

## Entwurf und Optimierung von nanoelektronischen Halbleiterbauelementen unter Nutzung der Prozess- und Bauelementesimulation

von T. Herrmann, S. Flachowsky, R. Illgen, W. Klix, R. Stenzel, J. Höntschel, M. Horstmann und A. Wei

### 1 Einleitung

Der rasante Fortschritt in der Halbleitertechnologie während der letzten Jahrzehnte war von einer stetigen Verkleinerung der Bauelementeabmessungen begleitet, welche zu einer kontinuierlichen Erhöhung der Geschwindigkeit und Funktionalität der elektronischen Schaltkreise führt. Die Erfindung der integrierten Fertigung von Halbleiterbauelementen durch Jack Kilby 1958, der für seine Beiträge zur Erfindung von integrierten Schaltkreisen im Jahre 2000 den Nobelpreis erhielt, war der Ausgangspunkt für diese Entwicklung. Aus der Analyse des Wachstums von 1959 bis 1965 konnte Moore 1965 eine jährliche Verdopplung der Anzahl der Komponenten pro integrierter Schaltung für die künftige Entwicklung ableiten [1]. Dieses empirische Gesetz wurde im Jahre 1975 auf eine Verdopplung aller zwei Jahre angepasst [2] und stellt bis heute einen Leitfaden der wachsenden Halbleiterindustrie dar. Der weitere Bestand dieses Gesetzes erfolgt durch Planungen und Zielstellungen der Halbleiterindustrie selbst, festgelegt in der International Technology Roadmap for Semiconductors (ITRS) [3]. Daraus leiten sich die Notwendigkeit und die Herausforderungen zukünftiger Entwicklungsrichtungen ab. Eine zentrale Technologie für die Herstellung von integrierten Schaltungen ist die CMOS-Technologie (CMOS - complementary metal oxide semiconductor), deren Name sich aus den Grundbauelementen, dem MOSFET (metal oxide semiconductor field effect transistor) herleitet, die in den beiden komplementären Formen p-MOSFET und n-MOSFET, mit p- bzw. n-leitenden Kanalgebieten, zum Einsatz kommen.

Die Verwendung dieser Bauelemente und ihre fortschreitende Miniaturisierung, das heißt die Skalierung ihrer Dimensionen, bestimmen zu einem wesentlichen Anteil die Anforderungen an die Technologie und deren künftige Entwicklung. Die Skalierung der CMOS-Bauelemente ist für eine weitere Erhöhung der Leistungsfähigkeit der integrierten Schaltungen erforderlich aber auch in besonderem Maße für eine Erhöhung der Produktivität und des wirtschaftlichen Ertrages. Die Leistungssteigerung zeigt sich in erster Linie in einer Erhöhung der Arbeitgeschwindigkeiten der Schaltungen und in einer erhöhten Funktionalität durch die Integration einer größeren Anzahl von Bauelementen pro Schaltkreis. Dadurch sinken auch die Kosten je Funktion.

Seit der Einführung der CMOS-Technologie für die Herstellung von integrierten Bauelementen stellt der MOSFET die wichtigste Grundstruktur im Fertigungsprozess dar. Der planare MOSFET (Abb. 1) besteht aus den hochdotierten Bereichen Source und Drain, eingebettet in das

Siliziumsubstrat und voneinander getrennt durch das Kanalgebiet, einem schmalen Bereich entgegengesetzter Dotierung. Flachere Bereiche hoher Dotierung, die Source-/Drain-Erweiterungen (SDE – source/drain extensions), verbinden Source und Drain mit dem Kanal. Über dem Kanal befindet sich eine Oxidschicht, die die Gateelektrode (Steuerelektrode) vom Kanalbereich elektrisch isoliert. Die Source- und Draingebiete sind durch Silizidkontakte mit der Schaltung verbunden. Die Gateelektrode besteht gewöhnlich aus dotiertem Polysilizium, das ebenfalls über ein Silizid kontaktiert ist. Die Gateelektrode ist von Abstandshaltern (Spacern) umgeben (Oxid- und/oder Nitrid-Bereiche), die in der Technologie der selbstjustierenden Masken benutzt werden.

In der Digitaltechnik wird der MOSFET als elektronischer Schalter genutzt. Dabei steuert die Spannung zwischen Gate und Source den Strom am Drain. Durch eine kapazitive Kopplung reichert die Gate-Source-Spannung Ladungsträger an der Silizium-Oxid-Grenzfläche an und erzeugt damit einen niederohmigen Inversionskanal, der Source und Drain elektrisch verbindet.

Die angestrebte Skalierung des MOSFET zur Erhöhung der Packungsdichte kann durch eine Reduzierung der Gatelänge ( $L_G$ ) erfolgen. Dies führt zu einem geringeren Widerstand im Kanalgebiet und durch eine Verringerung der Kapazitäten zu kürzeren Schaltzeiten. Für eine optimale Funktionsweise des Bauelementes müssen weitere Strukturparameter und die Betriebsspannung mit skaliert werden. Bei einer weiteren Skalierung mit Gateoxid-dicken kleiner als 1 nm gelangt man an physikalische

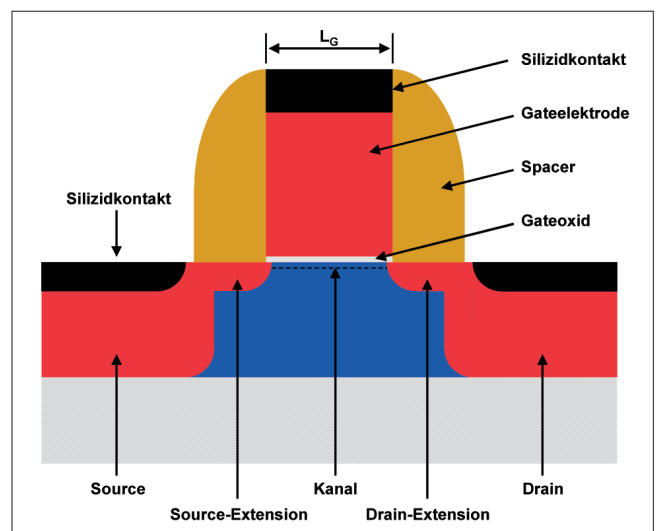


Abb. 1: Schematische Darstellung eines MOSFET

Grenzen, so dass die Entwicklung in Richtung immer kleinerer Gate-Längen bei gleich bleibender Gateoxid-dicke geht. Die Herausforderung für zukünftige Technologien ist deshalb eine weitere Steigerung der Leistungsfähigkeit und Zuverlässigkeit unter Berücksichtigung einer realisierbaren Fertigung.

Alternativ setzt man in der Mikroelektronik zunehmend auf neue Materialien und veränderte Transistorgeometrien, um die Begrenzung durch die Gate-Längenskalierung zu umgehen. Eine Technik zielt dabei auf die Verbesserung der Ladungsträgerbeweglichkeit im Kanal des MOS-Transistors durch Verwendung von verspanntem Silizium. Die mechanische Verspannung wird dabei bewusst durch den Einsatz von Werkstoffen mit abweichenden mechanischen Materialeigenschaften erzeugt.

Eine weitere Möglichkeit, die Gate-Längenskalierung voranzutreiben und dabei leistungsfähigere Transistoren zu erhalten, bietet die Verwendung von FinFET-Strukturen. Hierbei umschließen zwei beziehungsweise drei Gateelektroden eine Siliziumfinne, welche das Kanalgebiet darstellt. Mehrere Gatekontakte führen zu besseren Verarmungs- und Inversionseigenschaften, die wiederum schneller schaltende MOSFETs ermöglichen.

Die fortschreitende Entwicklung der Mikroprozessoren erfordert steigende Investitionen in Forschung und Entwicklung. Die Simulation kann dabei helfen, die Anzahl von Experimenten auf Silizium-Wafern zu verringern und die Entwicklungszeit für neue Halbleitertechnologien zu verkürzen. Dies führt zu einer deutlichen Kostenreduzierung und einer schnelleren Einführung des Produkts. Die steigende Rechenleistung ermöglicht zudem eine schnellere und genauere Simulation immer komplexer werdender Systeme. Des Weiteren können die Auswirkungen von Prozessschwankungen bestimmt und Lösungsansätze zu deren Verringerung untersucht werden. Simulationen bieten darüber hinaus einen Einblick in das innerelektronische Verhalten der Bauelemente und unterstützen das Verständnis der physikalischen Effekte. Die folgenden Abschnitte werden sich ausgehend von konventionellen MOSFETs mit Simulationen zur Verbesserung der Transistoreigenschaften beschäftigen. Dabei stehen die mechanische Verspannung des Siliziums zur Erhöhung der Ladungsträgerbeweglichkeit und neue Transistorstrukturen im Vordergrund.

## 2 Simulation des innerelektronischen Verhaltens der Halbleiterbauelemente

Bei der Simulation von Halbleiterbauelementen können verschiedenste Modelle genutzt werden, die unterschiedlich exakt sind und spezifischen Einschränkungen unterliegen. Dementsprechend ergeben sich zum Teil sehr hohe Rechenzeiten bei der Anwendung. Die für die klassische Halbleitersimulation benötigten Grundgleichungen ergeben sich aus den Maxwell'schen Gleichungen und

aus der Boltzmann-Transportgleichung. Aus Ersteren kann die Poisson-Gleichung hergeleitet werden, welche das elektrostatische Potential im Halbleiter ausgehend von den vorhandenen Ladungen beschreibt. Die Kontinuitätsgleichungen für Elektronen- und Löcherstromdichten ergeben sich ebenfalls aus den Maxwell'schen Gleichungen und beschreiben den Zusammenhang zwischen Stromdichte, der Ladungsträgerverteilung und deren Generation beziehungsweise Rekombination. Die unterschiedlichen Generations- und Rekombinationsvorgänge werden in separaten Gleichungen beschrieben. Die Boltzmann-Transportgleichung stammt aus der kinetischen Gastheorie und beschreibt die statistische Verteilung von Teilchen in einem Medium. Im Halbleiter sind diese Teilchen die Ladungsträger, welche ein sogenanntes Ladungsträgergas bilden. Mit Hilfe der Momenten-Methode kann die Boltzmann-Transportgleichung vereinfacht werden, wobei die Transport- und Energieerhaltungsgleichungen entstehen. Die Transportgleichungen vervollständigen das Drift-Diffusions-Modell. Die Stromdichten werden durch die Gradienten des elektrostatischen Potentials und der Ladungsträgerdichten beschrieben. Als zusätzliche Parameter gehen die Ladungsträgerbeweglichkeiten ein, die mit weiteren Modellen beschrieben werden. Das Drift-Diffusions-Modell ist das am häufigsten genutzte Gleichungssystem zur Beschreibung des inner-elektronischen Verhaltens im Halbleiter. Die Energieerhaltungsgleichungen erweitern das Drift-Diffusions-Modell zum hydrodynamischen Modell und beziehen die Ladungsträgerenergie mit in das Gleichungssystem ein.

Das entstehende partielle Differentialgleichungssystem wird mit Hilfe von Finite-Differenzen- oder Finite-Elemente-Methoden gelöst. Als Ergebnis liegen die örtlich zwei- oder dreidimensionale Verteilung von Potential- und Ladungsträgerdichten vor, aus denen die Klemmenströme und deren Zeitverlauf berechnet werden können. Die Struktur des zu berechnenden Bauelementes kann durch eine vorhergehende Simulation des Herstellungsprozesses erzeugt oder direkt vorgegeben werden.

Die fortschreitende Miniaturisierung bis in den Nanometerbereich lässt die klassischen Simulationsmodelle an ihre Grenzen stoßen. Im Inversionskanal eines MOSFET können in Folge der starken Inversion Ladungsträgerquantisierungseffekte auftreten, die nicht mehr mit den klassischen Modellen nachgebildet werden können. Hier schaffen Modellerweiterungen Abhilfe, aus denen das Quanten-Drift-Diffusions-Modell und das Quanten-Hydrodynamische Modell entstehen.

Die folgenden Simulationen wurden mit dem Bauelementesimulator SIMBA [4] und dem TCAD- (Technology Computer Aided Design) Programmsystem der Firma SYNOPSIS [5] durchgeführt.

## 3 Konventioneller MOSFET

Unter konventionellen MOSFETs sind die derzeit in Massenproduktion hergestellten planaren MOSFETs mit einem Gatekontakt zusammengefasst (Abb. 1). Wie schon in der Einleitung beschrieben, entsprechen die MOSFETs elektronischen Schaltern. Die Höhe der Drain-Source-Spannung unterscheidet den linearen und den Sättigungsbereich, welche unterschiedliche Eigenschaften haben. Mit einer Spannung von 0 V zwischen Gate und Source ist der MOSFET ausgeschaltet und es fließt nur ein geringer Sperrstrom ( $I_{D,off}$ ) am Drain, wie in Abb. 2 zu sehen. Steigt die Spannung zwischen Gate und Source, steigt auch der Drainstrom. Ist die Betriebsspannung erreicht, fließt am Drain der Sättigungsstrom ( $I_{D,sat}$ ) beziehungsweise der lineare Strom ( $I_{D,lin}$ ). Dieser Strom entlädt die Gatekapazität des darauf folgenden Transistors und schaltet diesen ebenfalls ein. Wie schnell dieser Vorgang geschieht, hängt von der Größe des Sättigungsstroms und der Größe der Gatekapazität ab. Ziel ist ein hoher Sättigungsstrom bei kleiner Gatekapazität und kleinen Leckströmen, um einen möglichst schnellen Schaltvorgang zu erreichen. Neben der Gatekapazität existieren weitere parasitäre Kapazitäten, die für genauere Betrachtungen mit in die dynamischen Simulationen einbezogen werden müssen. Aus der reinen Bauelementesimulation wird eine kombinierte Bauelemente- und Netzwerksimulation mit zusätzlichen diskreten Bauelementen. Hier können auch mehrere Transistoren gekoppelt und damit ganze Inverterketten simuliert werden.

In Folge von Variationen bei den Lithographie- und Ätzprozessen zur Gatestrukturierung schwanken die Gatelänge und somit auch die elektrischen Parameter der Transistoren. Ein wichtiger Parameter, dessen Änderung nur in bestimmten Grenzen zugelassen wird, ist die Schwellspannung. Sie gibt an, bei welcher Gate-Source-Spannung der Transistor zu leiten beginnt. Sind die Schwellspannungen der Einzeltransistoren zu unterschiedlich, kann die Funktion der Inverterkette nicht mehr gewährleistet werden. Daher ist die Schwankung der

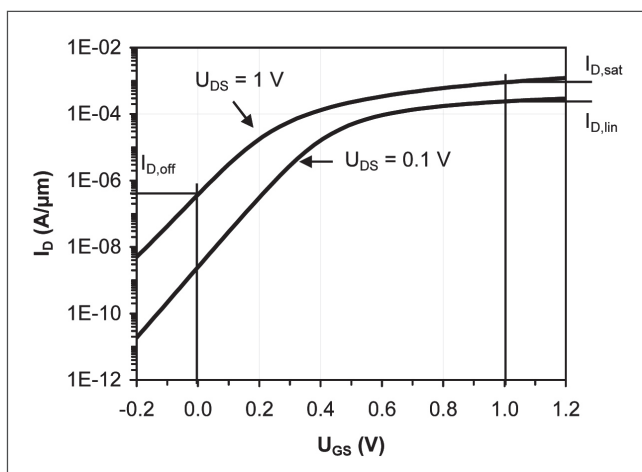


Abb. 2: Transferkennlinie eines konventionellen MOSFETs

Schwellspannung bei variiertem Gatekontakt ein weiterer Indikator für die Verbesserung der Transistoreigenschaften, wobei diese gering sein sollte. Wird der Sperrstrom über dem Sättigungsstrom für verschiedene Gatekontakte dargestellt, ergibt sich die sogenannte Universalkurve. Hier kann der Sättigungsstrom bei konstantem Sperrstrom direkt abgelesen werden.

## 4 Verspanntes Silizium

Das Konzept des verspannten Siliziums beruht auf einer Änderung der Ruhelage der Atome im Kristallgitter des Siliziums durch eine von außen einwirkende Kraft. Diese mechanische Verspannung führt zu einer Verformung des Kristalls und es kommt zu einer Modifizierung der elektronischen Bandstruktur, wodurch die elektrischen Eigenschaften des Siliziums erheblich verändert werden. Abhängig von der Stärke und Richtung in die der Kristall verspannt wird, lassen sich die Ladungsträgerbeweglichkeiten stark verbessern. Bereits eine Verformung des Siliziumkristalls um 1 % führt zu einer Verdopplung der Elektronenbeweglichkeit. Dieses Beispiel verdeutlicht das Potential, das im verspannten Silizium steckt.

Obwohl der Effekt der Verspannung auf die Ladungsträgerbeweglichkeit in Silizium schon seit Jahrzehnten bekannt ist [6], erlauben erst jüngste Fortschritte in der Kristallwachstumstechnologie, verspannt abgeschiedene Schichten gezielt für die Bauelementverbesserung auszunutzen. Prinzipiell profitiert der n-MOSFET von einer Zugverspannung im Kanal und der p-MOSFET von einer Druckverspannung, so dass in der CMOS-Technologie die Transistoren komplementär verspannt werden müssen. Es haben sich verschiedene Methoden entwickelt, um die Verspannung in den Transistorkanal einzuprägen. Dabei dienen zum einen epitaktisch abgeschiedene Siliziumverbindungen (SiGe, SiC) als Verspannungsquelle oder intrinsisch verspannte Siliziumnitridfilme, die auf den Transistor abgeschieden werden und das darunterliegende Silizium verspannen.

### 4.1 SILIZIUM-GERMANIUM

Die Element-IV-Halbleiter Silizium und Germanium sind in allen Mischungsverhältnissen vollständig miteinander mischbar, wobei die Gitterkonstantenabweichung der beiden Elemente  $\sim 4,2\%$  beträgt. Dies lässt sich zur Erzeugung mechanischer Verspannung ausnutzen, um die Ladungsträgerbeweglichkeiten im Kanal des Transistors zu erhöhen. Die Integration der Silizium-Germanium-Legierung (SiGe) in den Transistor ist in Abb. 3 dargestellt. Nach der Gate-Strukturierung werden mit Hilfe eines Ätzschritts Mulden in den zukünftigen Source/Drain-Gebieten des Transistors erzeugt. Auf dem restlichen Silizium im Boden der Mulden wird nun durch eine selektive Epitaxie SiGe abgeschieden. Das neue SiGe-Material nimmt dabei die kleinere Gitterkonstante des



Siliziumsubstrates an und erfährt eine intrinsische Druckverspannung in der Wachstumsebene und senkrecht dazu eine Zugverspannung (Abb. 4).

Die verspannten SiGe Source/Drain-Gebiete üben nun eine Kraft auf die umliegenden Gebiete aus und induzieren eine Druckverspannung im Kanalgebiet des Transistors. Die entstehenden Verspannungswerte liegen im GPa-Bereich (Abb. 5). Da diese Verspannungskonstellation die Löcherbeweglichkeit verbessert [7], wird das eingebettete SiGe beim p-MOSFET angewendet.

Der Einfluss der Verspannung auf die Ladungsträgerbeweglichkeit spiegelt sich in den Sättigungsströmen des Transistors wider. Abb. 6 zeigt die Universalcurven für p-MOSFETs ohne SiGe und für solche mit 20 % Ge bzw. 30 % Ge im SiGe. Die Verschiebung der Universalcurve der Transistoren mit SiGe im Vergleich zu denen ohne SiGe zu höheren Sättigungsströmen bei gleich bleibenden Sperrströmen zeigt eine Verbesserung des Transistorleistungsfähigkeit um ~30 %.

Man beobachtet für den p-MOSFET mit höherem Ge-Gehalt eine stärkere Verbesserung durch die größere Gitterkonstantenabweichung des SiGe mit mehr Ge-Gehalt. Dies wird durch Prozess- und Bauelementesimulationen bestätigt, welche einen linearen Zusammenhang zwischen dem Ge-Anteil im SiGe und der erzeugten Verspannung und entsprechender  $I_{D,sat}$ -Verbesserung zeigen (Abb. 7). Der weiteren Erhöhung des Ge-Gehaltes sind in der Praxis Grenzen gesetzt. In den stärker verspannten SiGe-Schichten kommt es durch Ionenimplantationen und Ausheilprozesse eher zu einer Relaxation der Verspannung und es bilden sich Kristalldefekte, die die elektrischen Eigenschaften des Bauelementes erheblich verschlechtern.

Neben dem Germaniumgehalt im SiGe beeinflusst der Abstand des SiGe zum Kanal die Leistungsfähigkeit des Transistors erheblich. Aus Abb. 5 ist erkennbar, dass die stärksten Verspannungen im Silizium am Rand zu den SiGe-Gebieten auftreten und zur Mitte des Kanals hin abnehmen. Bringt man das SiGe näher an den Kanal, rücken auch die stark verspannten Gebiete mehr ins

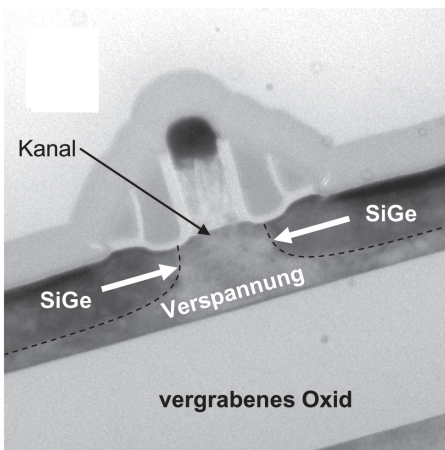


Abb. 3: TEM-Querschnittsbild eines SOI-Transistors (p-MOSFET) mit eingebettetem SiGe [7]

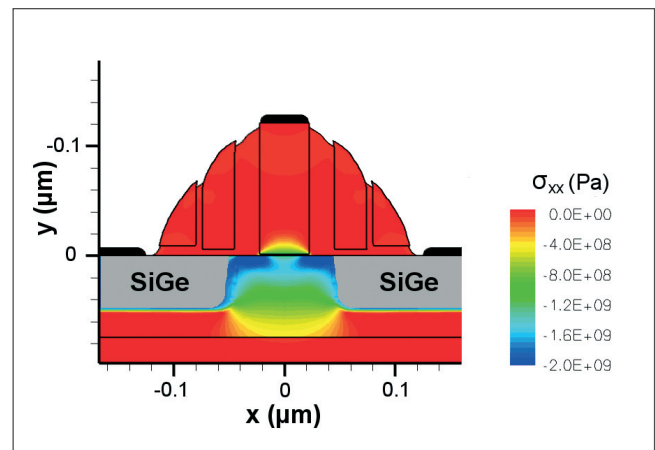


Abb. 5: Verspannungsprofil  $\sigma_{xx}$  in einem p-MOSFET durch eingebettetes SiGe (im Kanalgebiet herrscht Druckverspannung, negative Werte)

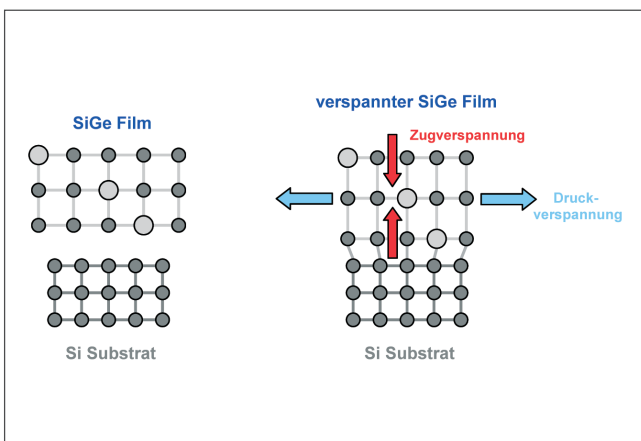


Abb. 4: Verformung eines epitaktischen SiGe-Filmes auf dem Si-Substrat und die resultierenden Kräfte

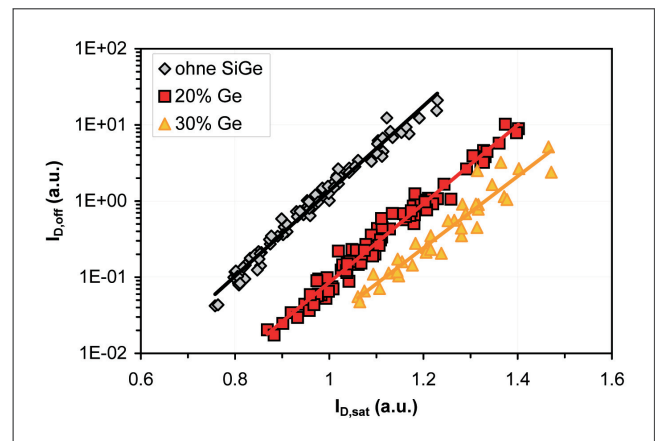


Abb. 6: Universalcurven von p-MOSFETs ohne SiGe und mit SiGe (20 % bzw. 30 % Ge)

Zentrum des Kanals, wo die erhöhte Verspannung einen größeren Einfluss auf den Ladungstransport ausübt. Dies ist in Abb. 8 dargestellt. Für kleiner werdende Abstände steigt die Verspannung in horizontaler sowie in vertikaler Richtung an. Gleichzeitig nimmt der Sättigungsstrom exponentiell zu.

## 4.2 VERSPANNTE NITRIDFILME

Verspannte Nitridschichten stellen eine weitere Möglichkeit für die gezielte Verspannung des Transistorkanals dar. Sie werden nach der Silizierung (Kontaktbildung) auf dem Bauelement abgeschieden und üben eine Kraft auf den darunterliegenden Transistor aus [8]. Abhängig von den Prozessbedingungen bei der Abscheidung können Nitridfilme mit intrinsischer Zug- oder Druckverspannung erzeugt werden. Dabei führen druckverspannte Schichten zu einer verbesserten Löcherbeweglichkeit im p-MOSFET und zugverspannte Schichten zu einer Steigerung der Elektronenbeweglichkeit im n-MOSFET [9]. Die resultierende Verspannungsverteilung durch einen 80 nm dicken

Nitridfilm mit 1.2 GPa intrinsischer Zugverspannung stellt Abb. 9 dar.

Mit Hilfe der Simulation kann der Einfluss einer veränderten Filmdicke untersucht werden (Abb. 10). Dabei ist eine Sättigung im Drainstromgewinn für Dicken größer als 140 nm erkennbar. Eine kontinuierliche Erhöhung der intrinsischen Verspannung des Nitridfilmes zeigt einen linearen Zusammenhang mit der Verbesserung des Sättigungsstromes (Abb. 11). Prozessbedingte Anforderungen setzen aber einer kontinuierlichen Erhöhung dieser beiden Parameter Grenzen. Weiterhin ist die durch den Nitridfilm entstehende Verspannung im Kanal stark von der Geometrie des Transistors, wie Polygatehöhe, Spacerweiten und der Entfernung zu benachbarten Transistoren abhängig. Dieses Zusammenspiel mehrerer Faktoren auf das Verhalten des Transistors erfordert eine Optimierung vieler Prozesse, wobei die Simulation sinnvoll als Werkzeug eingesetzt wird.

Für den p-MOSFET verwendet man ebenfalls einen verspannten Nitridfilm, allerdings mit einer intrinsischen

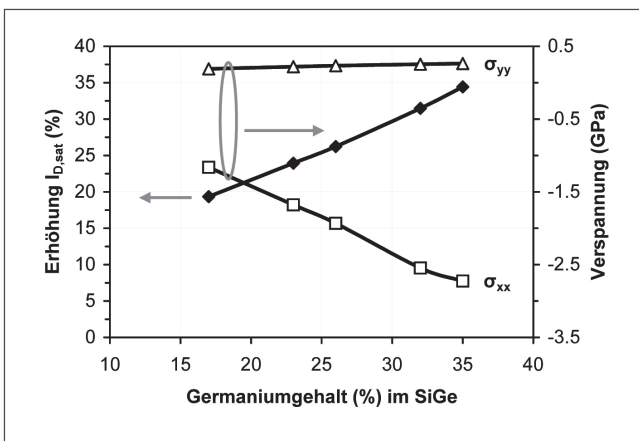


Abb. 7: Einfluss des Ge-Gehaltes im SiGe auf den Sättigungsstrom im Vergleich zu einem p-MOSFET ohne SiGe

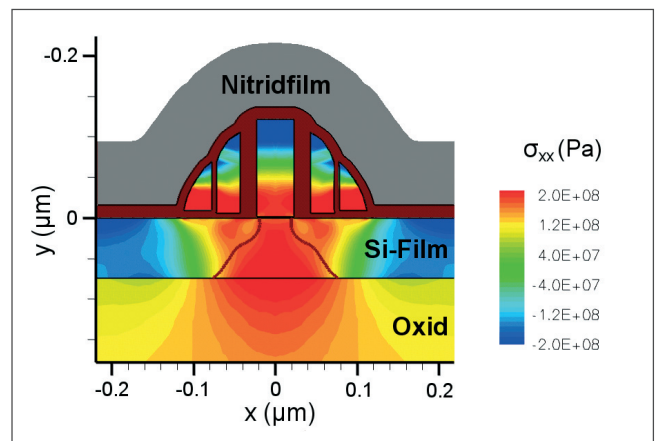


Abb. 9: Verspannung  $\sigma_{xx}$  in einem n-MOSFET durch einen zugverspannten Nitridfilm

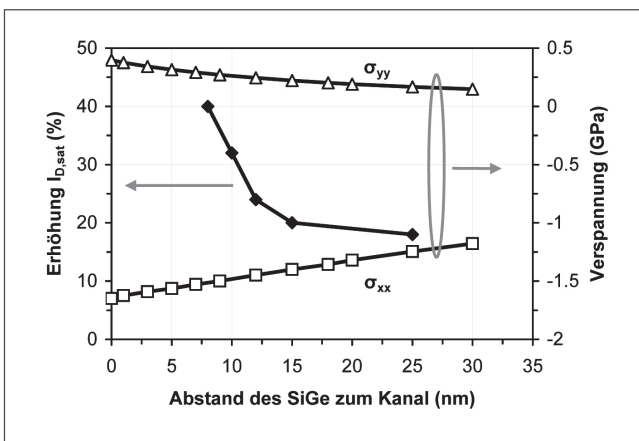


Abb. 8: Einfluss des Abstandes zwischen SiGe und Kanal auf die Verspannung und den Sättigungsstrom

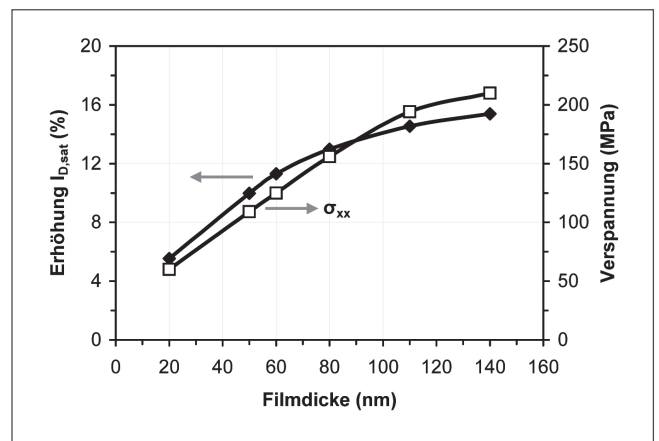


Abb. 10: Einfluss der Filmdicke auf die Verspannung im Kanal in x-Richtung und auf den Sättigungsstrom

Druckverspannung. Dies erlaubt eine Kombination mit dem SiGe, wodurch sich die Sättigungsströme additiv verbessern [8].

### 5 FinFETs

Die Skalierung der Gatelänge ohne eine weitere Skalierung der Gateoxiddicke führt dazu, dass das Gate immer weniger Steuerwirkung auf den Kanal hat und der Transistor nicht mehr ausgeschaltet werden kann. Die minimale Gateoxiddicke wird dabei durch die stärker werdenden Gateleckströme bestimmt und kann nicht entsprechend der Skalierungsvorschriften abgedünnt werden. Ein zweites Gate gegenüber dem Ersten kann die Gatesteuerwirkung vergrößern, wenn der dazwischen liegende Siliziumsteg entsprechend dünn ist. Des Weiteren können durch den zweiten Inversionskanal höhere Ströme erreicht werden, die aber auch mit einer größeren Gatekapazität einhergehen. Das zweite Gate kann auf verschiedenen Wegen realisiert werden. Hier soll nur die Variante des FinFETs vorgestellt werden.

FinFETs gelten als Favoriten unter den neuartigen MOSFET-Strukturen, die laut ITRS [3] im Jahre 2013 die konventionellen Transistoren ablösen sollen. Die ITRS spricht allgemein von Double-Gate-Transistoren, welche die FinFETs mit einschließen. Der Herstellungsprozess der FinFETs ähnelt dem herkömmlicher MOSFETs. Somit wird die Prozessintegration erleichtert [10], [11].

FinFETs bestehen aus einer Siliziumfinne, die das Kanalgebiet darstellt, und einem Gate, welches die Finne umschließt. An ihren beiden Enden geht die Finne in das Source-beziehungsweise Draingebiet über, wie in Abb. 12 dargestellt. Das Gate kann an zwei Seiten oder an drei Seiten anliegen, wobei es sich dann um ein Double-Gate-beziehungsweise Triple-Gate-FinFET handelt. Die Finnenhöhe ( $h_{Fin}$ ) entspricht der Gateweite von konventionellen MOSFETs und wird technologisch durch das Verhältnis zur Finnenweite ( $w_{Fin}$ ) begrenzt. Über die Gateweite kann

bei den konventionellen Transistoren die Höhe des Gesamtstroms bestimmt werden. Bei FinFETs wird dies durch die Aneinanderreihung der einzelnen Finnen zu einer so genannten Kammstruktur erreicht. Der Inversionskanal bildet sich an beiden Seiten über die gesamte Höhe der Finne aus. Am Triple-Gate-FinFET befindet sich zusätzlich an der Oberseite der Finne ein Inversionskanal, der sich über die Finnenweite ausbildet. Die Source- und Draingebiete besitzen eine größere Weite im Vergleich zur Finne um deren Widerstand möglichst gering zu halten. Da beim FinFET in allen drei Koordinatenrichtungen funktionsrelevante Effekte auftreten, muss er im Gegensatz zum konventionellen Transistor in allen drei Dimensionen simuliert werden. Dies erhöht die Anzahl der Punkte des Diskretisierungsgitters und damit den Rechenaufwand.

Die Ausgangsstruktur des hier dargestellten FinFETs besteht aus einer 10 nm breiten und 50 nm hohen Finne, die von einem 1 nm dicken Oxid bedeckt wird. Das Verhältnis von Finnenweite zu Höhe wurde so gewählt, dass die Finne während der Herstellung nicht umkippt. Unter der Finne befindet sich ein 100 nm dickes vergrabenes Oxid mit darunter liegendem Substrat. Das Gate ist vollständig metallisiert. Die Source- und Draingebiete bestehen aus einem Teil der Siliziumfinne sowie weiteren 50 nm epitaktisch aufgewachsenem Silizium. Zwischen dem Gate und den Source- und Draingebieten befinden sich Oxid-Spacer mit einer Länge von 10 nm, welche die Gebiete kapazitiv koppeln. In Abb. 13 ist die Dotierkonzentration des FinFETs bei einem Schnitt in der x-y-Ebene bei  $z = 0$  nm dargestellt. Abb. 14 zeigt einen Schnitt in der y-z-Ebene bei  $x = 20$  nm für dasselbe Bauelement. Da der FinFET symmetrisch ist, reicht es aus, die dargestellte Hälfte zu simulieren und anschließend die Ströme zu verdoppeln. Neben der Gatelänge von 20 nm wurden zwei weitere Gatelängen von 26 nm beziehungsweise 16 nm simuliert, so dass die Universalkurve gebildet werden kann.

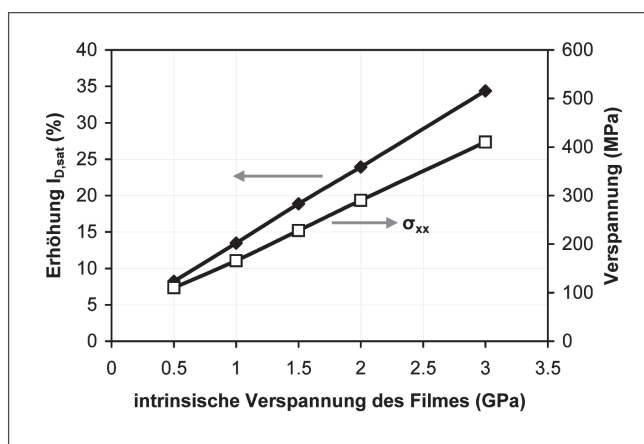


Abb. 11: Einfluss der intrinsischen Verspannung des Nitridfilmes auf die Verspannung im Kanal in x-Richtung und auf den Sättigungsstromgewinn

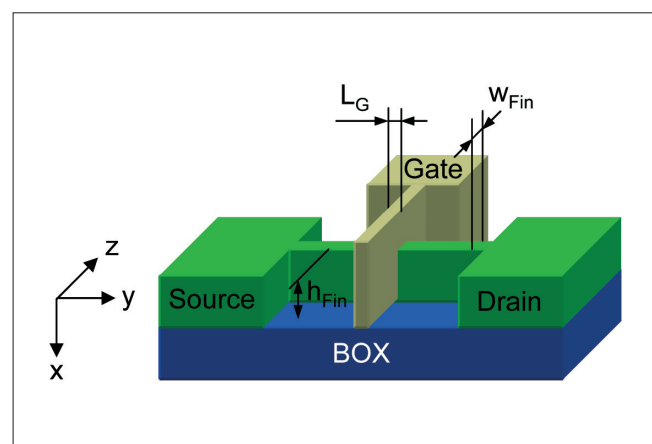


Abb. 12: Struktur des FinFETs

In Abb. 15 ist die Elektronendichte in der Finne dargestellt. Auf Grund der hohen Inversionsladungsträgerdichte und der geringen Finnenabmessungen kommt es am Gateoxid zu einer Quantisierung der Ladungsträgerenergiezustände in dessen Folge das Maximum der Ladungsträgerdichte von der Grenzfläche in die Finne verschoben wird. Die hohe Ladungsträgerdichte in der oberen Ecke der Finne wird durch die Überlagerung der elektrischen Felder vom oberen und seitlichen Gate verursacht.

Eine höhere Finne ergibt einen größeren Gesamtstrom des FinFETs, jedoch verringert sich der auf die Weite des Gates normierte Strom. Die Gateweite setzt sich beim Double-Gate-FinFET aus der zweifachen Finnenhöhe zusammen und beim Triple-Gate-FinFET kommt noch die Finnenweite hinzu. Werden die normierten Ströme in der Universalkurve dargestellt, zeigt sich eine Verringerung des Sättigungsstromes für eine größere Finne. Der Grund hierfür ist der abnehmende Einfluss der hohen Ladungsträgerdichte in der oberen Ecke der Finne auf den Gesamtstrom. Dieser wird anteilig immer geringer

und damit sinkt der normierte Sättigungsstrom des höheren FinFETs. Eine Veränderung der Schwellspannung konnte nicht festgestellt werden.

Anders verhält es sich für die Veränderung der Finnenweite. Diese wurde von 10 nm auf 14 nm und 18 nm vergrößert. Dabei lässt die Steuerwirkung des Gates auf den Kanal nach, da die gegenüberliegenden Gates voneinander entfernt werden und die Schwellspannung sinkt, wie in Abb. 16 dargestellt. Die Variation der Gatelänge führt darüber hinaus zu größeren Schwankungen der Schwellspannung. Wird die Schwellspannung des Transistors mit nomineller Gatelänge ( $L_G = 20$  nm) über die Kanaldotierung angepasst (mit \* gekennzeichnet), bleibt die stärkere Schwellspannungsschwankung erhalten. In der Universalkurve (Abb. 17) steigen sowohl die Sperr- als auch die Sättigungsströme bei einer größeren Finnenweite, jedoch verringern sich die Sättigungsströme bei angepasster Schwellspannung.

FinFETs zeigen eine gute Skalierbarkeit, sind aber anfällig gegenüber Schwankungen im Herstellungsprozess. Das Verhältnis aus Finnenweite und -höhe bedarf einer Optimierung, wobei die Prozessschwankungen berücksichtigt werden müssen.

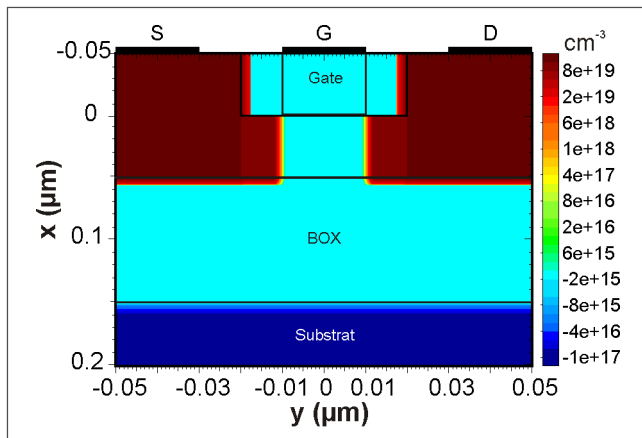


Abb. 13: Dotierungskonzentration als Schnitt in der x-y-Ebene

## 6 Zusammenfassung

Die dargestellten Beispiele für die Weiterentwicklung der CMOS-Technologie ermöglichen leistungsfähigere Schaltkreise jedoch bei erhöhtem Herstellungsaufwand. Die Simulation kann die bereits genutzte Technologie weiter optimieren beziehungsweise neue Verfahren und Strukturen untersuchen. Dazu müssen die genutzten physikalischen Modelle ständig angepasst und weiter entwickelt werden.

Eine Methode zur Leistungssteigerung ist die hier vorgestellte mechanische Verspannung des Siliziumkanals,

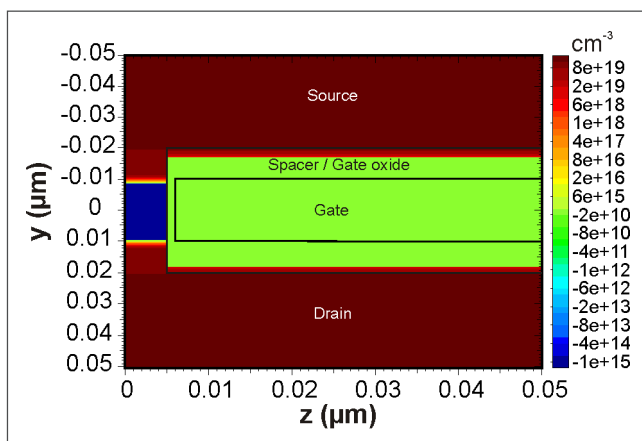


Abb. 14: Dotierungskonzentration als Schnitt in der y-z-Ebene (halbe Struktur)

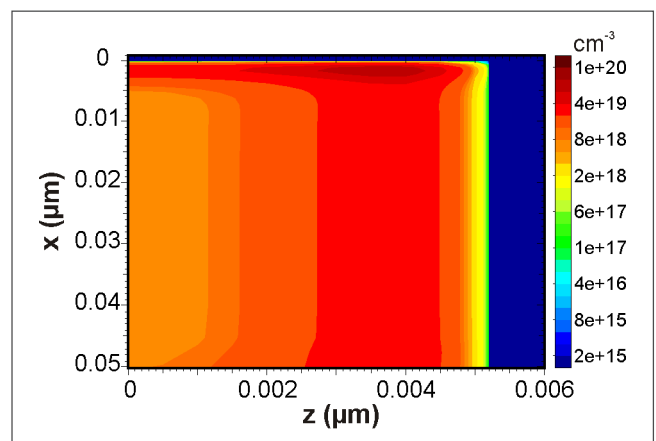


Abb. 15: Ladungsträgerdichte in der Finne bei eingeschaltetem Transistor



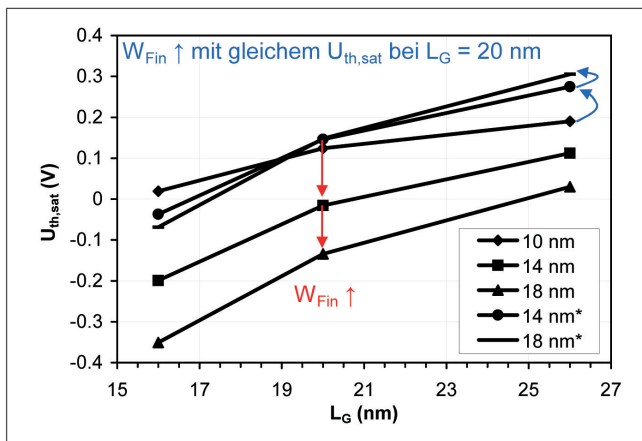


Abb. 16: Verlauf der Schwellspannung für verschiedene Finnenweiten

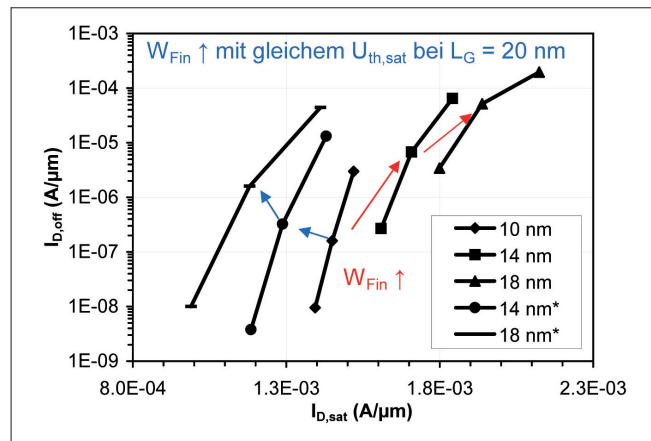


Abb. 17: Universalkurven der verschiedenen Finnenweiten

die eine erhöhte Ladungsträgerbeweglichkeit hervorruft. Im Gegensatz dazu nutzen FinFETs die Vorteile der verbesserten Schalteigenschaften durch eine neue Gateanordnung.

Das dieser Publikation zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 01M3156 und 01BM3167B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren, die experimentellen Daten wurden von AMD FAB 36 zur Verfügung gestellt.

### Literatur und Quellen

[1] G.E. Moore, „Cramming more components onto integrated circuits“, Electronics, Vol. 38, No. 8, 1965, S. 114-117

[2] G.E. Moore, „Progress in digital integrated circuit“, IEDM Techn. Dig., 1975, S. 11

[3] „Process Integration, Devices, and Structures“, ITRS 2006, www.itrs.net

[4] R. Stenzel, W. Klix, „SIMBA – ein universeller 3D-Bauelementesimulator“, Berichte und Informationen, HTW-Dresden, 2/1999, 1/2000, S. 51-56

[5] www.synopsys.com

[6] C.S. Smith, „Piezoresistance Effect in Germanium and Silicon“, Phys. Rev., Vol. 94, No. 1, 1954, S. 42-49

[7] M. Horstmann, A. Wei, T. Kammler et al. „Integration and Optimization of Embedded-SiGe, Compressive and Tensile Stressed Liner Films, and Stress Memorization in Advanced SOI CMOS Technologies“, IEDM Techn. Dig., 2005, S. 233 - 236

[8] A. Wei, T. Kammler, J. Höntschel et al., „Combining Embedded and Overlayer Compressive Stressors in Advanced SOI CMOS Technologies“, Int. Conf. On Solid State Devices and Materials, 2005, S. 32-33

[9] M. Wiatr, Th. Feudel, A. Wei et al., „Review on Process-induced Strain Techniques for Advanced Logic Technologies“, Int. Conf. On Advanced Thermal Processing of Semiconductors, 2007, S. 19-29

[10] E.J. Nowak, I. Aller, T. Ludwig et al., „Turning Silicon on its Edge“, IEEE Circuits & Devices Magazine, Jan/Feb 2004, S. 20-30

[11] B. Yu, L. Chang, S. Ahmed et al., „FinFET Scaling to 10nm Gate Length“, IEDM Techn. Dig., 2002, S. 251 - 254

Verfasser:

Dipl.-Ing. (FH) Tom Herrmann  
 Dipl.-Ing. (FH) Stefan Flachowsky  
 Dipl.-Ing. (FH) Ralf Illgen  
 Prof. Dr.-Ing. habil. Wilfried Klix  
 Prof. Dr.-Ing. habil. Roland Stenzel  
 Arbeitsgruppe Nanoelektronik, FB Elektrotechnik  
 Hochschule für Technik und Wirtschaft Dresden (FH)

Dr.-Ing. Jan Höntschel  
 Dr.-Ing. Manfred Horstmann  
 Ph.D. Andy Wei  
 AMD FAB 36 LLC & Co. KG, Germany