicht auf die experimentell vorliegende Streusituation zutreffen.

Kürzlich wurde eine Variante dieses Rekonstruktionsverfahrens veröffentlicht [197, 241], in der der Multislice-Zyklus in *umgekehrter* Reihenfolge durchlaufen wird. In diesem Fall wird die Verfeinerung des projizierten Streupotentials aus der Differenz zwischen einer ebenen Welle und der *Eintrittswellenfunktion* berechnet, die sich

aus der experimentell bestimmten Austrittswellenfunktion ergibt, wenn man die Multislice-Iteration mit dem genäherten Streupotential in umgekehrter Reihenfolge von der Austrittsfläche des streuenden Objekts zur Eintrittsfläche durchläuft.

Kapitel 9

Theorie der stochastischen Rekonstruktionsverfahren

Eine Übersicht der existierenden Phasenrekonstruktionsverfahren und der Potentialrekonstruktionsverfahren, die im vorangehenden Kapitel gegeben wurde, zeigt, daß die jeweiligen Verfahren mindestens von einer der folgenden Annahmen ausgehen:

- Die dynamische Beugung ist durch die Näherung des schwachen Phasenobjekts beschreibbar.
- Die dynamische Beugung ist durch eine Zwei-Blochwellen-Näherung beschreibbar.
- Der Kontrasttransfer ist linear.
- Es existiert eine ausreichend gute Schätzung des zu suchenden atomaren Kristallpotentials, beziehungsweise einer zu suchenden Austrittswellenfunktion vor, die als Anfangsnäherung genutzt werden kann.

In sehr vielen Anwendungen trifft jedoch keine der genannten Bedingungen zu, so daß der Erfolg einzelner Rekonstruktionsverfahren in Frage gestellt ist. In einigen wenigen Fällen sind entsprechende „Fehlschläge" auch publiziert worden [211, 197, 241].

In Hinblick auf diese Situation wurde in der vorliegenden Arbeit ein neuartiger Ansatz untersucht, der die Rekonstruktion einer beliebigen Austrittswellenfunktion aus elektronenmikroskopischen Bildern und die Rekonstruktion eines beliebigen atomaren Kristallpotentials aus einer Austrittswellenfunktion ermöglichen soll [247, 227]. Der Leitgedanke solcher allgemeiner Rekonstruktionsverfahren ist im Falle der Phasenrekonstruktion das Testen einer sehr großen Anzahl von Austrittswellenfunktionen mit Hilfe einer Bildberechnung im Rahmen der Kontrastübertragung unter partiell kohärenter Beleuchtung und einem anschließenden numerischen Bildvergleich mit einem experimentellen Bild. Aus der Menge der untersuchten Austrittswellenfunktionen ist dann jene als die Lösung des Rekonstruktionsproblems anzusehen, deren zugehörige Bilder auf der Basis eines Bildvergleichskriteriums die beste Übereinstimmung mit den experimentellen Bildern zeigt.

Analog ist im Fall der Potentialrekonstruktion eine große Anzahl atomarer Kristallpotentiale zu testen, indem man jeweils mit einem der in Kapitel 7 dargestellten Verfahren die zugehörigen Austrittswellenfunktionen berechnet, und sie mit Hilfe eines numerischen Kriteriums mit einer zuvor rekonstruierten experimentellen Austrittswellenfunktion vergleicht. Aus der Menge der untersuchten atomaren Kristallpotentiale ist dann jenes als die Lösung des Rekonstruktionsproblems anzusehen, dessen zugehörige Austrittswellenfunktion auf der Basis des Vergleichskriteriums die beste Übereinstimmung mit der experimentellen Austrittswellenfunktion zeigt.

Es ist naheliegend, wie bei der Bestimmung der Beugungs- und Abbildungsparameter, Probendicke t und Defokus Z (Abschnitt 8.3.1), zunächst ein vollständiges Durchprobieren aller möglichen Austrittswellenfunktionen, beziehungsweise aller möglichen atomaren Kristallpotentiale ins Auge zu fassen. Einer Fourierdarstellung einer Austrittswellenfunktion und eines Kristallpotentials zum Beispiel bei einem Kristall in einer Zonenachsenorientierung ist jedoch zu entnehmen, daß die Zahl N_P der beteiligten Potentialkoeffizienten $U_{\vec{g}}$ und die Zahl N_S der beteiligten Strahlen $\Psi_{\vec{g}}$ sehr groß ist. Die beiden zugehörigen „Suchräume“ des entsprechenden Rekonstruktionsproblems, die durch die komplexen Koordinaten $U_{\vec{g}}$ und $\Psi_{\vec{g}}$ aufgespannt werden, haben dann die Dimensionen $2N_\mathrm{P}$ und $2N_\mathrm{S}$. N_P und N_S sind fatalerweise so groß, daß ein „vollständiges“ Abtasten der Suchräume in äquidistanten Schritten entlang der Koordinatenachsen, wie es bei der Defokus- und Probendickenbestimmung durchgeführt wurde, praktisch jede Rechenkapazität sprengt.

Ein Ausweg aus dieser Situation sind *stochastische Suchalgorithmen*, die den Suchraum an einer geringeren Zahl von *zufällig* gewählten Punkten abtasten. Parallel zu frühen Anwendungen stochastischer Algorithmen konnte gezeigt werden, daß bei hochdimensionalen Optimierungsproblemen und bei beschränkter Rechenkapazität, also einer begrenzten Anzahl von auszuwertenden Punkten im Suchraum, das zufällige Verteilen der Punkte im Hinblick auf den statistischen Fehler des Ergebnisses und im Hinblick auf eine gute Näherungslösung des Optimierungsproblems günstiger ist, als die Punkte auf einem äquidistanten Gitter anzuordnen.

Stochastische Suchalgorithmen haben häufig eine zusätzlich wichtige Eigenschaft, die spezialisierten und effizienteren Verfahren, wie den in Kapitel 8 vorgestellten, fehlt: Sie kombinieren die *globale* Suche nach einer gut passenden Näherungslösung des Suchproblems und die *lokale* Suche nach einer optimierten Lösung, unter Benutzung einer schon gefundenen Näherungslösung. Dadurch werden die Algorithmen weniger anfällig für „schlecht“ passende Näherungslösungen, sie behalten aber gleichzeitig eine gewisse Effizienz bei der Optimierung.

Im Rahmen dieser Arbeit wurden zwei dieser stochastischen Suchverfahren zur Lösung der Rekonstruktionsaufgaben eingesetzt. Das erste Suchverfahren ist das sogenannte *Simulated Annealing*, das zweite ein *genetischer Optimierungsalgorithmus*. Beide Verfahren werden in den folgenden Abschnitten ausführlich beschrieben.

9.1 Vergleich von Austrittswellenfunktionen

Die Bewertungsfunktionen des bei einer Phasenrekonstruktion benutzten quantitativen Bildvergleichs wurden schon im Abschnitt 8.2 dargestellt. Hier ist noch anzufügen, daß bei der Bewertung einer Defokusserie aus N Bildern ebensoviele Bildvergleiche anzustellen sind. Die N zugehörigen Werte der Bewertungsfunktion R_i, $i = 1 \ldots N$ werden zu einem einzigen Wert $E(\Psi)$ zusammengefaßt, der das arithmetische Mittel der Einzelwerte ist:

$$E(\Psi) = \frac{1}{N} \sum_i R_i. \tag{9.1}$$

Bis zu dieser Stelle wurden noch keine Vergleichskriterien für Wellenfunktionen eingeführt, die bei einer Potentialrekonstruktion benötigt werden. Die Darstellung kann nach den Ausführungen des Abschnitts 8.2 jetzt kurz gefaßt werden. Im Prinzip sind die Vergleichskriterien zweier komplexer Austrittswellenfunktionen genau gleich aufgebaut, wie beim Vergleich zweier reeller Bilder. In der Realraum- und in der Reziprokraumversion lautet das Kleinstquadratkriterium:

$$R = \sum_i |\psi_{\mathrm{sim}}(\vec{r}_i) - \psi_{\mathrm{exp}}(\vec{r}_i)|^2 \tag{9.2}$$

und

$$R = \sum_{\vec{g}} |\Psi_{\mathrm{sim}}(\vec{g}) - \Psi_{\mathrm{exp}}(\vec{g})|^2 . \tag{9.3}$$

Der Kreuzkorrelationskoeffizient hat in beiden Versionen die Gestalt

$$R = \frac{\sum_i \psi_{\mathrm{sim}}(\vec{r}_i)\psi^*_{\mathrm{exp}}(\vec{r}_i)}{\sqrt{\sum_i |\psi_{\mathrm{sim}}(\vec{r}_i)|^2}\sqrt{\sum_i |\psi_{\mathrm{exp}}(\vec{r}_i)|^2}} \tag{9.4}$$

und

$$R = \frac{\sum_{\vec{g}} \Psi_{\mathrm{sim}}(\vec{g})\Psi^*_{\mathrm{exp}}(\vec{g})}{\sqrt{\sum_{\vec{g}} |\Psi_{\mathrm{sim}}(\vec{g})|^2}\sqrt{\sum_{\vec{g}} |\Psi_{\mathrm{exp}}(\vec{g})|^2}}. \tag{9.5}$$

Die Austrittswellenfunktionen müssen natürlich beide gleich normiert sein:

$$\sum_i |\psi_{\mathrm{sim}}(\vec{r}_i)|^2 = 1 \tag{9.6}$$

und

$$\sum_i |\psi_{\mathrm{exp}}(\vec{r}_i)|^2 = 1, \tag{9.7}$$

beziehungsweise

$$\sum_{\vec{g}} |\Psi_{\mathrm{sim}}(\vec{g})|^2 = 1 \tag{9.8}$$

und

$$\sum_{\vec{g}} |\Psi_{\mathrm{exp}}(\vec{g})|^2 = 1. \tag{9.9}$$

Darüber hinaus gibt es noch zwei notwendige Erweiterungen die daraus resultieren, daß eine Austrittswellenfunktion eine komplexe Funktion ist.

Die aus einer Bildintensität rekonstruierte Austrittswellenfunktion kann noch einen Phasenfaktor enthalten, der in der konventionellen Beugungs- und Abbildungsgeometrie eines Transmissionselektronenmikroskops nicht meßbar ist. Diese Tatsache äußert sich zum Beispiel dadurch, daß in der Beschreibung der dynamischen Beugung durch die Schrödingergleichung (Abschnitt 7.1) ein Phasenfaktor $\exp(2\pi i z U_0/2K)$ ausgeklammert und weggelassen werden kann, dessen Ursache die Propagation der gebeugten Wellen durch das *mittlere innere Potential* des Kristalls ist. Dieser Phasenfaktor ist nur meßbar, wenn man, wie in der Elektronenholographie [228, 209], einen Referenzstrahl zum Vergleich heranzieht, der nur durch das Vakuum verläuft und der nicht durch den Kristall gebeugt wird. Aus diesem Grund ist in den hier diskutierten Vergleichskriterien eine zusätzliche Normierung der simulierten und der experimentellen Austrittswellenfunktionen notwendig, die sinnvollerweise so durchgeführt wird, daß der Bezugspunkt die Phase des Nullstrahls ist.

Im allgemeinen sind auch nach der Normierung der Nullstrahlphase der Austrittswellenfunktionen die Zähler der Terme in (9.4) und (9.5) damit auch die gesamten Kreuzkorrelationskoeffizienten komplexe Größen. Zur Konstruktion einer sinnvollen Bewertungsfunktion, die aus einer einzigen reellen Größe besteht, ist dann zum Beispiel der Absolutbetrag des Kreuzkorrelationskoeffizienten geeignet. Diesem Wert läßt sich dann in Analogie zum Bildvergleichskriterium ein „Korrelationswinkel" mittels

$$\alpha = \arccos \left| R^2 \right| \tag{9.10}$$

zuordnen.

9.2 Simulated Annealing

Die numerische Methode des *Simulated Annealing* ist aus Grundprinzipien der statistischen Mechanik abgeleitet. Die Methode wurde von Metropolis und Mitarbeitern in einer Arbeit über ein Teilchensystem im thermischen Gleichgewicht eingeführt [230]. Eine Weiterentwicklung wurde von Kirkpatrick und Mitarbeitern bei der Anwendung der Methode auf Probleme der kombinatorischen Optimierung durchgeführt [224, 225].

Das Prinzip des Simulated Annealing ist auf die Beobachtung gegründet, daß ein Vielteilchensystem bei der Temperatur T ein Minimum der freien Energie annimmt. Die Suche nach einer Austrittswellenfunktion, die durch N Strahlen beschrieben werden kann, und die eine Lösung des Phasenrekonstruktionsproblems darstellt, sowie die Suche nach einem Kristallpotential, das durch N Fourierkomponenten beschrieben wird, und das eine Lösung des Potentialrekonstruktionsproblems darstellt, kann wie eine Suche nach dem Minimum der freien Energie eines N-Teilchen-Systems behandelt werden. Dazu ist im Falle der Phasenrekonstruktion das Bildvergleichsmaß, und im Falle der Potentialrekonstruktion das Vergleichsmaß zweier Austrittswellenfunktionen als freie Energie E einer N-Teilchen-Konfiguration zu interpretieren, die

im ersten Fall aus dem Satz der Fourierkomponenten $\Psi(\vec{g})$ und im zweiten Fall aus dem Satz der Potentialkoeffizienten $U_{\vec{g}}$ besteht.

In der statistischen Mechanik ist die Wahrscheinlichkeit π_s, daß sich ein System in einer Konfiguration s befindet, durch die Boltzmann-Verteilung

$$\pi_s = \frac{\exp\left(-\frac{E_s}{kT}\right)}{\sum_{w \in S} \exp\left(-\frac{E_w}{kT}\right)} \tag{9.11}$$

gegeben, wenn sich das System im thermischen Gleichgewicht befindet. S bezeichnet die Menge aller möglichen Konfigurationen, E die Energie einer bestimmten Konfiguration, T die Temperatur und k die Boltzmann-Konstante. Ein Übergang von einer Anfangskonfiguration i zu einer Folgekonfiguration f tritt dann mit einer Wahrscheinlichkeit

$$p_{i \to f} = \frac{\pi_f}{\pi_i} = \exp\left(-\frac{E_f - E_i}{kT}\right) \tag{9.12}$$

ein. Für $E_f \leq E_i$ findet der Übergang in jedem Fall statt, für $E_f > E_i$ nur mit einer Wahrscheinlichkeit $p < 1$.

Bei der Temperatur T befindet sich das System im Konfigurationsraum auf einer Schale mit einer Dicke proportional zu kT. Bei einer kleineren Temperatur befindet sich das System auf einer tieferen Schale mit kleinerer Oberfläche, gleichzeitig nimmt die Dicke der Schale ab. Im Grenzfall $T \to 0$ erreicht das System einen Zustand, der aus wenigen Konfigurationen oder gar einer einzigen Konfiguration niedrigster Energie besteht. Eine solcher Zustand heißt *Grundzustand*.

Ein Simulated-Annealing-Algorithmus besteht dementsprechend aus folgenden Komponenten: einer Repräsentation einer zu optimierenden Konfiguration, einem Generator für eine Anfangskonfiguration, einem Generator für neue Folgekonfigurationen, einer Bewertungsfunktion E, die einer Konfiguration eine Energie zuordnet, einer Funktion, die die Übergangswahrscheinlichkeit $p_{i \to f}$ berechnet, und einer Temperatursteuerung, die das Absenken der Temperatur T von einem Anfangswert auf den Wert Null kontrolliert.

In der Implementation des Simulated Annealing für die Phasenrekonstruktion einer Austrittswellenfunktion besteht eine zu optimierende Konfiguration aus einem Satz von reellen Zahlen, die die Amplituden und Phasen der Fourierkomponenten $\Psi(\vec{g})$ einer Test-Austrittswellenfunktion repräsentieren. In der Implementation für die Potentialrekonstruktion besteht die zu optimierende Konfiguration aus Amplituden und Phasen des Arguments $\sigma \Phi_P(\vec{g}) \Delta z$ der Transmissionsfunktion (7.37), das proportional zum Potentialkoeffizienten $\Phi_P(\vec{g})$ des projizierten Potentials ist.

Eine Anfangskonfiguration der Amplituden- und Phasenwerte wird in beiden Fällen mit Hilfe eines Zufallszahlengenerators erzeugt. Eine neue zu testende Folgekonfiguration wird erzeugt, indem man in einer Kopie der Vorgängerkonfiguration einen Amplituden- oder Phasenwert eines zufällig gewählten Strahls durch eine Zufallszahl ersetzt. Die Energie der Konfigurationen wird mit Hilfe der Bildvergleichskriterien aus Abschnitt 8.2 berechnet. Konkret wird für den Bildvergleich in dieser Arbeit der dimensionslose Kreuzkorrelationskoeffizient (8.10) benutzt, und für den

Vergleich zweier Wellenfunktionen wird das Kleinstquadrat-Kriterium (9.3) verwendet. Dementsprechend wird auch die Größe kT durch einen dimensionslosen Wert ersetzt. Die Übergangswahrscheinlichkeit ist dann der Gleichung (9.12) zu entnehmen. Die Entscheidung über die Annahme einer Folgekonfiguration wird mit Hilfe eines Zufallszahlengenerators getroffen, der Zufallszahlen zwischen Null und Eins erzeugt. Eine Folgekonfiguration wird dann angenommen, wenn die Übergangswahrscheinlichkeit größer als die Zufallszahl ist, ansonsten wird sie zurückgewiesen. Nach einer bestimmten Anzahl von erfolgreichen Übergängen wird angenommen, daß sich das System in einem thermischen Gleichgewicht befindet. Dann wird die Temperatur durch die Ersetzung $T \longleftarrow cT$, $c < 1$, um einen kleinen Betrag reduziert. Die gesamte Sequenz wird solange wiederholt, bis die Temperatur bis nahe Null gefallen ist. In Abbildung 9.1 ist eine kompakte Zusammenfassung des Simulated-Annealing-

```
Setze Anfangstemperatur;
Erzeuge Anfangskonfiguration;
Berechne Wert der Evaluationsfunktion für Anfangskonfiguration;
Wiederhole
   Wiederhole
     Erzeuge neue Konfiguration;
     Berechne Wert der Evaluationsfunktion für neue Konfiguration;
     Berechne Übergangswahrscheinlichkeit zur neuen Konfiguration;
     Entscheide, ob neue Konfiguration angenommen wird;
     Wenn ja:
       Ersetze aktuelle Konfiguration durch neue Konfiguration
   bis thermisches Gleichgewicht erreicht ist;
   Senke Temperatur um kleinen Betrag
bis Grundzustand erreicht ist.
```

Abbildung 9.1: Kompakte Darstellung des Simulated-Annealing-Algorithmus.

Algorithmus dargestellt.

Das langsame Verkleinern der Temperatur hat zur Folge, daß sich das System ständig nahe des thermischen Gleichgewichts befindet, so daß die eingangs geforderte Balance zwischen der globalen Suche nach neuen Konfigurationen minimaler Energie und der Optimierung einer schon nahe eines Minimums befindlichen Konfiguration besteht. Es erweist sich jedoch, daß ein Simulated-Annealing-Algorithmus nicht gut geeignet ist, eine schon gefundene Konfiguration minimaler Energie auch numerisch effizient zu optimieren. Im Prinzip kann man dann das Simulated Annealing abbrechen und die Optimierung mit anderen effizienten lokalen Suchverfahren fortsetzen. Es ist der Idee der stochastischen Suche angemessen, das Simulated Annealing für die Endoptimierung durch einen parallel ablaufenden Suchprozeß zu erweitern. In der vorliegenden Anwendung besteht diese Erweiterung in einer lokalen Suche auf der Basis des sogenannten *Hill-Climbing*. Dabei wird bei einer Konfigurationsänderung nicht eine beliebige neue Komponente einer Konfiguration eingesetzt, sondern es

werden nur kleine Änderungen einer zufällig ausgewählten Komponente der schon gefundenen, fast optimalen Konfiguration vorgenommen. Durch diese Maßnahme wird die Konvergenz der Endoptimierung beschleunigt, ohne daß das stochastische Suchprinzip verletzt wird.

9.3 Genetische Algorithmen

Das Prinzip genetischer Suchalgorithmen ist von einem vereinfachten Modell der Evolution abgeleitet, die als ein Optimierungsprozeß in der belebten Natur angesehen werden kann. Das Ziel dieses Optimierungs- und Suchprozesses ist die ständige Anpassung und Optimierung einer Art an eine vorhandene Umwelt. Genetische Algorithmen wurden zuerst von Holland entwickelt [216], der herausfand, daß mittels genetischer Suchverfahren unter gewissen Umständen fast-optimale Lösungen eines mathematischen Optimierungsproblems bestimmt werden können.

Ein einfaches Modell der Evolution, das für die Lösung von Optimierungsproblemen ausreichend ist, besteht aus folgenden Komponenten: eine Repräsentation der zu optimierenden Eigenschaften durch „Chromosomen", eine Anzahl von „Lebewesen", die zusammen eine Art bilden, bei der jedes Lebewesen eine unterschiedliche Chromosomen-Konfiguration repräsentiert, eine Bewertungsfunktion, die den Grad der Anpassung eines Lebewesens an seine Umwelt mißt, und genetische Operatoren, wie zum Beispiel Mutation und *Crossing-Over*, die eine Veränderung der Chromosomen von einer Generation zur nächsten bewirken [203].

In genetischen Algorithmen ändert der Mutationsoperator einzelne *Gene* eines Chromosoms, wobei Gene als in Chromosomen angeordenete Informationseinheiten angesehen werden können. Durch die Änderung einzelner Gene erhalten die Lebewesen und damit die gesamte Population eine Anzahl neuer Fähigkeiten und Eigenschaften. Das Crossing-Over wird ausgeführt, indem einzelne Gensequenzen von einem Chromosom in ein anderes kopiert werden, so daß eine schon existierende Information in einer neuen Weise kombiniert wird. Die Wahrscheinlichkeit, mit der ein Lebewesen der Population für ein Crossing-Over, und damit für eine Weitergabe seiner Eigenschaften an die nächste Generation ausgewählt wird, hängt von seinem individuellen Anpassungsgrad an die Umwelt ab. Die Crossing-Over-Operation ist derjenige Vorgang, der die Stärke eines genetischen Algorithmus ausmacht, denn die ständige Neukombination von Eigenschaften einzelner Individuen, unter besonderer Berücksichtigung gut angepaßter Individuen, steigert die Wahrscheinlichkeit, mit der wachsenden Zahl von Generationen ein beinahe perfekt angepaßtes Lebewesen zu erhalten.

Es gibt durch die jahrelange Forschung auf dem Gebiet der genetischen Optimierungsverfahren mittlerweile eine kaum noch zu überblickende Anzahl von Implementationsmöglichkeiten und Varianten, die etwa die Genrepräsentation der zu optimierenden Größen oder die Erfindung neuer genetischer Operatoren betreffen. Ein Übersicht ist zum Beispiel der Artikelsammlung [203] zu entnehmen.

In der in dieser Arbeit entwickelten Implementation eines genetischen Suchalgorithmus zur Phasenrekonstruktion einer Austrittswellenfunktion besteht die Chro-

mosomenrepräsentation aus der Sequenz von Amplituden und Phasen der Fourierkomponenten einer Austritswellenfunktion. Ene Anfangspopulation wird durch Belegen aller Gensequenzen mit gleichmäßig verteilten Zufallszahlen erzeugt. Die Mutationsoperation wird genau gleich wie eine Konfigurationsänderung beim Simulated Annealing durchgeführt (Abschnitt 9.2). Es werden gleichmäßig verteilte Zufallszahlen für die Wahl eines bestimmten Individuums, eines Gens seiner Chromosomenrepräsentation, und für die Entscheidung über das Ersetzen eines Amplituden- oder Phasenwertes benutzt. Das Crossing-Over wird ebenfalls durch Zufallszahlen kontrolliert. Der erste Partner ist stets das bestangepaße Individuum einer Population, und der zweite Partner ist ein zufällig gewähltes Mitglied der restlichen Population. Die Länge der Gensequenz, die zwischen beiden Partnern ausgetauscht wird, variiert zufällig von einem einzigen Gen bis zur Länge der vollständigen Gensequenz. Die Länge, sowie der Anfangs- und Endpunkt des Genabschnitts werden ebenfalls durch Zufallszahlen vorgegeben. Um die Populationsgröße von Generation zu Generation konstant zu halten, ersetzt einer der beiden Nachkommen das am schlechtesten angepaßte Individuum der Population. Abbildung 9.2 zeigt eine kompakte Darstellung

```
Erzeuge Anfangspopulation von Individuen;
Bewerte jedes Individuum mittels Bewertungsfunktion;
Wiederhole
   Mit einer Wahrscheinlichkeit W_P:
     Wähle zwei Individuen der Population aus;
     Führe Crossing-Over mit beiden Individuen aus;
     Bewerte den Nachkommen;
     Ersetze ein Individuum der Population durch den Nachkommen
   Mit einer Wahrscheinlichkeit W_M:
     Wähle ein Individuum der Population aus;
     Wähle ein Gen im Chromosom des Individuums;
     Ändere das Gen zufällig;
     Bewerte das Individuum neu
bis die Population optimiert ist.
```

Abbildung 9.2: Kompakte Darstellung des genetischen Suchalgorithmus.

eines genetischen Optimierungsalgorithmus.

Ein genetischer Algorithmus ist in ähnlicher Weise wie das Simulated Annealing schlecht geeignet, eine numerisch effiziente Optimierung des am besten angepaßten Individuums in der Endphase eines Suchprozesses durchzuführen. Deshalb wurde in der vorliegenden Implementation des genetischen Suchalgorithmus ein weiterer Mutationsoperator eingesetzt, der analog zur Hill-Climbing-Funktion des Simulated Annealing ein Variieren eines Gens um einen schon gut angepaßten Wert ermöglicht. Da dieser zweite Mutationsoperator gleichberechtigt zu den anderen Operatoren eingesetzt wird, bleibt das Prinzip der genetischen Optimierung erhalten, ohne daß eine spezialisierte lokale Optimierung durchgeführt werden müßte.

Kapitel 10

Anwendung stochastischer Phasenrekonstruktionsverfahren

In diesem Kapitel werden vier Rekonstruktionsexperimente beschrieben, die zur Untersuchung der Eigenschaften stochastischer Rekonstruktionsverfahren durchgeführt wurden. Im ersten Teil wird eine Phasenrekonstruktion einer periodischen Austrittswellenfunktion sowohl mittels eines Simulated-Annealing-Algorithmus als auch mittels eines genetischen Algorithmus behandelt, die von einer simulierten Defokusserie elektronenmikroskopischer Bilder der intermetallischen Verbindung Ni_4Mo ausgeht. Im zweiten Teil werden beide Rekonstruktionsalgorithmen auf eine experimentell aufgenommene Defokusserie elektronenmikroskopischer Bilder des gleichen Materials angewendet. Den Abschluß bildet eine Diskussion der Eigenschaften beider Algorithmen.

Die intermetallische Verbindung Ni_4Mo [252, 226, 237, 196] ist eine Ordnungslegierung, die unterhalb einer Temperatur von 860°C in der $D1_a$-Struktur kristallisiert. Die Legierung hat die Grundstruktur einer kfz-Matrix mit einer Gitterkonstante von 0.36 nm. Eine Projektion von 2×2 $D1_a$-Einheitszellen entlang der [001]-Richtung ist in der Abbildung 10.1 dargestellt. Sie zeigt die typische Aufeinanderfolge von vier Nickelatomen und einem Molybdänatom entlang einer ⟨100⟩-Richtung des kfz-Gitters und die relative Verdrehung der projizierten kfz- und $D1_a$-Einheitszellen um einen Winkel von 18°. Das Kontrastverhalten hochaufgelöster Abbildungen von Ni_4Mo in [001]-Projektion wurde in einer Arbeit von Thust und Urban [246] mit Hilfe eines quantitativen numerischen Bildvergleichs ausführlich untersucht.

10.1 Anwendung auf eine simulierte Defokusserie

Zur Untersuchung der Eigenschaften stochastischer Rekonstruktionsalgorithmen ist es wünschenswert, eine vollständige Kontrolle aller Eingabe- und Ausgabegrößen zu haben. Insbesondere ist es zur Bewertung der Leistungsfähigkeit der Rekonstruktionsalgorithmen von Vorteil, wenn die aus einer Defokusserie von Bildern zu rekonstruierende Austrittswellenfunktion schon vorher bekannt ist. Eine solche Situation ist gegeben, wenn man den Rekonstruktionsalgorithmen eine simulierte Defokusse-

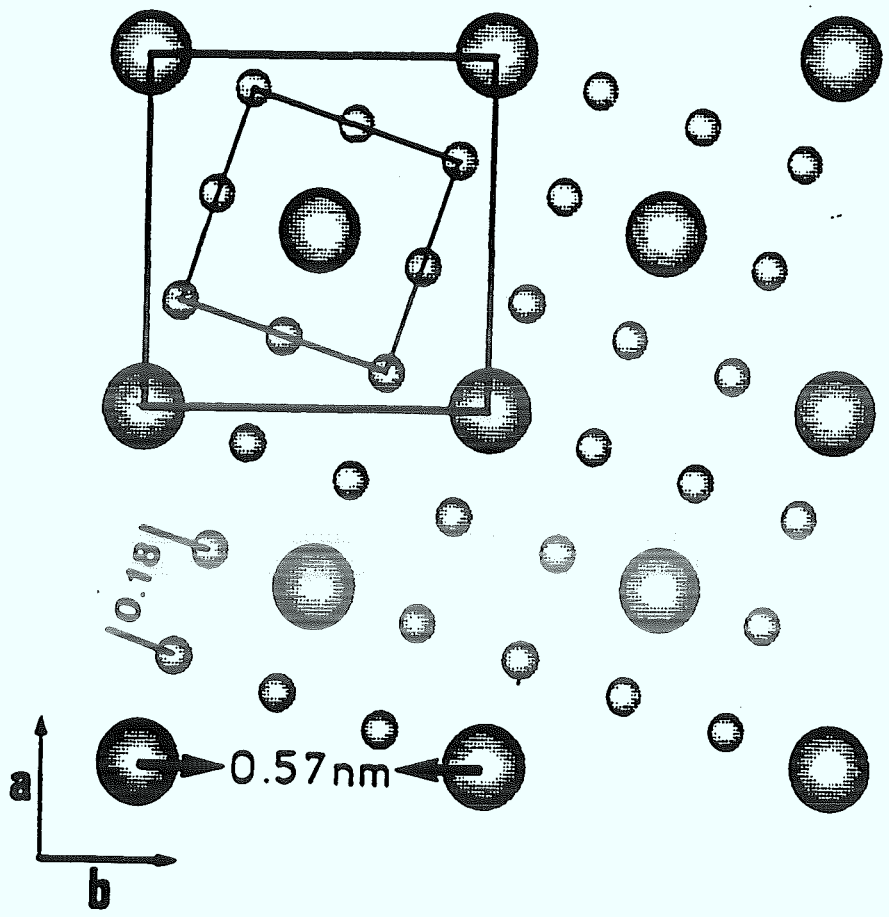


Abbildung 10.1: Projektion von 2×2 Einheitszellen der $D1_a$-Struktur von Ni_4Mo. Mo-Atome sind durch große, und Ni-Atome durch kleine Kugeln repräsentiert. Die $D1_a$-Einheitszelle ist durch dicke, und die kfz-Einheitszelle durch dünne Linien markiert.

rie von Bildern übergibt, die vorher aus einer bekannten Austrittswellenfunktion berechnet wurde. Die rekonstruierte Austrittswellenfunktion ist dann auf objektive und numerische Weise mit der bekannten Austrittswellenfunktion vergleichbar.

Zur Vorbereitung der folgenden Rekonstruktionsexperimente wurde ein Paar simulierter Eingabebilder von Ni_4Mo in [001]-Orientierung mit unterschiedlichen Defokussierungen von $\Delta Z = -35$ nm und $\Delta Z = -70$ nm berechnet. Die beiden resultierenden Bilder, die jeweils 4×4 projizierte Einheitszellen zeigen, sind in Abbildung 10.2 dargestellt. Die Simulation wurde für eine Beschleunigungsspannung von 400 kV, eine Probendicke von 25 nm, eine Konstante der sphärischen Aberration von $C_s = 1.0$ nm, einen Semikonvergenzwinkel von 0.9 mrad, eine Defokusstreuung von 10 nm und einen Blendendurchmesser von 15 nm^{-1} durchgeführt. Die gewählte Probendicke und die Defokussierungen führen zu Beugungs- und Abbildungsbedingungen, die weit entfernt von der schwachen Phasenobjektnäherung und von linearen Abbildungsbedingungen sind.

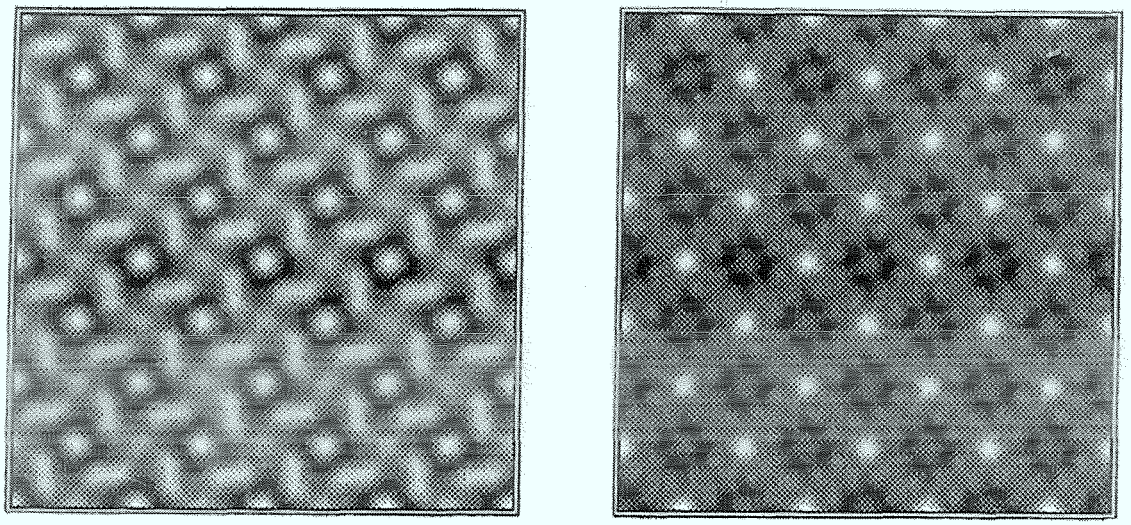
10.1.1 Rekonstruktion mittels Simulated Annealing

Die $D1_a$-Struktur hat entlang der [001]-Richtung eine vierzählige Symmetrie. Die 29 Strahlen, einschließlich des Nullstrahls, die bei einem Objektivblendendurchmesser von 15 nm^{-1} zur nichtlinearen Abbildung beitragen können, sind bei der dynamischen Beugung exakt entlang der [001]-Zonenachse nicht unabhängig, sondern sie können so in Gruppen eingeteilt werden, daß nur acht unabhängige Strahlen übrig bleiben. Diese Symmetriebeziehung kann dazu genutzt werden, durch das Verkleinern der Zahl der zu rekonstruierenden Strahlen einer Austrittswellenfunktion die Rechenzeit zu reduzieren.

Das kristallographische Zusatzwissen kann den Rekonstruktionsalgorithmen jedoch „vorenthalten" werden, indem man diese trotzdem alle 29 Strahlen einer Austrittswellenfunktion rekonstruieren läßt. Durch dieses Vorgehen kann einerseits die Abhängigkeit der Rechenzeit von der Strahlzahl studiert werden, andererseits können etwaige Rekonstruktionsfehler sichtbar gemacht werden, die sich in einer Verletzung der vierzähligen Symmetrie auswirken.

Abbildung 10.3 (A) zeigt die Amplitude (oben) und die Phase (unten) der Austrittswellenfunktion, die zur Berechnung der beiden Eingabebilder (Abbildung 10.2) benutzt wurde. Diese Wellenfunktion wird im folgenden die „erzeugende" Wellenfunktion genannt. Die Darstellung zeigt jeweils 4×4 projizierte Einheitszellen, die Grauwertskala wurde den Minima und Maxima von Amplitude und Phase angepaßt, und die Phasenwerte wurden auf eine Nullstrahlphase von Null normiert. Während die Apmlitude die Säulenkonfiguration nicht direkt wiedergibt, ist die Verteilung der Phasenwerte, die niedrige Werte an den Nickelpositionen annimmt, direkt mit der Ni_4Mo-Struktur korreliert. In diesem Fall gibt die Phaseninformation genau die Sequenz von vier Nickelatomen und einem Molybdänatom wieder, die in den Eingabebildern (Abbildung 10.2) nicht zu erkennen ist.

Die aus den Eingabebildern mittels des Simulated-Annealing-Algorithmus rekonstruierte Austrittswellenfunktion ist in Abbildung 10.3 (B) dargestellt. Sie zeigt das Ergebnis der 29-Strahlen-Rekonstruktion (siehe oben), während das Ergebnis



1 2

Abbildung 10.2: Simulierte Eingabebilder von Ni_4Mo in einer [001]-Projektion. Die beiden Bilder zeigen jeweils 4×4 projizierte Einheitszellen. Bild 1 wurde für eine Probendicke von 25 nm und einen Defokus von -35 nm simuliert, Bild 2 für die gleiche Probendicke und einen Defokus von -70 nm.

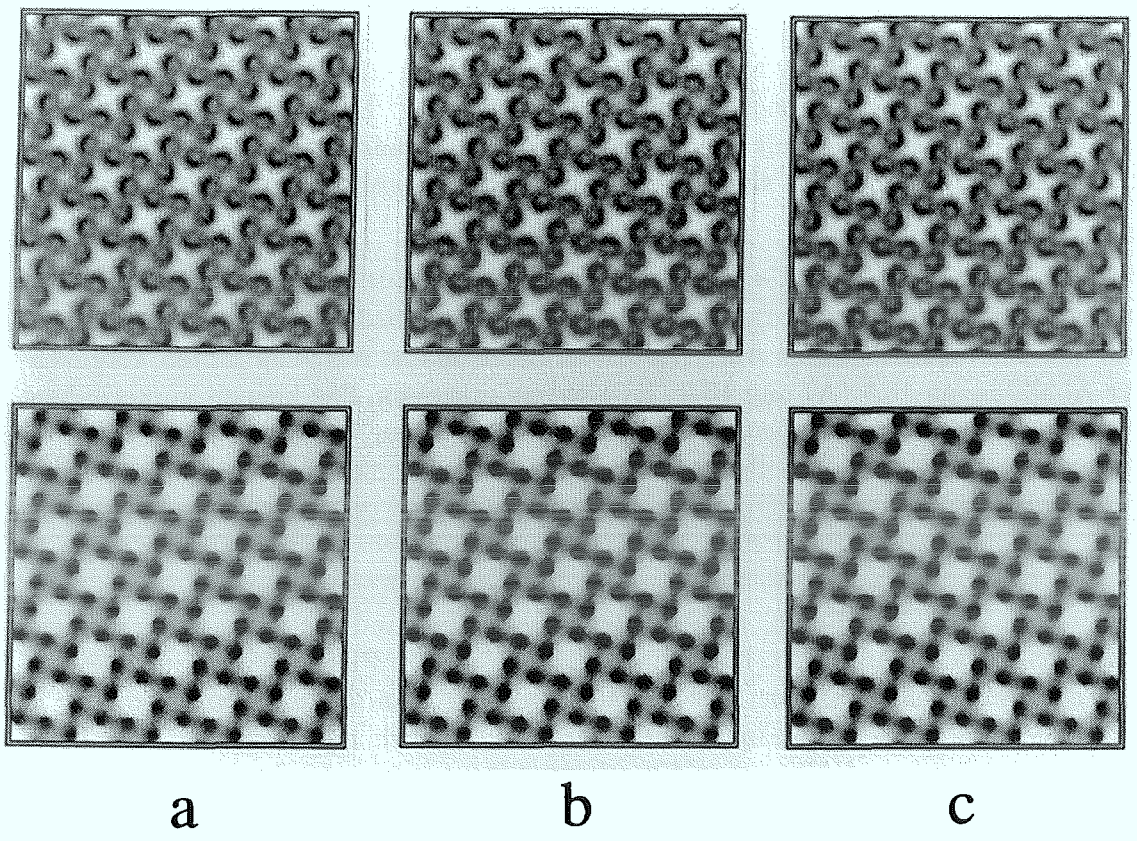


Abbildung 10.3: Austrittswellenfunktionen von Ni_4Mo[001] für eine Probendicke von 25 nm. Jedes Teilbild zeigt 4×4 Einheitszellen. Die obere Reihe zeigt die Amplituden der Wellenfunktionen, die untere Reihe die Phasen. (A) Erzeugende Wellenfunktion, die zur Berechnung der in Abbildung 10.2 gezeigten Bilder benutzt wurde. (B) Aus den Eingabebildern mit Hilfe des Simulated Annealing rekonstruierte Wellenfunktion, 29-Strahl-Modus. (C) Aus den Eingabebildern mit Hilfe des genetischen Suchalgorithmus rekonstruierte Wellenfunktion, 29-Strahl-Modus.

(hkl)	A_i	A_o	$\Delta A/A_i$ [%]	φ_i [Grad]	φ_o [Grad]	$\varphi_o - \varphi_i$ [Grad]
(000)	1.0000	1.0000	0.00	0.00	0.00	0.00
(110)	0.3341	0.3341	0.00	83.54	83.55	0.01
(200)	0.2481	0.2482	0.04	63.43	63.43	0.00
(220)	0.1287	0.1286	−0.08	30.48	30.48	0.00
(310)	0.1062	0.1062	0.00	25.38	25.38	0.00
(130)	0.2542	0.2542	0.00	223.99	224.00	0.01
(400)	0.0386	0.0384	−0.56	227.24	226.93	−0.31
(330)	0.0111	0.0103	−7.10	233.71	223.79	−9.92

Tabelle 10.1: Vergleich der erzeugenden Wellenfunktion mit der durch den Simulated-Annealing-Algorithmus rekonstruierten Wellenfunktion. Die Amplituden und Phasen der erzeugenden Wellenfunktion sind mit „i" bezeichnet, die der rekonstruierten Wellenfunktion mit „o". Die Amplitude des transmittierten Strahls ist auf 1 normiert, und die Phasen sind auf die Phase des transmittierten Strahls bezogen.

der 8-Strahlen-Rekonstruktion, die eine gleichaussehende Austrittswellenfunktion zur Folge hat, nicht extra dargestellt ist. Ein Unterschied zwischen der erzeugenden und den rekonstruierten Austrittswellenfunktionen ist durch optischen Vergleich der Abbildungen 10.3 (A) und 10.3 (B) nicht zu erkennen. Die kleinen numerischen Unterschiede der Amplituden- und Phasenwerte einzelner Strahlen der 8-Strahlen-Rekonstruktion sind in Tabelle 10.1 zusammengefaßt. Außer für die {400}- und die {330}-Strahlen sind die relativen Fehler der Amplitudenwerte typischerweise kleiner 0.1% und die Absolutabweichung der Phasenwerte typischerweise kleiner 0.1°.

Das Konvergenzverhalten des Simulated-Annealing-Algorithmus ist in der Abbildung 10.4 für die 8-Strahlen-Rekonstruktion (links) und die 29-Strahlen-Rekonstruktion (rechts) dargestellt. Die doppeltlogarithmischen Plots zeigen die Abhängigkeit der Bewertungsfunktion $E(\Psi)$ in Abhängigkeit vom Temperaturparameter T während eines typischen Rekonstruktionsablaufs. Die Simulated-Annealing-Rekonstruktion beginnt jeweils bei einem Korrelationswert um 90°, der das Fehlen jeglicher Übereinstimmung zwischen den zu vergleichenden Bildern anzeigt. Während des Rekonstruktionsprozesses wird die anfänglich zufällige Test-Austrittswellenfunktion mehr und mehr modifiziert, so daß sie Bilder erzeugt, die immer besser mit den Eingabebildern übereinstimmen. Der Rekonstruktionsvorgang wird bei einer Abnahme des Bewertungsparameters $E(\Psi)$ auf einen Wert von etwa 10^{-2} Grad abgebrochen, weil der zugehörige Wert des Kreuzkorrelationskoeffizienten schon so nahe bei 1 liegt, daß weitere Verbesserungen der Korrelation zwischen den Testbildern und den Eingabebildern mit der verwendeten achtstelligen Fließkommadarstellung nicht mehr aufgelöst werden können.

Aus den Konvergenzplots (Abbildung 10.4) ist außerdem abzulesen, daß die Fluktuationen der Bewertungsfunktion $E(\Psi)$ im Fall der 8-Strahlen-Rekonstruktion größer ist, als bei der 29-Strahlen-Rekonstruktion. Der Grund dafür ist der größere Einfluß einer Amplituden- oder Phasenänderung eines Strahls in der 8-Strahlen-Rekonstruktion, weil dieser Fehler ja wegen der vierzähligen Symmetrie auch viermal in die Bildberechnung eingeht. Es gibt einen Zusammenhang zwischen der Bewer-

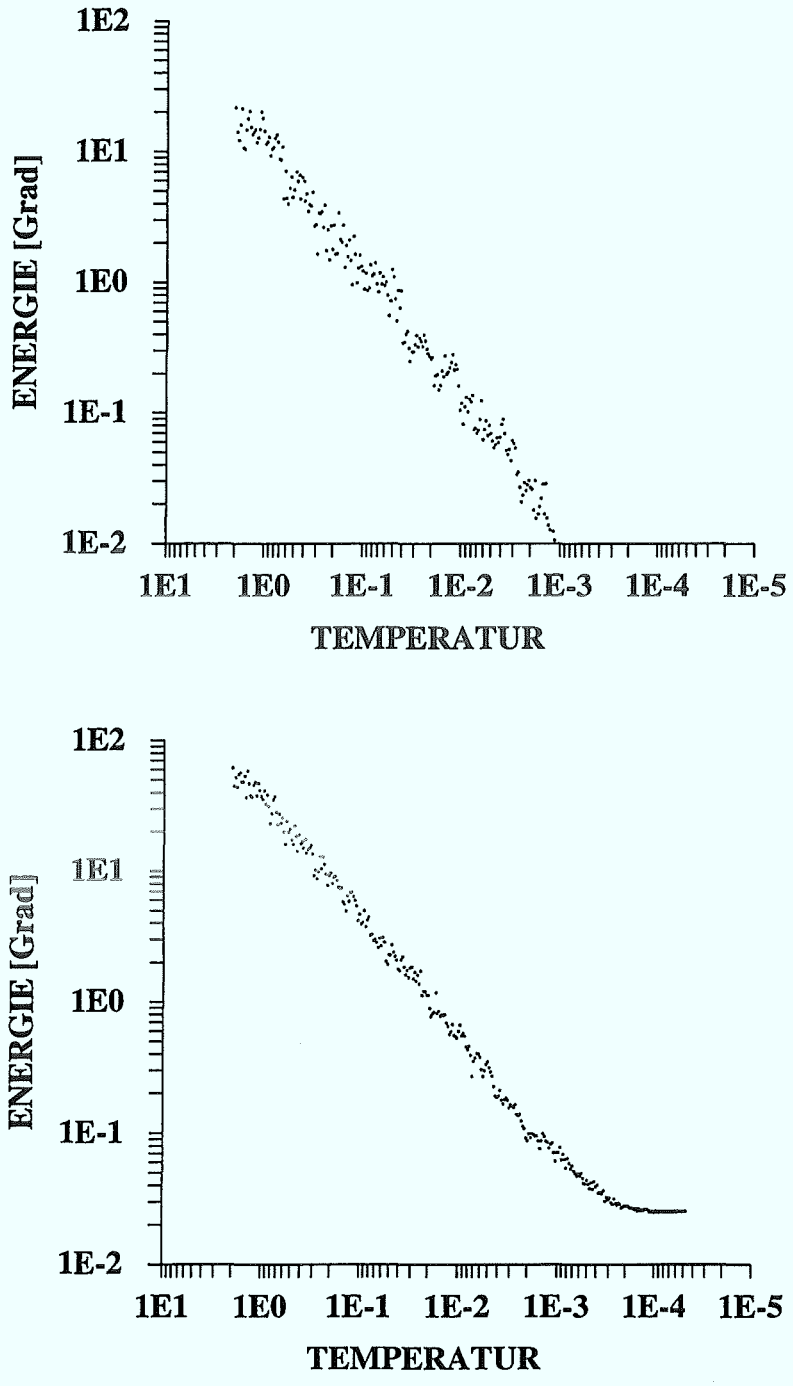


Abbildung 10.4: Konvergenzplots der Simulated-Annealing-Rekonstruktion für den 8-Strahl-Modus (oben) und den 29-Strahl-Modus (unten). Es ist doppeltlogarithmisch der Wert der Bewertungsfunktion E in Abhängigkeit vom Temperaturkontrollparameter T aufgetragen.

tungsfunktion und dem Temperaturparameter von der Form $E(\Psi) \propto T^{\kappa}$, mit $\kappa \approx 1$. Die Steigung ist in beiden Konvergenzplots etwa gleich, jedoch unterscheiden sich die benötigten Rechenzeiten in beiden Fällen. Auf einer DEC 3500 ALPHA Workstation benötigte die Rekonstruktion im 8-Strahlen-Fall etwa 1 bis 2 Minuten Rechenzeit, während für den 29-Strahlen-Fall etwa 10 Minuten Rechenzeit benötigt wurden.

Die benötigte Rechenzeit wird dadurch mitbestimmt, daß der Annealing-Prozeß nur so schnell durchgeführt werden kann, daß das System bei jedem Temperaturschritt im thermischen Gleichgewicht bleibt. Wird die Akühlrate zu groß gewählt, dann „friert" das System in einem metastabilen Zustand ein, aus dem heraus bei weiterem Abkühlen der Grundzustand nicht mehr erreichbar ist. Dieser zum raschen Abschrecken einer Schmelze analoge Prozeß ist natürlich unerwünscht, denn er bedeutet ja auf das Rekonstruktionsproblem übertragen, daß der Optimierungsprozeß bei einer Austrittswellenfunktion zum Stillstand kommt, die nicht Lösung des Rekonstruktionsproblems ist. Es zeigte sich, daß, wegen der größeren Zahl der zu optimierenden Parameter, bei der 29-Strahl-Rekonstruktion eine wesentlich langsamere Abkühlrate benutzt werden mußte, als bei der 8-Strahl-Rekonstruktion. Diese langsamere Abkühlrate findet ihren Niederschlag in einer längeren Rechenzeit.

In einem weiteren Rekonstruktionsexperiment wurde untersucht, ob die Rekonstruktion der Austrittswellenfunktion auch möglich ist, wenn nur *eines* der beiden Eingabebilder (Abbildung 10.2) benutzt wird. Für das Experiment wurde das zweite Bild (Abbildung 10.2 (2)) ausgewählt, weil dort die höheren Bildfrequenzen stärker vertreten sind. Mit Benutzung des Simulated-Annealing-Algorithmus im 8-Strahl-Modus konnte gezeigt werden, daß die Rekonstruktion der Austrittswellenfunktion mit der gleichen Genauigkeit möglich ist, wie im Fall der Rekonstruktion aus zwei Eingabebildern. Allerdings wurde eine etwa zehnfach längere Rechenzeit benötigt, weil wegen eines häufig drohenden „Einfrierens" des Suchprozesses eine relativ niedrige Abkühlrate gewählt werden mußte. Die Frage nach der Zahl der für eine Rekonstruktion benötigten Eingabebilder wird in der Diskussion weiter bearbeitet.

10.1.2 Rekonstruktion mittels eines genetischen Algorithmus

Die gleichen Rekonstruktionsexperimente, die im vorigen Abschnitt dargestellt sind, wurden mit einem genetischen Suchalgorithmus durchgeführt. Um einen Vergleich der beiden stochastischen Rekonstruktionsalgorithmen durchführen zu können, wurden die gleichen Eingabebilder wie bei der Rekonstruktion durch den Simulated-Annealing-Algorithmus verwendet. Die Rekonstruktion wurde ebenfalls im 8-Strahl-Modus, der die vierzählige Symmetrie von Ni_4Mo in einer [001]-Projektion ausnutzt, und im 29-Strahl-Modus durchgeführt.

Die rekonstruierte Austrittswellenfunktion ist in Abbildung 10.3 (C) dargestellt. Die dort gezeigte Lösung der 29-Strahl-Rekonstruktion hat das gleiche Aussehen, wie die nicht gezeigte 8-Strahl-Rekonstruktion. Die erzeugende Austrittswellenfunktion Abbildung 10.3 (A) und die rekonstruierte Wellenfunktion unterscheiden sich äußerlich nicht, die kleinen numerischen Abweichungen der 8-Strahl-Rekonstruktion sind in Tabelle 10.2 zusammengefaßt. Mit Ausnahme der {400}- und der {330}-Strahlen

(hkl)	A_i	A_o	$\Delta A/A_i$ [%]	φ_i [Grad]	φ_o [Grad]	$\varphi_o - \varphi_i$ [Grad]
(000)	1.0000	1.0000	0.00	0.00	0.00	0.00
(110)	0.3341	0.3340	−0.03	83.54	83.52	−0.02
(200)	0.2481	0.2479	−0.06	63.43	63.43	0.00
(220)	0.1287	0.1289	0.15	30.48	30.50	0.02
(310)	0.1062	0.1063	0.04	25.38	25.39	0.01
(130)	0.2542	0.2540	−0.06	223.99	223.98	−0.01
(400)	0.0386	0.0399	3.36	227.24	227.59	0.35
(330)	0.0111	0.0070	−36.77	233.71	240.32	6.61

Tabelle 10.2: Vergleich der erzeugenden Wellenfunktion mit der durch den genetischen Suchalgorithmus rekonstruierten Wellenfunktion. Die Amplituden und Phasen der erzeugenden Wellenfunktion sind mit „i" bezeichnet, die der rekonstruierten Wellenfunktion mit „o". Die Amplitude des transmittierten Strahls ist auf 1 normiert, und die Phasen sind auf die Phase des transmittierten Strahls bezogen.

sind die relativen Fehler der Amplituden kleiner als 0.1%, die absoluten Fehler der Phasenwerte sind kleiner als 0.1°. Die Rechenzeit für die 8-Strahl-Rekonstruktion betrug auf einer DEC 3500 ALPHA Workstation 1 bis 2 Minuten, während die 29-Strahl-Rekonstruktion eine Rechenzeit von 10 Minuten benötigte.

Die Kovergenzplots der 8-Strahl-Rekonstruktion (links) und der 29-Strahl-Rekonstruktion (rechts) sind in Abbildung 10.5 gezeigt, mit einer logarithmischen Auftragung der Bewertungsfunktion $E(\Psi)$ des bestangepaßten Individuums in Abhängigkeit von der Zahl der Generationen. Benachbarte Punkte in den Konvergenzplots haben einen zeitlichen Abstand von 100 Generationen.

Eine charakteristische Eigenschaft des benutzten genetischen Algorithmus ist der rapide Abfall der Bewertungsfunktion während der ersten Generationen. Bei der 8-Strahl-Rekonstruktion ist der Abfall so rasch, daß nur ein Punkt oberhalb 20° eingetragen ist. Die Konvergenzkurve der 29-Strahl-Rekonstruktion hat ebenfalls einen sehr steil abfallenden Verlauf. Dieses Verhalten wird durch die hohe Effizienz des Crossing-Over-Operators verursacht (Abschnitt 9.3), der eine massive Verbreitung optimierter Gensequenzen in der Population bewirkt. Diese lokale Optimierung vorhandener Gensequenzen führt jedoch zu einer Verarmung des Genpools der Population, so daß eine ähnliche Situation eintritt, wie beim „Einfrieren" des Simulated Annealing in einem metastabilen, nichtoptimalen Zustand. Dieses Verhalten der genetischen Optimierung wird durch Einstellen einer geeignet großen Mutationsrate vermieden. Sie verhindert einerseits die Dominanz einer einzige Sorte von Gensequenzen, andererseits stellt sie die globale Suche nach neuen, möglicherweise besser angepaßten Gensequenzen sicher. Bei zu hoher Mutationsrate wird andererseits die in der Population gespeicherte Information über optimale Gensequenzen von Generation zu Generation irreversibel zerstört. Die Effizienz eines genetischen Rekonstruktionsalgorithmus ist also von einer ausgwogenen Balance zwischen Crossing-Over und Mutation abhängig, die gleichzeitig das Verhältnis zwischen globaler Suche nach einer Lösung der Rekonstruktionsaufgabe und der lokalen Optimierung einer schon gefundenen Näherungslösung bestimmt. Der genetische Suchalgorithmus ist

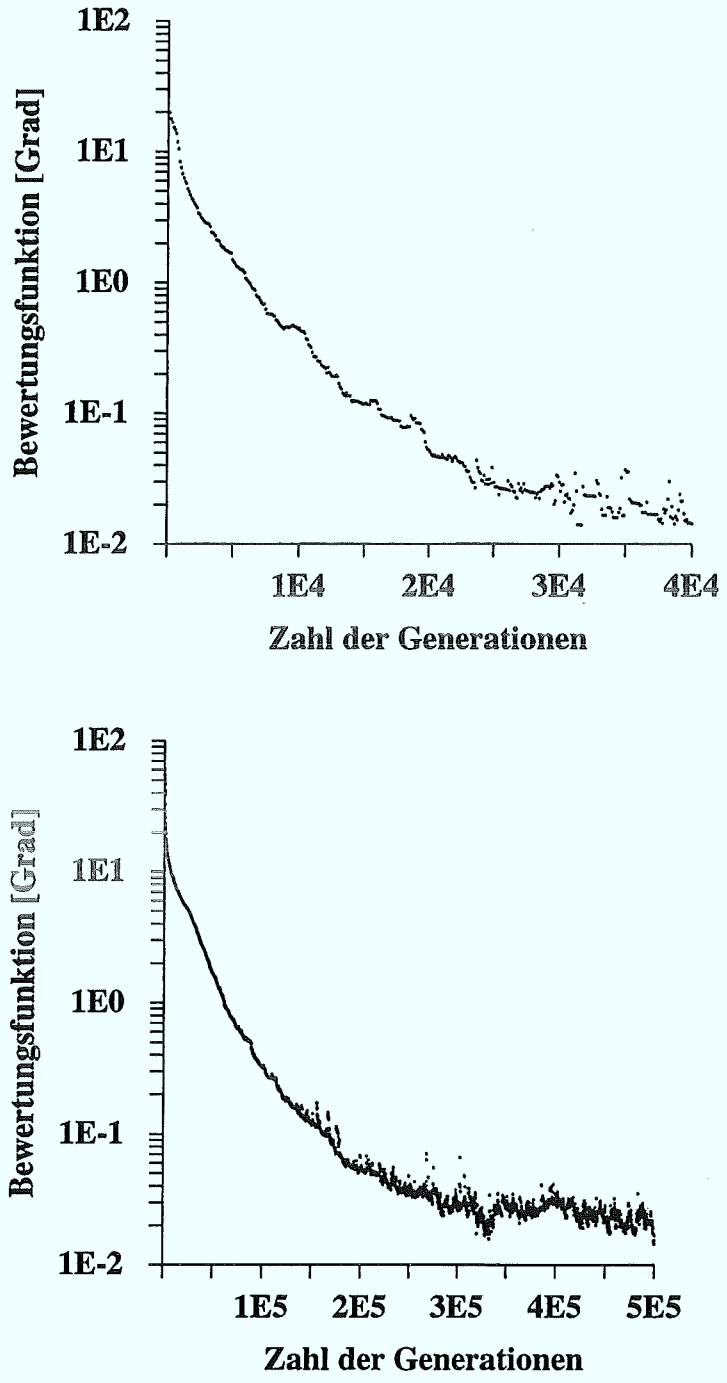


Abbildung 10.5: Konvergenzplots des genetischen Rekonstruktionsalgorithmus für den 8-Strahl-Modus (oben) und den 29-Strahl-Modus (unten). Der Logarithmus der Bewertungsfunktion E des besten Individuums ist gegen die Zahl der Generationen aufgetragen.

am effektivsten, wenn genügend große Sprünge in der Bewertungsfunktion zugelassen werden.

10.2 Anwendung auf eine experimentelle Defokusserie

10.2.1 Aufbereiten der experimentellen Defokusserie

Beide stochastische Rekonstruktionsalgorithmen, Simulated Annealing und der genetische Suchalgorithmus, wurden auf eine experimentelle Defokusserie von Ni_4Mo in [001]-Projektion angewendet. Die vier Bilder der Defokusserie wurden mit einem JEOL 4000EX bei einer Beschleunigungsspannung von 400 kV und einer Vergrößerung von 500.000× auf Kodak SO-163 Filme aufgenommen [246]. Die experimentellen Abbildungsparameter waren die gleichen, die bei der Erzeugung der simulierten Defokusserie benutzt wurden (Abschnitt 10.1).

Die auf den Photoplatten gespeicherte Bildinformation wurde in einer Reihe von Verarbeitungsschritten zur Anwendung des quantitativen numerischen Bildvergleichs aufbereitet, der während der Rekonstruktionsverfahren durchgeführt wird. Zuerst wurden die Bildintensitäten, die exakt derselben Probenstelle zugeordnet waren, mittels einer CCD-Kamera auf ein Gitter von 512 × 512 Pixeln digitalisiert, und die Grauwertskala wurde um die nichtlineare Empfindlichkeit der Photoplatten auf unterschiedliche Elektronenintensitäten korrigiert. Aus den digitalisierten Bildausschnitten wurden jeweils Gebiete von 4×4 projizierten Einheitszellen ausgewählt und auf ein neues Gitter von 256 × 256 Pixeln projiziert. Die Einheitszellen wurden dabei so angeordnet, daß die Eckpunkte mit Molybdänpositionen zusammenfallen.

Die so digitalisierte Bilderserie wurde anschließend rauschgefiltert. Dabei werden die nichtperiodischen Bildanteile durch einen Bragg-Filter unterdrückt, der im Reziprokraum nur die periodischen Raumfrequenzen durchläßt, die mit der 4 × 4-Perodizität des Bildausschnitts verträglich sind. Gleichzeitig wird dadurch der Rauschanteil, der sich auf den periodischen Bildfrequenzen befindet, um einen Faktor 4 reduziert.

Die rauschgefilterte Bildinformation enthält Kontrastanteile, die von der idealen vierzähligen Symmetrie abweichen, die eine Abbildung von Ni_4Mo in der [001]-Projektion haben sollte. Diese Abweichungen sind mit hoher Wahrscheinlichkeit durch eine Fehlorientierung der elektronenmikroskopischen Probe oder durch eine Strahlverkippung zur optischen Achse des Mikroskops verursacht. Der numerische Bildvergleich mit simulierten, perfekt vierzählig-symmetrischen Bildern während der stochastischen Rekonstruktion erfordert ein Wiederherstellen der vierzähligen Symmetrie der experimentellen Bilder. Das Symmetrisieren der Bildinformation wurde im vorliegenden Fall durch Mitteln der durch die Symmetrie verknüpften Bildamplituden und durch eine geeignete Wahl der Phasenwerte erreicht [253]. Bei der obigen Wahl des Nullpunkts der Bildkoordinaten können die Bildfrequenzen bei perfekter Symmetrie nur Phasenwerte von Null oder π annehmen. Das Wiederherstellen der Symmetrie ist erfolgreich, solange die vorliegende Asymmetrie nicht so groß ist, daß

die gestörten Phasenwerte, die von Null oder π abweichen, nicht wieder einem der beiden Fälle zugeordnet werden können.

Das Ergebnis der drei Bildverarbeitungsschritte, die digitalisierten Bildausschnitte von 4×4 Einheitszellen, die rauschgefilterten Bilder und die symmetrisierten Bilder, sind für die beiden zur stochastischen Rekonstruktion verwendeten Bilder in Abbildung 10.6 gezeigt. Mit Hilfe eines quantitativen numerischen Bildvergleichs konnten die Defokuswerte beider Bilder zu -48 nm und -57 nm bestimmt werden.

10.2.2 Rekonstruktion der Austrittswellenfunktion

Nach der Aufbereitung der experimentellen Bildinformation durch Rauschfilterung und Symmetrisierung ist eine Rekonstruktion der Austrittswellenfunktion mittels des Simulated Annealing und des genetischen Suchalgorithmus im 8-Strahl-Modus möglich, der durch Berücksichtigung der vierzähligen Symmetrie eine effizientere Rekonstruktion ermöglicht.

Amplituden (oben) und Phasen (unten) der rekonstruierten Austrittswellenfunktion als Ergebnis des Simulated-Annealing-Algorithmus (A) und als Ergebnis des genetischen Algorithmus (B) sind in Abbildung 10.7 dargestellt. Es sind keine Unterschiede zwischen den beiden gefundenen Austrittswellenfunktionen erkennbar. Die Darstellung der Wellenfunktionen ist auf eine Auflösung von 0.18 nm begrenzt, die Bewertungsfunktion nahm in beiden Fällen einen Minimalwert von 3.8° an. Der Grund für diesen relativ hohen Wert wird in der Diskussion erörtert.

Die rekonstruierte Amplitudenverteilung ist eng mit der projizierten Struktur der kfz-Matrix der $D1_a$-Struktur korreliert. Die Amplitudenminima befinden sich nahe den Atomsäulenpositionen, und die Amplitudenmaxima befinden sich nahe den Tunnelpositionen. Schwache Kontrastdetails der Amplitudenverteilung spiegeln die Sequenz von jeweils vier Nickelatomen und einem Molybdänatom entlang der $\langle 100 \rangle$-Richtungen der kfz-Matrix wider. Die Überstruktur der in die Nickelmatrix eingebauten Molybdänatome wird in der Phasenverteilung deutlich sichtbar. Phasenmaxima in den großen kreisförmigen Gebieten korrespondieren mit Molybdänpositionen.

Neben der Rekonstruktion aus den experimentellen Bildern wurde zum Vergleich eine Dickenserie von Austrittswellenfunktionen mit Hilfe eines Standardsimulationspakets (EMS, [243]) berechnet. Durch Vergleich der rekonstruierten Austrittswellenfunktion und aller Austrittswellenfunktionen der simulierten Dickenserie konnte die beste Übereinstimmung bei einer Probendicke von 14 nm gefunden werden. Die zugehörige Amplitude (oben) und Phase (unten) der simulierten Austrittswellenfunktion ist in Abbildung 10.7 (C) gezeigt. Die wichtigsten Kontrastdetails der simulierten Amplitudenverteilung haben eine große Ähnlichkeit mit der Amplitudenverteilung der rekonstruierten Austrittswellenfunktion, die 4×1-Sequenz der Nickel- und Molybdänatome ist jedoch nicht in der simulierten Amplitudenverteilung enthalten. Die Phasenverteilung stimmt in beiden Fällen perfekt überein. Es konnte mittels Durchmustern der simulierten Dickenserie keine andere Probendicke gefunden werden, die eine ähnlich gute Übereinstimmung zwischen Simulation und Rekonstruktion zeigte.

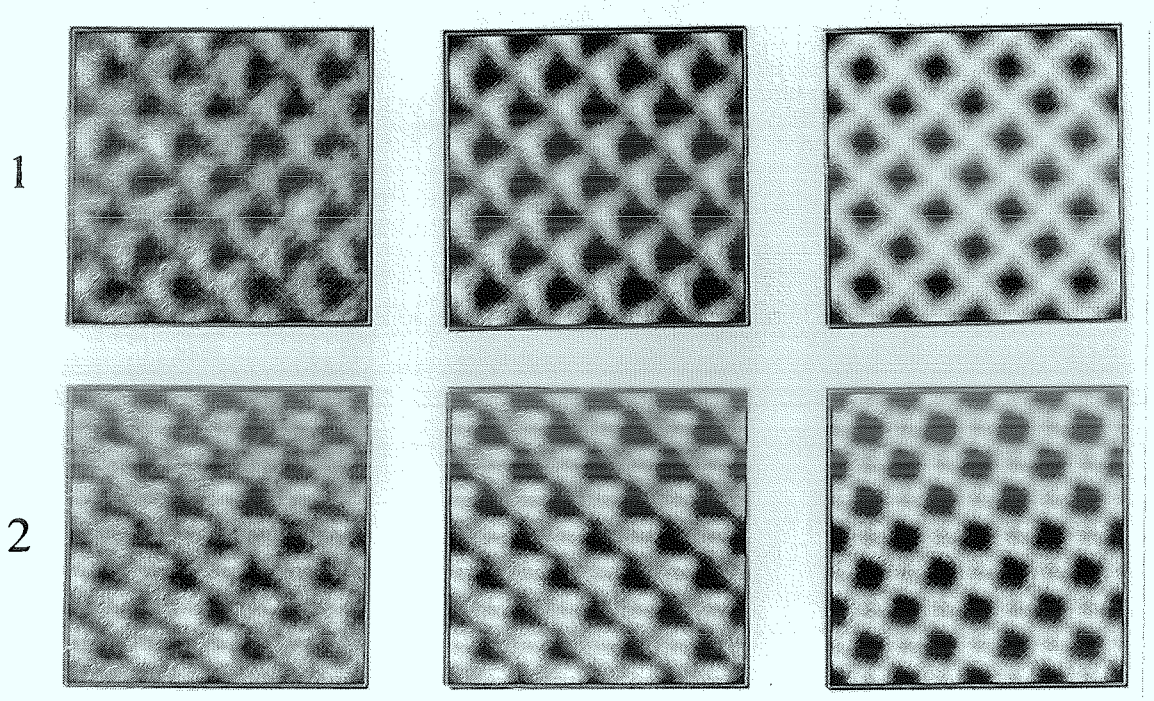


Abbildung 10.6: Aufbereitung zweier Bilder einer Defokusserie von $Ni_4Mo[001]$. Die Teilbilder zeigen jeweils 4×4 Einheitszellen. Die Ecken korrespondieren mit Molybdänpositionen. (Links) Eingabebilder 1 und 2 (256×256 Pixel). (Mitte) Eingabebilder 1 und 2 nach dem Rauschfiltern. (Rechts) Eingabebilder 1 und 2 nach der Symmetriekorrektur.

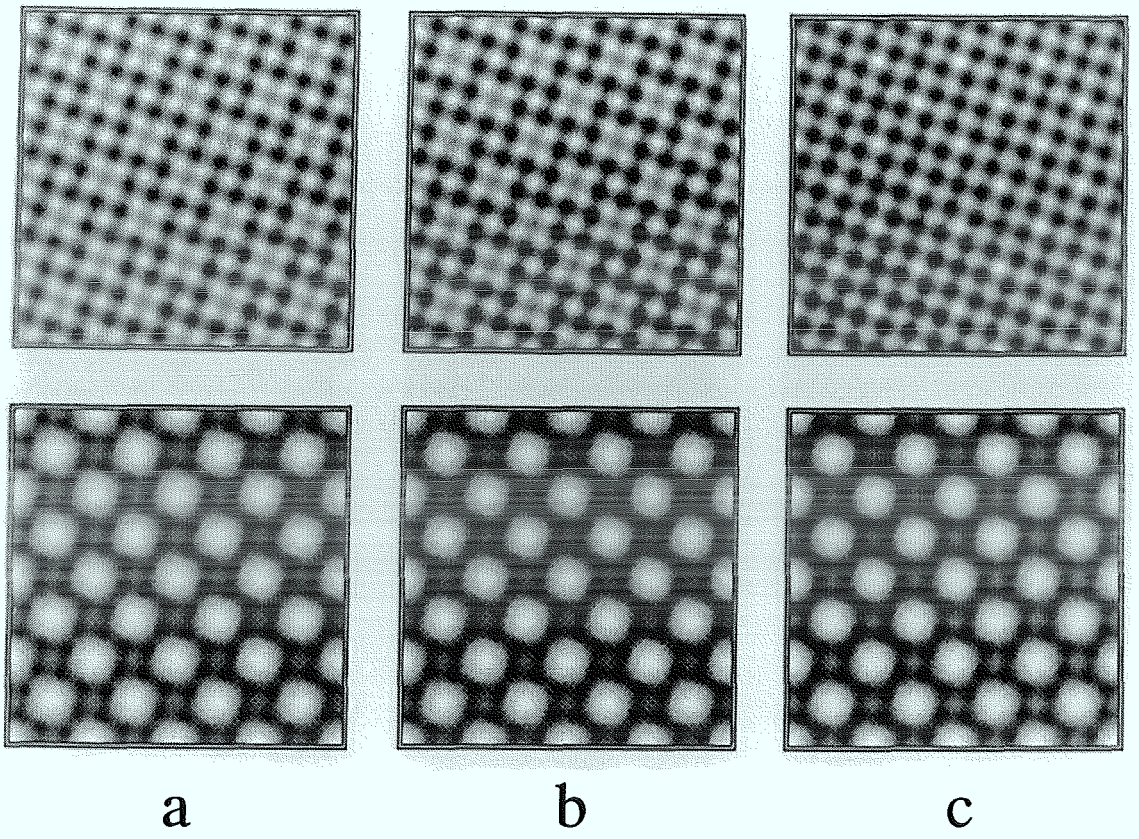


Abbildung 10.7: (A) Amplitude (oben) und Phase (unten) der mit Hilfe des Simulated Annealing aus den aufbereiteten Eingabebildern 1 und 2 rekonstruierten Austrittswellenfunktion. (B) Amplitude und Phase der rekonstruierten Austrittswellenfunktion als Ergebnis des genetischen Suchalgorithmus. (C) Die bestpassendste Austrittswellenfunktion, die mittels eines Standardsimulationspakets gefunden wurde. Die zugehörige Probendicke beträgt 14 nm.

10.3 Diskussion

Die Anwendung der stochastischen Suchverfahren, Simulated Annealing und des genetischen Suchalgorithmus, wurde in der vorliegenden Arbeit auf die Phasenrekonstruktion periodischer Bildinformation beschränkt. Diese Beschränkung ist grundsätzlich nicht durch die Struktur der stochastischen Rekonstruktionsalgorithmen verursacht, weil die Struktur unabhängig von der Art der zu optimierenden Funktion ist. Der Grund für die Behandlung periodischer Bildinformation ist die Tatsache, daß diese im Reziprokraum durch einen relativ kleinen Satz von Fourierkomponenten dargestellt werden kann. So können zum Beispiel Rekonstruktionsprobleme, deren Zahl von Freiheitsgraden nicht größer als etwa 100 ist, noch auf schnellen Workstations praktisch interaktiv gelöst werden. Die Information nichtperiodischer Bildintensitäten ist im Fourierraum wesentlich weiter ausgebreitet, so daß die Zahl der zu berücksichtigenden Fourierkomponenten, und damit der Freiheitsgrade in der Rekonstruktion, sehr viel größer sein kann. Unglücklicherweise steigt die benötigte Rechenzeit massiv mit der Zahl der zu berücksichtigenden Freiheitsgrade an.

Der erste Grund für den Anstieg der Rechenzeit ist die aufwendigere Bildberechnung, weil die Zahl der zu berechnenden Interferenzterme im ungünstigsten Fall quadratisch mit der Zahl der Strahlen ansteigt (siehe zum Beispiel Gleichung (7.40)). Der zweite Grund ist die mit der Zahl der freien Parameter ansteigende Zahl der notwendigen Berechnungen der Bewertungsfunktion, die ja unter anderem die Bildberechnungen enthält. Der funktionale Zusammenhang zwischen der Zahl der Bewertungen und der Zahl der Parameter ist a priori nicht festzulegen. Er hängt von der jeweiligen Anwendung ab, die zum Beispiel zur Vermeidung des „Einfrierens" der Suche beim Simulated Annealing unterschiedliche Abkühlraten erfordert (Abschnitt 10.1.1). Im günstigsten Fall ist die Abhängigkeit linear, nämlich dann, wenn die Parameter unabhängig voneinander in die Bewertungsfunktion eingehen. Im ungünstigsten Fall kann die Abhängigkeit exponentiell sein, nämlich bei einer vollständigen Kopplung aller Parameter untereinander. Im untersuchten Fall der Rekonstruktion mittels eines genetischen Suchalgorithmus stieg zum Beispiel die Zahl der benötigten Bewertungen beim 29-Strahl-Fall gegenüber dem 8-Strahl-Fall um einen Faktor Zehn an, während die Zahl der Strahlen nur um einen Faktor 4 größer war.

Die zahlreichen Anwendungen des Simulated-Annealing-Algorithmus und des genetischen Algorithmus auf die Phasenrekonstruktion im Rahmen dieser Arbeit zeigte, daß ihre Effizienz bei optimal eingestellten Kontrollparametern etwa gleich ist. Beide Algorithmen sind relativ einfach zu implemetieren, es gibt jedoch zwei wichtige Unterschiede, die die Einstellung der Kontrollparameter und die Erhöhung der Rechengeschwindigkeit betreffen.

Beim Simulated Annealing sind bei einer Phasenrekonstruktion zwei Kontrollparameter anzupassen, nämlich die Anfangstemperatur und die Abkühlrate. Die Anfangstemperatur kann in einem kurzen Testlauf dadurch bestimmt werden, daß die ersten Werte des Bildvergleichsmaßes fast bei 90° liegen sollten, damit die Suche nach der Austrittswellenfunktion mit einer mehr oder weniger unkorrelierten Startlösung beginnt. Sollte der Suchvorgang in einem metastabilen Zustand „ein-

frieren", dann muß die Rekonstruktion mit einer kleineren Abkühlrate wiederholt werden. Das Einstellen beider Parameter ist der Erfahrung nach einfach zu handhaben. Beim genetischen Suchalgorithmus ist die richtige Balance zwischen globaler Suche und lokaler Optimierung in der hier untersuchten einfachen Implementation von der Populationsgröße und der Mutationsrate pro Generation abhängig. Die Erfahrung zeigte, daß diese Parameter in manchen Fällen weit schwieriger abzustimmen waren, als die Kontrollparameter des Simulated Annealing.

Ein zweiter Umstand, der den Simulated-Annealing-Algorithmus zur leistungfähigeren der beiden Methoden macht, ist die Tatsache, daß von einer zur nächsten Test-Austrittswellenfunktion nur eine Änderung in *einer* Fourierkomponente vorgenommen wird, während beim genetischen Suchalgorithmus wegen des häufigen Crossing-Over erheblich größere Änderungen in mehreren Austrittswellenfunktionen einer Generation stattfinden. Im Fall der Phasenrekonstruktion ist der rechenaufwendigste Schritt die Bildberechnung, die wegen ihrer mathematischen Gestalt (Gleichung (7.40)) erheblich beschleunigt werden kann, wenn eine neu zu berechnendes Bild I_f sich von einem schon existierenden Bild I_i nur in einer Fourierkomponente der Raumfrequenz $\vec{d}$ der Wellenfunktion unterscheidet. Der Zusammenhang ist leicht einzusehen, wenn man in den Ausdruck für die Differenz beider Bilder die entsprechende Integration über alle beteiligten Interferenzterme einsetzt. In der Differenz bleiben nur solche Terme übrig, die die beiden sich unterscheidenden Strahlen $\Psi_i(\vec{d})$ und $\Psi_f(\vec{d})$ enthalten:

$$\begin{aligned} I_f(\vec{g}) - I_i(\vec{g}) \;=\; & (\Psi_f(\vec{d}) - \Psi_i(\vec{d}))\; \Psi_i^*(\vec{d}-\vec{g})\; T(\vec{d}, \vec{d}-\vec{g}, Z) + \\ & + (\Psi_f^*(\vec{d}) - \Psi_i^*(\vec{d}))\; \Psi_i(\vec{g}+\vec{d})\; T(\vec{g}+\vec{d}, \vec{d}, Z). \end{aligned} \tag{10.1}$$

Durch Abzählen der Summanden für N Raumfrequenzen $\vec{g}$ sieht man, daß $2N$ Interferenzterme neu zu berechnen sind, wenn nur eine einzige Komponente der Austrittswellenfunktion geändert wird. Dagegen erfordert eine volle Bildberechnung die Auswertung von N^2 Interferenztermen. Die Bildberechnung kann also erheblich beschleunigt werden, wenn man lediglich die Fourierkomponenten des Differenzbildes ausrechnet, und dann die entsprechenden Änderungen an den Bildfrequenzkomponenten des schon existierenden Bildes vornimmt. Diese Technik des *Differential Updating* akkumuliert jedoch bei jedem Schritt numerische Rundungsfehler, so daß in bestimmten Abständen eine volle Bildberechnung notwendig ist. Dieser zusätzliche Rechenaufwand fällt jedoch kaum ins Gewicht, weil der Erfahrung nach nur nach einigen tausend Schritten eine volle Bildberechnung einzuschieben ist. Wegen der erheblich massiveren Änderungen, die eine Testwellenfunktion während einer Rekonstruktion durch einen genetischen Algorithmus erfährt, ist das Differential-Updating dort nicht effizient durchführbar.

In Abschnitt 10.1.1 wurde ausgeführt, daß die Rekonstruktion auch unter Verwendung nur eines einzigen Eingabebildes erfolgreich durchgeführt werden konnte. Es nun zu diskutieren, ob es wirklich unter allen Umständen möglich ist, eine Austrittswellenfunktion eindeutig aus einem einzigen Eingabebild zu rekonstruieren. Diese Frage ist wegen der Komplexität der nichtlinearen Abbildung einer Austrittswellenfunktion unter partiell kohärenter Beleuchtung (Abschnitt 7.3.5) generell sehr

schwierig zu beantworten. Eine ausführliche analytische Betrachtung über die Existenz mehr als einer Lösung des Phasenrekonstruktionsproblems, wie die in [218] gegebene, war nicht Gegenstand der vorliegenden Arbeit. Vom Standpunkt der Rekonstruktion durch die beschriebenen stochastichen Suchverfahren ist jedoch eine empirische statistische Aussage zu machen, die aus der wiederholten Anwendung der Rekonstruktionsverfahren auf ein jeweils gleiches Eingabebild folgt. Da die Abtastung des Suchraums durch die stochastischen Rekonstruktionsverfahren durch Zufallsereignisse gesteuert wird (Abschnitt 9), sollte eine Doppellösung mit einer bestimmten, nichtverschwindenden Wahrscheinlichkeit auftreten. Es wurde als Ergebnis dieser Wiederholungsexperimente stets dieselbe Austrittswellenfunktion rekonstruiert, so daß zum Beispiel nach einer zehnmaligen Rekonstruktion die Existenz einer möglichen Doppellösung zumindest unwahrscheinlich ist.

Ob, neben der Eindeutigkeit der Lösung, auch wirklich immer eine Rekonstruktion der Fourierkomponenten einer Austrittswellenfunktion möglich ist, hängt in erster Linie davon ab, wie stark einzelne Fourierkomponenten die einzelnen Bildfrequenzen beeinflussen. Zum Beispiel haben Fourierkomponenten, die jenseits des Informationslimits des Elektronenmikroskops liegen, schon eine so starke Dämpfung erfahren, daß sie den Bildkontrast nicht mehr beeinflussen. Sie sind deshalb auch grundsätzlich nicht mehr durch ein Rekonstruktionsverfahren zu ermitteln. Daneben sind beim nichtlinearen Kontrasttransfer durchaus auch Situationen vorstellbar, in denen Raumfrequenzen abgebeugter Strahlen zwar innerhalb des Informationslimits liegen, die jedoch wegen der Existenz von Transferlücken weder über die linearen Interferenzterme noch über die nichtlinearen Interferenzterme zum Bildkontrast beitragen. Ein einfaches Beispiel einer solchen Transferlücke ist eine Raumfrequenz $\vec{g}$, für die die Aberrationsfunktion $\chi(\vec{g})$ (Gleichung (7.43)) gerade gleich einem Vielfachen von π ist, so daß die lineare Interferenz gerade ausgelöscht ist. Es ist nicht auszuschließen, daß der Ausschluß „wichtiger" Raumfrequenzen ein Scheitern der Rekonstruktion zur Folge haben kann. Das Problem der fehlenden Informationsübertragung durch Transferlücken kann natürlich durch die Verwendung meherer Bilder behoben werden, die bei verschiedenen Defokuswerten aufgenommen wurden.

Der geringere Einfluß von Strahlen der Austrittswellenfunktion, deren Raumfrequenzen nahe dem Informationslimit liegen, kann schon aus den numerischen Rekonstruktionsergebnissen abgelesen werden, die in den Tabellen 10.1 und 10.2 dargestellt sind. Inbesondere die Fehler der rekonstruierten Amplituden und Phasen der {330}- und der {400}-Strahlen zeigen, daß die Information dieser Raumfrequenzen weniger dominant in den Bildfrequenzen enthalten ist, als die Information der niedrigindizierten Strahlen.

Der Einfluß verschiedener Raumfrequenzen der Austrittswellenfunktion auf den Bildkontrast wurde durch ein zusätzliches numerisches Experiment untersucht, bei dem die Änderung des Bildvergleichsmaßes in Abhängigkeit von aufgeprägten Änderungen einer bekannten Austrittswellenfunktion beobachtet wurde. Als Ausgangspunkt wurde eine der perfekt rekonstruierten Austrittswellenfunktionen von Ni_4Mo genommen, und sukzessive an allen Fourierkomponenten $\Psi(\vec{g})$ entweder die Amplitude oder die Phase um ein kleinen Betrag geändert. Aus der so leicht geänderten Austrittswellenfunktion wurde jeweils ein Bild berechnet, und dieses mit dem korre-

(hkl)	$A+10\%$	$A-10\%$	$\varphi+5°$	$\varphi-5°$
(000)	4.548°	5.184°	5.545°	5.211°
(110)	4.979°	5.138°	4.305°	4.273°
(200)	4.332°	4.652°	4.057°	4.495°
(220)	1.240°	1.263°	1.953°	1.936°
(310)	1.401°	1.378°	0.888°	0.890°
(130)	2.778°	2.810°	2.114°	2.090°
(400)	0.046°	0.042°	0.034°	0.034°
(330)	0.005°	0.001°	0.002°	0.006°

Tabelle 10.3: Änderung der Bewertungsfunktion durch Fehler der Strahlamplituden der Austrittswellenfunktion von ±10 % und durch Fehler der Strahlphasen von ±5°. Die Werte wurden jeweils für einen gemeinsamen Fehler der durch die Symmetrie verbundenen Strahlen berechnet.

spondierenden Bild der ungestörten Austrittswellenfunktion korreliert. Die entsprechenden Werte des Bildvergleichsmaßes sind für alle Strahlen der Austrittswellenfunktion in Tabelle 10.3 dargestellt. Für die niedrigindizierten Strahlen bewirkt eine Änderung der Amplitudenwerte um ±10 % ein Ansteigen der Bewertungsfunktion auf etwa 1° bis 5°, eine Änderung der Phasenwerte um ±5° bewirkt einen Anstieg in der gleichen Größenordnung. Für die {330}- und die {400}-Strahlen ist die entsprechende Änderung jedoch um etwa zwei bis drei Größenordnungen kleiner. Dieses numerische Resultat bestätigt den vorher diskutierten schwachen Einfluß von Strahlen mit Raumfrequenzen nahe dem Informationslimit.

Aus diesem Resultat sind zwei Folgerungen zu ziehen: Der beobachtete Genauigkeitsverlust der rekonstruierten Austrittswellenfunktion bei hohen Raumfrequenzen ist kein Artefakt der stochastischen Rekonstruktionsverfahren, sondern er hat seine Ursache in der Dämpfung hoher Raumfrequenzen durch den Einfluß der partiell kohärenten Beleuchtung. Experimentelle Bedingungen, wie zum Beispiel Rauschen, können die Information über hohe Raumfrequenzen der Austrittswellenfunktion so weit überdecken und verfälschen, daß eine erfolgreiche Rekonstruktion solcher Strahlen nicht möglich ist. Insbesondere die vorher diskutierte Rekonstruktion aus einem einzigen Eingabebild kann unter solchen Bedingungen ebenfalls zu einem nicht eindeutigen Ergebnis führen.

In diesen Zusammenhang ist auch das Ergebnis der Rekonstruktion aus zwei Eingabebildern der experimentell aufgenommen Defokusserie zu stellen (Abschnitt 10.2). Der minimal erreichte Wert des Bildvergleichsmaßes betrug in diesem Fall 3.8°. Aus den Rekonstruktionsexperimenten, die mit simulierten Bildern als Eingabe durchgeführt wurden (Abschnitt 10.1), ist bekannt, daß die benutzten stochastischen Rekonstruktionsverfahren grundsätzlich in der Lage sind, Korrelationswerte der zur bestpassendsten Austrittswellenfunktion gehörenden Bilder zu erreichen, die bis zu zwei Größenordnungen kleiner sind. Daraus ist die Schlußfolgerung zu ziehen, daß die stochastische Rekonstruktion zwar die bestpassendste Austrittswellenfunktion ermittelt, diese aber bezüglich der beiden experimentellen Eingabebilder eine Kompromißlösung darstellt. Die beiden Eingabebilder enthalten teilweise widersprüchli-

che Information über die Austrittswellenfunktion, die zum Beispiel daher stammen kann, daß Rauschen die Bildfrequenzen teilweise verfälscht, daß die beiden Bilder Abbildungen der elektronenmikroskopischen Probe unter leicht verschiedenen Strahlverkippungen oder Probenkippungen darstellen, oder daß die benutzte Defokusdifferenz beider Bilder nicht präzise genug bestimmbar ist. Diese Fehler können unter anderem auch dafür verantwortlich sein, daß die Amplitude der rekonstruierten Austrittswellenfunktion nicht perfekt mit der bestpassendsten simulierten Austrittswellenfunktion übereinstimmt (Abbildung 10.7).

Das Problem einer nicht genau bekannten Defokusdifferenz betrifft die bei der Rekonstruktion vorzugebenden Defokuswerte der Eingabebilder. Es konnte gezeigt werden, daß die Defokusdifferenz, die bei einem Rekonstruktionsexperiment ein zunächst frei wählbarer Parameter ist, durch wiederholte Rekonstruktion der Austrittswellenfunktion mit verschiedenen Defokusdifferenzen so weit optimiert werden kann, daß ein möglichst geringer Widerspruch zwischen den experimentellen Bildern und den zur rekonstruierten Austrittswellenfunktion gehörenden Bildern besteht [245]. Bei konsistenter Bildinformation, zum Beispiel im Falle einer simulierten Defokusserie, ist die Defokusdifferenz sogar exakt zu bestimmen.

Es wäre darüber hinaus wünschenswert, durch eine analoge Serie von Rekonstruktionen der Austrittswellenfunktion den absoluten Wert des Defokus zum Beispiel des ersten Bildes einer Defokusserie zu bestimmen. Eine Betrachtung des mathematischen Modells der nichtlinearen Kontrastübertragung unter partiell kohärenter Beleuchtung (Abschnitt 7.3.4) zeigt jedoch, daß der Wert des absoluten Defokus lediglich in den Dämpfungsterm $E_S(\vec{k}, \vec{k}', Z)$ des Transmissionskreuzkoeffizienten eingeht, der bei kleinen Änderungen des Defokus nur schwach variiert. Der Propagationsterm $\exp\left(-2\pi i\chi(\vec{k}, Z)\right)$ des Transmissionskreuzkoeffizienten, der einzelnen abgebeugten Strahlen eine zusätzliche Phasenverschiebung aufprägt, hat bei einer Defokusänderung ΔZ lediglich die Wirkung eines zusätzlichen Phasenfaktors $\exp\left(-\pi i\lambda k^2\, \Delta Z\right)$ in der Austrittswellenfunktion, der wegen der Betragsquadratbildung in der Bildebene nicht meßbar ist. Es ist deshalb nicht zu erwarten, daß ein falsch gewählter absoluter Defokus bei der Rekonstruktion aus einer experimentellen Defokusserie eine genauso gravierende Inkonsistenz der Eingabedaten verursacht, wie eine falsche Defokusdifferenz, so daß die Bestimmung des absoluten Defokus nicht mit der gleichen Präzision, oder überhaupt nicht möglich ist.

Das Problem, eine genaue Kenntnis der Abbildungsparameter, wie zum Beispiel des absoluten Defokus oder der Konstanten der sphärischen Aberration, C_s, zu haben, ist allen Rekonstruktionsverfahren gemeinsam. Während die sphärische Aberration eine Gerätekonstante ist, die separat gemessen werden kann, ist die Bestimmung des absoluten Defokus aus der Bildinformation der Strukturabbildung eines stark streuenden Objektes ein bisher ungelöstes Problem. Es ist deshalb gängige Praxis, eine Defokusbestimmung entweder aus einer konventionellen Bildsimulation eines benachbarten periodischen Bildausschnitts einer bekannten Struktur abzuleiten, oder den Defokus aus Diffraktogrammen einer benachbarten, dünnen amorphen Probenstelle durch Anwenden der schwachen Phasenobjektnäherung zu gewinnen (Abschnitt 7.1.2). Die beiden Hilfsmethoden sind vom Standpunkt der Phasenrekonstruktion aus unbefriedigend, weil sie ein bestimmtes Maß an Zusatzinformation

benötigen, so daß die Entwicklung eines strukturunabhägigen Verfahrens zur Defokusbestimmung immer noch eine große Herausforderung darstellt.

Kapitel 11

Anwendung der stochastischen Potentialrekonstruktion

In diesem Kapitel wird die Anwendung eines Simulated-Annealing-Algorithmus (Abschnitt 9.2) auf die hochauflösende Rekonstruktion des projizierten Kristallpotentials aus simulierten Austrittswellenfunktionen von Galliumarsenid in der [110]-Projektion beschrieben. Das Konvergenzverhalten und die Genauigkeit des Rekonstruktionsalgorithmus wird bei drei verschiedenen Probendicken untersucht. Im ersten Abschnitt wird die Simulation der Austrittswellenfunktionen mit Hilfe des Multislice-Verfahrens (siehe Abschnitt 7.2) beschrieben. In den folgenden Abschnitten wird dargestellt, daß die Rekonstruktion inbesondere unter stark nichtlinearen Beugungsbedingungen möglich ist, die weit von der schwachen Phasenobjektnäherung (Abschnitt 7.1.2) entfernt sind. Es wird gezeigt, daß bei größeren Probendicken eine Konvergenz des Rekonstruktionsalgorithmus gegen eine alternative Lösung des Rekonstruktionsproblems eintritt. Abschließend wird die Frage nach der Genauigkeit und der Eindeutigkeit einer Potentialrekonstruktion diskutiert.

11.1 Simulation der Austrittswellenfunktionen von GaAs[110]

Der III-V-Verbindungshalbleiter Galliumarsenid hat eine Kristallstruktur vom Zinkblende-Typ mit einer Gitterkonstante der kubischen Einheitszelle von $a = 0.5654$ nm. In einer [110]-Projektion kann der Kristallstruktur eine tetragonale Einheitszelle mit den Kantenlängen 0.5654 nm und 0.4000 nm zugeordnet werden. Diese in der Abbildung 11.1 dargestellte Struktur weist die bekannte Hantelstruktur benachbarter projizierter Gallium- und Arsensäulen auf, wobei die Hantelachse entlang einer [001]-Kristallrichtung verläuft. Der kleinste Abstand zwischen den Gallium- und den Arsensäulen beträgt in dieser Projektion $a/4 = 0.1414$ nm, der dem Informationslimit von 0.12 nm eines Elektronenmikroskops vom Typ JEOL 4000EX sehr nahe kommt [199].

Mit Hilfe des Multislice-Unterprogramms, das Bestandteil des Simulationspakets EMS ist, wurden Austrittswellenfunktionen von GaAs[110] für drei verschiedene

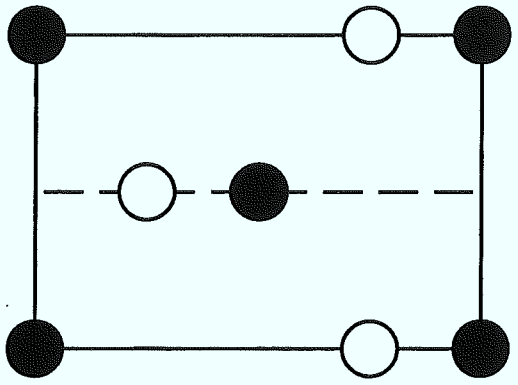


Abbildung 11.1: Projektion der Kristallstruktur von GaAs in der [110]-Projektion. Die Positionen von Galliumatomen sind durch dunkle Kreise, die von Arsenatomen durch helle Kreise angedeutet. Die Kantenlängen der projizierten Einheitszelle betragen 0.4000 nm und 0.5654 nm.

Probendicken berechnet (Abbildung 11.2). Die Simulationen wurden durchgeführt für eine Beschleunigungsspannung von 400 kV und eine Slice-Dicke von 0.4 nm. Die Fläche der tetragonalen Einheitszelle wurde mit einer Dichte von 64×64 Pixeln abgetastet, was einer Ortsauflösung von jeweils 0.0088 nm $\times$ 0.0063 nm entspricht. Für die atomaren Streufaktoren wurde die Doyle-Turner-Parametrisierung benutzt [204], und für die Debye-Waller-Faktoren der Gallium- und Arsenatome wurde der Wert $B = 0.006$ nm^2 verwendet. Die Simulation wurde für die Absorption Null durchgeführt.

Die zur ersten Austrittswellenfunktion (Abbildung 11.2 (oben)) gehörende Probendicke beträgt 0.8 nm. Dies korrespondiert mit einer Streusituation, in der der transmittierte Strahl noch eine Intensität nahe 1 hat, und in der die abgebeugten Strahlen nur schwach angeregt sind. Die zur zweiten Austrittswellenfunktion (Abbildung 11.2 (Mitte)) gehörende Probendicke beträgt 5.6 nm. Dies korrespondiert mit einer Streusituation, in der die meisten abgebeugten Strahlen eine maximale Intensität annehmen, und in der der transmittierte Strahl einen relativ niedrigen Wert erhält. Die zur dritten Austrittswellenfunktion (Abbildung 11.2 (unten)) gehörende Probendicke beträgt 11.2 nm. Bei dieser Probendicke durchlaufen die meisten abgebeugten Strahlen ein Intensitätsminimum, während der transmittierte Strahl eine Intensität nahe 1 erhält. Die dynamische Beugung führt bei den letzten beiden Probendicken zu Bedingungen, die weit entfernt von der schwachen Phasenobjektnäherung sind.

Ein Vergleich der in Abbildung 11.1 dargestellten projizierten Struktur mit den drei verschiedenen Austrittswellenfunktionen (Abbildung 11.2) zeigt, daß die Gestalt

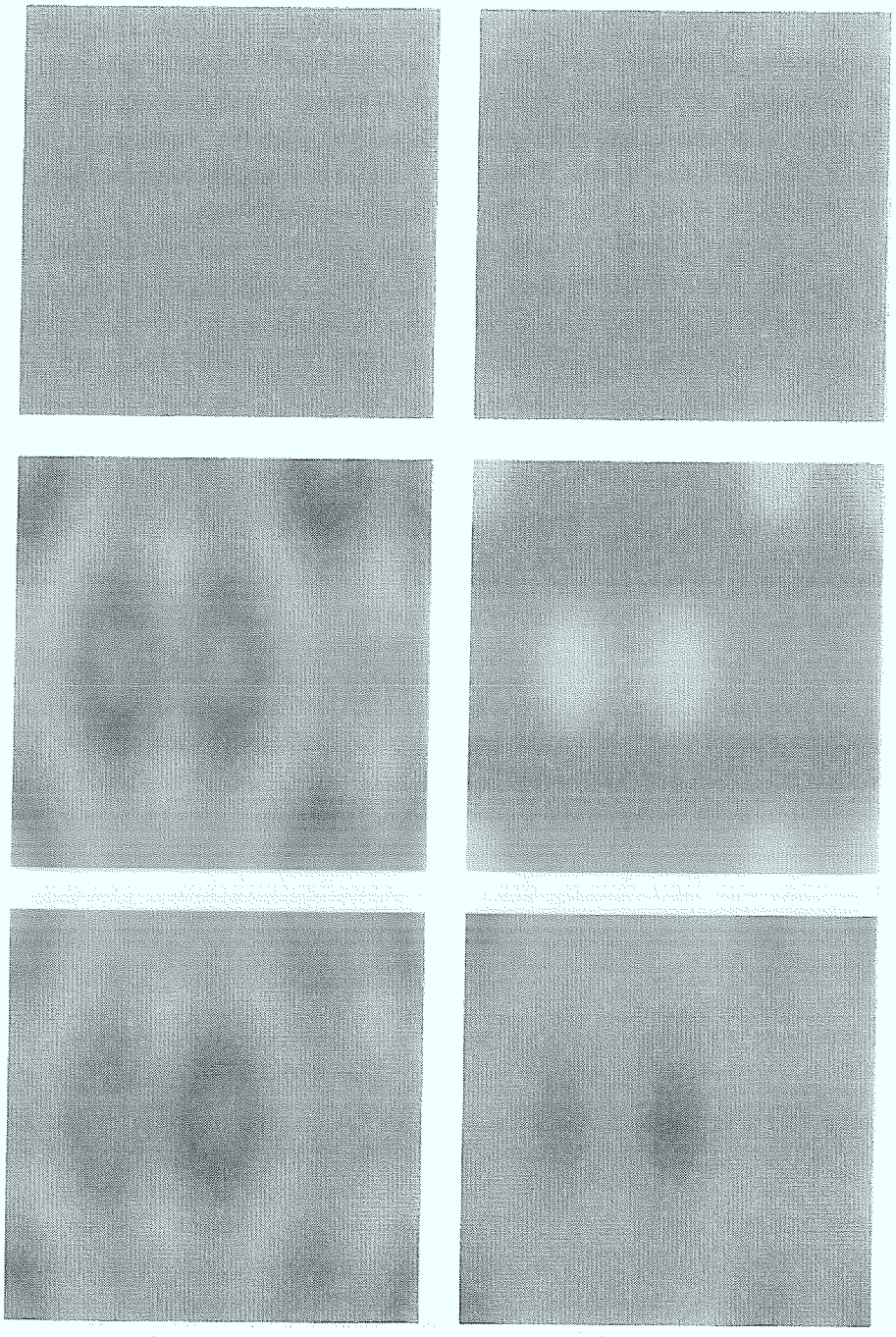


Abbildung 11.2: Simulierte Austrittswellenfunktionen von GaAs in der [110]-Projektion mit einer Ortsauflösung von 0.133 nm. Die linke Spalte zeigt die Amplitude, die rechte Spalte die Phase. Die drei verschiedenen Probendicken sind: (oben) $t = 0.8$ nm, (Mitte) $t = 5.6$ nm, (unten) $t = 11.2$ nm.

der Wellenfunktion mit der projizierten Struktur korreliert ist, daß aber gleichzeitig auch Kontrastdetails auftreten, die nicht direkt mit der Säulenstruktur in Verbindung stehen. Grundsätzlich enthält eine Austrittswellenfunktion eines stark streuenden Objektes nur bei sehr kleinen Probendicken eine direkte Strukturinformation. Bei größeren Probendicken ist die Strukturinformation mehr oder weniger in der lokalen Amplitude oder Phase der Wellenfunktion verschlüsselt.

Die Auflösung der graphischen Darstellung der Austrittswellenfunktion in Abbildung 11.2 wurde auf einen bestimmten Wert begrenzt, weil eine aus elektronenmikroskopischen Bildern rekonstruierte Austrittswellenfunktion in jedem Fall noch das Informationslimit des Elektronenmikroskops enthält (Abschnitt 10.3). In allen der in diesem Kapitel beschriebenen Rekonstruktionsexperimente wurde ein Informationslimit des Mikroskops von 0.133 nm angenommen. Dies entspricht einer Auflösung von 7.5 nm^{-1} im reziproken Raum.

11.2 Rekonstruktion des projizierten Kristallpotentials

Die in diesem Abschnitt beschriebenen Rekonstruktionsexperimente bestehen aus zwei Schritten: In einem ersten Schritt wird das projizierte Potential mit Hilfe des Simulated-Annealing-Algorithmus (Abschnitt 9.2) allein aus den simulierten Austrittswellenfunktionen (Abschnitt 11.1) ermittelt. In einem zweiten Schritt wird die rekonstruierte Lösung mit dem projizierten Potential verglichen, das ursprünglich der Ausgangspunkt zur Berechnung der Austrittswellenfunktionen gewesen ist. Da der Rekonstruktionsalgorithmus keine direkte Information über das ursprüngliche projizierte Potential hatte, kann auf diese Weise die Qualität der rekonstruierten Lösung beurteilt werden.

Zur Erleichterung der späteren Diskussion wird zunächst eine Beschreibung des ursprünglichen projizierten Potentials gegeben. Die Abbildung 11.3 zeigt die Phase der Transmissionsfunktion (Gleichung (7.37)), die proportional zum projizierten Potential pro Längeneinheit, $\Phi_P(\vec{r})$, einer Einheitszelle von GaAs[110] ist. Die Slice-Dicke wurde auf $\Delta z = 0.4$ nm gesetzt, und die Auflösung der graphischen Darstellung wurde gleich 0.056 nm (links) und 0.133 nm (rechts) gesetzt. Ein Vergleich mit der in Abbildung 11.1 dargestellten projizierten Kristallstruktur zeigt, daß die Orte der Gallium- und Arsensäulen mit Orten eines höheren Phasenwinkels korrespondieren, während an den Orten der Tunnelpositionen niedrige Werte des Phasenwinkels vorliegen. Die Transmissionsfunktion spiegelt außerdem die Spiegelsymmetrie der Kristallstruktur wider, die in der Abbildung 11.1 durch eine gestrichelte Linie angedeutet ist.

Das aus der Austrittswellenfunktion rekonstruierte projizierte Potential bei einer Probendicke von 0.8 nm ist in Abbildung 11.4 (oben) dargestellt. Der zugehörige Wert der Bewertungsfunktion (Gleichung (9.3)) ist $E = 1.32 \times 10^{-3}$ bei einem Wert des Temperaturparameters von 1.33×10^{-6}. Bei diesem niedrigen Wert der Bewertungsfunktion sind bei beiden Auflösungen der graphischen Darstellung nur minimale Unterschiede zwischen dem erzeugenden Potential und dem erfolgreich re-

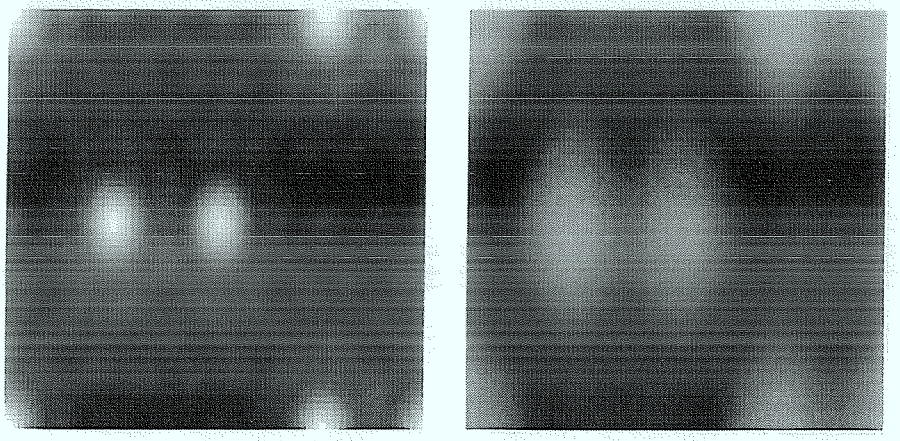


Abbildung 11.3: Phase der Transmissionsfunktion von GaAs[110], die proportional zum projizierten Kristallpotential ist. (links) Die Auflösung der graphischen Darstellung beträgt 0.056 nm. (rechts) Die Auflösung der graphischen Darstellung beträgt 0.133 nm.

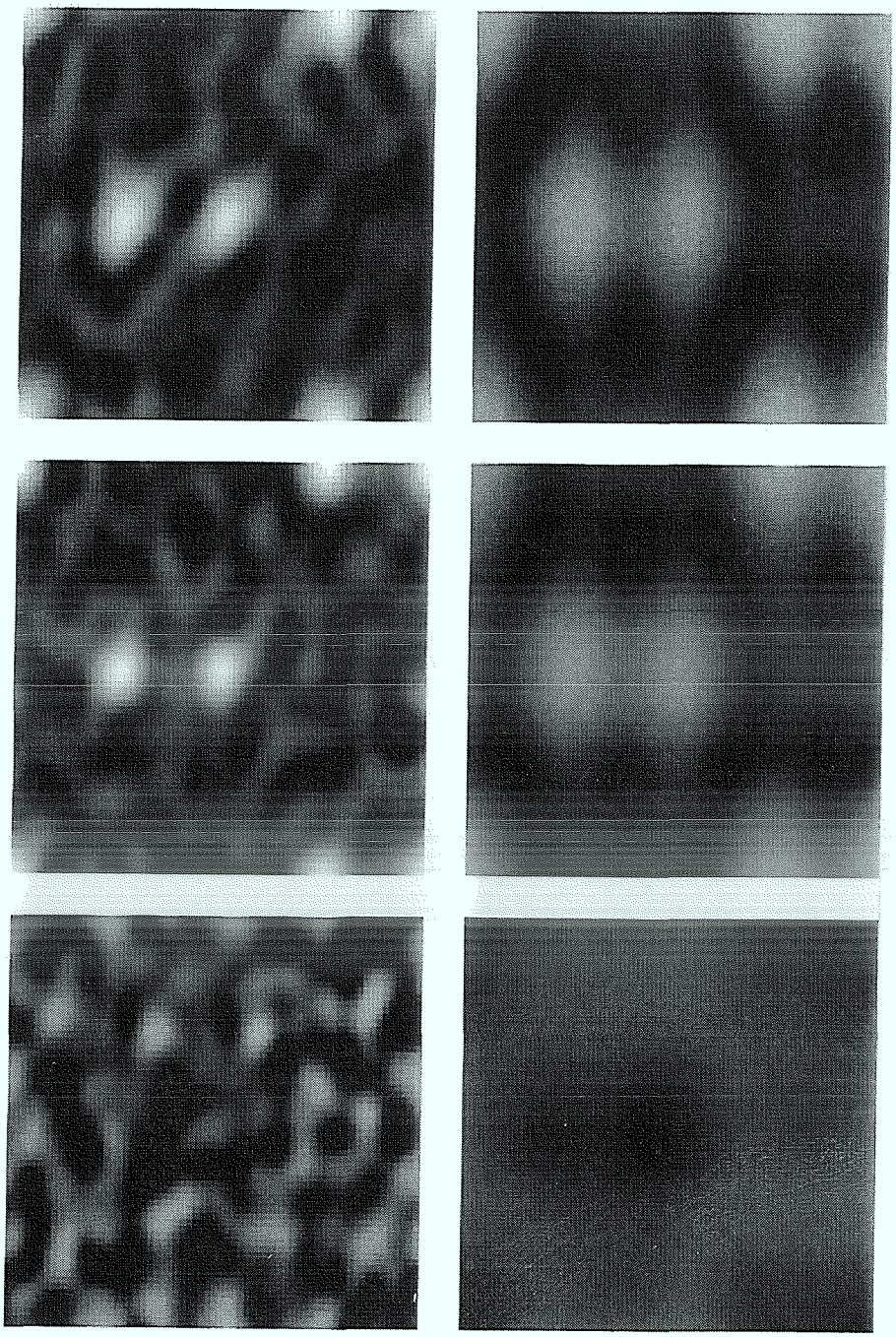


Abbildung 11.4: Das aus den simulierten Austrittswellenfunktionen rekonstruierte projizierte Potential von GaAs[110] für verschiedene Probendicken. (oben) Rekonstruktion bei einer Probendicke von 0.8 nm. (Mitte) Rekonstruktion bei einer Probendicke von 5.6 nm. (unten) Rekonstruktion bei einer Probendicke von 11.2 nm.

konstruierten Potential zu erkennen. Bei der Auflösung von 0.133 nm, die gleich dem in der Austrittswellenfunktion enthaltenen Informationslimit ist, herrscht eine perfekte Übereinstimmung zwischen dem erzeugenden und dem rekonstruierten Potential. Bei der Auflösung von 0.056 nm herrscht eine gute Übereinstimmung bezüglich der rekonstruierten Orte der Atomsäulen und der lokalen Höhe des Potentials. Zusätzlich sind kleine Rekonstruktionsfehler sichtbar, die sich in der leicht unterschiedlichen Form der Phasenmaxima und in der Anwesenheit eines Hintergrundrauschens niedriger Intensität ausdrücken.

Das aus der Austrittswellenfunktion rekonstruierte projizierte Potential bei einer Probendicke von 5.6 nm ist in Abbildung 11.4 (Mitte) dargestellt. Der zugehörige Wert der Bewertungsfunktion ist $E = 1.96 \times 10^{-3}$ bei einem Wert des Temperaturparameters von 9.06×10^{-7}. Trotz dieses niedrigen Wertes der Bewertungsfunktion sind Unterschiede zwischen dem erzeugenden Potential und dem erfolgreich rekonstruierten Potential zu erkennen. Bei der Auflösung von 0.133 nm herrscht eine perfekte Übereinstimmung zwischen dem erzeugenden und dem rekonstruierten Potential. Bei der Auflösung von 0.056 nm herrscht eine gute Übereinstimmung bezüglich der rekonstruierten Orte der Atomsäulen, der lokalen Höhe des Potentials, und der Form der Phasenmaxima. Zusätzlich sind kleine Rekonstruktionsfehler sichtbar, die sich in der Anwesenheit eines Hintergrundrauschens niedriger Intensität ausdrücken.

Die numerischen Unterschiede zwischen dem erzeugenden und den beiden erfolgreich rekonstruierten Potentialen wurden durch die Fouriertransformation der Phase der zugehörigen Transmissionsfunktionen und einen anschließenden Vergleich der Fourierkoeffizienten durchgeführt. Die Tabelle 11.1 zeigt die Listen der Fourierkoeffizienten, die bis einschließlich der {333}-Koeffizienten aufgeführt sind. Für die Rekonstruktion bei der Probendicke von $t = 0.8$ nm zeigen die Fourierkoeffizienten eine sehr gute Übereinstimmung bis zum Koeffizienten (004). Für die Rekonstruktion bei der Probendicke von $t = 5.6$ nm ist die Präzision der Fourierkoeffizienten deutlich niedriger. Oberhalb des (004)-Koeffizienten, dessen Raumfrequenz relativ nahe am Informationslimit der Eingabewellenfunktionen liegt, ist in beiden Fällen ein deutlicher Genauigkeitsverlust der rekonstruierten Fourierkoeffizienten zu beobachten. Das Ausmaß des beobachteten Hintergrundrauschens der rekonstruierten Potentiale (Abbildung 11.4 (oben), (Mitte)) kann durch Ablesen der Amplitudenwerte der kinematisch verbotenen Koeffizienten in Tabelle 11.1 abgeschätzt werden, die ja im Idealfall einer perfekten Rekonstruktion Null sein müßten.

Das aus der Austrittswellenfunktion rekonstruierte projizierte Potential bei einer Probendicke von 11.2 nm ist in Abbildung 11.4 (unten) dargestellt. Der zugehörige Wert der Bewertungsfunktion ist $E = 3.76 \times 10^{-3}$ bei einem Wert des Temperaturparameters von 1.31×10^{-5}. Trotz dieses relativ niedrigen Wertes sind bei beiden Auflösungen der graphischen Darstellung massive Unterschiede zwischen dem erzeugenden und dem rekonstruierten Potential zu erkennen. Bei der Auflösung von 0.133 nm variiert das rekonstruierte projizierte Potential nur schwach über der Fläche der Einheitszelle, mit einem flachen Minimum an der Hantelposition der Gallium- und Arsensäulen. Bei der höheren Auflösung von 0.056 nm werden schmale Maxima des Phasenwinkels der Transmissionsfunktion an Orten sichtbar, die nicht mit Säulenpositionen der GaAs[110]-Struktur korreliert sind. Darüber hinaus ist ein

	Fourierkoeffizienten $\sigma\,\Phi_P(hkl)\,\Delta z$					
	erzeugendes Potential		Rekonstruktion $t = 0.8$ nm		Rekonstruktion $t = 5.6$ nm	
(hkl)	$A \times 10^2$	ϕ[rad]	$A \times 10^2$	ϕ[rad]	$A \times 10^2$	ϕ[rad]
(001)	0.000	–	0.021	1.68	0.007	−2.75
$(\bar{1}10)$	0.000	–	0.022	−1.67	0.005	0.65
$(\bar{1}11)$	1.690	−0.84	1.675	−0.83	1.428	−0.81
$(\bar{1}1\bar{1})$	1.690	0.84	1.433	0.89	1.497	0.87
(002)	0.120	−3.14	0.128	−3.10	0.097	−2.69
$(\bar{1}12)$	0.000	–	0.024	−0.71	0.008	−1.53
$(\bar{1}1\bar{2})$	0.000	–	0.059	−0.49	0.010	0.51
$(\bar{2}20)$	1.574	0.00	1.559	−0.01	1.268	−0.03
(003)	0.000	–	0.009	1.42	0.017	−0.99
$(\bar{2}21)$	0.000	–	0.018	−0.92	0.019	−2.23
$(\bar{2}2\bar{1})$	0.000	–	0.007	−2.70	0.011	−2.12
$(\bar{1}13)$	0.942	0.82	0.954	0.82	0.743	0.73
$(\bar{1}1\bar{3})$	0.942	−0.82	0.892	−0.86	0.620	−0.75
$(\bar{2}22)$	0.046	3.14	0.044	3.13	0.042	−1.73
$(\bar{2}2\bar{2})$	0.046	−3.14	0.041	−2.77	0.079	1.71
(004)	1.069	0.00	1.070	−0.01	0.815	−0.02
$(\bar{2}23)$	0.000	–	0.008	2.12	0.024	−1.96
$(\bar{2}2\bar{3})$	0.000	–	0.016	−0.62	0.015	0.87
$(\bar{1}14)$	0.000	–	0.379	−0.23	0.070	−1.26
$(\bar{1}1\bar{4})$	0.000	–	0.074	−3.11	0.167	−0.71
$(\bar{3}30)$	0.000	–	0.001	−1.64	0.020	−2.40
$(\bar{3}31)$	0.678	−0.81	1.002	−0.81	0.648	−0.89
$(\bar{3}3\bar{1})$	0.678	0.81	0.590	0.25	1.071	0.82
$(\bar{3}32)$	0.000	–	0.333	−1.83	0.056	−1.93
$(\bar{3}3\bar{2})$	0.000	–	0.063	−1.91	0.017	1.50
$(\bar{2}24)$	0.821	0.00	1.496	−0.25	1.222	−0.18
$(\bar{2}2\bar{4})$	0.821	0.00	0.716	−0.42	1.045	−0.03
(005)	0.000	–	0.136	2.14	0.130	0.30
$(\bar{1}15)$	0.535	−0.80	0.596	−1.24	0.760	−0.91
$(\bar{1}1\bar{5})$	0.535	0.80	0.706	1.06	0.425	1.07
$(\bar{3}33)$	0.536	0.80	0.984	0.75	0.974	0.62
$(\bar{3}3\bar{3})$	0.536	−0.80	0.605	−1.32	0.792	−1.15

Tabelle 11.1: Vergleich des erzeugenden Potentials, das zur Berechnung der simulierten Austrittswellenfunktionen diente, mit den rekonstruierten Potentialen bei den Probendicken $t = 0.8$ nm und $t = 5.6$ nm. Die Amplituden und Phasen der Fourierkoeffizienten $\sigma\Phi_P(hkl)\,\Delta z$ sind mit A und ϕ bezeichnet.

Hintergrundrauschen über der gesamten Fläche der projizierten Einheitszelle zu beobachten.

Das bemerkenswerte Ergebnis der Rekonstruktion bei dieser Probendicke ist ein projiziertes Potential, das relativ zum mittleren Potential *negative* Werte an den Atomsäulenpositionen besitzt, aber dennoch die korrekte Austrittswellenfunktion liefert! Die Frage, warum die Rekonstruktion gegen diese „falsche" zweite Lösung konvergiert, wird im späteren Diskussionsteil behandelt.

Das Konvergenzverhalten des Simulated-Annealing-Algorithmus wird durch den in Abbildung 11.5 gezeigten Plot wiedergegeben, der der Rekonstruktion der Austrittswellenfunktion bei der Probendicke $t = 5.6$ nm entnommen ist. Die doppeltlogarithmische Auftragung zeigt die Abhängigkeit der Bewertungsfunktion $E(q(\vec{r}))$ vom Temperaturparameter T, der während einer Rekonstruktion langsam zu kleinen Werten hin abgesenkt wird. Der Rekonstruktionsprozeß startet bei Werten von $E \approx 10^{-1}$, die ein Fehlen jeglicher Korrelation zwischen der eingegebenen Austrittswellenfunktion und einer Austrittswellenfunktion ausdrückt, die aus einem zufällig gewählten projizierten Potential berechnet wurde. Während des Annealing-Prozesses wird dieses Testpotential mehr und mehr modifiziert, so daß es jeweils eine immer besser korrelierte Austrittswellenfunktion reproduziert. Schließlich wird der Annealing-Prozeß bei kleinen numerischen Werten der Bewertungsfunktion abgebrochen, weil während der abschließenden Optimierung der schon gefundenen Lösung keine großen Fortschritte mehr gemacht werden.

Die Potentialrekonstruktion auf der Basis eines Simulated-Annealing-Algorithmus benötigte im Fall der kleinen Probendicken $t = 0.8$ nm und $t = 5.6$ nm Rechenzeiten von einigen Stunden, im Fall der größeren Probendicke $t = 11.2$ nm eine Rechenzeit von einigen Tagen auf einer Silicon Graphics Workstation, die mit einer R4400-CPU ausgerüstet war.

11.3 Diskussion

In der vorliegenden Arbeit wurde der Simulated-Annealing-Algorithmus eingesetzt, um das projizierte Potential aus einer Austrittswellenfunktion auf der relativ kleinen Fläche einer projizierten Einheitszelle von GaAs in der [110]-Projektion zu rekonstruieren. In diesem Fall kann eine Austrittswellenfunktion und ein projiziertes Potential durch einen relativ kleinen Satz von Fourierkoeffizienten beschrieben werden. Diese notwendige Beschränkung folgt jedoch nicht aus der Struktur des Rekonstruktionsalgorithmus selbst, sondern sie folgt aus dem Anwachsen des numerischen Berechnungsaufwands, der mit der Zahl der Strahlen und mit der Zahl der einbezogenen Potentialkoeffizienten ansteigt. Im Prinzip kann der stochastische Rekonstruktionsalgorithmus auch auf Austrittswellenfunktionen auf größeren Flächen angewendet werden.

Aus drei Gründen steigt der Rechenaufwand, wenn eine größere Zahl von Fourierkomponenten berücksichtigt wird. Der Rechenaufwand ist erstens wegen der Verwendung des Multislice-Verfahrens bei der Berechnung einer Austrittswellenfunktion aus einem projizierten Potential proportional zur Zahl der Slices, und damit pro-

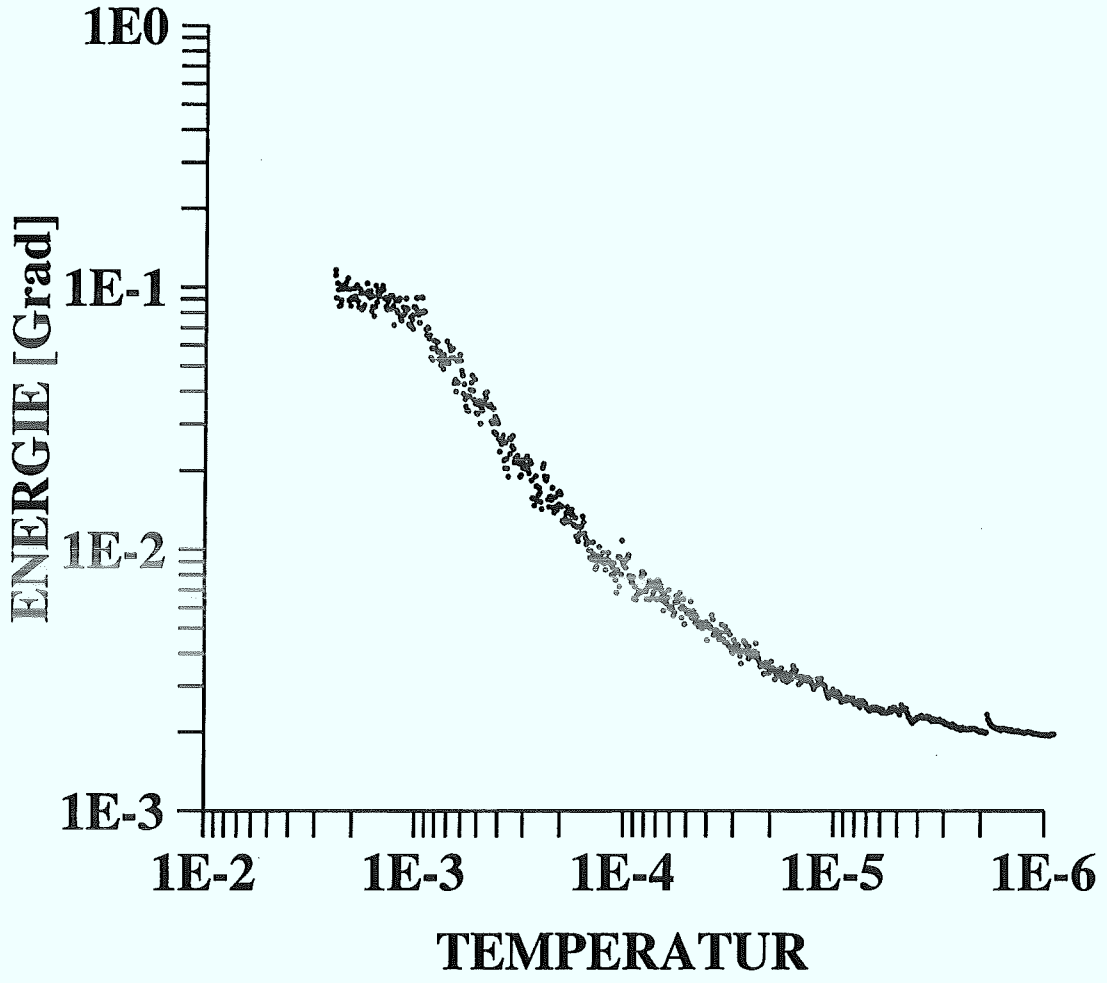


Abbildung 11.5: Konvergenzplot des Simulated-Annealing-Algorithmus der Rekonstruktion bei einer Probendicke von 5.6 nm. Der doppeltlogarithmische Plot enthält die Auftragung des Wertes der Bewertungsfunktion gegen den Temperaturkontrollparameter.

portional zur Probendicke. Zweitens erfordert jede Iteration des Multislice-Zyklus (Gleichung (7.35)) die Berechnung zweier Fouriertransformationen, deren Rechenaufwand bei der Verwendung eines FFT-Algorithmus proportional zu $N \log_2 N$ ist, wenn N Strahlen in die Rechnung einbezogen werden. Der dritte Grund ist die Notwendigkeit, daß in einem Simulated-Annealing-Algorithmus eine zunehmend kleinere „Abkühlrate" benutzt werden muß, je mehr zu bestimmende Potentialkoeffizienten, also Unbekannte, in die Rekonstruktion einbezogen werden. Die Abkühlrate muß so niedrig eingestellt werden, daß das „Einfrieren" des Systems in einem lokalen Minimum der Bewertungsfunktion verhindert wird. Dieser Effekt kann den Rechenaufwand im günstigsten Fall nur linear mit der Zahl der Unbekannten ansteigen lassen, im ungünstigsten Fall jedoch exponentiell. Die genaue Abhängigkeit ist schwierig anzugeben, denn sie hängt in der Hauptsache vom Kopplungsgrad der Unbekannten in der Bewertungsfunktion ab.

Das nächste zu diskutierende Thema ist das Problem der erreichbaren Genauigkeit eines rekonstruierten projizierten Potentials. Im vorigen Abschnitt wurde schon gezeigt, daß es eine Abhängigkeit in der Präzision der rekonstruierten Potentialkoeffizienten gibt, die von der Raumfrequenz abhängt. Bis zum (004)-Koeffizienten, der mit einer Ortsauflösung von 0.1414 nm korrespondiert, wurde eine gute Übereinstimmung zwischen dem erzeugenden Potential und dem rekonstruierten Potential beobachtet. Potentialkoeffizienten, deren zugehörige Raumfrequenzen jenseits dieses Auflösungsvermögens liegen, zeigen eine wesentlich schlechtere Übereinstimmung. Dieses Ergebnis wird durch die Auswertung der Amplituden der kinematisch verbotenen Potentialkoeffizienten bestätigt. Die entsprechenden Koeffizienten sind klein, jedoch ungleich Null, mit Amplituden, die jenseits des (004)-Koeffizienten stark ansteigen.

Diese Beobachtung kann eingeordnet werden, wenn man berücksichtigt, daß die Austrittswellenfunktion nur mit einer Ortsauflösung bekannt ist, die höchstens gleich dem Informationslimit M des Elektronenmikroskops ist. Einerseits wird eine Strukturinformation, deren Auflösung innerhalb des Informationslimits liegt, über mindestens *einen* Streuprozeß von einem Strukturfaktor $V_{\vec{g}}$, $|\vec{g}| \leq M$, in die Austrittswellenfunktion übertragen, zum Beispiel vom transmittierten Strahl $\Psi(0)$ in einen abgebeugten Strahl $\Psi(\vec{g})$. Andererseits wird eine Strukturinformation, deren Auflösung jenseits des Informationslimits liegt, über mindestens *zwei* Streuprozesse von einem Strukturfaktor $V_{\vec{g}}$, $|\vec{g}| > M$, in die Austrittswellenfunktion übertragen, zum Beispiel zuerst vom transmittierten Strahl $\Psi(0)$ zu einem abgebeugten Strahl $\Psi(\vec{g})$ und von dort dann zu einem abgebeugten Strahl $\Psi(\vec{g}')$, mit $|\vec{g}'| < M$. Somit ist es möglich, daß Strukturinformation *jenseits* des Informationslimits des Mikroskops *indirekt* in der Austrittswellenfunktion enthalten ist.

Grundsätzlich ist diese indirekte Strukturinformation schwächer in der Austrittswellenfunktion vertreten, als die direkte Information, die ihren Ursprung in Strukturdetails bei niedrigeren Raumfrequenzen hat. Der hauptsächliche Grund dafür ist die Abnahme der atomaren Streufaktoren mit zunehmender Raumfrequenz. Im Fall der direkten Übertragung von Strukturinformation unterhalb des Informationslimits sind vorwiegend große Strukturfaktoren beteiligt, während bei der Übertragung indirekter Strukturinformation in die Austrittswellenfunktion außerdem kleine

$g_{\max}$ [nm^{-1}]	$E(q(\boldsymbol{r}))$
100	9.8×10^{-7}
90	2.2×10^{-6}
80	1.3×10^{-5}
70	7.1×10^{-5}
60	3.7×10^{-4}
50	1.4×10^{-3}
40	5.6×10^{-3}
30	1.7×10^{-2}
20	3.7×10^{-2}
10	8.5×10^{-2}

Tabelle 11.2: Änderung der Bewertungsfunktion nach Ausblenden hochfrequenter Potentialanteile. Die Strukturfaktoren $\Phi_{\mathrm{P}}(\vec{g})$ des erzeugenden Potentials wurden für $|\vec{g}| > g_{\max}$ auf Null gesetzt.

Strukturfaktoren beteiligt sind. Aus diesen Betrachtungen folgt, daß die beobachtete Abnahme der Genauigkeit der Strukturfaktoren mit zunehmender Raumfrequenz eine direkte Folge der Tatsache ist, daß die Streuprozesse jenseits des Informationslimits von 0.133 nm weniger Einfluß auf die Austrittswellenfunktion haben, als die Streuprozesse innerhalb des Informationslimits.

Für diese Schlußfolgerung gibt es unter Benutzung der in der Potentialrekonstruktion verwendeten Bewertungsfunktion $E(q(\vec{r}))$ (Gleichung (9.3)) auch einen unabhängigen, objektiven numerischen Beleg. Zu diesem Zweck wurde eine Serie von Austrittswellenfunktionen von GaAs[110] für eine Probendicke von 5.6 nm berechnet, wobei jeweils die Strukturfaktoren jenseits einer bestimmten Abschneidefrequenz $g_{\max}$ auf Null gesetzt wurden. Anschließend wurden die Austrittswellenfunktionen mit der Austrittswellenfunktion verglichen, die aus dem vollen projizierten Potential berechnet worden war (Abschnitt 11.1). Die zugehörigen Werte der Bewertungsfunktion sind für eine unterschiedliche Größe der Abschneidefrequenz in der Tabelle 11.2 zusammengestellt.

Bei hohen Raumfrequenzen führt die Änderung des projizierten Potentials nur zu einer leichten Zunahme der Bewertungsfunktion, während die Änderung bei niedrigeren Raumfrequenzen eine sehr starke Zunahme zur Folge hat. Daraus folgt, daß der Einfluß der hochfrequenten Strukturfaktoren schwächer ist, als der Einfluß der niedrigfrequenten Strukturfaktoren. Deshalb ist auch im Ergebnis der Rekonstruktion ein größerer Fehler in den hochfrequenten Strukturfaktoren möglich. Es wird an dieser Stelle betont, daß dieses Resultat lediglich auf dem Gebrauch einer Bewertungsfunktion beruht, und es ist somit völlig unabhängig von einem konkreten Rekonstruktionsverfahren. Dies impliziert, daß auch die Ergebnisse anderer Potentialrekonstruktionsverfahren von diesem Umstand beeinflußt werden.

Die Diskussion über die erreichbare Genauigkeit einer Potentialrekonstruktion kann folgendermaßen zusammengefaßt werden: (1) Der Genauigkeitsverlust von rekonstruierten Strukturfaktoren bei höheren Raumfrequenzen wird durch die Natur der dynamischen Beugung, und nicht durch die Struktur eines Rekonstruktionsal-

gorithmus bewirkt. (2) Experimentelle Bedingungen, wie zum Beispiel die Existenz eines Informationslimits oder von Rauschen in einer experimentell ermittelten Austrittswellenfunktion können Strukturdetails eines rekonstruierten projizierten Potentials bei hohen Raumfrequenzen ebenfalls weiter verfälschen.

Als nächstes ist zu diskutieren, warum die stochastische Rekonstruktion bei größerer Probendicke nicht gegen das korrekte, erzeugende projizierte Potential konvergierte. Bei der Rekonstruktion der Austrittswellenfunktion bei einer Probendicke von 11.2 nm konvergierte der Simulated-Annealing-Algorithmus gegen ein projiziertes Potential, das sich erheblich vom erzeugenden Potential unterscheidet, das aber trotzdem die korrekte Austrittswellenfunktion reproduziert.

Zunächst hängt der Erfolg einer Rekonstruktion mit Hilfe des Simulated Annealing entscheidend davon ab, daß die Abkühlrate genügend klein gewählt wird (Abschnitt 9.2). In den Rekonstruktionsexperimenten bei kleineren Probendicken konnte die Abkühlrate stets so eingestellt werden, daß die korrekte Lösung des Rekonstruktionsproblems gefunden werden konnte. Daß das „Versagen" des Rekonstruktionsalgorithmus bei $t = 11.2$ nm nicht auf eine zu große Abkühlrate zurückzuführen ist, konnte in einem zusätzlichen Rekonstruktionsexperiment dadurch ausgeschlossen werden, daß testweise eine viel kleinere Abkühlrate benutzt wurde, die in einer hundertfach längeren Rekonstruktionsdauer resultierte. Trotzdem wurde wieder das gleiche „falsche" Potential als Lösung des Rekonstruktionsproblems ermittelt. Ein mögliches Problem mit einer ungünstig gewählten Abkühlrate konnte ebenfalls dadurch sicher ausgeschlossen werden, daß die Rekonstruktion bei den verschiedenen Versuchen nicht etwa gegen verschiedene Lösungen konvergierte, die jede für sich ein lokales Minimum der Bewertungsfunktion im Konfigurationsraum darstellt. Die rekonstruierte Lösung ist nämlich völlig unabhängig von einer konkret gewählten, genügend kleinen Abkühlrate.

Eine mögliche und plausible Erklärung für das Auftreten einer „Doppellösung" der vorliegenden Rekonstruktionsaufgabe ist die Tatsache, daß das Rekonstruktionsproblem zur Klasse der inversen Streuprobleme gehört, die mehr als eine Lösung haben können. Diese Interpretation deckt sich mit Beobachtungen, die in den Arbeiten [211, 197, 241] veröffentlicht wurden. In allen praktischen Fällen ist das Informationslimit eines Elektronenmikroskops nicht ausreichend groß genug, so daß eine hochauflösende Potentialrekonstruktion stets ein unterbestimmtes Anpaßproblem ist. Dies ist leicht einzusehen, indem man die Zahl der in einer rekonstruierten Austrittswellenfunktion meßbaren Fourierkoeffizienten der großen Zahl von Potentialkoeffizienten gegenüberstellt, die zu einer ausreichend genauen Beschreibung der dynamischen Beugung notwendig sind.

Im Rahmen der stochastischen Rekonstruktion ist eine besondere Eigenschaft der beiden möglichen Lösungen des Rekonstruktionsproblems abzuleiten. Da der Simulated-Annealing-Algorithmus im Fall der Probendicke $t = 11.2$ nm stets nur gegen eine der beiden Lösungen konvergierte, folgt direkt, daß die zweite Lösung durch die Wahrscheinlichkeitsverteilung (9.12) stark bevorzugt wird. Mit Berücksichtigung der Bewertungsfunktion $E(q(\vec{r}))$ folgt daraus, daß das zugehörige Minimum der Bewertungsfunktion für die zweite Lösung eine wesentlich breitere Fläche im Suchraum einnimmt, als das der korrekten Lösung des erzeugenden Potentials.

Die Unterbestimmtheit der Rekonstruktionsaufgabe resultiert unter anderem auch daher, daß das Potential aus nur einer einzigen Austrittswellenfunktion bestimmt werden muß, die nur bei einer einzigen Probendicke bekannt ist. Im Gegensatz zur Phasenrekonstruktion einer Austrittswellenfunktion aus einer Defokusserie von Bildern ist nämlich bei der Potentialrekonstruktion das Analogon einer *Dickenserie* von Austrittswellenfunktionen, die zumindest die Unterbestimmtheit durch das Vergrößern des Satzes von Meßdaten aufheben könnte, experimentell nicht realisierbar. Zwar ist es grundsätzlich möglich, eine Dickenserie von Austrittswellenfunktionen entlang einer keilförmig gedünnten, elektronenmikroskopischen Probe zu ermitteln, dies setzt allerdings voraus, daß die – unbekannte – projizierte Kristallstruktur entlang des Keils konstant ist. Gegenwärtig kann also das Problem der Doppellösungen einer hochauflösenden Potentialrekonstruktion nicht umgangen werden.

Schließlich gibt es noch ein weiteres Problem, das die Notwendigkeit betrifft, den wichtigen Beugungsparameter der Probendicke vor der Rekonstruktion kennen zu müssen. Dieses Problem ist allen Potentialrekonstruktionsverfahren gemeinsam. Es ist bekannt, daß kleine Fehler in der Probendickenbestimmung toleriert werden können, weil zum Beispiel der Effekt einer zu kleinen Probendicke durch ein etwas stärkeres Potential kompensiert werden kann, und umgekehrt. Dieser Zusammenhang kann mit Hilfe von Gleichung (7.37) eingesehen werden, indem man beachtet, daß Probendicke und projiziertes Potential dort als Produkt auftreten. Zu große Fehler in der Schätzung der Probendicke sind jedoch nicht zulässig, weil sonst die Phasendrehungen aufgrund des Propagationsterms (7.36) fehlerhaft werden. Das Problem der strukturunabhängigen Probendickenbestimmung eines stark streuenden Objektes ist bisher noch ungelöst, so daß man sich hier in der gleichen Situation befindet, wie bei der Bestimmung des absoluten Defokus bei der Phasenrekonstruktion einer Austrittswellenfunktion (Abschnitt 10.3).

Anhang A

Theorie der Zweistrahlbeugung

Die Zweistrahlbeugung ist ein Spezialfall der in Kapitel 7 dargestellten Vielstrahlbeugung hochenergetischer Elektronen an einem projizierten atomaren Kristallpotential. Die Vielstrahlbeugung verschlüsselt die Strukturinformation über das streuende Potential auf nichtlineare Weise (Gleichung (7.30)):

$$\Psi = \exp(2\pi \mathrm{i} \boldsymbol{A} z)\ \Psi', \tag{A.1}$$

wobei Ψ' den Spaltenvektor der Fourierkomponenten der Eintrittswellenfunktion, Ψ den Spaltenvektor der Fourierkomponenten der Austrittswellenfunktion, und z die Probendicke bezeichnet. Die Matrix $\boldsymbol{A}$ enthält die Anregungsfehler $s_{\vec{g}}$ der Strahlen $\Psi_{\vec{g}}$ und die Fourierkoeffizienten des projizierten Potentials $U_{\vec{g}}$:

$$A_{\vec{g},\vec{h}} = \begin{cases} s_{\vec{g}}, & \vec{g} = \vec{h} \\ U_{\vec{g}-\vec{h}}/2K, & \vec{g} \neq \vec{h} \end{cases}, \tag{A.2}$$

mit $K = 1/\lambda$ der reziproken Wellenlänge der Elektronen.

Im Fall der Zweistrahlbeugung sind nur der transmittierte Strahl und ein abgebeugter Strahl mit der Raumfrequenz $\vec{g}$ an der Beugung beteiligt, so daß das obige Gleichungssystem nur die zweikomponentigen Vektoren Ψ, Ψ' und die 2×2-Matrix $\boldsymbol{A}$ enthält. Durch Differenzieren beider Seiten der Gleichung (A.1) nach der Probendicke z erhält man eine lineare Beziehung zwischen der Wellenfunktion Ψ und ihrer Ableitung an der Stelle z:

$$\frac{\mathrm{d}\Psi}{\mathrm{d}z} = 2\pi \mathrm{i} \boldsymbol{A} \Psi. \tag{A.3}$$

Dies ist die *Howie-Whelan-Differentialgleichung der Zweistrahlbeugung* [214]. Sie erhält ihre übliche Gestalt, wenn der Koordinatenursprung des betrachteten Kristalls so gewählt wird, daß $U_{\vec{g}}$ eine reelle Größe wird, und $K/U_{\vec{g}}$ durch die *Extinktionslänge* $\xi_{\vec{g},0}$ ersetzt wird:

$$\frac{\mathrm{d}\Psi_0}{\mathrm{d}z} = \frac{\pi \mathrm{i}}{\xi_{\vec{g},0}} \Psi_{\vec{g}} \tag{A.4}$$

$$\frac{\mathrm{d}\Psi_{\vec{g}}}{\mathrm{d}z} = \frac{\pi \mathrm{i}}{\xi_{\vec{g},0}} \Psi_0 + 2\pi \mathrm{i} s_{\vec{g}} \Psi_{\vec{g}}, \tag{A.5}$$

mit der Anfangsbedingung $\Psi_0(z=0)=1$ und $\Psi_{\vec{g}}(z=0)=0$ für eine einfallende ebene Welle.

Für $s_{\vec{g}}=0$ hat das Differentialgleichungssystem eine periodische Lösung mit einer Dickenperiode gleich der Größe der Extinktionslänge $\xi_{\vec{g},0}$, so daß nach Durchlaufen der Probendicken $z=n\,\xi_{\vec{g},0}$, $n=1,2,3,\ldots$, der abgebeugte Strahl ausgelöscht ist, und der transmittierte Strahl wieder die Intensität Eins hat. Für konstantes $s_{\vec{g}}\neq 0$ zeigt die Lösung das gleiche Verhalten, jedoch ist jetzt die Dickenperiode gleich der Größe der effektiven Extinktionslänge

$$\xi_{\vec{g}}=\frac{\xi_{\vec{g},0}}{\sqrt{1+s_{\vec{g}}^2\,\xi_{\vec{g},0}^2}}. \tag{A.6}$$

Die explizite Lösung lautet:

$$\Psi_0(z) = \cos\left(\frac{\pi z}{\xi_{\vec{g}}}\right)-\mathrm{i}\cos\beta\,\sin\left(\frac{\pi z}{\xi_{\vec{g}}}\right) \tag{A.7}$$

$$\Psi_{\vec{g}}(z) = \mathrm{i}\sin\beta\,\sin\left(\frac{\pi z}{\xi_{\vec{g}}}\right), \tag{A.8}$$

mit $\cot\beta=s_{\vec{g}}\,\xi_{\vec{g},0}$.

Die Intensität des abgebeugten Strahls beträgt

$$I_{\vec{g}}(z)=|\Psi_{\vec{g}}(z)|^2=\left(\frac{\pi}{\xi_{\vec{g}}}\right)^2\frac{\sin^2(\pi s_{\mathrm{eff}}z)}{(\pi s_{\mathrm{eff}})^2}, \tag{A.9}$$

mit $s_{\mathrm{eff}}=\sqrt{s_{\vec{g}}^2+1/\xi_{\vec{g}}^2}$.

Seine eigenständige Bedeutung erhält das Howie-Whelan-Differentialgleichungssystem für den Fall eines gestörten Kristalls, dessen Gitterverzerrung durch ein ortsabhängiges Verschiebungsfeld $\vec{R}$ beschrieben werden kann (Abschnitt 2.1). Eine mit der Gitterverzerrung verknüpfte, mögliche lokale Verdrehung des Kristalls hat eine lokale Änderung des Anregungsfehlers zur Folge, der die Größe

$$\vec{g}\cdot\frac{\mathrm{d}\vec{R}}{\mathrm{d}z} \tag{A.10}$$

hat. Bei der numerischen Berechnung von $\Psi_0(z)$, $\Psi_{\vec{g}}(z)$ durch die Integration des Gleichungssystems (A.7), (A.8) ist dann statt des Anregungsfehlers $s_{\vec{g}}$ der lokale Anregungsfehler

$$s_{\vec{g}}(z)=s_{\vec{g}}+\vec{g}\cdot\frac{\mathrm{d}\vec{R}}{\mathrm{d}z} \tag{A.11}$$

einzusetzen.

Literaturverzeichnis

[1] M.S. Abrahams, J. Blanc und C.J. Buiocchi, Appl. Phys. Letters **21** (1972) 185.

[2] M.S. Abrahams, C.J. Buiocchi und G.H. Olsen, J. Appl. Phys. **46** (1975) 4259.

[3] M.S. Abrahams, L.R. Weisberg, C.J. Buiocchi und J. Blanc, J. Mater. Sci. **4** (1969) 223.

[4] S. Adachi, J. Appl. Phys. **53** (1982) 8775.

[5] T.G. Anderson, Z.G. Chen, V.D. Kulakovski, A. Uddin und J.T. Vallin, Appl. Phys. Letters **51** (1987) 752.

[6] H. Asai und K. Oe, J. Appl. Phys. **54** (1983) 2052.

[7] Y. Ashizawa, S. Akbar, W.J. Schaff, L.F. Eastman, E.A. Fitzgerald und D.G. Ast, J. Appl. Phys. **64** (1988) 4065.

[8] J.M. Ballingall, P. Ho, P.A. Martin, G.J. Tessmer, T.H. Yu, N. Lewis und E.L. Hall, J. Electron. Mater. **19** (1990) 509.

[9] S.D. Berger, D. McMullen, J. Yuan und L.M. Brown, Inst. Phys. Conf. Ser. **78** (1985) 137.

[10] D.K. Biegelsen, F.A. Ponce, A.J. Smith und J.C. Tramontana, J. Appl. Phys. **61** (1988) 1856.

[11] K.R. Breen, P.N. Uppal und J.S. Ahearn, J. Vacuum Sci. Technol. B **7** (1989) 758.

[12] L.M. Brown und G.R. Woolhouse, Phil. Mag. **21** (1970) 329.

[13] J.M. Burgers, Proc. Kon. Ned. Akad. Wetenschap. **42** (1939) 293.

[14] L. Carlson und C.G. Van Essen, J. Phys. E **7** (1974) 98.

[15] A.C. Chami, E. Ligeon, J. Fontenille und R. Danielou, J. Appl. Phys. **62** (1987) 3718.

[16] K.H. Chang, P.K. Bhattacharya und R. Gibala, J. Appl. Phys. **66** (1989) 2993.

[17] D. Cherns und M.J. Stowell, Scripta Metall. **7** (1973) 489.

[18] A.K. Chin, J. Electrochem. Soc. **82** (1982) 369.

[19] A.K. Chin, H. Temkin und R.J. Roedel, Appl. Phys. Letters **34** (1979) 476.

[20] A.K. Chin, C.L. Zipfel, B.H. Chin und M.A. DiGiuseppe, Appl. Phys. Letters **42** (1983) 1031.

[21] S. Christiansen, M. Albrecht, H. P. Strunk und P. O. Hansson, Appl. Phys. Lett. **66** (1995) 574.

[22] S.N.G. Chu, S. Nakahara, M.E. Twigg, L.A. Koszi, E.J. Flynn, A.K. Chin, B.P. Segner und W.D. Johnston Jr., J. Appl. Phys. **63** (1988) 611.

[23] T.E. Crumbaker, H.Y. Lee, M.J. Hafich und G.Y. Robinson, Appl. Phys. Letters **54** (1989) 140.

[24] B.D. Cullity, Elements of X-ray Diffraction (Addison-Wesley, Reading, 1978) 268.

[25] U. Das und P.K. Bhattacharya, J. Appl. Phys. **58** (1985) 341.

[26] S.M. Davidson, W.R.L. Clarke, T.J. Cumberbatch, E. Huang und S. Myhajlenko, Inst. Phys. Conf. Ser. **52** (1980) 351.

[27] S.M. Davidson und A. Rasul, Inst. Phys. Conf. Ser. **36** (1977) 65.

[28] D.G. Deppe, D.W. Nam, N. Holonyak Jr., K.C. Hsieh, R.J. Matyi, H. Schichijo, J.E. Epler und H.F. Chung, Appl. Phys. Letters **51** (1987) 1271.

[29] H.D. Dietze, Z. Phys. **132** (1952) 107.

[30] B. Dodson, Appl. Phys. Letters **53** (1988) 37.

[31] B.W. Dodson und J.Y. Tsao, Appl. Phys. Letters **51** (1987) 1325.

[32] B.W. Dodson und J.Y. Tsao, Appl. Phys. Letters **52** (1988) 852.

[33] A.V. Drigo, A. Aydinli, A. Carnera, F. Genova, C. Rigo, C. Ferrari, P. Franzosi und G. Salviati, J. Appl. Phys. **66** (1989) 1975.

[34] T.J. Drummond, W. Kopp, R. Fischer, H. Morkoç, R.E. Thorne und A.Y. Cho, J. Appl. Phys. **53** (1982) 1238.

[35] D.J. Dunstan, P. Kidd, P.F. Fewster, N.L. Andrew, R. Grey, J.P.R. David, L. González, Y. González, A. Sacedón und F. González-Sanz, Appl. Phys. Lett. **65** (1994) 839.

[36] D.J. Dunstan, R.H. Dixon, P. Kidd, L.K. Howard, V.A. Wilkinson, J.D. Lambkin, C. Jeynes, M.P. Halsall, D. Lancefield, M.T. Emeny, P.J. Goodhew, K.P. Homewood, B.J. Sealy und A.R. Adams, J. Crystal Growth **126** (1993) 589.

[37] D.J. Eaglesham, E.P. Kvam, D.M. Maher, C.J. Humphreys, G.S. Green, B.K. Tanner und J.C. Bean, Appl. Phys. Letters **53** (1988) 2083.

[38] D.J. Eaglesham, D.M. Maher, E.P. Kvam, J.C. Bean und C.J. Humphreys, Phys. Rev. Letters **62** (1989) 187.

[39] D.J. Eaglesham und M. Cerullo, Phys. Rev. Lett. **64** (1990) 1943.

[40] S.P. Edirisinghe, A. Staton-Bevan, P.N. Fawcett und B.A. Joyce, J. Vac. Sci. Technol. B **13** (1995) 967.

[41] B. Elman, E.S. Koteles, P. Melman, C. Jagannath, J. Lee und D. Dugger, Appl. Phys. Letters **55** (1989) 1659.

[42] M. Enatsu, M. Shimizu, T. Mizuki, K. Sugawara und T. Sakurai, Japan. J. Appl. Phys. **26** (1987) 1468.

[43] P.M. Enquist, L.P. Ramberg, P.E. Najjar, W.I. Schaff, K.L. Kavanagh, G.W. Wicks und L.F. Eastman, J. Crystal Growth **81** (1987) 378.

[44] L. Esaki und R. Tsu, IBM J. Res. Develop. **14** (1970) 61.

[45] J.D. Eshelby, Solid State Phys. **3** (1956) 79.

[46] J.D. Eshelby, in: Dislocations in Solids, Hrsg. F.R.N. Nabarro, (North-Holland, Amsterdam, 1979), 167.

[47] J.D. Eshelby, W.T. Read und W. Shockley, Acta Met. 1 (1953) 251.

[48] S.F. Fang, K. Adomi, S. Iyer, H. Morkoç, H. Zabel, C. Choi und N. Otsuka, J. Appl. Phys. **68** (1990) 31.

[49] R. Fischer, T.J. Drummond, J. Klenn, W. Kopp, T.S. Henderson, D. Perrachione und H. Morkoç, IEEE Trans. Electron Devices ED-31 (1984) 1028.

[50] E.A. Fitzgerald, J. Vacuum Sci. Technol. B **7** (1989) 782.

[51] E.A. Fitzgerald, Mater. Sci. Rep. **7** (1991) 87.

[52] E.A. Fitzgerald, D.G. Ast, P.D. Kirchner, G.D. Pettit und J.M. Woodall, J. Appl. Phys. **63** (1988) 693.

[53] E.A. Fitzgerald, P.D. Kirchner, G.D. Pettit, J.M. Woodall und D.G. Ast, Dislocations and Interfaces in Semiconductors (TMS, Warrendale, 1988) 173.

[54] E.A. Fitzgerald, P.D. Kirchner, R.E. Proano, G.D. Pettit, J.M. Woodall und D.G. Ast, Appl. Phys. Letters **52** (1988) 1496.

[55] E.A. Fitzgerald, G.P. Watson, R.E. Proano, D.G. Ast, P.D. Kirchner, G.D. Pettit und J.M. Woodall, J. Appl. Phys. **65** (1989) 2220.

[56] E.A. Fitzgerald, Y.-H. Xie, D. Brasen, M.L. Green, J. Michel, P.E. Freeland und B.E. Weir, J. Electron. Mater. **19** (1990) 949.

[57] E.A. Fitzgerald, Y.H. Xie, M.L. Green, D. Brasen, A.R. Kortan, J. Michel, Y.J. Mii und B.E. Weir, Appl. Phys. Letters **59** (1991) 811.

[58] E.A. Fitzgerald, Y.H. Xie, J. Michel, P.E. Freeland und B.E. Weir, Mater. Res. Soc. Sym. Proc. **160** (1990) 59.

[59] F.C. Frank, Phil. Mag. **42** (1951) 809.

[60] F.C. Frank und J.H. van der Merwe, Proc. Roy. Soc. A **198** (1949) 205.

[61] F.C. Frank und J.H. van der Merwe, Proc. Roy. Soc. A **198** (1949) 216.

[62] F.C. Frank und J.H. van der Merwe, Proc. Roy. Soc. A **200** (1949) 125.

[63] F.C. Frank und J.H. van der Merwe, Proc. Roy. Soc. A **201** (1949) 261.

[64] P. Franzosi, G. Salviati, F. Genova, A. Stano und F. Taiariol, Mater. Letters **3** (1985) 425.

[65] P. Franzosi, G. Salviati, F. Genova, A. Stano und F. Taiariol, J. Crystal Growth **75** (1986) 521.

[66] J. Frenkel und T. Kontorova, Phys. Z. Sowj. **13** (1938) 1.

[67] L.B. Freund, A. Bower und J.C. Ramirez, Mater. Res. Soc. Sym. Proc. **130** (1989) 139.

[68] L.B. Freund, Adv. Appl. Mech. **30** (1994) 1.

[69] I.J. Fritz, Appl. Phys. Letters **51** (1987) 1080.

[70] I.J. Fritz, P.J. Gourley und L.R. Dawson, Appl. Phys. Letters **51** (1987) 1004.

[71] I.J. Fritz, S.T, Picraux, L.R. Dawson, T.J. Drummond, W.D. Laidig und N.G. Anderson, Appl. Phys. Letters **46** (1985) 987.

[72] M. Gal, P.J. Orders, B.F. Usher, M.J. Joyce und J. Tann, Appl. Phys. Letters **53** (1988) 113.

[73] M. Gal, P.C. Taylor, B.F. Usher und P.J. Orders, J. Appl. Phys. **62** (1987) 3898.

[74] G.H. Gilmer und M.H. Grabow, J. Met. **39** (1987) 19.

[75] P.L. Gourley, R.M. Biefield und L.R. Dawson, Appl. Phys. Letters **47** (1985) 482.

[76] P.L. Gourley, I.J. Fritz und L.R. Dawson, Appl. Phys. Letters **52** (1988) 377.

[77] M. Grundmann, U. Lienert, J. Christen, D. Bimberg, A. Fischer-Colbrie und J.N. Miller, J. Vacuum Sci. Technol. B **8** (1990) 751.

[78] P. Haasen, Acta Metall. **5** (1957) 598.

[79] P. Haasen, Physikalische Metallkunde, 2. Aufl. (Springer, Berlin, Heidelberg, New York, Tokyo, 1984).

[80] W. Hagen und H. Strunk, Appl. Phys. **17** (1978) 85.

[81] M. Hastenrath, K. Lohnert und K. Kubalek, Inst. Phys. Conf. Ser. **61** (1981) 47.

[82] T. Henderson, M.I. Aksun, C.K. Peng, H. Morkoç, P.C. Chao, P.M. Smith, K.-H.G. Duh und L.F. Lester, IEEE Electron Device Letters EDL-7 (1986) 649.

[83] J.P. Hirth und J. Lothe, Theory of Dislocations, 2. Aufl. (Wiley, New York, 1982).

[84] Y. Hishida, K. Yoneda, N. Matsunami und N. Itoh, J. Appl. Phys. **62** (1987) 4460.

[85] S. Hiyamizu, T. Fujii, K. Nanbu, S. Maekawa und T. Hisatsugu, Surface Sci. **86** (1979) 137.

[86] E.M. Horl und E. Mugschl, Proceedings of the 5th European Conference on Electron Microscopy, (1972) 502.

[87] D.C. Houghton, D.D. Perovic, J.-M. Baribeau G.C. Wheatherly, J. Appl. Phys. **67** (1990) 1850.

[88] D. Hull, Introduction to Dislocations, 2. Aufl. (Pergamon Press, Oxford, 1975).

[89] R. Hull, J.C. Bean, D.J. Werder und R.E. Leibenguth, Appl. Phys. Letters **52** (1988) 1605.

[90] K. Ishida, M. Akiyama und S. Nishi, Japan. J. Appl. Phys. **26** (1987) 163.

[91] H. Ito und T. Ishibashi, Japan. J. Appl. Phys. **25** (1986) 421.

[92] S.S. Iyer, G.L. Patton, J.M.C. Stork, B.S. Meyerson und D.L. Harame, IEEE Trans. Electron Devices **36** (1989) 2043.

[93] S.C. Jain, Adv. Electron. Phys. Suppl. **24** (London, Academic Press, 1994)

[94] W.A. Jesser und D. Kuhlmann-Wilsdorf, Phys. Stat. Sol. **19** (1967) 95.

[95] W.A. Jesser und J.W. Matthews, Phil. Mag. **15** (1967) 1097.

[96] W.A. Jesser und J.W. Matthews, Phil. Mag. **17** (1968) 461.

[97] W.A. Jesser und J.W. Matthews, Phil. Mag. **17** (1968) 475.

[98] W.A. Jesser und J.W. Matthews, Phil. Mag. **17** (1968) 595.

[99] E. Kasper und H.-J. Herzog, Thin Solid Films **44** (1975) 357.

[100] H. Kato, N. Iguchi, S. Chika, M. Nakayama und N. Sano, Japan. J. Appl. Phys. **25** (1986) 1327.

[101] H. Kato, N. Iguchi, S. Chika, M. Nakayama und N. Sano, J. Appl. Phys. **59** (1986) 588.

[102] S. Katsumoto, A. Yamamoto und M. Yamaguchi, Japan. J. Appl. Phys. **24** (1985) 636.

[103] Y. Kohama, Y. Tukuda und M. Seki, Appl. Phys. Letters **52** (1988) 380.

[104] Y. Kohama, Y. Watanabe und Y. Fukuda, Japan. J. Appl. Phys. **26** (1988) 1944.

[105] S. Komiya, S. Yamazaki, Y. Kishi, I. Umebu und T. Kotani, J. Crystal Growth **61** (1983) 362.

[106] A.M. Kosevitch, in: Dislocations in Solids, Hrsg. F.R.N. Nabarro, (North-Holland, Amsterdam, 1979), 33.

[107] H. Kroemer, T.-Y. Liu und P.M. Petroff, J. Crystal Growth **95** (1989) 96.

[108] D. Kuhlmann-Wilsdorf, Phys. Rev. **120** (1960) 773.

[109] L.D. Landau und E.M. Lifshitz, Theory of Elasticity, (Pergamon, New York, 1959).

[110] H.J. Leamy, J. Appl. Phys. **53** (1982) 51.

[111] W. Lee und C.G. Fonstad, IEEE Electron Device Letters EDL-17 (1986) 683.

[112] J.C.M. Li, Acta Metall. **8** (1960) 296.

[113] J.C.M. Li, Acta Metall. **9** (1961) 384.

[114] J. Lothe, in: Elastic Strain Fields and Dislocation Mobility, Hrsg. V.L. Indenbom und J. Lothe, (North-Holland, Amsterdam, 1992), 175.

[115] J. Lothe, in: Elastic Strain Fields and Dislocation Mobility, Hrsg. V.L. Indenbom und J. Lothe, (North-Holland, Amsterdam, 1992), 329.

[116] R.M. Lum, J.K. Klingert, B.A. Davidson und M.G. Lamont, Appl. Phys. Letters **51** (1987) 36.

[117] O. Madelung, Physics of III-V Compounds, (Wiley, New York, 1964).

[118] P.M.J. Marée, J.C. Barbour, J.F. van der Veen, K.L. Kavanagh, C.W.T. Bulle-Lieuwma und M.P.A. Viegers, J. Appl. Phys. **62** (1987) 4413.

[119] J. Marek, R. Geiss, L. Glassman und M. Scott, J. Electrochem. Soc. **132** (1985) 1502.

[120] J.W. Matthews, J. Vacuum Sci. Technol. **12** (1975) 126.

[121] J.W. Matthews, Epitaxial Growth, Part B (Academic Press, New York, 1975).

[122] J.W. Matthews, in: Dislocations in Solids, Vol. 2. Dislocations in Crystals, Hrsg. F.R.N. Nabarro (North-Holland, Amsterdam 1979) 503.

[123] J.W. Matthews und A.E. Balkeslee, J. Crystal Growth **27** (1974) 118.

[124] J.W. Matthews und A.E. Balkeslee, J. Crystal Growth **29** (1975) 273.

[125] J.W. Matthews und A.E. Balkeslee, J. Crystal Growth **32** (1974) 265.

[126] J.W. Matthews und A.E. Blakeslee, J. Vacuum Sci. Technol. **14** (1977) 989.

[127] J.W. Matthews, A.E. Blakeslee und S. Mader, Thin Solid Films **33** (1976) 253.

[128] J.W. Matthews und J.L. Crawford, Thin Solid Films **5** (1970) 187.

[129] J.W. Matthews, S. Mader und T.B. Light, J. Appl. Phys. **41** (1970) 3800.

[130] F.R.N. Nabarro, Proc. Phys. Soc. **59** (1947) 256.

[131] Y. Nannichi und I. Hayashi, J. Crystal Growth **27** (1974) 126.

[132] H. Okamoto, Japan. J. Appl. Phys. **26** (1987) 315.

[133] A. Okamoto, H. Toyoshima und K. Ohata, Japan. J. Appl. Phys. **26** (1987) 539.

[134] P.J. Orders und B.F. Usher, Appl. Phys. Letters **50** (1987) 980.

[135] T.P. Pearsall und J.C. Bean, IEEE Electron Device Letters EDL-7 (1986) 308.

[136] R.E. Peierls, Proc. Phys. Soc. **52** (1940) 23.

[137] S.J. Pennycook, Inst. Phys. Conf. Ser. **61** (1981) 55.

[138] S.J. Pennycook, L.M. Brown und A.J. Craven, Phil. Mag. A **41** (1980) 589.

[139] S.J. Pennycook, A.J. Craven und L.M. Brown, Inst. Phys. Conf. Ser. **39** (1977) 69.

[140] R. People, IEEE J. Quantum Electron. QE-22 (1986) 1696.

[141] R. People und J.C. Bean, Appl. Phys. Letters **47** (1985) 322.

[142] R. People und J.C. Bean, Appl. Phys. Letters **49** (1986) 229.

[143] D.D. Perovic, G.C. Wheatherly, J.-M. Baribeau und D.C. Houghton, Thin Solid Films **183** (1989) 141.

[144] D.D. Perovic und G.C. Weatherly, Ultramicroscopy **35** (1991) 271.

[145] D.D. Perovic und D.C. Houghton, Inst. Phys. Conf. Ser. No 146 (1995) 117.

[146] P.M. Petroff, Inst. Phys. Conf. Ser. **61** (1981) 501.

[147] P.M. Petroff, D.V. Lang, J.L. Strudel und R.A. Logan, Scanning Electron Microsc. **1** (1978) 325.

[148] P.M. Petroff, R.A. Logan und A. Savage, J. Microsc. **118** (1980) 255.

[149] J.F. Petroff und M. Sauvage, J. Crystal Growth **43** (1978) 628.

[150] G. Pfefferkorn, W. Brocker und M. Hastenrath, Scanning Electron Microsc. **1** (1980) 251.

[151] S.T. Picraux, L.R. Dawson, G.C. Osbourn, R.M. Biefield und W.K. Chu, Appl. Phys. Letters **43** (1983) 1020.

[152] S.T. Picraux, L.R. Dawson, G.C. Osbourn und W.K. Chu, Appl. Phys. Letters **43** (1983) 930.

[153] G. Pindoria, R.F. Houghton, M. Hopkinson, T. Whall, R.A.A. Kubiak und E.H.C. Parker, J. Vacuum Sci. Technol. B **8** (1990) 21.

[154] K. Ploog, J. Crystal Growth **79** (1986) 887.

[155] R.H.M. van der Leur, A.J.G. Schellingerhout, F. Tuinstra und J.E. Mooij, J. Appl. Phys. **64** (1988) 3043.

[156] S.-M. Liu, M.B. Das, C.K. Peng, J. Klenn, T. Henderson, W. Kopp und H. Morkoç, J. Crystal Growth **81** (1987) 359.

[157] D.C. Radulescu, W.J. Schaff, L.F. Eastman, J.M. Ballingall, G.O. Ramseyer und S.D. Hersee, J. Vacuum Sci. Technol. B **7** (1989) 111.

[158] K. Rajan, R. Devine, W.J. Moore und P. Maigne, J. Appl. Phys. **62** (1987) 1713.

[159] L.P. Ramberg, P.M. Enquist, Y.-K. Chen, F.E. Najjar, L.F. Eastman, E.A. Fitzgerald und K.L. Kavanagh, J. Appl. Phys. **61** (1987) 1234.

[160] L. Reimer, Transmission Electron Microscopy (Springer, Berlin, 1984).

[161] S.H. Roberts, Inst. Phys. Conf. Ser. **61** (1981) 51.

[162] J.J. Rosenberg, M. Benlamri, P.D. Kirchner, J.M. Woodall und G.D. Pettit, IEEE Electron Device Letters EDL-6 (1985) 491.

[163] A.S. Sahli, H.J. Kim, R.F. Davis und G.A. Rozgonyi, Appl. Phys. Letters **46** (1985) 419.

[164] M. Shinohara, T. Ito, K. Yamada und Y. Imamura, Japan. J. Appl. Phys. **24** (1985) 711.

[165] M. Shinohara, T. Ito, K. Yamada und Y. Imamura, Japan. J. Appl. Phys. **24** (1985) 818.

[166] I.N. Stranski und L. Krastanow, Akad. Wiss. Wien, Math.-Naturwiss. Kl IIb **146** 797.

[167] A.N. Stroh, Phil. Mag. **3** (1958) 625.

[168] H. Strunk, W. Hagen und E. Bauser, Appl. Phys. **18** (1979) 67.

[169] S.M. Sze, The Physics of Semiconductor Devices (Wiley, New York, 1981) 182.

[170] M. Tachikawa und M. Yamaguchi, Appl. Phys. Letters **56** (1990) 484.

[171] H. Temkin, C.L. Zipfel und V.G. Keramidas, J. Appl. Phys. **52** (1981) 5377.

[172] A.D. Trigg, Scanning Electron Microsc. **3** (1985) 1011.

[173] C.G. Tuppen und C.J. Gibbings, Thin Solid Films **183** (1989) 133.

[174] C.G. Tuppen, C.J. Gibbings, M. Hockly und S.G. Roberts, Appl. Phys. Letters **56** (1990) 54.

[175] O. Ueda, H. Imai, T. Fujiwara, S. Yamakoshi, T. Sugawara und T. Yamaoka, J. Appl. Phys. **51** (1980) 5316.

[176] O. Ueda, I. Umebu, S. Yamakoshi und T. Kotani, J. Appl. Phys. **53** (1982) 2991.

[177] O. Ueda, K. Wakao, S. Komiya, A. Yamaguchi, S. Isozumi und I. Umebu, J. Appl. Phys. **58** (1985) 3996.

[178] O. Ueda, K. Wakao, A. Yamaguchi, S. Isozumi und S. Komiya, J. Appl. Phys. **57** (1985) 1523.

[179] O. Ueda, S. Yamakoshi, T. Sanada, I. Umebu, T. Kotani und O. Hasegawa, J. Appl. Phys. **53** (1982) 9170.

[180] J.H. van der Merwe, J. Appl. Phys. **34** (1963) 117.

[181] J.H. van der Merwe, J. Appl. Phys. **34** (1963) 123.

[182] V. Volterra, Ann. Ecole Norm. Super. **24** (1907) 400.

[183] J. Weertman und J. R. Weertman, Elementary Dislocation Theory (Macmillan, New York, 1964).

[184] S.-L. Weng, J. Appl. Phys. **66** (1989) 2217.

[185] G.J. Whaley und P.I. Cohen, J. Vacuum Sci. Technol. B **6** (1988) 625.

[186] G.J. Whaley und P.I. Cohen, Appl. Phys. Letters **57** (1990) 144.

[187] T. Won, G. Mums, M.S. Unlu, H. Unlu, J. Chyi und H. Morkoç, J. Appl. Phys. **62** (1987) 3860.

[188] B.G. Yacobi und D.B. Holt, J. Appl. Phys. bf 59 (1986) 1.

[189] K. Yagi, K. Takayanagi, K. Kabayashi und G. Honjo, J. Crystal Growth **9** (1971) 84.

[190] M. Yamaguchi, A. Yamamoto und Y. Itoh, J. Appl. Phys. **59** (1986) 1751.

[191] S. Yamakoshi, M. Abe, O. Wada, S. Komiya und T. Sakurai, IEEE J. Quantum Electron. QE-17 (1981) 167.

[192] S. Yamazaki, Y. Kishi, K. Nakajima, A. Yamaguchi und K. Akita, J. Appl. Phys. **53** (1982) 4761.

[193] S. Yamazaki, K. Nakajima, S. Komiya, Y. Hishi und K. Akita, J. Appl. Phys. **55** (1983) 3478.

[194] I. Yonenaga und K. Sumino, Phys. Stat. Sol. (a) **131** (1992) 663.

[195] T.E. Zipperian, L.R. Dawson, G.C. Osbourn und I.J. Fritz, IEDM Proceedings (1983) 696.

[196] S. Banerjee, K. Urban und M. Wilkens, Acta Met. **32**, No. 3 (1984) 299.

[197] M.J. Beeching und A.E.C. Spargo, Ultramicroscopy **52** (1993) 243.

[198] H.A. Bethe, Ann. Physik (Leipzig) **87** (1928) 55.

[199] A. Bourret und J.M. Penisson, JEOL news **25** E (1987) 2.

[200] W.M.J. Coene und T.J.J. Denteneer, Ultramicroscopy **38** (1991) 225.

[201] W. Coene, G. Janssen, M. Op de Beeck und D. Van Dyck, Phys. Rev. Lett. **69** (1992) 3743.

[202] J.M. Cowley und A.F. Moodie, Acta Cryst. **10** (1957) 609.

[203] L. Davis, Genetic Algorithms and Simulated Annealing (Pitman, London, 1987).

[204] P.A. Doyle und P.S. Turner, Acta Cryst. A **24** (1968) 390.

[205] A.J.J. Drenth, A.M. Huiser und H.A. Ferwerda, Opt. Acta **22**, No. 7 (1975) 615.

[206] J. Frank, Optik **38**, No. 5 (1973) 519.

[207] K. Fujiwara, J. Phys. Soc. Japan **16** (1961) 2226.

[208] K. Fujiwara, J. Phys. Soc. Japan **17**, Suppl. B II (1962) 118.

[209] M. Gajdardziska-Josifovska, M.R. McCartney, W.J. de Ruijter, D.J. Smith, J.K. Weiss und J.M. Zuo, Ultramicroscopy **50** (1993) 285.

[210] R.W. Gerchberg und W.O. Saxton, Optik **35**, No. 2 (1972) 237.

[211] M.A. Gribelyuk, Acta Cryst. A **47** (1991) 715.

[212] M.A. Gribelyuk und J.L. Hutchison, Ultramicroscopy **45** (1992) 127.

[213] R. Henderson, J.M. Baldwin, K.H. Downing, J. Lepault und F. Zemlin, Ultramicroscopy **19** (1986) 147.

[214] P. Hirsch, A. Howie, R. Nicholson, D.W. Pashley und M.J. Whelan, Electron Microscopy of Thin Crystals, (Krieger Publishing Company, Malabar, Florida, 1977).

[215] D. Hofmann und F. Ernst, Ultramicroscopy 53 (1994) 205.

[216] J.H. Holland, Adaption in Natural and Artificial Systems, (University of Michigan, Ann Arbor, 1975).

[217] J.J. Hu und F.H. Li, Ultramicroscopy **35** (1991) 339.

[218] A.M.J. Huiser, A.J.J. Drenth und H.A. Ferwerda, Optik **45**, No. 4 (1976) 303.

[219] K. Ishizuka, Ultramicroscopy **5** (1979) 55.

[220] R. Kilaas und R. Gronsky, Ultramicroscopy **11** (1985) 289.

[221] E.J. Kirkland, Ultramicroscopy **15** (1984) 151.

[222] E.J. Kirkland, Scanning Microsc. Suppl. **2** (1988) 139.

[223] E.J. Kirkland, B.M. Siegel, N. Uyeda und Y. Fujiyoshi, Ultramicroscopy **17** (1985) 87.

[224] S. Kirkpatrick, J. Statistical Phys. **34** (1984) 975.

[225] S. Kirkpatrick C.D. Gelatt und M.P. Vecchi, Science **220** (1984) 975.

[226] K.H. Lee, K. Hiraga, D. Shindo und M. Hirabayashi, Acta Met. **36**, No. 3 (1988) 641.

[227] M. Lentzen und K. Urban, Ultramicroscopy **62** (1996) 89.

[228] H. Lichte, Ultramicroscopy **20** (1986) 293.

[229] Y. Liu, S. Xiang, H. Fan, D. Tang, F. Li, Q. Pan, N. Uyeda und Y. Fujiyoshi, Acta Cryst. A **46** (1990) 459.

[230] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller und E. Teller, J. Chem. Phys. **21** (1953) 1087.

[231] D.L. Misell, J. Phys. D (Appl. Phys.) **6** (1973) 2200.

[232] G. Möbus und M. Rühle, Optik **93**, No. 3 (1993) 108.

[233] G. Möbus und M. Rühle, Ultramicroscopy **56** (1994) 54.

[234] M.A. O'Keefe und R. Kilaas, Scanning Microsc. Suppl. **2** (1988) 225.

[235] M. Op de Beeck, Dissertation, Antwerpen 1994.

[236] A. Ourmazd, F.H. Baumann, M. Bode und Y. Kim, Ultramicroscopy **34** (1990) 237.

[237] E. Ruedl, P. Delavignette und S. Amelinckx, Phys. Status Solidi **28** (1968) 305.

[238] O. Scherzer, J. Appl. Phys. **20** (1949) 20.

[239] P. Schiske, in: Image Processing and Computer Aided Design in Electron Optics (Academic Press, New York, 1973) S. 82.

[240] P.G. Self, M.A. O'Keefe, P.R. Buseck und A.E.C. Spargo, Ultramicroscopy **11** (1983) 35.

[241] A.E.C. Spargo, M.J. Beeching und L.J. Allen, Ultramicroscopy **55** (1994) 329.

[242] J.C.H. Spence, Acta Cryst. A **49** (1993) 231.

[243] P.A. Stadelmann, Ultramicroscopy **21** (1987) 131.

[244] L. Sturkey, Acta Cryst. **10** (1957) 858.

[245] A. Thust, Dissertation, RWTH Aachen, Aachen 1994.

[246] A. Thust und K. Urban, Ultramicroscopy **45** (1992) 23.

[247] A. Thust, M. Lentzen und K. Urban, Ultramicroscopy **53** (1994) 101.

[248] D. Van Dyck, J. Microscopy **119** (1980) 141.

[249] D. Van Dyck und M. Op de Beeck, in: Proc. 12th Int. Congr. on Electron Microscopy, Seattle, 1990, Vol. 1, S. 26.

[250] D. Van Dyck, M. Op de Beeck und W. Coene, Optik **93**, No. 3 (1993) 103.

[251] D. Van Dyck, J. Danckaert, W. Coene, E. Selderslaghs, D. Broddin, J. Van Landuyt und S. Amelinckx, in: Computer Simulation of Electron Microscope Diffraction and Images, Hrsg: W. Krakow und M. O'Keefe (The Minerals, Metals & Materials Society, 1989), S. 107.

[252] G. Van Tendeloo und S. Amelinckx, Phys. Status Solidi (a) **27** (1975) 93.

[253] D.N. Wang, S. Hovmöller, L. Kihlborg und M. Sundberg, Ultramicroscopy **25** (1988) 303.

DANKSAGUNG

Bei Herrn Prof. Dr. K. Urban bedanke ich mich herzlich für die interessante Themenstellung, für die Möglichkeit, diese Arbeit am Institut für Mikrostrukturforschung des Forschungszentrums Jülich durchführen zu können, sowie für die engagierte wissenschaftliche Unterstützung.

Weiter danke ich Frau Prof. Dr. D. Gerthsen (jetzt Universität Karlsruhe) für die intensive wissenschaftliche Zusammenarbeit, für ihre ständige Diskussionsbereitschaft und für ihre vielfältigen Anregungen zur Untersuchung der partiellen Relaxation von Übergittern.

Besonders herzlich danke ich Herrn Dr. A. Thust für die spannende Zeit während unserer Untersuchungen der stochastischen Phasenrekonstruktionsverfahren, für die Möglichkeit, einige interessante Experimente im Philips Research Lab. (Eindhoven) durchführen zu können, und für den angenehmen Stil, wissenschaftlich zu diskutieren.

Außerdem danke ich Herrn Dr. A. Förster (Institut für Schicht- und Ionentechnik, Forschungszentrum Jülich) für die freundliche Überlassung der zahlreichen Proben, einschließlich aller gewünschten Herstellungsdaten, sowie für hilfreiche Diskussionen zur Molekularstrahlepitaxie.

Schließlich bedanke ich mich herzlich bei Frau G. Waßenhoven und Frau R. Schieren für die sorgfältige Ausführung der zahlreichen Photoarbeiten, sowie bei Herrn J. Hanßen für stets gute Arbeitsbedingungen an den Mikroskopen und Hilfe bei kleinen technischen Problemen.

Allen Institutsmitgliedern, und insbesondere den zahlreichen Doktoranden und Diplomanden, danke ich nicht nur für wissenschaftliche Anregungen, sondern auch für viel Spaß und Vergnügen, die Basis einer wirklich angenehmen Atmosphäre. Zuletzt danke ich Herrn K. Chr. Tillmann, meinem Büro- und „Fehlpassungs"-Kollegen, für den anregenden westfälisch-rheinischen Disput, und für die ständige britische Tea-time mit „Lord Nelson", „Sir Winston" und „Lady Margaret".

FORSCHUNGSZENTRUM JÜLICH GmbH

KFA

Jül-3324
Dezember 1996
ISSN 0944-2952