Berechnung von STM-Profilkurven und von Quantenbillards endlicher Wandhöhe

von der Fakultät für Naturwissenschaften
der Technischen Universität Chemnitz-Zwickau
genehmigte Dissertation
zur Erlangung des akademischen Grades
„doctor rerum naturalium“
(Dr. rer. nat.)

vorgelegt von

Hartmut Sbosny

geboren am 14. Juni 1962 in Eisenhüttenstadt

eingereicht am 10. April 1995

Gutachter: Prof. Dr. Michael Hietschold, Chemnitz
Prof. Dr. Manfred Kleber, München
Prof. Dr. Günther Wilkening, Braunschweig

Tag der Verteidigung: 20. Oktober 1995

Bibliographische Beschreibung

HARMUT SBOSNY
Berechnung von STM-Profilkurven und von Quantenbillards endlicher Wandhöhe
Dissertation, Technische Universität Chemnitz-Zwickau, deutsch

Referat

Die Arbeit befaßt sich mit zweierlei: Zum einen wird der STM-Abbildungsprozeß simuliert, indem Probe und Spitze durch zweidimensionale Sommerfeld-Metalle frei wählbarer Geometrie beschrieben werden und der Tunnelstrom im Transfer-Hamiltonian-Formalismus bestimmt wird. Die Berechnung der Eigenzustände der Elektroden erfolgt numerisch durch Diskretisierung der Schrödingergleichung im Differenzenverfahren. Über die geometrische Entfaltung der erhaltenen Konstantstromprofile mit der Spitzengeometrie werden der Vergleich zum geometrischen (mechanischen) Abtasten gezogen und Möglichkeiten einer Vermessung von Spitze und Probe diskutiert.

Zum anderen wird durch Berechnung von Eigenzuständen in großen zweidimensionalen Potentialkästen (Quantenbillards) endlicher Wandhöhe der Frage nachgegangen, welchen Einfluß klassisch verbotene Gebiete (Außenraum, Tunnelbarriere) auf Eigenfunktionen in semiklassisch großen Systemen haben. Betrachtet wird insbesondere ein Gesamtsystem bestehend aus zwei Potentialkästen, die über eine Tunnelbarriere koppeln („Quantenbillards endlicher Wandhöhe im Tunnelkontakt“). Bei einer Reihe von Zuständen zeigen sich Scars, die aus der Barriere austreten und in diese zurücklaufen. Das Gesamtsystem ist in hohem Maße nichtintegrabel, „sichtbar“ wird dieses aber nur für Bahnen entweder des Kontinuums oder für komplexe Orbits. Eine semiklassische Beschreibung dieses Phänomens mit der gegenwärtigen, auf klassischen Orbits fußenden Theorie periodischer Bahnen ist nicht mehr möglich. Die Einbeziehung komplexer Orbits oder Bahnen des Kontinuums („ungebundener Orbits“) wird durch diese Ergebnisse angemahnt.

Stichwörter

Rastertunnelmikroskopie, Transfer-Hamiltonian-Formalismus, geometrische Entfaltung, Quantenbillards, Tunnelkontakt, Scars, Theorie periodischer Orbits, Differenzenverfahren, Lösung eines großen Eigenwertproblems für nichtdominante Eigenwerte

Von einem gewissen Punkt
an gibt es keine
Rückkehr mehr. Dieser Punkt
ist zu erreichen
F. Kafka

Danksagung

Danken möchte ich Herrn Prof. M. Hietschold für die freundliche Betreuung dieser Arbeit, sein nicht nachlassendes Interesse an ihrer Fertigstellung und die Unterstützung jedweder Art, die er mir während ihrer Entstehungszeit zuteil werden ließ.

In der herzlichen und zutiefst humorvollen Atmosphäre des Strukturlabors habe ich mich ausgesprochen wohlgefühlt, wofür ich allen Mitarbeiterinnen und Mitarbeitern ein herzliches Dankeschön sagen möchte.

Herrn Dr. L. Koenders von der PTB in Braunschweig bin ich zu Dank verpflichtet für sein anhaltendes Interesse am Stand der STM-Rechnungen, was kein geringer Ansporn war.

Herr Dr. P. Blaudeck war so liebenswürdig, meine betriebsgefährdend großen Datenmengen auf dem Server `physikus` mit nachsichtiger Gelassenheit drei Jahre (mehr oder minder) zu übersehen.

Inhaltsverzeichnis

1 Einführung **1**

1.1 STM und Quantenbillards? . . . 1
1.2 Der Transfer-Hamiltonian-Formalismus . . . 3
1.3 Zur Theorie des Rastertunnelmikroskops . . . 7
1.3.1 Zugänge den Transfer-Hamiltonian benutzend . . . 7
1.3.2 Zugänge nicht-störungstheoretischer Art . . . 8
1.3.3 Die klassisch-geometrische Entfaltung . . . 9
1.4 Quantenbillards . . . 9

2 Modell des Rastertunnelmikroskops **12**

2.1 Beschreibung der Elektroden . . . 12
2.2 Der Tunnelstrom . . . 14

3 Periodische Bahnen **17**

3.1 Einige Grundtatsachen . . . 17
3.2 Semiklassische Spurformeln . . . 19
3.3 Semiklassische Wellenfunktionen . . . 21
3.4 Ein spekulativer Ausflug . . . 23

4 Berechnung gebundener Zustände in mehrdimensionalen Potentialmulden **28**

4.1 Problemstellung . . . 28
4.2 Diskretisierung im Differenzenschema . . . 29
4.2.1 Vorbemerkungen . . . 29
4.2.2 Prinzipielles zum Differenzenschema . . . 30
4.2.3 Diskretisierung in zwei und drei Dimensionen . . . 32
4.3 Diskrektisierungsfehler . . . 35
4.3.1 Allgemeine Formulierung . . . 35
4.3.2 Fourierdarstellung . . . 42
4.3.3 Vollständige Auswertung im eindimensionalen Fall . . . 42
4.3.4 Zwei Ideen für eine a posteriori-Korrektur . . . 43
4.3.5 Quantitative Resultate . . . 44
4.3.6 Resümee . . . 50
4.4 Zur Lösung des Eigenwertproblems . . . 51
4.4.1 Grundprinzip . . . 51
4.4.2 Über die Matrix-mal-Vektor-Algorithmen . . . 53

5 Billardzustände **56**
5.1 Vorbemerkungen . . . 56
5.2 Exakte Resultate . . . 56
5.3 Test des numerischen Verfahrens . . . 57
5.4 Das nichtseparable Rechteck . . . 58
5.5 Zustände in Spitzengeometrien . . . 61
5.6 Zwei Quantenbillards im Tunnelkontakt . . . 62
5.7 Resümee und Ausblick . . . 64

6 Konstantstromprofile **67**
6.1 Vorbemerkungen . . . 67
6.2 Abbildung eines Grabens . . . 67
6.3 Abbildung einer Kerbe . . . 72
6.4 Abbildung eines Dreispitz . . . 74
6.5 Diskussion . . . 74
6.6 Resümee und Ausblick . . . 77

7 Zusammenfassung **79**

A Exakte Gitterlösungen **81**
A.1 Der unendlich hohe Potentialtopf als heuristisches Prinzip . . . 82
A.2 Der eindimensionale eingebettete Potentialtopf . . . 83
A.3 Verallgemeinerungen . . . 86

B Lösung des Eigenwertproblems **90**
B.1 Vorbemerkungen . . . 90
B.2 Vektor- und Teilraumiteration . . . 91
B.3 Berechnung beliebiger Eigenwerte . . . 93
B.4 Ritz-Technik und die Grundverfahren . . . 94
B.5 Die Tschebyscheff-Iteration . . . 97
B.6 Das schließliche Verfahren im Ganzen . . . 98
B.6.1 Grundschema . . . 98
B.6.2 Die Prozedur RITZIT . . . 99
B.7 Zur Konvergenzgeschwindigkeit . . . 101
B.7.1 Allgemeine Formulierung . . . 101
B.7.2 Formulierung für die quadratische Iteration . . . 102
B.7.3 Quantitative Auswertung an einem lösbaren Modellsystem . . . 104

C Linienmatrizen **108**
C.1 Definitionen . . . 108
C.2 Multiplikation zweier Linienmatrizen . . . 110
C.3 Quadrate der Diskretisierungsmatrizen . . . 114

D Die Bilder der Wellenfunktionen **116**
D.1 Spitze mit $R = 1$ Å und $\alpha = 90^o$ . . . 116
D.2 Spitze mit $R = 4$ Å und $\alpha = 90^o$ . . . 118
D.3 Spitze mit $R = 8$ Å und $\alpha = 90^o$ . . . 119
D.4 Spitze mit $R = 1$ Å und $\alpha = 120^o$ . . . 119

D.5 Ein Zweikastensystem: Rechteck + Kreis 120
D.5.1 20 Zustände bei einem Abstand von $d = 2$ Å 120
D.5.2 Ein Scar-Zustand bei einem Abstand von $d = 0$ Å 122
D.5.3 Ein Zustand bei einem Abstand von $d = 10$ Å 122

E Zur Software **123**

Literaturverzeichnis **129**

Abbildungsverzeichnis **130**

Symbole und Abkürzungen **131**

Selbständigkeitserklärung **132**

Thesen **133**

Lebenslauf **135**

Kapitel 1

Einführung

1.1 STM[1] und Quantenbillards?

Der Titel dieser Arbeit mag zunächst verwundern, haben Rastertunnelmikroskopie und Quantenbillards doch wenig miteinander gemein. Der Zusammenhang erhellt sich jedoch rasch, weiß man um Ausgangspunkt und ursprünglichen Ansatz der Arbeit.

Spätestens seitdem das Rastertunnelmikroskop [15] auch als Meßinstrument für Nanostrukturen (Nanometrologie) Interesse erfährt [40, 44], stellt sich die Frage nach der quantitativen Kalibrierung der Spitzen im Bereiche weniger Nanometer und der anschließenden, auf dieser Skala eben auch im Absolutwert aussagekräftigen Vermessung manifest nichtplanarer Probenstrukturen akut. Klarerweise tritt die Geometrie von Spitze und Probe hierbei in eine viel entschiedenere Position als beispielsweise bei der Abbildung atomar ebener Oberflächen.

Die gebräuchliche und in vielen Fällen auch hinreichende Standardtheorie der Rastertunnelmikroskopie von Tersoff/Hamann [78] besitzt den diesbezüglich allerdings nicht unkritischen Mangel, daß ihr die Annahme einer ideal planaren Probenoberfläche zugrundeliegt. Gewissermaßen das andere Extrem stellt die Reisssche Entfaltung [68] dar, die den Abbildungsprozeß als ein rein geometrisches (mechanisches) Abtasten annimmt, beliebige Geometrien erfassen könnte, aber von allen quantenmechanischen Effekten absieht. Um den hier interessierenden Zwischenbereich von 1 bis 10 nm, in dem, etwas lax gesprochen, Quantenmechanik *und* Geometrie wichtig sind, haben sich vor allem Lalouyoux et. al. [47, 46] verdient gemacht, die Tunnelströme und -dichten für verschiedene zylindersymmetrische[2] Anordnungen aus je zwei geeignet geformten Sommerfeld-Metallen berechneten. Allerdings konnte das „Scannen“ dort nicht simuliert werden, da dies die Zylindersymmetrie zerstört hätte. Die Schwierigkeiten einer theoretischen Beschreibung rühren letztlich daher, daß die nur noch numerisch mögliche vollständige quantenmechanische Behandlung geometrisch beliebig geformter Elektroden schon zweidimensional schnell an rechentechnische Grenzen führt.

Vor diesem Hintergrund und mit der ausdrücklichen Orientierung auf Fragen nach dem Einfluß der Geometrie von Spitze und Probe nähern wir uns dem rastertunnelmikroskopischen Abbildungsprozeß in dieser Arbeit in folgender Weise: Beide Elektroden werden

[1] engl. Abkürzung für Scanning Tunneling Microscope oder ∼ Microscopy

[2] ein numerisch also zweidimensionales Problem

durch zweidimensionale, räumlich begrenzte (mesoskopisch) große Potentialkästen endlicher Wandhöhe und frei wählbarer Geometrie beschrieben. Die Elektroden sind beliebig voreinander positionierbar, und der Tunnelstrom zwischen ihnen wird im Transfer-Hamiltonian-Formalismus [3] bestimmt. Die hierzu erforderlichen Eigenzustände der einzelnen Kästen berechnen wir numerisch, indem wir den jeweiligen Potentialkasten in ein größeres Grundgebiet einbetten und in diesem die stationäre Schrödingergleichung (SGL) auf einem Gitter diskretisieren. Mit diesem Ansatz sind praktisch alle interessierenden (zweidimensionalen) Geometriekonstellationen „Spitze vor Probe“ zugänglich.

Betrachtet man die obigen Elektroden aus einem etwas anderen Blickwinkel, offenbart sich eine überraschende Nähe zu den sogenannten Quantenbillards – zweidimensionalen Kästen mit eigentlich unendlich hohen Wänden – an denen bevorzugt Fragen der Integrabilität bzw. Nichtintegrabilität und den sich daraus ergebenden Konsequenzen für eine semiklassische Quantenmechanik studiert werden (Stichwort: Quantenchaos) [59, 60, 35, 16, 76]. Einen Zugang zu diesen Phänomenen gestattet allein die semiklassische Theorie periodischer Bahnen [28, 31, 12] (auch § 3 und die dortigen Zitate). Wählen wir unsere Elektroden hinreichend groß, kommt die semiklassische Problematik auch hier zum Tragen, und wir haben es offenbar mit Quantenbillards endlicher Wandhöhe zu tun. Tatsächlich zeigten die gefundenen Wellenfunktionen dann auch alle in diesem Zusammenhang bekannten Phänomene, von denen die spektakulärsten sicher die Scar-Zustände[3] [35] sind.

Die Besonderheit unserer Billards, nämlich das Vorhandensein quantenmechanisch erlaubter, klassisch aber verbotener Gebiete (Außenraum, Tunnelbarriere) förderte bei gezielteren numerischen Experimenten – „Quantenbillards im Tunnelkontakt[4]“ – dann Erscheinungen zu Tage, die mit der derzeitigen Theorie, die ganz auf klassischen Orbits fußt, aus eben diesem Grunde nur unbefriedigend bzw. überhaupt nicht beschrieben werden könnten. Dies wäre noch kaum erwähnenswert, muß eine semiklassische Theorie nicht alles erklären können, der springende Punkt ist aber, daß diese Erscheinungen ihrer ganzen Natur nach semiklassisch sind – „Scars, die eine Barriere durchqueren“ –, d. h. trotz allem durch eine Theorie periodischer Bahnen erklärt werden sollten! Die Darlegung dieser numerischen Resultate, die eine Ausdehnung der Theorie auf nichtklassische Orbits letztlich massiv anmahnen, wird einen zweiten Schwerpunkt bilden.

Die Arbeit enthält damit zwei, nimmt man die notwendigen, ausführlicheren Betrachtungen zum numerischen Verfahrens hinzu, sogar drei ihrem Umfange nach gleichrangige Komplexe. Gegliedert ist sie wie folgt:

Noch im Einführungskapitel (§ 1) wird das Rüstzeug für die STM-Problematik bereitgestellt: § 1.2 beschreibt den Transfer-Hamiltonian-Formalismus, § 1.3 gibt einen kurzen Überblick über wichtige, insbesondere hier relevante Arbeiten zur STM-Theorie.

§ 2 erörtert das verwendete Modell des Rastertunnelmikroskops gründlicher.

In § 3 wird die semiklassische Theorie periodischer Orbits im Umriß dargelegt. Dieses Kapitel schließt mit einem spekulativen Ausflug.

In § 4 beschreiben wir das numerische Verfahren im Überblick und behandeln jene Aspekte detailiert, die mit der Diskretisierung zusammenhängen. (Die Lösung des Eigenwertproblems wurde in den Anhang verbannt.) Viel Raum wird der Analyse der Diskretisierungs-

[3] Wellenfunktionen, deren Intensitäten entlang klassisch periodischer Bahnen zentriert sind

[4] Um Mißverständissen vorzubeugen: Untersucht werden die Gesamtzustände eines Systems aus zwei benachbarten Billiards, die über eine Tunnelbarriere gekoppelt sind; mit der störungstheoretischen Berechnung des Tunnelstroms bei der STM-Problematik hat dies nichts zu tun.

fehler eingeräumt, weil durch die schwierige Sachlage bei der Lösung des großen Eigenwertproblems das Diskretisierungsgitter über Gebühr weder fein sein sollte noch konnte und eine genauere Kenntnis des Einflusses der Gitterschrittweite auf die verschiedenen Aspekte der Gitterlösungen (Eigenwerte, Verhalten im Innenraum, Verhalten im Außenraum) erstrebenswert erschien.

In § 5 werden dann die gefundenen Billardzustände vorgestellt, wobei der neuartigen Problematik „Quantenbillards endlicher Wandhöhe“ das Hauptaugenmerk gilt.

§ 6 präsentiert schließlich die berechneten Konstantstromprofile für drei idealtypische Probengeometrien und verschiedene Spitzen. Insbesondere werden diese Profile je mit der aktuellen Spitzengeometrie nach Reiss [68] entfaltet und die Ergebnisse diskutiert.

Die Anhänge sind umfangreich geworden. Bis auf den Bildteil stehen sie alle mehr oder minder im direkten Bezug zum numerischen Verfahren.

In Anhang A werden exakte Gitterlösungen (Lösungen des diskretisierten Problems) abgeleitet, die bei der erwähnten Abschätzung der Diskretisierungsfehler in § 4 äußerst hilfreich sind.

Anhang B beschreibt die iterative Lösung des großen Eigenwertproblems. In einem längeren Exkurs werden zunächst verschiedene Aspekte vektoriterativer Eigenwertverfahren dargelegt und die schließlich verwendete Variante begründet. Wichtig ist die quantiative Abschätzung der Konvergenzgeschwindigkeit in B.7, die der besonderen Situation einer Iteration mit einer quadrierten Matrix Rechnung trägt.

Anhang D listet die Bilder der Wellenfunktionen auf, in Anhang C studieren wir eine spezielle Matrixform unter dem Blickwinkel eines schnellen Matrix-mal-Vektor-Algroithmus für Potenzen dieser Matrix, und zum Schluß gibt Anhang E wenigstens einen kurzen, verbalen Einblick in die rein programmtechnische Seite.

Zur Vereinfachung der Schreibweise werden in dieser Arbeit (außer in § 3) atomare Einheiten mit $\hbar = 2m_e = e^2/4\pi\epsilon_0 = 1$ verwendet; Längen sind dann in Bohrschen Radien (a_0) und Energien in Rydberg (Ryd) gegeben:

$$\begin{aligned} 1\, a_0 &= \frac{4\pi\epsilon_0}{e^2}\frac{\hbar^2}{m_e} \approx 0.5292 \times 10^{-10}\, m \\ 1\,\mathrm{Ryd} &= \frac{e^2}{4\pi\epsilon_0}\frac{1}{2a_0} \approx 2.18 \times 10^{-18}\, J = 13.606\,\mathrm{eV}\ . \end{aligned}$$

1.2 Der Transfer-Hamiltonian-Formalismus

Die Kleinheit des Tunnelstroms, genauer, dessen zumeist geringe Störung des Gleichgewichtszustandes, rechtfertigt in vielen Fällen einen störungstheoretischen Zugang. Das bekannteste Verfahren in dieser Richtung ist der Transfer-Hamiltonian-Formalismus (THF), der auf Bardeen [3] zurückgeht, und der auch den meisten Arbeiten zur STM-Theorie zugrundeliegt. Da wir ebenfalls mit dem THF arbeiten wollen, soll dieser Zugang zunächst skizziert werden, wobei etwas ausführlicher das Tunnelmatrixelement zur Sprache kommen muß, auf das wir in 2.2 infolge einer Besonderheit unseres Modellansatzes nochmals einzugehen haben. Im Sinne größtmöglicher Einfachheit beschränken wir uns hier auf das Einteilchenbild.

Der Grundgedanke der störungstheoretischen Behandlung ist die Annahme, daß die mit dem Tunnelstrom einhergehenden Veränderungen in den Elektroden näherungsweise

vernachlässigbar sein sollten. Die Eigenfunktionen und Zustandsdichten zweier im Tunnelkontakt befindlicher Elektroden können dann beschrieben werden, als wären beide Elektroden noch vollständig voneinander separiert und jede für sich im Gleichgewicht. Erst über das Hinzufügen eines Stör- oder Transferterms $\hat{H}_T$ im Hamiltonian werden Übergänge gewissermaßen nachträglich induziert; $\hat{H}_T$ ist dabei im Einteilchenbild einfach die Potentialänderung, die aus der Sicht eines (linken) Ausgangszustandes durch das Hinzuschalten des zweiten (rechten) Elektrodenpotentials hervorgerufen wird.

Seien nun $\{\psi_\nu\}$ und $\{\psi_\mu\}$ die als bekannt vorausgesetzten Eigenfunktionssysteme der linken bzw. rechten Elektrode, deren Spektren quasikontinuierlich angenommen werden,

$$\hat{H}_L\psi_\nu = E_\nu\psi_\nu\,, \quad \hat{H}_L = -\boldsymbol{\nabla}^2 + U_L\,, \qquad \langle\psi_\nu\,|\,\psi_{\nu'}\rangle = \delta_{\nu,\nu'}\,, \tag{1.1}$$

$$\hat{H}_R\psi_\mu = E_\nu\psi_\mu\,, \quad \hat{H}_R = -\boldsymbol{\nabla}^2 + U_R\,, \qquad \langle\psi_\mu\,|\,\psi_{\mu'}\rangle = \delta_{\mu,\mu'}\,. \tag{1.2}$$

U_L und U_R sind die jeweiligen Einteilchenpotentiale. Wir betrachten einen linken Ausgangszustand ψ_ν und schalten $\hat{H}_T = \Delta U_L$ ein. In erster Ordnung interessieren nur die direkten Übergänge in die rechte Elektrode, weshalb die Störungen von ψ_ν auch nur nach den „rechten“ ψ_μ entwickelt werden:

$$\Psi(t) = a(t)\,\psi_\nu\,e^{-iE_\nu t} + \sum_\mu b_{\mu\nu}(t)\,\psi_\mu\,e^{-iE_\mu t} \tag{1.3}$$

Weiter wird wie in der üblichen zeitabhängigen Störungsrechnung im kontinuierlichen Spektrum vorgegangen: Nach dem Einsetzen in die zeitabhängige SGL des Gesamtsystems,

$$i\frac{\partial}{\partial t}\Psi(t) = \hat{H}\,\Psi(t), \qquad \hat{H} = \hat{H}_L + \Delta U_L\,, \tag{1.4}$$

und mit den sukzessiven Ansätzen

$$a(t) = 1 + a^{(1)}(t) + \ldots\,, \qquad b_{\mu\nu}(t) = 0 + b^{(1)}_{\mu\nu}(t) + \ldots \tag{1.5}$$

folgen die Übergangsraten $W_{\mu\nu} = \frac{\partial}{\partial t}|b_{\mu\nu}|^2$ in erster Ordnung zu

$$W_{\mu\nu} \;=\; 2\pi\,|M_{\mu\nu}|^2\,\delta(E_\mu - E_\nu). \tag{1.6}$$

$W_{\mu\nu}$ beziffert die Übergänge je Zeiteinheit von ψ_ν nach ψ_μ und $eW_{\mu\nu}$ wäre der elektrische Strom. In der Deltafunktion kommt die Energieerhaltung zum Ausdruck (elastisches Tunneln).

Das Übergangsmatrixelement $M_{\mu\nu}$ besitzt zunächst die Form

$$M_{\mu\nu} \;=\; \langle\psi_\mu|\,\hat{H} - E_\nu\,|\psi_\nu\rangle\,, \tag{1.7}$$

worin effektiv wegen (1.1), (1.2) nur die Störpotentiale bleiben:

$$M_{\mu\nu} \;=\; \langle\psi_\mu|\,\Delta U_L\,|\psi_\nu\rangle \;=\; \langle\psi_\mu|\,\Delta U_R\,|\psi_\nu\rangle\,. \tag{1.8}$$

Im zweiten Gleichheitszeichen wurde die Hermitizität von $\hat{H}$,

$$\big\langle\psi_\mu\big|\hat{H} - E_\nu\big|\psi_\nu\big\rangle = \big\langle(\hat{H} - E_\nu)\psi_\mu\big|\psi_\nu\big\rangle,$$

sowie $E_\nu = E_\mu$, $\hat{H} = \hat{H}_R + \Delta U_R$ und ΔU_R reell benutzt. ($\hat{H}$ kann alternativ zu $\hat{H}_L + \Delta U_L$ auch aus $\hat{H}_R$ und ΔU_R aufgebaut werden, wobei ΔU_R dann die Störung aus der Sicht eines rechten Zustandes ψ_μ darstellt.) Nach (1.8) hat man die Wahl, entweder über das Gebiet

ΔU_L oder über ΔU_R zu integrieren, im wesentlichen also über das Gebiet einer kompletten Störelektrode, was bei komplizierten inneren Potentialen oft ein mühseliges Vorhaben sein dürfte.

Eine alternative Fassung des Matrixelementes und eigentlich die üblicherweise mit dem THF assoziierte, bei der die Wellenfunktionen explizit nur noch in der Barriere erforderlich sind, war daher einer der wesentlichen Punkte in BARDEENs Originalarbeit [3]: Unter Vernachlässigung kleiner ΔU_L-Anteile wird die Integration in (1.7) auf ein Gebiet Ω, das, sagen wir, rechts der Barrierenmitte liegt, beschränkt. Nunmehr kann zur Symmetrisierung von $M_{\mu\nu}$ der in Ω kleine Term $\langle\psi_\nu|\,\hat{H}-E_\mu\,|\psi_\mu\rangle^*$ in (1.7) subtrahiert werden[5],

$$M_{\mu\nu} \;=\; \int_\Omega \mathrm{d}\tau\,\left[\psi_\mu^*(\hat{H}-E_\nu)\psi_\nu \;-\; \psi_\nu(\hat{H}-E_\mu)\psi_\mu^*\right] \tag{1.9}$$

($\mathrm{d}\tau$ – Volumenelement), wonach für $E_\mu = E_\nu$ unter dem Integral nur die von der kinetischen Energie herrührenden Operatoren verbleiben:

$$M_{\mu\nu} \;=\; -\int_\Omega \mathrm{d}\tau\,\left[\psi_\mu^*\boldsymbol{\nabla}^2\psi_\nu \;-\; \psi_\nu\boldsymbol{\nabla}^2\psi_\mu^*\right]\,. \tag{1.10}$$

Mit Hilfe des GAUSSschen Satzes wird $M_{\mu\nu}$ schließlich in ein Oberflächenintegral über die Ω berandende Fläche ? überführt:

$$M_{\mu\nu} \;=\; -\iint_\Gamma \mathrm{d}\underline{\mathbf{A}}\,\left[\psi_\mu^*\boldsymbol{\nabla}\psi_\nu \;-\; \psi_\nu\boldsymbol{\nabla}\psi_\mu^*\right]\,. \tag{1.11}$$

Die Fläche ? kann im Rahmen der gemachten Näherungen dabei meist so verbogen werden, daß nur der in die Barriere liegende Teil einen Beitrag liefert. (Eine Ausnahme findet sich in unserem Modell, Abschnitt 2.2.)

Der Schritt von (1.10) zu (1.11) wurde hier für den dreidimensionalen Fall angeschrieben – im Eindimensionalen reduziert sich (1.10) auf eine einfache partielle Integration und im Zweidimensionalen hätte man

$$M_{\mu\nu} \;=\; -\oint_C \mathrm{d}s\,\underline{\mathbf{n}}\,\left[\psi_\mu^*\boldsymbol{\nabla}\psi_\nu \;-\; \psi_\nu\boldsymbol{\nabla}\psi_\mu^*\right]\,, \tag{1.12}$$

wo $\boldsymbol{\nabla}$ der zweidimensionale Nabla-Operator ist, ds das Bogenelement und $\underline{\mathbf{n}}$ der Normalenvektor der berandenden Kurve C.

Zur Gültigkeit: Die Störpotentiale sind beim Transfer-Hamiltonian-Formalismus nicht klein gegenüber dem Ausgangssystem, sondern von vergleichbarer Größe. Die Anwendbarkeit der Störungsrechnung, d. h. die Kleinheit der Matrixelemente beruht hier vielmehr darauf, daß die Wellenfunktionen durch den raschen Abfall im Außenraum bereits exponentiell klein sind (sein müssen) im gestörten Gebiet.

Diese Bedingung ist in Gefahr bei zu schmalen oder zu niedrigen Barrieren, und auch, wenn die gesamte Barriere von der Störung erfaßt wird. Letzteres wäre z. B. der Fall, wenn versucht wird, die zusätzliche Barrierenabsenkung, die durch die Bildkraft der zweiten Elektrode entsteht, über das Störpotential zu berücksichtigen. (D. h. wenn die Wellenfunktionen

[5] Der hinzugefügte Term ist von gleicher Größeordnung wie der durch die Beschränkung auf Ω vernachlässigte. Wann aber ist die Umformung exakt? Es wird praktisch $\Delta U_L = 0$ gesetzt außerhalb Ω und $\Delta U_R = 0$ innerhalb. In Strenge ist dies möglich, wenn es in der Barriere einen Bereich (oder wenigstens eine Fläche, einen Punkt im Eindimensionalen) gibt, zu dessen Linken ΔU_L verschwindet und zu dessen Rechten ΔU_R. Innerhalb dieses Bereiches sollte das BARDEENsche Matrixelement genommen werden. Für Rechteckbarrieren ist die Umformung auf der gesamten Barrierenbreite exakt.

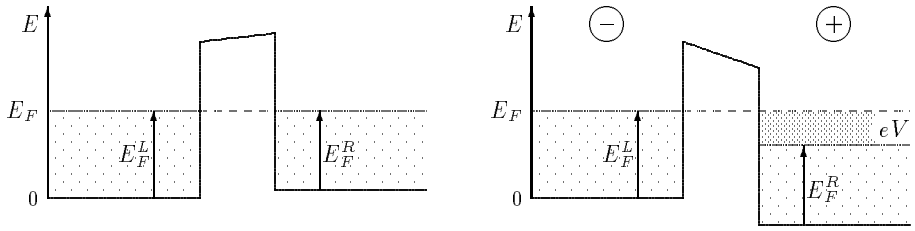


Abbildung 1.1: *Links Potential und Besetzungswahrscheinlichkeiten im stromlosen Gleichgewichtszustand bei der Tunnelspannung $V = 0$, rechts für $V > 0$, wobei die eingezeichnete Polarität gilt.*

für den Außenraum einer Einzelelektrode berechnet werden, die Matrixelemente aber für den Außenraum eines Zwei-Elektrodenproblems.) Besser wäre es in einem solchen Fall, das Potential der Einzelelektrode bereits an das Gesamtpotential bis zum Barrierenscheitel anzuschmiegen und danach geeignet fortzusetzen.

Letztlich sollte die Anwendbarkeit des THF immer dann in Frage gestellt sein, wenn die Tunnelströme nicht mehr als klein betrachtet werden können, eine natürlich mehr literarische Umschreibung des Problems. Als quantitatives Maß kann die Betragsgröße des Störterms in (1.3) gegen Eins genommen werden. Es ist an dieser Stelle aber anzumerken, daß im Gegensatz zu seiner weiten Verwendung (und seinem guten Funktionieren) die Beziehung des THF zu formal exakten Theorien nicht gut verstanden ist. Um FEUCHTWANG und CUTLER [26] aus dem Jahre 1987 zu zitieren: „To date no consistent, let alone unique, solution of this troublesome problem has been found". Eine Herleitung des THF aus einer exakten Darstellung heraus in [65] erforderte unerwartet diffizile Annahmen.

Der Tunnelstrom (als Summe von Hin- und Rückstrom) folgt aus den Übergangsraten (1.6) schließlich durch Summation über alle Anfangs- und Endzustände, wobei es deren Besetzungswahrscheinlichkeiten zu berücksichtigen gilt; die Kleinheit des Tunnelstroms rechtfertigt in der Regel die Gleichgewichtsverteilung (Fermi-Verteilung):

$$f(E) = \frac{1}{e^{(E-E_F)/k_BT} + 1}. \tag{1.13}$$

Im stromlosen Gleichgewichtszustand, bei der Tunnelspannung $V = 0$, kompensieren sich Hin- und Rückstrom exakt, was gleichbedeutend ist mit einem gemeinsamen Fermi-Niveau E_F in beiden Elektroden (Abb. 1.1, links). Eine externe Tunnelspannung $V \neq 0$ verschiebt dann ausgehend von dieser Gleichgewichtslage beide Elektroden energetisch gegeneinander um den Betrag eV (Abb. 1.1, rechts). Wie diese Verschiebung formal zu berücksichtigen ist, hängt davon ab, auf welche Energieskalen E_ν und E_μ bezogen werden. Mißt man E_ν und E_μ wie oben bei der Ableitung des THF in einer gemeinsamen Skala – andernfalls wäre dort nicht $\delta(E_\mu - E_\nu)$ entstanden – fließt eV über die Argumente der Fermi-Funktionen ein; bleibt die Skala dabei z. B. wie in Abb. 1.1 fest mit der linken Elektrode (mit den E_ν) verbunden, entsteht:

$$I = 2\pi e \sum_{\mu,\nu} |M_{\mu\nu}|^2 \left[f(E_\nu) - f(E_\mu + eV) \right] \delta(E_\mu - E_\nu) \,. \tag{1.14}$$

Diese Stromformulierung[6] ist im übrigen unabhängig von der Art und Weise, wie die Raten

ermittelt wurden, und nicht an den THF gebunden. Auch ein exakt lösbares elastisches Streuproblem kann auf diese Weise dargestellt werden, sofern nur eine geeigente Deklaration von Anfangs- und Endzuständen gelingt.

Im Falle $T = 0$ reduziert sich der Bereich des elastischen Tunnelns auf ein definiertes Energiefenster:

$$I = 2\pi e \int_{E_F - eV}^{E_F} \mathrm{d}E \sum_{\mu,\nu} |M_{\mu\nu}|^2 \delta(E_\mu - E)\delta(E_\nu - E). \tag{1.15}$$

Gleichung (1.15) zeigt folgendes: Im spektroskopischen STM-Mode wird bei einer festen Gleichspannung V_G der differentielle Leitwert, $\partial I/\partial V|_{V_G}$, gemessen. Die entsprechende Differentiation von (1.15) offenbart, daß (unter Vernachlässigung einer geringfügigen Spannungsabhängigkeit des Matrixelementes) der Wechselstrom dann aus der Umgebung des Energiekanals $E_F + eV_G$ kommt. Je nach V_G lassen sich auf diese Weise also gezielt Zustände detektieren, die bis zu mehreren Elektronenvolt ober- oder unterhalb der Fermienergie liegen können.

Bei kleinen Tunnelspannungen, $eV \ll E_F$, entsteht schließlich

$$I = 2\pi e^2 V \sum_{\mu,\nu} |M_{\mu\nu}|^2 \delta(E_\mu - E_F)\, \delta(E_\nu - E_F), \tag{1.16}$$

der wohl am häufigsten gewählte Startpunkt für einen theoretischen Zugang zur Rastertunnelmikroskopie.

1.3 Zur Theorie des Rastertunnelmikroskops

Die Literatur zur STM-Theorie ist zu umfangreich, um an dieser Stelle einen vollständigen Überblick zu geben. Wir konzentrieren uns auf das für uns Wesentliche.

1.3.1 Zugänge den Transfer-Hamiltonian benutzend

Eines der historisch ersten und ob seiner Erklärungserfolge und der schönen Unmittelbarkeit seiner Aussage zum interpretatorischen Standard bei hochaufgelösten STM-Bildern avancierten STM-Modelle war das Modell von Tersoff und Hamann [77, 78], das ein konstantes Barrierenpotential sowie eine atomar ebene Probenoberfläche zugrundelegt und die Spitze durch ein sphärisches Kastenpotential der Tiefe V_0 idealisiert, in der außerdem nur s-Zustände ($l = 0$) zugelassen sind. Die Wellenfunktionen der Spitze mit dem Radius R_0 und dem Mittelpunkt bei $\underline{\mathbf{r}}_0$ nehmen im Außenraum dann die einfache Gestalt an

$$\psi_\nu^{(t)}(\underline{\mathbf{r}}) = N_t \frac{e^{-\kappa|\mathbf{r}_0 - \mathbf{r}|}}{\kappa|\underline{\mathbf{r}}_0 - \underline{\mathbf{r}}|} \quad , \quad |\underline{\mathbf{r}}_0 - \underline{\mathbf{r}}| > R_0, \tag{1.17}$$

[6] Noch hierzu eine Anmerkung: Die Zahlenwerte der E_μ zu festem μ variieren in (1.14) also mit eV. Dies mag nicht unbedingt als die natürlichste Konvention erscheinen, werden die Eigenzustände beim THF doch für jede Elektrode separat und in der Regel ohne Berücksichtigung einer externen Tunnelspannung V berechnet. Man kann die Energien E_ν und E_μ auch in jeweils eigenen, mit der entsprechenden Elektrode fest verbundenen Skalen angeben, in denen auch die Fermi-Energien E_F^L und E_F^R und Fermiverteilungen f_L und f_R der separaten Elektroden gegeben sind (Abb. 1.1). Zur Unterscheidung schreiben wir dann einmal gestrichene Größen E'_ν und E'_μ. Die Verschiebung um eV erscheint in diesem Falle in der Deltafunktion:

$$I = 2\pi e \sum_{\mu,\nu} |M_{\mu\nu}|^2 \left[f_L(E'_\nu) - f_R(E'_\mu) \right] \delta(E'_\mu - E'_\nu - eV)\,.$$

mit $\kappa \equiv \sqrt{V_0 - E}$ und N_t als Normierungsfaktor. Die Probenoberfläche wird planar angesetzt, im Außenraum ($z > 0$) ist das Potential konstant V_0 und im Inneren ($z < 0$) beliebig periodisch. Für den Außenraum werden die Probenwellenfunktionen dann ganz allgemein durch Entwicklung nach zweidimensionalen Blochwellen dargestellt:

$$\psi_\mu^{(s)}(\underline{\mathbf{r}}) = \Omega_s^{-1/2} \sum_{\mathbf{G}} a_\mu(\underline{\mathbf{G}}) e^{-\sqrt{\kappa^2 + (\mathbf{G}+\mathbf{K})^2}\cdot z} e^{i(\mathbf{G}+\mathbf{K})\mathbf{R}} \quad , \quad z > 0. \tag{1.18}$$

mit Ω_s als Normierungsfaktor und κ wie oben. $\underline{\mathbf{G}}$ und $\underline{\mathbf{K}}$ sind hier reziproke Gittervektoren bzw. Blochvektoren des Oberflächengitters, und der Ortsvektor ist zerlegt parallel und senkrecht zur Oberfläche, $\underline{\mathbf{r}} = \underline{\mathbf{R}} + z\underline{\mathbf{e}}_z$ Nach Einsetzen von (1.17) und (1.18) in (1.11) und (1.16) faktorisiert das Matrixelement in einen Spitzen- und Probenanteil, und es entsteht:

$$I = \text{const} \cdot e^2 V \sum_\mu |\psi_\mu^{(s)}(\underline{\mathbf{r}}_0)|^2 \delta(E_\mu - E_F) \quad , \quad eV \ll E_F. \tag{1.19}$$

Profile konstanten Tunnelstroms wären demnach Linien konstanter lokaler Zustandsdichte der Probe am Krümmungsmittelpunkt $\underline{\mathbf{r}}_0$ der Spitze. Dieses Resultat würde im übrigen exakt auch von einem $\delta(\underline{\mathbf{r}}_0)$–funktionsartigen Spitzezustand geliefert und kann insofern als die ultimative Auflösungsgrenze des STM-Prinzips angesehen werden. Es ist zu betonen: Die Standardtheorie von Tersoff/Hamann setzt bezüglich der mesoskopischen Oberflächeninhomogenitäten voraus, daß deren Krümmungsradien auf der Probenoberflächen viel größer sind als diejenigen auf der Sondenspitze ($R_{\text{Probe}} \gg R_{\text{Spitze}}$).

Chung u. a. [19] berücksichtigten im Standardmodell auch Spitzenzustände mit $l \neq 0$. Die erhaltenen analytischen Zusatzterme erwiesen sich allerdings als zu wenig übersichtlich, um hernach praktische Bedeutung zu erlangen. Leavens und Aers [53] formulierten für das Standardmodell die Stromdichte nach Lang [49] und berechneten Stromdichteverteilungen insbesondere für eine Monoschicht Graphit.

Lang [49, 50, 51] ergänzte den Apparat des THF zunächst um den allgemeinen Ausdruck der Stromdichte [49] und berechnete hernach selbige sowie Konstantstromprofile zwischen zwei selbstkonsistent behandelten halbunendlichen Jelliumoberflächen, auf denen je ein Atom absorbiert war (Spitzenatom scannt Probenatom ab). Überraschend damals die negative Korrugation bei der Paarung Na-Spitze über He-Atom.

Hietschold und Sbosny [36] studierten in einem weitgehend analytisch behandelbaren einfachen Sommerfeld-Modell insbesondere die Tunnelstrombeiträge einzelner Spitzen- und Probenzustände bei der Abbildung von Kanten und Ecken.

1.3.2 Zugänge nicht-störungstheoretischer Art

Über den THF hinaus wurde der Tunnelstrom in verschiedenen Arbeiten auch auf nichtstörungstheoretische Weise berechnet. Das prinzipielle Vorgehen ähnelt dem bei der Ableitung des exakten Durchlaßkoeffizienten für die eindimensionale Rechteckbarriere: Betrachtet wird ein stationäres Streuproblem unter der Annahme, daß die Gesamtwellenfunktion asymptotisch für $z \to \pm\infty$ aus einfallenden und reflektierten bzw. nur aus transmittierten ebenen z-Wellen bestehen sollen. Natürlich ist ein Modell Voraussetzung, das asymptotisch bzgl. z separiert und dort eine von z unabhängige Zerlegungen nach ebenen Wellen gestattet (in z konstantes oder streng periodisches Potential). Die Schrödingergleichung wird numerisch für das Gesamtsystem gelöst und der Tunnelstrom letztlich aus den Amplitudenverhältnissen der asymptotischen Wellenfelder bestimmt.

Laloyaux u. a. [47, 46] studierten verschiedene axialsymmetrische Kombinationen aus je zwei planaren Sommerfeld-Metallen, auf denen Erhebungen (Halbkugel, Zylinder, Gauß-Peak) oder eine gaußförmige Vertiefung angebracht waren. Bestimmt wurden die Wellenfunktionen vermöge einer Finite-Elemente-Methode, wobei das volle äußere elektrostatische und Bildkraftpotential Berücksichtigung fand. Berechnet wurden Stromdichteverteilungen und Strom-Abstands-Charakteristiken.

Interessanterweise zeigten dabei zusätzlich im THF durchgeführte Vergleichsrechnungen [46], daß der THF die nicht-störungstheoretischen Resultate überraschend genau reproduzieren konnte, sogar bis hinab zu Abständen von 1 Å, während allgemein ein Zusammenbrechen der Transfer-Hamiltonian-Approximation auf kurzen Distanzen erwartet wurde. Blieb insbesondere das Bildpotential unberücksichtigt (das zwischen einem Ein- und einem Zwei-Elektrodensystem deutlich verschieden ausfällt), waren die Stromresultate nahezu ununterscheidbar.

Im Hinblick auf die Variierbarkeit der Elektrodengeometrien sind diese Arbeiten durchaus mit der unsrigen vergleichbar, allerdings konnte dort der Abbildungsprozeß (das „Scannen“) nicht simuliert werden, da dies die Zylindersymmetrie zerstört hätte.

Doyen und Drakova u. a. behandelten in mehreren Arbeiten die STM-Abbildung flacher oder atomar gestufter Metalloberflächen und den Einfluß von Adsobatatomen [22, 23, 45]. Der Tunnelstrom wurde hierbei – außer in [22] – nach Lippmann [55, 56] über das verallgemeinerte Ehrenfestsche Theorem bestimmt.

1.3.3 Die klassisch-geometrische Entfaltung

Eine rein klassisch-geometrische Betrachtungsweise, bei der von der Quantennatur der Tunnelphänomene abgesehen wird, praktizierten Reiss u. a. [69, 68] zur Entfaltung mesoskopisch rauher STM-Abbildungen. Da wir dieses geometrische Verfahren später als Referenz heranziehen möchten, seien die Details für den zweidimensionalen Fall einmal dargestellt:

Angenommen, Probe und Spitze würden im klassischen Sinne unter mechanischer, punktförmiger Berührung aufeinander abgleiten. Der Berührungspunkt wäre der Punkt zusammenfallender Körpertangenten und der Anstieg eines Konstantstromprofils käme dem Anstieg dieser Tangenten gleich. (Er kann unter dieser Annahme dem experimentellen STM-Profil unmittelbar entnommen werden). Ist nun die Geometrie der Spitze bekannt – und das ist hierbei die Voraussetzung –, läßt sich für jeden einzelnen Bildpunkt anhand der Tangente der zugehörige Berührungspunkt auf dem Spitzenkörper rekonstruieren (Abb. 1.2). Damit können die Bildpunkte, die ja nur dann mit der Probenoberfläche identisch wären, wenn der Kontakt grundsätzlich am Spitzenapex erfolgen würde, in der „wahren“ Berührungspunkte der Probe transformiert werden. In drei Dimensionen wären statt der Tangenten die Tangentialebenen zu nehmen.

1.4 Quantenbillards

An dieser Stelle soll der Begriff des Quantenbillards eingeführt werden, eine gründlichere Darstellung der hiermit im Zusammenhang stehenden semiklassischen Theorie findet sich in Kapitel 3.

Klassische konservative System können in zwei Klassen eingeteilt werden: *integrable* und *nichtintegrable*. Integrable Systeme mit N Freiheitsgraden besitzen N Konstanten der

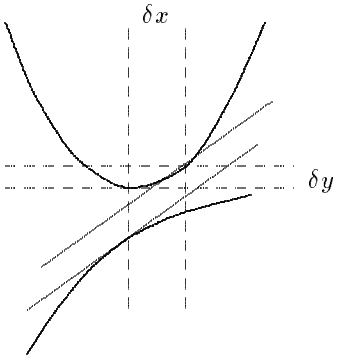


Abbildung 1.2: *Die* Reiss*sche Korrektur [68]. Eine mechanische Berührung zwischen Probe und Spitze erfolgt am Punkt zusammenfallender Körpertangenten und i. allg. nicht am Spitzenapex. Die Tangente wird dem STM-Profil (der unteren dicken Vollinie) entnommen und damit der Berührungspunkt auf der Spitze rekonstruiert. Um dessen Abstand vom Apex ist nun auch der STM-Meßpunkt zu korrigieren, um zum wahren Probenpunkt zu gelangen. Im linken Beispiel also um δx nach rechts und um δy nach oben.*

Bewegung. Die Trajektorien liegen auf einer N-dimensionalen Fläche im $2N$-dimensionlen Phasenraum, alle Bahnen sind entweder periodisch oder quasiperiodisch und der Abstand benachbarter Phasenraumpunkte wächst nur nach einem Potenzgesetz mit der Zeit.

In nichtintegrablen Systemen ist demgegenüber die Zahl der Bewegungskonstanten kleiner als N. Man unterscheidet nochmals in vollständig chaotische (ergodische) Systeme, bei denen die Menge der periodischen Bahnen vom Maß Null ist (gleichwohl es unendlich viele gibt) und in denen eine typische Bahn mit der Zeit den gesamten zugänglichen Phasenraum überstreicht, und in solche, in denen der Phasenraum sowohl Bereiche chaotischer Bewegung als auch solche quasiperiodischer Bewegung (KAM-Tori) enthält. Von chaotischem Verhalten spricht man, wenn die Dynamik sehr empfindlich von den Anfangsbedingungen abhängt. Benachbarte Bahnen laufen mit der Zeit exponentiell auseinander.

Systeme mit einem Freiheitsgrad sind immer integrabel. Die einfachsten Systeme, die Chaos zeigen können, sind zweidimensional und hierunter wiederum besonders einfach die sogenannten Billards. Unter einem klassischen Billard versteht man ein von unendlich hohen Wänden umrandetes ebenes Gebiet, in dem sich ein punktförmiges Teilchen bewegt und an den Rändern elastisch reflektiert wird. Die Randgeometrie entscheidet, ob ein integrables oder nichtintegrables System vorliegt. Typische Vertreter des ersteren sind das Rechteck- oder das Kreisbillard, zur zweiten Klasse gehören das Sinai- und das Stadionbillard.

Wird statt des klassischen nun ein quantenmechanisches Teilchen betrachtet (Quantenbillard), hat man die stationäre SGL $[\boldsymbol{\nabla}^2 + k^2]\Psi = 0$ mit Dirichletschen Randbedingungen $\Psi|_\Gamma = 0$ zu lösen. Von besonderem Interesse erscheint, auf welche Weise sich im semiklassischen Grenzfall die integrablen oder chaotischen Eigenschaften des korrespondierenden klassischen Systems niederschlagen. Der Begriff des Quantenchaos ist durchaus noch nicht fest, kann man doch den für die klassische Definition zentralen Bahnbegriff nicht ohne weiteres übertragen[7].

Es zeigte sich, daß sich integrable und chaotische Quantenbillards insbesondere anhand der statistischen Eigenschaften ihrer Eigenwertspektren charakterisieren lassen. Untersucht wurde vor allem die Nächste-Nachbar-Statistik (NNS) – die Verteilung der Abstände benachbarter Eigenwerte, $s_i = (E_{i+1} - E_i)/\bar{s}$; $\bar{s}$ ist der mittlere Abstand zweier Eigenwerte.

In klassisch-integrablen Systemen macht die Existenz invarianter Tori eine semiklassische Quantisierung nach dem WKB-Verfahren möglich. Die erhaltenen Energienieveaus sind weitgehend unkorreliert, neigen zur „Klumpung“ – die Wahrscheinlichkeit für einen

[7] Der Bahnbegriff kehrte in Gestalt des Pfadintegrals in die Quantenmechanik zurück, wenngleich nicht mehr auf der alten klassischen, unmittelbar anschaulichen Stufe der Realität.

Eigenwert bei E ist unabhängig davon, ob sich noch weitere bei E befinden – und zeigen im wesentlichen eine Poisson-Verteilung [14]:

$$P(s) = e^{-s}. \tag{1.20}$$

Auf nichtintegrable Systeme ist die WKB-Quantisierung nicht anwendbar. Die semiklassische Theorie operiert dort wesentlich mit dem Begriff der periodischen Bahnen (Kapitel 3) und ist noch keinewegs abgeschlossen. Zahlreiche Untersuchungen belegen indes, daß die NNS chaotischer Systeme sehr gut durch Ensembles von Zufallsmatrizen beschrieben werden können (für einen Überblick siehe [8, 17]). Insbesondere verhalten sich die Eigenwerte von Systemen mit Zeitumkehrinvarianz wie die eines Gausssschen Orthogonalen Ensembles (GOE), dessen NNS die Form einer Wigner-Verteilung besitzt:

$$P(s) = \frac{\pi}{2}\, s\, e^{-\frac{\pi}{4}s^2}. \tag{1.21}$$

Numerisch berechnete Eigenwertespektren des Stadionbillards [59, 60] und des Sinaibillards [18] ließen sich gut auf eine Wigner-Verteilung abbilden.

Zur Frage der Billard-Eigenfunktionen vergleiche man 3.3.

Kapitel 2

Modell des Rastertunnelmikroskops

2.1 Beschreibung der Elektroden

Das Hauptaugenmerk dieser Rechnungen galt dem Einfluß der Geometrie von Spitze und Probe auf den Abbildungsprozeß in der Rastertunnelmikroskopie. Auf einer mesoskopischen Skala ist dieser an stark zerklüfteten Oberflächen offensichtlich [69, 68], bei der atomaren Abbildung einer ideal (atomar) ebenen Probe eher vernachlässigbar, wie liegen aber die Verhältnisse zwischen beiden Extremen? Einsichten wurden erhofft für solche Fragen wie die nach einer Bestimmbarkeit charakteristischer Spitzenparameter anhand geeigneter, prägnanter Teststrukturen oder einem möglichst genauen Ausmessen der Breite von Gräben oder Linien. Zu diesem Zweck sollten die geometrischen Modellparameter so frei als möglich wählbar sein, was nach sich zog, daß numerisch gegebenen Grenzen bei der Lösung der Schrödingergleichung dann in Form starker Vereinfachungen bezüglich der elektronischen und atomaren Struktur der Elektroden Tribut zu zollen war.

Konkret beschreiben wir die Elektroden durch geeignet geformte Potentialkästen endlicher räumlicher Ausdehnung. Im Innen- wie im Außenraum herrscht konstantes Potential, das an der Oberfläche um den Betrag der Kastentiefe $V_0 \equiv E_F + \Phi_A$ auf den Vakuumwert springt, E_F ist die Fermienergie und Φ_A die Austrittsarbeit. Die Potentialkästen sind bis zur Fermi-Energie gefüllt mit einem wechselwirkungsfreien Elektronengas. Bekannt ist diese einfache Beschreibungsweise unter dem Namen *Sommerfeld-Modell* oder *Sommerfeld-Metall.* Die Barrierenverhältnisse zweier derartiger Elektroden im Tunnelkontakt bei der Tunnelspannung Null zeigt Abb. 2.1, wobei wir konkret dann immer die von der Größenordnung her für Metalle typischen Werte $\Phi_A = 4\,\mathrm{eV}$ und $E_F = 4\,\mathrm{eV}$ angenommen haben.

Der Tunnelstrom wird über den Transfer-Hamiltonian-Formalismus (THF) bestimmt. Folglich sind (nur) die Eigenzustände der einzelnen Elektroden erforderlich. Aufgrund der Beliebigkeit der Elektrodenformen muß das numerische Verfahren zur Berechnung der Eigenzustände dabei auf jeglichen Symmetrievorteil oder dergleichen verzichten und arbeitet sehr allgemein. Die Elektroden werden in ein hinreichend größeres Grundgebiet, auf dessen Rand $\psi = 0$ gefordert wird, eingebettet (Abb. 2.1), und in diesem die SGL im Differenzenschema auf einem äquidistanten Gitter diskretisiert. Sämtliche Schritte bis hin zur numerischen Erprobung wurden für zwei und für drei Dimensionen durchgeführt, die praktischen

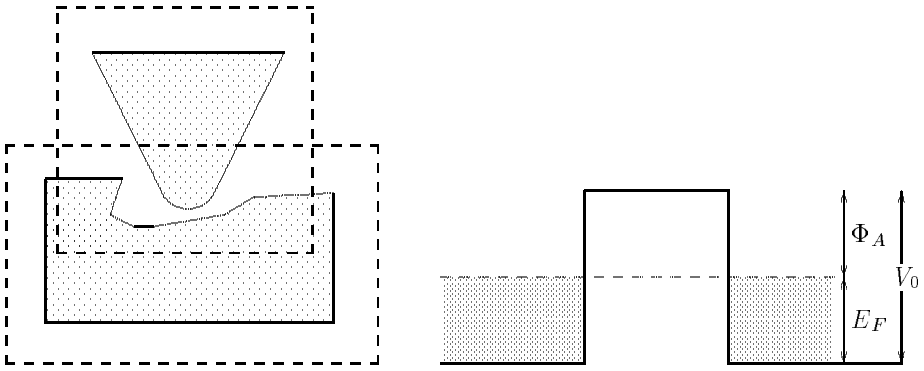


Abbildung 2.1: *Die beiden STM-Elektroden werden durch (zweidimensionale) räumlich begrenzte Potentialkästen der Tiefe V_0 beschrieben. Deren Geometrie ist frei wählbar. Zur Berechnung der Zustände wird jede Elektrode in ein größeres Grundgebiet eingebettet (gestrichelte Rechtecke links) und in diesem die SGL diskretisiert. Rechts ist das Barrierenpotential gezeigt.*

Berechnungen auf einer IBM-Workstation AIX Risc 6000 mußten allerdings auf zweidimensionale Systeme beschränkt bleiben. Das Verfahren wird detailierter noch darzustellen sein und stellt nicht mehr, aber auch nicht weniger als die Grundlage dieses Unterfangens dar.

Tatsächlich lassen sich damit gebundene Zustände nicht nur in Kästen, sondern in beliebigen Potentialmulden berechnen. Theoretisch wären statt der Rechteck- mithin auch „weichere“ und realistischere Barrierenformen möglich gewesen – etwa, indem für jede Anordnung der beiden Elektroden zunächst das elektrostatische und Bildkraftpotential bestimmt wird und dieses Gesamtpotential dann in einer für die Störungstheorie des THF geeigneten Weise den Einzelelektroden auferlegt wird – bis zum Barrierenscheitel z. B. das Gesamtpotential und danach fortgesetzt mit einem konstanten Außenwert. Abgesehen vom außerordentlichen Mehraufwand dieses Vorgehens – für jede Position wären dank des sich ändernden Potentials sämtliche Eigenzustände beständig neu zu berechnen – sind zwei Dinge anzumerken:

(i) Durch die Verquickung von geometrischen und Potentialeffekten würde die Interpretation der resultierenden STM-Profile erschwert: Wir benutzen immerhin Einteilchenzustände in allseitig begrenzten, mesoskopisch großen Elektroden, in denen die Kohärenz der Wellenzüge durch keinerlei (mit Vielteilcheneffekten zusammenhängendem) Dämpfungsmechanismus gestört wird. Geringfügige Änderungen am Rand können globale Änderungen der Zustände zur Folge haben. In Quantenbillards mit unendlich hohen Wänden werden bei ähnlichen Verhältnissen zwischen Wellenlänge und Systemabmesssungen derartige Phänomene beobachtet ([60] und auch Kapitel 5). Beim gegenwärtigen Kenntnisstand ist nicht vorhersagbar, ob ein zu chaotischem Verhalten neigender Billiardzustand[1] auf ein sich änderndes Bildkraftpotential nicht mit einem völlig neuen Aussehen in der stromwichtigen barrierennahen Region antwortet. (Im Gegenteil scheint unser numerisches Verfahren diesen Aspekt erstmals zugänglich zu machen.) Für eine STM-Betrachtung wäre dies insofern hinderlich, als die fixierte, aber an sich willkürlich gewählte Gesamtausdehnung unserer Elektroden

[1] Gemeint ist die Tatsache, daß es ganz offensichtlich einzelne Zustände gibt, die auf eine Störung besonders empfindlich reagieren, vgl. Kapitel 5.

eine pathologische Betonung erführe. Die Beschränkung auf Kastenpotentiale dient daher in jedem Fall der Konzentration auf die geometrische Fragestellung.

(ii) Das physikalisch sicher befriedigenste Vorgehen, nämlich in selbstkonsistenter Weise Vielteilcheneffekte durch simultane Lösung der Schrödinger- und der Poissongleichung im Dichtefunktional- oder Hartree-Fock-Formalismus zu berücksichtigen, führt an planaren Oberflächen zu Potentialverläufen, die dem Rechteckpotential wieder stärker ähneln als das reine Einteilchen-Bildkraftpotential [52].

2.2 Der Tunnelstrom

Die Berechnung des Tunnelstroms erfolgt im Transfer-Hamiltonian-Formalismus (THF). Auf eine Besonderheit ist hinsichtlich des Matrixelementes zu verweisen: Entsprechend der gewählten numerischen Randbedingung klingen die Wellenfunktionen bei uns nicht im Unendlichen ab, sondern verschwinden bereits außerhalb des Einbettungsgebietes. Sei L_a die Außenraumlänge, d. h. der Mindestabstand jedes Punktes der Elektrodenbegrenzung zum Einbettungsrand. Das Tunnelmatrixelement nach (1.8) liefert für alle Elektrodenabstände $d > L_a$ Null, da ab dieser Entfernung die Probenwellenfunktionen im Gebiet der Spitze verschwunden sind und umgekehrt. Im Gegensatz dazu würde das Matrixelement nach (1.11) Beiträge liefern, solange ein Überlapp der Wellenfunktionen in der Barriere existiert, also bis zu Abständen von $d < 2L_a$.

Dieser offensichtliche Widerspruch rührt daher, daß die naive Anwendung des Gauss-schen Satzes hier auf Schwierigkeiten stößt: Auf dem Rand des Einbettungsgebietes besitzt die Wellenfunktion einen Knick (außerhalb konstant Null), sie ist dort nicht stetig differenzierbar. In einer elektrodynamischen Analogie entspräche dies einer Oberflächenladung. Um konkret den Schritt von (1.9) zu (1.10) gehen zu können, darf die Oberfläche des Integrationsgebietes Ω deshalb den Einbettungsrand nicht schneiden, d. h. Ω muß von vornherein auf die Schnittmenge der Einbettungsgebiete beschränkt bleiben. Der rückwärtige Teil der Fläche ? , der normalerweise im Unendlichen liegt und von dort keine Beiträge liefert, kann hier also nicht vernachlässigt werden, er würde sich im obigen Falle ebenfalls in der Barriere befinden[2].

Aus diesem Grunde ist es in unserem Modell viel einfacher, das Matrixelement nicht aus den differentiellen Bardeenschen Formen (1.11) oder (1.12), sondern direkt über (1.8), d. h. durch Integration über das Innere einer Störelektrode, zu bestimmen, zumal die Kastenpotentialen dem keinerlei Schwierigkeiten entgegensetzen:

$$M_{\mu\nu} \;=\; -V_0 \iint\limits_{\text{Spitze}} \mathrm{d}\tau\; \psi_\nu^*(\underline{\mathbf{r}})\, \psi_\mu(\underline{\mathbf{r}}) \;=\; -V_0 \iint\limits_{\text{Probe}} \mathrm{d}\tau\; \psi_\nu^*(\underline{\mathbf{r}})\, \psi_\mu(\underline{\mathbf{r}}), \tag{2.1}$$

($\mathrm{d}\tau \equiv \mathrm{d}x\,\mathrm{d}y$ oder $\mathrm{d}\tau \equiv \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$).

Die Wellenfunktionen liegen nach der numerischen Lösung tabelliert als Werte eines äquidistanten Gitters vor. Für die Gebietsintegrationen in (2.1) sind die Stützstellen daher vorgegeben und es wurde nach der Simpson-Regel verfahren.

[2] Der entscheidende Term auf der Rückwand von ? in (1.11), falls ψ_μ einmal als der abklingende Zustand angenommen wird, ist $\psi_\nu^* \nabla \psi_\mu$. Im Eindimensionalen fällt ψ_μ bei ungehindertem Abklingen asymptotisch mit $e^{-\kappa x}$, d.h. $\nabla\psi_\mu \sim e^{-\kappa x} \overset{x\to\infty}{\longrightarrow} 0$. Wird ψ_μ dagegen vorher auf 0 gezwungen, sagen wir bei $x = b$, fällt ψ_μ mit $\sinh\kappa(x-b)$ (vgl. A.2), d. h. $\nabla\psi_\mu \sim \cosh\kappa(x-b) \overset{x\to b}{\longrightarrow} 1$. $\psi_\nu^* \nabla\psi_\mu$ verschwindet also auch nicht auf dem Einbettungsrand, gleichwohl ψ_μ selbst dort verschwindet.

Es erhebt sich die Frage nach der Mindestgröße der Elektroden und nach den erforderlichen Außenraumlängen L_a, die beide aus verständlichen numerischen Gründen nicht größer als nötig sein sollten.

Allermindenstens ist von den Elektrodenabmessungen zu fordern, daß die Rück- und Seitenwände keinen unmittelbaren Einfluß mehr auf die Krümmung der Wellenfunktion in der barrierennahen Region ausüben. Trotzdem besitzt der einzelne Zustand natürlich eine fixierte Phasenlage, die bei wenigen Zuständen in der LDOS zu einer artifiziellen lateralen Welligkeit führt und, wie Versuche ergaben, bei etwa 10 Zuständen weitgehend ausschmiert. Diese 10 Zustände sollten nun – das Bild eines quasikontinuierlichen Spektrums vor Augen – in einem relativ kleinen Intervall an der Fermie-Kante liegen, wodurch letztlich die Mindestgröße der Elektroden bestimmt wird. Typischerweise wurden Probengeometrien in die Oberseite eines 80×50 Å^2 großen Rechtecks eingebracht, und Spitzen erhielten eine z-Ausdehnung von 40 Å. (Zum Vergleich beträgt die de Broglie-Wellenlänge bei 4 eV ca. 6 Å.) Die 10 Energiewerte lagen je nach Elektrodengröße dann in einem Intervall $E_F \pm \Delta E$ mit $\Delta E = 10 \ldots 50$ meV.

Zur Stromberechnung nach (1.16) werden also 10 der Fermienergie $E_F = 4$ eV nächstbenachbarte Eigenzustände jeder Elektrode herangezogen und die insgesamt 100 Matrixelemente aufsummiert (jeder Spitzenzustand mit jedem Probenzustand kombiniert). Die Gleichheit der Energien ist dabei natürlich nur näherungsweise erfüllt.

Nun zur Außenraumlänge L_a. Der Einfluß von L_a auf die Eigenwerte ist ab $L_a \approx 5$ Å vernachlässigbar. Für die Berechnung des Tunnelstroms vermöge (2.1) muß jedoch ein hinreichender Überlapp mit dem Inneren der Gegenelektrode existieren. Die Außenraumlänge auf der der Gegenelektrode zugewandten Seite entscheidet also über die maximal möglichen Elektrodenabstände in den STM-Profilen.

Als ein guter Anhaltspunkt, bis zu welchen Abständen die THF-Näherung brauchbar ist, kann für unsere Kästen die analytisch behandelbare eindimensionale (symmetrische) Rechteckbarriere dienen. Der exakte Duchlaßkoeffizient lautet [48]

$$D_{\text{exakt}} = \frac{4\,k^2\kappa^2}{(k^2+\kappa^2)^2 \sinh^2(\kappa d) + 4k^2\kappa^2}, \tag{2.2}$$

während der THF auf

$$D_{\text{THF}} = \frac{16\,k^2\kappa^2}{(k^2+\kappa^2)^2} e^{-2\kappa d} \tag{2.3}$$

führt (d =Abstand). Letzteres ist gerade (2.2) in der Näherung $e^{-2\kappa d} \ll e^{2\kappa d}$. Mit unseren Daten $E = 4\,\text{eV}$ und $V_0 = 8\,\text{eV}$ finden wir folgende Verhältnisse:

d	1 Å	2 Å	3 Å	4 Å	5 Å
D_{exakt}	$4,04 \cdot 10^{-1}$	$6,42 \cdot 10^{-2}$	$8,52 \cdot 10^{-3}$	$1,10 \cdot 10^{-3}$	$1,42 \cdot 10^{-4}$
$D_{\text{THF}}/D_{\text{ex}}$	1,274	1,033	1,004	1,001	1,000

Die Genauigkeit des THF bis hinab zu $d = 2$ Å ist hier zweifellos auf die abstandsunabhängige konstant hohe Barriere zurückzuführen. Bei mehr Dimensionen dürften die Verhältnisse indes ähnlich liegen, erinnert sei an die Testrechnungen in [46] (Abschnitt 1.3.2), wo insbesondere für Rechteckbarrieren die mit dem THF erzielten und die nicht-störungstheoretischen Resultate sogar bis zu Abständen von 1 Å fast ununterscheidbar waren. Mit unseren Kastenbarrieren von 4 eV können wir also ohne weiteres Tunnelströme bis herunter zu

Abständen von 2 Å ernst nehmen; die relativ hohe Barriere erzwingt ein rasches Abklingen und ist ein Kompromiß zwischen den „wahren“ Verhältnissen und der Bestrebung, den numerischen Apparat nicht über Gebühr durch allzu große Einbettungsgebiete zu belasten.

Für die Außenraumlängen wurden letztlich $L_a = 10$ Å auf der der Gegenelektrode zugewandten Seite und $L_a = 8$ Å sonst gewählt. Bei Abständen von 6 Å existiert damit zur Stromberechnung noch ein Überlapp von 4 Å mit der Gegenelektrode, wobei zu bedenken ist, daß der Strom im Gefolge der abklingenden Wellenfunktionen dort bereits sehr klein ist und rasch gegen Null strebt.

Kapitel 3

Periodische Bahnen

3.1 Einige Grundtatsachen

In der traditionell im Operatorkalkül formulierten Quantenmechanik wird die zeitliche Entwicklung eines Hilbertraum-Zustandes $|\Psi(t)\rangle$ durch die Schrödingergleichung (SGL)

$$i\hbar\frac{\partial}{\partial t}\,|\Psi(t)\rangle = \hat{H}\,|\Psi(t)\rangle\,. \tag{3.1}$$

beschrieben[1]. Deren Lösungen stellt man in der Ortsdarstellung

$$\Psi(\underline{\mathbf{q}},t) = \langle\,\underline{\mathbf{q}}\,|\,\Psi(t)\rangle \tag{3.2}$$

häufig vermöge eines Propagators $K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ dar,

$$\Psi(\underline{\mathbf{q}}_b,t) = \int \mathrm{d}\underline{\mathbf{q}}_a K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)\Psi(\underline{\mathbf{q}}_a,0), \tag{3.3}$$

worin aus mathematischer Sicht zunächst zum Ausdruck kommt, daß den Lösungen des Anfangswertproblems (3.1) die Gestalt einer Integralgleichung mit zu findendem Integralkern $K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ gegeben werden kann. Im mehr anschaulich physikalischen Sinne transportiert $K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ die *Wahrscheinlichkeitsamplitude* eines infinitesimalen Volumenelementes vom Raumzeitpunkt $(\underline{\mathbf{q}}_a,0)$ zum Raumzeitpunkt $(\underline{\mathbf{q}}_b,t)$ und wird auch als *Übergangsamplitude* bezeichnet. Die Schreibweise $K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ besagt schon, daß wir uns hier auf zeittranslationsinvariante Probleme beschränken wollen.

Für ein stationäres System kann (3.1) formal integriert werden:

$$|\Psi(t)\rangle = e^{-\frac{i}{\hbar}\hat{H}t}\,|\Psi(0)\rangle\,. \tag{3.4}$$

Der Propagator lautet nun explizit

$$K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t) = \left\langle\underline{\mathbf{q}}_b\right|e^{-\frac{i}{\hbar}\hat{H}t}\left|\underline{\mathbf{q}}_a\right\rangle \tag{3.5}$$

und nimmt entwickelt nach Energieeigenfunktionen ϕ_n, in der Spektraldarstellung, die Gestalt an:

$$K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t) = \sum_n \phi_n(\underline{\mathbf{q}}_b)\phi_n^*(\underline{\mathbf{q}}_a)e^{-\frac{i}{\hbar}E_n t}\,. \tag{3.6}$$

[1] Es ist dies die basisunabhängige Ket-Schreibweise.

Aus der SGL abgeleitete Observablen sind in dem gleichen Sinne invariant bezüglich Zeitumkehr wie die korrespondierenden klassischen Größen, insbesondere also in zeitunabhängigen Systemen ohne Magnetfeld. Beschreibt dann $|\Psi(\underline{\mathbf{q}},t)|^2$ die Ausbreitung eines Wellenpaketes für $t>0$, existiert als formale Lösung immer auch das rückwärts in der Zeit laufende Wellenpakt $|\Psi(\underline{\mathbf{q}},-t)|^2$. $K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ in (3.5) ist demgemäß der allgemeine, uneingeschränkte Propagator, der Lösungen sowohl in positiver wie in negativer Zeitrichtung erfaßt.

Um dem Aspekt der Kausalität Rechnung zu tragen, wird gewöhnlich der retardierte oder kausale Propagator $K_r(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ eingeführt, der für $t<0$ verschwindet:

$$K_r(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)=\Theta(t)K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t). \tag{3.7}$$

Dessen Fouriertransformation, infolge der Stufenfunktion $\Theta(t)$ auf den positiven Zeitast beschränkt, gibt die kausale energieabhängige Greensche Funktion:

$$G(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,E) = -\frac{i}{\hbar}\int_0^\infty K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)\,e^{\frac{i}{\hbar}Et}\mathrm{d}t \tag{3.8}$$

$$G(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,E+i0) = \sum_n \frac{\phi_n(\underline{\mathbf{q}}_b)\phi_n^*(\underline{\mathbf{q}}_a)}{E-E_n+i0} \tag{3.9}$$

(„$+i0$“ steht für das einzufügende infinitesimale Dämpfungsglied zur Unterdrückung der unphysikalischen Fluktuationen für $t\to\infty$), aus der über die folgenden Bildungen dann Dichtematrix

$$\varrho(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,E) = -\frac{1}{\pi}\mathrm{Im}\left[G(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,E+i0)\right] \tag{3.10}$$

$$= \sum_n \phi_n(\underline{\mathbf{q}}_b)\phi_n^*(\underline{\mathbf{q}}_a)\delta(E-E_n), \tag{3.11}$$

lokale Energiezustandsdichte (LDOS)

$$\varrho(\underline{\mathbf{q}},E) = \varrho(\underline{\mathbf{q}},\underline{\mathbf{q}},E) = \sum_n |\phi_n(\underline{\mathbf{q}})|^2\delta(E-E_n) \tag{3.12}$$

und globale Zustandsdichte

$$D(E) = \int \mathrm{d}\underline{\mathbf{q}}\;\varrho(\underline{\mathbf{q}},E) = \sum_n \delta(E-E_n) \tag{3.13}$$

folgen. Mitunter bedient man sich auch direkt der Spur von $G(\underline{\mathbf{q}},\underline{\mathbf{q}},E)$,

$$g(E) = \int \mathrm{d}\underline{\mathbf{q}}\;G(\underline{\mathbf{q}},\underline{\mathbf{q}},E), = \sum_n \frac{1}{E-E_n}, \tag{3.14}$$

und erhält die Energieniveaus als Pole der sogenannten Resolvente $g(E)$.

Im Zeitpropagator sind alle Informationen über ein Quantensystem enthalten. Die vorstehende Auflistung sollte zeigen, wie sich daraus die für stationäre Quantensysteme zentralen Observablen $\varrho(\underline{\mathbf{q}},E)$ und $D(E)$ bzw. $g(E)$ gewinnen lassen.

In der FEYNMANschen Formulierung der Quantenmechanik [27] ist dem allgemeinen, uneingeschränkten Zeitpropagator $K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)$ der SCHRÖDINGERschen Theorie ein Funktionalintegral äquivalent – in kartesischen Koordinanten und für ein Teilchen in D Dimensionen mit dem Hamiltonoperator $\hat{H}=\hat{\mathbf{p}}/2m+V(\underline{\mathbf{q}},t)$ ausgeschrieben:

$$K(\underline{\mathbf{q}}_b,\underline{\mathbf{q}}_a,t)=\lim_{n\to\infty}\left(\frac{m}{2\pi i\hbar\epsilon}\right)^{Dn/2}\int \mathrm{d}\underline{\mathbf{q}}_1\ldots\int \mathrm{d}\underline{\mathbf{q}}_{n-1}\exp\left\{\frac{i}{\hbar}\sum_{i=0}^{n-1}S_{i+1,i}\right\}, \tag{3.15}$$

$$S_{i+1,i} \equiv \frac{m}{2}\frac{(\underline{\mathbf{q}}_{i+1}-\underline{\mathbf{q}}_i)^2}{\epsilon} - \epsilon V(\underline{\mathbf{q}}_i, t_i), \quad \epsilon \equiv \frac{t}{n}, \quad \underline{\mathbf{q}}_0 = \underline{\mathbf{q}}_a, \quad \underline{\mathbf{q}}_n = \underline{\mathbf{q}}_b,$$

Die gebräuchliche, allerdings rein symbolische Umschreibung von (3.15) lautet

$$K(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t) = \int_{\mathbf{q}(0)=\mathbf{q}_a}^{\mathbf{q}(t)=\mathbf{q}_b} \mathcal{D}\underline{\mathbf{q}}(\cdot) \exp\left\{\frac{i}{\hbar} S[\underline{\mathbf{q}}(\cdot), t]\right\}, \tag{3.16}$$

mit der die übliche Lesart des Funktionalintegrals betont wird: Die quantenmechanische Übergangsamplitude resultiert aus der Superposition sämtlicher Trajektorien $\underline{\mathbf{q}}(\tau)$, die für $\tau = 0$ bei $\underline{\mathbf{q}}_a$ starten und für $\tau = t$ bei $\underline{\mathbf{q}}_b$ enden, wobei jede Bahn mit dem komplexen Phasenfaktor $e^{\frac{i}{\hbar}S}$ beiträgt; $S(\underline{\mathbf{q}}, t)$ ist hierbei die im klassichen Sinne entlang jeder Trajektorie zu nehmende Wirkungsfunktion

$$S[\underline{\mathbf{q}}(\cdot), t] = \int_0^t \left[\frac{m}{2}\left(\frac{\partial \underline{\mathbf{q}}}{\partial \tau}\right)^2 - V(\underline{\mathbf{q}}, \tau)\right] \mathrm{d}\tau, \tag{3.17}$$

in der der Bahnbegriff offensichtlich ein wohldefinierter ist. Dem Hamiltonschen Variationsprinzip $\delta S/\delta \underline{\mathbf{q}} = 0$, aus dem in der kanonischen Mechanik die klassischen Bahnkurven $\underline{\mathbf{q}}_{kl}$ hervorgehen, werden in (3.16) natürlich auch nur diese extremalen Trajektorien gerecht, alle anderen führen zu klassisch nicht erlaubten Wirkungen, tragen aber ebenso zu $K(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t)$ bei.

FEYNMANs Pfadintegral-Formulierung kommt ohne Operatoren aus und ist in gewisser Weise anschaulicher, da klassische Bahnen wieder vorgestellt werden dürfen, wenngleich nur noch im Sinne von 'alle Bahnen zugleich'. Für semiklassische Näherungen ist die Pfadintegration in jedem Falle besonders geeignet.

3.2 Semiklassische Spurformeln

Das bekannte WKB-Verfahren [48, 12] zur Konstruktion einer semiklassischen Näherung für die Eigenzustände und die aus ihr resultierenden Bohr-Sommerfeld-Quantisierungsbedingungen sind in ihren Anwendungen auf separable Systeme beschränkt [13]. Die Bohr-Sommerfeld-Bedingungen lassen sich auf integrable Systeme erweitern – sie heißen dann EKB-Bedingungen[2] und das Verfahren Torusquantisierung, jedoch nicht darüber hinaus.

Im Gegensatz zu den Eigenzuständen der zeitunabhängigen SGL kann der Zeitpropagator der vollen SGL für beliebige Systeme semiklassisch genähert werden. Er folgt asymptotisch für $\hbar \ll S$ aus der Entwicklung der Pfadintegraldarstellung (3.16) um den klassischen Pfad. Das mathematische Werkzeug zur Auswertung der hierbei auftretenden Integrale über schnell oszillierende Funktionen ist die quadratische Entwicklung der Phasen an ihren stationären Punkten (Sattelpunktsmethode), engl. Stationary Phase Approximation (SPA). Fouriertransformation und Spurbildung (Ortsintegration) im Sinne von (3.13) oder (3.14) liefern anschließend semiklassische Näherungen für die Zustandsdichte $D(E)$ bzw. die Pole der Resolvente $g(E)$. Die Formeln werden wegen der Spurbildung häufig Spurformeln genannt und ihnen gemeinsam ist, daß es sich bei ihnen stets um Summen über alle klassischen periodischen Bahnen handelt. Die wesentlichen Grundlagen wurden von GUTZWILLER in vier bedeutenden Arbeiten aus den Jahren 1967–1971 gelegt [28, 29, 30, 31].

[2] EKB = Einstein-Kramers-Brillouin; zum Unterschied separabel – integrabel z. B. [32, S. 44]

Die semiklassische Näherung des Zeitpropagators $K(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t)$ führt auf die Formel [28, 12]:

$$K(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t)_{\mathrm{sc}} = \frac{1}{\sqrt{2\pi i\hbar}^D} \sum_r \left| \det \left(\frac{\partial^2 S_r}{\partial \underline{\mathbf{q}}_a \partial \underline{\mathbf{q}}_b} \right) \right|^{1/2} \exp \left\{ i \frac{S_r}{\hbar} - \frac{\mu_r \pi}{2} \right\} \tag{3.18}$$

Zu summieren ist über alle klassischen Pfade des Systems, die in der Zeit t von $\underline{\mathbf{q}}_a$ nach $\underline{\mathbf{q}}_b$ laufen, S_r ist die zugehörige klassische Wirkung. Die Determinante heißt VAN VLECK-Determinante (vgl. [43]) und ist bis auf das Vorzeichen mit der Dichte der klassischen Pfade, die von $\underline{\mathbf{q}}_a$ ausgehend nach der Zeit t bei $\underline{\mathbf{q}}_b$ eintreffen, identisch. Der diskrete Phasenfaktor $\mu_r \pi/2$ mit μ_r als sogenanntem MASLOV- oder MORSE-Index beschreibt die Phasensprünge, wenn die Bahn durch kritische Punkte läuft; μ_r zählt hierbei die Anzahl der kritischen Punkte auf dem r-ten Pfad entsprechend ihrer Vielfachheit. Ein kritischer (oder konjugierten) Punkt – die Verallgemeinerung der eindimensionalen Umkehrpunkte – ist ein Punkt, an dem die Dichte der klassischen Pfade unendlich wird. Weil das Bündel der klassischen Pfade einer Kontinuitätsgleichung genügt, muß die Dimension des ursprünglich D-dimensionalen Bündels an einem kritischen Punkt reduziert werden, und zwar gerade um jene erwähnte Vielfachheit des kritischen Punktes. Die Untermannigfaltigkeiten der kritischen Punkte im $\underline{\mathbf{q}}$-Raum heißen in Analogie zur Optik Kaustiken.

Gemäß (3.8) kann aus $K_{\mathrm{sc}}(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t)$ durch Fouriertransformation ein dann semiklassischer Ausdruck für $G(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, E)$ gewonnen werden [28, 12]:

$$G_{\mathrm{sc}}(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, E) = \frac{1}{i\hbar\sqrt{2\pi i\hbar}^{D-1}} \sum_r |\Delta_r|^{1/2} \exp \left(i \frac{W_r}{\hbar} - \frac{\mu_r \pi}{2} \right) \tag{3.19}$$

$$\Delta_r \equiv - \left(\frac{\partial^2 W_r}{\partial E^2} \right)^{1-D} \det \left(\frac{\partial^2 W_r}{\partial E^2} \frac{\partial^2 W_r}{\partial \underline{\mathbf{q}}_a \partial \underline{\mathbf{q}}_b} - \frac{\partial^2 W_r}{\partial \underline{\mathbf{q}}_a \partial E} \frac{\partial^2 W_r}{\partial \underline{\mathbf{q}}_b \partial E} \right) \tag{3.20}$$

Die Summe erstreckt sich über alle klassischen Pfade der Energie E, die $\underline{\mathbf{q}}_a$ und $\underline{\mathbf{q}}_b$ verbinden; W_r ist die verkürzte klassische Wirkung des r-ten Pfades,

$$W_r \;=\; S_r(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t_0) + E\, t_0 \;=\; \int_{\mathbf{q}_a}^{\mathbf{q}_b} \underline{\mathbf{p}}_r(\underline{\mathbf{q}})\, \mathrm{d}\underline{\mathbf{q}}, \tag{3.21}$$

mit t_0 als der klassischen Laufzeit von $\underline{\mathbf{q}}_a$ nach $\underline{\mathbf{q}}_b$. Bei der Ausführung des Fourierintegrals in stationärer Phase gehen Pfade mit der Laufzeit Null verloren, d. h. (3.19) gilt nicht mehr, wenn $\underline{\mathbf{q}}_a$ und $\underline{\mathbf{q}}_b$ eng zusammenrücken; zu einem auch dann gültigen Ausdruck siehe [12].

Um zur Zustandsdichte (3.13) zu gelangen, ist in $G_{\mathrm{sc}}(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, E)$ $\underline{\mathbf{q}}_a = \underline{\mathbf{q}}_b$ zu setzen – an dieser Stelle erscheinen in der Theorie erstmals geschlossene, aber noch keine periodischen Bahnen – und über $\underline{\mathbf{q}}_a$ zu integrieren. Integration in Sattelpunktsnäherung führt zur Stationaritätsbedingung $\underline{\mathbf{p}}_a = \underline{\mathbf{p}}_b$, d. h. unter allen zum Ausgangsort $\underline{\mathbf{q}}_a$ zurückkehrenden Pfaden werden nur jene berücksichtigt, die wieder mit dem ursprünglichen Impuls in die Bahn zurücklaufen. Erst das und nur das sind periodische Bahnen (Orbits).

Wie erwähnt, werden von der semiklassischen Entwicklung Pfade der Länge Null nicht erfaßt. Deren Beiträge müssen der Zustandsdichte gesondert hinzugefügt werden, die man daher in zwei Anteile aufspaltet,

$$D(E) = D_0(E) + D_{\mathrm{osc}}(E). \tag{3.22}$$

D_0 ist der sogenannte THOMAS-FERMI-Anteil,

$$D_0(E) = \int \frac{\mathrm{d}^D\underline{\mathbf{q}}\, \mathrm{d}^D\underline{\mathbf{p}}}{(2\pi\hbar)^D} \delta\left(E - H(\underline{\mathbf{p}}, \underline{\mathbf{q}})\right) \tag{3.23}$$

der den mittleren semiklassischen Untergrund eben jener Bahnen der Länge Null liefert.

Für den oszillierende Anteil $D_{\rm osc}$ fand GUTZWILLER [31] die Formel:

$$D_{\rm osc}(E) = \sum_{r}^{\rm Orbits} \frac{T_r}{\pi\hbar} \sum_{n=1}^{\infty} \frac{1}{|\det(\mathbf{M}_r^n - \mathbf{I})|^{1/2}} \cos\left[n\left(\frac{S_r}{\hbar} - \frac{\mu_r \pi}{2}\right)\right]. \tag{3.24}$$

Zu summieren ist über alle primitiven klassischen Orbits (r) und alle Mehrfachdurchläufe (n) dieser Orbits. T_r ist die Periodendauer des r-ten primitiven Orbits. $\mathbf{M}_r^n$ wird Monodromiematrix genannt. Sie beschreibt die Stabilität eines Orbits, genauer die Änderungen $\delta\underline{\mathbf{q}}_b^-, \delta\underline{\mathbf{p}}_b^-$ der Endpunkte einer Bahn, wenn im Phasenraum die Anfangspunkte senkrecht zur Bahn um $\delta\underline{\mathbf{q}}_a^-, \delta\underline{\mathbf{p}}_a^-$ variiert werden,

$$\begin{pmatrix} \delta\underline{\mathbf{q}}_b^- \\ \delta\underline{\mathbf{p}}_b^- \end{pmatrix} = \mathbf{M} \begin{pmatrix} \delta\underline{\mathbf{q}}_a^- \\ \delta\underline{\mathbf{p}}_a^- \end{pmatrix}. \tag{3.25}$$

Die Stabilitätseigenschaften der Orbits werden anhand der Eigenwerte von $\mathbf{M}_r^n$ klassifiziert, wobei man zwischen stabilen (parabolischen) – nur in diesem Falle sind die Bewegungsgleichungen integrabel – und instabilen (elliptischen und hyperbolischen) Orbits unterscheidet [32, S. 84]. Ein Orbit kann weiterhin isoliert oder nichtisoliert liegen. In vollständig chaotischen Systemen sind alle Orbits isoliert und instabil.

Gl. (3.24) wird heute als GUTZWILLERs Spurformel bezeichnet. Mit ihr wurde die außergewöhnliche Bedeutung klassisch-periodischer Bahnen für quantenmechanische Spektren zum ersten Mal in dieser Allgemeinheit offenbar, und zugleich rückte die Stabilität der Bahnen, im Gegensatz etwa noch zur Sachlage beim WKB-Verfahren, als zusätzliches Kriterium ins Blickfeld. Vor allem jedoch konnten semiklasssische Quantisierungsbedingungen zum ersten Mal auch für nichtintegrable Systeme abgeleitet werden. Verallgemeinerung erfuhren die Quantisierungsbedingungen später durch MILLER [64].

Tatsächlich erwies sich die Summe dann aber als nahezu pathologisch, da sie bei nichtintegrablen Systemen für beliebige reelle Energien nicht absolut konvergiert (in chaotischen Systemen kann die Länge periodischer Orbits beliebig groß werden). Erst in jüngster Zeit wurden durch Resummationsverfahren einige vielversprechende Fortschritte zur Behebung dieser Schwierigkeit erzielt [10, 11, 41, 21].

3.3 Semiklassische Wellenfunktionen

Verglichen mit dem Stand bei semiklassischen Eigenwertspektren herrscht auf der Ebene der Wellenfunktionen weit weniger Klarheit. Semiklassische Wellenfunktionen in integrablen Systemen (assoziiert mit der klassischen Bewegung auf einem N-dimensionalen Torus im $2N$-dimensionalen Phasenraum) sollten sich deutlich von denen in ergodischen Systemen (assoziiert mit der chaotischen Bewegung auf einer $(2N-1)$-dimensionalen Energieschale im Phasenraum) unterscheiden. Erstere zeigen in der lokal gemittelten Wahrscheinlichkeitsdichte

$$\int_{\Delta V} \mathrm{d}\underline{\mathbf{q}} |\Psi(\underline{\mathbf{q}})|^2 \tag{3.26}$$

eine erhöhte Intensität an den Kaustiken (den Grenzen des klassisch erlaubten Gebiets), und räumlich sind die verschiedenen Teile (etwa die Knotenlinien) stark miteinander korreliert [7].

Für Wellenfunktionen in chaotischen Systemen wurde in der Erwartung, daß hochangeregte Zustände mit Wigner-Funktionen korrespondieren sollten, die auf der Energieschale homogen sind, ein irreguläres Aussehen ohne nennenswerte räumliche Korrelationen vermutet [7]. Die konkrete Form der Wellenfunktionen allerdings war unbekannt.

Erste numerische Experimente von McDonald und Kaufmann [59] schienen diese Hypothese zu bestätigen: die Kontenlinien hochangeregter Zustände eines Stationbillards zeigten einen irregulären mäanderförmigen Verlauf. Indes wurden schon damals auch andere Wellenfunktionen beobachtet [58] (zitiert nach [16]), jedoch erst 1988 veröffentlicht [60].

Daher war es Heller [35], der den überraschenden Tatbestand erstmals publik machte, daß Wellenfunktionen in chaotischen Systemen häufig entlang instabiler periodischer Orbits konzentriert sind bzw. sein müßten. Diese Erscheinungen wurden von ihm „Scars“ (zu dt. „Narben“) genannt. Zugleich gab er eine erste, noch mehr heuristische Erklärung für dieses Phänomen anhand der semiklassischen Propagation eines Gaußschen Wellenpaketes. In der Folge fand man Scars auch in anderen chaotischen Systemen [80, 25], sie konnten darüberhinaus sogar direkt in Mikrowellenexperimenten [76, 75, 73] und indirekt an Wassserstoffatomen in Mikrowellenfeldern [38] beobachtet werden.

Es erscheint zunächst nicht zwingend, daß periodische Bahnen auch für die einzelne Wellenfunktion eine Sonderstellung einnehmen sollten. Die lokale Zustandsdichte $\varrho(\underline{\mathbf{q}}, E)$ gemäß (3.12) stellt sich in der Fomulierung über den Zeitpropagator zwar dar als Fourierintegral über alle geschlossene Bahnen,

$$\varrho(\underline{\mathbf{q}}, E) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}t \; K(\underline{\mathbf{q}}, \underline{\mathbf{q}}, t) \; e^{\frac{\mathrm{i}}{\hbar} Et} \tag{3.27}$$

keinesfalls jedoch nur über periodische. Erst die in semiklassischer Näherung ausgeführte Ortsintegration, sprich Spurbildung in (3.13), die von $\varrho(\underline{\mathbf{q}}, E)$ zur Zustandssumme $D(E)$ führt, projizierte bei der Herleitung der Gutzwiller-Formel unter allen geschlossenen Bahnen gerade die periodischen heraus.

Erforderlich ist daher wenigstens eine lokale Ortsmittelung wie in (3.26), um semiklassisch genähert dann auch in Ausdrücken für die Wellenfunktion zu periodischen Bahnen zu gelangen. Das Scar-Phänomen läßt hier aber zumindest Raum für die Vermutung, daß periodische Orbits Wellenfunktionen auch ganz elementar betreffen und Eigenzustände möglicherweise sogar ganz, d. h. ohne lokale Mittelung auf periodischen Bahen zurückführbar sein könnten (vgl. den folgenden Abschnitt). Unabhängig davon wird zunehmend der Versuch unternommen, neben den Energiespektren nun auch die einzelne semiklassische Wellenfunktion im Rahmen einer Theorie periodischer Orbits zu beschreiben. Wie bei den Spurformeln liegen die Hauptschwierigkeiten dabei in den Konvergenzproblemen einer Summe über alle Orbits.

Bogomolny [16] stellte als erster eine Beziehung zwischen den Wellenfunktionen und den periodischen Bahnen eines chaotischen Systems her. Aus der obengenannten Schwierigkeit heraus betrachtete er energetisch und lokal gemittelte Ausdrücke von Eigenzuständen,

$$\int_{\Delta V} \mathrm{d}\underline{\mathbf{q}} \, \frac{1}{N} \sum_{\{n\}} |\Psi_n(\underline{\mathbf{q}})|^2,$$

wobei die Summe über die Eigenzustände eines schmalen Energiefensters $[E_0 - \Delta E, E_0 + \Delta E], \Delta E \ll E_n$ läuft. Die Ortsmittelung führt in semiklassischer Näherung zu periodischen Bahnen und die Energiemittelung dazu, daß nur kurze Orbits beitragen, wodurch die

Konvergenz einer Orbit-Summe garantiert wird. Er gelangt zu einer Formulierung für die gemittelte Wellenfunktion, in der sich diese aus den Beiträgen einzelner Orbits aufgebaut darstellt. Die Energieabhängigkeit der einzelnen Beiträge ist infolge der semiklassisch großen Wirkungsphase beträchtlich; ändert sich der Energieschwerpunkt E_0, verschieben sich die relativen Gewichte der Orbits. In der gemittelten Wellenfunktion können bei verschiedenen Energien also ganz verschiedene periodische Bahnen dominieren.

BERRY [9] dehnte die BOGOMOLNYsche Theorie auf den Phasenraum aus und leitete semiklassische Ausdrücke für die spektrale Wignerfunktion her.

AGAM und FISHMAN [1] gelang es schließlich erstmals, Ausdrücke für individuelle Eigenfunktionen zu erhalten. Ausgehend von einer Resummationsvorschrift von BERRY und KEATING [11] finden sie Darstellungen für individuelle Wignerfunktionen in chaotischen Systemen, deren Projektion auf den Ortsraum dann die einzelnen Eigenfunktionen liefern. Insgesamt ist das Verfahren ausgesprochen kompliziert. Von denselben Autoren werden in [2] semiklassische Kriterien für das Auftreten von Scars in chaotischen Systemen diskutiert.

3.4 Ein spekulativer Ausflug

Dieser Abschnitt enthält einige Überlegungen zum Teil recht spekulativer Natur zum Verhältnis von Quantentheorie und periodischen Bahnen. Man nehme ihn als Anregung insbesondere für einen möglichen Ausbau einer Theorie periodischer Oribts. Zugleich werden die in Kapitel 5 für eine anschauliche Deutung der numerischen Resultaten nützlichen Bilder hier motiviert.

Wir beginnen mit einem erstaunenswerten Ergebnis aus GUTZWILLERs erster Arbeit [28]. Dessen ursprüngliches Ziel war eigentlich gewesen, auf semiklassischem Wege genäherte zwar, aber analytische Ausdrücke für Atom- und Moleküleigenfunktionen zu finden, weshalb er insbesondere die von Energie und Impuls abhängige Greensche Funktion untersuchte,

$$F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E) = \int \mathrm{d}\underline{\mathbf{q}}_a \int \mathrm{d}\underline{\mathbf{q}}_b \, G(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, E) \, e^{\frac{i}{\hbar}(\mathbf{p}_a \mathbf{q}_a - \mathbf{p}_b \mathbf{q}_b)} \tag{3.28}$$

$$= \sum_n \frac{\chi_n(\underline{\mathbf{p}}_b)\chi_n^*(\underline{\mathbf{p}}_a)}{E - E_n}, \tag{3.29}$$

deren Pole die Eigenwerte und deren zugehörige Residuen die Produkte der Eigenzustände in Impulsdarstellung

$$\chi_n(\underline{\mathbf{p}}) = \int \mathrm{d}\underline{\mathbf{q}} \, \phi_n(\underline{\mathbf{q}}) \, e^{\frac{i}{\hbar}\mathbf{p}\mathbf{q}} \tag{3.30}$$

liefern. Sein Hauptargument für $F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$ (und gegen $G(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, E)$) war, daß es im Falle eines typischen Atom- oder Molekülpotentials immer eine klassische Trajektorie gibt, die bei gegebener negativer Energie $E < 0$ zwei Impulse $\underline{\mathbf{p}}_a$ und $\underline{\mathbf{p}}_b$, die sich so ja nur in ihrer Richtung unterscheiden können, miteinander verbindet, während eine solche Tajektorie für zwei Koordinaten $\underline{\mathbf{q}}_a$ und $\underline{\mathbf{q}}_b$ nur existiert, wenn beide Punkte im (selben) klassisch erlaubten Gebiet liegen. Von $F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$ versprach er sich daher bessere Approximationseigenschaften, vor allem mit Blick auf die in klassisch verbotenen Gebieten liegenden Ausläufer der Eigenzustände.

Das Ergebnis seiner Rechnung war verblüffend: Die semiklassische Approximation des Wasserstoffproblems (des Keplerproblems) lieferte für den Propagator (3.28) zwar zunächst nur einen genäherten Ausdruck $\tilde{F}(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$, dessen Pole und zugehörige Residuen stellten

dann jedoch überhaupt keine Näherungen mehr dar, sondern erwiesen sich als die exakten quantenmechanischen *Eigenwerte* und *Eigenzustände*[3]. Dieses Ergebnis erscheint ausgesprochen mysteriös, wird hier doch in Gestalt der Wellenfunktionen die Quantenmechanik selbst beliebig tief im klassisch verbotenen Gebiet durch eine semiklassische Entwicklung vollständig reproduziert[4]. Die gleiche Exaktheit findet sich dann auch beim harmonischen Oszillator, wo sie zumindest mathematisch verständlicher wird, da die semiklassische Näherung (3.18) des Pfadintegrals (3.16) bis zu quadratischen Gliedern im Potential exakt ist, wenngleich die begrifflichen Schwierigkeiten bleiben.

Ebenso führen die Vorteile von $F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$, die in [29] zudem relativiert wurden, hier nicht weiter. Eine semiklassische Entwicklung (quadratischer Ordnung) zentriert sich im Ortsraum wie im Impulsraum immer um den klassischen Pfad und ist nur in einer begrenzten Umgebung der klassischen Lösung exakt (und fängt nicht beliebig große nichtklassische Impulse ein, wie es tief im verbotenen Gebiet erforderlich wäre.

An dieser Stelle sei ein radikaler Umkehrschluß als Hypothese erlaubt:

> *Die Observablen eines stationären Quantensystems im Sinne der herkömmlichen Quantentheorie lassen sich grundsätzlich immer exakt zurückführen auf eine Summe über alle – verallgemeinerten – periodischen Bahnen dieses Systems.* (3.31)

Unter *verallgemeinerten* periodischen Bahnen verstehen wir hierbei entweder komplexe Orbits (mit komplexen Wirkungen) oder solche klassisch-reellwertigen, die als „ungebundene Orbits“ über den unendlich fernen Punkt laufen. Letztlich sind natürlich auch komplexe Orbits, für die Umkehrpunkte nicht existieren, ungebunden. Unten werden wir ein mögliches Kriterium angegeben, um aus der Menge der ungebundenen Bahnen die „Orbits“ auszuwählen, wesentlich an dieser Stelle ist, daß für eine vollständige Rekonstruktion der Quantenmechanik prinzipiell Bahnen vonnöten sind, die klassisch verbotene Bereiche entweder durchqueren können (komplex) oder überqueren (ungebunden reell). Und da eine diesbezügliche Entscheidung hier offenbleiben muß, verwenden wir den Begriff der *verallgemeinerten Orbits* als einen, der beide Möglichkeiten einschließt.

Die exakte Lösbarkeit des Keplerproblems und des harmonischen Oszillators durch einen semiklassischen Ansatz würde sich in das Schema (3.31) dann in der Weise einordnen, daß jene beiden Systeme gerade diejenigen zwei sphärischen Potentialformen sind, bei denen jede klassische (gebundene) Lösung auch streng periodisch ist [48, S. 39], wodurch bei GUTZWILLERs semiklassischer Approximation des Propagators $F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$ gewissermaßen automatisch eine derartige Summe über alle periodischen Bahnen ausgeführt wurde. Verallgemeinerte Orbits wären bei diesen beiden Systemen insofern noch nicht zwingend, als hier auch klassische Orbits (reellwertig und gebunden) existieren, die den gesamten Ortsraum durchlaufen.

Zu betonen ist, daß falls eine Identität wie (3.31) bestehen sollte, sich diese unmittelbar nur an Beziehungen wie (3.8) oder (3.28) für die Propagatoren oder (3.27) für die Zustandsdichte und nur mittelbar am einzelnen Eigenzustand offenbaren kann, da Orbits aller

[3] Im Lichte der aus $\tilde{F}(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$ folgenden exakten Eigenwerte und Eigenzustände sollte man bei $\tilde{F}(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$ eigentlich nicht von einer Näherung sprechen, sondern lediglich von einer in der komplexen Ebene „etwas anderen“, physikalisch aber gleichwertigen Darstellung des Propagators $F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$. Sind die exakten Eigenwerte und Eigenzustände einmal bekannt, kann rückwirkend über (3.29) ja auch immer $F(\underline{\mathbf{p}}_b, \underline{\mathbf{p}}_a, E)$ rekonstruiert werden.

[4] Dagegen gelang eine Lösung des für die Quantenmechanik so wichtigen Wasserstoffproblems im echten zeitgegitterten Formalismus des Funktionalintegrals erst vor wenigen Jahren [24, 43].

Energien vonnöten sind, während der einzelne Eigenzustand erst hinterher aus dem Propagator herausprojiziert wird. Das Entstehen der exakten quantenmechanischen Ausdrücke für die klassisch verbotenen Gebiete hätte man sich dabei so vorzustellen, daß klassische Orbits mit $E > V(\underline{\mathbf{q}})$ – beim Keplerproblem z. B. mit $E \approx 0$ – diese Bereiche durchlaufen und die entsprechenden Informationen in den Propagator einbringen. (Wir formulieren dies so vage, weil GUTZWILLERs semiklassische Approximation des Keplerproblems direkt ja eben zunächst nur einen genäherten – oder mit Fußnote 3 „einen etwas anderen" – Propagator lieferte, aus dem dann aber die exakten Eigenwerte und Zustände folgten[5]).

Das Pfadintegral nun zeigt, daß der Bahnbegriff, wenngleich in einer abstrakteren Form und dann stets im Sinne von „alle Bahnen zugleich", auch der Quantenmechanik immanent ist. Und tatsächlich erscheint es gerade von diesem Standpunkt aus nicht unnatürlich, eher schon von einiger Zwangsläufigkeit, daß in einem stationären System, in dem gewissermaßen per Definition alles schon unendlich lange läuft, jede nichtperiodische Bahn, jeder nie wiederkehrende Vorgang gegenüber allem Wiederkehrenden gänzlich zur Bedeutung Null herabsinken muß. Es ist genaugenommen erst diese Vorstellung des *unendlichfachen* Aufsummierens *identischer* (Pfadintegral-)Beiträge, die für periodische Bahnen in stationären Quantensysteme nicht nur eine hervorgehobene, sondern eine ganz und gar singuläre Stellung in dieser Allgemeinheit nahelegt, und die obige Hypothese (3.31) im wesentlichen trägt.

Angenommen nun, ein Teilchen bewege sich in einem stationären Quantensystem tatsächlich auf allen periodischen Bahnen zugleich und auf keinen weiteren sonst („Teilchen" natürlich im mehr abstrakteren Sinne des Pfadintegrals). Weil der positive und der negative Umlaufsinn gleichberechtigt sind, bewege es sich außerdem in beiden Umlaufrichtungen zugleich. Letzteres könnte man als Zugleich von positiver und negativer Zeitrichtung auffassen und es *Simultanität der Zeitpfeile* nennen. Die Darstellung der Zustandsdichte (3.27) kann im Grunde ebenfalls schon als eine solche Simultanität von Vorwärts- und Rückwärtsausbreitung gesehen werden.

Würde man, um Hypothese (3.31) etwas zu konkretisieren, die semiklassischen Formeln, bei denen stets über alle periodischen Bahnen und diese jeweils gewichtet mit ihrer Stabilitätsmatrix (Monodromiematrix) summiert wird, als direkte Vorlage wählen, erhielte man etwa folgendes Bild: *Ein Teilchen bewegt sich auf allen periodischen Bahnen zugleich, wobei jede Bahn mit der Stabilitätsmatrix gewichtet ist.* Dieses Bild ist nicht sehr attraktiv, da der Stabilität einer Bahn kaum die „Aura" einer primären Eigenschaft anhaftet, die bereits auf der „untersten" Ebene erscheinen sollte. Eher weist eine Eigenschaft wie die Stabilität auf das Vorliegen einer äußeren Störung hin, die dem System diese Antwort abverlangt. Es böte sich an, das Auftreten der Stabilitätsmatrix mit dem quantenmechanischen Meßprozeß selbst in Zusammenhang zu bringen: *Ein Teilchen bewegt sich auf allen periodischen Bahnen zugleich. Der Meßprozeß stört das System dergestalt, daß jede Bahn mit ihrer Stabilitätsmatrix gewichtet in die Messung eingeht.* Wie diese Störung auszusehen hätte, wäre natürlich der springende Punkt. Klar ist zumindest, daß man sich in diesem Moment bereits nicht mehr auf dem Boden der herkömmlichen Quantentheorie bewegt, da man aus deren Sicht jetzt mit verborgenen Parametern operiert.

[5] In diesem Zusammenhange auffallend, daß zu jenem Bild, das mit der Pfadintegraldarstellung (3.16) des Zeitpropagators $K(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, t)$ gewöhnlich verbunden wird – „Ein Quantenteilchen, das von $(\underline{\mathbf{q}}_a, t_a)$ nach $(\underline{\mathbf{q}}_b, t_b)$ läuft, nimmt alle Pfade $\underline{\mathbf{q}}(t)$ zugleich." – kein Pendant, kein ebenso eindeutig „klassisch" formuliertes Bild für den energieabhängigen Propagator $G(\underline{\mathbf{q}}_b, \underline{\mathbf{q}}_a, E)$ existiert, bei dem die Begriffe Ort und Energie statt Ort und Zeit vorkommen müßten.

Die obige Vorstellung einer simultanen Bewegung in beiden Umlaufrichtungen wurde dort mit einer „Simultanität der Zeitpfeile“ assoziiert. Eine solche Simultanität stellt im Grunde die formal vollständigste Verkörperung jedweder Reversibilität dar: Der zeitgespiegelte Vorgang erscheint in den Gleichungen nicht nur als mathematisch gleichwertige Möglichkeit, sondern als sogar simultan stattfindend. Irreversible Vorgänge, etwa im Sinne der Thermodynamik oder auch der quantenmechanische Meßprozeß, könnten in Fortschreibung dieses Gedanken dann als Brechung dieser Symmetrie erscheinen – ein formales Prinzip fraglos nicht ohne Reiz.

Der Gedanke eines simultanen Umlaufs in beiden Richtungen stellt jedenfalls anheim, das Besondere periodischer Bahnen gegenüber allen sonstigen geschlossenen Bahnen außer in dem Umstand, daß die Bewegung $\underline{\mathbf{q}}(t)$ nach jeweils der vollen Periode T wieder mit dem ursprünglichen Geschwindigkeitsvektor $\dot{\underline{\mathbf{q}}}(t)$ in die Bahn zurückläuft,

$$\underline{\mathbf{q}}(t+T) = \underline{\mathbf{q}}(t) \quad \Leftrightarrow \quad \dot{\underline{\mathbf{q}}}(t+T) = \dot{\underline{\mathbf{q}}}(t) \; , \tag{3.32}$$

außerdem auch darin zu sehen, daß nach jeweils der halben Periode $T/2$ bereits der positive und der negative Umlauf, $\underline{\mathbf{q}}_+(t)$ und $\underline{\mathbf{q}}_-(t)$, zusammenstoßen und zwar mit entgegengesetzten Geschwindigkeitsvektoren:

$$\underline{\mathbf{q}}_+(t+T/2) = \underline{\mathbf{q}}_-(t+T/2) \quad \Leftrightarrow \quad \dot{\underline{\mathbf{q}}}_+(t+T/2) = -\dot{\underline{\mathbf{q}}}_-(t+T/2) \tag{3.33}$$

Hierüber läßt sich recht zwanglos ein Kriterium aufstellen, um selbst unter Bahnen, die aus dem Unendlichen kommen und im Unendlichen verschwinden, noch die „Periodischen“ auszuwählen, nämlich als jene, die obiger Anschlußbedingung für $T/2 \to \infty$ genügen:

$$\begin{array}{l} \lim\limits_{t\to\infty} \underline{\mathbf{q}}_+(t) \to \infty \\ \lim\limits_{t\to\infty} \underline{\mathbf{q}}_-(t) \to \infty \end{array} \quad \Leftrightarrow \quad \lim_{t\to\infty} \dot{\underline{\mathbf{q}}}_+(t) = -\lim_{t\to\infty} \dot{\underline{\mathbf{q}}}_-(t) \tag{3.34}$$

Das „Zusammenstoßen“ im Unendlichen veranschaulicht man sich für zwei Raumdimensionen am besten anhand einer Konstruktion wie der Riemanschen Zahlenkugel, bei der die Punkte der Ebene eindeutig auf eine Kugeloberfläche projiziert werden und alle im Unendlichen gelegenen Punkte auf dem Nordpol der Kugel koalizieren. Ungebundene Orbits im obigen Sinne wären hierdann Bahnen, die auf der Riemanschen Zahlenkugel den Nordpol knickfrei (stetig differenzierbar) durchlaufen, diesen also aus zwei diametral entgegengesetzten Richtungen erreichen. In drei Dimensionen hätte man sich den flachen dreidimensionalen Raum auf die Oberfläche einer vierdimensionalen Kugel projiziert vorzustellen. Bei zentralsymmetrischen Problemen wie dem Keplerproblem z. B. wären solche ungebundenen Orbits nur freie Bahnen mit dem Drehimpuls Null.

Dienlich wird eine solche Erweiterung auf ungebundene Orbits, wenn etwa im Sinne von Hypothese (3.31) oder auch für einen weiterführenden semiklassischen Ansatz Orbits in Betracht zu ziehen sind, die klassisch verbotene Gebiete durch- bzw. überqueren und bis Unendlich laufen, um hierüber z. B. die Wellenfunktionen einschließlich evanescenter Schwänze für *endlich* hohe Potentialkästen auf periodische Bahnen zurückzuführen. Angeregt wird diese Ausdehnung (auch schon für einen semiklassischen Ansatz) durch unsere numerischen Resultate in 5.4 und 5.6. Dabei stellt sich die vermeintliche Hauptschwierigkeit des ehrgeizigen vollständigen Programms (3.31) – die exakte Beschreibung der evanescenten, bis Unendlich reichenden Ausläufer der Wellenfunktionen – bereits scharf beim simplen Beispiel des eindimensionalen Potentialtopfes endlicher Höhe; eine exakte Rekonstruktion dieses Systems mittels verallgemeinerter Orbits sollte als erstes versucht werden.

Das Wesentliche dieses Abschnitt noch einmal zusammengefaßt: Ausgehend von einer durch das Pfadintegral insistierten Immanenz eines Bahnbegriffs auch in der Quantenmechanik und getragen von der Vorstellung, daß in einem stationären – einem unendlich lange laufenden – System alles Periodische gegenüber allem Nichtperiodischen von singulärer Bedeutung sein sollte, wurde die vollständige und grundsätzliche Zurückführbarkeit der stationären Quantenmechanik auf die Gesamtheit der periodischen Bahnen eines Systems als Möglichkeit erwogen. Die aus Gründen der Symmetrie hierdann auf jeder Bahn anzunehmende simultane Bewegung in beiden Umlaufrichtungen kann als eine Simultanität der Zeitpfeile aufgefaßt werden, die – als Prinzip genommen – den formal vollständigsten Ausdruck für die Reversibilität eines Systems darstellen würde. Der Gedanke eines simultanen Umlaufs in beiden Richtungen ermöglicht zudem eine zwanglose Ausdehnung des Begriffs der periodischen Bahnen auch auf solche, die über den unendlich fernen Punkt laufen.

Kapitel 4

Berechnung gebundener Zustände in mehrdimensionalen Potentialmulden

4.1 Problemstellung

Dieses Kapitel befaßt sich mit der numerischen Lösung der zeitfreien Schrödingergleichung

$$\left[\frac{\partial^2}{\partial \underline{\mathbf{r}}^2} + V(\underline{\mathbf{r}})\right] \psi(\underline{\mathbf{r}}) = E\psi(\underline{\mathbf{r}}) \;, \qquad \underline{\mathbf{r}} \in \underline{\mathbf{R}}^d \tag{4.1}$$

für beliebig geformte zwei- oder dreidimensionale Potentialmulden. Insbesondere sind mesoskopische Systemabmessungen, also etwa $L > 10\lambda$, wenn λ die de Broglie-Wellenlänge und L eine typische Systemlänge ist, ins Auge gefaßt, wodurch rechentechnische Leistungsgrenzen in praxi schnell erreicht sind. Konkret werden wir (4.1) dann ein Kastenpotential

$$V(\underline{\mathbf{r}}) \;=\; V_0 \;\Theta\big(f(\underline{\mathbf{r}})\big) \;=\; \begin{cases} 0 & \text{für } \underline{\mathbf{r}} \in \Omega_K \\ V_0 & \text{für } \underline{\mathbf{r}} \notin \Omega_K \,, \end{cases} \tag{4.2}$$

zugrundelegen, bei dem $\Theta(f)$ die Stufenfunktion ist und und $f(\underline{\mathbf{r}}) = 0$ die Berandung des Kastengebietes Ω_K in einem d-dimensionalen Ortsraum. Die Höhe des Randes beträgt V_0, d. h. der Energienullpunkt liegt auf dem Kastenboden. (Von der Kasteneigenschaft wird im numerischen Verfahren selbst kein Gebrauch gemacht.) Gesucht sind nunmehr Eigenfunktionen und Energieeigenwerte gebundener Zustände, und zwar speziell in der Weise, daß die zu einer fixierten, ansonsten aber beliebig wählbaren Energie $\epsilon < V_0$ (z. B. der Fermienergie) nächstbenachbart gelegenen zu finden sind. Die Frage nach jeweils nur einigen Eigenlösungen ist numerisch vernünftig oder eher die einzig mögliche Art zu fragen angesichts der Vielzahl von Zuständen in einem großen System. Alle Eigenfunktionen wären sukzessive dann durch „Durchstimmen“ von ϵ zu erhalten.

4.2 Diskretisierung im Differenzenschema

4.2.1 Vorbemerkungen

Numerische Lösungen von Differentialgleichungen (DGL) verlangen eine Diskretisierung (also eine Approximation) des differentiellen Problems, die üblicherweise entweder im Gefolge einer Basissatzentwicklung im Funktionenraum (Fourierraum) oder direkt im Raum der unabhängigen Variablen (dem Ortsraum hier) vonstatten geht.

Ist anzunehmen, daß die gesuchten (und a priori ja immer unbekannten) Lösungen hinreichend gut nach einem geeigneten Basissatz entwickelbar sein sollten, sind Basissatzentwicklungen oft das Mittel der Wahl. Als hierfür prädestiniert erweisen sich durch einfache Symmetrien geprägte Aufgaben, denn da sich numerisch immer nur endlich viele Entwicklungsglieder berücksichtigen lassen, die Basissätze also unvollständig bleiben, sollten bereits die Ansatzfunktionen wesentliche (Symmetrie-)Eigenschaften der späteren Lösung wiederspiegeln. Andernfalls geraten entweder die Diskretisierungsfehler zu groß oder es müssen unverhältnismäßig viele Glieder mitgenommen werden, weshalb für unsere beliebig geformten Elektroden dieses Vorgehen weniger geeignet ist.

Demgegenüber ersetzen direkte Diskretisierungverfahren den Ortsraum in einem begrenzten Gebiet durch eine diskrete Struktur, wobei die erreichbare Größe des Gebietes von der Rechenkapazität und der Feinheit der Maschen abhängt. Die gebräuchlichsten Verfahren für partielle DGL sind das Differenzenverfahren und die Finite Elemente Methode (FEM).

Ersteres überführt die partielle DGL auf einem fiktiven Gitter in eine Differenzengleichung und damit in ein algebraisches Gleichungssystem, das bei der SGL (auch für die FEM) natürlich zugleich ein Eigenwertproblem sein wird.

Die FEM [61] zerlegt das Grundgebiet in kleine endliche (finite) Elemente, auf denen einfachste Funktionen, z. B. Polynome zweiten und dritten Grades, definiert werden, deren Koeffizienten dann über ein geeigentes Funktional zu bestimmen sind. Die FEM, die in der moderenen Ingenieurmathematik eine enorme Bedeutung besitzt, ist die mächtigere Methode, da eine Anpassung der Maschenweite an lokale Gegebenheiten oder sogar während des Rechnung viel leichter zu bewerkstelligen ist als im Differenzenverfahren.

Jedoch gilt es folgendes zu bedenken: Die uns interessierenden Lösungen sind hochangeregete Schwingungen (und keine Verformungen oder Spannungen), d. h. die minimale Maschenweite wird durch die Wellenlänge bestimmt und muß ohnehin im gesamten Gebiet in etwa konstant gehalten werden. Ein dichteres Netz böte sich an höchstens zur getreueren Wiedergabe der Elektrodengeometrie. Deren genaue Form ist allerdings auf der Ebene der Diskretisierungsschrittweite relativ unerheblich, da ein etwas gröberer Polygonzug ebensogut für eine prinzipielle Betrachtung zur Rastertunnelmikroskopie taugt, wie ein feinerer.

Vor allen Dingen aber steht uns die Hauptschwierigkeit, die Lösung des mächtigen Eigenwertproblems, noch bevor. Die Diskretisierungsmatrix sollte hierzu von einfachster Gestalt sein, weshalb variable Schrittweiten gänzlich ausgeschlossen werden müssen. Die FEM besäße nun keine Vorteile mehr und wir wählen das konzeptionell einfachere Differenzenverfahren.

Anzumerken bleibt, daß speziell für die Helmholtz-Gleichung $(\boldsymbol{\nabla}^2 + k^2)\psi(\underline{\mathbf{r}}) = 0$, also für Gebiete konstanten Potentials, eine weitere numerische Möglichkeit dadurch existiert, daß Funktionswerte $\psi(\underline{\mathbf{r}})$ im Inneren dort vollständig durch Integrale über den Gebietsrand dargestellt werden können. Ridell [67] entwickelte auf dieser Basis ein Verfahren zur Lösung

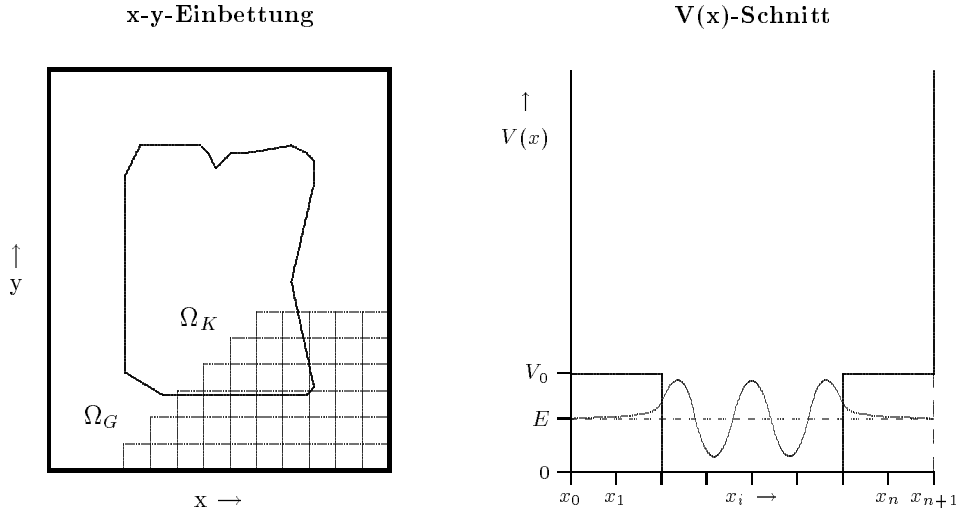


Abbildung 4.1: *Einbettung eines 2d-Potentialkastens Ω_K in ein rechteckiges Grundgebiet Ω_G mit unendlich hohen Wänden (links) und dessen Diskretisierung auf einem äquidistanten Gitter (rechts).*

der zweidimensionalen Helmholtz-Gleichung im Inneren beliebiger Polygone, mit dem auch die meisten der bisherigen numerischen Resultate zu Quantenbilliards erzielt wurden. Der Vorteil ist die reduzierte Dimension des Datenraums (Linie statt Fläche, Fläche statt Volumen etc.). Diese Möglichkeit entfällt jedoch, wenn das Potential wie bei uns im Grundgebiet ortsabhängig wird.

4.2.2 Prinzipielles zum Differenzenschema

In (4.1) liegt eine elliptische Randwertaufgabe vor, deren originale Randbedingung das Verschwinden gebundener Lösungen im Unendlichen fordert. Die numerische Behandlung muß hingegen auf einem endlichem Gitter stattfinden. Wir erzwingen ein Verschwinden deshalb bereits im Endlichen, indem der Potentialkasten in ein größeres Grundgebiet Ω_G eingebettet wird, das von unendlich hohen Wänden begrenzt ist und dessen Berandung überall hinreichend weit von der Berandung des interessierenden Potentialkastens entfernt bleibt (Abb. 4.1). Auf dem Rand von Ω_G ist dann $\psi = 0$ zu setzen. Dieses künstliche Grundgebiete ermöglicht, daß die gebundenen Zustände außerhalb des Kastens Ω_K praktisch ungestört abklingen können, und liefert gleichzeitig die numerisch erforderliche Randbedingung für das Verschwinden der Wellenfunktionen innerhalb eines endlichen Gebiets.

Das Grundschema des Differenzenverfahrens sei zunächst in einer Dimension skizziert: Über das Einbettungsgebiet Ω_G, in diesem Falle die Länge L_x, wird ein Raster von äquidistanten Stützstellen $x_{i+1} = x_i + h$ mit der Schrittweite h gelegt. Die Ränder von Ω_G und auch von Ω_K sind im einfachsten Falle dahingehend zu korrigieren, daß sie auf den ihnen am nächsten befindlichen Rasterpunkten zu liegen kommen, was zu den neuen Gebieten Ω_G^* und Ω_K^* Anlaß gibt. Wir erhalten n innere Stützstellen $x_1, \ldots, x_n$ und die Randpunkte x_0 und x_{n+1} mit den zugehörigen Funktionswerten $\psi_i \equiv \psi(x_i)$ – den „Auslenkungen“ – und dem Potential $V_i \equiv V(x_i)$. Der zweite Differentialquotient am Punkte x_i ist gemäß

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x) \;\rightarrow\; \frac{\psi_{i-1} - 2\psi_i + \psi_{i+1}}{h^2}\,, \qquad x_{i+1} = x_i + h \tag{4.3}$$

in einen Differenzenquotienten zu überführen, woraus eine Abhängigkeit unmittelbar benachbarter Gitterpunkte resultiert und für die eindimensionale SGL das schwach gekoppelte Gleichungssystem

$$\frac{1}{h^2}\left[(2+V_i h^2)\psi_i - \psi_{i-1} - \psi_{i+1}\right] = E\psi_i, \qquad (i = 1, \ldots, n) \tag{4.4}$$

mit der Matrixdarstellung

$$\frac{1}{h^2}\begin{pmatrix} 2+\tilde{V}_1 & -1 & & & \\ -1 & 2+\tilde{V}_2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2+\tilde{V}_{n-1} & -1 \\ & & & -1 & 2+\tilde{V}_n \end{pmatrix}\begin{pmatrix} \psi_1 \\ \\ \vdots \\ \\ \psi_n \end{pmatrix} = E\begin{pmatrix} \psi_1 \\ \\ \vdots \\ \\ \psi_n \end{pmatrix} \tag{4.5}$$

entsteht ($\tilde{V}_i \equiv V_i h^2$). Dank der homogenen Randbedingungen $\psi_0 = \psi_{n+1} = 0$ treten hier die Randpunkte des Einbettungsgebietes überhaupt nicht mehr in Erscheinung, die i. allg. sonst bei ψ_1 und ψ_n einfließen würden. (Ein fester inhomogener Rand wäre im gegebenen Fall allerdings keine korrekt gestellte Aufgabe, da ein gänzlich anderer Lösungstyp mit $E = 0$ und ohne Eigenzustände die Folge wäre [6].)

Die Lösung von (4.5) verlangt nunmehr die Lösung des Eigenwertproblems

$$\mathbf{A}\underline{\mathbf{u}} = E\underline{\mathbf{u}}, \qquad \underline{\boldsymbol{u}} \equiv (\psi_1, \ldots, \psi_n)^T, \tag{4.6}$$

wo $\mathbf{A} \in \mathbf{S}^{n,n}$ die symmetrische Diskretisierungsmatrix ist, E der Eigenwert und $\underline{\mathbf{u}}$ ein n-dimensionalen Spaltenvektor, in dem die gesuchten Funktionswerte der inneren Punkte des Einbettungsgebietes zusammengefaßt sind.

Man könnte versucht sein, im Außenraum der Elektroden (im monoton abklingenden Teil der Wellenfunktion) die Schrittweite größer und damit ortsabhängig zu gestalten ($h \to h_i = x_{i+1} - x_i$), um Diskretisierungspunkte zu sparen. Der Differenzenoperator wäre in

$$\psi_i'' = \frac{\frac{\psi_{i+1}-\psi_i}{h_{i+1}} - \frac{\psi_i - \psi_{i-1}}{h_i}}{(h_{i+1}+h_i)/2} = \frac{\psi_{i+1}}{h_{i+1}H_i} - 2\frac{\psi_i}{h_{i+1}h_i} + \frac{\psi_{i-1}}{h_i H_i},$$

mit $H_i \equiv (h_{i+1} + h_i)/2h$ abzuwandeln, infolgedessen die Diskretisierungsmatrix jedoch sofort unsysmmetrisch würde. Da sich Eigenwertberechnungen aber wesentlich vereinfachen, wenn Matrizen symmetrisch sind, wir andererseits eher zu mesoskopische Abmessungen der Potentialkästen streben, bei denen der relative Anteil des Außenraums zurückgeht, legen wir hinfort nur noch konstante Schrittweiten (äquidistante Gitter) zugrunde, und zwar sowohl innerhalb einer Koordinatenrichtung als auch bezüglich der verschiedenen Raumachsen[1],

$$h_x = h_y = h_z = h = \text{konst.} \tag{4.7}$$

Daß zwanglos daraus symmetrische Diskretisierungsmatrizen resultieren, bleibt ursächlich natürlich eine Folge des symmetrischen Differentialoperators in (4.1).

[1] Ein Differenzenschema für (4.1) mit konstanten, jedoch unterschiedlichen h_x, h_y und h_z ließe die Symmetrie zwar ebenfalls unangetastet, würde in zwei und drei Dimensionen aber nicht mehr das Herausziehen eines einheitlichen Faktors $1/h^2$ wie in (4.5) gestatten, was den Matrizen aber gerade ihre extrem einfache und vom numerischen Standpunkt äußerst vorteilhafte Gestalt verleiht.

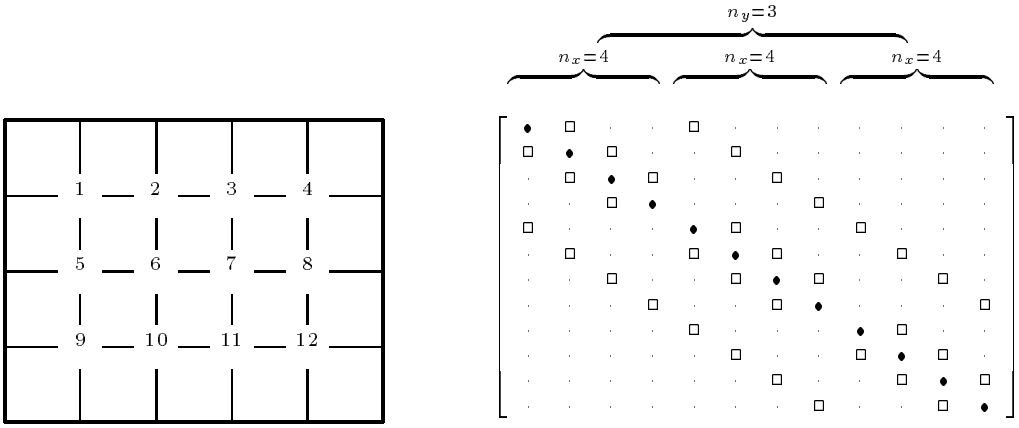


Abbildung 4.2: *Links das Beispiel eines rechteckigen Grundgebietes mit* 4×3 *inneren Punkten, die fortlaufend (lexikographisch) durchnumeriert sind. Rechts die daraus resultierende, mit* h^2 *multiplizierte Eigenwertmatrix. Die Hauptdiagonalememente "•" lauten* $a_{ii} = 4 + V_i h^2$ *und "□" steht für "−1".*

4.2.3 Diskretisierung in zwei und drei Dimensionen

In zwei Dimensionen (kartesische Koordinaten) wird

$$\frac{\partial^2}{\partial \underline{\mathbf{r}}^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \qquad \text{und} \qquad \psi(\underline{\mathbf{r}}) = \psi(x, y). \tag{4.8}$$

Wir wählen im einfachsten Fall ein rechteckiges Grundgebiet und ein äquidistantes Gitter der Schrittweite h, wobei n_x und n_y die Zahlen der inneren Punkte entlang der x-Achse bzw. y-Achse seien (Abb. 4.2). Die Koordinaten der inneren Gitterpunkte lauten

$$\underline{\mathbf{r}}_{ij} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} i \cdot h \\ j \cdot h \end{pmatrix}, \qquad 1 \le i \le n_x, \quad 1 \le j \le n_y, \tag{4.9}$$

und wir schreiben $\psi_{ij} \equiv \psi(x_i, y_i)$ und $V_{ij} \equiv V(x_i, y_j)$. Vereinbarungsgemäß soll die Wellenfunktion auf dem Rand identisch verschwinden:

$$\psi_{ij} = 0 \qquad \text{für } i \in \{0, n_x{+}1\} \text{ oder } j \in \{0, n_y{+}1\}. \tag{4.10}$$

Wird der zweidimensionale Laplaceoperator auf dem Gitter durch

$$\frac{1}{h^2}[-4\psi_{ij} + \psi_{i-1,j} + \psi_{i+1,j} + \psi_{i,j-1} + \psi_{i,j+1}] \tag{4.11}$$

approximiert, anschaulicher auch als „Fünfpunktstern"

$$\frac{1}{h^2}\begin{bmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{bmatrix} \tag{4.12}$$

bekannt, resultiert für die stationäre SGL das Differenzenschema

$$\frac{1}{h^2}\Big[(4 + V_{ij}h^2)\psi_{ij} - \psi_{i-1,j} - \psi_{i+1,j} - \psi_{i,j-1} - \psi_{i,j+1}\Big] = E\psi_{ij} \tag{4.13}$$

mit $1 \leq i \leq n_x$ und $1 \leq j \leq n_y$. (4.13) stellt ein System von

$$n = n_x \cdot n_y \tag{4.14}$$

schwach gekoppelten Gleichungen für die n unbekannten inneren Funktionswerte ψ_{ij} und den Eigenwert E dar.

Um (4.13) in der herkömmlichen Matrixformulierung

$$\mathbf{A}\underline{\mathbf{u}} = E\underline{\mathbf{u}} \tag{4.15}$$

mit einer $n \times n$-Matrix $\mathbf{A}$ und einem n-dimensionalen Eigenvektor $\underline{\mathbf{u}}$ aufzuschreiben, ist man genötigt, die zweifach indizierten ψ_{ij} durch einen einfach indizierten Vektor $\underline{\mathbf{u}}$ darzustellen. Das bedeutet, daß die (inneren) Punkte in irgendeiner Weise durchnumeriert werden müssen.

Bei einer fortlaufenden (einer lexikographischen) Numerierung der Gitterpunkte entsprechend Abb. 4.2 gilt die Zuordnung

$$u_\ell = \psi_{ij} \quad \text{mit} \quad \ell = i + (j-1)\, n_x, \tag{4.16}$$

und wir würden bspw. für das Grundgebiet aus Abb. 4.2 mit 4×3 inneren Punkten die dort gezeigte Eigenwertmatrix erhalten.

Im Grunde sind hier beliebige Numerierungen denkbar, die resultierenden Matrizen erweisen sich aus numerischer Sicht jedoch keinesfalls als äquivalent, sondern unterscheiden sich topologisch in der Anordnung der Nichtdiagonalelemente und besitzen im allg. auch unterschiedliche Konditionen und Konvergenzeigenschaften. Wir wollen künftig so verfahren, daß wir mit ψ_{ij} den durch den Gitterpunkt $\underline{\mathbf{r}}_{ij}$ eindeutig lokalisierten Wert der Wellenfunktion bezeichnen, während $\underline{\mathbf{u}} = \{u_\ell\}$ für jenen eindimensionalen Vektor stehen soll, der aus der (in gewisser Weise willkürlichen) Zuordnung der Indizes $\{i, j\} \rightarrow \{\ell\}$ hervorgeht, der aber letztlich Eingang in den numerischen Prozeß findet. Wie in Abschnitt 4.4 auszuführen bleibt, benutzten wir aus Gründen eines schnellen Matrix-mal-Vektor-Algorithmus im weiteren stets die lexikographische Zuordnung.

In ganz analoger Weise gilt für ein quaderförmiges Grundgebiet in drei Dimensionen für den inneren Gittervektor

$$\underline{\mathbf{r}}_{ijk} = \begin{pmatrix} x_i \\ y_i \\ z_k \end{pmatrix} = \begin{pmatrix} i \cdot h \\ j \cdot h \\ k \cdot h \end{pmatrix}; \qquad \begin{array}{l} 1 \leq i \leq n_x; \\ 1 \leq j \leq n_y; \\ 1 \leq k \leq n_z; \end{array} \qquad n = n_x n_y n_z. \tag{4.17}$$

und der Laplaceoperator $\boldsymbol{\nabla}^2\psi(x, y, z)$ wird mit $\psi_{ijk} \equiv \psi(x_i, y_i, z_k)$ durch

$$\frac{1}{h^2}\left[-6\psi_{ijk} + \psi_{i-1,j,k} + \psi_{i+1,j,k} + \psi_{i,j-1,k} + \psi_{i,j+1,k} + \psi_{i,j,k-1} + \psi_{i,j,k+1}\right] \tag{4.18}$$

approximiert. Definiert man den einfach indizierten n-dimensionalen Vektor $\underline{\mathbf{u}}$ lexikographisch gemäß

$$u_\ell = \psi_{ijk} \quad \text{mit} \quad \ell = i + (j-1)\, n_x + (k-1)\, n_x n_y, \tag{4.19}$$

würde bspw. für $n_x = 4$, $n_y = 3$ und $n_z = 2$ die in Abb. 4.3 skizzierte Eigenwertmatrix entstehen.

Insgesamt ist für die Diskretisierungsmatrix $\mathbf{A}$ folgendes charakteristisch:

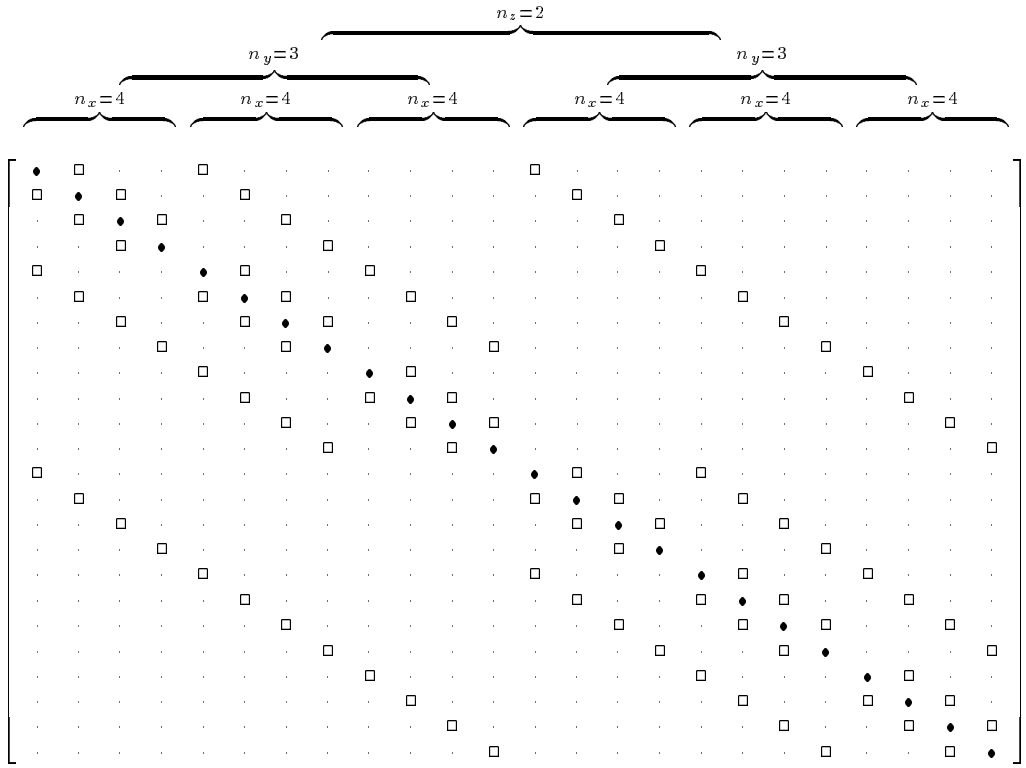


Abbildung 4.3: *Beispiel einer Eigenwertmatrix für ein dreidimensionales quaderförmiges Grundgebiet mit* $4 \times 3 \times 2$ *lexikographisch numerierten inneren Punkten. Die Hauptdiagonalememente "•" lauten* $a_{ii} = 6 + V_i h^2$ *und "□" steht für "*-1*".*

- **A** ist äußerst schwach besetzt als unmittelbare Folge der Diskretisierung einer (partiellen) DGL zweiter Ordnung.
- Für $h_{x|y|z} = \text{const}$ wird **A** symmetrisch.
- Für $h_x = h_y = h_z = h$ kann h darüberhinaus aus **A** herausgezogen werden, so daß Nichtdiagonalelemente nur die Werte 0 oder -1 annehmen. Den Matrix-mal-Vektor-Algorithmen bei der nachfolgenden Eigenwertbestimmung kommt das sehr zu gute.
- Das Potential geht nur in die Hauptdiagonale ein. Ohne Änderungen können somit auch beliebig „weiche" Potentialmulden beschrieben werden, sofern den gebundenen Zuständen in den Grenzen des Grundgebietes nur ein weitgehend ungestörtes Abklingen ermöglicht wird. Darüberhinaus lassen sich in ein Grundgebiet auch mehrere Potentialmulden eingeprägen.
- Die Form des Einbettungsgebietes und die Art der Numerierung bestimmen die Topologie von **A**, d. h. die Anordnung der Nichtdiagonalelemente. Bei rechteckiger und kubischer Einbettung, wie wir sie verwenden, und lexikographischer Numerierung entstehen die besonders einfachen *Linienmatrizen*.

Abschließend zur Diskretisierung der Geometrie, d. h. zur Approximation der stetigen Kastengeometrie Ω_K durch eine solche des Gitters, Ω_K^*, die bei uns über eine Diskretisierung des Potentials vonstatten geht und an jedem Gitterpunkt $\underline{\mathbf{r}}_{ij}$ auf die binäre Entscheidung

$$\textbf{if } \underline{\mathbf{r}}_{ij} \in \Omega_K \textbf{ then } V_{ij} = 0 \textbf{ else } V_{ij} = V_0 \tag{4.20}$$

zurückgeführt wurde. (Prinzipiell bestände durch „Verschmierung" des Potentials auch die Möglichkeit eines „Ditherings".)

Für ein Vorgehen wie (4.20) sind zwei Fragestellungen denkbar: (i) Wie ist Ω_K auf einem gegebenen Gitter der Schrittweite h zu positionieren, damit Ω_K^* möglichst gut Ω_K approximiert? (ii) Was ist die „eigentliche" stetige Entsprechung eines gegebenen diskretisierten Potentials V_{ij} (einer Geometrie Ω_K^*)?

Jede fundierte Antwort sähe sich der keineswegs trivialen Aufgabe gegenüber, ein geeignetes Maß für den Fehler $|\Omega_K^* - \Omega_K|$ zu finden[2]. Wir sind pragmatischer vorgegangen: Im Sinne der obigen Frage (i) wurden die Geometrien auf dem Gitter lediglich verrückt, um eventuelle Symmetrien von Ω_K in Ω_K^* zu wahren, um ansonsten nach (4.20) zu verfahren. Welcher stetigen Geometrie mit dem derart gewonnenen Ω_K^* dann in irgendeinem strengen Sinne entsprochen wurde, kann bei unserer STM-Fragestellung als unerheblich angesehen werden.

4.3 Diskrektisierungsfehler

4.3.1 Allgemeine Formulierung

Innerhalb unseres numerischen Verfahrens existieren zwei Fehlerquellen:

[2] Nichttrivial vor allem deshalb, weil das eigentliche Kriterium die resultierenden Wellenfunktionen sein sollten. Der Geometrie eines Quantendrahtes kann z. B. exakt entsprochen sein, ihrer anderen Dispersionsrelation wegen können Gitterlösungen jedoch an einer bestimmten Energie einen evanescenten Verlauf hervorbringen, wo die kontinuierliche Lösung oszillieren würde. Die entsprechende Fragestellung vorausgesetzt, wäre der Fehler mit der exakten Geometrie quasi unendlich, wohingegen ein etwas breiterer Draht (oder eine etwas höhere Energie) das qualitativ richtige Ergebnis geliefert hätte.

1. Die Differenzenapproximation der SGL auf einem diskreten Gitter verursacht die hier so bezeichneten Diskretisierungsfehler, die systematischer Natur sind.

2. Durch das Rechnen mit endlicher Genauigkeit und endlichen Abbruchschranken bei der anschließenden Lösung des aus der Diskretisierung hervorgegangenen Matrixeigenwertproblems entstehen Fehler rein numerischer Herkunft, die prinzipiell zwar vermeidbar wären, mit realistischem Aufwand bei derart großen Problemen in der Praxis aber nicht zu unterbinden sind und, wie sich zeigen wird, den Vertrauensbereich tatsächlich spürbar einschränken.

Fehler der ersten Art werden hier behandelt, solche der zweiten kommen in 4.4.2 und B.6.2 zur Sprache.

Ein wesentlicher Grund für die folgende, für eine gewöhnliche numerische Anwendung ungewöhnlich ausführliche Analyse der Diskretisierungsfehler war die schwierige Sachlage bei der numerischen Lösung des großen Eigenwertproblems (vgl. 4.4), infolgedessen das Diskretisierungsgitter über Gebühr weder fein sein sollte noch konnte und eine genauere Kenntnis des Einflusses der Schrittweite auf die verschiedenen Aspekte der Gitterlösungen (Eigenwerte, Verhalten im Innenraum, Verhalten im Außenraum) erstrebenswert erschien.

Stärker anwendungsorientiert geschriebene Textbücher zum Differenzenverfahren (z. B. [79, 61]) behandeln Diskretisierungsfehler in der Regel für normale DGL, es fehlen dagegen die diesbezüglichen Aussagen zu Eigenwertproblemen, die in unserem Fall gerade erforderlich wären. Mathematisch tiefergehende Darstellungen wie [33] berücksichtigen zwar auch diese (ebenda, Kap. 11.3), zur Gewinnung schärferer Aussagen bei einer konkreten Aufgabe sind die Theoreme aber sehr allgemein und eher ungeeignet. Wir wählen daher im folgenden einen Zugang, der die Diskretisierungsfehler, aufgeschlüsselt nach Eigenwert- und Zustandsfehlern, strikt aus der Sicht der exakten (kontinuierlichen) Lösungen beschreibt, d. h. wir setzen diese als bekannt und suchen die Gitterlösungen in deren Abhängigkeit zu bestimmen. Die Besonderheiten der Problemstellung (Quantenbillards) sind dadurch gut einzubringen.

Zur Notation: Aus Gründen der Lesbarkeit beschränken wir uns in den allgemeinen Formulierungen auf zwei Dimensionen. Neben exakten Kontinuums- und diskreten Gitterzuständen, $\psi(x, y)$ und ψ_{ij}, sind die zugehörigen Eigenwerte E und E_h zu unterscheiden, was der untere, die Schrittweite anzeigende Index leisten soll. Ein äquidistantes Gitter der Schrittweite h nennen wir h-Gitter. Gegebenenfalls ist eine auf dem Gitter des Grundgebietes Ω_G definierte Punktmenge wie $\{\psi_{ij}\}$ als eindimensionaler Vektor $\{u_\ell\}$ aufzufassen, was im Konkreten immer eine bestimmte Abbildungsvorschrift $\{ij\} \to \{\ell\}$ der Indexmengen verlangt. Unten ist die genaue Form dieser Vorschrift jedoch irrelevant, weshalb wir statt $\{u_\ell\}$ die abstraktere Ket-Schreibweise $|\psi_{ij}\rangle$ verwenden, wenn insbesondere die eindimensionale Sicht gemeint ist. Dabei gelte: $\langle \psi_{ij} \,|\, f_{ij} \rangle = \sum_{ij} \psi_{ij} f_{ij}$. Die zum Grundgebiet Ω_G gehörige Indexmenge $\{ij\}$ der Gitterpunkte sei I_G.

Die exakten Eigenfunktionen $\psi(x, y)$ des Kontinuums und die Gittereigenvektoren $|\psi_{ij}\rangle$ des diskretisierten Problems sind direkt zunächst gar nicht miteinander vergleichbar, da letztere lediglich eine diskrete Punktmenge verkörpern. Als natürliches Fehlermaß bietet sich immerhin an, die Abweichungen an den Gitterpunkten (x_i, y_j) zu nehmen, d. h. $\psi(x_i, y_i)$ und ψ_{ij} gegenüberzustellen. Allerdings sind nun Eigenzustände immer nur bis auf einen willkürlichen Faktor, über den gewöhnlich im Sinne einer Normierung verfügt wird, eindeutig bestimmt, so daß die Frage nach einem punktualen „Fehler“ $\psi(x_i, y_j) - \psi_{ij}$ erst dann

überhaupt einen Sinn bekommt, wenn die Antwort unabhängig wird von diesen willkürlichen Faktoren, d. h. das Verhältnis der Längen von $|\psi_{ij}\rangle$ und $|\psi(x_i, y_j)\rangle$ eine Festsetzung hierin mit erfährt. Letztlich sind also die Vektoren $|\psi_{ij}\rangle$ und $|\psi(x_i, y_j)\rangle$ als Ganzes in Relation zu bringen.

Wir stellen unser Vorgehen zunächst schematisch dar: Es sei

$$\hat{\mathbf{D}}_{ij} \, |\psi_{ij}\rangle = E_h \, |\psi_{ij}\rangle \tag{4.21}$$

das Matrixeigenwertproblem des Differenzenschemas mit $\hat{\mathbf{D}}_{ij}$ als Diskretisierungsmatrix, dem die Gittervektoren $|\psi_{ij}\rangle$ zum Eigenwert E_h genügen. Wie bemerkt, sehen wir $|\psi_{ij}\rangle$ und E_h als unbekannt an und setzen nur voraus, daß eine solche Eigenwertidentität mit irgendeinem E_h existiert. Demgegenüber gehorchen die exakten $\psi(x, y)$ zum Eigenwert E im gesamten Grundgebiet Ω_G der SGL, symbolisch $\hat{\mathbf{D}}\psi = E\psi$, was im Besonderen dann natürlich auch an allen Gitterpunkten gelten muß: $\hat{\mathbf{D}}\psi(x_i, y_j) = E\psi(x_i, y_j) \; \forall (ij) \in I_G$. Das sind N lokal erfüllte Differentialidentitäten an den N Gitterorten, die man in möglichster Analogie zu (4.21) auch „vektoriell“ schreiben kann[3]:

$$\hat{\mathbf{D}} \, |\psi(x_i, y_j)\rangle = E \, |\psi(x_i, y_j)\rangle \, . \tag{4.22}$$

Um den gesuchten Zusammenhang zwischen $\{\psi(x, y), E\}$ und $\{|\psi_{ij}\rangle \, , E_h\}$ herzustellen, ersetzen wir in der Eigenwertidentität des Differenzenschemas (4.21) die ψ_{ij} durch die exakten Amplitudenwerte $\psi(x_i, y_i)$ d. h. wir tauschen in (4.21) $|\psi_{ij}\rangle$ gegen einen Vektor $|\psi(x_i, y_j)\rangle$. Diese Substitution muß im allgemeinen die Identität zerstören, da eben der Vektor $|\psi_{ij}\rangle$ dieses Matrixeigenwertproblem löst und nicht $|\psi(x_i, y_j)\rangle$. Um trotzdem eine Identität zu bewahren (um weiter ein Gleichheitszeichen schreiben zu können), fügen wir simultan mit dieser Ersetzung an jedem Gitterort ein zunächst unbekanntes Kompensationsglied C_{ij} hinzu:

$$\hat{\mathbf{D}}_{ij} \, |\psi(x_i, y_j)\rangle + |C_{ij}\rangle = E_h \, |\psi(x_i, y_j)\rangle \, . \tag{4.23}$$

Da neben $|C_{ij}\rangle$ auch E_h als gesucht, also als variabel betrachtet wird, ist die Identitätsforderung allein noch nicht hinreichend für eine eindeutige Bestimmung beider Größen. ($|C_{ij}\rangle$ könnte auf Kosten von E_h beliebig große Anteile in Richtung $|\psi(x_i, y_j)\rangle$ enthalten.) Eindeutigkeit wird aber erreicht durch die Nebenbedingung

$$|C_{ij}\rangle - |\psi(x_i, y_j)\rangle \quad \Leftrightarrow \quad \| \, |C_{ij}\rangle \, \| \to \min \, . \tag{4.24}$$

Beide Formulierungen sind hier äquivalent. Implizit erfährt hierüber dann auch das angesprochene, a priori willkürliche Verhältnis der Normen von $|\psi_{ij}\rangle$ und $|\psi(x_i, y_j)\rangle$ seine Festsetzung.

In (4.23) überführen wir nunmehr den Differenzenausdruck $\hat{\mathbf{D}}_{ij}\psi$ durch Entwicklung von $\psi(x_i \pm h, y_j)$, $\psi(x_i, y_j \pm h)$ usf. in eine vollständige Taylorreihe um den Gitterort (x_i, y_j) – $\hat{\mathbf{D}}_{ij}\psi = \hat{\mathbf{D}}\psi|_{ij} + F_{ij}$ – in den exakten Differentialausdruck $\hat{\mathbf{D}}\psi$ an der Stelle (x_i, y_j) plus einen Restterm F_{ij}, der die höheren Ableitungen von ψ enthält:

$$\hat{\mathbf{D}} \, |\psi(x_i, y_j)\rangle + |F_{ij}\rangle + |C_{ij}\rangle = E_h \, |\psi(x_i, y_j)\rangle \, . \tag{4.25}$$

Der Vektor $|F_{ij}\rangle$ wird orthogonal zerlegt in einen Vektor in die Eigenwertrichtung $|\psi(x_i, y_j)\rangle$ (mit dem Proportionalitätsfaktor ΔE_h) und den hierzu senkrecht stehenden $|f_{ij}\rangle$, d. h. der

[3] Wir verwenden hier die abkürzenden Schreibweisen $\hat{\mathbf{D}}\psi(x_i, y_j)$ statt $\hat{\mathbf{D}}\psi(x, y)|_{x_i, y_j}$ bzw. später $\nabla^2\psi(x_i, y_j)$ statt $\nabla^2\psi(x, y)|_{x_i, y_j}$; $\hat{\mathbf{D}} \, |\psi\rangle$ wäre in Strenge zu lesen: $|(\hat{\mathbf{D}}\psi)_{ij}\rangle$.

Eigenwertanteil in $|F_{ij}\rangle$ wird abgespalten:

$$|F_{ij}\rangle = |f_{ij}\rangle + \Delta E_h\,|\psi(x_i,y_j)\rangle \quad \text{mit} \quad \langle f_{ij}\,|\,\psi(x_i,y_j)\rangle \stackrel{!}{=} 0. \tag{4.26}$$

F_{ij}, f_{ij} und ΔE_h sind hier vollständig ausgedrückt durch die exakten ψ, und d. h. hier als bekannt anzusehen. Aus (4.25) entsteht:

$$\hat{\mathbf{D}}\,|\psi(x_i,y_j)\rangle + |f_{ij}\rangle + |C_{ij}\rangle = (E_h - \Delta E_h)\,|\psi(x_i,y_j)\rangle\,. \tag{4.27}$$

Der Vergleich mit (4.22) zeigt nun folgendes: Gleichung (4.27) geht gerade dann in (4.22) über, erfüllt sich also in ihrer (a priori mit Hilfe der C_{ij} vorausgesetzten) Identität, wenn $E = E_h - \Delta E_h$ und $|C_{ij}\rangle = -\,|f_{ij}\rangle$. Wegen $|f_{ij}\rangle - |\psi(x_i,y_j)\rangle$ ist damit zugleich die zum Zwecke der Eindeutigkeit erhobenene Nebenbedingung (4.24) erfüllt. Wir können nunmehr eine eindeutige Aussage über die Eigenwert- und die lokalen Zustandsfehler in folgender Weise machen: Bei der Ersetzung $\psi_{ij} \rightarrow \psi(x_i,y_i)$ im Differenzenschema, also $\hat{\mathbf{D}}_{ij}\psi_{ij} \rightarrow \hat{\mathbf{D}}_{ij}\psi(x_i,y_j)$ auf der linken Seite und $E_h\psi_{ij} \rightarrow E_h\psi(x_i,y_j)$ auf der rechten, bleibt eine Identität, nämlich jetzt (4.22), genau dann gewahrt, wenn der unbekannte Gittereigenwert $E_h = E + \Delta E_h$ betrug, und auf der linken Seite die lokalen Kompensationsglieder $C_{ij} = -f_{ij}$ gemäß

$$\hat{\mathbf{D}}_{ij}\psi_{ij} = \hat{\mathbf{D}}_{ij}\psi(x_i,y_j) - f_{ij}\,. \tag{4.28}$$

im ursprünglichen Differenzenschema der ψ_{ij}, verglichen gegen eines mit den $\psi(x_i,y_j)$, absorbiert („verborgen“) gewesen waren. Letztere Aussage über die lokalen Fehler der Gitterlösungen $|\psi_{ij}\rangle$ werden wir dann am konkreten Fall näher diskutieren. Es läßt sich jedoch schon festhalten, daß f_{ij} hier keinen punktualen Fehler, $\psi_{ij} - \psi(x_i,y_j)$, angibt, sondern den Unterschied zwischen zwei Differenzenoperationen am Punkt $\underline{\mathbf{r}}_{ij}$, ausgeführt einmal an $|\psi_{ij}\rangle$ und einmal an $|\psi(x_i,y_j)\rangle$.

Um nunmehr in den konkreten Fall einzutreten, schreiben wir uns zunächst das Matrixeigenwertproblem (4.13) des Differenzenschemas, dem die Gittervektoren $|\psi_{ij}\rangle$ zum Eigenwert E_h genügen, noch einmal kompakter in der Form

$$[-\boldsymbol{\nabla}^2_{ij} + V_{ij}]\,\psi_{ij} = E_h\psi_{ij} \qquad \forall(ij) \in I_G; \tag{4.29}$$

$\boldsymbol{\nabla}^2_{ij}$ symbolisiert jetzt den Fünfpunktstern und $V_{ij} \equiv V(x_i,y_j)$. Die exakten $\psi(x,y)$ zum Eigenwert E befriedigen andererseits im gesamten Grundgebiet Ω_G die SGL (4.1), also insbesondere auch an allen Gitterpunkten (x_i,y_j),

$$[-\boldsymbol{\nabla}^2 + V_{ij}]\,\psi(x_i,y_j) = E\psi(x_i,y_j) \qquad \forall(ij) \in I_G, \tag{4.30}$$

womit das Analogon zu (4.22) gegeben wäre. Nochmals anzumerken, daß im Gegensatz zu (4.29) dies kein echtes (gekoppeltes) System von Gleichungen ist, sondern es sind N an jedem Gitterpunkt lokal („für sich“) bestehende differentielle Identitäten. Der folgende Schritt zu (4.23), d. h. die Ersetzung der ψ_{ij} im Differenzenschema durch die exakten Amplitudenwerte $\psi(x_i,y_j)$ unter simultaner Hinzufügung eines die Identität wahrenden Kompensationsgliedes C_{ij}, gibt dann:

$$[-\boldsymbol{\nabla}^2_{ij} + V_{ij}]\,\psi(x_i,y_j) + C_{ij} \;=\; E_h\psi(x_i,y_j) \qquad \forall(ij) \in I_G. \tag{4.31}$$

Entwicklung von $\psi(x_i \pm h, y_i \pm h)$ in eine vollständige Taylorreihe um $\psi(x_i,y_i)$ überführt den Fünfpunktstern in den Laplace-Operator plus den Restterm F_{ij}:

$$[-\boldsymbol{\nabla}^2 + V_{ij}]\,\psi(x_i,y_j) + F_{ij} + C_{ij} \;=\; E_h\psi(x_i,y_j) \qquad \forall(ij) \in I_G, \tag{4.32}$$

$$F_{ij} = -\frac{2}{h^2}\sum_{n=2}^{\infty}\frac{h^{2n}}{(2n)!}\left(\frac{\partial^{2n}}{\partial x^{2n}}+\frac{\partial^{2n}}{\partial y^{2n}}\right)\psi(x_i,y_j)\,. \tag{4.33}$$

Für $h \to 0$ strebt $F_{ij} \to 0$ und es konvergiert das Differenzenschema wie zu fordern gegen das exakte Problem, wobei der in h führende Fehlerterm mit h^2 skaliert. Allgemein resultiert F_{ij} hier ja daraus, daß die exakte Lösung der Differenzenoperation unterworfen und anschließend in den exakten Differentialausdruck plus diesen Rest transformiert wurde:

$$\hat{\mathbf{D}}_{ij}\psi(x_i,y_j) = \hat{\mathbf{D}}\psi(x_i,y_j) + F_{ij}\,. \tag{4.34}$$

Konkret also:

$$-\boldsymbol{\nabla}^2_{ij}\psi(x_i,y_j) = -\boldsymbol{\nabla}^2\psi(x_i,y_j) + F_{ij}\,. \tag{4.35}$$

($V_{ij}\,\psi(x_i,y_j)$ ist auf beiden Seiten identisch und fällt heraus.) F_{ij} verkörpert damit konkret also gerade den Unterschied zwischen differentieller und Differenzenkrümmung, und zwar unserem Vorgehen gemäß ausgedrückt durch Ableitungen eben der exakten $\psi(x,y)$ am Gitterpunkt $\underline{\mathbf{r}}_{ij}$.

An dieser Stelle eine Zwischenbemerkung, um den Zusammenhang zwischen ψ_{ij} und $\psi(x,y)$, wie er hier durch C_{ij} bzw. F_{ij} vermittelt wird, noch etwas auszuleuchten: Man betrachte die ähnlich angelegte, doch simplere Poissongleichung, $-\boldsymbol{\nabla}^2 u = \varrho$, die diskretisiert anstelle des Eigenwertproblems (4.29) nur ein gewöhnliches Gleichungssystem nach sich zöge und anstelle von (4.32) dann nur:

$$-\boldsymbol{\nabla}^2 u(x_i,y_j) + F_{ij} + C_{ij} \;=\; \varrho(x_i,y_j) \qquad \forall (ij) \in I_G\,.$$

Der Vergleich mit $-\boldsymbol{\nabla}^2 u = \varrho$ zeigt jetzt sofort $C_{ij} = -F_{ij}$; es gibt hier kein Zerlegungs- oder Eindeutigkeitsproblem, da die rechte Seite fix und bekannt ist. Anschaulicher noch, wenn man bei der Ersetzung $u_{ij} \to u(x_i,y_j)$ zunächst auf die Bewahrung einer Identität vermittels des Kompensationsgliedes C_{ij} verzichtet, und statt eines nun nicht mehr statthaften "=" einen vorläufigen Platzhalter, etwa ein "$\leftrightarrow$" schreibt:

$$-\boldsymbol{\nabla}^2 u(x_i,y_j) + F_{ij} \;\leftrightarrow\; \varrho(x_i,y_j) \qquad \forall (ij) \in I_G\,.$$

Man sieht: eine Identiät nach dem Übergang $u_{ij} \to u(x_i,y_j)$ bleibt gerade dann gewahrt, wenn F_{ij} im ursprünglichen Differenzenschema $-\nabla^2_{ij}u_{ij} = \varrho_{ij}$ „absorbiert" gewesen war, und beim Übergang $u_{ij} \to u(x_i,y_j)$ gewissermaßen „desorbiert" wird. F_{ij} kann dabei im ursprünglichen Differenzenschema enthalten gedacht werden sowohl in u_{ij} als auch in ϱ_{ij}. Das gibt dann zwei prinzipiell mögliche Lesarten eines Fehlers F_{ij}, je nachdem, ob dieser auf der linken Seite u_{ij} oder auf der rechten ϱ_{ij} angerechnet wird:

I. Die Gitterlösung u_{ij} wird als fehlerbehaftet angesehen, und zwar, weil beim Einsetzen der exakten Lösung in das Differenzenschema nur dann eine Identität gewahrt bleibt, wenn simultan auf der linken Seite F_{ij} subtrahiert (C_{ij} addiert) wird, gemäß

$$-\boldsymbol{\nabla}^2_{ij}u_{ij} = -\boldsymbol{\nabla}^2_{ij}u(x_i,y_j) - F_{ij}\,.$$

Die als Differenzenquotient genommene (positive) Krümmung von u_{ij} fällt also um F_{ij} höher aus als die *entsprechend genommene* von $u(x_i,y_j)$. Man beachte, daß kein wirklicher Vergleich mit der exakten *differentiellen* Krümmung $\boldsymbol{\nabla}^2 u(x_i,y_j)$ möglich ist.

II. Die Gitterlösung u_{ij} approximiert an allen $\underline{\mathbf{r}}_{ij}$ punktweise exakt ein gestörtes Kontinuumsproblem $-\boldsymbol{\nabla}^2\tilde{u} = \tilde{\varrho}$ mit der modifizierten rechten Seite $\tilde{\varrho}(x_i, y_j) = \varrho(x_i, y_j) - F_{ij}$. Man beachte, daß $\tilde{\varrho}$ nur an den Gitterpunkten eindeutig bestimmt ist.

Ein gewisses Pendant zu (II) wäre bei der SGL ein um f_{ij} modifizierter Potentialterm $\tilde{V}(\underline{\mathbf{r}}_{ij})\psi_{ij}$. Doch soll uns im weiteren nur Lesart (I) beschäftigen.

Zurück zu unserem Eigenwertproblem, wo nun $|F_{ij}\rangle$ gemäß (4.26) zu zerlegen ist. Zum Behelf der Übersichtlichkeit kürzen wir ab:

$$\Psi_{ij} \equiv \psi(x_i, y_j) \tag{4.36}$$

Die Zerlegung (4.26) liest sich damit:

$$|F_{ij}\rangle = |f_{ij}\rangle + \Delta E_h\,|\Psi_{ij}\rangle \quad \text{mit} \quad \langle f_{ij}\,|\,\Psi_{ij}\rangle \stackrel{!}{=} 0. \tag{4.37}$$

Erst die Nebenbedingung $\langle f_{ij}\,|\,\Psi_{ij}\rangle = 0$ macht diese Zerlegung eindeutig und ermöglicht die Bestimmung von ΔE_h:

$$\Delta E_h = \frac{\langle \Psi_{ij}\,|\,F_{ij}\rangle}{\langle \Psi_{ij}\,|\,\Psi_{ij}\rangle}. \tag{4.38}$$

Selbiges Resultat wäre im übrigen auch entstanden, wenn statt $\langle f_{ij}\,|\,\Psi_{ij}\rangle = 0$ die Extremalforderung $\|\,|f_{ij}\rangle\,\| \to \min$ zur Nebenbedingung erhoben worden wäre, mit anderen Worten, die Zerlegung (4.37) ist zugleich diejenige mit der kleinsten euklidischen Norm für $|f_{ij}\rangle$ (das Äquivalent zu $\|\,|C_{ij}\rangle\,\| \to \min$).

Mit (4.37) gewinnt (4.32) die Gestalt

$$[-\boldsymbol{\nabla}^2 + V_{ij}]\,\psi(x_i, y_j) + f_{ij} + C_{ij} \;=\; (E_h - \Delta E_h)\,\psi(x_i, y_j) \quad \forall (ij) \in I_G\,. \tag{4.39}$$

Und das kommt genau dann der Identität (4.30) gleich – vgl. die Argumentation zu (4.27) –, wenn $E_h - \Delta E_h = E$ gesetzt und f_{ij} dem Differenzenschema auf der linken Seite im Sinne von (4.28) angelastet wird, was also schließlich zu folgender Aussage führt: Die Differenzenapproximation liefert einen um ΔE_h abweichenden Gittereigenwert

$$E_h = E + \Delta E_h\,, \tag{4.40}$$

mit ΔE_h gemäß (4.38), und einen solcherart gestörten Gitterzustand ψ_{ij}, daß für diesen die Differenzenoperation $[-\boldsymbol{\nabla}^2_{ij} + V_{ij}]$ am Punkt $\underline{\mathbf{r}}_{ij}$ einen um f_{ij} niedrigeren Wert zur Folge hat als die gleichartige Operation an $\psi(x_i, y_j)$:

$$[-\boldsymbol{\nabla}^2_{ij} + V_{ij}]\,\psi_{ij} \;=\; [-\boldsymbol{\nabla}^2_{ij} + V_{ij}]\,\psi(x_i, y_j) - f_{ij}\,. \tag{4.41}$$

Es wird auch offenbar, daß nur die orthogonale Zerlegung (4.37) in Gleichung (4.39) zu einem vollständigen „Ausfällen“ der Identität (4.30) und damit zu eindeutigen *Störungen* f_{ij} und ΔE_h führt, was eine Korrespondenz zwischen ψ_{ij} und $\psi(x, y)$ herzustellen erst gestattet. Mit anderen Worten, die Eindeutigkeitsbedingung (4.24) ist zugleich die hier einzig mögliche und die obige Abbildung $\{\psi(x, y), E\} \to \{|\psi_{ij}\rangle\,, E_h\}$ liefert die tatsächlichen Gitterlösungen.

Wir betrachten einen Spezialfall von (4.39): Besitzt die kontinuierliche Lösung analog den Eigenzuständen eines unendlich hohen rechteckigen Kastens – $\psi(x, y) \propto \sin(k_\nu x)\sin(k_\mu y)$ – im gesamten Grundgebiet Ω_G die Eigenschaft

$$\left(\frac{\partial^{2n}}{\partial x^{2n}} + \frac{\partial^{2n}}{\partial y^{2n}}\right)\psi(x, y) = c_n\psi(x, y), \quad x, y \in \Omega_G, \quad n = 2, 3, \ldots \tag{4.42}$$

mit beliebigen reellen Zahlen c_n, erlaubt F_{ij} also die Darstellung

$$|F_{ij}\rangle = \Delta E_h\,|\Psi_{ij}\rangle \quad \Longleftrightarrow \quad |f_{ij}\rangle = |0\rangle\,, \tag{4.43}$$

folgt aus (4.41) wegen $|f_{ij}\rangle = |0\rangle$ nunmehr $\psi_{ij} = \psi(x_i, y_j)$, d. h. ψ_{ij} approximiert $\psi(x, y)$ an den Gitterpunkten sogar exakt und der gesamte Diskretisierungsfehler schlägt im Eigenwert E_h zu Buche. Man vergleiche hierzu das auch auf einem Gitter analytisch lösbare Beispiel des unendlich hohen Kastens aus Abschnitt A.1[4]. In allen sonstigen von (4.43) abweichenden Fällen wird ψ_{ij} dagegen immer fehlerbehaftet sein.

Die etwas unanschauliche Aussage (4.41) verschärft sich bei einem Quantenbilliard mit unendlich hohen Wänden und $V_{ij} = 0$ im Inneren zu

$$-\,\boldsymbol{\nabla}_{ij}^2\psi_{ij} \;=\; -\boldsymbol{\nabla}_{ij}^2\psi(x_i, y_j) - f_{ij}\,, \tag{4.44}$$

d. h. hier ist f_{ij} vollständig als Unterschied der Differenzenkrümmungen anzusehen. Bei einem endlich hohen, jedoch sehr großen Potentialtopf, bei dem V_{ij} im größten Teil von Ω_G (dem Kastenboden) denselben Wert besitzt und so $|V_{ij}\Psi_{ij}\rangle$ zumindest „fast“ in Richtung $|\Psi_{ij}\rangle$ zeigt, kann f_{ij} wegen $|f_{ij}\rangle - |\Psi_{ij}\rangle$ noch „fast vollständig“ der Krümmung zugeschrieben werden.

Die Gleichungen (4.33), (4.37) und (4.38) erlauben im Prinzip für alle Probleme mit analytisch bekannten Wellenfunktionen $\psi(x, y)$ eine vollständige Rekonstruktion der Gitterlösungen bzw. – dem äquivalent – eine exakte Angabe der Diskretisierungfehler. (Alle im Kontinuum analytisch lösbaren Fälle sind es also auch auf einem Gitter.) Dagegen sind bei den eigentlich interessanten nur noch numerisch zugänglichen Problemen in der Regel zwar gewisse plausible Annahmen über $\psi(x, y)$ möglich, wohl aber kaum wie in (4.33) gefordert über $(\partial_x^{2n} + \partial_y^{2n})\psi(x, y)$.

In einer Dimension lassen sich diese Differentiationen immerhin vollständig auf das Potential übertragen, da dort d^{2n}/dx^{2n} bis auf das Vorzeichen mit der n-ten Potenz des Operators der kinetischen Energie zusammenfällt und die aus der rekursiven Vorschrift

$$\begin{aligned}\boldsymbol{\nabla}^{2n}\psi(\underline{\mathbf{r}}) \;&=\; -\boldsymbol{\nabla}^{2n-2}(-\boldsymbol{\nabla}^2)\psi(\underline{\mathbf{r}}) \;=\; -\boldsymbol{\nabla}^{2n-2}\left[E - V(\underline{\mathbf{r}})\right]\psi(\underline{\mathbf{r}}) \\ &=\; -\left[E - V(\underline{\mathbf{r}})\right]\boldsymbol{\nabla}^{2n-2}\psi(\underline{\mathbf{r}}) \;+\; \psi(\underline{\mathbf{r}})\,\boldsymbol{\nabla}^{2n-2}V(\underline{\mathbf{r}})\end{aligned}$$

folgende Formel

$$\boldsymbol{\nabla}^{2n}\psi(\underline{\mathbf{r}}) = \psi(\underline{\mathbf{r}})\left\{(-1)^n\left[E - V(\underline{\mathbf{r}})\right]^n \;+\; \sum_{k=1}^{n-1}\boldsymbol{\nabla}^{2k}V(\underline{\mathbf{r}})\right\} \tag{4.45}$$

genutzt werden kann. Eine ähnlich einfache Auswertung in zwei (oder mehr) Dimensionen scheitert indes daran, daß

$$\frac{\partial^{2n}}{\partial x^{2n}} + \frac{\partial^{2n}}{\partial y^{2n}} \;=\; \boldsymbol{\nabla}^{2n} - \sum_{k=1}^{n-1}\binom{n}{k}\left(\frac{\partial^2}{\partial x^2}\right)^k\left(\frac{\partial^2}{\partial y^2}\right)^{n-k} \tag{4.46}$$

gerade keine Potenz von $\boldsymbol{\nabla}^2$ darstellt, sondern schwer zugängliche Zusatzglieder aufwirft.

[4] Im Eindimensionalen gilt hier für die Gitterlösungen (ohne Normierungsfaktor): $\psi_i = \sin(k_\nu x_i)$ mit $k_\nu = \frac{\pi}{L}\nu$, $\nu = 1, 2, \ldots, n$ (also die exakte Übereinstimmung an den Gitterpunkten: $\psi_i = \psi(x_i)$) und $E_h = \frac{2}{h^2}[1 - \cos k_\nu h]$. Wird andererseits der Fehlerterm F_i nach (4.33) mit Hilfe der exakten Eigenfunktionen berechnet, erlaubt dieser die Darstellung $F_i = \Delta E_h\,\psi(x_i)$ mit

$$\Delta E_h \;=\; -\frac{2}{h^2}\sum_{n=2}^{\infty}(-1)^n\frac{(k_\nu h)^{2n}}{(2n)!} \;=\; -\frac{2}{h^2}\left[\cos k_\nu h - 1 + \frac{(k_\nu h)^2}{2}\right],$$

was geradewegs (4.40) bestätigt: $E + \Delta E_h = k_\nu^2 + \Delta E_h = \frac{2}{h^2}[1 - \cos k_\nu h] = E_h$!

4.3.2 Fourierdarstellung

Einen für bestimmte Fragen günstigeren Zugang bietet mitunter die Fourierdarstellung

$$\psi(x,y) \;\equiv\; \mathcal{F}^{-1}\big[\phi(k_x,k_y)\big] \;\equiv\; \iint_{-\infty}^{+\infty} \mathrm{d}k_x \mathrm{d}k_y \; \phi(k_x,k_y)\; e^{ik_x x} e^{ik_y y} \tag{4.47}$$

mit der Transformierten

$$\phi(k_x,k_y) \;\equiv\; \mathcal{F}\big[\psi(x,y)\big] \;\equiv\; \frac{1}{(2\pi)^2} \iint_{\Omega_G} \mathrm{d}x\,\mathrm{d}y\; \psi(x,y)\; e^{-ik_x x} e^{-ik_y y}. \tag{4.48}$$

Man kann direkt mit (4.35) starten,

$$\begin{aligned} F_{ij} &= \boldsymbol{\nabla}^2 \psi(x_i,y_j) - \boldsymbol{\nabla}^2_{ij}\psi(x,y) && (4.49)\\ &= -(E-V_{ij})\psi(x_i,y_j) - \boldsymbol{\nabla}^2_{ij}\psi(x,y), && (4.50) \end{aligned}$$

also mit dem Differenzenschema, und die Transformationseigenschaft

$$\begin{aligned} \psi(x\pm h,y) &= \mathcal{F}^{-1}\big[e^{\pm ik_x h}\phi(k_x,k_y)\big] \\ \psi(x,y\pm h) &= \mathcal{F}^{-1}\big[e^{\pm ik_y h}\phi(k_x,k_y)\big] \end{aligned}$$

nutzen. Insgesamt entsteht

$$F_{ij} = \psi(x_i,y_j)\left\{\frac{4}{h^2} - \big[E-V_{ij}\big]\right\} + \frac{2}{h^2}\mathcal{F}^{-1}\Big[\big(\cos k_x h + \cos k_y h\big)\,\phi(k_x,k_y)\Big]. \tag{4.51}$$

Statt $\psi(x,y)$ und $(\partial_x^{2n}+\partial_y^{2n})\psi(x,y)$ sind nun Abschätzungen von $\psi(x,y)$ und $\phi(k_x,k_y)$ vonnöten. Insbesondere in der Anwendung auf mesoskopisch große Potentialmulden sollten sich über (4.51) gewisse Fortschritte erzielen lassen, da das Charakteristikum dieser Probleme – große Bereiche oszillierender Moden, schmale evanescente Ränder – eine hilfreiche Parametrisierung der Fouriertransformierten $\phi(k_x,k_y)$ verspricht. Ein Nahziel wird dabei stets das Finden der Darstellung (4.37) für F_{ij} sein, um Eigenwert- und Zustandsfehler zu separieren.

4.3.3 Vollständige Auswertung im eindimensionalen Fall

In einer Dimension kann der Fehlerterm

$$F_i = -\frac{2}{h^2}\sum_{n=2}^{\infty} \frac{\mathrm{d}^{2n}}{\mathrm{d}x^{2n}}\psi(x_i)\frac{h^{2n}}{(2n)!} \tag{4.52}$$

vollständig auf E, $\psi(x_i)$ und Ableitungen des Potentials $V(x)$ zurückgeführt werden, da $-d^2/dx^2$ mit dem Operator der kinetischen Energie zusammenfällt, was die Anwendung von (4.45) gestattet. Die Summe dort über $(E-V_i)^n$ läßt sich zu einer Kosinusreihe ergänzen,

$$\sum_{n=2}^{\infty}(-1)^n(E-V_i)^n\frac{h^{2n}}{(2n)!} \;=\; \cos\sqrt{E-V_i}h \;-\; 1 \;+\; \frac{(E-V_i)h^2}{2!},$$

so daß insgesamt entsteht:

$$F_i \;=\; \psi(x_i)\left\{\frac{2}{h^2}\Big[1-\cos\sqrt{E-V_i}h\Big] - (E-V_i) \;-\; \frac{2}{h^2}\sum_{n=2}^{\infty}\frac{h^{2n}}{(2n)!}\sum_{k=1}^{n-1}\frac{\mathrm{d}^{2k}}{\mathrm{d}x^{2k}}V(x_i)\right\}. \tag{4.53}$$

Speziell an allen Gitterpunkten mit konstantem oder linearem Potentialverlauf wird F_i durch Wegfall der Doppelsumme dann besonders einfach.

Betrachten wir insbesondere ein Kastenpotential mit $V(x) = 0$ innen und $V(x) = V_0$ außen und nutzen im Außenraum die Relation $\cos(iz) = -\cosh(z)$, folgt

$$F_i = \begin{cases} \psi(x_i)\left\{\frac{2}{h^2}\left[1-\cos\sqrt{E}h\right]-E\right\}, & \{\ldots\}<0 \quad \text{(innen)} \\ \psi(x_i)\left\{\frac{2}{h^2}\left[\cosh\sqrt{V_0-E}h-1\right]+V_0-E\right\}, & \{\ldots\}>0 \quad \text{(außen)} \end{cases}$$

und damit für (4.38)

$$\Delta E_h = \langle \Psi_i \mid \Psi_i \rangle^{-1} \left\{ \sum_i^{\text{innen}} \underbrace{(\ldots)}_{<0} \Psi_i^2 + \sum_i^{\text{außen}} \underbrace{(\ldots)}_{>0} \Psi_i^2 \right\} \tag{4.54}$$

Alle Gitterpunkte im Innenraum tragen (mit Ψ_i^2 gewichtete) negative Werte zum Eigenwertfehler bei und alle Gitterpunkte im Außenraum positive. Liegt die Gitterschrittweite fest, fällt ΔE_h daher monoton, je breiter ein Kasten wird, von einer bestimmten Kastenbreite an wird ΔE_h negativ sein und $|\Delta E_h|$ hernach stetig zunehmen, um asymptotisch in den Eigenwertfehler des unendlich hohen Kastens zu münden, der überhaupt keine Außenraumanteile besitzt. Hieraus darf allerdings nicht auf einen gleichermaßen wachsenden Fehler von ψ_i geschlossen werden, der im Gegenteil stetig geringer wird, da F_i mehr und mehr der Form (4.43) zustrebt (vgl. auch Tabelle 4.1 auf Seite 50.)

Nebenbei lassen sich auf diese Weise auch die in Anhang A mehr durch heuristische Überlegungen gewonnenen exakten Gitterresultate reproduzieren, was sich am raschesten für den unendlich hohen Kasten demonstrieren läßt. Hier wird

$$\Delta E_h = \frac{2}{h^2}\left[1-\cos\sqrt{E}h\right]-E \tag{4.55}$$

und über (4.40) folgt das bekannte Gitterresultat:

$$E_h = \frac{2}{h^2}\left[1-\cos\sqrt{E}h\right], \quad E = k^2 = (\frac{\pi}{L}\nu)^2, \quad \nu = 1,2,\ldots,n. \tag{4.56}$$

4.3.4 Zwei Ideen für eine a posteriori-Korrektur

Bisher wurden die Gitterlösungen aus der Sicht der exakten Lösungen betrachtet. In der Praxis sind diese natürlich gerade gesucht, und nur die numerischen Gitterresultate $|\psi_{ij}\rangle$ und E_h stehen zur Verfügung.

Für den eindimensionalen Fall wurde oben gezeigt, daß bei einem *endlich* hohen Potentialtopf die Genauigkeit der Gitterzustände $|\psi_i\rangle$ stetig zunimmt, je breiter der Topf wird, daß von einer bestimmten Kastenbreite an die Eigenwertwertapproximationen E_h zugleich dann aber immer schlechter werden. Es erhebt sich die Frage, ob aus der im mesoskopischen Fall sehr hohen Genauigkeit der Gitterzustände $|\psi_i\rangle$ nicht Kapital für eine Nachbesserung des Eigenwertes E_h gezogen werden kann.

Zwei Ideen für eine Nachbesserung des Eigenwertes scheinen dem Autor diskussionswürdig und werden hier eindimensional formuliert:

1. Man betrachte den Gitterzustand $|\psi_i\rangle$ als hinlängliche Näherung für $|\Psi_i\rangle$ und approximiere mit diesem den Fehlerterm F_i, was angesichts der geforderten Ableitungen

beliebiger Ordnung natürlich nur näherungsweise möglich ist:

$$F_i \;\approx\; -\frac{2}{h^2}\frac{\partial^4}{\partial x^4}\psi(x_i)\frac{h^4}{4!} \;\approx\; -\frac{h^2}{12}\frac{\boldsymbol{\nabla}^2_{i+1}\psi_{i+1} - 2\boldsymbol{\nabla}^2_i\psi_i + \boldsymbol{\nabla}^2_{i-1}\psi_{i-1}}{h^2} \equiv \tilde{F}_i.$$

Nur der in h führende Term $O(h^2)$ wird also berücksichtigt und die verbliebene 4. Ableitung dann durch einen Differenzenausdruck approximiert. Mit Hilfe von $\tilde{F}_i$ bestimme man nunmehr genähert ΔE_h,

$$\Delta E_h = \frac{\langle \boldsymbol{\Psi}_i \mid F_i \rangle}{\langle \boldsymbol{\Psi}_i \mid \boldsymbol{\Psi}_i \rangle} \;\approx\; \frac{\langle \psi_i | \tilde{F}_i \rangle}{\langle \psi_i \mid \psi_i \rangle} \equiv \Delta\tilde{E}_h,$$

um anschließend den ursprünglichen Gittereigenwert E_h wegen (4.40) gemäß $\tilde{E}_h = E_h - \Delta\tilde{E}_h$ zu korrigieren. (Anmerkung: Die Glieder $\boldsymbol{\nabla}^2_i\psi_i$ in $\tilde{F}_i$ lassen sich durch $\psi(x_i), f_i$ und $V_i\psi_i$ darstellen.)

2. Man betrachte den Gitterzustand $|\psi_i\rangle$ als hinlängliche Näherung für die Wellenfunktion, den zugehörigen Eigenwert des Differenzenschemas aber verwerfe man und berechne diesen stattdessen über eine nachträgliche Fourieranalyse von $|\psi_i\rangle$ (im speziellen Fall der homogenen Randbedingungen über eine reine Sinusanalyse):

$$|\psi_i\rangle = \sum_{n=1}^{N} c_n sin(k_n x) \quad \Longrightarrow \quad \tilde{E}_h = \sum_{n=1}^{N} c_n k_n^2.$$

Der Hintergedanke ist, daß für den unendlich hohen Potentialkasten ein Fourieransatz (im Falle homogener Randbedingungen am effektivsten ein reiner Sinusansatz) nicht nur die Auslenkungen exakt liefert (was auch schon das Differenzenschema tut), sondern zugleich auch die exakten Eigenwerte.

Der Grad der Verbesserung ließe sich anhand bekannter Lösungen leicht quantifizieren. Einer einfachen Übertragung auf mehr Dimensionen steht allerdings vorerst im Wege, daß das zugrundegelegte asymptotische Verhalten $\psi_i \rightarrow \psi(x_i)$ für $L \rightarrow \infty$ bei den tatsächlich interessanten nichtseparablen Probleme in zwei und mehr Dimensionen, also $\psi_{ij} \rightarrow \psi(x_i, y_j)$ für $L_{x|y} \rightarrow \infty$, nur für gewisse Spezialfälle (rechteckiger Kasten endlicher Höhe) naheziegen scheint. Eine genauere Analyse des Einflusses der Kastengeometrie wäre erforderlich.

4.3.5 Quantitative Resultate

Zum Erhalt quantitativer Resultate bedienen wir uns der in Anhang A gefundenen exakten eindimensionalen Gitterlösungen für die Potentialschwelle und den Potentialtopf endlicher Höhe. Neben dem Eigenwertfehler ΔE_h gilt es dabei die Fehler der Gitterzustände $|\psi_i\rangle$ zu analysieren – die Vektoren $|f_{ij}\rangle$ bzw. $|f_i\rangle$ von oben. Das Vorliegen einer Dispersionsrelation in unseren Beispielen sowohl im Kontinuierlichen als auch auf dem Gitter – $E(k)$ und $E_h(k_h)$ – ermöglicht, dies auf elegante Weise anhand von Wellenzahl k_h und Abklingkonstante κ_h – oder besser noch, weil schon die Übereinstimmung in der Dimension einen anschaulicheren Bezug zum Diskretisierungsparameter *Schrittweite* herstellt – anhand der inversen Größen Wellenlänge $\lambda_h = 2\pi/k_h$ und Abklinglänge κ_h^{-1} zu tun. Insbesondere werden wir nach den relativen Abweichungen von den Kontinuumswerten E, λ, κ^{-1} fragen:

$$\frac{\Delta E_h}{E} = \frac{E_h - E}{E}, \qquad \frac{\Delta\lambda_h}{\lambda} = \frac{\lambda_h - \lambda}{\lambda}, \qquad \frac{\Delta\kappa_h^{-1}}{\kappa^{-1}} = \frac{\kappa_h^{-1} - \kappa^{-1}}{\kappa^{-1}}. \tag{4.57}$$

Es ist anzumerken, daß auf einem Gitter i. allg. nicht die Dispersionsrelation des Kontinuums gilt, speziell für Rechteckpotentiale nicht $E = k^2$, sondern (A.15), weshalb über die Fehler von E_h und k_h getrennt buchzuführen ist.

Der Problematik der endlichen Kästen am angemessensten wäre, für einen vorgegebenen Potentialtopf die kontinuierlichen Lösungen mit denen der verschiedenen Gitter zu vergleichen. Unglücklicherweise existiert hierbei nicht nur ein einziges mögliches Vergleichsregime, da wir es mit diskreten Zuständen zu tun haben, bei denen sich mit der Schrittweite immer das komplette Zahlentripel $(E_h, \lambda_h, \kappa_h)$ ändert und lediglich die Knotenzahl (die Nummer des Zustandes) als Invariante auftritt. Worin hierbei das Problem liegt, zeige ein Beispiel: Praktisch wird bei der Eigenwertberechnung die Energie in Gestalt von ϵ vorgegeben. Wir wählen daher die Energie als Abzisse und wollen nach dem Fehler der Abklingkonstante κ_h mit wachsender Schrittweite h fragen. Die Eigenenergien der Gitterzustände variieren jedoch ebenfalls mit h (vgl. Abb. 4.6), d. h. die direkte Differenz $\kappa_h - \kappa$ für einen herausgegriffenen Eigenzustand fester Knotenzahl enthält sowohl den Effekt der Energie als auch den der Schrittweite. Um den Einfluß der Schrittweite zu extrahieren, müßte κ_h auf die Energie von κ normiert (interpoliert) werden. Je nachdem, ob man hierbei z. B. nach der kontinuierlichen Dispersionsrelation $E = \kappa^2 + V_0$ oder nach der der Gitterlösungen (A.15, A.17) verfährt, bedient man unterschiedliche Fragestellungen und erhält auch (leicht) unterschiedliche Antworten.

Die Schwierigkeiten einer Energienormierung bei λ_h und κ_h lassen sich noch umgehen, wenn zunächst nur das Problem eines freien Teilchens (auf einem unendlichen Gitter) betrachtet wird, das an einer Potentialschwelle der Höhe V_0 reflektiert wird ($0 < E < V_0$, $E_h = E$). Auf einem h-Gitter lauten hier die Relationen $k_h(E)$ und $\kappa_h(E)$ für die Wellenzahl des oszillierenden bzw. die Abklingkonstante des evanescenten Teils, die Gln. (A.31) und (A.32) aus Anhang A:

$$k_h(E) = \frac{1}{h} \arccos\left[1 - \frac{h^2}{2}E\right] \tag{4.58}$$

$$\kappa_h(E) = \frac{1}{h} \operatorname{arccosh}\left[1 + \frac{h^2}{2}(V_0 - E)\right], \tag{4.59}$$

während im Kontinuum bekanntlich $k(E) = \sqrt{E}$ und $\kappa(E) = \sqrt{V_0 - E}$ gilt.

Abb. 4.4 zeigt die relativen Fehler für λ_h und κ_h^{-1} für $V_0 = 8$ eV über der Energie. Gegenüber dem Kontinuum wird λ auf dem Gitter unter- und κ^{-1} überschätzt; für $E = 4$ eV und $h = 0.4$ Å finden wir beispielsweise λ_h um ca. 1% verringert und κ_h^{-1} um 1% erhöht.

Dieses Verhalten kann unmittelbar aus dem Differenzenschema heraus verstanden werden: In A.3 zeigen wir, daß eindimensionale Gitterlösungen für stückweise konstante Potentiale formal dieselbe Gestalt besitzen wie im Kontinuumsfall, konkret also $\sin(k_h x_i + \delta)$ im oszillierenden Mode ($E > V = 0$) und $\exp(-\kappa_h x_i)$ im evanescenten ($E < V_0, x_i > 0$), wobei x_i den Ort des i-ten Gitterpunktes bezeichnet und δ einen hier nebensächlichen Phasenfaktor, der in der folgenden Gleichung unterdrückt wird. Das Differenzenschema (4.4) liest sich für die oszillierende Lösung

$$h^{-2}\left[2 \sin k_h x_i - \sin k_h(x_i - h) - \sin k_h(x_i + h)\right] = E \sin k_h x_i.$$

Enwicklung von $\sin k_h(x_i \pm h)$ in eine Tayloreihe liefert

$$k_h^2 \; - \; k_h^4 \frac{h^2}{12} \; + \; \ldots (-1)^{j+1} k^{2j} \frac{h^{2j-2}}{(2j)!/2} \ldots \; = \; E,$$

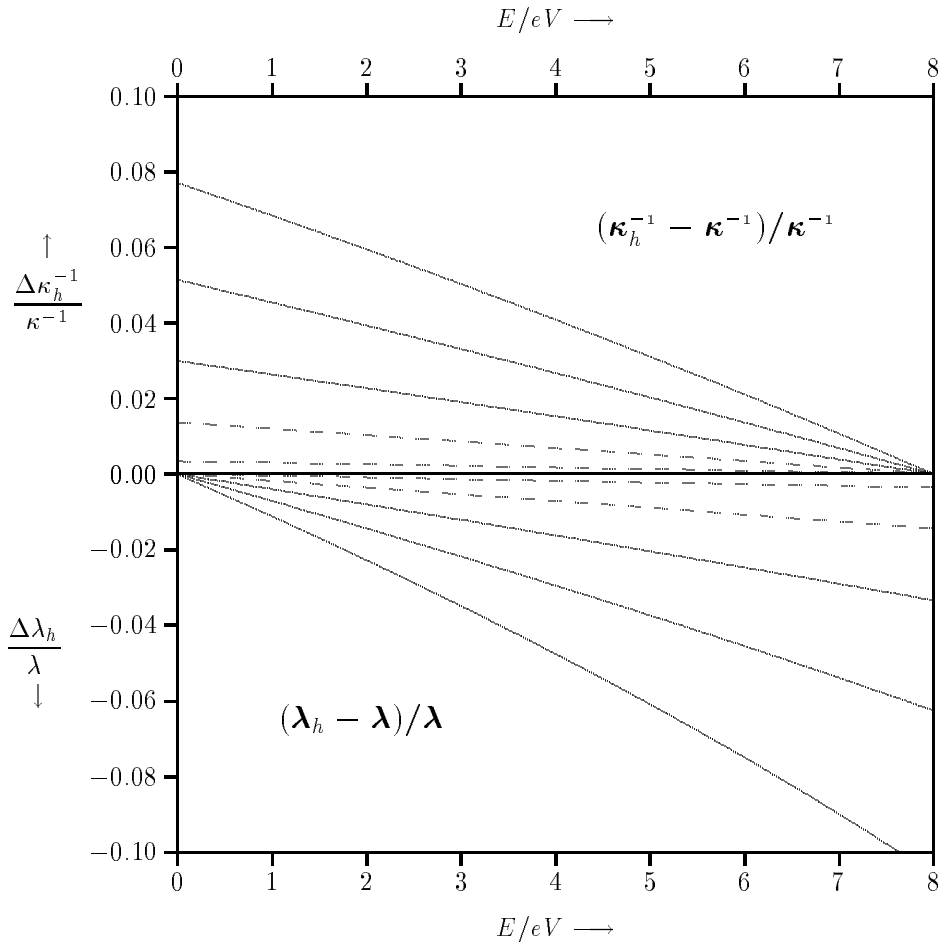


Abbildung 4.4: *Für ein freies Teilchen, das an einer Potentialschwelle der Höhe $V_0 = 8$ eV reflektiert wird ($0 < E < V_0$), sind Wellenlänge λ_h und evanescente Abklinglänge κ_h^{-1} der Gitterlösungen (Lösungen des diskretisierten Problems) mit ihren kontinuierlichen Pendants λ und κ^{-1} verglichen. Dargestellt sind die relativen Abweichungen in Abhängigkeit von der Energie. Die einzelnen Kurven gehören (mit wachsendem Abstand von der Nullinie) dabei zu den Gitterschrittweiten $h = 0.2, 0.4$ (gestrichelt), $0.6, 0.8, 1$ Å (durchgezogen).*

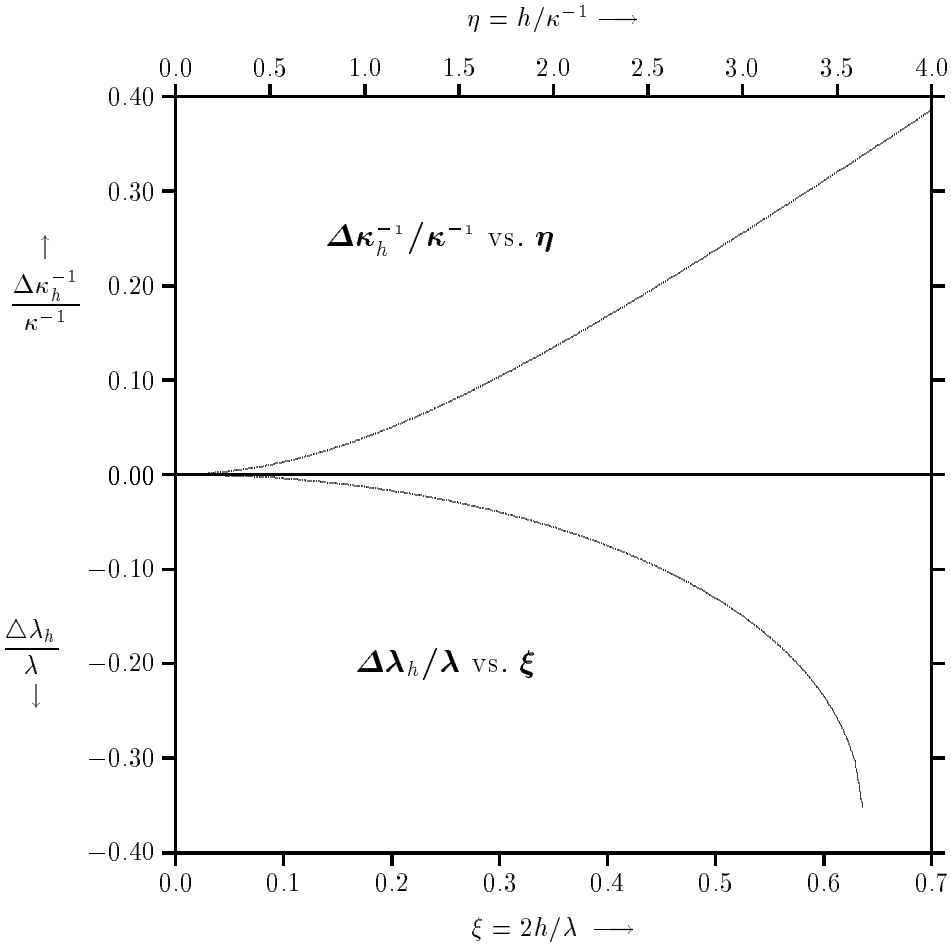


Abbildung 4.5: *Für das gleiche Problem wie in Abb. 4.4 sind die relativen Änderungen diesmal dargestellt in Abhängigkeit von den Verhältnissen Schrittweite zu exakter halber Wellenlänge, $\xi \equiv 2h/\lambda$, bzw. Schrittweite zu exakter Abklinglänge, $\eta \equiv h/\kappa^{-1}$ – die Gln. (4.60) und (4.61).*

während selbiges Vorgehen praktiziert mit $\exp(-\kappa_h x_i)$

$$\kappa_h^2 + \kappa_h^4 \frac{h^2}{12} + \ldots + \kappa_h^{2j} \frac{h^{2j-2}}{(2j)!/2} \ldots = V_0 - E$$

nach sich zieht. Der in h führende Korrekturterm ist im ersten Falle negativ – zur Kompensation (um zur gleichen Energie zu gelangen) muß k_h geringfügig gegenüber k erhöht sein –, im zweiten Falle dagegen positiv, weshalb κ_h kleiner zu sein hat als κ. Das Vorzeichen der Korrektur wäre für jede Gitterfunktion $f(x_i)$ mit $f^{(4)}f > 0$ und $f''f < 0$ bzw. $f''f > 0$ entstanden. Wir können schlußfolgern, daß zu gegebener Energie Wellenlängen oszillierender Gittermoden grundsätzlich kleiner sind als im Kontinuum, während monotone (evanescente) Moden langsamer abklingen.

Ferner resultiert die gegenläufige Tendenz in Abb. 4.4 für $\Delta\lambda_h$ und $\Delta\kappa_h^{-1}$ entlang der Energieachse simplerweise aus dem Umstand, daß λ bei kleinen Energien maximal wird, das Verhältnis h/λ somit besonders günstig ausfällt, während sich die Lage für κ^{-1} gerade entgegengesetzt darstellt.

Instruktiv scheint daher auch, die obigen Differenzen nicht nur über der Energie, sondern auch über den dimensionslosen Zahlen $\xi \equiv 2h/\lambda$ bzw. $\eta \equiv h/\kappa^{-1}$ aufzutragen, die Schrittweite also ins Verhältniss zu setzen zu exakter (halber) Wellenlänge $\lambda/2$ bzw. zu exakter Abklinglänge κ^{-1}. Entsprechend $Eh^2 = k^2h^2 = \pi^2\xi^2$ bzw. $(V_0 - E)h^2 = \kappa^2h^2 = \eta^2$ fällt die Energie in diesen Relationen dann ganz heraus und wir finden (Abb. 4.5):

$$\frac{\lambda_h - \lambda}{\lambda} = \frac{k - k_h}{k_h} = \frac{\pi\xi - \arccos\left[1 - \pi^2\xi^2/2\right]}{\arccos\left[1 - \pi^2\xi^2/2\right]}, \qquad \xi \equiv \frac{h}{\lambda/2}, \tag{4.60}$$

$$\frac{\kappa_h^{-1} - \kappa^{-1}}{\kappa^{-1}} = \frac{\kappa - \kappa_h}{\kappa_h} = \frac{\eta - \operatorname{arccosh}\left[1 + \eta^2/2\right]}{\operatorname{arccosh}\left[1 + \eta^2/2\right]}, \qquad \eta \equiv \frac{h}{\kappa^{-1}}. \tag{4.61}$$

Wegen der auf einem h-Gitter durch $E_{\max} = 4/h^2$ nach oben beschränkten Energie – vgl. (A.6) auf Seite 82 und den anschließenden Text – muß infolge des $Eh^2/2$-Terms in (4.58) mit $E_{\max}h^2/2 = 2 \equiv \pi^2\xi_{\max}^2/2$ auch ξ durch $2/\pi \approx 0.64$ nach oben beschränkt bleiben[5], während η beliebige Werte annehmen kann. Aus Abb. 4.5 ist dann beispielsweise zu entnehmen, daß für $h = \lambda/4$ der Periodenfehler 12% beträgt, während eine vergleichbare Abweichung in der evanescenten Mode für $h \approx 2\kappa^{-1}$ auftritt.

Allerdings diskretisieren wir kein Streuproblem, sondern räumlich begrenzte Kästen mit diskreten Spektren, die diffizilere Probleme aufwerfen. Abb. 4.6 zeigt für einen Potentialtopf von 80 Å Breite, in dem 36 gebundene Zustände existieren und die Nummer 26 mit der Fermienergie von 4 eV korrespondieren würde, die relativen Eigenwertfehler von E_h, λ_h und κ_h^{-1} aufgetragen über der Knotenzahl der Zustände. D.h. hier werden ohne Energienormierung bei λ_h und κ_h^{-1} die Eigenwerte der Zustände direkt gegenübergestellt.

Von der Eigenwertgleichung (A.25) werden primär die k_h geliefert und zwar generell größer als die exakten k (die λ_h sind wie bereits bemerkt also zu klein). Die Fehler von λ_h und k_h liegen dabei in der Größenordnung $h/$*Kastenbreite* und werden somit auch absolut geringer, je breiter ein Kasten wird, was klarer dann Tabelle 4.1 dokumentiert. Zum Verständnis betrachte man auch Abb. A.1 auf Seite 87.

Aus den k_h folgen κ_h und E_h anschließend über (A.19) und (A.15). Bei kleinen Knotenzahlen besitzen Gitterlösungen in Abb. 4.6 eine geringfügig höhere Energie als ihre kontinuierlichen Pendants und weichen danach merklich nach unten ab. Erstaunlich mutet der

[5]Beachte: ξ setzt h in Bezug zur Kontinuumswellenlänge λ und nicht zu der des Gitters λ_h (letztere wäre durch $2h$ nach unten beschränkt), weshalb $\xi_{\max}$ auch $2/\pi$ und nicht 2 ist.

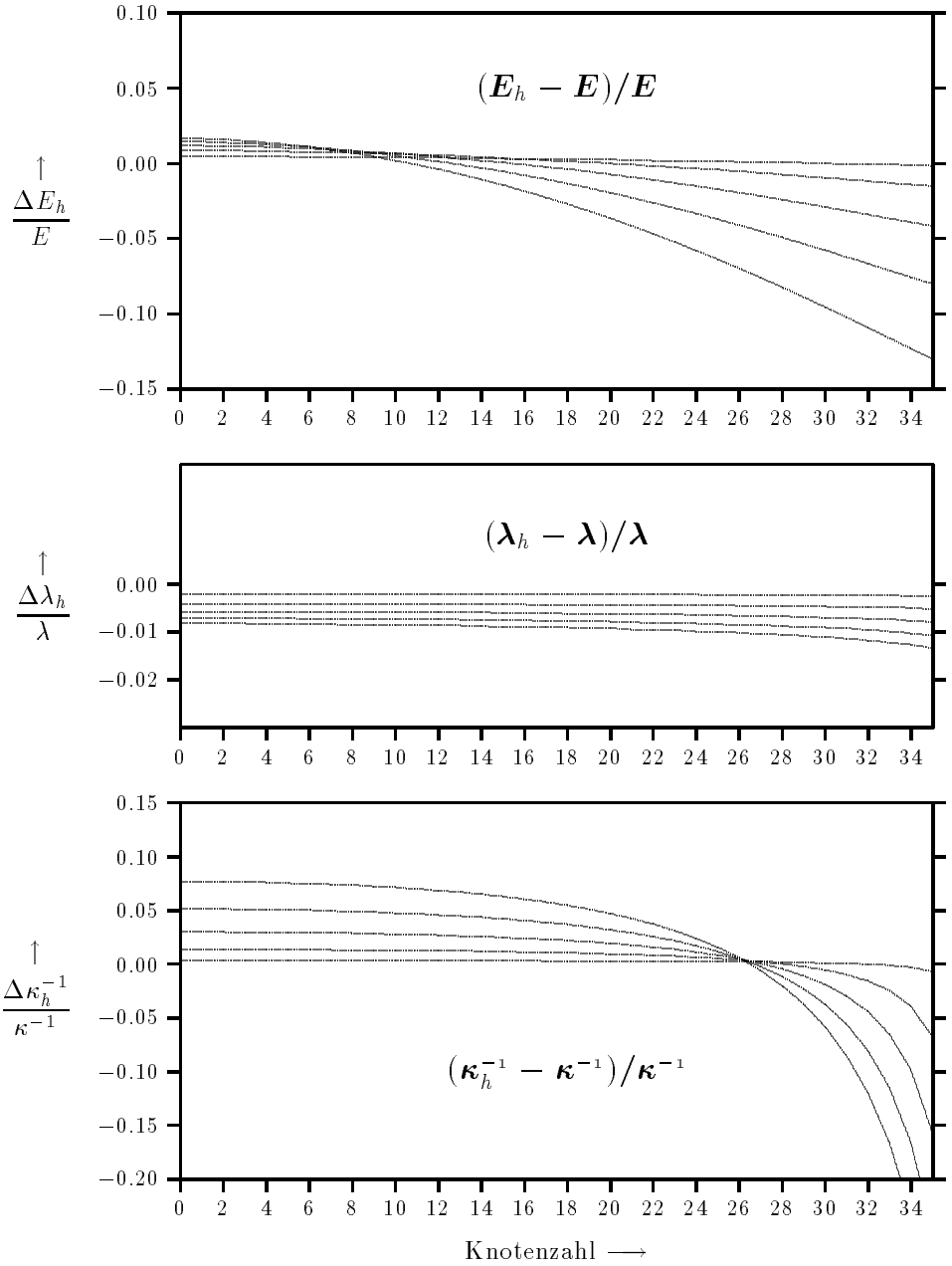


Abbildung 4.6: *Für einen Potentialtopf der Breite* 80 Å *und der Höhe* 8 *eV mit 36 gebundenen Zuständen wurden die Gitterlösungen mit ihren kontinuierlichen Pendants verglichen. Aufgetragen sind über der Knotenzahl der Zustände die relativen Abweichungen der Gitterwerte* E_h, λ_h *und* κ_h^{-1} *von den kontinuierlichen Werten* E, λ *und* κ^{-1} *für die Schrittweiten* $h = 0.2, 0.4, 0.6, 0.8, 1$ Å *(mit wachsendem Abstand von der Nullinie). Stetige Kurvenzüge verbinden hier die jeweils 36 diskreten Werte. Man beachte die unterschiedliche Skalierung der Ordinaten – der maximale Fehler bei* λ *beträgt ca.* 1.3%, *der bei* κ^{-1} *über* 20%*!*

Fehler in Prozent						
Breite →		20 Å	40 Å	80 Å	160 Å	320 Å
$h = 0.2$ Å	λ	−0.88	−0.46	−0.24	−0.12	−0.06
	κ^{-1}	0.88	0.48	0.23	0.11	0.06
	E	1.42	0.55	0.11	−0.12	−0.23
$h = 0.4$ Å	λ	−1.70	−0.89	−0.46	−0.23	−0.12
	κ^{-1}	1.67	0.80	0.39	0.20	0.11
	E	2.01	0.27	−0.54	−0.96	−1.17
$h = 0.8$ Å	λ	−3.18	−1.68	−0.85	−0.43	−0.21
	κ^{-1}	2.93	0.97	0.53	0.35	0.26
	E	0.49	−2.80	−4.11	−4.77	−5.10
$h = 1.6$ Å	λ	−5.44	−2.86	−1.44	−0.72	−0.36
	κ^{-1}	4.08	0.61	1.07	1.31	1.44
	E	−13.38	−18.29	−19.31	−19.83	−20.09

Tabelle 4.1: *Relative Fehler in Prozent der Gittereigenwerte eines eindimensionalen Kastens bei Variation von Kastenbreite und Schrittweite, registriert jeweils für denjenigen Eigenzustand, dessen Energie im kontinuierlichen Fall der halben Wandhöhe am nächsten kommt* ($E \approx V_0/2 = 4$ *eV*).

Koinzidenzpunkt bei κ_h^{-1} an, der, wie sich zeigte, unabhängig von V_0 und der Kastenbreite mehr oder weniger deutlich ausgeprägt stets nahe jener Knotenzahl erscheint, an der die Energie des kontinuierlichen Zustandes die halbe Wandhöhe erreicht ($E \approx V_0/2$), und zwar um so prägnanter, je breiter ein Kasten wird (je weniger die k_h von k differieren).

Tabelle 4.1 zeigt darüberhinaus das Verhalten der Diskretiersierungsfehler bei Variation von Kastenbreite und Schrittweite, und zwar jeweils für denjenigen Eigenzustand, dessen Energie im kontinuierlichen Fall der halben Wandhöhe (Fermi-Energie) am nächsten kommt ($E \approx V_0/2 = 4$ eV). Die mit $h/$*Kastenbreite* skalierenden λ_h-Abweichungen wurden bereits erwähnt, die κ_h liegen hier wegen $E \approx V_0/2$ stets in der Nähe jenes Koinzidenzpunktes, wodurch deren Fehler relativ klein bleiben, sich dadurch aber auch – zumindest für $h = 0.8$ Å und 1.6 Å – einer einfachen Tendenzaussage weitgehend entziehen. Die E_h schließlich fallen gegenüber E monoton, wenn Schrittweite und/oder Kastenbreite wachsen, was mit Hilfe von (4.54) und der anschließenden Argumentation auch gut verstanden werden kann.

4.3.6 Resümee

Insgesamt ist der Zusammenhang zwischen Gitterschrittweite, Eigenwertfehlern und Zustandsfehlern bereits bei diesen einfachen Beispielen recht komplex, und ein Hochrechnen auf zwei und mehr Dimensionen kann nur mit Vorbehalt erfolgen. Für die typische Anwendung eines Kastens von 80×50 Å^2 und $h = 0.4$ Å legt Tabelle 4.1 für Zustände mit $E \approx 4$ eV immerhin Fehler unter 1% für E_h, λ_h und κ_h^{-1} nahe. Hinzuzufügen, daß bei einer Eigenwertstatistik (die in dieser Arbeit nicht erfolgte) weniger die absolute Genauigkeit der Energieeigenwerte von Interesse wäre, als die Wahrung der Niveauabstände und die Frage, in welchem Umfang eine Diskretisierung Niveaukreuzungen hervorruft.

4.4 Zur Lösung des Eigenwertproblems

Die Details wurden nach Anhang B verbannt, an dieser Stelle wird das Verfahren im Überblick vorgestellt und Anhang C motiviert.

4.4.1 Grundprinzip

Zu lösen ist das Eigenwertproblem (4.15) der großen schwachbesetzten Diskretisierungsmatrix $\mathbf{A}$ für einige wenige (10 ... 30) Eigenwerte λ_j, und zwar speziell für diejenigen, die zu ϵ nächstbenachbart liegen. Äquivalent ist dies der Aufgabe, die betragskleinsten Eigenwerte von $\mathbf{A} - \epsilon\mathbf{I}$ zu finden.

Die Größe von $\mathbf{A}$ gestattet hier von vornherein nur die sogenannten vektoriterativen Verfahren, die mit einer Matrix-mal-Vektor-Vorschrift, $\underline{\mathbf{w}} = \mathbf{A}\underline{\mathbf{v}}$, auskommen und die Matrix ansonsten als Black Box behandeln. Deren Grundprinzip ist das Folgende: Sei $\{\underline{\mathbf{u}}_i\}$ ein vollständiges Orthonormalsystem der Matrix $\mathbf{A}$ mit $\mathbf{A}\underline{\mathbf{u}}_i = \lambda_i\underline{\mathbf{u}}_i$, wobei o.B.d.A.

$$|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$$

angenommen werde. Ein beliebiger Startvektor $\underline{\mathbf{v}}^{(0)}$, den man sich immer im System der $\underline{\mathbf{u}}_i$ entwickelt denken kann, $\underline{\mathbf{v}}^{(0)} = \sum_i c_i\,\underline{\mathbf{u}}_i$, wird gemäß der Iterationsvorschrift $\underline{\mathbf{v}}^{(k+1)} = \mathbf{A}\underline{\mathbf{v}}^{(k)}$ wiederholt mit $\mathbf{A}$ multipliziert:

$$\begin{array}{ccccccc}
\underline{\mathbf{v}}^{(1)} & = & \mathbf{A}\underline{\mathbf{v}}^{(0)} & = & \displaystyle\sum_i c_i\mathbf{A}\underline{\mathbf{u}}_i & = & \displaystyle\sum_i c_i\lambda_i\underline{\mathbf{u}}_i \\
\underline{\mathbf{v}}^{(2)} & = & \mathbf{A}\underline{\mathbf{v}}^{(1)} & = & \displaystyle\sum_i c_i\mathbf{A}^2\underline{\mathbf{u}}_i & = & \displaystyle\sum_i c_i\lambda_u^2\underline{\mathbf{u}}_i \\
\vdots & & \vdots & & & & \vdots
\end{array}$$

Für $|\lambda_1| > |\lambda_2|$ wird nach hinreichend vielen, sagen wir k, Iterationen in der rechtsstehenden Summe der Beitrag $c_1\lambda_1^k\underline{\mathbf{u}}_1$ des betragsgrößten Eigenwertes λ_1 dominieren, wodurch sich aus $\underline{\mathbf{v}}^{(k)}$ letztlich Näherungen für λ_1 und $\underline{\mathbf{u}}_1$ gewinnen lassen. Vermöge einer geeigenten Shift, z. B. $\mathbf{A} - \lambda_n\mathbf{I}$, ließe sich eine Konvergenz auch gegen λ_n erreichen, ansonsten aber gegen keinen anderen Eigenwert. Spezialfälle wie $|\lambda_1| = |\lambda_2|$ brauchen an dieser Stelle nicht zu interessieren, da von praktischer Relevanz ohnehin nur die in Anhang B beschriebenen simultanen Varianten der Vektoriteration sind, bei denen stets mehrere Vektoren zugleich iteriert werden.

Offensichtlich ist jedoch, daß diese Verfahren einen spektralen Abstand der gesuchten Eigenwerte von allen übrigen zur Voraussetzung haben (und desto größer dieser, desto besser die Konvergenz), und daß auf direktem Wege deshalb immer nur dominante (betragsgrößte) Eigenwerte bestimmt werden können, niemals auf direktem Wege jedoch solche aus der Mitte des Spektrums – wie bei uns vonnöten, aus der Umgebung von ϵ. Ist dies gefordert, arbeitet man gewöhnlich mit der Inversen Vektoriteration,

$$\underline{\mathbf{v}}^{(k+1)} = (\mathbf{A} - \epsilon\mathbf{I})^{-1}\underline{\mathbf{v}}^{(k)}, \tag{4.62}$$

weil die dominanten Eigenwerte der Iterationsmatrix $(\mathbf{A} - \epsilon\mathbf{I})^{-1}$, also die betragsgrößten $(\lambda_i - \epsilon)^{-1}$, gerade mit den gesuchten und am dichtesten bei ϵ gelegenen λ_i zusammenfallen. Praktisch wird man natürlich nie derart große Matrizen wirklich invertieren, sondern löst stattdessen bei jedem Iterationsschritt das Gleichungssystem

$$(\mathbf{A} - \epsilon\mathbf{I})\underline{\mathbf{v}}^{(k+1)} = \underline{\mathbf{v}}^{(k)} \tag{4.63}$$

nach $\underline{\mathbf{v}}^{(k+1)}$ auf. Die Koeffizientenmatrix $\mathbf{A} - \epsilon\mathbf{I}$ ist dabei indefinit, da eine Verschiebung des Nullpunktes in das Spektrum hinein zwangläufig negative und positive Eigenwerte zur Folge hat. Prinzipiell muß in unserem Falle sogar numerische Nichtregularität angenommen werden, da die Verschiebung um ϵ obendrein gewissermaßen „blind“ zum unbekannten und recht dichten $\{\lambda_i\}$-Spektrum erfolgt, und ϵ einem der λ_i beliebig nahe kommen kann. Aus diesem Grunde sind die verschiedenen schnellen (direkten) Lösungsverfahren, die für positiv definite Matrizen und da insbesondere für rechteckige Grundgebiete existieren [66, S. 848], hier nicht anwendbar. Als einzige Möglichkeit käme wegen der Größe von $\mathbf{A}$ eine *iterative* Auflösung von (4.63) in Frage.

Allgemein ist die Theorie der *iterativen* Lösung großer schwachbesetzter Gleichungssysteme [34, 57] weit ausgebaut für positiv definite Matrizen, d.h. relativ sicher, klar und allgemeingültig dort hinsichtlich zu wählender Iterationsparameter. Erprobte Codes sind verfügbar [66]. Gleiches gilt nicht für den indefiniten, speziell den nichtregulären Fall. Ohne einschränkende Positivitätsbedingungen arbeiten dort nur die modernen Mehrgitterverfahren [34, S. 278], allerdings ist in den Konvergenzbeweisen, die die große Schnelligkeit dieser Verfahren anzeigen, wenigstens implizit stets eine Regularitätsannahme enthalten [34, S. 331]. Ohne diese Annahme kommt man nur zu wesentlich schwächeren Aussagen (ebenda) und auch können im indefiniten Fall Stabilitätsprobleme auftreten [34, S. 301]. Ohnehin gibt es nicht das Mehrgitterverfahren schlechthin, sondern spezialisierte Varianten für die verschiedenen Aufgaben, wobei der nichtreguläre Fall allsamt wenig untersucht ist. (Selbst bei positiv definiten Aufgaben müssen die optimalen Iterationsparameter bei neuen Matrixbelegungen mehr oder weniger 'per Hand' eingestellt werden [66, S. 862].) Eine eigene Implementation eines Mehrgitterverfahrens für den nichtdefiniten Fall ohne oder mit nur sehr schwachen Reguläritätsvoraussetzungen – eher Gegenstand wohl der derzeitigen mathematischen Forschung – zur Lösung von (4.63) ließ daher erhebliche Komplikationen erwarten.

Um diesen Schwierigkeiten aus dem Wege zu gehen, wurde zur Eigenwertbestimmung ganz auf die Inverse Vektoriteration verzichtet und stattdessen mit der quadrierten Form

$$\mathbf{B} = (\mathbf{A} - \epsilon\,\mathbf{I})^2 - \rho\mathbf{I} \tag{4.64}$$

iteriert, wozu ein $\mathbf{A}\underline{\mathbf{v}}$–Algorithmus genügt. Die Verschiebung um ρ ist dabei so zu wählen, daß die kleinsten Eigenwerte von $(\mathbf{A} - \epsilon\,\mathbf{I})^2$ zu dominanten von $\mathbf{B}$ werden. Der große Vorteil ist der Wegfall des zweiten, inneren Iterationsverfahrens zur Gleichungsauflösung und dessen jedmaliger Justage bei neuen Matrixbelegungen. Die Form des Grundgebietes, die die Matrixtopologie bestimmt und bei der iterativen Lösung großer Gleichungssysteme eine empfindliche Rolle spielt, betrifft hier nur noch die $\mathbf{A}\underline{\mathbf{v}}$–Vorschriften und kann insofern recht frei variiert werden. Der nicht unkritische Nachteil allerdings ist, daß das Spektrum von $\mathbf{B}$ verglichen mit dem von $\mathbf{A}$ für eine Vektoriteration bedeutend ungünstiger wird (sehr schwache Konvergenz, bis zu 30.000 Iterationen bei unseren Anwendungen, Problem der Akkumulation von Rundungsfehlern in den Iterierten). Näher dargelegt ist dies in B.7, wo wir die Konvergenzproblematik für eine Iteration mit (4.64) näher untersucht haben.

Unter den verschiedenen vektoriterativen Eigenwertverfahren wurde die *Simultane Iteration* gewählt, speziell die aus der Litaratur bekannte Variante RITZIT von RUTISHAUSER [70, 71], die Teilraumiteration und Tschebyscheff-Beschleunigung kombiniert. Zur Optimierung der Tschebyscheff-Beschleunigung sind gewisse Vorabinformationen über das Eigenwertspektrum förderlich, über die wir dank geeigneter exakter Vergleichslösungen (An-

hang A) in ausreichendem Maße aber verfügen. Zur Minderung der Fehlerakkumulation mußte RITZIT allerdings auf Kaskadensummation umgestellt werden.

4.4.2 Über die Matrix-mal-Vektor-Algorithmen

An den Matrix-mal-Vektor-Vorschriften, in denen das Programm bei einer typischen Anwendung 99% bis 99.9% seiner Zeit verbringt, entscheiden sich zentral Rechenzeit und Aufwand. Es ist angebracht, diesen Algorithmen mit Sorgfalt zu begegnen. Folgende, in der Regel miteinander konkurierende Aspekte spielen eine Rolle:

1. Operationsbedarf
2. Lokalität[6]
3. Fehlerakkumulation
4. Parallelisierbarkeit

Zur Lokalität diese Bemerkung: Bei einem Gitter von 500×500 Punkten und 8-Byte-Zahlen belegt der einzelne Vektor 2.000.000 Byte. Für ein einfaches Skalarprodukt $\underline{\mathbf{w}} = \underline{\mathbf{a}}\,\underline{\mathbf{v}}$, wobei $\underline{\mathbf{a}}$ die Hauptdiagonale der Matrix sein kann, muß folglich ein Adressraum von sechs Megabyte realiter durchlaufen werden. Es ist leicht vorstellbar, daß bereits bei einem Skalarprodukt ein externes Medium bemüht werden muß, und daß ein von der Zahl der Operationen her optimaler $\mathbf{A}\underline{\mathbf{v}}$–Algorithmus in der Praxis dann deutlich langsamer sein kann als ein diesbezüglich schwächerer, der in seinen Speicherzugriffen aber lokaler arbeitet.

Nur die Hauptdiagonale der Diskretisierungsmatrix $\mathbf{A}$ enthält Informationen über das Innere des numerischen Grundgebietes, nur dort geht das Potential ein. Demgegenüber sind alle von Null verschiedenen Nichtdiagonalelemente vom Wert -1 und deren Anordnung wird allein durch die Geometrie des Grundgebietes bestimmt[7]. Aus Platzgründen halten wir daher nur die Hauptdiagonale von $\mathbf{A}$ im Speicher, während Nebendiagonalelemente innerhalb der Matrix-mal-Vektor-Algorithmen „on line“ zu rekonstruieren sind. Zumindest kann die Verschiebung $\mathbf{A} - \epsilon\mathbf{I}$ in (4.64), im Gegensatz zur Verschiebung mit ρ, dann immer bereits vor dem Rechengang physisch ausgeführt werden ($\mathbf{A} - \epsilon\mathbf{I} \equiv \mathbf{A}_\epsilon$) und wir benötigen folglich eine Vorschrift für

$$\underline{\mathbf{w}} = (\mathbf{A}_\epsilon^2 - \rho\mathbf{I})\underline{\mathbf{v}} \quad , \quad \mathbf{A}_\epsilon \equiv \mathbf{A} - \epsilon\mathbf{I}, \tag{4.65}$$

Die einfachste Lösung bietet hier zweifellos das rekursive Zurückführen auf zwei einfache $\mathbf{A}_\epsilon\underline{\mathbf{v}}$–Operationen plus anschließender Verschiebung, wozu ein Hilfsvektor $\underline{\mathbf{h}}$ benötigt wird:

$$\underline{\mathbf{w}} = \mathbf{A}_\epsilon(\mathbf{A}_\epsilon\underline{\mathbf{v}}) - \rho\underline{\mathbf{v}} \quad : \quad \text{(i)}\ \ \underline{\mathbf{h}} = \mathbf{A}_\epsilon\underline{\mathbf{v}}\,, \quad \text{(ii)}\ \ \underline{\mathbf{w}} = \mathbf{A}_\epsilon\underline{\mathbf{h}} - \rho\underline{\mathbf{v}}\,. \tag{4.66}$$

Es ist zweckmäßig, die zweite Multiplikation $\mathbf{A}_\epsilon\underline{\mathbf{h}}$ rückwärts auszuführen, weil nach (i) die unteren Teile von $\underline{\mathbf{h}}$ und diag($\mathbf{A}$) im Speicher stehen, die bei einem Seitenwechsel sonst verworfen werden müßten.

[6] Betriebssysteme mit virtueller Speicherverwaltung (Paging) nutzen die Erfahrung, daß zeitlich aufeinanderfolgender Programmcode überwiegend in benachbarten Adressbereichen abläuft bzw. auf Daten aus benachbarten Adressräumen zugreift (Prinzip der Lokalität) und laden jeweils nur aktuell benötigte Speicherseiten in den Kernspeicher. Größere Adressprünge im Datenbereich können daher zu häufigen Seitenwechseln mit deutlichen Effektivitätsverlusten führen.

[7] und natürlich durch die Art der Numerierung der Gitterpunkte

Um im folgenden Mehrdeutigkeiten vorzubeugen und den algorithmischen Aspekt etwas hervorzuheben, verwenden wir fortan für alle *nichtrekursiven* Basisoperationen kalligraphische Buchstaben: $\mathcal{A}\underline{\mathbf{v}}$, $\mathcal{B}\underline{\mathbf{v}}$, $\mathcal{B}^r\underline{\mathbf{v}}$. $\mathcal{A}\underline{\mathbf{v}}$ steht für $\mathbf{A}\underline{\mathbf{v}}$ oder $\mathbf{A}_\epsilon\underline{\mathbf{v}}$, $\mathcal{B}\underline{\mathbf{v}}$ führt nichtrekursiv $\mathbf{B}\underline{\mathbf{v}}$ aus, $\mathcal{B}^r\underline{\mathbf{v}}$ desgleichen die r-te Potenz $\mathbf{B}^r\underline{\mathbf{v}}$.

Sei ξ_A die Zahl der Nichtnullelemente einer Zeile bzw. Spalte in $\mathbf{A}$, welche identisch ist mit der Zahl der Summanden in einer $\mathcal{A}\underline{\mathbf{v}}$-Prozedur – z. B. ist $\xi_A = 5$ in zwei Dimensionen (Abb. 4.2) und $\xi_A = 7$ in drei (Abb. 4.3) –, so beträgt der unmittelbare Aufwand für einen Iterationsschritt (4.66), also für den Algorithmus $\mathcal{A}(\mathcal{A}\underline{\mathbf{v}}) - \rho\underline{\mathbf{v}}$:

$$\text{Aufwand für (4.66)} \;\sim\; n \cdot \big\{2\,[\,\text{opm} + (\xi_A - 1)\text{ops}\,] + \text{opms}\big\} \tag{4.67}$$

mit $n = \dim(\underline{\mathbf{v}})$ und den üblichen Abkürzungen: opm = Gleitkomma-Multiplikation, ops = Gleitkomma-Summation, opms = opm + ops. Traditionell werden Indexoperationen nicht mitgezählt. In (4.67) wurde berücksichtigt, daß die Nichtdiagonalelemente "-1" in $\mathbf{A}$ keine Multiplikation erfordern. Hinzu käme der Verwaltungsaufwand zur Rekonstruktion der Positionen der Nichtdiagonalelemente, der bei einer komplizierten Grundgebietsgeometrie recht beachtlich sein kann. Alles in allem dürfte (4.67) kaum zu unterbieten sein.

Zur Fehlerakkumulation: Werden N Zahlen in der üblichen rekursiven Weise aufsummiert, wachsen die relativen Rundungsfehler $F \sim (N-1)$ [42, Kap. 2]. Verbesserungen lassen sich erreichen durch Summation in aufsteigender Größe und es gibt die günstigere Kaskadensummation, $F \sim \log_2(N)$, in beiden Fällen müssen allerdings alle Summanden simultan zur Verfügung stehen (ebenda und vgl. auch B.6.2, insbesondere Abb. B.1). Für einen Iterationsschritt (4.66) beträgt bei rekursiver Summation innerhalb von $\mathcal{A}\underline{\mathbf{v}}$ der Fehler somit $F \sim 1+2(\xi_A-1)$ (die vordere Eins stammt von der ρ-Verschiebung) und für einen k-schrittigen Iterationsblock

$$\underline{\mathbf{w}}^{(k)} = \mathbf{B}^k\underline{\mathbf{v}} \tag{4.68}$$

dann

$$F \;\sim\; k \cdot [1 + 2(\xi_A - 1)] \;=\; k \cdot [2\xi_A - 1]. \tag{4.69}$$

Diese Fehlerakkumulation ist bei, sagen wir, $k \approx 10.000$ sehr wohl in Betracht zu ziehen und stellt eine ernste Schranke dar, leider kann aber die lineare Fehlerfortpflanzung zwischen zwei rekursiven Matrix-mal-Vektor-Multiplikationen, $\mathcal{A}(\mathcal{A}\underline{\mathbf{v}})$ oder $\mathcal{B}(\mathcal{B}\underline{\mathbf{v}})$, prinzipiell nicht unterbunden werden, da nach der inneren Multiplikation definitiv das erste Mal aufsummiert werden muß. Allein innerhalb der elementaren $\mathcal{A}\underline{\mathbf{v}}$–Prozedur bestände die Möglichkeit einer Einflußnahme[8], wobei dort der Effekt z. B. einer Kaskadensummation wegen der geringen Summandenzahl aber ganz unwesentlich wäre.

Die $\mathcal{A}\underline{\mathbf{v}}$–Algorithmen sind mit Sicherheit die am leichtesten zu implementierenden. Es gibt nun drei Gründe, darüberhinaus auch nichtrekursive, d. h. direkt quadrierende Prozeduren $\mathcal{B}\underline{\mathbf{v}}$ oder überhaupt höher potenzierende $\mathcal{B}^r\underline{\mathbf{v}}$ (r=1,2,...) ins Kalkül zu ziehen.

1. Die Parallelisierbarkeit von (4.66) ist gering, weil Schritt (ii) definitiv auf die Fertigstellung von (i) warten muß. Man wird einwenden, daß die natürliche Form der Parallelisierung einer simultanen Vektoritertation, bei der p Vektoren $\underline{\mathbf{v}}_j\,(j = 1, \ldots, p)$ zur gleichen Zeit iteriert werden, die parallele Ausführung der $\mathcal{A}\underline{\mathbf{v}}_j$–Operationen sein sollte. Hierzu müssen jedoch große, zum Teil Kopien der gleichen – etwa diag($\mathbf{A}$) –

[8] Natürlich könnte man die Mantissenzahl der Iterierten auch generell erhöhen, was aber zu bezahlen wäre mit erheblich mehr Speicherbedarf, größeren Datenströmen, häufigeren Seitenwechseln, längeren Rechenzeiten. Bewegt man sich bereits am Limit, ist dies nicht praktikabel.

Datenmengen durch jeden Prozessor geschleust werden und es kann sehr wohl vorteilhafter sein, auf lokalerer Ebene in den $\mathcal{B}^r\underline{\mathbf{v}}$–Prozeduren zu parallelisieren.

2. Es könnten $\mathcal{B}^r\underline{\mathbf{v}}$–Algorithmen existieren, die lokaler arbeiten oder mit weniger Operationen auskommen als (4.66).

3. Die Fehlerakkumulation kann verringert werden.

Der letzte Punkt bedarf einer Erklärung. Wie erwähnt, lassen sich Rundungsfehler nur innerhalb einer elementaren Matrix-mal-Vektor-Prozedur mindern, während über rekursive Aufrufe hinweg Fehler linear akkumuliert werden. Der Grundgedanke wäre nun, vermöge höherpotenzierender Prozeduren $\mathcal{B}^r\underline{\mathbf{v}}$ die Zahl der rekursiven Matrix-mal-Vektor-Multiplikationen in (4.68) zu verringern, was i. allg. natürlich mit zusätzlichen Operationen in den $\mathcal{B}^r\underline{\mathbf{v}}$ bezahlt wird, in denen dann aber eine optimierte Summationsreihenfolge und eine Kaskadensummation besser greift. Der Effekt zumindest für die Kaskadensummation kann abgeschätzt werden:

Sei ξ_{B^r} in Analogie zu ξ_A die Zahl der Summanden in einer $\mathcal{B}^r\underline{\mathbf{v}}$–Prozedur, so lauten die Fehler für einen k-schrittigen Iterationsblock (4.68) dann je nach Algorithmus genähert[9]:

$$\mathcal{A}(\mathcal{A}\underline{\mathbf{v}}) - \rho\underline{\mathbf{v}} \quad : \quad F \sim k \cdot [1 + 2\log_2(\xi_A)], \tag{4.70}$$

$$\mathcal{B}\underline{\mathbf{v}} \quad : \quad F \sim k \cdot \log_2(\xi_B), \tag{4.71}$$

$$\mathcal{B}^r\underline{\mathbf{v}} \quad : \quad F \sim k/r \cdot \log_2(\xi_{B^r}). \tag{4.72}$$

Eine Verbesserung mit (4.72) gegenüber (4.69) tritt ein, wenn der Quotient

$$Q_r = \frac{1/r \cdot \log_2(\xi_{B^r})}{2\xi_A - 1} \quad , \quad r = 1, 2, \ldots$$

unter Eins fällt, wegen $Q_r \sim 1/r$ ist der Effekt allerdings nicht sehr groß. Durch elementare Multiplikation oder mit Hilfe des Satzes C.1 auf Seite 112 findet man beispielsweise für unsere Diskretisierungsmatrix in zwei Dimensionen mit $\xi_A = 5$ (Abb. 4.2) für die Matrix $\mathbf{B}$ ($r = 1$) die Zahl $\xi_{B^1} = 13$ und für $\mathbf{B}^2$ ($r = 2$) die Zahl $\xi_{B^2} = 41$, und damit, wenn einmal im Sinne der Fußnote 9 exakt verfahren wird, $Q_1 = 2/3$, $Q_2 = 1/2$, also eine Halbierung der relativen Rundungsfehler mit $\mathcal{B}^2\underline{\mathbf{v}}$.

Diese mehr abstrakten Überlegungen können den konkreten Versuch mit einem real arbeitenden $\mathcal{B}^r\underline{\mathbf{v}}$–Algorithmus natürlich nicht ersetzen. Angesichts der Wichtigkeit der Matrix-mal-Vektor-Operationen bei den extrem schwachen Konvergenzraten unserer Anwendung und natürlich die Einfachheit der Diskretisierungsmatrizen vor Augen wurde daher speziell einmal für die von uns dann so bezeichneten Linienmatrizen – Verallgemeinerungen der Matrixformen in Abb. 4.2 und Abb. 4.3 – ein Formalismus entwickelt (Anhang C), mit dem solche $\mathcal{B}^r\underline{\mathbf{v}}$–Prozeduren relativ geradlinig entworfen werden können. Konkret experimentiert wurde dann mit verschiedenen $\mathcal{B}\underline{\mathbf{v}}$–Prozeduren. Leider stand ein Parallelrechner nicht zur Verfügung, so daß der vermeintliche Hauptvorteil noch nicht hinterfragt ist. Auf einem Einprozessor-Rechner (RISC 6000 mit 16 MB RAM) benötigten $\mathcal{B}\underline{\mathbf{v}}$-Prozeduren wenigstens das 1.5fache an Zeit wie (4.66). Gleichwohl das Resultat also zunächst negativ ausfiel, wurde mit Blick auf eine spätere Parallelisierung der Formalismus in Anhang C trotzdem dokumentiert.

[9] Strenggenommen gilt $F \sim \log_2(N)$ nur für N als Zweierpotenz, während für beliebiges N Parallelkaskaden notwendig sind. Z. B. für $N = 13 = 2^3 + 2^2 + 1$: $F \sim 3 + 2 + 1 = 6$

Kapitel 5

Billardzustände

5.1 Vorbemerkungen

Berechnet man Billardzustände – worunter wir hier in Erweiterung des üblichen Sprachgebrauchs gebundene Zustände sowohl in zweidimensionalen Kästen mit unendlich hohen als auch mit endlich hohen Wänden verstehen wollen – in irgendeiner willkührlich vorgegebenen Kastengeometrie, so wird man nahezu auch beliebige Zustandsformen finden können. Bei entsprechend kurzen Wellenlängen unterscheiden sich energetisch benachbarte Zustände ein und desselben Billards in ihrem Aussehen meist vollkommen und minimale Änderungen des Systems haben oft drastische und globale Änderungen einer herausgegriffenen Eigenfunktion zur Folge [60]. Der einzelne Zustand im semiklassischen Limes könnte daher leicht als pathologisches Objekt eingestuft werden, gäbe es nicht zugleich erfolgversprechende Ansätze, selbst die einzelne (semiklassische) Wellenfunktion im Rahmen einer Theorie periodischer Bahnen zu verstehen (vgl. Abschnitt 3.3 und die dortigen Zitate).

Nach einem Überblick zu bekannten exakten Lösungen und unseren numerischen Tests werden im folgenden Zustände vorgestellt, die in den zur Simulation des STM-Abbildungsprozeßes benutzten Potentialkästen von der Form einer Spitze gefunden wurden. Unser besonderes Augenmerk gilt in 5.4 und 5.6 dann zwei Systemen – dem nichtseparablen Rechteck und zwei Quantenbillards im Tunnelkontakt –, bei denen die Besonderheit der Problematik „Quantenbillards endlicher Wandhöhe“ exemplarisch zum Ausdruck kommt.

5.2 Exakte Resultate

Tatsächlich gibt es nur sehr wenige im Rahmen einer Einteilchen-SGL analytisch behandelbare Quantensysteme mit gebundenen Zuständen. Setzt man weiterhin voraus, daß die Bindung in Potentialkästen mit steilen Wänden erfolgen soll, so reduziert sich deren Anzahl auf die folgenden Beispiele:

- eindimensionaler Potentialkasten beliebiger Wandhöhe
- rechteckige Potentialkästen mit unendlich hohen Wänden in beliebiger ganzzahliger Dimension
- kreis- und kugelförmige Potentialkästen beliebiger Wandhöhe

- dreieckige Potentialkästen mit unendlich hohen Wänden für die Innenwinkelkombinationen $(\frac{\pi}{3}, \frac{\pi}{3}, \frac{\pi}{3})$, $(\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{4})$ und $(\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{6})$ [39], siehe auch [54].

Daneben existieren natürlich noch jene simplen Erweiterungen auf höhere Dimensionen, bei denen sich die zusätzlichen Koordinaten absepariern lassen.

Besonders gravierend erscheint in diesem Zusammenhang, daß die SGL bereits für den sehr einfach anmutenden Fall des zweidimensionalen rechteckigen Potentialkastens *endlicher* Wandhöhe nicht mehr separiert werden kann, also für das Potential

$$V_R(x,y) = \begin{cases} 0, & \text{innen: } (0 < x < a) \wedge (0 < y < b); \\ V_0, & \text{außen,} \end{cases} \tag{5.1}$$

da für dieses – im Gegensatz zu dem des bekanntlich separablen *unendlich* hohen Kastens – keine Darstellung mehr in der gehörigen additiven Gestalt eines reinen x-Terms plus eines reinen y-Terms existiert. Um dies zu sehen, d. h. um zu einer analytischen Darstellung für $V_R(x,y)$ zu gelangen, defnieren wir uns zunächst eine „Topffunktion“

$$T(x,a) \equiv \Theta(x-a) + \Theta(-x) = \begin{cases} 0\,, & 0 < x < a; \\ 1\,, & \text{sonst;} \end{cases} \tag{5.2}$$

die einen eindimensionalen Topf der Tiefe 1 und der Breite a beschreibt. Superposition nun je eines solchen x- und y-Topfes je der Wandhöhe V_0 und der Breiten a bzw. b gemäß

$$V_{\text{sep}}(x,y) = V_0\left[T(x,a) + T(y,b)\right] = \begin{cases} 0, & \text{innen;} \\ 2V_0, & x \notin (0,a) \wedge y \notin (0,b); \\ V_0, & \text{sonst,} \end{cases} \tag{5.3}$$

liefert sodann ein – infolge seiner Additivität natürlich – seperables Potentialgebilde V_{sep}, das sich allerdings vom „echten“ Rechteck (5.1) noch insofern unterscheidet, als sich in den Gebieten diagonal über den Ecken, in denen beide Koordinaten außen liegen, der Wert $2V_0$ findet statt V_0. Den echten rechteckigen Potentialtopf gibt dann erst

$$V_R(x,y) = V_{\text{sep}}(x,y) - V_0\,T(x,a)\,T(y,b)\,, \tag{5.4}$$

wo durch Subtraktion eines x-y-Produktterms genau dieser Unterschied korrigiert ist. Dieser Produktterm zerstört aber die Separabilität.

5.3 Test des numerischen Verfahrens

Die Potentialkästen in dieser Arbeit, sofern sie zur STM-Simulation benutzt wurden, erhielten wie in 2.1 erwähnt die Tiefe $V_0 = 8$ eV und die Fermi-Energie wurde mit $E_F = 4$ eV angesetzt, womit die Austrittsarbeit also ebenfalls $\Phi_A = 4$ eV betrug. Die Energieeigenwerte der hier gezeigten Wellenfunktionen liegen dann immer in unmittelbarer Nähe zu $E_F = 4$ eV.

Die Diskretisierung erfolgte auf einem äquidistanten Gitter der Schrittweite $h = 0.4$ Å in einem rechteckigen Einbettungsgebiet, wobei die Symmetrien der Ausgangsgeometrien gewahrt wurden. Im numerischen Verfahren selbst spielten diese Symmetrien keine Rolle (hier lag immer das volle Problem vor), sie ermöglichten aber eine einfache Kontrolle der numerischen Genauigkeit. Gemäß unseren Ausführungen in 4.3.5 sollten bei einer Schrittweite von 0.4 Å Diskretisierungsfehler bei Energieeigenwerten und Eigenzuständen unter 1% bleiben.

Getestet wurde das Eigenwertverfahren an einem $80 \times 50\,\text{Å}^2$ großen Rechteck mit dem separablen Potentialgebilde (5.3), für das neben den Kontinuums- auch die Gitterlösungen analytisch bekannt sind (Anhang A). Nach ca. 20.000 Iterationen waren die exakten Gittervektoren mit einer Genauigkeit von etwa 10^{-4}, die Eigenwerte mit etwa 10^{-7} erreicht, wobei simultan i. allg. 20 Vektoren plus 4 oder 6 weitere zur Konvergenzbeschleunigung iteriert wurden. Abb. 5.1 zeigt auf der linken Seite die Betragsquadrate $|\psi(x,y)|^2$ einiger dieser Zustände an der Fermi-Grenze mit offensichtlich dem erwarteten Verhalten: der korrekten Symmetrie, einer sinusartigen Oszillation innen und einem exponentiellen Abklingen außen, wobei das Abklingen in Richtung der kürzeren Wellenlänge (und somit höheren kinetischen Energie) korrekterweise langsamer erfolgt als quer dazu. Die leichten Streifen sind darstellungsbedingt.

Bei den $|\Psi|^2$ zeigenden Halbtonbildern stehen helle Pixel für eine hohe Aufenthaltswahrscheinlichkeit und dunkle für eine geringe, mit Ausnahme des Untergrundes $|\psi|^2 \approx 0$, der zur besseren Kontrastierung der Elektrodenränder und der Knotenlinien wieder weiß eingefärbt wurde. In Abb. 5.1 markiert ferner der hellgraue Randbereich (entspricht dort der Untergrundfarbe) einmal das gesamte numerische Einbettungsgebiet.

Um das Verfahren auch an einem weniger ideal mit der Gitter- und Grundgebietsgeometrie korrespondierenden System zu testen, wurden kreisförmige Potentialtöpfe hinzugezogen. Abb. 5.2 zeigt für einen Kreis mit dem Radius $R = 20$ Å einen Zustand relativ hoher Drehimpulsquantenzahl, der zugleich ein schönes Beispiel für die Korrespondenz zwischen klassischer Mechanik und semiklassischer Quantenmechanik im Falle eines separablen Systems liefert: Ein klassisches Teilchen entsprechend hohen Drehimpulses bewegt sich auf einer Rosettenbahn, die den zentrumsnahen Bereich ausspart (Drehimpulsbarriere) – die Wellenfunktion geht dort gegen Null.

5.4 Das nichtseparable Rechteck

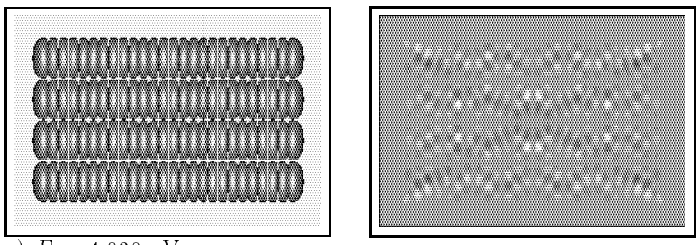
Im Anschluß an das separable Potentialgebilde (5.3) wurde ein „echter“ Potentialkasten von $80 \times 50\,\text{Å}^2$ nach (5.1) bzw. (5.4) betrachtet, der bereits ein nichttriviales und analytisch nicht mehr verifizierbares Resultat aufwirft. Die Quantenzahlen k_x und k_y existieren nur noch näherungsweise und sind – für eine semiklassische Betrachtung besonders interessant – allein durch den Potentialverlauf in einem klassisch nicht erlaubten Gebiet gekoppelt (das klassische Problem bei dieser Energie ist separabel, das quantenmechanische nicht[1]).

Rein optisch wären die gefundenen Zustände allerdings nicht von den bereits in Abb. 5.1 links gezeigten zu unterscheiden, weshalb der dortigen Abbildung nur noch die Differenzen der Wahrscheinlichkeitsdichten,

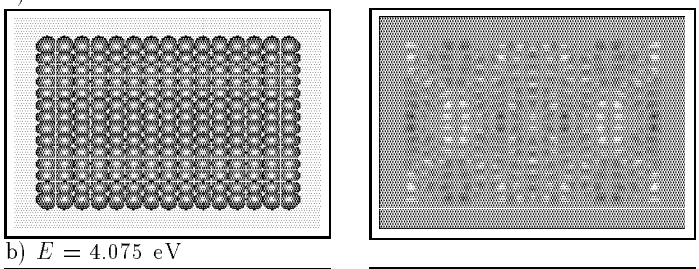
$$|\psi_n^{(5.4)}(\underline{\mathbf{q}})|^2 \; - \; |\psi_n^{(5.3)}(\underline{\mathbf{q}})|^2 \;, \quad \underline{\mathbf{q}} \equiv (x,y), \tag{5.5}$$

der Zustände beider Systemen hingefügt wurden. Wir bemerken, daß eigentlich erst eine solche Konstellation – Beibehaltung der Randgeometrie, Änderungen nur im Außenraum – den ganz direkten, quantitativen Vergleich der Zustände gestattet; der Übergang *integrabel* $\Leftrightarrow$ *nichtintegrabel* dagegen an einem unendlich hohen Billard vollzogen, erfordert die Verbiegung des Randes und erlaubt den Vergleich nur noch mittelbar [60].

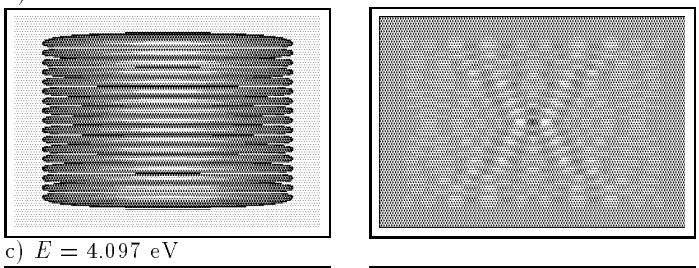
[1] Weil der Hamiltonoperator auch die evanescenten Raumbereiche beschreiben muß, die in der klassischen Hamiltonfunktion für $E < V_0$ gar nicht vorkommen



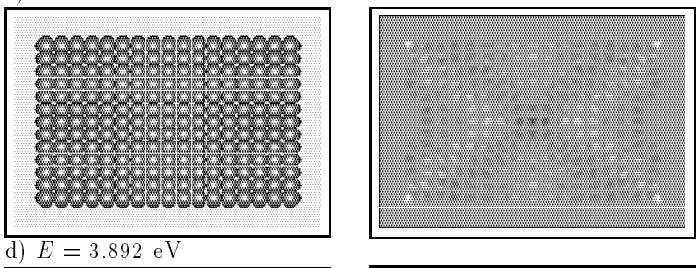
a) $E = 4.038$ eV



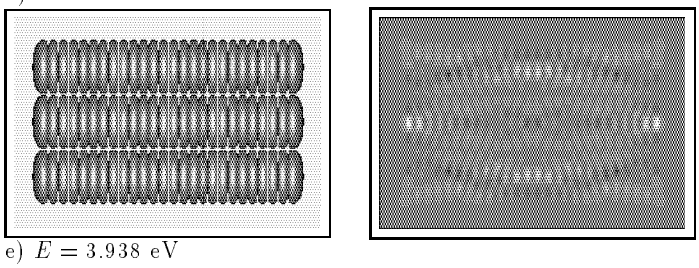
b) $E = 4.075$ eV



c) $E = 4.097$ eV



d) $E = 3.892$ eV



e) $E = 3.938$ eV

Abbildung 5.1: *Links das Betragsquadrat $|\psi(x,y)|^2$ einiger Zustände des separablen Potentialgebildes (5.3) bzw. – da optisch in dieser Form nicht unterscheidbar – auch des „echten“, nichtseparablen Kastens (5.4). Rechts die Differenzen (5.5) einander entsprechender Zustände beider Systeme. Positive Differenzen sind dabei hell, negative dunkel dargestellt und der Nullpunkt liegt im grauen Bereich.*

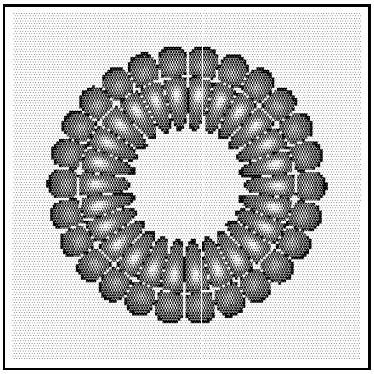


Abbildung 5.2: $|\psi(x,y)|^2$ *für einen Eigenzustand mit* $E \approx 4$ eV *in einem kreisförmigen Potentialtopf mit dem Radius* $R = 20\,\text{Å}$ *und* $V_0 = 8$ eV. *Gegenüber Abb. 5.1 ist die Pixelgröße hier verdoppelt.*

Bemerkenswert erscheint das unterschiedliche Verhältnis der Differenzenmuster zu ihren Ausgangswellenfunktionen – erkennbare Korrelationen zumindest der Knotenlinien wie in Abb. 5.1 (b) bis hin zu erstaunlicher Andersartigkeit in Abb. 5.1 (c) – und bemerkenswert insbesondere, daß bei manchen Zuständen diese Differenzen die Form von periodischen Bahnen (Scars) annehmen, Abb. 5.1 (d) und (e). Das heißt, gleichwohl das System als Ganzen nur sehr schwach nichtintegrabel geworden ist und die Zustände im wesentlichen und in ihrer Topologie gänzlich unverändert bleiben, zeigen die Änderungen mancher Zustände in Gestalt der Scars – Scars treten nur in chaotischen Systemen auf und sind an die Existenz instabiler Orbits gekoppelt – ganz unmittelbar an, daß der Übergang *integrabel* $\Leftrightarrow$ *nichtintegrabel* stattgefunden hat, gewissermaßen die Änderungen am System „chaotisch" waren. Warum jedoch gerade Zustand (e) „mit einem Scar antwortet" und nicht dagegen der sehr ähnliche Zustand (b), ist unbekannt. Eine weiterführende Untersuchung sollte die Eigenfunktionen einmal bei einem allmählichen Anwachsen der die Separabilität zerstörenden Anteile des Außenraumpotentials studieren.

Festzuhalten ist, daß energetisch benachbarte Zustände qualitativ sehr unterschiedlich auf die kleine, hier die Separabilität des Problems berührende Änderung des Außenraumpotentials[2] reagieren, und daß zweitens diese Reaktion als „Verlust" oder „Einbau" spezifischer Orbits verstanden werden könnte. In ähnliche Richtung weist das in 3.3 referierte Ergebnis von Bogomolny [16], der – allerdings für ein vollständig chaotisches System – zeigte, daß in energetisch gemittelten Wellenfunktionen bei unterschiedlichen Energien Beiträge ganz unterschiedlicher Orbits dominieren können.

Neu an diesem numerischen Experiment ist, daß der Übergang *integrabel* $\Leftrightarrow$ *nichtintegrabel* in einem nur quantenmechanisch zugänglichen Raumbereich stattfindet, d. h. die gewöhnlich betrachteten klassisch-reellen Orbits (bis zur Energie V_0) und deren Klassifikation anhand der Stabilitätsmatrix (3.25) ändern sich gar nicht.

Im herkömmlichen WKB-Verfahren lassen sich Wellenfunktionen ohne weiteres in den Außenraum hinein fortsetzen [12], in einer Theorie, die Zustände aus periodischen Bahnen aufbaut, ist dies durchaus ein Problem. Zwar können für eine endlich breite Tunnelbarriere bei Einführung einer imaginären Wirkung periodische Lösungen im Inneren der Barriere gefunden werden (die Betonung liegt immer auf periodisch), die Schwierigkeit im vorliegenden Fall besteht jedoch darin, daß der klassisch verbotene Außenraum sich bis Unendlich erstreckt, d. h. komplexe Bahnen verschwinden dort nichtperiodisch im Unend-

[2]Klein ist strenggenommen nicht die Potentialänderung an sich, aber die Auswirkungen für die Zustände sind es, weil nur die evanescenten Ausläufer betroffen sind und vor allem durch die Größe der Kästen das relative Gewicht des Außenraumes sehr klein ist.

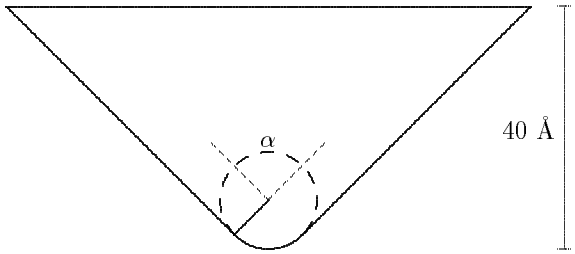


Abbildung 5.3: *Maßstabsgetreues Abbild einer Spitze mit dem Radius* 8 Å *und dem Flankenwinkel* $\alpha = 90^o$.

lichen. Ein denkbarer Ausweg wäre die periodische Wiederkehr über den unendlich fernen Punkt etwa im Sinne des Kriteriums (3.34), möglich vielleicht auch auf der Ebene der Propagatordarstellung[3] die Hinzuziehung von „Orbits" oberhalb V_0, die mit quantenmechanischen Ressonanzen korrespondieren könnten und ebenfalls über Unendlich laufen müßten. Ein möglicher Ausweg wären also die in 3.4 so bezeichneten verallgemeinerten Orbits[4].

Im übrigen gaben diese 80×50 Å^2 großen Rechtecke dann die Basis ab für die STM-Probengeometrien, die durch Einbringen von Stufen, Gräben, Ausstülpungen etc. in einer der 80 Å langen Kastenseiten entstanden.

5.5 Zustände in Spitzengeometrien

Als Grundform für die Spitzen diente ein (zweidimensionaler) gerader Kegel der Höhe 40 Å, der am Apex mit einen Radius abschließt. (Abb. 5.3). Radius R und Flankenwinkel α sind die zu variierenden Geometrieparameter. Betrachtet wurden Spitzen mit $\alpha = 60^0$, 75^0 90^0, 115^0, 120^o und $R = 0, 1, 4, 8, 20$ Å.

Für $R = 0$ entarten diese Spitzen zu Dreiecken und wären für $\alpha = 90^o$ bzw. $\alpha = 60^o$ exakt geometriegleich zu jenen in 5.2 aufgelisteten dreieckigen Kästen mit $(\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{4})$ bzw. $(\frac{\pi}{3}, \frac{\pi}{3}, \frac{\pi}{3})$, die analytisch behandelbar sind. (Letzteres träfe hier allerdings nicht zu, da unsere Kästen nur endlich hoch sind.)

Einen intuitiven Gesamteindruck vermitteln die 20 in Anhang D.1 en bloc gezeigten Zustände einer Spitze mit $R = 1$ Å und $\alpha = 90^o$. Mit fortlaufender Nummer wächst dort der energetische Abstand zur Besetzungsgrenze $E_F = 4$ eV. Die Zustände sind, wie eine Vergleichsrechnung zeigte, noch nahezu identisch mit denjenigen eines $(\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{4})$-Dreiecks – der kleine Radius führt noch zu keinen topologischen Änderungen.

Für $R = 4$ Å bleibt die Topologie der meisten Zustände erhalten, doch nehmen manche bereits neuartige, durch längere Knotenlinien gekennzeichnete Formen an (D.2).

Für $R = 8$ Å – die jeweiligen Übergangsbereiche wurde nicht studiert – finden sich solche Zustände dann verstärkt und ausgeprägter (D.3 a), doch ebenso besitzen einige nach wie vor noch eine sehr geordnete, fast ursprüngliche Topologie (D.3 (b)).

Im Bild der periodischen Bahnen käme man zu dem Schluß, daß die durch den Radius ermöglichten neuen Klassen periodischer Bahnen von einigen Zustände sehr leicht „einge-

[3] Zur einzelnen Energieeigenfunktion tragen im PO-Formalismus natürlich nur Orbits exakt dieser Energie bei; Orbits anderer Energien wären nur über den Propagator über ein geeignetes semiklassisches Ausschmieren einzubeziehen.

[4] Anhaltspunkt einer tieferen Analyse müßte sein, daß beim separablen Potentialgebilde (5.3) im Energiebereich $V_0 < E < 2V_0$ instabile reelle Orbits existieren, die zwischen den diagonalen Ecken hin und her laufen, sowie die vier „Superscars" in den Kanälen von jeweils der Kastenbreite, die sich in x- und y-Richtung je nach Unendlich erstrecken. Alle diese Orbits entstammen aber bereits dem Bereich des Kontinuums!

baut“ werden, während andere hiergegen eine bemerkenswerte Immunität besitzen.

D.4 zeigt schließlich einen Zustand in einer Spitze mit $\alpha = 120^o$ und $R = 1$ Å, der stark an ein Moire-Muster erinnert.

5.6 Zwei Quantenbillards im Tunnelkontakt

Unser numerisches Verfahren zur Berechnung der Wellenfunktionen gestattet prinzipiell beliebige Potentiale im Grundgebiet. Folglich lassen sich auch Gesamtwellenfunktionen für kombinierte Systeme bestimmen, die aus mehreren und nur durch eine kleine Tunnelbarriere voneinander getrennten Kästen bestehen (bzw. hierüber Kontakt haben). Wählt man die Kästen hinreichend groß, kommt die semiklassische Problematik der Quantenbillards zum Tragen und man hat es offenbar mit Quantenbillards im Tunnelkontakt zu tun.

Zu fragen wäre, inwieweit das semiklassische Konzept periodischer Bahnen noch tragen kann, denn wie schon in 5.4 müßten periodische Bahnen jetzt klassisch verbotenen Gebiete durch- oder überqueren. Als besondere „Provokation“ kann bei einem derartigen numerischen Experiment die Barriere auch beliebig niedrig gewählt werden, so daß die semiklassische Beschreibung zwar für das Kasteninnere adäquat sein sollte, nicht aber a priori unter der Barriere, wo ein solcher Zugang gewöhnlich hohe und breite Barrieren verlangt.

Als erste Annäherung an ein derartiges Mehrkastenproblem wird hier die Kombination aus einem Rechteck von 40×30 Å^2 und einem Kreis mit dem Radius 20 Å vorgestellt. Gewählt wurde eine bzgl. x spiegelsymmetrische Anordnung (vgl. die Bilder in Anhang D.5) und variiert wurde der y-Abstand d gemäß $d = 0, 1, 2, 5, 10$ Å. Wir bestimmten wieder Zustände mit Energien um 4 eV, setzen die Tiefe der Kästen diesmal jedoch nicht mit $V_0 = 8$ eV, sondern mit $V_0 = 5$ eV an (Barrierenhöhe damit nur 1 eV), um eine Kopplung über einen größeren Abstandsbereich hinweg zu ermöglichen. Mit Kreis und Rechteck wurden bewußt zwei Formen gewählt, deren Eigenzustände gut bekannt, ausgesprochen einfach und zudem signifikant verschieden sind, um Änderungen und eine gegenseitige Beeinflussung bereits vom Augenschein her feststellen zu können.

Einzeln sind die Billards (bis zur Energie $E \leq V_0$) klassisch integrabel, der Kreis ist es auch quantenmechanisch, während das Rechteck quantenmechanisch schwach nichtintegrabel ist (vgl. 5.4). In den Einzelbillards existieren keine Scars, da die Einzelsysteme nicht chaotisch sind.

Als Grundlage der Diskussion sind in Anhang D.5.1 20 Zustände des Systems bei einem Abstand der Kästen von $d = 2$ Å gezeigt. Diskutiert wird zunächst im Sinne einer Störungtheorie, aus der Sicht der Orginalzustände der einzelnen Kästen, wobei es dienlich ist, einzufügen, daß eine eindimensionale Rechteckbarriere gleicher Höhe ($V_0 = 5$ eV, $E = 4$ eV) und Breite (d) die folgenden exakten Durchlaßwahrscheinlichkeiten (2.2) zur Folge hätte:

d / Å	1	2	5	10
$D_{1\mathrm{d}}$	6.9×10^{-1}	3.0×10^{-1}	1.5×10^{-2}	9.1×10^{-5}

(5.6)

Die Durchlaßwahrscheinlichkeiten werden in unserem Falle geringer sein, da die Barriere außerhalb der Symmetrieachse breiter als d ist. (Auch $d = 0$ ist bei uns ein quantenmechanisches Tunnelproblem!)

Eine Kopplung über die Barriere hinweg wird begünstigt, wenn sich zu einem Originalzustand der einen Elektrode ein energetisch nahegelegener Partnerzustand in der Gege-

nelektrode findet. Weiterhin entscheidet die Form der Wellenfunktionen (vereinfacht: der y-Impuls in Richtung der Gegenelektrode) über die Stärke der Wechselwirkung. Exemplarisch hierfür zeigt der Zustand (i) keinerlei Intensität im Kreis (zumindest bei dieser Graustufenauflösung nicht), da der dominierende Rechteckzustand zum einen energetisch isoliert liegt und zum anderen aufgrund seiner minimalen y-Impulskomponente auch kaum Anteile im Kreis induzieren kann.

In den Bildern in D.5.1 lassen sich grob zwei Typen von Gesamtwellenfunktionen unterscheiden:

A Die Intensität ist auf beide Billards annähernd gleichverteilt.
Störungstheoretische Interpretation: Zur Energie des Gesamtzustandes finden sich Originalzustände in beiden Kästen. Die Gesamtwellenfunktion stellt sich dar als Linearkombination dieser Zustände.

B Die Intensität ist auf ein Billard konzentriert.
Störungstheoretische Interpretation: Zur Energie des Gesamtzustandes paßt nur ein Orginalzustand (oder auch mehrere bei Entartung) dieses einen Billards. Die Gesamtwellenfunktion besteht im wesentlichen aus diesem Zustand plus dessen „Ausläufern“ in die Gegenelektrode.

Die störungstheoretische Sicht ist natürlich nur partiell geeignet, da die Kopplung bei einem Abstand von $d = 2$ Å nach (5.6) nicht unbedingt als klein bezeichnet werden kann. Eine genauere Analyse müßte Niveauverschiebungen und Entartung berücksichtigen und auch die Ergebnisse einer Störungsrechnung einmal explizit dagegenstellen.

In dieser ersten Betrachtung richten wir das Augenmerk auf die besondere Form der „Ausläufer“ bei den Wellenfunktionen vom Typ B – die Bilder D.5.1 d, e, g, h, k, l, o, q–t. In den dort von der Gesamtwellenfunktion nur mit schwacher Intensität bedachten Kästen existiert offenbar kein energetisch passender Originalzustand – aus der Sicht des gegenüberliegenden Originalzustandes ist dieser Kasten gewissermaßen „leer“ – und es erhebt sich die Frage, welche Gestalt die Wellenfunktion in diesen Gebieten annimnt.

Nun, die Bilder legen hier den Schluß nahe, daß sich in diesen Fällen Scars herausbilden und zwar entlang solcher periodischer Bahnen, die aus der Barriere austreten und in diese zurückkehren. Dabei entscheidet die Gestalt der mit dem Hauptgewicht versehenen – der *dominanten* – Teilwellenfunktion in der Gegenelektrode, welche periodischen Bahnen zum Tragen kommen, konkret, unter welchen Winkeln diese die Barriere verlassen. Beispielsweise zeigt Bild (f) im Kreis annähernd einen drehimpulslosen s-Zustand und anschaulich dann auch ganz folgerichtig im Rechteck einen Scar, der senkrecht zur Barriere steht. Die größeren Drehimpulse in (k) und (o) führen zu schräger auslaufenden Scars, die in den diagonalen Ecken des Rechtecks reflektiert werden, und der größte Drehimpuls in Bild (s) schließlich geht mit einem bereits an den Seitenwänden des reflektierenden Scar einher. Die Scars im Kreis (und der dominanten Wellenfunktion im Rechteck) sind weniger markant, jedoch erkennt man unschwer in den Bildern (d) und (g) dem Kreis eingeschriebene Siebenecke, d. h. Bahnen, die siebenmal am Umfang reflektieren, und in Bild (e) sowohl Fünf- als auch Sechsecke.

Was ist hieran bemerkenswert? Erklärungen für Scars finden sich in der Literatur bisher stets an instabile klassische Orbits gekoppelt, an Systeme also, die bereits klassisch nichtintegrabel sind, an periodische Bahnen, um diesen Punkt herauszustreichen, die „Kenntnis“ vom Gesamtsystem und dessen Nichtintegrabilität besitzen. Solche globalen Orbits lassen sich im vorliegenden Fall nur erhalten, wenn der Begriff der periodischen Bahn (und der

der Stabilitätsmatrix) entweder auf komplexe Bahnen (komplexe Wirkungen) verallgemeinert wird, die die Barriere durchqueren können, oder auf solche klassisch-reellwertigen des Kontinuums mit Energien $E > V_0$, die als ungebundene Orbits über den unendlich fernen Punkt laufen. Da die Antwort gegenwärtig offen ist, sprechen wir wie schon in 3.4 und 5.4 wieder von verallgemeinerten Orbits. Als Schwierigkeit tritt hinzu, daß eine konsistente Beschreibung hier auch den Einfluß der dominanten Teilwellenfunktion erfassem müßte, die ja oben offensichtlich die Richtung der Scars bestimmt. Im Grunde wäre die vollständige Rückführung der quantenmechanischen Resultate auf verallgemeinerte Orbits zu leisten.

Eine Hypothese darüber, ob die „Ausläufer" der Wellenfunktionen vom Typ B *grundsätzlich* die Scar-Gestalt annehmen, ist anhand der wenigen Beispiele nicht gerechtfertigt. Daß eine Scar-Gestalt überhaupt zustandekommt und welche periodische Bahn dann besetzt wird, könnte immerhin so verstanden werden: Die Originalzustände der Einzelelektroden werden durch für sie jeweils spezifische Bahnengemische (aus verallgemeinerten Orbits) konstituiert, in denen bestimmte Bahnen dominieren. Bei Annäherung der Elektroden koppeln diese Bahnen bevorzugt an ihnen ähnliche Orbits des Gesamtsystems, die ihrerseits in der Summe wieder die Gestalt der Gesamtwellenfunktion ausmachen. Bei Wellenfunktionen vom Typ B okkupieren die so angeregten Gesamtorbits in der Gegenelektrode ein vormals „leeres" Gebiet und sind in diesem – da allein – auch trotz ihrer geringen Intensität sichtbar; bei Wellenfunktionen vom Typ A entstehen dagegen in beiden Elektroden Bahnengemische, in denen im wesentlichen noch die Orginalzustände dominieren.

Der Sprachgebrauch ist offensichtlich der einer Störungstheorie periodischer Bahnen, wie ja die ganze Argumentation im Geiste eigentlich das Bild jener in 3.4 hypothetisch aufgeworfenen Quantenmechanik verallgemeinerter Orbits schon in sich trägt.

Zwei bemerkenswerte Zustände seien noch aufgeführt:

D.5.2 zeigt bei einem Abstand $d = 0$ einen in beiden Elektroden voll ausgebildeten Scar-Zustand – einen tunnelnden Scar, denn auch $d = 0$ ist hier ein Tunnelproblem.

In D.5.3 (a) ist links für den Abstand $d = 10$ Å ein Gesamtzustand zu sehen, bei dem die Intensität fast ganz im Kreis und zwar in einem s-Zustand konzentriert ist – ein Spitzenzustand getreu dem STM-Modell von Tersoff/Hamann [78]. Die Intensitätsverteilung im Rechteck bleibt bei dieser Auflösung unsichtbar. Rechts wurden für beide Innenräumen und für den Außenraum jeweils separate Graustufenskalen verwandt. (Der Kreiszustand erscheint dadurch etwas größer.) Im Rechteck wird nunmehr der „Ausläufer" des drehimpulslosen s-Zustandes sichtbar – ein Scar in Richtung Kreismittelpunkt!

Zum Abschluß sei nochmals betont, daß die annähernd ideale Spiegelsymmetrie bzgl. der x-Achse in den Bildern stets das Resultat der numerischen (iterativen) Lösung des Eigenwertproblems war, wobei der Iterationsprozeß generell mit zufallsbelegten Startwellenfunktionen begonnen wurde, d. h. numerische Artefakte auf dieser, der „sichtbaren" Ebene können ausgeschlossen werden.

5.7 Resümee und Ausblick

Einen Zugang zu den Phänomenen in nichtintegrablen semiklassischen Quantensystemen bietet allein die Theorie periodischer Orbits. Dabei ist die Rolle der Integrabilität bzgl. ihres Einflüsses auf die statistische Verteilung der Eigenwerte bisher besser verstanden als die Auswirkungen auf die Gestalt der einzelnen Wellenfunktionen. Immerhin können als offenkundigster Ausdruck eines chaotischen Systems Scar-Zustände angesehen werden, die

mit instabilen Orbits korrespondieren. Anzufügen, daß die Theorie sich gegenwärtig nur gewöhnlich-klassischer Orbits (gebunden und reell) bedient.

In 5.4 wurde nun anhand des nichtseparablen Rechtecks endlicher Wandhöhe V_0 der Übergang von einem integrablen zu einem schwach nichtintegrablen System studiert, wobei die Besonderheit darin bestand, daß die Zerstörung der Integrabilität in einem für $E < V_0$ nur quantenmechanisch zugänglichen Raumbereich erfolgte. Bemerkenswerterweise besaßen bei einem Teil der Zustände die hierdurch hervorgerufenen Änderungen (nicht die Zustände selbst!) in den Wahrscheinlichkeitsdichten das Aussehen von Scars. Nimmt man den nach bisherigem Verständis zwingenden Zusammenhang zwischen Scars und einem chaotischen System als Maß, läßt dieses Ergebnis die Sicht zu, daß der integrabilitätszerstörende („chaotische") Charakter der Potentialänderung hier einen direkten Ausdruck in den Änderungen der Wellenfunktionen findet. (Eine weiterführende Frage wäre, ob dies auch in der Eigenwertstatistik zu Tage tritt[5].) Eine Beschreibung in einer Theorie periodischer Bahnen müßte diese Differenzen als Differenzen von Orbits darstellen. Dabei wären in jedem Falle jene Orbits mit $V_0 < E < 2V_0$ heranzuziehen, die dort aber bereits dem Bereich des Kontinuums(!) entstammen. Kleine Störungen, wie sie in der Monodromie-Matrix (3.25) betrachtet werden, führen zu ungebundenen Bahnen. Die grundsätzliche Einbeziehung komplexer oder ungebundener Orbits, kurz, verallgmeinerter Orbits, liegt hier, vorsichtig gesprochen, nahe.

In 5.6 – zwei Quantenbillards im Tunnelkontakt – fanden sich dann in Kästen, die für sich genommen integrabel bzw. nur schwach nichtintegrabel wären, Scars schließlich auch direkt in den Zuständen (nicht nur in deren Änderungen), und zwar entlang solcher Trajektorien, die aus einem Kasten in den anderen laufen, die Tunnelbarriere folglich durch- oder überqueren müssen. Hier nun wäre eine Theorie verallgemeinerter Orbits endgültig gefordert. Die Tatsache, daß die Scars ihrer Richtung nach stets als „Ausläufer" einer in diesen Fällen dominanten Teilwellenfunktion der Gegenelektrode erscheinen – wenn die Scar-Elektrode keinen energetisch passenden Originalzustand besitzt, „leer" ist –, würde plausibel in einer Art Störungstheorie periodischer Bahnen: *Wellenfunktionen sind jeweils spezifische Bahnengemische verallgemeinerter Orbits, deren einzelne Orbits bei einer Annäherung der Kästen bevorzugt an ihnen ähnliche Orbits des Gesamtsystems (der Gesamtwellenfunktion) koppeln.*

Hat man es also mit einem bevorzugten Auskoppeln einzelner Bahnen oder Bahnensätze aus einem Billard B1, die in Billard B2 dann als Scar sichtbar werden, zu tun? Würde hier B2 um B1 herumgeführt – numerisch oder sogar experimentell (siehe unten) denkbar –, könnte der Bahnengehalt eines einzelnen Originalzustandes Ψ_n mit der Energie E_n aus B1 sogar gemessen werden. Wie kritisch ist die Form von B2? Klarerweise darf es dort keinen eigenen Zustand bei E_n geben, um den Scar nicht zu überdecken. Genügt aber, daß der auszukoppelnde Scar im wesentlichen zur Auskoppelstelle zurücklaufen kann (ist also nur die engere Umgebung des Scars wichtig), oder sind in B2 die dem Scar entfernteren Gebiete gleichrelevant? (Immerhin muß auch ein Scar als Lösung der SGL sämtliche Randbedingungen erfüllen.)

Weitere numerische Untersuchungen bieten sich an: Eigenwertstatistik, Bewegung der Eigenwerte bei Bewegung der Kästen gegeneinander (Niveaukreuzungen?) und Frage nach

[5] Die Nächste-Nachbar-Statistik eines semiklassisch großen nichtseparablen Rechtecks dürfte sich nur unwesentlich von einer Poisson-Verteilung unterscheiden, möglicherweise offenbart aber eine Statistik der Eigenwert*änderungen* gegenüber dem separablen System (5.3) Anzeichen des vollen Chaos (Wigner-Verteilung?)

den hiermit einhergehenden Verschiebungen in den Gesamtwellenfunktionen bzgl. deren Verteilung auf die einzelnen Kästen; Differenzen dabei jeweils zu den Originalzuständen. Vergleich mit den expliziten Ergebnissen einer Störungsrechnung. Untersuchung von Kästen mit identischen Spektren, z. B. zwei gleichgroße, gegeneinander verdrehte Rechtecke (geraten bei bestimmten Orientierungen identische Bahnen über die Tunnelbarriere hinweg in Resonanz?).

Auch diesbezügliche Experimente mit zweidimensionalen Mikrowellenbillards [76, 75, 74] sind vorstellbar, z. B. indem zwei Billards über einen schmalen Kanal, in dem bei der betrachten Energie nur evanescente Lösungen möglich sind (Tunnelbarriere), gekoppelt werden. Die Situation entspricht nicht ganz der obigen, da hier wieder unendlich hohe Wände vorliegen würden. Nichtsdestotrotz, man wähle insbesondere zwei integrable Billards ohne Scar-Zustände und rege im ersten einen Zustand Ψ_n bestimmter Energie E_n an (der dominante). Sind die Abmessungen des zweiten Billards dergestalt, daß sich bei E_n dort kein Originalzustand befindet, sollte man bei bestimmten Stellungen der Billards zueinander (Winkel zwischen Kopplungskanal und den Bahnen) im zweiten Billard – einem an sich integrablen(!) – Scar-Zustände finden können, die aus dem Kanal austreten und in diesen zurücklaufen. Eine vorherige Simulation, z. B. mit dem hier verwandten Verfahren, sollte eine günstige Anordnung vorgeben können.

Kapitel 6

Konstantstromprofile

6.1 Vorbemerkungen

Der Wahl einer zu untersuchenden zweidimensionalen Spitze-Probe-Geometriekonstellation sind im verwendeten Modell fast keine Grenzen gesetzt. Wir konzentrieren uns hier auf drei idealtypische Probenformen: einen Graben, eine Kerbe und eine dreieckige Spitze (Dreispitz). Vorgestellt werden Simulationsrechnungen zur Abbildung dieser Strukturen im Modus konstanten Tunnelstroms sowie die geometrische Entfaltung der STM-Profile mit der aktuellen Spitzenform nach dem in Abschnitt 1.3.3 erläuterten Verfahren von Reiss [68]. Die Diskussion konzentriert sich dann auf diese beiden Punkte. Einige Ergebnisse bzgl. des Grabens finden sich bereits in [72].

Als Spitzengeometrie diente die in Abb. 5.3 gezeigte (zweidimensionale) Kegelform mit aufgesetztem Radius, wobei Radien von $R = 1, 4, 8, 20$ Å und Flankenwinkel von $\alpha = 75^o$, 90^o, 105^o, 120^o in Betracht gezogene Parameter waren.

Für die Berechnung des Tunnelstroms zwischen zwei voreinander plazierten Elektroden im Transfer-Hamiltonian-Formalismus wurden jeweils die 10 der Fermi-Energie $E_F = 4$ eV nächstbenachbarten Zustände aus beiden Elektroden herangezogen. Mit diesen 10 Zuständen berechneten wir auch die Linien konstanter lokaler Zustandsdichte (LDOS) bei E_F oberhalb der Probenstrukturen (Abb. 6.1), die nach dem Standardmodell von Tersoff/Hamann [78] den Konturen konstanten Tunnelstroms gleichzusetzen wären.

6.2 Abbildung eines Grabens

Ein Graben von 20 Å Breite und 10 Å Tiefe, eingelassen in die längere Seite eines 80×50 Å^2 großen Rechtecks, stellt die abzubildende Probenstruktur dar. Die Linien konstanter LDOS in Abb. 6.1 geben die Grabenform gut zu erkennen, auch z. B. im Vergleich mit denjenigen der Kerbe, gleichwohl insbesondere die scheinbare Grabenbreite mit wachsendem Abstand zur Oberfläche deutlich abnimmt.

Die Abbildungen 6.2 und 6.3 zeigen dann im jeweils oberen Teil berechnete Konstantstromprofile und zwar für drei unterschiedliche 90^o-Spitzen mit den Radien $R = 1, 8$ und 20 Å. Diese Konturen reflektieren kaum mehr die Grabenform, sondern sind weitgehend dominiert von der Geometrie der Spitze. Unterstrichen wird dies durch deren große Ähnlichkeit mit den zusätzlich eingetragenen Abtastkurven (gepunktet), die die Profile zeigen,

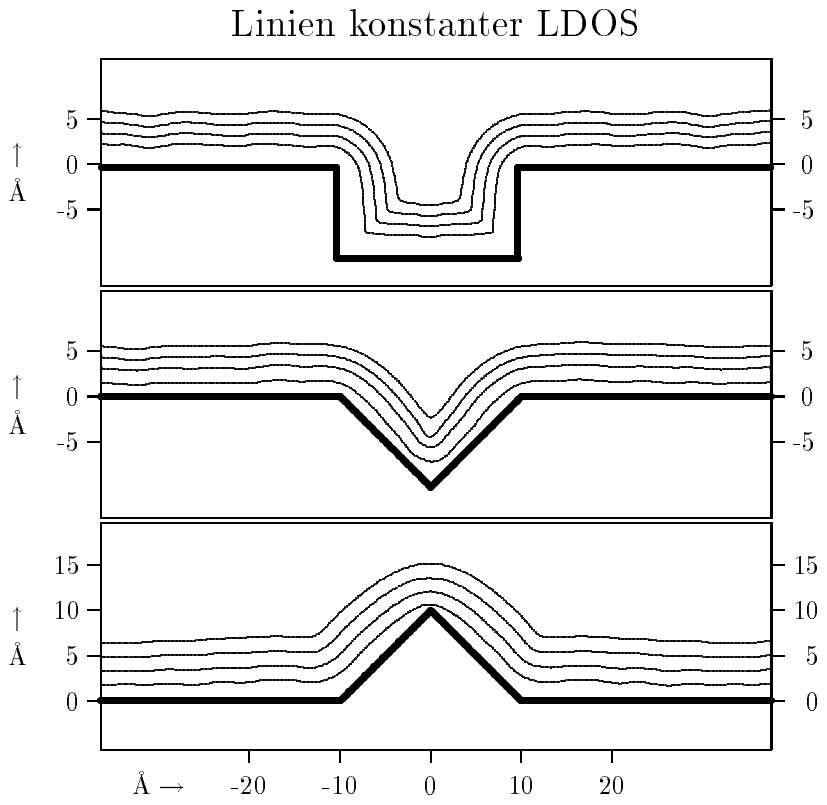


Abbildung 6.1: *Linien konstanter lokaler Zustandsdichte bei $E_F = 4$ eV über einem Graben, einer Kerbe und einem Dreispitz, jeweils* 20 Å *breit und* 10 Å *tief bzw. hoch. Mit diesen Formen wurde jeweils die lange Seite eines* 80×50 Å^2 *großen Rechtecks modelliert.*

wie sie bei rein mechanischer Berührung, im Falle des geometrischen Kontakts zwischen Probe und Spitze, entstehen würden. Auch die rapide Abnahme der Tiefe der STM-Profile mit wachsendem Spitzenradius ist vergleichbar mit der diesbezüglichen bei den Berührungskurven.

Alle STM-Profile wurden anschließend nach Reiss [68] mit der aktuellen Spitzengeometrie entfaltet. Die daraus resultierenden Kurven, die wir hier einfach die Reissschen Kurven nennen wollen, sind je darunter dargestellt. Diese Entfaltung betrachtet eine STM-Abbildung als ein rein mechanisches Aufeinanderabgleiten von Probe und Spitze unter punktförmiger Berührung, d. h. wäre eine Konstantstrom-Abbildung tatsächlich nur ein solches mechanisches Abtasten, müßten die dergestalt zurückgerechneten Kurven die Probenkontur an all jenen Punkten exakt reproduzieren, an denen ein geometrischer Kontakt mit dem Spitzenkörper hätte stattfinden können. Differenzen zur wirklichen Probengeometrie geben somit Aufschluß über die Unterschiede zwischen einem mechanischen Abtasten und einer echten (quantenmechanischen) Konstantstrom-Abbildung. Um den Effekt der Entfaltungprozedur zu verdeutlichen, wurden für jeden tatsächlich vorliegenden – jeden tatsächlich berechneten – Punkt eines STM-Profils (gelegen auf dem Diskretisierungsgitter im Raster der Schrittweite 0.4 Å) dessen neue, entfaltete Position nur durch einen einzelnen, kräftigeren Punkt markiert und nicht wie bei den STM-Profilen durch stetige Kurvenzüge verbunden. Häufungspunkte und ebenso an relevanten Daten arme Kurvenabschnitte blei-

ben besser sichtbar.

Voraussetzung für eine eindeutige Entfaltung ist allerdings, daß die Lage des Berührungspunktes auf der Spitze aus dem Anstieg des STM-Profils allein eindeutig bestimmt werden kann, was in zwei Dimensionen in der Regel gesichert ist bei Spitzenformen, die im Kontaktbereich einen stetig sich ändernden Anstieg aufweisen. Bei unseren Spitzen ist dies erfüllt für alle Berührungspunkte entlang des Radius, nicht jedoch für solche, die dem Spitzenschaft zuzuordnen sind, auf dem der Anstieg konstant bleibt. Alle derartige Punkte haben wir im gegebenen Fall dann vereinfachend dem Übergangspunkt 'Radius–Schaft' zugewiesen. (Im Hinblick auf eine solche Entfaltung ist ein kegelförmiges Spitzenmodell nicht ideal.) Das ist auch der Grund, warum in Abb. 6.2 die Reisschen Kurven selbst für die mechanische Berührung nicht strikt an den Oberkanten des Grabens enden – im Gegensatz zu Abb. 6.3, wo kein Schaftkontakt erfolgt und alle Berührungspunkte eindeutig zuordbar bleiben –, sehr wohl die Spitzen auch in Abb. 6.2 bei einem Abgleiten nie das Grabeninnere, sondern nur die oberen Kanten berühren. Für $R = 1$ Å in Abb. 6.2 bleibt die Entfaltungsprozedur dadurch auch fast ohne Nutzen, da in der Nähe der interessierenden Grabenflanken nahezu alle Profilpunkte mit einer Schaftberührung korrespondieren.

Nicht so für $R = 8$ Å und $R = 20$ Å. Die meisten STM-Profilpunkte aus dem Grabeninneren werden dort in Richtung der oberen Grabenränder befördert. Die laterale Position der Ränder, aus denen die Grabenbreite hervorgeht, kann überraschend genau anhand dieser Häufungszonen – wir sprechen nachfolgend etwas vereinfachend von *Häufungspunkten* – rekonstruiert werden, selbst (oder eher gerade) für das sehr ausgeschmierte Profil in Abb. 6.3, was zuallererst natürlich auf die Ähnlichkeit zwischen den STM- und den Abtastprofilen zurückverweist.

Zugleich sind einige charakteristische Eigenarten der Reisschen Kurven zu vermerken:

In der Grabenregion verbleiben zwar nur wenige Punkte, diese aber systematisch. Für $R = 20$ Å in Abb. 6.3, wo die gesamte Entfaltung eindeutig möglich ist, korrelieren diese Restpunkte genähert mit dem Spitzenradius, was die durchgezogene Linie unterstreicht. Für $R = 8$ Å sind die aus einer Schaftberührung herkommenden Reisschen Punkte natürlich auszuklammern (zu erkennen daran, daß sie parallel zum Schaft verlaufen), bei den eindeutig entfalteten, die sich den obigen zur Grabenmitte hin anschließen, zeigt sich diese Korrelation aber ebenfalls. Geliefert werden diese Restpunkte stets aus der unmittelbaren Umgebung der Minima der STM-Profile.

Zweitens, die laterale Position der Häufungspunkte skaliert geringfügig mit der Höhe der STM-Profile über der Probenoberfläche und zwar dergestalt, daß die scheinbare Grabenbreite mit wachsender Scanhöhe abnimmt (für $R = 8$ Å stärker, für $R = 20$ Å schwächer). Ein Blick auf die LDOS-Konturen in Abb. 6.1, wo dieses laterale „Zulaufen" der Probenstruktur mit wachsender Höhe noch ausgeprägter vorliegt, verrät die Herkunft dieses Effektes. Die hier interessante Frage ist eigentlich, warum in den Reisschen Kurven für die stumpfere Spitze ($R = 20$ Å) dieses Skalierungsverhalten der LDOS am wenigsten zum Tragen kommt, konkret, warum die Häufungspunkte für $R = 20$ Å die lateralen Positionen der Grabenkanten fast schon unabhängig von der Scanhöhe *gut* wiederspiegeln, während dies für $R = 8$ Å sichtlich weniger der Fall ist (und am wenigsten für die LDOS-Konturen).

Offenbar hiermit im Zusammenhang steht eine ähnliche Tendenz bei den Konstantstromprofilen: Mit wachsendem Spitzenradius gewinnen diese an *relativer* Schärfe – verglichen nämlich mit den Abtastprofilen: Während in Abb. 6.2 für $R = 1$ Å und $R = 8$ Å die STM-Profile nahezu parallel mit den Berührungskurven verlaufen, zeigen diese in Abb. 6.3

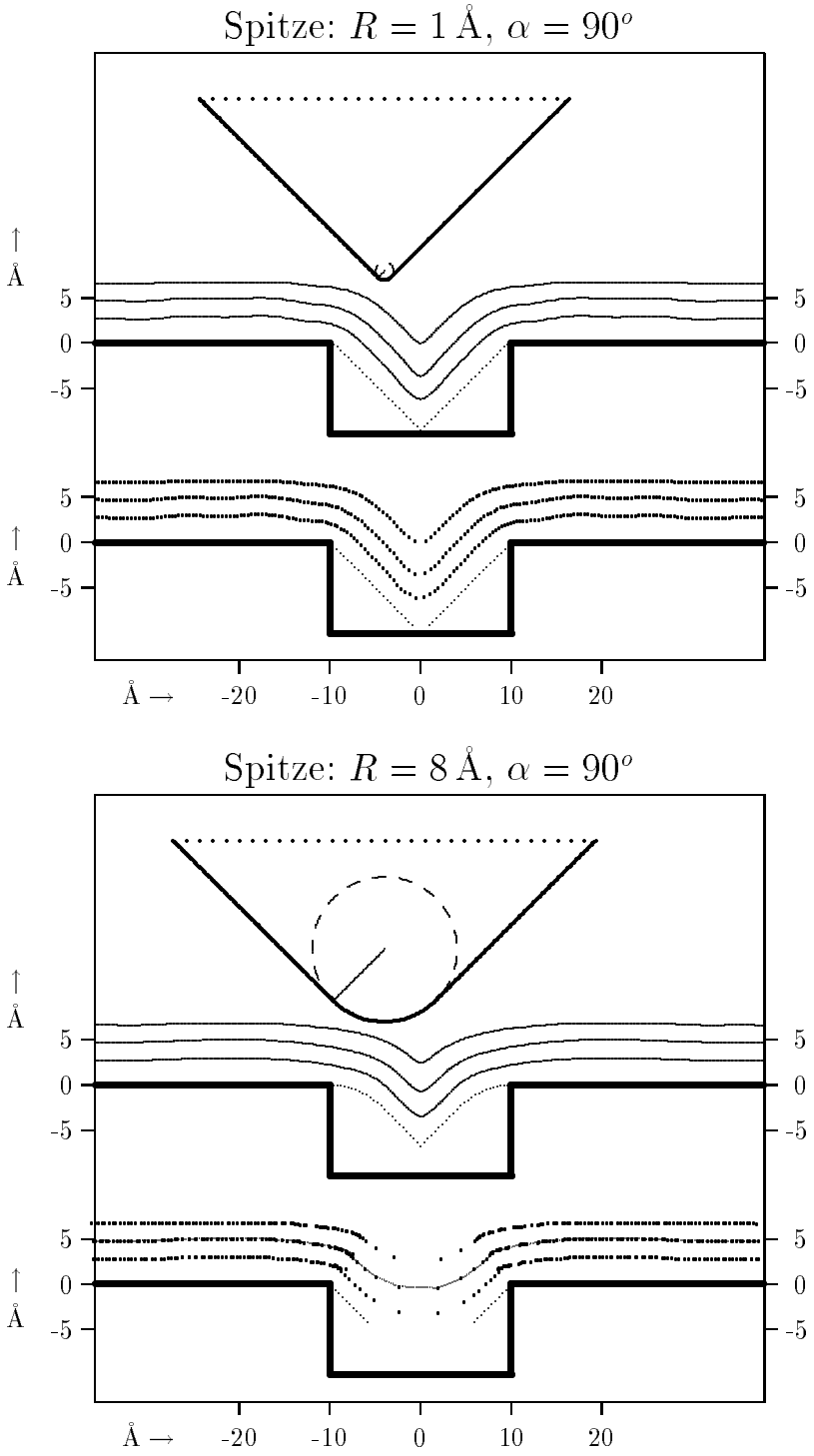


Abbildung 6.2: *Jeweils oben die Konturen konstanten Tunnelstroms für Spitzen mit* $\alpha = 90^o$ *und* $R = 1$ Å *bzw.* $R = 8$ Å. *Gepunktet die Kurven mechanischer Berührung. Jeweils darunter die nach* Reiss *[68] mit der Spitzengeometrie entfalteten Profile.*

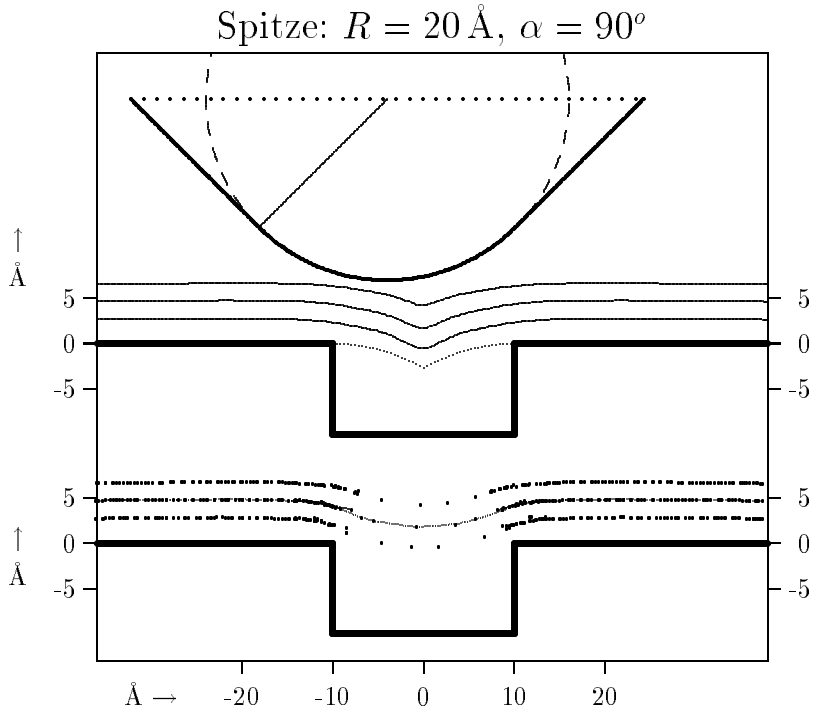


Abbildung 6.3: *Wie Abb. 6.2, jedoch für eine Spitze mit* $R = 20$ Å.

für $R = 20$ Å eine klar tiefere Einprägung als das Abtastprofil.

Qualitativ verstehen läßt sich dieser Effekt mit Hilfe eines Arguments, das die absoluten Tunnelflächen in Rechnung stellt:

Man betrachte die Abbildung einer rechtwinkligen, parallel zur y-Achse verlaufenden Kante $K(x, z)$, die bei $x = 0$ senkrecht abfällt; als Kastenpotential geschrieben: $K(x, z) = V_0\,\Theta(x)\,\Theta(z)$. Abgebildet werde durch zwei Spitzen $\mathcal{S}_1$ und $\mathcal{S}_2$ mit gleicher y-Ausdehnung (z. B. auch zweidimensional). In einer gemeinsamen Ausgangshöhe z_0 über dem planaren oberen Probengebiet fließe der zu jeder Spitze gehörige Strom über einen effektiven Strompfad, der durch die Querschnittsfläche A_1 bzw. A_2 charakterisiert sei. Die Pfade werden dort im wesentlichen senkrecht, über den untersten, der Probenebene zugewandten Apexbereich verlaufen. Die Größe der Querschnitte ist bestimmt durch die Form der Spitzen, $\mathcal{S}_2$ sei die stumpfere: $A_2 > A_1$. Beim Umfahren der Kante wandert der Strompfad am Schaft einer Spitze nach oben und auf seiten der Probe hin zur Kante. Naturgemäß wird $\mathcal{S}_2$ ein flacheres Profil geben. Erweist sich die Kante aber als schärfer als die Spitze selbst, wird in allen Positionen, in denen der Strompfad über die Kante läuft, dessen Querschnitt nicht mehr von der Spitze, sondern von der Probe beschränkt. Gegenüber der Ausgangsstellung wird dieser damit eingeschnürt und zwar relativ um so stärker, je stumpfer eine Spitze ist. Um dies auszugleichen, muß die stumpfere Spitze *relativ* gesehen tiefer eintauchen als die schärfere.

In gewisser Näherung kann das Konstanthalten des Tunnelstroms entlang eines STM-Profils also zusammengesetzt gedacht werden aus dem Konstanthalten der Länge des Strompfades als dominierendem Effekt und, diesem überlagert, als Effekt zweiter Ordnung, dem Konstanthalten einer absoluten effektiven Tunnelfläche A_{eff}, dem Querschnitt des Strom-

pfades.

Der springende Punkt ist hier der Gegensatz zum geometrischen (mechanischen) Abtasten: Die Spitzenposition spiegelt dort nämlich nicht *auch* die Konstanz einer „Kontaktfläche“ A_{eff}, sondern lediglich die Verhältnisse bzgl. eines mathematischen Punktes; ob die Berührung punktförmig oder flächenhaft erfolgt, bleibt ohne Belang. Die gewissermaßen große „Berührungsfläche“ über dem planaren Probenabschnitt erzwingt keine ebensolche in Stellungen gegenüber den scharfen Kanten, die Spitze gleitet schwerelos, ohne die „Last“ A_{eff}. Die Folge ist, daß bei unseren Beispielen die stumpfere Spitze im Konstantstrommodus tatsächlich tiefer in den Graben eintaucht als die schärfere – „tiefer“ natürlich nur relativ – im Vergleich, wie oben bemerkt, mit den Kurven für die mechanische Berührung.

Dieser Unterschied kommt nun zum Tragen, wenn im Reissschen Verfahren die STM-Profile wie Abtastprofile entfaltet werden. Bei der betrachteten Konstellation aus 20×10 Å^2 Graben und 90^o-Spitzen gewinnen die STM-Profile generell an Schärfe gegenüber den Abtastkurven, je stumpfer die Spitzen werden, und die Entfaltung kompensiert (gegebenefalls auch überkompensiert) hierdann in gewissem Umfang wieder das Verwässern der Grabenstruktur in der LDOS; für $R = 20$ Å im Beispiel offensichtlich soweit, daß die Häufungspunkte fast unabhängig von der Scanhöhe über den Grabenkanten zu liegen kommen.

6.3 Abbildung einer Kerbe

Anstelle des Grabens nun eine Kerbe von 10 Å Tiefe und 20 Å Breite, also 90^o Öffnungswinkel. Abb. 6.1 zeigt die Linien konstanter LDOS bei E_F und Abb. 6.4 die Profile konstanten Tunnelstroms für zwei 90^o-Spitzen mit $R = 8$ Å und $R = 20$ Å. Die Abtastkurven sind identisch mit denen des Grabens. Im Vergleich zu diesen zeigen sich dann auch die wesentlichen Änderungen:

Für $R = 8$ Å verlaufen die STM-Profile diesmal flacher als die Abtastkurve und erst für $R = 20$ Å genähert parallel. Die oben diskutierte Tendenz einer zunehmenden (relativen) Schärfe der Konstantstromprofile gegenüber den Abtastkurven mit stumpfer werdenden Spitzen ist hier also ebenfalls gegenwärtig, jedoch infolge der geringeren Einschnürung des Strompfades an den schräg (und nicht senkrecht) abfallenden Kerbflanken dahingehend modifiziert, daß für kleinere Radien die STM-Profile noch weniger scharf sind als die Berührungskurven und erst für $R > 20$ Å tiefere Einprägungen als diese an einer 45^0-Schräge hinterlassen würden (in diesem Beispiel aufgrund der dicht folgenden aufsteigenden Kerbflanke aber kaum werden[1]).

Dementsprechend skaliert die laterale Position der Häufungspunkte (Lage der Kerbränder?) mit der Höhe für $R = 20$ Å nicht weniger ausgeprägt als z. B. die scheinbare Kerbbreite in den LDOS-Konturen und für $R = 8$ Å sogar stärker.

Für die in der Kerbregion verbliebenen Reissschen Punkte ist indes auch hier wiederum eine Korrelation mit dem Spitzenradius zu verzeichnen.

[1] Ein hier sicheres Beispiel wäre eine Quaderspitze mit größerer Breite als die Kerbe: im Abtastprofil keine Einprägung, im Konstantstromprofil immer eine nichtverschwindende; $R \to \infty$ ist dagegen etwas pathologisch, weil auch die Tunnelfläche dann gegen Unendlich strebt.

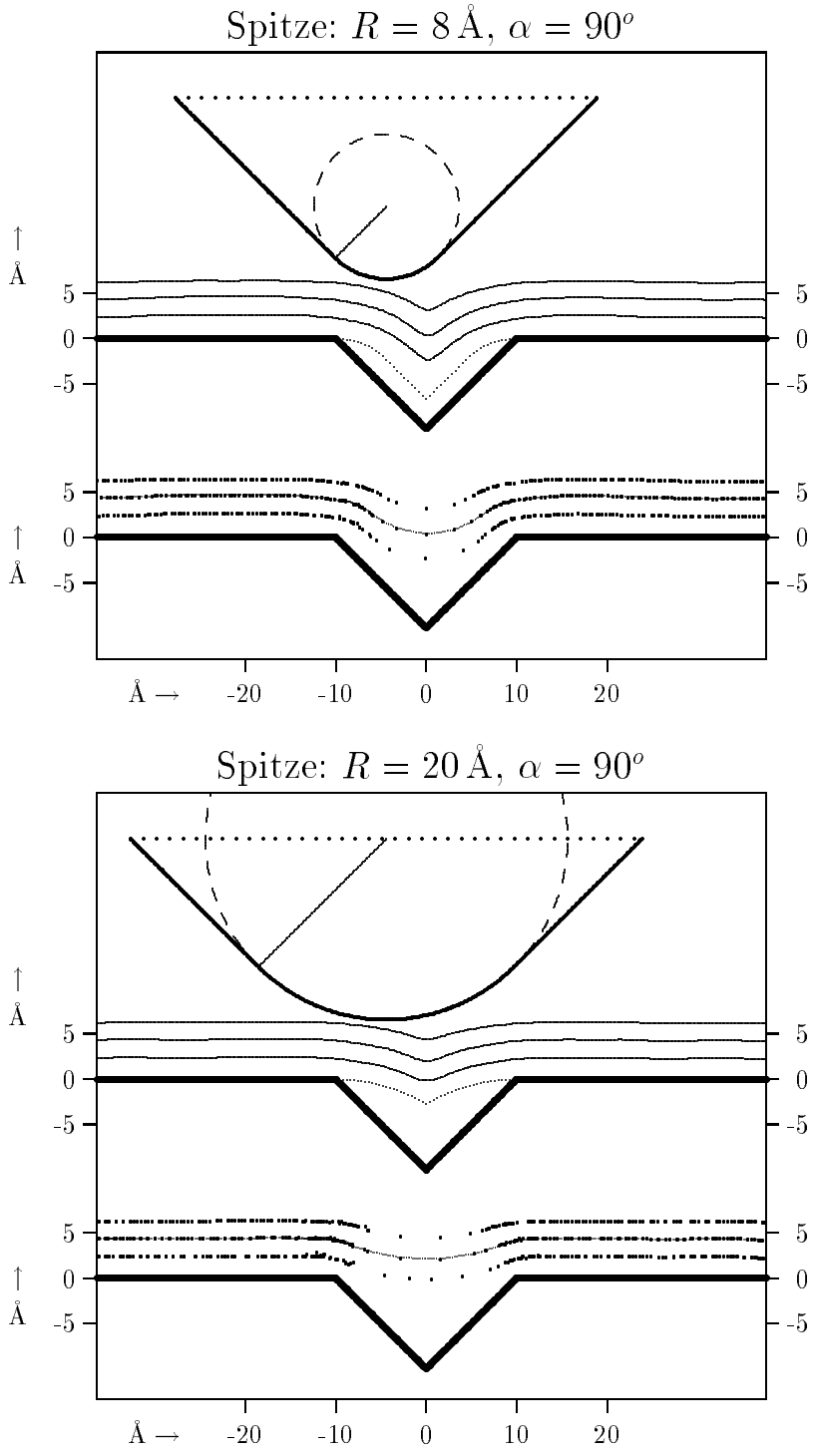


Abbildung 6.4: *Konstantstromprofile über einer Kerbe für die eingezeichneten Spitzen mit* $\alpha = 90^o$ *und* $R = 8$, Å *bzw.* $R = 20$ Å. *Darunter die mit der Spitzengeometrie entfalteten Profile.*

6.4 Abbildung eines Dreispitz

Nun das zur Kerbe inverse Problem – eine dreieckige Spitze (Dreispitz) von 10 Å Tiefe und 20 Å Fußbreite (Abb. 6.5). Man bemerke, daß unmittelbar über dem Dreispitz die Profile für $R = 20$ Å nicht wie für $R = 8$ Å oder für die LDOS grundsätzlich konvex verlaufen (von oben gesehen), sondern streckenweise flach und sogar leicht konkav. Hier greift wieder das Argument der absoluten Tunnelflächen: direkt über dem Dreispitz entsteht für die $R = 20$ Å Spitze eine wesentlich größere relative Einschnürung des Strompfades, während am Fuße, seitlich daneben, eine Aufweitung erfolgt, was in diesem Falle einen qualitativ anderen Kurvenverlauf zur Folge hat (Vorzeichenwechsel des Anstiegs).

In den Reisschen Kurven schlägt dies zu Buche: jene für $R = 8$ Å bleiben moderat, nähern sich den LDOS-Konturen aus Abb.6.1; bei $R = 20$ Å verursachen die Vorzeichenwechsel im STM-Anstieg direkt über dem Dreispitz einige Konfusion.

Die Lücken in den Reisschen Kurven geben Auskunft, daß aus den Regionen am Fuße des Dreispitz wenig Informationen in das STM-Profil eingeflossen sind. Auch hier verbleiben in den Lücken wieder Restpunkte, die genähert mit den Spitzenradien korrelieren. Zugleich wird deutlich, daß man die gering einbezogenen – die „unsichtbaren“ – Probenregionen nicht vertikal (in z-Richtung) unter den Lücken anzusiedeln hat, sondern eher senkrecht zur Kurventangente (Tangente an die Kurve der Restpunkte).

6.5 Diskussion

Bezüglich der konkreten numerischen Resultate ist folgendes festzuhalten:

1. Betrachtet wurden Spitze-Probe-Paarungen, bei denen die Krümmungsradien in der LDOS beider Elektroden vergleichbar waren, aufgrund der scharfen Kanten zum Teil sogar für die Probe kleiner. Bei allen drei Paarungen existierten größere Probenabschnitte, die von der Spitze infolge geometrischer Restriktionen nicht berührt werden konnten – mechanisch „unsichtbare“ Gebiete. (Den Zusatz *mechanisch* lassen wir im folgenden weg.)

2. In erster Ordnung erweisen die Konstantstromprofile sich als sehr ähnlich den geometrischen Abtastkurven (dominiert von der Spitzenform).

3. Tatsächlich gewinnen die Konstantstromprofile gegenüber den Abtastkurven an Schärfe (an Tiefe), je stumpfer die Spitzen werden. Qualitativ verstehen läßt sich dieser Fakt mit Hilfe des Arguments der absoluten Tunnelfläche. Dabei hängt es von der Probenkontur ab – z. B. senkrecht abfallend oder im Winkel von 45^o –, ab welchem Spitzenradius Konstantstromprofile auch absolut schärfer sind als diese. Beim betrachteten Graben von 20×10 Å^2 und (den 90^o-Spitzen) quasi für alle Radien $R > 0$, bei der Kerbe, besser, an einer 45^0-Schräge, erst für $R > 20$ Å.

4. Häufungspunkte in den Reisschen Kurven kennzeichnen bei Abtastprofilen den Rand eines „unsichtbaren“ Gebietes und kommen auch exakt auf diesem zu liegen. Bei den Konstantstromprofilen variierte (u. a. als Folge von Punkt 3) die laterale Position der Häufungspunkte mit dem Spitzenradius, dem Winkel der abfallenden Flanke und der Scanhöhe (Skalierungsverhalten der Häufungspunkte). Beim Graben wurden durch die Reisssche Entfaltung Ausschmiereffekte in der Proben-LDOS in gewisser Hinsicht

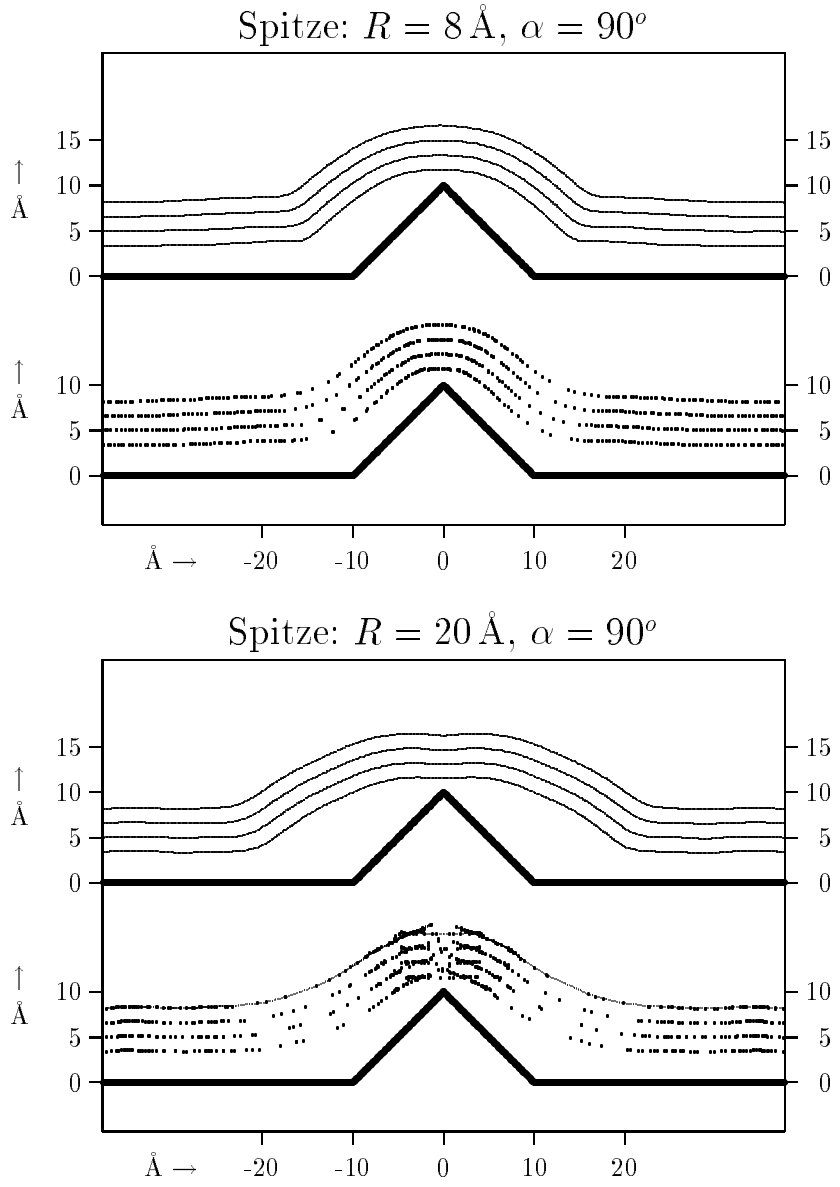


Abbildung 6.5: *Konstantstromprofile über einem Dreispitz für Spitzen mit* $\alpha = 90^o$ *und* $R = 8$, Å *bzw.* $R = 20$ Å. *Darunter die mit der Spitzengeometrie entfalteten Profile.*

kompensiert – die Häufungspunkte gaben für $R = 20$ Å die Grabenränder nahezu unabhängig von der Scanhöhe gut wieder, ganz im Gegensatz zum Grabenabbild in der LDOS.

5. Lücken in den REISSschen Kurven weisen bei Abtastprofilen die „unsichtbaren" Gebiete aus. In den REISSschen Kurven der Konstantstromprofile verblieben in diesen Lücken systematisch einige Punkte, die für alle drei Probenformen, Graben, Kerbe und Dreispitz, genähert mit dem aktuellen Spitzenradius korrelierten.

Ein Deutung von Punkt 5 ist schwierig, da mögliche Einflußgrößen und Abhängigkeiten aus den wenigen Beispiele noch nicht klar hervortreten. Das Typische zunächst beim Scannen über ein „unsichtbares" Gebiet G_0 ist, daß der relevante Strom nicht über G_0 fließt, und die Kehrseite für die Spitze dabei, daß nicht mehr der den Strom gewöhnlich tragende Apexbereich die Position der Spitze bestimmt, sondern höhergelegene Schaftbereiche. D. h. der Spitzenapex bewegt sich oberhalb von G_0 nicht mehr, wie sonst zumindest näherungsweise der Fall, entlang einer Kontur konstanter LDOS der Probe, sondern in Zonen auch weit geringerer LDOS und durchläuft insgesamt einen beträchtlichen LDOS-Gradienten.

Man könnte annehmen, daß an den Rändern von G_0, wenn der Strompfad (und damit die Kontrolle über die Spitzenposition) plötzlich[2] wieder auf den Apexbereich übergeht, dieser LDOS-Gradient entlang des Spitzenapex für die Spitzenpositionierung wichtig wird, d. h. in diesem Moment besonders viele Information über den Spitzenapex in ein Konstantstromprofil einfließen. Dies wäre zumindest ein Fingerzeig, warum ausgerechnet die „unsichtbaren" Gebiete in den REISSschen Kurven Informationen über den Spitzenradius zu liefern scheinen. Allerdings stammen die REISSschen Restpunkte bei Kerbe und Graben gerade nicht vom Rand, sondern aus der unmittelbaren Umgebung der Minima der STM-Profile, also aus der Mitte der „unsichtbaren" Gebiete! Das sind die Positionen, in denen der stromführende Pfad von einer Schaftseite auf die andere wechselt. Es liegt der Gedanke nahe, hierin den entscheidenden Umstand zu vermuten. Dieser Gedanke bekäme sogar ein anschauliches Gewand, dürfte man sich den Wechsel des Strompfades nicht als ein Springen, sondern als ein rasches Umlaufen des Pfades um den Spitzenapex herum vorstellen. Nichtsdestotrotz, eine stichhaltige Begründung für Punkt 5 steht aus.

Immerhin zeichnet sich folgendes, für eine experimentelle Verwertung interessantes Wechselverhältnis zwischen REISSschen Häufungspunkten und REISSsche Restpunkten ab, die beide, das sei nochmals betont, prägnant nur bei der Abbildung „unsichtbarer" Gebiete auftreten können:

- Häufungspunkte zeigen sich (natürlich) nur an konvex gekrümmten Probenformen wie abfallenden Flanken, REISSsche Restpunkte dagegen nur an konkav gekrümmten.
  (Die Krümmungsradien in der LDOS der Probe waren dabei in den gerechneten Beispielen sowohl konkav als auch konvex kleiner als die der Spitze.)

- Das Skalierungsverhalten der Häufungspunkte ist abhängig auch vom Winkel der Flanke (allerdings nicht sehr ausgeprägt), das der Restpunkte dagegen nicht.
  (Die Korrelation der Restpunkte mit dem Radius zeigte sich sowohl an einem Graben, einer 90^0-Kerbe als auch am Fuße eines im Winkel von 135^o ansteigenden Dreispitzes, also weitgehend unempfindlich gegen den (globalen) Öffnungswinkel der konkaven Form. Wich-

[2]Der Krümmungsradius der Proben-LDOS ist, sagen wir, spürbar kleiner als der der Spitzen-LDOS.

tig scheint hier allein der Fakt des Umspringens des Strompfades von einer Schaftseite zur anderen.)

- Geliefert werden Häufungs- und Restpunkte von unterschiedlichen Abschnitten eines Konstantstromprofils.
  (Dies erlaubt prinzipiell eine getrennte Entnahme der Daten aus einem gemessenen STM-Profil.)

Ziel weiterer Untersuchungen könnte sein:

- Wie ist das genauere Skalierungsverhalten der Häufungspunkte?
  (Betrachtung reiner Schrägen oder Stufen. Wie tief muß eine Stufe, wie schmal darf gegebenenfalls dann ein Graben sein, damit dieses zu Tage tritt? Wieviel stärker konvex gekrümmt muß die LDOS der Probenkante sein?)

- Welcher genauere quantitative Zusammenhang besteht zwischen Restpunkten und den Gegebenheiten beim Umspringen des Strompfades innerhalb eines konkav gekrümmten „unsichtbaren“ Gebietes?
  (Wie viel stärker konkav gekrümmt muß die LDOS der Probe im Gebiet des Umspringes sein? Was wäre bei unsymmetrischen Spitzen? Interessant ist folgendes: Verbreitert man einen Graben und ist der Grabenboden weiterhin „unsichtbar“, muß das hypothetische Umlaufen der Strompfades um den Apex herum mit höhrere Relativgeschwindigkeit zum Apex erfolgen, da die beiden stromführenden Schaftpunkte höher und in der Regel damit weiter auseinanderliegen. Hat also die Grabenbreite Einfluß auf die Restpunkte?)

Sollten Häufungspunkte und Restpunkte tatsächlich unterschiedliche Gegebenheiten einer Konstantstrom-Abbildung wiederspiegeln und letztlich auch unabhängig voneinander skalieren – Überfahren einer abfallenden Flanke und abhängig von deren Winkel auf der einen, Umspringen des Strompfades und abhängig z. B. von der Distanz der stromführenden Schaftpunkte auf der anderen Seite –, ließen sich, zumindest theoretisch, an einem einzigen Graben zwei Spitzenparameter wie Radius R und Offnungswinkel α nach folgender Weise bestimmen: Abgebildet wird ein Graben bekannter Breite und hinreichender Tiefe („unsichtbarer“ Grabenboden!) in unterschiedlichen Scanhöhen. Das Skalenverhalten der Häufungspunkte mit der Scanhöhe und desgleichen das der Restpunkte liefern hierdann letztlich zwei Gleichungen, um R und α in einer inversen Reisschen Entfaltung zu ermitteln.

6.6 Resümee und Ausblick

Angesichts der unendlichen Vielfalt möglicher Spitze-Probe-Geometrien sollten diesbezügliche Untersuchungen nach verallgemeinerungsfähigen Aussagen streben; im Idealfall ständen am Schluß einfache Skalengesetze für gewisse, experimentell zugängliche Größen, die man realistischerweise aber wohl dann nur mit eingeschränkter Gültigkeit für spezielle Situationen erwarten darf. Voraussetzung wäre, daß sich in einem Wechsel von Hypothesen und numerischer Überprüfung letztlich einige Kenngrößen, die nicht ad hoc gegeben werden können, als signifikant herausarbeiten lassen.

Bei den hier als Elektroden verwendeten Potentialkästen (Sommerfeld-Metalle) mit klar definierter, stetiger Randgeometrie bot sich der Vergleich zu einem geometrischen (mechanischen) Abtasten und der hierauf aufbauenden Reisschen Entfaltung an, also ein

Vergleich gegen den klassisch-geometrischen Grenzfall. Die Entfaltung arbeitet differentiell, wodurch bestimmte Eigenschaften eines flachen und zunächst nicht sehr aussagekräftigen STM-Profils fokussiert werden. Wichtig ist, daß die Geometrien von Probe und Spitze gleichberechtigt eingehen, so daß ein Fit an experimentelle Daten wahlweise Parameter der Spitze oder der Probe zum Ziele haben kann, je nachdem, über welche Seite größere Klarheit herrscht bzw. welche man bestimmen möchte.

Die Probengeometrien waren in bewußtem Gegensatz zur planaren Probenoberfläche des Standardmodells [78] gewählt. Folgerichtig zeigten die Konstantstromprofile nur geringe Änlichkeit mit den LDOS-Konturen der Probe. Die Profile waren dominiert von der Geometrie der Spitze und ähnelten Abtastprofilen. Aus den lateralen Positionen der Häufungspunkte in den Reisschen Kurven konnte dadurch die Lage der Graben- und Kerbränder auch gut rekonstruiert werden. In zweiter Ordnung zeigten diese lateralen Positionen an einer abfallenden Flanke ein typisches Skalierungsverhalten mit der Scanhöhe, dem Informationen über das Steilheit der Flanke und den Spitzenradius innewohnen.

Überraschenderweise verblieben in den mechanisch „unsichtbaren“ Gebieten systematisch einige Reissche Restpunkte, die nahezu unabhängig von der Form des entsprechenden Probenabschnitts mit dem Spitzenradius korrelierten und mit dem Wechsel des Strompfades von einer Schaftseite der Spitze zur anderen in Zusammenhang gebracht werden können. Dies wäre für eine experimentelle Bestimmung des Spitzenradius zweifellos das interessanteste Ergebnis und sollte in der im vorigen Abschnitt skizzierten Richtung weiter hinterfragt werden.

Mögliche Schwierigkeiten einer experimentellen Nutzung könnten sein, daß der Effekt bei den Häufungspunkten nicht sehr ausgeprägt ist, daß, zweitens, die Reisschen Restpunkte lediglich einem kleinen Abschnitt des Konstantstromprofils entstammen (dies sollte bei der heute erreichbaren lateralen Auflösung allerdings kein wirkliches Hindernis sein) und daß, drittens, die gesamte Reissche Entfaltung ein differentielles Verfahren darstellt, bei dem Störungen und Rauscheinflüsse verstärkt werden.

Von Interesse erscheint, der Reissche Entfaltungsprozedur einmal nicht die geometrischen Konturen, sondern die der LDOS zugrundezulegen, was ein Abgleiten zweier LDOS-Konturen zum Bezug nehmen würde. Allerdings bestände dann die zusätzliche Freiheit, welche LDOS-Konturen von Probe und Spitze bei welchem Abstand zu kombinieren sind.

Die Betrachtungen sind schließlich zweidimensional. Eine Ausdehnung auf drei Dimensionen für Elektroden mit ähnlich großen Linearabmessungen ist numerisch derzeit kaum realisierbar, mit den jetzigen Aufwand vergleichbar hingegen wären Grundgebiete von 10.000 $Å^3$, also z. B. $25 \times 20 \times 20$ $Å^3$. Sicher ist immerhin, daß eine dreidimensionale Spitze (mit ansonsten identischer x-z-Form) von einer Kante stets ein etwas flacheres Profil liefern wird, weil die dort zusätzlich erforderliche Krümmung der Spitzenwellenfunktionen in y-Richtung den bei gegebener Tunnelenergie maximal möglichen k_z-Impuls reduziert: über der Ebene tragen die Spitzenzustände mit den größten k_z den Hauptteil des Stromes, jenseits der Kante bricht dieser Anteil zusammen, wobei relativ gesehen dieser Anteil für die zweidimensionale Spitze größer ist (wegen der größeren k_z) und ein tieferes Eintauchen erzwingt [36].

Insgesamt haben sich hier einige neue Anhaltspunkte herauskristallisiert, mit denen konkrete experimentelle Ziele wie die Kalibrierung von Spitzen am 'corner hole' der Si (7×7) [44] gezielter numerisch und auch experimentell in Angriff zu nehmen sein sollten.

Kapitel 7

Zusammenfassung

In dieser Arbeit wurden zwei Themenkreise behandelt: Die Simulation des STM-Abbildungsprozesses im Rahmen eines zweidimensionalen Modells, worin beide Elektroden durch Sommerfeld-Metalle frei wählbarer Geometrie beschrieben werden und der Tunnelstrom im Transfer-Hamiltonian-Formalismus [3] bestimmt wird, und zweitens die numerische Berechnung von Zuständen in Quantenbillards endlicher Wandhöhe.

Bei den rastertunnelmikroskopischen Betrachtungen standen Fragen nach dem Einfluß der Geometrie von Probe und Spitze im Vordergrund. Es waren Geometriekonstellation gewählt worden, die im deutlichen Kontrast zur ideal planar angenommenen Probenoberfläche der Standardtheorie von Tersoff/Hamann [78] standen und die demgemäß auch größere Bereiche „unsichtbarer" Gebiete aufwiesen, sprich Gebiete, die einer geometrischen Spitzenberührung nicht zugänglich wären. Berechnet wurden Konstantstromprofile für die Abbildung dreier idealtypischer Probenformen – Graben, Kerbe und Dreispitz – mit Spitzen unterschiedlicher Radien. Durch die anschließende Entfaltung der erhaltenen Profile mit der aktuellen Spitzengeometrie nach Reiss [68] erfolgte insbesondere die Gegenüberstellung zum geometrischen (mechanischen) Abtasten, und die Diskussion konzentrierte sich auf Möglichkeiten, hierüber genauere Aussagen zur Probengeometrie und zur Kalibrierung von Spitzen zu gewinnen.

Die Konstantstromprofile erwiesen sich als dominiert von der Form der Spitzen und ähnelten Abtastprofilen. Folgerichtig konnte auch die Reisssche Entfaltung in Gestalt von Häufungspunkten die laterale Position der Ränder der „unsichtbaren" Gebiete näherungsweise gut wiedergeben. Dem überlagert zeigten die lateralen Positionen der Häufungspunkte ein Sklalierungsverhalten mit der Scanhöhe, dem Informationen über den Flankenwinkel und den Spitzenradius innewohnen. Erstere Abhängigkeit folgt der Tendenz der LDOS-Konturen der Probe, letztere wird qualitativ verständlich aus dem prinzipiellen Unterschied zwischen einem geometrischen Abtasten und einer Konstantstrom-Abbildung: an scharfen Probenregionen erfährt der Strompfad gegenüber planaren Gebieten eine Einschnürung, die relativ um so größer ist, je stumpfer die Spitze ist, ein Effekt, der bei einem Abtasten unter punktförmigem Kontakt prinzipiell nicht auftreten kann; Konstanstromprofile gewinnen gegenüber Abtastprofilen an Schärfe, je stumpfer die Spitzen werden.

Überraschenderweise verblieben in den entfalteten Kurven über den „unsichtbaren" Gebieten stets einige Punkte, die mit dem Spitzenradius korrelierten, und die mit dem Umspringen des Strompfades am Spitzenschaft in Zusammenhang gebracht werden können. Für die Kalibrierung von Spitzen wäre dies ein besonders interessanter Aspekt. Die Untersu-

chungen insgesamt lassen hoffen, daß mit dieser Entfaltung, die ein differentielles Verfahren darstellt und bestimmte Eigenschaften eines STM-Profils fokussiert, das Ausmessen von Parametern der Proben- oder Spitzengeometrien auf einer feineren Skala gelingen könnte als bisher.

Der spannenste Moment der Arbeit war zweifellos die am System „zwei Quantenbillards endlicher Wandhöhe im Tunnelkontakt“ numerisch zu Tage geförderte Existenz von Scars, die eine Tunnelbarriere durchquerten. In dem Zwei-Billardsystem “Rechteck+Kreis“ fanden sich ausgeprägt besonders im Gebiet des Rechtecks Scars, die aus der Barriere austreten und in diese zurücklaufen. Der springende Punkt ist, daß Scars einerseits nach bisherigem Verständnis nur in hinreichend nichtintegrablen (chaotischen) Systemen auftreten können und an instabile Orbits gekoppelt sind, daß andererseits die Theorie periodischer Orbits, die als einzige für die bisherigen Scar-Phänomene (Scars entlang klassischer Orbits) eine befriedigende semiklassische Beschreibung liefern kann, gegenwärtig nur auf klassischen Orbits fußt. Das System „Rechteck+Kreis“ ist natürlich in hohem Maße nichtintegrabel, „sichtbar“ wird diese Nichtintegrabilität aber nur für Bahnen entweder des Kontinuums oder für komplexe Orbits. Eine semiklassische Beschreibung unserer Resultate im Rahmen einer Theorie periodischer Orbits würde die Einbeziehung komplexer Orbits oder Bahnen des Kontinuums („ungebundener Orbits“) erfordern!

Die Konsequenzen einer solchen Erweiterung sind schwerlich abzusehen. Der Verfasser neigt hier durchaus zu der Ansicht, daß am Ende die vollständige Rückführung der stationären Quantenmechanik auf verallgemeinerte periodische Orbits stehen könnte. Einige Gedanken hierzu wurden in Abschnitt 3.4 geäußert, wo auch die Idee eines „ungebundenen Orbits“ konkreter gefaßt ist.

Schließlich bestand ein wesentlicher Teil der Arbeit im Entwurf und der Absicherung eines numerischen Verfahrens zur Berechnung höherangeregter gebundener Zustände in beliebig geformten, großen Potentialmulden, wozu aufgrund der schwierigen Sachlage bei der Lösung des großen Eigenwertproblems eine ausführliche Analyse der Diskretisierungsfehler und Untersuchungen zur Konvergenzgeschwindigkeit gehörten.

Anhang A

Exakte Gitterlösungen

Als Gitterlösungen bezeichnen wir Lösungen der auf einem fiktiven Gitter über ein Differenzenschema diskretisierten SGL, also Lösungen des Matrixeigenwertproblems $\mathbf{A}\underline{\mathbf{u}} = E\underline{\mathbf{u}}$ mit der $n \times n$-Diskretisierungsmatrix $\mathbf{A}$ und dem n-dimensionalen, die inneren Gitterpunkte beschreibenden Eigenvektor $\underline{\mathbf{u}}$ (Abschnitt 4.2). Ein äquidistantes Gitter der Schrittweite h wird nachfolgend h-Gitter genannt.

Exakte (analytische) Gitterlösungen sind hilfreich bei der Spektrenabschätzung der Diskretisierungsmatrizen. Sie bieten darüberhinaus dank der ihnen innewohnenden parametrischen Abhängigkeit von der Gitterschrittweite die einzige Möglichkeit, einmal exakt und detailiert jene mit einer Diskretisierung einhergehenden Fehler zu studieren (Abschnitt 4.3). Man bekommt fundiertere Informationen über die benötigte Feinheit des Gitters und ist etwas besser davor gefeit, die sich anschließende numerische Lösung des Eigenwertproblems mit einer die Güte der Diskretisierung weit übersteigenden Genauigkeit zu betreiben.

Aufgrund der engen Analogie zur numerischen Situation werden wir hier insbesondere den eindimensionalen eingebetteten Potentialtopf endlicher Höhe studieren. „Eingebettet“ will besagen, daß der Potentialtopf in ein größeres Grundgebiet mit unendlich hohen Wänden eingebettet liegt, wodurch das Verschwinden der Wellenfunktion auf dem Gebietsrand erzwungen wird und gebundene Zustände im Außenraum nach einer endlichen Länge abgeklungen sein müssen. Dies entspricht der Situation, die bei der numerischen Behandlung in einem endlichen Grundgebiet vorliegt.

Analytisch bekannt sind die exakten Gitterlösungen des *unendlich hohen* Potentialtopfes beliebiger Dimension [34, S.74][1]. Weil wir von diesen Lösungen das heuristische Prinzip für unser Vorgehen abgezogen haben, seien diese eindimensional jetzt als erstes kurz rekapituliert.

[1] In [34] wird kein (quantenmechanisches) Schwingungsproblem, sondern die numerische Randwertaufgabe $\Delta u = f(x, y)$ (Poisson-Gleichung) auf einem Einheitsquadrat behandelt. Die angeführten Eigenwerte und -vektoren der Diskretisierungsmatrix von Δu sind aber nichts anderes als die Eigenschwingungen eines unendlich hohen Potentialtopfes.

A.1 Der unendlich hohe Potentialtopf als heuristisches Prinzip

Der Innenraum des unendlich hohen Kastens habe die Breite L und sei mit n äquidistanten Stützstellen x_i der Schrittweite h überzogen,

$$x_i = i \cdot h, \qquad h = \frac{L}{n+1}, \qquad (1 \le i \le n).$$

Das die Matrix $\mathbf{A}$ begründende Differenzenschema der eindimensionalen SGL lautet bei zunächst beliebigem inneren Potential $V_i \equiv V(x_i)$

$$h^{-2}[(2 + V_i h^2)u_i - u_{i-1} - u_{i+1}] = Eu_i, \qquad (i = 1, \ldots, n) \tag{A.1}$$

und vereinfacht sich im Spezialfall des unendlich hohen Kasten gemäß $V_i = 0\ \forall i$.

Die n linear unabhängigen und orthonormalen Eigenvektoren der Matrix $\mathbf{A}$ zum Eigenwert E_ν sind die Vektoren $\underline{\mathbf{u}}^\nu$ mit den Komponenten

$$(\underline{\mathbf{u}}^\nu)_i = \sqrt{\frac{h}{2L}} \sin(k_\nu x_i), \quad k_\nu = \frac{\nu\pi}{L}, \quad x_i = hi \quad (1 \le \nu, i \le n), \tag{A.2}$$

wobei der Vorfaktor $\sqrt{h/(2L)}$ die Normierung $\|\underline{\mathbf{u}}^\nu\| = 1$ sicherstellt. Begründung: Für $\mathbf{A}\underline{\mathbf{u}}^\nu$ erhält man im inneren Gitterpunkt $x = x_i$ (den Gitterindex bei x_i lassen wir im folgenden wieder fort, wenn dieser dem Text nach offensichtlich ist)

$$(\mathbf{A}\underline{\mathbf{u}}^\nu)(x) = h^{-2}\sqrt{\frac{h}{2L}}\left[2\sin k_\nu x - \sin k_\nu(x-h) - \sin k_\nu(x+h)\right] \tag{A.3}$$

und weiter mit Hilfe des Sinusadditionstheorems

$$(\mathbf{A}\underline{\mathbf{u}}^\nu)(x) = \sqrt{\frac{h}{2L}} \sin(k_\nu x)\, \frac{2}{h^2}[1 - \cos k_\nu h], \tag{A.4}$$

also gerade die Form $\mathbf{A}\underline{\mathbf{u}}^\nu = E\underline{\mathbf{u}}^\nu$. Aus (A.4) kann der zugehörige Eigenwert

$$E_\nu \;=\; \frac{2}{h^2}[1 - \cos k_\nu h] \;=\; \frac{4}{h^2}\sin^2\frac{k_\nu h}{2} \qquad (1 \le \nu \le n) \tag{A.5}$$

daher unmittelbar abgelesen werden, der in der Grenze $h \to 0$ ordnungsgemäß gegen die Kontiunuumslösungen $E_\nu = k_\nu^2 = (\pi\nu/L)^2$ konvergiert. Der maximale Eigenwert entsteht für $\nu = n$:

$$E_n \;=\; \frac{4}{h^2}\cos^2\frac{\pi}{2(n+1)}\ . \tag{A.6}$$

Auf einem unendlich ausgedehnten Gitter ($n \to \infty$) – das Problem eines freien Teilchens – wird daraus $E_{\max} = 4/h^2$. Dies ist die größte auf einem h-Gitter darstellbare Energie überhaupt und aus $k_{\max} = \lim\limits_{n\to\infty} k_n = \lim\limits_{n\to\infty} \frac{\pi n}{h(n+1)}$ oder aus $4/h^2 = 2/h \cdot [1 - \cos(k_{\max} h)]$ folgen[2] dann auch mit

$$k_{\max} = \pi/h \quad \text{bzw.} \quad \lambda_{\min} = 2h \tag{A.7}$$

als zugehöriger maximaler Wellenzahl bzw. kleinster darstellbarer Wellenlänge die bekannte Aussage des SHANNONschen Abtasttheorems.

Gleichung (A.4) erfordert noch eine Zusatzüberlegung: In den randnahen Punkte $x = h$ und $x = a - h$ ist jeweils ein Nachbar kein innerer Gitterpunkt mehr und die Summanden

[2] Aber nicht aus $E_{\max} = k_{\max}^2$, denn das ist nicht die Dispersionsrelation auf einem Gitter!

u_{i-1} bzw. u_{i+1} dürften nicht auftreten. Da auf dem Rand aber $u_0 = u_{n+1} = 0$ gilt, bleibt (A.4) auch dort gültig.

Wir halten fest, daß die Gittereigenvektoren $\underline{\mathbf{u}}^\nu$ beim unendlich hohen Potentialtopf formal dieselbe Gestalt – $\sin(k_\nu x)$ – annehmen wie die Eigenfunktionen des kontinuierlichen Problems (hier stimmen sogar die Wellenzahlen k_ν überein), daß jedoch die Energieeigenwerte E_ν^{kont} und E_ν^{Gitt} sehr wohl voneinander differieren. Versuchsweise werden wir daher die Gittereigenvektoren des endlichen Potentialtopfes in der Gestalt der zugehörigen kontinuierlichen Lösungen ansetzen und die Eigenwerte anschließend über das Differenzenschema zu bestimmen versuchen.

A.2 Der eindimensionale eingebettete Potentialtopf

Für einen endlichen eingebetteten Potentialtopf wie in Abb. 4.1 auf Seite 30 lauten für $E < V_0$ die symmetrischen bzw. antisymmetrischen Ansätze der Kontinuumslösungen, wenn wir ein spiegelsymmetrisches Koordinatensystem

$$V(x) = \begin{cases} V_0\,\Theta(|x| - a) & |x| < b \\ \infty & |x| = b \end{cases} \tag{A.8}$$

zugrunde legen,

$$\begin{array}{lll} -b \le x \le -a: & -a \le x \le a: & a \le x \le b: \\ A\sinh\kappa(x+b) & \begin{array}{c} B\cos kx \\ B\sin kx \end{array} & C\sinh\kappa(x-b) \end{array} \tag{A.9}$$

$$\kappa = \sqrt{V_0 - k^2}, \quad (k, \kappa > 0), \tag{A.10}$$

die korrekt bei $x = \pm b$ verschwinden. Aus den Stetigkeitsforderungen bei $x = \pm a$ folgen für k die transzendenten Gleichungen

$$\frac{\kappa}{k}\coth\kappa L_a = \begin{cases} \tan ka & \text{(symmetrisch)} \\ -\cot ka & \text{(antisymmetr.)} \end{cases} \tag{A.11}$$

$$L_a \equiv b - a > 0, \tag{A.12}$$

die über $E = k^2$ dann auf die Energiewerte führen; L_a ist hier die Außenraumlänge.

Während beim nichteingebetteten symmetrischen Potentialtopf ($L_a \to \infty$) für beliebige Werte von V_0 und a zumindest der gebundene Grundzustand immer existiert, besteht hier wegen

$$\lim_{k \to \sqrt{V_0}} \frac{\kappa}{k}\coth\kappa L_a = (\sqrt{V_0}\,L_a)^{-1}, \quad \kappa = \sqrt{V_0 - k^2}$$

für dessen Auftreten die Grenzbedingung $(\sqrt{V_0}\,L_a)^{-1} < \tan\sqrt{V_0}\,a$.

Für das diskrete Problem setzen wir im Sinne unseres heuristischen Prinzips die Gittereigenvektoren der gebundenen Zustände in derselben Form an wie (A.9),

$$\begin{array}{lll} -b \le x_i \le -a: & -a < x_i < a: & a \le x_i \le b: \\ A\sinh\kappa(x_i+b) & \begin{array}{c} B\cos kx_i \\ B\sin kx_i \end{array} & C\sinh\kappa(x_i-b), \end{array} \tag{A.13}$$

allerdings soll und kann für k und κ $(k, \kappa > 0)$ jetzt nur deren Unabhängigkeit von x_i, nicht mehr jedoch die Relation (A.10) vorausgesetzt werden. Bestimmt man über die Normierung

z. B. die Konstante A, verbleiben, da je nach Symmetrie $C = \pm A$ gilt, als freie Parameter noch B, k und κ. Ferner wird über die (diskretisierte) Lage der Potentialkante dergestalt verfügt, daß an den Stützstellen $x_i = \pm a$ jeweils das Außenpotential V_0 beginnt:

$$V_i = \begin{cases} 0 & \text{für } |x_i| < a \\ V_0 & \text{für } a \leq |x_i| < b. \end{cases} \tag{A.14}$$

Für jeden inneren Gitterpunkt $|x| < a - h$, dessen *Nachbarn ebenfalls* zum Innenbereich gehören, liefert das Differenzenschema (A.1) bei gleichem Vorgehen wie im Anschluß an (A.3) erneut den „inneren" Eigenwert

$$E^{\text{innen}} = \frac{2}{h^2}[1 - \cos kh]\,. \tag{A.15}$$

Für einen äußerer Gitterpunkt $|x| > a + h$, der *einschließlich* seiner beiden Nachbarn im Außenbereich liegt, liest sich $(\mathbf{A}\underline{\mathbf{u}})(x)$ hingegen

$$h^{-2}C\left[(2 + V_0h^2)\sinh\kappa(x - b) - \sinh\kappa(x - b - h) - \sinh k(x - b + h)\right].$$

Mit $\sinh(x \pm y) = \sinh(x)\cosh(y) \pm \cosh(x)\sinh(y)$ entsteht

$$(\mathbf{A}\underline{\mathbf{u}})(x) = C\sinh\kappa x\,\frac{1}{h^2}\left[2 + V_0h^2 - 2\cosh\kappa h\right]$$

und der „äußere" Eigenwert lautet folglich:

$$E^{\text{außen}} = V_0 + \frac{2}{h^2}[1 - \cosh\kappa h]. \tag{A.16}$$

Die Eigenwertausdrücke (A.15) und (A.16) des inneren und äußeren Bereichs müssen bei einem herausgegriffenen Eigenvektor natürlich übereinstimmen, weshalb die Identitäten

$$\frac{2}{h^2}\left[-\cos kh + \cosh\kappa h\right] = V_0 \tag{A.17}$$

bzw.

$$\frac{4}{h^2}\left[\sin^2(\frac{kh}{2}) + \sinh^2(\frac{\kappa h}{2})\right] = V_0 \tag{A.18}$$

zu fordern sind. Diese stellen den Zusammenhang zwischen k und κ her und sind das Gitteranalogon zur Relation $k^2 + \kappa^2 = V_0$, in die sie für $h \to 0$ korrekterweise übergehen. $\kappa = \kappa(k)$ erhalten wir daher aus (A.17) über die inverse Kosinushyperbolikusfunktion:

$$\kappa(k) = \frac{1}{h}\ln\left(q + \sqrt{q^2 - 1}\right) \tag{A.19}$$

$$q \equiv \frac{V_0h^2}{2} + \cos kh \tag{A.20}$$

Im Übergangsbereich, an den Punkten $x = a - h$ und $x = a$, liefert das Differenzenschema zwei weitere Gleichungen, die innere und äußere Ansätze miteinander verknüpfen und die noch unbestimmten Paramter k und B festsetzen. Wir betrachten die symmetrische Lösung: Am Punkt $x = a - h$ muß gemäß $(\mathbf{A}\underline{\mathbf{u}})(a - h) = E\underline{\mathbf{u}}(a - h)$ die Identität

$$\frac{1}{h^2}\left[2B\cos k(a - h) - B\cos k(a - 2h) - C\sinh\kappa(a - b)\right] = EB\cos k(a - h).$$

erfüllt sein. Wird der zu $\cos k(a-h)$ passende „innere" Eigenwert E nach (A.15) eingetragen, entsteht

$$B\cos ka = C\sinh\kappa(a - b), \tag{A.21}$$

und analog am Punkt $x = a$ aus der Forderung $(\mathbf{A}\underline{\mathbf{u}})(a) = E\underline{\mathbf{u}}(a)$, wenn diesmal $E^{außen}$ nach (A.16) und die Additionstheoreme der hyperbolischen Funktionen zu Rate gezogen werden,

$$B\cos k(a-h) = C\sinh\kappa(a-b-h). \tag{A.22}$$

(A.21) und (A.22) zusammengenommen beinhalten offenbar die „Stetigkeit“ des Differenzenquotienten $\frac{1}{h}[\underline{\mathbf{u}}(a) - \underline{\mathbf{u}}(a-h)]$ an der Potentialkante, wie eine Subtraktion beider Gleichungen bestätigt. Dieses Äquivalent zur Stetigkeit der ersten Ableitung im Kontinuumsfall wurde hier allerdings nicht primär gefordert, sondern folgte algebraisch aus der bei einem Eigenwertproblem zu fordernden Konstanz des Eigenwertausdruckes an allen Gitterpunkten, sprich der Identität (A.17) (und natürlich aus dem Ansatz (A.13)).

Eine Division von (A.21) und (A.22) liefert schließlich die tranzendente Bestimmungsgleichung für k,

$$\frac{\cos k(a-h)}{\cos ka} = \frac{\sinh\kappa(L_a+h)}{\sinh\kappa L_a}, \tag{A.23}$$

wobei $L_a = b-a$ und $\kappa = \kappa(k)$ entsprechend (A.19) zu substituieren sind. Für den Vergleich mit der Kontinuumsgleichung (A.11), den Grenzübergang $h \to 0$, als auch für die praktische Berechnung von k ist (A.23) allerdings noch wenig geeignet. In (A.23) werden deshalb die Zählerfunktionen noch vermöge ihrer Additionstheoreme expandiert:

$$\tan(ka)\sin(kh) + \cos(kh) = \cosh(\kappa h) + \coth(\kappa L_a)\sinh(\kappa h). \tag{A.24}$$

Mittels (A.17) kann nun $\cosh(\kappa h)$ direkt, $\sinh(\kappa h)$ vermöge $\sqrt{\cosh^2 - 1}$ und κ in $\tanh(\kappa L_a)$ über (A.19) substituiert werden. Eine analoge Betrachtung für den antisymmetrischen Ansatz führt zusammengefaßt dann auf die folgende transzendente Gleichung für k:

$$F(k) = \begin{cases} \tan ka & \text{(symmetrisch)} \\ -\cot ka & \text{(antisymmetr.)} \end{cases} \tag{A.25}$$

$$F(k) = \frac{1}{\sin kh}\left\{\frac{V_0h^2}{2} + \coth\left[\frac{L_a}{h}\ln\left(q+\sqrt{q^2-1}\right)\right]\sqrt{q^2-1}\right\}. \tag{A.26}$$

Hat man darüber k bestimmt, findet sich κ und E hernach über (A.19) bzw. (A.15). Im Gegensatz zum speziellen Fall des unendlich hohen Potentialtopfes unterscheiden sich hier also Gitter- und Kontinuumslösungen außer in den Energieeigenwerten auch hinsichtlich Wellenzahl k und der Abklingkonstante κ.

An (A.25) läßt sich darüberhinaus gut der Übergang zur Kontinuumsvariante (A.11) für $h \to 0$ verifizieren[3], wodurch zusammen mit der gleich im Anschluß demonstrierten Existenz reeller k-Lösungen der versuchsweise Ansatz (A.13) damit bestätigt ist.

[3] Hierfür genügt es, alle q-relevanten Größen in $F(k)$ in der Ordnung $O(h)$ zu betrachten. Es gilt für $h \to 0$: $q \approx 1 + O(h^2)$ und weiter

$$\sqrt{q^2-1} \approx \sqrt{\left(\frac{V_0h^2}{2} + 1 - \frac{(kh)^2}{2} + O(h^4)\right)^2 - 1} = h\sqrt{V_0 - k^2} + O(h^2),$$

$$\ln\left(q+\sqrt{q^2-1}\right) \approx h\sqrt{V_0-k^2} + O(h^2),$$

Für coth[ln(...)] erhalten wir folglich

$$\lim_{h\to 0}\coth\left(\frac{L_a}{h}h\sqrt{V_0-k^2}\right) = \coth\left(L_a\sqrt{V_0-k^2}\right).$$

Hinzu tritt der Faktor

$$\lim_{h\to 0}\frac{\sqrt{q^2-1}}{\sin kh} = \lim_{h\to 0}\frac{h\sqrt{V_0-k^2}}{kh} = \frac{\sqrt{V_0-k^2}}{k},$$

Ferner entsteht für $L_a \to \infty$ die Lösung des nichteingebetteten Kastens auf einem unendlichen h-Gitter:

$$F_\infty(k) = \frac{1}{\sin kh}\left\{\frac{V_0 h^2}{2} + \sqrt{q^2-1}\right\}. \tag{A.27}$$

In Abb. A.1 sind $F(k)$ und $F_\infty(k)$ für die drei Schrittweiten $h = 0, 0.8$ und 1.35 Å dargestellt. Die Kastenbreite $2a$ findet sich nur in den zugehörigen trigonometrischen Funktionen $\tan(ka)$ und $\cot(ka)$ wieder; eingtragen sind in Abb. A.1 die für $2a = 10$ Å. Je größer die Gitterweite h, desto höher liegen die Eigenwerte k. Während hierbei für $h = 0.8$ Å die Anzahl gebundener Zustände noch mit der des kontinuierlichen Grenzfalls $h = 0$ übereinstimmt, ist für $h = 1.35$ Å ein zusätzlicher $E < V_0$-Zustand zu konstatieren. Dieser hat seine Ursache in der mit wachsendem h immer flacher verlaufende Dispersionsrelation $E(k)$ nach (A.15) – vergleiche Abb. A.2 –, in deren Folge Wellenzahlen, die $E = k^2$ gehorchend energetisch oberhalb V_0 einzuordnen wären, hier dann darunterbleiben. Darüberhinaus kompensiert die gegenüber $E = k^2$ flachere Dispersionsrelation (A.15) bei der Energie in gewissem Umfang wieder den Diskretisierungsfehler, den die generell zu großen k-Werte hervorrufen.

Für h-Werte, die die in Abb. A.1 illustrierten noch übersteigen, würden die Kurven $F(k)$ und $F_\infty(k)$ ohne weiteren Schnittpunkt schließlich gegen Unendlich streben – mathematischer Ausdruck des Umstandes, daß der maximale Eigenwert des Gitters selbst unterhalb V_0 zu liegen gekommen ist und generell keine gebundenen $E < V_0$-Zustände mehr existieren (d. h. in einem Differenzenschema dieser Schrittweite nicht mehr darstellbar sind). Zu den maximalen Gittereigenwerten vergleiche auch Abb. A.2.

Weil mit zunehmender Kastenbreite die tan- und cot-Funktionen immer steiler verlaufen, reduzieren sich die Unterschiede in den k-Eigenwerten auf den verschiedenen Gittern, d. h. die absoluten Diskretisierungfehler bzgl. k oder λ werden geringer, je breiter ein Kasten wird.

Ferner liefern in Abb. A.1 die eingebetteten Varianten $F(k)$ erwartungsgemäß stets die höheren Eigenwerte gegenüber $F_\infty(k)$, wobei zur Verdeutlichung des Effekts die außergewöhnlich kurze Außenraumlänge $L_A = 1$ Å gewählt wurde. Demgegenüber hätten Werte $L_a \geq 8$ Å, wie sie bei den numerischen Rechnungen später tatsächlich zum Einsatz kamen, in Abb. A.1 praktisch keine Differenzen mehr erkennen lassen.

A.3 Verallgemeinerungen

Das im vorhergehenden Abschnitt praktizierte Verfahren ist offensichtlich auf beliebige stückweise konstante eindimensionale Potentiale, die auch im Kontinuumsfall immer analytisch lösbar sind, übertragbar. Stückweise konstant heißt im Differenzenschema (A.1), es liegen wenigstens drei benachbarte Gitterpunkte immer auf gleichem Potential. Setzen wir die Gittereigenvektoren in jedem Teilstück mit dem Potential $V_\alpha = \text{const}$, $\alpha = 1, 2, \ldots$ in der Form

$$u_i = A_\alpha e^{k_\alpha x_i} + B_\alpha e^{-k_\alpha x_i} \tag{A.28}$$

---

so daß als Endresultat

$$\lim_{h\to 0} F(k) = \frac{\sqrt{V_0 - k^2}}{k} \coth\left(L_a\sqrt{V_0 - k^2}\right).$$

entsteht. Wegen (A.10) ist das gerade die linke Seite der Eigenwertgleichung (A.11).

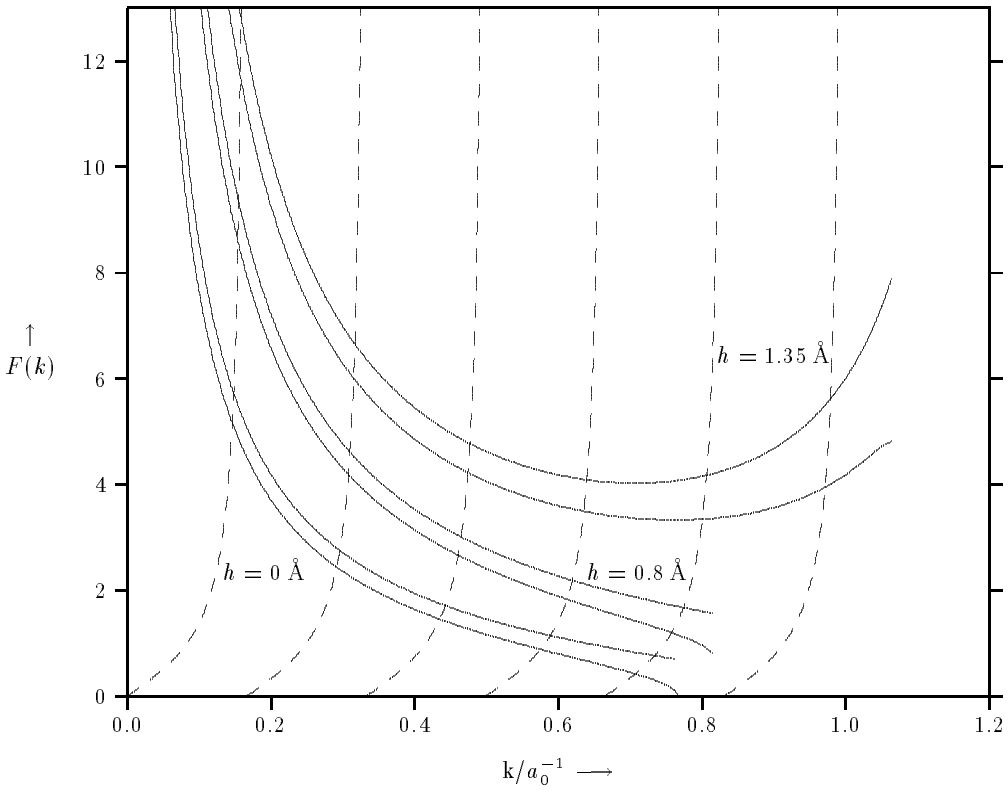


Abbildung A.1: *Die Funktionen $F(k)$ und $F_\infty(k)$ – die Gln. (A.26) und (A.27) – der transzendenten Eigenwertgleichung (A.25) für die Gitterlösungen des eindimensionalen Potentialtopfes. Die drei Kurvenpaare stehen von unten nach oben für die folgenden Schrittweiten:* 0 Å *(der kontinuierliche Grenzfall),* 0.8 Å *und* 1.35 Å. *Dabei zeigt bei jedem Paar die untere Kurve jeweils $F_\infty(k)$, d. h. den nichteingebetteten Potentialtopf auf einem unendlichen Gitter, und die obere Kurve $F(k)$ für die hier betont kurz gewählte Außenraumlänge $L_a =$* 1 Å. *Die Kastenbreite $2a$ selbst spiegelt sich nur in den trigonometrischen Funktionen* $\tan(ka)$ *und* $\cot(ka)$ *wieder, die beispielhaft für $2a =$* 10 Å *eingetragen sind, und aus deren Schnittpunkten mit $F(k)$ die k-Eigenwerte folgen. Weil mit wachsendem a die tan- und cot-Funktionen immer steiler verlaufen, reduzieren sich die Unterschiede der k-Eigenwerte auf verschiedenen Gittern, d. h. die absoluten Diskretisierungfehler bzgl. k oder λ werden geringer, je breiter ein Kasten wird,*

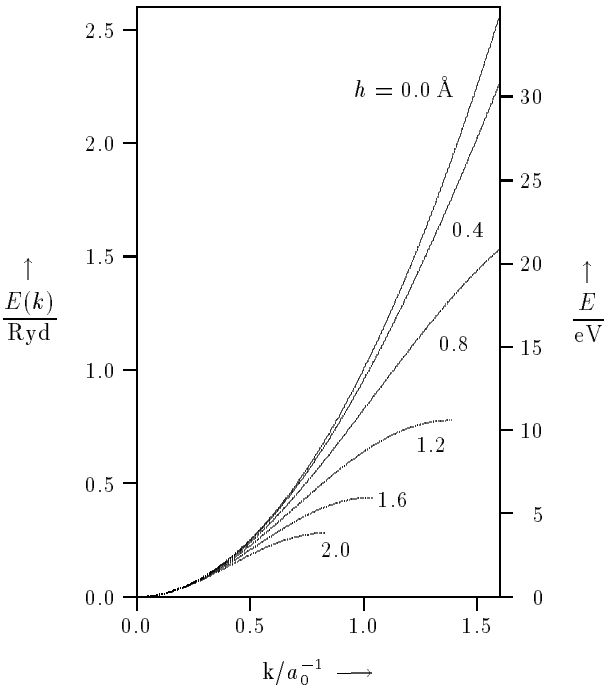


Abbildung A.2: *Die Dispersionsrelation* $E(k) = 2/h^2 \cdot [1 - \cos kh]$ *der Gitterlösungen des eindim. Potentialtopfes bzw. überhaupt von Gitterlösungen in Gebieten mit stückweise konstantem* $V = 0$ *(A.29) für die Schrittweiten* $h = 0 \ldots 2$ Å. *Die Kurven enden gegebenenfalls eher, wenn der maximale Gittereigenwert (A.7) erreicht ist.*

an ($k \to i\kappa$ ist evident), liefert das Differenzenschema jeweils die Eigenwertdarstellung

$$E = V_\alpha + \frac{2}{h^2}[1 - \cos k_\alpha h]. \tag{A.29}$$

Da E in allen Bereichen übereinstimmen muß, folgt

$$V_1 - \frac{2}{h^2}\cos k_1 h = V_2 - \frac{2}{h^2}\cos k_2 h = \ldots \tag{A.30}$$

als der Zusammenhang zwischen den verschiedenen Wellenzahlen k_α – das Gitteranalogon zu $V_1 + k_1^2 = V_2 + k_2^2 = \ldots$

Bezeichnen wir den Gitterpunkt x_i, an dem das Potential von Wert $V_{\alpha-1}$ auf den Wert V_α springt, mit a_α,

$$V(x_i) = V_\alpha \quad \text{für} \quad a_\alpha \le x_i < a_{\alpha+1},$$

lauten die Anschlußbedingungen an den Potentialkanten, z. B. bei a_1

$$\begin{aligned} A_1 e^{k_1 a_1} \quad &+ B_1 e^{-k_1 a_1} \quad &= \quad A_2 e^{k_2 a_1} \quad &+ B_2 e^{-k_2 a_1} \\ A_1 e^{k_1(a_1-h)} &+ B_1 e^{-k_1(a_1-h)} &= \quad A_2 e^{k_2(a_1-h)} &+ B_2 e^{-k_2(a_1-h)} \end{aligned}$$

und sinngemäß für a_2 etc. Damit existieren zur Bestimmung der Konstanten A_α, B_α ebensoviele Gleichungen wie im Kontinuumsfall aus der Stetigkeit von Wellenfunktion und erster Ableitung resultieren, so daß das weitere Vorgehen prinzipiell wie dort erfolgen kann (gebundene Zustände, Streuzustände auf einem unendlichen Gitter, etc.). Die algebraische Handhabung wird allerdings durch den komplizierteren E-k-κ-Zusammenhang, (A.29) und (A.30), erschwert, wie ein Blick auf die einander entsprechenden Gleichungen (A.11) und (A.25) zur Bestimmung von k im vorigen Abschnitt illustriert.

Exemplarisch sei die bei der Abschätzung der Diskretisierungsfehler (Abschnitt 4.3) interessante Reflexion an einer Potentialschwelle der Höhe V_0 angeführt. Die Energie $0 < E < V_0$ ist kontinuierlich, die zugehörigen $k(E)$-Werte im Gebiet $V = 0$ folgen aus (A.29),

$$k(E) = \frac{1}{h}\arccos\left[1 - \frac{h^2}{2}E\right], \tag{A.31}$$

und die Ersetzung $k \to i\kappa$ in (A.29) liefert die Abklingkonstante $\kappa(E)$ des evanescenten Teils ($E < V_0$):

$$\kappa(E) = \frac{1}{h}\text{arccosh}\left[1 + \frac{h^2}{2}(V_0 - E)\right]. \tag{A.32}$$

Weiterhin existieren im Mehrdimensionalen einfache Lösungsformen im Falle separabler Potentiale, wie folgendes zweidimensionale Beispiel zeigt: Zerfällt in kartesischen Koordinaten das Potential in eine Summe zweier jeweils nur von einer Koordinate abhängigen Summanden,

$$V(x_i, y_j) = V_x(x_i) + V_y(y_j), \tag{A.33}$$

ist die Eigenlösung u_{ij} das dyadische oder Tensorprodukt $\underline{\mathbf{u}} \otimes \underline{\mathbf{v}}$ zweier eindimensionaler Lösungsvektoren $\underline{\mathbf{u}}(x_i)$ und $\underline{\mathbf{v}}(y_j)$, komponentenweise dargestellt

$$u_{ij} = u_i\, v_j\,. \tag{A.34}$$

Begründung: Das Differenzenschema in zwei Dimensionen

$$\frac{1}{h^2}\left\{[4 + V(x_i, y_j)h^2]u_{ij} - u_{i-1,j} - u_{i+1,j} - u_{i,j-1} - u_{i,j+1}\right\} = Eu_{ij} \tag{A.35}$$

kann mit (A.33) und (A.34) geschrieben werden als

$$\begin{aligned} h^{-2}\left\{[2 + V_x(x_i)h^2]u_i - u_{i-1} - u_{i+1}\right\} v_j \quad &+ \\ h^{-2}\left\{[2 + V_y(y_i)h^2]v_j - v_{j-1} - v_{j+1}\right\} u_i \quad &= \quad (E_x + E_y)u_i v_j\,, \end{aligned} \tag{A.36}$$

d. h. wir erhalten links je die eindimensionalen Differenzenschemata für u_i und v_j, weshalb auch die Eigenwerte $E = E_x + E_y$ sich additiv aufbauen. (Das gilt natürlich für beliebige lineare Differenzengleichungen, in die das Potential linear eingeht und nicht nur für (A.35).) Anwendungsbeispiele sind der mehrdimensionale unendlich hohe Potentialtopf oder auch das aus der Überlagerung eindimensionaler Kästen endlicher Höhe entstehende Potentialgebilde (5.3) auf Seite 57, das für Testrechnungen und Abschätzungen der Eigenwertspektren verwendet wurde.

Anhang B

Lösung des großdimensionierten Eigenwertproblems

B.1 Vorbemerkungen

Dieses Kapitel beschreibt die numerische Lösung des großdimensionierten schwachbesetzten Eigenwertproblems $\mathbf{A}\underline{\mathbf{u}} = \lambda\underline{\mathbf{u}}$, $\mathbf{A} \in \mathbf{S}^{n,n}$, wie es aus der Diskretisierung der SGL hervorgegangen ist. Gesucht sind Eigenlösungen in der Umgebung einer frei wählbaren Energie ϵ, so daß die konkrete Aufgabe lautet, einige der betragskleinsten Eigenwerte und zugehörigen Eigenvektoren der Matrix $\mathbf{A} - \epsilon\,\mathbf{I}$ zu bestimmen.

Die Größe der Matrix beschränkt hier die Auswahl von vornherein auf die sogenannten vektoriterativen Verfahren, die nicht explizit auf $\mathbf{A}$ operieren, sondern mit einer Matrix-mal-Vektor-Regel, $\underline{\mathbf{w}} = \mathbf{A}\underline{\mathbf{v}}$, auskommen. Die Matrix kann als „Black Box" betrachtet werden und braucht physisch gar nicht vorhanden sein, was bei den großen, schwachbesetzten und einfach strukturierten Matrizen, wie sie insbesondere bei der Diskretisierung partieller DGL auftreten, einen numerischen Zugang oft überhaupt erst ermöglicht. Man unterscheidet dabei projektive Verfahren, die sukzessive die Projektion auf den Unterraum der gesuchten Eigenvektoren verbessern, und Relaxationsverfahren, die ein geeignet gewähltes Funktional maximieren.

Eine Schwierigkeit birgt natürlich die immense Dimension von $\mathbf{A}$, ein 300×300–Gitter führt bereits zu $n = 90.000$, ein solches in drei Dimensionen auf die interessante Zahl 27 Millionen.

Fast schwerwiegender noch, weil verfahrensinherent, wiegen die folgenden zwei Punkte: Zum einen ist die Konvergenz der iterativen Verfahren auf den spektralen Abstand der gesuchten Eigenwerte von allen übrigen gegründet und droht folglich zu verschwinden bei sehr dichten Spektren oder größeren Eigenwerthaufen. Hinzukommt, daß auf vektoriterativem Wege direkt immer nur Eigenwerte an den Spektrumsrändern bestimmt werden können, daß es jedoch ungleich aufwendiger wird, wenn beliebige Bereiche des Spektrums gefordert sind (wenn wie bei uns ϵ frei „durchgestimmt" werden soll). Verfügt man dann nicht über einen schnellen Löser des Gleichungssystems $(\mathbf{A} - \epsilon\mathbf{I})\underline{\mathbf{x}} = \underline{\mathbf{v}}$, um mit der inversen Vektoriteration

arbeiten zu können, ist man zur Iteration mit der quadrierten Form $(\mathbf{A}-\epsilon\mathbf{I})^2-\rho\mathbf{I}$ genötigt (Abschnitt B.3), wodurch sich das Konvergenzverhalten jedoch dramatisch verschlechtert (Abschnitt B.7).

Das Zusammentreffen beider Umstände in unserer Aufgabe führt letztlich zu einer außerordentlich schwachen Konvergenz mit bis zu 30 000 erforderlichen $\mathbf{A}\underline{\mathbf{v}}$–Iterationen. Das Problematische daran, neben der Zeitfrage, ist, daß die Akkumulation der Rundungsfehlern innerhalb der Iterierten $\underline{\mathbf{v}}$ bedrohliche Ausmaße anzunehmen beginnt, daß diese Akkumulation jedoch nicht mehr wie z. B. noch bei einem Skalarprodukt mit den bekannten, kostengünstigen Methoden wie *Aufsummierung der Zwischensummen in höherer Genauigkeit* oder *Kaskadensummation* (Abb. B.1) unterbunden werden kann. Die Zwischenresultate sind hier die Iterierten selbst, d. h. eine Verdopplung der Mantissenzahl verdoppelt nahezu den Speicherbedarf der gesamten Aufgabe, von der vermehrten Rechenzeit zu schweigen. Der Aufwand würde schneller noch wachsen als die Zahl der Operationen, da der Aufwand pro Operation ebenfalls steigt.

Für die in Betracht kommenden Algorithmen werden wir damit die Grenzen hinsichtlich Konvergenzgeschwindigkeit und Genauigkeit tangieren. Das einzelne Verfahren lediglich als „Black-Box“ zu sehen, scheint an dieser Stelle kaum mehr gerechtfertigt, und eine etwas eingehendere Betrachtung der einzelnen Verfahren tut not, um zu einer begründeten Auswahl und zu hinreichend präzisen Aussagen über Konvergenz und Fehlerakkumulation zu gelangen. Entsprechend der Problemstellung werden wir uns hierbei auf die spezielle symmetrische Eigenwertaufgabe beschränken.

B.2 Vektor- und Teilraumiteration

Diese iterativen Verfahren approximieren diejenigen Eigenvektoren der Matrix $\mathbf{A}\in\mathbf{S}^{n,n}$, die zu sogenannten dominanten, d. h. betragsgrößten Eigenwerten gehören. Die Algorithmen haben selbst keinen allzugroßen direkten Anwendungsbereich, sind jedoch die Grundlage der meisten leistungsfähigeren Verfahren und lassen deren wesentlichsten Merkmale gut erkennen.

Im folgenden seien die Eigenwerte $\{\lambda_i\}$ von $\mathbf{A}$ nach absteigenden Beträgen gemäß

$$|\lambda_1|\geq|\lambda_2|\geq\ldots|\lambda_{n-1}|\geq|\lambda_n| \tag{B.1}$$

geordnet und $\{\underline{\mathbf{u}}_i\}$ sei ein orthonormales System zugehöriger Eigenvektoren. Die Eigenwerte $\{\lambda_1,\ldots,\lambda_p\}$ heißen dominant, wenn

$$|\lambda_{p+1}|<|\lambda_p| \tag{B.2}$$

gilt. Den p dominaten Eigenwerten entspricht der durch die zugehörigen Eigenvektoren aufgespannte Teilraum

$$\mathcal{S}_p=\mathcal{S}_p(\mathbf{A})\equiv\mathrm{span}\{\underline{\mathbf{u}}_1,\ldots,\underline{\mathbf{u}}_p\}, \tag{B.3}$$

wobei $\{\underline{\mathbf{u}}_1,\ldots,\underline{\mathbf{u}}_p\}$ offensichtlich eine orthonormale Basis von $\mathcal{S}_p$ bildet und $\mathcal{S}_p$ die Dimension p besitzt. Aus $\mathbf{A}\underline{\mathbf{u}}_i=\lambda_i\underline{\mathbf{u}}_i\in S_p$ $(i=1,\ldots,p)$ folgt außerdem, daß $\mathcal{S}_p$ invariant unter der durch $\mathbf{A}$ vermittelten linearen Abbildung ist ($\mathcal{S}_p$ heißt invarianter Teilraum von $\mathbf{A}$). Umgekehrt können jedem invarianten Teilraum der Dimension p genau p Eigenvektoren und Eigenwerte von $\mathbf{A}$ zugeordnet werden. Verfügt man daher über eine genügend gute Approximation von S_p, wird man erwarten dürfen, aus dieser wegen der Trennung der

p dominanten Eigenwerte von den übrigen auch genügend gute Approximationen für die Eigenpaare $\{\lambda_i, \underline{\mathbf{u}}_i\}$ $(i = 1, \ldots, p)$ beschaffen zu können.

Wie gelangt man nun zu Approximationen für $\mathcal{S}_p$. Hierzu beachte man, daß die Eigenwertzerlegung von $\mathbf{A}$,

$$\begin{aligned} \mathbf{A} \;=\; \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \;&=\; \left[\lambda_1 u_1 u_1^T + \ldots + \lambda_p u_p u_p^T\right] \\ &\quad + \left[\lambda_{p+1} u_{p+1} u_{p+1}^T + \ldots + \lambda_n u_n u_n^T\right] \end{aligned} \tag{B.4}$$

mit $\mathbf{U} = (\underline{\mathbf{u}}_1, \ldots, \underline{\mathbf{u}}_n)$, $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ und $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ die analoge Zerlegung

$$\begin{aligned} \mathbf{A}^k \;=\; \mathbf{U}\boldsymbol{\Lambda}^k\mathbf{U}^T \;&=\; \left[\lambda_1^k u_1 u_1^T + \ldots + \lambda_p^k u_p u_p^T\right] \\ &\quad + \left[\lambda_{p+1}^k u_{p+1} u_{p+1}^T + \ldots + \lambda_n^k u_n u_n^T\right] \end{aligned} \tag{B.5}$$

für die k-te Potenz von $\mathbf{A}$ nach sich zieht. Für einen geeigneten Startvektor $\underline{\mathbf{v}}$, der im System der $\{\underline{\mathbf{u}}_j\}$ die Darstellung

$$\underline{\mathbf{v}} = \sum_j c_j \underline{\mathbf{u}}_j \quad \text{mit} \quad c_j = \underline{\mathbf{u}}_j^T \underline{\mathbf{v}} \tag{B.6}$$

besitzt und dessen erste Entwicklungskoeffizienten $c_1, \ldots, c_p$ nicht sämtlich verschwinden, stellt die Iterierte $\underline{\mathbf{v}}^{(k)} = \mathbf{A}^k \underline{\mathbf{v}}$ demzufolge für genügend großes k „fast“ eine Linearkombination der ersten p Eigenvektoren dar, d. h. $\mathbf{A}^k \underline{\mathbf{v}}$ liegt „fast“ in $\mathcal{S}_p$. Die Iteration wird dabei selbsverständlich rekursiv gemäß $\underline{\mathbf{v}}^{(k+1)} = \mathbf{A}\underline{\mathbf{v}}^{(k)}$ ausgeführt.

Die zugehörige Eigenwertnäherung μ liefert der sogenannte Rayleigh-Quotient, der für einen nicht notwendigerweise normierten Vektor $\underline{\mathbf{v}}$ allgemein die Form

$$\mu \;=\; R(\underline{\mathbf{v}}) \;\equiv\; \frac{\underline{\mathbf{v}}^T \mathbf{A} \underline{\mathbf{v}}}{\underline{\mathbf{v}}^T \underline{\mathbf{v}}} \tag{B.7}$$

besitzt und die im Sinne des Residualkriteriums [42, S. 360]

$$\|\mathbf{A}\underline{\mathbf{v}} - \mu\underline{\mathbf{v}}\| \rightarrow \text{Minimum} \tag{B.8}$$

bestmögliche Eigenwertapproximation μ für eine gegebene Eigenvektornäherung $\underline{\mathbf{v}}$ darstellt.

Im einfachsten Fall $p = 1$ ist nur ein Vektor zu iterieren, der dann die Richtung $\underline{\mathbf{u}}_1$ oder $-\underline{\mathbf{u}}_1$ approximiert. Einschließlich einer Über- oder Unterlauf vermeidenden Normierung nach jedem Iterationsschritt lautet das Basisverfahren der Vektoriteration damit:

S1 (Iteration): Berechne $\underline{\mathbf{w}}^{(k+1)} = \mathbf{A}\underline{\mathbf{v}}^{(k)}$
S2 (Normierung): Setze $\underline{\mathbf{v}}^{(k+1)} = \underline{\mathbf{w}}^{(k+1)} / \|\underline{\mathbf{w}}^{(k+1)}\|$

Für $c_1 = \underline{\mathbf{u}}_1^T \underline{\mathbf{v}}^{(0)} \neq 0$ konvergieren die $\underline{\mathbf{v}}^{(k)}$ gemäß

$$0 \leq \tan\varphi_k \leq \sigma^k \tan\varphi_0 \quad \text{mit} \quad \varphi_k = \sphericalangle\,(\underline{\mathbf{u}}_1, \underline{\mathbf{v}}^{(k)}) \in [0, \pi/2)\,, \tag{B.9}$$

d. h. linear mit dem Konvergenzfaktor $\sigma = |\lambda_2/\lambda_1| < 1$ gegen $\underline{\mathbf{u}}_1$ und die Rayleigh-Quotienten $\mu_k = R(\underline{\mathbf{v}}^{(k)})$ im Sinne des Abstandskriteriums $|\lambda_1 - \mu_{k+1}| = \sigma^2 |\lambda_1 - \mu_k|$ mit σ^2 – also schneller als die $\underline{\mathbf{v}}^{(k)}$ – gegen den Eigenwert λ_1 [42, S. 376].

Sind die ersten p Eigenpaare $\{\lambda_i, \underline{\mathbf{v}}_j\}$ gesucht, die dann nicht mehr notwendigerweise getrennt liegen müssen, ist die simultane Iteration mehrerer Vektoren gewöhnlich am günstigsten, d. h. es werden direkt Approximationen $\mathcal{V}_k$ für den p-dominaten Teilraum $\mathcal{S}_p$ berechnet. In der einfachsten Form kann hierzu ein Startteilraum $\mathcal{V}_0$ der Dimension

p vermöge $\mathcal{V}_{k+1} = \mathbf{A}\mathcal{V}_k$ mit $\mathbf{A}$ iteriert werden, wobei der Teilraum $\mathcal{V}_k$ durch die Spalten einer geeigneten Matrix $\mathbf{V}_k$,

$$\mathcal{V}_k = \mathcal{R}(\mathbf{V}_k), \quad \mathbf{V}_k \equiv (\underline{\mathbf{v}}_1^{(k)}, \ldots, \underline{\mathbf{v}}_p^{(k)}) \in \mathbf{R}^{n,p}, \tag{B.10}$$

aufgespannt wird und die Spalten von

$$\mathbf{W}_{k+1} \equiv (\underline{\mathbf{w}}_1^{(k+1)}, \ldots, \underline{\mathbf{w}}_p^{(k+1)}) = \mathbf{A}\mathbf{V}_k = (\mathbf{A}\underline{\mathbf{v}}_1^{(k)}, \ldots, \mathbf{A}\underline{\mathbf{v}}_p^{(k)}) \tag{B.11}$$

dann den Teilraum $\mathcal{V}_{k+1}$ representieren. Allerdings würde bei naiver Ausführung die Basis $\mathbf{V}_k$ mit wachsendem k zunehmend parallelisiert, weshalb mit einer orthonormalen $\mathbf{V}_k$ gearbeit wird (*Teilraumiteration mittels orthonormaler Basen*):

S1 (Iteration): Berechne $\mathbf{W}_{k+1} = \mathbf{A}\mathbf{V}_k$

S2 (Orthonormalisierung): Bestimme spaltenorthonormales $\mathbf{Q}_{k+1} \in \mathbf{R}^{n,p}$ und obere Dreiecksmatrix $\mathbf{R}_{k+1} \in \mathbf{R}^{p,p}$ so, daß $\mathbf{W}_{k+1} = \mathbf{Q}_{k+1}\mathbf{R}_{k+1}$

S3 Setze $\mathbf{V}_{k+1} = \mathbf{Q}_{k+1}$

Die Konvergenz entspricht noch der der einfachen Vektoriteration, doch läßt sich zeigen [70], daß in den Teilräumen $\mathcal{V}_k$ Richtungen existieren, die näher an den Eigenrichtungen $\underline{\mathbf{u}}_i$ liegen als die $\underline{\mathbf{v}}_i^{(k)}$ in (B.2) und individuell mit dem Faktor $|\lambda_{p+1}/\lambda_i|$ konvergieren könnten. Diese zu extrahieren, bedarf es allerdings der im übernächsten Abschnitt erwähnten Ritz-Technik. Zuvor sind einige Bemerkungen grundsätzlicher Natur zu vektoriterativen Verfahren angebracht.

B.3 Berechnung beliebiger Eigenwerte

Folgende drei Grundoperationen, die auch miteinander kombiniert werden können, lassen die Eigenvektoren einer Matrix $\mathbf{A}$ bekanntlich unverändert, verändern aber die Eigenwerte:

- Skalare Multiplikation (α reell): $\alpha\mathbf{A} \rightarrow \alpha\lambda_i$
- Spektralverschiebung (β reell): $\mathbf{A} - \beta\mathbf{I} \rightarrow \lambda_i - \beta$
- Potenzierung (k ganz): $\mathbf{A}^k \rightarrow \lambda_i^k$

Die für die Konvergenzgeschwindigkeit maßgeblichen Quotienten $\sigma_j = |\lambda_{p+1}/\lambda_j|$ ($j = 1, \ldots, p$) können daher durch eine geeignete Spektralverschiebung, $\mathbf{A} - \rho\mathbf{I}$, minimiert werden. So werden für $\lambda_1 \geq \ldots \geq \lambda_p > \lambda_{p+1} \geq \ldots \geq \lambda_n$ (keine Beträge!) die σ_j für $\rho_{\text{opt}} = \frac{1}{2}(\lambda_{p+1} + \lambda_n)$ minimal und die minimalen Werte lauten

$$\sigma_j^{\text{opt}} = \frac{1-\tau_j}{1+\tau_j} \quad \text{mit} \quad 0 < \tau_j \equiv \frac{\lambda_j - \lambda_{p+1}}{\lambda_j - \lambda_n} \leq 1. \tag{B.12}$$

Auch lassen sich für $\lambda_1 \geq \ldots \geq \lambda_{n-p-1} > \lambda_{n-p} \geq \ldots \geq \lambda_n$ mit Hilfe von

$$\mathbf{A} - \rho\mathbf{I} \quad \text{mit} \quad \rho \geq \frac{1}{2}(\lambda_1 + \lambda_{n-p-1}) \tag{B.13}$$

die p kleinsten Eigenwerte in dominante verwandeln, man beachte jedoch, daß durch einfache Spektralverschiebungen nur Eigenwerte an den Enden des Spektrums zu dominanten gemacht werden können.

Sind dagegen Eigenwerte aus der Mitte des Spektrums, etwa in der Nachbarschaft eines vorgegebenen Wertes ϵ gesucht (das sind die betragskleinsten von $\mathbf{A} - \epsilon\,\mathbf{I}$), bietet sich die inverse Vektoriteration

$$\mathbf{W}_{k+1} = (\mathbf{A} - \epsilon\,\mathbf{I})^{-1}\mathbf{V}_k \tag{B.14}$$

an, die die Eigenpaare zu den betragsgrößten $(\lambda_i - \epsilon)^{-1}$ liefert, also gerade die zu ϵ nächstbenachbarten λ_i. Insbesondere die *Inverse Rayleigh-Quotienten Iteration*,

$$\underline{\mathbf{v}}^{(k+1)} = (\mathbf{A} - \mu^{(k)}\mathbf{I})^{-1}\underline{\mathbf{v}}^{(k)}, \quad \mu^{(k)} = \left\langle v^{(k)}\middle|\mathbf{A}\middle|v^{(k)}\right\rangle / \left\langle v^{(k)}\middle|v^{(k)}\right\rangle,$$

bei der nach jeder Iteration die Matrix individuell für jede Iterierte zur jeweils letzten Eigenwertnäherung $\mu^{(k)}$ hin verschoben wird (und dadurch immer singulärer wird), ist mit einer kubischen Konvergenz für die Eigenwerte das bestkonvergierendste diesbezügliche Verfahren überhaupt [42, S. 395].

Die direkte Invertierung von $\mathbf{A} - \epsilon\mathbf{I}$ ist gewöhnlich unvertretbar. Stattdessen wird $\mathbf{W}_{k+1}$ durch Lösen des Gleichungssystems

$$(\mathbf{A} - \epsilon\,\mathbf{I})\mathbf{W}_{k+1} = \mathbf{V}_k \tag{B.15}$$

bestimmt. Dabei hat man in Rechnung zu stellen, daß die Koeffizientenmatrix $\mathbf{A} - \epsilon\,\mathbf{I}$ i. allg. indefinit ist, da die Verschiebung des Nullpunktes in das Spektrum hinein zwangsläufig positive und negative Eigenwerte zur Folge hat.

Eine zweite Möglichkeit, Eigenwerte aus der Mitte des Spektrums zu finden, bietet die Iteration mit der quadrierten Form $(\mathbf{A} - \epsilon\,\mathbf{I})^2$. Die kleinsten Eigenwerte von $(\lambda_i - \epsilon)^2$ sind gerade die gesuchten λ_i, die am dichtesten bei ϵ liegen, und diese können immer durch eine geeignete Shift, $(\mathbf{A} - \epsilon\,\mathbf{I})^2 - \rho\mathbf{I}$ mit hinreichend großem $\rho > \frac{1}{2}\max\{(\lambda_i - \epsilon)^2 : \forall i\}$, in dominante verwandelt werden.

Aus den in 4.4.1 genannten Gründen werden wir hier den zweiten Weg über die quadrierte Form $(\mathbf{A} - \epsilon\,\mathbf{I})^2 - \rho\mathbf{I}$ beschreiten.

B.4 Ritz-Technik und die Grundverfahren

Es erhebt sich die Frage, wie den durch die Matrizen $\mathbf{Q}_k$ in (B.2) representierten Teilraumapproximationen $\mathcal{V}_k = \mathcal{R}(\mathbf{Q}_k)$ möglichst gute Näherungen $(\mu_j, \underline{\mathbf{z}}_j)$ für die gesuchten Eigenpaare von $\mathbf{A}$ zugeordnet werden können. Die Antwort gibt der Rayleigh-Ritz-Algorithmus [42, S. 384]:

Sei $\mathcal{V} = \mathcal{R}(\mathbf{Q})$ ein p-dimensionaler Teilraum, der durch die Spalten der spaltenorthonormalen Matrix $\mathbf{Q} \in \mathbf{R}^{n,p}$ aufgespannt wird. Man bilde die symmetrische $(p \times p)$-Matrix

$$\mathbf{P} \equiv \mathbf{Q}^T\mathbf{A}\,\mathbf{Q} \in \mathbf{S}^{p,p}, \tag{B.16}$$

die Projektion von $\mathbf{A}$ auf $\mathcal{V} = \mathcal{R}(\mathbf{Q})$ oder auch Ritz-Ansatz von $\mathbf{A}$ zur Basis $\mathbf{Q}$ heißt, und suche anschließend deren Eigenwertzerlegung

$$\mathbf{P} = \mathbf{X}^T\mathbf{M}\mathbf{X} \tag{B.17}$$

mit $\mathbf{M} = \mathrm{diag}(\mu_1, \ldots, \mu_p)$ und orthogonalem $\mathbf{X} \in \mathbf{R}^{p,p}$. Mit Hilfe von $\mathbf{X}$ kann aus dem Unterraum $\mathcal{R}(\mathbf{Q})$ dann die besondere Basis (Ritz-Basis)

$$\mathbf{Z} \equiv \mathbf{Q}\mathbf{X} \in \mathbf{R}^{n,p} \tag{B.18}$$

mit $\mathcal{R}(\mathbf{Z}) = \mathcal{R}(\mathbf{Q})$ ausgewählt werden[1], die die Quadratsumme der Residuen

$$\|\mathbf{A}\,\mathbf{Z} - \mathbf{Z}\mathbf{M}\|_F^2 = \sum_{j=1}^{p} \|\mathbf{A}\underline{\mathbf{z}}_j - \underline{\mathbf{z}}_j\,\mu_j\|^2 \rightarrow \text{Minimum} \tag{B.19}$$

minimiert, d. h. $\{\mathbf{M},\ \mathbf{Z}\}$ sind im Sinne dieses Kriteriums die bestmöglichen Eigenpaare von $\mathbf{A}$ im Teilraum $\mathcal{R}(\mathbf{Q})$, Beweis z. B. [42, S. 385].

Die Ritz-Eigenpaare $\{\mathbf{M},\ \mathbf{Z}\}$ besitzen eine Reihe besonderer Eigenschaften, von denen wir eine erwähnen wollen: Sind $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ die Eigenwerte von $\mathbf{A}$ und $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ die Eigenwerte des Ritz-Ansatzes $\mathbf{P}$ nach (B.16), so gilt die Ungleichung

$$\lambda_i \geq \mu_i\,, \qquad i = 1,\ \ldots,\ p.$$

Um die p größten Eigenwerte von $\mathbf{A}$ zu berechnen, folgt daraus die Möglichkeit, die Summe der p Eigenwerte des Ritz-Ansatzes durch Variation der Basen $\mathbf{Q}$ oder $\mathbf{V}$ über p-dimensionalen Teilräumen $\mathcal{R}(\mathbf{Q}) = \mathcal{R}(\mathbf{V})$ zu maximieren. Dies führt zu dem Bewertungsfunktional

$$\mu(\mathbf{V}) \equiv \frac{\text{Sp}\,(\mathbf{V}^T\mathbf{A}\mathbf{V})}{\mathbf{V}^T\mathbf{V}}, \tag{B.20}$$

das eine Verallgemeinerung des Rayleigh-Quotienten darstellt, und das als maximalen Wert die Summe der p größten Eigenwerte von $\mathbf{A}$ besitzt.

Eine Reihe von Eigenwertverfahren, die zumeist für das allgemeine Eigenwertproblem $\mathbf{A}\underline{\mathbf{x}} = \lambda\mathbf{B}\underline{\mathbf{x}}$ entwickelt wurden, beruhen tatsächlich auf der Maximierung des Funktionals $\mu(\mathbf{V})$ (Relaxationsprinzip). Die allgemeine Iterationsvorschrift lautet [63, S. 21]

$$\mathbf{V}_{k+1} = \mathbf{V}_k\mathbf{K}_1 + \mathbf{S}\mathbf{K}_2 \tag{B.21}$$

wobei $\mathbf{V}_k \in \mathbf{R}^{n,\,p}$ die Iterierte, $\mathbf{S} \in \mathbf{R}^{n,\,m}$ eine vorgegebene Suchrichtung und $\mathbf{K}_{1/2}$ entsprechende Auswahlmatrizen der Ordnung $(p \times p)$ bzw. $(m \times p)$ sind. Die Verfahren unterscheiden sich in der Wahl der Suchrichtung[2] $\mathbf{S}$ und in der Konstruktion der Matrizen $\mathbf{K}_{1/2}$. Sie sind so ausgelegt, daß $\mu(\mathbf{V}_{k+1}) \geq \mu(\mathbf{V}_k)$ gewährleistet, d. h. beweisbar ist. Von praktischer Relevanz sind nach [63] das „simultane Gradientenverfahren mit optimaler“ oder „pseudooptimaler Auswahl von $\mathcal{R}(\mathbf{V}_{k+1})$“ und entsprechende Verfahren der konjugierten Gradienten.

Prinzipiell fällt hierunter auch der Lanczos-Algorithmus [79], der eine gwisse Sonderstellung insofern einnimmt, als die Suchrichtung nicht nur aus der letzten Iterierten $\mathbf{V}_k$, sondern aus $s \geq 1$ vorangegangen Iterierten $\mathbf{V}_k, \ldots, \mathbf{V}_{k-s-1}$ bestimmt wird. Die theortisch sehr gute Konvergenz wird in der Praxis allerdings dadurch unterminiert, daß bei simultanen Varianten des Lanczos-Algorithmus wesentliche Stabilitätsprobleme auftreten und geeignete quantitative Abbruchskriterien schwer zu finden sind [79][63, S. 37].

Demgegenüber stehen Algorithmen, die nicht explizit die Vergrößerung des Funktionals $\mu(\mathbf{V}_k)$ zum Ziel haben, sondern bei denen wie bei der Teilraumiteration (B.2) bewiesen werden kann, daß die Projektion der Teilräume $\mathcal{V}_k$ auf den Unterraum span($\underline{\mathbf{u}}_1,\ \ldots\ \underline{\mathbf{u}}_p$) der

[1] Für nichtorthonormale Basen $\mathbf{V}$ von $\mathcal{V}$ lautet die Ritz-Projekion $\mathbf{P} \equiv \mathbf{V}^T\mathbf{A}\mathbf{V}/(\mathbf{V}^T\mathbf{V})$, die nach der Eigenwertzerlegung (B.17) dann zu nichtorthogonalen Ritz-Basen $\mathbf{Z}{=}\mathbf{V}\mathbf{X}$ führt, die aber gleichfalls (B.19) befriedigen.

[2] $\mathbf{S} = \mathbf{grad}\,\mu(\mathbf{V}_k)$ bezeichnet man als optimale Suchrichtung, daneben existieren Verfahren mit näherungsweise optimalem $\mathbf{S}$ und Verfahren der Koordinatengruppenrelaxation. Letztere allerdings scheinen kaum praktische Bedeutung erlangt zu haben.

gesuchten Eigenvektoren schrittweise verbessert wird. Der wichtigste Vertreter ist hier die bekannte Simultane Iteration, die auf BAUER [5] und RUTISHAUSER [70, 71] zurückgeht und durch ihre Einfachheit und Kompaktheit besticht. Der Iterationsablauf kann durch

S1 $\mathbf{W}_{k+1} = \mathbf{A}\mathbf{V}_k$
S2 Konstruktion der $(p \times p)$-Matrix $\mathbf{T}$
S3 $\mathbf{V}_{k+1} = \mathbf{W}_{k+1}\mathbf{T}$

dargestellt werden, wobei verschieden Varianten sich dann hinsichtlich Schritt S2 unterscheiden. Z. B. kann hier der Rayleigh-Ritz-Algorithmus eingeschoben werden mit $\mathbf{X}$ gemäß (B.17) als $\mathbf{T}$. Die Konvergenz gegen die Eigenvektoren von $\mathbf{A}$ läßt sich bereits für beliebiges nichtsinguläres $\mathbf{T}$ und nichtdefektives $\mathbf{A}$ zeigen [63].

Entscheidend ist, daß bei geeignetem $\mathbf{T}$ der individuelle Konvergenzfaktor der Vektoren $\underline{\mathbf{v}}_i^{(k)}$ (der Spalten von $\mathbf{V}_k$) je Schritt auf den Wert $|\lambda_{p+1}/\lambda_i|$ gegenüber $|\lambda_{i+1}/\lambda_i|$ bei der einfachen Teilraumiteration verbessert werden kann. Nunmehr wird es auch günstiger, einige Vektoren mehr als nur die gesuchten zu iterieren, $p = s + m$, s sei die Zahl der gesuchten, m die der zusätzlich Iterierten, da die Iterationslücke $|\lambda_i - \lambda_{p+1}|, i = 1, \ldots, s$ und damit die Konvergenzgeschwindigkeit der gesuchten Eigenpaare dadurch gewöhnlich erhöht wird.

Die in Rede stehenden Anwendung wird die theoretische Grenze der Verfahren tangieren, weshalb wichtigstes Auswahlkriterium die zu erwartende Konvergenzgeschwindigkeit sein sollte, um Iterationsbedarf und damit Rundungsfehler zu minimieren. Während bei den projektiven Verfahren wie der Simultanen Iteration entsprechende Abschätzungen gut zugänglich sind, sind für Relaxationsverfahren Aussagen zur Konvergenzgeschwindigkeit schwierig bzw. in allgemeiner Form nicht bekannt [63]. Letztlich könnten nur Testrechnungen sicheren Aufschluß geben.

In [63] werden alle relevanten Verfahren anhand ausgewählter Fallbeispiele miteinander verglichen. Relaxationsverfahren benötigen einen etwas höheren Aufwand, besitzen gegenüber den projektiven Methoden aber den Vorteil, daß sie keine Spektrumsinformationen für eine optimale Konvergenz benötigen und im allgemeinen schneller konvergieren. Das Ausmaß des Konvergenzvorsprungs hängt allerdings ab von der konkreten Gestalt der Matrix.

Insbesondere in der Anwendung auf Schwingungsprobleme wurden mit der Simultanen Iteration gute Ergebnisse erzielt [62, 63, 4, 37]. Für ein optimal verschobenes Spektrum (B.12) – das Intervall der „unerwünschten" Eigenwerte liegt symmetrisch zum Nullpunkt – erreicht die Simultane Iteration in Kombination mit der Tschebyscheff-Beschleunigung (vgl. den folgenden Abschnitt) hier die Werte der Relaxationsverfahren. Auch wird in [63, S. 63] gezeigt, daß beim speziellen Eigenwertproblem die asymptotische Konvergenzgeschwindigkeit des mächtigen „simultanen Gradientenverfahrens mit optimaler Wahl der Suchrichtung" identisch wird mit der einer „optimal verschobenen" Simultanen Iteration.

In Anbetracht der relativ einfachen Implementation des Algorithmus und der Tatsache, daß wir dank geeigneter exakter Vergleichslösungen über ausreichende Spektrumsinformationen verfügen, legen wir unserem Verfahren daher die Simultane Iteration zugrunde und zwar speziell die Variante RITZIT von RUTISHAUSER [70, 71], die in einer sorgfältigen Ausführung Simultane Iteration und Tschebyscheff-Beschleunigung miteinander kombiniert. Bevor das Gesamtverfahren besprochen wird, skizzieren wir deshalb noch kurz die Tschebyscheff-Iteration, die entscheidend die Konvergenzgeschwindigkeit von RITZIT bestimmt.

B.5 Die Tschebyscheff-Iteration

Eine k-fache Iteration $\mathbf{V}_{s+k} = \mathbf{A}^k \mathbf{V}_s$ mit dem Konvergenzfaktor $|\lambda_{p+1}/\lambda_p|^k$ kann als Spezialfall einer Polynomiteration k-ter Ordnung $\mathbf{V}_{s+k} = P_k(\mathbf{A})\mathbf{V}_s$ aufgefaßt werden. Fragt man verallgemeinernd nach einem Polynom k-ter Ordnung, das im Verbund mit einer geeigneten linearen Transformation $\lambda \rightarrow \tilde{\lambda}$ den für die Konvergenzgeschwindigkeit dann maßgeblichen Wert $|P_k(\tilde{\lambda}_{p+1})/P_k(\tilde{\lambda}_p)|$ minimiert, wird man auf die Tschebyscheff-Polynome

$$T_k(x) \equiv \frac{1}{2}\left[\left(x+\sqrt{x^2-1}\right)^k + \left(x-\sqrt{x^2-1}\right)^k\right], \quad (k \geq 0) \tag{B.22}$$

mit den Eigenschaften

$$\begin{aligned}
T_k(x) &= \cos(k \arccos x) \quad \forall x \in [-1,+1] && \text{(B.23)}\\
T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x) \quad \text{(3-Term-Rekursion)} && \text{(B.24)}\\
|T_k(x)| &\leq 1 \quad \forall x \in [-1,+1] && \text{(B.25)}\\
|T_k(x)| &> 1 \quad \forall x \notin [-1,+1] \quad \text{(wächst stark mit } k\text{)} && \text{(B.26)}
\end{aligned}$$

geführt, deren k reelle Nullstellen alle im Intervall $[-1, 1]$ liegen. Die $T_k(x)$ sind dadurch ausgezeichnet, daß bei ihnen das Maximum des absoluten Betrages im Intervall $[-1, 1]$ den kleinsten Wert annimmt, der bei einem reellen Polynom k-ter Ordnung und gleichem höchsten Koeffizienten überhaupt möglich ist [20, S. 75].

Angenommen, die „unerwünschten" Eigenwerte $\lambda_{p+1}, \ldots, \lambda_n$ liegen alle im Intervall $[a, b]$ und alle gesuchten $\lambda_1, \ldots, \lambda_p$ außerhalb von $[a, b]$. Mit Hilfe der linearen Transformation $[a, b] \rightarrow [-1, 1]$

$$\tilde{\mathbf{A}} = \frac{2}{b-a}\left[\mathbf{A} - (a+b)\,\mathbf{I}\right] \tag{B.27}$$

werden die „unerwünschten" Eigenwerte in das Intervall $[-1, 1]$ transformiert und die $\lambda_1, \ldots, \lambda_p$ bleiben außerhalb. Eine nunmehr mit $\tilde{\mathbf{A}}$ durchgeführte Tschebyscheff-Iteration

$$\mathbf{V}_{s+k} = T_k(\tilde{\mathbf{A}})\mathbf{V}_s \tag{B.28}$$

die erstmals von RUTISHAUSER in [70, 71] beschrieben wurde, verbessert aufgrund der ausgezeichneten Eigenschaft der $T_k(x)$ die Konvergenzgeschwindigkeit dann zum Teil beträchtlich. Die rekursive Darstellung (B.24) der $T_k(x)$ ermöglicht zudem eine sehr ressourcensparende Implementierung, da ein kompletter Schritt $T_k(\tilde{\mathbf{A}})\mathbf{V}_s$ für beliebiges k bereits mit zwei zusätzlichen Vektoren der Dimension n realisiert werden kann.

Ein Nachteil allerdings ist, daß Spektrumsinformationen über das Intervall $[a, b]$ vonnöten sind, die i. allg. zu Beginn noch nicht vorliegen und erst im Laufe der Iteration gewonnen werden. In der Realisierung von RUTISHAUSER in der Prozedur RITZIT [71] für indefinite $\mathbf{A}$ wird dies wie folgt gelöst: Die ohnenhin anfallenden Eigenwertapproximationen μ_p des p-ten Vektors $\underline{\mathbf{v}}_p$ (der letzten Spalte von $\mathbf{V}_s$) liefern wegen $e \equiv |\mu_p| \leq |\lambda_p|$ eine untere Schranke für die Beträge der dominanten Eigenwerte, d. h. diese liegen stets außerhalb $[-e, e]$. Vereinfachend wird daher das symmetrische Intervall $[-e, e]$ als das Intervall der „unerwünschten" Eigenwerte $[a, b]$ angesehen und der linearen Transformation (B.27) zugrundegelegt, die sich damit auf $\tilde{\mathbf{A}} = \frac{1}{e}\mathbf{A}$ reduziert und zur rekursiven Darstellung

$$\begin{aligned}
\mathbf{V}_{s+1} &= \tfrac{1}{e}\mathbf{A}\mathbf{V}_s\\
\mathbf{V}_{s+j} &= \tfrac{2}{e}\mathbf{A}\mathbf{V}_{s+j-1} - \mathbf{V}_{s+j-2}, \quad j = 2, 3, \ldots
\end{aligned} \tag{B.29}$$

führt. Die Wahl $[-e,\, e]$ ist optimal, wenn $[a,\, b]$ tatsächlich symmetrisch zum Nullpunkt liegt, und um so schlechter, je exzentrischer $[a,\, b]$ ausfällt, also z. B. für $\lambda_i \geq 0\ \forall i$.

In der von uns angestrebten Anwendung kann $[a,\, b]$ nun allerdings schon vorher hinreichend genau bestimmt werden, so daß wir $[a,\, b]$ bereits zu Beginn durch eine entsprechende Verschiebung $\mathbf{A} - \frac{1}{2}(a+b)\,\mathbf{I}$ symmetrisieren können. In diesem Fall wird die symmetrische Implementation $[-e,\, e] \rightarrow [-1,\, 1]$ sogar zur besseren Variante, da gegenüber (B.27) die interne Verschiebung entfällt, die bei jedem Rekursionschritt (B.29) sonst hinzukäme[3].

Die Konvergenzrate $T_k(\tilde{\lambda}_{p+1})/T_k(\tilde{\lambda}_j)$ der j-ten Spalte von $\mathbf{V}_{s+k}$ nach einer k-schrittigen $T_k(\frac{1}{e}\mathbf{A})\mathbf{V}_s$-Iteration beträgt mit (B.22) und wenn näherungsweise $\lambda_{p+1} = e$ gesetzt wird:

$$\frac{T_k(1)}{T_k(\tilde{\lambda}_j)} = \frac{2}{\left(\tilde{\lambda}_j + \sqrt{\tilde{\lambda}_j^2 - 1}\right)^k + \left(\tilde{\lambda}_j - \sqrt{\tilde{\lambda}_j^2 - 1}\right)^k}, \qquad \tilde{\lambda}_j \equiv \frac{\lambda_j}{e} > 1. \tag{B.30}$$

Rutishauser approximiert diesen Ausdruck durch $\left(\tilde{\lambda}_j - \sqrt{\tilde{\lambda}_j^2 - 1}\right)^k$ und kann hernach den Konvergenzfaktor in der üblichen Weise 'pro Schritt' angeben. In unseren Fällen sehr schwacher Konvergenz ist diese Abschätzung jedoch etwas grob, weshalb der vollständige Ausdruck einmal angeschrieben wurde.

B.6 Das schließliche Verfahren im Ganzen

B.6.1 Grundschema

Zu finden sind einige wenige Eigenpaare $\{\lambda_j,\, \underline{\mathbf{v}}_j\}$ mit $\lambda_j \approx \epsilon$ der Matrix $\mathbf{A} \in \mathbf{S}^{n,\,n}$. Wird der Energienullpunkt auf dem Boden der Potentialkästen gelegt, ist $\mathbf{A}$ positiv definit, und wir indizieren

$$0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n. \tag{B.31}$$

Gemäß den Erörterungen in 4.4.1 bestimmen wir die Eigenlösungen um ϵ durch Berechnung der kleinsten Eigenwerte der quadrierten Matrix $(\mathbf{A} - \epsilon\,\mathbf{I})^2$. Als iteratives Eigenwertverfahren hierzu dient uns die Prozedur RITZIT, die auf direktem Wege allerdings nur die dominanten Eigenpaare einer Matrix liefert, weshalb die gesuchten Eigenwerte zuvor noch vermöge der Verschiebung

$$\mathbf{B} = (\mathbf{A} - \epsilon\,\mathbf{I})^2 - \rho\,\mathbf{I} \tag{B.32}$$

in dominante zu verwandeln sind. Um keine Konvergenzgeschwindigkeit zu verschenken, sollte der Wert ρ dabei so klein wie möglich gewählt werden, d. h. nur so groß, daß die gesuchten Eigenwerte gerade zu dominanten werden. In jedem Falle muß

$$\rho > \frac{1}{2}\max\{(\lambda_j - \epsilon)^2 : \forall j\}. \tag{B.33}$$

sein. Weil praktisch immer $\lambda_n/2 > \epsilon$, sogar $\lambda_n/2 \gg \epsilon$, gewährleistet ist, hat man ρ am maximalen Eigenwert λ_n von $\mathbf{A}$ auszurichten, zweckmäßigerweise über

$$\rho \;=\; \eta(\lambda_n - \epsilon)^2 \;\approx\; \eta(\lambda_{\max} - \epsilon)^2 \quad \text{mit} \quad \eta > 0.5, \tag{B.34}$$

[3]Dann nämlich $\mathbf{V}_{s+j} = \frac{4}{b-a}\mathbf{A}\mathbf{V}_{s+j-1} - 2\frac{b+a}{b-a}\mathbf{V}_{s+j-1} - \mathbf{V}_{s+j-2}$. Der tiefere Grund, warum Verschiebungen von $\mathbf{A}$ vorher meist billiger sind, besteht darin, daß alle diese iterativen Verfahren intern mit einer Matrix-mal-Vektor-Regel auskommen sollen (und können) und nicht an $\mathbf{A}$ manipulieren dürfen, während extern die Hauptdiagonale oftmals physisch vorhanden ist.

also mit Hilfe eines – möglichst knappen – Faktors η, der folglich eine Schätzung der Eigenwertabstände nahe ϵ beinhaltet. Bei der Dichte des Spektrums genügt meist $\eta \approx 0.55$.

Für den maximalen Eigenwert λ_n in (B.34) bedarf es wie dort angedeutet nur einer Näherung $\lambda_{\max}$, die entweder auf numerischem Wege durch „grobe“ Iteration mit $\mathbf{A}$ erhalten werden kann oder aber wie im Falle der rechteckigen und kubischen Grundgebiete schneller noch aus der analytischen Gitterlösung des unendlich hohen Kastens (A.6):

$$\lambda_{\max} = \frac{4}{h^2}\left[\cos^2\frac{\pi}{2(n_x+1)} + \cos^2\frac{\pi}{2(n_y+1)} + \ldots\right] + V_0; \tag{B.35}$$

$n_{x,y,\ldots}$ sind hier die Zahlen der inneren Gitterpunkte je Koordinatenachse und gegebenenfalls hat man um den n_z–Term der dritten Dimension zu erweitern. Mit dem V_0–Addenten ist dies sogar eine strenge obere Schranke, da jede in die V_0–Potentialebene eingeprägte Mulde die Energie nur verringern kann.

Für das Gesamtverfahren erhalten wir damit folgendes Schema:

Schema des Eigenwert-Gesamtverfahrens:

- S1 Diskretisierung: Belege Matrix $\mathbf{A}$ entsprechend gewähltem Potential und Geometrie
- S2 Bestimme näherungsweise maximalen Eigenwert von $\mathbf{A}$ anhand einer exakten Vergleichslösung oder durch „grobe“ Iteration: $\lambda_{\max} \approx \lambda_n$
- S3 Verschiebung zur Energie ϵ: $\mathbf{A}_\epsilon \equiv \mathbf{A} - \epsilon\,\mathbf{I}$
- S4 Quadrierung: $\mathbf{A}_\epsilon^2$
- S5 Verwandle minimale Eigenwerte von $\mathbf{A}_\epsilon^2$ in dominante: $\mathbf{B} \equiv \mathbf{A}_\epsilon^2 - \rho\,\mathbf{I}$ mit $\rho = \eta(\lambda_{\max} - \epsilon)^2$ und hinreichend großem $\eta > 0,5$
- S6 Berechne die p dominanten Eigenpaare $\{\beta_j, \underline{\mathbf{v}}_j\}$ von $\mathbf{B}$ (Prozedur RITZIT)
- S7 Bestimme die Eigenwerte von $\mathbf{A}$ über die Rayleigh-Quotienten $\lambda_j = \underline{\mathbf{v}}_j^T\,\mathbf{A}\,\underline{\mathbf{v}}_j$

Die Prozedur RITZIT benötigt eine Matrix-mal-Vektor-Regel $\underline{\mathbf{w}} = \mathbf{B}\underline{\mathbf{v}}$. Die Problematik dieser Algorithmen wurde in 4.4.2 erörtert. Üblicherweise wird die Operation $\mathbf{B}\underline{\mathbf{v}}$ auf zwei rekursive Aufrufe von $\mathbf{A}\underline{\mathbf{v}}$ zurückgeführt, so daß nur die erste Verschiebung $\mathbf{A} - \epsilon\,\mathbf{I}$ (der Schritt S3) auch tatsächlich physisch ausgeführt werden kann, während Quadrierung und Verschiebung mit ρ (die Schritte S4 + S5) virtuell von der Matrix-mal-Vektor-Regel $\underline{\mathbf{w}} = \mathbf{B}\underline{\mathbf{v}}$ übernommen werden müssen, mit der RITZIT operieren soll.

B.6.2 Die Prozedur RITZIT

Die original in ALGOL notierte Prozedur RITZIT wurde nach C umgeschrieben. Dabei konnte der Ressourcenbedarf für die Tschebyscheff-Iteration von den ursprünglich drei auf die Mindestzahl von zwei zusätzlichen Vektoren verringert werden. Das Schema von RITZIT wird nur soweit dargeboten, wie es für eine Konvergenzaussage notwendig ist, für Einzelheiten sei auf die Originalliteratur [70, 71] verwiesen.

Grundschema der Prozedur RITZIT:

Berechnung der p dominanten Eigenpaare $\{\beta_i, \underline{\mathbf{v}}_i\}$ der Matrix $\mathbf{B}$ mit Hilfe einer Matrix-mal-Vektor-Regel.

- S1 Initialisierung: Wähle spaltenorthonormales $\mathbf{V}_0$. Setze $k = 0,\ m = 2$
- S2 Tschebyscheff-Iteration: $\mathbf{V}_{k+m-1} = T_{m-1}(\mathbf{B})\mathbf{V}_k$
- S3 Orthonormalisiere Spalten von $\mathbf{V}_{k+m-1}$

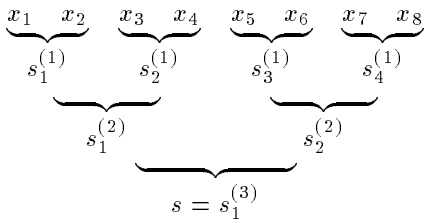


Abbildung B.1: *Grundprinzip der Kaskadensummation am Beispiel $n = 8$, die allgemein für $n = 2^p$ nur eines zusätzlichen REAL-Feldes der Länge p bedarf. Für beliebiges n sind entsprechend benachbarte Kaskaden auszuführen. Kaskadensummationen sind prädestiniert für Parallelverarbeitung.*

S4 Ritz-Schritt: Konstruiere $(p \times p)$-Matrix $\mathbf{T}$ und setze $\mathbf{V}_{k+m} = \mathbf{V}_{k+m-1}\mathbf{T}$
S5 Justiere m, Abbruchtests
S6 $k = k + m$, goto S1

Die Zahl m legt hier die Ordnung der Tschebyscheff-Polynome T_{m-1} fest, d. h. mit anderen Worten die Intervalle, in denen die aufwendigeren Ritz-Schritte samt Orthonormierung eingeschoben werden. Zu Beginn der Iteration ist $m = 2$ und wächst dann stark, bei schwacher Konvergenz ungefähr mit dem Faktor 2. Die Matrix $\mathbf{T}$ des Ritz-Schrittes ist so konstruiert, daß die individuelle Konvergenz der Vektoren mit dem Quotienten $|\beta_{p+1}/\beta_j|$ – nach einem Tschebyscheff-Gesamtschritt also mit dem Faktor (B.30) – erfolgt.

Ein bisher wenig erwähnter, jedoch sehr wichtiger Punkt ist die durch die hohe Dimension der Vektoren nicht mehr zu vernachlässigende Fehlerakkumulation bei Summationen über n, worunter alle Skalarprodukte $<\underline{\mathbf{w}}|\underline{\mathbf{v}}>$ leiden und in ganz besonderem Maße natürlich die Schmidt-Orthogonalisierung. Werden n Zahlen x_i in der üblichen rekursiven Form

FOR $i := 1$ **TO** n **DO** $s := s + x_i$

aufsummiert, wachsen die Rundungsfehler proportional zu $(n-1)\nu$ mit ν als relativem Rundungsfehlerniveau der Computerarithmetik und $\nu = 0.5 \cdot 10^{1-t}$ bei t dezimalen Mantissenstellen und symmetrischer Rundung [42, Kapitel 2]. Tatsächlich ließen sich mit 14–16 Mantissenstellen (8-Byte-Zahlen) bei unseren Dimensionszahlen $n > 10^4$ auf diese Weise überhaupt keine vernünftigen Resultate mehr erzielen, gleichwohl das Verfahren konvergierte, d. h. Testgrößen wie Rayleigh-Quotienten usw. stagnierten. Die Wellenfunktionen widersprachen jedoch der Anschauung, ließen angeforderte Symmetrien vermissen oder zeigten Unstetigkeiten. Wenigstens blieben die Residuen $\mathbf{A}\underline{\mathbf{v}} - \lambda\underline{\mathbf{v}}$ als numerisch keineswegs unproblematisches Gütekriterium oberhalb 10^{-2} eV.

Abhilfe schaffte die sogenannte binäre oder Kaskadensummation [42, S. 93], deren Grundprinzip Abb. B.1 illustriert. Die Rundungsfehler[4] wachsen hier nur proportional zu $\log_2(n)\nu$, für $n = 2^{17} = 131\,072$ also 17ν statt $131\,071\nu$. RITZIT wurde daher auf Kaskadensummation umgestellt.

[4]Genauer wird die Computerrealisierung einer Summation zum Fehlerniveau ν durch $s = \sum_{i=1}^{n} x_i \left(1 + \varepsilon_i^{(n)}\right)$ beschrieben, wobei die $\varepsilon_i^{(n)}$ die relativen Störungen sind, die die einzelnen x_i insgesamt erleiden. Bei der Standardsummation genügt $\varepsilon_i^{(n)}$ der Ungleichung $|\varepsilon_i^{(n)}| \leq (n-1)\nu$, während für Kaskadensummation $|\varepsilon_i^{(n)}| \leq \log_2(n)\nu$ gilt [42, Kap. 2].

B.7 Zur Konvergenzgeschwindigkeit

B.7.1 Allgemeine Formulierung

Zum Problematischen an praktischen Abschätzungen der Konvergenzgeschwindigkeit gehört, daß die diesbezüglichen theoretische Aussagen zumindestens der relevanten Eigenwerte des Spektrums (z. B. der gesuchten) bedürfen, währenddessen solche Eigenwerte gerade erst berechnet werden sollen. Insbesondere die Vorhersage der für die Konvergenz entscheidenden *Iterationslücke* – Abständen über 4 bis 10 benachbarte Eigenwerte hinweg – birgt viel Unsicherheit, da dort z. B. hineinspielt, ob man ein reguläres und hinsichtlich der Nächste-Nachbar-Statistik (NNS) damit Poisson-verteiltes, oder ein nichtreguläres und damit eher Wigner-verteiltes, steiferes Spektrum erwartet. Insofern wird bei gewissen Anwendungen erst das „numerische Experiment" das letzte Wort sprechen.

Eine zusätzliche Schwierigkeit – und mit der Hauptgrund für die folgende, etwas ausführlichere Analyse – bereitet in unserem Fall die Tatsache, daß mit der quadrierten Matrix (B.36) iteriert wird, Eigenwertschätzungen aber nur für die unquadrierte, ursprüngliche Matrix $\mathbf{A}$ sinnvoll möglich sind, so daß sämtliche Aussagen dementsprechend übertragen werden müssen.

Zur Rekapitulation: Gesucht waren Eigenwerte λ_k nahe ϵ der positiv definiten Matrix $\mathbf{A}$ mit der Indizierung (B.31), iteriert wird jedoch mit der Matrix

$$\mathbf{B} = (\mathbf{A} - \epsilon\mathbf{I})^2 - \rho\mathbf{I}, \tag{B.36}$$

deren Eigenwerte β_j im folgenden nach absteigenden Beträgen geordnet angenommen werden, wobei die ersten p Eigenwerte dominant sein sollen:

$$|\beta_1| \geq \ldots \geq |\beta_p| > |\beta_{p+1}| \geq \ldots \geq |\beta_n| \,. \tag{B.37}$$

Eine Konvergenzbetrachtung hat demgemäß das $\{\beta_j\}$-Spektrums zu untersuchen.

Abschätzungen der Konvergenzgeschwindigkeit nehmen das Streben der Eigenwertnäherungen $\beta_j^{(\ell)}$ (ℓ ist der Iterationsindex) gegen die exakten Eigenwerte β_j zum Ausgangspunkt und werden gewöhnlich in Form eines asymptotischen Konvergenzquotienten ξ_j gegeben [42, Kap. 13]:

$$\frac{|\beta_j - \beta_j^{(\ell+1)}|}{|\beta_j - \beta_j^{(\ell)}|} \leq \xi_j \,, \quad j = 1, \ldots, p. \tag{B.38}$$

(B.38) besagt, daß der Startfehler mit jeder Iteration um den Faktor ξ verringert wird, mit anderen Worten, soll das Verfahren solange laufen, bis die relativen Änderungen des Fehlers nur noch $\Sigma = \xi^N$ betragen, sind etwa $N = \log(\Sigma)/\log(\xi)$ Iterationen erforderlich. Die Größe $-\log_{10}(\Sigma)$ kann als Anzahl gültiger Dezimalstellen interpretiert werden.

Die individuelle Konvergenz der Iterierten $\underline{\mathbf{v}}_j^{(\ell)}$ gegen die Eigenvektoren $\underline{\mathbf{u}}_j$ wird anhand des Kleinwerdens der Norm $\|\underline{\mathbf{u}}_j - \underline{\mathbf{v}}_j^{(\ell)}\|$ oder des Winkels $\varphi_\ell = \sphericalangle\,(\underline{\mathbf{u}}_j, \underline{\mathbf{v}}_j^{(\ell)}) \in [0, \pi)$ beschrieben:

$$\frac{\|\underline{\mathbf{u}}_j - \underline{\mathbf{v}}_j^{(\ell+1)}\|}{\|\underline{\mathbf{u}}_j - \underline{\mathbf{v}}_j^{(\ell)}\|} \leq \sigma_j \qquad \text{oder} \qquad \frac{\tan\varphi_{\ell+1}}{\tan\varphi_\ell} \leq \sigma_j \,. \tag{B.39}$$

Bei der Simultanen Iteration gilt nun $\xi_j = \sigma_j^2$ [70], d. h. die Eigenwerte konvergieren theoretisch doppelt so schnell. Wir haben uns hier an den „langsameren" σ_j zu orientieren, da wir die Eigenvektoren ebenfalls benötigen. Auch werden bei der Simultanen Iteration zur

Konvergenzbeschleunigung stets einige Vektoren mehr als nur die gesuchten iteriert. So sei s die Zahl der gesuchten und m, typischerweise zwischen 2 und 10, die Zahl der zusätzlich Iterierten:

$$p = s + m\,. \tag{B.40}$$

Die Konvergenzrate (B.30) der Prozedur RITZIT erfordert nunmehr die Quotienten

$$|\beta_{p+1}/\beta_j| < 1\,, \quad j = 1,\,\ldots,\,s \le p\,, \tag{B.41}$$

der ersten s unter den p Iterierten, wobei wir uns auf den Quotienten von β_s,

$$\frac{|\beta_{p+1}|}{|\beta_s|} \;\equiv\; 1 - \frac{\Delta\beta_{s,m}}{|\beta_s|} \;\equiv\; 1 - R, \tag{B.42}$$

konzentrieren können, da unter den gesuchten Eigenwerten keiner schlechter als dieser konvergiert. Für die *Iterationslücke* im $\{\beta_j\}$-Spektrum wurde hier das Symbol

$$\Delta\beta_{s,m} \;\equiv\; |\beta_s| - |\beta_{s+m+1}| \;\;\ge\;\; 0 \tag{B.43}$$

eingeführt und mit

$$R \;=\; \Delta\beta_{s,m}/|\beta_s| \;\approx\; \Delta\beta_{s,m}/\rho \tag{B.44}$$

die Differenz des Konvergenzquotienten zu Eins bezeichnet. Die hierdurch implizierte Trennung in Eigenwertabstände $\Delta\beta_{s,m}$ und Absolutgrößen $|\beta_j|$ in (B.42) erscheint zweckmäßig, da die $\Delta\beta_{s,m}$ durch die physikalischen λ-Eigenwertabstände nahe ϵ determiniert sind, wohingegen die absoluten Zahlenwerte der $|\beta_j|$ aufgrund von $\lambda \approx \epsilon$ im wesentlichen $\rho = \eta(\lambda_{\max} - \epsilon)^2$, also reine Diskretisierungsgrößen verkörpern. Insbesondere geht der maximale Gittereigenwert $\lambda_{\max}$ mit $1/h^2$, so daß praktisch immer wie in (B.44) $\Delta\beta/|\beta_j| \approx \Delta\beta/\rho$ gesetzt werden kann.

Ein Zahlenbeispiel soll die Verhältnisse verdeutlichen: Für rechteckige und kubische Einbettungsgebiete und nicht zu große Schrittweiten ist $\lambda_{\max}$ nach (A.6) nahezu exakt. Mit $h = 0.4$ Å entsteht im zweidimensionalen Fall $\lambda_{\max} \approx 198$ eV und somit $\rho \approx 3.92 \times 10^4\ (\mathrm{eV})^2$, während für angenommenes $|\Delta\lambda| \approx 0.1$ eV ein $\Delta\beta \approx 0.01\,(\mathrm{eV})^2$ zu erwarten ist. Die damit implizierte Größenordnung $R \approx 10^{-7}$ stimmt derart bedenklich, daß wir unbedingt verläßlichere Aussagen beschaffen müssen, um die Grenzen des Verfahrens zu bestimmen.

B.7.2 Formulierung für die quadratische Iteration

Obzwar das $\{\beta_j\}$-Spektrum der quadrierten Matrix letztlich über die Konvergez entscheidet, wird eine Schätzung oder ein lösbares Modellsystem primär die Eigenwerte des ursprünglichen (physikalischen) $\{\lambda_k\}$-Spektrums bereitstellen, z. B. innerhalb eines Energieintervalls

$$\lambda_k \in [E_{\min},\, E_{\max}], \tag{B.45}$$

in dem die gesuchten λ_k erwartet werden. Die β_j werden entsprechend (B.36) dann eindeutige Funktionen dieser λ_k sein. Allerdings sollte die Zuordnung $\lambda_k \rightarrow \beta_j$ für unsere Zwecke von den j-Indizes der β_j her gedacht werden, da die Iteration primär die β_j, nämlich $\beta_1 \ldots \beta_s$, liefert. Eine solche Zuordnung schreiben wir wie folgt:

$$\beta_j \;=\; (\lambda_{\beta(j)} - \epsilon)^2 - \rho\,. \tag{B.46}$$

Hier wurde formal die Index-Zuordnungsvorschrift $k = \beta(j)$ eingeführt, die für einen β_j–Wert, der im Sinne von $\beta_j = f(\lambda_k, \epsilon)$ von λ_k abstammt, eben diesen Index k bereitstellt. $\beta(j)$ wird immer mit einem Klammernpaar erscheinen, so daß Verwechslungen mit den Eigenwerten β_j ausgeschlossen sind. Eine gleichnamige Indizierung – $\beta_j(\lambda_j)$ – verbot sich oben, da (B.31) und (B.37) simultan gelten sollen und über eine Indexmenge $\{j\}$ ja nur einmal, im Sinne einer Ordnung der β_j oder der λ_j, verfügt werden kann.

Die konkrete Abbildung $\beta_j \rightarrow \lambda_{\beta(j)}$ ist natürlich ϵ–abhängig. Die Zweideutigkeit, wenn ϵ in die Mitte zweier λ_k–Werte fällt, ist nur scheinbar, da die Kenntnis des $\{\lambda_k\}$-Spektrums ja bereits angenommen wird. In jedem Fall zieht $|\beta_j| \geq |\beta_{j+1}|$ die Ungleichung $|\lambda_{\beta(j)} - \epsilon| \leq |\lambda_{\beta(j+1)} - \epsilon|$ nach sich, so daß insbesondere $\lambda_{\beta(1)}$ den oder einen der zu ϵ nächstbenachbarten λ_k bezeichnet. Eine fortlaufende Indexfolge $j = 1, 2, \ldots$ bei β_j findet jedoch im allg. keine Entsprechung bei den $\lambda_{\beta(j)}$–Indizes, da für die zugehörigen λ_k's lediglich deren wachsende Entfernung zu ϵ gewiß ist, sie aber (in a priori nicht bekannter Abfolge) wechselseitig ober- als auch unterhalb ϵ liegen können. Ein Indexabstand m bei β_{j+m} kann daher innerhalb $\{\lambda_k\}$ gleich oder größer m sein, niemals aber kleiner.

Der optimale Wert für die Verschiebung ρ in (B.36) kann jetzt ganz in Analogie zu (B.12) formal exakt als Funktion der λ_k angegeben werden:

$$\rho_{\rm opt} = \frac{1}{2}\left[(\lambda_n - \epsilon)^2 + (\lambda_{\beta(p+1)} - \epsilon)^2\right] \quad \text{für} \quad \epsilon \leq \lambda_n/2. \tag{B.47}$$

Die zugehörigen optimalen $|\beta_{p+1}/\beta_s|$ bzw. R–Werte lauten

$$\left|\beta_{p+1}/\beta_s\right|_{\rm opt} = \frac{1-\tau_s}{1+\tau_s} \qquad \text{und} \qquad R_{\rm opt} = \tau_s^2 \tag{B.48}$$

mit

$$0 < \tau_s \equiv \frac{(\lambda_{\beta(s)} - \epsilon)^2 - (\lambda_{\beta(p+1)} - \epsilon)^2}{(\lambda_{\beta(s)} - \epsilon)^2 - (\lambda_n^2 - \epsilon)^2} \leq 1, \tag{B.49}$$

und für $\Delta\beta_{s,m}$ kann geschrieben werden:

$$\Delta\beta_{s,m} = (\lambda_{\beta(s)} - \epsilon)^2 - (\lambda_{\beta(p+1)} - \epsilon)^2. \tag{B.50}$$

Bemerkung: Aufgrund des quadratischen $\beta_j(\lambda_k)$–Zusammenhangs wird unabhängig von der Zahl m die Konvergenz tendenziell verbessert, je größer man s wählt (je mehr Eigenwerte simultan gesucht werden), da mit zunehmendem Abstand der $\lambda_{\beta(s)}$ vom ϵ–Scheitel in (B.46) gewöhnlich auch $\Delta\beta_{s,m}$ und – etwas abgeschwächt – τ_s anwachsen (immer vorbehaltlich der Möglichkeit, daß der „Puffer“ $\beta_{s+1}, \ldots, \beta_{p+1}$ komplett in einen Eigenwerthaufen geraten kann). Dieser Effekt steht in Konkurenz zur sinkenden Güte der Orthogonalisierung bei zunehmender Vektorenzahl und sollte quantifiziert werden.

Im nächsten Abschnitt werden wir konkrete Konvergenzdaten an einem lösbaren, zweidimensionalen Modellsystem erheben. Angesichts der im allgemeinen statistischen Phänomenologie der Eigenwertverteilung in derartigen Systemen erscheint eine statistische Auswertung förderlich. Das Beobachtungsintervall (B.45) werden wir daher etwas großzügiger wählen,

$$\lambda_k \in [0, V_0], \quad V_0 = 8 \text{ eV},$$

(es soll also die Berechnung beliebiger gebundener Zustände in Betracht gezogen werden), und innerhalb von $[0, V_0]$ für verschiedene s und m dann die minimalen Iterationslücken $\Delta\beta_{s,m}^{\rm min}$ suchen, definiert wie folgt:

$$\Delta\beta_{s,m}^{\rm min} \equiv \min(\Delta\beta_{s,m}) \;\forall\; \lambda_{\beta(1)}, \lambda_{\beta(p+1)} \in [0,\, V_0] \;\wedge\; \forall\epsilon. \tag{B.51}$$

Der Zusatz „$\forall\epsilon$“ hier besagt, daß zur Ermittlung von $\Delta\beta_{s,m}^{\min}$ alle möglichen ϵ zuzulassen sind – die $\beta_j(\lambda_k)$–Relation (B.46) hängt ja auch von ϵ ab, die genaue Positionierung von ϵ innerhalb des in der Praxis zunächst natürlich immer unbekannten $\{\lambda_k\}$-Spektrums erfolgt aber gewissermaßen „blind“ und der ungünstigste Fall ist in Rechnung zu stellen. Anders ausgedrückt, für $\Delta\beta_{s,m}^{\min}$ werden alle vollständig in $[0, V_0]$ gelegenen Sätze $\{\lambda_k, \ldots, \lambda_{k+p+1}\}$ miteinander verglichen und der am schlechtesten konvergierende ausgewählt, wobei ϵ den für die Konvergenz jeweils ungüngstigsten Wert ϵ^* annehmen soll.

Diesen ungüngstigsten Wert ϵ^* für einen herausgegriffenen Satz erhält man wie folgt: Sei

$$S_p = \{\lambda_{k+1}, \ldots, \lambda_{k+p}\}, \quad p > 1, \quad 0 < \lambda_1 \leq \lambda_2 \leq \ldots,$$

jener p-fache Satz, der im $\{\beta_j\}$-Spektrum die dominanten Eigenwerte $\{\beta_1, \ldots, \beta_p\}$ hervorbringt. Der *erste nichtdominante* Eigenwert $\beta_{p+1} \Leftrightarrow \lambda_{\beta(p+1)}$ wird entweder links oder rechts von S_p liegen, also λ_k oder λ_{k+p+1} sein. Wir ermitteln ϵ^* zunächst für den Fall, daß der links gelegene λ_k dieser Eigenwert ist, $\lambda_{\beta(p+1)} = \lambda_k$. Der hierfür zulässige Wertebereich für ϵ ist offenbar

$$(\lambda_k + \lambda_{k+p})/2 \;\leq\; \epsilon \;\leq\; (\lambda_k + \lambda_{k+p+1})/2\,. \tag{B.52}$$

Bei Verletzung der ersten Ungleichung gerät λ_k selbst unter die dominanten Eigenwerte, bei Verletzung der zweiten wird dagegen λ_{k+p+1} zum ersten nichtdominanten. Die Iterationslücke

$$\Delta\beta_{s,m} = (\lambda_{\beta(s)} - \epsilon)^2 - (\lambda_k - \epsilon)^2, \quad 0 < \lambda_k < \lambda_{\beta(s)} \tag{B.53}$$

wird hier nun um so kleiner, je kleiner ϵ wird, und der nach (B.52) kleinstmögliche Wert ist

$$\epsilon_L^* = (\lambda_k + \lambda_{k+p})/2\,. \tag{B.54}$$

Werden diese Überlegungen sinngemäß mit λ_{k+p+1} als erstem nichtdominanten Eigenwert für S_p wiederholt, findet man

$$\epsilon_R^* = (\lambda_{k+p+1} + \lambda_{k+1})/2\,. \tag{B.55}$$

Die resultierenden $\Delta\beta_{s,m}$ beider Varianten sind schließlich zu vergleichen, um das tatsächliche ϵ^* für S_p zu erhalten.

B.7.3 Quantitative Auswertung an einem lösbaren Modellsystem

Um nun zu einigen verläßlichen konkreten Zahlenwerten zu gelangen, nutzen wir das separable und dem rechteckigen Kasten der Höhe V_0 sehr ähnliche zweidimensionale Potentialgebilde (5.3). Separabilität geht bekanntlich mit einer Possion-Verteilung (1.20) der Nächste-Nachbar-Abstände einher (kleine Abstände werden bevorzugt, das Spektrum neigt zur Klumpung), so daß im Hinblick auf die Konvergenz hier ein eher besonders ungünstigstes Beispiel vermutet werden darf.

Für ein Rechteck der Abmessung 80×50 Å^2 mit 675 Eigenwerten in $[0, V_0] = [0, 8]$ eV ist das Ergebnis einer solchen Statistik, wie sie im obigen Abschnitt vorbereitet wurde, in Abb. B.2 gezeigt. Dargestellt ist das Anwachsen der minimalen Iterationslücken $\Delta\beta_{s,m}^{\min}$ mit der Zahl der gesuchten Iterierten (s). Kurvenparameter ist die Zahl der zusätzlich Iterierten (m). Offensichtlich wird dieses Anwachsen durch eine merkliche statistische Streuung moduliert, die inbesondere für $m = 2, 4$ mit ganz erheblichen relativen Schwankungen für $\Delta\beta_{s,m}^{\min}$ verbunden ist. Dies mag zunächst verwundern, werden die zu vergleichenden Vektorsätze

doch für alle s aus ein und demselben Intervall $[0, V_0]$ entnommen, d.h. das Intervall ist abgeschlossen, es kommen keine Spektrumsbereiche hinzu, wenn s wächst und irgendein einzelner, minimaler Eigenwertabstand, der nur von m abhängt (z.B. einer der Abstände, die unten „Randpuffer“ genannt werden), steht jedem Vektorsatz zur Verfügung. Man könnte daher annehmen, daß die große Anzahl konkurierender Vektorsätze für einen hinreichend glatten Verlauf des Minimums (B.51) sorgen sollte. Dem ist offensichtlich nicht so.

Folgendes ist hier zu bedenken: Durch die Quadrierung (B.36) wird der Gesamtpuffer aus allen m zusätzlich Iterierten aufgesplittet auf die beiden Ränder des Satzes S_p. Diese zwei Randbereiche, im folgenden Randpuffer genannt, werden für das Maß $\Delta\beta_{s,m}^{\min}$ nun dergestalt miteinander korreliert, daß zwischen den beiden Randpuffern zum einen die feste Anzahl von s Eigenwerten ($s = 1, \ldots, 40$) liegen muß, und die Minimumsbedingung (B.51) dann jene Sätze aussondiert, bei denen beide Randpuffer *simultan* am kleinsten werden (simultan am stärksten „klumpen“). Und dies ist, sagen wir, für $m = 4$ und zwei Übernächste-Nachbar-Abstände als Randpuffer in einer Distanz $s = 20$ offenbar stärker der Fall als in einer Distanz $s = 19$.

Daß andererseits $\Delta\beta_{s,m}^{\min}$ für $m = 2$ relativ gleichmäßig mit s wächst, ist wiederum insofern plausibel, als hier die beiden Randpuffer im wesentlichen durch Nächste-Nachbar-Abstände gestellt werden, die bei einem separablen Problem bekanntlich Possion-verteilt und unkorreliert sind.

Wir haben es hier offenbar mit typischen Spektrumseigenschaften bezüglich der Korrelation zweier Kurzabstände, charaktersiert durch m, in fixierter größerer Distanz, charakterisiert durch s, zu tun. Solche Eigenschaften wurden nach Wissen des Autors bisher nicht untersucht (es gäbe jenseits unserer speziellen Fragestellung dafür wohl auch keinen Grund), berühren die Thematik dieser Arbeit allerdings nur mittelbar und wurden nicht weiterverfolgt.

Zurück zu Abb. B.2: Legt man bei der Berechnung der R-Werte (B.44) aus den $\Delta\beta_{s,m}^{\min}$ die optimalen Verschiebungen ρ_{opt} nach Gl. (B.47) zugrunde, so kombiniert man offensichtlich die ungünstigste Möglichkeit bei der Wahl des Vektorsatzes und von ϵ mit der günstigesten bei der Wahl von ρ. Die daraus resultierenden R_{opt} sind dann wie folgt zu interpretieren: $q = 1 - R_{\text{opt}}$ ist der beste zu erwartende Konvergenzquotient unter der Maßgabe, einen p-fachen Vektorsatz ($p = s + m$) beliebig aus dem Eigenwertintervall $[0, V_0]$ herausgreifen zu wollen und außerdem noch ϵ beliebig ungünstig wählen zu dürfen.

Tatsächlich unterscheiden sich die ρ_{opt} für verschiedene $\Delta\beta_{s,m}^{\min}$ dann nur ganz minimal und die Kurven “R_{opt} über s“ würden optisch gänzlich parallel zu denen von $\Delta\beta_{s,m}^{\min}$ in Abb. B.2 verlaufen. Daher wurde auf eine gesonderte Darstellung verzichtet und in Abb. B.2 lediglich auf der rechten Seite der andere Ordinatenmaßstab für R_{opt} eingetragen.

Anhand der Abb. B.2 entnommenen R_{opt}-Werte können aus Abb. B.3 nun die totalen Konvergenzraten eines k-schrittigen Tschebyscheff-Blockes innerhalb der Prozedur RITZIT abgelesen werden, woraus letztlich die erforderlichen Iterationszyklen folgen.

Für $R = 5 \times 10^{-7}$ wäre laut Abb. B.3 bspw. nach einem 10^4-schrittigen Tschebyscheff-Block der Startfehler auf den 10^{-4}ten Teil gefallen und entsprechend 4 Dezimalstellen als gültig anzusehen (8 bei den Eigenwerten). Allerdings beginnt RITZIT nicht mit einem solchen Block, sondern mit einem von 1. Ordnung und erhöht deren Ordnung nur in etwa durch Verdopplung. Für einen abschließenden T_k-Block mit $k = 2^{13} = 8192$ wären also ca.

$$1 + 2 + 2^2 + 2^3 + \ldots + 2^{13} = 16383$$

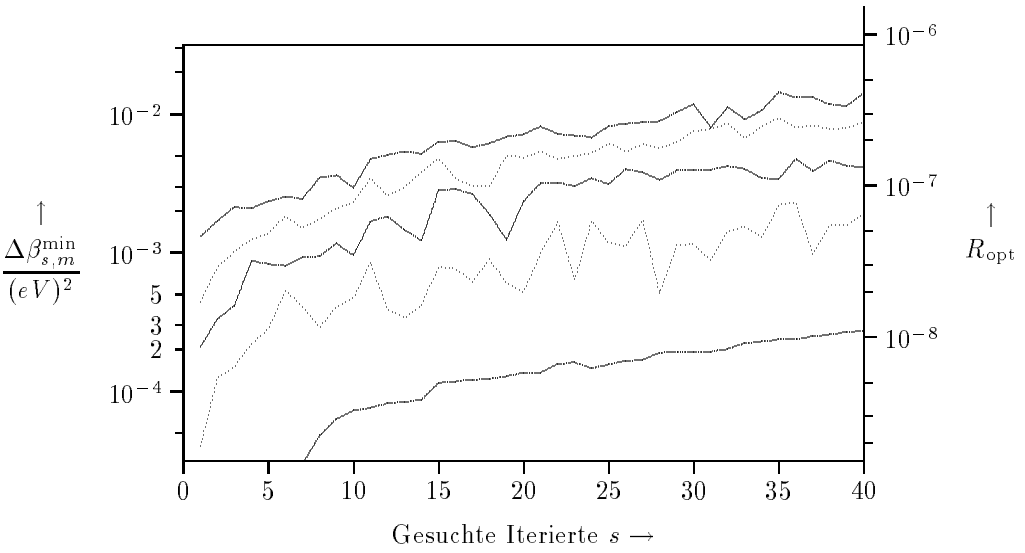


Abbildung B.2: *Statistik der minimalen Iterationslücken $\Delta\beta^{\min}_{s,m}$ (linke Ordinate) und der zugehörigen $R_{\rm opt}$ (rechte Ordinate) im Energieintervall $[0, V_0] = [0, 8]$ eV für das Modellsystem (5.3), speziell in der Größe 80×50 Å^2. Die Kurven gehören von unten nach oben zu $m = 2, 4, 6, 8, 10$. Siehe Text.*

Iteration vonnöten, deren Einzelkonvergenzraten T_k^{-1} sich dann multiplizieren.

Für die von uns zu Beginn praktizierte Kombination $s = 20$ und $m = 4$ (später $m = 6$) liefert Abb. B.2 allerdings $R \approx 5 \times 10^{-8}$ und aus Abb. B.3 würde $T_k^{-1} \approx 0.1$ für $k = 10^4$ folgen, was praktisch Nichtkonvergenz bedeuten würde.

Tatsächlich zeigten entsprechende Testrechnungen an ebendiesem System (und auch an anderen) jedoch deutlich bessere Konvergenzraten, vergleichbar etwa mit derjenigen für $R \approx 5 \times 10^{-7}$ und ≈ 20.000 Iterationen. Diese Diskrepanz ist nur zu einem geringen Teil auf das großzügig gewählte Intervall $[0, V_0] = [0, 8]$ eV zurückzuführen – für $[0, 5]$ eV steigen die entsprechenden Werte für $\Delta\beta^{\min}_{s,m}$ und $R_{\rm opt}$ nur um ca. 10%.

Der Hauptgrund dürfte sein, daß die verfügbaren theoretischen Aussagen zur Konvergenzgeschwindigkeit wie (B.9), (B.30) generell nur untere Schranken verkörpern, die darüberhinaus auch nur asymptotisch (d. h. für $\underline{\mathbf{v}}_i^\ell$ nahe $\underline{\mathbf{u}}_i$ bzw. $\ell \rightarrow \infty$ mit ℓ als Iterationsindex) gelten [42, Kap. 3][70]. Es finden sich in der Literatur immer wieder Hinweise, z. B. [42, Kap. 13][63], daß im Vergleich mit praktischen Rechnungen derartige Abschätzungen zu pessimistisch ausfallen. Nichtsdestotrotz ist die Tendenz dieser Aussagen natürlich gültig.

Nimmt man an, daß die speziell für das untersuchte separable Modellsystem von 80×50 Å^2 gefundenen Ergebnisse näherungsweise auch für andere zweidimensionale Systeme vergleichbarer Größe representativ sein sollten, wobei nichtseparable System mit steiferen Spektren etwas günstigere Konvergenzeigenschaften aufweisen dürften, kann folgendes Resümee gezogen werden:

- Die Konvergenz ist extrem schwach, eine Simultane Iteration ohne Tschebyscheff-Beschleunigung z. B. würde vollkommen versagen (vgl. die gepunktete Linie in Abb. B.3).

- Zur Zahl m der zusätzlich Iterierten:
  Nach Abb. B.2 stellt $m = 4$ das absolute Minimum dar, besser ist $m = 6$ oder $m = 8$.
  Für $m \leq 8$ wird die relative statistische Streuung der Konvergenzrate bzgl. s deutlich

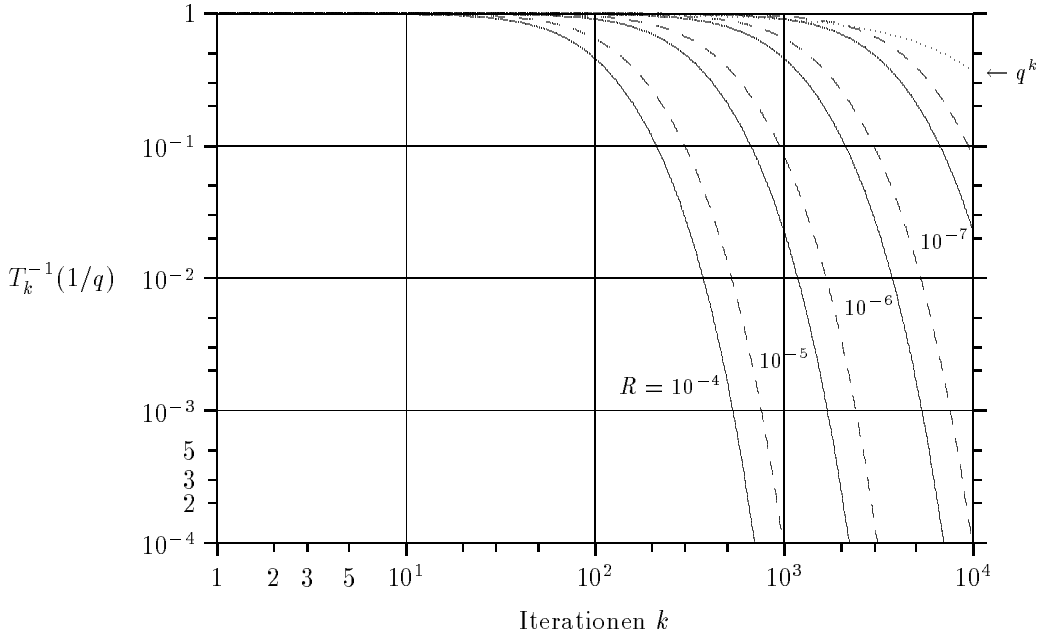


Abbildung B.3: *Die mit der Tschebyscheff-Beschleunigung erreichbaren Konvergenzraten* $T_k^{-1}(1/q)$ *gemäß (B.30) in Abhängigkeit von der Zahl der Iterationen* k *innerhalb eines Tschebyscheff-Blocks, also der Ordnung der Tschebyscheff-Polynome, wobei* $q = 1 - R = |\beta_{p+1}/\beta_j|$, $(j = 1, \ldots, p)$ *der Konvergenzquotient des gesuchten Eigenwertes ist. Durchgezogene Linien stehen für* $R = 10^{-4} \ldots 10^{-7}$ *und gestrichelte für die Zwischenwerte,* 5×10^{-5} *etc. Punktiert oben rechts zum Vergleich für* $R = 10^{-4}$ *die „naive" Konvergenzrate* q^k *einer Simultanen Iteration ohne Tschebyscheff-Beschleunigung.*

geringer.

- Zur Zahl s der gesuchten Iterierten:
  Für $m \leq 6$ wird die Zunahme der Konvergenzrate mit s weitgehend von statistischen Schwankungen überdeckt, ab $m = 8$ sind diese Schwankungen deutlich kleiner. Insgesamt ist die Konvergenzverbesserung aber viel geringer als bei einer Erhöhung von m um den gleichen Wert.

Anhang C

Linienmatrizen

C.1 Definitionen

Die Diskretisierung der zeitfreien SGL führte in Abschnitt 4.2 zu Eigenwertmatrizen, die Nichtnullelemente nur entlang einiger diagonaler Linien besaßen (Abb. 4.2 und 4.3). Matrizen mit dieser Eigenschaft bezeichnen wir hier als *Linienmatrizen*, wobei wir uns – Eigenwertaufgaben angemessen – nachfolgend auf deren quadratische Formen beschränken. Hervorgegangen ist dieses Kapitel aus der Suche nach einer schnellen und speicherplatzsparenden Matrix-mal-Vektor-Regel für quadrierte bzw. überhaupt für potenzierte Diskretisierungsmatrizen (vgl. 4.4.2). Dementsprechend wird auf eine für numerische Belange günstige Notation geachtet und insbesondere die Multiplikation zweier Linienmatrizen behandelt, mit deren Hilfe die Quadrate der Diskretisierungsmatrizen dann leicht gefunden werden können.

Um die Darstellungen zu vereinfachen, werden in diesem Kapitel Matrixelemente und Vektoren mit Null beginnend indiziert, d. h. für die Elemente a_{ij} einer (n,n)–Matrix gilt: $i,j = 0, \ldots, n-1$. Der Grund ist, daß Subdiagonalen sich auf natürliche Weise durch ihren Abstand von der Hauptdiagonalen charakterisieren lassen und dieser Abstand bei obiger Konvention geradewegs zusammenfällt mit dem Index in der nullten Zeile bzw. Spalte, was die Gleichungen nicht unmerklich entlastet.

In einer Matrix nennen wir einzelne parallel zur Hauptdiagonalen verlaufende Linien schlechthin 'Linien' und verstehen sodann unter einer (quadratischen) *Linienmatrix* $\mathbf{A} \in \mathbf{R}^{n,n}$ eine Matrix, deren Nichnullelemnte sich vollständig auf einigen wenigen Linien anordnen lassen, die bildlich also nur von wenigen Diagonalen durchzogen ist. Für die computerinterne Darstellung einer Linienmatrix bietet sich dann an, ist bei unseren großen Dimensionen sogar unumgänglich, nur die einzelnen Linien (als eindimensionale Vektoren) zu speichern, und es ist das Ziel der folgenden Ausführungen, benötigte Matrixoperationen auf Operationen zwischen Linien zurückzuführen.

Der Begriff 'Linienmatrix' soll also auf eine hier naheliegende und anzustrebende linienorientierte Betrachtungsweise – von den Diagonalen her – im Gegensatz zur sonstigen spalten- und zeilenorientierten hinweisen. Er ist per Definition, ähnlich dem der 'Bandmatrix', weder scharf noch exklusiv und folglich auf beliebige Matrizen anwendbar, denen vollbesetzt dann z. B. $2n-1$ Linien zuzuordnen wären. Notwendigerweise sind die folgenden Ausführungen auch ganz unabhängig von der Linienzahl, gleichwohl ein linienorientiertes Vorgehen nur bei schwach besetzten Matrizen sinnvoll sein wird.

Linien, die von Matrixrand zu Matrixrand durchgezogen erscheinen (jedes Element kann ein Nichtnullelemente sein), werden 'gefüllt' genannt und 'unvollständig gefüllt' solche, bei denen ein Teil der Elemente aus *topologischen* Gründen identisch verschwindet. Relevanz erlangt diese Klassifizierung erst bei Operationen mit Lininematrizen, wohingegen in einer separat betrachteten Matrix jede unvollständig gefüllte Linie ebensogut auch gefüllt (mit entsprechend vielen Nullen) genannt werden kann. Für die Ausgangsmatrizen werden vereinfachend alle Linien als gefüllt angenommen.

Zur Parametrisierung der Linien führen wir einen Positionsparameter p ein, der proportional zum Abstand einer Linie zur Hauptdiagonalen ist und der positiv oder negativ zählt, je nachdem, ob sich eine Linie ober- oder unterhalb der Hauptdiagonalen befindet (Abb. C.1). Offensichtlich gilt

$$p \;=\; \text{Spalte} - \text{Zeile}. \tag{C.1}$$

Die Gesamtheit der Linienpositionen in $\mathbf{A}$ enthalte der Satz S_A,

$$S_A = \{p_k\} = \{p_1, p_2 \ldots, p_m\}\,; \qquad m \le 2n-1, \tag{C.2}$$

so daß die Linieneigenschaft von $\mathbf{A}$ wie folgt beschrieben werden kann:

$$a_{ij} = \sum_{p \in S_A} a_{ij}\,\delta_{j,i+p}\;, \qquad (-n < p < n). \tag{C.3}$$

(C.3) ist quasi eine Abbildung der a_{ij} auf sich selbst, um dergestalt die Nichtnullelemente von $\mathbf{A}$ auszusondern: Überstreicht p an jeder i-j-Indexkombination alle je vorhandenen Linienpositionen, verbleiben vermöge des Kronecker-Symbols nur die Linienelemente, während Nichtlinienelemente ausgeblendet werden. Man beachte, daß hier trotz der Summe nach wie vor nur das jeweilige Element a_{ij} reproduziert oder eine Null erzeugt wird.

Weiter wird eine interne Indizierung der Linien benötigt, um diese als eindimensionale Vektoren $\underline{\boldsymbol{\ell}}_p^A \in \mathbf{R}^{n-|p|}$ auffassen zu können. Abb. C.1 zeigt den Bezug der $\underline{\boldsymbol{\ell}}_p^A$ zu den Matrixelementen: der innere Linienindex läuft im oberen Dreieck synchron mit der Matrixzeile, im unteren synchron mit der Matrixspalte, und es gilt

$$\ell_p^A(i) \;=\; \begin{cases} a_{i,i+p}\,; & p \ge 0;\;\; i = 0, \ldots, n-p-1; \\ a_{i-p,i}\,; & p \le 0;\;\; i = 0, \ldots, n+p-1. \end{cases} \tag{C.4}$$

Bei der umgekehrten Zuordnung der $\underline{\boldsymbol{\ell}}_p^A$ zu festen, herausgegriffenen a_{ij} treten dadurch in oberen und unteren Linien unterschiedliche innere Indizes zu Tage,

$$a_{ij} \;=\; \ell_p^A(i)\,\delta_{j,i+p} \qquad \text{für } i \le j \;\Leftrightarrow\; p \in [0,n), \tag{C.5}$$

$$a_{ij} \;=\; \ell_p^A(j)\,\delta_{j,i+p} \qquad \text{für } i \ge j \;\Leftrightarrow\; p \in (-n,0], \tag{C.6}$$

weshalb obere und untere Linien im weiteren getrennt geführt werden. Für die Hauptdiagonale $\underline{\boldsymbol{\ell}}_0^A$, diesbezüglich indifferent, wird vereinbart, sie dem oberen Dreieck, sprich dem Bereich positiver p, zuzuordnen, um für diese kein zusätzliches Szenario zu begründen.

Nach dem Einbinden der eindimensionalen Formen (C.5) und (C.6) in Gleichung (C.3) entsteht dann die hier verwendete linienorientierte Formulierung einer Matrix:

$$a_{ij} \;=\; \begin{cases} \displaystyle\sum_{p \in [0,n)} \ell_p^A(i)\,\delta_{j,i+p} & \text{für } i \le j; \\ \displaystyle\sum_{p \in (-n,-1]} \ell_p^A(j)\,\delta_{j,i+p} & \text{für } i > j. \end{cases} \tag{C.7}$$

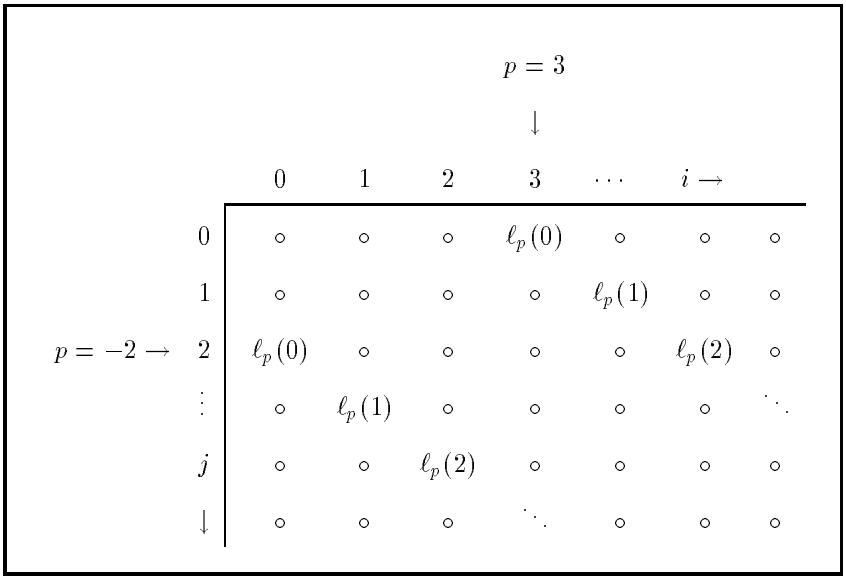


Abbildung C.1: *Beispiel für die Indizierung einer Linienmatrix. Positive p bezeichnen Linienpositionen oberhalb der Hauptdiagonalen, negative unterhalb. Die Linien selbst werden durch eindimensionale Vektoren $\underline{\boldsymbol{\ell}}_p$ mit eigener interner Indizierung beschrieben. Sämtliche Indizes beginnen mit Null.*

C.2 Multiplikation zweier Linienmatrizen

Gegeben seien zwei (n,n)-Linienmatrizen $\mathbf{A}$ und $\mathbf{B}$ und gesucht ist ihr Produkt $\mathbf{C} = \mathbf{AB}$ unter der Maßgabe, die Elemente $c_{ik} = \sum_j a_{ij} b_{jk}$ auf die Linien der Ausgangsmatrizen zurückzuführen. Es wird keine Kommutativität vorausgesetzt und $\mathbf{B}$ meint stets die rechte Matrix. Die Linienpositionen von $\mathbf{A}$ beschreibe $p \in S_A$ und entsprechendes leiste $q \in S_B$ für $\mathbf{B}$.

In der Lesart (C.3) können wir das Resultat sofort angeben:

$$\begin{aligned} c_{ik} &= \sum_j \sum_{p \in S_A} a_{ij}\, \delta_{j,i+p} \sum_{q \in S_B} b_{jk}\, \delta_{k,j+q} \\ &= \sum_{p \in S_A} \sum_{q \in S_B} a_{i,i+p}\, b_{k-q,k}\, \delta_{i+p,k-q}. \end{aligned} \tag{C.8}$$

Nachdem im zweiten Schritt die Summe über j abgegolten wurde, wird nur noch über von Linien besetzte Matrixelemente summiert und das verbleibende Kronecker-Symbol $\delta_{k,i+p+q}$ bringt gerade zur Geltung, daß Beiträge zu c_{ik} nur entstehen, wenn neben dem $\mathbf{A}$-Linienelement $a_{ij} = a_{i,i+p}$ auch sein Widerpart b_{jk} einer Linie in $\mathbf{B}$ angehört.

Etwas mehr Aufwand ist allerdings vonnöten, wenn sich im Ergebnis die eindimensionalen Vektoren $\underline{\boldsymbol{\ell}}^A_p$ und $\underline{\boldsymbol{\ell}}^B_q$ wiederfinden sollen, da dann in obere und untere Vektoren aufzuschlüsseln ist. Zunächst komplettieren wir den Ausgangspunkt durch die b_{jk} in der Notation (C.7)

$$b_{jk} = \begin{cases} \displaystyle\sum_{q \in [0,n)} \ell^B_q(j)\, \delta_{k,j+q} & \text{für } j \le k; \\ \displaystyle\sum_{q \in (-n,-1]} \ell^B_q(k)\, \delta_{k,j+q} & \text{für } j > k. \end{cases} \tag{C.9}$$

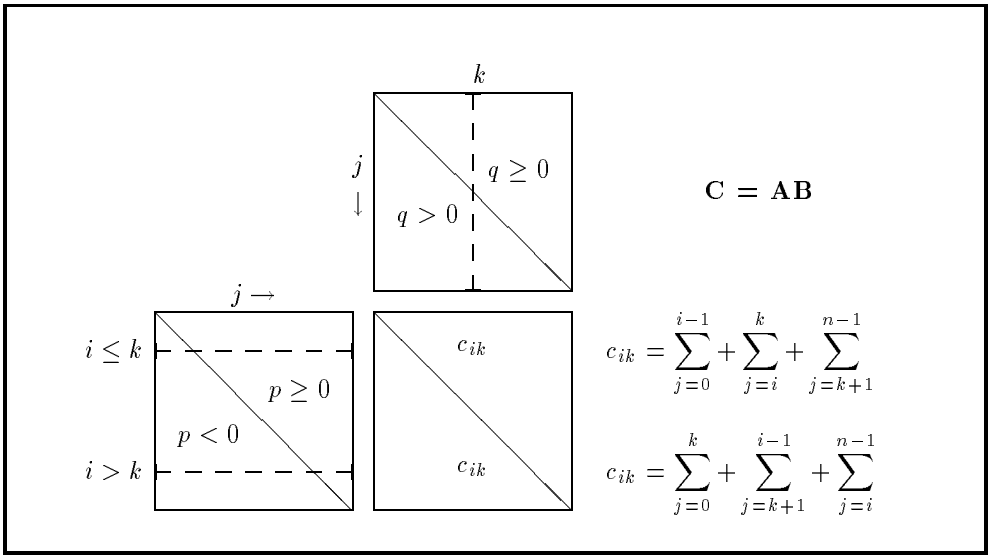


Abbildung C.2: *Die Summe* $c_{ik} = \sum_j a_{ij}\, b_{jk}$ *wird in drei Terme unterteilt, in denen die* a_{ij} *und* b_{jk} *entweder oberen oder unteren Linien* ($p, q \geq 0$ *oder* $p, q < 0$) *entstammen. Für* $i \leq k$ *wechseln zuerst die* a_{ij} *den Bereich, für* $i > k$ *dagegen die* b_{jk}*; vereinbarungsgemäß zählt die Hauptdiagonale selbst als obere Linie.*

Vorteilhafterweise werden bei der Summation $c_{ik} = \sum_j a_{ij} b_{jk}$ die Bereiche $i \leq k$ und $i > k$ getrennt behandelt, da der Summationsindex zum einen zuerst in **A**, zum anderen zuerst in **B** die Diagonale überschreitet, sprich, p und q in unterschiedlicher Reihenfolge das Vorzeichen wechseln (Abb. C.2). In beiden Fällen genügt dann eine Dreiteilung,

$$\boxed{i \leq k}: \quad c_{ik} = \underbrace{\sum_{j=0}^{i-1} a_{ij} b_{jk}}_{p<0,\, q\geq 0} + \underbrace{\sum_{j=i}^{k} a_{ij} b_{jk}}_{p\geq 0,\, q\geq 0} + \underbrace{\sum_{j=k+1}^{n-1} a_{ij} b_{jk}}_{p\geq 0,\, q<0} \tag{C.10}$$

$$\boxed{i > k}: \quad c_{ik} = \underbrace{\sum_{j=0}^{k} a_{ij} b_{jk}}_{p<0,\, q\geq 0} + \underbrace{\sum_{j=k+1}^{i-1} a_{ij} b_{jk}}_{p<0,\, q<0} + \underbrace{\sum_{j=i}^{n-1} a_{ij} b_{jk}}_{p\geq 0,\, q<0} \tag{C.11}$$

wobei mit den unteren Klammern angedeutet ist, aus welchen Linien die a_{ij} und b_{jk} entstammen.

Wie in (C.8) soll nun die Summe über j auf eine solche über wirklich vorhandene Linien in **A** reduziert werden, was mit der Frage einhergehen wird, ob auch das entsprechende Element b_{jk} einer **B**-Linie angehört. Um dieses Vorhaben in der Diktion der Gleichungen (C.7) und (C.9) zu bewerkstelligen, muß das dort involvierte Überstreichen der Positionsparameter auf paarweise sich ausschließende Intervalle beschränken werden, damit jede der möglichen p–q–Kombination jeweils nur in einem einzigen j–Bereich auftritt und nirgendwo mehrfach gezählt wird. Aus (C.1) mit

$$\begin{aligned} p &= j - i \\ q &= k - j \end{aligned}$$

folgen unmittelbar die jedem Summationsbereich zugehörigen p–q–Intervalle[1],

$\boxed{i \le k}$:

$$\begin{array}{lllll} j=0,\ldots,i-1; & \longrightarrow & p\in[-i,-1], & q\in(k-i,k] \\ j=i,\ldots,k; & \longrightarrow & p\in[0,k-i], & q\in[0,k-i] \\ j=k+1,\ldots,n-1; & \longrightarrow & p\in(k-i,n-i), & q\in(k-n,-1]; \end{array} \tag{C.12}$$

$\boxed{i > k}$:

$$\begin{array}{lllll} j=0,\ldots,k; & \longrightarrow & p\in[-i,k-i], & q\in[0,k] \\ j=k+1,\ldots,i-1; & \longrightarrow & p\in(k-i,-1], & q\in(k-i,-1] \\ j=i,\ldots,n-1; & \longrightarrow & p\in[0,n-i), & q\in(k-n,k-i]. \end{array} \tag{C.13}$$

die hier als geordnete Intervalle der Form [min, max] zu verstehen sind, d. h. die leer bleiben für min > max, wodurch sich das jedmalige Hinzufügen von $p<0$ oder $p\ge 0$ usw. erübrigt.

Nunmehr kann in allen Glieder von (C.10) und (C.11) wie in (C.8) geradlinig über j absummiert werden und wir erhalten zusammengefaßt

$$\boxed{i \le k}: \quad c_{ik} = \left\{ \begin{array}{l} \displaystyle\sum_{p\in[-i,-1]} \sum_{q\in(k-i,k]} \ell_p^A(i+p)\,\ell_q^B(i+p) \\ \quad + \displaystyle\sum_{p\in[0,k-i]} \sum_{q\in[0,k-i]} \ell_p^A(i)\,\ell_q^B(i+p) \\ \quad + \displaystyle\sum_{p\in(k-i,n-i)} \sum_{q\in(k-n,-1]} \ell_p^A(i)\,\ell_q^B(k) \end{array} \right\} \delta_{k,i+p+q}\ , \tag{C.14}$$

$$\boxed{i > k}: \quad c_{ik} = \left\{ \begin{array}{l} \displaystyle\sum_{p\in[-i,k-i]} \sum_{q\in[0,k]} \ell_p^A(i+p)\,\ell_q^B(i+p) \\ \quad + \displaystyle\sum_{p\in(k-i,-1]} \sum_{q\in(k-i,-1]} \ell_p^A(i+p)\,\ell_q^B(k) \\ \quad + \displaystyle\sum_{p\in[0,n-i)} \sum_{q\in(k-n,k-i]} \ell_p^A(i)\,\ell_q^B(k) \end{array} \right\} \delta_{k,i+p+q}\ . \tag{C.15}$$

Die Elemente der Matrix **C=AB** sind damit wie angestrebt auf die Linien von **A** und **B** zurückgeführt. Eine gewisse Unsymmetrie zwischen (C.14) und (C.15) bleibt der Tatsache geschuldet, daß die Hauptdiagonale dem oberen Dreieck zugeschlagen wurde.

Tatsächlich besitzt auch **C** wieder die Topologie einer Linienmatrix, wie Abb. C.3 illustriert und was näher ausgeführt wird im folgenden Satz:

Satz C.1 (Topologie von C) *Gegeben seien die beiden Linienmatrizen* $\mathbf{A},\mathbf{B}\in\mathbf{R}^{n,n}$ *mit den Positionssätzen* S_A, S_B, *den Positionsparametern* $p\in S_A$, $q\in S_B$ *und den Linien* $\underline{\boldsymbol{\ell}}_p^A\in\mathbf{R}^{n-|p|}$ *und* $\underline{\boldsymbol{\ell}}_q^B\in\mathbf{R}^{n-|q|}$. *Dann gilt:*

(i) Die Produktmatrix $\mathbf{C}=\mathbf{AB}$ *besitzt ebenfalls wieder die Linieneigenschaft, dergestalt, daß jede mögliche p–q–Kombination der Ausgangslinien mit* $p+q<n$ *eine* **C**–*Linie* $\underline{\boldsymbol{\ell}}_r^C$ *an der Position* $r=p+q$ *zur Folge hat und auch nur in diese hinein schreibt.*

[1] Es gibt (bei festem i und k) zu jedem j natürlich immer nur ein einziges p und q, formal gesehen aber würde z. B. eine Intervallschachtelung für p zusammen mit einer Aufteilung $q\ge 0$, $q<0$ genügen. Daß hier beide Parameter eingegrenzt werden, dient der Lesbarkeit der Gleichungen und betont die Gleichstellung von p und q.

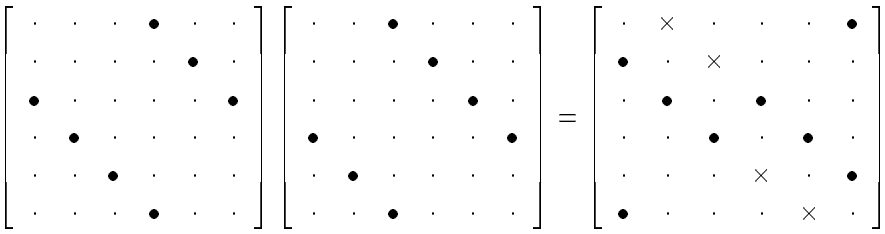


Abbildung C.3: *Beispiel einer Matrixmultiplikation* $\mathbf{C} = \mathbf{AB}$ *mit* $n = 6$ *und* $\mathcal{S}_A = \{3, -2\}$, $\mathcal{S}_B = \{2, -3\}$. *In* $\mathbf{C}$ *entstehen Linien an den Positionen* $r = 1, 5, -1, -5$, *wobei für* $r = 1$ $(p - q = 3 - 2)$ *und* $r = -1$ $(p - q = 2 - 3)$ *die mit "*$\times$*" gekennzeichneten Elemente nicht beschrieben werden (unvollständig gefüllte Linien entsprechend Satz C.1).*

(ii) Werden die Ausganglinien $\underline{\boldsymbol{\ell}}_p^A$ *und* $\underline{\boldsymbol{\ell}}_q^B$ *als gefüllt vorausgesetzt, dann sind auch die* $\mathbf{C}$*–Linien vollständig gefüllt, außer für* $p \cdot q < 0$. *In diesem Falle werden in der* $\mathbf{C}$*–Linie genau* $s \equiv \min(|p|, |q|)$ *Elemente nicht beschrieben, und zwar sind dies für* $p < 0$ *die ersten* s *Elemente,*

$$\ell_r^C(0) \ldots \ell_r^C(s-1),$$

und für $p > 0$ *die letzten* s *Elemente,*

$$\ell_r^C(n - |r| - s - 1) \ldots \ell_r^C(n - |r| - 1).$$

Beweis: Behauptung (i) über die Linieneigenschaft von $\mathbf{C}$ geht analytisch aus den Gleichungen (C.14) und (C.15) hervor, die beide die Form

$$c_{ik} = \left(\sum_{p,\,q} \ell_p^A \, \ell_q^B + \sum_{p,\,q} \ell_p^A \, \ell_q^B + \sum_{p,\,q} \ell_p^A \, \ell_q^B + \right) \delta_{k,\,i+p+q}$$

aufweisen, so daß sich infolge des Kronecker-Symbols Nichtnullelemente in $\mathbf{C}$ nur auf Parallelen zur Hauptdiagonalen im Abstand $p + q$ bilden können und außerdem jede $(p + q < n)$–Paarung genau zu einer $\mathbf{C}$–Linie an der Position $p + q$ kombiniert.

Aussage (ii) kann auf anschaulichem Wege fast unmittelbar aus Darstellungen wie Abb. C.3 entnommen werden, weshalb sich eine Begründung anhand von (C.14), (C.15) hier erübrigt. q. e. d.

Bemerkung: Die zweite Aussage des obigen Satzes betrachtet $\mathbf{C}$–Linien, wie sie einer einzigen p–q–Paarung entspringen würden; schreiben dagegen auch andere p–q–Kombinationen in diese Linie, überlagern sich die Beiträge entsprechend. □

Um schließlich aus den Matrixelementen c_{ik} noch die eindimensionalen Linien $\underline{\boldsymbol{\ell}}_r^C$ mit $r = p + q$ zu separieren, denken wir uns (C.14) und (C.15) im Sinne der Identität

$$\sum_{p,\,q} \delta_{k,\,i+p+q} \; = \sum_{r} \sum_{p,\,q} \delta_{r,\,p+q} \, \delta_{k,\,i+r}$$

expandiert[2] und vergleichen sodann mit der Entsprechung

$$c_{ik} \quad = \quad \begin{cases} \displaystyle\sum_{r \in [0,\,n)} \ell_r^C(i) \, \delta_{k,\,i+r} & \text{für } i \le k, \\ \displaystyle\sum_{r \in (-n,\,-1]} \ell_r^C(k) \, \delta_{k,\,i+r} & \text{für } i > k, \end{cases}$$

[2] $\delta_{k,\,i+p+q}$ erfüllt in (C.14), (C.15) zwei Funktionen: Herausprojizieren der neuen Linienpositionen $c_{i,\,i+r}$ und Klammerung der p–q–Kombinationen zu $(r = p + q)$–Paarungen. Beide Informationen werden in der Identität quasi aufgespalten.

wonach (C.14) sofort auf

$$\boxed{r \geq 0}: \quad \ell_r^C(i) = \left\{ \begin{array}{ll} \displaystyle\sum_{p\in[-i,-1]} \sum_{q\in(r,i+r]} & \ell_p^A(i+p)\,\ell_q^B(i+p) \\ + \displaystyle\sum_{p\in[0,r]} \sum_{q\in[0,r]} & \ell_p^A(i)\,\ell_q^B(i+p) \\ + \displaystyle\sum_{p\in(r,n-i)} \sum_{q\in(i+r-n,-1]} & \ell_p^A(i)\,\ell_q^B(i+r) \end{array} \right\} \delta_{r,p+q} \tag{C.16}$$

führt und (C.15) in analoger Weise zu einem ähnlichen Ausdruck $\ell_r^C(k)$ für $r < 0$, wobei in diesem wie angedeutet der innere Index formal zunächst k statt i lauten würde.

Wir geben hier noch einen weiteren Satz, der gewisse Bedeutung für Diskretisierungsmatrizen besitzt.

Satz C.2 (unvollständig gefüllte Linien in $\mathbf{A}^2$) *Ist $\mathbf{A} \in \mathbf{R}^{n,n}$ eine Linienmatrix mit dem Positionssatz S_A, deren Linien sämtlich als gefüllt vorausgesetzt werden, so gilt: In der Quadrierten $\mathbf{A}^2$ können nur dann unvollständig gefüllte Linien auftreten, wenn in $\mathbf{A}$ Kombinationen von Linienpositionen $p, q \in S_A$ existieren, die folgenden zwei Bedingungen genügen: (i) $p \cdot q < 0$ und (ii) $|p| + |q| > n$.*

Beweis: Bedingung (i) ist nach Satz C.1 notwendig für das Auftreten unvollständig gefüllter Linien im Produkt $\mathbf{C} = \mathbf{AA}$. Existiert nun in $\mathbf{A}$ eine (p, q)–Kombination mit $p \cdot q < 0$, so ist auch immer die gespiegelte Paarung ($p' = q$, $q' = p$) vorhanden, die beide in die $\mathbf{C}$–Linie $\underline{\ell}_r^C \in \mathbf{R}^{n-|r|}$ mit $r = p + q = p' + q'$ schreiben, wobei wir o.B.d.A. $p < 0$ setzen. Von den $n - |r|$ Elementen der Linie $\underline{\ell}_r^C$ läßt nach Satz C.1 die Kombination ($p < 0$, $q > 0$) die ersten s Elemente unberührt, und ($p' > 0$, $q' < 0$) die letzten s, wobei $s = \min(|p|, |q|) = \min(|p'|, |q'|)$. Die $\mathbf{C}$–Linie bleibt nur dann unvollständig, wenn sich die beiden nicht gefüllten Bereiche überlappen, also $2s > n - |r|$ gilt oder

$$2\min(|p|, |q|) > n - |p + q|.$$

Wegen $p \cdot q < 0$ können wir für $|p| < |q|$ schreiben $|p + q| = |q| - |p|$ und für $|p| > |q|$ entsprechend $|p + q| = |p| - |q|$. In beiden Fällen erhält man $|p| + |q| > n$. q. e. d.

C.3 Quadrate der Diskretisierungsmatrizen

Um nun vermöge (C.16) oder (C.14) zur konkreten Gestalt z. B. einer quadrierten Matrix $\mathbf{C} = \mathbf{A}^2$ oder eines Matrix-mal-Vektor-Algorithmus $\mathbf{A}^2\underline{\mathbf{v}}$ zu gelangen, hat man zunächst alle p-q-Kombinationen aufzuschreiben, die entstehenden $\mathbf{C}$-Linien zu registrieren und die Einzelbeiträge zu jeder $\mathbf{C}$-Linie schließlich aufzusummieren. Vereinfachenderweise existiert bei den Diskretisierungsmatrizen nach Abb. 4.2 und 4.3 zu jeder Nebendiagonallinie in $\mathbf{A}^2$ nur jeweils eine solche p-q-Kombination, und außerdem sind die $\mathbf{A}$ und damit auch alle Potenzen von $\mathbf{A}$ symmetrisch. Insgesamt erfordert der Schritt von Gl. (C.16) zum schließlichen Matrix-mal-Vektor-Algorithmus durchaus noch eine gehörige Portion Kleinarbeit, der Weg dahin ist aber geradlinig, weshalb wir auf das Listing eines konkreten derartigen Algorithmus an dieser Stelle verzichten wollen. Zumal, wie in 4.4.2 erwähnt, erst parallelisierte Varianten von eigentlichen Interesse wären, die man ohnehin neu zu schreiben hätte.

Erwähnt sei lediglich Folgendes: Die Positionssätze für die drei Raumdimensionen lauten für $n_x, n_y, n_z > 1$:

$$\text{1d:} \quad S_A = \{0, \pm 1\}; \qquad n = n_x \tag{C.17}$$

$$\text{2d:} \quad S_A = \{0, \pm 1, \pm n_x\}; \qquad n = n_x n_y \tag{C.18}$$

$$\text{3d:} \quad S_A = \{0, \pm 1, \pm n_x, \pm n_x n_y\}; \qquad n = n_x n_y n_z\,. \tag{C.19}$$

In keinem der Postionssätze existieren (p,q)-Paarungen, die simultan $pq < 0$ und $|p| + |q| > n$ genügen, um nach Satz C.2 eine unvollständig gefüllte Linie in der Quadrierten zu hinterlassen. D. h. $\mathbf{A}^2$ besteht ausschließlich aus vollen Linien. Ob und welche Linienbereiche dann in höheren Potenzen von $\mathbf{A}$ leer bleiben, für Matrix-mal-Vektor-Vorschriften durchaus von Belang, läßt sich ebenfalls mit Hilfe von Satz C.2 feststellen.

Den eigentlichen Matrix-mal-Vektor-Algorithmen $\underline{\mathbf{u}} = \mathbf{A}\underline{\mathbf{v}}$ mit beliebiger Matrix $\mathbf{A}$ liegt dann der Zusammenhang

$$u_i = \ell_p^A(i)\, v_{i+p} \quad , \quad (p \geq 0) \tag{C.20}$$

$$u_{i+|p|} = \ell_{|p|}^A(i)\, v_i \quad , \quad (p < 0) \tag{C.21}$$

für obere bzw. untere Linien zugrunde.

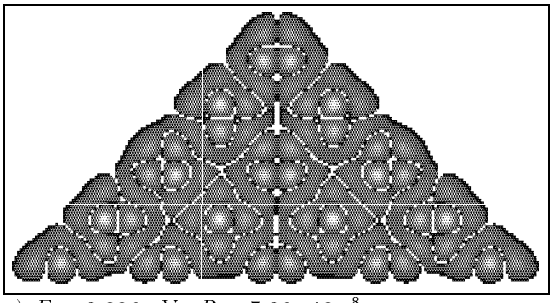
Anhang D

Die Bilder der Wellenfunktionen

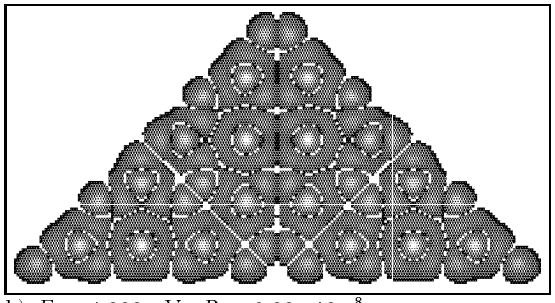
D.1 Spitze mit $R = 1$ Å und $\alpha = 90^o$

Die Betragsquadrate $|\psi(x,y)|^2$ von 20 Zuständen, die der Fermi-Energie $E_F = 4\,\mathrm{eV}$ nächstbenachbart liegen. Unter jedem Bild ist die Energie angegeben und das numerische Residuum des Eigenwertproblems $R = ||\mathbf{A}\underline{\mathbf{v}} - E\underline{\mathbf{v}}||$.

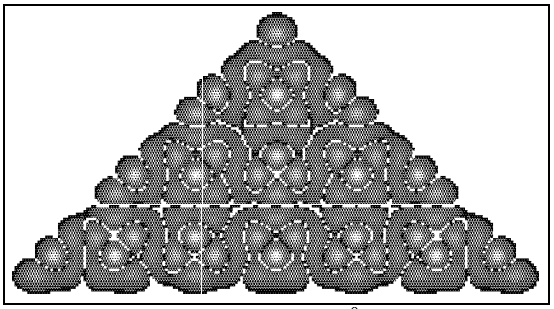
Numerische Daten:
$V_0 = 8$ eV.



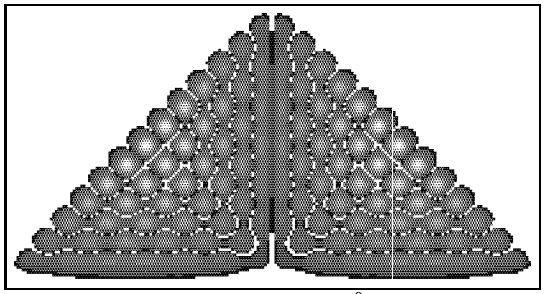
a) $E = 3.996$ eV, $R = 7.36 \cdot 10^{-8}$



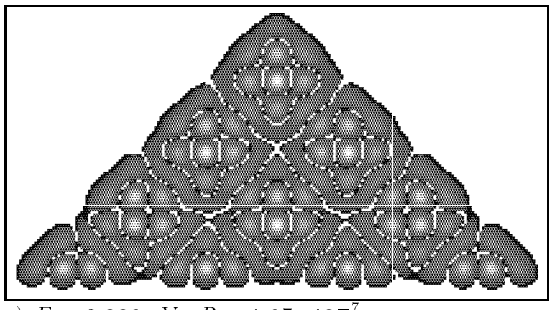
b) $E = 4.009$ eV, $R = 6.09 \cdot 10^{-8}$



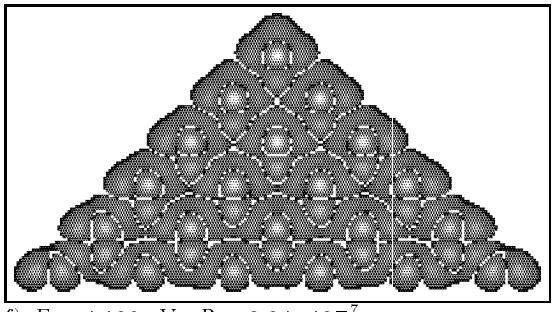
c) $E = 4.016$ eV, $R = 5.37 \cdot 10^{-8}$



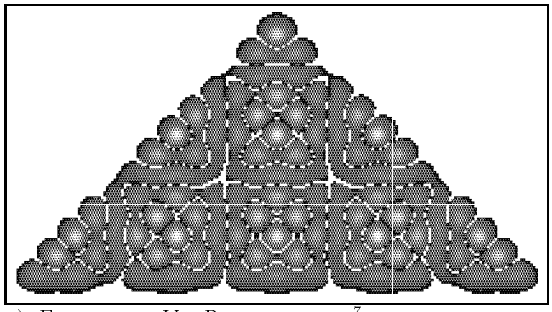
d) $E = 4.060$ eV, $R = 5.14 \cdot 10^{-8}$



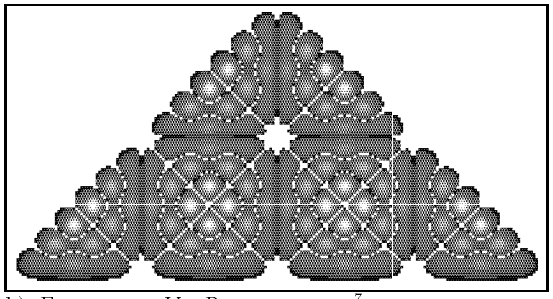
e) $E = 3.906$ eV, $R = 1.67 \cdot 10^{-7}$



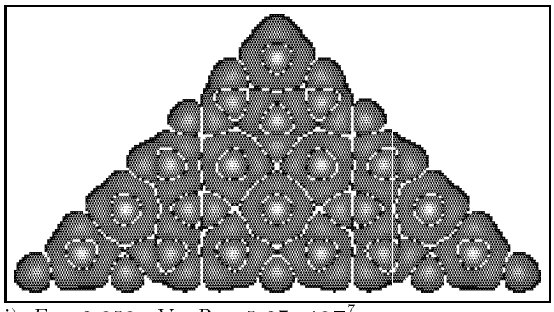
f) $E = 4.106$ eV, $R = 2.04 \cdot 10^{-7}$



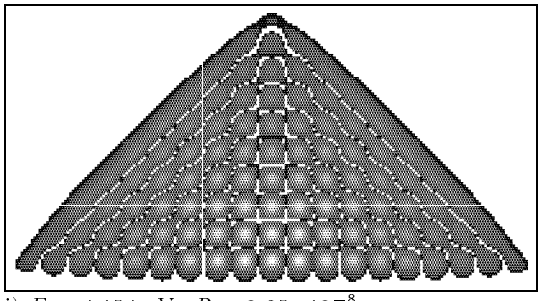
g) $E = 3.892$ eV, $R = 4.00 \cdot 10^{-7}$



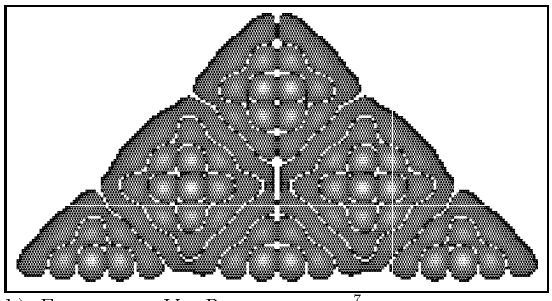
h) $E = 4.122$ eV, $R = 2.95 \cdot 10^{-7}$



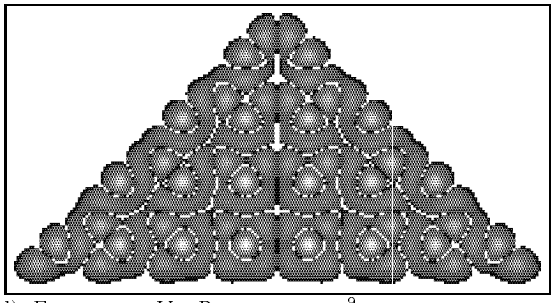
i) $E = 3.859$ eV, $R = 5.37 \cdot 10^{-7}$



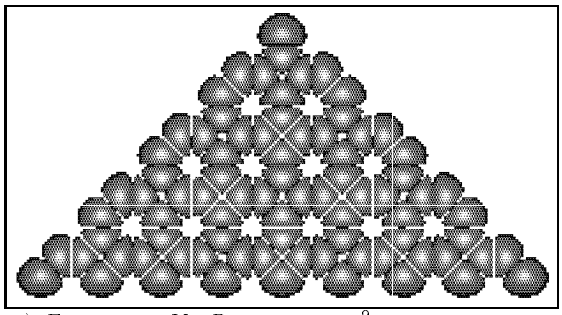
j) $E = 4.154$ eV, $R = 8.35 \cdot 10^{-8}$



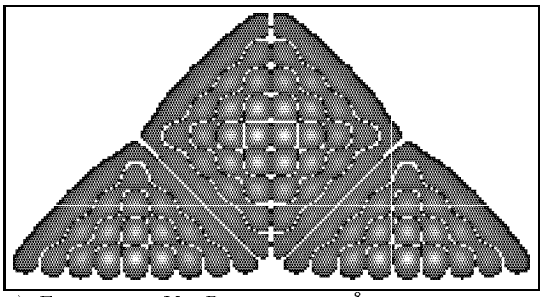
k) $E = 3.836$ eV, $R = 3.31 \cdot 10^{-7}$



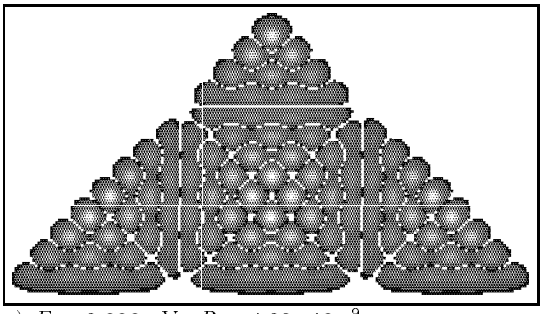
l) $E = 3.824$ eV, $R = 2.89 \cdot 10^{-9}$



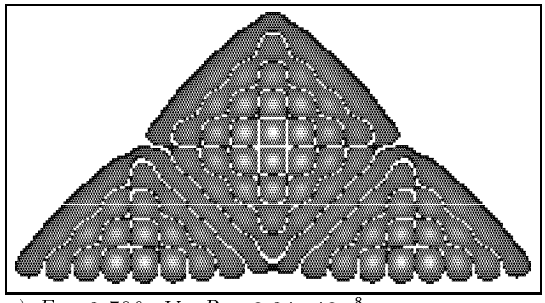
m) $E = 4.180$ eV, $R = 6.95 \cdot 10^{-9}$



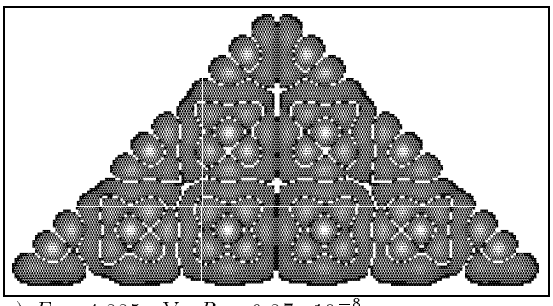
n) $E = 4.184$ eV, $R = 2.18 \cdot 10^{-8}$



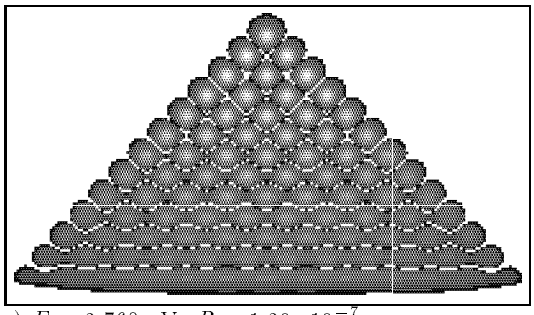
o) $E = 3.809$ eV, $R = 4.32 \cdot 10^{-9}$



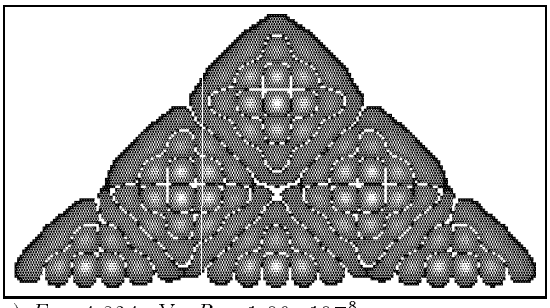
p) $E = 3.786$ eV, $R = 9.31 \cdot 10^{-8}$



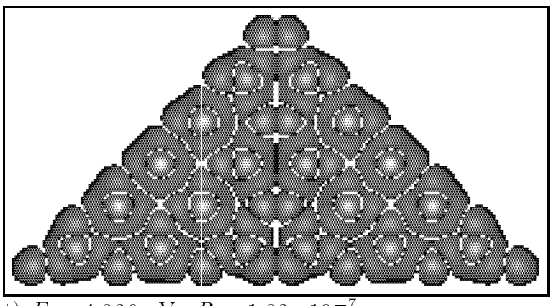
q) $E = 4.225$ eV, $R = 6.27 \cdot 10^{-8}$



r) $E = 3.768$ eV, $R = 1.30 \cdot 10^{-7}$

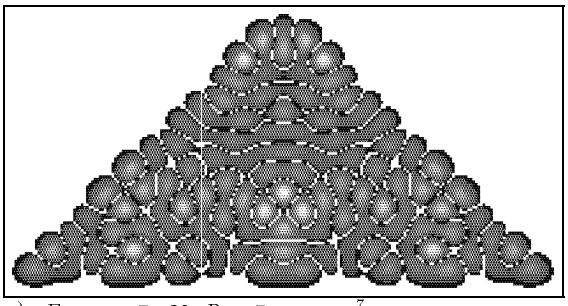


s) $E = 4.234$ eV, $R = 1.86 \cdot 10^{-8}$



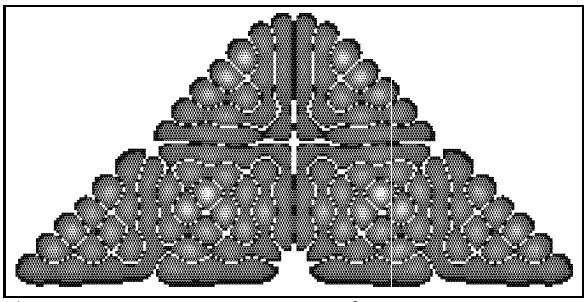
t) $E = 4.236$ eV, $R = 1.23 \cdot 10^{-7}$

D.2 Spitze mit $R = 4$ Å und $\alpha = 90^o$

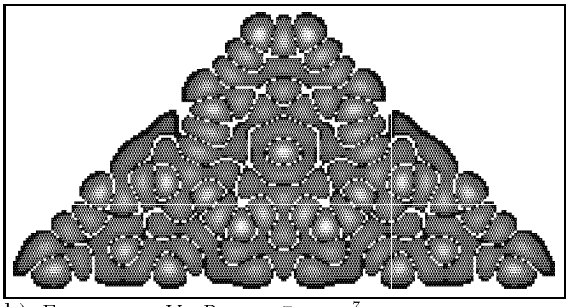


a): $E = 4.147$ eV, $R = 7.05 \cdot 10^{-7}$

D.3 Spitze mit $R = 8$ Å und $\alpha = 90^o$

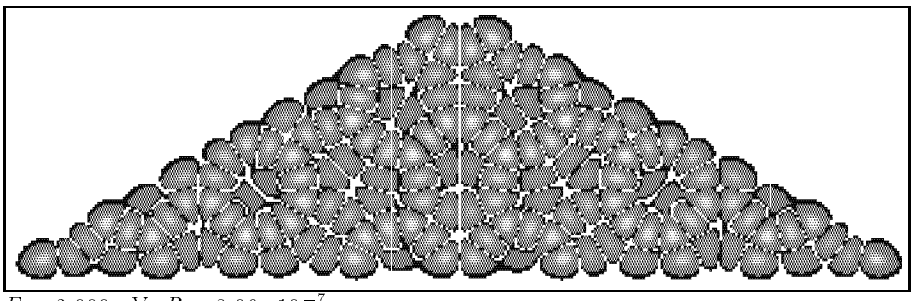


a) $E = 4.158$ eV, $R = 1.57 \cdot 10^{-6}$



b) $E = 3.884$ eV, $R = 2.37 \cdot 10^{-7}$

D.4 Spitze mit $R = 1$ Å und $\alpha = 120^o$



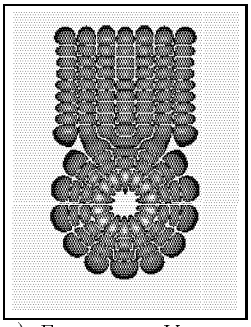
$E = 3.989$ eV, $R = 2.06 \cdot 10^{-7}$

D.5 Ein Zweikastensystem: Rechteck + Kreis

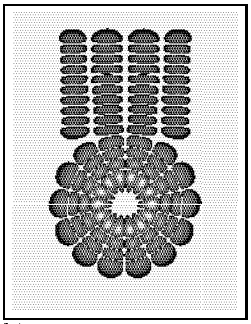
Die Betragsquadrate $|\psi(x,y)|^2$ von 20 Zuständen an der Fermi-Kante $E_F = 4\,\mathrm{eV}$ für ein kombiniertes System aus zwei Potentialkästen, die nur durch eine Tunnelbarriere voneinander getrennt sind: einem Kreis vom Radius 20 Å und einem Rechteck der Größe $40 \times 30\,\text{Å}^2$. Die Kastentiefe beträgt 5 eV, d. h. die Barrierenhöhe hier nur 1 eV. Betrachtete y-Abstände waren $d = 0, 1, 2, 5, 10\,\text{Å}$.

D.5.1 20 Zustände bei einem Abstand von $d = 2$ Å

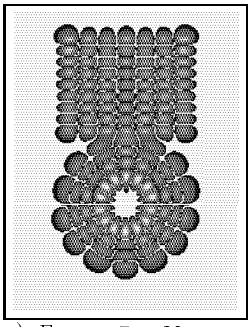
Numerische Daten:
$V_0 = 5.0$ eV. $\epsilon = 4.0$ eV. Simultan iteriert: $20 + 6$.
Grundgebiet: $70.0 \times 93.2\,\text{Å}^2$. $h = 0.4\,\text{Å} \quad \rightarrow \quad 174 \times 232 = 40368$ Punkte.
Iterationen: 11871. Rechenzeit: 6 h 30′ 59″. System: AIX R6000.



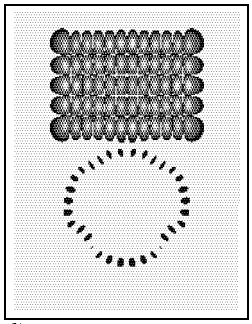
a) $E = 3.993$ eV, $R = 8.33 \cdot 10^{-7}$



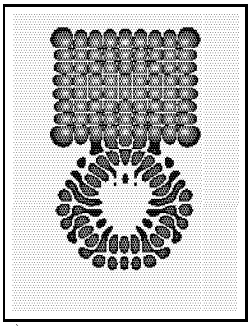
b) $E = 3.992$ eV, $R = 2.75 \cdot 10^{-6}$



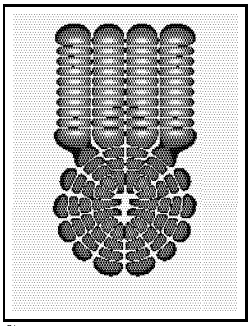
c) $E = 3.979$ eV, $R = 5.30 \cdot 10^{-7}$



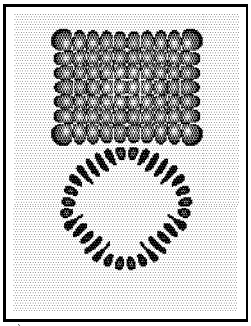
d) $E = 3.955$ eV, $R = 1.76 \cdot 10^{-6}$



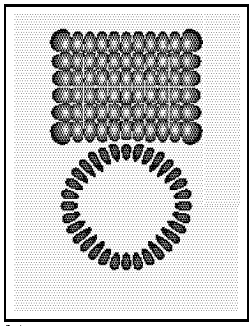
e) $E = 4.056$ eV, $R = 2.50 \cdot 10^{-7}$



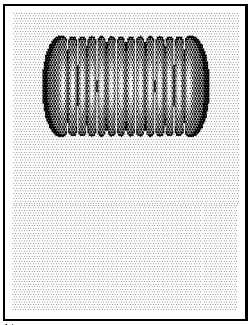
f) $E = 3.943$ eV, $R = 3.26 \cdot 10^{-6}$



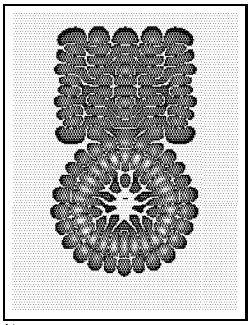
g) $E = 3.917$ eV, $R = 7.47 \cdot 10^{-6}$



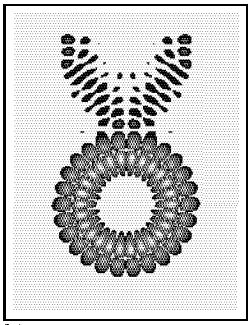
h) $E = 3.883$ eV, $R = 3.47 \cdot 10^{-8}$



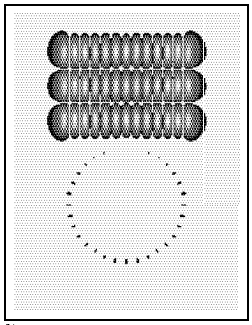
i) $E = 4.119$ eV, $R = 6.77 \cdot 10^{-9}$



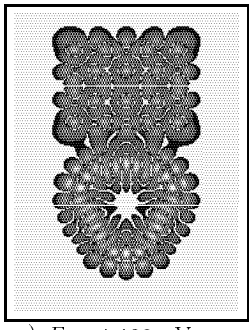
j) $E = 4.123$ eV, $R = 2.19 \cdot 10^{-8}$



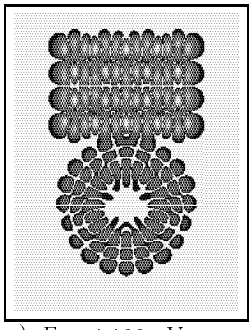
k) $E = 4.124$ eV, $R = 3.47 \cdot 10^{-8}$



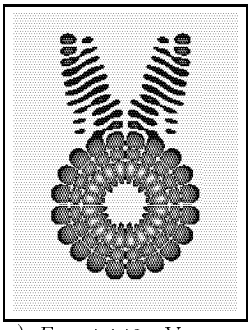
l) $E = 3.875$ eV, $R = 1.62 \cdot 10^{-7}$



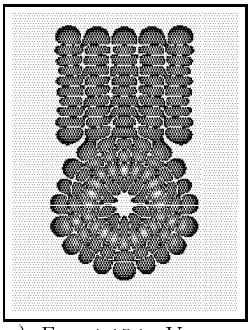
m) $E = 4.130$ eV,
$R = 9.79 \cdot 10^{-8}$



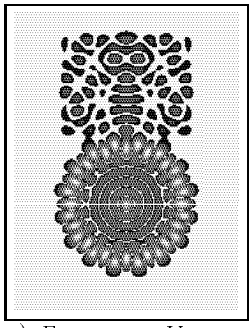
n) $E = 4.133$ eV,
$R = 7.45 \cdot 10^{-8}$



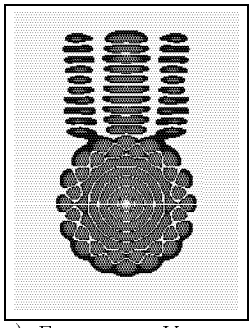
o) $E = 4.143$ eV,
$R = 5.50 \cdot 10^{-8}$



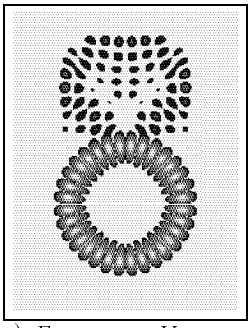
p) $E = 4.154$ eV,
$R = 3.37 \cdot 10^{-7}$



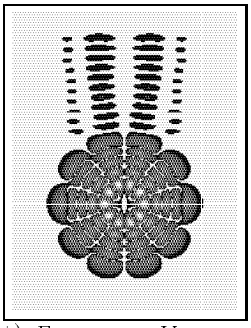
q) $E = 3.833$ eV,
$R = 1.45 \cdot 10^{-6}$



r) $E = 3.829$ eV,
$R = 7.62 \cdot 10^{-6}$

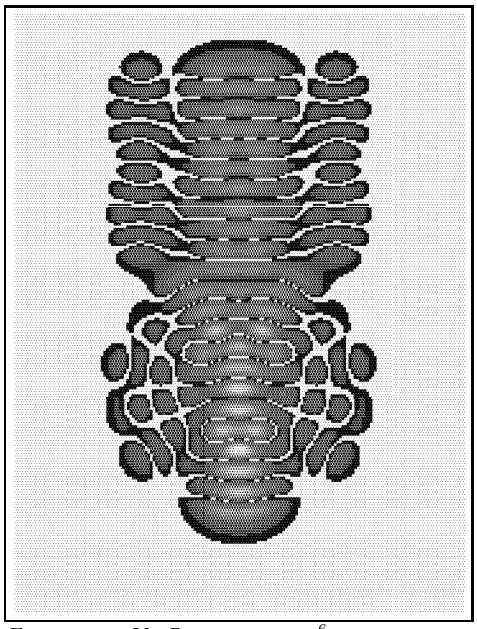


s) $E = 3.816$ eV,
$R = 1.68 \cdot 10^{-6}$



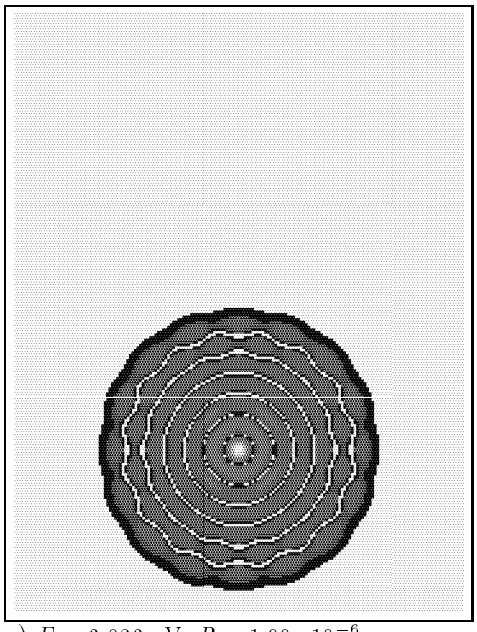
t) $E = 4.184$ eV,
$R = 1.70 \cdot 10^{-6}$

D.5.2 Ein Scar-Zustand bei einem Abstand von $d = 0$ Å

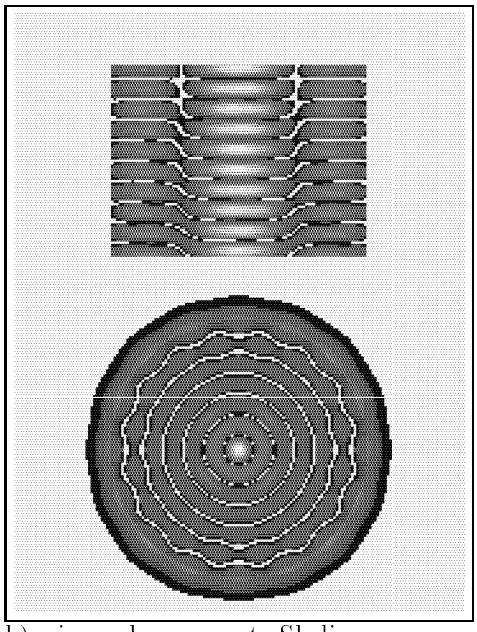


$E = 3.901$ eV, $R = 4.56 \cdot 10^{-6}$

D.5.3 Ein Zustand bei einem Abstand von $d = 10$ Å



a) $E = 3.823$ eV, $R = 1.90 \cdot 10^{-6}$



b) wie a, aber separate Skalierung.

Anhang E

Zur Software

An dieser Stelle soll ein kurzer Überblick über die entstandene Software gegeben werden, jedoch kein Programmlisting oder dergleichen.

Geschrieben wurden sämtliche Programme in C. Dienst-, Test- und Hilfsprogrammen eingeschlossen, entstand schätzungsweise Code im Umfang von 50.000 Zeilen. Die Berechung eines STM-Konstantstromprofils erfolgt dabei in zwei Schritten:

1. Berechnung der Wellenfunktionen von Probe und Spitze – Programm `WFCLC`
2. Berechnung des Tunnelstroms und der Konstantstromprofile – Programm `STM`

Im STM-Programm werden jeweils eine Spitzen- und eine Probenelektrode zusammengeführt, was sich schematisch so darstellt:

$$\boxed{\text{Spitze}} \Rightarrow \underbrace{\underbrace{\underbrace{\phi_1, \phi_2, \ldots, \phi_n}_{\text{Satz } \{\phi_i\}} \quad \underbrace{\psi_1, \psi_2, \ldots, \psi_m}_{\text{Satz } \{\psi_j\}}}_{\{\phi_i\} \leftrightarrow \{\psi_j\}}}_{\Uparrow \; \boxed{\text{STM}}} \Leftarrow \boxed{\text{Probe}} \tag{E.1}$$

Zur Stromberechnung und zur Spitze-Probe-Positionierung müssen stets Informationen über beide Elektroden zur Verfügung stehen (Zustände, Potential, Geometrie, Einbettungsgebiet). Für den Einzelfall genommen ist diese Aussage natürlich trivial, da bei der vorliegenden Problemstellung aber prinzipiell die Kombination jeder Elektrode mit jeder denkbar und als Möglichkeit auch offenzuhalten ist, sollte die Beschreibung der verschiedenen Elektroden in der gesamten Software in einer einheitlichen und streng systematisierten Weise erfolgen. Zweckmäßigerweise werden insbesondere die in Schritt 1 von `WFCLC` gelieferten Wellenfunktionen immer schon im endgültigen – d e m – Dateiformat abgelegt, welches die später benötigten Elektrodeninformationen vollständig und in einer vorab festzulegenden, systematisierten Form enthält. Aus der Sicht der Programmentwicklung und -änderung resultiert hieraus eine recht enge – und aus dieser Sicht auch recht beschwerliche – Verzahnung der verschiedenen Software-Teile.

Als entscheidend bei der Entwicklung erwiesen sich somit zwei Dinge: Die systematische und für die gesamte Software verbindliche Beschreibung der Elektrodengeometrien in einer Geometrieklasse und das Bereitstellen eines flexiblen Dateiformates zur Speicherung

der Wellenfunktionen, in dem neben den Daten auch komplexe Elektrodeninformationen untergebracht werden können, und das auch auf spätere Erweiterungen vorbereitet ist. Beide Probleme werden im Rahmen eines Objektes `WaveFunction` gelöst, das vereinfacht so dargestellt sei:

$$\boxed{\texttt{WaveFunction}} = \boxed{\text{Daten}} + \boxed{\text{Geometrie}} + \boxed{\text{WF-Methoden}}.$$

Und die wichtigsten WF-Methoden (Funktionen) sind eben das Schreiben und Lesen des nämlichen WF-Dateiformates. `WaveFunction` ist das Elementarobjekt der Software. Es behandelt den einzelnen Eigenzustand einer Elektrode im Sinne der nachfolgenden STM-Aufgaben vollständig, d.h. alle später benötigten Elektrodeninformationen sind hierin bereits enthalten. Dieses Objekt ist gemeinsamer Ausgangspunkt aller Programme, die auf Wellenfunktionen zugreifen, und das WF-Dateiformat kann quasi als die Existenzform dieses Objektes auf einem Massenspeicher angesehen werden.

Ein Blick auf das Schema E.1 verrät, in welcher hierarchischen Ordnung ausgehend von `WaveFunction` nunmehr einzelne Elektroden und schließlich ein den Abbildungsprozeß simulierendes „STM“ aufzubauen sind:

$$\boxed{\text{Wellenfunktion: } \psi_i} \rightarrow \boxed{\text{Wellenfunktionssatz: } \{\psi_i\}} \rightarrow \boxed{\text{STM: } \{\psi_i\} + \{\phi_j\}}$$

Aus der Sicht der jeweils höheren Ebene stellen sich die untergeordneten Einheiten dar als weitgehend eigenständig und abgeschlossen – gewöhnlich der ideale Ausgangspunkt für einen objektorientierten Programmansatz, insbesondere, um für jede Einheit (jedes Objekt) eine enge Kapselung von Daten und auf diese wirkende Methoden zu erreichen. Nun wurden die Programme nicht im objektorientierten C++, sondern in C geschrieben, da zu Beginn der Arbeiten auf dem Entwicklungsrechner kein C++ Compiler zur Verfügung stand, und auch eine möglichst weite Portierbarkeit auf andere Systeme/Umgebungen angestrebt wurde, in denen fast immer C, aber nicht immer C++ implementiert ist. Der objektorientierte Ansatz erweist sich hier jedoch als derart zweckmäßig, daß wichtige Eigenschaften objektorientierten Programmierens wie Kapselung und Vererbung in C nachgebildet wurden, was in der zeigerorientierten Sprache C im übrigen keine größeren Schwierigkeiten bereitet[1]. Speziell die Kapselung von Daten und Methoden ist bei solchen komplexen Programmen keine Frage der Ästhetik, sondern der Fehlervermeidung und damit der Entwicklungszeit. (Gleiches in FORTRAN hätte wesentlich länger gebraucht.)

Einige Bemerkungen abschließend zum WF-Dateiformat. Zu jeder Wellenfunktion sind neben den eigentlichen Datenfeldern eine ganze Reihe von jeweils sehr spezifischen Informationen über die zugehörige Elektrode zu speichern, und auch neue Elektrodenformen mit bis dato unbekanntem Informationsbedarf müssen aufgenommen werden können. Die simpelste Lösung wäre hier sicher je eine Daten- und eine Informationsdatei pro Wellenfunktion. Hat man aber mit mehreren hundert Wellenfunktionen zu tun, die auf unterschiedlichen Plattformen berechnet wurden, auf verschiedenen Massenspeichern aufbewahrt sind, zur

[1] Das Wichtgste in Kurzform: Die Kapselung wird erreicht durch Strukturen (Typ `struct`), die als Elemente neben Daten auch Zeiger auf Funktionen enthalten – die Methoden der Objekte. Die Strukturen sind zu initialisieren, wobei die Funktionszeiger auf „normale“ Funktionen gerichtet werden. In der Parameterliste jeder Methode ist beim Aufruf dann stets ein Zeiger auf die eigene Struktur zu übergeben, damit die Methode Zugriff auf das eigene Objekt erhält, wie dies im übrigen auch bei echten objektorientierten Sprachen der Fall ist – der `this`-Zeiger in C++, der `self`-Zeiger in Turbo-Pascal –, nur daß dies dort der Compiler von sich aus erledigt.

STM-Berechnung übers Netz zusammengeführt werden müssen, etc., geht leicht die Übersicht verloren und erfahrungsgemäß verschwindet nach geraumer Zeit entweder die Daten- oder die Informationsdatei (oder beide). Eine Datei pro Wellenfunktion ist anzustreben – im Dateikopf die Informationen und im Rumpf die Datenfelder.

Hier tritt folgende Schwierigkeit auf: Zahlen werden bei den verschiedenen Prozessortypen nicht auf einheitliche Weise codiert, d. h. die Interpretation der einzelnen Bits (meist nur der Bytes) z. B. einer Double-Zahl ist unterschiedlich. Will man die Rechenleistung verschiedener Plattformen nutzen, kann man bei der Speicherung der Ergebnisse diesen Umstand nicht ignorieren. Für die Datenfelder selbst ist dieses Problem leicht zu beherrschen, da dort eine Vertauschung meist nur der Byteordnung von einfacher Art vorzunehmen ist. Alle anderen Informationen (also auch alle Zahlen) im Dateikopf müssen dagegen als Strings gespeichert werden. Nur so ist letztlich Unabhängigkeit möglich bzgl. solcher Fragen, wie der Byteordnung, der internen Darstellung einer Struktur oder der Zahl der Bytes einer Integerzahl.

Die WF-Datei ist daher eine Binärdatei, die im Dateikopf alle sekundären Informationen als Strings enthält, und im Rumpf die Felddaten im Binärformat. Um dabei größte Elastizität zu wahren, sind die Informationen im Dateikopf zu Blöcken zusammengefaßt, die jeweils durch eine innerhalb der Datei eindeutige Blockmarke eingeleitet werden – z. B. Block A: Felddimensionen, B: numerische Parameter, C: Geometrie, Z: binäre Felddaten, etc. Blöcke können auch geschachtelt werden. Innerhalb eines Blockes geht jeder Informationen dann nochmals eine – in diesem Block eindeutige – Marke voran: `E=`...für die Energie. Dadurch können später beliebige Zusatzinformationen sogar in die Blöcke eingefügt werden. Ältere WF-Dateien enthalten diese Informationen natürlich nicht, das Lesen dieser Dateien (mit einem aktuallisierten Programm) ist aber problemlos, da die entsprechenden Marken nicht erscheinen und die entsprechenden Informtionen folglich auch nicht erwartet werden. Als Nebeneffekt sind die Dateiköpfe „im Notfall“ so auch einmal im Rohformat, mit Hilfe eines Texteditors, zu deuten.

Literaturverzeichnis

[1] O. Agam und S. Fishman, J. Phys. A: Math. Gen. **26**, 2113 (1993).

[2] O. Agam und S. Fishman, Phys. Rev. Lett. **73**, 806 (1994).

[3] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).

[4] K.-J. Bathe und E. L. Wilson, Int. J. Num. Meth. Engng. **6**, 213 (1973).

[5] F. L. Bauer, ZAMP **8**, 214 (1967).

[6] G. Berendt und E. Weimar, *Mathematik für Physiker, Bd. II, 2. Auflage* (VCH Verlagsgesellschaft mbH, Weinheim, 1990).

[7] M. V. Berry, J. Phys. A: Math. Gen. **10**, 2083 (1977).

[8] M. V. Berry, Ann. Phys. **131**, 163 (1981).

[9] M. V. Berry, Proc. R. Soc. London A **423**, 219 (1989).

[10] M. V. Berry und J. P. Keating, J. Phys. A: Math. Gen. **23**, 4839 (1990).

[11] M. V. Berry und J. P. Keating, Proc. R. Soc. London A **437**, 151 (1992).

[12] M. V. Berry und K. E. Mount, Rep. Prog. Phys **35**, 315 (1972).

[13] M. V. Berry und M. Tabor, Proc. R. Soc. London A **349**, 101 (1976).

[14] M. V. Berry und M. Tabor, Proc. R. Soc. London A **356**, 375 (1977).

[15] G. Binnig, H. Rohrer, C. Gerber, und E. Weibel, Phys. Rev. Lett. **49**, 57 (1998).

[16] E. B. Bogomolny, Physica D **31**, 169 (1988).

[17] O. Bohigas und M. J. Giannoni, in *Mathematical and computational methods in nuclear physics*, Vol. 209 of *Lecture Notes in Physics*, edited by J. S. Dehesa, J. M. G. Gomez, und A. Polls (Springer-Verlag, New York, 1984), pp. 1–99.

[18] O. Bohigas, M. J. Giannoni, und C. Schmit, Phys. Rev. Lett. **52**, 1 (1984).

[19] M. S. Chung, T. E. Feuchtwang, und P. H. Cutler, Surf. Sci. **187**, 559 (1987).

[20] R. Courant und D. Hilbert, *Methoden der mathematischen Physik, Bd. I, Zweite Auflage* (Springer-Verlag, Berlin, Heidelberg, New York, 1968).

[21] P. Cvitanovic und B. Eckhardt, Phys. Rev. Lett. **63**, 823 (1989).

[22] G. Doyen und D. Drakova, Surf. Sci. **178**, 375 (1986).

[23] G. Doyen, D. Drakova, E. Kopatzki, und R. J. Behm, J. Vac. Sci. Technol. A **6**, 327 (1988).

[24] I. H. Duru und H. Kleinert, Phys. Letters B **84**, 30 (1979).

[25] B. Eckhardt, G. Hose, und E. Polak, Phys. Rev. A **39**, 3776 (1989).

[26] T. E. Feuchtwang, P. H. Cutler, und N. M. Minkovsky, Phys. Scr. **35**, 132 (1987).

[27] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[28] M. C. Gutzwiller, J. Math. Phys. **8**, 1979 (1967).

[29] M. C. Gutzwiller, J. Math. Phys. **10**, 1004 (1969).

[30] M. C. Gutzwiller, J. Math. Phys. **11**, 1791 (1970).

[31] M. C. Gutzwiller, J. Math. Phys. **12**, 343 (1971).

[32] M. C. Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer Verlag, New York, Berlin, Heidelberg, 1990).

[33] W. Hackbusch, *Theorie und Numerik elliptischer Differentialgleichungen* (B. G. Teubner, Stuttgart, 1986).

[34] W. Hackbusch, *Iterative Lösung großer schwachbesetzter Gleichungssysteme* (B. G. Teubner, Stuttgart, 1991).

[35] E. J. Heller, Phys. Rev. Lett. **53**, 1515 (1984).

[36] M. Hietschold und H. Sbosny, Ultramicroscopy **42–44**, 200 (1992).

[37] A. Jennings, in *Sparse matrices and their uses*, edited by I. S. Duff (Acad. Press, London, New York, Toronto, Sydney, San Francisco, 1981), pp. 109–138.

[38] R. V. Jensen, M. M. Sanders, M. Saraceno, und B. Sundaram, Phys. Rev. Lett. **63**, 2771 (1989).

[39] C. Jung, Can. J. Phys. **58**, 719 (1980).

[40] O. Jusko, X. Zhao, und G. Wilkening, Technisches Messen **61**, 376 (1994).

[41] J. P. Keating, Proc. R. Soc. London A **436**, 99 (1992).

[42] A. Kiełbasiński und H. Schwetlick, *Numerische lineare Algebra* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1988).

[43] H. Kleinert, *Pfadintegrale in Quantenmechanik, Statistik und Polymerphysik* (BI-Wiss.-Verlag, Mannheim, Leipzig, Wien, Zürich, 1993).

[44] R. Köning und L. Koenders, Technisches Messen **61**, 382 (1994).

[45] E. Kopatzki, G. Doyen, D. Drakova, und R. J. Behm, J. Microscopy **152**, 687 (1988).

[46] T. Laloyaux, I. Derycke, J.-P. Vigneron, P. Lambin, und A. A. Lucas, Phys. Rev. B **47**, 7508 (1993).

[47] T. Laloyaux, A. A. Lucas, J.-P. Vigneron, P. Lambin, und H. Morawitz, J. of Microscopy **152**, 53 (1988).

[48] L. D. Landau und E. M. Liftschitz, *Lehrbuch der Theoretischen Physik, Band III, Quantenmechanik* (Akademie-Verlag, Berlin, 1988).

[49] N. D. Lang, Phys. Rev. Lett. **55**, 230 (1985).

[50] N. D. Lang, Phys. Rev. B **34**, 5947 (1986).

[51] N. D. Lang, Phys. Rev. B **36**, 8173 (1987).

[52] N. D. Lang und W. Kohn, Phys. Rev. B **1**, 4555 (1970).

[53] C. R. Leavens und G. C. Aers, Phys. Rev. B **38**, 7357 (1988).

[54] R. L. Liboff, J. Math. Phys. **35**, 596 (1994).

[55] B. A. Lippmann, Phys. Rev. Lett. **15**, 11 (1965).

[56] B. A. Lippmann, Phys. Rev. Lett. **16**, 135 (1966).

[57] G. Maess, *Iterative Lösung linearer Gleichungssysteme*, Nr. 238 Bd. 52 der Reihe Nova Acta Leopoldina (Deutsche Akademie der Naturforscher Leopoldina, Halle (Saale), 1979).

[58] S. W. McDonald, *Wave Dynamics of Regular and Chaotic Rays*, Ph.D.Thesis, University of California, 1983.

[59] S. W. McDonald und A. N. Kaufman, Phys. Rev. Lett. **42**, 1189 (1979).

[60] S. W. McDonald und A. N. Kaufman, Phys. Rev. A **37**, 3067 (1988).

[61] U. Meißner und A. Menzel, *Die Methode der finiten Elemente* (Springer Verlag, Berlin, 1989).

[62] A. Meyer, Wiss. Inf. TH Karl-Marx-Stadt **36**, (1983).

[63] A. Meyer, *Die simultane Berechnung einiger Eigenwerte und Eigenvektoren beim großdimensionierten allgemeinen Matrixeigenwertproblem*, Dissertation B, Technische Hochschule Karl-Marx-Stadt, 1985.

[64] W. H. Miller, J. Chem. Phys. **63**, 996 (1975).

[65] J. B. Pendry, A. B. Prêtre, und B. C. H. Krutzen, J. Phys. C: Condens. Matter **3**, 4313 (1991).

[66] W. H. Press, S. A. Teukolsky, W. T. Vetterling, und B. P. Flannery, *Numerical Recipes in Fortran, Second Edition* (Cambridge University Press, Cambridge, 1992).

[67] J. R. J. Riddell, J. Computational Physics **31**, 21 (1979).

[68] G. Reiss, F. Schneider, J. Vancea, und H. Hoffmann, Appl. Phys. Lett. **57**, 867 (1990).

[69] G. Reiss, J. Vancea, H. Wittmann, J. Zweck, und H. Hoffmann, J. Appl. Phys. **67**, 1156 (1990).

[70] H. Rutishauser, Num. Math. **3**, 4 (1969).

[71] H. Rutishauser, Num. Math. **16**, 205 (1970).

[72] H. Sbosny, L. Koenders, und M. Hietschold, Thin Solid Films (1995), im Druck.

[73] S. Sridhar, Phys. Rev. Lett. **67**, 785 (1985).

[74] S. Sridhar und E. J. Heller, Phys. Rev. A **46**, 1728 (1992).

[75] J. Stein, *Experimentelle Bestimmung von Billard-Eigenfunktionen*, Dissertation, Universität Marburg, 1993.

[76] J. Stein und H.-J. Stöckmann, Phys. Rev. Lett. **68**, 2867 (1992).

[77] J. Tersoff und D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983).

[78] J. Tersoff und D. R. Hamann, Phys. Rev. B **31**, 805 (1985).

[79] W. Tornig und P. Spellucci, *Numerische Mathematik für Ingenieure und Physiker* (Springer Verlag, Berlin, 1988).

[80] D. Wintgen und A. Hönig, Phys. Rev. Lett. **63**, 1467 (1989).

Abbildungsverzeichnis

1.1 Barrierenpotential 6
1.2 Die Reisssche Korrektur 10

2.1 Das STM-Modell 13

4.1 xy-Einbettungsgebiet 30
4.2 Beispiel einer lexikographischen Numerierung der Gitterpunkte 32
4.3 Beispiel einer Eigenwertmatrix in drei Dimensionen 34
4.4 Diskretisierungsfehler über der Energie 46
4.5 Diskretisierungsfehler über ξ und η 47
4.6 Diskretisierungsfehler für einen Potentialtopf 49

5.1 Zustände des nichtseparablen Rechtecks 59
5.2 Zum Test: Zustand im Kreis 60
5.3 Maßstabsgetreues Abbild einer Spitze mit $R = 8$ Å und $\alpha = 90^o$ 61

6.1 LDOS-Profile von Graben, Kerbe und Dreispitz 68
6.2 Konstantstromprofile eines Grabens für $R = 1,8$ Å 70
6.3 Konstantstromprofile eines Grabens für $R = 20$ Å 71
6.4 Konstantstromprofile einer Kerbe für $R = 8,20$ Å 73
6.5 Konstantstromprofile eines Dreispitz für $R = 8,20$ Å 75

A.1 Die Funktion F(k) 87
A.2 E(k) für verschiedene Schrittweiten 88

B.1 Grundprinzip der Kaskadensummation 100
B.2 Statistik der minimalen Iterationslücken $\Delta\beta_{s,m}^{\min}$ 106
B.3 Konvergenzraten der Tschebyscheff-Beschleunigung 107

C.1 Indizierung einer Linienmatrix 110
C.2 Aufspaltung der Summe $c_{ik} = \sum_j a_{ij}\, b_{jk}$ 111
C.3 Beispiel einer Multiplikation von Linienmatrizen 113

Symbole und Abkürzungen

$\mathbf{R}^n$	:	Menge der reellen n-dimensionalen Vektoren
$\mathbf{R}^{n,m}$	:	Menge der reellen $(n \times m)$-Matrizen
$\mathbf{S}^{n,n}$	:	Menge der symmetrischen reellen $(n \times n)$-Matrizen
$\mathbf{I}$	:	Einheitsmatrix
$\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$	:	Matrizen;
α, β, γ	:	Eigenwerte der Matrizen $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$
α, R	:	Flankenwinkel und Radius der Spitzen
D	:	Raumdimension
V_0	:	Wandhöhe der Potentialkästen
Ω_K, Ω_K^*	:	inneres und diskretisiertes inneres Gebiet eines Potentialkastens
Ω_G	:	Grundgebiet der Diskretisierung
h	:	Schrittweite des Diskretisierungsgitters
$\boldsymbol{\nabla}, \boldsymbol{\nabla}_{ij}$	:	Nabla-Operator des Kontinuums und des Gitters
F_{ij}, f_{ij}	:	Diskretisierungsfehler am Gitterort $\underline{\mathbf{r}}_{ij}$
E, E_h	:	Energie der kontinuierlichen und der Gitterzustände
λ, λ_h	:	Wellenlänge der kontinuierlichen und der Gitterzustände
κ, κ_h	:	evanescente Abklinglänge der kontinuierlichen und der Gitterzustände
E_F	:	Fermi-Energie
k_B	:	Boltzmannkonstante
T	:	absolute Temperatur
ϱ	:	lokale Zustandsdichte
ρ	:	Verschiebung in $(\mathbf{A} - \epsilon\mathbf{I})^2 - \rho\mathbf{I}$
DGL	:	Abk. für Differentialgleichung
LDOS	:	Abk. für local density of states (lokale Zustandsdichte)
SGL	:	Abk. für Schrödingergleichung
THF	:	Abk. für Transfer-Hamiltonian-Formalismus

Selbständigkeitserklärung

Ich erkläre, daß ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Chemnitz, den 06. 04. 1995 Hartmut Sbosny

Thesen

1. Der STM-Abbildungsprozeß wurde im Rahmen eines zweidimensionalen Modells simuliert, worin Probe und Spitze durch Sommerfeld-Metalle frei wählbarer Geometrie beschrieben wurden und die Bestimmung des Tunnelstroms im Transfer-Hamiltonian-Formalismus erfolgte. Berechnet wurden Konstantstromprofile für die Abbildung dreier idealtypischer Probengeometrien – Kerbe, Graben, Dreispitz – mit Spitzen unterschiedlicher Radien. Hierbei existierten größere Bereiche „unsichtbarer" Gebiete – Gebiete, die einer geometrischen Spitzenberührung nicht zugänglich wären.

2. Über eine Entfaltung der Konstantstromprofile mit der aktuellen Spitzengeometrie nach Reiss erfolgte die Gegenüberstellung insbesondere zum geometrischen (mechanischen) Abtasten. Die Diskussion konzentrierte sich auf Möglichkeiten, aus diesem Vergleich genauere Aussagen zur Probengeometrie und zur Kalibrierung von Spitzen zu gewinnen.

3. Die Konstantstromprofile waren dominiert von der Form der Spitzen und ähnelten Abtastprofilen. Folgerichtig konnte die Reisssche Entfaltung in Gestalt von Häufungspunkten die laterale Position der Ränder der „unsichtbaren" Gebiete näherungsweise gut wiedergeben.

4. Dem überlagert zeigten die lateralen Positionen der Häufungspunkte ein Sklalierungsverhalten mit der Scanhöhe, dem Informationen über den Flankenwinkel und den Spitzenradius innewohnen. Erstere Abhängigkeit folgt der Tendenz der LDOS-Konturen der Probe, letztere wird qualitativ verständlich aus dem prinzipiellen Unterschied zwischen einem geometrischen Abtasten und einer Konstantstrom-Abbildung: an scharfen Probenregionen erfährt der Strompfad gegenüber planaren Gebieten eine Einschnürung, die relativ um so größer ist, je stumpfer die Spitze ist, ein Effekt, der bei einem Abtasten unter punktförmigem Kontakt prinzipiell nicht auftreten kann; Konstanstromprofile gewinnen *relativ zu* Abtastprofilen an Schärfe, je stumpfer die Spitzen werden.

5. In den entfalteten Kurven verblieben über den „unsichtbaren" Gebieten stets einige Punkte, die mit dem Spitzenradius korrelierten, und die mit dem Umspringen des Strompfades am Spitzenschaft in Zusammenhang gebracht werden können. Für die Kalibrierung von Spitzen ein besonders interessanter Aspekt.

6. Einen semiklassischen Zugang zu den Phänomenen in stationären nichtintegrablen Quantensystemen bietet nur die Theorie periodischer Orbits, wobei diese Theorie gegenwärtig auf gewöhnlichen klassischen Orbits fußt. Als markantes Signum eines chaotischen Systems hinsichtlich der Wellenfunktion gelten Scar-Zustände (zentriert um einen klassischen Orbit).

7. Verglichen wurden die Eigenzustände $\Psi_n^{(1)}$ eines nichtseparablen Rechtecks (2d-Topf) der Wandhöhe V_0 (schwach nichtintegrabel) und die Zustände $\Psi_n^{(2)}$ eines geometriegleichen separablen Potentialgebildes, wie es aus der Überlagerung eindimensionaler

Kästen entsteht. Der Unterschied *integrabel* $\Leftrightarrow$ *nichtintegrabel* wäre hier für klassische Orbits mit $E < V_0$ unsichtbar, weil Differenzen zwischen beiden Potentialen nur im Außenraum bestehen. Die Zustände $\Psi_n^{(1)}$ und $\Psi_n^{(2)}$ zeigten sich nahezu identisch. Jedoch nahmen die Differenzen $|\Psi_n^{(1)}|^2 - |\Psi_n^{(2)}|^2$ bei einem Teil der Zustände das Aussehen von Scars an. Eine Beschreibung in einer Theorie periodischer Orbits müßte entweder komplexe Orbits oder Bahnen (Orbits) aus dem Bereich des Kontinuums oberhalb V_0 einbeziehen!

8. Für ein Zwei-Billardsystem „Rechteck + Kreis“ – zwei benachbarte Potentialkästen endlicher Wandhöhe, die über eine Tunnelbarriere koppeln – wurden Gesamtzustände berechnet. Ausgeprägt insbesondere im Rechteck fanden sich Scars, die aus der Barriere austreten und zu dieser zurücklaufen. Das System „Rechteck + Kreis“ ist in hohem Maße nichtintegrabel, „sichtbar“ wird diese Nichtintegrabilität jedoch nur für Bahnen entweder des Kontinuums oder für komplexe Orbits. Eine semiklassische Beschreibung dieses Phänomens im Rahmen einer Theorie periodischer Bahnen würde eine Ausweitung auf komplexe Orbits oder Bahnen des Kontinuums („ungebundene Orbits“) erfordern!

9. Zur numerischen Lösung der zeitfreien Schrödingergleichung für beliebig geformte Potentialmulden wurde diese im Differenzenverfahren diskretisiert. Für die resultierende große Diskretisierungsmatrix $\mathbf{A}$ war das Eigenwertproblem für einige wenige Eigenwerte $\lambda_i \approx \epsilon$ aus der Mitte des Spektrums zu lösen. Die Sachlage dabei ist schwierig, weil die in solchen Fällen übliche, gut konvergierende Inverse Vektoriteration hier aufgrund der Größe von $\mathbf{A}$ die iterative(!) Lösung eines indefiniten Gleichungssystems unter sehr schwachen Regularitätsvoraussetzungen erfordern würde. Zur Umgehung dieses Problems wurde mit der quadrierten und geeignet verschobenen Matrix $(\mathbf{A} - \epsilon\mathbf{I})^2 - \rho\mathbf{I}$ direkt vektoriteriert. Der Preis ist eine außergewöhnlich schwache Konvergenz.

10. Eine Analyse der Diskretisierungsfehler ergab, daß mit dem benutzten Diskretisierungsgitter diese bzgl. Eigenwert, Verhalten der Wellenfunktionen im Innenraum und Verhalten im Außenraum unter 1% bleiben sollten.

Lebenslauf

14. 6. 1962	geboren in Eisenhüttenstadt,
1969 – 1977	Besuch der Polytechnischen Oberschule in Dessau
1977 – 1981	Besuch der Erweiterten Oberschule in Dessau
3. 7. 1981	Abitur
1981 – 1983	18monatiger Grundwehrdienst
1983 – 1988	Studium der Physik an der Technischen Universität Karl-Marx-Stadt
1986	Gaststudium an der Universität Leipzig
1. 8. 1988	Physik-Diplom zum Thema „Experimentelle und theoretische Untersuchugen zu steuerbaren Tunnelübergängen“
1989 – 1992	Forschungsstudium an der TU Chemnitz
1993	Mitarbeit am Projekt „Berechnung von STM-Profilkurven“ zwischen der PTB Braunschweig und der TU Chemnitz
seit 1994	verschiedene wiss. Anstellungen an der TU Chemnitz-Zwickau am Lehrstuhl Analytik an Festkörperoberflächen
10. 4. 1995	Einreichung vorliegender Dissertation