



Graz University of Technology

Institute of Electronics

Masterarbeit

BILDANALYSE ZUR REKONSTRUKTION
VON INTEGRIERTEN SCHALTKREISEN IM
NANOMETERBEREICH

Udo Bachernegg

Graz, Österreich, Juni 2011

unter Anleitung von
Prof. Dr. Wolfgang Pribyl
Dipl. Phys. Bernhard Lippmann

Erklärung

Ich erkläre an Eides Statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Ich versichere, dass ich dieses Diplomarbeitsthema bisher weder im In- noch im Ausland (einem Beurteiler oder einer Beurteilerin) in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Graz, den 1. September 2011

Zusammenfassung

Das Design eines konkurrenzfähigen Produkts kann viele Jahre an Entwicklungsaufwand und viele Millionen Euro an Kosten mit sich bringen. Das Schützen der aus diesem Prozess hervorgegangenen Erfindungen durch Patente zählt zu einem notwendigen Vorgang, um das geistige Eigentum und damit die Marktstellung eines Unternehmens zu sichern.

In der folgenden Arbeit wird ein Konzept zur Rückgewinnung von Designinformation und Designparametern aus Integrierten Schaltungen vorgestellt, um patentrelevante Fragestellungen zu beantworten. Die Schwerpunkte liegen hierbei in der Evaluierung von subpixelgenauen Bildregistrierungsmethoden für dreidimensionale Mosaike mit zehntausenden von Bildern sowie in der Evaluierung von bildverarbeitungsgestützten Methoden zur (semi)automatischen Extraktion von topologischer Information. Die Kombination der gewonnenen Daten gestattet die Ableitung von schaltungsbezogenen Informationen und das Erstellen einer planaren Netzliste. Durch Auffinden von Hierarchien ist es zudem möglich, eine modulare Darstellung der Netzliste zu gewinnen. Als Bildmaterial dienen Aufnahmen von Rasterelektronenmikroskopen sowie Bilder von optischen Mikroskopen. Aufgrund der geringen Strukturgrößen auf den Proben sowie der enormen Komplexität aktueller Designs entstehen bei der Abbildung große Mengen an Daten, die verarbeitet werden müssen. Hierzu wurden in der vorliegenden Arbeit effiziente Methoden entwickelt, die einen wirtschaftlichen Nutzen überhaupt erst ermöglichen.

Abstract

To design a competitive integrated circuit can involve many years of development effort and expenses of many million Euros. The protection of the intellectual property emerged from this process and, with that, a company's market position, is a necessary procedure.

In the present thesis a concept for the reverse engineering of design information and design parameters from integrated circuits is introduced, in order to answer questions relevant to patents. The focus lies on the evaluation of image registration methods with sub pixel accuracy for three-dimensional mosaics with ten thousands of images, as well as on the evaluation of methods based on image processing for (semi) automatic extraction of topologic information. The combination of the data obtained allows derivations from circuit related information and the creation of a planar netlist. Additionally, it is possible to obtain a modular presentation of the netlist by detecting hierarchies. Images from a scanning electron microscope, as well as from optical microscopes, serve as graphic material. As a result of the small structures on the samples, as well as the vast complexity of current designs, the imaging produces a huge amount of data that needs to be processed. For this, the following thesis presents effective methods to challenge that task.

Danksagungen

Mein besonderer Dank gilt meinem herausragenden Betreuer Bernhard, der sich sehr für mich einsetzte und von dem ich sehr viel lernen konnte. Ausserdem Dr. Roland Perko sowie Prof. Werner Grogger für die ausgezeichnete fachliche Unterstützung.

Prof. Pribyl sowie Gerald danke ich für Organisation und Management.

Meinen Eltern Maria und Erwin für die bedingungslose Unterstützung und den Zuspruch meine ganze Studienzeit hindurch. Meinem Bruder Oliver und meinen leider viel zu selten gesehenen Geschwistern Lisa und Marko.

Tante Gerti für die Hilfestellung bei jeglicher Art von Verletzungen und Tante Regina für die Unterstützung bei kochtechnischen Fragen.

Meiner lieben Freundin ♡Lisa♡ und unserem Hund Molly, die es teilweise nicht leicht mit mir hatten ;-).

Des weiteren möchte ich mich bei meinen Freunden und Freundinnen Patrik, Flo, Markus, Thomas, Martin, Gogo, Slu, CC, Patrik, Gustl, Oliver, Christian, Tobi, Tina, Petra, Roli&Sabine, Kai&Melanie, Chrissi&Martina, Klaus&Gudi, Flo, Eva, Murla, Hoagie, Kurti, Lo&Alice, Hans Peter und seinem Hund Fritz, Flurl, Jakob, Georg, Kristina, Manu, Martin, Andi, Andrea&Hannes, Astrid, ... für die unzähligen lustigen Stunden bedanken, die mich erfolgreich vom Schreiben meiner Diplomarbeit abgehalten haben :-).

Meinen Kollegen von der ReadRF-Crew Edi, Emsi, Markus und Matthias für die erheiternde Zeit abseits der Arbeit (Party ON!), Albert, Koro, Walter für kompetente und freundliche Hilfestellung bei kniffligen Fragestellungen sowie Roi, Simon, Hannes und Szabolcs für das konstruktive Miteinander.

Auch meinen Kollegen aus Deutschland Herbert, Anja, Sabine, Kathi sei für die anregenden Diskussionen gedankt.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Stand der Technik	3
1.3	Aufbau der Arbeit	5
2	Schichtstruktur	6
2.1	Einleitung	6
2.2	Schichtaufbau	7
2.3	Schichtabtrag	9
3	Abbildungsverfahren	16
3.1	Motivation	16
3.2	Abbildungsverfahren der Prozessanalyse	17
3.3	Abbildungsverfahren der Schaltungsanalyse	17
3.3.1	Lichtmikroskopie	18
3.3.2	Rasterelektronenmikroskopie	19
3.3.3	Verwendete Abbildungsgeräte und Parameter	28
4	Bildanalytisch gestütztes Reverse Engineering	30
4.1	Einleitung	30
4.2	Überblick	30
5	Stitching	33
5.1	Einleitung	33
5.2	Überblick	35
5.3	Merkmalsextraktion und Matching	37
5.3.1	Auf Merkmalen basierende Methoden	37
5.3.1.1	Stand der Technik	37
5.3.2	Direkte Methoden	42
5.3.2.1	Einführung	42
5.3.2.2	Stand der Technik	46
5.3.2.3	Subpixel-Genauigkeit	48
5.3.3	Gewählte Verfahren	49

5.4	Koordinatentransformationen	51
5.4.1	Kalibrierung and Kameramodell	51
5.4.2	Struktur der Transformation	52
5.4.3	Abschätzung der Chi Ebenentransformation	55
5.4.4	Abschätzung der globalen Chiptransformation	59
5.5	Genauigkeitsanalysen	64
5.5.1	Lokale Bildtransformation	64
5.5.2	Chi Ebenentransformation	68
5.5.3	Globale Chiptransformation	72
5.6	Radiometrische Korrekturen	73
5.6.1	Korrektur der Bildhelligkeit	73
5.6.2	Überblendung	75
6	Schaltungsextraktion	78
6.1	Einleitung	78
6.2	Stand der Technik	80
6.3	Leiterbahnenextraktion	81
6.3.1	Bildvorverarbeitung	83
6.3.2	Segmentierung	84
6.3.2.1	Kantenbasierte Segmentierung	84
6.3.2.2	Modellbasierte Segmentierung	91
6.3.3	Vektorisierung	97
6.4	Kontaktextraktion	97
6.5	Extraktion von Bauteilen	102
6.6	Routing	102
6.7	Netzliste	104
6.8	Zusammenfassung and weiterführende Fragestellungen	105
7	Zellenextraktion	106
7.1	Einleitung	106
7.2	Extraktion digitaler Standardzellen	107
7.2.1	Mustererkennung	108
7.2.2	Ableitung von Kennzahlen	110
7.2.3	Standardzellenstatistiken	114
7.2.4	Clock-Tree	114
7.3	Extraktion von Speicherzellen	115
7.4	Verteiltes Rechnen	116
A	Chipsan Analysis Tool	117
A.1	Projektkonfiguration	117
A.2	Stitching	121
A.3	Leitungs- und Kontaktextraktion	126
A.4	Zellenextraktion	131

Literaturverzeichnis

133

Abbildungsverzeichnis

1.1	Beziehung zwischen forward engineering und reverse engineering.	3
2.1	Typischer Schichtaufbau eines IC	10
2.2	Schichtaufbau eines IC unter dem TEM	11
2.3	Durch Skelettätzen freigelegte Metallisierungsebene	12
2.4	Freilegung von Metallschichten	15
3.1	Präparation einer TEM-Lamelle durch eine FIB	18
3.2	Gegenüberstellung der Auflichtmikroskopie und der Laser- Konfokalmikroskopie	20
3.3	Interferenzerscheinungen an dünnen Schichten	21
3.4	Prinzip eines Rasterelektronenmikroskops	23
3.5	Interaktionen eines Elektronenstrahls mit Materie	24
3.6	Elementbestimmung einer Flash Speicherzelle durch EDX	25
3.7	Typisches Energiespektrum emittierter Elektronen	26
3.8	Eindringtiefe von Elektronen in Materie	27
3.9	Simulation eines in die Probe eindringenden Elektronenstrahls	28
3.10	Raith 200 Chip-Scanner	29
4.1	Der Bildanalyse RE Workflow	31
5.1	Extraktion von POIs am Siemensstern zur Evaluierung der vorgestellten Detektoren	43
5.2	Korrelationskoeffizienten der NCC und PC	46
5.3	Subpixelgenauigkeit durch Regression	50
5.4	Koordinatensysteme und überführende Transformationen	52
5.5	Typen von planaren 2D Transformationen	53
5.6	Verknüpfung lokaler Bildtransformationen und daraus resultierende Fort- pflanzungsfehler	56
5.7	Minimaler Spannbaum zur Abschätzung der Chipebenentransformation	57
5.8	Chipebenen nach globaler Registrierung	65
5.9	Entropie in Bildern	67

5.10	Histogramm der NCC Koeffizienten	68
5.11	Histogramm der Residuen von Bildregistrierungen durch die NCC	68
5.12	Optimierung der Chi Ebenentransformation	69
5.13	Fehlerenergie nach und vor Optimierung der Chi Ebenenkoordinaten	70
5.14	Graph der zusammenhängenden lokalen Bildtransformationen	71
5.15	Beispiele für zirkuläre Transformationsketten	72
5.16	Histogramm der Residuen von zirkulären Transformationen	73
5.17	Residuum der globalen Chiptransformation	74
5.18	Überblendungsfunktion mit variierenden Parametern	76
5.19	Zweiband Überblendung	77
6.1	Y Diagramm des Hardwareentwurfs	79
6.2	Präparationsartefakte durch Überätzung	82
6.3	3D Kantenprofil von Leitungen unter einem REM	87
6.4	Kantenbenennung	88
6.5	Leitungsextraktion einer Metallebene	91
6.6	Leiterbahnextraktion aus verschiedenen Chipbereichen	92
6.7	Probenpräparation durch Sputtern	93
6.8	Analyse des Kantenprofils	93
6.9	Leitungsextraktion mit der Hough-Transformation	95
6.10	Vektorisierung der Segmentsgrenzen	96
6.11	Kontaktextraktion durch Schwellwertbildung	99
6.12	Kontaktextraktion mit Hilfe der Distanztransformation	100
6.13	Kontaktextraktion durch die Hough-Transformation	101
6.14	Gegenüberstellung der Ebenen AA und GC	103
7.1	Standardzellen auf Polysilicium.	109
7.2	Standardzellenextraktion für 149 Zelltypen	111
7.3	Ausschnitt aus einer Standardzellenextraktion für 149 Zelltypen	112
7.4	Effektive Flächenausnutzung in einem Standardzellenbereich	113
7.5	Statistik über Zellhäufigkeit und Zellfläche	115
A.1	GUI zur Projektkonfiguration	118
A.2	Bildnummern und Aufnahmereihenfolge einer Bilderserie	119
A.3	GUI zur Bildregistrierung	122
A.4	Korrelationskoeffizienten der Debug-Ausgabe	126
A.5	GUI zur Leitungs- und Kontaktextraktion	128
A.6	GUI zur Zellenextraktion	132

Tabellenverzeichnis

2.1	Die CMOS Prozesstechnologie	9
2.2	Schichten und Verwendung findende Stoffe	12
2.3	Trockenätzverfahren	13
5.1	Hierarchie von planaren 2D Transformationen	54

If I have seen further it is only by standing on
the shoulders of giants

Sir Isaac Newton

Kapitel 1

Einleitung

Schon in frühen Tagen der Ingenieurwissenschaften spielte der Umkehrprozess der Synthese - die Analyse - eine wichtige Rolle beim Verstehen der Abläufe von Natur und Umwelt. Im Prinzip gründet der Prozess des Erschaffens von Wissen auf dem Verständnis von Gegebenem. Durch Beobachtungen der Umwelt und deren Interpretation entstanden formale Beschreibungen und daraus objektive Grundlagen, aus denen Schlussfolgerungen und damit fundiertes Wissen geschaffen werden konnte. Wenn auch anfängliche Modelle nur von geringer Genauigkeit waren und der Abstraktion nicht standhalten konnten, war es aufgrund der darauf gründenden Logik dennoch möglich, durch Falsifikation deren Richtigkeit zu widerlegen oder durch Einbringen neuer Erkenntnisse diese zu verbessern. Ob dieser Vorgang jemals abgeschlossen sein kann, d.h. die Richtigkeit der Modelle erwiesen werden kann, ist aufgrund der subjektiven menschlichen Wahrnehmung zu bezweifeln. Was jedoch bleibt, sind Modelle die für viele technische Anwendungen ausreichend genau sind. Die wahre Natur der Welt wird uns wohl dennoch verborgen bleiben.

Inhaltsverzeichnis

1.1 Motivation	1
1.2 Stand der Technik	3
1.3 Aufbau der Arbeit	5

1.1 Motivation

Das Design eines konkurrenzfähigen Produkts kann viele Jahre an Entwicklungsaufwand und viele Millionen Euro an Kosten mit sich bringen. Das Schützen der, aus diesem Prozess hervorgegangenen, Erfindungen durch Patente zählt zu einem notwendigen Vorgang um das geistige Eigentum und damit die Marktstellung eines Unternehmens zu sichern. Der Mangel an dieser Möglichkeit würde zu einer Verzerrung des Wettbewerbs führen, da sich hohe Investitionskosten in Innovation - die ein Mitbewerber quasi umsonst erlangen kann - nicht in wirtschaftlichem Erfolg niederschläge. Was aus unternehmerischer Sicht

eine Notwendigkeit darstellt, widerspricht jedoch der Idee des freien Marktes. Startup-Unternehmen wird es zunehmend schwerer gemacht, in neue Märkte einzusteigen, da diese mit einer Fülle von Patenten abgesichert sind, oft auch genau aus diesem Grund. Dies führt zu Monopolisierungen, Hemmung von Innovation und schlussendlich zum Schaden für die Allgemeinheit.

Nichts desto trotz hat sich ein System etabliert, in dem Patente einen wichtigen Stellenwert einnehmen und das Schützen derer einen mindestens genauso hohen. Durch die laufenden Patenterhaltungskosten muss auch in Kontrolle von Patentverletzungen investiert werden. In manchen Bereichen kann dies durch eine einfache Inspektion erfolgen, in anderen gestaltet es sich schwieriger und es müssen Möglichkeiten geschaffen werden, ein Konkurrenzprodukt näher zu durchleuchten.

Das Hauptaugenmerk von *reverse engineering* liegt auf dem Prozess, Rückschlüsse von der Implementierung bzw. der physikalischen Realisierung auf das Design zu gewinnen. Die zugrundeliegenden Konzepte sind in vielen Bereichen anwendbar [?], darunter Softwareentwicklung [?] und Mikroelektronik [?, ?]. Im speziellen Ablauf und vor allem im dafür notwendigen Aufwand unterscheiden sich die Anwendungen jedoch erheblich. Im Abschnitt 1.2 wird ein kurzer Überblick über *reverse engineering* im Kontext von Integrierten Schaltungen gegeben.

Reverse engineering ist für viele Firmen in der Halbleiterindustrie lebensnotwendig, um auf dem hart umkämpften Markt bestehen zu können. Laut Chipworks, einem professionellen kanadischen *reverse engineering* Unternehmen, wird diese Analysemethode von 95% aller großen Halbleiterfirmen eingesetzt. Professionelle Firmen wie Chipworks, Semiconductor Insights, Ixento u.a. bieten erfolgreich Prozess- und Schaltungsanalysen für Fremdmuster an. Im Angebot stehen *product teardowns* (Identifikation des Produkts, Analyse der Leiterplatte sowie deren Komponenten), *system level analysis* (Funktionalität, Timing, Signalpfad), *process analysis* (Struktur und Materialanalyse, Lagenaufbau) und *circuit extraction* (Komponenten, Zusammenschaltung, Netzliste). *Reverse engineering* Kunden fallen zum einen in die Gruppe der an patentrelevanten Themen Interessierten und zum anderen in die Gruppe der an der technischen Realisierung Interessierten. Technische Kunden kommen aus dem Bereich der Produktion, der Entwicklung, dem strategischen Marketing oder *Benchmarking* [?]. *Benchmarking* bietet die Möglichkeit einer kontinuierlichen vergleichenden Analyse des eigenen Unternehmens mit den Mitbewerbern, indem aus den Analysedaten Kennzahlen abgeleitet werden. Dieser Vergleich, der Rückschlüsse auf die aktuelle und zukünftige Marktstellung liefert, beeinflusst mittel- und langfristige Entscheidungen sowie die strategische Ausrichtung des Unternehmens.

Ein weiterer Anlass, *reverse engineering* zu betreiben, liegt bei Produkten vor, die bereits vom Markt genommen wurden oder bei Produkten von Firmen, die nicht weiter bestehen. In vielen Bereichen werden Mikrochips für spezielle Aufgaben in kleinen Stückzahlen produziert. Bei fehlender Verfügbarkeit sowie mangelnder Dokumentation, ist es oft wirtschaftlicher, den Mikrochip zu rekonstruieren als das komplette System neu zu entwerfen [?, ?]. Ein Überblick in den Rekonstruktionsworkflow von VLSI Chips wird in

[?] gegeben.

1.2 Stand der Technik

Allgemein versteht man unter *reverse engineering* den Prozess des Entwickelns einer Menge an Spezifikationen für ein komplexes System durch eine systematische Untersuchung von Mustern des Systems [?]. Diese Definition aus einer anerkannten Grundlagenarbeit über *reverse engineering* stellt das Gegenstück zum sogenannten *forward engineering* dar. *Forward engineering* entspricht dem traditionellen Ansatz, der ausgehend von Spezifikationen das Design bis hin zur Implementierung wiedergibt. Abbildung 1.1 zeigt die Beziehungen dieser beiden Prozesse über Transformationen in die jeweilig andere Domäne auf.

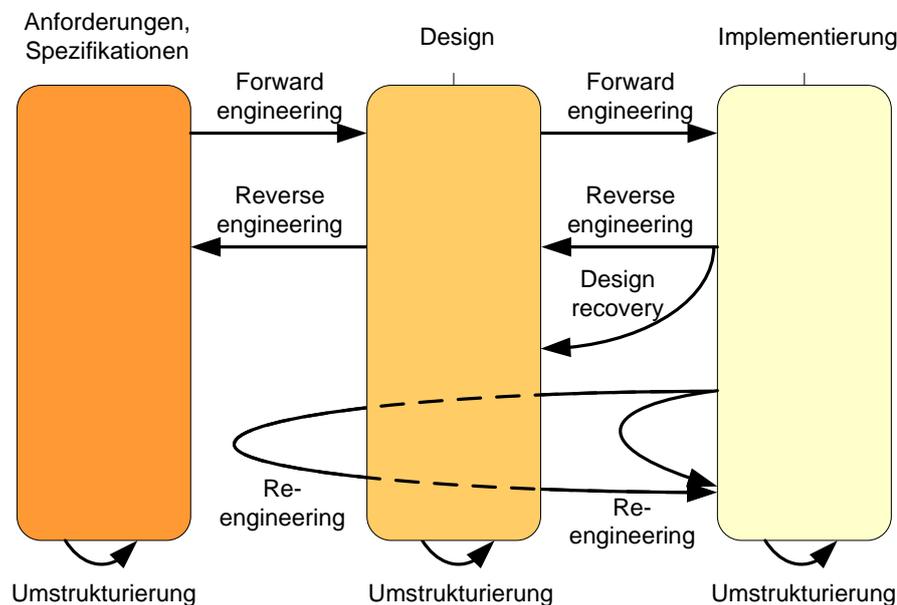


Abbildung 1.1: Ausgehend von Anforderungen an ein Produkt und die daraus definierten Spezifikationen entstehen durch den Prozess des *forward engineering* Design und Implementierung. Beim inversen Prozess wird abhängig von der gewünschten Information zwischen *reengineering*, *design recovery* und *reverse engineering* unterschieden. Diese Abbildung ist angelehnt an [?].

Im Zentrum dieser Arbeit steht *design recovery* von Hardware respektive die Extraktion von Designinformationen aus Integrierten Schaltungen.

Um Zugriff auf die in Silicium realisierten Strukturen zu bekommen, ist eine Reihe von Vorabanalysen notwendig, um die für den Schichtabtrag erforderlichen Parameter zu ermitteln. Dazu gehört die Prozessanalyse bzw. Technologieanalyse, auf die im Folgenden kurz eingegangen wird.

Prozessanalysen finden in vielen Unternehmen *in house* statt, da die dazu notwendige Ausrüstung in jeder *Fab* zum Zwecke der Prozesskontrolle und zur Fehleranalyse (FA) vorhanden ist. Ein Überblick über die verwendeten Abbildungsverfahren in der FA kann in [?]

gefunden werden. Typischerweise werden durch einen fokussierten Ionenstrahl (*focused ion beam* FIB) sehr dünne Blättchen aus der Probe geschnitten und mit einem Transmissionsmikroskop analysiert. Dabei wird der Schichtaufbau der Probe ersichtlich und es können genaue Ausmessungen der Strukturen vorgenommen werden. Schicht- und Oxiddicken, die die Performance der Bauelemente wesentlich bestimmen, liefern wichtige Kennwerte um eigene Schwächen und Optimierungspotential finden zu können. Desweiteren gestattet der Einsatz von Röntgendetektoren Rückschlüsse auf Materialeigenschaften und die Ermittlung von Statistiken über den atomaren Aufbau des untersuchten Bereiches.

Schaltungsextraktion gewinnt mit steigender Komplexität der Technologien zunehmend an Bedeutung. In früheren Zeiten wurden Mikrochips mit gerade einmal einer Metallisierungsebene und mit einer Strukturbreite von $1\ \mu\text{m}$ und darüber gefertigt. Dadurch konnte eine Begutachtung durch ein optisches Mikroskop schon zu dem gewünschten Ziel führen. Heutzutage bestehen Mikrochips aus teilweise 10 Metalllagen und mehr, sowie Strukturbreiten von unter $45\ \text{nm}$. Dies schafft neue Herausforderungen, da der Zugriff auf die verborgenen Lagen nur durch raffinierte Probenpräparation und Abtragung der Ebenen möglich ist. Die geringe Strukturbreite erfordert desweiteren den Einsatz von Elektronenmikroskopen. Nichtinvasive Ansätze über Abbildungen durch Röntgenstrahlung scheitern an mangelnder geometrischer Auflösung, die im Bereich von $10\ \mu\text{m}$ liegt. Es gibt jedoch Bestrebungen von diversen Forschungsteams das Auflösungsvermögen von Elektronenmikroskopen mit der Durchdringungsleistung von Röntgenstrahlung zu kombinieren [?]. Durch diese neuen Randbedingungen ist es möglich, Produkte mit immer steigender Komplexität zu entwerfen und dem Mooreschen Gesetz gerecht zu werden [?]. Aber nicht nur die Komplexität hat im Laufe der Zeit zugenommen sondern auch die Vielfalt der verwendeten Materialien [?]. Eine Schaltungsextraktion und die dazu notwendigen Präparationen bestehen aus folgenden Phasen:

- Inspektion der Probe durch Röntgenstrahlung um Information über Position und Lage des Die und der Bonddrähte zu bekommen. Manche Chipgehäuse können auch zwei oder mehrere Dies enthalten.
- Entfernen des Gehäuses (*package*) mittels Schliff und Ätzung. Dabei wird das Die freigelegt und kann falls erforderlich komplett aus dem Gehäuse entfernt werden.
- Nun besteht die Möglichkeit durch ein optisches Mikroskop einen *ersten Blick* auf das Die zu werfen. Falls die oberste Metallisierungsebene nicht zum Sichtschutz flächig ausgelegt wurde, kann der Chip grob in Partitionen eingeteilt werden. Die nachfolgenden Analysen sind oft nicht für alle Bereiche notwendig und können schon in dieser Phase geplant werden.
- *Abtragen* der Schichten (*delayering*) durch Nass- und Trockenätzung mit anschließender Digitalisierung durch einen Chipscanner. Dieser ermöglicht durch eine Kombination aus Rasterelektronenmikroskop (REM) und motorisierter Probenhalterung das Die mäanderförmig auf eine Bilderserie abzubilden.

- Nun kann mit der eigentlichen *Analyse* und mit der Rückgewinnung der Schaltung begonnen werden. Grundlegend kann zwischen manueller, rechnerunterstützter, semiautomatischer und automatischer Analyse unterschieden werden. Eine zunehmende Automatisierung durch bildanalytische Methoden setzt eine entsprechende Qualität des Bildmaterials voraus, die jedoch nur mit optimaler Probenpräparation erreichbar ist. Dies ist in der Praxis, bedingt durch den komplexen Schichtaufbau und daher notwendige selektive Ätzung, nur schwer zu erreichen.

1.3 Aufbau der Arbeit

Einleitend wird in Kapitel 2 eine kurze Einführung in den CMOS Prozessablauf gegeben und erläutert, wie Mikrostrukturen durch sukzessives Abtragen der einzelnen Schichten freigelegt werden können.

Um eine qualitative Interpretation des Bildmaterials zu ermöglichen, wird in Kapitel 3 ein kurzer Überblick über die verwendeten Abbildungsverfahren und ein Einblick in die Methodologie der Technologieanalyse gegeben.

Kapitel 4 gibt eine Übersicht über die Kernthemen der Arbeit. Dabei wird die Aufgabenstellung in logisch abgegrenzte Module eingeteilt und bestehende Datenabhängigkeiten dargestellt.

In Kapitel 5 wird die Struktur der zugrundeliegenden Bildtransformationen untersucht und eine Klassifizierung des Bildinhaltes vorgenommen. Im Anschluss werden Methoden zur Mosaikerstellung (*stitching*) evaluiert und die implementierten Algorithmen beschrieben. Das Hauptaugenmerk liegt hierbei darauf, Ergebnisse zu erhalten, die den hohen Anforderungen an die geometrische Genauigkeit gerecht werden. Dazu werden *ground truth* generiert, mit denen die Qualität der verwendeten Algorithmen objektiv erfasst werden kann. Abgeschlossen wird das Kapitel mit einem Überblick über radiometrische Anpassungen des Bildinhaltes, um einen nahtlosen Übergang von Bild zu Bild zu gewährleisten.

Kapitel 6 befasst sich mit der Segmentierung von Metallisierungsebenen, mit dem Ziel einer polygonalen Beschreibung. Dabei werden wiederum verschiedene Ansätze verfolgt, die Vor- und Nachteile, abhängig vom zugrundeliegenden Bildmaterial, haben. Mit dem Ziel, aus der Beschreibung eine Netzliste generieren zu können, folgt ein Abschnitt über *routing*, welches Verbindungen innerhalb einer Lage sowie zwischen den Lagen berücksichtigt.

Die Analyse von Standardzellenbereichen und die daraus ableitbaren Schlussfolgerungen werden in Kapitel 7 behandelt. Um die aufwendigen Berechnungen in wirtschaftlicher Zeit durchzuführen wird eine *client-server* Lösung mit *multithreading* vorgeschlagen.

Abschließend folgt eine Zusammenfassung sowie ein Ausblick auf noch offene Themen. In Anhang A befindet sich eine Anleitung zur entstandenen Analysesoftware "Chipscan Analysis Tool".

Life is like an onion: you peel off layer after layer and then you find there is nothing in it.

James Huneker

Kapitel 2

Schichtstruktur

Zusammenfassung — In diesem Kapitel wird der Schichtaufbau von integrierten Schaltungen in CMOS Technologie beschrieben. Dabei erfolgt die Definition der einzelnen Schichten, die in der Analyse Anwendung finden. Um das Bildmaterial der Schichtabbildungen in Hinblick auf die Schaltungsanalyse interpretieren zu können, bedarf es eines Grundwissens über die physikalische Realisierung. Dazu werden diverse REM Abbildungen kommentiert um einen Bezug zwischen Physik und Bildmaterial herzustellen. Desweiteren werden Methoden des Schichtabtrags beschrieben um die dabei auftretenden Probleme besser verstehen zu können. Mit einem grundlegenden Verständnis der Verfahren, die vom unbekanntem Chip bis hin zu den Aufnahmen führen, ist es möglich, gezielte Wünsche an den Präparator zu stellen und damit passende Bildeigenschaften für eine robuste Bildanalyse zu erreichen.

Inhaltsverzeichnis

2.1	Einleitung	6
2.2	Schichtaufbau	7
2.3	Schichtabtrag	9

2.1 Einleitung

Seit der Erfindung der ersten integrierten Schaltung (*integrated circuit* (IC)) durch Jack Kilby und Robert Noyce unterliegt das Gebiet der Halbleitertechnologie einer rasanten Entwicklung. Neben fortlaufenden Verbesserungen der Defektdichte leiten sich heute eine Vielzahl an Spezialisierungen in Form von Epitaxie-Schichten, *silicon on isolater* (SOI) und verspannte *Si* Schichten von den ursprünglichen Verfahren ab. Alle Entwicklungen haben zum Ziel, die Leistungsanforderungen und Kostenanforderungen der Halbleiterindustrie zu befriedigen. Die Schritte, die von einer Idee bis hin zu einem

Produkt führen, umfassen interdisziplinäre Forschung und Entwicklung auf höchstem Niveau. Dabei musste und wird man auch in Zukunft immer wieder Schranken überwinden müssen, die in der Vergangenheit für unüberwindbar galten. Dabei zeigte die Vergangenheit exponentielles Wachstum in der Miniaturisierung integrierter Strukturen als auch bei der Bauelementendichte. Diese Entwicklung führte zu faszinierenden Möglichkeiten, komplexe Abläufe auf engstem Raum kostengünstig zu realisieren und ebnete damit den Einzug der Mikroelektronik in viele Bereiche des alltäglichen Lebens. Die ersten integrierten Schaltkreise bestanden aus wenigen Dutzend Transistoren und einer geringen Anzahl an lateralen Schichten. Durch ausgereifere Prozessmethoden und Planarisierungstechniken können heute Milliarden Bauelemente in einem vielschichtigen Aufbau integriert werden. Dem Aufbau der Schichtstrukturen soll sich dieses Kapitel widmen.

2.2 Schichtaufbau

Silicium hat die einzigartige Eigenschaft, dass in einer oxidierenden Atmosphäre eine dünne, stark gebundene Schicht von amorphem Silicium abgeschieden wird. Diese kann zum Schutz vor Verunreinigungen und noch wichtiger, zur Ausbildung von Dielektrika genutzt werden. Mit diesem Werkstoff wurde der Bau planarer Halbleiter möglich, d.h. Halbleiterstrukturen, die schichtweise aufgebaut sind. Ein entscheidender Vorteil ist, dass die Oxidschicht selektiv geätzt werden kann, was gezielte Verbindungen zu den untereinanderliegenden Schichten ermöglicht. Außerdem bleibt die Kristallfläche eben, was viele störende Oberflächeneffekte eliminiert. Mit der Planartechnik wurde es möglich, durch fotolithographische Verfahren und Diffusion aus der Gasphase viele baugleiche Transistoren auf einer einzelnen Siliciumscheibe aufzubauen.

Heutzutage ist es möglich, viele Milliarden Transistoren und andere Bauelemente wie Widerstände, Kondensatoren, Induktivitäten oder auch Mikro-Elektro-Mechanische Systeme (MEMS) auf engstem Raum zu integrieren. Die Herstellung einer integrierten Schaltung beruht im Grunde auf einer ausgeklügelten Abfolge von Dotierung, Schichtabscheidungen für leitende und isolierende Schichten sowie deren laterale Strukturierung durch Lithographie und Ätzung. Mit der Entwicklung der Halbleiterindustrie ist eine hohe Entwicklungsleistung zum Erreichen von Homogenität, Selektivität, Reinheit und Prozesskompatibilität der verwendeten Stoffe verknüpft. Desweiteren erfordert jede Prozesstechnologie spezifisches, hochgenaues Prozessequipment. Es gelingt dadurch, Schichten aufzutragen und zu strukturieren, deren Abmessungen nur wenige Atomlagen umfassen. Wesentliche Herausforderungen liegen dabei in der Homogenität, der Reproduzierbarkeit und der Defektfreiheit.

Die folgenden Beschreibungen des Schichtaufbaues beschränken sich auf die CMOS Technologie, welche vorherrschend Verwendung findet. Weiters werden in dieser Arbeit die Begriffe Schicht, Ebene und Lage austauschbar verwendet. Die in Klammer angeführten Kurzbezeichnungen der Lagen werden für eine spätere Referenzierung

herangezogen.

Wie schon oben erwähnt ist ein Mikrochip mehrschichtig aufgebaut. Die aktiven Gebiete (*active area* - **AA**) eines Mikrochips befinden sich auf unterster Ebene und bilden die *source/drain* Gebiete der primären Bauelemente - den Transistoren - ab. Darüber befindet sich, durch eine dünne Oxidschicht (Gateoxid) getrennt, eine Schicht aus Poly-Silicium (**Poly1**), welche zur Steuerung des sich darunter befindlichen Kanals dient. Unter Zuhilfenahme einer weiteren Poly-Silicium Schicht (**Poly2**) können persistente Speicherzellen wie EEPROM und Flash oder Kondensatoren realisiert werden. Die Verbindung der Transistoren mit den Metallisierungsebenen erfolgt durch die Kontaktebene (*gate contact* - **GC**), welche weitgehend aus Wolfram gefertigt wird.

Die Metallschichten (**M1...Mn**) bestehen meist aus Aluminium, da dieses Metall die nötigen Voraussetzungen für gute Integration mit sich bringt. Dazu zählen gute Haftung auf Siliciumdioxid und den Zwischenoxiden, ein hoher Leitwert, gute Kontaktierbarkeit und eine leichte Strukturierung durch Trockenätzverfahren. Da bei direktem Kontakt und ab einer Temperatur von etwa 200°C Silicium in Aluminium diffundiert, ist es notwendig die Stoffe durch eine Barriere zu trennen. Dies kann z.B. durch Titan oder Titanitrid erfolgen. Um den Kontaktwiderstand klein zu halten, wird zusätzlich zwischen Barriere und Silicium eine Schicht Titansilicid eingebracht. Da Aluminium bei kleinen Strukturgrößen zunehmend schlechter werdende Eigenschaften aufweist, wird entweder eine Aluminium-Kupfer Legierung oder überhaupt reines Kupfer verwendet. Auch in Punkto Elektromigration schneidet Kupfer wesentlich besser ab. Zur Kontaktierung zwischen den Ebenen wird bei Aluminium üblicherweise Wolfram eingesetzt, bei Kupfer kann der Stoff selbst dazu verwendet werden.

Zwischen den einzelnen Prozessschritten wird der Chip durch Planarisierungsverfahren planarisiert, um optimale Bedingungen für die Lithographie zu gewährleisten. Die oberste Schicht dient hierbei der Passivierung, also dem Schutz vor Korrosion und Verunreinigungen und wird meist nicht mehr planarisiert. Zwischen den funktionalen Schichten werden Isolationsschichten (*Inter Layer Dielectrics* (ILD)) eingebracht, die je nach Lage und Prozesstechnologie aus Silicate-Glass-Mischungen wie SiO_2 , Phosphosilicate glass (PSG) oder Borophosphosilicate glass (BPSG) bestehen.

Um einen grundlegenden Überblick über die in einem Mikrochip vorkommenden Schichten und den Aufbau dieser zu erhalten, wird in Tabelle 2.1 ein sehr vereinfachter CMOS Prozessablauf skizziert. Ausgangspunkt stellt hierbei ein monokristallines Siliciumsubstrat dar, welches je nach Prozesstechnologie großflächig schwach p bzw. schwach n dotiert wurde.

Ein IC besteht somit aus einer Vielzahl an Stoffen. Abbildungen 2.1 und 2.2 zeigen einen Schnitt durch einen Mikrochip. In Tabelle 2.2 sind verschiedene Schichten und die

Prozessschritt	Erläuterung
Substrat	Basis aus monokristallinem Silicium mit schwacher Dotierung.
Wanne	Herstellung der Grunddotierung für n-Kanal und p-Kanal Transistoren.
Isolation	Elektrische Isolation benachbarter Bereiche durch SiO_2 (STI, LOCOS).
Bauteile	Einstellung der Bauteil-Eigenschaften durch Kanaldotierung, Gateoxiddicke und Gatematerial.
Bauteilanschlüsse	Dotierprofil zum Anschluss des Kanalgebietes, niederohmige Verbindungen zu den Verdrahtungsebenen (meist aus Wolfram).
Lokale Verbindungen	Kurze Verdrahtungslängen innerhalb einer Schaltung.
Verdrahtung	Allgemeine Verdrahtung, je nach Komplexität der Schaltung. Derzeit bis zu 10 Ebenen.
Globale Verdrahtung	Niederohmige Verdrahtung mit relaxierten Anforderungen an die Strukturgröße. Für lange Verbindungen und Leistungsversorgung der Schaltungsblöcke.
Passivierung	Kapselung der Schaltung gegenüber Umwelteinflüssen und dadurch Sicherstellen der Langzeitfunktionalität.

Tabelle 2.1: Die CMOS Prozesstechnologie. Als Basis dient ein schwach dotiertes Substrat aus Silicium. Durch Prozessschritte wie Dotierung, Lithographie, Ätzung und Schichtabscheidung entstehen integrierte aktive und passive Bauelemente sowie die Verbindungen dieser. Damit sich die Bauelemente nicht gegenseitig beeinflussen oder damit keine parasitären Bauelemente entstehen, ist es notwendig, Isolationsschichten einzubringen. Transistoren, die in den untersten (aktiven) Schichten aufgebaut werden, werden durch Einbringen des Feldoxides getrennt, um Leckströme unter benachbarten Transistoren zu vermeiden. Bei neueren Prozesstechnologien werden die aktiven Bauelemente durch eine Grabenisolation getrennt.

dafür typischen Stoffe aufgelistet.

2.3 Schichtabtrag

Um zu verstehen, wie die Schichten eines unbekanntes ICs aufgebaut sind, ist es notwendig die Schichtstruktur zu analysieren. Dies kann durch Inspektion eines vertikalen Schnittes durch den IC wie in Abschnitt 3.2 beschrieben erfolgen. Mit den Analyseergebnissen werden die Methoden und Parameter für den Schichtabtrag bestimmt. Ist es wichtig die verwendeten Materialien genau zu kennen, kann durch Analysemethoden der Röntgenspektroskopie (EDX) eine Elementarteile-Statistik erzeugt werden. Sind nun die verwendeten Materialien bekannt, können durch Ätzverfahren die Schichten sukzessive abgetragen werden. Dazu bedient man sich je nach Schicht und Präparationsziel unterschiedlicher Verfahren. Zur Charakterisierung der Ätzverfahren werden folgende Größen definiert:

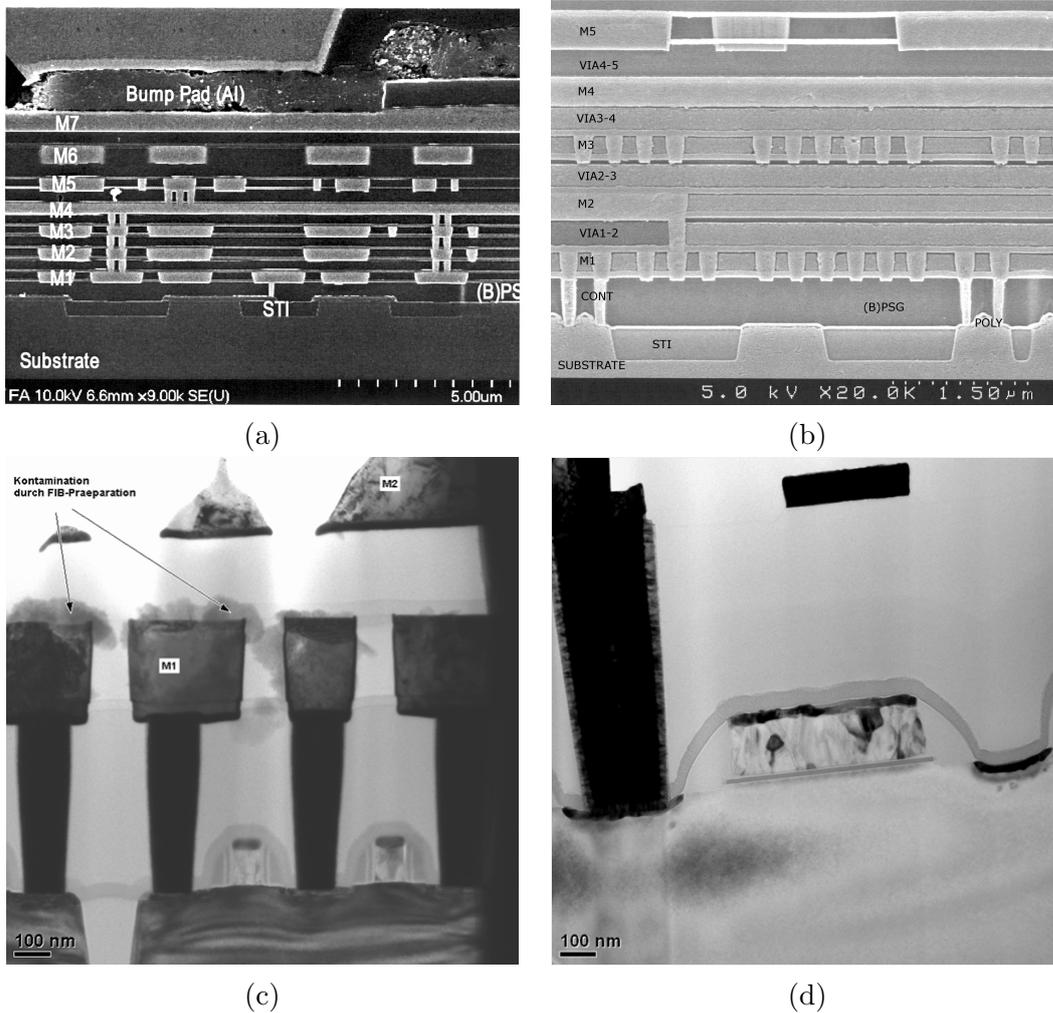


Abbildung 2.1: Typischer Schichtaufbau eines IC. Durch einen gut gewählten vertikalen Schnitt durch den Chip kann der Schichtaufbau sichtbar gemacht werden ((a) und (b)). Die oberste Lage (Passivierung) dient zum Schutz vor Korrosion und Verschmutzung. Die darunterliegenden Metalllagen stellen die Verbindungen der Bauelemente der untersten Schichten her. Zwischen den Metalllagen befinden sich Oxide zur Isolation, sowie VIAs, die die Metalllagen elektrisch leitend verbinden. Zwischen der untersten Metallschicht und dem aktiven Gebiet dient einerseits eine dicke Schicht Oxid zur Isolation, andererseits stellen Kontakte die Verbindung mit den Transistoren her. Eine Grabenisolation (*shallow trench isolation (STI)*) zwischen den Transistoren soll einen unerwünschten Stromfluss zwischen benachbarten einzelnen Transistoren verhindern. (Ältere Prozesstechnologien verwendeten hierzu ein Feldoxid, welches durch das LOCOS Verfahren eingebracht wurde.) Um höhere Auflösungen zu erreichen und die Transistoren einzeln untersuchen zu können, werden dünne Blättchen mit einem FIB präpariert und durch eine TEM analysiert. Die Bilder (c) und (d) zeigen eine so durchleuchtete Probe. Die Transistoren befinden sich zwischen den Kontakten, die M1 mit dem Substrat verbinden. In Bild (d) wurde ein Transistor vergrößert dargestellt. Das Gate aus Poly-Silicium befindet sich mittig. Zwischen Gate und Substrat ist das Gateoxid sichtbar, welches eine Höhe von nur wenigen Nanometern hat.

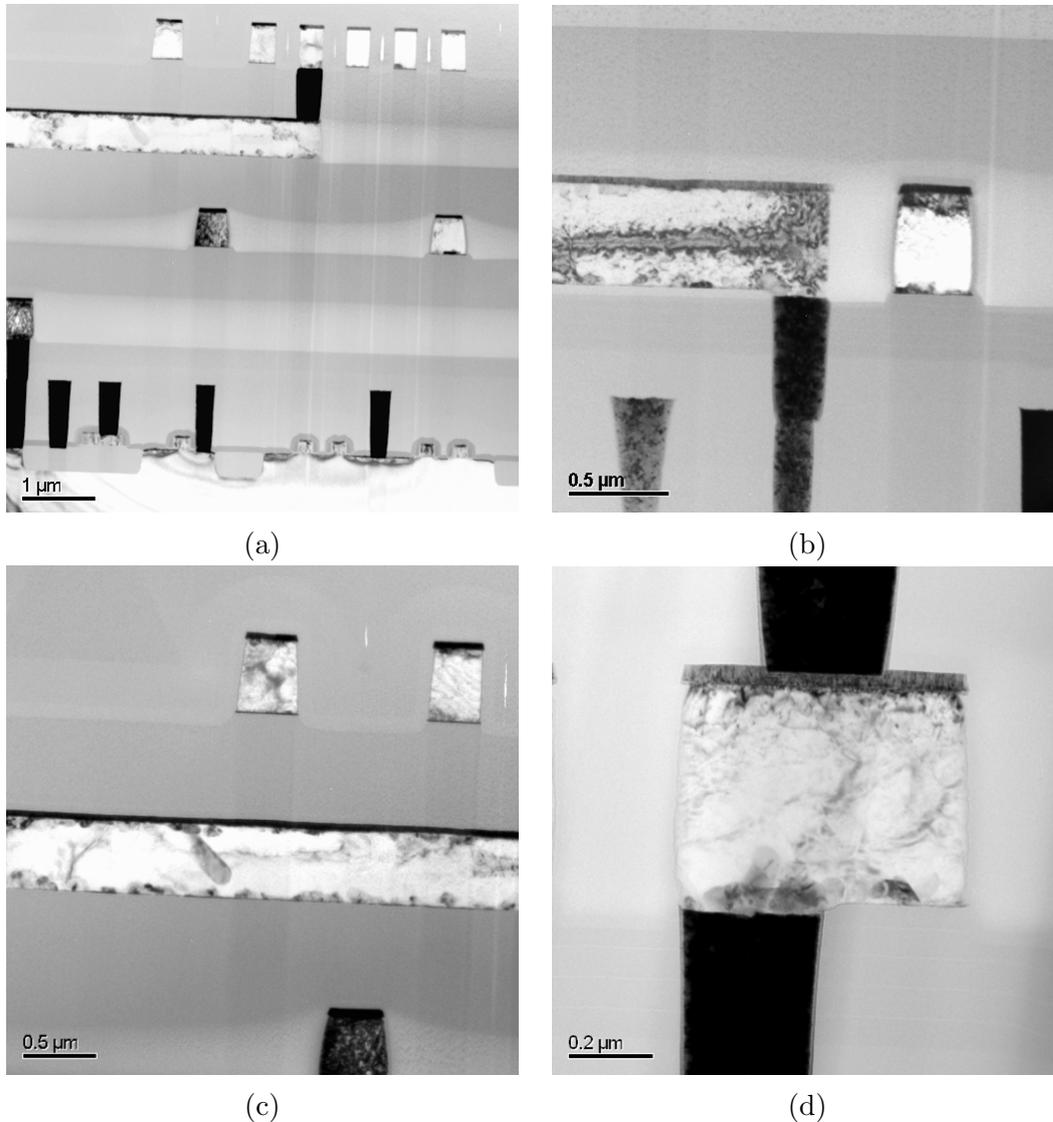


Abbildung 2.2: Wie auch Abbildung 2.1 zeigen diese TEM Aufnahmen den Schichtaufbau einer integrierten Schaltung. Hierbei liegt der Fokus jedoch auf den Leiterbahnen und den verbindenden Kontakten. Die dunklen Bereiche zeigen Kontakte aus Wolfram, dessen Ordnungszahl von 74 im Vergleich zu Aluminium mit 13 wesentlich höher ist und damit in einem TEM dunkler abgebildet wird. In Bild (d) ist eine Leiterbahn mit 2 Kontakten zu den benachbarten Ebenen zu sehen. Als Grenzschicht wurde eine Schicht Titannitrid (TiN) als Barriere aufgebracht, um die Diffusion von Wolfram in Aluminium zu unterbinden.

Schicht	Material
Substrat	Silicium (<i>Si</i>), Bor (<i>B</i>), Phosphor (<i>P</i>)
Gateoxid	Siliciumdioxid (SiO_2)
Gate	Polysilicium (<i>Poly-Si</i>)
Gate Kontakt (GC)	Wolfram (<i>W</i>)
Bariere	Titan (<i>Ti</i>), Titanitrit (<i>TiN</i>), Titansilicid (<i>TiSi</i>)
Metall	Aluminium (<i>Al</i>), Kupfer (<i>Cu</i>)
VIA	Wolfram (<i>W</i>), Aluminium (<i>Al</i>), Kupfer (<i>Cu</i>)
ILD	Silicatglassmischungen (PSG, BPSG, USG, FSG)
Passivierung	Siliciumnitrid (Si_3N_4), Siliciumdioxid (SiO_2)

Tabelle 2.2: In einer integrierten Schaltung kommen eine Vielzahl an Stoffen zum Einsatz. Diese Auflistung erhebt keinen Anspruch auf Vollständigkeit, sondern soll eine Gegenüberstellung der Schichten und der darin vorkommenden Stoffe wiedergeben. Für den Schichtabtrag, der in Abschnitt 2.3 beschrieben wird, ist es wichtig die Zusammensetzung der Schichten zu kennen, um ein geeignetes Verfahren wählen zu können.

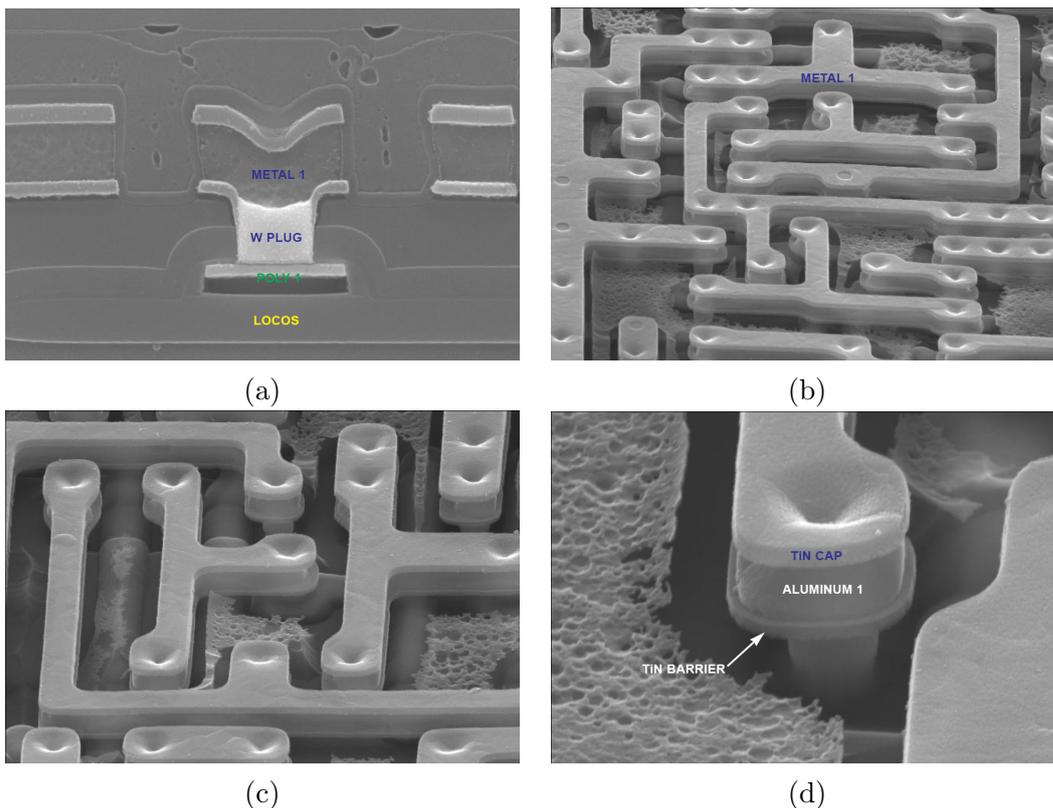


Abbildung 2.3: Durch Skelettätzen freigelegte Metallisierungsebene. Bild (a) zeigt einen Querschnitt durch eine ältere Prozesstechnologie mit LOCOS-Technik. Bild (b) bis (d) die durch Skelettätzen freigelegte erste Metallisierungsebene.

Methoden	Type	Anisotropie	Selektivität
Ionenstrahlätzen	physikalisch	anisotrop	gering
Sputtern	physikalisch	anisotrop	gering
Reaktives Ionenstrahlätzen	phys./chemisch	anisotrop, mit isotropem Anteil	steuerbar
Reaktives Ionenätzen	phys./chemisch	anisotrop, mit isotropem Anteil	gering
Plasmaätzen	phys./chemisch	isotrop, mit isotropem Anteil	hoch
Barrelätzen	chemisch	isotrop	hoch

Tabelle 2.3: Trockenätzverfahren können physikalischer, chemischer oder gemischter Natur sein. Bei physikalisch unterstützten Methoden werden Atome durch den Beschuss mit energiereichen Ionen aus der Probe herausgelöst. Die dabei in die Gasphase übergehende Materie wird in einer Vakuumkammer abgeführt. Um eine bessere Selektivität zu erreichen, kann das physikalische Verfahren durch ein Ätzgas chemisch unterstützt werden. Tabelle: [?]

- Die *Ätzrate* gibt den Materialabtrag pro Zeit an und damit die Abtragsgeschwindigkeit des zu ätzenden Stoffes. Die Angabe erfolgt in $\mu\text{m}/\text{min}$.
- Die *Selektivität* entspricht dem Quotienten der Ätzrate des zu ätzenden Stoffes zur Ätzrate eines anderen Stoffes. Beim strukturerhaltenden Abtrag, wie z.B. dem Entfernen von nichtleitenden Schichten (Skelettätzen), ist hohe Selektivität zwischen Siliciumdioxid und Aluminium/Kupfer/Titan erforderlich (Abbildung 2.4 zeigt eine freigelegte Metallschicht mit den darauf befindlichen Kontakten).
- Die *Anisotropie* beschreibt die Richtungsabhängigkeit des Ätzverfahrens und wird durch 1 minus dem Verhältnis der horizontalen zur vertikalen Ätzrate bestimmt. Verfahren mit einer Anisotropie von 1 werden als *anisotrope* bezeichnet. Verfahren mit einer Anisotropie gegen 0 als *isotrope*.

Es existieren eine Reihe von Ätzverfahren, die sich methodisch unterscheiden. Je nach gewünschter Information und den zu entfernenden Stoffen wird ein Verfahren gewählt, das den Anforderungen gerecht wird. Mit den oben definierten Größen können die Verfahren klassifiziert werden.

- *Trockenätzen*. Dies kann rein physikalisch, chemisch oder physikalisch -chemisch erfolgen. Beim rein physikalischen Ätzen wird Material durch gerichteten Beschuss mit Ionen anisotrop abgetragen (*sputter*, *ionmilling*). Die chemisch unterstützte Methode erzielt, durch Einbringen eines Ätzgases (Plasma) in die Vakuumkammer, hohe

Selektivität bei isotroper Ätzwirkung. Durch die Kombination der Ionenstrahlmethode mit einem sich in einer separaten Kammer befindlichen Plasmas kann eine anisotrope selektive Ätzwirkung erreicht werden. Dabei werden im Vergleich zur rein physikalischen Methode geringere Energien gewählt um Sputtereffekte zu vermeiden. *Reaktives Ionenätzen* (*reactive ion etching* (RIE)) zählt zu einer speziellen Form des Trockenätzens, wobei die Probe kapazitiv an eine RF-Quelle gekoppelt ist und sich negativ auflädt. Eine Bias-Spannung sorgt dabei für den gerichteten Beschuss. Dabei werden Ätzteilchen (Ionen), die in einem Gasplasma erzeugt wurden auf die Probe beschleunigt und lösen darauf eine chemische Reaktion aus. Das Ätzprodukt ist flüchtig und kann mit Hilfe eines Vakuums abgeführt werden. Die Ätzwirkung ist anisotrop und selektiv. Selektivität, Anisotropie und Ätzrate können durch Prozessparameter eingestellt werden. Eine Gegenüberstellung einiger Trockenätzmethoden und deren Eigenschaften sind in Tabelle 2.3 angeführt.

- *Nasschemisches Ätzen*. Diese Methode beruht auf der Auflösung von Substanzen und die Überführung in eine wässrige Lösung. Durch die richtige Wahl der Ätzlösung kann hohe Selektivität erreicht werden. Jedoch ist der Prozess schlecht kontrollierbar und es werden oft darunterliegende Schichten angegriffen. Das Ätzprofil ist typischerweise isotrop, da aufgrund der geringen freien Weglänge eine gerichtete Wirkung nur schwer erzielt werden kann. Jedoch kann durch Ausnutzung der von der Kristallstruktur abhängigen Ätzrate, Anisotropie erreicht werden. So lassen sich Kristallorientierungen mit Miller-Index 100 und 110 (z.B. senkrechte Wände) schneller ätzen wie 111 Orientierungen. Wichtige Parameter sind neben der Zusammensetzung der Ätzlösung, die Temperatur der Ätzlösung sowie die Ätzdauer. Ein anschließendes Spülen mit Wasser und Aceton stoppt die Ätzwirkung.
- *Mechanisches polieren*. Die Oberfläche der Probe wird hierbei durch ein feinkörniges Poliermittel und evtl. Zugabe von reaktiven Zusätzen (chemisch-mechanisch) schleifend abgetragen. Dadurch kann die Abtragtiefe sehr gut kontrolliert werden. Die Methode ist jedoch nicht selektiv, da im Prinzip nur die obersten Schichten abgetragen werden. Ecken werden durch die Schleifbewegung abgerundet. Ein weiteres Problem stellt das plane Einspannen von kleinen Proben dar.

Alle diese Verfahren finden auch beim Schichtaufbau Verwendung und sind daher gut bekannt.

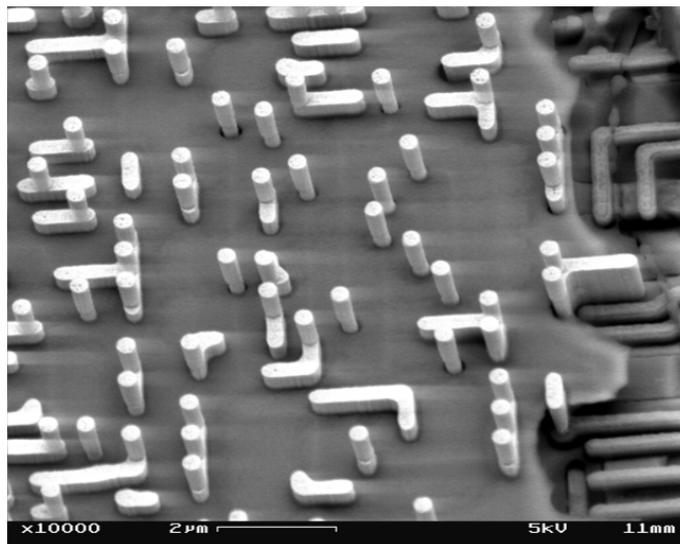


Abbildung 2.4: Freilegung von Metallschichten. Die Freilegung der Metallschichten vom Siliciumdioxid erfolgt mit Ätzverfahren hoher Selektivität. Auf dieser REM Aufnahme aus dem Logikbereich einer unteren Metallisierungsebene wurde das Oxid entfernt und damit die Metallschicht inklusive der sich darauf befindlichen Kontakte freigelegt. Ein großflächiges gleichmäßiges Entfernen, ohne das Oxid der darunterliegenden Schichten anzugreifen, ist jedoch nur schwer erreichbar. Um dieses Problem zu umgehen, wird das Metall nicht vollständig vom Oxid befreit. Durch Erhöhung der Beschleunigungsspannung des REM kann das Metall auch unter dem Oxid kontrastreich abgebildet werden. Ein weiterer Vorteil besteht darin, dass bei dieser Methode Kontakte nach der Abbildung in einem REM heller erscheinen als Leiterbahnen, was der Extraktion der Kontakte durch Methoden der Bildanalyse zugute kommt.

Measure what is measurable, and make measurable what is not so.

Galileo Galilei

Kapitel 3

Abbildungsverfahren

Zusammenfassung — In diesem Kapitel werden typische Abbildungsverfahren, die bei der Analyse von Nanostrukturen integrierter Schaltungen Anwendung finden, vorgestellt. Nach einer kurzen allgemeinen Darstellung und Motivation wird auf den Ablauf der Prozessanalyse eingegangen, der zur Technologiebestimmung und damit als Grundlage für darauffolgende Analyseschritte dient. Da Abbildungen aus der Elektronenmikroskopie die Datengrundlage der durch Bildanalyse gestützten Schaltungsanalyse bilden, wird darauf näher eingegangen und das Entstehen der Aufnahmen beschrieben.

Inhaltsverzeichnis

3.1 Motivation	16
3.2 Abbildungsverfahren der Prozessanalyse	17
3.3 Abbildungsverfahren der Schaltungsanalyse	17

3.1 Motivation

Die Fähigkeit feine Strukturen zu unterscheiden ist für das menschliche Auge sehr begrenzt. Diese als Auflösungsvermögen bezeichnete Kenngröße wird als Winkelabstand zweier gerade noch getrennt wahrzunehmenden punktförmigen Strahlungsquellen definiert. Der Mensch versuchte schon früh, dieser naturgegebenen Begrenzung durch technische Hilfsmittel zu begegnen. Daraus entsprangen eine Reihe von Erfindungen, die einerseits versuchten, weit entfernte Objekte näher erscheinen zu lassen und andererseits, den Mikrokosmos für das menschliche Auge zugänglich zu machen und damit einen Blick in die Struktur der Materie zu ermöglichen. Der Grundgedanke - Vergrößerung - ist hierbei derselbe, nur unterscheiden sich die jeweiligen Realisierungen. Die Mikroebene der Natur bietet eine Fülle an fantastischen Strukturen, die ohne moderne technische Hilfsmittel nicht erschlossen werden könnten.

Die Lichtmikroskopie nutzt von der Probe reflektierte elektromagnetische Strahlung, die in einem lichtsensitiven Sensor akkumuliert und anschließend elektronisch ausgewertet wird. Aufgrund der Wellennatur des Lichtes, das im sichtbaren Bereich Wellenlängen zwischen 380 nm und 780 nm besitzt, ist das Auflösungsvermögen optischer Mikroskope beschränkt. Um tiefer in die Struktur der Materie blicken zu können, muss die Frequenz des Lichtes erhöht werden. Da optische Linsen bei höheren Auflösungen auch höheren Anforderungen an die Fertigungsgenauigkeit gerecht werden müssen, sind der optischen Mikroskopie technische Grenzen gesetzt. Alternativ können auch andere Elementarteilchen für Untersuchungen verwendet werden. Eine Möglichkeit stellen Elektronen dar, die im Gegensatz zu Photonen den Vorteil haben, eine elektrische Ladung zu besitzen und damit durch eine elektrische Spannung gezielt steuerbar sind. Eine Realisierung dieser alternativen Variante stellt das Rasterelektronenmikroskop (REM) und das Transmissionselektronenmikroskop (TEM) dar. Bei beiden Mikroskopen werden Elektronen auf eine Probe beschleunigt und die durch das Auftreffen verursachten Reaktionen untersucht.

3.2 Abbildungsverfahren der Prozessanalyse

Um die in Kapitel 2 angeführten Ebenen Analysezwecken zugänglich zu machen, erfordert es Mikroskope, deren Auflösungsvermögen wesentlich höher ist, als die kleinste vorkommende Strukturbreite. Da es sich hierbei um Größen im Nanometerbereich handelt, scheiden optische Verfahren aus. Um Informationen über die verwendete Prozesstechnologie zu gewinnen, wie etwa den horizontalen Schichtaufbau und die am Wafer abgeschiedenen Materialien sowie deren genauen Dicken, ist es notwendig, Schnitte durch die Probe zu legen. Dazu werden durch einen fokussierten Ionenstrahl (*focused ion beam* (FIB)) möglichst dünne Blättchen aus der zu untersuchenden Probe geschnitten und anschließend mit einem TEM durchleuchtet. Ein FIB besteht aus einer Ionenquelle, die zum atomaren Abtragen und zum Auftragen von Materie aus der Gasphase dient, sowie einem REM zur Prozesskontrolle. In Abbildung 3.1 ist der Präparationsvorgang einer TEM-Lamelle im Zeitraffer dargestellt.

3.3 Abbildungsverfahren der Schaltungsanalyse

Nach einer grundlegenden Material- und Schichtanalyse liegen nun Daten über Schichtdicken und derer atomarer Zusammensetzung vor. Mit diesem Wissen können durch selektives Abtragen einzelne Lagen freigelegt und anschließend durch ein hochauflösendes Mikroskop abgebildet werden. Abhängig von den kleinsten vorkommenden Strukturen wird zwischen optischen und nicht optischen Verfahren entschieden.

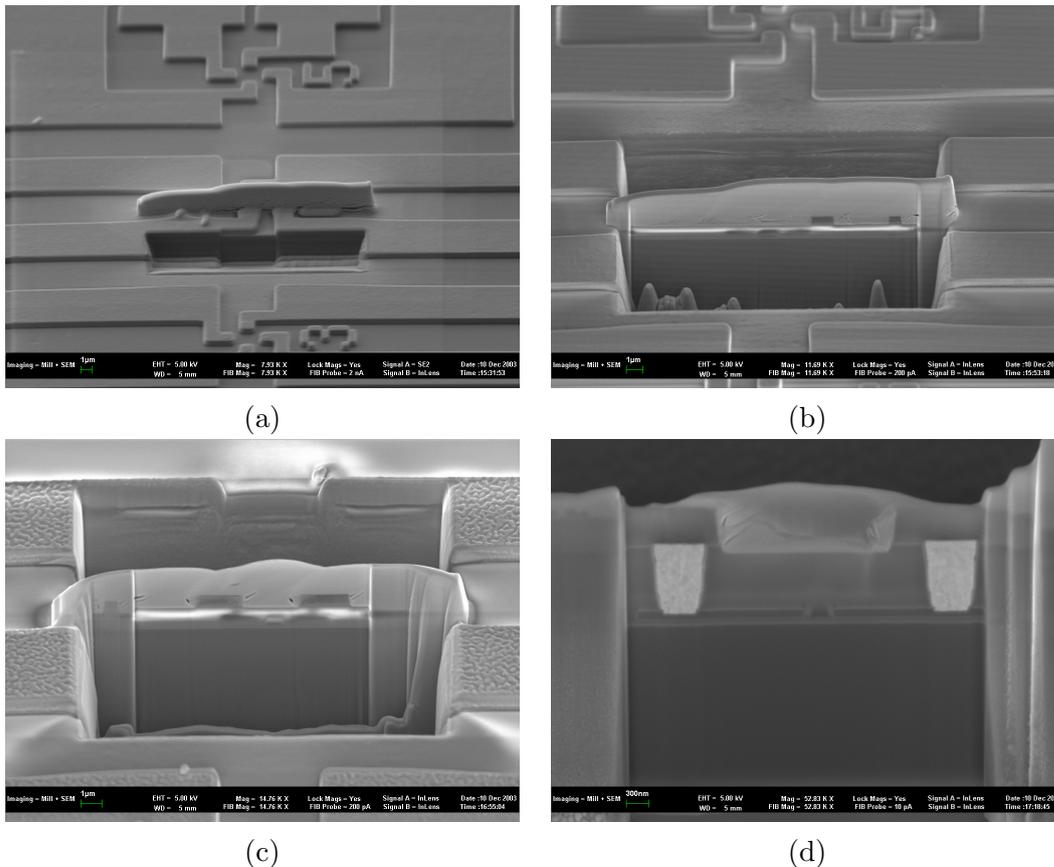


Abbildung 3.1: Präparation einer TEM-Lamelle durch eine FIB. Durch einen fokussierten Ionenstrahl wird eine Probenstelle beidseitig freigelegt und die so erzeugte Lamelle auf einen Nadelmanipulator befestigt. Damit ist es möglich, sehr dünne Probenblättchen von wenigen Nanometern zu erzeugen, die mit einem TEM analysiert werden können. In der Technologieprozessanalyse findet diese Methode zur Bestimmung des Schichtaufbaues Verwendung. Abbildungen (a)-(c) zeigen verschiedene Stadien dieses Präparationsprozesses. In Abbildung (d) ist das Blättchen in Nahaufnahme zu sehen. Im unteren Bereich befindet sich das Siliciumsubstrat. Die hellen Bereiche stellen Kontakte aus Wolfram dar, die Poly-Silicium und Metal 1 verbinden. Zwischen den Kontakten befindet sich der Speichertransistor, welcher von *sidewall-spacern* umgeben ist.

3.3.1 Lichtmikroskopie

Optische Verfahren sind durch den Wellencharakter des Lichtes in der Auflösung begrenzt. Nach der Abbeschen Auflösungsgrenze von

$$d_{min} = \frac{\lambda}{2n \sin \alpha} = \frac{\lambda}{2NA} \quad (3.1)$$

mit der Lichtwellenlänge λ , der Brechzahl n , dem halben Öffnungswinkel α des Objektivs und der sich daraus ergebende numerischen Apertur NA , hängt die Auflösungsgrenze direkt proportional von der Lichtwellenlänge ab und begrenzt die theoretische Auflösung

eines Lichtmikroskops bei Verwendung von violetterem Licht ($\lambda \approx 400 \text{ nm}$) auf etwa $0.2 \mu\text{m}$. In der Praxis gelten jedoch Größen unter $0.5 \mu\text{m}$ als nicht auflösbar. Die numerische Apertur ist eine Kenngröße für die Bündelung von Lichtstrahlen in optischen Systemen und beschreibt das Auflösungsvermögen eines Objektivs. Um robuste Bildanalyse betreiben zu können, muss die geometrische Auflösung deutlich höher sein als die kleinsten vorkommenden Strukturen. Daher ist diese Art der Abbildung für neuere Prozesstechnologien eher ungeeignet. Um jedoch einen ersten Überblick zu bekommen ist dies durchaus ausreichend.

Ein wichtiger Faktor bei der Lichtmikroskopie ist die Art und Weise der Beleuchtung der Probe. Bei der konventionellen Lichtmikroskopie wird die Probe entweder durch nicht-kohärentes Licht großflächig oder über eine Kondensorlinse fokussiert beleuchtet. Ein damit verbundener auflösungsbegrenzender Effekt ist, dass auch Licht aus nichtfokussierten Bereichen das Auge bzw. den Bildsensor erreicht. Dadurch wirkt die Abbildung bei erhöhter Vergrößerung zunehmend diffus.

Die konfokale Mikroskopie versucht diesen Nachteil zu überwinden, indem sie die Probe durch kohärentes Licht im Rasterverfahren punktuell zur Strahlung anregt, abbildet und im Anschluss die einzelnen Intensitätsmessungen zu einem Gesamtbild zusammenfügt. Um den oben beschriebenen Nachteil der Überlagerung von Licht aus nichtfokussierten Bereichen zu minimieren, wird vor dem Detektor zusätzlich eine Lochblende platziert. Eine Realisierung dieses Verfahrens stellt das *confocal laser scanning microscope* (CLSM) dar. In Abbildung 3.2 wird eine Gegenüberstellung der beiden beschriebenen Verfahren wiedergegeben. Weiter besteht die Möglichkeit durch Messung des Reflektionsfaktors Rückschlüsse über die Oberflächenbeschaffenheit zu ziehen.

Bei der Abbildung integrierter Schaltungen kann es durch die dünnen, transparenten Oxidschichten zu Interferenzen kommen. Der in der Physik unter Interferenzen gleicher Neigung bekannte wellenbedingte Effekt entsteht durch Überlagerung der Reflektionen an den verschiedenen Mediumsübergängen. In Abbildung 3.3 sind diese Interferenzerscheinungen exemplarisch dargestellt.

3.3.2 Rasterelektronenmikroskopie

Da neue Prozesstechnologien, die im Stande sind Strukturgrößen unter 40 nm abzubilden, höhere Anforderungen an das geometrische Auflösungsvermögen des Mikroskops stellen und das Auflösungsvermögen eines Lichtmikroskop aufgrund des Wellencharakters des Lichtes beschränkt ist, müssen alternative Verfahren verwendet werden, um den Anforderungen gerecht zu werden. Jedoch auch Materie besitzt Welleneigenschaften. Die Zuordnung von im Sinne der klassischen Physik typischen Teilcheneigenschaften, wie Lokalisierbarkeit, Energie, Impuls usw. zu Wellen legt aus Symmetriegründen die Idee nahe, umgekehrt den Materieteilchen auch Welleneigenschaften zuzuordnen. *De Broglie* postulierte 1924, dass zwischen dem Impuls $p = mv$ der Teilchen und der Wellenlänge λ der den Teilchen zugeordneten Materiewellen derselbe Zusammenhang

$$p = mv = \frac{h}{\lambda} \quad (3.2)$$

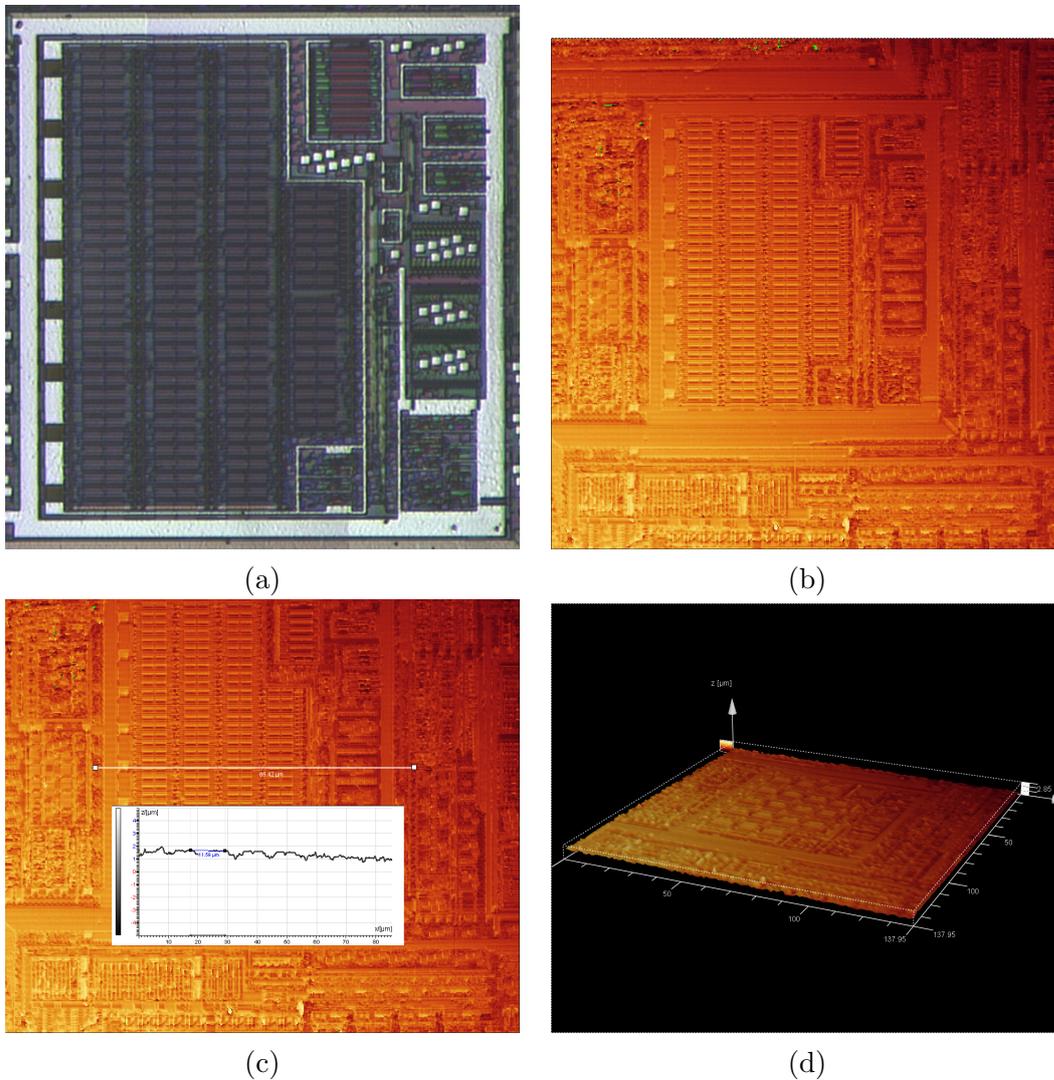


Abbildung 3.2: Gegenüberstellung der Auflichtmikroskopie und der Laser-Konfokalmikroskopie. Beide Aufnahmen wurden mit einem 100-fach Objektiv aufgenommen (entspricht einer Vergrößerung von 1000-fach). Abbildung (a) zeigt die Aufnahme eines Chipbereiches durch ein gewöhnliches Lichtmikroskop (PolyVar Met / Reichert-Jung). In Abbildung (b) wurde derselbe Chipbereich durch ein Konfokalmikroskop (Leica TCS SP2) aufgenommen. Durch die Reduzierung des Streulichtes wird auch in z -Richtung gute Auflösung erzielt. Abbildungen (c) und (d) zeigen die durch 3D-Rekonstruktion ermittelte Topologie des Chipausschnittes aus Abbildung (b).

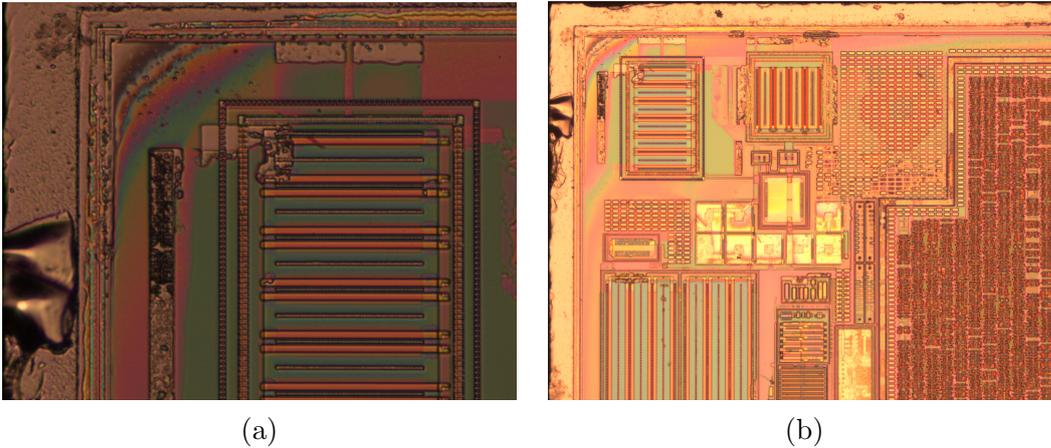


Abbildung 3.3: Interferenzerscheinungen an dünnen Schichten. Durch die dünnen Oxidschichten, die unter der Kohärenzlänge von weißem Licht liegen, interferieren die von den Medien reflektierten Lichtstrahlen. Dieser Effekt äußert sich bei der nachfolgenden Bildanalyse als störend.

wie bei Licht besteht. Solange die Geschwindigkeit der Teilchen im nichtrelativistischen Bereich bleibt, gilt für ein durch die Spannung U beschleunigtes Elektron mit der Ladung e und der Masse m_e :

$$v_e = \sqrt{\frac{2eU}{m_e}} \quad (3.3)$$

Für höhere Beschleunigungsspannungen muss der relativistische Energiesatz angewandt werden [?]. Nach Einsetzen von Gleichung 3.3 in Gleichung 3.2 erhält man

$$\lambda = \frac{h}{p} = \frac{h}{m_e v_e} = \frac{h}{\sqrt{2em_e U}}. \quad (3.4)$$

Für ein bewegtes Elektron mit einer kinetischen Energie von $E = 1 \text{ eV}$ gilt daher $\lambda = 12.2 \text{ \AA}$. Die Wellenlänge λ des Elektrons verhält sich indirekt proportional zu \sqrt{E} und nimmt damit bei höherer Beschleunigung ab. Für eine typische Beschleunigungsspannung von 10 keV besitzt ein Elektron eine *De Broglie*-Wellenlänge von 0.12 \AA und liegt damit unter dem Atomradius von Silicium mit 1.1 \AA . Damit sind es hier weniger Interferenz- und Beugungseffekte, die die Auflösung begrenzen sondern vielmehr die komplexen Interaktionen zwischen Elektronen und Materie.

Die Transmissionselektronenmikroskopie ermöglicht geometrische Auflösungen bis unter den Nanometerbereich und stellt damit ein gut geeignetes Instrument zur Untersuchung dar. Dieses erstmals 1931 von *Knoll* und *Ruska* gebaute Elektronenmikroskop, für welches *Ruska* erst im Jahre 1986 den Nobelpreis erhielt, verwendet im Gegensatz zur Lichtmikroskopie ein grundlegend anderes Prinzip. Im Mittelpunkt dieser Technik stehen Elektronen, die auf die Probe beschleunigt und Reaktionen in derselben hervorrufen. Um die aus diesem Verfahren hervorgehenden Intensitätsbilder qualitativ zu verstehen, soll im Folgenden näher auf das Verfahren eingegangen werden. Ein umfassender Überblick über dieses faszinierende Verfahren wird in [?] gegeben.

Das Prinzip eines Rasterelektronenmikroskops ist in Abbildung 3.4(a) schemenhaft dargestellt. Durch Aufheizen der Kathode (thermische Emission) im Vakuum oder durch Anlegen eines hohen elektrischen Feldes (Feldemission) wird die Kathode soweit angeregt, dass die Austrittsarbeit überschritten und neben Photonen auch Elektronen emittiert werden. Diese freien Elektronen können nun durch Anlegen einer positiven Spannung an der Anode gezielt beschleunigt und gesteuert werden. Durch die Beschleunigung nehmen Elektronen Energie auf, die wie in der Teilchenphysik üblich, in Elektronenvolt (eV) angegeben wird. Diese Energie, die typischerweise zwischen 1 keV und 50 keV liegt, ist einer der primären Parameter, welcher vom Bediener an die jeweilige Messaufgabe angepasst werden muss. Der Einsatz eines Hochvakuums ist hierbei wichtig, um die emittierten Elektronen auf dem Weg zur Probe so geringfügig wie möglich durch die sich in der Kammer befindlichen Moleküle zu stören.

Die für die Fokussierung des Elektronenstrahls notwendige Optik unterscheidet sich grundlegend von der, wie sie in der Lichtmikroskopie verwendet wird. Die Optik dient hierbei nicht der Vergrößerung sondern lediglich der Fokussierung des Teilchenstrahles. Wahlweise können magnetische Felder oder elektrische Felder eingesetzt werden. Dieses Prinzip findet auch in Kathodenstrahlröhren Anwendung, die früher in Fernsehern (magnetische Ablenkung) oder Oszilloskopen (elektrische Ablenkung) eingesetzt wurden. Wenn ein bewegtes Elektron in ein magnetisches Feld eintritt, wirkt darauf die Lorentzkraft $\vec{F} = e(\vec{v} \times \vec{B})$, die das Elektron auf eine zyklotronische Bahn zwingt. Dies ist auch der Grund, weshalb Abbildungen oft gedreht erscheinen. Bei Erhöhung der Beschleunigungsspannung und der damit verbundenen Zunahme an Energie der Elektronen muss die Linsenstärke angepasst werden, d.h. die Stärke des magnetischen Feldes erhöht werden. Da der auf die Probe auftreffende Strahl einen möglichst kleinen Durchmesser besitzen soll, werden wie in Abbildung 3.4 zu sehen ist, mehrere Linsen kaskadiert. Die wiederholte Fokussierung führt zu einer Vergrößerung des Winkels und zur Abnahme der Intensität. Im Vergleich zu thermisch emittierten Elektronen weisen durch Feldemission emittierte Elektronen durch die kleinere Austrittsfläche kleinere Strahldurchmesser auf. Nach der Fokussierung durch das Linsensystem treffen die (primär) Elektronen auf der Probe auf und regen wiederum (sekundär) Elektronen und Photonen zur Emission an (Abbildung 3.5). Die von der Probe emittierte elektromagnetische Strahlung und abgelenkte/herausgeschlagene Teilchen tragen Information aus der Wechselwirkung des Primärelektronenstrahls mit der Probe mit sich. Durch entsprechende Detektoren können diese ausgewertet und zur Analyse des untersuchten Bereichs herangezogen werden. Das emittierte Energiespektrum der detektierten Elektronen besteht aus Sekundärelektronen (*secondary electrons* SE), rückgestreuten Elektronen (*backscattered electrons* (BSE)) und Auger Elektronen (*auger electrons* AE).

Durch die Energieabgabe der Primärelektronen an die Elektronen in der Probe werden aus Atomen Elektronen herausgelöst. Stammt ein Elektron aus einer inneren Schale des Atoms, so wird dieses von einem höher energetischen Elektron nachbesetzt. Die Differenzenergie wird in der Regel durch Abgabe eines Photons mit einer der Differenzenergie entsprechenden Röntgenfrequenz $\nu = E/h$ (Plank-Einstein Gleichung) nachbesetzt.

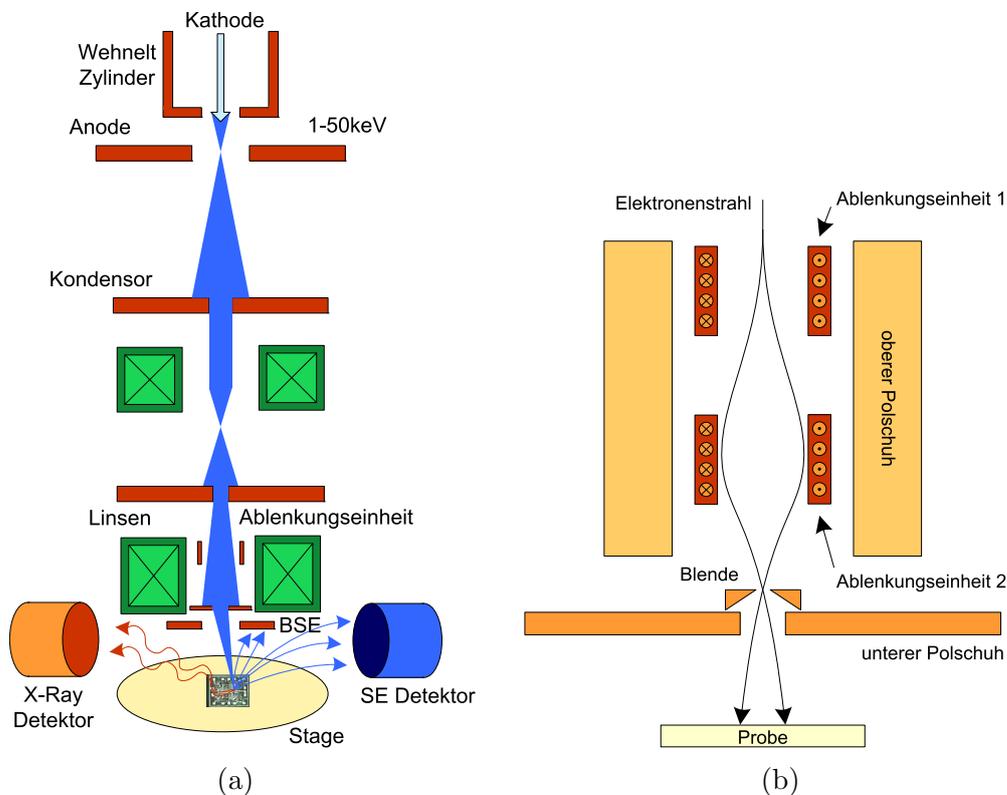


Abbildung 3.4: Prinzip eines Rasterelektronenmikroskops. Durch Beheizung oder Anlegen eines elektrischen Feldes treten an der Kathode Elektronen aus und werden in Richtung Anode beschleunigt. Die Intensität, also die Anzahl der austretenden Elektronen und die Fokussierung kann durch eine negative Spannung am Wehneltzylinder gesteuert werden. Die Potentialdifferenz zwischen Kathode an Anode bestimmt die Energie der Elektronen. Durch Kaskadierung von Linsensystemen wird der Durchmesser des Primärstrahls an den Kreuzungspunkten (*cross-over*) verkleinert. Allerdings sinkt damit auch das Signal-Rausch-Verhältnis. Nach der Ablenkung des Strahles auf den gewünschten Probereich werden SE und BSE durch eine elektrische Spannung in Richtung Detektor abgezogen. Da die Energie der Elektronen quadratisch mit der Geschwindigkeit verknüpft ist, beschreiben Elektronen mit kleinerer Energie stärker gebogene Trajektorien bei gleicher Saugspannung. Damit lassen sich SE und BSE trennen und durch unterschiedliche Detektoren zählen (vgl. Massenspektrometer). Röntgenstrahlen, die bei der Nachbesetzung von Elektronen aus inneren Schalen emittiert werden, können durch Halbleiterdetektoren gemessen werden und geben Rückschlüsse auf die Elementverteilung (EDX).

In Abbildung (b) ist die Fokussierung eines Elektronenstrahls durch magnetische Linsen näher dargestellt. Auf bewegte Elektronen wirkt in einem magnetischen Feld die Lorentzkraft normal zur Bewegungsrichtung des Elektrons und normal zur Richtung des Feldes. Eine Elektronenlinse besteht aus einem axialen, rotationssymmetrischen magnetischen Feld, mit welchem die Elektronen auf eine zyklotronische Bahn gelenkt werden. Durch den Spulenstrom kann die Stärke des Feldes gesteuert werden. Eine nachgestellte Blende gewährleistet einen definierten Konvergenzwinkel.

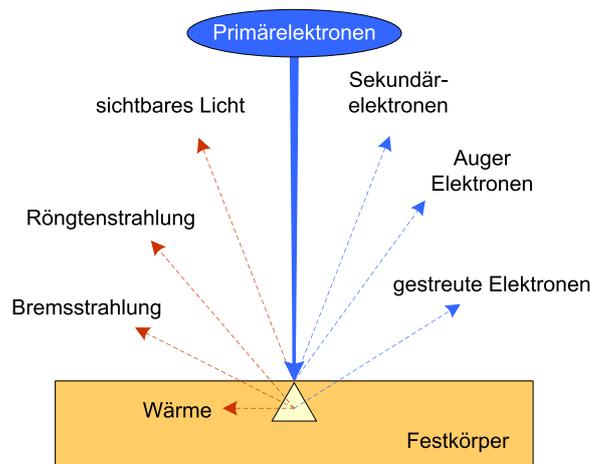


Abbildung 3.5: Von der Elektronenquelle emittierte Primärelektronen werden mit hohen Spannungen hin zur Probe beschleunigt und regen wiederum Materie und elektromagnetische Strahlung zur Emission an. Die von der Probe abgelenkten und herausgeschlagenen Teilchen (z.B. SE) werden durch eine hohe Spannung in Richtung Detektor abgezogen. Durch die Reaktionen der Elektronen mit den Atomen der Probe und durch Abbremsung der Elektronen wird Röntgenstrahlung emittiert. Die bei der Ionisierung entstehende Strahlung enthält dabei eine typische Signatur der Atome.

Abbildung 3.6 zeigt die Analyse einer Flash Speicherzelle durch die Energiedispersive Röntgenspektroskopie (EDX).

Jedoch besteht auch die Möglichkeit, die Energie an andere Elektronen abzugeben, welche die Atome als so genannte Auger Elektronen verlassen (Auger-Meitner Effekt). In Abbildung 3.7 ist das Spektrum der von der Probe emittierten Elektronen zu sehen:

- Das ausgeprägte Maximum im niederenergetischen Bereich bis ca. 50 eV rührt hauptsächlich von *Sekundärelektronen*. Die Interpretation von Sekundärelektronen stellt die am häufigst verwendete Methode dar, Rückschlüsse über die Materialbeschaffenheit zu ziehen.
- *Auger Elektronen* besitzen Energien zwischen 50 eV und 2 keV und kommen hauptsächlich von der Oberfläche der Probe. Da die Oberfläche meist durch Oxidation konterminiert ist, muss diese zuvor gereinigt und anschließend in einer Ultrahochvakuum-Kammer untersucht werden.
- Das breite Spektrum der *rückgestreuten Elektronen* umfasst per Definition den Bereich von 50 eV bis hin zur, für die Messung verwendeten, Primärenergie. Durch Mehrfachkollisionen innerhalb der Probe und die damit verbundenen Energieverluste entsteht eine breite Verteilung. Das Maximum am oberen Rand des Spektrums resultiert aus elastischer, die kinetische Energie erhaltende, Streuung.

Da SE von oberflächennahen Schichten der Probe stammen, bilden diese die Topo-

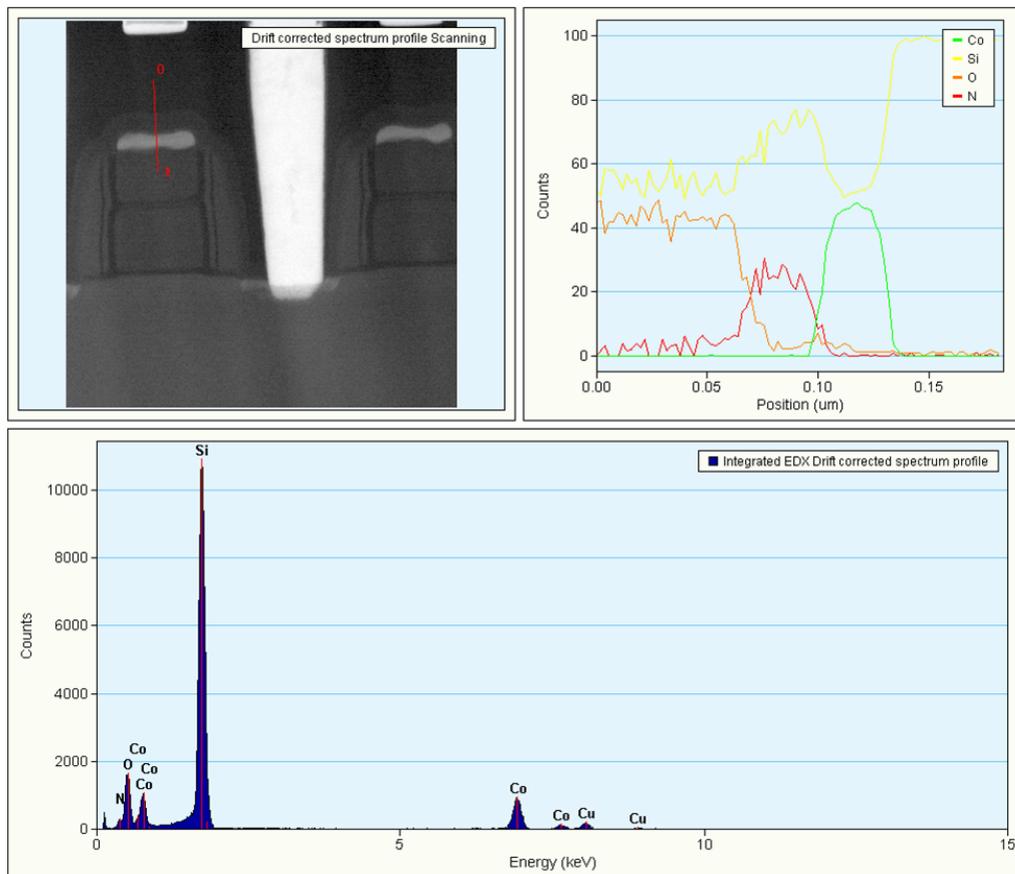


Abbildung 3.6: Elementbestimmung einer Flash Speicherzelle durch EDX. Durch die energiedispersive Röntgenspektroskopie lässt sich die Elementzusammensetzung einer Probe qualitativ und quantitativ bestimmen. Durch die Energieabgabe der Primärelektronen an die Elektronen in der Probe werden aus den Atomen Elektronen herausgelöst und dadurch ionisiert. Stammt dieses Elektron aus einer inneren Schale des Atoms, so wird dieses von einem höherenergetischen Elektron aus einer Schale mit höherem Energieniveau nachbesetzt. Die Energiedifferenz wird in der Regel durch Abgabe eines Photons mit einer der Energiedifferenz entsprechenden Röntgenfrequenz nachbesetzt. Anhand dieser Energiedifferenzen, die mit der Ordnungszahl der Elemente zusammenhängen, kann über eine Energiestatistik die Elementzusammensetzung qualitativ bestimmt werden. Im vorliegenden Beispiel wurde die Konzentration der Elemente Kobalt (*Co*), Silicium (*Si*), Sauerstoff (*O*) und Stickstoff (*N*) entlang einer Strecke durch die Speicherzelle gemessen. Die Messergebnisse in (b) zeigen die Elementstatistik von Marker #0 in Richtung Marker #1 von dem SEM Abbild in (a). Die Schicht von Position $0\ \mu\text{m}$ bis $0.07\ \mu\text{m}$ besteht aus Siliciumoxid und dient der Isolation des Transistors. Die dünne Schicht von $0.07\ \mu\text{m}$ bis $0.1\ \mu\text{m}$ enthält Siliciumnitrid, welches als *hard-mask* fungiert. Position $0.1\ \mu\text{m}$ bis $0.13\ \mu\text{m}$ besteht aus Kobaltsilizid und soll die Leitfähigkeit des Übergangs Gate-Kontakt verbessern. Ab Position $0.13\ \mu\text{m}$ dominiert Silicium, welches das Gate des Transistors bildet. Abbildung (c) zeigt die Energiestatistik der Röntgenstrahlung bei Marker #1.

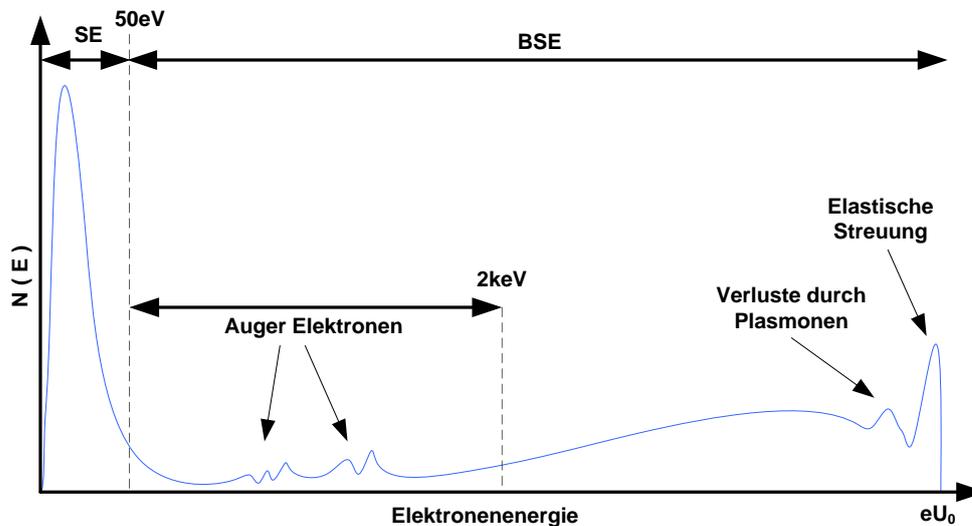


Abbildung 3.7: Typisches Energiespektrum emittierter Elektronen. Das Energiespektrum besteht aus Sekundärelektronen (SE), rückgestreuten Elektronen (BSE) und Auger Elektronen (AE). Als SE bezeichnet man Elektronen, die eine Energie unter 50 eV besitzen und durch inelastische Stöße von der Probenoberfläche entspringen. Das Maximum, also die wahrscheinlichste Energie eines SE, befindet sich zwischen $2 - 5\text{ eV}$. AE können bei der Nachbesetzung eines inneren Elektrons durch ein energetisch höheres entstehen. Dabei wird die Energiedifferenz statt durch Abgabe eines Röntgenquant an ein Elektron übergeben, welches das Atom als AE verlässt. AE kommen ebenfalls von den obersten Atomsschichten der Probe und lassen unter entsprechenden Voraussetzungen die Bestimmung des Atom-Typs zu. Elektronen mit Energien über 50 eV werden als BSE bezeichnet und sind Primärelektronen, die durch Mehrfachkollisionen unter der Probenoberfläche an Energie verloren haben. Bei Austritt aus der Probe können sie SE (sogenannte SE_2) anregen, die sich dem SE Spektrum überlagern. Die Austrittsfläche ist dabei weitaus größer als bei SE, weshalb nicht die Auflösung eines SE Bildes erreicht werden kann.

grafie der Probe am besten ab. Die austretenden Elektronen mit Energien bis zu 50 eV werden durch Anlegen einer positiven Spannung in Richtung Detektor gezogen und im Detektor durch einen Sensor erfasst. Das Abziehen der Elektronen in Richtung Detektor verursacht durch die Unebenheiten auf der Probe Schattenbildungen. Da anhand von Mehrfachkollisionen durch BSE auch SE (sogenannte SE_2) angeregt werden, sind diese den gewünschten SE überlagert. Diese machen allerdings nur einen kleinen Prozentsatz ($\approx 10\%$) des Gesamtsignales aus. Durch den oberflächennahen Ursprung der SE, der bei Leitern ca. $0.5 - 1.5\text{ nm}$ und bei Isolatoren bei ca. $10 - 20\text{ nm}$ beträgt, werden geometrische Auflösungen von $1 - 20\text{ nm}$ erreicht.

In einem REM werden dem Detektor zugeneigte Flächen und Kanten mit hoher Helligkeit abgebildet. Dieser *Kanteneffekt* begründet sich durch die vermehrte Emission von Elektronen an kleinen Erhebungen und Vertiefungen, da sich große Teile des Wechselwirkungsbereiches in der Nähe der Probenoberfläche

befinden. Ein weiterer Effekt, der von der Orientierung des Detektors relativ zur Probe abhängig ist, ist die *Abschattung*. Von Bereichen, die in Richtung Detektor durch die Probe verdeckt sind, werden weniger Elektronen gezählt als von Bereichen die dem Detektor frei zugewandt sind. *Diese Effekte sind auch für die Analyse des Bildmaterials wichtig, da er dessen Erscheinungsbild maßgeblich prägt.*

Durch die allmähliche Abnahme der kinetischen Energie der Elektronen durch Wechselwirkungen mit Materie ergibt sich eine charakteristische Eindringtiefe. Die Eindringtiefe in die Probe ist abhängig von der Energie des Elektronenstrahles und der Probenbeschaffenheit und bewegt sich in der Größenordnung von 100 nm bis $10\text{ }\mu\text{m}$. Eine empirische Formel für die Eindringtiefe kann mit

$$R = \frac{0.0276AE_0^{5/3}}{Z^{0.89}\rho} \quad (3.5)$$

angegeben werden, wobei R die Eindringtiefe in μm , A das Atomgewicht der Probe in $[\text{g/mol}]$, E_0 die Primärenergie in eV , Z die Ordnungszahl der Materie und ρ die Dichte in $[\text{g/cm}^3]$ darstellt [?]. Für Elemente, die in der Mikroelektronik häufig Verwendung finden, ist die Eindringtiefe in Abbildung 3.8 als Funktion der Beschleunigungsspannung aufgetragen.

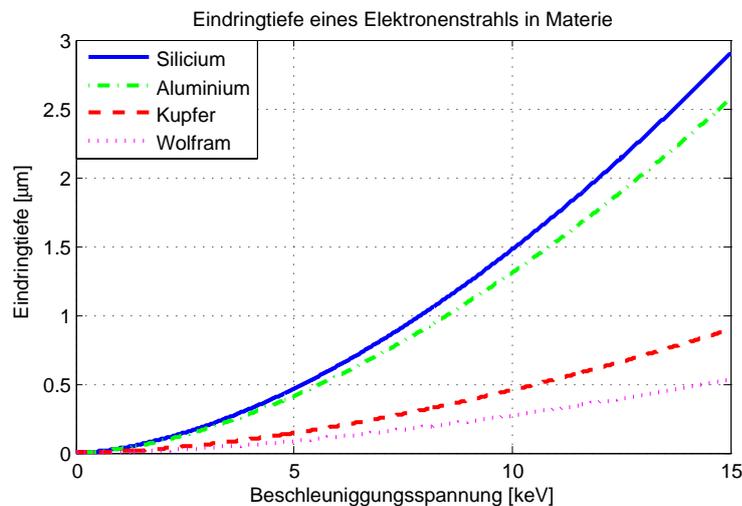


Abbildung 3.8: Eindringtiefe von Elektronen in Materie. Die Eindringtiefe eines Elektronenstrahls in die Probe ist von der Energie des auftreffenden Strahles, sowie von der atomaren Zusammensetzung der Probe abhängig. Für in der Prozesstechnologie häufig verwendete Elemente, ist die charakteristische Eindringtiefe über die Beschleunigungsspannung nach der empirischen Gleichung 3.5 gegenübergestellt. Bei schwereren Elementen mit höheren Ordnungszahlen dringt der Strahl nicht so tief in die Probe ein.

Ein Modell für die Simulation der Trajektorien eines in die Probe eindringenden Elektronenstrahls, wurde in [?] vorgestellt. Um eine bildliche Vorstellung von den Wechsel-

wirkungen in der Probe zu bekommen, wurde in Abbildung 3.9 eine Simulation mit der Software *CASINO* für verschiedene Materialien und Beschleunigungsspannungen durchgeführt.

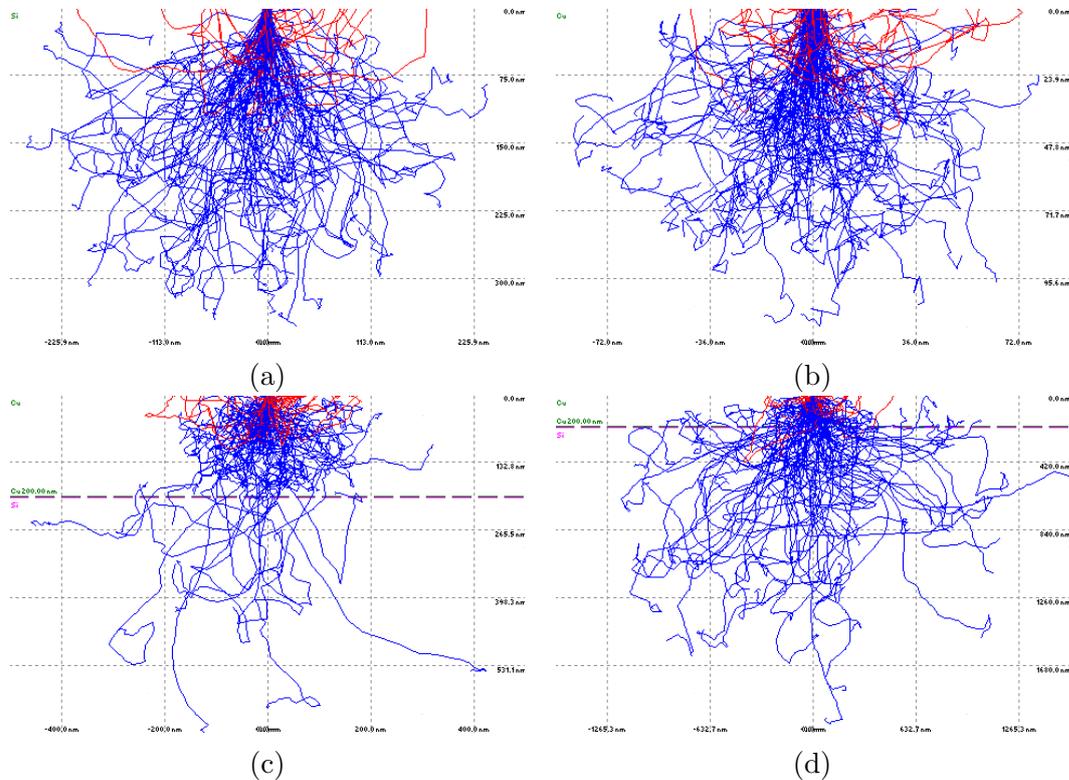


Abbildung 3.9: Simulation eines in die Probe eindringenden Elektronenstrahls mit verschiedenen Primärenergien. Die roten Trajektorien stellen Elektronen dar, die wieder aus der Probe austreten (BSE). In Abbildung (a) wurde der Primärstrahl mit einer Energie von 5 keV auf ein Siliciumsubstrat beschleunigt. Die Eindringtiefe beträgt in etwa 300 nm . Grundlage für die Abbildungen (b)-(d) bildete eine 200 nm dicke Kupferschicht mit einem darunter liegenden Siliciumsubstrat. Die Primärenergie wurde in 5 keV Schritten von 5 keV bis 15 keV erhöht. Bei einer Primärenergie von 5 keV (b) ruft der Primärstrahl Wechselwirkungen bis in eine Tiefe von etwa 100 nm hervor, bleibt also innerhalb der Kupferschicht. Ein Primärstrahl mit einer Energie von 10 keV (c) ruft bereits Wechselwirkungen im sich unter dem *Cu* befindlichen *Si*-Substrat hervor. Bei weiterer Erhöhung auf 15 keV (d) treten bereits Elektronen, die aus dem Substrat kommen, wieder an die Oberfläche und werden vom Elektronendetektor gezählt. Zu bemerken ist hierbei, dass die Abbildungen (a)-(d) in unterschiedlichen Maßstäben dargestellt wurden.

3.3.3 Verwendete Abbildungsgeräte und Parameter

Zur Abbildung der Chipebenen im Lichtmikroskop wurde das Mikroskop *Polyvar Met* von *Reichert-Jung* verwendet. Die Elektronenmikroskop-Bilder wurden mit einem Chipscanner der Firma *Raith GmbH* erzeugt (Abbildung 3.10). Dieser besteht aus einer Kombination

aus Elektronenmikroskop und einem, im Nanometerbereich verstellbaren, Probetisch. Um Bereiche automatisch anfahren zu können, wird die absolute Position durch ein Laserinterferometer bestimmt. Typischerweise entstehen die Bilderserien durch mäanderförmige Rasterung der Chipoberfläche. Die *working distance* ist durch den hochpräzisen Probetisch auf 2 nm fixiert. Die Beschleunigungsspannung wurde je nach Dicke der auf dem Metall verbleibenden Oxidschicht zwischen $5 - 20\text{ keV}$ festgesetzt. Um auch bei schräg liegendem Chip über die ganze Ebene hinweg fokussierte Bilder zu erhalten, kann eine Fokusebene definiert werden.

Um moderne Mikrochips strukturgenaue abzubilden, bedarf es einer geometrischen Auflösung von $5 - 10\text{ nm}$ pro Pixel. Um eine Chipfläche von einem mm^2 in einer Auflösung von 10 nm pro Pixel abzubilden, benötigt es $(10^{-3} \times 10^{-3}) / (10 \times 10^{-8} \times 10 \times 10^{-8}) = 10^{10}$ Pixel. Bei einem Bildsensor mit einem Megapixel entstehen damit pro Ebene 10000 Bilder. Diese Datenmengen stellen eine große Herausforderung an die Bildanalyse dar.

Die für die Arbeit verwendeten Aufnahmen wurden von Mitarbeitern der Firma Infineon Technologies erstellt und zu Analysezwecken zu Verfügung gestellt.



Abbildung 3.10: Der Raith 200 Chip-Scanner. Die für die Arbeit verwendeten REM Bilder wurden mit einem Raith 200 Chipscanner erzeugt. Dieser besteht aus einem REM, einem im Nanobereich verstellbaren Probetisch und einem Laserinterferometer zur Bestimmung der absoluten Position des Probetisches. In Verbindung mit einer vorhergehenden Kalibrierung ist es möglich, eine Chipebene automatisch und verzerrungsfrei abzubilden.

Vision is the art of seeing what is invisible to others.

Jonathan Swift

Kapitel 4

Bildanalytisch gestütztes Reverse Engineering

4.1 Einleitung

Bildanalyse dient in vielen industriellen Bereichen der Automatisierung von Abläufen, welche eine optische Inspektion erfordern. Die Einsatzmöglichkeiten sind dabei sehr vielfältig und reichen von Qualitätskontrolle über Biometrie bis hin zur Steuerung von industriellen Robotern. In dieser Arbeit soll Bildanalyse genutzt werden, um Designinformation aus Abbildungen von integrierten Schaltungen (*integrated circuits* (IC)) zu extrahieren und damit Analysezwecken zugänglich zu machen. Jedes Halbleiterunternehmen führt im Rahmen der Prozesskontrolle fortlaufend Analysen der eigenen Produkte durch, um Prozessparameter optimal regeln zu können. Designanalysen von Fremdprodukten, diese beinhalten Schaltungsanalysen und Layoutanalysen, werden jedoch in der Regel aufgrund der Wirtschaftlichkeit an Fremdfirmen übertragen. Mit der fortlaufenden Miniaturisierung der Strukturgrößen integrierter Schaltungen und der damit einhergehenden Steigerung der Kosten des Prozessequipments, besteht der Trend dahin *fabless*, d.h. ohne eigene Produktionsstätte, zu produzieren und sich damit auf Design Know-How zu konzentrieren. Dadurch gewinnt das Thema Designanalyse an Stellenwert. Motivationsgründe warum es sinnvoll oder sogar für das Unternehmen überlebensnotwendig ist, Schaltungsanalyse zu betreiben, wurden bereits in Abschnitt 1.1 geschildert. Bei der enormen Komplexität aktueller ICs ist es jedoch unmöglich, mittelgroße bis große Schaltungsbereiche manuell zu analysieren. Dies fordert Automatisierung.

4.2 Überblick

In der folgenden Arbeit wird ein Konzept zur Rückgewinnung von Designinformation und Designparametern aus Integrierten Schaltungen vorgestellt. Der Schwerpunkt liegt hierbei

in der Evaluierung von bildverarbeitungsgestützten Methoden zur (semi) automatischen Extraktion. Als Bildmaterial dienen Aufnahmen von Rasterelektronenmikroskopen sowie Bilder von optischen Mikroskopen. Aufgrund der geringen Strukturgrößen auf den Proben sowie der enormen Komplexität aktueller Designs entstehen bei der Abbildung große Mengen an Daten, die verarbeitet werden müssen. Zwangsläufig besteht der Bedarf an effizienten und optimierten Methoden, die einen wirtschaftlichen Nutzen überhaupt erst möglich machen. Abbildung 4.1 zeigt einen Überblick über den Workflow, der von der Aufnahme über die verschiedenen Phasen der Extraktion bis hin zur planaren Schaltung führt. In dieser Arbeit wurde folgenden Problemstellungen Rechnung getragen:

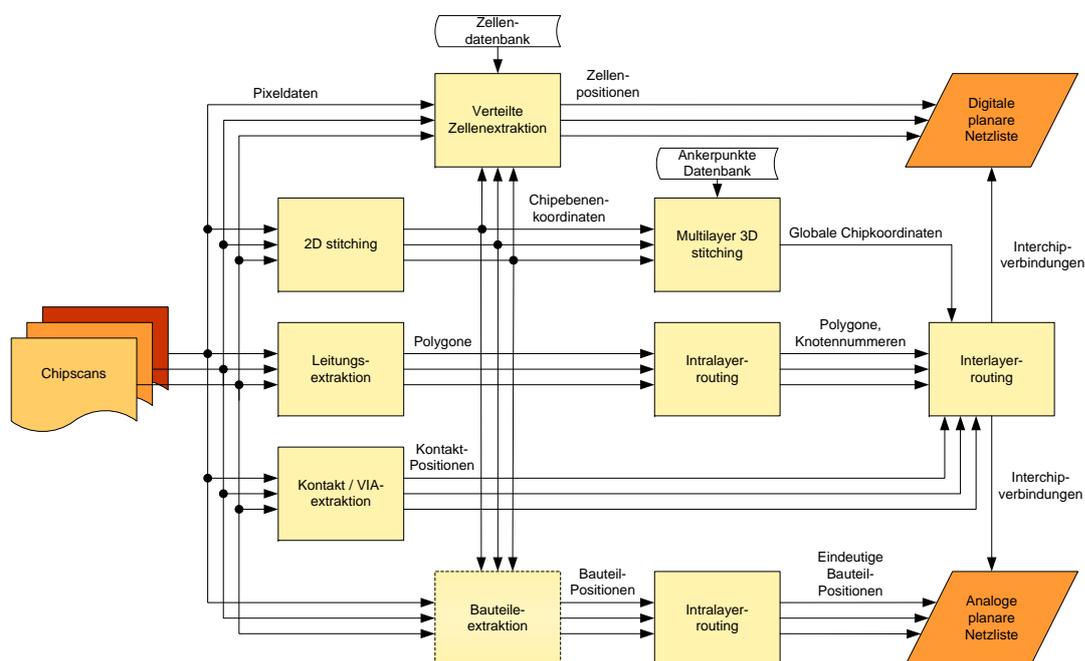


Abbildung 4.1: Der Bildanalyse RE Workflow. Datengrundlage für die Analyse bilden Aufnahmeserien der Chipebenen. Nicht alle Berechnungen erfordern dabei registrierte Bilder und können damit parallel zur Registrierung durchgeführt werden. Die Extraktion von Leitungen, Kontakten und VIAs kann ohne Zuhilfenahme der Nachbarbilder durchgeführt werden, da der immer vorhandene Überlapp ein späteres Zusammenführen gewährleistet (Intralayer-Routing). Standardzellen und Bauteile können hingegen über mehrere Bilder verteilt sein und damit bei einzelner Betrachtung nicht gefunden werden. Der Schritt des Interlayer-Routings verbindet Zellen und Bauteile durch Leitungen und VIAs aller Lagen und ermöglicht damit die Erstellung von planaren Netzlisten.

- Teilbilder einer Aufnahmereihe müssen zu einem Gesamtbild zusammengesetzt werden. Dies muss mit sehr hoher Genauigkeit erfolgen, da die resultierenden Gesamtbilder als Datengrundlage der nachfolgenden Analysen dienen und Abweichungen von wenigen μm bereits zu einem Versatz führen, der eine weitere Verarbeitung

unmöglich macht. Aus diesem Grund wurde ein vergleichsweise hoher Evaluierungsaufwand zu dieser Problemstellung betrieben. (Kapitel 5)

- Nach der planaren Registrierung der Chipebenen ist es notwendig, die Ebenen dreidimensional auszurichten (*3D stitching*). Hierzu wird eine Ebene als Referenzebene definiert und die übrigen durch Auffinden von Korrespondenz zur jeweiligen Nachbarebene ausgerichtet. (Kapitel 5)
- Die Extraktion von passiven und aktiven *Bauteilen* wie Widerstände, Kapazitäten, Spulen, Dioden und Transistoren. Diese Analysen werden auf den unteren Ebenen des Mikrochips durchgeführt, auf denen Bauelemente physikalisch realisiert sind. (Kapitel 6.5)
- *Speicherzellenanalyse* zur Abschätzung der Anzahl der RAM/ROM Zellen und zur Gewinnung von Information über Redundanz und Zellenaufbau. (Kapitel 7)
- Analyse von *Standardzellen* und Ableiten von Kennzahlen wie *utilisation*, Gatteräquivalent und Statistiken, die auf die Syntheseoptionen rückschließen lassen. (Kapitel 7)
- Extraktion von *Maskendaten* der Metallisierungsebenen und die Überführung in eine approximative polygonale Beschreibung. Durch das Einbringen von Vorwissen wie Design-Rules können die, durch nichtideale Präparierung entstandenen, Fehlerkennungen detektiert werden (Kapitel 6).
- Durch den schichtweisen Aufbau eines Mikrochips ist es erforderlich, durch Auffinden von *Interlayerkontakten* einen Zusammenhang zwischen den unterschiedlichen Lagen herzustellen. Dies setzt eine präzise Ausrichtung (*global alignment*) der einzelnen Lagen untereinander voraus (Kapitel 6).
- In der *Routingphase* wird die polygonale Beschreibung weiter vereinfacht und damit eine Darstellung als Graph von Knoten ermöglicht. Durch graphentheoretische Algorithmen können Plausibilitätstests durchgeführt und Fehler erkannt werden (Kapitel 6).

Kapitel 5

Stitching

Zusammenfassung — In diesem Abschnitt werden verschiedene Methoden zur Bildregistrierung von mehrlagigen Bildern vorgestellt und evaluiert. Der Schwerpunkt liegt dabei auf der Vermeidung von Fortpflanzungsfehlern sowie auf der Erzielung möglichst hoher Genauigkeit für Bilderserien mit mehreren Gigapixel. Dabei wurde ein neuer Ansatz zur konsistenten Ausrichtung mehrerer Chiplagen entwickelt, welcher die hohen praktischen Anforderungen erfüllt.

Inhaltsverzeichnis

5.1	Einleitung	33
5.2	Überblick	35
5.3	Merkmalsextraktion und Matching	37
5.4	Koordinatentransformationen	51
5.5	Genauigkeitsanalysen	64
5.6	Radiometrische Korrekturen	73

5.1 Einleitung

Die Strukturichte moderner Mikrochips steigt zunehmend an und fordert damit an allen Fronten höchste Aufbietung. Dies schließt auch die Analyse nicht aus. Um Schaltungsanalyse betreiben zu können, müssen die Lagen eines Mikrochips strukturgenaue abgebildet werden. Da das Sehfeld der Abbildungsgeräte jedoch eingeschränkt ist, muss das Gesamtbild in eine Reihe von Bildern aufgeteilt werden. Ein solcher *chipscan* erfordert je nach Chip- und Strukturgröße enorme Datenmengen, die in wirtschaftlicher Zeit verarbeitet werden müssen. Durch den mehrlagigen Aufbau eines Mikrochips ist weiter gefordert, dass die Einzelbilder auch über die Lagen hinweg präzise ausgerichtet sein müssen, d.h. die Bildregistrierung mit Berücksichtigung der benachbarten Lagen dreidimensional durchgeführt werden muss.

Fehler in der automatischen Registrierung einer großen Menge von Teilbildern können fehlerhafte Ausgangsdaten für die nachfolgende Analyse liefern. Überschreiten die Fehler das vorgegebene Toleranzintervall, so führt dies nicht nur zu einer teilweisen fehlerhaften Interpretation sondern kann auch die gesamte Schaltungsanalyse zunichte machen. Diese Forderung stellt hohe Ansprüche an Präzision und Effizienz der involvierten Algorithmen. Der Prozess mit dem Ziel, einen nahtlosen Übergang an den Einzelbildgrenzen zu erreichen, wird als *image stitching*, Bildregistrierung (*image registration*) oder auch *image mosaicing* bezeichnet. In dieser Arbeit werden diese Begriffe austauschbar verwendet.

Ein Gefühl für die zu verarbeitenden Datenmengen soll ein kurzes Rechenbeispiel geben. Ein derzeit aktueller Desktop Prozessor von Intel (Modell i7-980X) mit 6 Kernen wird mit einer kleinsten Strukturgröße von 32 nm gefertigt und besitzt eine Fläche von $42.5\text{ mm} \times 45\text{ mm}$. Soll nun der gesamte Chip strukturgenaue abgebildet werden, so sollte die auf einen Pixel abgebildete Fläche nicht größer sein als $\frac{32}{5}\text{ nm} \times \frac{32}{5}\text{ nm}$ um überhaupt noch vernünftige Bildanalyse betreiben zu können. Setzt man nun die Fläche eines Pixel in Relation zur Gesamtfläche des Chips, so erhält man eine Auflösung von $6\,640\,625\text{ Pixel} \times 6\,875\,000\text{ Pixel}$. Bei einer Größe eines Einzelbildes von 1 MPixel erhält man die gewaltige Anzahl von $45\,652\,96$ Bildern. Bei einer durchschnittlichen Dateigröße von 500 kB ergibt dies einen Speicherbedarf von 22 TB pro Lage. Nun ist oft nicht der ganze Chip von Interesse sondern man beschränkt sich auf ausgesuchte Bereiche, in etwa zur Überprüfung von Patentansprüchen. Aber auch ein Bereich von einem mm^2 benötigt bei einer Strukturgröße von 32 nm in etwa $24\,000$ Teilbilder und damit 12 GB an Daten ohne den für die Bildregistrierung notwendigen Bildüberlapp zu berücksichtigen.

Weiter werden auch in punkto Registrierungsgenauigkeit sehr hohe Anforderungen gestellt. Ein aus 100×100 Teilbildern á 1 MPixel bestehendes Mosaik darf an keiner Stelle eine höhere Abweichung als die halbe Strukturgröße aufweisen um keine Fehlverbindungen zu benachbarten Lagen zu verursachen. Da eine halbe Strukturbreite in der Regel auf $5 - 10\text{ Pixel}$ abgebildet wird, erfordert dies eine Registrierungsgenauigkeit von $0.05 - 0.1\text{ Pixel}$ von Bild zu Bild. Bei mangelnder Bildqualität und Bildern mit wenig Informationsgehalt ist diese Forderung meist nicht einzuhalten und erfordert deshalb die Einbeziehung der Nachbarlagen zur Positionsbestimmung.

Auch in anderen Fachgebieten besteht der Bedarf nach performanten, hochgenauen Registrierungsmethoden, die im Grunde vielseitig einsetzbar sind. Heutige Geoinformationssysteme wie Google's Maps oder Microsoft's Pentant Bing Maps bieten jedem der Zugang zum Internet hat, die Möglichkeit, kostenlos ferne Länder zu erkunden, die Reise-route für den nächsten Urlaub zu planen, berühmte Bauwerke von oben zu betrachten und selbst erstellte GPS Daten zu visualisieren. Bei hoch aufgelösten Gebieten reicht die geometrische Auflösung bis weit unter 1 m^2 per Pixel. Die Einzelbilder, die in Summe hunderte Terabyte umfassen, müssen dabei konsistent zueinander ausgerichtet und entzerrt werden, damit das Gesamtbild keinen Versatz beinhaltet. Dies stellt eine große Herausforderung an die Algorithmen und deren Genauigkeit sowie Robustheit dar.

5.2 Überblick

Das nahtlose Zusammenfügen von sich überlagernden Bilderserien ist ein wohlbekanntes Problem der Bildverarbeitung. Die damit verbundenen Algorithmen sind größtenteils allgemeiner Natur und gut verstanden. Die Anwendungen sind dabei vielfältig und stellen ihre eigenen Randbedingungen, weshalb die Thematik noch immer zu einem aktiven Forschungsgebiet zählt. Dazu zählen Anwendungen wie das Erstellen von Panorambildern [?, ?, ?], die virtuelle Mikroskopie, in der das Sehfeld durch zuvor aufgenommene Bilder erweitert wird [?, ?, ?, ?, ?], Anwendungen in der erweiterten Realität (*augmented reality*) oder virtuellen Realität (*virtual reality*) [?, ?] und Anwendungen in der eine Reihe von Luftbildaufnahmen aus mehreren Sensoren in Echtzeit verarbeitet werden [?, ?]. Ausgezeichnete Einführungen zum Thema können in [?, ?] gefunden werden.

Grundlegend lassen sich *image stitching* Algorithmen in zwei Gruppen einteilen:

- Auf *Merkmale basierende Methoden* (*feature based methods*) extrahieren Bildmerkmale, die möglichst große Invarianz zu den zugrundeliegenden Bildtransformationen aufweisen. Diese als *point of interest* (POI) bezeichneten Merkmale sollen eindeutig durch ihre Umgebung gekennzeichnet sein, um damit das Auffinden in unterschiedlichen Aufnahmen derselben Szene zu ermöglichen. In einer anschließenden *matching* Phase werden Merkmale aus verschiedenen Bildern verglichen, um Ausreißer zu eliminieren und die Parameter der abbildenden Transformation zu ermitteln.
- *Direkte Methoden* (*area based methods*) vergleichen Bildausschnitte unterschiedlicher Aufnahmen durch ein Ähnlichkeitsmaß. Ist die Struktur der zugrundeliegenden Bildtransformation rein translatorischer Natur, können durch dieses Verfahren sehr robuste Ergebnisse erhalten werden. Bei komplexen Transformationen muss erhöhter Aufwand betrieben werden.

Um einen Überblick über den Ablauf der Bildregistrierung zu erlangen, wird der Prozess in dieser Arbeit in Phasen unterteilt, auf die im Anschluss näher eingegangen wird. Je nach Anwendung können auch Teile einer Phase oder ganze Phasen entfallen, da sich die Anforderungen an geometrische und radiometrische Genauigkeit unterscheiden. Allgemein können die Phasen folgendermaßen eingeteilt werden (für die Definition der Koordinatensysteme und derer Transformationen siehe Abschnitt 5.4) :

- *Beseitigung von Nichtlinearitäten in der Abbildung*. Da Bildtransformationen in Folge als linear angenommen werden, müssen Nichtlinearitäten, die durch Unzulänglichkeiten des abbildenden Systems entstanden sind, korrigiert werden.

In diesem Schritt muss ein Fehlermodell gewählt und deren Parameter durch eine Kalibrierungsphase ermittelt werden. (Abschnitt 5.4.1)

- *Detektion von Merkmalen.* Das Auffinden von eindeutigen, unterscheidbaren Objekten wie Kanten, Ecken, Konturen oder Kreuzungspunkten. Bei ausgedehnten Merkmalen kann der Schwerpunkt des Merkmals als Kontrollpunkt verwendet werden. Dieser Schritt kann bei Verwendung einer direkten Methode entfallen. (Abschnitt 5.3.1)
- *Struktur der Transformationen.* Die Struktur der Transformation (lokal sowie global) muss im Vorfeld ermittelt werden, um die Anzahl der zu ermittelnden Parameter zu begrenzen und damit unnötige Fehlerquellen auszuschließen. (Abschnitt 5.4.2)
- *Lösung des Korrespondenzproblems.* Dabei werden die Merkmale zweier Bilder verglichen und Korrespondenzen ermittelt. Ohne das Vorwissen der Position im Mosaik ist es notwendig, jedes Bild mit jedem zu vergleichen. Aus den Korrespondenzen können im Anschluss die lokalen Bildtransformationen ermittelt werden. (Abschnitte 5.3.1 und 5.3.2)
- *Abschätzung der Chi Ebenentransformation.* Durch die Verknüpfung von lokalen Bildtransformationen können Chi Ebenentransformationen berechnet werden, die die Bilder vom lokalen Bildkoordinatensystem in das Chi Ebenenkoordinatensystem überführen. Dabei ist das Problem der Fehlerfortpflanzung zu berücksichtigen und es sind geeignete Methoden zu entwickeln, um diese Fehler gering zu halten. (Abschnitt 5.4.3)
- *Abschätzung der globalen Chiptransformation.* Nach der planaren Registrierung der Chi Ebenen ist es weiters notwendig, die Ebenen dreidimensional auszurichten. Hierzu wird eine Ebene als Referenzebene definiert und die übrigen durch Auffinden von Korrespondenz zur jeweiligen Nachbarebene ausgerichtet. Dabei werden zwei Methoden vorgeschlagen. Die erste, indirekte Methode berechnet eine affine Abbildung auf Basis der ermittelten Chi Ebenenkoordinaten, um eine lineare Verzerrung zu korrigieren. Durch Registrierungsfehler sind die einzelnen Lagen jedoch in der Regel nichtlinear verzerrt und können durch die indirekte Methode nicht konsistent ausgerichtet werden. Durch Einbeziehung sog. Ankerpunkte berechnet der zweite, direkte Ansatz aus den lokalen Bildtransformationen aller Lagen direkt die globalen Chipkoordinaten, er bestimmt also die Positionen aller Einzelbilder aller Lagen in einem Berechnungsschritt. (Abschnitt 5.4.4)
- *Bildtransformation.* Die gefundene Transformation vom lokalen Bildkoordinatensystem in das globale Chipkoordinatensystem wird nun auf die Bilddaten angewendet. Die dabei entstehenden Pixelintensitäten mit nichtganzzahligen Pixelpositionen müssen durch Interpolationsalgorithmen berechnet werden (*resampling*). (Abschnitt 5.5.1)

- *Überblendung sich überlappender Bereiche.* Da sich die Farb- und Helligkeitsintensität durch verschiedene Aufnahmebedingungen unterscheiden kann, folgt ein abschließender Abgleich durch Überblendung der überlappenden Bereiche. Dies gewährleistet unter Voraussetzung einer genauen Bildtransformation einen nahtlosen Übergang von Bild zu Bild. (Abschnitt 5.6)

5.3 Merkmalsextraktion und Matching

5.3.1 Auf Merkmalen basierende Methoden

Die in der Literatur als *points of interest* (POI) bekannten Merkmale basieren auf lokal eindeutigen Merkmalen, die folgende Eigenschaften ausweisen sollten:

- *Hohe Reproduzierbarkeit.* Damit POI wiedergefunden werden können, wird hohe Reproduzierbarkeit angestrebt. POI sollen dabei insensitive auf Helligkeitsschwankungen und Bildrauschen sein.
- *Hoher Informationsgehalt.* Die Umgebung des POI soll reich an Information und damit eindeutig identifizierbar sein. Dies ermöglicht eine robuste und genaue Lokalisierung, auch unter verschiedensten Abbildungen.

5.3.1.1 Stand der Technik

Die in der Literatur beschriebenen POI Detektoren lassen sich in drei Kategorien einteilen: konturbasierende, intensitätsbasierende und auf einem parametrischen Modell basierende Methoden [?]. Da in dieser Arbeit intensitätsbasierende Detektoren Anwendung finden, werden die beiden anderen Kategorien nur kurz umrissen.

Konturbasierende Methoden extrahieren in einer ersten Phase Konturen und suchen anschließend nach Krümmungsmaxima entlang des Konturkammes oder approximieren die Kontur durch ein Polygon, um Schnittpunkte dieser zu finden. Signifikante Änderungen entlang der Konturen werden in [?] durch den *curvature primal sketch* beschrieben. Dies ermöglicht eine skalierungsunabhängige Klassifizierung von Objekten und erlaubt weiters die approximative Rekonstruktion durch Spline-Kurven. In anderen Arbeiten werden POI als Maxima der Kontur beschreibenden B-Splines, durch Schnittpunkte von in die Kontur gefitteten Linien oder auch durch Unterteilung in stückweise konvexe und konkave Konturabschnitte beschrieben.

Modellbasierende Methoden passen die Parameter eines zuvor gewählten intensitätsbasierenden Modells an einen Ausschnitt des Bilds an. Diese Methode ist nur mit Vorwissen über die im Bild vorhandenen Objekte anwendbar.

Intensitätsbasierende Methoden berechnen ein Maß für das Vorhandensein eines POI auf Basis der umliegenden Grauwerte. Einer der ersten signalbasierenden POI Detektoren wurde von [?] vorgestellt. Dieser Detektor basiert auf der Autokorrelationsfunktion eines Signals und misst die Signaldifferenz zwischen einem, dem Punkt umliegenden Fenster und einem, in den Richtungen der 8-Nachbarschaft verschobenen Fenster, also die lokalen Änderungen in der Umgebung des Punktes. Mit dem Bild $\mathbf{I}(x, y)$, einer Fensterfunktion $\mathbf{W}(x, y)$ und dem Verschiebungsvektor $\begin{pmatrix} u & v \end{pmatrix}^T$ kann man den Unterschied $\mathbf{E}(u, v)$ folgendermaßen spezifizieren:

$$\mathbf{E}(u, v) = \sum_{x, y} \mathbf{W}(x, y) |\mathbf{I}(x + u, y + v) - \mathbf{I}(x, y)|^2 \quad (5.1)$$

Mit $u = -1 \dots 1, v = -1 \dots 1$ und einer Fensterfunktion $\mathbf{W}(x, y)$, die in einer spezifizierten Region den Wert Eins und außerhalb dieser den Wert Null besitzt. Wenn das Minimum der Teilergebnisse $\min(\mathbf{E})$ einen festgesetzten Schwellwert übersteigt, gilt der Punkt als markant. Da nur Richtungen in 45° Schritten berücksichtigt werden, entsteht eine anisotropische Antwort, d.h. es besteht eine ungewünschte Winkelabhängigkeit. Um diese Winkelabhängigkeit zu beseitigen und damit einen rotationsinvarianten Detektor zu erhalten, verwenden [?] eine analytische Erweiterung durch eine Taylorreihe im Fenstermittelpunkt:

$$\mathbf{I}(x + u, y + v) = \mathbf{I}(x, y) + \frac{\partial \mathbf{I}(x, y)}{\partial x} u + \frac{\partial \mathbf{I}(x, y)}{\partial y} v + O(x, y) \quad (5.2)$$

Bei Abbruch nach dem linearen Term und mit $\mathbf{I}_x = \frac{\partial \mathbf{I}}{\partial x}$ und $\mathbf{I}_y = \frac{\partial \mathbf{I}}{\partial y}$ folgt nach Substitution von Gleichung 5.2 in Gleichung 5.1:

$$\mathbf{E}(u, v) = \sum_{x, y} \mathbf{W}(x, y) |\mathbf{I}_x u + \mathbf{I}_y v|^2 \quad (5.3)$$

$$= \sum_{x, y} \mathbf{W}(x, y) |\mathbf{I}_x^2 u^2 + 2\mathbf{I}_x \mathbf{I}_y uv + \mathbf{I}_y^2 v^2| \quad (5.4)$$

$$= \begin{pmatrix} u & v \end{pmatrix} \begin{bmatrix} \sum_{\mathbf{W}} \mathbf{I}_x^2 & \sum_{\mathbf{W}} \mathbf{I}_x \mathbf{I}_y \\ \sum_{\mathbf{W}} \mathbf{I}_x \mathbf{I}_y & \sum_{\mathbf{W}} \mathbf{I}_y^2 \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \quad (5.5)$$

$$= \begin{pmatrix} u & v \end{pmatrix} \mathbf{A} \begin{pmatrix} u \\ v \end{pmatrix} \quad (5.6)$$

Der hierbei ermittelte Tensor \mathbf{A} , der die zweiten Momente des Bildes \mathbf{I} enthält und in direkter Verbindung mit der Autokorrelationsfunktion (Gleichung 5.1) steht, wird als *structure tensor* bezeichnet und stellt die Grundlage für eine Reihe von Detektoren dar. Als Fensterfunktion $\mathbf{W}(x, y)$ wird oft statt einer binären Funktion eine gauß'sche Verteilungsfunktion verwendet, um die Rauschneigung des Detektors zu vermindern und genauer definierte Maxima zu erhalten.

Für eine praktische Berechnung des *structure tensor* können die ersten partiellen Ableitungen des Bildes I durch Faltung mit der abgeleiteten gauß'schen Verteilungsfunktion

gewonnen werden.

$$g_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5.7)$$

Damit lässt sich der Gradient des Bildes durch

$$\mathbf{I}_x = \mathbf{I} \star \frac{\partial g_\sigma}{\partial x}, \mathbf{I}_y = \mathbf{I} \star \frac{\partial g_\sigma}{\partial y} \quad (5.8)$$

approximieren.

Wobei der Operator \star die Faltung darstellt. Durch das Vektorprodukt von $(\mathbf{I}_x \ \mathbf{I}_y)^T$ mit sich selbst, lässt sich der *structure tensor* \mathbf{A}_σ konstruieren.

$$\mathbf{A}_\sigma = \nabla \mathbf{I} \star \nabla \mathbf{I}^T = \begin{bmatrix} \mathbf{I}_x^2 & \mathbf{I}_x \mathbf{I}_y \\ \mathbf{I}_x \mathbf{I}_y & \mathbf{I}_y^2 \end{bmatrix} \quad (5.9)$$

Da \mathbf{A}_σ nur einen von Null verschiedenen Eigenwert aufweist und dadurch ausschließlich eindimensionale intrinsische Information (Kanteninformation) besitzt, muss diese durch eine Glättung in der Umgebung verteilt werden. Dadurch erhält jeder Punkt Beiträge von Kanten verschiedener Orientierung und damit 2D Information.

Da die Analyse des *structure tensors* im Zentrum der folgenden Ausführung steht, werden dessen Eigenschaften kurz angeführt:

- \mathbf{A} ist symmetrisch und positiv definit.
- Die Eigenwerte λ_1 und λ_2 sind positiv.

Die Eigenvektoren \mathbf{v}_1 und \mathbf{v}_2 der Matrix \mathbf{A}_σ geben Auskunft über die dominanten lokalen Orientierungen und die Eigenwerte λ_1 und λ_2 über die Stärke der jeweiligen Orientierung. Eine Klassifikation kann nun über die Eigenwerte wie folgt erfolgen:

- ($\lambda_1 \ll, \lambda_2 \ll$) Wenn beide Eigenwerte klein sind, liegt ein mehr oder weniger homogener Bereich vor.
- ($\lambda_1 \gg, \lambda_2 \ll$) \vee ($\lambda_1 \ll, \lambda_2 \gg$) Ist einer der Eigenwerte groß und der andere klein, handelt es sich um eine nur in eine Richtung ausgedehnte Struktur, d.h. um eine Kante.
- ($\lambda_1 \gg, \lambda_2 \gg$) Sind beide Eigenwerte groß, liegt also eine große Variation in beide Richtungen vor, so handelt es sich um eine Ecke.

Die Fläche der durch die Eigenvektoren und Eigenwerte aufgespannten Ellipse ist proportional zur Determinante $\det(\mathbf{M}) = \lambda_1 \lambda_2$, die als Maß für die „Eckigkeit“ (*cornerness*) herangezogen werden kann. Aufgrund von Geschwindigkeitsgründen wird die explizite Berechnung der Eigenwerte in der Praxis nicht durchgeführt.

[?] nutzen in ihrer Arbeit die Kovarianzmatrix $\mathbf{C} = \mathbf{A}^{-1}$. Hohe Gradienten der Autokorrelationsmatrix entsprechen kleinen Varianzen der Kovarianzmatrix. Durch die Eigenwertanalyse der Kovarianzmatrix lässt sich eine Fehlerellipse konstruieren, die zur Bestimmung der *cornerness* herangezogen wird. Eine kleine runde Fehlerellipse entspricht einer hohen *cornerness*. Die Größe der Fehlerellipse $\frac{1}{K_{Foerstner}}$ kann wie folgt definiert werden:

$$\frac{1}{K_{Foerstner}} = \frac{1}{\lambda_{C1} + \lambda_{C2}} = \frac{1}{\text{trace}(\mathbf{C})} = \frac{\det(\mathbf{A})}{\text{trace}(\mathbf{A})} \quad (5.10)$$

Die Richtung der dominanten Hauptachse der Fehlerellipse kann aus

$$\tan 2\phi = \frac{2\mathbf{A}_{12}}{\mathbf{A}_{11} - \mathbf{A}_{22}} \quad (5.11)$$

abgeleitet werden. Liegt im Fenster eine Kante, wird dessen Richtung durch ϕ wiedergegeben. Eine Kennzahl für die Rundheit der Ellipse kann wie folgt berechnet werden:

- Das Verhältnis der beiden Eigenwerte. Dabei ist das Verhältnis unabhängig davon, ob \mathbf{A} oder \mathbf{C} zur Berechnung der Eigenwerte herangezogen wird. Das Verhältnis lässt sich als *signal to noise* (SNR) einer Kante interpretieren, da ein Eigenwert die Varianz in Kantenrichtung beschreibt und der Andere die Varianz normal dazu.

$$SNR^2 = \frac{\lambda_1}{\lambda_2} \quad (5.12)$$

Ein Nachteil dieser Methode ist, dass die Eigenwerte explizit berechnet werden müssen.

- Ein anderes Maß für die Rundheit der Ellipse kann durch direkte Berechnung aus den Komponenten der Matrix gewonnen werden:

$$q = 1 - \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} = \frac{\text{trace}(\mathbf{A})^2}{4\det(\mathbf{A})} = 1 - \frac{SNR^2 - 1}{SNR^2 + 2} \quad (5.13)$$

Durch die Normalisierung erreicht man, dass $q \in [0, 1]$. Wenn $q = 1$ wird die Ellipse zum Kreis, da die Eigenwerte identisch sind und $SNR = 1$ ist. Ist $q = 0$ liegt das Fenster an einer idealen Kante. Das SNR kann mit Wissen von q durch $SNR = \frac{1+\sqrt{1-q}}{1-\sqrt{1-q}}$ berechnet werden [?].

In ähnlicher Weise wurden eine Reihe von Funktionen als Maß für die *cornerness* eines Fensterbereiches vorgeschlagen. Einer der bekanntesten Detektoren von [?] ist wie folgt definiert:

$$K_{Harris\&Stephens} = \det(\mathbf{A}) - k * \text{trace}(\mathbf{A})^2 \quad (5.14)$$

durch passende Wahl der empirischen Konstante k (typischerweise $k \in [0.04, 0.15]$) kann anhand des Vorzeichens zwischen Kante ($K_{Harris\&Stephens} < 0$) und Ecke ($K_{Harris\&Stephens} \geq 0$) unterschieden werden. Der zweite Term unterdrückt Punkte mit

nur einem großen Eigenwert.

[?], die ihre Arbeit aus dem Bereich *feature tracking* motivieren, begründen ihre Wahl damit, dass zwei große Eigenwerte entweder eine Ecke, *salt and pepper* Rauschen oder eine andere Textur, die man gut verfolgen kann, repräsentieren. Unter der Voraussetzung, dass die Eigenwerte das Rauschkriterium erfüllen, also genügend groß sind, ist die \mathbf{A} in der Regel gut konditioniert.

$$K_{Shi\&Tomas} = \min(\lambda_1, \lambda_2) \quad (5.15)$$

Eine Serie anderer Aufsätze nutzen als Grundlage für die Entwicklung eines POI Detektors die Hesse Matrix, also die partiellen zweiten Ableitungen:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \mathbf{I}}{\partial x^2} & \frac{\partial^2 \mathbf{I}}{\partial x \partial y} \\ \frac{\partial^2 \mathbf{I}}{\partial y \partial x} & \frac{\partial^2 \mathbf{I}}{\partial y^2} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{xx} & \mathbf{I}_{xy} \\ \mathbf{I}_{yx} & \mathbf{I}_{yy} \end{bmatrix} \quad (5.16)$$

Aufgrund der Kommutativität des Differenzialoperators gilt $\mathbf{I}_{xy} = \mathbf{I}_{yx}$. [?] verwendet die Determinante der Hesse Matrix als Maß für die *cornerness*.

$$K_{Beaudet} = \det(\mathbf{H}) = \mathbf{I}_{xx} * \mathbf{I}_{yy} - \mathbf{I}_{xy}^2 \quad (5.17)$$

[?] schlagen vor, Konturmaxima gleicher Intensität (*isophotes*) als POI zu verwenden. Um die Rauschneigung zu mindern, werden zusätzlich die Gradienten zur Berechnung hinzugezogen.

$$K_{Kitchen\&Rosenfeld} = \frac{\mathbf{I}_{xx} * \mathbf{I}_y^2 + \mathbf{I}_{yy} * \mathbf{I}_x^2 - 2\mathbf{I}_{xy}\mathbf{I}_x\mathbf{I}_y}{\mathbf{I}_x^2 + \mathbf{I}_y^2} \quad (5.18)$$

[?] charakterisierten die Krümmungsantwort durch Anwendung des Laplace-Operators in Verbindung mit dem Gradienten:

$$K_{Wang\&Brady} = (\Delta \mathbf{I})^2 - s * |(\nabla \mathbf{I})|^2 \quad (5.19)$$

Ein grundlegend anderer Ansatz verwendet Punkte in denen Frequenzkomponenten maximal in Phase sind. Dieses Modell wird als *local energy model* bezeichnet und wurde von [?] und [?] entwickelt. Es wurde weiters gezeigt, dass diesem Ansatz ein plausibles Modell der menschlichen Wahrnehmung zugrundeliegt [?].

In [?] wird der Standpunkt vertreten, dass das Hauptaugenmerk der traditionellen Kantendetektoren, die versuchen, sprunghafte Kanten zu finden, falsch gesetzt wurde. Bilder enthalten eine Vielzahl an Kantentypen, die damit nicht gefunden werden können. Ein großes Spektrum an kantendarstellenden Funktionen kann durch die Fourierreihe

$$s(x) = \sum_{n=0}^{\infty} \frac{1}{(2n+1)^p} \sin[(2n+1)x + \phi] \quad (5.20)$$

beschrieben werden. Wobei p die exponentielle Abnahme der Amplitude der Harmonischen und ϕ den Winkel, in der die Phasengleichheit auftritt, angibt [?].

Verschiedene Methoden wurden zur Beurteilung der Qualität eines POI Detektors vorgeschlagen. Darunter finden sich Ansätze über „Bodenwahrheiten“ (*ground truths*), visuelle Inspektion, Lokalisierungsgenauigkeit, theoretische Analyse, Reproduzierbarkeit und Informationsinhalt [?]. Ein etabliertes Maß für die Qualität eines POI Detektors ist die Reproduzierbarkeit (vorgeschlagen in [?]), die wiedergibt, wie gut ein POI unter Änderungen der Belichtungsbedingungen oder der Rotation wiedergefunden werden kann. Die Reproduzierbarkeit eines POI misst die Anzahl der gefundenen Punkte in gemeinsamen Bildabschnitten zweier Bilder im Vergleich zur Gesamtanzahl der gefundenen POIs. [?] kommen zum Schluss, dass der in [?] vorgeschlagene Operator die stabilsten Ergebnisse unter den untersuchten Operatoren liefert.

Eine Evaluierung der vorgestellten Detektoren wird anhand des Siemensstern in Abbildung 5.1 gegeben.

5.3.2 Direkte Methoden

5.3.2.1 Einführung

Bei direkten *matching* Methoden entfällt der Schritt der Merkmalsextraktion, da die Intensitätswerte des Bildes direkt zur Registrierung der Bilder herangezogen werden. Bei auf Korrelation basierenden Algorithmen wird typischerweise ein rechteckiges Gebiet (*template matrix*) aus einem Bild gewählt und durch ein Ähnlichkeitsmaß (*similarity measure*) mit dem anderen Bild (*search matrix*) verglichen. Eine Limitierung dieses Ansatzes liegt darin begründet, dass Rechtecke nur Translationsinvarianz aufweisen und sich bei komplexeren Abbildungen verformen. Dadurch ist der grundlegende Ansatz nur auf Translationsprobleme anwendbar. Durch Wahl eines runden Ausschnittes und durch Verwendung von Polarkoordinaten kann der Ansatz aber auch auf relativ zueinander rotierte Bildausschnitte ausgeweitet werden. Wenn die *template matrix* und die *search matrix* genügend Informationsgehalt aufweisen und für die Anwendung nur einfache Transformationen eine Rolle spielen, können damit jedoch sehr robuste Ergebnisse erzielt werden. Grundlegend lässt sich zwischen Methoden, die im Ortsbereich und Methoden, die im Frequenzbereich arbeiten, unterscheiden.

Ein klassischer im Ortsbereich definierter Ansatz verwendet die **Kreuzkorrelation** (*cross correlation* (CC)) als Ähnlichkeitsmaß, um zwei Bildausschnitte miteinander zu vergleichen. Die Kreuzkorrelation ist durch die ℓ_2 Norm motiviert. Für eine *search matrix* $\mathbf{I}(x, y)$ und eine in einer Region R definierten *template matrix* $\mathbf{T}(x, y)$ ist die ℓ_2 Norm folgendermaßen definiert:

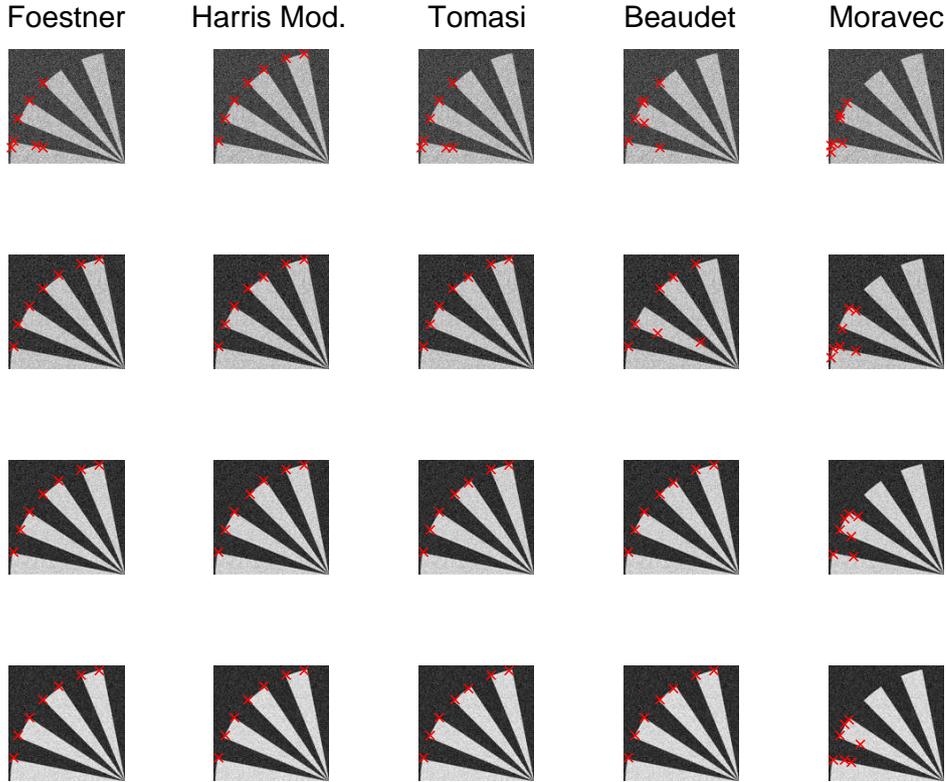


Abbildung 5.1: Extraktion von POIs am Siemensstern zur Evaluierung der vorgestellten Detektoren. Diese Abbildung zeigt die Besten sieben POIs der Detektoren *Foerstner*, *Harris mod.*, *Tomasi*, *Beaudet* und *Moravec* am Beispiel des Siemenssterns. Mit Absinken des SNR werden Unterschiede zwischen den Detektoren sichtbar. *Harris mod.* zeigte die beste Rauschunempfindlichkeit bei Beaufschlagung mit gauß'schem Rauschen. Das SNR sinkt dabei von den unteren Bildern zu den oberen Bildern.

$$\ell_2^2(u, v) = \sum_{(x,y) \in R} |\mathbf{I}(x, y) - \mathbf{T}(x - u, y - v)|^2 \quad (5.21)$$

$$= \sum_{(x,y) \in R} |\mathbf{I}^2(x, y) - 2\mathbf{I}(x, y)\mathbf{T}(x - u, y - v) + \mathbf{T}^2(x - u, y - v)| \quad (5.22)$$

Wobei der Term $\mathbf{T}^2(x - u, y - v)$ konstant ist und der Term $\mathbf{I}^2(x, y)$ als konstant angenommen wird. Da diese Terme nur einen Bias des Ergebnisses verursachen, werden diese in Folge nicht berücksichtigt. Weiteres kann der Faktor 2 weggelassen werden. Nach Streichung dieser Terme erhält man die Kreuzkorrelationsfunktion, die auch durch den Faltungsoperator \star beschrieben werden kann.

$$cc(u, v) = \sum_{(x,y) \in R} |\mathbf{I}(x, y)\mathbf{T}(x - u, y - v)| = \mathbf{I} \star \mathbf{T} \quad (5.23)$$

Durch die Darstellung des *matching* Problems als Faltung kann dieses effizient durch die schnelle Fourier-Transformation (FFT) in $O(n \log n)$ gelöst werden. Es sollte jedoch berücksichtigt werden, dass die FFT eine zyklische Faltung darstellt und deshalb Fehler verursacht. Um eine lineare Faltung mit Hilfe der FFT durchführen zu können, muss der Randbereich ausreichend durch Nullen erweitert werden (*zero padding*) oder Verfahren wie Overlap-Save oder Overlap-Add angewandt werden.

Trotz der effiziente Berechnung im Frequenzraum bleiben bei dieser Art der Korrelation eine Reihe von Nachteilen:

- Der Wertebereich von $cc(u, v)$ hängt vom Wertebereich der Intensitäten sowie von der Anzahl der Pixel der Menge R ab. Um dieser Abhängigkeit entgegenzuwirken, muss eine Normalisierung durchgeführt werden.
- Änderungen der Belichtungszeit oder der Empfindlichkeit des Aufnahmesystems gehen direkt in den Kennwert ein. Durch Abzug des über alle Pixel eines Teilbildes gerechneten Mittelwertes (*zero mean normalization*) kann diese Abhängigkeit beseitigt werden.
- Wenn die Bildenergie $\mathbf{I}^2(x, y)$ über das Bild variiert, wird dadurch das Ergebnis verfälscht.

Die normalisierte Kreuzkorrelation (*normalized cross-correlation*) γ , die als eine Standardmethode in der Bildanalyse gilt, versucht diese Nachteile durch die oben erwähnten Verbesserungen zu überwinden.

$$\gamma(u, v) = \frac{\sum_{(x,y) \in R} (\mathbf{I}(x, y) - \bar{\mathbf{I}}_{u,v})(\mathbf{T}(x - u, y - v) - \bar{\mathbf{T}})}{\sqrt{\sum_{(x,y) \in R} (\mathbf{I}(x, y) - \bar{\mathbf{I}}_{u,v})^2 (\mathbf{T}(x - u, y - v) - \bar{\mathbf{T}})^2}} \quad (5.24)$$

Da sich die Funktion durch diese Modifikationen nicht durch eine alleinige schnelle Faltung lösen lässt, wurden Verbesserungen vorgeschlagen, die in Abschnitt 5.3.2.2 beschrieben werden.

Eine oft verwendete, im Frequenzbereich ansetzende Methode stellt die **Phasenkorrelation** (*phase correlation* (PC)) dar. Grundlage dieser bildet das Verschiebungstheorem im Frequenzbereich. Das *matching* Problem für eine rein translatorische Abbildung kann für zwei Bilder \mathbf{I} und \mathbf{T} im Ortsbereich durch

$$\mathbf{T}(x, y) = \mathbf{I}(x + \Delta x, y + \Delta y) \quad (5.25)$$

beschrieben werden. Wobei die Translation $(\Delta x \ \Delta y)^T$ gesucht wird. Durch Fouriertransformation der Bilder erhält man:

$$\mathcal{F}(\mathbf{T}(x, y)) = \mathcal{F}(\mathbf{I}(x + \Delta x, y + \Delta y)) = e^{j(\omega_x \Delta x + \omega_y \Delta y)} \mathcal{F}(\mathbf{I}(x, y)) \quad (5.26)$$

Wenn \mathcal{T} die Fouriertransformierte von \mathbf{T} kennzeichnet, und \mathcal{I} die Fouriertransformierte von \mathbf{I} kann Gleichung 5.26 mit

$$\mathcal{T}(\omega_x, \omega_y) = e^{j(\omega_x \Delta x + \omega_y \Delta y)} \mathcal{I}(\omega_x, \omega_y) \quad (5.27)$$

angeschrieben werden. Das Kreuzenergiespektrum \mathcal{E} (*cross power energy spectrum*) der Signale $\mathbf{I}(x, y)$ und $\mathbf{T}(x, y)$ ist definiert durch:

$$\mathcal{E}(\omega_x, \omega_y) = \mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{T}(\omega_x, \omega_y)^* \quad (5.28)$$

Wobei der Operator \cdot das elementweise Matrizenprodukt kennzeichnet. Die Normalisierung des Kreuzenergiespektrums führt zu:

$$\frac{\mathcal{E}(\omega_x, \omega_y)}{|\mathcal{E}(\omega_x, \omega_y)|} = \frac{\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{T}(\omega_x, \omega_y)^*}{|\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{T}(\omega_x, \omega_y)^*|} \quad (5.29)$$

Durch Substitution von Gleichung 5.27 in Gleichung 5.29 bekommt man:

$$\frac{\mathcal{E}(\omega_x, \omega_y)}{|\mathcal{E}(\omega_x, \omega_y)|} = \frac{\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{I}(\omega_x, \omega_y)^* e^{-j(\omega_x \Delta x + \omega_y \Delta y)}}{|\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{I}(\omega_x, \omega_y)^*| \underbrace{|e^{-j(\omega_x \Delta x + \omega_y \Delta y)}|}_{=1}} \quad (5.30)$$

$$= \frac{\mathcal{I} \cdot (\omega_x, \omega_y) \mathcal{I}(\omega_x, \omega_y)^*}{\underbrace{|\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{I}(\omega_x, \omega_y)^*|}_{=1}} e^{-j(\omega_x \Delta x + \omega_y \Delta y)} \quad (5.31)$$

$$= e^{-j(\omega_x \Delta x + \omega_y \Delta y)} \quad (5.32)$$

Dies führt zu dem Ergebnis, dass das normalisierte Kreuzenergiespektrum, der durch die Translation verursachten Phasenverschiebung, dem Frequenzbereich entspricht:

$$e^{-j(\omega_x \Delta x + \omega_y \Delta y)} = \frac{\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{T}(\omega_x, \omega_y)^*}{|\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{T}(\omega_x, \omega_y)^*|} \quad (5.33)$$

Durch Anwendung der inversen Fouriertransformation auf Gleichung 5.33 ergibt sich als analytisches Ergebnis ein Dirac Impuls mit einer Energie von 1 an den Koordinaten der gesuchten Translation $(\Delta x \quad \Delta y)^T$:

$$\delta(x - \Delta x, y - \Delta y) = \mathcal{F}^{-1} \left(\frac{\mathcal{I} \cdot (\omega_x, \omega_y) \mathcal{T}(\omega_x, \omega_y)^*}{|\mathcal{I}(\omega_x, \omega_y) \cdot \mathcal{T}(\omega_x, \omega_y)^*|} \right) \quad (5.34)$$

Bei realen Bildern verteilt sich jedoch ein großer Teil der Energie auf das Grundrauschen. Ein spezielles Augenmerk sollte auch der Randbehandlung gelten, da die Fouriertransformation wiederum aus Performancegründen durch die zyklische FFT berechnet wird.

Einer der Vorteile dieses Ansatzes liegt in der Robustheit gegenüber Intensitätsrauschen und nicht gleichförmigen, zeitveränderlichen Beleuchtungsstörungen. Dies kann damit erklärt werden, dass nur die Phaseninformation zur Korrelationsberechnung

herangezogen wird und die Amplitude der Frequenzen, in denen sich diese Störungen widerspiegeln, durch eine Normalisierung dieser ausgeblendet wird. Desweiteren liefert die PC im Vergleich zur NCC ein steileres Maximum, welches eine genauere Lokalisierung ermöglicht.

Eine qualitative Gegenüberstellung der Koeffizienten einer NCC und PC ist in Abbildung 5.2 ersichtlich.

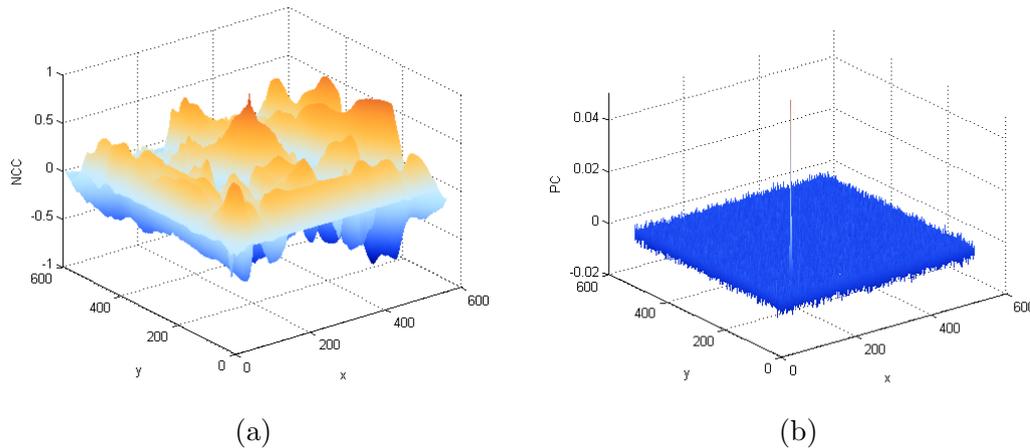


Abbildung 5.2: Korrelationskoeffizienten der NCC und PC. In Abbildung (a) sind die Korrelationskoeffizienten der NCC als Höhenmodell dargestellt. Maxima entsprechen Punkten hoher Korrelation und weisen damit hohe Ähnlichkeit zum *template* auf. Abbildung (b) zeigt die inverse Fouriertransformation des Kreuzleistungsspektrums einer PC. Das globale Maximum ist hierbei stärker ausgeprägt und ermöglicht somit eine genauere Lokalisierung.

Verbesserungsmöglichkeiten in Punkto Robustheit, Genauigkeit und Performance werden in Abschnitt 5.3.2.2 besprochen.

5.3.2.2 Stand der Technik

Direkte Methoden haben durch ihre breite Anwendung bei vielen Problemstellungen der Bildanalyse eine lange Historie und sind gut verstanden. Im Folgenden wird zwischen Korrelationsmethoden unterschieden, die im Ortsbereich definiert sind und Methoden, die sich das Verschiebungstheorem im Frequenzbereich zu Nutze machen. Diese Einteilung ist allerdings unscharf, da viele im Ortsbereich definierte Methoden aus Performancegründen auch die Frequenzdarstellung nutzen.

Die in Abschnitt 5.3.2.1 vorgestellte normalisierte Kreuzkorrelation, die ihre Wurzeln in der Signalverarbeitung hat, kann in der, durch Gleichung 5.24 dargestellten Form, nicht ausschließlich durch eine schnelle Faltung berechnet werden. In [?] wird vorgeschlagen, die Berechnung des Nenners durch Summentabellen zu beschleunigen.

In [?] wird das Template als eine gewichtete Summe von rechteckförmigen Basisfunktionen dargestellt und die Korrelation für die Basisfunktionen anstatt für das gesamte Template berechnet. Dadurch lässt sich die Berechnung des Zählers der NCC beschleunigen.

Eine weitere Laufzeitoptimierung konnte in [?] durch die Beschränkung auf Anwendungen, in denen beide Bilder die gleiche Größe aufweisen, erreicht werden.

Die Motivation des Aufsatzes von [?] liegt in der Erkenntnis, dass eine hohe Genauigkeit nur im Umfeld des Maximums notwendig ist. Gleichung 5.24 wird mittels Monte-Carlo Methoden gelöst, was jedoch das Auffinden des globalen Maximums nicht garantiert. Weiteres wird statt der rechenintensiven ℓ_2 Norm die ℓ_1 Norm (*sum of absolute differences (SAD)*) eingesetzt.

Auf Bildpyramiden basierende Ansätze starten die Suche auf einer niederen Auflösung und verfeinern die Suche nur an Positionen hoher Korrelation. Dadurch kann zwar ein großer Performancegewinn erzielt werden, jedoch besteht die Gefahr, ein Maximum zu verfehlen. Es existieren eine Reihe von Vorschlägen, um diese Nachteile durch sog. *backtracking* zu umgehen. Ein Einblick in die Geschichte dieser Klasse von Methoden kann in [?] erhalten werden.

Um die durch *zero padding* entstehenden Diskontinuitäten am Bildrand abzumindern, wird in [?] vorgeschlagen, ein Blackman Fenster [?] einzusetzen. Dies führt jedoch zu Informationsverlust, der das Auffinden von Mustern zum Rand hin erschwert. In [?] wird deshalb vorgeschlagen, das Suchbild sowie das Template durch fünf abfallend gewichtete Pixel in jeder Richtung zu erweitern. Dadurch wird die Energie im Maximum erhöht und Maxima, die durch die Diskontinuitäten entstehen, signifikant abgeschwächt.

Eine Vielzahl von Anwendungen erfordern Registrierungsgenauigkeiten die unter einem Pixel liegen müssen. Eine Möglichkeit dies zu erreichen besteht darin, die Auflösung der Bilder durch einen Interpolationsalgorithmus zu erhöhen und die nachfolgende Registrierung mit einem pixelgenauen Algorithmus durchzuführen. Durch Reskalierung auf die Originalauflösung und Division der ermittelten Translationen durch den inversen Skalierungsfaktor kann Subpixelgenauigkeit erzielt werden. Da sich dabei für eine Verdopplung der Genauigkeit der Aufwand vervierfacht, ist dieser Ansatz in Punkto Rechenaufwand sowie Speicheraufwand sehr uneffizient.

Eine alternative Methode wie in [?] vorgeschlagen, passt die Parameter einer analytischen Modellfunktion den Korrelationskoeffizienten in der Umgebung um das Maximum an. Durch Berechnung des Maximums der Funktion kann durch sehr geringen Rechenaufwand hohe Genauigkeit erzielt werden. [?] verwenden hierzu ein Paraboloid und erzielen Genauigkeiten unter 1/16 eines Pixels bei Verwendung der NCC.

Eine Reihe weiterer Arbeiten erweitern die Phasenkorrelation (PC) um Subpixelgenauigkeit [?, ?, ?]. In [?] wird eine analytische Untersuchung des Maximums durchgeführt und

festgestellt, dass dieses sehr gut durch eine zweidimensionale *sinc* Funktion approximiert werden kann [?].

Für einen Überblick über verschiedenste Registrierungsmethoden kann [?] empfohlen werden.

5.3.2.3 Subpixel-Genauigkeit

Die vorgestellten Verfahren liefern pixelgenaue Translationen. Um Genauigkeiten von einem Bruchteil eines Pixels zu erhalten, müssen die Verfahren erweitert werden. Eine sehr effiziente Möglichkeit zur Verbesserung stellt die Klasse der Methoden dar, die eine analytische Funktion in die Umgebung des Maximums des Korrelationskoeffizienten einpasst. Durch Differenzieren der Funktion und anschließendes Null setzen kann dadurch das Maximum subpixelgenau ermittelt werden. Da sich die Koeffizienten der NCC in der Regel nicht sprunghaft ändern, kann ein zweidimensionales Polynom zweiter Ordnung an die Umgebung des Maximums angepasst werden, welches die Umgebung gut beschreibt. In kartesischen Koordinaten ist diese Funktion durch

$$c_{x,y} = c(x, y) = ax^2 + by^2 + cxy + dx + ey + f \quad (5.35)$$

definiert. Die Funktion $c(x, y)$ hat genau ein Maximum, dass durch

$$\frac{\partial c(x, y)}{\partial x} = 2ax + cy + d \stackrel{!}{=} 0 \quad (5.36)$$

$$\frac{\partial c(x, y)}{\partial y} = 2by + cx + e \stackrel{!}{=} 0 \quad (5.37)$$

gefunden werden kann. In Matrizenform können die Gleichungen 5.36 und 5.37 über

$$\begin{bmatrix} 2a & c \\ c & 2b \end{bmatrix} \begin{pmatrix} x_{max} \\ y_{max} \end{pmatrix} = \begin{pmatrix} -d \\ -f \end{pmatrix} \quad (5.38)$$

$$\begin{pmatrix} x_{max} \\ y_{max} \end{pmatrix} = \begin{bmatrix} 2a & c \\ c & 2b \end{bmatrix}^{-1} \begin{pmatrix} -d \\ -f \end{pmatrix} = \frac{1}{2ab - c^2} \begin{bmatrix} 2b & -c \\ -c & 2a \end{bmatrix} \begin{pmatrix} -d \\ -f \end{pmatrix} = \begin{pmatrix} \frac{cf - 2bd}{4ab - c^2} \\ \frac{cd - 2af}{4ab - c^2} \end{pmatrix} \quad (5.39)$$

gelöst werden.

Zur Ermittlung der sechs Koeffizienten der Gleichung 5.35 wird eine 3×3 Umgebung

$$\mathbf{C} = \begin{bmatrix} c_{-1,-1} & c_{0,-1} & c_{1,-1} \\ c_{-1,0} & c_{0,0} & c_{1,0} \\ c_{-1,1} & c_{0,1} & c_{1,1} \end{bmatrix} \quad (5.40)$$

um das Maximum $c_{0,0}$ der Korrelationsfunktion gewählt.

Jeder Koeffizient aus dem Vektor $\mathbf{c} = (c_{-1,-1} \ c_{0,-1} \ c_{1,-1} \ c_{-1,0} \ c_{0,0} \ c_{1,0} \ c_{-1,1} \ c_{0,1} \ c_{1,1})^T$ liefert eine Gleichung für das Gleichungssystem [?]

$$\mathbf{Ax} = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} c_{-1,-1} \\ c_{0,-1} \\ c_{1,-1} \\ c_{-1,0} \\ c_{0,0} \\ c_{1,0} \\ c_{-1,1} \\ c_{0,1} \\ c_{1,1} \end{pmatrix} \quad (5.41)$$

Wobei der Vektor $\mathbf{x} = (a \ b \ c \ d \ e \ f)^T$ die zu ermittelten Koeffizienten der Funktion aus Gleichung 5.35 darstellt. Mit neun Gleichungen und sechs Unbekannten ist das Gleichungssystem überbestimmt und kann mit der Methode, wie in Abschnitt 5.4.3 beschrieben, gelöst werden. In Abbildung 5.3 sind die Koeffizienten nach der Verfeinerung dargestellt.

Die Annahme, dass die Korrelationskoeffizienten in der Umgebung des Maximums einem Polynom zweiter Ordnung folgen, ist willkürlich und begrenzt damit die erreichbare Genauigkeit im Mittel auf ungefähr $1/18 \text{ Pixel}$ (siehe Abbildung 5.11).

Alternative Aufsätze wie etwa in [?, ?] versprechen höhere Genauigkeiten bei moderatem Mehraufwand. Diese wurden in dieser Arbeit jedoch nicht näher betrachtet und bieten daher Raum für Verbesserungen.

5.3.3 Gewählte Verfahren

Für optische Mikroskopbilder wurde auf Grund der Ergebnisse aus [?] und nach Versuchen mit verschiedenen Detektoren der Ansatz von [?] als Basis gewählt. Um die Rauschempfindlichkeit zu verringern, wurden die Ableitungen der Matrix \mathbf{A} durch einen abgeleiteten Gauß-Kern ermittelt. Um Mehrfacherkennungen zu vermeiden, wurde *non maximum suppression* auf das Ergebnis des Detektors angewandt.

Für Abbildungen des SEM wurde ein direkter Ansatz gewählt, da das Vorhandensein von POI oft nicht gegeben ist. Die Wahl fiel hierbei auf die NCC, mit der in [?] vorgeschlagenen Optimierung zur effizienten Berechnung des Nenners. Der Zähler wurde durch eine schnelle Faltung im Frequenzbereich berechnet. Um die für die Problemstellung notwendige Subpixelgenauigkeit zu erreichen, wurde das Maximum der Korrelation, wie in [?] beschrieben, durch Einpassen eines Polynoms zweiter Ordnung mit zwei Unbekannten verfeinert. Zahlreiche Versuche zeigten, dass die NCC im Mittel robustere Ergebnisse lie-

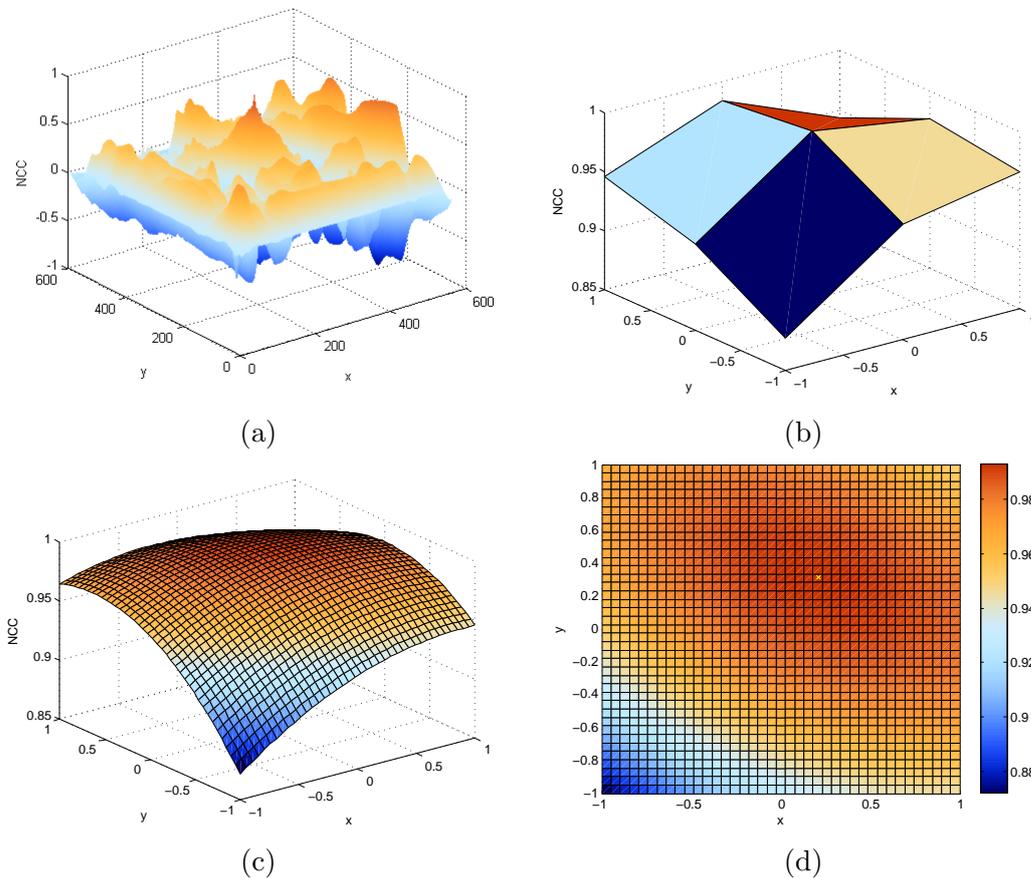


Abbildung 5.3: Subpixelgenauigkeit durch Regression. Bild (a) zeigt die normalisierten Korrelationskoeffizienten einer Bildregistrierung. In etwa mittig befindet sich das globale Maximum, welches die Position der größten Ähnlichkeit widerspiegelt. In Bild (b) wurde nur die 3×3 Umgebung um das globale Maximum dargestellt. Allgemein wird das Maximum nicht pixelgenau zu liegen kommen, sondern innerhalb der 3×3 Umgebung. Durch Modellierung der Umgebung des Maximums durch eine analytische Funktion lässt sich das Maximum subpixelgenau bestimmen. Man sollte jedoch beachten, dass die Form der Funktion (Polynom zweiter Ordnung mit zwei Unbekannten) willkürlich gewählt wurde und damit die Genauigkeit auf in etwa $1/18$ Pixel beschränkt. Bild (c) zeigt die Funktion nach Schätzung der sechs Parameter. Bild (d) stellt die Funktion von oben dar. Die Verfeinerung ergibt einen Offset von $(0.22 \ 0.31)^T$ Pixel vom ursprünglich pixelgenauen Maximum.

ferte als die PC, bei welcher bei sehr stark verrauschten Bildern oftmals kein eindeutiges Maximum zu finden war.

5.4 Koordinatentransformationen

Um eine präzise Formulierung zu gewährleisten, werden die in dieser Arbeit verwendeten Koordinatensysteme wie folgt definiert:

- *Lokale Bildkoordinaten.* Jedes Teilbild einer Chipebene hat sein eigenes *lokales Bildkoordinatensystem*. Wie in der Bildverarbeitung üblich, befindet sich der Koordinatenursprung in der linken oberen Ecke. Im Gegensatz zu der in der Mathematik üblichen Notation, befindet sich die positive y -Halbebene unter dem Koordinatenursprung.
- *Chipebenenkoordinaten.* Eine Ebene eines Chips besitzt ihr eigenes Koordinatensystem. Im ersten Schritt besteht die Notwendigkeit, Transformationen vom lokalen Bildkoordinatensystem in das *Chipebenenkoordinatensystem* zu finden, um aus den einzelnen Chipebenen ein über die Ebenen konsistentes dreidimensionales Mosaik zu bilden.
- *Globale Chipkoordinaten.* Das dem Chipebenenkoordinatensystem übergeordnete *globale Chipkoordinatensystem* bildet die Basis für Analysen des gesamten Mikrochips. Dabei wird ein Chipebenenkoordinatensystem als Referenzsystem definiert.

In Abbildung 5.4 werden die in dieser Arbeit verwendeten Koordinatensysteme und die überführenden Transformationen dargestellt.

Im Folgenden wird auf die Struktur der Transformationen eingegangen, die eine Überführung zwischen den oben definierten Koordinatensystemen ermöglichen.

5.4.1 Kalibrierung and Kameramodell

Nichtlinearitäten in der Abbildung durch optische oder elektromagnetische Systeme, die durch lichtideale Linsen (Linsenverzeichnungen) verursacht werden, erfordern eine Korrektur durch ein für die Anwendung ausreichend genaues Fehlermodell. Die Modellparameter können durch Abbildung von exakt bekannten Strukturen ermittelt werden. In der optischen Bilderfassung werden deshalb oft sog. *targets* wie Schachbrettmuster, die definierte mit hoher Genauigkeit zu detektierende Ecken enthalten, aus unterschiedlichen Blickwinkeln und Abständen abgebildet und daraus die intrinsischen Kamera-Parameter wie Brennweite (*focal length*), Bildmittelpunkt (*principal point*), Schiefe (*skew*) und Parameter für rotationssymmetrische Abbildungsfehler abgeleitet.

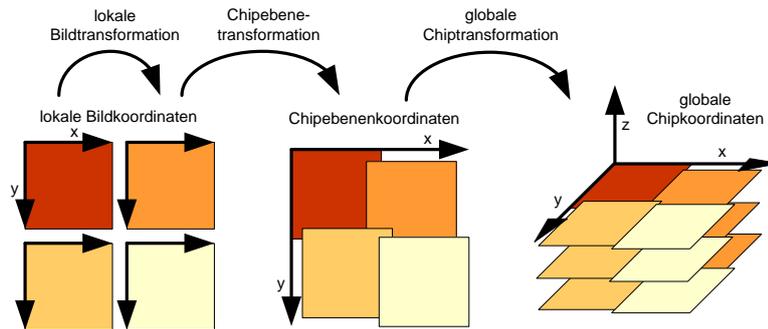


Abbildung 5.4: Koordinatensysteme und überführende Transformationen. Jedes Bild eines Chipecbenenausschnittes hat sein eigenes lokales Bildkoordinatensystem. Durch Verkettung der mittels Bild zu Bild Registrierung gefundenen lokalen Bildtransformationen lässt sich für jedes Bild eine Chipecbene-Transformation berechnen, die die lokalen Bildkoordinaten auf die jeweilige Chipecbene überführt. Durch Finden von Korrespondenzen zwischen den Ebenen wird für jede Ebene eine globale Chiptransformation ermittelt, mittels derer die Chipecbenenkoordinaten auf die globalen Chipkoordinaten abgebildet werden.

Ein oft verwendetes Kameramodell, welches die Abbildung auf die Bildebene beschreibt, unterteilt die Abbildungsparameter in intrinsische und extrinsische. Die intrinsischen Parameter werden durch eine im Vorfeld durchgeführte Kalibrierung ermittelt, die externen ergeben sich durch die Anordnung der Kamera zu dem abzubildenden Motiv. Für eine sehr gute Herleitung dieser Zusammenhänge kann [?] empfohlen werden. Da die Bilder für diese Arbeit auf einem bereits kalibrierten System aufgenommen wurden, beschränkt sich der folgende Teil, ausschließlich auf die extrinsischen Kameraparameter.

5.4.2 Struktur der Transformation

Die extrinsischen Parameter des Kamerasystems definieren die Position und Ausrichtung der Kamera relativ zum Weltkoordinatensystem. Nach einer vorhergehenden Entzerrung der Abbildungsfehler durch eine Kalibrierung kann die extrinsische Orientierung mathematisch durch die isomorphe Transformationsmatrix \mathbf{T} beschrieben werden. In homogenen Koordinaten kann die Transformation wie folgt dargestellt werden:

$$\begin{pmatrix} x' \\ y' \\ h \end{pmatrix} = \mathbf{T} * \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (5.42)$$

Dabei stellt der Vektor $\begin{pmatrix} x & y & 1 \end{pmatrix}^T$ den zweidimensionalen Punkt vor der Transformation und der Vektor $\begin{pmatrix} x' & y' & h \end{pmatrix}^T$ den transformierten Punkt dar. Abhängig von der Anzahl

der Freiheitsgrade (*degrees of freedom* (DOF)) wird zwischen unterschiedlichen Typen von Transformationen unterschieden. In Abbildung 5.5 sind die planaren zweidimensionalen Transformationstypen dargestellt. Jeder Typ besitzt geometrische Erhaltungsgrößen, die in Tabelle 5.1 aufgelistet sind. Transformationen mit n DOF sind jeweils Verallgemeinerungen von Transformationen mit $< n$ DOF. Diese hierarchische Natur spiegelt sich auch in den geometrischen Erhaltungsgrößen wider (siehe Tabelle 5.1). Für eine ausführliche Darstellung der Transformationshierarchien und deren Eigenschaften kann auf [?] verwiesen werden.

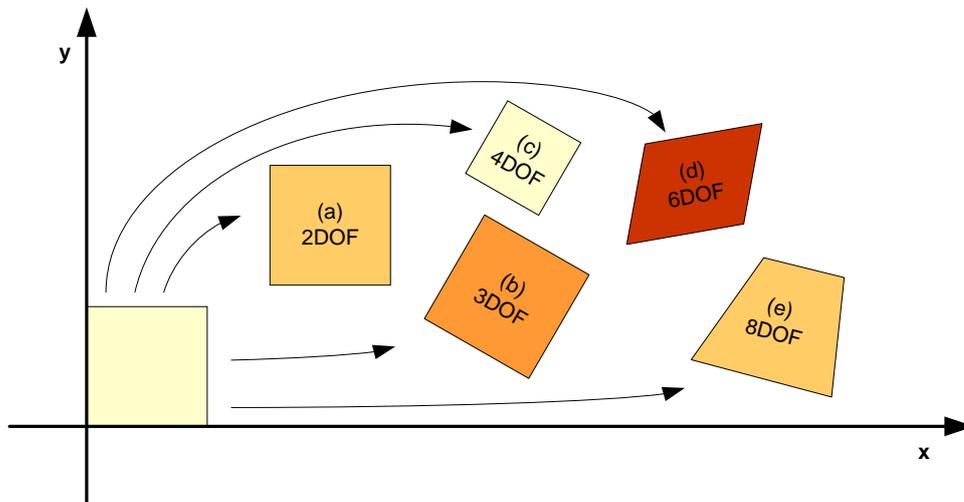


Abbildung 5.5: Verschiedene Typen von planaren 2D Transformationen in Abhängigkeit der Freiheitsgrade. Unterschieden wird mit steigender Anzahl an Freiheitsgraden zwischen reiner Translation (a), rigider (euklidischer) Transformation (b), Ähnlichkeitstransformation (c), affiner Transformation (d) und projektiver Transformation (e).

Die Verwendung eines hochgenauen verstellbaren Probenstisches ermöglicht es, die gesuchte Transformation für optische sowie SEM Bilder auf eine einfache Translation zu begrenzen. Für dieses Problem mit zwei Freiheitsgraden genügt zur exakten Lösung der Parameter $\begin{pmatrix} t_x & t_y \end{pmatrix}^T$ genau ein korrespondierendes Punktepaar. Wenn mehrere Punktepaare zur Verfügung stehen, kann die Robustheit durch Methoden der robusten Statistik verbessert werden. Desweiteren kann dadurch die Bestimmung der Translation mit höherer Genauigkeit erfolgen.

Da ein Mikrochip mehrlagig aufgebaut ist und die Probe zum Abtragen einer Schicht aus dem Mikroskop entfernt werden muss, bedarf es dennoch komplexerer Transformationen, um die Zusammenhänge zwischen den Lagen zu beschreiben. Versuche zeigten, dass eine affine Transformation mit 6 Freiheitsgraden diese Relationen am Besten wiedergibt.

Eine Transformation, die einzig durch die *Translation* bestimmt ist, hat in homogenen

Transformation	Matrix \mathbf{T}	DOFs	Erhaltungsgrößen	Symbol
Translation	$\begin{bmatrix} \mathbf{I} & t \\ \mathbf{0}^T & 1 \end{bmatrix}$	2	Orientierung, Längen, Winkel, ...	
Rigide	$\begin{bmatrix} \mathbf{R} & t \\ \mathbf{0}^T & 1 \end{bmatrix}$	3	Längen, Winkel, ...	
Ähnlichkeit	$\begin{bmatrix} s\mathbf{R} & t \\ \mathbf{0}^T & 1 \end{bmatrix}$	4	Winkel, Kollinearität, ...	
Affine	$\begin{bmatrix} \mathbf{A} & t \\ \mathbf{0}^T & 1 \end{bmatrix}$	6	Kollinearität, Verhältnis von Distanzen	
Projektiv	$\begin{bmatrix} \mathbf{H} \end{bmatrix}$	8	Geraden	

Tabelle 5.1: Hierarchie von planaren 2D Transformationen. Die verschiedenen Transformationsgruppen sind ineinander verschachtelt. Komplexere Transformationen stellen Verallgemeinerungen von einfacheren dar.

Koordinaten folgende Form:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + t_x \\ y + t_y \\ 1 \end{pmatrix} \quad (5.43)$$

Nach Umformung können die Parameter mit Hilfe von korrespondierenden Punktepaaren ermittelt werden.

$$\begin{pmatrix} x' - x \\ y' - y \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (5.44)$$

Jedes Punktepaar liefert hierbei zwei Zeilen der Matrix. Da das Gleichungssystem für mehr als ein Punktepaar überbestimmt ist und damit allgemein keine eindeutige Lösung besitzt, muss die Lösung durch eine Näherungsmethode bestimmt werden.

Eine *rigide* Transformation ist definiert durch die Translation $\mathbf{t} = (t_x \ t_y)^T$, die Rotation ϕ sowie der Skalierung λ und besitzt damit 4 Freiheitsgrade:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \lambda\mathbf{R}(\phi) & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} \lambda \cos \phi & \lambda \sin \phi & t_x \\ -\lambda \sin \phi & \lambda \cos \phi & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (5.45)$$

$$= \begin{bmatrix} a & b & c \\ -b & a & d \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax + by + c \\ -bx + ay + d \\ 1 \end{pmatrix} \quad (5.46)$$

Dabei kann die Skalierung durch $\lambda = \sqrt{a^2 + b^2}$ und der Winkel durch $\phi = \arctan(\frac{b}{a})$ berechnet werden. Analog wie im Falle von Gleichung 5.43 kann das Gleichungssystem

umgeformt werden, sodass die gesuchten Parameter ermittelt werden können:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} x & y & 1 & 0 \\ y & -x & 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \quad (5.47)$$

Die *affine* Transformation berücksichtigt neben den Parametern der rigiden Transformation die Scherung m . Daraus ergeben sich in Summe sechs DOF:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \lambda \cos \phi & m\lambda \cos \phi + \lambda \sin \phi & t_x \\ -\lambda \sin \phi & \lambda \cos \phi - m\lambda \sin \phi & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (5.48)$$

$$= \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax + by + c \\ dx + ey + f \\ 1 \end{pmatrix} \quad (5.49)$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} \quad (5.50)$$

5.4.3 Abschätzung der Chi Ebenentransformation

Mit der lokalen Bild zu Bild Registrierung wurden lokale Bildtransformationen zwischen benachbarten Bildkoordinatensystemen ermittelt. Um Chi Ebenenkoordinaten zu erhalten, müssen die lokalen Bildtransformationen geeignet kombiniert werden. Im einfachsten Fall können diese reihen- bzw. spaltenweise verknüpft werden. Dabei dient ein festgelegtes lokales Bildkoordinatensystem als Anker und definiert damit den Koordinatenursprung des Chi Ebenenkoordinatensystems. Dieses Verfahren wird in Abbildung 5.6 skizziert.

Ein Nachteil dieser Methode liegt darin, dass nicht alle ermittelten lokalen Bildtransformationen berücksichtigt werden können und damit suboptimale Ergebnisse erzielt werden. Vor allem bei großen Bilderserien führt dieser naive Ansatz zur Fehlerfortpflanzung, welche ein akkurates *stitchen* unmöglich macht.

Um die immer vorhandene Fehlerfortpflanzung zu minimieren, wurden eine Reihe von Methoden entwickelt, die versuchen, alle ermittelten und als gut befundenen lokalen Bildtransformationen in die Berechnung einfließen zu lassen.

- Ein graphenbasierender Ansatz wie in [?], beschrieben versucht die Fortpflanzungsfehler durch Finden eines minimalen Spannbaumes, der alle Bildausschnitte einer

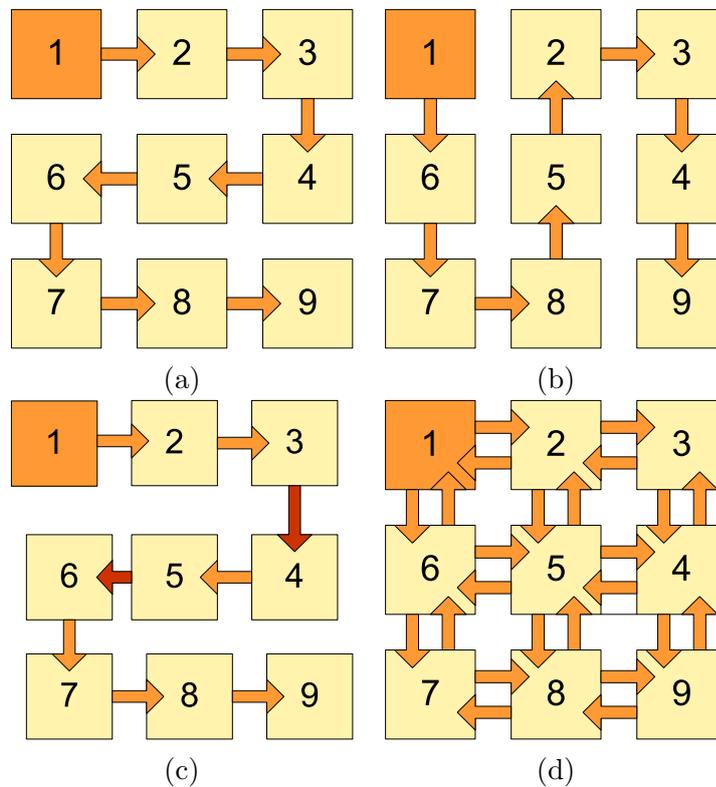


Abbildung 5.6: Verknüpfung lokaler Bildtransformationen und daraus resultierende Fortpflanzungsfehler. Die Berechnung der Chi Ebenenkoordinaten erfolgt auf Basis der lokalen Bildtransformationen. Eine naive reihenweise Verknüpfung wie in (a) oder spaltenweise Verknüpfung wie in (b) führt dabei zu nicht vermeidbaren Fortpflanzungsfehlern. Abbildung (c) skizziert diese Problematik am Beispiel (a). Tritt bei der Registrierung von Bild 3 und Bild 4 ein Fehler auf, so entstehen Unstetigkeiten zwischen der ersten und zweiten Spalte. Ähnliches gilt bei einem Fehler zwischen Bild 5 und 6. Eine Möglichkeit dieses Problem zu vermeiden zeigt Abbildung (d). Hierbei finden alle ermittelten Transformationen zur Berechnung der Chi Ebenenkoordinaten Verwendung.

Ebene verbindet, zu verringern. Hierbei bildet ein mit Bildregistrierungsfehlern gewichteter Graph die Grundlage des Algorithmus. In der Implementation wird der Spannggraph durch den von Kruskal vorgeschlagenen Algorithmus [?] ermittelt. Ein Teil des Graphen wird in Abbildung 5.7 dargestellt. Die Gewichte wurden aus den normalisierten Kreuzkorrelationskoeffizienten γ der NCC abgeleitet. Bei der Phasenkorrelation (PC) kann die Energie des Maximums herangezogen werden. Bei POI basierenden Ansätzen der mittlere Rückprojektionsfehler (Residuum).

- Das Problem kann durch ein lineares Gleichungssystem formuliert werden, welches die lokalen Bildtransformationen einer Ebene beschreibt. Einer der Vorteile dieses Ansatzes liegt in der Miteinbeziehung von mehr Bildtransformationen als für die exakte Lösung notwendig wäre.

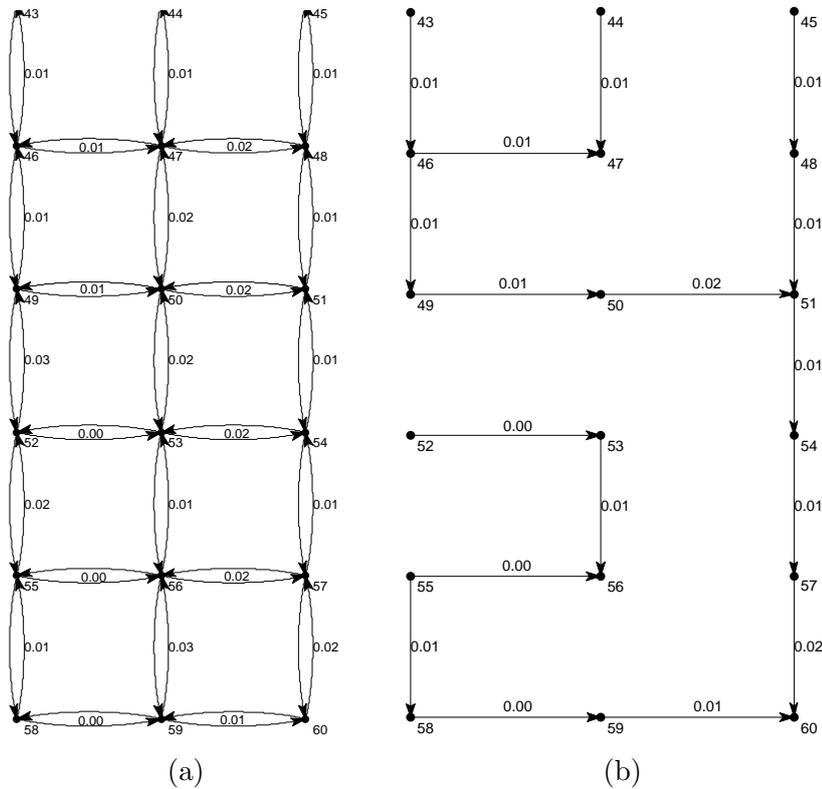


Abbildung 5.7: Minimaler Spannbaum zur Abschätzung der Chipebenentransformation. Die Fortpflanzungsfehler können durch Finden eines minimalen Spannbaumes, der alle Bildausschnitte einer Ebene verbindet, verringert werden. Die Knoten des Baumes repräsentieren die einzelnen Bilder und die Kanten die paarweise ermittelten geometrischen Relationen. Als Kantengewicht wird ein Fehlermaß der Registrierungsqualität herangezogen. Bild (a) zeigt einen Ausschnitt eines Graphen. Neben jedem Knoten ist der Bild-Index angeführt, neben jeder Kante das Fehlermaß $(1 - \gamma)$ der paarweisen Registrierung. Bild (b) zeigt einen Ausschnitt des ermittelten minimalen Spannbaumes, welcher die Summe der Registrierungsfehler minimiert.

Für die Ebene i mit N Bildern und M Transformationen ($M \approx 2N$) liefert dies die überbestimmten linearen Gleichungssysteme $\mathbf{t}_x^{(i)} = \mathbf{B}_x^{(i)} \mathbf{x}^{(i)}$ sowie $\mathbf{t}_y^{(i)} = \mathbf{B}_y^{(i)} \mathbf{y}^{(i)}$ welche durch Ausgleichsrechnung gelöst werden können. Dabei stellen die Vektoren $\mathbf{t}_x^{(i)}$ und $\mathbf{t}_y^{(i)}$ die lokalen Bildtransformationen der Ebene i in x und y Richtung dar. Die unbekannt Vektoren $\mathbf{x}^{(i)}$ und $\mathbf{y}^{(i)}$ die x und y die Positionen der Bilder im Chipebenenkoordinatensystem. Die Matrizen $\mathbf{B}_x^{(i)}, \mathbf{B}_y^{(i)} \in M(N \times M, \mathbb{R})$ besitzen Rang $N - 1$ und sind damit bis auf eine Variable bestimmt. Um vollen Rang zu erhalten muss noch die Position eines Bildes fixiert werden. Dazu wird $x_0^{(i)} = 0$ sowie $y_0^{(i)} = 0$ gewählt und damit der Koordinatenursprung in Bild 0 gelegt. Dieser Zusammenhang kann durch die Matrizen $\mathbf{A}_x^{(i)}, \mathbf{A}_y^{(i)} \in M(N \times M + 1, \mathbb{R})$ dargestellt werden:

$$\begin{pmatrix} 0 \\ t_{x,0}^{(i)} \\ t_{x,1}^{(i)} \\ \dots \\ t_{x,M-1}^{(i)} \\ t_{x,M}^{(i)} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \begin{pmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \dots \\ x_{N-1}^{(i)} \\ x_N^{(i)} \end{pmatrix} = \mathbf{A}^{(i)} \mathbf{x}^{(i)} \quad (5.51)$$

Analog lässt sich auch für die y -Koordinaten ein Gleichungssystem konstruieren.

Zur Lösung des überbestimmten Gleichungssystems kann die gauß'sche Normalgleichung verwendet werden, der die Methode der kleinsten Fehlerquadrate zugrunde liegt. Gesucht ist die Lösung $\hat{\mathbf{x}}$ die folgende Forderung erfüllt:

$$\min \|\mathbf{t} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 \Leftrightarrow \forall \mathbf{x} \|\mathbf{t} - \mathbf{A}\hat{\mathbf{x}}\|_2^2 \leq \|\mathbf{t} - \mathbf{A}\mathbf{x}\|_2^2 \quad (5.52)$$

Zur Bestimmung von $\hat{\mathbf{x}}$ kann folgendermaßen vorgegangen werden:

$$\mathbf{t} = \mathbf{A}\hat{\mathbf{x}} \quad (5.53)$$

Multiplikation mit \mathbf{A}^T liefert eine invertierbare quadratische $(M+1) \times (M+1)$ Matrix.

$$\mathbf{A}^T \mathbf{t} = \mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} \quad (5.54)$$

Lösen nach $\hat{\mathbf{x}}$ liefert:

$$(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{t} = (\mathbf{A}^T \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{A}) \hat{\mathbf{x}} = \hat{\mathbf{x}} \quad (5.55)$$

Die Matrix $\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ wird als *Pseudoinverse* bezeichnet. Mit diesem Wissen lassen sich für die Ebene i die Chi Ebenenkoordinaten $\left(x_j^{(i)} \quad y_j^{(i)}\right)^T$ der Bilder $j = 1 \dots N$ durch Lösen der Gleichungen

$$\mathbf{x}^{(i)} = \mathbf{A}_x^{(i)\dagger} \mathbf{t}_x^{(i)} \quad (5.56)$$

$$\mathbf{y}^{(i)} = \mathbf{A}_y^{(i)\dagger} \mathbf{t}_y^{(i)} \quad (5.57)$$

bestimmen.

Da die Matrix $\mathbf{A}^T \mathbf{A}$ sehr groß werden kann, jedoch sehr dünn besetzt ist, kann diese effizient gespeichert und durch spezielle Methoden der linearen Algebra invertiert werden. Um numerische Stabilität zu gewährleisten, sollte die Pseudoinverse über die Singulärwertzerlegung berechnet werden.

Eine weitere Verbesserung der Ergebnisse kann durch die Gewichtung des Gleichungssystems 5.53 erzielt werden. Die Gewichtungsfaktoren geben die Signifikanz

der einzelnen Gleichungen wieder und fließen dadurch direkt in die Summe der Fehlerquadrate ein. Die Verwendung eines höheren Gewichtes bedeutet damit, dass die jeweilige Gleichung genauer angenommen wird, als eine Gleichung mit niedrigerem Gewicht. Bei direkten Registrierungsmethoden können die normierten Korrelationskoeffizienten (bzw. die Spektralenergien) als Gewichtungsfaktoren verwendet werden. Formal kann dieser Zusammenhang durch eine Multiplikation des Gleichungssystems 5.53 mit der Diagonalmatrix \mathbf{W} dargestellt werden.

$$\mathbf{W}\mathbf{t} = \mathbf{W}\mathbf{A}\hat{\mathbf{x}} \quad (5.58)$$

- Eine Formulierung als nichtlineares Optimierungsproblem wurde als dritte Möglichkeit evaluiert. Die durch eine reihenweise bzw. spaltenweise Verkettung entstehenden Abweichungen (Fehlerfortpflanzung) zu den lokalen Bildtransformationen können durch Minimierung einer Fehlerfunktion verringert werden. Mit dem Vektor der Fehlerfunktionen $\mathbf{e}(\mathbf{X})$ und der $2 \times n$ Matrix \mathbf{X} der Chipebenenkoordinaten kann das Problem für N lokale Transformation folgendermaßen formuliert werden:

$$\min \|\mathbf{e}(\mathbf{X})\|_2^2 = \min (e_1(\mathbf{X})^2 + e_2(\mathbf{X})^2, \dots, e_N(\mathbf{X})^2) \quad (5.59)$$

Die zu minimierende Fehlerfunktion e_i berechnet sich aus der Differenz der zu optimierenden Chipebenenkoordinaten und den durch die Bildregistrierung gewonnenen lokalen Transformationen $\mathbf{t}_{i \rightarrow j}$.

$$e_i(\mathbf{X}) = \|(\mathbf{x}_j - \mathbf{x}_k) - \mathbf{t}_{i,(\mathbf{I}_k \rightarrow \mathbf{I}_j)}\|^2 \quad i = 1 \dots N \quad (5.60)$$

Die initialen Startkoordinaten \mathbf{X}_0 können entweder durch simple Verkettung wie oben beschrieben oder durch lösen des linearen Gleichungssystems berechnet werden.

Eine Analyse der Balancierung der Transformationen und den dadurch optimierten Fehlerverteilungen kann in Abschnitt 5.5.2 gefunden werden.

5.4.4 Abschätzung der globalen Chiptransformation

Nach dem planaren *stitching* der Chipebenen ist es weiters notwendig, die Ebenen dreidimensional auszurichten. Hierzu wird eine Ebene als Referenzebene definiert und die übrigen durch Auffinden von Korrespondenz zur jeweiligen Nachbarebene ausgerichtet. Elektrische Verbindungen zwischen den Chipebenen werden in Mikrochips durch s.g. VIAs (*Vertical Interconnect Access*) realisiert. Typischerweise sind VIAs aus Wolfram und können unter der Voraussetzung einer präzisen Präparation im Elektronenmikroskop mit hohem Kontrast abgebildet werden. Ein robustes Auffinden dieser Verbindungen durch den Einsatz der Bildanalyse ist dennoch eine Herausforderung.

In Abschnitt 6.4 wird auf diesen Punkt näher eingegangen. Auch mit dem Wissen der VIA Positionen ist es schwierig die korrespondierenden Punkte auf den damit verbundenen Ebenen zu finden. Die Enden von Leiterbahnen von einer, dem VIA darüberliegenden Ebene können als Kandidaten für ein Punktmatching verwendet werden. Da jedoch hohe Genauigkeit gefordert ist, wird ein manueller Ansatz durch optische Inspektion gewählt.

Die Schwierigkeit Punktpaare mit Sicherheit aufzufinden, zeigt sich überdies auch bei der manuellen Methode. Nach diversen Versuchen mit verschiedenen Transformationstypen hat sich die affine Transformation (Gleichung 5.49) als die akkurateste herausgestellt, um zwei benachbarte Lagen zueinander auszurichten. Zur exakten Lösung der sechs Freiheitsgrade werden 3 Punktpaare benötigt. Um die geforderte Robustheit und damit eine präzise Ausrichtung zu erreichen, werden deutlich mehr Punktpaare aus gleichmäßig über den Chip verteilten Bereichen zur Berechnung hinzugezogen. Jedes Punktpaar liefert 2 Zeilen der Matrix in Gleichung 5.50. Das hochgradig überbestimmte Gleichungssystem kann wie in Abschnitt 5.4.3 beschrieben durch Regression gelöst werden. Da jedoch die Methode der kleinsten Quadrate sehr sensitiv auf Ausreißer ist, wurden Methoden aus der robusten Statistik verwendet.

Eine Möglichkeit besteht im Einsatz des (*Random Sample Consensus*) Algorithmus [?], der durch zufällige Stichproben aus dem Punktpaare-Pool das Model (in diesem Fall die affine Transformation) exakt löst und darauffolgend ein Fehlermaß wie z.B. die euklidische Distanz zu den restlichen Punkten im Pool bestimmt. Wenn dieses Maß unter einem festgelegten Schwellwert ϵ liegt, gilt das Model als gut. Diese Methode wurde durch Anwendungen in der Bildanalyse motiviert und kann mit Daten, die eine hohe Anzahl an Ausreißern aufweisen, umgehen. Als Erweiterung können die Parameter der Transformation mit Punktpaaren der besten n Stichproben durch die Methode der kleinsten Quadrate bestimmt werden.

Es ist oft unnötig und unpraktikabel zur jeder möglichen Samplekonfiguration die jeweilige Transformation und die dazugehörigen Residuen zu berechnen. Es ist hinreichend, die Anzahl der benötigten Samples N genügend hoch zu wählen, sodass mit einer Wahrscheinlichkeit p zumindest eine zufällige Stichprobe mit s Punkten keinen *outlier* enthält. Üblicherweise wird $p = 0.99$ gewählt. Angenommen, w ist die Wahrscheinlichkeit, dass ein Punkt ein *inlier* ist, dann entspricht $\epsilon = 1 - w$ der Wahrscheinlichkeit, dass ein Punkt ein *outlier* ist [?]. Dann werden zumindest N Stichproben benötigt damit gilt:

$$(1 - w^s)^N = 1 - p \Rightarrow N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)} \quad (5.61)$$

Ist w bzw. ϵ nicht bekannt, dann kann ein adaptiver Algorithmus zur Bestimmung von N verwendet werden. Ausgangspunkt stellt dabei eine *worst case* Abschätzung der Wahrscheinlichkeit w dar, die bei Vorhandensein größerer Konsistenz verbessert wird.

Stellt die ermittelte Transformation $\mathbf{T}_{i \rightarrow j}$ die globale Chiptransformation von Ebene i auf Ebene j dar, können die zur Ebene j passenden Pixelposition $\mathbf{x}^{(i)}$ der

Ebene i durch

$$\mathbf{x}'^{(i)} = \mathbf{T}_{i \rightarrow j} \mathbf{x}^{(i)} \quad (5.62)$$

ermittelt werden.

Eine exakte Ausrichtung zweier Ebenen setzt allerdings voraus, dass die Berechnung der Chi Ebenenkoordinaten keine lichtlinearen Verzerrungen innerhalb der Ebenen verursacht hat. Es scheint einleuchtend, dass diese Voraussetzung aufgrund der enormen Anzahl an Bildern und deren teilweise mangelnden Bildqualität nicht garantiert werden kann. Um dennoch konsistente Verhältnisse über mehrere Lagen zu erreichen und damit eine Schalungsanalyse zu ermöglichen, wurde ein Konzept erdacht, welches sich dieser Problematik annimmt.

In diesem Ansatz werden den manuell definierten Korrespondenzen zwischen benachbarten Lagen besondere Affinität zugewiesen. Diese nun im folgenden als *Ankerpunktpaare* bezeichneten Korrespondenzen gewährleisten durch besondere Berücksichtigung eine möglichst präzise Ausrichtung über alle Lagen hinweg. Der mathematische Ansatz unterscheidet sich grundlegend von der oben ausgeführten Methode, da diese auf den fixen, bereits durch das Gleichungssystem 5.58 ermittelten Chi Ebenenkoordinaten aufbaut. Um Ankerpunktpaare zu berücksichtigen müssen diese bereits bei der Lösung des Gleichungssystems berücksichtigt werden. Da es sich bei dem Gleichungssystem jedoch nur um translatorische Abbildungen handelt, die Lagen jedoch auch relativ zueinander rotiert abgebildet werden könnten, muss dies gesondert berücksichtigt werden. Um die Darstellung jedoch zu vereinfachen, wird die Rotation zunächst mit Null angenommen und nachfolgend verallgemeinert. Ausgangspunkt stellt das Gleichungssystem 5.63 welches für die x Koordinaten der Lage i wie folgt aufgebaut ist:

$$\mathbf{W}^{(i)} \mathbf{t}_x^{(i)} = \mathbf{A}^{(i)} \mathbf{W}^{(i)} \mathbf{x}^{(i)} \quad (5.63)$$

$$\begin{pmatrix} 0 \\ t_{x,0}^{(i)} w_0^{(i)} \\ t_{x,1}^{(i)} w_1^{(i)} \\ \dots \\ t_{x,M-1}^{(i)} w_{M-1}^{(i)} \\ t_{x,M}^{(i)} w_M^{(i)} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -w_0^{(i)} & w_0^{(i)} & 0 & \dots & 0 & 0 & 0 \\ 0 & -w_1^{(i)} & w_1^{(i)} & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -w_{M-1}^{(i)} & w_{M-1}^{(i)} & 0 \\ 0 & 0 & 0 & \dots & 0 & -w_M^{(i)} & w_M^{(i)} \end{bmatrix} \begin{pmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \dots \\ x_{N-1}^{(i)} \\ x_N^{(i)} \end{pmatrix} \quad (5.64)$$

Analog kann für die Ebene j ein Gleichungssystem aufgestellt werden:

$$\mathbf{W}^{(j)} \mathbf{t}_x^{(j)} = \mathbf{A}^{(j)} \mathbf{W}^{(j)} \mathbf{x}^{(j)} \quad (5.65)$$

Die beiden Gleichungssysteme sind ohne Einführung der Ankerpunktpaare voneinander unabhängig, können aber trotzdem in einem Gleichungssystem gelöst werden.

$$\begin{pmatrix} \mathbf{W}^{(i)} \mathbf{t}_x^{(i)} \\ \mathbf{W}^{(j)} \mathbf{t}_x^{(j)} \end{pmatrix} = \begin{bmatrix} \mathbf{A}^{(i)} \mathbf{W}^{(i)} \\ \mathbf{A}^{(j)} \mathbf{W}^{(j)} \end{bmatrix} \begin{pmatrix} \mathbf{x}^{(i)} \\ \mathbf{x}^{(j)} \end{pmatrix} \quad (5.66)$$

Um die Lagen nun in Beziehung zu setzen, werden K Ankerpunkte $\mathbf{b}_k^{(i \rightarrow j)}$, $k = 0 \dots K - 1$ zwischen den Lagen i und j eingeführt. Diese sind von den Koordinaten der jeweiligen Lagen $\mathbf{x}^{(i)}$ und $\mathbf{x}^{(j)}$ bzw. $\mathbf{y}^{(i)}$ und $\mathbf{y}^{(j)}$ abhängig. Als Ankerpunkte können die RANSAC-*inlier* der Punktkorrespondenzen zwischen den Lagen verwendet werden. Als maximales Residuum, das *inlier* von *outlier* unterscheidet, wird der maximal zu erwartende Fehler gewählt. Dieser kann durchaus höher gewählt werden als die Hälfte der minimalen Strukturbreite, da der Fehler durch Einflüsse wie fehlerhafte Registrierungen und Fehlerfortpflanzungen vergrößert wird. Ziel dieses Schrittes ist das Auffinden von falschen, aber als richtig angenommenen Korrespondenzen.

Für einen Ankerpunkt $\mathbf{b}_k^{(i \rightarrow j)}$ gilt:

$$\mathbf{b}_k^{(i \rightarrow j)} = \begin{pmatrix} x_s^{(i)} & y_s^{(i)} \end{pmatrix}^T + \mathbf{t}_{b_k}^{(i)} \quad (5.67)$$

sowie

$$\mathbf{b}_k^{(i \rightarrow j)} = \begin{pmatrix} x_t^{(j)} & y_t^{(j)} \end{pmatrix}^T + \mathbf{t}_{b_k}^{(j)} \quad (5.68)$$

Damit wird über die Ankerpunkte $\mathbf{b}_k^{(i \rightarrow j)}$ ein Zusammenhang zwischen den beiden Lagen i und j hergestellt. Die Vektoren $\mathbf{t}_{b_k}^{(i)}$ und $\mathbf{t}_{b_k}^{(j)}$ geben den Abstand zu den Positionen $\begin{pmatrix} x_s^{(i)} & y_s^{(i)} \end{pmatrix}^T$ und $\begin{pmatrix} x_t^{(j)} & y_t^{(j)} \end{pmatrix}^T$ der jeweiligen Lagen wieder. Jeder Ankerpunkt trägt dabei zwei weitere Gleichungen zum Gleichungssystem 5.66 bei.

In Matrizenform können diese Zusammenhänge folgendermaßen geschrieben werden:

$$\begin{pmatrix} t_{b_0}^{(i)} \\ t_{b_0}^{(j)} \\ \dots \\ t_{b_K}^{(i)} \\ t_{b_K}^{(j)} \end{pmatrix} = \begin{bmatrix} \dots & -1 & \dots & 0 & \dots & 1 & \dots \\ \dots & 0 & \dots & -1 & \dots & 1 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & \dots & \dots & \dots & 1 \\ \dots & \dots & 0 & \dots & -1 & \dots & 1 \end{bmatrix} \begin{pmatrix} \mathbf{x}^{(i)} \\ \mathbf{x}^{(j)} \\ \mathbf{b}^{(i \rightarrow j)} \end{pmatrix} \quad (5.69)$$

oder auch

$$\begin{pmatrix} \mathbf{t}_b^{(i)} \\ \mathbf{t}_b^{(j)} \end{pmatrix} = \mathbf{B}^{(i \rightarrow j)} \begin{pmatrix} \mathbf{x}^{(i)} \\ \mathbf{x}^{(j)} \\ \mathbf{b}^{(i \rightarrow j)} \end{pmatrix} \quad (5.70)$$

Durch Gewichtung der Gleichungen 5.67 und 5.68 können Abweichungen von diesen eine höhere Bedeutung zugewiesen werden. Die Gewichte gehen dabei quadratisch in den Fehler ein und müssen abhängig von der Anzahl der Gleichungen gesetzt werden. Eine mögliche Wahl der Gewichte wäre $w_{b_k} = \sqrt{c * N}$. Wobei N die Anzahl der Gleichungen wiedergibt und c eine frei wählbare Konstante ist. In der Praxis hat sich eine Wahl von $c = 2$ als günstig erwiesen.

Durch die Kombination der Gleichungssysteme 5.66 und 5.70 kann nun ein System an Gleichungen angegeben werden, welches eine gemeinsame Berechnung der Einzelbildpositionen zweier miteinander verbundenen Lagen ermöglicht (Gleichungssystem 5.71).

$$\begin{pmatrix} \mathbf{W}^{(i)}\mathbf{t}^{(i)} \\ \mathbf{W}^{(j)}\mathbf{t}^{(j)} \\ \mathbf{W}^{(i \rightarrow j)}\mathbf{t}_b^{(i)} \\ \mathbf{W}^{(i \rightarrow j)}\mathbf{t}_b^{(j)} \end{pmatrix} = \begin{bmatrix} \mathbf{A}^{(i)}\mathbf{W}^{(i)} \\ \mathbf{A}^{(j)}\mathbf{W}^{(j)} \\ \mathbf{B}^{(i \rightarrow j)}\mathbf{W}^{(i \rightarrow j)} \end{bmatrix} \begin{pmatrix} \mathbf{x}^{(i)} \\ \mathbf{x}^{(j)} \\ \mathbf{b}^{(i \rightarrow j)} \end{pmatrix} \quad (5.71)$$

Dieser Ansatz kann für eine beliebige Anzahl an Lagen verallgemeinert werden. Dazu ist es nur notwendig, weitere Chipebenentransformationen und Ankerpunktgleichungen in das System einzubringen. Gleichungssystem 5.72 beschreibt die geometrischen Relationen für L Lagen.

$$\begin{pmatrix} \mathbf{W}^{(0)}\mathbf{t}^{(0)} \\ \mathbf{W}^{(1)}\mathbf{t}^{(1)} \\ \dots \\ \mathbf{W}^{(L-1)}\mathbf{t}^{(L-1)} \\ \mathbf{W}^{(L)}\mathbf{t}^{(L)} \\ \mathbf{W}^{(0 \rightarrow 1)}\mathbf{t}_b^{(0)} \\ \mathbf{W}^{(0 \rightarrow 1)}\mathbf{t}_b^{(1)} \\ \dots \\ \mathbf{W}^{(L-1 \rightarrow L)}\mathbf{t}_b^{(L-1)} \\ \mathbf{W}^{(L-1 \rightarrow L)}\mathbf{t}_b^{(L)} \end{pmatrix} = \begin{bmatrix} \mathbf{A}^{(0)}\mathbf{W}^{(0)} \\ \mathbf{A}^{(1)}\mathbf{W}^{(1)} \\ \dots \\ \mathbf{A}^{(L-1)}\mathbf{W}^{(L-1)} \\ \mathbf{A}^{(L)}\mathbf{W}^{(L)} \\ \dots \\ \mathbf{B}^{(0 \rightarrow 1)}\mathbf{W}^{(0 \rightarrow 1)} \\ \mathbf{B}^{(L-1 \rightarrow L)}\mathbf{W}^{(L-1 \rightarrow L)} \end{bmatrix} \begin{pmatrix} \mathbf{x}^{(0)} \\ \mathbf{x}^{(1)} \\ \dots \\ \mathbf{x}^{(L-1)} \\ \mathbf{x}^{(L)} \\ \mathbf{b}^{(0 \rightarrow 1)} \\ \dots \\ \mathbf{b}^{(0 \rightarrow 1)} \end{pmatrix} \quad (5.72)$$

Das dadurch entstehende Gleichungssystem kann dabei sehr groß werden. Bei sieben Ebenen ($L = 7$) mit jeweils 10.000 Einzelbildern resultiert ein System mit in etwa 70.000 Unbekannten und 280.000 Gleichungen. Durch die dünne Besetzung der Matrix kann das System dennoch mit relativ geringem Aufwand gelöst werden.

Abbildung 5.8 zeigt vier Lagen eines mit der oben beschriebenen Methode registrierten IC.

Im allgemeinen Fall besitzen globale Chiptransformation mehr als zwei Freiheitsgrade, da die Probe zum Zwecke der Präparation aus dem REM entfernt werden muss. Sind nun die Lagen i und j zueinander um den Winkel $\phi^{(i \rightarrow j)}$ gedreht, so müssen die lokalen Chiptransformationen sowie auch die Einzelbilder selbst zuvor um diesen Winkel gedreht werden. Ähnliches gilt auch, wenn die Bilder mit unterschiedlicher geometrischer Auflösung aufgenommen worden sind. Dabei ist der relative Skalierungsfaktor zu ermitteln und eine Lage dementsprechend anzupassen. Zur Ermittlung der Transformationsparameter können die Punktkorrespondenzen (Ankerpunktpaare) zwischen den Lagen i und j , wie im Anfang dieses Abschnittes beschrieben, benutzt

werden. Dazu werden die Transformationsparameter einer rigiden Abbildung durch Gleichung 5.47 und den RANSAC-*inlier* der Punktkorrespondenzen berechnet und der translatorische Anteil auf Null gesetzt. Als globales Koordinatensystem wird die Referenzebene definiert und darauf die restlichen Lagen ausgerichtet. Hierzu werden die lokalen Chiptransformationen durch die ermittelte rigide Abbildung transformiert. Beim Zusammensetzen der Bilder ist weiters darauf zu achten, dass die Einzelbilder ebenfalls dieser Transformation unterworfen werden müssen.

5.5 Genauigkeitsanalysen

Um die evaluierten Verfahren einer quantitativen Genauigkeitsanalyse zugänglich zu machen, wurden für die einzelnen Transformationstypen Methoden entwickelt, die eine Abschätzung der Genauigkeit der Verfahren liefern.

5.5.1 Lokale Bildtransformation

Die lokalen Bildtransformationen stellen die Grundlage aller übergeordneten Transformationen dar. Bei Bilderserien von mehreren zehntausend Bildern muss diese mit sehr hoher Genauigkeit erfolgen, um die geforderten Fehlertoleranzen einhalten zu können. Eine pixelgenaue Bildtransformation sollte unter optimalen Bedingungen einen maximalen Fehler von 0.5 Pixel verursachen. In einer Bilderserie mit 100×100 Bildern liefert diese Methode einem maximalen Fehler von $100 * \sqrt{0.5^2 + 0.5^2} \approx 70 \text{ Pixel}$. Bei einer Pixelgröße von 10 nm ergibt dies einen Fehler von 700 nm , was die kleinste Strukturgröße von aktuellen Mikrochips bei weiten überschreitet. In dieser Abschätzung wurde allerdings davon ausgegangen, dass alle Registrierungen pixelgenau berechnet wurden. Von dieser Voraussetzung kann in der Praxis jedoch nicht ausgegangen werden, da Überlappungsbereiche vorhanden sind, in denen nur wenig oder gar keine Information vorhanden ist. Diese Überlappungsbereiche gilt es zu erkennen und gesondert zu betrachten. Die restlichen informationsreichen Bildbereiche müssen subpixelgenau registriert werden, um den Fehler so gering wie möglich halten zu können.

Eine Klassifizierung der Überlappungsgebiete soll im ersten Schritt eine Abschätzung liefern, ob überhaupt mit genauen Resultaten zu rechnen ist.

Da den SEM Bildern ein teilweise hoher Anteil an gauß'schem Rauschen überlagert ist und manche Registrierungsverfahren sehr rauschsensitiv sind, ist es sinnvoll ein Maß für die Rauschintensität abzuleiten. Da die Intensität des Bildrauschens innerhalb einer Bilderserie meist ähnlich bleibt, kann dies anhand eines Einzelbildes erfolgen.

Ein geeignetes Maß zur Abschätzung des Rauschanteiles stellt die *Entropie* einer Aufnahme eines annähernd homogenen Chipbereiches dar. Die Entropie ist eine statistische Maßzahl, die den Grad an Zufälligkeit wiedergibt. Wenn ein Bild $G = 2^b$ verschiedene

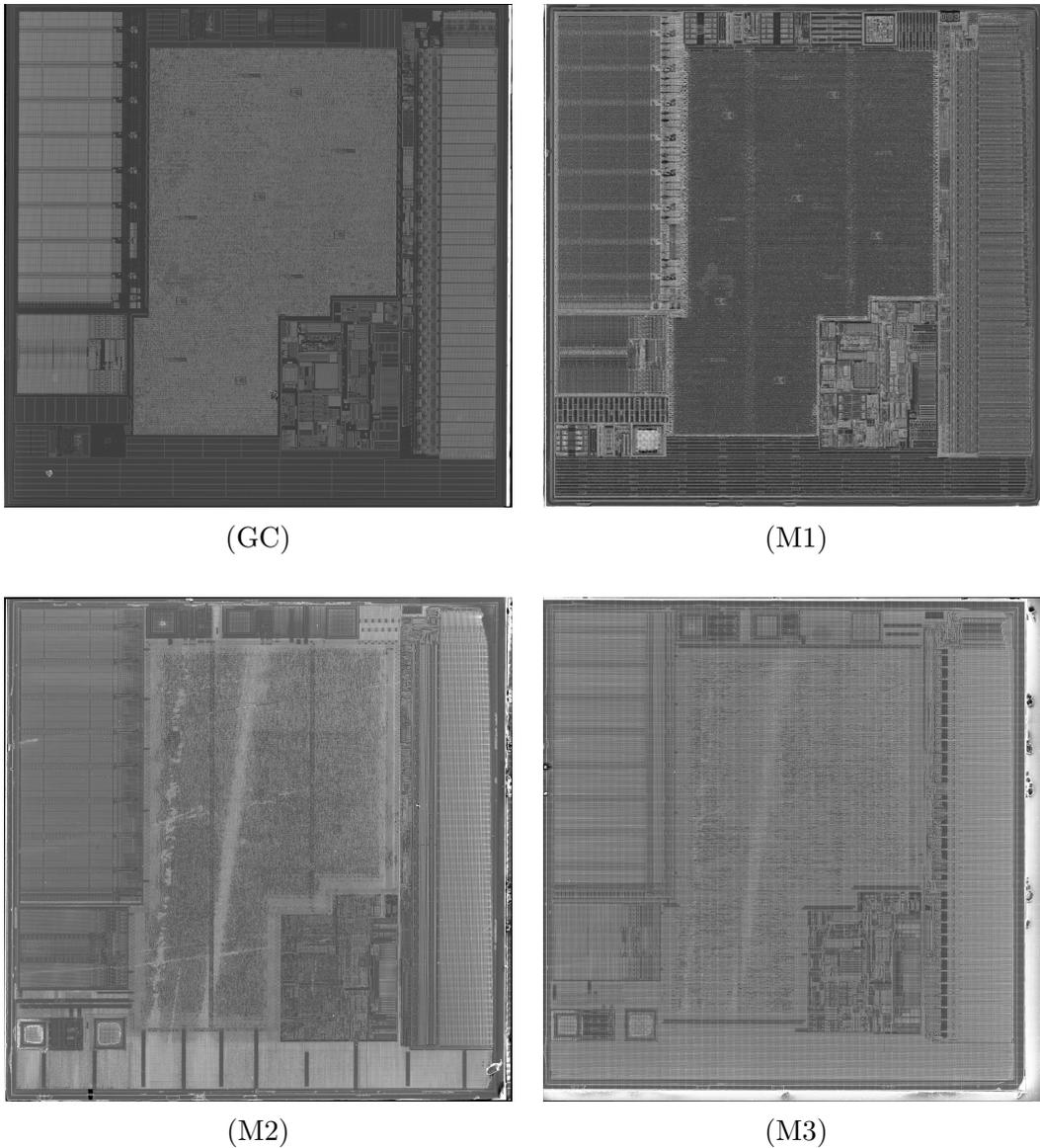


Abbildung 5.8: Chipebenen nach globaler Registrierung. Die hier gezeigten Mosaik wurden durch Lösen von Gleichung 5.72 berechnet. Bild (GC) besteht aus 82×82 Kacheln mit jeweils 1000×1000 Pixel (ergibt in etwa eine Auflösung von 82000×82000 Pixel oder 6.7 GigaPixel) und zeigt die Transistorebene des IC. Die Abbilder der ersten beiden Metallisierungsebenen (M1) und (M2) bestehen aus jeweils 83×83 Teilbildern mit einer Bildauflösung von in etwa 6.9 GigaPixel . Die dritte Ebene (M3) wurde mit 64×64 Teilbildern (4 GigaPixel) aufgenommen. Als lokale Registrierungsmethode wurde die NCC verwendet.

Intensitätsstufen (Grauwerte) besitzt, die durch b Bit darstellbar sind, und die Intensitätsstufe i die Wahrscheinlichkeit P_i besitzt dann kann die Entropie H_e (ohne Korrelationen zwischen den Intensitätsstufen zu berücksichtigen) durch

$$H_e = - \sum_{i=0}^{G-1} P_i \log_2(P_i) \quad (5.73)$$

angegeben werden. Eine exakte Berechnung der Wahrscheinlichkeiten P_i ist jedoch nicht möglich, da die statistischen Eigenschaften des Bildes unbekannt sind. Jedoch lässt sich P_i und damit die Entropie über das Histogramm abschätzen [?]. Sei h_i die Anzahl der Pixel mit der Intensität i im Bild \mathbf{I} der Größe $M \times N$, so lässt sich die Wahrscheinlichkeit P_i durch

$$\tilde{P}_i = \frac{h_i}{MN} \quad (5.74)$$

abschätzen. Daraus folgt für die geschätzte Entropie $\tilde{H}_e \in [0, \log_2(G)]$ mit Gleichung 5.73

$$\tilde{H}_e = - \sum_{i=0}^{G-1} \tilde{P}_i \log_2(\tilde{P}_i) \quad (5.75)$$

Die minimale Entropie von $\tilde{H}_e = 0$ wird bei einem homogenen Bild gemessen, die maximale von $\tilde{H}_e = b$ bei einem Bild mit gleichverteiltem Histogramm. Die Abschätzung kann jedoch nur eine grobe Näherung darstellen, da hohe Korrelation zwischen den Intensitätsstufen besteht. Bei einem Farbbild können die Farbkanäle entweder getrennt betrachtet oder das Bild in ein Grauwertebild konvertiert werden.

Hohe Entropie ist gleichbedeutend mit hoher Zufälligkeit, was bei der Abbildung einer homogenen Fläche Rückschlüsse auf den Rauschanteil zulässt. In Abbildung 5.9 wurden Aufnahmen von drei unterschiedlichen Bilderserien gegenübergestellt.

Nach der Klassifikation der Bildausschnitte wird davon abhängig ein passender Registrierungsalgorithmus gewählt. Bei einer Rauschintensität von $\tilde{H}_e > 5$ stellte sich die NCC im Vergleich zur PC als robuster heraus und gilt damit als Mittel der Wahl für stark verrauschte SEM Bilder.

Bei direkten Methoden, können die Korrelationskoeffizienten zur Beurteilung der Qualität der Registrierungen benutzt werden. Die normalisierte Kreuzkorrelation liefert Werte zwischen $[-1, 1]$. In Abbildung 5.10 ist ein Ausschnitt des Histogramm der NCC Koeffizienten dargestellt. Da nicht alle Transformationen für die Berechnung der Chipebenen-Transformationen notwendig sind, sollte ein Schwellwert gesetzt werden, anhand dessen entschieden wird, ab wann eine Transformation aufgrund des Korrelationskoeffizienten als genau eingestuft wird. Dieser Wert kann entweder festgelegt oder durch statistische Methoden berechnet werden. Bei Normalverteilung kann man den Mittelwert μ und die Standardabweichung σ durch Parameterschätzung ermitteln. Die Überprüfung der Verteilung auf Normalverteilung könnte beispielsweise mit dem *Shapiro-Wilk* Test erfolgen.

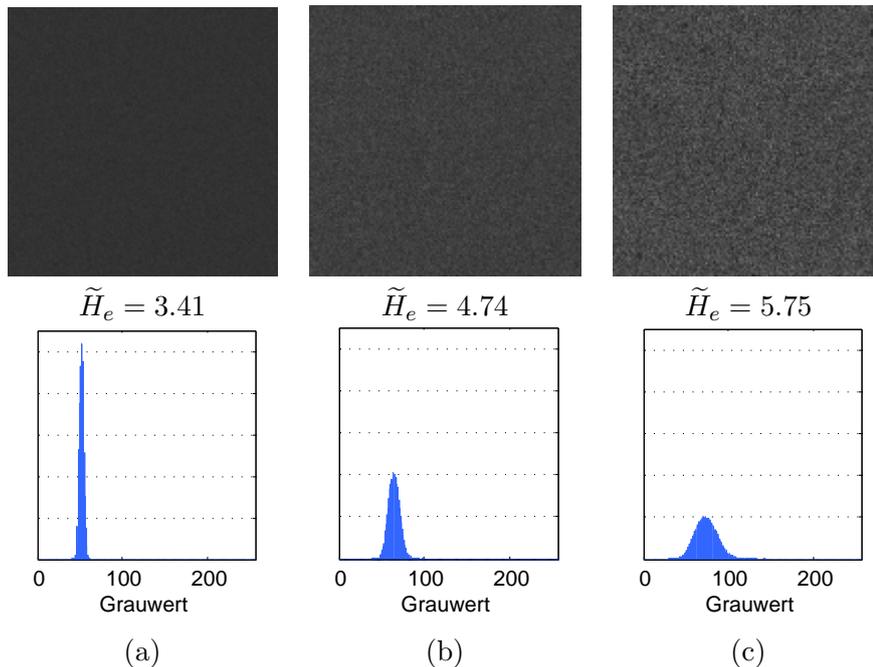


Abbildung 5.9: Entropie in Bildern. Rauschen erschwert die Registrierung von Bildern. Diese drei Ausschnitte sind von unterschiedlichen Bilderserien eines SEM entnommen. Dabei wurden Bildausschnitte gewählt, die eine nahezu homogene Oberfläche wiedergeben. Der höhere Rauschanteil von Bild (c) spiegelt sich in einer höheren Entropie sowie in einem höherem σ des normalverteilten Histogramms wieder. Ein Bild mit konstantem Grauwert besitzt eine Entropie von $\tilde{H}_e = 0$. Die maximale Entropie von $\tilde{H}_e = \log_2(G) = b = 8$ wird bei Gleichverteilung der Grauwerte erreicht.

In der Praxis hat sich gezeigt, dass ein hoher fester Schwellwert von $Th > 0.98$ meist zufriedenstellende Ergebnisse liefert. Durch manuelles Feintuning unter Zuhilfenahme des Histogramms können die Ergebnisse allerdings verbessert werden. Die automatische Wahl des Schwellwertes wird durch die vielfältigen Verteilungen erschwert.

Um genaue Aussagen über die Genauigkeit der evaluierten Registrierungsverfahren treffen zu können, müssen die Transformationen bekannt sein (*ground truths*). Da dies in der Anwendung nicht möglich ist (sonst müsste man keine Registrierung durchführen), wurde ein Bilderpaar erzeugt, indem Bild \mathbf{I}_1 mit einer bekannten Translation $\mathbf{t}_{1 \rightarrow 2}$ in Bild \mathbf{I}_2 transformiert wurde. Um auch Verschiebungen im Subpixelbereich zu berücksichtigen, müssen die neuen Intensitätswerte durch Interpolation (z.B. bikubisch) von den umliegenden abgeschätzt werden. In Abbildung 5.11 wurden durch die NCC unter einer randomisierter Translation \mathbf{t}_i 1000 lokale Bildtransformationen \mathbf{t}'_i mit $i = 0 \dots 999$ ermittelt und das Residuum $res_i = \|\mathbf{t}_i - \mathbf{t}'_i\|_2$ als Histogramm dargestellt. Der mittlere Fehler beträgt in etwa $\frac{1}{18}$ eines Pixels.

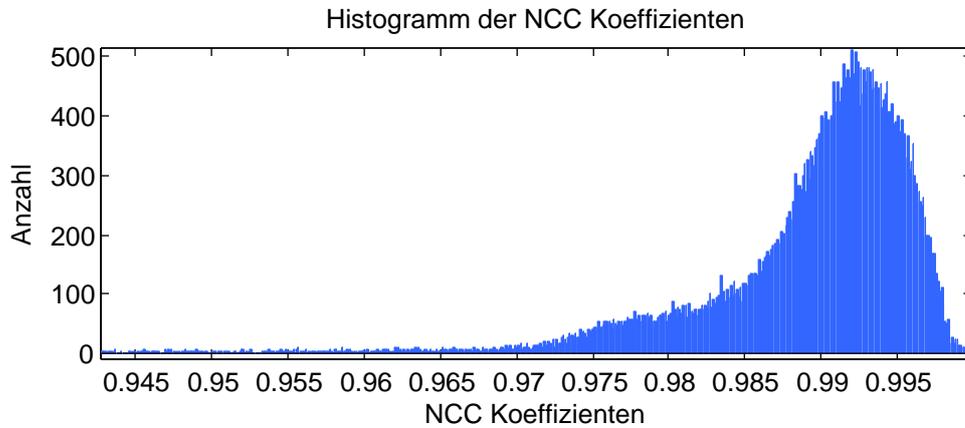


Abbildung 5.10: Histogramm der NCC Koeffizienten. Die NCC Koeffizienten einer Registrierung mit 53834 Transformation sind in diesem Histogramm dargestellt. Der größte Teil der Koeffizienten ist um 0.993 verteilt, was in Summe einer sehr guten Registrierung entspricht. Ein Schwellwert bestimmt, ab wann eine Transformation als genau genug angenommen wird.

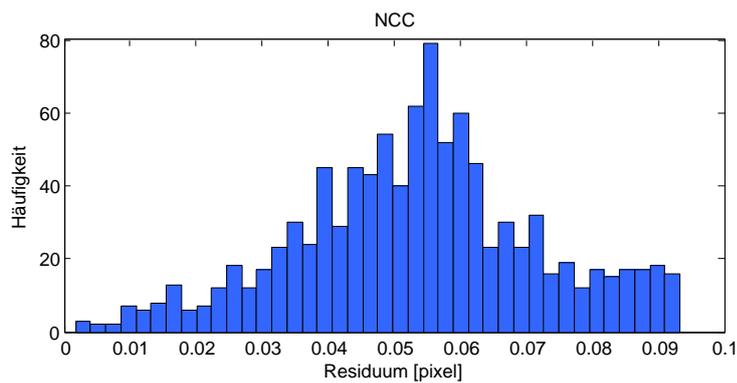


Abbildung 5.11: Histogramm der Residuen von Registrierungen durch die NCC. Es wurden 1000 synthetische Bilderpaare erzeugt und mit 10 dB gauß'schem Rauschen beaufschlagt. Die Registrierung erfolgte subpixelgenau durch Einpassen eines zweidimensionalen Polynoms zweiter Ordnung an die Umgebung des Maximums der Korrelationsfaktoren (siehe 5.3.2.3). Die Residuen sind um 0.05 Pixel verteilt was der mittleren theoretisch erreichbaren Genauigkeit entspricht.

5.5.2 Chipebenentransformation

Zur Evaluierung wurden ideale lokale Bildtransformationen mit gauß'schem Rauschen ($\sigma = 0.08$) beaufschlagt und durch naive Verkettung dieser ein initiales Mosaik erstellt. Die nachfolgende Optimierungsphase versucht den Fehler zwischen den ursprünglichen lokalen Bildtransformationen und den Differenzen der berechneten Chipebenenkoordinaten zu minimieren. Die Resultate der Methode der kleinsten Quadrate und die der nichtlinearen Optimierung liefern dabei ähnliche Resultate. In [Abbildung 5.12](#) sind die Kacheln eines Mosaiks vor und nach der Optimierung dargestellt und [Abbildung 5.13](#) zeigt die

Verteilung der Fehler.

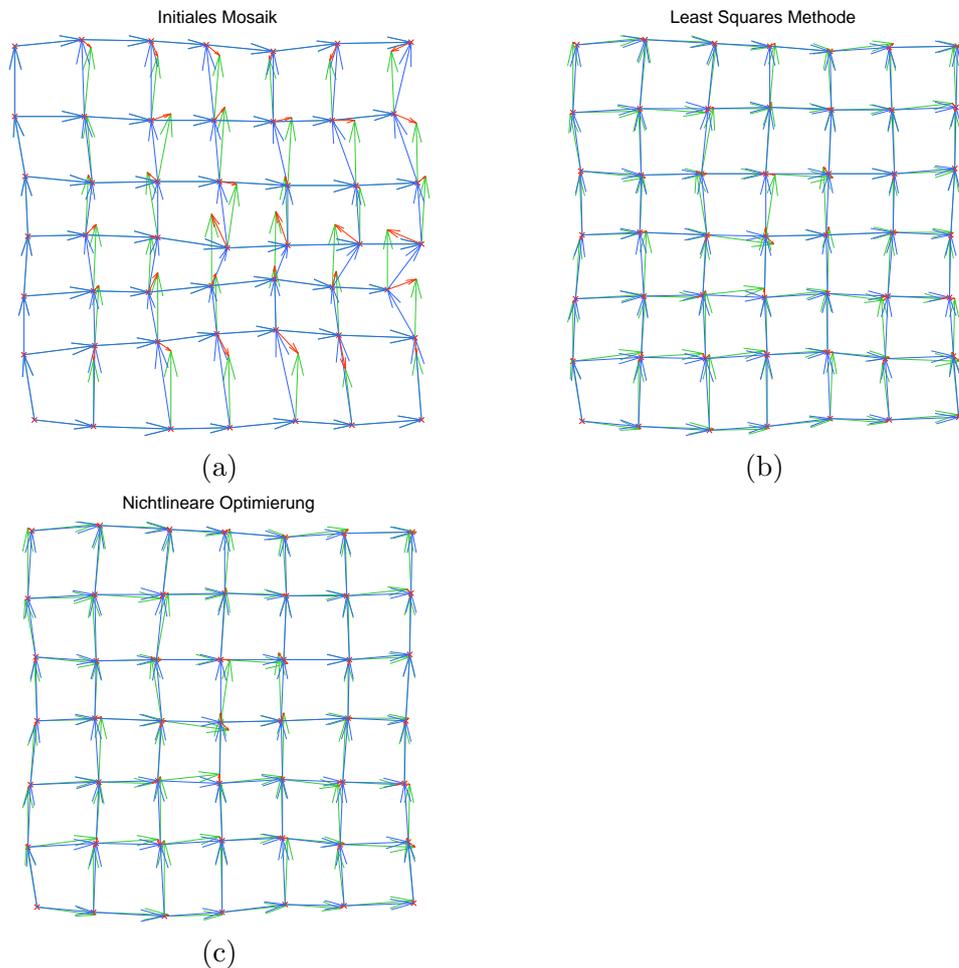


Abbildung 5.12: Optimierung der Chipecbenentransformation. Ideale lokale Bildtransformationen wurden mit gauß'schem Rauschen belegt und durch mäanderförmige Verkettung dieser wurde ein initiales Mosaik (a) erstellt. Grüne Kanten entsprechen dabei den lokalen Bildtransformationen, blaue Kanten werden durch Differenzbildung der errechneten Chipecbenenkoordinaten gewonnen. Der rote Pfeil gibt die Distanz der beiden Vektoren wieder und entspricht somit dem Fehler, der minimiert werden soll. Die direkte Ermittlung der Chipecbenenkoordinaten über die gauß'sche Normalgleichung liefert Mosaik (b). Die Fehler wurden hierbei durch die Methode der kleinsten Quadrate verkleinert. Mosaik (c) bedient sich eines nichtlinearen Optimierungsverfahrens, wobei als Startwert die Koordinaten von (a) verwendet wurden. Durch Balancieren der Chipecbenenkoordinaten wird der Fehler minimiert.

Durch das Entfernen von lokalen Bildtransformationen kommt es meist vor, dass nicht alle Bilder durch eine Verkettung der verbleibenden Transformationen erreicht werden können. Stellt man diesen Zusammenhang als Graph dar, in dem die Bilder einer Chipecbene die Knoten und die lokalen Bildtransformationen die Kanten darstellen, würde dies

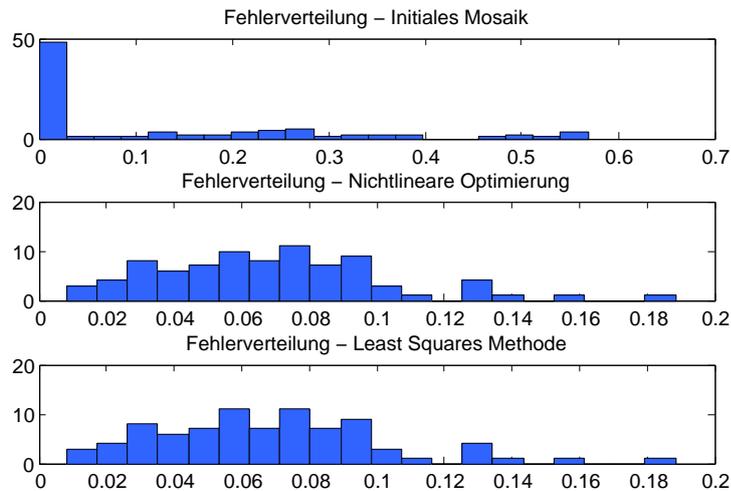


Abbildung 5.13: Fehlerenergie nach und vor Optimierung der Chi Ebenenkoordinaten. Die in Abbildung 5.12 gezeigten Mosaik wurden hinsichtlich der Fehlerverteilung untersucht. Das initiale Mosaik 5.12(a) zeigt eine breite Verteilung mit Fehlern bis zum 0.55-fachen der Bildgröße. Das direkte Lösen durch die gauß'sche Normalgleichung, sowie die Minimierung der Translationsdifferenz verbessert die Fehlerverteilung wesentlich. Die beiden Methoden liefern hierbei sehr ähnliche Ergebnisse. Der maximale Fehler sinkt dabei auf das unter 0.2-fache der Bildgröße.

bedeuten, dass der Graph in Teilgraphen aufbricht. In Abbildung 5.14 sind die Teilgraphen, die sich nach der Wahl des NCC Schwellwertes von $TH = 0.98$ (siehe Abbildung 5.10) ergeben, im Mosaik farbig dargestellt.

Um einen zusammenhängenden Graphen zu erhalten, können nun die initialen Chi Ebenenkoordinaten, die das Laser Interferometer wie in Abschnitt 3.3.3 beschrieben liefert, verwendet werden. Um nicht Fehler in der Abschätzung der Chi Ebenentransformation (Abschnitt 5.4.3) zu verursachen, werden die Teilgraphen jeweils nur an einer Stelle miteinander verbunden. Durch die Ausgleichsrechnung würden Mehrfachverbindungen Fehler in allen Teilgraphen verursachen.

Um wie in Abschnitt 5.5.1 die berechneten Positionen mit *ground truths* vergleichen zu können, wurde ein Datensatz mit bekannten Positionen synthetisch erstellt. Im ersten Schritt wurden für den Datensatz die Chi Ebenenkoordinaten berechnet und ein Mosaik erzeugt. Im zweiten Schritt wurde das erstellte Mosaik mit einem virtuellem $N \times M$ Raster in Kacheln eingeteilt. Das Raster errechnet sich dabei durch die Anzahl der Kacheln in x und y Richtung und einem mit einer normalverteilten Unsicherheit belegtem Überlapp. Bildunstetigkeiten die durch Registrierungsfehler verursacht wurden haben keinen Einfluß auf der Ergebnis, da diese in beiden Überlappungsgebieten vorkommen. Um diese unnatürlichen Sprünge jedoch zu vermeiden, wurde das Raster ungleich dem Ursprungsraster gewählt. Durch Ausschneiden der Teilbilder aus der *ground truth* Serie kann ein Referenzprojekt angelegt werden. Um Verschiebungen im Subpixelbereich zu

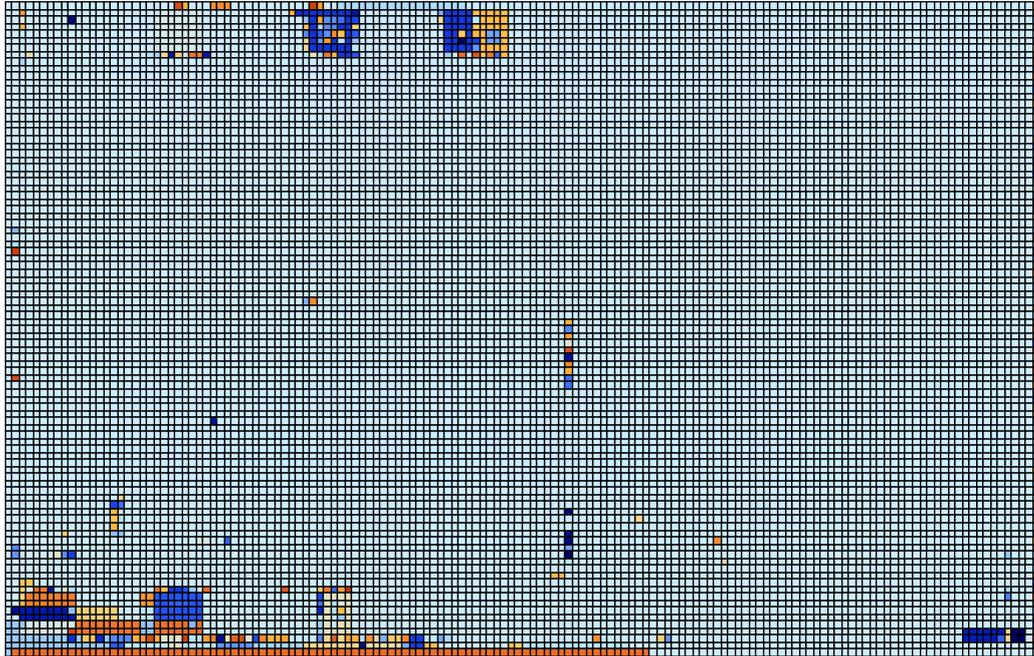


Abbildung 5.14: Graph der zusammenhängenden lokalen Bildtransformationen. Nach Aussortierung der als ungenau angenommenen lokalen Bildtransformationen ist meist nicht mehr jedes Bild von jedem aus durch Verkettung der lokalen Bildtransformationen erreichbar. Dieser Zusammenhang lässt sich durch einen Graphen beschreiben, der in mehrere Teilgraphen aufbricht. In dieser Abbildung wurden die Teilgraphen eines Mosaikes durch verschiedene Farben dargestellt. Die Anzahl der entstehenden Teilgraphen hängt eng mit der Wahl des Schwellwertes (Abschnitt 5.5.1) zusammen. Entstehen zu viele Teilgraphen, so sollte der Schwellwert herabgesetzt werden. Vor allem an den Chiprändern befinden sich meist durch die Präparation entstandene Artefakte die aus der Fokusebene ragen und damit verschwommen abgebildet werden. Da diese Bilder jedoch keine Information über den Aufbau den Chips beinhalten ist dies unkritisch.

berücksichtigen, wurden die Pixelintensitäten durch bikubische Interpolation ermittelt.

Als ein weiteres gut geeignetes Qualitätsmaß, um die Qualität von Chipebenentransformationen beurteilen zu können, haben sich *zirkuläre Transformationsketten* erwiesen. Zirkuläre Transformationsketten, wie in Abbildung 5.15 skizziert, haben die Eigenschaft, dass sie geschlossen sind. Startet die Kette der Länge N bei Position \mathbf{p}_0 des Bild ind_0 so lässt sich die Kette durch $\mathbf{p}_i = \mathbf{T}_{ind_{i-1} \rightarrow ind_i} \mathbf{p}_{i-1}$ mit $i = 1 \dots N$ berechnen. Um die Bedingung einer Kette zu erfüllen muss gelten $ind_0 = ind_N$. Das Residuum kann durch $res_{cir} = \|\mathbf{p}_N - \mathbf{p}_0\|_2$ berechnet werden. Sind die beteiligten Transformationen ideal gilt $res_{cir} = 0$. In Abbildung 5.16 ist ein Histogramm der Restfehler einer Registrierung mit 58405 zirkulären Transformationen dargestellt. Die lokalen Bildtransformationen wurden durch die NNC mit Subpixelgenauigkeit ermittelt. Die Chipebenenkoordinaten können durch lösen von Gleichungssystem 5.58 bestimmt werden.

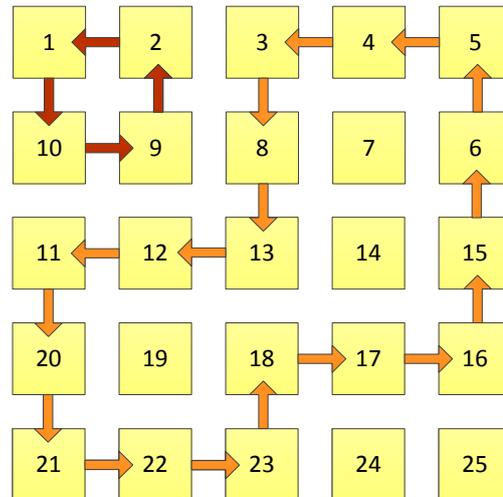


Abbildung 5.15: Beispiele für zirkuläre Transformationsketten. Verkettet man die gefundenen Transformationen auf solche Weise, sodass Anfangs- und Endbild die selben sind, lassen sich über das Residuum Rückschlüsse auf die Genauigkeit der beteiligten Transformationen schließen. Dabei sind unterschiedliche Formen denkbar. Die einfachste Form stellte die Verkettung von vier benachbarten Bildern dar (Kette $1 \rightarrow 10, 10 \rightarrow 9, 9 \rightarrow 2, 2 \rightarrow 1$). Anfangs-/Endpunkt sind innerhalb der Kette frei wählbar.

Zu beachten gilt es hierbei jedoch, dass ein Residuum von Null nicht eindeutig bedeutet, dass die gefundenen Transformationen ideal sind. Wird in Abbildung 5.15 eine gesamte Zeile oder Spalte verschoben, so bleibt das Residuum unverändert.

5.5.3 Globale Chiptransformation

Die globale Chiptransformation $\mathbf{T}_{i \rightarrow j}$, die das Chipebenenkoordinatensystem von Ebene i in das Chipebenenkoordinatensystem von Ebene j überführt, wurde aus k Korrespondenzen $(\mathbf{p}_k^{(i)}, \mathbf{p}_k^{(j)})$ mittels RANSAC ermittelt. Nun soll eine numerische Abschätzung der Genauigkeit erfolgen. Da eine affine Transformation mit sechs DOF zur exakten Lösung genau drei Punktpaare benötigt und zur Berechnung ein Vielfaches dieser verwendet wurden, kann aus dem Verhältnis *inlier* zu *outlier* auf die Qualität der Transformation geschlossen werden. Weiters kann über das Residuum der *inlier* eine Fehlerabschätzung numerisch angegeben werden.

$$\mathbf{res}_k^{i \rightarrow j} = \|\mathbf{T}_{i \rightarrow j} \mathbf{p}_k^{(i)} - \mathbf{p}_k^{(j)}\|_2 \quad (5.76)$$

$$r = \|\mathbf{res}_k\|_2 \quad (5.77)$$

Zu beachten ist hierbei, dass in das Fehlermaß sowohl Fehler aus der Bestimmung der Korrespondenzen auf den beiden Chipebenen, Fehler aus der Chipebenentransformation (Abschnitt 5.5.2) sowie Fehler der lokalen Bildtransformationen (Abschnitt 5.5.3) einfließen. Der Einsatz von robusten statistischen Methoden gewährleistet jedoch, dass inkonsistente Korrespondenzen nicht zur Berechnung miteinbezogen werden. In Abbildung

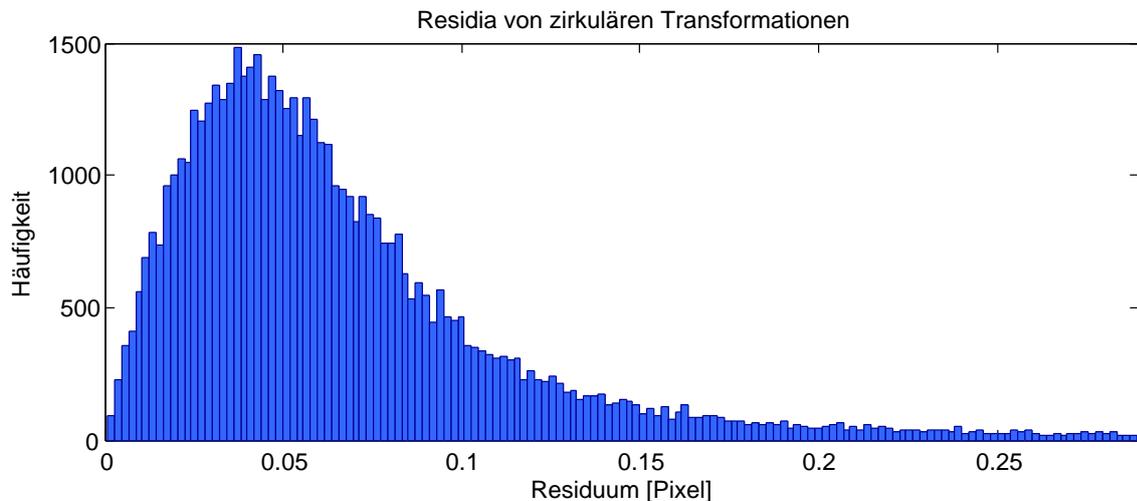


Abbildung 5.16: Histogramm der Residuen von zirkulären Transformationen. Die Residua von zirkulären Transformationen wurden durch Verkettung von jeweils vier Transformationen ($T_{\downarrow}T_{\rightarrow}T_{\uparrow}T_{\leftarrow}$) berechnet (siehe Abbildung 5.15). Der Hauptteil liegt um 0.05 Pixel verteilt, welches der mittleren Genauigkeit des eingesetzten Bildregistrierungsverfahrens entspricht. Von den 58 405 4-Tupel Transformationen liegen 93.6% innerhalb eines Fehlers von $[0, 0.2]$, 99.9% innerhalb von $[0, 1]$. Der maximale Fehler beträgt 1.9 Pixel mit einer Häufigkeit von 2. Die lokalen Bildtransformationen wurden durch die NCC ermittelt, die Chipebenenkoordinaten durch die LSM.

5.17 wurden die Residua für 60 über die Chi Ebene verteilte Korrespondenzen ermittelt und gegenübergestellt.

5.6 Radiometrische Korrekturen

Durch variierende Parameter des Aufnahmesystems können innerhalb einer Bildserie Helligkeitsschwankungen auftreten, die eine automatische Analyse erschweren. Die dabei auftretenden Unstetigkeiten an den Bildgrenzen können in weiteren Analyseschritten zu Fehlerkennungen führen. Abhängig von den Anforderungen der Problemstellung kann dieser Schritt jedoch entfallen.

Im Folgenden werden Verfahren vorgestellt, um die Konsistenz der radiometrischen Eigenschaften innerhalb einer Serie zu verbessern.

5.6.1 Korrektur der Bildhelligkeit

Unter der Annahme, dass bei dem Abbildungsverfahren ein linearer Zusammenhang zwischen Belichtungs- bzw. Detektionszeit und Helligkeit eines Pixels besteht, können durch Korrekturen der Helligkeitsmittelwerte Variationen in der Belichtungs- bzw. Detektionszeit innerhalb einer Bilderserie korrigiert werden. Für eine Bilderserie mit N Bildern, den

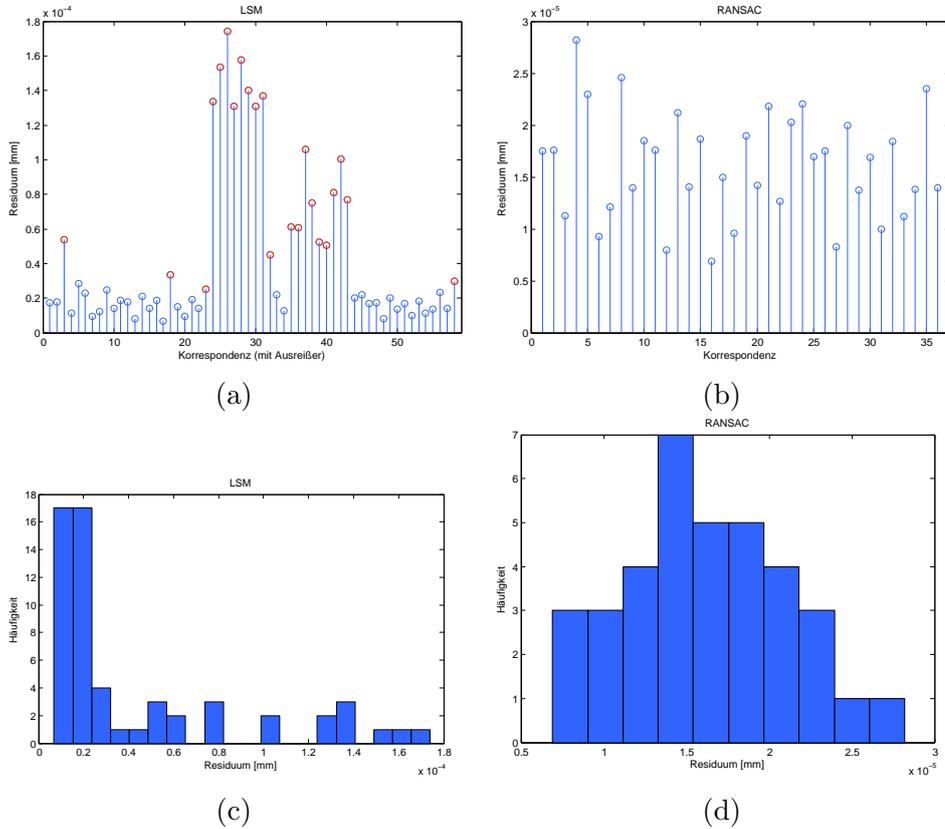


Abbildung 5.17: Für 60 über die Chipebene verteilte Korrespondenzen wurde mit den in Abschnitt 5.4.4 vorgestellten Methoden eine affine Abbildung zwischen den Punktpaaren berechnet. Durch Unsicherheiten in der Punktwahl und durch Verzerrungen, die von Fehlern des *matching* stammen, weisen bestimmte Punktpaare ein erhöhtes Residuum auf. Das Residuum der einzelnen Korrespondenzen ist in Diagramm (a), ein Histogramm derselben in (c) dargestellt. Um Ausreißer zu identifizieren und um diese nicht bei der Berechnung mit einzubeziehen, wurde der RANSAC Algorithmus eingesetzt. Um unter den Tolleranzgrenzen von $\frac{90 \text{ nm}}{2}$ zu bleiben wurde das Residuum, ab wann ein Punktepaar als *inlier* bezeichnet werden kann, auf $r_{max} = 40 \text{ nm}$ und die Wahrscheinlichkeit, dass sich ein Ausreißer in den Punktpaaren befindet auf $\epsilon = 0.4\%$ gesetzt. Mit $N = 19$ Samples ergibt sich die Wahrscheinlichkeit, dass zumindest ein Sample keinen Ausreißer beinhaltet zu $p = 0.99$ (Gleichung 5.61). Ausreißer, die beim RANSAC Verfahren nicht zur Abschätzung der Transformation beitragen, wurden in (a) rot markiert. Im Vergleich zum LSM Verfahren sinkt der maximale Fehler von $\max(\mathbf{res}_{LSM}) = 175 \text{ nm}$ auf $\max(\mathbf{res}_{RANSAC}) = 28 \text{ nm}$ und der Mittelwert von $\tilde{res}_{LSM} = 45 \text{ nm}$ auf $\tilde{res}_{RANSAC} = 16 \text{ nm}$. Das Verhältnis $\frac{\#inlier}{\#outlier}$ beträgt $\frac{36}{24} = 1.5$.

Anpassungsfaktoren g_i , den Medianen m_{ij} sowie der Menge O der überlappenden Bilder i und j kann das Problem folgendermaßen formuliert werden:

$$g_i m_{ij} - g_j m_{ji} \stackrel{!}{=} 0 \quad (i, j) \in O \quad (5.78)$$

Nach Verankerung eines Anpassungsfaktors und Umschreibung in Matrizenform kann das Gleichungssystem wie in Abschnitt 5.4.3 beschrieben gelöst werden. Das Gleichungssystem weist dabei eine ähnliche Struktur wie in Gleichung 5.51 auf. Die Komponenten ungleich 0 der Matrix \mathbf{A} , werden dabei durch die entsprechenden Mittelwerte ersetzt.

5.6.2 Überblendung

Um auch nichtlineare Intensitätsschwankungen sowie Winkelabhängigkeiten (wie z.B. Abschattungen in der Elektronenmikroskopie oder Reflektionen an dünnen Schichten in der optischen Mikroskopie) in den Überlappungsgebieten ausgleichen zu können, wird eine Überblendungsfunktion definiert, die ein Bild auf das andere überführt. Eine einfache Möglichkeit, das Überlappungsgebiet R von Bild i nach Bild j als linear gewichtete Summe dieser darzustellen, liefert folgende Gleichung:

$$\mathbf{I}_{ij}(x, y) = \alpha \mathbf{I}_i + (1 - \alpha) \mathbf{I}_j \quad x, y \in \mathbb{R}, \alpha \in [0, 1] \quad (5.79)$$

Um mehr Flexibilität zu erhalten, kann der Faktor α mit einem Exponenten $n \in [1, \infty)$ versehen werden, der die Glätte des Übergangs bestimmt. Damit weiters gilt $\alpha^n + (1 - \alpha)^n \stackrel{!}{=} 1$ muss die Gleichung normalisiert werden.

$$\mathbf{I}_{ij}(x, y) = \frac{\alpha^n \mathbf{I}_i + (1 - \alpha)^n \mathbf{I}_j}{\alpha^n + (1 - \alpha)^n} \quad x, y \in \mathbb{R}, \alpha \in [0, 1] \quad (5.80)$$

Gleichung 5.79 ist ein Spezialfall dieser Gleichung mit $n = 1$. In Abbildung 5.18(a) ist Gleichung 5.80 für verschiedene Parameter n dargestellt. Masken mit unterschiedlichem n sind in Abbildung 5.18(b)-(d) ersichtlich.

Allgemein ist es jedoch nicht optimal, den gesamten Überlappungsbereich zu überblenden. Bei großen Überlappungsbereichen führt dies zu verschwommenen Kanten (*ghosting*), die eine Kantenerkennung bei späteren Analyseverfahren negativ beeinträchtigen können. Eine passende Fenstergröße sollte größer als das größte dominante Merkmal sein, jedoch kleiner als das doppelte dessen. Ohne Vorwissen, ist diese Merkmalsgröße jedoch schwer zu bestimmen. [?] schlägt vor, den in den Frequenzbereich transformierten Überlappungsbereich beider Bilder in Oktaven einzuteilen und die gewichtete Überblendung für die Frequenzbänder getrennt durchzuführen. Eine Implementierung dieses Ansatzes ist durch den Einsatz von laplace'schen Bildpyramiden auch im Ortsbereich möglich.

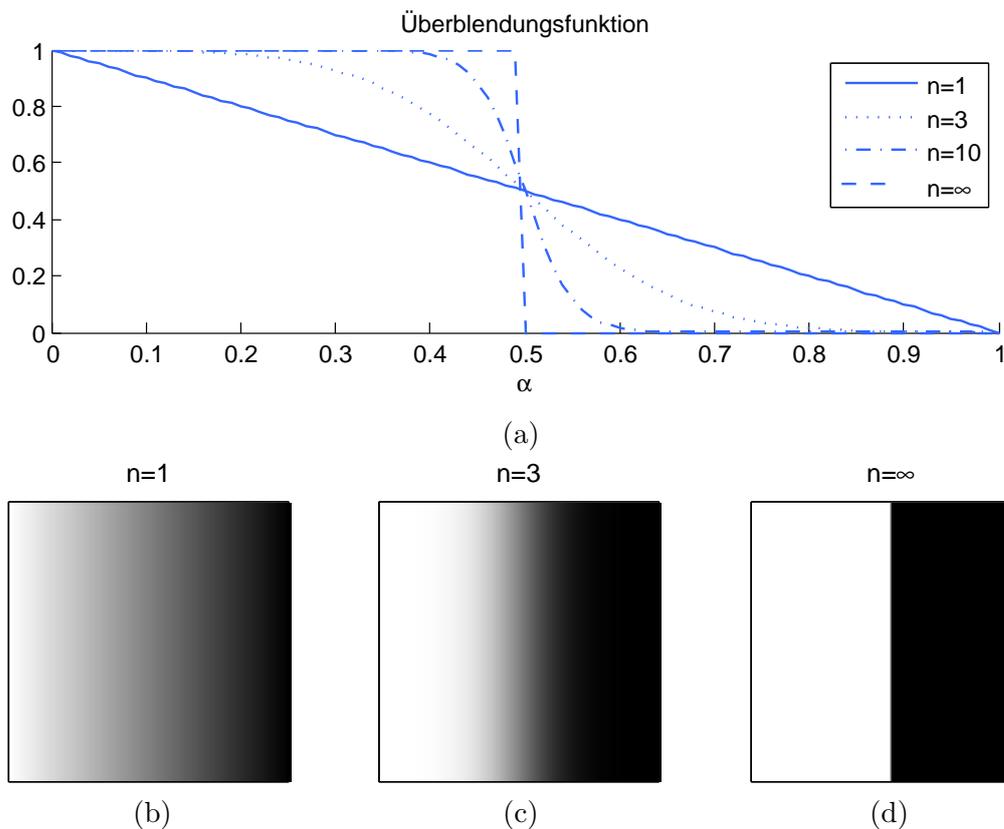


Abbildung 5.18: Überblendungsfunktion mit variierenden Parametern. In Gleichung 5.80 kann durch den Parameter n die Glattheit der Überblendungsfunktion angepasst werden. In Abbildung (a) ist die Überblendungsfunktion für verschiedene Exponenten n dargestellt. Abbildung (b) bis (d) zeigt die berechneten Transparenzmasken für einen horizontalen Übergang.

Die in dieser Arbeit implementierte Methode verwendet einen ähnlichen Ansatz wie in [?] vorgeschlagen. Dabei werden nur zwei Frequenzbänder (*two-band blending*) betrachtet. Tiefe Frequenzen werden langsam durch eine gewichtete Summe mit Gleichung 5.80 ($n = 3$) übergeführt, hohe Frequenzen durch eine sprunghafte binär gewichtete Summe ($n = \infty$). Dadurch wird gewährleistet, dass hohe Frequenzen erhalten bleiben und dennoch ein nahtloser Übergang entsteht. In Abbildung 5.19 sind die beiden Frequenzbänder anhand eines Chipausschnittes gegenübergestellt. Zur Tiefpass-Filterung wurde ein 5×5 Gausskern mit $\sigma = 1.5$ gewählt. Der hochfrequente Anteil des Bildes wurde durch Subtraktion des ursprünglichen Bildes mit dem durch den Gausskern gefilterten separiert.

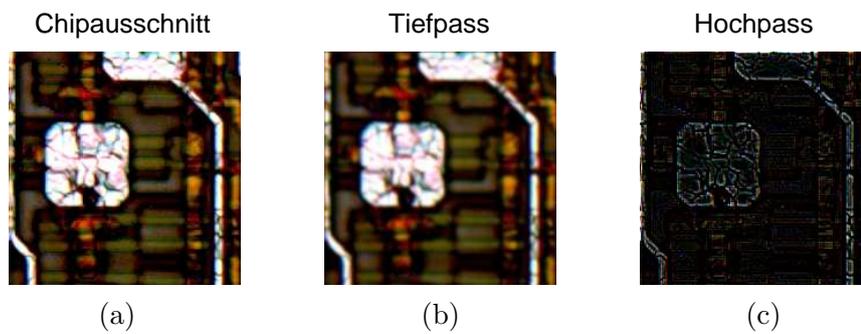


Abbildung 5.19: Das Überlappungsgebiet wird durch Filterung in einen Hochfrequenzanteil ($\lambda < 2$ Pixel) und einen Tieffrequenzanteil ($\lambda > 2$ Pixel) aufgeteilt. Diese werden getrennt durch Gleichung 5.80 überblendet. Da die hohen Frequenzen für Kanten wichtig sind, wird hierbei ein sprunghafter Übergang gewählt ($n = \infty$). Für tiefe Frequenzen ein glatter ($n = 3$).

The most beautiful experience we can have is the mysterious.

Albert Einstein

Kapitel 6

Schaltungsextraktion

Zusammenfassung — In diesem Kapitel wird der Begriff der Schaltungsextraktion eingeführt. Nach einer kurzen Einleitung und einem Abriss der bisherigen Forschungsergebnisse wird mit der Darstellung einer automatischen, bildanalysegestützten Rekonstruktion von Layoutdaten fortgefahren. Dabei werden Verfahren zur Segmentierung und Vektorisierung von Leiterbahnen, zur Kontaktextraktion und *routing* vorgestellt. Da die Bildeigenschaften sehr von der Präparation und dem Abbildungsverfahren abhängen, wurden verschiedene Methoden evaluiert sowie Vor- und Nachteile aufgezeigt. Ein kurzer Abschnitt über Netzlisten und Möglichkeiten einer Partitionierung runden dieses Kapitel ab.

Inhaltsverzeichnis

6.1	Einleitung	78
6.2	Stand der Technik	80
6.3	Leiterbahnenextraktion	81
6.4	Kontaktextraktion	97
6.5	Extraktion von Bauteilen	102
6.6	Routing	102
6.7	Netzliste	104
6.8	Zusammenfassung and weiterführende Fragestellungen	105

6.1 Einleitung

Ist das Schaltungsdesign einer integrierten Schaltung nahezu abgeschlossen, wird im nächsten Schritt die physikalische Implementierung - das Layout - erstellt. Dabei werden Bauteile und Leitungen in Form von geometrischen Primitiven abgebildet, die in weiterer Folge für die Erstellung der Masken verwendet werden.

Schaltungsextraktion stellt nun den umgekehrten Ablauf dar, in dem aus einer gegebenen physikalischen Repräsentation relevante Maskendaten extrahiert und daraus Rückschlüsse auf die Schaltung gezogen werden. Die nach dem Schichtabtrag und nach der Abbildung durch ein Mikroskop vorliegenden Bilderserien dienen hierbei als Datenbasis. Dabei sind einerseits die in den unteren Schichten realisierten Bauteile von Interesse, andererseits die in den Metallschichten befindliche Verdrahtung sowie die Kontakte zwischen den Ebenen. Die Kombination dieser topologischen Daten gestattet die Ableitung von schaltungsbezogenen Informationen und das Erstellen einer planaren Netzliste. Durch Auffinden von Hierarchien ist es zudem möglich, eine modulare Darstellung der Netzliste zu gewinnen. Eine im Hardwareentwurf oft verwendete Visualisierung dieser Zusammenhänge stellt das Y-Diagramm aus Abbildung 6.1 dar.

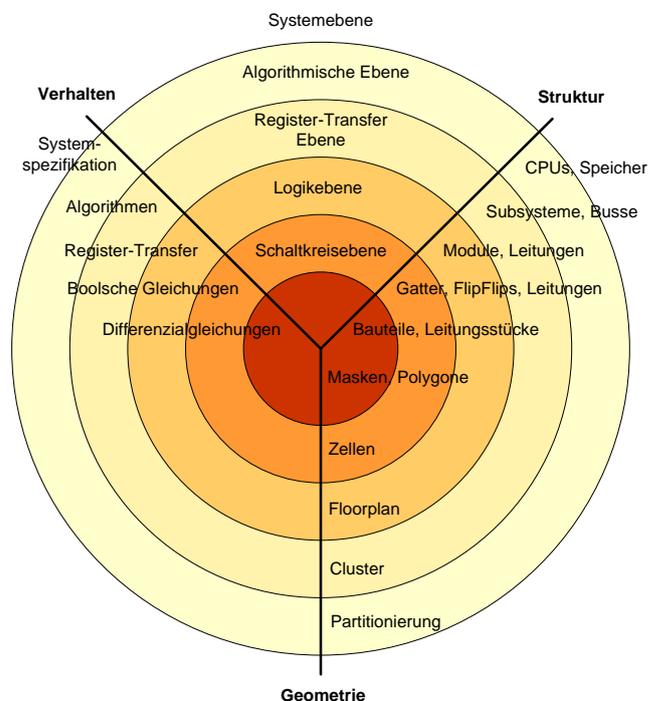


Abbildung 6.1: Das Y Diagramm [?] gibt verschiedene Sichtweisen auf den Hardwareentwicklungsprozess wieder. Es wird zwischen Verhalten, Struktur und Geometrie unterschieden. Im klassischen *forward engineering* (Abschnitt 1.2) geht man *top-down*, also von Außen nach Innen, vor. Im Gegensatz dazu geht man beim *reverse engineering bottom-up*, d.h. von Innen nach Außen, vor. Die hier vorgestellten Algorithmen sind in den beiden innersten Ringen angesiedelt.

Die Kontrolle von Prozessparametern und die Fehleranalyse sind integraler Bestandteil jeder Fertigungsschiene. Durch vertikale Schnitte durch den Mikrochip ist es möglich, Informationen über abgeschiedene Schichtdicken zu ermitteln und gegebenenfalls Prozessparameter dahingehend anzupassen, die gewünschten Eigenschaften zu erreichen. Durch diese Methoden werden nicht nur eigene Prozesstechnologien untersucht sondern auch je-

ne von anderen Mitbewerbern beurteilt. Dies dient einerseits zur Bestandaufnahme von Stärken und Schwächen der eigenen Verfahren sowie des Patentschutzes (siehe Abschnitt 1.1).

Viel Know-how steckt allerdings in Patenten, die schaltungstechnischer Herkunft sind. Diese zu überprüfen übersteigt jedoch die Möglichkeiten einer Technologieanalyse und fordert neue Analysemethoden, welche die Topologie des Mikrochips untersucht. Mit neuen Prozesstechnologien und den damit einhergehenden kleineren Strukturbreiten steigen die Kosten für eine Fertigungsschiene. War es früher üblich, dass jedes Halbleiterunternehmen eigene Fertigungswerke besaß, besteht heutzutage ein Trend hin zu einer fabriklosen (*fab-less*) Ausrichtung. Um hierbei das geistige Eigentum, welches einen Großteil des Unternehmenwertes ausmacht, zu schützen, wird in Zukunft vermehrt Schaltungsanalyse betrieben werden müssen.

Da nichtinversive Verfahren, wie Röntgenabbildungen durch das mangelnde Auflösungsvermögen ausscheiden, muss der Mikrochip wie in Abschnitt 2.3 beschrieben, Schicht für Schicht abgetragen werden. Die Strukturen von jeder für die Analyse notwendigen Schicht wird mit Hilfe der Mikroskopie auf eine Serie von Bildern übertragen. Um eine Analyse der dabei anfallenden Datenmengen, die hunderte Gigabyte umfassen, zu ermöglichen, bedarf es im ersten Schritt einer *Datenreduktion*. Durch Klassifikation des ursprünglich aus Pixel bestehenden Bildinhaltes in Regionen mit gemeinsamen Eigenschaften kann der wesentliche Bildinhalt kompakter dargestellt werden. Diese Klassifikation wird als Segmentierung bezeichnet und steht im Mittelpunkt dieses Kapitels.

6.2 Stand der Technik

Die Segmentierung von Bildern ist ein wichtiger Schritt in der Bildanalyse-Pipeline. Das Hauptziel besteht dabei in der Einteilung in zusammenhängende Segmente, die einen engen Zusammenhang mit den darauf abgebildeten realen Objekten aufweisen. Für eine Einführung in diese Thematik können [?, ?, ?] empfohlen werden.

Da das zugrundeliegende Problem der Segmentierung allgemeiner Natur ist, existieren eine Fülle von Methoden und Algorithmen, auf die zurückgegriffen werden kann. Es wird zwischen pixelbasierenden, regionenbasierenden, kantenbasierenden, modellbasierenden und texturorientierten Verfahren unterschieden. Die Klasse der pixelbasierenden Verfahren treffen für jeden Pixel die Entscheidung, zu welchem Segment er zuzuordnen ist. Kantenorientierte Verfahren nutzen die Eigenschaft, dass Segmente typischerweise durch Helligkeitsänderungen getrennt sind. Durch Ableitung des Bildes kann das Kantenbild gefunden werden, in welchem Kantenzüge Grenzen zwischen den Segmenten darstellen. Ziel der regionenorientierten Verfahren ist das Auffinden von homogenen, zusammenhängenden Regionen. Modellbasierende Ansätze versuchen durch Zusatzwissen über die sich im Bild befindlichen Objekte, diese zu finden. Zeichnen sich Segmente durch einheitliche Texturen aus, so können texturorientierte Verfahren, welche die Struktur als Homogenitätskriterium

verwenden, eingesetzt werden.

Es existieren eine Reihe von Arbeiten, die sich mit der Segmentierung von Metalllagen integrierter Schaltungen befassen. [?] und [?] verwenden ein simples globales Schwellwertverfahren und schließen dabei auftretende Löcher durch morphologische Operationen. Ein modellbasierender Ansatz, durch die Hough-Transformation, findet in [?] und [?] Anwendung. In [?, ?, ?] wird das Resultat eines Schwellwertverfahrens mit dem einer Kantenextraktion kombiniert, um die Robustheit zu steigern. Zur Ermittlung der Segmentgrenzen findet in [?] ein Kantenfilter Verwendung. Aufgrund einer Hystereseffunktion wird entschieden, ob es sich um eine relevante Kante handelt.

Anwendungen aus dem Bereich der Qualitätskontrolle, wie die automatische Überprüfung von Photomasken benötigen eine Kombination aus einem hoch auflösenden Abbildungssystem und hohem Durchsatz. SEM Abbilder besitzen zwar ein hohes Auflösungsvermögen, liefern jedoch bei kurzer Integrationszeit stark verrauschte Bilder. Standardverfahren zur Kantendetektion liefern bei Bildern mit niedrigem Signal-Rausch-Verhältnis viele Fehlerkennungen, die eine robuste Segmentierung erschweren. In [?] wird daher vorgeschlagen, Kanten durch einen *ridge signal* Detektor zu finden, da dieser typische REM Kanten besser nachbildet und *false positives* unterdrückt werden. In [?] werden verschiedene Typen von REM Abbildungen unterschieden. Bilder, die nur Kanteninformation beinhalten werden als 1D Abbildungen bezeichnet. Bilder, in denen Leitungen flächig erscheinen, also Kanteninformation, Vorder- und Hintergrundinformation enthalten, als 2D Bilder. 2D Bilder wurden dabei durch die *watershed transformation* segmentiert, 1D Bilder durch einen kantenbasierenden Ansatz. In [?] werden zwei Methoden zur Auffindung von VIAs durch *blob detection* und *template matching* vorgestellt.

6.3 Leiterbahnenextraktion

Eine Metallisierungsebene besteht in erster Näherung aus leitenden und isolierenden Materialien. Ziel einer Segmentierung dieser Art von Ebene ist das Erkennen der Topologie der leitenden Materialien. Dabei soll eine Klassifikation in leitende und isolierende Materialien vorgenommen und durch Regionen beschrieben werden. Diese klassische, jedoch für die Anwendung zu genaue Beschreibung, kann durch Polygone approximiert werden. Den Grad der Approximation und die damit verbundene Anzahl an Stützstellen soll dabei konfigurierbar sein, um an die Genauigkeitsanforderungen der jeweiligen Aufgabenstellung angepasst werden zu können.

Die Erstellung des Layouts ist an eine Reihe von Regeln (*design rules*) gebunden, um eine fehlerfreie Realisierung unter Einhaltung der Spezifikationen der Prozesstechnologie gewährleisten zu können. Diese Regeln beinhalten minimale Leiterbahnbreiten und Abstände zwischen diesen und können zur Verifikation des Extraktionsergebnisses genutzt werden. Weiters ist es üblich, Leiterbahnen nur mit Winkeln von 45° und 90° zu layouten. Dies ergibt acht mögliche Orientierungen, in der eine Leiterbahn verlaufen kann. Allerdings

ist dabei zu berücksichtigen, dass aufgrund der geringen Strukturgrößen Randeffekte bei diversen Herstellungsschritten zu verschliffenen Ecken führen. Auch kann beim Schichtabtrag oft nicht optimale Selektivität gewährleistet werden. Ein quadratisches Leitungsstück mit minimalen Abmessungen erscheint so z.B. meist als rundes Objekt.

Eine robuste Leiterbahnenextraktion setzt eine saubere Präparation voraus. Das Entfernen der Zwischen-Dielektrika (IMD) muss dabei über den Chip hinweg so homogen wie möglich erfolgen, um nicht mehrere Lagen gleichzeitig sichtbar zu machen. Abbildung 6.2 zeigt Ausschnitte eines Chips, in denen durch eine zu lang andauernde Ätzung darunterliegende Schichten freigelegt wurden. Zu entscheiden, ob eine Leiterbahn zur aktuell analysierenden Lage gehört, ist dadurch nicht immer möglich und führt in Folge zu Fehlerkennungen. Präparationsrückstände können weiters Kurzschlüsse zwischen Leiterbahnen verursachen und verhindern damit eine vollautomatische Layoutrekonstruktion.

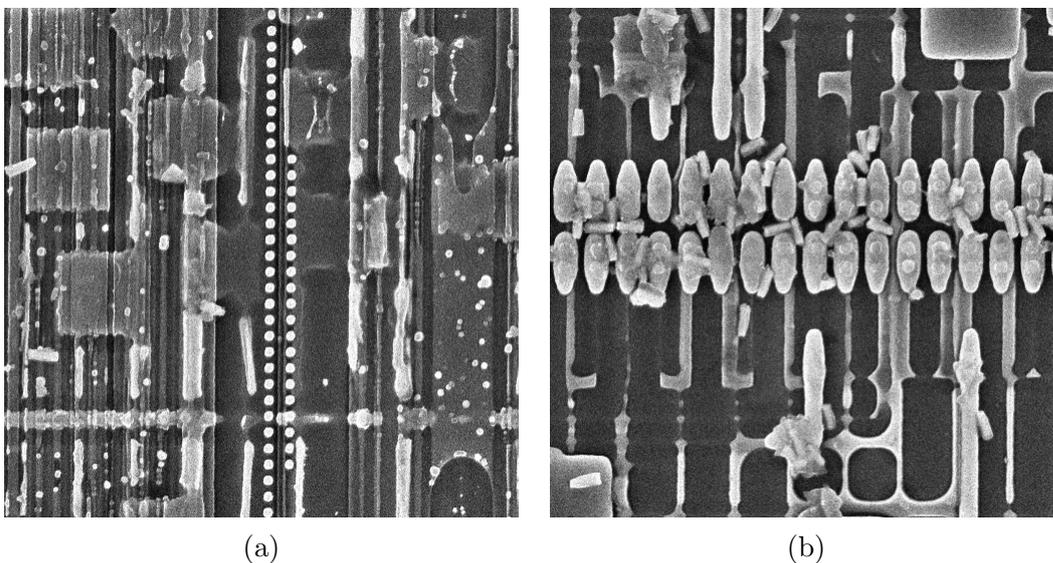


Abbildung 6.2: Präparationsartefakte durch Überätzung. Bei der Präparation der Probe wurden durch Überätzung darunterliegende Schichten sichtbar. Dies macht eine robuste Leitungsextraktion unmöglich. Weitere Probleme, welche durch Überätzung entstehen, sind abgetrennte Kontakte, die über den Chip verstreut zum Liegen kommen (Abbildung (b)). Da diese Kontakte ähnliche Materialeigenschaften aufweisen wie Leiterbahnen, sind diese nur schwer von Leitungen zu unterscheiden und verursachen Kurzschlüsse zwischen benachbarten Leiterbahnen. Mit dem Wissen der dem Prozess zugrundeliegenden *design rules* können Kurzschlüsse zwar identifiziert werden, jedoch bedarf es in der Regel manueller Nacharbeit um diese zu beseitigen.

Um auch Schaltungsanalysen von neueren Prozesstechnologien durchführen zu können, wurden für die vorliegende Arbeit die Lagen durch ein REM abgebildet. Die Aufnahmen weisen auch bei kleinen Technologiegrößen ausreichende geometrische Auflösung auf und bilden dabei nur die oberen Atomsschichten ab.

Abbildungen eines SEM zeigen eine spezielle Charakteristik, die durch die Abbildungs-

technik bedingt ist. An kleinen Erhebungen und Vertiefungen ruft der Elektronenstrahl eine verstärkte Emission von Elektronen hervor, da von mehreren Seiten Elektronen in Richtung Detektor abgezogen werden. Dies führt zu einer Betonung von Kanten an Erhebungen und Vertiefungen. Da bei einem selektiven Ätzvorgang genau diese Erscheinungen zu Tage treten, wirkt der Kanteneffekt positiv auf den Kontrastunterschied zwischen den verschiedenen Materialien. Diese Eigenheit von REM Aufnahmen wurde bei der Bildanalyse berücksichtigt, um optimale Ergebnisse erzielen zu können.

Ziel der Segmentierung des Bildinhaltes ist eine Klassifikation in leitende und nichtleitende Bereiche. Um die Komplexität des anschließenden *routing* zu reduzieren und um eine Schnittstelle zu Standardsoftware herzustellen, werden die gefundenen leitenden Regionen in eine approximative polygonale Beschreibung übergeführt. Damit besteht die Möglichkeit, das Layout in ein Standardformat der Industrie wie GDSII (*Graphical Design Station II*) oder DEF (*Design Exchange Format*) zu exportieren. Auf die einzelnen Schritte der Bildanalyse, um aus den vorliegenden Eingabebildern zu deren Vektorbeschreibung zu gelangen, soll nun näher eingegangen werden.

6.3.1 Bildvorverarbeitung

Aufnahmen eines REM haben ein sehr schlechtes Signal-zu-Rausch Verhältnis (SNR). Nur durch Integration vieler Aufnahmen ist es möglich ein höheres SNR zu erzielen. Das Spektrum des verbleibenden additiven Rauschanteils entspricht durch seine vielfältige und zufällige Natur einer Gaußverteilung und ist unkorreliert. Eine einfache Möglichkeit diesen Rauschanteil zu unterdrücken, stellt eine Tiefpaßfilterung durch einen gauß'schen Faltungskern dar. Dabei muss jedoch beachtet werden, dass nicht nur der Rauschanteil hochfrequente Anteile in das Spektrum des Bildes einbringt. Starke Kanten enthalten ebenfalls hohe Frequenzen, die bei der Entrauschung gedämpft werden. Dies führt zu einer kleiner werdenden Antwort eines Kantentfilters und damit zu einer möglichen Nichterkennung der Kante, falls die Schwellwerte der Algorithmen zu niedrig gesetzt sind. Sinnvoll erscheint die Standardabweichung σ_g sowie die Größe des Glättungskerns als Funktion der minimal im Bild vorkommenden Strukturgröße zu definieren. Enthält ein Bild wesentliche Strukturen, die nur wenige Pixel ausgedehnt sind, sollte ein kleineres σ_g gewählt werden um die Strukturgrenzen nicht zu sehr zu dämpfen. Ein interessanterer Ansatz um kantenerhaltend zu filtern wird in [?] vorgestellt. Dabei wird vorgeschlagen, dass neben der räumlichen Distanz auch die Farbdistanz zur Berechnung des neuen Farbwertes herangezogen werden sollte.

Für manche Algorithmen ist es sinnvoll, die Bilder in der Vorverarbeitung einer Kontrastspreizung (Histogrammstreckung) zu unterziehen und damit zu gewährleisten, dass der gesamte Wertevorrat an Pixelintensitäten ausgenutzt wird. Mit dieser Voraussetzung kann mit absoluten, globalen Schwellwerten gearbeitet werden. Eine weitere Methode den Kontrast zu verbessern, ist die Histogrammäqualisation (*histogram equalization*), welche über eine Reskalierung der Intensitäten hinausgeht. Im Gegensatz zur Hi-

stogrammstreckung wird das Histogramm mit einer nichtlinearen Übertragungsfunktion beaufschlagt, die eine Gleichverteilung der Intensitäten anstrebt. Zu beachten ist allerdings, dass beide Verfahren auch Intensitätsrauschen verstärken, weshalb das Bild zuvor durch einen Tiefpaß gefiltert werden sollte.

6.3.2 Segmentierung

Sind die Bilder weitgehend entrauscht, kann mit der eigentlichen Segmentierung des Bildinhaltes begonnen werden. Da die Literatur zu dieser speziellen Anwendung relativ wenig bereitstellt, wurden verschiedene Segmentierungsansätze evaluiert und die Ergebnisse miteinander verglichen. Als Ergebnis liefern die Methoden eine Beschreibung des Grenzverlaufes zwischen den Segmenten. Idealerweise sollten diese Segmente geschlossen sein, um eine Klassifikation in leitende/nichtleitende Segmente treffen zu können. Aufnahmen von Metalllagen durch ein REM zeigen je nach Präparationstechnik ein sehr unterschiedliches Erscheinungsbild. Teilweise sind Leitungen sehr flächig zu sehen, teilweise nur die Begrenzungskanten der Leiterbahnen. Aufgrund dieser Tatsache sind erhöhte Anforderungen an die Segmentierungsmethode gestellt, da verschiedene Fälle abgedeckt werden müssen. Ein kantenbasierender Ansatz hat im Vergleich zu einem regionenbasierenden Ansatz den Vorteil, eine höhere Erscheinungsvielfalt abzudecken und wird deshalb in dieser Arbeit Vorzug gegeben. Die geometrischen Eigenschaften der Leiterbahnen legen ebenfalls einen modellbasierenden Ansatz wie z.B. die *hough transformation* nahe. Die Evaluation der beiden Ansätze wird im Folgenden beschrieben.

6.3.2.1 Kantenbasierte Segmentierung

Abgrenzungen zwischen Leiter wie Aluminium oder Kupfer und Isolatoren wie SiO_2 zeigen aufgrund ihrer unterschiedlichen atomaren Eigenschaften hohen Kontrast in REM Aufnahmen. Der hohe Kontrast in den Übergängen, der durch den Kanteneffekt des REM noch verstärkt wird, kann ausgenutzt werden um Leiterbahnen von isolierenden Bereichen zu trennen. Zum Auffinden dieser Übergänge, die im Bild als Kanten in Erscheinung treten, existieren eine Vielzahl an Standardverfahren. Der Ansatz in [?] gilt noch heute als einer der besten und robustesten Kantendetektoren für sprunghafte, mit gauß'schen Rauschen behaftete Kanten (*step edges*). Motiviert wurde die Kantendetektion von Untersuchungen der menschlichen Wahrnehmung in der Neurophysiologie. Die Idee des *raw primary sketch* geht auf [?] zurück und stellt eine frühe Stufe der menschlichen visuellen Wahrnehmung dar. Der Canny Kantendetektor wurde im Hinblick auf die Optimierung folgender Kriterien entworfen:

1. *Detektion*. Die Wahrscheinlichkeit für die Nichterkennung einer echten Kante sowie die Wahrscheinlichkeit für eine Detektion einer falschen Kante soll minimiert werden. Da beide Wahrscheinlichkeiten monoton steigende Funktionen des Signal-zu-Rausch (SNR) Verhältnisses sind, entspricht dieses Kriterium einer Maximierung des SNR.

$$SNR = \frac{|\int_{-W}^W \mathbf{G}(-x)\mathbf{h}(x)dx|}{n_0\sqrt{\int_{-W}^W \mathbf{h}^2(x)dx}} \quad (6.1)$$

2. *Lokalisierung.* Die durch den Operator gefundene Kante sollen so nah wie möglich an der Mitte der richtigen Kante sein.

$$Localization = \frac{|\int_{-W}^W \mathbf{G}'(-x)\mathbf{h}'(x)dx|}{n_0\sqrt{\int_{-W}^W \mathbf{h}'^2(x)dx}} \quad (6.2)$$

3. *Anzahl der Antworten.* Eine richtige Kante soll nicht mehrfach erkannt werden.

Zu Beginn des mehrstufigen Algorithmus steht eine Rauschunterdrückung wie in Abschnitt 6.3.1 beschrieben. Bei der Größe des Faltungskerns muss ein Kompromiss zwischen Kantenerhaltung und Rauschunterdrückung gefunden werden. In der Praxis liefern Kerne mit der Größe 5×5 gute Ergebnisse.

Nach der Glättung des Bildes werden die Gradienten des Bildes berechnet. Sei das Bild eine stetige Funktion \mathbf{I} dann ist die Ableitung gegeben durch

$$\nabla \mathbf{I} = \left(\frac{\partial \mathbf{I}}{\partial x} \quad \frac{\partial \mathbf{I}}{\partial y} \right)^T \quad (6.3)$$

Der Betrag $|\nabla \mathbf{I}|$ des Gradienten gibt dabei die Stärke der Steigung an, der Winkel $\phi = \arctan\left(\frac{\partial \mathbf{I}/\partial x}{\partial \mathbf{I}/\partial y}\right)$ die Richtung. Da es sich allerdings um eine diskrete Funktion \mathbf{I} handelt, muss der Gradient in Gleichung 6.3 durch Differenzbildung approximiert werden. Daher gilt

$$\frac{\partial \mathbf{I}}{\partial x} \approx \frac{\mathbf{I}(x+n, y) - \mathbf{I}(x-n, y)}{2n} \quad (6.4)$$

sowie

$$\frac{\partial \mathbf{I}}{\partial y} \approx \frac{\mathbf{I}(x, y+n) - \mathbf{I}(x, y-n)}{2n} \quad (6.5)$$

Der Wert von n muss dabei klein genug gewählt werden, um die Ableitung gut zu approximieren jedoch groß genug, um nichtrelevante Änderungen im Bild zu vernachlässigen. Üblicherweise wählt man $n = 1$. Die Gleichungen 6.4 und 6.5 können durch eine Faltung mit den Kernen $\mathbf{h}_x = [-1, 0, 1; -1, 0, 1; -1, 0, 1]$ und $\mathbf{h}_y = \mathbf{h}_x^T$ realisiert werden. Um jedoch bessere Isotropie zu erzielen, wird der Sobel Operator [?] mit $\mathbf{h}_x = [-1, 0, 1; -2, 0, 2; -1, 0, 1]$ verwendet. Nach der Faltung mit den Eingabebildern erhält man die beiden Kantenbilder

$$\mathbf{G}_x = \mathbf{I} \star \mathbf{h}_x \quad (6.6)$$

$$\mathbf{G}_y = \mathbf{I} \star \mathbf{h}_y. \quad (6.7)$$

Die Kantenstärken ergeben sich aus

$$|\mathbf{G}| = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}. \quad (6.8)$$

Die Kantenorientierungen können durch

$$\angle \mathbf{G} = \arctan\left(\frac{\mathbf{I}_y}{\mathbf{I}_x}\right) \quad (6.9)$$

berechnet werden.

Algorithmus 1 Non-maximum suppression

Eingabe: Kantenbeträge $|\mathbf{G}|$.

Ausgabe: lokale Maxima von $|\mathbf{G}|$.

- 1: Berechnung der Kantenrichtungen für jeden Pixel mit einem Gradientenbetrag größer ϵ nach Gleichung 6.9. Im Folgenden werden diese Pixel als Kantenpixel bezeichnet.
 - 2: Quantisierung der Kantenrichtungen in acht Schritten entsprechend der 8-Pixel-Nachbarschaft.
 - 3: Der Betrag jedes Kantenpixels wird mit den zwei Beträgen der normal zum Kantenpixel anliegenden Pixel verglichen. Ist einer der Beträge größer, gehört der aktuelle Pixel nicht zum Maximum und dessen Betrag wird deshalb auf 0 gesetzt.
-

Durch Setzen eines Schwellwertes wäre es nun möglich zu bestimmen, ab wann eine Kantenantwort als Kante interpretiert werden soll. Wie jedoch in Abbildung 6.3(b) ersichtlich, besteht eine Kantenantwort auf eine nicht ideale sprunghafte Kante aus einem ausgedehnten Gebirgskamm. Per Definition steht der Gradient normal zur Kante und sollte in der Nähe der Mitte den größten Betrag aufweisen. Die Methode, diese lokalen Maxima aufzufinden, nennt man *non-maximum suppression* (NMS). Wenn Echtzeitfähigkeit gefordert ist, können effiziente Algorithmen, wie in [?] beschrieben, eingesetzt werden, die den Standardalgorithmus um den Faktor drei bis vier beschleunigen. Da die Problemstellung jedoch keinen Anspruch auf Echtzeitfähigkeit stellt und die NMS nur einen kleinen Bruchteil des Gesamtberechnungsaufwandes ausmacht, wurde in der vorliegenden Arbeit auf einen Standardansatz zurückgegriffen. Mit der Beobachtung, dass das Maximum des Gradientenbetrages entlang des Kamms verläuft und der Betrag vom Maximum normal zur Kantenrichtung abnimmt, kann Algorithmus 1 abgeleitet werden.

Algorithmus 2 Hysterese der Kantenantwort

Eingabe: lokale Maxima in Kantenrichtung von $|\mathbf{G}|$.

Eingabe: Gradientenbetragsschwellwerte Th_1, Th_2 .

Ausgabe: Binärbild aller relevanten Kantenpixel.

- 1: Alle Kantenpixel mit einem Gradienten größer als Th_2 werden als relevante Kantenpixel markiert.
 - 2: Alle Kantenpixel mit einem Gradienten im Intervall $[Th_1, Th_2]$ gelten als relevante Kantenpixel, wenn sie zu einem relevanten Kantenpixel benachbart sind (8-Pixel-Nachbarschaft).
 - 3: Wiederhole Schritt 1-2 bis sich keine Änderungen mehr ergeben.
-

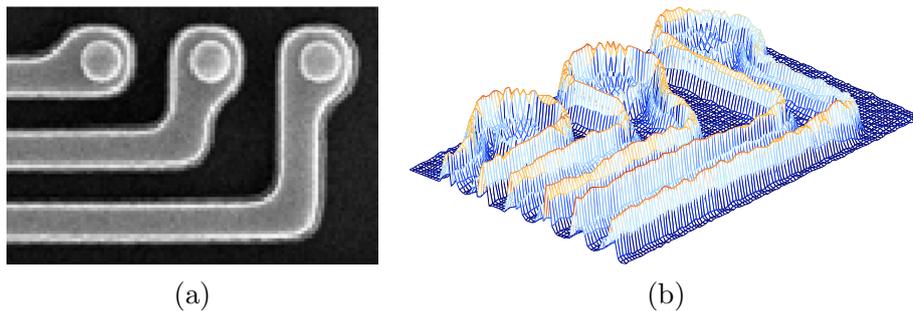


Abbildung 6.3: 3D Kantenprofil von Leitungen unter einem REM. In Abbildung (a) ist eine Aufnahme von drei Leitungsenden zu sehen. Abbildung (b) zeigt das Kantenbild, also die erste Ableitung des Bildes. Die Höhe des Kantenkamms und damit die maximale Steigung der Kante kann innerhalb eines Kamms variieren. Durch Hysterese konnten diese Schwankungen berücksichtigt werden. Weiters kam es durch den für REM typischen ausgeprägten Kanteneffekt oft zu Mehrfacherkennung einer Kante. Dies könnte durch geeignete Maßnahmen unterdrückt werden.

Das resultierende Gradientenbild enthält nun nur mehr Kanten, die eine Breite von einem Pixel aufweisen und damit die Grenzen der Segmente präzise wiedergeben. Da jedoch nur Kanten ab einer gewissen Stärke als Segmentgrenze in Frage kommen, wird noch ein Schwellwert benötigt, anhand dessen diese Unterscheidung getroffen werden kann. Um das Aufbrechen von Kanten durch Fluktuationen um den Schwellwert zu vermeiden, wird von [?] vorgeschlagen, eine Hysterese zu verwenden. Dieser, durch das Toleranzintervall $[Th_1, Th_2]$ definierte Bereich erlaubt Schwankungen innerhalb eines Kantenzuges und dient damit der Robustheit gegenüber nichtidealer Einflüsse wie Intensitätsrauschen. Brechen Kanten trotz Hysterese auf, können Methoden aus der mathematischen Morphologie genutzt werden, um diese zu schließen. In Algorithmus 2 wird der verwendete Ansatz dargestellt und Algorithmus 3 fasst die Schritte zum Auffinden von Kanten anhand des Kantendetektors nach Canny zusammen.

Algorithmus 3 Canny Kantendetektion

Eingabe: Intensitätsbild \mathbf{I}

Ausgabe: Binärbild der relevanten Kanten

- 1: Tiefpassfilterung von \mathbf{I} mit einem Gaußkern wie in Abschnitt 6.3.1 beschrieben.
 - 2: Bildung des Gradienten \mathbf{G}_x und \mathbf{G}_y des Bildes \mathbf{I} durch Gleichung 6.6 und 6.7.
 - 3: Auffinden der lokalen Maxima durch *non-maximum suppression* (Algorithmus 1).
 - 4: Robuste Klassifizierung der relevanten Kantenpixel anhand einer Hysterese (Algorithmus 2).
-

Das Resultat der Segmentierungsphase ist ein Binärbild, welches idealerweise leitende und nichtleitende Strukturen durch ein Pixel breite Konturen abgrenzt (Abbildung 6.4). Um Leiterbahnen jedoch durch geschlossene Pixelzüge darstellen zu können, bedarf es einer Interpretation der Konturgrenzen sowie etwaiger Knotenpunkte. Um diese Einteilung

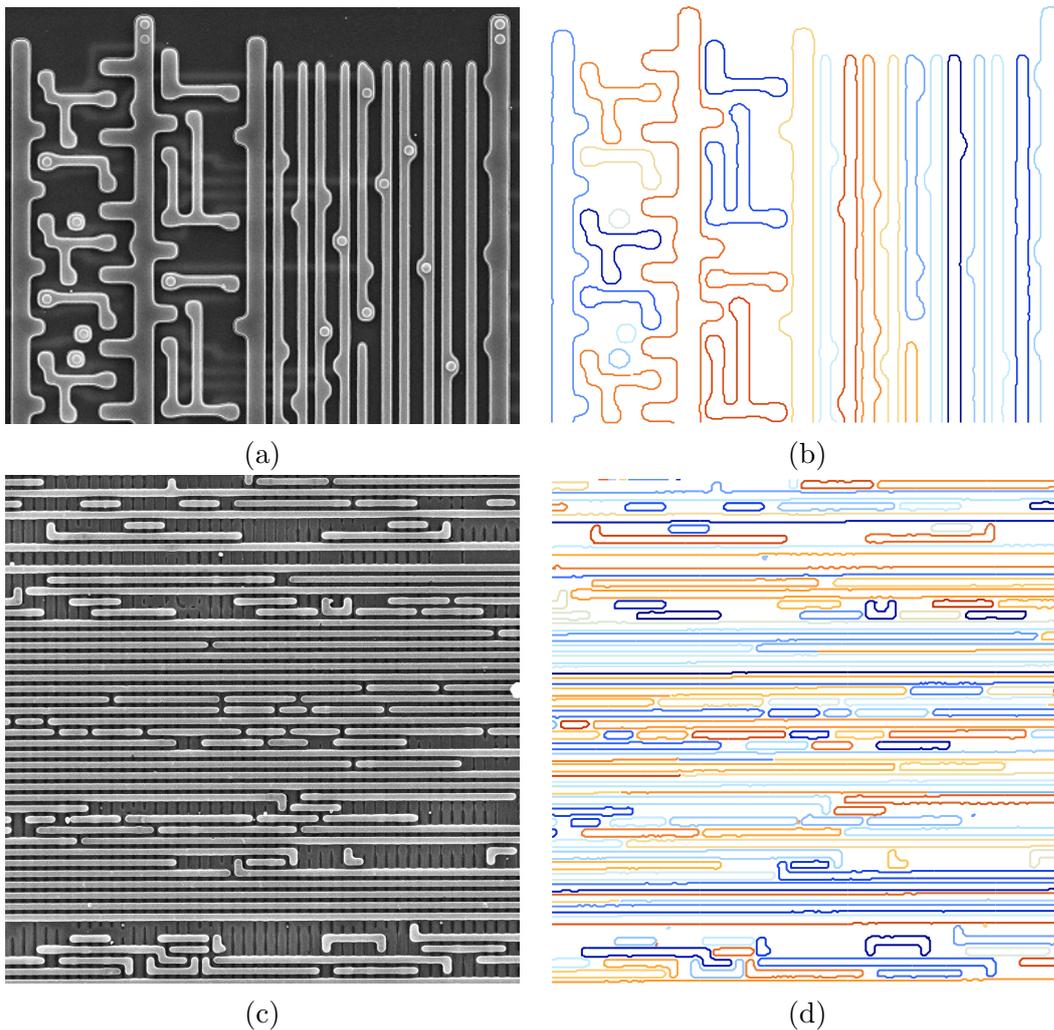


Abbildung 6.4: Nach der Kantenextraktion erfolgt die eindeutige Benennung der Segmente. Die in Abbildung (b) und (d) ersichtlichen Segmente wurden dabei farblich codiert.

zu treffen, reicht es aus, die 8-Pixel-Nachbarschaft jedes Kantenpixels zu untersuchen. Eine Klassifikation in Kantenendpixel, Kantenknotenpixel und normaler Kantenpixel kann anhand der 8-Nachbarschaft getroffen werden. Ein normaler Kantenendpixel ist dadurch gekennzeichnet, dass sich in seiner Nachbarschaft genau zwei $1 \rightarrow 0$ bzw. $0 \rightarrow 1$ Übergänge befinden. Ein normaler Kantenpixel kommt auf genau vier und ein Kantenknotenpixel auf sechs oder acht. Befindet sich in der Nachbarschaft kein Übergang, handelt es sich um einen isolierten Kantenpixel und kann ignoriert werden.

Sei $\mathbf{x}(n)$ mit $n = 0 \dots 7$ ein Vektor, der die Nachbarschaftspixel im Uhrzeigersinn enthält, dann ist die Summe des Absolutbetrages seiner Ableitung (Differenzbildung) gleich der Anzahl der Übergänge. Algorithmus 4 berechnet für ein gegebenes Binärbild eine Liste an Knotenpunkten und Endpunkten, die für die anschließende Kantenverfolgung not-

Algorithmus 4 Klassifikation der Kantenpixel

Eingabe: Binärbild \mathbf{B} **Ausgabe:** Kantenknotenliste, Kantenendpunktliste

```

1: for all  $\mathbf{B}(x, y) = 1$  do
2:    $\mathbf{v}(n) = \text{getNeighborhoodPixels}(\mathbf{B}, x, y)$ 
3:    $\text{transitions} = |\text{diff}(\mathbf{v}(n))|$ 
4:   if  $\text{transitions} = 6$  or  $\text{transitions} = 8$  then
5:      $(x, y)$  zur Kantenknotenliste hinzufügen
6:   else if  $\text{transitions} = 2$  then
7:      $(x, y)$  zur Kantenendpunktliste hinzufügen
8:   else
9:     Kantenpixel in  $(x, y)$  ist normaler Kantenpixel
10:  end if
11: end for

```

wendig sind. Ziel der Kantenverfolgung ist die Benennung der einzelnen Kantenzüge, um zusammenhängende Kantenpixel zu finden. Anfang und Endpunkt eines Kantenzuges sind per Definition entweder ein Kantenknotenpunkt oder Kantenendpunkt nach Algorithmus 4.

Die Kantenverfolgung durchläuft jeden noch nicht klassifizierten Kantenpixel und sucht nach allen damit verbundenen Kantenpixel, bis auf einen Knoten oder Endpixel gestoßen wird. Diese, Algorithmus 5 folgende Strategie vergibt für jeden Kantenzug eine Identifikationsnummer, anhand welcher dieser eindeutig bestimmt ist. Damit sind zusammenhängende Kanten eindeutig bestimmt und können weiter verarbeitet werden. Handelt es sich um einen geschlossenen Kantenzug, so handelt es sich mit hoher Wahrscheinlichkeit um ein in sich geschlossenes Leitungsstück. Grenzen Leitungen an den Rand eines Einzelbildes, so ist nicht immer klar, wo der offene Kantenzug zu schließen ist. Dies hängt sehr stark von der Präparation der Ebene ab. Bei flächigem Erscheinungsbild der Leiterbahn kann das Bild nach außen hin durch einen Rahmen mit mittlerer Hintergrundhelligkeit ergänzt werden. Eine anschließende Tiefpassfilterung glättet dabei auftretende Unstetigkeiten. Nun ist die Kantensuche imstande, auch an den Bildrändern geschlossene Kantenzüge zu finden.

Abbildung 6.5 zeigt eine so extrahierte Lage als Übersicht und Abbildung 6.6 Ausschnitte dieser.

Ist das Bild präparationsbedingt (z.B. durch Sputtern eines elektrisch leitenden Materials auf die Oberfläche) sehr stark kantenlastig geprägt, so entstehen Probleme am Bildrand, da Leitungen nicht vom isolierenden Oxid zu unterscheiden sind. Damit sind an den Bildrand grenzende Leitungen nicht geschlossen und werden verworfen. Ein Beispiel ist in Abbildung 6.7 aufgezeigt.

Um die Kontur dieser Leitungen schließen zu können, muss bekannt sein, ob sich die Leiterbahn links oder rechts von der Kante befindet. Eine Analyse des Kantenprofils nor-

Algorithmus 5 Kantenverfolgung

Eingabe: Binärbild \mathbf{B} , Kantenknotenliste \mathbf{j} , Kantenendpunktliste \mathbf{e} **Ausgabe:** Labels \mathbf{L}

```

1:  $\mathbf{L} \leftarrow \mathbf{0}$ 
2: for all  $\mathbf{B}(x, y) = 1$  and  $\mathbf{L}(x, y) = 0$  do
3:   repeat
4:      $\mathbf{x}(n) = \text{getNeighborhoodCoordinates}(x, y)$ 
5:     if  $\exists n \mathbf{x}(n) \in (\mathbf{j} \cup \mathbf{e})$  then
6:       break
7:     end if
8:      $\mathbf{v}(n) = \text{getNeighborhoodPixels}(\mathbf{B}, x, y)$ 
9:     for  $i = 0$  to  $7$  do
10:      if  $\mathbf{v}(i) = 1$  then
11:        return  $i$ ;
12:      end if
13:    end for
14:   until
15: end for

```

mal zur Kante, wie in Abbildung 6.8 zeigt, dass Kanten hin zur Leiterbahn meist etwas langsamer abfallen. Eine eindeutige Einteilung ist aufgrund dieser Feststellung jedoch nicht zu treffen. Was weiters auffällt, ist der in Abschnitt 3.3.2 beschriebene REM typische Abschattungseffekt. Ränder einer Leiterbahn, welche dem Detektor abgewandt sind, weisen im Profil ein zusätzliches Minimum in der Nähe des Maximums auf. Dieses Kriterium wurde in dieser Arbeit verwendet um die Leiterbahn zugeneigte Seite zu ermitteln. Als problematisch erweisen sich jedoch Kanten, die in der Richtung des Detektors verlaufen und damit keinen Abschattungseffekt aufweisen.

Eine mögliche Fehlerquelle der vorgestellten kantenbasierenden Segmentierungsmethode sind die an die Leiterbahn grenzenden VIAs, die den Kantenzug in die „irre führen“ können und damit eine fehlerhafte Beschreibung des Umrisses liefern. Gilt der Kantenzug als nicht geschlossen, d.h. sind Kantenanfangs- und endpunkt nicht unter einer minimalen Distanz entfernt, so wird dieser verworfen.

Abhilfe können dabei alternative Segmentierungsmethoden oder entsprechende Berücksichtigung bei der Präparation schaffen. Allgemein kann man sagen, dass es sich schwierig gestaltet, eine generalisierte Segmentierungsmethode zu entwickeln, die alle möglichen Variationen abdecken kann. Es erscheint sinnvoller, spezialisierte Methoden je nach Bildtyp anzuwenden.

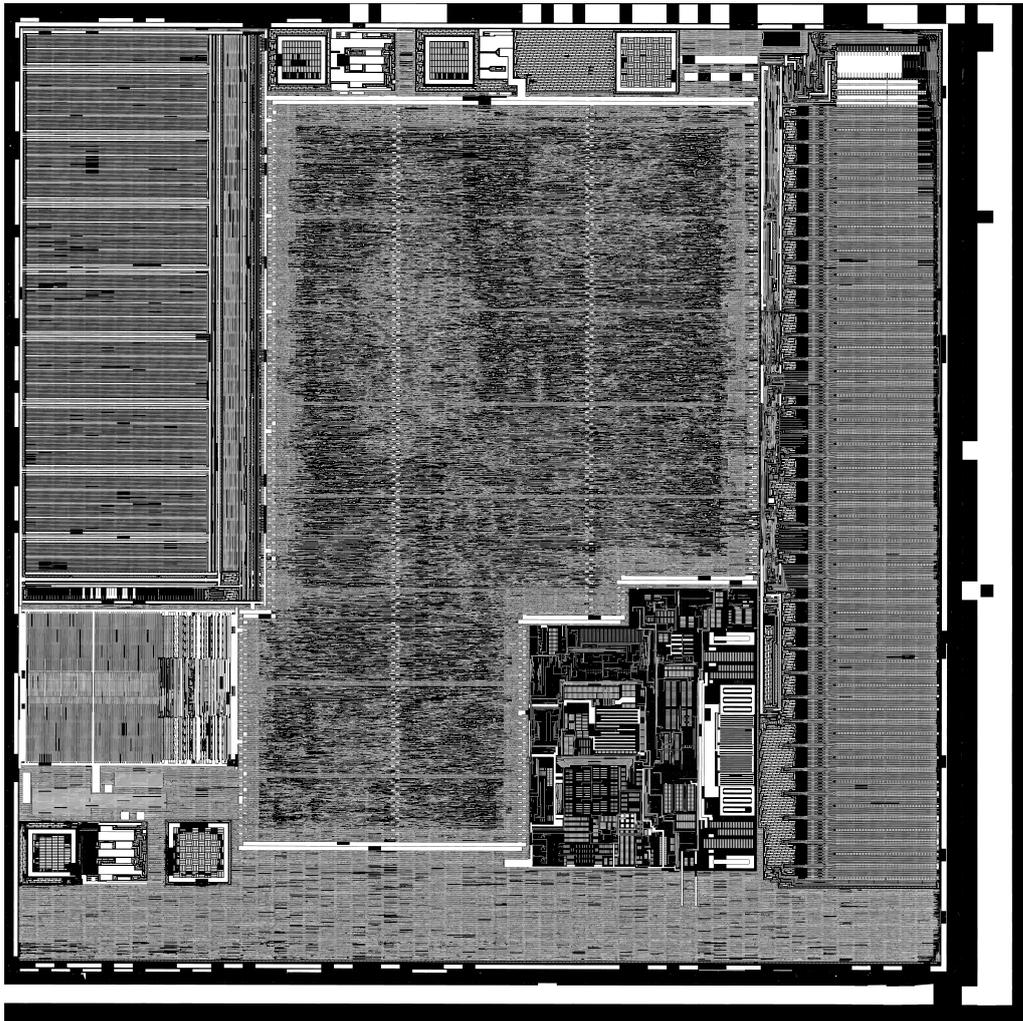


Abbildung 6.5: Leitungsextraktion einer Metallebene. Auf einem aus 4098 Einzelbilder bestehenden Mosaik einer Metallisierungsebene mit einer Fläche von $\approx 2\text{mm}^2$ wurden Leitungen extrahiert und ein polygonales Modell mit 264387 Polygonen generiert. Nach einem Datelexport in ein Standard-Layoutformat, kann das polygonale Modell in Standardsoftware der Industrie, wie Cadence Virtuoso importiert werden. Ein *Intel Core i5-750* Prozessor benötigte für Berechnung 5.6 Stunden.

6.3.2.2 Modellbasierte Segmentierung

Da es sich bei den zu untersuchenden Nanostrukturen um bekannte geometrische Objekte handelt, die zwar eine große Vielfalt ausweisen, dessen Primitive sich jedoch gleichen, kann durch Modellierung der geometrischen Zusammenhänge eine Methode zur Extraktion entworfen werden. Die Hough-Transformation in ihrerer klassischen Form [?, ?] bietet die Möglichkeit, parametrisierte Kurven in Binärbilder aufzufinden. Die generalisierte Hough-Transformation [?]

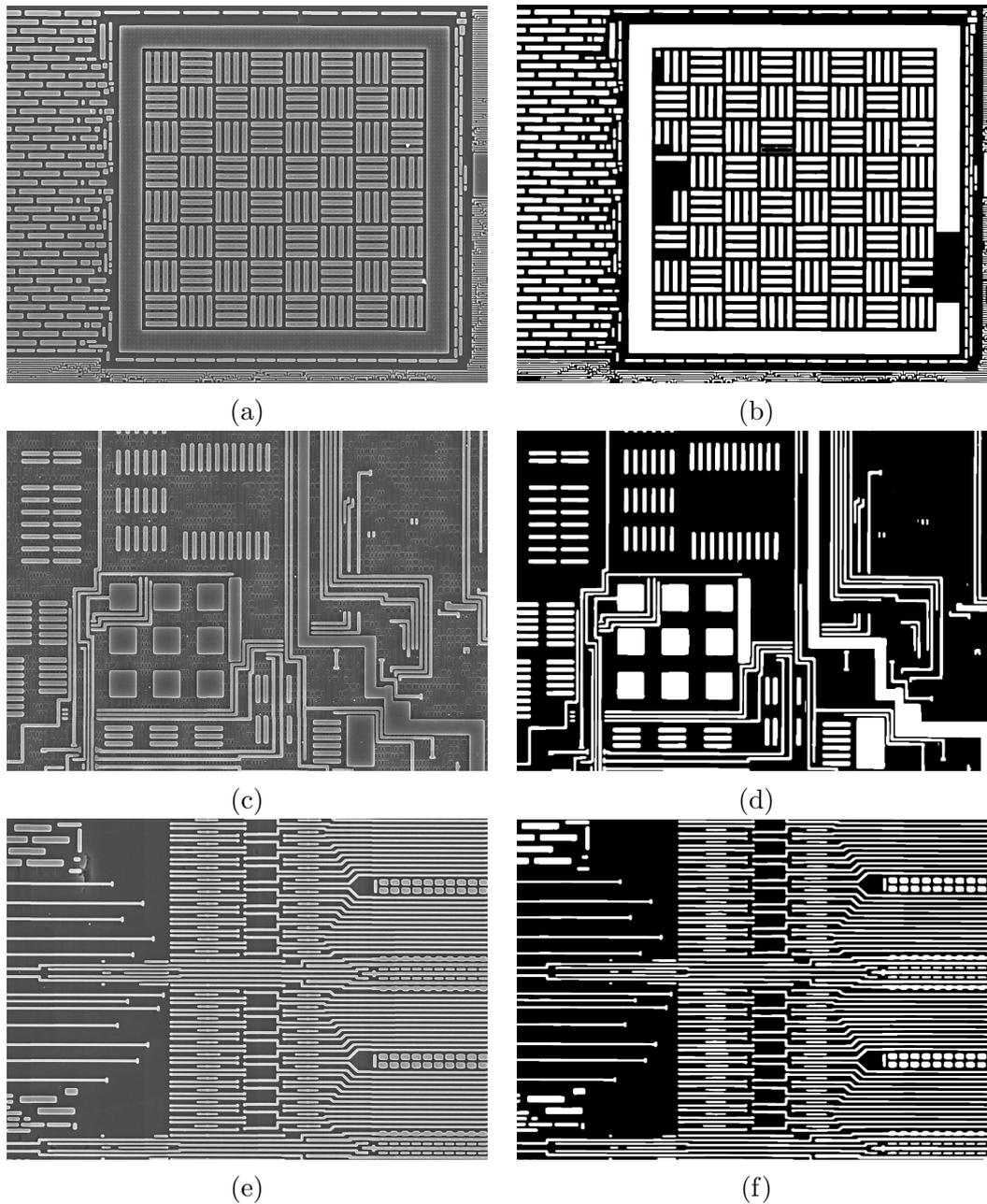


Abbildung 6.6: Leiterbahnenextraktion aus verschiedenen Chipbereichen. Die rechten Abbildungen stellen das aus den Polygondaten rekonstruierte Layout dar. Abbildung (a) bzw. (b) zeigen ein Anschlusspad, Abbildung (c) und (d) die Verdrahtung von 8+1 Bipolartransistoren einer Bandgap-Struktur und Abbildung (e) bzw. (f) Adressleitungen von Flash Speicherzellen.

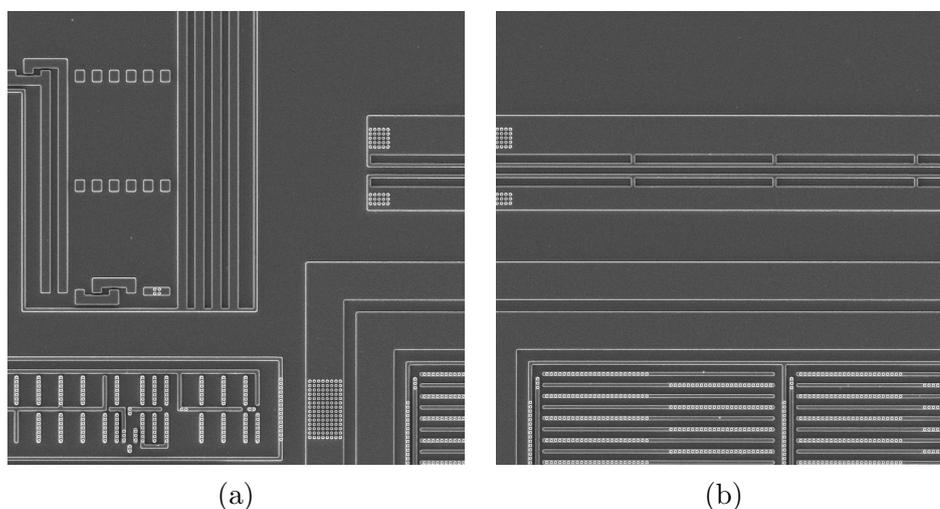


Abbildung 6.7: Probenpräparation durch Sputtern. Damit sich nichtleitende Proben durch den Elektronenbeschuss nicht positiv aufladen, kann durch Sputtern eine dünne leitende Schicht aufgebracht werden. Dies wirkt sich aber negativ auf den Kontrast zwischen Leiter und Nichtleiter aus. Die Topologie wird dabei rein durch den Kanten effekt der REM abgebildet, welches eine eindeutige Bestimmung von leitenden/nichtleitenden Segmenten erschwert. Die an den Bildrand grenzenden Leitungen sind durch den fehlenden Kontrast zum Hintergrund (Nichtleiter) nicht abgegrenzt und liefern bei einer kantenbasierenden Segmentierung nicht geschlossene Segmente.

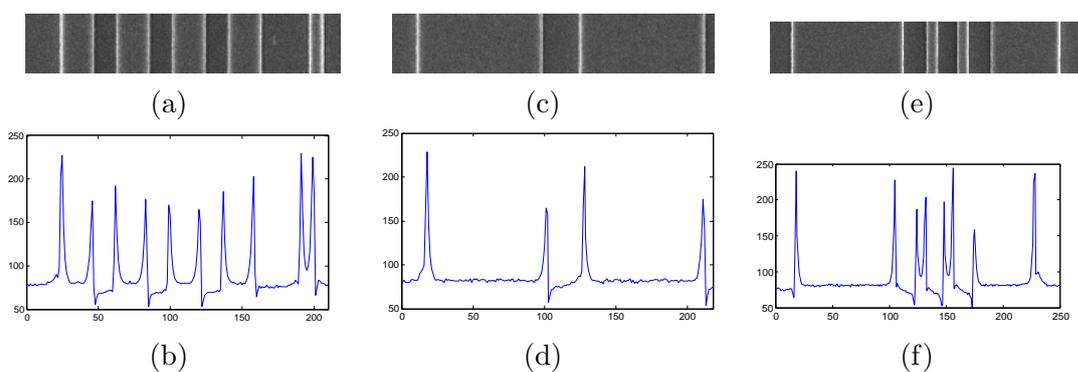


Abbildung 6.8: Analyse des Kantenprofils. Abbildung (a) und (c) zeigen vertikale Leitungen aus Abbildung 6.7. In Abbildung (b) und (d) sind die zugehörigen Kantenprofile zu sehen. Auffällig sind die nach den Maxima befindlichen Minima, welche von Elektronenabschattungen verursacht werden. Abbildung (e) zeigt (um 90° gedrehte) horizontale Leitungen, welche in Abbildung (f) gleiche Eigenschaften zeigen. Nach dieser Beobachtung befand sich der Detektor bei der Aufnahme links oberhalb der Probe.

verallgemeinert diesen Ansatz auf beliebige Formen sowie auf Grauwertbilder. In dieser Arbeit soll die klassische Hough-Transformation evaluiert werden, um dessen Leistungsfähigkeit im Bezug auf die Leiterbahnenextraktion abschätzen zu können.

Ausgangspunkt für die Segmentierung bildet ein durch Algorithmus 3 berechnetes binäres Kantenbild \mathbf{B} . Ziel des Segmentierungsschrittes ist wiederum das Finden einer Einteilung in leitende und nichtleitende Gebiete. Wie schon in Abschnitt 6.3 erwähnt, werden Leitungszüge durch Leitungsstücke realisiert, die in 45° und 90° Winkeln relativ zueinander orientiert sein können. Durch Auffinden von Geraden bestimmter Orientierungen und einer darauffolgenden Klassifikation der durch die Schnittpunkte der Geraden definierten rechteckigen Gebiete kann eine Einteilung getroffen werden. Die Grundidee der klassischen Hough-Transformation ist die Überführung eines Binärbildes in einen durch die Modellparameter aufgespannten Raum (Parameterraum) \mathbf{A} . Dabei werden für jeden Kantenpixel $\mathbf{B}(x, y) = 1$ die Parameter des Modells ermittelt, diskretisiert und der Wert an der betreffenden Koordinate des als Akkumulator dienenden Parameterraums inkrementiert. Hohe Werte im Parameterraum entsprechen somit Parametersätzen, welche durch die Gleichungen des Modelles häufig erfüllt werden. Aufgrund der in erster Näherung rechteckigen Geometrie von Leiterbahnen wird versucht, die Segmentgrenzen durch Geraden abzustecken.

Die Hesse'sche Normalform einer Geraden ist definiert durch

$$\rho = x \cos(\phi) + y \sin(\phi). \quad (6.10)$$

Für alle Kantenpixel werden nun die jeweiligen Positionen $\begin{pmatrix} x & y \end{pmatrix}^T$ in Gleichung 6.10 eingesetzt. Für alle Werte ϕ und ρ , für welche die Gleichung erfüllt ist, werden die entsprechenden Akkumulatorzellen des Parameterraumes inkrementiert. Dafür wird ϕ in $\Delta\phi$ Schritten diskretisiert und alle zugehörigen ρ berechnet (und ebenfalls diskretisiert). Ein Kantenpixel entspricht damit einem Sinusoid im Parameterraum und ein Punkt im Parameterraum einer Geraden im Bild.

Im nächsten Schritt können Geraden durch eine Suche nach Häufungspunkten im Parameterraum gefunden werden. Abbildung 6.9(c) zeigt eine so durchgeführte Analyse. Durch Kantenrauschen sind die Maxima im Parameterraum verschmiert und werden zur Vermeidung von Mehrfachdetektion gefiltert. Ein enormer Vorteil der gegebenen Geometrie ist die Beschränkung auf wenige zu untersuchende Richtungen, da dadurch Fehldetektionen verringert werden können.

Die gefundenen Geraden unterteilen das Bild in ein nicht equidistantes Raster. Eine Möglichkeit der Klassifikation ist, für jedes abgeteilte Rechteck dessen mittleren Intensitätswert zu bestimmen und auf Grund diesem zu entscheiden, ob es sich um ein Leitungsstück handelt. Voraussetzung dafür ist, dass die Leiterbahnen flächig gefüllt sind. Abbildung 6.9(d) zeigt ein so segmentiertes Bild.

Eine Schwierigkeit dieser modellbasierenden Segmentationsmethode liegt in der Erken-

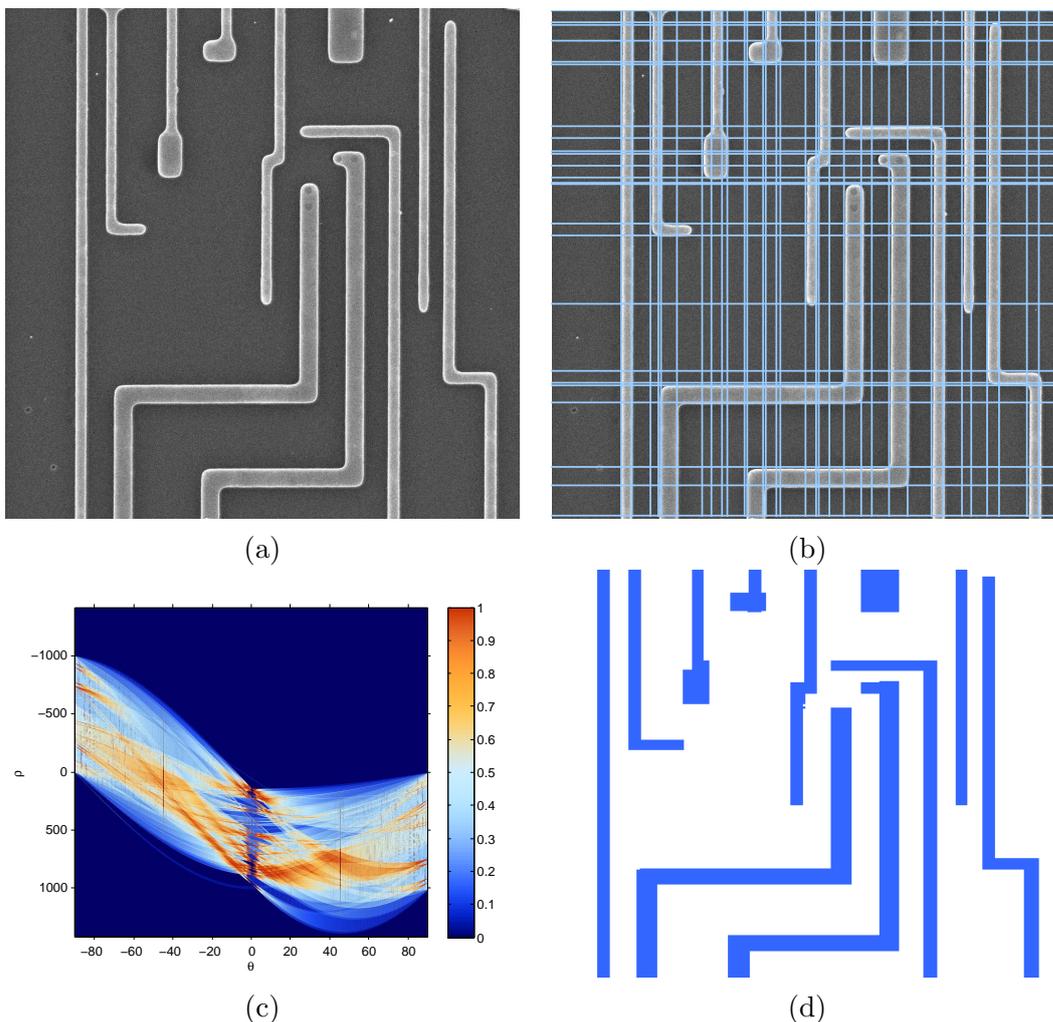


Abbildung 6.9: Leitungsextraktion mit der Hough-Transformation. Vorwissen über die geometrische Form der Leiterbahnen kann dazu genutzt werden, diese aufzufinden. In diesem Beispiel wurde die Suche nach Geraden mit Orientierungen von 0° und 90° mit einer Toleranz von $\pm 1^\circ$ beschränkt. Abbildung (b) zeigt die in Abbildung (a) gefundenen dominanten Geraden durch die Hough-Transformation unter Verwendung der Gleichung 6.10. Abbildung (c) zeigt den dazugehörigen normalisierten Parameterraum und Abbildung (d) die Rekonstruktion der Leiterbahnen.

nung von kurzen Leitungsstücken. Diese können oft nur sehr wenige Pixel ausmachen und werden durch andere zufällige Konstellationen überdeckt. Da diese Methode die Leitungen nur in erster Näherung betrachtet, können Abrundungen an den Leitungsenden nicht berücksichtigt werden und stellen damit Fehlerquellen dar. Eine Ausweitung auf komplexere Geometrie mit evtl. Zuhilfenahme der generalisierten Hough-Transformation wurde in dieser Arbeit nicht durchgeführt, bietet jedoch Potential zur Verbesserung.

Algorithmus 6 Leitungsextraktion durch Hough-Transformation**Eingabe:** Intensitätsbild \mathbf{I} , VIA-Radius R_{VIA} , Schwellwert Th_{metal} **Ausgabe:**

- 1: Berechnung des Kantenbildes \mathbf{B} aus Bild \mathbf{I} durch Algorithmus 3.
- 2: Hough-Transformation unter Verwendung von Gleichung 6.10 zur Erstellung des Parameterraums $\mathbf{A}(\phi, \rho)$.
- 3: Tiefpassfilterung von $\mathbf{A}(\phi, \rho)$.
- 4: Begrenzung auf relevante Richtungen durch partielles Null setzen von $\mathbf{A}(\phi, \rho)$.
- 5: Suche nach lokalen Maxima in $\mathbf{A}(\phi, \rho)$ liefert dominante Geraden in Bild \mathbf{I} .
- 6: Unterteilung des Bildes \mathbf{I} durch die gefundenen Geraden in rechteckige Ausschnitte.
- 7: Klassifikation der Ausschnitte anhand deren mittlerer Intensitätswerte. Die Berechnung des Schwellwertes kann mit der Methode nach Otsu erfolgen [?].

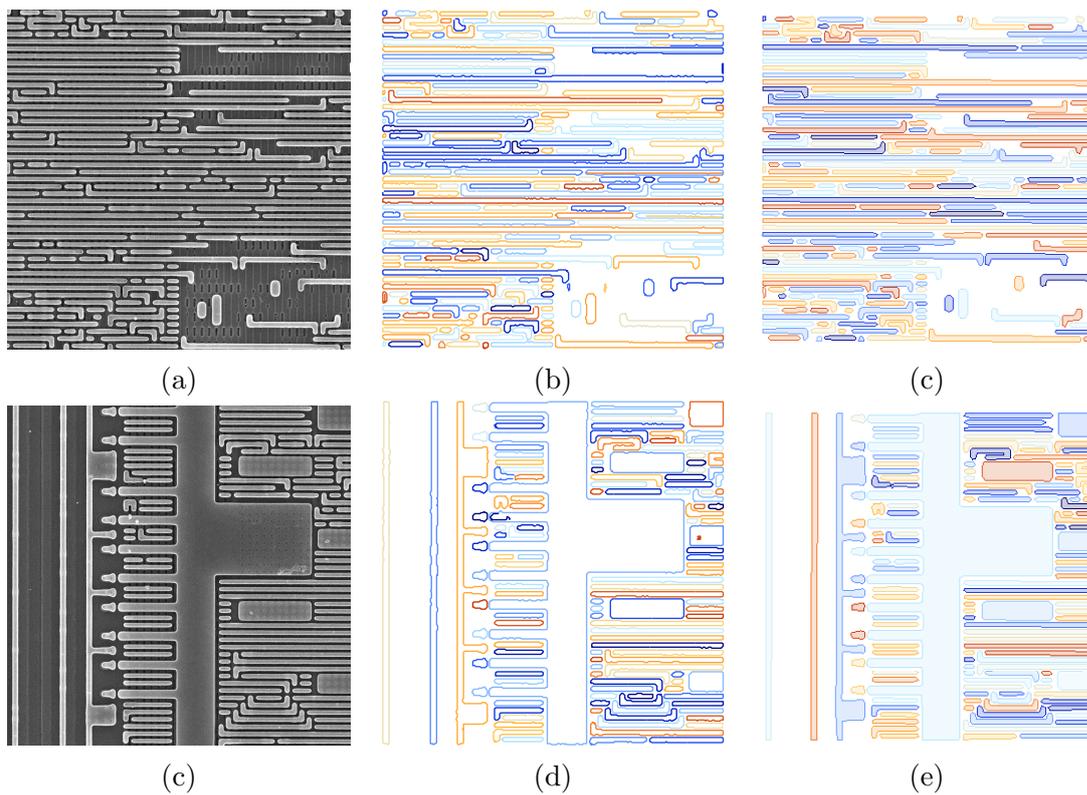


Abbildung 6.10: Vektorisierung der Segmentgrenzen. Die durch Kantendetektion und Kantenverfolgung ermittelten Konturen aus Abbildung (b) und (d) werden durch Algorithmus 7 in eine Vektordarstellung umgewandelt (Abbildung (c) und (e)). Einerseits reduziert sich die zur Beschreibung notwendige Datenmenge und andererseits dient dies der Abstraktion für nachfolgende Algorithmen. Der maximal tolerierbare Abstand zur Kontur wurde auf $d_{max} = 1$ gesetzt.

6.3.3 Vektorisierung

Aus der Segmentierung resultiert neben dem gewonnenen Bildverständnis auch eine enorme Datenreduktion, die wiederum die Laufzeit darauffolgender Algorithmen reduziert. Konturen stellen aber eine für die Anforderung zu genaue Beschreibung der Segmentgrenzen dar und müssen zum Zwecke einer effizienten Weiterverarbeitung abstrahiert werden. Eine für die Problemstellung zweckmäßige Beschreibung durch simple Polygone, erfordert das Einpassen von Liniensegmenten in die bestehende Beschreibung. Es soll weiters möglich sein, durch einen Parameter den Approximationslevel einzustellen, welcher eine Abwägung zwischen Genauigkeit und Anzahl der Stützstellen trifft.

Eine Maßzahl für die Genauigkeit der Approximation kann über eine maximale tolerierbare Distanz d_{tol} der Liniensegmente zu den Konturpixeln definiert werden. Eine maximale Distanz von Null entspricht damit einer exakten Nachbildung der Kontur durch Liniensegmente. Mit steigendem d_{tol} sinken die benötigten Stützstellen, aber auch die Genauigkeit der Nachbildung. Wird d_{tol} zu hoch gewählt, kann dies zu einer Entartung der ursprünglichen Form führen und damit Fehler in der Netzliste verursachen. Algorithmus 7 bildet die Kontur $\mathbf{c}[n], n = 0 \dots N - 1$ von N Pixel durch Vorgabe einer maximalen Abstandstoleranz d_{tol} durch M Liniensegmente $\mathbf{l}[m], m = 0 \dots M - 1$ nach. Dabei gilt $M \leq N$. Beginnend mit der Geraden durch Anfangspunkt $\mathbf{c}[n_{start} = 0]$ und Endpunkt $\mathbf{c}[n_{end} = N - 1]$ wird der maximale Abstand d_{max} der Gerade zur Kontur bestimmt. Ist dieser kleiner als d_{tol} , entspricht die Kontur also in etwa einer Geraden, so wird diese durch ein Liniensegment ersetzt. Ist die Toleranz höher als erlaubt, wird der Endpunkt so lange in Richtung Startpunkt verschoben, bis das Toleranzkriterium erfüllt ist oder der Endpunkt den Startpunkt erreicht. Nach Setzen des neuen Startpunktes auf den alten Endpunkt kann der Algorithmus wiederholt werden bis er mit $n_{start} = N - 1$ terminiert. Als Resultat erhält man eine Liste von Liniensegmenten, welche die Kontur approximiert.

6.4 Kontaktextraktion

Die vertikale Verbindung der Ebenen erfolgt durch Kontakte (VIAs), die für das Auffinden der Knotenpunkte einer Schaltung erforderlich sind. Aufgrund dessen, dass VIAs meist aus Wolfram gefertigt sind, werden sie in einem SEM typischerweise mit höherer Intensität abgebildet und können rein optisch gut unterschieden werden. Auch einer automatisierten Extraktion kommt dies zugute, da die Kontakte anhand einer einfachen Klassifizierung durch Schwellwertbildung des Intensitätsbildes gefunden werden können. Wie stark sich ein Kontakt von einer Leiterbahn hervorhebt, ist sehr vom Ziel und dem Gelingen der Präparation abhängig. Sollen die Leiterbahnen mit hohem Kontrast abgebildet werden, muss das isolierende Siliciumdioxid bis in die Nähe der Leiterbahn angetragen werden. Durch anisotropen Materialabtrag können Kontakte freistehen und sich durch physikalische Einwirkungen von der Metallbahn lösen. Diese dann teilweise über den Chip verstreuten Kontakte erfordern zur korrekten

Algorithmus 7 Vektorisierung**Eingabe:** Abstandstoleranz d_{tol} , Konturpixelpositionen $\mathbf{c}[n]$ mit $n = 0 \dots N - 1$ **Ausgabe:** Liste \mathbf{l} der Liniensegmenten

```

1:  $N \leftarrow \text{length}(c)$ 
2:  $n_{start} \leftarrow 0$ 
3:  $n_{end} \leftarrow N - 1$ 
4: repeat
5:    $d_{max} \leftarrow \text{maxLineDev}(\mathbf{c}, n_{start}, n_{end})$ 
6:   if  $d_{max} \leq d_{tol}$  or  $n_{start} = n_{end}$  then
7:      $\text{addLine}(\mathbf{l}, n_{start}, n_{end})$ 
8:      $n_{start} \leftarrow n_{end}$ 
9:      $n_{end} \leftarrow N - 1$ 
10:  else
11:     $n_{end} \leftarrow n_{end} - 1$ 
12:  end if
13: until  $n_{start} = N - 1$ 

```

Klassifizierung komplexere Methoden, die jedoch nicht näher betrachtet werden. Ansätze aus der Objekterkennung erscheinen hier sinnvoll, sprengen jedoch den Rahmen dieser Arbeit.

Nach Filterung des Bildes wie in Abschnitt 6.3.1 beschrieben und einer Kontrastnormalisierung (Kontrastspreizung) ist es möglich, durch ein einfaches pixelbasierendes Segmentationsverfahren, wie der globalen Schwellwertbildung, die zu den Kontakten zugehörigen Regionen zu finden (Abbildung 6.11). Ein nachfolgendes *labeling* teilt den einzelnen Regionen eine eindeutige Kennung zu. Der Mittelpunkt eines Kontaktes kann durch den Schwerpunkt der Region definiert werden.

Ist ein geringer ohm'scher Widerstand zwischen zwei Leiterbahnen unterschiedlicher Lagen gefragt, wie es unter anderem für die Stromversorgung gefordert ist, werden üblicherweise viele Kontakte blockweise angeordnet. Bei der Abbildung dieser Bereiche in einem REM kann es vorkommen, dass einzelne VIAs verbunden erscheinen und durch eine simple globale Schwellwertbildung nicht getrennt werden können. Eine Möglichkeit, diese Regionen zu trennen, ist die Anwendung der euklidischen Distanztransformation (*distance transform*). Dabei wird jedem Pixel $\mathbf{B}(x, y) = 1$ (allen VIA-Pixel) ein Wert zugeordnet, der der kleinsten Distanz zum nächstliegenden Pixel $\mathbf{B}(u, v) = 0$ (dem Hintergrund) entspricht. Die restlichen Hintergrundpixel besitzen die Distanz 0 zu sich selbst. Anhand der Distanzmatrix \mathbf{D} erhalten Pixel, die sich in der Mitte einer geschlossenen Region befinden, höhere Werte als am Rand liegende. Eine grafische Darstellung der Distanzwerte als Funktion der Bildkoordinaten enthält Abbildung 6.12(c). Nach Normalisierung der Distanzmatrix und setzen eines Schwellwertes können

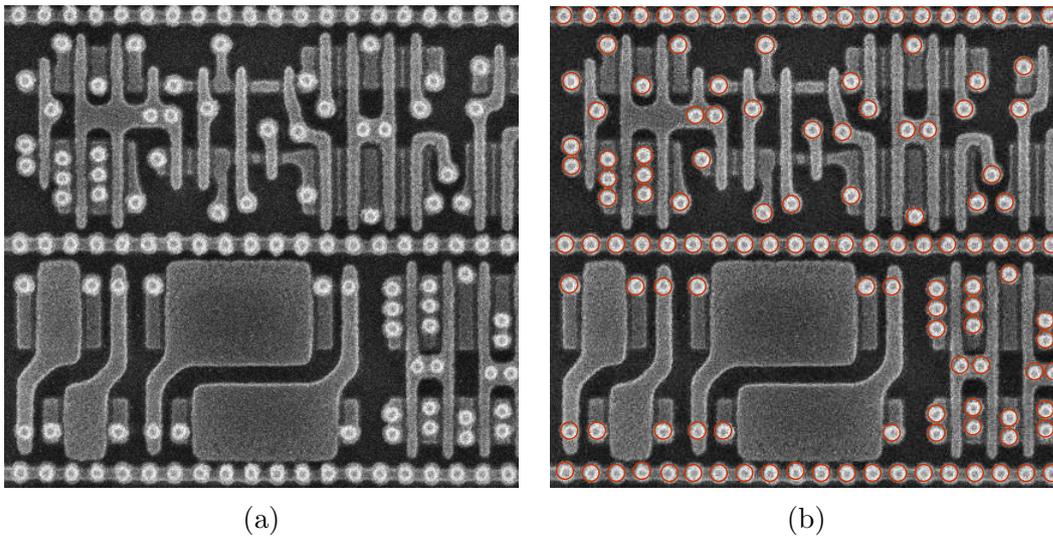


Abbildung 6.11: Ist genügend Kontrast zwischen Kontakt und Leiterbahn gegeben, können die Kontakte durch eine einfache Schwellwertbildung gefunden werden. Mittelpunkt des Kontaktes stellt dabei der Schwerpunkt der gefundenen Region dar. Um robuste Ergebnisse über die Bilderreihe zu erhalten, muss eine Normalisierung der Intensitätswerte erfolgen. Unter der Voraussetzung, dass VIAs die höchste Helligkeit aufweisen, kann der Schwellwert meist auf einen fixen Wert zwischen $[0.9,1]$ manuell festgelegt werden.

nun zusammenhängende Regionen getrennt werden. Obwohl für die topologische Analyse der Schaltungsknoten die blockweise Zusammenschaltung mehrerer Kontakte rein redundanter Natur wäre, ist es dennoch vorstellbar, dass sich bei entsprechender Platzierung der Kontakte ein Schwerpunkt außerhalb des zusammenhängenden Blockes ergibt und damit das Einfügen eines nicht existierenden Knotens in die Schaltung erfolgt. Es existieren effiziente Algorithmen zur Berechnung der Distanztransformation, die eine lineare Laufzeitkomplexität $O(n)$ besitzen [?, ?, ?].

In der Regel werden VIAs in den Layoutdaten quadratisch abgebildet. Jedoch durch Beugungseffekte, Fertigungstoleranzen sowie präparationsbedingte Einwirkungen erscheinen Ecken von Leiterbahnen und VIAs immer abgerundet. Diese geometrische Eigenschaft von Kontakten kann dazu genutzt werden Randbedingungen festzulegen. Ein mögliches Maß für die *Rundheit* einer Region kann durch das Verhältnis zwischen seiner Fläche zur Referenzfläche $r^2\pi$ definiert werden. Um jedoch mehr Spielraum zu eckigerem Aussehen zu haben, wurde als Maß das Verhältnis zwischen der Länge der Hauptachse und der Länge der Nebenachse gewählt. Die durch die Haupt- und Nebenachse aufgepannte Ellipse besitzt dabei die selben zweiten Momente wie die Region. Ein Verhältnis von 1 kann damit einer runden oder quadratischen Gestalt entsprechen.

Durch die kreisartige Form der VIAs, welche auch in Kantenbildern ersichtlich ist, wäre weiters der Einsatz einer Hough-Transformation für Kreise denkbar. Ein entscheidender Vorteil dieser Methode sind relaxierte Anforderungen an den Kontrast zwischen Leiterbahn

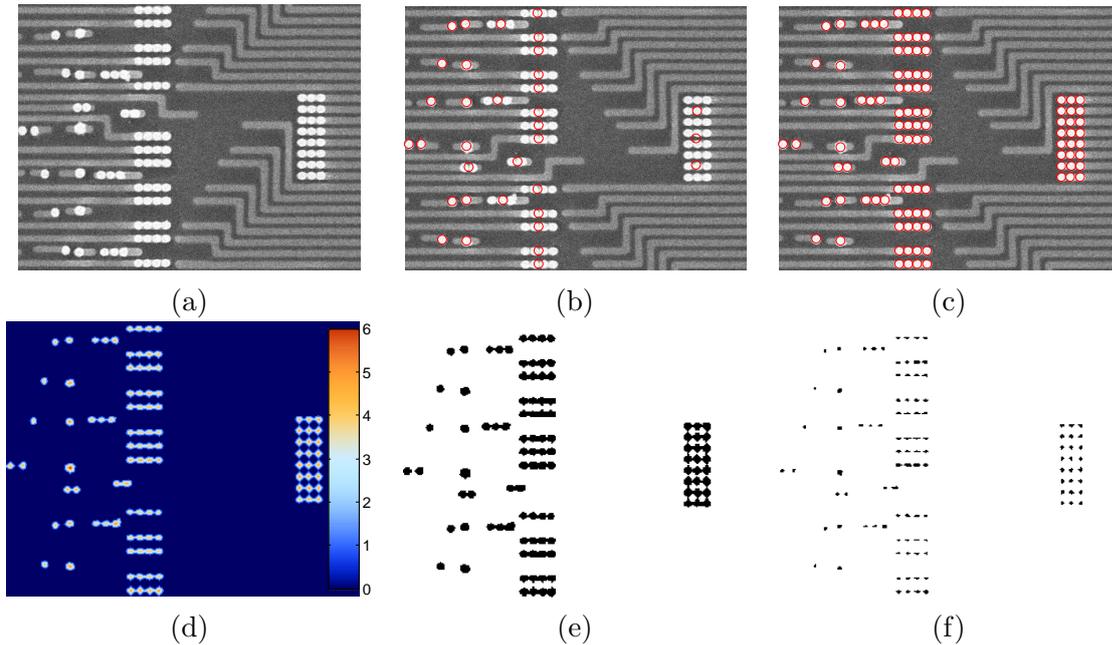


Abbildung 6.12: Kontaktextraktion mit Hilfe der Distanztransformation. Um den Kanteneffekt eines REM zu mindern und damit die Leiterbahnen flächiger erscheinen zu lassen, muss über der Metallleitung eine Oxidschicht verbleiben und für ausreichenden Kontrast mit höheren Beschleunigungsspannungen gearbeitet werden. Die Kontakte aus Wolfram werden dadurch mit hoher Intensität abgebildet und können einander überstrahlen. Die dadurch auftretenden Kurzschlüsse zwischen den Kontakten können zu Fehlern in der Netzliste führen. Ausgangsbild (a) wurde durch einfache Schwellwertbildung segmentiert ((b) und (e)). Dabei wurden Segmente fälschlicherweise verbunden. Mit Hilfe der Distanztransformation ist es möglich, diese wieder zu trennen ((c) und (f)). Bild (d) zeigt die Distanzmatrix \mathbf{D} , anhand welcher eine Auftrennung der Segmente ermöglicht wird.

Algorithmus 8 VIA Extraktion durch globale Schwellwertbildung

Eingabe: Intensitätsbild \mathbf{I} , Schwellwerte Th_{VIA} , Th_{dist} , Th_r

Ausgabe:

- 1: Normalisierung der Intensitäten des Bildes \mathbf{I} auf $[0, 1]$.
 - 2: Globale Schwellwertbildung durch Th_{VIA} liefert das Binärbild \mathbf{B} .
 - 3: Distanztransformation auf \mathbf{B} liefert die Distanzmatrix \mathbf{D} .
 - 4: Normalisierung der Distanzmatrix \mathbf{D} auf $[0, 1]$.
 - 5: Erneute Schwellwertbildung von \mathbf{D} mittels Th_{dist} liefert das Binärbild \mathbf{B}_D .
 - 6: Auffinden aller Regionen in \mathbf{B}_D und Berechnung derer Hauptkomponenten durch die Hauptkomponentenanalyse. Das Verhältnis der Eigenwerte $r = \lambda_1/\lambda_2$ mit $\lambda_1 \geq \lambda_2$ und damit $r \geq 1$ liefert den Kennwert für die Rundheit.
 - 7: Für alle Regionen, für welche die Rundheitsbedingung erfüllt ist, d.h. $r < Th_r$ gilt, berechnen sich die Positionen der VIAs durch Finden der Schwerpunkte der Regionen.
-

und Kontakt, was mehr Spielraum bei der Präparation schafft. Die klassische Hough-Transformation wurde bereits in Abschnitt 6.3.2.2 vorgestellt. Verwendet man nun statt Gleichung 6.10, die Gleichung

$$(x - m_x)^2 + (y - m_y)^2 = r^2 \quad (6.11)$$

eines Kreises, mit dem Mittelpunkt $(m_x \ m_y)$ und dem Radius r , so kann ein dreidimensionaler Parameterraum $\mathbf{A}(m_x, m_y, r)$ berechnet werden. Legt man einen fixen Radius $r = R_{VIA}$ fest, so reduziert sich der Raum auf zwei Dimensionen. Da der Radius eines VIA innerhalb einer Aufnahmeserie in etwa konstant ist, kann dies ohne Einschränkung erfolgen. Abbildung 6.13 zeigt eine so durchgeführte VIA Extraktion. Als problematisch erweist sich, dass der Radius der Kontakte meist nur wenige Pixel ausmacht und dadurch die lokalen Maxima des Parameterraums verrauscht sind. Weiters ist bei der Suche nach kleinen Kreisen die Wahrscheinlichkeit höher, dass zufällige Pixel-Konstellationen kreisartige Strukturen bilden.

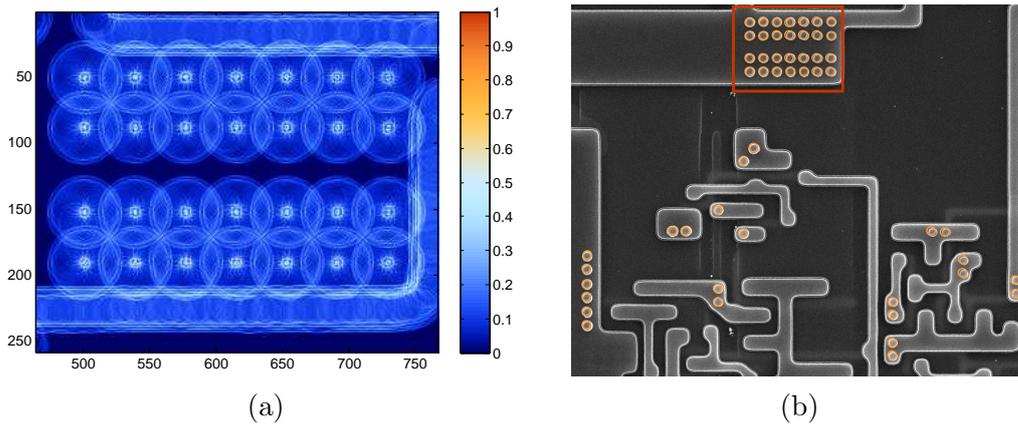


Abbildung 6.13: Kontaktextraktion durch die Hough-Transformation. Ist nicht genügend Kontrast zwischen Leiterbahn und VIA vorhanden, scheidet der Einsatz eines globalen Schwellwertverfahrens aus. Die Hough-Transformation kann bei ausreichender Größe der VIAs genutzt werden, um diese aufzufinden. In Abbildung (a) ist der Parameterraum $\mathbf{A}(m_x, m_y)$ für ein Teilbild von Abbildung (b) zu sehen. R_{VIA} beträgt 13 Pixel.

Wie schon bei der Segmentierung von Leiterbahnen ist auch hier die Wahl des Algorithmus sehr vom Datenmaterial abhängig. Kann ausreichender Kontrast zwischen Leiterbahn und VIA gewährleistet werden, so bietet sich das einfache und schnelle globale Schwellwertverfahren aus Algorithmus 8 an. Alternativ kann bei genügend großer geometrischer Auflösung ($R_{VIA} > 10$) die Hough-Transformation verwendet werden.

Algorithmus 9 VIA Extraktion durch Hough-Transformation

Eingabe: Intensitätsbild \mathbf{I} , VIA-Radius R_{VIA} **Ausgabe:**

- 1: Berechnung des Kantenbildes \mathbf{G} aus Bild \mathbf{I} durch Algorithmus 2.
 - 2: Hough-Transformation unter Verwendung von Gleichung 6.11 (mit $r = R_{VIA}$) zur Erstellung des Parameterraums $\mathbf{A}(m_x, m_y)$.
 - 3: Tiefpassfilterung von $\mathbf{A}(m_x, m_y)$.
 - 4: Suche nach lokalen Maxima in $\mathbf{A}(m_x, m_y)$.
-

6.5 Extraktion von Bauteilen

Eine große Herausforderung stellt das Auffinden und die Klassifikation von Bauteilen dar. Die primären Bauteile - CMOS Transistoren - bestehen, sehr vereinfacht beschrieben, aus n und p dotierten Gebieten und einer sich darüber befindlichen, isolierten Schicht aus Polysilicium. Die Unterscheidung zwischen n und p Gebieten ist nur durch aufwendige (und kostenintensive) Verfahren möglich (vgl. SIMS, AFM) und beschränkt sich damit auf kleinere ausgewählte Gebiete. Eine Differenzierung zwischen NMOS und PMOS Transistoren kann deshalb nur eingeschränkt, aufgrund der Verschaltung getroffen werden. Abgesehen von dieser Tatsache zeigen Transistoren je nach verwendeter Prozesstechnologie und Präparation sehr unterschiedliches Aussehen. Dies erschwert den Entwurf einer allgemeinen Extraktionsmethode enorm.

Prinzipiell sind Transistoren in den untersten Lagen des Mikrochips angesiedelt. Die dotierten *source/drain* Gebiete sind auf der Ebene *active area* (AA) ersichtlich. Die *gate* Elektrode und die Kontakte zur ersten Metallisierungsebene (M1) sind auf *gate contact* (GC). Abbildung 6.14 zeigt eine Gegenüberstellung der Ebenen AA und GC desselben Chipausschnittes.

Leider konnte auf diesen interessanten Abschnitt nicht eingegangen werden, da es über den Rahmen dieser Arbeit hinausgeht. Die Problemstellung wird somit weiterführenden Arbeiten überlassen.

6.6 Routing

Die in polygonaler Form vorliegenden Layoutdaten wurden für jedes Teilbild einzeln ermittelt und erfordern daher einer Zusammenführung. Da sich benachbarte Bilder überlappen, werden an den Bildrand grenzende Polygone auch in benachbarten Bildern gefunden und können somit durch einen Überlappungstest in Verbindung gebracht werden. Schneiden sich Polygone, so bekommen sie die selbe Spannungsknotennummer zugewiesen und besitzen somit das selbe elektrische Potential. Der Test auf Überlappung

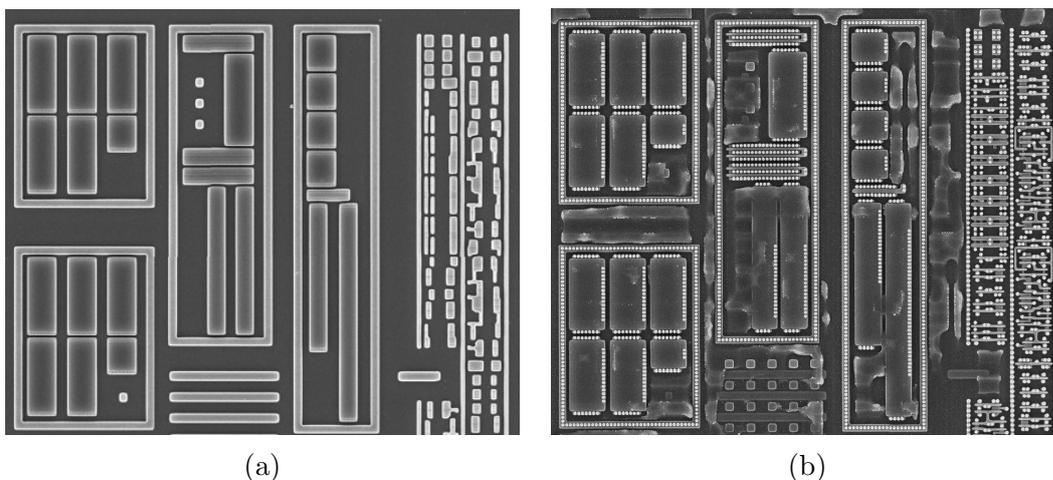


Abbildung 6.14: Gegenüberstellung der Ebenen AA und GC. In Abbildung (a) ist ein Ausschnitt aus der Ebene *active area* eines analogen Bereiches zu sehen, in Abbildung (b) der selbe Ausschnitt auf der Ebene *gate contact*. Im linken Teil der Bilder befinden sich durch zwei *guard rings* eingerahmte *source-gate* Kapazitäten, mittig ebenfalls *source-gate* Kapazitäten sowie vier Transistoren und im rechten Teil zwei Spalten Standardzellen. Abbildung (b) weist erheblich, durch den Präparationsprozess bedingte, Verunreinigungen auf, die eine automatische Extraktion erschweren.

von Polygonen kann auf eine Schnittbestimmung von Liniensegmenten rückgeführt werden. Wenn n die Anzahl der Liniensegmente und k die Anzahl der Schnittpaare darstellt, bedeutet das für den naiven Ansatz, der alle Kanten eines Polygons mit den Kanten der umliegenden Polygone auf Überschneidung testet, eine Laufzeitkomplexität von $O(n^2)$. Eine Klasse effizienterer Ansätze nutzt das *sweep line* Verfahren und kommt damit der theoretischen unteren Schranke sehr nahe. In [?] und [?] werden Ansätze vorgestellt, die eine Laufzeitkomplexität von $O((n + k) \log n)$ besitzen und damit eine wesentliche Verbesserung darstellen. Neuere Arbeiten wie [?] kommen auf $O(n \log(n + k))$, sind jedoch in der Umsetzung aufwendiger. Die Methode nach [?] findet aufgrund seiner praktischen Vorteile auch heute noch breite Verwendung und gilt auch in dieser Arbeit als Mittel der Wahl.

Verbindungen zwischen den Ebenen sind durch Kontakte realisiert. Die Kontaktextraktion aus Abschnitt 6.4 liefert Positionen in Form von zweidimensionalen Punkten. Mit dem Wissen der minimalen Strukturgröße einer Ebene kann die Ausdehnung der Kontakte durch Umrahmung beschrieben werden. Um die oben beschriebene Methodik beizubehalten und damit Konformität zu gewährleisten, bietet sich an, ein Quadrat mit der minimalen Strukturgröße als Seitenlänge und die Kontaktposition als Mittelpunkt zu verwenden.

6.7 Netzliste

Die Beschreibung eines elektrischen Designs erfolgt in der Regel in Form einer Netzliste welche die verwendeten Komponenten und deren Verbindungen untereinander definiert. Komponenten können einfache Bauteile wie Transistoren oder Widerstände sein, aber auch komplexere Module (*black boxes*) wie digitale Zellen. Als Standard hat sich die SPICE (*Simulation Program with Integrated Circuit Emphasis*) Netzliste etabliert, für welche viele EDA (*Electronic Design Automation*) Softwarepakete Schnittstellen bieten und es sich damit als ein attraktives Format für den Schaltungsexport anbietet. Ein weiterer Vorteil des Dateiformates, welches die Handhabung wesentlich erleichtert, ist die Kodierung im ASCII Format. Die für den Export der schaltungsrelevanten Daten notwendige Syntax gestaltet sich dabei sehr einfach. Im Prinzip besteht die Netzliste aus einer Auflistung von Komponenten, derer Anschlüsse und Parameter. Die Abbildung der Verbindungen erfolgt durch die Benennung der Anschlüsse mit einer eindeutigen Knotennummer.

Eine flache Netzliste beinhaltet keinerlei Information über die Hierarchien der Schaltung und ist damit für das Verständnis größerer Schaltungen ungeeignet. Durch eine Partitionierung anhand des Layouts kann eine rudimentäre Zwei-Ebenenhierarchie geschaffen werden. Bauteile, die sich lokal nahe befinden, sind auch meist in der Schaltung nur wenige Knoten entfernt. Zudem sind Subschaltungen meist durch *gard rings* getrennt und können somit leicht gefunden werden. Dies trifft zumindest für analoge Schaltungsbereiche zu. Bei digitalen Schaltungen, die durch Standardzellen realisiert werden, gestaltet sich eine Partitionierung jedoch schwieriger und ein Auffinden von Subschaltungen muss hier auf einem höheren Abstraktionslevel erfolgen. Durch semantische Interpretation anhand von Grammatiken sowie durch den Vergleich mit bekannten Grundschaltungen können Subschaltungen gefunden werden. Ein semantischer Vergleich von Subschaltungen mit *high-level* Komponenten wird in [?, ?] vorgestellt. Dabei wurden aus unbekanntem digitalen Schaltungen, Komponenten wie ALUs, Addierer, Multiplexer, Dekoder und Enkoder extrahiert und damit eine funktionale Beschreibung der digitalen Schaltungen erstellt. In [?, ?, ?] werden Ansätze vorgeschlagen, die auf heuristischer Optimierung einer graphentheoretischen Darstellung beruhen. Da jedoch eine weitere Ausführung dieser Thematik den Rahmen dieser Arbeit bei weitem sprengen würde, wird an dieser Stelle auf die Literatur verwiesen.

Das Importieren von Netzlisten in Standardsoftware für Schaltungssimulation oder Schaltungssynthese eröffnet vielseitige Möglichkeiten. Durch Verifikationsalgorithmen kann die Plausibilität der Zusammenschaltung analysiert werden und Fehler können durch Kurzschlüsse oder fehlende Verbindungen gefunden werden. Die graphische Darstellung der Schaltung ermöglicht ein Platzieren der Bauteile und deren Verbindungen, um ein Verständnis über die Funktionalität zu erlangen. Bei Voraussetzung einer korrekten Netzliste kann durch digitale und analoge Simulation ein tieferes, quantitatives Verständnis über die Schaltung gewonnen werden. Ohne ein hierarchisches Modell sind

die Simulationsergebnisse jedoch von geringem Wert.

6.8 Zusammenfassung and weiterführende Fragestellungen

In diesem Kapitel wurde ein durch Bildanalyse gestütztes Verfahren zur Segmentierung und Vektorisierung von Leiterbahnen und zur Extraktion von Kontakten aus REM Aufnahmen integrierter Schaltkreise vorgestellt. Unter der Voraussetzung einer sauberen Präparation lieferten die Methoden robuste Ergebnisse.

Durch Präparationsartefakte verursachte Fehlerkennungen bleiben jedoch nicht aus und müssen in einer *review* Phase korrigiert werden. Durch das Festlegen von Randbedingungen wie *layout design rules* könnten Verletzungen automatisch gefunden und damit mögliche Fehler detektiert werden. Der *design rule check* (DRC) ist ein Standardfeature jeder Verifikationssoftware für Layouts und findet im Design Anwendung. Mit dem Export in das GDSII oder DEF Dateiformat können die extrahierten Daten in DRC fähige Software eingelesen und analysiert werden. Das dazu benötigte Regelwerk muss für die jeweilige Prozesstechnologie angepasst werden.

Der vorgestellte regionenbasierende Ansatz zur Kontaktextraktion aus Abschnitt 6.4 erfordert ausreichend Kontrast zwischen Leiterbahnen und Kontakten. Dies muss bei der Präparation berücksichtigt werden, um korrekte Ergebnisse zu erhalten. Der Ansatz über die Hough-Transformation setzt hohe geometrische Auflösung und rundes Aussehen der Kontakte voraus. Im Mittel wurden durch den regionenbasierenden Ansatz robustere Ergebnisse erzielt.

Im Rahmen dieser Arbeit wurde das Layout mehrerer Mikrochips rekonstruiert. Es zeigte sich theoretisch wie praktisch, dass die Laufzeitkomplexität der vorgestellten Verfahren im Mittel linear von der Anzahl der Bilder und dessen Auflösung und damit linear von der Fläche des Mikrochips abhängt. Die Extraktion von Layout und Kontakten benötigt pro Metallebene durchschnittlich 2.8 Stunden/mm^2 auf einem *Intel Core i5-750* Prozessor. Damit benötigt die Rekonstruktion der Metalllagen eines fünfflagigen Mikrochips mit einer Fläche von 2 mm^2 in etwa 28 Stunden und bleibt damit wirtschaftlich und praktisch realisierbar.

The digital revolution is far more significant than the invention of writing or even of printing.

Douglas Engelbart

Kapitel 7

Zellenextraktion

Zusammenfassung — In diesem Kapitel werden Methoden vorgestellt, welche die Analyse von digitalen Layouts ermöglicht. Dabei werden Standardzellen aus dem Layout extrahiert und Kennwerte sowie Statistiken abgeleitet. Mit der Kombination der in Kapitel 6 entwickelten Algorithmen zur Leiterbahnextraktion ermöglicht dies die Rekonstruktion von digitalen Netzlisten.

Inhaltsverzeichnis

7.1	Einleitung	106
7.2	Extraktion digitaler Standardzellen	107
7.3	Extraktion von Speicherzellen	115
7.4	Verteiltes Rechnen	116

7.1 Einleitung

Digitaldesign ist die vorherrschende Entwurfstechnik in der Mikroelektronik. Durch die fortschreitende Minimalisierung der Prozesstechnologien wurde es möglich, komplexe Abläufe mit Hilfe von digitalen Entwurfsverfahren auf engstem Raum physikalisch abzubilden. Diese neuen Möglichkeiten verdrängen die analoge Schaltungstechnik, die bisher für viele Anwendungsbereiche als Mittel der Wahl galt. Dennoch ist die analoge Schaltungstechnik nicht wegzudenken, da die Schnittstelle zur Umwelt meist analoger Natur ist. Ein großer Anteil eines modernen *mixed-signal* IC besteht aus digitalen Zellenbereichen, die logische Funktionen realisieren oder zur Speicherung von Daten dienen. Die logischen Funktionen sind dabei in Zellen (Standardzellen) gekapselt, um den Entwurfsprozess zu abstrahieren. Eine regelmäßige Anordnung dieser Zellen in Spalten oder Reihen vereinfacht eine automatische Verbindung (*routing*) der Zellen zu funktionalen Einheiten (Module). Zwischen den Spalten bzw. Reihen verlaufen Leitungen für die Versorgungsspannung, die die Versorgung der Zellen ohne zusätzlichen Verdrahtungsaufwand gewährleisten. Die Funktionalität der Module wird durch eine Beschreibungssprache

wie VHDL oder Verilog festgelegt. Diese Abstraktion ermöglicht einerseits den Entwurf und die Simulation komplexer Funktionen und andererseits eine Trennung von Technologie und Design und erleichtert damit die Wiederverwendbarkeit von bereits bestehenden Entwürfen.

Die Schnittstelle der analogen zur digitalen Welt bilden Analog zu Digital Umsetzer. Der Trend, immer früher in die digitale Welt zu wechseln, setzt sich immer weiter fort, sodass die digitale Schaltungstechnik immer mehr an Bedeutung gewinnt.

Aus einer digitalen Netzliste kann eine Reihe von Rückschlüssen auf das Design oder den Syntheseprozess abgeleitet werden. Digitale Standardzellenbibliotheken beinhalten viele tausende verschiedene Zellen, die sich aus einer Menge von Grundzellen ableiten. Dabei variiert die Stärke der Ausgangstreiber, die Anzahl an Eingängen, der Strombedarf und eine Reihe anderer Parameter. Bei der Synthese werden Parameter teilweise festgesetzt oder richten sich nach Anforderungen und Randbedingungen. Dadurch findet nur ein kleiner Teil des zur Verfügung stehenden Zellenrepertoires in die Realisierung digitaler Logik Eingang. In der Regel besteht ein digitaler Chipbereich aus ein paar hundert verschiedenen Zellen. Dazu zählen einfache logische Gatter, komplexere Funktionen wie Multiplexer sowie Speicherzellen, wie etwa Latches und diverse Arten von *flip-flops*. Durch Klassifikation der Zellen eines digitalen Bereiches können Kennwerte wie *utilization* (Verhältnis von der durch Zellen belegten Fläche zur Gesamtfläche), *Gatteräquivalenz* (Fläche einer NAND oder NOR Zelle, zu der durch Zellen belegten Fläche), Anzahl der Speicherzellen oder auch allgemein Statistiken über Zellenhäufigkeiten abgeleitet werden. Mit Hilfe dieser Kennwerte können Rückschlüsse auf die Performance der Synthesemethoden gezogen und die Komplexität der digitalen Schaltung abgeschätzt werden.

7.2 Extraktion digitaler Standardzellen

Üblicherweise befinden sich die Verdrahtungen innerhalb einer Zelle auf Metallisierungsebene 1 (M1). Ab der Metallisierungsebene 2 (M2) werden die einzelnen Zellen miteinander verbunden. Dabei erfolgt das *routing* je Ebene meist nur in eine Richtung um Freiheitsgrade einzuschränken und um die Konvergenz der Algorithmen zu garantieren. Je nach Synthesetool ist es jedoch auch möglich, kurze Verbindungen zwischen benachbarten Zellen auf M1 zu ziehen. Dabei entfällt die Kontaktierung auf eine höhere Metallisierungsebene.

Um nun eine Klassifikation der sich in einem Digitalbereich befindlichen Standardzellen treffen zu können, müssen Lagen betrachtet werden, die für die Zellen typische Muster beinhalten. Da sich auf den oberen Metallisierungsebenen die Verbindungen zwischen den Zellen befinden, scheiden diese aus. Gut geeignet erscheinen dagegen die unteren aktiven Ebenen, in welchen sich die Transistoren befinden. Da auch Kontakte auf die erste Metallisierungsebene sichtbar sein sollen, bietet sich hierfür die Ebene *gate contact* (GC) an. Beispiele hierfür findet man in Abbildung 7.1.

Eine alternative Möglichkeit zur Erkennung bieten die Strukturen auf M1. Da diese jedoch wie oben erwähnt auch Verbindungen zu Nachbarzellen beinhalten können, unterscheiden sich gleiche Zellen geringfügig und verfälschen dadurch die Ergebnisse. Diese Feststellung impliziert eine Schwierigkeit. Werden Zellen nicht auf Transistorebene analysiert sondern als *black boxes* betrachtet, können variierende Verbindungen innerhalb der Zelle nicht berücksichtigt werden.

Eine Möglichkeit diese Schwierigkeit zu umgehen ist, die Standardzelle nur bis zur Ebene GC als Einheit zu sehen und alle Metallisierungsebenen durch Leitungsextraktion zu rekonstruieren. Da dies zusätzliche Verarbeitungsschritte und damit Fehlerquellen mit sich bringt, sollten die Bestrebungen dahin gehen, etwaige Verbindungen zwischen benachbarten Zellen auf M1 aufzufinden. Da aufgrund von *design rules*, abhängig von der jeweiligen Ebene, Mindestabstände einzuhalten sind, können Standardzellen nicht auf Stoß plaziert werden. Befindet sich im Raum zwischen zwei benachbarten Zellen eine Leitung auf M1, so kann daraus geschlossen werden, dass dies einer Verbindung zweier Zellen entspricht. In Frage kommen dabei Verbindungen zwischen dem Ausgang einer Zelle und dem Eingang einer Nachbarzelle (Serienschaltung) oder eine Verbindung zweier Eingänge benachbarter Zellen (Parallelschaltung).

7.2.1 Mustererkennung

Eine der zentralen Problemstellungen dieses Kapitels ist das Auffinden der Standardzellen in einem ausgewählten Chipbereich. Dabei sind verschiedenste Ansätze aus der Bildanalyse denkbar, wie das Anwenden von Methoden der Mustererkennung, Segmentierung, [TODO]. Grundlegend muss entschieden werden, ob die Klassifikation aufgrund von Grammatiken erfolgt oder aufgrund von Ähnlichkeitsmaßen. Das Anwenden von Grammatiken erfordert eine Überführung in eine beschreibende Darstellung, wie es in etwa bei der Leitungsextraktion in Kapitel 6.3 erfolgt ist. Aufgrund dieser Beschreibung kann durch Regeln oder dem Vergleich mit bestehenden Beschreibungen aus einer Datenbank eine Klassifikation durchgeführt werden. Jedoch gestaltet sich die Segmentierung der unteren, strukturreichen Ebenen erheblich schwerer als die der Metallisierungsebenen. Auf den Abbildungen der Ebene GC sind Polysilizium (*gates*), die Dotierungsgebiete (*source/drain*) sowie die Kontakte nach M1 mit jeweils verschiedenem Kontrast zu sehen. Aufgrund dieser Schwierigkeiten fiel die Wahl auf eine direkte Methode. Die bereits in Kapitel 5.3 erarbeiteten Methoden zur Bildregistrierung können auch auf diese Problemstellung angewandt werden. Dabei wurde zwischen merkmalsbasierenden und direkten Methoden unterschieden. Zum Unterschied zur Bildregistrierung bestehen jedoch relaxierte Anforderungen an die Genauigkeit, da eine pixelgenaue Lokalisierung ausreichend ist. Korrelationsmethoden wie in Abschnitt 5.3.2 beschrieben, haben sich für rein translatorische Problemstellungen als robust erwiesen und sollen auch hier Anwendung finden.

Korrelationsmethoden vergleichen vorgegebene Suchbilder (*templates*) mit einem Bild (*search matrix*) und errechnen über ein Ähnlichkeitsmaß Korrelationskoeffizienten,

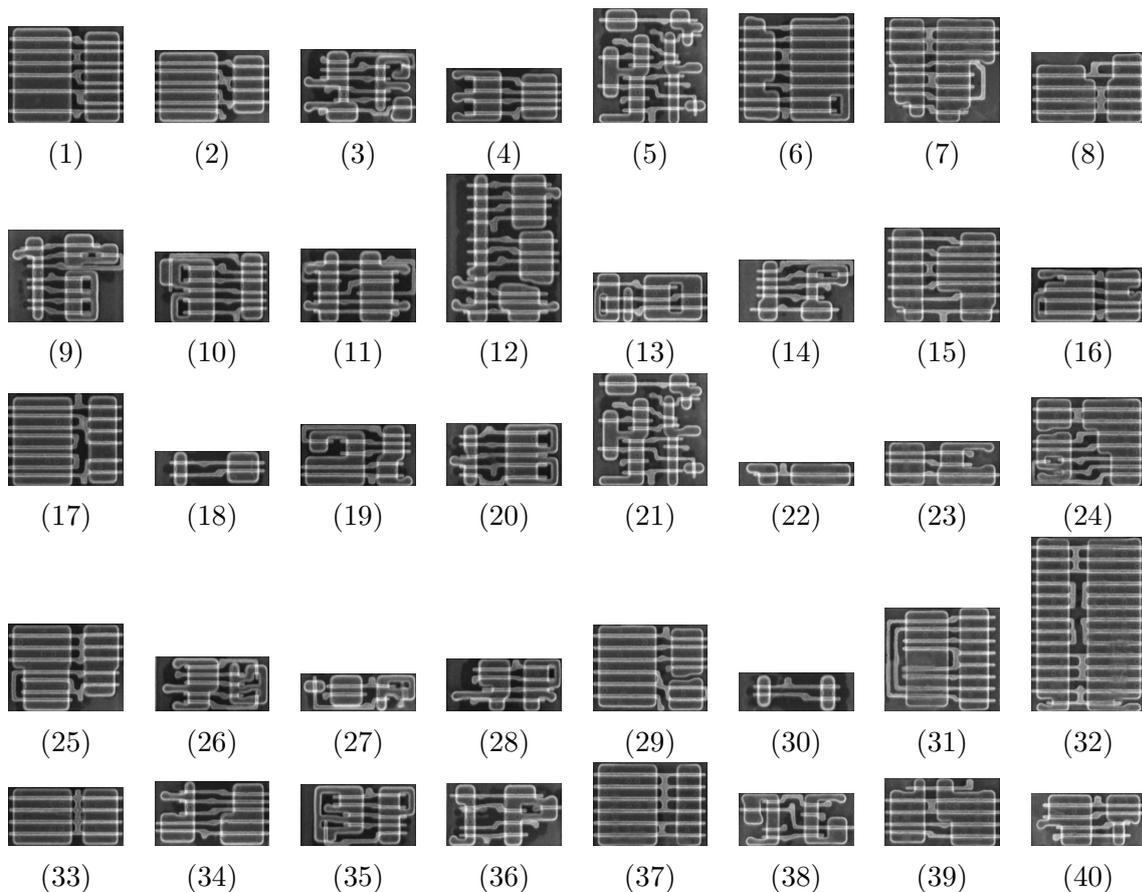


Abbildung 7.1: Standardzellen auf Polysilicium. Für die Suche nach Standardzellen eignen sich Ebenen, die für Zellen charakteristische Strukturen beinhalten. Dafür bietet sich die Ebene Polysilicium an, da hier Transistoren direkt sichtbar sind. Die hier gezeigten Standardzellen wurden per Hand ausgeschnitten und in eine Datenbank eingefügt. Der Typ der Zelle muss durch Schaltungs-*reverse-engineering* gefunden werden. Dabei besitzen PMOS Transistoren aufgrund ihrer geringeren Stromergiebigkeit größere Weiten als NMOS Transistoren und können dadurch in vielen Fällen unterschieden werden.

anhand derer entschieden werden kann, wie gut das Suchbild mit den jeweiligen Bildpositionen des Bildausschnittes übereinstimmt. Im Anwendungsfall der Standardzellen-Lokalisierung erfordert dies eine Bibliothek mit allen vorkommenden Standardzellen als Pixelintensitätsbilder. Der Aufbau der Standardzellenbibliothek kann dabei automatisiert oder manuell per Hand erfolgen. In dieser Arbeit wurde der manuelle Ansatz gewählt, um eine definierte Datenbasis sicherzustellen. Mit geschultem Auge sind die einzelnen Zellen unter Zuhilfenahme der Ebenen AA, GC und M1 abgrenzbar und können durch ein Bildbearbeitungsprogramm ausgeschnitten und in die Standardzellenbibliothek aufgenommen werden. Durch *reverse engineering* der Zellen kann die Schaltung ermittelt und damit auf die Funktionalität geschlossen werden. In der

Zelle gefundene VIAs von M1 auf M2 beschreiben Schnittstellen (Ein- und Ausgänge) für die Verbindung zwischen den Zellen und werden ebenfalls in die Standardzellenbibliothek aufgenommen. Damit kann eine Standardzelle als *black box* behandelt werden, in welcher nur deren Funktionalität und Ein- und Ausgänge zur weiteren Analyse Verwendung finden. Um das *reverse engineering* der Zellen zu automatisieren, müssen auch die Transistoren der Zellen und derer Verbindungen innerhalb der Zelle extrahiert werden. Durch graphenbasierende Ansätze wie in [?], [?] oder [?] beschrieben, kann die Funktion der Schaltung klassifiziert werden. Diese Möglichkeit wird in dieser Abhandlung jedoch nicht weiter verfolgt und damit zukünftigen Arbeiten überlassen.

Die manuell aufgebaute Standardzellenbibliothek enthält eine auf Pixelintensitäten basierende Beschreibung der Ebenen AA, GC und M1 sowie die Positionen der Ein- und Ausgänge auf M1. Durch Anwenden von Korrelationsmethoden wie der NCC oder PC (siehe Abschnitt 5.3.2) kann nun nach Zellen gesucht werden. Zu beachten ist hierbei, dass eine Zelle in vier unterschiedlichen Orientierungen (Spiegelungen) vorkommen kann. Da Korrelation variant auf Spiegelungen ist, müssen alle möglichen gespiegelten Varianten einer Zelle berücksichtigt werden. Dies erhöht den Suchaufwand um den Faktor vier. Mit einem merkmalsbasierenden Ansatz könnte hierbei Invarianz erreicht werden. In Abbildung 7.2 und 7.3 wurde ein Digitalteil nach 149 Zelltypen durchsucht und die gefundenen Zellen farblich umrahmt. Die Lokalisierung erfolgte dabei durch eine Suche nach lokalen Maxima in den ermittelten Korrelationskoeffizienten. Da eine gefundene Position meist mehrere lokale Maxima ausweist, wurde das Maximum mit dem höchsten Korrelationskoeffizienten verwendet und die restlichen Maxima zur Vermeidung von Mehrfacherkennungen unterdrückt.

Vor allem bei kleinen Zellen kann es vorkommen, dass eine größere Zelle eine ähnliche Struktur beinhaltet. Dies trifft vor allem bei Invertern zu, da viele Zellen einen nachgeschalteten Puffer beinhalten. Um Mehrdeutigkeiten zu vermeiden, fiel im Zweifelsfall die Entscheidung zugunsten der größeren Zelle aus.

7.2.2 Ableitung von Kennzahlen

Mit Kenntnis der digitalen Standardzellen und deren Positionen ist es möglich, verschiedene Kennzahlen zu ermitteln. Eine sehr nützliche Kennzahl bildet die *Flächenausnutzung* (*utilization*) ab. Diese definiert sich aus dem Verhältnis der Gesamten für die digitale Logik verwendeten Fläche zur durch Standardzellen belegten Fläche. Eine *utilization* von 1 bedeutet demnach, dass die gesamte Fläche, also alle Zeilen bzw. Spalten, mit Standardzellen ausgefüllt ist. Dieser ideale Wert ist in der Praxis jedoch nicht erreichbar, da die Verdrahtung der Zellen in der Regel mehr Fläche beansprucht. Typische Werte der *utilization* bewegen sich zwischen 0.6 und 0.8. Jede digitale Standardzellenbibliothek beinhaltet sog. *filler cells* verschiedener Größen, die in Freiräumen platziert werden können. Zweck dieser Zellen ist die Stützung der Versorgungsspannung durch Source-Gate Kapa-

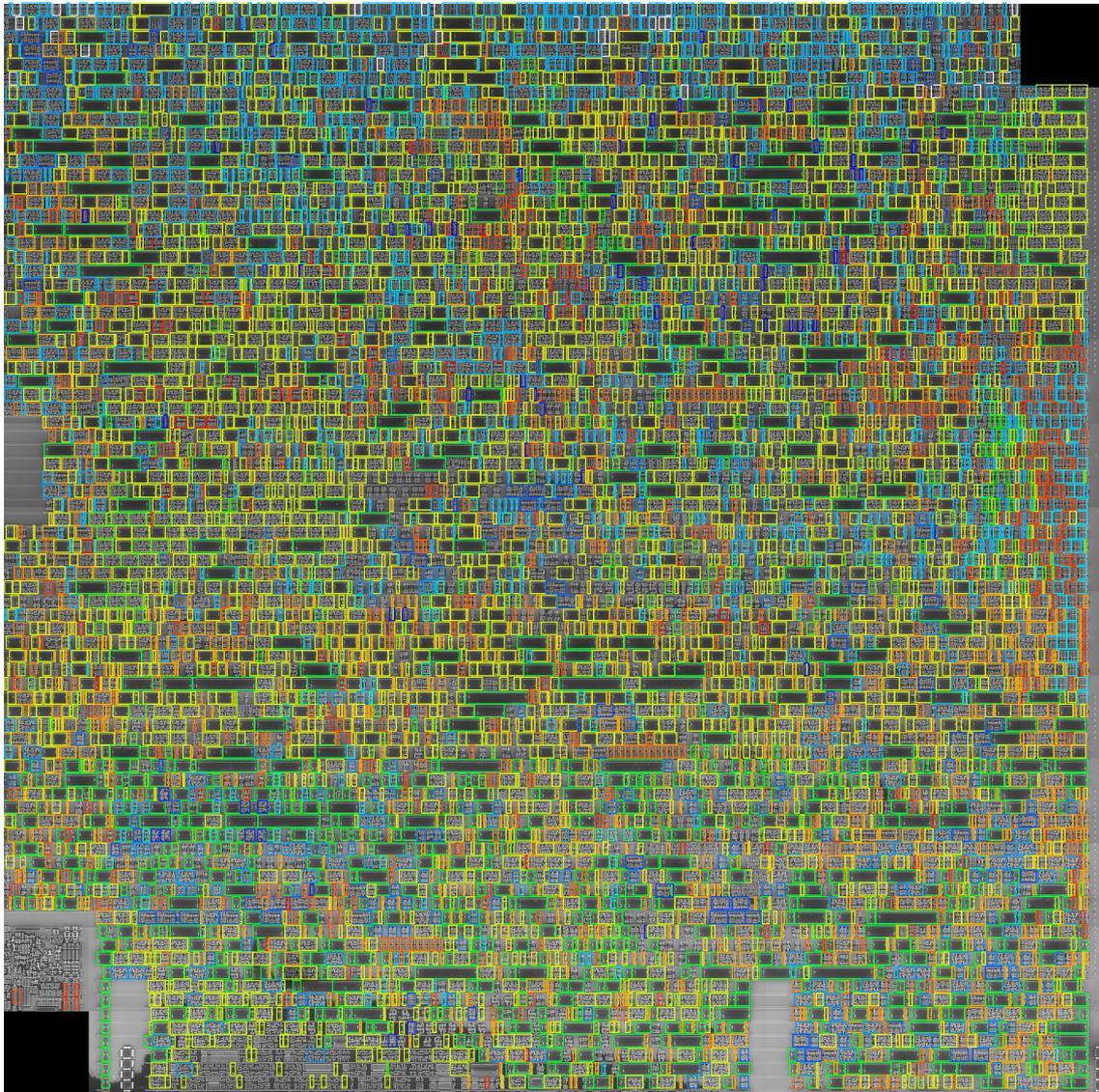


Abbildung 7.2: Standardzellenextraktion für 149 Zelltypen. In einem digitalen Chipbereich wurde nach 149 unterschiedlichen Typen von Standardzellen gesucht. Dabei wurden 21614 Zellen gefunden. Jeder Zelltyp kann dabei in vier Orientierungen (Spiegelungen) realisiert sein. Da der Matchingalgorithmus variant bezüglich Spiegelungen ist, bedeutet dies den vierfachen Rechenaufwand. Ein kleiner Ausschnitt des Bereiches ist in [Abbildung 7.3](#) vergrößert dargestellt. Statistiken über Flächen und Zellenhäufigkeiten befinden sich in [Abbildung 7.5](#).

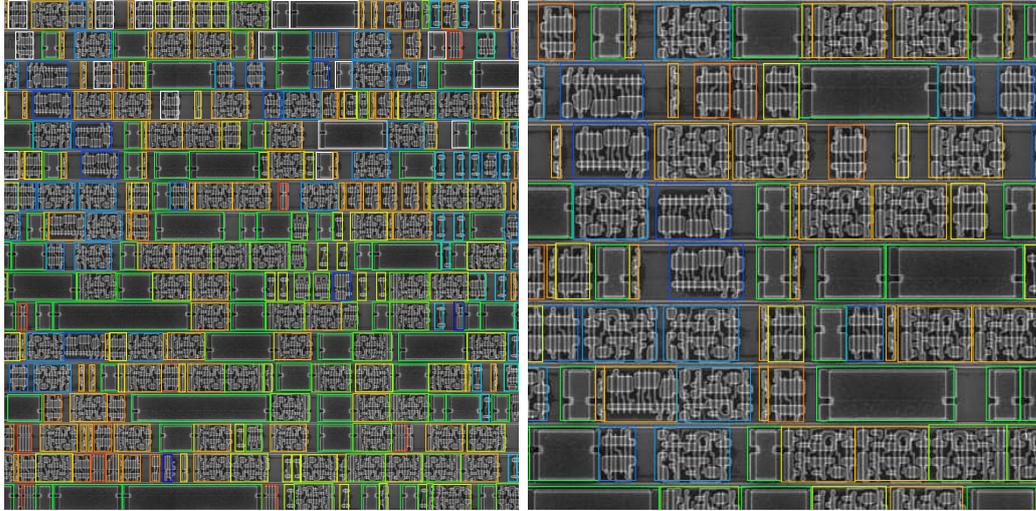


Abbildung 7.3: Ausschnitt aus einer Standardzellenextraktion für 149 Zelltypen. Zwei Bereiche aus Abbildung 7.2 wurden vergrößert dargestellt und Zellen farblich umrahmt. Etwa 37% der Chipfläche bestehen aus sog. *filler cells* (siehe Abschnitt 7.2.2), die als Stützkapazitäten zwischen den Versorgungsleitungen fungieren.

zitäten. Die Kapazitäten sorgen für die Bereitstellung von Ladungen bei Schaltvorgängen und gewährleisten dadurch eine saubere Versorgungsspannung. *Filler cells* zählen zwar zu Standardzellen, fließen jedoch nicht in die Berechnung der durch Zellen belegten Fläche A_{used} ein. Weiters ist die Verdrahtung der Zellen auf höheren Ebenen an ein festes Raster gebunden. Dadurch können Freiräume zwischen den Zellen entstehen, die ab einer Abmessung von $\text{Raster}/2$ ebenfalls der Leerfläche zuzuordnen sind. Aufgrund dieser Feststellungen kann die *utilization* u eines Digitalbereiches folgendermaßen definiert werden:

$$u := \frac{A_{digital}}{A_{used}} = 1 - \frac{A_{digital}}{A_{unused} + A_{filler}} \approx 1 - \frac{A_{digital}}{A_{filler}} \quad (7.1)$$

mit der Fläche $A_{digital}$ des gesamten Digitalbereiches, der Fläche A_{filler} aller *filler cells* und der ungenutzten Fläche A_{unused} . $A_{digital}$ kann durch Abmessung des digitalen Flächenbedarfs auf Polysilizium ermittelt werden, A_{filler} und A_{unused} durch Auffindung und Klassifikation.

Die Kennzahl der *Gatteräquivalenz* dient der Beurteilung der Komplexität der digitalen Schaltung. Mit Kenntnissen der Spezifikation des Mikrochips kann daraus auf die Effizienz der Implementierung rückgeschlossen werden. Die Gatteräquivalenz g einer digitalen Schaltung berechnet sich aus der Fläche $A_{digital}$ des gesamten Digitalbereiches, der *utilization* u und der Fläche eines einfachen Gatters wie A_{nand} oder A_{nor} zu:

$$g := \frac{uA_{digital}}{A_{nand}} \quad (7.2)$$

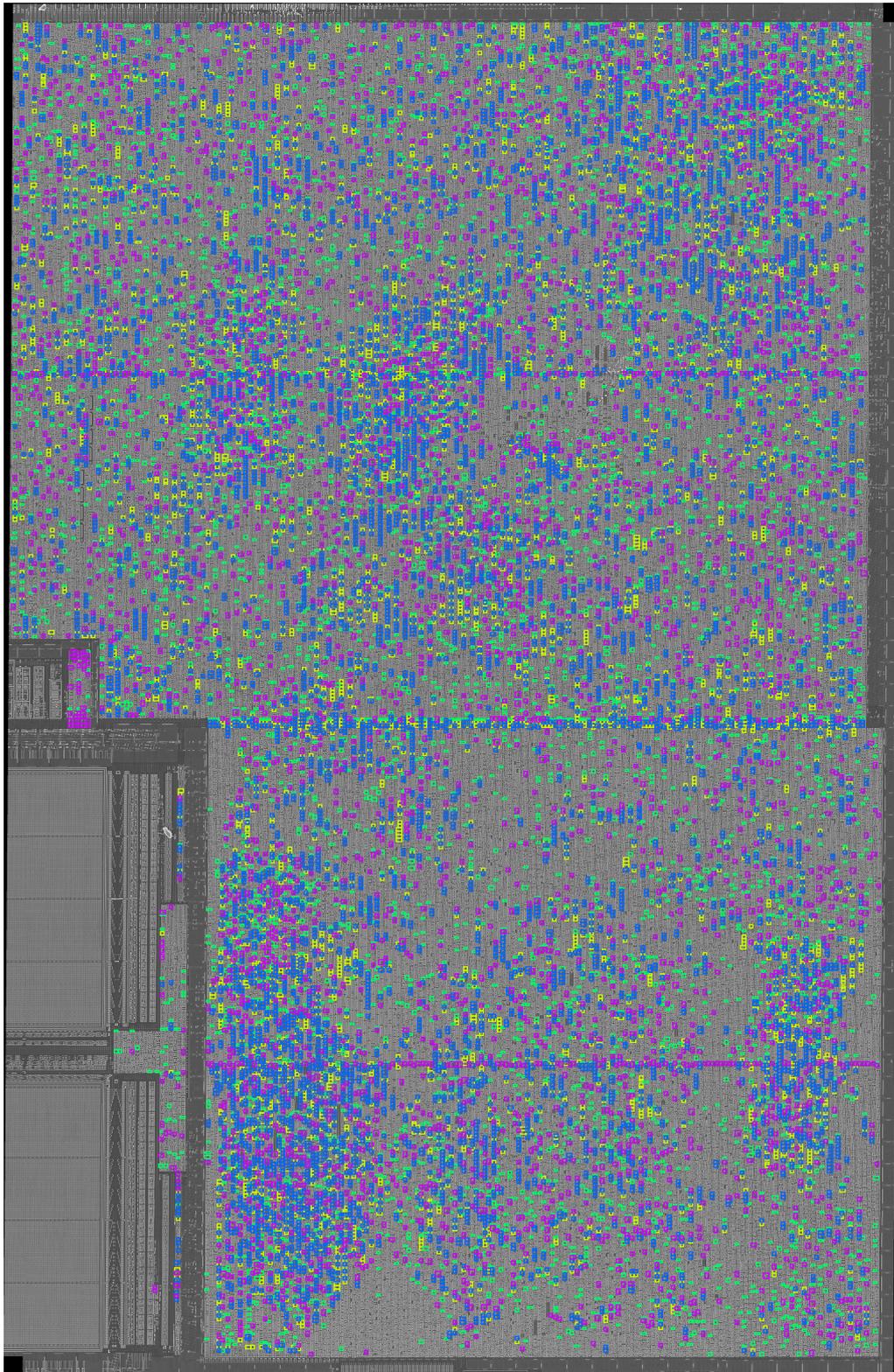


Abbildung 7.4: Effektive Flächenausnutzung in einem Standardzellenbereich. Aufgrund des Flächenbedarfs auf den Verdrahtungsebenen kann nicht die gesamte Standardzellenfläche zur Unterbringung von logischen Zellen genutzt werden. Die sich dabei ergebende Leerfläche wird mit sog. *filler cells* ausgefüllt, welche zwischen die Spannungsversorgung geschaltene *source-gate* Kapazitäten darstellen und damit der Stabilität der Schaltung dienen. Die durch Definition 7.1 festgelegte Kennzahl der *utilization* kann damit abgeschätzt werden.

Da jedoch oft nicht alle Funktionen des Mikrochips offiziell spezifiziert sind oder manche Funktionen experimenteller Natur sein können, kann dieser Wert verfälscht sein. Der Umfang dieser zusätzlichen Funktionen betrifft allerdings nur einen geringen prozentuellen Anteil der Gesamtfläche, da das Hauptaugenmerk des Entwurfs eines sich in Serienproduktion befindlichen Mikrochips in der effizienten Ausnützung der Fläche liegt.

7.2.3 Standardzellenstatistiken

Ist die Zellenbank aufgebaut und sind die Zellen gefunden, können auch ohne Kenntnis der Verdrahtung hilfreiche Informationen abgeleitet werden. Zum einen sind das die in Abschnitt 7.2.2 vorgestellten Kennzahlen und zum anderen Statistiken über Häufigkeiten und Verteilungen der Zellen. Abbildung 7.5a zeigt eine Häufigkeitsverteilung von 149 Zellentypen. Dabei ist die Wahrscheinlichkeit hoch, dass es sich bei oft vorkommenden Zellentypen um einfache grundlegende Gatter wie Inverter, NAND, NOR handelt. Dies kann eine vorab Klassifikation auch ohne Ermittlung der Schaltung einer Zelle ermöglichen. Die in Abbildung 7.5b dargestellten Flächen können ebenfalls genutzt werden, um Zellen vage zu klassifizieren. Eine ebenfalls interessante Kennzahl ist die Anzahl der verwendeten *flipflops*. Aufgrund ihrer schaltungsbedingten Rückkopplung sind diese Zellen auch ohne aufwendiges Schaltungs-*reverse engineering* relativ einfach anhand des Layouts zu klassifizieren.

7.2.4 Clock-Tree

Synchrone Logik erfordert eine stabile und saubere Zeitreferenz, damit eine definierte Datenübergabe zu jedem Zeitzyklus gewährleistet werden kann. An die Zeitbasis sind erhöhte Anforderungen in Punkto Genauigkeit, Verzögerung und Leistungsbedarf gestellt. In digitalen Schaltkreisen erfolgt die Verteilung des Taktsignals durch ein sog. *clock distribution network* bzw. einen *clock tree*. Diese bestehen zumeist aus einem Oszillator in einer Phasenregelschleife (PLL) und einem Verteilungsnetzwerk, bestehend aus einer Menge von kaskadierten Puffern, welche eine fristgerechte Verbreitung des Taktsignals sicherstellen. Demnach erfordert die kosteneffiziente Implementierung dieser Strukturen besondere Aufmerksamkeit. Eine sehr gute Darstellung dieser Thematik gibt [?].

Jede synchrone Speicherzelle (*flipflop*) besitzt einen Takteingang, der direkt mit dem *clock tree* verbunden ist. Durch Rückverfolgung der Takteingänge kann das zugrundeliegende Verteilungsnetzwerk extrahiert werden. Dazu bedarf es der Auffindung der Speicherzellen und der mit dem Takteingang verbundenen Pufferzellen. Pufferzellen werden in der Regel durch mehrstufige Inverter realisiert, deren Stromergiebigkeit vom Eingang zum Ausgang zunimmt. Damit wird hohe Stromergiebigkeit des Ausganges und geringe kapazitive Belastung des Eingangs gewährleistet.

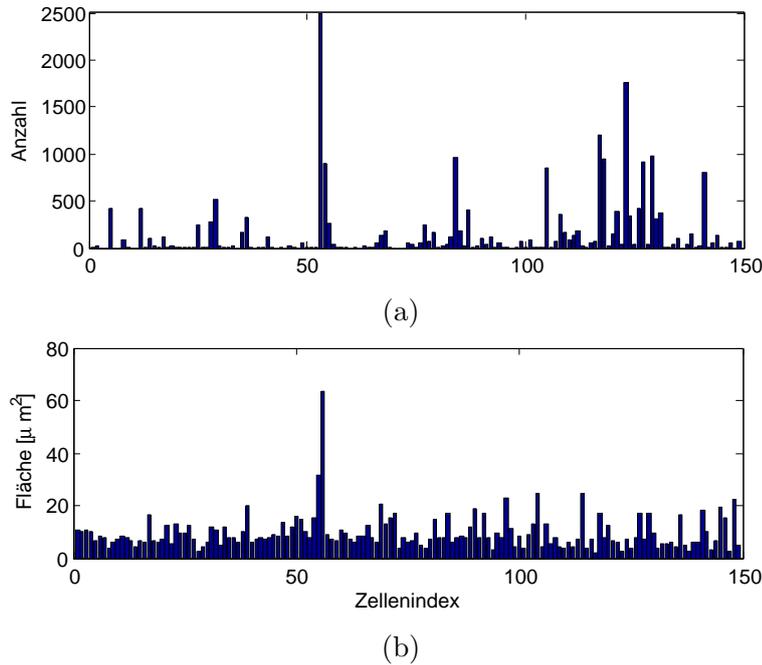


Abbildung 7.5: Statistik über Zellhäufigkeit und Zellfläche. Eine vage vorab Klassifikation der Zelltypen kann anhand der Fläche erfolgen. Den größten Platzbedarf beanspruchen in der Regel Speicherzellen (alle Arten von *flipflops*) gefolgt von komplexen Schaltungen wie Addierer oder logische Zellen mit vielen Eingängen. Bild (a) zeigt wie oft jede der 149 Zellen in Abbildung 7.2 aufgefunden wurde. Bild (b) zeigt die jeweilige Zellengröße. Nach der vorab Klassifikation in komplexe und einfache Zellen kann durch Schaltungs-*reverse engineering* die Funktion bestimmt werden. Dabei können die in Kapitel 6 vorgestellten Methoden angewendet werden.

7.3 Extraktion von Speicherzellen

Ein großer Flächenanteil von mixed-signal oder rein digitalen Mikrochips wird durch Speicherzellen (ROM, EEPROM, Flash) belegt. Diese Zellen werden zwar nicht zu Standardzellen gezählt, können jedoch aufgrund ihrer Regelmäßigkeit und ihres wiederholten Auftretens im selben Rahmen behandelt werden. Speicherzellen werden durch Reihen- und Spaltenleitungen adressiert. Die Zellen sind dabei je nach Speichertyp in Blöcken organisiert. Zum Zwecke der Datenintegrität werden pro Block ein oder mehrere Bits zur Fehlererkennung angefügt. Diese interne Organisation, welche ohne genaue Kenntnis der Implementation unbekannt ist, kann durch die Extraktion ermittelt werden. Auch die genaue Anzahl an Zellen und damit die Zelldichte kann eine nützliche Informationen darstellen.

7.4 Verteiltes Rechnen

Direkte Registrierungsmethoden stellen sehr rechenintensive Operationen dar. Auch bei Verwendung von effizienten Methoden wie in Abschnitt 5.3.2.2 beschrieben, benötigt das Auffinden von ein paar hundert Zellen in mehreren tausend Bildern erheblichen Rechenaufwand und kann viele Tage in Anspruch nehmen. Jedoch ist aufgrund der geringen Datenabhängigkeiten hohe Parallelisierbarkeit gegeben. Dies kann einerseits lokal in einem Mehrprozessorsystem durch den Einsatz von *threads* oder andererseits durch Verteilung der Gesamtaufgabe in einem Rechnerverbund erreicht werden. Auch die Kombination beider Möglichkeiten ist gültig. Der erreichbare *speedup* ist dabei durch das Gesetz von Amdahls [?] begrenzt. Dies besagt, dass die nicht parallelisierbaren sequenziellen Programmanteile den maximal durch Parallelisierung erreichbaren Geschwindigkeitszuwachs beschränkt. Da jedoch die Zellen unabhängig voneinander gesucht werden können und damit der parallelisierbare Anteil bei weitem überwiegt, ist gute Parallelisierbarkeit gegeben.

Zur Beschreibung des verwendeten Ansatzes werden folgende Begriffe definiert:

- Ein *task* beinhaltet die Suche einer Zelle in einem Bild.
- Ein *work package* legt die Suche mehrerer Zellen in einem Bild fest und besteht damit aus mehreren *tasks*.
- Der *work package pool* definiert das Gesamtproblem als die Suche von mehreren Zellen in mehreren Bildern und enthält daher mehrere *work packages*.

Die Wahl fiel auf eine Client-Server Lösung über TCP/IP. Der Server verwaltet dabei einen *work package pool*, der an anfragende Clients verteilt wird und stellt die notwendigen Daten bereit. Hat ein *client* sein *work package* abgearbeitet, so werden die Ergebnisse an den Server gesendet. Jeder *client* wiederum startet abhängig von der Anzahl der Prozessoren ein oder mehrere *threads*, die die Abarbeitung der *tasks* übernehmen.

Der erreichbare *speedup* hängt dabei von der Anzahl der zu suchenden Zelltypen ab. Bei einer kleinen Anzahl kommt der Mehraufwand stärker zu tragen, bei einer großen Anzahl skaliert die Gesamtlaufzeit gut mit der Anzahl der Rechner und Prozessoren.

Anhang A

Chipscan Analysis Tool

Im Rahmen dieser Diplomarbeit wurde die Software Chipscan Analysis Tool (CAT) erstellt. Zweck war das Schaffen einer Umgebung für die Evaluierung der in dieser Arbeit vorgestellten Algorithmen. Als Programmiersprache wurde MATLAB gewählt, da diese ein schnelles Prototyping ermöglicht. Teilweise wurden Routinen jedoch in C++ ausgelagert, um höhere Performance zu erzielen oder Datenbankverbindungen herzustellen. Unter Verwendung des MATLAB Compilers kann aus dem plattformunabhängigen Interpreter-Quellcode eine ausführbare Datei erzeugt werden, die auch ohne die Entwicklungsumgebung lauffähig ist. Jedoch benötigt man die MATLAB Laufzeitumgebung, die im Vorfeld installiert werden muss. Der Benutzer hat die Wahl, eine grafische Benutzeroberfläche (GUI) zu verwenden, kann die jeweiligen Routinen jedoch auch ohne GUI starten. Die dafür benötigten Parameter sind in den jeweiligen MATLAB Quellcode-Dateien beschrieben.

Der **Start** von CAT erfolgt durch Starten des Skripts *run.m* aus der Entwicklungsumgebung oder durch Aufruf der ausführbaren Datei *cat.exe*. Nach einer kurzen Initialisierungsphase erscheint das in Abbildung A.1 dargestellte Formular, welches der Projektkonfiguration dient.

A.1 Projektkonfiguration

Jeder Chipebene ist ein eigenes Projekt zugeordnet. Dies hat historische Gründe und kann in zukünftigen Versionen entfallen. (Aktuelle Version bei Erstellung des Dokuments ist 0.1.80.)

Vor dem Erstellen eines neuen Projektes muss durch die Radiobuttons **B** ausgewählt werden, ob die Bilder laut SSC Datei einen Überlapp besitzen. Ist dies der Fall, so ist *SSC with overlap* zu wählen. Dies trifft für SSC Dateien zu, die durch den Raith Chipscanner erzeugt wurden. Wurden die Bilder durch ein optisches Mikroskop erstellt, so sind die absoluten Positionen im Mosaik und damit der Überlapp nicht genau bekannt. Bei größeren

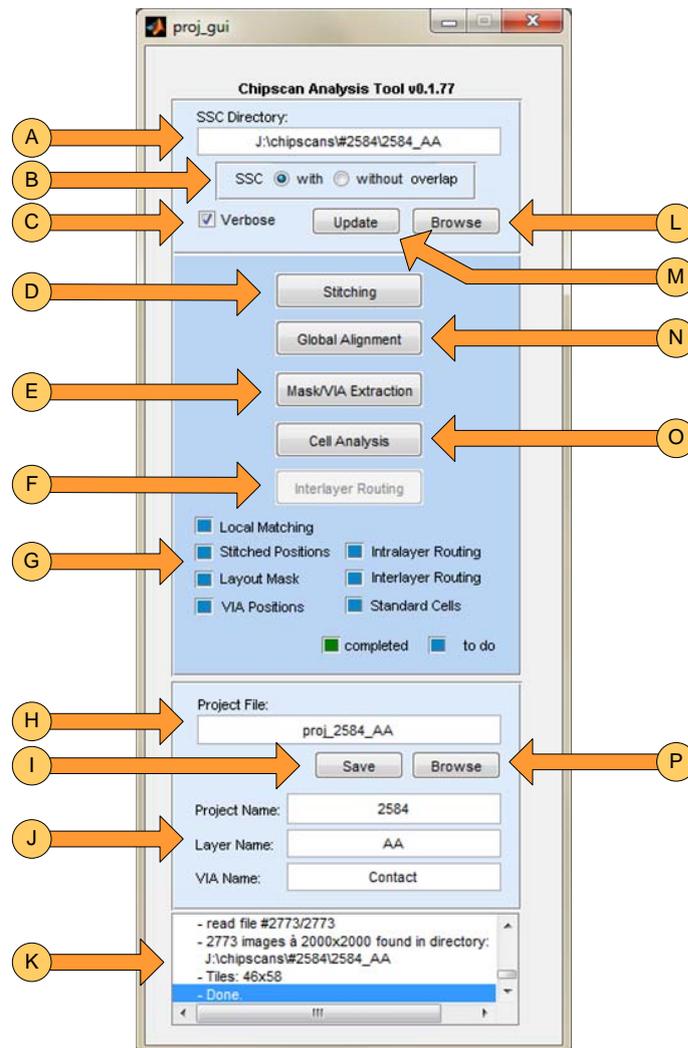


Abbildung A.1: GUI zur Projektkonfiguration. Diese Benutzeroberfläche bietet Funktionen zum Erstellen, Laden und Verwalten eines Projektes sowie zum Starten der jeweiligen Analysen.

Mosaiken kann dies zu Fehlern bei der Berechnung der Kachelanzahl führen. Deshalb sollte bei optischen Aufnahmen immer die Option *SSC without overlap* gewählt werden.

Zum Erstellen eines neuen Projektes muss durch Button **L** ein Verzeichnis ausgewählt werden, in dem sich die Bildinformations-Dateien (SSC Dateien) befinden. Diese werden nach Auswahl eingelesen und in die internen Datenstrukturen übernommen. Der gewählte Pfad erscheint dabei in Textfeld **A** und kann in späterer Folge z.B. bei einer Verschiebung der Dateien in ein anderes Verzeichnis angepasst werden. Übernommen wird die Änderung durch Drücken von Button **M**.

Tritt beim Einlesen der Bildinformations-Dateien ein Fehler auf, so kann mit aktivierter Checkbox **C** (Verbose) die Mosaikanordnung grafisch dargestellt werden. Abbildung [A.2](#)

zeigt so einen Graphen. Knoten stellen dabei Bilder dar und Übergänge visualisieren die Aufnahmereihenfolge (aufgrund des Bildindizes in den SSC-Dateinamen).

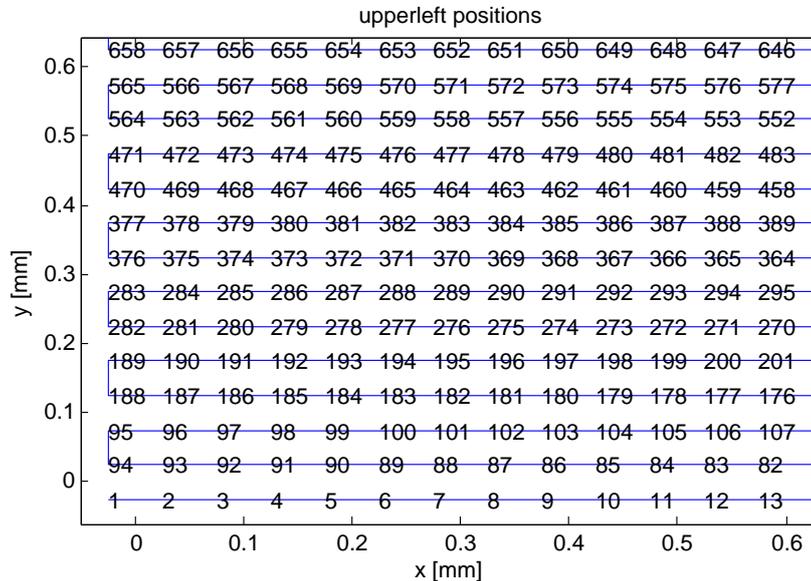


Abbildung A.2: Bildnummern und Aufnahmereihenfolge einer Bilderserie. Dieser Graph gibt einen Ausschnitt eines Mosaiks mit 46×52 Einzelbildern wieder. Die Indizes entsprechen den internen Bildnummern und die Übergänge zwischen ihnen der Aufnahmereihenfolge. In dieser Bilderserie wurde der Chip beginnend bei Index 0 mäanderförmig gerastert. Ein Bild hat eine geometrische Auflösung von $0.05 \times 0.05 \text{ mm}$ und bei einer Größe von $4MPixel$ eine Pixelabmessung von in etwa 25 nm . Treten bei der Indexvergabe Fehler auf, so können diese anhand dieses Graphen eruiert werden.

Durch Button **I** kann die Projektdatei unter dem Namen von Textbox **I** gespeichert werden. Dies erfolgt nach aufwendigeren Berechnungen automatisch oder kann bei Bedarf manuell erfolgen. Die Projektdatei wird im MATLAB *.mat* Format abgelegt und kann auch durch den MATLAB Befehl `"load projectname.mat"` manuell geladen werden.

Nach dem Erstellen des Projektes muss dem Projekt ein Projektname, ein Name der Ebene und ein Name der VIA Ebene zugeordnet werden. Dies dient nicht nur der besseren Übersicht und Organisation sondern definiert auch Schlüsselnamen, welche in einem späteren Datei-Export Verwendung finden.

Das Laden eines bereits bestehenden Projektes erfolgt durch Button **P**. Hat sich der in Editbox **A** ersichtliche Pfad zu den Bildern geändert, so kann dieser wie oben beschrieben angepasst werden. Nach erfolgreichem Laden wird der Projektstatus in Bereich **G** angezeigt. Ein grünes Quadrat steht für bereits erledigte Analysen, ein blaues für nicht ausgeführte. Bei grüner Statusanzeige können folgende Informationen abgelesen werden:

- *Stitched Positions*: Die lokale und globale Bildregistrierung wurde durchgeführt und

die Chipebenen-Transformationen ermittelt. Damit liegen Daten über die absoluten Positionen der Bilder innerhalb des Mosaiks vor. Diese Information ist Grundlage für weitere Analysen. Es besteht weiter die Möglichkeit Übersichtsbilder des gesamten Chips zu erstellen.

- *Layout Mask*: Es liegen polygonale Daten über extrahierte Metallmasken vor. Diese Daten können im Formular *mask extraction* (Button **E**) zur späteren Verarbeitung in ein Layout-Format exportiert werden. Liegen Fehlerkennungen vor, so können diese im genannten Formular editiert werden.
- *VIA Positions*: Es wurde der Schritt der VIA Extraktion ausgeführt. Die gefundenen Positionen können wie beim Maskenlayout im Formular *mask extraction* exportiert werden.
- *Intralayer Routing*: Die in einer Ebene bildweise gefundenen Polygone werden bei Überlappung zusammengeführt und bekommen eine Knotennummer. Dieser Schritt und die Extraktion der VIAs ist die Grundlage für das später folgende *interlayer routing*.
- *Interlayer Routing*: Den durch das *intralayer routing* generierten Knoten mehrerer Ebenen werden mit Hilfe der VIA-Positionen und den globalen Chipkoordinaten globale Knotennummern zugewiesen. Damit werden Verbindungen über mehrere Ebenen berücksichtigt.
- *Standard Cells*: Eine Standardzellenanalyse wurde durchgeführt. Die Positionen der Standardzellen können in eine SQL Datenbank exportiert werden und damit weiterverarbeitender Software zur Verfügung gestellt werden. Mit den globalen Knotennummern können digitale Netzlisten geschrieben werden.

Zum Erstellen oder Laden eines Projektes bieten die Buttons **D**, **N**, **E**, **O** und **F** die Möglichkeit, Formulare für die jeweiligen Analysen zu öffnen. Die Einteilung der Analysen steht dabei in engem Zusammenhang mit der gerade besprochenen Statusanzeige.

- *Stitching*: Dieses Formular bietet Möglichkeiten zur Bildregistrierung, also der genauen Lokalisierung der Einzelbilder einer Bilderserie innerhalb eines Mosaiks.
- *Mask/Via Extraktion*: Extraktion der Leiterbahnen einer Metallebene und Auffindung der Kontakte zur darüber liegenden Ebene. Diese Analyse kann auch ohne vorherige Bildregistrierung ausgeführt werden.
- *Cell Analysis*: Die Zellenanalyse sucht vordefinierte Standardzellen oder Speicherzellen innerhalb einer repräsentativen Ebene wie AA, GC oder M1.
- *Global Alignment*: Globale Registrierung mehrerer Ebenen mit Hilfe von Ankerpunkten. Dieser Schritt kann auch als *3D alignment* oder *3D stitching* bezeichnet werden.

A.2 Stitching

Unter Voraussetzung eines geladenen Projektes ist es möglich, die Benutzeroberfläche für die Bildregistrierung zu starten (Abbildung A.3). Zweck dieses Formulars ist die Konfiguration des *workflows* für die Bildregistrierung. Nach dem Start wird das Bild mit dem Index 1 in den Rahmen **F** (*Center*) geladen. Falls dieses Bild über Nachbarbilder verfügt, werden die entsprechenden Rahmen **G** (*Lower*) und **K** (*Right*) besetzt. Durch Drücken eines Button der Buttongruppe **E** kann innerhalb der Bilderserie navigiert werden. Zweck dieser Anzeige ist die Adjustierung der Überlapp-Parameter in den Textfeldern **L**. Bei Verwendung von Bilddaten aus dem Raith Chipscanner wird der Überlapp aus den SSC Dateien berechnet und ist damit vorkonfiguriert. Ein rotes bzw. gelbes Rechteck in den Bildern kennzeichnet den Überlapp nach rechts bzw. unten. Bei großen Abweichungen der gemessenen Koordinaten (laut SSC-Dateien) und den Tatsächlichen, sollten die Parameter nachkorrigiert werden, um optimale Ergebnisse zu erhalten. Bei optischen Aufnahmen ohne Überlappinformation muss die Konfiguration auf jeden Fall erfolgen. Dabei starten die roten und gelben Rechtecke mit einer Fläche von Null.

Die Parameter der Textfelder **L** haben folgende Bedeutung:

- *Additional X Overlap*: Anpassung des Überlapps zum sich rechts befindlichen Bild. Dabei werden die Werte zu den vorkonfigurierten Werten addiert (anhand der SSC Dateien). Bei Bilderserien vom Raith Chipscanner kann dieser Wert deshalb auf $\begin{pmatrix} 0 & 0 \end{pmatrix}$ belassen werden. Optische Bilder starten mit einem vorkonfigurierten Überlapp von $\begin{pmatrix} 0 & Y_{MAX} \end{pmatrix}$ und erfordern manuelle Konfiguration. Dabei wird zum vorkonfigurierten Überlapp der zusätzliche Überlapp aus dem Textfeld "Additional X Overlap" addiert. Besitzt das rechte Bild einen Überlapp von x Pixel so kann der Wert auf $\begin{pmatrix} x & 0 \end{pmatrix}$ gesetzt werden. Die Summe aus vorkonfiguriertem und zusätzlichem Überlapp entspricht dann $\begin{pmatrix} x & Y_{MAX} \end{pmatrix}$. Weiter ist noch zu erwähnen, dass auch negative Werte gültig sind, dh. es besteht die Möglichkeit, den vorkonfigurierten Y-Überlapp von Y_{MAX} durch Setzen eines negativen y zu verkleinern. Ein Beispiel hierfür wäre ein Überlapp von $\begin{pmatrix} x & -y \end{pmatrix}$ der zu einem Gesamtüberlapp von $\begin{pmatrix} x & Y_{MAX} - y \end{pmatrix}$ führt.
- *Additional Y Overlap*: Anpassung des Überlapps zum sich unterhalb befindlichen Bild. Dabei gilt ähnliches wie für den Überlapp in X-Richtung. Besitzt das untere Bild einen Überlapp von y Pixel, so kann der Wert auf $\begin{pmatrix} 0 & y \end{pmatrix}$ gesetzt werden. In Summe entspricht dies einem Überlapp von $\begin{pmatrix} X_{MAX} & y \end{pmatrix}$. Für Bilder vom Raith Chipscanner gilt wiederum ein Wert von $\begin{pmatrix} 0 & 0 \end{pmatrix}$.
- *Additional Search Overlap*: Sind die Überlappungen in x/y Richtung konfiguriert, kann das Suchfenster im zentralen Bild angepasst werden. Dies bietet die Möglichkeit,

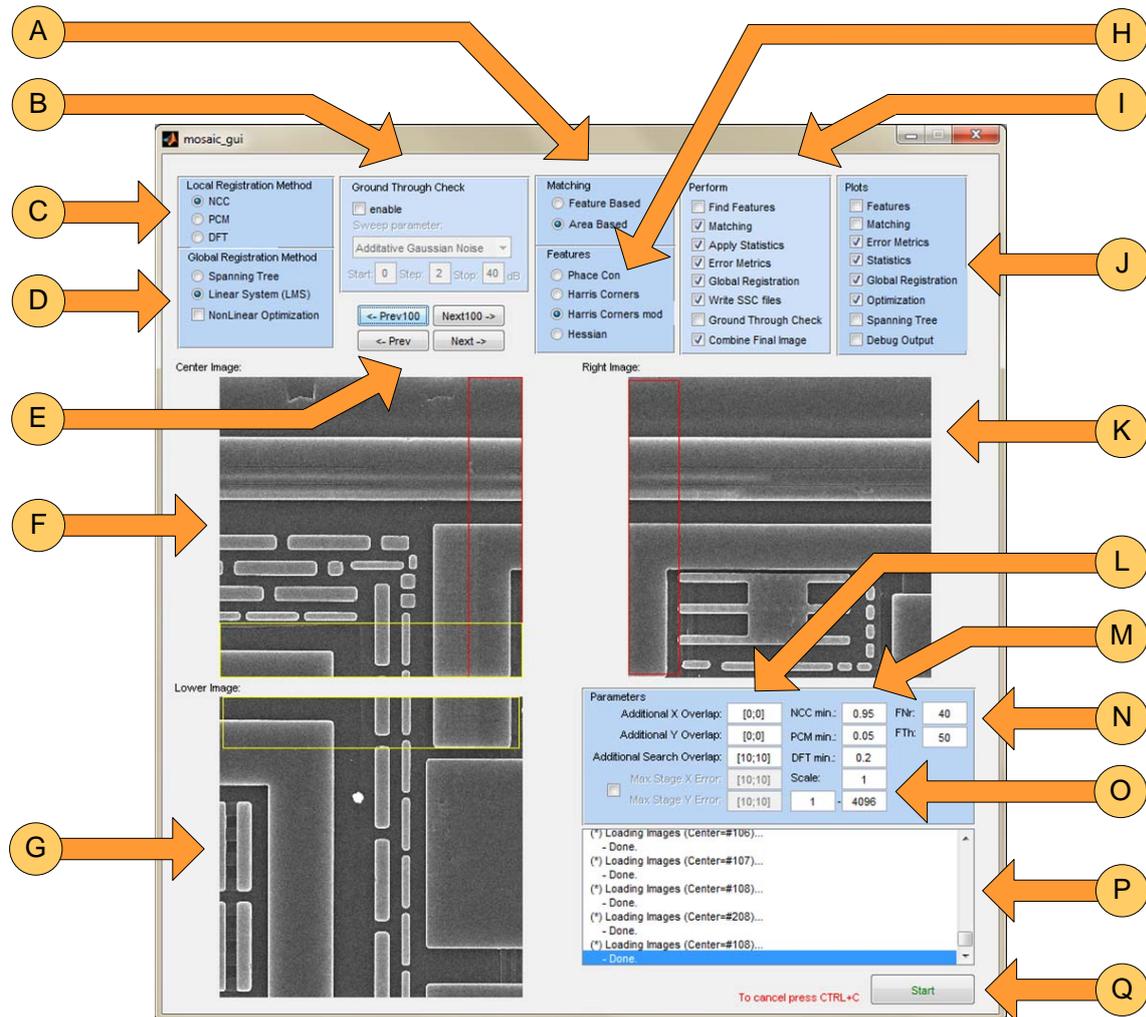


Abbildung A.3: GUI zur Bildregistrierung.

Abweichungen des Überlapps innerhalb einer Bilderserie zu korrigieren. Als Standardeinstellung werden 10 Pixel vorgeschlagen. Vor allem bei optischen Bildern, bei denen meist der Probenstisch manuell verschoben wird, sollte man diesem Parameter besondere Aufmerksamkeit widmen. Um gute Ergebnisse zu erzielen, sollte man wie weiter unten beschrieben sicherstellen, dass sich der im rechten bzw. unterem Rechteck befindliche Bildinhalt immer deutlich innerhalb der entsprechenden Rechtecke des zentralen Bildes befindet.

Allgemein sollte der Überlapp so konfiguriert sein, dass sich der im roten bzw. gelben Rechteck befindliche Bildinhalt des rechten bzw. unteren Bildes innerhalb der zugehörigen Rechtecke des zentralen Bildes befindet. In der Praxis hat es sich als günstig erwiesen, die Bilderserie durch Buttongruppe **E** zu prüfen, um etwaige Verschiebungen

innerhalb der Serie zu erkennen und Parameter dahingehend anzupassen. Insbesondere gilt dies bei Verwendung der lokalen Registrierungsmethode NCC, da bei fehlerhafter Konfiguration die Subpixel-Genauigkeit sinkt. Um ein Gefühl für die Parameter zu bekommen, empfiehlt es sich, Änderungen vorzunehmen und die Auswirkungen auf die graphische Ausgabe zu beobachten.

Der grundsätzliche Ablauf des Registrierungsverfahrens kann in der Gruppe **E** (*Perform*) festgelegt werden. Die Reihenfolge der Parameter entspricht dabei dem chronologischen Ablauf des Algorithmus. Die Optionen haben folgende Bedeutung:

- *Matching*: Durchführung der lokalen Registrierung. Die Wahl der Methode erfolgt je nach Konfiguration der Parameter in **A**, **C** und **H**. Dabei werden die relativen Positionen der Nachbarbilder berechnet. Dieser Schritt gilt als aufwendigster und kann bei Neuberechnung der globalen Registrierung entfallen. Dies ermöglicht die Nachjustierung von Parametern nachfolgender Schritte, ohne die lokale Registrierung neu berechnen zu müssen.
- *Statistics*: Eliminieren der als ungenau angenommenen Resultate der lokalen Registrierung und Berechnung von statistischen Kenngrößen. Diese Option sollte immer aktiviert sein.
- *Global Registration*: Globale Registrierung unter Verwendung der Resultate der lokalen Registrierung nach Methode **D**. Nach Ausführung dieses Schrittes ist die genaue Position eines Bildes innerhalb des Mosaiks bekannt. Nun können neue angepasste Koordinaten-Dateien geschrieben oder ein Übersichtsbild erstellt werden.
- *Write SSC*: Schreiben der Koordinatendateien einer Bilderserie. Dabei werden die zuvor berechneten absoluten Bildpositionen in die entsprechenden Koordinatendateien geschrieben. Das Verzeichnis entspricht dabei dem im Projektformular festgelegten.
- *Ground Through Check*: Ausführen von *ground through* Vergleichen um die erreichte Genauigkeit bei Verwendung spezieller Datensätze zu verifizieren. Dies hat für eine praktische Verwendung keine Bedeutung und sollte deaktiviert sein.
- *Combine Final Image*: Erstellen eines Übersichtsbildes der gesamten Chipebene. Da diese Bilder sehr groß ausfallen können, gibt Parameter **O** die Möglichkeit, Auflösung sowie Anzahl der Bilder zu beschränken.

Nicht alle Datenabhängigkeiten wurden bei der Überprüfung der möglichen Konfiguration berücksichtigt. Bei ungeeigneter Wahl kann dies zu Abbrüchen führen. Die Standardeinstellung ist jedoch für die meisten Anwendungen gut geeignet. Wurde bereits eine lokale Registrierung (*Matching*) durchgeführt und sollen nun geänderte Parameter nach Auswahl **L** berücksichtigt werden, sollte die Option *Matching* deaktiviert

werden, um unnötige Rechenzeit zu vermeiden. Da dieser Schritt einen Hauptteil der Rechenzeit ausmacht, ist die Deaktivierung bei großen Bilderserien unbedingt erforderlich.

Nachfolgend werden die Parameter des Registrierungsverfahrens beschrieben:

Auswahlgruppe **A** definiert die lokale Registrierungsmethode. Hier stehen merkmalsbasierende Ansätze nach Auswahl **H** und direkte Verfahren nach Auswahl **C** zur Wahl. Merkmalsbasierende Ansätze liefern meist nur bei optischen Bildern gute Ergebnisse, da diese in der Regel reich an Information sind. Bei Aufnahmen eines REM liefern direkte Verfahren im Mittel bessere Ergebnisse.

Gruppe **C** stellt die verfügbaren direkten lokalen Registrierungsmethoden zur Wahl. Als Standardeinstellung wird die NCC vorgeschlagen. Es werden dabei folgende Optionen geboten (vgl. Abschnitt 5.3.2):

- *NCC*: Normalisierte Kreuzkorrelation mit Subpixel-Genauigkeit (1/20 eines Pixels) durch Interpolation der Korrelationskoeffizienten. (vgl. Abschnitt 5.3.2.3)
- *PCM*: Phasenkorrelation mit Pixel-Genauigkeit. (vgl. Abschnitt 5.3.2.1)
- *DFT*: Phasenkorrelation mit Subpixel-Genauigkeit durch *upsampling* des Bildes nach [?]. Genauigkeit liegt im Bereich von 1/10 eines Pixels.

In Punkto Robustheit und Genauigkeit hat sich in der Praxis die Methode der NCC als Mittel der Wahl erwiesen. Die weiteren Methoden sind zumeist experimenteller Natur und können für zukünftige Verbesserungen Anreiz bieten. Robuste subpixelgenaue Registrierung wurde jedoch nur mit sehr schwankenden Ergebnissen erzielt.

Durch die Parameter in Gruppe **M** können Schwellwerte gesetzt werden, die bestimmen, ab welcher Höhe des Korrelationskoeffizienten die dazugehörigen Transformationen als genau genug bewertet werden können, um in weitere Berechnungen einzufließen. Eine Anpassung der Parameter sollte aufgrund der weiter unten beschriebenen Diagramme erfolgen.

Fällt die Wahl auf einen merkmalsbasierenden Ansatz, so kann in Gruppe **H** der zu verwendende Merkmalstyp gewählt werden. Es stehen folgende Optionen zur Wahl:

- *Phase Con*: Phase Congruency Detektor nach [?].
- *Harris Corner*: Harris Corner Detektor nach [?].
- *Harris Corner mod*: Modifizierte Variante des Harris Corner Detektors wie in [?] vorgeschlagen.
- *Hessian*: Berechnung der Eckigkeit durch die Hesse Matrix. (siehe Abschnitt 5.3.1)

Für eine nähere Beschreibung siehe 5.3.1.1. Die Konfiguration des Schwellwertes für die erforderliche Signifikanz eines Merkmals sowie die maximal zu verwendende Anzahl an Merkmalen kann in Gruppe **N** vorgenommen werden.

Auswahlgruppe **D** legt die globale Registrierungsmethode für den Bündelausgleich fest. Nach der lokalen Registrierung, bei der die relativen Bildtransformationen ermittelt wurden, folgt nun die Berechnung der absoluten Positionen innerhalb des Mosaiks. Da das Problem überbestimmt ist, muss die Lösung mit Ausgleichsrechnung bestimmt werden. Dazu stehen folgende Möglichkeiten zur Wahl:

- *Spanning Tree*: Lösung durch Aufbau eines minimalen Spannbaumes, welcher die lokalen Bildtransformationen unter Berücksichtigung eines Fehlermaßes in Beziehung setzt.
- *Linear System*: Die lokalen Transformationen werden in ein gewichtetes lineares Gleichungssystem eingefügt und die Lösung per Ausgleichsrechnung (LSM) bestimmt. Man erreicht dadurch eine Aufteilung der Abweichungen auf das gesamte Mosaik.
- *Nonlinear Optimization*: Formulierung als nichtlineares Optimierungsproblem.

In Abschnitt 5.4.3 werden die evaluierten Methoden näher beschrieben. Die Methoden *Linear System* und *Nonlinear Optimization* liefern sehr ähnliche Ergebnisse, wobei jedoch nur erstere auch bei sehr großen Mosaiken (> 5.000 Bilder) praktikable Laufzeiten besitzt. Die *Spanning Tree* Methode liefert im Mittel Mosaik schlechterer Qualität und sollte daher nicht verwendet werden.

Zwischen der Ausgabe von diversen Diagrammen kann in Auswahlgruppe **J** gewählt werden:

- *Features*: Bildliche Darstellung der extrahierten Merkmale.
- *Matching*: Informationen zur lokalen Registrierung. Ausgabe der Bilder der Überlappungsbereiche sowie dreidimensionale Darstellung der Korrelationskoeffizienten. (siehe Abbildung A.4)
- *Statistics*: Häufigkeitsverteilung der Korrelationskoeffizienten anhand derer die Parameter in Gruppe **M** angepasst werden können (siehe Abbildung 5.10). Fehlerstatistik der zirkulären Transformationen (siehe Abschnitt 5.5.2).
- *Global Registration*: Ausgabe des *connected component graphs* (CCG). Ausgabe der in Abschnitt 5.5.2 und Abbildung 5.14 beschriebenen grafischen Darstellung der in Verbindung stehenden lokalen Transformationen.
- *Spanning Tree*: Grafische Ausgabe des ermittelten Spannbaumes.

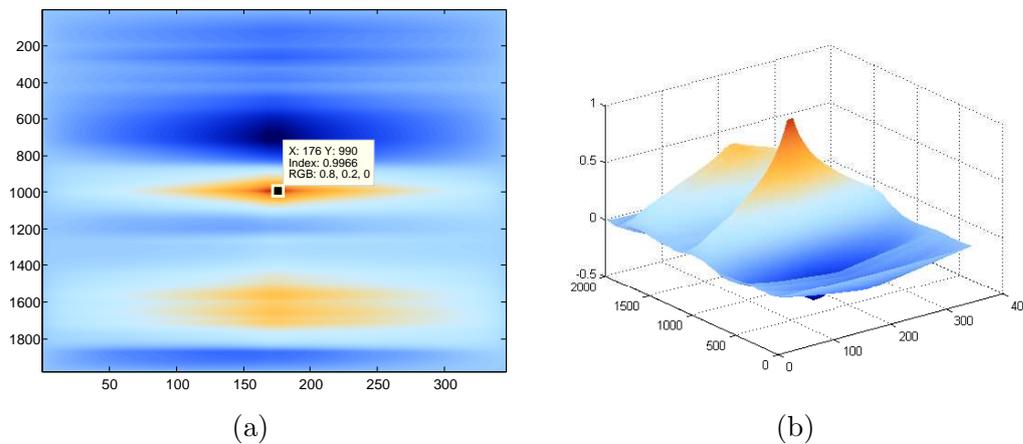


Abbildung A.4: Korrelationskoeffizienten der Debug-Ausgabe. Das globale Maximum der Korrelationsfunktion sollte sich, wie hier dargestellt, immer innerhalb des angezeigten Ausschnittes befinden. Ist dies nicht der Fall, so ist der Überlappungsbereich in den Textfeldern **L** mit hoher Wahrscheinlichkeit fehlerhaft konfiguriert.

Abschließend seien noch die Parameter des Schrittes *Combine Final Image* (Gruppe **I**) erwähnt. Bei großen Bilderserien ist es notwendig, einen Skalierungsfaktor $\in [0, 1]$ zu wählen oder die Anzahl der beitragenden Bilder durch Festsetzen eines Start und Endindex zu begrenzen. Für Bilderserien mit bis zu 3000 Bildern ist ein Skalierungsfaktor von 0.5 empfohlen, um eine Bildgröße von 200MB nicht zu überschreiten. Für Bilder mit bis zu 6000 Bildern wird ein Skalierungsfaktor von 0.3 empfohlen. Bei höherer Bildanzahl sollte der Parameter entsprechend angepasst werden. Bei zu hoher Wahl führt ein "Out of memory" Fehler zu Abbruch des Programmes (Ausgabe **P**).

A.3 Leitungs- und Kontaktextraktion

Das Formular der Leitungs- und Kontaktextraktion aus Abbildung A.5 bietet die Möglichkeit, Leiterbahnen und Kontakte (VIAs) aus Metallisierungsebenen zu extrahieren. Dazu stehen mehrere Algorithmen zur Wahl, um ein möglichst breites Spektrum an Bildtypen abdecken zu können. Zur Auswahl (**C**) stehen:

- *Edge Linking*: Abhängig von der Option *Adaptive Thresholding* kann zwischen kantenbasierender (vgl. 6.3.2.1) (*Adaptive Thresholding* deaktiv) und regionenbasierender Segmentierung unterschieden werden. Bei nicht flächigem Aussehen der Leiterbahnen muss die Option deaktiviert werden.
- *Hough Transform*: Auffindung der Leitungsgrenzen durch die Hough Transformation. Nähere Informationen können in Abschnitt 6.3.2.2 nachgelesen werden.

- *Split and Merge*: Klassische Split and Merge Segmentierung. Liefert in der Regel übersegmentierte Ergebnisse und ist nur bei flächigen Leiterbahnen anwendbar.
- *Graph based*: Es gilt ähnliches wie bei *Split and Merge*.

Generell liefert die Methode *Edge Linking* mit deaktivierter Option *Adaptive Thresholding* gute Ergebnisse bei einer Vielzahl an Bildtypen. Wichtig hierbei ist die korrekte Wahl der Parameter **J**, **K** und **L**, die weiter unten beschrieben werden. Kritisch sind allerdings Leitungen mit ausgeprägtem Kanteneffekt, da hierbei die Kante doppelt erkannt wird. Durch Anpassung der Parameter **J** und **K** kann dieser Effekt verringert werden. Ist die Leiterbahn flächiger Natur können auch mit aktivierter Option *Adaptive Thresholding* gute Ergebnisse erzielt werden. Die alternativen Methoden sind wie schon bei der Bildregistrierung größtenteils experimenteller Natur und teilweise noch nicht ausgereift. Die *Hough Transform* liefert bei kurzen Leiterbahnen unzureichende Resultate.

Die Parameter **J** legen die Hysterese des Canny Kantendetektors fest. Mit Hilfe dieser Parameter wird entschieden, ob eine Kante signifikant ist. Eine signifikante Kante muss mindestens einen Kantenpixel mit einer Kantenstärke größer *Canny Max Threshold* besitzen. Die restlichen zur Kante gehörenden Pixel müssen eine Kantenstärke von mindestens größer *Canny Min Threshold* aufweisen. In Abschnitt 6.3.2.1 wird das Verfahren ausführlich beschrieben.

Parametersatz **K** legt die Eigenschaften des gauß'schen Filterkerns fest. Jedes Bild wird in einem Vorverarbeitungsschritt durch einen Tiefpass gefiltert um Intensitätsrauschen zu dämpfen. Der Filterkern sollte dabei nicht zu groß gewählt werden um Kanten zu erhalten. Gültige Werte für die Größe des Kerns sind $2n+1$.

Parametersatz **L** bezieht sich auf die Vektorisierung (bzw. Polygonalisierung) und legt dessen Randbedingungen fest. Dabei haben die Parameter folgende Bedeutung:

- *Discard Contours Less Than*: Kantenzüge unter dieser Mindestlänge werden verworfen. Viele durch Intensitätsrauschen verursachte Kantenfehlerkennungen können dadurch gefiltert werden. Mit der kleinsten vorkommenden Strukturbreite d_{min} sollte dieser Wert $d_{min}\pi$ nicht unterschreiten. (Sehr kleine Strukturen haben abgerundete Ecken und erscheinen deshalb kreisförmig.)
- *Maximum Deviation From Original Edge*: Definiert die maximal erlaubte Abweichung der polygonalen Beschreibung von der Kantenbeschreibung in Pixel. Je kleiner dieser Wert gesetzt wird, desto genauer wird die Kantenbeschreibung approximiert. Allerdings steigt damit auch die Anzahl der notwendigen Stützstellen. Ein zu groß gewählter Wert kann zur Entartung der ursprünglichen Form führen. Hierbei ist ein Kompromiss zwischen Genauigkeit und Anzahl der Stützstellen zu treffen.

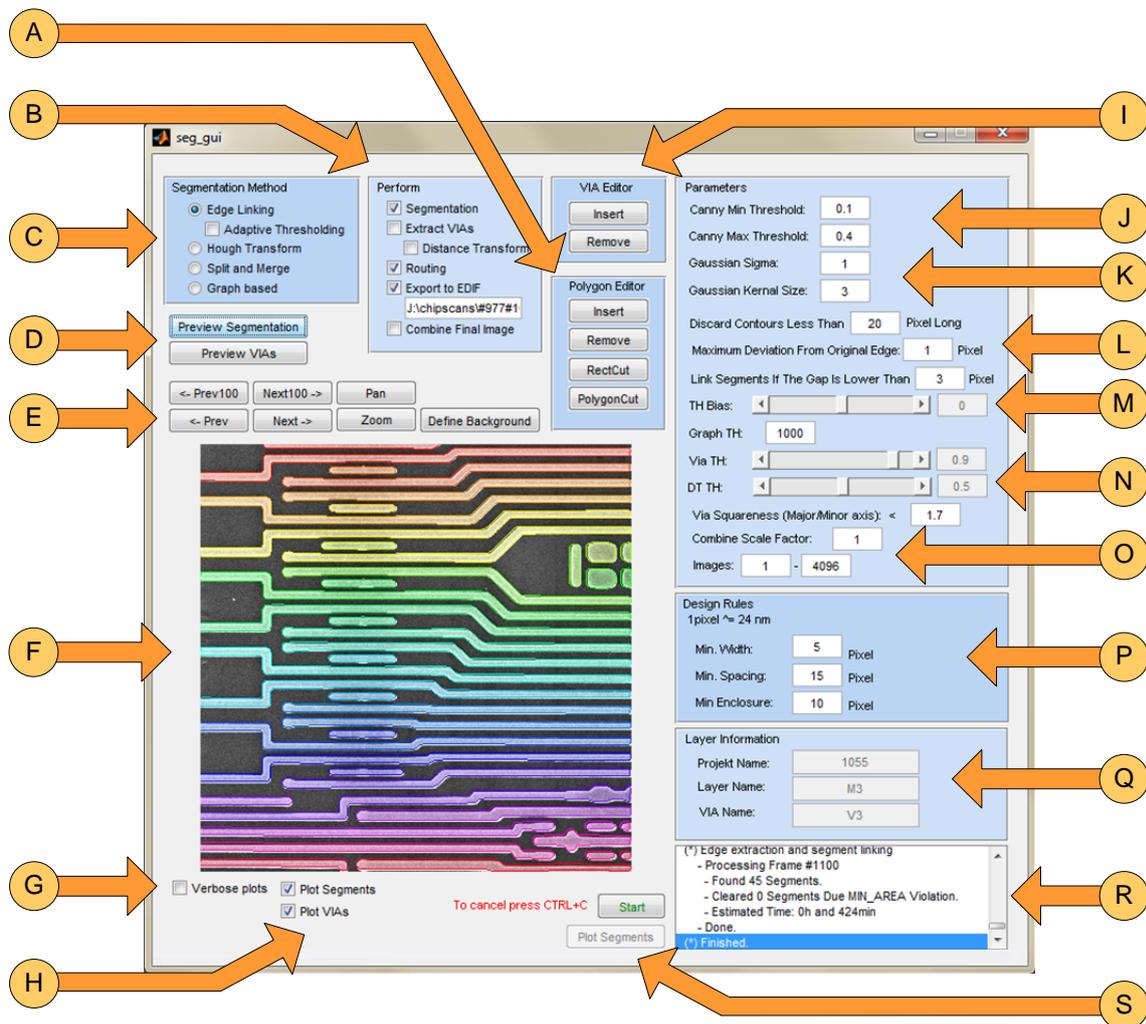


Abbildung A.5: GUI zur Leitungs- und Kontaktextraktion.

- *Link Segments If The Gab Is Lower Than*: Es werden nur Kantenzüge verarbeitet, die geschlossen sind. Kantenzüge, bei denen Anfangs- und Endpunkte nicht zusammenfallen werden jedoch verworfen. Befinden sich Anfangs- und Endpunkt eines Kantenzuges in einer Distanz kleiner d_{gab} , so kann durch diesen Parameter veranlasst werden, diesen zu schließen.

Eine nähere Beschreibung findet man in Abschnitt 6.3.3.

Um die gewählte Methode testen zu können, stehen die Preview-Buttons **D** zu Verfügung. In der Regel sollte dies immer vor dem Start der Analyse erfolgen, um sicherzustellen, dass die gewählte Methode unter den definierten Parametern zufriedenstellende Ergebnisse liefert. Entsprechen die Resultate nicht dem gewünschten

Ergebnis, so sollte Option **G** zugeschaltet werden (siehe weiter unten).

Wie schon bei der Bildregistrierung bietet Auswahl **B** die Möglichkeit, den durch Button *Start* (Buttongruppe **R**) zu startenden Ablauf zu konfigurieren. Dabei stehen folgende Schritte zur Auswahl:

- *Segmentation*: Dabei werden mit Hilfe der gewählten Segmentationsmethode **C** Leiterbahnen extrahiert und als Polygonmodell gespeichert. Dieser Schritt ist der rechenaufwendigste und kann je nach Bildanzahl und Rechner mehrere Stunden in Anspruch nehmen. Eine Ausgabe des aktuellen Status ist aus Listbox **Q** zu entnehmen.
- *Extract VIAs*: Diese Option startet die Kontaktextraktion für die aktuelle Ebene. Anwendung findet dabei der in Abschnitt 6.4 dargestellte Algorithmus. Die Implementation der Hough Transformation für Kreise ist durch diese GUI nicht wählbar. Voraussetzung hierbei ist ausreichender Kontrast zwischen Leiterbahnen und VIAs. Die Option *Distance Transform* kann gewählt werden, falls sich VIAs durch eine einfache Schwellwertbildung als nicht trennbar erweisen.
- *Routing*: Zuweisung von Knotennummern zu den einzelnen Polygonen. Dieser Schritt erfolgt bildübergreifend und berücksichtigt über mehrere Bilder verlaufende Leiterbahnen. Damit wird jeder gefundenen Leiterbahn eine eindeutige Kennung zugewiesen. Diese Daten stellen die Ausgangsdaten für ein späteres lagenübergreifendes *routing* dar.
- *Export to EDIF*: Mit Wahl dieser Option wird das berechnete Polygonmodell durch Aufruf eines externen Konvertierungsprogrammes, im Dateiformat EDIF, unter dem angegebenen Dateinamen abgelegt. Dies ermöglicht die Importierung in professionelle Software wie *Cadence Virtuoso* oder auch *AutoCAD*.
- *Combine Final Image*: Die extrahierten Polygone werden unter Verwendung der in Gruppe **O** festgelegten Parameter in ein Binärbild gezeichnet. Das erzeugte Bild im PNG Format ist im Verzeichnis **B** (Abbildung A.1) unter *img_combined_scale_x_poly.png* zu finden.

Durch Festlegen von der sich in Gruppe **P** befindlichen *design rules* können Randbedingungen für die konforme Koexistenz von Leiterbahnen festgelegt werden. Die minimale Weite fordert hierbei eine minimale Fläche, die eine Leiterbahn erfüllen muss. Die restlichen beiden Parameter werden in Version 0.1.80 noch nicht berücksichtigt.

Die in **P** angezeigten Informationen entsprechen den in Formular A.1(J) konfigurierten und werden bei der Exportierung in ein externes Dateiformat berücksichtigt.

Option *Verbose Plots* (**G**) erlaubt die Ausgabe von Debug-Informationen um Auswirkungen von Parameteranpassungen zu verfolgen. Bei Verwendung der kantenbasierenden Segmentierung (Auswahl **C**, *Edge Linking* ohne *Adaptive Thresholding*) wird in etwa das Kantenbild des Canny Detektors ausgegeben. Wurden zu wenige oder zu viele Kanten gefunden, sind die Schwellwerte **J** falsch gesetzt und müssen angepasst werden. Die Debug-Ausgaben der Kontaktextraktion zeigen die durch globale Schwellwertbildung gefundenen VIA-Regionen.

Sollen bereits extrahierte Daten in Ausgabe **F** angezeigt werden, kann dies durch die Optionen **H** bestimmt werden.

Die Parameter zur Kontaktextraktion nach Algorithmus 8 befinden sich in Optionsgruppe **N** und haben folgende Bedeutung:

- *Via TH*: Globaler Schwellwert $\in [0, 1]$ zur Segmentierung der VIAs. Werden zu wenige VIAs gefunden, so kann dieser Wert nach unten angepasst werden. Dabei ist ausreichend Kontrast zwischen Leiterbahn und VIA gefordert. Debug-Ausgaben können durch die Option *Verbose Plots* (**G**) zugeschaltet werden.
- *DT TH*: Können Kontakte durch eine globale Schwellwertbildung nicht getrennt werden, so besteht die Möglichkeit, die Regionen mit Hilfe der Distanztransformation zu treffen. Der Parameter *DT TH* gibt dabei den globalen Schwellwert an, der auf die normalisierte Distanzmatrix angewandt wird. Für eine genauere Beschreibung siehe Abschnitt 6.4.
- *VIA Geometry*: Randbedingung betreffend der Geometrie, die eine VIA-Region zu erfüllen hat. Dabei ist gefordert, dass der Quotient zwischen der Länge der Hauptachse und der Länge der Nebenachse einen bestimmten Wert nicht überschreitet. Siehe dazu Abschnitt 6.4 oder Algorithmus 8. Ein rundes oder quadratisches Objekt besitzt einen Wert von 1. Werte zwischen 1.3 und 1.7 haben sich in der Praxis als günstig erwiesen. Soll diese Randbedingung nicht berücksichtigt werden, kann dies durch Eingabe eines hohen Wertes unterbunden werden.

Die nach Durchlauf der Analyse vorhandenen Daten können bei Wahl der Option *Plot Segments* im Rahmen **F** für jedes Einzelbild durchgesehen werden. Ist ein Übersichtsbild des gesamten Analysebereiches gewünscht, so kann dies durch Button *Plot Segments* (**S**) in einem extra Fenster angezeigt werden.

Sollen fehlerhafte Erkennungen gelöscht oder fehlende Polygone zugefügt werden, so bietet hierzu der Polygon-Editor (**A**) rudimentäre Funktionen. Dazu stehen folgende Möglichkeiten zur Wahl:

- *Insert*: Einfügen eines Polygons im aktuellen Bild **F**. Durch Drücken der linken Maustaste werden die Eckpunkte des (simplen) Polygons festgelegt. Drücken der rechten Maustaste beendet die Eingabe und übernimmt das Polygon.
- *Remove*: Löschen eines Polygons durch Angabe eines sich im Polygon befindlichen Punktes. Das Setzen des Punktes erfolgt durch die rechte Maustaste.
- *RectCut*: Aufspannen eines Rechteckes zum Löschen der sich darin befindlichen Polygone. Die linke Maustaste legt den linken oberen Eckpunkt des Rechteckes fest. Die rechte Maustaste den rechten unteren Eckpunkt. Das Rechteck ist dabei parallel zu den Achsen ausgerichtet.
- *PolyCut*: Definition eines Polygons zum Löschen der sich darin befindlichen Polygone. Ähnlich wie beim Einfügen eines neuen Polygons werden die Eckpunkte durch Drücken der linken Maustaste festgelegt. Die rechte Maustaste beendet die Eingabe.

Auch für Kontakte gibt es die Möglichkeit der Nacharbeit durch Buttongruppe **I**:

- *Insert*: Einfügen eines neuen Kontaktes im aktuellen Bild **F**. Die Definition der Position erfolgt durch Drücken der rechten Maustaste.
- *Remove*: Löschen eines fehlerhaften Kontaktes. Die Wahl des Kontaktes erfolgt durch Drücken der rechten Maustaste. Um die Eingabe zu erleichtern wird der Kontakt gelöscht, der sich am Nächsten an der angegebenen Position befindet.

A.4 Zellenextraktion

Die Suche nach digitalen Standardzellen erfolgt durch das Formular in Abbildung A.6. Dieser Schritt erfordert eine zuvor erstellte Datenbank, welche das Aussehen der Zellen sowie Informationen über die Positionen der Ein- und Ausgabe Anschlüsse bereitstellt. Es wird dabei zwischen drei Tabellen unterschieden:

- *Definition table*: Tabelle aller Standardzellen. Diese Tabelle stellt das Bild der Standardzelle sowie Informationen über die Anschlüsse bereit.
- *Object table*: In dieser Tabelle werden die Positionen und Orientierungen der gefundenen Standardzellen eingetragen. Ein Eintrag erfolgt nur bei entsprechender Auswahl in Gruppe **C**.
- *Drawing table*: Informationen über Darstellungsparameter für den späteren Import in externer Software.

Mit Angabe des zur Datenbank führenden Pfades sowie der Tabellennamen in **L** werden die Daten durch Drücken von Button **M** übernommen. Dabei wird die

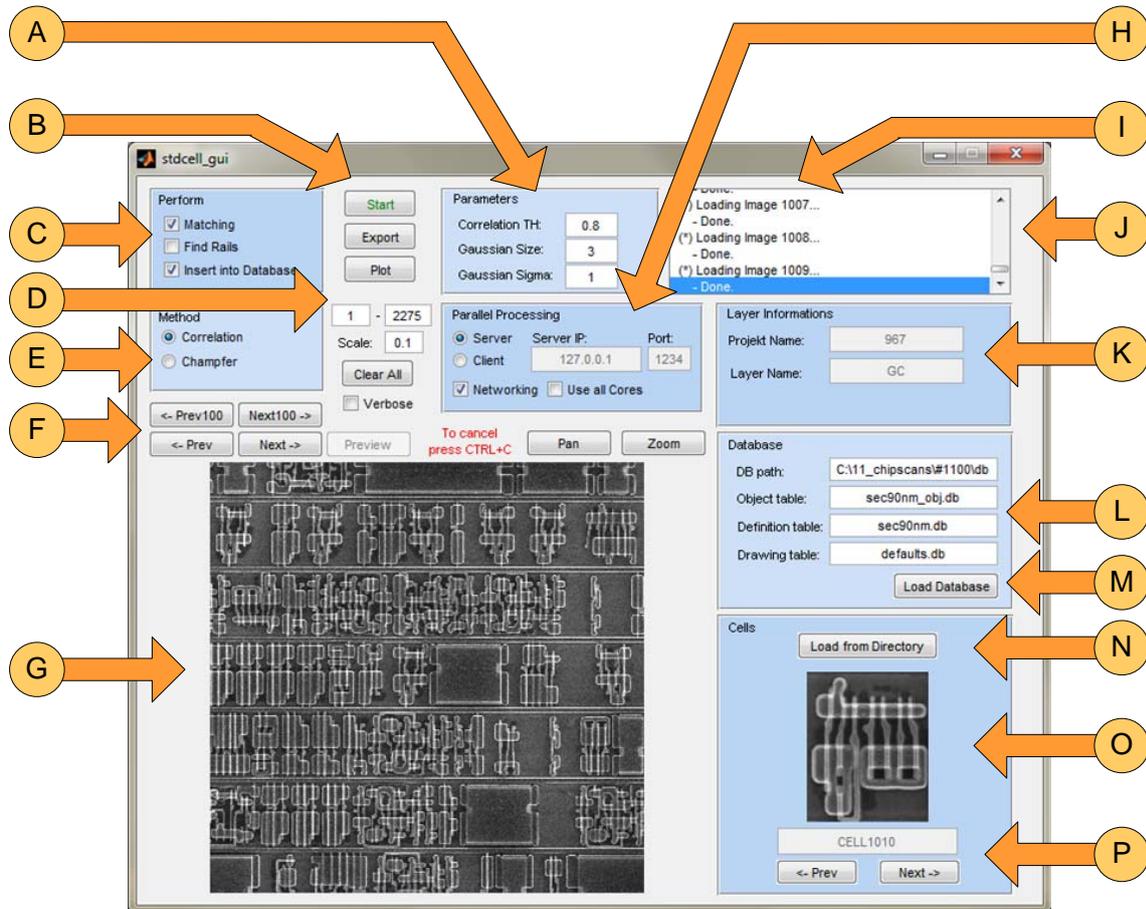


Abbildung A.6: GUI zur Zellenextraktion.

Definitionstabelle ausgelesen und die Bilder der darin enthaltenen Standardzellen in das Verzeichnis *stdcells* abgelegt und in Auswahl **O** angefügt.

Stehen die Datenbanken nicht zur Verfügung oder soll die Analyse nur experimentell durchgeführt werden, so können die Bilder auch manuell durch Button **N** geladen werden. Zu beachten ist allerdings, dass damit ein späterer Export in eine Datenbank nicht mehr möglich ist.

Die Konfiguration des Analyse-Ablaufes erfolgt durch Auswahl **C**. Hier stehen folgende Optionen zur Wahl:

- *Matching*: Durchführen der Standardzellensuche. Die zu verwendende Methode kann in Auswahl **E** gewählt werden.
- *Find Rails*: Rein experimentelle Option. Steht in Version 0.1.80 nicht zu Verfügung.

- *Insert into Database*: Einfügen der gefundenen Zellen in die durch **L** festgelegte ObjektTabelle. Soll kein Eintrag erfolgen, muss die Option abgewählt werden. Die extrahierten Daten werden damit in einer internen Datenstruktur gespeichert und können in einem späteren Schritt in die Datenbank übernommen werden. Dazu wird die Analyse mit alleiniger Wahl dieser Option gestartet.

Die zu verwendende Suchmethode wird in Auswahl **E** festgelegt. In der Version 0.1.80 hat nur die direkte Korrelationsmethode praktische Bedeutung. Eine Suche durch *chamfer matching* steht zwar zur Auswahl, wurde in den letzten Versionen aber nicht mitgezogen und führt deshalb zu Fehlern. In zukünftigen Versionen wird dem Rechnung getragen.

Das Starten der Analyse erfolgt durch Drücken des Button *Start* (**B**). Je nach Größe der Bilderserie und Anzahl der Standardzellen kann dieser Prozess viele Stunden in Anspruch nehmen.

Um den Aufwand auf mehrere Prozessoren und/oder Rechner aufzuteilen, können die Optionen in Gruppe **H** verwendet werden. Für den Einsatz auf einem Rechner unter Zuhilfenahme aller zu Verfügung stehenden Kerne muss die Option *Networking* deaktiviert und *Use all Cores* aktiviert sein.

Bei Verwendung eines Rechnerverbunds muss ein Rechner als Server fungieren (siehe Abschnitt 7.4). Die restlichen Rechner werden unter der Angabe der Server IP als Client gestartet. Dies ermöglicht ein verteiltes Rechnen, um auch größere Analysebereiche in vertretbarer Zeit abarbeiten zu können.

Literaturverzeichnis

- [Ahmed and Jafri, 2008] Ahmed, J. and Jafri, M. N. (2008). Improved phase correlation matching. In *ICISP '08: Proceedings of the 3rd international conference on Image and Signal Processing*, pages 128–135, Berlin, Heidelberg. Springer-Verlag.
- [Amdahl, 1967] Amdahl, G. M. (1967). Validity of the single processor approach to achieving large scale computing capabilities. In *AFIPS '67 (Spring): Proceedings of the April 18-20, 1967, spring joint computer conference*, pages 483–485, New York, NY, USA. ACM.
- [Appleton et al., 2005] Appleton, B., Bradley, A., and Wildermoth, M. (2005). Towards optimal image stitching for virtual microscopy. In *DICTA '05: Proceedings of the Digital Image Computing on Techniques and Applications*, page 44, Washington, DC, USA. IEEE Computer Society.
- [Argyriou and Vlachos, 2006] Argyriou, V. and Vlachos, T. (2006). A study of sub-pixel motion estimation using phase correlation. In *British Machine Vision Conference*, pages 387–397.
- [Argyriou and Vlachos, 2007] Argyriou, V. and Vlachos, T. (2007). On the estimation of subpixel motion using phase correlation. *Electron. Imaging*, 16.
- [Asada and Brady, 1986] Asada, H. and Brady, M. (1986). The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):2–14.
- [Avery and Al, 2002] Avery, C. and Al, S. (2002). Reverse engineering complex application-specific integrated circuits (asics). In *Diminishing Manufacturing Sources and Material Shortages Conference*.
- [Bailey, 2004] Bailey, D. G. (2004). An efficient euclidean distance transform. In *In Combinatorial Image Analysis, IWCI 2004*, pages 394–408.

- [Balaban, 1995] Balaban, I. J. (1995). An optimal algorithm for finding segments intersections. In *SCG '95: Proceedings of the eleventh annual symposium on Computational geometry*, pages 211–219, New York, NY, USA. ACM.
- [Ballard, 1981] Ballard, D. H. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122.
- [Barnea and Silverman, 1972] Barnea, D. I. and Silverman, H. F. (1972). A class of algorithms for fast digital image registration. *IEEE Trans. Comput.*, 21(2):179–186.
- [Beaudet, 1978] Beaudet, P. (1978). Rotationally invariant image operators. In *Proceedings of the 4th international joint conference on Pattern Recognition*, pages 579–583.
- [Bentley and Ottmann, 1979] Bentley, J. L. and Ottmann, T. A. (1979). Algorithms for reporting and counting geometric intersections. *IEEE Trans. Comput.*, 28(9):643–647.
- [Biggerstaff, 1989] Biggerstaff, T. J. (1989). Design recovery for maintenance and reuse. *IEEE Computer*, 22(7):36–49.
- [Blythe et al., 1993] Blythe, S., Fraboni, B., Lall, S., H Ahmed, H., and de Riu, U. (1993). Layout reconstruction of complex silicon chips. In *IEEE Journal of Solid-State Circuits*, volume 28.
- [Bradley et al., 2005] Bradley, A. P., Wildermoth, M., and Mills, P. (2005). Virtual microscopy with extended depth of field. In *DICTA '05: Proceedings of the Digital Image Computing on Techniques and Applications*, page 35, Washington, DC, USA. IEEE Computer Society.
- [Breu et al., 1995] Breu, H., Gil, J., Kirkpatrick, D., and Werman, M. (1995). Linear time euclidean distance transform algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:529–533.
- [Briechle and Hanebeck, 2001] Briechle, K. and Hanebeck, U. D. (2001). Template matching using fast normalized cross correlation. In *Proceedings of SPIE: Optical Pattern Recognition XII*, volume 4387, pages 95–102.
- [Brown and Gottesfeld, 1992] Brown, M. and Gottesfeld, L. (1992). A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376.
- [Brown and Lowe, 2007] Brown, M. and Lowe, D. (2007). Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision*, 74(1):59–73.
- [Brown and Lowe, 2003] Brown, M. and Lowe, D. G. (2003). Recognising panoramas. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, page 1218, Washington, DC, USA. IEEE Computer Society.

- [Büttgenbach, 1994] Büttgenbach (1994). *Mikromechanik. Einführung in Technologie und Anwendungen*. Teubner Studienbücher Physik.
- [Burt and Adelson, 1983] Burt, P. J. and Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2(4):217–236.
- [Calvin R. Maurer et al., 2003] Calvin R. Maurer, J., Qi, R., and Raghavan, V. (2003). A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:265–270.
- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.
- [Chen, 1995] Chen (1995). Quicktime vr: an image-based approach to virtual environment navigation. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 29–38, New York, NY, USA. ACM.
- [Chikofsky and Cross, 1990] Chikofsky, E. J. and Cross, J. H. (1990). Reverse engineering and design recovery: A taxonomy. *IEEE Softw.*, 7(1):13–17.
- [Chisholm et al., 1999] Chisholm, G. H., Eckmann, S. T., Lain, C. M., and Veroff, R. L. (1999). Understanding integrated circuits. *IEEE Des. Test*, 16(2):26–37.
- [Czichos and Hennecke, 2008] Czichos, H. and Hennecke, M. (2008). *Hütte - Das Ingenieurwissen*. Springer-Verlag.
- [Davis, 1998] Davis, J. (1998). Mosaics of scenes with moving objects. In *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 354, Washington, DC, USA. IEEE Computer Society.
- [Doom et al., 1997] Doom, T., White, J., Wojcik, A., and Chisholm, G. (1997). Identification of functional components in combinational circuits. Technical report, Division of Information Science, Argonne National Laboratory.
- [Doom et al., 1998] Doom, T., White, J., Wojcik, A., and Chisholm, G. (1998). Identifying high-level components in combinational circuits. In *In Proceedings of the 1998 Great Lakes Symposium on VLSI*, pages 313–318.
- [Doudkin et al., 2005] Doudkin, A., Inyutin, A., and Vatkin, M. (2005). Objects identification on the color layout images of the integrated circuit layers. In *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pages 69–72.

- [Doudkin et al., 2001] Doudkin, A., Sadykhov, R., Selikhanovich, A., and Vershok, D. (2001). Contour extraction algorithms for lsi circuit video image processing. In *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pages 69–72.
- [Doudkin and Vershok, 2003] Doudkin, A. and Vershok, D. (2003). Computer-aided inspection of some design rules of integrated circuit layers. volume 2, pages 232–235. IEEE Computer Society.
- [Drouin et al., 2007] Drouin, D., Couture, A., Joly, D., Tastet, X., Aimez, V., and Gauvin, R. (2007). Casino v2.42: a fast and easy-to-use modeling tool for scanning electron microscopy and microanalysis users. *Scanning*, 29(3):92–101.
- [Duda and Hart, 1972] Duda, R. O. and Hart, P. E. (1972). Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15.
- [Feng et al., 2005] Feng, H., Ye, J., and Pease, R. (2005). Reconstruction of pattern images from scanning electron microscope images. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 23 of *J. Vac. Sci. Technol.*
- [Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395.
- [Forsyth and Ponce, 2003] Forsyth, D. and Ponce, J. (2003). *Computer Vision: A Modern Approach*. Prentice Hall.
- [Friedman, 2001] Friedman, E. G. (2001). Clock distribution networks in synchronous digital integrated circuits. In *Proc. IEEE*, pages 665–692.
- [Frischholz and Spinnler, 1993] Frischholz, R. and Spinnler, K. (1993). Class of algorithms for realtime subpixel registration. In *Proceedings of SPIE, Vol*, pages 50–59.
- [Förstner and Gülch, 1987] Förstner, W. and Gülch, E. (1987). A fast operator for detection and precise location of distinct points, corners and centres of circular features. *Conference*, pages 281–305.
- [Gajski et al., 1992] Gajski, D. D., Dutt, N. D., Wu, A. C., and Lin, S. Y. (1992). *High-level synthesis: introduction to chip and system design*. Kluwer Academic Publishers, Norwell, MA, USA.
- [Gleason et al., 1990] Gleason, Hunt, and Jatko (1990). Subpixel measurement of image features based on paraboloid surface fit. In *Proc. Machine Vision Systems Integration in Industry, SPIE*, pages 135–144.

- [Gonzalez and Woods, 2008] Gonzalez, R. C. and Woods, R. E. (2008). *Digital image processing*. Prentice Hall, Upper Saddle River, N.J.
- [Guizar et al., 2008] Guizar, M., Thurman, S., and Fienup, J. (2008). Efficient subpixel image registration algorithms. *Opt. Letters*, 33:156–158.
- [Harris and Stephens, 1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151.
- [Hartley and Zisserman, 2000] Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049.
- [Hough, 1962] Hough, P. (1962). Method and means for recognizing complex patterns. In *US Patent*.
- [Jang and Suk, 2008] Jang, H. L. and Suk, I. Y. (2008). An effective image segmentation technique for the sem image. In *IEEE International Conference on Industrial Technology*.
- [Kanaya and Okayama, 1972] Kanaya, K. and Okayama, S. (1972). Penetration and energy-loss theory of electrons in solid targets. *J. Phys.*, D5:43–58.
- [Kitchen and Rosenfeld, 1982] Kitchen, L. and Rosenfeld, A. (1982). Gray level corner detection. *PRL*, 1(2):95–102.
- [Kovesi, 2002] Kovesi, P. D. (2002). Edges are not just steps. In *Proceedings of the Fifth Asian Conference on Computer Vision*, pages 822–827. Melbourne.
- [Kruskal, 1956] Kruskal, J. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–50.
- [Lagunovsky et al., 1998] Lagunovsky, D., Ablameyko, S., and Kutas, M. (1998). Recognition of integrated circuit images in reverse engineering. *International Conference on Pattern Recognition*, 2:1640.
- [Lan and Mohr, 1998] Lan, Z. and Mohr, R. (1998). Direct linear sub-pixel correlation by incorporation of neighbor pixels' information and robust estimation of window transformation. *Mach. Vis. Appl.*, 10(5-6):256–268.
- [Langumovsky et al., 1996] Langumovsky, D., Ablameyko, S., and Kutas, M. (1996). Extraction of topological features of integrated circuit from grey-scale image. In *IAPR Workshop on Machine Vision Applications*, pages 271–274.
- [Lewis, 1995] Lewis, J. (1995). Fast normalized cross-correlation.

- [Madisetti et al., 1999] Madisetti, V. K., Yong-Kyu, J., Moinul, H. K., Jeongwook, K., and Theodore, F. (1999). Reengineering legacy embedded systems. *IEEE Design and Test of Computers*, 16:38–47.
- [Marr and Hildreth, 1980] Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 207(1167):187–217.
- [Masalskis and Navickas, 2008] Masalskis, G. and Navickas, R. (2008). Reverse engineering of cmos integrated circuits. In *Electronics and electrical Engineering*, volume 88.
- [Moore, 1965] Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8):114117.
- [Morevec, 1977] Morevec, H. P. (1977). Towards automatic visual obstacle avoidance. In *IJCAI'77: Proceedings of the 5th international joint conference on Artificial intelligence*, pages 584–584, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- [Morrone and Owens, 1987] Morrone and Owens, R. A. (1987). Feature detection from local energy. *Pattern Recogn. Lett.*, 6(5):303–313.
- [Morrone et al., 1986] Morrone, M., Ross, J., Burr, D., and Owens, R. (1986). Mach bands are phase dependent. *Nature*, 324(2):250–253.
- [Morrone and Burr, 1988] Morrone, M. C. and Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Royal society of London*, 235(1280):221–245.
- [Neubeck and Van Gool, 2006] Neubeck, A. and Van Gool, L. (2006). Efficient non-maximum suppression. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 850–855, Washington, DC, USA.
- [Nii et al., 2006] Nii, Sanuki, and Okayama (2006). A 45nm high performance bulk logic platform technology (cmos6) using ultra high na(1.07) immersion lithography with hybrid dual-damascene structure and porous low-k beol. In *Electron Devices Meeting, 2006. IEDM '06. International*.
- [Oppenheim et al., 1999] Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (1999). *Discrete-time signal processing (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- [Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66.
- [Pease et al., 2006] Pease, R., Feng, H., and Ye, J. (2006). Segmentation-assisted edge extraction algorithms for sem images. In *Society of Photo-Optical Instrumentation*

Engineers (SPIE) Conference Series, volume 6349 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*.

- [Perko, 2004] Perko, R. (2004). *Computer Vision for Large Format Digital Aerial Cameras*. PhD thesis, Graz, University of Technology, Austria.
- [Reimer, 2000] Reimer, L. (2000). Scanning electron microscopy: Physics of image formation and microanalysis, second edition. *Measurement Science and Technology*, 11(12):1826.
- [Rekoff, 1985] Rekoff, M. (1985). On reverse engineering. *IEEE Transactions on Systems, Man and Cybernetics*, 3/4:244–252.
- [Romero et al., 2007] Romero, E., Gómez, F., and Iregui, M. (2007). Virtual microscopy in medical images: a survey. In *Modern Research and Educational Topics in Microscopy*. Badajoz, Spain.
- [Rubanov, 2003] Rubanov, N. (2003). Subislands: the probabilistic match assignment algorithm for subcircuit recognition. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 22(1):26–38.
- [Rubanov, 2005] Rubanov, N. (2005). An efficient subcircuit recognition using the nonlinear graph matching. In *SBCCI '05: Proceedings of the 18th annual symposium on Integrated circuits and system design*, pages 44–49, New York, NY, USA. ACM.
- [Rubanov, 2006] Rubanov, N. (2006). A high-performance subcircuit recognition method based on the nonlinear graph optimization. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 25(11):2353–2363.
- [Schmid et al., 2000] Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *Int. J. Comput. Vision*, 37(2):151–172.
- [Shamos and Hoey, 1976] Shamos, M. I. and Hoey, D. (1976). Geometric intersection problems. In *SFCS '76: Proceedings of the 17th Annual Symposium on Foundations of Computer Science*, pages 208–215, Washington, DC, USA. IEEE Computer Society.
- [Shekarforoush et al., 2002] Shekarforoush, Zerubia, J., and Berthod, M. (2002). Extension of phase correlation to subpixel registration. *IEEE Trans. on Image Processing*, 11(3):188–200.
- [Sobel and Feldman, 1968] Sobel, I. and Feldman, G. (1968). A 3x3 isotropic gradient operator for image processing. Never published but presented at a talk at the Stanford Artificial Project.
- [Sobh and Patel, 2007] Sobh, K. E. and Patel, S. (2007). Reverse engineering of vlsi chips: A roadmap. *Journal of Engineering and Applied Sciences*, 2(2):290–298.

- [Soden et al., 1997] Soden, J. M., Anderson, R. E., and Henderson, C. L. (1997). Ic failure analysis: Magic, mystery, and science. *IEEE Des. Test*, 14(3):59–69.
- [Sonka et al., 2007] Sonka, M., Hlavac, V., and Boyle, R. (2007). *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering.
- [Stone et al., 2003] Stone, H., Tao, B., and McGuire, M. (2003). Analysis of image registration noise due to rotationally dependent aliasing. *Journal of Visual Communication and Image Representation*, 14(2):114–135.
- [Sun et al., 2005] Sun, C., Beare, R., Hilsenstein, V., and Jackway, P. (2005). Mosaicing of microscope images. In *DICTA '05: Proceedings of the Digital Image Computing on Techniques and Applications*, page 50, Washington, DC, USA. IEEE Computer Society.
- [Szeliski, 1994] Szeliski, R. (1994). Image mosaicing for tele-reality applications. CRL-Report CRL-94-02, Cambridge Research Laboratory.
- [Szeliski, 2006] Szeliski, R. (2006). Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104.
- [Takita et al., 2003] Takita, K., Aoki, T., Yoshifumi, S., Huguchi, T., and Kobayashi, K. (2003). High-accuracy subpixel image registration based on phase-only correlation. *IEICE Trans. Fundamentals*, 86.
- [Thévenaz et al., 2006] Thévenaz, P., Lambiel, D., and Unser, M. (2006). A strategy based on maximum spanning trees to stitch together microscope images. *SPIE the international society for optical engineering*, 7(2).
- [Thévenaz and Unser, 2007] Thévenaz, P. and Unser, M. (2007). User-friendly semiautomated assembly of accurate image mosaics in microscopy. *Microscopy Research and Technique*, 70(2):135–146.
- [Thibault et al., 2008] Thibault, P., Dierolf, M., Menzel, A., Bunk, O., David, C., and Pfeiffer, F. (2008). High-Resolution Scanning X-ray Diffraction Microscopy. *Science*, 321:379–382.
- [Tian and Huhns, 1986] Tian, Q. and Huhns, M. (1986). Algorithms for subpixel registration. *Comput. Vision Graph. Image Process.*, 35(2):220–233.
- [Tomasi and Kanade, 1991] Tomasi, C. and Kanade, T. (1991). Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University.
- [Tomasi and Manduchi, 1998] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. pages 839–846.

- [Torrance and Dick, 2007] Torrance, R. and Dick, J. (2007). Reverse engineering in the semiconductor industry. In *IEEE'07: Custom Integrated Circuits Conference*, pages 363–381, Berlin, Heidelberg. Springer-Verlag.
- [Torrance and James, 2009] Torrance, R. and James, D. (2009). The state-of-the-art in ic reverse engineering. In *CHES '09: Proceedings of the 11th International Workshop on Cryptographic Hardware and Embedded Systems*, pages 363–381, Berlin, Heidelberg. Springer-Verlag.
- [Tsai and Lin, 2003] Tsai, D. and Lin, C. (2003). Fast normalized cross correlation for defect detection. *Pattern Recogn. Lett.*, 24(15):2625–2631.
- [Vershok, 2001] Vershok, D. (2001). The algorithm of segmentation of grayscale images. In *Proc. of the 2nd int. Conference on Neural Networks and Artificial Intelligence*, pages 143–146.
- [Wang and Brady, 1991] Wang, H. and Brady, J. (1991). Corner detection for 3d vision using array processors. *BARNAIMAGE91*. Barcelona, Spain.
- [Zebedin et al., 2006] Zebedin, L., Klaus, A., Gruber Geymayer, B., and Karner, K. (2006). Towards 3d map generation from digital aerial images. *PandRS*, 60(6):413–427.
- [Zitova and Flusser, 2003] Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.