

TECHNISCHE UNIVERSITÄT DRESDEN

**Modellierung von Transistoren mit lokaler  
Ladungsspeicherung für den Entwurf von  
Flash-Speichern**

**Rico Srowik**

von der Fakultät Elektrotechnik und Informationstechnik der  
Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

DOKTORINGENIEURS  
(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender: Prof. Dr.-Ing. habil. U. Frühauf

Gutachter: Prof. Dr.-Ing. habil. R. Schüffny  
Prof. Dr.-Ing. habil. W. Reinhold  
Prof. Dr.-Ing. T. Mikolajick

Tag der Einreichung: 22. 03. 2007

Tag der Verteidigung: 28. 01. 2008



*Pour Sonia, Kilian et Alena.  
Je vous aime beaucoup.*



## Kurzfassung

In dieser Arbeit werden Speichertransistoren mit Oxid-Nitrid-Oxid-Speicherschicht und lokaler Ladungsspeicherung untersucht, die zur nichtflüchtigen Speicherung von Informationen genutzt werden. Charakteristisch für diese Transistoren ist, dass an beiden Enden des Transistorkanals innerhalb der Isolationsschicht Informationen in Form von Ladungspaketen unabhängig und getrennt voneinander gespeichert werden.

Für das Auslesen, Programmieren und Löschen der Speichertransistoren werden die physikalischen Hintergründe diskutiert und grundlegende Algorithmen zur Implementierung dieser Operationen auf einer typischen Speicherfeldarchitektur aufgezeigt.

Für Standard-MOS-Transistoren wird ein Kurzkanal-Schwellspannungsmodell abgeleitet und analytisch gelöst. Anhand dieser Modellgleichung werden die bekannten Kurzkanaleffekte betrachtet. Weiterhin wird ein Modell zur Berechnung des Drainstroms von Kurzkanaltransistoren im Subthreshold-Arbeitsbereich abgeleitet und gezeigt, dass sich die Drain-Source-Leckströme bei Kurzkanaltransistoren vergrößern.

Die Erweiterung des Schwellspannungsmodells für Standard-MOS-Transistoren auf den Fall der lokalen Ladungsspeicherung innerhalb der Isolationsschicht erlaubt die Ableitung eines Schwellspannungsmodells für Oxid-Nitrid-Oxid-Transistoren mit lokaler Ladungsspeicherung. Dieses Modell gestattet die qualitative und quantitative Diskussion der Erhöhung der Schwellspannung durch die lokale Injektion von Ladungsträgern beim Programmervorgang. Weiterhin ist es mit diesem Modell möglich, die Trennung der an beiden Kanalenden des Transistors gespeicherten Informationen beim Auslesevorgang qualitativ zu erklären und diese Bittrennung in Abhängigkeit von der Drainspannung zu berechnen. Für Langkanalspeichertransistoren wird eine analytische Näherungslösung des Schwellspannungsmodells angegeben, während das Kurzkanalverhalten durch die numerische Lösung der Modellgleichung bestimmt werden kann.

Für Langkanalspeichertransistoren wird ein Subthreshold-Modell zur Berechnung des Drainstroms abgeleitet. Dieses Modell zeigt, dass sich die Leckströme von programmierten Speichertransistoren im Vergleich zu Standard-MOS-Transistoren gleicher Schwellspannung vergrößern. Die Ursache dieses Effekts, die Verringerung der Subthreshold-Steigung von Transistoren im programmierten Zustand, wird analysiert.

Für einige praktische Beispiele wird die Anwendung der hergeleiteten Modellgleichungen beim Entwurf von Flash-Speichern demonstriert.



## **Abstract**

In this work, memory transistors with an oxide-nitride-oxide trapping-layer and local charge storage, which are used for non-volatile information storage, are examined. Characteristic for these transistors is an independent and separated storage of information by charge packages, located at both sides of the transistor channel, in the insulation layer.

The physical backgrounds for reading, programming and erasing the memory transistors are discussed, and basic algorithms are shown for implementing these operations on a typical memory array architecture.

For standard MOS-transistors a short channel threshold model is derived and solved analytically. By using these model equations, the known short channel effects are considered. Further, a model for calculating the drain current of short channel transistors in the subthreshold operation region is derived. This model is used to show the increase of drain-source leakage currents in short channel transistors.

By extending the standard MOS-transistor threshold voltage model for local charge storage in the insulation layers, the derivation of a threshold voltage model for oxide-nitride-oxide transistors with local charge storage is enabled. This model permits the quantitative and qualitative discussion of the increase in threshold voltage caused by local injection of charges during programming. Furthermore, with this model, the separation of the information, which are stored at both sides of the transistor channel, in the read-out operation is explained qualitatively, and the bit separation is calculated dependent on the drain voltage. For long channel memory transistors an analytical approximation of the threshold voltage model is given, whereas the short channel behaviour can be determined by solving the model equation numerically.

For long channel memory transistors, a subthreshold model for calculating the drain current is derived. This model shows the increase in leakage current of programmed memory transistors in comparison to standard MOS-transistors. The root cause of this effect, the reduced subthreshold swing of transistors in the programmed state, is analysed.

The application of the derived model equations for the development of flash memories is demonstrated with some practical examples.





# Inhaltsverzeichnis

|                                                                            |           |
|----------------------------------------------------------------------------|-----------|
| <b>Verzeichnis der verwendeten Symbole</b>                                 | <b>XI</b> |
| <b>1 Motivation</b>                                                        | <b>1</b>  |
| <b>2 Ladungsspeicherung in ONO-Schichten</b>                               | <b>3</b>  |
| 2.1 Nichtflüchtige Speicher . . . . .                                      | 3         |
| 2.2 ONO-Transistoren mit lokaler Ladungsspeicherung . . . . .              | 6         |
| 2.2.1 Auslesen der Speicherzelle . . . . .                                 | 7         |
| 2.2.2 Programmierung der Speicherzelle . . . . .                           | 8         |
| 2.2.3 Löschen der Speicherzelle . . . . .                                  | 11        |
| 2.3 Aufbau eines Speicherfelds . . . . .                                   | 13        |
| 2.4 Algorithmen . . . . .                                                  | 15        |
| 2.4.1 Lesealgorithmus . . . . .                                            | 16        |
| 2.4.2 Programmieralgorithmus . . . . .                                     | 18        |
| 2.4.3 Löschalgorithmus . . . . .                                           | 20        |
| 2.5 Multilevel-Speicher . . . . .                                          | 20        |
| <b>3 Schwellspannung von MOS-Transistoren</b>                              | <b>23</b> |
| 3.1 Grundlagen zur Modellierung der Schwellspannung . . . . .              | 23        |
| 3.1.1 Bändermodell . . . . .                                               | 23        |
| 3.1.2 Definition der Schwellspannung . . . . .                             | 25        |
| 3.2 Elektrostatisches Potential . . . . .                                  | 26        |
| 3.2.1 Herleitung der Differentialgleichung des Potentialverlaufs . . . . . | 26        |
| 3.2.2 Lösung der Differentialgleichung des Potentialverlaufs . . . . .     | 34        |
| 3.2.3 Diskussion des Potentialverlaufs . . . . .                           | 35        |
| 3.3 Schwellspannungsmodell . . . . .                                       | 39        |
| 3.3.1 Kurzkanaleffekt . . . . .                                            | 39        |
| 3.3.2 Drain Induced Barrier Lowering . . . . .                             | 41        |
| 3.4 Subthreshold-Verhalten . . . . .                                       | 42        |
| 3.4.1 Subthreshold-Transferkennlinie . . . . .                             | 43        |
| 3.4.2 Subthreshold-Steigung . . . . .                                      | 48        |
| <b>4 Schwellspannung von ONO-Speichertransistoren</b>                      | <b>50</b> |
| 4.1 Modell der Flachbandspannung . . . . .                                 | 50        |
| 4.2 Oberflächenpotential . . . . .                                         | 56        |

|          |                                                                      |            |
|----------|----------------------------------------------------------------------|------------|
| 4.2.1    | Differentialgleichung des Potentialverlaufs . . . . .                | 56         |
| 4.2.2    | Diskussion des Potentialverlaufs des Langkanaltransistors . . .      | 60         |
| 4.2.3    | Diskussion des Potentialverlaufs eines Kurzkanaltransistors . .      | 67         |
| 4.2.4    | Grenzen und Erweiterungsmöglichkeiten des Potentialmodells .         | 69         |
| 4.3      | Schwellspannungsmodell . . . . .                                     | 72         |
| 4.3.1    | Programmierung des Langkanaltransistors . . . . .                    | 72         |
| 4.3.2    | Nachbarbitunterdrückung bei Langkanaltransistoren . . . . .          | 75         |
| 4.3.3    | Programmierung eines Kurzkanaltransistors . . . . .                  | 82         |
| 4.3.4    | Kurzkanalverhalten . . . . .                                         | 84         |
| 4.3.5    | Nachbarbiteffekt und Bittrennung bei Kurzkanaltransistoren . .       | 85         |
| 4.4      | Subthreshold-Langkanalmodell . . . . .                               | 88         |
| 4.4.1    | Subthreshold-Transferkennlinie . . . . .                             | 88         |
| 4.4.2    | Subthreshold-Steigung . . . . .                                      | 90         |
| 4.4.3    | Grenzen und Erweiterungen des Subthreshold-Modells . . . . .         | 92         |
| <b>5</b> | <b>Anwendung beim Entwurf von Flash-Speichern</b>                    | <b>95</b>  |
| 5.1      | Dimensionierung der Drainspannung . . . . .                          | 95         |
| 5.2      | Verringerung des Lesefensters . . . . .                              | 97         |
| 5.3      | Parasitäre Subthreshold-Ströme im Lesevorgang . . . . .              | 99         |
| 5.4      | Subthreshold-Leckstrom von stark programmierten Transistoren . . . . | 101        |
| <b>6</b> | <b>Zusammenfassung</b>                                               | <b>104</b> |
|          | <b>Literaturverzeichnis</b>                                          | <b>106</b> |
| <b>A</b> | <b>Transistormodelle für die Bauelementesimulation</b>               | <b>114</b> |
| A.1      | Langkanal-MOS-Transistor . . . . .                                   | 114        |
| A.2      | Kurzkanal-MOS-Transistor . . . . .                                   | 119        |
| A.3      | Langkanal-ONO-Transistor . . . . .                                   | 124        |
| A.4      | Kurzkanal-ONO-Transistor . . . . .                                   | 129        |
| <b>B</b> | <b>MINIMOS-NT-Steuerdateien</b>                                      | <b>134</b> |
| <b>C</b> | <b>Numerische Lösung des Schwellspannungsmodells</b>                 | <b>136</b> |

# Verzeichnis der verwendeten Symbole

|               |                                                                                    |
|---------------|------------------------------------------------------------------------------------|
| $A^*$         | Effektive Richardson-Konstante                                                     |
| $C_{ox}$      | Kapazitätsbelag des Gateoxids einer MOS-Struktur; $C_{ox} = \epsilon_{ox}/t_{ox}$  |
| $\vec{E}$     | Elektrische Feldstärke                                                             |
| $E_{x,y}$     | x- bzw. y-Komponente der elektrischen Feldstärke                                   |
| $\vec{j}$     | Elektrische Stromdichte                                                            |
| $j_{x,y}$     | x- bzw. y-Komponente der elektrischen Stromdichte                                  |
| $k$           | Boltzmannkonstante; $k = 1,381 \cdot 10^{-23}$ As/K                                |
| $L$           | Effektive Transistorlänge                                                          |
| $l_c$         | Charakteristische Länge für Kurzkanaleffekte                                       |
| $n_0$         | Kenngroße $(d\psi_S/dV_g)^{-1}$                                                    |
| $n_i$         | Eigenleitungsichte von Silizium; $n_i = 1,5 \cdot 10^{10} \text{ cm}^{-3}$ (300 K) |
| $N_A$         | Akzeptordichte                                                                     |
| $N_A^-$       | Akzeptordichte, ionisiert                                                          |
| $N_D$         | Donatordichte                                                                      |
| $N_D^+$       | Donatordichte, ionisiert                                                           |
| $q$           | Betrag der Elementarladung eines Elektrons; $q = 1,602 \cdot 10^{-19}$ As          |
| $Q$           | Elektrische Ladung                                                                 |
| $Q'$          | Elektrische Linienladungsdichte                                                    |
| $Q_{inj}''^*$ | Ladungsdichte der in das Nitrid injizierten Ladungen                               |
| $Q_{ox}''$    | Ladungsdichte im Gateoxid einer MOS-Struktur                                       |
| $Q_{ss}''$    | Zwischenschichtladungsdichte der Oxid-Halbleiter-Grenzfläche                       |
| $SS$          | Subthreshold-Steigung                                                              |
| $T$           | Temperatur                                                                         |
| $t_{ono}^*$   | Effektive Dicke des ONO-Stapels                                                    |
| $t_{ox}$      | Oxiddicke von Standard-MOS-Transistoren                                            |
| $t_{dep}$     | Tiefe der Raumladungszone in MOS-Transistoren;<br>Verarmungslänge                  |
| $t_{dep,L}$   | Tiefe der Raumladungszone in Langkanal-MOS-Transistoren                            |
| $t_j$         | Tiefe der Source-/ Drain-Gebiete                                                   |
| $U_T$         | Temperaturspannung; $U_T = kT/q$                                                   |
| $V$           | Externes Potential                                                                 |
| $V_{ab}$      | Spannung zwischen den Kontakten a und b                                            |
| $V_{bi}$      | Diffusionsspannung                                                                 |
| $V_{fb}$      | Flachbandspannung                                                                  |

## Verzeichnis der verwendeten Symbole

---

|                          |                                                                                                                                                                                                      |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\Delta V_{fb,s(d)}$     | Maximalwert der source- (drain-) seitigen Verschiebung der Flachbandspannung durch die Programmierung des ONO-Speichertransistors; Programmierniveau; Injektionsniveau                               |
| $\Delta V_{fb,s(d),min}$ | Minimalwert von $\Delta V_{fb,s(d)}$ , der notwendig ist, um ein lokales Minimum im Potentialverlauf auszubilden                                                                                     |
| $\Delta V_{fb,s(d),0}$   | Minimalwert von $\Delta V_{fb,s(d)}$ , ab dem eine Verschiebung der Schwellspannung eines ONO-Speichertransistors stattfindet; Definiert für die quadratische Näherung der Schwellspannungsgleichung |
| $V_{fb,0}$               | Flachbandspannung des initialen ONO-Speichertransistors                                                                                                                                              |
| $\Delta V_M$             | Unterschied der Langkanalschwellspannung eines unprogrammierten Transistors zur maximalen Rückwärts-Schwellspannung ohne Nachbarbiteffekt; Bittrennung                                               |
| $V_{ox}$                 | Spannung über dem Gateisolator                                                                                                                                                                       |
| $V_{th}$                 | Schwellspannung                                                                                                                                                                                      |
| $V_{th,fwd}$             | Schwellspannung in der Vorwärts-Lese-Betriebsart                                                                                                                                                     |
| $V_{th,L}$               | Langkanalschwellspannung                                                                                                                                                                             |
| $V'_{th,L}$              | Extrapolierte Langkanalschwellspannung                                                                                                                                                               |
| $V'_{th0,L}$             | Extrapolierte Schwellspannung eines langen ONO-Speichertransistors im initialen Zustand                                                                                                              |
| $V_{th,rev}$             | Schwellspannung in der Rückwärts-Lese-Betriebsart                                                                                                                                                    |
| $V_{th,s(d)}$            | Schwellspannung eines ONO-Speichertransistors, die sich durch die Verwendung des source- (drain-) seitigen lokalen Minimums des Potentialverlaufs ergibt                                             |
| $V_V$                    | Breite der Schwellspannungsverteilung                                                                                                                                                                |
| $V_\Delta$               | Abstand der Schwellspannungsverteilung der programmierten und der gelöschten Bits eines Speicherfelds; Lesefenster                                                                                   |
| $W_c$                    | Energie des Leitungsbands                                                                                                                                                                            |
| $W_{EA}$                 | Austrittsarbeit                                                                                                                                                                                      |
| $W_F$                    | Fermienergie                                                                                                                                                                                         |
| $W_g$                    | Bandabstand                                                                                                                                                                                          |
| $W_i$                    | Energie des Eigenleitungsniveaus                                                                                                                                                                     |
| $W_m$                    | Aktivierungsenergie des Metallgates                                                                                                                                                                  |
| $W_s$                    | Aktivierungsenergie des Siliziumsubstrats                                                                                                                                                            |
| $W_v$                    | Energie des Valenzbands                                                                                                                                                                              |
| $x_{min}^\Delta$         | Lage des Minimalwerts der Potentialverschiebung $\Delta\psi_S^i$                                                                                                                                     |
| $\gamma$                 | Substratsteuerfaktor; Bodyeffekt-Koeffizient                                                                                                                                                         |
| $\delta$                 | Effektive Kanalweite                                                                                                                                                                                 |
| $\epsilon_{nit}$         | Absolute Dielektrizitätskonstante des Nitrids                                                                                                                                                        |
| $\epsilon_{ox}$          | Absolute Dielektrizitätskonstante des Gateoxids                                                                                                                                                      |
| $\epsilon_{si}$          | Absolute Dielektrizitätskonstante von Silizium                                                                                                                                                       |

---

|                    |                                                                                                                                                     |
|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| $\eta$             | Fittingparameter                                                                                                                                    |
| $\lambda_{s,d}$    | Ausdehnung der lokalen Änderungen der Flachbandspannung                                                                                             |
| $\Lambda$          | Mittlere freie Weglänge eines Ladungsträgers                                                                                                        |
| $\mu$              | Beweglichkeit                                                                                                                                       |
| $\bar{\mu}$        | mittlere Beweglichkeit                                                                                                                              |
| $\rho$             | Raumladungsdichte                                                                                                                                   |
| $\rho_{inj}$       | Raumladungsdichte der in das Nitrid injizierten Ladungen                                                                                            |
| $\rho_{inj}^*$     | Effektive Raumladungsdichte der in das Nitrid injizierten Ladungen unter Berücksichtigung der lateralen Komponente des elektrischen Felds im Nitrid |
| $\varphi$          | Makropotential                                                                                                                                      |
| $\phi$             | Elektrostatistisches Potential im Halbleiter, bezogen auf das Fermi-niveau im Halbleiterinneren                                                     |
| $\Phi_{MS}$        | Metall-Halbleiter-Kontaktspannung                                                                                                                   |
| $\psi$             | Elektrostatistisches Potential im Halbleiter, bezogen auf Eigenleitungsniveau im Halbleiterinneren                                                  |
| $\psi_0$           | Oberflächenpotential bei Beginn der starken Inversion; $\psi_0 = V_s + V_{bi}$                                                                      |
| $\psi_S$           | Oberflächenpotential des Halbleiters                                                                                                                |
| $\psi_S^i$         | Initiales Oberflächenpotential eines ONO-Transistors                                                                                                |
| $\Delta\psi_S^i$   | Verschiebung des initialen Oberflächenpotentials eines ONO-Transistors durch die injizierten Ladungen                                               |
| $\Delta\psi_S$     | Unterschied zum Oberflächenpotential eines langen Transistors                                                                                       |
| $\Delta\psi_{S,L}$ | Unterschied zum Oberflächenpotential eines langen, unprogrammierten ONO-Transistors                                                                 |
| $\Psi_{FB}$        | Fermipotential des Substrates                                                                                                                       |



# 1 Motivation

Als nichtflüchtige Speicher (*non-volatile memories*) werden Halbleiterspeicher bezeichnet, in denen die gespeicherte Information auch nach dem Abschalten der Stromzufuhr erhalten bleibt. Für diese Speicherklasse gibt es einen weiten Anwendungsbereich. Gerade in jüngster Vergangenheit ist die Nachfrage nach nichtflüchtigen Speichern durch neuartige Anwendungen, wie USB-Speicherstifte, SD- oder Compact-Flash-Speicherkarten oder tragbare Musikabspielgeräte, innerhalb kürzester Zeit stark angestiegen. Aufgrund dieser starken Nachfrage bauen die Halbleiterhersteller zurzeit ihre Produktionskapazitäten aus. Mit dem Ausbau der Produktionskapazitäten und der Massenproduktion der Speicherchips geht aber auch eine Verschärfung des Wettbewerbs unter den Herstellern einher. Der Wettbewerbsdruck zwingt die Halbleiterhersteller dazu, Speichertechnologien einzuführen, die die Entwicklung von Produkten mit dem geringstmöglichen Preis pro gespeicherter Information erlauben.

Die Suche nach Speicherprinzipien mit hoher Speicherdichte führte zur Wiederentdeckung der Transistoren mit Oxid-Nitrid-Oxid-Speicherschicht (sogenannter *SONOS-Speicher*) in Verbindung mit der Entwicklung des Prinzips der lokalen Ladungsspeicherung, das darauf beruht, dass in jedem Transistor an beiden Kanälen Informationen unabhängig voneinander gespeichert werden können. Dieses Prinzip wurde von der Firma Saifun unter der Bezeichnung NROM<sup>TM</sup>-Technologie entwickelt und an verschiedene Halbleiterhersteller lizenziert. Auf dem NROM<sup>TM</sup>-Prinzip basierende Produkte werden zum Beispiel unter den Bezeichnungen Twin Flash<sup>TM</sup> (Qimonda), NBit<sup>TM</sup> (Macronix), Mirrorbit<sup>TM</sup> (Spansion) oder micro flash (Tower Semiconductor) angeboten. Weitere Firmen, die diese Speichertechnologie lizenziert haben, sind unter anderem Sony, NEC und SMIC. Die NROM<sup>TM</sup>-Technologie ist also mit Massenprodukten auf dem Markt der nichtflüchtigen Speicher vertreten.

Obwohl die Technologie der lokalen Ladungsspeicherung bereits in kommerziellen Produkten angewendet wird, sind die physikalischen Hintergründe der Ladungsspeicherung, der Zuverlässigkeit und Alterung, des Datenhaltevermögens, der Wiederbeschreibbarkeit und des Strom-Spannungs-Verhaltens der Speichertransistoren noch unzureichend erforscht. Modelle, die die zugehörigen physikalischen Prozesse beschreiben, sind derzeit Gegenstand intensiver Forschungen und kaum für eine Anwendung in der Produktentwicklung verfügbar.

Typischerweise beginnt der System- und Schaltungsentwurf von Demonstratorsystemen, die die Ausbeutefähigkeit einer neuen Technologie nachweisen sollen, und von nachfolgenden kommerziellen Einführungsprodukten jedoch schon, wenn sich die zugehörige Technologiegeneration noch in einer frühen Phase ihrer Entwicklung befindet.

Trotzdem müssen bereits in dieser frühen Entwurfsphase viele wichtige Entscheidungen getroffen werden, wie zum Beispiel die Auswahl der Speicherfeldarchitektur oder die Auswahl der Schaltungskonzepte. Zusätzlich werden den System- und Schaltungsentwicklern bereits in der Konzeptphase verlässliche Abschätzungen zu Produktkenngrößen wie Schreib-, Lese- und Löschgeschwindigkeiten und den zugehörigen Verlustleistungen abverlangt. Fehlentscheidungen in der Konzeptauswahl oder Auswirkungen von ungenauen Abschätzungen der Produktparameter sind in späteren Entwurfsphasen kaum noch korrigierbar.

Die System- und Schaltungsentwickler benötigen deshalb Modelle, mit denen die elektrischen Eigenschaften der Speichertransistoren schon mit Hilfe weniger bekannter Technologieparameter abgeschätzt werden können. Diese Technologieparameter können dabei zum Teil durch Extrapolation aus vorherigen Technologiegenerationen gewonnen werden. Die Verwendung von physikalischen Modellen erlaubt es dem Designer, wichtige Systemkenngrößen, wie die Arbeitspunkte der Speichertransistoren oder eine eventuelle gegenseitige Beeinflussung der beiden in einem Transistor gespeicherten Informationen, zu einem sehr frühen Zeitpunkt des Entwurfsprozesses zu bestimmen und damit gegebenenfalls Einfluss auf die Entwicklung des Speicherbauelements zu nehmen.

Die wichtigste Kenngröße von nichtflüchtigen Speichertransistoren ist die Transistorschwellspannung, da die Höhe dieser Schwellspannung die gespeicherte Information repräsentiert. Außerdem ist ein Schwellspannungsmodell Voraussetzung für die Beschreibung der Strom-Spannungs-Kennlinie der Speichertransistoren. Aus diesem Grund wird in dieser Arbeit ein physikalisches Schwellspannungsmodell für Oxid-Nitrid-Oxid-Speichertransistoren mit lokaler Ladungsspeicherung entwickelt. Besondere Beachtung soll dabei der Anwendung dieses Modells für den Systementwurf von Flash-Speichern geschenkt werden.

In der folgenden Arbeit wird zunächst in das Gebiet der nichtflüchtigen Speicher eingeführt. Dabei wird besonders auf Speichertransistoren mit Oxid-Nitrid-Oxid-Speicherschicht und lokaler Ladungsspeicherung eingegangen. Anhand dieses Speicherprinzips werden die physikalischen Mechanismen, die das Lesen, Programmieren und Löschen der Speicherzellen erlauben, erläutert und der Aufbau von typischen Speicherfeldern beschrieben. Im nächsten Kapitel wird ein Modell zur Berechnung der Schwellspannung von Standard-MOS-Transistoren abgeleitet und ein Modell der Strom-Spannungs-Kennlinie im Subthreshold-Arbeitsbereich der Transistoren erstellt. Anschließend wird das Schwellspannungsmodell auf Speichertransistoren mit lokaler Ladungsspeicherung erweitert und es werden Schlussfolgerungen zum Programmierverhalten und zu eventuellen Beeinflussungen der Schwellspannung eines Bits durch die gespeicherte Ladung des Nachbarbits gezogen. Weiterhin wird die Strom-Spannungs-Kennlinie im Subthreshold-Bereich für programmierte Langkanalspeichertransistoren abgeleitet und diskutiert. Anhand einiger praktischer Beispiele wird die Anwendung der Modelle beim System- und Schaltungsentwurf gezeigt.



## 2 Ladungsspeicherung in ONO-Schichten

In diesem Kapitel erfolgt zunächst eine Einteilung der nichtflüchtigen Speicher nach dem Aufbau der Speichertransistoren. Für die Unterart der SONOS-Speicher mit lokaler Ladungsspeicherung werden anschließend die Programmierung, das Auslesen und das Löschen in ihren physikalischen Mechanismen und in der technischen Realisierung diskutiert. Im Anschluss wird die Zusammenschaltung einzelner Speicherzellen zu großen Speicherfeldern und die Ausführung von Programmier-, Lese- und Löschalgorithmen gezeigt. Den Abschluss bildet ein Ausblick auf die Multilevel-Speicherung.

### 2.1 Nichtflüchtige Speicher

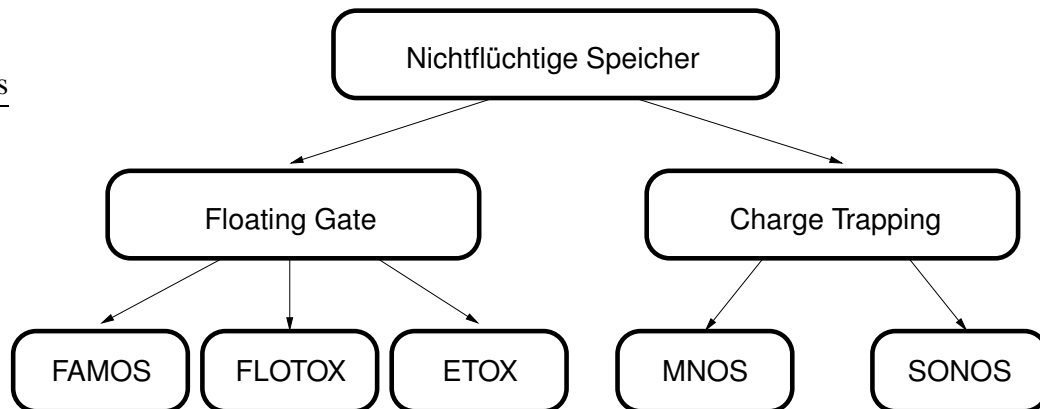
Im Laufe der Entwicklung der Halbleiterspeicher wurden verschiedene physikalische Prinzipien zur nichtflüchtigen Speicherung von Informationen implementiert. Alle diese Verfahren beruhen darauf, dass in einer MOS-Transistorstruktur elektrische Ladungen in eine Speicherschicht, die sich zwischen dem Gate und dem Substrat befindet, eingebracht werden und dass das Abfließen dieser Ladungen mit Hilfe isolierender Schichten verhindert wird. Wie in Abschnitt 3.1 gezeigt werden wird, verändert die so gespeicherte Ladung die Schwellspannung des Transistors. Da die Schwellspannung den Drainstrom beeinflusst, kann man durch Messung des Drainstroms auf die Schwellspannung rückschließen und damit die gespeicherte Information auslesen.

Im Allgemeinen werden bei der sogenannten Programmierung des Transistors Elektronen in die Speicherschicht eingebracht, wodurch die Schwellspannung ansteigt. Es ist üblich, dem Zustand der erhöhten Schwellspannung den logischen Pegel „0“ zuzuordnen. Werden die gespeicherten Ladungen entfernt und der niedrige Schwellspannungswert eingestellt, spricht man vom Löschen des Transistors. Die niedrige Schwellspannung hat im Allgemeinen den logischen Wert „1“.

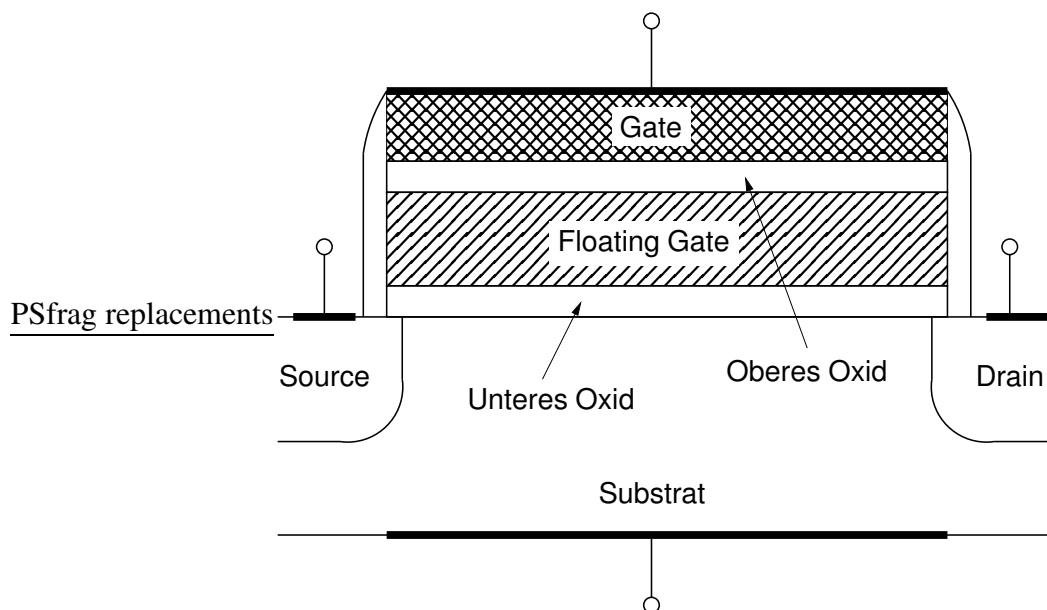
Entsprechend ihrem Aufbau werden nichtflüchtige Speicher in zwei Hauptklassen unterteilt [BB98]. Die Floating-Gate-Transistoren besitzen, wie in Abb. 2.2 gezeigt wird, ein zusätzliches, isoliertes Gate. Dieses „schwebende“ (floating) Gate besteht für gewöhnlich aus leitfähigem Polysilizium. Das Floating Gate ist von isolierenden Schichten aus Siliziumoxid umgeben. Im Allgemeinen besitzen Floating-Gate-Transistoren ein zusätzliches Steuergate (Control Gate).

Die ersten funktionsfähigen Floating-Gate-Speichertransistoren nutzten für die Programmierung die Injektion von hochenergetischen Elektronen, die durch einen Lawinendurchbruch erzeugt wurden (*Floating gate Avalanche MOS transistor* - FAMOS) [FB74]. Die derzeit am weitesten verbreiteten *FLO*ating *gate* *Thin OXid*- (FLOTOX) Transisto-

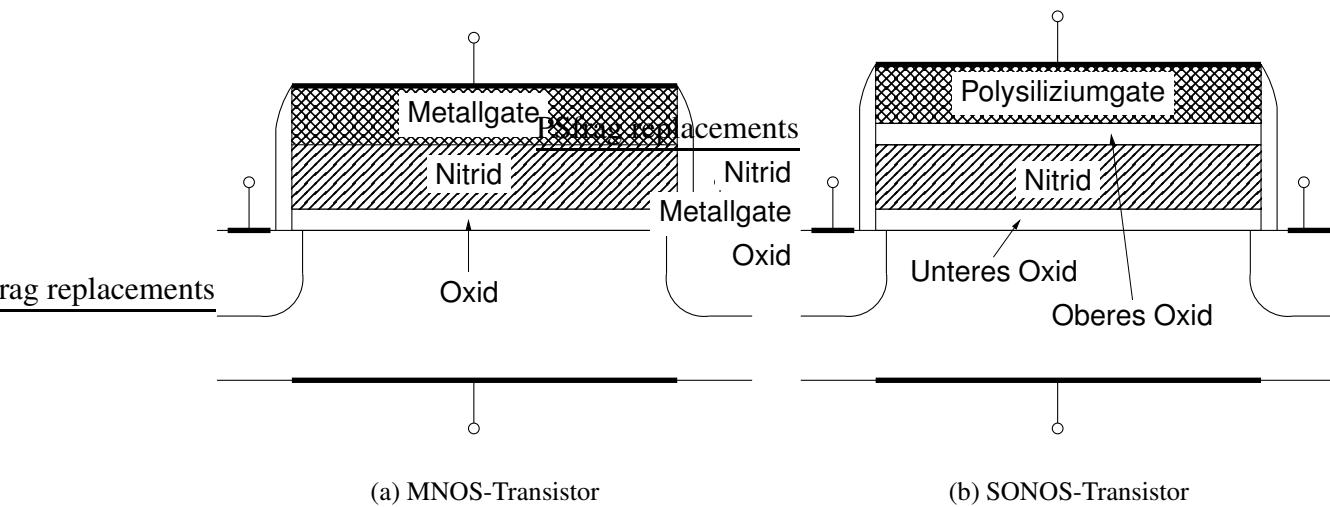
PSfrag replacements



**Abbildung 2.1:** Einteilung der nichtflüchtigen Speicher nach deren Aufbau



**Abbildung 2.2:** Allgemeiner Aufbau eines Floating-Gate-Transistors

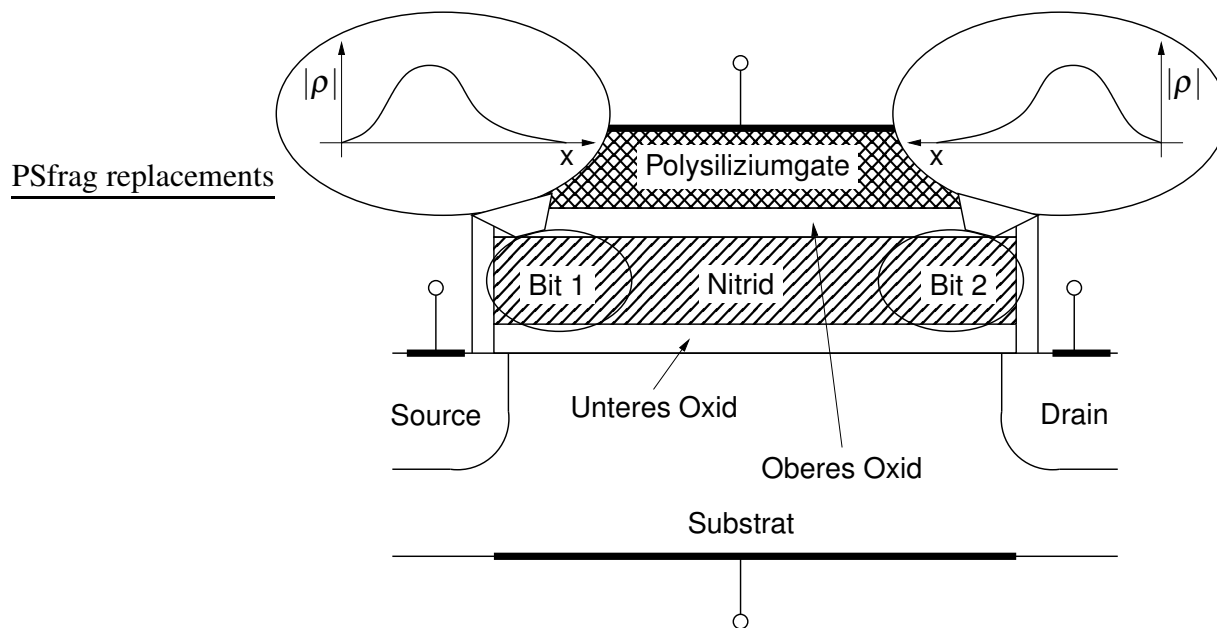


**Abbildung 2.3:** Gegenüberstellung des Aufbaus von MNOS- und SONOS-Speichertransistoren

ren und *EPROM Tunnel Oxid*- (ETOX<sup>TM</sup>) Transistoren werden jedoch gewöhnlicherweise durch Fowler-Nordheim- (FN-) Tunnelung des unteren, dünnen Oxids programmiert und gelöscht. In einigen Floating-Gate-Speichern werden auch hochenergetische Kanalelektronen für den Programmiervorgang verwendet. Dieser physikalische Mechanismus wird in Abschnitt 2.2.2 diskutiert.

Floating-Gate-Speicher sind die gebräuchlichsten nichtflüchtigen Speicher und dominieren gegenwärtig den Markt. Floating-Gate-Transistoren sind technisch so weit entwickelt, dass sogar mehrere Schwellspannungsniveaus programmiert und über eine lange Datenhaltezeit stabil ausgelesen werden können. Diese Multibit-Speicherung erlaubt eine weitere Erhöhung der effektiven Speicherdichte.

Die zweite Hauptklasse der nichtflüchtigen Speicher sind die Charge-Trapping-Speicher. Bei dieser Speicherart wird die injizierte Ladung in einer elektrisch nichtleitenden Schicht, meist in Siliziumnitrid ( $\text{Si}_3\text{Ni}_4$ ), gespeichert. In dieser Speicherschicht oder an deren Nitrid-Oxid-Grenzschicht befinden sich sogenannte Haftstellen (*Traps*), die mit Ladungsträgern besetzt werden können und die diese Ladungsträger dann, unter entsprechenden Bedingungen, wieder emittieren. In der ursprünglichen Form waren diese Transistoren als Metall-Nitrid-Oxid-Silizium (MNOS-) Stapel aufgebaut [FB70]. Da jedoch festgestellt wurde, dass diese Transistoren die eingefangenen Ladungen über den Metall-Nitrid-Übergang nach und nach verlieren können, wurden zur Erhöhung der Datenhaltezeit eine weitere Oxidschicht zwischen dem Nitrid und dem Gateanschluss eingefügt und diese Transistoren mit einem Polysilizium-Gate ausgestattet. Es ergibt sich eine Silizium-Oxid-Nitrid-Oxid-Silizium- (SONOS-) Schichtenfolge. Diese beiden Arten der Charge-Trapping-Speicher sind in Abb. 2.3 einander gegenübergestellt.



**Abbildung 2.4:** Aufbau eines SONOS-Transistors mit lokaler Ladungsspeicherung

In den MNOS-Speichern wurden bei der Programmierung die Elektronen durch Fowler-Nordheim- (FN-) -Tunnelung aus dem Substrat in das Nitrid eingebracht. Für das Löschen wurde das FN-Tunneln von Löchern aus dem Substrat in das Nitrid bzw. der Elektronen aus dem Nitrid in das Gate verwendet [LS72]. SONOS-Speicher wenden ebenfalls häufig FN-Tunnelung für die Programmierung an, teilweise werden aber auch hochenergetische Kanalelektronen verwendet [CCS<sup>+</sup>98]. Für das Löschen sind ebenfalls verschiedene Mechanismen verbreitet, wie zum Beispiel das FN-Tunneln von Löchern bzw. Elektronen oder die Injektion von hochenergetischen Löchern aus dem Substrat.

## 2.2 ONO-Transistoren mit lokaler Ladungsspeicherung

Auf der Suche nach Speicherprinzipien, die eine hohe Informationsdichte und damit einen niedrigen Preis pro gespeichertem Bit erlauben, entwickelte Boaz Eitan einen speziellen SONOS-Speichertransistor, der sich dadurch auszeichnet, dass in jedem Transistor zwei Bits gespeichert werden können [EPB<sup>+</sup>99] [EPB<sup>+</sup>00].

Wie in Abb. 2.4 gezeigt wird, besitzt diese Transistorart den üblichen SONOS-Schichtenaufbau. Im Unterschied zu den gewöhnlichen SONOS-Speichertransistoren werden die injizierten Ladungen aber nicht homogen in lateraler Richtung im Nitrid gespeichert. Vielmehr befinden sich in diesen Speichertransistoren die Ladungspakete, die die beiden gespeicherten Bits repräsentieren, räumlich voneinander getrennt über dem source- bzw. drainseitigen Kanalende in der Nitrid-Speicherschicht.

Dieses neuartige Speicherprinzip gewinnt derzeit stark an Beachtung und wird auch schon für kommerzielle Massenprodukte eingesetzt [MDS<sup>+</sup>02] [SEB<sup>+</sup>04] [KIM<sup>+</sup>05] [LAC<sup>+</sup>04] [TZW<sup>+</sup>06] [FKR<sup>+</sup>06] [CPL<sup>+</sup>06]. Im Folgenden soll sich auf diese SONOS-Speicher mit lokaler Ladungsspeicherung konzentriert werden.

### 2.2.1 Auslesen der Speicherzelle

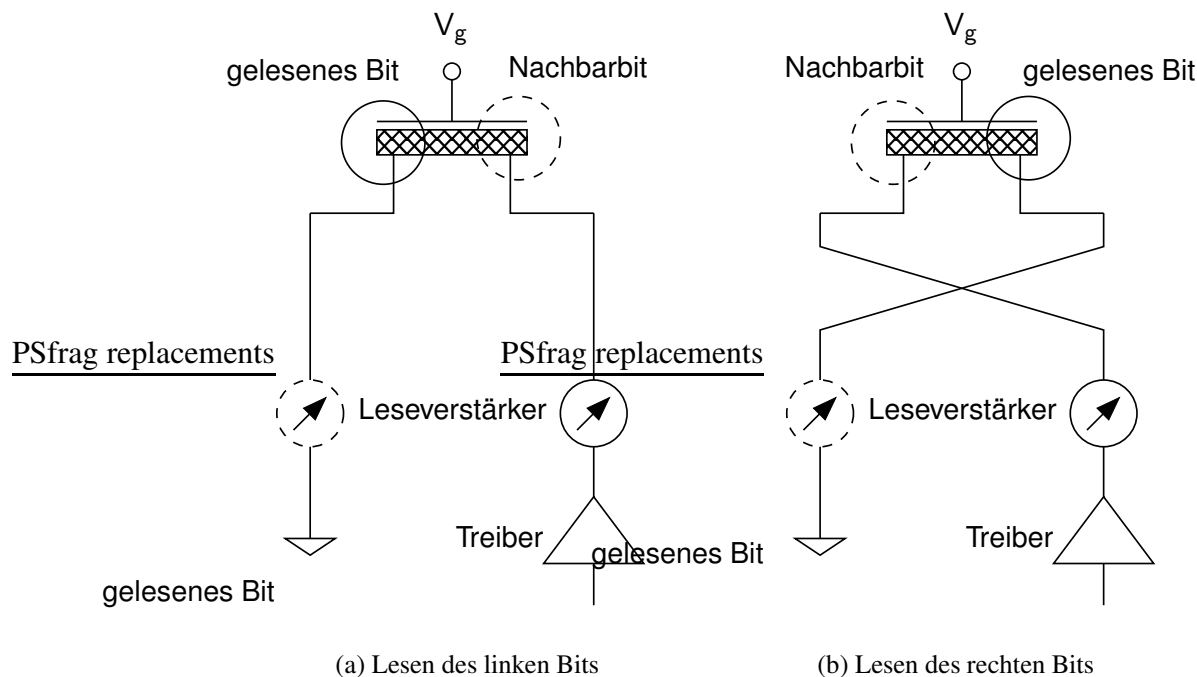
Wie in den letzten Abschnitten erwähnt wurde, wird durch das Programmieren eine hohe Schwellspannung und durch das Löschen eine niedrige Schwellspannung des Speichertransistors eingestellt. Beim Lesen soll der Schwellspannungszustand festgestellt werden und eine Zuordnung zu einem logischen Wert stattfinden.

Bei der Leseoperation wird ausgenutzt, dass der Drain-Source-Strom eines MOS-Transistors unter anderem von der Schwellspannung abhängt. An den Speichertransistor wird eine Drainspannung von 1...2 V und eine Gatespannung von  $\sim 4$  V angelegt [LAC<sup>+</sup>04] [SEB<sup>+</sup>04] [KIM<sup>+</sup>05]. Das Source wird auf einem Potential von  $< 0,2$  V gehalten.

Der Drain-Source-Strom wird gewöhnlich gemessen, indem er in einen äquivalenten Spannungswert umgesetzt wird. Diese Umsetzung kann durch einen ohmschen Widerstand (*DC-Sensing*) oder durch die Integration des Stroms auf einer Kapazität (*AC-Sensing*) erfolgen. Weiterhin gibt es die beiden schaltungstechnischen Möglichkeiten, den Strom im Drain-Zweig (*Drain-Side-Sensing*) oder im Source-Zweig (*Source-Side-Sensing*) zu messen. Der gemessene Strom wird dann mit einem Referenzstrom verglichen. Ist der Messstrom größer als der Referenzwert und die Schwellspannung des Speichertransistors also gering, wird ein gelöschter Zustand detektiert und dem ausgelesenen Bit folglich der logische Wert „1“ zugeordnet. Ist der Messstrom im umgekehrten Fall kleiner als der Referenzstrom, wird der logische Wert „0“ ausgegeben. Die Strommessung und Detektion des logischen Werts findet im Leseverstärker (*Sense Amplifier*) statt.

Bei ONO-Transistoren mit lokaler Ladungsspeicherung sind Informationen an beiden Kanälen gespeichert. Beim Lesen wird die Information desjenigen Bits ausgewertet, das sich an der Sourceseite befindet, also auf dem niedrigeren Potential liegt. Das andere, sogenannte *Nachbarbit*, liegt auf der Drainseite auf einem hohen Potential und seine Auswirkung auf den Messstrom wird somit im Idealfall unterdrückt. Sollen beide Bits nacheinander gelesen werden, müssen also die beiden Kanalkontakte nacheinander die Rolle des Drain- und Sourceanschlusses übernehmen (Abb. 2.5).

Beim Auslesen des interessierenden Bits kann das Nachbarbit das Messergebnis störend beeinflussen. Aus diesem Grund wird in Abschnitt 4.3.2 der physikalische Hintergrund der Nachbarbitunterdrückung untersucht und ein Modell zur Quantifizierung der Nachbarbitunterdrückung hergeleitet.



**Abbildung 2.5:** Auslesen einer ONO-Speicherzelle

### 2.2.2 Programmierung der Speicherzelle

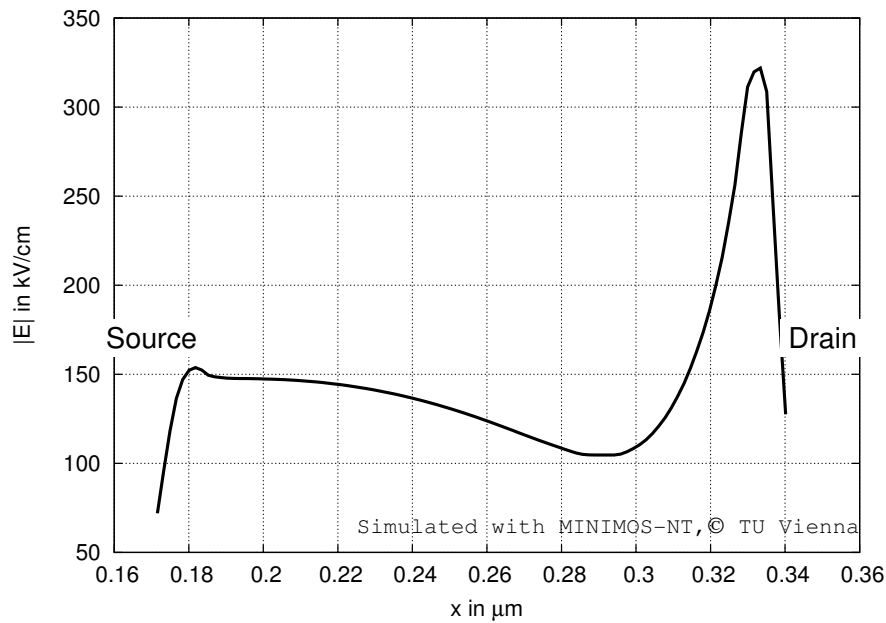
Wie auch einige andere nichtflüchtige Speicher werden die ONO-Speicher mit lokaler Ladungsspeicherung durch hochenergetische („heiße“) Kanalelektronen programmiert (*Channel Hot Electrons* – CHE).

Zur Erzeugung dieser heißen Elektronen wird an den Transistor eine Drain-Source-Spannung von zum Beispiel  $\sim 5\text{ V}$  angelegt. Die Gate-Source-Spannung liegt etwa bei  $\sim 9\text{ V}$  [KIM<sup>+</sup>05]. Da der Transistor folglich im Bereich der starken Inversion betrieben wird, bildet sich an der Substratoberfläche ein leitfähiger Elektronenkanal aus.

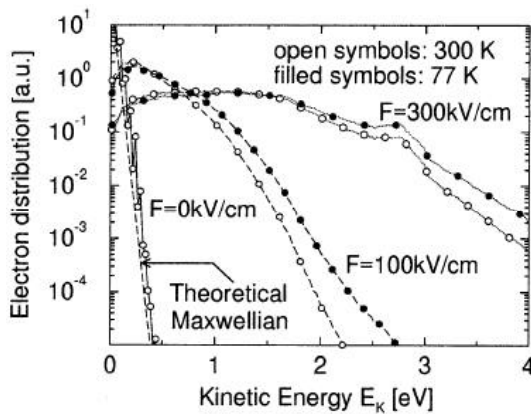
In Abb. 2.6 ist ein typischer Feldstärkeverlauf entlang des Elektronenkanals eines MOS-Transistors gezeigt. Charakteristisch ist der starke Anstieg der elektrischen Feldstärke in einem schmalen Bereich am drainseitigen Kanalende. Die Elektronen werden entsprechend auf ihrem Weg vom Source-Anschluss zum Drain in diesem Bereich stark in Drainrichtung beschleunigt und nehmen dabei Energie auf. Durch stochastische Interaktionen mit Phononen, Stoßionisation oder Interaktionen der Ladungsträger untereinander kommt es dabei zu zufälligen Energie- und Richtungsänderungen der Ladungsträger.

In Abb. 2.7 ist eine Simulation der sich einstellenden Energie- und Richtungsverteilung aus [SF99] gezeigt. Es ist erkennbar, dass bei hohen Feldstärken durch die Beschleunigungs- und Stoßprozesse die Verteilung der Energien der Population der Kanalelektronen im Vergleich zur Maxwell-Verteilung des thermodynamischen Gleichge-

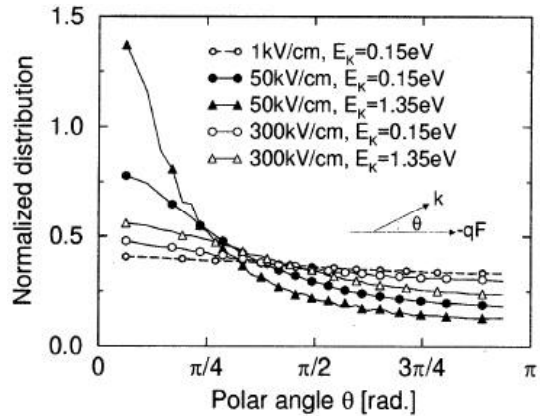
PSfrag replacements



**Abbildung 2.6:** Typischer Verlauf der Feldstärke im Kanal eines SONOS-Transistors bei der Programmierung mit heißen Kanalelektronen



(a) Energieverteilung

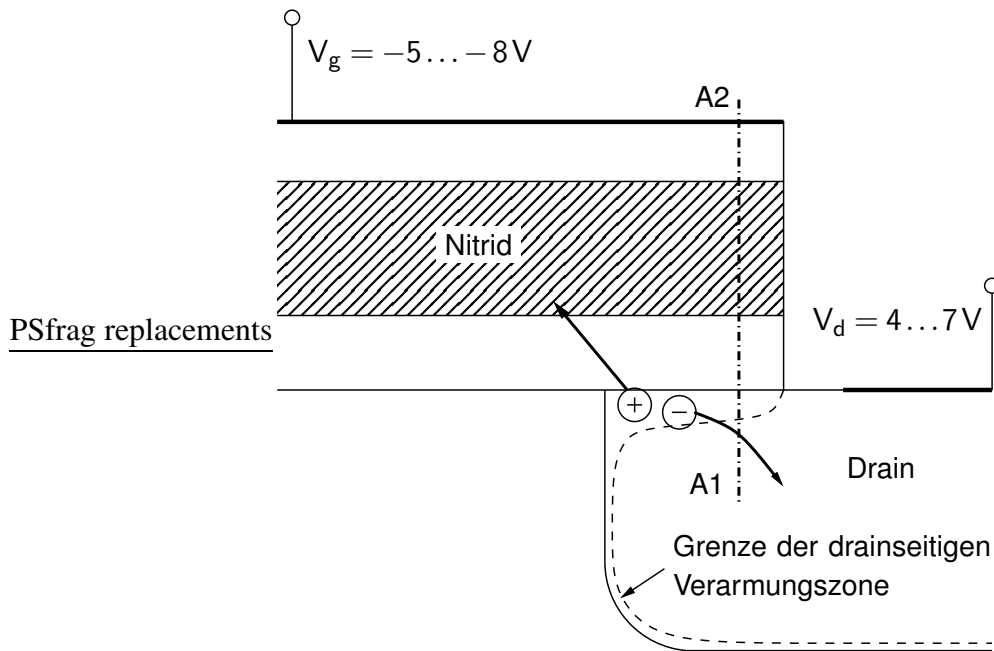


(b) Verteilung der Richtung der Ladungsträgerbewegung

**Abbildung 2.7:** Verteilungen der Energien und Richtungsvektoren der Kanalelektronen bei verschiedenen Feldstärken aus [SF99] (S. 183, Abb. 4.17); Mit freundlicher Genehmigung von Springer Science and Business Media ©1999 Springer/Kluwer Academic Publishers







**Abbildung 2.9:** Ausbildung einer Verarmungszone im Draingebiet eines ONO-Transistors beim Löschvorgang

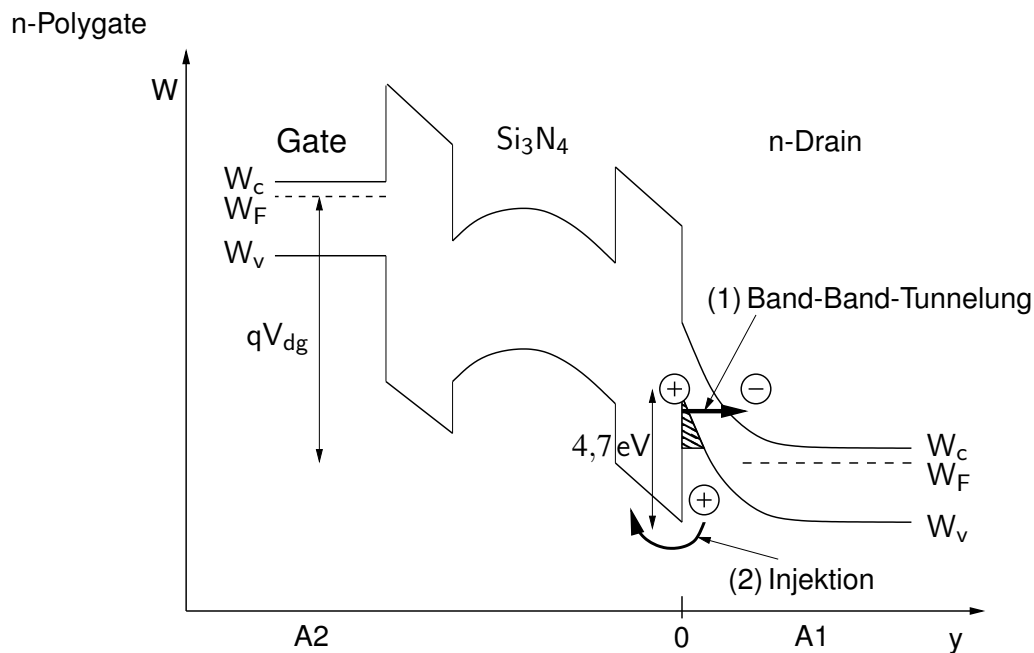
die Injektion benutzt. Die anderen Kanalelektronen tragen nur zum Drainstrom bei und führen damit zu einem hohen Leistungsbedarf bzw. zu einer geringen Programmier-effizienz.

### 2.2.3 Löschen der Speicherzelle

Der dominierende physikalische Effekt beim Löschen einer NROM-Speicherzelle ist nach [LPE04] das Löschen mit heißen Löchern. Dabei werden zuerst Elektronen-Loch-Paare erzeugt und dann wird den Löchern Energie zugeführt, so dass einige Löcher in das Nitrid injiziert werden können.

Um die Löschbedingung einzustellen, wird am Gate ein negatives Potential von ungefähr  $-5 \dots -8 \text{ V}$  angelegt. Am Drain liegt eine positive Drainspannung von  $4 \dots 7 \text{ V}$  an, während das Source im Allgemeinen unbeschaltet (floating) oder auf einem niedrigen Potential gehalten wird [KIM<sup>+</sup>05]. Wie in Abb. 2.9 dargestellt wurde, bildet sich in der Drain-Gate-Überlappungszone an der Oberfläche des n-Draingebiets aufgrund der hohen Drain-Gate-Spannung und der durch die Programmierung eingebrachten negativen Raumladung eine Verarmungszone aus.

In Abb. 2.10 wurde entlang der in Abb. 2.9 eingetragenen Schnittlinie durch die Verarmungszone A1–A2 das Bänderdiagramm aufgetragen. Dabei wird ersichtlich, dass aufgrund der hohen Spannungen die Bänder in der Verarmungszone sehr stark verbogen sind. Ist die Bandverbiegung der Valenzbandkante aber größer als der Bandabstand, wie

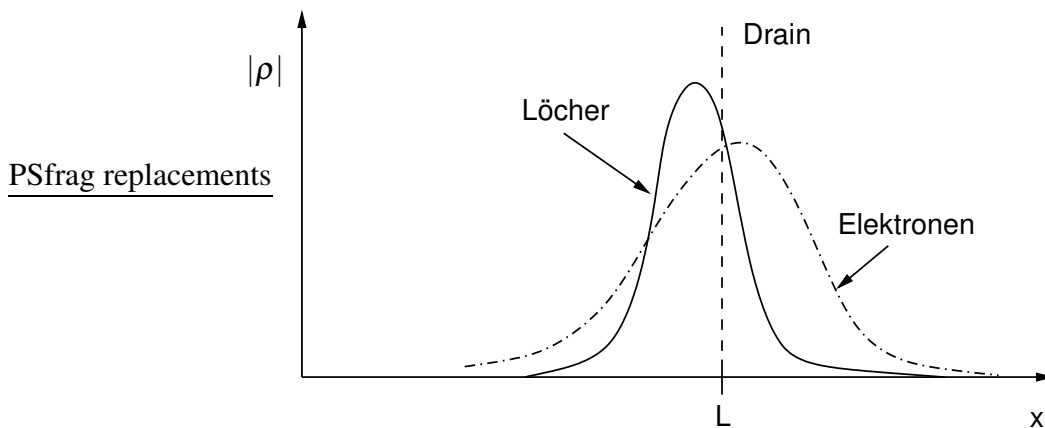


**Abbildung 2.10:** Bändermodell der ONO-Speicherzelle im Löschbetrieb; (1) Erzeugung der Löcher durch Band-Band-Tunnelung, (2) Injektion heißer Löcher in das Nitrid

im schraffiert markierten Bereich der Abbildung gezeigt wird, können bei hohen Feldstärken die Valenzelektronen in das Leitungsband tunneln. Folglich werden Elektronen-Loch-Paare erzeugt. Dieser physikalische Vorgang wird als Band-Band-Tunnelung bezeichnet [CL87][CCC<sup>+</sup>87].

Die generierten Elektronen bewegen sich zum Drainanschluss, während die Löcher durch die hohe Drain-Gate- bzw. Drain-Source-Spannung in Richtung der Silizium-Oxid-Grenzfläche und des Sourceanschlusses beschleunigt werden. Analog zu den beschriebenen Vorgängen bei der Programmierung können zufällig einige Löcher eine Energie, die größer als 4,7 eV ist, erlangen [BB98], so dass sie, falls sie in Richtung des ONO-Stapels gestreut werden, die Energiebarriere der Silizium-Oxid-Grenzfläche überwinden können. Diese Löcher können dann Haftstellen im Nitrid besetzen und die negative Ladung der bei der Programmierung injizierten Elektronen kompensieren.

Aufgrund der unterschiedlichen Feldverteilungen und der Verschiedenartigkeiten der physikalischen Mechanismen der Ladungsträgergeneration und -beschleunigung beim Löschvorgang im Vergleich zur Programmierung sind die lateralen Verteilungen der im Nitrid eingefangenen Löcher und Elektronen unterschiedlich, insbesondere befinden sich ihre Schwerpunkte an verschiedenen lateralen Positionen [ISS02]. Da die Ladungsträger eine sehr geringe Beweglichkeit innerhalb des Nitrids besitzen, findet keine kurzfristige Rekombination der Ladungsträger statt. Das Löschen eines programmierten ONO-Transistors beruht vielmehr auf der elektrischen Neutralisation der Löcher- und Elektronenpakete. Nach den Modellen [LSDBE02] [Jan03] [JES<sup>+</sup>04] führt eine langsame



**Abbildung 2.11:** Löschen und Programmieren erzeugt unterschiedliche Verteilungen der Löcher bzw. Elektronen.

laterale Umverteilung und der Ausgleich der Elektronen- und Löcherpopulationen allerdings zu einem allmählichen Absinken der Schwellspannung der programmierten Bits. Dieser Vorgang begrenzt die Datenhaltezeit der programmierten Speicherzellen (*Speicherungsverlust - Retention Loss*).<sup>1</sup>

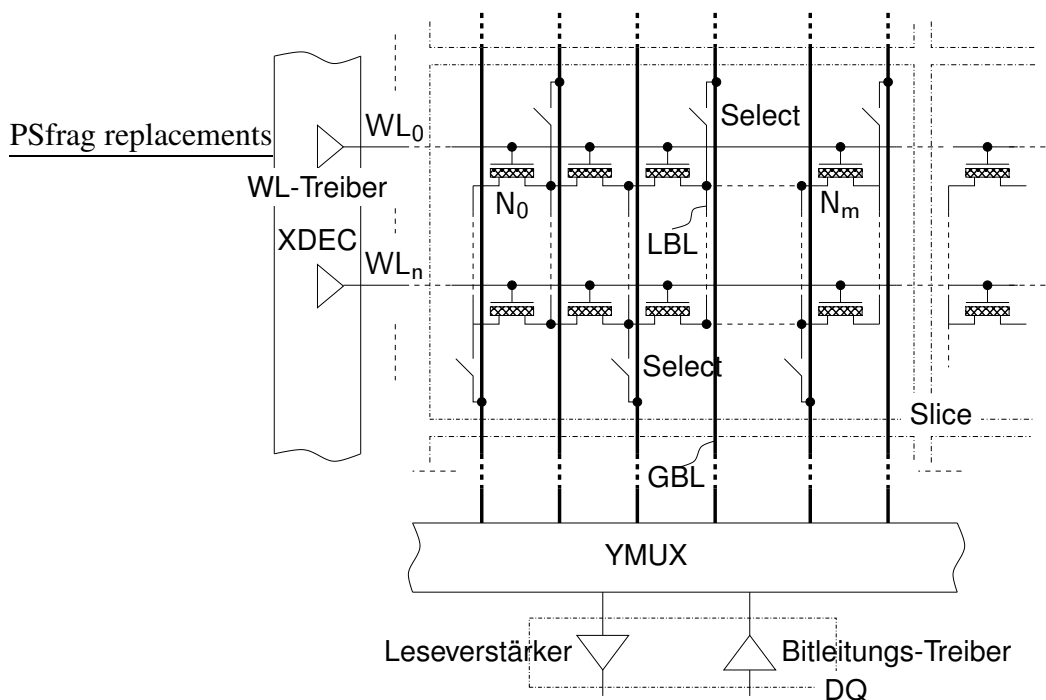
Ein ONO-Speichertransistor, der noch nicht programmiert oder gelöscht wurde, wird als Transistor im *initialen Zustand* oder *initialer Transistor* bezeichnet. Aufgrund der nicht übereinstimmenden Verteilungen der Löcher und Elektronen wird ein zunächst initialer ONO-Transistor nach erfolgter Programmierung und anschließendem Löschen nicht wieder exakt in den initialen Zustand zurückversetzt. Im Folgenden werden deshalb, wo notwendig, der initiale und der gelöschte Zustand unterschieden.

## 2.3 Aufbau eines Speicherfelds

Nachdem der Aufbau und die Funktionsweise der ONO-Speicherzelle beschrieben wurden, wird nun auf den Aufbau eines Speicherfelds (*Array*) eingegangen. Dabei wird sich auf die für Speichertransistoren mit lokaler Ladungsspeicherung typische Architektur des *Virtual Ground Arrays* beschränkt.

In Abb. 2.12 ist dargestellt, dass in dieser Arraystruktur einige der Speichertransistoren an ihren Gateanschlüssen mittels der sogenannten *Wortleitung* (WL) horizontal verbunden sind. Eine Anzahl von  $m + 1$  Speichertransistoren, zum Beispiel  $m = 32$  oder  $m = 65$  (inkl. Fehlerkorrekturbits), sind dabei jeweils durch die Verbindung ihrer Kanalkontakte in Serie geschaltet. Diese Transistoren bilden dabei typischerweise eine strukturelle und logische Einheit, die sogenannte Speicherseite oder *Page*. Speicherseiten, die hinterein-

<sup>1</sup> Andere Ansätze sehen als Ursache des Speicherverlusts einen Elektronenverlust durch das untere Oxid [THS06].



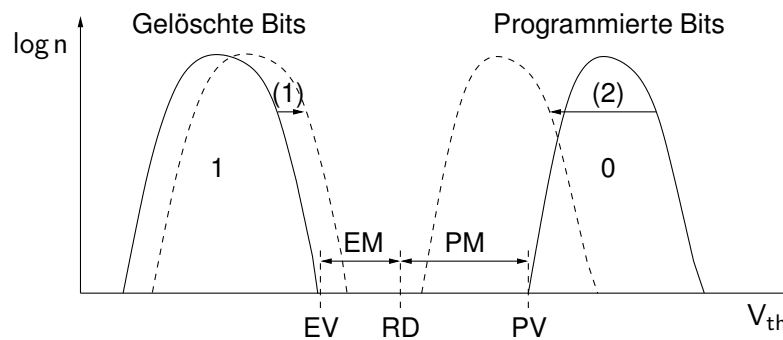
**Abbildung 2.12:** Aufbau eines Speicherfelds; Die Transistoren innerhalb eines Slices sind als *Virtual Ground Array* verschaltet.

ander entlang einer Wortleitung angeordnet sind, werden für gewöhnlich durch Isolation der Kanalkontakte der Endtransistoren elektrisch voneinander getrennt.

Die untereinander liegenden Speicherseiten von  $n + 1$  Wortleitungen, zum Beispiel  $n = 255$  oder  $n = 511$ , sind durch *Lokale Bitleitungen* (LBL) miteinander verbunden. Die  $(m + 1) \times (n + 1)$  Speichertransistoren bilden eine weitere strukturelle Einheit, die sogenannte *Slice*. Die Lokalen Bitleitungen der Slice können durch Schaltertransistoren (Select Transistors) von den *Globalen Bitleitungen* elektrisch getrennt werden. Dadurch beeinflussen hohe Drainspannungen, die beim Programmieren und Löschen auf einer Globalen Bitleitung auftreten, nicht diejenigen Slices, auf die bei diesen Operationen nicht zugegriffen wird. Charakteristische Störungen, sogenannte *Bitleitungsstörungen*, werden unterdrückt. Die Globalen Bitleitungen verbinden die Slices untereinander in vertikaler Richtung.

In Flash-Speichern werden beim Löschen ganze Slices oder große Untergruppen der Slices gleichzeitig gelöscht. Diese Blöcke haben in Abhängigkeit von der Speichergröße bei Daten-Flash-Speichern zum Beispiel eine Größe von 128 Kbyte. Dadurch unterscheiden sich die Flash-Speicher von den EEPROMs, in denen Bits einzeln gelöscht werden können.

Der Y-Multiplexer (YMUX) wählt in Abhängigkeit von der Speicheradresse des Bits, auf das im Array zugegriffen werden soll, zwei Globale Bitleitungen aus, die die für



**Abbildung 2.13:** Typische Häufigkeitsverteilung der Schwellspannungen in einem Speicherfeld und typische Verschiebungen innerhalb der Verteilung durch Raum-Temperatur-Drift (1) und Speicherverluste (Retention Loss) (2)

die jeweilige Betriebsart des Arrays notwendigen Source- und Drainspannungen übertragen. Dafür werden die entsprechenden Bitleitungen mit dem Bitleitungs-Treiber und im Lesebetrieb mit dem Leseverstärker verbunden.

Die Auswahl der Speicherseite geschieht im X-Dekoder (XDEC). Im X-Dekoder sorgen die Wortleitungs-Treiber dafür, dass die aktivierte Wortleitung mit der betriebsartspezifischen Gatespannung beschaltet wird, während die nichtadressierten, inaktiven Wortleitungen durch entsprechende Gatepotentiale deaktiviert werden.

Die Baugruppe, die den Bitleitungs-Treiber und Leseverstärker enthält, wird häufig vereinfachend als *DQ* bezeichnet. Um den Datendurchsatz zu erhöhen, werden Speichertransistoren in mehreren Slices gleichzeitig gelesen, programmiert oder gelöscht. Ein Flash-Speichersystem hat deshalb für gewöhnlich eine größere Anzahl von *DQs* und damit eine Vielzahl von Bitleitungs-Treibern und Leseverstärkern, die parallel betrieben werden.

## 2.4 Algorithmen

Die Lese-, Programmier- oder Löschoptionen auf dem beschriebenen Speicherfeld erfordern komplexe Algorithmen. Ziel der Algorithmen ist, dass die Daten so in das Speicherfeld eingeschrieben werden, dass der Dateninhalt jederzeit zuverlässig gelesen werden kann und dass bei Bedarf eine Veränderung der Daten durch Löschen und Neuprogrammierung möglich ist.

Aufgrund der physikalischen Prozesse mit den beschriebenen stochastischen Komponenten, die bei der Programmierung und dem Löschen angewendet werden, sowie aufgrund anderer systematischer und zufälliger Abweichungen sind die Schwellspannungen der Speicherzellen, die zu einem logischen Pegel gehören, in ihrer Häufigkeit um einen Zentralwert zufällig verteilt. Ein typisches Beispiel dieser Häufigkeitsverteilung wird in Abb. 2.13 gezeigt. Neben den Verteilungen unmittelbar nach dem Programmieren und

dem Löschen sind auch die typischen Verschiebungen der Schwellspannungsverteilungen nach langer Datenhaltezeit qualitativ dargestellt: Die Schwellspannung der gelöschten Bits steigt durch die sogenannte *Raum-Temperatur-Drift* kurz nach dem Löschen leicht an [BPE02], während die Schwellspannung der programmierten Bits im Vergleich dazu durch den Speicherverlust (Retention Loss) stärker absinkt [Jan03].

Bei der Leseoperation muss nun sicher entschieden werden können, zu welcher Verteilung und damit zu welchem logischen Pegel ein gelesenes Bit gehört. Um diese Sicherheit während der gesamten Datenhaltezeit und unter dem Einfluss verschiedenster Störungen garantieren zu können, werden Sicherheitsabstände (*Margins*) zwischen den Verteilungen eingefügt. Der Lese-Referenzwert  $RD$  muss so gewählt werden, dass er zwischen den beiden Verteilungen liegt. Der Abstand der Lesereferenz zum höchsten Schwellspannungswert der Verteilung der gelöschten Bits  $EV$  wird als *Erase Margin* (EM) bezeichnet und bestimmt die Sicherheit, mit der logische „1“-Werte gelesen werden können. Der Abstand der Lese-Referenz zum niedrigsten Schwellspannungswert der Population der programmierten Bits  $PV$  wird *Program Margin* (PM) genannt und bestimmt die Sicherheit des Erkennens einer gespeicherten logischen „0“. Typische Werte für den Erase Margin und den Program Margin sind 500 mV bzw. 1100 mV [SLCE06].

Zur Ausführung der komplexen Algorithmen, die die Erzeugung einer solchermaßen „sicheren“ Verteilung der Schwellspannung erlauben, sind in Flash-Systemen entweder Finite-State-Maschinen oder programmgesteuerte Mikrocontroller implementiert. Die Grundzüge dieser Algorithmen werden im Folgenden erläutert.

### 2.4.1 Lesealgorithmus

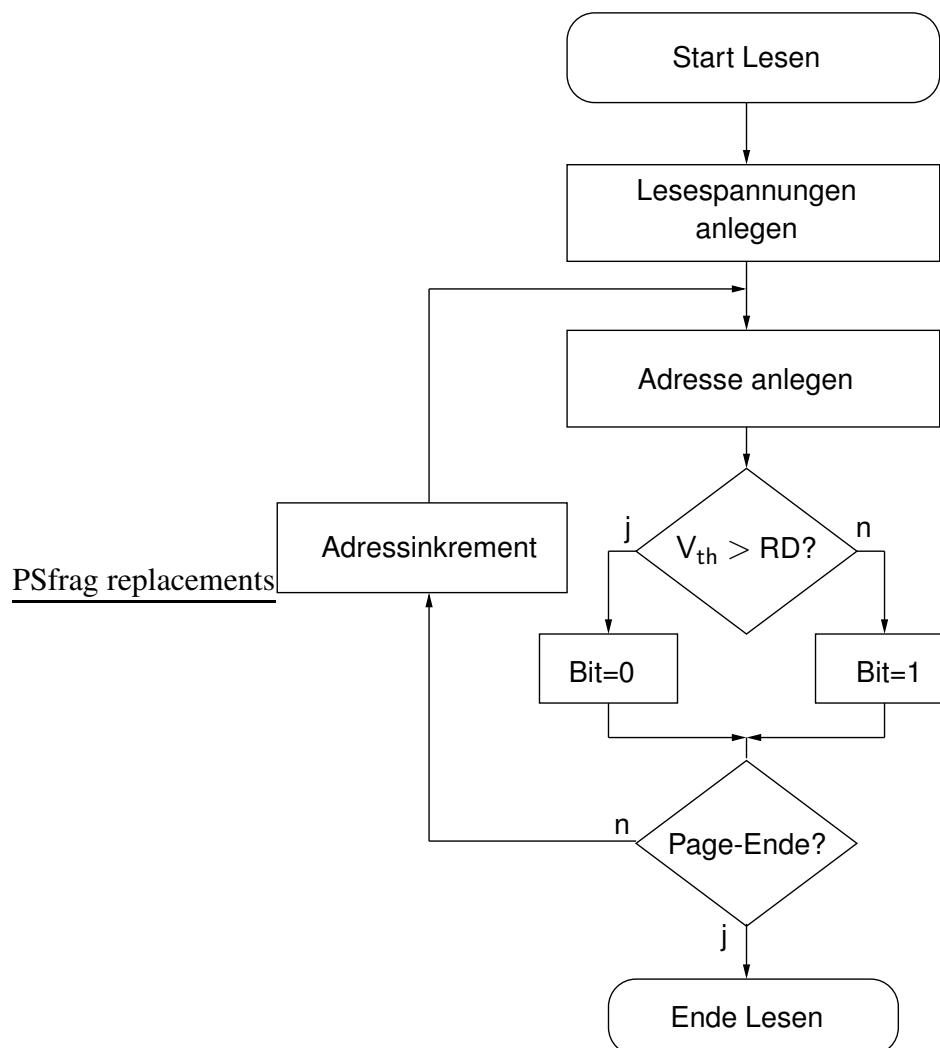
Das Auslesen einer Speicherseite des Speicherfelds beginnt, wie in Abb. 2.14 dargestellt, mit dem Einstellen der notwendigen Wortleitungs- und Bitleitungsspannung für den Lesebetrieb. Zu diesem Zweck werden im Schaltkreis die entsprechenden Ladungspumpen und Spannungsregler eingeschaltet. Durch das Anlegen der Speicheradresse wird mittels der X-Dekoder-, YMUX- und Select-Transistoren die zu aktivierende Speicherseite und der auszulesende Speichertransistor, meist zunächst der erste Transistor der Speicherseite, ausgewählt. Dabei wird der YMUX derart angesteuert, dass das auszulesende Bit mit dem niedrigeren Potential verbunden wird, also an der Sourceseite des Speichertransistors anliegt. Wie in Abschnitt 2.2.1 beschrieben wurde, erfolgt nun das Auslesen des Transistors im Leseverstärker durch einen Vergleich mit dem Lese-Referenzwert  $RD$ .<sup>2</sup>

Das Auslesen der Speicherseite erfolgt parallel in einer Vielzahl von DQs. Die gelesenen Bits aller DQs können, wie in Code-Flash-Speichern üblich, sofort über die Datenausgänge ausgegeben werden oder, wie bei Daten-Flash-Speichern verbreitet, in einem Pufferspeicher gesammelt werden.

Als nächster Schritt wird die neue Adresse angelegt, typischerweise durch das Inkre-

---

<sup>2</sup>Obwohl technisch eigentlich ein Vergleich mit einem Referenzstrom erfolgt, werden die Referenzwerte meist als äquivalente Schwellspannungswerte angegeben.



**Abbildung 2.14:** Algorithmus für das Auslesen einer Speicherseite (Page); RD ist die Lese-Referenzspannung.

mentieren der aktuellen Adresse. Dadurch wird innerhalb der Speicherseite das nächste Bit und gegebenenfalls der nächste Speichertransistor ausgewählt. Der Vergleich mit der Lese-Referenz, die Ausgabe bzw. Zwischenspeicherung der DQ-Daten und das Adressinkrement wiederholen sich, bis die komplette Speicherseite ausgelesen ist. Als Ergebnis stehen bei Daten-Flash-Speichern nun die Daten der Speicherseite in den Pufferspeichern für die Ausgabe über die Datenleitungen des Schaltkreises zur Verfügung.

Es soll darauf hingewiesen werden, dass der Vergleich im Leseverstärker auch gegen andere Referenzwerte als den Lese-Referenzwert  $RD$  erfolgen kann. Diese Option wird in den nachfolgend beschriebenen Programmier- und Löschalgorithmen verwendet.

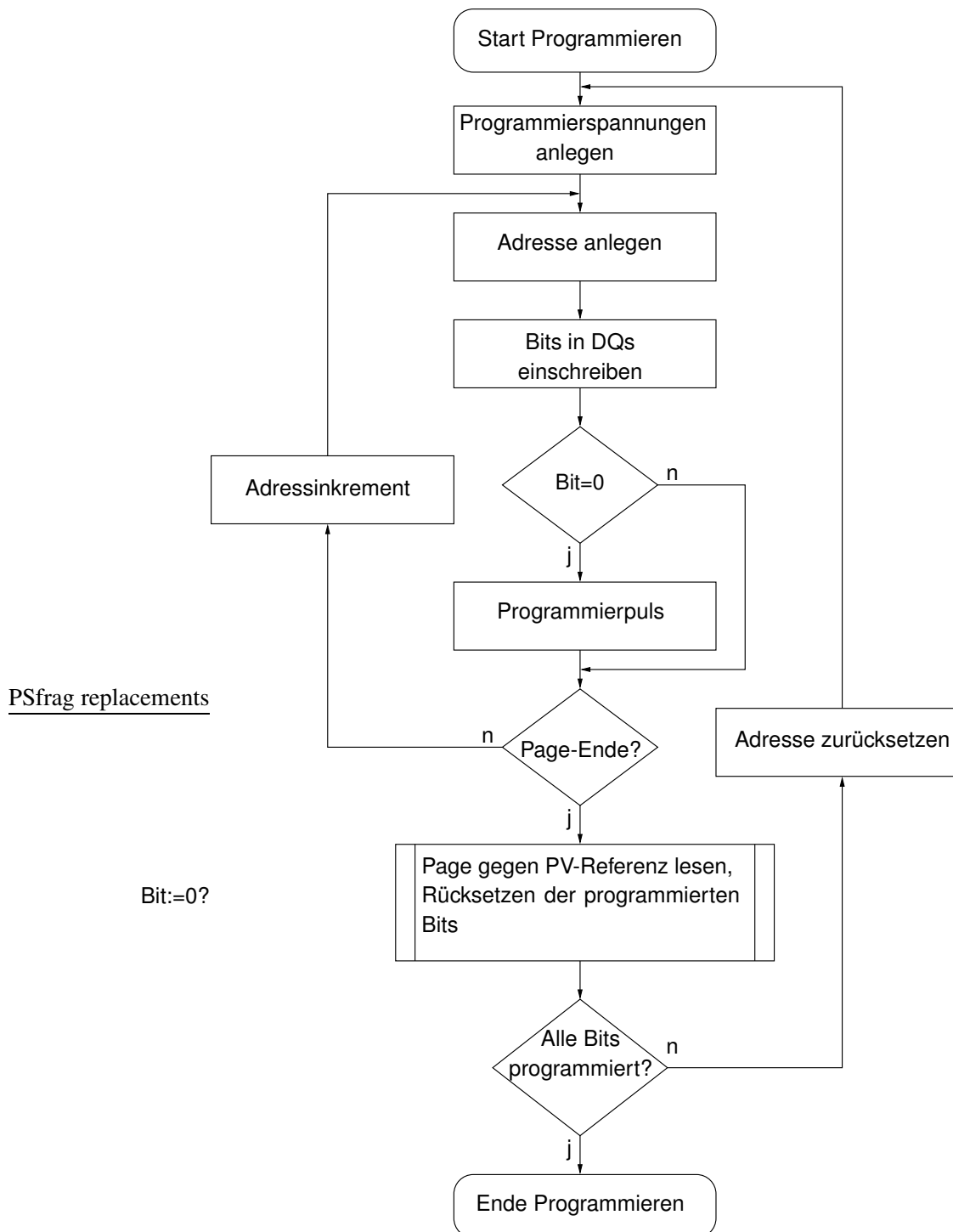
### 2.4.2 Programmialgorithmus

Zur Vorbereitung der Speicherseiten-Programmierung werden die Daten der zu programmierenden Speicherseite in einen Pufferspeicher eingeschrieben. Zu Beginn des Programmialgorithmus werden die Ladungspumpen und Spannungsregulatoren eingeschaltet und auf die notwendigen Wortleitungs- und Bitleitungsspannungen eingestellt. Durch Anlegen der Speicheradresse wird der betreffende Speichertransistor, gewöhnlich der erste Transistor der adressierten Speicherseite, mit Hilfe der X- und Y-Dekoder und der Select-Transistoren ausgewählt. Das zu programmierende Bit liegt auf der Drainseite des Transistors. Aus dem Pufferspeicher werden die Datenwerte der parallel programmierten, adressierten Bits gelesen und in die zugehörigen DQs eingeschrieben. Soll in einer DQ ein Bit programmiert werden, ist sein korrespondierender logischer Datenwert üblicherweise „0“. In diesem Fall wird an den Transistor ein Drainspannungspuls von  $\sim 5\text{ V}$  und  $\sim 200\text{ ns}$  Dauer angelegt [MDS<sup>+</sup>02]. Anschließend wird die Adresse inkrementiert, dadurch das nächste Bit ausgewählt und, nach dem Schreiben des zugehörigen Datenwerts aus dem Pufferspeicher in die DQ, gegebenenfalls programmiert. Dieser Vorgang wiederholt sich bis zum letzten zu programmierenden Datenwort, meist bis zum Ende der Speicherseite.

Nachdem nun ein erster Programmierzyklus abgearbeitet wurde, wird die komplette Speicherseite, wie in Abschnitt 2.4.1 beschrieben, ausgelesen (*Program-Verify*). Im Unterschied zum normalen Speicherseiten-Lesealgorithmus wird aber mit einer anderen Referenz, der Program-Verify-Referenz  $PV$ , verglichen. Mit Hilfe dieses Vergleichs wird festgestellt, ob die Schwellspannung eines zu programmierenden Bits bereits einen vorgeschriebenen Minimalwert überschritten hat. Ist der PV-Schwellspannungswert überschritten, kann der zum Bit korrespondierende Datenwert im Pufferspeicher auf den logischen Wert „1“ gesetzt werden, da in diesem Fall keine weiteren Programmierpulse für das Bit generiert werden. Anderenfalls bleibt der logische Datenwert des Bits im Pufferspeicher unverändert.

Nachdem die Adresse auf das erste zu programmierende Bit der Speicherseite zurückgesetzt wurde, wird die Programmier-Lese-Sequenz wiederholt ausgeführt bis im Endergebnis jedes der zu programmierenden Bits der Speicherseite das geforderte PV-Niveau überschritten hat.





**Abbildung 2.15:** Algorithmus für das Programmieren einer Speicherseite (Page); PV ist die Program-Verify-Referenzspannung.

### 2.4.3 Löschalgorithmus

Ähnlich zu den Lese- und Programmialgorithmen werden beim Löschen zuerst die Ladungspumpen und Spannungsregler eingeschaltet und auf die negative Wortleitungs- und positive Bitleitungsspannung eingestellt. Alle zum Block gehörenden Wortleitungen werden aktiviert und mit der Wortleitungsspannung beschaltet. Die beiden Bits des Speichertransistors werden in zwei aufeinander folgenden Schritten gelöscht, wobei die jeweils zu löschenden Bits an den Drainseiten der Transistoren anliegen und parallel mit dem Bitleitungstreiber verbunden werden, während die anderen Seiten der Transistoren im Allgemeinen unbeschaltet bleiben. Durch einen drainseitigen Löschpuls von  $4 \dots 7 \text{ V}$  und einer Dauer von ungefähr  $\sim 250 \mu\text{s}$  wird die erste Gruppe von Bits gelöscht und nach Vertauschung von Source und Drain die zweite Gruppe [BPE02].

Anschließend wird der ganze Block gelesen und mit einer Erase-Verify-Referenz *EV* verglichen. Bei diesem Lesevorgang ist allerdings nicht der konkrete Dateninhalt relevant, sondern nur, ob mindestens ein Bit des Blocks eine Schwellspannung hat, die größer als die EV-Referenz ist. Ist dies der Fall, erhält der komplette Block einen weiteren doppelseitigen Löschpuls. Das Löschen ist beendet, wenn alle Bits des Blocks eine Schwellspannung unterhalb der EV-Referenzschwellspannung haben.

## 2.5 Multilevel-Speicher

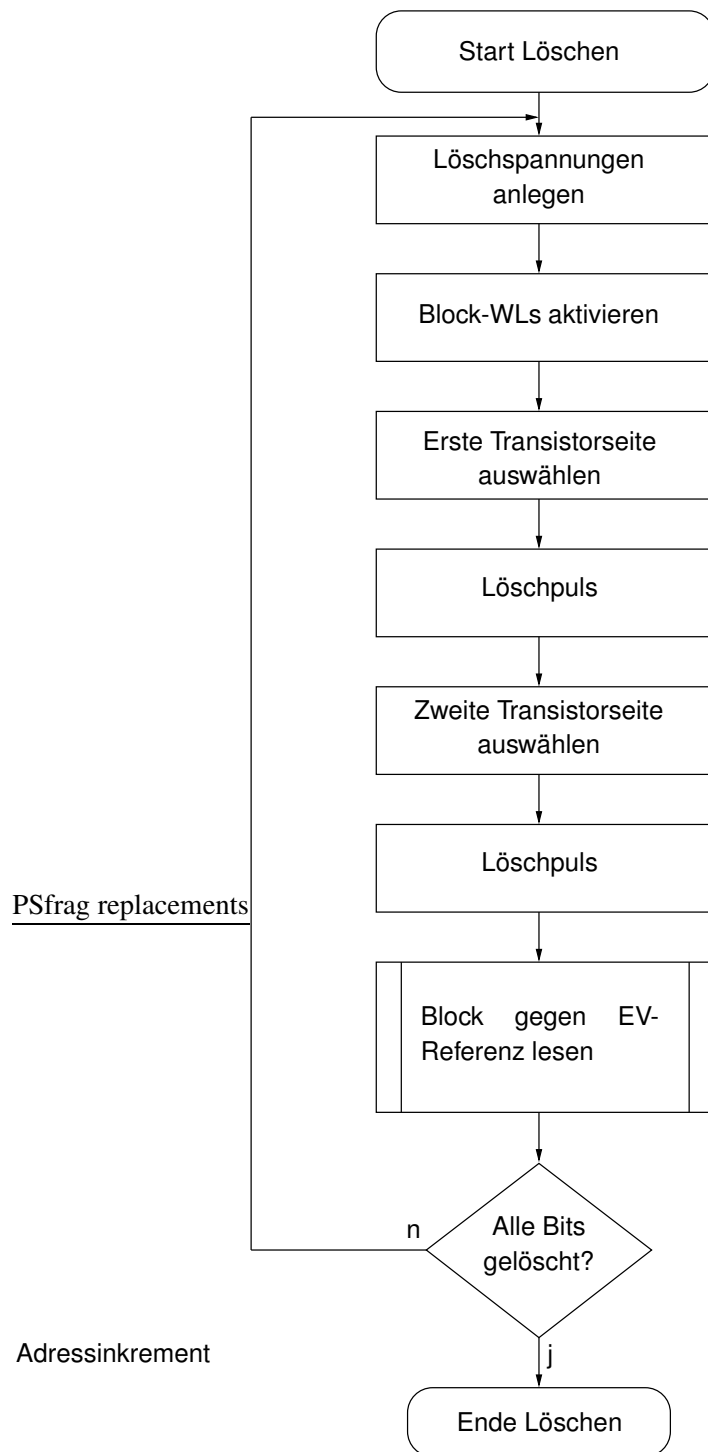
Um die Speicherdichte immer weiter zu erhöhen, wird aktiv an der Entwicklung der Multilevel-Speicher auf der Basis der ONO-Speichertransistoren mit lokaler Ladungsspeicherung gearbeitet. Bei dieser Technologie können auf jeder Seite der Speichertransistoren mehrere Bits gespeichert werden [PLS<sup>+</sup>06].

In Abb. 2.17 wird eine typische Häufigkeitsverteilung der Schwellspannungen eines Multilevel-Produkts gezeigt [PLS<sup>+</sup>06][SLCE06], in dem pro Transistorseite zwei Bits gespeichert sind. Aufgrund der beiden gespeicherten Bits sind vier Populationen in den Häufigkeitsverteilungen der Schwellspannungen zu erkennen. Jeder Population ist eine Bitkombination zugeordnet. Weil beim Auslesen zwischen den Populationen unterschieden werden muss, sind drei Lese-Referenzwerte  $RD1 \dots RD3$  vorhanden.

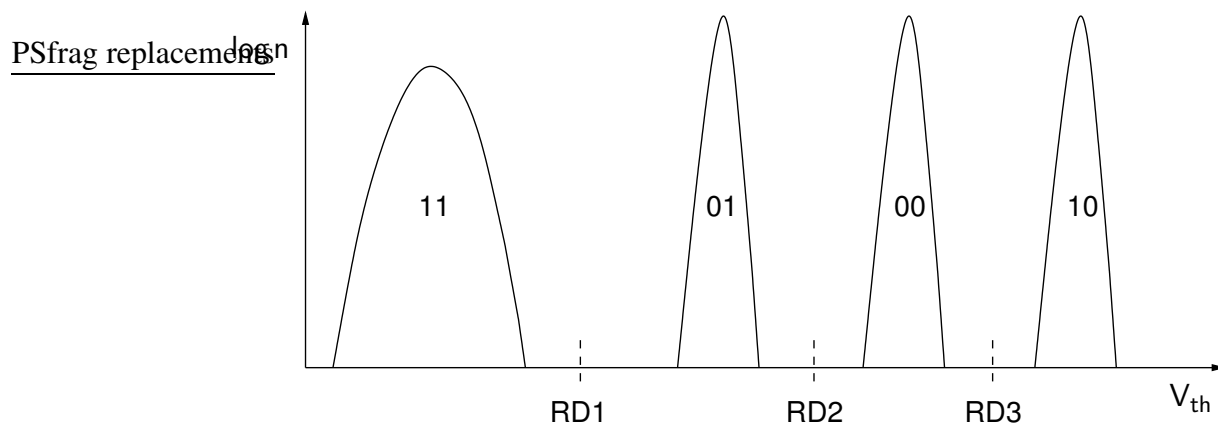
Da also bei Multilevel-Speichern die Anzahl der Schwellspannungspopulationen exponentiell mit der Bitzahl steigt, die maximalen Schwellspannungswerte aber nicht beliebig gesteigert werden können, müssen die Schwellspannungsverteilungen schmaler werden und/oder die Sicherheitsabstände zwischen den Verteilungen sinken.

Um bei der Programmierung derartig schmale Verteilungen zu erhalten, werden Algorithmen eingesetzt, die zunächst aggressiv mit großen Schwellspannungsinkrementen die Zielschwellspannung grob einstellen und dann mit kleineren Schwellspannungsinkrementen eine Feinjustage der Zielschwellspannung erlauben [PLS<sup>+</sup>06].

Beim Lesen werden Fehlererkennungsmechanismen eingesetzt. Wird zum Beispiel ein Lesefehler durch eine verschobene Schwellspannungsverteilung erkannt, wird versucht,



**Abbildung 2.16:** Algorithmus für das Löschen eines Blocks; EV ist die Erase-Verify-Referenzspannung.



**Abbildung 2.17:** Typische Häufigkeitsverteilung der Schwellspannungen in einem Multilevel-Speicher

die Lese-Referenz so nachzuführen, dass sie sich wieder optimal in der Lücke zwischen den Verteilungen befindet [PLS<sup>+</sup>06].

Trotz dieser fortschrittlichen Mechanismen ist die genaue Kenntnis der physikalischen Mechanismen, die eine Verteilungsverbreiterung oder Verteilungsverschiebung bewirken, unerlässlich und deshalb sind einige dieser Effekte Gegenstand dieser Arbeit.

## 3 Schwellspannung von MOS-Transistoren

In diesem Abschnitt werden die halbleiterphysikalischen Modellierungsmethoden zur Berechnung der Schwellspannung von konventionellen kurzkanaligen MOS-Transistoren und die dazugehörigen mathematischen Lösungsmethoden vorgestellt und bewertet. Ausgehend von dieser Untersuchung wird ein Modell zur Berechnung der Schwellspannung von Standard-MOS-Transistoren mit kurzen Kanallängen eingeführt und Schlussfolgerungen diskutiert.

### 3.1 Grundlagen zur Modellierung der Schwellspannung

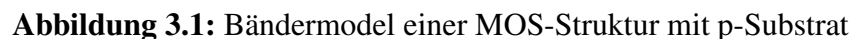
#### 3.1.1 Bändermodell

Das Bändermodell ist eine Veranschaulichung der Energie der Halbleiterelektronen. In Abb. 3.1 wird das Bändermodell einer Metall-Oxid-Halbleiter-Struktur (MOS-Struktur) gezeigt. Der Halbleiter sei p-dotiert, so dass sich die Fermienergie  $W_F$  unterhalb der Eigenleitungsenergie  $W_i$  befindet. Der Halbleiter sei in x-Richtung unendlich ausgehend, um Randeffekte auszuschließen. Der Halbleiter befinde sich im thermodynamischen Gleichgewicht, es gelte  $V_{gb} = 0$ .

Im Bändermodell Abb. 3.1 sind der Verlauf des Eigenleitungsniveaus  $W_i$ , der Energien der Leitungs- und Valenzbänder  $W_c$  bzw.  $W_v$ , des Fermi-niveaus und des Makropotentials  $\phi$  entlang eines Schnittes parallel zur y-Achse dargestellt.

Aufgrund des Unterschieds der Austrittsarbeiten der Elektronen aus dem Metall  $W_m$  und dem Halbleiter  $W_s$  sowie aufgrund von Ladungen an der Grenzfläche Oxid-Halbleiter  $Q''_{ss}$  oder von Ladungen im Isolator  $Q''_{ox}$  verschieben sich die Energieniveaus im Halbleiter an der Grenzfläche zum Oxid so weit, bis sich über dem Gateisolator eine Spannung  $V_{ox}$  einstellt, bei der die Gesamtladungsbilanz des Halbleiters ausgeglichen ist. Weit im Inneren des Halbleiters sind die energetischen Verhältnisse aber weiter ungestört, so dass sich im Allgemeinen eine Verbiegung der Bänder an der Halbleiteroberfläche, selbst bei  $V_{gb} = 0$ , ergibt. Diese Bandverbiegung kann kompensiert werden, indem eine entsprechende Gate-Bulk-Spannung, die Flachbandspannung  $V_{fb}$ , extern angelegt wird. Unter Einfluss dieser Bandverbiegung entsteht im Halbleiter, auch im thermodynamischen Gleichgewicht, eine Raumladungszone.

Den potentiellen Energien des Bändermodells können auch elektrostatische Potentiale zugeordnet werden. Ebenso wie bei den potentiellen Energien, ist die Lage des Bezugsniveaus der Potentiale beliebig. Wir definieren das Potential  $\psi$  auf das Eigenleitungsniveau und das Potential  $\phi$  auf das Fermi-niveau im Halbleiterinneren bezogen, woraus sich für



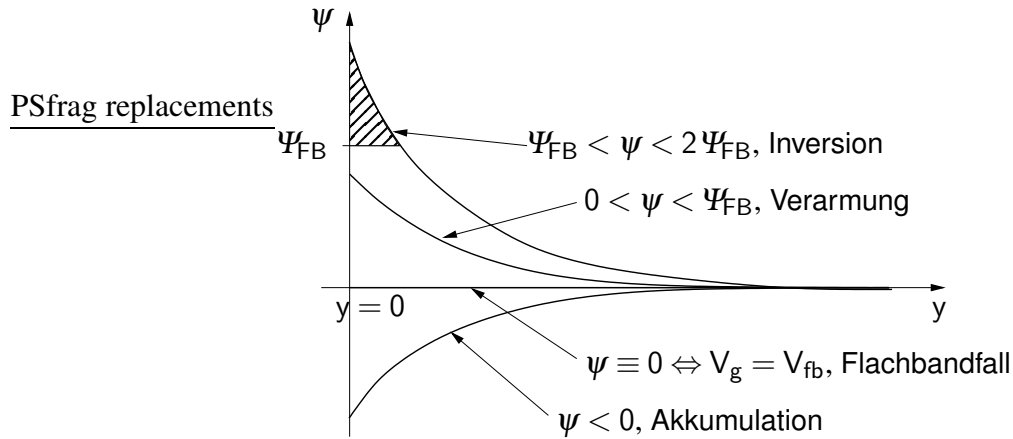


Abbildung 3.2: Betriebsfälle einer MOS-Struktur mit p-Substrat

3. **Verarmung:** Das Potential ist positiv, erreicht aber nicht das Fermi-niveau  $\Psi_{FB}$ . In diesem Fall überwiegt in der Raumladungszone immer noch die gleiche Ladungsträgerart wie innerhalb des Halbleiters, allerdings mit niedrigerer Dichte als dort.
4. **Inversion:** Das Potential ist größer als  $\Psi_{FB}$ . Es gibt in diesem Gebiet eine höhere Dichte der Minoritätsladungsträger des Halbleiterinneren als der entsprechenden Majoritätsladungen. Bei einem p-Halbleiter überwiegen also in dem Inversionsgebiet die Elektronen. Sind sie in genügender Dichte vorhanden, kann in nMOS-Transistoren ein Stromtransport einsetzen.

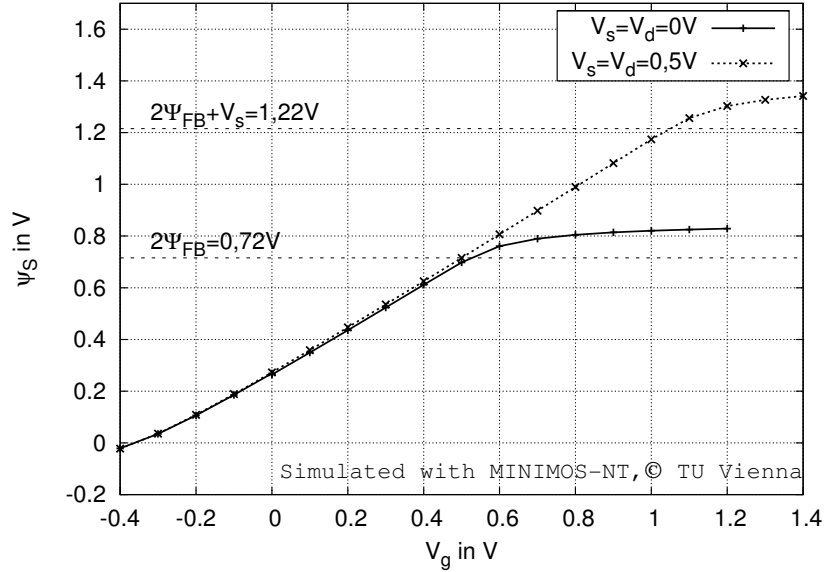
### 3.1.2 Definition der Schwellspannung

Bildet sich eine Inversionsschicht aus, nimmt die Dichte der die Inversionsschicht bildenden Ladungsträger gemäß Gl. (3.4) exponentiell mit dem Halbleiterpotential zu. Wie in Abb. 3.3 gezeigt, erreicht das Potential einen Sättigungswert, wenn sich im MOS-Transistor die Inversionsschicht ausbildet. Wie in [Tsi99] diskutiert wurde und in Abb. 3.3 erkennbar ist, liegt dieser Sättigungswert näherungsweise, bis auf einige Vielfache der Thermospannung  $U_T$ , beim doppelten Wert des Fermi-niveaus des Substrates  $\Psi_{FB}$ .

Es ist üblich, die Schwellspannung Source-referenziert zu definieren. Wir berechnen die Schwellspannung also als die Gate-Source-Spannung, die notwendig ist, um den Sättigungswert in Abb. 3.3 zu erreichen. Trotz der erwähnten nur annähernden Bestimmung des Sättigungswerts ist es üblich, die Schwellspannung  $V_{th}$  wie folgt zu definieren:

$$\psi_0 := 2\Psi_{FB} + V_s, \quad (3.5)$$

$$V_{th} := (V_g - V_s) \big|_{\psi_{S,min}=\psi_0}. \quad (3.6)$$



**Abbildung 3.3:** Simulation des Oberflächenpotentials  $\psi_S$  in Abhängigkeit vom Gatepotential  $V_g$  für  $V_s = 0V$  und  $V_s = 0,5V$  für einen Langkanaltransistor

Dabei wird weiterhin von der Annahme ausgegangen, dass sich die maximale Bandverbiegung im Halbleiter an der Oxid-Halbleiter-Grenzfläche einstellt und sich damit im Allgemeinen der Inversionskanal an der Halbleiteroberfläche ausbildet.<sup>1</sup> Aus diesem Grund wird die Definition der Schwellspannung gemäß Gl. (3.6) auf das Oberflächenpotential  $\psi_S \equiv \psi(y=0)$  bezogen [Tsi99][Pau72][ML86].

Zur Ermittlung der Schwellspannung muss also zunächst der Verlauf des Oberflächenpotentials  $\psi_S(x)$  bekannt sein. Als zweiter Schritt wird danach das Minimum des Oberflächenpotentials  $\psi_{S,min}$  berechnet und dieses zur Bestimmung der Schwellspannung des Transistors benutzt.

## 3.2 Elektrostatisches Potential

### 3.2.1 Herleitung der Differentialgleichung des Potentialverlaufs

In Abb. 3.4 wird ein Schnittbild eines MOS-Transistors gezeigt. Zur Ermittlung der Schwellspannung dieser Anordnung muss zunächst das elektrostatische Potential berechnet werden. Der Potentialverlauf in Halbleitern wird durch die Poissongleichung beschrieben:

$$\text{div grad } \psi = -\frac{\rho}{\varepsilon}. \quad (3.7)$$

<sup>1</sup>Die Gültigkeit dieser Annahme wird für den Spezialfall der ONO-Transistoren im Abschnitt 4.2.4 eingehender untersucht.



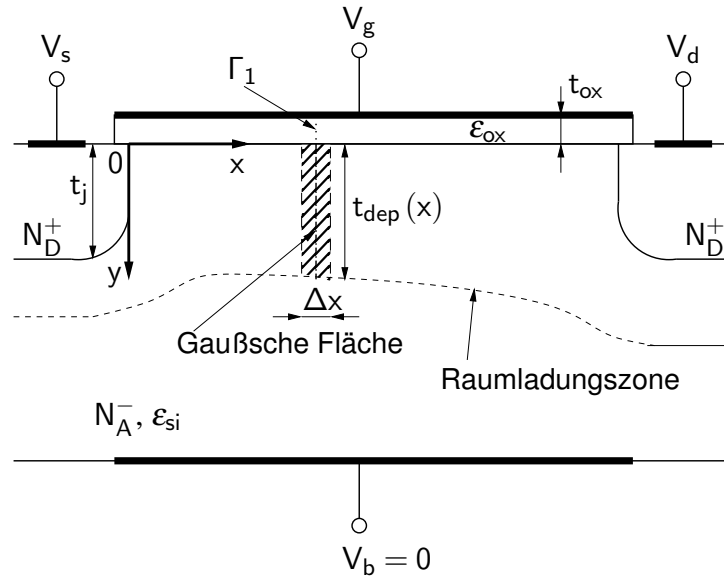


Abbildung 3.4: Schnittbild eines MOS-Transistors

Unter Annahme eines p-dotierten Silizium-Halbleiters, einer Störstellenerschöpfung und einer Verarmung in der Raumladungszone unterhalb des Gates ( $p, n \ll N_A$ ) lässt sich die Gleichung Gl. (3.7) für das Raumladungsgebiet  $0 \leq y < t_{dep}$  folgendermaßen vereinfachen:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{q N_A}{\epsilon_{si}}. \quad (3.8)$$

Neben numerischen Lösungsverfahren (zum Beispiel in [GD80]) werden bei einer analytischen Lösung der Poissonsgleichung verschiedene halbleiterphysikalische Modellierungsmethoden und angepasste mathematische Lösungsmethoden angewendet.

Bei den Charge-Sharing-Ansätzen wird die Poissonsgleichung meist für Langkanaltransistoren, für die die Annahme  $\partial \psi / \partial x = 0$  gilt, gelöst. Der Einfluss von Source- bzw. Drain-Gebieten in Kurzkanaltransistoren wird bei diesen Verfahren durch eine Reduktion der effektiven Bulkladung  $q N_A$  berücksichtigt. Dazu werden die ortsfesten Ladungen in den Raumladungszonen in der Nähe von Source bzw. Drain entweder den Source-/Drain-Gebieten oder dem Bulk-Gebiet zugeordnet. Diese Zuordnung der Ladungen geschieht dabei im Allgemeinen durch eine willkürliche geometrische Aufteilung. Da sich die effektive Bulkladung durch diese teilweise Zuordnung der Ladungen zu Source bzw. Drain verringert, wird in den Charge-Sharing-Ansätzen mit Hilfe eines Reduktionsfaktors die berechnete Langkanal-Schwellschpannung korrigiert [Yau74][Lee73][AS82][Tay78]. Diese Modelle erwiesen sich für die Modellierung von kurzen Transistoren als zu ungenau [RM82].

Andere Verfahren nutzen zur Lösung der Differentialgleichung Gl. (3.8) eine Separation der Gleichung in homogenen und inhomogenen Anteil durch entsprechende An-

satzfunktionen und lösen die sich ergebenden partiellen Differentialgleichungen [RM82] [KB86]. Trennung der Variablen kann ebenfalls als ein direktes Lösungsverfahren für Gl. (3.8) verwendet werden [PK84]. Greensche Funktionen werden in [LW87] angewendet. Alle diese mathematischen Verfahren führen ebenso wie die Analyseverfahren in [PSM85][ADM95] [FTW98] auf komplizierte unendliche Reihen als Lösungsfunktion und sind somit nur schwer weiterführend mathematisch behandelbar.

Eine weitere Klasse von Lösungsverfahren nutzt Methoden der Variationsrechnung zur Lösung von Gl. (3.8) und erhält damit ebenfalls sehr komplexe Lösungen [PM89].

Eine Methode zur Transformation der zweidimensionalen Form der Poissongleichung in eine eindimensionale Form wird von [SMP88] verwendet. Dazu wird ein quadratischer Verlauf des Oberflächenpotentials vorausgesetzt und ein nichtlineares Koordinatensystem eingeführt, dass die Poissongleichung in beschränkten Bereichen nahe den Achsen dieses Koordinatensystems in eine einfachere, eindimensionale Form überführt. Zur Lösung der sich ergebenden Gleichungen werden allerdings aufwendige iterative Verfahren benötigt.

Sehr erfolgreich wurden quasi-zweidimensionale Lösungsverfahren angewendet, bei denen die partielle Differentialgleichung durch geeignete Annahmen oder Ansatzfunktionen auf eine oder mehrere gewöhnliche Differentialgleichungen abgebildet wird, welche dann im Allgemeinen einfacher lösbar sind. Diese Verfahren führen dabei häufig auf relativ einfache, analytische Lösungen, die einer weiteren mathematischen Behandlung zugänglich sind, wobei die Lösungen sehr gut mit numerischen Simulationen korrelieren [TA79][CT91][FS93][LHH<sup>+</sup>93]. Aus diesem Grund basiert die folgende Herleitung auf diesem Verfahren. Der Herleitungsweg entspricht dabei im Wesentlichen [CT91].

Zur Lösung der Poissongleichung wählen wir unter Bezugnahme auf Abb. 3.4 die folgenden Randbedingungen:

$$\psi(0, y) = V_{bi} + V_s; \quad 0 \leq y < t_j, \quad (3.9a)$$

$$\psi(L, y) = V_{bi} + V_d; \quad 0 \leq y < t_j, \quad (3.9b)$$

$$\left. \frac{\partial \psi(x, y)}{\partial y} \right|_{y=t_{dep}} = 0, \quad (3.9c)$$

$$\psi(x, t_{dep}) = 0. \quad (3.9d)$$

Dabei bezeichnet  $V_{bi}$  die Diffusionsspannung  $V_{bi} = U_T \ln(N_A N_D / n_i^2)$ ,  $t_j$  ist die Tiefe der Source- bzw. Drain-Gebiete und  $L$  ist die Transistorlänge.

Zunächst integrieren wir die Poissongleichung Gl. (3.8) zweimal formal über  $y$ :

$$\psi(x, y) = \frac{q N_A}{2 \epsilon_{si}} y^2 + C_1 y + C_2 - \int_0^y \int_0^{y_1} \frac{\partial^2 \psi(x, y_2)}{\partial x^2} dy_2 dy_1. \quad (3.10)$$

Aus der Randbedingung Gl. (3.9c) ergibt sich:

$$C_1 = -\frac{qN_A}{\epsilon_{si}} t_{dep} + \int_0^{t_{dep}} \frac{\partial^2 \psi(x, y_1)}{\partial x^2} dy_1. \quad (3.11)$$

Eingesetzt in Gl. (3.10) und unter Anwendung der Randbedingung Gl. (3.9d) erhalten wir:

$$C_2 = \frac{qN_A}{2\epsilon_{si}} t_{dep}^2 - t_{dep} \int_0^{t_{dep}} \frac{\partial^2 \psi(x, y_1)}{\partial x^2} dy_1 + \int_0^{t_{dep}} \int_0^{y_1} \frac{\partial^2 \psi(x, y_2)}{\partial x^2} dy_2 dy_1. \quad (3.12)$$

Damit ergibt sich schließlich für  $0 \leq y < t_{dep}$ :

$$\begin{aligned} \psi(x, y) = & \frac{qN_A}{2\epsilon_{si}} (t_{dep} - y)^2 - (t_{dep} - y) \int_0^{t_{dep}} \frac{\partial^2 \psi(x, y_1)}{\partial x^2} dy_1 + \\ & + \int_0^{t_{dep}} \int_0^{y_1} \frac{\partial^2 \psi(x, y_2)}{\partial x^2} dy_2 dy_1 - \int_0^y \int_0^{y_1} \frac{\partial^2 \psi(x, y_2)}{\partial x^2} dy_2 dy_1. \end{aligned} \quad (3.13)$$

Eine Betrachtung der Gl. (3.13) lässt erkennen, dass man die Gleichung für Langkanaltransistoren aufgrund von  $\partial^2 \psi / \partial x^2 = 0$  auf den bekannten Langkanal-Potentialverlauf

$$\psi(x, y) = \psi(y) = \frac{qN_A}{2\epsilon_{si}} (t_{dep} - y)^2 \quad (3.14)$$

reduzieren kann.

Neben der Funktion  $\partial^2 \psi(x, y) / \partial x^2$  benötigt man zur Lösung dieser Gleichung die Tiefe der Raumladungszone  $t_{dep} = t_{dep}(x)$ . Für die Bestimmung von  $t_{dep}(x)$  betrachten wir die schraffierte Fläche in Abb. 3.4. Wenden wir das Gaußsche Gesetz auf diese Fläche an und beachten wir dabei, dass sich aufgrund der oben gemachten Näherungen nur Akzeptorladungen in der Gaußschen Fläche befinden, so kann man für kleine  $\Delta x$  formulieren:

$$\begin{aligned} Q' = -qN_A \Delta x t_{dep} = & \epsilon_{si} \int_0^{t_{dep}} E_x(x + \Delta x, y_1) dy_1 - \epsilon_{si} \int_0^{t_{dep}} E_x(x, y_1) dy_1 + \\ & + \epsilon_{si} E_y(x, t_{dep}) \Delta x - \epsilon_{si} E_y(x, 0) \Delta x. \end{aligned} \quad (3.15)$$

Dabei bezeichnen  $E_x$  und  $E_y$  die x-Komponente bzw. y-Komponente des elektrischen Feldes  $\vec{E}(x, y)$  und es gilt  $E_x = -\partial \psi / \partial x$  und  $E_y = -\partial \psi / \partial y$ . Es ergibt sich damit für  $\Delta x \rightarrow 0$  und unter Beachtung der Randbedingung Gl. (3.9c):

$$\frac{qN_A t_{dep}}{\epsilon_{si}} = \int_0^{t_{dep}} \frac{\partial^2 \psi(x, y_1)}{\partial x^2} dy_1 + E_y(x, 0). \quad (3.16)$$

Für die Oxid-Halbleiter-Grenzfläche bei  $y = 0$  gilt aufgrund des Gaußschen Gesetzes:

$$\epsilon_{si} E_y(x, 0) = \epsilon_{ox} E_{ox} + Q''_{ss}, \quad (3.17)$$

wobei die Feldstärke  $E_{ox}$  konstant innerhalb des Oxids ist und eine Integration dieser Feldstärke entlang des Wegs  $\Gamma_1$  aus Abb. 3.4 ergibt:

$$\int_{\Gamma_1} E_{ox} dy = E_{ox} t_{ox} = V_g - \Phi_{MS} - \psi(x, 0). \quad (3.18)$$

Aus Gl. (3.16) – Gl. (3.18) erhält man schließlich zusammen mit Gl. (3.13) das folgende allgemeine Gleichungssystem zur Berechnung des Potentialverlaufs:

$$\frac{q N_A t_{dep}}{\epsilon_{si}} = \int_0^{t_{dep}} \frac{\partial^2 \psi(x, y_1)}{\partial x^2} dy_1 + \frac{\epsilon_{ox}}{\epsilon_{si}} \cdot \frac{V_g - V_{fb} - \psi(x, 0)}{t_{ox}}, \quad (3.19a)$$

$$\begin{aligned} \psi(x, y) = & \frac{q N_A}{2 \epsilon_{si}} (t_{dep} - y)^2 - (t_{dep} - y) \int_0^{t_{dep}} \frac{\partial^2 \psi(x, y_1)}{\partial x^2} dy_1 + \\ & + \int_0^{t_{dep}} \int_0^{y_1} \frac{\partial^2 \psi(x, y_2)}{\partial x^2} dy_2 dy_1 - \int_0^y \int_0^{y_1} \frac{\partial^2 \psi(x, y_2)}{\partial x^2} dy_2 dy_1, \end{aligned} \quad (3.19b)$$

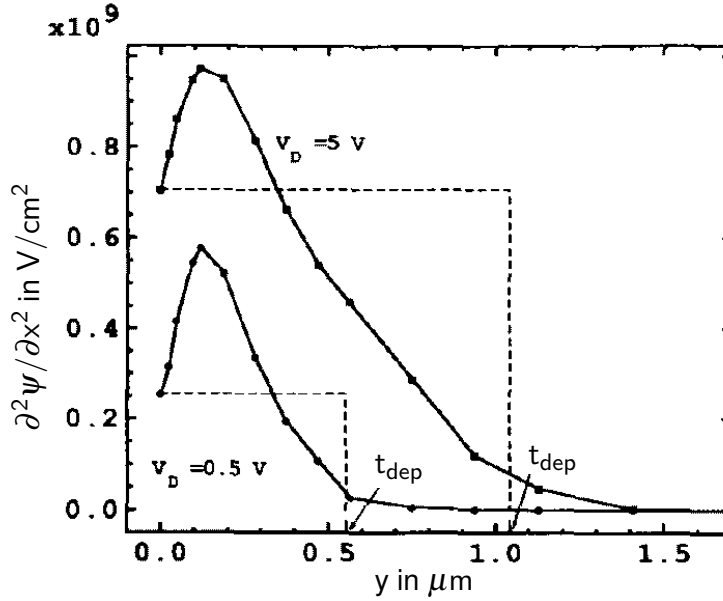
wobei die Flachbandspannung  $V_{fb}$  der Definition aus Gl. (3.2) entspricht.

Ein Vergleich von Gl. (3.19a) und Gl. (3.19b) lässt erkennen, dass beide Ausdrücke über  $\partial^2 \psi(x, y) / \partial x^2$  und  $t_{dep} = t_{dep}(x)$  gekoppelt sind. Zur Lösung dieses Systems ist es zweckmäßig, eine Ansatzfunktion für  $\partial^2 \psi(x, y) / \partial x^2$  zu finden, die es erlaubt, die Gleichungen weiterführend mathematisch zu behandeln, gleichzeitig aber Modellierungsergebnisse liefert, die mit numerischen Simulationsergebnissen gut korrelieren. In [CT91] wurde dazu der Ansatz:

$$\frac{\partial^2 \psi(x, y)}{\partial x^2} = \frac{\partial^2 \psi(x, 0)}{\partial x^2} = \frac{d^2 \psi_S(x)}{dx^2} \quad (3.20)$$

gewählt. Das bedeutet, dass angenommen wird, dass sich die Krümmung des Potentialverlaufs in x-Richtung über die gesamte Tiefe der Raumladungszone nur wenig ändert und dass sie mit der Krümmung an Oxid-Halbleiter-Grenzfläche übereinstimmt. Obwohl sich diese Wahl nicht streng rechtfertigen lässt und Abb. 3.5 zeigt, dass die Krümmung doch entlang der y-Richtung variiert, kann man ebenfalls in dieser Abbildung erkennen, dass der Mittelwert des Krümmungsverlaufs über die gesamte Tiefe der Raumladungszone gut mit dem entsprechenden Mittelwert der Ansatzfunktion übereinstimmt. Das Einfach-Integral in Gl. (3.19b) und das Integral in Gl. (3.19a) entsprechen aber genau diesem Mittelwert. Insofern können diese Integrale gut angenähert werden. Die

PSfrag replacements



**Abbildung 3.5:** Verlauf von  $\partial^2\psi/\partial x^2$  und Approximation für verschiedene Drain-Spannungen [CT91]; Mit freundlicher Genehmigung ©1991 IEEE

Doppelintegrale in Gl. (3.19b) werden aber durch den gewählten Ansatz weniger gut repräsentiert. Die sich aus dem Ansatz ergebende Lösung muss deshalb unbedingt mit numerischen Simulationsergebnissen verglichen werden.

Anzumerken ist, dass auch in der Modellbildung nach [FS93] und [LHH<sup>+</sup>93] der Mittelwert der Krümmung des Potentials über die Tiefe der Raumladungszone mit dem Krümmungswert des Oberflächenpotentials gleichgesetzt und zur Berechnung des Oberflächenpotentials verwendet wird. Die Ansätze nach [CT91], [LHH<sup>+</sup>93] und [FS93] führen deshalb im Weiteren auf strukturell ähnliche Gleichungen.

Der Vorteil der Ansatzfunktion Gl. (3.20) liegt darin, dass die Krümmung des Potentialverlaufs **nicht** mehr von der Koordinate  $y$  abhängig ist und dass damit die Integrale einfach gelöst werden können. Es ergibt sich aus Gl. (3.19a) und Gl. (3.19b) folgendes Gleichungssystem:

$$\frac{d^2\psi_S(x)}{dx^2} + \frac{2}{(t_{dep} - y)^2} \cdot \psi(x, y) = \frac{qN_A}{\epsilon_{si}} \quad (3.21a)$$

$$\frac{d^2\psi_S(x)}{dx^2} + \frac{\epsilon_{ox}}{\epsilon_{si}} \cdot \frac{V_g - V_{fb} - \psi_S(x)}{t_{ox} t_{dep}} = \frac{qN_A}{\epsilon_{si}}. \quad (3.21b)$$

Um dieses Gleichungssystem zu lösen, berechnen wir das Oberflächenpotential, also das Potential für  $y = 0$ . Aus den erwähnten Gründen ist dieser Potentialverlauf auch für die weiterführende Schwellspannungsmodellierung maßgebend. Mit  $y = 0$  und  $\psi_S(x) =$

$\psi(x, 0)$  vereinfacht sich das Gleichungssystem zu:

$$\frac{d^2 \psi_S(x)}{dx^2} + \frac{2}{t_{dep}^2} \cdot \psi_S(x) = \frac{q N_A}{\epsilon_{si}} \quad (3.22a)$$

$$\frac{d^2 \psi_S(x)}{dx^2} + \frac{\epsilon_{ox}}{\epsilon_{si}} \cdot \frac{V_g - V_{fb} - \psi_S(x)}{t_{ox} t_{dep}} = \frac{q N_A}{\epsilon_{si}}. \quad (3.22b)$$

Ein Vergleich dieses Gleichungssystems mit den Herleitungen in [LHH<sup>+</sup>93] und [FS93] zeigt, dass in beiden Arbeiten ebenfalls zur Berechnung des Potentialverlaufs Gl. (3.22b) verwendet wird. Der Unterschied zu diesen beiden Vorgehensweisen liegt in der unterschiedlichen Modellierung der Tiefe der Raumladungszone  $t_{dep}(x)$ .

In [LHH<sup>+</sup>93] wird für den Parameter  $t_{dep}$  die Verarmungslänge von Langkanaltransistoren bei konstantem Oberflächenpotential von  $\psi_S = 2\psi_{FB} + V_{sb}$  verwendet. Damit wird  $t_{dep}$  als unabhängig von  $x$  modelliert. Da diese Näherung nur sehr ungenau ist, wird ein Fittingparameter  $\eta$  eingeführt.

In [FS93] wird die Verarmungslänge  $t_{dep}$  ebenfalls aus einer Langkanalbetrachtung abgeleitet. Allerdings wird die Abhängigkeit von  $\psi_S(x)$  und damit die Ortsveränderlichkeit von  $t_{dep}$  im Rahmen einer linearen Näherung berücksichtigt.

Im vorhergehenden Gleichungssystem kann die Verarmungslänge aus Gl. (3.22a) hergeleitet werden:

$$t_{dep} = \sqrt{\frac{2 \epsilon_{si} \psi_S}{q N_A}} \cdot \frac{1}{\sqrt{1 - \frac{\epsilon_{si}}{q N_A} \frac{d^2 \psi_S}{dx^2}}}, \quad (3.23)$$

$$= t_{dep,L} \cdot \frac{1}{\sqrt{1 - \frac{\epsilon_{si}}{q N_A} \frac{d^2 \psi_S}{dx^2}}}, \quad (3.24)$$

wobei  $t_{dep,L}$  die Verarmungslänge von Langkanaltransistoren darstellt. Eine Beschreibung des Oberflächenpotentials gemäß Gl. (3.22a) und Gl. (3.22b) modelliert also sowohl die Spannungsabhängigkeit von  $t_{dep}$ , als auch die Auswirkung von Kurzkanaleffekten auf die Verarmungslänge und stellt damit eine Erweiterung der Ansätze von [FS93] und [LHH<sup>+</sup>93] dar.

Durch Einsetzen von Gl. (3.24) in Gl. (3.22b) und durch anschließende Umstellung erhält man folgende Differentialgleichung:

$$\frac{\epsilon_{ox}}{t_{ox}} (V_g - V_{fb} - \psi_S) = q N_A \sqrt{1 - \frac{\epsilon_{si}}{q N_A} \frac{d^2 \psi_S}{dx^2}} \cdot t_{dep,L}(\psi_S). \quad (3.25)$$

[CT91] verzichtet auf die Lösung dieser Differentialgleichung und auf eine damit verbundene Diskussion der Folgerungen. Da hier jedoch eine weiterführende Analyse durchgeführt werden soll, wird Gl. (3.25) in dieser Arbeit näherungsweise gelöst. Dazu wird Gl. (3.25) um  $d^2\psi_s/dx^2 = 0$  und  $\psi_s = \psi_{s0}$  linearisiert. Man erhält:

$$\boxed{\frac{d^2\psi_s(x)}{dx^2} + \frac{2\epsilon_{ox}}{\epsilon_{si}} \cdot \frac{V_g - V_{fb} - n_0\psi_s(x)}{t_{ox} t_{dep,L}(\psi_{s0})} = \frac{qN_A}{\epsilon_{si}}}, \quad (3.26)$$

wobei  $n_0$  entsprechend der Definition in [Tsi99] die inverse Steigung der in Abb. 3.3 dargestellten Funktion  $\psi_s = f(V_g)$  im Punkt  $\psi_s = \psi_{s0}$  ist:

$$n_0 := \left( \frac{d\psi_s}{dV_g} \right)^{-1} \bigg|_{\psi_s=\psi_{s0}} = 1 + \frac{\gamma}{2\sqrt{\psi_{s0}}}, \quad (3.27)$$

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2\epsilon_{si}qN_A}. \quad (3.28)$$

Um die Plausibilität von Gl. (3.26) zu überprüfen, wenden wir diese Gleichung auf Langkanaltransistoren an, für die  $d^2\psi_s/dx^2 = 0$  gilt. Wir wählen  $\psi_{s0} = \psi_0 := 2\psi_{FB} + V_{SB}$ . Nach dem Umstellen erhalten wir:

$$\frac{\gamma}{2}\sqrt{\psi_0} + \left(1 + \frac{\gamma}{2\sqrt{\psi_0}}\right)\psi_s = V_g - V_{fb}. \quad (3.29)$$

Dieser Zusammenhang kann auch direkt aus der für Langkanaltransistoren bekannten Beziehung [Tsi99]:

$$V_g - V_{fb} = \psi_s + \gamma\sqrt{\psi_s} \quad (3.30)$$

durch Linearisierung um  $\psi_0$  erhalten werden. Damit ist also sichergestellt, dass Gl. (3.26) nicht der Langkanaltheorie widerspricht.

In [LHH<sup>+</sup>93] hat die Differentialgleichung zur Berechnung des Oberflächenpotentials die gleiche mathematische Struktur. Unterschiedlich zu Gl. (3.26) sind jedoch die Faktoren  $n_0$  und 2 im zweiten Summanden dieser Gleichung. Statt dessen wird in jener Analyse ein allgemeiner Fitting-Parameter  $\eta$  verwendet. Ursache dieses Unterschieds ist, wie schon analysiert, die Vernachlässigung der Ortsabhängigkeit und des Einflusses von Kurzkanaleffekten auf die Verarmungslänge in [LHH<sup>+</sup>93]. Ein Vergleich ergibt:

$$\eta \approx 2n_0. \quad (3.31)$$

Für Standard-MOS-Transistoren ist  $n_0 = 1,1 \dots 1,3$  und damit  $\eta \approx 2,2 \dots 2,6$ . Als Vergleich können die weit verbreiteten BSIM-Modelle [BSI], die auf [LHH<sup>+</sup>93] basieren, dienen. In diesen Modellen ist der entsprechende Parameter DVT1 standardmäßig äquivalent zu  $\eta = 3,55$  eingestellt. Die theoretische Vorhersage nach Gl. (3.26) stimmt damit gut mit dem empirisch gewonnenen Standardwert des BSIM-Parameters DVT1 überein.

### 3.2.2 Lösung der Differentialgleichung des Potentialverlaufs

Bei der Lösung der linearisierten Differentialgleichung Gl. (3.26) können aus Gl. (3.9a) und Gl. (3.9b) folgende Randbedingungen abgeleitet werden:

$$\psi_S(x=0) = V_s + V_{bi}, \quad (3.32a)$$

$$\psi_S(x=L) = V_d + V_{bi}. \quad (3.32b)$$

Es sei darauf hingewiesen, dass aus der Lösung der Differentialgleichung Gl. (3.26) unter den gegebenen Randbedingungen **nur** Lösungen des Gleichungssystems Gl. (3.21) für  $y=0$  folgen. Überträgt man trotzdem die daraus gefundene Lösung für  $\psi_S$  formal in Gl. (3.21a), ergibt sich zum Beispiel entlang des Grenzsegments von Source zum Bulk, also für  $x=0$  und  $0 < y < t_j$ , wobei  $t_j$  die Tiefe des Source-Gebietes bezeichnet:  $\psi(0,y) \neq \psi_S(0)$ . Damit ist die Randbedingung  $\psi(0,y) = V_s + V_{bi}$  für  $0 < y < t_j$  in Gl. (3.21a) **nicht** erfüllt. Ein formales Einsetzen der Lösung von Gl. (3.26) in Gleichung Gl. (3.21a) führt deshalb für  $y > 0$  und  $x \approx 0$  zu unphysikalischen Ergebnissen. Dies ist eine Konsequenz aus dem Ansatz Gl. (3.20) und dem quasi-zweidimensionalen Verfahren zur Lösung einer partiellen Differentialgleichung. Da wir hier jedoch nur an dem Oberflächenpotential interessiert sind, ist diese Diskrepanz unerheblich.

Die Lösung der Differentialgleichung unter den gegebenen Randbedingungen lautet:

$$\psi_S(x) = \psi_1 \exp(x/l_c) + \psi_2 \exp(-x/l_c) + \psi'_{S,L}. \quad (3.33)$$

Man kann aus dieser Gleichung ableiten, dass für  $L \gg l_c$  und  $x \gg l_c$  das Oberflächenpotential  $\psi_S(x) \rightarrow \psi'_{S,L}$  geht.  $\psi'_{S,L}$  entspricht damit dem (linearisierten) Langkanal-Oberflächenpotential:

$$\psi'_{S,L} = \frac{V_g - V_s - V'_{th,L} + n_0 \psi_0}{n_0}, \quad (3.34)$$

wobei  $V'_{th,L}$  der extrapolierte Wert der Langkanalschwellspannung ist. Dieser wird gewonnen, indem die Funktion  $\psi_S = f(V_g)$  der Langkanaltheorie Gl. (3.30) am Punkt  $\psi_{S0}$  linearisiert wird und mit dieser Linearisierung der Gatespannungswert, für den  $\psi_S = \psi_0$  gilt, berechnet wird:

$$\begin{aligned} V'_{th,L} &= V_{fb} + \psi_0 n_0 + \frac{\gamma}{2} \sqrt{\psi_{S0}} - V_s, \\ &= V_{fb} + \psi_0 - V_s + \frac{\gamma}{2} \left( \frac{\psi_0}{\psi_{S0}} + 1 \right) \sqrt{\psi_{S0}}. \end{aligned} \quad (3.35)$$

Wählt man  $\psi_{S0} = \psi_0$ , ist  $V'_{th,L} = V_{th,L}$ .

Auch der Ausdruck Gl. (3.34) wurde durch Linearisierung um  $\psi_{S0} = \psi_0$  gewonnen. Um die Genauigkeit des Potentialmodells für den Fall  $\psi_S \ll \psi_0$  zu erhöhen, substituieren wir  $\psi'_{S,L}$  formal mit dem Ausdruck für das exakte Langkanalpotential:  $\psi'_{S,L} \rightarrow \psi_{S,L}$ .



Damit ergibt sich nun endgültig:

$$\begin{aligned}
 \psi_S(x) &= \psi_1 \exp(x/l_c) + \psi_2 \exp(-x/l_c) + \psi_{S,L}, \\
 \psi_1 &= \frac{V_d + V_{bi} - (V_s + V_{bi} - \psi_{S,L}) \exp(-L/l_c) - \psi_{S,L}}{2 \sinh(L/l_c)}, \\
 \psi_2 &= V_s + V_{bi} - \psi_1 - \psi_{S,L}, \\
 \psi_{S,L} &= \left( -\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_g - V_{fb}} \right)^2 \approx \frac{V_g - V_s - V'_{th,L} + n_0 \psi_0}{n_0}, \\
 l_c &= \sqrt{\frac{\epsilon_{si} t_{ox} t_{dep,L}(\psi_{S0})}{2 n_0 \epsilon_{ox}}}.
 \end{aligned} \tag{3.36}$$

### 3.2.3 Diskussion des Potentialverlaufs

In Abb. 3.6 und Abb. 3.7 wird die Berechnung des Oberflächenpotentials nach Gl. (3.36) einer Simulation mit dem Programm MINIMOS-NT [Fis94] für verschiedene Transistorlängen und Arbeitspunkte gegenübergestellt. Zum Vergleich wird die Berechnung nach [LHH<sup>+</sup>93] ohne Fitting, das heißt mit  $\eta = 1$ , gezeigt. Man erkennt, dass das obige Modell vor allem bei kurzen Kanallängen besser mit den Vorhersagen der Halbleitersimulation übereinstimmt als das Modell nach [LHH<sup>+</sup>93]. Das Modell ist damit für eine Verwendung bei der Berechnung der Schwellspannung von Kurzkanaltransistoren geeignet.

Für die folgenden Analysen müssen aus den vorhergehenden Ergebnissen das Potentialminimum und der Anstieg des Potentialminimums in Abhängigkeit von der Schwellspannung berechnet werden. Mathematische Standardverfahren ergeben dafür unter Bezugnahme auf Gl. (3.36) für die x-Koordinate des Minimalwerts:

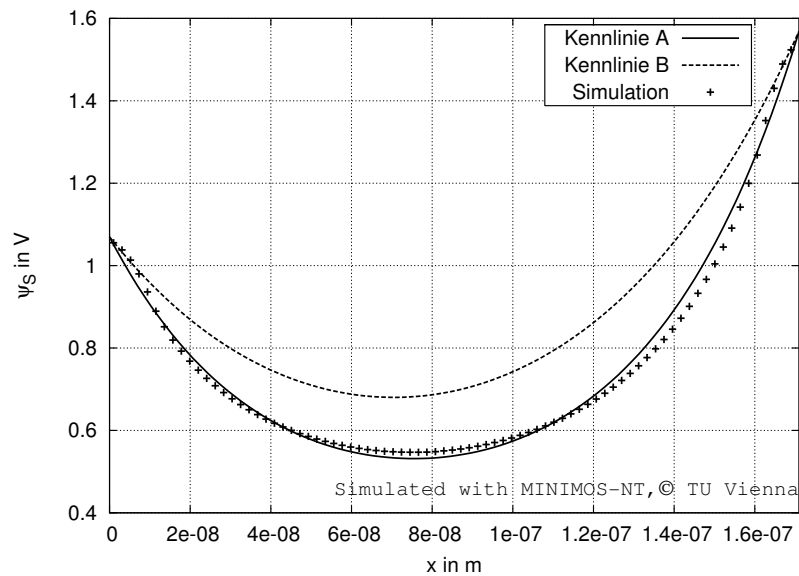
$$x_{min} = \frac{1}{2} l_c \ln \left( \frac{\psi_2}{\psi_1} \right). \tag{3.37}$$

Eingesetzt in Gl. (3.36) ergibt sich als zugehöriger Minimalwert:

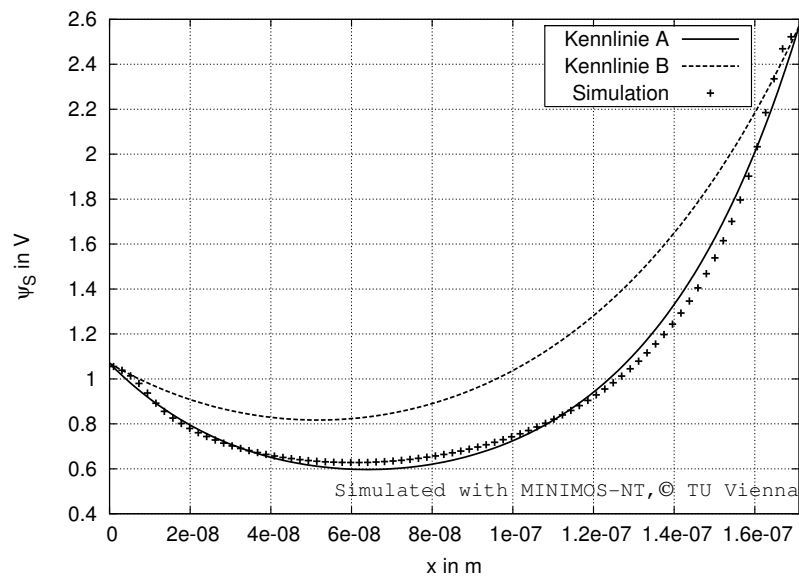
$$\begin{aligned}
 \psi_{S,min} &= 2 \underbrace{\sqrt{\psi_1 \psi_2}}_{=: \Delta \psi_S} + \psi_{S,L}.
 \end{aligned} \tag{3.38}$$

Für einige der nachfolgenden Herleitungen ist es zweckmäßig, folgende Näherung zu verwenden, die auf der Annahme  $V_{ds} \ll V_{bi} - \psi_{S,L} \implies x_{min} \rightarrow L/2$  beruht:

$$\Delta \psi_S \approx \frac{2(V_{bi} - \psi_{S,L}) + V_d + V_s}{2 \cosh(L/(2l_c))}. \tag{3.39}$$

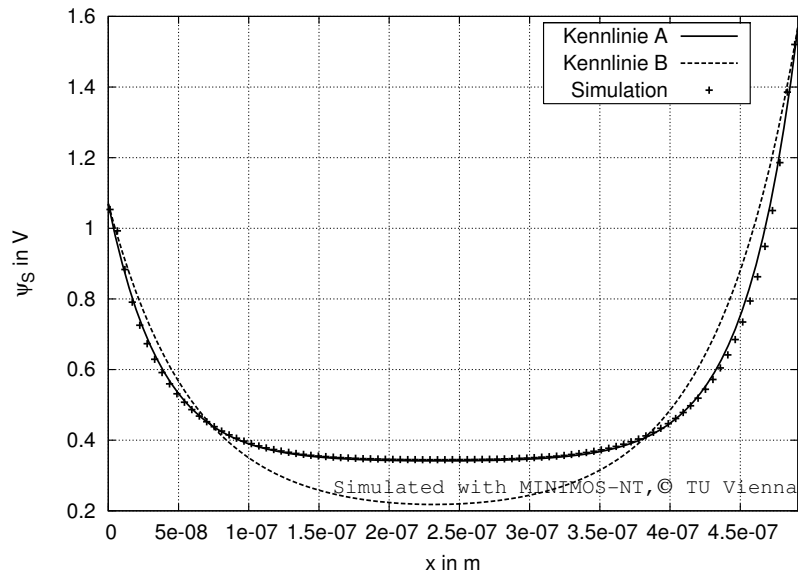


(a)  $L = 0,18 \mu\text{m}$ ;  $V_{gb} = 0,2 \text{ V}$ ,  $V_{ds} = 0,5 \text{ V}$

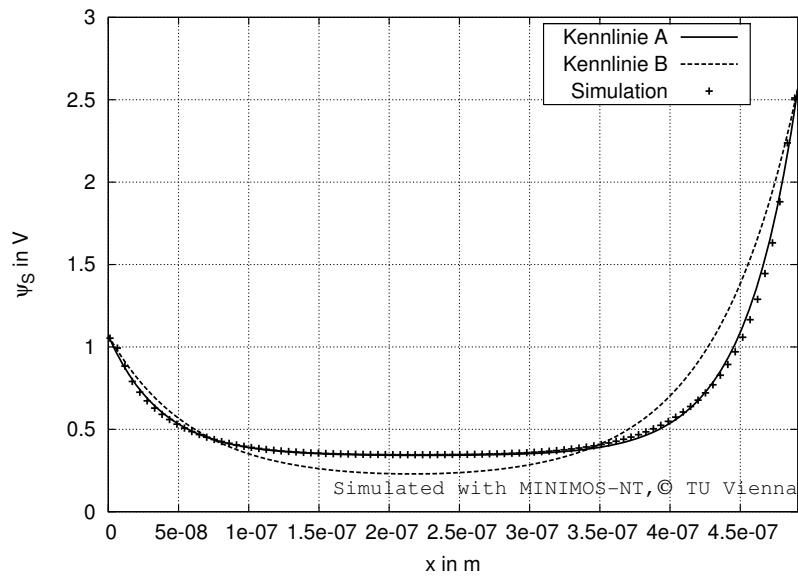


(b)  $L = 0,18 \mu\text{m}$ ;  $V_{gb} = 0,2 \text{ V}$ ,  $V_{ds} = 1,5 \text{ V}$

**Abbildung 3.6:** Numerische Simulation (MINIMOS-NT) und Vergleich der Berechnung des Oberflächenpotentials gemäß Gl. (3.36),  $\psi_{S0} = 2\psi_{FB}$ , (Kennlinie A) mit einer Berechnung nach [LHH<sup>+</sup>93],  $\eta = 1$  (Kennlinie B) für einen Transistor mit  $L = 0,18 \mu\text{m}$

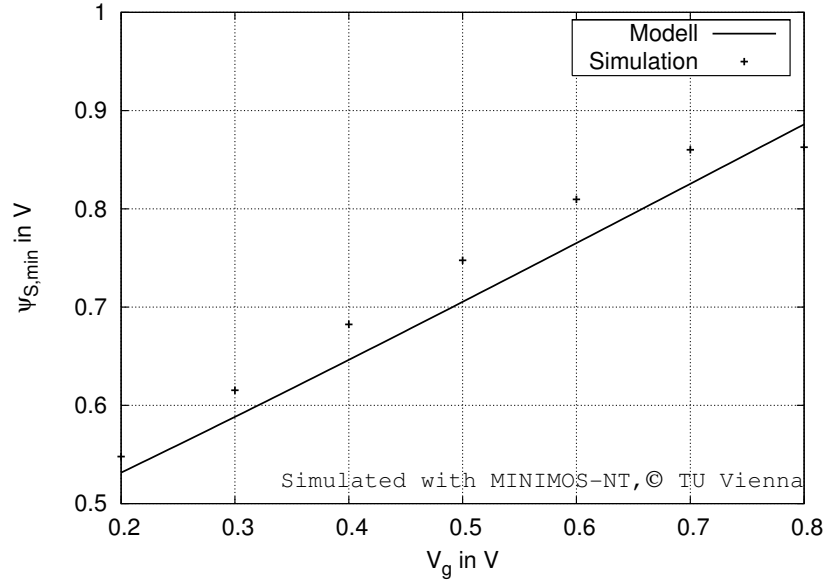


(a)  $L = 0.5 \mu\text{m}$ ;  $V_{gb} = 0.2 \text{ V}$ ,  $V_{ds} = 0.5 \text{ V}$



(b)  $L = 0.5 \mu\text{m}$ ;  $V_{gb} = 0.2 \text{ V}$ ,  $V_{ds} = 1.5 \text{ V}$

**Abbildung 3.7:** Numerische Simulation (MINIMOS-NT) und Vergleich der Berechnung des Oberflächenpotentials gemäß Gl. (3.36),  $\psi_{S0} = 2\psi_{FB}$ , (Kennlinie A) mit einer Berechnung nach [LHH<sup>+</sup>93],  $\eta = 1$  (Kennlinie B) für einen Transistor mit  $L = 0.5 \mu\text{m}$



**Abbildung 3.8:** Minimalwert des Kanalpotentials für einen Transistor mit  $L = 0,18\mu\text{m}$  und Vergleich mit numerischer Simulation

Ein Vergleich der Berechnung Gl. (3.38) mit den Ergebnissen einer numerischen Simulation in Abb. 3.8 zeigt einen geringen Offset der Kennlinie und einen geringen Unterschied in der Steigung der Kennlinie.

Die Abhängigkeit des Anstiegs des minimalen Kanalpotentials vom Anstieg des Gatepotentials ist ein Maß für die Effektivität der Gatesteuerung und kann mit folgendem Zusammenhang ermittelt werden, bei dessen Herleitung die Näherung Gl. (3.39) verwendet wurde:

$$\frac{d\psi_{S,min}}{dV_g} = \frac{d(\psi_{S,L} + \Delta\psi_S)}{dV_g} \approx \left(1 - \frac{1}{\cosh(L/(2l_c))}\right) \frac{d\psi_{S,L}}{dV_g}, \quad (3.40)$$

$$\frac{d\psi_{S,L}}{dV_g} = 1 - \frac{\gamma}{2\sqrt{\frac{\gamma^2}{4} + V_g - V_{fb}}} \approx \frac{1}{n_0}. \quad (3.41)$$

Aus dieser Gleichung ist ersichtlich, dass der Anstieg des Minimalwerts des Oberflächenpotentials  $d\psi_{S,min}/dV_g$  in Kurzkanaltransistoren geringer als in Langkanaltransistoren ( $d\psi_{S,min}/dV_g \rightarrow d\psi_{S,L}/dV_g$ ) ist. Folglich besitzt das Gatepotential in Kurzkanaltransistoren eine geringere Kanalsteuerfähigkeit. Konsequenzen aus diesem Ergebnis werden bei der Auswertung des Subthreshold-Verhaltens von Kurzkanaltransistoren in Abschnitt 3.4.2 aufgezeigt.

### 3.3 Schwellspannungsmodell

Aus den vorhergehenden Gleichungen zur Berechnung des Oberflächenpotentials kann ein Schwellspannungsmodell abgeleitet werden. Gemäß der Definition Gl. (3.6) ist die Schwellspannung diejenige Gate-Source-Spannung, für die  $\psi_{S,min} = \psi_0$  gilt. Setzen wir diese Definition in Gl. (3.38) ein und stellen dann nach  $V_g$  um, so ergibt sich unter Verwendung der Linearisierung Gl. (3.34) [Iñ95]:

$$\begin{aligned}
 V_{th} &= V_{th,L} - n_0 \psi_0 + \frac{-n_0 b + n_0 \sqrt{b^2 - 4ac}}{2a} \\
 a &= 2 (\cosh(L/l_c) - 1) - \sinh^2(L/l_c), \\
 b &= 2 (-\cosh(L/l_c) + 1) (2V_{bi} + V_d + V_s) + 2\psi_0 \sinh^2(L/l_c), \\
 c &= 2 (V_{bi} + V_s) (V_{bi} + V_d) \cosh(L/l_c) - (V_{bi} + V_d)^2 - (V_{bi} + V_s)^2 - \\
 &\quad - \psi_0^2 \sinh^2(L/l_c),
 \end{aligned} \tag{3.42}$$

Als Grenzwert der Schwellspannung für  $(L/l_c) \rightarrow \infty$  ergibt sich wie erwartet für einen langen Transistor:

$$\lim_{(L/l_c) \rightarrow \infty} V_{th} = V_{th,L}. \tag{3.43}$$

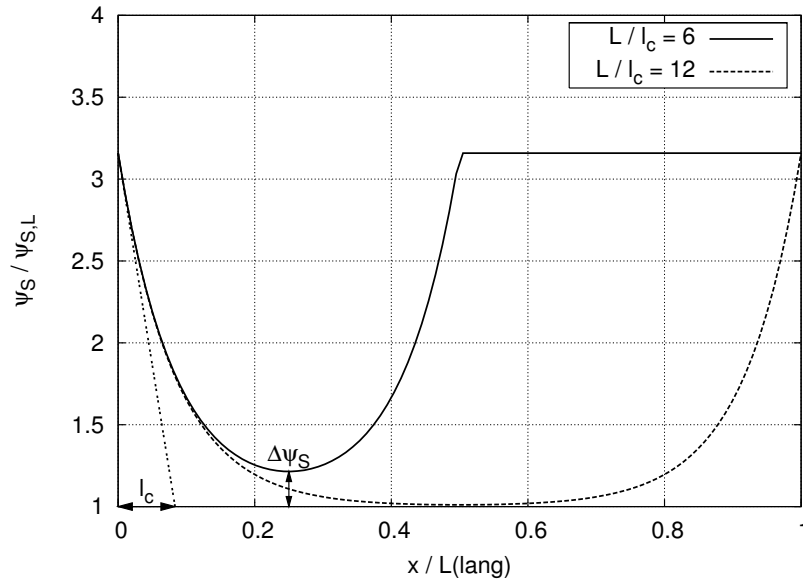
Aus Gl. (3.42) ist ersichtlich, dass die Schwellspannung  $V_{th}$  im allgemeinen Fall nicht der Langkanalschwellspannung entspricht, sondern sowohl von der Drain-Source-Spannung, als auch vom Verhältnis der effektiven Transistorlänge zur charakteristischen Länge  $L/l_c$  beeinflusst wird. Diese Effekte werden im Folgenden näher untersucht.

#### 3.3.1 Kurzkanaleffekt

In diesem Abschnitt wird die Abhängigkeit der Schwellspannung von der effektiven Transistorlänge untersucht. Zunächst sollen qualitative Aussagen gewonnen werden, die das Verständnis der physikalischen Hintergründe vereinfachen.

Wie aus Gl. (3.36) gefolgert werden kann, ist das Oberflächenpotential und damit auch die Schwellspannung vom Verhältnis der effektiven Kanallänge zur charakteristischen Länge  $L/l_c$  abhängig. Langkanaltransistoren sind demnach Transistoren, für die  $L \gg l_c$  gilt. Ist die Kanallänge in der gleichen Größenordnung wie die charakteristische Länge, handelt es sich um Kurzkanaltransistoren. Transistoren mit  $L < l_c$  sind nicht praktisch einsetzbar.

In Abb. 3.9 sind mit Hilfe des Modells Gl. (3.36) typische Verläufe des Oberflächenpotentials für Kurz- bzw. Langkanaltransistoren dargestellt. Um die wesentlichen physikalischen Vorgänge, die zu Kurzkanaleffekten führen, hervorzuheben, wählen wir  $V_d = V_s$ . Beide Transistoren haben, bis auf die Kanallänge, den gleichen Aufbau und die gleiche Dotierung.



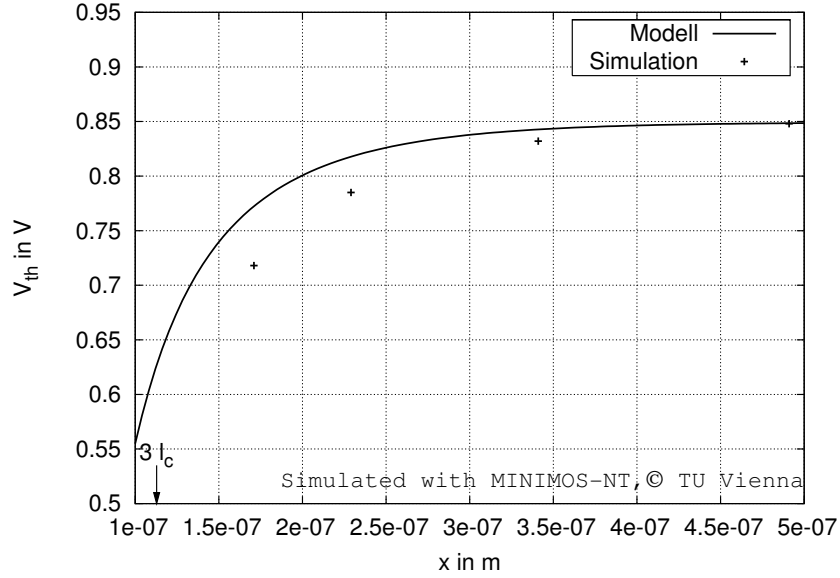
**Abbildung 3.9:** Oberflächenpotential, normiert auf das Langkanalpotential, für verschiedene Transistorlängen

Zunächst diskutieren wir den Potentialverlauf für Langkanaltransistoren. Man erkennt in Abb. 3.9, dass sich das Oberflächenpotential an den Kanalenden in Übereinstimmung mit den definierten Randbedingungen auf Source- bzw. Drain-Potential befindet. Vom Source-Potential bei  $x = 0$  fällt die Spannung dann nach einer Exponentialfunktion ab. Die Abfallrate ist abhängig von der charakteristischen Länge  $l_c$ . Das Oberflächenpotential erreicht dann bei  $x \gg l_c$  für den Langkanalfall  $L \gg l_c$  den stationären Wert  $\psi_{S,L}$ . Am drainseitigen Kanalende  $x = L$  steigt  $\psi_S$  auf den Wert des Drain-Potentials an. Analog zum sourceseitigen Kanalende geschieht dieser Anstieg nach einer Exponentialfunktion mit der Anstiegsrate  $l_c$ .

Verkürzen wir nun die Kanallänge  $L$  bei konstanter charakteristischer Länge, überlagern sich bei Transistorlängen von  $L \rightarrow l_c$  die source- und drainseitigen Exponentialfunktionen. Infolgedessen erreicht das Oberflächenpotential nicht mehr den stationären Langkanalwert  $\psi_{S,L}$ , sondern es bildet sich ein Minimum aus, dessen Wert höher ist, als  $\psi_{S,L}$ . Damit wird bei steigender Gatespannung der Einsetzpunkt zur Inversion  $\psi_0$  früher erreicht als bei Transistoren mit langem Kanal. Folglich ist die Schwellspannung des Kurzkanaltransistors gesunken.

Diese Verringerung der Schwellspannung für kürzer werdende Transistoren wird im Allgemeinen als *Kurzkanaleffekt* bezeichnet.

In Abb. 3.10 wurde mit Hilfe des Modells Gl. (3.42) die Schwellspannung in Abhängigkeit von der Kanallänge berechnet. Man erkennt, dass sich die Schwellspannung  $V_{th}$  für lange Transistoren der Langkanalschwellspannung  $V_{th,L}$  annähert. Ungenauigkeiten in der Modellierung des Potentialverlaufs im Besonderen in der Umgebung der Kontakte,



**Abbildung 3.10:** Abhängigkeit der Schwellspannung  $V_{th}$  von der Transistorlänge für  $V_{ds} = 0$  V

z.B. durch die Anwendung der Ansatzfunktion Gl. (3.20) und durch die Linearisierung der Differentialgleichung Gl. (3.25), führen im Beispiel zu einer leichten Unterschätzung der charakteristischen Länge  $l_c$ . Deshalb ist der Abfall der berechneten Schwellspannungskennlinie im Vergleich zur Simulation zu kleineren Kanallängen verschoben.

Wird für den Grenzfall  $V_{ds} \rightarrow 0$  die Näherungsformel Gl. (3.39) angewendet, kann man nach [LHH<sup>+</sup>93] folgende Näherung ableiten:

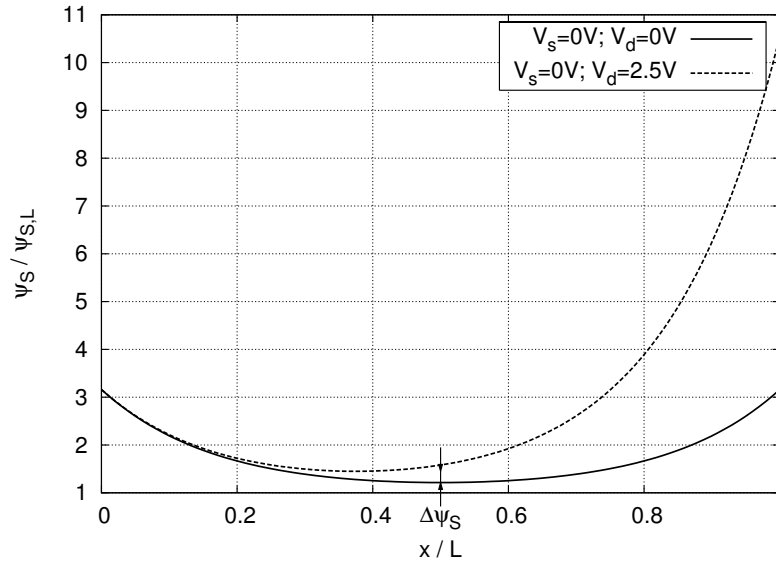
$$V_{th} \approx V_{th,L} - \frac{n_0}{2} \frac{2(V_{bi} - 2\Psi_{FB}) + V_{ds}}{\cosh(L/(2l_c)) - 1}. \quad (3.44)$$

### 3.3.2 Drain Induced Barrier Lowering

Aus Gl. (3.42) folgt neben der schon untersuchten Längenabhängigkeit auch eine Abhängigkeit der Schwellspannung von der Drain-Source-Spannung. Diese wird im Weiteren analysiert.

Für eine qualitative Betrachtung der physikalischen Effekte wird in Abb. 3.11 das Oberflächenpotential für einen kurzen Transistor bei  $V_d = V_s = 0$  V gezeigt. Aufgrund dieser Randbedingung ist das sourceseitige Oberflächenpotential  $\psi_s(x=0)$  gleich dem drainseitigem Potential  $\psi_s(L)$ . Der Potentialverlauf ist symmetrisch zu  $x = L/2$ .

Erhöhen wir nun z.B. nur das Drain-Potential, heben wir damit auch das drainseitige Oberflächenpotential an, während die Source-Seite auf unverändertem Potentialniveau liegt. Der Potentialverlauf „kippt“ also, dass heißt der Verlauf wird unsymmetrisch und das Potentialminimum verschiebt sich in Richtung Source-Kontakt. Wie in Abb. 3.11



**Abbildung 3.11:** Verlauf des Oberflächenpotentials für verschiedene Drain-Source-Spannungen für einen kurzen Transistor mit  $L/l_c = 6$

weiterhin zu erkennen ist, erhöht sich außerdem auch der Minimalwert des Oberflächenpotentials um  $\Delta\psi_S$ . Wie schon im vorherigen Abschnitt dargestellt, ist das gleichbedeutend mit einer Verringerung der Schwellspannung.

Der Effekt der Reduktion der Schwellspannung mit größer werdender Drain-Source-Spannung wird als *Drain Induced Barrier Lowering (DIBL)* bezeichnet.

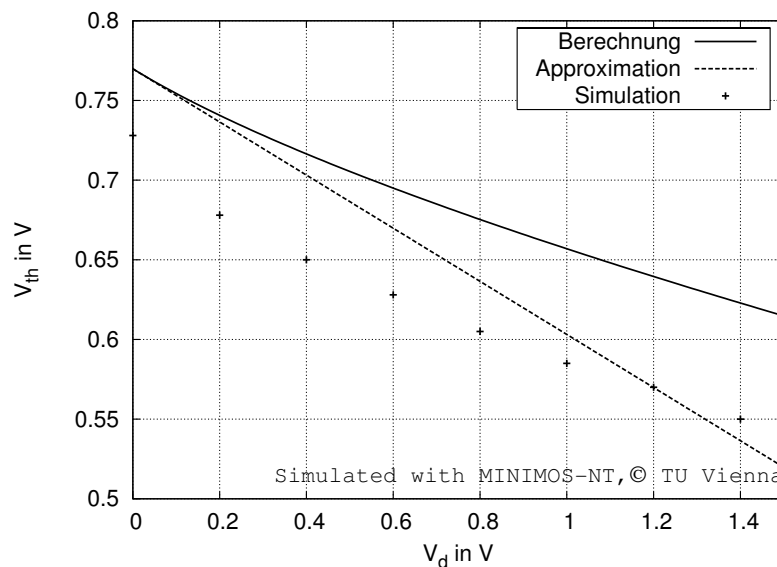
In Abb. 3.12 wurde mit Hilfe des Modells Gl. (3.42) die Schwellspannung in Abhängigkeit von der Drain-Source-Spannung berechnet. In derselben Abbildung wird zum Vergleich auch eine Berechnung mit der Näherungsformel Gl. (3.44) gezeigt. Man erkennt, dass diese Approximation nur für kleine Drain-Source-Spannungen  $V_{ds}$  zufriedenstellende Ergebnisse liefert. Für größere Drain-Source-Spannungen sollte Gl. (3.42) verwendet werden [Iñf95]. Die Verschiebung der berechneten Kennlinie im Vergleich zum Simulationsergebnis ist zum großen Teil auf die leichte Unterschätzung der charakteristischen Länge  $l_c$  zurückzuführen, die schon zu Abb. 3.10 diskutiert wurde.

## 3.4 Subthreshold-Verhalten

Als Subthreshold-Betriebsbereich von Transistoren wird der Bereich der Transistorkennlinie verstanden, für den  $V_{gs} < V_{th}$  gilt.

Subthreshold-Leckströme können neben funktionellen Problemen in dynamischen Logikschaltungen und in Speicher- und Bildsensoranwendungen auch eine erhöhte Ruhestromaufnahme von hochintegrierten Schaltkreisen verursachen. Die sogenannten Standby-Ströme von abgeschalteten Baugruppen sind auch wesentlich im Gesamtleistungs-





**Abbildung 3.12:** Abhängigkeit der Schwellspannung  $V_{th}$  von der Drain-Source-Spannung für einen Transistor mit  $L = 0,18 \mu\text{m}$

budget von Systemen mit mittlerer Verlustleistung (z.B. Work Stations) und limitierend für Systeme mit niedriger Verlustleistung (z.B. Mobilanwendungen) [FDN<sup>+</sup>01]. Subthreshold-Leckströme setzen daher im Allgemeinen eine untere Grenze für die Verringerung der Schwellspannung bei der Weiterentwicklung der Technologien und bestimmen häufig auch die erlaubten Variationen der Schwellspannung und damit das Prozessfenster.

Die genaue Modellierung des Transistorverhaltens im Subthreshold-Bereich ist damit besonders für die Beurteilung des Sperrvermögens kurzkanaliger Transistoren von Bedeutung. Aus diesem Grund wird in diesem Abschnitt die Subthreshold-Kennlinie abgeleitet und die Subthreshold-Steigung, ein Maß für den Übergang von Transistoren in den Sperrbereich, analysiert.

Da der Subthreshold-Bereich, außer im Bereich der moderaten Inversion bei  $V_{gs} \approx V_{th}$ , von  $n, p \ll N_A$  gekennzeichnet ist [Tsi99], können die Ergebnisse von Abschnitt 3.3 zur weiteren Analyse herangezogen werden.

### 3.4.1 Subthreshold-Transferkennlinie

Im Weiteren werden wir voraussetzen, dass die Subthreshold-Ströme durch Drift-Diffusionsströme erzeugt werden, die durch Gl. (3.45) beschrieben werden können.<sup>2</sup> Die

<sup>2</sup>Thermische Emissionsprozesse, die dem Richardson-Gesetz folgen, werden zunächst vernachlässigt. Diese Transportprozesse werden aber im Rahmen des Subthreshold-Modells der ONO-Speichertransistoren in Abschnitt 4.4.3 diskutiert.

weitere Herleitung entspricht dabei im Wesentlichen dem Vorgehen in [FS93].

Grundlage zur Berechnung der Drift-Diffusionsströme im p-Transistor ist die Transportgleichung der Elektronen:

$$\vec{j}_n = q\mu_n \left( n\vec{E} + U_T \text{grad} n \right). \quad (3.45)$$

Da wir für die weitere Berechnung nur die x-Komponente des Stromdichtevektors benötigen, können wir diese Gleichung wie folgt vereinfachen:

$$j_{n,x} = q\mu_n U_T \left( -\frac{n}{U_T} \frac{\partial \psi}{\partial x} + \frac{\partial n}{\partial x} \right). \quad (3.46)$$

Multiplizieren wir diese Gleichung mit dem Integrierenden Faktor  $\exp(-\psi/U_T)$ , können wir die rechte Seite als Totales Differential einer Funktion  $f$  schreiben und es gilt:

$$f(x) = n(x) \exp\left(-\frac{\psi(x)}{U_T}\right) + C, \quad (3.47)$$

$$j_{n,x} \exp\left(-\frac{\psi}{U_T}\right) = q\mu_n U_T \frac{df(x)}{dx}. \quad (3.48)$$

Für die Beweglichkeit wird im Rahmen unserer Untersuchung ein von der x-Koordinate unabhängiger Mittelwert  $\mu_n \rightarrow \bar{\mu}_n$  eingesetzt. Trennung der Variablen und anschließende Integration führen auf:

$$\int_{x=0}^L j_{n,x} \exp\left(-\frac{\psi}{U_T}\right) dx = q\bar{\mu}_n U_T (f(L) - f(0)). \quad (3.49)$$

Setzt man voraus, dass  $j_{n,x}$  unabhängig von  $x$  ist und unter Verwendung der Randbedingungen  $n(0) = n(L) = N_D$  und Gl. (3.9a)/ Gl. (3.9b) kann man schreiben:

$$j_{n,x} = q\bar{\mu}_n N_D U_T \frac{\exp\left(-\frac{V_d + V_{bi}}{U_T}\right) - \exp\left(-\frac{V_s + V_{bi}}{U_T}\right)}{\int_{x=0}^L \exp\left(-\frac{\psi}{U_T}\right) dx}. \quad (3.50)$$

Zur Lösung des Integrals im Nenner wird  $\psi(x, y)$  an den Punkten  $x = x_{min}$  und  $y = 0$  in eine Taylor-Serie entwickelt. Für die Approximation des Funktionsverlaufs über die x-Koordinate wird das quadratische Glied berücksichtigt, andernfalls wird die Taylorreihe

nach dem linearen Glied abgebrochen:

$$\begin{aligned} \psi(x, y) \approx & \underbrace{\psi(x_{min}, 0)}_{\psi_{S,min}} + \underbrace{\frac{\partial \psi}{\partial x} \Big|_{x_{min}, y=0}}_0 (x - x_{min}) + \frac{\partial \psi}{\partial y} \Big|_{x_{min}, y=0} y + \\ & + \underbrace{\frac{\partial^2 \psi}{\partial x^2} \Big|_{x_{min}, y=0}}_{d^2 \psi_S / dx^2} (x - x_{min})^2. \end{aligned} \quad (3.51)$$

Um von der Stromdichte  $j_{n,x}$  auf den gesuchten Drain-Source-Strom übergehen zu können, müssen beide Seiten von Gl. (3.50) über die Tiefe der Raumladungszone integriert werden. Dabei muss beachtet werden, dass aufgrund des gewählten Koordinatensystems die positive Richtung des Stromdichtevektors  $\vec{j}$  entgegengesetzt zum Strom  $I_{ds}$  gerichtet ist. Damit ergibt sich unter Verwendung der vorhergehenden Taylorentwicklung und mit  $V_{bi} = U_T \ln(N_A N_D / n_i^2)$ ,  $N_A = n_i \exp(\Psi_{FB} / U_T)$ ,  $\psi_{S,min} = \psi_{S,L} + \Delta\psi_S$ :

$$I_{ds} = \frac{q \bar{\mu}_n N_A W U_T \delta}{L v} \exp\left(\frac{\psi_{S,L} + \Delta\psi_S - 2\psi_{FB} - V_s}{U_T}\right) \left[1 - \exp\left(-\frac{V_{ds}}{U_T}\right)\right]. \quad (3.52)$$

Dabei entspricht  $\delta$  der effektiven Kanaltiefe in [FS93]. Es gilt entsprechend unserer Herleitung:

$$\begin{aligned} \delta &= \int_0^{t_{dep}} \exp\left(\frac{1}{U_T} \frac{\partial \psi}{\partial y} \Big|_{x_{min}, y=0} y\right) dx, \\ &= \frac{U_T}{\partial \psi / \partial y} \left[ \exp\left(\frac{1}{U_T} \frac{\partial \psi}{\partial y} t_{dep}\right) - 1 \right]. \end{aligned} \quad (3.53)$$

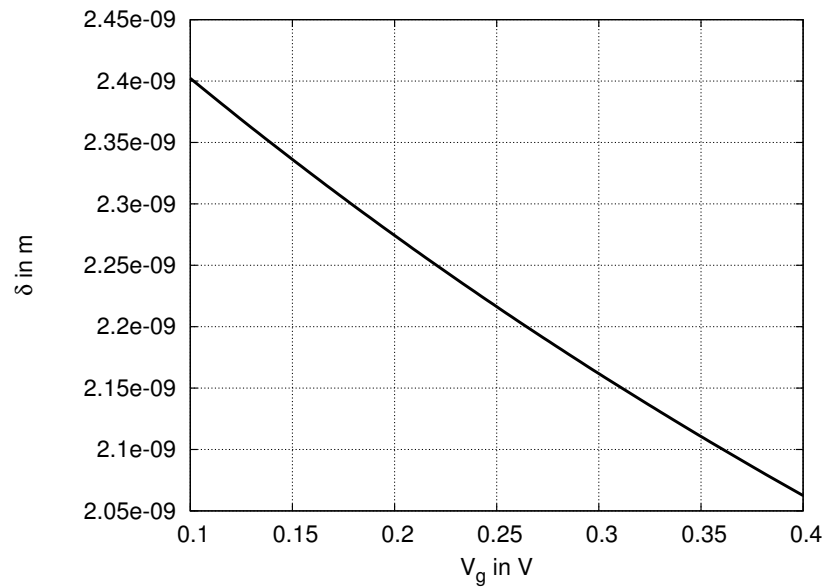
Zur Bestimmung des Ausdrucks  $\partial \psi / \partial y$  kann Gl. (3.21a) zusammen mit Gl. (3.24) verwendet werden:

$$\frac{\partial \psi}{\partial y} \Big|_{x_{min}, y=0} = -\frac{2\psi_{S,min}}{t_{dep}} = -\frac{q N_A}{\epsilon_{si}} t_{dep,L} \sqrt{1 - \frac{\epsilon_{si}}{q N_A} \frac{d^2 \psi_S}{dx^2}} \quad (3.54)$$

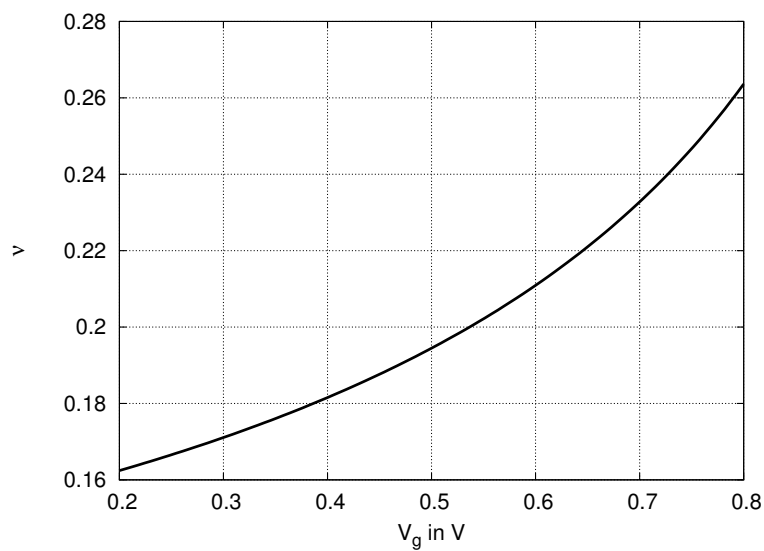
$$\approx -\frac{q N_A}{\epsilon_{si}} t_{dep,L}. \quad (3.55)$$

Damit ergibt sich schließlich die effektive Kanaltiefe zu:

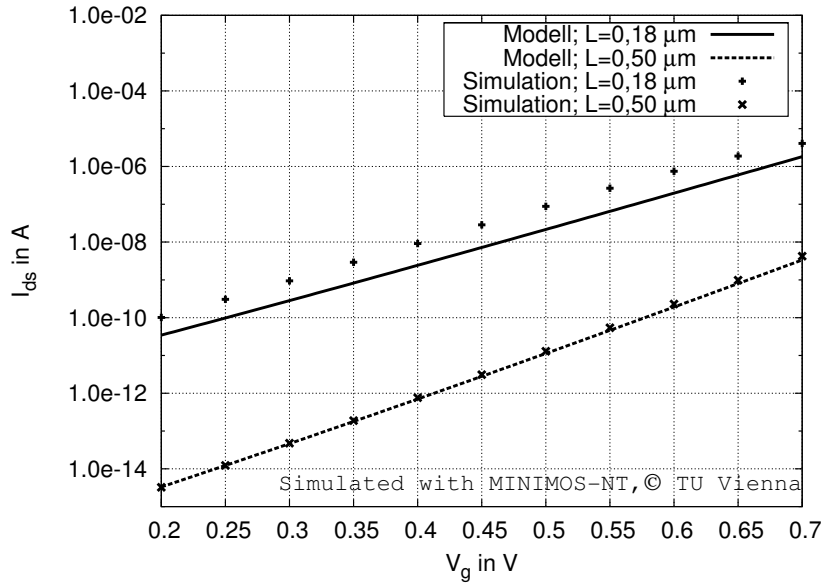
$$\delta \approx \frac{\epsilon_{si} U_T}{q N_A t_{dep,L}} \underbrace{\left[1 - \exp\left(-\frac{2\psi_{S,min}}{U_T}\right)\right]}_{\approx 1}. \quad (3.56)$$



**Abbildung 3.13:** Effektive Kanaltiefe  $\delta$  in Abhängigkeit vom Gatepotential  $V_g$



**Abbildung 3.14:** Funktion  $v$  in Abhängigkeit vom Gatepotential  $V_g$



**Abbildung 3.15:** Drainstrom für Transistoren unterschiedlicher Länge in Abhängigkeit vom Gatepotential

Die Exponentialfunktion kann dabei vernachlässigt werden, sobald das minimale Kanalpotential ein Vielfaches der Temperaturspannung beträgt. Für die schwache Inversion mit  $\Psi_{FB} < \psi_S < 2\Psi_{FB}$  ist dies in praktischen Fällen immer erfüllt.

Die Funktion  $v$  beschreibt den Einfluss der Kurzkanaleffekte:

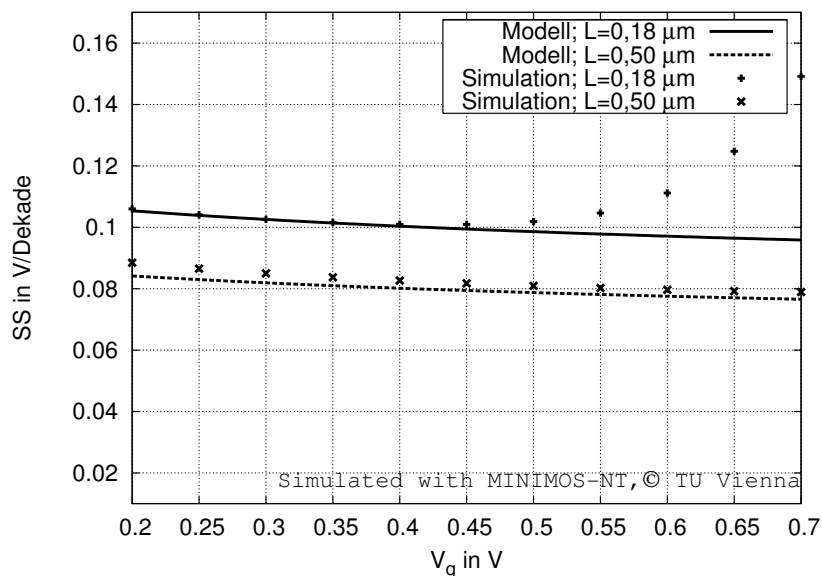
$$v = \frac{1}{L} \int_0^L \exp\left(-\frac{1}{U_T} \frac{d^2 \psi_S}{dx^2} (x - x_{min})^2\right) dx. \quad (3.57)$$

$$v = \sqrt{\frac{U_T}{\Delta \psi_S}} \frac{l_c}{L} \frac{\sqrt{\pi}}{2} \underbrace{\left[ \operatorname{erf}\left(\sqrt{\frac{\Delta \psi_S}{U_T}} \frac{L - x_{min}}{l_c}\right) - \operatorname{erf}\left(-\sqrt{\frac{\Delta \psi_S}{U_T}} \frac{x_{min}}{l_c}\right) \right]}_{\approx 2}. \quad (3.58)$$

Für lange Transistoren mit  $d^2 \psi_S / dx^2 = 0$  gilt:  $v \rightarrow 1$ .

In der Abb. 3.13 und Abb. 3.14 wird der Verlauf der Funktionen  $\delta$  bzw.  $v$  für Transistoren unterschiedlicher Kanallänge gezeigt. Man erkennt, dass beide Kenngrößen nur geringfügig mit dem Gatepotential variieren. Kombiniert man schließlich die hergeleiteten Gleichungen, erhält man die Kennliniengleichung eines MOS-Transistors im Subthreshold-Arbeitsbereich:

$$I_{ds} = \frac{\bar{\mu}_n W \varepsilon_{si} U_T^2}{L t_{dep,L} v} \exp\left(\frac{V_{gs} - V'_{th,L}}{n_0 U_T} + \frac{\Delta \psi_S}{U_T}\right) \left[1 - \exp\left(-\frac{V_{ds}}{U_T}\right)\right]. \quad (3.59)$$



**Abbildung 3.16:** Subthreshold-Steigung in Abhängigkeit vom Gatepotential für Transistoren unterschiedlicher Länge

In Abb. 3.15 wird ein Vergleich der Berechnung des Drainstroms nach Gl. (3.52) mit numerischen Simulationen gezeigt. Man erkennt eine gute Übereinstimmung der beiden Ergebnisse. Die berechnete Kurzkanalkennlinie ist um ungefähr 50 mV zu höheren Spannungen verschoben. Die Ursache dessen liegt in der schon zu Abb. 3.10 diskutierten Überschätzung der Kurzkanalschwellspannung.

#### 3.4.2 Subthreshold-Steigung

Wie schon erläutert wurde, ist die Subthreshold-Steigung  $SS$  ein wichtiges Maß für den Übergang des Transistors in den Sperrbereich. Deshalb soll aus der vorhergehenden Modellierung des Drainstroms auch der Zusammenhang zur Subthreshold-Steigung ermittelt werden.

Die Subthreshold-Steigung wird in V/Dekade gemessen. Ihre Definition ist entsprechend [GS02]:

$$SS = \left( \frac{d \log(I_{ds})}{dV_{gs}} \right)^{-1}. \quad (3.60)$$

Da die Funktionen  $v$  und  $\delta$  nur eine schwache Abhängigkeit von  $V_g$  zeigen, können sie bei der Berechnung der Subthreshold-Steigung vernachlässigt werden. Die Ausführung

der Differentiation liefert mit Gl. (3.40):

$$\boxed{SS = \frac{\ln(10) U_T}{\frac{d\psi_{S,min}}{dV_{gs}}} \approx \frac{\ln(10) U_T}{\left(1 - \frac{1}{\cosh(L/(2l_c))}\right) \frac{d\psi_{S,L}}{dV_{gs}}} .} \quad (3.61)$$

In Abb. 3.16 ist zu erkennen, dass die Subthreshold-Steigung für lange Transistoren mit hoher Genauigkeit mit den Ergebnissen der numerischen Simulation übereinstimmt. Demgegenüber ergeben sich bei kurzen Transistoren Abweichungen zum Modell. Für hohe Gatespannungen ist der Unterschied darin begründet, dass der Kurzkanaltransistor eine simulierte Schwellspannung von ca. 0,65 V hat und dass deshalb im Bereich von  $V_g \approx 0,6$  V die moderate Inversion einsetzt [VJD97]. In diesem Bereich steigt die Elektronendichte stark an und unsere bei der Herleitung der Potentialgleichung gemachte Näherung  $n \ll N_A$  ist nicht mehr gültig.

Die Subthreshold-Steigung beträgt für Langkanaltransistoren ca. 65 mV/Dekade und ist, wie direkt aus Gl. (3.16) gefolgert werden kann, geringer als für Kurzkanaltransistoren. Kurzkanaltransistoren weisen deshalb höhere Subthreshold-Leckströme auf als Transistoren mit langen Kanälen. Die Ursache dieser Verschlechterung des Sperrverhaltens liegt in dem verringerten Einfluss des Gates auf das Kanalpotential, der schon in Abschnitt 3.2.3 festgestellt worden war.

## 4 Schwellspannung von ONO-Speichertransistoren

In diesem Abschnitt werden die Erkenntnisse, die im vorherigen Kapitel über Standard-MOS-Transistoren gewonnen wurden, für eine Anwendung zur Modellierung von Transistoren mit ONO-Speicherschicht weiterentwickelt. Es wird ein Schwellspannungsmodell vorgestellt und Schlussfolgerungen zu den Besonderheiten von Speichertransistoren mit lokaler Ladungsspeicherung gezogen.

### 4.1 Modell der Flachbandspannung

Wie in Abschnitt 2.2.2 beschrieben wurde, geschieht die Programmierung der ONO-Transistoren mit lokaler Ladungsspeicherung mit „heißen“, also hochenergetischen Elektronen. Dazu werden die Kanalelektronen durch hohe elektrische Felder am drainseitigen Kanalende beschleunigt. Da die Ladungsträgererzeugung in einem örtlich begrenzten Gebiet am drainseitigen Kanalende geschieht, ist der Ladungsträgereinbau in der Nähe des Draingebiets lokalisiert. Durch diese lokale Verschiebung der Flachbandspannung werden das Kanalpotential und folglich auch das elektrische Verhalten der ONO-Speichertransistoren beeinflusst.

Zur Berechnung des Verlaufs der Flachbandspannung wird das Modell nach Abb. 4.1 verwendet. In der Abbildung wird der Aufbau des Oxid-Nitrid-Oxid-Stapels gezeigt. Abgebildet sind weiterhin ein angenommener Verlauf der Dichte der injizierten Ladungen sowie der prinzipielle Verlauf der Feldstärkekomponente  $E_y$ .

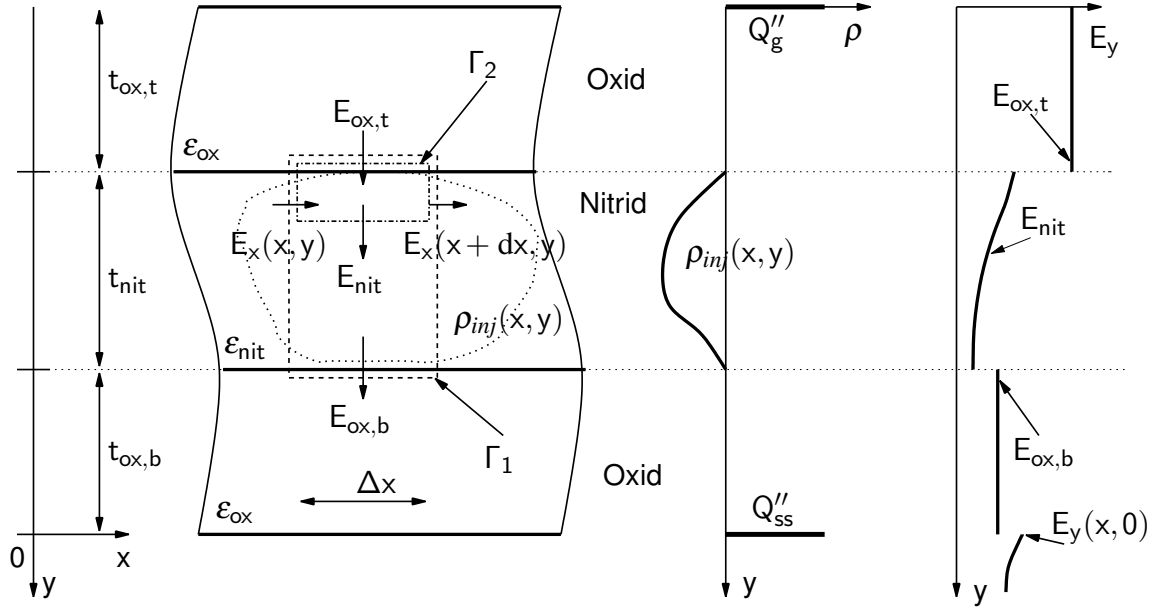
Im Weiteren werden die Feldstärkekomponenten in x-Richtung für das obere und untere Oxid vernachlässigt, so dass sich dort konstante Feldstärken  $E_{ox,t}$  und  $E_{ox,b}$  ergeben. Diese Annahme ist durch die in diesen Bereichen fehlende Raumladung gerechtfertigt. In der Nitridschicht ist eine Raumladung vorhanden, so dass hier im Allgemeinen die Komponente  $E_x$  nicht vernachlässigt werden kann.

Wie in Abb. 3.2 gezeigt wurde, ist die Flachbandspannung die Gate-Bulk-Spannung, für die  $\psi(y) \equiv 0$  und damit  $E_y(y) \equiv 0$  ( $y \geq 0$ ) gilt. Wird diese Bedingung auf Gl. (3.17) angewendet, ergibt sich folgende Definitionsgleichung der Flachbandspannung:

$$E_{ox,b}(V_g = V_{fb}) = -\frac{Q''_{ss}}{\epsilon_{ox}}. \quad (4.1)$$

Für die Ermittlung der Feldstärke  $E_{ox,b}$  in Abhängigkeit von der Gate-Bulk-Spannung wird analog zu Gl. (3.18) in y-Richtung über die Feldstärkekomponente  $E_y$  von der Ga-





**Abbildung 4.1:** Modell des ONO-Stapels mit Verlauf der injizierten Ladungsdichte  $\rho_{inj}(x,y)$  für die Herleitung des Flachbandspannungsmodells

teelektrode bis zur Substratoberfläche integriert:

$$V_g - \Phi_{MS} - \psi_S(x) = \int_0^{t_{ox,t}} E_{ox,t} dy' + \int_0^{t_{nit}} E_{nit}(y') dy' + \int_0^{t_{ox,b}} E_{ox,b} dy'. \quad (4.2)$$

Zur Eliminierung von  $E_{ox,t}$  wird die Gaußsche Fläche  $\Gamma_1$  aus Abb. 4.1 betrachtet. Der Gaußsche Satz liefert hierfür:

$$\begin{aligned} \Delta x \int_0^{t_{nit}} \rho_{inj}(x, y') dy' &= \epsilon_{ox} E_{ox,b} \Delta x - \epsilon_{ox} E_{ox,t} \Delta x + \\ &+ \epsilon_{nit} \int_0^{t_{nit}} E_x(x + dx, y') dy' - \epsilon_{nit} \int_0^{t_{nit}} E_x(x, y') dy'. \end{aligned} \quad (4.3)$$

Eine Umstellung und der Grenzübergang  $\Delta x \rightarrow 0$  liefern:

$$E_{ox,t} = E_{ox,b} - \frac{1}{\epsilon_{ox}} \int_0^{t_{nit}} \underbrace{\left( \rho_{inj}(x, y') - \epsilon_{nit} \frac{\partial E_x(x, y')}{\partial x} \right)}_{=: \rho_{inj}^*(x, y')} dy'. \quad (4.4)$$

Da der Verlauf von  $E_x(x, y')$  im Integral vor allem von der Verteilung von  $\rho_{inj}(x, y')$  abhängt, kann man den Einfluss von  $\partial E_x / \partial x$  auch dadurch berücksichtigen, dass man von

einer Ladungsverteilung  $\rho_{inj}(x, y')$  auf eine effektive Ladungsverteilung  $\rho_{inj}^*(x, y')$  übergeht. Aufgrund der Definition:

$$\int_0^{t_{nit}} \rho_{inj}^*(x, y') dy' =: Q_{inj}''^* \quad (4.5)$$

ergibt sich damit:

$$E_{ox,t} = E_{ox,b} - \frac{Q_{inj}''^*}{\epsilon_{ox}}. \quad (4.6)$$

Ein analoges Vorgehen liefert für die zweite Gaußsche Fläche  $\Gamma_2$  aus Abb. 4.1:

$$E_{nit}(x, y') = \frac{\epsilon_{ox}}{\epsilon_{nit}} E_{ox,t} + \frac{1}{\epsilon_{nit}} \int_0^{y'} \rho_{inj}^*(x, y'') dy''. \quad (4.7)$$

Werden die Zwischenergebnisse aus Gl. (4.6) und Gl. (4.7) in Gl. (4.2) eingesetzt, erhalten wir:

$$\begin{aligned} E_{ox,b} = & \frac{V_g - \Phi_{MS} - \psi_S(x)}{t_{ono}^*} + \frac{t_{ono}^* - t_{ox,b}}{t_{ono}^*} \frac{Q_{inj}''^*}{\epsilon_{ox}} - \\ & - \frac{1}{\epsilon_{nit} t_{ono}^*} \int_0^{t_{nit}} \int_0^{y'} \rho_{inj}^*(x, y'') dy'' dy'. \end{aligned} \quad (4.8)$$

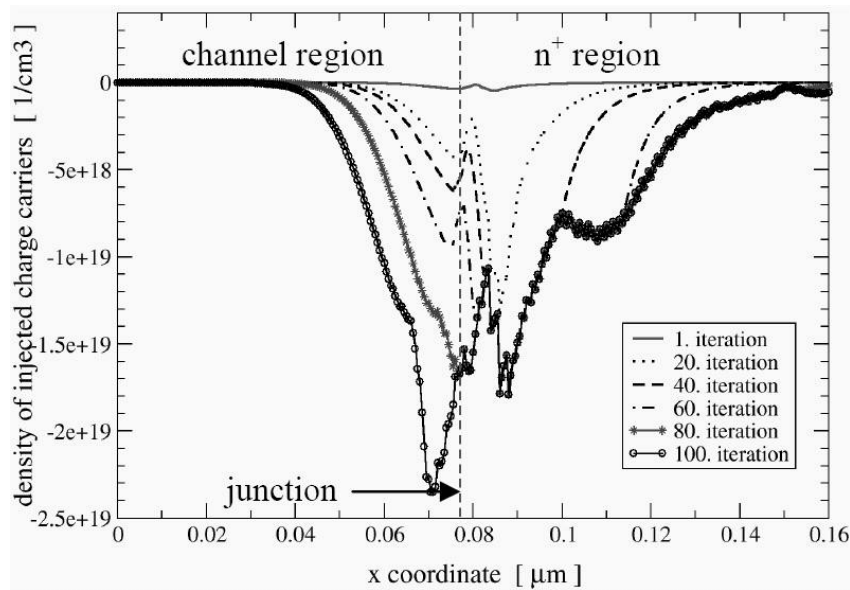
Dabei bezeichnet  $t_{ono}^*$  die elektrisch effektive Höhe des ONO-Stapels:

$$t_{ono}^* = t_{ox,t} + \frac{\epsilon_{ox}}{\epsilon_{nit}} t_{nit} + t_{ox,b}. \quad (4.9)$$

Aus der Definitionsgleichung Gl. (4.1) und unter Beachtung von  $\psi_S = 0$  ergibt sich damit abschließend für die Flachbandspannung:

$$\begin{aligned} V_{fb}(x) = & \underbrace{-\frac{Q_{inj}''^*(x)}{\epsilon_{ox}} (t_{ono}^* - t_{ox,b}) + \frac{1}{\epsilon_{nit}} \int_0^{t_{nit}} \int_0^{y'} \rho_{inj}^*(x, y'') dy'' dy'}_{\Delta V_{fb}(x)} + \\ & + \underbrace{\Phi_{MS} - \frac{Q_{ss}''}{\epsilon_{ox}} t_{ono}^*}_{V_{fb,0}}. \end{aligned} \quad (4.10)$$

Aus dieser Gleichung folgt, dass für den Flachbandspannungsverlauf und damit für das elektrische Verhalten des ONO-Speichertransistors die Verteilung der injizierten Elektronen in der Nitridschicht  $\rho_{inj}^*$  von großer Bedeutung ist. Von besonderem Interesse



**Abbildung 4.2:** Laterale Verteilung der in die Nitridschicht injizierten Ladungen aus [HDF<sup>+</sup>04]; Mit freundlicher Genehmigung ©2004 IEEE

ist dabei die laterale Verteilung [LVP<sup>+</sup>02][TR04]. Diese Verteilung kann mit indirekten Verfahren [SLSD<sup>+</sup>a][SLSD<sup>+</sup>b][LSDM<sup>+</sup>04][LSDBE01], mit numerischen Bauelementesimulationen mit Monte-Carlo-Methoden [ISS02] oder Simulationen mit ballistischen Transportmodellen [HDF<sup>+</sup>04] ermittelt werden.

In Abb. 4.2 wird gezeigt, dass die laterale Verteilung der injizierten Ladungsträger einen Maximalwert besitzt, der sich in der Nähe des metallurgischen Übergangs vom Drain zum Substrat befindet. Der Abfall der injizierten Ladungsträgerdichte in Richtung des Substrats und in Richtung des Drainkontakts erfolgt im Allgemeinen unterschiedlich stark. In [HDF<sup>+</sup>04] wird weiterhin gezeigt, dass die Injektionsstromdichte mit wachsendem Abstand zum metallurgischen Übergang näherungsweise exponentiell abfällt. Damit lässt sich die Annahme rechtfertigen, dass auch die Dichte der injizierten Ladungsträger einem exponentiell abfallenden Verlauf folgt.

Eine wesentliche Beobachtung, die mit Hilfe von numerischen Bauelementesimulationen in [LSDBE01] gewonnen wurde, ist, dass die injizierten Ladungen, die sich über dem Source- bzw. Draingebiet befinden, kaum Einfluss auf die Transfercharakteristik der Transistoren haben. Dieser Effekt ist verständlich, da in MOS-Transistoren die Kanalelektronen das elektrische Transistorverhalten steuern und deshalb auch vor allem die oberhalb des Kanals injizierten Ladungen elektrisch wirksam sind.

Aus diesen Eigenschaften folgt zusammenfassend, dass sich in ONO-Speichertransistoren die Ladungsdichte im Nitrid  $\rho_{inj}$  über dem Kanal exponentiell ändert und nur die über dem Kanalgebiet befindlichen injizierten Elektronen berücksichtigt werden müssen. Setzen wir voraus, dass der Einfluss der lateralen Feldstärkekomponente so gering ist,

dass auch die effektive Ladungsdichte  $\rho_{inj}^*$  näherungsweise einem exponentiellen Verlauf folgt, können wir den folgenden Ansatz wählen:

$$\rho_{inj}^*(x, y') = (\rho_{inj}^*(0, y') \exp(-x/\lambda_s) + \rho_{inj}^*(L, y') \exp(-(L-x)/\lambda_d)) . \quad (4.11)$$

Für die vertikale Verteilung der injizierten Ladungen kann als Ergebnis der Untersuchungen in [SHIT83] und [GPS98] angenommen werden, dass sich die Haftstellen vor allem an der Grenzfläche des Nitrids zum oberen Oxid befinden. In [LSH<sup>+</sup>04] wird präzisiert, dass sich diese Verteilung der Haftstellen nur dann einstellt, wenn die obere Oxidschicht durch eine thermische Oxidation des Nitrids erzeugt wurde. Wird das obere Oxid jedoch mittels eines Hochtemperatur-LPCVD-Verfahrens aufgebracht (*High Temperature Oxid*–HTO), sind die Grenzflächenhaftstellen nur in geringer Zahl vorhanden und die Haftstellen im Inneren der Nitridschicht dominieren. In diesem Fall kann angenommen werden, dass sich die Haftstellen im Zentrum des Nitrids konzentrieren.

Für unsere weitere Herleitung nehmen wir eine Haftstellenverteilung an der Grenzfläche des oberen Oxids zum Nitrid an und berechnen damit:

$$\int_0^{t_{nit}} \int_0^{y'} \rho_{inj}^*(x, y'') dy'' dy' = Q_{inj}''(x) t_{nit} . \quad (4.12)$$

Damit kann Gl. (4.10) berechnet werden und wir erhalten:

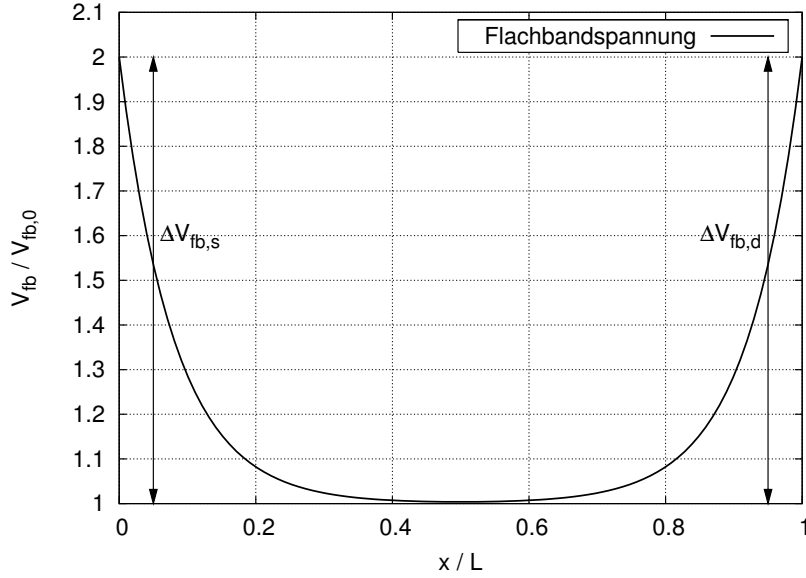
$$V_{fb}(x) = \underbrace{\Phi_{MS} - \frac{Q_{ss}''}{\epsilon_{ox}} t_{ono}^*}_{=: V_{fb,0}} - \underbrace{\frac{Q_{inj}''(x)}{\epsilon_{ox}} t_{ox,t}}_{=: \Delta V_{fb}(x)} . \quad (4.13)$$

Aufgrund des vorausgesetzten exponentiellen Verlaufs von  $\rho_{inj}^*$  kann man den Ansatz rechtfertigen, dass sich in ONO-Speichertransistoren, die sich im programmierten Zustand befinden, die Flachbandspannung über dem Kanal exponentiell ändert. Für den Verlauf der Flachbandspannung über dem Kanal eines programmierten ONO-Speichertransistoren wählen wir deshalb folgendes Modell für  $0 \leq x \leq L$ :

$$V_{fb}(x) = V_{fb,0} + \underbrace{\Delta V_{fb,s} \exp(-x/\lambda_s) + \Delta V_{fb,d} \exp(-(L-x)/\lambda_d)}_{\Delta V_{fb}(x)} .$$

(4.14)

Dabei bezeichnet  $V_{fb,0}$  die Flachbandspannung des Transistors im initialen Zustand, also vor dem Programmiervorgang. Die Parameter  $\Delta V_{fb,s}$  und  $\Delta V_{fb,d}$  kennzeichnen den Maximalwert der durch die Injektion der Ladungsträger in der Nähe des Source- bzw. Drain-Anschlusses hervorgerufenen lokalen Erhöhung der Flachbandspannung. Die Parameter  $\lambda_s$  und  $\lambda_d$  beschreiben den lateralen Abfall der Flachbandspannung. Es gilt die



**Abbildung 4.3:** Modellierter Verlauf der Flachbandspannung in programmierten ONO-Speichertransistoren

Beziehung:

$$\Delta V_{fb,s} = -\frac{Q_{inj}''(0)}{\epsilon_{ox}} t_{ox,t}, \quad (4.15)$$

$$\Delta V_{fb,d} = -\frac{Q_{inj}''(L)}{\epsilon_{ox}} t_{ox,t}. \quad (4.16)$$

Den initialen Zustand des Transistors erhält man, indem man  $\Delta V_{fb,s} = \Delta V_{fb,d} = 0$  setzt. In Abb. 4.3 wird der in Gl. (4.14) definierte Verlauf der Flachbandspannung gezeigt.

Neben dem exponentiellen Ansatz ist es auch möglich, die Flachbandspannung als stückweise konstant zu modellieren. Entsprechende Modelle wurden in [PBI<sup>+</sup>05] und [LHW<sup>+</sup>06] beschrieben. Dieser Ansatz hat jedoch den Nachteil, dass sich statt kontinuierlicher Gleichungen nur stückweise gültige Teillösungen der Gleichungen ergeben und deshalb stets Fallunterscheidungen vorgenommen werden müssen, um den zutreffenden Lösungsanteil zu bestimmen. Dadurch kann die Lösungsstruktur des Gesamtsystems unübersichtlich werden.

Es sei nochmals darauf hingewiesen, dass das Modell Gl. (4.14) den Einfluss der injizierten Ladungen auf das Kanalpotential beschreibt. Darin inbegriffen sind aufgrund des Übergangs zur effektiven Ladungsdichte  $\rho_{inj} \rightarrow \rho_{inj}^*$  nicht nur der Verlauf und die Stärke der Ladungsdichte, sondern auch die laterale Ausdehnung des durch die Ladung verursachten elektrischen Felds. Wäre zum Beispiel die injizierte Ladung eine exakte Punktladung, folgte aus elementaren elektrostatischen Gesetzen, dass die laterale Ausdehnung des elektrischen Felds der Punktladung an der Substrat-Isolator-Grenzfläche trotzdem

von Null verschieden wäre. Dieser elementare Effekt der Elektrostatik ist in dem Modell der Flachbandspannung Gl. (4.14) inbegriffen.

Nachdem ein Modell der Flachbandspannung eines ONO-Speichertransistors hergeleitet wurde, kann damit nun der Verlauf des Oberflächenpotentials ermittelt werden.

## 4.2 Oberflächenpotential

### 4.2.1 Differentialgleichung des Potentialverlaufs

Folgt man der Herleitung in Kapitel 3.2 unter Beachtung der im vorhergehenden Abschnitt dargestellten Besonderheiten von ONO-Speichertransistoren, erhält man eine Differentialgleichung analog zu Gl. (3.26) mit  $V_{fb} \rightarrow V_{fb}(x)$ ,  $t_{ox} \rightarrow t_{ono}^*$ :

$$\frac{d^2 \psi_S(x)}{dx^2} + \frac{2 \epsilon_{ox}}{\epsilon_{si}} \cdot \frac{V_g - V_{fb}(x) - n_0 \psi_S(x)}{t_{ono}^* t_{dep,L}(\psi_{S0})} = \frac{q N_A}{\epsilon_{si}}. \quad (4.17)$$

Der Verlauf von  $V_{fb}(x)$  ist dabei durch Gl. (4.14) gegeben. Für die Lösung dieser Differentialgleichung gelten dieselben Randbedingungen Gl. (3.32a) / Gl. (3.32b) wie bei Standard-MOS-Transistoren.

Unter diesen Randbedingungen kann man die Differentialgleichung Gl. (4.17) lösen und erhält mit  $\psi_{S,L}$  und  $l_c$  aus Gl. (3.36):

$$\begin{aligned} \psi_S(x) &= \psi_1^o \exp(x/l_c) + \psi_2^o \exp(-x/l_c) + \psi_{S,L} + \psi_3 \exp(-x/\lambda_s) + \psi_4 \exp(x/\lambda_d), \\ \psi_1^o &= \frac{V_d + V_{bi} - (V_s + V_{bi} - \psi_{S,L}) \exp(-L/l_c) - \psi_{S,L}}{2 \sinh(L/l_c)} + \\ &+ \psi_3 \frac{\exp(-L/l_c) - \exp(-L/\lambda_s)}{2 \sinh(L/l_c)} + \psi_4 \frac{\exp(-L/l_c) - \exp(L/\lambda_d)}{2 \sinh(L/l_c)}, \\ \psi_2^o &= V_s + V_{bi} - \psi_1^o - \psi_{S,L} - \psi_3 - \psi_4, \\ \psi_{S,L} &= \left( -\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_g - V_{fb,0}} \right)^2 \approx \frac{V_g - V_s - V'_{th0,L} + n_0 \psi_0}{n_0}, \\ \psi_3 &= \Delta V_{fb,s} \frac{\lambda_s^2}{l_c^2 - \lambda_s^2}, \\ \psi_4 &= \Delta V_{fb,d} \exp(-L/\lambda_d) \frac{\lambda_d^2}{l_c^2 - \lambda_d^2}. \end{aligned} \quad (4.18)$$

Dabei bezeichnet  $V'_{th0,L}$  die extrapolierte Schwellspannung eines langen ONO-Speichertransistors im initialen Zustand.

Es sei darauf hingewiesen, dass die Lösungsstruktur von Gl. (4.18) streng genommen nur für den Fall  $\lambda_s \neq l_c$  bzw.  $\lambda_d \neq l_c$  gilt. Der Spezialfall  $\lambda_{s(d)} = l_c$  wird in der Theorie der Differentialgleichungen als *Resonanzfall* bezeichnet. Obwohl bei der Programmierung der ONO-Transistoren dieser Spezialfall durchaus vorkommen kann, wird im Weiteren bei der Herleitung der verschiedenen Gleichungen, wenn nicht anders erwähnt, von  $\lambda_{s(d)} \neq l_c$  ausgegangen, um die Übersichtlichkeit der Zusammenhänge zu bewahren. Der Spezialfall  $\lambda_{s(d)} = l_c$  kann aber im Allgemeinen durch einen Grenzübergang  $\lambda_{s(d)} \rightarrow l_c$  aus der jeweiligen Gleichung gewonnen werden.

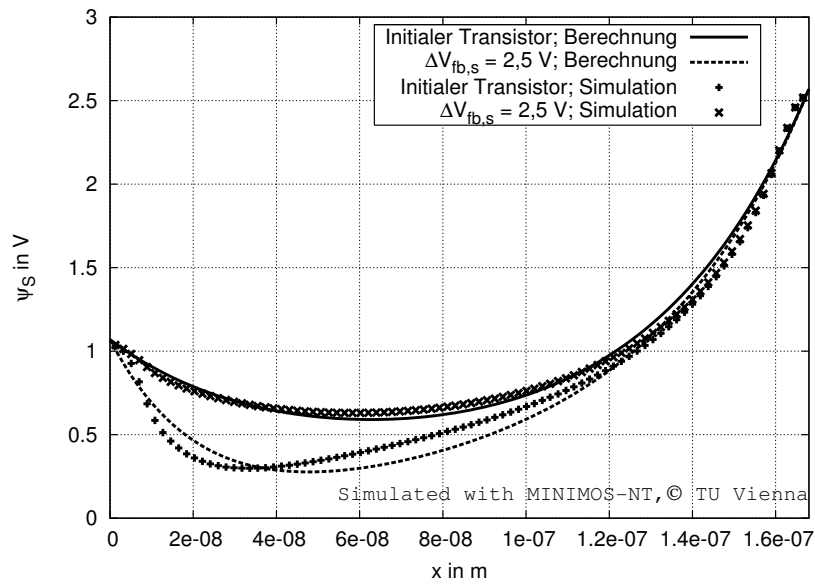
Beispielsweise ist in Gl. (4.18) ersichtlich, dass für den Spezialfall  $\lambda_s = \lambda_d = l_c$  die Koeffizienten  $\psi_3$  und  $\psi_4$  über alle Grenzen wachsen. Ein Grenzübergang nach der Bernoullischen Regel liefert jedoch aus Gl. (4.18) die richtige Lösung der Differentialgleichung für den Resonanzfall. Diese Lösung hat die Struktur:

$$\psi_S(x) = (A + Bx) \exp(-x/l_c) + (C + Dx) \exp(x/l_c) + E. \quad (4.19)$$

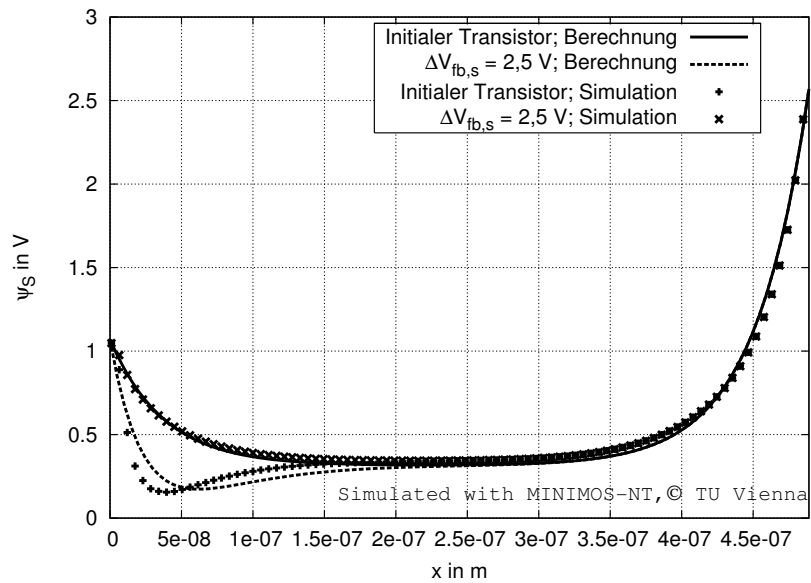
In Abb. 4.4 werden Ergebnisse der Berechnung nach Gl. (4.18) mit den Ergebnissen einer numerischen Bauelementesimulation eines ONO-Transistors verglichen. Man erkennt eine gute Übereinstimmung. Obwohl sowohl bei den Kennlinien der Kurzkanal-, als auch der Langkanaltransistoren für den programmierten Zustand  $\Delta V_{fb,s} = 2,5 \text{ V}$  ein weniger steiler Abfall des Oberflächenpotentials bei  $x \approx 0$  berechnet wird, stimmt die Höhe des berechneten Potentialminimums sehr gut mit der Bauelementesimulation überein.

Für eine qualitative Diskussion der Ergebnisse der Berechnung des Potentialverlaufs nach Gl. (4.18) wird in Abb. 4.5 der modellierte Verlauf des Oberflächenpotentials eines langen Transistors mit ONO-Speicherschicht für verschiedene Fälle gezeigt:

1.  $\Delta V_{fb,s} = 0$  und  $\Delta V_{fb,d} = 0$ : Der Transistor befindet sich im initialen Zustand. Der Potentialverlauf im ONO-Transistor ist äquivalent zu dem in Abschnitt 3.2 untersuchten Verlauf in Standard-MOS-Transistoren. Es gibt genau ein Potentialminimum. Dieses Minimum befindet sich im mittleren Kanalabschnitt. Die Lage und die Höhe des Minimums werden vor allem durch die Eigenschaften des Siliziumsubstrats und die Flachbandspannung bestimmt.
2.  $\Delta V_{fb,s} \neq 0$  oder  $\Delta V_{fb,d} \neq 0$ : An dem source- oder drainseitigen Kanalende wurden Elektronen in das Nitrid injiziert. Es gibt genau ein Potentialminimum. Bei einem geringen Injektionsniveau ist das Minimum unverändert zum initialen Transistor durch die Substrateigenschaften und die Flachbandspannung bestimmt und befindet sich im mittleren Kanalabschnitt. Erhöht man jedoch das Injektionsniveau, wird das Potentialminimum von der Injektionsladung bestimmt. Das Minimum liegt in diesem Fall wie gezeigt in der Umgebung der injizierten Ladung, also am source- bzw. drainseitigen Kanalende.



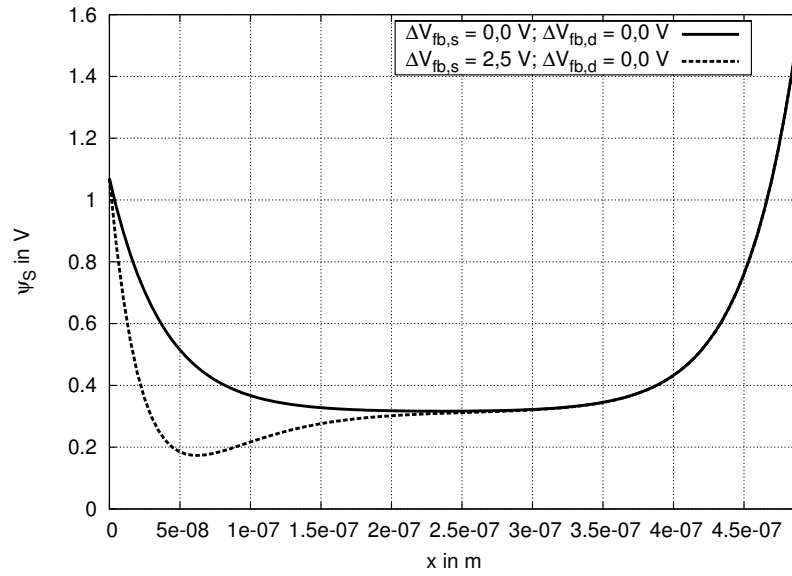
(a)  $L = 0,18 \mu\text{m}$ ;  $V_{gb} = 0,2 \text{ V}$ ,  $V_{ds} = 1,5 \text{ V}$



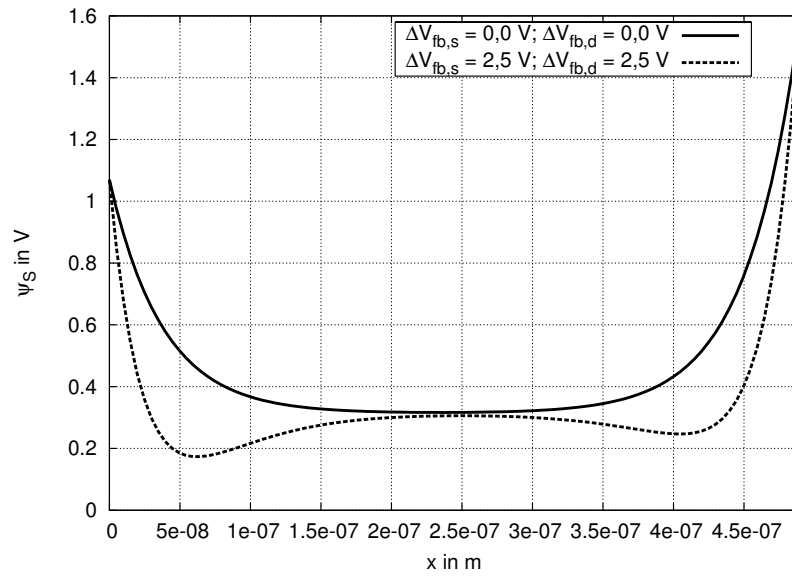
(b)  $L = 0,5 \mu\text{m}$ ;  $V_{gb} = 0,2 \text{ V}$ ,  $V_{ds} = 1,5 \text{ V}$

**Abbildung 4.4:** Vergleich der Berechnung des Oberflächenpotentials nach Gl. (4.18) mit den Ergebnissen einer numerischen Simulation für Transistoren mit unterschiedlichen Längen und Programmierzuständen





(a) Initialer Zustand mit  $\Delta V_{fb,s} = 0$  und Potentialverlauf für  $\Delta V_{fb,s} = 2,5 \text{ V}$



(b) Initialer Zustand mit  $\Delta V_{fb,s} = 0$  und Potentialverlauf für  $\Delta V_{fb,s} = \Delta V_{fb,d} = 2,5 \text{ V}$

**Abbildung 4.5:** Berechneter Verlauf des Oberflächenpotentials eines langen ONO-Transistors mit  $L = 0,5 \mu\text{m}$  für verschiedene Fälle der Injektion von Elektronen;  $V_g = 0,2 \text{ V}$ ;  $V_{ds} = 0,5 \text{ V}$

3.  $\Delta V_{fb,s} \neq 0$  und  $\Delta V_{fb,d} \neq 0$ : An beiden Kanalseiten wurden Elektronen in das Nitrid injiziert. Es gibt bei entsprechender Höhe des Injektionsniveaus zwei lokale Minima. Die lokalen Minima können sich in Abhängigkeit von der injizierten Ladungsmenge im mittleren Kanalbereich oder bei stärkerer Injektion an dem Source- und/oder Drainende des Bauelements befinden. Von der Höhe der beiden Injektionsniveaus  $\Delta V_{fb,s}$  und  $\Delta V_{fb,d}$  sowie von den Anschlusspotentialen  $V_s$  bzw.  $V_d$  ist abhängig, welches der beiden lokalen Minima das globale Minimum des Oberflächenpotentials im Kanalbereich bildet.

#### 4.2.2 Diskussion des Potentialverlaufs des Langkanaltransistors

Die weitere quantitative Diskussion von Lage und Höhe des Potentialminimums wird zunächst am Langkanaltransistor durchgeführt, da sich hier explizite analytische Zusammenhänge herstellen lassen. Diese allgemeinen Zusammenhänge lassen sich dann auf Kurzkanaltransistoren übertragen. Dazu muss Gl. (4.18) für den Langkanalfall vereinfacht werden, indem man in dieser Gleichung den Grenzübergang  $L \rightarrow \infty$  durchführt. Man erhält eine Gleichung, die das Verhalten des Oberflächenpotentials am sourceseitigen Kanalende beschreibt:

$$\psi_S(x) = \underbrace{(V_s + V_{bi} - \psi_{S,L} - \psi_3)}_{=: \psi_{2,L}^o} \exp(-x/l_c) + \psi_3 \exp(-x/\lambda_s) + \psi_{S,L}. \quad (4.20)$$

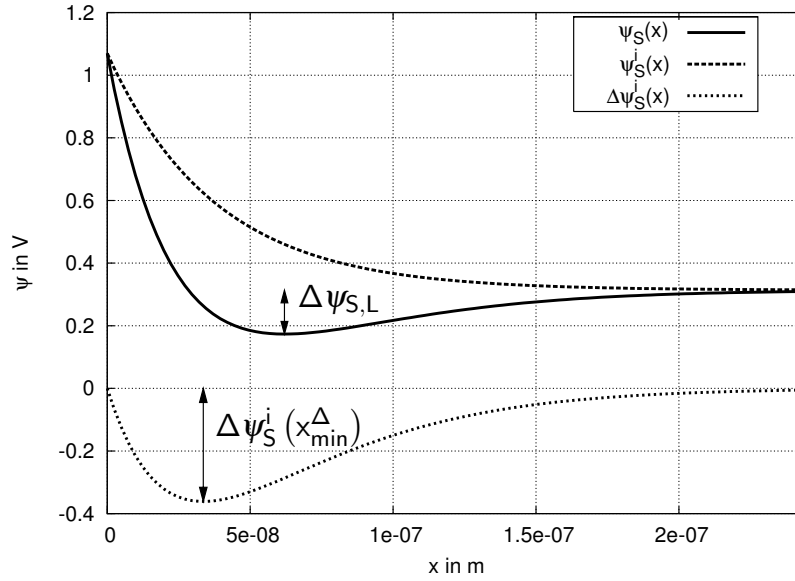
Zur Bestimmung des Potentialverlaufs in der Umgebung des Drainkontakts führt man zunächst zweckmäßigerweise in Gl. (4.18) die Koordinatentransformation  $x = L - x'$  durch und ermittelt dann den Grenzwert für  $L \rightarrow \infty$ :

$$\psi_S(x') = \underbrace{(V_d + V_{bi} - \psi_{S,L} - \psi_4')}_{=: \psi_{1,L}^o} \exp(-x'/l_c) + \psi_4' \exp(-x'/\lambda_d) + \psi_{S,L}, \quad (4.21)$$

$$\psi_4' = \Delta V_{fb,d} \frac{\lambda_d^2}{l_c^2 - \lambda_d^2}.$$

Eine Betrachtung der beiden Langkanalgleichungen lässt erkennen, dass Gl. (4.20) neben allgemeinen Transistorkenngrößen nur das Sourcepotential und das sourceseitige Injektionsniveau enthält, während Gl. (4.21) das korrespondierende drainseitige Potential und Injektionsniveau beinhaltet. Der jeweilige Potentialverlauf wird also nicht von den Spannungen und Programmierzuständen des jeweils anderen Kanalendes beeinflusst. Die beiden Gleichungen sind in dieser Hinsicht entkoppelt. Diese wesentliche Vereinfachung des Langkanalmodells werden wir in den nächsten Abschnitten benutzen.

Weiterhin ist zu erkennen, dass beide Gleichungen eine identische Struktur besitzen und damit Gl. (4.21) aus Gl. (4.20) durch die formalen Substitutionen:  $x \rightarrow x'$ ,  $\lambda_s \rightarrow \lambda_d$ ,  $V_s \rightarrow V_d$  und  $\psi_3 \rightarrow \psi_4'$  hervorgeht. Damit kann man Untersuchungen häufig auf das



PSfrag replacements

**Abbildung 4.6:** Verlauf des Oberflächenpotentials  $\psi_S(x)$  am sourceseitigen Kanalende eines Langkanaltransistors und Zerlegung in den initialen Potentialverlauf  $\psi_S^i(x)$  und die Verschiebung durch die injizierten Ladungen  $\Delta\psi_S^i(x)$ ;  $l_c = 37 \text{ nm}$ ;  $\lambda_s = 30 \text{ nm}$ ;  $V_g = 0,2 \text{ V}$ ;  $\Delta V_{fb,s} = 2,5 \text{ V}$

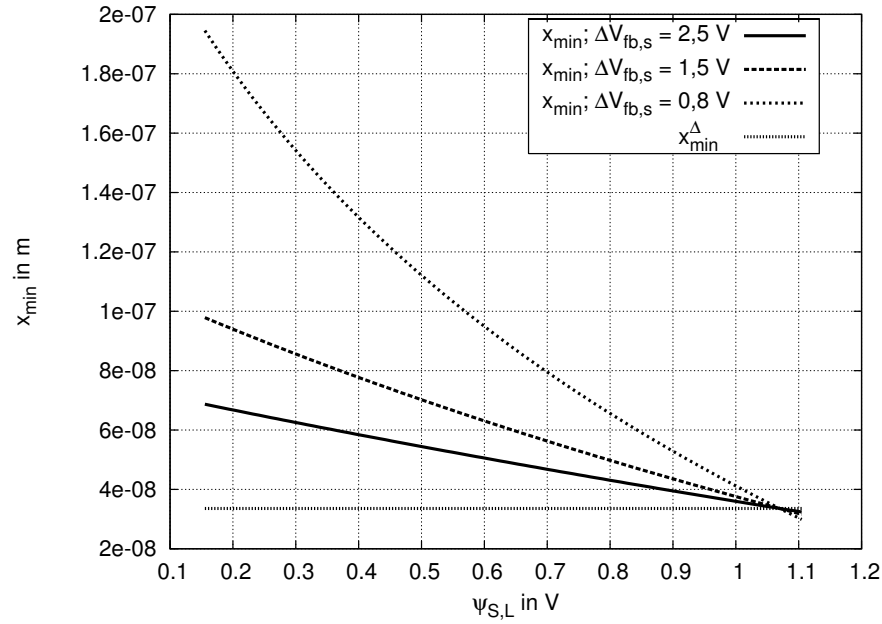
sourceseitige Kanalende beschränken und die erzielten Ergebnisse durch die Substitutionen auf die Umgebung des Draingebiets übertragen.

Für die weitere Analyse kann man die Gleichung Gl. (4.20) durch Umsortieren auf folgende Form bringen:

$$\begin{aligned} \psi_S(x, V_g) = & \underbrace{(V_s + V_{bi} - \psi_{S,L}(V_g)) \exp(-x/l_c) + \psi_{S,L}(V_g)}_{\psi_S^i(x, V_g)} + \\ & + \underbrace{\psi_3 (\exp(-x/\lambda_s) - \exp(-x/l_c))}_{\Delta\psi_S^i(x)}. \end{aligned} \quad (4.22)$$

Der erste Teil der Gleichung ist unabhängig vom Injektionsniveau und beschreibt den exponentiellen Abfall des Kanalpotentials an der Sourceseite eines MOS-Transistors. Dieser initiale Potentialverlauf  $\psi_S^i(x)$  ist unabhängig von der Programmierung des Transistors und nur durch Transistoreigenschaften, wie zum Beispiel die Kanaldotierung, und durch den Arbeitspunkt des Transistors bestimmt. Der initiale Potentialverlauf stimmt mit dem sourceseitigen Verlauf des Oberflächenpotentials eines langen Standard-MOS-Transistors bzw. eines langen ONO-Transistors ohne Programmierung, also im initialen Zustand, überein.

Der zweite Term der Gleichung beinhaltet die Verschiebung des initialen Potentialverlaufs durch die im Nitrid befindlichen Ladungen. In Abb. 4.6 wird der Gesamtpotential-



**Abbildung 4.7:** Lage des Potentialminimums eines Langkanaltransistors für verschiedene Injektionsniveaus  $\Delta V_{fb,s}$ ;  $l_c = 37 \text{ nm}$ ,  $\lambda_s = 30 \text{ nm}$

verlauf zusammen mit den beiden Einzelkomponenten dargestellt.

Eine Diskussion des Kurvenverlaufs ergibt für die Lage des Minimums von  $\Delta\psi_S^i(x)$  und damit also für den Ort der betragsmäßig größten Verschiebung des initialen Potentialverlaufs:

$$x_{min}^{\Delta} = \frac{l_c \lambda_s}{l_c - \lambda_s} \ln \left( \frac{l_c}{\lambda_s} \right), \quad (4.23)$$

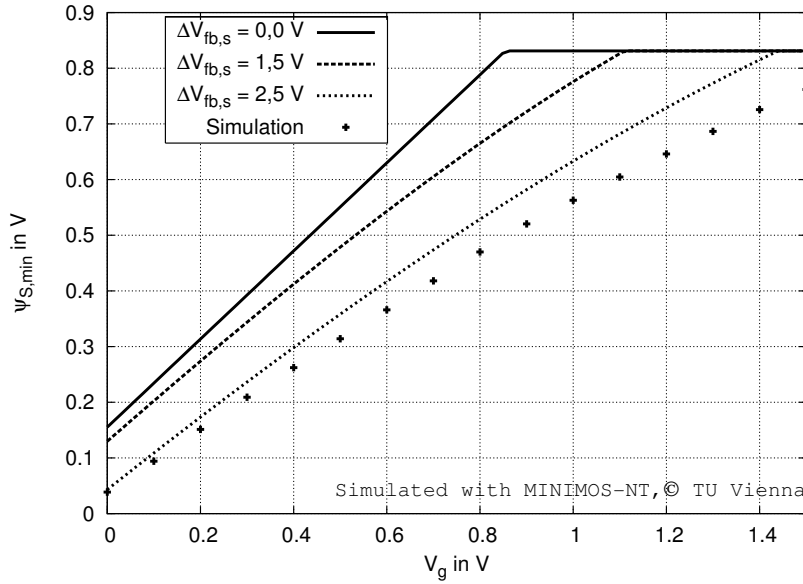
$$\Delta\psi_S^i(x_{min}^{\Delta}) = -\Delta V_{fb,s} \frac{\lambda_s}{l_c + \lambda_s} \left( \frac{\lambda_s}{l_c} \right)^{\frac{l_c}{l_c - \lambda_s}}. \quad (4.24)$$

Für die weitere Diskussion wird die Lage des Minimums des Potentialverlaufs benötigt. Einfache mathematische Standardverfahren ergeben für die x-Koordinate:

$$x_{min} = \frac{\lambda_s l_c}{l_c - \lambda_s} \ln \left( -\frac{l_c}{\lambda_s} \frac{\psi_3}{\psi_{2,L}^o} \right) \quad (4.25)$$

$$= \underbrace{\frac{\lambda_s l_c}{l_c - \lambda_s} \ln \left( \frac{l_c}{\lambda_s} \right)}_{x_{min}^{\Delta}} - \frac{\lambda_s l_c}{l_c - \lambda_s} \ln \left( -\frac{\psi_{2,L}^o}{\psi_3} \right). \quad (4.26)$$

In Abb. 4.7 wird der Verlauf der x-Koordinate des Minimums für verschiedene Injektionshöhen dargestellt. Man erkennt, dass bei steigendem Langkanaloberflächenpotential  $\psi_{S,L}$  und damit bei steigendem Gatepotential das Potentialminimum in Richtung



**Abbildung 4.8:** Minimalwert des Kanalpotentials eines Langkanal-ONO-Transistors in Abhängigkeit vom Gatepotential  $V_g$  für verschiedene Injektionsniveaus und Vergleich mit einer numerischen Bauelementesimulation mit  $\Delta V_{fb,s} = 2,5 \text{ V}$ . Die berechneten Funktionen wurden auf  $\psi_{S,min} = \psi_0 = V_s + 2\Psi_{FB}$  begrenzt.

des Sourcegebiets verschoben wird. Bei  $\psi_{S,L} = V_s - V_{bi}$  ist, unabhängig vom Injektionsniveau,  $x_{min} = x_{min}^\Delta$ . Ebenfalls ersichtlich ist, dass man den Funktionsverlauf gut durch eine Gerade annähern kann. Entwickelt man Gl. (4.25) am Schnittpunkt der Funktionenschar ( $x_{min}^\Delta$ ;  $\psi_{S,L} = V_s + V_{bi}$ ) in eine Taylorreihe, erhält man folgende Näherungsformel:

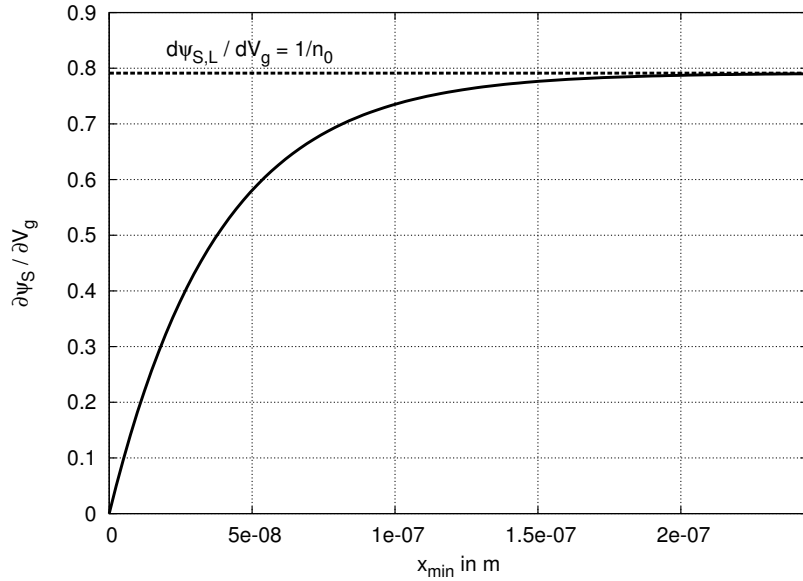
$$x_{min}(\psi_{S,L}) \approx x_{min}^\Delta + l_c \frac{\lambda_s + l_c}{\lambda_s} \frac{1}{\Delta V_{fb,s}} (V_s + V_{bi} - \psi_{S,L}) . \quad (4.27)$$

Der Minimalwert des Oberflächenpotentials eines langen ONO-Speichertransistors mit sourceseitiger Programmierung ist:

$$\psi_{S,min}(V_g) = \underbrace{\psi_{S,L}(V_g) - \Delta V_{fb,s} \frac{\lambda_s}{l_c + \lambda_s} \exp(-x_{min}(V_g)/\lambda_s)}_{\Delta\psi_{S,L}} , \quad (4.28)$$

wobei  $\Delta\psi_{S,L}$  die durch die Programmierung hervorgerufene Differenz vom Potential eines unprogrammierten, initialen Langkanaltransistors zum Minimalwert des Potentials eines programmierten Langkanaltransistors beschreibt. Diese Potentialdifferenz wurde bereits in Abb. 4.6 eingetragen.

Ein typischer Verlauf der Höhe des Potentialminimums über dem Gatepotential  $V_g$ , berechnet nach Gl. (4.28), wird in Abb. 4.8 gezeigt. Zum Vergleich sind die Ergebnis-



**Abbildung 4.9:** Die Funktion  $\partial \psi_S(x, V_g) / \partial V_g$  nach Gl. (4.30) zur Veranschaulichung des Einflusses des Gatepotentials auf das Kanalpotential in ONO-Speichertransistoren

se einer numerischen Bauelementesimulation eingetragen. Man erkennt, dass sich bei einer Erhöhung der Programmierung des Transistors die Steilheit der Kennlinie verringert. Die Auswirkungen dieses Effekts werden noch bei der Modellierung des Verlaufs der Subthreshold-Kennlinie zu diskutieren sein, da, genau wie bei Standard-MOS-Transistoren, auch bei ONO-Transistoren der Anstieg  $d\psi_{s,min}/dV_g$  einen direkten Einfluss auf die Subthreshold-Steigung hat.

Für die weitere Berechnung von  $d\psi_{s,min}/dV_g$  könnte direkt Gl. (4.28) verwendet werden, eine tiefere Einsicht in den physikalischen Hintergrund der Verringerung der Steilheit der Kennlinie  $\psi_{s,min}(V_g)$  kann aber erlangt werden, indem man von der ursprünglichen Gleichung Gl. (4.22) ausgeht:

$$\frac{d\psi_{s,min}}{dV_g} = \frac{d\psi_S(x_{min}(V_g), V_g)}{dV_g} = \underbrace{\frac{\partial \psi_S(x_{min}, V_g)}{\partial x}}_0 \frac{dx_{min}}{dV_g} + \underbrace{\frac{\partial \psi_S(x_{min}, V_g)}{\partial V_g}}_{\partial \psi_S^i / \partial V_g}, \quad (4.29)$$

$$= \frac{d\psi_{S,L}}{dV_g} (1 - \exp(-x_{min}/l_c)). \quad (4.30)$$

Für die Interpretation dieses Zusammenhangs sei auf Abb. 4.9 verwiesen, in der die Funktion nach Gl. (4.30) gezeigt wird. Diese Kennlinie stellt die Ortsabhängigkeit der Auswirkung von Gatespannungsänderungen auf die Änderung des Kanalpotentials dar. Diese Änderung des Kanalpotentials in Abhängigkeit vom Gatepotential kann als ein

Maß für die Effektivität der Steuerung des Kanals durch das Gatepotential interpretiert werden.

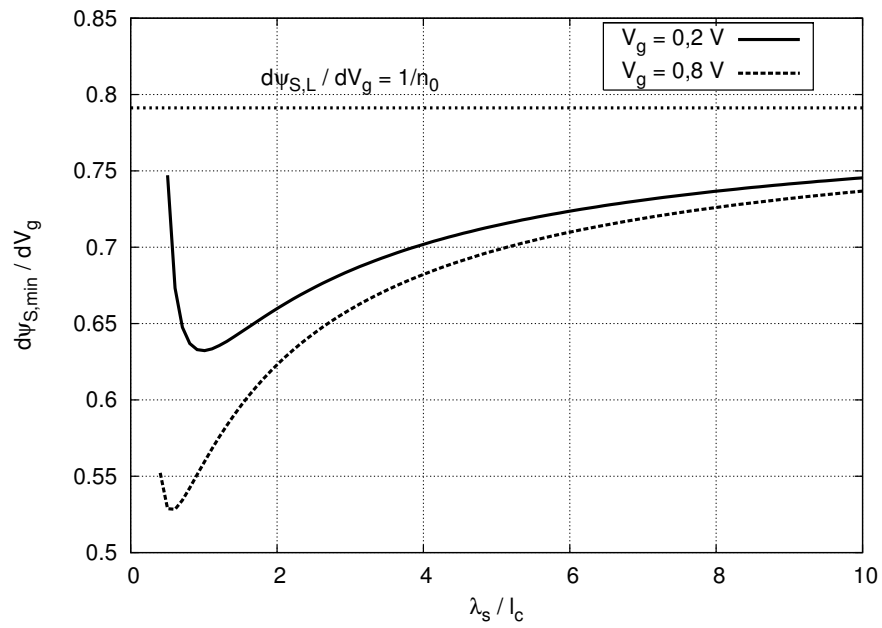
Unmittelbar am Sourcegebiet ist diese Funktion Null, da das Oberflächenpotential an dieser Stelle aufgrund der Randbedingung nur vom Sourcepotential und nicht vom Gatepotential abhängt. Im Inneren des Transistors, bei großen Werten der Abszisse, ist mit  $d\psi_S/dV_g \rightarrow 1/n_0$  die gleiche Effektivität der Gatesteuerung wie bei Langkanaltransistoren erreicht. Zwischen beiden Regionen gibt es einen Übergangsbereich, in dem die Gatesteuerung zwar vorhanden, im Vergleich zum Langkanaltransistor jedoch geringer ist.

Anhand der Abb. 4.9 kann man  $d\psi_{S,min}/dV_g$  dadurch ermitteln, dass man am Graphen den Funktionswert an der Stelle  $x = x_{min}$  abliest. Wie gezeigt wurde, ist bei ONO-Speichertransistoren die Lage des Minimums  $x_{min}$  sowohl vom Arbeitspunkt, als auch von der Programmierung  $\Delta V_{fb,s}$  und der Ausdehnung der Injektionszone  $\lambda_s$  abhängig. Es wurde ebenfalls gezeigt, dass sich dieses Minimum für große  $\Delta V_{fb,s}$ , kleine  $\lambda_s < l_c$  oder große  $V_g$  in Richtung des Sourcekontakts verschiebt. Da in der Umgebung des Sourcegebiets entsprechend Abb. 4.9 der Gateeinfluss abnimmt, verringert sich unter diesen Bedingungen die Steilheit von  $\psi_{S,min}(V_g)$ . Für große  $\lambda_s > l_c$  befindet sich hingegen das Minimum des Potentials im mittleren Kanalabschnitt. An dieser Stelle zeigt die Gatesteuerung aber Langkanalverhalten. Damit ist für große  $\lambda_s$  keine Verringerung des Anstiegs von  $\psi_{S,min}(V_g)$  im Vergleich zum Langkanalfall zu erwarten.

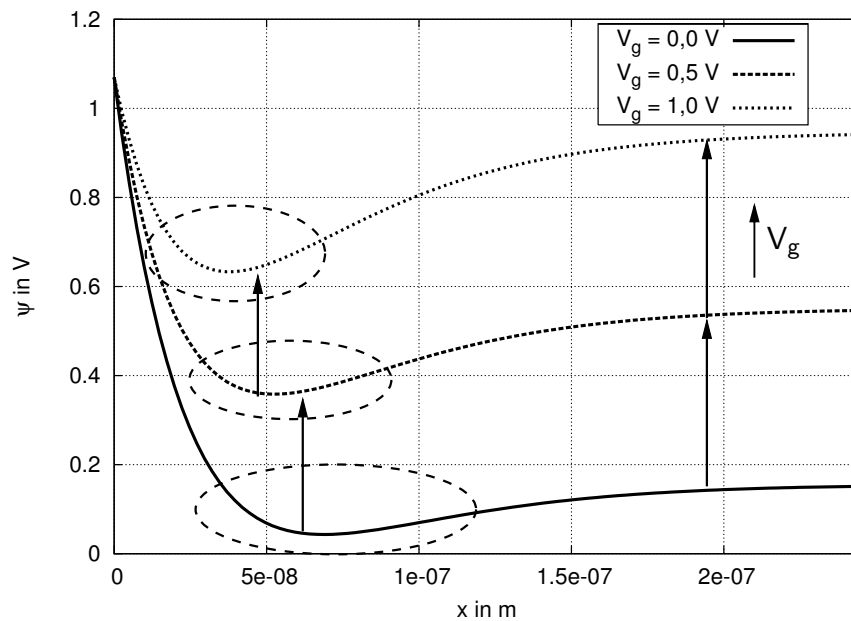
Weil, wie bei normalen MOS-Transistoren, ein direkter Zusammenhang zwischen dem Anstieg von  $\psi_{S,min}(V_g)$  und der Subthreshold-Steigung besteht, kann man mit diesen Überlegungen die Beobachtungen in [LSDBE01] und [LSDM<sup>+</sup>04] erklären, nach denen eine Vergrößerung der Subthreshold-Steigung von programmierten ONO-Speichertransistoren vor allem bei geringen Ausdehnungen der Injektionszone  $\lambda_s$  auftritt, während sich bei großen Ausdehnungen die Langkanalsteilheit einstellt.

In Abb. 4.10 ist der Verlauf von  $d\psi_{S,min}/dV_g$  über die auf die charakteristische Länge normierte Ausdehnung des Injektionsgebiets  $\lambda_s/l_c$  aufgetragen. Der berechnete Funktionsverlauf stimmt qualitativ mit den aus numerischen Simulationen gewonnenen Ergebnissen in [TR04] überein.

Geht man also davon aus, dass der Funktionswert von  $d\psi_{S,min}/dV_g$  in der Kanalmitte grösser als am sourceseitigen Ende ist, wird auch der in Abb. 4.11 gezeigte beulenartige Verlauf des Oberflächenpotentials in der Umgebung des Minimums am sourceseitigen Kanalende plausibel: Bei steigendem Gatepotential  $V_g$  steigt das Oberflächenpotential in der Kanalmitte aufgrund der an dieser Stelle größeren Steilheit  $d\psi_{S,min}/dV_g$  schneller als in der Umgebung des Sourcekontakts. Damit bildet sich mit steigendem Gatepotential die in [LSDBE01] beschriebene charakteristische Ausbeulung im Verlauf von  $\psi_S(x)$  aus. In dieser Sichtweise ist diese Ausbeulung nicht, wie dort behauptet, die Ursache der vergrößerten Subthreshold-Steigung, sondern ebenfalls eine direkte Folge des ortsabhängigen Einflusses der Gates auf den Kanal.



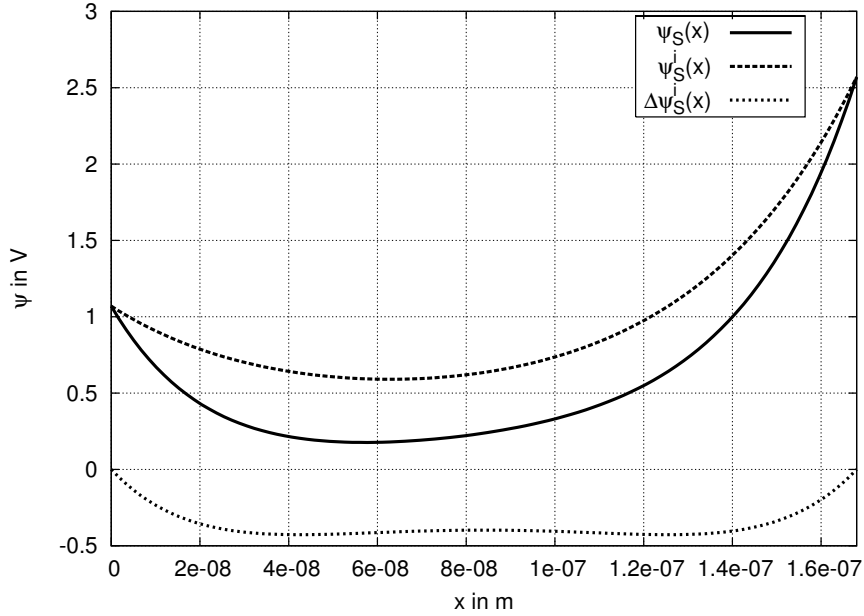
**Abbildung 4.10:** Die Funktion  $d\psi_{S,min}/dV_g$  in Abhängigkeit von der normierten Ausdehnung der Injektionszone  $\lambda_s/l_c$



PSfrag replacements

**Abbildung 4.11:** Charakteristische Form der Ausbildung des Minimums im Verlauf des Oberflächenpotentials bei steigendem Gatepotential





**Abbildung 4.12:** Verlauf des Oberflächenpotentials  $\psi_S(x)$  eines Kurzkanaltransistors und Zerlegung in den initialen Potentialverlauf  $\psi_S^i(x)$  und die Verschiebung durch injizierte Ladungen  $\Delta\psi_S^i(x)$ ;  
 $L = 0,18 \mu\text{m}$ ;  $l_c = 37 \text{ nm}$ ;  $\lambda_s = \lambda_d = 37 \text{ nm}$ ;  
 $V_g = 0,2 \text{ V}$ ;  $\Delta V_{fb,s} = \Delta V_{fb,d} = 2,5 \text{ V}$

### 4.2.3 Diskussion des Potentialverlaufs eines Kurzkanaltransistors

Nach der Herleitung und der Diskussion des Potentialverlaufs bei Langkanal-Speichertransistoren soll nun eine Verallgemeinerung für den Kurzkanalfall und für eine Injektion am drainseitigen Kanalende erfolgen. Zur Ableitung des initialen Potentialverlaufs  $\psi_S^i$  sowie der injektionsbedingten Verschiebung des Oberflächenpotentials  $\Delta\psi_S^i$  wird Gl. (4.18) umgestellt. Man erhält mit  $\psi_1$  und  $\psi_2$  aus Gl. (3.36):

$$\psi_S = \psi_S^i + \Delta\psi_S^i, \quad (4.31)$$

$$\psi_S^i = \psi_1 \exp(x/l_c) + \psi_2 \exp(-x/l_c) + \psi_{S,L}, \quad (4.32)$$

$$\begin{aligned} \Delta\psi_S^i = & \psi_3 \left( \exp(-x/\lambda_s) - \exp(-x/l_c) + \frac{\exp(-L/l_c) - \exp(-L/\lambda_s)}{\sinh(L/l_c)} \sinh(x/l_c) \right) + \\ & + \psi_4 \left( \exp(x/\lambda_d) - \exp(-x/l_c) + \frac{\exp(-L/l_c) - \exp(L/\lambda_d)}{\sinh(L/l_c)} \sinh(x/l_c) \right). \end{aligned} \quad (4.33)$$

In Abb. 4.12 werden typische Verläufe der beiden Komponenten des Oberflächenpotentials dargestellt.

Die lokalen Extrempunkte des Potentialverlaufs können als Nullstellen der folgenden Gleichung bestimmt werden:

$$0 = \psi_1^o z - \psi_2^o z^{-1} - \psi_3 \frac{l_c}{\lambda_s} z^{-\frac{l_c}{\lambda_s}} + \psi_4 \frac{l_c}{\lambda_d} z^{\frac{l_c}{\lambda_d}}, \quad (4.34)$$

$$x_i = l_c \ln z_i.$$

Wie aus dieser Struktur zu erkennen ist, kann diese Gleichung für den allgemeinen Fall nur mit numerischen Hilfsmitteln gelöst werden.

Wie schon erwähnt wurde, folgt aus den Betrachtungen des Potentialverlaufs, dass diese Gleichung für physikalisch sinnvolle Verläufe des Oberflächenpotentials bis zu drei Nullstellen im Intervall  $0 \leq x \leq L$  hat. Für den Fall einer Nullstelle beschreibt diese natürlich das globale Minimum:

$$\psi_{S,min} = \psi_S(x_{min}). \quad (4.35)$$

Gibt es zwei Nullstellen  $x_1$  und  $x_2$ , ist eine der Nullstellen ein lokales Minimum, die andere ein Wendepunkt. Man erhält dann zweckmäßigerweise den gesuchten globalen Minimalwert, indem man die dazugehörigen Werte des Oberflächenpotentials aus der Ursprungsgleichung Gl. (4.18) ermittelt und berechnet:

$$\psi_{S,min} = \min(\psi_S(x_1), \psi_S(x_2)). \quad (4.36)$$

Sind drei Nullstellen  $x_1 < x_2 < x_3$  vorhanden, sind  $x_1$  und  $x_3$  lokale Minima. Die dazugehörigen Werte des Oberflächenpotentials ermittelt man dann ebenfalls aus Gl. (4.18):

$$\psi_{S,min} = \min(\psi_S(x_1), \psi_S(x_3)). \quad (4.37)$$

Wie im Langkanalfall der Gl. (4.29) kann man zur Bestimmung des Einflusses des Gatepotentials berechnen:

$$\frac{d\psi_{S,min}}{dV_g} = \frac{\partial \psi_S(x_{min}, V_g)}{\partial V_g}, \quad (4.38)$$

$$= \left( 1 - \frac{\sinh(x_{min}/l_c) + \sinh((L - x_{min})/l_c)}{\sinh(L/l_c)} \right) \frac{d\psi_{S,L}}{dV_g}, \quad (4.39)$$

wobei für  $x_{min}$  diejenige Lösung aus Gl. (4.34) einzusetzen ist, für die sich das globale Potentialminimum  $\psi_{S,min}$  ergibt.

Man kann erkennen, dass sich auch im Kurzkanalfall eine Verringerung der Steilheit von  $\psi_{S,min}(V_g)$  dadurch ergibt, dass das globale Minimum in den Bereich der geringeren Gatesteuerung am Kanalende verschoben wird.

#### 4.2.4 Grenzen und Erweiterungsmöglichkeiten des Potentialmodells

Da die weiteren Herleitungen des Schwellspannungs- und Subthresholdmodells auf dem Modell der Potentialverteilung beruhen, sollen zunächst noch die Grenzen des Potentialmodells untersucht werden und anschließend ein Ansatz aufgezeigt werden, der es erlaubt, den Gültigkeitsbereich des Potentialmodells zu erweitern.

Zunächst sei wiederholt, dass bei dem Übergang von der partiellen Differentialgleichung der Potentialverteilung zur quasi-zweidimensionalen Näherung in Abschnitt 3.2.1 gemäß Gl. (3.20) die Ansatzfunktion:

$$\frac{\partial^2 \psi(x, y)}{\partial x^2} = \frac{d^2 \psi_S(x)}{dx^2} \quad (4.40)$$

gewählt wurde. Wird diese Ansatzfunktion in Gl. (3.19b) eingesetzt, erhält man für den Potentialverlauf nach wenigen Schritten:

$$\psi(x, y) = \psi_S(x) \left( \frac{t_{dep}(x) - y}{t_{dep}(x)} \right)^2. \quad (4.41)$$

Gemäß dieser Gleichung ist der Potentialverlauf in y-Richtung quadratisch. Der Scheitelpunkt der Parabel liegt bei  $y = t_{dep}$  mit  $\psi(x, t_{dep}) = 0$ . Ist folglich das Oberflächenpotential  $\psi_S > 0$ , liegt der Maximalwert des Potentialverlaufs entlang der y-Koordinate immer bei  $y = 0$ . Da sich der Elektronenkanal am Ort des maximalen Potentials befindet, beginnt sich unter diesen Bedingungen der Elektronenkanal bei Inversion an der Oberfläche des Halbleitersubstrats auszubilden. Unter diesen Voraussetzungen ist es daher gerechtfertigt, für die weitere Berechnung der Schwellspannung nur das Oberflächenpotential zu betrachten.

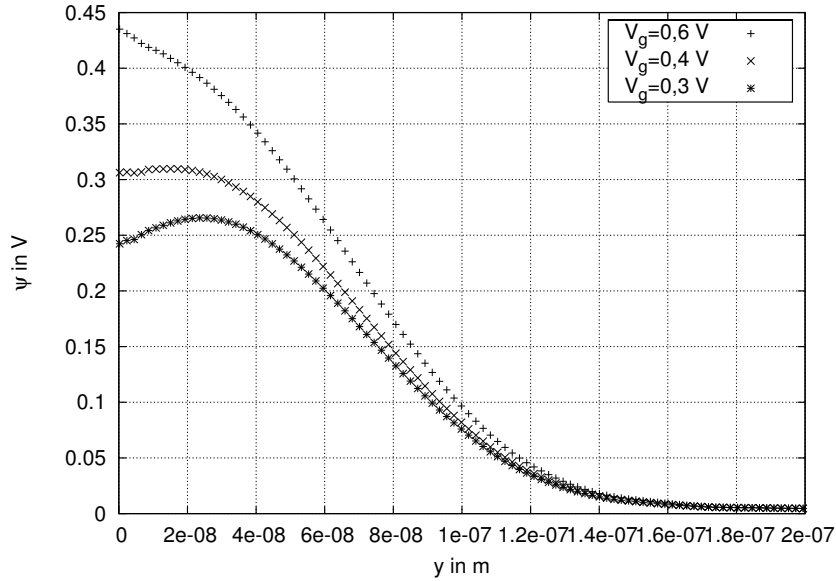
In Abb. 4.13 ist jedoch der Potentialverlauf entlang der y-Achse als Ergebnis einer numerischen Bauelementesimulation gezeigt. Offensichtlich ergeben sich unter gewissen Bedingungen bei sinkender Gatespannung immer stärkere Abweichungen vom quadratischen Verlauf des Modells. In dem in der Abbildung gezeigten Fall befindet sich bei einem Gatepotential von  $V_g = 0,3 \text{ V}$  das Potentialmaximum sogar deutlich unterhalb der Substratoberfläche. Das Modell kann diesen Effekt aufgrund des gewählten Ansatzes prinzipiell nicht darstellen und wird deshalb unter diesen Bedingungen ungenauere Vorhersagen liefern.

Zur Abschätzung des Grenzwerts, bis zu dem der Kanal an der Substratoberfläche ausgebildet wird, wird gefordert:

$$\frac{\partial \psi}{\partial y}(x, 0) = -E_y(x, 0) \stackrel{!}{<} 0. \quad (4.42)$$

Unter Berücksichtigung von Gl. (3.17) und Gl. (3.18) erhält man damit::

$$\boxed{V_g - V_{fb}(x) \stackrel{!}{>} \psi_S(x).} \quad (4.43)$$



**Abbildung 4.13:** Potentialverlauf entlang eines y-Schnittes in einem Abstand von  $x = 53 \text{ nm}$  vom Source-Kontakt in einem ONO-Transistor mit  $L = 0,5 \mu\text{m}$ ,  $\Delta V_{fb,s} = 2,5 \text{ V}$  und  $V_{ds} = 1,5 \text{ V}$

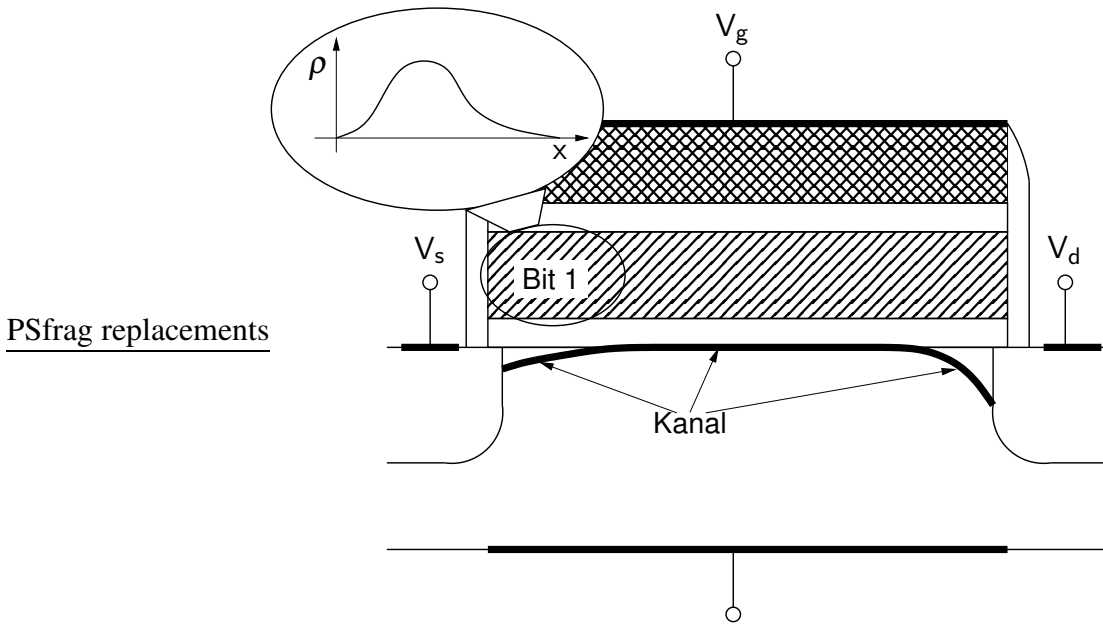
Diese Forderung ist um so schwieriger zu erfüllen, je größer die Flachbandspannung und/oder das Oberflächenpotential ist. Da das Oberflächenpotential in der Umgebung des Source- und besonders des Drainkontakts ansteigt, sind in diesen Gebieten Ungenauigkeiten in der Modellierung zu erwarten. Weiterhin folgt aus der Tatsache, dass in programmierten ONO-Transistoren die Flachbandspannung lokal erhöht ist aus der Forderung Gl. (4.43), dass stark programmierte ONO-Transistoren ungenauer als schwach programmierte ONO-Transistoren oder Standard-MOS-Transistoren modelliert werden.

Aus diesen Erkenntnissen lässt sich auch auf den qualitativen Verlauf des Elektronenkanals bei stark programmierten Transistoren schließen, der in Abb. 4.14 skizziert wurde: Aufgrund der starken Programmierung verschiebt sich der Elektronenkanal am sourceseitigen Kanalende in das Innere des Halbleitergebiets. Am drainseitigen Kanalende wird der Kanalverlauf aufgrund der hohen Drainspannung in das Substratinnere gebogen.

Um den Verlauf des Potentials nach Abb. 4.13 prinzipiell beschreiben zu können, ist mindestens ein Polynom dritter Ordnung erforderlich. Um dieses zu erhalten, kann zum Beispiel folgende Ansatzfunktion gewählt werden:

$$\frac{\partial^2 \psi(x, y)}{\partial x^2} = \frac{d^2 \psi_S(x)}{dx^2} \frac{t_{dep}(x) - y}{t_{dep}(x)}. \quad (4.44)$$

Dieser Ansatz setzt voraus, dass sich das Potential in x-Richtung am Ende der Verarmungszone nur wenig ändert und dort deshalb  $\partial^2 \psi(x, t_{dep}) / \partial x^2 = 0$  gilt. Zwischen dem Wert der Krümmung des Potentialverlaufs an der Oberfläche und am Ende der Ver-



**Abbildung 4.14:** Qualitativer Verlauf des Elektronenkanals in einem sourceseitig stark programmierten ONO-Transistor

mungszone wird dabei linear interpoliert.

Setzt man diese Ansatzfunktion in Gl. (3.19b) ein und löst man die Integrale, erhält man das kubische Polynom:

$$\psi(x, y) = \frac{q N_A}{2 \epsilon_{si}} \left( (t_{dep} - y)^2 - \frac{(t_{dep} - y)^3}{t_{dep}} \right) + \psi_S \left( \frac{t_{dep} - y}{t_{dep}} \right)^3 \quad (4.45)$$

Wie schon in Abschnitt 3.2.1 vorgeführt wurde, kann man diesen Ansatz weiterhin auf Gl. (3.19a) anwenden und das resultierende Differentialgleichungssystem für  $y = 0$  durch Elimination von  $t_{dep}$  in eine nichtlineare Differentialgleichung in  $\psi_S(x)$  umwandeln. Wie gezeigt wurde, kann man diese dann durch Linearisierung in eine lineare Differentialgleichung umwandeln und diese lösen. Die Lösung dieser Gleichung ist unter den gegebenen Näherungen möglich, obgleich sich unübersichtliche Resultate ergeben. Aus dieser Lösung und dem kubischen Polynom Gl. (4.45) kann man dann letztendlich den Maximalwert des Potentials in  $y$ -Richtung bestimmen.

Leider ist die mathematische Struktur dieser Lösung derart, dass sich die Gleichung bei der anschließenden Berechnung des Schwellspannungswerts nicht invertieren lässt. Aus diesem Grund wurde auf die weitere Analyse dieses Wegs trotz der potentiell höheren Modellierungsgenauigkeit verzichtet.

### 4.3 Schwellspannungsmodell

Wie aus den Ausführungen des vorherigen Abschnitts erkannt werden kann, ist bei der Definition der Schwellspannung von ONO-Transistoren zu beachten, dass der Potentialverlauf ein oder zwei lokale Minima aufweisen kann. Unter diesen Randbedingungen kann man die Schwellspannung von ONO-Speichertransistoren aus dem Potentialverlauf Gl. (4.18) wie folgt berechnen:

$$\psi_0 := V_s + 2\Psi_{FB} = \min(\psi_S(x_{min,i}, V_{th,i})) , \quad (4.46a)$$

$$0 = \psi_1^o z - \psi_2^o z^{-1} - \psi_3 \frac{l_c}{\lambda_s} z^{-\frac{l_c}{\lambda_s}} + \psi_4 \frac{l_c}{\lambda_d} z^{\frac{l_c}{\lambda_d}} , \quad (4.46b)$$

$$x_{min,i} = l_c \ln z_i .$$

Im allgemeinen Fall kann man dieses System nichtlinearer Gleichungen nur mit numerischen Methoden berechnen.

Im Weiteren soll eine Bezeichnung eingeführt werden, um die Schwellspannungen, die mit den zwei lokalen Minima assoziiert sind, einfacher beschreiben zu können. Dabei soll mit  $V_{th,s}$  diejenige Schwellspannung bezeichnet werden, die sich durch die Berechnung der Schwellspannung mit Hilfe des lokalen Minimums am sourceseitigen Kanalende ergibt. Verwendet man das drainseitige lokale Minimum, bezeichnet man die resultierende Schwellspannung mit  $V_{th,d}$ . Es gilt:

$$V_{th} = \max(V_{th,s}, V_{th,d}) . \quad (4.47)$$

#### 4.3.1 Programmierung des Langkanaltransistors

Um verhältnismäßig einfache Zusammenhänge zu erhalten, soll zuerst das Gleichungssystem Gl. (4.46) für den Langkanalfall mit  $L \rightarrow \infty$  betrachtet werden.

Wie schon im vorhergehenden Abschnitt nachgewiesen wurde, sind die Source- und Drainkontakte und auch die source- und drainseitigen Injektionen unter Langkanalbedingungen soweit räumlich voneinander getrennt, dass keine gegenseitigen Beeinflussungen stattfinden. Die Potentialverläufe an den jeweiligen Kanalenden und die zugeordneten Schwellspannungen  $V_{th,s}$  bzw.  $V_{th,d}$  können damit unabhängig voneinander mit Gl. (4.20) bzw. Gl. (4.21) berechnet werden und die beiden Teilergebnisse mit Hilfe von Gl. (4.47) zur Ermittlung von  $V_{th}$  kombiniert werden.

Aus dem allgemeinen Gleichungssystem Gl. (4.46) folgt daher für den Langkanalfall bei Betrachtung der Sourceseite:

$$\psi_0 = (V_s + V_{bi} - \psi_{S,L}(V_{th,s}) - \psi_3) z^{-1} + \psi_3 z^{-\frac{l_c}{\lambda_s}} + \psi_{S,L}(V_{th,s}) , \quad (4.48a)$$

$$0 = (V_s + V_{bi} - \psi_{S,L}(V_{th,s}) - \psi_3) z^{-1} + \psi_3 \frac{l_c}{\lambda_s} z^{-\frac{l_c}{\lambda_s}} , \quad (4.48b)$$

$$x_{min} = l_c \ln z .$$

Für die Drainseite ergibt sich entsprechend:

$$\psi_0 = (V_d + V_{bi} - \psi_{S,L}(V_{th,d}) - \psi'_4) z^{-1} + \psi'_4 z^{-\frac{l_c}{\lambda_d}} + \psi_{S,L}(V_{th,d}), \quad (4.49a)$$

$$0 = (V_d + V_{bi} - \psi_{S,L}(V_{th,d}) - \psi'_4) z^{-1} + \psi'_4 \frac{l_c}{\lambda_d} z^{-\frac{l_c}{\lambda_d}}, \quad (4.49b)$$

$$x'_{min} = l_c \ln z.$$

Wie schon festgestellt, können diese Gleichungen nur mit numerischen Mitteln gelöst werden. Um jedoch eine graphische Darstellung zum Beispiel der Funktion  $V_{th,s} = f(\Delta V_{fb,s})$  zu ermöglichen, kann man aus Gl. (4.48) eine parametrische Darstellung gewinnen, indem man  $z$  als Parameter auffasst. Dann besitzt diese Gleichung neben dem Parameter  $z$  nur die Variablen  $V_{th,s}$  und  $\Delta V_{fb,s}$ , die zueinander in einem linearen Zusammenhang stehen. Das Gleichungssystem kann deshalb leicht nach den beiden Variablen aufgelöst werden und man erhält mit  $\Delta V_{th,s} := V_{th,s} - V'_{th0,L}$  die Parameterdarstellung:

$$\Delta V_{fb,s}(z) = \frac{l_c^2 - \lambda_s^2}{\lambda_s^2} (V_s + V_{bi} - \psi_0) \frac{1}{f(z)} \quad (4.50a)$$

$$\Delta V_{th,s}(z) = n_0 \frac{l_c - \lambda_s}{\lambda_s} (V_s + V_{bi} - \psi_0) \frac{z^{-\frac{l_c}{\lambda_s}}}{f(z)} \quad (4.50b)$$

$$f(z) = 1 + \frac{l_c - \lambda_s}{\lambda_s} z^{-\frac{l_c}{\lambda_s}} - \frac{l_c}{\lambda_s} z^{\frac{\lambda_s - l_c}{\lambda_s}} \quad 1 \leq z < \infty.$$

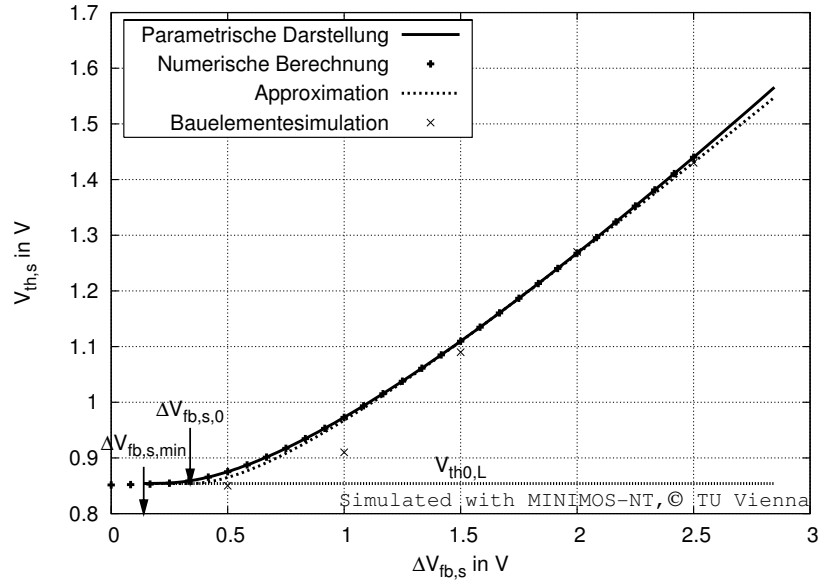
Bildet sich bei steigender Injektion ein lokales Minimum aus, hat Gl. (4.48) eine Lösung in  $0 \leq x_{min} < \infty$ . Damit ist der Wertebereich des Parameters  $z$  auf  $1 \leq z < \infty$  beschränkt. Das minimale Injektionsniveau, ab dem sich ein Minimum ausbildet, entspricht dem Punkt  $x_{min} \rightarrow \infty$ . An diesem Punkt ist  $\Delta V_{th,s} \rightarrow 0$ . Das zugehörige minimale Injektionsniveau, ab dem die Ausbildung des Minimums stattfindet und die Schwellspannungsverschiebung  $\Delta V_{th,s} > 0$  ist, ist also:

$$\Delta V_{fb,s,min} = \Delta V_{fb,s}(z \rightarrow \infty) = \frac{l_c^2 - \lambda_s^2}{\lambda_s^2} (V_s + V_{bi} - \psi_0). \quad (4.51)$$

Für  $x_{min} \rightarrow 0$  gehen  $\Delta V_{fb,s}$  und  $\Delta V_{th,s} \rightarrow \infty$ .

In Abb. 4.15 wird die parametrische Darstellung gezeigt und den Ergebnissen der numerischen Lösung des nichtlinearen Gleichungssystems Gl. (4.46) gegenübergestellt. Man erkennt, dass die parametrische Darstellung nur den Verlauf der Schwellspannung für  $\Delta V_{fb,s} > \Delta V_{fb,s,min} \Leftrightarrow \Delta V_{th,s} > 0$  beschreibt. Im Bereich  $\Delta V_{fb,s} \leq \Delta V_{fb,s,min}$  ist  $V_{th,s} = V'_{th0,L}$ .

Um eine näherungsweise Berechnung der Schwellspannung für den Langkanalfall zu ermöglichen, muß zunächst Gl. (4.28) vereinfacht werden. Dabei erweist es sich als zweckmäßig, die Exponentialfunktion derart durch eine quadratische Gleichung zu approximieren, dass beide Funktionen an einer Stelle  $x_0$  den gleichen Funktionswert und



**Abbildung 4.15:** Schwellspannung eines ONO-Speichertransistors mit  $L = 0,5 \mu\text{m}$ ; Vergleich der numerischen Lösung von Gl. (4.46) mit der parametrischen Darstellung Gl. (4.50), der Approximation Gl. (4.54) und einer numerischen Bauelementesimulation

den gleichen Anstieg haben und dass der Scheitelpunkt der Parabel auf der x-Achse liegt. Es ergibt sich:

$$\exp(-x) \approx \frac{\exp(-x_0)}{4} (2 + x_0 - x)^2 \quad x \leq x_0 + 2. \quad (4.52)$$

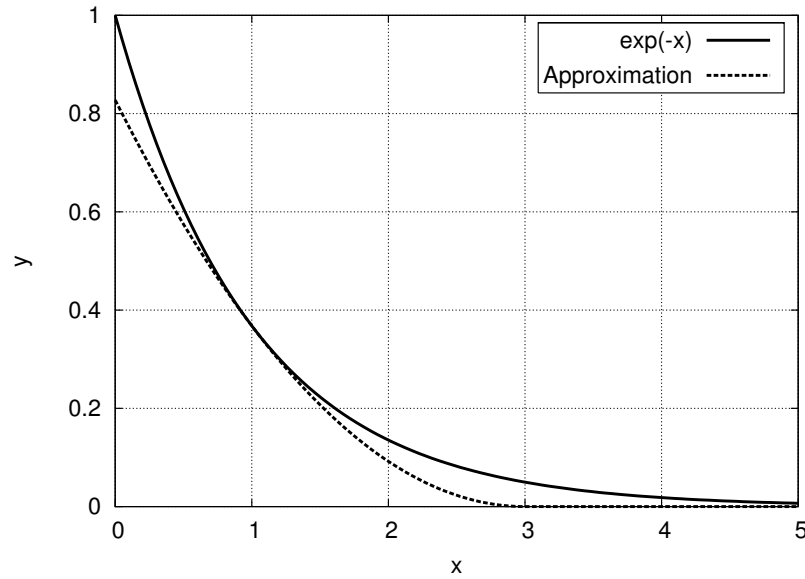
Beide Funktionen sind in Abb. 4.16 exemplarisch für  $x_0 = 1$  dargestellt. Wie gezeigt, wird für die Approximation nur der linke Parabelast der quadratischen Funktion, für den  $x \leq x_0 + 2$  gilt, verwendet.

Diese Näherung kann nun mit  $x \rightarrow x_{\min}/\lambda_s$  auf Gl. (4.28) angewendet werden und dabei als Stützstelle  $x_0 = x_{\min}^{\Delta}/\lambda_s$  gewählt werden. Anschließend wird dann die lineare Näherung für  $x_{\min}(\psi_{S,L})$  aus Gl. (4.27) eingesetzt und man erhält den Zusammenhang:

$$\psi_0 = \underbrace{\psi_{S,L} - \frac{\Delta V_{fb,s}}{4} \frac{\lambda_s}{l_c + \lambda_s} \exp\left(-x_{\min}^{\Delta}/\lambda_s\right) \left(2 + \frac{l_c}{\lambda_s} \frac{l_c + \lambda_s}{\lambda_s} \frac{V_s + V_{bi} - \psi_{S,L}}{\Delta V_{fb,s}}\right)^2}_{\approx \Delta\psi_{S,L}}. \quad (4.53)$$

Mit dem aus Gl. (4.18) gefolgerten Zusammenhang  $\psi_{S,L}(V_{th}) = (V_{th} - V_{th0,L})/n_0 + \psi_0$  kann diese quadratische Gleichung nach  $\Delta V_{th} := V_{th} - V_{th0,L}$  aufgelöst werden und man





**Abbildung 4.16:** Approximation der Funktion  $y = \exp(-x)$  durch eine quadratische Funktion mit der Stützstelle  $x_0 = 1$

erhält:

$$\frac{\Delta V_{th,s}}{n_0} \approx \begin{cases} -2a(\Delta V_{fb,s} - \Delta V_{fb,s,0}) + \frac{ab}{2}\Delta V_{fb,s} - \\ \quad - \sqrt{\frac{a^2 b^2}{4}\Delta V_{fb,s}^2 - 2a^2 b \Delta V_{fb,s}(\Delta V_{fb,s} - \Delta V_{fb,s,0})} \\ 0 \quad (\Delta V_{fb,s} < \Delta V_{fb,s,0}) \end{cases} \quad (4.54)$$

$$a = \frac{\lambda_s}{l_c} \frac{\lambda_s}{l_c + \lambda_s}; \quad b = 4 \frac{\lambda_s}{l_c} \exp\left(x_{min}^\Delta / \lambda_s\right); \quad \Delta V_{fb,s,0} = \frac{V_s + V_{bi} - \psi_0}{2a}$$

In Abb. 4.15 ist das Ergebnis dieser Näherung der numerischen Lösung der Schwellspannungsgleichung Gl. (4.46) gegenübergestellt. Nur im Bereich der sehr schwachen und der sehr starken Programmierung sind leichte Abweichungen zu erkennen. Die Näherungslösung liefert also sehr gut mit den exakten Lösungen übereinstimmende Resultate.

### 4.3.2 Nachbarbitunterdrückung bei Langkanaltransistoren

Nun soll untersucht werden, welche Mechanismen für die Nachbarbitunterdrückung verantwortlich sind und unter welchen Bedingungen beim Auslesen eines Bits die Information des Nachbarbits maskiert werden kann. Diese Unterdrückung ist notwendig, um die

source- bzw. drainseitigen Informationen voneinander trennen zu können und somit eine Zweibitspeicherung zu ermöglichen.

Analog zur Behandlung des sourceseitigen Kanalendes kann auch für den Drainbereich eine Parameterdarstellung zur Berechnung von  $V_{th,d}$  angegeben werden:

$$\Delta V_{fb,d}(z) = \frac{l_c^2 - \lambda_d^2}{\lambda_d^2} (V_d + V_{bi} - \psi_0) \frac{1}{f(z)} \quad (4.55a)$$

$$\Delta V_{th,d}(z) = n_0 \frac{l_c - \lambda_d}{\lambda_d} (V_d + V_{bi} - \psi_0) \frac{z^{-\frac{l_c}{\lambda_s}}}{f(z)} \quad (4.55b)$$

$$f(z) = 1 + \frac{l_c - \lambda_s}{\lambda_s} z^{-\frac{l_c}{\lambda_s}} - \frac{l_c}{\lambda_s} z^{\frac{\lambda_s - l_c}{\lambda_s}} \quad 1 \leq z < \infty.$$

Entsprechend Gl. (4.54) kann näherungsweise der folgende Zusammenhang zur expliziten Berechnung von  $\Delta V_{fb,s}$  verwendet werden:

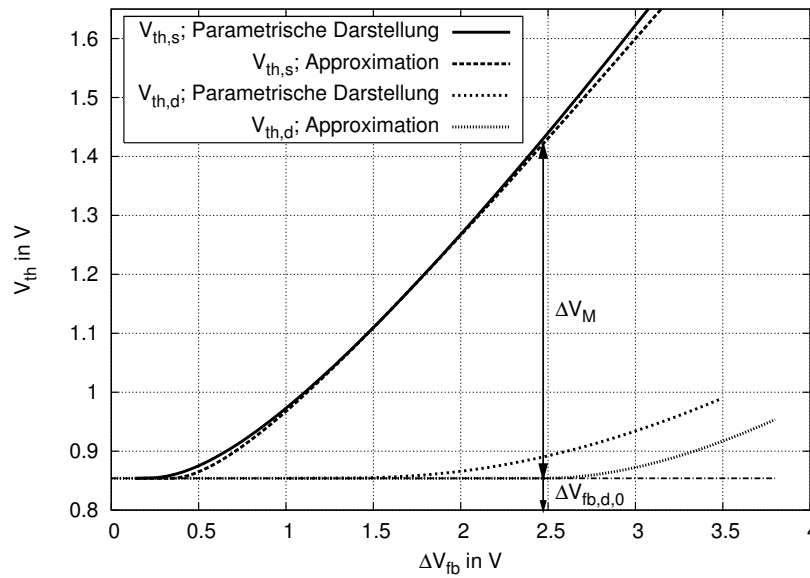
$$\frac{\Delta V_{th,d}}{n_0} \approx \begin{cases} -2a(\Delta V_{fb,d} - \Delta V_{fb,d,0}) + \frac{ab}{2} \Delta V_{fb,d} - \\ \quad - \sqrt{\frac{a^2 b^2}{4} \Delta V_{fb,d}^2 - 2a^2 b \Delta V_{fb,d} (\Delta V_{fb,d} - \Delta V_{fb,d,0})} \\ 0 \quad (\Delta V_{fb,d} < \Delta V_{fb,d,0}) \end{cases} \quad (4.56)$$

$$a = \frac{\lambda_d}{l_c} \frac{\lambda_d}{l_c + \lambda_d}; \quad b = 4 \frac{\lambda_d}{l_c} \exp\left(x_{min}^\Delta / \lambda_d\right); \quad \Delta V_{fb,d,0} = \frac{V_d + V_{bi} - \psi_0}{2a}$$

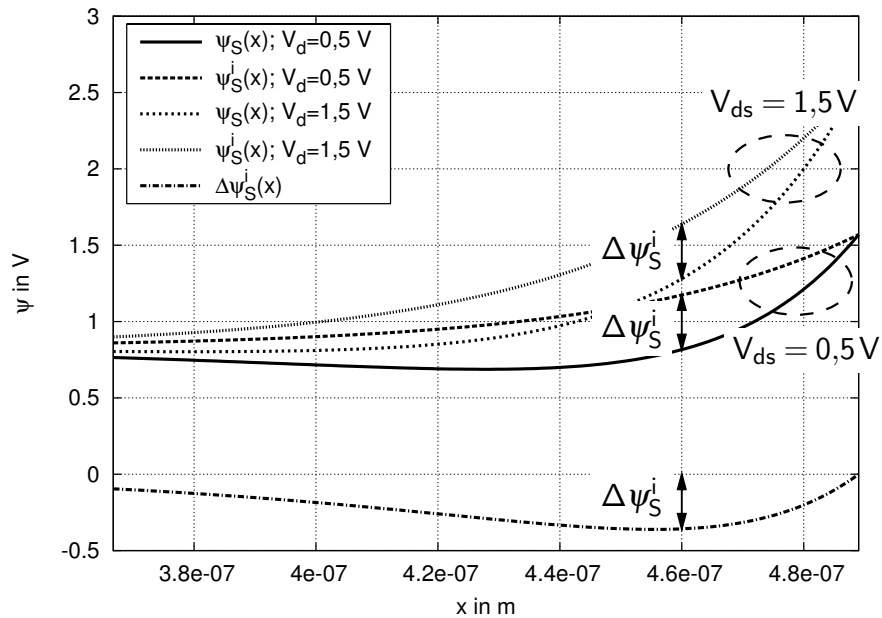
In Abb. 4.17 werden die Schwellspannungen  $V_{th,d}$  und  $V_{th,s}$  dargestellt. Zunächst sei darauf hingewiesen, dass  $V_{th,s}$  besser als  $V_{th,d}$  durch die Approximation beschrieben werden kann. Die Ursache dafür ist, dass sich bei hohen Drainspannungen das lokale Minimum in einem größeren Abstand vom Stützpunkt  $x_0 = x_{min}^\Delta / \lambda_d$  ausbildet und dass in diesem Bereich die Exponentialfunktion nur ungenau durch die quadratische Funktion approximiert wird.

Ein Vergleich der Kennlinien  $V_{th,d}$  und  $V_{th,s}$  der Abbildung, hier für  $\lambda_s = \lambda_d$ , lässt auch erkennen, dass die Kennlinie der drainseitigen Schwellspannung  $V_{th,d}$  zu höheren Injektionsniveaus  $\Delta V_{fb}$  verschoben ist und flacher verläuft. Das bedeutet, dass die am drainseitigen Ende des Kanals befindliche Ladung eine geringere Auswirkung auf die Schwellspannung des Transistors hat, als die sourceseitige Ladung. Aus diesem Grund ist auch die minimale Injektion, die notwendig ist, um am Drainende ein lokales Minimum im Potentialverlauf zu erzeugen, höher, als am sourceseitigen Kanalende:

$$\Delta V_{fb,d,min} = \frac{l_c^2 - \lambda_d^2}{\lambda_d^2} (V_d + V_{bi} - \psi_0) \geq \Delta V_{fb,s,min}. \quad (4.57)$$



**Abbildung 4.17:** Vergleich der Schwellspannungen  $V_{th,s}$  und  $V_{th,d}$  bei  $V_s = 0\text{V}$ ,  $V_d = 1,5\text{V}$  für einen ONO-Speichertransistor mit  $L = 0,5\mu\text{m}$ ; Eingetragen sind außerdem die Approximationen nach Gl. (4.54) und Gl. (4.56).



**Abbildung 4.18:** Verlauf des Oberflächenpotentials  $\psi_s$ , des initialen Potentials  $\psi_s^i$  und der Potentialverschiebung  $\Delta\psi_s^i$  bei verschiedenen Drainspannungen;  $\Delta V_{fb,d} = 2,5\text{V}$ ,  $V_s = 0\text{V}$

PSfrag replacements

Die Ursache dieser geringeren Empfindlichkeit kann anhand der Darstellung Abb. 4.18 erläutert werden. Gezeigt werden für verschiedene Drainspannungen jeweils der initiale Potentialverlauf  $\psi_S^i$ , also der Potentialverlauf ohne Programmierung, die von den Anschlusspotentialen unabhängige Potentialverschiebung durch die Programmierung  $\Delta\psi_S^i$  und der sich durch eine Überlagerung beider ergebende Gesamtpotentialverlauf  $\psi_S$ .

Bei entsprechenden Anschlussspannungen und Programmierung wird sich das Potentialminimum in der Nähe des Drainkontakts ausbilden. In diesem Gebiet ist das initiale Potential bei großen Drain-Source-Spannungen aber wesentlich höher als bei niedrigen Drainspannungen. Die Potentialverschiebung  $\Delta\psi_S^i$  hat damit bei der Überlagerung einen geringeren Einfluss auf den resultierenden Gesamtpotentialverlauf. Der resultierende Potentialverlauf wird also bei steigenden Drainspannungen immer weniger von der drainseitigen Programmierung beeinflusst. Infolgedessen kann es sogar soweit kommen, dass bei hohen Drain-Source-Spannungen an der Drainseite gar kein lokales Minimum im Potentialverlauf ausgebildet wird. Damit hat die drainseitige Programmierung keinen Einfluss mehr auf die Schwellspannung, so dass die Drainprogrammierung ideal unterdrückt wird.

Diese Maskierung der drainseitigen Programmierung durch eine Drainspannung ausreichender Höhe wird bei den Speichertransistoren mit lokaler Ladungsspeicherung angewendet, um an beiden Kanalenden Informationen in Form von injizierten Elektronen zu speichern. Beim Auslesen der Information an einem Kanalende wird das jeweils andere Kanalende auf ein entsprechend hohes Potential gesetzt, um eine Beeinflussung des Leseergebnisses durch die Programmierung des Nachbarbits zu vermeiden.

Aus Gl. (4.55) kann eine Parameterdarstellung der Abhängigkeit der drainseitigen Schwellspannungsdifferenz  $\Delta V_{th,d}(V_d)$  vom Drainpotential gewonnen werden. Mit dieser Gleichung kann unter anderem ermittelt werden, wie stark beim Auslesen der sourceseitigen Schwellspannung die drainseitige Programmierung in Abhängigkeit vom Drainpotential unterdrückt wird:

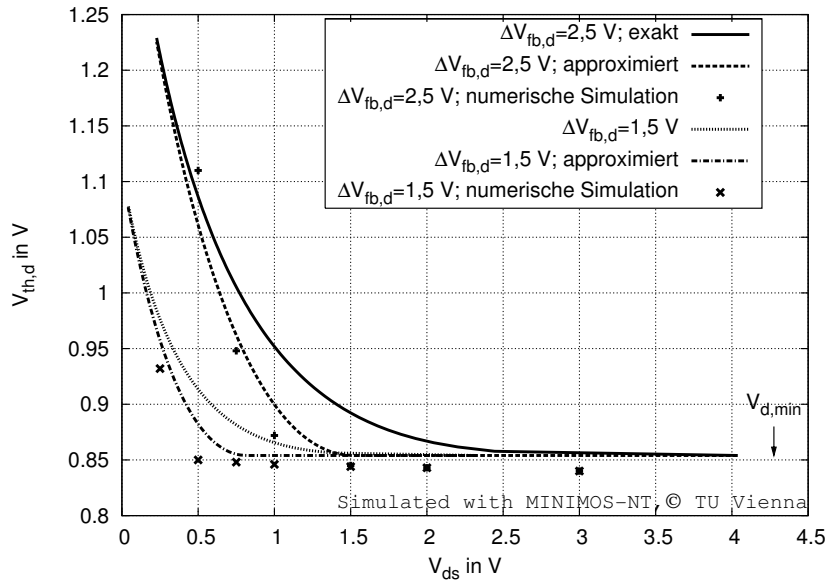
$$V_d(z) = \frac{\lambda_d^2}{l_c^2 - \lambda_d^2} \Delta V_{fb,d} f(z) + \psi_0 - V_{bi} \quad (4.58a)$$

$$\Delta V_{th,d}(z) = n_0 \frac{\lambda_d}{l_c + \lambda_d} \Delta V_{fb,d} z^{-\frac{l_c}{\lambda_s}}. \quad (4.58b)$$

In Abb. 4.19 wird dieser parametrische Funktionsverlauf und die explizite Darstellung in der Näherung nach Gl. (4.54) gezeigt. In dieser Darstellung ist zu erkennen, dass der Einfluss der drainseitigen Programmierung auf die Schwellspannung des Transistors mit zunehmender Drainspannung absinkt. Ab einer Drainspannung von:

$$V_d \geq V_{d,min} = \frac{\lambda_d^2}{l_c^2 - \lambda_d^2} \Delta V_{fb,d} \quad (l_c > \lambda_d) \quad (4.59)$$

bildet sich am Drainende kein lokales Minimum mehr aus und die Schwellspannung des Transistors ist von der drainseitigen Programmierung unbeeinflusst. In diesem Fall



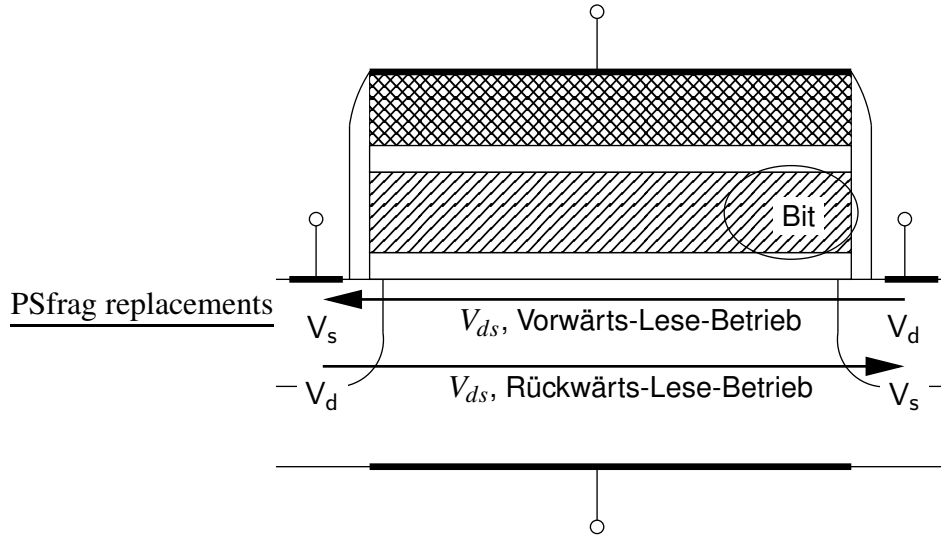
**Abbildung 4.19:** Abhängigkeit der Schwellspannung  $V_{th,d}$  von der Drain-Source-Spannung für verschiedene Injektionsniveaus und Approximation nach Gl. (4.54)

stellt sich die initiale Langkanalschwellspannung  $V_{th0,L}$  ein und die Programmierung der Drainseite wird folglich ideal unterdrückt. Für den Fall  $l_c \leq \lambda_d$  bildet sich für jede Drainspannung ein lokales Minimum aus.

In Abb. 4.19 erkennt man aber auch, dass schon bei wesentlich geringeren Drainspannungen als  $V_{d,min}$  die Abhängigkeit der Schwellspannung von der Draininjektion sehr stark reduziert ist und in praktischen Anwendungsfällen vernachlässigt werden kann.

Es sei nochmals darauf hingewiesen, dass diese Beeinflussung der Schwellspannung von der drainseitigen Programmierung auch bei ONO-Speichertransistoren mit  $L \rightarrow \infty$  auftritt. Damit ist sogar bei sehr langen Transistoren, bei denen sich die source- und drainseitigen Injektionen gegenseitig nicht direkt beeinflussen, eine minimale Drainspannung notwendig, um eine ausreichende Bittrennung beim Auslesen der Informationen zu gewährleisten.

Zur Charakterisierung der Bittrennung von Langkanal-ONO-Transistoren wird in [LHW<sup>+</sup>06] folgende Methode vorgeschlagen, deren praktische Relevanz noch in Abschnitt 5.1 aufgezeigt wird. Dafür wird in einem Speichertransistor nur eine Seite auf einen festen Wert programmiert, während die andere Seite im unprogrammierten, initialen Zustand verbleibt. Als *Rückwärts-Lese-Betrieb* (*Reverse-Read-Mode*) wird diejenige Betriebsart des Transistors bezeichnet, in der die injizierte Ladung am sourceseitigen Kanalende liegt. Die zugehörige Schwellspannung sei im Weiteren mit  $V_{th,rev}$  bezeichnet. Befindet sich die Ladung jedoch am drainseitigen Kanalende, bezeichnet man die Betriebsart als *Vorwärts-Lese-Betrieb* mit der zugehörigen Schwellspannung  $V_{th,fwd}$ . Damit



**Abbildung 4.20:** Rückwärts- und Vorwärts-Lese-Betrieb

ergeben sich für diese beiden Schwellspannungen folgende Definitionen:

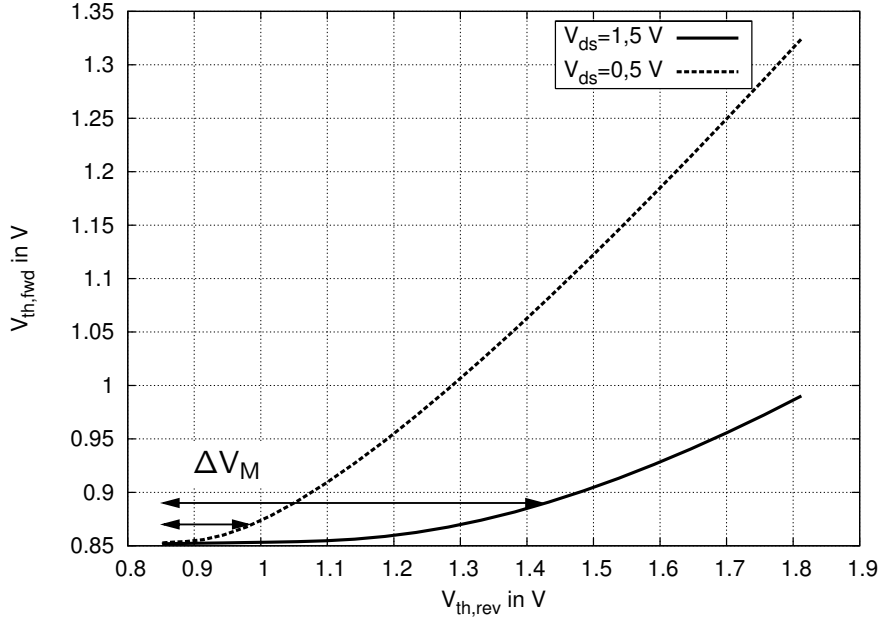
$$V_{th,rev} = V_{th}(\Delta V_{fb,s} = \Delta V_{fb}, \Delta V_{fb,d} = 0), \quad (4.60)$$

$$V_{th,fwd} = V_{th}(\Delta V_{fb,d} = \Delta V_{fb}, \Delta V_{fb,s} = 0). \quad (4.61)$$

Bei einem ONO-Transistor, in dem nur eine Seite programmiert wurde, kann also von der Vorwärts- in die Rückwärts-Lese-Betriebsart und umgekehrt übergegangen werden, indem die Source- und Drain-Anschlüsse ihre Rollen vertauschen, insbesondere also dadurch, dass die Anschlusspotentiale wechselseitig getauscht werden. Diese Umpolung der Anschlussspannung findet beim Auslesen der ONO-Speichertransistoren in Virtual-Ground-Arrays statt. Damit kennzeichnen beide Größen die Verschiebung der Schwellspannung eines unprogrammierten Bits durch die Programmierung des Nachbarbits, den sogenannten Nachbarbiteffekt.

Zur Illustration wird meist, wie in Abb. 4.21 gezeigt,  $V_{th,fwd}(\Delta V_{fb})$  über  $V_{th,rev}(\Delta V_{fb})$  aufgetragen. Typischerweise bestehen diese Kennlinien aus zwei Abschnitten. Der erste Abschnitt ist dadurch gekennzeichnet, dass die Vorwärts-Schwellspannung unabhängig von der Programmierung bei  $V_{th0,L}$  liegt. Steigt das Programmierniveau  $\Delta V_{fb}$  und damit die Rückwärts-Schwellspannung weiter an, erhöht sich aufgrund des Nachbarbiteffekts auch  $V_{th,fwd}$ . Wie schon erwähnt, ist der Übergang zwischen den beiden Regionen durch die beginnende Ausbildung des lokalen Minimums an der Drainseite gegeben und ist damit abhängig von der Drainspannung.

Die maximale Rückwärts-Schwellspannung, die erreicht werden kann, ohne dass sich die Vorwärts-Schwellspannung ändert, wird in [LHW<sup>+</sup>06] als Kenngröße der Bittrennung eingeführt und ist in Abb. 4.21 als  $\Delta V_M$  gekennzeichnet. Im strengen mathematischen Sinn ist diese Grenze bei  $\Delta V_{fb,min}$  gemäß Gl. (4.57) erreicht. Wie in der Diskus-



PSfrag replacements

**Abbildung 4.21:** Schwellspannung der Vorwärts-Lese-Betriebsart aufgetragen über die korrespondierende Rückwärts-Lese-Schwellspannung zur Illustration des Nachbarbiteffekts

sion von Abb. 4.18 jedoch festgestellt wurde, ist in diesem Bereich die Sensitivität der Schwellspannung auf Änderungen von  $V_d$  so gering, dass die für das Erreichen einer vorgegebenen Bittrennung notwendige Drainspannung für praktische Anwendungsfälle weit überschätzt werden würde. Deshalb wird im Folgenden für die Berechnung von  $\Delta V_M$  die Näherungsformel Gl. (4.56) und damit  $\Delta V_{fb,d,0}$  statt  $\Delta V_{fb,d,min}$  verwendet. Die Definition von  $\Delta V_M$  lautet somit:

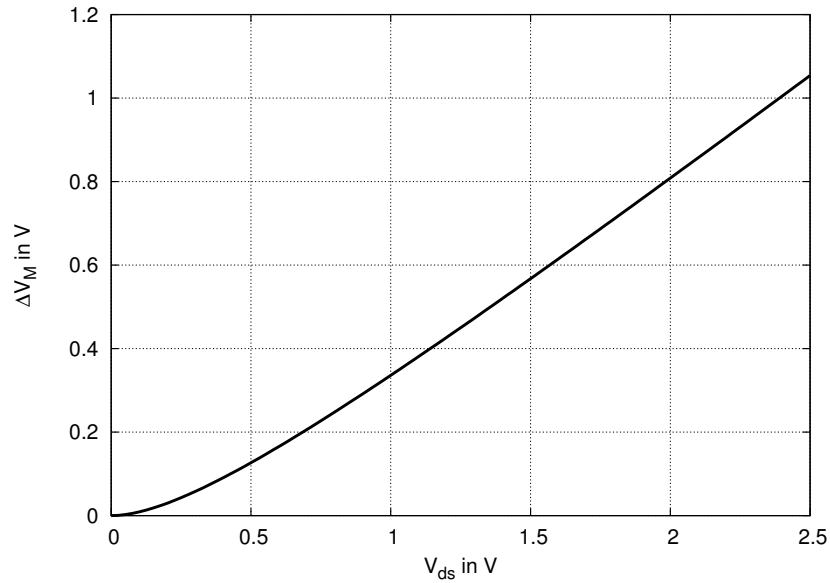
$$\Delta V_M := V_{th,s}(\Delta V_{fb,s} = \Delta V_{fb,d,0}) - V'_{th0,L}. \quad (4.62)$$

Das Vorgehen zur Berechnung von  $\Delta V_M$  kann anhand der Abb. 4.17 visualisiert werden. Zunächst wird  $\Delta V_{fb,d,0}$  an der Kennlinie  $V_{th,d} = f(\Delta V_{fb,d})$  ermittelt. Dieser Wert wird dann verwendet, um an der entsprechenden Kennlinie  $V_{th,s}(\Delta V_{fb,d,0})$  zu berechnen. Die gesuchte Kenngröße  $\Delta V_M$  ist der Abstand dieses Punkts zur Schwellspannung des unprogrammierten Transistors.

Unter Verwendung von Gl. (4.56) erhält man also:

$$\Delta V_{fb,d,0} = \frac{l_c(l_c + \lambda_d)}{2\lambda_d^2} (V_{ds} + V_{bi} - 2\Psi_{FB}). \quad (4.63)$$

Eingesetzt in Gl. (4.56) mit  $\lambda_s = \lambda_d = \lambda$  und unter Beachtung der Definition Gl. (4.62)



**Abbildung 4.22:** Parameter der Bittrennung  $\Delta V_M$  eines Langkanaltransistors; Bis auf einen geringen Offset ist  $\Delta V_M$  in guter Näherung proportional zu der anliegenden Drainspannung.

erhält man nach wenigen Umformungen:

$$\begin{aligned} \frac{\Delta V_M}{n_0} = & -V_{ds} + \frac{b}{4} (V_{ds} + V_{bi} - 2\Psi_{FB}) - \\ & - \sqrt{\frac{b^2}{16} (V_{ds} + V_{bi} - 2\Psi_{FB})^2 - \frac{b}{2} (V_{ds} + V_{bi} - 2\Psi_{FB}) V_{ds}}. \end{aligned} \quad (4.64)$$

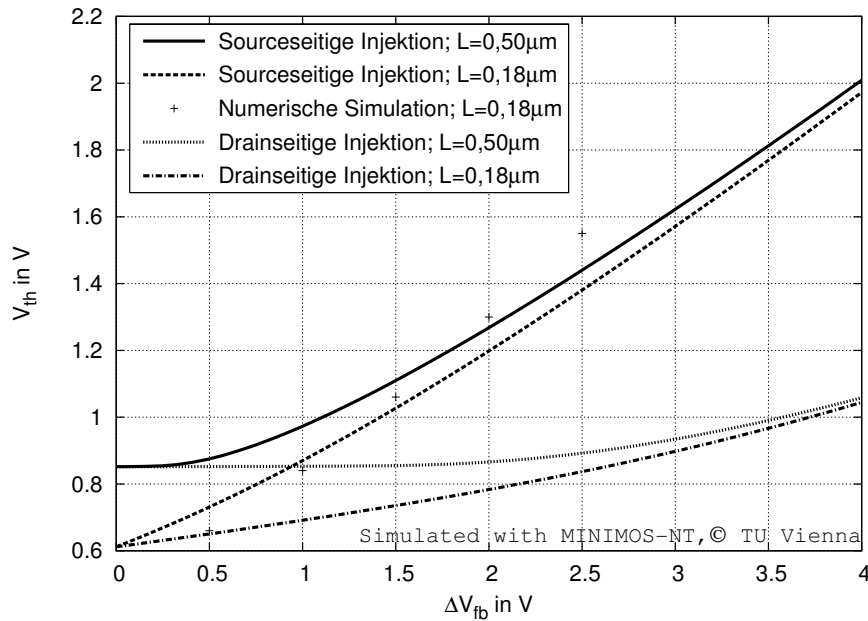
$$b = 4 \frac{\lambda}{l_c} \exp(x_{min}^\Delta / \lambda)$$

Der Verlauf von  $\Delta V_M$  wird in Abb. 4.22 gezeigt. Aus dieser Darstellung kann gefolgert werden, dass für praktisch relevante Drainspannungen bei Langkanaltransistoren die Bittrennung  $\Delta V_M$  bis auf einen geringen Offset in guter Näherung proportional zu der anliegenden Drainspannung wächst.

### 4.3.3 Programmierung eines Kurzkanaltransistors

Bei allen nachfolgenden Untersuchungen der Kurzkanaltransistoren wird auf die Verwendung von Parameterdarstellungen verzichtet. Obwohl sich auch für Kurzkanal-ONO-Transistoren diese Gleichungen herleiten lassen, sind sie jedoch im Allgemeinen so kompliziert, dass sie kaum für eine übersichtliche Diskussion der Effekte verwendet werden können. Statt dessen wird Gl. (4.46) direkt mittels numerischer Verfahren gelöst.





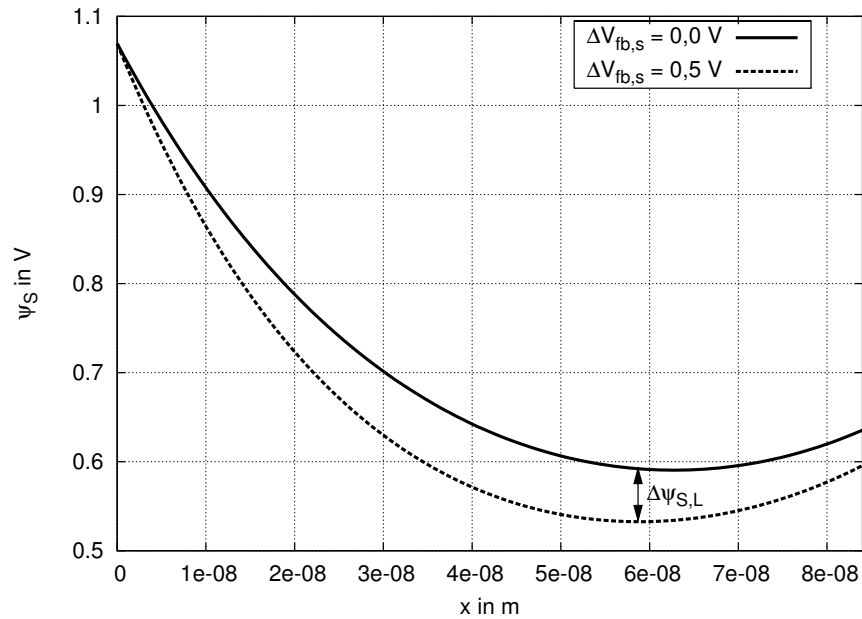
**Abbildung 4.23:** Verschiebung der Schwellspannung durch eine source- bzw. drainseitige Programmierung im Vergleich für Lang- und Kurzkanaltransistoren

Für die Beurteilung der Unterschiede im Programmierverhalten von Lang- und Kurzkanaltransistoren wird in Abb. 4.23 die Schwellspannung in Abhängigkeit von source- bzw. drainseitiger Injektion für Transistoren unterschiedlicher Längen nach numerischer Berechnung mittels Gl. (4.46) dargestellt.

Der augenfälligste Unterschied ist, dass bei Kurzkanaltransistoren auch bei geringerer Injektion schon eine deutliche Verschiebung der Schwellspannung stattfindet. Der für Langkanaltransistoren beobachtete Effekt, dass erst ab dem Injektionsniveau  $\Delta V_{fb,min}$  die Transistorschwellspannung beeinflusst wird, ist bei Transistoren mit kürzerer Kanallänge weniger stark ausgeprägt. Aus diesem Grund ist auch die Krümmung der Kennlinien im Bereich geringer Schwellspannungen bei Kurzkanaltransistoren weniger stark ausgeprägt.

Für eine hohe Injektion, also eine starke Programmierung, nähern sich die Kennlinien der Lang- und Kurzkanaltransistoren einander an. Die Ursache dafür ist, dass der Potentialverlauf in der Umgebung des Potentialminimums von der hohen Injektion dominiert wird. Kurzkanal- und DIBL-Effekte haben dementsprechend einen geringeren Einfluss.

Die Ursache der Verschiebung der Schwellspannung bei geringen Injektionsniveaus kann anhand der Abb. 4.24 erläutert werden, die den Potentialverlauf eines Kurzkanaltransistors in der Umgebung des Sourcekontakts für verschiedene Injektionsniveaus zeigt. Wie in Abschnitt 3.3.1 begründet wurde, ist bei Standard- und ONO-Kurzkanaltransistoren die Lage des Minimums und damit die Schwellspannung von dem Verlauf des Potentials in der Umgebung von Source- und Drainkontakten abhängig. Dieser Ver-



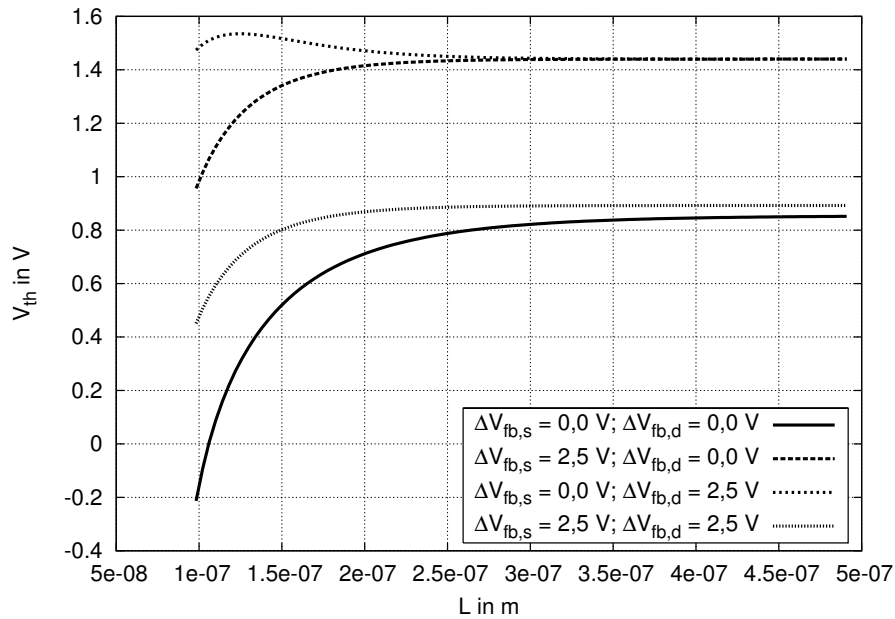
**Abbildung 4.24:** Ausschnitt aus dem Potentialverlauf eines Kurzkanaltransistors mit  $L = 0,18 \mu\text{m}$  im initialen Zustand und nach schwacher Programmierung

lauf wird selbst bei schwacher Programmierung merklich verändert. Daraus ergibt sich direkt, dass auch die Schwellspannung schon bei geringen Injektionsniveaus beeinflusst wird. Bei Langkanaltransistoren hingegen ist die Lage des Minimums bei geringem Injektionsniveau vom Potentialverlauf in der Umgebung der Kontakte unabhängig. Daher verändert eine geringfügige Modulation dieses Verlaufs die Schwellspannung nicht.

#### 4.3.4 Kurzkanalverhalten

Wie beim Standard-MOS-Transistor soll hier die Auswirkung der Verringerung der Kanallänge bei konstanten Technologieparametern untersucht werden. Nicht betrachtet werden hier die Veränderungen bei einer Skalierung der Technologiegeneration, da diese die Transistorparameter stark verändert.

Die Lösung von Gl. (4.46) liefert die Darstellung der Abb. 4.25. Wie bei Standard-MOS-Transistoren erkennt man den in Abschnitt 3.3.1 diskutierten Abfall der Schwellspannung bei verringerter Kanallänge. Auffällig ist, dass bei dem unprogrammierten Transistor mit  $\Delta V_{fb,s} = \Delta V_{fb,d} = 0$  der Schwellspannungsabfall schon bei größeren Transistorlängen beginnt. Die Ursache dessen liegt darin begründet, dass bei programmierten Transistoren der Potentialverlauf auch durch die injizierten Elektronen beeinflusst wird. Die Modulation des Potentialverlaufs durch die Source- und Drain-Potentiale, Ursache des Kurzkanaleffekts, hat im Vergleich dazu eine geringere Auswirkung auf den Gesamt-



**Abbildung 4.25:** Änderung der Schwellspannung in Abhängigkeit von der Kanallänge für verschiedene Programmierzustände;  $V_{ds} = 1,5 \text{ V}$

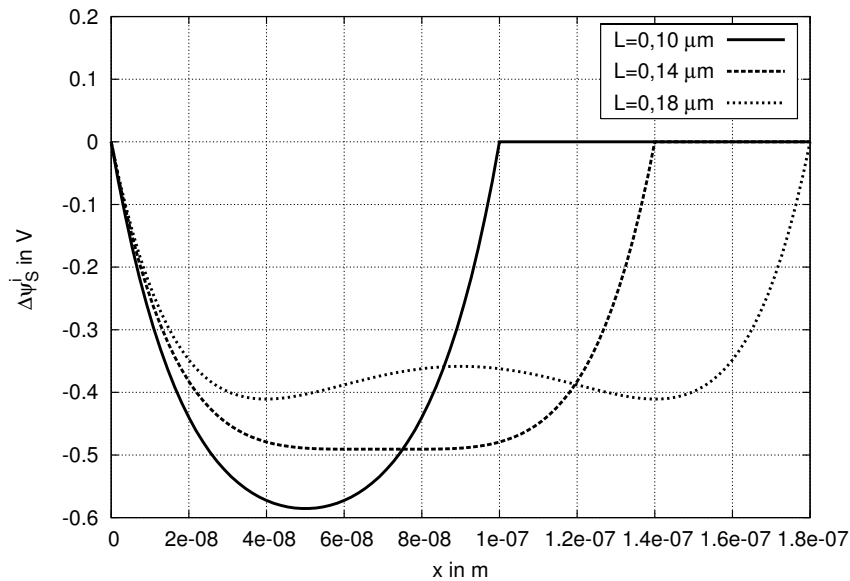
verlauf. Damit wird der Kurzkanaleffekt erst bei geringeren Kanallängen wirksam.

Im Gegensatz zu Standard-MOS-Transistoren erkennt man jedoch im Schwellspannungsverlauf für  $\Delta V_{fb,s} = \Delta V_{fb,d} = 2,5 \text{ V}$  im Bereich von  $L = 100 \dots 200 \text{ nm}$  einen leichten Anstieg der Schwellspannung. Zur Untersuchung dieses Effekts wurde in Abb. 4.26 die Verschiebung des initialen Potentialverlaufs  $\Delta \psi_s^i$  durch die Programmierung des Transistors, wie sie in Gl. (4.31) definiert wurde, dargestellt. Offensichtlich überlagern sich bei kurzen ONO-Transistoren mit beidseitiger Programmierung die source- und drainseitigen Potentialverschiebungen derart, dass der Betrag des Maximalwerts der Verschiebung ansteigt, wodurch die Schwellspannung erhöht wird. Dieser wechselseitige Durchgriff der beiden Injektionszonen bei kurzen Transistoren ist von der Drainspannung unabhängig und ist damit eine fundamentale Grenze für die Nachbarbittrennung.

Der Effekt des Anstiegs der Schwellspannung durch die erhöhte Potentialverschiebung und die Absenkung der Schwellspannung durch den Kurzkanaleffekt und das Drain Induced Barrier Lowering (DIBL) überlagern sich derart, dass sich je nach Stärke der Einzelkomponenten ein effektiver Anstieg oder Abfall der Schwellspannung bei veränderter Kanallänge ergeben kann.

#### 4.3.5 Nachbarbiteffekt und Bittrennung bei Kurzkanaltransistoren

Zur Untersuchung der Bittrennung bei Kurzkanaltransistoren soll zunächst wie bei Langkanaltransistoren die Veränderung der Schwellspannung durch eine drainseitige Pro-



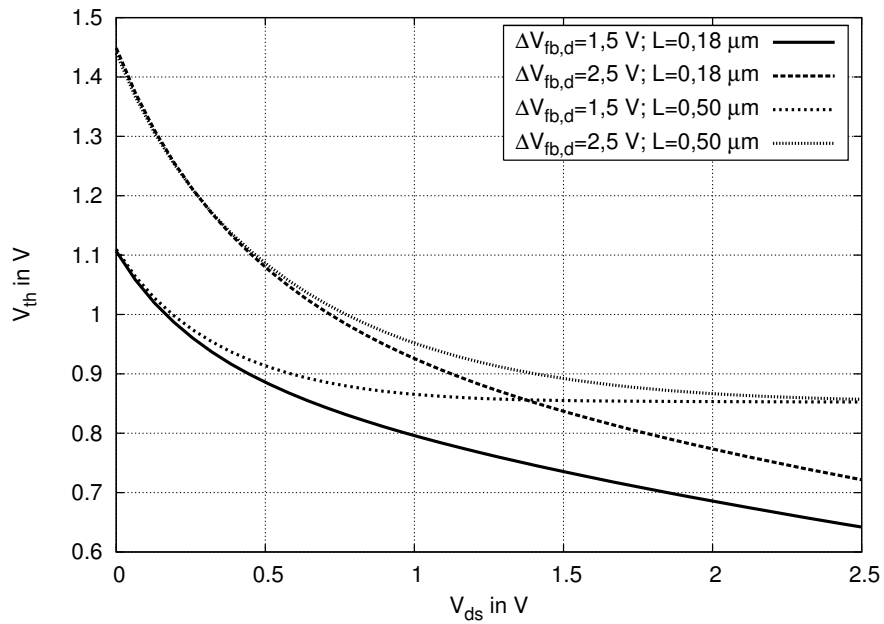
**Abbildung 4.26:** Verlauf der in Gl. (4.31) definierten Verschiebung des initialen Potentialverlaufs  $\Delta\psi_S^i$  in Folge der Programmierung für Transistoren mit unterschiedlichen Kanallängen;  $\Delta V_{fb,s} = \Delta V_{fb,d} = 2,5$  V

grammierung in Abhängigkeit von der Drain-Source-Spannung untersucht werden. Für diesen Zweck wurde in Abb. 4.27 der Verlauf der Schwellspannung  $V_{th}$  jeweils für einen Lang- und einen Kurzkanaltransistor für unterschiedliche Injektionsniveaus der Drainseite dargestellt.

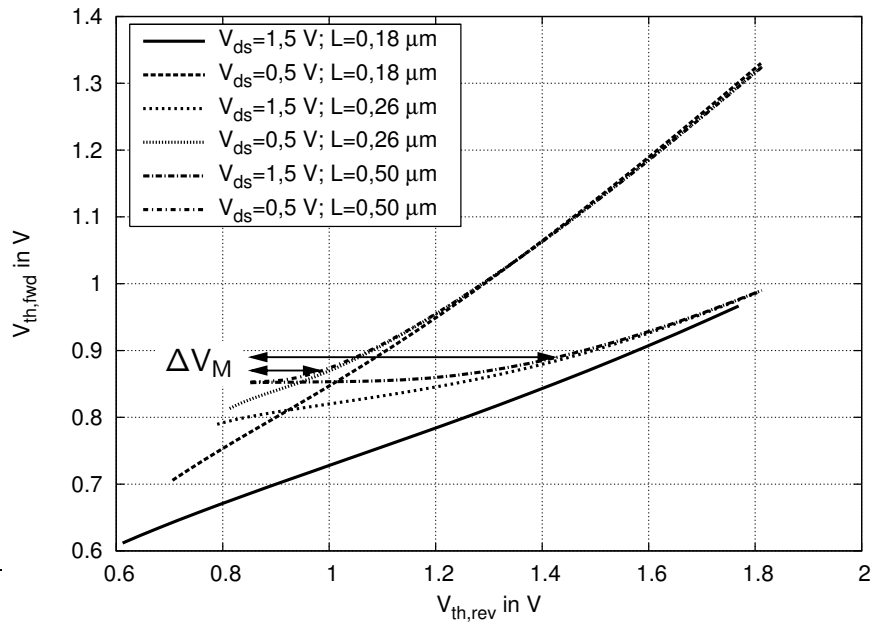
Man erkennt, dass wie beim Langkanaltransistor die Schwellspannung des Kurzkanaltransistors mit steigender Drainspannung sinkt. Im Gegensatz jedoch zum Langkanaltransistor, bei dem bei großen Drainspannungen eine konstante Schwellspannung von  $V'_{th0,L}$  erreicht wird, sinkt bei dem Kurzkanaltransistor die Schwellspannung aufgrund des Drain Induced Barrier Lowering (DIBL) mit steigender Drainspannung annähernd linear weiter und es stellt sich deshalb kein konstanter Schwellspannungswert ein. Im Bereich des DIBLs ist aber die Abfallrate der Schwellspannung geringer als bei  $V_{ds} = 0$ , so dass Kurzkanaltransistoren zweckmäßigerweise in diesem Bereich betrieben werden sollten. Die Grenze zwischen beiden Bereichen liegt ungefähr bei der gleichen Drainspannung, bei der auch die Schwellspannungskennlinie des Langkanaltransistors abflacht und sich dessen Schwellspannung dem Langkanalwert annähert. Lang- und Kurzkanaltransistoren sollten deshalb ungefähr bei den gleichen Drainspannungen betrieben werden.

Für die weitere Beurteilung des Nachbarbiteffekts wird in Abb. 4.28 die Rückwärts-Lese-Schwellspannung über der Vorwärts-Lese-Schwellspannung dargestellt. Eingetragen sind die Kennlinien für Transistoren unterschiedlicher Kanallängen.

Während die Langkanaltransistoren, wie schon beschrieben wurde, bei geringen Rückwärts-Lese-Spannungen einen Kennlinienabschnitt mit konstanter Vorwärts-Lese-Span-



**Abbildung 4.27:** Vergleich der Abhängigkeit der Schwellspannung von der Drain-Source-Spannung von Lang- und Kurzkanaltransistoren



**Abbildung 4.28:** Schwellspannung der Vorwärts-Lese-Betriebsart aufgetragen über die korrespondierende Rückwärts-Lese-Schwellspannung für Transistoren unterschiedlicher Kanallängen

nung aufweisen, sind Transistoren mit kürzerer Kanallänge auch in diesem Bereich sensitiv auf Änderungen der Programmierung des Nachbarbits und damit auf die Rückwärts-Lese-Spannung. Dieser Effekt wurde schon in Abb. 4.23 beobachtet und in Abschnitt 4.3.3 diskutiert.

Besonders bei dem Transistor mit mittlerer Kanallänge ist erkennbar, dass die Kennlinie in zwei annähernd lineare Bereiche mit unterschiedlicher Steigung unterteilt ist. Der Übergang zwischen beiden Abschnitten findet näherungsweise unabhängig von der Transistorlänge bei derselben Rückwärts-Lese-Spannung statt, bei der beim Langkanaltransistor der Bereich konstanter Vorwärts-Lese-Spannung endet und an der  $\Delta V_M$  definiert ist. Damit kann die Berechnung von  $\Delta V_M$  gemäß Gl. (4.64) in erster Näherung auch für kurzkanalige Transistoren verwendet werden. Wie an Abb. 4.28 deutlich wird, nähern sich die Steigungen der beiden Kennlinienabschnitte bei kürzer werdenden Kanallängen einander an.

## 4.4 Subthreshold-Langkanalmodell

Um das Subthreshold-Verhalten von programmierten ONO-Langkanaltransistoren zu untersuchen, können prinzipiell die Zwischenergebnisse der Herleitung des MOS-Subthreshold-Modells von Abschnitt 3.4 verwendet werden. Zu beachten ist jedoch, dass die Unterschiede im Potentialverlauf von programmierten ONO-Speichertransistoren berücksichtigt werden.

Wie in der Herleitung der Kennliniengleichung der Standard-MOS-Transistoren, wird zunächst die Gültigkeit der Drift-Diffusion-Transportgleichung Gl. (3.45) vorausgesetzt. Im Abschnitt 4.4.3 werden die Grenzen dieser Annahme diskutiert.

### 4.4.1 Subthreshold-Transferkennlinie

Für die Herleitung der Subthreshold-Transferkennlinie sei zunächst Gl. (4.28) wiederholt:

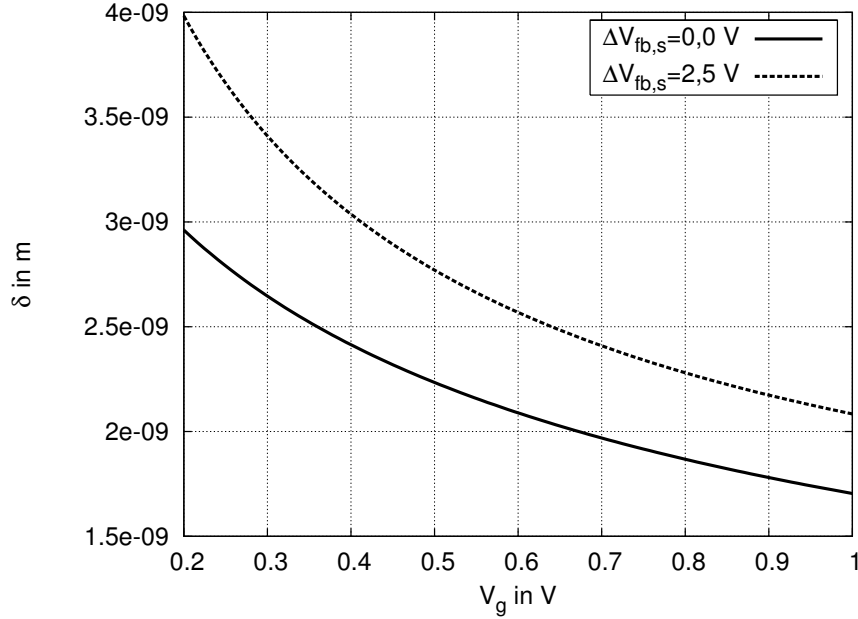
$$\psi_{S,min}(V_g) = \psi_{S,L}(V_g) - \underbrace{\Delta V_{fb,s} \frac{\lambda_s}{l_c + \lambda_s} \exp(-x_{min}(V_g)/\lambda_s)}_{\Delta\psi_{S,L}}. \quad (4.65)$$

Daraus ergibt sich direkt:

$$l_c^2 \frac{d^2 \psi_S}{dx^2} \Big|_{x_{min}} = -\frac{\lambda_s}{l_c} \Delta\psi_{S,L}. \quad (4.66)$$

Für die effektive Kanaltiefe  $\delta$  erhält man unverändert zu Gl. (3.56):

$$\delta(\psi_{s,min}) \approx \frac{\epsilon_{si} U_T}{q N_A t_{dep,L}(\psi_{S,min})} \underbrace{\left[ 1 - \exp\left(-\frac{2\psi_{S,min}}{U_T}\right) \right]}_{\approx 1}, \quad (4.67)$$



**Abbildung 4.29:** Effektive Kanaltiefen  $\delta$  von programmierten und unprogrammierten Langkanaltransistoren

wobei hier natürlich  $\psi_{S,min}$  aus Gl. (4.65) anzuwenden ist. Aus Abb. 4.29 ist ersichtlich, dass sich die Kanaltiefe eines programmierten ONO-Transistors nur in geringem Maße von der Kanaltiefe eines unprogrammierten Langkanaltransistors unterscheidet. Die Variation mit dem Gatepotential ist relativ gering.

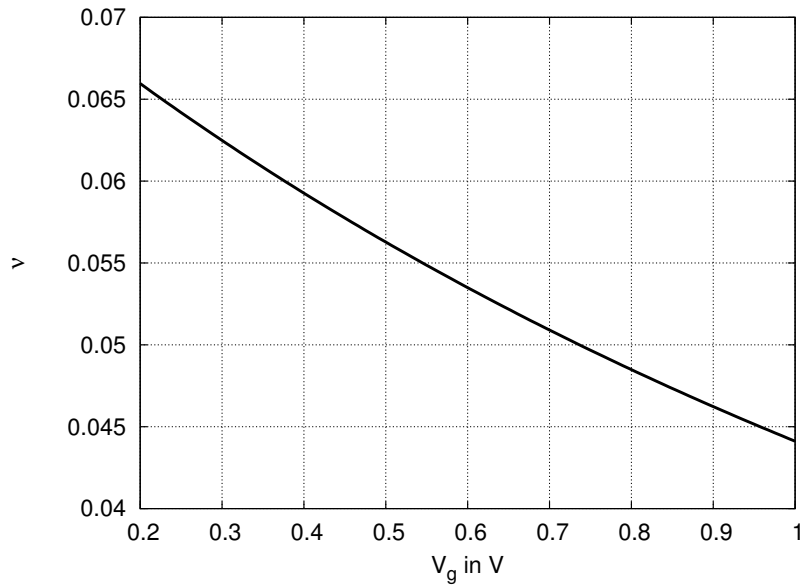
Der Faktor  $\nu$  beinhaltet die Auswirkung der Krümmung des Potentialverlaufs in der Nähe des Potentialminimums. Man erhält:

$$\nu = \frac{1}{L} \int_0^L \exp\left(-\frac{1}{U_T} \frac{d^2\psi_S}{dx^2} (x - x_{min})^2\right) dx. \quad (4.68)$$

$$\nu = \sqrt{-\frac{l_c}{\lambda_s} \frac{U_T}{\Delta\psi_{S,L}}} \frac{l_c}{L} \frac{\sqrt{\pi}}{2} \left[ \underbrace{\operatorname{erf}\left(\sqrt{-\frac{\lambda_s}{l_c} \frac{\Delta\psi_{S,L}}{U_T}} \frac{L - x_{min}}{l_c}\right) + \operatorname{erf}\left(\sqrt{-\frac{\lambda_s}{l_c} \frac{\Delta\psi_{S,L}}{U_T}} \frac{x_{min}}{l_c}\right)}_{\approx 2 \text{ für } (-\lambda_s/l_c \Delta\psi_{S,L} \gg U_T)} \right], \quad (4.69)$$

Für  $\Delta V_{fb,s} \rightarrow 0$  gilt wie beim unprogrammierten Langkanaltransistor  $\nu \rightarrow 1$ .

Aus Abb. 4.30 ist ersichtlich, dass bei mittleren oder großen Programmier-niveaus der Korrekturfaktor  $\nu$  sich zwar nur gering mit dem Gatepotential  $V_g$  ändert, aber einen wesentlichen Einfluss auf den Drain-Source-Strom hat. Der physikalische Hintergrund ist, dass der Bereich, in dem das Kanalpotential sein Minimum erreicht und der für die Sper-



**Abbildung 4.30:** Korrekturfaktor  $v$  eines programmierten Langkanal-ONO-Transistors mit  $\Delta V_{fb,s} = 2,5 \text{ V}$

ung des Transistors verantwortlich ist, bei starker Programmierung sehr schmal wird. Folglich sinkt die Fähigkeit zur Stromunterdrückung und der Drainstrom steigt an.

Kombiniert man die gewonnenen Zusammenhänge, erhält man mit  $\Delta \psi_{S,L}$  aus Gl. (4.65):

$$I_{ds} = \frac{\bar{\mu}_n W \epsilon_{si} U_T^2}{L t_{dep,L} v} \exp \left( \frac{V_{gs} - V'_{th0,L}}{n_0 U_T} + \frac{\Delta \psi_{S,L}}{U_T} \right) \left[ 1 - \exp \left( -\frac{V_{ds}}{U_T} \right) \right]. \quad (4.70)$$

Wie in Abb. 4.70 gezeigt wird, ist die Übereinstimmung des modellierten Drainstroms mit den Ergebnissen einer Bauelementesimulation zufriedenstellend.

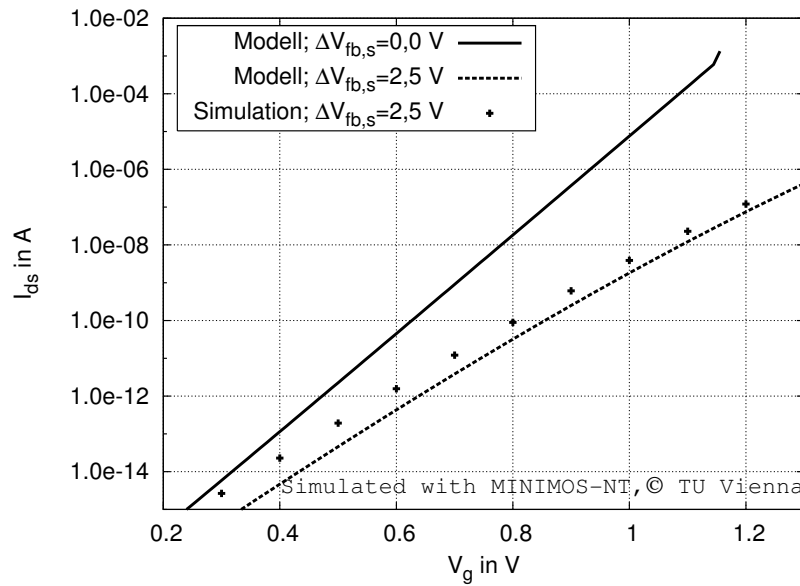
#### 4.4.2 Subthreshold-Steigung

Nachdem die Strom-Spannungs-Beziehung für den Subthreshold-Bereich bekannt ist, kann auch die Subthreshold-Steigung berechnet werden.

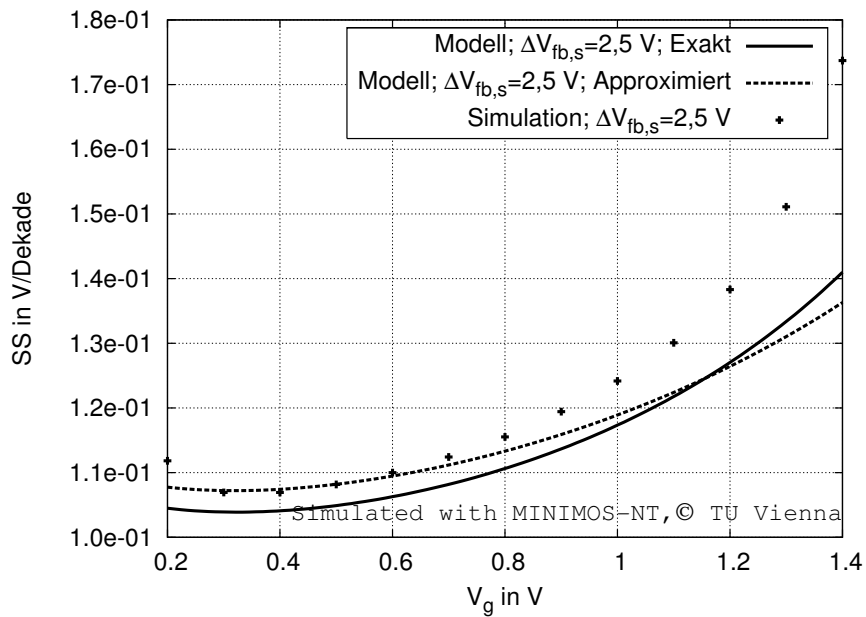
Da die Parameter  $\delta$  und  $v$ , wie beschrieben, im Vergleich zum exponentiellen Einfluss von  $\psi_{S,min}$  nur gering mit dem Gatepotential variieren, werden sie bei der Berechnung der Subthreshold-Steigung als konstant angenommen. Unter Anwendung der Definitionsgleichung der Subthreshold-Steilheit Gl. (3.60) auf die Kennliniengleichung Gl. (4.70) folgt nach Differentiation und unter Beachtung von Gl. (4.30):

$$SS = \frac{\ln(10) U_T}{(1 - \exp(-x_{min}(V_g)/l_c)) \frac{d\psi_{S,L}}{dV_{gs}}}. \quad (4.71)$$





**Abbildung 4.31:** Strom-Spannungs-Kennlinie eines programmierten und unprogrammierten ONO-Langkanaltransistors und Vergleich mit einer Bauelementesimulation



**Abbildung 4.32:** Subthreshold-Steigung eines programmierten ONO-Langkanaltransistors und Vergleich mit den Ergebnissen einer Bauelementesimulation; Neben der exakten Lösung nach Gl. (4.71) ist auch das Ergebnis der Approximation gemäß Gl. (4.71) eingetragen.

Wendet man jedoch den quadratischen Approximationsansatz Gl. (4.53) an, kann man die Subthreshold-Steigung näherungsweise angeben mit:

$$SS \approx \frac{\ln(10) U_T}{\left(1 - \exp(-x_{min}^\Delta / l_c)\right) \left(1 - \frac{\Delta V_{fb,s,0}}{\Delta V_{fb,s}} + \frac{l_c (l_c + \lambda_s)}{2 n_0 \lambda_s^2} \frac{V_{gs} - V'_{th0,L}}{\Delta V_{fb,s}}\right)} \frac{d\psi_{S,L}}{dV_{gs}}. \quad (4.72)$$

Die Funktionsverläufe sind in Abb. 4.32 dargestellt. Bemerkenswert ist, dass die Subthreshold-Steigung, im Unterschied zu ihrem Verhalten bei gewöhnlichen MOS-Transistoren, deutlich vom Gatepotential selbst abhängt. Je höher das Gatepotential wird, um so größer wird die Subthreshold-Steigung.

Der Vergleich des Modells mit den Resultaten einer Bauelementesimulation eines ONO-Langkanaltransistors im programmierten Zustand zeigt eine gute Übereinstimmung. Die Subthreshold-Steigung wurde vom Modell leicht unterschätzt. Der sehr starke Anstieg bei  $V_g = 1,4 \text{ V}$  in der Bauelementesimulation ist auf den Beginn der starken Inversion zurückzuführen, da die Schwellspannung des Transistors bei einer Programmierung von  $\Delta V_{fb,s} = 2,5 \text{ V}$  ca.  $1,45 \text{ V}$  beträgt.

#### 4.4.3 Grenzen und Erweiterungen des Subthreshold-Modells

Bei der Herleitung der Langkanal-Subthreshold-Gleichungen wurde für die Beschreibung des Ladungsträgertransports über die Potentialbarriere die Gültigkeit der Drift-Diffusion-Transportgleichung Gl. (3.45) vorausgesetzt. Die Gültigkeit dieser Gleichung ist aber an das Vorhandensein einer isotropen Maxwell-Verteilung der Ladungsträgergeschwindigkeiten gebunden [Ber85][RW88][Sze81]. Weil sich diese Gleichverteilung aufgrund der Gitterstöße einstellt, kann der Ladungsträgertransport über die Barriere nur dann durch die Drift-Diffusion-Gleichung beschrieben werden, wenn die Ladungsträger beim Überqueren der Barriere mit großer Wahrscheinlichkeit Gitterstöße ausführen. Die Barriere muss also hinreichend breit sein, um die Anwendung der Drift-Diffusion-Transportgleichung zu rechtfertigen.

Im anderen Grenzfall ist die Potentialbarriere dagegen so schmal, dass die Ladungsträger die Barriere ohne Gitterstöße und somit ohne Energieverlust überwinden können. In diesem Fall wird angenommen, dass aufgrund der Energieverteilung einige Ladungsträger auf beiden Seiten der Potentialbarriere eine genügend hohe Energie und entsprechende Richtungsvektoren besitzen, um die Barriere überwinden zu können. Dieser Transportprozess wird als *thermische Emission* bezeichnet. Da diese Ladungsträger das Kollektiv an der Barrierengrenze verlassen, ist die Geschwindigkeitsverteilung dicht an der Barrierengrenze nicht mehr isotrop und die Drift-Diffusion-Gleichung ist nicht mehr anwendbar [Ber85].

Wie gezeigt werden wird, ist in ONO-Transistoren die Breite der Potentialbarriere in der Umgebung des Barrierenmaximums von der Programmierung des Transistors abhängig. Bei geringer Programmierung ist die Barriere breit und der Ladungsträgertransport

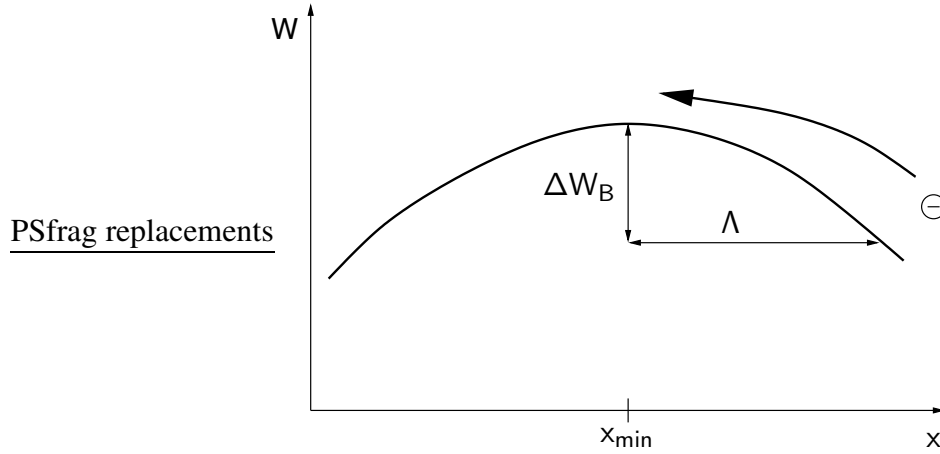


Abbildung 4.33: Ladungsträgertransport über eine Barriere

kann durch die Drift-Diffusion-Gleichung beschrieben werden. Bei starker Programmierung wird die Potentialbarriere dagegen so schmal, dass thermische Emissionsprozesse den Ladungsträgertransport über die Barriere bestimmen. In diesem Fall ist die Subthreshold-Transfergleichung Gl. (4.70) nicht mehr gültig und bedarf einer Modifikation.

In Abb. 4.33 ist der potentielle Energieverlauf im Halbleiter in der Umgebung der maximalen Barrierenhöhe dargestellt. Nach Bethe findet der Stromtransport über die Barriere vor allem dann durch thermische Emissionsvorgänge statt, wenn die Strecke, auf der der Ladungsträger beim Überqueren des Energiemaximums die Energiedifferenz  $\Delta W_B$  überwindet, kleiner ist als die mittlere freie Weglänge des Ladungsträgers  $\Lambda$ . Das modifizierte Bethe-Kriterium<sup>1</sup> ist damit [Ber85]:

$$\Delta W_B := |q (\psi_{S,min} - \psi_S(x = x_{min} \pm \Lambda))| > 2kT. \quad (4.73)$$

Der Potentialverlauf eines programmierten Langkanal-ONO-Transistors in der Umgebung des Potentialmaximums kann mit Gl. (4.66) durch folgende Taylorreihe näherungsweise beschrieben werden:

$$\psi_S(x) \approx -\frac{1}{2} \frac{\lambda_s}{l_c} \Delta\psi_{S,L} \left( \frac{x - x_{min}}{l_c} \right)^2 + \psi_{S,min}. \quad (4.74)$$

Die starke Programmierung und die daraus folgende große Potentialverschiebung  $\Delta\psi_{S,L}$  führen also zu einer starken Krümmung des Potentialverlaufs, folglich zu einer geringeren effektiven Barrierenbreite und zu zunehmenden thermischen Emissionsprozessen.

<sup>1</sup>Bethe ging in seiner Untersuchung von einer minimalen Energiedifferenz  $\Delta W_B > kT$  aus. Eine detaillierte Betrachtung in [Ber85] präzisiert jedoch die Grenze des Übergangs zur thermischen Emission auf Gl. (4.73).

Mit Gl. (4.74) kann man das Bethe-Kriterium für die Dominanz der thermischen Emission beim Ladungsträgertransport für den Sonderfall der programmierten ONO-Langkanaltransistoren näherungsweise folgendermaßen berechnen:

$$-\Delta\psi_{S,L} > 4U_T \frac{l_c}{\lambda_s} \left( \frac{l_c}{\Lambda} \right)^2. \quad (4.75)$$

In diesem Fall geht die Transferkennlinie Gl. (4.70) analog zu [FS93] über in:

$$I_{ds} = A^* T^2 W \delta \exp \left( \frac{V_{gs} - V'_{th0,L}}{n_0 U_T} + \frac{\Delta\psi_{S,L}}{U_T} \right) \left[ 1 - \exp \left( -\frac{V_{ds}}{U_T} \right) \right], \quad (4.76)$$

wobei  $A^*$  die effektive Richardson-Konstante der Elektronen in Silizium bezeichnet. Für eine Bewegung der Elektronen in die  $\langle 100 \rangle$ -Kristallrichtung ist diese Konstante  $A^* = 252 \text{ A/cm}^2 \text{ K}^2$  [Sze81].

Beim Vergleich dieser Kennliniengleichung mit Gl. (4.70) erkennt man, dass sich eine ähnliche exponentielle Struktur ergibt und lediglich der Vorfaktor verändert ist.

## 5 Anwendung beim Entwurf von Flash-Speichern

In diesem Kapitel soll demonstriert werden, wie die im Kapitel 4 gewonnenen Erkenntnisse angewendet werden können, um typische Problemstellungen zu lösen, die im Entwurfsprozess von hochintegrierten Flash-Speichern mit ONO-Speichertransistoren auftreten. Für diesen Zweck werden anhand einiger Beispiele, die für den praktischen Entwurf von Flash-Speichern besonders relevant sind, die Anwendung der hergeleiteten Formeln und Diagramme gezeigt.

### 5.1 Dimensionierung der Drainspannung

Bereits in der Konzeptphase werden von den System- und Schaltungsentwicklern verlässliche Abschätzungen des Leistungsbedarfs des Systems innerhalb der Betriebsphasen Lesen–Programmieren–Löschen abgefordert. Da das Lesen des Dateninhalts des Speicherfelds als Verify-Operation auch in den Betriebsarten Programmieren und Löschen eine Rolle spielt, ist die Abschätzung des Leistungsbedarfs des Lesevorgangs besonders wichtig.

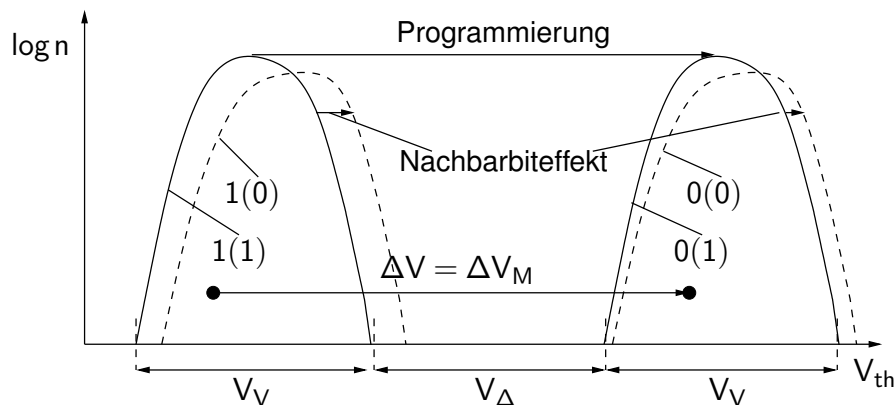
Wie gezeigt werden wird, haben im Lesebetrieb die durch das ständige Umladen der Globalen Bitleitung entstehenden dynamischen Ströme einen großen Anteil an der Gesamtverlustleistung. Eine optimale Dimensionierung der Bitleitungsspannung ist aus diesem Grund besonders wichtig.

#### Problemstellung

Gegeben sei ein Flash-Speicherfeld, das aus Langkanal-ONO-Speichertransistoren mit  $L = 0,5 \mu\text{m}$  aufgebaut ist. Die Breiten der Verteilungen der gelöschten und programmierten Bits seien  $V_V = 0,5 \text{ V}$ .

Wie hoch muss die Drainspannung gewählt werden, um den Nachbarbiteffekt unter Beachtung der Verlustleistungseffizienz optimal zu unterdrücken, wenn ein Abstand der gelöschten und programmierten Verteilungen von  $V_\Delta = 0,5 \text{ V}$  eingestellt werden soll?

Wie hoch ist in diesem Fall die aus der externen Spannungsversorgung aufgenommene mittlere dynamische Verlustleistung, die durch das Aufladen einer Bitleitungskapazität  $C = 1 \text{ pF}$  im Lesetakt von  $T_R = 100 \text{ ns}$  entsteht? Der Lesevorgang sei parallel in  $n = 128 \text{ DQs}$  ausgeführt und die externe Spannung betrage dabei  $V_{dd} = 3,3 \text{ V}$ .



**Abbildung 5.1:** Verschiebung der Verteilungen durch die Programmierung einzelner Bits und durch den Nachbarbiteffekt für den Fall einer homogenen Programmierung, bei der die relative Lage der Bits innerhalb der Verteilung bei der Programmierung nicht verändert wird. Gekennzeichnet ist der Dateninhalt der Bits und in Klammern der Dateninhalt der jeweiligen Nachbarbits.

### Lösung

Das Vorgehen zur Lösung des ersten Teils der Aufgabenstellung wird anhand der Abb. 5.1 erläutert. Dargestellt ist zunächst die Verteilung der gelöschten Bits mit dem logischen Wert „1“. Durch die Programmierung einiger gelöschter Bits entsteht die Verteilung der programmierten Bits. Durch die fortgesetzte Programmierung verschiebt sich die Verteilung der programmierten Bits zu höheren Schwellspannungen. Ab einer bestimmten programmierten Schwellspannung wird für die gelöschten Nachbarbits der Nachbarbiteffekt wirksam. Dieser Nebeneffekt äußert sich darin, dass die Schwellspannung eines gelöschten Nachbarbits durch die Programmierung der anderen Transistorseite unbeabsichtigt verschoben wird. Die Schwellspannungen derjenigen gelöschten Bits, die programmierte Nachbarbits haben, verschieben sich daher zu höheren Werten. Die Gesamtverteilung der gelöschten Bits mit programmierten bzw. gelöschten Nachbarbits wird folglich breiter.

In der Abbildung werden die Spannungsverhältnisse am Beispiel eines Transistors gezeigt, bei dem ein Bit gelöscht ist und das zweite Bit programmiert werden soll (Dateninhalt „0(1)“ bzw. „1(0)“). Wird das programmierte Bit ausgelesen, hat der Transistor folglich ein gelöschtes Nachbarbit. Beim Auslesen des gelöschten Bits ist das Nachbarbit programmiert. Diese beiden Betriebsarten des Transistors wurden aber in Abschnitt 4.3.2 als Vorwärts- bzw. Rückwärts-Lese-Betrieb bezeichnet und im Hinblick auf die Bittrennung untersucht. Dort wurde festgestellt, dass die maximale Schwellspannungsverschiebung des programmierten Bits des Transistors, ohne dass sich die Schwellspannung des gelöschten Bits im selben Transistor durch den Nachbarbiteffekt wesentlich ändert, dem Wert  $\Delta V_M$  entspricht. Folglich wird der Nachbarbiteffekt gerade noch unterdrückt, falls

die Schwellspannungsverschiebung  $\Delta V = \Delta V_M$  beträgt. Aus Abb. 5.1 geht weiterhin hervor, dass gilt:

$$\Delta V \stackrel{!}{=} \Delta V_M = V_V + V_\Delta = 1 \text{ V}.$$

$V_V$  beschreibt dabei die Breite der Verteilung und  $V_\Delta$  den Abstand der programmierten von der gelöschten Verteilung, das sogenannte *Lesefenster*. Für die Ableitung dieses Zusammenhangs wurde vorausgesetzt, dass die Programmierung homogen erfolgt, so dass alle programmierten Bits um denselben Betrag verschoben wurden und dass sich deshalb die relativen Lagen der Bits innerhalb der Verteilungen vor und nach dem Programmiervorgang nicht unterscheiden. Zur Berechnung der minimalen Drain-Source-Spannung, die für Langkanaltransistoren notwendig ist, um diese Bittrennung zu gewährleisten, kann nun Gl. (4.64) verwendet werden, die dann entsprechend nach  $V_{ds}(\Delta V_M)$  umgestellt werden muss. Alternativ kann man in unserem Fall auch direkt aus dem Diagramm Abb. 4.22 ablesen:

$$V_{ds}(V_M) \approx \underline{\underline{2,4 \text{ V}}}.$$

Die mittlere Leistungsaufnahme aus der externen Stromversorgung beträgt allgemein:

$$\bar{P} = V_{dd} \bar{I}.$$

Der mittlere dynamische Ladestrom durch das parallele Aufladen von  $n = 128$  Bitleitungen mit einer Bitleitungskapazität von jeweils  $C = 1 \text{ pF}$  von  $V_{ds} = 0 \text{ V}$  auf  $V_{ds} = 2,4 \text{ V}$  mit einer Periodendauer des Lesevorgangs von  $T_R = 100 \text{ ns}$  beträgt:

$$\bar{I} = nC \frac{V_{ds}}{T_R}.$$

Folglich ergibt sich die mittlere Leistungsaufnahme zu:

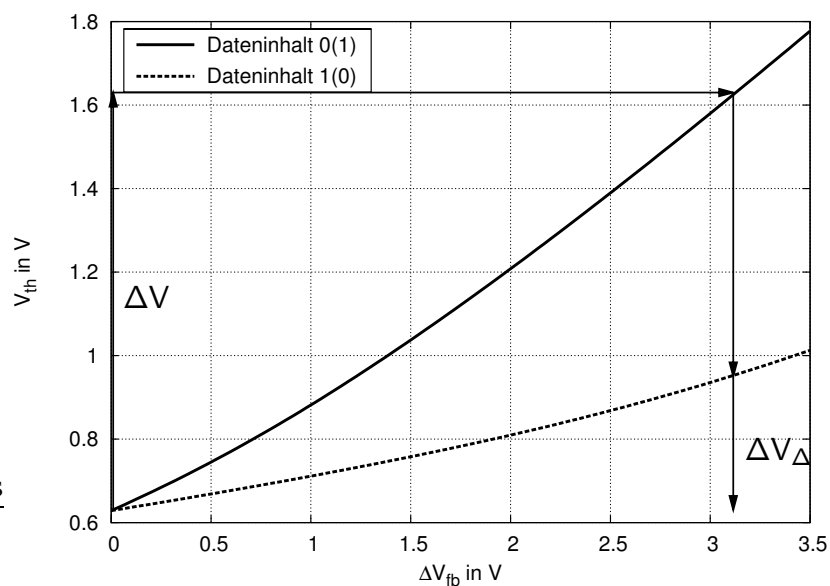
$$\begin{aligned} \bar{P} &= nC \frac{V_{dd} V_{ds}}{T_R} \\ &= \underline{\underline{10 \text{ mW}}}. \end{aligned}$$

## 5.2 Verringerung des Lesefensters

Teilweise ist es in einigen Anwendungsfällen nicht ohne Weiteres möglich, eine ausreichende Drainspannung zur Gewährleistung der Bittrennung bereitzustellen. Gründe hierfür könnten zum Beispiel eine zu hohe dynamische Verlustleistung oder die ungenügende Höhe der externen Versorgungsspannung sein. So würde die Bereitstellung einer Spannung von  $V_{ds} = 2,4 \text{ V}$  aus einer externen Versorgungsspannung von  $V_{dd} = 1,8 \text{ V}$  nur durch den Einsatz von Spannungsvervielfacherschaltungen möglich sein, was sich allerdings meist aus Gründen des Strom- oder Flächenbedarfs verbietet.

In diesem Fall ist es notwendig, die Reduzierung des Abstands der gelöschten und der programmierten Verteilung, also des Lesefensters  $V_\Delta$ , durch den Nachbarbiteffekt

PSfrag replacements



**Abbildung 5.2:** Verschiebung der Schwellenspannung  $V_{th}$  eines Kurzkanal-ONO-Transistors in Abhängigkeit von der Programmierung  $\Delta V_{fb}$  für  $V_{ds} = 1,3 \text{ V}$

abzuschätzen, um letztendlich beispielsweise die Wahl des Lese-Referenzwerts zu ermöglichen.

### Problemstellung

Gegeben sei ein Flash-Speicherfeld, das aus Kurzkanal-ONO-Speichertransistoren mit  $L = 0,18 \mu\text{m}$  aufgebaut ist. Um eine externe Versorgungsspannung von  $V_{ds} = 1,8 \text{ V}$  zu ermöglichen, wird die Drain-Source-Spannung auf  $V_{ds} = 1,3 \text{ V}$  festgelegt. Das ideale Lesefenster, also ohne die Auswirkungen des Nachbarbiteffekts, sei  $V_{\Delta} = 0,5 \text{ V}$  und die Verteilungsbreite  $V_V = 0,5 \text{ V}$ .

Wie groß ist die Verringerung des Lesefensters durch den Nachbarbiteffekt?

### Lösung

Für die Lösung dieser Aufgabenstellung sei zunächst wieder auf Abb. 5.1 verwiesen. Aus dieser Abbildung wird ersichtlich, dass durch den Nachbarbiteffekt die Verteilung der gelöschten Bits mit programmierten Nachbarbits (Dateninhalt „1(0)“) sowie die Verteilung der programmierten Bits mit ebenfalls programmierten Nachbarbits (Dateninhalt „0(0)“) zu höheren Schwellenspannungen verschoben werden. Das Lesefenster wird aber nur durch die Verschiebung der Verteilung der gelöschten Bits verringert. Für die Lösung der Aufgabenstellung ist deshalb auch nur diese Spannungsverschiebung von Interesse.

Da sich die Aufgabenstellung auf Kurzkanaltransistoren bezieht, muss Gl. (4.46) direkt auf numerischem Weg gelöst werden. In Abb. 5.2 ist diese Lösung für den entsprechenden ONO-Speichertransistor für  $V_{ds} = 1,3 \text{ V}$  und die Dateninhalte „1(0)“ und „0(1)“ als Funktion des Programmierlevels  $\Delta V_{fb}$  dargestellt.



Zunächst wird berechnet, wie groß die Spannungsverschiebung durch die Programmierung  $\Delta V$  sein muss, um das geforderte ideale Lesefenster von 0,5 V zu erhalten. Wie in der vorherigen Problemstellung im Abschnitt 5.1 erhält man unter Voraussetzung eines homogenen Programmiervorgangs:

$$\Delta V = V_V + V_\Delta = 1 \text{ V}.$$

Anhand der Abb. 5.2 kann das Programmierniveau  $\Delta V_{fb,1}$  abgelesen werden, das erforderlich ist, um diese Spannungsverschiebung einer Speicherzelle mit Dateninhalt „0(1)“ zu erreichen:

$$\Delta V_{fb,1} \approx 3,1 \text{ V}.$$

Nun kann an der Kennlinie des Dateninhalts „1(0)“ abgelesen werden, wie weit sich die Schwellspannung der gelöschten Bits dadurch verschiebt, dass die Nachbarbits auf das ermittelte Niveau programmiert werden. Diese Spannungsverschiebung entspricht der Verringerung des Lesefensters:

$$\Delta V_\Delta = V_{th}(\Delta V_{fb,1}) - V_{th}(0) \approx \underline{\underline{330 \text{ mV}}}.$$

## 5.3 Parasitäre Subthreshold-Ströme im Lesevorgang

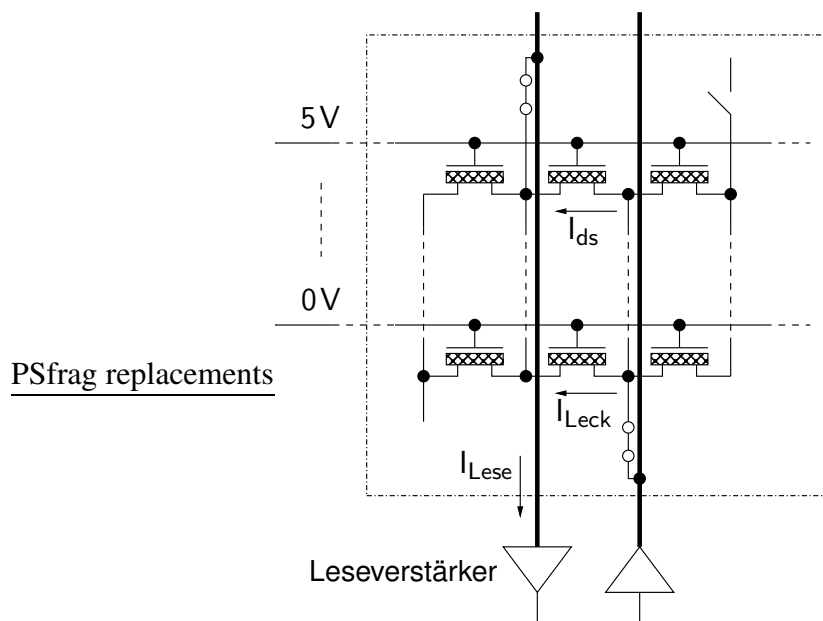
Als Folge der Tatsache, dass die Transistoren im Virtual-Ground-Speicherfeld innerhalb eines Slices nicht durch Schaltertransistoren voneinander getrennt sind, kommt es zu sogenannten Arrayeffekten. So kann es zum Beispiel beim Auslesen des Dateninhalts einer Speicherzelle dazu kommen, dass der Messstrom im Leseverstärker nicht exakt dem Drain-Source-Strom des Speichertransistors entspricht, sondern durch die nichtadressierten Transistoren gestört wird.

Ein Teil dieser Störströme ist auf den Subthreshold-Leckstrom der nichtadressierten Speicherzellen zurückzuführen.

### Problemstellung

Gegeben sei ein Flash-Speicherfeld mit  $(m + 1) = 1024$  Wortleitungen, das aus Kurzkanal-ONO-Speichertransistoren mit  $W = 0,2 \mu\text{m}$  und  $L = 0,18 \mu\text{m}$  aufgebaut ist. Die mittlere initiale Langkanalschwellspannung der verwendeten Transistoren betrage  $V'_{th0,L} = 0,85 \text{ V}$ . Die Verteilung der initialen Langkanalschwellspannungen sei symmetrisch um  $V'_{th0,L}$  mit einer Verteilungsbreite von  $\pm 0,25 \text{ V}$ . Die Spannung der nichtadressierten Wortleitung betrage  $0 \text{ V}$ . Die Bitleitungsspannung sei  $V_{ds} = 1,3 \text{ V}$ .

Wie groß ist der Fehler im Lesestrom, der durch die parasitären Leckströme der nichtadressierten Speichertransistoren hervorgerufen wird, wenn angenommen werden soll, dass im schlechtesten Fall alle betreffenden Transistoren des Slices den maximalen Leckstrom haben?



**Abbildung 5.3:** Störung des Lesestroms durch die Subthreshold-Leckströme einiger nichtadressierter Speichertransistoren

### Lösung

Zunächst sei festgestellt, dass prinzipiell nur die Transistoren innerhalb des adressierten Slices zum Messfehler beitragen können, da die Speichertransistoren der anderen Slices mit Hilfe der Select-Transistoren elektrisch abgetrennt wurden.

Die Gegebenheiten innerhalb des Slices sind in Abb. 5.3 dargestellt. Durch den für die Leseoperation ausgewählten Transistor fließt der zu messende Drain-Source-Strom  $I_{ds}$ . Alle Transistoren, die sich mit dem adressierten Transistor beide Lokale Bitleitungen teilen und die damit mit ihren Source-Drain-Kontakten parallel zum auszulesenden Transistor liegen, sind durch die Wortleitungsspannung von 0V ausgeschaltet. Die Subthreshold-Leckströme  $I_{Leck}$  aller dieser ausgeschalteten Transistoren überlagern sich jedoch dem Source-Drain-Strom  $I_{ds}$ , so dass der gemessene Strom  $I_{Lese}$  gestört ist. Bei allen anderen Transistoren des Slices ist mindestens ein Source- oder Drainkontakt mit Hilfe der Select-Schalter von der Globalen Bitleitung abgetrennt. Durch diese Transistoren können also keine Subthreshold-Leckströme fließen und die Transistoren tragen deshalb nicht zur Störung bei.

Zunächst sei festgestellt, dass der maximale Leckstrom mit der minimalen Schwellspannung verknüpft ist. Folglich ist der Leckstrom derjenigen Transistoren, bei denen beide Bits im initialen oder gelöschten Zustand sind, besonders hoch. Die minimale initiale Langkanalschwellspannung beträgt bei einer Verteilungsbreite von  $\pm 0,25$  V:

$$V_{th,min} = V'_{th0,L} - 0,25 \text{ V} = 0,65 \text{ V}.$$

Da für die Transistoren also der initiale Zustand betrachtet werden muss, kann man das

Kurzkanal-Subthreshold-Modell der Standard-MOS-Transistoren Gl. (3.59) anwenden. Man ermittelt zunächst unter Anwendung der Näherungsformel Gl. (3.39) mit  $V'_{th0,L} = V_{th,min}$  und  $V_{gs} = 0$  V:

$$\Delta\psi_S \approx \frac{2(V_{bi} - \psi_{S,L}(V_{th,min}) + V_d + V_s)}{2 \cosh(L/(2l_c))} = 0,22 \text{ V}.$$

Im nächsten Schritt kann  $v$  mittels Gl. (3.58) berechnet werden:

$$v \approx \sqrt{\frac{U_T}{\Delta\psi_S}} \frac{l_c}{L} \sqrt{\pi} = 0,13.$$

Damit ergibt sich der Drain-Source-Leckstrom eines Transistors mit Hilfe von Gl. (3.59) zu:

$$I_{Leck} \approx \frac{\bar{\mu}_n W \epsilon_{si} U_T^2}{L t_{dep,L} v} \exp\left(\frac{-V_{th,min}}{n_0 U_T} + \frac{\Delta\psi_S}{U_T}\right) = 80 \text{ pA}.$$

Der Fehler im Lesestrom ist die Summe der Leckströme aller  $m$  dem adressierten Transistor parallel geschalteten Transistoren und damit:

$$\sum_{i=1}^m I_{Leck} = m I_{Leck} \approx \underline{\underline{80 \text{ nA}}}.$$

Dieser Arrayeffekt ist folglich in diesem speziellen Fall gegenüber den typischen Drain-Source-Strömen von  $I_{ds} \gg 1 \mu\text{A}$  vernachlässigbar.

Wird die Verteilung der Langkanalschwellspannungen jedoch breiter, zum Beispiel durch eine Aufladung der Nitridschicht während des Herstellungsprozesses, können die parasitären Leckströme den Lesestrom messbar stören.

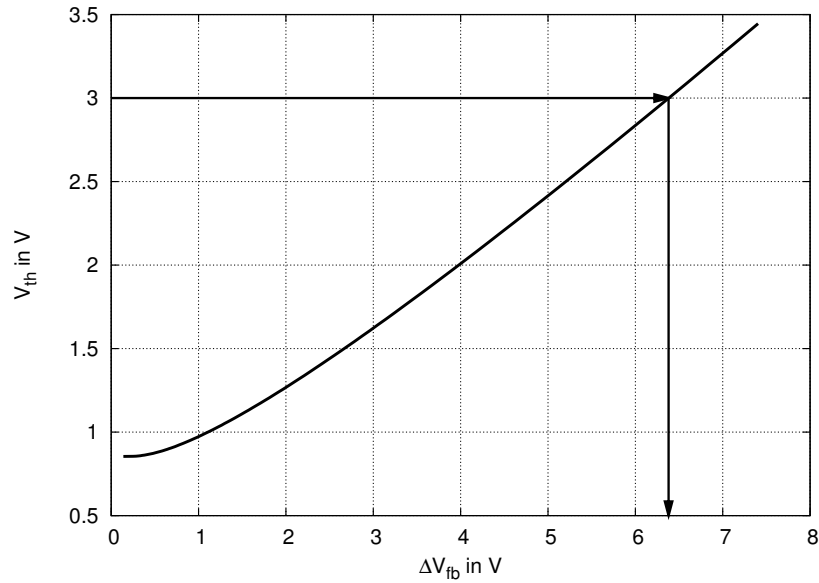
## 5.4 Subthreshold-Leckstrom von stark programmierten Transistoren

Einige spezielle Anwendungen erfordern, dass die ONO-Speichertransistoren sehr stark programmiert werden, um ihren Drainstrom so weit wie möglich zu unterdrücken. Man spricht dann gewöhnlicherweise von *überprogrammierten Transistoren*. Diese Transistoren werden dann häufig mit Gatespannungen unterhalb der Schwellspannung betrieben.

Aufgrund der starken Programmierung besitzen überprogrammierte Transistoren jedoch eine schlechtere Subthreshold-Steigung. In vielen Anwendungsfällen sollten deshalb die Subthreshold-Leckströme abgeschätzt werden, um deren Auswirkungen beurteilen zu können.

### Problemstellung

Gegeben sei ein Langkanal-ONO-Speichertransistor mit  $W = 1,0 \mu\text{m}$  und  $L = 0,5 \mu\text{m}$ . Die initiale Schwellspannung des Transistors betrage  $V'_{th0,L} = 0,85 \text{ V}$ . Der Transistor wurde auf  $V_{th} = 3 \text{ V}$  programmiert. Für das Auslesen wird eine Gate-Source-Spannung von



**Abbildung 5.4:** Schwellspannung eines Langkanal-ONO-Transistors mit  $L = 0,5 \mu\text{m}$  in Abhängigkeit von der Programmierung

$V_{gs} = 2,65 \text{ V} < V_{th}$  und eine Drain-Source-Spannung von  $V_{ds} \sim 2 \text{ V}$  verwendet.

Wie hoch ist der Auslesestrom?

Im Rahmen einer Schaltungssimulation wird für den überprogrammierten Transistor statt eines ONO-Modells ein Standard-MOS-Modell (z.B. BSIM) verwendet. Wie hoch ist in diesem Fall der simulierte Auslesestrom?

### Lösung

Im ersten Schritt muss das Programmierniveau des Transistors ermittelt werden. Zu diesem Zweck wird mit Hilfe der Parametergleichung Gl. (4.50), wie in Abb. 5.4 gezeigt, der Verlauf von  $V_{th}$  in Abhängigkeit von  $\Delta V_{fb}$  für einen der Aufgabenstellung entsprechenden Transistor aufgetragen. Aus dieser Darstellung kann man entnehmen:

$$\Delta V_{fb}(V_{th} = 3 \text{ V}) \approx 6,4 \text{ V}.$$

Mit diesem Wert kann man nun für  $V_{gs} = 2,65 \text{ V}$  mit Hilfe von Gl. (4.25) berechnen:

$$x_{min} = \frac{\lambda_s l_c}{l_c - \lambda_s} \ln \left( -\frac{l_c}{\lambda_s} \frac{\psi_3}{\psi_{2,L}^0} \right) = 18,6 \text{ nm},$$

und damit durch Anwendung von Gl. (4.28):

$$\Delta \psi_{S,L} = -\Delta V_{fb,s} \frac{\lambda_s}{l_c + \lambda_s} \exp(-x_{min}/\lambda_s) = -1,52 \text{ V}.$$

Vor der weiteren Berechnung des Subthreshold-Stroms muss nun abgeschätzt werden, ob der Ladungsträgertransport durch Drift- und Diffusionsprozesse oder durch thermische Emission erfolgt. Diese Abschätzung erfolgt mit Hilfe der Bedingung Gl. (4.75), wobei für die mittlere freie Weglänge der Elektronen in Silizium nach [NOY77]  $\Lambda \approx 9 \text{ nm}$  angenommen wird:

$$-\Delta\psi_{S,L} = 1,52 \text{ V} \not\geq 4 U_T \frac{l_c}{\lambda_s} \left( \frac{l_c}{\Lambda} \right)^2 = 2,29 \text{ V}.$$

Aus dieser Abschätzung folgt, dass die Subthreshold-Transferkennlinie durch die Drift und die Diffusion der Ladungsträger bestimmt wird.

Die Berechnung von  $v$  mittels Gl. (4.69) ergibt:

$$v = \sqrt{-\frac{l_c}{\lambda_s} \frac{U_T}{\Delta\psi_{S,L}}} \frac{l_c}{L} \sqrt{\pi} = 0,0195.$$

Nun sind alle notwendigen Zwischenergebnisse berechnet und man erhält schließlich als Auslesekstrom aus Gl. (4.70):

$$I_{ds} \approx \frac{\bar{\mu}_n W \epsilon_{si} U_T^2}{L t_{dep,L} v} \exp \left( \frac{V_{gs} - V'_{th0,L}}{n_0 U_T} + \frac{\Delta\psi_{S,L}}{U_T} \right) = \underline{\underline{58 \text{ nA}}}.$$

Dieser Subthreshold-Strom ist bemerkenswert hoch, da der Transistor eigentlich deutlich unterhalb seiner Schwellspannung betrieben wird.

Den Subthreshold-Strom eines Langkanal-MOS-Transistors mit  $V'_{th,L} = V_{th} = 3 \text{ V}$  erhält man mit  $\Delta\psi_S = 0$  und  $v = 1$  aus Gl. (3.59):

$$I_{ds} \approx \frac{\bar{\mu}_n W \epsilon_{si} U_T^2}{L t_{dep,L}} \exp \left( \frac{V_{gs} - V'_{th,L}}{n_0 U_T} \right) = \underline{\underline{2,6 \text{ pA}}}.$$

Bei einer Verwendung eines Standard-MOS-Modells wird also der Auslesekstrom in der Simulation um mehr als vier Größenordnungen unterschätzt.

## 6 Zusammenfassung

In dieser Arbeit wurde eine besondere Klasse von nichtflüchtigen Speichern untersucht, die in vielen kommerziellen Flashspeicher-Produkten eingesetzt wird. Kennzeichnend für diese Speicherart ist die ladungsspeichernde Oxid-Nitrid-Oxid-Isolationsschicht, in die lokal an beiden Enden des Speichertransistors beim Programmervorgang Elektronen injiziert werden. Aufgrund der lokal begrenzten Ladungsverteilungen können in diesen Transistoren zwei Bits unabhängig voneinander gespeichert werden.

Es wurde dargestellt, dass diese Transistoren durch Injektion hochenergetischer Kanalelektronen programmiert und durch Injektion hochenergetischer Löcher gelöscht werden. Die Speichertransistoren mit lokaler Ladungsspeicherung sind zur Realisierung von Flashspeichersystemen typischerweise in einer Virtual-Ground-Array-Speicherfeldarchitektur verschaltet. Für die Ausführung der Lese-, Programmier- und Löschoperationen sind komplexe Algorithmen abzuarbeiten.

Für Standard-MOS-Transistoren wurde mit Hilfe eines quasi-zweidimensionalen Ansatzes die zweidimensionale Poisson-Gleichung in eine gewöhnliche Differentialgleichung überführt. Die Lösung dieser Differentialgleichung ermöglichte die Berechnung des Verlaufs des Oberflächenpotentials. Von besonderem Interesse war dabei der Minimalwert des Potentialverlaufs, da anhand dieses Punkts die Schwellspannung des Transistors bestimmt werden konnte. Mit dem abgeleiteten Schwellspannungsmodell wurde gezeigt, dass die Schwellspannung in Kurzkanaltransistoren bei reduzierten Kanallängen und steigenden Drainspannungen aufgrund des Einflusses der elektrischen Felder der Kanalkontakte absinkt. Mit Hilfe der Kenntnisse des Potentialverlaufs im Transistorinneren konnte außerdem die Transferkennlinie im Subthreshold-Bereich bestimmt werden. Es wurde dargestellt, dass sich die Subthreshold-Steigung und damit das Sperrvermögen von Kurzkanaltransistoren dadurch verschlechtert, dass die Kanalkontakte im Vergleich zum Gateanschluss an Einfluss auf das Kanalpotential gewinnen.

Um die Verschiebung der Schwellspannung bei der Programmierung von ONO-Speichertransistoren beschreiben zu können, wurde das Modell der Flachbandspannung erweitert. Es wurde ein Ansatz gewählt, bei dem die lokale Injektion der Elektronen durch einen exponentiell zur Kanalmitte abfallenden Verlauf der Flachbandspannung modelliert wird. Unter Verwendung dieses Ansatzes kann die Differentialgleichung des Potentialverlaufs für programmierte ONO-Transistoren gelöst werden. Wie im Fall der Standard-MOS-Transistoren kann bei Kenntnis der Minimalwerte dieses Verlaufs die Schwellspannung des Speichertransistors berechnet werden. Für Langkanaltransistoren konnte durch Anwendung von Näherungen eine explizite Lösung der Schwellspannungsgleichung angegeben werden. Für Kurzkanaltransistoren konnte die Lösung durch die

---

Anwendung von numerischen Verfahren gewonnen werden.

Das abgeleitete Schwellspannungsmodell beschreibt die Verschiebung der Schwellspannung eines ONO-Speichertransistors durch die lokale Injektion von Elektronen. Mit diesem Modell konnte weiterhin gezeigt werden, dass die Schwellspannung eines gespeicherten Bits durch die Einflüsse des Nachbarbits im Transistor verändert werden kann. Es konnte demonstriert werden, dass diese Störung der Schwellspannung auch bei Langkanaltransistoren beobachtet werden kann und dass die Trennung der beiden im Transistor gespeicherten Bits beim Auslesen durch eine Erhöhung der Drainspannung verbessert wird.

Weiterhin konnte aus dem Potentialverlauf die Subthreshold-Kennlinie von programmierten Langkanal-ONO-Transistoren abgeleitet werden und gezeigt werden, dass sich die Subthreshold-Steigung durch das Programmieren der Transistoren verschlechtert. Als Ursache dafür wurde erkannt, dass sich das Potentialminimum durch die Programmierung in den Bereich des Felds des Sourcekontakts verschiebt und sich folglich der Einfluss des Gatepotentials auf die Höhe des Potentialminimums verringert.

Anhand einiger praktischer Beispiele konnte gezeigt werden, dass die qualitative und quantitative Kenntnis dieser Effekte notwendig ist, um optimierte Flash-Speicherschaltungen und -systeme entwerfen zu können.

## Literaturverzeichnis

- [ADM95] AGRAWAL, Bhavna ; DE, Vivek K. ; MEINDL, James D.: Three-Dimensional Analytical Subthreshold Models for Bulk MOSFET's. In: *IEEE Transactions on Electron Devices* 42 (1995), Nr. 12, S. 2170–2180
- [AS82] AKERS, L. A. ; SANCHEZ, J. J.: Threshold Voltage Models of Short, Narrow and Small Geometry MOSFET's: A Review. In: *Solid-State Electronics* 25 (1982), Nr. 7, S. 621–641
- [BB98] BROWN, William D. ; BREWER, Joe E.: *Nonvolatile Semiconductor Memory Technology*. IEEE Press, Piscataway, 1998
- [Ber85] BERZ, F.: The Bethe Condition For Thermionic Emission Near An Absorbing Boundary. In: *Solid-State Electronics* 28 (1985), Nr. 10, S. 1007–1013
- [BPE02] BLOOM, Ilan ; PAVAN, Paolo ; EITAN, Boaz: NROM<sup>TM</sup>—a new technology for non-volatile memory products. In: *Solid-State Electronics* 46 (2002), S. 1757–1763
- [BSI] Department of Electrical Engineering and Computer Sciences, University of Southern California, Berkeley: *BSIM MOSFET Model*. – <http://www-device.eecs.berkeley.edu>
- [CCC<sup>+</sup>87] CHEN, J. ; CHAN, T. Y. ; CHEN, J. C. ; KO, P. K. ; HU, Chenming: Sub-breakdown Drain Leakage Current in MOSFET. In: *IEEE Electron Device Letters* EDL-8 (1987), Nr. 11, S. 515–517
- [CCS<sup>+</sup>98] CHANGA, Kuo-Tung ; CHEN, Wei-Ming ; SWIFT, Craig ; HIGMAN, Jack M. ; PAULSON, Wayne M. ; CHANG, Ko-Min: A New SONOS Memory Using Source-Side Injection for Programming. In: *IEEE Electron Device Letters* 19 (1998), Nr. 7, S. 253–255
- [CL87] CHANG, Chi ; LIEN, Jih: Corner-Field Induced Drain Leakage In Thin Oxide MOSFETs. In: *Electron Device Meeting (EDM 1987)*. Washington, USA, 1987, S. 714–717
- [CPL<sup>+</sup>06] CHOI, Byung Y. ; PARK, Byung-Gook ; LEE, Jong D. ; SHIN, Hyungcheol ; LEE, Yong K. ; SUNG, Suk-Kang ; LEE, Se-Hoon ; CHAE, Heesoon ; LEE, Jong J. ; BAI, Keun H. ; KIM, Dong-Dae ; KIM, Dong-Won ; LEE,



- Choong-Ho ; PARK, Donggun: TWIn SONOS TransistOR (TWISTOR) for 2-bit/cell SONOS Memory Technology. In: *21st IEEE Non-Volatile Semiconductor Memory Workshop (NVSMW 2006)*. Monterey, USA, 2006
- [CT91] CONTI, Massimo ; TURCHETTI, Claudio: On the Short-Channel Theory for MOS-Transistor. In: *IEEE Transactions on Electron Devices* 38 (1991), Nr. 12, S. 2657–2661
- [EPB<sup>+</sup>99] EITAN, Boaz ; PAVAN, Paolo ; BLOOM, Ilan ; ALONI, Efraim ; FROMMER, Aviv ; FINZI, David: Can NROM, a 2-bit, Trapping Storage NVM Cell, Give a Real Challenge to Floating Gate Cells? In: *Proceedings of International Conference on Solid State Devices and Materials (SSDM)*. Tokyo, Japan, 1999, S. 522–524
- [EPB<sup>+</sup>00] EITAN, Boaz ; PAVAN, Paolo ; BLOOM, Ilan ; ALONI, Efraim ; FROMMER, Aviv ; FINZI, David: NROM: A Novel Localized Trapping, 2-Bit Nonvolatile Memory Cell. In: *IEEE Electron Device Letters* 21 (2000), Nr. 11, S. 543–545
- [FB70] FROHMAN-BENTCHKOWSKY, Dov: The Metal-Nitride-Oxide-Silicon (MNOS) Transistor-Characteristics and Applications. In: *Proceedings of the IEEE* 58 (1970), Nr. 8, S. 1207–1224
- [FB74] FROHMAN-BENTCHKOWSKY, Dov: FAMOS - A New Semiconductor Charge Storage Device. In: *Solid-State Electronics* 17 (1974), Nr. 6-B, S. 517–529
- [FDN<sup>+</sup>01] FRANK, David J. ; DENNARD, Robert H. ; NOWAK, Edward ; SOLOMON, Paul M. ; TAUR, Yuan ; WONG, Hon-Sum P.: Hot-Electron Emission in N-Channel IGFET's. In: *Proceedings of the IEEE* 89 (2001), Nr. 3, S. 259–288
- [Fis94] FISCHER, Claus: *Bauelementsimulation in einer computergestützten Entwurfsumgebung*, Technische Universität Wien, Diss., 1994. – <http://www.iue.tuwien.ac.at/phd/fischer>
- [FKR<sup>+</sup>06] FUKS, David ; KIV, Arnold ; ROIZIN, Yakov ; GUTMAN, Micha ; AVICHAIL-BIBI, Rachel ; MAXIMOVA, Tatyana: The Nature of HT  $V_t$  Shift in NROM Transistors. In: *IEEE Transactions on Electron Devices* 53 (2006), Nr. 2, S. 304–313
- [FS93] FJELDLY, Tor A. ; SHUR, Michael: Threshold Voltage Modeling and the Subthreshold Regime of Operation of Short-Channel MOSFET's. In: *IEEE Transactions on Electron Devices* 40 (1993), Nr. 1, S. 137–145

- [FTW98] FRANK, David J. ; TAUR, Yuan ; WONG, Hon-Sum P.: Generalized Scale Length for Two-Dimensional Effects in MOSFET's. In: *IEEE Electron Device Letters* 19 (1998), Nr. 10, S. 385–387
- [GD80] GREENFIELD, James A. ; DUTTON, Robert W.: Nonplanar VLSI Device Analysis Using the Solution of Poisson's Equation. In: *IEEE Journal of Solid-State Circuits* SC-15 (1980), Nr. 4, S. 585–597
- [GPS98] GRITSENKO, V. A. ; PETRENKO, I. P. ; SVITASHEVA, S. N.: Excess silicon at the  $\text{Si}_3\text{N}_4/\text{SiO}_2$  Interface. In: *Applied Physics Letters* 72 (1998), Nr. 4, S. 462–464
- [GS02] GWOZIECKI, R. ; SKOTNICKI, T.: Physics of the subthreshold slope - initial improvement and final degradation in short CMOS devices. In: *32rd Conference on European Solid-State Device Research, ESSDERC '02*. Florence, Italy, 2002, S. 639–642
- [HCT<sup>+</sup>02] HE, Ping ; CHEN, Wensong ; TIAN, Lilin ; LIU, Litian ; LI, Zhijian: Unified MOSFET Scaling Theory Using Variational Method. In: *Proc. of 23rd Internatinal Confonference on Microelectronics, MIEL 2002*. Niš, 2002, S. 491–494
- [HDF<sup>+</sup>04] HAGENBECK, R. ; DECKER, S. ; FISCHER, J.-M. ; ISLER, M. ; LAU, F. ; MIKOLAJICK, T. ; TEMPEL, G. ; HAIBACH, P.: Modeling and Simulation of Electron Injection during Programming in Twin Flash<sup>TM</sup> Devices Based on Energy Transport and the Non-Local Lucky Electron Concept. In: *10th International Workshop on Computational Electronics, IWCE-10*. Purdue University, West Lafayette, Indiana, USA, 2004, S. 155–156
- [Iñí95] IÑÍGUEZ, Benjamin: Comments on Threshold Voltage Model for Deep-Submicrometer MOSFET's. In: *IEEE Transactions on Electron Devices* 42 (1995), Nr. 9, S. 1712
- [ISS02] INGROSSO, G. ; SELMI, L. ; SANGIORGI, E.: Monte-Carlo Simulation of Program and Erase Charge Distributions in NROM<sup>TM</sup> Devices. In: *32rd Conference on European Solid-State Device Research, ESSDERC '02*. Florence, Italy, 2002, S. 187–190
- [Jan03] JANAI, Meir: Data Retention, Endurance and Acceleration Factors of NROM Devices. In: *Proceedings of the 41st Annual IEEE Internat. Reliability Physics Symposium (IRPS'03)*. Dallas, USA, 2003, S. 502–505
- [JES<sup>+</sup>04] JANAI, Meir ; EITAN, Boaz ; SHAPPIR, Assaf ; LUSKY, Eli ; BLOOM, Ilan ; COHEN, Guy: Data Retention Reliability Model of NROM Nonvolatile

- Memory Products. In: *IEEE Transactions on Device and Materials Reliability* 4 (2004), Nr. 3, S. 404–415
- [KB86] KENDALL, J. D. ; BOOTHROYD, A. R.: A Two-Dimensional Analytical Threshold Voltage Model for MOSFET's with Arbitrarily Doped Substrates. In: *IEEE Electron Device Letters* EDL-7 (1986), Nr. 7, S. 401–403
- [KIM<sup>+</sup>05] KAMIENSKI, E. G. S. ; ISLER, M. ; MIKOLAJICK, T. ; LUDWIG, C. ; SCHULZE, N. ; NAGEL, N. ; RIEDEL, S. ; WILLER, J. ; KÜSTERS, K.-H.: An Overview of Twin-Flash<sup>TM</sup> Technology. In: *Proceedings of the 6th Annual Non-Volatile Memory Technology Symposium (NVMTS 2005)*. Dallas, USA, 2005, S. 5–10
- [LAC<sup>+</sup>04] LE, Binh Q. ; ACHTER, Michael ; CHNG, Chin G. ; GUO, Xin ; CLEVELAND, Lee ; CHEN, Pau-Ling ; BUSKIRK, Michael V. ; DUTTON, Robert W.: Virtual-Ground Sensing Techniques for a 49-ns/200-MHz Access Time 1.8-V 256-Mb 2-Bit-per-Cell Flash Memory. In: *IEEE Journal of Solid-State Circuits* 39 (2004), Nr. 11, S. 2014–2022
- [Lee73] LEE, H. S.: An Analysis of the Threshold Voltage for Short-Channel IG-FET's. In: *Solid-State Electronics* 16 (1973), S. 1407–1417
- [LHH<sup>+</sup>93] LIU, Zhi-Hong ; HU, Chenming ; HUANG, Jian-Hui ; CHAN, Tung-Yi ; JENG, Min-Chie ; KO, Ping K. ; CHENG, Y. C.: Threshold Voltage Model for Deep-Submicrometer MOSFET's. In: *IEEE Transactions on Electron Devices* 40 (1993), Nr. 1, S. 86–94
- [LHW<sup>+</sup>06] LUE, Hang-Ting ; HSU, Tzu-Hsuan ; WU, Min-Ta ; HSIEH, Kuang-Yeu ; LIU, Rich ; LU, Chih-Yuan: Studies of the Reverse Read Method and Second-Bit Effect of 2-Bit/Cell Nitride-Trapping Device by Quasi-Two-Dimensional Model. In: *IEEE Transactions on Electron Devices* 53 (2006), Nr. 1, S. 119–125
- [LPE04] LARCHER, Luca ; PAVAN, Paolo ; EITAN, Boaz: On the Physical Mechanism of the NROM Memory Erase. In: *IEEE Transactions on Electron Devices* 51 (2004), Nr. 10, S. 1593–1599
- [LS72] LUNDSTRÖM, K. I. ; SVENSSON, Christer M.: Properties of MOS-Structures. In: *IEEE Transactions on Electron Devices* ED-19 (1972), Nr. 6, S. 826–836
- [LSDBE01] LUSKY, Eli ; SHACHAM-DIAMAND, Yosi ; BLOOM, Ilan ; EITAN, Boaz: Characterization of Channel Hot Electron Injection by the Subthreshold Slope of NROM<sup>TM</sup> Device. In: *IEEE Electron Device Letters* 22 (2001), Nr. 11, S. 556–558

- [LSDBE02] LUSKY, Eli ; SHACHAM-DIAMAND, Yosi ; BLOOM, Ilan ; EITAN, Boaz: Electrons Retention Model for Localized Charge in Oxide-Nitride-Oxide (ONO) Dielectric. In: *IEEE Electron Device Letters* 23 (2002), Nr. 9, S. 556–558
- [LSDM<sup>+</sup>04] LUSKY, Eli ; SHACHAM-DIAMAND, Yosi ; MITENBERG, Gill ; SHAPPIR, Assaf ; BLOOM, Ilan ; EITAN, Boaz: Investigation of Channel Hot Electron Injection by Localized Charge-Trapping Nonvolatile Memory Devices. In: *IEEE Transactions on Electron Devices* 51 (2004), Nr. 3, S. 444–451
- [LSH<sup>+</sup>04] LUE, Hang-Ting ; SHIH, Yen-Hao ; HSIEH, Kuang-Yeu ; LIU, Rich ; LU, Chih-Yuan: A Transient Analysis Method to Characterize the Trap Vertical Location in Nitride-Trapping Devices. In: *IEEE Electron Device Letters* 25 (2004), Nr. 12, S. 816–818
- [LVP<sup>+</sup>02] LARCHER, Luca ; VERZELLESI, Giovanni ; PAVAN, Paolo ; LUSKY, E. ; BLOOM, Ilan ; EITAN, Boaz: Impact of Programming Charge Distribution on Threshold Voltage and Subthreshold Slope of NROM Memory Cells. In: *IEEE Transactions on Electron Devices* 49 (2002), Nr. 11, S. 1939–1946
- [LW87] LIN, Pole-Shang ; WU, Ching-Yuan: A New Approach to Analytically Solving The Two-Dimensional Poisson's Equation and Its Application in Short-Channel MOSFET Modeling. In: *IEEE Transactions on Electron Devices* ED-34 (1987), Nr. 9, S. 1947–1956
- [MDS<sup>+</sup>02] MAAYAN, Eduardo ; DVIR, Ran ; SHOR, Joseph ; POLANSKI, Yan ; SOFER, Yair ; BLOOM, Ilan ; AVNI, Dror ; EITAN, Boaz ; COHEN, Zeev ; MEYASSED, Moshe ; ALPERN, Yair ; PALM, Herbert ; KAMIENSKI, Elard S. ; HAIBACH, Patrick ; RIEDEL, D. Caspary S. ; KNÖFLER, Roman: A 512Mb NROM Flash Data Storage Memory with 8MB/s Data Rate. In: *IEEE Internat. Solid-State Circuits Conference (ISSCC 2002)*. San Francisco, USA, 2002, S. 100–101
- [Mel06] MELDE, Thomas: *Untersuchungen an Zellstrukturen für NROM basierte Flash Speicher*, Technische Universität Dresden, Diplomarbeit, 2006
- [ML86] MÖSCHWITZER, Albrecht ; LUNZE, Klaus: *Halbleiterelektronik: Lehrbuch*. VEB Verlag Technik, Berlin, 1986
- [NOY77] NING, T. H. ; OSBURN, C. M. ; YU, H. N.: Emission probability of hot electrons from silicon into silicon dioxide. In: *Applied Physics Letters* 48 (1977), Nr. 1, S. 286–293
- [Pau72] PAUL, Reinhold: *Feldeffekttransistoren*. VEB Verlag Technik, Berlin, 1972

- [PBI<sup>+</sup>05] PERNIOLA, Luca ; BERNARDI, Sandrine ; IANNACCONE, Giuseppe ; MASSON, Pascal ; SALVO, Barbara D. ; GERARDI, Cosimo: Analytical Model of the Effects of a Nonuniform Distribution of Stored Charge on the Electrical Characteristics of Discrete-Trap Nonvolatile Memories. In: *IEEE Transactions on Nanotechnology* 4 (2005), Nr. 4, S. 360–368
- [PK84] POOLE, D. R. ; KWONG, D. L.: Two-Dimensional Analytical Modeling of Threshold Voltages of Short-Channel MOSFET's. In: *IEEE Electron Device Letters* EDL-5 (1984), Nr. 11, S. 443–446
- [PLS<sup>+</sup>06] POLANSKY, Yan ; LAVAN, Avi ; SAHAR, Ran ; DADASHEV, Oleg ; BETSER, Yoram ; COHEN, Guy ; MAAYAN, Eduardo ; EITAN, Boaz ; CHEN, Tim Chang-Ting ; CHANG, Chin-Hung ; LIAO, Chun-Yu ; CHEN, Chung-Kuang ; HO, Wen-Chiao ; SHIH, Yite ; LU, Wenpin ; TING, Wenchi ; NI, Ful-Long ; KU, Yen-Hui J. ; LU, Chih-Yuan: A 4 bits/cell NROM 1Gb Data Storage Memory. In: *IEEE Internat. Solid-State Circuits Conference (ISSCC 2006)*. San Francisco, USA, 2006
- [PM89] PIMBLEY, Joseph M. ; MEINDL, James D.: MOSFET Scaling Limits Determined by Subthreshold Conduction. In: *IEEE Transactions on Electron Devices* 36 (1989), Nr. 9, S. 1711–1721
- [PSM85] PFIESTER, James R. ; SHOTT, John D. ; MEINDL, James D.: Performance Limits of CMOS ULSI. In: *IEEE Journal of Solid-State Circuits* SC-20 (1985), Nr. 1, S. 253–263
- [RM82] RATNAKUMAR, K. N. ; MEINDL, James D.: Short-Channel MOST Threshold Voltage Model. In: *IEEE Journal of Solid-State Circuits* SC-17 (1982), Nr. 5, S. 937–948
- [RW88] RHODERICK, E. H. ; WILLIAMS, R. H.: *Metal-Semiconductor Contacts*. Second Edition. Clarendon Press, Oxford, 1988
- [SEB<sup>+</sup>04] SOFER, Yair ; EDAN, Mori ; BETSER, Yoram ; GROSSGOLD, Meir ; MAAYAN, Eduardo ; EITAN, Boaz: A 55mm<sup>2</sup> 256Mb NROM Flash Memory with Embedded Microcontroller Using an NROM-Based Program File ROM. In: *IEEE Internat. Solid-State Circuits Conference (ISSCC 2004)*. San Francisco, USA, 2004, S. 48–49
- [SF99] SELMI, Luca ; FIEGNA, Claudio: Physical Aspects of Cell Operation and Reliability. In: CAPPELLETTI, Paolo (Hrsg.) ; GOLLA, Carla (Hrsg.) ; OLIVIO, Piero (Hrsg.) ; ZANONI, Enrico (Hrsg.): *Flash Memories*. Kluwer Academic Publishers, Dordrecht, 1999, S. 153–239

- [SHIT83] SUZUKI, Eiichi ; HAYASHI, Yutaka ; ISHII, Kenichi ; TSUCHIYA, Tatsuo: Traps created at the interface between the nitride and the oxide on the nitride by thermal oxidation. In: *Applied Physics Letters* 42 (1983), Nr. 7, S. 608–610
  
- [SLCE06] SHAPPIR, Assaf ; LUSKY, Eli ; COHEN, Guy ; EITAN, Boaz: NROM Window Sensing for 2 and 4-bits per cell products. In: *21st IEEE Non-Volatile Semiconductor Memory Workshop (NVSMW 2006)*. Monterey, USA, 2006, S. 68–69
  
- [SLSD<sup>+</sup>a] SHAPPIR, Assaf ; LEVY, David ; SHACHAM-DIAMAND, Yosi ; LUSKY, Eli ; BLOOM, Ilan ; EITAN, Boaz: *Spatial Characterization of Localized Charge Trapping and Charge Redistribution in the NROM<sup>TM</sup> Device*. <http://www-leti.cea.fr/commun/europe/adamant/events/abstracts/a5-02.pdf>,
  
- [SLSD<sup>+</sup>b] SHAPPIR, Assaf ; LEVY, David ; SHACHAM-DIAMAND, Yosi ; LUSKY, Eli ; BLOOM, Ilan ; EITAN, Boaz: *Spatial Characterization of Localized Charge Trapping and Charge Redistribution in the NROM<sup>TM</sup> Device*. <http://www-leti.cea.fr/commun/europe/adamant/events/slides/p5-02.pdf>,
  
- [SMP88] SKOTNICKI, Tomasz ; MERCKEL, Gerard ; PEDRON, Thierry: The Voltage-Doping Transformation: A New Approach to the Modeling of MOSFET Short-Channel Effects. In: *IEEE Electron Device Letters* 9 (1988), Nr. 3, S. 109–112
  
- [Sze81] SZE, S. M.: *Physics of Semiconductor Devices*. Second. John Wiley & Sons, Inc., New York, 1981
  
- [TA79] TOYABE, Toru ; ASAI, Shojiro: Analytical Models of Threshold Voltage and Breakdown Voltage of Short-Channel MOSFET's Derived from Two-Dimensional Analysis. In: *IEEE Journal of Solid-State Circuits* SC-14 (1979), Nr. 2, S. 375–383
  
- [Tay78] TAYLOR, Geoffrey W.: Subthreshold Conduction MOSFET's. In: *IEEE Transactions on Electron Devices* ED-25 (1978), Nr. 3, S. 337–350
  
- [THS06] TEMPEL, G. ; HAGENBECK, R. ; STRASSBURG, M.: Quantitative Model for Data Retention Loss at NROM Nitride Charge Trapping Devices after Program / Erase Cycling. In: *21st IEEE Non-Volatile Semiconductor Memory Workshop (NVSMW 2006)*. Monterey, USA, 2006, S. 78–80

- [TR04] TOMAR, Bhawna ; RAO, V. R.: Sub-threshold Swing Degradation due to Localized Charge Storage in SONOS Memories. In: *Proc. of 11th IPFA*. Taiwan, 2004, S. 251–253
- [Tro75] TROUTMAN, R. R.: Subthreshold Slope for Insulated Gate Field-Effect Transistors. In: *IEEE Transactions on Electron Devices* ED-22 (1975), Nr. 11, S. 1049–1051
- [Tsi99] TSIVIDIS, Yannis: *Operation and modeling of the MOS transistor*. Second. McGraw-Hill, Boston, London, 1999
- [TZW<sup>+</sup>06] TSAI, Wen-Jer ; ZOUS, Nian-Kai ; WANG, Tahui ; KU, Yen-Hui J. ; LU, Chih-Yuan: A Novel Operation Method to Avoid Overerasure in a Scaled Trapping-Nitride Localized Charge Storage Flash Memory Cell and Its Application for Multilevel Programming. In: *IEEE Transactions on Electron Devices* 53 (2006), Nr. 4, S. 808–814
- [VJD97] VANDAMME, E. P. ; JANSEN, Ph. ; DEFERM, L.: Modeling the Subthreshold Swing in MOSFET's. In: *IEEE Electron Device Letters* 18 (1997), Nr. 8, S. 369–371
- [Yau74] YAU, L. D.: A Simple Theory To Predict the Threshold Voltage of Short-Channel IGFET's. In: *Solid-State Electronics* 17 (1974), S. 1059–1063

## Anhang A

### Transistormodelle für die Bauelementesimulation

Für die numerische Bauelementesimulation wurde der Simulator MINIMOS-NT benutzt [Fis94]. Die notwendigen Dateien, die die Transistorgeometrien und -dotierungen beschreiben, wurden mit dem zum Simulator gehörenden Programm MAKEDEVICE erzeugt.

Die verwendeten Transistorgeometrien, Dotierungen und Gitter wurden auf Basis der zum Programmpaket gehörenden Beispiele durch geeignete Skalierungen erzeugt.

#### A.1 Langkanal-MOS-Transistor

|                                                  |                                   |
|--------------------------------------------------|-----------------------------------|
| Art                                              | n-Kanal-MOS-Transistor            |
| nominelle Kanallänge $L$                         | $0,5\,\mu\text{m}$                |
| effektive Kanallänge $L_{eff}$                   | $0,46\,\mu\text{m}$               |
| Oxiddicke $t_{ox}$                               | $10\,\text{nm}$                   |
| Tiefe der Source-/Drain-Gebiete $t_j$            | $73\,\text{nm}$                   |
| Substratdotierung $N_A$                          | $8 \cdot 10^{16}\,\text{cm}^{-3}$ |
| Source-/Drain-Dotierung $N_D$                    | $10^{21}\,\text{cm}^{-3}$         |
| Langkanalschwellspannung $V_{th,L}$              | $0,854\,\text{V}$                 |
| Fermipotentiale $\Psi_{FB}$                      | $0,416\,\text{V}$                 |
| Flachbandspannung $V_{fb}$                       | $-0,416\,\text{V}$                |
| Diffusionsspannung $V_{bi}$                      | $1,070\,\text{V}$                 |
| Inverser Anstieg des Oberflächenpotentials $n_0$ | $1,264$                           |
| Charakteristische Länge $l_c$                    | $37,8\,\text{nm}$                 |
| Mittlere Elektronenbeweglichkeit $\bar{\mu}_n$   | $0,08\,\text{m}^2/\text{Vs}$      |



---

*MAKEDEVICE Input-File*

---

```

2      #include <mkdev.mkd>
3
4      //-----
4      // make_mos.mkd
4      // This is an example input-deck for makedevice creating a MOS device.
6      //-----
7
8      // default units are SI units, except for :
8      // o) length (default unit um)
10     // o) concentration (default unit cm-3)
11
12     Input
13     {
14         type          = "mos";
15     }
16
17     Output
18     {
19         outputPif      = "./mos_05u.pif";
20         outputEps      = "./mos_05u.eps";
21     }
22
23     Geometry
24     {
25         Mos
26         {
27             gateLength      = 0.5 um;          // gate length
28             aux gateLength_orig = 0.25 um;      // orig gate length
29             aux L            = gateLength;
30             aux L_orig       = gateLength_orig;
31             oxideThickness   = 10 nm;          // oxide thickness
32
33             sourceWellLength = 1./3.*L_orig;    // length of the Source contact
34             drainWellLength  = 1./3.*L_orig;    // length of the Drain contact
35             sourceSpacerLength = 1./3.*L_orig;  // length of the Source spacer
36             drainSpacerLength = 1./3.*L_orig;   // length of the Drain spacer
37             bulkThickness    = 0.4 um;          // bulk semiconductor thickness
38             contactThickness = 0.05 um;        // contact thickness
39         }
40     }
41
42     Doping
43     {
44         +Bulk : SegmentDefaults
45         {
46             constAcceptor = 8e16 "cm-3";      // bulk acceptor doping
47             constDonor    = 1e12 "cm-3";      // bulk donor doping
48
49             Peak
50             {
51                 +Source : PeakDefaults
52                 {
53                     aux L = ~Geometry.Mos.gateLength;
54                     aux L_orig = ~Geometry.Mos.gateLength_orig;
55                     aux SW = ~Geometry.Mos.sourceWellLength;
56                     aux SG = ~Geometry.Mos.sourceSpacerLength;
57
58                     on      = yes;              // ~usePeak1; switch on or off,
59                     mode    = "gauss";          // gauss, cosine, or pearson mode
60                     dopType = "donor";          // acceptor or donor doping
61
62                     N        = 1e21 "cm-3";    // peak doping value

```

```

        x          = 0 um;           // x-position of the peak
64      xLength    = SW + SG;         // x-length of the peak, must be >= 0
        y          = 0 um;           // y-position of the peak
66      yLength    = 0 um;           // y-length of the peak, must be >= 0
        xSigLeft   = L_orig/60.;      // left x-sigma of the peak
68      xSigRight  = xSigLeft;        // right x-sigma of the peak
        ySigUpper  = L_orig/15.;      // upper y-sigma of the peak
70      ySigLower  = ySigUpper;       // lower y-sigma of the peak
    }
+Drain : Source
{
74      aux L      = ~Geometry.Mos.gateLength;
        aux L_orig = ~Geometry.Mos.gateLength_orig;
76      aux SW     = ~Geometry.Mos.sourceWellLength;
        aux SG     = ~Geometry.Mos.sourceSpacerLength;
78      aux DW     = ~Geometry.Mos.drainWellLength;
        aux DG     = ~Geometry.Mos.drainSpacerLength;
80
        on        = yes;
82      x          = SW + SG + L;      // x-position of the drain peak
        xLength    = DW + DG;         // x-length of the peak, must be >= 0
84    }
}
}
}
Grid
{
90  {
    Mos
    {
92      {
        Dop
94      {
            aux L      = ~Geometry.Mos.gateLength;
96            aux L_orig = ~Geometry.Mos.gateLength_orig;

            useSimGrid = yes;          // whether to use the simulation grid
                                         // for the doping grid,
100                                         // if off, all following parameters are not
                                         // required
102            vDistDop  = L_orig/100.; // distance of vertical doping grid lines
                                         // in the channel and doping regions
104            vDist     = L_orig/5.;   // distance of vert. dop. grid lines elsewhere
            lDistCh     = L/50.;       // distance of lateral doping grid lines
106                                         // in the channel region
            lDistSD     = L/5.;        // distance of lateral doping grid lines
108                                         // in the Source and Drain regions
        }
    }
110  Sim
    {
112        aux L      = ~Geometry.Mos.gateLength;
            aux L_orig = ~Geometry.Mos.gateLength_orig;
114        aux SW     = ~Geometry.Mos.sourceWellLength;
            aux SG     = ~Geometry.Mos.sourceSpacerLength;
116        aux BT     = ~Geometry.Mos.bulkThickness;
            aux Nx     = ~Doping.Bulk.Knots.Nx;
118        aux Ny     = ~Doping.Bulk.Knots.Ny;

120        /* keywords for vertical grid <> y direction */
            /* ----- */
122
            name = "SimulationGrid";    // simulation grid name
124
            vNumberOx = 10;             // number of vert. simulation grid lines

```

---

```

126          vIncCont      = 1.6;          // in the oxide
127          vDistCh       = 0.2 nm;      // incremental factor in the contact
128          // distance of vertical simulation grid
129          // lines in the channel region
130          vNumberCh      = 15;          // number of vert. simulation grid lines
131          // in the channel region
132          vInc           = 1.5;          // vertical incremental factor elsewhere
133          vDistMin       = L/150.;      // minimum distance of vert. simulation
134          // grid lines outside the channel region
135          vDistMax       = BT;          // maximum distance of vert. simulation
136          // grid lines

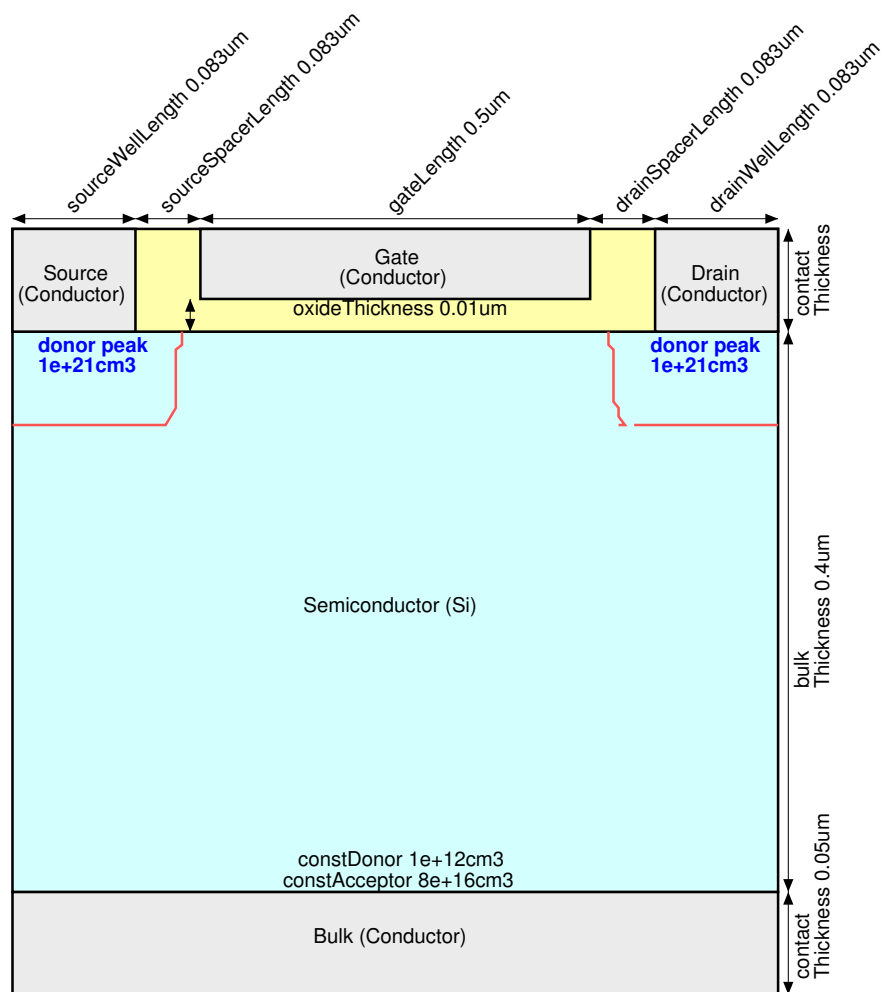
137          /* keywords for lateral grid <> x direction */
138          /* ----- */
139          chBorderLeft   = SW+SG-10 nm;  // left border of the channel region
140          chBorderRight  = SW+SG+L+10 nm; // right border of the channel region
141          lDistChBorder  = L/80.;        // distance of lat. simulation grid lines
142          // at the left or right channel border
143          lDistChMiddle  = L/20.;        // distance of lat. simulation grid lines
144          // in the middle of the channel
145          lDistMin       = L/100.;      // minimum distance of lat. simulation
146          // grid lines
147          lDistMax       = (SW + SG)/10.; // maximum distance of
148          // lateral simulation grid lines

149          /* knots keywords for simulation grid */
150          /* ----- */
151          lNumberKnots   = 3 * Nx;       // minimum number of lateral grid lines
152          // in the knots region
153          vNumberKnots   = 3 * Ny;       // minimum number of vertical grid lines
154          // in the knots region
155          xticks         = [0];          // array of additional gridpoints in x-dir
156          yticks         = [0];          // values in um
157          }
158          // array of additional gridpoints in y-dir
159          // values in um
160          }
161          Refinement
162          {
163          +Refinement1 : LateralRefinementDefaults
164          {
165          aux L = ~Geometry.Mos.gateLength;
166          aux SW = ~Geometry.Mos.sourceWellLength;
167          aux SG = ~Geometry.Mos.sourceSpacerLength;

168          on
169          = no;
170          leftBorder = SW+SG+0.7*L; // the left border of the refinement region
171          rightBorder = SW+SG+0.9*L; // the right border of the refinement region
172          distance = L/80.; // the refined maximum grid distance
173          }
174          +Refinement2 : VerticalRefinementDefaults
175          {
176          on
177          = no;
178          upperBorder = 0 nm; // the upper of the refinement region
179          lowerBorder = -15. nm; // the lower border of the refinement region
180          distance = 0.5 nm; // the refined maximum grid distance
181          }
182          }
183          }

```

---



**Abbildung A.1:** Geometrie und Dotierung des Langkanal-MOS-Transistors

## A.2 Kurzkanal-MOS-Transistor

|                                                  |                                   |
|--------------------------------------------------|-----------------------------------|
| Art                                              | n-Kanal-MOS-Transistor            |
| nomielle Kanallänge $L$                          | 0,18 $\mu\text{m}$                |
| effektive Kanallänge $L_{\text{eff}}$            | 0,14 $\mu\text{m}$                |
| Oxiddicke $t_{\text{ox}}$                        | 10 nm                             |
| Tiefe der Source-/Drain-Gebiete $t_j$            | 73 nm                             |
| Substratdotierung $N_A$                          | $8 \cdot 10^{16} \text{ cm}^{-3}$ |
| Source-/Drain-Dotierung $N_D$                    | $10^{21} \text{ cm}^{-3}$         |
| Langkanalschwellspannung $V_{th,L}$              | 0,854 V                           |
| Fermipotentia $\Psi_{FB}$                        | 0,416 V                           |
| Flachbandspannung $V_{fb}$                       | -0,416 V                          |
| Diffusionsspannung $V_{bi}$                      | 1,070 V                           |
| Inverser Anstieg des Oberflächenpotentials $n_0$ | 1,264                             |
| Charakteristische Länge $l_c$                    | 37,8 nm                           |
| Mittlere Elektronenbeweglichkeit $\bar{\mu}_n$   | 0,08 $\text{m}^2/\text{Vs}$       |

```

----- MAKEDEVICE Input-File -----
#include <mkdev.mkd>

2
//-----
4 // make_mos.mkd
// This is an example input-deck for makedevice creating a MOS device.
6 //-----

8 // default units are SI units, except for :
// o) length (default unit um)
10 // o) concentration (default unit cm^-3)

12 Input
{
14     type          = "mos";
}

16 Output
18 {
    outputPif      = "./mos_018u.pif";
20     outputEps    = "./mos_018u.eps";
}

22

24 Geometry
{
26     Mos
    {
28         gateLength      = 0.18 um;           // gate length
        aux gateLength_orig = 0.25 um;         // orig gate length
    }
}

```

```

30     aux L          = gateLength;
31     aux L_orig      = gateLength_orig;
32     oxideThickness  = 10 nm;          // oxide thickness

34     sourceWellLength = 1./3.*L_orig;    // length of the Source contact
35     drainWellLength  = 1./3.*L_orig;    // length of the Drain contact
36     sourceSpacerLength = 1./3.*L_orig;  // length of the Source spacer
37     drainSpacerLength = 1./3.*L_orig;    // length of the Drain spacer
38     bulkThickness    = 0.4 um;          // bulk semiconductor thickness
39     contactThickness  = 0.05 um;        // contact thickness
40 }
41 }
42
43 Doping
44 {
45     +Bulk : SegmentDefaults
46     {
47         constAcceptor = 8e16 "cm^-3";    // bulk acceptor doping
48         constDonor    = 1e12 "cm^-3";    // bulk donor doping
49
50     Peak
51     {
52         +Source : PeakDefaults
53         {
54             aux L = ~Geometry.Mos.gateLength;
55             aux L_orig = ~Geometry.Mos.gateLength_orig;
56             aux SW = ~Geometry.Mos.sourceWellLength;
57             aux SG = ~Geometry.Mos.sourceSpacerLength;
58
59             on      = yes;                // ~usePeak1; switch on or off,
60             mode    = "gauss";            // gauss, cosine, or pearson mode
61             dopType = "donor";            // acceptor or donor doping
62
63             N        = 1e21 "cm^-3";      // peak doping value
64             x         = 0 um;              // x-position of the peak
65             xLength   = SW + SG;           // x-length of the peak, must be >= 0
66             y         = 0 um;              // y-position of the peak
67             yLength   = 0 um;              // y-length of the peak, must be >= 0
68             xSigLeft  = L_orig/60.;        // left x-sigma of the peak
69             xSigRight = xSigLeft;          // right x-sigma of the peak
70             ySigUpper = L_orig/15.;        // upper y-sigma of the peak
71             ySigLower = ySigUpper;         // lower y-sigma of the peak
72         }
73         +Drain : Source
74         {
75             aux L = ~Geometry.Mos.gateLength;
76             aux L_orig = ~Geometry.Mos.gateLength_orig;
77             aux SW = ~Geometry.Mos.sourceWellLength;
78             aux SG = ~Geometry.Mos.sourceSpacerLength;
79             aux DW = ~Geometry.Mos.drainWellLength;
80             aux DG = ~Geometry.Mos.drainSpacerLength;
81
82             on      = yes;
83             x        = SW + SG + L;        // x-position of the drain peak
84             xLength  = DW + DG;            // x-length of the peak, must be >= 0
85         }
86     }
87 }
88
89 Grid
90 {
91     Mos

```

```

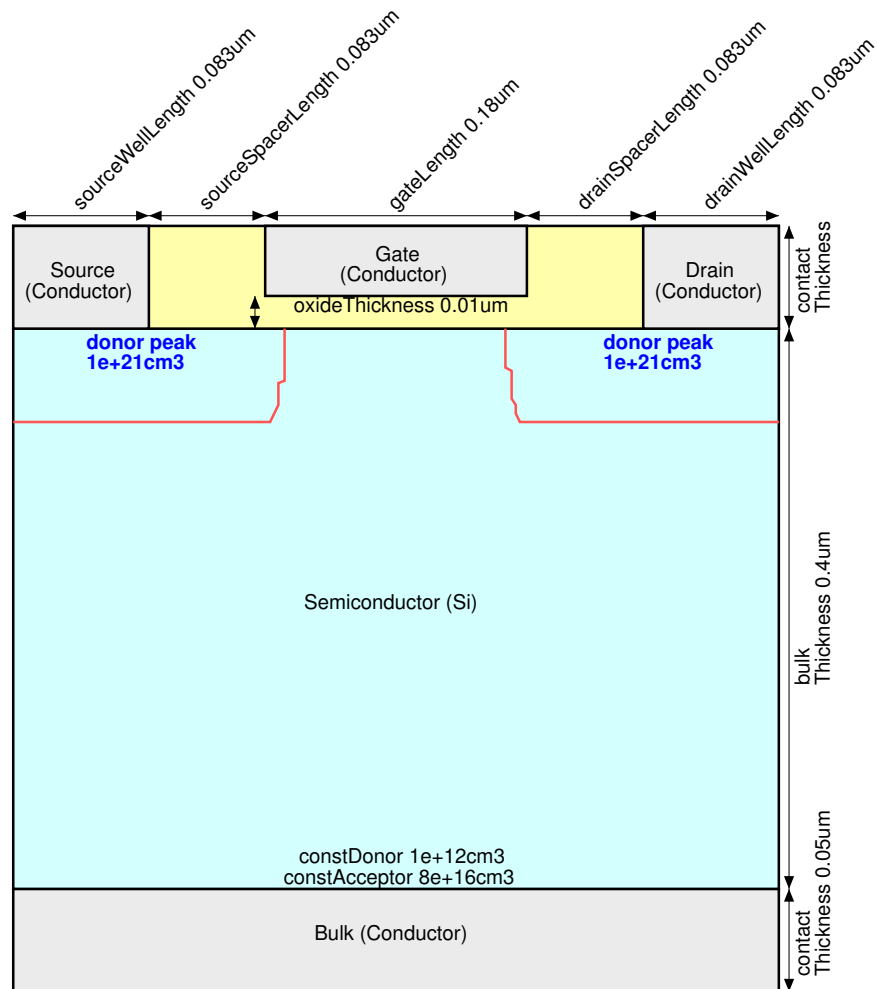
94     {
95         Dop
96         {
97             aux L      = ~Geometry.Mos.gateLength;
98             aux L_orig = ~Geometry.Mos.gateLength_orig;
99
100             useSimGrid = yes;      // whether to use the simulation grid
101                                     // for the doping grid,
102                                     // if off, all following parameters are not
103                                     // required
104             vDistDop   = L_orig/100.; // distance of vertical doping grid lines
105                                     // in the channel and doping regions
106             vDist      = L_orig/5.;  // dist. of vert. dop. grid lines elsewhere
107             lDistCh     = L/50.;     // distance of lateral doping grid lines
108                                     // in the channel region
109             lDistSD     = L/5.;      // distance of lateral doping grid lines
110                                     // in the Source and Drain regions
111         }
112         Sim
113         {
114             aux L      = ~Geometry.Mos.gateLength;
115             aux L_orig = ~Geometry.Mos.gateLength_orig;
116             aux SW     = ~Geometry.Mos.sourceWellLength;
117             aux SG     = ~Geometry.Mos.sourceSpacerLength;
118             aux BT     = ~Geometry.Mos.bulkThickness;
119             aux Nx     = ~Doping.Bulk.Knots.Nx;
120             aux Ny     = ~Doping.Bulk.Knots.Ny;
121
122             /* keywords for vertical grid <> y direction */
123             /* ----- */
124
125             name = "SimulationGrid"; // simulation grid name
126
127             vNumberOx   = 10;        // number of vert. simulation grid lines
128                                     // in the oxide
129             vIncCont     = 1.6;       // incremental factor in the contact
130             vDistCh      = 0.2 nm;   // distance of vertical simulation grid
131                                     // lines in the channel region
132             vNumberCh    = 15;       // number of vert. simulation grid lines
133                                     // in the channel region
134             vInc         = 1.5;       // vertical incremental factor elsewhere
135             vDistMin     = L/150.;    // minimum distance of vert. simulation
136                                     // grid lines outside the channel region
137             vDistMax     = BT;        // maximum distance of vert. simulation
138                                     // grid lines
139
140             /* keywords for lateral grid <> x direction */
141             /* ----- */
142
143             chBorderLeft = SW+SG-10 nm; // left border of the channel region
144             chBorderRight = SW+SG+L+10 nm; // right border of the channel region
145             lDistChBorder = L/80.;      // distance of lat. simulation grid lines
146                                     // at the left or right channel border
147             lDistChMiddle = L/20.;      // distance of lat. simulation grid lines
148                                     // in the middle of the channel
149             lDistMin      = L/100.;     // minimum distance of lat. simulation
150                                     // grid lines
151             lDistMax      = (SW + SG)/10.; // maximum distance of
152                                     // lateral simulation grid lines
153
154             /* knots keywords for simulation grid */
155             /* ----- */

```

```
156         lNumberKnots = 3 * Nx;           // minimum number of lateral grid lines
                                           // in the knots region
158         vNumberKnots = 3 * Ny;           // minimum number of vertical grid lines
                                           // in the knots region
160
161         xticks        = [0];              // array of additional gridpoints in x-dir
162         yticks        = [0];              // values in um
163     }                                     // array of additional gridpoints in y-dir
164     }                                     // values in um
165     Refinement
166     {
167         +Refinement1 : LateralRefinementDefaults
168         {
169             aux L = ~Geometry.Mos.gateLength;
170             aux SW = ~Geometry.Mos.sourceWellLength;
171             aux SG = ~Geometry.Mos.sourceSpacerLength;
172
173             on      = no;
174             leftBorder = SW+SG+0.7*L; // the left border of the refinement region
175             rightBorder = SW+SG+0.9*L; // the right border of the refinement region
176             distance  = L/80.;        // the refined maximum grid distance
177         }
178         +Refinement2 : VerticalRefinementDefaults
179         {
180             on      = no;
181             upperBorder = 0 nm;        // the upper of the refinement region
182             lowerBorder = -15. nm;     // the lower border of the refinement region
183             distance  = 0.5 nm;        // the refined maximum grid distance
184         }
185     }
186 }
```

---





**Abbildung A.2:** Geometrie und Dotierung des Kurzkanal-MOS-Transistors

### A.3 Langkanal-ONO-Transistor

|                                                         |                                  |
|---------------------------------------------------------|----------------------------------|
| Art                                                     | n-Kanal-ONO-Transistor           |
| nominelle Kanallänge $L$                                | $0,5 \mu\text{m}$                |
| effektive Kanallänge $L_{eff}$                          | $0,46 \mu\text{m}$               |
| Obere Oxiddicke $t_{ox,t}$                              | 4 nm                             |
| Nitridicke $t_{nit}$                                    | 4 nm                             |
| Untere Oxiddicke $t_{ox,b}$                             | 4 nm                             |
| Tiefe der Source-/Drain-Gebiete $t_j$                   | 73 nm                            |
| Substratdotierung $N_A$                                 | $8 \cdot 10^{16} \text{cm}^{-3}$ |
| Source-/Drain-Dotierung $N_D$                           | $10^{21} \text{cm}^{-3}$         |
| Langkanalschwellspannung $V_{th,L}$                     | 0,854 V                          |
| Fermipotential $\Psi_{FB}$                              | 0,416 V                          |
| Diffusionsspannung $V_{bi}$                             | 1,070 V                          |
| Inverser Anstieg des Oberflächenpotentials $n_0$        | 1,264                            |
| Charakteristische Länge $l_c$                           | 37,8 nm                          |
| Ausdehnung der Injektionsgebiete $\lambda_s, \lambda_d$ | 30 nm                            |
| Mittlere Elektronenbeweglichkeit $\bar{\mu}_n$          | $0,08 \text{m}^2/\text{Vs}$      |

```

_____ MAKEDEVICE Input-File _____
#include <mkdev.mkd>

2
//-----
4 // make_mos.mkd
// This is an example input-deck for makedevice creating a MOS device.
6 //-----

8 // default units are SI units, except for :
// o) length (default unit um)
10 // o) concentration (default unit cm^-3)

12 Input
{
14     type          = "mosStack3";
}

16 Output
18 {
    outputPif      = "./nrom_05u.pif";
20     outputEps     = "./nrom_05u.eps";
}

22

24 Geometry

```

```

26 {
27   MosStack3
28   {
29     gateLength      = 0.5 um;           // gate length
30     aux gateLength_orig = 0.25 um;       // orig gate length
31     aux L            = gateLength;
32     aux L_orig       = gateLength_orig;
33     oxideThickness   = 12 nm;           // total oxide thickness = (4+4+4)nm
34
35     sourceWellLength = 1./3.*L_orig;     // length of the Source contact
36     drainWellLength  = 1./3.*L_orig;     // length of the Drain contact
37     sourceSpacerLength = 1./3.*L_orig;   // length of the Source spacer
38     drainSpacerLength = 1./3.*L_orig;    // length of the Drain spacer
39     bulkThickness     = 0.4 um;          // bulk semiconductor thickness
40     contactThickness  = 0.05 um;        // contact thickness
41     upperOxideThickness = 4 nm;         // upper oxide thickness
42     middleOxideThickness = 4 nm;        // nitride thickness
43   }
44 }
45
46 Doping
47 {
48   +Bulk : SegmentDefaults
49   {
50     constAcceptor = 8e16 "cm^-3";       // bulk acceptor doping
51     constDonor    = 1e12 "cm^-3";       // bulk donor doping
52
53   Peak
54   {
55     +Source : PeakDefaults
56     {
57       aux L = ~Geometry.MosStack3.gateLength;
58       aux L_orig = ~Geometry.MosStack3.gateLength_orig;
59       aux SW = ~Geometry.MosStack3.sourceWellLength;
60       aux SG = ~Geometry.MosStack3.sourceSpacerLength;
61
62       on      = yes;                     // ~usePeak1; switch on or off,
63       mode    = "gauss";                 // gauss, cosine, or pearson mode
64       dopType = "donor";                 // acceptor or donor doping
65
66       N       = 1e21 "cm^-3";           // peak doping value
67       x       = 0 um;                   // x-position of the peak
68       xLength = SW + SG;                 // x-length of the peak, must be >= 0
69       y       = 0 um;                   // y-position of the peak
70       yLength = 0 um;                   // y-length of the peak, must be >= 0
71       xSigLeft = L_orig/60.;             // left x-sigma of the peak
72       xSigRight = xSigLeft;              // right x-sigma of the peak
73       ySigUpper = L_orig/15.;            // upper y-sigma of the peak
74       ySigLower = ySigUpper;             // lower y-sigma of the peak
75     }
76     +Drain : Source
77     {
78       aux L = ~Geometry.MosStack3.gateLength;
79       aux L_orig = ~Geometry.MosStack3.gateLength_orig;
80       aux SW = ~Geometry.MosStack3.sourceWellLength;
81       aux SG = ~Geometry.MosStack3.sourceSpacerLength;
82       aux DW = ~Geometry.MosStack3.drainWellLength;
83       aux DG = ~Geometry.MosStack3.drainSpacerLength;
84
85       on      = yes;
86       x       = SW + SG + L;             // x-position of the drain peak
87       xLength = DW + DG;                 // x-length of the peak, must be >= 0
88     }
89   }
90 }

```

```

88     }
89   }
90 }

92 Grid
93 {
94   MosStack3
95   {
96     Dop
97     {
98       aux L      = ~Geometry.MosStack3.gateLength;
99       aux L_orig = ~Geometry.MosStack3.gateLength_orig;
100
101       useSimGrid = yes;      // whether to use the simulation grid
102                               // for the doping grid,
103                               // if off, all following parameters are not
104                               // required
105       vDistDop   = L_orig/100.; // distance of vertical doping grid lines
106                               // in the channel and doping regions
107       vDist      = L_orig/5.;   // distance of vert. dop. grid lines elsewhere
108       lDistCh    = L/50.;      // distance of lateral doping grid lines
109                               // in the channel region
110       lDistSD    = L/5.;       // distance of lateral doping grid lines
111                               // in the Source and Drain regions
112     }
113     Sim
114     {
115       aux L      = ~Geometry.MosStack3.gateLength;
116       aux L_orig = ~Geometry.MosStack3.gateLength_orig;
117       aux SW     = ~Geometry.MosStack3.sourceWellLength;
118       aux SG     = ~Geometry.MosStack3.sourceSpacerLength;
119       aux BT     = ~Geometry.MosStack3.bulkThickness;
120       aux Nx     = ~Doping.Bulk.Knots.Nx;
121       aux Ny     = ~Doping.Bulk.Knots.Ny;
122
123       /* keywords for vertical grid <> y direction */
124       /* ----- */
125
126       name = "SimulationGrid";      // simulation grid name
127
128       vNumberOx   = 10;             // number of vert. simulation grid lines
129                                     // in the oxide
130       vIncCont    = 1.6;            // incremental factor in the contact
131       vDistCh     = 0.2 nm;        // distance of vert. simulation grid
132                                     // lines in the channel region
133       vNumberCh   = 15;            // number of vert. simulation grid lines
134                                     // in the channel region
135       vInc        = 1.5;            // vertical incremental factor elsewhere
136       vDistMin    = L/150.;        // minimum distance of vert. simulation
137                                     // grid lines outside the channel region
138       vDistMax    = BT;            // maximum distance of vert. simulation
139                                     // grid lines
140
141       /* keywords for lateral grid <> x direction */
142       /* ----- */
143       chBorderLeft = SW+SG-10 nm;   // left border of the channel region
144       chBorderRight = SW+SG+L+10 nm; // right border of the channel region
145       lDistChBorder = L/80.;        // distance of lat. simulation grid lines
146                                     // at the left or right channel border
147       lDistChMiddle = L/20.;        // distance of lat. simulation grid lines
148                                     // in the middle of the channel
149       lDistMin      = L/100.;       // minimum distance of lat. simulation
150                                     // grid lines

```

```

152         lDistMax      = (SW + SG)/10.; // maximum distance of
                                         // lateral simulation grid lines

154         /* knots keywords for simulation grid          */
         /* ----- */

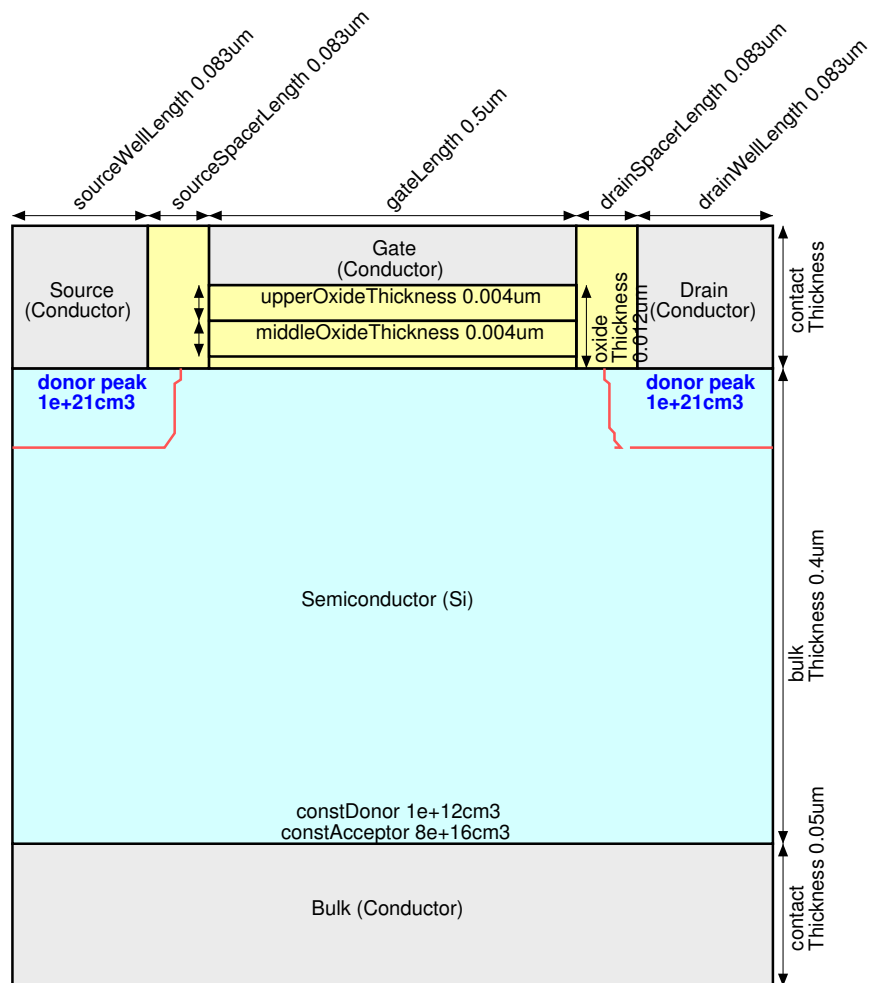
156         lNumberKnots  = 3 * Nx;          // minimum number of lateral grid lines
158                                         // in the knots region
         vNumberKnots  = 3 * Ny;          // minimum number of vertical grid lines
160                                         // in the knots region

162         xticks       = [0];             // array of additional gridpoints in x-dir.
         yticks       = [0];             // values in um
164     }                                     // array of additional gridpoints in y-dir.
                                         // values in um
166     Refinement
    {
168         +Refinement1 : LateralRefinementDefaults
        {
170             aux L = ~Geometry.MosStack3.gateLength;
172             aux SW = ~Geometry.MosStack3.sourceWellLength;
174             aux SG = ~Geometry.MosStack3.sourceSpacerLength;

174             on          = no;
176             leftBorder  = SW+SG+0.7*L; // the left border of the refinement region
176             rightBorder = SW+SG+0.9*L; // the right border of the refinement region
178             distance    = L/80.;      // the refined maximum grid distance
        }
178         +Refinement2 : VerticalRefinementDefaults
180         {
182             on          = no;
182             upperBorder = 0 nm;        // the upper of the refinement region
184             lowerBorder = -15. nm;     // the lower border of the refinement region
184             distance    = 0.5 nm;     // the refined maximum grid distance
        }
186     }
    }
188

```

---



**Abbildung A.3:** Geometrie und Dotierung des Langkanal-ONO-Transistors

## A.4 Kurzkanal-ONO-Transistor

|                                                         |                                   |
|---------------------------------------------------------|-----------------------------------|
| Art                                                     | n-Kanal-ONO-Transistor            |
| nomielle Kanallänge $L$                                 | $0,18\ \mu\text{m}$               |
| effektive Kanallänge $L_{\text{eff}}$                   | $0,14\ \mu\text{m}$               |
| Obere Oxiddicke $t_{\text{ox},t}$                       | 4 nm                              |
| Nitridicke $t_{\text{nit}}$                             | 4 nm                              |
| Untere Oxiddicke $t_{\text{ox},b}$                      | 4 nm                              |
| Tiefe der Source-/Drain-Gebiete $t_j$                   | 73 nm                             |
| Substratdotierung $N_A$                                 | $8 \cdot 10^{16}\ \text{cm}^{-3}$ |
| Source-/Drain-Dotierung $N_D$                           | $10^{21}\ \text{cm}^{-3}$         |
| Langkanalschwellspannung $V_{\text{th},L}$              | 0,854 V                           |
| Fermipotential $\Psi_{FB}$                              | 0,416 V                           |
| Flachbandspannung $V_{fb}$                              | -0,416 V                          |
| Diffusionsspannung $V_{bi}$                             | 1,070 V                           |
| Inverser Anstieg des Oberflächenpotentials $n_0$        | 1,264                             |
| Charakteristische Länge $l_c$                           | 37,8 nm                           |
| Ausdehnung der Injektionsgebiete $\lambda_s, \lambda_d$ | 30 nm                             |
| Mittlere Elektronenbeweglichkeit $\bar{\mu}_n$          | $0,08\ \text{m}^2/\text{Vs}$      |

```

_____ MAKEDEVICE Input-File _____
#include <mkdev.mkd>

2
//-----
4 // make_mos.mkd
// This is an example input-deck for makedevice creating a MOS device.
6 //-----

8 // default units are SI units, except for :
// o) length (default unit um)
10 // o) concentration (default unit cm^-3)

12 Input
{
14     type          = "mosStack3";
}

16
Output
18 {
    outputPif       = "./nrom_018u.pif";
20     outputEps      = "./nrom_018u.eps";
}

22

```

```

24 Geometry
{
26   MosStack3
    {
28       gateLength      = 0.18 um;           // gate length
        aux gateLength_orig = 0.25 um;       // orig gate length
30       aux L           = gateLength;
        aux L_orig      = gateLength_orig;
32       oxideThickness  = 12 nm;           // total oxide thickness = (4+4+4)nm

34       sourceWellLength = 1./3.*L_orig;     // length of the Source contact
        drainWellLength  = 1./3.*L_orig;     // length of the Drain contact
36       sourceSpacerLength = 1./3.*L_orig;   // length of the Source spacer
        drainSpacerLength = 1./3.*L_orig;    // length of the Drain spacer
38       bulkThickness    = 0.4 um;          // bulk semiconductor thickness
        contactThickness  = 0.05 um;        // contact thickness
40       upperOxideThickness = 4 nm;         // upper oxide thickness
        middleOxideThickness = 4 nm;        // nitride thickness
42   }
}

44 Doping
{
46   {
        +Bulk : SegmentDefaults
48     {
        constAcceptor = 8e16 "cm^-3";       // bulk acceptor doping
50     constDonor      = 1e12 "cm^-3";       // bulk donor doping

52     Peak
        {
54         +Source : PeakDefaults
            {
56             aux L = ~Geometry.MosStack3.gateLength;
                aux L_orig = ~Geometry.MosStack3.gateLength_orig;
58             aux SW = ~Geometry.MosStack3.sourceWellLength;
                aux SG = ~Geometry.MosStack3.sourceSpacerLength;

60             on      = yes;                 // ~usePeak1; switch on or off,
62             mode     = "gauss";            // gauss, cosine, or pearson mode
                dopType = "donor";           // acceptor or donor doping

64             N        = 1e21 "cm^-3";      // peak doping value
                x        = 0 um;             // x-position of the peak
66             xLength  = SW + SG;           // x-length of the peak, must be >= 0
                y        = 0 um;             // y-position of the peak
68             yLength  = 0 um;             // y-length of the peak, must be >= 0
                xSigLeft = L_orig/60.;       // left x-sigma of the peak
                xSigRight = xSigLeft;        // right x-sigma of the peak
70             ySigUpper = L_orig/15.;       // upper y-sigma of the peak
                ySigLower = ySigUpper;       // lower y-sigma of the peak
72             }
        }
74     +Drain : Source
        {
76         aux L = ~Geometry.MosStack3.gateLength;
            aux L_orig = ~Geometry.MosStack3.gateLength_orig;
78         aux SW = ~Geometry.MosStack3.sourceWellLength;
            aux SG = ~Geometry.MosStack3.sourceSpacerLength;
80         aux DW = ~Geometry.MosStack3.drainWellLength;
            aux DG = ~Geometry.MosStack3.drainSpacerLength;

82         on      = yes;
            x      = SW + SG + L;           // x-position of the drain peak

```



```

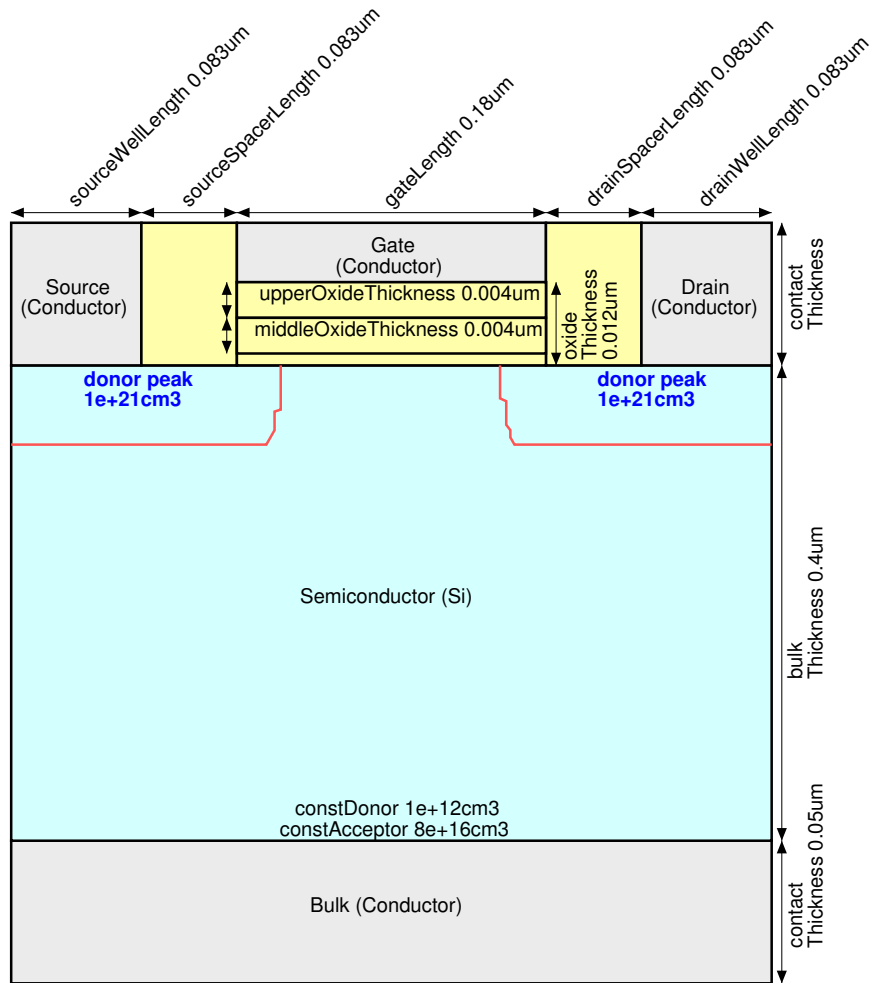
86         xLength   = DW + DG;           // x-length of the peak, must be >= 0
87     }
88 }
89 }
90 }

92 Grid
93 {
94     MosStack3
95     {
96         Dop
97         {
98             aux L      = ~Geometry.MosStack3.gateLength;
99             aux L_orig = ~Geometry.MosStack3.gateLength_orig;
100
101             useSimGrid = yes;           // whether to use the simulation grid
102                                         // for the doping grid,
103                                         // if off, all following parameters are not
104                                         // required
105             vDistDop   = L_orig/100.;   // distance of vertical doping grid lines
106                                         // in the channel and doping regions
107             vDist      = L_orig/5.;     // distance of vert. dop. grid lines elsewhere
108             lDistCh    = L/50.;        // distance of lateral doping grid lines
109                                         // in the channel region
110             lDistSD    = L/5.;         // distance of lateral doping grid lines
111                                         // in the Source and Drain regions
112         }
113     Sim
114     {
115         aux L      = ~Geometry.MosStack3.gateLength;
116         aux L_orig = ~Geometry.MosStack3.gateLength_orig;
117         aux SW     = ~Geometry.MosStack3.sourceWellLength;
118         aux SG     = ~Geometry.MosStack3.sourceSpacerLength;
119         aux BT     = ~Geometry.MosStack3.bulkThickness;
120         aux Nx     = ~Doping.Bulk.Knots.Nx;
121         aux Ny     = ~Doping.Bulk.Knots.Ny;
122
123         /* keywords for vertical grid <> y direction */
124         /* ----- */
125
126         name = "SimulationGrid";        // simulation grid name
127
128         vNumberOx   = 10;               // number of vert. simulation grid lines
129                                         // in the oxide
130         vIncCont    = 1.6;              // incremental factor in the contact
131         vDistCh     = 0.2 nm;          // distance of vertical simulation grid
132                                         // lines in the channel region
133         vNumberCh   = 15;              // number of vert. simulation grid lines
134                                         // in the channel region
135         vInc        = 1.5;              // vertical incremental factor elsewhere
136         vDistMin    = L/150.;          // minimum distance of vert. simulation
137                                         // grid lines outside the channel region
138         vDistMax    = BT;              // maximum distance of vert. simulation
139                                         // grid lines
140
141         /* keywords for lateral grid <> x direction */
142         /* ----- */
143
144         chBorderLeft = SW+SG-10 nm;     // left border of the channel region
145         chBorderRight = SW+SG+L+10 nm; // right border of the channel region
146         lDistChBorder = L/80.;          // distance of lat. simulation grid lines
147                                         // at the left or right channel border
148         lDistChMiddle = L/20.;         // distance of lat. simulation grid lines

```

```
150         lDistMin      = L/100.;           // in the middle of the channel
                                           // minimum distance of lat. simulation
                                           // grid lines
152         lDistMax      = (SW + SG)/10.;    // maximum distance of
                                           // lateral simulation grid lines
154
155         /* knots keywords for simulation grid      */
156         /* ----- */
157
158         lNumberKnots    = 3 * Nx;           // minimum number of lateral grid lines
                                           // in the knots region
160         vNumberKnots    = 3 * Ny;           // minimum number of vertical grid lines
                                           // in the knots region
162
163         xticks          = [0];              // array of additional gridpoints in x-dir.
164         yticks          = [0];              // values in um
                                           // array of additional gridpoints in y-dir.
                                           // values in um
166     }
167     Refinement
168     {
169         +Refinement1 : LateralRefinementDefaults
170         {
171             aux L = ~Geometry.MosStack3.gateLength;
172             aux SW = ~Geometry.MosStack3.sourceWellLength;
173             aux SG = ~Geometry.MosStack3.sourceSpacerLength;
174
175             on      = no;
176             leftBorder = SW+SG+0.7*L; // the left border of the refinement region
177             rightBorder = SW+SG+0.9*L; // the right border of the refinement region
178             distance  = L/80.;        // the refined maximum grid distance
179         }
180         +Refinement2 : VerticalRefinementDefaults
181         {
182             on      = no;
183             upperBorder = 0 nm;        // the upper of the refinement region
184             lowerBorder = -15. nm;     // the lower border of the refinement region
185             distance  = 0.5 nm;        // the refined maximum grid distance
186         }
187     }
188 }
```

---



**Abbildung A.4:** Geometrie und Dotierung des Kurzkanal-ONO-Transistors

## Anhang B

### MINIMOS-NT-Steuerdateien

Das folgende Beispiel zeigt eine Steuerdatei des Bauelementesimulators MINIMOS-NT zur Simulation der Kennlinie  $I_{ds} = f(V_{gs})$  eines ONO-Transistors mit  $\Delta V_{fb,s} = 2,5V$ :

```
_____ MINIMOS-NT Input-File _____
#include <defaults.ipd>

2
Extern {
4     aux defaultStepString = if (firstStep, "", "s" + step);
    }
6

8 Device : DeviceDefaults
{
10     Input { file = "nrom_05u"; }

12     Output { file = "nrom_05u_0V_subth_out_sourceside_15";
        attributes = defaultInitAttributesDD;
14         //attributes = "";
        }

16     Log
        {
18         easModels = true;
        easEvalModels = false;
20         physModels = true;
        checkPhys = true;
22     }

24     +Source = 0.0 V;
    // +Drain = step(0.0 V, 1.5 V, 0.5 V, pri = 2, name = "Vd");
26     +Drain = 1.5 V;
    +ControlGate = step(0.0 V, 1.8 V, 0.1 V, pri = 1, name = "Vg");
28     +Bulk = 0.0 V;

30     Phys {
    // PolyGate defined! -> Obsolete: +Gate { Contact { Ohmic { Ew = -0.55 eV; }}}
32         srh = "Semi*";
        iChBd = "LowerOxide_Semiconductor";
34         +LowerOxide {
            interfaceCharge = "Pure";
36             InterfaceCharge {
                Pure {
38
                ext x = 0.0;
40                aux x0 = 0.177; // x-coordinate for x=0
                aux xL = 0.659; // x-coordinate for x=L
42                aux Qinj_0 = -5.333e12; // =^ 2.5V
                aux ls = 0.03;

44                Di = if((x>=x0) && (x<=xL),
```

---

```
46             Qinj_0*exp((-x+x0)/ls)*1.0 "cm^-2", 0.0);
48         }
50     }
52 }

54 Iterate { Scheme : SchemeDefaults.DD; }

56 Curve
57 {
58     file = "ids_vds_mos_05u_subth_out_sourceside_15";
59     Response
60     {
61         +Id = output("Device", "I", "Drain");
62     }
63 }
```

---

## Anhang C

# Numerische Lösung des Schwellspannungsmodells

Zur numerischen Lösung des Schwellspannungsmodells Gl. (4.46) wurde das Programm OCTAVE, ein frei verfügbarer MATLAB-Klon, verwendet. Die folgende Datei dokumentiert den Lösungsalgorithmus:

```
##### Octave Input-File #####
2 #
  # This Octave script is used to calculate the threshold volt. of ONO-transistors
4 #
  #####
6
  #####
8 # Physical constants
  #####
10 global q = 1.602e-19;      # Elementary charge in As
  global e0 = 8.854e-12;     # Permittivity of free space in As/Vm
12
  #####
14 # Device and material parameters
  #####
16
  global Vfb = -0.4155;      # Flat band voltage in V
18 global phi_f = 0.4155;    # Fermi potential in V
  global phi_bi = 0.4155+0.654; # Build in potential in V
20 global toxb = 4.0e-9;     # Bottom oxid thickness
  global toxt = 4.0e-9;     # Top oxid thickness
22 global tnit = 4.0e-9;     # Nitrid thickness

24 global eox = 3.9;         # Relative permittivity of oxid
  global enit = 7.4;         # Relative permittivity of nitrid
26 global esi = 11.9;        # Relative permittivity of silicon
  global Na = 8e22;          # Substrate doping in m^-3
28 global L_x0_s = 0.172e-6;  # Short transistor's length
  global L_x1_s = 0.340e-6;
30 global L_s = L_x1_s - L_x0_s;
  global L_x0 = 0.170e-6;    # Long transistor's length
32 global L_x1 = 0.661e-6;
  global L = L_x1-L_x0;      # Device length in m
34

  #####
36 ## Define operating point and ONO charging
  #####
38
  global Vs = 0;             # Source potential
40 global Vd = 0;             # Drain potential
```

---

```

42 global DVfbs = 2.5;                # ONO charging at source
43 global DVfbd = 2.5;                # ONO charging at drain
44 global la=30e-9;                   # Length of injection zone at source
45 global lb=30e-9;                   # Length of injection zone at drain
46
47 #####
48 # Derived parameters
49 #####
50
51 global tono = toxb+eox/enit*tnit+toxt;          # Effective ONO thickness in m
52 global xd = sqrt(2*e0*esi*2*phi_f/(q*Na));      # Depletion length at phi_0
53 global phi_s0 = 2.0*phi_f;                      # Linearisation point
54
55 global Vth0 = Vfb+q*Na*xd*tono/(e0*eox)+2*phi_f; # Long channel Vth in V
56 global Cox = e0*eox/tono;                      # Cap. density in As/Vm^-2
57 global g = sqrt(2*e0*esi*q*Na)/Cox;            # Gamma in sqrt(V)
58 global a = tono*esi/eox*0.5*sqrt(2*e0*esi*phi_s0/(q*Na));
59 global n = 1+g/(2*sqrt(phi_s0));
60 global lc=sqrt(a/n);                          # Characteristic length
61
62 global op_Vg;
63
64 #####
65 # Function definitions
66 #####
67 function y = psi_3
68     global DVfbs;
69     global la lc;
70
71     y = DVfbs*la^2/(lc^2-la^2);
72 endfunction
73
74 function y = psi_4
75     global DVfbd;
76     global L lb lc;
77
78     y = DVfbd*lb^2/(lc^2-lb^2)*exp(-L/lb);
79 endfunction
80
81 function y = psi_1_o(psi_sl)
82
83     global Vfb phi_bi;
84     global Vs Vd;
85     global L lc
86     global la lb;
87
88     y = (Vd+phi_bi-(Vs+phi_bi)*exp(-L/lc))/(2*sinh(L/lc)) \
89         + psi_3 * (exp(-L/lc)-exp(-L/la))/(2*sinh(L/lc)) \
90         + psi_4 * (exp(-L/lc)-exp(-L/lb))/(2*sinh(L/lc)) \
91         + psi_sl * (exp(-L/lc)-1)/(2*sinh(L/lc));
92 endfunction
93
94 function y = psi_2_o(psi_sl)
95     global Vs Vd phi_bi;
96
97     y= Vs + phi_bi - psi_1_o(psi_sl) - psi_sl - psi_3 - psi_4;
98
99 endfunction
100
101 function y = psi_sl(Vg)
102     global Vs;
103     global q Na e0 esi a n Vfb;

```

```

104         b = q*Na/(e0*esi);
106         y = -b*a/n+(Vg-Vfb)/n;
108     endfunction

110     function y = psi_s(x)
111     ## Definition of equation system for psi_sl and zmin(xmin) calculation
112     ## x(1)=x, x(2)=Vg

114         global lc la lb;

116         y = psi_1_o(psi_sl(x(2)))*exp(x(1)/lc)\
            + psi_2_o(psi_sl(x(2)))*exp(-x(1)/lc)\
118             + psi_sl(x(2)) + psi_3*exp(-x(1)/la) + psi_4*exp(x(1)/lb);
120     endfunction

120     function y = dpsismin_dVg(x)
122     # x(1)=x, x(2)=Vg
123         global n lc L;
124
125         dpsl_sl_dVg=1/n;
126         y = (1 - (sinh(x(1)/lc)+sinh((L-x(1))/lc))/sinh(L/lc)) * dpsl_sl_dVg;
127     endfunction

128     function y = calc_xmin(x)
129         global op_Vg;
130         global lc la lb;
131         y = psi_1_o(psi_sl(op_Vg))*exp(x(1)/lc)\
132             - psi_2_o(psi_sl(op_Vg))*exp(-x(1)/lc)\
133             - psi_3*lc/la*exp(-x(1)/la) + psi_4*lc/lb*exp(x(1)/lb);
134     endfunction

136     function y = xmin(Vg)
137         global op_Vg;
138         global L;
139
140         op_Vg = Vg;
141
142         [y(1,:),inf1]=fsolve("calc_xmin", 0.0*L);
143         if (inf1!=1)
144             y(1)=0;
145         endif
146         [y(2,:),inf2]=fsolve("calc_xmin", 1.0*L);
147         if (inf2!=1)
148             y(2)=L;
149         endif
150     endfunction

152     #####
153     # Algorithm for solution of decoupled equation system
154     #####

156     function y = calc_Vth(Vg,L_loc,op)
157     # Startwerte x(1)=Vth(0)
158     # op(1)=Vs, op(2)=Vd, op(3)=DVfbs, op(4)=DVfbd
159     # Rueckgabe y(1)=psi_sl; y(2)=x_min; y(3)=info --> y(3)=0 --> Konvergiert

160
161         global phi_f;
162         global DVfbs DVfbd;
163         global Vd Vs;
164         global L;
165
166

```





```
230 endfor
    data1=[data1,Vth'];
232 save -text "Vth_Vds_longshort.dat" data1
```

---