

Testen eines Floating-Gate Analogspeichers in 65 nm Single-Poly Technologie

Bachelorarbeit
zur Erlangung des akademischen Grades
Bachelor of Science

vorgelegt von
Stanislaw Hüll
geboren in Rostow am Don, Russland

Fakultät für Physik und Astronomie
Universität Heidelberg
2014

1. Gutachter : Prof. Dr. Karlheinz Meier

2. Gutachter : Dr. Johannes Schemmel

Datum des Einreichens der Arbeit: 03. März 2014

Kurzfassung

Die vorliegende Arbeit erläutert die Untersuchungen an Floating-Gate Speicherzellen die in einem 65 nm Single-Poly Prozess implementiert sind. Orientiert am Vorgängermodell, das in einem 180 nm-Herstellungsprozess entwickelt und hergestellt worden ist (siehe [1]), wird der Frage nachgegangen inwiefern ein Einsatz als analoger Parameterspeicher für neuromorphe Hardware in Frage kommt. Anhand zehn untersuchter Speicherzellen wird die prinzipielle Funktionsfähigkeit bestätigt und grundlegende Eigenschaften wie Speicherzeit und Dauerbelastbarkeit untersucht; tieferegehende Analysen können aufgrund eines Fehlers im Entwurf der Schaltung nicht durchgeführt werden. Der potentielle Einsatz als Analogspeicher ist aufgrund der Ergebnisse dieser Arbeit durchaus denkbar, stichhaltige Aussagen erfordern jedoch weitere Untersuchungen.

Abstract

This thesis presents the results of tests on 65 nm single-poly floating gate memory cells. A previous version, implemented in 180 nm technology ([1]), is successfully being applied as analog parameter memory in neuromorphic hardware. Through the examination of ten test memory cells it is thus inspected whether these are also suitable for operation in this realm. The memory is, in principle, found to be functional and some basic properties like the storage period and resilience to long-term duty are analyzed; due to erroneous chip design more sophisticated testing could not be conducted. According to the results of this thesis a possible utilization as analog memory is imaginable, however, more investigation is needed for an accurate statement.

Inhaltsverzeichnis

1	Einleitung	2
2	Grundlagen	4
2.1	Aufbau des Floating-Gate Speichers	4
2.2	Beschreiben des Floating-Gates	6
2.3	Auslesen des Wertes	8
2.4	Speicherzeit	9
3	Experimenteller Aufbau	10
3.1	Testchip mit Floating-Gate Speicherzellen	10
3.2	Setup zum Erzeugen von Programmierpulsen	13
4	Messungen und Ergebnisse	17
4.1	Lade- und Entladezeitmessungen	17
4.1.1	Messmethode	17
4.1.2	Resultate	17
4.2	Speicherzeitmessungen	19
4.2.1	Messmethode	19
4.2.2	Beispielmessung der Speicherzeit	22
4.2.3	Resultate	23
4.3	Stresstest	29
4.3.1	Messmethode	29
4.3.2	Resultate	29
4.4	Lichtempfindlichkeit	30
4.4.1	Messmethode	30
4.4.2	Resultate	31
5	Zusammenfassung und Ausblick	32
	Quellenverzeichnis	33

Kapitel 1

Einleitung

Die Notwendigkeit große Datenmengen elektronisch zu speichern verhalf dem Floating-Gate Speicher in den letzten Jahrzehnten zu großer Verbreitung. In seiner bekanntesten Ausführung ist er als digitaler Flash-Speicher überall dort vertreten, wo platzsparend und mit geringem Stromverbrauch nichtflüchtig gespeichert werden muss, wie z.B. in Mobiltelefonen und Digitalkameras. Auch in Form von aktuellen Solid-State-Drives findet dieses Prinzip Einsatz.

Sehr spezielle Anwendung findet der Floating-Gate Speicher in der Disziplin der neuromorphen Hardware, einem aktuellen Forschungsbereich der Neurowissenschaft. Große neuronale Netzwerke werden durch Emulation untersucht, d.h. einzelne elektrische Schaltkreise, deren Ströme den Ionenfluß in Neuronen repräsentieren, werden zu großskaligen Netzwerken geschaltet um im Verbund zu arbeiten. Dieses Vorgehen bietet gegenüber der Simulation den entscheidenden Vorteil, dass die Größe und Komplexität der zu untersuchenden Strukturen nicht durch Rechenleistung beschränkt wird.

Das analoge VLSI-System¹ der Heidelberger Forschungsgruppe Electronic Vision(s) vereint bspw. etwa 200k Neuronen auf einem einzigen Wafer (s.[2]). Rund zwanzig individuell einstellbare Parameter pro Neuron bieten dem Experimentierenden ein hohes Maß an Anpassungsfähigkeit. Der hieraus resultierenden hohe Speicherbedarf wird durch Floating-Gate Speicher gedeckt, der hier allerdings als Analogspeicher eingesetzt wird.

In seiner üblichen Anwendung tritt der Floating-Gate Speicher als digitaler Speicher² auf; eine Speicherzelle repräsentiert ein einzelnes Bit und stellt in Verbindung mit weiteren Bits eine abgeschlossene Speichereinheit dar. Im Falle analoger neuromorpher Hardware ist es allerdings notwendig Analogspeicher zu implementieren, da alle auftretenden Ströme und Spannungen als analoge Signale verwendet werden, also in Form ihrer physikalischen Erscheinung. Die Funktionsweise des Floating-Gate Speichers (Kap. 2) bietet grundlegend die Möglichkeit neben zwei binären Zuständen auch stufenlos zu speichern. In Verbindung mit der kleinen Größe und der geringen Leistungsaufnahme ist der Floating-Gate Speicher daher für den großskaligen Einsatz als Analogspeicher auf neuromorpher Hardware prinzipiell geeignet.

¹„Very Large Scale Integration“; Angabe der Integrationsdichte bei integrierten Schaltungen.

²auch als „binärer“ Speicher bezeichnet.

Das Heidelberger System mitsamt dem verwendeten Floating-Gate Speicher wurde in einem 180 nm-Prozess³ hergestellt. Ein Nachfolgemodell in kleinerer 65 nm-Technologie ist aktuell in Planung, wobei man sich hier vorerst gegen den Einsatz von Floating-Gate Speicher entschieden hat. Im aktuellen System sind Probleme im Zusammenhang mit der Schreibzeit und der Genauigkeit des Speichers beobachtet worden ([3], Kap. 3 und [4], Kap. 9), sodass stattdessen ein neu entwickelter kapazitiver Analogspeicher verwendet werden soll (siehe [5]). Floating-Gate Speicher bleibt aber dennoch für den Einsatz in analoger Hardware für zukünftige Projekte interessant.

Ziel dieser Arbeit ist es daher Floating-Gate Speicher auf Basis von 65 nm-Technologie zu untersuchen und eine Einschätzung über eine mögliche Verwendung als Analogspeicher zu geben. Es werden der Aufbau und Funktionsweise von Floating-Gate Speicher in Single-Poly Technologie erläutert und der experimentelle Aufbau beschrieben mit dem Untersuchungen an einigen Speicherzellen durchgeführt worden sind. Abschließend werden die Messergebnisse diskutiert und ein Fazit als Ergebnis der Arbeit präsentiert.

³kleinste laterale Strukturgröße die im Prozess verwendet wird.

Kapitel 2

Grundlagen

Dieses Kapitel gibt einen kurzen Überblick über den Aufbau und das Prinzip des Floating-Gate Speichers in Single-Poly – Technologie wie er in neuromorpher Hardware verwendet wird. Floating-Gate Speicher wird als Flash-Speicher üblicherweise in Dual-Poly Technologie hergestellt. Die Single-Poly – Realisierung wird erstmals in [6] beschrieben.

Genauere Ausführungen zu den hier erwähnten physikalischen Mechanismen der Lade- und Entladeprozesse sind in [1] beschrieben. Die großskalige Implementierung analogen Floating-Gate Speichers auf neuromorpher Hardware wird in [4], Kap. 9 beschrieben.

2.1 Aufbau des Floating-Gate Speichers

Eine Floating-Gate Speicherzelle in Single-Poly Technologie besteht aus zwei unterschiedlich dimensionierten selbstsperrenden p-Typ MOSFET-Transistoren und einem ebenfalls selbstsperrenden n-Typ MOSFET-Auslesetransistor¹ (Abb. 2.1).

Die Gates aller drei Transistoren sind miteinander verbunden. Da ansonsten keine weiteren Anschlüsse existieren gibt es keinen direkten Zugriff auf das Gate; es „floatet“.

Source-, Drain- und Bulkanschluss innerhalb der pMOS sind kurzgeschlossen. Die Transistoren agieren daher in der dargestellten Schaltung als Kondensatoren. Dies ergibt sich daraus, dass bezüglich des Gates Drain, Source und Bulk eine Parallelschaltung dreier Kapazitäten darstellen die in ihrer Ersatzschaltung als einzelne Kapazität in Erscheinung tritt (Abb. 2.2). Die Transistoren werden daher im Folgenden mit ihren Kapazitäten C_L bzw. C_S identifiziert. Die üblichen Transistorströme entfallen da keine Potentialdifferenz² zwischen Drain und Source existiert.

Der Drain-Source Strom I_{DS} des Auslesetransistors hängt bei gegebenem Drain-Potential V_{DD} nur von der Gate-Source Spannung ab und wird daher in der vorliegenden Konfiguration vollständig durch Potential des Floating-Gates V_{FG} bestimmt. Setzt man die Ausgangskennlinien des Transistors als bekannt voraus, so kann jedem gemessenen Strom sein korrespondierendes Gate-Potential zugeordnet werden. Dieser Mechanis-

¹Die Transistoren werden im Folgenden kurz mit *pMOS* bzw. *nMOS* abgekürzt.

²Alle Potentiale beziehen sich auf das Massepotential der Schaltung.

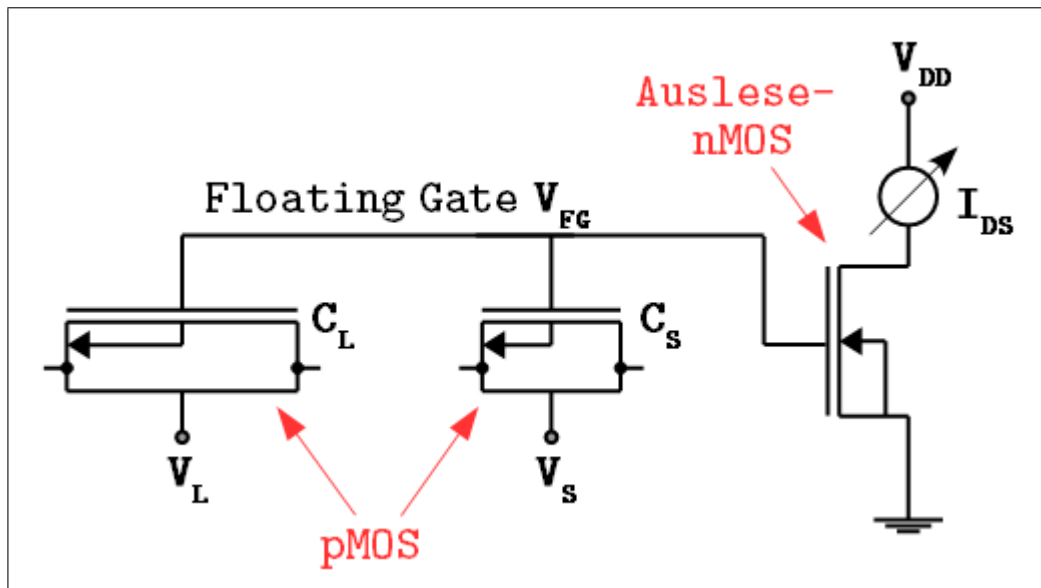


Abbildung 2.1: Aufbau einer Floating-Gate Speicherzelle in Single-Poly Technologie mit großer (C_L) und kleiner (C_S) Kapazität und Auslesetransistor.

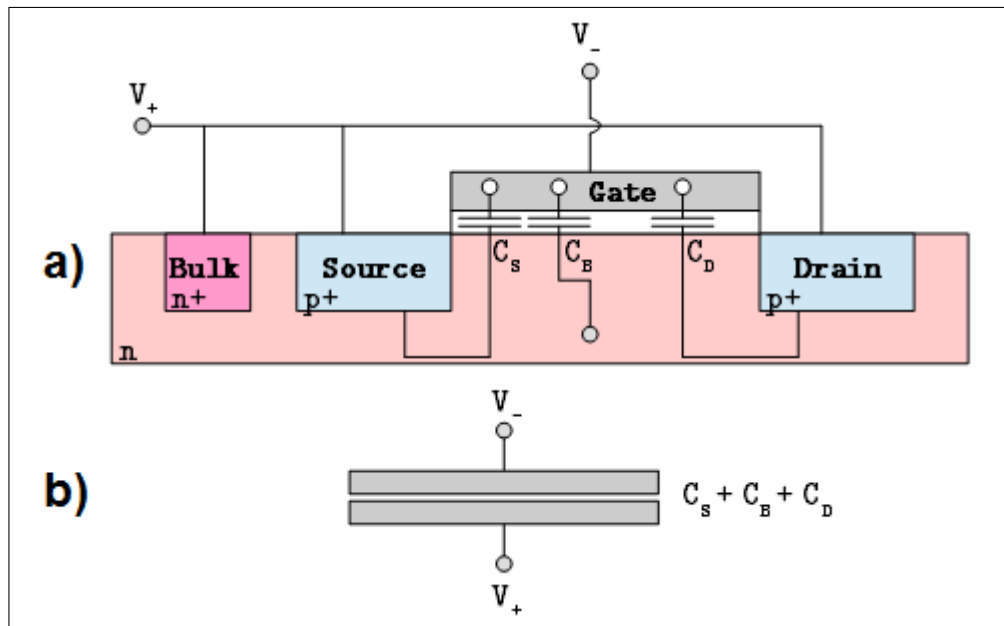


Abbildung 2.2: a) Querschnitt durch den pMOS mit entsprechenden Dotierungen. Bezüglich des Gates stellen die Bereiche der Drain-, Source- und Bulkregion Kapazitäten dar. Liegen alle drei Anschlüsse auf dem gleichen Potential V_+ , so kann das Verhalten der Schaltung anhand ihrer Ersatzschaltung b) untersucht werden.

mus gestattet es also den auf dem Floating-Gate gespeicherten Analogwert in Form des Potentials V_{FG} auszulesen. Die quantitative Abhängigkeit des Drain-Stroms vom Gate-Potential wird in Kap. 3 durch Simulation gewonnen.

2.2 Beschreiben des Floating-Gates

Um das Gate-Potential ohne direkten Zugriff nachhaltig zu verändern werden die quantenmechanischen Eigenschaften des Elektrons als Ladungsträger verwendet.

Die Lösung der Schrödingergleichung liefert das theoretische Konzept für Elektronen Potentialbarrieren zu überwinden, deren Höhe die kinetische Energie der Ladungsträger übersteigt und klassisch eine Reflexion des Teilchens zur Folge hätte. Eine Potentialdifferenz zwischen Floating-Gate und dem Substrat der pMOS führt daher je nach Vorzeichen der Spannung zu einer Zu- oder Abnahme der Ladung auf dem Gate und damit zu einer Änderung des Potentials.

Die pMOS – Floating-Gate Anordnung stellt einen Halbleiter-Oxid-Halbleiter Übergang dar weshalb die genauen physikalischen Zusammenhänge wie Ladungs- und elektrische Feldverteilung innerhalb des Materials von den Dotierungsdichten und dem Betrag der Spannung abhängen ([7]). Exakte quantitative Aussagen über den Tunnelstrom können daher nicht ohne erheblichen Aufwand gemacht werden.

Aus diesem Grunde wird die Annahme der exponentiellen Abhängigkeit der Tunnelstromdichte von der anliegenden Spannung verwendet ([3]). Genauere Untersuchungen ([1]) zeigen einen komplizierteren Zusammenhang. Für die Zwecke dieser Arbeit ist es allerdings völlig hinreichend festzuhalten, dass die Abhängigkeit des Tunnelstroms von der Spannung wesentlich stärker ist als von der Transistorfläche. Diese Aussagen gelten für beliebig orientierte Spannungen.

Um einen Floating-Gate Speicher zu beschreiben wird von der unterschiedlichen Dimensionierung der beiden pMOS-Transistoren Gebrauch gemacht. Die Gateflächen unterscheiden sich um ein Vielfaches was zu einem großen Unterschied zwischen den Kapazitäten führt. Der größere der beiden Transistoren sei mit C_L bezeichnet, der kleinere mit C_S und die jeweiligen Potentiale der von außen zugängliche Kontakte mit V_L bzw. V_S , entsprechend Abb. 2.1.

Wird V_S auf ein hohes Potential gebracht, während V_L geerdet ist, so verhindert die große Kapazität C_L , dass das Gate-Potential V_{FG} dem anliegenden Potential V_S folgt. V_{FG} koppelt also sehr viel stärker an C_L als an C_S . Man erhält somit eine Spannung $U = V_S - V_{FG}$ an der kleinen Kapazität C_S , wobei V_S auf höherem Potential liegt. Dies führt zum Abfluß von Elektronen vom Gate und daher zur Erhöhung des Potentials; das Gate wird geladen.

Zum Entladen wird das Potential V_L am großen pMOS erhöht während V_S auf dem Massepotential liegt. Die starke kapazitive Kopplung des Floating-Gates an C_L sorgt dafür, dass V_{FG} dem angelegten Potential folgt. Als Resultat erhält man wiederum eine Spannung $U = V_{FG} - V_S$ an der kleinen Kapazität C_S , allerdings mit umgekehrtem Vorzeichen. Elektronen fließen auf das Gate und entladen es.

Die Spannungsverhältnisse im Moment des Erzeugens eines Potentials können durch einfaches Anwenden der Ladungserhaltung aus der Schaltung in Abb. 2.1 gewonnen werden. Befindet sich auf dem Gate die Ladung Q so ist das Gate-Potential durch

$$V_{FG} = \frac{V_S \cdot C_S + V_L \cdot C_L + Q}{C_S + C_L} \quad (2.1)$$

gegeben. Die wirkende Programmiervoltage U ist daher im Falle des Ladens des Gates

$$U_{Laden} = V_S - V_{FG} = V_S \cdot \left(1 - \frac{C_S}{C_S + C_L}\right) - \frac{Q}{C_S + C_L}.$$

Der Tunnelstrom steigt daher mit erhöhtem Verhältnis C_L/C_S und nimmt für sonst feste Parameter mit steigender Ladung Q , also mit höherem Gate-Potential V_{FG} , ab. Im Falle des Entladens ist

$$U_{Entladen} = V_{FG} - V_S = \frac{V_L \cdot C_L + Q}{C_S + C_L};$$

Die Ladespannung steigt mit größerem Kapazitätsverhältnis C_L/C_S sowie mit größerem Gate-Potential V_{FG} bzw. Ladung Q .

Durch die beschriebenen Prozesse kann V_{FG} über die Dauer und Höhe der angelegten Spannungen theoretisch stufenlos festgelegt werden. In der Praxis werden wiederholte Programmierpulse fester Höhe zum Programmieren verwendet.

Es ist zu beachten, dass während des Ladens bzw. Entladens das Vorzeichen der sehr viel kleineren Spannung an C_L relativ zu C_S negativ ist, was zu Tunnelströmen in die Gegenrichtung führt. Die Tatsache, dass der Tunnelstrom wesentlich stärker von der Potentialdifferenz abhängt als von der Fläche stellt aber sicher, dass das Laden und Entladen bei passender Dimensionierung der Kondensatorfläche wie gewünscht stattfindet.

Ein anschauliches Bild der beschriebenen Vorgänge liefert die Betrachtung aus Sicht der Ladungsverteilung auf dem Gate. Behandelt man die Transistoren wie gewöhnliche Plattenkondensatoren, so ist bspw. um C_L auf ein hohes Potential zu bringen eine gewisse positive Ladung nötig die auf die äußere Kondensatorplatte gebracht wird (Abb. 2.3). Auf dem Gate fließen nun so lange Elektronen auf die gegenüberliegende Platte bis das elektrische Feld im Gate kompensiert ist. Die Ladungserhaltung sorgt für eine

entgegengesetzte Ladung auf dem anderen Plattenpaar und damit für ein elektrisches Feld an C_S mit umgekehrtem Vorzeichen. Das elektrische Feld eines Plattenkondensators ist der Ladungsträgerdichte auf den Platten proportional, sodaß auf dem kleinen Kondensator ein entsprechend dem Flächen- bzw. Kapazitätsverhältnis stärkeres Feld zu einem höheren Tunnelstrom führt. Der Fall des Ladens des Gates ist völlig analog mit negativem Vorzeichen der elektrischen Felder.

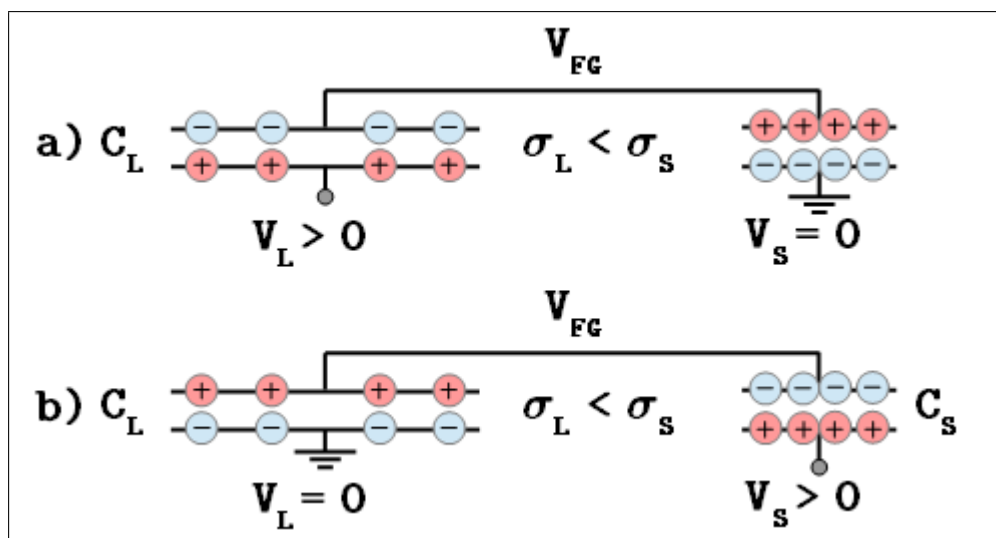


Abbildung 2.3: Ladungsverteilung der Speicherzelle beim Entladen (a) und Laden (b) des Gates. σ bezeichnet den Betrag der Ladungsdichte an den jeweiligen Kondensatoren. Die hier dargestellten Ladungsverhältnisse gehen für eine simple Darstellung von einem neutralen Gate aus.

In der vorhergehenden Diskussion der Programmierspannungen wurde die Rolle des Auslesetransistors ignoriert. Dieser besitzt eine eigene Kapazität und hat damit, je nach seiner Größe und den anliegenden Potentialen, eine nicht zu vernachlässigende Rolle. Sein Einfluß auf die Untersuchungen dieser Arbeit wird in den folgenden Kapiteln bei Bedarf erläutert. In der üblichen Verwendung von Floating-Gate Speicher wird der Auslesetransistor derart dimensioniert, dass sein Einfluss auf die Schaltung ignoriert werden kann ([8]).

2.3 Auslesen des Wertes

Um einen gespeicherten Analogwert abzurufen muss ein Auslesemechanismus implementiert werden. Sollen Spannungen bspw. an Neuronschaltungen übergeben werden, so wird der Auslesetransistor als Teil eines Sourcefolgers konzipiert der im Wesentlichen das Gate-Potential reproduziert. Im Falle von Strömen als Parameter kann der Strom des Auslesetransistors direkt verwendet werden. Die extrahierten Größen können

mittels weiterer Schaltungen nach Bedarf skaliert werden. V_L und V_S sind während des Auslesens geerdet. Genaueres zur Extraktion von Analogwerten aus dem Floating-Gate Speicher ist in [4] dargestellt. Für die Zwecke dieser Arbeit wird der gespeicherte Analogwert dem Gatepotential gleichgesetzt und durch Messen des Stromes durch den Auslesetransistors abgerufen.

2.4 Speicherzeit

Wurde das Floating-Gate geladen, so existiert im normalen Lesebetrieb zwischen den Kapazitäten und dem Gate eine Spannung die zu parasitären Tunnelströmen führt. Dieser Leckstrom, der eine Entladung des Gates verursacht, steigt exponentiell mit dünner werdendem Gate-Oxid der beteiligten Transistoren ([1]).

Dieses ungewünschte Tunneln von Ladung durch das Gate-Oxid der Transistoren lässt sich grundsätzlich nicht vermeiden. Es muss daher die Frage nach der Zeitdauer gestellt werden, während der ein Floating-Gate seinen gespeicherten Wert beibehält.

Die Benutzereingabe von Parametern erfolgt in der Regel digital. Die Werte werden durch Digital-Analog Konverter in analoge Signale umgewandelt und an eine Programmierschleife übergeben. Diese besteht aus einem Regelkreis, der Soll- und Istwert vergleicht und bei Diskrepanz je nach Bedarf einen Lade- oder Entladepuls an den Speicher weitergibt. Dies wird wiederholt bis der Abgleich erfolgreich ist (Genauerer s. [4]).

Digital-Analog Konverter besitzen allerdings eine endliche Auflösung; üblich sind 10 Bit ([8]). Der Betriebsbereich des Floating-Gate Potentials wird also in 1024 verschiedene Werte unterteilt. Definitionsgemäß ist die Speicherzeit daher durch die Dauer bestimmt, in der das Gate-Potential bzw. der Auslestestrom seinen ursprünglichen Wert beibehält.

Es ist zusätzlich zu erwähnen, dass sich die Fläche der Kapazitäten nicht auf die Speicherzeiten auswirkt, da der Quotient aus Ladung und Fläche näherungsweise konstant ist. Mit steigender Fläche steigt der Tunnelstrom gleichermaßen wie die Ladung, die vorher zum Erreichen des Potentials auf das Gate gebracht werden muss.

Kapitel 3

Experimenteller Aufbau

Dieses Kapitel beschreibt die verwendeten Floating-Gate Speicherzellen und untersucht die auftretenden Programmierspannungen anhand von Simulationen. Anschließend wird der Aufbau zur Erzeugung von Programmierpulsen erklärt, die für spätere Messungen verwendet werden. Alle weiteren Messmethoden werden im Rahmen der Messungen selbst erläutert.

3.1 Testchip mit Floating-Gate Speicherzellen

Zur Untersuchung stehen einige Testchips mit jeweils mehreren Floating-Gate Speicherzellen zur Verfügung. Alle für die Testzellen verbauten Transistoren sind für eine Gate-Oxid Spannung von höchstens 2.5 V konzipiert. Zuvor durch Simulation untersuchte Modelle mit dünnerem Gate-Oxid für 1.2 V Spannungen zeigen einen zu hohen Tunnelstrom als dass sie für analogen Floating-Gate Speicher verwendet werden könnten ([8]). Für die in dieser Arbeit untersuchten Zellen liegen allerdings keine Simulationsmodelle vor, sodass experimentelle Untersuchungen durchgeführt werden.

Die Messungen werden an zehn Zellen auf zwei verschiedenen Chips durchgeführt. Abb. 3.1 zeigt einen Testchip mit entsprechenden Pads zum Anschluss von V_L , V_S und V_{DD} . Der Aufbau der Zellen ist wie in Abb. 2.1 bereits gezeigt.

Fünf der Testzellen besitzen ein doppelt so großes C_L um den Einfluss verschiedener Kapazitätsverhältnisse C_L/C_S untersuchen zu können. Die größere Variante wird mit Typ I bezeichnet, die kleinere mit Typ II.

Die Auslesetransistoren werden bei $V_{DD} = 2.5$ V Drain-Source Spannung betrieben. Sie sind derart dimensioniert, dass bei einem Gate-Potential von $V_{FG} = 2.5$ V ein Drain-Source Strom von $I_{DS} = 1$ mA fließt. Dies sind gleichzeitig die zur Verwendung vorgesehenen Maximalwerte der jeweiligen Größe. Mit einer Auflösung von 10 bit ist die Schrittweite der Analogwerte daher ca. 2.5 mV. Abb. 3.2 zeigt den Verlauf des Stroms durch den Auslesetransistor in Abhängigkeit des Gatepotentials. Diese Kurve dient zum Umrechnen des gemessenen Stroms auf den entsprechenden Analogwert V_{FG} .

Die verwendete Programmierspannung des Vorgängermodells in 180 nm Prozesstechnik liegt bei 11 V. Aufgrund des etwas dünneren Gate-Oxids werden daher für die vor-

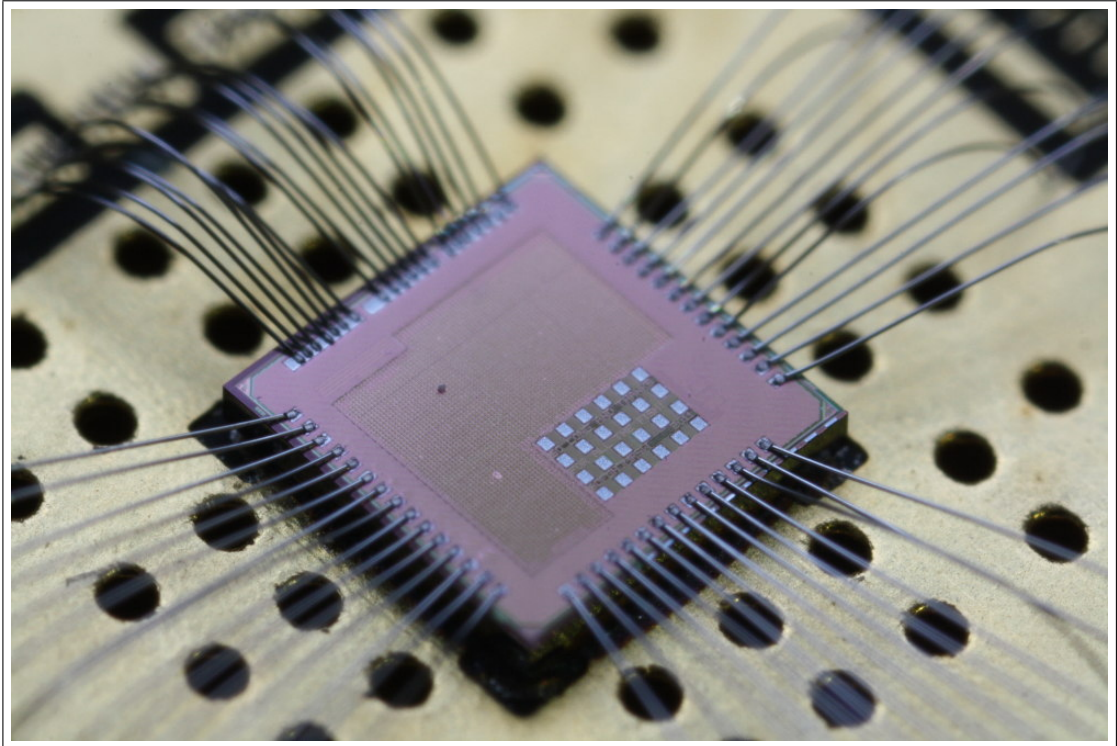


Abbildung 3.1: Archivfoto eines Testchips mit enthaltenen Floating-Gate Zellen. Einige der freien Pads stellen die Kontakte für Auslesetransistor, C_L und C_S dar.

liegenden Floating-Gate Zellen die Untersuchungen bei 10 V begonnen.

Die Dimensionierung des Auslesetransistors richtet sich bei den verwendeten Testschaltungen nach einem hohen und damit gut messbaren Ausgangsstrom. Die Größenordnung der Ströme in den großskalig verwendeten Floating-Gates liegt im Bereich von einigen Mikroampere. Tab. 3.1 stellt die Gateflächen aller drei Transistoren dar.

	Fläche C_L [μm^2]	Fläche C_S [μm^2]	Fläche Auslesetransistor [μm^2]
Typ I	5.12	0.32	10.24
Typ II	2.56	0.32	10.24

Tabelle 3.1: Gateflächen der Transistoren der Floating-Gate Zelle von Typ I und Typ II.

Aus der Tabelle geht hervor, dass die Fläche des Auslesetransistors und damit einhergehend auch seine Kapazität (i. F. als Auslesekapazität bezeichnet) wesentlich größer ist als die der beiden Programmiertransistoren. Aus diesem Grunde muss sein Einfluss auf die Schaltung untersucht werden.

Genaue, quantitative Abschätzungen über die Auswirkungen lassen sich nicht ohne

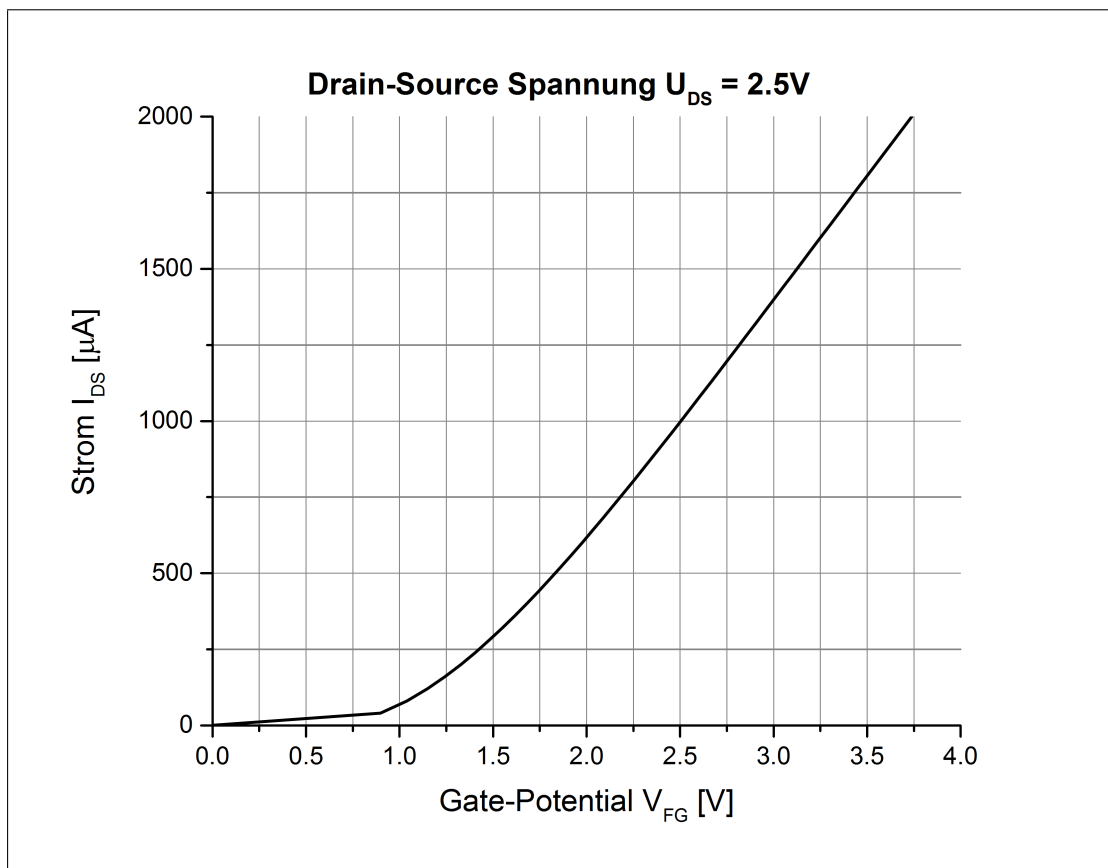


Abbildung 3.2: Simulation des Drain-Stroms des Auslesetransistors in Abhängigkeit des Gate-Potentials bei 2.5 V Versorgungsspannung.

Weiteres machen, da die Kapazitäten der Transistoren, insbesondere die des Auslesetransistors abhängig von den vorliegenden Potentialen ist ([7]). Die genaue Ladungsverteilung im Oxid und Silizium der Transistoren ist von den wirkenden elektrischen Kräften abhängig und das Substrat des Auslesetransistors ist bei anliegendem Drain-Potential nicht äquipotent, da hierüber die Drain-Source Spannung abfällt.

Zur Vereinfachung der Diskussion an späterer Stelle sei, wenn von der Kapazität des Auslesetransistors die Rede ist, immer jene Kapazität gemeint, die ein Plattenkondensator an seiner Stelle besitzen müsste um die selbe kapazitive Kopplung auf das System zu haben. Hiermit ist der effektive Mittelwert über die Gate-Fläche des Auslesetransistors gemeint. Ein solcher Effektivwert muss daher, wie eben begründet, vom Gate-Potential V_{FG} anhängig sein. Analog hierzu wird auch vom Potential des Auslesetransistors gesprochen (nicht zu verwechseln mit dem Drain-Potential V_{DD}). Über das effektive Potential lässt sich aussagen, dass dessen Wert zwischen 0 V und dem Drain-Potential V_{DD} liegen muss.

Um quantitative Aussagen über die auftretenden Gate-Potentiale machen zu können wird auf Simulationen zurückgegriffen. Die verwendete Simulationssoftware ist *Cadence VIRTUOSO* wobei die industriell gängigen Simulationsmodelle (*BSIM 4.5*) eingesetzt werden. Hiermit kann die Wirkung des Auslesetransistors untersucht werden.

Einen deutlichen Effekt besitzt dieser auf den Entladevorgang. Abb. 3.3 zeigt in der Simulation die anliegende Entladespannung $U_{Entladen}$ bei $V_L = 10\text{ V}$ Entladepulsen für eine Typ II-Zelle (kleines C_L) mit bzw. ohne Auslesekapazität. Bei einem anfänglichen Gate-Potential V_{FG} von 1 V erhält man unter Vernachlässigung des Auslesetransistors eine Programmierspannung von $U_{Entladen} = 9.45\text{ V}$, für das reale Setup mit Auslesetransistor ist dagegen $U_{Entladen} = 3.16\text{ V}$. Die Diskrepanz zwischen dem idealen Modell und den real vorliegenden Testzellen ist also enorm. Die Simulation mit einer Typ I Zelle ändert dies nicht erwähnenswert.

Diese Unterschiede resultieren aus der Tatsache, dass bei dem Versuch, das Gate zu entladen, der starken kapazitiven Kopplung des Floating-Gates an das hohe Potential V_L eine stärkere Kopplung hin zum wesentlich niedrigeren Potential am Auslesetransistor überlagert wird (10 V an C_L im Vergleich zu Werten kleiner als dem Drain-Potential von 2.5 V am Auslesetransistor). Das resultierende Gate-Potential fällt daher deutlich kleiner aus was eine kleinere Programmierspannung an C_S zur Folge hat. Die Entladespannungen sind also deutlich geringer und es ist mit längeren Entladezeiten zu rechnen als im idealen Fall.

Beim Ladevorgang sorgen die große Kapazität und das niedrige Potential des Auslesetransistors dafür, dass das Gate-Potential, das ohnehin durch die Kopplung an C_L niedrig gehalten wird, nie dem anliegenden Potential V_S nahe kommt. Abb. 3.4 zeigt für eine Typ II Floating-Gate Zelle, ausgehend von 1 V Gate-Potential die Ladespannung mit und ohne Auslesekapazität. Es zeigt sich, dass die große Auslesekapazität in der Simulation sogar zu einer etwas höheren Programmierspannung führt. Es ist also zu erwarten, dass der Ladevorgang bei den vorliegenden Zellen nicht gestört wird.

3.2 Setup zum Erzeugen von Programmierpulsen

Zum Generieren von Programmierpulsen einstellbarer Länge und Höhe wird eine Schaltung bestehend aus FPGA¹, Spannungswandler und Operationsverstärker (OP) verwendet. FPGAs sind integrierte Schaltkreise, die als programmierbare logische Schaltungen eingesetzt werden. Sie eignen sich gut um Spannungspulse festdefinierter Dauer zu erzeugen. Der FPGA ist auf einem entsprechenden Entwicklungsboard montiert welches Schnittstellen zur externen Kommunikation besitzt und einige Drucktasten und Schalter zur Verfügung stellt.

Die Längen üblicher Programmierpulse liegen im Bereich zwischen Mikro- und eini-

¹Field Programmable Gate Array

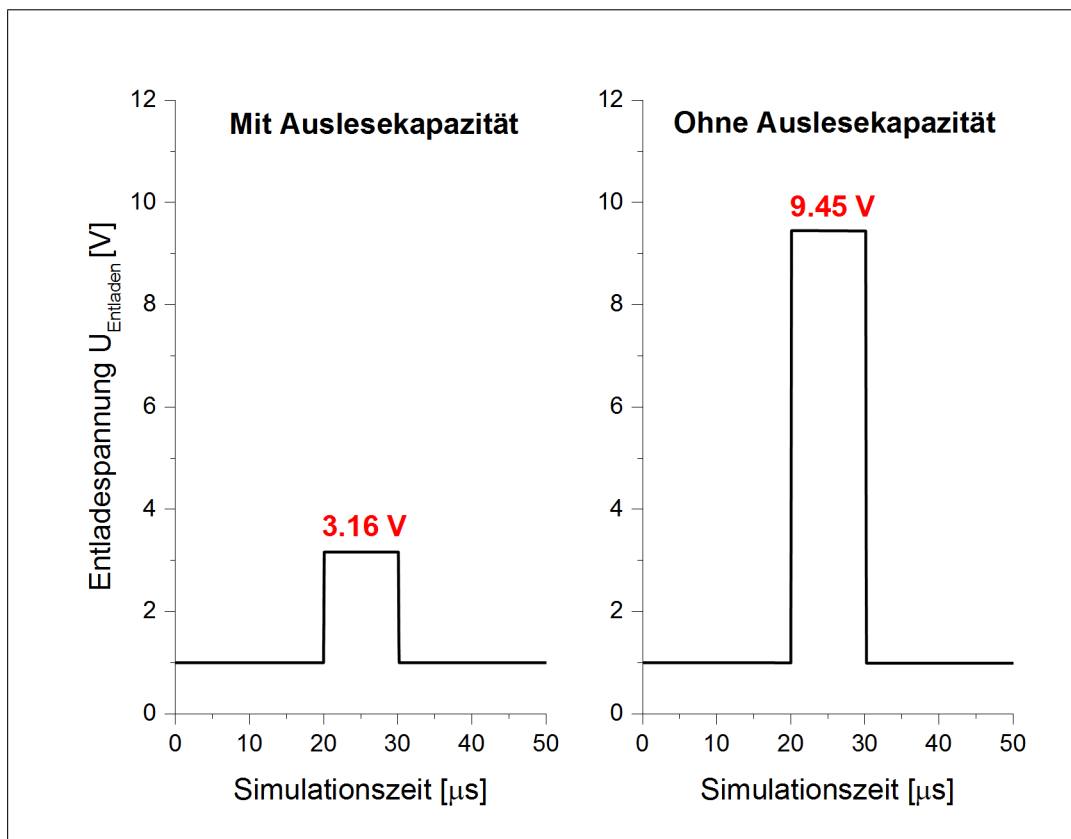


Abbildung 3.3: Simulation (Typ II) der resultierenden Entladespannung bei 10 V Pulsen mit realistischer (links) bzw. idealer (rechts) kapazitiver Kopplung durch den Auslesetransistor. Das anfängliche Potential des Floating-Gates liegt bei 1 V. Der Puls beginnt bei 20 μs und dauert 10 μs an.

gen Zehn Nanosekunden ([8]) und müssen daher elektronisch generiert werden. Die hohe Taktrate von 100 MHz des FPGA gestattet diese Zeitaufösung und ist exakt. Bei Verwendung von Computern kann es vorkommen, dass während einiger Prozessor-takte Hintergrundberechnungen durch das Betriebssystem durchgeführt werden. Diese werden zur Laufzeit der Software nicht registriert und führen schlussendlich zu Ungenauigkeiten in der Zeitmessung.

Um einen FPGA verwenden zu können muss dieser vorher entsprechend konfiguriert werden. Das hierfür erstellte VHDL-Design² ermöglicht es auf Knopfdruck einen Rechteckpuls zu erzeugen, sowie dessen Länge anhand von Schaltern auf dem Board festzulegen.

²Im Zusammenhang mit FPGAs wird nicht von „Programmen“ gesprochen, da es sich nicht um übliche sequenzielle Software handelt. VHDL ist die verwendete Hardwarebeschreibungssprache.

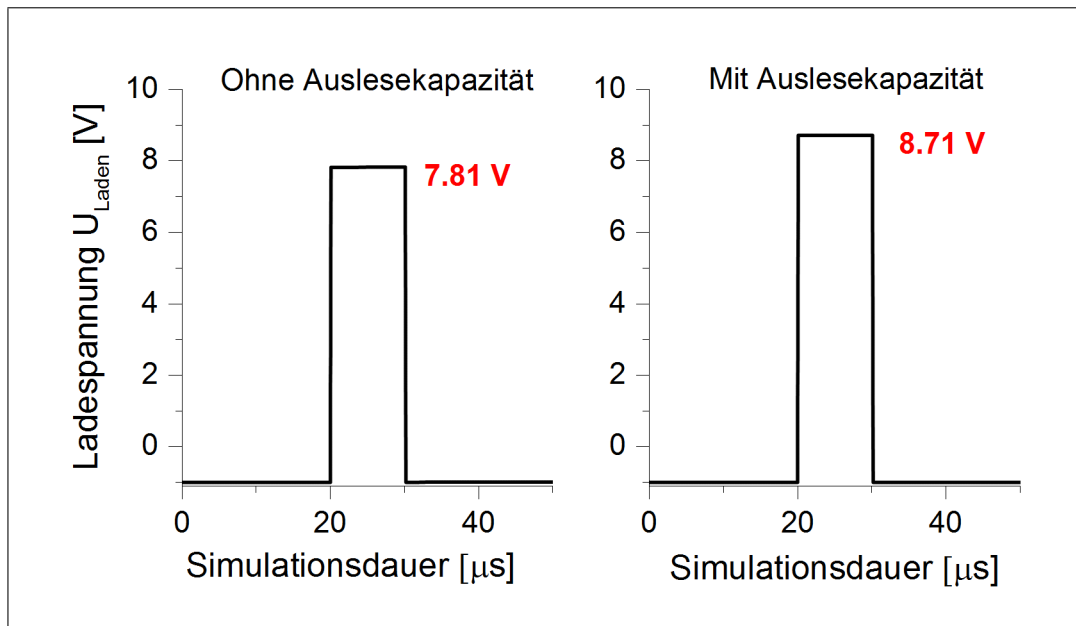


Abbildung 3.4: Simulation (Typ II) der resultierenden Ladespannung bei 10 V Pulsen mit idealer (links) bzw. realistischer (rechts) kapazitiver Kopplung durch den Auslesetransistor. Das anfängliche Potential des Floating-Gates liegt bei 1 V. Der Puls beginnt bei $20 \mu\text{s}$ und dauert $10 \mu\text{s}$ an.

Ein FPGA besitzt als digitale Logikeinheit nur die Ausgangssignale High-State („1“) und Low-State („0“) welche als $V_1 = 2.5 \text{ V}$ bzw. $V_0 = 0 \text{ V}$ an einem Logik-Pin abgegriffen werden können. Um wirksame Programmierspannungen zu erzeugen wird das Ausgangssignal des FPGA daher lediglich als Steuersignal eines Operationsverstärkers an dessen nicht-invertierenden Eingang verwendet. Der invertierende Ausgang wird auf $1/2 V_1 = 1.25 \text{ V}$ fixiert. Wird nun ein Spannungspuls am FPGA erzeugt, so ist das Potential des nicht-invertierenden Eingangs höher als das des invertierenden und am Ausgang des Operationsverstärkers liegt seine Versorgungsspannung an. Außerhalb des Pulses sind die Verhältnisse umgekehrt und man erhält das Massepotential am Ausgang. Die Spannungspulse werden also auf die Höhe der Versorgungsspannung des OP verstärkt und können anschließend an den Chip zum Programmieren der Floating-Gates weitergeleitet werden.

Abb. 3.5 stellt den entsprechenden Schaltplan dar. Ein Labornetzteil versorgt einen linearen Spannungsregler (*LM 317-220*) mit einer Spannung von $V_{\text{Netzteil}} = 16 \text{ V}$. Der Output (V_{out}) des Spannungswandlers kann über ein Potentiometer (R_2) kontrolliert werden und stellt die Versorgung für einen Rail-To-Rail Operationsverstärker (*TS 912IN*) dar. Ein Spannungsteiler an einem 2.5 V-Versorgungspin des FPGA liefert die gewünschten 1.25 V am invertierenden Eingang. Aus Gründen der einfacheren Bedienbarkeit wird für C_L und C_S jeweils ein separater Operationsverstärker mit entsprechenden FPGA-Pins verwendet (nicht in Abbildung enthalten).

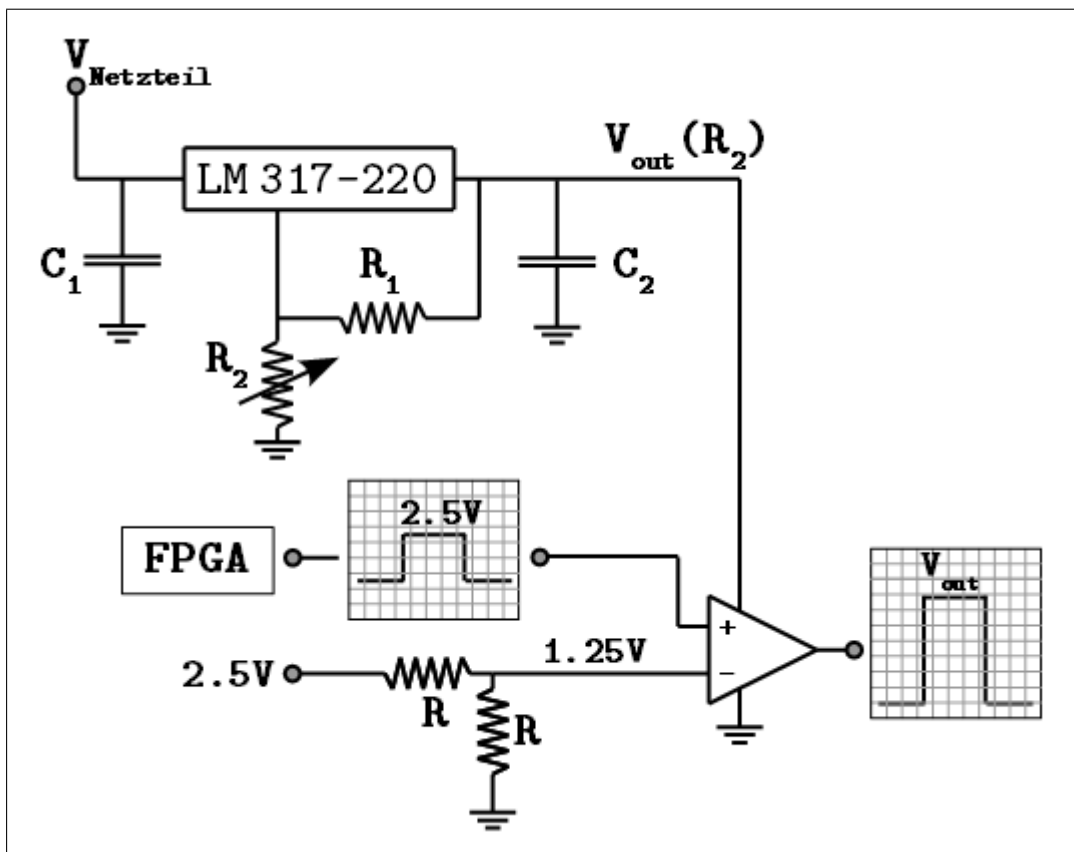


Abbildung 3.5: Schaltplan des Aufbaus zur Erzeugung von Rechteckpulsen. Ein linearer Spannungsregler (*LM 317-220*) liefert die durch ein Potentiometer (R_2) einstellbare Versorgungsspannung V_{out} eines OP, der die von einem FPGA erzeugten Rechtecksignale auf diesen Wert verstärkt.

Kapitel 4

Messungen und Ergebnisse

4.1 Lade- und Entladezeitmessungen

4.1.1 Messmethode

Das Speichern eines analogen Wertes auf dem Floating-Gate erfolgt wie beschrieben mittels eines Programmierzyklus der durch Vergleich zwischen Ist- und Sollwert wiederholt je nach Anforderung Lade- oder Entladepulse an das Floating-Gate gibt. Ein solcher Regelkreis steht für diese Arbeit nicht zur Verfügung, sodass zur Abschätzung der Schreibzeit des Floating-Gate Speichers jene Dauer bestimmt wird, die benötigt wird um das Gate-Potential von ihrem Minimum auf den Maximalwert bzw. in die Gegenrichtung zu laden. Hierbei wird die Länge der Pulse angepasst bis ein passender Wert gefunden ist.

4.1.2 Resultate

Die Messungen ergeben Lade- und Entladezeiten die drastisch von der erwarteten Größenordnung abweichen. Die benötigte Pulsdauer zum vollständigen Laden des Gates auf $V_{FG} = 2.5 \text{ V}$ bei 10 V Pulshöhe liegt im Bereich von Sekunden. Diese Zeiten werden üblicherweise benötigt um einige Tausend Floating-Gates des im 180 nm -Prozess gefertigten Wafersystems (s. Einleitung) zu beschreiben ([8]). Es wäre zu erwarten, dass bei den vorliegenden Floating-Gates aufgrund der geringeren Dicke des Gate-Oxids die Schreibdauer geringer ist.

Darüberhinaus wird beobachtet, dass die Floating-Gates mit kleinem C_L nicht unterhalb eines minimalen Potentials entladen werden können. Dieses Minimum ist für 10 V Pulse etwa 1.7 V und fällt mit fallender Pulshöhe (bei 8 V Pulshöhe auf ca. 1.6 V). Es muss daher davon ausgegangen werden, dass nicht alle Annahmen über die Funktionsweise der vorliegenden Floating-Gates korrekt sind. Der große Auslesetransistor liefert eine teilweise Erklärung für die Beobachtungen:

In Kap. 3 wurden bereits anhand von Abb. 3.3 Auswirkungen des großen Auslesetransistors auf den Entladevorgang beschrieben. Die starke Kopplung des Floating-Gates an die Auslesekapazität führt zu einem niedrigeren Gate-Potential und einer Behinderung des Entladevorgangs. Aus der Abbildung lässt sich darüberhinaus erkennen, dass

das zuvor vernachlässigbare Tunneln in die ungewünschte Gegenrichtung auf der Seite des großen pMOS nicht mehr vernachlässigt werden kann. In der Simulation aus Abb. 3.3 beträgt die Spannung an der großen Kapazität $U_L = 10 - 3.16 \text{ V} = 6.84 \text{ V}$ und ist damit wesentlich höher als die eigentliche Entladespannung $U_{Entladen} = 3.16 \text{ V}$. Es ist also zu erwarten, dass bei dem Versuch das Gate in diesem Zustand zu entladen ein Ladevorgang stattfindet.

Betrachtet man die vorliegende Konfiguration von Kapazitäten und ihren Potentialen so wird dies unmittelbar klar. Nimmt man für eine einfachere Darstellung das Potential der Auslesekapazität als 0 V an, was in Anbetracht der 10 V Programmierpulse eine akzeptable Näherung ist, so besteht die Floating-Gate Zelle aus einer großen Kapazität $C_A + C_S$ und einem relativ hierzu kleineren C_L . Wird nun in der Absicht das Gate zu entladen ein Spannungspuls an V_L gegeben, so erhält man exakt die im vorigen Kapitel zum Laden des Floating-Gates beschriebene Konfiguration: Die kleine Kapazität liegt auf einem hohem Potential, die große Kapazität ist geerdet.

Nichtsdestotrotz wird zum Teil ein Entladen der Zellen beobachtet. Dieser Entladeprozess erklärt sich dadurch, dass es einen Bereich für die Höhe der Programmierpulse an V_L gibt, die das Gate-Potential derart anheben, dass der dadurch erhöhte entladende Tunnelstrom durch den Auslesetransistor und C_S gegenüber dem ladenden Tunnelstrom durch C_L dominiert (Abb. 4.1). Erhöht man die Pulshöhe weiter, so sorgt die starke Kopplung des Gates an die Auslesekapazität dafür, dass sich die Potentialdifferenz an C_L stärker erhöht als am Auslesetransistor und an C_S . Letztendlich übertrifft die exponentielle Abhängigkeit des Tunnelstroms von der Spannung die lineare Abhängigkeit zur Fläche und durch das kleinere C_L fließt mehr pos. Ladung auf das Gate als durch den großen Auslesetransistor abfließen kann. Es gibt daher zu jedem Gate-Potential eine maximale Pulshöhe an V_L oberhalb welcher statt entladen geladen wird. Umgekehrt existiert bei fester Pulshöhe an V_L ein minimales Gate-Potential das nicht unterschritten werden kann.

Inwiefern sich die Größe der großen Kapazität C_L auf die beschriebenen Effekte auswirkt kann Abb. 4.2 entnommen werden. Bei fester Pulshöhe ist die die Kopplung durch das kleinere C_L geringer und so steigt der ungewünschte Tunnelstrom durch C_L schneller an. Bei kleinem C_L ist also auch die maximale Pulshöhe, die für das Entladen verwendet werden kann, niedriger und das minimale Gate-Potential bei fester Pulshöhe größer.

Die Beobachtungen im Rahmen des Entladevorgangs können daher zumindest qualitativ anhand der Rolle des Auslesetransistor beschrieben werden. Unklar bleibt dagegen weshalb die benötigten Ladezeiten derart lang sind und weshalb jene Floating-Gates mit großem C_L komplett entladen werden können anstatt dass sich die untere Potentialgrenze des Gates nur etwas verringert. Diese Befunde werden an dieser Stelle nicht weiter verfolgt, da nicht klar ist, ob die Beobachtungen für eine reale Floating-Gate Speicherzelle mit passender Dimensionierung aller Bauteile überhaupt relevant sind.

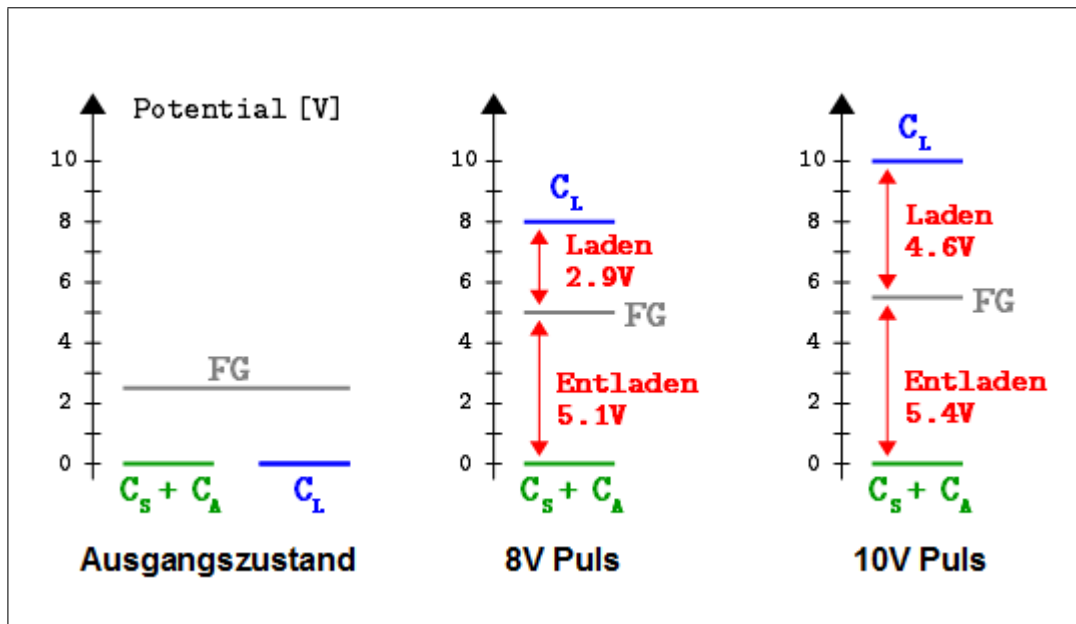


Abbildung 4.1: Simulation (Typ I) der Potentialverhältnisse bei anfänglichem Gate-Potential von 2.5 V für Spannungspulse an V_L von 8 V und 10 V. Für den ersten Fall ist die entladende Spannung am Auslesetransistor und C_S wesentlich größer als die ladende an C_L . Erhöht man die Pulse auf 10 V so nähern sich die Potentiale dem Gleichgewicht, bei dem der Ladestrom durch C_L den Entladestrom durch C_S und den Auslesetransistor ausgleicht. Bei Pulsspannungen darüber überwiegt der Ladevorgang vollständig.

4.2 Speicherzeitmessungen

4.2.1 Messmethode

Da durch den Wechsel zum kleineren 65 nm Herstellungsprozess nicht nur die laterale Strukturgröße verringert wird, sondern auch die Dicke des Gate-Oxids der Transistoren kleiner ist, muss mit kürzeren Speicherdauer der Speicherzellen im Vergleich zum Vorgängermodell gerechnet werden.

Die Speicherzeitmessungen werden für Ströme durch den Auslesetransistor bei 1 mA sowie bei 0.25 mA durchgeführt. Diese Werte entsprechen Abb. 3.2 nach Floating-Gate Potentialen von 2.5 V und ca. 1.4 V. Es werden also Speicherzeiten für den Bereich maximalen sowie mittleren Floating-Gate Potentials gemessen.

Setzt man den dynamischen Verwendungsbereich des Floating-Gate Potentials von 0 – 2.5 V fest, so sind die Messungen der Speicherzeit bei maximalem Wert hinreichend. Es ist klarerweise zu erwarten, dass bei hohen Potentialdifferenzen die Entladung am schnellsten vonstattengeht, sodass die Speicherzeiten durch den Leckstrom bei hohen Potentialen limitiert werden.

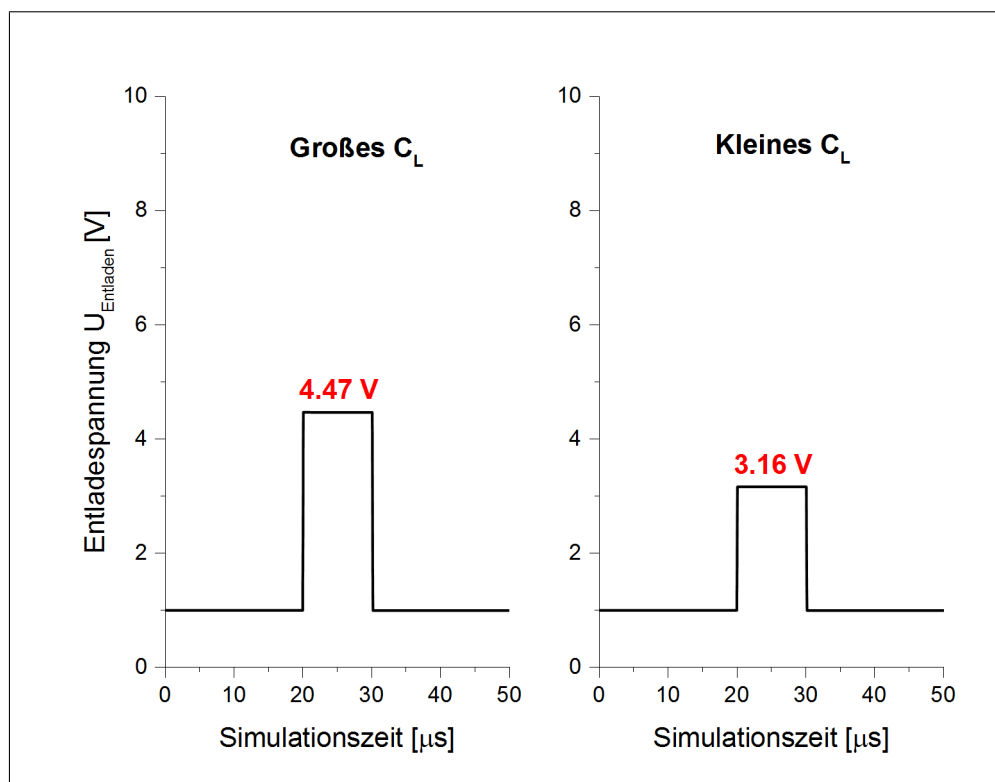


Abbildung 4.2: Simulation der resultierenden Entladespannung bei 10 V Pulsen bei anfänglichem Gate-Potential von 1 V für Typ I und II. Der Auslesetransistor wurde in der Simulation beachtet. Der Puls beginnt bei 20 μs und dauert 10 μs an.

Sollte sich die Speicherzeit für niedrigere Potentiale als genügend lang herausstellen, so könnte sich durch Einschränken des Potentialbereiches hin zu niedrigeren Werten ein Gewinn an Speicherzeit ergeben. Bei gleichbleibender Auflösung bedingt dies allerdings eine feinere Unterteilung des Potentials was eine höhere Messgenauigkeit beim Auslesen des Stromes erfordert und die Schrittweite der Analogwerte verringert. Ob die kleinere Schrittweite der Analogwerte den kleineren Entladestrom vollständig kompensiert, oder ob wirklich bessere Speicherzeiten zu erwarten sind kann durch Messen der Entladezeit ermittelt werden.

Erwartungsgemäß sollte sich die Speicherzeit bei kleinerem Verwendungsbereich des Gate-Potentials erhöhen, da die Schrittweite linear mit dem maximalen Gate-Potential steigt, hingegen der Tunnelstrom aber exponentiell.

Zu diesem Zwecke wird neben dem maximalen Gate-Potential auch bei dem genannten Wert von ca. 1.4 V gemessen. Abb. 3.2 stellt dar, dass bei diesen Potentiale ein hinreichend linearer Zusammenhang zwischen Strom und Potential besteht, was die Extraktion der Speicherzeit aus den gemessenen Strömen vereinfacht. Da die Genauig-

keit der Simulation nicht unmittelbar festzustellen ist, wurden niedrigere Werte nicht gewählt um nicht in den Bereich der Kennlinie zu gelangen, indem ein komplizierterer Zusammenhang zwischen dem Auslesestrom und dem Gate-Potential besteht. Darüberhinaus ist erkenntlich, dass für niedrigere Werte der Strom des Auslesetransistors stark abfällt, was die Genauigkeit der Messungen beschränken würde.

Vor dem Beginn einer Messung wird das Gate-Potential der zu untersuchenden Speicherzelle mittels der in Kap. 3 beschriebenen Schaltung durch wiederholte Programmpulse erhöht, bis der gewünschte Strom durch den Auslesetransistor fließt. Anschließend wird der zeitliche Verlauf des Stromes gemessen.

Dies wird mittels eines Digital-Multimeters (*KEITHLEY 2100 6 1/2 Digit Multimeter*) durchgeführt, welches über die USB-Schnittstelle mittels C++ Software angesteuert wird. Für einen Messpunkt werden 25 Einzelmessungen durchgeführt, deren Werte gemittelt und der statistische Fehler des Mittelwertes bestimmt; die Abweichung der Einzelwerte vom Mittel ist ohne Bedeutung, da diese statistischer Natur ist („Rauschen“) und nicht durch Änderungen des Gate-Potentials verursacht wird. Die benötigte Zeit zum Messen und Übertragen der 25 Einzelwerte ist kleiner als 100 ms sodass die Änderung des Stromes während der Messung vernachlässigt werden kann. Nach einer Wartezeit von etwa 1 s wird der nächste Messpunkt aufgenommen. Dieser Prozess wird wiederholt bis die gewünschte Messdauer verstrichen ist.

Zum Bestimmen der Änderung des Gate-Potentials aus der gemessenen zeitlichen Abhängigkeit des Auslesestromes wird aus der simulierten Transistorkennlinie 3.2 für beide Bereiche des Auslesestromes die Steigung der Strom-Potential Kennlinie mittels linearem Fit entnommen. Aus der Steigung des zeitlichen Stromverlaufes kann auf den zeitlichen Abfall des Gate-Potentials umgerechnet werden. Die Umrechnungsfaktoren K sind wie folgt:

Auslesestrom I_{DS}	Umrechnungsfaktor $K_I := \frac{dV_{EG}}{dI_{DS}}$
1 mA	1267 V/A
0.25 mA	1870 V/A

Die Werte sind durch Simulation gewonnen, die Angabe eines Fehlers entfällt daher. Hier muss aber erwähnt werden, dass die Genauigkeit der Kurve je nach Transistor schwankt. Herstellungsvarianzen und unterschiedliche Temperaturen verändern die Strom-Spannungs Charakteristik der Transistoren was zu Ungenauigkeiten in den gemessenen Speicherzeiten führt. Ziel dieser Messung ist allerdings nur das Ermitteln einer Größenordnung der Speicherzeit; Präzisionsmessungen sind aufgrund des falsch dimensionierten Auslesetransistors zwecklos.

4.2.2 Beispielmessung der Speicherzeit

Der beschriebene Messvorgang wird nun beispielhaft an einer Floating-Gate Zelle bei hohem Gate-Potential durchgeführt.

Abb. 4.3 stellt sechs Messungen des zeitlichen Verlaufes des Auslesestromes von Floating Gate #10 dar. Die Kurven lassen einen Abfall von rund 1 ‰/min erkennen. Bei einer geforderten Auflösung von 10bit entspricht dies also einer Speicherzeit in der Größenordnung von Minuten.

Einen genaueren Wert erhält man durch Berechnung der mittleren Steigung \bar{m} der Graphen. Tab. 4.1 stellt die durch linearen Fit gemessenen Steigungen der einzelnen Kurven dar. Die Fehler der Fits sind in Anbetracht der Variation der Steigungen untereinander zu vernachlässigen. Eine Abschätzung für den Fehler erhält man daher aus der Standardabweichung s , die ebenfalls angegeben ist.

Kurve (Abb. 4.3)	Steigung [$\mu\text{A}/\text{min}$]
1	-0.764
2	-0.725
3	-0.898
4	-0.758
5	-0.740
6	-0.805
Mittelwert \bar{m}	-0.782
Standardabw. s	0.064

Tabelle 4.1: Steigungen der Strommessungen für Floating-Gate #10 (Abb. 4.3). Die Fitgenauigkeiten sind wesentlich geringer als die Standardabweichungen und können vernachlässigt werden.

Der Abfall des Gate-Potentials ist somit

$$\frac{dV_{FG}}{dt} = (\bar{m} \pm s) \mu\text{A}/\text{min} \cdot \frac{10^{-6}\text{A}}{\mu\text{A}} \cdot \frac{1 \text{ min}}{60 \text{ s}} \cdot K_{1mA} = (-16.5 \pm 1.4) 10^{-6} \text{ V/s},$$

woraus eine Speicherzeit Δt von

$$\Delta t = \frac{2.5 \text{ mV}}{dV_{FG}/dt} = (151 \pm 13) \text{ s}$$

folgt.

4.2.3 Resultate

Speicherzeit für hohes Gate-Potential $V_{FG} \approx 2.5 V$

Die beschriebene Prozedur wird für alle zehn Speicherzellen bei hohem Gate-Potential durchgeführt. Die folgende Tabelle stellt die Ergebnisse der Messungen dar:

# Floating-Gate Zelle	Speicherzeit Δt [s]	Ausgewertete Messkurven
1	255 ± 21	5
2	476 ± 39	5
3	143.0 ± 4.6	7
4	–	5
5	327 ± 29	5
6	–	3
7	–	3
8	189.9 ± 5.0	5
9	1600 ± 274	5
10	151 ± 13	6

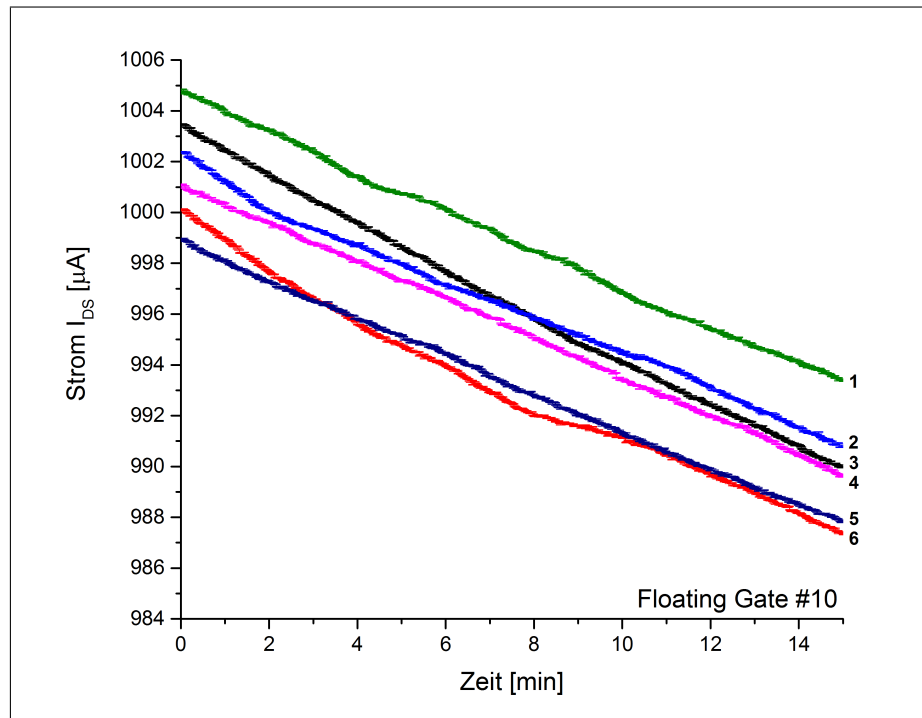


Abbildung 4.3: Messungen des Auslesestromes in Abhängigkeit der Zeit von Floating-Gate #10 bei hohem (hohes Gate-Potential).

Die Speicherzeiten der untersuchten Floating-Gate Zellen erstrecken sich also über einige Minuten bis hin zu einer halben Stunde. Unterschiede zwischen Zellen mit großem und kleinem C_L konnten nicht beobachtet werden.

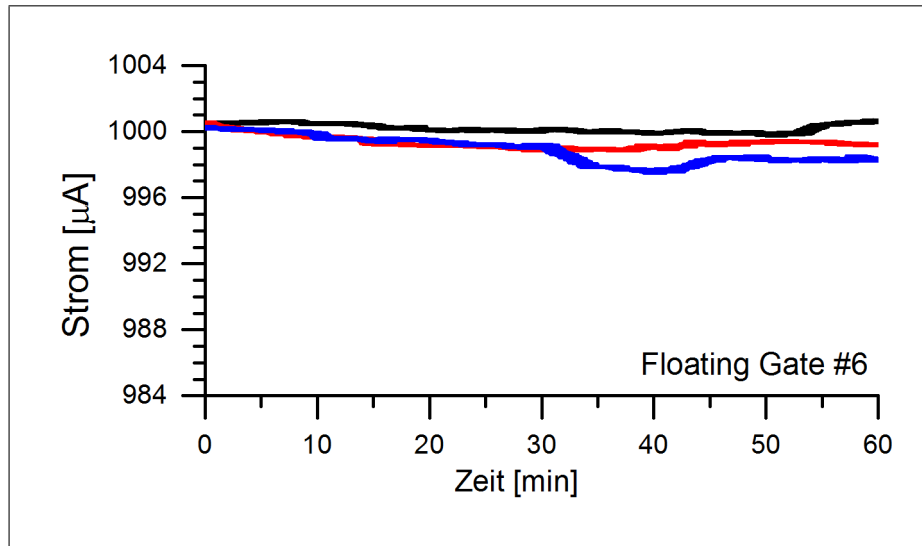


Abbildung 4.4: Messungen des Auslesestromes in Abhängigkeit der Zeit von Floating-Gate #6 (hohes Gate-Potential).

Die Messungen für die Speicherzellen #4, #6 und #7 liefern keine brauchbaren Ergebnisse; auch wiederholte Messungen bei einer Messzeit von 1 h blieben erfolglos. Abb. 4.4 zeigt für Floating-Gate #6 den Verlauf des Auslesestromes für die einstündigen Messungen. Die auch in Abb. 4.3 beobachtbaren Unregelmäßigkeiten im Abfall der Kurve führen bei der hier vorliegenden sehr niedrigen Steigung zu großen Unsicherheiten bei der Datenauswertung die keine sinnvolle Berechnung zulassen.

Genauere Aussagen können aus Langzeitmessungen gewonnen werden, die aber aus Zeitgründen nicht durchgeführt wurden. Die Speicherzeitmessungen werden daher anhand der Messergebnisse der restlichen Zellen diskutiert.

Speicherzeit für mittleres Gate-Potential $V_{FG} \approx 1.4 V$

Für die Speicherzeitmessungen bei mittlerem Gate-Potential stehen aufgrund der in Kap. 4.1 beschriebenen Beobachtungen nur die Hälfte der vorhandenen Floating-Gate Zellen zur Verfügung, da die restlichen Zellen nicht auf das gewünschte Gate-Potential programmiert werden können. Die Untersuchung der restlichen Speicherzellen gemäß dem bereits angewandten Verfahren liefert sehr unterschiedliche Ergebnisse.

Abb. 4.5 stellt den Verlauf des Auslesestromes zweier Floating-Gate Zellen dar, bei welchen das Messergebnis mit den grundsätzlichen Erwartungen übereinstimmt. Der

Abfall der Messkurve ist kleiner als bei hohem Gate-Potential und die resultierende Speicherzeit ist entsprechend höher. Die Speicherzeiten betragen 15.7 h bzw. 18.2 h und sind damit wesentlich länger als die zuvor bestimmten Werte bei hohem Gate-Potential. Die aus den Fits errechneten relativen Fehler sind kleiner als 1 %.

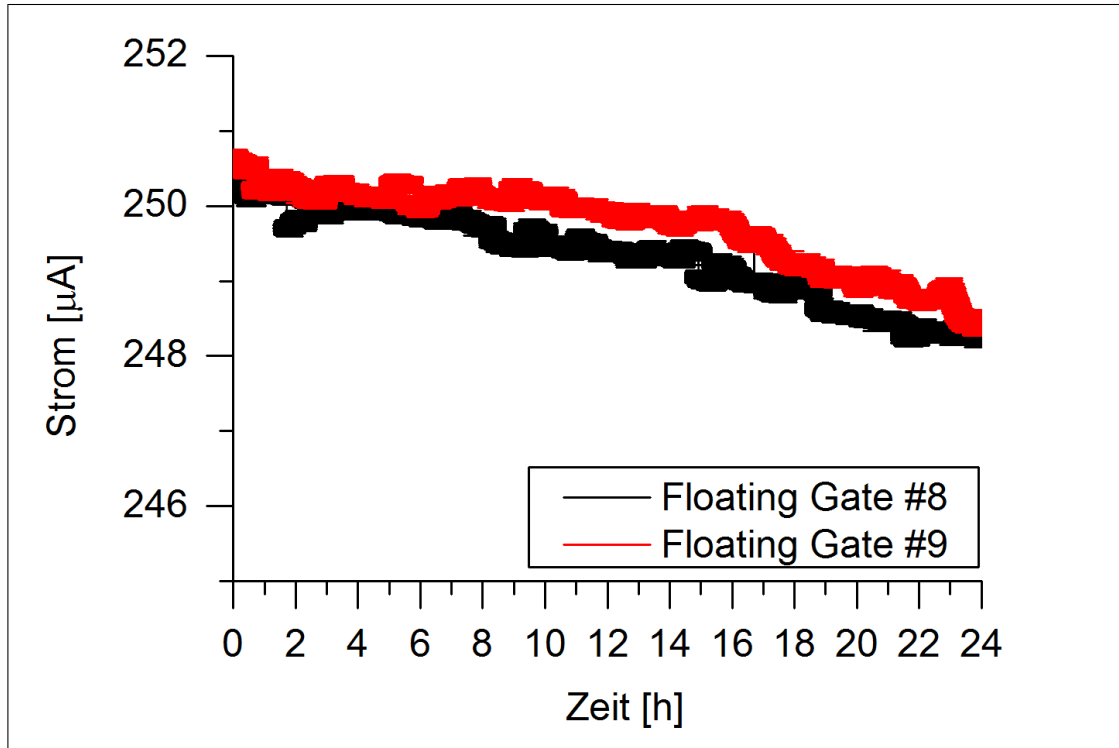


Abbildung 4.5: Messungen des Auslesestromes in Abhängigkeit der Zeit von Floating Gate #8 und #9 (mittleres Gate-Potential). Die Kurve wirkt aufgrund der vielen überlappenden Fehlerbalken und des schmalen dargestellten Stromintervalls dick.

Messungen an den restlichen Zellen zeigen allerdings, dass sich bisher unbeachtete Einflüsse deutlich auf die Konsistenz auswirken. Abb. 4.5 zeigt die Messergebnisse an zwei weiteren Floating-Gates. Hier ist ein deutlicher Anstieg des Auslesestromes zu erkennen dessen Ursprung nicht anhand der üblichen Funktionsweise erklärt werden kann. Es ist insbesondere unklar ob der gemessene Verlauf des Auslesestromes auch die zeitliche Entwicklung des Gate-Potentials wiedergibt.

Der einzig bekannte Weg das Gate-Potential im Auslese-Modus zu erhöhen stellt das Tunneln durch den Auslesetransistor dar. Wie bereits beschrieben kann nicht ohne Weiteres angegeben werden, wie die Feldverteilung entlang des Substrats ist, daher ist die hinreichend genaue Angabe eines effektiven Potentials (siehe Kap. 3) und einer effektiven Kapazität nicht möglich. Sollte das effektive Potential größer sein als das Gate-Potential (ca. 1.4 V) so wird das Gate durch den Auslesetransistor geladen. In diesem Falle driftet das Gate-Potential zu einem Gleichgewicht, bei dem das Laden

durch den Auslesetransistor das Entladen durch C_L und C_S ausgleicht.

In der Tat ähnelt der Verlauf der oberen Messkurve aus Abb. 4.5 einem beschränkten Wachstum gegen eine Schranke. Eine Erklärung dafür, dass dieser Effekt nicht in gleichem Maße an allen Speicherzellen beobachtet wird, sind mögliche herstellungsbedingte Unterschiede zwischen den Zellen, die das vermutete Gleichgewichtspotential des Gates verschieben.

Die beobachteten Unregelmäßigkeiten in den Messungen variieren ebenso für ein und die selbe Speicherzelle. Abb. 4.6 zeigt zwei Messkurven der selben Floating-Gate Zelle die im Abstand von einigen Tagen aufgenommen wurden. Die erste Messung zeigt bis auf kleine Abweichungen einen linearen Abfall des Auslesestromes während die zeitlich nachfolgende Messung einen ansteigenden Verlauf zeigt.

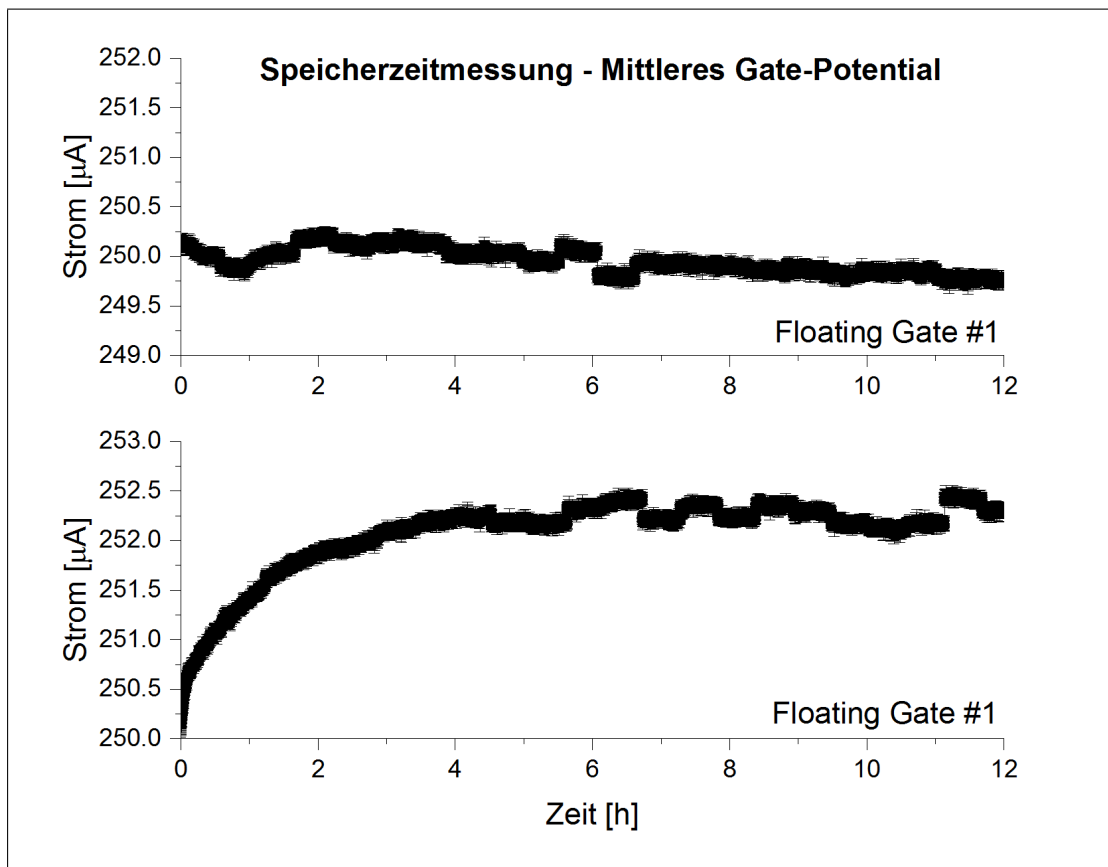


Abbildung 4.6: Messungen des Auslesestromes in Abhängigkeit der Zeit von Floating-Gate #1 (mittleres Gate-Potential). Die Messungen wurden mit einem Abstand von einigen Tagen aufgenommen.

Als weitere Quelle systematischer Abweichungen ist die Temperaturabhängigkeit des

Auslesetransistors zu beachten. Der Strom hängt in einem inversen Verhältnis mit der Temperatur zusammen, wie der Simulation in Abb. 4.7 entnommen werden kann. Die Zunahme des Stromes bei einer Abnahme der Temperatur von drei Kelvin beträgt für den Bereich mittleren Gate-Potentials rund $2.5 \mu\text{A}$. Solche Temperaturänderungen sind durchaus realistisch; das Ausschalten von Messgeräten in der Umgebung der Floating Gates oder das Abkühlen des Labors bei Nacht führen zum Sinken der Raumtemperatur. Der Einfluss der Temperatur liefert daher eine mögliche Erklärung für Schwankungen im Auslesestrom wie bspw. in Abb. 4.6o. innerhalb der ersten zwei Stunden zu beobachten ist und für den Anstieg des Stromes im selben Bild unten unmittelbar nach Beginn der Nachtmessung.

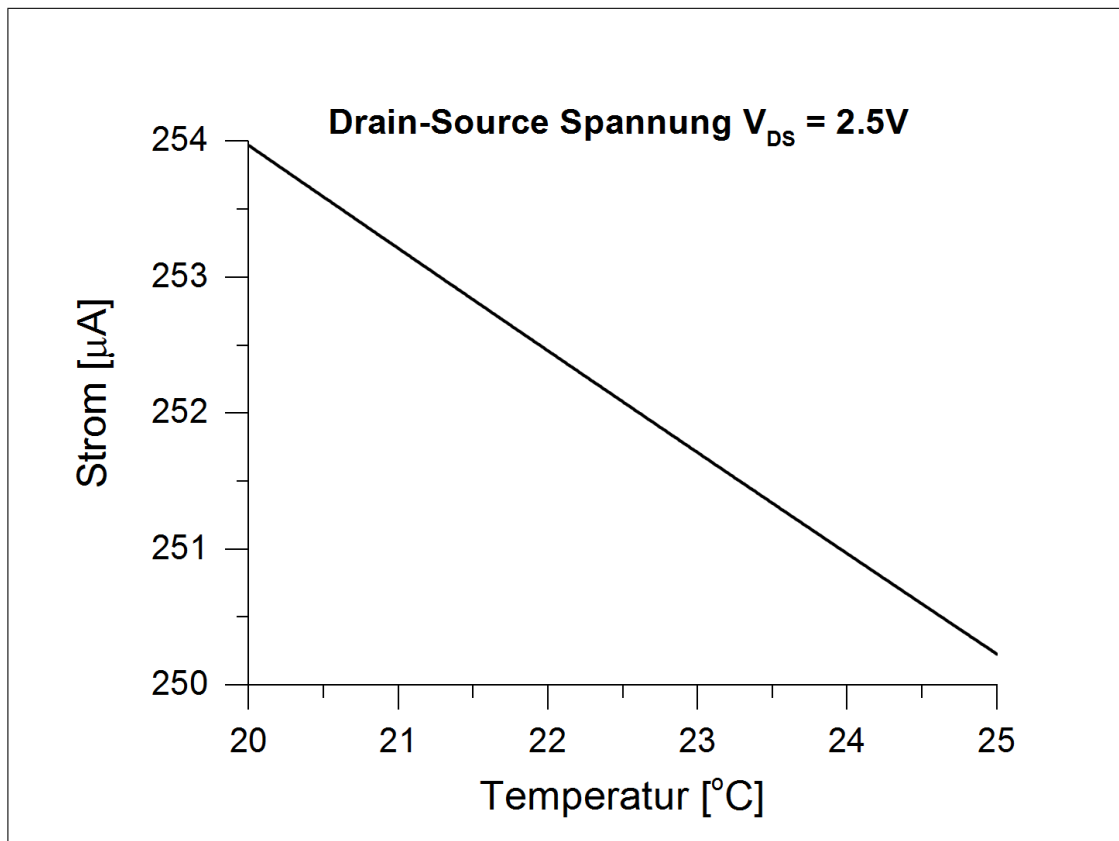


Abbildung 4.7: Simulierte Temperaturabhängigkeit des Auslesestromes (mittleres Gate-Potential). Der Gradient beträgt im dargestellten Bereich rund $-0.8 \mu\text{A}/\text{K}$.

Viele Strommessungen weisen Sprünge im Kurvenverlauf auf (Abb. 4.8). Sofern mehrere dieser Sprünge aufeinanderfolgen steigt und fällt der Auslesestrom im Wechsel und die Zeitdauer zwischen zwei Sprüngen beträgt rund 30 Minuten. Die Höhe variiert stark und reicht von großen Sprüngen von $0.7 \mu\text{A}$ hin zu beliebig kleinen.

Die Tatsache, dass diese Beobachtung gleichermaßen an allen Floating-Gates beobachtet wird, insbesondere dass die Abstände zwischen Sprüngen sich nicht zwischen den Floating-Gates auf verschiedenen Chips unterscheidet, deutet darauf hin, dass der Ursprung dieser Erscheinung außerhalb der Testchips und der Floating-Gates zu suchen ist. Die Spannungsversorgung der Schaltung kann als Quelle ausgeschlossen werden sodass nur das FPGA-Board den Fehler verursachen kann; Genaueres konnte nicht festgestellt werden.

Da keine Präzisionsmessungen durchgeführt werden, wird der Einfluss der Sprünge ignoriert. Die Tatsache, dass die Sprünge in ihrer Richtung symmetrisch sind, begründet die Annahme, dass sie im zeitlichen Mittel eine vernachlässigbare Rolle spielen. Als denkbare Fehlerquelle kann ein Aussenden von unerwünschten Pulsen durch den FPGA als Folge eines Programmierfehlers genannt werden. Aus Zeitgründen wird dies nicht weiter verfolgt.

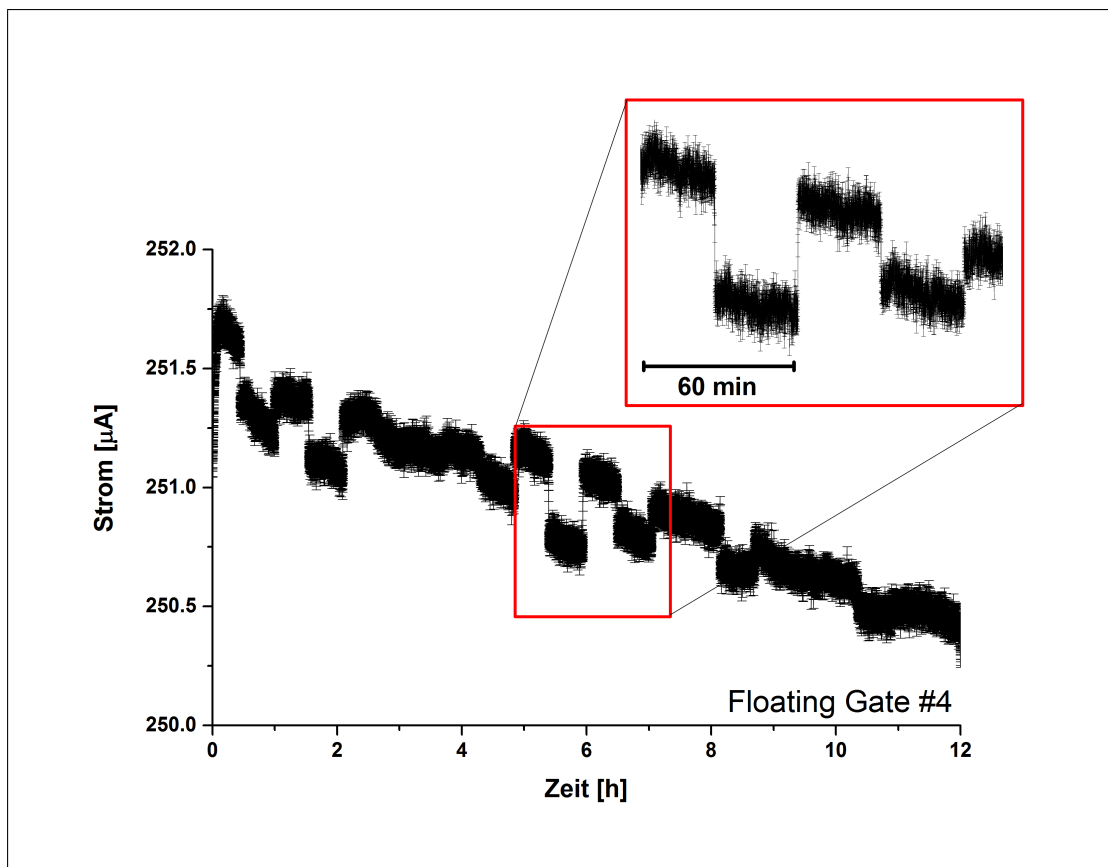


Abbildung 4.8: Auftretende Sprünge im Auslesestrom vergrößert dargestellt.

4.3 Stresstest

4.3.1 Messmethode

Das aus einem Ladepuls resultierende Gate-Potential liegt im Bereich von 8 V und mehr, wie durch Simulation in Kap. 3.1 ermittelt wird. Die Spezifikation der verwendeten Transistoren sieht allerdings eine maximale Spannung von 2.5 V vor die über dem Gate-Oxid anliegt. Schäden am Oxid sind daher in Betracht zu ziehen, falls mit diesen hohen Spannungen gearbeitet wird ([8]). Solche Schäden führen zu einem erhöhten Tunnelstrom durch das Gate und daher zu schneller Entladung des Gates. Sollten also kürzere Speicherzeiten beobachtet werden, so kann als Ursache eine Beschädigung des Gate-Oxids in Betracht gezogen werden.

Um dies zu untersuchen wird eine Zelle über die Zeitdauer von 24 h mit 10 V Programmiervspannung kontinuierlich geladen und entladen. Aufgrund den im ersten Absatz dieses Kapitels beschriebenen langsamen Lade- und Entladevorgänge werden Pulse der Dauer von 5 s verwendet. Die maximal auftretenden Ströme durch den Auslesetransistor liegen nach einem Ladevorgang bei ca. 1.5 mA, was einem Gate-Potential von rund 3.5 V entspricht (s. Abb. 3.2). Die nachfolgende Entladung bringt das Gate-Potential auf 0 V.

Im Anschluss wird erneut die Speicherzeit gemessen und mit vorher ermittelten Werten verglichen.

4.3.2 Resultate

Abb. 4.9 zeigt die Ergebnisse des Auslesestromes in Abhängigkeit der Zeit vor und nach dem Stresstest.

Für die untersuchte Floating-Gate Speicherzelle konnte in den zuvor durchgeführten Speicherzeitmessungen keine Speicherdauer ermittelt werden (FG #4, s. Speicherzeitmessung), da die Entladung des Gates derart klein war, dass nach 15 min kein signifikanter Abfall im Auslesestrom gemessen werden konnte. Die Speicherzeit ist nach dem Stresstest drastisch gefallen und beträgt nur ca. 60 s.

Die in der Abbildung im oberen Graphen gemachten Messungen wurden unmittelbar vor Beginn des 24 h-Programmierzklus durchgeführt, die Messungen unten im Bild direkt im Anschluss. Aufgrund von ähnlichen Temperaturbedingungen ist davon auszugehen, dass Schäden am Gate-Oxid zum schnelleren Entladen des Gates führen. Erneute Speicherzeitmessung zu einem späteren Zeitpunkt bestätigen diesen Verdacht.

Dieser Befund schließt die Verwendung der Speicherzellen als Parameterspeicher in neuromorpher Hardware bei den verwendeten Spannungen aus. Im Rahmen des großskaligen Einsatzes werden von den Floating-Gate Zellen vielen Millionen Lade- und

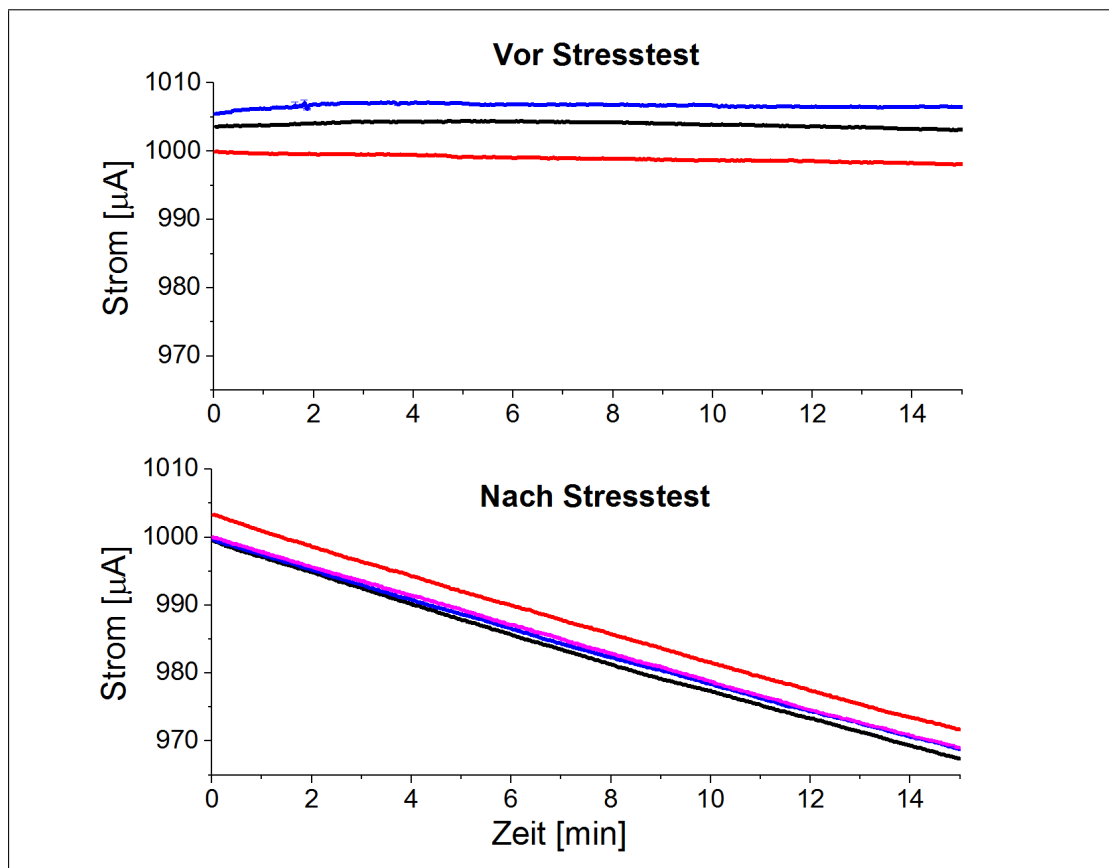


Abbildung 4.9: Stresstest: Gemessener Auslesestrom von FG #4 vor und nach 24-stündigem Laden und Entladen bei 10 V Pulsen der Dauer 5 s (hohes Gate-Potential).

Entladezyklen gefordert ([8]). Eine deutlicher Abfall in der Speicherdauer tritt hier jedoch bereits nach einigen Tausend Schreibvorgängen auf.

4.4 Lichtempfindlichkeit

4.4.1 Messmethode

Die Vorgängerversion des Floating-Gate Speichers zeigt eine Lichtabhängigkeit der Speicherdauer ([4], Kap. 9). Der photoelektrische Effekt wird hierbei als Ursache vermutet, wobei Elektronen durch Absorption von Photonen genügend Energie erhalten um die Potentialbarriere des Gate-Oxids zu überwinden und zu einer Entladung des Gates führen.

Hierzu wird die Speicherzeit bei direktem Lichteinfall auf den Chip und bei Dunkelheit

durch Benutzen eines lichtundurchlässigen Deckels untersucht. Es werden 15-minütige Messungen bei hohem Gate-Potential durchgeführt.

4.4.2 Resultate

Das Ergebnis der Messungen ist eine signifikante Änderung der Speicherzeit von $7.2 \pm 2.0\%$, wobei die Speicherzeit bei Lichteinfall geringer ist. Der bekannte Effekt kann also auch bei Floating-Gates in 65 nm-Technologie beobachtet werden.

Kapitel 5

Zusammenfassung und Ausblick

Es wurde analoger Floating-Gate Speicher in Single-Poly 65 nm-Technologie getestet. Die grundlegende Funktionsfähigkeit konnte gezeigt werden; wiederholtes Erzeugen von Lade- bzw. Entladepulsen erlaubt ein beliebig feines Einstellen des Gate-Potentials.

Wird eine Auflösung von 10 bit verwendet, so liegt die gemessene Speicherzeit bei Gate-Potentialen um 2.5 V in der Größenordnung von Minuten. Wird das maximale Potential halbiert, so kann bereits mit Speicherzeiten von Stunden gerechnet werden. Durch Anpassen des dynamischen Potentialbereiches lässt sich die Dauer der Speicherung also deutlich verbessern. Dieser Aspekt bietet daher durchaus das Potential einer Verwendung des Speichers in neuromorpher Hardware.

Es wurden große Variationen der Speicherzeiten bei verschiedenen Zellen ermittelt. Unklar ist, ob diese Unterschiede durch häufiges Laden und Entladen vor den eigentlichen Speicherzeitmessungen verursacht wurden. Ein Stresstest zeigt eine deutliche Änderung in der Speicherzeit bereits nach einigen Tausend Programmierzyklen. Es wird daher vermutet, dass Schäden am Gateoxid bei den verwendeten Spannungen auftreten.

Weitere Untersuchungen zur Funktionalität des Speichers sollten daher bei niedrigeren Pulsspannungen durchgeführt werden. Dem konnte in dieser Arbeit aufgrund der Rolle des großen Auslesetransistors nicht nachgegangen werden. Erste Änderungen im Auslestrom werden erst ab einer Pulshöhe von rund 8.5 V beobachtet und sämtliche Spannungen, die das Programmieren des Floating-Gates in sinnvollen Zeiten ermöglichen, liegen jenseits von 9 V.

Ebenso kann keine sinnvolle Angabe über die Schreibdauer der Floating-Gate Zellen gemacht werden. Die gemessenen Lade- und Entladezeiten sind aufgrund der falschen Dimensionierung des Auslesetransistors stark verzerrt und liegen weit abseits von Werten, die man in Anbetracht der bekannten Schreibdauern des Vorgängermodells in 180 nm-Technologie erwarten würde.

Als Fazit folgt aufgrund dieser Ausführungen, dass der Speicher in seiner jetzigen Form nicht unmittelbar einsetzbar ist. Die limitierenden Eigenschaften sind hierbei die geringe Speicherdauer von Minuten und der vermutete Verfall der Gateoxids innerhalb weniger Tausend Schreibvorgänge. Weitere Untersuchungen sollten sich daher darauf konzentrieren, ob die benötigte Pulsspannung verringert und ob der Einsatzbe-

reich des Gate-Potentials hin zu niedrigeren Werten verschoben werden kann. Sollte dies möglich sein, so kann der Einsatz des Speichers in analoger Hardware in Erwägung gezogen werden. Alternativ kann analysiert werden, inwiefern bei Floating-Gate Speicher mit stabilerem Gateoxid einer anderen Herstellungsoption die bestehenden Probleme auftreten.

Quellenverzeichnis

- [1] Jan-Peter Loock. Evaluierung eines floating gate analogspeichers für neuronale netze in single-poly umc 180nm cmos-prozess, 2006.
- [2] Sebastian Millner, Andreas Grübl, Karlheinz Meier, Johannes Schemmel, and Marc-Olivier Schwartz. A vlsi implementation of the adaptive exponential integrate-and-fire neuron model. *Advances in Neural Information Processing Systems*, 23:1642–1650, 2010.
- [3] Alexander Kononov. Testing of an analog neuromorphic network chip, 2011.
- [4] Sebastian Millner. Development of a multi-compartment neuron model emulation, 2012.
- [5] M. Hock, A. Hartel, J. Schemmel, and K. Meier. An analog dynamic memory array for neuromorphic hardware. In *Circuit Theory and Design (ECCTD), 2013 European Conference on*, pages 1–4, Sept 2013.
- [6] K. Ohsaki. A single poly eeprom cell structure for use in standard cmos processes. *IEEE Journal*, 29, 1994.
- [7] David L. Pulfrey. *Understanding Modern Transistors and Diodes*. Cambridge University Press, 2010.
- [8] M. Hock. Mündliche kommunikation.

Erklärung

Ich versichere hiermit, dass ich diese Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen verfasst habe.

Ort, Datum

Unterschrift