Bauelemententwicklung

am Beispiel SiGe

Christoph Bromberger



A. M. D. G.



ulm university universität
uulm

Bauelemententwicklung am Beispiel SiGe

DISSERTATION

zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

(Dr.-Ing.)

der Fakultät für Ingenieurwissenschaften
und Informatik der Universität Ulm

von

Dipl.-Phys. Christoph Bromberger

aus Aalen

Gutachter:	Prof. Dr.-Ing. Hermann Schumacher
	Prof. Dr. Bernd Tillack
Amtierender Dekan:	Prof. Dr.-Ing. Klaus Dietmayer

Ulm, 04.07.2011

Vorwort

So gewahrten die alten portugiesischen Seefahrer,
die das große Endmeer fürchteten und dennoch weiterfuhren,
endlich nicht Ungeheuer und Abgründe,
sondern Wunderstrände und nie geschaute Gestirne.

Fernando Pessoa, AUS DEN GEDICHTEN DES ÀLVARO DE CAMPOS [0.1]

Die vorliegende Arbeit skizziert ein neuartiges Hochfrequenzbauelement, mitsamt einem Fertigungsablauf, der sich als optionales Modul in eine HBT[1]-Technologie einfügt. Das vorgeschlagene Bauelement, eine vertikal integrierte Kaskodenstufe, ergänzt die vorhandenen HBTs hin zu höheren Frequenzen und Leistungen.

Je nach Neigung der Autoren nähern sich Darstellungen der Technologieintegration von der Warte der Physik, aus der Sicht der Schaltungstechnik oder von der Fertigungstechnologie her. Hervorragende Arbeiten stehen aus allen drei Fachdisziplinen zur Verfügung. Während sich die Teildisziplinen im allgemeinen merklich auseinanderentwickelt haben, erwächst vor allem der Hochfrequenzintegration ein großer Mehrwert aus der Zusammenschau der unterschiedlichen Blickwinkel.

Die Inspiration zu den hier vorgestellten Arbeiten liegt in den Anforderungen der Schaltungsentwicklung. Das Bauelementkonzept emöglicht einen Einblick in die zugrundeliegende Physik. Der Entwurf eines Herstellungsverfahrens schließt den Kreis zu Experiment und Fertigungsablauf. Der Gegenstand der Arbeit bietet sich daher an, en passant aufzuzeigen, wie sich die Teilaspekte gegenseitig bedingen und befruchten.

Teil I der Arbeit bietet die obligate knappe Übersicht über den Bipolartransistor.

In Deutschland kommen auf jeden Maurerlehrling drei Architekturstudenten[2]. Um die Gewichte anders zu verteilen, betone ich in Teil II, bei der Einführung in die Technologiearchitektur, die handwerklichen Aspekte: Ist die Spannungsfestigkeit eines Transistors vorgegeben, sind es gerade die „trivialen“ handwerklichen Aspekte der Prozeßintegration, die die Hochfrequenzeigenschaften begrenzen. Als Beispiel dient die Notwendigkeit, Kollektor, Basis und Emitter elektrisch zu kontaktieren. Hierbei wird das Experiment, namentlich die Fehleranalyse, eine wichtige Rolle spielen.

In Teil III folgt ein Überblick über die physikalische Simulation: Die Simulation von Prozessen und Bauelementen ist ein wichtiger Wegweiser in der Entwicklungsphase einer Technologiearchitektur.

Teil IV befaßt sich mit der Schaltungstechnik. Hier werden Verfahren und Lösungen vorgestellt, mit denen die Schaltungstechnik die Begrenzungen durch die Bauelementphysik ein Stück weit hinausschieben kann. Aus den Anforderungen aus der Schaltungstechnik und den Randbedingungen der Technologiearchitektur erwächst die Idee zur vertikalen Integration einer Kaskodenschaltung.

---

[1] Alle Abkürzungen werden im Schlagwortverzeichnis erläutert

[2] 37416 Architekturstudenten zum 31. 12. 2005 (www.bak.de),
11167 Auszubildende im Maurerhandwerk 2005 (www.zdh.de)

Der nachfolgende, zentrale Teil V skizziert einen Herstellungsablauf für die vertikal integrierte Kaskode. Der Herstellungsablauf fügt sich als optionales Modul in die SiGe2-Technologiearchitektur von Atmel ein.

Der abschließende Teil VI weist anhand eines Kompaktmodells, dessen Parameter aus der physikalischen Simulation extrahiert werden, den schaltungstechnischen Nutzen des vorgeschlagenen Bauelements nach. Es schließen sich die Zusammenfassung (Teil VII) und ein kurzer Anhang (Teil VIII) an, in welchem ich (aus Gründen der Vollständigkeit) kurz die Prozeßsimulation anreiße.

Aus meinen beiden Jahren bei Herrn Prof. Schumacher sind über 20 Patenteinreichungen hervorgegangen. An passenden Stellen nehme ich durch ein Ⓟ, auf die laufenden Nummer im Teil X, „Patenteinreichungen“, bezug.

Dem Rezensenten einer Dissertation wird man nicht zumuten, den Honig wissenschaftlicher Originalität zwischen weiten Brachen des Standes der Technik zusammenzutragen. Auf der anderen Seite ist eine Arbeit, die zum „Wissen der Menschheit“ beitragen will, auf eine gewisse Verständlichkeit angewiesen. Ich habe versucht, zwischen dem Telegrammstil einer wissenschaftlichen Veröffentlichung und der epischen Breite eines Lehrbuches zu vermitteln.

Eine Berechtigung erfährt dieses Vorgehen auch aus der Notwendigkeit, ein „Gefühl“ für den dargestellten Gegenstand zu vermitteln. Dieses Gefühl war erforderlich zur Fortschreibung des status quo ante, seine Vermittlung möge als notwendiger Teil der Arbeit erachtet werden.

Die entstandene Schrift ist, wie Äsops Fledermaus zwischen Vögeln und Landtieren, zwischen Experiment, Bauelementphysik und Schaltungstechnik angesiedelt. Der Kürze der Darstellung geschuldet, wie auch der synoptischen Zielrichtung, läuft die Arbeit in Teilen sicher Gefahr, in ein Konvolut aus Trouvaillen abzugleiten. Auch wird manches, was beim Überfliegen zunächst als Unsinn erscheint, bei näherer Betrachtung Unsinn bleiben.

Anderes aber mag, wenngleich nicht neu, so doch aus einem neuen Blickwinkel beleuchtet sein, und, wenn nicht in sich selbst, so doch in der Arbeit des Lesers Früchte tragen. Und so...

> *... kann ich mich nicht enthalten, hier am Anfang Tölpel und Toren zu bitten, wenn sie irgendwohin kommen, wo sie diese Verse und diese Worte hören ... daß sie sie ungeschoren lassen.*
>
> Willem, VAN DEN VOS RYNAERDE – REINART FUCHS [0.2]

Abstract

Device Design using SiGe as an Example

The work at hand suggests a new high frequency device. The parent technology of the proposed device is SiGe of Atmel, Heilbronn.

The original line of attack was an experimental 'hit and run'. Inherent shortcomings within the parent technology oppose this heuristic approach: the high-frequency capability is no longer limited by the transistor itself, but by the necessity to establish electrical contact to the transistor's regions.

Improvements are proposed in the form of a modified manufacturing architecture. To verify this architecture, extensive experimental pilots on solid-phase epitaxial overgrowth of silicides by silicon are still lacking. The present work does not produce a new off-the-shelf device, but rather provides a well-founded investment proposal, closing the circle back towards experiment.

From experiment through theory and back towards experiment, I followed my whims as they occured, embracing five years of semiconductor industry experience – a very satisfying round-up of a chapter in my professional development!

Is the outcome satisfactory from a scientific perspective, too?

Enhancements to SiGe Technology

Latent shortcomings – they first came to light in this thesis – in the silicidation of the exterior base region of the SiGe2-transistor bar the way to miniaturization. I propose a means to circumvent this roadblock.

In the course of refining the SiGe process architecture, several contributions to process integration have been made[3]. Among these, there is a new method to quantify oxidation-induced strain in silicon without the need for micro-Raman scattering.

Physical Simulation

Earlier, while working on high-voltage integrated DMOS I did profit a lot from presimulation (c.f. chapter 18). In the status quo ante of this work, neither at Atmels's nor anywhere else in industry has there been any noticeable success on SiGe device simulation. Tying up a man-year of ressources to establish simulation proficiency requires the necessary 'allies'. The year was distinct in accepting coalitions and keeping many an in-house contact warm.

The output speaks in favor of the risk taken: Starting from very basic experimental data such as SEM cross-sections and SIMS profiles all electrical parameters are reproduced well within the experimental uncertainty. To my knowledge, this is (at least) an industry first!

[3] Being day-to-day business, I took them for non-essentials and mainly left them out.

Neccessary for that success have been

- material parameters adopted from Hendrik Mau [III.7];
- knowledge of the experiment: as an example, SIMS, spreading resistance, and X-ray diffraction offer complementary information on depth profiles;
- skilfull grid definition;
- in-depth consideration of device physics targeting a thorough understanding of transistor behaviour, especially within the collector region.

With respect to the last issue, let me refer to fig. 6.5. Whereas Johnson ([I.3]) detects, 'doubling transit frequency of a high-voltage semiconductor device cuts breakdown voltage by half', I demonstrate 'doubling transit frequency of a high-frequency SiGe transistor requires three times the current carrying capability'[4]!

Simulating the electrical realities within the device's cross-section confirmed the worries on miniaturisability of the current technology.

Circuit Technology

A third strand approaches SiGe from the application side. Targeting mobile communication market, circuit requirements lead to solutions by circuitry, resolving inherent limitations in device technology.

As well new, as startling is the result that bipolar cascodes cock a snook at the Johnson limit. Given a suitable source impedance at the cold base, the cascode is defying all competition: Superior to common emitter by a factor of three in breakdown voltage, endowed with twice the gain and considerably higher transit frequency, and on-chip matchable up to outstandingly high output powers. Furthermore, and most surprisingly, the cascode offers first-rate efficiencies.

Being used as a power stage, the laterally integrated cascode circuitry is difficult to wire: Excessive amounts of leads with all kinds of parasitics cut down high-frequency suitability. Application needs led me to suggest a new type of high-frequency devices[5]:

The Vertically Integrated Cascode Stage

Two stacked transistors provoke the instant objection of being a thyristor, very fast and showing fair blocking ability – but not really noticed for best-in-class linearity... This injection will be dealt with.

A less obvious and far more valid doubt is the challenge to contact buried regions. The solution put forward for the vertically integrated cascode structure is advantageous for SiGe-HBTs too: the use of silicides, burried beneath solid-phase epitaxial silicon.

---

[4] Sections 6.1 to 6.3 show how to draw conclusions from the current carrying cabability on breakdown voltage for short collector diodes.

[5] In the course of the examination, a patent officer raked up power devices of a similar making; in the end, the new concept was held to be innovative and the patent was granted.

An excessive literature survey serves to explore the aptitude of the individual silicides. Certain, in themselves quite marginal features of existing technology architectures render the present approach almost plausible, hinting that the use of lateral solid-phase expitaxy of silicon on silicide was practicable.

Based on this presumption, a process flow is proposed targeting the vertically integrated cascode structure. The process flow fits beautifully as an optional module into Atmel SiGe technology architecture.

Linking the Strands

Semiconductor physics leads to doping constraints to keep the thyristor-like structure from firing. Adding geometrical requirements from the process flow, there is sufficient information for device simulation. Proficiency in physical simulation offers the possibility to predict the electrical properties of the vertically integrated cascode structure. Simulation results proove the superiority to the constituting bipolar transistors.

Using simulated device behaviour, a large-signal model was compiled, scalable in geometry and doping levels. With the help of the model the doping profile is refined further. Based on the model, circuit simulation affirms paramount compression behaviour as well as pre-eminent efficiency.

Starting from experimental challenges in process integration ('How to improve Atmel's SiGe?'), passing the realms of theory, circuitry and simulation, the work circles back to a new and exciting experimental question: the task of lateral solid phase epitaxy of silicon on silicide.

Evidence is lacking, as well with the proposed device, as with the proposed production flow: With a daunting gesture, I've added a grain of sand to the mountains[6].

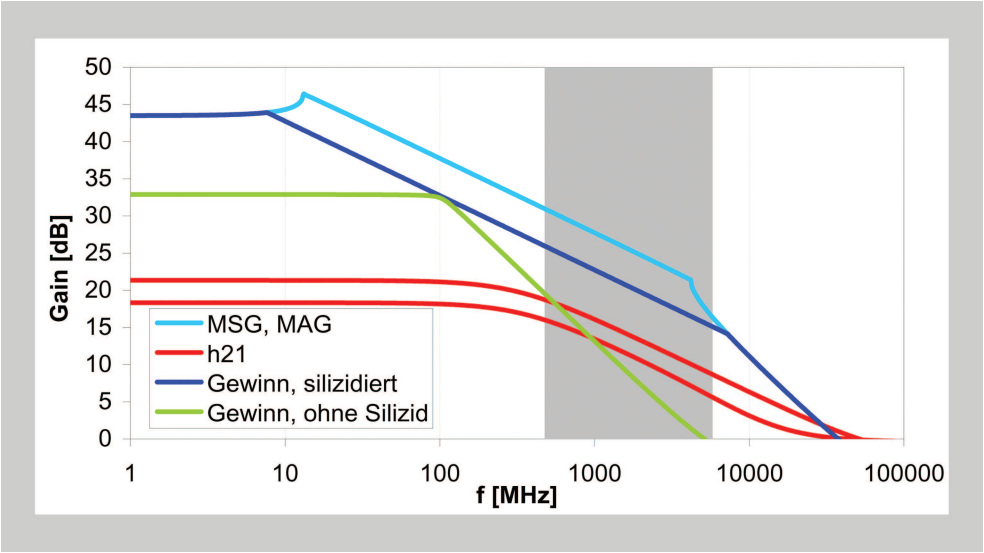
Hopefully, the work will not degenerate into a whale-size wit without experimental answers. But wit and phantasy will allways be inspiration and expiration of all scientific work. And in the end, it will not be Phantasos, master of wit, but his oncle Thanatos, brother of sleep, answering all questions.

---

[6] Even without experimental proof-of-concept, the innovational content of the work at hand is not insignificant: My two years at Prof. Schumacher's bred over 20 patent applications. By nature, mostly applications, centering around technology architecture and device structures left their marks. Eight patents on circuitry (including power electronics) did not gain due recognition here.

One further invention on electronic design is being used in the coming ADS power amplifier design guide. Ideas on measuring technology are being pursued in co-operation with Rohde & Schwarz and the LTE chair of Prof. Weigel and are being funded by the state of Bavaria. Insight on the mathematical structure of lossless matching led to deeper understanding of Doherty amplifiers, being exploited together with NXP in Nijmegen.

I Einleitung



Come hether friend, I am ashamed to hear that what I hear of you . . . You have almost attayned to the age of nyne yeeres, at least eight and a halfe, and seeing that you knowe your dutie, if you neglect it you deserve greater punishment then he which through ignorance doth it not. Think not that the nobilitie of your Ancestors doth free you to doe all that you list, contrary-wise, it bindeth you more to followe vertue.

Pierre Erondell,

THE FRENCH GARDEN: FOR ENGLISH LADIES OR GENTLEWOMEN TO WALKE IN [I.1]

Bipolar- wie auch Hetero-Bipolartransistor (HBT) setze ich als im Prinzip bekannt voraus. Das einleitende Kapitel dient lediglich der Hinführung an das Thema.

Die Verfügbarkeit schneller und komfortabler Simulationswerkzeuge verstellt leicht den Blick für einfache Lösungswege, die eher dem Verständnis des untersuchten Gegenstandes dienen. Weniger geistferne, „altmodische“ Lösungswege ergeben häufig besser nachvollziehbare Lösungen und fördern das Verständnis eher – „The nobilitie of our Ancestors bindeth us to followe vertue“.

In die Untersuchung des Transistorverhaltens steige ich daher, nach einem kurzen Abriß zum HBT, mit einer analytischen Behandlung des Hybrid-Pi-Modells ein. Der geneigte Leser möge das einführende Kapitel betrachten wie den Vorortzug, an den sich im weiteren Verlauf der Reise schnellere und komfortablere Verbindungen anschließen; auch in einer Vorlesung spielt es die selbe Rolle wie der Vorortzug: die des Lumpensammlers, bis nach den Semesterferien alle Studenten wieder an Bord sind. Gleichzeitig mag es als Anregung für Übungsaufgaben dienen.

1 Silizium-npn-Transistor

Zur Hinführung gehe ich vom herkömmlichen Silizium-npn-Transistor aus. Ich charakterisiere ihn mit drei Sorten Kenngrößen, welche a) das Gleichstromverhalten wiedergeben, b) die Hochfrequenzeignung messen, respektive c) das Kompressionsverhalten darstellen .

1.1 Das Gleichstromverhalten

Ein bewährtes Bild aus der Schaltungstechnik ist die Idee, ein Bipolartransistor verstärke einen in die Basis eingeprägten Strom. Leider verstellt dieses Bild den Blick für die physikalischen Vorgänge.

Die Majoritätsladungsträger in der Basis sind Löcher. Um „einen (konstanten) Löcherstrom in die Basis einprägen zu können“, müssen wir, bei gesperrter Kollektordiode, den Löcherüberschuß über den Emitter abführen. Hierfür ist erforderlich, die Emitterdiode in Vorwärtsrichtung zu polen.

Bei geöffneter Emitterdiode fließt gleichzeitig ein Elektronenstrom vom Emitter aus in die Basis. Dieser ist jedoch nur eingeschränkt eine **Folge** des Löcherstromes in die

Basis. Zugegeben, es ist ein Henne-Ei-Problem, Potentialdifferenz und Löcherstrom sind untrennbar aneinandergekoppelt. Bei der theoretischen Behandlung der Vorgänge in der vertikal integrierten Kaskodenstufe im Abschnitt 10.1 werden wir jedoch besser fahren mit der Betrachtungsweise, **der Elektronen- wie der Löcherstrom seien unabhängige Folgen der Potentialdifferenz**. Der Löcherstrom ist dann lediglich der Preis, den wir zahlen müssen, um die Potentialdifferenz aufrecht zu erhalten, die dem Elektronenstrom zugrundeliegt.

Die erste Aufgabe beim Bau eines Bipolartransistors ist, das Bauelement so zu gestalten, daß der Elektronenstrom deutlich größer als der vom Elektronenfluß unabhängige Löcherstrom ist – dies ist, wie weiter unten gezeigt wird, das grundlegende Motiv hinter dem HBT.

Zu beiden Seiten einer in Vorwärtsrichtung gepolten Raumladungszone herrscht ein Überschuß an Minoritätsladungsträgern. Während die Potentialdifferenz zwischen Basis und Emitter die Minoritätsladungsträgerkonzentrationen an den Grenzen der Raumladungszone bestimmt, ist der Abfluß der Überschuß-Minoritätsladungsträger durch die Diffusion in den neutralen Bauelementbereichen bestimmt.

Für die kurze Diode ist der Diffusionsstrom proportional zu $1/(N_X \cdot L_X)$, wobei N_X die Dotierung und L_X die Ausdehnung des Gebietes X, entlang des Stromflusses gemessen, bezeichnet (zur Herleitung s. S. 76): Die Abhängigkeit von N_X rührt von der Proportionalität der Überschußladungsträgerkonzentration zu der Gleichgewichts-Minoritätsladungsträgerdichte, welche ihrerseits proportional ist zum Kehrwert der Dotierung. Die Überschußladungsträgerkonzentration resultiert in einem Konzentrationsgradienten, proportional zum Kehrwert des Abstandes L_X bis zu der Stelle, in dem die Ladungsträgerkonzentration ins Gleichgewicht gesetzt wird.

Dies gilt für den Elektronen- wie für den Löcherstrom. Die Stromverstärkung eines npn-Transistors als Verhältnis des Elektronen- zum L“ocherstrom ist gleich dem Verhältnis der Dotierstoffintegrale $\int N(x)dx$ in Basis und Emitter. Eine hohe Verstärkung erfordert eine kleine „Gummelzahl“ der Basis (zu Ehren des Erfinders der Ladungskontrollrelationen firmieren Dotierstoffintegrale über neutrale Halbleiterbereiche bei Bipolartransistoren als Gummel-Zahlen).

Wenn die Stromverstärkung von der Gummelzahl der Basis abhängt, wird sie beeinflußt von allen Vorgängen, die die Gummelzahl verändern. Der wichtigste Einfluß ist das Kollektorpotential: Ändert sich der Sperrspannung über der Basis-Kollektor-Diode, so atmet die Basis-Kollektor-Raumladungszone, und zwar sowohl basis- wie kollektorseitig. Dehnt sich die Raumladungszone in die Basis hinein aus, ändert sich das Dotierstoffintegral über die neutrale Basis, und mithin die Verstärkung des Transistors: Je höher die Kollektorspannung, desto kleiner ist die Gummelzahl der Basis, und desto höher wird die Verstärkung.

Es stellt sich heraus, daß sich sämtliche Ausgangskennlinien zu jeweils festem Basisstrom, durch Geraden angenähert, in einem Punkt auf der negativen Spannungsachse schneiden. Der Abstand dieses Punktes vom Ursprung ist bekannt als Early-Spannung: Je größer die Early-Spannung eines Transistors, desto besser die Konstanz der Verstärkung mit der Kollektorspannung.

Die relative Änderung der Basis-Gummelzahl mit der Kollektorspannung wird um so größer, je kleiner das Dotierstoffintegral über die gesamte Basis ist: Je höher die Verstärkung, desto kleiner ist die Early-Spannung.

Eine ausreichende Basis-Gummelzahl ist daher Vorraussetzung für eine hohe Early-Spannung. Um gleichzeitig eine befriedigende Verstärkung zu erzielen, ist man auf eine hohe Gummelzahl des Emitters angewiesen. Die Emitterdotierung findet jedoch ihre Grenze nicht erst in der Aktivierbarkeit der Dotanden, bzw. der Löslichkeit der Dotanden in Silizium: Eine hohe Dotierung verringert die Bandlücke. Die Minoritätsladungsträgerdichte ist nicht nur umgekehrt proportional zur aktiven Dotierstoffkonzentration, sondern hängt auch exponentiell vom Wert der Bandlücke ab. Durch eine sinkende Bandlücke steigt bei höchsten Emitterdotierungen der Basisstrom wieder an.

Man findet, daß für einen gut konstruierten Transistor das Produkt aus den beiden wesentlichen Gleichstrom-Kenngrößen eines Transistors, Stromverstärkung und Earlyspannung, weitgehend unabhängig ist von technologischen Maßnahmen zur weiteren Optimierung des Transistoraufbaus.

1.2 Die Hochfrequenzeignung

Zwei wesentliche Maßzahlen für das Hochfrequenzverhalten sind die Transitfrequenz f_T und die maximale Schwingfrequenz f_{max}.

Die Transitfrequenz wird durch Lauf- und Ladezeiten bestimmt. Ein wichtiger Anteil der Laufzeiten, die Basis-Laufzeit, bezeichnet dabei eigentlich die Umladezeit der Basis-Minoritätsladungsträgerpopulation. Die Basis-Laufzeit wird kürzer bei dünnerer Basis – daher der anschauliche Name.

Die Umladezeiten der Sperrschicht werden begrenzt vom Verhältnis aus Stromdichte und Kapazitätsbelag. Wie im Abschnitt 6.4 gezeigt, sinkt die erreichbare Ladezeit mit zuehmender Kollektordotierung.

Entsprechend ist es nicht möglich, das Gleichstrom- und das Hochfrequenzverhalten gemeinsam zu optimieren: Ein hohes f_T erfordert eine dünne Basis und einen hochdotierten Kollektor – und erzwingt daher ein miserables Early-Verhalten.

1.3 Das Kompressionsverhalten

Das Kompressionsverhalten, also der Gang der Verstärkung mit der Ausgangsleistung, ist ein besonders heikler Punkt: Im allgemeinen nimmt die Verstärkung eines Transistors mit 20 dB pro Frequenzdekade ab. In Arbeitspunkten, in welchen bereits die Gleichstrom-Leistungsverstärkung gering ist, ist daher auch die Verstärkung bei einer gegebenen Frequenz schlecht.

Wird ein Transistor um einen Arbeitspunkt herum ausgesteuert, so durchläuft er während jeder Schwingungsperiode unterschiedliche Stellen im Ausgangskennlinienfeld. Die Großsignalverstärkung als Verhältnis der integralen Aus- und Eingangsleistung läßt sich durch das Integral über die lokale Kleinsignalverstärkung entlang der Ausgangskennlinie darstellen.

Variiert die Niederfrequenzverstärkung entlang der Ausgangskennlinie beträchtlich, wird die Verstärkung bei der Betriebsfrequenz in jeder Periode des HF-Signales stark vom Mittelwert abweichen: ein „schlechtes“ Gleichstrom-Ausgangskennlinienfeld, insbesondere eine niedrige Early-Spannung, führt zu einem suboptimalen Kompressionsverhalten, und bereits die Nicht-Konstanz der Niederfrequenzverstärkung über dem Ausgangskennlinienfeld führt zu einer Verzerrung des HF-Signales.

2 Wirkprinzip des HBT

Der grundlegende Gedanke hinter dem Hetero-Bipolartransistor (HBT) ist, den Emitter mit einer größeren Bandlücke als die Basis auszuführen (sehr übersichtlich [I.2]).

Man kann die Bandlücke in der Basis verringern, etwa, indem man der Basis eines Siliziumtransistors 20 % Germanium zulegiert. Germanium ist, ebenso wie Silizium, ein Gruppe-IV-Halbleiter; anstatt im Silizium als Dotand zu wirken, wie die Fremdstoffe, welche man üblicherweise beimengt, bildet es mit Silizium in beliebigen Mischungsverhältissen Halbleiter-Mischkristalle, deren Bandlücken zwischen den Werten für Silizium (1.12 eV) und Germanium (0.66 eV) vermitteln. (Auch die Gitterkonstante variiert mit dem Germaniumgehalt. Dies begrenzt den versetzungsfrei einzubringenden Germaniumgehalt in der Basis des SiGe-HBT.)

Welchen Vorteil erhält man aus unterschiedlichen Bandlücken in Basis und Emitter? Elektronen, welche vom Emitter in die Basis gelangen, sind dort Minoritätsladungsträger. Der Diffusionsstrom der Minoritätsladungsträger hängt vom Minoritätsladungsträgergradienten, der Minoritätsladungsträgergradient von der Gleichgewichts-Minoritätsladungsträgerdichte und diese wiederum exponentiell vom Kehrwert der Bandlücke ab. Durch einen kleineren Wert der Bandlücke in der Basis wird der Diffusionsstrom von Elektronen in der Basis, und damit der Strom von Elektronen, die den Emitter verlassen, exponentiell erhöht. Gleichzeitig bleiben die Verhältnisse im Emitter, welche den Basisstrom bestimmen, gleich: Die Gleichstromverstärkung steigt exponentiell mit dem Unterschied in den Bandlücken, gemessen in Einheiten von $k_B\,T$.

In der Folge ist kein hohes Verhältnis der Dotierstoffkonzentrationen in Emitter und Basis mehr nötig, um eine hohe Verstärkung zu erzielen. Dies ist im Allgemeinen aber kein Designziel: Die Durchbruchspannung BV_{CE0} bei Betrieb des Transistors in Emitterschaltung sinkt mit hoher Stromverstärkung. Eine unnötig hohe Stromverstärkung führt, auf dem Umweg über eine verringerte Gummelzahl der Basis, zu einem schlechten Early-Verhalten.

Tatsächlich nutzt man daher die zusätzliche Freiheit im Design durch den Einbau von Germanium in die Basis nicht, um die Verstärkung erhöhen, sondern um die Dotierstoffverhältnisse in Basis und Emitter umkehren zu können:

- Durch eine hohe Basis-Gummelzahl werden direkt das Early-Verhalten und (über den Basis-Bahnwiderstand) die maximale Schwingfrequenz verbessert sowie (über das bessere Gleichstrom-Kennlinienfeld) das Kompressionsverhalten.

- Die höhere Basisdotierung erlaubt, die Kollektordotierung anzuheben, ohne das Early-Verhalten zu verschlechtern; die hierdurch ermöglichte höhere Stromdichte verbessert die Transitfrequenz.
- Durch die niedrigere Emitterdotierung sinkt die Emitterkapazität bei einer niedrigen Basisspannung; f_T bei kleinen Stromdichten steigt an, was insbesondere den Bau schneller LNAs erleichtert, zumal das Rauschverhalten ebenfalls von dem niedrigen Basis-Bahnwiderstand profitiert. Gleichzeitig reduziert die niedrige Emitterdotierung die Tunnelrate am Basis-Emitter-Übergang und erleichtert ein sauberes Sperren der Emitterdiode.

2.1 SiGe und das Johnson-Limit

Ausgehend von der Sättigungsgeschwindigkeit, der Weite der gesperrten Raumladungszone und der Durchbruchfeldstärke von Silizium hat E. O. Johnson 1965 ([I.3]) gezeigt, daß hochsperrende Silizium-Leistungtransistoren höchstens mit 200 GV/s schalten können[7].

Es erscheinen immer wieder Aufsätze, deren Autoren mit Hilfe von SiGe „das Johnson-Limit durchbrochen“ haben[I.4].

Die 200 GV/s wurden mit Blick auf niedrigdotierte, hochsperrende Silizium-Vertikaltransistoren hergeleitet. Die Durchbruchfeldstärke im Halbleiter steigt, aufgrund der Coulomb-Streuung, mit der Dotierung an. Für Spannungsfestigkeiten unter 10 V ergibt sich in Silizium ein Zahlenwert von weit über 500 GV/s. Entsprechend liegt es nicht an der Germanium-Beimengung in der Basis, daß die HF-SiGe-HBTs schneller als mit 200 GV/s schalten, sondern an der zur f_T-Optimierung erforderlichen hohen Kollektordotierung. (*... das sind die rechten Mohren, die sich über die schwarzen Flecken in der Sonne freuen!.* Arno Schmidt [I.5])

[7] Das Johnson-limit ist eigentlich das materialabhängige Produkt der Basis-Kollektor-Durchbruchspannung mit der Transitfrequenz der Stromverstärkung. Häufig wird es auch zur Abschätzung der erzielbaren Transitfrequenz in Abhängigkeit von der Emitter-Kollektor-Durchbruchspannung herangezogen wobei der Einfluß der Stromverstärkung durch empirisch motivierte Faustformeln berücksichtigt wird.

3 Das Hybrid-Pi-Modell

In Vorbereitung auf die technologischen Grundlagen im nächsten Kapitel untersuche ich, wieweit eine Veränderung einzelner Bauelementparameter (wie z.B. des Basis-Bahnwiderstandes) einen Einfluß auf das Bauelementverhalten hat. Als Vehikel bediene ich mich des Hybrid-Pi-Modells.

Zum Zweck der Schaltungssimulation werden elektronische Bauelemente in Form von Kompaktmodellen wiedergegeben. Kompaktmodelle beschreiben das Verhalten des realen Bauelements, indem sie es in Form einer Ersatzschaltung aus numerisch leicht beherrschbaren idealisierten Bauelementen nachbilden.

Zu einem besonders einfachen Modell, einem Kleinsignalmodell, gelangt man, indem man das „wahre“ Bauelementverhalten um einen Arbeitspunkt herum entwickelt und nur die Glieder bis zur linearen Ordnung behält. Die Gültigkeit des Kompaktmodells wird um so schlechter, je weiter man das Bauelement aus dem betrachteten Arbeitspunkt heraus aufsteuert.

Das Hybrid-Pi-Modell ist ein solches Kleinsignalmodell. Als lineares Modell läßt es sich einfach in Gleichungsform bringen und analytisch behandeln. Aus der Lösung lassen sich bereits weitgehende Schlüsse für die Weiterentwicklung der zugrundeliegenden Technologie ziehen.

Abb. 3.1 zeigt die Schaltungstopologie des Hybrid-Pi-Modells:

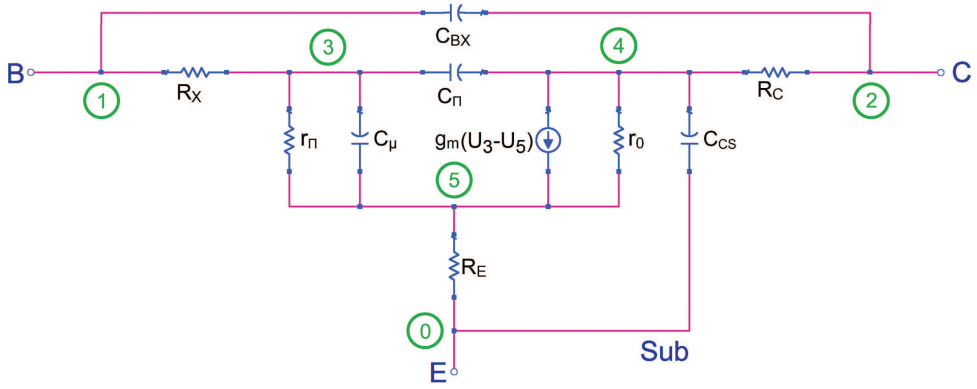


Abb. 3.1: Schaltbild des Hybrid-Pi-Modells in Emitterschaltung

R_x ist vor allem der Anschlußwiderstand durch die äußere Basis und der innere Basis-Bahnwiderstand. Da C_{BX} an Knoten 1 angreift, enthält R_x nicht den Zuleitungswiderstand irgendwelcher Metallbahnen. Dies ist einleuchtend, da das Modell zur Beschreibung des Transistors in einer integrierten Schaltung extrahiert wurde.

r_π ist der differentielle Widerstand der in vorwärtsgespannten Basis-Emitter-Diode, $dI_3 \,/\, dU_3 = I_S/U_T \exp(U_{35}/U_T)$, und C_μ die Basis-Emitter-Sperrschichtkapazität.

g_m bezeichnet die Steilheit des Transistors, g_m=I_4/U_{35} für U_{45}=0. C_π steht für die Kollektor-Basis- und C_{CS} für die Kollektor-Substrat-Kapazität.

Mit Hilfe des Kollektor-Emitter-Widerstandes r_0 wird insbesondere die Rückwirkung vom Ausgang auf den Eingang des Transistors modelliert.

3.1 Lösung des Hybrid-Pi-Modells

Die Schaltung aus Abb. 3.1 enthält ausschließlich lineare Bauelemente und läßt sich daher direkt in eine lineare Gleichung überführen. Da von den sechs Knoten einzig die drei Tore Ströme aufnehmen, hat die Admittanzgleichung die Form

$$\left[\begin{array}{c|c} & \\ \widetilde{Y}_{11} & \widetilde{Y}_{12} \\ & \\ \hline & \\ \widetilde{Y}_{21} & \widetilde{Y}_{22} \\ & \end{array}\right] \left[\begin{array}{c} U_0 \\ U_1 \\ U_2 \\ \hline U_3 \\ U_4 \\ U_5 \end{array}\right] = \left[\begin{array}{c} I_0 \\ I_1 \\ I_2 \\ \hline 0 \\ 0 \\ 0 \end{array}\right] \quad .$$

Um zu der Y-Matrix zu gelangen, wird ein Bauelement zwischen den Knoten a und b in der Y-Matrix wiedergegeben durch den Eintrag seines Leitwerts an der Stelle (a,b) der Matrix. Sämtliche Leitwerte in der selben Zelle der Matrix Y werden addiert.

Jeder Widerstand R_{ab} und jede Kapazität C_{ab} zwischen den Knoten a und b wird also ersetzt durch Matrixeinträge (Konvention: aus einem Knoten herausfließende Ströme werden positiv gezählt)

$$R_{\mathrm{ab}}: \begin{array}{c} \\ \mathrm{a} \\ \mathrm{b} \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{a} & \mathrm{b} \end{array} \\ \left[\begin{array}{cc} 1/R & -1/R \\ -1/R & 1/R \end{array}\right] \end{array}, \qquad \mathrm{C}_{\mathrm{ab}}: \begin{array}{c} \\ \mathrm{a} \\ \mathrm{b} \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{a} & \mathrm{b} \end{array} \\ \left[\begin{array}{cc} i\omega C & -i\omega C \\ -i\omega C & i\omega C \end{array}\right] \end{array} .$$

In Leitwert-Darstellung liest sich die gesteuerte Quelle g_m zwischen Knoten 4 und 5 mit Steuerspannung zwischen Knoten 3 und 5

$$g_m: \begin{array}{c} \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cc} 3 & 5 \end{array} \\ \left[\begin{array}{cc} g_m & -g_m \\ -g_m & g_m \end{array}\right] \end{array} .$$

Die homogenen Gleichungen, Zeile 3 bis 6 der Admittanzgleichung, werden nun aufgelöst, um U_3 bis U_5 als Funktionen von U_0 bis U_2 darzustellen. Einsetzen in die ersten drei Zeilen liefert die 3x3 Y-Matrix der Schaltung und damit den gesuchten Zusammenhang zwischen den Strömen und Spannungen an den Toren.

Die Summe der externen Ströme ist Null, die Zeilen der 3x3-Y-Matrix sind nicht unabhängig von einander. Oder, ausgehend von den Potentialen: Während in die Impedanzgleichung formal drei Spannungen eingehen, sind physikalisch nur zwei Spannungsdifferenzen von Bedeutung. Man kann auf das Potential eines beliebigen Ports referenzieren, aus dem Dreitport wird ein Zweitor.

Um die Y-Matrix invertieren zu können, muß man sie daher zunächst durch Streichen der zu dem Bezugsport gehörigen Zeilen und Spalten zu einer 2x2-Matrix verkürzen.

Im Beispiel lege ich mich auf die Emitterschaltung fest. Der Emitterstrom wird als abhängig betrachtet und die erste Zeile der Impedanzgleichung gestrichen. Referenzieren der Spannungen auf Emitterpotential entspricht dem Streichen der ersten Spalte. Die verbleibende 2x2 Y-Matrix ist invertierbar. Ich verwende sie, um auf die bekannte Weise Strom- und Leistungsverstärkung zu bestimmen.

Für Frequenzen, in denen das Modell nicht bedingungslos stabil ist, läßt sich die maximale stabile Leistungsverstärkung MSG nur unmittelbar auf dem Stabilitätskreis erzielen. Ein stabil unter Produktionsbedingungen zu erzielender (Leistungs-) Gewinn liegt erfahrungsgemäß 5 dB tiefer. Als „erzielbare Leistungsverstärkung“ trage ich daher das Minimum aus MSG-5 dB und MAG über der Frequenz auf.

Das Aufstellen der 6x6-Admittanzmatrix auf die beschriebene Weise ist einfach. Ebenso läßt sich mit Hilfe von Maple daraus leicht die 2x2-Y-Matrix des Hybrid-Pi in Emitterschaltung gewinnen. Auch die Bestimmung von Strom- und Leistungsverstärkung unter Anwendung der einschlägigen Formeln gelingt Maple in Sekunden.

Ein Ausdruck der so erhaltenen analytischen Lösung erstreckt sich über mehrere Druckseiten: In der vollen Allgemeinheit ist sie nur innerhalb von (bspw.) Maple beherrschbar. Nachfolgend gebe ich daher, anstelle der allgemeinen Lösung, zunächst einen im weiteren Verlauf wichtigen Spezialfall an, und stelle den Einfluß einzelner Parameter des Hybrid-Pi-Modells anhand von Grafiken der frequenzabhängigen Verstärkungen dar.

3.2 Spezialisierung der analytischen Lösung

Zur Beschreibung der elektrischen Eigenschaften eines Transistors in einem gegebenen Arbeitspunkt mit Hilfe des Hybrid-Pi-Modells ist es erforderlich, aus der Messung Zahlenwerte für die Modellparameter zu extrahieren. Als ein Beispiel gebe ich eine - m.E.n. neue - Methode die (arbeitspunktabhängigen) Diodenkapazitäten und „physikalischen“ Widerstände $C_\pi, C_\mu, R_X, R_C, R_E$ zu bestimmen. Zu diesem Zweck spezialisiere und vereinfache ich die mit Hilfe von Maple erhaltene analytische Lösung:

- Wie sich nachfolgend zeigen wird, kann man bei dem Aufbau der betrachteten Transistoren C_{CS} für den Einzeltransistor getrost und in jedem Falle vernachlässigen. Für einen im weiteren Verlauf der Arbeit vorgestellten verbesserten HBT gilt dies auch für C_{BX}.
- In ausgeschaltetem Zustand (das heißt, wenn entweder nur die Basis- oder nur die Kollektorspannung von Null verschieden ist) gilt darüberhinaus g_m=0 und die Leitwerte von r_π und r_μ verschwinden ebenfalls.

Es ergibt sich $$Z_{off} = \begin{bmatrix} R_X + R_E + \frac{1}{j\omega C_\pi} & R_E + \frac{1}{j\omega C_\pi} \\ R_E + \frac{1}{j\omega C_\pi} & R_X + R_E + \frac{1}{j\omega C_\pi} + \frac{1}{j\omega C_\mu} \end{bmatrix}.$$

- Ich betrachte den Grenzfall kleiner Frequenzen und entwickle nach ω:

$$\lim_{\substack{\omega \approx 0 \\ r_0, r_\pi \to \infty \\ C_{BX}, C_{CS}, g_m = 0}} \mathbf{Y} = j\omega\, \mathbf{j_y} + \omega^2\, \mathbf{r_y}\ , \qquad \mathbf{j_y} = \begin{pmatrix} C_\mu + C_\pi & -C_\mu \\ -C_\mu & C_\mu \end{pmatrix}$$

$$\mathbf{r_y} = \begin{pmatrix} R_X(C_\mu + C_\pi)^2 + R_C C_\mu^2 + R_E C_\pi^2 & -R_X C_\mu (C_\pi + C_\mu) - R_C C_\mu^2 \\ -R_X C_\mu (C_\pi + C_\mu) - R_C C_\mu^2 & C_\mu^2 (R_X + R_E) \end{pmatrix}$$

Die Bedingung, unter der ich die Y-Matrix spezialisiert habe, war, daß jeweils nur die Basis- oder nur die Kollektorspannung von Null verschieden ist. Durch Messung des Niederfrequenverhaltens der Leitwerte des Transistors in ausgeschaltetem Zustand steht somit ein einfaches Verfahren zur Verfügung, um die Spannungsabhängigkeit der Sperrschichtkapazitäten zu messen: Sechs der Einträge aus $\mathbf{j_y}$ und $\mathbf{r_y}$ sind unabhängig und lassen sich eindeutig nach R_X, R_E, R_C, C_π und C_μ auflösen.

3.3 Zielgrößen der Technologieentwicklung

In diesem Kapitel verwende ich das Hybrid-Pi-Modell eines bestehenden SiGe1-Transistor, um Hinweise für die Technologieoptimierung zu erhalten. Anstatt die Hybrid-Pi-Parameter aus der Messung zu extrahieren, schummel ich ein wenig und bediene mich des hervoragenden SiGe1-Designkits von Herrn Dr. Kraus, Atmel (unveröffentlicht, Anfragen an Fa. Atmel, Heilbronn).

Um zu Zahlenwerten zu gelangen, extrahiere ich aus einem skalierbaren Großsignalmodell mit Hilfe von ELDO die Hybrid-Pi-Parameter für einen Transistor mit selektiv implantiertem Kollektor mit 5.6 µm Emitterlänge und 1.2 µm Emitterbreite in SiGe1-Technologie, im f_{max}-optimalen Arbeitspunkt: Mit den gegebenen Zahlenwerten läßt

C_π	C_μ	C_{BX}	C_{CS}	g_m
1.29 pF	37.1 fF	29.2 fF	50.8 fF	0.45422
R_E	R_C	r_0	R_X	r_π
1.7 Ω	9.3 Ω	2.1 kΩ	26.6 Ω	301.5 Ω

sich der Einfluß der Parameter auf den Frequenzgang der Verstärkung untersuchen. Zu diesem Zweck stelle ich, ausgehend von der vollen analytischen Lösung, in Maple den Frequenzverlauf der Verstärkungen dar:

3.3.1 Der Basis-Bahnwiderstand und die maximale Schwingfrequenz

In Abb. 3.2 sind für drei Werte von R_X Stromverstärkung (rot) und erzielbare Leistungsverstärkung (blau) – das Minimum aus MSG-5 dB und MAG, s.o. – sowie MAG / MSG (lichtblau) über der Frequenz aufgetragen. Neben dem Basis-Bahnwiderstand des Modells habe ich den halben und den doppelten Wert angesetzt:

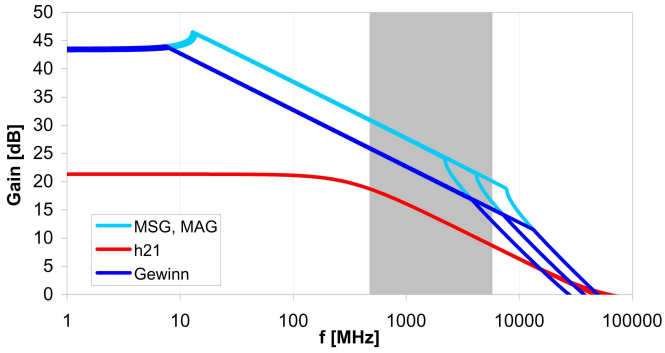


Abb. 3.2: Variation des Basis-Bahnwiderstands: 50%, 100%, 200% vom Modellwert. (Grafik mit Maple aus der analytischen Lösung gewonnen; schraffierter Bereich: Ziel-Anwendungsfrequenzen der Technologie)

Zunächst läßt sich aus der Grafik ablesen, daß eine Halbierung des Basis-Bahnwiderstandes R_X der SiGe1-Technologie die maximale Schwingfrequenz f_{max} um 30 % erhöht. Dies ist eine große Hilfe für Anwendungen oberhalb der Mobilfunk- und Wimax-Frequenzen.

Tatsächlich war die Hauptzielrichtung für die Weiterentwicklung von SiGe1 zu SiGe2, den Massenmarkt Mobilfunk für eine siliziumbasierte Technologie zu erschließen. Im hierfür relevanten Frequenzbereich – er ist in Abb. 3.2 grau hinterlegt – hat die Verbesserung von f_{max} keinerlei Einfluß auf die Leistungsverstärkung! Dies zeigt gleichzeitig, daß die maximale Schwingfrequenz nicht in jedem Fall eine gute Zielfunktion für die Technologieoptimierung darstellt.

Nähere Untersuchung auch der übitrgien Modellparameter ergibtm, daß im betrachteten Frequenzbereich – in der Umgebung der gegebenen Parameterwerte – die Stromverstärkung lediglich durch C_π, die Leistungsverstärkung dagegen durch C_μ, C_{BX} und R_E beeinflußt wird.

3.4 Grenzen des Modells, I: Der Kollektorwiderstand

Insbesondere bietet für SiGe1 der Kollektorwiderstand im Rahmen des Hybrid-Pi-Modells kein Verbesserungspotenzial. Dies ist ein Artefakt des auf einen Einzeltransistor beschränkten Modells – innerhalb einer integrierten Schaltung ist der Kollektorwiderstand sehr wohl wichtig!

Bei einem zweistufigen Verstärker bildet der Kollektorwiderstand der ersten Stufe mit dem Eingangswiderstand der zweiten einen Spannungsteiler. Die Eingangsspannung der zweiten Stufe erscheint exponentiell im Ausgangssignal. Während der Kollektorwiderstand für den Einzeltransistor kaum Auswirkungen zeigt, geht er – in dB! – linear in die Verstärkung der zweistufigen Schaltung ein.

Ein Detail der Simulationen von Herrn Dr. Ertugrul Sönmez zu einem zweistufigen LNA in SiGe2RF für 24 GHz läßt sich in diesem Zusammenhang verstehen: Bei Variation des Kollektorwiderstandes im Mextram-Modell findet man für jedes Ω Kollektorwiderstand einen Verlust von 1 dB Verstärkung (mündliche Mitteilung).

3.5 Grenzen des Modells, II: Die Kollektorkapazität

Als nächstes wird die Kollektorkapazität variiert. Die elektrischen Eigenschaften bleiben vollkommen unverändert; C_{CS} ist in SiGe1 bereits so gering, daß die Eigenschaften des Einzelbauelements von anderen Größen dominiert werden. Auch hier zeigen sich die Grenzen des auf einen Einzeltransistor begrenzten Modells: Die Substratkapazität hat Aspekte, die im Wortsinne über den einzelnen Transistor „hinausreichen“: Sie schafft eine Verkoppelung zwischen benachbarten Bauelementen[8]. Nicht wegen der Eigenschaften des einzelnen Transistors, sondern um das Verhalten der Gesamtschaltung beherrschen zu können, muß die Substratkapazität gering gehalten werden.

3.6 Kollektorkapazität und SOI

Eine bekannte Möglichkeit, um die Kollektorkapazität zu verringern, wird besonders bei hochintegrierten CMOS-Technologien angewandt: der Einsatz von SOI-Wafern. Ein SOI-Wafer besteht aus einem Handling-Wafer und einer dünnen, monokristallinen Bauelementschicht, getrennt durch eine Oxidlage. Während bei Wafern ohne vergrabene Oxidschicht („Bulk“) die Bauelemente durch einen gesperrten pn-Übergang gegen das Substrat isoliert sind, sind sie auf SOI dielektrisch vom Substrat getrennt.

Der Kapazitätsbelag der Substratdiode wird beim Bulk-Wafer von der Weite der Raumladungszone bestimmt. Die Weite der Raumladungszone, und mithin die Sperrschichtkapazität, variiert mit der Wurzel der Dotierung: Je höher die Subtratdotierung, desto enger die Raumladungszone (bei gegebener Sperrspannung) und desto höher die Kapazität.

Beim Übergang von Bulk- zu SOI-Wafern wird die Sperrschichtkapazität der Substratdiode durch eine Oxid-Kapazität ersetzt[9]. Die Oxid-Kapazität läßt sich durch eine hohe Oxiddicke reduzieren. Dies findet seine Grenze in den unterschiedlichen Ausdehnungskoeffizienten von Silizium und Oxid: Oberhalb von etwa 0.5 µm dicken vergrabenen Oxidschichten verbiegt sich der Wafer beim „Rapid Thermal Annealing“, einer wichtigen Methode, Dotierungsschäden auszuheilen und Dotierstoffe im Silizium zu aktivieren.

Hochintegrierte CMOS-Technologien arbeiten, um verschiedenen Skalierungseffekten zu begegnen, mit Substratwiderständen bis hinunter zu 0.1 Ωcm. 500 nm Oxid haben die selbe Kapazität wie die Sperrschicht bei 3 Ωcm spezifischem Substratwiderstand kurz vor dem Durchbruch. Der Übergang zu SOI-Material reduziert hier die Substratkapazität deutlich.

---

[8] Zur Verdeutlichung dieser Verkopplung gibt es ausgezeichnete Arbeiten aus dem Kompetenzzentrum „Integrated Circuits in Communication“, etwa [I.6].

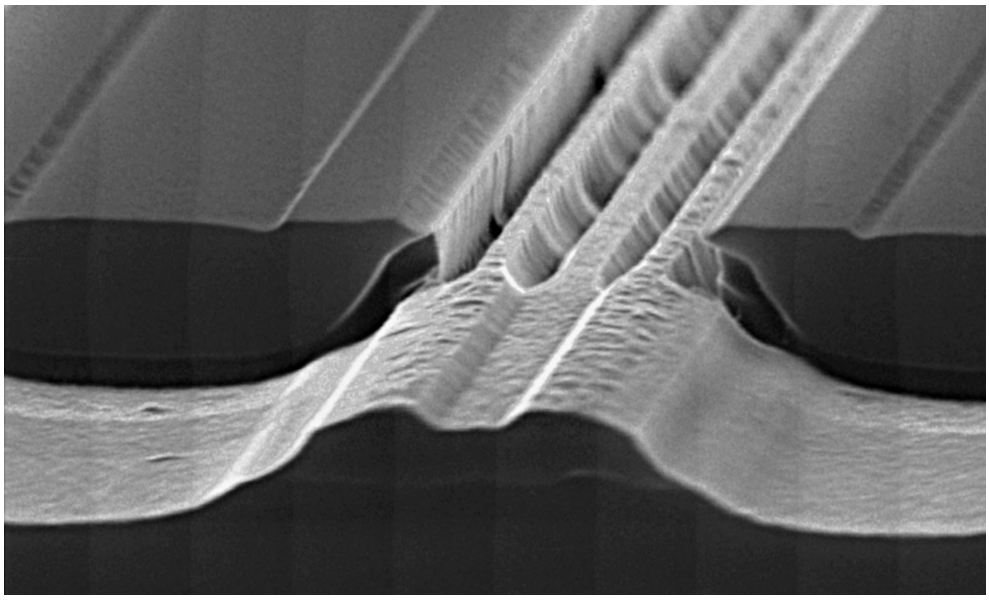
[9] Dies gilt cum grano salis, bildet sich doch auch unterhalb der vergrabenen Oxidschicht eine Raumladungszone aus. Um das Übersprechen zwischen benachbarten Bauelementen reduzieren zu können und so die Packungsdichte zu erhöhen, wird jedoch der Handling-Wafer im allgemeinen hochdotiert ausgeführt, so daß die Substratkapazität tatsächlich im wesentlichen von der Oxidkapazität bestimmt wird.

Analoge HF-Technologien dagegen verwenden Substratwiderstände von 20-1000 Ωcm. Der Übergang zu SOI-Wafern **erhöht** bei einem niederohmigen Handlingwafer die Substratkapazität beträchtlich. Bei einem hochohmigen Handling-Wafer dagegen sind, in erster Näherung, die Oxid- und die Sperrschichtkapazität in Serie als Substratkapazität wirksam. Die Oxidschicht reduziert die Substratkapazität lediglich unwesentlich.

Tatsächlich jedoch verspricht der Übergang zu SOI an anderer Stelle auch für die Hochfrequenz deutliche Vorteile. So ist es möglich, das vergrabene Oxid als Sandwich aus zwei durch die Oxidlage getrennten Silizidschichten auszuführen. Isolationsgräben, welche in der Oxidschicht enden, schneiden höchst leitfähige Kollektoranschlußbahnen aus der oberen Silizidschicht. Die untere Silizidschicht spreizt die Abwärme der Transistoren und stellt eine nahezu ideale Masselage zur Hochfrequenzisolation zur Verfügung. Durch deep trenches auf die Waferoberseite gezogen, dient sie der quasi induktionsfreien Masseanbindung.

Ein derartiger Waferquerschnitt wird vorgeschlagen in [I.7], während sich ein BMBF-Projekt mit Namen Isosurf (Isolation schemes for ultra high RF-circuits) der Untersuchung vergrabener Oxid- und Silizidschichten widmet, unter Beteiligung insbesondere des FZ Jülich, der TU Chemnitz, dem Fraunhofer IZM und der Firma Atmel [I.8].

II Grundlagen



Nun sind die electrischen Versuche aber, wie ein jeder weiß, von der Art, daß auch ein Ungelehrter, und der mit Maschinen nur ein wenig umzugehen gelernt hat, sie gar leicht vor sich anstellen kann. (...) Die Herrn Gelehrten, und am allermeisten meine Hoechst-geehrteste Herrn, werden also leicht so billig seyn, und glauben, daß auch mir nicht unmöglich gefallen, die electrischen Versuche so gut, als ein anderer, anzustellen.

Akronym F. W. [II.1]

SCHREIBEN AN DIE KÖNIGLICHE ACADEMIE DER WISSENSCHAFTEN ZU BERLIN, DIE UNTERSUCHUNG DER ELECTRIZITÄT / UND DEN VON IHR BESTIMMTEN PREISZ BETREFFEND.

Die frühen kommerziell erhältlichen Bipolartransistoren waren zumeist vertikal aufgebaut. Als Beispiel ging man z.B. aus von einem n-dotierten Siliziumplättchen. Ein ganzer Siliziumwafer, in der Größe eines Hosenknopfes, diente als Kollektor. Mit zwei Diffusionen wurden vollflächig eine Basis- und eine Emitterschicht definiert, indem bspw. das Siliziumplättchen zunächst mit einem Boratglas belegt und getempert wurde. Aus dem Glas diffundiert Bor in das Silizium. Anschließend wurde Arsentrioxid aufgesputtert und erneut getempert. Die unterschiedlichen Temperaturbeaufschlagungen und Diffusionsraten führen zur Ausbildung einer Doppelschicht, oberflächlich n-, darunter p-dotiert. Zuletzt wurde auf der Oberseite mit Brenzkatechin die n-dotierte Schicht am Rand entfernt, um die Basis, sowie die dotierten Schichten auf der Unterseite weggeschliffen, um den Kollektor anschließen zu können.

Eine bessere Kontrolle über die Eigenschaften von Basis und Kollektor erlaubt ab den sechziger Jahren die Epitaxie: Basis und Emitter werden epitaktisch aufgebracht. Silizium-Leistungsbauelemente mit Sperrspannungen von einigen 100 Volt werden auch heute noch nach diesem Verfahren hergestellt.

Im Gegensatz hierzu stehen die integrierten Technologien, bei denen eine große Anzahl von Bauelementen auf einem Siliziumplättchen nebeneinander angeordnet ist. Voraussetzung hierfür ist die Möglichkeit, den Kollektor ausreichend niederohmig von der Seite her anschließen zu können. Dies ist die wesentliche Aufgabe der vergrabenen Schichten: In das jungfräuliche Silizium werden lokal, zum Teil noch durch Diffusion, meist aber mit Hilfe maskierter Implantationen, hochdotierte Kollektor-Anschlußbahnen eingebracht und von einer sog. Kollektorepitaxie überdeckt („buried layers“). Die Kollektorepitaxie übernimmt die Rolle der aktiven Bauelementschicht, die früher der knopfgroße Wafer innehatte. Die niederohmige, vergrabene Schicht erlaubt, den Kollektoranschluß neben das Aktivgebiet zu verlegen. Der Wafer selber dient im wesentlichen als Trägermaterial und als Getter für Gitterstörungen.

Auch bei der hier untersuchten Technologiefamilie – SiGe1 und SiGe2 von Atmel – werden zunächst vergrabene Kollektoranschlußbahnen definiert und anschließend eine Kollektorepitaxie aufgebracht. Die Kollektorschicht verbleibt nur in wenigen, aktiven Gebieten. Auf dem größten Teil des Schaltkreises wird sie durch Trockenätzung entfernt und durch eine aufgebrachte Oxidschicht ersetzt. Anschließend erfolgt eine zweite Epitaxie, mit der Basis- und Emitterschicht aufgebracht werden.

Nach Abschluß der Siliziumprozessierung wird der Wafer mit einer oxidischen Passivierung versehen. Anschließend werden in der Passivierungsschicht Kontaktfenster geöffnet und Metallbahnen zum Anschluß von Basis und Emitter definiert. Kontaktfenster und Metallbahnen sind lithografisch strukturiert, mit kleinsten Strukturgrößen von 0.8µm. Dies erlaubt nicht, die metallischen Kontakte von Emitter und Basis hinreichend eng nebeneinander anzuordnen. Deswegen wird vor Abscheiden der Passivierung eine quasi metallische Silizidbahn eingebracht, die äußere Basis. Sie kontaktiert die innere Basis unmittelbar neben dem Emitter und stellt eine niederohmige Verbindung zum weiter außen gelegenen Kontaktfenster her.

Einiges an Aufwand steckt dabei in der lithografieunabhängigen, zum Emitter selbstjustierten Definition dieses Basisanschlußgebietes. Nach der Basis- und Emitterepitaxie schützt zunächst ein Nitridblock das spätere Emitterfenster. Er dient als Hartmaske für Implantation und Silizidierung. Anschließend wird Oxid aufgebracht und planarisiert, mit dem Nitridblock als Polierstop. Nach Entfernen des Nitridblocks wird, mit dem Oxid als Hartmaske, ein Emitteranschluß aus hochdotierten Polysilizium in die Hohlform eingebracht, die der Nitridblock hinterlassen hat. Die Oxidhartmaske ist, ohne Justier- oder Lithographietoleranzen, das Negativ der Nitrid-Hartmaske: Äußere Basis und Emitter sind zueinander selbstjustiert.

Atmels SiGe2-HBTs existieren, neben einer großen Vielfalt an lateralen Geometrievarianten, in 2x2 unterschiedlichen Bauformen: Zwei unterschiedlichen Kollektordotierungen und zwei verschiedenen Kollektordicken. Beide Kollektordotierungen (SiC und NonSiC; SiC steht hierbei für „selektiv implantierter Kollektor“) stehen auf dem selben Wafer zur Verfügung. Bei den beiden Kollektordicken handelt es sich jedoch um getrennte Technologievarianten, SiGe2RF und in SiGe2Power.

Allen Bauformen ist der selbe prinzipielle Querschnitt gemein, wie er in Abb. 4.1 gezeigt ist. Anhand dieses Bildes werden in Kapitel 4, „Technologiearchitektur“, zwei Herstellunsgschritte etwas feiner aufgelöst dargestellt. Sie werden im weiteren Verlauf der Arbeit wichtig für das Verständnis des Herstellungsverfahrens des vorgeschlagenen neuen Bauelementes. Während die Technologiearchitektur für die Planung des Herstellunsgablaufes verantwortlich ist, nimmt die Prozeßintegration die Aufgaben des Bauleiters wahr: Das Zusammenspielen der verschiedenen Gewerke im Kleinen sicherzustellen. Anhand von Fehleranalysen aus dem Umfeld einer Silizium-Hochfrequenztechnologie führt Kapitel 5 in die Prozeßintegration ein.

Die Schwerpunkte sind nicht von der Neuheit des Dargestellten bestimmt, sondern sollen ein Gefühl für die wechselseitigen Abhängigkeiten der Einzelprozesse innerhalb einer Technologiearchitektur, sowie der Bauelementeigenschaften von den Einzelprozessen, schaffen. Die Darstellung ist daher nur dort mit originären Beiträgen angereichert, wo mir diese dienlich erschienen, ein solches Gefühl zu befördern.

4 Technologiearchitektur

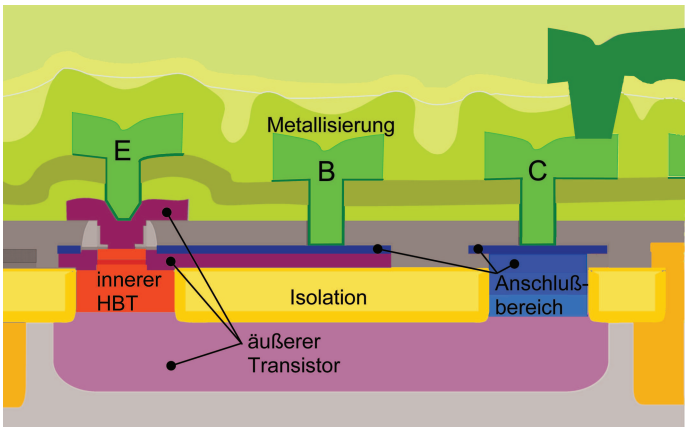


Abb. 4.1: Querschnitt durch einen SiGe-HBT

Der Herstellungsablauf einer integrierten Schaltung gliedert sich in Module, wie „Abscheiden der Kollektorepitaxie“ oder „Aufbringen eines Intermetalldielektrikums mit ebener Oberfläche auf einen unebenen Untergrund“.

In einer 0.8 µm -Technologie begrenzen nicht so sehr der innere Transistor (in der Abbildung rot) als vielmehr die Notwendigkeit, den eigentlichen Transistor anzuschließen, die elektrischen Eigenschaften der Bauelemente. Neben dem äußeren Transistor (violett) und der Metallisierung (grün) sind hier vor allem der Anschlußbereich (blau) und die Isolation (gelb) zu nennen.

Ich nähere mich der Technologiearchitektur daher über zwei Module aus der Peripherie des inneren SiGe-Transistors von Atmel[10]: Die Grabenisolation („Shallow Trench Isolation“, STI, gelb) und die zum Emitter selbstjustierte Silizidierung der äußeren Basis („Basisanschlußmodul“, blau).

4.1 Die Shallow Trench Isolation, STI

Das STI-Modul besteht aus zwei Hauptteilen: dem Eintiefen der Isolationsgräben und dem ebenen Auffüllen der Feldgebiete mit Oxid.

4.1.1 Das Eintiefen der Isolationsgräben

Nach dem Aufwachsen[11] der Kollektorepitaxie[12] wird die Kollektorschicht in einem Trockenätzprozeß größtenteils wieder entfernt. Im HBT verbleiben als aktive Gebiete lediglich der Kollektor-Anschlußbereich (blau) und der innere Kollektor (rot).

---

[10] Für einen Überblick über die bestehenden SiGe-Technologien kann ich den Leser auf [II.2] verweisen; zu Fertigungsverfahren wie zu Bauelementverhalten ist [II.3] uneingeschränkt, [II.4] zumindest in Maßen empfehlenswert.

[11] In der Halbleiterei wird „aufwachsen“ sowohl transitiv als auch intransitiv verwendet.

[12] n-Si, $2{\cdot}10^{16}$ cm^{-3} P, 450 nm bzw. 550 nm in SiGe2RF bzw. SiGe2Power.

„Trockenätzen“ bezeichnet einen Sputterprozeß, bei dem der Materialabtrag unterschiedliche Komponenten aufweist. Zunächst ist es, im Gegensatz zu nasschemischen Verfahren, ein physikalischer Prozeß: Aus einem Niederdruckplasma werden Ionen auf die zu bearbeitende Oberfläche hin beschleunigt und sputtern dort Material ab. Durch Anlegen geeigneter elektrischer Felder treffen die Ionen senkrecht auf die Oberfläche, so daß entstehende Vertiefungen fast lotrechte Wände aufweisen.

Als rein phyiskalischer Prozeß gefahren, unterscheidet sich die Sputterrate zwischen unterschiedlichen Materialien nur unwesentlich. Entsprechend schwer ist es, geeignete Masken zu finden, namentlich für die Erstellung tiefer Ätzgräben. Durch eine geschickte Wahl der Gaschemie gelingt es, eine gewisse Selektivität im Ätzabtrag mit im wesentlichen senkrechten Ätzgräben zu verbinden. So erzielt man bei der Strukturierung einer Nitridschicht[13] mit einer Lackmaske eine Selektivität von etwa 1:3, das Nitrid wird dreimal so schnell wie die Lackmaske abgetragen.

Dem Unterätzen der Lackmaske durch die chemische Abtragskomponente – dem isotropen Ätzangriff auf die senkrechten Grabenwände – wirken die Abbauprodukte der Lackmaske entgegen, welche sich als hochgradig inertes Polymer auf den senkrechten, dem physikalischen Abtrag wenig ausgesetzten, Grabenwänden ablagert.

Dem Trockenätzen geht eine Temperung der Lackmaske bei etwa 150°C für zwei Stunden, das „Lackhärten“, voraus, welches die Trockenätzrate des Lackes reduziert. Im Anschluß an das Trockenätzen muß der polymerisierte Lack im Sauerstoffplasma entfernt werden („Veraschen“) – sowohl an der Oberfläche, wo er einem kontinuierlichen Ionenbeschuß ausgesetzt war, als auch in den senkrechten Grabenwänden, wo sich während der Trockenätzung Polymer abgelagert hat. Erst nach Entfernen der Polymerschicht läßt sich der Wafer nasschemisch ablacken.

Bei Eintiefen der Isolationsgräben im STI-Modul müssen etwa 0.6µm (SiGe2RF) bzw. 0.8µm (SiGe2Power) Silizium abgetragen werden. Dies gelingt nicht mit einer Lackmaske, sondern erfordert für die Trockenätzung eine Hartmaske aus Nitrid.

Nitrid unterscheidet sich im Ausdehnungskoeffizienten stark von Silizium. Um einen nitridbedeckten Wafer thermisch oxidieren zu können müssen Silizium und Nitrid durch eine dünne Oxidschicht, ein sog. Haftoxid, getrennt sein[14]. Ohne diese Schicht platzen beim Tempern kleine Nitridblöcke ab, während große durch die thermische Verspannung das Kristallgefüge des Siliziumwafers schädigen. Das Oxid wird bereits bei moderaten Temperaturen duktil und bildet eine Gleitschicht, in der die thermischen Verspannungen abgebaut werden.

Daher werden zur Eintiefung der Isolationsgräben auf die Kollektorepitaxie nacheinander ein dünnes Haftoxid, eine Hartmaske aus Nitrid und eine Maske aus Photolack aufgebracht. Die Lackmaske wird gehärtet, belichtet und nasschemisch strukturiert. Sie dient als Maske für das Trockenätzen der Nitridschicht. Nach Veraschen, Ablacken

[13] Spricht man unspezifisch von „Oxid“, so ist SiO_2, bei „Nitrid“ Si_3N_4 gemeint.

[14] Auch biden sich bei direktem Kontakt zwischen Mitrid und Silizium nichtstöchiometrische Stickstoffverbindungen, welche sich nicht nasschemisch entfernen lassen.

und einem kurzen BHF[15]-Dip zur nasschemischen Entfernung der offenliegenden Haftoxidreste befindet sich eine strukturierte, hochbeständige Hartmaske aus Nitrid auf dem Wafer, welche erlaubt, trockenchemisch die Isolationsgräben einzutiefen. Nach der Ätzung findet sich die Kollektorepitaxie nur noch unter den nitridabgedeckten „Aktivgebieten".

Am Boden der Isolationsgräben ist das Kristallgefüge im Silizium erheblich gestört und das Material durch Einträge aus dem Niederdruckplasma verunreinigt. Ätzschäden und Verunreinigungen werden in einer sog. Liner-Oxidation entfernt: Die Siliziumoberfläche wird thermisch oxidiert, wobei die Ofenatmosphäre mit geringen Mengen HCl versetzt ist. Das Chlor bindet alkalische Verunreinigungen fest in das Oxid ein[16].

Auch der Liner-Oxidationsprozeß birgt Schwierigkeiten: Das Oxid nimmt das 2.2-fache Volumen des Silizium ein, aus dem es gebildet wird. Selbst ein dünnes thermisches Oxid erfordert eine sorgfältige Abrundung der Kante zwischen Grabenboden und Grabenwand, also eine fortgeschrittene Prozeßführung bei der Trockenätzung, damit der Oxidkeil das angrenzende Kristallgefüge nicht sprengt.

Das Silizium-Bulkmaterial des nackten Wafer, welches in den Grabenböden offenliegt, ist niedrig p-dotiert. Die Raumladungszonen zweier gleichnamiger, gegen das Substrat vorgespannter Aktivgebiete dehnen sich weit in das Silizium hinein aus. Damit es nicht zum Durchgriff kommt, wird die Dotierung der Grabenböden durch eine sog. Channel-Stopper-Implantation angehoben (die Implantation erfolgt maskiert, es gibt Bauelemente wie die lateralen pnp-Transistoren, in denen der Grabenboden eine eigene Dotierung aufweisen muß). Das Lineroxid dient als Implantationsoxid für die Channel-Stopper-Implantation[17] und wird anschließend nasschemisch entfernt.

Ihren Abschluß findet die Grabenätzung mit einer zweiten Lineroxidation zur Passivierung des Grabenbodens.

---

[15] „BHF" steht für buffered HF. Es ist eine Mischung aus NH_3, HF und Wasser. Durch die Verdünnung der Flußsäure lassen sich stark reduzierte Ätzraten von Siliziumdioxid sowie ein verschwindender Abtrag von Silizium erzielen.

[16] Namentlich das „allgegenwärtige" Natrium diffundiert andernfalls leicht durch das Oxid. Dies führt beispielsweise zu einer Verschiebung der Schwellspannung im Betrieb eines MOS-Transistors. Bei der LOCOS-Oxidation hilft die HCl-Beimischung darüber hinaus gegen den white-ribbon-Effekt. Höhere Beimischungen als 5 % führen jedoch zur Segregation von HCl-Gas innerhalb der Oxidschicht sowie zwischen Oxid und Siliziumoberfläche.

[17] Genau in <100> Richtung eintreffende Ionen dringen bei der Implantation unerwünscht tief in den Siliziumkristall ein („channeling"). Man begegnet diesem durch eine kleine Abweichung der Implantation von der senkrechten (typischerweise 7°), sowie durch Implantation durch ein etwa 20nm dünnes Implantationsoxid, welches die Ionen bereits ein wenig aus der Senkrechten ablenkt. Gleichzeitig bleiben niederenergetische Sekundärionen, welche von den Strahlblenden im Implanter abgesputtert werden, im Implantationsoxid stecken und werden im Anschluß an die Implantation zusammen mit dem Oxid entfernt.

4.1.2 Das Auffüllen der Isolationsgräben

Nun wird der STI-Graben mit TEOS verfüllt[18]. Das TEOS wird isotrop auf dem ganzen Wafer abgeschieden, bedeckt also auch die Nitrid-Hartmaske. Diese findet nun eine zweite Aufgabe, sie dient als Polierstop beim Chemisch-Mechanischen Polieren:

Beim Chemisch-Mechanische Polieren (CMP) wird der Wafer mechanisch planarisiert, indem er über ein hartes, ebenes, mit einer Polieremulsion bedecktes Polierpad bewegt wird, ähnlich dem Schleifen von Brillengläsern[19]. Die Zusammensetzung der Emulsion ist so gewählt, daß sie den Abtrag eines Teiles der auf dem Wafer befindlichen Strukturen unterstützt. So läßt sich durch eine geeignete Naßchemie beispielsweise der Abtrag von Oxid gegenüber dem von Nitrid verstärken.

Auch die konzeptuell einfache CMP hat ihre Tücken. So werden bspw. bevorzugt Kanten abgetragen. Material auf großen „Hochflächen“ oder in weiten „Tiefebenen“ wird weniger angegriffen als vereinzelte „Berge“. Nahe beeinanderliegende „Bergkämme“ werden vom Pad nicht als Einzelstrukturen aufgelöst und durchlaufen den Polierprozeß, ähnlich wie Hochebenen, weitgehend unverändert.

Die parallelen Emitterfingern großer Transistoren sind parallele Bergkämme. Es ist nötig, vor der CMP die „Bergspitzen“ aus Oxid zu öffnen. Hierzu dient ein BHF-Dip mit einer zu den Nitridblöcken inversen und geeignet gesizten[20] Lackmaske. Das über den Kanten der Nitridmaske verbleibende Oxid ist der CMP zugänglich.

Nach der CMP liegen auf dem Wafer nitridbedeckte Aktivgebiete und oxidverfüllte Gräben nebeneinander. Das offenliegende Nitrid wird in kochender Phosphorsäure, das darunterliegende Haftoxid in einem kurzen BHF-Dip entfernt. Jetzt, am Ende des STI-Moduls, sind offenliegende Aktivgebiete aus monokristallinem Silizium von oxidischen Isolationsgebieten, den Feldflächen, getrennt, während die Oberfläche aus Oxid und Silizium weitgehend eben ist.

Nach Verfüllen der Isolationsgräben und Planarisieren des Feldoxides folgt die epitaktische Abscheidung eines dünnen Schichtpaketes, aus welchem später die Basis und der Emitter geformt werden: Zunächst wenige nm n-dotiertes Silizium, die als Startschicht im wesentlichen den Boden bereitet für die folgende Schicht; eine 20 nm dicke, bordotierte SiGe-Schicht als aktive Basis; schließlich eine arsendotierte Silizium-Emitterschicht von 70 nm Dicke.

Auf den freiliegenden Siliziuminseln wachsen die Schichten epitaktisch, über den Feldgebieten polykristallin auf.

---

[18] Aus **T**etra-**E**thyl-**O**rtho**S**ilan wird Oxid in einem CVD-Prozeß abgeschieden.

[19] Das CMP-Verfahren ging aus einer Kneipenwette zwischen einem Amateurastronom und einem Ingenieur bei Intel hervor: Gegenstand der Wette war denn auch tatsächlich die Politur optischer Linsen...

[20] „Sizing“ bezeichnet ein Verfahren, bei dem auf dem Rechner bei der Maskengenerierung die Strukturgrößen relativ zu den gezeichneten Strukturgrößen verändert werden. Erfolgt, bei gegebener Maske, die Änderung der Strukturgrößen durch eine abweichende Belichtung des Lackes, spricht man von „Auf“- oder „Abbelichten“.

Auch diese Schichten verbleiben nur zu einem kleinen Teil auf dem Wafer, im wesentlichen dort, wo die Basis des Transistors entsteht: Die innere Basis des Transistors (in Abb. 4.1 dunkelrot) wird aus der monokristallinen SiGe-Schicht, der Basisanschluß aus den polykristallinen Bereichen geformt vermittels eines in der Idee sehr schönen Prozeßmoduls, dessen Tücken sich erst im weiteren Verlauf zeigen:

4.2 Emitterblock und Basisanschluß

Wie zum Schutz der Aktivgebiete bei der Definition der Isolationsgräben wird eine Nitrid-Hartmaske geformt: Die Stellen, an denen später die Emitter ausgebildet werden, werden durch nitridische Opferstrukturen bedeckt, die Emitterblöcke. Die Emitterblöcke sind lithografisch auf die Aktivgebiete zentriert und ringsum von monokristallinem Silizium umgeben.

Über die Emitterblöcke wird eine Schicht Oxid aufgebracht. Die Abscheidung in einem plasmaunterstützen CVD-Prozeß (PE-TEOS, **p**lasma **e**nhaced TEOS) erfolgt isotrop, senkrechte und waagrechte Kanten werden mit einer Oxidschicht selber Dicke bedeckt. Wird der Oxidfilm per Trockenätzung anisotrop abgetragen, verschiebt sich die Oxidoberfläche senkrecht nach unten. Am Ende der Ätzung verbleiben von der Oxidschicht lediglich kleine Flügel an allen senkrechten Kanten, die „Spacer“.

Der Nitridblock mit Oxidspacern dient als Hartmaske, wenn vermittels einer hochdosigen Borimplantation der Widerstand im Basisanschlußgebiet („äußere Basis“, violette Schicht in Abb. 4.1) verringert wird[21]. Die Implantation erfolgt weitgehend im Polykristallinen; gleichzeitig werden Teile des monokristallinen Bereiches um den Emitterblock herum umdotiert, so daß der pn-Übergang im Monokristallinen liegt.

Um den Basis-Bahnwiderstand weiter zu verringern, wird das freiliegende Silizium der Basisanschlußregion silizidiert. Die Diskussion dieses Prozeßschrittes erfolgt in Kapitel 5, „Prozeßintegration“ . Das Silizid ist in Abb. 4.1 dunkelblau dargestellt.

Das Silizid wird mit einer Oxidschicht bedeckt, und die Oxidschicht CMP-planarisiert. Der Nitridblock dient auch hier als Polierstop. Anschließend wird die nitridische Opferstruktur entfernt: Die Oxidschicht weist zur äußeren Basis selbstjustierte, senkrecht berandete „Löcher“ auf, in die der Emitteranschluß in Form einer hoch arsendotieren Schicht amorphen Siliziums (α-Si) eingebracht wird.

Die Oxidspacer auf der Außenseite der Nitridblöcke reichen nicht aus, um genügend Abstand zwischen die äußere Basis und den Emitteranschluß zu bringen. Auch ist ein kleineres Emitterfenster erforderlich, als lithographisch direkt erzeugt werden kann. Vor Abscheiden des Emitter-α-Si werden daher an den Innenkanten der Emitter-Hohlform Innenspacer eingebracht.

Dies erfordert ein anderes Vorgehen als bei den Außenspacern: Die Hohlform darf nicht zu weit verfüllt werden, um den Emitterwiderstand nicht zu erhöhen und das Einbringen des α-Si nicht zu behindern. Das Mittel der Wahl sind Nitrid-L-Spacer:

[21] Während der Implantation sind Teile anderer Bauelemente durch eine zusätzliche Lackmaske abgedeckt; für den HBT ist dies ohne Belang.

In den Emitteröffnungen befindet sich noch das ursprüngliche Haftoxid, welches den Emitterblock vom darunterliegenden Silizium getrennt hat. Nach Aufbringen von 45 nm Nitrid werden, durch Abscheiden und Zurückätzen von 200 nm amorphen Siliziums (α-Si), zunächst α-Si-Innenspacer in der Emitteröffnung ausgeformt. Diese dienen als selbstjustierte Hartmaske zur Strukturierung der Nitridschicht und werden anschließend entfernt. Es verbleiben L-förmige Innenspacer aus Nitrid in der Emitter-Hohlform. Nach Abscheiden des Emitter-α-Si in die Hohlform entsteht so ein zur äußeren Basis selbstjustierten Emitteranschluß (oberer violetter Bereich in Abb. 4.1).

5 Prozeßintegration

Die Technologiearchitektur entsteht im Gespräch und am Schreitbisch. Probleme, die bei der **Umsetzung** der Technologiearchitektur auftreten, lassen sich nicht am Schreibtisch lösen. Sie sind Gegenstand der **Prozeßintegration**.

Die Untersuchung der prozessierten Wafer erfolgt ex posteriori. Auch während meiner Zeit bei Prof. Schumacher habe ich derartige Analysen für Atmel durchgeführt. Auf drei meiner Analysen greife ich hier zurück, um die Aufgabe der Prozeßintergation zu illustrieren[22].

5.1 Oxidation und Verspannung

Bei der Darstellung des STI-Moduls habe ich eine Frage ignoriert, welche sich eigentlich aufdrängt: Gehärtete Lackmaske zur Strukturierung der Nitridmaske, Nitridmaske zum Ätzen der Grabenisolation, Haftoxid zwischen Nitrid und Silizium – wer denkt sich eigentlich so etwas aus? Selbst die schlichte Hartmaske für die Trockenätzung erscheint viel zu komplex in Konzeption und Umsetzung, als daß ein klar denkender Ingenieur auf eine solche Idee verfallen könnte.

Die Frage vernachlässigt die Wurzeln des Verfahrens. Sie liegen in der „lokalen Oxidation von Silizium“, LOCOS, und sind jedem Halbleiteringenieur intim vertraut. Neben der Verwendung von Spacern war in der Vergangenheit die lokale Oxidation das am weitesten verbreitete Prozeßmodul in der Siliziumtechnik überhaupt. Das elegante und robuste Verfahren wurde zwischen 1967 und 1970 von Herrn Dr. Else Kooi am Philips Natuurkundig Laboratorium in Eindhoven entwickelt.

Zwischen 1970 und 2000 ist es das vorherrschende Isolationsverfahren im CMOS-Bereich: Die flach implantierten CMOS-Source- bzw. -Drain-Gebiete in einer gegennamig dotierten Wanne lassen sich gegeneinander isolieren, indem die Source-Drain-Implantation durch ein Dickoxid abgeschirmt wird. In den letzten Jahren wird das LOCOS-Verfahren durch eine STI-Grabenisolation abgelöst, wie sie auch bei SiGe2 Einsatz findet. Was sich jedoch am LOCOS-Verfahren lernen läßt, findet unmittelbare Anwendung bei der STI, und zwar sowohl bei der Bildung der Hartmaske, als auch bei der Lineroxidation zur Passivierung der Ätzflächen.

[22] Das dritte Standbein der Halbleiterentwicklung, die Einzelprozeßentwicklung, klammere ich aus; sie liegt außerhalb meines Erfahrungsbereichs.

Bei der Oxidation von Silizium diffundiert das Oxidationsmittel durch eine bereits gebildete Oxidschicht, um an der darunterliegenden Grenzfläche mit dem Silizum zu reagieren. Die Idee hinter dem LOCOS-Verfahren ist, eine strukturierte Nitridmaske zu verwenden, um bei einem Ofenprozess die Oxidbildung auf die freiliegende Siliziumoberfläche zu beschränken. Das Oxidationsmittel diffundiert hierbei auch ein Stück weit unter das Nitrid. Es bildet sich ein charakteristischer Oxid-Querschnitt aus, der „LOCOS-Vogelschnabel“ (bird's beak, Abb. 5.1, 5.2).

Die Nitridkanten werden durch das wachsende Oxid hochgebogen. Das verformte Nitrid übt große Kräfte auf darunterliegendes Silizium aus. Werden die Spannungen zu groß, lösen sie sich in Versetzungslinienscharen: Das Kristallgitter wird zerstört.

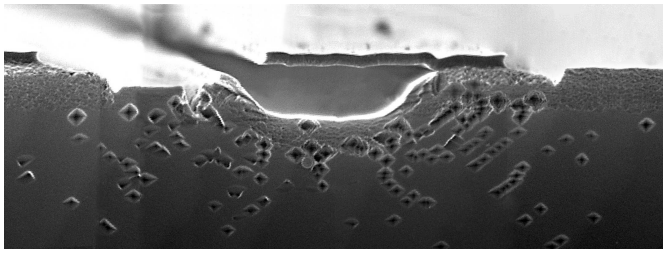


Abb. 5.1: Bruchstelle mit LOCOS-Vogelschnabel

Während Flußsäure Silizium kaum angreift, führt die Beimischung von Schwermetallionen zum Ätzangriff auf **Gitterstörungen**, bzw. auf das eine Gitterstörung umgebende verspannte Silizium. Bei der Sirtl-Beize, bspw., wird zur Fluß- ein wenig Chromsäure zugesetzt ($HF : H_2O : CrO_3 \approx 1 : 0{,}4 : 0.2$) und die zu untersuchende Probe 10 Sekunden getaucht. Abb. 5.1 zeigt einen gesirtlten Bruch mit oxidbedingten Versetzungslinien. In Bildmitte ist die nitridbedeckte bird's-beak-Form zu erkennen (das Oxid ist durch die Präparation entfernt). Die rautenförmigen Ätzgruben sind Austrittspunkte von Versetzungslinien in die Bruchfläche – die Prozessführung während der Oxidation in einem Hochdruckreaktor war offensichtlich ungeeignet!

Dr. Kooi hat die Verspannungen verringert, indem er zwischen Nitrid und Silizium eine „Gleitschicht“ aus Oxid dazwischengeschaltet hat. Die Dicke des Haftoxides bestimmt dabei die Form des Vogelschnabels: Um ein platzsparendes Oxidmodul mit steilen Oxidkanten zu erhalten, braucht man ein dünnes Haftoxid. Je dünner das Haftoxid, desto höher werden die Verspannungen. Das Verfahren bedarf daher einer sorgfältigen Abstimmung zwischen der Dicke des Haftoxides, dem Temperaturverlauf und dem Sauerstoffpartialdruck während der Oxidation.

Mit Hilfe der Sirtl-Beize lassen sich Versetzungen, also Störungen im Kristallgitter, einfach dekorieren. Zur Optimierung eines Oxidationsmoduls ist es extrem hilfreich, auch Verspannungen quantifizieren zu können, welche noch nicht zu einer Zerstörung des Kristallgitters geführt haben. Hierfür steht seit einigen Jahren die prohibitiv aufwendige Micro-Raman-Streuung zur Verfügung.

Ich habe mir ein deutlich einfacheres Verfahren ausgedacht. Es ist zwar nicht im selben Maße quantitativ, aber dennoch aussagekräftig genug. Ich nutze aus, daß Sirtl

nicht wirklich Versetzungen, sondern die Spannungsfelder um eine Versetzung herum angreift.

Ich belege einen Wafer nach der Feldoxidation mit amorphem Silizium (α-Si) und breche ihn. Bei einer anschließenden Sirtl-Beize wird das maximal gestörte α-Si sehr schnell gelöst. An der Grenzfläche zwischen Oxid und Silizium bildet sich aber eine dünne Schicht aus nicht stöchiometrischen Oxiden. Diese Schicht führt zu einigen Sekunden Ätzverzug, so daß bei vorsichtiger Präparation das α-Si entfernt wird, das Oxid jedoch erhalten bleibt.

Zur Illustration, wie sich mit dem Verfahren Verspannungen dekorieren lassen, entferne ich in einem Teil der Probe die Verspannungen künstlich: Ich tauche die Bruchstelle vor dem Sirtln in BHF. Von der Bruchkante aus wird das Feldoxid etwa 1 µm tief herausgelöst:

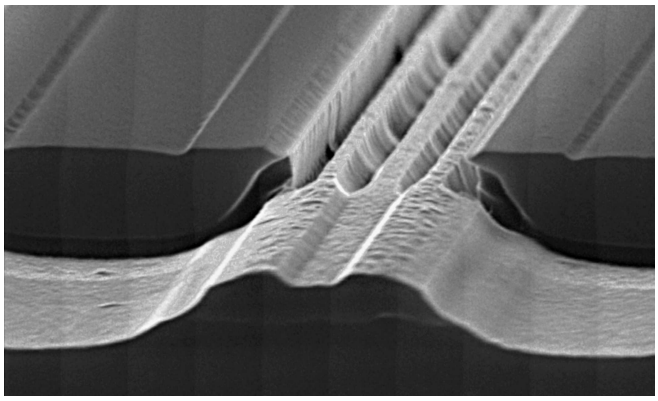


Abb. 5.2: REM-Aufnahme nach Spannungsätze

In der REM-Aufnahme Abb. 5.2 sieht man, rechts und links im Hintergrund, CMP-planarisierte LOCOS-Strukturen. Im Vordergrund ist das Oxid durch den BHF-Dip vor der Sirtl-Beize entfernt worden.

Wo das Oxid erhalten ist, sind, im Anschluß an die Vogelschnäbel, die verspannten Bereiche in Form von Ätzgräben im Silizium herauspräpariert worden. Mit dem Oxid verschwinden jedoch auch die Verspannungen: Wo ich das Oxid entfernt habe, findet kein Angriff auf das Silizium statt.

Bei Prozeßvariationen ändert sich der Grad der Verspannung und damit die Ausdehung des Ätzgrabens: Es steht daher eine einfache und billige Meßlatte für die Verspannungen zur Verfügung!

Derselben Problematik sehen wir uns auch bei der Liner-Oxidation bei SiGe2 gegenüber: Liegt zwischen Grabenboden und Grabenwand ein rechter Winkel, führt bereits ein dünner „Oxidkeil“, der sich in den Winkel schiebt, zu hohen Verspannungen. Es bedarf einer sorgsam mit dem Oxidationsrezept abgestimmten Verrundung des Grabenbodens bei der Trockenätzung, um die Verspannungen zu beherrschen.

Zur Illustration geben ich ein Beispiel simulierter Verspannungen wieder, das einem SILVACO-Prospekt [II.5] entnommen ist (Abb. 5.3):

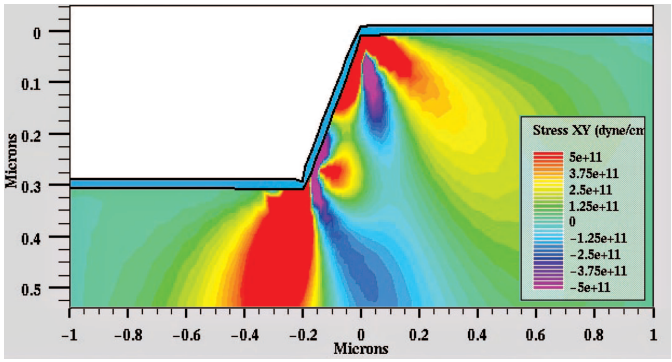


Abb. 5.3: Simulation oxidativer Zug- und Druckverspannungen nach einer Lineroxidation eines offenen, scharfwinklig berandeten Grabens im Silizium. Rote und violette Töne stehen für maximale Druck- bzw. Zugspannung

5.2 Metallsystem

Als nächstes greife ich das Metall-Wiederholmaß in SiGe2 auf.

Die Atmel'sche SiGe-Technologie stellt zwei Metallagen zur Verfügung. Die erste Metallage ist in Abb. 5.4 grün, die Intermetalldielektrika olivgrün bis gelb dargestellt. Von besonderem Interesse ist hier die erste Oxidschicht oberhalb der Bahnen.

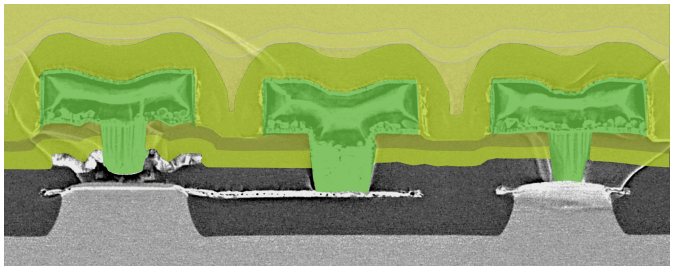


Abb. 5.4: Metallisierung und IMD

Den rasterelektronenmikroskopischen Querschnitt durch eine SiGe2-Transistor hat mir dankenswerter Weise Herr Holger Höhnemann von Atmel zur Verfügung gestellt.

Im Gegensatz zu meinen eigenen Aufnahmen an einem Gerät aus den frühen 70er Jahren zeichnet sich nicht nur die Kristallstruktur im polykristallinen äußeren Emitter und in der Metallisierung sehr klar ab. Auch ist die SiGe-Schicht bereits ohne jede Probenpräparation deutlich sichtbar.

Bei einem Blick auf den Transistor-Querschnitt aus Abb. 5.4 erscheint der Abstand zwischen den Metallkontakten unsinnig hoch: Die Kontaktstruktur ist ein vielfaches so groß wie der Transistor selber. Tatsächlich ist dieser Abstand erforderlich, wie sich anhand von Fertigungsproblemen in UHF5s – einer bipolaren Silizium-Hochfrequenztechnologie von Atmel – veranschaulichen läßt. In UHF5s sind spontan blumenkohlartige Ausblühungen oberhalb des Intermetalldielektrikums entstanden. Im weiteren Herstellungsverlauf bewirken derartige „hillocks" Abrisse in darüberliegenden Metallbahnen.

Nach einem Bruch mittig durch ein Blumenkohl-Röschen (Abb. 5.5) zeigt sich im REM, daß unterhalb des Röschens die Lücke zwischen zwei Leiterbahnen bei Abscheidung des Intermetall-Dieletrikums nicht aufgefüllt, sondern lediglich abgedeckt worden ist. Im verbleibenden Hohlraum – einem sog. „void“ – verbleiben Bearbeitungsrückstände, im vorliegenden Fall Wasser von der Reinigung der Wafer zwischen zwei Prozeßschritten. Das ausdampfende Wasser beeinflußt lokal die Abscheidung des Intermetalloxids im CVD-Reaktor.

Die Vermeidung von Voids erzwingt einen Mindestabstand der Metallbahnen, damit sich die Zwischenräume lückenlos ausfüllen lassen. Kontaktlochmaße und Justiertoleranzen wiederum erfordern eine Mindestbreite der Metallbahnen. Bei SiGe2 ergibt dies ein Metall-Pitch (=Wiederholmaß), das sehr viel größer ist als die Breite des inneren Transistors.

Insbesondere läßt sich der metallische Basiskontakt nicht nahe an die Emitterkante rücken. Den Abstand alleine mit Polysilizium (eigentlich eine polykristalline Dreifachschicht Si/SiGe/Si) zu überbrücken, verbietet sich, wie nachfolgend gezeigt wird, aufgrund des Schichtwiderstandes.

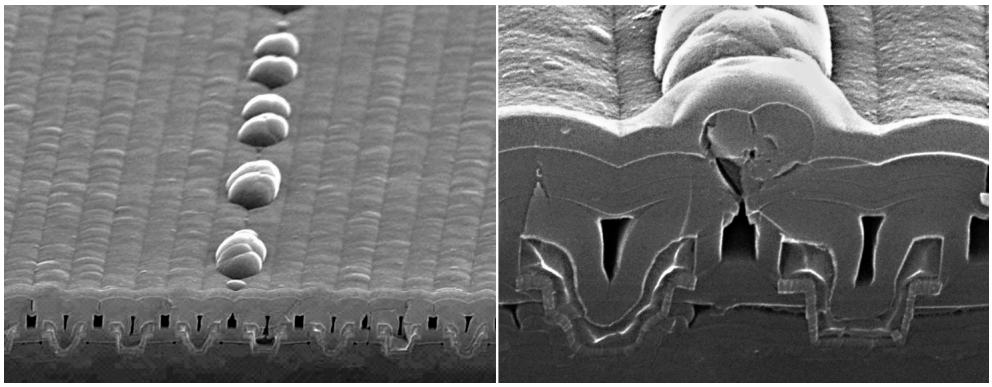


Abb. 5.5: Bruchbilder durch Ausblühungen, links: Übersicht, rechts: Detail.

5.3 Silizidierung, Basisanschluß und Randdiode

Wird ein titanbedeckter Siliziumwafer erhitzt, beginnen Silizium und Titan bei 400°C zu TiSi, und bei 500°C zu $TiSi_2$ zu reagieren[23]. Man muß das Disilizid auf 650°C erhitzen, damit es sich von einer metastabilen Modifikation („C49-Phase“, spez. Widerstand: 60 bis 90 µΩ/cm) dauerhaft in die stabile und niederresistive Modifikation umgewandelt („C54-Phase“, 13 bis 16 µΩ/cm).

Titan und SiO_2 reagieren oberhalb von 700°C zu niederstöchiometrischem Rutil, einem schlechten Isolator, der sich nicht selektiv entfernen läßt (bereits die Titansabscheidung muß sauerstofffrei erfolgen). Genauso bildet sich durch die Reaktion von Titan mit Silizium-Nitrid schwer lösliches Titannitrid. Die Umwandlung von Titan und Silizium zu C54-$TiSi_2$ wird daher nach der Bildung von C49-$TiSi_2$ unterbrochen und die auf Oxid und Nitrid verbleibenden Titanreste selektiv entfernt, bevor, in einem zweiten RTA-Schritt bei 850°C, die C54-Phase gebildet werden kann.

Geeignete Lösungsmittel für das verbleibende Titan sind 1 NH_4OH:1 H_2O_2:5 H_2O, die „Titan-Ätze“ 20 H_2O:1 H_2O_2:1 HF oder 2 H_2SO_4:1 H_2O_2 (80°C, 5 min).

Um den Schichtwiderstand des Basisanschlußgebietes zu verringern, wird die obere Siliziumlage in Titansilizid, $TiSi_2$, umgewandelt. Im REM-Bruchbild (Abb. 5.6) ist die verbleibende Si/SiGe Doppelschicht violett und das Titansilizid blau hervorgehoben:

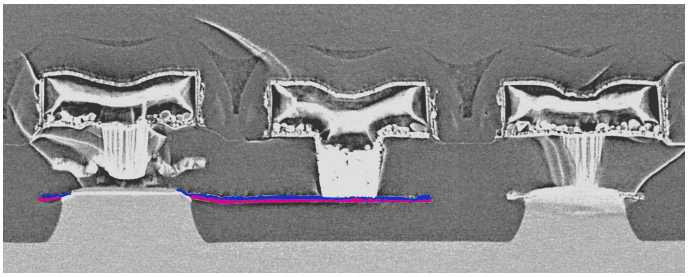


Abb. 5.6: Silizidierter äußerer Basisanschluß

Mithilfe der analytischen Lösung des Hybrid-Pi-Modells (S. 6) läßt sich die Bedeutung der Silizidierung leicht quantifizieren. Ein Transistor mit 5.6 µm Emitterlänge weist eine Breite der äußeren Basis von ca. 8 µm auf. Eine 70 nm dicke Silizidschicht mit einem spezifischen Widerstand von 15 µΩcm und einer Breite von 8 µm ergibt einen Bahnwiderstand von 26.8 Ω – gleich dem äußeren Basis-Bahnwiderstand R_X des Modells, s. S. 9. Bei einer aktivierten Dotierstoffkonzentration von $1.65{\cdot}10^{20}$ cm^{-3} ergibt sich in monokristallinem Silizium ein spezifischer Widerstand von 0.5 mΩcm. Mit diesem Wert als untere Grenze für Polysilizium stellt eine 70 nm dicke Poly-Schicht einen Vorwiderstand von mindestens 900 Ω dar.

Der Stromverstärkung und dem Leistungsgewinn beim silizidierten Transistor werden in Abb. 5.7 die Werte für einen hypothetischen Transistor ohne Silizidierung gegenübergestellt. Durch einen Verzicht auf die Silizidierung fällt die maximale Schwingfrequenz um eine Größenordnung ab – die Bedeutung der Silizidierung ist offensichtlich!

[23] Eine hervorragende Einführung in die Bildung von Titansilizid findet sich z. B. in [II.6].

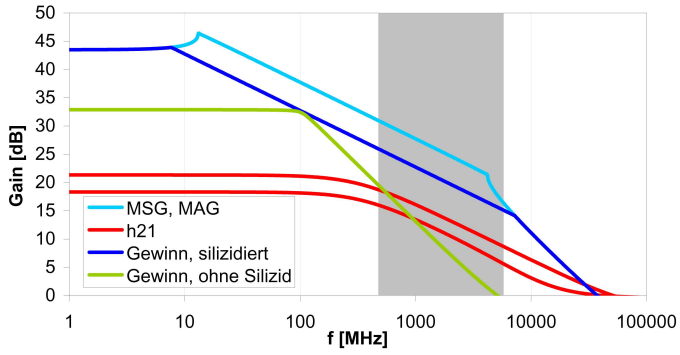


Abb. 5.7: Basis silizidiert vs. Polysilizium; „Gewinn" ist der erzielbare Leistungsgewinn
Die roten Kurven zeigen den Einfluß der Silizidierung auf h_{21}.

Wie mit Silizium reagiert Titan auch mit Germanium zur homologen Verbindung C54:$TiGe_2$. Die thermische Stabilität des Titandigermanids ist jedoch gering: Während zunächst ein $Ti(Si_{1-x}Ge_x)_2$-Germanosilizid mit dem Silizium-Germanium-Verhältnis des Ausgangshalbleiters gebildet wird, fällt bei weiterer Temperaturbeaufschlagung $Si_{1-y}Ge_y$ aus, während der Germanium-Gehalt des Mischsilizids abnimmt. Bei Temperaturen, wie sie für die Umwandlung des $TiSi_2$ in die niederresistive Modifikation erforderlich sind, zerfällt das Germanosilizid in eine zweiphasige Schicht aus Germanosilizidkörnern, getrennt von SiGe-Prezipitaten [II.7].

Untersuchungen von Dr. Andreas Schüppen und Harry Dietrich, Atmel, zur Silizidierung der äußeren Basis des SiGe-HBT haben ergeben, daß sich der Übergangswiderstand vom Titansilizid in die SiGe-Schicht durch eine dergestalt gestörte zweiphasige Zwischenschicht stark erhöht (mündliche Mitteilung A. Schüppen): Die Silizidierungsfront darf nicht bis auf die SiGe-Schicht der Basis hinunterreichen.

Hierauf beruht m.E.n. eine wesentliche Schwäche der SiGe2 Technologiearchitektur.

Peinlicherweise ist mir diese Schwäche erst anhand der simulierten Dotierstoffverteilung bewußt geworden. Die Dotierstoffverteilung im Silizium habe ich durch Simulation des Herstellungsablaufs erhalten. Sie erscheint verläßlich. Die Diffusion von Bor aus der verspannten und schlecht charakterisierten Kristallstruktur der SiGe-Basis läßt sich kaum simulieren. Das Bor- und das Germaniumprofil in der Basis habe ich daher aus SIMS-Daten abgeschätzt und nachträglich in den Querschnitt eingefügt.

In Abbildung 5.8 überlagere ich REM-Querschnitt und simulierte Dotierstoffverteilung. pn-Übergänge sind lila dargestellt.

Konzentrieren wir die Betrachtung auf den monokristallinen inneren Emitter, das, aufgrund der farbkodierten Dotierstoffverteilung, grüne Rechteck unterhalb des rot dargestellten Polysiliziumemitters. Unterhalb des Emitters liegt, aufgrund der höheren Dotierstoffkonzentration gelb dargestellt, die dünne Basisschicht, vom Emittergebiet getrennt durch einen pn-Übergang.

Zunächst wird der Emitter durch seine hohe Dicke selbst ohne Vorwärtsspanung der Basisdiode nur zu einem kleinen Teil ausgeräumt. Der neutrale Emitter mit einer

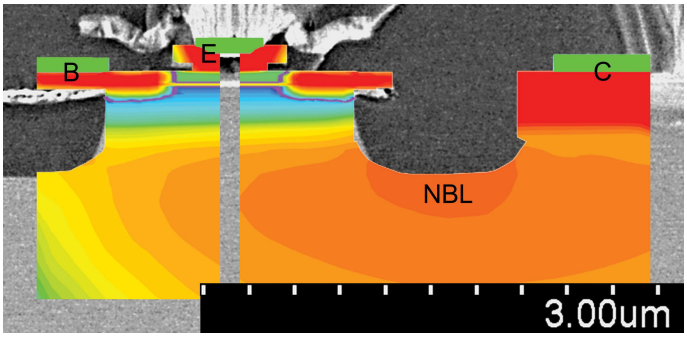


Abb. 5.8: Querschnitt durch den Transistor mit simulierter Dotierstoffverteilung (farbkodiert)

Dotierung von $2{\cdot}10^{18}\,\mathrm{cm}^{-3}$ stellt 40% des Emitterwiderstandes (vgl. Modelldaten auf S. 6) und verringert die Leistungsverstärkung um 2 dB!

Der Übergang zu einem Mono-Silizid wie NiSi – man muß im weiteren Prozeßablauf auf die geringere Temperaturbeständigkeit des Nickelsilizids Rücksicht nehmen – reduziert die verbrauchte Menge an Silizium bei einer gegebenen Silizid-Schichtdicke. Durch den höheren spezifischen Widerstand würden bei gleichbleibendem Schichtwiderstand jedoch nur 5 nm Siliziumschichtdicke eingespart.

Die Grenzfläche zwischen dem n-Gebiet des Emitters und dem p-Gebiet der Basis besteht zum zweiten und wesentlichen nicht nur aus der waagrechten SiGe-Diode zwischen Emitter und Basis. Parallelgeschaltet ist vielmehr eine – senkrecht orientierte – Siliziumdiode zwischen Mono-Emitter und äußerer Basis!

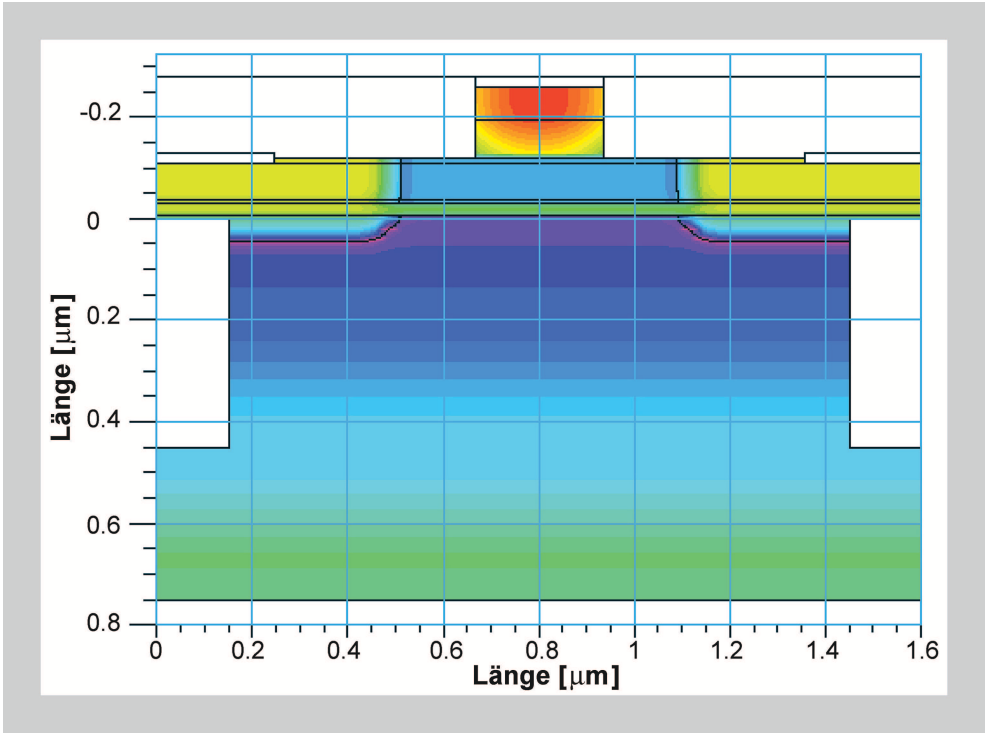
Die in Flußrichtung gepolte Siliziumdiode trägt nicht zur Transistorwirkung bei. Im Gegenteil, sie erhöht den Basis- nicht aber den Kollektorstrom, und reduziert so die Verstärkung. Darüberhinaus erhöht sie die Emitterkapazität nicht unerheblich.

Skalieren wir im Geiste die lateralen Abmessungen des SiGe-Transistors. Da die Silizidierungsfront nicht bis auf die SiGe-Schicht hinabreichen darf, ändert sich nichts an der Siliziumdicke, die für die Silizidierung vorzuhalten ist. Während die aktive Emitterdiode kleiner wird, bleibt die parasitäre Randdiode unverändert!

Transferiert man den Prozeß auf eine Fertigungsstraße, die Basisdimensionen von 350 nm mal 800 nm zuläßt. Hier macht die Randdiode fast 60 % der Fläche der aktiven Diode aus. Die zusätzliche Emitterkapazität reduziert das ohne Randdiode erreichbare f_{max} um 30 %!

Silizium zur in-situ Silizidierung vorzuhalten, macht eine laterale Skalierung zur Verbesserung der Hochfrequenzeigenschaften hinfällig. Ohne neues Basisanschlußmodul ist Atmels SiGe-Techologie in der zweiten Technologiegeneration ausgereizt!

III Bauelementsimulation



So kam auch ich aus eitler Ruhmessucht auf den Gedanken, der Nachwelt ein Weniges zu hinterlassen, um nicht als Einziger von der Freiheit des Fabulierens ausgeschlossen zu bleiben; allein, da ich aber nichts Wesentliches zu erzählen hatte – ich hatte ja nichts Erwähnenswertes geleistet – verlegte ich mich auf die Erfindung, was in meinem Falle viel verzeihlicher ist als bei den anderen: Ich werde nähmlich in dem einen Punkt die Wahrheit sprechen, wenn ich sage, dies alles seien – nichts als Hirngespinste!

Lukian, WAHRE GESCHICHTEN [III.1]

Neben Kopf und Experiment ist unser wichtigstes Werkzeug die Simulation. Weil aber die Wahrheit ein wertvolles Gut ist, und als solches sparsam dosiert, ist die physikalische Simulation mit Vorsicht zu genießen. Die Karte ist nicht das Gelände.

Nicht von ungefähr habe ich der Simulation die Behandlung einer einfachen Ersatzschaltung als Modellbild, die Technologiearchitektur als Rahmen und das Experiment als dringend nötige Richtschnur vorangestellt. Ohne die experimentelle Absicherung in jedem kleinen Schritt, ohne beständige Plausibilitätskontrolle anhand von Intuition und Erfahrung führt die Simulation unbedingt in die Irre, und wir können, wie Lukian, unsere Resultate in erster Näherung sämtlich als Hirngespinst bezeichnen.

Die Prozeßsimulation modelliert Vorgänge wie Implantation und Diffusion von Dotierstoffen. Mit dem Herstellunsgablauf als Eingabegröße liefert sie realistische Bauelementquerschnitte, wie die Dotierstoffverteilung aus Abb. 5.8. Die vorliegende Arbeit läuft auf ein vorgeschlagenes neues SiGe-Bauelement zu. Die Prozeßsimulation spielt keine große Rolle, und ich verbanne sie nolens volens in den Anhang, Kap. 18.

Die Bauelementsimulation berechnet, ausgehend von gemessenen, simulierten oder erträumten Bauelementquerschnitten die elektrischen Eigenschaften.

Wie bei der Schaltungssimulation liegt die Gefahr darin, sich von den leistungsfähigen Werkzeugen blenden zu lassen. Die Simulation ist keine Methode zur Erzeugung von Zahlenwerten, sondern die Aufgabe, die Zahlenwerte zu interpretieren.

Der Unterschied zwischen Zahlenwert und Simulationsergebnis läßt sich besonders einfach an einem leicht defizitären Modell herausarbeiten. Zu diesem Zweck greife ich noch einmal auf das Hybrid-Pi-Modell zurück.

Um ein gut konstruiertes Bauelement in einer elektrischen Eigenschaft zu verbessern, müssen für gewöhnlich an anderer Stelle Abstriche gemacht werden. Ein klassisches derartiges Eigenschaftspaar sind die Hochfrequenzeignung und die Sperrfähigkeit.

Gemäß dem Hybrid-Pi-Modell (s. S. 6) wird ein Transistor schneller, wenn wir die Kollektorkapazität C_π reduzieren. Die Verringerung des Kapazitätsbelages der Basis-Kollektor-Raumladungszone erfordert eine Verringerung der Kollektordotierung. Da mit sinkender Dotierung die Durchbruchspannung ansteigt, verbessert man gleichzeitig Hochfrequenzeignung und Sperrfähigkeit. Wir verweisen das Simulationsergebnis ins Reich der Lukianischen Hirngespinste und nähern uns nachfolgend der Bauelmentsimulation von einer stärker physikalisch geprägten Warte aus.

6 Bauelementsimulation, I: Der Kollektor des HBT

6.1 Durchbruchmechanismen

Ausgesprochen wichtig für den Betrieb eines Transistors ist seine Spannungsfestigkeit. Modell und Experiment vergleichend, verwende ich die Spannungsfestigkeit, um die Vorstellungen von den Vorgängen im Kollektor zu überprüfen.

Je nach Betriebsart sind beim Bipolartransistor zwei Sperrspannungen von Belang: Bei vorgegebenem Basispotential bricht der Transistor, wenn der Multiplikationsfaktor M im Feld der Kollektor-Raumladungszone den Wert eins erreicht. Die Sperrfähigkeit ist gleich der Dioden-Druchbruchspannung zwischen Basis und Kollektor (Basis-Kollektor-Durchbruchspannung, BV_{CB0}). Bei eingeprägtem Basisstrom – und damit floatendem Basispotential – beträgt sie typischerweise die Hälfte bis ein Drittel dieses Wertes (Basis-Emitter-Durchbruchspannung, BV_{CE0}). Warum?

Prägt man in die Basis eines npn-Transistors einen (Löcher-) Strom ein, wird dieser durch einen um die Stromverstärkung β höheren (Elektronen-) Strom durch die Kollektor-Raumladungszone beantwortet. Dieser Strom unterliegt im Feld der Kollektor-Raumladungszone der Ladungsträgermultiplikation. Die bei der Multiplikation erzeugten Löcher werden durch das Feld in der Raumladungszone in die Basis abgesogen. Sie stehen, zusätzlich zu dem von außen eingeprägten Strom, zur Verstärkung zur Verfügung – das heißt, sie verschieben kapazitiv das Basispotential, bevor sie in den Emitter abfließen können – und erhöhen wiederum den Kollektorstrom um einen Beitrag $M\beta$. Der wirksame Multiplikationsfaktor erreicht den Wert 1 bereits, wenn der Multiplikationsfaktor in der Raumladungszone erst $1/\beta$ beträgt: Die Durchbruchspannung bei vorgegebenem Basisstrom liegt bei der Hälfte bis einem Drittel der Diodendurchbruchspannung.

Nachfolgend entwerfe ich, zur Einführung in die Simulation, ein einfaches Modell für den Durchbruch der Kollektordiode. Der wesentliche Punkt, auf den ich abziele, ist, neben dem Modell selbst, das die Basis liefert für das Verständis der Vorgänge im Kollektor, die Bedeutung des Experiments und seiner Interpretation, um die erforderlichen Eingabegrößen der Simulation herauszuarbeiten. Als Beispiel für das Experiment und seine Deutung bietet sich die Auslegung von SIMS-Daten an, um zu Tiefenprofilen der Dotierstoff-Verteilung zu gelangen.

6.2 Kollektorprofile

Bei der Sekundärionen-Massenspektroskopie (SIMS) werden Testfelder von typischerweise 600µm mal 600µm untersucht, welche im Tiefenprofil wichtigen Bauelement-Bereichen nachempfunden sind: Mit einem (Sauerstoff- oder Cäsium-) Primärionenstrahl wird Material von der Oberfläche eines SIMS-Feldes abgetragen. Die Sekundärionen werden in einem Quadrupol-Massenspektrometer nach Ionensorten getrennt und schließlich gezählt. Es läßt sich ein ausgezeichnetes Tiefenprofil der Dotierstoffkonzentration gewinnen. Um zu unverfälschten Meßergebnissen zu gelangen, muß man die Messung um einige Eigenheiten der SIMS-Apparatur korrigieren:

– Die SIMS-Felder müssen von (nativem) Oxid befreit und dünn mit amorphem Silizium bedeckt sein, damit Anlaufeffekte nicht die eigentliche Messung verfälschen.

– Eine Stufe in der Dotierstoffkonzentration wird im SIMS-Profil mit endlicher Steilheit wiedergegeben. Je nach Energie der Primärionen beträgt der scheinbare Dotierstoffgradient zwischen 3 nm und 7 nm pro Konzentrationsdekade: Die Eindringtiefe der Primärionen sowie die Sekundärionenkaskade führen zu einer „Durchmischung“ der oberflächennahen Probenschichten.

– Ein Teil des Probenmaterials wird durch den Ionenbeschuß auch weiter in die Tiefe geschossen („knock-in“). Bei einer hohen Konzentration der nachzuweisenden Substanz erscheint dadurch die abnehmende Flanke zusätzlich verschmiert.

– Es muß ein abgeschätzter apparativer Untergrund abgezogen werden.

(Weitere Eigenheiten der SIMS-Messung lernen wir bei näherer Betrachtung des Basisprofiles in Kap. 7 kennen.)

Ein typisches Beispiel für eine korrigierte SIMS-Messung – den Querschnitt durch einen Transistor mit selektiv implantiertem Kollektor – zeigt Abb. 6.1.

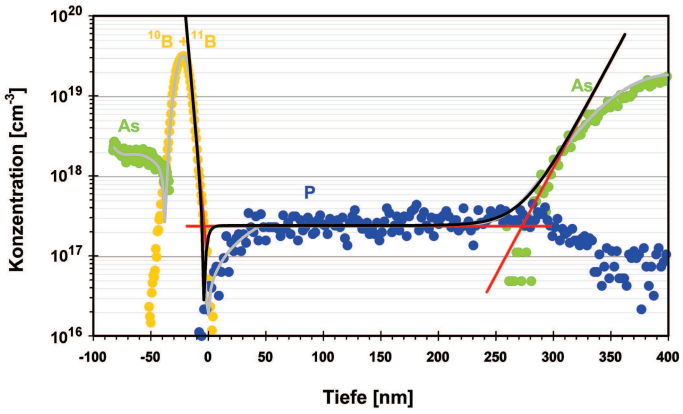


Abb. 6.1: SIMS-Profil

Die Oberfläche der Probe liegt links. Es schließt sich der arsendotierte Emitter an. Der Anstieg des Arsen-Profils zur Oberfläche der Probe hin ist kein Anlaufeffekt: Bei der Gasphasenepitaxie schwimmt Arsen während der Epitaxie auf.

Die kollektorseitigen Dotierstoffprofile lassen sich mit den rot eingezeichneten Kurven fitten, aus denen sich die schwarze Nettodotierung ergibt:

$$P_0 \exp\left(-\frac{x+3.7}{2.7}\right) \text{ (Basis-Flanke)}$$

$$N_0 = 2.384 \cdot 10^{17} \text{ (Phosphor-Plateau)}$$

$$N_0 \exp\left(\frac{x-273}{16.2}\right) \text{ (Buried-Layer-Flanke)}$$

6.3 Stromloser Kollektor

Das SIMS-Profil läßt sich ausgezeichnet heranziehen, um eine Vorstellung von den Vorgängen in der stromfreien Raumladungszone zu gewinnen: Die Feldstärke berechnet sich als erstes Integral, die Spannung als zweites Integral über die Raumladungsdichte, mit der Randbedingung, daß auf n- wie auf p-Gebiet das selbe Potential abgebaut wird. Als Anschlußbedingung ist ferner die Normalkomponente der dielektrischen Verschiebung an der Grenzfläche stetig.

Für die Dielektrizitätskonstante setze ich ein Stufenprofil an, wobei ich linear zwischen den Werten von Silizium und Germanium interpoliere und für den p-dotierten Basisbereich einen homogenen Germaniumgehalt von 19 % annehme (de facto hat es keinen Einfluß, wenn wir überall den Silizium-Wert verwenden). Es bietet sich ferner an, die Dotierstoffprofile $N(x)$ und $P(x)$ in Einheiten der Elementarladung pro Kubik-Nanometer anzusetzen. $P(x)$ wähle ich positiv, $N(x)$ negativ.

Die maximale Feldstärke in Abhängigkeit von der angelegten Spannung berechnet sich, dank der integrablen Fitformeln an die SIMS-Kurven, ohne großen Aufwand mit MAPLE (`q` steht für den Betrag der Elementarladung, `e_b` und `e_c` für die Dielektrizitätskonstanten in Basis und Kollektor):

```
Yn:=unapply(solve(int(P(x),x=-yp..0)=int(N(x),x=yn..0),yn),yp);
U_n:=y_n->q/e_c*int(int(N(x),x=y_n..y),y=0..y_n);
U_p:=y_p->q/e_b*int(int(P(x),x=-y_p..y),y=0..y_p);
Yp:=U0->fsolve(U_n(Yn(k))+U_p(k)=U0,k=0..100);
E_n0:=U0->1e7*q/e_c*int(N(y),y=Yn(Yp(U0))..0);
```

Auf n- wie auf p-Gebiet wird das selbe Potential abgebaut. Hieraus wird, in Abhängigkeit von der Weite der Raumladungszone `y_p` im Basisgebiet, die Ausdehnung `Yn` der Raumladungszone in den Kollektor bestimmt.

`U_n` und `U_p` sind – als doppelte Integrale über die Raumladungsdichte – die Potentiale, die über dem kollektor- bzw. dem basisseitigen Anteil der Raumladungszone abfallen, wenn diese eine Ausdehnung von `y_n` bzw. `y_p` aufweisen.

Die Potentialanteile und die Randbedingung zusammenfassend, bestimmt `Yp` die Weite der Basis-Raumladungszone in Abhängigkeit von der anliegenden Spannung. In der letzten Zeile wird hieraus die Feldstärke am metallurgischen pn-Übergang in Abhängigkeit von der anliegenden Spannung bestimmt.

An Testlosen wurden Kollektordicke und -dotierung variiert (für die Postenbetreuung geht mein herzlicher Dank an Frau Hannelore Wagner!). Jede Kollektorvariante wurde auf vier Wafer aufgebracht. Je Wafer standen neun PCM-Felder („process control monitor“) zur Messung der Bauelementeigenschaften zur Verfügung, jedoch keine SIMS- oder spreading-resistance-Strukturen: Als Näherungswerte für Dicke und Dotierung der Kollektorepitaxie skaliere ich die oben ermittelten SIMS-Profile.

Für jedes der Dotierstoffprofile wird das Maximum der Feldstärke am Basis-Kollektor-Durchbruch berechnet. In Abb. 6.2 blau eingetragen sind die aus dem Experiment abgeleiteten Durchbruchfeldstärken. Die gefundenen Werte vergleichen sich freundlich mit dem grün dargestellten Literaturwert für Silizium ([III.2], S. 102), $E_m \approx \frac{4{\cdot}10^5 \mathrm{V/cm}}{1-log(N/1{\cdot}10^{16})/3}$:

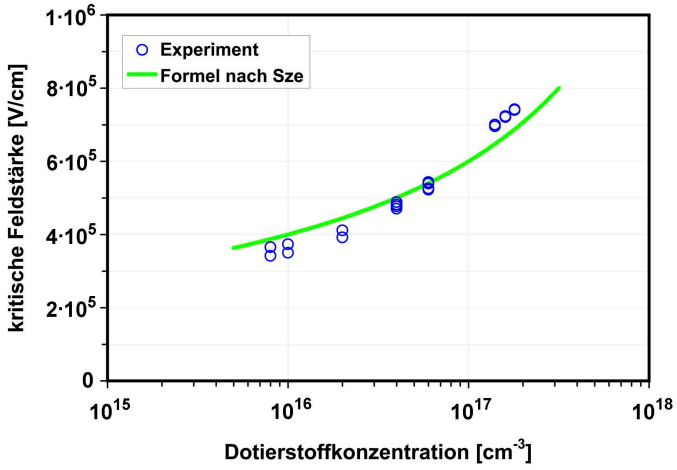


Abb. 6.2: Kritische Feldstärke: Experiment und Literatur

Zuversichtlicher geworden, bestimme ich eine dem Experiment nicht zugängliche Größe: das Maximum der Feldstärke für den Fall, daß genau die gemessene Emitter-Kollektor-Durchbruchbruchspannung anliegt. Die Feldstärke am Emitter-Kollektor-Durchbruch mit der Feldstärke am Diodendurchbruch ins Verhältnis setzend, erhalte ich für alle untersuchten Transistoren einen Quotienten von 2.1±0.1.

Mit diesem Quotienten kann ich für ein beliebiges Kollektorprofil die Emitter-Kollektor-Durchbruchspannungen abschätzen (der erhaltene Wert gilt selbstredend nur für den betrachteten Basisaufbau). Für die vier im Experiment untersuchten Dicken der Epitaxieschicht (450 nm, 500 nm, 650 nm, 700 nm) zeigt Abb. 6.3 den theoretischen Verlauf von BV_{CB0} (blau), BV_{CE0} (orange) sowie die Diodendurchbruchspannung einer „langen Diode“ (d.h., beliebig dicke Epitaxie), aufgetragen über der Kollektordotierung. Als Punkte sind die zugehörigen Messwerte als Mittelwerte über die jeweils 36 Messungen der Durchbruchspannungen eingetragen. Das denkbar einfache Modell liefert mehr als befriedigende Resultate. Entsprechend erlaubt die Simulation, für eine gegebene Epitaxie-Schichtdicke den Verlauf der Durchbruchspannung über der Dotierung vorherzusagen, und für das Experiment einen sinnvollen Schätzwert für die Dotierung festzulegen. Aus den Kurvenverläufen lassen sich sogar zu einer gewünschten Betriebsspannung Schätzwerte für Epidicke und -dotierung erhalten.

Betrachten wir zunächst die grüne Kurve, die Diodenduchbruchspannung der „langen Diode“: Sie fällt mit steigender Dotierung ab. Oberhalb einer Dotierstoffkonzentration von $1{\cdot}10^{17}\,\mathrm{cm}^{-3}$ flacht die Kurve ab, da in Silizium mit zunehmender Dotierung die Durchbruchfeldstärke ansteigt.

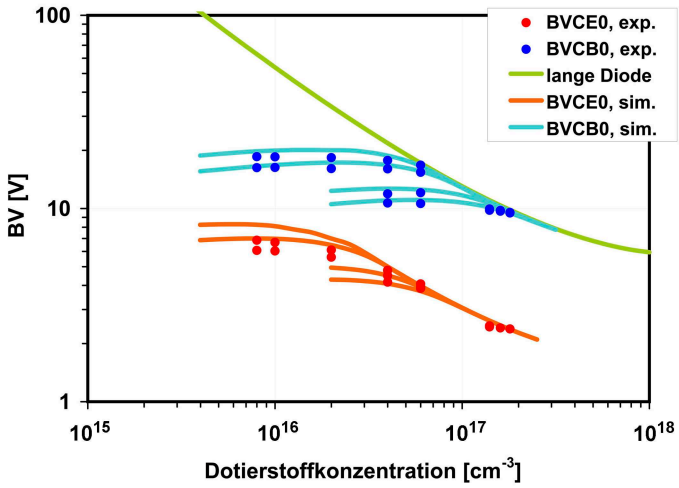


Abb. 6.3: Durchbruchspannungen in Experiment und Simulation

Die Diodendurchbruchspannungen (blau) zu den höchsten Kollektordotierungen liegen auf der Kurve für die „lange Diode“: Hier greift die Raumladungszone auch am Durchbruch nicht bis auf die vergrabene Schicht durch.

Bei niedriger Kollektordotierung stößt die Kollektorraumladungszone an der vergrabenen Schicht an. Die Diodendurchbruchspannung fällt unter den Wert für die lange Diode. Man mache sich dies anhand der Feldverteilung klar, verglichen mit der „langen Diode“. Der Durchbruch erfolgt jeweils, wenn an einem Punkt die Durchbruchfeldstärke erreicht wird. Das über der Raumladungszone abfallende Potential ist das Integral über die Feldstärke. Im Falle der langen Diode bildet der Feldstärkeverlauf ein Dreieck mit dem höchsten Punkt am metallurgischen pn-Übergang. Greift die Raumladungszone bis auf die vergrabene Schicht durch, wird die „rechte“ Spitze des Dreiecks abgeschnitten. Zu einem gegebenen Feldstärkemaximum gehört dann eine kleinere Fläche unter der Feldstärkekurve, die Durchbruchspannung sinkt.

Die Gleichungen sind korrekt gelöst. Der Vergleich mit dem Experiment geglückt. Das Resultat läßt sich intuitiv nachfollziehen – und dennoch bedarf es der kritischen Würdigung. Wie wir sehen werden, ist es, wegen der angenommenen Stromlosigkeit, für den Betrieb des Transistors vollständig irrelevant!

Ich habe, zu Illustrationszwecken, einen typischen Simulantenfehler emuliert: Die Verwendung eines nicht problemangemessenen Modells. Das Modell beschreibt die stromlose Diode. Dies ist richtig für die Betriebsbedingungen, unter denen die Durchbruchspannung gemessen wird. Deswegen fällt auch der Vergleich mit dem Experiment so vielversprechend aus. Im realen Betrieb des Transistors kann man jedoch den Einfluß der beweglichen Ladungsträger auf die Raumladungsdichte nicht vernachlässigen! Dies ist Gegenstand des nächsten Abschnitts:

6.4 Stromdurchflossener Kollektor

Stromdurchflossen wird der Transistor interessanter: Ein Teil der angelegten Spannung fällt an den Bahnwiderständen ab. Wird hierdurch der Basis-Kollektor-Übergang in Vorwärtsrichtung gepolt, wird der Transistor gleichzeitig vorwärts und rückwärts betrieben – er ist tief im Sättigungsbereich.

Die Elektronengeschwindigkeit sättigt bei moderaten Feldstärken. Ein Elektronenstrom ist mit einer bestimmten Dichte beweglicher Ladungsträger verbunden. Neben den ortsfesten Ionenrümpfe trägt auch die Raumladungsdichte der beweglichen Ladungsträger zum elektrischen Feld bei. Sie ist der ortsfesten Ladung der positiven Ionenrümpfe in der Kollektor-Raumladungsdichte entgegengesetzt.

Das linke Teilbild zeigt die Situation im stromlosen Transistor. Die anliegende Sperrspannung führt zur Ausbildung einer Raumladungszone; die Ladungsdichte im basisseitigen Teil der Raumladungszone ist negativ, während kollektorseitig positive Ionenrümpfe zurückbleiben. Die Feldverteilung, rot, als Integral über der Raumladungsdichte, weist ein Maximum am metallurgischen Basis-Kollektor-Übergang auf.

Abb. 6.4 skizziert in der oberen Reihe die Raumladungsdichten im npn-Transistor unter vier Betriebsbedingungen. Waagrecht verläuft die Ortskoordinate. B, C und NBL bezeichnen Basis, Kollektor und vergrabene Schicht. Als grüne Rechtecke sind idealisierte Raumladungsdichten wiedergegeben, in rot die resultierende Feldverteilung.

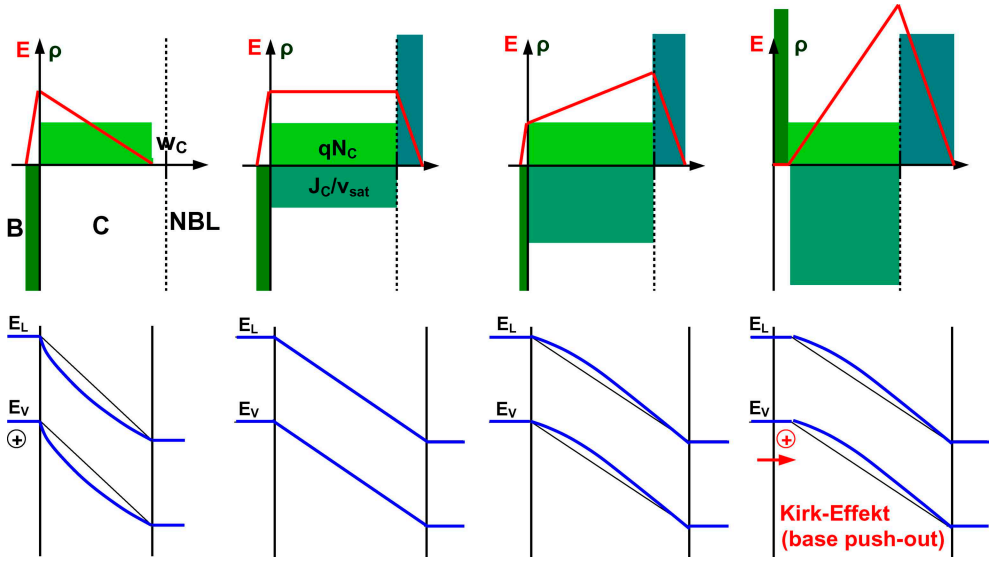


Abb. 6.4: Schemazeichnung Kirk-Effekt
Teilbilld eins: Der stromlose Transistor, Teilbild zwei: Moderate Stromdichten, Teilbild drei: Hohe Stromdichte, Teilbild vier: Einsetzen des Kirk-Effektes (Die Bandschemata vernachlässigen basisseitige Raumladungszonen, sonst wäre die Ableitung der Bandkanten stetig).

Bei kleinen Stromdichten (zweites Teilbild) tritt zur positiven Raumladungsdichte der Ionenrümpfe im Kollektor die negative Raumladungsdichte der bewegten Elektronen. Dargestellt ist die Situation bei der Kompensationsstromdichte $J_C{=}qN_Cv_{sat}$, wenn also die Raumladungsdichte der Ionenrümpfe gerade kompensiert wird. Die Feldverteilung im Kollektor verläuft in diesem Fall eben.

Erhöht man den Strom (Teilbild drei), kehrt sich die Neigung des Feldverlaufs um. Das Maximum der Feldstärke wandert an die Grenze zur vergrabenen Schicht.

Ein bei höchster Stromdichte verschwindendes Feld übersetzt sich in einen ebenen Valenzbandkantenverlauf. Die Löcher sind nicht mehr an die metallurgische Basis gebunden[24]: Die neutrale Basis weitet sich in den Kollektor hinein aus (**Kirk-Effekt** [III.2] oder „base push-out"). Die Basis-Gummelzahl wächst, Kollektorstrom und Verstärkung brechen zusammen, und f_T sinkt aufgrund der steigenden Basislaufzeit.

Die untere Reihe von Abb. 6.4 skizziert den Bandverlauf zwischen Basis und Kollektor: Herrscht in Teilen einer „Raumladungszone" die Ladungsdichte Null (zweites Teilbild; diese Situation findet sich auch bspw. in einer pin-Diode), so ist dort das Feld konstant, die Valenzbandkante, die den Potentialverlauf widerspiegelt, hängt linear vom Ort ab. Herrscht eine konstante Raumladungsdichte, verläuft die Feldstärke linear, und die Bandkanten nehmen die Form einer Parabel an. Je nach Polarität der Ladung ist die Parabel dabei nach oben oder nach unten geöffnet (Teilbilder 1 und 3). Verschwindet schließlich das Feld, das die Löcher an die Basis bindet, verlaufen die Bandkanten dort eben, die Löcher können sich frei bewegen (Teilbild 4).

In SiGe dagegen werden bei verschwindendem Feld die Löcher vom Valenzbandkantensprung am Verlassen der Basis gehindert – es kann sich tatsächlich ein negatives elektrisches Feld ausbilden. Das Feld bremst die Bewegung der Elektronen durch die Raumladungszone. Im Bänder-Modell bildet sich ein Buckel im Leitungsband aus, der die Bewegung der Elektronen behindert (heterojunction barrier effect, HBE; „parasitäre Leitungsband-Barriere"). Anders als beim konventionellen Kirk-Effekt[25] werden nicht die Gummelzahl erhöht und der Kollektorstrom verringert. Simulationen legen zwei andere Effekte nahe ([III.4]; ausgezeichnet [III.5] (vgl. 6.6), den simulierten Leitungsbandverlauf zeigt [III.6]):

- Durch eine höhere Ladung in der Basis – aufgrund einer Polarisationswolke aus gebremsten Elektronen und an den kollektorseitigen Rand der Basis gezogenen Löchern – steigt auch hier die Transitzeit merklich an, f_T bricht zusammen.
- Durch die höhere Dichte an freien Elektronen wie Löchern wächst die Rekombinationsrate; als Folge steigt der Basisstrom, die Verstärkung sinkt.

Während im herkömmlichen Transistor mit Überhandnehmen der Kompensation der Kollektorstrom zurückgeht, steigt im HBT der Basisstrom. In beiden Fällen brechen bei hohen Stromdichten Verstärkung und Transitfrequenz ein.

[24] Ein negatives Feld würde daher durch Umverteilung der Löcher kompensiert.

[25] Manche Autoren zieren sich, die Vorgänge im HBT mit dem selben Namen zu belegen; in meinen Augen ist dies Spiegelfechterei.

6.5 Der selektiv implantierte Kollektor

Um die Transitfrequenz zu erhöhen, müssen beim Schalten die Kapazitätsbeläge schneller umgeladen werden. Folgt man dem Hybrid-Pi-Modell, muß hierfür die Kollektorkapazität verringert, und somit der Kollektor niedriger dotiert werden.

Dies kann sinnvoll sein: In dem Fall, daß die Kollektor-Raumladungszone im Betrieb nicht vollständig ausgeräumt ist (vgl. S. 34), stellt der neutrale Anteil der Kollektorepitaxie einen unnötigen Vorwiderstand dar. Überwiegt der Einfluß dieses Widerstandes, kommt es zur Sättigung. Die effektive Basis-Kollektor-Sperrspannung sinkt bei steigendem Strom, die Raumladungszone zieht sich nur noch weiter zusammen.

Verringert man nun die Dotierung des Kollektors soweit, daß die Raumladungszone bis auf die buried layer durchgreift, beseitigt man den Vorwiderstand und verschiebt die Sättigung hin zu höheren Stromdichten. Ein – durch die größere Weite der Raumladungszone – kleinerer Kapazitätsbelag wird durch eine höhere Stromdichte umgeladen, der Transistor schaltet schneller.

Was, wenn jedoch umgekehrt die Kompensation dominiert? Um nun den Zusammenbruch der Verstärkung aufgrund der Kirk-Effekts zu höheren Stromdichten hin zu verschieben, muß die Kollektordotierung angehoben werden!

Dies ist der Gedanke hinter der selektiven Kollektorimplantation. In einer Technologie ohne Grabenisolation werden benachbarte Transistoren durch pn-Übergänge gegeneinander isoliert. Die Fläche der Basis-Kollektor-Diode ist wesentlich größer als der aktive, stromtragende Bereich. Eine hohe Kollektordotierung, und mithin ein erheblicher Kapazitätsbelag, ist nur in einer kleinen, aktiven Transistorregion erforderlich. Erzielt man diese vermittels einer Implantation durch die Emitteröffnung, gelingt es, Dotierung und Kapazität im Randbereich geringzuhalten, ohne die Stromtragfähigkeit im aktiven Teil zu opfern.

Bei SiGe2 ist der Name „SiC-Implantation“ jedoch irreführend. Richtig wäre, davon zu sprechen, daß durch die Implantation eine zweite Epi-Konzentration zur Verfügung gestellt wird. In einer STI-Technologie wie SiGe2 wird der Randbereich der Basis-Kollektor-Diode durch die Grabenätzung weggeschnitten. Gleichzeitig erfolgt die SiC-Implantation nicht erst durch die Emitteröffnung, sondern bereits vor der Grabenätzung. Die auf das Eintiefen der Gräben folgende Lineroxidation zur Beseitigung der Ätzschäden erzeugt eine gewisse Zahl an Siliziumatomen auf Zwischengitterplätzen (Interstitials). Ein Überschuß an Interstitials befördert die Diffusion namentlich der leichteren Dotierstoffe stark[26].

Normalerweise verteilen sich die Interstitials, welche selbst eine ausgesprochen hohe Diffusionsrate haben, schnell in der Tiefe des Bulkmaterials, wo sie durch zu diesem Zweck bei der Waferherstellung eingebrachte Sauerstoffprezipitate gebunden werden. Die voluminösen und hochkonzentrierten Arsenionen in der vergrabenen Schicht

[26] Dies ist der Hintergrund hinter der transient enhaced diffusion, TED („vorübergehend überhöhter Diffusion“) im Anschluß an eine Implantation, ebenso wie, im vorliegenden Beispiel, der oxidation enhanced diffusion, OED.

von SiGe2 verstopfen jedoch gewissermaßen die Poren, durch die sonst die Interstitials wegdiffundieren. Die Interstitials sind in den Mesa-Strukturen der Aktivgebiete gefangen. Schon unter dem Einfluß der verhältnismäßig geringen Temperaturbeaufschlagung während der dünnen Lineroxidation wird das Dotierstoffprofil der SiC-Implantation perfekt eingeebnet (vgl. das SIMS-Profil in Abb. 6.1). Es liegt kein in irgendeiner Weise „selektiv“ implantierter Kollektor vor, mit einer hochdotierten zentralen Region und niedrigdotierten Randpartien, sondern die Dotierstoffkonzentration wird im gesamten Kollektorgebiet gleichmäßig angehoben – gerade, als stünde lokal eine Epitaxie mit einer anderen Dotierung zur Verfügung.

Gewinnt nun bei SiGe2 die Sättigung oder die Kompensation? Müßten wir, zu der gegebenen Schichtdicke, die Kollektordotierung eventuell anheben oder absenken?

6.6 Die Verhältnisse bei SiGe2

Als Beispiel für einen Transistor ohne selektiv implantierten Kollektor (Non-SiC) betrachte ich einen SiGe2RF 20x0.5µm^2-Transistor im f_T-optimalen Arbeitspunkt[27] zwischen 6 und 10 mA. Bei $1{\cdot}10^7$ cm/s Sättigungsgeschwindigkeit entspricht dies einer Elektronendichte[28] von $6.2{\cdot}10^{16}$ cm^{-3} : Der mit $2{\cdot}10^{16}$ cm^{-3} dotierte Kollektor wird deutlich überkompensiert, die Raumladungszone dehnt sich in die buried layer hinein aus.

Der 20x0.5 µm^2 Transistor mit selektiv implantiertem Kollektor (SiC) weist eine deutlich höhere Transitfrequenz von 77 GHz auf. Es findet sich ein breites f_T-Optimum bei Strömen zwischen 10 und 35 mA. Die Dichte der bewegten Elektronen, $2.2{\cdot}10^{17}$ cm^{-3} , entspricht in etwa der Kollektordotierung. Auch beim SiC-Transistor wird sich im Betrieb die Raumladungszone bis zur vergrabenen Schicht hin ausdehnen: Der Vorwiderstand aufgrund der nicht-ausgeräumten Epitaxieschicht verschwindet im Betrieb.

Der Vergleich der optimalen Stromdichten mit der Kollektordotierung ist ein Hinweis, aber noch kein Beweis, daß tatsächlich der Kirk-Effekt die Transitfrequenz der SiGe-Transistoren begrenzt.

Aus der Datenbank von Herrn Dr. Kraus, Atmel, habe ich die bestehenden f_T-Messungen an $0.8{\cdot}10$ µm^2 (Kontaktanordnung C - E - B) Transistoren ausgelesen. Zum Zeitpunkt der Untersuchung waren 32 SiGe2-Lose gemäß dem Produktionsrezept[29] gelaufen. In Abb. 6.5 ist waagrecht die jeweils erreichte Transitfrequenz und senkrecht der optimale Strom aufgetragen.

---

[27] Laut Kompaktmodell. Dies gilt für Strom- wie auch für Leistungsverstärkung im Frequenzbereich zwischen 500 MHz und 5.45 GHz. Ich untersuche hier Transistor N20005E11xy, und verweise bezüglich der Namenskonvention auf das SiGe-Designkit.

[28] Während das Kollektorgebiet 1.2 µm breit ist, läßt die extrinsische Basis dem Strom nur einen Bereich von 0.5 µm Breite, vgl. Abb. 5.8. In der Kollektorschicht wird sich der Strom tendenziell wieder ein wenig aufweiten. Inwiefern dies Auswirkungen auf die Kompensation hat, zeigt im Einzelfall die physikalische Simulation.

[29] Einbezogen sind SiC- und NonSiC-Transistoren aus SiGe2RF und SiGe2Power. Von 934 fT-Messungen an 112 Wafern habe ich lediglich 4 Werte als Messfehler geklammert; etwas

Abb. 6.5 illustiert die Bedeutung der erzielbaren Stromdichte für die Hochfrequenzeigenschaften. Die Auswertung stärkt die Interpretation des Kirk-Effektes als begrenzendem Faktor: Je höher die Kollektordotierung, desto höher die Transitfrequenz[30]. Um nachzuweisen, daß tatsächlich der Kirk-Effekt die Transitfrequenz begrenzt, trage

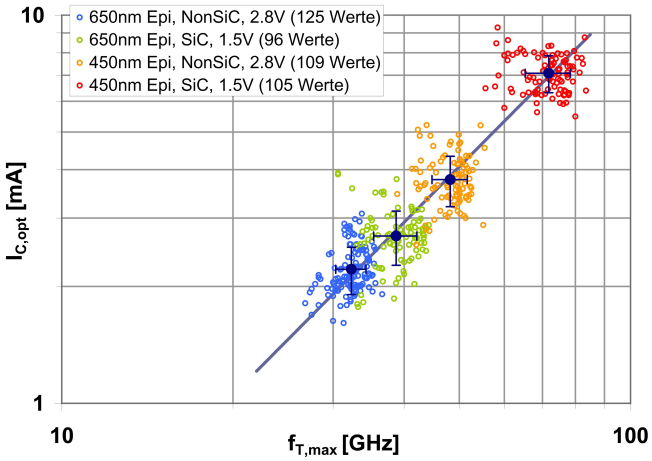


Abb. 6.5: Kollektorstrom bei der maximalen Transitfrequenz

ich die logarithmierten Basis- und Kollektorströme über der **inneren** Basis-Emitter-Spannung auf. Bei einer Auftragung über der angelegten Potentialdifferenz, wie etwa beim Gummel-Plot, maskiert der Einfluß der Bahnwiderstände den darzustellenden Effekt. Lee [III.5] ist mir hier knapp zuvorgekommen, ich reproduziere seine Grafik im linken Teilbild von Abb. 6.6. In dieser Auftragung ist offensichtlich, daß der Basisstrom ansteigt, und nicht der Kollektorstrom abbricht.

Im rechten Teilbild der Abb. 6.6 stelle ich, für drei Kollektorspannungen, dem Basisstrom die Transitfrequenz eines SiGe2RF-NonSic-Transistor gegenüber. Die optimalen Transitfrequenzen liegen genau bei der Stromdichte, bei welcher der Basisstrom abknickt: Die Transitfrequenz wird durch den Kirk-Effekt, bzw. die Herausbildung einer parasitären Leitungsbandbarriere limitiert.

---

aufwendig war die Postenzettel-Recherche im Archiv, um nachzuvollziehen, welche Lose den Standardprozeß erhalten haben und einbezogen werden konnten: In über 50 Losen wurden während des Einfahrens der Technologie Prozeßvariationen abgetestet.

[30] Wie man aus Abb. 6.5 ersieht, erfordert die zehnfache Transitfrequenz die 30-fache Stromdichte und mithin Kollektordotierung. Bei einer Steigerung der Dotierung um einen Faktor 30 sinkt die Sperrfähigkeit auf 18 % (und nicht etwa auf 10 %) ab. Die Proportionalität des Johnson-limit hat für niedersperrende Transistoren keine Gültigkeit (vgl. S. 5).

Für die Schaltungstechnik, welche auf eine gegebene Kollektordotierung festgelegt ist, ist ein anderer Multiplikationsfaktor wichtiger: Die f_T-optimale Stromdichte eines SiGe2-Transistors steigt um das 2.4-fache pro Frequenzdekade.

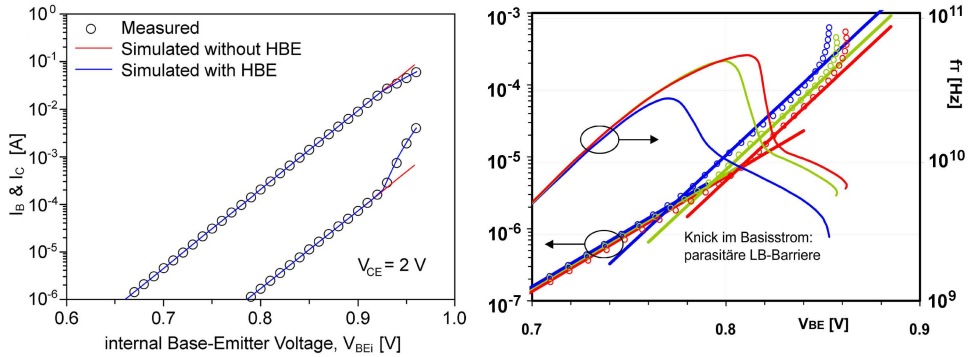


Abb. 6.6: Linkes Teilbild: Gummel-Plot nach Lee [III.5]; HBE = heterojunction barrier effect, vgl. S. 36. Rechtes Teilbild: Basisstrom und Transitfrequenz (rechtes Teilbild) über der inneren Basis-Emitter-Spannung bei SiGe2: Der Fingerabdruck der parasitären Leitungsbandbarriere. Die Farben entsprechen drei verschiedenen Kollektrospannungen.

6.7 Zusammenfassung

Zum einen habe ich untersucht, welche physikalischen Effekte im SiGe2-Kollektor eine Rolle spielen, quasi: wie sich der Transistor in der freien Wildbahn verhält.

Daneben kam es mir aber darauf an, anhand dieser Untersuchung zu zeigen, in welchen Fußstricken wir uns bei der Simulation verheddern können: An vielen Stellen hatten wir die Möglichkeit, einen Zahlenwert als Simulationsergebnis zu akzeptieren; stets hätten wir im später im Experiment Schiffbruch erlitten. Dies gilt um so stärker, wenn wir im nächsten Abschnitt eine Abstraktionsebene höher steigen: zu den Simulationswerkzeuge von SILVACO.

7 Bauelementsimulation, II: Die Basis des HBT

Im Anhang, Kapitel 18, stelle ich den Nutzen der **Technologie**simulation vor: Im Vorfeld des Experiments lassen sich Auswirkungen von Technologieänderungen abschätzen und allfällige Überraschungen abfedern.

Die Technologiesimulation ist besonders zuverlässig, wo grobe Strukturen betrachtet werden: Wird die anfänglich kleinräumige Dotierstoffverteilung einer Implantation durch einen ausgedehnten Temperaturschritt zu einer tiefen Wanne eingetrieben, liefert die Simulation ausgezeichnete Ergebnisse.

Bei kleinräumigen Strukturen, bei denen die Diffusion darüberhinaus durch ein inhomogenes Matrixmaterial und durch Verspannungen des Halbleiters beeinflußt wird, gibt es keine bauelement-unabhängigen Materialkonstanten. Technologiesimulation ist hier . . . sportlich. Auch sind die Optimierungsmöglichkeiten der Technologie durch die Simulation eingeschränkt, da das Basisprofil von den Anforderungen einer reproduzierbaren Gasphasenepitaxie bestimmt wird.

Meine **Bauelementsimulation**, die ich im aktuellen Kapitel vorstelle, geht daher von experimentell bestimmten Profilen und Querschnitten aus.

Zu den wichtigsten Meßmethoden zählen:

XRD: Unter streifendem Einfall wird ein Röntgenbeugungsmuster aufgenommen. Parallel wird am Rechner für ein angenommenes Germaniumprofil das resultierende Beugungsbild simuliert. Das Profil in der Simulation wird angepaßt, bis das errechnete mit dem gemessenen Muster übereinstimt. Die XRD-Messung erfolgt aufgrund der schlechten Fokussierbarkeit der Röntgenstrahlung und des erforderlichen streifenden Einfalls auf unstrukturierten Testwafern.

SIMS: Mit Sekundärionen-Massenspektroskopie lassen sich fein tiefenaufgelöste Dotierstoffprofile erstellen. Die Messung erfolgt an SIMS-Feldern mit typischerweise 600 µm Kantenlänge. Die SIMS-Felder stehen nur auf besonderen Test-Maskensätzen für die Technologieentwicklung zur Verfügung.

REM: Auf den Testmaskensätzen sind auch langestreckte Bruchstrukturen realisiert, die es erleichtern, bspw. quer durch einen Transistor zu brechen. Der Bruch wird naßchemisch präpariert, um den Kontrast der Strukturen im Elektronenmikroskop zu erhöhen und mit einem Edelmetall besputtert. Im geeichten REM läßt sich der Bauelementquerschnitt ausmessen.

REM-Brüche lassen sich, mit etwas Geschick, auch an Produktscheiben ohne dedizierte Bruchstrukturen erstellen.

Elektrische Messungen: Als Test ex posteriori dienen elektrische Messungen. Für die Messung müssen die Transistoren mit geeigneten Kontaktpads zur Verfügung gestellt werden. Dies geschieht auf Test-Maskensätzen oder in einzelnen „Process Control Monitors“, PCM, welche in geringer Zahl auf jedem Produktwafer einbelichtet werden.

Nicht nur vom Standpunkt der Simulation her, sondern auch experimentell stellt jede neue Technologie den Simulanten auf eine völlig neue Ebene seines Unwissens. Beispielsweise wird die korrekte Interpretation der Basisprofile aus der SIMS-Messung ausgesprochen vielschichtig. Wenn die Basis aus SiGe besteht treten zu den beschriebenen Eigenheiten der SIMS-Messung an Silizium insbesondere hinzu:

- Bei der SIMS-Messung werden quadrupol-massenspektrometrisch die Sekundärionen einzelner Ionenspezies gezählt. Die Tiefeninformation – und damit gleichzeitg das Eichnormal, um die Zahl der Ionen in eine Konzentration umrechnen zu können – stammt aus der bekannten Sputterrate, mit der die Primärionen die Probenoberfläche abtragen.

  Im SiGe-Transistor sind die Dotierstoffe nicht in eine homogene Matrix eingebettet. Ein reiner Silizium- und ein SiGe-Mischkristall werden unterschiedlich schnell abgetragen. Dies führt zu einer falschen Tiefeninformation und damit einer falsch angenommenen Dotierstoffkonzentration. Diesen Matrix-Effekt versucht man herauszurechnen, indem man die Abtragsrate an homogenen SiGe-Mischkristallen mißt. Aus den gezählten Germanium-Ionen wird nun selbstkonsistent die Germanium-Konzentration und die Abtragsrate bestimmt und danach das Dotierstoffprofil kalibriert. Die hierbei erreichte Genauigkeit hat ihre Grenzen unter anderem in den Verspannungen in der Si-SiGe-Si-Schichtfolge, welche ebenfalls die Abtragsrate beeinflussen.

- Die außer Haus erstellten SIMS-Profile habe ich daher anhand einer unabhängigen Messung des Germanium-Profils nachkalibriert anhand von XRD-Messungen auf blanken Wafern nachkalibriert, welche in der selben Epitaxiefahrt belegt worden sind.

- Mit Rest-Wasserstoff in der SIMS-Anlage verbinden sich Germanium-Ionen zu einem Ge-H-Komplex mit der selben Masse wie As. Er führt zu einem „Arsen-Nachweis“ auch in der Basis – ohne daß sich dort tatsächlich Arsen befände.

Die experimentell gewonnenen Querschnittsdaten lassen sich, ähnlich wie mit einem CAD-Programm, in den Bauelementsimulator ATLAS eingeben. Bequemer ist die Strukturdefinition mit dem speziellen Tool „DevEdit“ . Vor allem bei der Gitterdefinition ermöglicht DevEdit eine größere Flexibilität (Kapitel 19 gibt ein Beispiel für DevEdit-Eingabefiles).

Sechs anhand von XRD-Daten korrigierte Bor- und Germaniumprofile werden nachfolgend jeweils auf den maximalen Gehalt normiert und übereinandergelegt. Selbst diese, mit hohem apparativen Aufwand gewonnen prima-facie-Meßdaten sind für die Simulation noch ohne Wert, solange sie nicht, ausgehend von plausiblen Annahmen über die echte Dotierstoffverteilung, „geschönt“ werden: Die Interpretation der Messung, die bewußte und kontrollierte Abweichung vom Meßwert an sich, ist eine unabdingbare Kernkompetenz des Simulanten.

7.1 Basisprofil

Abb. 7.1 zeigt das gemessene Bor- und Germaniumprofil in der Basis. Die abgeschiedenen Bor- und Germaniumgehalte sind als Stufenprofile angedeutet. Während die Flanken nach bestem Wissen symmetrisch abgeschieden werden, springt an den gemessenen Profilen eine starke Asymmetrie ins Auge.

Eine überschlägige Berechnung der Diffusion zeigt, daß sich die unterschiedlichen Flankenformen nicht mit der Temperaturbeaufschlagung während der Epitaxie erklären lassen. Statt dessen beruht die Assymmetrie auf dem „knock-in“ (s. S. 31) während der SIMS-Messung: In der Sekundärionenkaskade werden Dotierstoffe tiefer in die Probe hinein verfrachtet. Für das verwendete SIMS-Verfahren gibt der Dienstleister (Evans, München) eine Stufenauflösung von 2 nm pro Konzentrationsdekade an. Tatsächlich erscheinen die in der Messung gezeigten 3 nm Anstieg pro Dekade also auch mit einem perfekten Stufenprofil vereinbar!

In Abwesenheit verläßlicher Daten über die Steilheit der Stufe nehme ich ein symmetrisches Basisprofil an: Die obere SIMS-Flanke wird nach unten gespiegelt (Abb. 7.1, schraffierte Linien, nicht gefüllte Dreiecke). Obwohl ich über eine Reihe hochaufgelöster SIMS-Profile verfüge, muß ich doch den Verlauf des Basisprofils raten!

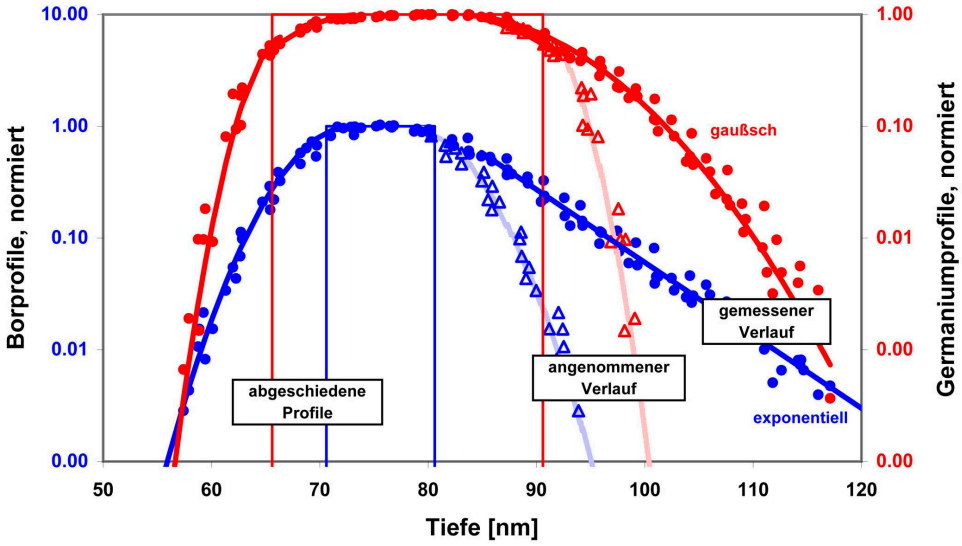


Abb. 7.1: Normierte Basisprofile, getempert

7.2 Physikalische Modelle und Materialparameter

Eine wichtige Voraussetzung für die erfolgreiche Simulation ist die Wahl problemangepaßter Modelle. Bewährt hat sich bei SiGe, konzentrations– und feldabhängige Ladunsgträgerbeweglichkeiten sowie das band-gap narrowing nach den Klaassen'schen Modellen zu berücksichtigen (**models conmob fldmob kla bgn**). Weiters muß die Konzentrationsabhängigkeit der Shockley-Read-Hall-Rekombination sowie, aufgrund der hohen auftretenden Stromdichten, die Auger-Rekombination berücksichtigt werden (**models consrh auger bgn**).

Wie wir bei der Modellierung des Kollektors gesehen haben, stellen die beweglichen Ladungsträger im Kollektor einen wichtigen Anteil der Raumladung. Als numerisches Lösungsverfahren wird daher sogenannte Newton-Methode gewählt, was bei Silvaco für eine selbstkonsistente Betrachtung der Dynamik und Raumladungsdichten der festen und beweglichen Ladungsträger steht (**method newton**).

Der Devicesimulator ATLAS von Silvaco kennt von Hause aus nur Silizium. Ich teile das Bauelement in Regionen ein und gebe für jede Region konstante Materialparameter vor, welche einer bestimmten Germaniumkonzentration entsprechen.

Während sich die Eigenschaften monokristalliner SiGe-Einkristalle mit den bekannten experimentellen Methoden bestimmen lassen (es stehen lediglich weniger Daten als bspw. für Silizium oder GaAs zur Verfügung) ist man für verspannte Heterostrukturen auf first-principles Monte-Carlo-Simulationen angewiesen. Gute Erfahrungen habe ich mit dem Parametersatz aus [III.7] gemacht. Die dort gegebenen Daten werden polynomisch gefittet, mit dem Germaniumgehalt als unabhängiger Variablen. Die Fit-Formeln finden sich im Anhang, s. S. 121 ff.).

7.3 Diskretisierung

Für die Simulation wird der Bauelementquerschnitt trianguliert. Das verwendete Gitter zeigt Abb. 7.2. Die Definition des Gitters erfordert einige Sorgfalt[31] :

- Jede Regionen-Grenze muß als Linie von Gitterknoten ausgeführt sein.
- Die nötige feine Tiefenauflösung des Germaniumprofils erfordert einen geringen vertikalen Abstand zweier Gitterebenen.
- Ein hohes Seitenverhältnis der Dreiecke behindert die Konvergenz. In der Tiefe fein abgestufte Gitter dürfen deswegen auch lateral nicht zu grob aufgelöst sein.
- Dies resultiert in einer großen Anzahl von Gitterpunkten. Die Simulationszeit steigt quadratisch mit der Zahl der Gitterpunkte. Für den betrachteten Querschnitt ergab sich ein Minimum der Rechenzeit bei etwa 19600 Gitterpunkten.

[31] In meinen ersten Simulationen wollte ich das Durchbruchverhalten lateraler HVMOS mit LOCOS-Feldoxid modellieren. Auf zwei Weisen diskretisiert, ergaben sich für den selben Querschnitt Durchbruchspannungen von 20 V bzw. 100 V!

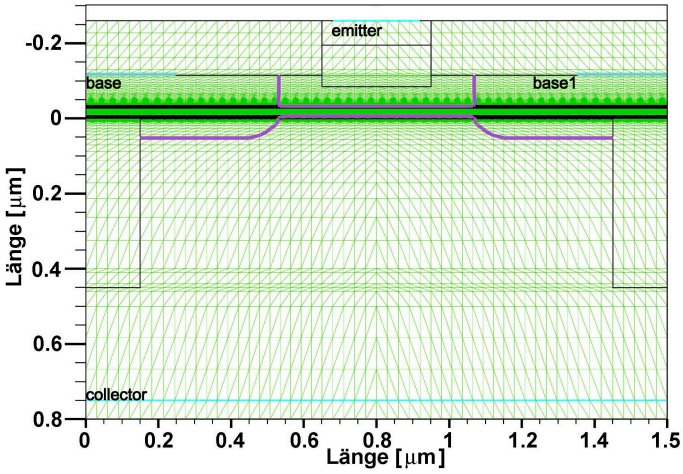


Abb. 7.2: Diskretisierung des Querschnitts

7.4 Geometrie

Ausgehend vom geschätzten Basisprofil (Abschnitt 7.1), von dem am Bruchbild vermessenen Querschnitt, von der in Abschnitt 6.2 angegebenen Kollektordotierung sowie von einer Technologiesimulation der Dotierstoffverteilung in der äußeren Basis und im Poly-Emitter setze ich ein idealisiertes, zweidimensionales Bauelementprofil zusammen[32]. Abb. 7.3 zeigt die farbkodierte Nettodotierung:

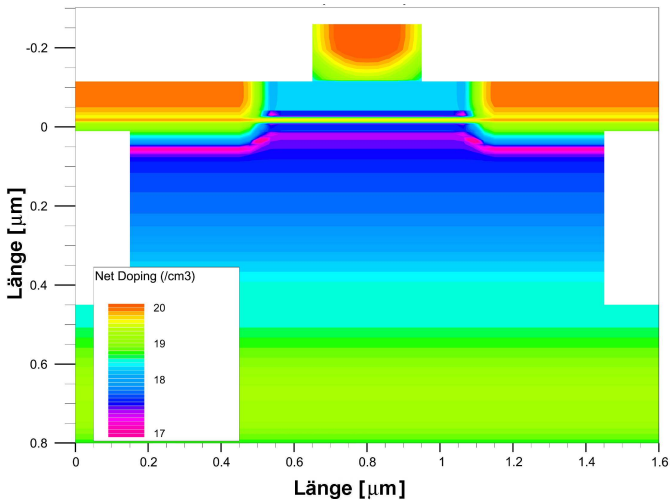


Abb. 7.3: Dotierstoffverteilung für die HBT-Simulation

[32] Die translationsinvariante Darstellung der vergrabenen Schicht entspricht einem Transistor mit Kollektoranschlüssen zu beiden Seiten des Emitters.

7.5 Materialparameter, oder: über die Funktion von Regeln

Für ein „schlecht definiertes“ Material wie SiGe ist die Versuchung groß, durch Drehen an den Materialparametern die Simulationsergebnisse mit der Messung in Einklang zu bringen. Hiervor kann ich nicht genügend warnen! Es wird andernfalls **unmöglich**, mehr als nur jeweils einen einzigen Transistor nachzusimulieren.

Einige wenige, meist minder wichtige, Parameter unterliegen jedoch prozeßbedingten Schwankungen und müssen an die konkreten Herstellungsbedingungen angepaßt werden. Meist betreffen solche „variablen Konstanten“ die Prozeßsimulation:

- Bei oxidierenden Eintreibschritten in Wasserdampfathmosphäre kann die Bildungsrate von Zwischengitteratomen an der Oxidgrenzfläche und damit die Diffusionsrate der Dopanden vom konkreten Ofenrezept abhängen. Man paßt die Parameter an durch den Vergleich der Simulation mit SIMS-Profilen.
- Ein gleiches gilt für die Akzeptorsegregationsrate, welche bspw. den Anteil implantierten Bors bestimmt, welcher sich nach der Oxidation im Oxid, anstatt im Silizium befindet („piling“).
- Die Aktivierungsrate von Arsen im Polysilizium-Emitter ist abhängig davon, „wie kristallin“ oder „wie amorph“ das Polysilizium ist. Sie wird anhand von SIMS- und Spreading-Resistance-Messungen abgeglichen.

Andere Parameter haben einen unmittelbareren Einfluß auf die elektrischen Eigenschaften und müssen bei der Bauelementsimulation berücksichtigt werden:

- Die Rekombinationsrate von Elektronen und Löchern an der Grenzfläche zwischen dem mono- und dem polykristallinen Emitter hängt von der Qualität der Grenzfläche ab.Vor der Polyabscheidung auf der Monosiliziumoberfläche verbleibendes natives Oxid segregiert zu Oxidinseln. Die große innere Oberfläche führt zu einer sehr effektiven Rekombination. Wird hingegen auf eine saubere Kristalloberfläche amorphes Silizium aufgebracht, setzt dieses bereits bei geringer Temperaturbeaufschlagung das Kristallgitter per Festphasenepitaxie monokristallin fort. Es kommt nicht zu einer erhöhten Rekombination an der „Grenzfläche“, der Emitter ist elektrisch dicker.

  Eine erhöhte Rekombinationsrate an der Mono-Poly-Grenzfläche habe ich, angelehnt an [III.8], bei der Nachsimulation von Atmels UHF6-Trasistoren berücksichtigen müssen, um die elektrischen Parameter befriedigend wiederzugeben.
- Einen ähnlichen Effekt – erhöhte Rekombination aufgrund von Störungen – glauben einzelne Autoren[33] an der Basis-Kollektor-Grenzfläche ihrer SiGe-Transistoren gefunden zu haben.

[33] siehe z. B. [III.9] (Experiment) und [III.10] (Simulation); interessanter Weise hat ein Teil der Autoren von [III.10] ein Jahr zuvor, in [III.11], das Basisstromverhalten auf Rekombination in der neutralen Basis zurückgeführt.

Aufgrund der hohen Ladunsgträgerlebensdauern in Silizium ist die Rekombination im Volumenmaterial i. A. zu vernachlässigen. Eine Ausnahme hiervon ist bspw. die Dotierung mit Gold in der Driftzone von Hochspannungsbauelementen. Hier werden bewußt lokalisierte Zustände innerhalb der Bandlücke geschaffen, um die Ladungsträgerlebensdauer zu verringern, und so die Durchbruchspannung auf Kosten des Einschaltwiderstandes zu erhöhen.

Auch Störungen im Kristallgitter können zu einer erhöhten Rekombination führen. Welche Auswirkungen hätte eine solche Rekombination im verspannten Kristallgitter am kollektorseitigen Rand der Basis?

Stünden Minoritäts- und Majoritätsladungsträger zur Verfügung, trüge die Rekombination zum Basisstrom bei. Würde der trap-behaftete Bereich dagegen ausgeräumt, ginge der Rekombinationsstrom zurück. Mit Basis und Emitter auf Masse fiele also der Basis-Leckstrom mit steigender Kollektorspannung ab.

Um zu untersuchen, ob eine solche Rekombination im Volumenmaterial bei SiGe2 eine Rolle spielt, habe ich in einem frühen Entwicklungszustand der SiGe2-Technologie den Leckstrom in die geerdete Basis in Abhängigkeit von der Kollektorspannung (mit Emitter auf Masse) bzw., der Emitterspannung (mit geerdetem Kollktor) gemessen. Aufgrund der kleinen Ströme ist eine hohe Integrationszeit erforderlich.

Es trat eine starke starke Abhängigkeit des Basisstromes von der **Emitterspannung im invers betriebenen Transistor** zu Tage. Hieraus läßt sich auf die Existenz von Rekombinationszentren zwischen Basis und Emitter schließen.

Messungen an verschiedenen Transistorgeometrien zeigen eine starke Abhängigkeit vom Formfaktor: Je höher der Randanteil des Transistors, desto höher der Einfluß der Emitterspannung auf den Basisstrom. Die Rekombinationszentren liegen daher bevorzugt entlang des Umfanges des Basis-Emitter-Diode, also im Übergangsbereich zwischen der äußeren Basis und dem Emitter.

Wir haben also in einem völlig praxisfernen Betriebszustand einen erhöhten Basis-Leckstrom – na und? Hat dieser irgendeinen Einfluß auf den Betrieb des Transistors? Um dies zu klären, nutze ich die Simulation.

Zunächst stelle ich die Ladungsträger-Lebensdauer im Grenzgebiet zwischen der äußeren Basis und dem Emitter so ein, daß die Simulation die gemessene Abhängigkeit des Basis-Leckstromes von der Emitterspannung reproduziert: Abb. 7.4 zeigt den Basis-Leckstrom, normiert auf den Wert bei geerdetem Emitter, $I_B(V_E)/I_B(V_E = 0)$, aufgetragen über der Emitterspannung. Rot ist der gemessene Leckstrom dargestellt, blau der simulierte Leckstrom für unterschiedliche Ladungsträgerlebensdauern. Aus dem Vergleich zwischen der Messung und den Simulationsergebnissen läßt sich die Ladungsträgerlebensdauer im Grenzgebiet bestimmen.

Abb. 7.5 zeigt den Einfluß, welchen die Berücksichtigung unterschiedlicher physikalischer Effekte auf die simulierte Stromverstärkung hat.

Blau dargestellt ist der simulierte Verstärkungsverlauf ohne und grün mit Berücksichtigung der Feldabhängkeit der Elektronenbeweglichkeit, **models fldmob**.

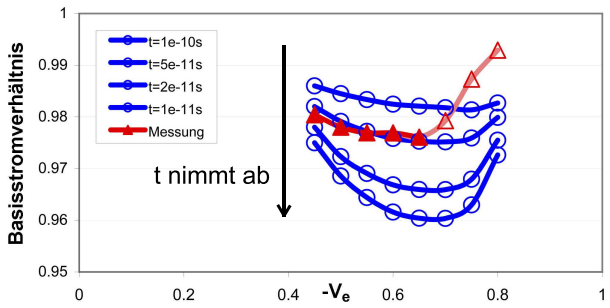


Abb. 7.4: Anpassung der Basis-Emitter-Rekombinationsrate, parametrisiert mit der Ladungsträger-Lebensdauer an der Grenzfläche: Basisstromverhältnis aus dem spannungsgesteuerten Ausgangskennlinienfeld in Basisschaltung

Der violett dargestellt Verstärkungsverlauf ergibt sich, wird zusätzlich zum Modell **fldmob** die aus dem Vergleich mit dem Experiment bestimmte Lebensdauer am Rand der äußeren Basis berücksichtigt.

Rot ist der gemessene Gang der Verstärkung wiedergegeben. Die gute Übereinstimmung der violetten, nicht jedoch der grünern Kurve mit dem Experiment zeigt, daß der starke Abfall der Verstärkung ein Artefakt einer gestörten Übergangszone zwischen äußerer Basis und Emitter darstellt.

Nach Verbesserungen im Silizidierungsmodul wird heute tatsächlich der grün dargestellte Gainverlauf auch im Experiment realisiert: Die Simulation kann manchmal selbst bei der Klärung „schwerstexperimenteller Dreckeffekte“ helfen!

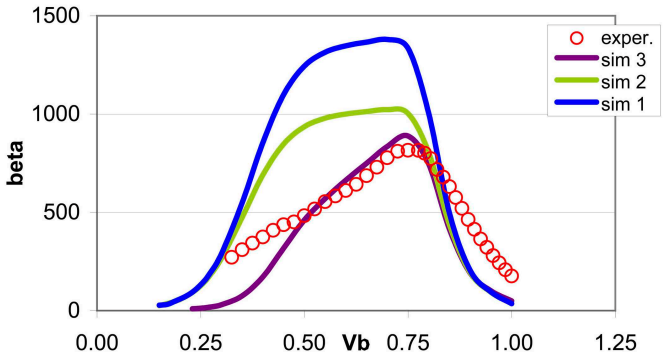


Abb. 7.5: Anpassung der Basis-Kollektor-Rekombinationsrate: B-Verlauf in Experiment und Simulation
– blau: **method newton**
– grün: + **models fldmob**
– lila: + Ladungsträgerlebensdauer am Rand der äußeren Basis reduziert

7.6 Nachsimulation der Technologie

Während die Voraussimulation in der Entwicklungsphase Hinweise geben kann auf Schnitzer im Herstellungsablauf, reift die Nachsimulation gemeinsam mit der Technologie. Für den letzten Abgleich mit dem Experiment sind ein stabiler Technologiezustand und verläßliche Meßwerte unabdingbar.

Aufgrund des hohen Meßaufwandes wird meist nur auf zwei Wafern pro Los das Hochfrequenzverhalten ausgewertet. Dies erlaubt keine statistisch relevante Aussage. Das Gleichstromverhalten wird jedoch automatisiert in vielen PCM-Feldern gemessen. Dies erlaubt Aussagen mit guter Statistik.

Die Early-Spannung und die Verstärkung lassen sich auf Kosten des jeweils anderen Parameters verbessern. Gleichzeitig reagieren beide empfindlich auf Schwankungen in der Basisepitaxie und bei der Prozessierung der Basis-Emitter-Grenzfläche, welche auch das Hochfrequenzverhalten beeinflußen.

Bei meinen Untersuchungen hat sich herausgestellt, daß das geometrische Mittel von Earlyspannung und Verstärkung (die „Güte“ des Transistors) ein Indikator ist für die Stabilität des Hochfrequenzverhaltens und mithin für die Aussagekraft einiger weniger HF-Messungen auf einem Wafer. Den Vergleich des Frequenzverhaltens zwischen Experiment und Simulation habe ich daher erst vorgenommen, als die Streuung in der Güte innerhalb eines Loses ausreichend stark abgenommen hatte.

Für ein solches stabiles Los habe ich SIMS-Profile erstellen lassen, Brüche angefertigt und im REM ausgemessen sowie f_T-Messungen angestoßen. Ich danke Frau Kraus für die geduldige und bereitwillige Unterstützung. Abb. 7.6 zeigt in grau den Verlauf der Transitfrequenz über dem Kollektorstrom für drei Wafer aus dem untersuchten Los.

Dunkelblau dargestellt ist die nachsimulierte Transitfrequenz unter der Annahme eines Gauß'schen Borprofils der Basis, das mit der symmetrisierten SIMS-Kurve innerhalb der Meßtoleranzen vereinbar ist. Eine Abschätzung der Bordiffusion während der Epitaxie läßt vermuten, daß in Basismitte ein kleines Bor-Plateau verbleibt, welches von symmetrischen „error function“-Kurven flankiert wird. Die Simulation mit einem solchen kleinen Plateau im Basisprofil – anhand der SIMS-Daten lassen sich beide Borverläufe nicht voneinander unterscheiden – ergibt den rot dargestellten f_T-Verlauf in ausgezeichneter Übereinstimmung mit dem Experiment.

Meines Wissens nach ist dies die erste SiGe-Nachsimulation, die sämtliche elektrischen Messungen innerhalb der experimentellen Streuung reproduzieren kann – und dies liegt nicht etwa an hohen experimentellen Streuungen!

7.7 Optimierungsmöglichkeiten

Die gute Nachsimulation der ausgereiften Technologie ermöglicht, die Grenzen für eine Weiterentwicklung abzuschätzen. Am Rechner – auch, wenn der Simulant hierbei nach Gehör spielt – lassen sich innerhalb von Stunden oder Tagen Auswirkungen von Technologievariationen abschätzen, deren experimentelle Klärung einen vollen

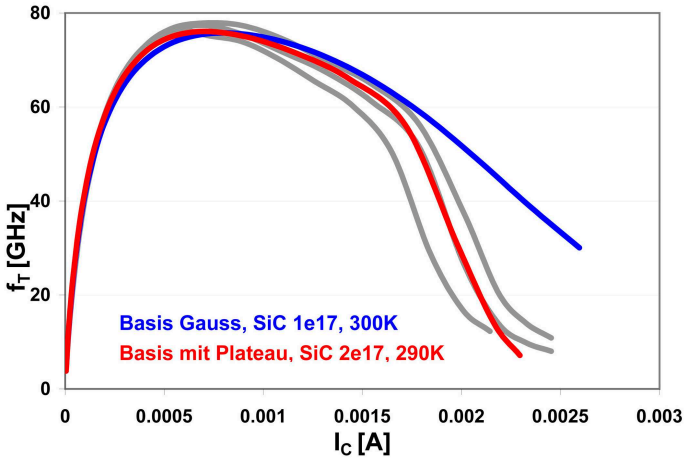


Abb. 7.6: Transitfrequenz vs. Kollektorstrom (Experiment und Nachsimulation

Technologie-Durchlauf (während der Entwicklungsphase zwischen 3 und 9 Monate) erfordern würde.

Ausgehend von SiGe2RF skaliere ich die Basisweite aggressiv (5 nm statt 10 nm abgeschiedene, bordotierte Basis), gehe mit dem Emitterprofil hart an die Grenze der experimentellen Reproduzierbarkeit und erhöhe den Germaniumgehalt von 19 % auf 25 %. Auch das Kollektorprofil wurde optimiert.

Die erzielte Verbesserung der elektrischen Parameter ist kaum der Rede wert, die Transitfrequenz steigt um 20 %. Der bestehende Technologieaufbau scheint, im gewünschten Sperrspannungsbereich, weitgehend ausgereizt!

Die SiGe-Simulation kann ich jedoch, scharf geschliffen, meinem wissenschaftlichen Werkzeugkoffer einverleiben; ihre Hilfe brauche ich im weiteren Verlauf der Arbeit, um die Eignung eines neu vorgeschlagenen Bauelementkonzepts beurteilen zu können.

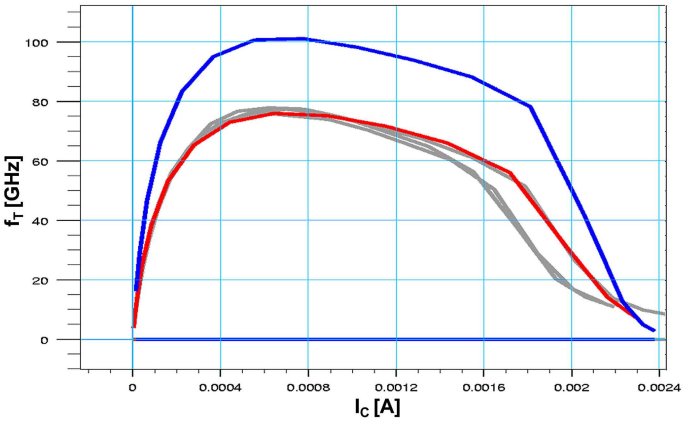
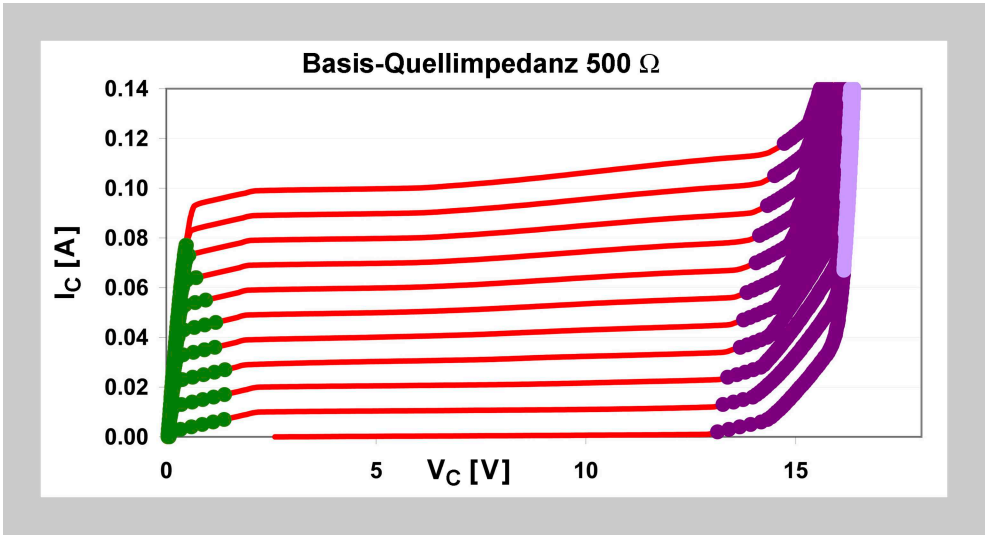


Abb. 7.7: Transitfrequenz vs. Kollektorstrom (grau: Experiment, rot: Nachsimulation, blau: Simulation mit optimiertem Prozeßablauf

IV Schaltungstechnik – Anforderungen und Lösungen



On such a full sea are we now afloat;
And we must take the current when it serves,
Or lose our ventures.

William Shakespeare, JULIUS CESAR [IV.1]

SiGe stößt als neue Technologie in einen reifen Markt. Kostenvorteil und Integrationsgewinnn gegenüber den III/V-Halbleitern kann sie nur ausspielen, wo sie bestehende Anwendungen ohne Systemmodifikation verdrängen kann. In eine GaAs-Anwendung paßt SiGe jedoch nur mit Mühe:

Aus einer Vielzahl von Gründen kann sich die Fußpunktimpedanz einer mobilen Antenne dramatisch ändern. Neben der Antennenumgebung spielen hier beispielsweise die Federkontakte eine Rolle, über die im Mobilfunkgerät Antenne und Leistungsverstärker miteinander verbunden sind. Bei Stößen lösen sich die Federkontakte kurzzeitig, der Ausgangstreiber ist leerlauffest auszulegen. Während GaAs-PAs aufgrund der hohen Durchbruchspannung die kurzzeitige Totalreflektion des Ausgangssignales unbeschadet überstehen, weisen SiGe-Antennenverstärker nicht die nötige „ruggedness“ auf.

Für SiGe bietet sich zum einen an, zu höheren Frequenzen, in unerschlossene Märkte auszuweichen – neue Schaltungssysteme können auf die Stärken von SiGe zugeschnitten werden. Ein künftiger Massenmarkt, der die hohe Integrationsdichte nutzen kann, ist das 77 GHz Automobilradar.

Zum zweiten lassen sich auf schaltungstechnischem Wege die Schwachstellen der Technologie umgehen, um zu einem konkurrenzfähigen Ausgangstreiber für Mobilfunk-Handgeräte zu gelangen. Diesen Weg möchte ich nachfolgend auskundschaften.

Bisherige war unsere Stoßrichtung, mit den Transistoren selbst der Ströme Herr zu werden. Häufig läßt sich jedoch mit den Mitteln der Schaltungstechnik erreichen, was für die Bauelementphysik allein unlösbar erscheint. Der f_T-Vervielfacher, den ich im Anhang behandle, ist ein bekanntes Beispiel.

Im vorliegenden Teil zeige ich, daß sich mit der Kaskodenschaltung gleichzeitig Betriebsspannung, Betriebsfrequenz und Ausgangsleistung über das mit den Einzelbauelementen Erzielbare hinaus erhöhen lassen. Allerdings sehen wir uns vor unlösbare Layoutprobleme gestellt, wenn wir einen Ausgangstreiber mit vielen Emitterfingern durch eine Kaskode ersetzen wollen.

Ich schlage daher ein neues Hochfrequenzbauelement vor, die vertikal integrierte Kaskodenstufe. Ein Großteil der Anschlußproblematik wird aus dem Schaltungslayout zurück in die Bauelemententwicklung gespiegelt; die Kollektorschichten beider Teiltransistoren der Kaskodenstufe lassen sich getrennt optimieren; das Versprechen der Kaskode von höherer Betriebsspannung, Betriebsfrequenz und Ausgangsleistung sollte sich für die Schaltungsentwicklung nutzbar machen lassen.

8 Die Kaskodenschaltung

8.1 Effizienz von Basis- und Emitterschaltung

Die Lastimpedanz, welche ein großer Ausgangstreiber im Effizienzoptimum sehen muß, liegt deutlich unter 50 Ω. Je höher die Last im Effizienzoptimum, das heißt, je höher der Spannungshub im Betrieb, desto leichter fällt die Anpassung. Ein hoher Spanungshub erfordert eine hohe Sperrfähigkeit. Daher betrachte ich nachfolgend neben der Emitter- auch die Basisschaltung.

Für das Automobilradarband bei 24 GHz vergleiche ich Basis- und Emitterschaltung in SiGe2Power anhand des HICUM-Modells des NonSiC-Transistor mit 24.3 µm mal 0.9 µm Emitterfläche, N243090210. Bei dieser Frequenz sind beide Schaltungstopologien unbedingt stabil, was das Konvergenzverhalten der load-pull-Simulationen erheblich verbessert.

Zunächst suche ich effizienzoptimierende Betriebsbedingungen. Brute force macht das keinen Spaß. Ⓟ 22 umreißt ein einfaches Verfahren, um effizienzoptimierte Arbeitspunkte im Ausgangskennlinienfeld zu finden.

Tabelle 8.1 zeigt die Betriebsbedingungen im Effizienzoptimum der Emitterschaltung; für die Basisschaltung wählen wir Betriebsbedingungen nach Tabelle 8.2; sie liegen leicht abseits vom Effizienzmaximum. Die Gründe werden später klar. (I_{C0} bezeichnet den Ruhestrom im Arbeitspunkt. Mit wachsender Eingangsleistung wird der Transistor dynamisch aufgesteuert. I_{C1dB} bezeichnet die nullte Fourier-Komponente des Kollektorstromes bei 1dB Kompression der Leistungsverstärkung.)

↑ $Z_L = (32.6 + j\,80.8)\,\Omega$		V_{CE}	$= 3.6\,V$	PAE	$= 24.6\,\%$
CE		I_{C0}	$= 4.5\,mA$	P_{in}	$= 8.7\,dBm$
↓ $Z_S = (10 - j\,5)\,\Omega$		I_{C1dB}	$= 14\,mA$	P_{out}	$= 12.9\,dBm$
		P_{dc}	$= 49\,mW$	G_{0dBm}	$= 5.2\,dB$

Tabelle 8.1: Effizienzoptimierende Betriebsbedingungen in Emitterschaltung

↑ $Z_L = (37.5 + j\,119)\,\Omega$		V_{CE}	$= 6.3\,V$	PAE	$= 35\,\%$
CB		V_B	$= 0.775\,V$	P_{in}	$= 6.7\,dBm$
↓ $Z_S = (6 - j\,6)\,\Omega$		I_{C1dB}	$= 17\,mA$	P_{out}	$= 15.9\,dBm$
		P_{dc}	$= 106mW$	$G_{0\,dBm}$	$= 10.2\,dB$

Tabelle 8.2: (Nahezu) effizienzoptimierende Betriebsbedingungen in Basisschaltung

Die Basisschaltung weist eine um zehn Prozentpunkte höhere Effizienz und eine um 5 dB höhere Verstärkung auf!

Damit ist noch keine Aussage getroffen, wie groß jeweils die Transistoren gewählt, und welche Ausgangsleistungen erzielt werden können. Dies ist in erster Linie eine Frage der Anpaßbarkeit der Transistoren:

8.2 Erzielbare Lastimpedanzen

In der Theorie läßt sich vermittels zweier geeigneter Reaktanzen der Ursprung auf jeden beliebigen (passiven) Punkt des Smith-Charts abbilden. In praxi stehen diesem Vorhaben verschiedene Einschränkungen im Wege.

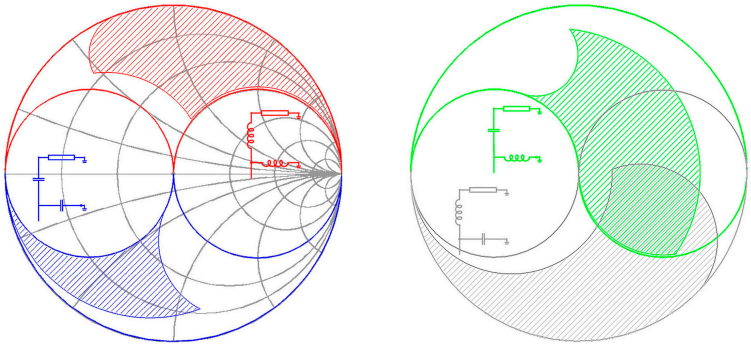


Abb. 8.1: L-Anpaßnetzwerke: Anpaßbereiche on-chip (schraffiert) & off-chip (umrandet)

Betrachten wir zunächst für einige einfache Anpaßschaltungen, auf welchen Bereich sich der Ursprung der Reflektionsfaktorebene abbilden läßt, wenn beliebige Reaktanzwerte zur Verfügung stehen. Diese Bereiche sind, in Abb. 8.1, für die dargestellten Anpaßschaltungen farbig umrandet (zur Klarheit sind nur dem linken Einheitskreis Impedanzkoordinaten unterlegt).

In integrierter Form stehen Reaktanzwerte von 0.1 bis 10 nH und von 0.1 bis 10 pF zur Verfügung. Hierdurch wird der Bereich, auf den sich 50 Ω abbilden lassen, auf die in Abb. 8.1 schraffiert dargestellten Teile des Smith-Diagramms beschränkt.

In Abb. 8.2 sind die Anpaßbereiche „on-chip“ für alle nichttrivialen Netze aus zwei Reaktanzen farbig umrahmt.

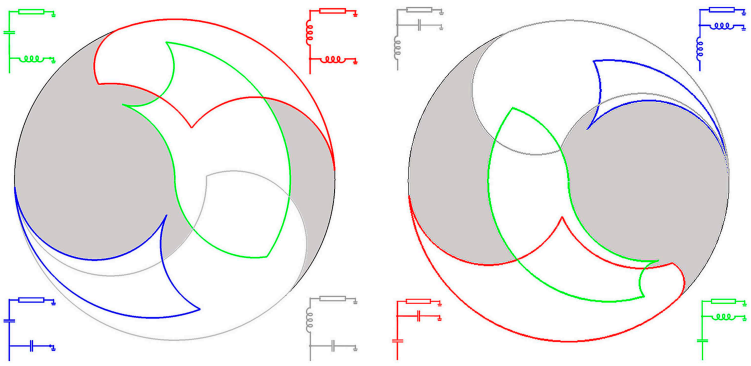


Abb. 8.2: L-Anpaßnetzwerke: Anpaßbereiche on-chip

8.3 Anpaßbarkeit und Fertigungstoleranzen

Die zur Anpassung verwendeten Reaktanzen unterliegen Fertigungsschwankungen (bei SiGe2 $\leq$10 % des Sollwertes). Entsprechend streuen die erzielten Impedanzwerte. Ein Maß für die Streuung ist der Reflektionsfaktor, der sich zwischen der Sollimpedanz und real erzielter Impedanz ergibt. Angegeben in dB und als Absolutwert wird diese Größe gemeinhin als „Anpassung“ bezeichnet. Um sprachliche Untiefen à la „Anpassung der Anpassung“ zu umschiffen, spreche ich statt dessen von „Fehlanpassung“.

Ich betrachte eine Schaltung in Gegenwart von Fertigungstoleranzen als anpaßbar, wenn Fertigungsschwankungen zu höchstens -10 dB Rückflußdämpfung von der Sollimpedanz führen können.

Das Smith-Chart links in Abb. 8.3 zeigt, farbkodiert, für die angegebenen Anpaßnetze die Fehlanpassungen aufgrund einer Abweichung der Reaktanzwerte um 10%. Eingefärbt sind die anpaßbaren Bereiche. Sie sind nochmal deutlich kleiner als die mit den verfügbaren Impedanzwerten insgesamt zugänglichen Bereiche: Zu dem grau dargestellten Anpaßnetz findet sich **kein einziger Punkt im Smith-Chart**, auf den sich der Ursprung mit der geforderten Tolleranz gegenüber Fertigungsschwankungen abbilden läßt!

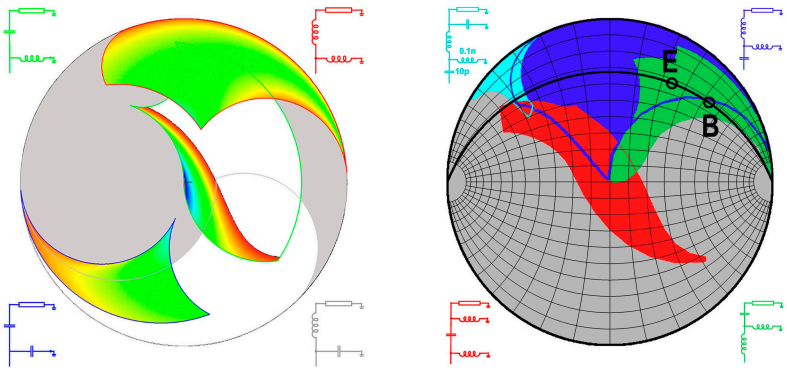


Abb. 8.3: Links: L-Anpaßnetzwerke on-chip: Beschneidung der anpassbaren Impedanzbereiche durch Fertigungstoleranzen
Rechts: Π- und T-Anpaßnetze on-chip: Anpaßbarer Bereich (die blaue Linie markiert den Rand des blau gehöhten Anpaßbereichs der blau dargestellten Anpaßtopologie)

Nachdem bereits die auf dem Wafer zur Verfügung stehenden Reaktanzwerte zu einer Einschränkung des anpaßbaren Bereichs geführt haben, stellt die Forderung nach Toleranz gegen Fertigungsschwankungen eine weit härtere Einschränkung dar.

Zieht man drei Reaktanzen zur Anpassung heran, geht also zu Π- und T-Anpaßnetzen über, erweitert sich der anpaßbare Bereich. Die hinreichend genau erzielbaren Reflektionsfaktoren im induktiven Teil des Smith-Charts zeigt, für alle dreigliedrigen Anpaßschaltungen, das rechte Teilbild aus Abb. 8.3.

Über die Anpaßbereiche habe ich, in Form dünner schwarzer Linien, ein Netz aus „Leistungskoordinaten“ der folgenden Art gelegt:

- Skaliert man die Emitterlänge mit der gewünschten Ausgangsleistung, ändert sich der problemangepaßte Last-Leitwert proportional. Die Leitwert-Gerade bildet sich auf einen Kreisbogen im Smith-Chart ab.
- Ein Transistor verschwindender Größe erfordert einen Leerlauf, unendliche Ausgangsleistung dagegen einen Kurzschluß als Last.
- Der problemangepaßte Lastreflektionsfaktor zu einer gegebenen Emitterlänge (und mithin erzielbaren Ausgangsleistung) bildet einen dritten Punkt auf dem Kreisbogen. Dies legt den Kreisbogen fest, auf welchem sich die problemangepaßten Lastreflektionsfaktoren bei Veränderung der Transistorgröße bewegen.

Als „waagrechte“ Koordinatenlinien habe ich einige solcher Kreisbögen eingezeichnet. Der Abstand zwischen zwei „senkrechten“ Koordinatenlinien entlang der „waagrechten“ Kreisbögen entspricht dabei 1 dB Differenz in der erzielbaren Ausgangsleistung, bzw. einem Faktor von 1.25 in der Emitterlänge.

8.4 Anpaßbare Ausgangsleistung

Ich trage die effizienzoptimierenden Lastreflektionsfaktoren aus Tabelle 8.1 und 8.2 in das Smith-Chart mit Leistungskoordinaten, Abb. 8.3, ein. Ein **B** markiert die Basis-, ein **E** die Emitterschaltung, als fetter scharzer Bogen ist der zugehörige Leistungskreis eingezeichnet[34].

Nun läßt sich für die Basis- und die Emitterschaltung die on-chip anpaßbare maximale Ausgangsleistung ablesen: Bei 24 GHz und im Effizienzoptimum erlaubt die Emitterschaltung, Ausgangsleistungen zwischen 6 mW und 100 mW, die Basisschaltung dagegen Leistungen von 20 mW bis 320 mW on-chip an 50 Ω anzupassen.

Verglichen mit der Emitterschaltung erlaubt die Basisschaltung, neben einer um 10 Prozentpunkte höheren Effizienz und einer um 5 dB höheren Verstärkung, 5 dB mehr Ausgangsleistung an 50 Ω abzugeben!

[34] Bei der Basisschaltung bin ich in Tab. 8.2 leicht von der effizienzoptimierenden Last abgewichen, damit die Anpaßbedingungen zu beiden Schaltungstopologien auf den selben Leistungskreis fallen.

8.5 Von der Basisschaltung zur Kaskodenstufe

Bei dem Bipolartransistor sind, aufgrund des exponentiell mit der Eingangsspannung ansteigenden Eingangsstroms, die problemanpassten Eingangsimpedanzen kleiner als die optimalen Lastimpedanzen. Dies ist ein unglückliches Verhältnis: Aufgrund der Verstärkung reicht für die Vorstufe jeweils eine kleinere Emitterlänge aus als für die nachfolgende Stufe, so daßsich das Impedanzverhältnis weiter verschlechtert.

Da die Basisstufe bei selber Größe eine um weitere 5 dB höhere Lastimpedanz erfordert, ist sie als Vorstufe ungeeignet. Ich versuche, die Basisstufe als Ausgangstransistor mit einer Treiberstufe in Emitterschaltung anzusteuern.

Nach Tabelle 8.1 und 8.2 kann eine Emitterstufe im Effizienzoptimum eine vier mal so große Basisstufe treiben. Die Kompression der Einzelstufen addiert sich, so daß die Vorstufen weiter im Back-off betrieben werden, es verbleibt ein Größenverhältnis von, sagen wir, 1:2. Um die Effizienz der Hauptstufe zu optimieren, muß die Vorstufe treiben, was auch immer ihr die effizienzangepaßte Endstufe an Impedanz präsentiert. Dies beeinträchtigt die Leistung der Vorstufe, und wir setzen die selbe geometrische Größe für Vor- wie für Endstufe an.

Ein Anpaßnetzwerken verbrät einen Teil der Leistung. Besonders bei hohen Frequenzen ist man bestrebt, die Leitungslänge zu minimieren. Ich verzichte auf eine dezidierte Anpaßstruktur zwischen Vor- und Endstufe, und füge beide, DC-gekoppelt, zusammen: Aus der Basisschaltung als Ausgangstreiber ist eine Kaskodenstufe geworden!

Von der Kaskodenschaltung erwarten wir eine schlechte Effizienz: Der Spannungsoffset im Ausgangskenlinienfeld begrenzt den Spannungshub. Betrachten wir trotzdem die Schaltung nach Abb. 8.4. Der Arbeitspunkt von Transistor Q1, dem Eingangstransistor einer Kaskodenstufe, wird über den Stromspiegel Q3 mit Helferlein Q4 eingestellt. Die Basis des Impedanzwandlers Q2 ist über eine Spannungsquelle mit Innenwiderstand auf 2.8 V gelegt. Die Quellimpedanz wird durch einen symmetrischen Spannungsteiler mit zwei Widerständen der Größe R modelliert.

Die Transistoren sind mit hart brechenden Dioden zwischen Basis und Kollektor dekoriert; der harte Durchbruch (bei 4.7 V in Q1, bei 9.6 V in Q2) läßt die für die Argumentation wesentlichen Punkte in den Ausgangskennlinienfeldern der Abb. 8.5 klarer hervortreten.

Nach der reinen Lehre wird das Potential der Basis des Impedanzwandlers einer Kaskodenstufe durch eine Spannungsquelle mit verschwindendem Innenwiderstand festgelegt. Bei einem Widerstand nahe Null im Spannungsteiler erhält man das bekannte Ausgangskennlinienfeld (Abb. 8.5, oberes Teilbild). Dunkelgrün markiert sind die Punkte, bei denen das Kollektorpotential des Impedanzwandlers Q2 um mindestens die intrinsische Diodenspannung unter seinem Basispotential liegt: Der Impedanzwandler ist in Sättigung. Purpurn dargestellt sind Punkte, bei denen die Sperrspannung der Basis-Kollektor-Diode in Q2 die Diodendurchbruchspannung (im Beispiel 9.6 V) erreicht.

Für das Ausgangskennlinienfeld im mittleren Teilbild von Abb. 8.5 wurde ein Spannungsteiler mit 2x500 Ω gewählt. Hier ergibt sich ein vollkommen anderes Bild.

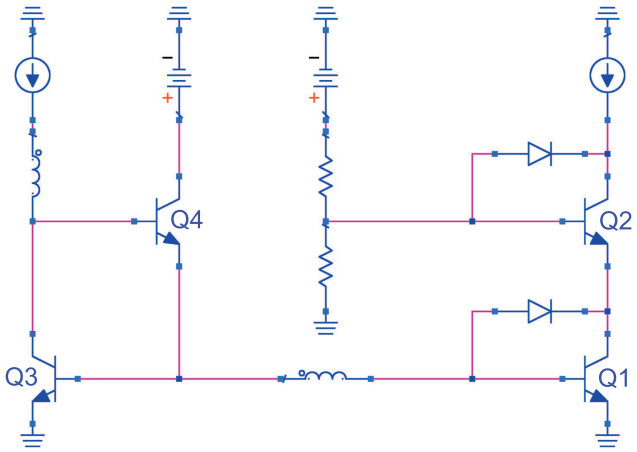


Abb. 8.4: Schematische Kaskodenschaltung

Betrachten wir zuerst die Situation nahe der Sättigung: Beginnt sich die Kollektordiode von Q2 zu öffnen, steigt der Basisstrom stark an, der Spannungsteiler wird belastet und die Basisspannung fällt ab. Die Kollektorspannung kann nicht unter das Basispotential sinken, die hochohmige Quelle gibt der Basis die Möglichkeit, dem Kollektor auszuweichen. Ohne daß Q2 sättigt, füllen die Ausgangskennlinien den ganzen Spannungsbereich bis zur Sättigung von Transistor Q1.

Gleichzeitig hat sich das Bild bei hohen Kollektorspannungen verändert. In der Kollektor-Raumladungszone des Impedanzwandlers Q2 unterliegen die Ladungsträger einem feldabhängigen Multiplikationsfaktor. Gibt man das Basis**potential** vor, fließen generierte Löcher durch den Basisanschluß ab. Steigt die Generationsrate, sinkt der von außen in die Basis einzuspeisende Löcherstrom. Oberhalb der Emitter-Kollektor-Durchbruchspannung zieht man gar überschüssige Löcher aus der Basis ab: Der Basisstrom ändert sein Vorzeichen („base-current reversal“).

Ein „umgekehrter“ Basisstrom hebt, bei hochohmiger Basisversorgung, das Basispotential von Transistor Q2 an. Die hochohmige Versorgung erlaubt der Basis, bei hoher Kollektorspannung dem Kollektor zu folgen, der Durchbruch in Transistor Q2 verschiebt sich zu höheren Spannungen.

Die Potentiale an der Basis von Transistor Q2 und am Kollektor von Transistor Q1 unterscheiden sich stets um die (stromabhängige) Basis-Emitter-Spannung. Steigt die Basispannung in Transistor Q2 zu weit an, bricht der Treibertransistor Q1 durch.

Nach klassischer Exegese wird die Treiberstufe Q1 mit vorgegebenem Basisstrom betrieben. Tatsächlich gibt der Stromspiegel in Abb. 8.4 jedoch die Basis-Emitter-**Spannung** vor, auch Q1 bricht erst bei der vollen Basis-Kollektor-Durchbruchspannung. Ich habe den Kollektor-Durchbruch von Q1 bei 4.7 V durch die parallelgeschaltete, steile Diode akzentuiert, und in Abb. 8.5 mit fliederfarbenen Punkten markiert.

Beim dritten Ausgangskennlinienfeld, unteres Teilbild von Abb. 8.5, wurde die hochohmige Basisanbindung mit 2x10 kΩ übertrieben.
Am oberen Ende des Ausgangskennlinienfeldes wird bei der Emitter-Kollektor-Durchbruchspannung in Transistor Q2 mit der Umkehrung des Basisstromes die Basis von

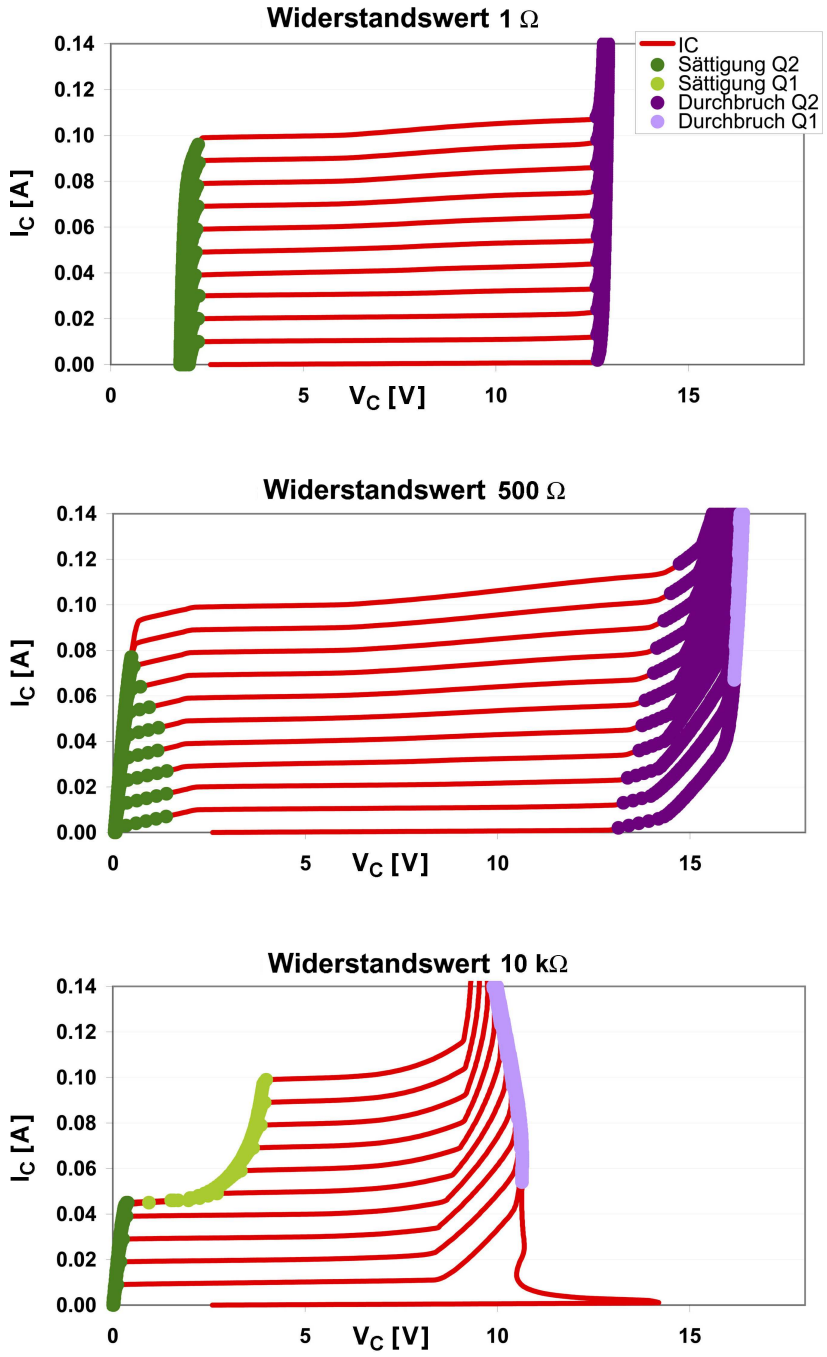


Abb. 8.5: Ausgangskennlinien der Kaskodenschaltung von Abb. 8.4 für unterschiedliche Widerstandwerte R im Spannungsteiler

Q2 steil nach oben gezogen: Der Kollektor von Q1 folgt der Kollektorspannung von Q2 im Abstand BV_{CE0}. Q1 bricht, bevor die Diodendurchbruchspannung des Transistors Q2 erreicht ist. Das Ausgangskennlinienfeld zeigt nur noch fliederfarbene, keine purpurnen Punkte mehr.

Bei kleineren Kollektorspannungen fließt der Basisstrom „in die richtige Richtung“. Die Spannung am hochohmigen Spannungsteiler bricht bei moderaten Basisströmen zusammen. Der Kollektor des Treibertransistors Q1 folgt der Basis des Impedanzwandlers im Abstand einer Basis-Emitter-Spannung. Dies führt zu Sättigung im Treibertransistor Q1. Im Ausgangskennlinienfeld habe ich die Punkte, an denen Q1 sättigt, hellgrün markiert. (Wieder wähle ich als Kriterium für die Sättigung, daß das Kollektorpotential um die intrinsische Diodenspannung unter dem Basispotential liegt.)

Das Gleichstromverhalten zusammenfassend läßt sich festhalten: Der Spannungsoffset im Ausgangskennlinienfeld hängt nicht von der Topologie, sondern von der Beschaltung der Kaskodenstufe ab; es erscheint daher möglich, Kaskodenschaltungen als effiziente Ausgangstreiber einzusetzen. **Geeignet versorgt, erweitert die Kaskode gegenüber der Basisschaltung die Spannungsfestigkeit und den erzielbaren Spannungshub noch einmal deutlich, grob überschlagen auf die Summe der Basis-Kollektor-Durchbruchspannungen beider Teiltransistoren.**

8.6 Hochfrequenzeignung der Kaskode

Im vorhergehenden Kapitel habe ich MEXTRAM-Modelle verwendet, welche ich in Kap. 14 aus simulierten Querschnitten extrahiere. Der harte Durchbruch der parallelgeschalteten Dioden erlaubt, die Effekte in den Ausgangskennlinienscharen deutlicher herausarbeiten. In diesem Kapitel rechne ich mit HICUM-Transistormodellen (N24309B110H aus SiGe2Power) und verzichte auf die Dioden.

Um das Frequenzverhalten von verschiedener Schaltungsvarianten arbeitspunktunabhängig zu vergleichen, bestimme ich die höchste Verstärkung im gesamten Ausgangskennlinienfeld und trage sie über der Frequenz auf. Ich vergleiche die Emitterschaltung und zwei Kaskodenschaltungen mit einer niederohmigen, rsp. einer moderat hochohmigen Spannungsquelle an der Basis von Transistor Q2 (vgl. Abb. 8.4, 8.5).

Die maximalen Strom-, bzw. Leistungsverstärkungen sind in Abb. 8.6 wiedergegeben.

Vergleichen Sie die grüne und die blaue MAG-, bzw. die rote und die blaue H_{21}-Kurve. Bei gleicher Transitfrequenz fällt die Stromverstärkung der Emitterstufe mit 20dB pro Frequenzdekade, bei der Kaskode jedoch mit 40dB: **Die untersuchte Kaskode verhält sich wie ein zweistufiger Verstärker mit geringer Fehlanpassung zwischen beiden Stufen!**

Dies ist, nach dem hohen Spannungshub, die zweite Überraschung, die die Kaskode bereithält. Hören wir Kaskode, denken wir: „Spannungsoffset!“ Tatsächlich erlaubt die Kaskodenstufe, geeignet betrieben, einen ausgesprochen hohen Spannungshub, und einen Kollektor-Swing bis nahe an Masse. Hören wir Kaskode, denken wir: „Miller-Effekt!“ und mit der Betonung des Miller-Effekts geht die Betrachtung der Kaskodenschaltng als einstufiger Verstärker plus Impedanzwandler einher. Dabei macht es –

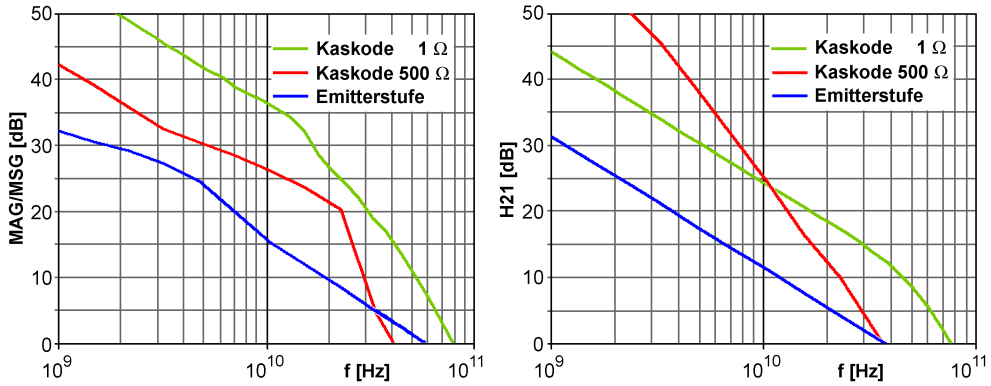


Abb. 8.6: MAG und H_{21} für Kaskoden-, rsp. Emitterschaltung

zumindest bei der betrachteten SiGe-Kaskode; für MOSFETs mögen andere Gesetze gelten – viel mehr Sinn, die Kaskode als zweistufigen Verstärker aufzufassen.

Das Zweite, was es an Abb. 8.6 hervorzuheben gilt, ist, daß sich mit der Kaskode die Transitfrequenz verdoppeln, und sogar die maximale Schwingfrequenz deutlich erhöhen läßt[35]!

Da die Kaskode gleichzeitig einen erheblich höheren Spannungshub gestattet, schlägt sie Haken um das Johnson-Limit[36]!: Wie versprochen, erweitert die Schaltungstechnik die Möglichkeiten, die die Bauelementphysik allein zur Verfügung stellen kann[37]!

8.7 Frequenzverhalten und Effizienz

In Abb. 8.6 habe ich das Maximum der Verstärkung über das gesamte Ausgangskennlinienfeld gebildet und über der Frequenz aufgetragen. Nun bilde ich zu einer festen Frequenz das Verstärkungsmaximum nur über den Ausgangsstrom, und trage dieses über der Ausgangsspannung auf.

Abb. 8.7 zeigt, für die Schaltung aus Abb. 8.4 bei 5.45 GHz, die Abhängigkeit der Leistungsverstärkung von der Ausgangsspannung, für drei Widerstandswerte im Spannungsteiler. Die Leistungsverstärkung wurde als max_outer(max_gain(S)) im Data

[35] Das Simulationsergbenis ist mit Vorsicht zu genießen: Ein zugrundeliegender Mechanismus in einer Ersatzschaltung erschließt sich mir nicht. Der Versuch, die Funktionsweise zu begreifen, hat jedoch zu einem neuartigen fT-Doppler, Ⓟ 6, geführt.

[36] Natürlich ist die crux, daß die gewinnmaximierende Impedanz zu klein ist, um den Spannungshub voll auszunutzen, während die Verstärkung durch Wahl effizienzoptimierender Abschlußbedingungen erheblich reduziert wird.

[37] Dies ist nichts Ehrenrühriges. Die Physik schreibt keinen unmittelbaren Zusammenhang zwischen der inversen Signallaufzeit und der Verstärkung vor - man halte sich prominente Gegenbeispiele wie etwa den Laser oder die Wanderfeldröhre vor Augen.

Display von ADS aus einem feinaufgelösten Ausgangskennlinienfeld mit S-Parameterschar bestimmt. Wird die Basis von Transistor Q2 niederohmig versorgt, zeigt die

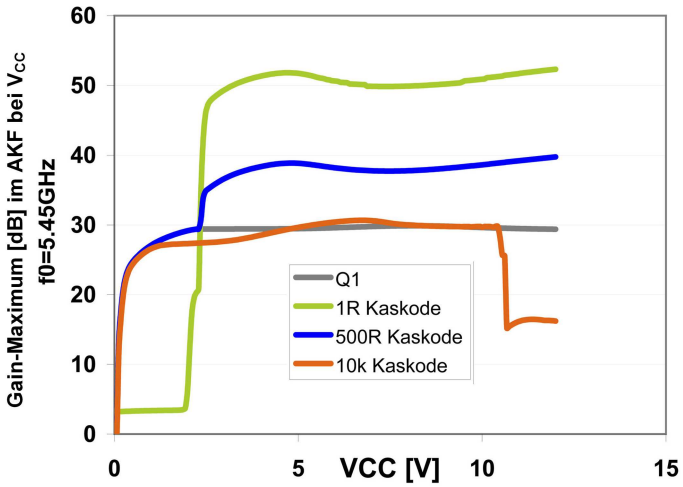


Abb. 8.7: MAG-Addition

Schaltung das Verhalten eines zweistufigen Verstärkers mit Spannungsoffset (grüne Kurve): Die hohe Verstärkung bricht abrupt zusammen, wenn die Ausgangsspannung unter 2 V sinkt.

Bestimmt man den Verstärkungsverlauf von Transistor Q1 allein ergibt sich der grau wiedergegebene Frequenzgang. Das Verhalten bei hochohmiger Spannungsversorgung (orange) gleicht dem der Emitterstufe allein: Der Impedanzwandler trägt weder Verstärkung noch Spannungsfestigkeit bei.

Der interessante Fall ist wieder der einer moderat hochohmigen Basisversorgung des impedanzwandelnden Transistors Q2 (blau). Anders als bei der Lehrbuch-Kaskode bricht die Verstärkung unterhalb von 3.3 V - 0.7 V = 2.6 V nicht abrupt zusammen; vielmehr verstärkt Transistor Q1 das Signal alleine, und Q2 reicht es durch. Oberhalb von 2.6 V kommt auch Q2 aus der Reserve. Aufgrund der „nachgiebigen“ Basis ist seine Verstärkung jedoch geringer als in reiner Basisschaltung. Der Einbruch der Verstärkung unterhalb von ca. 2.6 V führt zu einer Beimischung der dritten Harmonischen. Für die ins Auge gefaßte schmalbandige Anwendung – Ausgangstreiber für Mobilfunk-Handgeräte – kann man diese bei Ein-Träger-Modulation leicht wegfiltern.

Die Effizienz einer Kaskodenschaltung wird hiernach bei einer geeigneten Ansteuerung nicht von einem Spannungsoffset im Ausgangskennlinienfeld, sondern nur vom Energiegehalt der dritten Harmonischen beeinträchtigt (welcher sich, bei entsprechender Auslegung der Anpaß-strukturen, mit geeigneter Phasenlage wieder in den Verstärker zurückreflektieren und teilweise wiederverwerten läßt).

9 Lateral Integrierte Kaskodenschaltungen

Der Optimierung lateral integrierter Kaskodenschaltungen sind enge Grenzen gesetzt. Die Emitterstufe ist von der Lastspannung abgeschirmt. Um ihre Frequenzeigenschaften zu optimieren, ist eine minimale Sperrfähigkeit des Transistors wünschenswert. Gleichzeitig erfordert die Basisstufe eine hohe Spannungsfestigkeit, um den benötigten Spannungshub, und damit die Anpaßbarkeit einer hohen Ausgangsleistung an 50 Ω, zu gewährleisten.

Für die Kaskodenschaltung bei 1 W Ausgangsleistung ergibt die Simulation optimale Kollektordicken von 300 nm bei einer Dotierung von $1{\cdot}10^{16}$ cm^{-3} (Basisstufe) und von 70 nm / $7{\cdot}10^{17}$ cm^{-3} für die Emitterstufe. Beide Transistoren auf einem Stück Silizium zu realisieren, sprengt die Möglichkeiten der bestehenden Technologiearchitektur.

Das vorliegenden Kapitel umreißt konventionelle Lösungen. Die Untersuchung der handwerklichen Einschränkungen werden im nachfolgenden Teil der Arbeit zum Vorschlag für ein neuartiges Bauelement führen: die vertikal integrierte Kaskodenstufe.

9.1 Pedestal

Wie oben gezeigt (s. S. 38), bietet die SiC-Implantation einen Weg zu unterschiedlichen Kollektordotierungen in der selben Epitaxieschicht.

Bei SiGe2RF stellt die Dicke der Kollektorepitaxie einen Kompromiß dar zwischen SiC- und des NonSiC-Transistor: Die Ladungsdichte der Elektronen, die sich mit Sättigungsgeschwindigkeit durch die Raumladungszone bewegen, kompensiert die Ladungsdichte der ortsfesten Ionen. Die Netto-Dotierung sinkt, die Raumladungszone dehnt sich aus und greift bis auf die vergrabene Schicht durch, wo die hohe Ionendichte wieder zu einem nichtverschwindenden Feldgradienten führt: Im Hochstrombetrieb belastet den SiC-Kollektor keine unausgeräumte Kollektorschicht mehr mit einem zusätzlichen Bahnwiderstand (vgl. Abschnitt 6.6).

Es läßt sich leicht abschätzen, daß es bei einer Kollektordicke von 300 nm und einer Dotierung von $7{\cdot}10^{17}$ cm^{-3} nicht zur Kompensation kommt: steuert man den Transistor auf, gewinnt die Sättigung, die Raumladungszone schrumpft, anstatt sich auszudehnen.

Es ergibt sich daher, zusätzlich zur SiC-Implantation, die Aufgabe, die vergrabene Schicht nach oben fortzusetzen. Ein solcher „Sockel“, englisch pedestal, läßt sich auf verschiedene Weisen erzeugen. Einige untersuche ich nachfolgend auf ihre Eignung.

9.1.1 Implantationen

Der SiC-Implantation verwandt ist die „buried layer extension“: Eine tiefe Implantation setzt die vergrabene Schicht nach oben fort. Läßt sich der Bereich von 70 nm bis 250 nm Tiefe hoch mit Donatoren dotieren? Zunächst schätze ich die für die Kaskode erforderliche Dotierstoffkonzentration ab:

Eine Basisstufe mit 24x0.9 µm^2 Emitterfläche erfordert einen Realteil der Quellimpedanz von 6 Ω (s. Tabelle 8.2, S. 52). Dies dient als Schätzwert für den Realteil der Eingangsimpedanz.

Der Pedestal des zugehörigen Treibertransistors hat einen Querschnitt von 25x1.5 µm^2. Der Bahnwiderstand des Kollektoranschlusses bildet mit der Eingangsimpedanz der Basisstufe einen Spannungsteiler. Das Verhältnis beider Widerstandsbeiträge geht exponentiell in die Verstärkung der Kaskode ein.

Bei realistischen 5 Ω Bahnwiderstand trägt der Pedestal maximal ein weiteres Ohm bei. Es folgt ein spezifischer Widerstand von 20 mΩcm und eine moderate Dotierstoffkonzentration von $1.3{\cdot}10^{18}$ cm^{-3} .

Mithilfe der Tabelle auf S. 390 im „Ryssel“ [IV.2][38] findet man als Lösung eine Doppelimplantation Phosphor, $4.48{\cdot}10^{13}$ cm^{-2}/100 keV und $9.43{\cdot}10^{13}$ cm^{-2}/200 keV. Die Implantation erfolgt vor der Basis-Epitaxie und wird per RTA ausgeheilt.

Implantationsdosen und -Energien sind realistisch (für eine Darstellung der Implantationstechnik s. [IV.3], dem die folgende Zusammenstellung entnommen ist): In Abb. 9.1 ist waagrecht, für verschiedenen Implantertechnologien, die Energie der implantierten Ionen dargestellt (sie gibt ein Maß für die Eindringtiefe) und senkrecht die für einige Beispielanwendungen erforderliche Dosis. Die drei verwendeten Sorten Implanter sind farbkodiert dargestellt: Rot die Hochdosis-Implanter, schwarz die Mittelspannungsanlagen, blau die Hochenergieimplanter. Die erforderlichen Implantationen lassen sich leicht mit einem Mittelströmer erzielen.

Läßt sich auch eine Maske angeben, mit der sich die Implantation auf einen Teil des Wafers eingrenzen läßt? Ein Implantationsprofil hat ein Maximum bei einer energieabhängigen Eindringtiefe; es streut um dieses Maximum herum. Streuung und Eindringtiefe sind materialabhängig. Es darf höchstens auf dem 5-σ-Niveau durch den Lack durchimplantiert werden. Die Lackdicke kann ferner um 10 % vom Sollwert abweichen. Aus den Tabellen aus Landolt-Börnstein [IV.4] ergibt sich, daß die Lackdicke mindestens dreimal die Eindringtiefe in Silizium betragen muß.

Bei einer typischen Dicke einer HIPR-Lackmaske von 1.31 µm läßt sich die gewünschte Eindringtiefe von 250 nm also leicht maskieren. Auch hält der Lack leicht die gewünschte Implantationsdosis aus, ohne zu verbrennen.

Auf dem Papier scheint die „Pedestal-Implatation“ zu funktionieren. Den Strich durch die Rechnung machen die Implantationsschäden. Hochdosige Implantationen amorphisieren die implantierte Siliziumschicht, niederdosige nicht. Die hier erforderliche Dosis liegt gerade dazwischen: Im Maximum des Implantationsprofils wird „ein wenig“ amorphisiert, die entstehenden Kristallschäden sind schwer auszuheilen.

Mit den zur Verfügung stehenden RTA-Prozessen war ich tatschlich nicht in der Lage, mit einer ausreichenden Sicherheit die Bildung von Versetzungslinien aufgrund der

[38] Die Bedeutung einer intimen Kenntnis der Implantationsschäden für das Erstellen einer Technologiearchitektur macht [IV.2] unverzichtbar für den Technologen. 20 Jahre alt und seit langem vergriffen, werden antiquarisch 300 bis 500 US-Dollar für ein Exemplar gezahlt.

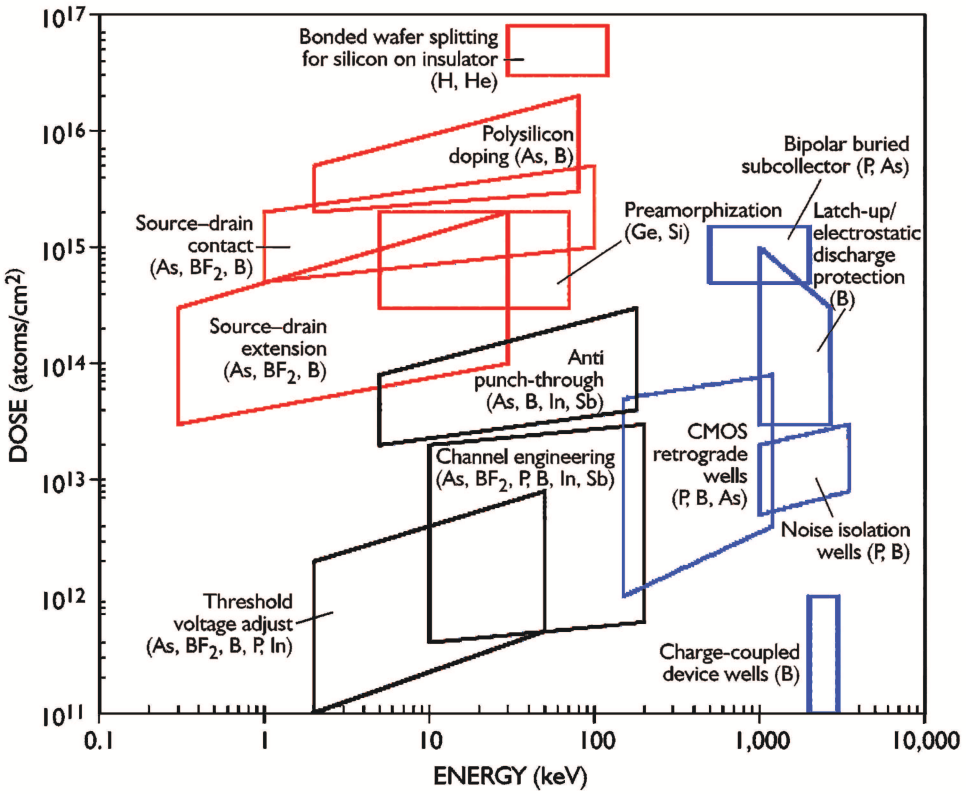


Abb. 9.1: Implantationsdosen vs. Energie, nach [IV.3]

Implantationsschäden zu vermeiden. Die Versetzungen werden epitaktisch durch die Basisschicht fortgesetzt und schließen den betroffenen Transistor kurz.

Leichter als in einer arsendotierten buried layer lassen sich die Schäden nach einer hochdosigen Bor-Implantation herauspräparieren, namentlich, wenn auf die Implantation eine Oxidation folgt. Die Implantationsschäden reifen hierbei zu sogenannten Oxidstapelfehlern (OISF, oxidation induced stacking faults), die nach einer Defektätzung als charakteristische „Strich-“ oder „Hantel-“ strukturen erscheinen (die als Beispiel herangezogene Präparation, Abb. 9.2, habe ich während der Technologieentwicklung von UHF6 für Atmel erstellt).

Eine Alternative liegt in der Implantation von Arsen oder Antimon: Durch die höheren Atomgewichte reichen die gewünschten Dosen zur Amorphisierung aus. Dem steht zunächst ein anderes Schädigungsmuster entgegen: Der EOR oder End-of-Range-Damage: Nur in der Nähe des Maximums des Dotierstoffprofils führt die deponierte Dosis zur Amorphisierung. Während des RTA setzt sich das Siliziumgitter von oben und von unten in die amorphe Zone hinein fort. Wo beide Kristallisationsfronten aufeinandertreffen, verbleiben Imperfektionen im Kristallgitter.

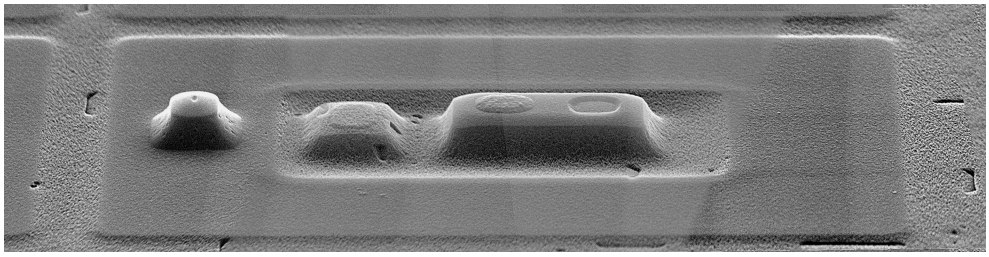


Abb. 9.2: Oxidation-induced stacking faults, OISF

Dem EOR kann man ausweichen, indem man auch die SiC-Implantation mithilfe schwerer Dotanden ausführt und den inneren Kollektor amorphisiert.

Wir sind in der Beschleunigungsspannung, je nach Wetter[39], auf maximal 250 keV begrenzt. Für Produktionszwecke beschränken sich die Kollegen aus der Implantation gerne auf 220 keV, die sie unabhängig von der Luftfeuchtigkeit erzielen. Die tiefe Implantation von Arsen oder Antimon erfordert auf den zur Verfügung stehenden Mittelströmern die Verwendung zweifach geladener Ionen. Zwar reichen 440 keV/Ion aus, um Arsen in die erforderliche Tiefe zu befördern. Die vorhandenen Ionenquellen liefern jedoch einen verschwindenden Strahlstrom an zweifach geladenen Ionen, und lassen eine Dosis von $1{\cdot}10^{14}\,\mathrm{cm}^{-2}$ aus wirtschaftlichen Gründen nicht zu.

9.1.2 Epitaxievarianten

Leichter läßt ein Pedestal sich mithilfe der Epitaxie erzeugen: Hierbei wird entweder die Kollektorepitaxie unterbrochen, der Pedestal hochdosig implantiert und nach Ausheilen die Epitaxie fortgesetzt Ⓟ 16. Eine zweite Lösung, Ⓟ 17, die ich nachfolgend beschreibe, ist das Aufbringen einer dicken Kollektorepitaxie mit anschließendem selektivem Abtrag des überschüssigen Kollektormaterials am Ort des niedersperrenden Transistors.

Ohne Ätzstop läßt sich die Tiefe eines Abtrags schwer einstellen. Als Ätzstop kommen für die kostengünstige Naßchemie zwei Verfahren in Frage:

– Während der Kollektorepitaxie wird eine dünne SiGe-Schicht mit einem Germaniumgehalt von 20 % abgeschieden. KOH, TMAH oder Brenzkatechin tragen Silizium ab, nicht jedoch, bei ausreichenden Germaniumgehalt, SiGe. In der Heilbronner Fab steht ein KOH-Ätzprozess zur Verfügung[40].

[39] Kein Witz! Selbst unter klimatisierten Reinraumbedingungen!

[40] TMMA steht für Tetramethyl-Ammoniumhydroxid, $(CH_3)_4NOH$. „Brenzkatechin“ bedeutet eigentlich eine Mischung aus Ethylendiamid, Brenzkatechin (englisch: Pyrocatechol), Pyrazin und Wasser. Andere Bezeichnungen sind daher EPPW oder Ethylendiamin-Pyrocatechol-Dissolution, EPD. EPD löst Silizium bei Temperaturen um die 100°C. Die Selektivität liegt bei Brenzkatechin und 15 % Germanium bei 50:1, und steigt bis zu ei-

Anschließend kann die SiGe-Schicht selektiv zum darunterliegenden Silizium entfernt werden: Eine ganze Anzahl HNO_3 –, H_2O_2– oder NH_4 –haltiger Ätzbäder läßt sich so einstellen, daß SiGe mit einer wesentlich höheren Rate als Silizium abgetragen wird [IV.6]. Die in der Kollektor-Raumladungszone der Basisstufe verbleibende SiGe-Schicht beeinflußt das elektrische Verhalten nicht.

– Alternativ trägt eine Mischung aus einem Teil Flußsäure (HF), drei Teilen Salpetersäure (HNO_3) und acht Teilen Ameisensäure (CH_3COOH) Silizium mit einer stark dotierungsabhängigen Ätzrate ab. Für hoch Phosphor-dotiertes Silizium (ρ<10 mΩcm) ergeben sich Ätzraten von 1 µm/min, um bei einer um eine Größenordnung niedrigeren Dotierung (ρ>100 mΩcm) auf ein Hundertstel dieses Wertes abzufallen [IV.7]. Die Ätzung kommt daher bei einer Dotierstoffkonzentration von $5{\cdot}10^{17}$ cm^{-3} faktisch zum Stillstand[41].

  Die tiefe Flanke eines Implantationsprofils fällt erheblich steiler ab als die der Oberfläche zugewandte. Eine maskierte Phopshor-Implantation stellt also eine ausgezeichnete Methode dar, eine Epitaxieschicht bis in eine wohldefinierte Tiefe dem Abtrag durch dotierungsabhängige Ätzlösungen zugänglich zu machen.

  Bei der dotierungsabhängigen Ätzung wird also zunächst eine für die Basisstufe optimierte Kollektorepitaxie aufgebracht. Wo später Transistoren in Emitterschaltung ausgebildet werden, wird durch die Implantation von Phosphor eine Opferschicht präpariert und diese naßchemisch entfernt.

Nach Aufwachsen eines Implantationsoxides, welches gleichzeitig Verunreinigungen aus einer alkalischen Ätzlösung binden kann, wird mittels Implantation das Dotierstoffniveau im Kollektor der Emitterstufe eingestellt.

So gut geeignet das beschriebene Verfahren ist, um zwei in ihrer Dicke und Dotierung stark unterschiedliche Kollektorschichten getrennt optimieren zu können, hat es für die hier verfolgten Zwecke einen gravierenden Nachteil:

Ein wichtiger Schritt im SiGe-Herstellungsablauf ist das Chemisch-Mechanische Polieren. Es trägt Kanten und Spitzen ab, während Hoch- oder Tiefebenen unverändert bleiben. An Stellen, wo sich Berge und Täler abwechseln, werden nur die Bergspitzen abgeschliffen. Mit dem beschriebenen Herstellungsverfahren lassen sich daher hoch- und niedersperrende Transistoren **in getrennten Bereichen** der Chips realisieren. Gleichzeitig ist es aber nicht möglich, die Basis- und die Emitterstufe einer Kaskodenschaltung ineinander zu verschachteln.

Eine blockweise Anordnung der Basis- und der Emitterstufe einer Kaskode erhöht (qua Leistungslänge) die Emitter-Induktivität der Basisstufe ebenso, wie (qua Zahl

---

nem Germaniumgehalt von 30 % auf 1800:1. Bei KOH steigt sie von 350:1 auf 1280:1 für Germaniumgehalte zwischen 20 % und 30 % [IV.5].

[41] Der Vollständigkeit halber weise ich darauf hin, daß in der Mikrostrukturtechnik eine andere Sorte dotierungsabhängiger Ätzung weiter verbreitet ist: Die Ätzung mit anisotropen Basen, wie KOH, stoppt auf hoch bordotierten Schichten, siehe bspw. [IV.8].

der Emitterfinger an einer Leitungsbahn) die induzierte Spannung pro nH an dieser Induktivität: Die Basisstufe erfährt eine starke Emittergegenkopplung. Hierdurch werden die Verstärkung und, in Folge, die Effizienz, reduziert.

9.2 Emitter-Down-Konfiguration

Der Emitter eines Bipolartransistors muß nicht unbedingt geometrisch oberhalb der Basis angeordnet sein. Frühe Beispiele finden sich z. B. in [IV.9], [IV.10]. Jüngst sind Transistoren in „emitter-down“ oder „reverse orientation“ beispielsweise an der Universität Delft für Mobilfunkanwendungen untersucht worden [IV.11]. Hier bieten sie insbesondere den Vorteil eines induktionsarmen Emitteranschlusses.

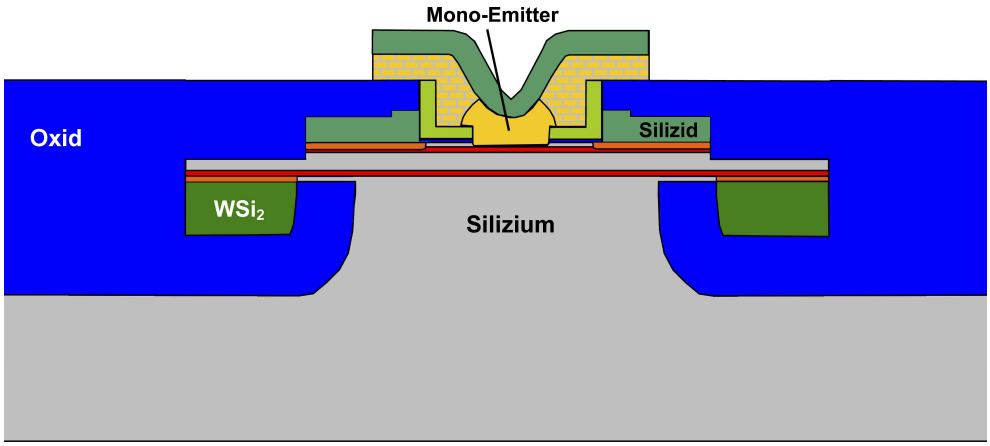
Ordnen wir in Gedanken zwei Emitterfinger parallel an, je einen aus der Basis- und einen aus der Emitterstufe. Der Emitter der Basisstufe liege dabei **unterhalb** der Basis, der Transistor stehe also auf dem Kopf. Die Emitterstufe dagegen sei herkömmlich mit geometrisch obenliegendem Emitter ausgeführt. Der untenliegende Emitter der Basisstufe und der untenliegende Kollektor der Emitterstufe seien durch die buried layer direkt miteinander verbunden.

Der Kollektor der niedersperrenden Treiberstufe läßt sich hervorragend 70 nm dick mit einer Dotierstoffkonzentration von $7{\cdot}10^{17}\,\mathrm{cm}^{-3}$ ausführen – er ist damit beinahe perfekt geeignet als Emitter der Basisstufe! Beide können tatächlich gemeinsam in einem Epitaxieschritt abgeschieden werden. Anschließend wird der STI-Graben tief in die buried layer hinein eingetieft (dies schafft die nötige Dicke des Grabenoxides, um die Kapazität zwischen der äußeren Basis und der buried layer zu reduzieren).

Nach der Planarisierung wird die p-SiGe-Basisepitaxie abgeschieden und abschließend eine 250 nm dicke n-Siliziumschicht mit einer Dotierstoffkonzentration von $1{\cdot}10^{16}\,\mathrm{cm}^{-3}$ aufgewachsen. Wo die letzte Siliziumschicht zu einem Emitter umgeformt werden soll, wird, wie oben, mit grob 200 keV Phosphor implantiert (wo das Phosphor bis in die Basis durchdringt, schafft es nicht, diese nennenswert umzudotieren).

So gut sich das Konzept für die Realisierung der Kaskode eignet (Ⓟ 10) liegt der Nachteil für eine **integrierte Schaltung** doch auf der Hand: Der hochsperrende Transistor ist in emitter-down-Konfiguration ausgeführt. Soll er außerhalb der Kaskode in Emitterschaltung verwendet werden, ist der Emitter lediglich mit der intrinsischen Spannung gegen das Substrat isoliert. Die kleine Weite der Raumladungszone und die großflächige Substratdiode belasten ihn mit einer hohen Kapazität und verschlechtern die Hochfrequenzeigenschaften.

V Weiterentwicklung von SiGe2



Von diesem Augenblick an verließ ihn schon nicht mehr das Gefühl einer verblüffenden Entdeckung, ähnlich dem Funken im Pulvermörser Bertholletts – einer dieser seelischen Einstürze, unter denen sich blitzend das Feuer herausreißt.

Alexandr Grin, DAS PURPURSEGEL [V.1]

10 Vertikal integrierte Kaskode

Es heißt, während sich ein guter Wissenschaftler durch Kreativität auszeichne, gelinge es einem guten Ingenieur, seine Aufgaben mit bewährten Mitteln zu lösen. Mir drängt sich der Gedanke auf, ein schlechter Ingenieur zu sein – meine Ideen werden zunehmend ... einfallsreich: Was spricht dagegen, die beiden Transistoren übereinander anzuordnen? Nach der Abscheidung von Kollektor, Basis und Emitter des unteren Transistors beläßt man den Wafer in der Epitaxiekammer und scheidet eine zweite Basis- und eine zweite Emitterschicht ab. Dies spart Fläche, und die Verdrahtung zwischen beider Stufen erübrigt sich auch.

Die entstehende Struktur – npnpn – ist bekannt: Zwei npn-Transistoren, durch einen pnp-Transistor miteinander verbunden, bilden einen Thyristor. Zwar leidlich schnell, aber nicht sonderlich linear: Einmal eingeschaltet, muß man die Betriebsspannung abklemmen, um ihn wieder zu sperren. Auch, wenn es zunächst nicht danach aussieht, kommt dieser Ansatz der idealen Kaskode dennoch schon sehr nahe.

In der Leistungselektronik dienen thyristorartige Strukturen tatsächlich zur Realisierung von Kaskodenschaltungen. Man geht hier den einfachsten Weg, ein Zünden des Thyristors zu verhindern: Schiere Masse! Schichtdicke und Dotierung in der Basis des pnp sind so hoch gewählt, daß die Schleifenverstärkung aus npn und pnp kleiner als eins wird. Bei den lateralen Strukturgrößen der Hochfrequenz ist dies nicht möglich.

Abb. 10.1 zeigt den Bauelement-Querschnitt aus einem Patent von SGS-Thomson [V.2]; n-Gebiete sind blau, p-Gebiete rot und Kontakte grün gehöht:

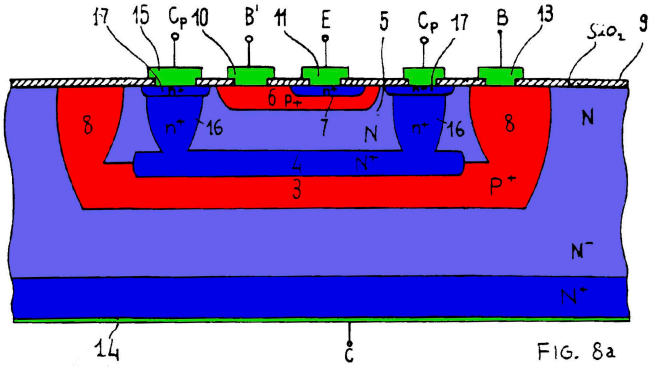


Abb. 10.1: Bauelement aus EP493854, Prio 31.12.1990

Das Patent von SGS-Thomson erwähnt die Schleifenverstärkung nicht und eröffnet Philips N.V. damit einen Weg, es zu umgehen. Indem Philips in [V.3] feststellt, das von

Thomson vorgestellte Bauelement neige zu thyristorähnlichem Verhalten, bekommt sie die Möglichkeit, es zu „verbessern“. Als Mesa-Struktur ausgeführt (Abb. 10.2), verläuft jeder Weg zwischen den beiden p-dotierten Gebieten durch eine hoch-n-dotierte Schicht (Bezugsziffer 12). Laut Philips dient diese Schicht dazu, daß die Ladungsträger „substantially entirely recombine“ – der Wirkmechanismus erscheint wenig stichhaltig, und läßt sich tatsächlich leicht mit einer kleinen Simulation widerlegen.

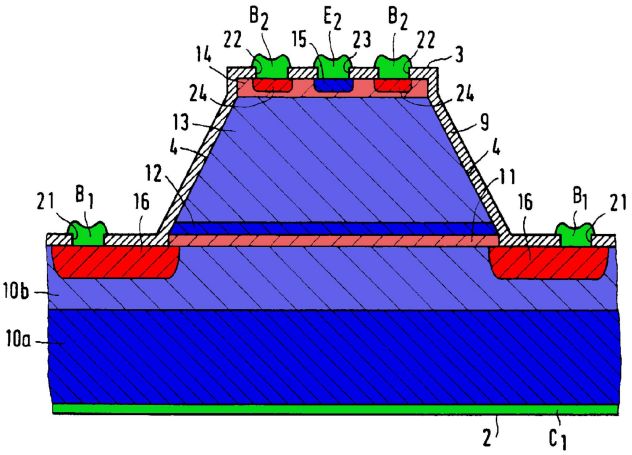


Abb. 10.2: Bauelement aus EP605920, Prio 28.12.1992

In einer von Texas Instruments eingereichten Schutzschrift werden die Ladungsträger tatsächlich zu einer vollständigen Rekombination gezwungen: In der vertikal intergrierten I^2L-Struktur aus [V.4] sind beide Teiltransistoren durch eine Silizidschicht, also quasi eine metallische Elektrode, voneinander getrennt (Abb. 10.3):

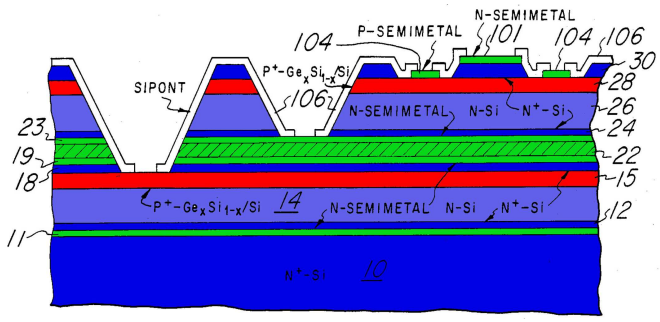


Abb. 10.3: Bauelement aus US4771013, Prio 01.08.1986

Derartige Strukturen lassen sich z.B. durch Waferbonden, durch die heteroepitaktische Abscheidung von Wechselschichten von Silizium mit bestimmten gitterangepaßten Siliziden ($CoSi_2$,$ErSi_2$) oder durch Allotaxie herstellen: Nach epitaktischer Abscheidung einer einkristallinen Siliziummatrix mit Silizidausscheidungen und einer Silizium-Deckschicht, welche die monokristalline Struktur fortsetzt, formiert sich unter einer aggressiven Hochtemperaturbehandlung eine durchgehende Silizidschicht, die zwei einkristalline Siliziumbereiche trennt [V.5], [V.6] (Vgl. auch Ⓟ 20).

10.1 Vertikale Integration in SiGe

Keimzelle meiner Arbeit war die Überlegung, der Parasit solle sich recht effektiv unterdrücken lassen, wenn die Basen der vertikal integrierten Kaskodenschaltung in SiGe ausgeführt werden. Beim Verlassen der Basis – sie stellt den Emitter des parasitären pnp dar – laufen die Löcher gegen den Valenzbandkantensprung. Dies trennt sozusagen den Emitterkontakt des pnp durch.

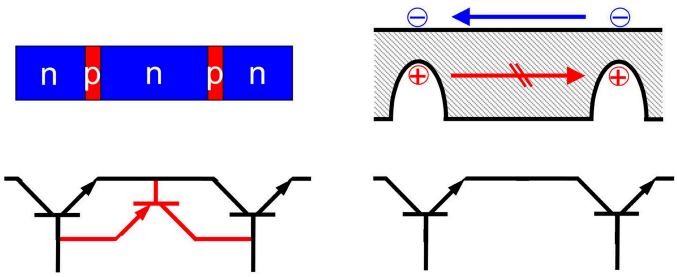


Abb. 10.4: Thyristor

Links in Abb. 10.4 ist noch einmal die npnpn-Struktur abgebildet. Im Silizium ist ihre Ersatzschaltung die eines Thyristors. Mit SiGe in den Basen ergibt sich das oben rechts gezeigte Flachbanddiagramm – sowie, als Ersatzschaltung, eine Kaskode!

Die entstehende vertikal integrierte Heterokaskode ist Gegenstand der Einreichungen Ⓟ 8, Ⓟ 9, Ⓟ 18 und Ⓟ 21.

In zwei prägenden Jahren bei meinem verehrten Lehrer Prof. Harro Heuser habe ich seine Skepsis übernommen gegenüber den „drei klassischen Beweismethoden: Überredung, Überzeugung und Unterdrückung“. An der Idee, den Parasiten durch einen Hetero-Übergang zu unterdrücken, kann man sich leicht berauschen. Auch scheinen überschlägige Simulationen sie zu bestätigen. Allerdinge ist sie vollständig Unfug!

Nachfolgend betrachte ich zunächst nocheinmal die Verhältnisse im Transistor, dann die Unterschiede zum Hetero-Transistor, um schließlich zum Thyristor überzugehen. Erst dabei klärt sich die Frage nach der latch-up-Neigung des vorgeschlagenen Bauelements abschließend.

10.1.1 Halbleiterphysik

Startpunkt ist die Boltzmann-Gleichung, [III.2], mit der Dichte n, p der Elektronen und Löcher, der intrinsischen Ladungsträgerkonzentration n_i, dem intrinsischen Fermilevel E_i und den Ferminiveaus E_F^n und E_F^p für Elektronen und Löcher:

$$n = n_i \, exp\left(\frac{E_F^n - E_i}{kT}\right) \quad , \quad p = n_i \, exp\left(\frac{E_i - E_F^p}{kT}\right). \tag{10.1}$$

Im Gleichgewicht gilt $E_F^n = E_F^p$, und das Produkt der Elektronen- und der Löcherdichte ist gleich dem Quadrat der intrinsischen Ladungsträgerkonzentration n_i.

Durch Anlegen einer Spannung V ändern sich die Ferminiveaus E_F^n und E_F^p. Mit der Ladung $-q$ eines Elektrons gilt $-qV = E_F^p - E_F^n$, und das Produkt der Ladungsträgerkonzentrationen wird

$$np = n_i^2 \, exp\left(\frac{E_F^p - E_F^n}{kT}\right)$$

Aufgrund der Ladungsneutralität ändern sich unter Spannung die Konzentrationen der Majoritätsladungsträger unwesentlich. Für die Minoritäten n_p, p_n (d. h., Elektronen im p-dotierten, und Löcher im n-dotierten Halbleiter) erhält man

$$n_p = \frac{n_i^2}{p_p} exp\left(\frac{qV}{kT}\right) \quad , \quad p_n = \frac{n_i^2}{n_n} exp\left(\frac{qV}{kT}\right). \tag{10.2}$$

Im Rahmen der Shockley'schen Näherung, wenn insbesondere die Grenzen der Raumladungszone scharf verlaufen, so daß die über der Raumladungszone abfallende Spannung wohldefiniert ist, stellen Glgn. 10.2 die Randbedingungen für die Minoritätladungsträgerkonzentrationen am Rand der neutralen Gebiete dar.

An den Grenzflächen zwischen Halbleiter und Kontakte rekombinieren sämtliche Überschuß-Minoritätsladungsträger. Hier gelten als Randbedingungen $n_p = p_n = n_i$.

Nichtgleichgewichtskonzentrationen von Minoritätsladungsträgern entwickeln sich im neutralen Halbleiter diffusiv und mit einer endlichen Lebensdauer. Betrachtet man die Ausdehnung der neutralen Gebiete im Transistor, gemessen senkrecht zum Rand der Raumladungszonen, als klein gegenüber der Diffusionslänge, wird der Verlauf der Minoritätsladungsträgerdichte linear, wobei die Randbedingungen den Minoritätsladungsträgerverlauf in den neutralen Gebieten festlegen.

Aus der Kontinuitätsgleichung und der Einstein-Relation (mit minuskulen Approximationen, s. z. B. [III.2], S. 86 f.) folgt die Diffusionsstromdichte als (D_p, D_n: Diffusionskonstante von Löchern bzw. von Elektronen):

$$J_p = -qD_p \frac{\partial n_p}{\partial x} \quad , \quad J_n = qD_n \frac{\partial p_n}{\partial x}. \tag{10.3}$$

Mit dem Verlauf der Ladungsträgerkonzentration kennt man auch die Elektronen- und Löcherstromdichte in den neutralen Gebieten des Halbleiters!

10.1.2 Der Transistor

Abb. 10.5 zeigt die Ladungsträger-Konzentrationsverläufe im Querschnitt eines stromtragenden npn-Transistors:

Die Basis wird auf ein höheres Potential als der Emitter gelegt, der Kollektor auf ein höheres Potential als die Basis. Wie aus den Randbedingungen (Glgn. 10.2) ersichtlich, ergeben sich Minoritätsladungsträgergradienten und, gemäß Glgn. 10.3, ein Elektronenstrom in der neutralen Basis in Richtung auf den Kollektor sowie ein von

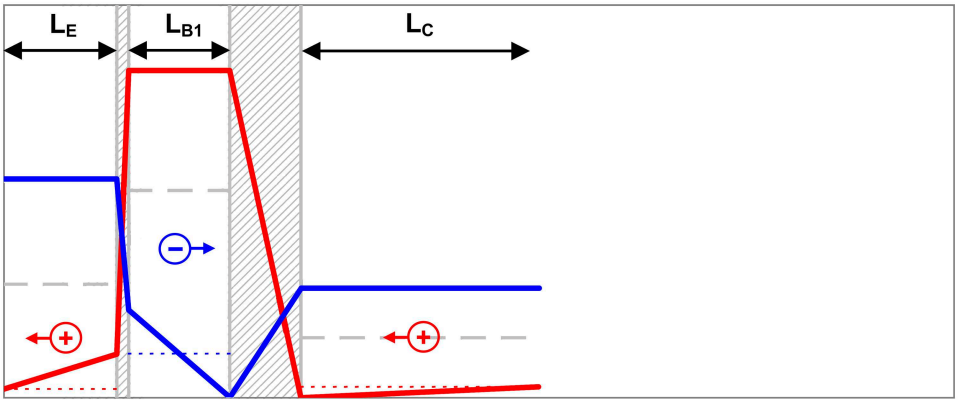


Abb. 10.5: Die Konzentration beweglicher Ladungsträger im Transistor im Betrieb. Dargestellt ist waagrecht die Ortskoordinate entlang eines Schnittes durch den aktiven Transistor und senkrecht die Ladungsträgerkonzentration in logarithmischer Darstellung. Die schraffierten Bereiche kennzeichnen die Raumladungszonen zwischen Emitter (durch L_E gekennzeichnet), Basis (L_{B1}) und Kollektor (L_C). Blau ist die Elektronenkonzentration, rot die Löcherkonzentration dargestellt. Grau strichliert die intrinsischen Ferminiveaus, rot bzw. blau punktiert die Minoritätladungsträgerdichten im thermodynamischen Gleichgewicht.

der Basis weggerichteter Fluß von Löchern im neutralen Emitter. Die Löcher, welche die Basis, sowie die Elektronen, die den Emitter verlassen, müssen nachgeliefert werden. Was geschieht, wenn der „Löcher-Nachschub“, sprich: der Basisstrom, plötzlich unterbrochen wird? Der Konzentrationsgradient von Löchern im Emitter besteht zunächst noch, und mit ihm der Diffusionsstrom. Dieser muß sich nun aus den Löchern am Rand der neutralen Basis speisen. Teile der Basis werden ionisiert, und die Raumladungszone wird weiter. Mit anderen Worten, es wird die (Vorwärts-) Spannung über der Basis-Emitter-Diode abgebaut. Aus physikalischer Sicht ist man auf den Nachschub der aus der Basis abfließenden Löcher angewiesen, um das thermodynamische Ungleichgewicht (=die Potentialdifferenz) und damit den Diffusionsstrom aufrechterhalten zu können.

Nicht weniger gültig erscheint beim Transistor die umgekehrte Betrachtungsweise: Ein Fluß von Löchern in die Basis hinein erzeugt einen Ladungsüberschuß, wodurch die Raumladungszone, von der neutralen Basis ausgehend, weiter mit Löchern aufgefüllt wird. Mit der Verengung der Raumladungszone wird eine Potentialdifferenz generiert, und mithin ein Fluß von Elektronen durch die Basis zum Kollektor. Diese Betrachtungsweise entspricht eher dem elektrotechnischen Bild, ein in die Basis „eingeprägter“ Strom werde durch den Transistor verstärkt.

(Das Verhalten von Kaskode und Thyristor läßt sich jedoch leichter darstellen, geht man von anliegenden Potentialen aus, anstatt von fließenden Strömen.)

10.1.3 Der HBT

Was ändert sich demgegenüber beim Hetero-Transistor? Beim intrinsischen Halbleiter liegt das Fermi-Niveau waagrecht in der Mitte der jeweiligen Bandlücke. Dies legt den Verlauf der Valenz- und der Leitungsbandkanten fest. Die intrinsische Ladungsträgerkonzentration ergibt sich aus dem Abstand der Bandkanten vom Ferminiveau, eine Zunahme der Bandlücke unterdrückt sie exponentiell.

Die Besetzungsdichten der Zustände werden durch den Abstand zum Ferminiveau gegeben. Die Dotierung ändert mithin den Bandkantenverlauf, wobei zunächst die Besetzung den Abstand des Majoritätsladungsträgerbandes von der Fermikante bestimmt, worauf die Bandlücke die Lage und Besetzung des Minoritätenbandes festlegt.

Ändert man mithin in Teilbereichen des Halbleiters den Wert der Bandlücke, werden hiervon (bei Vernachlässigung des Einflußes der stöchiometrischen Variation auf die Zustandsdichte) ausschließlich Lage und Besetzung des Minoritätsbandes beeinflußt. Die intrinsische Ladungsträgerkonzentration geht ein in die Randbedingungen der Glgn. 10.2: Der Wert der Bandlücke beeinflußt die Nichtgleichgewichts–Minoritätsladungsträgerkonzentration, und damit die Diffusionsströme.

Nun ist klar, wie der Konzentrationsverlauf der Majoritäts- und der Minoritätsladungsträger durch den Valenzbandkantensprung zwischen Basis und Emitter beeinflußt werden. Völlig analog zu Abb. 10.5 skizzieren wir die Ladungsträgerkonzentrationen im Querschnitt eines stromtragenden npn-SiGe-HBTs, Abb. 10.6.

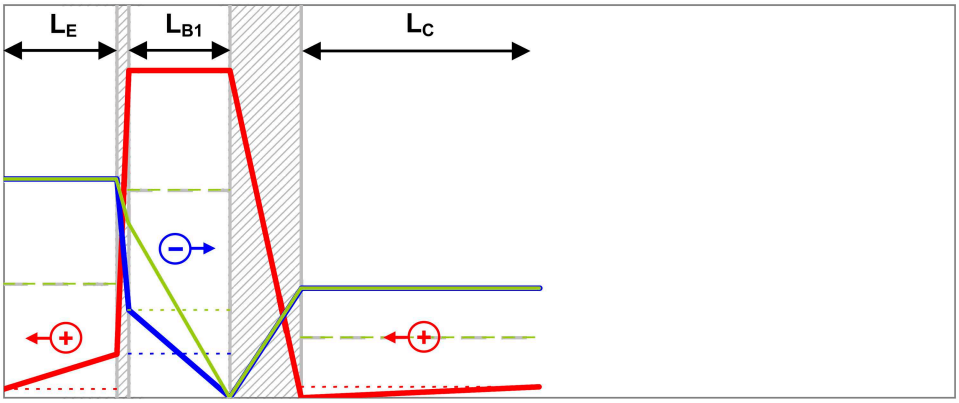


Abb. 10.6: Ladungsträgerkonzentration im HBT im Betrieb.
Wie in Abb. 10.5 sind zunächst, rot, blau und grau die Verhältnisse im Silizium-Transistor wiedergegeben. Im Vergleich dazu in grün die Elektronenkonzentration (durchgezogen), das Fermi-Niveau (strichliert) sowie die intrinsische Elektronendichte (punktiert) im SiGe-HBT.

Während es zunächst offensichtlich erschien, daß durch den Valenzbandkantensprung die Bewegung der Löcher von der Basis in den Emitter behindert wird, sieht man

nun, daß der Löcher-Strom nicht im geringsten beeinflußt wird! (Hierfür wäre ein erheblich höherer Valenzbandkantensprung in der Größenordnung der aufgeprägten Potentialdifferenz, plus etwa $5\,k_BT$, erforderlich.)

Zum Beweis lassen sich zwei Vergleichtransistoren heranziehen, von gleichem Aufbau und gleicher Dotierung, beide aus Silizium, jedoch einmal mit, und das zweite Mal ohne Beimengung von Germanium in der Basis. Die zugehörigen Gummel-Plots zeigen, daß die Basisströme in beiden Bauelementen identisch sind, während der Emitterstrom im SiGe-HBT gegenüber demjenigen im Siliziumtransistor stark erhöht ist. Durch Germaniumeinbau und Bandkantensprünge wird mithin tatsächlich nicht der Fluß von Löchern in den Emitter behindert, sondern umgekehrt der Diffusionsstrom der Elektronen in der neutralen Basis verstärkt [V.7]!

Es sieht also zunächst einmal schlecht aus für mein projektiertes Bauelement – bin ich doch gestartet mit dem Vorsatz, den parasitären pnp-Transistor durch den Einfluß des Valenzbandkantensprunges zu unterdrücken.

Abweichend von der Reihenfolge, in welcher ich hier meine Ergebnisse präsentieren, habe ich natürlich schon früh im Entwicklungszyklus ausführliche physikalische Bauelementsimulationen durchgeführt. Diese zeigen tatsächlich keinerlei Latch-Up-Neigung – wie ist dies zu verstehen?

10.1.4 Kaskode und Thyristor

Gehen wir über zum Querschnitt durch die vertikal integrierte SiGe-Kaskodenstruktur (Abb. 10.7; die Darstellungskonventionen entsprechen denen der Abb. 10.5): Ein linker Transistor aus Emitter (E), Basis 1 (B1) und Zwischenbasisbereich (Z) als Kollektor und ein rechter Transistor aus dem Zwischenbasisbereich als Emitter, Basis 2 (B2) und Kollektor (C). Die Basis B1 ist die Steuerelektrode der Kaskodenstruktur. B2 wird mit einem festen Potential beaufschlagt.

Im Betrieb folgt aus der Nichtgleichgewichtsverteilung der Minoritäten im Zwischenbasisbereich eine Löcher-Diffusion von B2 nach B1. Die aus B2 abfließenden Löcher werden vermittels einer geeigneten externen Quelle nachgeliefert.

Genauso ergibt sich ein Diffusionsstrom von Löchern aus der Basis B1 durch den Emitter. **Dieser** Abfluß von Löchern wird zum Teil durch eine Quelle an der Basis B1 ausgeglichen, zum anderen Teil aber durch den Löcherstrom aus B2 durch den Zwischenbasisbereich.

Kann der Strombedarf der Basis B1 zur Gänze durch den Diffusionsstrom im Zwischenbasisbereich beglichen werden, so läßt sich das Bauelement nicht mehr durch Abtrennen der externen Versorgung an B1 abschalten: Die Steuerelektrode verliert ihre Steuerungsfähigkeit, die Kaskodenstruktur zündet als Thyristor.

Wann ist dieser Betriebszustand zu gewärtigen? Aus Glgn. 10.1 bis 10.3 und dem Bauelementschnitt erhält man sofort die Stromdichten in den Halbleiterbereichen. Ich wähle die anliegenden Potentialdifferenzen alle deutlich größer als die thermische Spannung $V_{th} \approx 0.0259\,\mathrm{V}$. Unter der Einschränkung auf $L_{B1} \cdot N_{B1} = L_{B2} \cdot N_{B2} \equiv$

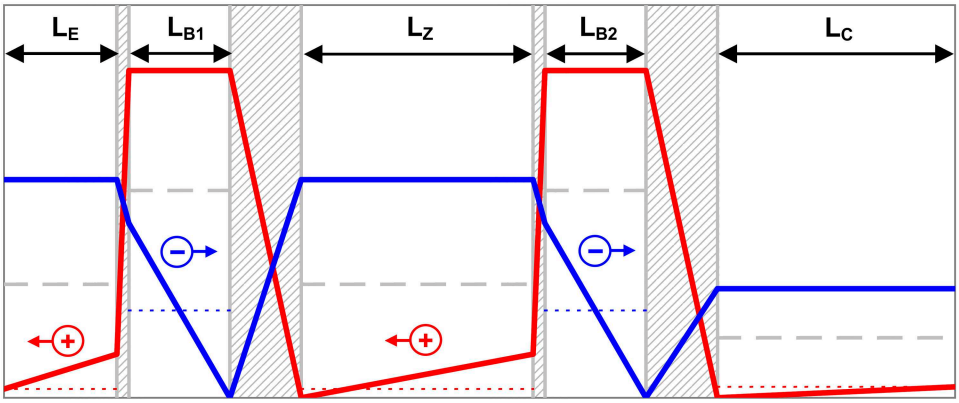


Abb. 10.7: Ladungsträgerverlauf in der vertikal integrierten Kaskodenstruktur im aktiven Vorwärtsbetrieb (Darstellunsgkonventionen wie in Abb. 10.5).

$L_B \cdot N_B$ folgt aus der Kontinuität $j^{\,n}_{B1} = j^{\,n}_{B2}$, und so $V_{B1\,E} = V_{B2\,Z}$. Die Ausdrücke für die Diffusionsströme vereinfachen sich dann mit dem Kürzel $A = q \exp\left(\frac{V_{B1\,E}}{V_{th}}\right)$:

$$\begin{aligned} J^{\,p}_{E} &= -A\tfrac{D_p \cdot n^2_{i,Si}}{L_E \cdot N_E} \quad , \quad J^{\,p}_{Z} = -A\tfrac{D_p \cdot n^2_{i,Si}}{L_Z \cdot N_Z} \quad , \\ J^{\,n}_{B1} &= +A\tfrac{D_n \cdot n^2_{i,SiGe}}{L_B \cdot N_B} \quad , \quad J^{\,n}_{B2} = J^{\,n}_{B1} \quad , \\ J^{\,p}_{C} &\approx 0 \quad . \end{aligned}$$

(L_x, N_x stehen für Länge und Dotierstoffkonzentration des nicht-ausgeräumten Anteils von Halbleiterbereich x, V_{xz} für die Spannung zwischen den Halbleiterbereichen x und z, $j^{\,y}_{x}$ für den Diffusionsstrom der y-Ladungsträger ($y = n, p$) in Halbleiterbereich x sowie q für den Betrag der Elementarladung.)

Es folgt sofort für die Verstärkungen $\beta^{\,n}_{E,B1,Z}$ des npn- und $\beta^{\,p}_{B2,Z,B1}$ des pnp-HBTs,

$$\begin{aligned} \beta^{\,n}_{E,B1,Z} &= \tfrac{D_n}{D_p}\left(\tfrac{n_{i,SiGe}}{n_{i,Si}}\right)^2 \tfrac{L_E\,N_E}{L_B\,N_B} \propto \exp\left(+\tfrac{\Delta\,E_G}{k_B\,T}\right) \\ \beta^{\,p}_{B2,Z,B1} &= \tfrac{D_p}{D_n}\left(\tfrac{n_{i,Si}}{n_{i,SiGe}}\right)^2 \tfrac{L_B\,N_B}{L_Z\,N_Z} \propto \exp\left(-\tfrac{\Delta\,E_G}{k_B\,T}\right) \end{aligned}$$

Man sieht, daß die Verstärkung des pnp-Transistors tatsächlich durch den Hetero-Übergang verringert wird. Dennoch hat mich meine Anschauung getrogen: Der Grund für die geringe Verstärkung des pnp ist keine Valenzbandbarriere. Schneidet man in Gedanken den pnp-Transistor frei, wird klar, daß der Elektronenstrom in der Kaskode als Basisstrom des pnp-Transistors gelten muß (Betriebsrichtung von links nach rechts) . Als Diffusionsstrom in B2 ist er proportional zu $n^2_{i,SiGe}$ – nicht aufgrund eines

verringerten Löcher-, sonderen eines erhöhten Elektronenstromes sinkt die Verstärkung des pnp mit wachsendem Germaniumgehalt!

Der Basisstrom gleicht den Diffusionsstrom im Emitter aus; er ist hiernach proportional zur Ladungsträgerkonzentration im Emitter und unabhängig vom Germaniumgehalt in der Basis!

Ist das jetzt die ganze Miete? Nein: Wir haben gesehen, daß B1 im Thyristorbetrieb ihre Steuerungsfähigkeit verliert, wenn der Diffusionsstrom im Zwischenbasisbereich den gesamten Basisstom für B1 zur Verfügung stellen kann. Wie aus Abb. 10.7 sofort ersichtlich, ist dies der Fall, wenn (bei gleicher Dotierung) der Emitter dicker ausfällt als der Zwischenbasisbereich. In diesem Fall herrscht im Zwischenbasisbereich ein höherer Löcher-Gradient, und in die Basis B1 werden von rechts mehr Löcher nachgeliefert, als wir nach links abführen.

Dies ist die wichtige Richtschnur für das Bauelementdesign, nicht der Germaniumgehalt! Betrachtet man den Thyristor anhand seines npn-/pnp-Ersatzschaltbildes, zündet er offensichtlich, wenn die Schleifenverstärkung

$$\beta^{\,n}_{E,B1,Z} \cdot \beta^{\,p}_{B2,Z,B1} = \frac{L_E \cdot N_E}{L_Z \cdot N_Z} \tag{10.4}$$

größer wird als Eins. Dies bedeutet aber gerade $L_E \cdot N_E \geq L_Z \cdot N_Z$!

Meine vertikal integrierte Kaskodenstruktur profitiert hiernach von den üblichen Vorteilen des Germaniumeinbaus in der Basis (vgl. S. 4), ein etwaiges thyristorähnliches Verhalten wird jedoch – auch in Gegenwart von Germanium – einzig durch das Verhältnis der Gummelzahlen von Emitter und Zwischenbasisbereich bestimmt!

Nachdem über das Wirkprinzip der vertikalen Integration Klarheit herrscht, kehre ich zur Prozeßarchitektur zurück.

11 Realisierungsmöglichkeiten

Aufbauend auf Vorschlägen zur Verbesserung der HBT-Technologie skizziere ich in diesem Kapitel, wie sich die vertikal integrierte Kaskodenstruktur als optionales Modul innerhalb einer verbesserten SiGe-Technologie zur Verfügung stellen läßt.

11.1 Der SiGe2-Standardprozeß

Ich rufe anhand der Querschnittbilder des SiGe2-Transistors in unterschiedlichen Bearbeitungsstufen (Abb. 11.1) den Fertigungsablauf bei der Siliziderung der äußeren Basis noch einmal ins Gedächtnis:

Nitridblock: Auf die Epitaxie der p-Basis- und n-Emitterschichten folgt die Definition eines Nitridblocks als „Platzhalter“, wo später der Emitter entstehen soll. Nitridblock und Silizium sind durch ein dünnes Oxid getrennt. Zum einen läßt das Oxid die Spannungen der Nitridschicht abgleiten, zum zweiten dient es als Ätzstop und zur Endpunktdetektion während der Trockenätzung des Nitrids.

Über den Nitridblock wird eine Oxidschicht abgeschieden und trockenchemisch (in einem zeitgesteuerten Ätzprozess) zurückgeätzt. Die Abscheidung erfolgt isotrop, die Rückätzung anisotrop, so daß am Rand des Nitridblocks ein Oxidrest, ein sogenannter Spacer, verbleibt. Der Spacer dient bei der nachfolgenden Implantation als Abstandhalter zum Nitridblock (Abb. 11.1 obere Skizze).

Basisanschluß: Nach Implantation des äußeren Basisanschlußgebietes und naßchemischem Entfernen des Implantationsoxides wird Titan abgeschieden und mit dem Silizium zur Reaktion gebracht (vgl. Abschnitt 5.3, S. 25).

Isolationsoxid: Es folgt ein Verfahrensschritt, der die Isolation von Basis- und Emitteranschluß gegeneinander sicherstellt: Nach Abscheidung einer dicken Oxidschicht wird das Oxid über den Nitridblöcken mittels einer geeignet skalierten, zur Nitridblock-Maske inversen Lackmaske per Trockenätzen entfernt (zweites Querschnittsbild) und die verbleibenden „Oxidkämme“ an den Rändern der Nitridblöcke per CMP planarisiert, wobei der Nitridblock als Polierstop dient.

Emitteranschluß: Die Oberfläche ist nun eben. Mit kochender Phosphorsäure werden die offenliegenden Nitridblöcke entfernt. In die nach Entfernen der nitridischen Opferstruktur verbleibenden Hohlräume werden L-förmige Innenspacer und der Poly-Emitter eingebracht. Anschließend wird hoch n-dotiertes Polysilizium abgeschieden, silizidiert und strukturiert. Nach der Strukturierung reicht das Polyzid zu beiden Seiten des Emitterblocks bis auf das Isolationsoxid hinaus (Querschnitt drei).

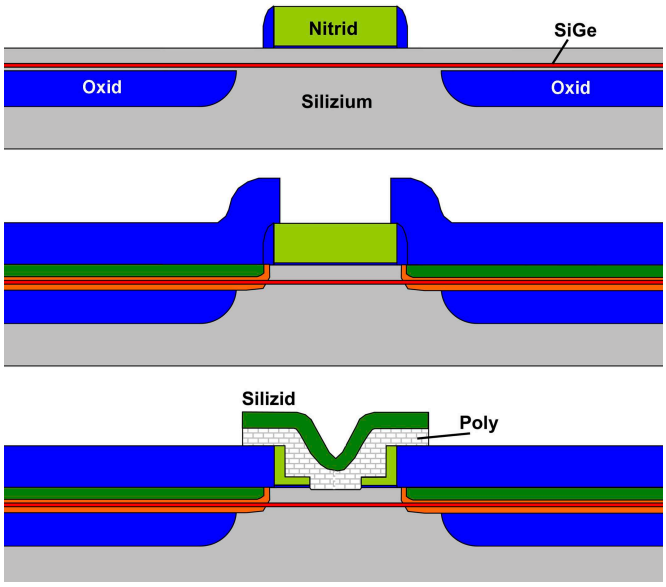


Abb. 11.1: Fertigungsablauf Basisanschluß SiGe2

11.2 Bekannte Alternativen beim Basisanschluß

Bei der Betrachtung des SiGe2-Prozeßablaufes kommt man immer wieder zurück auf die Dicke der abgeschiedenen Emitterschicht: Während in der Basisanschlußregion ausreichend Fleisch für die Silizidierung vorgehalten werden muß, schadet die „dicke“ Randdiode im Emittergebiet. Die Mitbewerber gehen bei Silizidierung und Basisanschluß entsprechend teils andere Wege [V.8]:

Bei der epitaktischen Siliziumabscheidung werden Prekursorengase an der heißen Waferoberfläche gespalten. Die Reaktion kann in beide Richtungen erfolgen[42]. Je nach den Anteilen von Silan und Wasserstoff überwiegen Abscheidung oder Rückätzung der offenliegenden Siliziumflächen. Die Epitaxie nahe am Gleichgewicht zwischen Abscheidung und Abtragung zu fahren, ist Voraussetzung, daß die abgeschiedene Schicht in einer hohen Qualität aufwächst. Bringt man das Schichtwachstum fast zum Stillstand, widersteht das Silizium nur im einkristallinen Verbund der Abtragung: Auf Oxidflächen kann kein Polysilizium aufwachsen („selektive Epitaxie“).

Unter Verwendung selektiver Epitaxie gelangt man leicht zu eleganten Technologiearchitekturen. Das Verfahren selbst ist jedoch aufwendig, langsam und teuer. Ich schlage einen einfacheren Weg vor, um Silizium – und letztlich Silizid – ausschließlich neben den Nitridblöcken aufzubringen. Den Prozeßablauf illustriere ich anhand der Querschnitte in Abb. 11.2.

---

[42] Derselbe Berthollet, der bei Alexandr Grin in [V.1] vorkommt, hat 1803 in seinem »Essai de statistique chimique« zuerst auf diesen Aspekt chemischer Reaktionen hingewiesen.

11.3 Ein verbessertes Basisanschlußmodul

11.3.1 Neues Silizidmodul (Ⓟ 5, Ⓟ 7)

Nach Definition der Nitridblöcke wird eine Doppelschicht aus p^+-α-Silizium und Oxid abgeschieden. Beide Materialien lassen sich mit einer Inversmaske trockenätzen und einfach per CMP planarisieren. Nach der Planarisierung liegt seitlich der Nitridblöcke α-Si offen (linkes Teilbild in Abb. 11.2): Nach Abscheiden des Emitteranschlusses ergäbe sich ohne weitere Maßnahmen ein Basis-Emitter-Kurzschluß.

Während sich das α-Si prinzipiell (leidlich, aufgrund der schrägen Kanten) per Trockenätzen eintiefen läßt, stehen hierfür weder ein Endpunkt noch ein Ätzstop zur Verfügung: Es ist zunächst keine Ätzung bis in eine definierte Tiefe möglich.

Ein einfaches Verfahren, Schichten auch ohne Ätzstop oder Endpunktsignal bis in eine wohldefinierte Tiefe abzutragen, lege ich in Ⓟ 5 offen. Auf unser Beispiel übertragen, wird nach der Oxid/Poly-CMP eine Titanschicht wohldefinierter Dicke abgeschieden und mit dem offenliegenden α-Si zur Reaktion gebracht (rechtes Teilbild). Titan, Titansilizid und Oxid werden selektiv entfernt, es verbleibt eine Siliziumschicht ausschließlich seitlich neben den Nitridblöcken!

Die selektive Abtragung erfolgt naßchemisch: $2H_2SO_4 : H_2O_2$ ätzt Oxid mit einem Zehntel, Nitrid und Polysilizium mit einem Tausendstel der Titanätzrate und entfernt, nach meinen Versuchen, auch TiN. Zur selektiven Entfernung von Titan ist auch $NH_4OH : H_2O_2 : 5\ H_2O$ geeignet.

Wie sich zeigen wird, sind bei diesem Prozeß keine Oxid-Spacer an den Außenkanten der Nitridblöcke erforderlich. Nach Entfernen des nicht reagierten Titans wird das offenliegende Oxid trockenchemisch entfernt.

Das Titansilizid wird bspw. durch Kochen in HCl oder Tauchen in $NH_3 : H_2O_2$ selektiv entfernt. Die nun offenliegende Polysiliziumschicht läßt sich nach Standardverfahren silizidieren und dotieren, die weitere Prozessierung folgt dem SiGe2-Ablauf.

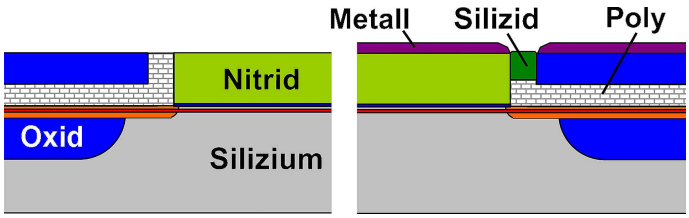


Abb. 11.2: Querschnitt-Skizzen zum Silizidmodul

11.3.2 Neues Emittermodul

Die abgeänderte Prozeßführung bei Herstellung des Basisanschlusses verschafft Bewegungsfreiheit zur Überarbeitung des Emitters. Ich schlage vor, die Emitterschicht nur 15 nm dünn als niedrig **p**-dotiertes Silizium aufzuwachsen sowie die äußere Basis durch Diffusion mit dem Silizid als Dotierstoffquelle zu definieren[43].

Bei der Haftoxidation vor Nitridabscheidung, bei Bildung des Implantationsoxides auf der äußeren Basis sowie (im Emitterfenster) beim Öffnen der L-Spacer werden jeweils einige Nanometer der p-dotierten Schicht abgetragen. Unterhalb des silizidischen Basisanschlusses wird eine evtl. verbleibende dünne p^--Zone durch Ausdiffusion aus dem borübersättigten Polyzid des Basisanschlusses geschlossen.

In das Emitterfenster wird hoch n-dotiertes α-Si abgeschieden. Bei dem Temperschritt zur Emittersilizidierung nimmt dieses in Teilen die Kristallstruktur des Mono-Emitters an. Gleichzeitig diffundieren Donatoren in die p^--Schicht. Ein etwaige verbleibender p^--Rest beeinträchtigt aufgrund der niedrigen Dotierung die Basisweite nicht, reduziert aber die Basis-Emitter-Kapazität.

Abb. 11.3 skizziert den Querschnitt bei abgeändertem Silizid- und Emittermodul:

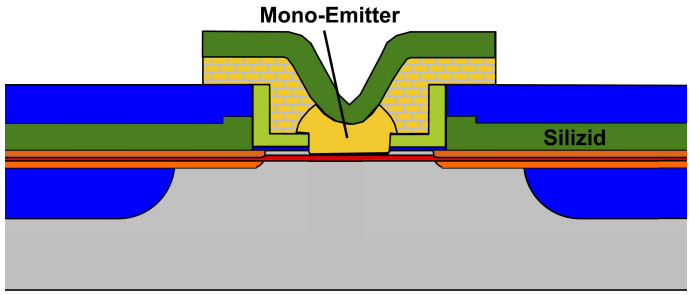


Abb. 11.3: Der HBT nach dem vorgeschlagenen Herstellungsablauf

Der wesentliche Vorteil des vorgeschlagenen Emitteraufbaus ist das vollständige Verschwinden der parasitären Diode zwischen äußerer Basis und Emitter. Bei der durch den neuen Herstellungsablauf ermöglichten geringen Dicke der Silizium-Emitterschicht verringert sich auch die laterale Streuung bei der Implantation der äußeren Basis: Man kann auf die Oxidspacer an den Nitridblöcken verzichten.

Dies erleichtert ein naßchemisches Entfernen des Abdeckoxides aus Abb. 11.2 – die dünnen Spacer-Hohlräume, die andernfalls nach der Naßchemie verbleiben würden, wären nur schwer mit TEOS oder BPSG aufzufüllen – und gestaltet das vorstehend beschriebene Silizidierungsmodul einfacher und wirtschaftlicher.

[43] Auf diese Möglichkeit, hat mich Herr Peter Brandl aufmerksam gemacht. Soweit ich weiß, geht sie letztlich auf eine Prozeßidee aus dem IHP in Frankfurt/Oder zurück.

11.4 Erste Gehversuche in vertikaler Integration

Versucht man, vertikal integrierte Kaskoden mit den Mittel von SiGe2 herzustellen, gelangt man zu dem in Abb. 11.4 skizzierten Prozeßablauf [V.9]: Auf eine erste SiGe-Epitaxie mit Silizium-Starterschicht, Basis- und Emitterabscheidung wird eine Oxidmaske aufgebracht und strukturiert. Es folgt die zweite SiGe-Epitaxie.

Nach Defninition einer Nitrid- und einer Lackmaske – den Bearbeitungszustand zeigt das erste Teilbild der Abb. 11.4 – wird per Trockenätzung strukturiert. Die große offene Fläche des Grabenoxides ermöglicht die Endpunktdetektion, die Oxidmaske schützt die Basisanschlußregion der unteren Basis vor der Abtragung (zweites Teilbild). Nach Entfernen der Lackmaske und Aufbringung von Oxidspacern werden die Basisanschlußgebiete implantiert (Teilbild 3).

Eine offensichtliche Engstelle des Prozeßablaufes ist in Teilbild 3 violett eingekreist: Bei der Epitaxie wächst über den vergrabenen Oxiden polykristallines Material auf, schraffiert dargestellt. Gesperrte pn-Übergänge kommen in polykristallinen Bereichen zu liegen – das Bauelement kann so nicht funktionieren!

Dieses Problem läßt sich – weitgehend – vermeiden, wählt man die Oxidmaske so dick, daß die Basisanschlußimplantation sicher im Oxid stoppt (Teilbild 4). Auch dieser Ablauf hat seine Tücken: Zum einen werden, aufgrund der vorzuhaltenden Justiertoleranzen, die Randanteile des Bauelements viel zu groß. Zum zweiten sinkt in den in Teilbild 4 eingekreisten Bereichen die Gummelzahl des Parasiten, es erscheint schwierig, ein Zünden des Thyristors sicher auszuschließen.

Das weiter oben vorgestellte verbesserte Basisanschlußmodul für SiGe2, bei welchem eine dünne äußere Basis aus Silizium per Diffusion aus einer aufliegenden Silizidschicht dotiert wird, vermeidet eine Ausbreitung des p-dotierten Anschlußbereiches der oberen Basis in die Zwischenbasischicht, und mithin ein Zünden des Thyristors.

Auch mit den dort entwickelten Methoden äußerst anspruchsvoll gestaltet sich dagegen der Anschluß der unteren Basis. Von den verschiedenen geometrisch möglichen Arrangements scheint der Anschluß **von unten** die eleganteste Technologiearchitektur zu ermöglichen. Hier ist erforderlich, den Basisanschluß bereits **vor** der SiGe-Epitaxie einzubringen.

Dafür bedarf es zweier bei Atmel so nicht zur Verfügung stehender Einzelprozesse: dem Aufbringen eines hochtemperaturfesten Silizids sowie der lateralen Festphasenepitaxie von Silizium. Beide Einzelprozessen werde ich nachfolgend anhand der Literatur auf ihre Umsetzbarkeit hin abprüfen.

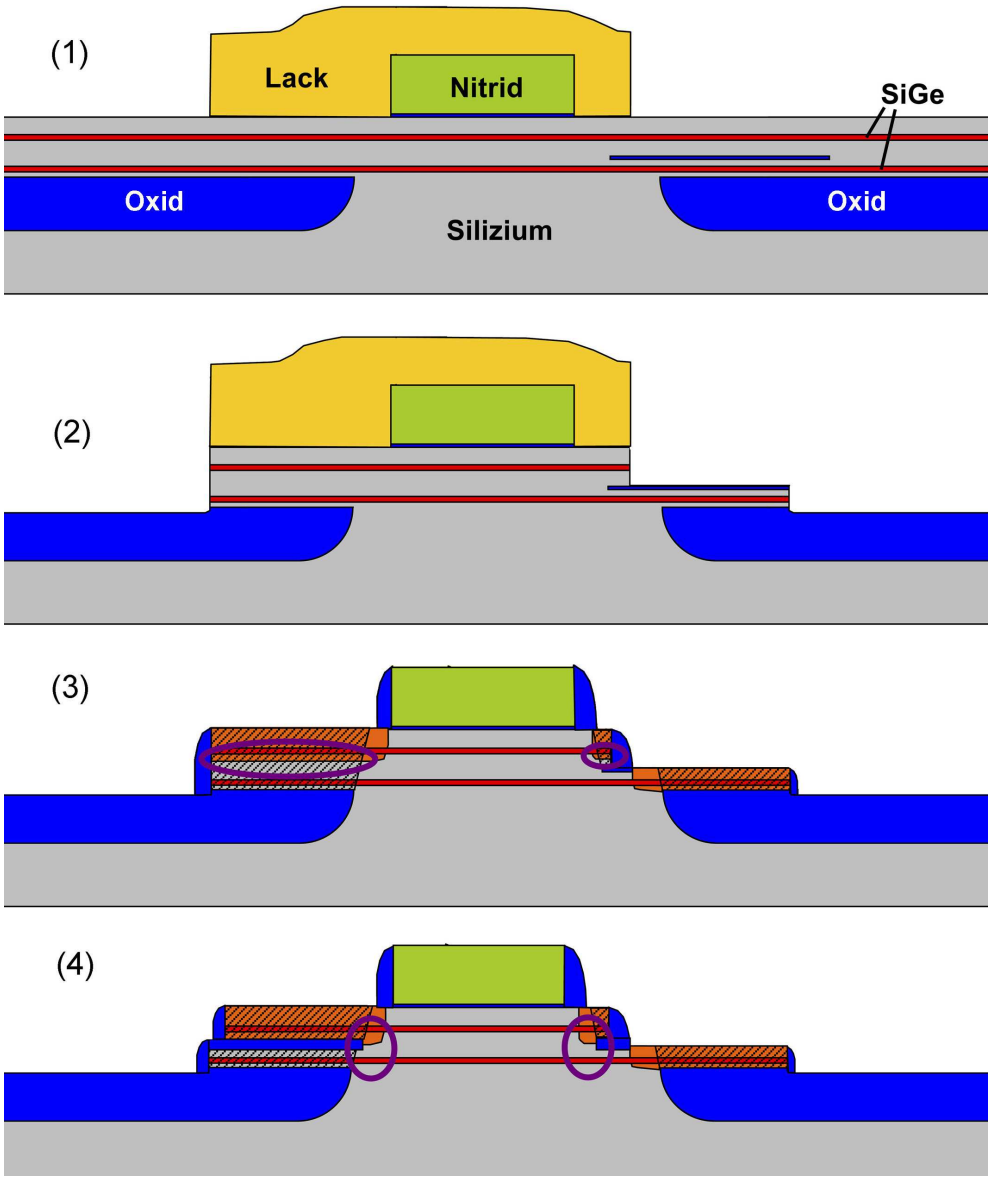


Abb. 11.4: Vertikal integrierte Kaskode, erster Anlauf

11.5 Silizide

11.5.1 Temperaturbeständigkeit der Silizide

„Im Prinzip“ sind viele Silizide hochtemperaturfest. Manche, wie $NbSi_2$ [V.10] oder $MoSi_2$ [V.11] dienen gar als keramische Hochtemperaturwerkstoffe! Für die Halbleiterei gelten jedoch andere Bedingungen: So ist bei Nickel ein Monosilizid die niederresistive Form. In Kontakt mit Silizum reagiert es oberhalb von 750°C zu höherresistiven Verbindungen. Andere, wie $MoSi_2$ und WSi_2, oxidieren bei erhöhten Temperaturen. Viele Disilizide agglomerieren unter Temperaturbeaufschlagung (ein Beispiel ist $TiSi_2$): Silizium breitet sich entlang der Silizid-Korngrenzen aus, die Silizide verinseln und der Leitfähigkeitsgewinn durch die Silizidauflage geht verloren.

In manchen Fällen läßt sich die Agglomeration unterdrücken. So wächst $CoSi_2$ auch heteroepitaktisch auf Silizium[44]. Ohne Korngrenzen keine Agglomeration: Während der Widerstand polykristalinen Kobaltsilizids bei 750°C anzusteigen beginnt [V.24], übersteht epitaktisches $CoSi_2$ 950°C [V.25][45]. Dies ist die höchste Temperaturbeständigkeit unter den Standard-Siliziden $TiSi_2$, $CoSi_2$ und NiSi[46].

Tabelle 11.1 listet, angelehnt an die Übergangsmetallreihe des Periodensystems, Silizide mit ihren Bildungstemperaturen T_F und den Stabilitätsgrenzen T_S (in °C) sowie den spezifischen Widerständen der niederresisitiven Phasen (in µΩcm) auf. Die eingeschränkt nutzbaren Silizide sind ausgegraut.

Aus dem gesamten Silizidbildner-Zoo verbleiben lediglich vier mögliche hochtemperaturbeständige Silizide, **$MoSi_2$**, **WSi_2**, **$TaSi_2$** und **HfSi**:

---

[44] Scheidet man zur Bildung von Kobaltsilizid zunächst eine dünne Titanschicht ab [V.12], [V.13], bildet dieses mit dem Sauerstoff aus dem Restgas der Anlage Oxide [V.14]. Das Oxid dient als Diffusionsbarriere für das Kobalt. Die Reaktion mit Silizium erfolgt langsam und ergibt eine heteroepitaktische $CoSi_2$-Schicht. Auch Wolfram eignet sich als Diffusionsbarriere [V.15]. Die Barriereschicht schwimmt auf und kann im Anschluß an die Silizidierung naßchemisch selektiv entfernt werden [V.16].

Auch lassen sich epitaktische Schichten in einem RTA-Verfahren erzeugen, wenn die Kobaltschicht vor der Silizidierungsreaktion mit einer dünnen Schicht Titan oder Titannitrid bedeckt worden ist [V.17], [V.18].

Weitere Silizide, die sich heteroepitaktisch auf <100>-Silizium aufbringen lassen, sind $FeSi_2$ [V.19], $ReSi_2$ [V.20], $CoSi_2$ [V.21], $ErSi_2$ [V.22] sowie, für <111>-Si, $CrSi_2$ [V.23] (Die Verwendung von Eisen in einer Silizium-Fab sollte man dabei sehr vorsichtig abwägen...)

[45] Kobalt reagiert oberhalb von 600°C, analog zu Titan, mit Si_3N_4 zu CoN und CoSi [V.26]. Wie bei Titan sollte auch hier die Silizidierung der germaniumhaltigen Basis vermieden werden [V.27]. Schließlich ist es schlecht möglich, Kobaltsilizid trockenzuätzen [V.28]: Während sich $TiSi_2$ leicht in flüchtige Reaktionsprodukte wie TiF oder $TiCl_2$ überführen läßt, ist $CoSi_2$ chemisch träge und besitzt keine gasförmigen Abbauprodukte [V.29].

[46] Auf Anwendung und Technologie der drei Arbeitspferde **NiSi**, **$CoSi_2$** und **$TiSi_2$** gehen das Buch [V.30] oder der Übersichtsartikel [V.31] ausführlich ein.

	$TiSi_2$ T_F=650 T_S<900 ρ=13··25	VSi_2	$CrSi_2$ T_F=450 ρ=60	$MnSi_2$	$FeSi_y$	**$CoSi_2$** T_F=450 T_S<750 (epi: 950) ρ=18··25	**NiSi** T_F=400 T_S<750(?) ρ=10··15
$ErSi_2$	$ZrSi_2$ T_F=600	$NbSi_2$	**$MoSi_2$** T_F=1000 T_S<1000 ρ=100	$TcSi_x$	$RuSi_y$	RhSi T_F=300	Pd_2Si T_F=200
	HfSi T_F=550 T_S<1000 ρ=45··50	**$TaSi_2$** T_F=1000 T_S<1000 ρ=35··38	**WSi_2** T_F=650 T_S<1000 ρ=70	$ReSi_y$	$OsSi_y$	IrSi T_F=300	$PtSi_2$ T_S<750 ρ=28··35

Tabelle 11.1: Übersicht über die Silizide: Bildungstemperatur T_F, Stabilitätsgrenze T_S in °C sowie spezifischer Widerstand ρ in $\mu\Omega$ cm

- Die Hälfte der Silizide ist halbleitend (Silizide von Cr, Mn, Fe, Ru, Re, Os, Ir, Mg, Ca, Ba [V.32]) und für unsere Zwecke weniger geeignet[47].
- Wenige Daten liegen im Zusammenhang mit der Halbleiterei für die Silizide von Zr, V, Nb sowie, selbstredend, Tc, vor.
- Die Silizide von Rh, Pd, Ir, Pt sowie Zr haben auf Silizium eine ungenügende Temperaturstabilität [V.35].
- Erbium ist in Silizium ein tiefer Donator.

In oxidierender Atmosphäre wird Hafnium-Disilizid bei 700°C explosionsartig in das Silikat umgewandelt; es ist nur eingeschränkt nutzbar.

Das Oxidationsverhalten der drei verbleibenden temperaturbeständigen Silizide hängt vom Untergrund ab. Auf SiO_2 reagieren $MoSi_2$, $TaSi_2$ und WSi_2 zu ihren Oxiden:

- $MoSi_2$ bildet oberhalb von 800°C in oxidierender Atmosphäre MoO_3. Das Oxid ist oberhalb von 900°C flüchtig [V.36].
- WSi_2 weist eine geringere Stabilität gegenüber Sauerstoff auf als $MoSi_2$ [V.37].
- $TaSi_2$ oxidiert zu SiO_2 und Ta_2O_5 [V.38].

[47] Halbleitende Silizide sind von potentiellem Nutzen als IR-Photodetektoren auf Siliziumbasis, da sie eine kleine, direkte Bandlücke aufweisen ([V.33] und [V.34]).

Liegen diese Silizide bei der Oxidation aber auf Silizium auf, diffundiert das Silizium aufgrund seiner höheren Sauerstoffaffinität durch die Silizide hindurch, um an der Silizidoberfläche zu SiO_2 zu reagieren. Die Silizide selber werden nicht beeinflußt [V.39], [V.40], [V.41]. Dieses Verhalten spiegelt sich im ternären Phasendiagramm [V.42] wider:

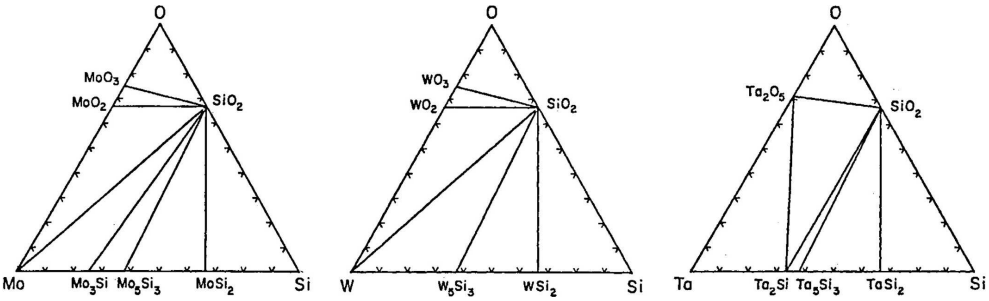


Abb. 11.5: Ternäre Phasendiagramme

Phasen, die mit einer Linie verbunden sind, sind gegeneinander chemisch stabil. Überschreitet man eine solche „tie-line“, findet eine Reaktion der beteiligten Konstituenten statt: So läßt sich $MoSi_2$ beliebig mit SiO_2 mischen, ist aber nicht stabil gegenüber MoO_2: Verändert man von $MoSi_2$ nach MoO_2 das Verhältnis Si /O, schneidet man die Linien, auf denen Mo_5Si_3, Mo_3Si bzw. Molybdän selbst mit SiO_2 im Gleichgewicht stehen. Es findet eine Reaktion statt, bei der **der Silizium-Partner aus dem Silizid abgereichert und oxidiert** wird. Erst jenseits von $(MoSi)O_2$ wird das Silizium vom Molybdän aus der Bindung mit dem Sauerstoff verdrängt.

Mit anderen Worten: Es wird bevorzugt Silizium oxidiert, und Molybdän erst „verstoffwechselt“, wenn kein Silizium zur Verfügung steht. In Gegenwart eines Siliziumreservoirs wird auf $MoSi_2$ zwar Quarz gebildet, das abgereicherte Silizium jedoch aus der darunterliegenden Schicht nachgeliefert.

Bei einer sauerstofffreien Prozeßführung sind alle vier Hochtemperatursilizide auf Silizium wie auf Oxid bis 1000°C stabil.

Schließlich ist die hohe Bildungstemperatur von $TaSi_2$ und $MoSi_2$ (1000°C) einschränkend zu erwähnen[48]: Beide Silizide werden durch Sputtern oder CVD-Verfahren, nicht jedoch durch eine in-situ-Reaktion, auf Silizium aufgebracht [V.43].

11.5.2 SADS, Silicide as diffusion source

Der Übergang zwischen dem quasi-metallischen Silizid und einem moderat dotierten Halbleiter ist vom Schottky-Typ. Die an das Silizid angrenzenden Siliziumbereiche

[48] Folgen wir der Mehrheitsmeinung in meinem persönlichen Umfeld, gewissermaßen: Die Latrinenparole als wissenschaftliche Methode. Abweichend gibt Sze [III.2] (S. 292) eine $TaSi_2$-Bildungstemperatur von 750°C an.

müssen also bis an die Aktivierungsgrenze dotiert werden, um einen metallischen Kontakt zu erzielen.

Bei einer hochdosigen Implantation in das Silizium wird das Kristallgitter geschädigt. Es muß im Anschluß an die Implantation regeneriert werden. Die Ausheilung läßt sich umgehen, verwendet man eine altertümliche Diffusion, um Dotierstoffe in das Silizium einzubringen. Tatsächlich kommen manche Silizide als Dotierstoffquelle in Betracht.

So gibt $CoSi_2$ Bor und Arsen mit einer hohen Quellstärke an Silizium ab [V.44], [V.45], [V.46], [V.47], [V.48]. Bei anderen Silizid/Dotiersoff-Paarungen gilt dies nicht. So bindet $TiSi_2$ Bor in Form von TiB_2 fest an sich.

Die Quellstärke eines Silizides als Dotierstoffquelle wird durch zwei Faktoren begrenzt: Durch die unterschiedlichen Affinitäten der Dotierstoffe zu Silizid bzw. Silizium, sowie durch die Rate, mit welcher ein Silizid die oberflächliche Abreicherung der Dotanden aufgrund der Segregation ins Silizium ausgleichen kann, sprich: durch die Beweglichkeit des Dotierstoffes im Silizid.

In Abwesenheit von Daten zur relativen Affinität betrachte ich nur die Diffusionsraten und habe, ausgehend von [V.42], S. 303 ff., die Diffusionslängen der Dotierstoffe in den Siliziden abgeschätzt (Tabelle 11.2; die Zahlenwerte gelten für eine Diffusion von 20 s Dauer bei der Stabilitätsgrenze T_S).

Silizid	T_f	T_s	Fremdstoff	$\sqrt{Dt}$ [nm] nach 20 sec.@T_s
$TiSi_2$	650°C	900°C	B	36
			P	2000
			As	2828
$CoSi_2$	450°C	950°C	B	4743
			P	1917
			As	899
$MoSi_2$	1000°C	1000°C	B	253
			P	?
			As	109
$TaSi_2$	1000°C	1000°C	B	30
			P	66
			As	74
WSi_2	650°C	1000°C	B	6982
			P	5341
			As	4361

Tabelle 11.2: Diffusionslängen typischer Siliziumdotanden in Siliziden

$TiSi_2$:B ist als Dotierstoffquelle ungeeignet, denn Bor diffundiert in Titandisilizid lediglich 36 nm weit. $CoSi_2$ eignet sich für Bor und Arsen als Dotierstoffquelle, und tatsächlich diffundieren diese 4700 bzw. 900 nm durch das Kobaltdisilizid.

Tabelle 11.2 können wir entnehmen, daß $MoSi_2$ und $TaSi_2$ als ergiebige Diffusionsquellen ausscheiden, während die diffusionsoffenen Silizide $CoSi_2$, WSi_2 (und für Donatoren auch $TiSi_2$) möglicherweise als Feststoffquellen in Frage kommen.

Im Vergleich zwischen $CoSi_2$ und WSi_2 sprechen die Bildungstemperatur (450°C statt 650°C) und der spezifische Widerstand (20 µΩcm statt 70 µΩcm) eher für $CoSi_2$. Erst

zwischen 750°C und 1000°C muß WSi_2 an seine Stelle treten.

Wolfram hingegen besitzt flüchtige Verbindungen, wie Wolframhexaflourid, WF_6. Anders als für $CoSi_2$ stehen für WSi_2 daher Trockenätzverfahren zur Verfügung. Beispielsweise stellt [V.49] einen einfachen Trockenätzprozess für WSi_2 und Silizium mit guter Selektivität zu Oxid (und vermutlich Nitrid) vor.

11.5.3 Einschub: p-Dotanden in Silizium

Während das vorgeschlagene Emittermodul die laterale Skalierbarkeit sicherstellt, ist noch eine Anmerkung zur vertikalen Skalierung angebracht. Bei einer Basisweite deutlich unter den aktuell verwendeten etwa 20 nm gerät man in die Zwickmühle: Will man einen Basisanschluß per SADS niederohmig ausformen, nachdem die SiGe-Schicht abgeschieden ist, so soll der Dotierstoff aus dem Silizid, nicht jedoch aus der Basis ausdiffundieren.

Ich habe, ausgehend von den Literaturwerten der Diffusionskonstanten, für die p-Dotanden Diffusionsprofile berechnet (außer Thallium): Abb. 11.6 gibt die Diffusionsweite nach einer Stunde auf Zieltemperatur wieder, bezogen auf Bor als Standard.

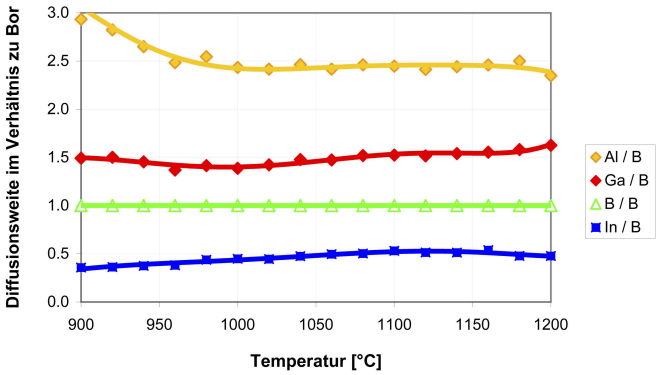


Abb. 11.6: Diffusionsweite im Verhältnis zu Bor

Mit Gallium und Aluminium stehen durchaus agile, mit Indium auch ein diffusionsträger Akzeptor zur Verfügung. Weswegen werden sie kaum eingesetzt?

Der Grund ist ihre geringe Aktivierbarkeit. Bei 1100°C ist Gallium im Gleichgewicht nur bis zu einer Konzentration von $4{\cdot}10^{19}\,cm^{-3}$ [V.50], Aluminium bis $4{\cdot}10^{18}\,cm^{-3}$ [IV.2], S. 233, Indium sogar nur bis $1{\cdot}10^{17}\,cm^{-3}$ [V.51] aktivierbar. Die gewünschte Bor-Konzentration in der SiGe2-Basis liegt, zum Vergleich, bei $3{\cdot}10^{19}\,cm^{-3}$, im Basisanschlußgebiet oberhalb von $1{\cdot}10^{20}\,cm^{-3}$.

In der Basis eines SiGe-HBT kann es vorteilhaft sein, Kohlenstoff zuzugeben. Bor diffundiert unter Zuhilfenahme von Siliziumatomen auf Zwischengitterplätzen. Die Diffusion von Kohlenstoff bindet Zwischengitteratome. Ein Kohlenstoffgradient behindert so die Bordiffusion. Kohlenstoff bildet nun aber mit Indium in Silizium De-

fektpaare. Die Aktivierungsenergie eines solchen Paares liegt erheblich unter der des einzelnen Indium-Atoms.

Nach neueren Untersuchungen [V.52], [V.53] läßt sich durch Doppelimplantation von Indium und Kohlenstoff eine aktive Indium-Konzentration von $3{\cdot}10^{20}$ cm^{-3} erreichen! Das nötige Quentchen Abenteuerlust vorausgesetzt, kann es lohnen, für eine zukünftige Technologiegeneration Versuchstransistoren mit einer SiGeC:In-Basis und Bor als Dotand im Basisanschlußgebiet herzustellen.

11.5.4 Elektromigrationsfestigkeit der Silizide

Aluminium mit kleinen Beigaben Silizium und Kupfer ist das Standardmaterial bei der Metallisierung. AlSiCu beginnt bei einer DC-Stromdichte[49] über $5{\cdot}10^5$ A/cm^2 zu fließen: Die strömenden Elektronen reißen die Metallatome mit. Ausgehend von Korngrenzen – hier ist der Querschnitt der Leiterbahnen reduziert – wird Material abtransportiert, ein fortschreitender Prozeß, der zum Versagen der Metallisierung führt.

Je höher die Betriebsfrequenz eines elektronischen Bauelementes, desto höher wird die erforderliche DC-Stromdichte (Abb. 6.3). Tatsächlich erreicht bereits bei SiGe2RF der Arbeitspunktstrom die Stromtragfähigkeit der Metallisierung!

Bei den Siliziden ist die Datenbasis mager. [V.54] stellt fest, die Elektromigrationsfestigkeit von Tantalsilizid sei erheblich höher als von Polysilizium. [V.55] gibt für mesoskopische, heteroepitaktische $CoSi_2$-Drähte eine Elektromigrationsfestigkeit von $1.3{\cdot}10^7$ A/cm^2 an, mehr als den zwanzigfachen Wert von AlSiCu. Hierbei – mesoskopisch, heteroepitaktisch – handelt es sich sicherlich um eine obere Grenze.

Nichtsdestominder spricht nichts gegen die Eignung der Silizide zum Transport oder zur Spreizung der anfallenden Ströme.

11.5.5 Zusammenfassung Silizide

Als im Vergleich zu $TiSi_2$ höher temperaturfeste Silizide kommen WSi_2 und heteroepitaktisches $CoSi_2$ infrage. Beide lassen sich in situ erzeugen und scheinen als Festkörper-Diffusionsquellen geeignet. $CoSi_2$ zeichnet sich durch einen geringen spezifischen Widerstand aus, während für WSi_2 geeignete Trockenätzprozesse zur Verfügung stehen.

---

[49] Ein Wechselstromanteil zählt nur zu 10% zu einem Gleichtrom hinzu, Elektromigrationsfestigkeitsuntersuchungen unter reiner Wechselstrombelastung kenne ich nicht.

11.6 LEO, lateral epitaxial overgrowth

Die Rekristallisation amorphen Siliziums spielt an vielen Stellen eine Rolle:

- Ein durch eine Implantation oberflächlich amorphisierter Bereich rekristallisiert ausgehend vom unterliegenden Saatkristall.
- Scheidet man amorphes Silizium über einer Oxidschicht ab und tempert, kristallisiert das α-Si spontan oder ausgehend von der Grenzfläche zum Oxid, welche Kristallisationskeime zur Verfügung stellt, zu einer polykristallinen Schicht.
- Auch existieren Fremdstoffe, die die Kristallisationsgeschwindigkeit, nicht aber die Keimbildungsrate im Silizium erhöhen [V.56]. Am bekanntesten ist Nickel:

  Bei der Herstellung von Dünnschichttransistoren oder grobkristallinen Solarzellen wird auf ein Flachglassubstrat eine nickeldotierte Schicht amorphen Siliziums aufgebracht. Der Nickelzusatz senkt den Temperaturbereich, in dem grobkristallines Polysilizium entsteht, unter die Erweichungstemperatur des kostengünstigen Glassubstrates [V.57].

Eine besonders interessante Anwendung fand die Rekristallisation amorphen Siliziums im industriellen Maßstab, bevor SOI-Substrate kommerziell erhältlich waren [V.58]: Nach Aufbringen cgfssdddddftgtrdsx«yyolzsdfgkl.n[50] ... nach Aufbringen einer Oxidschicht und Öffnen von Saatfenstern im Oxid wird amorphes Silizium abgeschieden. Beim anschließenden Tempern breitet sich eine Kristallisationsfront, ausgehend von den Saatfenstern, durch das α-Si aus. Monokristallines Silizium wächst auf dem Oxid (für ein frühes Review der Technik s. [V.59]; ausführlich auf die Prozeßentwicklung geht z. B. [V.60] ein.)

Einen wesentlichen Unterschied gibt es zum Ausheilen nach einer Implantation: Die kristallisierende Siliziumschicht liegt einer Oxidschicht auf. Die Grenzfläche stellt eine Störung dar und erleichtert die Keimbildung, beziehungsweise dient selbst als Kristallisationskeim. An Stellen, an denen sich Kristallite gebildet haben, kommt die laterale Festphasenepitaxie zum Stillstand.

Mit undotiertem α-Si lassen sich etwa 4 µm Oxid lateral in guter Kristallqualität überwachsen [V.61], [V.62], [V.63], [V.64], [V.65], [V.66]. In hoch phosphor-, bor- oder germaniumdotiertem α-Si wandert die Kristallisationsfront dagegen 20 bis 40 µm weit, bevor sie an den spontan gebildeten Kristalliten zum Stillstand kommt. Die laterale Ausdehnung des monokristallinen Bereichs beträgt maximal die zehnfache Schichtdicke [V.67], [V.68], [V.69], [V.70], [V.71], [V.72].

Nach Kristallisation einer Doppelschicht aus einer dünnen, undotierten und einer dickeren, hochdotierten Lage amorphen Siliziums läßt sich die hochdotierte kristalline Lage selektiv entfernen (s. S. 65 f.). Es verbleibt eine dünne, undotierte, monokristalline Siliziumschicht, die sich 20 bis 40 µm weit über das Oxid erstreckt [V.72].

[50] der Einwurf stammt von meiner Tochter Anne; ihre erste Zeile, die im Druck erscheint!

11.6.1 Festphasenepitaktische Überwachsung von Silizid

Lassen sich auch Silizide festphasenepitaktisch überwachsen? Anders als Oxid weisen sie eine starke Neigung zur kristallinen Modifikation auf. Amorphes Kobaltsilizid beispielsweise rekristallisiert bereits bei etwas über 100°C [V.73]. Um Silizide durch Ionenbeschuß zu amorphisieren, muß die Waferrückseite während der Implantation mit Flüssigstickstoff gekühlt werden!

Die höchste Rekristallisationstemperatur (500°C) weist $TaSi_2$ auf [V.73]. Bei dieser Temperatur wächst der monokristalline Bereich bei der Festphasenepitaxie von undotiertem Silizium erst um etwa 1nm/min [V.59]. Man kann also davon ausgehen, daß bei der Silizium-Festphasenepitaxie Silizide stets in kristalliner Form vorliegen. Die Kristallite sind mögliche Kristallisationskeime für das amorphe Silizium.

In einem Patent zum selbstjustierten Silizidieren der Source-/Draingebiete in CMOS [V.74] verbirgt sich ein äußerst interessanter Hinweis, daß sich auch Silizide per lateraler Festphasenepitaxie mit monokristallinem Silizium überwachsen lassen sollten. Das Verfahren im Einzelnen:

- Im Kontakt mit Luft bildet sich bei Raumtemperatur auf Siliziumoberflächen ein natives, unterstöchiometrisches Oxid. Auf dieses natürliche Oxid werden 50 nm Titan und 100 nm amorphen Siliziums abgeschieden.
- Während einem Temperschritt – 30 Minuten bei 450°C unter Stickstoff – bilden sich aus der Titan-Silizium-Doppelschicht 90 nm $TiSi_2$. Das natürliche Oxid verhindert die Reaktion des Titans mit dem darunterliegenden Mono-Silizium. Eine 75 nm dicke Siliziumschicht verbleibt auf dem Silizid.
- Anschließend wird die Temperatur auf 550°C erhöht und für drei Stunden gehalten. Zunächst wird der im natürlichen Oxid nur schwach gebundene Sauerstoff im Silizid gelöst.

  Anschließend diffundiert das verbleibende amorphe Silizium durch die Silizidschicht, um epitaktisch auf dem darunterliegenden Silizium aufzuwachsen.

Treibende Kraft hinter dem dritten Schritt ist die Differenz der freien Enthalpie zwischen amorphem und kristallinem Silizium, er ist thermodynamisch nur möglich, wenn dem Silizid **amorphes** Silizium aufliegt: Ohne Kontakt zu einem Saatfenster ist eine α-Si-Schicht auf Titansilizid auch nach 3 Stunden bei 550°C noch amorph!

Ausgehend von einem Saatfenster erzielt [V.71] dagegen in 3 Stunden bei 560°C eine monokristalline Oxidüberwachsung von 6 µm.

Es folgt, daß sich Silizide, von einem Saatfenster aus, um mindestens 6 µm per lateraler Festphasenepitaxie monokristallin überwachsen lassen.

11.7 Die Idee der Silizidierung des Grabenbodens

Der Schichtwiderstand vergrabenener Siliziumschichten beträgt mehr als 10Ω/□ . 150 nm $TiSi_2$ haben dagegen gerade einmal 1Ω/□ ! Es wird daher immer wieder vorgeschlagen, durch Silizidierung den Bahnwiderstand des externen Kollektors zu reduzieren. Eine nähere Betrachtung läßt eine Realisierbarkeit fraglich erscheinen.

Silizidiert man zwischen den Aktivgebieten den Grabenboden, verringert dies zunächst den Bahnwiderstand nicht: Bei einer typischen buried-layer-Dotierung von $1{\cdot}10^{19}$ cm^{-3} beträgt der Übergangswiderstand 0.1 Ωcm^2. Bei einer Kontaktfläche von 1x20 μm^2 ergibt sich ein Kontaktwiderstand von 500 kΩ. Der Strom bleibt im Silizium. Zunächst müßte man daher das Silizid als Diffusionsquelle verwendet, um die Oberflächenkonzentration der Dotanden im Silizium zu erhöhen (Ⓟ 12).

Ein Silizid als Auflage auf dem Grabenboden wird **parallel zur Grenzfläche** bestromt. Ich teile, unter Vernachlässigung des Silizidwiderstandes, die Kontaktfläche in N Übergangswiderstandselemente auf, verbunden durch N Teile des Silizium-Bahnwiderstandes. In der Grenze $N \to \infty$ ergibt sich (mit dem Kontaktwiderstand R_{cont}, dem Square-Widerstand R_{sq} der Siliziumschicht, dem spezifischen Kontaktwiderstand ρ_c sowie Breite b und Länge L des Kontaktfensters)

$$bR_{cont} = \sqrt{R_{sq}\rho_c}coth\left(L\sqrt{R_{sq}/\rho_c}\right)$$

Erst nach einer Strecke von $L_c{=}1.5\sqrt{\rho_c/R_{sq}}$ (typischerweise 1 bis 2 µm) fließt der Strom im Silizid. Zwischen zwei Emitterfingern muß der Strom zweimal die Grenze zwischen Silizium und Silizid überwinden. Bei einem typischen Abstand von 5 µm zwischen zwei Emitterfingern ist er weiterhin kaum im Silizid unterwegs. Man müßte daher weiters erreichen, daß die Grenzfläche zwischen Silizium und Silizid senkrecht zur Stromrichtung liegt. Dies leistet das eher spekulative Vorratspatent Ⓟ 11:

Bereits nach Eintreiben der vergrabenen Schichten, also noch vor der Kollektorepitaxie, wird ein 500 nm tiefer Graben in das Silizium geätzt. Während die späteren Aktivgebiete noch mit der Nitrid-Hartmaske bedeckt sind, werden Grabenboden und -wände isotrop mit Metall belegt und silizidiert. Die Silizidschicht erstreckt sich anschließend auf dem Grabenboden und zieht sich seitlich die Grabenwände hinauf.

Das Silizid wird mit Bor oder Phosphor implantiert und strukturiert. Um die Ebenheit der Mesa-Fläche aus Silizium und Silizid zu verbessern, folgt ein mechanischer Polierschritt mit anschließender HCl-Oxidation.

Danach werden als Kollektorepitaxieschicht 300 nm α-Si abgeschieden und per Festphasenepitaxie kristallisiert. Die weitere Prozessierung folgt dem SiGe2-Prozeß, beginnend mit der Ätzung der STI-Gräben. Den Querschnitt nach Aufbringen der Hartmaske zeigt Abb. 11.7; grau ist das monokristalline Silizium dargestellt, grün das Silizid, blau das Oxid und kariert das auf den Silizidschichten gewachsene Polysilizium.

Die HBT-Struktur in einem späten Fertigungszustand zeigt Abb. 11.8.

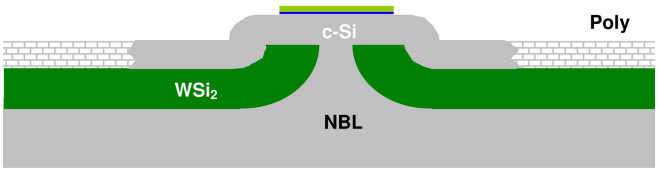


Abb. 11.7: Fertigungsablauf buried-layer-Silizidierung

Die Festphasenepitaxie ähnelt dem Ausheilen nach einer Implantation, lateral müssen nur minimale Distanzen überbrückt werden. Allerdings erfordert die Technologiearchitektur die Verfügbarkeit eines temperaturbeständigen Silizids als Diffusionsquelle: Das mit Silizium und Oxid abgedeckt WSi_2 wird bei der späteren SiGe-Epitaxie Temperaturen von mehr als 960°C ausgesetzt.

Eine 100 nm dicke Siliziumschicht und 39.5 nm Wolframauflage reagieren zu 100.7 nm WSi_2 (Dichte 9.88g/cm^3 bei Raumtemperatur, [V.42]). Die Silizidierung mit Kobalt geht dagegen mit einer vierprozentigen Volumenkontraktion einher. Eine geeignete Beimengung von Kobalt zum Wolfram führt zu einer unverspannten Silizium-Silizidgrenzfläche bei Raumtemperatur.

Die Crux im Prozeßablauf liegt in unterschiedlichen thermischen Ausdehnungskoeffizient von Silizid und Silizium: Im Verbund mit der geringen Duktilität der Silizide selbst bei erhöhten Temperaturen wird das Siliziumgefüge in Gegenwart der für eine senkrechte Bestromung der Grenzfläche erforderlichen Siliziddicke unter Temperaturbeaufschlagung zerstört (s. Oxidverspannungen, S. 20 ff.).

Ich habe das hypothetische Modul einer in Silizium vergrabenen Silizidschicht hier lediglich als einführenden Exkurs ausgebreitet: Der Grundgedanke des per lateraler Festphasenepitaxie unter monokristallinem Silizium vergrabenen (im folgenden Beispiel: ebenen und von warmduktilem Oxid umgebenen) Silizides läßt sich nutzen, um bei der vertikal integrierten Kaskodenstruktur die äußere Basis des unteren Transistors **von unten her** anzuschließen:

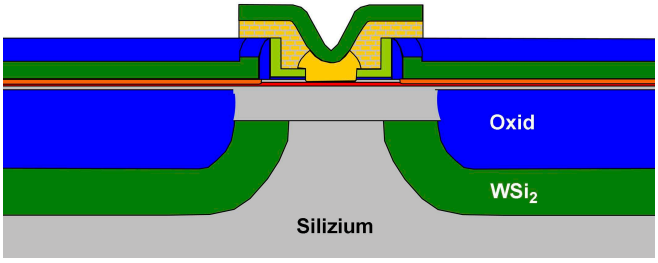


Abb. 11.8: HBT-Querschnitt mit silizidierter buried layer

11.8 Der Anschluß der unteren Basis

Nach Ätzen des Isolationsgrabens wird dieser mit einer Doppelschicht aus Oxid und Polysilizium verfüllt. Das Polysilizium wird im weiteren Verlauf silizidiert werden, um die untere Basis der vertikal integrierten Kaskodenstruktur von unten her anzuschließen. Zunächst jedoch wird es strukturiert und von einer Schicht TEOS bedeckt. Das Schichtpaket wird mit einer Inversmaske zurückgeätzt, um planarisieren zu können. Den Bauelementquerschnitt in diesem Bearbeitungszustand zeigt Abb. 11.9.

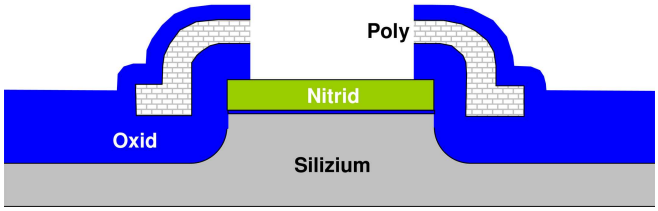


Abb. 11.9: Anschluß der unteren Basis

Im Anschluß an die CMP wird das offenliegende Polysilizium mit Wolfram silizidiert. Die Hartmaske wird entfernt und nach Abscheiden amorphen Siliziums die Struktur per Festphasenepitaxie mit monokristallinem Silizium überwachsen. Schließlich werden per Gasphasenepitaxie die untere Basis und ein Teil des Zwischenbasisbereichs gebildet.

An dieser Stelle wird die Epitaxie unterbrochen. Nach Abscheiden und Strukturieren eines dünnen Oxides wird in zwei Bädern zunächst das offenliegende Silizium selektiv zur unterliegenden SiGe-Schicht und anschließend das SiGe selektiv zum darunterliegenden Silizium entfernt (zur Badchemie s. S. 65 f.). Das Oxid schützt die Stellen, an denen vertikal integrierte Kaskodenstrukturen entstehen sollen, vor dem Ätzangriff.

Auf Abscheiden des restlichen Zwischenbasisbereichs aus Silizium, der oberen Basis und der dünnen p^--Siliziumschicht für den Emitter folgen das oben beschriebene Silizid- und Emittermodul. Man erhält den Querschnitt aus Abb. 11.10.

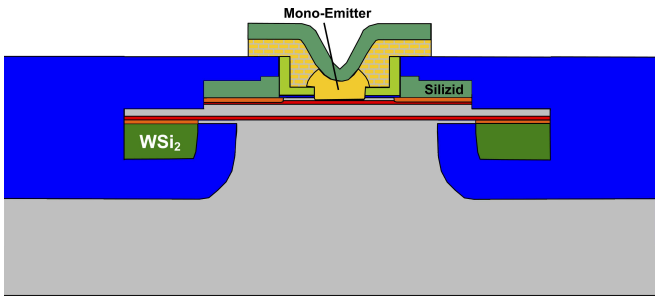
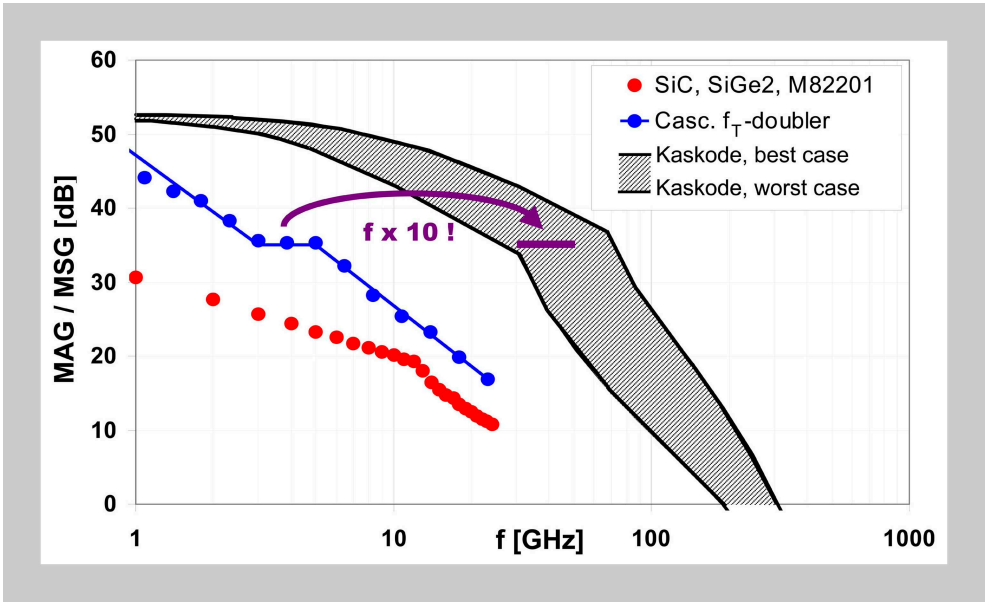


Abb. 11.10: Integrierte Kaskodenstufe mit silizidierter buried layer

VI Nutzen des neuen Bauelementes



Selbst, wenn sich der Bauer sorgt und handelt,
wo die Saat in Sommer sich verwandelt,
reicht er niemals hin. Die Erde schenkt.

Rainer Maria Rilke, SONETTE AN ORPHEUS [VI.1]

Ausgehend von SiGe2 aus dem Hause Atmel habe ich in die Technologieintegration eingeführt. Anhand praktischer Beispiele habe ich gezeigt, wie sich Technologieintegration und die Bauelemententwicklung mit der Simulation unterstützen lassen. Aus Anforderungen der Schaltungstechnik ist die Idee für ein neuartiges Hochfrequenzbauelement erwachsen. Aufbauend auf der Erfahrung in der Technologieintegration wurde ein möglicher Herstellungsablauf für das neuartige Bauelement skizziert.

Als Vorbereitung für die Investitionsentscheidung im Technical Board der Firma Atmel untersuche ich nachfolgend, ob sich mit dem vorgeschlagenen Bauelement die Effizienz von Leistungsverstärkern für Mobilgeräte nach dem Bluetooth-Standard verbessern läßt. Dies gibt mir gleichzeitig die Möglichkeit, meine praktische Erfahrung in der Modellbildung und der Schaltungsentwicklung auszubauen.

Zunächst entwerfe ich hierfür ein einfaches Kompaktmodell aus zwei npn-Transistoren als Ersatzschaltung für das vorgeschlagene Bauelement und bestimme die Modellparameter anhand der Simulationen. Anschließend wird das Modell in ADS implementiert und ein Großsignal-Leistungsverstärker simuliert.

12 Simulation der integrierten Kaskode

In Kapitel 7 habe ich die physikalische npn-Simulation vorgestellt: Ausgehend von experimentell ermittelten oder aus Prozeßablauf und Bauelementgeometrie abgeschätzten Querschnitten verwende ich ATLAS und DevEdit (beide SILVACO). DevEdit erlaubt, auf einfache Weise Bauelementquerschnitte und -diskretisierungen zu definieren. Zu beiden Programmen existieren hervorragende Handbücher, auf die ich bezüglich der technischen und physikalischen Grundlagen verweise.

Bei der HBT-Simulation definiere ich Regionen einheitlicher Zusammensetzung, in denen die Materialparameter an den jeweiligen Germaniumgehalt angepaßt sind. Das kontinuierliche Germaniumprofil wird also ersetzt durch ein hinreichend fein abgestuftes Stufenprofil. Die Materialparameter für SiGe habe ich aus [III.7] bezogen.

Geometrie- und Dotierungsvarianten spiele ich durch, indem ich Bauelementgeometrie und -diskretisierung in Abhängigkeit geeigneter Parameter definiere; um die Auswirkungen einer Geometrievariation auf das Bauelementverhalten zu bestimmen, genügt es dann, lediglich an einer Stelle in einem Vorspann der Eingabefiles die Geometrieparameter mit geeigneten Werten zu versehen.

Derselben Strategie bediene ich mich bei der Simulation der vertikal integrierten Kaskode. Für Einzelheiten verweise ich auf das Eingabefile im Anhang, Kapitel 19.

Auf die simulierten „Meßwerte“ gehe ich im Rahmen der Modellierung und der Schaltungssimulation in den nächsten Kapiteln ein. Für hier möge ein Querschnitt genügen, Abb. 12.1, welcher den Dotierstoffverlauf (schwarz) wiedergibt, sowie für unterschiedliche Arbeitspunkte die Elektronen- (Rot-) und die Löcherdichten (Blautöne):

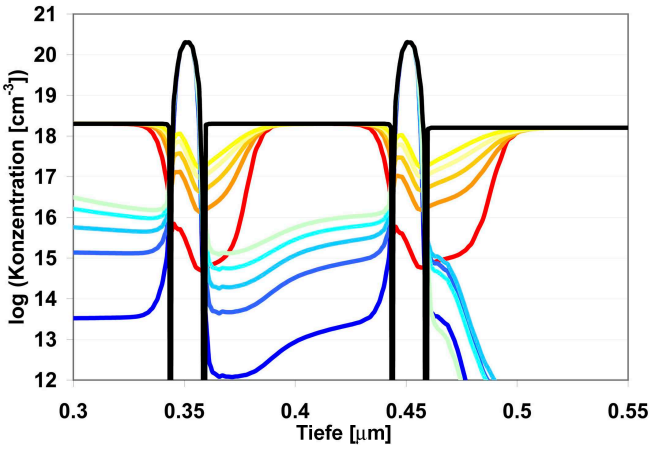


Abb. 12.1: Dotierstoff- und Ladungsträger-Verläufe der vertikal integrierten Kaskodenstufe (Basisspannungen 0.7·0.8·0.85·0.9·0.95 V, Farbkodierung selbsterklärend.)

13 Ergebnisse der physikalischen Simulation

Es wurde eine große Anzahl physikalischer Bauelementsimulationen durchgeführt, um zu einem in Geometrie und Dotierung skalierbaren Kompaktmodell für die Schaltungssimulation zu gelangen. Aus dem Herstellunsgablauf (s. S. 94) folgen hierbei die erzielbaren Bauelementgeometrien und Dotierstoffprofile, aus dem Kollektormodell (s. Kapitel 6.3) die optimale Kollektordicke bei gegebener Kollektordotierung:

$$\text{Kollektordicke [nm]} = 1246 - 61.28 \log\left(\text{ Kollektordotierung } [\text{cm}^{-3}]\right)$$

Das Eingabefile für die Simulationen findet sich im Anhang, s. S. 121. Es enthält die Dicke und die Dotierung des Kollektors sowie die lateralen Bauelementabmessungen als freie Parameter. Bevor ich in die Parameterextraktion einsteige, fasse ich die Ergebnisse der physikalischen Simulation zusammen und vergleiche sie mit dem Stand der Technik in SiGe2RF.

Abb. 13.1 zeigt in rot den Frequenzgang der Leistungsverstärkung eines minimalen Transistors mit selektiv implantiertem Kollektor. Mit schaltungstechnischen Tricks läßt sich der mit dem Einzelbauelement erzielbare Frequenzverlauf verbessern[51]. Blau dargestellt ist die gemessene Leistungsverstärkung eines neuartigen f_T-Dopplers aus Ⓟ 6.

Trägt man schließlich die Leistungsverstärkung der unterschiedlichen simulierten Kaskodenstrukturen im jeweils optimalen Arbeitspunkt über der Frequenz auf, so liegen sämtliche Werte innerhalb des grau schraffierten Bereichs: Die vertikal integrierte Kaskode zeigt dieselbe Verstärkung wie der bereits ausgezeichnete f_T-Doppler – aber bei einer um den Faktor 10 höheren Frequenz!

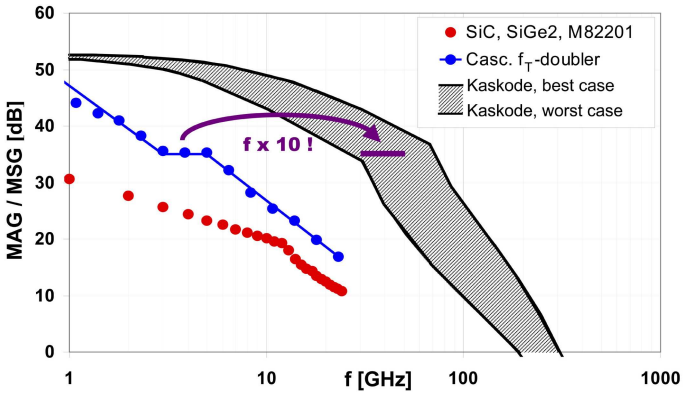


Abb. 13.1: Frequenzverhalten

Mit diesem vielversprechenden Zwischenergebnis starte ich in die Erstellung eines Kompaktmodells:

---

[51] Anders als zum Teil in der Literatur angenommen, gilt dies tatsächlich auch für die Leistungs- und nicht nur für die Stromverstärkung!

14 Kompaktmodell und Parameterextraktion

Auch, wenn eine physikalische Simulation der Bauelementeigenschaften zur Verfügung steht, ist für eine schnelle Schaltungssimulation ein Kompaktmodell erforderlich: Die weiter unten gezeigte Load-Pull-Optimierung der Kaskodenstruktur hätte mithilfe der physikalischen Simulation 40 Jahre CPU-Zeit erfordert.

Vom Aufbau her handelt es sich bei der vertikal integrierten Kaskodenstufe um ein thyristorähnliches Bauelement. Die Physik ist jedoch eine andere. Die bestehenden Thyristormodelle sind auf hochsperrende, nichtlineare Bauelemente ausgelegt, während hier ein kaskodenartiges, lineares Hochfrequenzbauteil charakterisieren werden soll. Ich modelliere daher die beteiligten npn- und pnp-Transistoren getrennt. Die Teiltransistoren werden mit Hilfe von MEXTRAM-Modellen beschrieben[52].

Die Kapazitäten zwischen den Basen und dem Zwischenbasisgebiet werden dabei fälschlicherweise doppelt, jeweils einmal in einem npn- und einmal im pnp-Modell, gezählt. Entsprechend können wir keine ideale Übereinstimmung zwischen der physikalischen Simulation und den Ergebnissen aus dem Kompaktmodell erwarten.

Die Extraktion der MEXTRAM-Modellparameter beginnt mit der Bestimmung der Kapazitäten und Bahnwiderstände. Nach [VI.3] bietet sich hier vor allem die Verwendung eines LCR-Messplatzes an – für die Simulation ist dieses Verfahren natürlich nur eingeschränkt tauglich!

14.1 Diodenkapazitäten und Bahnwiderstande

Ich greife einmal mehr auf die analytischen Lösung des Hybrid-Pi-Modells (s. S. 6) zurück. Anhand der Hybrid-Pi-Lösung lassen sich Näherungen zur Abschätzung der Diodenkapazitäten und Bahnwiderstände angeben.

Zunächst vereinfache ich die Lösung. Hierbei nutze ich aus, daß die Kollektor-Substrat-Kapazität bei SiGe2 keine wesentliche Rolle für das Verhalten des innneren Bauelements spielt. Ein Selbes gilt, mit dem simulierten Querschnitt im Blick, für die äußere Basiskapazität. Entsprechend setze ich C_{bx} und C_{CS} gleich Null.

Angelehnt an die „off-state S-Parameter method“ [VI.3] zur Bestimmung der Kapazitäten betrachte ich den Transistor nur in ausgeschaltetem Zustand (d. h., entweder nur die Basis- oder nur die Kollektorspannung sind von Null verschieden). Hier sollte das Hybrid-Pi-Modell eine ausgezeichnete Näherung darstellen. Ich kann auch g_m nullsetzen und die Widerstände r_0 und r_π nach unendlich gehen lassen.

Die analytische Lösung, nach Potenzen von ω entwickelt, vereinfacht sich dann im

[52] Für die Einführung in die MEXTRAM-503-Extraktion geht mein herzlicher Dank an Herrn Dr. Ertugrul Sönmez vom ICIC, Ulm. Für diejenigen, die auf Ertugruls Unterstützung verzichten müssen, gibt [VI.2] eine Beschreibung des Extraktionsverfahrens.

Grenzfall kleiner Frequenzen (in MEXTRAM-konformer Nomenklatur) zu:

$$\lim_{\substack{\omega \approx 0 \\ r_0, r_\pi \to \infty \\ C_{bx}, g_m = 0}} \mathbf{Y} = j\omega\, \mathbf{j_y} + \omega^2\, \mathbf{r_y}\,, \qquad \mathbf{j_y} = \begin{pmatrix} C_{jc} + C_{je} & -C_{jc} \\ -C_{jc} & C_{jc} \end{pmatrix}$$

$$\mathbf{r_y} = \begin{pmatrix} r_b(C_{jc} + C_{je})^2 + r_c C_{jc}^2 + r_e C_{je}^2 & -r_b C_{jc}(C_{je} + C_{jc}) - r_c C_{jc}^2 \\ -r_b C_{jc}(C_{je} + C_{jc}) - r_c C_{jc}^2 & C_{jc}^2 (r_b + r_e) \end{pmatrix}$$

Aus einer einfachen S-Parameter-Simulation ergibt sich sofort die Spannungsabhängigkeit der Diodenkapazitäten und der Bahnwiderstände! (Das vorgeschlagene Extraktionsverfahren ist m. E. n. neu; möglicherweise bietet es sich auch, zumindest als Startpunkt, für die Modellierung realer Bauelemente an.)

14.1.1 „Messung“

Ähnlich wie mit einem gut ausgestatteten Meßplatz für das reale Bauelement, lassen sich mit ATLAS sämtliche interessierenden „Meßwerte“ am simulierten Bauelement bestimmen. Zur Auswertung der dergestalt in Form von .log-files erhaltenen Werte hat sich das nachfolgend skizzierte Verfahren bewährt:

Zunächst wird das Bauelement in den gewünschten Arbeitspunkt gefahren. Anschließend erstellt man vermittels einer Kommandosequenz wie

```
log outf=S.log width=1 gains inport=base outport=collector master
solve prev ac.analysis freq=1e6 fstep=1.3 mult.f nfstep=53 auto
```

ein ATLAS-logfile. Hierin folgt einem header beispielsweise eine Liste der Form:

f	3	base	emitter	collector	...	
p	63	40	811	911	...	812
d	1.00E+08	0.505091	-0.0157547	0.0102605	...	0.000183986
d	2.00E+08	0.504382	-0.031488	0.0102683	...	0.000367738
...	...	...	...	...	...	...

Die erste Zeile (**f**) deklariert die Zahl der Elektroden und ihre Namen in der gewählten Reihenfolge. Daran schließt sich eine Zeile (**p**) mit Schlüsseln an, die erklären, welche Einträge in den folgenden Spalten stehen. Der erste Eintrag, im Beispiel **63**, gibt die Zahl der Spalten. Über die Darstellung des .log-files in Tonyplot, sowie den Vergleich unterschiedlicher .log-files habe ich die folgende Konkordanz erstellt, um die Schlüssel für den Spalteninhalt interpretieren zu können. Die Tabelle gibt jeweils den Schlüssel, gefolgt von seiner Auflösung in Klartext, an:

2	B Voltage	20	B Current	66	Current gain
3	E Voltage	21	E Current	67	Unil. pwr g.
4	C Voltage	22	C Current	68	Transd. gain
40	Frequency	1011	re(H11)	1111	im(H11)
203	k	1012	re(H12)	1112	im(H12)
3000	Gms (dB)	1021	re(H21)	1121	im(H21)
3001	Gma (dB)	1022	re(H22)	1122	im(H22)
811	re(S11)	831	re(Y11)	931	im(Y11)
812	re(S12)	832	re(Y12)	932	im(Y12)
821	re(S21)	841	re(Y21)	941	im(Y21)
822	re(S22)	842	re(Y22)	942	im(Y22)
911	im(S11)	851	re(Z11)	951	im(Z11)
912	im(S12)	852	re(Z12)	952	im(Z12)
921	im(S21)	861	re(Z21)	961	im(Z21)
922	im(S22)	862	re(Z22)	962	im(Z22)
111	C B>B	121	C E>B	131	C C>B
112	C B>E	122	C E>E	132	C C>E
113	C B>C	123	C E>C	133	C C>C
211	G B>B	221	G E>B	231	G C>B
212	G B>E	222	G E>E	232	G C>E
213	G B>C	223	G E>C	233	G C>C

Die übrigen Zeilen in .log-file, die mit dem Bezeichner **d** beginnen, enthalten die gesuchten Zahlenwerte in der Reihenfolge der Schlüssel. Als bspw. .citi-file formatiert, lassen sich die Daten in ADS einlesen.

Drei Anmerkungen zu Eigenheiten des Simulators ATLAS 5.3.36 sind angebracht:

Zunächst bestimmt ATLAS die Phase der Admittanzparameter falsch: Das Produkt der Y- mit der Z-Matrix ergibt die negative Einheitsmatrix. Es wurde wohl irgendwo tief im Kernel die „elektronische“ und die „physikalische“ Zeitabhängigkeit, exp(-iωt) bzw. exp(+jωt), verwechselt. . . . Entsprechend muß man den **negativen** Imaginärteil der ausgegebenen Y-Parameter für die Kapazitätsbestimmung heranziehen.

Beim Vergleich mit den von ATLAS angegebenen Kapazitätswerten fällt auf, daß diese fälschlich stets auf eine Bauelementweite von 1 µm bezogen sind, während die Y-Parameter korrekt mit der angegebenen Bauelementweite skalieren.

Ferner möchte ich darauf hinweisen, daß die von SILVACO berechneten Verstärkungen (wie bspw. MAG oder h_{21}) allesamt um einen Faktor 2.3 zu hoch sind. Der Vergleich mit den S-Parametern zeigt, daß ATLAS den zehnfachen **natürlichen** Logarithmus als Dezibel-Wert ausgibt – wir verkneifen uns ein Grinsen, da die ATLAS-Entwickler im allgemeinen hervorragende Arbeit geleistet haben!

14.1.2 Parameteranpassung und Vergleich mit den „Meßdaten“

Sind die spannungsabhängigen Kapazitätswerte bekannt, wird ihr Verlauf mithilfe der MEXTRAM-Formeln angefittet. Das Standardverfahren nach Paaschens [VI.3], ein

numerischer „best-fit“, ist m. E. n. zur Bestimmung der Diffusionsspannung untauglich:

Hält man die Diffusionsspannung V_{diff} fest und bestimmt die verbleibenden Modellparameter so, daß sie die mittlere quadratische Abweichung zwischen dem „gemessenen“ und dem gefitteten Kapazitätsverlauf minimieren, erhält man eine Kurve „mittlerer Fehler über Diffusionsspannung“. Das vorgeschlagene Extraktionsverfahren wählt den Wert der Diffusionsspannung, der den mittleren Fehler minimiert.

Die Zuweisung ist nicht exakt. Ein geringfügig abweichender Wert der Diffusionsspannung führt im allgemeinen nur zu einem geringfügig höheren Fehler. Ein einfaches Verfahren, die Zuverlässigkeit der dergestalt bestimmten Diffusionsspannung abzuschätzen, wählt den Konfidenzbereich als die Menge aller Ordinatenwerte, für die die mittlere Abweichung höchstens doppelt so hoch ist wie im Minimum. Im Falle der hier betrachteten Bauelemente erstreckt sich der V_{diff}-Konfidenzbereich von 0.6 V bis 2.5 V – was, verglichen selbst mit der gröbsten theoretischen Abschätzung, vollkommen indiskutabel ist!

Ich wähle daher die theoretische, von Dotierung und Germaniumgehalt abhängige Diffusionsspannung und bestimme lediglich die verbleibenden Modellparameter durch Fehlerminimierung (die übrigen Konfidenzintervalle für die Kapazitätsparameter sind, im betrachteten Beispiel, schmal).

Abb. 14.1 zeigt die simulierte Kollektorkapazität für unterschiedliche Kollektordotierungen, aufgetragen über der Basis-Kollektor-Spannung. Die aus den Y-Parametern erhaltenen „Meßwerte“ sind als Punkte dargestellt, durchgezogen wird der best-fit nach der in MEXTRAM verwendeten analytischen Formel wiedergegeben.

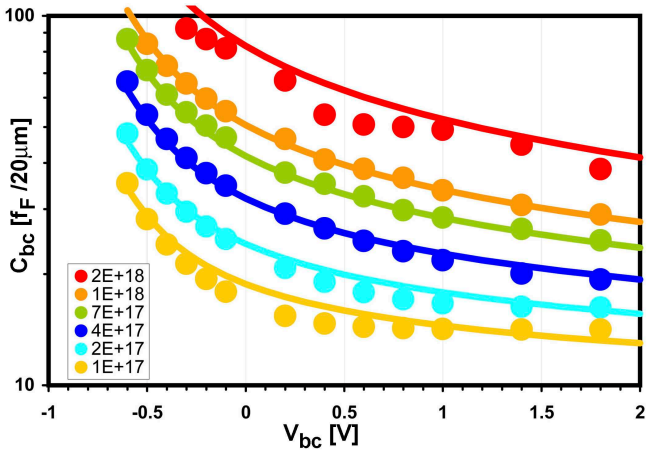


Abb. 14.1: Die Kollektorkapazität in Simulation und Modell (die Zahlenwerte in der Kurvenbeschriftung geben die Kollektordotierung pro Kubikzentimeter an).

Die Simulation mehrerer unterschiedlicher Transistoren erlaubt, zwischen den Parametern zu interpolieren und so die statistischen Unwägbarkeiten zu verringern. Die grafische Auftragung der Modellparameter für die Kollektorkapazität (Abb. 14.2) zeigt, wie die Interpolation den numerischen Fehler verringert.

Um für den Sättigungsstrom, Abb. 14.3, zu verlässlichen Werten zu gelangen, werden

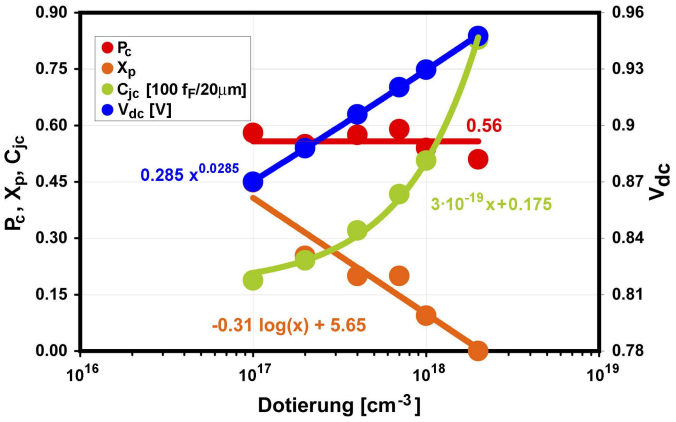


Abb. 14.2: Dotierungsabhängigkeit der Parameter, die die Kollektorkapazität beschreiben

die aus dem Verhalten der in Vorwärts- und Rückwärtsrichtung betriebenen Transistoren extrahierten Werte für unterschiedliche Kollektordotierungen interpoliert (rein heuristisch entscheide ich mich für ein Sättigungsverhalten vom (n-1)/(n+1)-Typ).

Andere der (dotierungsabhängigen) Parameter lassen sich mit geringer Unsicherheit extrahieren. Bestimmt man bei diesen Parametern die Zahlenwerte für verschiedene Bauelementgeometrien, ergeben sich glatte Kurven. Dies ist beipielsweise der Fall für die Bahnwiderstände (Abb. 14.4).

Auch die Parameter, die die Kapazitätsbeläge beschreiben, sind, über der Emitterweite aufgetragen, hervorragend glatt. Ich verzichte auf die Darstellung im Einzelnen, und gehe über zum Vergleich der ATLAS-Simulation mit den Ergebnissen des Kompaktmodells.

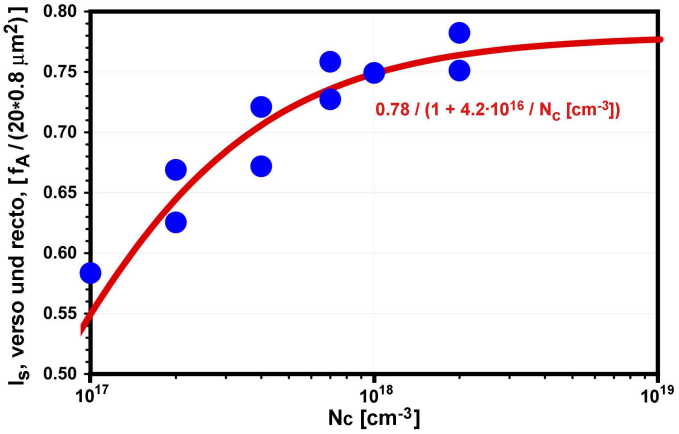


Abb. 14.3: Interpolation des Sättigungsstromes

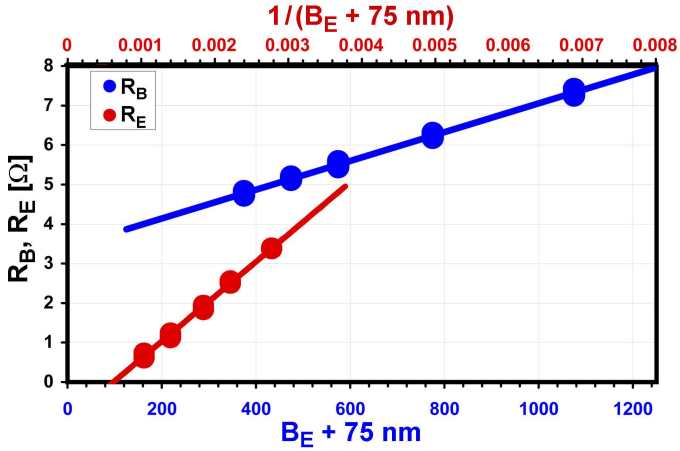


Abb. 14.4: Verlauf von R_E und R_B über der Breite der Emitters

14.1.3 Vergleich Simulation – Kompaktmodell

Die DC-Parameter aus der physikalischen Simulation der Teiltransistoren lassen sich gut mit MEXTRAM-Modellen reproduzieren: Abb. 14.5 zeigt die Ausgangskennlinienfelder zu Kollektordotierungen von $2{\cdot}10^{17}$ cm^{-3} bzw. $7{\cdot}10^{17}$ cm^{-3}. Die gewählten Dotierungen sind relevant für den Teiltransistor in Basis-, rsp. in Emitterschaltung.

Abb. 14.6 zeigt die S-Parameter für $2{\cdot}10^{17}$cm^{-3} bzw. $7{\cdot}10^{17}$cm^{-3} Kollektordotierung. Die Übereinstimmung erscheint dem frühen Entwicklungsstadium angemessen.

Beim Zusammenbau der Modelle der Teiltransistoren führe ich einen zusätzlichen Fehler durch die doppelte Wertung des Kapazitätsbelages zwischen den Basen und dem Zwischenbasisbereich ein (sowohl in einem npn-, als auch einem pnp-Modell). Abb. 14.7 vergleicht die physikalische Simulation der Leistungsverstärkung der Kaskodenstufe (rot) mit der Wiedergabe durch das Kompaktmodell (blau). Die Übereinstim-

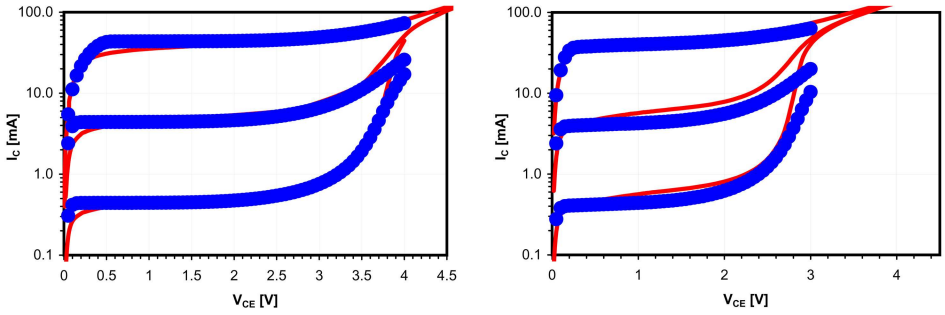


Abb. 14.5: Ausgangskennlinienfeld, N_C=$2{\cdot}10^{17}$cm^{-3} (links) bzw. $7{\cdot}10^{17}$cm^{-3} Kollektordotierung (rechts): Rot dargestellt sind die Ergebnisse der physikalischen Simulation, blau die S-Parameter aus dem Kompaktmodell.

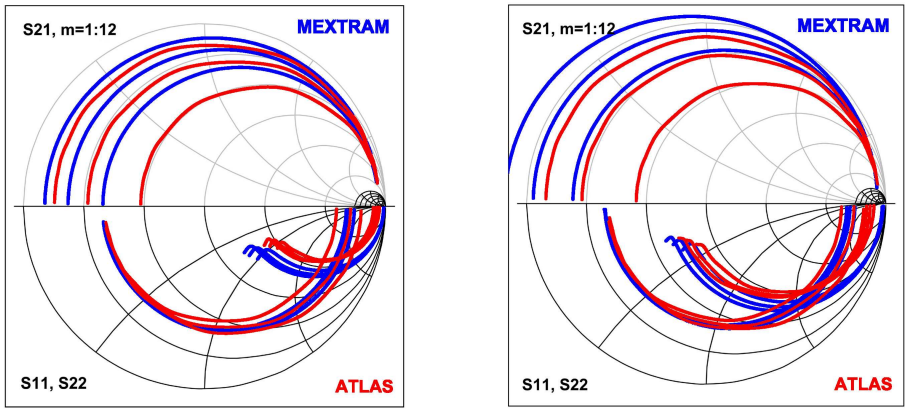


Abb. 14.6: S-Parameter für $2{\cdot}10^{17}cm^{-3}$ (links) bzw. $7{\cdot}10^{17}cm^{-3}$ Kollektordotierung (rechts): Rot dargestellt sind die Ergebnisse der physikalischen Simulation, blau die S-Parameter aus dem Kompaktmodell.

mung erscheint ausreichend, um anhand der Simulation eine grobe Abschätzung des Verhaltens des projektierten Bauelementes in einer elektronischen Schaltung vornehmen zu können:

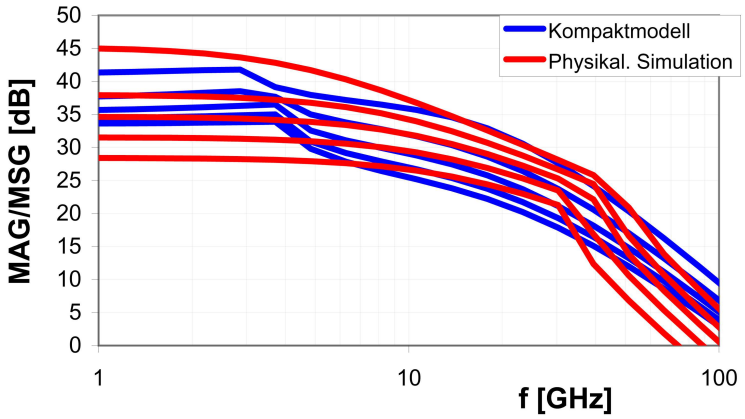


Abb. 14.7: Leistungsverstärkung der integrierten Kaskodenstufe

15 Schaltungssimulation

Das vorliegende Kapitel faßt die Simulation der vertikal integrierten Kaskodenstufe anhand des Kompaktmodelles (Kollektordotierung: $7{\cdot}10^{17}\,\mathrm{cm}^{-3}$ / $2{\cdot}10^{17}\,\mathrm{cm}^{-3}$) zusammen und vergleicht sie mit Simulationen einer Emitterschaltung anhand des HiCuM-Modells des Transistors aus dem SiGe2-Power-Designkit ohne selektiv implantierten Kollektor (Kollektordotierung: $2{\cdot}10^{16}\,\mathrm{cm}^{-3}$) [53].

Das wesentliche Hindernis für den Einsatz der SiGe2-Transistoren als Ausgangstreiber im Mobilfunkbereich ist die fehlende „ruggedness“: Im effizienzoptimierten Betrieb an 3.4 V verbleibt nur ein minimaler Overhead zwischen der höchsten dynamisch erreichten Spannung und der Durchbruchspannung der Bauelemente. Bei Lastschwankungen verändert sich die Form der dynamischen Lastlinien, und der Overhead reicht nicht aus, um eine Zerstörung der Bauelemente unter Fehlanpassung zu verhindern.

Aufgrund des Übergangs von der Emitter-Kollektor-Durchbruchspannung bei der Emitterschaltung zur Basis-Kollektor-Durchbruchspannung bei der Kaskode erlaubt die Kaskode selbst noch bei einer höheren Sperrfähigkeit eine deutlich höhere Kollektordotierung und mithin Stromtragfähigkeit und Transitfrequenz als der Transistor.

15.1 Gleichstrom-Kennlinienfeld und ruggedness

Abb. 15.1 zeigt das Gleichstromverhalten von vertikal integrierter Kaskode (blau) und SiGe-Transistor (grün). Die Reserve zwischen der höchsten im Normalbetrieb auftretenden Spannung und der Durchbruchspannung des SiGe2-Power-Transistors wird in Abb. 15.1 durch den kurzen roten Pfeil dargestellt. Der lange gelbe Pfeil stellt den Spannungs-Overhead für die vertikal integrierte Kaskodenstufe dar. Durch den Übergang zu einer Kaskodenschaltung läßt sich die ruggedness deutlich erhöhen!

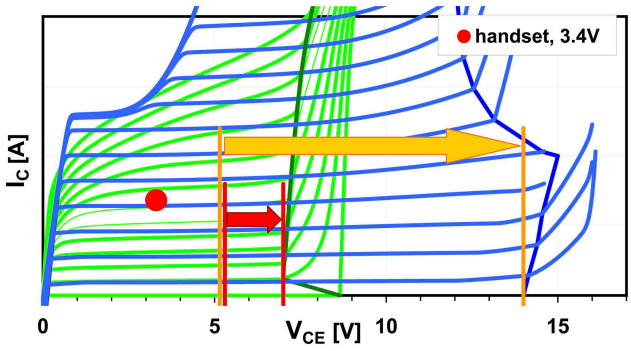


Abb. 15.1: Ruggedness Blau: Ausgangskennlinienfeld zur vertikal integrierten Kaskode, grün: zum Non-SiC Transistor von SiGe2-Power in Emitterschaltung. Durch Pfeile angedeutet ist der Abstand von der Betriebs- zur Durchbruchspannung.

[53] Eine lateral integrierte Kaskodenstruktur ist aufgrund der erforderlichen Verdrahtung für Leistungsverstärker ungeeignet.

15.2 Kirk- und Early-Effekt in Kaskode und Transistor

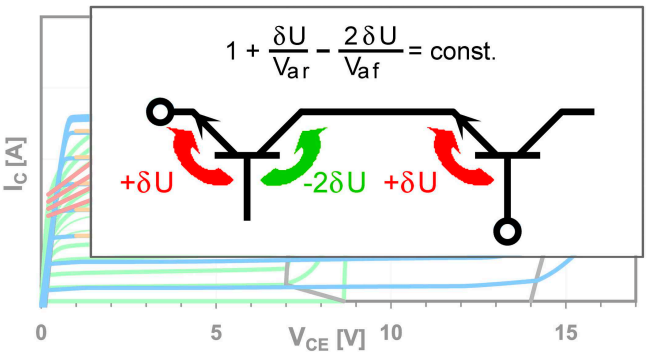


Abb. 15.2: Zum Early-Verhalten der Kaskodenstufe: Early- und reverse Early-Effekt schwächen einander ab.

Steigt die Kaskoden-Eingangsspannung um δU, erzwingt der geänderte Strom auch in der Basisstufe eine geänderte Basis-Emitter-Spannung. Hierdurch verringert sich das Potential im Zwischenbasisbereich, d. h. am Kollektor der Eingangsstufe, um -δU: In der Eingangsstufe sinkt die Basis-Kollektor-Spannung um 2δU, wenn die Basis-Emitter-Spannung um δU zunimmt.

Mithilfe des skalierbaren Kompaktmodells findet man, daß sich bei einer Dotierung von $2{\cdot}10^{18}\,\mathrm{cm}^{-3}$ im Emitter und $7{\cdot}10^{17}\,\mathrm{cm}^{-3}$ im Zwischenbasisbereich Early- und reverse Early effect aufheben: Man erhält ideales Early-Verhalten (Abb. 15.3, 15.2).

Bei dem Transistor ohne selektiv implantierten Kollektor degradiert durch den Kirk-Effekt die Verstärkung bei moderaten Stromdichten (vgl. [III.5], [VI.4]). Auch tragen das Basisprofil und die Ausformung der äußeren Basis zu einer schlechten Early-Spannung bei (vgl. [VI.4]).

Bei der Kaskodenstufe ist die Weite bei der geometrisch unteren höher als bei der oberen Basis. Der Strom wird aufgespreizt, die Stromdichte im Kollektor verringert. Das Ausgangskennlinienfeld ist auch bei kleinen Spannungen nicht verrundet (Abb. 15.3).

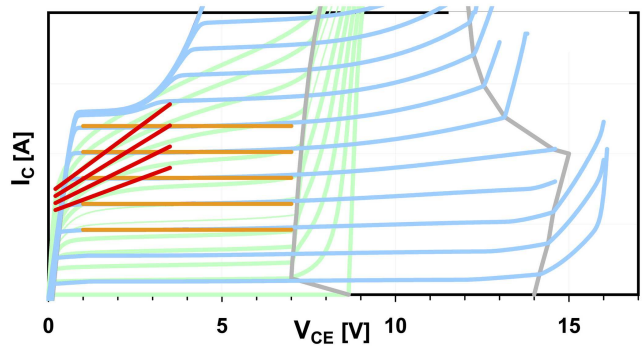


Abb. 15.3: Kirk-Effekt: Die stärkere Aufspreizung des Stromes Im Kollektor verbessert das Early-Verhalten der integrierten Kaskode bei kleinen Spannungen (orange) im Vergleich zum Transistor (rot)

15.3 Kompressionsverhalten von Kaskode und Transistor

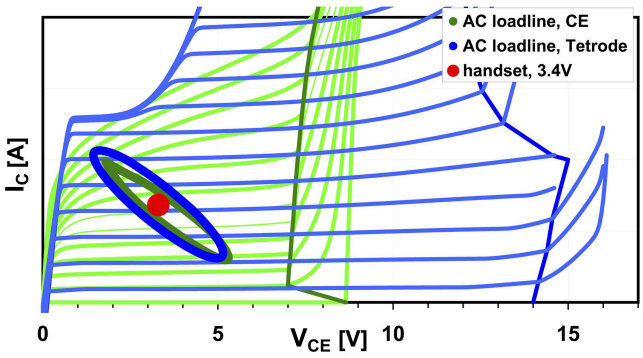


Abb. 15.4: Early-Verhalten: Konsequenzen „runde“, verzerrungsarme HF-Kennlinie bei der Kaskode (blau), obertonreiche Kennlinie (grün) im Transistor

Die hohe Early-Spannung ist ein gutes Vorzeichen für einen verzerrungsarmen Hochfrequenzbetrieb. Tatsächlich weist die vertikal integrierte Kaskodenschaltung ein auffällig ärmeres Obertonspektrum auf als die nämliche Kennlinie des SiGe-HBT in Emitterschaltung, vgl. die wohlgerundete HF-Ausgangskennlinie der Kaskode in Abb. 15.4.

Dies übersetzt sich unmittelbar in ein freundliches Kompressionsverhalten im selben Arbeitspunkt, Abb. 15.5.

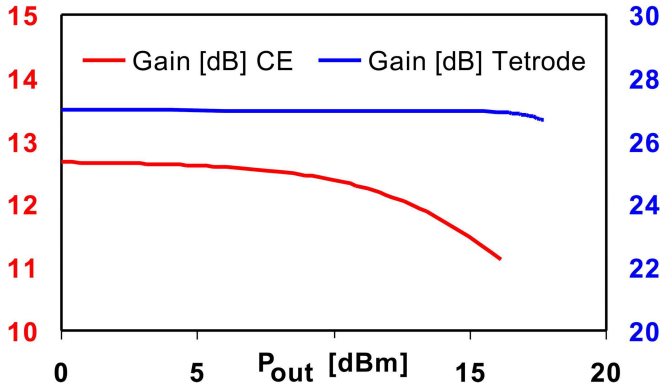


Abb. 15.5: Kompressionsverhalten Emitter-Schaltung („CE“) bzw. vertikal integrierte Kaskodenschaltung („Tetrode“)

VII Zusammenfassung

So eine Arbeit wird eigentlich nie fertig; man müßte sie für fertig erklären, wenn man nach Zeit und Umständen das Möglichste getan hat.

Goethe, ITALIENISCHE REISE [VII.1]

Die vorliegende Arbeit kreist um ein neu vorgeschlagenes Hochfrequenz-Bauelement.

Der urspüngliche Ansatz war schwerstexperimentell, quasi nach dem „hit and run"-Prinzip. Interessanterweise stehen einer solchen Bastel-Lösung die selben Schwierigkeit im Wege wie einer Fortentwicklung des SiGe-HBT: Die Hochfrequenzeignung wird nicht mehr allein durch den inneren Transistor bestimmt, sondern durch die Notwendigkeit begrenzt, die einzelnen Transistorregionen elektrisch zu kontaktieren!

Für diese Aufgabe habe ich eine mögliche Lösung in Form einer Technologiearchitektur angegeben. Voraussetzung für die Umsetzung ist der Erfolg umfangreicher experimenteller Vorarbeiten zur festphasenepitaktischen Überwachsung von Siliziden mit Silizium. Und so steht am Ende der Arbeit kein fertiges Bauelement, sondern ein wohl-ausgefeilter Investitionsvorschlag – der Kreis schließt sich zurück zum Experiment.

Auf dem Weg vom Experiment durch die Theorie zurück zu experimentellen Fragestellungen konnte ich mein Steckenpferd, par force, einmal quer durch die Technologie-Entwicklung reiten – in meinen Augen ein höchst befriedigender Abschluß eines Ausbildungsabschnitts...

Fällt der Ertrag auch in wissenschaftlicher Hinsicht befriedigend aus?

16.4 Weiterentwicklung der SiGe-Technologie

Ausgangspunkt der Arbeit ist die SiGe-Prozeßintegration. Hier kann ich auf eine verborgene Schwachstelle hinweisen: Die Silizidierung der äußeren Basis steht einer Miniaturisierung – wie sie mit Blick auf die maximale Schwingfrequenz geboten ist – im Wege (im weiteren Verlauf der Arbeit schlage ich einen Herstellungsablauf für den Basisanschluß vor, der die bestehende Schwachstelle abmildert).

Im Zusammenhang mit der SiGe-Prozeßintergation stehen kleinere originäre Arbeiten zur Prozeßintegration, wie das ab S. 20 vorgestellte Verfahren, oxidationsbedingte Verspannungen ohne Micro-Raman-Streuung quantifizieren zu können.

16.5 Physikalische Simulation

Mit hochsperrenden integrierten DMOS habe ich gute Erfahrungen mit der Voraussimulation gemacht (vgl. Kap. 18). Mit SiGe-Transistoren lag im status quo ante weder bei Atmel, noch in einem anderen Industrieunternehmen nennenswerte Simulationserfahrung vor.

Es dauerte über ein Jahr, die Simulationsfähigkeit herzustellen. Dieses Jahr war geprägt von intensiver Koalitionsarbeit und laufender Vorleistung vielen innerbetrieblichen Aufpunkten gegenüber: Ein Mannjahr Entwicklungsressourcen in die Forschung

binden zu können, erfordert ein beträchtliches Defensivgewicht. Die Ergebnisse sprechen für sich: Ausgehend von einem Minimum an experimentellen Daten (Bruchbild und SIMS-Profil) kann ich sämtliche elektrischen Parameter von Atmels SiGe-Technologie innerhalb der experimentellen Unsicherheit reproduzieren. Meines Wissens nach ist dies, zumindest in industriellem Umfeld, ein Novum.

Wesentlich hierfür waren

- die von Hendrik Mau [III.7] übernommenen Materialparameter,
- ein gutes Verständnis des Experiments, v. a. der SIMS-Profilierung,
- wie stets in der Simulation, eine problemangepaßte Gitterdefinition,
- ausführliche physikalische Voruntersuchungen, um das Verhalten des SiGe-Transistors – vor allem im Kollektorbereich – besser zu verstehen.

  Besonders hinweisen möchte ich auf Abb. 6.5, die erlaubt, für Hochfrequenztransistoren eine Aussage zu treffen analog zu Johnsons Grenzwert: Während Johnson für den hochsperrenden Transistor zeigt, Verdoppeln der Betriebsfrequenz halbiere die **Spannungsfestigkeit**, ist beim Hochfrequenz-Bipolartransistor die dreifache **Stromtragfähigkeit** erforderlich, um die Betriebsfrequenz verdoppeln zu können[54].

Die Nachsimulation bestätigt die Bedenken bezüglich einer weiteren SiGe-Miniaturisierung, ausgehend von der bestehenden Technologiearchitektur.

16.6 Schaltungstechnik

In einem dritten Handlungsstrang nähere ich mich der SiGe-Technologie von der Anwendung her. Ausgehend von den schaltungstechnischen Anforderungen der Zielanwendung „Mobilkommunikation“ entwickle ich schaltungstechnische Lösungen, um die Beschränkungen der Technologie umgehen zu können.

Ein neues und m.E.n. unerwartetes und interessantes Ergebnis betrifft die Kaskodenschaltung aus zwei Bipolartransistoren: Einen geeigneten Quellwiderstand der Gleichspannungsversorgung der kalten Basis vorausgesetzt, schlägt die Kaskodenschaltung Haken um das Johnson-Limit! Sie ist mit Blick auf die Sperrfähigkeit der Emitterschaltung deutlich überlegen, weist eine höhere Verstärkung und Transitfrequenz auf und erlaubt, eine deutlich höhere HF-Ausgangsleistung an $50\,\Omega$ anzupassen. Und dies – das ist vielleicht besonders überraschend – auch noch mit ausgezeichneter Effizienz!

Die herkömmliche, lateral integrierte Kaskodenschaltung bereitet als Leistungsbauelement Schwierigkeiten durch die erforderliche exzessiver Verdrahtung, deren Parasiten die Hochfrequenztauglichkeit erheblich beschneiden dürften. Ferner lassen sich nicht beide Transistoren getrennt voneinander auf die gewünschten Sperrspannungsbereiche hin optimieren. Dies führt zum Konzept der vertikalen Integration:

[54] In den Abschnitten 6.3 und 6.4 führe ich vor, wie sich hieraus bei kurzer Kollektordiode Aussagen über die Sperrfähigkeit treffen lassen.

16.7 Vorschlag eines neuartigen HF-Bauelments

Die Transistoren geometrisch übereinanderliegend anzuordnen, ergibt eine thyristorähnliche Struktur, zwei npn-Transistoren, deren Basen durch einen parasitären pnp-Transistor miteinander verbunden sind. Einfache physikalische Betrachtungen zeigen, wie sich die Neigung zu thyristorähnlichem Verhalten beherrschen läßt.

Wie sich im Laufe der Arbeit herausgestellt hat, sind Bauelemente mit einem vergleichbaren Wirkprinzip aus der Leistungselektronik bekannt. Ich weise als erster auf die vielversprechenden Eigenschaften im Hochfrequenzbereich hin. Zum anderen entwerfe ich neue Wege der Realisierung, vor allem bei der Kontaktierung der vergrabenen Bauelementstrukturen.

Dies führt uns zum weiteren Handlungsstrang: der Technologiearchitektur für das vorgeschlagene Bauelement.

16.8 Technologiearchitektur

Um die vergrabenen Bauteilstrukturen anschließen zu können, schlage ich die Verwendung vergrabener Silizidstrukturen vor. Umfangreiche Literaturarbeiten dienen dazu, die zu diesem Zweck geeigneten Silizide herauszuarbeiten, sowie im Vergleich mit anderen Technologiearchitekturen die Möglichkeit eines lateralen, festphasenepitaktischen Überwachsens von Siliziden mit monokristallinem Silizium, ausgehend von monokristallinen Saatfenstern, plausibel zu machen.

Darauf aufbauend entwerfe ich eine mögliche Technologiearchitektur, einen Herstellungsablauf für die vertikal integrierte Kaskodenstruktur, welcher sich als Modul in die atmel'sche SiGe-Arbeitsgangfolge einfügt.

16.9 Verknüpfung der einzelnen Handlungsstränge

Die Erfahrung mit der physikalischen Simulation des SiGe-HBT erlauben, die elektrischen Eigenschaften der vertikal intergrierten Kaskodenschaltung vorauszusimulieren. Die aus dem Herstellungsablauf folgenden Bauelementgeometrien und die aus der physikalischen Betrachtung zum Zünden des Thyristors folgenden Dotierstoffprofile dienen als Eingabegrößen für die Simulation. Die Simulation zeigt die Überlegenheit des Frequenz- und Durchbruchverhaltens des projektierten Bauelements gegenüber dem Hetero-Bipolartransistor.

Ausgehend von dem simulierten Bauelementverhalten erstelle ich ein skalierbares Großsignalmodell. Anhand des Modells werden die Dotierstoffverhältnisse im Bauelement weiter optimiert und die Eigenschaften in der Schaltung untersucht. Die Schaltungssimulation verspricht insbesondere ein überragendes Kompressionsverhalten, verbunden mit einer ausgezeichneten Effizienz.

Ausgehend von experimentellen Fragestellungen der Prozeßintegration („Wie verbessere ich die Atmel'sche SiGe-HBT-Technologie?“) schließt sich am Ende der Kreis zu einer völlig neuartigen und hochinteressanten experimentellen Fragestellung, nämlich

der, ob es möglich sei, Silizide, ausgehend von einem Saatfenster, festphasenepitaktisch mit monokristallinem Silizium zu überwachsen.

Freilich steht hier, wie auch bei dem vorgeschlagenen Bauelement als solchem, der Nachweis aus: Mit gewaltiger Gebärde habe ich dem Berg ein Sandkorn hinzugefügt. Aber auch ohne experimentellen Nachweis ist der Innovationsgehalt der vorliegenden „Machbarkeitsstudie“ nicht unerheblich: Aus meinen beiden Jahre bei Herrn Prof. Schumacher in Ulm sind über 20 Patenteinreichungen erwachsen[55]. Und so kann die Arbeit vielleicht auch ohne das Experiment, den unbestechlichen Arbiter Elegantiae, bestehen, ohne zur Gänze zum Witz im Walfischformat zu verkommen.

Wenn nicht, so waren Witz, Phantasie und Spieltrieb seit jeher Anfang und Ende aller wissenschaftlichen Arbeit. Und es ist nicht Phantasos, der Sohn des Schlafes, sondern Thanatos, des Schlafes Bruder, der Antwort gibt auf alle Fragen.

17 Danksagung

Eine Doktorarbeit unterliegt der Fiktion, von der angegebenen Literatur abgesehen, selbständig erstellt worden zu sein. Dies ist natürlich Unsinn! In einem viel höheren Maße als aus der Literatur speist sie sich aus der Erfahrung der Kollegen, von denen ich mein Handwerk gelernt habe und der kritischen Begleitung der Gesprächspartner, die die ärgsten Auswüchse zurechtstutzen.

Zu nennen ist hier als erstes Herr Dr. Jürgen Arndt. Als ich bei Atmel angefangen habe, war er Leiter der Technologieentwicklung. Ich verdanke seiner Unterstützung und Anleitung mehr, als ich hier ausdrücken kann – nicht nur in technologischer Hinsicht! Und nur mithilfe seiner Unterstützung noch aus seinem Ruhestand heraus (zwischen Arbeit auf dem Bauernhof, Cellospiel und Basketball), hat die Firma Atmel mir ermöglicht, meine Bauelementideen im Rahmen einer Promotion weiterzubetreiben.

Als Feierabendprojekt gestartet, hätte die vorliegende Arbeit nicht zustande kommen können ohne die politische Unterstützung durch die Dres. Bertram Koch, Volker Dudek und Matthias Bopp. Auch ihnen gilt mein Dank. Vor allem Bertram hat mir in meiner Zeit bei Atmel ungewöhnliche Freiheit gelassen, meine eigenen Wege zu gehen.

Ohne die moralische und pädagogische Unterstützung durch Herrn Michel Sagebiel bei der Load-Pull-Simulation hätte ich unmöglich den vom Technical Board geforderten Effizienznachweis erbringen können. Vielen Dank Dir, Michael!

---

[55] Naturgemäß berücksichtigt die vorliegende Arbeit vor allem Einreichungen mit Schwerpunkten in der Technologieintegration und den elektronischen Bauelementen. Dagegen finden acht schaltungstechnische Einreichungen (darunter ein Patent aus der Leistungselektronik) keinen Niederschlag.

Eine weitere Erfindung zu elektronischen Entwurfverfahren wird im neuen „power amplifier design guide“ von ADS genutzt, und, von der Bayerischen Forschungsstiftung gefördert, in Zusammenarbeit mit Rohde und Schwarz und dem LTE in Erlangen auf ihren Nutzen in der HF-Meßtechnik hin untersucht. Diese Erfindung wird an anderer Stelle vorgestellt werden.

Ganz wichtig beim Versuch, meine Arbeit in eine lesbare Form zu bringen, war mein Freund und Kollege Jürgen Häfner, der das „Lektorat“ übernommen hat. Schon die Ausarbeitung von Ⓟ 23 , einer Keimzelle der vorliegenden Arbeit, wäre nicht möglich gewesen ohne seine Simulationsunterstützung – herzlichen Dank für Geduld und Überstunden!

Wissenschaft findet im Gespräch statt. Meine Opfer diesbezüglich, vor allem die Herren Peter Brandl und Michael Sagebiehl haben mir durch ihre Zeit, ihre Unterstützung und ihr offenes Ohr Vieles leichter und angenehmer und erfolgreicher gestaltet – vielen Dank Euch!

Das selbe gilt natürlich für die Kollegen am Kompetenzzentrum integrierte Schaltungen in der Kommunikationstechnik in Ulm: den Dres. Andreas Trasser, Ingmar Kallfass, Peter Abele, Martin Häfele, Sebastian Chartier, Jochen Dederer, Nabil Alomari, um – in willkürlicher Reihung und in parte pro toto – nur einige zu nennen.

Herausheben möchte ich hier Herrn Bernd Schleicher für seine fantastische Korrekturarbeit!

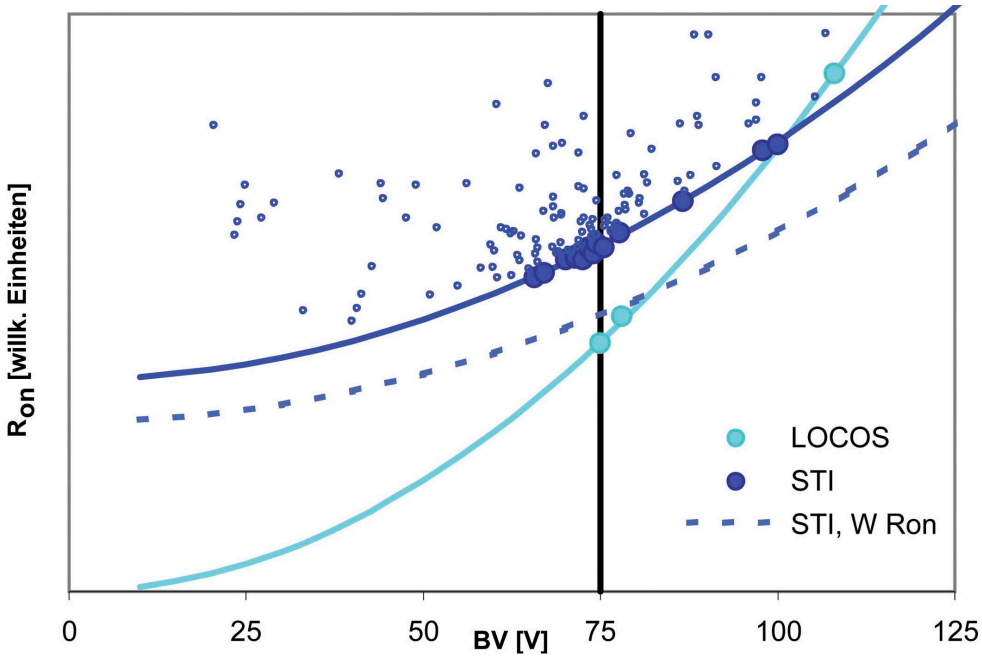
Auch dem Korreferenten, Hern Prof. Bernd Tillack, gilt mein Dank.

Vor allem aber richtet sich mein Dank natürlich an Prof. Dr.-Ing. Hermann Schumacher. Dem forschungspolitischen Rahmen geschuldet, kämpft mein Doktorvater mit unermüdlichem Einsatz um die Mittel, mit denen er unserer Gruppe überhaupt erst die kritische Masse sichern kann. Um so mehr hat mich die Freiheit erstaunt, mit der ich meinen eigenen Themen nachgehen konnte!

Nicht nur am Institut, sondern auch in meiner Familie war ich während meiner Promotion von vielen der sonst üblichen Aufgaben freigestellt. Vor allem hat eine solche Arbeit ihre Höhen und Tiefen, und allzuoft haben meine Frau Daniela, genauso wie Mara, Jakob und Anne, auf den gerechten Anteil von Papas Geduld verzichten müssen. In dieser Hinsicht freue ich mich besonders, daß meine Promotion – manchmal schon fast wider Erwarten – vor dem Abschluß steht!

Zuletzt danke ich meinen Eltern für ihre Unterstützung. Ich weiß nicht, ob ich ohne die tex-nische Unterstützung durch meinen Vater und ohne die Geduld meiner Mutter mit uns beiden mit der Verschriftlichung meiner Arbeit jemals zum Abschluß gekommen wäre: Herzlichen Dank an Euch beide!

VIII Anhang



18 Technologiesimulation

Die Halbleiterei ist konservativ. Von einer Generation zur nächsten sinken die Strukturgrößen, während der grundlegende Aufbau, die Techologiearchitektur, erhalten bleibt. Das Paradebeispiel dafür ist die CMOS-Logik: 1967 wurde von Dr. Else Kooi am Philips Natuurkundig Laboratorium in Eindhoven die Lokale Oxidation von Silizium, LOCOS, entwickelt (vgl. Abschnitt 5.1). 25 Jahre lang stand „integrierte Logik“ synonym für „LOCOS-Isolation“.

Die Idee hinter dem LOCOS-Verfahren ist, eine strukturierte Nitridmaske zu verwenden, um bei einem Ofenprozess die Oxidbildung auf die freiliegende Siliziumoberfläche zu beschränken. Während des Oxidwachstums diffundiert das Oxidationsmittel durch das Oxid und reagiert an der Grenzfläche mit Silizium. Die Diffusion ist nicht gerichtet, das Nitrid wird vom Oxid unterkrochen. Es bildet sich eine charakteristische Oxidkante aus, der „Vogelschnabel“.

Ein CMOS-Chip ist in parallele Streifen aus n- und p-Wannen aufgeteilt. In den n-Wannen werden PMOS, in den p-Wannen NMOS ausgebildet. Die Übergangsbereiche zwischen zwei Wannen werden zur Verdrahtung genutzt. Die Source- bzw. Drain-Gebiete der Transistoren bestehen aus flachen, implantierten Halbleitergebieten in einer gegennamig dotierten Wanne. Zwei Draingebiete sind durch ein LOCOS-Oxid gegeneinander isoliert, welches die Source-Drain-Implantation abschirmt. Das Oxid verbleibt anschließend auf dem Wafer.

Die LOCOS-Oxidation erfordert einen Mindestabstand zweier Sorce/Drain-Gebiete: Es muß dick genug sein, um die Implantation abzuschirmen, wobei die Oxidkanten flach auslaufen. Während die Transistoren mit jeder Generation schrumpfen, läßt sich die LOCOS-Isolation nicht beliebig verkleinern: Bereits bei 0.8 µm ground rule macht die Isolation den größeren Teil der Chipfläche aus!

Mitte der neunziger Jahre hat sich deshalb die CMOS-Architektur grundlegend geändert: Das LOCOS-Oxid wurde durch ein Grabenoxid mit senkrechten Kanten („STI“, Shalow-Trench-Isolation) ersetzt, die Wannenisolation durch eine dielektrische Isolation („deep trenches“, tiefe, oxidgefüllte Schlitze im Halbleiter).

Neben reinem CMOS gibt es Technologien, die CMOS-Logik, HVMOS-Schalter und Bipolartransistoren verbinden („BCD“ für „Bipolar-CMOS-DMOS“). In BCD ist der Logikanteil begrenzt, der Übergang von LOCOS zu STI fand später statt.

Ich befasse mich nachfolgend mit den Schwierigkeiten, in einer BCD-Technologie von der Wannenisolation mit LOCOS auf Trenches und SOI umzustellen. Der wichtigste Punkt ist, daß die STI-Isolation gleichzeitig dem HVMOS als Feldoxid dient. Bei solch einem stark abgeänderten Bauelement oder Herstellungsablauf kann die Simulation ihre Stärken voll ausspielen: Ein Fertigungsdurchlauf erfordert ein halbes Jahr, die Simulation wenige Tage. Viele der Überraschungen lassen sich im Vorfeld des Experiments abfedern.

18.1 Hochsperrende laterale MOSFETS

Um die Aufgabe des Feldoxides zu beschreiben, entwickelt Abb. 6.1 den HVMOS schrittweise aus dem CMOS. Die linke n^+-Region stellt jeweils die Drain, die rechte die Source eines Tranistors dar. Blau sind Gate- bzw. Feldoxide dargestellt. Schwarze durchgezogene Linien bezeichnen pn-Übergänge, in warmen Farbtönen sind Äquipotentiallinien eingezeichnet.

Im hochintegrierten **CMOS** (Abb. 18.1, CMOS) konkurrieren verschiedene Durchbruchmechanismen. Ich beschränke mich auf den **Lawinendurchbruch** der Draindiode bei Transistoren mit Sperrspannungen BV von, typischerweise, bis zu 12 V.

Um höhere Spannungsfestigkeiten zu erreichen, verringert man die Wannendotierung und umgibt die Drain mit einer moderat dotierten „Driftzone“ (Abb. 18.1, DMOS für „diffused MOS“, ursprünglich D^2MOS, Double Diffused MOS: Drain und Driftzone wurden durch zwei Diffusionen eingebracht). Die Driftzone erhöht die Dioden-Durchbruchspannung.

Die Spannung wird auf beiden Seiten des pn-Übergangs gleichmäßig abgebaut. Bis zum pn-Übergang an der Kante des Gateoxides ist die Hälfte der Sperrspannung abgefallen. Die Durchbruchfestigkeit thermischen Oxids liegt bei 1 V/nm, bis zu einem Drittel dieser Feldstärke sind die Oxide dauerfest. Die Gateoxiddicke anzuheben erhöht gleichzeitig die Schwellspannung. Ihre praktischen Grenzen finden **DMOS** daher in der **Spannungsfestigkeit des Gateoxids** bei etwa 20 V.

Um die Sperrfähigkeit weiter zu erhöhen (Abb. 18.1, HVMOS), schließt man an das dünne Gateoxid drainseitig ein Dickoxid an. Man zieht das Gate auf das Dickoxid. Die Gate-Unterseite ist eine Äquipotentialfläche. Als „Feldplatte“ drückt sie die übrigen Äquipotentialflächen vom Dünnoxid weg: Bis zum eigentlichen Gateoxid ist erheblich mehr als die Hälfte der Drainspannung abgebaut, so daß ein 100 V-HVMOS mit der selben Gateoxiddicke auskommt wie ein 5 V-CMOS!

Unter dem n-HVMOS ist zumeist eine vergrabene p^+-Schicht angeordnet, die dazu beiträgt, ein Zünden des parasitären npn-Transistors („Latch-up“) des Bauelementes zu vermeiden. Die Sperrspannung wird zum einen begrenzt von der Durchbruchspannung von der Driftzone auf die vergrabene Schicht.

Vor allem begrenzt bei **HVMOS** der **Einschaltwiderstand**, R_{on}, die Spannungsfestigkeit: Schaltet man drei MOSFETs in Reihe, verdreifachen sich BV und R_{on}. Um den ursprünglichen Einschaltwiderstand beizubehalten, muß sich auch die Gatelänge verdreifachen: Der Flächenbedarf steigt mit dem Quadrat der Durchbruchspannung.

Die wirtschaftliche Grenze, bis zu der HVMOS eingesetzt werden und die technologische Grenze, bis zu der hin sie sich sinnvoll integrieren lassen, liegt bei etwa 120 V. Bei höheren Sperrspannungen wird man zur Zeit auf den Aufbau der Schaltung aus diskreten Bauelementen zurückgreifen.

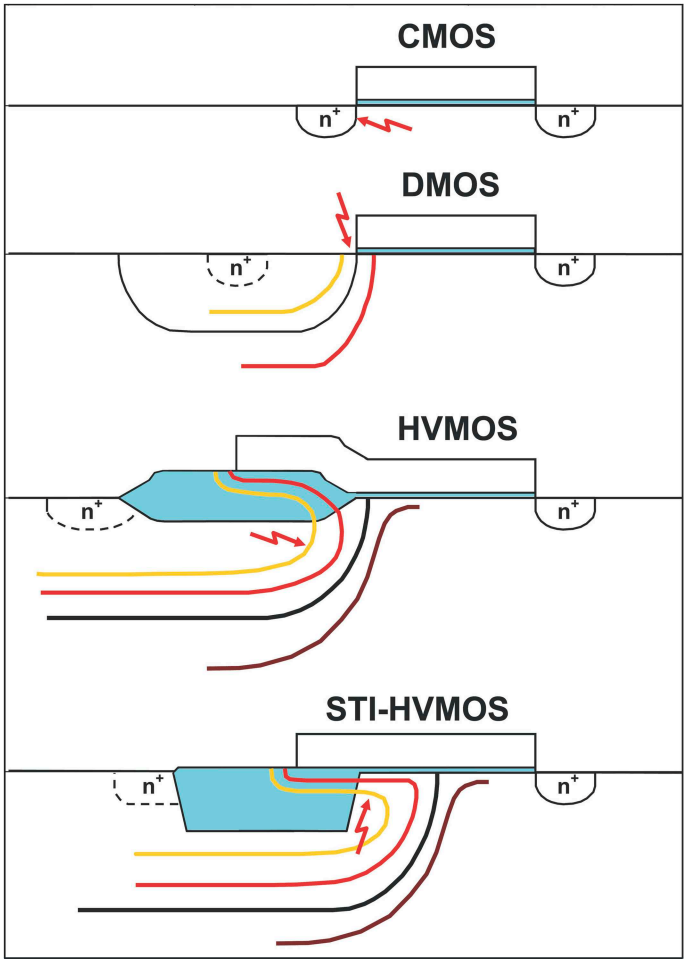


Abb. 18.1: Durchbruchmechanismen bei MOS-Transistoren

18.2 Übergang von LOCOS- auf STI-Oxid

Eine grobe Technologiearchitektur mit geschätzter Bauelementgeometrie dient als Startpunkt für die **Technologiesimulation**. Hierbei wird im Rechner nachgestellt, wie sich bei Implantationen und Ofenschritten beipielsweise die Dotanden im Silizium verteilen. Das Ergebnis ist ein (typischerweise) zweidimensionaler Bauelementquerschnitt. Ausgehend vom Querschnitt mit der Dotierstoffverteilung folgt die **Bauelement-Simulation**, insbesondere die Simulation der Durchbruchspannung.

Für eine feste Bauelementgeometrie, aber unterschiedliche Dotierstoffkonzentrationen in der Driftzone, erwartet man das folgende Verhalten der Durchbruchspannung zwischen Drain und Source: Bei hochdotierter Driftzone kommt es zum Lawinendurchbruch, wobei die Durchbruchspannung mit sinkender Dotierung steigt. Für sehr niedrige Konzentrationen greift dagegen die Raumladungszone bis auf die Source durch. In

diesem Regime fällt die Sperrspannung mit der Konzentration. Dazwischen liegt ein Maximum der Durchbruchspannung, in dessen Nähe ein sinnvoller Bauelementaufbau erreicht wird. Dieses Verhalten zeigt die grüne Kurve in Abb. 18.2 für die schließlich gefundene Technologiearchitektur.

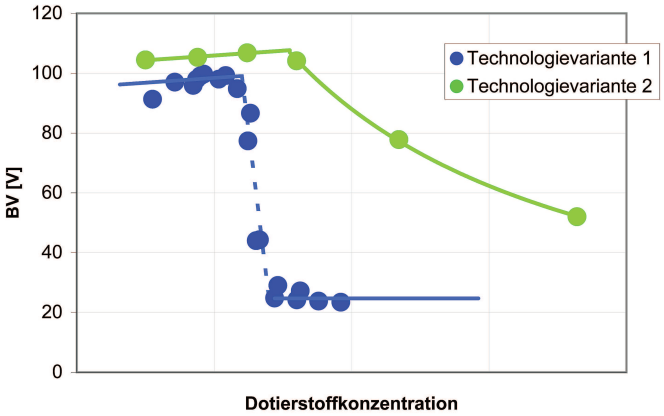


Abb. 18.2: Sperrspannung vs. Driftdotierung für zwei Technologievarianten

Der erste Prozeßablauf, mit dem ich in die Simulation gegangen bin, verhält sich radikal anders (blaue Kurve): Auf den Durchgriff folgt abrupt ein offensichtlich völlig anderer Durchbruchmechanismus, der, unabhängig von der Dotierung der Driftzone, zu einer Spannungsfestigkeit von 25 V führt! (Die Unterschiede im Prozeßablauf zwischen der grünen und der blauen Kurve sind dabei auf den ersten Blick marginal.)

Nun beginnt die eigentliche Arbeit des Simulanten: Wie kann der STI-HVMOS so untypisch früh brechen?

Die zugrundeliegende Physik ist bekannt vom RESURF-Effekt (Reduced Surface Field): Zwei „benachbarte“ Äquipotentiallinien haben einen gegebenen Abstand, gemessen senkrecht zum pn-Übergang. Tritt ein ebener pn-Übergang unter einem flachen Winkel an eine Silizium-Oberfläche, steigt der Abstand zweier Äquipotentiallinien, gemessen entlang der Siliziumoberfläche. Die Feldkomponente parallel zur Oberfläche ist daher kleiner als bei senkrechtem Austritt des pn-Übergangs.

Denselben Fall findet man beim LOCOS-HVMOS, mit vertauschten Rollen (Abb. 18.1, HVMOS): Neben dem pn-Übergang ist auch die Gateoberfläche eine Äquipotentiallinie zu 0 V. Durch den flachen Winkel, in dem der LOCOS-Schnabel ausläuft, ist das Feld entlang der Silizium-Oxid-Grenzfläche reduziert. Anders beim STI-HVMOS: Die Grenzfläche zwischen Silizium und Grabenoxid steht senkrecht auf der Gateoberfläche. Ist die Dotierung entlang der Grabenwand hoch, so daß die anliegende Spannung über einer kurzen Strecke abgebaut wird (und die Entfernung zum pn-Übergang groß), treten hohe Felder entlang der senkrechten Grabenkante auf – Felder, die die Durchbruchfeldstärke überschreiten können! Diese „parasitären“ Felder verlaufen senkrecht zum metallurgischen pn-Übergang und helfen noch nicht einmal beim Abbau der anliegende Sperrspannung. Ohne sorgfältige Kontrolle der Prozeßführung, vor allem von Neigung und Dotierung der Grabenwand, erhält man anstelle

des gewünschten Lawinendurchbruchs bei ca. 80 V einen Feldplattendurchbruch bei 25 V, unabhängig von der Dotierung in der Driftzone, unterhalb des STI-Oxids.

Hätte erst das Experiment das absonderliche Verhalten gezeigt, wären Zeitplan oder gar Erfolg der Technologieentwicklung ernsthaft in Frage gestellt gewesen. Aufgrund der Simulationsergebnisse konnten wir während der Konzeptionsphase auf die Herausforderungen reagieren. Die Erkenntnisse aus meinen Simulationen sind in insgesamt vier Patenteinreichungen eingeflossen, Ⓟ 1, Ⓟ 2, Ⓟ 3 und Ⓟ 4.

Von Herrn Dr. Arndt von Atmel stammt der Hinweis, daß es sich bei dem gefundenen Durchbruchmechanismus um ein bekanntes Phänomen handelt, den Feldplattendurchbruch oder „Insulated Gate Controlled Breakdown“. Er wird in einem Bauelement, der Feldplattendiode, genutzt.

18.3 Verfeinerung der Technologiearchitektur

Unter Vermeidung eines Feldplattendurchbruchs kann die Vorbereitung der Experimente beginnen: Eine große Anzahl von Technologie- und Bauelementvarianten wird simuliert. Zu beachten sind dabei Einschränkungen durch zahlreiche weitere Bauelemente neben dem HVMOS.

Besondere Rücksicht erfordert dabei das Wannenmodul: Zwei gegennamig dotierte Wannen für die Erzeugung von NMOS und PMOS sind durch einen bis auf die vergrabene Oxidschicht reichenden, mit einer Doppelschicht aus Oxid und Polysilizium verfüllten, Graben gegeneinander isoliert. Erfährt der Wafer vor der Eintiefung der Grabenstrukturen eine hohe Temperaturbeaufschlagung, fließen beide Wannen ineinander. Ein CMOS weist dann nahe am Graben eine andere Schwellspannung auf als in der Mitte der Wanne.

Weitere interessante Punkte sind die Einstellung einer einheitlichen und für Hochtemperaturanwendungen geeigneten Schwellspannung der n- wie der p-MOS, sowie ein parasitäres Bauelement, der Feld-MOS: Ohne ausreichend dicke dielektrischen Strukturen und eine ausreichende Dotierstoffkonzentration im Silizium schaltet ein „MOSFET“ mit einer 100 V-Metallbahn als Gateelektrode auf.

Nach einer großen Anzahl von Optimierungsläufen der Simulation ergibt sich das Bild aus Abb. 18.3. Dargestellt sind die Einschaltwiderstände über der Durchbruchspannung (dunkelblaue Kreise). Die durchgezogene dunkelblaue Kurve extrapoliert auf den erzielbaren Einschaltwiderstand der STI-HVMOS. Zum Vergleich sind, hellblau, die Werte des bestehenden status quo ante bei Atmel – HVMOS in LOCOS-Technologie – eingetragen. Wie oben angedeutet, ergibt sich eine quadratische Abhängigkeit des Einschaltwiderstandes von der Durchbruchspannung.

Im STI-HVMOS wird die Sperrspannung im wesentlichen waagrecht unter dem STI-Graben abgebaut, der Strom muß im eingeschalteten Zustand jedoch auch entlang der Grabenwand aufwärts fließen. Der Einschaltwiderstand ist mit STI-Feldoxid höher. Die senkrechte Kante der Feldgebiete erlaubt jedoch, die übrigen lateralen Bauelementabmessungen gegenüber dem LOCOS-HVMOS zu reduzieren: Das Produkt $R_{on} \cdot W$

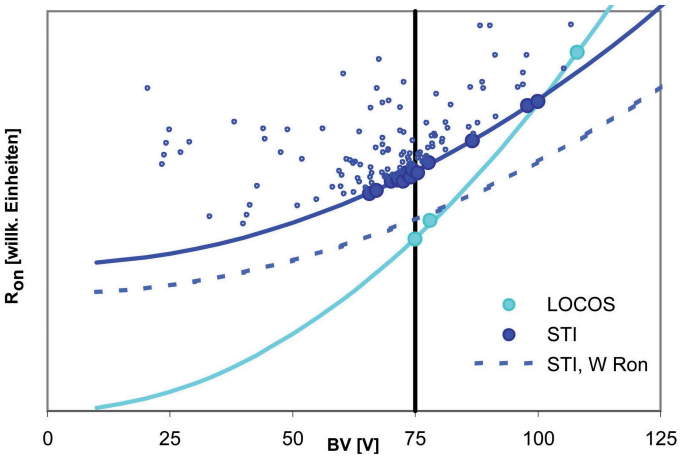


Abb. 18.3: Durchbruchspannungen bei FeldMOS

aus Einschaltwiderstand und Gateweite[56] erscheint in der gewünschten Spannungsfestigkeitsklasse (senkrechte schwarze Linie) konkurrenzfähig.

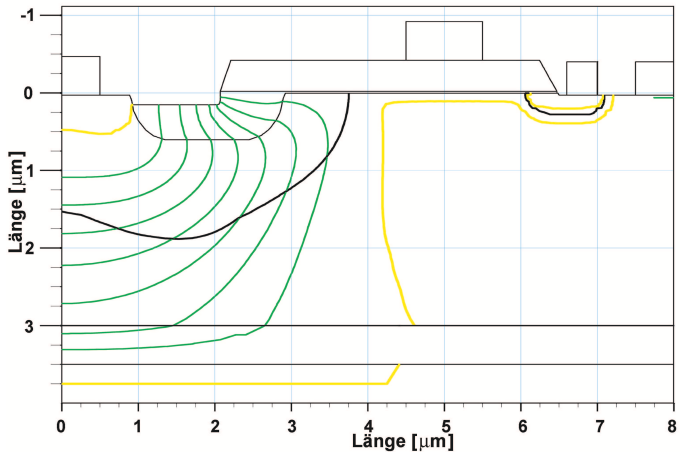


Abb. 18.4: Querschnitt und Feldverlauf STI-FeldMOS

Den Bauelementquerschnitt mit Äquipotentiallinien in ausgeschaltetem Zustand zeigt Abb. 18.4. Die Äquipotentiallinien schneiden den (schwarz dargestellten) metallurgischen pn-Übergang. Hier wird auch der Einfluß des vergrabenen SOI-Oxides genutzt, um die Feldstärke entlang der senkrechten STI-Kante zu verringern.

[56] Anders als im Bipolarbereich bezeichnet „Gate**länge**“ die Ausdehnung eines MOSFETs quer, „Weite“ diejenige längs des Stromflußes.

19 Eingabefiles zur SiGe-Simulation

```
# In einem ersten Schritt parametrisiere ich die
# Querschnittsgeometrie dergestalt, daß sie sich durch
# eine kleine Anzahl von Größen darstellen läßt.
# Diese Größen sind z.B. Dotierstoffgehalte, Dicken
# epitaktischer Schichten und die lateralen Abmessungen. Die
# Parameter lege ich in Quellfiles ab, die vom Eingabeskript
# aus aufgerufen werden:

source set_Dotierung.in # Dotierstoff-Gehalte
source set_1_Bor_Profil.in # Basis-Form
source set_2_mesh_und_devedit.in # Gitter-Parameter
source Define_1_Ge_x_Stufen.in # Lage der Ge-Stufen
source Define_2_Ge_Profil.in # Ge-Gehalte

# Die Definition der Materialparameter für die verschiedenen
# Regionen in Abhängigkeit vom Germanium-Gehalt findet sich in

source Define_3_Material.in

 # Für jede Region enthält dieses File die folgenden Angaben:
 #
 # Zunächst wird ein Fehler im ATLAS- Parametersatz behoben
 # (für die Zustandsdichten bei 300K werden per default die
 # Zahlenwerte bei 0K verwendet):
 #
 # set M_NV300_0=(1e19)*1.04
 # set M_NC300_0=(1e19)*2.89
 #
 # Die vom Germanium-Gehalt abhängigen Größen werden jeweils
 # in der folgenden Form definiert:
 #
 # set x_h=$‘x_01‘‘
 # set funktion_Nv=1-6.3934*$‘x_h‘‘
 # +29.821*$‘x_h‘*$‘x_h‘‘
 # -109.29*$‘x_h‘*$‘x_h‘*$‘x_h‘‘

 # +288.40*$‘x_h‘*$‘x_h‘*$‘x_h‘*$‘x_h‘‘
 # -470.43*$‘x_h‘*$‘x_h‘*$‘x_h‘*$‘x_h‘*$‘x_h‘‘
 # +334.66*$‘x_h‘*$‘x_h‘*$‘x_h‘*$‘x_h‘*$‘x_h‘*$‘x_h‘‘
 # set funktion_Nc=1-0.2648*$‘x_h‘+0.0705*$‘x_h‘*$‘x_h‘‘
 # set nv_01=$‘funktion_Nv‘‘
 # set nc_01=$‘funktion_Nc‘‘

 # set M_eps_01=11.8+(16-11.8)*$‘‘ x_01‘‘
 # set M_EG300_01=1.1242-0.7499*$‘‘ x_01‘‘
 # set M_EGalpha_01=(4.73+(4.77-4.73)*$‘‘ x_01‘‘ )*1e-4
 # set M_EGbeta_01=636+(235-636)*$‘‘ x_01‘‘
 # set M_NC300_01=$‘‘ M_NC300_0‘‘ *$‘‘ nc_01‘‘
 # set M_NV300_01=$‘‘ M_NV300_0‘‘ *$‘‘ nv_01‘‘
```

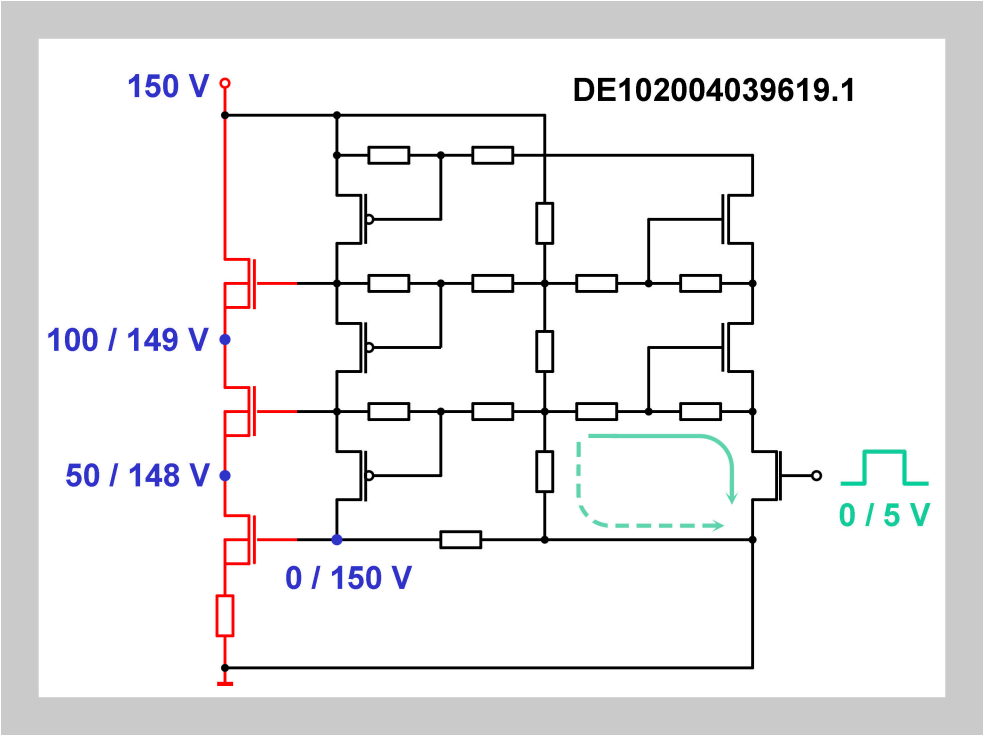
```
 # set vsatn_01=1e7*(1.03-$‘‘ x_01‘‘ /0.2*(1.03-1))
 # set vsatp_01=1e7*(1.03-$‘‘ x_01‘‘ /0.2*(1.03-0.798))

# Aus den Querschnittsparametern werden Parameter abgeleitet,
# die bei der Definition des Gitters helfen:

source Define_ 4_ Geometriehilfen.in

# Anschließend wechsele ich zu ATLAS:

go ATLAS simflags =‘‘-V 5.3.36.C‘‘

# Ich definiere in ATLAS zunächst Regionen, um in Devedit
# mit einem geeigneten mesh-Refinement beginnen zu können.
# Die Re-Definition der Regionen in Devedit
# kann dann mit einem problemangemessenen Gitter starten:
# (Nur mit einem problemangemessenen Gitter erzeugt Devedit
# glatte Grenzen zwischen den Regionen.)

# Das File Mesh.in enthält die Gitterdefinition, die ich in
# Define_ 4-Geometriehilfen.in aus der Querschnitts-
# parametrisierung gewonnen haben.

source Mesh.in

# Regions.in enthält - ebenfalls ausgehend von den
# Querschnits-Parametern - die Definition der Regionen:

source Regions.in

struct outfile=Test0.str

########################################
########################################

go devedit simflags =V 2.4.6.R‘‘

# Die direkte Übergabe der Regionen zwischen ATLAS und
# Devedit funktioniert schlecht; ich lösche daher die
# Struktur, und lese anschließend das gespeicherte
# .str-file ein.

clear
init infile=Test0.str

# Devedit_ meshrefinement.in schreibt die laterale und
# vertikale Auflösung der Triangulationen sowie
# zweckdienliche Seitenverhältnisse der Dreiecke vor:

source Devedit_ meshrefinement.in
mesh
```

```
# Regions.in erzählt DevEdit noch einmal, was es laut
# Handbuch eigentlich von ATLAS überommen haben sollte:
# Welche Region wo liegt und wie sie jeweils heißt.
# (Besonders unangenehm ist, daß DevEdit und ATLAS bei
# der Definition der Regionen eine vollkommen unterschiedliche
# Syntax verwenden!).

source Regions_DevEdit.in

struct outfile=dev_0.str

# Ich gehe zurück zu ATLAS. Dort werden einzelne Regionen
# als Elektroden gekennzeichnet (wieder klappt die Übergabe
# aus DevEdit nicht). Anschließend werden, ausgehend von
# der Bauelement-Parametern, die Dotierstoffprofile definiert:

go ATLAS simflags =V 5.3.36.C‘‘

########################################
########################################

source Electrodes.in
source Doping.in

# Zuletzt setze ich in den einzelnen Regionen die physi-
# kalischen Parameter auf die oben definierten Zahlenwerte:


source Set_parameters.in

# Ich schreibe problemangemessene physikalische Modelle vor,

models conmob fldmob kla consrh auger bgn temperature=290 /
 hcte impact n.concannon p.concannon

# und lege die in den log-files abzulegenden Ausgabe-
# größen fest:

output Band.Param con.band val.band e.field e.temp h.temp /
 e.velocity h.velocity impact e.mobility h.mobility /
 qfn qfp j.electron j.hole j.conduc j.disp j.total/
 permittivity recomb traps taurn taurp u.auger u.srh

# Schließlich wähle ich geeignete numerische Verfahren:

method newton autonr trap atrap=0.5 climit=1e-4 /
 maxtraps=10 itlimit=100

# Die Simulation beginnt mit

solve init
struct outfile=Test.str

# Anschließend werden die einzelnen Elektroden auf den
# gewünschten Arbeitspunkt gerampt. Im Interesse einer
```

```
# beschleunigten Konvergenz kann es hierbei sinnvoller sein,
# die Elektroden reihum ein wenig in der Spannung anzuheben,
# anstatt alle nacheinander von Null auf die Sollspannung zu
# bringen.

log outfile=log1.log j.electron j.hole sim.time
solve vbase=0.01
solve vbase=0.025
solve vbase=0.1 vstep=0.1 vfinal=0.7 name=base
struct outfile=T1.str master
log off

# ...

# Ist der gewünschte Arbeitspunkt erreicht, führe ich
# S-Parameter-Simulationen durch und bin mit der
# Simulation fertig:

log outfile=Gummel.log j.electron j.hole sim.time
solve prev
struct outfile=Gummel_0p70.str master
log off

log outf=fmax_B0p70.log width=1 gains inport=base
     outport=collector master
solve prev ac.analysis freq=1e6 fstep=1.3 mult.f nfstep=53 auto
log off
log outfile=Gummel.log j.electron j.hole sim.time append

solve vbase=0.71
solve vbase=0.72
solve vbase=0.75

struct outfile=Gummel_0p75.str master
log off
log outf=fmax_B0p75.log width=1 gains inport=base
     outport=collector master
solve prev ac.analysis freq=1e6 fstep=1.3 mult.f nfstep=53 auto

 # usf.
```

IX Patenteinreichungen

Christoph Bromberger



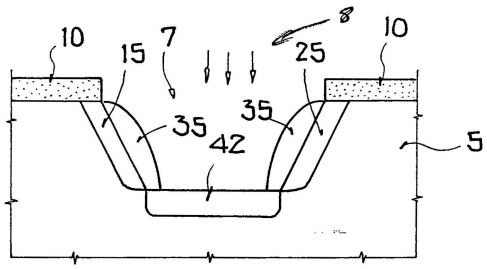
Aber man kann dann wohl verlangen, daß der Kritiker für den Aussagebereich der Theorie adäquate Alternativen entwickelt und sich nicht mit dem Hinweis [darauf] begnügt, [daß] im Verblendungszusammenhang des Spätkapitalismus die Wirklichkeit nicht begriffen werden könne.

Niklas Luhmann, SOZIALE SYSTEME [IX.1]

(P) 1. Verfahren zur Dotierung eines Halbleiterkörpers

C. BROMBERGER, F. DIETZ, V. DUDEK, M. GRAF, J. HERRFURTH, M. KLAUSSNER
EP000001271628, [07.06.2002], DE000010131704, US000006933215, US000006933215

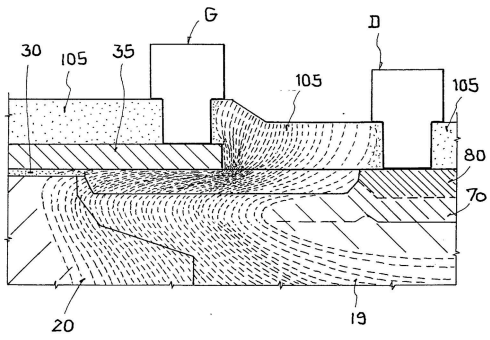
Bei den bisher bekannten Verfahren ist bei einer grabenförmigen Struktur die Dotierung der Seitenwände an die Dotierung des Bodenbereiches gekoppelt. Nach dem neuen Verfahren läßt sich bei einer grabenförmigen Struktur die Dotierung der Seitenwände unabhängig von der Dotierung des Bodenbereichs einstellen. Ferner läßt sich auch die Konzentration der Dotierung zwischen den einzelnen Seitenwänden unterschiedlich einstellen.



(P) 2. Verfahren zur Herstellung eines DMOS-Transistors

C. BROMBERGER, F. DIETZ, V. DUDEK, M. GRAF, J. HERRFURTH, M. KLAUSSNER
EP000001271638, [07.06.2002], DE000010131705, JP002003060204, US000006780713, US020030001198

Bei den bisher bekannten Verfahren ist bei einer grabenförmigen Struktur die Dotierung der Seitenwände an die Dotierung des Bodenbereiches gekoppelt. Nach dem neuen Verfahren läßt sich bei einer grabenförmigen Struktur die Dotierung der Seitenwände unabhängig von der Dotierung des Bodenbereiches einstellen. Ferner läßt sich auch die Dotierung zwischen den Seitenwänden unterschiedlich einstellen. Damit lassen sich bei dem DMOS Transistoren bei niedrigen Einschaltwiderständen hohe Durchbruchspannungen erzeugen und den Flächenverbrauch insbesondere von Treiberstrukturen verringern.

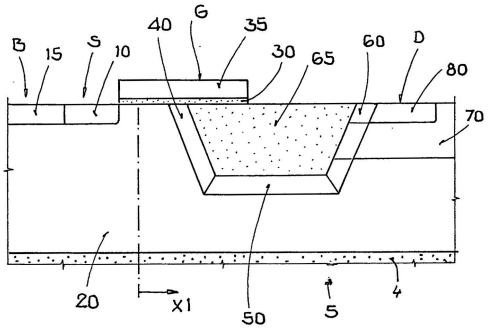


(P) 3. Verfahren zur Herstellung eines DMOS-Transistors

C. BROMBERGER, F. DIETZ, V. DUDEK, M. GRAF, J. HERRFURTH, M. KLAUSSNER
EP000001271638, [07.06.2002], DE000010131706, JP002003060205, US000006878603, US020030003669

Verfahren zur Herstellung eines DMOS-Transistors (100) mit einem Halbleiterkörper (5),

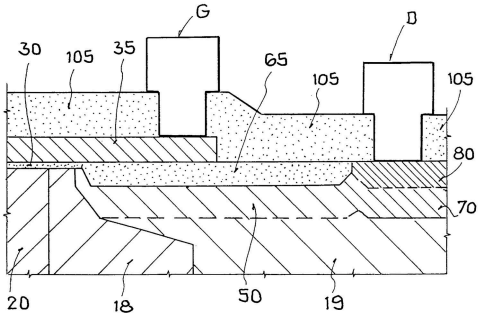
- der eine Oberflächenschicht mit einem Sourcebereich (10) und einem Drainbereich (80) eines zweiten Leitfähigkeitstyps und einen ersten den Sourcebereich (10) umschließenden Wannenbereich (20) eines ersten Leitfähigkeitstyps aufweist, und
- auf der Oberfläche der Oberflächenschicht des Halbleiterkörpers (5) ein Gatebereich (35) ausgebildet ist, der sich beginnend von dem Sourcebereich (10) über wenigstens einen Teil des Wannenbereichs (20) erstreckt, wobei
- ausgehend von der Oberfläche des Halbleiterkörpers (5) in der Oberflächenschicht eine grabenförmige Struktur erzeugt wird, und
- im Bodenbereich (50) der grabenförmigen Struktur eine Dotierung des zweiten Leitfähigkeitstyps mit einem ersten Konzentrationswert erzeugt wird, und
- in der sourceseitigen Seitenwand (40) der grabenförmigen Struktur eine Dotierung des zweiten Leitfähigkeitstyps mit



(P) 4. Verfahren zur Herstellung eines DMOS-Transistors

C. BROMBERGER, F. DIETZ, V. DUDEK, M. GRAF, J. HERRFURTH, M. KLAUSSNER
EP000001271639, [07.06.2002], DE000010131707, US000006806131, US020030003638

Bei den bisher bekannten Verfahren ist bei einer grabenförmigen Struktur die Dotierung der Seitenwände an die Dotierung des Bodenbereiches gekoppelt. Nach dem neuen Verfahren läßt sich bei einer grabenförmigen Struktur die Dotierung der Seitenwände unabhängig von der Dotierung des Bodenbereiches einstellen. Ferner läßt sich auch die Dotierung zwischen den Seitenwänden unterschiedlich einstellen. Damit lassen sich bei den DMOS Transistoren bei niedrigen Einschaltwiderständen hohe Durchbruchspannungen erzeugen und den Flächenverbrauch insbesondere von Treiberstrukturen verringern.

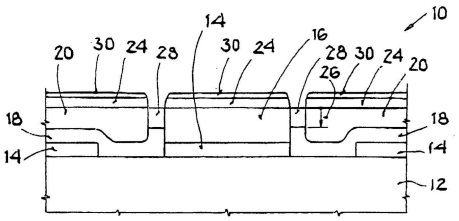


Beschreibung

(P) 5. Verfahren zur Strukturierung eines Halbleiterbauelements

C. BROMBERGER
DE102004013047, [10.03.2004], US020050202669

Vorgestellt wird ein Verfahren zur Strukturierung einer sich in lateraler Richtung erstreckenden ersten Schicht (18) in einem Halbleiterbauelement (10) mit Hilfe einer reaktionsfähigen zweiten Schicht (24), die mit der zu strukturierenden ersten Schicht (18) erste Reaktionsprodukte (28) bildet, die durch einen selektiv auf die ersten Reaktionsprodukte (28) wirkenden Materialabtrag entfernt werden. Das Verfahren zeichnet sich dadurch aus, dass die Strukturierung in vertikaler Richtung erfolgt.



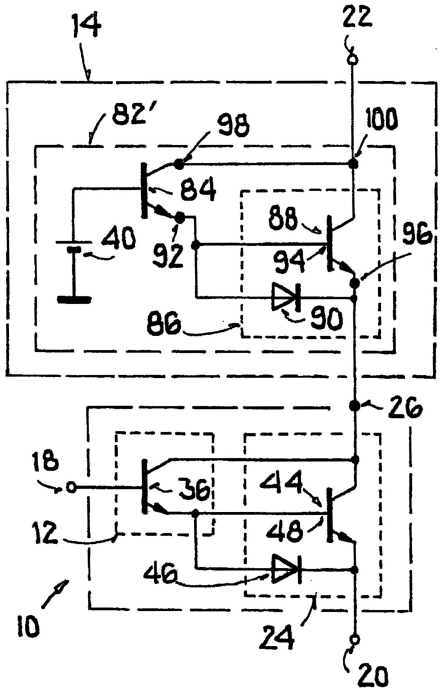
(P) 6. Schaltung zur Erhöhung der Transitfrequenz...

M. JOODAKI, J. BERNTGEN, P. BRANDL, B. KRAUS, C. BROMBERGER
DE102004017165, [01.04.2004], JP002005295568, US000007154340, US020050218995

Vorgestellt wird eine Schaltung (10) mit einem Eingangsverstärker (12) und einem zweiten Verstärker (14), der für eine gegenüber einer Transitfrequenz des Eingangsverstärkers (12) erhöhte Transitfrequenz der Schaltung (10) sorgt, wobei die Schaltung (10) einen mit einem Steuereingang (16) des Eingangsverstärkers (12) gekoppelten Steuereingang (18) und einen ersten Stromanschluss (20) und einen zweiten Stromanschluss (22) aufweist. Die Schaltung (10) zeichnet sich dadurch aus, dass sie einen weiteren Verstärker (24) aufweist, der mit dem zweiten Verstärker (14) in Reihe liegt und der von dem Eingangsverstärker (12) gesteuert wird.

Beschreibung:

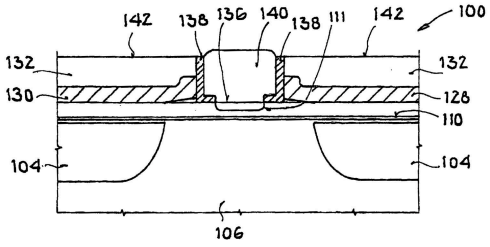
Die Erfindung betrifft eine Schaltung mit einem Eingangsverstärker und einem zweiten Verstärker, der für eine gegenüber einer Transitfrequenz des Eingangsverstärkers erhöhte Transitfrequenz der Schaltung sorgt, wobei die Schaltung einen mit einem Steuereingang des Eingangsverstärkers gekoppelten Steuereingang und einen ersten Stromanschluss und einen zweiten Stromanschluss aufweist.



(P) 7. Verfahren zur Herstellung des Basisanschlusses...

C. BROMBERGER
DE102004017166, [01.04.2004],US020050221570

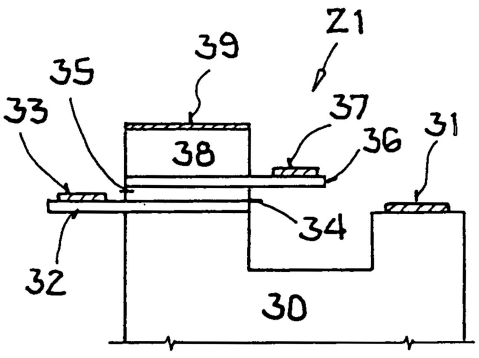
Die Erfindung betrifft ein Verfahren zur Herstellung eines Basisanschlusses eines Bipolartransistors. Das Verfahren zeichnet sich durch folgende Schritte aus: Erzeugen einer Halbleiterstruktur (100), die insbesondere einen Basisbereich enthält; Erzeugen einer zu benachbarten Bereichen selektiv entfernbaren dreidimensionalen Opferstruktur auf der Halbleiterstruktur (100); Abscheiden einer ersten Halbleiterschicht; Abscheiden einer zweiten Schicht aus dielektrischem Material; partielles Freilegen der ersten Halbleiterschicht durch partielles Abtragen der zweiten Schicht; Abscheiden einer ersten Reaktionsschicht, die mit dem Material der ersten Halbleiterschicht selektiv zu benachbarten Bereichen entfernbare Reaktionsprodukte bildet; Entfernen der ersten Reaktionsprodukte, des verbleibenden, nicht mit Material der ersten Halbleiterschicht zur Reaktion gekommenen, Materials der ersten Reaktionsschicht sowie der zweiten Schicht; Abscheiden einer zweiten Reaktionsschicht, die mit dem Material der ersten Halbleiterschicht eine niederresistive Verbindung bildet; Entfernen des verbleibenden, nicht mit Material der ersten Halbleiterschicht zur Reaktion gekommenen, Materials der zweiten Reaktionsschicht; Abscheiden einer weiteren dielektrischen Schicht (132); Freilegen der Opferstruktur durch partielles Abtragen der weiteren dielektrischen Schicht



(P) 8. Verfahren zur Verbesserung elektrischer Eigenschaften...

C. BROMBERGER
DE102004037186, [30.07.2004], EP000001654768, US020060145299, WO002005020330

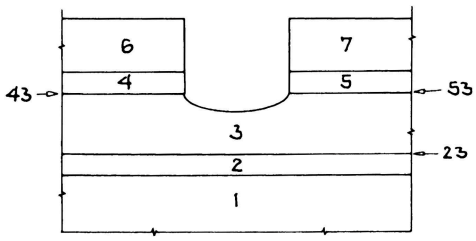
Bei den bisher bekannten Verfahren zur Verbesserung elektrischer Eigenschaften aktiver Bipolarbauelemente wird die Steuerfähigkeit eines Eingangssignals über ein Ausgangssignal stark beeinträchtigt oder es wird, vor allem im Hochfrequenzbereich, das transiente Verhalten bei gegebener Sperrfähigkeit nur geringfügig verbessert. Bei dem neuen Verfahren werden dreischichtige durch fünfschichtige Halbleiteranordnungen ersetzt, während gleichzeitig mithilfe eines Heteroüberganges die Neigung fünfschichtiger Halbleiteranordnungen zu thyristorähnlichem Verhalten unterdrückt wird. Durch das neue Verfahren werden insbesondere die Hochfrequenzeigenschaften und die Sperrfähigkeit aktiver Bipolarbauelemente verbessert, während die Steuerfähigkeit eines Eingangssignals über ein Ausgangssignal weitgehend erhalten wird.



Ⓟ 9. Verfahren zur Integration von drei Bipolartransistoren...

C. BROMBERGER
DE102004037252, [31.07.2004], EP000001628340, US020060022306

Bei den bisher bekannten Verfahren der Integration von Bipolartransistoren wird die Steuerfähigkeit eines Eingangssignals über ein Ausgangssignal stark beeinträchtigt oder es ist, vor allem im Hochfrequenzbereich, die Verbesserung des transienten Verhaltens der integrierten Bipolartransistoren bei gegebener Sperrfähigkeit limitiert. Bei dem neuen Verfahren wird mit Hilfe eines Hetero-Überganges die Neigung mehrschichtiger Halbleiteranordnungen mit den drei Bipolartransistoren zu thyristorähnlichem Verhalten unterdrückt. Durch das neue Verfahren werden insbesondere die Hochfrequenzeigenschaften und die Sperrfähigkeit der Verschaltung der drei Bipolartransistoren flexibler gestaltet, während die Steuerfähigkeit eines Eingangssignals über ein Ausgangssignal



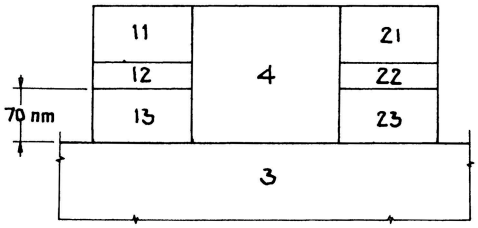
Ⓟ 10. Kaskode, Kaskodenschaltung und Verfahren...

C. BROMBERGER
DE102004038699, [10.08.2004], EP000001626441, US020060033183

Kaskode einer Hochfrequenzschaltung, mit einem ersten Transistor mit

- einem ersten Basishalbleitergebiet,
- einem ersten Kollektorhalbleitergebiet und
- einem ersten Emitterhalbleitergebiet
  und einem zweiten Transistor mit
- einem zweiten Basishalbleitergebiet,
- einem zweiten Kollektorhalbleitergebiet und
- einem zweiten Emitterhalbleitergebiet, dadurch gekennzeichnet, dass das erste Emitterhalbleitergebiet des ersten Transistors und das zweite Kollektorhalbleitergebiet des zweiten Transistors bezüglich einer Waferoberfläche geometrisch oben liegen und

das erste Kollektorhalbleitergebiet des ersten Transistors und das zweite Emitterhalbleitergebiet des zweiten Transistors bezüglich der Waferoberfläche geometrisch unten liegen.
Beschreibung:
[0001] Die vorliegende Erfindung betrifft eine Kaskode, eine Kaskodenschaltung und ein Verfahren zur vertikalen Integration von zwei Bipolartransistoren zu einer Kaskodenanordnung.

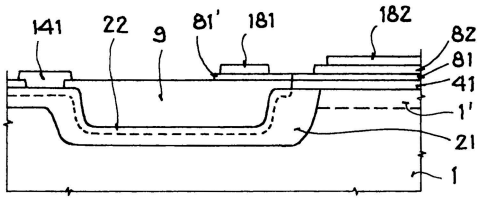


(P) 11. Verfahren zur Herstellung einer Halbleiteranordnung...

C. BROMBERGER
DE102004048332, [05.10.2004], US020060071214

Verfahren zur Herstellung einer Halbleiteranordnung, indem

- eine Silizidschicht formiert wird,
- eine amorphe Halbleiterschicht sowohl auf der Silizidschicht als auch auf einen offenliegenden, an die Silizidschicht angrenzenden, monokristallinen Halbleiterbereich aufgebracht wird und
- in einer nachfolgenden Temperaturbehandlung die amorphe Halbleiterschicht, ausgehend von dem als Kristallisationskeim wirkenden, offenliegenden, monokristallinen Halbleiterbereich, kristallisiert wird, so daß die Silizidschicht zumindest teilweise durch eine kristallisierte, monokristalline Halbleiterschicht bedeckt wird.



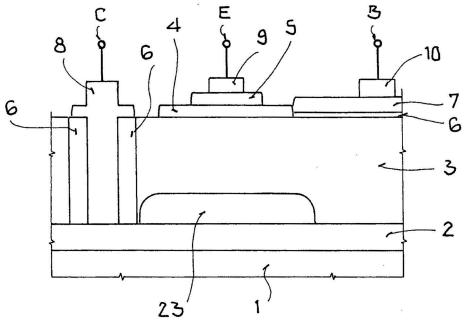
(P) 12. Halbleitergegenstand und Verfahren zur Herstellung

C. BROMBERGER
DE102004050740, [19.10.2004], EP000001650793, US020060081843

Verfahren zur Herstellung eines Halbleitergegenstandes, indem

- eine Silizidschicht (2) aufgebracht wird,
- in die Silizidschicht (2) ein Fremdstoff eingebracht wird, der in einem Halbleitergebiet (3) als Dotand wirkt,
- die Silizidschicht (2) zumindest teilweise unterhalb des an die Silizidschicht angrenzenden, monokristallinen Halbleitergebietes (3) angeordnet wird, so dass die Silizidschicht (2) unter einer Schicht aus dem monokristallinen Halbleitergebiet (3) zumindest teilweise vergraben wird, und
- mittels eines späteren Hochtemperaturschrittes der als Dotand wirkende Fremdstoff zumindest teilweise in das angrenzende monokristalline Halbleitergebiet (3) aus der zu mindest teilweise vergrabenen Silizidschicht (2) eindiffundiert wird.

Beschreibung:
[0001] Die vorliegende Erfindung betrifft einen Halbleitergegenstand und ein Verfahren zur Herstellung des Halbleitergegenstandes.

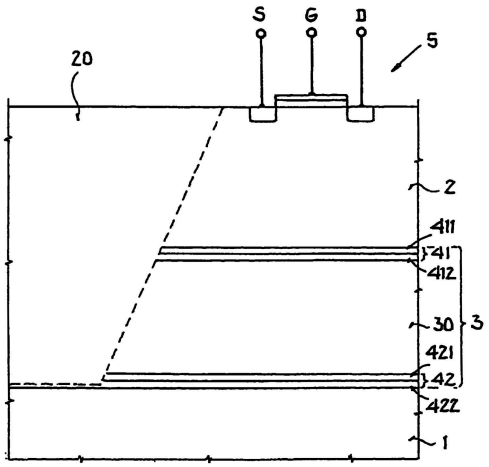


Ⓟ 13. Halbleiteranordnung und Verfahren zur Herstellung...

C. BROMBERGER
DE102004053016, [03.11.2004]

Halbleiteranordnung mit

- einem ersten, monokristallinen Halbleitergebiet,
- einem elektrischen Isolator, der auf das
- erste, monokristalline Halbleitergebiet aufgebracht ist, einem zweiten, monokristallinen Halbleitergebiet, das aus einem amorphen Halbleitermaterial, ausgehend von einer freigelegten Oberfläche des ersten, monokristallinen Halbleitergebietes als Saatfenster zumindest teilweise kristallisiert ist,
  wobei
- das zweite monokristalline Halbleitergebiet teilweise den elektrischen Isolator bedeckt und
- der elektrische Isolator einen Wärmeleitwert größer 20 W/mK aufweist.

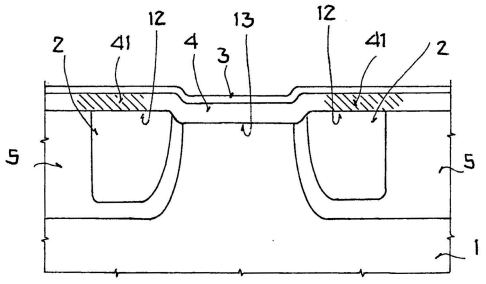


Ⓟ 14. Halbleiteranordnung und Verfahren zur Herstellung...

C. BROMBERGER
DE102004053394, [05.11.2004], EP000001655782, US020060097243

Verfahren zur Herstellung einer Halbleiteranordnung, insbesondere eines Hochfrequenzbipolartransistors, indem folgende Verfahrensschritte in Reihenfolge erfolgen, so dass

- ein Dielektrikum auf einem monokristallinen, ersten Halbleitergebiet eines ersten Leitungstyps erzeugt wird,
- in die Silizidschicht (2) ein Fremdstoff eingebracht wird, der in einem Halbleitergebiet (3) als Dotand wirkt,
- eine Silizidschicht abgeschieden und derart strukturiert wird, dass die Silizidschicht durch das Dielektrikum von dem ersten
- Halbleitergebiet isoliert ist, ein zweites Halbleitergebiet eines zweiten Leitungstyps zur Ausbildung eines Basisbereichs auf das erste Halbleitergebiet und auf die Silizidschicht derart aufgebracht, dass das zweite Halbleitergebiet mit einer ersten Grenzfläche auf dem ersten Halbleitergebiet und mit einer zweiten Grenzfläche auf der Silizidschicht aufliegt.

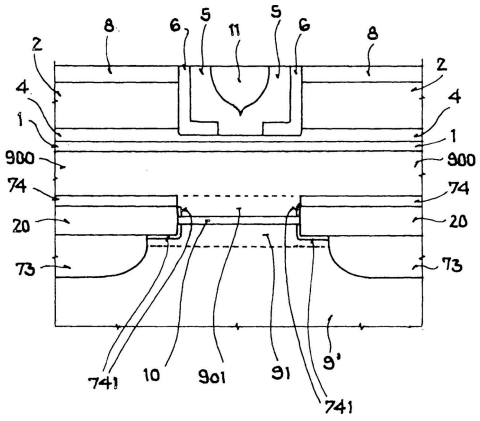


Ⓟ 15. Verfahren zur Herstellung einer vertikal integrierten...

C. BROMBERGER
DE102004053393, [05.11.2004], CN000001790636, EP000001655783, US020060099770

Verfahren zur Herstellung einer vertikal integrierten Kaskodenstruktur, wobei

- ein erster Basisbereich (10) und ein zweiter (1) Basisbereich ausgebildet wird,
- zwischen dem ersten Basisbereich (10) und dem zweiten Basisbereich (1) ein Zwischenbasisbereich (900) ausgebildet wird,
- eine an den ersten Basisbereich (10) angrenzende Anschlussschicht (20) aufgebracht wird, die aus einem Silizidmaterial ausgebildet ist, das mit zumindest einem Fremdstoff versetzt ist, welcher in einem an das Silizidmaterial angrenzen-

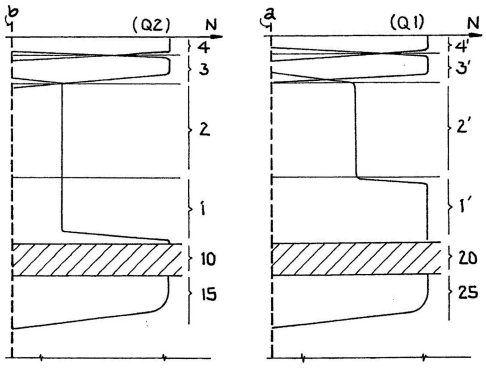


Ⓟ 16. Integrierte Schaltung und Verfahren zur Herstellung...

C. BROMBERGER
DE102004055183, [16.11.2004], EP000001657752, US020060105532

Verfahren zur Herstellung einer integrierten Schaltung auf einem Halbleiterplättchen, wobei

- ein erster Bipolartransistor (Q1) mit einem ersten Kollektorbereich (11, 2') eines ersten Leitungstyps und
- ein zweiter Bipolartransistor (Q2) mit einem zweiten Kollektorbereich (1, 2) dieses ersten Leitungstyps erzeugt werden, indem
- der erste Kollektorbereich (11, 2') durch eine erste Kollektorepitaxie und nachfolgend eine zweite Kollektorepitaxie aufgewachsen wird und
- der zweite Kollektorbereich (1, 2) auch durch diese erste Kollektorepitaxie und diese zweite Kollektorepitaxie aufgewachsen wird und
- in den ersten Kollektorbereich (11, 2') nach dieser ersten Kollektorepitaxie und vor dieser zweiten Kollektorepitaxie Dotanden des ersten Leitungstyps derart eingebracht werden, dass eine erste Dotierstoffkonzentration in einer durch diese erste Kollektorepitaxie aufgewachsenen ersten Epitaxieschicht (11) des ersten Kollektorbereichs (11, 2') eine zweite Dotierstoffkonzentration in einer durch diese erste Kollektorepitaxie



(P) 17. Integrierte Schaltung und Verfahren zur Herstellung...

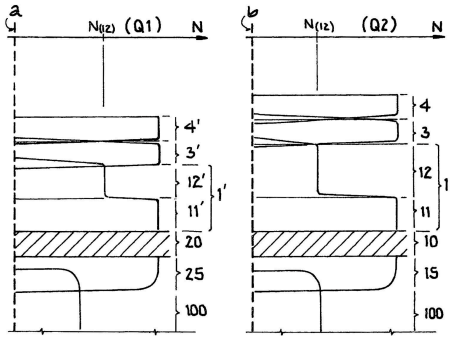
C. BROMBERGER
DE102004055213, [16.11.2004], EP000001657745, US020060105535

Integrierte Schaltung auf einem Halbleiterplättchen, mit

- einem ersten Bipolartransistor (Q1) und
- einem zweiten Bipolartransistor (Q2), wobei
- der erste Bipolartransistor (Q1) einen

  ersten, durch zu mindest eine Epitaxieschicht aufgewachsenen Kollektorbereich (1')

  eines ersten Leitungstyps aufweist und
- der zweite Bipolartransistor (Q2) einen zweiten, durch diese Epitaxieschicht



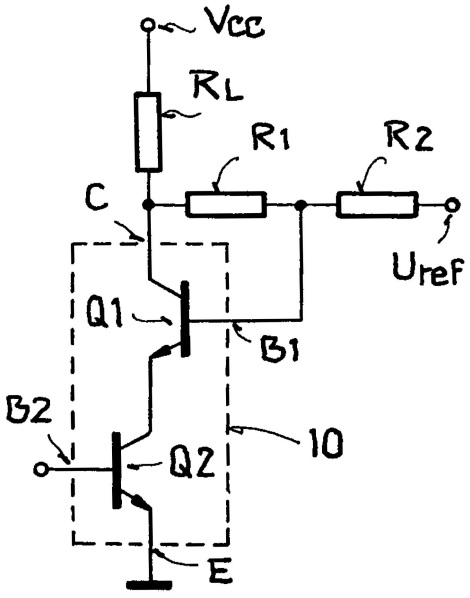
(P) 18. Verstärkerschaltung

C. BROMBERGER
DE102004062135, [23.12.2004], WO002006069736

Verstärkerschaltung zur Verstärkung eines Eingangssignals mit einer vertikal integrierten Kaskode, die

- ein Kollektorhalbleitergebiet eines Kollektors,
- ein erstes, an das Kollektorhalbleitergebiet

  biet einer ersten Basis,
- ein zweites Basishalbleitergebiet einer zweiten Basis,
- ein sowohl an das erste Basishalbleitergebiet als auch an das zweite Basishalbleitergebiet angrenzendes Zwischenbasishalbleitergebiet und
- ein an das zweite Basishalbleitergebiet angrenzendes Emitterhalbleitergebiet eines Emitters aufweist, wobei
- ein Signaleingang mit der zweiten Basis verbunden ist und
- die erste Basis sowohl mit einer, von dem Eingangssignal unabhängigen Spannungsquelle als auch mit dem Kollektor elektrisch gekoppelt ist.



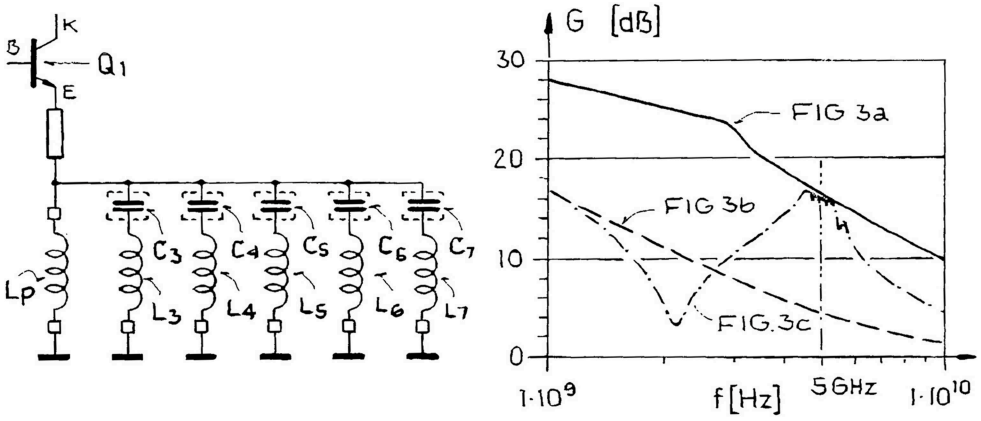
Die Verstärkerschaltung wird vorzugsweise in

(P) 19. Hochfrequenzanordnung

C. BROMBERGER
DE102005008195, [23.02.2005], EP000001696487, US020060186532

Hochfrequenzanordnung mit

- einem integrierten Hochfrequenzschaltkreis,
- einem ersten Bondpad, der über eine erste elektrische Zuleitung, insbesondere einen Bonddraht und/oder einen Lötbump, mit einem Gehäuseanschluss und/oder einem weiteren Schaltkreis elektrisch verbunden ist, wobei
- das erste Bondpad an ein Dielektrikum grenzt, so dass das erste Bondpad mit dem Dielektrikum und einem elektrisch leitfähigen Gebiet des integrierten Hochfrequenzschaltkreises eine erste Kapazität bildet, und



(P) 20. Verfahren zur Integration von zwei Bipolartransistoren...

C. BROMBERGER
DE102005009725, [03.03.2005], EP000001701386, US020060199348

Verfahren zur Integration von zwei Bipolartransistoren in einen Halbleiterkörper (1), indem

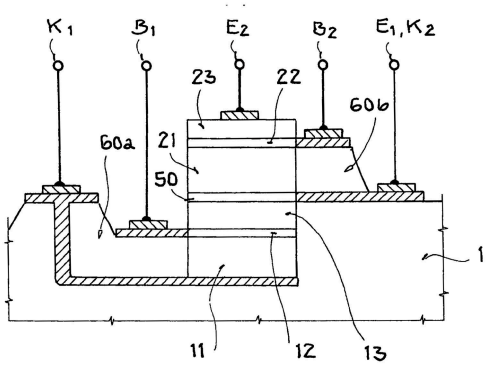
- für den ersten Bipolartransistor

a) ein erstes Emitterhalbleitergebiet (13),

b) ein erstes Basishalbleitergebiet (12) und

c) ein erstes Kollektorhalbleitergebiet (11) erzeugt werden,

- auf den ersten Bipolartransistor eine Rekombinationsschicht (50) aufgebracht wird, die an das erste Emitterhalbleitergebiet (13) oder das erste Kollektorhalbleitergebiet (11) grenzt und derartig ausgebildet wird, dass Ladungsträger an der Rekombinationsschicht (50) rekombinieren und,



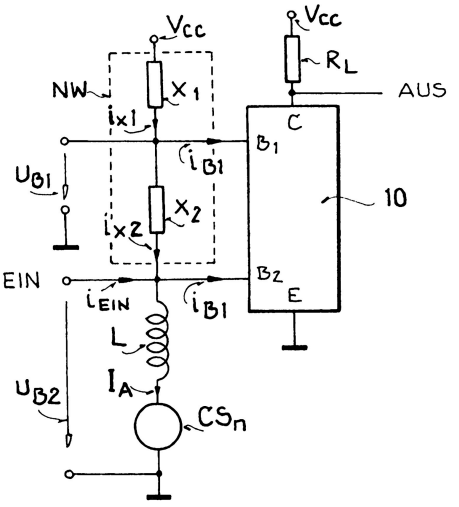
(P) 21. Verstärkerschaltung

C. BROMBERGER
DE102005013385, [23.03.2005], WO002006099993

Verstärkerschaltung zur Verstärkung eines Eingangssignals (i_{EIN}) mit einer vertikal integrierten Kaskode (10), die

- ein Kollektorhalbleitergebiet (1)

  eines Kollektors (C),
- ein erstes, an das Kollektorhalbleiterge-
- ein Signaleingang (EIN) mit der zweiten Basis (B2) verbunden ist und
- die erste Basis (B1) über ein Netzwerk (NW) derart mit der zweiten Basis (B2) verbunden ist, dass eine Kleinsignalspannung (u−B1) an der ersten Basis (B1) mit einer Kleinsignalspannung (u−B2) an der zweiten Basis (B2) und/oder ein Kleinsignalstrom (iB1) durch die erste Basis (B1) mit einem Kleinsignalstrom (iB2) durch die zweite Basis (B2) gekoppelt ist.



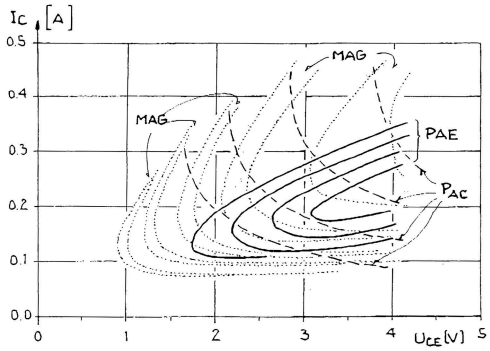
(P) 22. Verfahren zum Entwurf einer Schaltung...

C. BROMBERGER
DE102005016459, [11.04.2005], EP000001713014, US020060230365

Verfahren zum Entwurf einer Schaltung, insbesondere mit einem aktiven Bauelement, vorzugsweise eines Hochfrequenzschaltkreises, indem

- eine Vielzahl (Gnm) von Lastlinien (G)

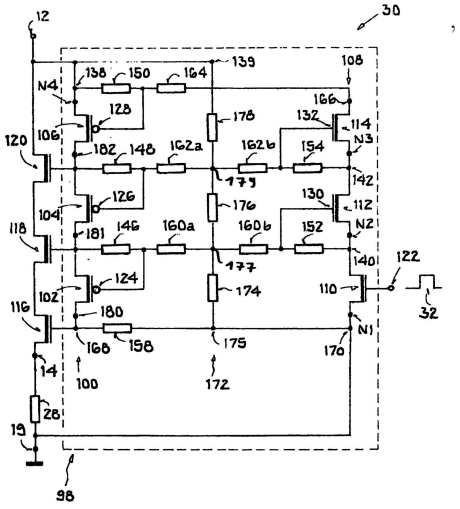
  zumindest näherungsweise ermittelt wird,
- ein Verlauf einer Kleinsignal-Kenngrösse entlang einer jeden Lastlinie (G) zumindest näherungsweise bestimmt;
- ein Bereich (B) einer jeden Lastlinie (G) wird als eine durch eine charakteristische, von der Kleinsignal-Kenngröße abhängende Eigenschaft ausgezeichnete Teilmenge der Lastlinie (G) bestimmt;
- eine der Lastlinie (G) zugeordnete



(P) 23. Schaltungsanordnung mit aktiven Bauelementen...

C. BROMBERGER, V. DUDEK
EP000001624571, [02.08.2005], DE102004039619

Vorgestellt wird eine integrierte Schaltungsanordnung (10) mit einem ersten Schaltungsanschluss (12), einem zweiten Schaltungsanschluss



,

wobei die aktiven Bauelemente (16, 18) jeweils einen Eingangsanschluss (20, 22) und einen Ausgangsanschluss (24, 26) und eine vorgegebene maximale Sperrspannung zwischen Eingangsanschluss (20, 22) und Ausgangsanschluss (24, 26) aufweisen und wobei ein Maximalwert eines zwischen dem ersten Schaltungsanschluss (12) und dem zweiten Schaltungsanschluss (14) erzielten Spannungshubes größer als die vorgegebene maximale Sperrspannung ist. Die Schaltungsanordnung (10) zeichnet sich dadurch aus, dass ein Eingangsanschluss (22) eines n-ten aktiven Bauelements (18) mit einem Ausgangsanschluss (24) eines (n - 1)-ten aktiven Bauelements (16) verbunden ist und dass die Schaltungsanordnung

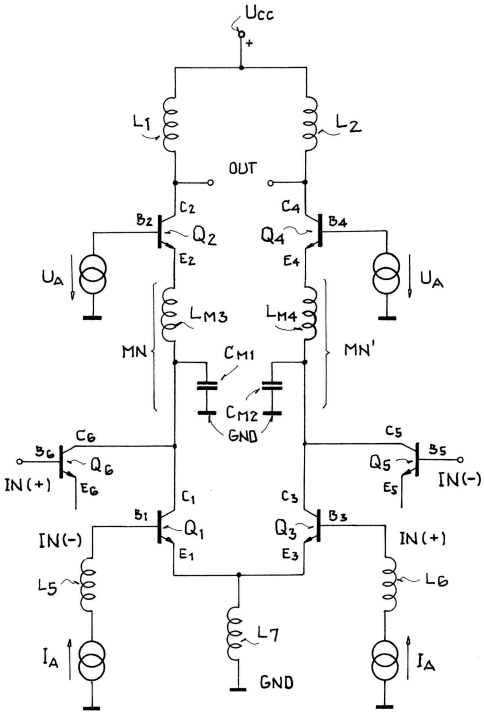
(P) 24. Selektivverstärker

C. BROMBERGER
DE102005062764, [28.12.2005]

Selektivverstärker

- mit einem ersten Transistor (Q1),

  der einen ersten Emitter (E1), eine erste Basis (B1) und einen ersten Kollektor (C1) aufweist,
- mit einem zweiten Transistor (Q2), der einen zweiten Emitter (E2), eine zweite Basis (B2) und einen zweiten Kollektor (C2) aufweist, und
- mit einem Anpassungsnetzwerk (MN), bei dem der erste Transistor (Q1) in Emitterschaltung verschaltet ist und das Eingangssignal (IN) auf die erste Basis (B1) des ersten Transistors (Q1) wirkt,
- bei dem der zweite Transistor (Q2) in Basisschaltung verschaltet ist und das Ausgangssignal (OUT) am zweiten Kollektor (C2) des zweiten Transistors (Q2)

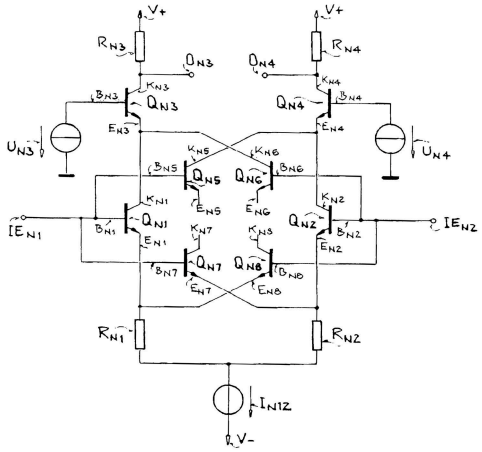


(P) 25. Kaskoden-Differenzverstärker und Differenzverstärker

C. BROMBERGER
DE102005062767, [28.12.2005]

Kaskoden-Differenzverstärker

- mit einem ersten Transistor (Q_{P1}, Q_{N1}) mit einer ersten Basis (B_{P1}, B_{N1}), einem ersten Kollektor (K_{P1}, K_{N1}) und einem ersten Emitter (E_{P1}, E_{N1}), wobei die erste Basis (B_{P1}, B_{N1}) mit einem ersten Eingang (IE_{P1}, IE_{N1}) verbunden ist,
- mit einem zweiten Transistor (Q_{P2}, Q_{N2}) mit einer zweiten Basis (B_{P2}, B_{N2}), einem zweiten Kollektor (K_{P2}, K_{N2}) und einem zweiten Emitter (E_{P2}, E_{N2}), wobei die zweite Basis (B_{P2}, B_{N2}) mit einem zweiten Eingang (IE_{P2}, IE_{N2}) verbunden ist,
- mit einem dritten Transistor (Q_{P3}, Q_{N3}) mit einer dritten Basis (B_{P3}, B_{N3}), einem dritten Kollektor (K_{P3}, K_{N3}) und einem dritten Emitter (E_{P3}, E_{N3}), wobei der dritte Transistor (Q_{P3}, Q_{N3}) mit dem ersten Transistor (Q_{P1}, Q_{N1}) zu einer ersten Kaskodeanordnung verbunden ist,
- mit einem vierten Transistor (Q_{P4}, Q_{N4})

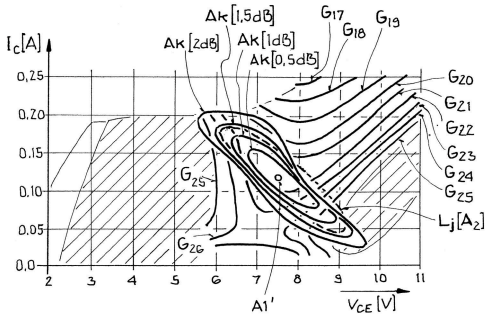


(P) 26. Schaltung

C. BROMBERGER
DE102005062766, [28.12.2005]

Schaltung mit einem ersten Transistor in Basisschaltung,

- bei der der erste Transistor (QN3, Q1) in Basisschaltung in einem derartigen Arbeitspunkt (A1, A-T) mit derartigen Abschlussbedingungen betrieben ist, dass ein Gang einer Leistungsverstärkung(G1(L1)) über einer Aussteuerung positiv ist,
- wobei der Arbeitspunkt (A1, AI') in einem Bereich eines lokalen Minimums der Kleinsignal-Leistungsverstärkung (G1(L1)) entlang einer zu den Abschlussbedingungen gehörigen Ausgangskennlinie des Transistors (QMS, Q1) in Basisschaltung eingestellt ist.



Schlagwortverzeichnis

ATLAS	Bauelemente Simulator	der Firma Silvaco
BCDMOS	Atmel-Prozeß	0.8 µm Hochspannungs-Prozess
BiCMOS2	Atmel-Prozeß	verbindet 1 µm Bipolar- und CMOS-Technologie
BHF	Buffered HF	Mischung aus NH_4F, HF, H_2O, mit stark [sehr stark] reduzierter Ätzrate von SiO_2 [Silizium]
BLAZE	Heterostruktur Simulator	der Firma Silvaco
BMFT	Bundesministerium	für Forschung und Technologie
CAD	Computer Aided Design	am PC konstruieren
CMP	Chemisch-Mechanisches Polieren	
CMOS	Complementary Metal Oxid Semiconductor	
CVD	Chemical Vapour Deposition	Vakuum-Auftragprozess
DMOS	Diffused Metal Oxid Semiconductor	
ELDO	Programm zur Schaltungssimulation	
EOR	End-of-Range-Damage	
FET	Field Effect Transistor	
HBT	Hetero Bipolar Transistor	
HIPR	Lackmaske aus HIPR-6512	photoresist (Hunt Chemical)
HVMOS	High Voltage MOS	
IMD	Inter-Metal Dielectric	Abstandsschicht
IZM	Institut für Zuverlässigkeit und Mikrointegration	Fraunhofer Institut
LEO	Lateral Epitaxial Overgrowth	
LNA	Low Noise Amplifier	
LPCVD	Low Pressure CVD	
LOCOS	Local Oxidation of Silicon	
MAG	Maximum Available Gain	
MAPLE	Computeralgebrasystem	
MESA	engl. f. „Tafelberg“	hochstehendes Aktivgebiet
MSG	Maximum Stable Gain	
MOSFET	Metal Oxide Semiconductor FET	
NMOS	N-channel FET	
OISF	Oxygen-Induced Stacking Faults	

PCM	Process Control Monitor	ein Satz von Teststrukturen zur Fertigungsüberwachung
PECVD	Plasma Enhanced Chemical Vapor Deposition	
PE-TEOS	Plasma Enhaced Tetra-Ethyl-Ortho-Silanoxid	
Pinranha	$H_2SO_5 = H_2SO_4 + H_2O_2$ starkes Oxidationsmittel, frißt alles Organische ratzeputz auf	
PMOS	P-channel Metal-Oxide Semiconductor FET	
REM	Raster Elektronen Mikroskop	
RESURF	Reduced Surface Field	
RTA	Rapid Thermal Annealing	kurzzeitige Temperatur-behandlung
RTO	Rapid Thermal Oxidation	
SADS	Silicide As Diffusion Source	
SiGe	Silicon Germanium	
SiGe2RF	Atmel-Prozeß	SiGe-HF-Technologie
SIMS	Sekundär-Ionen Massen-Spektrometer	
SOI	Silicon On Insulator	
STI	Shallow Trench Isolation	
TED	Transient-Enhanced Diffusion	
TEOS	Tetra-Ethyl-Ortho-Silanoxid	
TMAH	Tetramethyl-Ammoniumhydroxid	(CH_3) 4NOH
UHF5s	Atmel-Prozeß	Silizium-HF-Technologie
UHF6	Atmel-Prozeß	Silizium-HF-Technologie
XRD	X Ray Diffraction	Röntgendiffraktometrie

Literaturverzeichnis

[0.1] FERNANDO PESSOA: Dichtungen. Aus dem Portugiesischen von Georg Rudolf Lind, Fischer, Frankfurt a. M. (1965)

[0.2] WILLEM, Van den vos Rynaerde – Reinart Fuchs, deutsch von A. BERTELOOT und H.-L. WORM, N. G. Elwert Verlag, Marburg (1982)

[I.1] P. ERONDELL: The French garden: for English Ladyes and Gentlewomen to walke in. Or, a Sommer dayes labour. Being an instruction for the attayning vnto the knowledge of the French Tongue: wherein for the practise thereof, are framed thirteene Dialogues French and English, concerning diuers matters from the rising in the morning till Bed-time. Also the Historie of the Centurion mencioned in the Gospell: in French Verses. Which is an easier and shortter Methode then [illeg.] beene yet set forth, to bring the louers of the French tongue to the perfection of the same. Zitiert nach: M.ST. CLARE BYRNE (Ed.): The Elizabethan Home Discovered in Two Dialogues by Claudius Hollyband and Peter Erondell, London, 1925

[I.2] H. KROEMER, Quasi-electric Fields and Band Offsets: Teaching Electrons new Tricks, Nobel Lecture (2000). See e.g., http:// nobelprize.org/ nobel_prizes/ physics/ laureates/ 2000/ kroemer-lecture.pdf

[I.3] E. O. JOHNSON, Physical Limitations on Frequency and Power Parameters of Transistors, RCA Rev., pp. 163-177 (1965).

[I.4] A. OURMAZD, F. WEINL: Annual Report (Sept. 2000 – Aug. 2001) des Instituts für Halbleiterphysik, Frankfurt (Oder) (2001) *„Fig. 25 demonstrates the figure of merit exceeding the theoretical frequency and power limit described by Johnson in 1965.“*

[I.5] A. SCHMIDT: Aus dem Leben eines Fauns, Sämtliche Romane und Erzählungen 1946-1964, Haffmanns Verlag 2000, S. 301

[I.6] E. SÖNMEZ, S. CHARTIER, A. TRASSER, H. SCHUMACHER: Isolation Issues in multifunctional Si/SiGe ICs at 24 GHz, 2005 IEEE MTT-S International Microwave Symposium digest, Vol. 1, S. 169-172 (2005)

[I.7] J.S. HAMEL, S. STEFANOU, M. BAIN, B.M. ARMSTRONG, H.S. GAMBLE: Substrate cross-talk suppression capability of silicon-on-insulator substrates with buried ground planes (GPSOI), IEEE Microw. Guid. Wav. Lett., vol. 10, pp. 134-135 (2000); zitiert nach: S. HALL, B. ECCLESTON: Silicon-germanium for ULSI, Journal of Telecommunications and Information Technology, No. 3/4, Vol. 2000, S. 3-9, ed. National Institute of Telecommunications, Warschau, Polen (2000)

[I.8] S. ZIMMERMANN, Q.T. ZHAO, H. HÖHNEMANN, M. WIEMER, C. KAUFMANN, S. MANTL, V. DUDEK, T. GESSNER: Different approaches to integrate patterned buried $CoSi_2$ layers in SOI substrates, Proceedings of the ninth european workshop on materials for advanced metallization 2005, Microelectronic Engineering, Vol. 82, No. 3-4, S. 454-459 (2005)

[II.1] Akronym F.W.: Schreiben an die königliche Academie der Wissenschaften zu Berlin, Die Untersuchung der Electrizität / und den von ihr bestimmten Preisz betreffend., Prag, den 15. Jan. 1745.

[II.2] E. SÁNCHEZ, M. FERNÁNDEZ-BARCIELA, O. MOJÓN: A report on microwave SiGe HBT technology and applications, Report GDAF/ES/20050501 des Departamento de Teoría de la Señal y Comunicaciones, Escuela Técnica Superior de Ingenieros de Telecomunicación, UNIVERSIDAD DE VIGO (ESPAÑA), www.tsc.uvigo.es/DAF/Investigacion/index (2005)

[II.3] J.D. CRESSLER, G. NIU: Silicon-Germanium Heterojunction Bipolar Transistors, Artech House Publishers, ISBN-13: 978-1580533614 (2003)

[II.4] P. ASHBURN: SiGe Heterojunction Bipolar Transistors, Wiley & Sons (2003)

[II.5] SSuprem4 2D Process Simulation Software, www.silvaco.com/products/vwf/athena/ss4/ssuprem4_ hires.pdf

[II.6] R.W. MANN, L.A. CLEVENGER, P.D. AGNELLO, F.R. WHITE: Silicides and local interconnections for high-perfomanec VLSI applications, IBM J. Res. Develop. Vol. 39, No. 4, S. 403-417 (July 1995)

[II.7] T.A. JAIN and C.R. KAO: Phase Equilibria of the Si-Ge-Ti System Relevant to the Reactions between SiGe Alloys and Ti, Journal of Electronic Materials, Vol. 27, No. 7, S. 842-847 (1998)

[III.1] LUKIAN: Wahre Geschichten

[III.2] S.M. SZE: Physics of Semiconductor Devices, J. Whiley & Sons (1981)

[III.3] C.T. Kirk, Theory of transistor cutoff frequency falloff at high current densities, IRE Trans. Elect. Dev., Vol. 3, S. 164-170 (1964)

[III.4] A. J. JOSEPH, J.D. CRESSLER, D.M. RICHEY, G. NIU: Optimization of SiGe HBTs for operation at high current densities, IEEE Trans. Electron Devices, Vol. 46, No. 7, S. 1347-1354 (1999)

[III.5] K. LEE, D-H. CHO, K.-W. PARK, K. BUMMAN: Improved VBIC Model for SiGe HBTs With an Unified Model of Heterojunction Barrier Effects, IEEE Trans. Electron. Devices, Vol. 53, No. 4, S. 743-752 (2006)

[III.6] M.-W. HSIEH, Y.-M. HSIN, K.-H. LIANG, Y.-J. CHAN, D. TANG: RF Power Characteristics of SiGe HBTs at Cryogenic Temperatures, IEEE Trans. Elec. Dev., Vol. 53, No. 6, S. 1452-1458 (Juni 2006), hierin Abb. 1.

[III.7] H. MAU: Anpassung und Implementation des Energietransportmodells zur vergleichenden Simulation mit dem Drift-Diffusionsmodell an SiGe-Heterobipolartransistoren, Dissertation, TU Ilmenau (1997)

[III.8] L. DEFERM, S. DECOUTERE, C. CLAEYS, G. DECLERCK, J. HEYMAN, D. BEERNAERT: The influence of interface trap density on the performance of bipolar devices, Proceedings of the 1989 Bipolar Circuits and Technology Meeting (BCTM1989), S. 136-139 (1990)

[III.9] T. TAKAGI, K. YUKI, K. TOYODA, Y. KANZAWA, K. KATAYAMA, K. NOZAWA, T. SAITOH, M. KUBO: Reduction of Neutral Base Recombination in Narrow Band-gap SiGeC Base Heterojunction Bipolar Transistors, Proceedings of the 2000 IEEE Bipolar / BiCMOS Circuits and Technology Meeting BCTM, S. 114-117 (2000). Die Effekte, die die Autoren unter dem Begriff „Neutral Base Recombination“ subsummieren, firmieren im allgemeinen unter „Interface Trap Density“; im Lichte der Debatte, welcher von beiden Prozessen im SiGe-Transistor vorherrscht, sollte man beide Begriffe eher auseinanderhalten...

[III.10] G. NIU, J.D. CRESSLER, A.J. JOSEPH: Quantifying Neutral Base Recombination and the Effect of Collector-Base Junction Traps in UHV/CVD SiGe HBT's, IEEE Trans. Elec. Dev., Vol. 45, No. 12, S. 2499-2504 (1998)

[III.11] A. J. JOSEPH, J.D. CRESSLER, D.M. RICHEY, R.C. JAEGER, D.L. HARAME: Neutral Base Recombination and its Influence on the Temperature Dependence of Early Voltage and Cuurent Gain – Early Voltage Product in UHV/CVD SiGe Heterojunction Bipolar Transistors, IEEE Trans. Elec. Dev., Vol 44, No. 3, S. 404-413 (1997)

[IV.1] W. SHAKESPEARE: Julius Cesar, Act 4, Scene 3 (1599)

[IV.2] H. RYSSEL, I. RUGE: Ion Implantation, John Wiley & Sons, ISBN 0 471 10311 X (1986)

[IV.3] L. RUBIN, J. POATE: Ion Implantation in Silicon Technology, The industrial physicist, June/July 2003, S. 12-15 (2003)

[IV.4] LB. Lackdicke und Eindringtiefe der Implantationen

[IV.5] J.O. CHU, D.R. DIMILIA, L. HUANG: Layer transfer of low defect SiGe using an etch-back process, United States Patent 6,890,835, eingereicht durch die International Business Machines Corporation am 19. Oct. 2000

[IV.6] G. TARASCHI: Strained Si-on-insulator development accelerates, Compound Semiconductor, April 2003

[IV.7] Y. SUMIMOTO et al., Extended Abstracts of the Electrochemical Society, 1972, 72-1, p.74; zitiert nach: H. ISHIWARA, T. DAN, K. FUKAO: Selective Surface Doping Method of P Atoms in Lateral Solid Phase Epitaxy..., Jp. J. Appl. Phys. Vol. 31 (1992), pp. 1695-1701

[IV.8] H. SEIDEL et al.: Anisotropic etching of crystalline silicon in alkaline solutions, J. Electrochem. Soc. Vol. 137, S. 3612-3632 (1990)

[IV.9] M.G. ADLERSTEIN: Heterojunction Bipolar Transistors And Method Of Manufacture, United States Patent 4,939,562, eingereicht durch Raytheon Company am 7. April 1989

[IV.10] P.M. ENQUIST, D.B. SLATER, Jr.: Symmetric Self-Aligned Processing, United States Patent 5,318,916, eingereicht durch das Research Triangle Institute am 31. Juli 1992

[IV.11] L. VAN DEN OEVER: The Inverse SiGe Heterojunction Bipolar Transistor, PhD thesis, Delft University of Technology, promotors: J.W. SLOTBOOM, L.K. NANVER, ISBN 90-9015612-7 (2002)

[V.1] A. GRIN, Das Purpursegel, Aus dem Russischen von L. KLEMENTINOWSKAJA, Cotta's Bibliohek der Moderne, Vol. 51, ISBN 3-608-95429-5 (1986) (Claude Louis Berthollet, 1748-1822, entdeckte 1786 das Kaliumchlorat, ein starkes Oxidationsmittel. Er erhöhte durch Kaliumchloratzusatz Brisanz und Treibwirkung französischen Schießpulvers – was nicht nur zahlreichen Soldaten, sondern auch einigen Arbeiter in der Pulvermühle von Essonnes das Leben kostete...)

[V.2] S. PUZZOLO, R. ZAMBRANO, M. PAPARO: Integrated structure of a bipolar power transistor and a low voltage bipolar transistor in the emitter switching or semibridge configurations and associated manufacturing process, Patent EP 0 493 854, eingereicht durch SGS-Thomson Microelectronics S.r.I., Priorität 31. Dez. 1990

[V.3] R.W.C. DEKKER, H.G.R. MAAS, D.J. GRAVESTEIJN, M.P.J. VERSLEIJEN: Cascode circuit structure with epitaxial transistors comprising a low-level base connection, Patnet EP 0 605 920, eingereicht durch Philiips Electronics N.V., Priorität 28. Dez. 1992

[V.4] P.A. CURRAN: Process of making a double Heterojunction 3-D I^2L Bipolar Transistor with a Si/Ge Superlattice, United States Patent 4,771,013, eingereicht durch Texas Instruments Inc. am 1. Aug. 1986

[V.5] S. MANTL, M. HACKE, L. BAY, L. KAPPIUS, ST. MESTERS: Self-ordering of $CoSi_2$ precipitates and epitaxial layer growth of $CoSi_2$ on Si(100), Thin Solid Films, Vol. 321, No. 1-2, S. 251-255 (1998)

[V.6] S. MANTL: Molecular beam allotaxy: a new approach to epitaxial heterostructures, Journal of Physics D: Applied Physics, Vol. 31, No. 1, S. 1-17 (1998)

[V.7] J.D. CRESSLER, G. NIU: Silicon-Germanium Heterojunction Bipolar Transistors, Artech House Publishers, ISBN-13: 978-1580533614 (2003): Dort Fig. 4.5, S. 102

[V.8] IEEE IEDM 2003, Technical Digest, Art. 5.3.1

[V.9] P. BRANDL: Integrierte Halbleiter-Kaskodenschaltung für Hochfrequenzanwendungen, Patenteinreichung DE102004044835, Anmelder ATMEL Germany GmbH (2004)

[V.10] R.P.DICKS, X. WU: Development of Niobium/ Niobium-Silicide Based Alloys for Ultra High Temperature Applications, 17th Advanced Aerospace Materials & Processes (AeroMat) Conference & Exposition, (2006)

[V.11] A.K. VASUDEVAN, J.J. PETROVIC (eds.): High Temperature Structural Silicides, Proceedings of the First Workshop, Gaithersburg, MD, USA, 4-6 November 1991, North-Holland, ISBN: 0-444-89753-4, x + 278 pages (1992)

[V.12] A. VANTOMME, M.-A. NICOLET, N.D. THEODORE: Epitaxial $CoSi_2$ films on Si(100) by solid-phase reaction, J. Appl. Phys., Vol. 75, No. 8, S. 3882-3891 (1994)

[V.13] J. CARDENAS, S.-L. ZHANG, B.G. SVENSSON, C.S. PETERSSON: On the fromtion of inhomogeneities in epitxial $CoSi_2$ layers grown from the interaction of Co/Ti bilayers with Si <100> substrates, J. Appl. Phys., Vol. 80, No. 2, S. 762-768 (1996)

[V.14] T.I. SELINDER, D.J. MILLER, K.E. GRAY: *Insitu* x-ray diffraction study of the role of annealing ambient in epitaxial $CoSi_2$ growth from Co/Ti bilayers on Si(001), App. Phys. Lett., Vol. 67, No. 11, S. 1597-1599 (1995)

[V.15] A.Z. MOSHFEGH, S.J. HASHEMIFAR, O. AKHAVAN: The growth of $CoSi_2$ thin film in Co/W/Si(100) multilayer structures, Solid State Communications, Vol. 128, S. 239-244 (2003)

[V.16] S.L. HSAI, T.Y.TAN, P. SMITH, G.E. MCGUIRE: Resistance and structural stabilities (sic!) [?] of epitaxial $CoSi_2$ films on (001) Si substrates, Journal of Applied Physics, Vol. 72, No. 5, S. 1864-1873 (1992)

[V.17] G. ROTERS: Effect of Various Co/TiN and Co /Ti Layer Stacks and the Salicide Rapid Thermal Porcess Conditions on Cobalt Silicide Formation, 16tes Treffen der Benutzergruppe RTP, Micronas-IMTEK, Freiburg (13.05.2004)

[V.18] G.B. KIM, J.S. KWAK, H.K. BAIK, S.-M. LEE: Effect of Ti-Capping thickness on the dormation of an oxide-interlayer-mediated-epitaxial $CoSi_2$ film by *ex situ* (sic!) annealing, J. Appl. Phys., Vol. 85, No. 3, S. 1503-1507 (1999)

[V.19] J.E. MAHAN, K.M. GEIB, G.Y. ROBINSON, R.G. LONG, Y. XINGUAN, G. BAI, M.-A. NICOLET, M. NATHAN: Epitaxial Films of Semiconducting $FeSi_2$ on (001) Silicon, Appl. Phys. Lett. 56, No. 21, S. 2126-2129 (1990)

[V.20] J.E. MAHAN, J.P. BECKER, R.G. LONG: $ReSi_2$ Thin Film Infrared Detectors, J. Vacuum Sci. Tech. A13, No. 3, S. 1133-1135 (1995)

[V.21] B. GEBHARDT: Wachstum epitaktischer $CoSi_2$-Schichten durch Reaktion metallischer Doppelschichten mit Si(100), Dissertation, TU Chemnitz, 2000

[V.22] A. TRAVLOS, N. SALAMOURAS, E. FLOUDA: Epitaxial erbium silicide films on (100) silicon: growth, structure and electrical properties, Appl. Surf. Sci., 120, No. 3, S. 355-364 (Dec. 1997)

[V.23] J.E. MAHAN, A.VANTOMME, M.-A. NICOLET, R.G. LONG: Reactive Deposition Epitaxy of CrSi 2, Appl. Surf. Sci., 73, S. 146 (1993)

[V.24] J.-Y. TSAI, C.M. OSBURN, S.L. HSIA: Silicidation strategy of sub-0.1 µm junction for deep submicron devices, MRS symposium proceedings“ Silicide Thin Films – Fabrication, Properties, and Applications.“ Vol. 648, S. 245-250 (1995)

[V.25] J.B. LASKY et al. IEEE Trans. Elec. Dev., 38, 262 (1991).

[V.26] T. NGUYEN, H.L. HO, D.E. KOTECKI, T.D. NGUYEN: Reaction study of cobalt and silicon nitride, J. Mater. Res., Vol. 8, No. 9, S. 2354 (1993)

[V.27] K. RIM, J. CHU, K.A. JENKINS, T. KANARSKY, K. LEE, A. MOCUTA, H. ZHU, R. ROY, J. NEWBURY, J. OTT K. PETRARCA, P. MOONEY, D. LACEY, S. KOESTER, K. CHAN, D. BOYD, M. IEONG, H.-S. WONG: Characteristics and Device Design of Sub-100nm strained Si N- and PMOSFETs, 2002 Symposium on VLSI Technology, Digest of Technical Papers, S. 98 [?] (2002)

[V.28] F. FRACASSI et al., J. Electrochem. Soc., 143, 701 (1996)

[V.29] A.E. Morgan et al., J. Electrochem. Soc., 134, 925-935 (1987)

[V.30] L.J.CHEN (ed.): Silicide Technology for Integrated Circuits, 304 pages, IEE (2004)

[V.31] S.-L. ZHANG, M. ÖSTLING: Metal Silicides in CMOS Technology: Past, Present, and Future Trends, Critical Reviews in Solid State and Material Sciences, Vol. 28, No. 1, S. 1-129, (2003)

[V.32] R. HULL, R.M. OSGOOD, H. SAKAKI, A. ZUNGER Semiconducting Silicides, 348 pages, Springer (2000)

[V.33] J.E.MAHAN: Narrow bandgap semiconducting silicides intrinsic infrared detectors on a silicon chip : final report, SuDoc NAS 1.26:190882, NASA (1990), download: http:// ntrs.nasa.gov/ archive/ nasa/ casi.ntrs.nasa.gov/ 19930004242_ 1993004242.pdf

[V.34] S. MANTL, B. SCHULLER, R. CARIUS: Semiconducting silicides: New materials for silicon based optoelectronics, 13th Japan-Germany Forum on Information Technology, Shimagun, Mie, Japan (2000)

[V.35] www.icknowledge.com/ misc_ technology/ Periodic.pdf

[V.36] J. CHEN, C. LI, Z. FU, X. TU, M. SUNDBERG, R. POMPE: Low temperature oxidation of a $MoSi_2$-based material, Mat. Sci. and Engineering A261, S. 239-244 (1999)

[V.37] G.V. SAMSONOV, V.A. LAVRENKO, L.A. GLEBOV: Oxidation of tungsten disilicide in oxygen, Journal Powder Metallurgy and Metal Ceramics, Vol. 13, No. 3, S. 232-235, (1974)

[V.38] K.C.SARASWAT, R.S. NOWICKI, J.F. MOULDER: Thermal oxidation of tantalum silicide in O_2 and H_2O, Appl. Phys. Lett., Vol. 42, IS. 12, S. 1127-1129 (1982)

[V.39] B.L. CROWDER, S. ZIRINSKY: Method for providing a metal silicide layer on a substrate, United States Patent 4180596 (1977)

[V.40] R. GOMEZ-SAN ROMAN, R. PEREZ-CASERO, J. PERRIERE, J.P. ENARD, J.M. MARTINEZ-DUART: Oxidation of tantalum silicide thin films in an RF oxygen plasma, ICV-12: International Vacuum Congress No. 12: Applications of surface science and electronic materials, Den Haag, 1992, Appl. surf. sci., Vol. 70-71 No. 1-4 S. 479-482 (1993)

[V.41] R. PÉREZ-CASERO, J. PERRIÈRE, J.P. ENARD, J.M. MARTÍNEZ-DUART: Plasma oxidation mechanisms in tungsten silicide thin films,Journal of Applied Physics, Vol. 78, No. 1, S. 514-518 (1995)

[V.42] K. MEAX, M. VAN ROSSUM: Properties of Metal Silicides, Electronic Materials Informations Service (Ed.), Inspec Data Review Series Vol. 14, 334 S. (1995)

[V.43] www.semiconductorglossary.com/ default.asp?searchterm=tantalum% 2C+Ta

[V.44] F. LA VIA, C. SPINELLA, E. RIMINI: Arsenic and Boron diffusion in silicon from implanted cobalt silicide layers, Semicond. Si. Technolog., Vol. 10, S1362-1367, (1995)

[V.45] G.B, KIM, J.S. KWAK, H.K. BAIK, S.-M. LEE: Characteristics of diodes prepared using epitaxial $CoSi_2$ as a boron diffusion source, J. Vac. Sci. Technolog. B, Vol. 18, No. 5, S. 2576-2578 (2000)

[V.46] J.-C. HU, Q.-Z. HONG, J.A. KITTL, I.-C. CHEN: Source/drain formation using cobalt silicide as diffusion source for deep sub-micron nMOS, Proc. SPIE – Int. Soc. Opt. Eng. (USA), Vol. 3506, S. 103 (1998)

[V.47] K. MAEX et al., J. Appl. Phys., 66, 5327 (1989)

[V.48] F.C. SHONE et al., Tech. Dig. Int. Elec. Dev. Meet., (1986), p. 407

[V.49] R. BASHIR, A.E. KABIR, F. HEBERT, C. BRACKEN: Tungsten silicide and tungsten polycide anisotropic dry etch process for highly controlled dimensions and profiles, J. Vac. Sci. Technol. B, Vol. 16, No. 4, S. 2118-2120 (1998)

[V.50] F.A. TRUMBORE, Bell Syst. Tech. J., Vol. 39, S. 205 ff. (1960), zitiert nach [V.52]

[V.51] A. MÜHLBAUER, in: O. MADELUNG, M. SCHULZ, and H. WEISS (eds.): Numerical Data and Functional Relationships in Science and Technology, Vol. 17, Chap. 6, Springer, (1984), zitiert nach [V.52]

[V.52] H. BOUDINOV, J.P. DE SOUZA, C.K. SAULB: Enhanced electrical activation of indium coimplanted with carbon in a silicon substrate, J. Appl. Phys., Vol.86, No. 10, S. 5909-5911 (1999)

[V.53] S. SCALESE, M. ITALIA, A. LA MAGNA, G. MANNINO, V. PRIVITERA, M. BERSANI, D. GIUBERTONI, M. BAROZZI, S. SOLMI, P.PICHLER: Diffusion and electrical activation of indium in silicon, J. Appl. Phys., Vol. 93, No. 12, S. 9773-9782 (2003)

[V.54] W.HASSE, J. SCHULTE, J. GRAUL: Electromigration of silicon and phosphorus in tantalum polycide interconnections, IEEE Transactions on Electron Devices, Vol. 34, IS. 3, S. 659-663, (Mar 1987)

[V.55] N.M. ZIMMERMANN, J.A. LIDDLE, A.E. WHITE, K.T.SHORT: Transport in submicrometer buried mesotaxial cobalt silicide wires, Appl. Phys. Lett. , Vol. 62, No. 4, S. 387-389 (1993)

[V.56] W. LEHNHÄUSER: Glasuren und ihre Farben, Wilhelm Knapp Verlag (1973)

[V.57] Y. TAKEMURA: Semiconductor Device Having Transistors With Different Orientations Of Crystal Channel Growth With Respect To Current Carrier Direction, United States Patent 4,939,562, eingereicht durch Semiconductor Engineering Laboratory, Priorität 27. Aug. 1993

[V.58] D.L. LEUNG, R.C. COLE, D.M. COBERT, J.F. KNUDSEN, J.P. HURRELL, D.C. MAYER, R. NEWMAN: A Complementary Metal Oxide Semiconductor Process in Epitaxially Regrown Silicon over Oxide, J. Electrochem. Soc., Vol. 138, No. 12, S. 3771-3777 (1991)

[V.59] A.V. ZOTOV, V.V. KOROBTSOV: Present Status of Solid Phase Epitaxy of Vacuum-Deposited Silicon, J. Crystal Growth, Vol. 98, S. 519-530 (1989)

[V.60] J.L. GLENN, Jr., G.W. NEUDECK: A Double Self-Aligned Silicon Bipolar Transistor Utilizing Selectively-Grown Single Crystal Extrinsic Contacts, Electrical and Computer Engineering (ECE) Technical Reports, TR-EE 92-46, Purdue Libraries, http:// docs.lib.purdue.edu/ ecetr/ 267 (1992)

[V.61] M. SASAKI, T. KATOH, H. ONODA, N. HIRASHITA: Lateral solid phase epitaxy of Si over SiO_2 patterns and its application to silicon-on-insulator transistors, Appl. Phys. Lett. Vol. 49, No. 7, S. 397-399 (1986)

[V.62] K. KUSUKAWA, M. OHKURA, M. MONIWA, M. MIYAO: Effects of thin SiO_2 capping layer on silicon-on-insulator formation by solid-phase epitaxy, Appl. Phys. Lett., Vol. 60, No. 1, S. 80-81 (1992)

[V.63] K. KUSUKAWA, M. MONIWA, E. MURAKAMI, T. WARABISAKO, M. MIYAO: Grown-facet-dependent characteristics of silicon-on-insulator by lateral solid phase epitaxy, Appl. Phys. Lett., Vol. 52, No. 20, S. 1681-1683 (1988)

[V.64] H. ISHIWARA, H. YAMAMOTO, S. FURUKAWA, M. TAMURA, T. TOKUYAMA: Lateral solid phase epitaxy of amorphous Si films on Si substrates with SiO_2 patterns, Appl. Phys. Lett., Vol. 43, No. 11, S. 1028-1030 (1983)

[V.65] M. MONIWA, K. KUSUKAWA, E. MURAKAMI, T. WARABISAKO, M. MIYAO: Influence of Si film thickness on growth enhancement in Si lateral solid phase epitaxy, Appl. Phys. Lett., Vol. 52, No. 21, S. 1788-1790 (1988)

[V.66] K.KUSUKAWA, M. MONIWA, M. OHKURA, E. TAKEDA: Enhancement of lateral solid phase epitaxy over SiO_2 using a densified and thinned amorphous Si layer, Appl. Phys. Lett., Vol. 56, No. 6, S. 560-562 (1990)

[V.67] H. ISHIWARA, A. TAMBA, S. FURUKAWA: Lateral solid phase epitaxy of amorphous Si films onto nonplanar SiO_2 patterns of Si substrates, Appl. Phys. Lett., Vol. 48, No. 12, S. 773-775 (1986)

[V.68] T. DAN, H. ISHIWARA, S. FURUKAWA: Lateral solid phase epitaxy of amorphous Si films by selective surface doping method of P atoms, Appl. Phys. Lett., Vol. 53., No. 26, S. 2626-2628 (1988)

[V.69] H. ISHIWARA, M. TANAKA, S. FURUKAWA: Lateral solid phase epitaxy in selectively P-doped amorphous Si films, Appl. Phys. Lett., Vol. 49, No. 20, S. 1363-1365 (1986)

[V.70] M. MIYAO, M. MONIWA, K. KUSUKAWA, W. SINKE: Low-temperature SOI (Si-on-insulator) formation by lateral solid-phase epitaxy, J. Appl. Phys.,Vol. 64, No. 6, S. 3018-3023 (1988)

[V.71] J.-H. OH, C.-J. KIM, H. ISHIWARA: Enhanced Growth Mechanism in Lateral Solid-Phase Epitaxy of Si-Films Simultaneously Doped with P and Ge Atoms, Jpn. J. Appl. Phys., Vol. 35, pt. 1, No. 3, S. 1605-1619 (1996)

[V.72] H. ISHIWARA, T. DAN, K. FUKAO: Selective Surface Doping Method of P Atoms in Lateral Solid Phase Epitaxy, Jpn. J. Appl. Phys. Vol. 31, pt. 1, No. 6A, S. 1695-1701 (1992)

[V.73] C.A. HEWETT, s. S. LAU, I. SUNI, L.S. HUNG: Phase transformations in ion-irradiated silicides, J. Appl. Phys. Vol. 57, No. 4, S. 1089-1096 (1985)

[V.74] S. SAITO: Method of fabricating Semiconductor Device Having a Salicide Structure, USP 5,759,899, (eingereicht durch NEC in Japan am 30. 01. 1995)

[VI.1] R.M. RILKE: Sonette an Orpheus, (I, 12)

[VI.2] W.J. KLOOSTERMANN, J.A.M. GEELEN, D.B.M. KLAASEN: Efficient Parameter Extraction for the MEXTRAM Model, Proceedings of the 1995 IEEE Bipolar / BiCMOS Circuits and Technology Meeting (BCTM1995), S. 70-73 (1996)
s. auch: W.J. KLOOSTERMANN, J.A.M. GEELEN: Parameter extraction methodology for the mextram bipolar transistor model, Philips Electronics National Laboratory Unclassified Report 003/96 (1996)

[VI.3] J.C.J. PAASCHENS, W.J. KLOOSTERMANN: Derivation of the Model Equations for MEXTRAM, level 503, Philips Electronics National Laboratory Unclassified Report NL-UR 2002/808 (2002)

[VI.4] J.S. DUNN et al.: Foundation of RF CMOS and SiGe BiCMOS technologies, IBM J. Res. & Dev., Vol. 47, No. 2/3, (März/Mai 2003) s. hier insbes. S. 107.

[VII.1] J.W. von GOETHE: Italienische Reise, Neapel, Capua, »Tasso«

[IX.1] N. LUHMANN: Soziale Systeme, Suhrkamp Taschenbuch Wissenschaft, S.9 (1984)

[Busch] W. BUSCH: Der fliegende Frosch. Aus: Hernach, Verlag Joachim, München (1908)

Wenn einer, der mit Mühe kaum
gekrochen ist auf einen Baum
schon meint, daß er ein Vöglein wär -
so irrt sich der!

Wilhem Busch, DER FLIEGENDE FROSCH [Busch]