



Holger Zoglauer

Entwicklung und Testergebnisse eines Prototypensystems
für die Wafer-Scale-Integration

Diplomarbeit

HD-KIP-09-28

Fakultät für Physik und Astronomie
Ruprecht-Karls-Universität Heidelberg

Diplomarbeit
im Studiengang Physik

vorgelegt von
Holger Zoglauer
geboren in Gera

April 2009

Entwicklung und Testergebnisse eines Prototypensystems für die Wafer-Scale-Integration

Die Diplomarbeit wurde ausgeführt von Holger Zoglauer am
Kirchhoff-Institut für Physik
Universität Heidelberg
unter der Betreuung von
Prof. Dr. Karlheinz Meier

Entwicklung eines Prototypensystems für die Wafer-Scale-Integration

Die vorliegende Arbeit beschreibt den Aufbau eines Prototypensystems zur Integration von analogen Schaltungen auf Waferebene. Das fertige System wird neuronale Netzwerkchips in hoher Stückzahl und hoher Dichte miteinander verbinden. Erste Anforderungen sind die Erzeugung von Verbindungen zwischen Retikeln auf einem Wafer, Verbindungen vom Wafer weg zu Peripheriegeräten, sowie der mechanische Aufbau inklusive der räumlichen Ausrichtung der einzelnen Elemente zueinander. Die notwendigen Verbindungen zwischen benachbarten Retikeln werden per Postprozessierung mit einem Punkt-zu-Punkt-Abstand von $9,7\ \mu\text{m}$ hergestellt. Die vertikalen Verbindungen, welche vom Wafer weg führen, werden nach einer selbst entwickelten Idee über Elastomerverbinder hergestellt. Dafür werden auf einem Wafer von 20 cm Durchmesser etwa 350 Elastomere verwendet, die insgesamt circa 17500 Verbindungen erzeugen. Die Realisierung dieser Idee wurde getestet und Messergebnisse werden vorgestellt. Ein entscheidender Teil des entwickelten Systems ist die Justage der einzelnen Komponenten. Mit einer selbst entwickelten und von der institutseigenen Werkstatt gefertigten Justiervorrichtung können mit Hilfe einer eigens entworfenen optischen Justagemethode Genauigkeiten von $50\ \mu\text{m}$ in x- und y-Richtung, sowie $(3 \cdot 10^{-3})^\circ$ gegen eine Verdrehung erreicht werden.

Development of a wafer scale integration prototype

The following thesis describes the setup of a prototype for the wafer-scale integration of analog circuits. The final system will incorporate a high number of neural network chips without separating a wafer in single dies. First requirements are the connection of reticles on the wafer, the connection between the wafer and external electrical components, an adequate mechanical setup and the alignment of the components against each other. With post-processing, adjacent reticles are connected on the wafer with a line pitch of $9.7\ \mu\text{m}$. The vertical connectivity between the wafer to a controller board is implemented with elastomeric connectors. On a wafer with a diameter of 20 cm, about 350 elastomers realise approximately 17500 connections. The signal transmission performance with these elastomeric connectors was tested and test results are presented. An essential part of the wafer scale integration is the alignment of the different components. A setup and an optical method to adjust the components was specially developed and built by the mechanic workshop of the institute. With the setup it is possible to align the components with an accuracy of $50\ \mu\text{m}$ in x- and y-direction and to adjust the contortion within $(3 \cdot 10^{-3})^\circ$.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Das FACETS-Projekt	2
1.3	Übersicht	3
2	Testsystem - Überblick	5
2.1	Wafertträger	6
2.2	Dichtungsringe	6
2.3	Wafer	7
2.4	Elastomerverbinder	7
2.5	Positionierungsmaske	8
2.6	Hauptplatine	10
2.7	Deckel	10
3	Eigenschaften des Testwafer und der Postprozessierung	13
3.1	Short Loop Wafer	13
3.2	Postprozessierung des Wafers	13
3.2.1	Postprozessierung: Verfahren	14
3.2.2	Postprozessierung: Strukturen	16
3.2.3	Postprozessierung: Ergebnisse	18
3.2.3.1	Struktur 1	18
3.2.3.2	Struktur 2	20
3.2.3.3	Struktur 3	21
3.2.3.4	Struktur 4	22
3.2.4	Postprozessierung: Bewertung	23
4	Prinzip der Justage der einzelnen Komponenten	27
4.1	Justiervorrichtung	28
4.2	Optische Justage	29
4.3	Elektrische Justage	30
4.4	Einstellung des Abstandes zwischen Wafer und Platine	32
4.5	Zusammenfassung	33
5	Aufbau der Hauptplatine	35
5.1	Anforderungen an die Testplatine	35
5.2	Realisierbarkeit der Platine	36
5.3	Design der Testplatine	37

6 Messungen der Elastomere	41
6.1 Technische Daten	41
6.2 Messmethode	42
6.3 Zuverlässigkeit	43
6.4 Elektrisches Verhalten	48
6.5 Diskussion	50
7 Zusammenfassung und Ausblick	53
Literaturverzeichnis	55
Danksagungen	58

Kapitel 1

Einleitung

1.1 Motivation

Die Natur dient dem Menschen schon immer als Vorbild für neue Entwicklungen. Es ist schwierig alltägliche Gegenstände aus unserem Leben zu finden, welche kein Gegenstück in der Natur haben. Diese Erkenntnis machen sich viele Wissenschaftler und Ingenieure zu Nutze. So versucht man schon seit langem einen Faden herzustellen, welcher ähnliche Elastizitätsmodule und spezifische Gewichte wie die von vielen Spinnen erzeugten Spinnfäden hat. Auch die Haut eines Haifisches ist ein Beispiel, in dem der Mensch versucht von der Natur entwickelte Konzepte zu seinem Vorteil zu nutzen.

Die nachfolgende Arbeit ist Teil eines Projektes, welches das menschliche Gehirn als Vorbild hat. Das Gehirn zu verstehen ist von großem Interesse. Es kann dazu beitragen neurologische und psychiatrische Erkrankungen zu verstehen und zu heilen. Es könnte die Anbindung von technischen Geräten an das menschliche Nervensystem ermöglichen, bzw. weiterentwickeln. Das menschliche Gehirn ist eine sehr leistungsstarke und zugleich effiziente Rechenmaschine, mit positiven Eigenschaften wie beispielsweise Lernfähigkeit und Fehlertoleranz. Aufgrund der hohen Komplexität des menschlichen Gehirns wird es jedoch auch in den nächsten Jahrzehnten nicht möglich sein seine Funktionen in vollem Umfang zu verstehen. Es ist allerdings realistisch sich einzelne Stärken zu Nutze zu machen. So kann man beispielsweise versuchen das Prinzip der Datenverarbeitung im Gehirn nachzuahmen. Um dieses Ziel zu erreichen, müssen weitere und tiefergehende Erkenntnisse gesammelt werden. Man muss zunächst einmal das Prinzip, welches hinter der Informationsverarbeitung im Gehirn steckt, erkennen. Hierzu ist die Vermessung von einzelnen Neuronen und deren Verbindung zu anderen sowie von großen Netzwerken von Neuronen, also großen Arealen des Gehirns ein wichtiger Punkt. Es sind schon einige Erkenntnisse aus solchen Messungen bekannt. So weiß man, dass es verschiedenen Arten von Neuronen gibt, dass das Gehirn in unterschiedliche Bereiche eingeteilt ist, welche für bestimmte Aufgaben verantwortlich sind und dass im Gehirn um 10^{11} Neuronen und 10^{14} Synapsen vorhanden sind. Im weiteren benötigt man Modelle von Neuronen und Netzwerken welche das gemessene Verhalten erklären können. Anschließend kann man versuchen diese Modelle zu überprüfen. Dies kann man mit herkömmlichen Computern machen, solange die Modelle auf Gleichungen basieren, welche numerisch lösbar sind. Allerdings stößt man hier sehr schnell an Grenzen. Das Gehirn verarbeitet seine Daten parallel in mehreren Neuronen. Ein Computer verarbeitet seine Informationen seriell. Als Folge müssen viele Werte zwischengespeichert werden. Selbst bei Verwendung von Cluster Computing bleibt die Aufgabe ab gewissen Netzwerkgrößen sehr zeitaufwendig. Unterm Strich ist die Simulation von neuronalen Netzwerken auf digitalen Computern mit einem hohen Maß an Rechenkapazität, Speicherkapazität, Zeit, Energiebedarf und als Folge mit hohen Kosten

verbunden. Alternativ dazu und weiter dem Gehirn als Vorbild folgend, können neuronale Netze auch in Hardware abgebildet werden. Dabei werden verschiedene Neuronenmodelle mit Hilfe analoger Schaltkreise nachgebildet und die Kommunikation zwischen den Neuronen digitalisiert. Ein Vorteil dieser Methode ist außerdem, dass die resultierende Schaltung mit erhöhter Geschwindigkeit im Vergleich zum biologischen Vorbild arbeitet und somit die zeitliche Entwicklung eines Netzes in kürzeren Zeiten beobachten kann. Diese Tatsache ist eine sehr nützliche Eigenschaft, da es bei der Verwendung von neuronalen Netzen viele einstellbare Parameter gibt. Um große Bereiche dieser Parameter abzudecken, ist es erforderlich die Systeme lange bzw. oft durchlaufen zu lassen.

Neuronale Netze zu verstehen und zu simulieren bzw. emulieren ist eine sehr komplexe Aufgabe und benötigt ein hohes Maß an Kompetenz und Zusammenarbeit vieler verschiedener Fachrichtungen. Das FACETS¹-Projekt versucht diese Zusammenarbeit herzustellen.

1.2 Das FACETS-Projekt

Das FACETS-Projekt ist ein von der EU finanziertes Forschungsprojekt, das im Bereich neuronaler Wissenschaft anzusiedeln ist. Es wurde für einen Forschungszeitraum von vier Jahren geplant. 13 Universitäten aus ganz Europa sind Teil dieses Projektes. Sie sind in 10 Arbeitsgruppen eingeteilt. Sechs dieser Gruppen bearbeiten wissenschaftliche Kernthemen. Die anderen vier führen die gewonnenen Erkenntnisse zusammen und erledigen administrative Aufgaben. Die sechs Kernthemen umfassen biologische Experimente auf zellulärer und Netzwerkebene, die Modellierung von einzelnen Zellen sowie von Netzwerken von Neuronen und Synapsen und dem Bau von neuronaler Hardware. Auch beim Bau von neuronaler Hardware gibt es sowohl das Ziel ein Abbild einzelner Neuronen vom biologischen Vorbild möglichst genau nachzubauen, als auch den Versuch möglichst große Netze von Neuronen und Synapsen zu implementieren.

Die Universität Heidelberg ist neben der Projektleitung auch mit forschendem Engagement ein Teil vom FACETS-Projekt. Die Electronic Vision(s) Gruppe der Universität Heidelberg arbeitet zusammen mit der technischen Universität Dresden an der Implementierung von großskaligen neuronalen Netzwerken in Hardware. Der Aufgabenbereich der Universität Heidelberg ist dabei in zwei Abschnitte Stage 1 und Stage 2 unterteilt worden. In Stage 1 wurde ein Chip (Spikey²) mit analogen und digitalen Bauteilen hergestellt, auf welchem 384 Neuronen und (384*256) Synapsen realisiert sind. Auf dem Spikey wird durch einen analogen Schaltungsentwurf ein Neuronenmodell emuliert. Die Kommunikation zwischen Neuronen im Gehirn findet nach derzeitigen Forschungsstand digital statt. Deshalb werden Informationen zwischen Neuronen auch im Spikey digital umgesetzt. Neuronale Netze laufen auf dem Spikey bis zu 100 000mal schneller als ihr biologisches Gegenstück. Der Spikey-Chip wurde so konzipiert, dass er mit anderen Spikeys zusammengeschlossen werden kann und man somit größere Netzwerke erhält. Aufgrund der großen Anzahl an Verbindungen, welche man mit solchen Netzen erreichen will, ist die Bandbreite beim Zusammenschluss von mehreren Spikeys ein stark limitierender Faktor.

Um dem entgegen zu wirken wird in Stage 2 ein weiterer Chip (HICANN³) entwickelt. Dieser Chip realisiert 512 Neuronen und 128k Synapsen ((8), (9), (3)). Beim Entwurf des HICANN-Chips fließen Erkenntnisse und Erfahrungen ein, welche mit dem Spikey gesammelt wurden. Zusätzlich wird bei der Entwicklung darauf geachtet, dass man den HICANN

¹Fast Analog Computing with Emergent Transient States

²the spike - der Zacken. Die Kommunikation zwischen Neuronen ist pulsorientiert. D. h. ein Neuron kann durch verschiedene Eingangssignale eine gewisse Schwelle erreichen und sendet daraufhin einen Ausgangspuls. Weitere Informationen sind beispielsweise in (4) zu finden.

³High Input Count Analog Neural Network

auf einem Wafer direkt mit anderen HICANNs zusammen schalten kann. Ziel des ganzen ist der Zusammenschluss vieler HICANNs auf einem Wafer. Dabei erreicht man eine hohe Dichte an Neuronen und Synapsen mit nur noch kurzen digitalen Verbindungen, wodurch sich die Bandbreite für die Kommunikation stark erhöht.

Es gibt verschiedene Prozesse zur Herstellung eines Wafers. Der HICANN-Chip wird auf einem Wafer hergestellt, auf dem Strukturauflösungen bis 180 nm möglich sind. Die Waferfläche ist in Rechtecke eingeteilt⁴, welche man als Retikel bezeichnet. Feine Strukturen bis 180 nm können nur innerhalb eines Retikels erzeugt werden. Um Retikel miteinander zu verbinden muss man auf andere Verfahren zurückgreifen, welche keine so hohe Auflösung mehr umsetzen können. Auf einem Retikel kann man acht HICANN-Chips unterbringen. Die HICANNs werden sowohl untereinander auf dem Wafer, als auch mit Peripheriegeräten nach außen verbunden. Die Verbindung von HICANNs eines Retikels ist nicht sehr problematisch, da hierfür sehr kleine Strukturgrößen verwendet werden können. Um die Konnektivität zwischen HICANNs verschiedener Retikel zu herzustellen, muss man 2048 Kontakte zwischen benachbarten Retikeln verlegen. Bei einer Kantenlänge von 20 mm müssen die Leitungen mit einem Mittelpunktsabstand von max 9,7 μm ⁵ angeordnet werden. Um die Spannungsversorgung und Kommunikation mit den HICANNs nach außen herzustellen, müssen pro HICANN etwa 50 Signale vom Wafer weg geführt werden. Normalerweise erreicht man dies bei einzelnen Chips durch Bonden. Auf dem Wafer sollen circa 350 HICANNs Platz finden. Daraus folgen in der Größenordnung um 17500 Kontakte. Eine so große Anzahl an Verbindungen mittels Bonden fehlerfrei und dauerhaft herzustellen ist kaum möglich. Daher wurde eine andere Kontaktmethode konzipiert, die Elastomerverbinder⁶ verwendet. Diese Elastomere haben dabei eine Länge von 18 mm und eine Querschnittsfläche von 1 mm auf 1 mm. Es werden insgesamt circa 350 Elastomere benötigt um die 17500 Verbindungen herstellen zu können.

Im fertigen System soll der Wafer in einen Aluträger eingelassen werden, welcher sowohl zum Abtransport der Abwärme als auch zur mechanischen Fixierung des Wafers dienen soll. Darüber wird eine Platine gelegt, welche die Signale vom Wafer über die Elastomere entgegen nimmt und weiterverarbeitet. Auch die Spannungsversorgung wird auf der Platine verteilt. Über der Platine befindet sich eine weitere Aluplatte, welche zusammen mit dem Aluträger das System zusammen hält. In Abbildung 1.1 ist ein Bild des geplanten Systems zu sehen. In dem Bild sind auf der Hauptplatine noch weitere kleinere Platinen zu sehen. Auf diesen Platinen befinden sich DNCs⁷ und FPGAs. Dabei steuert ein FPGA vier DNCs an und ein DNC acht HICANNs. Man erhält daraus 12 FPGA-Platinen. Der Abtransport der von den DNCs und FPGAs erzeugten Wärme wird über Heatpipes an einen Kühlkörper auf der Oberseite gewährleistet. Neben den 12 FPGA-Platinen kann man in Bild 1.1 zwei weitere Aufsteck-Platinen sehen. Diese sind für die Erzeugung und Verwaltung der verschiedenen benötigten Versorgungsspannungen notwendig.

1.3 Übersicht

In dieser Arbeit soll ein Testsystem vorgestellt werden, welches benutzt wird, um die Eigenschaften und die Zuverlässigkeit der einzelnen Komponenten zu testen. Zusätzlich wird

⁴Die Ursache hierfür wird in Kapitel 3 erläutert

⁵Die Retikel werden durch Postprozessierung miteinander verbunden. Dieses Verfahren wird in Kapitel 3 erklärt.

⁶Funktionsweise und technische Daten stehen in Kapitel 6

⁷Digital Network Chip. Der Chip ist ein wichtiges Verbindungsstück in der Kommunikation zu den HICANNs. Er wurde zusammen mit der angeschlossenen FPGA-Platine von der TU Dresden entwickelt. Für weitere Information siehe (11), (12) und (13).

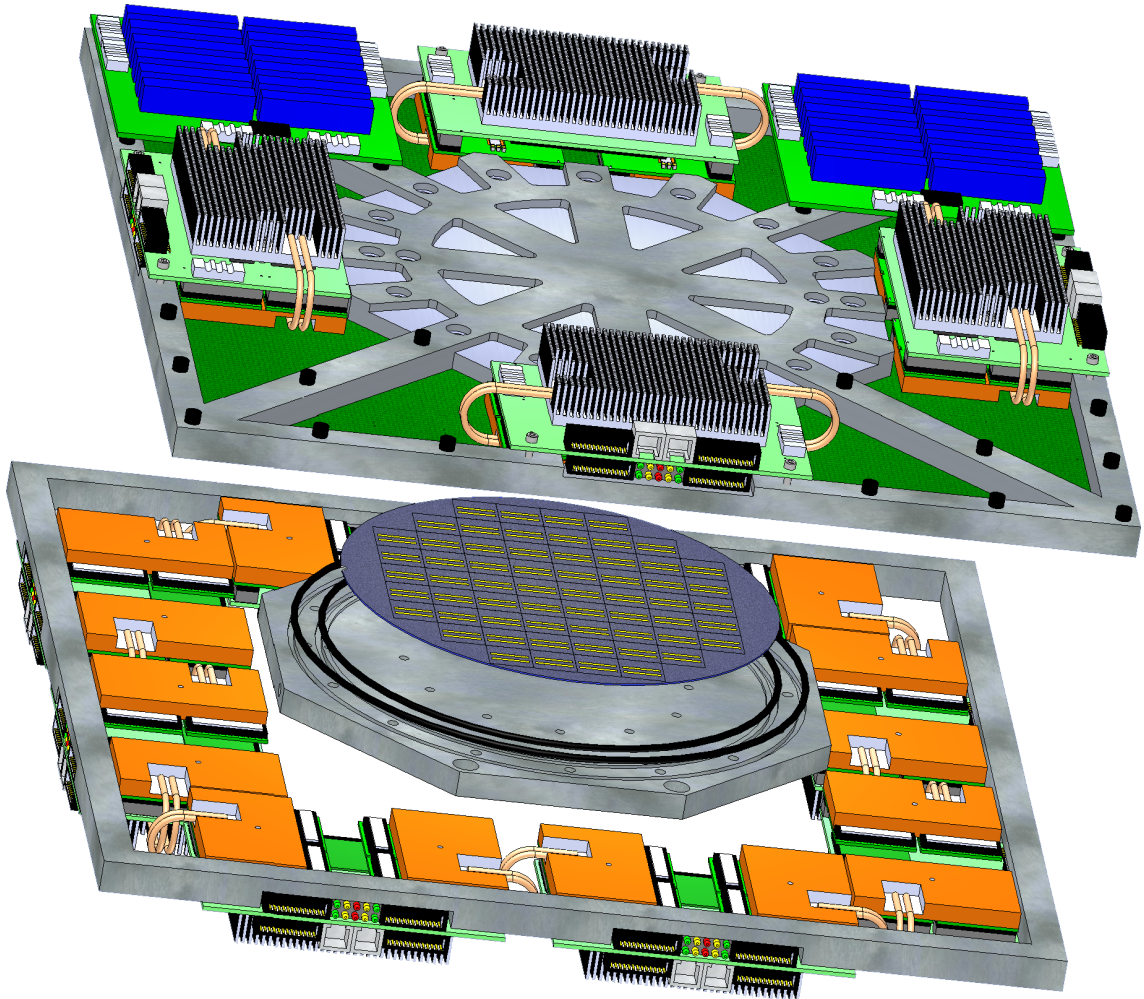


Abbildung 1.1: Das Bild zeigt den Aufbau des fertigen Wafer-Scale-Integrationsystems, wie er zur Zeit geplant ist. Die Zeichnung wurde mit Solidworks2008 erstellt. Vielen Dank an Dan Husmann de Oliveira für diese Darstellung. Dieses und weitere Bilder sind auch auf der FACETS-Homepage unter www.facets-project.org zu finden.

die verwendete Mechanik vorgestellt und eine Justage entwickelt, welche man auch für das finale System benötigt.

Im zweiten Kapitel wird der Aufbau des Testsystem aufgezeigt. Dieser entspricht prinzipiell dem Aufbau aus Abbildung 1.1. Das dritte Kapitel beschreibt die Eigenschaften des Testwafers und der Postprozessierung. Kapitel 4 stellt die Mechanik für die Justage, sowie zwei verschiedene Justage-Methoden vor. Der Entwurf und die technische Grenzen bei der Herstellung der Hauptplatine werden in Kapitel 5 vorgestellt. Vor der Zusammenfassung der gewonnen Erkenntnisse werden in Kapitel 6 die vertikalen Verbindungen thematisiert.

Kapitel 2

Testsystem - Überblick

Im folgenden Kapitel wird ein Testsystem für das Wafer-Scale-Integrationssystem vorgestellt. Es werden die Grundfunktionen der einzelnen Bauteile erläutert. Abbildung 2.1 zeigt eine Explosionszeichnung des Gesamtsystems.

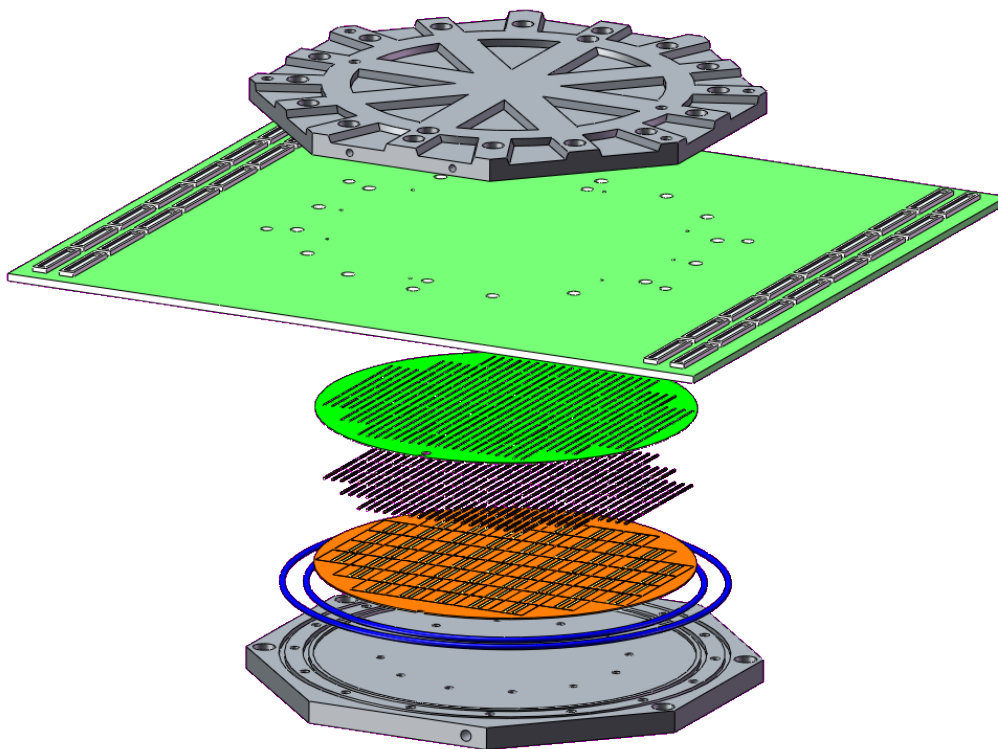


Abbildung 2.1: Mit Solidworks erstellte Explosionszeichnung des Gesamtsystems.

In der Abbildung 2.1 kann man von oben nach unten sehen:

- **Deckel**, der mit dem Waferträger das System zusammen hält.
- **Hauptplatine** mit 33 Steckern zum Anschluss der Testelektronik.
- **Positionierungsmaske** zur Ausrichtung der Elastomerverbinder.
- **Elastomerverbinder**, bis zu 480 Stück.
- **Wafer** mit Postprozessierungsstrukturen.

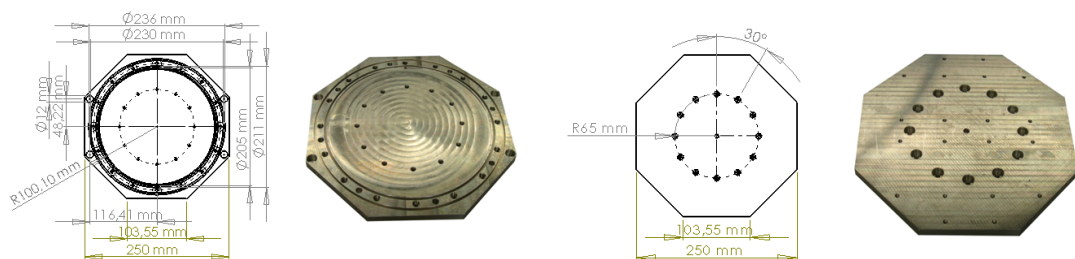
- **Dichtungsringe**, welche in den Waferträger eingelassen werden.
- **Waferträger**, in dem der Testwafer liegt.

Die einzelnen Teile werden im Folgenden in z-Richtung von unten nach oben, also vom Waferträger bis zum Deckel, vorgestellt.

2.1 Waferträger

Der Waferträger besteht aus einer 12 mm dicken Aluminiumplatte. Er hat eine achteckige Form, bei der gegenüberliegende Kanten einen Abstand von 250 mm haben. In der Mitte ist eine kreisförmige Vertiefung mit einem Durchmesser von 200,2 mm eingelassen. In diese Vertiefung wird der Testwafer eingelegt. Zur groben Justage gegen Verdrehung ist am Rand ein Stift vorhanden, der nach Definition den Süden in der xy-Ebene markiert. Die Vertiefung ist 0,7 mm tief.

Außerhalb der Vertiefung liegen zwei ebenfalls kreisförmige Nute. Ihre Durchmesser be-



(a) Foto in Draufsicht und nebenstehend Solidworkszeichnung mit Bemaßung

(b) Foto von Unterseite und nebenstehend Solidworkszeichnung mit Bemaßung

Abbildung 2.2: Draufsicht und Sicht auf Unterseite vom Waferträger. Beschreibung der Ausfräsungen, Nuten und Löcher stehen im Text.

tragen 205 mm bzw. 230 mm. Sie haben eine Breite von jeweils 3 mm und sind 2,5 mm tief. Sie dienen als Führung für die Dichtungsringe.

Zwischen den zwei Nuten liegen 16 Bohrungen inklusive M6-Gewindeinsätze. Diese Löcher gehen durch das gesamte System bis zum Deckel. In diese Löcher werden Schrauben eingesetzt, welche das System zusammen pressen.

In der Vertiefung ist ebenfalls ein Ring mit Bohrlöchern. Diese haben ein M6-Gewinde. Sie werden benötigt um den Wafer bei Bedarf wieder aus dem Waferträger zu bekommen. Dies spielt insbesondere dann eine Rolle, wenn sich eine Wärmeleitpaste zwischen Waferträger und Wafer befindet. Um die dann vorhandenen Adhäsionskräfte zu überwinden kann über diese Löcher mit Hilfe eines extra angefertigten Stempels ein gleichmäßig verteilter Druck auf den Wafer ausgeübt werden.

Auf der Unterseite befinden sich noch weitere Gewindebohrungen zur optionalen Befestigung von Kühlsystemen. Es gibt noch andere Bohrungen im Waferträger, welche für die Justage des Gesamtsystems benötigt werden und auf die in Kapitel 4 näher eingegangen wird.

2.2 Dichtungsringe

Da die Vertiefung tiefer ist als der Wafer dick (Waferdicke ca. 0,77 mm), entsteht beim Zusammenschrauben ein Hohlraum. Dieser wird hauptsächlich von den Elastomeren eingenommen. Das übrige Volumen soll möglichst von Sauerstoff frei gehalten werden, damit

die Kontakte des Wafers nicht oxidieren und somit die Leitfähigkeit beeinträchtigt wird. Hierzu wird das Volumen möglichst gut abgedichtet und anschließend mit Stickstoff durchspült, so dass sich das Volumen mit Stickstoff füllt. Zum Abdichten zwischen Waferträger und Hauptplatine werden dazu die Dichtungsringe verwendet. Die Dichtungsringe haben eine Größe, die gerade zu den entsprechenden Nuten im Waferträger passt. Die Löcher auf der Unterseite des Waferträgers sowie die Löcher, die zum Befüllen des Stickstoff verwendet werden, werden mit Schrauben verschlossen, in deren Kopf ein kleiner Dichtungsring eingelassen ist.

2.3 Wafer

Der verwendete Wafer ist eine Siliziumscheibe ohne aktive Strukturen. Als Testwafer wird ein sogenannter Short Loop Wafer¹ benutzt. Es sind mehrere Schritte notwendig um Strukturen auf einem Wafer aufzubringen. Prinzipiell verfährt man dabei ähnlich zu dem Postprozessierungsverfahren, das in Kapitel 3 vorgestellt wird. Im Gegensatz zur Postprozessierung werden die feineren Strukturen auf dem Wafer nicht auf einmal belichtet, sondern mit Hilfe eines sogenannten Step-and-Repeat-Verfahrens aufgebracht. Dabei wird eine Fläche von maximal (25×25) mm² belichtet. Anschließend wird eine benachbarte Fläche belichtet. Dies wird solange wiederholt, bis der gesamte Wafer belichtet wurde. Als Ergebnis besteht die Waferfläche aus aneinander gereihten Rechtecken. Eine solche rechteckige Fläche wird als Retikel bezeichnet. Auf dem verwendeten Testwafer wurden 60 Retikel aufgebracht, welche jeweils nur einige horizontale Verbindungen und kleine Pads an den Rändern aufweisen. Horizontale Verbindung bedeutet, dass sie bezogen auf den oben definierten Süden von West nach Ost verlaufen. Die kleinen Pads liegen entsprechend auch jeweils am westlichen bzw. östlichen Rand.

Der Testwafer wurde hauptsächlich zum Test der nachträglich aufgebrachten Verbindungen zwischen den Retikeln, der sogenannten Postprozessierung, entworfen. Während dieses Schrittes wurden auch $(2 \times 0,19)$ mm große Pads aufgebracht. Mit Hilfe dieser Postprozessierungspads können die vertikalen Verbindungen über Elastomerverbinder zur Hauptplatine getestet werden.

In Kapitel 3 werden sowohl die Herstellungstechnik der Postprozessierung und die verwendeten Strukturen, als auch die Ergebnisse der Tests der Postprozessierung diskutiert.

2.4 Elastomerverbinder

Die Elastomerverbinder sind auf dem ersten Blick einfache Streifen aus einem elastischen Material. Genauer betrachtet wechseln sich zwei verschiedene Materialien der Länge nach ab. Das eine Material ist ein elastischer Kunststoff. Das andere ist ebenfalls ein Kunststoff, jedoch sind in diesem Silberkugeln eingelassen. Auf Druck pressen sich die Kugeln aneinander und verbinden somit die Unterseite des Elastomerverbinders mit der Oberseite elektrisch. In Abb. 2.3 wird dieses Prinzip veranschaulicht. Fragen wie, mit welchem Pitch kann man Signale über den Elastomer schicken, welche mechanischen und elektrischen Belastungen verträgt er, wie fehlertolerant gegen Verschiebung und Kompression ist er, werden in Kapitel 6 aufgegriffen.

¹Ein Short Loop Wafer ist ein Wafer ohne aktive Strukturen und nur einer Metallage. Short Loop Wafer werden normalerweise für Packaging-Tests verwendet.

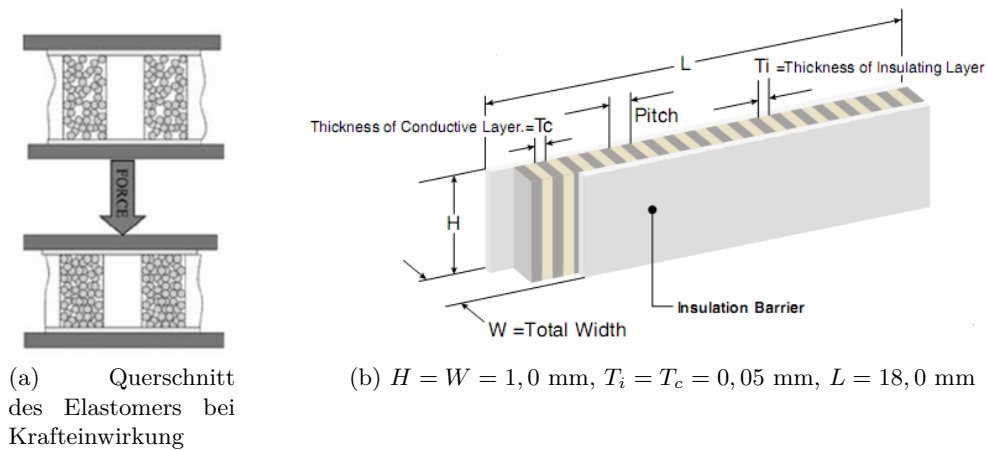


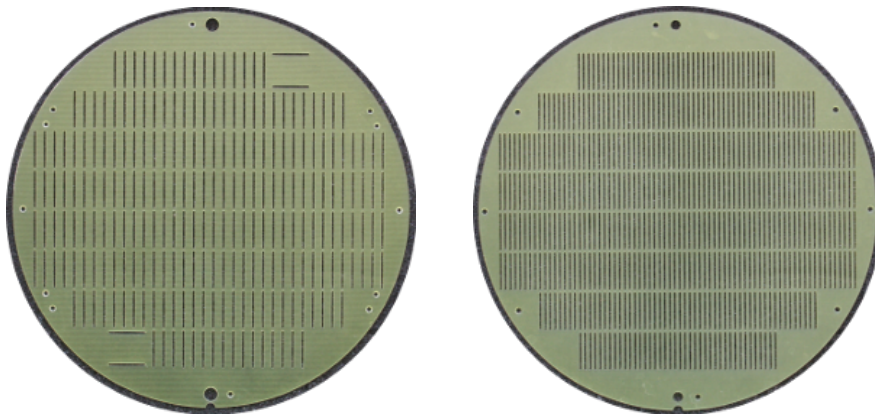
Abbildung 2.3: Elastomerverbinder ZEBRA[®] Ag (1.0 - IB2 x 18 x 1.0) von der Firma Fujipoly.

2.5 Positionierungsmaske

Die zuvor beschriebenen Streifenverbinder liegen zwischen Wafer und Hauptplatine. Es werden 236 bzw. 480 Stück benötigt. Die Elastomere können prinzipiell kleine Verschiebungen in x- und y-Richtung ausgleichen, sind jedoch sehr anfällig gegen Verdrehungen. Es ist daher nötig die Elastomere in bestimmten Grenzen vor zu Justieren. Dabei gibt es die Möglichkeit sie auf dem Wafer zu legen und die Platine aufzusetzen oder die Elastomere an der Platine zu befestigen und anschließend auf den Wafer aufzulegen. Da der spätere Wafer sehr empfindlich und teuer sein wird, ist es besser die Elastomere an der Platine auszurichten. Man muss dabei jedoch beachten, dass die Verbinder beim Zusammensetzen des Systems an der Unterseite der Hauptplatine sitzen werden.

Für diese Anforderungen wurde eine Positionierungsmaske entwickelt. Sie besteht aus einem 0,55 mm dicken Verbundstoff, welcher auch zur Herstellung von Platinen verwendet wird. Die Maske ist also ein elektrisch nicht leitendes Material. In die Maske werden Schlitze gefräst, in welche die Elastomerverbinder gesteckt werden. Man kann somit die Elastomere relativ zueinander und durch Ausrichten der Schablone zur Hauptplatine auch zu den Strukturen auf der Hauptplatine positionieren. Das Ausrichten zwischen Schablone und Hauptplatine wird über kleine Senkkopfschrauben erreicht. Diese werden von der Unterseite durch die Maske und die Platine in den Deckel geschraubt. Durch die angesenkten Bohrungen in der Maske und genügend genauen Durchgangsbohrungen in der Platine werden die Elastomere ausreichend gut ausgerichtet. Ausreichend gut bedeutet dabei, dass wir eine Toleranz in x-Richtung von 0,5 mm und in y-Richtung von ca. 0,2 mm haben. Die Toleranz in x-Richtung ergibt sich aus den 1,2 mm breiten Kontaktierflächen auf der Hauptplatine, bei nur 1 mm breiten Elastomeren. Die Toleranz in y-Richtung erhält man dadurch, dass nicht die gesamte Länge des Elastomers zur Kontaktierung vorgesehen ist, da nur knapp aufliegende Elastomere am Rand eine unzuverlässigen Kontaktierung bedeuten. In welchem Rahmen diese Toleranzen in y-Richtung zulässig sind wird auch in Kapitel 6 besprochen.

Abbildung 2.4 zeigt zwei Photos von verschiedenen Masken. Der Grund für zwei verschiedene Masken liegt zum einen darin, dass zu Beginn die Maske mit 236 Ausfräsungen gefertigt wurde, um das erdachte Prinzip zu testen ohne jedoch eine Instabilität der Schablone zu riskieren. Zum anderen wurde die Form und Größe der Schlitze auf die zuerst



(a) Maske für 236 Elastomere. 4 auf 58 Retikeln plus 2 auf 2 Retikeln für die elektrische Justage.

(b) Maske für 480 Elastomere. 8 auf 60 Retikeln.

Abbildung 2.4: Fotos von zwei unterschiedlichen Masken. Beide wurden vom Technikon des Kirchhoff-Instituts für Physik der Universität Heidelberg gefertigt.

gelieferten Elastomere angepasst. Wie sich später dann heraus stellte war dies eine Fehlproduktion und die Nachbestellung einer größeren Stückzahl beinhaltete etwas schmalere Elastomere (siehe Abb. 2.5). In Abbildung 2.5 ist zu erkennen, dass die seitliche Isolierung bei dem einen Elastomer doppelt so dick ist, wie bei dem anderen. Abbildung 2.6 zeigt die Form und Größe der unterschiedlichen Ausfräsungen. Die gebogenen Form der

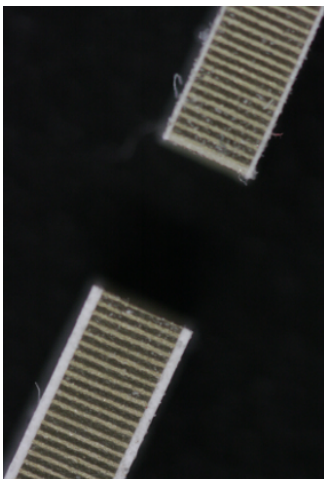
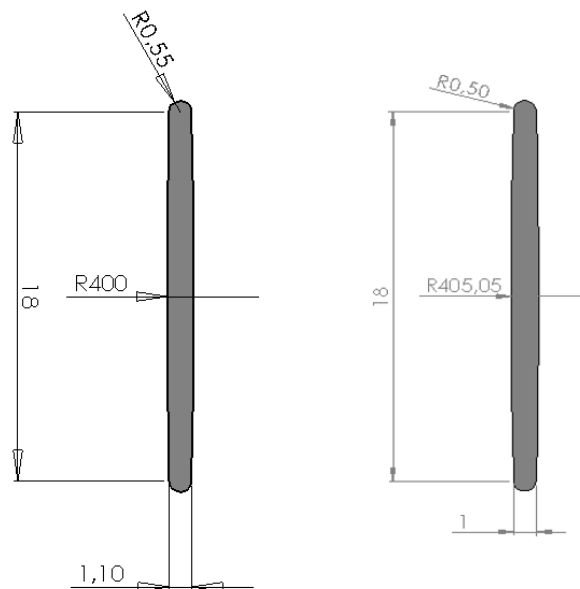


Abbildung 2.5: Foto der verschiedenen Elastomere.



(a) Ausfräsung für die zuerst bestellten Elastomere.

(b) Ausfräsung für die nachbestellten Elastomere.

Abbildung 2.6: Solidworkszeichnung inkl. Bemaßung für die zwei verschiedenen Ausfräsungsformen.

Längsseiten dient als zusätzlicher Raum für die gepressten Elastomere. Diese Form der Ausfräsung gewährt bei guter Positionierung zum einen den Halt der Elastomere auf der Unterseite der Platine und zum anderen Platz für die seitliche Ausdehnung der Elastomere beim Zusammendrücken.

Wie zuvor kurz erwähnt wird die Schablone durch die Platine an den Deckel geschraubt. Man muss hierbei daran denken, dass dadurch eine Verbindung zwischen dem abgeschlossenen, sauerstofffreien Zwischenraum zwischen Wafer und Platine und dem Deckel besteht. Es ist also notwendig auch diese Verbindung abzudichten. Zum Abdichten wird eine flüssige Gewindedichtung im Durchgangsloch bzw. Gewinde im Deckel für die Befestigungsschrauben verwendet, welche nach kurzer Trockenzeit aushärtet.

2.6 Hauptplatine

Die Hauptplatine ist ein PCB (printed circuit board), welches in enger Zusammenarbeit mit der Firma Würth Elektronik entworfen und gefertigt wurde. Von den oben genannten 236 bzw. 480 Elastomeren werden jeweils nur 16mal 2 Elastomerverbinder zur Signalübertragung zwischen Testwafer und Hauptplatine verwendet. Dies liegt daran, dass es sehr aufwendig ist die Leitungen so zu verlegen, dass die Platine noch gut zu fertigen ist, also der Aufwand und die Kosten so niedrig wie möglich gehalten werden, um dennoch einen vernünftigen Test zu gewährleisten. Als Mittelweg wurde eine 6-lagige Platine entwickelt, welche jeweils 60 Leitungen als differentielle Pärchen von 16-Retikeln nach außen an Stecker führt. Das ganze geschieht auf 3 Signallagen, welche jeweils mit Powerlagen abgeschirmt werden. Die Powerlagen sind auf dem Testboard einfache Kupferflächen. Ein genauerer Aufbau der Platine und ihrer Lagen wird in Kapitel 5 vorgestellt.

2.7 Deckel

Als Abschluss in diesem Aufbau wird oben ein Deckel aufgesetzt. Dieser besteht ebenfalls wieder aus einer 12 mm dicken Aluminiumplatte. Er hat einen Ring Durchgangslöcher mit Versenkung für die Köpfe der Schrauben, die zum Zusammenhalt des Systems und seiner Komponenten und für einen ausreichenden Anpressdruck der Elastomere sorgen. Die Form des Deckels ist analog zum Waferträger ein Achteck. Abbildung 2.7 zeigt Fotos und Bemaßungen des Deckels.

Die Strukturen auf der Unterseite des Deckels waren in einem älteren Konzept für DNC's

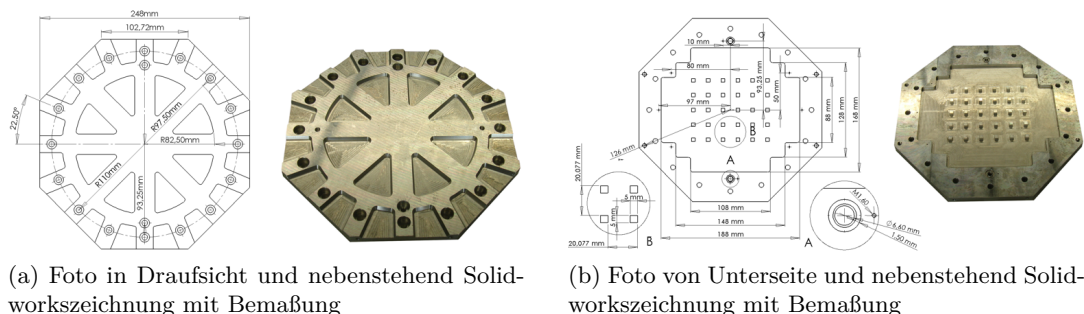


Abbildung 2.7: Draufsicht und Sicht auf Unterseite vom Deckel. Beschreibung der Ausfräsungen und Löcher stehen im Text.

(digital network chip) vorgesehen, welche direkt über dem Wafer und der Hauptplatine im Deckel versenkt werden sollten. Dieses Konzept wurde allerdings geändert. Diese Chips,

welche für die Kommunikation mit externen Geräten unabdinglich sind und darüber hinaus auch zur Kommunikation innerhalb des Wafers verwendet werden können, werden im momentanen Konzept außerhalb des Deckels und Waferträgers sowohl auf Ober- wie auch Unterseite der Hauptplatine verschoben. Der Platz, der sich durch diese Vertiefungen ergibt wird jetzt für kleinere Bauteile, vor allem Blockingkapazitäten und Leistungstransistoren verwendet.

Kapitel 3

Eigenschaften des Testwafer und der Postprozessierung

Wafer sind Siliziumscheiben, welche als Trägermaterial für integrierte Schaltungen dienen. Übliche Durchmesser sind 100 mm, 200 mm oder 300 mm. Auf ihnen werden mit Hilfe unterschiedlicher Prozesse minimale Gatebreiten bis 45 nm MOSFETs (metal oxide semiconductor field-effect transistor) durch Belichtung aufgebracht. Dies läuft im sogenannten Step-and-Repeat-Verfahren ab, bei dem einzelne Teilflächen nacheinander belichtet werden. Die Flächen, die auf einmal belichtet werden, sind auf wenige Quadratzentimeter beschränkt. Man bezeichnet diese Flächen als Retikel. Für den finalen 200 mm-Durchmesser Wafer, welcher von UMC (united microelectronics cooperation) in einem 180 nm Prozess hergestellt wird, wurden 20 mm auf 20 mm große Retikel gewählt. Diese Größe ergibt sich daraus, dass der entwickelte HICANN-Chips in der Entwicklungsphase in einer kleinen Stückzahl produziert wird, wobei eine maximale Größe von 5 auf 10 mm möglich. Bei einer Anordnung von jeweils 4 HICANN-Chips in 2 Reihen auf einem Retikel ergeben sich somit 20 auf 20 mm.

3.1 Short Loop Wafer

Der Testwafer hat einen Durchmesser von 200 mm. Er wurde aus Kostengründen ein Short Loop Wafer hergestellt. Die Strukturen auf dem Testwafer wurden nur auf einer Metallage, Metal6, aufgebracht. Metal6 bedeutet dabei, dass es sich um die sechste von sechs Metallagen aus dem UMC-Prozess handelt. Abbildung 3.2 zeigt die Strukturen dieses Testwafers vor der Postprozessierung. Ein Retikel hat 8 Padleisten, zu je 60 Pads. Jedes zweite Pad einer Reihe ist über horizontal verlaufende Metal6-Leitungen mit seinen Nachbarn verbunden. Die Dicke der Metal6-Lage und somit die Höhe der Leitungen beträgt 0,8 μm . Am rechten und linken Ende eines Retikels enden ausgewählte Leitungen an kleineren Pads. Diese werden später über Postprozessierung mit dem Nachbarretikel verbunden.

3.2 Postprozessierung des Wafers

Um die einzelnen Retikel miteinander zu verbinden und somit später die gesamte Fläche des Wafers effektiv nutzen zu können, ist ein zusätzlicher Fertigungsschritt erforderlich. Dieser wird als Postprozessierung bezeichnet. Dabei werden die kleineren Pads eines Retikels (siehe Abbildung 3.2 und 3.3) mit denen des benachbarten Retikels über die

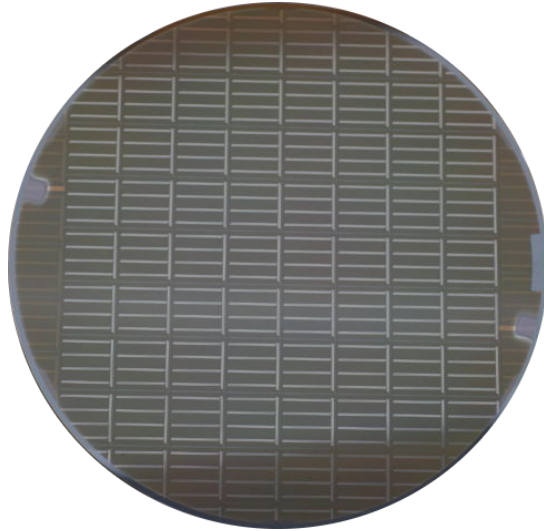


Abbildung 3.1: Photo eines Wafers ohne Postprozessierung. Der Wafer besteht aus 60 ganzen Retikeln. Zusätzlich befinden sich am Rand abgeschnittene Retikel. Der zuvor definierte Süden des Wafers liegt im Bild rechts.

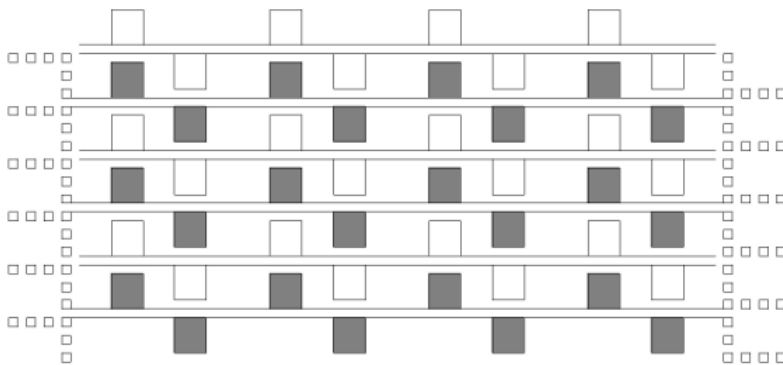
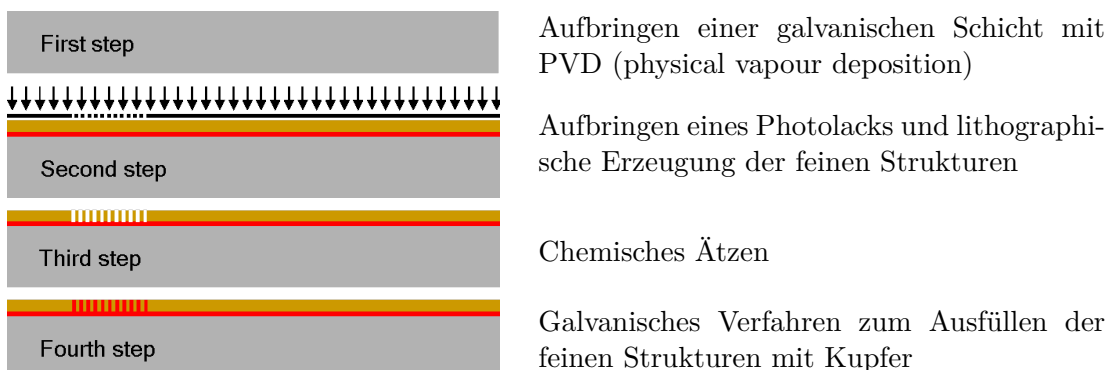


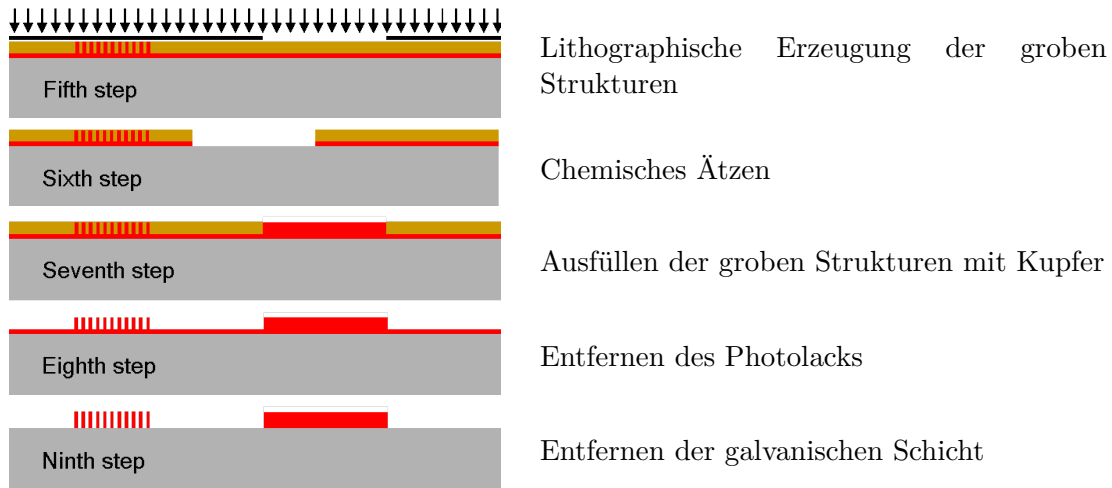
Abbildung 3.2: Schema eines Retikels. Im Gegensatz zu den weißen werden die grauen Pads via Postprozessierung über die kleineren Pads rechts bzw. links mit dem nächsten Retikel verbunden.

77 μm breite Lücke zwischen den Retikeln miteinander verbunden. Die genaue Vorgehensweise und die Ergebnisse werden in den folgenden vier Unterabschnitten beschrieben.

3.2.1 Postprozessierung: Verfahren

Für die Postprozessierung werden verschiedene Masken angefertigt, mit welchen man jeweils den gesamten Wafer belichten kann. Die folgenden neun Verfahrensschritte werden dabei verwendet:





Die feinen Strukturen der Postprozessierung werden verwendet um die Retikel miteinander zu verbinden, während die groben Strukturen die Pads für die Verbindung senkrecht nach oben vom Wafer weg bilden. Die durch dieses Verfahren entstandene Pads haben eine Kupferoberfläche. Diese Kupferoberfläche, welche zur Kontaktierung dient, verändert durch Oxidation ihre Kontakteigenschaften. Es ist daher nicht vorteilhaft sie aus Kupfer zu fertigen.

Aus diesem Grund wurde ein zweites Postprozessierungsverfahren entwickelt, bei dem die Oberfläche zur Kontaktierung vergoldet wird. Dazu wird nach Schritt 6 für das galvanische Verfahren kein Kupfer sondern Nickel/ Gold verwendet. Abbildung 3.3 ist eine schematische Darstellung eines Retikels mit Postprozessierungsstrukturen. Wie man sieht werden

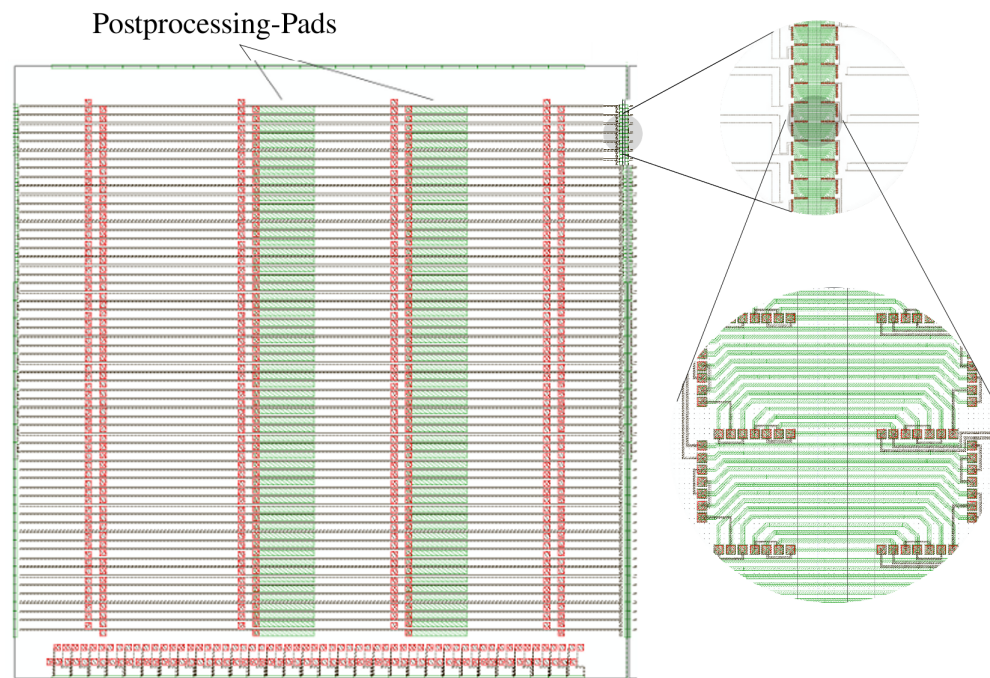


Abbildung 3.3: Schema eines Retikels nach der Postprozessierung. In der Mitte sind die Postprozessierungspads zu sehen, am Rand rechts und links die Verbindungen zum benachbarten Retikel, welche teilweise auch vergrößert dargestellt sind.

nur zwei der 8 Padleisten an große Pads angeschlossen. Die anderen Pads können allerdings auch über Messnadeln kontaktiert werden.

3.2.2 Postprozessierung: Strukturen

Aus Abbildung 3.3 kann man auch erkennen, dass die feinen Postprozessierungsstrukturen nicht gleich sind. Es wurden insgesamt sechs verschiedene Strukturen getestet. Die Strukturen unterscheiden sich in Linienbreite, -abstand und Padgröße. In der Tabelle 3.2 werden diese Strukturen aufgelistet. Abbildung 3.4 zeigt den Testwafer mit zwei vergrößerten Ausschnitten. Da bei den Strukturen 5 und 6 die Pads bei der Postprozessierung ineinander

STRUKTUR	BESCHREIBUNG	VERBINDET PINREIHE
Struktur1	Die Padgröße beträgt $8\ \mu\text{m} \times 8\ \mu\text{m}$ und der Leitungsabstand beträgt $8\ \mu\text{m}$. Vergl. „Abb. „	1 bis 6
Struktur2	Padgröße $8\ \mu\text{m} \times 8\ \mu\text{m}$; Leitungsabstand $6\ \mu\text{m}$	7 bis 14
Struktur3	Padgröße $8\ \mu\text{m} \times 8\ \mu\text{m}$; Leitungsabstand $5\ \mu\text{m}$	15 bis 20
Struktur4	Padgröße $5\ \mu\text{m} \times 5\ \mu\text{m}$; Leitungsabstand $6\ \mu\text{m}$	21 bis 40
Struktur5	Padgröße $4\ \mu\text{m} \times 4\ \mu\text{m}$; Leitungsabstand $5\ \mu\text{m}$	41 bis 50
Struktur6	Padgröße $4\ \mu\text{m} \times 4\ \mu\text{m}$; Leitungsabstand $4\ \mu\text{m}$	51 bis 60

Tabelle 3.2: Übersicht über die verschiedenen Strukturen

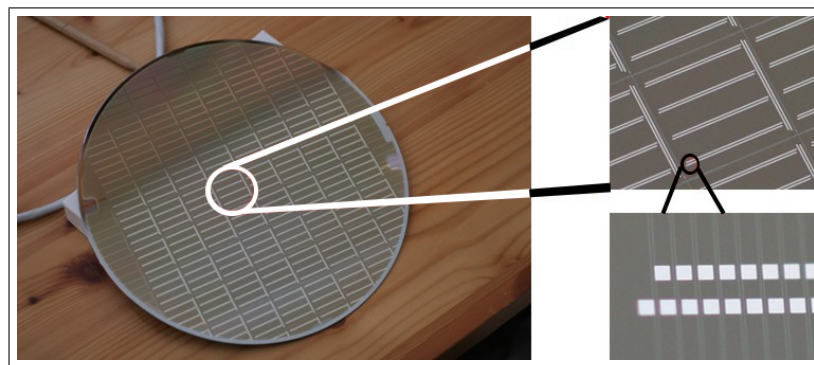


Abbildung 3.4: Foto des Testwafers mit Vergrößerungsausschnitten eines Retikels (oben rechts) und der Kontaktflächen für Messnadeln (unten rechts).

gelaufen sind, verursachen diese alle einen Kurzschluss und somit ist keine Auswertung dieser Strukturen möglich. Man erkennt dies in Abbildung 3.5. Abbildung 3.6 zeigt zum Vergleich Struktur 1. Aufgrund dieser Kurzschlüsse werden bei den folgenden Messungen nur die Postprozessierungswiderstände an den Pins 1 bis 39 gemessen. Tabelle 3.2 zeigt welche Pins über welche Strukturen miteinander verbunden sind. Eine Ausnahme bilden hier die Übergänge zum letzten Retikel einer Reihe. Hier gibt es durch einen konzeptionellen Fehler eine Überlappung verschiedener Strukturen, was zu keiner Verbindung zwischen den Retikeln bei den Strukturen 2 und 3 führt. Abbildung 3.9 veranschaulicht dies. Bei den Strukturen 1 bis 3 wird nur jeder zweite Pin angeschlossen, da die Verbindung zwischen den Pads nicht eine einfache Verbindung über die Retikelgrenze ist, sondern in einem bestimmten Muster verlegt wird. Bei den größeren Strukturen hat dies zur Folge, dass nicht

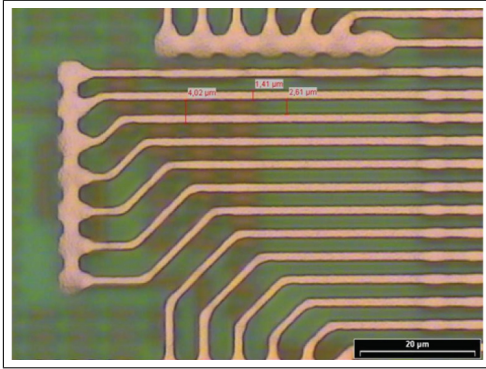


Abbildung 3.5: Foto der Pads von Struktur 6

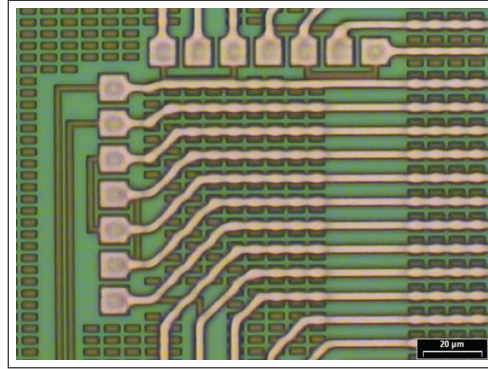


Abbildung 3.6: Foto der Pads von Struktur 1

genügend Platz ist um jeden Pin mit seinem Gegenüber am nächsten Retikel zu verbinden. Abbildung 3.7 zeigt eine Retikelverbindung zwischen den Pins 2, also von Struktur 1. In Abbildung 3.8 sieht man eine solche aus Struktur 5.

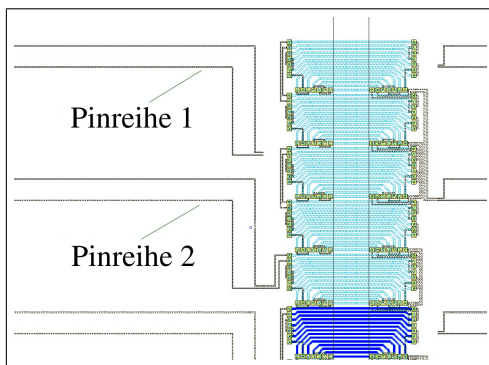


Abbildung 3.7: Schema der Postprozessierung von Struktur 1. Die dunkleren Linien der Postprozessierung verbinden Pinreihe 4.

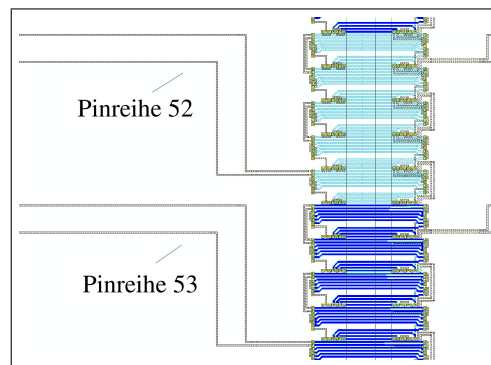


Abbildung 3.8: Schema der Postprozessierung von Struktur 5. Die dunkleren Linien verbinden Pinreihe 51 bzw. Pinreihe 53.

Bei der Messung werden vier Nadeln verwendet. Die Nadeln werden so angeordnet, dass sie auf den Pads n , n' , n'' und n° aufliegen (Abbildung 3.10). Es wird der Widerstand zwischen n und n'' und der Widerstand zwischen n' und n° gemessen. Die Messung zwischen n und n' dient zur Überprüfung auf einen möglichen Kurzschluss.

Bei den folgenden Widerständen handelt es sich um den gemessenen Widerstand einer 2-Draht-Widerstandsmessung. Das heißt, es wird der Widerstand von Pad zu Pad zuzüglich dem Widerstand der Messspitzen gemessen und nicht nur der Widerstand der Postprozessierungsstruktur. Der Widerstand von den Pads bis zur Postprozessierung kann auf $(105 \pm 3) \Omega$ für die Strukturen 1 bis 3 und auf $(118 \pm 4) \Omega$ für die Struktur 4 abgeschätzt werden. Dies ergibt sich aus der Länge der Metal6-Leitungen l bei gegebener Metal6-Dicke d , -breite b und spezifischem Widerstand ρ .

$$R = \rho * \frac{l}{d * b}$$

Wobei $d = 0,8 \mu\text{m}$, $b = 2 \mu\text{m}$ und $l = 6,4 \text{ mm}$ (bei den Strukturen 1 bis 3) bzw. $l = 7,2 \text{ mm}$

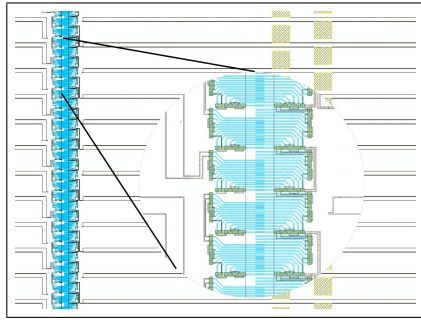


Abbildung 3.9: Schema der Postprozessierung. Man sieht den Fehler von Struktur 2 und 3 beim Übergang zum letzten Retikel.

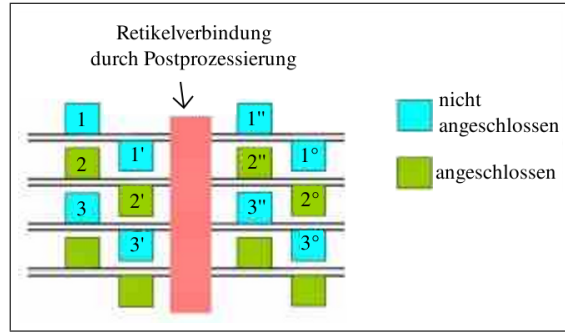


Abbildung 3.10: Darstellung zur Lage der vier Nadeln einer Messung.

(für Struktur 4) ist. Vom Hersteller wird ein Oberflächenwiderstand $\rho_f = (23 \pm 13) \frac{m\Omega}{sq}$ angegeben. Eine Vergleichsmessung von Metal6-Leitungen bekannter Länge auf dem Testwafer, hat ein $\rho_f = (32,8 \pm 1,0) \frac{m\Omega}{sq}$ ergeben. Umgerechnet in den spezifischen Widerstand ergibt dies $\rho = (26,2 \pm 0,8) \frac{m\Omega}{m}$. Dieser Wert ergibt in die obige Formel eingesetzt die oben stehenden Widerstände der Metal6-Leitungen bis zur Postprozessierung. Man erhält den tatsächlichen Postprozessierungswiderstand aus der Differenz der gemessenen Werten und dem eben abgeschätzten Widerstand. Der Widerstand der Messspitzen kann vernachlässigt werden, da er im Vergleich zum Metal6-Widerstand sehr klein ist (ca. 500 m Ω).

3.2.3 Postprozessierung: Ergebnisse

3.2.3.1 Struktur 1

Übergang auf:	B	C	D	E	F	G	H	I
1			143,4 ± 4,2	143,1 ± 3,6	142,3 ± 3,6	144,0 ± 3,7		
2		142,2 ± 3,2	143,5 ± 3,1	145,6 ± 3,5	144,6 ± 3,5	143,4 ± 3,3	141,6 ± 3,3	
3	139,7 ± 3,0	142,2 ± 2,9	146,0 ± 2,9	148,8 ± 3,1	148,1 ± 3,1	146,4 ± 3,0	144,1 ± 2,9	142,9 ± 3,1
4	141,3 ± 3,2	143,9 ± 3,0	146,1 ± 2,9	148,0 ± 2,9	146,4 ± 2,9	148,1 ± 3,1	145,6 ± 2,9	149,0 ± 10,7
5	139,4 ± 0,8	142,7 ± 0,7	141,8 ± 0,2	144,1 ± 0,1	145,1 ± 0,6	142,3 ± 0,1	142,3 ± 0,2	140,5 ± 0,1
6	142,2 ± 0,2	145,3 ± 0,4	146,3 ± 0,2	147,0 ± 0,2	146,2 ± 0,2	144,5 ± 0,2	142,1 ± 0,1	139,3 ± 0,3
7		146,1 ± 0,4	151,1 ± 0,4	151,1 ± 0,5	150,1 ± 0,2	148,2 ± 0,3	142,4 ± 0,2	
8			150,1 ± 0,5	156,3 ± 1,4	163,1 ± 2,3	163,0 ± 1,9		

Abbildung 3.11: Verteilung der mittleren gemessenen Widerstände bei Struktur 1.

Verbindungswiderstände Die Abbildung 3.11 zeigt, wie sich die Verteilung der mittleren Widerstände von Struktur 1 über die Fläche verteilt. Die Widerstände werden ten-

denziell zum linken und unteren Rand hin größer. Insbesondere in der letzten bzw. unteren Reihe werden die Widerstände deutlich größer. Die Ursache für diese Verteilung ist in dem Herstellungsprozess zu finden. Die Umgebungstemperaturen während der Herstellung der Postprozessierungsmaske für die kleinen Strukturen und während der Postprozessierung selber können unterschiedlich¹ sein. So ist es möglich, dass die Maske wegen thermischer Ausdehnung nicht ganz mit der Struktur des Wafers übereinstimmt. Als Folge sind die Schnittpunkte zwischen Wafer- und Maskenstruktur nicht immer deckungsgleich. In unserem Fall bedeutet das, dass die Pads der Postprozessierung nicht ganz, sondern nur teilweise auf den Pads des Wafers liegen (Abbildung 3.12) und sich dadurch der Kontaktwiderstand erhöht. Da die Maske allerdings auf einem bestimmten Punkt justiert wird, sind an manchen Stellen die Pads deckungsgleich (Abbildung 3.13). An Stellen mit niedrigen Verbindungswiderstände liegen die Pads somit genau übereinander, bei größeren Widerständen verschiebt die Maske die Postprozessierungspads gegenüber den Pads auf dem Wafer. Es sei hier noch einmal darauf hingewiesen, dass die angegebenen Widerstände

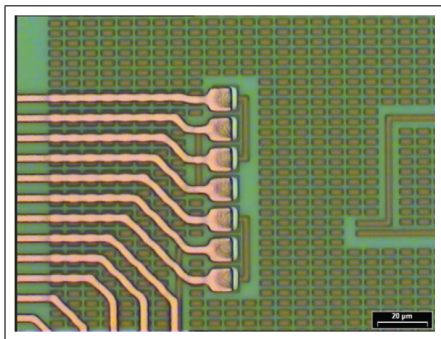


Abbildung 3.12: Pads sind nicht deckungsgleich aufgrund der Verschiebung zwischen Wafer und Maske.

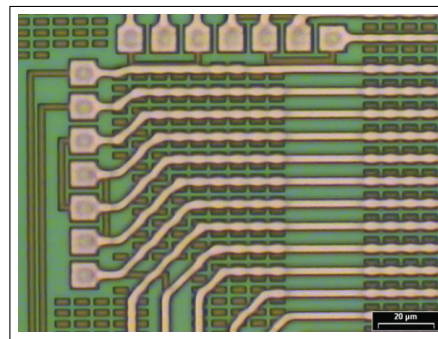


Abbildung 3.13: Deckungsgleiche Pads, das heißt keine Verschiebung zwischen Maske und Wafer.

den oben beschriebenen konstanten Widerstand von ca. $(105 \pm 3) \Omega$ beinhalten. Der tatsächliche Widerstand der Postprozessierungsstruktur verdoppelt sich also vom kleinsten Widerstand zum größten.

Bei genauerer Betrachtung der Werte fällt auf, dass die Fehler der Werte in der oberen Hälfte des Wafers größer sind. Die Ursache hierfür ist weniger in dem Aufbau des Wafers, der Postprozessierung oder im Konzept des Aufbaus, als vielmehr in der Messung selber zu finden. Wie oben beschrieben, wird mit vier Nadeln gemessen. Es stellte sich heraus, dass die Nadeln einen Einfluss auf die Widerstandsmessung haben können. So wurde die obere Hälfte des Wafers mit einer Nadelkonfiguration gemessen, bei der man für den Widerstand zwischen Pin n und n'' einen um einige Ohm (5.5Ω) erhöhten Widerstand maß. Bei der unteren Hälfte wurden andere Nadeln in der selben Anordnung verwendet und die Werte lagen alle im selben Bereich. Diese Tatsache ist bei allen Strukturen gleich, doch ist der Trend der höheren Fehler nicht so deutlich zu erkennen, da bei den feineren Strukturen die Widerstandswerte stärker schwanken und somit dieser systematische Fehler nicht maßgeblich ist.

¹Zur Abschätzung des minimalen Unterschiedes der Umgebungstemperatur kann man folgende Betrachtung machen. Die Postprozessierungsmaske besteht aus Natriumglas mit einem Wärmeausdehnungskoeffizienten von $\alpha = 3,9 * 10^{-6} \text{ K}^{-1}$ (aus [10], Kap. 5.1.3)). Über dem Waferdurchmesser von $l = 200 \text{ mm}$ entspricht eine Längenänderung $\Delta l = 1 \mu\text{m}$ ein Temperaturunterschied von $\Delta T = \frac{\Delta l}{l} * \alpha^{-1} = 1,3 \text{ K}$.

Kurzschlüsse Die Messung der Kurzschlußwiderstände fällt bei Struktur 1 sehr häufig positiv aus. So wird in allen Retikelreihen außer Reihe 3 in Pinreihe 6 ein Kurzschluss gemessen. Dass jeweils zwischen der gesamten Pinreihe (über alle benachbarten Retikel einer Reihe) ein Kurzschluss ist, liegt daran, dass auf einem Retikel die Pins einer Reihe miteinander verbunden sind und zwischen den Retikeln über Postprozessierung verbunden wurden. Es ist allerdings auffällig, dass der Kurzschluss immer nur in Pinreihe 6 gemessen wird. Dafür ist der konzeptionelle Fehler im Übergang zum letzten Retikel einer Reihe verantwortlich (vergl. Abbildung 3.9). Da die äußersten Pins des letzten Retikels einer Reihe über Postprozessierung mit dem Pin zwei Reihen tiefer verbunden sind (siehe Abbildung 3.15), wird Pinreihe 6 mit dem Strukturfehler beim Retikelübergang der Pinreihe 8 verbunden. Diese Pinreihe weist allerdings immer einen Kurzschluß auf, wie in Abbildung 3.9 zu sehen ist. Die Tatsache, dass Retikelreihe 3 keinen Kurzschluss bei Struktur 1 aufweist, lässt sich auch mit Abbildung 3.15 erklären. Aufgrund der fehlenden Verbindung zwischen Pinreihe 6 und 8 gibt es auch keinen Kurzschluß.

Ein weiterer Punkt, der für diese Verbindung am Ende einer Reihe als Ursache des Kurzschlusses spricht, ist, dass die Widerstandswerte der Kurzschlüsse zum Ende einer Reihe kleiner werden. Der Leitungswiderstand bis zum Kurzschluss wird geringer, also ist der Weg zum Kurzschluss kürzer.

3.2.3.2 Struktur 2

Übergang auf:	B	C	D	E	F	G	H
1			147,4 ± 3,5	147,2 ± 3,4	146,4 ± 3,4		
2		146,6 ± 3,2	148,5 ± 3,2	151,4 ± 3,4	149,2 ± 3,3	147,7 ± 3,1	
3	144,6 ± 2,8	146,4 ± 2,9	150,9 ± 2,9	154,5 ± 4,2	153,7 ± 3,0	151,4 ± 2,9	149,0 ± 3,0
4	146,8 ± 3,0	148,3 ± 3,0	150,0 ± 3,5	152,3 ± 6,2	151,7 ± 2,9	152,3 ± 2,8	150,0 ± 2,7
5	144,4 ± 0,3	148,2 ± 0,3	147,6 ± 0,4	149,9 ± 0,2	149,8 ± 0,3	147,8 ± 0,2	147,4 ± 0,3
6	147,4 ± 0,3	151,1 ± 0,2	152,0 ± 0,3	153,2 ± 0,4	151,9 ± 0,2	149,6 ± 0,2	147,1 ± 0,8
7		151,4 ± 0,2	157,3 ± 0,5	158,3 ± 0,7	157,8 ± 0,4	154,0 ± 0,4	
8			154,4 ± 0,5	166,9 ± 1,4	167,3 ± 1,3		

Abbildung 3.14: Verteilung der mittleren gemessenen Widerstände bei Struktur 2. Bei dem markierten Übergang wurde ein Messwert weggelassen. Hier gibt es eine offene Verbindung, welche in Abbildung 3.16 zu sehen ist.

Verbindungswiderstände Wie bei Struktur 1 weist auch Struktur 2 eine ähnliche Verteilung der mittleren Widerstände über die Fläche auf (Abbildung 3.14). Auch die Größenordnung der Unterschiede ist in etwa gleich groß, da bei diesen beiden Strukturen die Padgrößen gleich sind ($8 \mu\text{m} \times 8 \mu\text{m}$). Ein Unterschied zur Struktur 1 liegt darin, dass

in Struktur 2 ein Kontakt einen sehr hohen Widerstand aufweist ($R = 6,96 \text{ M}\Omega$, an Pin 10 beim Retikelübergang von C2 auf D2). Abbildung 3.16 zeigt die Ursache für die nicht vorhandene Verbindung. Ursache hierfür könnte ein Fehler in der Maske oder Verunreinigungen beim Belichten der Strukturen sein.

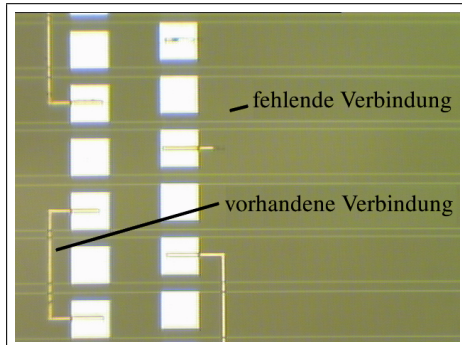


Abbildung 3.15: Das Foto zeigt Postprozessierungsverbindungen zwischen einer Pinreihe und ihrem übernächsten Nachbarn. In Retikelreihe 3 fehlt die Verbindung von Pinreihe 6 auf 8.

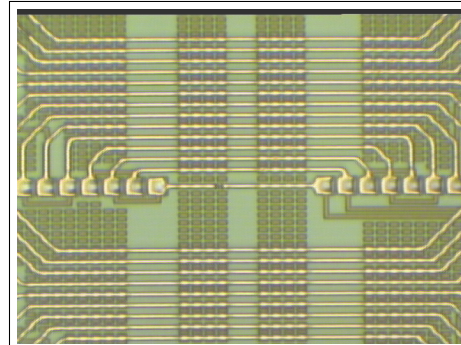


Abbildung 3.16: Foto der offenen Pinverbindung von Retikel C2 auf D2 in Pinreihe 10.

Kurzschlüsse Die Anzahl der Kurzschlüsse ist bei Struktur 2 wesentlich höher als bei Struktur 1. Bei genauerer Betrachtung fällt auf, dass nur in Retikelreihe 3 und 4 Kurzschlüsse auftreten und diese über alle Kontakte von Struktur 2. Anhand der Werte der Kurzschlusswiderstände kann man vermuten, dass der Kurzschluss zwischen D3 auf E3 und zwischen D4 auf E4 liegen muss. Eine Erklärung für dieses Verhalten konnte allerdings nicht gefunden werden.

3.2.3.3 Struktur 3

Verbindungswiderstände Auch bei dieser Struktur kann man eine Verteilung wie in den vorangegangenen Abschnitten beobachten (Abbildung 3.17). Es ist aufs Neue ein Unterschied der Größenordnung der Fehler zwischen oberer und unterer Waferhälfte zu erkennen. Die Unterschiede zwischen den Widerständen an verschiedenen Orten auf dem Wafer zeigen einen ähnlichen Verlauf und eine ähnliche Größenordnung wie bei den Strukturen 1 und 2. Dies liegt daran, dass man auch bei Struktur 3 eine Padgröße von $8 \mu\text{m} \times 8 \mu\text{m}$ wieder findet.

Offene Kontakte gibt es wie bei Struktur 1 nicht.

Kurzschlüsse Ähnlich zu Struktur 2 weisen in Retikelreihe 3 und 4 alle Postprozessierungsverbindungen einen Kurzschluss auf. Außerdem lassen sich anhand der Kurzschlusswiderstände die Kurzschlüsse auch hier zwischen den oben genannten Retikeln vermuten. Im Gegensatz zu Struktur 2, werden bei Struktur 3 noch zwei weitere Kurzschlüsse (Pinreihe 18 in Retikelreihe 5 und Pinreihe 16 in Retikelreihe 6) gemessen.

Der Kurzschluss in Retikelreihe 5 liegt vermutlich zwischen den ersten beiden Retikeln. Die genaue Ursache konnte nicht ermittelt werden.

Der Kurzschluss in Retikelreihe 6 hat einen kleinsten Widerstand von $12 \text{ M}\Omega$ und ist somit eigentlich kein Kurzschluss.

Übergang auf:	B	C	D	E	F	G	H
1			146,9 ± 3,1	147,6 ± 3,4	146,9 ± 3,5		
2		147,7 ± 3,2	149,4 ± 3,2	151,7 ± 3,3	149,9 ± 3,3	148,0 ± 3,3	
3	144,6 ± 2,9	146,6 ± 3,1	150,7 ± 3,7	155,0 ± 4,8	153,8 ± 3,0	151,7 ± 2,9	149,8 ± 2,9
4	146,5 ± 3,1	148,9 ± 3,1	150,4 ± 3,0	152,5 ± 5,8	151,9 ± 3,0	152,1 ± 2,8	150,5 ± 2,8
5	145,0 ± 0,5	148,4 ± 0,3	148,0 ± 0,3	150,9 ± 0,3	149,7 ± 0,1	148,1 ± 0,2	147,1 ± 0,3
6	148,5 ± 0,5	151,1 ± 0,2	153,2 ± 0,6	154,4 ± 0,5	151,9 ± 0,4	149,5 ± 0,1	146,9 ± 0,2
7		152,2 ± 0,5	157,3 ± 0,5	160,3 ± 0,7	159,1 ± 0,4	154,9 ± 0,8	
8			155,3 ± 0,6	167,4 ± 1,8	169,9 ± 1,8		

Abbildung 3.17: Verteilung der mittleren gemessenen Widerstände bei Struktur 3.

3.2.3.4 Struktur 4

Übergang auf:	B	C	D	E	F	G	H	I
1			18359,8 ± 75285,0	189,2 ± 9,4	4982,1 ± 18977,3	101191,5 ± 203289,0		
2		181,6 ± 21,4	173,9 ± 4,0	185,8 ± 6,7	203,2 ± 33,8	181,6 ± 7,6	869,8 ± 1694,6	
3	171,9 ± 22,1	173,6 ± 5,5	178,5 ± 5,1	200,7 ± 35,1	194,8 ± 18,1	206,0 ± 25,4	795,2 ± 1737,0	38520,4 ± 141530,8
4	179,2 ± 13,7	179,8 ± 8,9	171,0 ± 4,0	190,2 ± 4,4	174,7 ± 7,9	192,5 ± 10,5	192,6 ± 35,1	241,1 ± 231,8
5	185,6 ± 6,8	173,7 ± 2,6	174,0 ± 2,9	178,4 ± 2,6	167,1 ± 3,5	169,6 ± 2,8	193,3 ± 30,5	5361,8 ± 21697,9
6	208,9 ± 25,1	198,1 ± 4,5	203,5 ± 7,4	197,7 ± 6,8	197,0 ± 10,4	190,7 ± 18,3	265,8 ± 204,1	17721,9 ± 46985,2
7		449,4 ± 616,1	417,3 ± 239,0	6687,5 ± 27031,8	262,9 ± 16,1	297,5 ± 37,6	84897,2 ± 178943,0	
8			458783,0 ± 848979,9	342089,8 ± 741587,4	226837,7 ± 523508,1	584925,1 ± 932904,3		

Abbildung 3.18: Verteilung der mittleren gemessenen Widerstände bei Struktur 4.

Verbindungswiderstände Bei der kleinsten gemessenen Struktur zeigt die Verteilung der mittleren Widerstände über die Fläche (Abbildung 3.18) eine deutliche Abhängigkeit vom Ort. Man sieht hier deutlich, dass die Widerstände vor allem in der letzten Retikeldreiecke und im Übergang zum letzten Retikeldreiecke einer Reihe besonders groß werden. Dass die Unterschiede hier so groß werden, liegt daran, dass Struktur vier sehr kleine Padgrößen hat ($5 \mu\text{m} \times 5 \mu\text{m}$) und somit ein Verrutschen der Maske bei der Postprozessierung einen

nur geringen Padübergang zur Folge hat, was einen hohen Kontaktwiderstand bewirkt. Man sieht allerdings auch, dass wenn man das Problem mit dem Verrutschen der Maske lösen kann, die Widerstände zwar höher sind als in den zuvor beschriebenen Strukturen, sich die Werte der Widerstände aber auf einige Ohm gleichen.

Kurzschlüsse Kurzschlusswiderstände werden in verschiedenen Größen und an unterschiedlichen Stellen immer wieder gemessen. Es konnte kein Schema der Lage der Kurzschlüsse ausgemacht werden. Schließlich kommen in Struktur 4 auf 936 Verbindungen (in 72 Reihen) mindestens 16 Kurzschlüsse.

3.2.4 Postprozessierung: Bewertung

An den Histogrammen in Abbildung 3.19 bis 3.22 kann man erkennen, dass die Messungen der Widerstände einer Verteilung um einen Mittelwert entsprechen. Man kann aus ihnen darauf schließen, dass die Messungen repräsentativ für die jeweiligen Strukturen sind. Desweiteren erkennt man, dass die Breite der Verteilungen zu den feineren Strukturen hin zunimmt. Die Widerstände bei den gröberen Strukturen schwanken in einem kleineren Bereich und können somit genauer vorhergesagt werden.

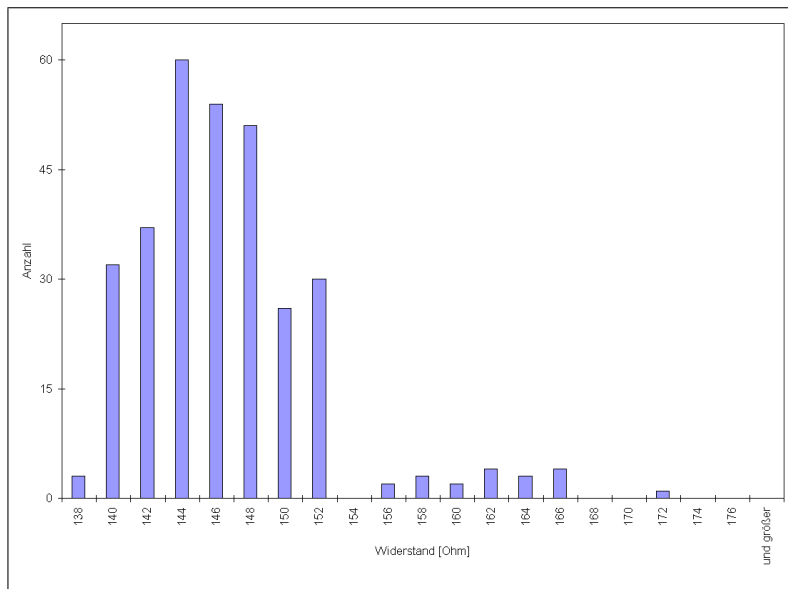


Abbildung 3.19: Histogramm der gemessenen Widerstände in Struktur 1

Als Fazit kann man festhalten, dass die Strukturen eins bis vier prinzipiell funktionieren. Aufgrund der thermischen Ausdehnung der Maske ist allerdings eine Struktur mit größeren Pads und vor allem größeren Padabständen notwendig. Somit kommt als minimale Struktur nur Struktur 3 in Frage. In Zusammenarbeit mit IZM² wurden für das endgültige System minimale Abstände zwischen den Padmitten von mindestens 9,5 μm und 15 μm auf 15 μm große Pads vereinbart.

²Fraunhofer Institut für Zuverlässigkeit und Mikroelektronik, Berlin

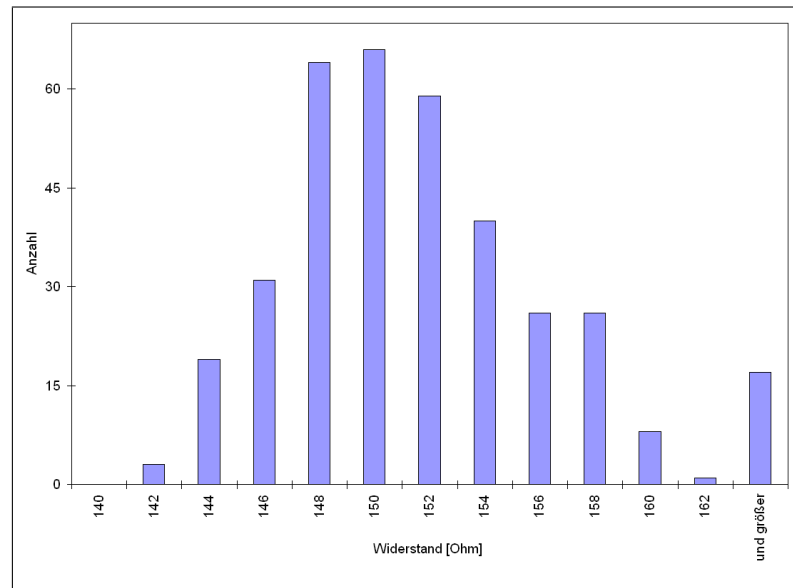


Abbildung 3.20: Histogramm der gemessenen Widerstände in Struktur 2

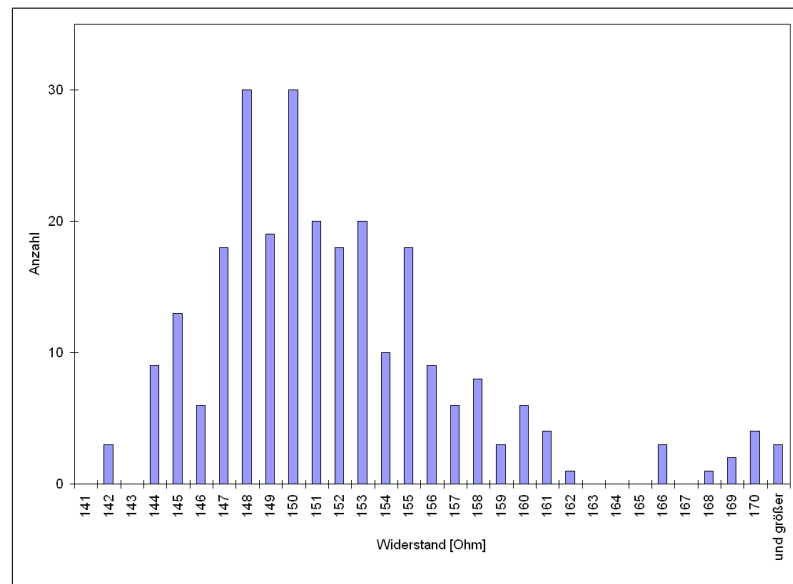


Abbildung 3.21: Histogramm der gemessenen Widerstände in Struktur 3

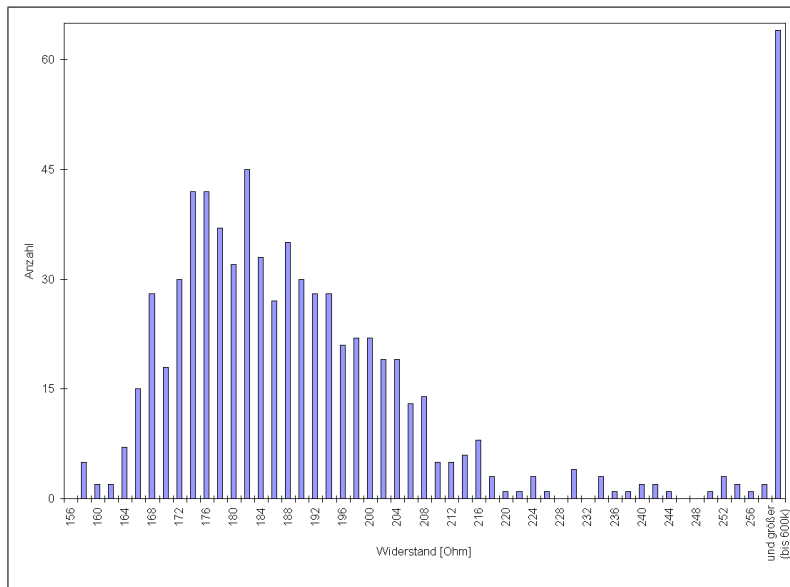


Abbildung 3.22: Histogramm der gemessenen Widerstände in Struktur 4

Kapitel 4

Prinzip der Justage der einzelnen Komponenten

Die Postprozessierung des Wafers wurde getestet und es hat sich als funktionierend für die gegebenen Anforderungen herausgestellt. Eine weitere Aufgabe besteht darin die Pads der Postprozessierung auf dem Wafer mit den entsprechenden Pads auf der Unterseite der Hauptplatine zu verbinden. Diese Verbindung wird über Elastomerverbinder (siehe Kap. 6) hergestellt. Da der Elastomerverbinder nur halb so breit ist wie die Pads auf dem Testwafer, auf denen er aufliegt, erfordert die horizontale Lage der Pads, welche in vertikalen Reihen angeordnet sind, keine sehr hohe Genauigkeit. Die genaue vertikale Lage der Elastomere muss auch nicht sehr exakt stimmen, da pro Pad zwei leitende Teile des Elastomers zur Verbindung bereit stehen und bei einer Verschiebung zwar ein Teil entfällt, allerdings durch das gleichmäßige Muster des Elastomers der nächste wieder zur Verbindung beiträgt. Die Position der Pads auf Wafer und Platine müssen jedoch sehr exakt übereinstimmen. Die folgenden drei Punkte sollen abschätzen wie hoch die Genauigkeit sein muss.

x-Richtung In x-Richtung sind die Pads 2 mm breit. Eine Verschiebung in dieser Richtung ändert nur die Lage der Elastomere auf den Pads ohne die Kontaktfläche zu verkleinern und hat sonst keine weiteren Auswirkungen. Die Ansprüche an eine hohe Genauigkeit in x-Richtung sind somit nicht sehr hoch. Als Mindestanforderung sollte hier gelten, dass die gesamte Breite eines Elastomerverbinders genutzt wird, um die Pads miteinander zu verbinden. Bei einer Padbreite von 2 mm und eine Elastomerbreite von 1 mm im nicht gepressten und geschätzten maximalen 1,5 mm im gepressten Zustand, sollte die x-Verschiebung kleiner als 250 μm sein.

y-Richtung In y-Richtung haben die Pads eine Breite von 190 μm und einen Abstand von 100 μm . Der Abstand zwischen zwei Padpaaren sollte größer als 50 μm bleiben, da sonst ein Kurzschluss über einen leitenden Streifen des Elastomerverbinders möglich ist. Darüber hinaus sollten immer zwei Streifen zumindest teilweise an der Kontaktierung beteiligt sein. Um dies zu gewährleisten sollte die y-Verschiebung kleiner 50 μm sein. Die folgende Grafik 4.1 veranschaulicht diese Abschätzung.

Verdrehung Eine Verdrehung sollte möglichst ganz vermieden werden. Am ehesten macht sich eine Verdrehung zwischen zwei weit entfernten Retikeln bemerkbar. Unter der Annahme, dass die Achsen von Wafer und Hauptplatine zusammenfallen, kann man eine Abschätzung, wie in Abbildung 4.2 gezeigt, machen. Als Einschränkung wird hier angenommen, dass sich die Pads zumindest noch zur halben Breite, also 100 μm überdecken

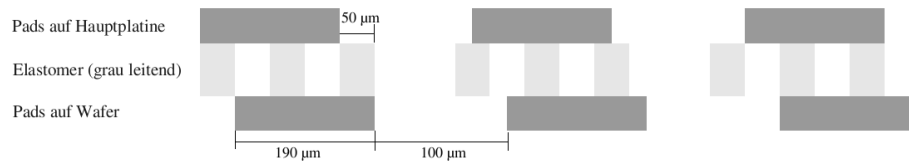


Abbildung 4.1: Die Grafik dient zur Abschätzung der maximal zulässigen y -Verschiebung. Bei der Situation ganz links ist nur ein Streifen des Elastomers an der Kontaktierung beteiligt. Alle gezeigten Varianten entsprechen einer Verschiebung um $50 \mu\text{m}$.

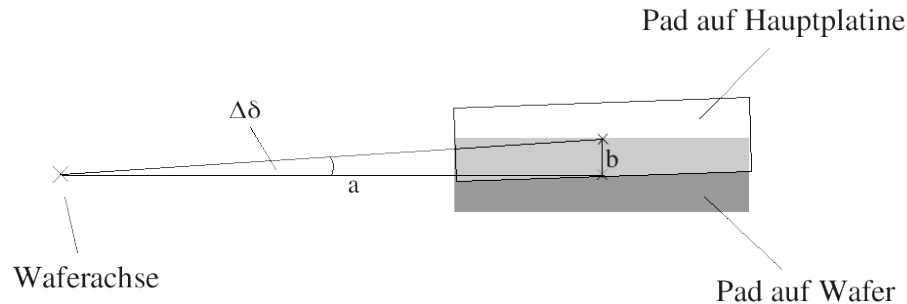


Abbildung 4.2: Grafik zur Abschätzung von $\Delta\delta$. b ist der Abstand der Padmitten. Er sollte kleiner $100 \mu\text{m}$ sein. a ist der Abstand des am weitesten entfernten Pads (92 mm).

müssen. Man erhält dann ein $\Delta\delta = 0,062^\circ$.

Der folgende Abschnitt stellt einen Aufbau vor, mit dem man die geforderten Genauigkeiten erzielen kann.

4.1 Justiervorrichtung

Um die zuvor genannten Genauigkeiten erreichen zu können, wurde das in Abbildung 4.3 zu sehende System entwickelt. In diesem System liegt der Waferträger samt Wafer auf der Grundplatte der Justagevorrichtung auf. Die Position des Waferträgers ist durch einen Stift auf der Grundplatte und einer Vertiefung im Waferträger gegeben. Platine, Deckel und Positionierungsschablone inklusive Elastomere werden zusammengeschaubt und auf den Waferträger gelegt. Diese Teile liegen auf 8 Federn auf, welche in den Waferträger eingelassen sind und dafür sorgen, dass die Elastomere vor dem Zusammenschaublen des Gesamtsystems über dem Wafer schweben. Zum Justieren kann man die x -, y -Richtung und Drehung um die z -Achse über Messschrauben¹ einstellen. Die Meßschrauben haben eine Ablesegenauigkeit von $0,002 \text{ mm}$. D. h., man kann die x - und y -Richtung über die Messschrauben auf $2 \mu\text{m}$ genau einstellen. Die Drehung um die z -Achse wird über eine Messschraube eingestellt, welche an einer am Waferträger befestigten Strebe angreift. Die Messschraube hat dabei einen Abstand von $296,5 \text{ mm}$ zum Drehpunkt des Waferträgers. Über die Genauigkeit der Messschraube und die Entfernung zum Drehpunkt kann man die Verdrehung mit einer Abweichung von $(0,39 \cdot 10^{-3})^\circ$ einstellen.

Nach dem Justieren über die drei Einstellmöglichkeiten können die einzelnen Komponenten vor dem Zusammenschaublen in der Justiervorrichtung fest eingespannt werden. Dies

¹Miniatur-Messschraube 148-142 von Mitutoyo. Skaleneinteilung $0,002 \text{ mm}$, Genauigkeit $2 \mu\text{m}$.

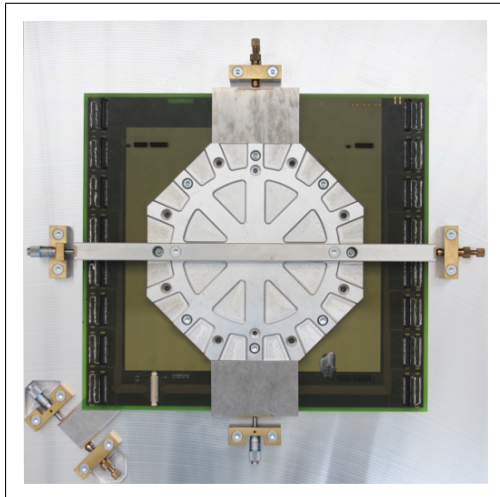


Abbildung 4.3: Foto zeigt das Prototypensystem innerhalb einer Justiervorrichtung. Die zusätzlich zu sehenden Platten bzw. Streben werden nur zur Justage an den Deckel bzw. den Waferträger montiert.

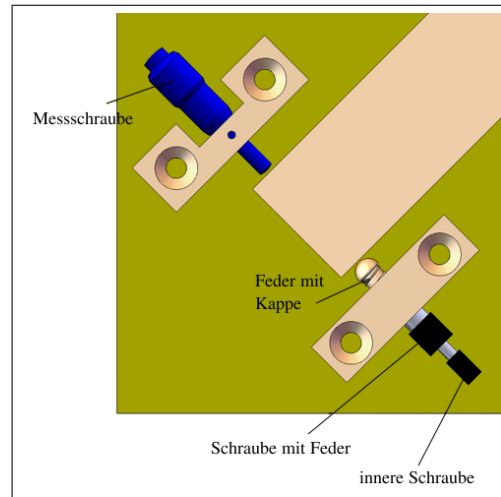


Abbildung 4.4: Das Bild zeigt das Prinzip zum Justieren und Einspannen der Verdrehung um die z-Achse. Das Prinzip für die Verschiebung in x- und y-Richtung funktioniert analog.

ist notwendig, um einer Dejustage beim Zusammenschrauben des Gesamtsystems vorzubeugen. Das hierzu verwendete Verfahren wird in Abbildung 4.4 gezeigt. Gegenüber der Messschraube befindet sich eine Schraube, auf deren Spitze eine Feder mit Kappe steckt. Diese Feder übt Druck aus und sorgt dafür, dass sich die zu justierende Richtung in beide Richtungen (plus und minus) sofort beim Drehen der Messschraube nachstellt. Die Schraube ist längs mit Innengewinde durchbohrt. In diesem Gewinde steckt eine zweite Schraube, welche zum Befestigen nach dem Justieren die Kappe der Feder fest an den zu verschiebenden Körper drückt und somit keine Bewegung in der entsprechenden Richtung (entlang der Schraubenachse) mehr zulässt. Nach dem Einspannen mit Hilfe der kleineren Schrauben kann das Gesamtsystem zusammenschraubt werden. Anschließend werden die kleinen Schrauben gelockert, das System aus der Justiervorrichtung entnommen und die Platten und Streben von Waferträger und Deckel abmontiert.

4.2 Optische Justage

Mit der entwickelten Justiervorrichtung kann man Wafer und Platine mit den gewünschten Genauigkeiten ausrichten. Das Problem, welches sich jetzt ergibt, besteht darin, dass man neue Strukturen relativ zu den vorhandenen auf Wafer und Platine benötigt, welche man dann zum Justieren benutzen kann. Die erste Möglichkeit besteht darin, zwei Muster zu wählen, deren relative Lage man während der Justage optisch kontrolliert und somit zum Ausrichten von Wafer und Platine verwendet.

Dazu wird auf dem Testwafer nachträglich ein Muster per Laser aufgebracht. Abbildung 4.5 zeigt die Form und Dimension des Musters auf dem Wafer. Dieses Muster ist mit einer Abweichung von weniger als $2 \mu\text{m}$ relativ zur Postprozessierung plaziert. Das Muster wird zweimal aufgebracht. Einmal im zuvor definierten Norden und einmal im Süden. Beide Stellen können dann durch Löcher in Deckel, Platine und Positionierungsmaske gesehen werden.

Als Gegenmuster werden zwei senkrecht zueinander liegende Drähte auf der Unterseite der Hauptplatine quer über die Löcher gelötet. Die gelöteten Drähte haben eine Dicke

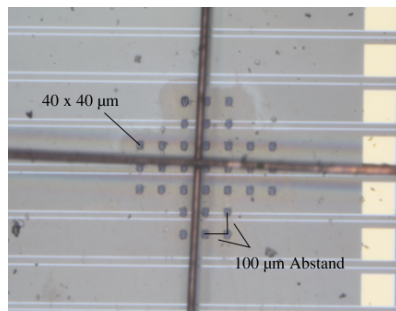


Abbildung 4.5: Foto zeigt die Ausrichtung von Vergleichsmuster auf Wafer zu gelöteten Drähten auf der Platine.

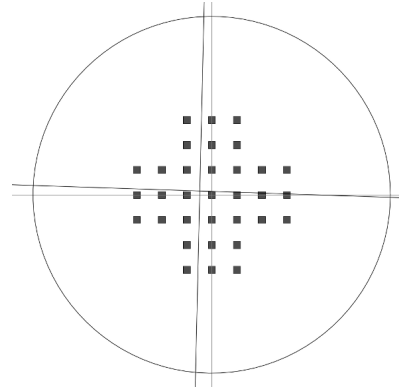


Abbildung 4.6: Grafik zeigt ein Beispiel für die Lage von optimal und tatsächlich gelöteten Drähten bei richtiger Justage.

von ungefähr $50 \mu\text{m}$. Die Lage der Lötstellen wird durch kleine Pads bestimmt. Zusätzlich wird nach dem Lötten die Lage der Lötstellen auf $10 \mu\text{m}$ genau relativ zum Muster auf der Platine ausgemessen. Die Verschiebung von der idealen Lage wird in einer kleinen Grafik dargestellt und zur Justage verwendet. Ein Beispiel für solch eine grafische Darstellung ist in Abbildung 4.6 zu sehen.

Die Genauigkeiten, welche man mit dieser Methode erreichen kann sind schwer abzuschätzen. Es ist zum einen schwierig die genaue Drahtstärke anzugeben, da die Dicke des gelöteten Drahtes nicht genau bekannt ist. Zum anderen ist die Form der gelaserten Quadrate nicht ganz quadratisch und teilweise kleiner als 40 auf $40 \mu\text{m}$. Außerdem liegen die beiden Markierungen beim Justieren nicht in der selben Ebene, sondern sind im besten Fall etwas mehr als einen Millimeter voneinander getrennt. Somit ist es schwierig bei ausreichender Vergrößerung sowohl das Muster auf dem Wafer, als auch die Drähte auf der Platine scharf zu stellen. Auf der anderen Seite kann man jedoch mit dem Auge die Lage der Drähte gut einschätzen und somit auch genauer als eine Drahtdicke von $50 \mu\text{m}$ justieren. Somit sollten die Verschiebungen in x - und y -Richtung gut einzustellen sein.

Für die Verdrehung muss man umrechnen, welche Verschiebung in Millimetern einer Verdrehung in Grad entsprechen würde. Ähnlich wie in vorherigen Rechnungen wird dabei auch wieder eine einfache trigonometrische Beziehung angewendet. Hierbei wird jetzt angenommen, dass der Wafer um $(60,0 \cdot 10^{-3})^\circ$ gedreht wird. Der Mittelpunkt des Musters auf dem Wafer hat einen Abstand von $93,25 \text{ mm}$ zur Drehachse des Wafers. Dies würde einer Verschiebung des Musters in x -Richtung von $98 \mu\text{m}$ entsprechen. Das bedeutet, dass man mit einer Genauigkeit von geschätzten $(30,0 \cdot 10^{-3})^\circ$ die Verdrehung einstellen kann. Allerdings muss man hierzu zwei sich gegenüber liegende Stellen zur optischen Justage verwenden, da die z -Achsen von Wafer und Hauptplatine nicht zwingend deckungsgleich sein müssen.

Abschließend kann man festhalten, dass sich die optische Justage in der vorgestellten Form eignet um Wafer und Platine mit den gegebenen Padgrößen und Abständen zu justieren.

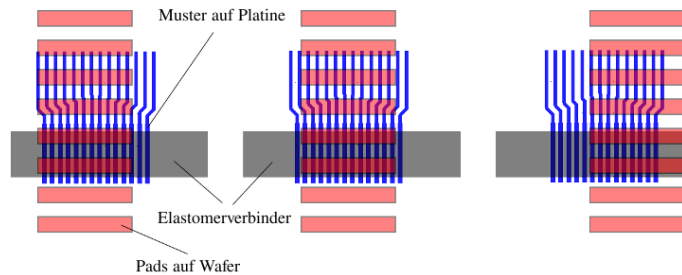
4.3 Elektrische Justage

Als zweite Methode die relative Lage zwischen Wafer und Platine zu bestimmen, kann die elektrische Justage verwendet werden. Auch hier werden wieder zwei Muster verglichen um die Verschiebung zu bestimmen. Als Nachteil sei hier gleich erwähnt, dass man zum

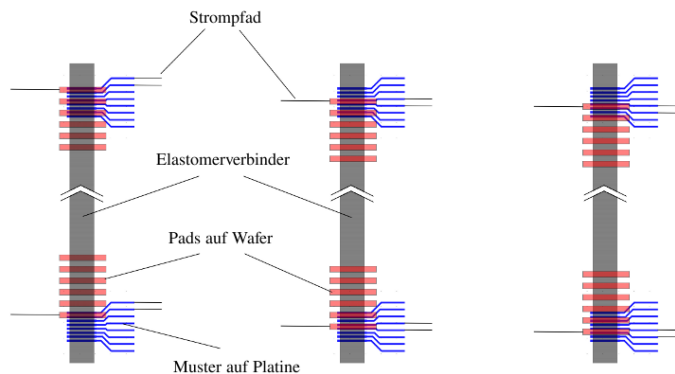
Vergleichen der Lagen das System zusammenschrauben muss. Ein Einstellen der korrekten Position würde im allgemeinen bedeuten, dass bei ausreichender Vorjustage das System zusammen geschraubt, gelockert und wieder auseinandergeschraubt werden müsste. Dies ist zum einen sehr zeitaufwendig und würde zum anderen einen erhöhten Verschleiß und mehr Fehlerquellen bedeuten. Die elektrische Justage soll daher nur zur Überprüfung des optischen Verfahrens dienen.

Prinzipiell müssen zwei Muster gewählt werden, deren Verschiebung zueinander von außen durch einen Stromfluß leicht zu messen ist. Einfach wäre es dabei natürlich zwischen zwei Zuständen zu unterscheiden: leitend oder nicht leitend. Es sollte also später eine Kombination von Pins bestimmt werden, die miteinander verbunden sind oder eben nicht, und aus der man dann die Verschiebung zwischen Testwafer und Platine bestimmen kann. Es gibt viele Möglichkeiten für ein Muster, welches sich in unserem Fall eignen würde.

Als die Idee für ein zweites Justageverfahren aufkam, lag die Struktur der Postprozessierung schon fest. Es wurde daher ein Muster entwickelt, welches als Gegenmuster auf der Platine verwendet werden konnte. Abbildung 4.7 zeigt das Prinzip der elektrischen Justage. Das Prinzip des Musters besteht darin, dass auf der Platine mehrere kleine Leitungen



(a) Muster zur Bestimmung der Verschiebung in x-Richtung.



(b) Muster zur Bestimmung der Verschiebung in y-Richtung.

Abbildung 4.7: Grafik zeigt das Prinzip der elektrischen Justage auf. Es werden jeweils drei Varianten für die x- und y-Richtung aufgezeigt. Die mittlere gibt je die optimale Lage an.

nebeneinander liegen, welche beim Zusammenschrauben des Systems über ein Elastomer und einem Pad auf dem Wafer elektrisch verbunden werden. Durch Messen, welche Linien mit welchen verbunden sind, kann man auf die relative Lage von Platine zu Wafer zurückschließen.

Die Breite der Linien beträgt genau wie der Raum zwischen den Linien $75 \mu\text{m}$. Somit

kann man hier eine Genauigkeit in x- und y-Richtung von 75 μm erreichen. Höhere Genauigkeiten durch schmalere Leitungen sind bis zu einer Genauigkeit von 50 μm möglich. Besser als die optische Justage kann man mit dieser Methode der elektrischen Justage allerdings nicht werden, da die Breite und der Abstand von Kontaktierstreifen eines Elastomers 25 μm beträgt.

Die Verschiebung in x- und y-Richtung wird je viermal auf dem Wafer gemessen. Dazu wurden 4 Retikel verwendet, jeweils das erste und letzte der obersten und untersten Reihe. Die y-Verschiebung wird je zweimal im ersten Retikel der obersten Retikelreihe und im letzten der unteren bestimmt. Selbes gilt für die x-Verschiebung auf den anderen beiden Retikeln. In Abbildung 5.2 aus Kapitel 5 sind die entsprechenden Retikel in einem Schema des Wafers eingezeichnet. Die betreffenden Retikel haben einen Abstand zum Mittelpunkt von 80,5 mm. Mit Hilfe einer ähnlichen Abschätzung wie im vorherigen Kapitel erhält man daraus eine Genauigkeit gegen Verdrehung von $(35,0 \cdot 10^{-3})^\circ$.

4.4 Einstellung des Abstandes zwischen Wafer und Platine

Die Elastomere werden ab einer bestimmten Kompression in z-Richtung leitend. Laut Hersteller wirken sie ab einer Kompression auf 95 % der Originalhöhe als elektrischer Leiter und sollten nicht auf weniger als 70 % gepresst werden. Diesen recht weiten Bereich² kann man benutzen um Unebenheiten durch verschieden hohe Pads und unterschiedlicher Wafer- und Platinendicke auszugleichen. Um diesen Bereich optimal nutzen zu können muss man einen Abstand einstellen, welcher Abweichungen nach oben und unten zulässt. Ein Abstand von 80 % der Ausgangshöhe der Elastomere erfüllt diese Aufgabe. Mit diesem Wert ist für größere Abstände mehr Spielraum als für kleinere. Es ist allerdings auch sinnvoll die Ausgangshöhe so zu wählen, da eine Abweichung von dem vom Hersteller angegebenen Bereich zu geringeren Kompressionen als auf 95 % zu keinem Kontakt führt. Eine Abweichung hin zu größeren Kompressionen unter 75 % führt hingegen nur zur einer Verschlechterung der Elastomereigenschaften. So erhöht sich zum Beispiel der Widerstand der Elastomere um einige Prozent und die Wahrscheinlichkeit für Kurzschlüsse nimmt zu. Neben dem Einstellen eines optimalen Abstands von möglichst genau 80 % der Originalhöhe der Elastomere, sollte auch die Möglichkeit bestehen Wafer unterschiedlicher Höhen in das System einzubauen und somit eine erhöhte Variabilität zu gewährleisten. Um diese Anforderung umsetzen zu können, wurde ein System konzipiert, indem speziell gedrehte Unterlegscheiben verwendet werden, um den gewünschten Abstand einzustellen. In diesem Konzept werden in den Waferträger runde Vertiefungen eingefräst. Diese Vertiefungen sind axialsymmetrisch mit den Gewindebohrungen für die Schrauben, welche das System zusammen halten. Abbildung 4.8 zeigt eine Zeichnung von Vertiefung und Schraubenlöchern. Die Kombination aus Höhe der Vertiefung und Dicke der Scheiben stellt den Abstand zwischen Wafer und Platine ein. Sehr dünne Scheiben sind schwer zu fertigen und weisen höhere Ungenauigkeit auf. Dickere Scheiben bedeuten, dass die Vertiefungen höher sein müssen. Durch zu tiefe Ausfräsungen wird die Länge der Gewinde verkürzt, welche die Befestigungsschrauben des Systems halten. Das Gewinde sollte jedoch nicht unnötig verkürzt werden. Als Mittelweg wurden für die Vertiefungen eine Höhe von 1,00 mm gewählt. Die Höhe der Unterlegscheiben wird je nach Waferdicke angepasst. Für sie ergibt sich folgende Beziehung:

$$d_d = d_w - d_{wt} + h_{optimal} + d_{dt} \quad \text{mit } h_{optimal} = 0,80 \text{ mm} \quad (4.1)$$

²Die verwendeten Elastomere haben eine Höhe von 1,0 mm. Bei einem Arbeitsbereich von 70 bis 95 % der originalen Höhe können sie Unterschiede im Abstand zwischen Wafer zu Hauptplatine von 250 μm ausgleichen.

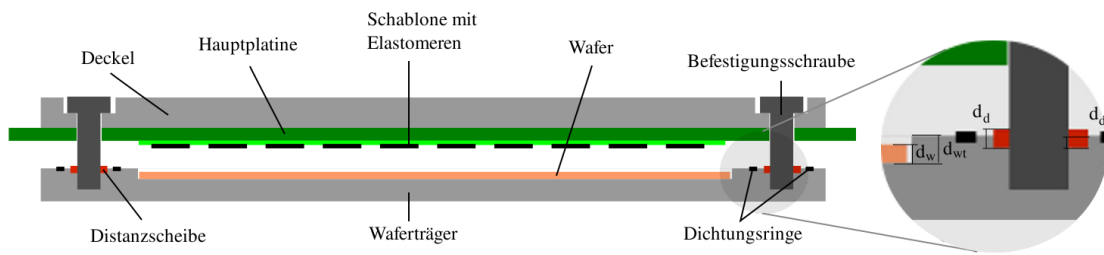


Abbildung 4.8: Die Grafik zeigt ein Schema von einem Querschnitt des Gesamtsystems. An den Löchern für die Schrauben, die zum Zusammenhalt des Gesamtsystems dienen, kann man Vertiefungen erkennen. In diesen Vertiefungen liegen Unterlegscheiben mit einer bestimmten Höhe. Mit dieser Höhe wird der Abstand zwischen Wafer und Hauptplatine eingestellt. $d_w \hat{=}$ Waferdicke; $d_{wt} \hat{=}$ Höhe der Wafervertiefung; $d_d \hat{=}$ Dicke der Distanzscheibe; $d_{dt} \hat{=}$ Höhe der Vertiefung für die Distanzscheibe

Außer d_w können alle Höhen mit einer Genauigkeit von 0,01 mm gefertigt werden. Aus der Kombination der eben vorgestellten Höhen darf der Wafer nicht dünner als 0,65 mm sein. Höhen über 0,85 mm sind auch nicht gut, da dann die Dichtungsringe zwischen Platine und Waferträger durch höhere ersetzt werden müssten.

4.5 Zusammenfassung

Die folgende Tabelle fasst noch einmal alle Genauigkeiten zusammen und gibt einen Überblick über geforderte und einhaltbare Toleranzen.

Aus der Tabelle erkennt man, dass sowohl die elektrische, als auch die optische Methode

	Δx in μm	Δy in μm	$\Delta\delta$ in $(10^{-3})^\circ$
geforderte Toleranz	250	50	60
Genauigkeit der Justiervorrichtung	2	2	0,39
Genauigkeit der optischen Methode	50	50	30
Genauigkeit der elektrischen Methode	75	75	35

zur Justierung verwendet werden kann. Darüber hinaus weist die Justiervorrichtung noch Potential auf, um genauer zu Justieren.

Mit einer ausgeklügelteren Technik bei der elektrischen Justage könnte man vielleicht auch hier noch höhere Genauigkeiten erzielen. Vor dem Hintergrund, dass die elektrische Justage nur zur Kontrolle verwendet werden sollte und die optische Justage nicht viel besser ist, ist es jedoch nicht notwendig die elektrische Justage zu verbessern.

Auch die optische Justage kann noch verbessert werden. Es gibt Überlegungen anstatt zwei sich kreuzenden Drähten drei Drähte nebeneinander zu löten, welche von drei weiteren gekreuzt werden. Zusätzlich soll das Muster auf dem Wafer erweitert werden. Man gewinnt dadurch die Möglichkeit leichter innerhalb der obigen Toleranzen zu justieren. Durch Verwendung feinerer Drähte und einem angepassten Gegenmuster kann die Genauigkeit weiter erhöht werden.

Die Variation in z-Richtung ist eine Eigenschaft, welche gut genutzt werden kann. Bisher wurde diese Methode nicht eingebaut. Man sollte das System aber entsprechend nachrüsten, um möglichst variabel zu bleiben und auf unvorhergesehene Änderungen reagieren zu

können. Nachfragen bei UMC ergaben allerdings auch die Erkenntnis, dass die Dicke von nicht extra gedünnten Wafern $(29 \pm 0,4)$ mil bzw. $(0,737 \pm 0,010)$ mm beträgt.

Kapitel 5

Aufbau der Hauptplatine

Auf dem Wafer im fertigen Wafer Scale Integration System werden verschiedene Schaltungen realisiert. Diese Schaltungen benötigen neben einer Spannungsversorgung Ein- und Ausgänge für Signale, welche sie erhalten, verarbeiten und wieder ausgeben. Die Hauptplatine ist dafür verantwortlich die entsprechenden Spannungen bereitzustellen und die Signale richtig zu verteilen bzw. weiter zu leiten.

Im Prototypensystem wird eine Platine verwendet, welche die 120 Kontaktflächen pro Retikel zu Steckern nach außen führen soll. Das Design der Platine sollte dabei alle Probleme, welche bei der späteren Hauptplatine auftreten könnten abdecken. Die folgenden Abschnitte umfassen die Anforderungen an die Platine, die Realisierbarkeit der Platine in Zusammenarbeit mit Würth Elektronik¹ und eine kurze Beschreibung der Platine, wie sie im Prototypensystem verwendet werden.

5.1 Anforderungen an die Testplatine

Wie eben schon erwähnt, besteht die Hauptaufgabe der Platine darin, die Pads der Retikel zu Steckern auf der Platine zu führen. Die erste Aufgabe, welche sich dabei stellt, ist ein Muster, welches mit der Postprozessierung des Wafer übereinstimmt, zu erstellen. Dies sollte kein Problem sein, da Platinen im Format 40 mm auf 36,6 mm mit Strukturgrößen von 75 μm gefertigt werden können. Eine andere Aufgabe besteht darin die Pads auf der Unterseite der Platine an den Rand zu routen. Hierbei sollte man sich nicht darauf beschränken alle Stecker auf der Unterseite zu plazieren. Unter der Voraussetzung, dass man einen Schichtaufbau entwirft, bei welchem abwechselnd auf einer Lage der sogenannten Signallagen die Pads geroutet werden und eine Lage als Powerlage definiert wird, würde man eine Platine mit 24 Lagen erhalten. Die Powerlagen sind zum einen zum Verteilen der verschiedenen Spannungen vorgesehen und dienen zum anderen als Referenz für die impedanzkontrollierten differentiellen Leitungen.

Die Dicke der Metallagen, also dem leitenden Teil der Lagen, beträgt 35 μm . Bei mehreren Lagen und ungeschicktem Routing könnte es passieren, dass die Dicke der Platine um einige Zehntel Millimeter variiert. Dies sollte allerdings vermieden werden, da die Platine die Elastomere auf den Wafer drückt und die Elastomere nach dem Anpressen von 1 mm Höhe auf 0,65 bis 0,8 mm Höhe gepresst werden sollten. Weniger Anpressung bedeuten kein Kontakt. Mehr Anpressdruck kann zu Kurzschlüssen und Beschädigung des Wafers führen. Um eine gleichmäßige Dicke zu gewährleisten muss das Routing der Platine also möglichst symmetrisch und gleichverteilt erfolgen.

¹Würth Elektronik GmbH & Co. KG, Werk Rot am See

5.2 Realisierbarkeit der Platine

Ein Board mit 24 Lagen ist zwar prinzipiell möglich, allerdings nicht in unserem Fall. Das liegt am Verfahren zur Herstellung von Viellagenplatten. Dabei werden zwei Lagen um einen Kern gepresst und belichtet. Dieser Vorgang wird für mehrere Lagenpaare wiederholt. Zum Schluss werden die entstandenen 2-Lagen-Platinen in einem weiteren Schritt zusammengepresst, wodurch relativ dicke Mehrkernplatten entstehen.

Für die verwendete Platine ist es wichtig, die oberste mit der untersten Lage verbinden zu können.

Es gibt prinzipiell vier Möglichkeiten verschiedene Lagen mit einander zu verbinden. Durch die Verwendung von Buried Vias, Blind Vias, Stacked Vias oder durchbohrten Vias. Abbildung 5.1 zeigt die verschiedenen Verbindungsmöglichkeiten. Bei allen Möglichkeiten muß

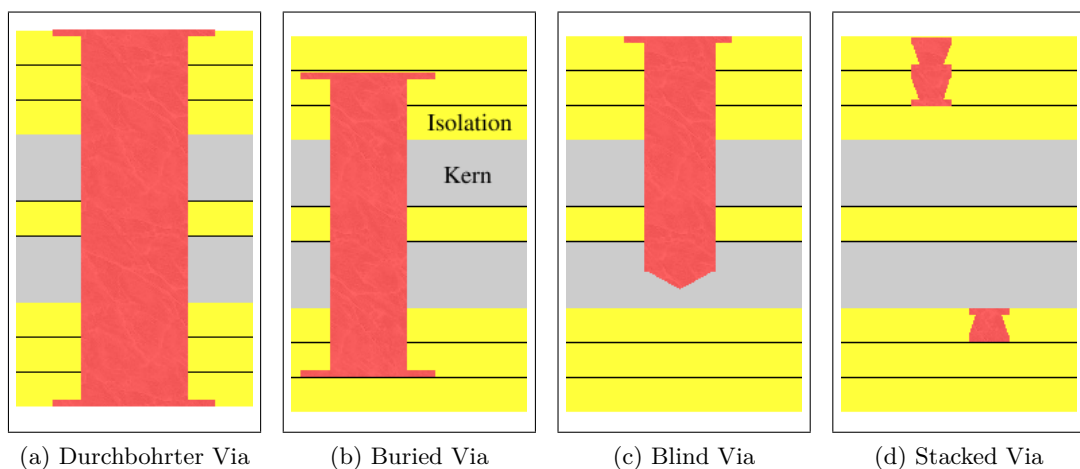


Abbildung 5.1: Bilder der vier verschiedenen Möglichkeiten um verschiedene Lagen einer Mehrlagenplatine elektrisch miteinander zu verbinden. Es können alle Lagen, durch die ein Via geht an das Via angeschlossen werden. Die elektrisch leitenden Lagen befinden sich zwischen zwei Isolierungen, zwischen Isolierung und Kern und auf der oberen und unteren Seite.

das sogenannte Aspekt-Ratio eingehalten werden. Dieses bildet das Verhältnis zwischen Durchmesser und Tiefe eines Vias. Das Aspekt-Ratio darf laut Würth Elektronik einen Wert von 1,25 nicht unterschreiten. Tiefe Vias haben nach diesem Verhältnis einen größeren Durchmesser als flachere Vias. Für die Platine sind Vias, welche die unterste mit der obersten Signallage verbinden und dabei möglichst wenig Platz benötigen sehr wichtig. Die Pads, welche in der obersten Signallage verlegt werden, müssen durch Vias direkt nach oben geführt werden. Da die Verbindungsdichte an dieser Stelle der Platine (direkt über dem Postprozessierungsmuster) sehr hoch ist, müssen die Vias sehr klein sein. Zusätzlich dürfen nicht alle Vias durch die gesamte Platinen gehen, da sonst auf allen Lagen Platz für das Verlegen von Leitungen verloren gehen würde. Als Alternative zur Herstellung von viellagigen Platinen gibt es die Möglichkeit um einen Kern zwei Lagen zu pressen und anschließend um diese 2-lagige Platine zwei neue Lagen. Es ist somit möglich einen Stacked Via zu erzeugen, welcher die zwei weit entfernten Lagen miteinander verbindet. Der Nachteil dieser Methode ist, dass die einzelnen Teile der Platine jetzt nicht 2mal gepresst werden, sondern bei n Lagen $\frac{n}{2}$ -mal. Das Problem dabei ist, dass das verwendete Material bei jedem Pressen spröder wird und ab einer bestimmten Anzahl an Pressvorgängen kaputt geht. In Absprache mit Würth Elektronik wurden 3 Pressungen für die

Testplatine vereinbart mit der Option von 6 Verpressungen für die endgültige Platine.
Drei Verpressungen bedeutet, dass zum Routen der Signalleitungen 3 Lagen zur Verfü-

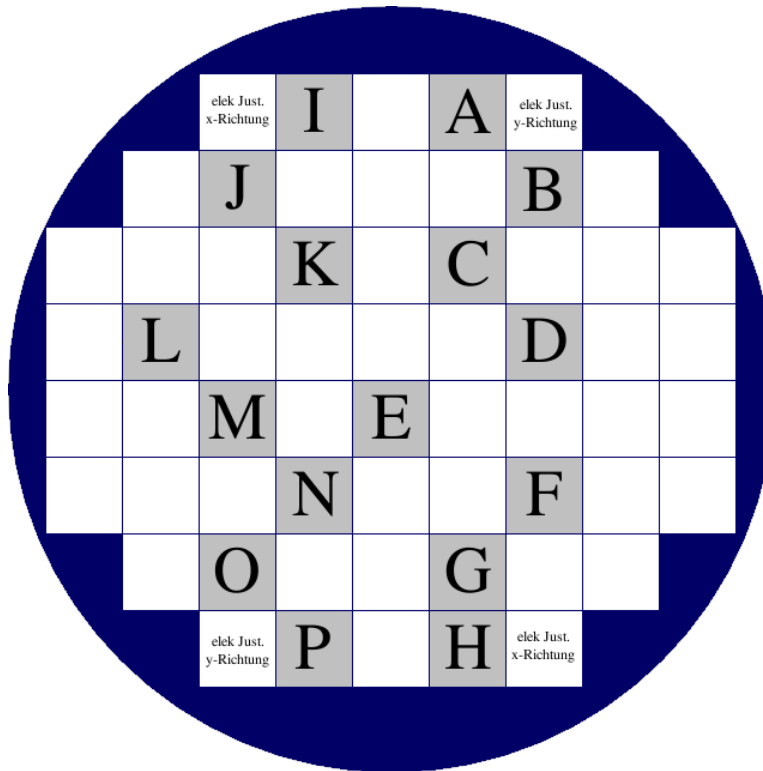


Abbildung 5.2: Bild zeigt die Auswahl der Retikel, welche auf der Testplatine nach außen zu Steckern geroutet werden.

gung stehen. Dies bedeutet als Einschränkung, dass nicht alle 60 sondern nur 16 Retikel nach außen geroutet werden können. Abbildung 5.2 zeigt die Verteilung der nach außen geführten Retikel. Zusätzlich zu den gezeigten Retikeln wurden jeweils noch die ersten beiden Leitungen jedes dieser Retikel zu einem extra Stecker geführt. An diesen Stecker wird später eine Messplatine angeschlossen, mit der man die relative Lage zwischen Wafer und Platine mit Hilfe einer Messung qualitativ abschätzen kann. Es wurden auch noch einige Leitungen für die elektrische Justage verlegt. Eine Übersicht des fertigen Designs wird im nächsten Abschnitt gezeigt.

5.3 Design der Testplatine

Das Design der fertigen Testplatine ist in Abbildung 5.3 und Abbildung 5.4 zu sehen. Die eine Abbildung zeigt die Oberseite der Platine, die andere die Unterseite. Das Routing über die zwei Signallagen im Inneren wird nicht gezeigt. Die Verteilung der Leitungen ist allerdings so gewählt, dass alle Lagen gleichmäßig über die Fläche für eine konstante Leiterplattendicke sorgen.

Wie oben schon erwähnt, werden 16 Retikel an Stecker geroutet. Ein Stecker hat 60 Kontakte. Ein Retikel hat 2 Pinreihen zu je 60 Pins. Das bedeutet, dass die Signale eines Retikels an zwei Stecker geführt werden. In Abbildung 5.3 sieht man in der Mitte eine Kupferfläche mit Durchgangslöchern zur Befestigung der Platine an den Deckel, der Positionierungsmaske an die Platine und für die Schrauben, welche das System zusam-

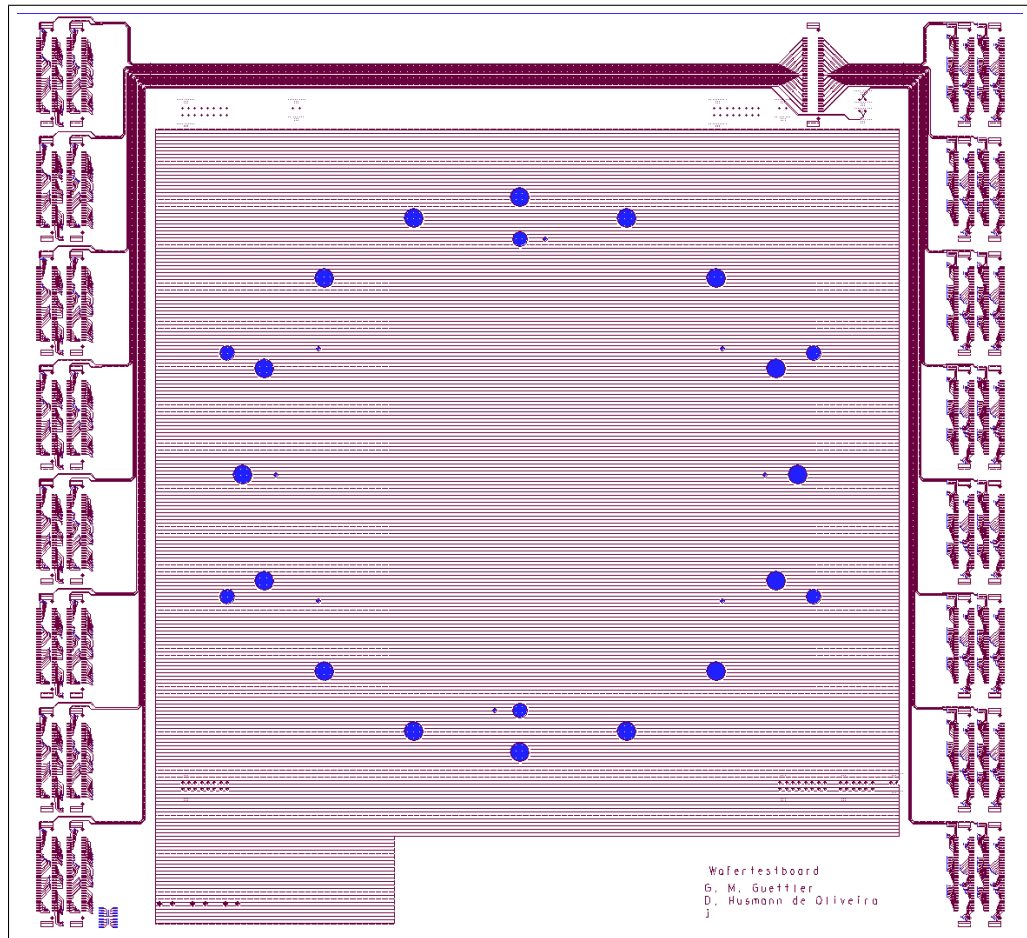


Abbildung 5.3: Bild zeigt das Layout der Oberseite der Testplatine.

menpressen. Links und rechts neben dieser Fläche sind die Footprints² der Stecker für die jeweiligen Retikel zu sehen. Oben rechts sind zwei Stecker nebeneinander, welche zu Retikel A gehören. Vergleiche hierzu Abbildung 5.2. Darunter folgen die zwei Stecker für die Retikel B, anschließend für Retikel C usw. Oben links werden die Stecker für Retikel I platziert. Gefolgt von den Steckern der Retikel J bis P. Es ist außerdem zu sehen, dass jeweils die ersten beiden Pins eines jeden Steckers an einen extra Stecker geführt werden. Dies ermöglicht das Messen einer Flächenverteilung in einem Messvorgang. Das Prinzip dieser Messung und der Messung an einem Retikel wird im nächsten Kapitel detailliert erklärt.

Die Unterseite der Testplatine ist in Abb. 5.4 zu sehen. Man erkennt passend zur Oberseite wieder die Durchgangslöcher zum Zusammenschrauben der einzelnen Komponenten. In der Mitte liegt das Muster, welches deckungsgleich zum Postprozessierungsmuster auf dem Testwafer ist. Abweichungen von diesem Muster sind auf 4 Retikeln in der Mitte und an 4 Retikeln an den Ecken zu sehen. Die Muster an den Ecken sind für die elektrische Justage vorgesehen. Die entsprechenden Leitungen wurden auf der Unterseite an außenliegende Pins verlegt. Die 4 Muster in der Mitte erfüllen keinen besonderen Zweck. Diese wurden beim Designen der Retikel für die elektrische Justage aus Symmetriegründen mit-

²Footprint bezeichnet den Teil des Platinenlayouts, welcher für alle Stecker gleich ist. In dem Footprint für die Stecker befinden sich die Kontaktflächen, auf die die Steckerkontakte gelötet werden. Zusätzlich werden die Kontakte so verteilt, dass sie an einer Seite parallel und mit in einem gleichmäßigen Abstand nebeneinander liegen.

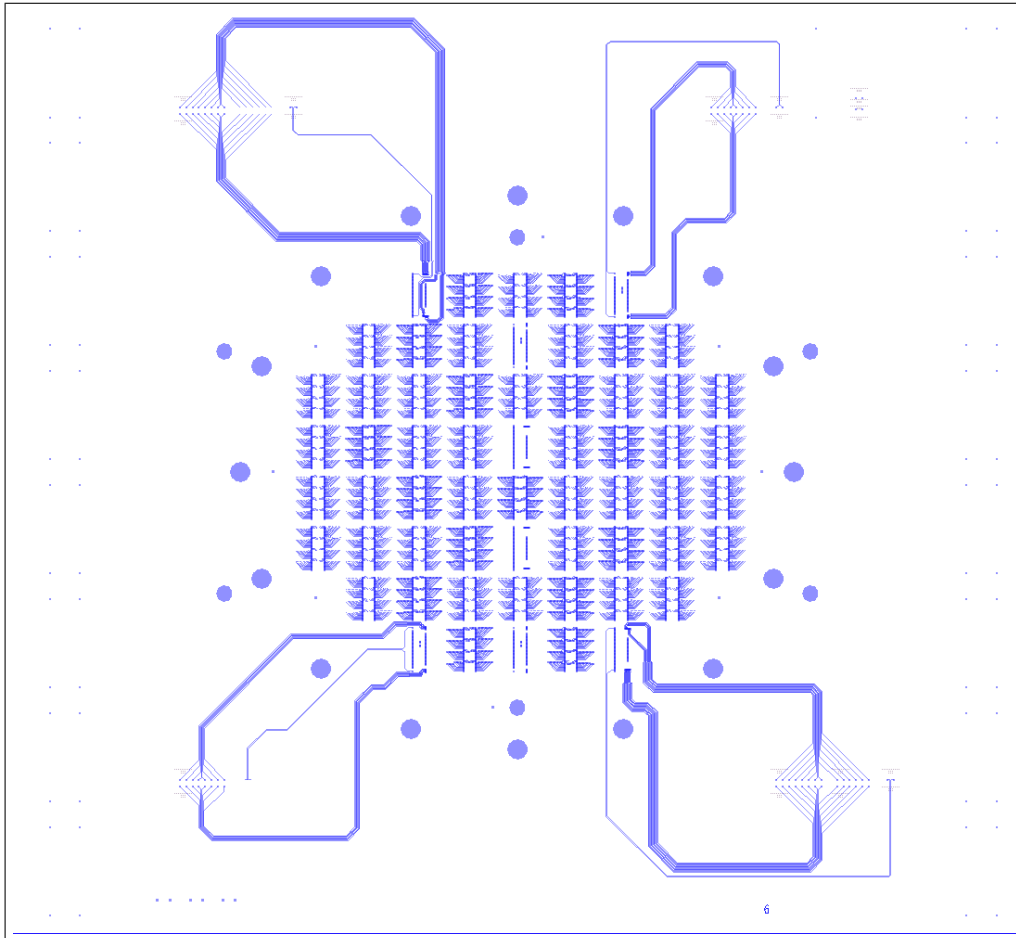


Abbildung 5.4: Bild zeigt das Layout der Unterseite der Testplatine.

verändert. Die Löcher auf der linken und rechten Seite gehen durch die gesamte Platine und dienen als Positionierungshilfen für die Stecker. Das Muster auf der Unterseite wurde nach dem Fertigen der Platine noch einmal ausgemessen und wies auf der gesamten Fläche keine Abweichungen von mehr als $10\ \mu\text{m}$ auf.

Kapitel 6

Messungen der Elastomere

Eine der wichtigsten Aufgaben des Testsystems besteht darin, die Zuverlässigkeit der vertikalen Verbindungen zwischen Wafer und Hauptplatine zu testen. Die vertikalen Verbindungen werden über Elastomere, welche unter Druck von oben elektrisch leitend werden, realisiert. Das Testsystem dient dazu herauszufinden, welche Verbindungsdichten man mit Hilfe dieser Elastomere erreichen kann. Dabei ist darauf zu achten, dass möglichst keine Ausfälle aufgrund von Nicht-Kontaktierungen oder Kurzschlüssen auftreten. In den folgenden Unterkapiteln werden die technischen Daten der Elastomere, die Messmethode zum Test der vertikalen Verbindungen und die Ergebnisse der Messungen vorgestellt und diskutiert.

6.1 Technische Daten

Die verwendeten Elastomere wurden bei Nucletron Technologies GmbH bestellt. Es handelt sich um ZEBRA[®] High Performance Silver Connectors der Firma Fujipoly[®]. Die Elastomerverbinder sind in verschiedenen Größen erhältlich. In Abbildung 2.3b ist eine Zeichnung eines Elastomers zu sehen. Tabelle 6.1 zeigt eine Auflistung der verschiedenen Größen, in denen die Elastomerverbinder bestellt werden können.

GRÖSSE	GRENZEN	TOLERANZEN	GEWÄHLT FÜR TESTSYSTEM
Dicke Isolation	0,050 mm	$\pm 0,025$ mm	
$T_i = T_c$	0,050 mm		
$T_i + T_c$	0,100 mm	+ 0,052 mm	
L	6,35 bis 127,00 mm	$\pm (0,12 \text{ bis } 0,64)$ mm	$(18,00 \pm 0,20)$ mm
H	1,00 bis 12,70 mm	$\pm (0,08 \text{ bis } 0,18)$ mm	$(1,00 \pm 0,08)$ mm
W	0,50 bis 2,54 mm	$\pm (0,08 \text{ bis } 0,13)$ mm	$(1,00 \pm 0,08)$ mm

Tabelle 6.1: Tabelle listet die möglichen Größen auf, in denen man die Elastomerverbinder bestellen kann.

Die Varianz der Breite W kann darin zu einem Problem werden. Die Ausfräsungen in der Positionierungsmaske wurden so geformt, dass ein Elastomer bei der Kompression noch Raum hat, um sich gleichmäßig zu verformen. Würde man die Form zu schmal wählen, könnten sich die Elastomere beim Pressen zu stark verformen und dadurch Kurzschlüsse erzeugen. Aus diesem Grund sitzen die Elastomere nicht sehr fest in der Positionierungs-

maske. Dies kann beim Zusammenbau dazu führen, dass einzelne Elastomere aus der Maske fallen. Das wiederum sollte absolut vermieden werden. Die Wahl der richtigen Größe für die Aussparungen der Positionierungsmaske ist schwierig. Es wurden 16 verschiedenen große Aussparungen getestet und drei Positionierungsmasken gefertigt. Bei der besten Variante vielen von den 456 Elastomeren einzelne bei der Justage aus der Schablone. Ein neue Maske mit kleineren Ausfräsungen wird derzeit hergestellt und konnte noch nicht getestet werden.

Der Widerstand eines Elastomerverbinders wird vom Hersteller durch folgende Gleichung gegeben.

$$R = \frac{H * 0,01 \Omega * m}{W * E_w} + 0.10 \Omega$$

E_w ist darin die Breite des schmaleren von zwei Pads, welche das Elastomere verbinden soll. Diese Breite beträgt bei uns $E_w = 190 \mu\text{m}$. Daraus folgt für die Elastomere im verwendeten Testsystem ein Elastomerwiderstand pro Pad von $R = 108 \text{ m}\Omega$.

Der Maximalstrom durch ein Elastomer ist vom Hersteller mit einer Stromdichte von $6,78 * 10^5 \frac{\text{A}}{\text{m}^2}$ angegeben. Gerechnet auf einen Kontaktstreifen eines Elastomers im Testsystem ergibt dies $I_{max} = 33,9 \text{ mA}$.

Die Angaben für Widerstand und Maximalstrom sind für Elastomere gemacht, welche im eingebauten Zustand eine Höhe zwischen 75 % bis 95 % der Originalhöhe haben. Bei niedrigerer Kompression steigt der Widerstand schnell an, bis der Elastomer nicht mehr kontaktiert. Bei höherer Kompression steigt der Widerstand auch an. In welchem Maße er dies tut wird vom Hersteller nicht genau spezifiziert. Um den Elastomer auf 80 % zu komprimieren, ist eine bestimmte Kraft notwendig. Diese Kraft berechnet sich aus:

$$F = \frac{H - H_1}{H} * W * L * \beta \quad \text{mit } \beta = 9,8 * 10^6 \frac{\text{kg}}{\text{m} * \text{s}^2}$$

H_1 ist darin die Höhe des Elastomers im komprimierten Zustand. Nach Einsetzen der entsprechenden Werte erhält man für die Kompression auf 80 % der Originalhöhe eine Kraft $F = 35,5 \text{ N}$. Die 16 Schrauben, die das Gesamtsystem zusammen pressen, müssen bei der Verwendung von etwa 350 Elastomeren den Deckel mit einer Kraft von rund 12,5 kN gegen den Waferträger pressen.

6.2 Messmethode

Um die für das Prototypensystem entscheidenden Eigenschaften des Elastomeres richtig bestimmen zu können, braucht man drei Tests. Im ersten Test muss man die Zuverlässigkeit der Kontaktierung bei gegebenem Padabstand und gegebener Kompression ermitteln. Hierzu wird der Kontaktwiderstand der Elastomere an unterschiedlichen Stellen auf dem Wafer und nach unterschiedlichen Justiervorgängen gemessen. Darüber hinaus werden die Messungen bei verschiedenen Temperaturen durchgeführt. Anschließend werden die Messungen miteinander verglichen und eine Aussage über die Zuverlässigkeit getroffen.

Neben dieser Zuverlässigkeitsmessung muss verifiziert werden, wie sich bei Kompressionen auf weniger als 75 % der Originalhöhe die Leitfähigkeit der Elastomere verhält. Außerdem sollte auch bestimmt werden, was bei Strömen über dem Maximalstrom passiert. Hierbei ist interessant, welche Effekte bei Stromspitzen und bei der Dauerbelastung mit Strömen größer dem Maximalstrom auftreten.

Für eine statistische Aussage muss eine große Zahl an Widerständen gemessen werden. Man benötigt daher eine automatisierte Messmethode. Dies lässt sich relativ einfach unter Verwendung von Multiplexern, welche die entsprechenden Widerstände an die Messelektronik schalten, realisieren. Die Messgenauigkeit, mit der die Widerstände bestimmt

werden, muss hoch genug sein, um den Widerstand der Elastomere auflösen zu können. Ein Problem hierbei ist, dass der Innenwiderstand der Multiplexer um 3 Größenordnungen größer ist als der der Elastomerverbinder. Die Schwankungen der Innenwiderstände der Multiplexer sind immer noch eine Ordnung größer. Es ist daher notwendig eine Schaltung zu entwickeln, welche zwei Pfade möglichst dicht an den zu messenden Widerstand heran führt. Ein Schema, welches diese Aufgabe erfüllt und auch umgesetzt wurde ist in Abbildung 6.1 zu sehen. Es werden also zwei Pfade an den Widerstand heran geführt. Durch den ersten fließt ein vorgegebener Strom, welcher den Messstrom durch den Widerstand vorgibt. Der zweite Pfad ist unabhängig von diesem Strom und führt lediglich die Potentiale vor und hinter dem Widerstand nach außen. Als Folge wird ein Strom durch den zu messenden Widerstand definiert und die Potentialdifferenz über diesen Widerstand gemessen. Daraus kann der Elastomerwiderstand berechnet werden.

Es wurde eine Messplatine mit zweimal neun 8-auf-1-Multiplexern gefertigt. Mit dieser ist es möglich die 60 Widerstände eines Retikels automatisiert zu messen. Ein gemessener Widerstand setzt sich dabei aus 5 Widerständen zusammen:

- Widerstand vom Stecker bis zum Pad auf der Unterseite der Hauptplatine.
- Widerstand von Elastomer 1, welcher Platine und Wafer miteinander verbindet.
- Metal6-Widerstand auf dem Wafer, welcher zwei benachbarte Pads verbindet.
- Widerstand von Elastomer 2, welcher das benachbarte Pad auf dem Wafer mit dem entsprechendem auf der Platine verbindet.
- Widerstand von diesem Pad zurück zum Stecker.

Der gesamte Widerstand, welcher sich daraus ergibt, wird von dem Metal6-Widerstand dominiert. Auch hier ist der gemessene Widerstand um 2 bis 3 Größenordnungen größer als der (doppelte) Elastomerwiderstand. Dies ist allerdings kein sehr großes Problem, da die Schwankungen der Zuleitungs- und Postprozessierungswiderstände nur durch Temperaturänderungen gegeben sind. Es kann somit durch einen Vergleich von mehreren Messungen auf den Elastomerwiderstand zurück gerechnet werden.

Zusätzlich wurde noch eine zweite Messplatine entwickelt. Diese wird an einen extra Stecker auf der Hauptplatine angeschlossen. An diesen Stecker wurden von jedem anderen Stecker jeweils die zwei ersten Pins gelegt. Es ist somit möglich innerhalb einer Messung einen Eindruck zu bekommen, wie sich die Bedingungen über den Wafer verändern. Beispielsweise der Anpressdruck oder eine mögliche Fehljustage. Abbildung 6.2 zeigt eine Bildschirmaufnahme des Labview-Programms¹, welches die Multiplexer auf der Messplatine ansteuert, die Messspannungen ausliest und in den richtigen Widerstand umrechnet. Das Prinzip der Schaltung ist analog zu dem in Abbildung 6.1. Bei der Flächenmessung wird allerdings nur zwischen 28 Widerständen geschaltet, wodurch sich die Anzahl der Multiplexer von 2mal neun auf 2mal fünf reduziert. Es werden nur 28 Elastomerübergänge, also 14 Retikel, gemessen, da die Art des Routing und die Anzahl an Pins pro Stecker dies nicht anders zulassen. Im weiteren wird auf die Flächenmessung nicht weiter eingegangen, da sich diese mehr als schnelle Visualisierung als als gute Messmethode herausgestellt hat.

6.3 Zuverlässigkeit

Bei den Zuverlässigkeitsmessungen sollten alle Messungen nach guten Justiervorgängen miteinander verglichen werden. Anfangs war es sehr schwierig eine gute Justage mit den

¹Informationen zu Labview sind unter (6) zu finden.

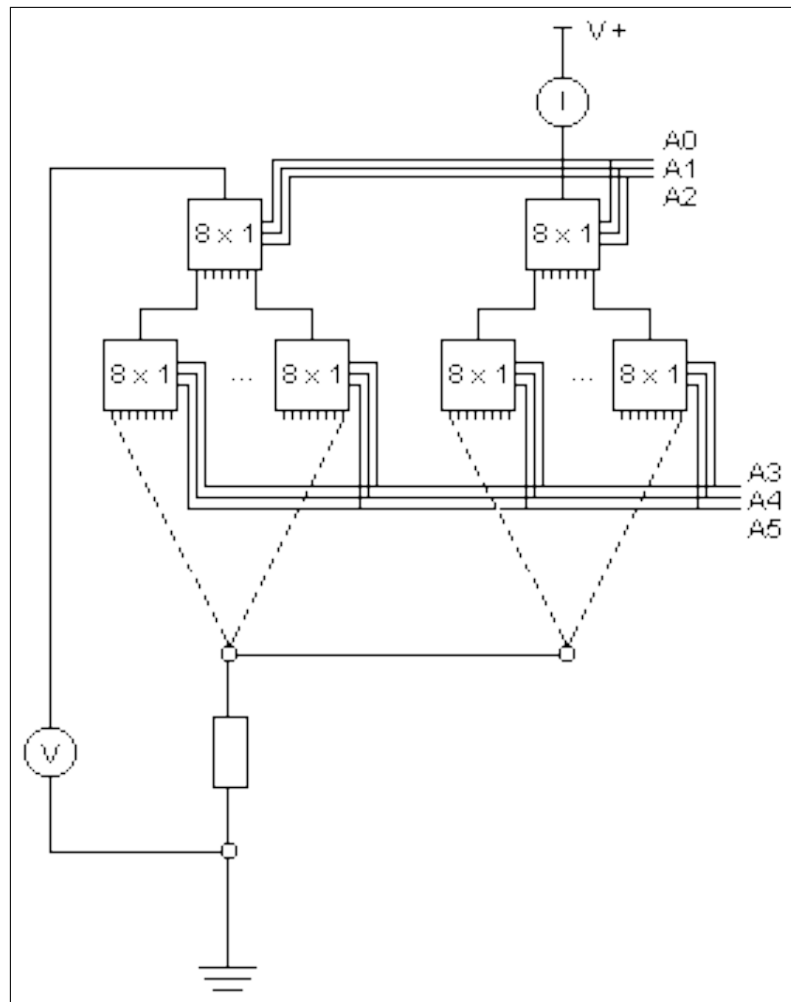


Abbildung 6.1: Schaltungs-Schema zur Messung der Elastomerwiderstände ohne Einfluss der Innenwiderstände der Multiplexer. Durch die Stromquelle wird ein konstanter Strom erzeugt. Dieser wird durch 1 + 8 Multiplexer an den richtigen Widerstand geleitet. Wird der selbe Widerstand an Strom- und Spannungspfad angeschlossen, kann man über den gemessenen Spannungsabfall und gegebenen Strom den Widerstand nach dem ohmschen Gesetz berechnen. Alle Widerstände sind auf einer Seite über Masse miteinander verbunden. Dies bringt ein paar Nachteile mit sich, hat allerdings den Vorteil, dass man nur halb so viele Multiplexer benötigt. Die Multiplexer sind in einer Hierarchie angeordnet. Jede der zwei Hierarchieebenen hat drei Eingänge (A0 bis A2 bzw. A3 bis A5) zum Schalten des richtigen analogen Ausganges.

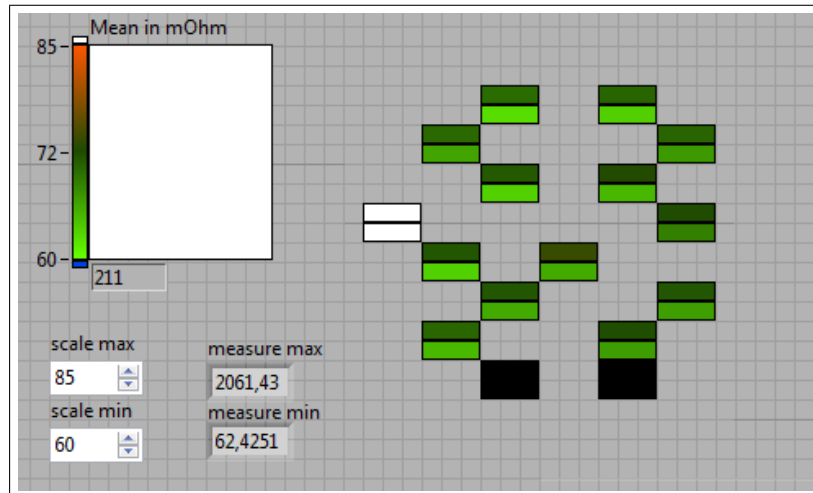


Abbildung 6.2: Das Bild zeigt ein Teil der Oberfläche des mit Labview erstellten Programms zum Ansteuern und Auswerten der Messplatine für die Flächenmessung. Oben links im Bild ist die verwendete Farbskala und der Mittelwert aus allen Einzelmessungen zu sehen. Widerstände, die außerhalb des angegebenen Bereichs liegen werden weiß ($R \geq 85 \Omega$) bzw. blau ($R \leq 60 \Omega$) eingefärbt. Auf der linken Seite sind zwei weiße Kästchen zu sehen. Ihr gemessener Widerstand entspricht dem größt möglichen Spannungsabfall. Die Elastomere kontaktieren dementsprechend nicht. Ursachen hierfür könnten ein verrutschter Elastomer, Verunreinigungen, Fehljustage oder ein Problem mit dem Stecker sein. Die vier schwarzen Kästen an der unteren Seite deuten an, dass diese Widerstände nicht gemessen werden.

zuvor angegeben Genauigkeiten zu erreichen. Dies lag an einer schlechten Ausleuchtung bei der optischen Justage. Das Problem hierbei ist, dass die Markierungen zum Justieren in einem ca. 12 mm tiefem 6 mm im Durchmesser großem Loch liegen. Bei der Aufnahme der Markierungen mit Hilfe eines Lupenobjektives ist es schwierig für eine ausreichende Beleuchtung zu sorgen. Dieses Problem wurde mit Hilfe einer Koaxialbeleuchtung von Falcon² gelöst.

Mit den anschließend durchgeführten Justiervorgängen konnten die benötigten Genauigkeiten erreicht werden und es wurden gute Ergebnisse erzielt. Abbildung 6.3 zeigt ein Ergebnis von den Messungen aller Retikel nach einer guten Justage. Ähnlich zu diesem Bild kann man auch die anderen Messungen zusammenfassen. Ein Großteil der Verbindungen wird gut kontaktiert. Die vier Ausfälle, welche im Bild noch extra markiert sind, befinden sich immer an der selben Stelle. Diese Nichtkontaktierungen haben ihre Ursache nicht in den Elastomeren, da auch bei erneuter Justage und neuen Elastomeren die Ausfälle an den selben Stellen vorkommen. Vielmehr liegt der Grund für diese Ausfälle an den Anschlusssteckern für das Messboard. In Bild 6.3 sind einige Kurzschlüsse zu erkennen. Die Anzahl an Kurzschlüssen entspricht etwa 1 % aller Verbindungen. Diese Tendenz zieht sich über alle Messungen hinweg. Die Lage der Kurzschlüsse variiert dabei von Justage zu Justage. Zwischen zwei Justiervorgängen hingegen ändert sich die Lage der Kurzschlüsse auch unter verschiedenen äußeren Bedingungen wie Lage oder Temperatur des Systems nicht.

Neben der Frage wie gut die Elastomere generell für die vertikale Verbindung geeignet sind, gibt es die Frage, welche Auswirkungen Temperaturschwankungen auf das System

²Falcon LED Lighting Ltd., Offenau. www.falcon-lighting.de

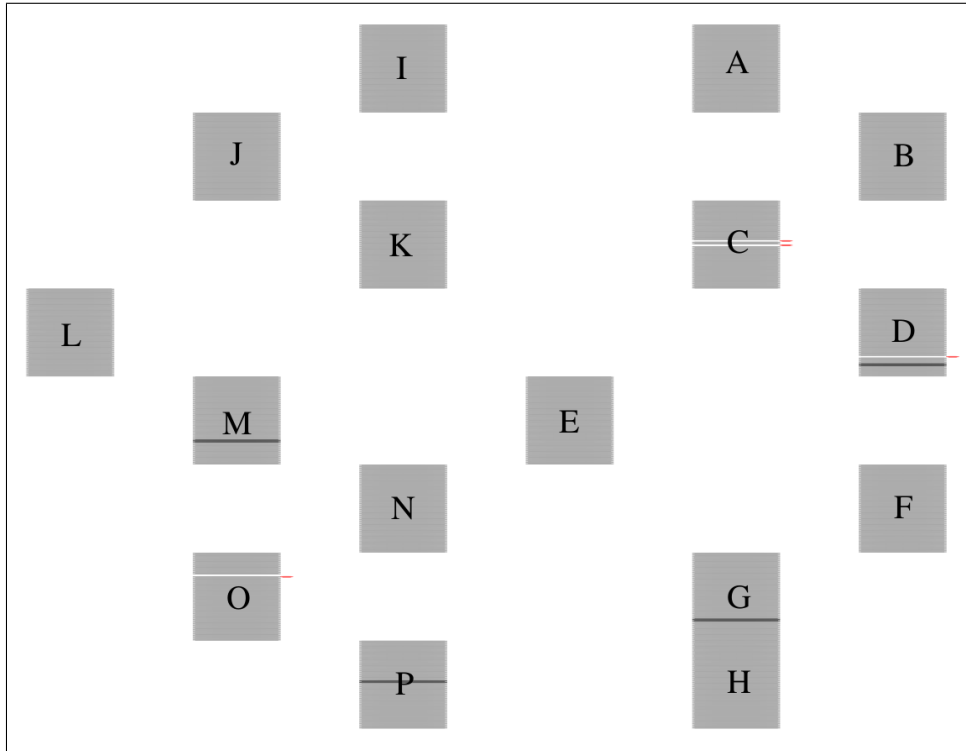


Abbildung 6.3: Das Bild zeigt die in Abbildung 5.2 benannten Retikel A bis P. In Grau werden gute Kontaktierungen dargestellt. Weiß kennzeichnet keine Verbindung und schwarz Kurzschlüsse.

haben. Das System wurde in einem Wärmeschrank unter verschiedenen Temperaturen ausgemessen. Im Wärmeschrank wurde dabei eine Temperaturkurve von 5 bis 70 °C gefahren. Abbildung 6.4 zeigt den Verlauf eines Widerstandes in Abhängigkeit von der Temperatur, welche am Waferträger gemessen wurde.

Die Schwankungen des gemessenen Widerstandes bei Temperaturen unter circa 20 °C sind noch nicht ganz verstanden. Die Schwankungen liegen im Bereich von 1,5 Ω. Dieser Wert liegt um eine Größenordnung höher als der Elastomerwiderstand. Die Ursache für die Schwankung ist demzufolge eher kein Effekt der Elastomere, deren Arbeitsbereich laut Hersteller auch von –40 °C bis 125 °C reicht. Eine Ursache für die Schwankung könnten die Multiplexer sein, welche den Messpfad auswählen. Diese befanden sich während der Messung auch im Wärmeschrank. Jedoch wurde der Temperaturbereich dieser Bauteile³ vom Hersteller mit –40 °C bis 85 °C angegeben. Den größten Anteil am gemessenen Widerstand haben die Metal6-Leitungen auf dem Wafer. Der Arbeitsbereich des Wafers wird vom Hersteller auf einen sehr großen Temperaturbereich bestimmt (–55 °C bis 125 °C). Somit scheint dieser Teil nicht Ursache der Schwankungen zu sein. Die Platine wurde auch für einen Temperaturbereich angegeben, welcher weit über dem im Test verwendeten reicht.

Bild 6.4 zeigt den Verlauf eines Widerstandes während dem Durchlaufen einer Temperaturkurve. Aus diesem Bild wird nicht ersichtlich, ob es während dieser Messung zu Kurzschlüssen oder Nichtkontaktierungen kam. In Abbildung 6.5 wird der Verlauf der gemessenen Widerstände angezeigt. Die Kurven im Diagramm folgen alle dem Temperaturverlauf aus Abbildung 6.4. Es sind insgesamt 60 Kurven. Jede Kurve zeigt den Wider-

³Multiplexer von Vishay: Vishay Siliconix DG408

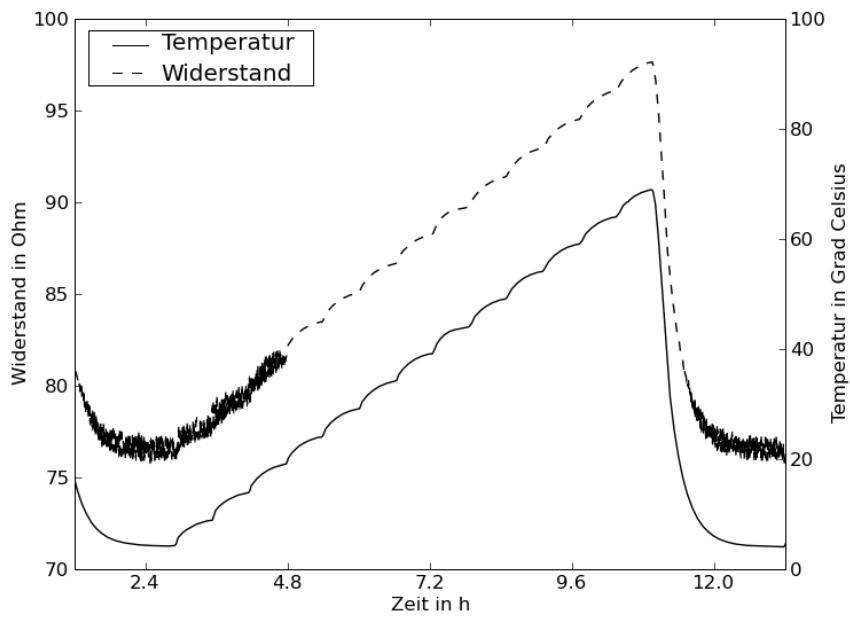


Abbildung 6.4: Das Diagramm zeigt den Verlauf von der Temperatur am Waferträger und dem Widerstand 29 an Retikel C über die Zeit.

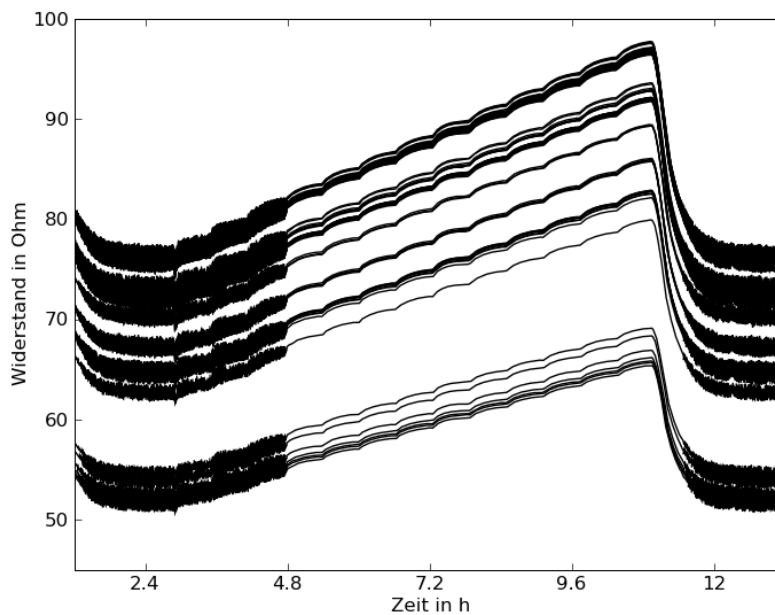


Abbildung 6.5: Im Diagramm sind 60 Widerstände eines Retikels beim Durchlaufen der Temperaturkurve zu sehen. Alle Widerstände folgen dem Temperaturverlauf aus Abbildung 6.4. Es tauchen keine Sprünge auf, die auf einen Kurzschluss (Halbierung des Widerstandwertes) oder eine Nichtkontaktierung (Widerstand $\gg 100 \Omega$) hinweisen würden.

stand eines Kontaktpfades. Der Offset zwischen den Kurven ist mit Unregelmäßigkeiten der Postprozessierungsstrukturen, Metal6-Leitungen und Zuleitungen auf der Hauptplatine zu begründen.

6.4 Elektrisches Verhalten

Neben der allgemeinen Zuverlässigkeit, Anfälligkeit gegenüber Kurzschlüssen sowie Nichtkontaktierungen und der Auswirkung unterschiedlicher Temperaturen ist die Frage der Strombelastbarkeit ein wichtiges Kriterium. Die Aufgabe bestand darin zu untersuchen, in wie weit die Stromangaben des Herstellers eingehalten werden müssen. Laut Hersteller ist ein Elastomerverbinder für Ströme bis 33,9 mA pro Kontaktstreifen ausgelegt. Die verwendeten Elastomere mit einer Länge von 18 mm haben etwa 180 Kontaktstreifen. Zur Strombelastung wurden unterschiedliche Tests durchgeführt. Anfangs wurde ein kleines Testsystem aufgebaut, indem vier Elastomere eingebaut wurden. In diesem Aufbau verbunden die Elastomere zwei Platinen miteinander. Der bei der Messung gemessene Widerstand setzte sich aus drei Widerstände in Reihe zusammen:

- Zuleitung zum Elastomer auf der oberen Platine
- Elastomerwiderstand
- Leitung zwischen Elastomer und Messpin auf der unteren Platine

Die Widerstände der Zuleitung lagen in der Größenordnung um 10Ω und konnten auf 1 % genau bestimmt werden. Bei einem Elastomerwiderstand von etwa $0,1 \Omega$ war es somit nur möglich den Widerstand der Elastomere mit einer Abweichung von 100 % abzuschätzen. Bei den Messungen wurden deshalb auch wieder nur die Änderungen des Gesamtwiderstandes untersucht. Ausschlaggebend für diese Messung war die Reproduzierbarkeit und die Suche nach möglichen Ausfällen. Abbildung 6.6 vergleicht exemplarisch die Ergebnisse einer Erst- und einer Wiederholungsmessung. In der Messung wurde die Spannung über den Strom an einer bestimmten Stelle im kleinen Testaufbau aufgetragen. Bis zu einem Strom von 40 mA ⁴ konnte man immer ein lineares Strom-Spannungsverhalten aufnehmen. Über 40 mA war dieses Verhalten in vielen Messungen nicht mehr linear. Es lies sich bei wiederholten Messungen auch nicht exakt reproduzieren. Allerdings wurde der spezifizierte Bereich bis 40 mA immer gleich gemessen. Ströme über 40 mA flossen in diesen Messungen immer nur wenige Sekunden. In einer von etwa 30 Messungen viel eine Kontaktierung ab einem Strom von 250 mA aus und lies sich nur durch Austausch gegen einen neuen Elastomerverbinder wieder herstellen.

In einem weiteren Test dieses kleinen Systems wurden Ströme bis 400 mA über einen Zeitraum von einer Stunde durch das Elastomer geschickt. Abbildung 6.7 zeigt wieder exemplarisch ein Ergebnis dieser Messungen. Die Entwicklung des Widerstandes über die Zeit ist recht konstant. Nach solch einer Messung wurde immer der lineare Bereich neu ausgemessen und konnte immer sehr genau reproduziert werden.

Ein letzte Stromtest wurde mit dem großen Testsystem durchgeführt. Hier bestand die Schwierigkeit zunächst darin, dass der Messstrom über den Wafer fließen musste. Der Wafer wurde jedoch in erster Linie zur Untersuchung der Postprozessierung entworfen. Die Verbindung zwischen zwei Pads auf dem Wafer besteht aus sehr schmalen Leitungen, welche nur wenige Milliampere Strom transportieren können. Aus diesem Grund wurden

⁴Alle folgenden Stromangaben beziehen sich auf den Strom durch einen Kontaktstreifen.

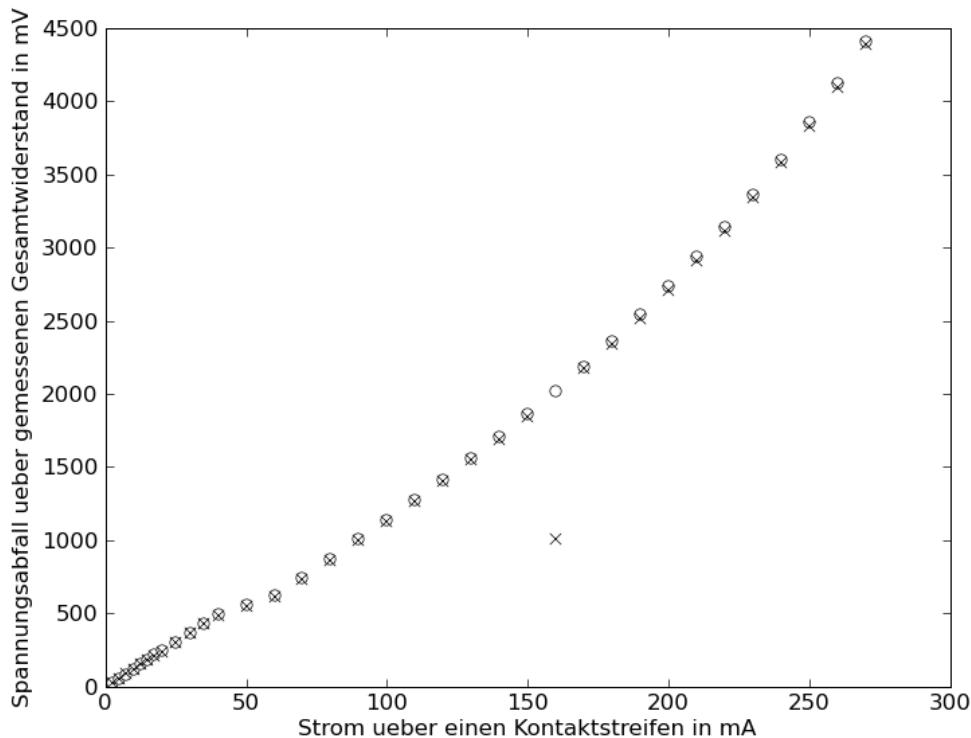


Abbildung 6.6: Spannungsabfall über den Widerstand an Pad D13 im kleinen Testsystem bei verschiedenen Strömen. Die Ströme sind auf einen Kontaktstreifen umgerechnet. Insgesamt waren 10 Kontaktstreifen an der Messung beteiligt. 'o' markiert die erste Messung, 'x' eine Wiederholungsmessung.

an ausgewählten Stellen auf dem Wafer jeweils drei Bonddrähte parallel zu den Metal6-Leitungen gezogen. Anschließend konnten über den Wafer Strombelastungstests abgewickelt werden. Insgesamt wurde ein erhöhter Strom über 12 Padreihen geleitet. Diese Pads wurden in 4 Gruppen zu je 3 benachbarten Padreihen eingeteilt. Alle folgenden Stromangaben beziehen sich hier auf ein 200 μm breites Pad, also 2 Kontaktstreifen. Zum Vergleich mit den obigen Werten müssen die folgenden Angaben dementsprechend halbiert werden. In Abbildung 6.8 ist das Ergebnis einer Messung zu sehen. Dieses Ergebnis steht repräsentativ für alle anderen Messungen, die ein ähnliches Ergebnis aufweisen. Bei diesen Messungen war ein annähernd lineares Verhalten bis 1,4 A zu beobachten. Dabei flossen durch die mittlere Padreihe 1,4 A und durch die äußeren jeweils 1 A. Aufgrund der benutzten Stromquellen konnten keine höheren Ströme getestet werden. Auch nach diesen Strombelastungstests konnte der Bereich für kleine Ströme identisch mit den Messungen vor der Belastung reproduziert werden. Außer in einem Fall gab es keine Veränderung in den Eigenschaften des Elastomers bedingt durch die hohen Ströme. Eine Verbindung brach bei einem Strom von 300 mA ab und konnte nicht mehr hergestellt werden. Dieses Problem lag jedoch weniger am Elastomer, als an der Verbindung auf dem Wafer. Hier sind vermutlich die gebondeten Drähte beim Zusammenbau beschädigt worden und unter Belastung durch den hohen Strom zerstört worden. Auch nach erneutem Zusammenbau mit neuen Elastomeren konnte hier kein Kontakt hergestellt werden.

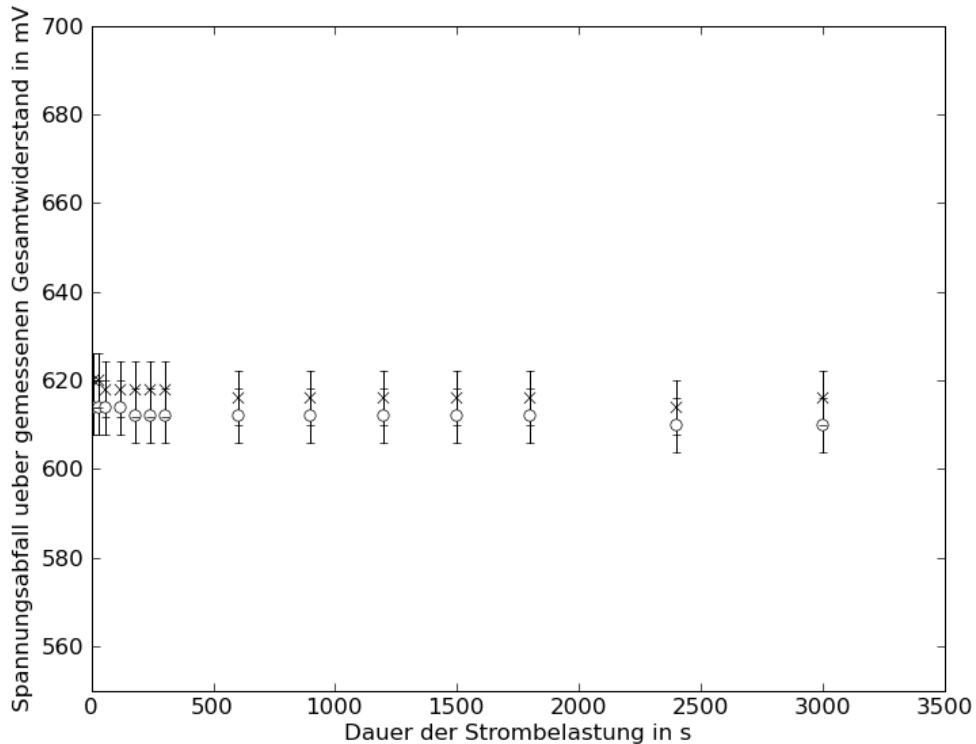


Abbildung 6.7: Zeitliche Änderung des Spannungsabfalls an Pad C23 im kleinen Testsystem. Während der Messung floß ein Strom von 400 mA pro Kontaktstreifen. An der Kontaktierung waren zwei Kontaktstreifen beteiligt. 'o' und 'x' markieren wieder Erst- und Wiederholungsmessung

6.5 Diskussion

Die Tests mit den Elastomeren haben verschiedene Erkenntnisse erbracht. Die Zuverlässigkeit der Elastomerverbinder hinsichtlich Kurzschlüssen und Nichtkontaktierungen fällt recht positiv aus. So gibt es bei guter Justage keine Ausfälle bei der Kontaktierung, die auf die Elastomere zurück zu führen wären. Kurzschlüsse treten in kleiner Anzahl auf. Diese Tatsache ist ein Problem, auf das man reagieren muss. Um Kurzschlüsse im finalen System möglichst zu vermeiden können zwei Maßnahmen getroffen werden. Zum einen wird man die minimale Padgröße und den minimalen Padabstand vergrößern. Im Testsystem wurden 190 μm breite Pads mit einem Abstand von 100 μm verwendet. Diese Zahlen werden für das endgültige System auf mindestens 200 μm sowohl für Padbreite, als auch für Padabstand erhöht. Darüber hinaus kann man auch die Größe Pads auf der Hauptplatine kleiner wählen als der Abstand der Pads auf dem Wafer ist. Diese Maßnahme würde Kurzschlüsse selbst bei einer schlechten Justage weitestgehend ausschließen und somit mögliche Beschädigungen verhindern.

Die Temperaturmessungen haben gezeigt das man das System in einem Bereich von 20 $^{\circ}\text{C}$ bis 70 $^{\circ}\text{C}$ betreiben kann. Niedrigere Temperaturen wird man im Betrieb kaum erreichen, da das System zum einen ausschließlich unter Laborbedingungen bedient wird und somit keinen Temperaturen unter 20 $^{\circ}\text{C}$ ausgesetzt ist. Zum anderen wird das System gekühlt. Man kann eine Abschätzung⁵ machen, nach der zwischen Kühlkörper und

⁵Für die Abschätzung wird angenommen, dass 1 kW an Abwärme vom Wafer wegtransportiert werden muss. Der verwendete Kühlkörper aus Aluminium hat dabei eine effektive Dicke von 1 cm. Zusätzlich muss die Dicke des Waferträgers aus Aluminium und die Dicke des Wafers aus kristallinem Silizium berücksichtigt

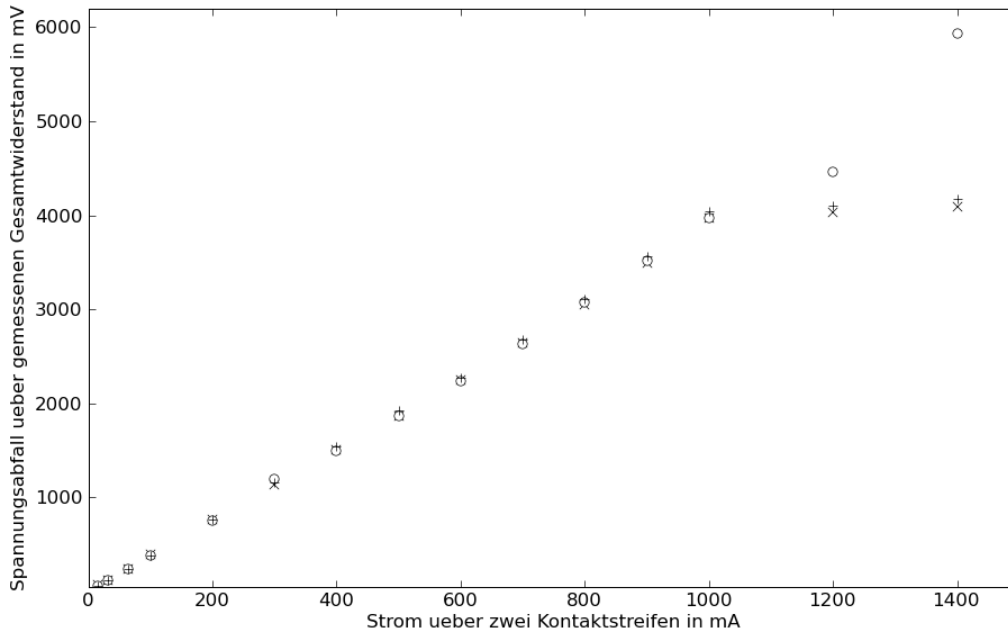


Abbildung 6.8: Das Diagramm zeigt die Abhängigkeit vom Spannungsabfall zum Strom bei der Messung über zwei parallele Kontaktstreifen. Es werden alle drei Pad-übergänge dargestellt. Mit 'x' und '+' werden die außen liegenden Übergänge dargestellt. 'o' markiert den dazwischen liegenden Übergang. Während der Strom für den mittleren Übergang dem Strom an der x-Achse entspricht, werden bei den äußeren Übergängen die letzten Punkte bei einem Strom von 1,0 A gemessen.

Waferoberfläche ein Temperaturkoeffizient um 15 K auftritt. Die Kühlung des Systems erfolgt über vorbei strömende Raumluft. Bei einer Raumtemperatur von maximal 30 °C kann man eine maximale Temperatur an den Elastomeren von 45 °C annehmen. Verglichen mit den Messergebnissen lässt diese Annahme noch genügend Sicherheit zu, so dass es mit zu hohen Temperaturen für die Elastomere keine Probleme geben sollte.

Die Strombelastungstests haben interessante Ergebnisse erbracht. Die vom Hersteller angegebenen 33,9 mA pro Kontaktstreifen stellten in allen Messungen kein Problem dar. Höhere Ströme waren in fast allen Fällen ohne negative Auswirkungen für die Elastomere. Allerdings fiel eine Verbindung durch Zerstörung des Elastomers bei einem Strom über 200 mA pro Kontaktstreifen aus. In einem zweiten Fall lag der Ausfall der Kontaktierung vermutlich nicht am Elastomer. Es kann allerdings nicht ausgeschlossen werden, dass auch hier der Elastomer bei etwa 150 mA pro Kontaktstreifen kaputt ging. Zusammenfassend ziehen hohe Ströme nicht zwangsläufig einen Defekt der Elastomere nach sich. Allerdings sollte man nicht mit Strömen planen, die über 40 mA pro Kontaktstreifen hinaus gehen. Bis zu dieser Grenzen verhielt sich der Spannungsabfall über die Elastomere linear zum Strom und konnte in verschiedenen Messungen reproduziert werden.

werden. Als Kühlfläche wird die Fläche eines runden Wafers mit 20 cm Durchmesser angenommen. Bei der Verwendung von $221 \frac{W}{K \cdot m}$ und $148 \frac{W}{K \cdot m}$ für die Wärmeleitfähigkeiten von Aluminium bzw kristallinem Silizium folgt daraus eine Temperaturdifferenz von ungefähr 6 K. Zu diesem Wert muss man noch den Übergang zwischen Kühlkörper und Wafertträger und zwischen Wafertträger und Wafer berücksichtigen. Außerdem sollte man berücksichtigen, dass die Abwärme nicht gleichmäßig über die gesamte Fläche verteilt entsteht.

Kapitel 7

Zusammenfassung und Ausblick

In der Arbeit wurde ein Prototypensystem für die Wafer-Scale-Integration vorgestellt. Wichtige Komponenten waren dabei die Postprozessierung des Wafers zum Verbinden der einzelnen Retikel, die vertikale Verbindung zwischen Wafer und Hauptplatine, mechanische Komponenten wie beispielsweise die Positionierungsmaske für die Elastomere und die Mechanik zum Ausrichten der einzelnen Komponenten.

Messungen ergaben, dass die Postprozessierung funktioniert und dass es möglich ist mit einem Punkt-zu-Punkt-Abstand von $9,5\ \mu\text{m}$ die Mindestanforderung von $9,7\ \mu\text{m}$ zu erfüllen. Mögliche Verschiebungen von Postprozessierungsstrukturen und Strukturen auf dem Wafer durch thermische Ausdehnung der Postprozessierungsmaske können dabei ausgeglichen werden. Ein Vergolden der Postprozessierungspads schützt die Pads vor Oxidation und gewährleistet somit gleichmäßig gute Kontaktoberflächen.

Bauteile wie Deckel, Waferträger und Positionierungsmaske wurden getestet. Der Deckel und der Waferträger erfüllten dabei ihren Zweck, das System zusammen zu halten ohne sich zu verformen. Eine Kontaktierung der Elastomere erfolgte bei jedem Zusammenschrauben des Gesamtsystems. Auch bei der Verwendung von knapp 456 Elastomeren funktionierten alle Elastomere und weder Deckel noch Waferträger verformten sich. Beim Zusammenschrauben ist jedoch darauf zu achten die 16 Schrauben gleichmäßig anzuziehen, um ein Verkanten des Deckels und eine daraus resultierende ungleichmäßige Pressung der Elastomere zu vermeiden. Die Positionierungsmaske kann durch ihre Befestigungsschrauben ausreichend gut zur Hauptplatine ausgerichtet werden. Die Dicke der Maske von $0,51\ \text{mm}$ ist klein genug für die Verwendung der Elastomere und groß genug um eine ausreichende Stabilität zu gewährleisten. Problematisch sind hier nur die Ausfräsungen in der Schablone. Diese sollten auf der einen Seite nicht zu klein sein, um zu starken Verformungen der Elastomere vorzubeugen, auf der anderen Seite müssen sie klein genug sein, damit die Elastomere während der Justage nicht aus der Maske fallen. Die momentan verwendete Maske weist keine optimale Lösung auf, da hier bei der Justage einige Elastomere aus der Maske fielen. Es ist notwendig eine weitere Positionierungsmaske herzustellen.

Die Elastomere erfüllen alle an sie gestellten Anforderungen. Es ist möglich die geforderte Leitungsdichte mit den Elastomeren herzustellen. Im finalen System können dazu Pads mit einer Breite und einem Abstand $\geq 200\ \mu\text{m}$ benutzt werden. Die Funktionsfähigkeit der Elastomere konnte in einem Temperaturbereich von $20\ ^\circ\text{C}$ bis $70\ ^\circ\text{C}$ gesichert werden. Strombelastungstests zeigten, dass die Elastomere Ströme bis $40\ \text{mA}$ pro Kontaktstreifen sicher verarbeiten können. Bei Strömen bis $200\ \text{mA}$ traten keine dauerhaften Schäden auf. Ströme über $200\ \text{mA}$ führten teilweise zu Ausfällen einzelner Bereiche der Elastomere. Es wurden allerdings auch Ströme bis $800\ \text{mA}$ pro Kontaktstreifen über eine Dauer von $10\ \text{min}$ gemessen, welche das Elastomer nicht beschädigten.

Zum Ausrichten der Komponenten wurde eine Justiervorrichtung entwickelt. Mit ihr und

einem optischen Verfahren zum Feststellen der relativen Lage von Wafer zu Hauptplatine ist es möglich die beiden Teile auf $50\ \mu\text{m}$ in x- und y-Richtung genau zueinander auszurichten. Zusätzlich kann die Verdrehung mit einer maximalen Abweichung von $(30 \cdot 10^{-3})^\circ$ eingestellt werden. Eine variable Einstellung der Höhe der Elastomere im gepressten Zustand ist derzeit nicht möglich. Mit dem verwendeten Testwafer werden die Elastomere auf $0,8\ \text{mm}$ gepresst. Für den Fall einer Änderung der Dicke verschiedener Wafer oder anderen nicht erwarteter Veränderungen, die diese Höhe beeinflussen, sollte die Möglichkeit nachgerüstet werden, die Endhöhe der Elastomere einzustellen. Ein Verfahren dies zu tun wurde vorgestellt und wird zukünftig getestet.

Erste Schritte für den Bau eines Systems für die Wafer-Scale-Integration wurden somit bewältigt. Eine ungeklärte Frage ist der Einfluss der Elastomerverbinder auf Hochfrequente Signale. Messungen hierzu folgen nach einer Präparierung des Testwafers, da mit dem Testwafer in der momentanen Version keine entsprechenden Tests möglich sind. Es stehen auch noch weitere Aufgaben an. Die Signale, welche vom Wafer über die Elastomere auf die Platine geführt wurden, müssen dort an die entsprechenden Stellen weitergeleitet werden. Hierbei kann man zwei Unterscheidungen treffen. Zum einen müssen Kommunikationssignale angesteuert und ausgewertet werden. Zum anderen müssen die Schaltungen des HICANN-Chips mit den entsprechenden Spannungen versorgt werden. Die Elektronik für die Kommunikation soll über Aufsteckeinheiten hergestellt werden. Diese Einheiten bestehen aus vier DNC-Platinen, einem FPGA-Board, einem Kühlsystem und zwei Netzwerkanschlüssen für den Anschluss eines PCs. Insgesamt werden 12 solcher Einheit auf die Hauptplatine aufgesteckt. Neben diesen Einheiten werden zwei bis vier sogenannte Powerplatinen aufgesteckt. Auf diesen Platinen wird die Verteilung und Kontrolle der Leistungsversorgung für den Wafer, die DNC-Chips und die FPGAs umgesetzt. Zusätzlich wird weitere Elektronik auf der Hauptplatine aufgebracht, welche zur Überwachung der Systemtemperatur, der Abschaltung einzelner Bereiche des Gesamtsystems, sowie zur Initialisierung des gesamten Systems verantwortlich ist. All diese Teile im System unterzubringen ist eine sehr komplexe Aufgabe. Abbildung 1.1 zeigt die aktuelle Planung, die diese Aufgabe umsetzen soll.

Neben der Komplexität durch die hohe Anzahl an Teilen, die zusammen an das System angeschlossen werden, ist die Kühlung des Systems eine anspruchsvolle Aufgabe. Das Prinzip der Kühlung des Systems ist ebenfalls in Abbildung 1.1 zu sehen. Der Wafer soll eine Leistung von maximal $1\ \text{kW}^1$ verbrauchen. Die Elektronik zur Steuerung der Kommunikation verbraucht auch etwa $1\ \text{kW}$ und bei der Erzeugung der benötigten Versorgungsspannungen entstehen rund $200\ \text{W}$. Grob abgeschätzt müssen $2,5\ \text{kW}$ Wärme pro Wafersystem abgeführt werden. Diese Wärme soll im aktuellen Konzept über Luftkühlung abtransportiert werden. Um dies zu gewährleisten soll die entstehende Wärme jeweils an einzelne Kühlkörper geführt werden, die sich in einem gemeinsamen Luftstrom befinden. Die kritischsten Stellen sind dabei die Wärmeentwicklung des Wafers und die der DNCs. Die Wärme des Wafers wird allerdings über eine sehr große Fläche an einen sehr großen Kühlkörper, der direkt an den Wafertäger montiert wird, abgeführt. Die Kühlung der DNCs ist etwas schwieriger, da diese zwischen der Hauptplatine und den FPGA-Platinen liegen. Um die Wärme der DNCs an Kühlkörper zu leiten, die sich im Luftstrom befinden, werden Heatpipes verwendet.

Da das gesamte System entwickelt wurde um große neuronale Netzwerke abbilden zu können, ist auch die Möglichkeit vorgesehen mehrere Wafersysteme zusammen zu schließen. Es

¹Die maximale Leistungsaufnahme der Schaltungen über den ganzen Wafer beträgt $1\ \text{kW}$. Dieser Fall tritt ein, wenn in den realisierten Netzwerken eine hohe Neuronenaktivität herrscht. In vielen Fälle wird dies jedoch nicht der Fall sein und es werden immer nur kleine Bereiche aktiv sein. Somit wird die durchschnittliche Leistungsaufnahme meistens nur einen Teil von $1\ \text{kW}$ betragen.

gibt daher auch ein Konzept mehrere Systeme in ein Rack² einzubauen. In ein Rack können dabei bis zu fünf Systeme eingebaut werden. Ein Problem, welches sich beim Zusammenbau eines Mehrwafer-Systems ergibt, ist die maximale Leistungsaufnahme aller Systeme zusammen. Hierfür sollte noch ein Konzept entwickelt werden, welches die entstehende Wärme sinnvoll nutzen kann, denn ein großer Vorteil von diesen analogen Schaltkreisen ist ihr geringer Leistungsverbrauch. Bei einem System, dessen großer Vorteil 'Low Power' sein soll, ist eine Verlustleistung von mehreren Kilowatt sehr fragwürdig, auch wenn ein sehr großes Netzwerk abgebildet wird und eine hohe Informationsverarbeitung in kurzer Zeit stattfindet. Trotzdem ist dies verglichen mit alternativen Ansätzen zur Simulation großer neuronaler Netzwerke (beispielsweise auf Rechen-Clustern) gut.

Prinzipiell ist die Wafer-Scale-Integration, welche nur durch die Verwendung von analoger Schaltungstechnik mit geringerem Stromverbrauch möglich geworden ist, eine gute Anwendung. Viele Probleme die anfangs auftraten wurden gelöst. Für die Probleme, die nach wie vor noch bestehen gibt es gute Lösungskonzepte, so dass es innerhalb des nächsten Jahres auch möglich sein wird ein erstes komplettes System in Betrieb zu nehmen. Die größte Hürde wird dabei der Zusammenbau und das fehlerfreie Zusammenspiel aller Komponenten sein.

²the rack - das Regal. Rack ist in der Elektronik ein Schrank, indem einzelne Bauteile wie Platinen zu einem Gesamtsystem verschaltet werden. Man kann Racks kaufen, welche schon Spannungsgeneratoren und ein Kühlsystem integriert haben.

Literaturverzeichnis

- [1] FACETS - Fast Analog Computing with Emergent Transient States, www.facets-project.org, 2008.
- [2] Falcon LED Lighting, www.falcon-lighting.de, 2009.
- [3] Fierres, J., J. Schemmel, and K. Meier: *Realizing biological spiking network models in a configurable wafer-scale hardware system*, in Proceedings of the 2008 International Joint Conference on Neural Networks (IJCNN), 2008.
- [4] Grübl, A.: *VLSI Implementation of a Spiking Neural Network*, Dissertation, Universität Heidelberg, 2007.
- [5] Husmann de Oliveira, D., H. Zoglauer: *A Prototype System for testing the reliability of the interplay of the components*, Universität Heidelberg, 2008.
- [6] LabVIEW - Laboratory Virtual Instrumentation Engineering Workbench, www.ni.com/labview, 2008.
- [7] Millner, S.: *An Integrated Operational Amplifier for a Large Scale Neuromorphic System*, Diplomarbeit (englisch), Universität Heidelberg, 2007.
- [8] Schemmel, J.: *Specification of the HICANN Microchip*, Universität Heidelberg, 2008.
- [9] Schemmel, J., J. Fierres, and K. Meier: *Wafer-scale integration of analog neural networks*, in Proceedings of the 2008 International Joint Conference on Neural Networks (IJCNN), 2008.
- [10] Skript zur Vorlesung Mikrolithographie I + II, SS91 + WS91/92, FH Köln.
- [11] Technische Universität Dresden, FACETS Deliverable D7-5: *Demonstrate the FPGA-based host interface as well as the communication between the individual components of the wafer-scale system*, vom internen FACETS Dokumenten-Server, 2008.
- [12] Technische Universität Dresden, FACETS Technical Report: *DNC specification*, vom FACETS-internen Dokumenten-Server, 2008.
- [13] Technische Universität Dresden, FACETS Technical Report: *HICANN - digital interface specification*, vom FACETS-internen Dokumenten-Server, 2008.
- [14] UMC - United Microelectronics Corporation, www.umc.com, 2009.

[15] Würth Elektronik, www.we-online.de, 2008.

Datenblätter

- [16] Binder Kälte-/Wärmetestkammer MK: *Betriebsanleitung - Temperaturprüfschränke für natürliche Simulation mit Bildschirm-Programmregler MB1*, 2006.
- [17] DMM Keithley 2001: *Model 2100 6 1/2-Digit Resolution Digital Multimeter User's Manual*, www.keithley.com, 2008.
- [18] Rhode & Schwarz - NGPT 35: *Beschreibung - Dreikanal-Stromversorgungsgerät, Typ NGPT*.
- [19] Tektronix - PS2521G: *User Manual - PS2520, PS2520G, PS2521 & PS2521G Programmable Power Supplies*, 1995.
- [20] Vishay Siliconix DG408/409: *8-Ch/Dual 4-Ch High-Performance CMOS Analog Multiplexers*, www.vishay.com, 2008.
- [21] ZEBRA[®]High Performance Silver Connectors: *Fujipoly[®]New High Performance Silver ZEBRA[®] Connector*, Fujipoly Data Sheet Number FPDS 01-34 / Version 2, 2002.

Danksagungen

Die Arbeit wurde in der Electronic Vision(s) Gruppe am Kirchhoff-Institut für Physik in Heidelberg durchgeführt. Die Betreuung und die Zusammenarbeit innerhalb der Gruppe könnten besser nicht sein. Somit danke ich der gesamten Gruppe für die Unterstützung während der Arbeit und beim Zusammenschreiben. Ganz besonderen Dank gilt,

- Dr. Johannes Schemmel, der mich in die Gruppe gebracht hat, das richtige Thema in einem ausführlichen Gespräch mit mir gefunden hat und mit fachlichem Rat immer helfen konnte.
- Dan Husmann, für die unzählig beantworteten Fragen und die Organisation der außerordentlichen Treffen.
- the Softie's room, allen voran Eric Müller für Hilfe bei den kleinsten Problemen.
- Prof. Dr Karlheinz Meier

Neben der Hilfe aus der Gruppe konnte man sich jederzeit auch auf andere Mitarbeiter des Instituts verlassen. Eine große Hilfe waren mir dabei,

- Ralf Achenbach
- das Technikon des Instituts
- Marcus Dorn

Ganz wichtig für mich während der Diplomarbeit und dem gesamten Studium waren neben guten Freunden meine Eltern. Vielen Dank für die moralische und finanzielle Unterstützung.

Zu guter Letzt und weil er mir gezeigt hat, wie man das Leben von der ruhigen Seite genießen kann noch einen großen Dank an meinen Hund, Knuddel!

Erklärung:

Ich versichere, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 14.04.2009

.....

(Unterschrift)