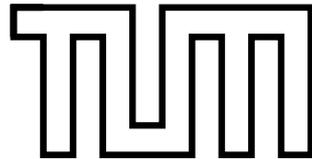


Harald Gunther Kuhn

Physikalische Modellbildung von IGCTs
für die Schaltungssimulation



Technische Universität München

München 2002

Lehrstuhl für Elektrische Antriebssysteme
Technische Universität München

Physikalische Modellbildung von IGCTs für die Schaltungssimulation

Harald Gunther Kuhn

Vollständiger Abdruck der von der Fakultät Elektrotechnik und Informationstechnik
der Technischen Universität München zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs
genehmigten Dissertation.

Vorsitzender: Univ. Prof. Dr. rer. nat. G. Wachutka

Prüfer der Dissertation:

1. Univ. Prof. Dr.-Ing., Dr.-Ing. h.c. D. Schröder
2. Univ. Prof. Dr.-Ing. R. Marquardt;
Universität der Bundeswehr München

Die Dissertation wurde am 18.02.2002 bei der Technischen Universität München
eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am
08.07.2002 angenommen.

Zusammenfassung

Die vorliegende Arbeit befaßt sich mit einem Modell für den Integrated Gate Commutated Transistor (IGCT), das zur rechnergestützten Schaltungssimulation verwendet werden kann. Dabei steht das dynamische Verhalten dieses Leistungsschalters im Vordergrund. Aufgrund bestimmter physikalischer Eigenschaften bipolarer Leistungshalbleiterbauelemente kann auf eine physikalisch basierte Modellbeschreibung nicht verzichtet werden. Die Simulationsergebnisse, die mit dem vorgestellten Modell durchgeführt wurden, stimmen dabei sehr gut mit entsprechenden Meßergebnissen überein.

Dabei ist die Serienschaltung von IGCTs einerseits ein interessanter Anwendungsfall für das Modell und andererseits selbst das Objekt einer weitgehenden Analyse. Es kann mit Hilfe der Simulation gezeigt werden, daß selbst kleinste Unterschiede zwischen den einzelnen Serien-IGCTs bzw. der zugehörigen Ansteuerung, zu erheblichen Asymmetrien in der Spannungsaufteilung beim Abschalten führen. Schließlich wird ein neues Schaltungskonzept zur aktiven Ansteuerung vorgestellt, mit dem diese Asymmetrien vermieden werden können. Auf Entlastungsnetzwerke kann dann weitestgehend verzichtet werden.

Abstract

The thesis presented here deals with a model for the Integrated Gate Commutated Transistor (IGCT), which can be used for circuit simulation. The main interest focusses on the dynamic behaviour of this power electronic device. According to the physical characteristics of bipolar power devices, it is necessary to choose a mathematical approach which is based on semiconductor physics. Calculations performed with the model presented here, show a good agreement with corresponding experimental results.

An interesting application example on the one hand and an object of a detailed analysis on the other hand is the series connection of IGCTs. It can be shown that even small differences between the IGCTs in series, or their gate drives respectively, yield considerable asymmetry regarding the blocking voltages of each single IGCT during turn off. Finally, a new concept for an active drive with a corresponding circuit, that ensures symmetry without applying large scale snubber circuits, is presented.

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als Wissenschaftlicher Mitarbeiter am Lehrstuhl für Elektrische Antriebssysteme der Technischen Universität München und wurde gefördert durch die Deutsche Forschungsgemeinschaft.

Dem Ordinarius des Lehrstuhls, Herrn Professor Dr.-Ing., Dr.-Ing. h.c. D. Schröder, danke ich sehr herzlich für die Anregung und Betreuung meiner Forschungstätigkeit, was neben dem hervorragend ausgestatteten Arbeitsumfeld am Lehrstuhl, erheblich zum guten Gelingen der Arbeit beitrug.

Für die Übernahme des Koreferats und für das entgegengebrachte Interesse an dieser Arbeit danke ich Herrn Professor Dr.-Ing. R. Marquardt (Universität der Bundeswehr München) und für die Übernahme des Prüfungsvorsitzes gebührt mein Dank Herrn Professor Dr. rer. nat. G. Wachutka.

Dankend erwähnen möchte ich besonders auch Herr Dr. Horst Grüning (damals ABB Industrie, Turgi), der mich einerseits mit Meßergebnissen zum hartgeschalteten IGCT und andererseits mit wertvollen Hinweisen zum besseren Verständnis dieses Bauelements, maßgeblich unterstützte. Weitere Meßergebnisse zum IGCT bekam ich von Herrn Dr. Jürg Waldmeyer (ABB Semiconductors, Lenzburg), bei dem ich mich an dieser Stelle ebenfalls herzlich bedanke.

Erwähnenswert ist außerdem das hervorragende Arbeitsklima am Lehrstuhl, zu dem ausnahmslos alle Mitarbeiter und Mitarbeiterinnen sowohl fachlich als auch menschlich beitrugen. Vor allem im ersten Jahr meiner Tätigkeit verhalf mir Herr Dr. rer. nat. A. Schlögl zu einer schnellen und effektiven Einarbeitung in den Themenkomplex physikalische Modellbildung. Ihm und allen anderen ehemaligen und aktuellen Lehrstuhlkollegen möchte ich ein herzliches Dankeschön für die gute Zusammenarbeit ausdrücken

Schließlich möchte ich mich bei meinen lieben Eltern bedanken, die mich bei meinem Ausbildungsweg jederzeit geduldig und motivierend unterstützten.

Inhaltsverzeichnis

0	Einführung	1
1	Physikalische Grundlagen	5
1.1	Zum Begriff Halbleiter	5
1.2	Ladungsträgerstatistik bei Halbleitern	6
1.2.1	Eigenleitungsträgerdichte	6
1.2.2	Trägerdichten bei Dotierung	10
1.2.3	Reduktion der Bandlücke bei hohen Dotierungen	13
1.3	Die Halbleiter-Grundgleichungen	14
1.3.1	Die Maxwellgleichungen	14
1.3.2	Die Transportgleichungen	15
1.3.3	Die Kontinuitätsgleichungen	15
1.4	Rekombinationsvorgänge	16
1.4.1	Der Begriff Lebensdauer	16
1.4.2	Rekombination mittels Rekombinationszentren	18
1.4.3	Auger-Rekombination	20
1.4.4	Bauelementemodelle und Lebensdauern	23
1.4.5	Stoßionisation und Lawineneffekt	24

1.5	Streuprozesse freibeweglicher Ladungsträgern	27
1.5.1	Arten von Streuprozessen	27
1.5.2	Driftgeschwindigkeiten und Beweglichkeiten	28
1.5.3	Streuung am Gitter und an Verunreinigungen	31
1.5.4	Elektronen-Löcher-Streuung I: Herkömmliche Theorie	34
1.5.5	Elektronen-Löcher-Streuung II: Theorie nach Mnatsakanov	35
1.6	Abrupte Halbleiterübergänge	39
1.6.1	Der pn -Übergang	39
1.6.2	Der $p^+\nu$ -Übergang	43
1.6.3	Der $n^+\nu$ -Übergang	43
2	Der Aufbau physikalischer Modelle	45
2.1	Modularer Modellaufbau am Beispiel der Leistungsdiode	45
2.2	Modellierung der Randzonen	48
2.2.1	Injektionsverhalten und Stromverläufe	48
2.2.2	Die transparente Anode	51
2.2.3	Temperaturverhalten	53
2.3	Modellierung der Driftzone	56
2.3.1	Beschreibung des Ladungsträgerverhaltens	56
2.3.2	Realisierung im Modell	62
2.4	Modellierung weiterer Eigenschaften	73
2.4.1	Modellierung exponentieller Dotierprofile	73
2.4.2	Modellierung von Lebensdauerprofilen	76
2.4.3	Der dynamische Lawineneffekt	78
2.5	Das Gesamtmodell	81

3	Das IGCT–Modell	85
3.1	Thyristor, GTO, IGCT	85
3.2	Das GCT–Modell: Modularer Aufbau	92
3.3	Modellierung der Pufferschicht	94
3.3.1	Allgemeines	94
3.3.2	Quasistatische Modellierung	95
3.3.3	Pufferzone als Teil der Driftzonenbeschreibung	97
3.3.4	Pufferschicht und Raumladungszone	99
3.4	Modellierung der p–Basis	101
3.5	Das Gesamtmodell	105
4	Validierung des IGCT–Modells	109
4.1	Modellanpassung	109
4.2	Validierung anhand von Messungen	113
4.2.1	Meßschaltung und gemessene GCT–Typen	113
4.2.2	Abschaltverhalten	114
4.2.3	Einschaltverhalten und Kennlinie	120
4.2.4	IGCT–Modell versus Modell aus der SABER–Bibliothek . . .	123
5	Analyse der Serienschaltung von IGCTs	127
5.1	Motivation für die Serienschaltung	127
5.2	Asymmetrisches Abschaltverhalten	128
5.2.1	Ursachen und Mechanismus	128
5.2.2	Vermeidung der Spannungsasymmetrie	137

6	Konzept der aktiven Ansteuerung von IGCTs	141
6.1	Das Regelungskonzept	141
6.1.1	Allgemeines	141
6.1.2	Von der Idee zum P-Regler	141
6.1.3	Simulationsergebnisse	144
6.1.4	Allgemeine Anforderungen	147
6.2	Schaltungstechnische Umsetzung	148
6.2.1	Schaltungsblöcke	148
6.2.2	Spannungserfassung und Differenzbildung	149
6.2.3	Der Leistungstreiber	152
6.2.4	Analoger adaptiver P-Regler	157
6.3	Simulationsergebnisse	159
6.4	Grenzen und mögliche Problemfelder	163
6.4.1	Aktive Ansteuerung und Erwärmungsverhaltens beim IGCT .	163
6.4.2	Aktive Ansteuerung und Redundanz	169
6.4.3	Aktive Ansteuerung und Pufferschicht	170
6.5	Diskussion und Bewertung	173
7	Zusammenfassung	175
7.1	Inhalt und Ergebnisse der Arbeit	175
7.2	Ausblick	177
A	Anhang	186
	Symbolverzeichnis	190

0 Einführung

Die vorliegende Arbeit wurde am Lehrstuhl für Elektrische Antriebssysteme¹ der Technischen Universität München durchgeführt. Zu einem antriebstechnischen System gehört neben anderen Komponenten, wie z.B. der Regelung oder einem Elektromotor, ein Stellglied, d.h. eine leistungselektronische Schaltung, die für den gewünschten Energiefluß zwischen den einzelnen Teilsystemen (z.B. Netz und Elektromotor) sorgt. Leistungselektronische Schaltungen finden aber gerade im Bereich hoher Leistung vor allem Verwendung bei der Energieversorgung, z.B. um Netze mit unterschiedlicher Netzspannung und Netzfrequenz zu koppeln (HVDC, FACTS, Frequenzumrichter).

Für solche Schaltungen existieren am Markt unterschiedliche Halbleiterbauelemente, die dem Schaltungsentwickler, abhängig vom Anwendungsfall, bzw. der gewünschten Leistungsklasse, zur Verfügung stehen. Für kleinere Leistungen, wie beispielsweise in der Automobilelektronik, d.h. im Bereich weniger kVA und gleichzeitig für Spannungen unter 100V, verwendet man in der Regel Leistungs-MOSFETs, die Schaltfrequenzen oberhalb von 10kHz zulassen. Für größere Leistungen und Spannungen ist der IGBT (Insulated Gate Bipolar Transistor) das Bauelement der Wahl. Selbst für Spannungen größer als 1kV schien der IGBT den GTO (Gate Turn Off Thyristor) fast vollständig zu verdrängen. Jedoch kann der etwa Mitte der neunziger Jahre entwickelte IGCT (Integrated Gate Commutated Thyristor), als eine neue, hartschaltende GTO-Variante, dem IGBT im Hochvoltbereich Paroli bieten [66]. Gerade im Leistungsbereich von 1MVA und größer, sind die Schaltverluste nicht größer als beim IGBT, bei gleichzeitig wesentlich geringeren Durchlaßspannungen [6], [7], [33].

Hinsichtlich der Schaltungsentwicklung und -auslegung wäre es nun von großem Nutzen, wenn man das Werkzeug Netzwerksimulation einsetzen könnte, um Antworten auf Fragestellungen hinsichtlich der maximalen Strom- und Spannungsbelastbarkeit, der maximalen Schaltfrequenz, der Temperaturbelastung, u.v.a.m., zu erhalten. Bekanntlich sparen aussagekräftige Simulationsläufe einen, oder mehrere Testschritte ein, was zu einer erheblichen Kostenreduktion in der Entwicklung führen kann, vor allem vor dem Hintergrund, daß der Preis für nur einen IGCT bereits im dreistelligen DM-Bereich liegt.

¹s.a. <http://www.eat.ei.tum.de>

Für diese Art von Simulationen benötigt man allerdings geeignete Modelle für einen der gängigen Netzwerksimulatoren, wie z.B. PSPICE oder SABER [2]. Man könnte nun die Auffassung vertreten, solche Modelle sind bereits in den Standardbibliotheken dieser Simulatoren reichlich vorhanden, wie das für die meisten Signalbauelemente tatsächlich der Fall ist. Diese Aussage stimmt aber nur zum Teil, sofern es sich nämlich um unipolare Leistungsbaulemente, wie z.B. Schottky-Diode und Leistungs-MOSFET handelt. Bei den bipolaren Leistungsbaulementen, also Leistungsdioden, IGBTs und Thyristoren, gibt es dagegen lediglich Modelle mit wenig, bzw. sehr eingeschränkter Aussagekraft bezüglich den oben genannten Fragestellungen.

Damit stellt sich die Frage, woran es liegen mag, daß geeignete Modelle für bipolare Leistungshalbleiter nur spärlich erhältlich sind. Leistungshalbleiter, die hohe Spannungen sperren bzw. blockieren können, brauchen dazu eine sehr lange, schwach n -dotierte Zone (Driftzone, bzw. ν -Zone). Bei bipolaren Bauelementen wird diese in Vorwärtspolung beidseitig (Diode, GTO, Thyristor), bzw. einseitig (IGBT) mit Überschlußladungsträgern überschwemmt. Die entsprechende Konzentration freibeweglicher Überschlußladungsträger bewegt sich durchschnittlich im Bereich zwischen 10^{15}cm^{-3} und 10^{17}cm^{-3} . Die Donatorkonzentration N_ν in dieser Zone beträgt dagegen in der Regel nur zwischen 10^{13}cm^{-3} und 10^{14}cm^{-3} , so daß $p \approx n \gg N_\nu$ gilt. Aus dieser Beziehung folgt, daß es sich um die sogenannte starke Injektion, bzw. Hochinjektion, handelt. Im Gegensatz hierzu, hat man es in den stark dotierten Randzonen, d.h. Anode und Kathode, und bei den meisten Signalbauelementen ausschließlich mit schwacher Injektion zu tun. Das Auftreten starker Injektion hat zur Konsequenz, daß die Halbleitergleichungen zur Beschreibung der Vorgänge in der Driftzone nicht mehr für schwache Injektion angenähert und somit nicht mehr stark vereinfacht werden können. Hinzu kommt, daß die Driftzonenlänge mit mehreren $100\mu\text{m}$ im Falle von Hochvoltbauelementen, die Längen aller anderen Zonen mit meist nur einigen wenigen μm bei weitem übersteigt. Somit muß beim Einschalten eine erhebliche Menge Ladungsträger in die Driftzone eingebracht werden. Umgekehrt kann beim Abschalten die Driftzone erst die volle Sperrspannung aufnehmen, wenn diese Überschlußladungen wieder durch Absaugen und/oder durch Rekombination verschwunden sind. Dieses Ein- und Ausbringen von Ladung wirkt sich auf das dynamische Verhalten, d.h. das Schaltverhalten und die Schaltverluste, erheblich aus. Mit anderen Worten: die interessierenden Strom- und Spannungstransienten werden maßgeblich von den Ladungsträgerverhältnissen in der ν -Zone bestimmt. Das zeitliche und räumliche Verhalten der Überschlußladungen wird von einer partiellen Differentialgleichung beschrieben, die numerisch gelöst werden muß.

Was bedeutet dies für die Modellierung? Es leuchtet ein, daß es wenig Sinn macht, ein Bauelement aufwendig zwei- oder dreidimensional numerisch zu beschreiben. Abgesehen davon, daß dies mit einem hohen Rechenaufwand nebst potentiell auftretenden Konvergenzproblemen verbunden wäre, ist dies bereits die Domäne der

sogenannten Bauelementesimulation (z.B. DESSIS [30], MEDICI). Andererseits liefern Modelle von Leistungshalbleiterbauelementen, die auf Ersatzschaltbildern aus passiven Elementen und gesteuerten Quellen basieren, meist sehr ungenaue Resultate für das tatsächliche Schaltverhalten (siehe z.B. [3]). Dies aber immerhin in sehr kurzer Rechenzeit. Daneben existieren noch andere Konzepte, von denen im folgenden zwei exemplarisch kurz erwähnt werden sollen. Zum einen wird versucht, das dynamische Verhalten der Überschußladungen mit einem Netzwerk aus RC-Kettenbrüchen anzunähern [67]. Zum anderen gibt es den sogenannten Lumped Charge-Ansatz: Hierbei wird die Driftzone in einige wenige Unterzonen aufgeteilt, deren Ladungen als eine einzige kompakte Punktladung aufgefaßt werden [41], [42]. Wie in Kapitel 4 für den Fall des GTO-Modells (s. [41]) gezeigt wird, liefert dieser Ansatz prinzipiell akzeptable Simulationsergebnisse, aber bestimmte Effekte wie z.B. die auftretende Spannungsasymmetrie bei seriengeschalteten unentlasteten IGCTs, können mit so einem Modell nicht berechnet werden (s. dazu Kap. 5). Schließlich kann man versuchen, einen Mittelweg zwischen einer aufwendigen mehrdimensionalen numerischen Lösung und den anderen vorgestellten Konzepten zu finden. Aufbauend auf die Arbeiten [45], [73] wird in der vorliegenden Arbeit das Konzept der physikalischen Modellierung verfolgt. Dabei wird lediglich die räumliche und transiente Ladungsträgerverteilung in der Driftzone numerisch berechnet, wobei eindimensionale Verhältnisse angenommen werden. Alle anderen relevanten physikalischen Gleichungen und Größen werden separat für jede einzelne Zone implementiert. Um eine direkte Implementierung von physikalischen Gleichungen zu ermöglichen, wird auf den Schaltungssimulator SABER zurückgegriffen, dessen analoge Beschreibungssprache MAST im Gegensatz zu PSPICE, bessere Möglichkeiten zu einer physikalischen Modellierung bietet. Die physikalischen Modelle, die wegen ihrer Mischung aus kompakter analytischer und numerischer Beschreibung auch Hybridmodelle genannt werden (s. [25]), liefern sehr gute Simulationsergebnisse in vertretbarer Rechenzeit. Weitergehende Übersichten über verschiedene Modellierungskonzepte finden sich u.a. in [70] für Leistungsdioden und in [63] für den IGBT. Physikalische Modelle für Leistungs-MOSFETs, Diode, IGBT und dem konventionellen GTO sind Gegenstand vorhergehender Arbeiten (vgl. u.a. [45], [73], [74]). In der vorliegenden Arbeit wird ein validiertes physikalisches Modell für den IGCT vorgestellt. Des weiteren wird gezeigt, wie das Modell zu Analyse- und Synthesezwecken eingesetzt werden kann. Zunächst aber werden im Kapitel 1 die halbleiterphysikalischen Gleichungen, soweit sie für die Modellierung wichtig sind, vorgestellt. Aspekte der Implementierung in MAST, d.h. Fragen zur Modellstruktur, Modularität, verwendeter Numerik, usw., werden in Kapitel 2 erläutert. Kapitel 3 widmet sich speziell dem IGCT-Modell. Vergleiche von simulierten und gemessenen Schaltverläufen und Kennlinien, einschließlich einer abschließenden Beurteilung, findet sich in Kapitel 4. In Kapitel 5 wird gezeigt, daß das Verhalten von seriengeschalteten, unentlasteten IGCTs, d.h. die auftretenden Spannungsasymmetrien beim Abschalten, mit

dem verwendeten Modell richtig berechnet werden kann. Als Konsequenz daraus wurde überlegt, ob es nicht möglich wäre, eine Schaltung zur intelligenten (aktiven) Gateansteuerung, die eine Spannungssymmetrierung ohne RC(D)–Beschaltung ermöglichen würde, zu entwickeln. Der daraus entstandene Ansatz wird in Kapitel 6 behandelt. Diese Arbeit schließt mit einer Zusammenfassung und einem Ausblick in Kapitel 7.

1 Physikalische Grundlagen

In diesem Kapitel soll nicht etwa die gesamte Halbleitertheorie umfassend dargestellt werden, sondern vielmehr ein Überblick gegeben werden, welche Teile davon bis zu welcher Tiefe für das Erstellen physikalischer Modelle bedeutsam sind. Wegen der großen Bedeutung im Bereich der Leistungshalbleiterbauelemente, wird besonders auch auf Temperaturabhängigkeiten physikalischer Größen und Gleichungen eingegangen.

1.1 Zum Begriff Halbleiter

Einer vereinfachenden Sichtweise zufolge, lassen sich Metalle, Halbleiter bzw. Isolatoren durch ihren spezifischen Widerstand unterscheiden. Dabei ist die Grenze zwischen Halbleitern und Isolatoren mit ca. $10^{-4}\Omega\text{cm}$ recht willkürlich gewählt [49], [61]. Zu einer besseren Möglichkeit, Halbleiter von Isolatoren abzugrenzen, kommt man, wenn man sich mit den Aussagen der Festkörperphysik, insbesondere bzgl. Kristallgitter in Verbindung mit dem daraus abgeleiteten Bändermodell, auseinandersetzt (s. z.B. [31]). Eine detaillierte Betrachtung des Bändermodells würde den Rahmen dieser Arbeit sprengen, weshalb im folgenden die wichtigsten Aussagen nur qualitativ zusammengefaßt werden. Unterstellt man nun, daß es in einem Kristallgitter eine bestimmte Anzahl freibeweglicher Elektronen gibt, die zu einem Stromfluß beitragen können, befinden sich diese in Wechselwirkung mit dem elektrostatischen Kräftefeld der sich wiederholenden Anordnung von Atomrümpfen im Gitter, oder mit anderen Worten: die potentielle Energie dieser Leitungselektronen ist periodisch bezüglich des Ortes. Während nun ein freies Elektron nur diskrete Energiewerte annehmen kann, ergibt die Lösung der Schrödinger-Gleichung für ein frei bewegliche Elektron in einem periodischen Potential, mehrere, von Lücken unterbrochene Energiebänder. Die Elektronen können nur die Energiezustände besetzen, die innerhalb eines Bandes liegen. Die Energielücken hingegen sind verbotene Zustände. Von besonderer Bedeutung sind dabei die obersten beiden Bänder, das sogenannte Leitungs- und Valenzband, die ebenfalls durch eine verbotene Energielücke getrennt sind. Beim Valenzband sind für $T=0\text{K}$ immer alle Zustände besetzt. Beim Metall ist für $T=0\text{K}$ außerdem das Leitungsband großenteils mit Elektronen besetzt. Diese

Elektronen sind innerhalb des Metallgitters frei beweglich und bilden das sogenannte Elektronengas. Sowohl bei Isolatoren als auch bei Halbleitern sind alle Zustände im Leitungsband für $T=0\text{K}$ unbesetzt, d.h. auch Halbleiter sind für $T=0\text{K}$ ideale Isolatoren¹. Im Falle von Halbleitern können sich für $T>0\text{K}$ mit einer gewissen Wahrscheinlichkeit, die sich nach der sogenannten Fermi-Verteilung richtet, Elektronen aus dem Valenzband die Energielücke überwinden und in das Leitungsband gelangen. Sowohl diese Elektronen als auch die im Valenzband zurückbleibenden Löcher sind nun frei beweglich und können zu einem Stromfluß beitragen. Diese Erscheinung nennt man Eigenleitung (s. Abschnitt 1.2.1) und tritt prinzipiell auch bei Isolatoren auf, allerdings wegen der meist viel größeren Bandlücke, in nur vernachlässigbar geringem Maße. Außerdem läßt sich bei Halbleitern durch gezieltes Verunreinigen mit Fremdatomen (Dotierung, s. Abschnitt 1.2.2) die Leitfähigkeit steigern und damit quasi metallähnliches Verhalten realisieren.

1.2 Ladungsträgerstatistik bei Halbleitern

1.2.1 Eigenleitungsträgerdichte

In Abschnitt 1.1 wurde festgestellt, daß sich bei Halbleitern, also auch und besonders für Silizium, für $T>0\text{K}$, Elektronen im Leitungsband befinden. Damit stellt sich nun die Frage nach ihrer Anzahl. Da es unmöglich ist, den Zustand jedes einzelnen Elektrons deterministisch vorherzusagen, bedarf es statistischer Aussagen. Die Verteilung klassischer Teilchen (z.B. in der Gastheorie) kann mit der sogenannten Boltzmann-Verteilung beschrieben werden. Für Teilchen, die dem Pauli-Prinzip gehorchen, d.h. jeder Energiezustand kann nur mit maximal einem Teilchen besetzt werden, wie das für Löcher und Elektronen der Fall ist, gilt die sogenannte Fermi-Dirac-Verteilung, die durch den Ausdruck

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)} \quad (1.1)$$

gegeben ist. Das Fermieniveau E_F entspricht dem Energieniveau, für das die Wahrscheinlichkeit, daß es mit einem Elektron besetzt wird, genau 0,5, d.h. 50% ist. Für undotierte, d.h. eigenleitende Halbleitermaterialien, liegt das Fermi-niveau ungefähr in der Mitte der Bandlücke zwischen Leitungs- und Valenzbandkante. Die Bandlücke bei Silizium beträgt etwa 1,1eV, wohingegen $1 k_B T$ bei $T=300\text{K}$ nur etwa 0,026eV ergibt. Für jedes Elektron, das einen Zustand E innerhalb des Leitungsbandes besetzt, gilt somit $E - E_F \gg k_B T$ (Abstand Leitungsbandkante zum Fermi-niveau ca.

¹Außer im Falle der Entartung. Zu diesem Begriff siehe Fußnote 4 auf S. 12.

0,55eV). Damit vereinfacht sich (1.1) für Elektronen zu

$$f_n(E) \approx \exp\left(\frac{-E - E_F}{k_B T}\right) \quad (1.2)$$

Dies entspricht der oben erwähnten Boltzmann-Verteilung. Für die Löcher erhält man nach analogen Überlegungen

$$f_p(E) = 1 - f_n(E) \approx \exp\left(\frac{-E_F - E}{k_B T}\right). \quad (1.3)$$

Die Zustandsdichte ist ein weiterer wichtiger Begriff zur Beschreibung der Ladungsträgerstatistik und leitet sich von der Theorie des freien Elektronengases ab (vgl. z.B. [31]). Sie gibt die Anzahl diskreter Energieniveaus im Intervall $[E, E+dE]$ pro Volumeneinheit an. In einem periodischen Potential, d.h. in einem Kristall, verhält sich ein freibewegliches Teilchen (Elektron oder Loch) mit der Masse m , wie ein Teilchen mit einer von m verschiedenen Masse m^* . Dabei ist m^* seine effektive Masse, die indirekt proportional zur Bandkrümmung im \vec{k} -Raum, d.h. der zweiten Ableitung der Energie $E(\vec{k})$ nach \vec{k} ist. Für Silizium kann zum einen Isotropie, d.h. $k_x = k_y = k_z$ und somit $k = |\vec{k}|$, und zum anderen im Bereich des Minimums der Leitungsbandkante E_C , bzw. im Bereich des Maximums der Valenzbandkante E_V (die Differenz ergibt die minimale Bandlücke E_G in Höhe der 1,1eV), näherungsweise ein quadratischer Energieverlauf $E(k)$ angenommen werden, so daß sich für die beweglichen Ladungsträger nahe den Bandkanten konstante effektive Massen ergeben. Unter der Annahme von Kugelsymmetrie, d.h. die Flächen gleicher Energie befinden sich auf den Kugelschalen, ergibt sich für die Zustandsdichte N

$$N(E) = \frac{4\pi (2m^*)^{3/2}}{h^3} \cdot \sqrt{E}. \quad (1.4)$$

Die Elektronendichte n_0 im Leitungsband im thermischen Gleichgewicht, errechnet sich mit (1.2) zu

$$n_0 = \int_{E_C}^{\infty} N(E) f_n(E) dE. \quad (1.5)$$

Setzt man nun (1.4) in (1.5) ein und führt die Integration durch, so erhält man für die Elektronendichte n_0

$$\begin{aligned} n_0 &= N_C \cdot \exp\left(-\frac{E_C - E_F}{k_B T}\right) \\ N_C &= 2 \left(\frac{2\pi m_n^* k_B T}{h^2}\right)^{3/2}. \end{aligned} \quad (1.6)$$

Analog erhält man mit (1.3), (1.4) und der Integration von $-\infty$ nach E_V , also über alle Zustände des Valenzbandes, die intrinsische Löcherdichte p_0 im thermischen Gleichgewicht

$$\begin{aligned} p_0 &= N_V \cdot \exp\left(-\frac{E_F - E_V}{k_B T}\right) \\ N_V &= 2 \left(\frac{2\pi m_p^* k_B T}{h^2}\right)^{3/2}. \end{aligned} \quad (1.7)$$

Hierbei heißen N_C und N_V äquivalente Zustandsdichten des Leitungs- bzw. des Valenzbandes.

Im thermische Gleichgewicht sind die Generationsrate $G(T)$ und die Rekombinationsrate $R(T)$ gleich. Es gilt also $R(T) = G(T)$ mit $R(T) = r(T) n_0 p_0$. Damit ergibt sich

$$n_0 p_0 = \frac{G(T)}{r(T)} = \text{const}(T). \quad (1.8)$$

Diese Beziehung sagt aus, daß sich die gegenläufigen Prozesse Generation und Rekombination im thermischen Gleichgewicht im Mittel soweit ausgleichen, daß das Produkt $n_0 p_0$ im Sinne eines Massenwirkungsgesetzes immer konstant ist und nur von der Temperatur T abhängt. Wird nun eine Ladungsträgersorte, z.B. durch Dotierung erhöht, sinkt die andere entsprechend ab. Im Fall der Eigenleitung hinterlassen die n_0 Elektronen, die vom Valenzband ins Leitungsband gelangt sind, exakt die gleiche Anzahl Löcher im Valenzband. Es gilt also

$$n_i = n_0 = p_0, \quad (1.9)$$

wobei n_i die sogenannte Eigenleitungsträgerdichte ist. Mit (1.9) folgt aus (1.8)

$$n_i^2 = n_0 p_0 \quad (1.10)$$

Setzt man nun (1.6) und (1.7) in (1.10) ein und zieht die Wurzel, bekommt man mit Hilfe des Ausdrucks für die Bandlücke $E_G = E_C - E_V$, folgende Formel zur Bestimmung der Eigenleitungsträgerdichte n_i

$$n_i = \sqrt{N_C N_V} \cdot \exp\left(\frac{-E_G}{2k_B T}\right). \quad (1.11)$$

Unter Berücksichtigung der Tatsache, daß die Bandlücke E_G , die kritischen Massen $m_{n,p}^*$ und die äquivalenten Zustandsdichten $N_{C,V}$, von der Temperatur abhängen (s.u.), zeigt (1.11), daß n_i^2 tatsächlich eine Konstante ist, die von der Temperatur abhängt, wie im Zusammenhang mit (1.8) behauptet wurde und deshalb eine thermisches Gleichgewichtskonzentration beschreibt. Wie in Abschnitt 1.6 und Kapitel 2 noch gezeigt wird, taucht n_i bzw. n_i^2 in diversen Formeln zur Bestimmung von

Ladungsträgerkonzentrationen immer wieder auf. Darin liegt der eigentliche Wert der Kenntnis von n_i für die Modellierung, denn die Eigenleitung selbst, hat keinerlei technische Bedeutung.² Die Gleichungen (1.6) und (1.7) können aufgrund (1.9) gleichgesetzt werden. Damit kann das Fermi-niveau E_{Fi} eines undotierten Halbleiters bestimmt werden zu

$$E_{Fi} = \frac{E_C + E_V}{2} + \frac{k_B T}{2} \ln \left(\frac{N_V}{N_C} \right). \quad (1.12)$$

Das Fermi-niveau liegt also wegen $N_C \neq N_V$ nicht exakt in der Mitte der Bandlücke, sondern um den Wert des zweiten (logarithmischen) Terms verschoben.

Das bisher Geschilderte basiert auf analytischen Herleitungen aus physikalischen Gleichungen und Gesetzen. Für bestimmte physikalische Größen, insbesondere wenn es um Temperaturabhängigkeiten geht, muß man sich bei der Modellbildung häufig empirischer Ausdrücke behelfen. Zur Bestimmung der Eigenleitung betrifft dies die effektiven Massen der freibeweglichen Ladungsträger inklusive ihrer Temperaturabhängigkeit einerseits und die Abhängigkeit der Bandlücke von der Temperatur und der Dotierung (sog. Bandgap Narrowing) andererseits, wobei auf das Bandgap Narrowing erst in Abschnitt 1.2.3 eingegangen wird.

Betrachtet man die Ausdrücke (1.6) und (1.7), erkennt man, daß bei bekannter Temperatur lediglich die effektiven Massen $m_{n,p}^*$ von Elektronen im Leitungsband und Löchern im Valenzband, zur Ermittlung von $N_{C,V}$ fehlen. In den Modellen findet sich eine Beschreibung der isotropen effektiven Elektronenmasse m_n^* und ihrer Temperaturabhängigkeit gemäß [26], wie sie beispielsweise auch im Bauelementesimulator DESSIS implementiert ist (s. [30]). Dabei wird m_n^* in einen temperaturabhängigen transversalen Teil ($m_t = 0,1905 m_0 E_G(0)/E(T)$) und einen temperaturunabhängigen longitudinalen Teil ($m_l = 0,9163 m_0$) zerlegt. Diese beiden Massen werden mit der Beziehung $m_n^* = \sqrt[3]{m_t^2 m_l}$ zur effektiven Elektronenmasse m_n^* verknüpft. Insgesamt ergibt sich für m_n^*

$$m_n^* = 1,0618 m_0 \sqrt[3]{\frac{E_G(0)}{E_G(T)}}. \quad (1.13)$$

Hierbei sind m_0 die Ruhemasse eines Elektrons und $E_G(T)$ die schwach temperaturabhängige Bandlücke gemäß (1.15) (s. unten). Die effektive Löcher-masse m_p^* ist durch den empirischen Ausdruck nach [40] gegeben:

$$m_p^* = m_0 \left(\frac{a + bT + cT^2 + dT^3 + eT^4}{1 + fT + gT^2 + hT^3 + iT^4} \right)^{2/3}, \quad (1.14)$$

Die genauen Werte der Koeffizienten a bis i finden sich u.a. in [30], [58].

²Abgesehen davon übertrifft die Konzentration freier Ladungsträger, die von Verunreinigungen selbst technisch reinsten Siliziums herrühren, die Eigenleitungskonzentration n_i bei weitem.

Die experimentellen Arbeiten [9] und [43], hatten zum Ergebnis, daß sich die Bandlücke in Abhängigkeit von der Temperatur sehr gut durch ein Polynom 2. Ordnung annähern lassen. In [30] wird eine solche Näherung für den Bereich 100K bis 415K angegeben. Sie lautet für E_G in eV:

$$E_G = 1.1689 + 2.35 \cdot 10^{-5} T - 7.9358 \cdot 10^{-7} T^2 + 6.9383 \cdot 10^{-10} T^3 \quad (1.15)$$

Damit sind nun alle Größen, die man zur Berechnung von $N_{C,V}$ und damit von n_i gemäß (1.11) benötigt, bekannt.

1.2.2 Trägerdichten bei Dotierung

Unter Dotieren versteht man das bewußte Verunreinigen eines vorher reinen Halbleitermaterials mit Fremdatomen, um dadurch die Leitfähigkeit gezielt einstellen zu können. Dabei unterscheidet man Donatoren und Akzeptoren. Der hier interessierende Halbleiter Silizium ist vierwertig, d.h. jedes Si-Atom im Kristall ist an vier weiteren Si-Atomen gebunden. Wird nun ein fünfwertiges Atom wie z.B. Phosphor oder Arsen in den Gitterverband eingebaut, wird es ebenfalls an vier Si-Atome gebunden. Damit bleibt aber ein nicht gebundenes Valenzelektron in der äußeren Schale des Fremdatoms übrig, das sich sehr leicht vom Dotieratom lösen läßt (Ionisation) und zur elektrischen Leitfähigkeit beitragen kann. Wegen dem fehlenden Elektron ist das ionisierte Fremdatom einfach positiv geladen. Somit gleicht das entstehende Elektronengas die Ladungen der ionisierten Atome gerade wieder aus, so daß die Neutralität (makroskopisch) gewahrt bleibt. Dotierstoffe, die in dieser Weise Elektronen abgeben, heißen Donatoren und das Halbleitermaterial wird n-leitend. Nun ist es auch möglich dreiwertige Atome wie z.B. Aluminium oder Bor in den Si-Kristall einzubauen. Um sich an vier Si-Atomen binden zu können, fehlt dem Dotieratom ein Valenzelektron. Dieses fehlende Elektron wird als Defektelektron oder Loch bezeichnet und kann nach dem Bändermodell als einfach positiv geladenes Teilchen mit einer effektiven Masse m_p^* (vgl. Abschnitt 1.2.1) interpretiert werden, daß sich ebenso wie Elektronen im Donatorenfall vom Dotieratom lösen läßt und somit zu einer Leitfähigkeit beitragen kann. Wegen dem fehlenden Loch ist das ionisierte Fremdatom einfach negativ geladen. Damit ist der Halbleiter weiterhin elektrisch neutral. Löcherspendende (= elektronenaufnehmende) Atome heißen Akzeptoren und das Halbleitermaterial wird entsprechend p-leitend.

Betrachtet sei im folgenden ein n-dotierter (p-dotierter) Halbleiterkristall mit der Donatorenkonzentration N_D (Akzeptorenniveau E_A).³ Das Energieniveau E_D (E_A)

³Die Konzentration von Si in einem Kristall beträgt ca. $5 \cdot 10^{22} \text{ cm}^{-3}$, d.h. selbst für sehr hohe Dotierstoffkonzentrationen von z.B. 10^{19} cm^{-3} ist durchschnittlich nur jeder 5000ste Gitterplatz mit einem Fremdatom besetzt.

des Donators (Akzeptors) liegt dabei oberhalb (unterhalb) der Bandmitte, meist nahe der Leitungsbandkante (Valenzbandkante). Der Abstand von E_D (E_A) zum Leitungsbandniveau E_C (Valenzbandniveau E_V) ist relativ klein (einige 10mV), so daß man für Raumtemperatur annimmt, daß die Donatoren (Akzeptoren) bereits vollständig ionisiert sind. Dann gilt $N_D^+ \approx N_D$ [$N_A^- \approx N_A$] und es gibt eine zur Eigenleitung zusätzliche Konzentration freibeweglicher Elektronen (Löcher), die genau der Dotierstoffkonzentration entspricht. Allerdings ist die Annahme der vollständigen Ionisation selbst bei Raumtemperatur für $N_{D,A} > 10^{16} \text{cm}^{-3}$ nicht mehr haltbar (vgl. u.a. [58]).

Wieviele der der Donatoratome (Akzeptoratome) nun tatsächlich ionisiert werden, hängt von der Temperatur T und insbesondere von der dotierstoffspezifischen effektiven Ionisationsenergie $E_{A,D}^*$ ab, die ihrerseits von der Temperatur und der Dotierstoffkonzentration $N_{A,D}$ abhängt. Einen zu Modellierungszwecken brauchbaren empirisch-analytischen Zusammenhang liefert [54]:

$$E_{A,D}^* = E^0 + a \cdot \left[\frac{T}{1 + \exp\left(\frac{T_1 - T}{28.8K}\right)} - \frac{T_0}{1 + \exp\left(\frac{T_2 - T}{28.8K}\right)} \right] \ln\left(\frac{N_0}{N_{A,D}}\right) \quad (1.16)$$

Tabelle 1.1 faßt die Parameter $E_{A,D}^0$, $T_{0,1,2}$, a und N^0 für wichtige Dotierstoffe zusammen.

Tabelle 1.1: Koeffizienten zu (1.16) für Phosphor P, Arsen As, Antimon Sb und Bor B.

Dotier- element	E^0 /meV	a / 10^{-4}meV	N^0 / 10^{18}cm^{-3}	T_0 /K	T_1 /K	T_2 /K
P	45.5	272	115	82.0	98	125
As	53.7	305	26.2	108	128	153
Sb	42.7	275	82.3	76.3	92.5	125
B	44.4	240	1.02	136	173	199

Mit Hilfe der Ionsisationsenergie für Donatoren E_D^* (Akzeptoren E_A^*) kann das Energieniveau der Donatoren (Akzeptoren) mittels $E_D = E_C - E_D^*$ [$E_A = E_V + E_A^*$] berechnet werden. Die Dichte besetzter, d.h. nicht ionisierter Donatoren (Akzeptoren), kann wiederum mit der Fermi-Dirac-Statistik zu $N_D^0 = N_D f(E_D)$ [$N_A^0 = N_A (1 - f(E_A))$] berechnet werden. Da $N_D^+ = N_D - N_D^0$ [$N_A^- = N_A - N_A^0$] gilt, bestimmt sich die Dichte der ionisierten Donatoren (Akzeptoren) demnach zu $N_D^+ = N_D (1 - f(E_D))$ [$N_A^- = N_A f(E_A)$]. Um aber die Ladungsträgerstatistik für solche

Fälle beschreiben zu können, muß die Gleichung (1.1) um den sogenannten Degenerationsfaktor g erweitert werden:

$$f(E) = \frac{1}{1 + \frac{1}{g} \exp\left(\frac{E - E_F}{k_B T}\right)} \quad (1.17)$$

Bei g handelt es sich um einen Gewichtungsfaktor der dem Quotient aus Anzahl quantenmechanischer Zustände geteilt durch Anzahl aufnehmbarer freier Ladungsträger beim betrachteten Energieniveau entspricht. In Abschnitt 1.2.1 galt demnach $g = 1$. Setzt man nun (1.17) in die Bestimmungsgleichungen für N_D^+ und N_A^- ein, so bekommt man

$$N_D^+ = \frac{N_D}{1 + g_D \exp\left(\frac{E_F - E_D}{k_B T}\right)} \quad (1.18)$$

und

$$N_A^- = \frac{N_A}{1 + g_A \exp\left(\frac{E_A - E_F}{k_B T}\right)}, \quad (1.19)$$

wobei $g_D = 2$ und $g_A = 4$ gilt. Da das Akzeptorniveau beim Valenzbandmaximum im \vec{k} -Raum (also für $k = 0$) doppelt aufgespalten⁴ ist, gilt $g_A = 2 \cdot g_D$. Mit Hilfe der Neutralitätsbedingung

$$n_0 + N_A^- = p_0 + N_D^+ \quad (1.20)$$

kann das Fermi-niveau bestimmt werden. Im folgenden sei lediglich der Fall der Dotierung mit Donatoren, d.h. $N_A = 0$, betrachtet. Setzt man dann neben (1.18) mit $E_D = E_C - E_D^*$, (1.6) und (1.7) mit $E_V = E_C - E_G$ in (1.20) ein, so erhält man

$$N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) = \frac{N_D}{1 + 2 \exp\left(\frac{E_D^* - (E_C - E_F)}{k_B T}\right)} + N_V \exp\left(-\frac{E_G - (E_C - E_F)}{k_B T}\right). \quad (1.21)$$

⁴ Für sehr hohe Dotierungen $N_{A,D} > 10^{19} \text{cm}^{-3}$ spalten sich das Donatorniveau, bzw. Akzeptorniveau weiter auf und bilden ein Band. Dabei wird entsprechend dieser Niveaufspaltung g immer größer und kann schließlich als gegen unendlich strebend betrachtet werden. In Gleichung (1.17) bedeutet dies, daß $f(E) = 1$ wird und deshalb die Zustände dieses Bandes voll besetzt sind. Da dieses Band mit dem Leitungs- bzw. dem Valenzband überlappt, sind dann immer und für jede Temperatur (auch für $T = 0\text{K}$) freibewegliche Ladungsträger vorhanden. Der Halbleiter zeigt somit metallisches Verhalten. Man spricht in diesem Zusammenhang von entarteten Halbleitern (vgl. [49]).

Mit Hilfe dieser Gleichung kann das Fermineau relativ zum Leitungsbandniveau berechnet werden. Hierbei ist anzumerken, daß für (1.6), $E_C - E_F \gg k_B T$ vorausgesetzt wurde⁵. Gerade für höhere Dotierdichten, d.h. für ca. 10^{18}cm^{-3} und höher, gilt diese Näherung nicht mehr und zur Berechnung von n_0 mittels (1.5) muß dann (1.1) verwendet werden. Dieses Integral muß dann numerisch gelöst werden und die Bestimmung des Fermineaus wird wesentlich komplizierter.

Kennt man nun das Fermineau läßt sich N_D^+ gemäß (1.18) berechnen. Gerade für hohe Dotierdichten und insbesondere fallende Temperaturen wird der prozentuale Anteil der ionisierten Störstellen an der Dotierstellendichte immer geringer. Deshalb nennt man diesen Effekt auch Ausfrieren von Ladungsträgern (carrier freeze-out). Wie oben bereits erwähnt spielt dieser Effekt selbst bei $T = 300\text{K}$ eine gewisse Rolle und muß in den Modellen gemäß obiger Beschreibung berücksichtigt werden.

Wegen (1.10) läßt sich (1.20) schreiben als

$$n_0 = \frac{n_i^2}{n_0} + N_D^+ \quad (1.22)$$

Dies ergibt eine quadratische Gleichung bezüglich n_0 . Ihre Lösung lautet

$$n_0 = \frac{N_D^+}{2} \left(1 + \sqrt{1 + \left(\frac{2n_i}{N_D^+} \right)^2} \right) \quad (1.23)$$

Selbst kleinste noch realisierbare Dotierdichten im Bereich 10^{12}cm^{-3} sind größer als n_i mit etwa $1,4 \cdot 10^{10} \text{cm}^{-3}$. Daher gilt für n_0 mit sehr guter Näherung für nahezu alle Dotierdichten

$$n_0 \approx N_D^+ \quad (1.24)$$

1.2.3 Reduktion der Bandlücke bei hohen Dotierungen

Wie oben beschrieben, beginnen sich bei höheren Dotierungen die Donatoren- bzw. Akzeptorenniveaus zu Bändern aufzuweiten und es kann zu Überlappungen mit dem Leitungs- bzw. Valenzband kommen. Neben anderen Effekten wird dieser Degenerationseffekt als Hauptursache für die Reduktion der Bandlücke um ein $\Delta E_G(N_{A,D}, T)$ angesehen. Die Diskussion über die richtige theoretische Beschreibung dieses Effekts ist schwierig und noch nicht abschließend geklärt. Deshalb behilft man sich empirisch abgeleiteter Gleichungen für ΔE_G . In der Anleitung für den Bauelementesimulator Dessis [30], werden eine ganze Reihe Formulierungen verschiedener Autoren für das im Englischen „Bandgap-Narrowing“ (BGN) genannte Phänomen vorgestellt. In

⁵Allgemein akzeptiert wird die Boltzmann-Näherung für $E_C - E_F \geq 4k_B T$, wobei $4k_B T$ ca. 100mV entsprechen.

[30] wird ferner eine Ableitung präsentiert, wie dieses ΔE_G auf die Ladungsträgerstatistik wirken soll, nämlich über eine Erhöhung der Eigenleitungsträgerdichte n_i . Gleichung (1.11) wird entsprechend erweitert, so daß sich

$$n_{i,eff} = n_i \cdot \exp\left(\frac{\Delta E_G}{2k_B T}\right). \quad (1.25)$$

ergibt. Aus den verschiedenen empirischen Modellen für $\Delta E_G(N_{A,D}, T)$ wurde die Beschreibung nach Slotboom/De Graaf [65] verwendet. Diese lautet

$$\Delta E_G = 9meV \cdot \left(\ln\left(\frac{N_{A,D}}{10^{17} \text{cm}^{-3}}\right) + \sqrt{\left[\ln\left(\frac{N_{A,D}}{10^{17} \text{cm}^{-3}}\right)\right]^2 + 0,5} \right). \quad (1.26)$$

Das so errechnete ΔE_G wird dann lediglich in Gleichung (1.25) eingesetzt. Diese Vorgehensweise ist nicht unumstritten, da E_G noch auf andere physikalische Größen wirkt. Im Modell wurden deshalb drei Alternativen implementiert, zwischen denen „umgeschaltet“ werden kann: (1) kein BGN, (2) BGN mit (1.26) in Verbindung mit (1.25) und schließlich (3) BGN, bei dem ΔE_G gemäß (1.26) außerdem auf die effektive Elektronenmasse in (1.13), folglich auf N_C und damit zusätzlich auf $n_{i,eff}$ wirkt.

1.3 Die Halbleiter–Grundgleichungen

1.3.1 Die Maxwellgleichungen

Die aus der Elektrodynamik bekannten Maxwellgleichungen gelten natürlich auch in der Halbleiterphysik. Diese bilden ein geschlossenes System, bestehend aus vier Gleichungen, zur Beschreibung elektromagnetischer Felder. Für homogenes und isotropes Halbleitermaterial lauten die Maxwellschen Gleichungen (vgl. z.B. [68])

$$\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{j} \quad \text{Durchflutungsgesetz} \quad (1.27)$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad \text{Induktionsgesetz} \quad (1.28)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad \text{Quellenfreiheit des } \vec{B}\text{-Feldes} \quad (1.29)$$

$$\vec{\nabla} \cdot \vec{D} = \rho \quad \text{Poisson-Gleichung} \quad (1.30)$$

wobei \vec{H} die magnetische Feldstärke, \vec{B} die magnetische Flußdichte, \vec{E} die elektrische Feldstärke, $\vec{D} = \epsilon_r \epsilon_0 \vec{E}$ die dielektrische Verschiebungsdichte, \vec{j} die Konvektionsstromdichte und ρ die elektrische Raumladungsdichte darstellt.

1.3.2 Die Transportgleichungen

Die Konvektionsstromdichten \vec{j}_p für Löcher und \vec{j}_n für Elektronen setzen sich bei Halbleitern jeweils aus zwei Komponenten, nämlich aus einem Driftterm \vec{j}_{drift} und einem Diffusionsterm \vec{j}_{diff} , zusammen. Wie bereits in der Einleitung erwähnt, genügt zur Modellierung meist eine eindimensionale mathematische Beschreibung. Damit lauten die Stromdichtegleichungen

$$j_p = j_{drift,p} + j_{diff,p} = q_0 \mu_p p E - q_0 D_p \frac{\partial p}{\partial x} \quad (1.31)$$

$$j_n = j_{drift,n} + j_{diff,n} = q_0 \mu_n n E + q_0 D_n \frac{\partial n}{\partial x} \quad (1.32)$$

wobei μ_n und μ_p die sogenannten Beweglichkeiten für Elektronen bzw. Löcher darstellen, die in Abschnitt 1.5 näher beschrieben werden. D_n und D_p sind die sogenannten Diffusionskoeffizienten, die über die Einstein-Relationen

$$D_n = \frac{k_B T}{q_0} \mu_n \quad \text{und} \quad D_p = \frac{k_B T}{q_0} \mu_p \quad (1.33)$$

mit den Beweglichkeiten verknüpft sind. Zur Berechnung der Gesamtstromdichte muß die Summe von (1.31) und (1.32) gebildet werden. Man erhält also für die Gesamtstromdichte an jedem beliebigen Ort x den Ausdruck

$$j_{ges} = j_p + j_n \quad (1.34)$$

Setzt man überall Flächengleichheit senkrecht zur Stromflußrichtung voraus, was ein wesentliches Charakteristikum eindimensionaler Modellierung ist, bedeutet dies, daß $j_{ges}(x) = const.$ gilt.

1.3.3 Die Kontinuitätsgleichungen

Bildet man die Divergenz von (1.27) ergibt sich aufgrund der Rechenregel (vgl. [10]) $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{H}) = 0$

$$\vec{\nabla} \cdot \left(\vec{j} + \frac{\partial \vec{D}}{\partial t} \right) = \vec{\nabla} \cdot \vec{j} + \frac{\partial}{\partial t} (\vec{\nabla} \cdot \vec{D}) = 0 \quad (1.35)$$

Setzt man (1.30) in (1.35) ein, ergibt sich die allgemeine Kontinuitätsgleichung

$$\vec{\nabla} \cdot \vec{j} + \frac{\partial \rho}{\partial t} = 0 \quad (1.36)$$

Mit $\rho_p = q_0 p$ und $\rho_n = -q_0 n$ lassen sich anstelle von (1.36) ladusträgerspezifische Kontinuitätsgleichungen schreiben. Dazu werden wiederum eindimensionale Verhältnisse angenommen. Ferner müssen Rekombinations- und Generationsvorgänge gemäß Abschnitt 1.4 berücksichtigt werden. Dies geschieht mit Hilfe der sogenannten Nettorekombinationsrate R_{netto} , die der Differenz von Rekombinations- und Generationsrate entspricht, d.h. es gilt $R_{netto} = R - G$ für Löcher. Ist R_{netto} negativ, überwiegt der Generationsprozeß. Damit erhält man

$$\frac{\partial p}{\partial t} = -R_{netto,p} - \frac{1}{q_0} \frac{\partial j_p}{\partial x} \quad (1.37)$$

$$\frac{\partial n}{\partial t} = -R_{netto,n} + \frac{1}{q_0} \frac{\partial j_n}{\partial x}. \quad (1.38)$$

Subtrahiert man die beiden Gleichungen voneinander erhält man (1.36). Betrachtet man also die Gesamtladung $\rho = q(p - n)$ fällt R_{netto} weg, d.h. die gespeicherte Ladung kann sich lediglich durch Zu- oder Abfluß ändern. Setzt man (1.31) in (1.37) und (1.32) in (1.38) ein, erhält man partielle Differentialgleichungen, die das zeitliche und örtliche Ladungsträgerverhalten im Halbleiter beschreiben (vgl. Kapitel 2). Die Nettorekombinationsrate R_{netto} hängt nicht zuletzt von den Trägerdichten selbst ab. Um also die Differentialgleichungen zu komplettieren, muß für R_{netto} ein geeigneter Ausdruck gefunden werden. Dies geschieht in Abschnitt 1.4.

1.4 Rekombinationsvorgänge

1.4.1 Der Begriff Lebensdauer

Bereits in Abschnitt 1.2.1 wurden die Begriffe Generation mit der Generationsrate $G(T)$ und Rekombination mit der Rekombinationsrate $R(T, n, p)$ eingeführt, denn für thermisches Gleichgewicht gleichen sich die beiden Vorgänge so aus, daß das Produkt aus Löcherdichte und Elektronendichte gemäß (1.10) genau $n_i^2(T)$ entspricht. Im folgenden sei ein homogenes Stück n-Typ-Halbleiter⁶ betrachtet. In ihm werden durch kurzzeitige Bestrahlung mit einem Lichtimpuls, zusätzliche Elektronen–Loch–Paare generiert, vorausgesetzt es handelt sich um einen direkten Halbleiter⁷. Die

⁶Bei dotierten Halbleitern lassen sich wegen der Dominanz einer Ladungsträgersorte bestimmte Näherungen durchführen, wie bei der Herleitung von (1.41) aus (1.40) zu sehen ist. Damit läßt sich die Größe Lebensdauer auf relativ einfache Art und Weise einführen.

⁷Bei direkten Halbleitern, wie z.B. Galliumarsenid, sind strahlende Übergänge, also Rekombination unter Abgabe eines Photons, vorherrschend. Silizium dagegen ist ein sogenannte indirekter Halbleiter, dessen dominanter nichtstrahlender Rekombinationsmechanismus in Abschnitt 1.4.2 beschrieben wird.

zusätzliche Elektronendichte, die sich beim, bzw. kurz nach dem Abschalten der Generationsquelle Licht ergibt, sei n' und wegen des paarweisen Auftretens gilt dann $n' = p'$. Damit ist das thermische Gleichgewicht gestört und es gilt mit $n = n_0 + n'$ und $p = p_0 + p'$

$$np > n_i^2 \quad (1.39)$$

Nachdem das Licht ausgeschaltet ist und damit die Störung des thermischen Gleichgewichts wegfällt, setzt ein Ausgleichsprozeß ein, um den Ladungsträgerüberschuß $np - n_i^2$ wieder abzubauen. Dieser Ausgleichsprozeß heißt Rekombination⁸ und er wirkt mit der spezifische Größe (vgl. (1.8)), dem sogenannten Rekombinationskoeffizienten $r(T)$. Damit ergibt sich für den Abbau des Minoritätsträgerüberschuß (hier der Löcher), die folgende Gleichung⁹:

$$\frac{dp'}{dt} = r(np - n_i^2). \quad (1.40)$$

Setzt man nun $n = n_0 + n'$ und $p = p_0 + p'$ ein, ergibt sich

$$\begin{aligned} \frac{dp'}{dt} &= r(n_0 + n')(p_0 + p') - rn_i^2 \\ &= r \underbrace{(p_0 n_0 - n_i^2)}_0 + rn_0 p' + rp_0 n' + rp' n' \end{aligned}$$

Bei dem betrachteten n-Typ-Halbleiter gilt $n_0 \approx N_D^+ \gg p_0, n', n'$, weswegen man alle Summanden in obiger Gleichung gegenüber $rn_0 p'$ vernachlässigen kann. Somit kann man schreiben:

$$\frac{dp'}{dt} = \frac{p'}{1/rn_0} = \frac{p'}{\tau_p} \quad (1.41)$$

Wie man sieht, handelt es sich hierbei um eine gewöhnliche lineare Differentialgleichung erster Ordnung. Als Lösung derselben ergibt sich ein exponentielles Abklingen der Überschußlöcherdichte mit der Zeitkonstante $1/rn_0$. Da $n_0 \approx N_D^+ = const$ gilt, handelt es sich tatsächlich um eine Konstante, die Minoritätsträgerlebensdauer τ_p (Index p für Löcher) heißt. Stellt sich abschließend die Frage wie schnell die Majoritätsträger (hier n) abklingen. Wären anfangs lediglich Überschußelektronen n' in den n-Typ-Halbleiter gelangt (z.B. durch Injektion), hätten diese die Tendenz mit der sogenannten Relaxationszeit $\tau_d = \epsilon_r \epsilon_0 / \sigma$, die meist um mindestens den Faktor 1000 geringer als die Minoritätsträgerlebensdauer ist, abzuklingen. Im vorliegenden Fall aber, also bei paarweiser Generation, klingen die Elektronen aufgrund der Neutralitätsbedingung (1.20) wie die Löcher, also mit der größeren Zeitkonstante τ_p ab.

⁸Eigentlich handelt es sich um die in Abschnitt 1.3.3 eingeführte Nettorekombination R_{netto}

⁹Entspricht (1.37) für $\partial/\partial x = 0$

1.4.2 Rekombination mittels Rekombinationszentren

Bei indirekten Halbleitern, wie beim hier betrachteten Silizium, liegen das Valenzbandmaximum und das Leitungsbandminimum im \vec{k} -Raum nicht direkt übereinander, d.h. zwischen ihnen liegt ein Δk . Damit ein Elektron direkt vom Leitungsband ins Valenzband gelangen kann, muß neben der Energieabgabe in Form eines Photons auch ein Impulsausgleich proportional zu Δk mit dem Gitter stattfinden. Doch dieses Ereignis ist bei indirekten Halbleitern eher selten, so daß Rekombinationsvorgänge überwiegend indirekt, über sogenannte Rekombinationszentren ablaufen und nichtstrahlend sind. Diese Rekombinationszentren sind zum einen Kristalldefekte, die mittels Bestrahlung erzeugt werden können und zum anderen spezielle Fremdatome, wie beispielsweise Gold, die mittels Dotierung eingebracht werden. Diese Rekombinationszentren befinden sich auf Energieniveaus, die sich innerhalb der Bandlücke befinden. Somit läuft der Rekombinationsvorgang in zwei Stufen ab: Ein Elektron wird von einer Störstelle eingefangen¹⁰. Fängt diese Störstelle ein Loch vom Valenzband oder gibt es das überschüssige Elektron ans Valenzband ab, ist der Rekombinationsvorgang vollständig. Analog kann mittels dieser Zentren ein Generationsmechanismus ausgelöst werden. Insgesamt ergeben sich vier Teilmechanismen, nämlich Elektroneneinfang, Löchereinfang, Elektronenabgabe und Löcherabgabe. Um eine Beschreibung des Gesamtmechanismus zu erhalten, muß für jeden dieser Teilmechanismen, die Ladungsträgerstatistik abgeleitet werden. Auf diese komplizierte, als Shockley–Read–Hall–Modell bekannte, mathematische Beschreibung, soll an dieser Stelle verzichtet werden (vgl. z.B. Anhang von [49]), denn zu Modellierungszwecken genügt das Ergebnis, wie es auch in [30] verwendet wird. Der Ausdruck für die Nettorekombination lautet demnach

$$R_{netto}^{SRH} = \frac{np - n_{i,eff}^2}{\tau_p(n + n_1) + \tau_n(p + p_1)}, \quad (1.42)$$

wobei sich die Parameter n_1 und p_1 zu

$$n_1 = n_{i,eff} \cdot \exp\left(\frac{E_T - E_{Fi}}{k_B T}\right) \quad (1.43)$$

$$p_1 = n_{i,eff} \cdot \exp\left(-\frac{E_T - E_{Fi}}{k_B T}\right) \quad (1.44)$$

bestimmen lassen. Hierbei ist E_T das maßgebliche Trapniveau. Für Gold oder Neutronenbestrahlung liegt dieses ungefähr in der Bandmitte, d.h. es fällt mehr oder weniger mit dem Eigenleitungsferminiveau E_{Fi} zusammen, so daß man gerne näherungsweise $n_1 \approx p_1 \approx n_{i,eff}$ in (1.42) verwendet. Allerdings ist zu bedenken,

¹⁰Deshalb heißen die Rekombinationszentren auch Haft- oder Fangstellen, bzw. traps (engl. für „Ladungsträgerfallen“)

daß z.T. schon kleine Werte für $E_T - E_{F_i}$ zu recht großen Werten der Exponentialfunktion in (1.43) und (1.44) führen. So gilt z.B. für $E_T - E_{F_i} = 100\text{meV}$ bereits $n_1 \approx 48 n_{i,eff}$.

Bei elektronischen, insbesondere bipolaren Halbleiterbauelementen, kommt es bei Anlegen von bestimmten Spannungen zu Stromfluß und zur gegenseitigen Injektion von Ladungsträgern zwischen den einzelnen in Art und/oder Stärke unterschiedlich dotierten Zonen (vgl. Abschnitt 1.6). Diese injizierten Überschußladungsträger sind Ausdruck dafür, daß das thermische Gleichgewicht im Bauelement gestört ist. Wird der Strom abgeschaltet, bauen Rekombinationsvorgänge die Überschußladungsträger ab¹¹.

Für die Betrachtung von Bauelementen sind nun zwei sehr häufig auftretende Spezialfälle bedeutsam, nämlich die sogenannte schwache und die starke Injektion. Im folgenden wird wieder eine Schicht n–Typ–Halbleiter betrachtet. Eine angrenzende p–Schicht injiziert Löcher in die n–Schicht. Damit stellt sich die Frage mit welcher Zeitkonstante die Minoritätsträger Löcher für die beiden genannten Fälle abgebaut werden:

Fall 1: schwache Injektion

Die Dotierdichte N_D sei in der Größenordnung 10^{18}cm^{-3} . Die Dichte der injizierten Löcher p direkt an der Grenze zur p–Schicht befindet sich realistisch im Bereich 10^{15}cm^{-3} – 10^{16}cm^{-3} . Da E_T etwa in der Bandmitte liegen soll, liegen n_1 und p_1 aus (1.42) ähnlich wie n_i in der Größenordnung 10^{10}cm^{-3} . Somit kann man die Näherung $n = n_0 \approx N_D \gg p \gg n_i, p_1, n_1$ auf (1.42) anwenden und erhält

$$R_{netto}^{SRH} = \frac{p}{\tau_p}. \quad (1.45)$$

Die Minoritätsträger Löcher werden also mit der Löcherlebensdauer τ_p abgebaut. Entsprechendes gilt bei Elektroneninjektion in stark dotierte p–Schichten, indem man in (1.45) einfach p durch n und τ_p durch τ_n ersetzt.

Fall 2: starke Injektion

Gerade bei Leistungshalbleitern gibt es schwach, meist n–dotierte Zonen mit N_D im Bereich von 10^{13}cm^{-3} . Die Dichte der injizierten Löcher liegt wieder im Bereich 10^{15}cm^{-3} – 10^{16}cm^{-3} , also etwa 1000–fach größer als die Dotierdichte. Um die Neutralität gemäß (1.20) zu wahren, müssen der Schicht entsprechend zusätzliche Elektronen über $n_0 \approx N_D$ hinaus zuffießen. Somit gilt die sogenannte Hochinjekti-

¹¹Zusätzlich werden Ladungsträger beim Abschalten eines Bauelements aufgrund eines im Zusammenhang mit der Spannungsaufnahme auftretenden Ausräumvorgangs abgebaut (vgl. dazu Kapitel 2)

onsbedingung $n \approx p \gg N_D, n_i, p_1, n_1$. Damit erhält man für (1.42)

$$R_{netto}^{SRH} = \frac{p}{\tau_p + \tau_n} = \frac{p}{\tau_h}. \quad (1.46)$$

Die effektiv wirksame Lebensdauer für die Löcher entspricht für Hochinjektion also der Summen aus den Minoritätsträgerlebensdauern für Elektronen und Löchern und heißt Hochinjektionslebensdauer τ_h . Dasselbe Ergebnis erhält man selbstverständlich wenn man von schwach p-dotierten Zonen ausgeht.

Die Lebensdauern für Löcher τ_p and Elektronen τ_n wurden in (1.42) quasi wie Konstanten behandelt. Tatsächlich sind sie aber von der Dotierdichte abhängig. Bis $N_{A,D} \approx 10^{16} \text{cm}^{-3}$ kann man die beiden Lebensdauern als konstant betrachten. Für größere Dotierdichten verringern sich die Lebensdauern relativ stark. Mangels Ausarbeitung theoretischer Grundlagen zur Quantifizierung dieses Effekts, bedient man sich einmal mehr empirischer Fitformeln. Beliebte ist hierbei die auch in [30] verwendete Scharfetter–Beziehung, die in ihrer einfachsten Form

$$\tau_{n,p}(N_{A,D}) = \frac{\tau_{n0,p0}}{1 + \frac{N_{A,D}}{N_{refA,D}}}. \quad (1.47)$$

lautet. Für $N_{refA,D}$ wählt man meist 10^{16}cm^{-3} . Außerdem sind Lebensdauern auch temperaturabhängig, wobei höhere Temperaturen mit längeren Lebensdauern einhergehen. Leider liegen auch hier nur wenige Forschungsarbeiten hinsichtlich einer detaillierten physikalischen Beschreibung vor, deren Ergebnisse besonders für tiefe Temperaturen mit Unsicherheiten behaftet zu sein scheinen. Für die vorliegenden Bauelemente wird hierzu die Formel

$$\tau(T) = \tau(300K) \cdot \left(\frac{T}{300K} \right)^\alpha \quad (1.48)$$

verwendet, wie sie auch in [25] zu finden ist. Der Parameter α kann hierbei als Fitparameter auf Basis von Messungen dienen, oder er wird standardmäßig auf den bewährten Wert von $\alpha = 1,5$ zurückgegriffen.

1.4.3 Auger–Rekombination

Wie in Abschnitt 1.4.2 bereits erläutert wurde, ist ein Rekombinationsvorgang bei indirekten Halbleitern nur möglich, wenn ein entsprechende Impulsabgabe des rekombinierenden Teilchens stattfinden kann. Alternativ zu festen Rekombinationszentren, bieten sich hierfür auch andere freibewegliche Teilchen des rekombinierenden Typs an. Kann also ein Teilchen seinen Impuls an eine entsprechendes „Nachbarteilchen“ des gleichen Typs abgeben, kommt es zur Rekombination im Sinne eines

direkten Band–Band–Übergangs. Dieser Rekombinationsmechanismus heißt Auger–Effekt und schließt einen hier nicht weiter beschriebenen Generationsmechanismus mit umgekehrtem Ablauf ein¹².

Im folgenden werden die relevanten Ausdrücke zur quantitativen Beschreibung des Auger–Effekts gemäß [64] hergeleitet. Betrachtet man nun die Elektronen (Löcher) im Leitungsband (Valenzband), so wird die Dichte derjenigen Elektronen (Löcher), die auf diese Weise mit einem Loch (Elektron) rekombinieren, proportional zur Anzahl der überschüssigen Elektronen–Loch–Paare $np - n_{i,eff}^2$ sein. Weiterhin ist sie proportional zur Dichte der Leitungselektronen n (Löcher p), da jedes einzelne rekombinierende Elektron (Loch) seinen Impuls an jeweils einem anderen sich im Leitungsband (Valenzband) befindlichen Elektron (Loch) abgeben muß. Ferner benötigt man zur Bestimmung der Netto–Auger–Rekombinationsraten für Elektronen bzw. Löcher noch einen Koeffizienten $C_{An,p}$, die sogenannte Einfangrate für Elektronen bzw. Löcher. Die gesamte Netto–Auger–Rekombinationsrate ergibt sich somit zu

$$R_{netto}^{Auger} = (C_{An} n + C_{Ap} p) (np - n_{i,eff}^2). \quad (1.49)$$

Die Einfangraten $C_{An,p}$ werden mit Hilfe von Anpassung an Meßdaten bestimmt (vgl. [64]). Dazu finden sich in der Literatur mehrere verschiedene Werte für die Parameter. Sie alle sind im Bereich $10^{-31} - 10^{-30} cm^6/s$ und ihre Temperaturabhängigkeit ist vergleichsweise gering.

Diese Gleichung läßt sich ähnlich wie (1.42) wiederum für die Fälle schwache und starke Injektion vereinfachen (vgl. S. 19). Wiederum wird stellvertretend ein n–Typ–Halbleiterstück betrachtet, in das Löcher injiziert werden.

Fall 1: schwache Injektion

Die Gleichgewichtselektronendichte n_0 , die bei vollständiger Ionisation der Donatordichte N_D entspricht, ist sehr viel größer als die Dichte der injizierten Löcher p . Mit $np > n_{i,eff}^2$ erhält man aus (1.49):

$$R_{netto}^{Auger} = C_{An} n_0^2 p = \frac{p}{\tau_{Ap}} \Leftrightarrow \tau_{Ap} = \frac{1}{C_{An} n_0^2} \quad (1.50)$$

Bei p–dotierten Halbleitern bekommt man entsprechend $\tau_{An} = 1/(C_{Ap} p_0^2)$. Die Auger–Minoritätsträgerlebensdauern $\tau_{Ap,n}$ hängen bei schwacher Injektion also nur von der Dotierstoffkonzentration (wegen $n_0, p_0 \approx N_{D,A}$) ab und sind umso geringer,

¹²Der direkt Band–Band–Übergang ist nicht der einzige Auger–Effekt. Unter anderen, wird vor allem dem sogenannten Trap–Assisted–Auger–Effekt Bedeutung zugemessen. Aufgrund der Beteiligung von Haftstellen ähnelt dessen theoretische Beschreibung der Shockley–Read–Hall–Theorie. Da sich wegen dieser Ähnlichkeit beide Effekte aber kaum voneinander trennen lassen, kann dieser Auger–Effekt in die SRH–Beschreibung gemäß Abschnitt 1.4.2 integriert werden, weshalb in dieser Arbeit auf eine detaillierte Beschreibung verzichtet wird.

je größer die Dotierdichte ist.

Fall 2: starke Injektion

Wiederum gelte die Hochinjektionsbedingung $n \approx p \gg N_D$. Damit vereinfacht sich (1.49) zu

$$R_{netto}^{Auger} = (C_{An} + C_{Ap})p^3 = C_A p^3 = \frac{p}{\tau_{Ah}} \Leftrightarrow \tau_{Ah} = \frac{1}{C_A p^2} \quad (1.51)$$

Die Auger–Hochinjektionslebensdauer τ_{Ah} hängt nun von der Dichte der injizierten Ladungsträger ab. Alle vorher behandelten Lebensdauern waren nur von Dotierungen oder der Temperatur abhängig, d.h. sie waren an jedem Ort eines Halbleiterbauelements konstant, soweit die Dotierdichten $N_{A,D}$ konstant waren. Da z.B. verschiedene Stromdichten mit unterschiedlichen Dichten injizierter Ladungsträger einher gehen, gilt dies für τ_{Ah} wegen (1.51) nicht mehr. Je höher also die Dichte der injizieren Ladungsträger ist, desto niedriger wird τ_{Ah} .

Damit stellt sich abschließend die Frage welchen Einfluß der Auger–Effekt im Vergleich zum dominanten Shockley–Read–Hall–Mechanismus auf das gesamte Rekombinationsverhalten hat. Da sich beide Mechanismen gegenseitig nicht beeinflussen, kann die effektiv wirksame Gesamtlebensdauer mittels

$$\frac{1}{\tau_{gesk}} = \frac{1}{\tau_{SRHk}} + \frac{1}{\tau_{Ak}} \quad (1.52)$$

berechnet werden, wobei die in Abschnitt 1.4.2 abgeleiteten Lebensdauern mit einem Index SRH versehen sind. Der Index k kann jeweils durch p,n bzw. h ersetzt werden, d.h. (1.52) gilt sowohl für die Minoritätslebensdauern als auch für die Hochinjektionslebensdauer. Für die Bauelementemodelle wird lediglich die Auger–Hochinjektionslebensdauer τ_{Ah} verwendet¹³. Die hierfür benutzte Formel inklusive Temperaturabhängigkeit lautet (vgl. [60] und [58]):

$$\tau_{Ah} = \frac{1}{10^{-30} \text{cm}^6 \text{s}^{-1} \cdot p^2} \left(\frac{T}{300\text{K}} \right)^{-0,6} \quad (1.53)$$

Damit läßt sich nun der Einfluß des Auger–Effekts abschätzen. Die SRH-Hochinjektionslebensdauer betrage beispielsweise $1\mu\text{s}$. Damit die Gesamtlebensdauer um 10% absinkt, d.h. auf $0,9\mu\text{s}$, muß τ_{Ah} gemäß (1.52) genau $10\mu\text{s}$ betragen. Setzt man dies in (1.53) ein, bekommt man für die injizierte Löcherdichte bei schwach n -dotierten Zonen $p = 10^{17} \text{cm}^{-3}$. Für alle geringere Löcherdichten spielt der Auger–Effekt eine geringere Rolle; für alle Löcherdichten darüber wird die Dominanz des SRH–Mechanismus geringer. Insgesamt läßt sich also sagen, daß der Auger–Effekt erst bei sehr starker Injektion von Ladungsträgern zur Geltung kommt.

¹³Zur Begründung siehe Abschnitt 1.4.4

1.4.4 Bauelementemodelle und Lebensdauern

In diesem Abschnitt soll im Vorgriff auf Kapitel 2 angedeutet werden, wie die oben vorgestellten Lebensdauern in den Bauelementemodellen Verwendung finden. In reinen Bauelementesimulatoren, wie z.B. Dessis [30], werden globale Minoritätsträgerlebensdauern τ_n und τ_p vom Anwender ausgewählt. Über (1.42), (1.47) und (1.48) wird dann an jedem Ort die effektiv wirkende Lebensdauer berechnet. Ähnliches gilt für den Auger-Effekt.

Dieses Vorgehen ist für die allerdings Schaltungssimulation unzuweckmäßig. Der Grund hierfür liegt an der Zielsetzung und damit auch an der Zielgruppe für die beiden verschiedenen Simulationsformen. Die Bauelementesimulation nutzen Bauelementeentwickler, die trivialerweise alle Designparameter der zu entwickelnden, bzw. zu optimierenden Bauelemente kennen. Dagegen wird die Schaltungssimulation von Schaltungsentwicklern verwendet, denen die Strukturdaten von bereits am Markt erhältlichen Bauelementen und oftmals auch die Zusammenhänge physikalischer Parameter, weitgehend unbekannt sind. Deshalb sollte der Parametersatz auch von physikalischen Modellen möglichst anschaulich sein, d.h. die Wirkung von Parametern auf das Simulationsergebnis sollen nachvollziehbar und die Anzahl der wichtigen Parameter auf möglichst Wenige begrenzt sein, um die Anpassung eines Modells an das reale Verhalten eines Bauelementes mit vertretbarem Aufwand zu ermöglichen.

Dabei sind gerade Lebensdauern schwierig zu handhaben, da sie entweder gar nicht, oder aufgrund der Natur der Meßverfahren nur ungenau bekannt sind. Ferner ist auch die Abhängigkeit von der Dotierdichte gemäß (1.47) mit Unsicherheiten behaftet. Selbst in [30] wird gefordert, die Lebensdauern und die Referenzdotierung in (1.47) als Fitparameter zu betrachten. Für die Modellierung hat sich folgende Vorgehensweise bewährt (vgl. Kapitel 2):

1. Für jede Zone wird festgelegt, ob überwiegend starke oder schwache Injektion vorliegt. Damit benötigt man prinzipiell nur noch die Vereinfachungen (1.45), (1.46), (1.50) und (1.51). Da „mittlere“ Injektion bei Leistungshalbleiterbauelementen im allgemeinen nur eine untergeordnete Rolle spielt, ist dieses Vorgehen durchaus sinnvoll.
2. Die Lebensdauern werden „entkoppelt“, d.h. jede Zone bekommt ihre eigene Lebensdauer. So wird für die schwach dotierte Driftzone, der Parameter Hochinjektionslebensdauer als wichtige Kenngröße zur Adaption der Speicherladung, unabhängig von den Lebensdauern von Anode, der Kathode, oder anderen Schichten, eingestellt. Des weiteren kann nach Wunsch die Hochinjektionslebensdauer ortsabhängig als sogenanntes Lebensdauerprofil eingegeben werden.

3. Die Minoritätsträgerlebensdauern in den Randzonen, d.h. Anode und Kathode, gehen neben anderen Parametern, wie Zonenweite, Dotierung und Diffusionskonstante in die sogenannte Sperrsättigungsstromsdichte $j_{n0,p0}$ ein. Wie in Abschnitt 2.2 erläutert wird, legt dieses $j_{n0,p0}$, die Injektionseffizienz der entsprechenden Zone eindeutig fest. Somit läßt sich die Injektionseffizienz von Anode und Kathode definieren, ohne die Strukturdaten der Zonen selbst kennen zu müssen.
4. Das Anpassen mittels $j_{n0,p0}$ hat einen Nachteil: die Anpassung gilt nur für die jeweilige Temperatur. Für jede andere Temperatur muß $j_{n0,p0}$ nachjustiert werden, da eine einfache Formel $j_{n0,p0} = f(T)$ nicht erhältlich ist. Wenn also Simulationen in einem weiten Temperaturbereich von Interesse sind, wie in [57], kommt man nicht umhin, die Strukturdaten zu kennen.
5. Zum Zwecke der Modellierung wurde ein Weg gefunden, 3. und 4. zu verbinden. Für die Parameter Dotierung und Zonenweite werden vom Modell Anfangswerte bereitgestellt, die geändert oder auch akzeptiert werden können. Die Anpassung der Injektionsfähigkeit für $T = 300K$ kann nun wieder mit $j_{n0,p0}(300K)$ vorgenommen werden. Für Anwender, für die Temperaturbetrachtungen uninteressant sind, ist der Anpassungsvorgang damit beendet. Für alle Anderen ist bedeutsam zu wissen, daß im Modell aus $j_{n0,p0}(300K)$ und den anderen Größen, die Lebensdauer der jeweiligen Randzone extrahiert wird. Diese wird dann für beliebige andere Temperaturen mittels (1.48) umgerechnet. Zusammen mit allen weiteren maßgeblichen Parametern, die ebenfalls soweit notwendig für die neue Temperatur berechnet werden, wird dann das nun gültige $j_{n0,p0}(T_{neu})$ berechnet (s. 2.2.3).
6. Wie aus 5. hervorgeht, kann auf eine Beschreibung des Auger-Effekt gemäß (1.50) verzichtet werden. In den Modellen ist der Auger-Effekt aber für die schwach dotierte Driftzone gemäß (1.51) implementiert.
7. Bei der Modellierung der schwach dotierten Driftzone wird für R_{netto} in (1.37) $p/\tau_{h,ges}$ gemäß (1.52) eingesetzt.

1.4.5 Stoßionisation und Lawineneffekt

Zum Abschluß dieses Abschnitts wird ein für leistungselektronische Bauelemente und damit auch für die Modellierung, sehr wichtiger Generationsvorgang beschrieben, nämlich die Stoßionisation, die in der Regel eine Ladungsträgermultiplikation zur Folge hat: der Lawineneffekt. Dieser Durchbrucheffekt erhöht den Strom und damit die auftretenden Verluste und kann sogar zur Zerstörung des betroffenen Bauelementes führen. Die Stoßionisation tritt bei Silizium auf, wenn die elektrische

Feldstärke einen kritischen Wert E_{crit} von ca. $2 \cdot 10^5 V/cm$ überschreitet. Unter Stoßionisation versteht man das Zusammenstoßen eines, von dem \vec{E} -Feld beschleunigten Ladungsträgers, mit einem anderen Ladungsträger, der an ein Gitteratom gebunden ist, mit der Folge, daß aufgrund der Abgabe von kinetischer Energie des stoßenden Teilchens an und einem Impulsausgleich mit dem gebundenen Teilchen, dieses ionisiert wird. Wird beispielsweise ein Elektron ionisiert, verbleibt gleichzeitig ein Loch, das ebenfalls freibeweglich ist. Folglich werden Elektronen–Loch–Paare generiert. Es handelt es sich also um einen Band–Band–Übergang, ähnlich dem Auger–Effekt, mit dem Unterschied, daß die dafür notwendige Energie keine thermische, sondern kinetische Energie ist. Die Grundvoraussetzung ist demnach das Vorhandensein eines Stroms. Wirklich bemerkbar macht sich der Effekt allerdings erst, wenn die Lawinenmultiplikation einsetzt, oder anders ausgedrückt: wenn ein Elektron (Loch) nicht nur ein, sondern mehrere Elektron–Loch–Paare und diese ihrerseits wiederum mehrere Elektron–Loch–Paare auf diese Weise generieren können. Die Weglänge, die ein Teilchen zwischen zwei Stößen zurücklegt, heißt frei Weglänge. Wichtig ist also, daß die Länge der Region mit großer Feldstärke, größer als diese frei Weglänge ist. Bei den schwach dotierten Driftzonen von bipolaren Bauelementen¹⁴ ist diese in der Regel immer der Fall.

Beim Lawineneffekt unterscheidet man gemäß den Betriebsbedingungen von Bauelementen zwischen dem statischen und dem dynamischen Lawineneffekt. Ersterer ist z.B. anhand der Sperrkennlinie einer Leistungsdiode sichtbar. Ab einer gewissen Sperrspannung, der sogenannten Durchbruchspannung U_{Br} , die proportional zur kritischen Feldstärke ist, steigt der Sperrstrom stark an. Die dazu notwendigen initialen Ladungsträger liefert der Sperrsättigungsstrom. Der dynamische Lawineneffekt tritt beim Abschalten bipolarer Bauelemente auf. Aufgrund von parasitären Induktivitäten kommt es am Bauelement zu Überspannungen, bei denen die elektrische Feldstärke ebenfalls den kritischen Wert erreichen bzw. übersteigen kann. Gleichzeitig fließt noch ein nicht unerheblicher Strom durch die Driftzone, so daß der Lawineneffekt ausgelöst wird. Dadurch wird der Abschaltstrom erhöht, bzw. das Abklingen zu Null zeitlich verzögert, so daß die Abschaltverlustleistung steigt. Da immer Elektronen–Loch–Paare erzeugt werden, die in Summe elektrisch neutral sind, tragen diese nicht zu einer weiteren Erhöhung der Raumladung und damit der elektrischen Feldstärke bei. Der dynamische Lawineneffekt begrenzt sich also selbst und verschwindet wenn die Überspannung abgeklungen ist. Solange das Bauelement die erhöhte Verlustleistung (sdichte) thermisch verträgt, schaltet das Bauelement im

¹⁴Bei unipolaren Bauelementen, wie dem Leistungs-MOSFET, ist die realisierbare Weite der Driftzone wegen der fehlenden Ladungsträgerüberschwemmung und des damit verbundenen hohen Bahnwiderstandes begrenzt. Damit tritt aber oftmals auch kein Lawineneffekt auf, was bedeutet, daß die maximale Feldstärke, Werte im Bereich $3 \cdot 10^5 V/cm$ erreichen darf, bis dann schließlich der Zener–Durchbruch eintritt. Damit kann das Bauelement mehr Spannung aufnehmen, als man mit dem Lawineneffekt als begrenzenden Faktor erwarten würde.

Normalfall ab¹⁵. Wie in den Kapiteln 3 und 4 gezeigt und näher erläutert wird, tritt genau dieser Fall, nämlich einwandfreies Abschalten trotz Lawineneffekts, für große Ströme bei hart und unentlastet abgeschalteten IGCTs ein.

Im folgenden wird gezeigt, wie der Lawineneffekt quantitativ beschrieben werden kann. Dazu wird eine Generationsrate G eingeführt (vgl. [49], [30]).

$$G = \alpha_n n |v_n| + \alpha_p p |v_p|. \quad (1.54)$$

Die Koeffizienten $\alpha_{n,p}$ heißen Ionisationskoeffizienten. Sie geben die Anzahl generierter Elektronen–Loch–Paare pro Längeneinheit an und sind natürlich stark vom elektrischen Feld abhängig. Zur Bestimmung dieser Koeffizienten werden in [30] wiederum empirische Modelle mehrerer Autoren angeboten, die sich voneinander nicht unerheblich in ihren Ergebnissen unterscheiden. Stellvertretend sei an dieser Stelle das beliebte Chynoweth–Gesetz [15] genannt, wobei die Parametrisierung auf Messungen von [52] zurückgehen. Die Ausdrücke für $\alpha_{n,p}$ lauten dann

$$\alpha_n = 7,03 \cdot 10^5 \text{ cm}^{-1} \gamma \exp\left(-\frac{1,23 \cdot 10^6 \text{ Vcm}^{-1} \gamma}{E}\right) \quad (1.55)$$

$$\alpha_p = 1,58 \cdot 10^6 \text{ cm}^{-1} \gamma \exp\left(-\frac{2,04 \cdot 10^6 \text{ Vcm}^{-1} \gamma}{E}\right) \quad (1.56)$$

$$\text{mit} \quad \gamma = \frac{0,84}{\tanh\left(\frac{63 \text{ meV } q_0}{2k_B T}\right)}$$

Setzt man nun (1.54) anstelle von R_{netto} in die Kontinuitätsgleichung (1.37) ein, erhält man für $\partial p/\partial t = 0$ (stationärer Fall)

$$\alpha_n \underbrace{q_0 n |v_n|}_{j_{n,drift}} + \alpha_p \underbrace{q_0 p |v_p|}_{j_{p,drift}} + \frac{\partial j_p}{\partial x} = 0. \quad (1.57)$$

Die in (1.57) als $j_{n,p,drift}$ bezeichneten Terme, entsprechen wegen $|v_{p,n}| = \mu_{p,n} E$ den Drifttermen der Transportgleichungen (1.31) und (1.32) (s. dazu Abschnitt 1.5). Beim Auftreten des Lawineneffekts sind aufgrund der hohen Feldstärke, die Driftterme gegenüber den Diffusionstermen dominant, so daß $j_{p,n} \approx j_{drift,p,n}$ gilt. Weiterhin gilt wegen (1.34): $j_n = j_{ges} - j_p$. Damit läßt sich die Gleichung (1.57) schreiben zu

$$\frac{\partial j_p}{\partial x} - (\alpha_p - \alpha_n) j_p = \alpha_n j_{ges}. \quad (1.58)$$

Hierbei handelt es sich um eine gewöhnliche, aber wegen $\alpha_{p,n} = f(E(x))$ nichtlineare Differentialgleichung bzgl. des Ortes. Die allgemeine Lösung von (1.58) ist ein relativ

¹⁵Eine Ausnahmen bildet das mögliche Wiedereinschalten von GTOs (siehe dazu z.B. [45]).

komplizierter Ausdruck, der an der Stelle des maximalen elektrischen Feldes, also am pn -Übergang (s. Abschnitt 1.6) vereinfacht werden kann (Herleitung s. z.B. [73]). Für den Multiplikationsfaktor M_{AV} , der die Verstärkung des Stroms aufgrund des Lawineneffekts angibt, folgt dann

$$M_{AV} = \frac{1}{w_{RLZ} \int_0 (\alpha_n + \alpha_p) dx}. \quad (1.59)$$

Das Integral in (1.59) wird für größer werdende Feldstärken immer größer und strebt für $E \geq E_{crit}$ gegen eins, d.h. M_{AV} geht gegen unendlich und der Strom wächst (theoretisch) über alle Grenzen.

In den Modellen von Vogler [73] wurde eine Beschreibung gemäß den (1.54) bis (1.59) implementiert. Nach Erfahrungen des Autors der vorliegenden Arbeit, ist diese Art der Modellierung des Lawineneffekts mit einer Reihe von Nachteilen verbunden. Einerseits wählt der Simulator immer kleinere Schrittweiten falls E_{crit} erreicht wird, was die Simulationsdauer z.T. inakzeptabel erhöht. Andererseits kann es dabei zu Konvergenzproblemen kommen, d.h. die Simulation bricht ab und obendrein ist die Genauigkeit des Ergebnisses oftmals unbefriedigend. Wie in Kapitel 2 noch genauer erläutert wird, kann anstatt (1.54) bis (1.59), eine Fitformel $M_{AV} = f(E)$ mit E_{crit} und einem Anpaßkoeffizienten als Parameter, verwendet verwendet. Damit läßt sich vor allem der dynamische Lawineneffekt hervorragend anhand von Meßkurven bei vertretbarem Simulationsaufwand adaptieren (s. dazu Kapitel 4).

1.5 Streuprozesse freibeweglicher Ladungsträgern

1.5.1 Arten von Streuprozessen

Die von einem elektrischen Feld beschleunigten, durch das Halbleitermaterial driftenden Ladungsträger, werden aufgrund von Streuprozessen gebremst. Dabei unterscheidet man zwischen vielen verschiedenen physikalischen Ursachen¹⁶. Diese kann man jedoch zu drei Hauptmechanismen zusammenfassen:

- (1) Thermische Gitterstreuung
- (2) Störstellenstreuung
- (3) Streuung an freien Ladungsträgern (Elektronen–Löcher–Streuung)

¹⁶Eine detaillierte Beschreibung dieser Prozesse findet sich in [5]

Zu (1): Die Gitteratome im Siliziumkristall führen abhängig von der Kristalltemperatur thermische Schwingungen aus. Die einzelnen Schwingungen sind dabei über die Bindungen miteinander verkoppelt. Sie breiten sich als elastisches Wellenpakete longitudinal oder transversal aus. Diesen Wellenpaketen kann eine Gruppengeschwindigkeit (Energiequantum) und ein Wellenvektor \vec{k} , d.h. ein Impuls, zugeordnet werden. Folglich verhalten sich die Wellenpakete bzgl. der Wechselwirkung mit Teilchen, selbst wie Teilchen im Sinne des Dualismus Teilchen–Welle. In Analogie zu den Photonen der elektromagnetischen Wellen, nennt man diese „Teilchen“ Phononen. Können diese Wellen(pakete) durch das elektrische Feld einer Lichtwelle angeregt werden, spricht man von optischen, ansonsten von akustischen Phononen. Die Wechselwirkung von bewegten Ladungsträger mit diesen Phononen heißt Gitterstreuung.

Zu (2): Die Atomrümpfe von Störstellen sind nach erfolgter Ionisation elektrisch geladen. Es liegt nahe, daß die geladenen Atomrümpfe, die sich in der Nähe befindlichen Ladungsträger je nach Vorzeichen anziehen oder abstoßen. Bei der Störstellenstreuung handelt es sich also um die klassische Coulomb–Streuung. Ihre Bedeutung nimmt mit wachsender Dotierdichte zu.

Zu (3): Bei Anlegen eines elektrischen Feldes, fließen Elektronen und Löcher in entgegengesetzte Richtungen. Dabei kann es zwischen ihnen zu Zusammenstößen kommen, zumal sie sich Elektronen und Löcher elektrostatisch anziehen. Dann spricht man von Elektronen–Loch–Streuung. Besonders bei tiefen Temperaturen, wenn (1) nur noch eine geringe Rolle spielt, kommt dieser Effekt zum Tragen. Aufgrund dieses Phänomens kann ein Ladungsträger sogar seine Flußrichtung umdrehen und sich somit entgegengesetzt zum elektrischen Feld bewegen (sog. Drag–Effekt).

1.5.2 Driftgeschwindigkeiten und Beweglichkeiten

Die in Abschnitt 1.5.1 genannten Streuprozesse führen dazu, daß die driftenden Ladungsträger immer wieder beschleunigt und nach jedem Stoß gebremst werden. Es macht demnach Sinn, von einer mittleren Geschwindigkeit in Flußrichtung \bar{v} auszugehen. Die Zeit, die durchschnittlich zwischen zwei Stößen vergeht, heißt freie Flugzeit τ_c und ist eine Konstante. Für Elektronen mit der effektiven Masse m_n^* läßt sich mittels Betrachtung der wirkenden Kräfte folgende Gleichung aufstellen:

$$m_n^* \frac{d\bar{v}_n}{dt} = \underbrace{-q_0 E}_{\text{Beschleunigen}} - \underbrace{m_n^* \frac{\bar{v}_n}{\tau_{c,n}}}_{\text{Bremsen}}. \quad (1.60)$$

Diese Gleichung ist folgendermaßen zu interpretieren: Ein Elektron habe die Anfangsgeschwindigkeit Null. Dann wird es durch das elektrische Feld beschleunigt

und die Geschwindigkeit wächst. Dies geschieht solange bis der zweite Term auf der rechten Seite von (1.60) den gleichen Wert einnimmt wie der erste Term. Dann ist ein Gleichgewicht erreicht und die mittlere Geschwindigkeit ändert sich nicht mehr, d.h. $d\bar{v}_n/dt = 0$. Damit folgt aus (1.60) für Elektronen und analog für Löcher

$$\bar{v}_n = -\frac{q_0 \tau_{c,n}}{m_n^*} E \quad \bar{v}_p = \frac{q_0 \tau_{c,p}}{m_p^*} E \quad (1.61)$$

Ferner gelten für die eindimensionalen Driftstromdichten¹⁷ nach dem lokalen Ohmschen Gesetz

$$\begin{aligned} j_{drift,n} &= -q_0 n \bar{v}_n = q_0 n \mu_n E = \sigma_n E \\ j_{drift,p} &= q_0 p \bar{v}_p = q_0 p \mu_p E = \sigma_p E \end{aligned} \quad (1.62)$$

Damit ergibt sich für die mittleren Driftgeschwindigkeiten

$$\bar{v}_n = -\mu_n E \quad \text{und} \quad \bar{v}_p = \mu_p E \quad (1.63)$$

Ein Vergleich von (1.61) mit (1.63) ergibt für die Beweglichkeiten

$$\mu_n = \frac{q_0}{m_n^*} \tau_{c,n} \quad \text{und} \quad \mu_p = \frac{q_0}{m_p^*} \tau_{c,p} \quad (1.64)$$

Demnach sind die Beweglichkeiten μ_p und μ_n ein Maß dafür, wie ungestört Ladungsträger durch ein Halbleiterstück driften können und damit auch eine wichtige Größe zur Bestimmung der spezifischen Leitfähigkeit σ . An dieser Stelle sei angemerkt, daß (1.64) nur zur Ableitung des Begriffs Beweglichkeit verwendet wird, aber keinesfalls zur quantitativen Bestimmung der Gesamtbeweglichkeiten μ_p und μ_n genutzt werden kann. Dazu sind, wie in den nachfolgenden Abschnitten beschrieben, jeweils Einzelbeweglichkeiten für die drei o.g. Streuprozesse aufzustellen und geeignet zu kombinieren.

Weiterhin implizieren (1.63) und (1.64) Unabhängigkeit vom elektrischen Feld für μ_p und μ_n . Diese Annahme ist bei Silizium aber nur für Feldstärken bis ungefähr 10^3V/cm gerechtfertigt [49]. Man beobachtet bei darüberliegenden Feldstärken, daß die Geschwindigkeiten nur noch unterlinear mit E anwachsen, bis schließlich ab ca. $E = 2 \cdot 10^4 \text{V/cm}$ für Elektronen und $E = 2 \cdot 10^4 \text{V/cm}$ für Löcher, eine obere Grenze, die sogenannte Sättigungsgeschwindigkeit $v_{sat,n,p}$, die bei ca. 10^7cm/s liegt, erreicht wird. Die mikroskopischen Ursachen für diesen Effekt sollen an dieser Stelle nicht diskutiert werden (s. dazu [68], [64]). Mathematisch kann die Feldabhängigkeit mit Hilfe einer Formel von Caughey–Thomas [14] beschrieben werden:

$$\mu_{n,p}(E) = \frac{\mu_{low,n,p}}{\left[1 + \left(\frac{\mu_{low,n,p} E}{v_{sat,n,p}} \right)^{\beta_{n,p}} \right]^{1/\beta_{n,p}}} \quad (1.65)$$

¹⁷Das sind die Driftterme aus (1.31) und (1.32)

Hierbei entspricht $\mu_{low,n,p}$ den in (1.64) beschriebenen feldunabhängigen Gesamtbeweglichkeiten. In [30] werden für die Parameter $v_{sat,n,p}$ und $\beta_{n,p}$ Werte von Canali [13] vorgeschlagen. Diese lauten für $T = 300K$:

$$\begin{aligned} v_{sat,n} &= 1,07 \cdot 10^7 \text{ cm/s} & \text{und} & & \beta_n &= 1,109 \\ v_{sat,p} &= 8,37 \cdot 10^6 \text{ cm/s} & \text{und} & & \beta_p &= 1,213 \end{aligned}$$

Setzt man (1.65) in (1.63) ein, erhält man für die Driftgeschwindigkeiten $v_{n,p}(E)$ Kurven, wie sie im Bild 1.1 aufgetragen sind. Dabei wurden für $\mu_{low,n,p}$ die Werte aus [20] verwendet, d.h. $\mu_{low,n} = 1430 \text{ cm}^2/(Vs)$ und $\mu_{low,p} = 495 \text{ cm}^2/(Vs)$ (s. S. 31).

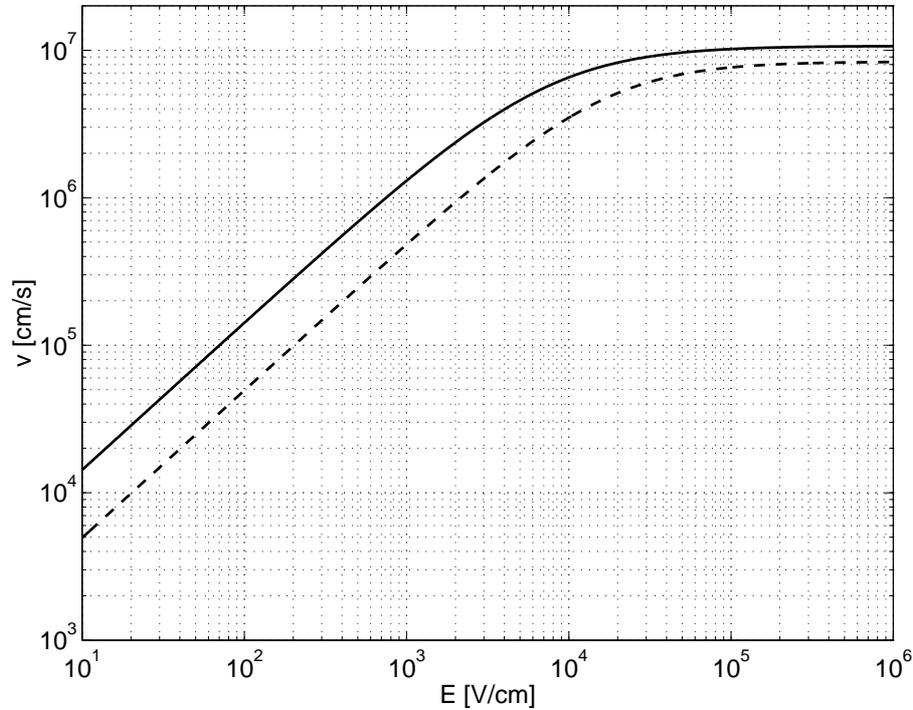


Bild 1.1: Driftgeschwindigkeiten v_n (durchgezogene Linie) und v_p (gestrichelte Linie) als Funktion der elektrischen Feldstärke für $T = 300K$

Die Sättigungsgeschwindigkeiten $v_{sat,p,n}$ sind temperaturabhängig, wofür Canali [13] folgende Ausdrücke nennt:

$$v_{sat,n} = v_{sat,n}(300K) \cdot \left(\frac{T}{300}\right)^{-0,87} \quad (1.66)$$

$$v_{sat,p} = v_{sat,p}(300K) \cdot \left(\frac{T}{300}\right)^{-0,52} \quad (1.67)$$

1.5.3 Streuung am Gitter und an Verunreinigungen

Gitterstreuung

Da die Energie der Phononen hauptsächlich von der Gittertemperatur abhängt (s. Abschnitt 1.5.1), ist die Gittertemperatur auch die Haupteinflußgröße zur Bestimmung der Teilbeweglichkeiten $\mu_{n,p}^L$ für die Gitterstreuung (oberer Index L für Lattice = Gitter) und zwar derart, daß mit steigender Temperatur die Beweglichkeiten abnehmen. Die Ableitung physikalisch korrekter Ausdrücke für die Phononenstreuung direkt aus der Halbleitertheorie ist schwierig und liefert quantitativ ungenaue Ergebnisse. Deshalb empfiehlt es sich, einmal mehr empirische Ausdrücke, die anhand von Meßdaten gefittet werden, zu Modellierungszwecken zu verwenden (vgl. [64]). In der Simulationspraxis benützt man hierfür häufig das einfache Potenzgesetz (z.B. auch in [30])

$$\mu_n^L = \mu_n^0 \left(\frac{T}{300K} \right)^{\alpha_n} \quad (1.68)$$

$$\mu_p^L = \mu_p^0 \left(\frac{T}{300K} \right)^{\alpha_p} \quad (1.69)$$

Es gibt eine große Anzahl von Arbeiten bezüglich der Fitparameter $\mu_{n,p}^L$ und $\alpha_{n,p}$ zur Modellierung der Teilbeweglichkeiten¹⁸. In den vorliegenden Bauelementemodellen wurden die Parameter aus [20] verwendet:

$$\begin{aligned} \mu_n^0 &= 1430 \text{cm}^2 / (\text{Vs}) & \text{und} & \quad \alpha_n = -2 \\ \mu_p^0 &= 495 \text{cm}^2 / (\text{Vs}) & \text{und} & \quad \alpha_p = -2,18 \end{aligned}$$

Störstellenstreuung

Zur mathematischen Beschreibung der Teilbeweglichkeiten $\mu_{n,p}^I$ für Störstellenstreuung (oberer Index I für Ionized Impurities = ionisierte Störstellen), wurde erstmals von Conwell und Weisskopf [16] ein brauchbares Modell anhand theoretischer Überlegungen, inklusive Temperaturabhängigkeit, hergeleitet. Ihre Formel lautet:

$$\mu_{n,p}^I = \frac{64 \sqrt{\pi} \epsilon_0^2 \epsilon_r^2 (2 k T)^{3/2}}{N_{A,D} q_0^3 m_{n,p}^{*1/2}} \left(\ln \left[1 + \left(\frac{12 \pi \epsilon_0 \epsilon_r k T}{q_0^2 N_{A,D}^{1/3}} \right)^2 \right] \right)^{-1} \quad (1.70)$$

Die gegenseitige Abschirmung benachbarter ionisierter Dotierstellen bewirkt eine Schwächung dieses Streueffekts, was sich dann in einer höheren Beweglichkeit niederschlägt. In (1.70) wird dies mit Hilfe des logarithmischen Terms modelliert.

¹⁸Eine Auflistung findet sich in [64]

Eigentlich müßten außerdem zusätzliche Abschirmungseffekte, die von den Ladungen der freien Elektronen und Löcher ausgehen, ebenfalls berücksichtigt werden. Deshalb wurde (1.70) von Brooks [11] dahingehend modifiziert (mehr dazu in [64]).

Kombinationen von Gitter- und Störstellenstreuung

Läßt man die Elektronen–Löcher–Streuung zunächst außer acht, kann man sich überlegen, in welcher Weise $\mu_{n,p}^L$, z.B. gemäß (1.68), bzw. (1.69) und $\mu_{n,p}^I$, z.B. gemäß (1.70), miteinander kombiniert werden müssen, damit man die in (1.31) und (1.32) wirkenden effektiven Gesamtbeweglichkeiten $\mu_{n,p}$ erhält. Da die Elektronen–Löcher–Streuung noch außen vor bleibt, seien $\mu_{n,p}$ im folgenden $\mu_{n,p}^{LI}$ genannt. Setzt man voraus, daß sich Phononenstreuung und Störstellenstreuung nicht gegenseitig beeinflussen, kann man die sogenannte Matthiesen–Regel

$$\frac{1}{\mu_{n,p}^{LI}} = \frac{1}{\mu_{n,p}^L} + \frac{1}{\mu_{n,p}^I} \quad (1.71)$$

anwenden. Da aber die beiden Effekte nicht völlig unabhängig voneinander sind, ist (1.71) für diesen Zweck ungeeignet. Anhand theoretischer Überlegungen haben Debye und Conwell [19] eine andere Beziehung für $\mu_{n,p}^{LI}$ abgeleitet¹⁹, die

$$\mu_{n,p}^{LI} = \mu_{n,p}^L \cdot (1 + X^2 \cdot [\text{Ci}(X) \cos(X) + (\text{Si}(X) - \pi/2) \sin(X)]) \quad (1.72)$$

lautet, wobei X durch

$$X = \left(\sqrt{\frac{6 \mu_{n,p}^L}{\mu_{n,p}^I}} \right)$$

gegeben ist. Abweichend von (1.71) und (1.72) verwendet man zum Zwecke der Simulation häufig einfacher zu handhabende semi–empirische Ausdrücke für $\mu_{n,p}^{LI}$, deren Parameter an Meßdaten angepaßt werden. In den Bauelementemodellen von Vogler [73] wird durchgängig auf die von Caughey und Thomas aufgestellte Gleichung [14]

$$\mu_{n,p}^{LI} = \mu_{n,p}^{min} + \frac{\mu_{n,p}^L - \mu_{n,p}^{min}}{1 + \left(\frac{N_{A,D}}{C_{n,p}^{ref}} \right)^{\beta_{n,p}}} \quad (1.73)$$

zurückgegriffen, wobei sich die temperaturabhängigen Parameter $\mu_{n,p}^{min}$, $C_{n,p}^{ref}$ und $\beta_{n,p}$ zu

¹⁹Ci(X) ist die Integralcosinus- und Si(X) die Integralsinusfunktion. Zur Definition s. [10]

$$\mu_n^{min} = \begin{cases} 80 \frac{cm^2}{Vs} \left(\frac{T}{300K} \right)^{-0,45} & : T > 200K \\ 80 \frac{cm^2}{Vs} \left(\frac{200}{300} \right)^{-0,45} \left(\frac{T}{200K} \right)^{-0,15} & : T \leq 200K \end{cases} \quad (1.74)$$

$$\mu_p^{min} = \begin{cases} 45 \frac{cm^2}{Vs} \left(\frac{T}{300K} \right)^{-0,45} & : T > 200K \\ 45 \frac{cm^2}{Vs} \left(\frac{200}{300} \right)^{-0,45} \left(\frac{T}{200K} \right)^{-0,15} & : T \leq 200K \end{cases} \quad (1.75)$$

$$C_n^{ref} = 1,12 \cdot 10^{17} cm^{-3} \left(\frac{T}{300K} \right)^{3,2} \quad (1.76)$$

$$C_p^{ref} = 2,23 \cdot 10^{17} cm^{-3} \left(\frac{T}{300K} \right)^{3,2} \quad (1.77)$$

$$\beta_{n,p} = 0,72 \left(\frac{T}{300K} \right)^{0,065} \quad (1.78)$$

bestimmen. Diese Formel liefert genügend genaue Ergebnisse und beinhaltet außerdem den Effekt des Ausfrierens von Ladungsträgern (s. Abschnitt 1.2.2 auf Seite 11ff). Für $\mu_{n,p}^L$ setzt man die Beweglichkeit, die man aus (1.68), bzw. (1.69) erhält, ein. Ganz auf die Verwendung von (1.68), bzw. (1.69) verzichtet das Beweglichkeitsmodell von Arora [4], das speziell für den Temperaturbereich 250–500K angepaßt wurde, aber auch für tiefere Temperaturen gute Ergebnisse liefert. Die zugehörigen empirischen Formeln dazu lauten

$$\mu_n^{LI} = 88 \left(\frac{T}{300K} \right)^{-0,57} + \frac{7,4 \cdot 10^8 (T/K)^{-2,33}}{1 + \left(\frac{N}{1,26 \cdot 10^{17} (T/300K)^{2,4}} \right)^{0,88} \left(\frac{T}{300K} \right)^{-0,146}} \quad (1.79)$$

und

$$\mu_p^{LI} = 54,3 \left(\frac{T}{300K} \right)^{-0,57} + \frac{1,36 \cdot 10^8 (T/K)^{-2,23}}{1 + \left(\frac{N}{2,35 \cdot 10^{17} (T/300K)^{2,4}} \right)^{0,88} \left(\frac{T}{300K} \right)^{-0,146}} \quad (1.80)$$

Dieses Beweglichkeitsmodell liefert zusammen mit der Elektronen–Löcher–Beschreibung nach Mnatsakanov (s. Abschnitt 1.5.5) genaue Resultate, so daß es aktuell

für die Modellierung der schwach dotierten Driftzone verwendet wird, wohingegen für die stark dotierten Randbereiche, weiterhin die Beweglichkeiten gemäß [14], d.h. (1.73) bis (1.78), implementiert sind.

1.5.4 Elektronen–Löcher–Streuung I: Herkömmliche Theorie

Wie bereits in Abschnitt 1.5.1 unter (3) ausgeführt, können Elektronen und Löcher, die bei Anlegen eines elektrischen Feldes gegeneinander fließen, einander streuen. Dieser Effekt spielt besonders dann eine Rolle, wenn besonders viele freibewegliche Ladungsträger vorhanden sind, also bei Hochinjektion ($p \approx n \gg N_{D,A}$). Damit ist er auch für die Modellierung bipolarer Leistungsbaulemente interessant. Außerdem wird der Effekt umso dominanter, je geringer die Temperatur und damit der Einfluß der Phononenstreuung ist.

Zu diesem Effekt, für den in der Literatur oft das Kürzel CCS (engl. carrier–carrier–scattering) verwendet wird, gibt es wieder mehrere, meist empirische Modelle (s. dazu z.B. [5], [30], [64]). Sie alle werden hier unter der Kategorie „Herkömmliche Theorie“ subsummiert und haben folgende Charakteristika:

- a) Es wird analog zu den beiden anderen Streueffekten, eine Teilbeweglichkeit μ^C definiert. Diese Beweglichkeit ist eine Funktion von den Ladungsträgerdichten n , p und der Temperatur T und zwar so, daß μ^C mit steigendem p und fallender Temperatur T , abnimmt.
- b) Die Teilbeweglichkeit μ^C wird mittels der Debye–Conwell–Formel (1.72) oder der Matthiesenregel (1.71), mit den anderen beiden Teilbeweglichkeiten $\mu_{n,p}^L$ und $\mu_{n,p}^I$, oder den kombinierten Beweglichkeiten $\mu_{n,p}^{LI}$, verknüpft. Somit führt μ^C zu einer Verkleinerung der Gesamtbeweglichkeit $\mu_{n,p}$.

Im Gegensatz zu den beiden anderen Teilbeweglichkeiten, hängt μ^C und damit auch die errechneten Gesamtbeweglichkeiten $\mu_{n,p}$ von der injizierten Ladungsträgerdichten p bzw. n ab. Dies hat für die Modellierung, insbesondere der schwach dotierten Driftzone mehrere Konsequenzen (vgl. Kapitel 2). So sind die Diffusionskonstanten gemäß den Einsteinrelationen (1.33) keine wirklichen Konstanten mehr, sondern abhängig von der Ladungsträgerdichte am jeweiligen Ort. Somit gilt also:

$$D_n(x) = \frac{k_B T}{q_0} \mu_n(p(x), n(x)) \quad \text{und} \quad D_p(x) = \frac{k_B T}{q_0} \mu_p(p(x), n(x)) \quad (1.81)$$

Dies führt letztlich zu einer wesentlich komplexeren mathematischen Beschreibung des Ladungsträgerverhaltens in der Driftzone (vgl. Abschnitt 2.3.1). In den Bauelementemodellen von Vogler [73] wird die häufig verwendete CCS–Beschreibung von

Dorkel und Leturcq [20] verwendet. Der zugehörige Ausdruck für μ^C , angepaßt an Messungen von [17] und [34], lautet:

$$\mu^C = \frac{1,04 \cdot 10^{21} \text{ cm}/(Vs) \cdot \left(\frac{T}{300K}\right)^{3/2}}{\sqrt{np} \cdot \ln \left(1 + 7,45 \cdot 10^{13} \text{ cm}^{-2} \cdot \left(\frac{T}{300K}\right)^2 \cdot (np)^{-1/3}\right)} \quad (1.82)$$

Wie oben in b) beschrieben, wird in den Bauelementemodellen [73] mit Hilfe der Matthiesen-Regel, μ^C mit $\mu_{n,p}^{LI}$ gemäß (1.73) zur Gesamtbeweglichkeit $\mu_{n,p}$ verrechnet.

In Abschnitt 1.5.1 unter (3) wurde bereits angedeutet, daß aufgrund der Elektronen-Löcher-Streuung, Ladungsträger ihre Flußrichtung umkehren können und sich damit entgegen der Feldrichtung bewegen (Drag-Effekt). Dies kann unter Umständen sogar dazu führen, daß die Gesamtbeweglichkeit μ_n bzw. μ_p negativ wird (vollständiger Drag-Effekt). Für den direkten Halbleiter Galliumarsenid wurde der vollständige Drag-Effekt bereits von Höpfel [29] experimentell nachgewiesen. Für den indirekten Halbleiter Silizium wurden bislang noch keine experimentellen Ergebnisse in dieser Hinsicht veröffentlicht, aber einer Abschätzung auf Basis theoretischer Überlegungen von Gresserov [27] zufolge, sollte der Drag-Effekt 40–50 mal stärker sein als bei Galliumarsenid. Setzt man nun (1.82) in (1.72) oder (1.71) ein, so erkennt man, daß die jeweilige Gesamtbeweglichkeit μ_n und μ_p zwar verringert wird (wie in b) behauptet), aber niemals negative Werte annehmen kann. Dieses wichtige Ergebnis läßt somit starke Zweifel an einer korrekten Beschreibung der Elektronen-Löcher-Streuung, mit Hilfe der herkömmlichen CCS-Modelle, aufkommen. Letztlich kann, wie in [57] gezeigt wird, die Elektronen-Löcher-Streuung mit den herkömmlichen Ansätzen nicht korrekt modelliert werden und führt sogar zu widersinnigen Simulationsergebnissen (vgl. auch [58]).

1.5.5 Elektronen-Löcher-Streuung II: Theorie nach Mnatsakanov

Mnatsakanov [46] wählt einen völlig anderen Ansatz als der in Abschnitt 1.5.4 Beschriebene. Zusätzlich zum eigentlichen Drift- und Diffusionsterm, werden in den Stromdichtegleichungen jeweils zusätzliche, den Drag-Effekt beschreibende Terme, eingeführt. Die Transportgleichungen können somit zu

$$j_p = q_0 \mu_p^p p E - q_0 D_p^p \frac{\partial p}{\partial x} - q_0 \mu_p^n n E - q_0 D_p^n \frac{\partial n}{\partial x} \quad (1.83)$$

$$j_n = \underbrace{q_0 \mu_n^n n E + q_0 D_n^n \frac{\partial n}{\partial x}}_{\text{Drift} + \text{Diffusion}} - \underbrace{q_0 \mu_n^p p E + q_0 D_n^p \frac{\partial p}{\partial x}}_{\text{Drag}} \quad (1.84)$$

geschrieben werden. Durch den Drag-Effekt wird also eine Teilchenstromdichte durch einen entgegengesetzt laufenden Teilchenstrom der anderen Ladungsträgerart geschwächt. Ein Vergleich von (1.83) mit (1.31) und (1.84) mit (1.32) ergibt für die Gesamtbeweglichkeiten

$$\mu_p = \mu_p^p - \frac{n}{p} \mu_p^n \quad \text{und} \quad \mu_n = \mu_n^n - \frac{p}{n} \mu_n^p \quad (1.85)$$

und für die effektiv wirkenden Diffusionskoeffizienten²⁰

$$D_p = D_p^p + D_p^n \quad \text{und} \quad D_n = D_n^n + D_n^p \quad (1.86)$$

An Stelle von μ^C zur Beschreibung der Wechselwirkung von Elektronen und Löchern, führt Mnatsakanov die Beweglichkeiten μ_{pn} zur Beschreibung der Streuung von Löchern an Elektronen und μ_{np} zur Beschreibung der Streuung von Elektronen an Löchern, ein. Mit Hilfe einer erneuten Anpassung an die Messungen von [17] und [34], ergibt sich

$$\mu_{pn} = \frac{1840 \text{ cm}/(Vs)}{n/3, 2 \cdot 10^{17} \text{ cm}^{-3} + \frac{n/4, 6 \cdot 10^{16} \text{ cm}^{-3}}{1 + (n+p)/7 \cdot 10^{16} \text{ cm}^{-3}}} \left(\frac{T}{300K} \right)^{1.5} \quad (1.87)$$

$$\mu_{np} = \frac{1840 \text{ cm}/(Vs)}{p/3, 2 \cdot 10^{17} \text{ cm}^{-3} + \frac{p/4, 6 \cdot 10^{16} \text{ cm}^{-3}}{1 + (n+p)/7 \cdot 10^{16} \text{ cm}^{-3}}} \left(\frac{T}{300K} \right)^{1.5} \quad (1.88)$$

Diese beiden Beweglichkeiten haben natürlich wieder die Eigenschaft, daß sie umso kleiner, d.h. dominanter sind, je tiefer die Temperaturen und je höher die Trägerdichten sind. Für ihr Verhältnis ergibt sich offensichtlich:

$$\frac{\mu_{pn}}{\mu_{np}} = \frac{p}{n} \quad (1.89)$$

Mit μ_{pn} , μ_{np} und $\mu_{n,p}^{LI}$ (z.B. gemäß Arora [4]; s. Abschnitt 1.5.3), lassen sich die in (1.83) und (1.84) eingeführten Beweglichkeiten darstellen als (s. [46], [47])

$$\begin{aligned} \mu_p^p &= \mu_p^{LI} \frac{\mu_{pn}(\mu_{np} + \mu_n^{LI})}{\Delta} \quad \text{und} \quad \mu_p^n = \frac{\mu_{pn} \mu_n^{LI} \mu_p^{LI}}{\Delta} \\ \mu_n^n &= \mu_n^{LI} \frac{\mu_{np}(\mu_{pn} + \mu_p^{LI})}{\Delta} \quad \text{und} \quad \mu_n^p = \frac{\mu_{np} \mu_n^{LI} \mu_p^{LI}}{\Delta} \end{aligned} \quad (1.90)$$

$$\text{mit} \quad \Delta = \mu_{pn} \mu_{np} + \mu_{pn} \mu_n^{LI} + \mu_{np} \mu_p^{LI}$$

²⁰Für $\partial p/\partial x = \partial n/\partial x$; dies ist zumindest näherungsweise immer erfüllt

Für die in (1.83) und (1.84) eingeführten Diffusionskoeffizienten ergibt sich dann gemäß den Einstein-Relationen (1.33)

$$D_p^p = \frac{k_B T}{q_0} \mu_p^p, \quad D_p^n = \frac{k_B T}{q_0} \mu_p^n, \quad D_n^n = \frac{k_B T}{q_0} \mu_n^n, \quad D_n^p = \frac{k_B T}{q_0} \mu_n^p \quad (1.91)$$

Setzt man nun nacheinander (1.90) in (1.91) und weiter in (1.85) und (1.86) ein, erkennt man sofort, daß die Einstein-Relationen (1.33) zur Berechnung der effektiv wirksamen Diffusionskoeffizienten $D_{n,p}$ aus den Gesamtbeweglichkeiten $\mu_{n,p}$ nicht mehr gelten! Im folgenden werden wieder die Spezialfälle schwache und starke Injektion betrachtet. Dabei wird sich der Knappheit der Darstellung wegen, einmal mehr auf n-Dotierung beschränkt.

Fall 1: schwache Injektion

Bei schwacher Injektion in n-dotierten Zonen gilt $p \ll N_D^+$. Für die Minoritätsträgerbeweglichkeit gilt dann nach (1.85)

$$\mu_p = \mu_p^p - \frac{N_D^+}{p} \mu_p^n \quad (1.92)$$

Der Faktor N_D^+/p wird für hohe Dotierungen sehr groß. Gleichzeitig sind die Beweglichkeiten μ_p^p und μ_p^n wegen der Störstellenstreuung schon recht klein, so daß μ_p negativ werden kann (vollständiger Drag-Effekt). Berechnungen mit Hilfe der in diesem Abschnitt vorgestellten Gleichungen und dem Beweglichkeitsmodell gemäß [4], ergeben für eine Dotierung von $N_D = 5 \cdot 10^{18} \text{ cm}^{-3}$ und Temperaturen unter 300K, negative Werte für μ_p (mehr dazu z.B. bei [58]). Dieses Ergebnis läßt sich auch gut veranschaulichen, indem man sich vorstellt, daß ein Strom Minoritätsträger geringer Dichte quasi vom Majoritätsträgerstrom hoher Dichte einfach mitgerissen wird und deshalb in die gleiche Richtung, d.h. elektrisch gesehen in die Gegenrichtung, fließt.

In Abschnitt 1.5.4 wurde behauptet, daß zum Zwecke der Modellierung, Elektronen-Löcher-Streuung vor allem für Hochinjektion in Verbindung mit niedrig dotierten Zonen, relevant ist. Diese Aussage scheint dem eben erhaltenen Ergebnis zu widersprechen. Dabei ist aber zu bedenken, daß für die Relevanz bezüglich der Modellierung, weniger der Wert einer Beweglichkeit selbst, sondern der Beitrag dieser Beweglichkeit zu einer interessierenden elektrischen Größe, wie z.B. einem Spannungsabfall, wichtig ist. Bei einer stark dotierten Zone, trägt die im Vergleich zur Majoritätsträgerdichte viel geringere Minoritätsdichte, nur unwesentlich zum elektrischen Widerstand dieser Zone bei. Bei einer niedrig dotierten Zone dagegen, wird der elektrische Widerstand hauptsächlich von den injizierten Ladungsträgern bestimmt, so daß Elektronen-Löcher-Streuung einen starken Einfluß auf den Spannungsabfall haben kann, obwohl die Gesamtbeweglichkeiten dadurch nur reduziert, nicht aber negativ werden. Auch wenn das Phänomen „negative Beweglichkeit“ bei schwacher Injektion nicht in den Bauelementemodellen explizit berücksichtigt wird,

so dient die Beleuchtung dieser Thematik hier als Indiz für die prinzipielle Richtigkeit des Ansatzes von Mnatsakanov.

Fall 1: starke Injektion

Bei starker Injektion in n-dotierten Zonen gilt $p \approx n \gg N_D^+$, d.h. $p/n = 1$. Die Gleichung (1.85) läßt sich dann vereinfachen zu

$$\mu_p = \mu_p^p - \mu_p^n \quad \text{und} \quad \mu_n = \mu_n^n - \mu_n^p. \quad (1.93)$$

Für die Beweglichkeiten, die die Wechselwirkung der Ladungsträger beschreiben, gilt demzufolge (s. (1.89))

$$\mu_{pn} = \mu_{np}. \quad (1.94)$$

Die Gleichung (1.90) lautet nun unter Beachtung von (1.94)

$$\begin{aligned} \mu_p^p &= \mu_p^{LI} \frac{\mu_{pn} + \mu_n^{LI}}{\Delta} \quad \text{und} \quad \mu_p^n = \frac{\mu_n^{LI} \mu_p^{LI}}{\Delta} \\ \mu_n^n &= \mu_n^{LI} \frac{\mu_{pn} + \mu_p^{LI}}{\Delta} \quad \text{und} \quad \mu_n^p = \frac{\mu_n^{LI} \mu_p^{LI}}{\Delta} \end{aligned} \quad (1.95)$$

$$\text{mit} \quad \Delta = \mu_{pn} + \mu_n^{LI} + \mu_p^{LI}$$

Mit diesem Ergebnis folgt für die Gesamtbeweglichkeiten aus (1.93)

$$\mu_p = \mu_p^{LI} \cdot \frac{\mu_{pn}}{\Delta} \quad \text{und} \quad \mu_n = \mu_n^{LI} \cdot \frac{\mu_{pn}}{\Delta} \quad (1.96)$$

und für die effektiv wirkenden Diffusionsgrößen aus (1.91)

$$D_p = D_p^p \frac{\mu_{pn} + 2\mu_n^{LI}}{\Delta} \quad \text{und} \quad D_n = D_n^n \frac{\mu_{pn} + 2\mu_p^{LI}}{\Delta} \quad (1.97)$$

Man erkennt aus (1.96), daß die Gesamtbeweglichkeiten nicht negativ, aber durch eine kleiner werdende Beweglichkeit μ_{pn} , ebenfalls kleiner wird. Auf die Ergebnisse (1.94) bis (1.97) wird im nächsten Kapitel zurückzukommen sein.

Weitere Überlegungen bezüglich der EHS-Theorie von Mnatsakanov und ein Vergleich der quantitativen Ergebnisse für Beweglichkeits- und Diffusionsgrößen mit denen der CCS-Theorie gemäß Dorkel und Leturcq [20] (s. Abschnitt 1.5.4) finden sich in der Dissertation von Schlögl [58].

1.6 Abrupte Halbleiterübergänge

1.6.1 Der pn -Übergang

In diesem und in den Abschnitten 1.6.2 und 1.6.3 werden die zur Modellierung wichtigen Formeln zu Halbleiterübergängen, besonders zur Beschreibung der Ladungsträgerinjektion, zusammengefaßt. Detaillierte Ableitungen hierzu finden sich in [49], [61] und [68]. An dieser Stelle und in den Modellen werden überwiegend abrupte Halbleiterübergänge betrachtet. Dies reicht für die Modellbildung meist aus, da sich auch nicht-abrupte Übergänge, zumindest näherungsweise als abrupte Übergänge mit einer durchschnittlichen Dotierdichte, darstellen lassen.

Zunächst sei ein pn -Übergang im thermodynamischen Gleichgewicht betrachtet, wobei die (hohen) Dotierdichten des n -Gebiets und des p -Gebiets etwa in der gleichen Größenordnung (etwa 10^{18} cm^{-3}) sind. Die Löcher im p -Gebiet nahe am Übergang, „sehen“ einen Mangel an Löchern im n -Gebiet, so daß Löcher in das n -Gebiet diffundieren. Auf diese Weise diffundieren zugleich Elektronen vom n -Gebiet in das p -Gebiet. Da die beiden Gebiete ursprünglich elektrisch neutral waren, bildet sich nun aufgrund dieser Diffusion von Ladungsträgern, eine Raumladungszone aus. Dadurch entsteht eine Übergangsspannung, die eine Ladungsträgerdrift verursacht, die der Ladungsträgerdiffusion entgegenwirkt. Im thermodynamischen Gleichgewicht gleichen sich Drift und Diffusion aus, so daß der resultierende Gesamtstrom sinnigerweise Null ist. Somit kann die am pn -Übergang anliegende Spannung, die sogenannte Diffusionsspannung U_D , berechnet werden, indem man die Stromdichtegleichungen (1.31) und (1.32) einfach zu Null setzt. Löst man dann beide Gleichungen jeweils nach der elektrischen Feldstärke auf und ersetzt die Diffusionskonstanten durch die Einsteinrelationen (1.33)²¹, erhält man für die elektrische Feldstärke:

$$E(x) = \frac{k_B T}{q_0} \frac{1}{p} \frac{dp}{dx} = -\frac{k_B T}{q_0} \frac{1}{n} \frac{dn}{dx} \quad (1.98)$$

Bekanntlich gilt weiterhin $E(x) = -dV(x)/dx$, so daß das ortsabhängige Potential $V(x)$ mittels Integration von $E(x)$ berechnet werden kann:

$$V(x) = -\frac{k_B T}{q_0} \int_{n_i}^{p(x)} \frac{dp}{p} = -\frac{k_B T}{q_0} \ln \left(\frac{p(x)}{n_i} \right) \quad (1.99)$$

$$V(x) = \frac{k_B T}{q_0} \int_{n_i}^{n(x)} \frac{dn}{n} = \frac{k_B T}{q_0} \ln \left(\frac{n(x)}{n_i} \right) \quad (1.100)$$

²¹Vorausgesetzt man kann die Elektronen-Löcher-Streuung vernachlässigen (vgl. dazu Abschnitt 1.5.5). Dies ist in den Bauelementemodellen für Zonen hoher Dotierdichten der Fall.

Löst man (1.99) und (1.100) nach den Ladungsträgerdichten auf, ergeben sich wieder die Boltzmann-Gleichungen:

$$p(x) = n_i \exp\left(-\frac{q_0 V(x)}{k_B T}\right) \quad (1.101)$$

$$n(x) = n_i \exp\left(\frac{q_0 V(x)}{k_B T}\right) \quad (1.102)$$

Betrachtet man (1.101) und (1.102), so erkennt man das im thermischen Gleichgewicht natürlich weiterhin (1.10), also $p(x)n(x) = n_i^2$ gilt. In genügend großer Entfernung vom Übergang gilt bekanntlich $p \approx N_A^-$ im p -Gebiet und $n \approx N_D^+$ im n -Gebiet. Dabei gilt für das jeweilige Potential $V(x) = V_p$ und $V(x) = V_n$. Die Differenz dieser Potentiale ergibt die Diffusionsspannung. Setzt man also V_p und N_A^- in (1.99), bzw. V_n und N_D^+ in (1.99) ein und bildet die Differenz erhält man:

$$U_D = \frac{k_B T}{q_0} \ln\left(\frac{N_A^- N_D^+}{n_i^2}\right) \quad (1.103)$$

Mittels der Schottkyschen Parabelnäherung²² erhält man für die Weite der Raumladungszone w_{RLZ} :

$$w_{RLZ} = \sqrt{\frac{2 \epsilon_0 \epsilon_r}{q_0} U_D \left(\frac{1}{N_A^-} + \frac{1}{N_D^+}\right)} \quad (1.104)$$

Legt man nun eine Sperrspannung U_R an den pn -Übergang, verbreitert sich diese Raumladungszone. Ihre Weite berechnet sich dann einfach zu

$$w_{RLZ} = \sqrt{\frac{2 \epsilon_0 \epsilon_r}{q_0} (U_D - U_R) \left(\frac{1}{N_A^-} + \frac{1}{N_D^+}\right)} \quad (1.105)$$

Damit kann nun die, für das dynamische Verhalten von Bauelementen wichtige Sperrschichtkapazität C_j , berechnet werden, indem man (1.105) in

$$C_j = \epsilon_0 \epsilon_r \frac{A}{w_{RLZ}} \quad (1.106)$$

einsetzt. Außerdem fließt im Sperrzustand ein geringer Strom, der sogenannte Sperr sättigungsstrom (mehr dazu in Kapitel 2). Wird der pn -Übergang in Durchlaßrichtung betrieben, verringert sich die Raumladungszonenweite und es werden Löcher (Elektronen) vom p -Gebiet (n -Gebiet) in das n -Gebiet (p -Gebiet) injiziert,

²²Bei abrupten Übergängen sind die Ladungen innerhalb der Raumladungszone jeweils auf beiden Seiten in guter Näherung konstant. Für das elektrische Feld ergibt sich durch Integration eine lineare Funktion. Das Integral über den linearen E -Feld-Verlauf ergibt dann einen parabelförmigen Potentialverlauf (s. u.a. [61])

die Minoritätsträgerdichte dort also angehoben. Der pn -Übergang wird leitend und es kann Strom fließen. Die Frage ist nun, wie die auftretende Ladungsträgerinjektion quantitativ beschrieben werden kann. Bild 1.2 oben zeigt einen pn -Übergang im thermodynamischen Gleichgewicht. Die Ladungsträgerstatistik wird durch (1.101) und (1.102), die letztlich den Gleichungen (1.6) und (1.7) entsprechen, beschrieben.

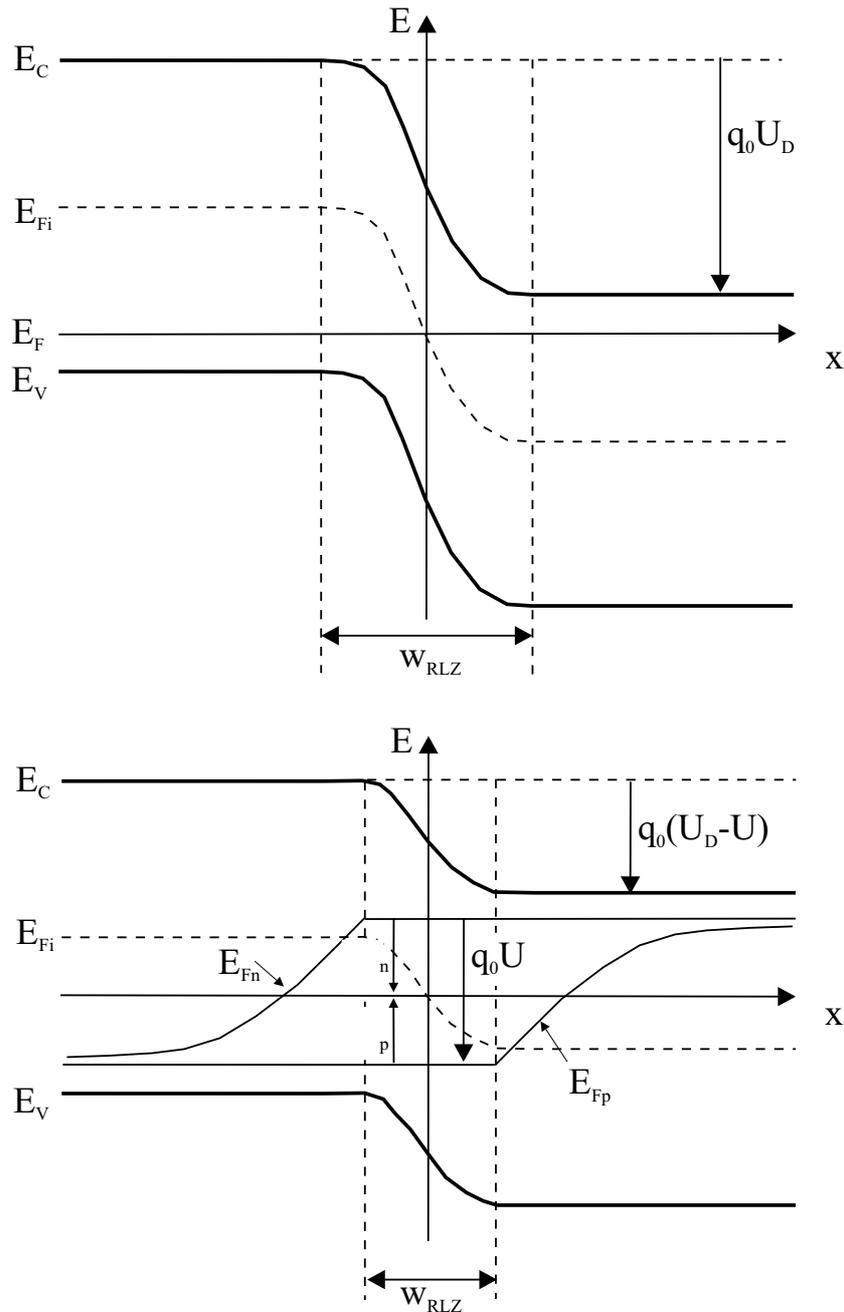


Bild 1.2: Bänderdiagramme für einen pn -Übergang im thermodynamischen Gleichgewicht (oben) und für Flußpolung (unten)

Bild 1.2 unten zeigt hingegen den pn -Übergang bei Durchlaßpolung. Das Fermi-niveau ist nun nicht mehr überall gleich, sondern spaltet sich auf in die sogenannten Quasi-Fermi-niveaus E_{Fp} (E_{Fn}), welche im p -Gebiet (n -Gebiet), genügend weit entfernt von der Raumladungszone, wieder konstant sind. Anstelle von (1.6) und (1.7) gilt nun

$$p = N_V \cdot \exp\left(-\frac{E_{Fp} - E_V}{k_B T}\right) \quad (1.107)$$

$$n = N_C \cdot \exp\left(-\frac{E_C - E_{Fn}}{k_B T}\right), \quad (1.108)$$

bzw. mit den Quasi-Fermipotentialen $\phi_p = -E_{Fp}/q_0$ und $\phi_n = -E_{Fn}/q_0$ und n_i gemäß (1.11)

$$p(x) = n_i \cdot \exp\left(\frac{q_0(\phi_p - V(x))}{k_B T}\right) \quad (1.109)$$

$$n(x) = n_i \cdot \exp\left(\frac{q_0(V(x) - \phi_n)}{k_B T}\right). \quad (1.110)$$

Für das Produkt aus $p(x)n(x)$ gilt mit $U = \phi_p - \phi_n$ an jeder Stelle x innerhalb und direkt am Rand der Raumladungszone:

$$p(x) \cdot n(x) = n_i^2 \exp\left(\frac{q_0 U}{k_B T}\right) \quad (1.111)$$

Für $U = 0$, d.h. für thermisches Gleichgewicht, erhält man wiederum (1.10). Für das p -Gebiet läßt sich nun die (injizierte) Minoritätsträgerdichte n_p am Rand der Raumladungszone gemäß (1.111) allgemein zu

$$n_p = \frac{n_i^2}{N_A^-} \exp\left(\frac{q_0 U}{k_B T}\right) \quad (1.112)$$

bestimmen. Wegen der Neutralitätsbedingung (1.20) paßt sich die Majoritätsträgerdichte p_p der Minoritätsträgerdichte n_p an, d.h.

$$p = n_p + N_A^- \quad (1.113)$$

Im Falle schwacher Injektion kann man in (1.113) n_p vernachlässigen und es gilt $p \approx N_A^-$. Für die Löcherdichte p_n am Rand der Raumladungszone an der Grenze zum (hochdotierten) n -Gebiet erhält man entsprechend:

$$p_n = \frac{n_i^2}{N_D^+} \exp\left(\frac{q_0 U}{k_B T}\right). \quad (1.114)$$

1.6.2 Der $p^+\nu$ -Übergang

Mit dem Begriff $p^+\nu$ -Übergang ist ein Übergang zwischen einem stark dotierten p -Gebiet ($N_A \geq 10^{18} \text{cm}^{-3}$) und einem schwach dotierten n -Gebiet ($N_\nu \leq 10^{14} \text{cm}^{-3}$), das im folgenden ν -Zone mit der Dotierung N_ν und der Weite w_ν genannt wird, gemeint. Dabei findet im p -Gebiet schwache und in die ν -Zone starke Injektion statt. Im Sperrfall vereinfacht sich (1.105) wegen $N_\nu \ll N_A^-$ zu²³

$$w_{RLZ} = \sqrt{\frac{2 \epsilon_0 \epsilon_r}{q_0} \frac{U_D - U_R}{N_\nu}}. \quad (1.115)$$

Der Zusammenhang zwischen Übergangsspannung und injizierter Löcherdichte am Rand der Raumladungszone entspricht (1.114):

$$p_\nu = \frac{n_i^2}{N_\nu} \exp\left(\frac{q_0 U}{k_B T}\right). \quad (1.116)$$

Wegen der Neutralitätsbedingung (1.20) gilt wiederum $n_\nu = p_\nu + N_\nu$. Dieser Zusammenhang vereinfacht sich bei Hochinjektion wegen $p_\nu \gg N_\nu$ weiter zur bekannten Hochinjektionsbedingung

$$n_\nu \approx p_\nu \quad (1.117)$$

1.6.3 Der $n^+\nu$ -Übergang

Auch bei einem Übergang zwischen zwei Schichten gleichartiger Dotierung aber mit unterschiedlichen Dotierstärken, wie z.B. $n^+\nu$ -Übergängen, finden Ausgleichsvorgänge in Form von Diffusion und Drift von Elektronen statt, so daß sich eine Diffusionsspannung einstellt. Setzt man in (1.102) $n(x) = N_\nu$ und $V(x) = V_{n\nu}$ für die ν -Zone und $n(x) = N_D^+$ und $V(x) = V_{nn^+}$ für das n^+ -Gebiet ein und bildet die Differenz $V_{nn^+} - V_{n\nu}$ so erhält man für die Diffusionsspannung $U_{Dn\nu}$

$$U_{Dn\nu} = \frac{k_B T}{q_0} \ln\left(\frac{N_D^+}{N_\nu}\right) \quad (1.118)$$

Auch hier kann durch Anlegen einer Flußspannung, eine Injektion von Elektronen vom n^+ -Gebiet in die ν -Zone erreicht werden. Da Elektronen auch in der ν -Zone Majoritäten sind, handelt es sich hier um eine Majoritätsträgerinjektion. Somit ist in (1.111) die Löcherdichte $p(x)$ (= Minoritätsträger) mit n_i^2/N_ν und $n(x)$ durch

²³Bei schwach dotierten Zonen kann man von vollständiger Ionisation ausgehen (vgl. S. 11 ff.)

n_ν (= Elektronendichte am Rand der Raumladungszone) zu ersetzen. Damit ergibt sich für n_ν

$$n_\nu = N_\nu \exp\left(\frac{q_0 U}{k_B T}\right). \quad (1.119)$$

Auch hier liegt in der Regel Hochinjektion vor, so daß wiederum (1.117) gilt.

2 Der Aufbau physikalischer Modelle

2.1 Modularer Modellaufbau am Beispiel der Leistungsdiode

Die Leistungsdiode ist das am einfachsten aufgebaute Leistungselement, bestehend aus einer p^+ -Zone (Anode), einer n^+ -Zone (Kathode) und einer vergleichsweise langen schwach n -dotierten Driftzone (ν -Zone). Diese Struktur zeigt Bild 2.1.

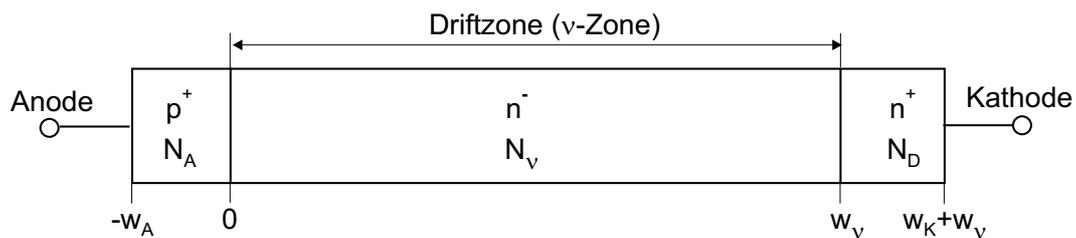


Bild 2.1: Grundstruktur einer Leistungsdiode

Dieses Bauelement wird nicht aktiv ein- oder ausgeschaltet und kann auf jeden Fall eindimensional beschrieben werden. Folglich eignet sich das Modell für die Leistungsdiode besonders gut, das Prinzip der hier vorliegenden physikalischen Modellierung zu erläutern. Außerdem läßt sich bereits mit den hier vorhanden drei Zonen, der modulare Aufbau des Modells, wie er von Vogler [73] vorgeschlagen wurde, veranschaulichen. Allerdings gibt es auch bei Leistungsdioden durchaus komplexere Strukturen, wie Pufferschichten, Lebensdauerprofilen, doppelt diffundierte Randzonen, usw., die, soweit es die beiden letztgenannten Punkte betrifft, in Abschnitt 2.4 beschrieben werden. Außerdem gibt es Dioden mit ausgeprägten 2D-Strukturen, wie z.B. die SPEED-Diode [56], bei der zusätzliche p^+ -Inseln, in die p -dotierte Anodenschicht eindiffundiert werden, mit dem Ziel eine stromdichteabhängige Injektionseffizienz der Anode zu erreichen. Derlei Bauelemente benötigen ein Modell mit zumindest teilweiser zweidimensionaler Beschreibung, was sich aber nur mit erheblichen Aufwand realisieren läßt [35].

Bild 2.2 zeigt das aus den Modulen bestehende Modell, mit Bezugnahme auf den Schaltungssimulator Saber und deren Modellierungssprache MAST.

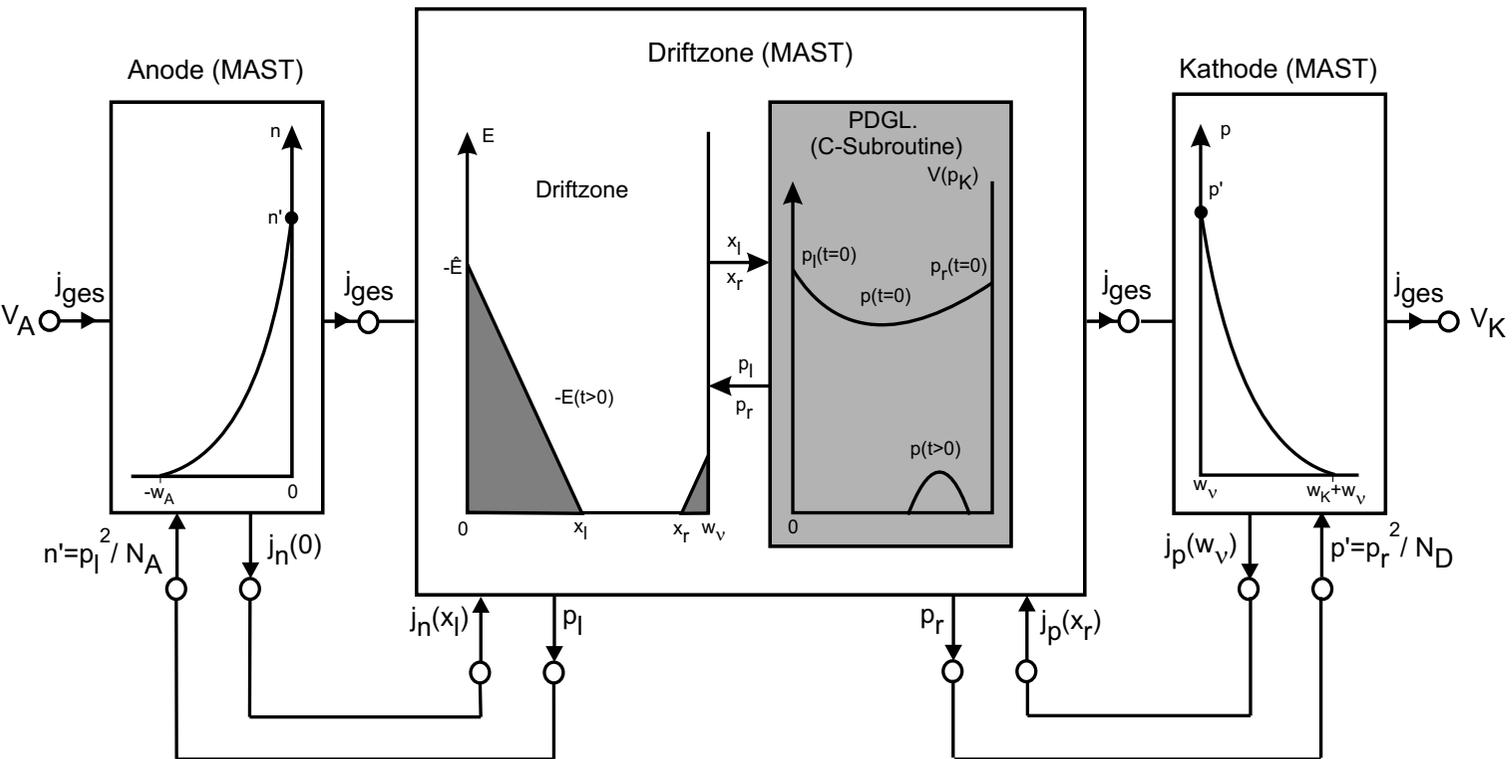


Bild 2.2: Schematische Darstellung der Komponenten des Leistungsdiodenmodells (Darstellung der Löcher- und E-Feldverteilung im Driftzonenmodul für Ausschalten)

Anode und Kathode bilden die sogenannten Randzonen, die für die Injektion von Ladungsträgern in die schwach dotierte Driftzone verantwortlich sind. Deshalb muß bei der Modellierung das Hauptaugenmerk auf das Injektionsverhalten dieser Randzonen gelegt werden. Die Randzonen können hierbei kompakt modelliert werden, d.h. auf eine Ortsdiskretisierung kann verzichtet werden. Die maßgeblichen Gleichungen werden wie in Bild 2.2 angedeutet, direkt in MAST implementiert. Die detaillierte Beschreibung der dazu notwendigen Gleichungen findet sich in Abschnitt 2.2.

Das Kernstück jeden bipolaren Leistungsbauelementemodells ist das Driftzonenmodul. Dem Anoden–Driftzonen–Übergang ($p^+\nu$ -Übergang) wird im folgenden (vgl. Bilder 2.1 und 2.2) die Koordinate $x = 0$ zugeordnet¹. Der Widerstand der Driftzone im Durchlaßfall, also bei Injektion, wird durch die injizierten Ladungsträger erheblich vermindert, was einerseits mit Blick auf statische Verluste ein großer Vorteil ist. Andererseits erhöhen sich, bedingt durch die Notwendigkeit, diese Ladungen beim Ein- bzw. Ausschalten, einbringen bzw. ausräumen zu müssen, die dynamischen Verluste massiv. Der richtigen Beschreibung des Ladungsträgerverhaltens kommt also eine zentrale Bedeutung zu. Wie bereits in der Einleitung dieser Arbeit angedeutet, kann auf die ortsdiskrete Darstellung der, dem Ladungsträgerverhalten zugrundeliegenden partiellen Differentialgleichung und deren numerische Lösung für jeden Zeitschritt, nicht verzichtet werden. Die implementierte Numerik befindet sich in einer, in der Programmiersprache C verfaßten Subroutine, die sich unproblematisch über eine dafür in SABER vorgesehene Schnittstelle an das eigentliche MAST-Modul für die Driftzone ankoppeln läßt. Bei Betrachtung des statischen Verhaltens der Diode ist bezüglich der ν -Zone folgendes zu sagen: Entweder es existiert nur ein Ladungsträgerprofil wie die Hängekurve $p(t = 0)$ in Bild 2.2 für den Ein-Zustand, oder eine Raumladungszone mit dreiecksförmigen bzw. trapezförmigen E-Feld-Verlauf im Aus-Zustand. Bei den dynamischen Übergängen ist dies grundlegend anders, wie dies in Bild 2.2 für den Fall des Übergangs vom Durchlaß in den Sperrzustand dargestellt ist: Für jeden Zeitpunkt während des Abschaltens, existiert ein Bereich $x(t) \in [x_l(t), x_r(t)]$ mit Restladung (Plasmagebiet) und einem Bereich $x(t) \in [0, x_l(t)[$ bzw. $]x_r(t), w_\nu]$ mit dreiecksförmigem E-Feldverlauf (Raumladungszone). Das Ausmaß der Raumladungszone wird im MAST-Driftzonenmodul bestimmt. Die beweglichen Grenzen $x_l(t)$ und $x_r(t)$ werden an die Subroutine übergeben. Darin wird die vorher definierte Anzahl der Stützstellen $N_{stütz}$ immer wieder auf den „aktuellen“ Bereich $[x_l(t), x_r(t)]$ aufgeteilt, so daß man insgesamt mit relativ wenigen Stützstellen auskommt und dadurch eine merkliche Erhöhung der Simulationsgeschwindigkeit erzielen kann. Die Berechnung der Verteilung des E-Feldes

¹Ganz korrekt müßte die Koordinate des Anodenrandes $x = 0 - \Delta_A$ und die Koordinate des Driftzonenbeginns $x = 0 + \Delta_\nu$ lauten, da selbst bei starker Injektion immer eine kleine Raumladungszone $w_{RLZ} = \Delta_A + \Delta_\nu$ ohne freibewegliche Ladungsträger verbleibt. Man kann jedoch in sehr guter Näherung die Stromdichten entlang dieser RLZ konstant setzen, d.h. $j_n(0 - \Delta_A) = j_n(0 + \Delta_\nu) = j_n(0)$ und somit den $p^+\nu$ -Übergang zu einer einzigen Koordinate $x = 0$ zusammenfassen.

fällt somit in den Bereich des MAST-Moduls, d.h. auf eine Ortsdiskretisierung wird hier verzichtet. Man behilft sich, indem man einen streng linearen, d.h. dreiecksförmigen E-Feld-Verlauf unterstellt. Zwar ist der Verlauf tatsächlich nicht ganz linear, doch stellt die o.g. Betrachtung eine ausreichend gute Näherung zur Berechnung der Sperrspannung dar. Eine detaillierte Beschreibung der Driftzone erfolgt in Abschnitt 2.3.

2.2 Modellierung der Randzonen

2.2.1 Injektionsverhalten und Stromverläufe

In Abschnitt 1.6 wurde festgestellt, daß an vorwärtsgepolten pn -Übergängen, Injektion von Ladungsträgern stattfindet. So injiziert die ν -Zone Elektronen in die Anodenschicht, wobei es sich dann bekanntlich um schwache Injektion handelt, d.h. $n(0)|_{Anode} = n_p \ll N_A^-$. Ferner gilt die Kontinuitätsgleichung für Elektronen (1.38), die im stationären Fall, d.h. $\partial n/\partial t = \partial n'/\partial t = 0$ und mit $R_{netto,n} = n'/\tau_n$

$$\frac{1}{q_0} \frac{dj_n}{dx} - \frac{n'}{\tau_n} = 0 \quad (2.1)$$

lautet². Für die Elektronenstromdichte gilt (1.32), wobei wegen $n' \ll N_A^-$ der Driftterm gegenüber dem Diffusionsterm vernachlässigt werden kann. Also gilt:

$$j_n = q_0 D_n \frac{\partial n}{\partial x} \quad (2.2)$$

Setzt man (2.2) in (2.1) ein, ergibt sich die Minoritätsträgerdiffusionsgleichung

$$\frac{d^2 n}{dx^2} - \frac{n'}{L_n^2} = 0, \quad (2.3)$$

wobei L_n die sogenannte Diffusionslänge für Elektronen darstellt, die zu

$$L_n = \sqrt{D_n \tau_n} \quad (2.4)$$

definiert ist. Die allgemeine Lösung dieser gewöhnlichen Differentialgleichung zweiter Ordnung lautet:

$$n'(x) = C_1 \sinh\left(\frac{x}{L_n}\right) + C_2 \cosh\left(\frac{x}{L_n}\right) \quad (2.5)$$

²Zur Festlegung der Minoritätsträgerlebensdauer τ_n bzw. τ_n siehe Anmerkungen 3,4 und 6 in Abschnitt 1.4.4 und Abschnitt 2.2.3

Hierbei sind C_1 und C_2 Konstanten, die mit geeigneten Randbedingungen festgelegt werden. Diese lauten im Koordinatensystem von Bild 2.2, $n'(0) = n_p$, mit n_p gemäß (1.112) und aufgrund der Tatsache, daß die Dichte der Minoritätsträger am Metall-Halbleiterübergang null sein muß, $n'(-w_A) = 0$. Die spezielle Lösung mit diesen Randwerten ergibt sich somit zu

$$n'(x) = \frac{n_p}{\sinh\left(\frac{w_A}{L_n}\right)} \sinh\left(\frac{x + w_A}{L_n}\right) \quad (2.6)$$

Für die Kathode ergibt sich für die Löcherdichte eine analoge Lösung. Bei (2.6) handelt es sich letztlich um eine bei $x = 0$, mit der Maximalkonzentration n_p , startende exponentiell abklingende Funktion, wie sie auch im linken Teil von Bild 2.3 gezeichnet ist.

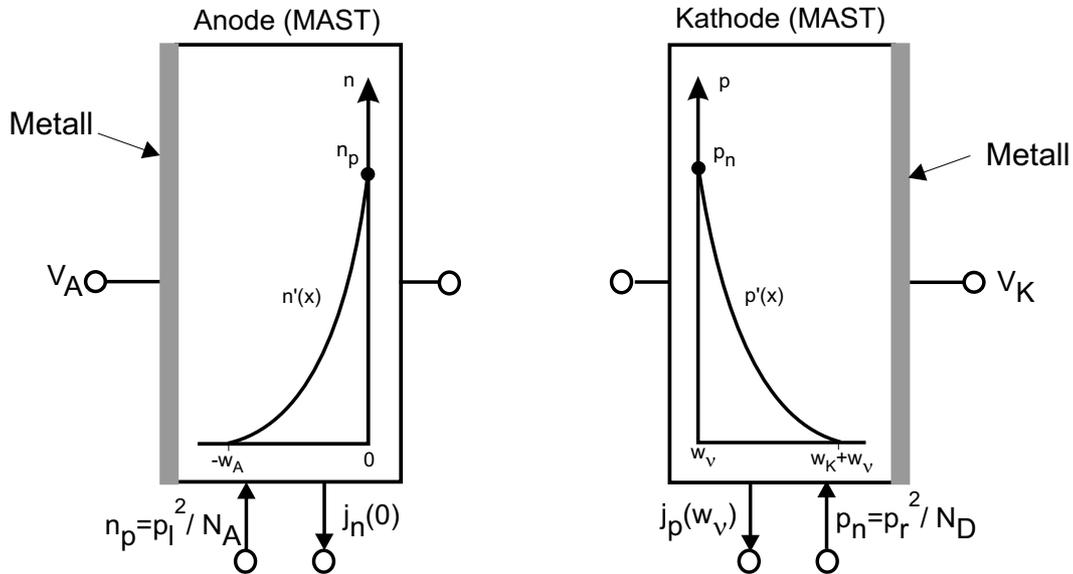


Bild 2.3: Die Randzonenmodule: links Anode und rechts Kathode

Allerdings klingt die Ladungsträgerdichtefunktion tatsächlich nur für $w_A \gg L_n$ exponentiell ab.³ Durch Einsetzen von (2.6) in (2.2), kann nun die Elektronenstromdichte für $-w_A < x < 0$ berechnet werden. Es ergibt sich

$$j_n(x) = \frac{q_0 D_n n_p}{L_n \sinh\left(\frac{w_A}{L_n}\right)} \cosh\left(\frac{x + w_A}{L_n}\right). \quad (2.7)$$

Für die weitere Analyse des Injektionsverhaltens der Anode ist vor allem die Elektronenstromdichte direkt am p^+n -Übergang, also bei $x = 0$, entscheidend. Letztlich

³Ist diese Bedingung verletzt, erhält man eine sogenannte transparente Randschicht, wie sie in Abschnitt 2.2.2 beschrieben wird.

gilt für das Ladungsträgerprodukt am $p^+\nu$ -Übergang⁴ $n_p N_A^- = p_l^2$. Ersetzt man nun n_p durch diesen Ausdruck ergibt sich für $j_n(0)$

$$j_n(0) = \frac{q_0 D_n}{L_n} \cdot \coth\left(\frac{w_A}{L_n}\right) \cdot \frac{p_l^2}{N_A^-}. \quad (2.8)$$

Erweitert man die rechte Seite von (2.8) mit n_i^2 , bekommt man

$$j_n(0) = j_{n0} \left(\frac{p_l}{n_i}\right)^2 \quad (2.9)$$

wobei j_{n0} die sogenannte Sperrsättigungsstromdichte der Anode ist. Sie ist definiert durch

$$j_{n0} = \frac{q_0 D_n n_i^2}{L_n N_A^-} \cdot \coth\left(\frac{w_A}{L_n}\right) \quad (2.10)$$

Im Falle Kathode lassen sich entsprechende Beziehungen ableiten. Hierbei bilden Löcher die Minoritätsträger und man erhält entsprechend eine Diffusionsgleichung wie (2.3), indem man einfach n durch p ersetzt (s. Bild 2.3 rechts). Nun interessiert die Löcherstromdichte am $n^+\nu$ -Übergang, also $j_p(w_\nu)$ und man erhält anstatt (2.9)

$$j_p(w_\nu) = j_{p0} \left(\frac{p_r}{n_i}\right)^2. \quad (2.11)$$

Hierbei ist j_{p0} die Sperrsättigungsstromdichte der Kathode⁵ und ist dito (2.10) definiert durch

$$j_{p0} = \frac{q_0 D_p n_i^2}{L_p N_D^+} \cdot \coth\left(\frac{w_K}{L_p}\right) \quad (2.12)$$

Damit kennt man nun die Gleichungen, die die Randzonen charakterisieren und es bietet sich an, deren Wirken innerhalb des Diodenmodells etwas näher zu betrachten. So erhält aus (2.9)

$$p_l = n_i \sqrt{\frac{j_n(0)}{j_{n0}}}, \quad (2.13)$$

indem man die Formel nach p_l umstellt. Hierbei kann man in guter Näherung $j_n(0)$ sowohl an der Anodengrenze als auch an der Driftzongrenze, also im gesamten Bereich der schmalen Raumladungszone, als konstant betrachten. Dies bedeutet, daß auch die Ladungsträgerverhältnisse am Driftzonenrand zu diesem $j_n(0)$ passen müssen, oder mit anderen Worten: Der Gradient $dp/dx(0)$ und die Ladungsträgerdichte $p(0) = p_l$ am Rand der ν -Zone müssen sich so einstellen, daß sich $j_n(0)$ ergibt

⁴ p_l : Löcherdichte auf der linken Seite der Driftzone (s. Bild 2.2); es gilt $p_l = n_l$ wegen Hochinjektion

⁵Bei der Signaldiode bildet die Summe aus j_{n0} und j_{p0} die Sperrstromdichte, aus der man durch Multiplikation mit der effektiv wirksamen Diodenfläche den Sperrstrom erhält.

(vgl. Abschnitt 2.3). Gleichung (2.13) bildet somit die Bestimmungsgleichung für die anodenseitige Löcherinjektion in die Driftzone⁶.

Betrachtet man (2.13) noch etwas genauer, fällt sofort auf, daß sich der Betrag der in die Driftzone injizierten Löcher p_l , indirekt proportional zur Höhe der Sperrsättigungsstromdichte für Elektronen j_{n0} verhält. Ein großes j_{n0} impliziert demnach neben einem großen statischen Sperrstrom, eine geringe Injektionsfähigkeit und damit zwar vergleichsweise große statische, aber geringere dynamische Verluste (schnelle Diode). Auch für die Beziehung zwischen j_{p0} und p_r an der Driftzonen-/Kathodengrenze ergibt sich die indirekte Proportionalität (s. (2.11)). Die beiden Randdichten p_l und p_r und das sich dazwischen ergebende Ladungsträgerprofil (s. Abschnitt 2.3), bestimmen aber das komplette statische und dynamische Verhalten der Leistungsdiode. Für den Modellentwickler und letztlich noch mehr für den Modellanwender, ergibt sich die komfortable Situation, daß sich das Injektionsverhalten von hoch dotierten Randzonen, durch eine einzige Größe, nämlich der Sperrsättigungsstromdichte der jeweiligen Minoritätsträger, d.h. j_{n0} oder j_{p0} , ausdrücken läßt. Folglich kann auf eine genaue Modellierung der Randzonen inklusive aller Strukturparameter, die gemäß (2.10) und (2.12 in $j_{n0,p0}$ eingehen, verzichtet werden, denn ein Diodenmodell läßt sich bequem mittels Variation der Parameter j_{n0} oder j_{p0} an reale Meßkurven anpassen⁷. Diese Methode wurde bereits erfolgreich in dem Diodenmodell von Vogler [73] angewendet. Leider gelten diese Aussagen wegen der nichtlinearen Abhängigkeit der Strukturparameter von der Temperatur, nur wenn man sich auf eine einzige Temperatur beschränkt. Geeignete Maßnahmen zur Modellbildung dazu, werden in Abschnitt 2.2.3 beschrieben.

2.2.2 Die transparente Anode

Wie man in Bild 2.3 links erkennt, wurde eine exponentieller Verlauf der Elektronendichte als Lösung der Diffusionsgleichung (2.3) eingezeichnet. Diese Lösung erhält man, wenn man annimmt, daß $w_A \gg L_A$ gilt. In der Regel ist diese Annahme bei Randzonen mit Längen im Bereich $w_a \geq 10\mu m$ und sehr hohen Dotierungen durchaus gerechtfertigt, denn hohe Dotierungen verringern die Minoritätslebensdauer τ_n (vgl. Scharfetter-Relation (1.47)) und damit wegen (2.4) die Diffusionslänge. Dies soll im folgenden kurz gezeigt werden. Für die sinh-Funktion gilt (vgl. [10]):

$$\sinh(y) = \frac{\exp(y) - \exp(-y)}{2} \approx \frac{1}{2} \exp(y) \text{ für } y \gg 1 \quad (2.14)$$

⁶Für die kathodenseitige Elektroneninjektion $n_r = p_r$ gelten analoge Aussagen.

⁷Der „Wert“ dieses Sachverhalts wird noch dadurch vergrößert, daß Bauelementhersteller in der Regel nur spärlich Strukturdaten ihrer Bauelemente an Dritte weitergeben.

Wendet man diese Näherung auf (2.6) an, erhält man für $n'(x)$

$$n'(x) = n_p \exp\left(\frac{x}{L_n}\right) \quad (2.15)$$

Für $j_n(0)$ gilt mit (2.15) eingesetzt in (2.2) demnach

$$j_n(0) = \frac{q_0 D_n n_p}{L_n} \exp\left(\frac{x}{L_n}\right) \quad (2.16)$$

Betrachtet man nun $j_n(-w_A)$ kurz vor dem Metallkontakt, so erhält man mit (2.16) $j_n(-w_A) = 0$ wegen $|-w_A| \gg L_n$, d.h. direkt am Metallkontakt gibt es nur noch einen Majoritätsträgerstrom j_p . Anders ausgedrückt: Der Metallkontakt kann den Elektronendiffusionstrom, der bei $x = 0$ maximal ist nicht „sehen“ (opaque Anode).

Für sehr kurze Anoden ist allerdings auch die Situation $w_A \ll L_n$ vorstellbar. Für diese Annahme, kann die sinh-Funktion mit

$$\sinh(y) \approx y \text{ für } y \ll 1 \quad (2.17)$$

angenähert werden (vgl. [10]). Damit ergibt sich für das hier gewählte Koordinatensystem

$$n'(x) = n_p \frac{x + w_A}{w_A}, \quad (2.18)$$

also eine Geradengleichung für $n'(x)$ als Lösung. Bild 2.4 rechts zeigt diese lineare Elektronenverteilung im Vergleich zur oben beschriebenen exponentiellen Lösung, die in Bild 2.4 auf der linken Seite zu sehen ist.

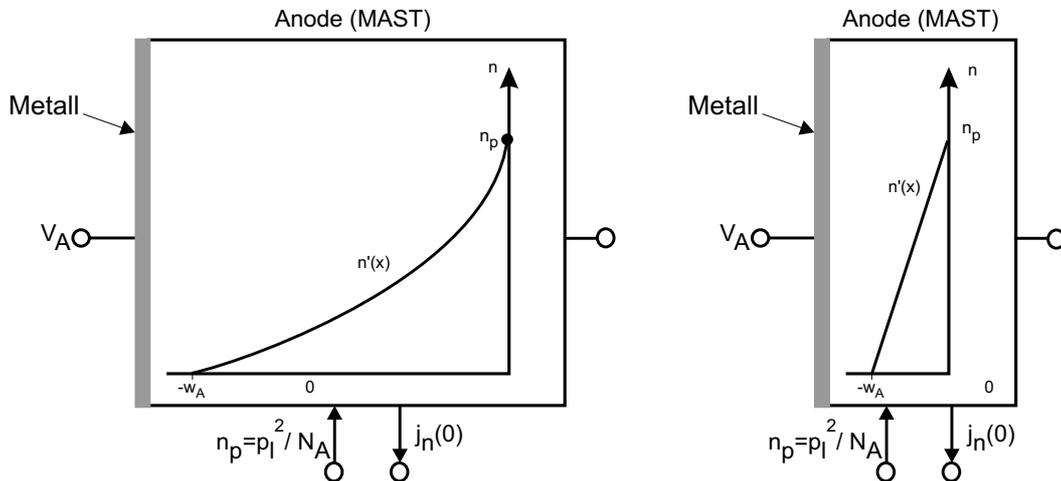


Bild 2.4: „normale“ Anode (links) vs. transparente Anode (rechts)

Berechnet man wiederum $j_n(0)$, indem man die Lösung (2.18) in (2.2) einsetzt, bekommt man

$$j_n(0) = \frac{q_0 D_n n_p}{w_A} = \text{const!} \quad (2.19)$$

Die Elektronenstromdichte ist also im gesamten Anodenbereich gleich, also auch und besonders an der Stelle $x = -w_A$. Weil nun der Anodenkontakt den Elektronenstrom für $x = 0$ „sehen“ kann, spricht man von einer transparenten (=durchsichtigen) Anode. Für die Sperrsättigungsstromdichte ergibt sich in diesem Fall aus (2.10) mit $\coth(y) = 1/y$ für $y \ll 1$:

$$j_{n0} = \frac{q_0 D_n n_i^2}{w_A N_A^-} \quad (2.20)$$

Die Sperrsättigungsstromdichte ist also unabhängig von der Diffusionslänge und damit auch von der oft unbekanntem bzw. schwer zu bestimmenden Minoritätsträgerlebensdauer. Weiterhin bieten transparente Anoden die Möglichkeit, mit schmalen Zonen und relativ niedrigen Dotierungen, hohe Sperrsättigungsstromdichten, zwecks Reduktion der Injektionsfähigkeit, zu erreichen. Genau aus diesem Grund wird die transparente Anode beim IGCT angewendet (s. dazu Kapitel 3).

2.2.3 Temperaturverhalten

Wie bereit am Ende von Abschnitt 2.2.1 erläutert, genügt j_{n0} (j_{p0}) zur vollständigen Beschreibung des Injektionsverhaltens der Anode (Kathode). Parametrisiert man ein Leistungsdiode Modell anhand bestimmter Meßkurven, aufgenommen für eine bestimmte (beliebige) Temperatur T_0 , so erhält man eben auch $j_{n0}(T_0)$ und $j_{p0}(T_0)$, gültig primär für diese eine Temperatur T_0 . Nun stellt sich die Frage, was passiert, wenn man dieses so parametrisierte Modell für eine andere Temperatur T verwendet. Um dann richtige Simulationskurven errechnen zu können, müßten sich $j_{n0,p0}(T)$ explizit durch $j_{n0,p0}(T_0)$ und T ausdrücken lassen, d.h. für den Fall der Anode müßte sich

$$j_{n0}(T) = f(T, j_{n0}(T_0)). \quad (2.21)$$

schreiben lassen. Hierbei bietet es sich an, im Modell $T_0 = 300K$ festzulegen, d.h. man legt bei der Modellanpassung immer $j_{n0,p0}(300K)$ fest und rechnet dann für beliebige Temperaturen T um. In den Modellen von Vogler [72], [73], wird ein im Vergleich zu (2.21) vereinfachter Ausdruck der Struktur

$$j_{n0}(T) = j_{n0}(300K) \cdot f(T) \quad (2.22)$$

verwendet. Damit ließ sich das Temperaturverhalten in gewissen Grenzen nachbilden. Allerdings zeigen neuere Untersuchungen in [57], [59], daß das Injektionsverhalten der Randzonen, beschrieben durch (2.21) bzw. (2.22), insbesondere für tiefere Temperaturen, nicht mehr stimmt. Dies ist auch unmittelbar einsichtig, wenn man sich (2.10) mit den temperaturabhängigen Parametern ansieht:

$$j_{n0} = \frac{q_0 D_n(T) n_i(T)^2}{L_n(T) N_A^-(T)} \cdot \coth\left(\frac{w_A}{L_n(T)}\right) \quad (2.23)$$

Hierbei gelten die Einstein–Relationen⁸, d.h.

$$D_n(T) = k_B T \mu_n(T)/q_0. \quad (2.24)$$

Weiterhin gilt mit (2.4):

$$L_n(T) = \sqrt{D_n(T) \tau_n(T)}. \quad (2.25)$$

Setzt man die Ausdrücke für die einzelnen Temperaturabhängigkeiten aus Kapitel 1 in (2.23) ein, erhält man eine Gleichung, die sich nicht in die Form (2.21) oder (2.22) bringen läßt.

Somit kommt man doch nicht daran vorbei, die Designgrößen der Randzonen $N_{A,K}$ und $w_{A,K}$ in dem Modell anzugeben. Problematisch bleibt die Verwendung von $\tau_{n,p}$, da diese Größe selbst Bauelementeherstellern oft unbekannt ist, bzw. die Berechnung ihrer Dotierabhängigkeit, z.B. gemäß der Scharfetter–Beziehung (1.47), eher mit Unsicherheiten behaftet zu sein scheint. Um die Konsequenzen dieser Problematik kurz zusammenzufassen: Der Aufwand zum Adaptieren eines solches Modells an reale Meßkurven steigt im Vergleich zum oben beschriebenen j_{n0} –Konzept erheblich. Deshalb wird auch in den vorliegenden Modellen auf dieses Konzept nicht verzichtet, sondern die Temperaturproblematik folgendermaßen angemessen berücksichtigt: Weiterhin bleibt $j_{n0,p0}(300K)$ der Aktionsparameter zum Anpassen der Injektionsfähigkeit der Randzonen, wobei zusätzlich $N_{A,K}$ und $w_{A,K}$ angegeben werden⁹. Dann wird mit Hilfe von (2.23) für $T = 300K$, τ_n extrahiert, anschließend auf die gewünschte Temperatur T_1 umgerechnet und zusammen mit allen anderen Größen jetzt für T_1 in (2.23) eingesetzt und daraus $j_{n0}(T_1)$ berechnet¹⁰.

Setzt man allerdings (2.25) in (2.23) ein, erkennt man sofort, daß τ_n sowohl im Argument des cosh–Terms als auch im Nenner des Bruch–Terms auftaucht. Dies bedeutet, daß sich (2.23) nicht explizit nach τ_n auflösen läßt. Glücklicherweise kann man im vorliegenden Fall recht gut nähern. Dazu wird (2.23) mit w_A erweitert, so daß sich

$$j_{n0} = \underbrace{\frac{q_0 D_n n_i^2}{w_A N_A^-}}_{S(.)} \cdot \underbrace{\frac{w_A}{L_n}}_y \cdot \underbrace{\coth\left(\frac{w_A}{L_n}\right)}_{\coth y} \quad (2.26)$$

ergibt. Hierbei werden also die Designparameter zu einem Faktor $S(.)$ zusammengefaßt und $y := w_A/L_n$ gesetzt. Das Verhältnis w_A/L_n (dito w_K/L_p) nimmt in der Realität meist kleine Werte im Bereich $0 < y \leq 10$ an. Man kann nun die Funktion $y \coth y$ sehr gut durch eine gebrochen rationale Funktion im Bereich $y > 0$

⁸Elektronen–Loch–Streuung ist bei den Randzonen nicht berücksichtigt, so daß die Einstein–Relationen gültig sind

⁹Ein Anwender, dem das Temperaturverhalten nicht interessiert, kann auf die Eingabe von $N_{A,K}$ und $w_{A,K}$ verzichten, weil das Modell dafür sinnvolle Defaultwerte zur Verfügung stellt.

¹⁰Die Aussagen an dieser Stelle und im folgenden gelten prinzipiell auch für die Kathode

annähern. Diese Näherung lautet:

$$y \coth y \approx \frac{y^2 + 1,08x + 1,73}{x + 1,76} \quad (2.27)$$

Bild 2.5 zeigt die Näherung (strichpunktierte Linie) im Vergleich zur Originalfunktion (durchgezogene Linie) und den dadurch verursachten relativen Fehler ϵ_{rel} in %, der sich innerhalb des Bereichs $\pm 5\%$ bewegt.

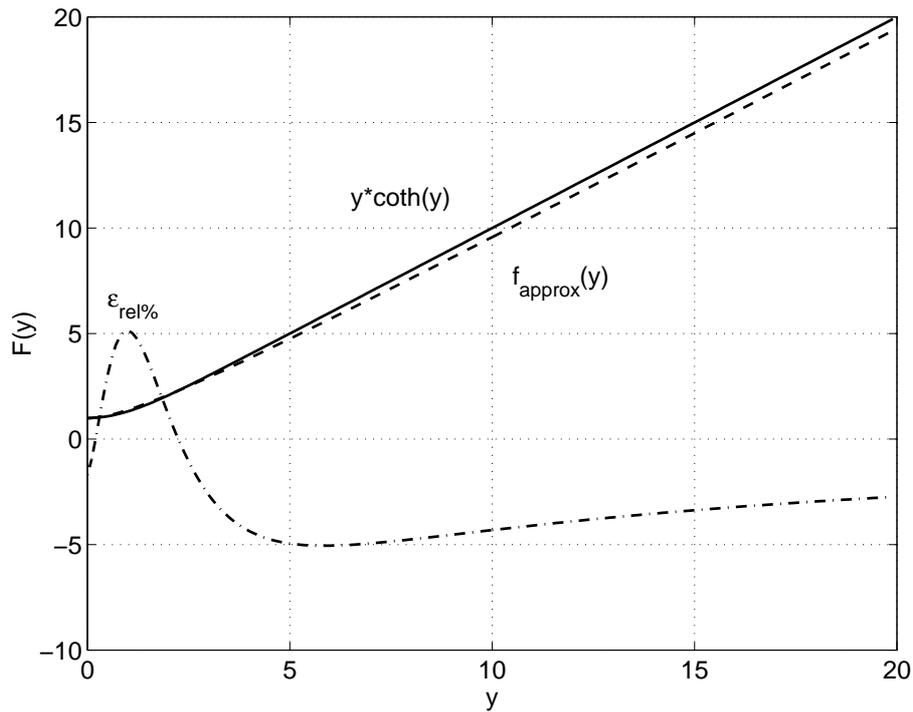


Bild 2.5: Graphische Darstellung der Funktion $y \coth(y)$ (durchgezogene Linie), einer geeigneten Näherungsfunktion (gestrichelte Linie) und des relativen Fehlers ϵ_{rel} in % (strichpunktierte Linie), den man mit dieser Näherung macht.

Setzt man diese Näherung in (2.26) ein, kann daraus eine quadratische Gleichung für y gebildet werden. Ihre zugehörige Lösung lautet:

$$y = \frac{-(1,08 S - j_{n0}) + \sqrt{(1,08 S - j_{n0})^2 + (7,04 j_{n0} - 6,92 S)}}{2 S} \quad (2.28)$$

Mit sinnvollen Werte für j_{n0} (j_{p0}) und den Designparametern, ergibt sich ein positives y . Dieses y kann dann zurücks substituiert werden, d.h. man erhält L_n (L_p) und daraus wiederum τ_n (τ_p). Damit hat man alle Parameter für $T = 300K$ und kann dann mit den Ausdrücken aus Kapitel 1, j_{n0} (j_{p0}) für jede beliebige Temperatur berechnen.

2.3 Modellierung der Driftzone

2.3.1 Beschreibung des Ladungsträgerverhaltens

Wie bereits erwähnt, hat man es bei schwach dotierten Driftzonen mit Hochinjektion zu tun, falls die Diode in Durchlaßpolung betrieben wird. Setzt man $n \approx p$ in (1.31) und (1.32) ein, ergibt sich für die Stromdichtegleichungen:

$$j_p = q_0 \mu_p p E - q_0 D_p \frac{\partial p}{\partial x} \quad (2.29)$$

$$j_n = q_0 \mu_n p E + q_0 D_n \frac{\partial p}{\partial x}. \quad (2.30)$$

Addiert man (2.29) und (2.30) zur Gesamtstromdichte j_{ges} , löst das Ergebnis nach dem elektrischen Feld auf und setzt den resultierenden Term in (2.29) und (2.30) ein, erhält man die sogenannten ambipolaren Stromdichtegleichungen:

$$j_p = \frac{1}{b+1} j_{ges} - q_0 D \frac{\partial p}{\partial x} \quad (2.31)$$

$$j_n = \frac{b}{b+1} j_{ges} + q_0 D \frac{\partial p}{\partial x}. \quad (2.32)$$

Hierbei ist b das sogenannte Beweglichkeitverhältnis.

$$b = \frac{\mu_n}{\mu_p} \quad (2.33)$$

Dabei ergibt sich der allgemeine Diffusionskoeffizient D aus obiger Herleitung zu

$$D = \frac{D_n + b D_p}{1 + b}. \quad (2.34)$$

Im folgenden wird gezeigt, daß sich je nachdem, ob man Elektronen–Löcher–Streuung (Fälle II und III) berücksichtigt oder nicht (Fall I), unterschiedliche Kontinuitätsgleichungen zur Beschreibung des Ladungsträgerverhaltens in der ν -Zone ergeben. Unterschiede ergeben sich besonders auch zwischen der herkömmlichen CCS–Beschreibung (Fall II) und der EHS-Theorie nach Mnatsakanov (Fall III).

Fall I: Keine Elektronen–Löcher–Streuung

Da bei Nichtberücksichtigung von EHS (vgl. Fall III), die Einstein–Beziehungen (1.33) gelten, folgt aus (2.33) auch $b = D_n/D_p$. Setzt man dies in (2.34) ein, erhält man die sogenannte ambipolare Diffusionskonstante D_A

$$D_A = 2 D_p D_n / (D_p + D_n) = D_p \frac{2b}{b+1} \quad (2.35)$$

Im vorliegenden Fall I gilt also $D = D_A$, weshalb in (2.31) und (2.32), D durch D_A ersetzt werden kann. Durch Einsetzen von (2.31) in (1.37), erhält man dann die gewünschte Kontinuitätsgleichung, wobei b und D_A Konstanten bzgl. des Ortes x sind. Somit erhält man die sogenannte ambipolare Diffusionsgleichung

$$\frac{\partial p}{\partial t} = D_A \frac{\partial^2 p}{\partial x^2} - \frac{p}{\tau_h}, \quad (2.36)$$

wobei gemäß (1.52)

$$\frac{1}{\tau_h} = \frac{1}{\tau_{h,SRH}} + \frac{1}{\tau_{h,A}} \quad (2.37)$$

gilt, da in Driftzonen neben der indirekten Rekombination (SRH-Beschreibung; s. Abschnitt 1.4.2), auch der Auger-Effekt (s. Abschnitt 1.4.3) eine Rolle spielen kann. Für den statischen Fall, d.h. $\partial p/\partial t = 0$, erhält man mit der Definition der ambipolaren Diffusionslänge $L_A = \sqrt{D_A \tau_h}$

$$\frac{d^2 p}{dx^2} - \frac{p}{L_A^2} = 0, \quad (2.38)$$

also eine ähnliche Gleichung wie (2.3), nur für Löcher. Die allgemeine Lösung entspricht (2.5), allerdings nun für $p(x)$. Die Randwerte für die Diode im Durchlaßfall, sind $p(0) = p_l$ (anodenseitige Injektion) und $p(w_\nu) = p_r$ (kathodenseitige Injektion). Als spezielle Lösung ergibt sich damit

$$p(x) = p_l \frac{\sinh\left(\frac{w_\nu - x}{L_A}\right)}{\sinh\left(\frac{w_\nu}{L_A}\right)} + p_r \frac{\sinh\left(\frac{x}{L_A}\right)}{\sinh\left(\frac{w_\nu}{L_A}\right)} \quad (2.39)$$

Damit erhält man eine Kurve für $p(x)$, die an ein durchhängendes Seil erinnert, das links an p_l und rechts an p_r aufgehängt ist und dessen Durchhang, d.h. die Lage und der Betrag des Minimums, von dem Verhältnis w_ν/L_A abhängt. So eine „Hängekurve“ ist u.a. in Bild 2.2 (s. Driftzonenmodul rechts für $p(t = 0)$) schematisch eingezeichnet. Die Lösung (2.39) gilt aber so nur bei statischem Durchlaßbetrieb. Wie in Abschnitt 2.1 beschrieben, können für Schaltvorgänge keine analytischen Ausdrücke als allgemeine Lösung angegeben werden, d.h. (2.36) muß numerisch gelöst werden. Die Kurve für $p(t > 0)$ in Bild 2.2 zeigt eine Löcherverteilung beim Übergang vom leitenden in den sperrenden Zustand für irgendeinen Zeitpunkt des dynamischen Vorgangs. Hierbei sind wegen der Ausdehnung der Raumladungszone (Spannungsaufnahme), $p(x_l) = 0$ statt $p(0)$ und $p(x_r) = 0$ statt $p(w_\nu)$ die Randwerte, oder mit anderen Worten: Die für die Randwerte maßgeblichen Grenzen $x_l(t)$ und $x_r(t)$ wandern. Daher bietet sich ein bewegliches Koordinatensystem an (s. Abschnitt 2.3.2).

Die Ladungsträgerdichte in der Driftzone ist maßgeblich für den Durchlaßwiderstand

dieser Zone verantwortlich. Addiert man (2.29) und (2.30), löst nach E auf und ersetzt die Diffusionskonstanten durch die Beweglichkeiten (Einstein-Beziehungen), erhält man

$$E(x) = \frac{j_{ges}}{q_0 \mu_p (b+1) p(x)} - \frac{k_B T}{q_0} \frac{b-1}{b+1} \frac{1}{p(x)} \frac{\partial p(x)}{\partial x} \quad (2.40)$$

Den Spannungsabfall im Plasmagebiet erhält man mittels Einsetzen der Lösung von (2.38) (im statischen Fall) in (2.40) und anschließender Integration über das Plasmagebiet.

Fall II: Elektronen–Löcher–Streuung gemäß herkömmlicher Theorie

Wie in Abschnitt 1.5.4 erläutert wurde, steigt die Relevanz der Elektronen–Löcher–Streuung mit zunehmender Dichte und abnehmender Temperatur. Somit sind die Beweglichkeiten μ_p und μ_n abhängig von der Ladungsträgerdichte p und damit vom Ort x . Folglich ist das Beweglichkeitsverhältnis $b = \mu_n(x)/\mu_p(x)$ keine Konstante mehr. Weiterhin ergibt sich aus den Einstein–Beziehungen (1.81), daß die Diffusionskonstanten D_n und D_p und damit auch D ebenso Funktionen von x sind. Somit ergibt sich durch Einsetzen von (2.31) in (1.37) die Diffusionsgleichung

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2} + \frac{dD}{dx} \frac{\partial p}{\partial x} - \frac{j_{ges}}{q_0} \frac{1}{(b+1)^2} \frac{db}{dx} - \frac{p}{\tau_h}, \quad (2.41)$$

die die Ableitungen dD/dx und db/dx enthält. Ansonsten gelten die bei Fall I gemachten Aussagen auch für den Fall II, außer daß der Rechenaufwand steigt. Besonders gilt weiterhin (2.40) zur Berechnung des Spannungsabfalls, sofern man b durch $b(x)$ ersetzt und berücksichtigt, daß μ^C in μ_p eingeht.

Fall III: Elektronen–Löcher–Streuung nach Mnatsakanov

In Abschnitt 1.5.5 waren die Transportgleichungen (1.83) und (1.84) Ausgangspunkt der Überlegungen. Hierbei wurden die Beweglichkeiten und die Diffusionskonstanten so zerlegt, daß die Elektronen–Löcher–Streuung in eigenen Drag–Termen erfaßt werden konnte. Diese Mnatsakanov–Beweglichkeiten (μ_n^n, μ_n^p, \dots) bestehen ihrerseits aus den Beweglichkeiten $\mu_{n,p}^{LI}$ (Gitter- und Störstellenstreuung) und den EHS–Beweglichkeiten $\mu_{np,pn}$ (s. dazu (1.87) bis (1.90)). Gleiches gilt für die Diffusionskonstanten (s. dazu (1.91)). Die Gleichungen (1.83) und (1.84) lassen sich durch geeignete Kombinationen der auftretenden Beweglichkeiten und Diffusionskonstanten (s. dazu (1.85) und (1.86)) in die Form (1.31) und (1.32) transformieren. Allerdings fällt auf, daß dann der einfache Zusammenhang von Beweglichkeiten und Diffusionskonstanten gemäß den Einstein–Relationen (1.33), zumindest was die letztendlichen Gesamtgrößen betrifft, nicht mehr gilt.

Zur Herleitung der Kontinuitätsgleichung¹¹ wird abweichend von den Fällen I und II, nicht sofort die Hochinjektionsnäherung $p \approx n$, sondern allgemein $n = p + N_\nu$ (Neutralitätsbedingung) angesetzt. Die Gleichungen (1.31) und (1.32) lauten dann

$$j_p = q_0 \mu_p p E - q_0 D_p \frac{\partial p}{\partial x} \quad (2.42)$$

$$j_n = q_0 \mu_n (p + N_\nu) E + q_0 D_n \frac{\partial p}{\partial x} \quad (2.43)$$

Diese können wieder zur Gesamtstromdichte j_{ges} addiert und nach dem elektrischen Feld umgestellt werden. Hierbei kann aufgrund der Ungültigkeit der Einstein-Beziehungen, neben dem Beweglichkeitsverhältnis $b = \mu_n/\mu_p$, ein eigenständiges Verhältnis der Diffusionskonstanten $c = D_n/D_p$ definiert werden. Somit ergibt sich

$$E = \frac{j_{ges} - q_0 D_p (c - 1) \frac{\partial p}{\partial x}}{q_0 \mu_p [(b + 1) p + b N_\nu]} \quad (2.44)$$

Im Gegensatz zum Fall I beinhaltet das Beweglichkeitsverhältnis b jetzt die Elektronen-Löcher-Streuung. Das Beweglichkeitsverhältnis ohne Berücksichtigung von EHS sei nun zu $b^{LI} = \mu_n^{LI}/\mu_p^{LI} = konst.$ definiert¹². Mit den Gleichungen (1.85), (1.89), (1.90) und der genannten Definition von b^{LI} ergibt sich für b :

$$b = \frac{\mu_n}{\mu_p} = b^{LI} \frac{\left(1 + \frac{N_\nu}{p}\right) \mu_{pn} + \frac{N_\nu}{p} \mu_p^{LI}}{\mu_{np} - \frac{N_\nu}{p} \mu_n^{LI}} \quad (2.45)$$

Weiterhin läßt sich für c ein ähnlicher Ausdruck finden, indem man die Gleichungen (1.86), (1.89), (1.90), und (1.91) verwendet. Man erhält dann

$$c = \frac{D_n}{D_p} = b^{LI} \left(1 + \frac{N_\nu}{p}\right) \frac{\mu_{pn} + 2 \mu_p^{LI}}{\mu_{np} + 2 \mu_n^{LI}} \quad (2.46)$$

Beide Verhältnisse sind also abhängig von der Löcherdichte p . Setzt man j_p gemäß (2.42), ohne sofort das E-Feld wie in den Fällen I und II zu eliminieren, in die

¹¹Eine ausführliche Herleitung findet sich in [48].

¹²Um Verwechslungen vorzubeugen sei darauf hingewiesen, daß es einen grundlegenden Unterschied in der Nomenklatur im Vergleich zur zugehörigen Literatur [46], [47], [48], [59] und [58] gibt: Dort werden alle Größen, die EHS beinhalten, mit einer Tilde versehen. Somit beziehen sich dort alle Größen ohne besondere Kennzeichnung, auf Größen, die nur von Gitter- und Störstellenstreuung beeinflusst werden. Hierfür wird in der vorliegenden Arbeit der obere Index LI (lattice, imputity) verwendet, wohingegen alle für das Modell relevanten Größen, im Fall III also mit EHS, keine besondere Kennzeichnung vorhanden ist. Beispielsweise beinhaltet D_p hier EHS, in den angegebenen Literaturstellen gerade eben nicht (dort \tilde{D}_p).

Kontinuitätsgleichung für Löcher (1.37) ein, erhält man nach einer etwas längeren Rechnung (vgl. [48])

$$\frac{\partial p}{\partial t} = D_p \frac{(b+c)p + b N_\nu}{(b+1) + b N_\nu} \frac{\partial^2 p}{\partial x^2} - \frac{b N_\nu}{(b+1)p + b N_\nu} E \mu_p \frac{\partial p}{\partial x} + \quad (2.47)$$

$$+ \left(\frac{\mu_n}{\mu_n^{LI}} - 1 \right) \left[\frac{D_p}{p} \frac{(b+c)p + b N_\nu}{(b+1) + b N_\nu} - \frac{D_A}{p} \right] \left(\frac{\partial p}{\partial x} \right)^2 - \frac{p}{\tau_h}. \quad (2.48)$$

Hierbei ist D_A die ambipolare Diffusionskonstante aus (2.35). In (2.35) muß b durch b^{LI} und D_p durch D_p^{LI} ersetzt werden, um formal mit der Nomenklatur für den Fall III kompatibel zu sein. Nun wird die Hochinjektionsbedingung $p \approx n \gg N_\nu$ angewendet. Für b und c gemäß (2.45) und (2.46) gilt dann unter Berücksichtigung von (1.94):

$$b = b^{LI} = konst. \quad \text{und} \quad c = b^{LI} \frac{\mu_{pn} + 2 \mu_p^{LI}}{\mu_{pn} + 2 \mu_n^{LI}} \quad (2.49)$$

Interessant ist dabei vor allem die Ortsunabhängigkeit von b , die sich aufgrund der Hochinjektion ergibt. Des weiteren kann mit der Definition von c , (2.34) umgeschrieben werden zu

$$D = D_p \frac{b+c}{b+1} \quad (2.50)$$

Nun kann b und c gemäß (2.49) in (2.44) eingesetzt werden. Das diesbezügliche Resultat, b und c gemäß (2.49) und die Hochinjektionsbedingung selbst kann anschließend auf (2.48) angewendet werden. Unter Berücksichtigung von (1.94), (1.96), (1.97) und (2.50) ergibt sich dann die zu Modellierungszwecken geeignete Kontinuitätsgleichung zur Beschreibung des Ladungsträgerverhaltens bei Hochinjektion. Sie lautet:

$$\begin{aligned} \frac{\partial p}{\partial t} = & \underbrace{D \frac{\partial^2 p}{\partial x^2}}_{\text{Diffusion}} - \underbrace{\frac{q_0}{k_B T} \frac{j_{ges}}{q_0 \mu_p^{LI} (b^{LI} + 1)} \frac{D N_\nu}{2 p^2} \frac{\partial p}{\partial x}}_{\text{Drift1}} + \\ & + \underbrace{\frac{b^{LI} - 1}{b^{LI} + 1} \frac{\mu_{pn}}{\mu_{pn} + \mu_n^{LI} + \mu_p^{LI}} \frac{D N_\nu}{2 p^2} \left(\frac{\partial p}{\partial x} \right)^2}_{\text{Drift2}} - \underbrace{\frac{p}{\tau_h}}_{R_{netto}} \end{aligned} \quad (2.51)$$

Ein Vergleich mit (2.36) läßt erkennen, daß nun zwei Terme, sogenannte Driftterme hinzugekommen sind. Die EHS-Beweglichkeit μ_{pn} geht hierbei in den Driftterm 2 ein. Damit stellt sich die Frage nach der Relevanz dieser neuen Kontinuitätsgleichung, da man offensichtlich auch mit (2.36) bzw. (2.41) befriedigende Simulationsergebnisse erzielen konnte. In [48] und [58] wird detailliert untersucht, bei welchen Kriterien die Driftterme relevant werden. Neben der Temperatur, der Stromdichte die die Driftzone durchfließt, steht vor allem das Verhältnis w_ν/L_A im Mittelpunkt der Betrachtungen. Ein sehr wichtiges Ergebnis ist, daß die Driftterme zu berücksichtigen sind, wenn für $w_\nu/L_A > 4,1$ gilt. Da die meisten Bauelemente kleinere

w_ν/L_A besitzen, genügt in der Regel (2.36). Doch gerade bei Hochspannungsdioden, die einigermaßen schnell schalten können sollen, ergeben sich w_ν/L_A -Verhältnisse, die in diese Größenordnung passen (Bsp.: $\tau_h = 5\mu s$ und $w_\nu = 500\mu m$ ergeben $w_\nu/L_A = 5, 13$).

Weiterhin kann die Hochinjektionsnäherung auf die Gleichung (2.44) zur Bestimmung des elektrischen Feldes angewendet werden. Berücksichtigt man zusätzlich (2.49), (1.94), (1.96) und (1.97) erhält man

$$E(x) = \frac{j_{ges}}{q_0 \mu_p^{LI} (b^{LI} + 1) p(x)} + \frac{\mathbf{j}_{ges}}{q_0 \mu_{pn} \mathbf{p}(\mathbf{x})} - \frac{k_B T}{q_0} \frac{b^{LI} - 1}{b^{LI} + 1} \frac{1}{p(x)} \frac{\partial p(x)}{\partial x} \quad (2.52)$$

Vergleicht man (2.52) mit (2.40) vom Fall I, erkennt man, daß die beiden Gleichungen bis auf den fett gedruckten Zusatzterm in (2.52), der die EHS-Beweglichkeit μ_{pn} enthält, völlig identisch sind. Im Gegensatz dazu, waren (2.40) und die entsprechende Gleichung für CCS (Fall II) nur formal identisch, denn b und μ_p enthielten die CCS-Beweglichkeit $\mu^C(p(x))$. Dies ist ein großer Unterschied zur EHS-Theorie, bei der eben ein Zusatzterm für die Erhöhung des Spannungsabfalls sorgt. Abweichend davon ging die CCS-Beweglichkeit μ^C im Fall II direkt in die Größen b und μ_p von (2.40) ein. Genau hier liegt der Schlüssel zu korrekten Simulationsergebnissen, wie im folgenden qualitativ in knapper Form diskutiert wird.

Vergleich der Modelle für Fall II und Fall III

An dieser Stelle soll in knapper Form, dargelegt werden, warum die Implementierung der oben und im Abschnitt 1.5.5 beschriebenen EHS-Theorie nach Mnatsakanov anstatt den herkömmlichen Theorien in den Modellen, sinnvoll und wichtig ist. Eine ausführliche Darstellung zu den folgenden Stichpunkten findet sich in [59] und in der Dissertation von Schlögl [58].

Für den Vergleich des Diodenmodells aus [59] mit dem Diodenmodell von Vogler [73] wurden an speziell präparierten Teststrukturen, Kennlinien und die Ladungsträgerkurven in der ν -Zone mittels eines aufwendigen optischen Meßverfahrens¹³, für verschiedene Temperaturen im Bereich 100–400K, gemessen. Das neue Diodenmodell unterscheidet sich vom Modell von Vogler [72] hauptsächlich in zwei Punkten:

- (1) Temperaturabhängigkeit der Injektionsfähigkeit der Randzonen: Statt eines einfachen Zusammenhangs $j_{n0,p0}(T) = j_{n0,p0}(300K) \cdot f(T)$ wird die in Abschnitt 2.2.3 beschriebene Methode angewendet.
- (2) Anwendung von EHS nach Mnatsakanov, d.h. die Kontinuitätsgleichung (2.51) und die Gleichung (2.52) zur Berechnung der Durchlaßspannung (Kennlinie)

¹³Nach dem Prinzip der Internen Laserdeflektion [18]

Dabei wurden beide Modelle für $T = 400K$ an Meßergebnisse angepaßt. Dabei waren in beiden Fällen die Unterschiede zur Realität entsprechend gering. Schon für Raumtemperatur jedoch gab es zwar noch tolerierbare, jedoch schon merkliche Abweichungen des Vogler-Modells. Für noch tiefere Temperaturen (bis $100K$) wurde der Spannungsabfall zwar wie erwartet größer, aber dem Betrag nach geringer als in der Messung. Dies impliziert, daß im Vergleich zum Experiment zu viele Ladungsträger in der Driftzone vorhanden sind. Die Simulation zeigt aber genau das Gegenteil, nämlich zuwenig Speicherladung. Dies ist ein Widerspruch, der der CCS-Theorie inhärent ist und somit nicht durch diese Theorie aufgelöst werden kann. Dagegen ermöglicht gerade der Zusatzterm in (2.52), einen höheren Spannungsabfall bei noch relativ viel Speicherladung. Folglich zeigten die Simulationen mit dem neuen Modell eine gute Übereinstimmung von Simulation und Messung, sowohl betreffs der Kennlinien als auch der Ladungsträgerprofile. Deshalb wird auch beim GCT-Modell die EHS-Theorie nach Mnatsakanov berücksichtigt. Abschließend seien die Gründe für die Abweichungen bei den Simulationen mit dem Vogler-Modell bei tiefen Temperaturen zusammengefaßt:

- Das Ladungsträgerniveau ist zu niedrig: Die Randwerte sind zu niedrig wegen der Verwendung einer nicht korrekten Beschreibung der Injektionsfähigkeit der Randzonen (s. oben (1)). Abweichungen im „Durchhang“ der Ladungsträgerfunktion ergeben sich aufgrund der Verwendung von (2.41) statt (2.51).
- Berechnung zu kleiner Spannungsabfälle wegen der Verwendung von (2.40) in Verbindung mit zu großen Werten der Löcherdichte (s. oben (2)).

2.3.2 Realisierung im Modell

Wie bereits erwähnt, muß die Kontinuitätsgleichung (2.51) numerisch gelöst werden. Dabei wird sich im folgenden ausschließlich auf die in den Modellen verwendeten numerischen Methoden beschränkt und auf eine allgemeine Betrachtung numerischer Verfahren verzichtet. Eine ausführliche Darstellung, Diskussion und Bewertung bezüglich Stabilität, etwaiger Konvergenzprobleme, Rechenzeitaufwand, usw., unterschiedlicher prinzipiell geeigneter numerischer Verfahren, findet sich u.a. in der Dissertation von Vogler [73].

Die ambipolare Diffusionsgleichung (2.36) ist eine lineare partielle Differentialgleichung vom parabolischen Typ:

$$\frac{\partial p(x, t)}{\partial t} = \alpha \frac{\partial^2 p(x, t)}{\partial x^2} + \beta p(x, t) \quad (2.53)$$

Bei solchen Gleichungen läßt sich eine Ortsdiskretisierung in n Abschnitte mit der Differenzenmethode durchführen. Mit einer geeigneten Zeitintegration läßt sich das

Randwertproblem im Zusammenhang mit (2.53), nach dem z.B. in [12] vorgeschlagenen Algorithmus lösen: Dabei wird durch die Ortsdiskretisierung aus der Differentialgleichung ein algebraische Gleichungssystem und die Löcherdichten $p(x_i)$ an den Stellen x_i mit $i = 1..n$, können für jeden Zeitpunkt t_j mittels Berechnung der Matrizen–Vektor–Gleichung unter Anwendung der Randwerte $p(x_0, t_j) = p_l(t_j)$ und $p(x_{n+1}, t_j) = p_r(t_j)$ direkt bestimmt werden. Allerdings wurde dann noch kein bewegliches Koordinatensystem berücksichtigt. Bei Anwendung von (2.51) sind die Verhältnisse wegen der Nichtlinearität (Quadrat der Ableitung von p nach x) hingegen komplizierter. Gleichung (2.51) kann formal zu

$$\frac{\partial p(x, t)}{\partial t} = f \left(p(x), \frac{\partial p(x, t)}{\partial x}, \frac{\partial^2 p(x, t)}{\partial x^2} \right) \quad (2.54)$$

geschrieben werden. Aufgrund der Nichtlinearität, muß mit dem Newton–Raphson–Verfahren linearisiert werden, d.h. anstatt einer Koeffizientenmatrix erhält man eine Jacobi-Matrix. Die $p(x_i, t_j)$ werden somit nicht direkt, sondern über ein iteratives Verfahren berechnet (s. [12]). Im folgenden wird genau dieser Ansatz zur Lösung von (2.51) beschrieben.

Örtliche Diskretisierung – Die Differenzenmethode

Die maßgebliche Kontinuitätsgleichung (2.51) kann zu

$$\frac{\partial p(x, t)}{\partial t} = \sum_{n=1}^4 \Gamma_n(p(x, t)) \quad (2.55)$$

geschrieben werden, d.h. die rechte Seite besitzt die vier Terme:

$$\Gamma_1(p(x, t)) = D \frac{\partial^2 p}{\partial x^2} \quad (2.56)$$

$$\Gamma_2(p(x, t)) = -\frac{q_0}{k_B T} \frac{j_{ges}}{q_0 \mu_p^{LI} (b^{LI} + 1)} \frac{D N_\nu}{2 p^2} \frac{\partial p}{\partial x} \quad (2.57)$$

$$\Gamma_3(p(x, t)) = \frac{b^{LI} - 1}{b^{LI} + 1} \frac{\mu_{pn}}{\mu_{pn} + \mu_n^{LI} + \mu_p^{LI}} \frac{D N_\nu}{2 p^2} \left(\frac{\partial p}{\partial x} \right)^2 \quad (2.58)$$

$$\Gamma_4(p(x, t)) = -\frac{p}{\tau_h} \quad (2.59)$$

Diese Ausdrücke hängen von $p(x, t)$ ab, d.h. sie sind noch immer ortskontinuierlich bzgl. x . Der Übergang zu ortsdiskreten Stützstellen x_i kann formal durch

$$p(x, t) \longrightarrow p(x_i, t) = p_i(t)$$

ausgedrückt werden. Zur Berechnung der Ableitungen an der Stelle x_i benützt man die benachbarten Stützstellen x_{i-1} und x_{i+1} und die entsprechenden Löcherdichten p_{i-1} und p_{i+1} . Insgesamt wird die Plasmazone in $n+2$ Stützstellen unterteilt, wobei $i=0$ und $i=n+1$ die Ränder mit den korrespondierenden Randwerten p_l und p_r bilden. Berechnet wird also $\vec{p} = (p_1, \dots, p_n)^T$.

Betrachtet man sich die Ladungsträgerverläufe, die sich im Falle der Diode als Lösung der Kontinuitätsgleichung (2.51) ergeben, erkennt man daß die Gradienten an den Randbereichen der Plasmakurve grundsätzlich vom Betrag größer sind als in der Mitte. Somit bietet es sich aus Genauigkeitsgründen an, mehr Stützstellen an den Rändern zu plazieren, als im Zentralbereich, wo auch mit einer geringen Stützstellenzahl eine hinreichend genaue Lösung gefunden werden kann. Dies bedeutet, daß die Abstände zwischen den Stützstellen im allgemeinen unterschiedlich sind. Zur Berechnung von p_i sind somit die Abstände

$$\begin{aligned} h_i &= x_i - x_{i-1} \\ h_{i+1} &= x_{i+1} - x_i \end{aligned} \quad (2.60)$$

relevant. Hiermit läßt sich die erste Ableitung der Löcherdichte nach dem Ort, an der Stelle x_i und zum Zeitpunkt t_j , nach der Methode der finiten Differenzen (siehe dazu u.a. [12]), unter Berücksichtigung der Tatsache, daß $h_i \neq h_{i+1}$ gilt, darstellen als

$$\frac{\partial}{\partial x}(p(x_i, t_j)) \approx \frac{p_{i+1}^j - p_{i-1}^j}{h_i + h_{i+1}}, \quad (2.61)$$

wobei mit p_i^j die Löcherdichte an der Stelle x_i zum Zeitpunkt t_j gemeint ist. Die zweite Ableitung lautet dementsprechend

$$\frac{\partial^2}{\partial x^2}(p(x_i, t_j)) \approx 4 \frac{h_i p_{i+1}^j - (h_i + h_{i+1}) p_i^j + h_{i+1} p_{i-1}^j}{(h_i + h_{i+1})(h_i^2 + h_{i+1}^2)} \quad (2.62)$$

Mithin lauten die Terme Γ_1 bis Γ_4 :

$$\Gamma_1(p_i^j) = 4 D_i^j \frac{h_i p_{i+1}^j - (h_i + h_{i+1}) p_i^j + h_{i+1} p_{i-1}^j}{(h_i + h_{i+1})(h_i^2 + h_{i+1}^2)} \quad (2.63)$$

$$\Gamma_2(p_i^j) = -\frac{q_0}{k_B T} \frac{j_{ges}}{q_0 \mu_p^{LI} (b^{LI} + 1)} \frac{D_i^j N_\nu}{2 p_i^{j^2}} \frac{p_{i+1}^j - p_{i-1}^j}{h_i + h_{i+1}} \quad (2.64)$$

$$\Gamma_3(p_i^j) = \frac{b^{LI} - 1}{b^{LI} + 1} \frac{\mu_{pn_i}^j}{\mu_{pn_i}^j + \mu_n^{LI} + \mu_p^{LI}} \frac{D_i^j N_\nu}{2 p_i^{j^2}} \left(\frac{p_{i+1}^j - p_{i-1}^j}{h_i + h_{i+1}} \right)^2 \quad (2.65)$$

$$\Gamma_4(p_i^j) = -\frac{p_i^j}{\tau_{h_i}^j} \quad (2.66)$$

Zeitintegration

Die linke Seite von (2.54), d.h. $\partial p/\partial t$, impliziert die Notwendigkeit einer Diskretisierung im Zeitbereich. Vogler [73] schlägt u.a. das sogenannte TR–BDF2–Verfahren vor. Dieses Integrationsverfahren liefert eine ausgezeichnete Genauigkeit bei globaler Stabilität. Allerdings benötigt es einen hohen Implementierungs– und Rechenzeitaufwand, da es sich um ein Zweischrittverfahren handelt. In dieser Hinsicht sind Einschrittverfahren vorteilhafter, aber auch etwas ungenauer. Vertretbar genau, bei gleichzeitig globaler Stabilität ist hierbei das Implizite Euler–Verfahren, das in den Modellen letztlich zur Anwendung kommt. Die (Teil)–Kontinuitätsgleichung für den Ortspunkt x_i , analog zu (2.54), lautet dann

$$\frac{p_i^j - p_i^{j-1}}{t_j - t_{j-1}} = f(p_{i-1}^j, p_i^j, p_{i+1}^j), \quad (2.67)$$

wobei die Ableitungen der rechten Seite durch die Ausdrücke (2.61) und (2.62) ersetzt werden. Charakteristisch für das implizite Euler–Verfahren ist, daß man zur Berechnung der Löcherdichte p_i^j zum Zeitpunkt t_j , die Löcherdichte p_i^{j-1} des vorangegangenen Zeitpunkts t_{j-1} benötigt. Alle Teil–Differentialgleichungen für die x_i mit $1 \leq i \leq n$ (bei $i = 0$ und $i = n + 1$ liegen die Randwerte!) können in Vektornotation zusammengefaßt werden zu

$$\frac{\partial \vec{p}}{\partial t} = \frac{\vec{p}^j - \vec{p}^{j-1}}{t_j - t_{j-1}} = f(\vec{p}^j). \quad (2.68)$$

Die Gleichungen (2.67) und (2.68) sind in dieser Form nur gültig, wenn die Stützstellen x_i sich im Zeitablauf nicht ändern. Diese Voraussetzung ist aber aufgrund der Verwendung eines beweglichen Koordinatensystems nicht erfüllt. Die dann nötigen Modifikationen werden im folgenden Unterabschnitt erläutert.

Bewegliches Koordinatensystem

Wie in Abschnitt 2.3.1 bereits erwähnt, erstreckt sich die räumliche Ausdehnung des Plasmagebiets nicht zwingend über die komplette ν –Zone. Vielmehr teilt sich die ν –Zone während des Abschaltens eines Bauelements in eine Plasma– und eine Raumladungszone auf. Dies bedeutet, daß die Grenzen der Plamazonen, also x_l und x_r , im Zeitablauf wandern. Damit ändern sich aber automatisch auch die $n + 2$ Stützstellen. Bild 2.6 zeigt eine solche Situation für eine (beliebige) Diode, die anfangs eingeschaltet ist und das Ladungsträgerprofil $p(t = 0)$ aufweist. Die Kurve $p(t > 0)$ zeigt die Löcherverteilung zu irgendeinem Zeitpunkt während des Abschaltvorgangs. Man sieht, daß sich die x_i (exemplarisch x_4 im Bild) für beide Löcherverteilungen jeweils an verschiedenen Orten befinden.

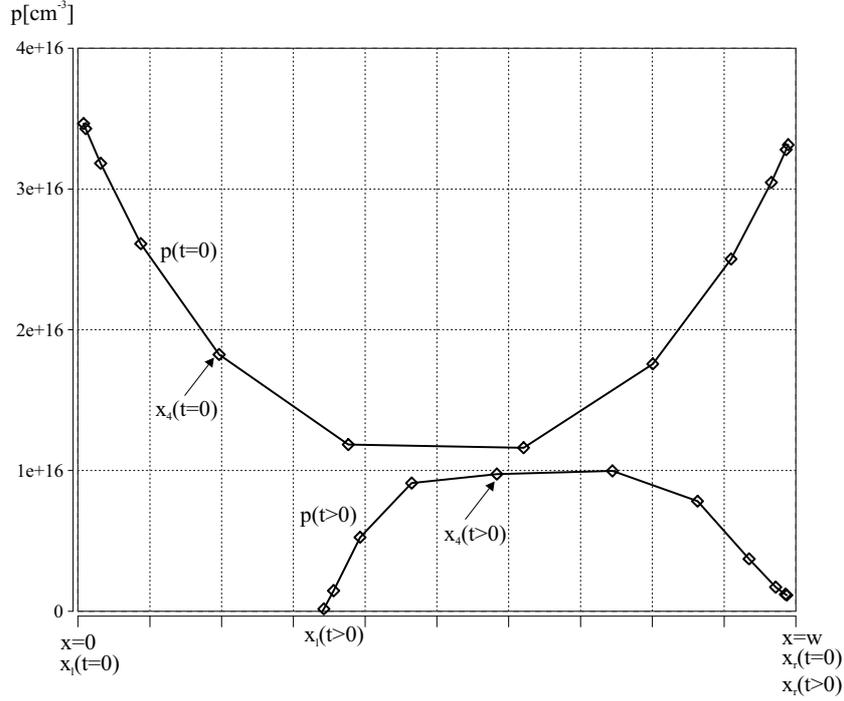


Bild 2.6: Löcherplasma einer Diode zu Beginn ($t = 0$) und während des Abschaltens ($t > 0$); Simulation mit 12 ($n = 10$) Stützstellen.

Diese Zeitabhängigkeit der Stützstellen x_i kann formal durch

$$x_i \longrightarrow x_i(t)$$

dargestellt werden. Die Stützstellen $x_i(t)$ lassen sich nun allgemein zu

$$x_i(t) = x_l(t) + \eta_i (x_r(t) - x_l(t)) \quad \text{mit} \quad \eta_i > \eta_{i-1}, \quad \eta_0 = 0, \quad \eta_{n+1} = 1 \quad (2.69)$$

berechnen. Hierbei legt η_i die Lage der Stützstellen bezüglich der Weite der Plasmazone $x_r(t) - x_l(t)$ fest. In den Modellen findet hierfür die Gleichung

$$\eta_i = \begin{cases} \eta_k & = 0.5 \cdot \left(\frac{2k}{n+1} \right)^{\alpha_S} \\ \eta_{n+1-k} & = 1 - 0.5 \cdot \left(\frac{2k}{n+1} \right)^{\alpha_S} \end{cases} \quad : \quad 0 < k \leq \frac{n+1}{2} \quad (2.70)$$

Anwendung. Hierbei werden mit der oberen Formel von (2.70) die Stützstellen bis zur Hälfte der Plasmazonenweite berechnet. Aus Symmetriegründen werden mittels der unteren Formel die Stützstellen für die andere Hälfte gleich mitberechnet. Je größer der Modellparameter α_S ist, desto mehr konzentrieren sich die Stützstellen an den Rändern der Plasmazone, d.h. die Abstände zwischen den Stützstellen in

Randnähe werden kleiner. Dies geschieht allerdings auf Kosten der Abstände im zentralen Bereich, die dann immer weiter werden. Für $\alpha_S = 1$ besitzen die Stützstellen den gleichen Abstand, d.h. $h_{i+1} = h_i$. Als günstig hat sich beispielsweise $\alpha_S = 3$ erwiesen.

Da offensichtlich $x_i^j \neq x_i^{j-1}$ gilt, berechnet man mittels des Zählers des Terms auf der linken Seite von (2.67), die Differenz der Löcherdichten im Zeitablauf, an zwei verschiedenen Positionen! Das Wandern von x_i im Zeitablauf $t_{j-1} \rightarrow t_j$ muß somit durch einen Zusatzterm erfaßt werden. Formal kann also

$$\underbrace{\frac{\partial p(x_m, t)}{\partial t}}_{\text{physikalisch}} = \underbrace{\frac{\partial p(x_i, t)}{\partial t}}_{\text{numerisch}} + \underbrace{\frac{\partial p(x_i, t)}{\partial x} \frac{dx}{dt}}_{W_i^j} \quad (2.71)$$

geschrieben werden. Dabei handelt sich bei x_m im linken Term von (2.71) um eine gedachte, im Zeitablauf ortsfeste Stützstelle. Somit repräsentiert dieser Term die in (2.67) tatsächlich auftretende Ableitung der Löcherdichte p nach der Zeit, an eben dieser fixen Stelle x_m . Für t_{j-1} gilt $m = i$, wohingegen zum Zeitpunkt t_j , $m \neq i$ ist. Der erste Term auf der rechten Seite vom Gleichheitszeichen in (2.71) entspricht der jetzt „falschen“ Zeitableitung von (2.67), die aufgrund der Abweichung von x_i^j zu x_m , um den Wanderungsterm W_i^j ergänzt werden muß, um korrekte Berechnungen zu erhalten (vgl. [73]). Folglich verändert sich (2.67) zu:

$$\frac{p_i^j - p_i^{j-1}}{t_j - t_{j-1}} + W_i^j = f(p_{i-1}^j, p_i^j, p_{i+1}^j), \quad (2.72)$$

Damit erhält man wiederum einen Satz von n Gleichungen, die in Vektornotation zusammengefaßt werden können

$$\frac{\partial \vec{p}}{\partial t} = \frac{\vec{p}^j - \vec{p}^{j-1}}{t_j - t_{j-1}} = \vec{f}(\vec{p}^j) - \vec{W}^j = \vec{f}^*(\vec{p}^j), \quad (2.73)$$

wobei $\vec{p}^j = (p_1, \dots, p_n)^T$ und $\vec{f}^* = (f_1^*, \dots, f_n^*)^T$ ist. Der Wanderungsterm an der Stelle x_i kann mit (2.69) und (2.61) und unter Berücksichtigung der Tatsache, daß jetzt h_i und h_{i+1} ebenfalls zeitabhängige Funktionen sind, geschrieben werden zu

$$W_i^j = \frac{\partial p(x_i, t)}{\partial x} \frac{dx}{dt} = \frac{p_{i+1}^j - p_{i-1}^j}{h_i^j + h_{i+1}^j} \left(\eta_i \frac{dx_r}{dt} + (1 - \eta_i) \frac{dx_l}{dt} \right) \quad (2.74)$$

Damit ist die Diskretisierung von (2.54) vollständig. Die Gleichung muß nun nach den p_i^j , d.h. nach \vec{p}^j aufgelöst werden. Im Falle der linearen ambipolaren Diffusionsgleichung (2.36) bzw. (2.53) wäre die Angelegenheit relativ einfach, denn statt (2.73) erhielte man eine Gleichung der Form $\mathbf{A} \vec{p}^j + \vec{b} = \mathbf{0}$, die einfach nach \vec{p}^j aufzulösen wäre. Wegen der Nichtlinearitäten in (2.51) bzw. (2.54) ergibt sich nach der Diskretisierung ein ebenfalls nichtlinearer Satz von algebraischen Gleichungen der Form (2.73). Die zur Lösung notwendige Linearisierung wird im folgenden Unterabschnitt erläutert.

Das Lösungsverfahren

Man kann einen Funktionsvektor $\vec{F}(\vec{p}^j)$ einführen, der folgendermaßen definiert ist

$$\vec{F}(\vec{p}^j) = \vec{f}^*(\vec{p}^j) - \frac{\partial \vec{p}}{\partial t} = \vec{0} \quad (2.75)$$

wobei $\vec{F} = (F_1, \dots, F_n)^T$ wegen (2.73) gleich dem Nullvektor sein muß. Somit entspricht die Lösung der diskretisierten Differentialgleichung (2.73) dem Nullstellenvektor (Wurzeln) dieses Funktionsvektors $\vec{F}(\vec{p}^j)$. Aufgrund seiner Robustheit und schnellen Konvergenz wird hier das Newton–Raphson–Verfahren verwendet (vgl. [10], [12]). Die k-te Näherung für \vec{p}^j , d.h. \vec{p}_k^j , erhält man über die Iterationsvorschrift

$$\vec{p}_k^j = \vec{\delta} + \vec{p}_{k-1}^j, \quad (2.76)$$

also mit Hilfe der k-1-ten Näherung. Den Vektor $\vec{\delta}$ erhält man durch Lösen des Gleichungssystems

$$\mathbf{J}(\vec{p}_{k-1}^j) \cdot \vec{\delta} = -F(\vec{p}_{k-1}^j). \quad (2.77)$$

$\mathbf{J}(\vec{p}_{k-1}^j)$ ist die Jacobi–Matrix, die der ersten Ableitung von $\vec{F}(\vec{p}^j)$ entspricht. Die Koeffizienten dieser $n \times n$ -Matrix lauten allgemein (z : Zeilenindex; s : Spaltenindex):

$$J_{zs} = \frac{\partial F_z}{\partial p_s}(\vec{p}_{k-1}^j) \quad (2.78)$$

Betrachtet man sich die Struktur einer Einzelgleichung $F_i(\vec{p}_{k-1}^j)$ des Gleichungssystems $F(\vec{p}_{k-1}^j)$ (vgl. (2.72)), so erkennt man, daß jeweils nur drei Koeffizienten einer Zeile von $\mathbf{J}(\vec{p}_{k-1}^j)$ ungleich Null sind¹⁴. Relevant sind also in jeder Zeile die drei Koeffizienten

$$\begin{aligned} J_{i,i-1} &= \frac{\partial F_i}{\partial p_{i-1}}(p_{i-1,k-1}^j, p_{i,k-1}^j, p_{i+1,k-1}^j) \\ J_{i,i} &= \frac{\partial F_i}{\partial p_i}(p_{i-1,k-1}^j, p_{i,k-1}^j, p_{i+1,k-1}^j) \\ J_{i,i+1} &= \frac{\partial F_i}{\partial p_{i+1}}(p_{i-1,k-1}^j, p_{i,k-1}^j, p_{i+1,k-1}^j) \end{aligned} \quad (2.79)$$

Konkret hat $\mathbf{J}(\vec{p}_{k-1}^j)$ damit folgendes Aussehen:

¹⁴Bei der ersten und der n -ten Zeile sind sogar nur zwei Koeffizienten verschieden von Null.

$$\begin{aligned}
& \mathbf{J}(\vec{p}_{k-1}^j) = \\
& = \begin{pmatrix}
J_{11} & J_{12} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
J_{21} & J_{22} & J_{23} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
0 & J_{32} & J_{32} & J_{34} & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
0 & 0 & J_{43} & J_{44} & J_{45} & 0 & \dots & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & & & & & \vdots \\
\vdots & \vdots & \vdots & & \ddots & \ddots & \ddots & & & & \vdots \\
\vdots & \vdots & \vdots & & & \ddots & \ddots & \ddots & & & \vdots \\
\vdots & \vdots & \vdots & & & & \ddots & \ddots & \ddots & & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & & \ddots & \ddots & \ddots & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & J_{n-1n-2} & J_{n-1n-1} & J_{n-1n} \\
0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & J_{nn-1} & J_{nn}
\end{pmatrix} \quad (2.80)
\end{aligned}$$

Wie aus (2.80) zu erkennen ist, besitzt $J(\vec{p}_{k-1}^j)$ Tridiagonalgestalt, d.h. nur die Haupt- und die angrenzenden Nebendiagonalen sind besetzt. Daran orientiert sich schließlich die Wahl des Lösungsverfahrens zur Lösung von (2.77). Für das wohlbekannte Gauß'sche Eliminationsverfahren benötigt man $(n^3 + 3n^2 - n)/3$ Multiplikationen und $(2n^3 + 3n^2 - 5n)/6$ Additionen. Aufgrund der vorliegenden tridiagonalen Struktur bietet sich das Crout'sche Reduktionsverfahren mit nur $5n - 4$ notwendigen Multiplikationen und $3n - 3$ Additionen an (vgl. [12]). Dadurch läßt sich eine wesentliche Erhöhung der Rechengeschwindigkeit realisieren und wird deshalb in den Modellen verwendet.

Damit ist ein (beliebiger) Iterationsvorgang beschrieben, von denen i.a. mehrere, ausgehend von einer Anfangslösung (z.B. der Lösung der Arbeitspunktanalyse oder der Lösung des vorangegangenen Zeitschritts) vonnöten sind, um eine Lösung für \vec{p}^j zu finden. Diese Lösung ist erreicht, wenn ein bestimmtes Abbruchkriterium erfüllt ist. So kann der Iterationsvorgang enden, wenn die maximale relative Änderung vom $k-1$ -ten zum k -ten Iterationsschritt, formal ausgedrückt durch $\max\{(\delta_{1,k}/p_{1,k-1}, \dots, \delta_{i,k}/p_{i,k-1}, \dots, \delta_{n,k}/p_{n,k-1})^T\}$, eine bestimmte Grenze unterschreitet, also sich die Lösung durch jeden weiteren Iterationsschritt kaum noch ändert.

Berechnung des statischen Spannungsabfalls

Ist die Löcherverteilung für einen bestimmten Zeitschritt (oder dem Arbeitspunkt) nach dem oben beschriebenen Verfahren berechnet, kann mit dessen Hilfe das elektrische Feld entlang der Plasmakurve mittels (2.52) berechnet werden. Mit der

bekanntes Ortsdiskretisierung kann (2.52) geschrieben werden zu:

$$E_i^j = \frac{j_{ges}}{q_0 \mu_p^{LI} (b^{LI} + 1) p_i^j} + \frac{j_{ges}}{q_0 \mu_{pn}^j p_i^j} - \frac{k_B T}{q_0} \frac{b^{LI} - 1}{b^{LI} + 1} \frac{1}{p_i^j} \frac{p_{i+1}^j - p_{i-1}^j}{h_i + h_{i+1}} \quad (2.81)$$

Allgemein kann der Spannungsabfall durch die Integration dieser elektrischen Feldstärke, d.h. mit

$$U = \int_{x_l}^{x_r} E(x) dx \quad (2.82)$$

bestimmt werden. Für jedes der $n+1$ Segmente $[x_{i-1}^j, x_i^j]$, kann durch Flächenbildung unterhalb der durch Geradenstücke verbundenen E_i die Teilspannung u_i nach dem sogenannten Trapezverfahren [12] berechnet werden:

$$U_i = E_{i-1}^j h_i^j + \frac{1}{2} h_i^j (E_i^j - E_{i-1}^j) = \frac{1}{2} h_i^j (E_i^j + E_{i-1}^j), \quad (2.83)$$

wobei an der Stelle x_l , $E_0 = 0$ und $U_0 = 0$ gilt. Diese U_i müssen nun über alle i , d.h. $i = [1, \dots, n+1]$ aufsummiert werden. Somit errechnet sich die Spannung U_ν zu:

$$U_\nu = \sum_{i=1}^{n+1} U_i \quad (2.84)$$

Berechnung der Sperrspannung

Vom Spannungsabfall über der Plasmazone (positives Vorzeichen, wenige Volt) zu unterscheiden ist die Spannungsaufnahme im Sperrbetrieb bzw. während des induktiven Abschaltens (negatives Vorzeichen, je nach Bauelement bis zu mehreren kV). Zu ihrer Berechnung benötigt man die Ausdehnung der Raumladungszone und die Funktion des elektrischen Feldes in Abhängigkeit vom Ort. Diese Funktion erhält man direkt aus der Poisson-Gleichung (1.30), die für eindimensionale Verhältnisse

$$\frac{dE}{dx} = \frac{\rho}{\epsilon_0 \epsilon_r} \quad (2.85)$$

lautet. Bild 2.7 links, zeigt die Verhältnisse für irgendeinen Zeitpunkt während des induktiven Abschaltens. Dabei verringert sich die Ladungsträgerdichte in der Driftzone aufgrund zweier Mechanismen:

1. Löcherextraktion (Elektronenextraktion) ausgehend von der negativ (positiv) gepolten Anode (Kathode) mit anschließendem Eindringen der Raumladungszone (RLZ) in die entstehenden „Freiräume“ $[0, x_l]$ und $[x_r, w_\nu]$ ¹⁵.
2. Durch Rekombination wie in Abschnitt 1.4 beschrieben.

¹⁵In Bild 2.7 ist der Einfachheit halber nur eine linksseitige RLZ eingezeichnet, d.h. es gilt $p_r \neq 0$ und somit $x_r = w_\nu$. Im allgemeinen gibt es bei Dioden aber zusätzlich eine rechtsseitige RLZ

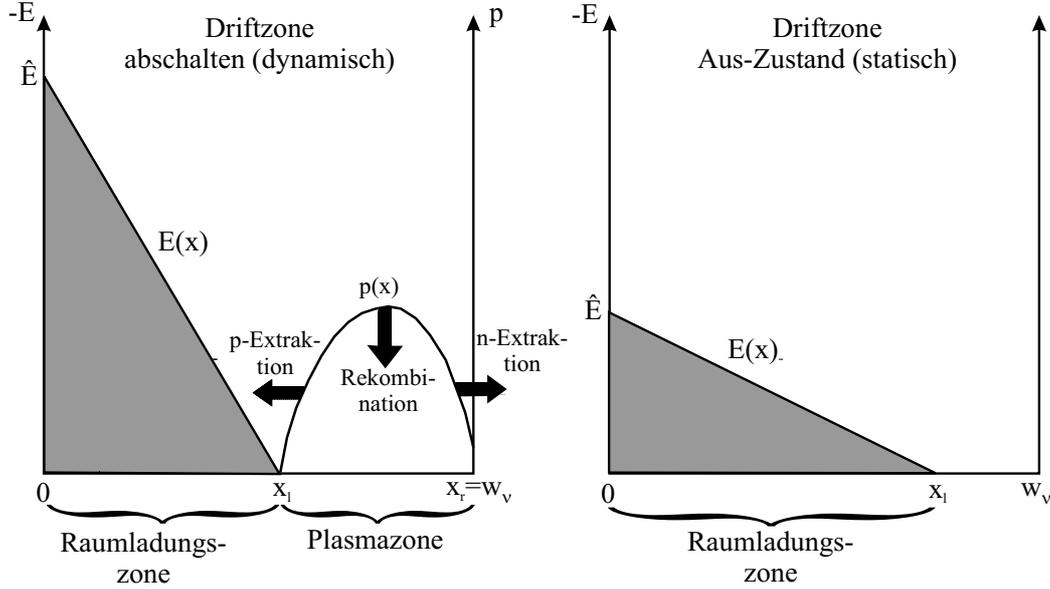


Bild 2.7: Aufteilung der Driftzone in eine Raumladungszone und eine Plasmazone während des induktiven Abschaltens einer Diode kurz vor Eintritt der rechtsseitigen RLZ (links). Verhältnisse in der Driftzone im Aus-Zustand (rechts)

In der Raumladungszone entspricht die Dichte fester Ladungen wegen der vollständigen Ionisierung der Störstellen genau N_v^+ . Im Falle des Abschaltens kommt aufgrund des z.T. noch hohen Sperrstroms eine zusätzliche in (2.85) zu berücksichtigende Ladung hinzu. Die Ladung in der Raumladungszone wird also zu $\rho = q_0 N_{eff}$ angenommen. Hierbei heißt N_{eff} effektive Dotierung, die sich wegen $j_n = q_0 n v_n$ und $j_p = q_0 p v_p$ (s. Abschnitt 1.5.2) zu

$$N_{eff} = N_v + p - n = N_v + \frac{j_p}{q_0 v_p} - \frac{j_n}{q_0 v_n} \quad (2.86)$$

ergibt. Da im gesamte Bereich der linksseitigen RLZ $x_{RLZ,l} \in [0, x_l]$, $p(0) = p(x_{RLZ,l}) = p(x_l) = 0$ gilt, ergibt (2.9) $j_n(0) = 0$. Damit ergibt sich gemäß (1.34) $j_p(0) \approx j_{ges}$, d.h. die Verschiebestromdichte j_{dis} (siehe unten (2.90)) wird bei der Berechnung von N_{eff} vernachlässigt. Da es in der RLZ keine freien Ladungsträger und damit folglich auch keine Ladungsträgergradienten gibt, kann für $j_p(x_{RLZ,l})$

$$j_p(x_{RLZ,l}) = j_p(0) = j_{ges} = const \quad (2.87)$$

geschrieben und nebst $j_n = 0$ zur Ermittlung von $N_{eff,l} = const$ in (2.86) eingesetzt werden. Rechtsseitig gilt für die RLZ $x_{RLZ,r} \in]x_r, w_v]$ Entsprechendes, d.h. $j_p(w_v) = j_p(x_r) = 0$ und $j_n(x_r) = j_n(x_{RLZ,r}) \approx j_{ges}$. Setzt man dies in (2.86) ein, erhält man $N_{eff,r}$, wobei die rechtsseitige RLZ trivialerweise nur dann auftreten kann, wenn $j_n/(q_0 v_n) > N_v^+$ gilt, weil nur dann ein E-Feld-Maximum $\hat{E}(w_v)$

mit negativem Vorzeichen entsteht, das von der Kathode abgebaut werden kann. Damit leuchtet auch unmittelbar ein, daß immer $N_{eff,l} > N_{eff,r}$ gilt. Man kann somit beobachten, daß die linksseitige RLZ, die vom pn -Übergang ausgeht, das Abschaltverhalten dominiert. Für die Driftgeschwindigkeiten $v_{p,n}$ kann man in ebenfalls guter Näherung die Sättigungsgeschwindigkeiten $v_{satn,p}$ annehmen. Damit umgeht man vor allem das Problem der Abhängigkeit von $v_{n,p}$ vom elektrischen Feld gemäß (1.65). Dies würde den Rechenaufwand erheblich steigern, weil man anstelle von (2.88) eine implizite Gleichung zur Berechnung des elektrischen Feldes erhielte.

Die Integration von (2.85) von $x = 0$ nach $x = x_l$ liefert die Funktion für das elektrische Feld in der linksseitigen RLZ:

$$E(x) = \frac{q_0 N_{eff}}{\epsilon_0 \epsilon_r} x - \hat{E} \quad \text{mit} \quad \hat{E} = \frac{q_0 N_{eff}}{\epsilon_0 \epsilon_r} x_l \quad (2.88)$$

Es handelt sich also um eine lineare Funktion und man spricht deshalb auch von einem E-Feld-Dreieck¹⁶. Die Fläche unter diesem Dreieck, d.h. die Integration von (2.88) über $x_{RLZ,l}$ ergibt die (linksseitige) Sperrspannung:

$$U_S = -\frac{1}{2} \frac{q_0 N_{eff}}{\epsilon_0 \epsilon_r} x_l^2 \quad (2.89)$$

Die entsprechenden Ausdrücke für die rechtsseitige RLZ erhält man entsprechend durch die einfache (E-Feld) bzw. doppelte Integration (Spannung) von (2.85) von $x = x_r$ nach $x = w_\nu$ mit $N_{eff,r}$.

Leistungsdioden werden meist für $j_{ges,max} \approx 100 A/cm^2$ ausgelegt. Die Rückstrom-(dichte)spitzen beim Abschalten liegen in ähnlichen Größenordnungen. Dies ergibt ein N_{eff} von ca. $6 \cdot 10^{13} cm^{-3}$, also gleich oder oft sogar größer als N_ν . Dies bedeutet, daß \hat{E} dynamisch schnell, sehr große Werte (große Überspannungen!), die die kritische Feldstärke von ca. $2 \cdot 10^5 V/cm$ übersteigen, annehmen kann. Folglich spielt der dynamische Avalanche-Effekt gerade bei Hochspannungs- bzw. Hochleistungsbauelementen eine große Rolle. Ist das Bauelement dann abgeschaltet (s. Bild 2.7 rechts), also im statischen Aus-Zustand, gilt wegen $j_p \approx 0$, $N_{eff,l} = N_\nu$ in (2.88) und (2.89). Das E-Feld-Dreieck verläuft demnach flacher als im Fall des Abschaltens. Eine rechtsseitige Raumladungszone gibt es dann auf keinen Fall, d.h. sie verschwindet in dem Moment, in die oben genannte Bedingung $j_n/(q_0 v_n) > N_\nu^+$ durch die Verringerung von $j_n(t)$ verletzt wird.

Da sich während des Abschaltens, das E-Feld-Dreieck im Zeitverlauf verändert, gilt $\partial E/\partial t \neq 0$. Dies bedeutet, daß bei den Stromdichtegleichungen die Maxwellsche Verschiebungsstromdichte berücksichtigt werden muß. Sie ist definiert zu

¹⁶Wenn das elektrische Feld die Kathode erreicht, spricht man vom sogenannten Punch Through, der je nach Bauelement verboten oder erlaubt, bzw. wie beim IGCT sogar erwünscht ist. In diesem Fall geht das Dreieck in ein Trapez über. Dazu mehr in Abschnitt 3.3.4

$j_{dis} = \partial D / \partial t$. Da wegen der Poisson-Gleichung $\partial D / \partial x = \rho$ und $\rho = const.$ gilt, läßt sich für j_{dis} an den Grenzen des Plasmagebiets x_l und x_r

$$j_{dis}(x_l) = q_0 N_{eff}(x_l) \frac{\partial x_l}{\partial t} \quad \text{und} \quad j_{dis}(x_r) = q_0 N_{eff}(x_r) \frac{\partial x_r}{\partial t} \quad (2.90)$$

schreiben.

2.4 Modellierung weiterer Eigenschaften

2.4.1 Modellierung exponentieller Dotierprofile

Zur Modellierung der Injektionsfähigkeit der Randzonen wird bisher auf konstante (effektive) Dotierungen N_A und N_D mit abrupten Übergängen zur Driftzone zurückgegriffen. Dies liefert zum einen genügend genaue Rechenergebnisse und zum anderen sind dafür die Gleichungen zur Bestimmung der beiden Sperrsättigungsstromdichten recht einfach (s. Abschnitt 2.2). Allerdings handelt man sich bei Nichtberücksichtigung realistischer Dotierprofile, Ungenauigkeiten in Bezug auf die effektiven Plasmazonenweiten $w_{eff}(p)$ ein, wie man in Bild 2.8 erkennen kann.

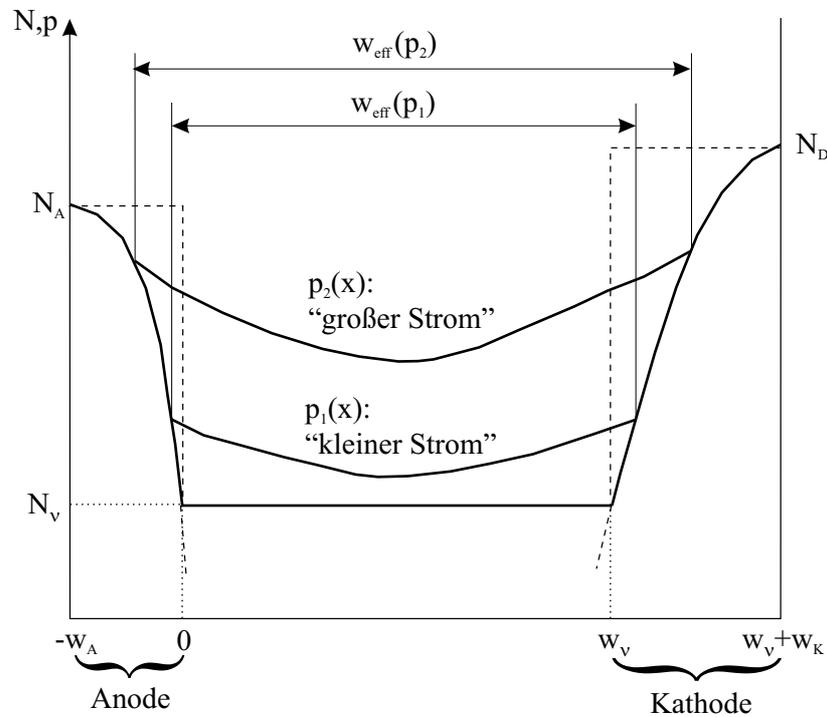


Bild 2.8: realistisches Dotierprofil einer Leistungsdiode; zwei Löcherverteilungen für zwei verschiedene Ströme in Vorwärtspolung

Das Bild zeigt qualitativ die Lage von zwei unterschiedlichen Löcherverteilungen $p_1(x)$ und $p_2(x)$ bei zwei verschiedenen großen Stromdichten. Betrachtet wird also der statische Durchlaßfall. Wie man in Bild 2.8 erkennt, sind die effektiven Plasmazonenweiten $w_{eff}(p)$ im Durchlaßfall stets größer als die technologische Driftzonenweite w_ν . Je höher die Abweichung $w_{eff} - w_\nu$ bezogen auf w_ν ist, desto größer wird der Fehler im Simulationsergebnis sein. Deshalb werden realistische Dotierprofile in den Modellen berücksichtigt und zwar ohne auf die Vorzüge der konstanten Dotierdichten mit abrupten Übergängen verzichten zu müssen! Dieser scheinbare Widerspruch läßt sich auflösen, indem man im Modell (a) die Berechnung der Injektionsfähigkeit der Randzonen und (b) die Bestimmung der effektiven Plasmazonenweite voneinander entkoppelt. Demnach werden für (a) abrupte Übergänge und konstante (effektive!) Dotierdichten und die entsprechenden Gleichungen aus Abschnitt 2.2 verwendet. Dagegen kann unabhängig davon, ein exponentielles Dotierprofil (Gauss-Profil) mit einer, i.d.R. von den effektiven Dotierdichten aus (a) abweichenden maximalen Akzeptorendichte \hat{N}_l (Donatorendichte \hat{N}_r) und den Dotierweiten w_l (w_r), in dem MAST-Driftzonenmodul (s. Bild 2.2) definiert werden. Vorgegeben wird also eine Weite w , aus der sich, die für das Sperrverhalten maßgebliche technologische Driftzonenweite w_ν , mittels $w_\nu = w - w_l - w_r$ berechnen läßt. Ebenfalls kann man die Plasmazonenweite $w_{eff}(p)$ bestimmen, wie weiter unten gezeigt wird. Zunächst sei auf einen weiteren Vorteil der genannten Entkopplung hingewiesen: Sie ermöglicht die Definition einer zweiten tiefen Dotierschicht, die sich an eine Randzone anschließt. Somit kann mit der Dotierprofil-Beschreibung im Driftzonenmodul

- I. das realistische Profil der Randzonen selbst
- II. **oder** eine zweite, an die Randzone anschließende, i.a. schwächer dotierte Schicht vom gleichen Typ (sogenannte tiefe Diffusion)

beschrieben werden. Diese beiden Möglichkeiten verdeutlicht exemplarisch Bild 2.9. Hierbei ist x_{jl} (x_{jr}) technologisch gesehen der genaue Ort des $p^+\nu^-$ -Übergangs (νn^+ -Übergangs). An die Anode mit konstantem Profil mit der Akzeptorendichte N_A und der Ausdehnung $[-w_A, 0]$ schließt sich eine zweite tiefe Diffusionsschicht mit der maximalen p-Dotierdichte $\hat{N}_l < N_A$ und der Ausdehnung $[0, x_{jl}]$ an (Variante II.). Auf der rechten Seite fällt die Diffusionsschicht hingegen mit der Kathode zusammen, d.h. x_{jr} entspricht w_ν und w entspricht $w_\nu + w_K$ (Variante I). Wohlgermerkt, die Berechnung des Injektionsverhaltens der Randzonen, geschieht in jedem Fall in den Randzonenmodulen gemäß Abschnitt 2.2. An dieser Stelle werden nur die Auswirkungen der Dotierprofile auf die Ladungsträgerkurve in der Driftzone, konkret auf die Ausdehnung der Plasmazone (w_{eff}) betrachtet. Modellierungstechnisch betrifft dies also nur das Driftzonenmodul.

Tiefe Schichten werden technologisch überwiegend durch Diffusionsvorgänge hergestellt, wohingegen gerade bei dünnen Randzonen auch Ionenimplantation zum

Einsatz kommt. In beiden Fällen erhält man ein Gauss-Profil, wobei im Falle der Diffusion von einer erschöpflichen Quelle, d.h. die Dichte des Dotierstoffs in der Quelle wird im Zeitablauf selbst geringer, ausgegangen wird¹⁷ [69].

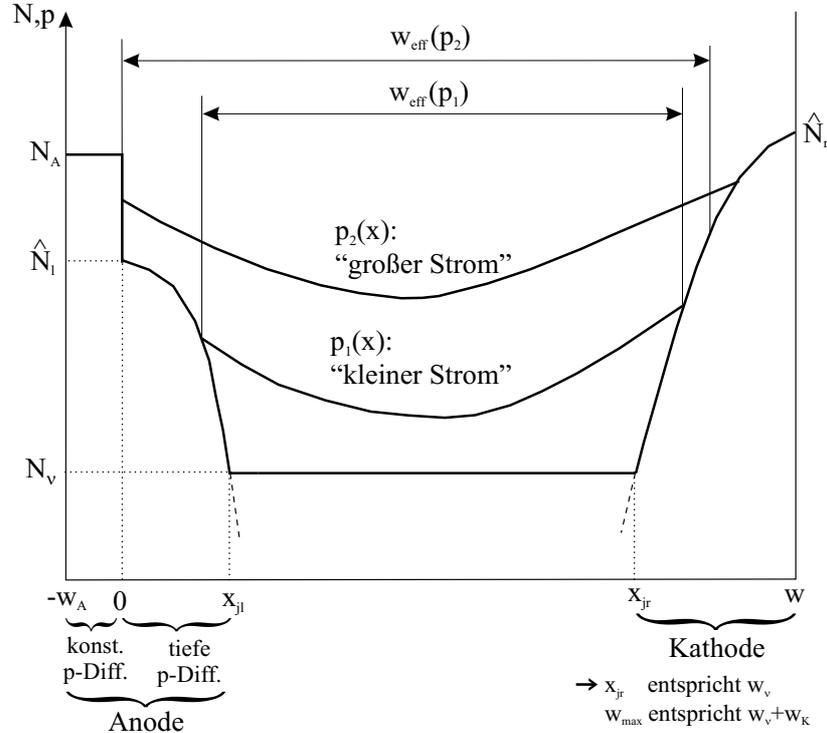


Bild 2.9: Beispiel für Dotierprofile im Driftzonenmodul: zusätzliche zweite Anodenschicht (linksseitig) und ein Kathodenprofil (rechtsseitig)

Mathematisch läßt sich ein Diffusionsprofil folgendermaßen ausdrücken:

$$N(x) = \hat{N} \exp\left(-\frac{x^2}{\sigma^2}\right) \quad (2.91)$$

Bezüglich der Bestimmung der Plasmaweite w_{eff} , bzw. deren Grenzen x_l und x_r , gibt es, abhängig davon, welche randzonenseitigen Löcherdichten $p_{l,r}$ sich ergeben, folgende Fälle zu unterscheiden:

- $x_l \geq x_{jl}$ bzw. $x_r \leq x_{jr}$ für $p_{l,r} = 0$
- $0 < x_l < x_{jl}$ bzw. $x_{jr} < x_r < w$ für $0 < p_{l,r} < \hat{N}_{l,r}$

Falls es sich um eine tiefe Diffusion handelt, kommt sinnvollerweise

- $x_l = 0$ bzw. $x_r = w$ für $p_{l,r} \geq \hat{N}_{l,r}$

¹⁷Wird dagegen mit einer unerschöpflichen Quelle, d.h. die Dotierstoffkonzentration in der Quelle verändert sich während des Dotiervorgangs nicht, dotiert, erhält man ein sogenanntes Erfc-Profil (Erfc= complementary error-function; Def. s. [10]). Dieser Fall wird hier nicht weiter behandelt.

hinzu. Die Fälle a) und c) sind bereits bekannt. Hier interessiert der Fall b), wenn also $p_{l,r}$ quasi „auf dem Profil liegt“. Per Definition soll $N_l(x_{jl}) = N_r(x_{jr}) = N_\nu$ gelten (vgl. Bild 2.9). Damit ergibt sich aus (2.91) für die linke Seite:

$$N_\nu = \hat{N}_l \exp\left(-\frac{x_{jl}^2}{\sigma_l^2}\right) \quad \text{und} \quad N_\nu = \hat{N}_r \exp\left(-\frac{(w - x_{jr})^2}{\sigma_r^2}\right) \quad (2.92)$$

Damit lassen sich die $\sigma_{l,r}$ zu

$$\sigma_l = \frac{x_{jl}}{\sqrt{\ln\left(\frac{\hat{N}_l}{N_\nu}\right)}} \quad \text{und} \quad \sigma_r = \frac{w - x_{jr}}{\sqrt{\ln\left(\frac{\hat{N}_r}{N_\nu}\right)}} \quad (2.93)$$

berechnen. Nun kann (2.93) in (2.91) eingesetzt werden und mit $p_l = N_l(x_l)$ und $p_r = N_r(x_r)$ ergibt sich daraus für x_l und x_r

$$x_l = \sigma_l \sqrt{\ln\left(\frac{\hat{N}_l}{p_l}\right)} \quad \text{und} \quad x_r = w - \sigma_r \sqrt{\ln\left(\frac{\hat{N}_r}{p_r}\right)} \quad (2.94)$$

Damit kann schließlich die Plasmaweite zu $w_{eff}(p) = x_r - x_l$ bestimmt werden.

2.4.2 Modellierung von Lebensdauerprofilen

In der homogen dotierten Driftzone, kann man von der Annahme einer über x konstanten SRH-Hochinjektionslebensdauer τ_h gemäß (1.46) ausgehen. Eine hohe Lebensdauer geht mit einer großen Menge Speicherladung Q_S einher¹⁸, was durchaus erwünscht sein kann, weil dann die Spannung und damit die statischen Verluste gering sind. Andererseits impliziert eine hohe Speicherladung, große Schaltverluste, da jeweils die Ladungsmenge beim Schalten entweder eingebracht, oder ausgeräumt werden muß. Der Bauelementehersteller hat es also mit einem Optimierungsgesetz zu tun, d.h. die optimalen Designparameter werden von der gewünschten Anwendung bestimmt. So gibt es neben Gleichrichterioden, die tatsächlich sehr hohe Lebensdauern von $\tau_h \geq 30\mu s$ haben, schnelle Dioden mit weitaus niedrigeren τ_h -Werten, um z.B. Umrichter mit hohen Taktfrequenzen zu realisieren.

Oftmals sollte die Hochinjektionslebensdauer ein bestimmtes Maß nicht unterschreiten, um die statischen Verluste in vertretbaren Grenzen zu halten. Dieses Minimum

¹⁸Die einfache Formel $Q_{S,n,p} = I \cdot \tau_{n,p}$ für die jeweilige Minoritätsspeicherladung in der Anode bzw. Kathode, gilt zwar für die Driftzone bei Leistungsdioden so nicht, gibt aber zumindest einen Hinweis auf die Proportionalität von Q_S zu τ_h

kann aber bereits zu groß sein, was die dynamischen Anforderungen betrifft. Dieses Problem kann gelöst werden, indem man die Lebensdauer nur lokal absenkt. Beispielsweise wirkt es sich bei Dioden positiv auf den Reverse-Recovery-Effekt (Abschaltverhalten) aus, wenn τ_h lokal in der Nähe der Anode, reduziert wird (s. Bild 2.10). Denn damit wird die Löcherdichte anodenseitig reduziert und folglich kann die Raumladungszone von der Anode her schneller eindringen. Damit erhält man eine verringerte Rückstromspitze einerseits und ein weiches Abklingen des Rückstroms andererseits, bzw. kann dadurch ein Abreißen des Stroms unter Umständen überhaupt erst verhindert werden.

Die Reduktion der Hochinjektionslebensdauer wird durch Vorhandensein zusätzlicher Rekombinationszentren realisiert (vgl. Abschnitt 1.4.2). Diese zusätzlichen Traps werden aber nicht direkt durch Dotierung eingebracht, sondern indirekt mit Bestrahlung erzeugt. Dafür stehen Gamma-, Elektronen-, bzw. Protonenbestrahlung zur Verfügung. Gemäß den zugehörigen physikalischen Zusammenhängen erhält man einen gaußförmigen τ -Profilverlauf wie in Bild 2.10 eingezeichnet.

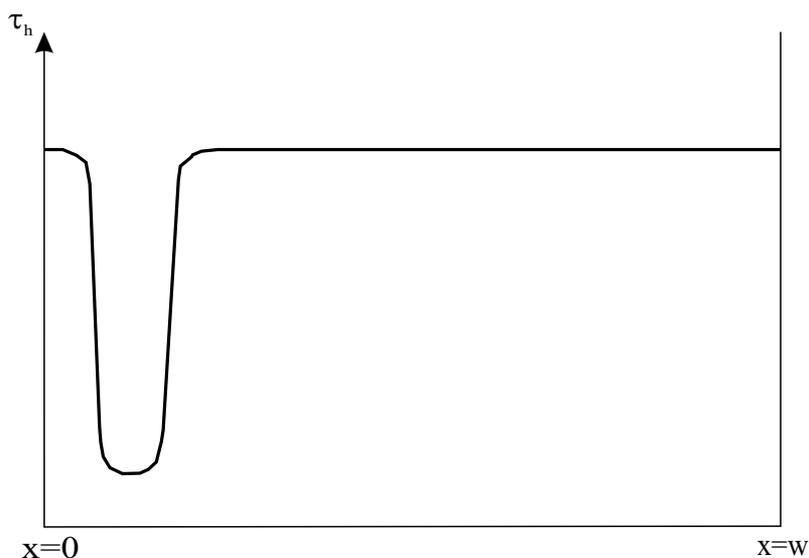


Bild 2.10: Lebensdauerprofil in der ν -Zone nahe der Anode.

In den Modellen ist eine abschnittsweise Definition von τ_h vorgesehen. Dabei läßt sich die Anzahl m der Stützstellen x_k und die zugehörige Lebensdauer $a_j \cdot \tau_h$ mit $j = [0, \dots, m-1]$ und $0 < a_j \leq 1$ frei wählen. Dabei bedeutet $m=0$, daß es nur eine konstante Lebensdauer gibt. Je größer die Anzahl m ist, desto feiner läßt sich ein Lebensdauertopf nachbilden. Ein Beispiel mit $m=12$ zeigt Bild 2.11. Die waagrechten Teilstücke im Bild entsprechen den Lebensdauern, die jeweils in den entsprechenden Abschnitten $[x_k, x_{k+1}]$ gelten. Jede der variablen Stützstellen x_k^j der Plasmazone (s. Abschnitt 2.3.2) kann dabei so einem Abschnitt $[x_k, x_{k+1}]$ zugeordnet werden.

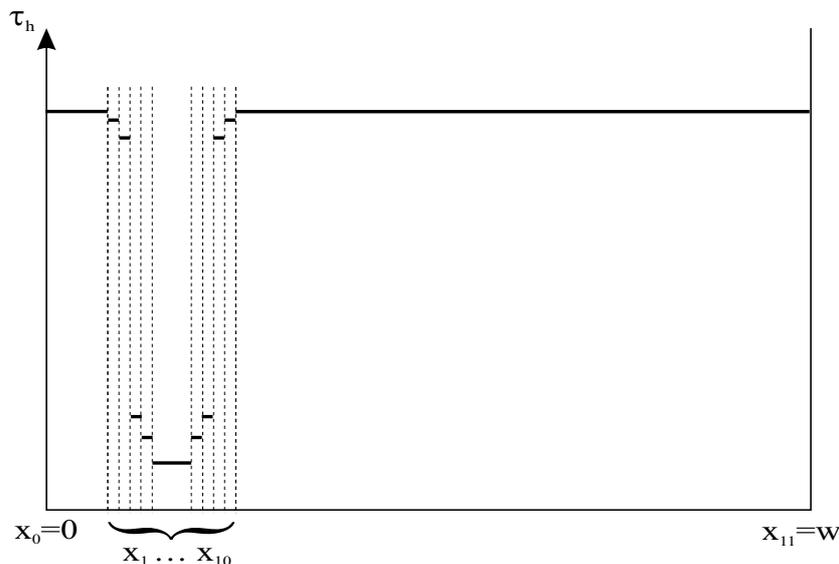


Bild 2.11: Abschnittsweise definierte Lebensdauer: Beispiel mit $m=12$.

2.4.3 Der dynamische Lawineneffekt

Gerade bei Hochvoltbauelementen, spielt der dynamische Lawineneffekt eine große Rolle und darf bei der Modellierung keinesfalls vernachlässigt werden. In Abschnitt 1.4.5 ist die Theorie zur Beschreibung des Lawineneffekts mit den Formeln (1.54) bis (1.59) zusammengefaßt. Die Implementierung dieser Gleichungen hat allerdings zur Folge, daß die Simulationen bei Einsetzen der Lawinenmultiplikation spürbar „gebremst“ werden. Meist enden diese Simulationen aufgrund von Konvergenzproblemen vorzeitig. Deshalb wird hier eine Fitmormel verwendet, die bei ausgezeichneter Konvergenz, gute Ergebnisse liefert, ohne die Simulationsdauern spürbar zu erhöhen:

$$M_{AV} = 1 + \frac{(\hat{E} - E_{crit})^2}{k_{AV}}. \quad (2.95)$$

Dabei ist M_{AV} der Multiplikationsfaktor. Bei E_{crit} handelt es sich um die kritische Feldstärke, bei der Lawinenmultiplikation einsetzt. Der Parameter k_{AV} gibt an, wie stark der Lawineneffekt wirkt. Beide Parameter können vom Anwender geändert werden. Dabei hat sich für $E_{crit} \approx 1,9 \cdot 10^5 V/cm$ und $k_{AV} \approx 1,5 \cdot 10^9 cm^2/V^2$ als sinnvoll erwiesen. In der RLZ kann j_p als konstant und j_n zu Null betrachtet werden (s.o.). Der Multiplikationseffekt läßt sich somit formal durch

$$j_p \longrightarrow M_{AV} \cdot j_p \quad (2.96)$$

darstellen. Der dynamische Lawineneffekt begrenzt sich selbst, da die entstehenden Elektronen–Loch–Paare in Summe elektrisch neutral sind und demnach nicht

zu einer weiteren Erhöhung der effektiven Dotierdichte gemäß (2.86) beitragen. Letztlich wird dadurch \hat{E} nicht wesentlich weiter erhöht. Hält das Bauelement der erhöhten Verlustleistung aufgrund des Lawineneffekts stand, wird es sicher abschalten. Bild 2.12 zeigt die die simulierten Abschalttransienten I_A (Anodenstrom) und U_{AK} (Anoden-Katodenspannung), den sogenannten Reverse-Recovery-Effekt.

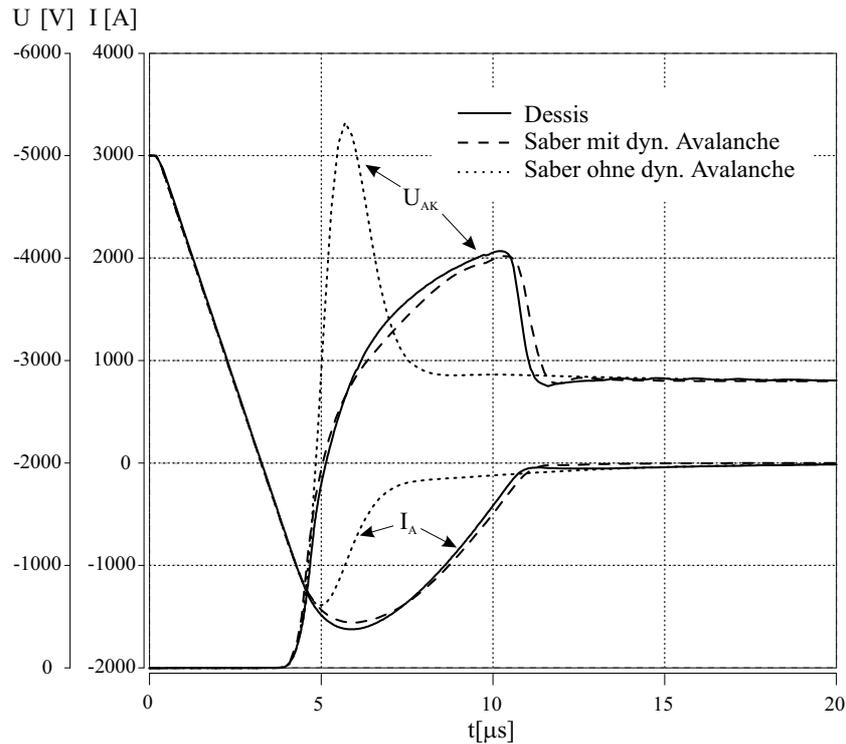


Bild 2.12: *Transienten eines Abschaltvorgangs einer Leistungsdiode, bei dem der dynamische Lawineneffekt wirkt: Bauelementesimulation (Dessis) vs. Schaltungssimulation (Saber) mit dem Modell.*

Für diese Simulation wurde die Struktur einer fiktiven Hochleistungsdiode in den Bauelementesimulator Dessis [30] eingegeben. Die zugehörigen Transienten beim Abschalten von $3kA$ gegen $2,8kV$ (s. Bild 2.12 durchgezogene Linie) wurden mit der Schaltungssimulation (Saber [2]), die sich auf das physikalische Modell stützt, verglichen (gestrichelte Linie in Bild 2.12). Man sieht in diesem Bild, daß mit (2.95), sehr gute Ergebnisse erzielt werden. Ergänzt wird Bild 2.12 durch eine Saber-Simulation ohne Berücksichtigung des Lawineneffekts (punktirierte Linie in Bild 2.12). Man sieht, daß durch den Lawineneffekt der Rückwärtsstrom nach Überschreiten der Rückstromspitze langsamer abklingt. Gleichzeitig steigt die Sperrspannung zwar langsamer, verbleibt jedoch länger oberhalb der statischen Sperrspannung. Insgesamt ergibt sich daher eine Erhöhung der Verlustleistungsdichte wie Bild 2.13

zeigt. Abschließend zeigt Bild 2.14 den zeitlichen Verlauf des maximale elektrische Feldes $\hat{E}(x_1)$, wie er sich für die beiden Saber-Simulationen ergibt.

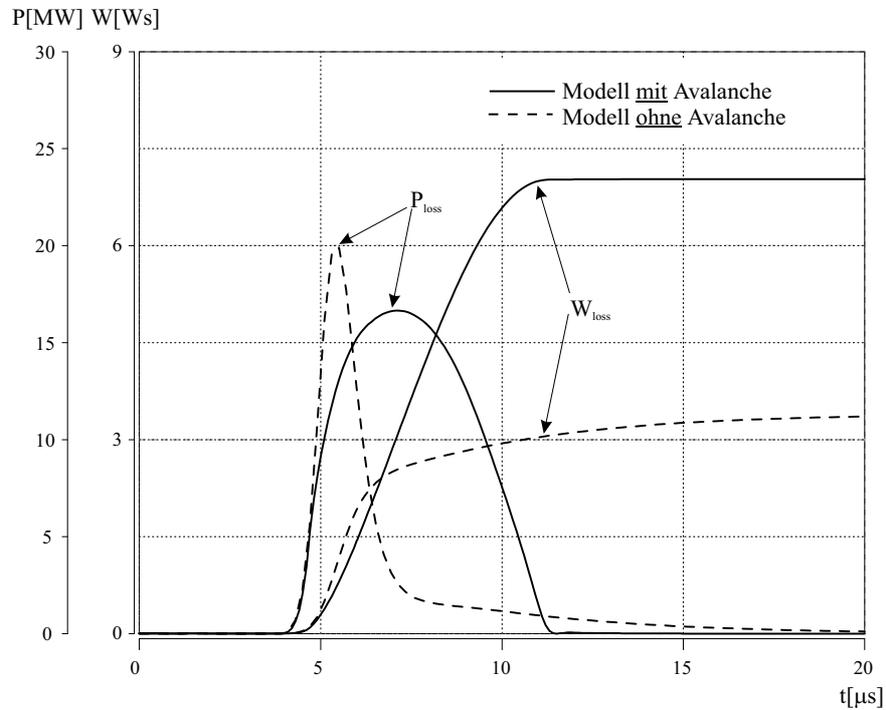


Bild 2.13: Verlustleistung bzw. Verlustenergie zu Bild 2.12

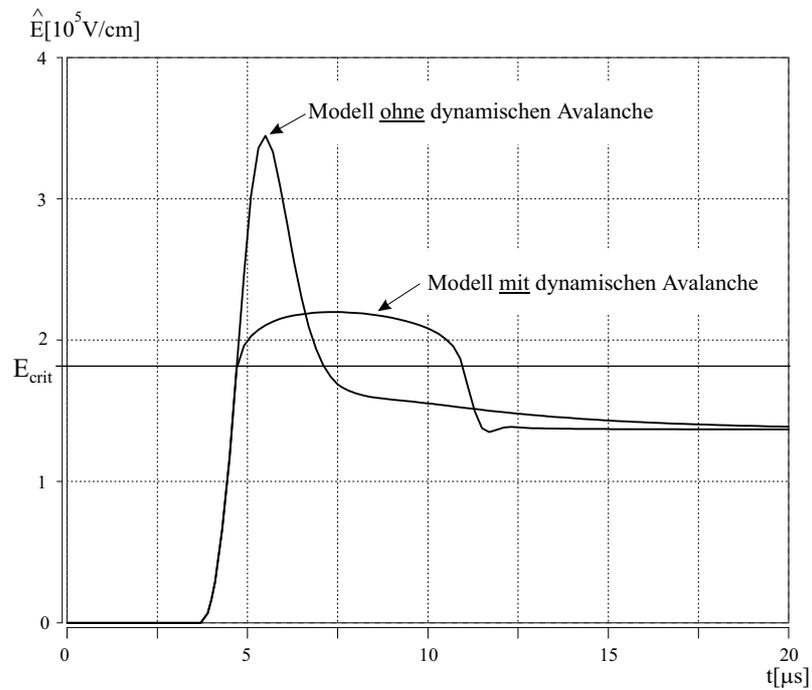


Bild 2.14: Maximales elektrisches Feld am $p^+\nu$ -Übergang zu Bild 2.12

Berücksichtigt man den Lawineneffekt im Modell nicht, ergibt sich eine unrealistisch hohe maximale elektrische Feldstärke \hat{E} von ca. $3,4 \cdot 10^5 V/cm$. Dagegen zeigt die Rechnung mit dem Avalanche-Effekt, daß dieses \hat{E} nur wenig größer als $2 \cdot 10^5 V/cm$ ist.

2.5 Das Gesamtmodell

Bislang wurden die Module einzeln beschrieben. Nun wird der Blick auf die Verkopplung der Module und auf die zugehörigen Übergabegrößen an den Schnittstellen gerichtet. Die Schnittstellen im gewählten Koordinatensystem befinden sich an den Driftzongrenzen $x = 0$ und $x = w_\nu$ (vgl. Bild 2.2). Es ist sofort einleuchtend, daß für einen beliebigen Ortspunkt $x^* \in [0, w_\nu]$ entweder eine Überschußladung $p(x^*) > 0$, d.h. x^* liegt im Plasmagebiet, oder ein E-Feld $E(x^*) < 0$, d.h. x^* liegt in der Raumladungszone, existiert. Gleiches gilt selbstverständlich an den Grenzen $x = 0$ und $x = w_\nu$. Dieses „Entweder oder“ bedeutet, daß sich die beiden Größen p_l (p_r) und x_l (x_r) modelltechnisch in eine auf die Driftzondotierung N_ν normierte Systemvariable m_l (m_r) überführen lassen, die dann aus den Strombilanzen bestimmt werden kann. Durch eine Fallunterscheidung wird festgelegt, ob sich bei $x = 0$ ($x = w_\nu$) Überschußladungsträger p_l (p_r) befinden, oder ob sich eine Raumladungszone der Länge x_l (x_r) anschließt (vgl. [73]). Im folgenden wird das Vorgehen anhand der linken Seite der Driftzone beschrieben, wobei die Beschreibung des exponentiellen Dotierprofils gemäß Abschnitt 2.4.1 eingeschlossen ist:

	Werte von m_l	Werte von p_l, x_l	Wertebereich
Fall 1	$m_l \geq \frac{\hat{N}_l}{N_\nu} + 1$	$p_l = (m_l - 1) N_\nu$ $x_l = 0$	$\hat{N}_l \leq p_l < N_A$
Fall 2	$1 < m_l < \frac{\hat{N}_l}{N_\nu} + 1$	$p_l = (m_l - 1) N_\nu$ x_l gemäß (2.93) & (2.94)	$0 < p_l < \hat{N}_l$ $0 < x_l < x_{jl}$
Fall 3	$0 < m_l \leq 1$	$p_l = 0$ $x_l = w_\nu - m_l \cdot w_\nu + x_{jl}$	$x_{jl} \leq x_l < w_\nu$

(2.97)

Im Fall 1 hat man an der Stelle $x = x_l = 0$ eine injizierte Löcherdichte in Höhe von p_l , wobei p_l größer als die maximale Dotierung \hat{N}_l der diffundierten Schicht ist. Im Fall 2 ist p_l kleiner als \hat{N}_l , d.h. x_l befindet sich entsprechend der Plasmazonenmodulation (vgl. Bild 2.9) zwischen 0 und der Grenze Diffusionszone/ ν -Zone bei $x = x_{jl}$. Fall 3 behandelt den Fall, daß eine Raumladungszone vorhanden ist, d.h. $p_l = 0$. Eine analoge Tabelle zu (2.97) kann auch für die rechte Seite erstellt werden.

Als nächstes soll die Verkopplung der Stromdichtegleichungen, die letztlich im Zusammenspiel mit den Systemvariablen $m_{l,r}$, die Randwerte $p_{l,r}$ bestimmt, beschrieben werden. Vereinfachend sei dazu angenommen, daß kein exponentielles Dotierprofil existieren soll, d.h. $x_{jl,r} = 0$. Für die randzonenseitigen Minoritätsträgerstromdichten an den Grenzen $x = 0$ und $x = w_\nu$ gelten (2.9) und (2.11). Solange $x_l = 0$ und $x_r = w_\nu$ und damit $p_{l,r} > 0$ ist, gilt dann $j_{n,p} > 0$. Beim Vorhandensein einer RLZ, gilt wegen $p_{l,r} = 0$, $j_n(0) = j_n(x_l) = 0$ bzw. $j_p(w_\nu) = j_p(x_r) = 0$. Während des Abschaltens müssen allerdings die Verschiebestromdichten gemäß (2.90) berücksichtigt werden, woraus dann $j_{n,p}(x_{l,r}) = j_{dis}(x_{l,r})$ folgt. Man kann also allgemein

$$j_n(x_l) = j_{n0} \left(\frac{p_l}{n_i} \right)^2 + j_{dis}(x_l) \quad (2.98)$$

$$j_p(x_r) = j_{p0} \left(\frac{p_r}{n_i} \right)^2 + j_{dis}(x_r) \quad (2.99)$$

schreiben. Die ν -zonenseitigen Randbedingungen lassen sich durch die ambipolare Stromdichtegleichungen (2.31) und (2.32) darstellen zu

$$j_p(x_l) = \frac{1}{b+1} j_{ges} - q_0 D \frac{\partial p}{\partial x}(x_l) \quad (2.100)$$

$$j_n(x_r) = \frac{b}{b+1} j_{ges} + q_0 D \frac{\partial p}{\partial x}(x_r). \quad (2.101)$$

Durch Addition von (2.98) mit (2.100) an der Stelle x_l und (2.99) mit (2.101) bei x_r , erhält man zwei Gleichungen für die Gesamtstromdichte j_{ges} . Da j_{ges} auch schon in (2.100) und (2.101) vorkommen, kann man die beiden Gleichungen nach j_{ges} umstellen und man bekommt:

$$j_{ges} = \left(\frac{b+1}{b} \right) \left[j_{n0} \left(\frac{p_l}{n_i} \right)^2 + j_{dis}(x_l) - q_0 D \frac{\partial p}{\partial x}(x_l) \right] \quad (2.102)$$

$$j_{ges} = (b+1) \left[j_{p0} \left(\frac{p_r}{n_i} \right)^2 + j_{dis}(x_r) + q_0 D \frac{\partial p}{\partial x}(x_r) \right]. \quad (2.103)$$

Wie in Abschnitt 2.3 beschrieben, läßt sich die Kontinuitätsgleichung (2.51) nur numerisch lösen. An den Rändern können die beiden Ladungsträgergradienten fol-

gendermaßen genähert werden (vgl. Abschnitt 2.3.2)

$$\frac{\partial p}{\partial x}(x_l) \approx \frac{p_1 - p_l}{h_1} \quad (2.104)$$

$$\frac{\partial p}{\partial x}(x_r) \approx \frac{p_r - p_n}{h_{n+1}}, \quad (2.105)$$

wobei wegen (2.60) $h_1 = x_1 - x_l$ und $h_{n+1} = x_r - x_n$ gilt. Nun hängt p_1 natürlich von seinem Nachbarkonzentration p_2 und dieser wiederum von p_3 usw., ab. Letztlich hängt also p_1 auch von p_r und dito p_n von p_l ab, so daß für (2.102) und (2.103) verkürzt

$$j_{ges} = f_1(p_l, p_r) \quad (2.106)$$

$$j_{ges} = f_2(p_l, p_r) \quad (2.107)$$

geschrieben werden kann. Man hat letztlich zwei Funktionen mit zwei Unbekannten. Im Modell wird mit Hilfe von $m_{l,r}$ und dem implementierten numerischen Verfahren für jeden Zeitschritt solange iteriert, bis (2.106) und (2.107) erfüllt sind.

Schließlich muß noch der Gesamtspannungsabfall U_F einer Diode in Durchlaßrichtung berechnet werden. Sie setzt sich zusammen aus den beiden Übergangsspannungen plus dem Spannungsabfall in der Driftzone längs des Plasmas gemäß (2.84), also

$$U_F = U_{jl} + U_{jr} + U_\nu \quad (2.108)$$

Für die Übergangsspannungen gelten die beiden Gleichungen

$$U_{jl} = \frac{k_B T}{q_0} \ln \left(\frac{p_l N_\nu}{n_i^2} \right) \quad (2.109)$$

$$U_{jr} = \frac{k_B T}{q_0} \ln \left(\frac{p_r + N_\nu}{N_\nu} \right), \quad (2.110)$$

die zwingend aus (1.116) und (1.119) folgen.

3 Das IGCT-Modell

3.1 Thyristor, GTO, IGCT

Die Leistungshalter Thyristor, GTO und auch der IGCT (Integrated Gate Commutated Thyristor) basieren alle auf der gleichen $p^+n^-pn^+$ -Vierschichtstruktur, wie sie Bild 3.1 zeigt.

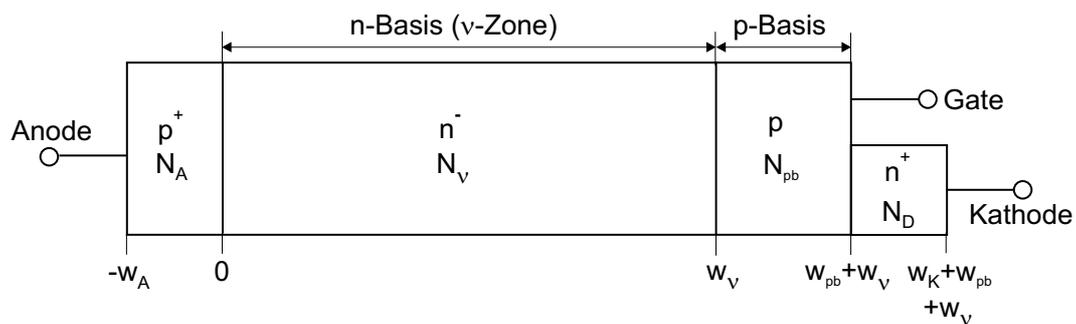


Bild 3.1: $p^+n^-pn^+$ -Thyristorstruktur

Zum Verständnis der prinzipiellen Funktion dieser Struktur kann das sogenannte Zwei-Transistor-Ersatzschaltbild, bestehend aus einem p^+n^-p -Transistor und einem n^+pn^- -Transistor (vgl. [49], [55], [61], [68]) herangezogen werden. Damit können aber nur die statischen Kennlinien beschrieben werden. Um das dynamische Verhalten beim Ein- bzw. Ausschalten verstehen zu können, muß das Verhalten der Speicherladungen in den schwach dotierten Zonen, insbesondere in der n^- -Basis, ähnlich wie in Kapitel 2, mathematisch modelliert werden. Die n^- -Basis bildet dabei die mittlere Zone des p^+n^-p -Transistors, des sogenannten Wide-Base-Transistors (WBT), der für das Schaltverhalten der Vierschichtstruktur von entscheidender Bedeutung ist.

Zunächst soll aber kurz darauf eingegangen werden, worin sich die drei genannten Vierschichtbauelemente eigentlich unterscheiden. So grenzen sich GTO und GCT vom Thyristor dadurch ab, daß sie über den Gateanschluß nicht nur ein- sondern auch ausgeschaltet werden können. Während der GTO dabei eine Ausschaltentlastung in Form eines RC(D)-Snubbers benötigt, kann auf diesen beim GCT verzichtet

werden (sogenannter Snubberless-Betrieb). Dies setzt allerdings hartes Abschalten mit $dI_G/dt > 1000A/\mu s$ voraus, d.h. die (parasitäre) Gateinduktivität muß sehr klein sein. Da diese Aufgabe keineswegs trivial ist, liefert der Hersteller die Gatesteuerung gleich mit und der GCT wird damit zum IGCT [28]. Im folgenden werden die Zusammenhänge und Gründe, die zu diesen unterschiedlichen Abschaltverhalten der genannten Vierschichtstrukturen führen und die damit zusammenhängenden Begriffe wie hartes Schalten, Filamentierung, etc., kurz erläutert.

Die Gesamtweite eines $4,5kV$ -GTO/GCT kann, je nachdem, ob eine Pufferschicht vorhanden ist oder nicht, leicht $500\mu m$ und mehr erreichen. Dabei beträgt der Durchmesser der Anoden- bzw. der Gate/Kathodenfläche einige cm , oder mit anderen Worten: Anders als Bild 3.1 impliziert, ist die laterale Ausdehnung bei weitem größer als die Ausdehnung in Stromflußrichtung¹. Eine zumindest qualitativ realitätsnähere Vorstellung eines Vierschichtbauelements zeigt Bild 3.2

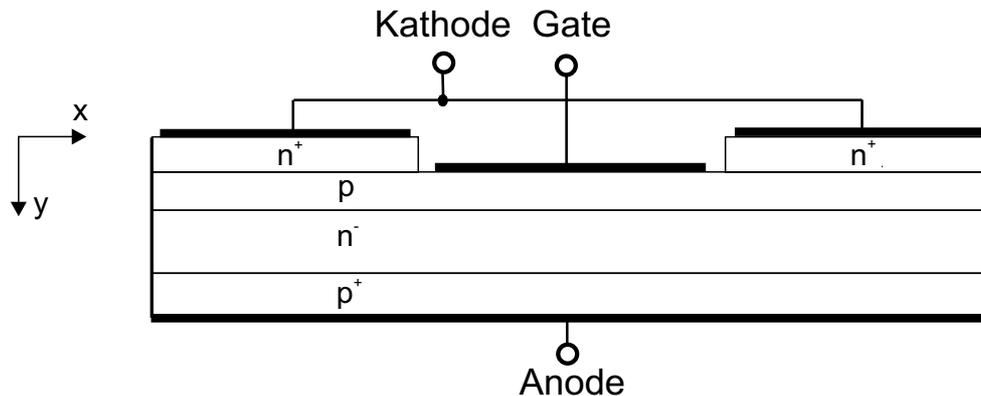


Bild 3.2: „realitätsnähere“ $p^+n^-p^-n^+$ -Thyristorstruktur

Man kann sich nun sehr gut vorstellen, daß das Gate bei Anlegen des negativen Abschaltesignals, die Überschußlöcherdichte aufgrund der kürzeren Reichweite, schneller aus der n^- -Basis in x -Richtung, als aus der p -Basis unterhalb der n^+ -Kathodenschicht in y -Richtung, absaugen wird. Daraus folgt, daß bestimmte Bereich der Kathode immer noch Elektronen injizieren, während in der Driftzone (n^- -Basis), bereits eine ausgedehnte Raumladungszone existiert, oder kurz: Der Gate/Kathoden-Übergang ist noch nicht vollständig gesperrt, während das Vierschichtbauelement bereits eine hohe Blockierspannung über den mittleren n^-p^- -Übergang aufgenommen hat. Dabei wird der Bereich unterhalb der Kathode, über den noch Strom fließt und der demnach Elektronen injiziert, immer mehr von der sich lateral ausdehnenden Raumladungszone eingeschnürt. Die noch verbleibenden dünnen Plasmagebiete, über die der u.U. noch hohe Kathodenstrom fließen muß,

¹Über die Auswirkungen auf die eindimensionale mathematische Modellierung wird noch zu reden sein.

heißen Filamente. Ihr Widerstand ist sehr hoch, was im Zusammenhang mit der Kathodenstromdichte zu einer lokal hohen Verlustleistungsdichte und damit verbunden, zu einer hohen lokalen thermischen Belastung führt, die das Bauelement schließlich zerstört. Dieser laterale Effekt heißt aufgrund der Einschnürung des Ladungsträgergebiets zu einzelnen leitenden dünnen Fäden, Filamentierung (siehe dazu u.a. [21]). Ein Abschalten über das Gate kann erreicht werden, wenn die Vierschichtelemente mit nicht nur einem Gate-Gebiet und einem Kathoden-Gebiet, sondern mit viele kleineren, sich abwechselnde Gate- und Kathodeninseln versehen werden. Ein solche Bauelementstruktur mit einer fein strukturierten Gate/Kathodenfläche zeigt Bild 3.3. Dabei wechseln sich Gate- mit Kathodenringen ab.

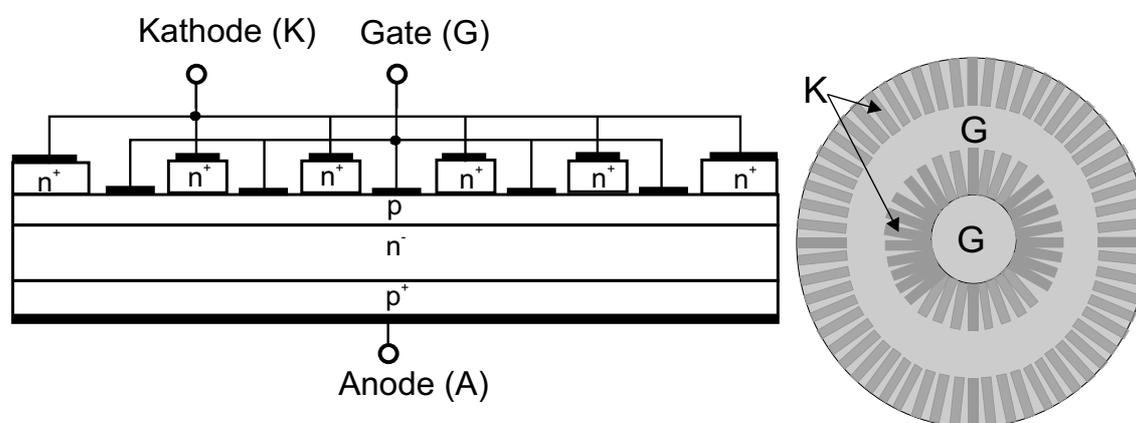


Bild 3.3: $p^+n^-p^-n^+$ -Bauelement mit strukturierter Gate/Kathodenfläche: Schnitt (links), Draufsicht (rechts)

Betrachtet man die Größenverhältnisse in Bild 3.3, läßt sich jetzt folgern, daß der laterale Einfluß des Gateanschlusses im Vergleich zu Bild 3.2 wesentlich ausgeprägter ist und damit ein homogenes Schalten überhaupt erst ermöglicht wird. Bei Thyristoren reicht dabei eine vergleichsweise geringe Anzahl breiter Gate- bzw. Kathodenringe aus, um über den Gateanschluß sicher einschalten zu können. Beim GTO dagegen ist eine höhere Anzahl von sich abwechselnden relativ schmalen Gate- und Kathodenringen vonnöten, wobei die Kathodengebiete eine Interdigitalstruktur bilden, d.h. kleine Kathodengebiete werden zu konzentrischen Ringstrukturen zusammengefaßt (s. Bild 3.3 rechts), damit ein Abschalten über das Gate möglich wird [1]. GTOs werden weich angesteuert, d.h. mit deutlich unter $100A/\mu s$ Gatestromanstieg. Dies hat zur Folge, daß bei einem Abschaltimpuls von ca. $U_{GK} = -20V$, die Gateinduktivität L_G größer gleich $100nH$ sein darf, was mit zumutbarem Aufwand zu erreichen ist. Die Abschaltverstärkung eines GTOs liegt im Bereich 3 bis 5, d.h. der negative Gatestrom erreicht immerhin bis zu 33% des abgeschalteten Anodenstroms. Da anfangs der Anodenstrom I_A noch nicht abfällt, d.h. $I_A = I_{Last}$, fließt der negative Gatestrom notwendigerweise über den Kathodenkontakt. Wenn der An-

odenstrom dann zu fallen beginnt, sinkt zwangsläufig auch der Kathodenstrom I_K . Dieser ist aber aber durchaus noch nicht Null, wenn der GTO, d.h. seine Driftzone, bereits merklich Blockierspannung aufnimmt. Dies führt aber genau zu der oben beschriebenen Filamentierung mit lokaler Überhitzung und Zerstörung des GTO. Um dies zu vermeiden, wird ein GTO immer mit einer RC(D)-Abschaltentlastung (Snubber) betrieben. Einen prinzipiellen Verlauf der Transienten eines GTO beim Abschalten mit Snubber zeigt Bild 3.4, wobei für diese Simulation ein einfacher Tiefsetzsteller verwendet wurde (vgl. Kapitel 4).

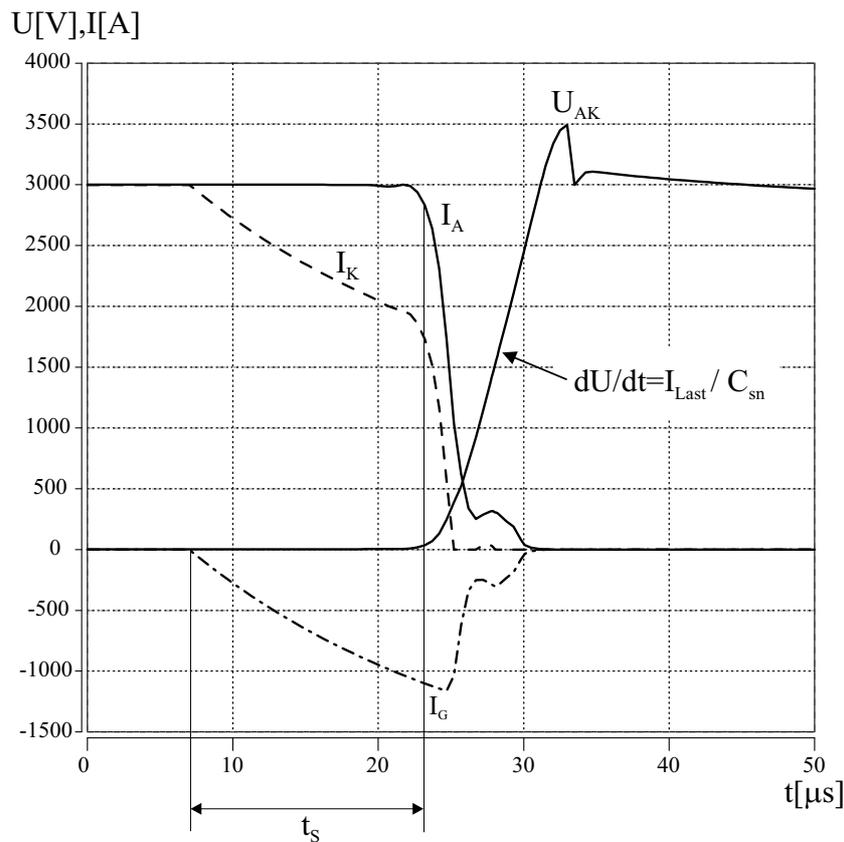


Bild 3.4: *Simulierte Transienten eines fiktiven GTO mit RCD-Snubber bei Abschalten von $I = 3\text{kA}$ gegen $U = 2,8\text{kV}$*

Die Kapazität C_{sn} des Snubbers übernimmt beim Abschalten nahezu den ganzen Anodenstrom und lädt sich dabei auf. Wegen $dU_{AK}/dt = I_{Last}/C_{sn}$ steigt die Spannung mit dieser von C_{sn} festgelegten Steigung im Vergleich zum unentlasteten Fall an und ermöglicht ein vorhergehendes Abklingen des Anoden- und Kathodenstroms, indem der Snubber den Laststrom temporär solange übernimmt, bis die Überspannung am GCT die Vorwärtspolung der Freilaufdiode ermöglicht und damit der Laststrom auf die Diode kommutieren kann. Die verzögerte Aufnahme der Blockierspannung ermöglicht das komplette Freiräumen der p -Basis von Überschlußladungsträgern,

bevor sich die Raumladungszone in die n^- -Basis ausdehnt. Damit wird ein laterales homogenes Abschaltverhalten erreicht, die Verlustleistungsdichte im GTO erheblich vermindert (vgl. Bild 3.6) und der Filamentierungseffekt dadurch vermieden. Die Diode² schließt den Snubberwiderstand R_{sn} beim Aufladen kurz. Liegt am GTO die Überspannung an, übernimmt die Freilaufdiode den Laststrom und C_{sn} wird über R_{sn} , der strombegrenzend wirkt, entladen. Insgesamt schaltet der GTO somit sicher ab. Allerdings werden die Verluste vom Leistungshalter damit größtenteils in den Snubberwiderstand verlagert und die Dauer eines Abschaltvorgangs verlängert sich³. Letztlich führte dieser Umstand zur weitgehenden Verdrängung des GTO durch den schaltverlustärmeren und schnelleren IGBT. Erst die Erfindung der GTO-Variante (I)GCT, der hartschaltend und unentlastet betrieben werden kann, führte zu einer Renaissance der Thyristorstruktur im Bereich mittlerer und hoher Leistung.

Auch der GCT besteht im Kern aus der gleichen Zonenfolge wie ein Thyristor, nur ermöglicht seine Gate/Kathodenstrukturierung⁴, die der des GTO ähnelt, hartes und unentlastetes Abschalten. Um mit einer möglichst kurzen n^- -Basis auszukommen, befindet sich zusätzlich eine n -dotierte Pufferschicht zwischen p^+ -Anodenschicht und der n^- -Basis. Es handelt sich also um ein Punch-Through-Bauelement ähnlich dem PT-IGBT und die Driftzone kann aufgrund der Möglichkeit der Aufnahme eines trapezförmigen E -Feldes, bei gleicher Zonenweite, mehr Spannung blockieren [32]. Dies wird durch den Nachteil des Verlusts der Rückwärtssperrfähigkeit erkauft. Man spricht deshalb von einem asymmetrischen GCT. Seit kurzem gibt es Veröffentlichungen betreffs der Entwicklung eines symmetrischen GCTs ohne Puffer (z.B. [75]), um prinzipiell auch Stromzwischenkreisumrichter mit GCTs betreiben zu können. Weiterhin werden zur Verringerung der Injektionseffizienz der Anode, im allgemeinen Anodenshots, d.h. n^+ -dotierte Bahnen im p^+ -Gebiet, verwendet. Ein asymmetrischer GCT mit Pufferschicht würde sich mit Anodenshots allerdings nur sehr schwer einschalten lassen [21] und man verwendet stattdessen eine transparente Anode mit einer Ausdehnung von nur $1 - 2\mu\text{m}$, um die Löcherinjektion in die Pufferschicht bzw. n -Basis wirksam zu begrenzen (vgl. dazu Abschnitt 2.2.2). In der vorliegenden Arbeit wird nur der asymmetrische GCT abgehandelt. Genauer gesagt, handelt es sich hierbei um die 3kA/4,5kV-Typen 5SHY35L4502 und 5SHY35L4503 der Firma ABB.

Die Transienten beim unentlasteten, harten Abschalten, unterscheiden sich dabei deutlich von denjenigen des GTO, der eben weich und entlastet abgeschaltet wird, erheblich. Für weiches Abschalten der GTOs der betrachteten Leistungsklasse, sind

²Die Snubberdiode D_{sn} wird aus Kostengründen meist weggelassen.

³Alleine die Speicherzeit t_S , die vom Beginn des Abschaltvorgangs bis zum 90%-Wert des Anodenstroms definiert ist (s. Bild 3.4), beträgt weit mehr als $10\mu\text{s}$.

⁴Auf die tatsächlich realisierten bzw. realisierbaren Varianten von Gate/Kathodenstrukturierungen wird in dieser Arbeit nicht weiter eingegangen, da sie für die 1D-Modellierung nicht maßgeblich sind.

Gatestromanstiege dI_G/dt von $40A/\mu s$ typisch, wohingegen man bei GCTs mit immerhin $dI_G/dt \geq 3000A/\mu s$ rechnen kann [1]. Bei negativen Spannungsimpulsen von ca. 15..20V, bedeutet dies, daß die Gateinduktivität L_G nur wenige nH betragen darf. Die Stromverstärkung I_A/I_G beträgt hierbei eins und der Gatestrom erreicht betragsmäßig den Wert des zu schaltenden Stromes I_A , d.h. in Verbindung mit dem steilen Gatestromansieg, wird der Anodenstrom sehr schnell vom Gateanschluß übernommen und der Kathodenstrom wird null, oder mit anderen Worten: Der Gate/Kathoden-Übergang wird sehr schnell und vor allem homogen gesperrt und die Vierschichtstruktur⁵ reduziert sich, funktionell betrachtet, zu einer Dreischichtstruktur, nämlich einem $p^+(n)n^-p$ -Transistor, dem oben erwähnten Wide-Base-Transistor (WBT). Das Abschaltverhalten wird im Bild 3.5 gezeigt.

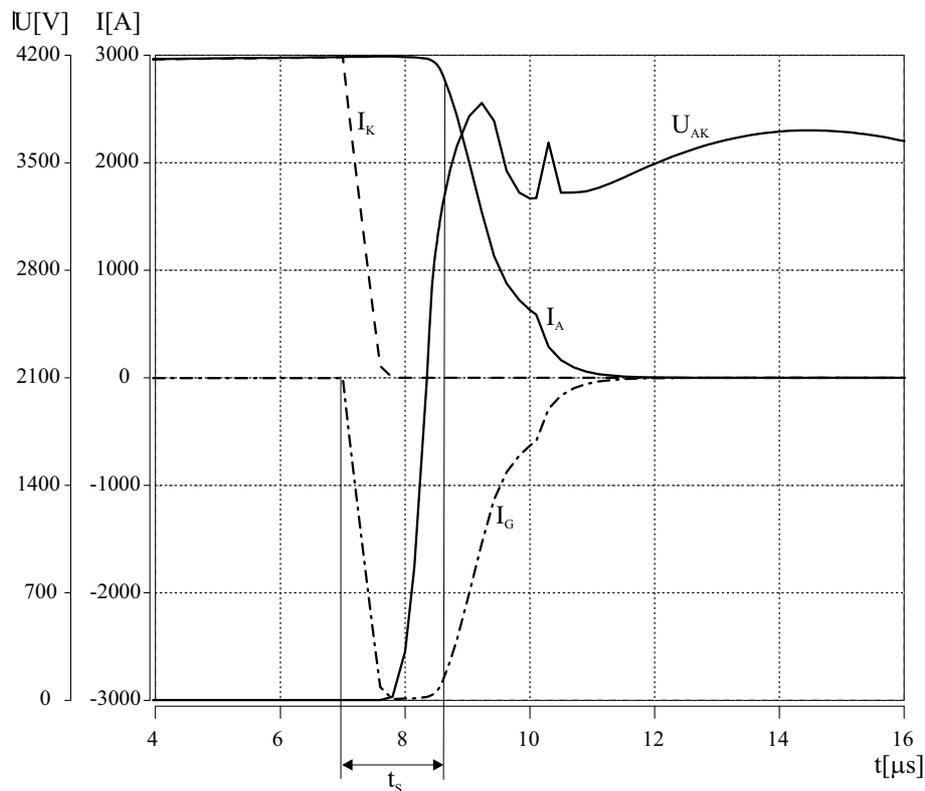


Bild 3.5: *Simulierte Transienten eines 3kA/4,5kV-IGCT bei unentlastetem Abschalten von $I = 3kA$ gegen $U = 2,8kV$*

Man sieht, daß der Kathodenstrom I_K bereits null ist, bevor der GCT Spannung aufnimmt, sich die Raumladungszone also ausdehnt. Der Filamentierungseffekt wird somit vermieden und der GCT kann erheblich größeren Verlustleistungsdichten standhalten, als der weichangesteuerte GTO. Die Verhältnisse betreffs der Verlustleistung

⁵mit Puffer sind es genau genommen fünf Schichten

können auch durch Betrachten der Bilder 3.4 und 3.5 abgeschätzt werden. Während beim Abschaltvorgang eines entlasteten GTOs niemals ein großer Anodenstrom und gleichzeitig eine bereits hohe Blockierspannung auftritt, geschieht dies beim unentlasteten GCT zwangsläufig, da erst die Überspannung am GCT ein Abkommutieren des Laststroms von der Anode zur Freilaufdiode ermöglicht. Weiterhin erhöht der dynamische Lawineneffekt die dynamischen Verluste im GCT. Dieser tritt auf, weil eine große Überspannung, d.h. Erreichen der kritischen Feldstärke E_{crit} , bei gleichzeitig noch hohem Strom, der durch den GCT fließt, die Lawinenmultiplikation außerordentlich begünstigt (vgl. hierzu die Abschnitte 1.4.5 und 2.4.3). Im Bild 3.5 erkennt man dies anhand des abnehmenden dU_{AK}/dt kurz vor dem ersten Maximum von U_{AK} bei ca. $t = 9,2\mu s$. Die zweite Spannungsspitze bei $t \approx 10,3\mu s$ tritt auf, wenn die Raumladungszone die Pufferschicht erreicht und das E-Feld von einem dreiecksförmigen in einen trapezförmigen Verlauf übergeht (vgl. Abschnitt 3.3.4). Dies bedeutet, daß sich die Fläche unterhalb der E-Feld-Gerade und damit die Blockierspannung, schlagartig erhöht. Der dadurch schneller abnehmende Anodenstrom bewirkt eine weitere Abnahme der effektiven Dotierung N_{eff} gemäß (2.86) und damit eine geringere Steigung von $E(x)$ und folglich eine Abnahme der Blockierspannung. Damit wird diese zweite Spannungsspitze quasi wieder abgebaut. Bild 3.6 zeigt nun die auftretende Verlustleistung und die dissipierte Energie eines einzelnen Abschaltvorgangs, jeweils für die beiden betrachteten Abschaltvorgänge des GTO und des GCT (Bilder 3.4 und 3.5)

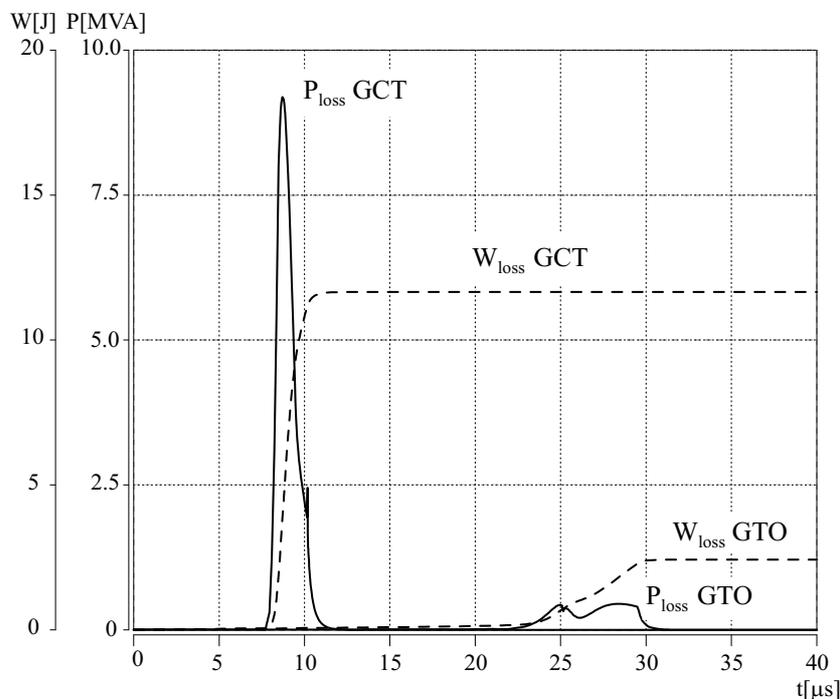


Bild 3.6: Vergleich von Verlustleistung und dissipierter Energie eines Schaltvorgangs: entlasteter GTO vs. unentlasteter GCT

Der scheinbare Nachteil beim GCT betreffs der Verlustleistung, fällt bei näherer Betrachtung kaum ins Gewicht, denn addiert man die dissipierten Energien des GTO und des Snubber-Widerstands R_{sn} , entspricht diese Summe in etwa der Verlustenergie des GCTs. Weiterhin führt ein Weglassen des RCD-Snubbers zu einer Reduktion der eingesetzten Bauelemente und damit der entstehenden Kosten⁶. Außerdem ist der gesamte Abschaltvorgang beim GCT wesentlich kürzer als beim GTO⁷. Alleine die Speicherzeit t_s ist mit ca. $2\mu s$ um ein Vielfaches kleiner als beim enlasteten GTO. Abschließend sei bemerkt, daß die Transienten $I_A(t)$ und $U_{AK}(t)$ beim GCT nicht von der äußeren Beschaltung, sondern von physikalischen, bauelementeninternen Vorgängen, bestimmt werden, was für die erforderliche Genauigkeit der Modellbeschreibung äußerst wichtig ist. Des weiteren kommt der eindimensionalen Modellbildung zugute, daß aufgrund des schnellen homogenen (lateralen) Abschaltens des Gate/Kathoden-Übergangs mit anschließender „Beschränkung“ der Betrachtung auf den WBT, laterale Effekte vernachlässigt werden können.

3.2 Das GCT-Modell: Modularer Aufbau

Das vorliegende GCT-Modell ist modular aufgebaut, wie aus Bild 3.7 ersichtlich wird und basiert auf dem GTO-Modell von Metzner [44], [45]. Ähnlich wie beim Diodenmodell (vgl. Bild 2.2), wird ein zentrales Modul von den beiden Randzonenmodulen Anode und Kathode eingerahmt. Die Modellierung der Randzonenmodule wurde bereits in Abschnitt 2.2 erläutert, so daß hierauf nicht mehr eingegangen werden muß. Das zentrale Modul beschreibt den Ein- und Auschaltmechanismus einerseits und das statische und dynamische Verhalten der Überschußladungsträger andererseits. Die p -Basis, über die elektrische Signale am Gateanschluß unmittelbar auf das Bauelement einwirken, bildet die Basis des n^-pn^+ -Transistors T_1 , der folglich immer zuerst schaltet. Erst danach reagiert aufgrund des sich ergebenden Kathodenstrom des T_1 , der dem Basisstrom von T_2 entspricht, der Wide-Base-Transistor (p^+nn^-p -Transistor) T_2 . Diese Transistoranordnung zeigt Bild 3.8. Das zentrale Modul beschreibt letztlich die Funktionsweise beider Transistoren plus der zugehörigen Ladungsträgerdynamik. Da sich aber der Löwenanteil der Überschußladungen in der schwach dotierten n^- -Zone, der Basis des WBT, befindet, heißt dieses Modul WBT-Modul. Wiederum wird das statische und dynamische Ladungsträgerverhalten in der n^- -Basis, durch die Kontinuitätsgleichung (2.51) beschrieben. Folglich ist auch bei diesem Modell das numerische Lösungsverfahren gemäß Abschnitt 2.3.2 implementiert.

⁶Kapazitäten im μF -Bereich für hohe Spannungen sind nicht billig. Gleiches gilt für die Dioden, die deswegen ohnehin meist weggelassen werden, was die Verlustsituation am R_{sn} weiter verschlechtert.

⁷Man beachte dazu den unterschiedlichen Maßstab der Zeitachsen in den Bildern 3.4 und 3.5

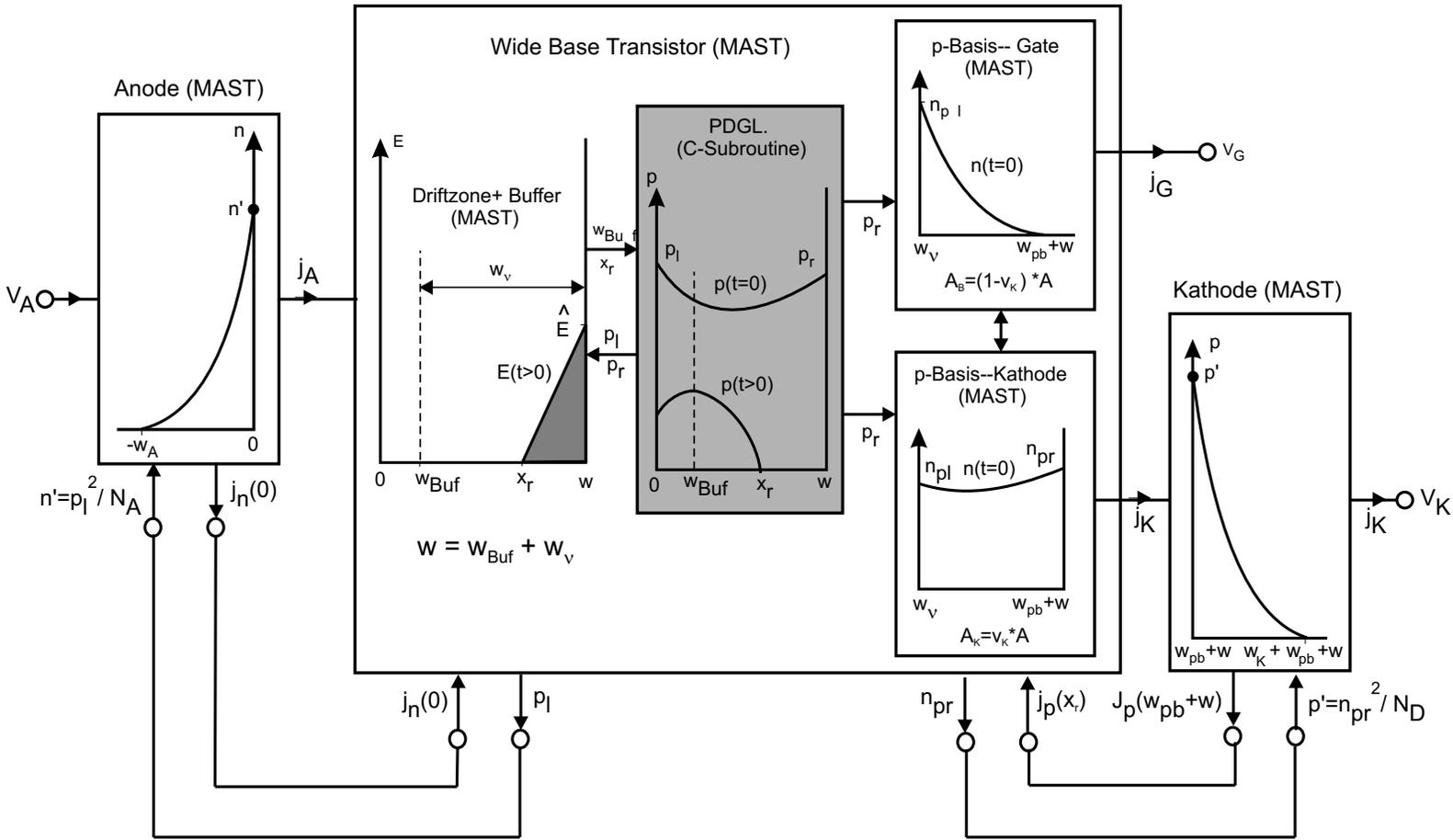


Bild 3.7: Schematische Darstellung der Komponenten des GCT-Modells
 (Darstellung der Löcher- und E-Feldverteilung im WBT-Modul für Ausschalten)

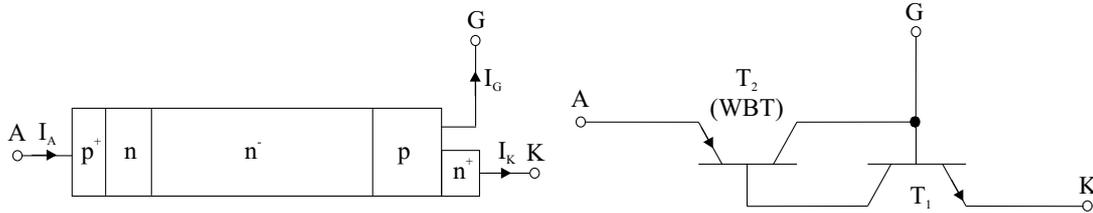


Bild 3.8: Zonenfolge und zugehöriges Zwei-Transistor-Ersatzschaltbild

Ein Unterschied zum Leistungsdiodenmodell in Bild 2.2 ist folgender: Da die Überschußladungen beim Abschalten nur über den Gateanschluß abgesaugt werden können und der p^+n -Übergang für $U_{AK} > 0$ immer vorwärtsgepolt ist, kann die Raumladungszone nur von der p -Basis aus, in die ν -Zone (n^- -Basis) vordringen. Dies bedeutet, daß immer $x_l = 0$ gilt. Der Verlauf des elektrischen Feldes im MAST-Modul, wird wiederum als lineare Funktion genähert und die aufgenommene Spannung durch Flächenberechnung berechnet. Die Beschreibung des Übergangs eines dreieckförmigen, zu einem trapezförmigen Feldverlauf aufgrund der Pufferschicht ist ebenfalls enthalten. Ferner kommen (2.95) und (2.96) gemäß Abschnitt 2.4.3 zur Beschreibung des dynamischen Lawineneffekts zur Anwendung. Zunächst aber wird die Modellierung der Pufferschicht und die Auswirkungen auf das elektrische Feld beschrieben (Abschnitt 3.3). In Abschnitt 3.4 folgt die Beschreibung, der für den Schaltmechanismus so wichtigen p -Basis. Fragen betreffs des Zusammenspiels der einzelnen Zonen(module) werden in Abschnitt 3.5 abgehandelt.

3.3 Modellierung der Pufferschicht

3.3.1 Allgemeines

Die Pufferschicht hat die Aufgabe als Stoppschicht für das elektrische Feld zu dienen, d.h. die sich während des Abschaltens von der p -Basis ausbreitende Raumladungszone dringt nicht wesentlich in die Pufferzone ein und verhindert somit den Punch Through mit der p^+ -Anode, der unweigerlich zu einem massiven Wiederanstieg des Stromes bei voller Blockierspannung und damit zur Zerstörung des Bauelements führen würde. Damit die Raumladungszone nur unwesentlich in den n -Puffer eindringt und anschließend ein E-Feld-Trapez in der ν -Zone entsteht, muß die Dotierdichte N_{Buf} merklich größer sein als die Dotierdichte der n^- -Basis N_ν . In der Praxis bewegt sich die Pufferdotierdichte im Bereich von 10^{16}cm^{-3} . Damit liegt man aber etwa in der gleichen Größenordnung wie die in den Puffer injizierte Löcherdichte. Man hat also weder schwache noch hohe Injektion, sondern so etwas wie

mittlere Injektion $p \approx N_{Buf}$, woraus für die Elektronendichte $n = p + N_{Buf}$ folgt⁸. Die sich prinzipiell einstellenden Ladungsträgerdichten im stromführenden Zustand im Puffer und den angrenzenden Zonen Anode bzw. n^- -Basis, zeigt Bild 3.9.

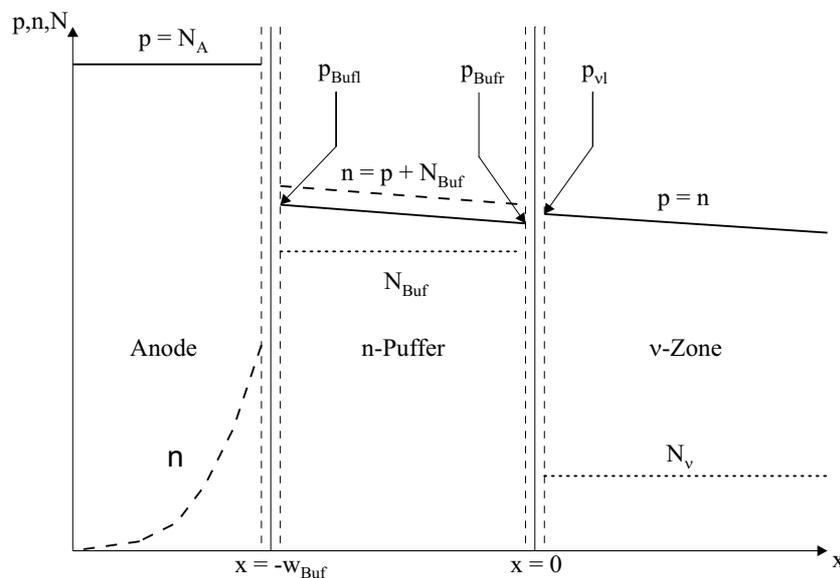


Bild 3.9: Ladungsträgerdichten im Puffer eines eingeschalteten GCT

Typische Zonenweiten für Puffer liegen bei einigen $10\mu m$. Da das elektrische Feld nicht in den Puffer eindringt, müssen die Überschussladungsträger dort einzig durch Rekombination mit der Pufferlebensdauer $\tau_{Buf} < \tau_h$, abgebaut werden. Betreffs der Modellierung kann man zwei Konzepte verfolgen:

- (1) Die quasistatische Modellierung: Das Verhalten der Überschussladungsträger im Puffer wird kompakt modelliert, d.h. die Ladungsträgerverläufe werden durch eine analytische Funktion beschrieben, die sich aus der Diffusionsgleichung für $\partial p/\partial t = 0$ (quasistatisch!) ergibt. Eine numerische Lösung mit Ortsdiskretisierung findet demnach nicht statt.
- (2) Die Pufferschicht wird „in die Driftzone“ integriert, d.h. eine numerische Lösung zur Berechnung der Ladungsträgerkurven im Numerik-Submodul ist vorgesehen.

3.3.2 Quasistatische Modellierung

Hierbei wird die Pufferzone ähnlich wie die Randzonen Anode und Kathode kompakt beschrieben [36]. Dazu benötigt man die Parameter Zonenweite w_{Buf} , die Dotierung

⁸folgt aus der Neutralitätsbedingung (1.20) bei Annahme vollständiger Ionisierung

N_{Buf} und die Löcherlebensdauer⁹ im Puffer τ_{pBuf} . Die Anode injiziert für positiven Stromfluß Löcher in die Pufferschicht, weshalb sich am linken Rand des Puffers die Löcherdichte p_{Buf_l} einstellt. Zur Bestimmung von p_{Buf_l} gilt anstelle von (2.9) am p^+n -Übergang

$$j_n(-w_{Buf}) = j_{n0} \frac{p_{Buf_l} (p_{Buf_l} + N_{Buf})}{n_i^2} \quad (3.1)$$

Die Löcherdichte am rechten Pufferzonenrand p_{Buf_r} und die linksseitige Löcherdichte in der ν -Zone $p_{\nu l}$ hängen über die Beziehung

$$p(0) n(0) = p_{Buf_r} (p_{Buf_r} + N_{Buf}) = p_{\nu l}^2 \quad (3.2)$$

zusammen. Aufgelöst nach p_{Buf_r} , ergibt sich

$$p_{Buf_r} = 0,5 \left(N_{Buf} + \sqrt{N_{Buf}^2 + 4p_{\nu l}^2} \right) \quad (3.3)$$

Die Ladungsträgerverteilung in der Pufferschicht wird ebenfalls durch eine Diffusionsgleichung ähnlich (2.36) beschrieben, nur eben jetzt allgemein für mittlere Injektion. Quasistatisch heißt, daß die Zeitableitung der Löcherdichte in der partiellen Differentialgleichung Null gesetzt wird und somit eine gewöhnliche Dgl, die analytisch lösbar ist, übrigbleibt. Mit den Randwerten p_{Buf_l} und p_{Buf_r} lautet diese Lösung mit $L_{Buf} = \sqrt{\tau_{pBuf} D_p}$ analog (2.39)

$$p(x) = p_{Buf_l} \frac{\sinh\left(\frac{-x}{L_{Buf}}\right)}{\sinh\left(\frac{w_{Buf}}{L_{Buf}}\right)} + p_{Buf_r} \frac{\sinh\left(\frac{x + w_{Buf}}{L_{Buf}}\right)}{\sinh\left(\frac{w_{Buf}}{L_{Buf}}\right)} \quad (3.4)$$

Nun ist diese Lösung tatsächlich nur für den Durchlaßfall exakt richtig¹⁰. Doch insbesondere für das Abschalten ist diese Lösung fehlerhaft, da (3.4) immer eine links gekrümmte Kurve zwischen den Randwerten $p_{Buf_l} > 0$ und $p_{Buf_r} > 0$, d.h. $\partial^2 p / \partial x^2 > 0$, impliziert, obwohl dies beim induktiven Abschalten nicht der Fall ist, wie in Abschnitt 3.3.3 (s. Bild 3.10) gezeigt wird. Dieser Fehler hält sich in der Regel in Grenzen, da der Anteil der Überschlußladungen im Puffer im Vergleich zur ν -Zone gering ist, sofern $w_{\nu} \gg w_{Buf}$ gilt. Mit (3.4) läßt sich die kompakte Ladung Q_{Buf} zu

$$Q_{Buf} = q_0 \int_{-w_{Buf}}^0 p(x) dx \quad (3.5)$$

bestimmen. Mit (1.34) folgt

$$j_p(-w_{Buf}) = j_{ges} - j_n(-w_{Buf}). \quad (3.6)$$

⁹wie bei allen Lebensdauern, verwendet man auch τ_{pbuf} als Fitparameter

¹⁰Dann gilt in der Tat $\partial p / \partial t = 0$

Die Beschreibung des Minoritätsträgerstroms an der Stelle $x = 0$, wird durch das zeitliche Verhalten der kompakten Überschußladung Q_{Buf} gemäß (3.5) bestimmt. Integriert man die Kontinuitätsgleichung für Löcher (1.37) mit $R_{netto,p} = p/\tau_{pBuf}$ über w_{Buf} und multipliziert mit q_0 durch, ergibt sich:

$$\frac{\partial}{\partial t} \int_{-w_{Buf}}^0 q_0 p dx = -\frac{1}{q_0} \int_{-w_{Buf}}^0 q_0 \frac{\partial j_p}{\partial x} dx - \int_{-w_{Buf}}^0 q_0 \frac{p}{\tau_{pBuf}} dx \quad (3.7)$$

Durch Lösen der Integrale ergibt sich mit (3.5):

$$\frac{dQ_{Buf}}{dt} = -j_p(0) + j_p(-w_{Buf}) - \frac{Q_{Buf}}{\tau_{pBuf}} \quad (3.8)$$

Die Löcherstromdichte $j_p(-w_{Buf})$ erhält man mit Hilfe von (3.6) und (3.1). Somit läßt sich die Löcherstromdichte an der Stelle $x = 0$ darstellen als

$$j_p(0) = j_p(-w_{Buf}) - \frac{Q_{Buf}}{\tau_{pBuf}} - \frac{dQ_{Buf}}{dt}. \quad (3.9)$$

Die Summe der Übergangsspannungen am p^+n -Übergang und $n\nu$ -Übergang (vgl. Abschnitt 1.6) ergibt sich zu

$$U_j(-w_{Buf}) + U_j(0) = \frac{k_B T}{q_0} \ln \left(\frac{p_{Buf} N_{Buf}^2}{p_{\nu} n_i^2} \right) \quad (3.10)$$

Damit sind alle Gleichungen bekannt, um die Ladungsträgervorgänge in der Pufferschicht zu beschreiben. In [36] wird gezeigt, daß ein Modell mit dieser Pufferschichtbeschreibung prinzipiell richtige Ergebnisse liefert. Allerdings ergeben sich Probleme hinsichtlich der Konvergenz, so daß dieser Ansatz letztlich zugunsten der Integration der Pufferschicht in das Modul zur Berechnung der Löcherdichten in der ν -Zone verworfen wurde.

3.3.3 Pufferzone als Teil der Driftzonenbeschreibung

Die Pufferzone als Teil der Driftzonenbeschreibung bedeutet einfach, daß der Bereich $x \in [0, w]$ in dem eine Ortsdiskretisierung im Numerikmodul stattfindet, in einen Bereich Pufferschicht mit $x \in [0, w_{Buf}]$ und einen Bereich ν -Zone mit $x \in]w_{Buf}, w]$ unterteilt wird (vgl. Bild 3.7). Für die zur Bestimmung der anodenseitigen Löcherinjektion p_{Buf} maßgebliche Elektronenstromdichte gilt (3.1), wobei bedingt durch das unterschiedliche Koordinatensystem im Vergleich zu Abschnitt 3.3.2, auf der linken Seite natürlich $j_n(0)$ statt $j_n(-w_{Buf})$ stehen muß. Die kathodenseitige Elektroneninjektion dagegen wird nach wie vor von der Löcherstromdichte $j_p(w)$ gemäß (2.11)

bestimmt. Diese Lösung ist programmiertechnisch sehr einfach zu implementieren und liefert ein sehr gutes Konvergenzverhalten. Allerdings fällt faktisch der Puffer-Driftzonen-Übergang (ein nn^- -Übergang, s. Abschnitte 3.3.2 und 1.6.3) weg, d.h. es wird stillschweigend $p_{Buf} = p_{vl}$ vorausgesetzt, was außer für Hochinjektion eigentlich nicht korrekt ist (vgl. dazu Bilder 3.9 und 3.12). Die dadurch bedingten Rechenfehler werden aber durch die Tatsache, daß der grundsätzliche Verlauf der Plasmakurven an der Stelle $x = w_{Buf}$ von der implementierten Numerik richtig berechnet wird, mehr als wettgemacht, wie die Simulationsergebnisse im Kapitel 4 zeigen. Bild 3.10 zeigt die simulierten Löcherverteilungen für induktives Abschalten von $I_A = 3kA$ gegen $U_{AK} = 2.8kV$. Jede einzelne Kurve entspricht der Löcherverteilung für einen von SABER festgelegten Zeitschritt. Zu Beginn, d.h. für $t = 0$, erhält man die typische Löcherverteilung, die dem eingeschalteten Zustand für $I = 3kA$ entspricht und sich als Lösung von (2.51) mit $\partial p/\partial t = 0$ ergibt. Man sieht, daß sich auch im Pufferbereich der Ladungsträgerverlauf am Anfang links gekrümmt ist, sich im Laufe des Abschaltens aber „umstülp“, d.h. daß sich Rechtskrümmung einstellt.

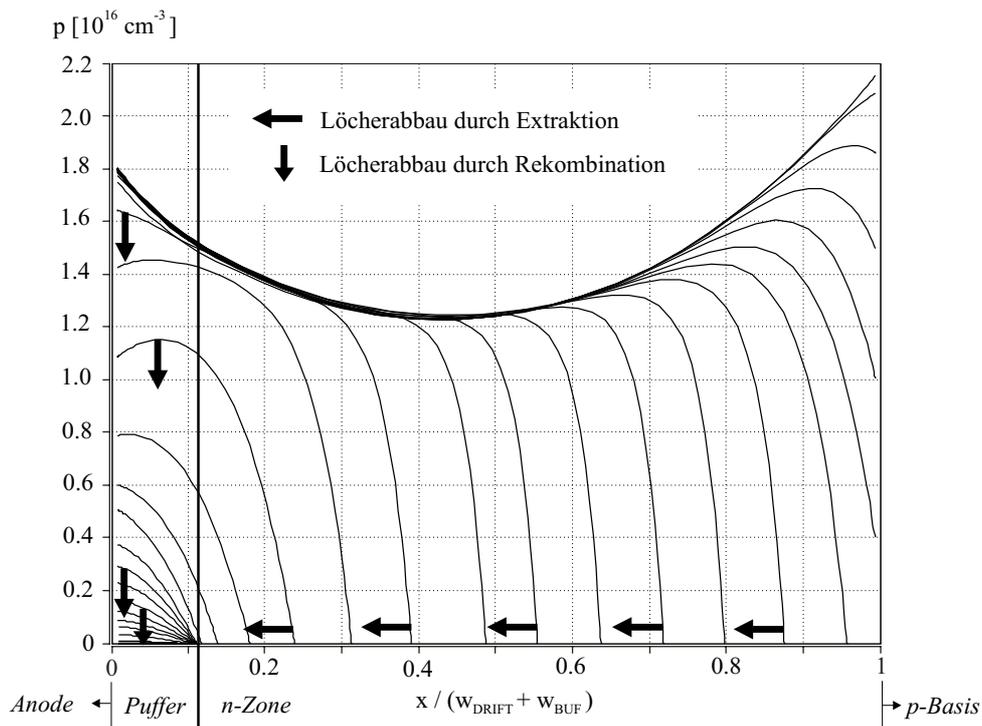


Bild 3.10: Simulierte Ladungsträgerkurven eines $3kA/4,5kV$ -IGCT bei unentlastetem Abschalten von $I = 3kA$ gegen $U = 2,8kV$

Für den Pufferbereich ist τ_{pBuf} die maßgebliche Lebensdauer, während für die n -Zone weiterhin die größere Hochinjektionslebensdauer τ_h gilt. Wie bereits in Abschnitt 2.4.2 erläutert, läßt sich die Lebensdauer in der Driftzone ortsabhängig definieren. Somit kann man $\tau_{pBuf} = a \cdot \tau_h$ mit $0 < a < 1$ für $x \in [0, w_{Buf}]$ eingeben.

3.3.4 Pufferschicht und Raumladungszone

Wie bereits in Abschnitt 3.2 erwähnt, werden die Löcher beim Abschalten lediglich über den Gateanschluß, nicht aber über den Anodenkontakt extrahiert und der zentrale pn -Übergang zwischen ν -Zone und p -Basis beginnt zu sperren. Folglich dehnt sich die Raumladungszone (RLZ) von $x = w$ kommend entgegen der positiven x -Richtung in die ν -Zone aus, oder mit anderen Worten: Die RLZ kommt „von rechts“ (vgl. Bild 3.7). In Bild 3.11 läßt sich während des Abschaltens, die Gesamtweite $w = w_{Buf} + w_\nu$ in zwei Bereiche unterteilen: Der Plasmabereich erstreckt sich über $0 < w_{Plasma} < x_r$, wohingegen die RLZ von x_r und w begrenzt wird. Die Aufnahme des elektrischen Feldes während und nach dem Abschaltvorgang läuft prinzipiell ähnlich wie bei der Leistungsdiode ab (s. dazu Abschnitt 2.3.2). Wiederum findet die Berechnung der RLZ in einem MAST-Modul, dem WBT-Modul, das dem Numerik-Modul übergeordnet ist, statt. Bisher wurde der Puffer nur bezüglich des Ladungsträgerverhaltens beschrieben, wogegen nun die Modellierung des elektrischen Feldes im Vordergrund steht. Um das Verständnis für die dynamischen Vorgänge diesbezüglich zu erleichtern, dient Bild 3.11, das qualitativ die Überschußlöcher- und E -Feld-Verteilungen für vier verschiedene repräsentative Zeitpunkte $t_{1,2,3,4}$, während eines harten und unentlasteten Abschaltvorgangs zeigt.

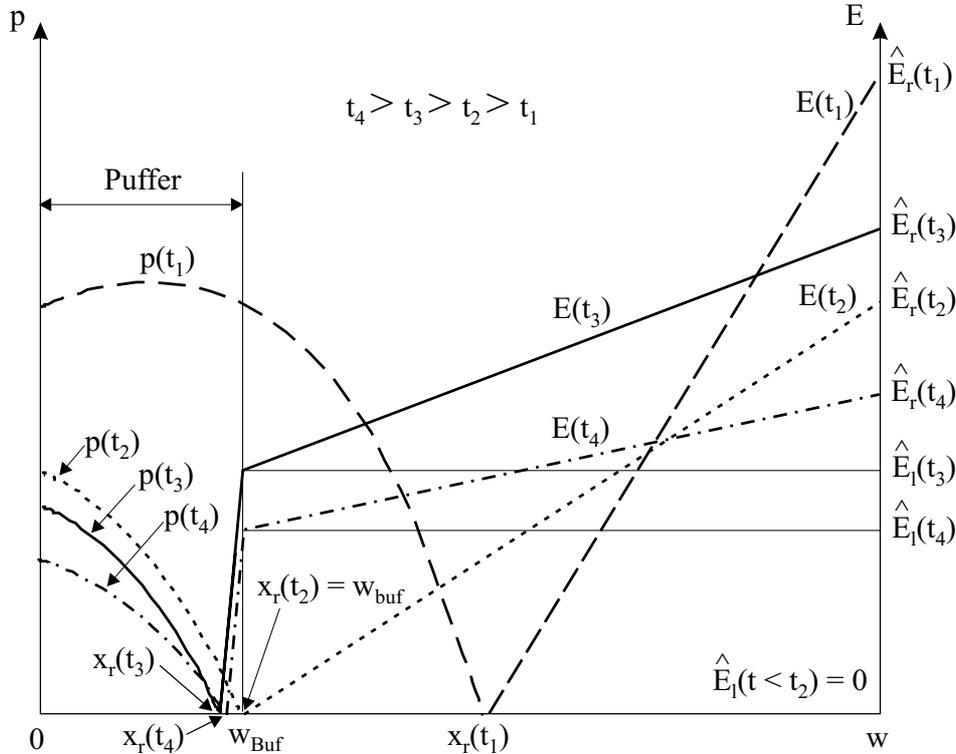


Bild 3.11: Darstellung zur Erläuterung des dynamischen Verhaltens des E -Feldes beim harten und unentlasteten Abschalten eines GCT.

Zum Zeitpunkt t_1 hat der Abschaltvorgang schon eingesetzt. Dabei wurde bereits ein beträchtlicher Teil der Überschußladungen ausgeräumt und eine Raumladungszone (RLZ) $w - x_r(t_1)$ ist bereits entstanden. Zu diesem Zeitpunkt ist der Strom durch den GCT noch relativ groß. Folglich ist die effektive Dotierung N_{eff} gemäß (2.86) ebenfalls groß¹¹ und folglich der E-Feld-Gradient entsprechend steil. Zum Zeitpunkt t_2 hat die RLZ die Pufferzone gerade erreicht, d.h. es gilt $x_r(t_2) = w_{Buf}$. Das E-Feld-Dreieck ist in der Regel nun wesentlich flacher, da einerseits die Überspannung schon wieder abklingt und andererseits der Strom durch den GCT und damit N_{eff} ebenfalls kleiner geworden ist. Im Bild 3.5 befindet sich t_2 bei etwa $10\mu s$, also unmittelbar vor der zweiten Spannungsspitze. Zum Zeitpunkt t_3 , der unmittelbar t_2 folgt, ist die Raumladungszone in den Puffer eingedrungen. Die Länge der RLZ im Puffer beträgt genau dann $w_{Buf} - x_r(t_3)$. In diesem Teilbereich der RLZ bestimmt die Pufferdotierung N_{Buf} und nicht mehr N_ν bzw. N_{eff} den E-Feld-Gradienten¹². Folglich erhält man wesentlich steilere E-Feld-Gradienten im Puffer im Vergleich zur Driftzone. Es entsteht also sehr schnell ein trapezförmiger Verlauf des Feldes in der Driftzone und ein entsprechend starkes Anwachsen der Fläche unter der E-Feld-Funktion. Die ist gleichbedeutend mit einer sprunghaften Erhöhung der Blockierspannung am GCT, was sich in der besagten Spannungsspitze in Bild 3.5 äußert. Die zusätzliche Überspannung an der parasitären Induktivität bewirkt ein stärkeres Abklingen des Stroms durch den GCT, was sich durch den Knick in der Kurve I_A im Bild 3.5 zum Zeitpunkt des Spannungspeaks bemerkbar macht. Dadurch wird aber N_{eff} und dE/dx in der Driftzone abgesenkt und die RLZ im Puffer wieder verkleinert. Dieser Mechanismus wird in Bild 3.11 durch $p(x, t_4)$ und $E(x, t_4)$ repräsentiert. Die maximale Ausdehnung der RLZ in den Puffer ist im Vergleich zu w_{Buf} gering, weshalb man beim Puffer berechtigterweise von einer Stoppschicht bzgl. der RLZ sprechen kann.

Im Modell wird in guter Näherung, zur Berechnung der Blockierspannung selbst, nur die Fläche unterhalb der E-Feld-Verteilung in der Driftzone betrachtet. Sofern es sich um einen dreiecksförmigen Verlauf handelt benötigt man neben der Ausdehnung der RLZ $w - x_r$, das maximale Feld \hat{E}_r bei $x = w$. Im Falle des Trapezes, beträgt die Ausdehnung der RLZ wegen der eben erwähnten Näherung fix w . Neben \hat{E}_r braucht man das linksseitige Feldmaximum \hat{E}_l bei $x = 0$, das sich mit Hilfe von (2.85) mit $w_{Buf} - x_r$ und N_{Buf} zu

$$\hat{E}_l = \begin{cases} 0 & : w_{Buf} < x_r(t) < w \\ \frac{q_0 N_{Buf}}{\epsilon_0 \epsilon_r} (w_{Buf} - x_r(t)) & : 0 < x_r(t) < w_{Buf} \end{cases} \quad (3.11)$$

¹¹Anstelle $j_p(0)$ wie bei der Diode, wird zwangsläufig $j_p(w)$ in (2.86) eingesetzt. Wiederum wird j_n in der RLZ vernachlässigt und demnach null gesetzt.

¹² N_{eff} beträgt maximal etwa $5 \cdot 10^{13} cm^{-3}$. Während N_{eff} im Vergleich zu $N_\nu \approx 1 \cdot 10^{13} cm^{-3}$ nicht vernachlässigt werden darf, spielt N_{eff} im Vergleich zu $N_{Buf} \approx 10^{16} cm^{-3}$ keine Rolle.

ergibt. Ebenfalls unter Anwendung von (2.85) mit $w - x_r$ und N_{eff} unter Berücksichtigung von \hat{E}_l gemäß (3.11), errechnet sich \hat{E}_r gemäß

$$\hat{E}_r = \begin{cases} \frac{q_0 N_{eff}(t)}{\epsilon_0 \epsilon_r} (w - x_r(t)) & : w_{Buf} < x_r(t) < w \\ \frac{q_0 N_{eff}(t)}{\epsilon_0 \epsilon_r} w + \hat{E}_l & : 0 < x_r(t) < w_{Buf} \end{cases} \quad (3.12)$$

Somit kann die aufgenommene Blockierspannung berechnet werden. Die entsprechenden Ausdrücke lauten:

$$U_{Block} = \begin{cases} \frac{1}{2} \hat{E}_r (w - x_r(t)) & : w_{Buf} < x_r(t) < w \\ \frac{1}{2} (\hat{E}_l + \hat{E}_r) w & : 0 < x_r < w_{Buf} \end{cases} \quad (3.13)$$

Die maximaler rechtsseitige Feldstärke \hat{E}_r erreicht bzw. überschreitet die kritische Feldstärke E_{crit} recht häufig im Falle hart abgeschalteter IGCTs, die bei Nennstrom und Nennspannung betrieben werden. Deshalb wird auch beim GCT-Modell der dynamische Lawineneffekt gemäß der Beschreibung in Abschnitt 2.4.3 implementiert, wobei in diesem Fall $j_p(w)$ mit dem Multiplikationsfaktor M_{AV} gemäß (2.95) multipliziert wird, da wiederum im Bereich der gesamten Raumladungszone $j_p(w) = j_p(x > x_r) \approx j_{ges}$ und $j_n(x > x_r) \approx 0$ gilt. Ferner muß bei der Berechnung der Stomdichte die Verschiebungsstromdichte, die sich aus (2.90) ergibt, berücksichtigt werden, wobei hier wegen $dx_l/dt = 0$, nur $j_{Dis}(x_r)$ maßgeblich ist.

3.4 Modellierung der p-Basis

Die p -Basis des IGCT entspricht zugleich der Basis eines n^-pn^+ -Transistors (T_1 in Bild 3.8), der über das Gate angesteuert wird und für das Schalten von IGCTs von großer Bedeutung ist. Zur Beschreibung des Transistoreffekts dient die Minoritätsträgerverteilung (Elektronenverteilung) in der p -Basis. Um die Modellkomplexität zu begrenzen, wird auf eine Ortsdiskretisierung in der p -Basis verzichtet, d.h. diese Zone wird quasistatisch modelliert. Die Diskussion um die Zulässigkeit dieser Maßnahme erfolgt weiter unten im Text auf Seite 103. Trotzdem bildet das p -Basis-Modul nicht einen monolithischen Block, sondern besteht gewissermaßen aus zwei Teilen. Dies wird notwendig, weil die p -Basis teilweise an die Kathodenschicht grenzt und der Rest die Gatefläche bildet. Durch diese Zweiteilung des Moduls (vgl. Bild 3.7) kann die eindimensionale Beschreibung aufrechterhalten werden. Die externen Modellströme, d.h. Anodenstrom I_A , Kathodenstrom I_K und Gatestrom I_G ,

v_K mit $0 < v_K < 1$. Für die Elektroneninjektion von der Kathode in die p -Basis gilt anstelle von (2.11)

$$j_{pK}(w + w_{pb}) = j_{p0} \frac{n_{pr} (n_{pr} + N_{pb})}{n_i^2}, \quad (3.15)$$

unabhängig davon, ob schwache, mittlere oder starke Injektion vorliegt. Bei Annahme einer ortsunabhängigen Dotierung N_{pb} und folglich einem abrupten νp -Übergang bei $x = w$ (vgl. Abschnitt 1.6) gilt für die Produkte aus Löcher- und Elektronendichte links und rechts vom betrachteten Übergang:

$$p_r^2 = n_{pl} (n_{pl} + \hat{N}_{pb}) \quad (3.16)$$

Gerade aber die p -Basisschicht wird technologisch meist durch eine tiefe Diffusion hergestellt (s. z.B. [51]). Folglich ist $N_{pb}(w)$ bereits sehr klein im Vergleich zur maximalen Akzeptorendichte $N_{pb}(w + w_p) = \hat{N}_{pb}$, so daß N_{pb} in (3.16) vernachlässigt werden kann und damit näherungsweise

$$p_r = n_{pl} \quad (3.17)$$

gilt¹³. Aus (1.38) kann eine Diffusionsgleichung der Bauart (2.36) abgeleitet werden, die sich für den quasistatischen Fall, d.h. $\partial n / \partial t = 0$, analytisch lösen läßt. Analog (2.39) ergibt sich für diese Lösung mit $L_{pb} = \sqrt{\tau_{npb} D_n}$ und den Randwerten n_{pl} und n_{pr}

$$n(x) = n_{pl} \frac{\sinh\left(\frac{w_{pb} + w - x}{L_{pb}}\right)}{\sinh\left(\frac{w_{pb}}{L_{pb}}\right)} + n_{pr} \frac{\sinh\left(\frac{x - w}{L_{pb}}\right)}{\sinh\left(\frac{w_{pb}}{L_{pb}}\right)}. \quad (3.18)$$

Diese Lösung impliziert mit $n_{pl,r} > 0$ wieder eine linksgekrümmte Kurve, d.h. $\partial n^2 / \partial x^2 > 0$. Diese Aussage entspricht den Verhältnissen im eingeschalteten Zustand und zumindest qualitativ auch während des Einschaltens. Anders sind die Verhältnisse beim Ausschalten, wo die Kurven im Zeitablauf ähnlich wie beim Puffer von der Linkskrümmung in eine Rechtskrümmung übergehen (vgl. [23], [61]). Dieser Fall wird aber durch die quasistatische Modellierung nicht berücksichtigt. Der Rechenfehler, der sich dadurch ergibt, kann allerdings aus folgenden Gründen verschmerzt werden:

1. Ähnlich wie beim Puffer, ist die p -Basis deutlich schmaler als die n -Basis, so daß die Menge an gespeicherten Ladungsträgern vergleichsweise gering ist.
2. Aufgrund des harten Abschaltens kommutiert der Transferstrom sehr schnell von der Kathode zur Anode, was ein sehr schnelles Ausräumen der p -Basis

¹³Gleiches kann übrigens auch für die Pufferschicht gelten, was dann die in Abschnitt 3.3.3 gemachte Annahme $p_{Bufr} = p_{vl}$ zusätzlich untermauert.

(3.19), erhält man die für j_K maßgebliche Elektronenstromdichte $j_{nK}(w + w_{pb})$

$$j_{nK}(w + w_{pb}) = j_{nK}(w) + \frac{Q_{pbK}}{\tau_{npb}} + \frac{dQ_{pbK}}{dt}. \quad (3.21)$$

Die Gesamtstromdichte an der Kathode j_K errechnet sich zu

$$j_K = j_K(w + w_{pb}) = j_{pk}(w + w_{pb}) + j_{nk}(w + w_{pb}) \quad (3.22)$$

also durch Addition von (3.15) und (3.21). Da auf der rechten Seite von (3.19) ebenfalls j_K steht, muß (3.22) noch nach j_K aufgelöst werden (s. Abschnitt 3.5). Damit ist die p -Basis ausreichend beschrieben.

3.5 Das Gesamtmodell

Die Verkopplung der Einzelmodule des GCT-Modells funktioniert im Grunde ähnlich wie beim Diodenmodell (s. Abschnitt 2.5). Allerdings sind einige Unterschiede zu beachten, auf die im folgenden eingegangen werden. Die rechtsseitige Löcherinjektion in die ν -Zone p_r , hängt unmittelbar mit der Elektroneninjektion von der Kathode in die p -Basis n_{pr} und den elektrischen Eigenschaften bzw. dem Verhalten der p -Basis zusammen. Für das Modell sind also die Variablen p_l an der Stelle $x = 0$ und n_{pr} an der Stelle $x = w + w_{pb}$ interessant, die so eingestellt werden, daß die Modellgleichungen, d.h. insbesondere die Transportgleichungen, erfüllt sind. Bekanntlich kann eine Raumladungszone beim vorwärtsblockierenden GCT nur rechtsseitig in die Driftzone eintreten, da der Anoden/Pufferübergang immer in Vorwärtspolung betrieben wird, d.h. eine Tabelle analog zu (2.97) umfaßt linksseitig lediglich Fall 1 und 2 (vgl. auch Bild 3.7).

Zur Beschreibung der verkoppelten Stromdichtegleichungen, kann man sich vorerst auf die linke Seite konzentrieren, da die kathodenseitigen Gleichungen bereits in Abschnitt 3.4 diskutiert werden. Die n -Pufferzone wurde in das numerische Driftzonenmodul integriert. Da für die Dotierdichte im Puffer $N_{buf}(x = 0) = \hat{N}_{Buf}$ gilt, also maximal ist, darf man nicht mehr von Hochinjektion ausgehen, sondern muß allgemein ansetzen. Weiterhin gilt wegen $x_l = 0$ (keine RLZ!), zwingend $j_{dis}(x_l) = 0$. Im Modell gilt weiterhin die vereinfachende Annahme, daß der Parameter N_{Buf} überall gleich ist und damit auch seinem Maximalwert \hat{N}_{Buf} entspricht. Man enthält anstelle von (2.98)

$$j_n(x_l = 0) = j_{n0} \frac{p_l (p_l + N_{Buf})}{n_i^2}. \quad (3.23)$$

Auch bezüglich der pufferseitigen Randbedingung, d.h. $j_p(x_l)$, darf die Hochinjektionsbedingung nicht mehr verwendet werden. Anstelle von (2.100) verwendet man

die Löcherstromdichtegleichung für allgemeine Injektion

$$j_p(x_l = 0) = \frac{p_l}{p_l(b+1) + bN_{Buf}} j_A - q_0 D_{AI}(x_l) \frac{\partial p}{\partial x}(x_l). \quad (3.24)$$

Man erhält diese Gleichung, indem man in (2.30) unter der Annahme, daß statt $n = p$, $n = p + N_{Buf}$ (allgemeine Injektion) gilt, p durch $p + N_{Buf}$ ersetzt. Hierbei ist D_{AI} ein Diffusionskoeffizient, der für allgemeine Injektion (AI) gilt. Er ist zu

$$D_{AI}(x_l) = \frac{p_l D_n + b(p_l + N_{Buf}) D_p}{p_l(b+1) + bN_{Buf}} \quad (3.25)$$

definiert und schließt (2.34) ein, wenn man die Hochinjektionsbedingung $p \gg N_{Buf}$ in (3.25) einsetzt. Der Koeffizient D_{AI} ist nicht mehr konstant bezüglich p wie bei starker oder schwacher Injektion, sondern eine Funktion von p .

Die Addition von (3.23) mit (3.24) ergibt die Anodenstromdichte j_A . Da j_A auch in (3.24) vorkommt, muß die entstehende Gleichung noch nach j_A aufgelöst werden. Gleichung (3.22) mit (3.15) plus (3.21), unter Berücksichtigung des Verschiebestroms $j_{dis}(x_r)$, beschreibt das Pendant dazu auf der Kathodenseite bei $x = w + w_{pb}$. Die beiden Bestimmungsgleichungen lauten also:

$$j_A = \left(\frac{p_l(b+1) + bN_{Buf}}{b(p_l + N_{Buf})} \right) \left[j_{n0} \frac{p_l(p_l + N_{Buf})}{n_i^2} - q_0 D_{AI} \frac{\partial p}{\partial x}(x_l) \right] \quad (3.26)$$

$$j_K = (b+1) \left[j_{p0} \frac{n_{pr}(n_{pr} + N_{pb})}{n_i^2} + \frac{Q_{pbK}}{\tau_{npb}} + \frac{dQ_{pbK}}{dt} + j_{dis}(x_r) - q_0 D_A \frac{\partial n}{\partial x}(w) \right]. \quad (3.27)$$

Letztlich lassen sich diese beiden Gleichungen wieder ähnlich verkürzen wie in Abschnitt 2.5, d.h.

$$j_A = f_1(p_l, n_{pr}) \quad (3.28)$$

$$j_K = f_2(p_l, n_{pr}). \quad (3.29)$$

Das Modell wird mit Hilfe der Systemvariablen $m_{l,r}$ solange iteriert und damit p_l und n_{pr} eingestellt, bis die beiden genannten Funktionen erfüllt sind.

Für den Spannungsabfall von der Anode zum Gate U_{AG} bei eingeschaltetem GCT erhält man

$$U_{AG} = U_{jl}(0) + U_{jr}(w) + U_\nu \quad (3.30)$$

mit U_ν gemäß (2.84). Die Gate-Kathoden-Spannung berechnet sich zu

$$U_{GK} = U_{jK}(w + w_{pb}) \quad (3.31)$$

Die Kirchhoff'sche Maschenregel fordert zwingend

$$U_{AK} = U_{AG} + U_{GK} = U_{jl} + U_{jr} + U_{jK} + U_{\nu} \quad (3.32)$$

Dabei gilt für die drei genannten Übergangsspannungen

$$U_{jl} = \frac{k_B T}{q_0} \ln \left(\frac{p_l N_{Buf}}{n_i^2} \right) \quad (3.33)$$

$$U_{jr} = \frac{k_B T}{q_0} \ln \left(\frac{p_r N_{\nu}}{n_i^2} \right) \quad (3.34)$$

$$U_{jK} = \frac{k_B T}{q_0} \ln \left(\frac{n_{pr} N_{pb}}{n_i^2} \right), \quad (3.35)$$

wobei diese drei Gleichungen alle aus (1.116) folgen, da es sich bei allen drei Übergängen um pn -Übergänge handelt.

4 Validierung des IGCT-Modells

4.1 Modellanpassung

In der Dissertation von Vogler [73], wurde die Vorgehensweise zur Parameterextraktion beim entsprechenden Dioden- und IGBT-Modell, in einem Ablaufplan zusammengefaßt. Dies ist aufgrund der höheren Komplexität beim vorliegenden GCT-Modell leider nicht möglich. Anstelle dieser Vorgehensweise, wird hier beschrieben, welchen Einfluß die wichtigsten Parameter, auf das Verhalten der GCT-Transienten haben. Sowohl beim GTO als auch beim GCT ist das Abschalten hinsichtlich der dissipierten Verluste, der kritischere Schaltvorgang. Außerdem werden beim IGCT aufgrund des harten Abschaltens ohne Entlastungsnetzwerk, die Transienten im wesentlichen vom Bauelement selbst und nicht von der äußeren Beschaltung bestimmt. Allgemein sei daran erinnert, daß das Abschaltverhalten im wesentlichen von der Menge und der örtlichen Verteilung der Ladungsträger, die im eingeschalteten Zustand in der Driftzone gespeichert werden, abhängt. Viele der maßgeblichen Modellparameter haben letztlich Einfluß auf diese Ladungsträgerverteilung.

In Anhang A ist der Kopf des Modells und die Bedeutung der wichtigsten Parameter beschrieben. Man kann diese Parameter in mehrere Kategorien unterteilen:

K1 Geometriedaten, wie die Anodenfläche A , das Verhältnis der Kathoden- zur Anodenfläche vk , die Pufferzonenweite w_{buf} , die Gesamtweite von Puffer plus Driftzone w und die Weite der p -Basis w_p .

Technologiedaten, wie die Dotierdichten der ν -Zone N_b , der Pufferschicht N_{buf} und der p -Basis N_p . Außerdem gehören die Parameter, die ein Diffusionsprofil beschreiben, d.h. x_{j_l} , N_{0_l} , x_{j_r} , N_{0_r} , in diese Kategorie.

K2 Fitparameter: Hierzu zählen vor allem die Sperrsättigungsströme i_{n0} und i_{p0} von Anode und Kathode zur Festlegung der Injektionsfähigkeit und damit der Randkonzentrationen der Überschußladungsträger. Weiterhin zählen die Hochinjektionslebensdauern τ_{uh0} und die Lebensdauer in der p -Basis τ_{aup} zu den Fitparametern. Die Pufferlebensdauer wird ortsabhängig eingegeben. Dazu benötigt man die Vektorparameter $x_{\tau}[i]$ und $\tau_{rel}[i]$. Damit können auch noch zusätzliche Lebensdauertöpfe, falls vorhanden, definiert werden.

- K3 Parameter zur Berücksichtigung physikalischer Effekte. Dazu zählen boolsche Parameter wie `aug`, `ehs` bzw. `freeze`, die den Augereffekt, die Elektronen-Löcher-Streuung, bzw. das Ausfrieren von Ladungsträgern, berücksichtigen oder nicht. Weiterhin zählen dazu die Parameter `Ecrit` und `avf`, die den Lawineneffekt in der Driftzone beschreiben. Den Durchbruch des Gate-Kathodenübergangs beschreibt `Ugk_av`. Der Parameter `fep`, legt fest, welchen (kleinen) Anteil der Blockierspannung die p -Basis aufnimmt.
- K4 Parameter zur Temperaturbeschreibung: dazu zählen die Struktur- und Technologiedaten der Randzonen `wa`, `wc`, `Na`, `Nk`. Dazu gehören weiterhin die Parameter `alpha_n`, `alpha_p`, `zeta_n`, `zeta_p`, und `gamma`, die in die Formeln diverser Temperaturabhängigkeiten eingehen. Die Auswahl zwischen zwei Temperatugesetzen betreffs $\tau(T)$, ermöglicht `tau_law`. Die globale Temperatur wird dabei mit `T` angegeben.
- K5 sonstige Parameter, wie `plot`, `pow`, `calc`, `fc` (Bedeutung s. Anhang A).

Zur Beschreibung der Einflußparameter dienen im folgenden die Markierungen in Bild 4.1.

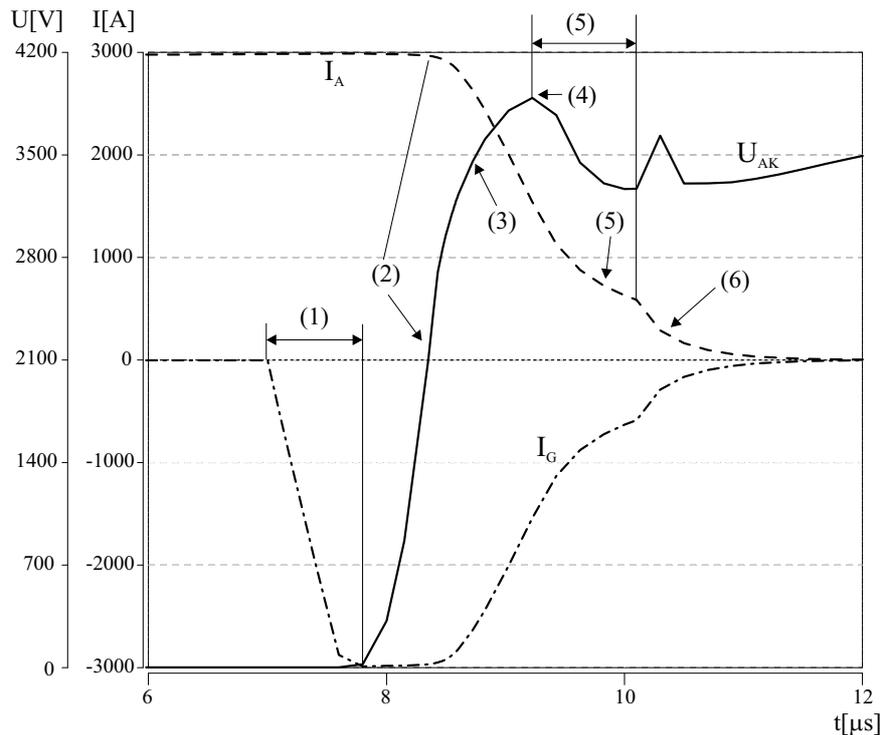


Bild 4.1: Abschaltverhalten eines 3kA/4,5kV-IGCT ohne Snubber; Markierungen zum Erläutern der Einflußparameter.

Die Parameter der Kategorie K1 sollten möglichst genau bekannt sein. Dies gilt besonders für die effektiven Flächen. Dabei können die Dotierdichten am ehesten

zur Feinjustierung genutzt werden. Im Mittelpunkt des eigentlichen Anpaßvorgangs stehen die Parameter der Kategorie K2. Für die meisten Parameter der Kategorie K3 können Standardwerte verwendet werden, wobei die Parameter **Ecrit** und **avf** gut zur Feinanpassung des Lawineneffekts verwendet werden können. Die Parameter der Kategorien K4 und K5 sind zur Anpassung des Modells weniger interessant und werden hier nicht weiter betrachtet. Außerdem muß man noch die äußere Beschaltung gut kennen. Im folgenden entsprechen die Numerierungen den Markierungen in Bild 4.1.

Zu (1): Hier ist die Zeitspanne gemeint, die zwischen dem Einsetzen des Abschaltimpulses und dem Beginn der Spannungsaufnahme vergeht¹. Die Spannungsaufnahme kann beginnen, wenn $p(x_r)$ gerade verschwunden ist, oder mit anderen Worten: Die rechtsseitige Randkonzentration zu Beginn des Abschaltvorgangs $p(x_r, t = 0) = p_r(t = 0)$ ist ein Maß für den genannten Zeitraum! Je höher $p_r(t = 0)$ ist, desto länger die Verzögerung. Die beiden Parameter zum Einstellen von p_r sind **ip0** und **taunp**. Mit **ip0** wird die Elektroneninjektion der Kathode in die p -Basis, d.h. n_{pr} , eingestellt. Je niedriger **ip0** ist, desto höher ist n_{pr} und damit mittelbar p_r . Die Lebensdauer **taunp** geht in die Diffusionslänge der p -Basis ein. Je höher **taunp** ist, desto größer wird p_r . Weitere Parameter, die p_r stark beeinflussen sind **Np** und **wp**, die aber bekannt sein sollten. Manchmal empfiehlt es sich dennoch, **Np** und **wp**, zur Feinjustierung zu verwenden. Dies kann nötig werden, wenn p_r über einem weiten Strombereich, z.B. $I_{Last} = 500 - 3000A$, richtig berechnet werden soll. Je kürzer man **wp** und je kleiner man **Np** wählt, desto größer ist p_r . Ferner spielt auch noch der statische Gatestrom I_G eine gewisse Rolle. Je größer dieser ist, desto mehr Elektronen werden injiziert und desto größer wird folglich p_r .

Zu (2): Der Spannungsanstieg dU/dt hängt von der gespeicherten Ladungsmenge ab. Neben dem bereits eingestellten Randwert p_r aus (1), spielt somit der linksseitige Randwert p_l und der Durchhang von $p(x)$ eine große Rolle. Dabei steigt die Spannung umso langsamer, je größer p_l ist und je geringer $p(x)$ durchhängt. Zum Einstellen von p_l (Löcherinjektion von der Anode in den Puffer) verwendet man **in0**, den Sperrsättigungsstrom der Anode. Je kleiner **in0** gewählt wird, desto größer wird p_l . Um ein betragsgroßes Minimum von $p(x)$, d.h. geringen Durchhang, zu bekommen, muß man **tauh0** möglichst groß wählen, wobei **tauh0** in einem gewissen Maße auch p_l und p_r beeinflusst. Die Summe aus Driftzonenweite und Pufferweite **w** muß bekannt sein und sollte entsprechend verwendet werden.

Zu (3): Falls der dynamische Avalanche wirkt, verringert sich der Spannungsanstieg dU/dt . Durch den Lawineneffekt wird auch das dI/dt verringert. Die Parameter zum Einstellen dieses Effekts sind **Ecrit** und **avf**. Verringert man **Ecrit**, beginnt der verminderte Spannungsanstieg früher, bzw. bei einer niedrigeren Spannung. Mit

¹nicht zu verwechseln mit der Speicherzeit, die anders definiert ist (s.S. 89 und [1]).

`avf` wird der Grad der Abflachung von dU/dt eingestellt, wobei sich die Abflachung umso stärker bemerkbar macht, je geringer `avf` ist.

Zu (4): Das Spannungsmaximum hängt natürlich auch von den beiden Parametern, die den Lawineneffekt beschreiben ab (s. (3)). Ansonsten ist es wieder die Löcherverteilung in der ν -Zone, die hier zum Tragen kommt. Wie man $p(x)$ verändern kann, wurde bereits im Punkt (2) erläutert. Ein weitere wichtige Einflußgröße ist die parasitäre Hauptkreisinduktivität, die man zumindest als Schätzung kennen sollte.

Zu (5): Der Abstand zwischen dem eigentlichen Spannungsmaximum und der zweiten Spannungsspitze, die von der Funktion der Pufferzone als Feldstoppschicht verursacht wird, hängt wieder von $p(x)$ ab. Dabei gilt: je höher $p(x)$, desto schwerer tut sich die Raumladungszone Richtung Puffer vorzudringen und die besagte Spitze tritt später auf. Auch der Parameter `Nb` (Dotierung der ν -Zone) hat hierauf einen starken Einfluß. Besonders das dI/dt kurz vor der Spitze läßt sich damit fein anpassen. Wird `Nb` verkleinert, wird die Spitze auf der t -Achse nach links verschoben und das dI/dt kurz vor der Spitze verkleinert (I -Kurve flacher). Dabei hat dann der Knickpunkt im Strom einen größeren Betrag auf der I -Achse.

Zu (6): Der Diffusionsschwanz (Tailstrom) wird bei einem Punch-Through-Bauelement von der Speicherladung in der Pufferschicht alleine bestimmt. Somit sind die Pufferparameter `wbuf`, `Nbuf` und τ_{Buf} dafür verantwortlich. Da p_l in den Puffer injiziert wird, spielt diese Größe auch hier eine entscheidene Rolle. Unter der Annahme, daß `wbuf` und `Nbuf` bekannt sind (Feineinstellung!), bestimmt L_{Buf} über τ_{Buf} , wie stark die Löcherdichte im Puffer zur ν -Zone hin abfällt. Je geringer τ_{Buf} ist, desto stärker fällt $p(x)$ ab und desto niedriger ist die Pufferladung insgesamt, was bedeutet, daß der Tailstrom schneller abklingt. Allerdings wird damit auch festgelegt, wieviel Ladung $p_{\nu l}$ am linken ν -Zonenrand übrigbleibt. Somit beeinflussen die Pufferparameter letztlich auch (2), (3), (4) und (5). Die Lebensdauer τ_{Buf} wird als Topf eingegeben. Als Parameter stehen dazu die Vektoren `x_tau[i]` und `tau_rel[i]` zur Verfügung.

Wie aus (1) bis (6) hervorgeht, steht und fällt der Anpassungsprozeß mit der Ladungsträgerverteilung $p(x)$. Damit wird aber schnell klar, daß die hier definierten Transientenabschnitte (1) bis (6) nicht unabhängig voneinander angepaßt werden können, worin die eigentliche Schwierigkeit der Anpaßarbeit liegt. Je genauer man anpassen möchte und umso weniger Informationen man über das Bauelement besitzt, desto mehr Zeit wird der Anpaßvorgang in Anspruch nehmen. Kennt man dagegen die Daten des GCTs recht gut und hat neben Anodenströme und Blockierspannungen auch die Gateströme als Meßkurven zur Verfügung, kann man mit ein wenig Übung und Erfahrung mit Anpaßdauern von unter einem Arbeitstag rechnen.

4.2 Validierung anhand von Messungen

4.2.1 Meßschaltung und gemessene GCT–Typen

Bild 4.2 zeigt die einfache Schaltung, mit der sowohl die Messungen², als auch die Simulationen durchgeführt wurden. Hierbei handelt es sich um einen Tiefsetzsteller mit Clampkreis. Der Clampkreis hat die Aufgabe, die große Induktivität L_S (einige μH), die zum Einschalten notwendig ist (turn on–snubber), bezüglich der Überspannung beim Abschalten kurzzuschließen. Die Last wird durch eine ideale Konstantstromquelle modelliert, da sich in dem kurzen Zeitraum während des Schaltens, der Laststrom aufgrund der hohen Lastinduktivität (einige mH) ohnehin nicht ändert. Weiterhin wird mit L_{rest} die unvermeidbare und für die Überspannung verantwortliche parasitäre Restinduktivität modelliert, deren Betrag meist wenige $100nH$ beträgt. Eine geeignet parametrisierte Freilaufdiode (Model nach [59]) ist ebenfalls enthalten³.

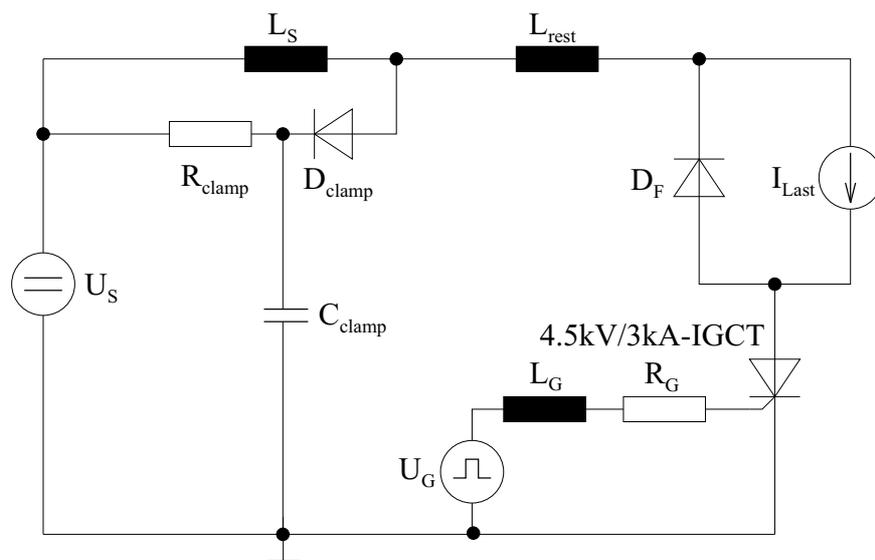


Bild 4.2: Meß- und Simulationsschaltung: Tiefsetzsteller mit Clampkreis.

Als Meßergebnisse stehen dem Autor transiente Ein- und Ausschaltverläufe bei verschiedenen Strömen gegen $U_S = 2800V$ und die Kennlinie zur Verfügung. „Device under Test“ waren dabei die beiden $3kA/4,5kV$ IGCT–Typen 5SHY35L4502 und 5SHY35L4503 der Firma ABB, die sich vor allem in der Menge der gespeicherten Ladungsträger bei gleichem Strom unterscheiden. Dabei enthält der Typ 5SHY35L4502

²durchgeführt bei ABB Semiconductors AG in Lenzburg, Schweiz

³Diese Aussage gilt auch für die Clampdiode

deutlich mehr Speicherladungen, was in niedrigeren Durchlaßspannungen resultiert. Dafür sind die Abschaltverluste erheblich höher als beim 5SHY35L4503.

4.2.2 Abschaltverhalten

Anhand von Messungen des Abschaltverhaltens wurden die Parameter für den IGCT 5SHY35L4503 in den oben beschriebenen sechs Schritten angepaßt. Dabei waren die Parameter der Kategorie K1 (s. Abschnitt 4.1) bekannt. Besonders günstig ist dabei die Kenntnis des Gatestroms, da man einerseits den genauen Zeitpunkt des Abschaltimpulses kennt und andererseits mit dem dI_G/dt , die effektive Gateinduktivität L_G gut abschätzen kann. Der Anpaßvorgang hat sich zunächst an der Messung für $I = 3kA$ (Bild 4.3) orientiert. Die erhaltene Erstanpassung stimmt dann für $3kA$, wird aber in der Regel für kleiner werdende Ströme ungenauer. Dies liegt an der noch ungenügenden Beschreibung der anoden- und kathodenseitigen Ladungsträgerinjektion. Im folgenden müssen also noch die richtigen Parametersätze (in_0 , $\tau_{rel}[i]$, N_{buf}) und (ip_0 , τ_{anp} , N_p , w_p) gefunden werden (Feinanpassung). Dazu wurde die Messung für $1,5kA$ verwendet (s. Bild 4.5). Das endgültige Anpaßergebnis zeigen die Bilder 4.3 bis 4.6, die die Simulationsergebnisse der Abschalttransienten verschiedener Ströme (durchgezogene Linien) im Vergleich zu den jeweils passenden Meßergebnissen (gestrichelte Linien) zeigen.

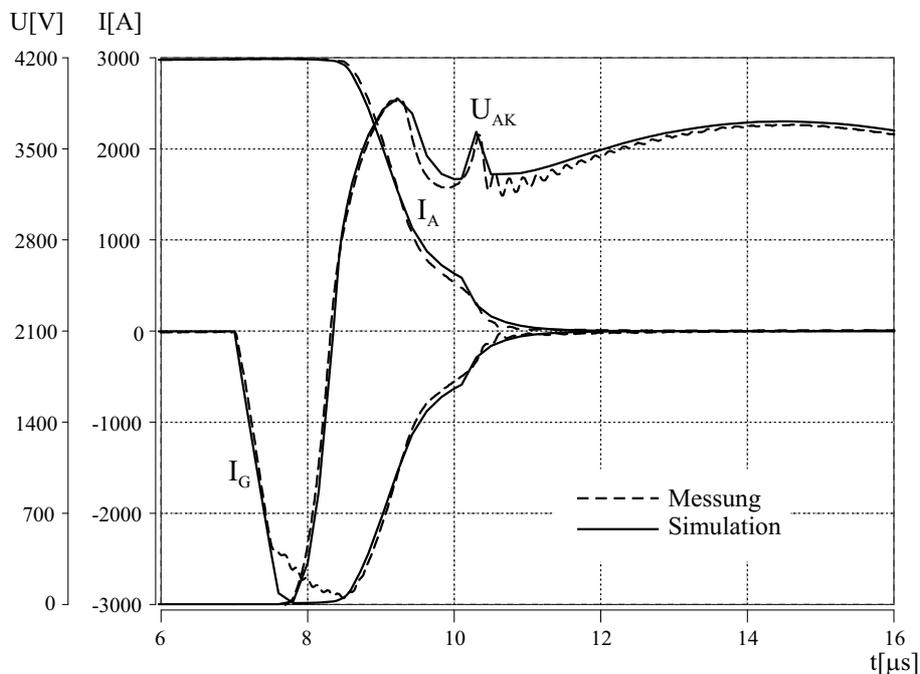


Bild 4.3: hartes unentlastetes Abschalten von $3kA$ gegen $2,8kV$ des GCT 5SHY35L4503: Messung vs. Simulation

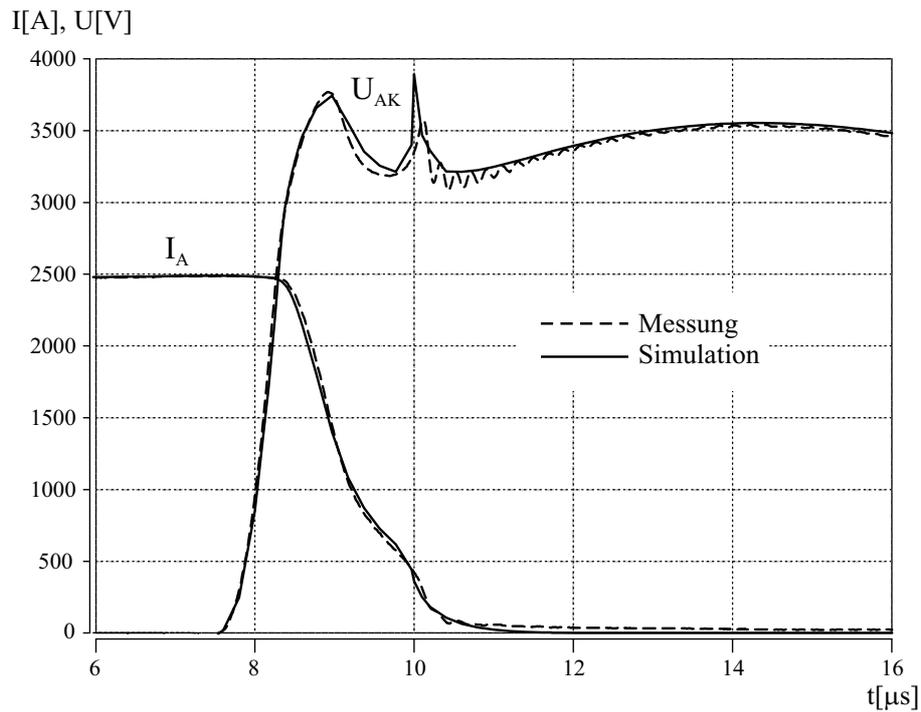


Bild 4.4: hartes unentlastetes Abschalten von 2,5kA gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

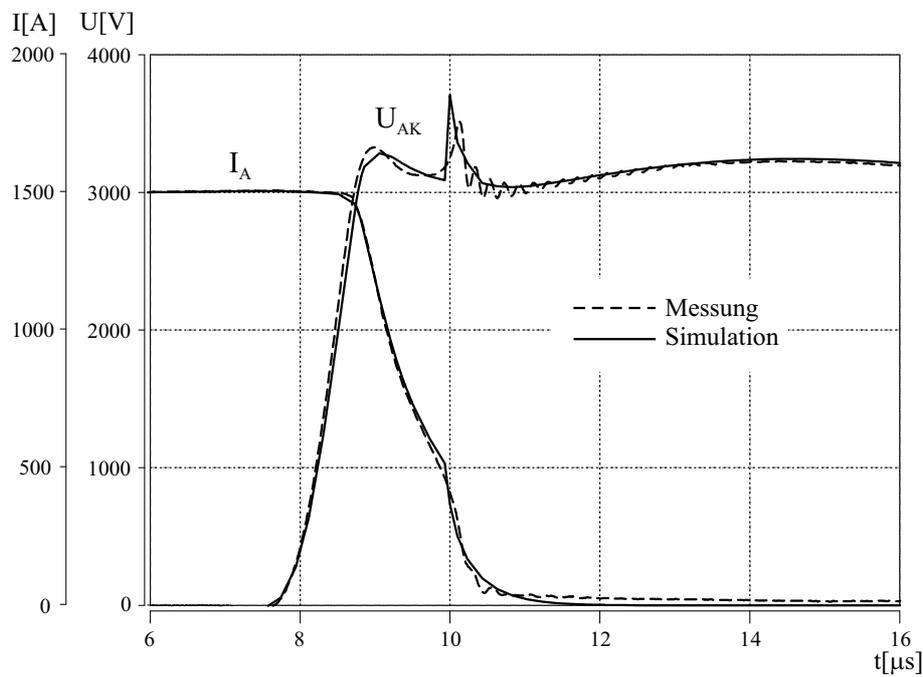


Bild 4.5: hartes unentlastetes Abschalten von 1,5kA gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

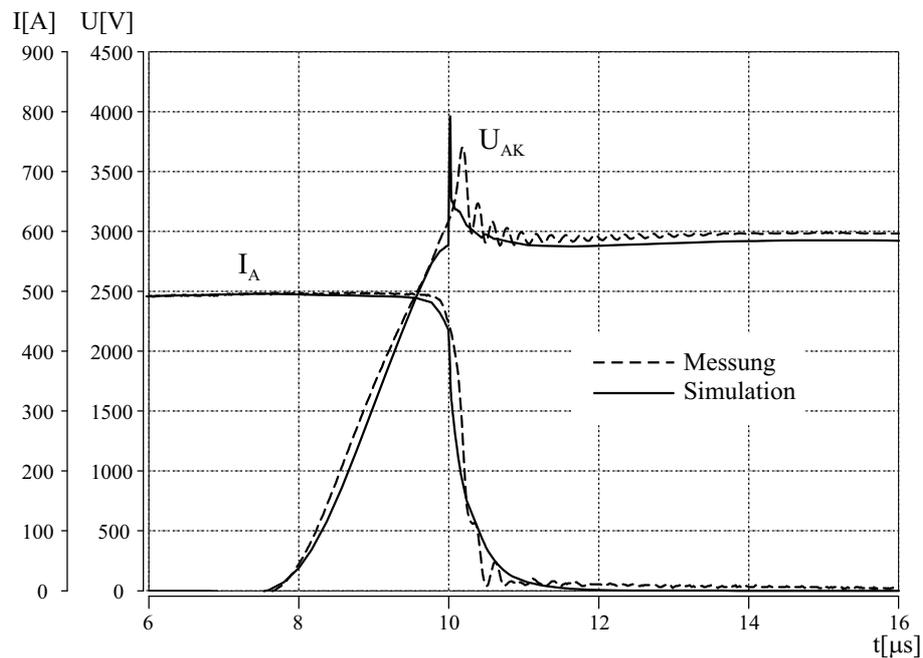


Bild 4.6: hartes unentlastetes Abschalten von 500A gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

Abschließend zeigt Bild 4.7 einen Vergleich der gemessenen und simulierten Verläufe für die dissipierte Energie eines Abschaltvorgangs für die vier betrachteten Ströme.

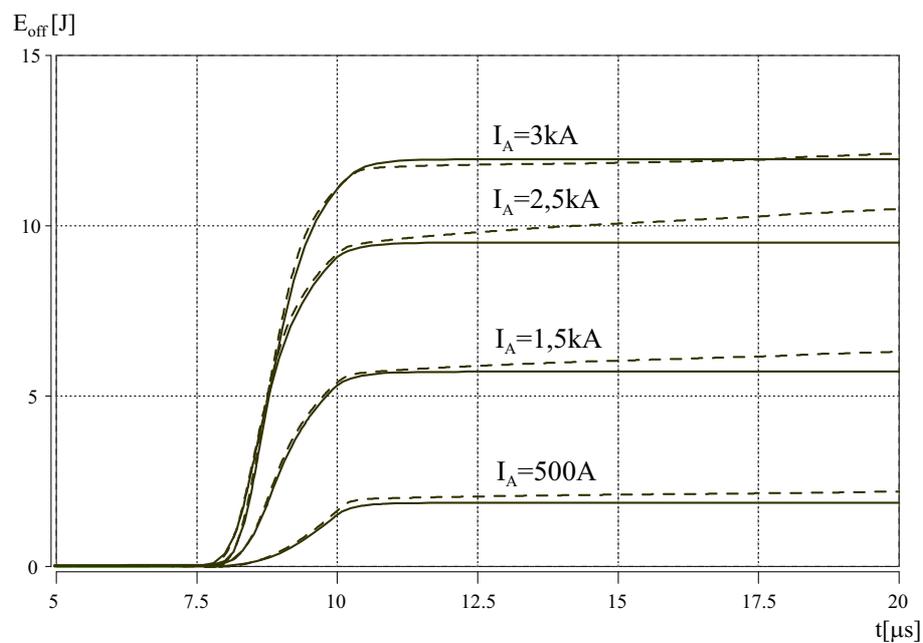


Bild 4.7: Energiedissipation je Abschaltvorgang zu den Bildern 4.3 bis 4.6

Man sieht in den Bildern, daß die Simulationsergebnisse mit dem so parametrisierten Modell über einen weiten Strombereich (500 – 3000A) hervorragend mit der Realität übereinstimmt. Etwaige Unterschiede in Bild 4.7 nach $t = 12\mu s$, beruhen darauf, daß in der Messung ein spürbarer Offset $I > 0$ (Meßfehler!) nach dem Abschaltvorgang verbleibt und $E_{off}(t)$ deshalb leicht linear ansteigt. Übrigens benötigt SABER auf einem PC mit einer 500MHz-CPU, ca. 20s für einen Abschaltvorgang im gezeigten Genauigkeitsgrad⁴.

Die Bilder 4.8 bis 4.11 zeigen die Abschalttransienten des IGCT 5SHY35L4502 für $I = 1kA, 2kA, 3kA$ und $3,3kA$ gegen $U = 2,8kV$. Dabei werden die Messungen von den gepunkteten und die Simulationen von den durchgezogenen Linien dargestellt. Für $I = 1kA$ (Bild 4.8) ist der Spannungsanstieg dU/dt vor dem (ersten) Maximum nahezu konstant und die aufgenommene Blockierspannung wächst somit linear. Folglich spielt der dynamische Lawineneffekt hier noch keine Rolle. Ab ca. $I = 2kA$ setzt dann der Lawineneffekt ein. Dazu zeigt Bild 4.9 eine zusätzliche Simulation (gestrichelte Linie), die man erhält, wenn man den Lawineneffekt im Modell nicht berücksichtigt, d.h. E_{crit} unrealistisch hoch wählt. Die Abweichungen zwischen den Simulationen mit und ohne Avalanche werden für noch größere Ströme immer augenfälliger (s. Bilder 4.10 und 4.11).

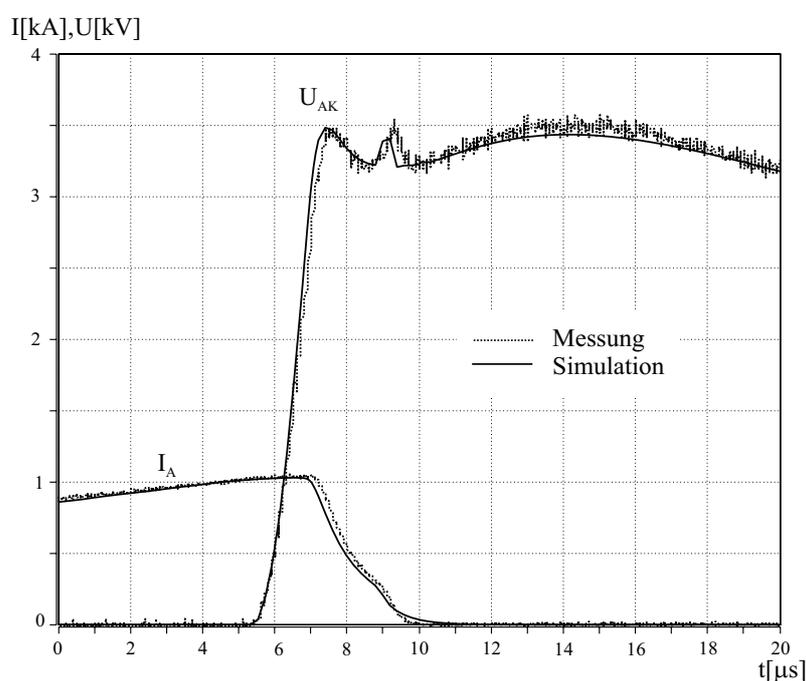


Bild 4.8: hartes unentlastetes Abschalten von 1kA gegen 2,8kV des GCT 5SHY35L4502: Messung vs. Simulation

⁴Um „schöne“ Kurven zu bekommen, wurden die Genauigkeitsparameter im Vergleich zur Voreinstellung erheblich erhöht. Ansonsten lassen sich zwar Simulationszeiten deutlich unter 10s realisieren, allerdings mit entsprechend größeren Zeitschrittweiten und weniger glatten Transienten.

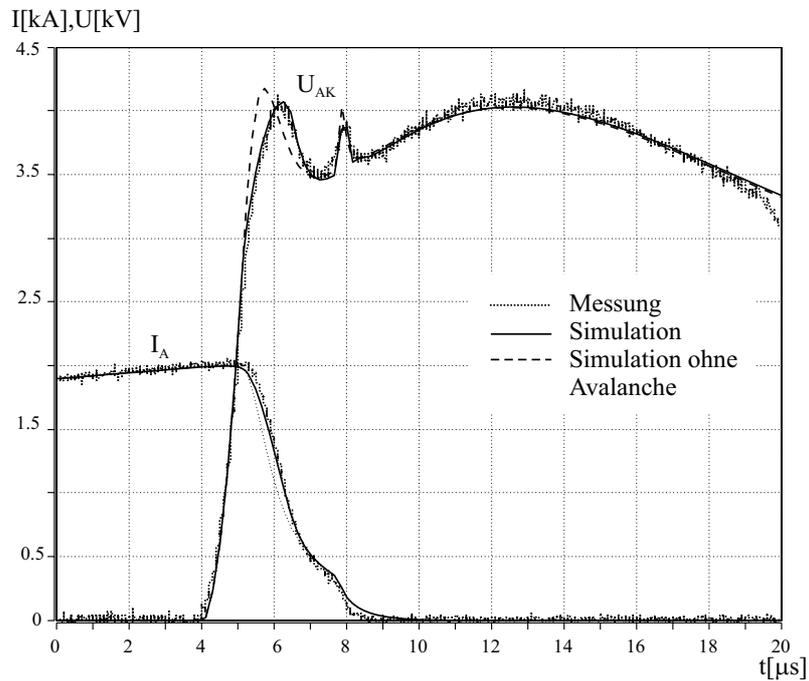


Bild 4.9: hartes unentlastetes Abschalten von 2kA gegen 2,8kV des GCT 5SHY35L4502: Messung vs. Simulation

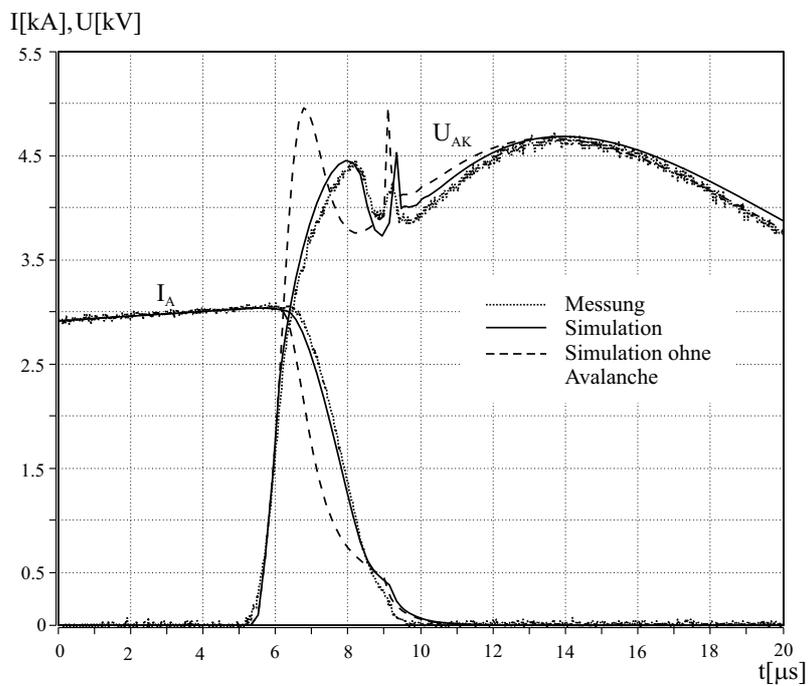


Bild 4.10: hartes unentlastetes Abschalten von 3kA gegen 2,8kV des GCT 5SHY35L4502: Messung vs. Simulation

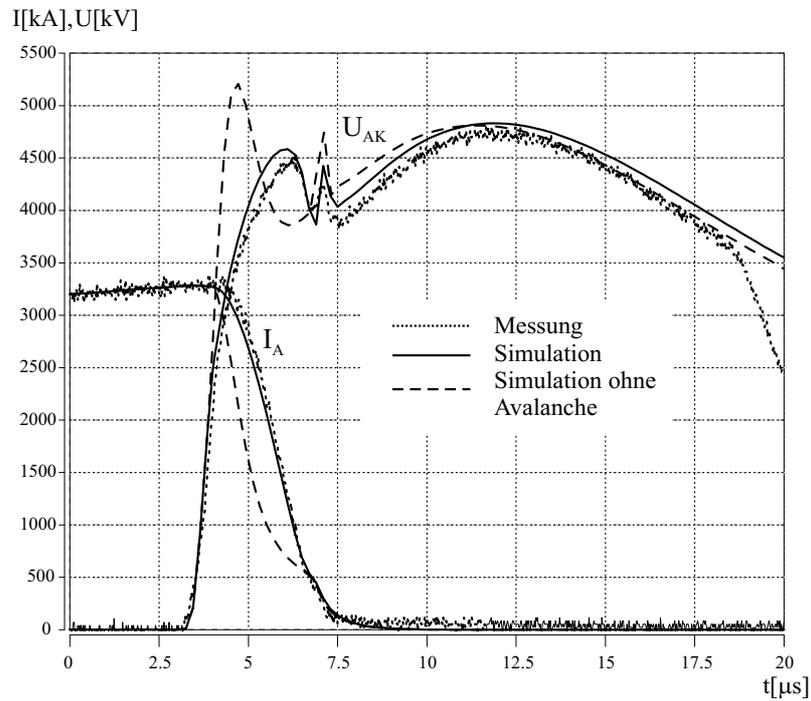


Bild 4.11: hartes unentlastetes Abschalten von $3,3\text{kA}$ gegen $2,8\text{kV}$ des GCT 5SHY35L4502: Messung vs. Simulation

Bild 4.12 zeigt den Vergleich der gemessenen und simulierten Verläufe für die dissipierte Energie eines Abschaltvorgangs, mit und ohne Lawineneffekt, für $I = 1, 2, 3\text{kA}$.

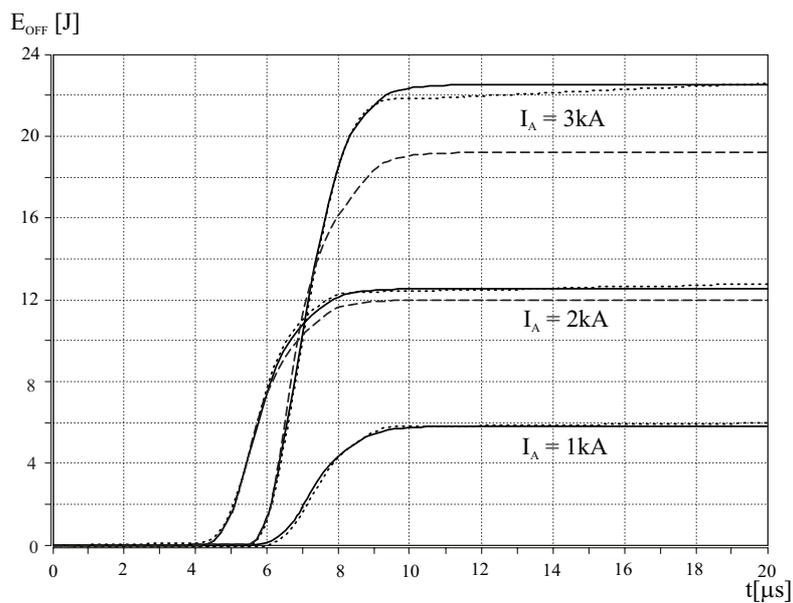


Bild 4.12: Energiedissipation je Abschaltvorgang zu den Bildern 4.8 bis 4.10

Vergleicht man Bild 4.12 mit Bild 4.7, stellt man fest, daß im IGCT 5SHY35L4502 erheblich höhere Verlustenergien auftreten, als beim IGCT 5SHY35L4503. Wie eingangs erwähnt liegt dies an der höheren Ladungsträgerdichte in der ν -Zone bei gleichem Strom. Je höher der betrachtete Strom ist, desto größer ist auch die prozentuale Abweichung bzgl. der Verluste der betrachteten IGCT-Typen. Dies liegt daran, daß beim IGCT 5SHY35L4502 neben der höher Überschußträgerdichte auch noch der dynamische Lawineneffekt stärker wirkt, als beim IGCT 5SHY35L4503.

Ähnlich wie oben, stimmen auch hier die Simulationen und die Messungen sehr gut überein, inklusive der Pufferbeschreibung (zweite Spannungsspitze). Weiterhin konnte anhand der Bilder 4.10, 4.11 und 4.12 gezeigt werden, wie wichtig die Modellierung des Lawineneffekts ist. Das Verhalten der Überschußladungsträger während des Abschaltens kann vom Modell ebenfalls ausgegeben werden (vgl. Bild 3.10). Abschließend sei noch bemerkt, daß aufgrund des nicht bekannten Gatestromverlaufs, hier die Anpaßprozedur etwas schwieriger und langwieriger war, als oben beim IGCT 5SHY35L4503.

4.2.3 Einschaltverhalten und Kennlinie

Der zeitliche Verlauf des Einschaltstroms ist weitgehend unabhängig vom Verhalten des IGCTs. Der Stromanstieg dI/dt wird von der Snubberinduktivität L_S (s. Bild 4.2) bestimmt. Der Überstrom entspricht dem Reverse Recovery-Strom der Freilaufdiode mit umgekehrten Vorzeichen. Somit ist die Freilaufdiode das kritische Bauelement des Einschaltvorgangs und auch an dessen Bedürfnisse orientiert sich die Dimensionierung der Snubberinduktivität L_S . Allerdings kann beim Einschalten mit einem vergleichsweise geringen Gatestrom, das Wechseln der Spannung vom GCT zur Snubberinduktivität L_S verlangsamt werden, d.h. das dU/dt wird kleiner, was auch den Verlauf des Überstroms beeinflußt. Im Bild 4.13 und 4.14, ist dies durch einen gestrichelten Kreis kenntlich gemacht. An dieser Stelle gibt es einen leichten Knick, d.h. das dI/dt wird geringer und der Überstrom klingt langsamer ab. Bei einem größeren Gatestrom ($\hat{I}_G > 1kA$), wird der Überstrom letztlich nur noch von der Diode bestimmt und entspricht dann einem reinen Reverse Recovery-Strom. Die Bilder 4.13 und 4.14 zeigen einen Vergleich gemessener und simulierter Transienten zweier Einschaltvorgänge für $I = 3kA$ bzw. $I = 1,5kA$. Wieder stimmen Experiment und Rechnung gut überein.

Bild 4.15 zeigt, wie sich die Überschußladung in Abhängigkeit von der Zeit mit wachsendem Strom aufbaut. Im Gegensatz zum Abschaltverhalten, wo auch rechtsgekrümmte Kurven vorkommen, sind beim Einschalten alle Plasmakurven wie im statischen Durchlaßfall, linksgekrümmt (vgl. Diskussion in Abschnitt 3.4). Bild 4.16 zeigt abschließend die Kennlinie des IGCT 5SHY35L4502.

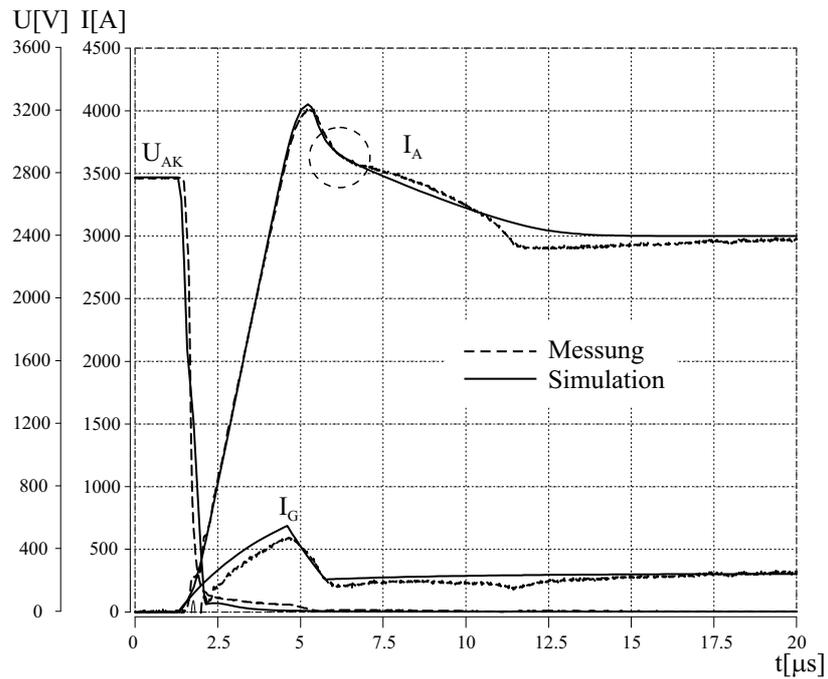


Bild 4.13: Einschalten von 3kA gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

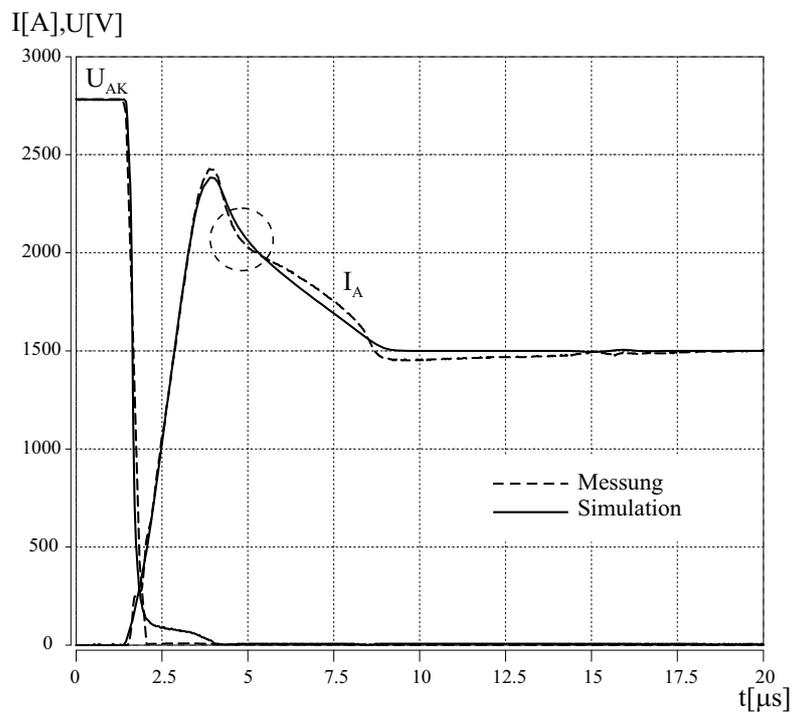


Bild 4.14: Einschalten von 1,5kA gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

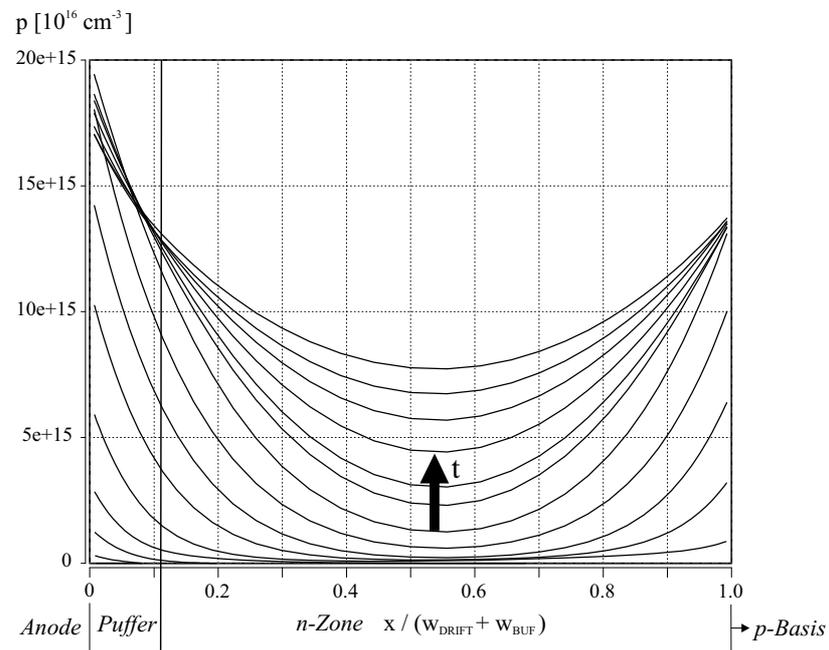


Bild 4.15: Plasmaverteilungen im Laufe des Einschaltvorgangs eines IGCT

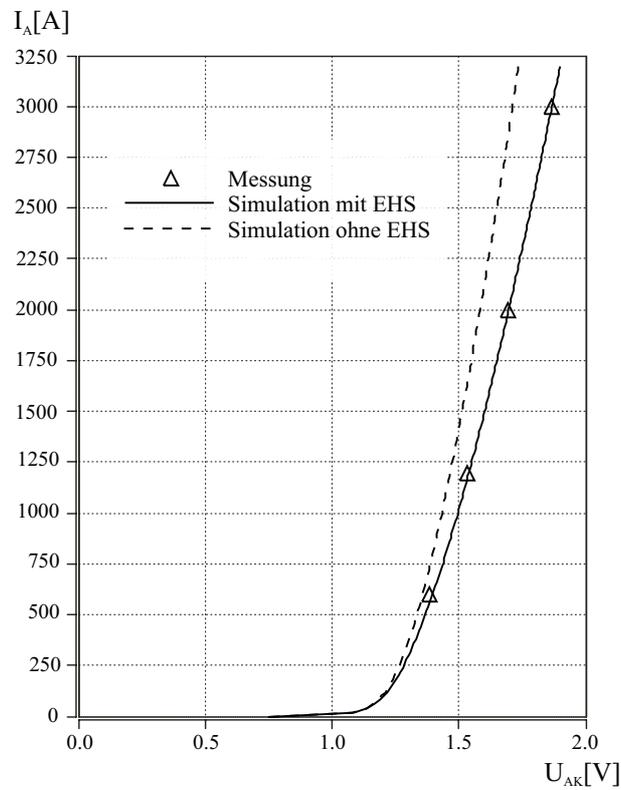


Bild 4.16: gemessene und simulierte Kennlinie des IGCT 5SHY35L4502

Dabei werden die Meßwerte durch die Dreieckssymbole gekennzeichnet. Die durchgezogene Linie entspricht der Simulation mit dem in Abschnitt 4.2.2 parametrisierten Modell. Zusätzlich ist noch ein kleiner Serienwiderstand (Kontaktwiderstand) von ca. $70\mu\Omega$ notwendig. Ferner ist noch eine zusätzliche Simulation abgebildet, bei der die Elektronen–Löcher–Sreuung nach Mnatsakanov (vgl. Abschnitte 1.5.5 und 2.3.1) deaktiviert wurde und von der Messung entsprechend abweicht (gestrichelte Linie). Der zu geringe Spannungsabfall ist auf den fehlenden Zusatzterm in der Bestimmungsgleichung für den statischen Spannungsabfall (2.52) zurückzuführen.

4.2.4 IGCT–Modell versus Modell aus der SABER–Bibliothek

Bisher wurde gezeigt, daß sich das physikalische Modell hervorragend an Messungen über einen weiten Strombereich anpassen läßt und damit aussagekräftige Simulationsergebnisse auch bei komplexeren Schaltungen, hinsichtlich den im Einführungskapitel erwähnten Fragestellungen, zu erwarten sind. Angesichts der hohen Modellkomplexität kann aber die Frage aufkommen, ob nicht einfachere Modellansätze genügen, um genügend genau simulieren zu können. Um dieser Frage nachzugehen, wurde das in der SABER–Bibliothek zur Verfügung gestellte GTO–Modell, bestmöglich an die Messungen des abgeschalteten IGCT 5SHY35L4503 angepaßt. Das Modell basiert auf dem Ansatz konzentrierter Ladungselemente (Lumped Charge–Approach) nach [41]. Dabei wird das Bauelement entlang der x–Achse in eine endliche Anzahl $i = m$ von Boxen (lumps) unterteilt, denen jeweils eine konzentrierte Ladung Q_i zugeteilt wird. Damit können zwischen den Boxen Stromdichtegleichungen ähnlich wie beim quasistatischen Ansatz aufgestellt werden (vgl. (3.9) bzw. (3.21)). Ferner werden geeignete Spannungsterme an den Boxenübergängen definiert. Ein einfaches Modell für den Lawineneffekt ist ebenfalls implementiert.

Bild 4.17 zeigt den Vergleich gemessener und simulierter Transienten des IGCT 5SHY35L4503 für $I = 3kA$ (vgl. Bild 4.3). Dabei wurde in erster Linie der Anodenstrom $I_A(t)$ angepaßt. Bei $U_{AK}(t)$ ergeben sich dann spürbare Abweichungen bei Beginn der Spannungsaufnahme und das dU/dt ist zu groß. Im Bereich kurz vor dem Maximum kann das dU/dt mit dem implementierten Avalanche–Modell etwas verkleinert werden. Wesentlich auffälliger aber sind die Unterschiede beim Gatestrom I_G . Die Zeitverzögerung zwischen Beginn des Abschaltimpulses und dem Beginn der Spannungsaufnahme ist praktisch Null, so daß dieser Impuls in der Rechnung etwa $1\mu s$ später beginnt, als im Experiment. Trotzdem ist das Ergebnis, sofern man überwiegend an den Ausgangstransienten $U_{AK}(t)$ und $I_A(t)$ interessiert ist, durchaus passabel, was die Übereinstimmung der berechneten und gemessenen dissipierten Energie in Bild 4.19 zusätzlich bestätigt. Verändert man allerdings bei dem so parametrisierten Modell die Ströme, wie in Bild 4.18 für $I = 1,5kA$ gezeigt wird, ergeben sich zum Teil starke Abweichungen.

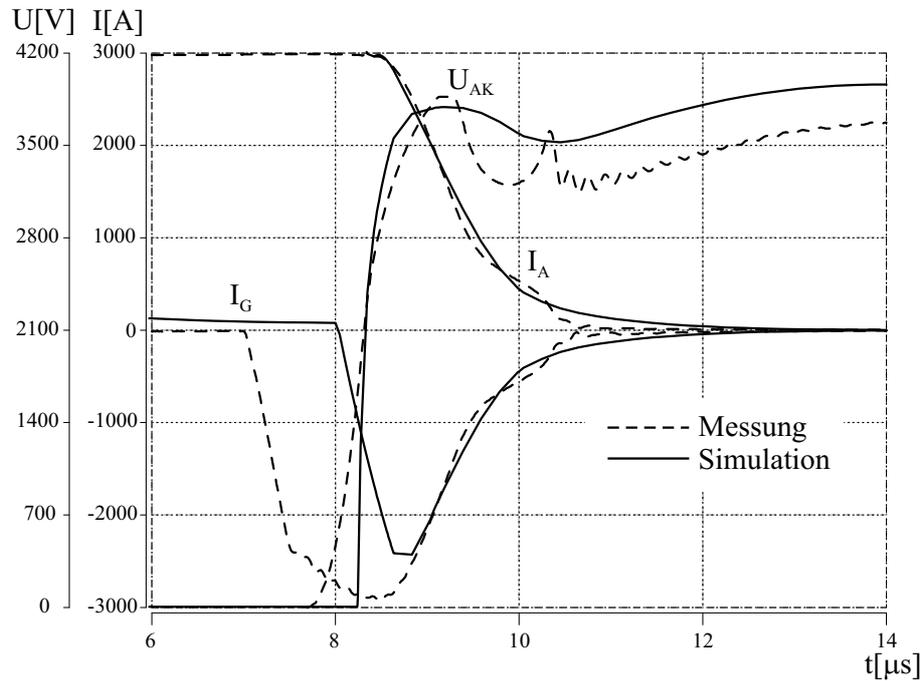


Bild 4.17: hartes unentlastetes Abschalten von 3kA gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

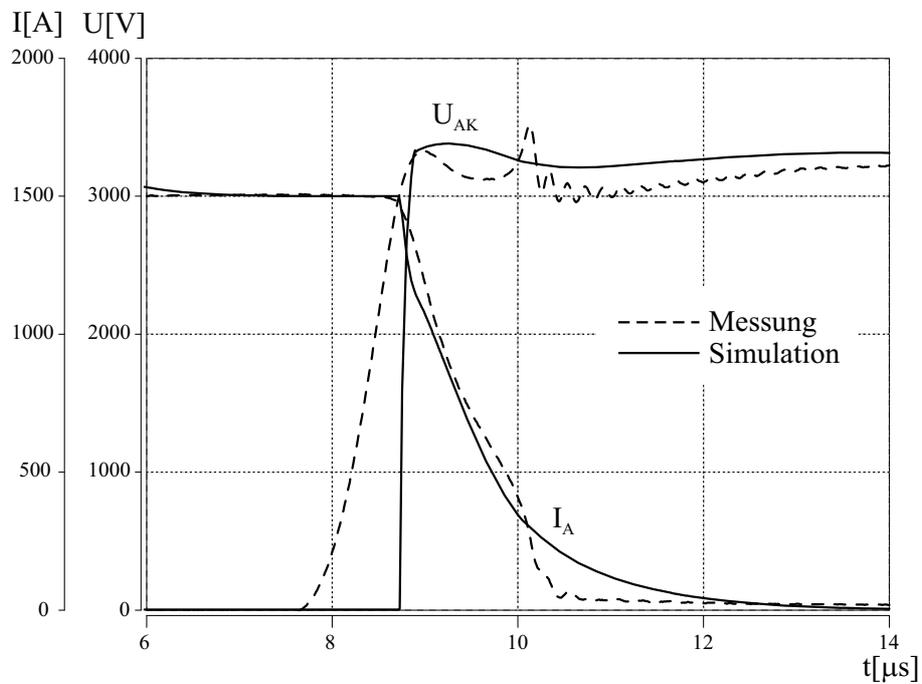


Bild 4.18: hartes unentlastetes Abschalten von 1,5A gegen 2,8kV des GCT 5SHY35L4503: Messung vs. Simulation

Zunächst wurde die Modellparameter etwas nachjustiert (!), um eine Übereinstimmung der gemessenen und simulierten Kurve für $I_A(t)$ zu erreichen. Der Unterschied im Spannungsanstieg ist hier noch viel größer als vorher für $I = 3kA$. Da bei $I = 1,5kA$ der Lawineneffekt noch nicht wirkt, wächst die berechnete Spannung fast senkrecht. Somit kann gesagt werden, daß dieses Modell nur für einen kleinen Strombereich parametrisiert werden kann. Diese Aussage wird durch einen Blick auf die Verhältnisse bzgl. der Verlustenegien in Bild 4.19 bestätigt, vor allem vor dem Hintergrund, daß ohne die oben erwähnte Nachjustierung der Modellparameter, der Unterschied noch größer wäre. Die zweite Spannungsspitze konnte wegen der fehlenden Pufferbeschreibung in diesem Modell, naturgemäß nicht berechnet werden.

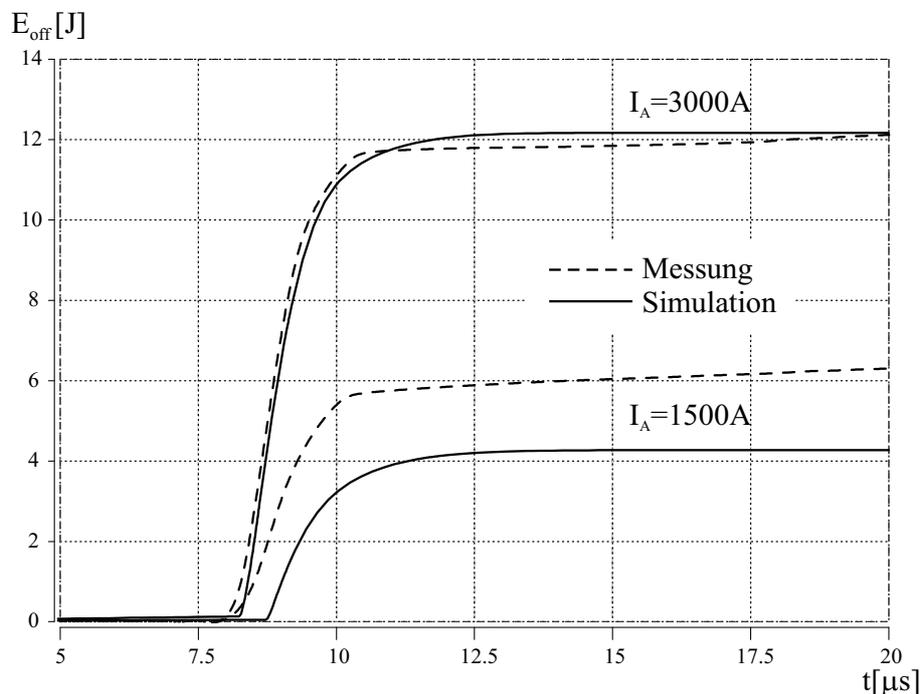


Bild 4.19: *Energiedissipation je Abschaltvorgang zu den Bildern 4.17 und 4.18*

Erwähnenswert ist außerdem, daß die sich Simulationsdauern der beiden Modelle gar nicht so stark unterscheiden, was zum einen daran liegt, daß in beiden Fällen dasselbe physikalische Modell für die beiden Dioden verwendet wird und zum anderen die Genauigkeitsanforderungen der Simulationen erhöht wurde. Letztlich hängt es von der Aufgabenstellung ab, ob man ein physikalisches Modell, ein einfacheres Modell, oder gar nur ideale Schalter zum Simulieren verwendet. Für Aussagen, ob ein Schaltungsdesign prinzipiell funktioniert oder nicht, reichen mit Sicherheit ideale Schalter aus. Möchte man grobe Abschätzungen für Überspannungen und -ströme erhalten, sollten wenigstens „einfache“ Modelle, wie das eben Beschriebene, verwen-

det werden. Für alle genaueren Analysen, wie z.B. Worst Case-Abschätzungen, um beispielsweise die Belastungsgrenzen von Bauelementen in bestehenden Schaltungen zu untersuchen, bieten sich die hochgenauen physikalischen Modelle an. Kombiniert man diese Modelle mit geeigneten Temperaturnetzwerken, können auch in dem für Hochleistungsbauelementen so wichtigem Gebiet der thermischen Belastungsgrenzen, aussagekräftige Rechenergebnisse erzielt werden.

In der Analyse von seriengeschalteten IGCTs, spielt das physikalische Modelle aufgrund seiner Genauigkeit in der Beschreibung, ebenfalls seine Überlegenheit aus. Denn die auftretenden Asymmetrieprobleme bezüglich der aufgenommenen Blockierspannung, können lediglich mit dem physikalischen Modell richtig analysiert werden. Folgerichtig kann das Modell dann auch zum Schaltungsdesign eingesetzt werden, wenn es darum geht, Lösungen zu finden, potentielle Asymmetrien zu vermeiden. Die Analyse der Serienschaltung und die Synthese eines Konzepts zur aktiven Ansteuerung zur Vermeidung von Spannungsunsymmetrien, unter Verzicht großer RC(D)-Entlastungsnetzwerke, sind Themen der nächsten beiden Kapitel.

5 Analyse der Serienschaltung von IGCTs

5.1 Motivation für die Serienschaltung

Die in Kapitel 4 betrachteten IGCT-Typen 5SHY35L4502 bzw. 5SHY35L4503 haben einen maximal abschaltbaren Strom von $3,5\text{kA}$ und eine maximal zulässige Blockierspannung von $4,5\text{kV}$. Dabei handelt es sich aber um eine maximal erlaubte Spannung inklusive auftretender Überspannungen. Für den Zwischenkreis ist eine Spannung U_{ZK} von maximal $2,8\text{kV}$ vorgesehen. Für höhere Spannungen gibt es am Markt noch $5,5\text{kV}$ -Typen mit einer maximal zulässigen Zwischenkreisspannung U_{ZK} von $3,3\text{kV}$, aber einem maximalen Abschaltstrom I_{off} von unter 2kA . Beim GTO gibt es noch etwas leistungsstärkere Bauelemente, wie z.B. den 6kV -Typ 5SGT30J6004 des gleichen Herstellers mit $U_{ZK} = 3,8\text{kV}$ und $I_{off} = 3\text{kA}$ ¹.

Für bestimmte Hochleistungstopologien, wie sie z.B. für Netzkupplungen eingesetzt werden [66], sind höhere Zwischenkreisspannungen gewünscht, als die Einzelelemente maximal zulassen. Ein Ausweg wären IGCTs noch höherer Blockierfähigkeit, wie beispielsweise der in [22] vorgeschlagene 10kV -Typ. Solche IGCTs bräuchten eine entsprechend lange schwach dotierte Driftzone, um diese hohe Spannung blockieren zu können. Driftzonen lassen sich aber nicht beliebig verlängern. Erstens läßt sich das Minimum der Überschußlöcherdichte im statischen Durchlaßfall, z.B. bei Nennstrom, nicht beliebig durch Erhöhung der Lebensdauer oder der Injektionseffizienz der Randzonen, vergrößern (Sättigungseffekt). Dies bedeutet, daß der lokale statische Widerstand im Bereich der Driftzonenmitte mitunter groß und damit der maximale statische Nennstrom begrenzt sein wird. Zweitens steigt mit der Länge der Driftzone, die Anfälligkeit für Ausfälle aufgrund der Einwirkung kosmischer Strahlung. Neben diesen Nachteilen, fällt noch die Tatsache ins Gewicht, daß die mit dem 10kV -IGCT erzielte Zwischenkreisspannung von etwa 6kV für viele Anwendungen immer noch zu niedrig ist. Das schlagkräftigste Argument gegen diesen HV-IGCT dürfte aber ein Wirtschaftliches sein: Die teure Entwicklung dürfte sich kaum lohnen in einem engen Markt, wie dem der Mittelspannungsanwendungen.

¹Stand Anfang 2000

Somit bleibt als Standardlösung, die Serienschaltung von IGCTs oder GTOs zur Erhöhung der Systemspannung übrig [50], [66]. Damit werden einerseits die oben erwähnten Nachteile eines potentiellen HV-IGCTs vermieden und andererseits wird Redundanz ermöglicht. Redundanz bedeutet, daß mehr IGCTs/GTOs je Phasenmodul in Reihe geschaltet werden, als eigentlich notwendig wären. Fällt nun ein IGCT aus, indem z.B. aufgrund der kosmischen Strahlung das Element seine Blockierfähigkeit verliert und sich wie ein kleiner Widerstand (Kurzschluß) verhält, kann die Spannung immer noch von den verbliebenen funktionsfähigen IGCTs blockiert werden. Für den Umrichter aus [66] spendiert man immerhin 50% Redundanz. Nachteilig an der Serienschaltung sind die höhere Bauelementanzahl und der damit verbundene erhöhte Platz- und Schaltungsaufwand. Weiterhin können die IGCTs ihr Potential nicht ganz ausschöpfen, da ein unentlasteter (snubberless) Betrieb nicht mehr möglich ist (siehe dazu Abschnitt 5.2). Die notwendigen RC-Entlastungsglieder treiben den Platz- und den Kostenaufwand zusätzlich in die Höhe.

5.2 Asymmetrisches Abschaltverhalten

5.2.1 Ursachen und Mechanismus

Zur Untersuchung der Spannungsasymmetrie bei unentlastetem Abschalten, bedient man sich im einfachsten Fall eines Tiefsetzstellers mit zwei seriengeschalteten IGCTs, wie ihn Bild 5.1 zeigt.

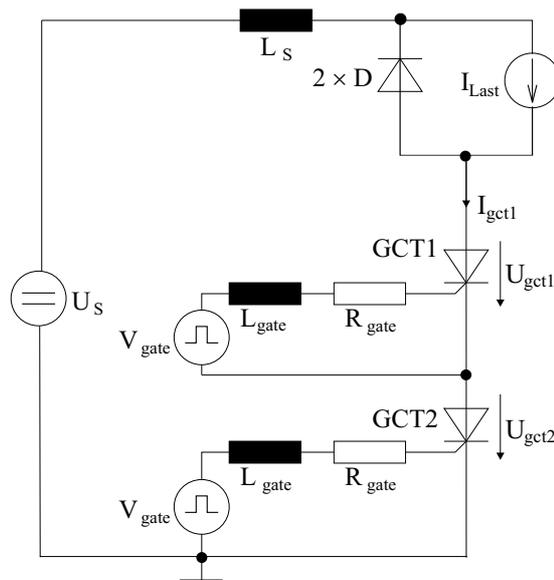


Bild 5.1: Simulationsschema: Tiefsetzsteller mit zwei IGCTs in Serie.

Zunächst sei der triviale Fall betrachtet, daß beide IGCTs und deren Ansteuerung vollkommen identisch sind. Ferner sei angenommen, daß die beiden Gatesignale ebenfalls identisch sind. Das Simulationsergebnis, das mit dem in Kapitel 3 vorgestellten Modell erzielt wird, zeigt Bild 5.2.

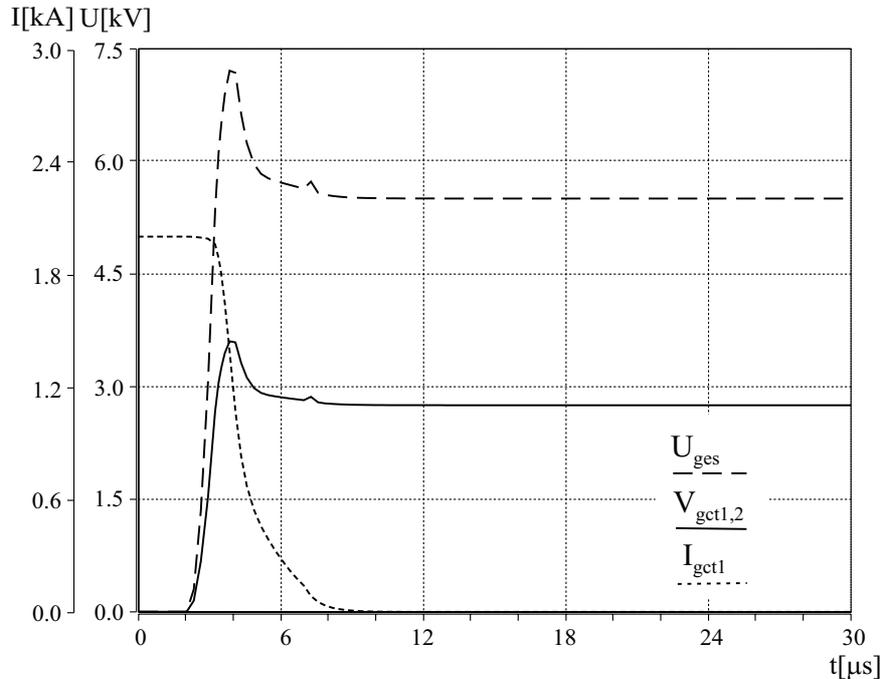


Bild 5.2: *SABER-Simulation: hartes unentlastetes Abschalten von 2kA gegen 5,5kV zweier identischer IGCTs in Serie*

Man erkennt, daß die beiden IGCT-Spannungen identisch sind und ihre Summe selbstverständlich die Gesamtspannung ergibt. Leider sind identische Verhältnisse unrealistisch und es stellt sich die Frage was passiert, wenn Differenzen zwischen einzelnen IGCTs und/oder den zugehörigen Gatekreisen auftreten. Betrachtet seien im folgenden zwei identische IGCTs und gleiche Gatesignale, aber die Gateinduktivität des IGCT1 sei um ca. 8% geringer. Dazu wurde das unentlastete Abschaltverhalten von $I = 3kA$ gegen $U = 2,8kV$ mit dem Bauelementesimulator DESSIS und einer zweidimensionalen IGCT-Struktur simuliert. Das Ergebnis zeigt Bild 5.3: Zunächst sind die von den beiden IGCTs aufgenommenen Spannungen noch ziemlich gleich. Ab $t \approx 8,3\mu s$ laufen die beiden Spannungen massiv auseinander, wobei der IGCT1 wegen der geringeren Gateinduktivität erwartungsgemäß mehr Spannung aufnimmt. Am Ende liegt nahezu die komplette Blockierspannung am IGCT1 an, was bei hoher Zwischenkreisspannung zur Zerstörung einzelner IGCTs führen kann².

²Deshalb wurde bewußt nur $2,8kV$ für die Gesamtspannung gewählt, um das Auftreten z.B. des Lawineneffekts zu vermeiden, was zu einer erheblichen Verlangsamung der Simulationsläufe und zu Konvergenzproblemen bei DESSIS führen kann.

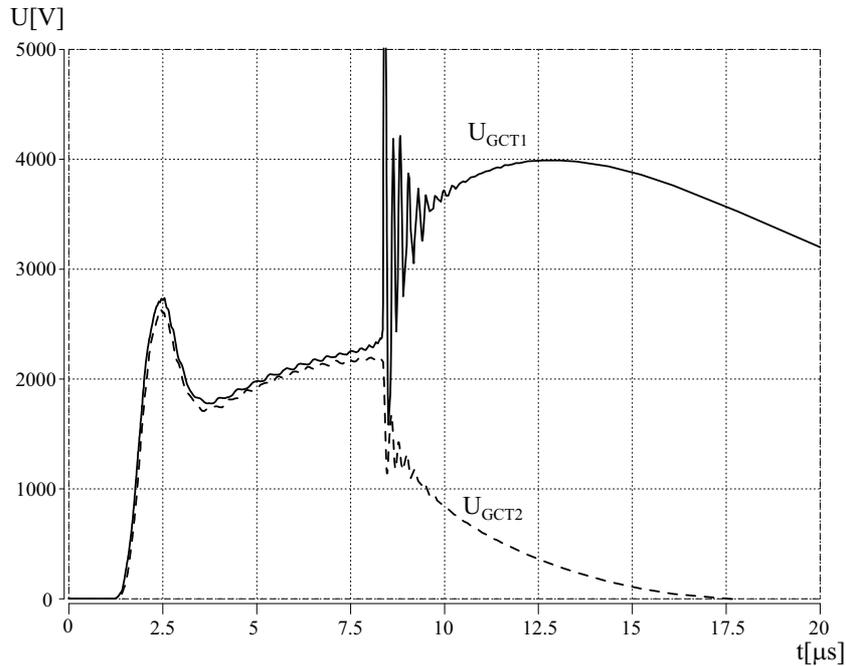


Bild 5.3: *DESSIS-Simulation: hartes unentlastetes Abschalten von 3kA gegen 2,8kV zweier IGCTs in Serie; $L_{gate1,2}$ unterschiedlich*

Im folgenden wird der Mechanismus des Auseinanderlaufens der Spannungen näher beleuchtet, wozu auch Bild 5.4 dient. Zunächst steigt der Gatestrom des IGCT1 wegen der kleineren Gateinduktivität schneller negativ an. Dadurch entsteht die Raumladungszone etwas eher und dehnt sich etwas schneller in Richtung Pufferschicht aus, als dies beim IGCT2 der Fall ist. Die jeweilige Blockierspannung ist zwingend durch die Ausdehnung der jeweiligen Raumladungszone, d.h. Driftzonenweite minus Restplasmazone, und dem Feldgradienten gemäß (2.85) bestimmt. Die Unterschiede betreffs der Ausdehnung der RLZ zwischen beiden IGCTs sind allerdings relativ gering, so daß der Unterschied in der aufgenommenen Spannung anfangs, d.h. für $t < 8,3\mu\text{s}$ im Bild 5.3, ebenfalls noch gering ist (s.a. Bild 5.4 oben). Bei etwa $t = 8,4\mu\text{s}$ ist die Driftzone des IGCT1 völlig von Überschußladungen ausgeräumt worden, während beim IGCT2 noch Plasma vorhanden ist. Damit ist der E-Feldverlauf vom IGCT2 immer noch dreiecksförmig durch die Raumladungszone und den Feldgradienten bestimmt. Die RLZ des IGCT1 entspricht hingegen der gesamten Driftzone und der E-Feldverlauf ist entsprechend trapezförmig wie in Abschnitt 3.3.4 beschrieben (s.a. Bild 5.4 unten). Deshalb laufen die Spannungen zu diesem Zeitpunkt so heftig auseinander. Die Restladung des IGCT2 wird nicht mehr ausgeräumt und mit dem Abklingen der Überspannung insgesamt, d.h. von $U_{GCT1} + U_{GCT2}$, weicht das E-Felddreieck des IGCT1 wieder Richtung p -Basis zurück und die Spannung am IGCT2 sinkt und zwar im Extremfall bis zu Null.

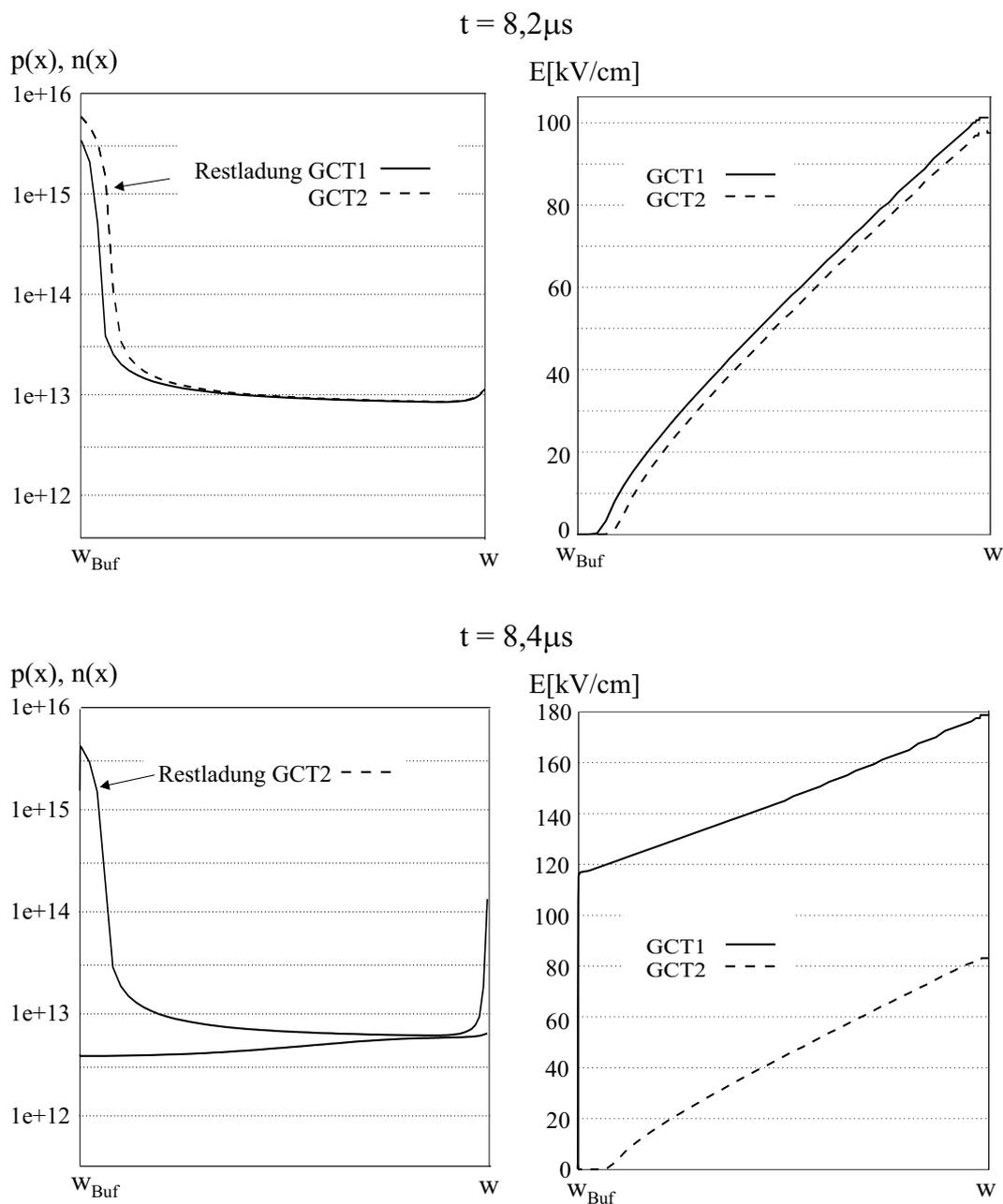


Bild 5.4: *DESSIS-Simulation: Überschlußladungsträger (links) und aufgenommenes elektrisches Feld (rechts) zum Zeitpunkt $t = 8.2\mu\text{s}$ (oben) und $t = 8.4\mu\text{s}$ (unten)*

Die Feldverläufe der beiden IGCTs für $t = 12\mu\text{s}$ zeigt Bild 5.5. Die Ausdehnung des Felddreiecks für den IGCT2 ist bereits recht klein. Die Restladung im IGCT2 klingt dabei lediglich durch Rekombination ab. Am Ende bleibt der dynamisch verursachte Spannungsunterschied auch im statischen Auszustand vorhanden, soweit keiner der

IGCTs durch die Spannungsbeanspruchung zerstört wird.

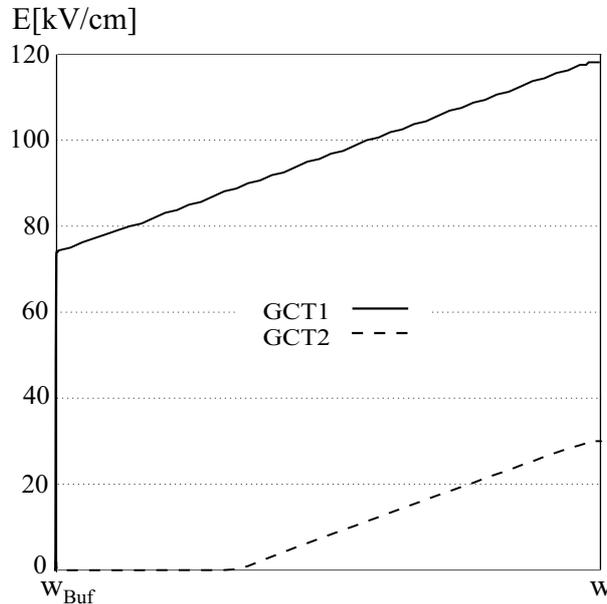


Bild 5.5: Aufgenommenes elektrisches Feld zum Zeitpunkt $t = 12\mu s$

Die Gründe, die zum Auftreten der Asymmetrie führen, lassen sich wie folgt zusammenfassen:

- (1) unvermeidliche Unterschiede in den genauen physikalischen und geometrischen Daten der einzelnen diskreten IGCTs, besonders der Lebensdauern.
- (2) Unterschiede in den jeweiligen Gatekreisen, d.h. unterschiedliche Gateinduktivitäten und/oder Gatewiderstände.
- (3) Zeitversetzte Abschaltimpulse, d.h. zwischen einzelnen Abschaltimpulsen existiert eine Zeitverzögerung Δt .
- (4) Beliebige Kombinationen von (1), (2) und (3).

In obiger DESSIS-Simulation wurde der Effekt von (2) behandelt. Im folgenden wird gezeigt, daß mit dem in Kapitel 3 vorgestellten und in Kapitel 4 validierten Modell, die asymmetrische Spannungsaufnahme simuliert werden kann. Daß bei identischen Verhältnissen keinerlei Asymmetrie auftritt, zeigt bereits Bild 5.2. Jetzt werde der IGCT2 $20ns$ später mit dem Abschaltimpuls beaufschlagt, wohingegen die restlichen Verhältnisse identisch bleiben (Grund (3)). Das entsprechende Simulationsergebnis zeigt Bild 5.6. Wieder ergibt sich ein ähnliches Bild wie bei der Dessimulation: Zunächst, d.h. für $t < 7\mu s$, ist der Unterschied zwischen U_{gct1} und U_{gct2} noch relativ gering. Für $t > 7\mu s$ laufen die Spannungen dann wieder stark auseinander, wobei am IGCT2 im Aus-Zustand immerhin $1kV$ Restspannung übrigbleiben. Interessant

ist in dem Zusammenhang, daß der abzuschaltende Strom I_{gct1} schon sehr klein ist, worauf am Ende dieses Kapitels noch zurückzukommen sein wird.

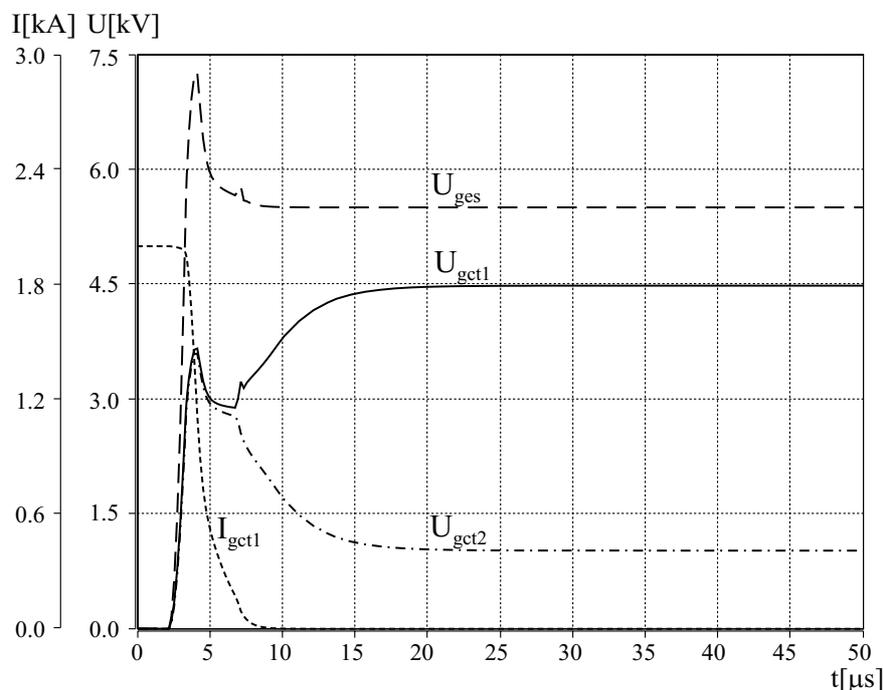


Bild 5.6: *SABER-Simulation: hartes unentlastetes Abschalten von 2kA gegen 5,5kV; $\Delta t = 20ns$, ansonsten sind die Verhältnisse identisch (Grund (3)).*

Auf jeden Fall bleibt festzuhalten, daß selbst minimale Abweichungen zwischen den Daten der einzelnen Serien-IGCTs und ihrer Gatekreise, zu großen Spannungsabweichungen führen können. Ein Beispiel für den Grund (1) zeigt Bild 5.7, wobei nun drei IGCTs in Serie betrachtet werden. Beim IGCT1 ist die Hochinjektionslebensdauer nur 1% geringer und beim IGCT3 nur 1% höher als beim IGCT2. Ansonsten sind die Verhältnisse wieder identisch. Trotzdem laufen die Spannungen auseinander und zwar so, daß eine statische Differenz von immerhin etwa 500V zwischen U_{gct1} bzw. U_{gct3} und der „Referenzspannung“ U_{gct2} verbleibt. Der IGCT mit der kleinsten Lebensdauer und der etwas geringeren Speicherladung nimmt am meisten Blockierspannung auf. Nun sei eine „Worst Case Situation“ betrachtet, die so in der Realität nicht oder selten vorkommen wird. IGCT1 und IGCT2 hat eine 10% geringere Hochinjektionslebensdauer und IGCT3 eine 10% höhere Lebensdauer als beim Parametersatz für den 5SHY35L4503 (s. Kapitel 4). Die Gateimpedanz ist bei IGCT1 und IGCT2 um 5% kleiner und beim IGCT3 um 5% größer als bei den Simulationen in Kapitel 4. Zusätzlich wird der IGCT3 30ns später mit dem Abschaltimpuls beaufschlagt. Das Simulationsergebnis zeigt Bild 5.8.

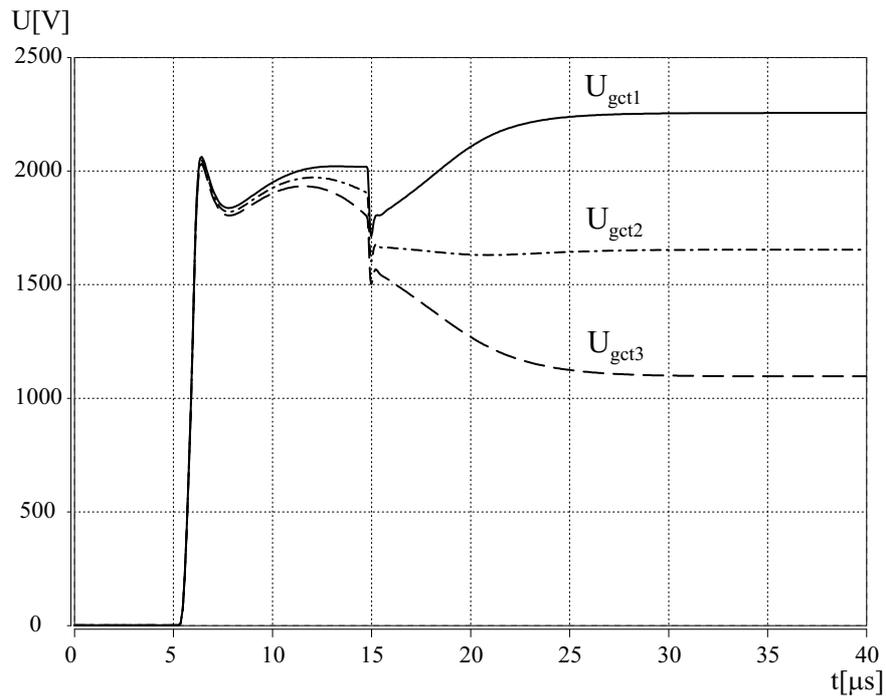


Bild 5.7: SABER-Simulation: hartes unentlastetes Abschalten von 3kA gegen 5kV; $\Delta t = 20\text{ns}$, ansonsten Verhältnissen identisch.

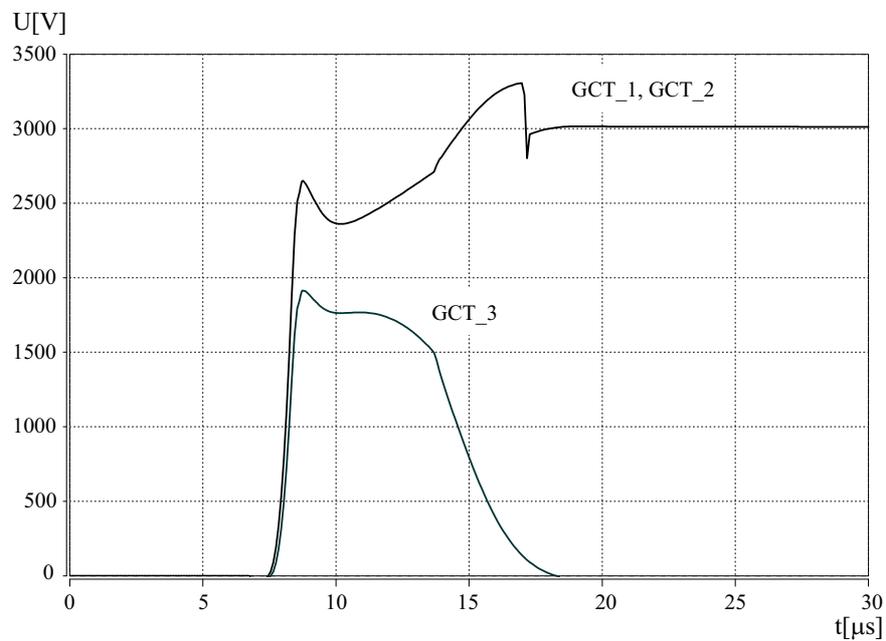


Bild 5.8: SABER-Simulation: hartes unentlastetes Abschalten von 3kA gegen 6kV; „Worst Case Szenario“

Alle Parameterunterschiede sind so, daß sie sich nicht gegenseitig kompensieren, d.h. alle drei Parameterdifferenzen ($\tau_H, L_{gate}, \Delta t$) sind so gewählt, daß der IGCT3 langsamer und weniger Spannung aufnimmt, als IGCT1 und IGCT2. Deswegen ergibt sich die große Spannungsdifferenz von über $500V$ bereits nach dem Spannungsmaximum. Für $t > 12,5\mu s$ laufen die Spannungen wieder sehr stark auseinander und U_{gct3} wird sehr schnell Null. Daß die Spannungen U_{gct1} und U_{gct2} übereinander liegen ist kein Zufall, da die Parameter der beiden IGCTs und den zugehörigen Gatekreisen genau gleich sind.

Das in dieser Arbeit vorgestellte Modell kann also sehr gut zur Analyse von serienschalteten IGCT-Elementen verwendet werden. Im Gegensatz dazu ist das Standard-Modell aus der SABER-Bibliothek gemäß [41] nicht geeignet die beim Abschalten auftretende Spannungsasymmetrie, zu analysieren. Um dies zu zeigen, wurde eine Serienschaltung, bestehend aus zwei IGCTs, einmal mit dem physikalischen und einmal mit dem Modell der SABER-Bibliothek simuliert. Dabei wird der Abschaltimpuls am IGCT2 um $40ns$ später appliziert. Sonst sind die Verhältnisse identisch. Die berechneten Transienten zeigt Bild 5.9.

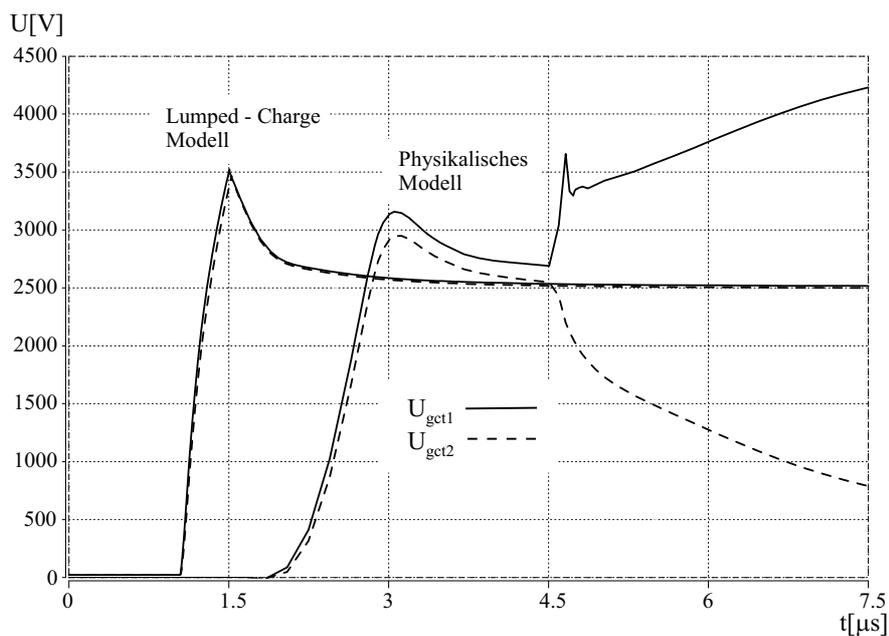


Bild 5.9: Vergleich physikalisches Modell versus „Lumped Charge-Modell“: hartes unentlastetes Abschalten von $2kA$ gegen $5kV$; $\Delta t = 40ns$, ansonsten Verhältnissen identisch.

Während das physikalische Modell aufgrund der physikalisch exakten Beschreibung der Spannungsaufnahme, den Asymmetrieeffekt richtig berechnet, findet dies mit dem Modell gemäß [41], nicht, bzw. nur marginal, statt. Man könnte nun gegen diese Argumentation folgendes einwenden: Die Beschreibung des Mechanismus am

Anfang dieses Abschnitts, mit der Feststellung, daß erst wenn das E-Feld den Puffer erreicht, die Spannung massiv auseinanderläuft, legt nahe, daß die Asymmetrie nur bei Punch-Through Bauelementen mit einer entsprechenden Pufferschicht auftritt. Da aber in dem GTO-Modell nach [41] gar keine Pufferschichtbeschreibung vorhanden ist, kann der Effekt zwangsläufig nicht berechnet werden.

Um zu untersuchen, ob der Puffer alleine für das stark asymmetrische Spannungsverhalten verantwortlich ist, wurde erneut eine Simulation mit dem physikalischen IGCT-Modell durchgeführt. Dabei wurde die Zwischenkreisspannung so klein gewählt, daß bei beiden IGCTs das E-Feld zu keinem Zeitpunkt den Puffer erreicht und damit immer ein dreiecksförmiger Feldverlauf vorliegt. Bild 5.10 zeigt die Transienten dieser Simulation mit einem $\Delta t = 30ns$ und sonst identischen Verhältnissen.

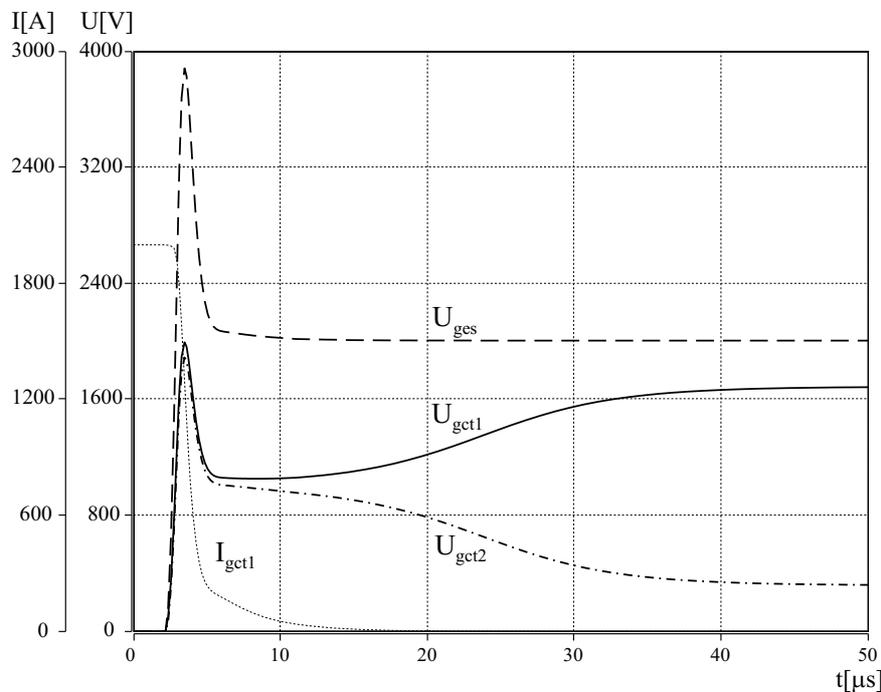


Bild 5.10: *Simulation zweier IGCTs in Serie: hartes unentlastetes Abschalten von 2kA gegen 2kV; $\Delta t = 30ns$*

Die besagte Asymmetrie tritt also auch ohne Pufferschicht auf, nur laufen die Spannungen deutlich langsamer auseinander. Vergleicht man Bild 5.6 mit Bild 5.10, stellt man fest, daß die statischen Endwerte jetzt erst ca. $20\mu s$ später erreicht werden. In beiden Fällen ergibt sich aber für die statischen Endwerte ein Verhältnis $U_{gct1} : U_{gct2}$ von etwa 4,5 : 1. Daraus kann nun eindeutig geschlossen werden, daß der Effekt der asymmetrischen Spannungsverteilung beim Abschalten seriengeschalteter IGCTs, GTOs oder auch IGBTs, nur mit einem physikalischen Modell, das eine möglichst exakte Beschreibung des Ladungsträgerabbaus und der damit zusammenhängenden

Ausbreitung der Raumladungszone (E-Feld) enthält, berechnet werden kann. Das vorliegende Modell erfüllt diese Kriterien und kann zu weiteren Analysen eingesetzt werden. So kann man potentielle Lösungsansätze zur Vermeidung dieser Asymmetrie simulativ untersuchen.

5.2.2 Vermeidung der Spannungsasymmetrie

Zunächst könnte man nur Widerstände parallel zu den IGCTs schalten. Dadurch wird eine entstehende Asymmetrie wieder abgebaut und zwar umso schneller, je kleiner der Widerstandswert ist. Ideal wären dazu Widerstände im Bereich von 100Ω oder kleiner. Diese Widerstände müssen andererseits groß genug sein ($R \gtrsim 50k\Omega$), damit sie gemäß $P_V = U^2/R$ nicht selbst zu viel Energie verbrauchen und den Wirkungsgrad eines Umrichters inakzeptabel verschlechtern. Aufgrund dieser Anforderung sind Parallelwiderstände allein zu groß, um das Entstehen der Asymmetrie und eine eventuelle Überbelastung eines oder mehrerer IGCTs zu verhindern.

Kleine Widerstände im Bereich von 10Ω können verwendet werden, indem man sie mit einem Kondensator in Reihe schaltet. Dieser verhält sich zu Beginn des Schaltvorgangs wie ein Kurzschluß, so daß die Widerstände wie kleine Parallelwiderstände wirken. Während des Ladevorgangs wächst der Gesamtwiderstand der RC-Kette an und ist am Ende unendlich groß. Wirkleistung wird an den Widerständen nur während des Schaltvorgangs verbraucht. Energetisch noch günstiger ist es, wenn man die RC-Kette durch eine Diode zu einem vollständigen RCD-Snubber erweitert. Aus Kostengründen beschränkt man sich in der Praxis aber gerade bei Hochleistungsumrichtern auf reine RC-Glieder (s z.B. [66]).

GTOs können ohnehin nur entlastet abgeschaltet werden, wofür man ebenfalls entsprechende RC-Glieder verwendet (vgl. z.B. [62]). Mit dieser Beschaltung erreicht man somit zwei Ziele, nämlich einerseits Entlastung und andererseits Symmetrierung der Spannungen. Wie noch zu zeigen ist, können die Kapazitäten zum Zwecke der Symmetrierung kleiner dimensioniert werden, als zum Entlasten. Damit bestimmt die notwendige Entlastung die Kapazitätsauswahl bei GTO-Umrichtern. Für $3kA/4, 5kV$ -Elemente muß man meist Kapazitäten von mindestens $C = 6\mu$ verwenden (vgl [1]).

Dagegen kann man sich bei den IGCTs auf die Symmetrierung der Spannungen beschränken. Die Analyse mittels SABER zeigt, daß die Kapazität umso höher sein muß, je größer die Abweichungen zwischen den IGCTs sind. Im Beispiel zu Bild 5.6, bei dem der einzige Unterschied ein kleineres Δt zwischen den Gatesignalen war, genügte bereits eine Kapazität von $C = 100nF$, um eine befriedigende Symmetrierung sicherzustellen, wie Bild 5.11 zeigt. Für das zweite Beispiel, das oben als „Worst Case“ bezeichnet wurde, benötigt man immerhin schon einen Snubberkondensator

von $C = 3\mu F$. Das zugehörige Simulationsschema zeigt Bild 5.12.

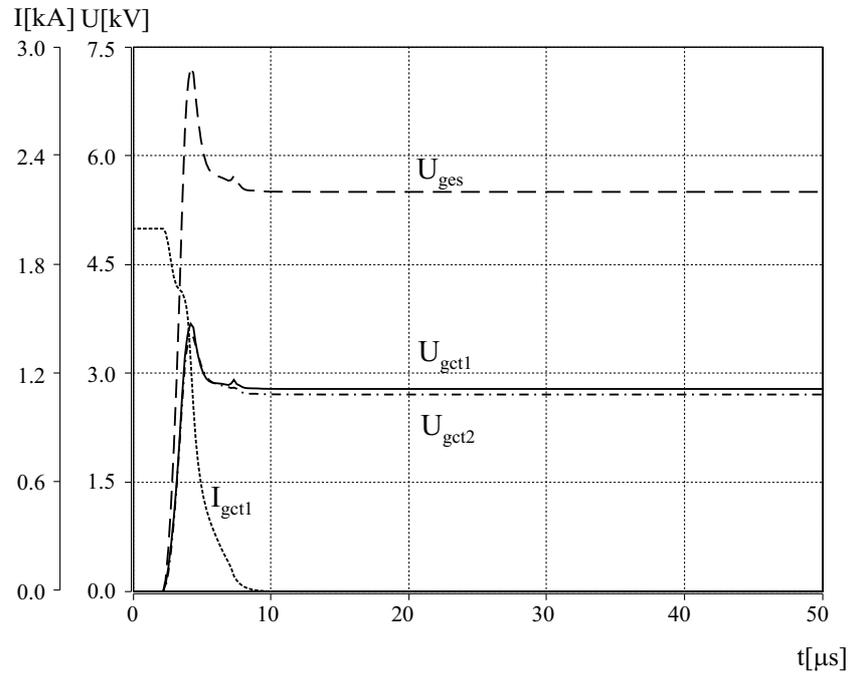


Bild 5.11: SABER-Simulation: hartes unentlastetes Abschalten von 2 kA gegen 5,5 kV; $\Delta t = 20 ns.$; RC-Netzwerk mit $C = 100 nF$

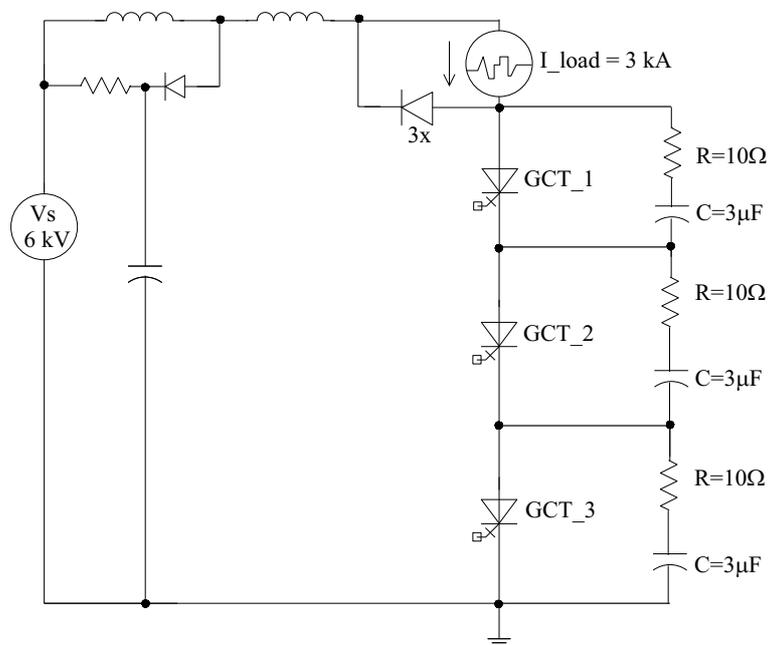


Bild 5.12: Schema mit 3 IGCTs in Serie; jeweils ein paralleles RC-Glied

Die Spannungstransienten der dazu passenden Simulation, sind im Bild 5.13 zu sehen. Dabei sind die Spannungen, die man mit dem RC-Snubber erhält, mit den durchgezogenen Linien gekennzeichnet, wohingegen die gestrichelten Linien nochmal die Kurven ohne Snubber zeigen.

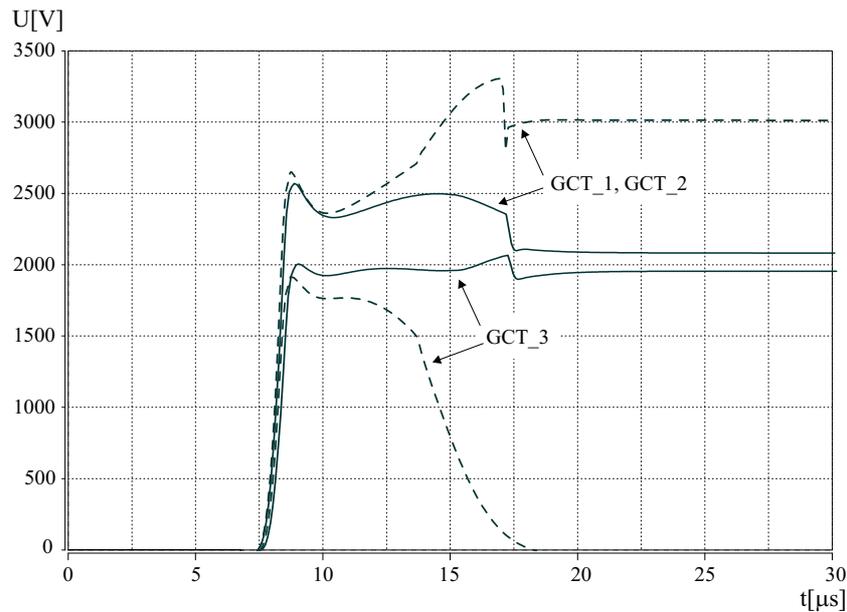


Bild 5.13: *SABER-Simulation: hartes unentlastetes Abschalten von 3kA gegen 6kV; „Worst Case Szenario“, RC-Netzwerk mit $C = 3\mu F$*

Beim GTO (!) benötigt man wegen $dU/dt = I_A/C$ eine große Snubberkapazität C um den Spannungsanstieg dU/dt in erforderlichem Maße, zum Zwecke der Entlastung, zu begrenzen. Wie die Bilder 5.6 und 5.10 zeigen, ist der Anodenstrom bereits sehr klein wenn die Spannung stark auseinanderläuft. Deshalb können die Snubberkapazitäten zum Zwecke der Symmetrierung bei IGCTs erheblich kleiner sein, als wie dies zum Entlasten von GTOs der Fall ist. Je größer die Abweichungen zwischen den einzelnen IGCTs ist, desto früher beginnt die (merkliche) Asymmetrie und desto höher ist der Reststrom. Deshalb reichen im Beispiel zu Bild 5.11 $C = 100nF$, wohingegen für das Beispiel in Bild 5.13 bereits $C = 3\mu F$ vonnöten sind.

Die beiden Bilder 5.11 und 5.13 zeigen gewissermaßen Extremfälle. Die Annahme, daß nur eine Verzögerungszeit in Höhe von $\Delta t = 20ns$ als einzige Abweichung auftritt, ist sicherlich eine unrealistische Annahme. Allerdings kann auch der „Worst Case“ durch geschickte Wahl der Bauelemente verhindert werden. Ein Kriterium dafür könnte beispielsweise eine (fast) gleiche Kennlinie verschiedener IGCTs sein, denn ist die Kennlinie annähernd gleich, ist mit hoher Wahrscheinlichkeit auch die

Verteilung der überschußladungen in der ν -Zone ähnlich. In der Praxis sind also Fälle, die zwischen den beiden Beschriebenen liegen, realistisch. Der Autor nimmt an, daß man mit Kapazitäten im Bereich von $C \approx 1\mu F$ sehr gut zurechtkommt, wenn man die einzelnen IGCTs entsprechend auswählt.

Nun benötigen Kapazitäten dieser Größe für die betrachtete Spannungsklasse immer noch relativ viel Platz und sind teuer. Deshalb ist es sinnvoll, sich Gedanken über mögliche Alternativen zu machen. Für Serienschaltungen von IGBTs wurden zwischenzeitlich aktive (intelligente) Ansteuerschaltungen entwickelt, die sozusagen ein gemeinsames Abschalten der Transistoren koordinieren und damit Spannungssymmetrie sicherstellen [24], [53]. Über ein Konzept für eine Schaltung gleicher Zielsetzung, jetzt aber für seriengeschaltete IGCTs, wird in Kapitel 6 berichtet.

6 Konzept der aktiven Ansteuerung von IGCTs

6.1 Das Regelungskonzept

6.1.1 Allgemeines

Am Ende des Kapitels 5 wurde bereits angedeutet, daß und vor allem warum RC-Glieder zur Spannungssymmetrierung nicht unbedingt optimal sind. Prinzipiell bietet der IGCT die Möglichkeit des unentlasteten Ausschaltens. Somit stellt sich die Frage, ob es nicht möglich ist, lediglich durch Nachregeln mit einer aktiven Gateansteuerung, eine homogene Aufteilung der Blockierspannung während des Schaltvorgangs zu erzielen¹. Bei Reihenschaltungen von IGBTs ist dieses Verfahren längst Stand der Technik [24], [53]. Allerdings ist der IGBT mit seinem MOS-Steuerkopf ein spannungsgesteuertes Bauelement, d.h. es fließen vergleichsweise kleine Ströme zum Umladen der Gatekapazitäten. Hingegen erreicht der Gatestrom beim stromgesteuerten IGCT, Beträge in Höhe des abzuschalteten Stroms (vgl. z.B. Bild 4.3). Somit sind Anforderungen, die an eine Schaltung für den genannten Zweck zu stellen sind, beträchtlich höher, als beim IGBT (s. Abschnitt 6.1.4). Als Istwert einer Regelung dient naturgemäß die gemessene Anoden-Kathoden-Spannung U_{AK} , als Sollwert der symmetrische Spannungswert und als Stellgröße der Gatestrom I_G .

6.1.2 Von der Idee zum P-Regler

Wie in Kapitel 3 erläutert, ist das Hartansteuern eine Voraussetzung für den unentlasteten Betrieb. Dabei ist der Kathodenstrom I_K bereits Null, d.h. der Gate-Kathoden-Übergang sperrt vollständig, bevor der IGCT merkbar Spannung aufnimmt. Da der Anodenstrom sich noch nicht verändert hat, entspricht der Ga-

¹Eigentlich müßte es im folgenden GCT statt IGCT heißen, da sich das I in der Abkürzung auf die vom Hersteller mitgelieferte Gateansteuerung bezieht, die hier durch die aktive Ansteuerung ersetzt wird. Aus „Gewohnheit“ wird aber weiterhin zumeist IGCT verwendet.

testrom dem Anodenstrom und ist maximal. Den sinnvollen Aktionsbereich eines Reglers zeigt die grau hinterlegte Fläche in Bild 6.1. Die linke Begrenzung des Aktionsbereichs wird vom Zeitpunkt des Beginns der Spannungsaufnahme festgelegt, denn vorher kann schlichtweg kein Istwert detektiert werden. Die rechte Grenze des Gesamtaktionsbereichs entspricht dem Zeitpunkt, zu dem die Stellgröße, d.h. I_G null wird. Zusätzlich ist noch ein kurzer Zeitraum von etwa $1\mu s$ Dauer zusätzlich schraffiert gekennzeichnet und wird hier Primärer Aktionsbereich genannt.

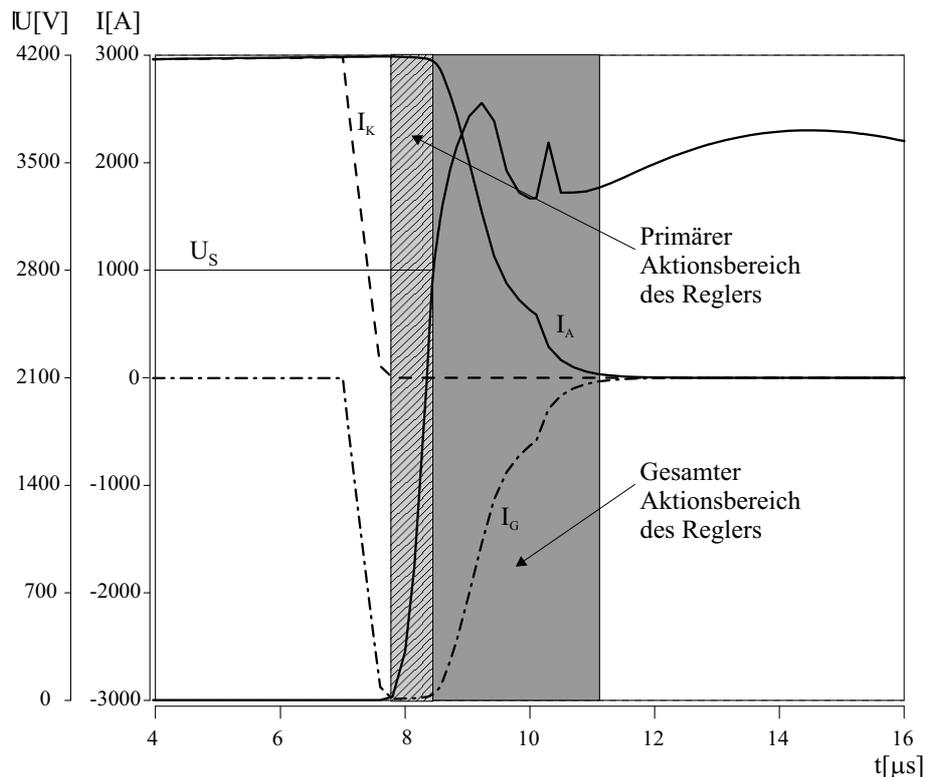


Bild 6.1: Sinnvoller Reglerbereich im Bezug auf die Transienten eines hart abgeschalteten IGCT.

Er erstreckt vom Zeitpunkt des Beginns der Spannungsaufnahme bis zu dem Zeitpunkt, an dem die Spannung zum ersten Mal den statischen Endwert(anteil) U_S erreicht. In diesem Zeitraum muß die Regleraktion beginnen. Danach beginnt die Überspannung, die zum Abkommutieren des Anodenstroms führt. Damit sinkt aber auch der Gatestrom zwangsläufig in gleicher Weise und der Abschaltvorgang ist quasi „beendet“, auch wenn die Ströme und die Spannung ihre statischen Endwerte noch nicht erreicht haben. Damit ist die Einflußnahme am Gate nur noch bedingt sinnvoll, z.B. im Sinne einer Nachjustierung. Wie in Abschnitt 6.3 gezeigt wird, findet die hauptsächliche Regleraktion tatsächlich im primären Aktionsbereich statt, wogegen die Feinanpassung im abklingenden Ast des Gatestroms geschieht.

Eine Regelung, die auf eine Sollwertabweichung der aufgenommenen Spannung U_{AK} im primären Aktionsbereich reagiert, kann den Betrag des Gatestroms I_G nur zwingen kleiner zu werden (vgl. Bild 6.1). Diese Erkenntnis hat zwei Konsequenzen: Erstens bedeutet eine Verkleinerung von I_G wegen $I_A = const$ zwangsläufig, daß I_K zumindest kurzzeitig wieder größer null wird, was man beim harten Schalten zur Vermeidung des Filamentierungseffekts eigentlich verhindern will. Allerdings sollte sich dies nicht allzusehr negativ bemerkbar machen, da I_k nur kurz größer null wird. Somit injiziert die Kathode nicht viele Elektronen in die p -Basis. Da diese auch nicht weit in die p -Basis eindringen können, entstehen vermutlich keine Ladungsfilamente. Auf diese Problematik wird detaillierter in Abschnitt 6.4.1 eingegangen, denn vorher soll untersucht werden, in welcher Größenordnung I_K überhaupt liegt. Zweitens kann ein IGCT wegen $I_{G_geregelt} < I_{G_ungeregelt}$ vom Regler nur gebremst, d.h. die Spannungsaufnahme verlangsamt werden. Eine geregelte Beschleunigung des Abschaltverhaltens ist wegen der Schranke $|I_G| \leq I_A$ nicht möglich.

Somit ist es sinnvoll genau die IGCTs zu bremsen, deren aufgenommene Spannung über dem Sollwert liegen. An allen anderen IGCTs ist der jeweilige Regler passiv. Dadurch, daß an den gebremsten IGCTs der Spannungsanstieg kleiner wird, vergrößert er sich bei den ungebremsten IGCTs automatisch. Es leuchtet unmittelbar ein, daß es vorteilhaft ist, wenn der Regler bereits bei kleinen Spannungsabweichungen von wenigen Volt anspricht. Betrachtet werde nun ein beliebiger IGCT i (im weiteren GCT $_i$ genannt) als Teil einer Serienschaltung, die aus n IGCTs besteht. Die Sollwertspannung beträgt dann zu jedem Zeitpunkt $U_{ges}(t)/n$. Die Sollwertabweichung ergibt sich damit zu $U\Delta_i = U_{GCT_i} - U_{ges}(t)/n$. Somit gilt für zu bremsende IGCTs $U\Delta_i > 0$. Möchte man diesen GCT $_i$ bremsen, kann der negative Gateimpuls $U_{Gi}(t)$, d.h. das ursprüngliche Signal zum Abschalten, mit dem positiven $U\Delta_i$ addiert werden, wobei $U\Delta_i$ noch mit einem Faktor K multipliziert werden kann. Damit hat man einen einfachen Proportionalen Regler (P-Glied) realisiert, wie ihn schematisch Bild 6.2 zeigt [39].

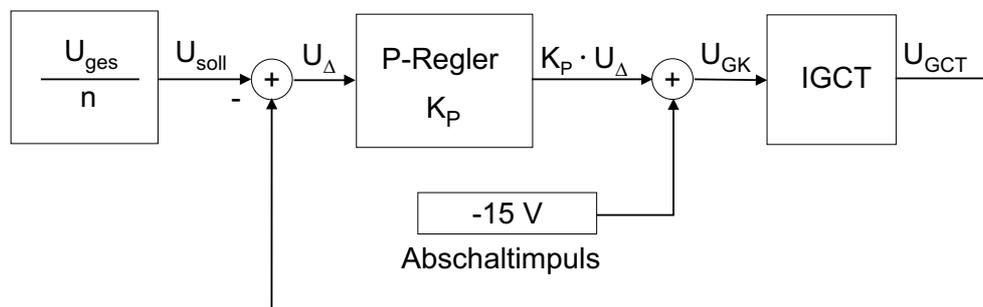


Bild 6.2: Aktive Gatesteuerung mit einem P-Regler: Signalflußplan

Nun könnte man meinen, daß einem so einfachen Reglerprinzip, eine einfache Strecke zugrundeliegt. Dem ist aber leider nicht so. Eine Untersuchung in [8] zeigte, daß sich

für das transiente Verhalten des IGCT keine einfache Strecke modellieren läßt. Wäre dies so, genügten einfache Verhaltensmodelle anstatt der aufwendigen physikalischen Modelle. Ferner kommt wegen der hohen Geschwindigkeitsanforderungen (vgl. Abschnitt 6.1.4) kein PI-, PID-Regler, oder irgend ein anderer Regler in Betracht. Alleine der P-Regler mit dem Verstärkungsfaktor K_P kommt in Frage. Variiert K_P im Zeitablauf nach einer gewünschten Charakteristik ist der P-Regler adaptiv. Es ist sinnvoll zu Beginn der Spannungsaufnahme mit einem großem Verstärkungsfaktor K_P zu arbeiten, wohingegen am Ende, wenn ein Reglereingriff unerwünscht ist, K_P kleiner zu wählen. Ab einem bestimmten Zeitpunkt kann K_P null werden, da eine bestimmte statische Abweichung ohnehin bleibt.

6.1.3 Simulationsergebnisse

Bild 6.3 zeigt das Simulationsschema für die aktive Ansteuerung von zwei IGCTs in Serie, basierend auf einem idealen P-Regler, der mit den gesteuerten Quellen PG1, und PG2 modelliert wird.

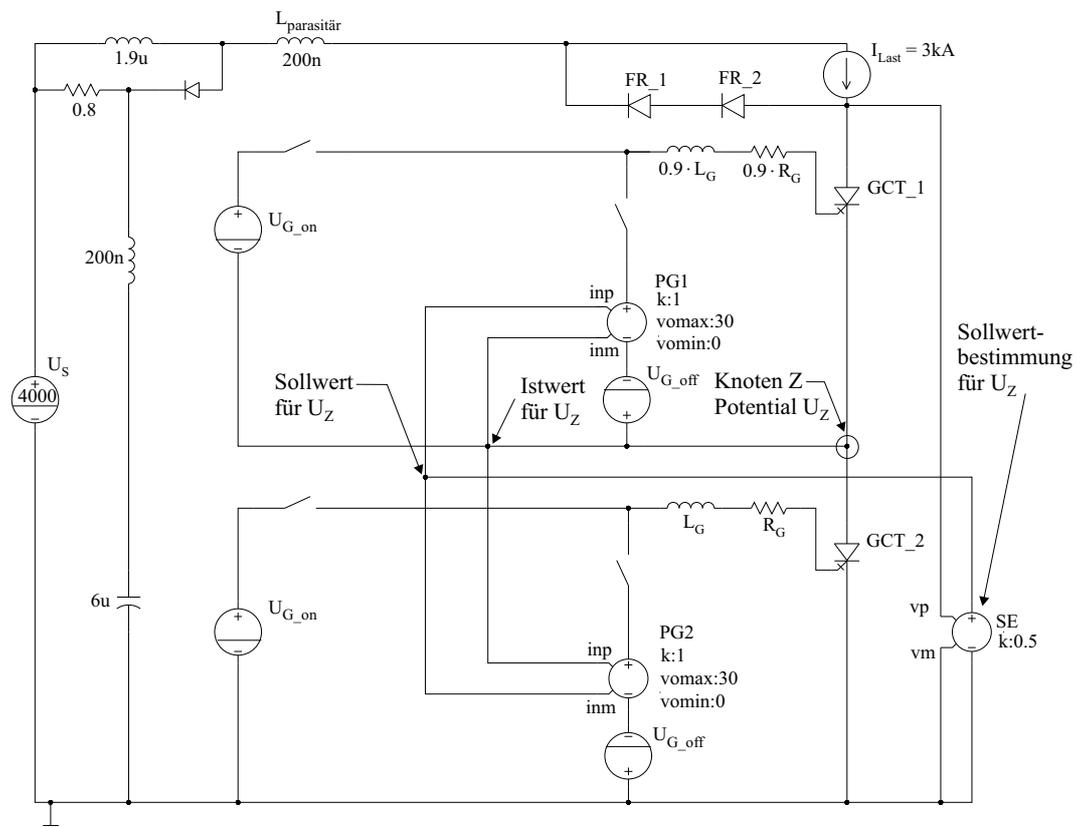


Bild 6.3: Aktive Gateansteuerung mit idealem P-Regler: Simulationsschema; $K_P = 1$

Die spannungsgesteuerte Spannungsquelle SE erfaßt die Summenspannung U_{ges} über beide IGCTs. Am Ausgang von SE erhält man dann U_{ges} multipliziert mit eins geteilt durch die Anzahl der IGCTs (hier $1/2$). Dieser Wert entspricht dem Sollwert des Potential am Knoten Z. Die Differenz aus diesem Sollwert minus dem Istwert U_Z liegt am Eingang der spannungsgesteuerten Spannungsquelle PG1, die dem P-Glied entspricht. Am Eingang von PG2 liegt dann genau die gleiche Spannungsdifferenz $U\Delta$ mit umgekehrtem Vorzeichen an. Am Ausgang von PG1,2 liegt die jeweilige Eingangsspannung $\pm U\Delta$ an, d.h. $K_P = 1$, soweit sie sich innerhalb der definierten Grenzen bewegt. Die untere Grenze beträgt hierbei $0V$ und die obere $30V$, d.h. bei $30V$ wird die Ausgangsspannung abgeschnitten. Die Ausgangsquelle von PG1,2 ist mit der jeweiligen Spannungsquelle U_{Goff} , die den Ausschaltimpuls liefert, in Reihe geschaltet. Somit addieren sich beide Spannungen U_{Goff} und $U\Delta$ zu einer effektiven Gatespannung $U_G = U\Delta + U_{Goff}$, wobei aufgrund der positiven Ausgangsspannung von PG1,2, U_G betragsmäßig verkleinert oder sogar positiv wird. Insgesamt kann sich die effektive Ausgangsspannung U_G aufgrund der Limitierung im Bereich $[-15V, +15V]$ bewegen, was dem Arbeitsbereich vieler Operationsverstärker(schaltungen) entspricht.

Angenommen, der obere GCT_1 schaltet schneller, so daß er etwas mehr Spannung aufnimmt als GCT_2. Dann gilt trivialerweise $U_Z < U_{ges}/2$ und $U\Delta$ am PG1 ist positiv. Folglich wird die effektive Gatespannung U_G erst betragsmäßig vermindert und dann schließlich positiv. Somit wird der GCT_1 gebremst. Am Ausgang von PG2 liegt dagegen wegen $-U\Delta$ am Eingang und dem unteren Limit 0 keine Summenspannung, sondern weiterhin alleine $U_G = U_{Goff}$ am Gate an.

In Bild 6.3 wurden L_G und R_G am GCT_1 um ca. 10% kleiner gewählt als am GCT_2. Weiterhin ist die Hochinjektionslebensdauer beim GCT_1 20% kleiner als beim GCT_2. Somit kann man erwarten, daß der GCT_1 wesentlich schneller schaltet und demnach dazu neigt, mehr Spannung aufzunehmen. Die simulierten Spannungstransienten zeigt Bild 6.4 und die dazugehörigen Gateströme Bild 6.5. Die Spannung an beiden Reglerausgängen plus dem jeweiligen U_{Goff} zeigt Bild 6.6. Zunächst, d.h. $t \geq 3,5\mu s$, ist die Spannung am GCT_1 tatsächlich größer als am GCT_2 (s. Bild 6.4). Folglich wird der Regler PG1 aktiv, was man in Bild 6.5 an dem Einbruch des Gatestrom des GCT_1 sieht. Wegen $dI_G/dt > 0$ muß demnach die Spannung an der parasitären Gateinduktivität des GCT_1 größer als null sein (Bild 6.6). Die Spannungsaufnahme wird am GCT_1 somit verlangsamt und am GCT_2 relativ dazu beschleunigt, bis $U_{GCT_2} > U_{GCT_1}$ ab $t \geq 4\mu s$ gilt. Dann beginnt PG2 in gleicher Art und Weise wie vorhin PG1 zu regeln. Letztlich wechseln sich PG1 und PG2 ab, bis sich das Ganze zur statische Spannungsdifferenz von deutlich unter $100V$ eingeschwungen hat. Bei Betrachtung von Bild 6.4 fällt auf, daß PG1 bzw. PG2 überwiegend im in Bild 6.1 definierten primären Aktionsbereich arbeiten. Beim idealen Regler wird im abklingenden Ast des Gatestroms kaum mehr nachgeregelt, außer beim Erreichen der Pufferschicht bei $t \approx 11,7\mu s$ (siehe Bild 6.4).

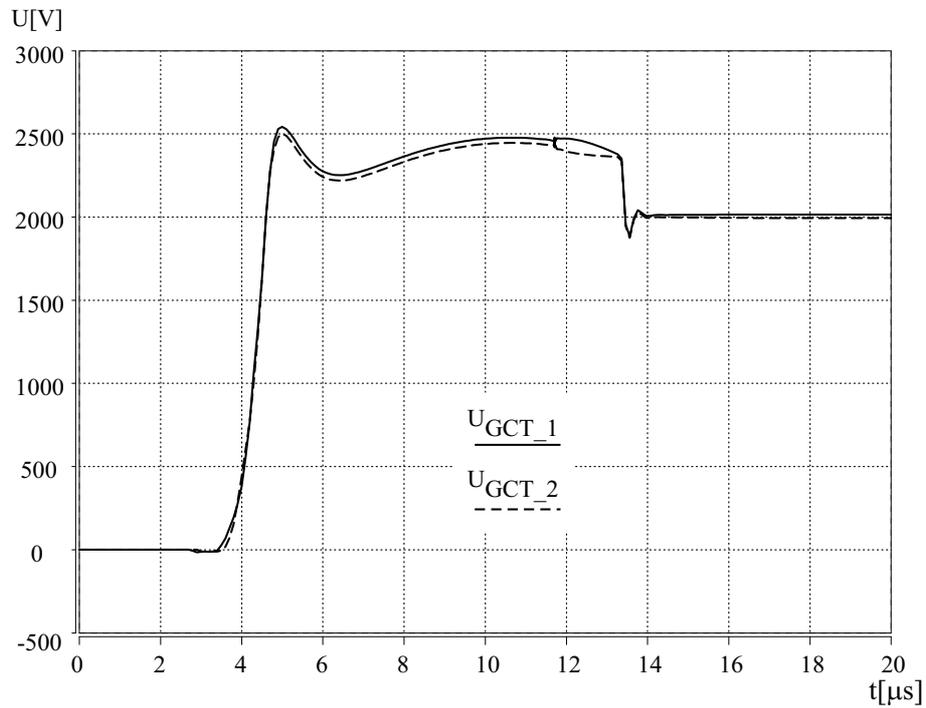


Bild 6.4: Simulation zum Schema in Bild 6.3: Spannungstransienten des geregelten Abschaltens.

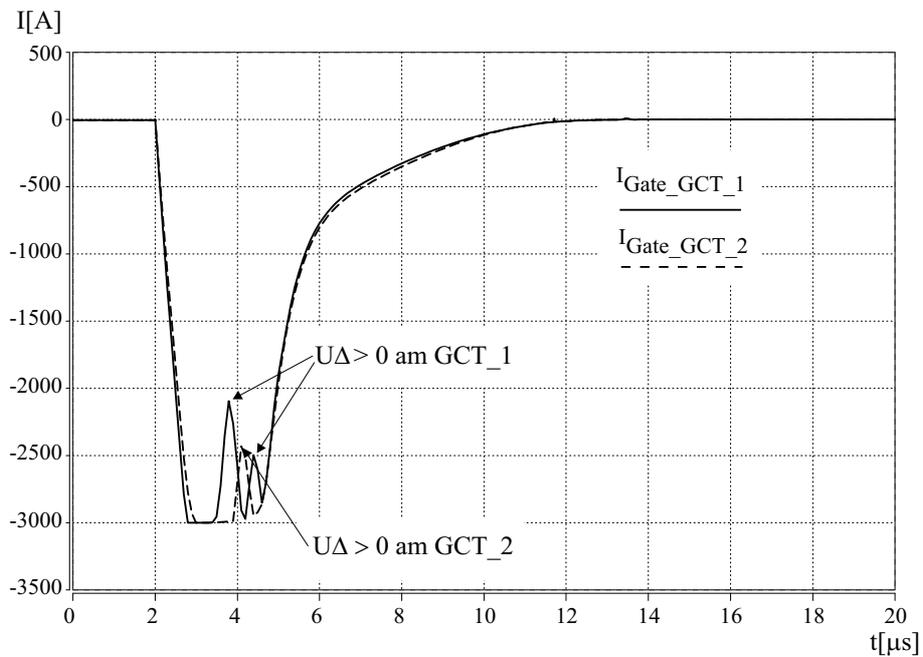


Bild 6.5: Simulation zum Schema in Bild 6.3: Verlauf des Gatestroms des geregelten Abschaltens.

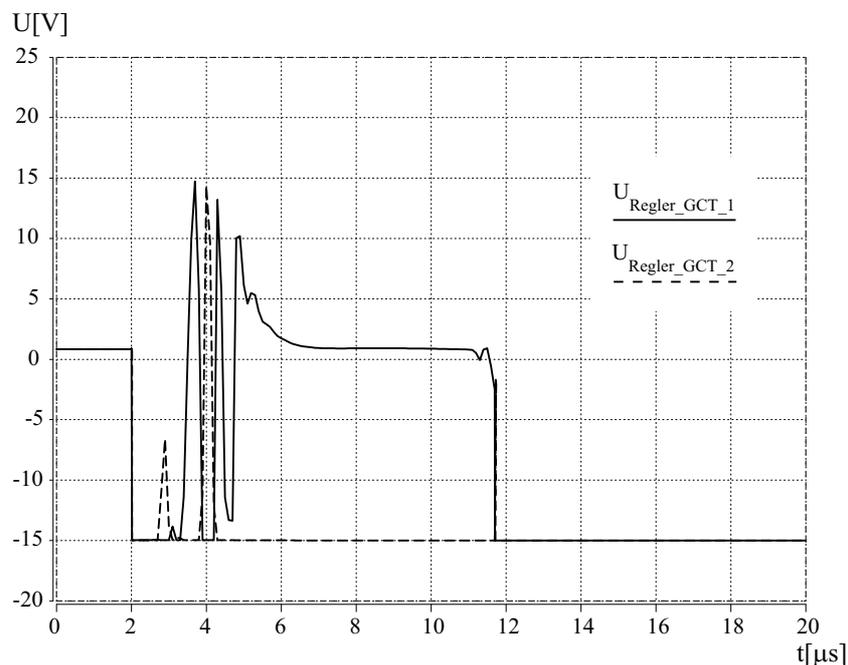


Bild 6.6: Simulation zum Schema in Bild 6.3: Ausgangsspannung des Reglers plus U_{Goff} der beiden IGCTs.

Wie man sieht, funktioniert diese Konzept sehr gut zur Symmetrierung der Blockierspannung. Allerdings bleibt zu untersuchen, ob der Einbruch des Gatestroms bis zu 800A nicht Schwierigkeiten hinsichtlich der Filamentierungsproblematik macht, da diese 800A zwangsläufig über die Kathode abfließen und der Gate-Kathoden-Übergang temporär wieder einschaltet (mehr dazu in Abschnitt 6.4.1). Im obigen Beispiel war $K_p = 1$. Dies bedeutet wegen $U_G = U_{Goff} + K_p \cdot U_\Delta$, daß $U_\Delta \geq |U_{Goff}|$ gelten muß, damit $U_G > 0$ gilt. Dies ist Voraussetzung dafür, daß $dI_G/dt = U_G/L_G > 0$ ist und damit I_G betragsmäßig kleiner werden kann, mit der Folge, daß der IGCT gebremst wird.

Bislang wurde die Meßwerterfassung und der P-Regler ideal mit gesteuerten Quellen modelliert. Gerade bei schnellen Schaltvorgängen, bei denen auch noch hohe Ströme fließen, sind derartige Schaltungen an hohe Anforderungen geknüpft. Diese Anforderungen sind Gegenstand des folgenden Abschnitts 6.1.4.

6.1.4 Allgemeine Anforderungen

Vor dem Hintergrund, daß die hier auftretenden Ströme und Spannungen sehr hoch sind, kann der Abschaltvorgang von IGCTs als sehr schnell bezeichnet werden. Insbesondere ist der primäre Aktionsbereich des Reglers mit etwa $1\mu s$ sehr

kurz. Da während dieses Zeitraums der Spannungsanstieg sehr schnell mit $dU/dt \approx 2-3kV/\mu s$ erfolgt, muß die Regelung gleichfalls sehr schnell sein, d.h. es darf nur wenig Zeit zwischen der Erfassung und Bestimmung der Spannungsdifferenz bis hin zur Reaktion des Stellgliedes verstreichen. Kritisch ist hierbei neben diversen Verzögerungszeiten, vor allem die Grenzfrequenz, die hauptsächlich vom Stellglied determiniert wird. Untersuchungen in [8] zeigen, daß die Grenzfrequenz deutlich größer als $1MHz$ sein muß. Damit der Regler möglichst früh in Aktion tritt, muß die Spannungserfassung bereits wenige Volt Spannungsunterschied detektieren können, was angesichts der hohen Spannungen, die zu subtrahieren sind, keine leichte Aufgabe ist. Im Abschnitt 6.2 wird eine auf dem vorliegenden Konzept beruhende, entwickelte Schaltung vorgestellt und per Simulation untersucht. Die Simulationen zeigen, daß die Schaltung diese Anforderungen erfüllen kann. Schon jetzt sei in diesem Zusammenhang darauf hingewiesen, daß diese Simulationen, auf aussagekräftigen Modellen mit realistischen Parametern für die einzelnen Komponenten beruhen und damit praxisrelevante Ergebnisse liefern.

6.2 Schaltungstechnische Umsetzung

6.2.1 Schaltungsblöcke

Die entwickelte Schaltung besteht für jeden einzelnen Serien-IGCT aus folgenden Funktionsblöcken: die Spannungserfassung und Differenzspannungsbildung, dem diskreten und analogen P-Glied und der Endstufe, die das jeweilige Gate mit dem notwendigen Strom, sowohl zum Ein- als auch zum Ausschalten zur Verfügung stellt. Außerdem kann mittels einer Gatestromerfassung an der Endstufe, der Gatestrom zur Adaption des Verstärkungsfaktors K_P des P-Reglers verwendet werden. Der Gatestrom ist ein Maß dafür, wie weit der Schaltvorgang im Zeitablauf bereits fortgeschritten ist. Damit kann z.B. ein unbeabsichtigtes Einschalten von IGCTs aufgrund der statischen Spannungsdifferenz am Ende des Abschaltens verhindert werden, indem K_P zu diesem Zeitpunkt null bzw. sehr klein eingestellt wird. Die Schaltung wird vervollständigt durch eine Ein/Aus-Logik und den Quellen, die die Impulse zum Ein- bzw. Ausschalten zur Verfügung stellen. Bild 6.7 zeigt den Blockschaltplan, der die genannten Funktionsblöcke, bzw. die Komponenten, die bereits aus den vorhergehenden Kapiteln bekannt sind, für einen repräsentativen IGCT, der wieder GCT_i genannt wird, enthält. Die Serienschaltung bestehe aus n IGCTs von denen der oberste GCT₁ und der unterste GCT_n genannt wird. Der GCT_i in Bild 6.7 ist ein beliebiger innerer IGCT aus dieser Serienschaltung, außer GCT₁ und GCT_n. Der Grund für diese Unterscheidung ist folgender: Die Erläuterungen zur Struktur der Funktionsblöcke orientiert sich am allgemeinsten Fall, d.h. den

inneren IGCTs. Bei GCT₁ bzw. GCT_n können zum Teil einfachere Strukturen verwendet werden, was unmittelbar einleuchtet, da z.B. die Kathode des GCT_n direkt an die Masse angeschlossen ist.

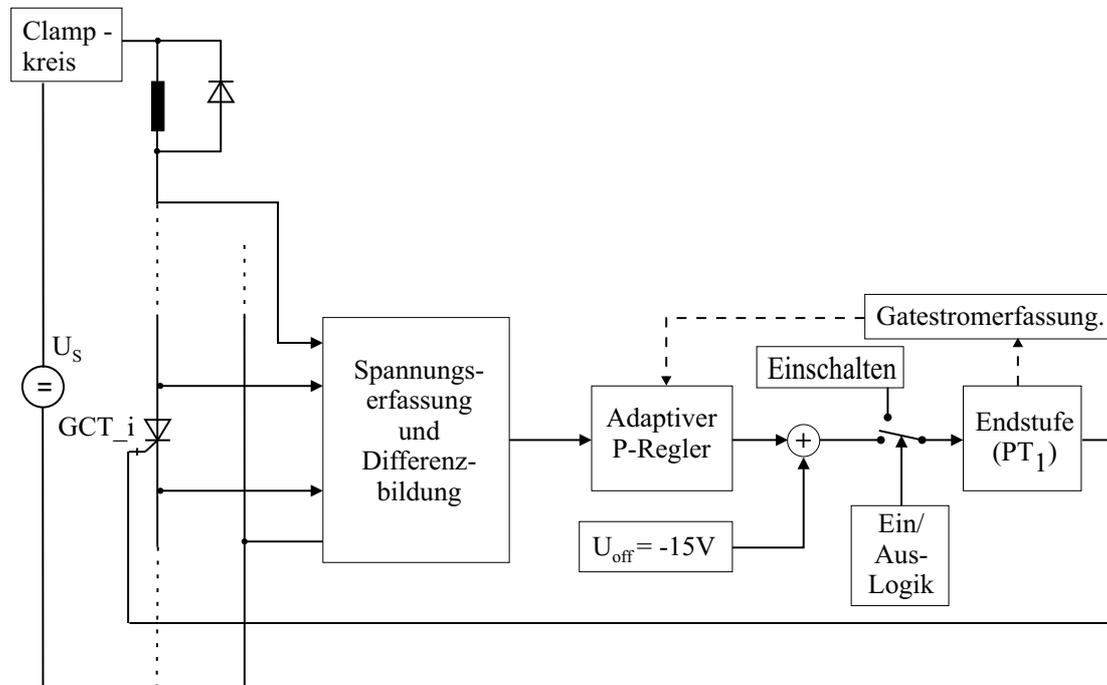


Bild 6.7: Umsetzung der Aktiven Gateansteuerung: Blockschaltbild

In den folgenden Abschnitten werden die Schaltungsprinzipien der einzelnen Funktionsblöcke erläutert. Die genaue schaltungstechnische Umsetzung findet sich detailliert in [8].

6.2.2 Spannungserfassung und Differenzbildung

Der Funktionsblock „Spannungserfassung und Differenzbildung“ umfaßt die Teilaufgaben Messung der Istspannung U_{GCT} , Bestimmung der Sollwertspannung U_{ges}/n und deren anschließenden Subtraktion, so daß man die Differenz $U\Delta$ erhält.

Grundsätzlich kommen für solche Aufgaben, Spannungssubtrahierer, die aus Operationsverstärkern (OPV) bestehen, in Frage. Allerdings erfüllen Schaltungen mit mehreren OPV hintereinander, nicht die strengen Zeitanforderungen, die gemäß Abschnitt 6.1.4 zu stellen sind. Wie oben erwähnt, müssen bereits Spannungsdifferenzen von wenigen Volt detektiert werden können. Voneinander abgezogen werden aber Spannungen von einigen Kilovolt. Wenn man bedenkt, daß es sich um

eine mehrkettige OPV-Schaltung handelt, bewirken Offsets am Anfang der Kette eine fehlerhafte Differenzbildung am Ende der Kette, zumal die Differenz verstärkter Offsets, größer sein kann, als die zu bestimmende Spannungsdifferenz. Ähnlich gravierend ist die Tatsache, daß das Signal zur Sollwertbildung von einer größeren Anzahl OPVs hintereinander prozessiert wird, als das detektierte Signal des Istwerts. Diese Laufzeitunterschiede führen am Subtrahierer zu einer nicht zeitgerechten Subtraktion und zu völlig falschen Spannungsdifferenzen. Es gibt zwar einerseits Operationsverstärker mit geringen Laufzeiten und andererseits OPV-Typen mit sehr kleinen Offsets und Bias-Strömen am Markt. Doch scheinen diese beiden Größen ein Optimierungsgegensatz zu sein, d.h. die Stärke des einen Typs ist jeweils die Schwäche des Anderen, so daß keine Operationsverstärker erhältlich sind, die beiden Anforderungen in genügendem Maße entsprechen. Letztlich muß dieser Ansatz also verworfen werden.

Alternativ wurde ein einfacher und effektiver Ansatz zur schnellen Fehlerspannungsbildung erdacht, der auf der Subtraktion äquivalenter Ströme an einem Knoten beruht. Die Schaltung dazu zeigt Bild 6.8 und besteht prinzipiell nur aus einem Stromspiegel und mehreren Widerständen. Zur besseren Übersichtlichkeit wird in dem Bild eine Serienschaltung bestehend aus nur drei IGCTs betrachtet. Erfasst werden soll die Spannungsabweichung des mittleren (inneren) IGCTs $U\Delta = U_{GCT} - U_{ges}/3$, oder allgemein $U\Delta = U_{GCT} - U_{ges}/n$. Die Spannung U_{ges} über alle IGCTs wird in die zwei Teilspannungen U_{ges1} und U_{ges2} zerlegt. Im allgemeinen Fall gelten folgende Definitionen, wobei die Spannungsdifferenz am GCT i betrachtet wird.

$$\begin{aligned} U_{ges} &= \sum_{j=1}^n U_{gctj} \\ U_{ges1} &= \sum_{j=1}^i U_{gctj} \\ U_{ges2} &= \sum_{j=i+1}^n U_{gctj} \end{aligned} \quad (6.1)$$

Im vorliegenden Beispiel kann in (6.1), i durch 2 und n durch 3 ersetzt werden. Mit der Masche M1 kann ein Strom I_1 proportional zu U_{ges1} generiert werden: Durch den Widerstand $R_1 = 3R$ (allgemein $R_1 = nR$) fließt zwingend $I_1 = U_{ges1}/3R$ (allgemein $I_1 = U_{ges1}/nR$). In ähnlicher Weise wird mit der Masche M2 der Strom I_2 proportional zu U_{ges2} erzeugt: Durch den Widerstand $R_2 = 3R$ (allgemein $R_2 = nR$) fließt zwingend $I_2 = U_{ges2}/3R$ (allgemein $I_2 = U_{ges2}/nR$). Durch den Stromspiegel wird I_1 1:1 gespiegelt, so daß aus dem Knoten s sowohl I_1 als auch I_2 herausfließt. Folglich fließt dem Knoten der Strom $I_{1+2} = I_1 + I_2$ zu, der gleichzeitig aus dem Knoten „Stromsubtrahierer“ herausfließt. Dieser Summenstrom läßt sich berechnen zu $I_{1+2} = (U_{ges1} + U_{ges2}/3R)$, was gleichbedeutend mit $U_{ges}/3R$ ist. Auf R normiert bekommt man also einen Strom entsprechend dem Sollwert $U_{ges}/3$ (allgemein U_{ges}/n).

Die Masche M3 erzeugt den Strom $I_3 = U_{GCT}/R$, der somit auf R normiert, dem Istwert U_{GCT} entspricht und in den Knoten „Stromsubtrahierer“ hineinfließt. Am Ausgang des Knoten „Stromsubtrahierer“ fließt der Differenzstrom $I\Delta = I_3 - I_{1+2}$, der normiert auf R , allgemein $U\Delta = U_{GCT} - U_{ges}/n$ entspricht. Damit hat man die gewünschte Regelgröße $U\Delta$ erhalten. Existiert keine Spannungsdifferenz, so fließt auch kein Ausgangsstrom $I\Delta$ aus dem Knoten „Stromsubtrahierer“ heraus.

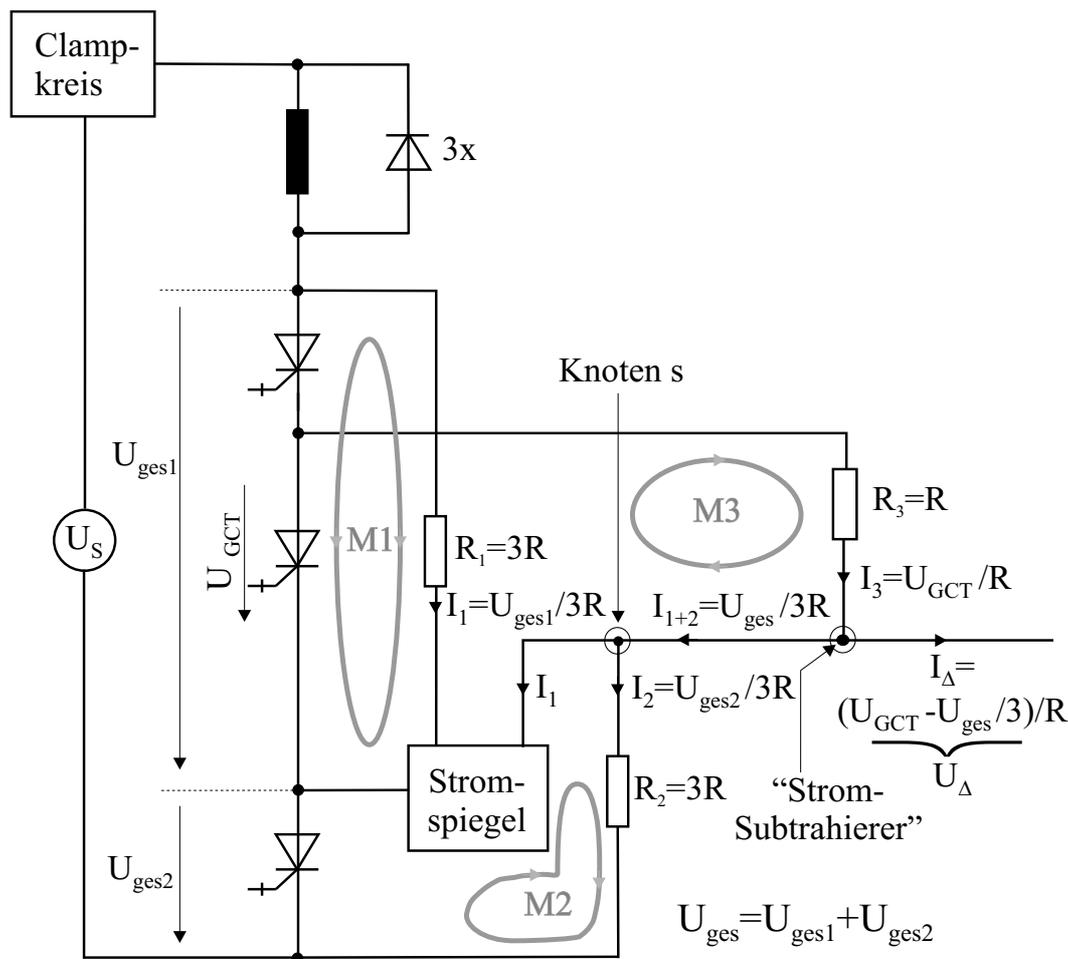


Bild 6.8: Struktur des „Stromsubtrahierers“.

Mit dieser Schaltung werden mit vergleichsweise geringem Aufwand all die Nachteile der OPV-Subtraktionsschaltungen vermieden. Es existieren keine Offsets, praktisch keine Laufzeiten und die Subtraktion am Knoten geschieht nahezu unendlich schnell. Als Nachteil ist zu erwähnen, daß die erlaubten Toleranzen der Widerstände nur sehr klein dürfen, da sonst falsche $I\Delta$ generiert werden. Die Auswahl der einzelnen Widerstände kann dadurch entsprechend aufwendig werden. Daß in den Widerständen $R_{1,2,3}$ auch im statischen Aus-Zustand auch Wirkleistung verbraucht wird, ist ein weiterer Nachteil. Um diese Verluste zu minimieren, sollte R möglichst groß gewählt

werden. Für das Beispiel mit 3 IGCTs in Serie hat sich der Wert $R = 200k\Omega$ als sinnvoll erwiesen.

Die tatsächliche Schaltung besteht aus den Widerständen und einem sehr schnellen Differenzverstärker, bestehend aus JFETs und zwei Operationsverstärkern anstelle des Stromspiegels. Die Schaltungen für den obersten und untersten IGCT sind noch etwas einfacher aufgebaut.

Der Ausgang der Schaltung liefert einen Fehlerstrom $I\Delta$, der in einem nachgeschalteten I/U–Verstärker, bestehend aus Operationsverstärkern, in die richtige Spannungsdifferenz $U\Delta$ zurückgewandelt wird. Dazu muß ein Verstärkungsfaktor gewählt werden, der beim vorliegenden I/U–Verstärker genau dem Widerstandswert R , im Beispiel also $200k\Omega$ entspricht. Der Verstärker geht bei $U = 30V$ in Sättigung, d.h. $U\Delta$ am Ausgang des I/U–Verstärkers kann Werte im Bereich $[0..30V]$ annehmen.

Weiterhin erfolgt die Addition mit der Abschaltspannung $-15V$ noch vor der eigentlichen Reglerschaltung. Dies ist möglich ist, weil es sich hierbei um einen linearen Regler handelt. Man muß sich allerdings vor Augen halten, daß dann mit $K_P \cdot (-15V + U\Delta)$ und nicht mit $-15V + K_P \cdot U\Delta$ geregelt wird, was aber der prinzipiellen Funktionalität und der Performance der Schaltung selbst keinen Abbruch tut. Dazu benötigt man einerseits eine Potentialverschiebung des Ausgangssignals $U\Delta$ des I/U–Verstärkers um $-15V$ und andererseits eine Stauchung der potentialverschobenen Spannung auf den Bereich $[-12..12V]$. Diese Stauchung erfolgt deshalb, weil der Leistungstreiber, der das Reglersignal weiterverarbeitet, einen Verstärkungsfaktor in Höhe von 1,25 besitzt und man dadurch am Ausgang des Leistungstreibers wieder auf den gewünschten Spannungsbereich $[-15..15V]$ gelangt.

Details zu den einzelnen beschriebenen Teilschaltungen und zu den verwendeten Komponenten, finden sich in [8].

6.2.3 Der Leistungstreiber

Da der Leistungstreiber das Signal zur Adaption des Verstärkungsfaktors K_P zur Verfügung stellt, wird die Beschreibung der Endstufe, der Erläuterung des Reglers vorgezogen. Der Leistungstreiber muß in der Lage sein, die z.T. sehr hohen Gateströme zur Verfügung zu stellen. Außerdem sei daran erinnert, daß die Anforderungen bzgl. der Grenzfrequenz sehr hoch sind. Mit einer Endstufe, die aus einer Parallelschaltung verfügbarer Leistungs–MOSFETs besteht, lassen sich einerseits die Anforderung bezüglich des Gatestroms erfüllen und andererseits Grenzfrequenzen von immerhin $3MHz$ realisieren. Dazu müssen die Gates der MOSFETs von einer geeigneten Treiberstufe, die den Gatekapazitäten genügend hohe Ströme zum Umladen der Gatekapazitäten (Parallelschaltung!) liefern kann, versorgt werden.

Die Ausgangsspannung des Reglers $K_P \cdot (-15V + U\Delta)/1,25$ ist gleichzeitig die Eingangsspannung am Leistungstreiber, die im folgenden einfach U_{in} genannt wird. Der Leistungstreiber muß die einwandfreie Funktion der Schaltvorgänge der IGCTs gemäß Tabelle 6.1 sicherstellen können. In der Tabelle 6.1 wird dem jeweiligen Eingangspegel, die Flußrichtung des benötigten Ausgangsstroms gegenübergestellt.

Tabelle 6.1: *Eingangsspannung und Ausgangstrom am Leistungstreiber verschiedener Schaltvorgänge*

Einschalten	$U_{in} > 0$	$I_G > 0$
Ausschalten	$U_{in} < 0$	$I_G < 0$
Symmetrieren (Regeln)	$U_{in} > 0$	$I_G < 0$, aber $dI_G/dt > 0$ (betragsmäßiger Rückgang von I_G)

Um den jeweils benötigten Strom I_G sicherstellen zu können, braucht man somit einen Linearverstärker, der insbesondere negativen Ausgangsstrom bei positivem Eingangspegel (und positivem Ausgangspegel) zum Regeln zuläßt. Dies ist notwendig, um $dI_G/dt > 0$ an der Gateinduktivität zu ermöglichen. Da der negative Gatestrom im Betrag nur abgesenkt wird aber negativ im Vorzeichen bleibt, benötigt man eine Verstärker, der quasi eine Art Vier-Quadranten-Betrieb zuläßt. Der Leistungstreiber besteht aus zwei Stufen, nämlich einer Endstufe, deren parallelgeschaltete MOSFETs von einer vorgeschalteten Treiberstufe getrieben werden.

Für die Endstufe eines Linearverstärker kommen dafür prinzipiell die drei Topologien in Betracht, die Bild 6.9 zeigt. Grundsätzlich sind n-Kanal MOSFETs den p-Kanal-Typen vorzuziehen, da diese einen erheblich niedrigeren Ein-Widerstand R_{on} besitzen, was damit zusammenhängt, daß die Elektronenbeweglichkeit um etwa den Faktor drei größer ist als die entsprechende Löcherbeweglichkeit (vgl. Abschnitt 1.5). Folglich muß man weniger n-Kanal MOSFETs parallelschalten, als dies für p-Kanal MOSFETs der Fall ist. Allein unter diesem Gesichtspunkt wäre Topologie A die Struktur der Wahl. Allerdings hat Topologie A einen gravierenden Nachteil: Es gibt kein gemeinsames Bezugspotential der Gate-Source-Signale und die Potentialdifferenz der beiden Bezugspotentiale hängt zudem vom Ausgang ab. Dies erfordert eine aufwendige Signalkonditionierung in der vorhergehenden Treiberstufe, um die für die Endstufe benötigten Gate-Source-Signale zu erhalten. Die dadurch bedingte erhöhte Komplexität der vorgeschalteten Treiberstufe führt dazu, daß die Anforderungen bezüglich Grenzfrequenz bzw. erlaubter Signallaufzeiten nicht mehr zu erfüllen sind. Bei der Topologie B leiten die parallelen n-Kanal MOSFETs den (höheren) Abschaltestrom des IGCT-Gate und die hochohmigeren p-Typen den (niedrigeren) Einschaltestrom. Auch die Potentialdifferenz der bei-

den Bezugspotentiale ist mit $2 \cdot U_{bb}$ konstant und unabhängig vom Ausgangspegel. Allerdings müssen die beiden Gate-Source-Signale zueinander um 180° phasenverschoben sein, was ebenfalls eine aufwendigere Treiberstufe erfordert. Da außerdem die Spannungsverstärkung der Endstufe B vom Betriebspunkt abhängt, ist auch dieses Konzept letztlich unbrauchbar.

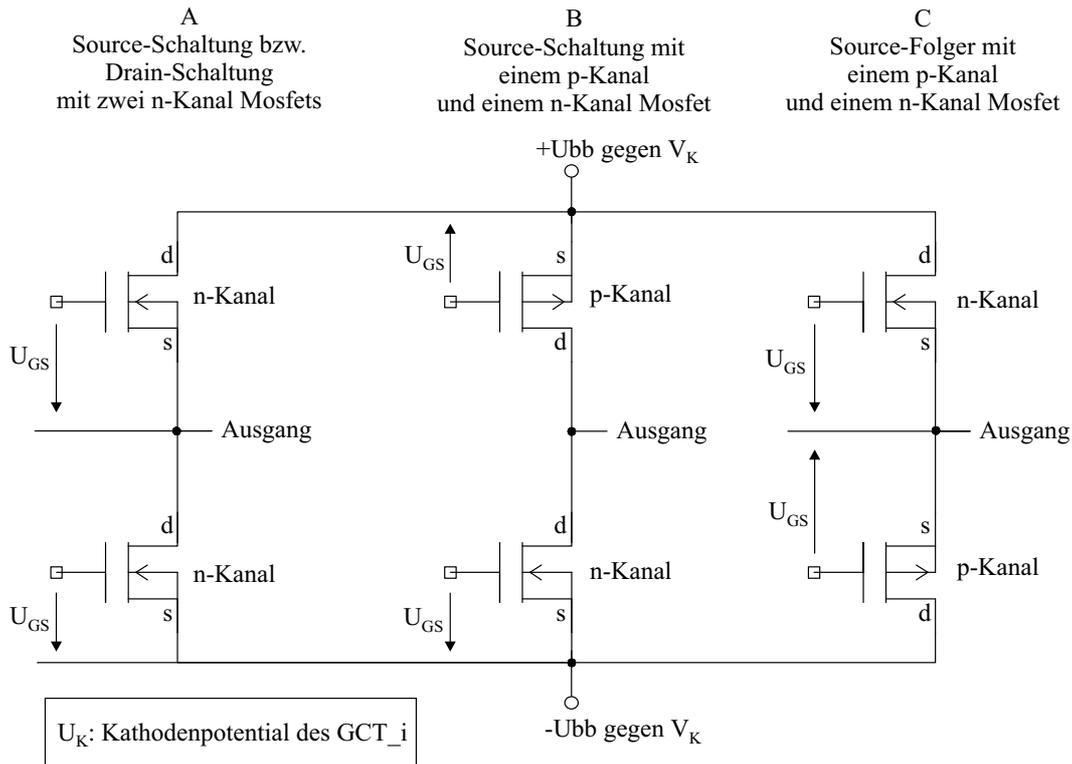


Bild 6.9: Mögliche Topologien für die Endstufe

Bei der Topologie C ist die Tatsache, daß der hohe Abschaltstrom über die hochohmigen p-Kanal-MOSFETs fließt, ein Nachteil. Andererseits kommt man aufgrund des gemeinsamen Bezugspunkts mit einem einzigen Gate-Source-Signal aus, was eine weniger komplexe Treiberstufe und damit die Einhaltung der oben genannten Anforderungen ermöglicht. Ferner ist die Spannungsverstärkung der Source-Folger konstant und etwa eins, d.h. $U_{out} \approx U_{in}$. Daher wurde die Topologie C für die Endstufe weiterverfolgt.

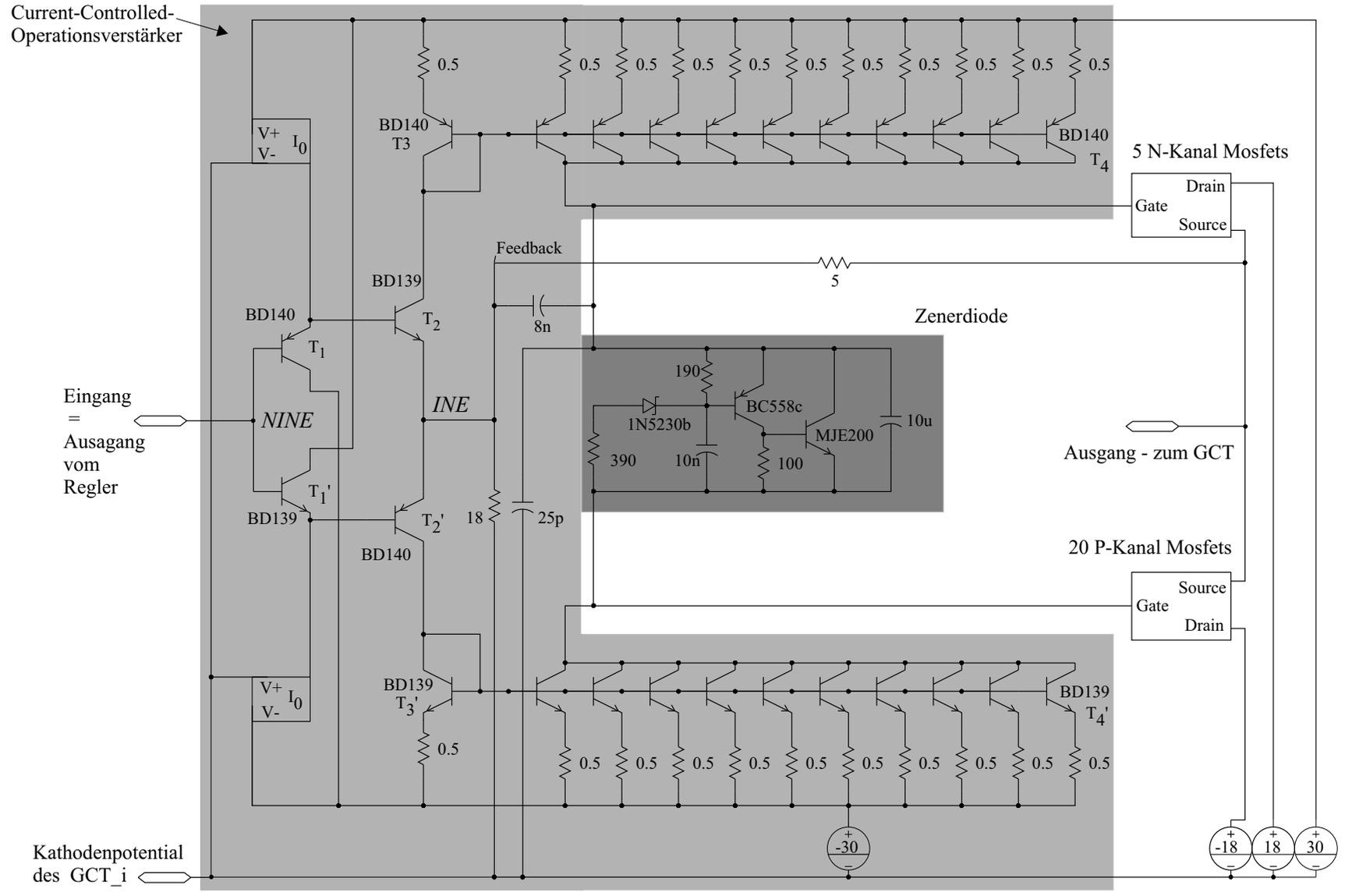
Der z.T. hohe negative Gatestrom (ca. $3kA$) beim Abschalten eines IGCTs fließt damit über die parallelen p-Kanal MOSFETs, wohingegen der positive Gatestrom (einige $100A$) zum Einschalten über die niederohmigen n-Kanal MOSFETs fließt. Beim Symmetrieren fließt wegen $U_{out} \approx U_{in}$ und $I_G < 0$ (s. Tabelle 6.1) der negative, betragsmäßig verringerte Gatestrom über die n-Kanal MOSFETs, die damit invers betrieben werden. Letztlich werden gute Ergebnisse erzielt, indem zwanzig

am Markt erhältliche p-Kanal MOSFETs und fünf n-Kanal MOSFETs parallelgeschaltet werden, wobei hier ein virtuelles Parallelschalten mit realistischen Modellen zu Simulationszwecken gemeint ist. Die Spannung U_{bb} wurde mit $18V$ etwas höher als das Regelspannungsmaximum von $15V$ gewählt, damit ein entsprechend großes U_{DS} einen genügend hohem Drainstrom gewährleistet.

Bei einer Parallelschaltung von MOSFETs sind die effektiven MOS-Kapazitäten, d.h. die einzelnen Gate-Source-Kapazitäten und die Gate-Drain-Kapazitäten (Miller-Kapazitäten), ebenfalls parallel geschaltet und addieren sich. Somit benötigt man einen relativ hohen Ladestrom, der von der Treiberstufe, die der Endstufe vorgeschaltet ist, geliefert werden muß. Eine Topologie, die diese Forderung bewerkstelligt, basiert auf dem sogenannten Current-Controlled-Operationsverstärker (CC-OPV) [71]. Im Bild 6.10, das den kompletten Leistungstreiber zeigt, ist der diskret aufgebaute CC-OPV grau hinterlegt, wobei die im Bild genannten Bipolartransistoren verwendet wurden. Die Transistoren T_1 , T'_1 , T_2 und T'_2 bilden einen Spannungsfolger, d.h. einen Impedanzwandler für den $U_{aus} = U_{in}$ gilt. Die wohl bekannteste Topologie für einen Spannungsfolger, ist der nicht-invertierende Verstärker, bestehend aus einem Operationsverstärker bei dem der invertierende Eingang direkt mit dem Ausgang des OPV verbunden wird, während die Eingangsspannung am nicht-invertierenden Eingang anliegt. (s. [71]). Im Bild 6.10 entspricht der nicht-invertierende Eingang (*NINE*) dem Eingang des CC-OPVs und ist mit den Basisanschlüssen von T_1 und T'_1 verbunden. Der invertierende Eingang *INE* befindet am Knoten zwischen den Emittern von T_2 und T'_2 und wird zum Ausgang des Leistungstreibers über einen Widerstand verbunden.

Die Impedanz des invertierenden Eingangs *INE* ist beim CC-OPV sehr niedrig. Dies bedeutet, daß mit den Stromquellen T_3 und T'_3 , ein Strom abhängig von der Spannung am *INE* erzeugt wird. Am *INE* liegt wegen der Spannungsfolgerfunktion bekanntlich die Regelgröße, so daß der erzeugte Strom prinzipiell zum Laden der MOS-Kapazitäten verwendet werden kann. Da dieser aber noch nicht ausreicht wird er mit den beiden Stromspiegeln, bestehend aus T_3 , T_4 und T'_3 , T'_4 übersetzt und verstärkt. Den Transistor T_4 (T'_4) erhält man, indem man 10 Bipolartransistoren vom selben Typen wie bei T_3 (T'_3) parallelschaltet, was zu einer Stromverstärkung im Verhältnis 10:1 führt. Damit erreicht man die notwendigen Ladeströme für die MOS-Kapazitäten bis zu einer Höhe von etwa $10A$, was ein ausreichend schnelles Schalten und damit die Einhaltung der Grenzfrequenz ($f_g \approx 3MHz$) gewährleistet. Ferner werden beim CC-OPV noch zwei Stromquellen benötigt, die den Ruhestrom I_0 für T_1 und T'_1 zur Verfügung stellt. Der Ruhestrom I_0 wird dabei abhängig vom Potentialunterschied $V_+ - V_-$ an der Quelle eingestellt. Der genaue diskrete Aufbau dieser Quellen wird hier nicht wiedergegeben (s. dazu [8]). Des weiteren ist eine Leistungszenerdiode, die aus einer Zenerdiode, zwei Bipolartransistoren, Widerständen und Kondensatoren besteht, in Bild 6.10 eingezeichnet und ebenfalls grau hinterlegt.

Bild 6.10: Leistungstreiber, bestehend aus Treiberstufe plus Endstufe



Diese Leistungsenerdiode sorgt für eine, bzgl. dem Mittelwert konstante Vorspannung an den MOS-Gateanschlüssen der Source-Folger-Endstufe. Diese Endstufe stellt bei entsprechender Vorspannung ebenfalls einen Spannungsfolger dar. Der Parallelkondensator $C = 25\text{pF}$ hält die Vorspannung während der Ladestromstöße konstant.

Insgesamt erfüllt der vorliegende Leistungstreiber die geforderte Grenzfrequenz von ca. 3MHz und die Eingangsspannung wird um den oben genannten Faktor 1,25 verstärkt. Schließlich enthält der Leistungstreiber, genauer gesagt die Endstufe, noch eine weitere Funktion: Sie stellt ein Signal, geeignet zur Adaption der Reglerverstärkung K_P zur Verfügung, mit dem der Fortgang des Abschaltvorgangs erfaßt werden kann. Hierbei handelt es sich um die Gate-Source-Spannung an den P-Kanal-MOSFETs, die direkt proportional zum jeweiligen Drainstrom ist. Die Drainströme der parallelen MOSFETs ergeben in Summe den negativen Gatestrom des IGCT. Somit ist diese Gate-Source-Spannung U_{GS} ein Maß für den Gatestrom des IGCT.

6.2.4 Analoger adaptiver P-Regler

Die Schaltung eines P-Reglers ist einfach aufgebaut: um $K_P < 1$ zu realisieren, reicht ein Spannungsteiler aus; für $K_P > 1$ kann man einen nichtinvertierenden Verstärker, bestehend aus einem OPV und zwei Widerständen verwenden und für $K_P = 1$ braucht man gar keine Schaltung. Etwas komplizierter ist ein P-Reglerschaltung bei dem K_P adaptiv eingestellt wird, wie dies Bild 6.11 zeigt. Wie oben bereits erläutert dient die Spannung U_{GS} der p-Kanal-MOSFETs zur Adaption von K_P .

Gemessen wird U_{GS} mit dem in Bild 6.11 grau hinterlegten Spannungssubtrahierer, der als Standard-Operationsverstärkerschaltung realisiert ist. Die Adaption geschieht mit der ebenfalls grau hinterlegten einstellbaren Zenerdiode. Der Verstärkungsfaktor K_P wird grundsätzlich über den Spannungsteiler, gebildet aus R_1 und R_2 eingestellt. Mit der einstellbaren Zenerdiode kann das Bezugspotential des Spannungsteilers abhängig von der Ausgangsspannung des OPVs variiert werden. Je positiver dieses Bezugspotential ist, desto größer ist die Verstärkung K_P . Entsprechend geringer wird K_P bei immer negativer werdenden Bezugspotential. Dabei sorgt das RRCD-Netzwerk für einen verzögerten Anstieg der Spannung U_{BE} des Transistors T_1 der einstellbaren Zenerdiode. Damit läßt sich das Absinken von K_P im Laufe des Abschaltvorgangs nach Wunsch verzögern, so daß eine optimale Funktion des adaptiven P-Reglers erzielt werden kann. Ferner befindet sich zwischen dem RRCD-Netzwerk und dem Transistor T_1 eine Zenerdiode, die zu einer Potentialverschiebung der Ausgangsspannung des OPVs führt.

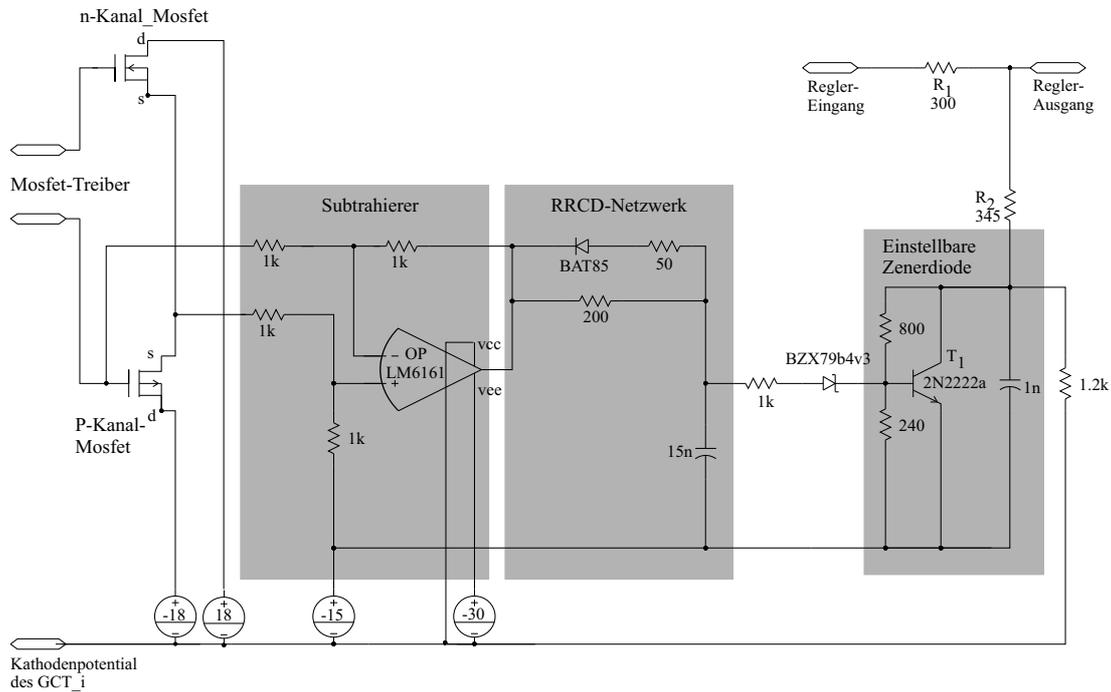


Bild 6.11: Schaltung des adaptiven P-Reglers

Am Ende des Abschaltvorgangs ist T_1 dann voll durchgesteuert und K_P so niedrig, daß ein unerwünschtes Wiedereinschalten des IGCTs vermieden wird.

Abschließend wird noch kurz auf das Einschaltnetzwerk eingegangen, um die Gesamtschaltung „Aktive Ansteuerung“ zu komplettieren. In Bild 6.12 wird die Schaltung zum Einschalten des IGCTs gezeigt. Sie befindet sich zwischen dem Ausgang des Reglers und dem Eingang des Leistungstreibers am Knoten „z“.

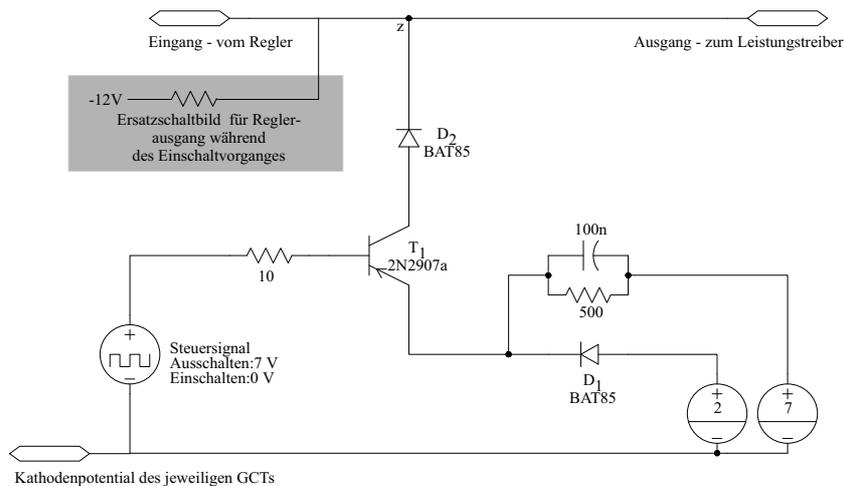


Bild 6.12: Netzwerk zum Einschalten des IGCTs

Erhält der Transistor T_1 kein positives Signal U_{BE} , ist er gesperrt und trennt die beiden positiven Quellen vom Potential am Knoten „z“, d.h. die Ausgangsspannung des Reglers liegt am Eingang des Treibers. Der Reglerausgang, der in Bild 6.12 als einfaches Ersatzschaltbild dargestellt ist, besitzt eine relativ große Impedanz von einigen 100Ω . Wird nun T_1 von der Einschaltelogik eingeschaltet, führt dies dazu, daß beide positiven Quellen niederohmig auf den Knoten „z“ aufgeschaltet werden. Damit wird der Treibereingang vom hochohmigen Reglerausgang getrennt. Zunächst dominiert die $7V$ -Quelle am Knoten „z“. Das RC -Netzwerk stellt sicher, daß dies nur kurzzeitig geschieht, d.h. ein Einschaltimpuls für den IGCT erzeugt wird. Somit wird nach kurzer Zeit der Parallelwiderstand von 500Ω wirksam und die $7V$ -Quelle wieder abgekoppelt. Damit wird die $2V$ -Quelle für das Potential am Knoten „z“ maßgeblich und der statische Ein-Zustand sichergestellt, bis aufgrund eines Abschaltimpulses, T_1 gesperrt wird.

Damit sind die wesentlichen Teilnetzwerke der Gesamtschaltung zum vorgeschlagenen Ansteuerungskonzept erläutert. Unter Verwendung realistischer Modelle mit entsprechenden Parametern, kann diese Ansteuerschaltung per Simulation untersucht werden. Die Ergebnisse dazu werden unten im Abschnitt 6.3 vorgestellt.

6.3 Simulationsergebnisse

Um die Leistungsfähigkeit der oben vorgestellten Schaltung zu demonstrieren, wird das „Worst Case Szenario“ aus Abschnitt 5.2 mit den drei IGCTs in Serie und jetzt mit der aktiven Ansteuerung simuliert. Die Hochinjektionslebensdauer und die Gateinduktivität weichen dabei beim GCT_3 erheblich von den beiden anderen IGCTs, deren Parameter gleich sind, ab. Ferner gibt es noch eine Zeitverzögerung der Abschaltimpulse des GCT_3². Alles in allem ist der GCT_13 der Langsamste von allen.

Das vereinfachte Simulationsschema zeigt Bild 6.13. Die aufgenommene Spannung der drei IGCTs zeigt Bild 6.14, während Bild 6.15 die zugehörigen geregelten Gateströme zeigt. Die Ausgangsspannung des Leistungstreibers (Stellgröße) während des Abschaltvorgangs zeigt Bild 6.16. Der Ausregelvorgang geschieht prinzipiell genauso wie in Abschnitt 6.1.2 im Zusammenhang mit dem idealen P-Regler beschrieben: zuerst wird also wieder GCT_1, dann wechselseitig alle drei IGCTs gebremst, bis die statische Regeldifferenz erreicht ist. Insbesondere fällt auf, daß im Vergleich zum idealen P-Glied gegen Ende des Schaltvorgangs stärker nachgeregelt wird (s. Bild 6.15). Im Vergleich zur Lösung mit RC-Snubber in Bild 5.13, verbleiben bedeutend niedrigere statische Spannungsdifferenzen im Aus-Zustand. Außerdem verbleibt eine kleine positive Spannung am Ausgang der aktiven Ansteuerung (Treiber)

²zu den Zahlenangaben siehe Seite 133

für den GCT_3 , die allerdings nicht zum Wiedereinschalten des GCT_3 ausreicht.

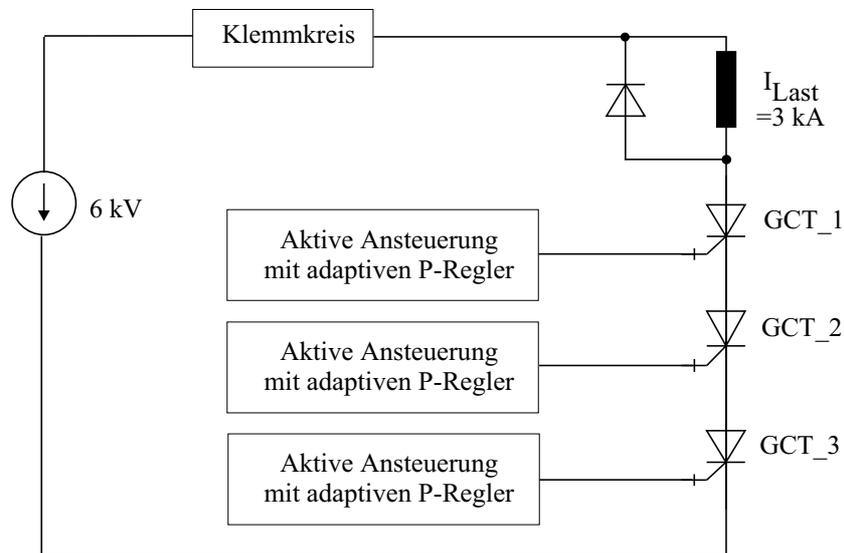


Bild 6.13: Schema mit 3 IGCTs in Serie, jeweils mit aktiver Ansteuerung

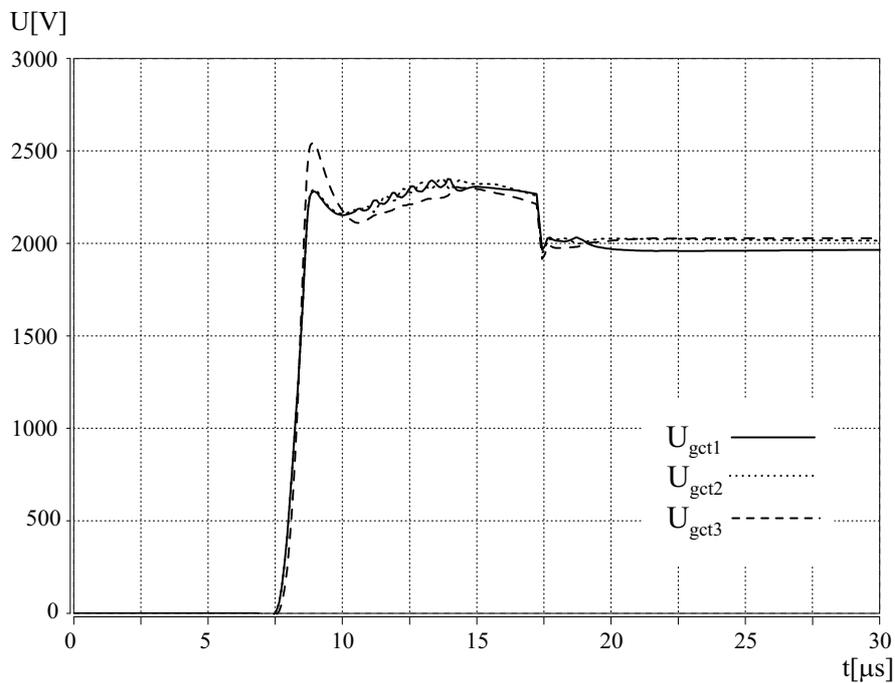


Bild 6.14: SABER-Simulation: Spannungsverläufe beim harten unentlasteten Abschalten von 3kA gegen 6kV mit aktiver Ansteuerung; „Worst Case Szenario“,

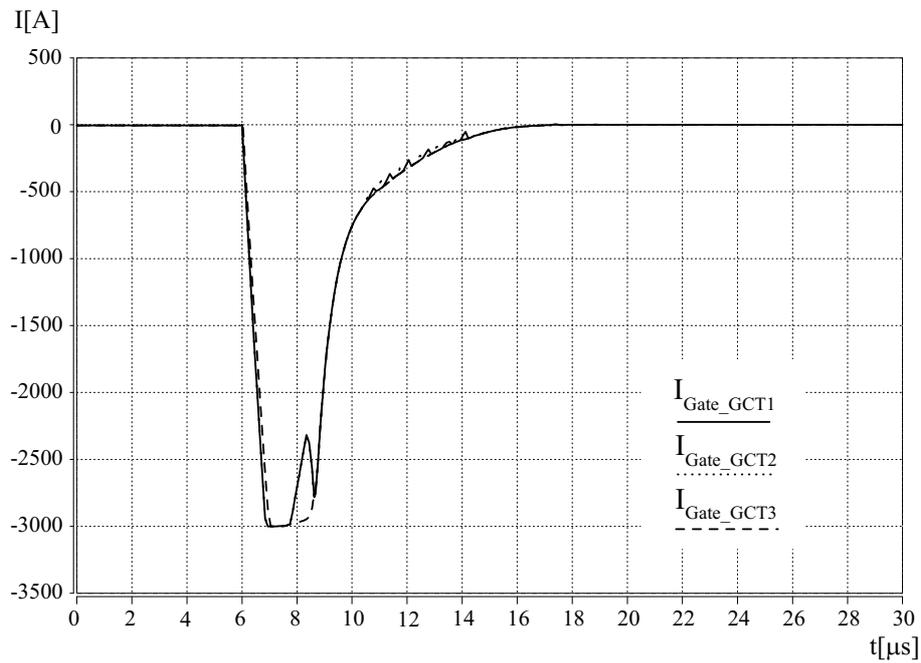


Bild 6.15: SABER-Simulation: Gateströme beim harten unentlasteten Abschalten mit aktiver Ansteuerung; „Worst Case Szenario“,

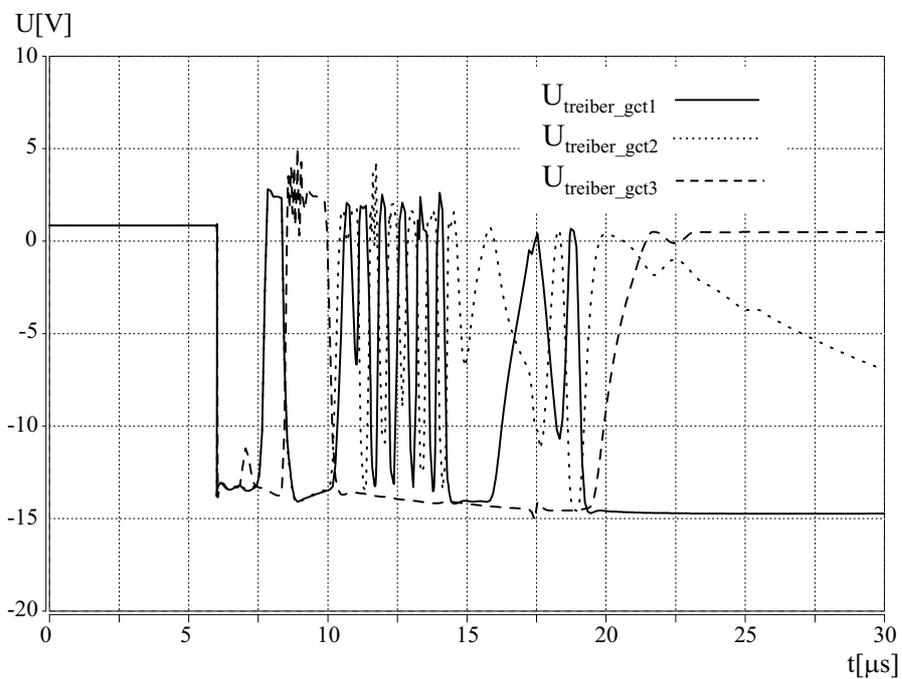


Bild 6.16: SABER-Simulation: zugehörige Ausgangsspannung des Treibers

Das Ergebnis ist also sehr vielversprechend. Während des gesamten Abschaltvorgangs, werden die Spannungen nahezu symmetrisch aufgenommen und die verbleibenden statischen Differenzen sind sehr gering. Abschließend zeigt Bild 6.17 die aufgenommenen Spannungen ohne Adaption des Verstärkungsfaktors, d.h. K_P ist konstant. Auch dann werden die Spannungen ausgeglet, doch sind die Spannungsdifferenzen während des Abschaltens größer.

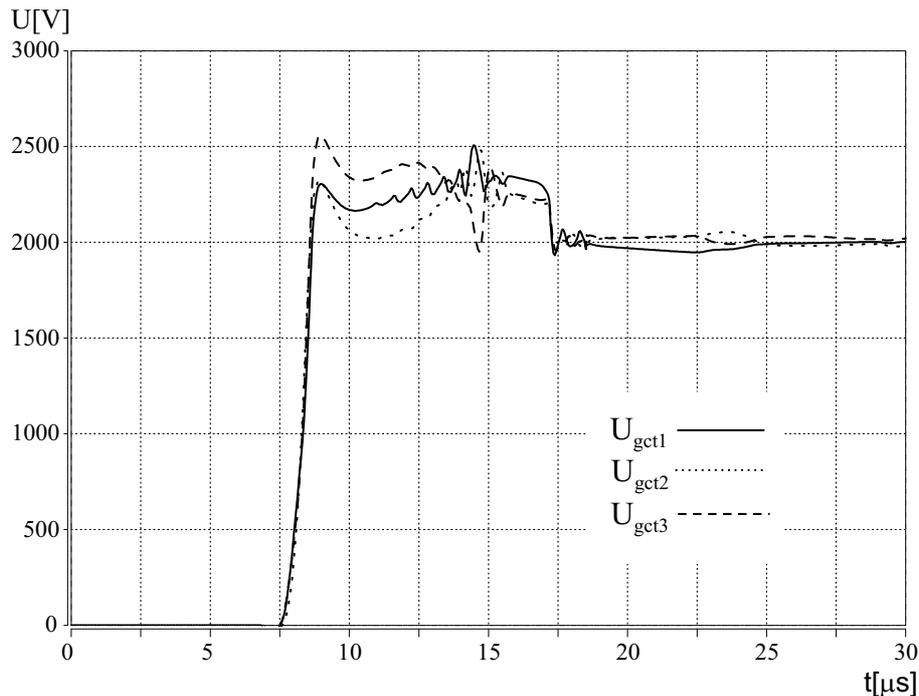


Bild 6.17: SABER-Simulation: Spannungsverläufe beim harten unentlasteten Abschalten von $3kA$ gegen $6kV$ mit aktiver Ansteuerung und $K_P = const.$; „Worst Case Szenario“,

Außerdem bewirken Spannungsdifferenzen gegen Ende des Abschaltvorgangs, stärkere Regleraktivitäten, da K_P nicht abnimmt, was im Extremfall zum ungewollten Wiedereinschalten eines IGBTs führen könnte. Selbst im statischen Aus-Zustand ist dann mit Regleraktivität zu rechnen.

Aus dem betragsmäßigen Gatestromrückgang um etwa $600\text{--}700A$ am GCT_1 bei $t \approx 8,5\mu s$ (s. Bild 6.15), folgt ein Kathodenstrom am GCT_1 in gleicher Höhe. Ob dies zu thermischen Problemen führt im Sinne der im Abschnitt 6.1.2 angedeuteten Filamentierungsproblematik, wird in Abschnitt 6.4.1 näher untersucht.

6.4 Grenzen und mögliche Problemfelder

6.4.1 Aktive Ansteuerung und Erwärmungsverhaltens beim IGCT

Es wurde bereits darauf hingewiesen, daß aufgrund des Regelmechanismus kurzzeitig wieder ein Kathodenstrom größer null fließen muß. Wie weiterhin in Kapitel 3 erläutert, beruht die Möglichkeit des unentlasteten Abschaltens darauf, daß der Gate-Kathoden-Übergang bereits vollständig sperrt, wenn der IGCT, genauer die ν -Zone, mit dem Aufnehmen der Blockierspannung beginnt. Dann ist der Kathodenstrom null und die Kathode kann keine Elektronen mehr in die p -Basis injizieren. Weiterhin kann sich die Raumladungszone ausbreiten, ohne daß Filamente, d.h. sehr dünne und damit hochohmige Elektronenkanäle entstehen, die sich bei $I_K > 0$ lokal überhitzen (Hot Spots) und zur Zerstörung des IGCTs führen.

Im folgenden wird untersucht, ob der kurzzeitig auftretende Kathodenstrom im Zusammenhang mit der Regelung, ebenfalls zu thermischen Problemen führt. Dazu werden die Ergebnisse entsprechender 2D-Simulationen mit DESSIS, unter Berücksichtigung der Temperatur, vorgestellt. Als Struktur dient dabei eine einzelne IGCT-Zelle, deren Geometrie an den Angaben in [21] angelehnt ist. Durch Eingabe eines Faktors wird die Zelle quasi vervielfacht und man erhält insgesamt eine realistische Fläche. Als Ausgangstemperatur dient dabei der realistische Wert von $400K$. Um den Simulationsaufwand bzgl. der Rechenzeit³ zu begrenzen und um Konvergenz zu erzielen, wird erstens die Temperatur am Anodenkontakt konstant bei $400K$ gehalten (ideale Kühlung) und zweitens das numerische Gitter (Grid) nicht allzu fein gewählt. Damit wird es aber passieren, einen möglichen Hotspot nicht genau zu treffen, weil an diesem Ort gerade kein Gitterpunkt ist. Die Ergebnisse ermöglichen somit keine genauen quantitativen, sondern vielmehr tendenzielle Aussagen.

Einmal mehr dient ein Tiefsetzsteller mit einem einzelnen IGCT zu Simulationszwecken. Der Regelvorgang wird nachgebildet, indem dem negativen Abschaltensignal für kurze Zeit ein vom Betrag größerer positiver Impuls überlagert wird. Zum Vergleich werden Simulationen mit harter und weicher Gateansteuerung gezeigt. Zunächst sind als Referenz die Transienten des harten Abschaltvorgangs in Bild 6.18 dargestellt. Danach folgen der „geregelte“ Abschaltvorgang in Bild 6.19 und der weiche Abschaltvorgang (soft turn off) in Bild 6.20⁴. Der jeweilige Temperaturtransient repräsentiert den Ort (Gitterpunkt) mit maximaler Temperatur. Ergänzt werden die drei Bilder noch durch die üblichen Transienten für U_{AK} , I_A , I_G und I_K .

³T \approx 24h mit einer SUN Sparc 20 Workstation

⁴Im Gegensatz zu dieser Simulation benötigt man beim weichen Schalten in der Praxis einen RC(D)-Snubber. In dieser Simulation ist aus Konvergenzgründen der Gatestromanstieg mit weit über $100A/\mu s$ immer noch sehr steil (Praxis beim GTO: $40A/\mu s$) und der maximale Gatestrom mit $2/3 \cdot I_A$ auch relativ groß (Praxis beim GTO: $\hat{I}_G \approx 1/3 \cdot I_A$)

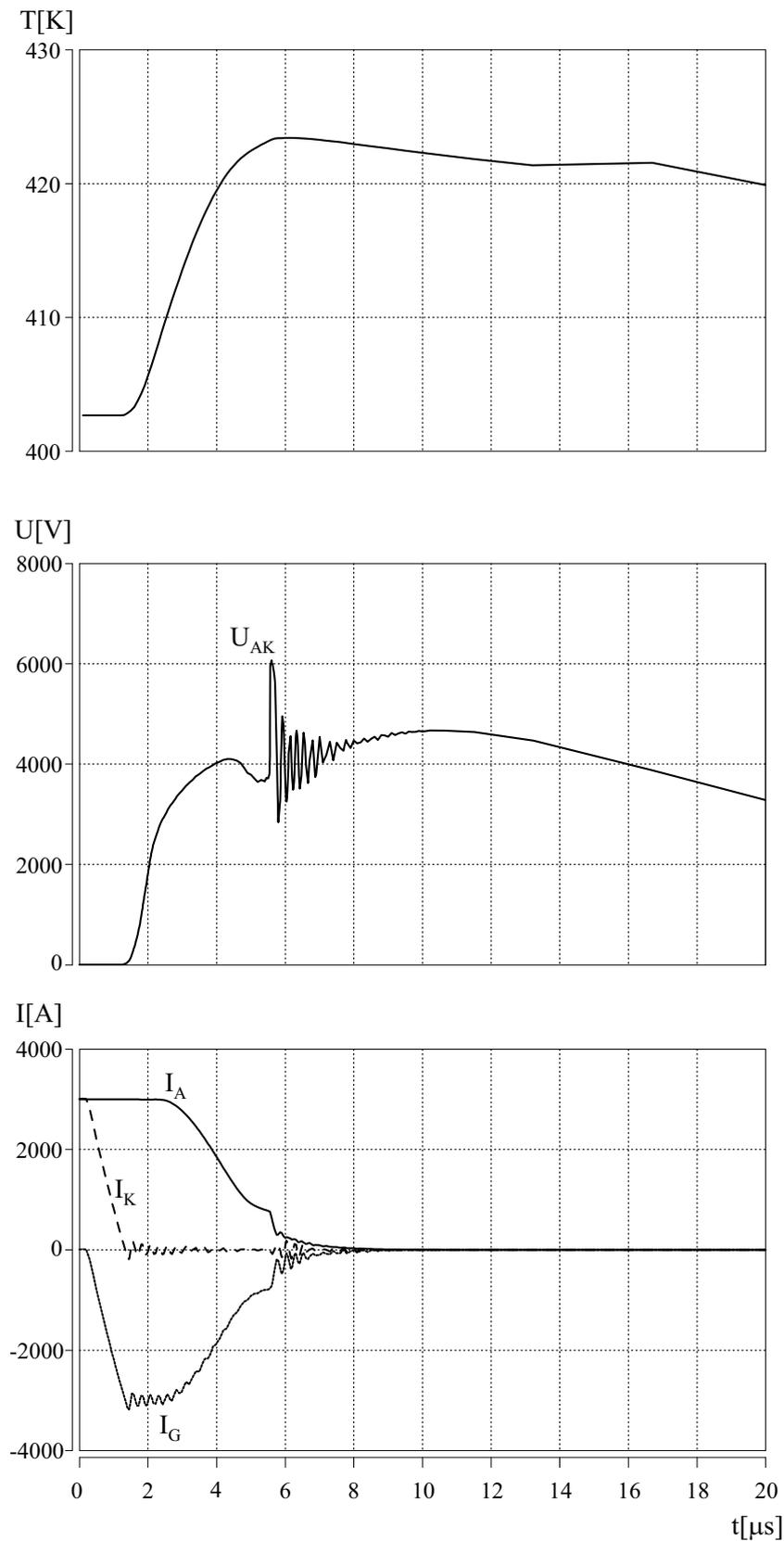


Bild 6.18: *DESSIS-Simulation: hartes Abschalten einer IGCT-Struktur; oben: Temperaturverlauf der heißesten Stelle, mitte: Spannungsaufnahme, unten: Ströme*

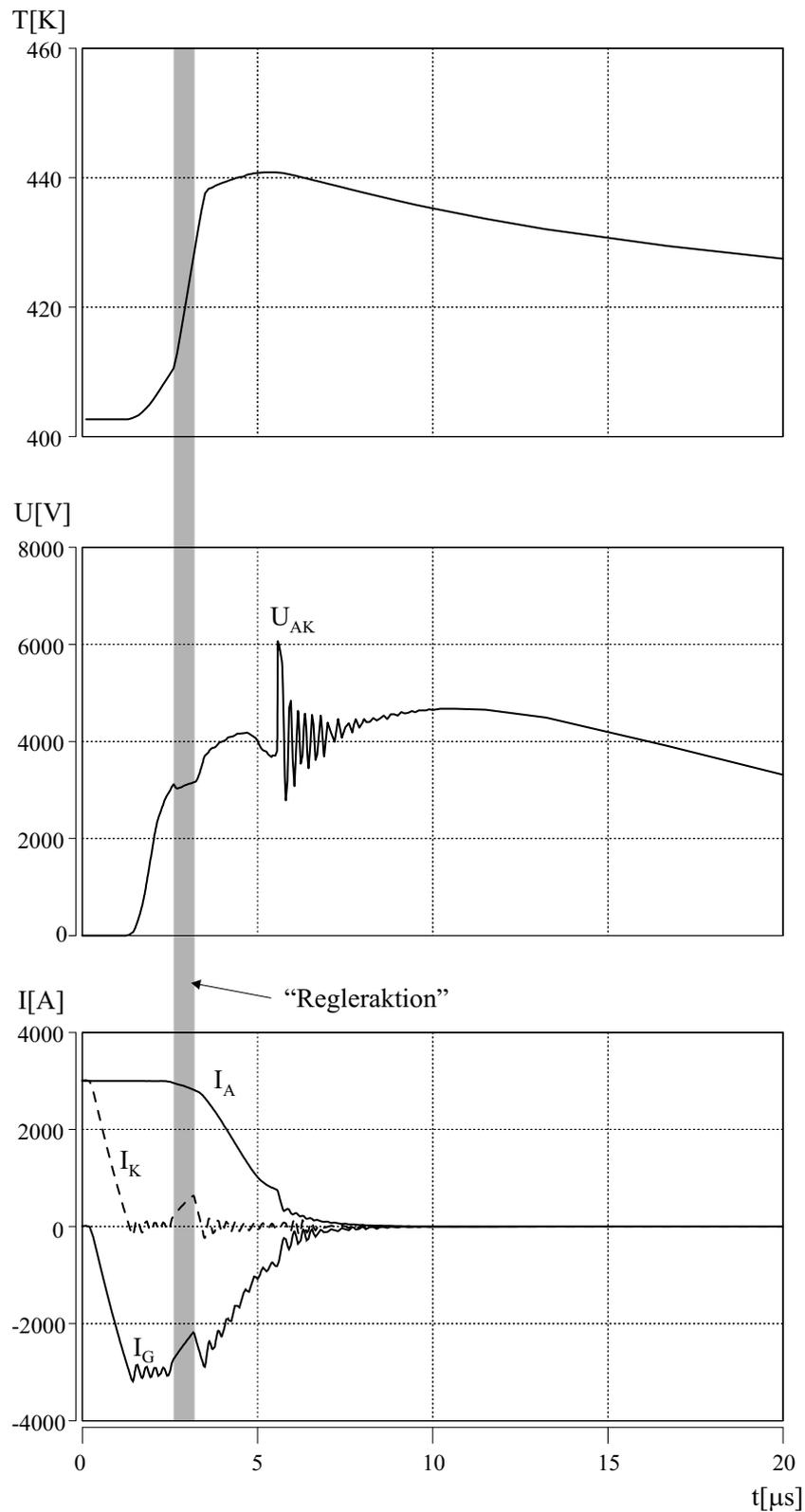


Bild 6.19: *DESSIS-Simulation: „geregeltes“ Abschalten einer IGCT-Struktur; oben: Temperaturverlauf der heißesten Stelle, mitte: Spannungsaufnahme, unten: Ströme*

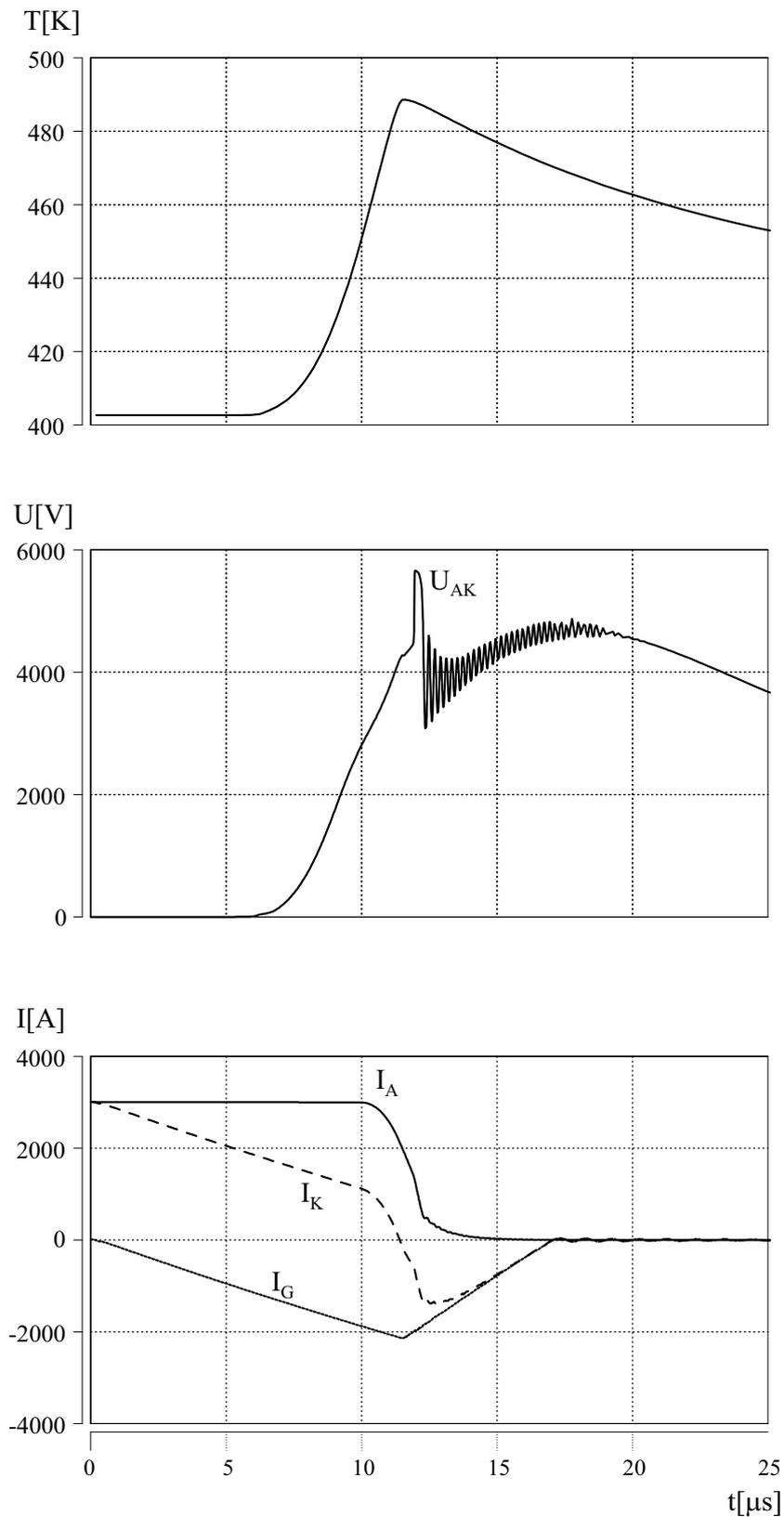


Bild 6.20: *DESSIS-Simulation: weiche Abschalten einer IGCT-Struktur; oben: Temperaturverlauf der heißesten Stelle, mitte: Spannungsaufnahme, unten: Ströme*

Beim harten Schalten ergibt sich erwartungsgemäß die kleinste Temperaturerhöhung auf ca. $420K$. Dagegen steigt die Temperatur für den geregelten Fall etwas stärker auf etwa $440K$ an. Jedoch wird die maximale Temperatur mit etwa $490K$ für das weiche Schalten noch viel größer. Dieser Wert dürfte sogar noch steigen, wenn man realistischere dI_G/dt für weiches Abschalten annimmt (s. Fußnote auf Seite 163). Im Bild 6.21 sind nochmal alle drei Temperaturverläufe zusammengestellt, wobei überall die Anfangstemperatur abgezogen ist. Zusätzlich zeigt eine der durchgezogenen Linien, die Differenz der Temperaturtransienten des hart abgeschalteten vom „geregelten“ Fall.

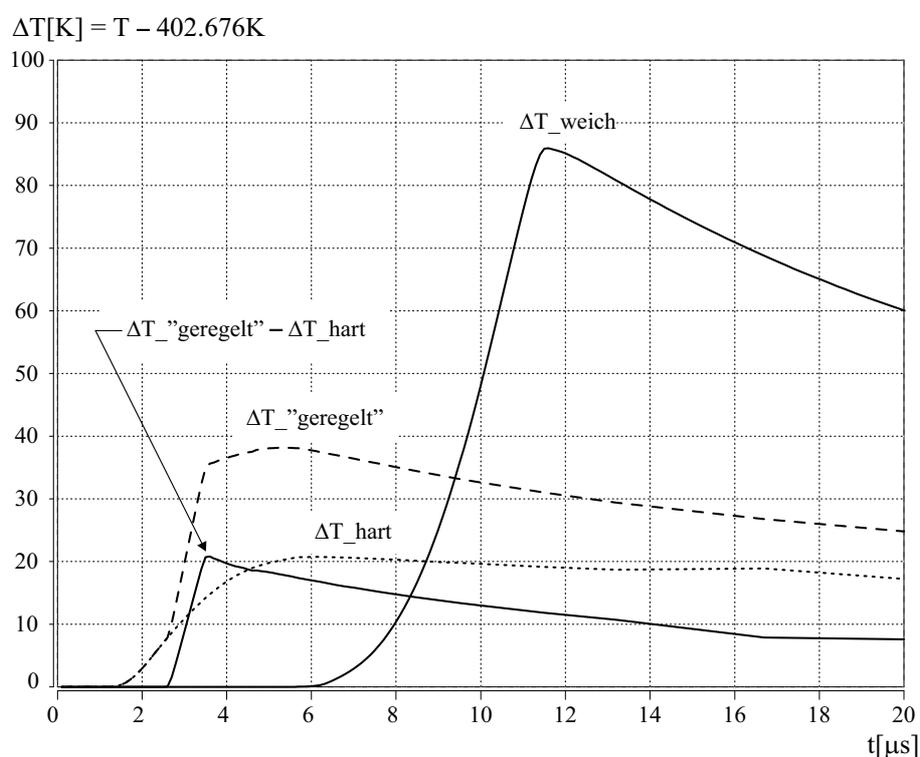
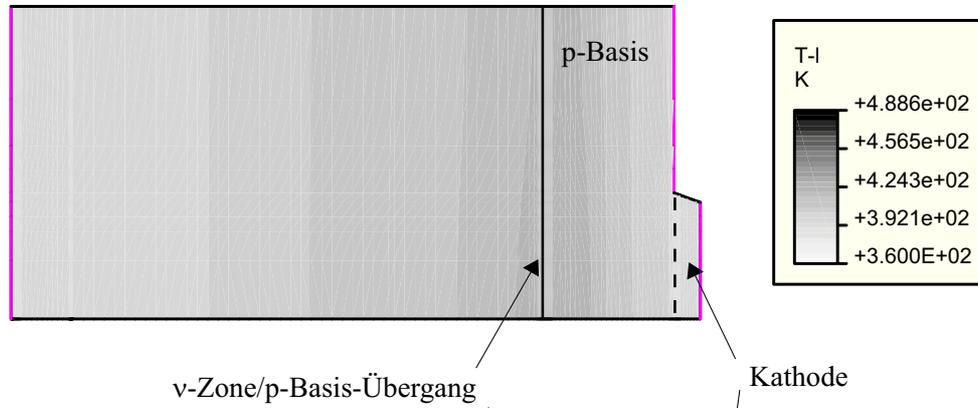


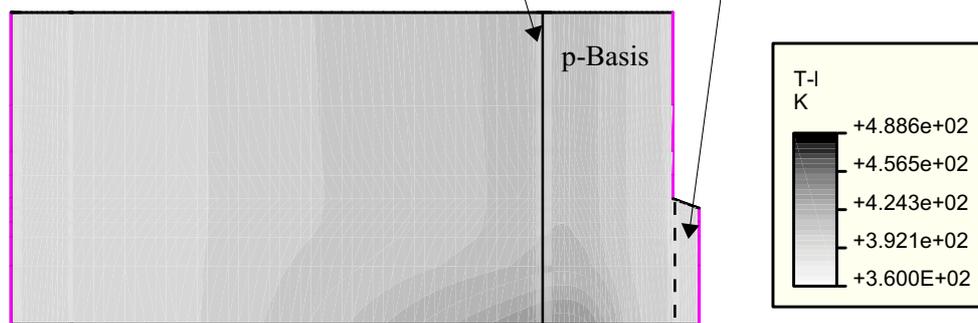
Bild 6.21: *DESSIS-Simulation: Vergleich der Temperaturtransienten*

Wie gesagt, können Maximaltemperaturen mit dieser Simulation sicherlich nicht exakt berechnet werden, zumal die Praxis zeigt, daß unter Umständen einzelne Zellen eines IGCTs/GTOs besonders anfällig sind, Filamente auszubilden. Festzuhalten bleibt, daß der Temperaturanstieg beim weichen Schalten deutlich größer ist als beim geregelten harten Schalten. Daher wird der Abschaltvorgang mit aktiver Ansteuerung nicht zur Beschädigung des IGCTs führen, wofür der abschließende Beweis allerdings nur mit experimenteller Validierung zu führen ist. Abschließend zeigt Bild 6.22 die örtliche Verteilung der Temperatur zu den obigen drei Simulationen.

a) hartes Schalten



b) hartes Schalten mit Gegenspannung



c) weiches (softes) Schalten

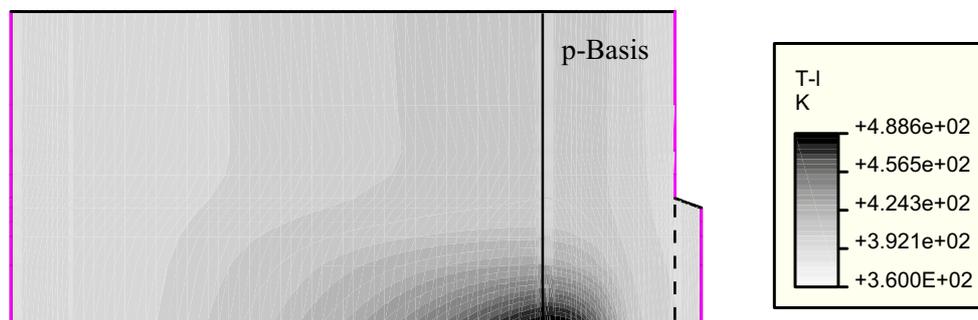


Bild 6.22: *DESSIS-Simulation: örtliche Temperaturverteilung zu den drei Simulationen*

Beim untersten Teilbild c) (weiches Schalten) sieht man deutlich die Filamentierung: An dem Punkt, der einerseits vom Gate lateral am weitesten entfernt ist, gibt es noch ein Gebiet mit Elektronen, das andererseits von der RLZ am Übergang ν -Zone/ p -Basis maximal zusammengedrückt wird, wodurch genau dort ein Hot Spot entsteht.

6.4.2 Aktive Ansteuerung und Redundanz

In Abschnitt 5.1 wurde auf den Begriff der Redundanz bereits kurz eingegangen. Verhält sich ein IGCT aufgrund einer Beschädigung wie ein Kurzschluß, wird aufgrund redundant vorhandener IGCTs die Funktionsfähigkeit der Serienschaltung sichergestellt, d.h. der Anteil der Blockierspannung jedes einzelnen der restlichen IGCTs bleibt kleiner gleich der maximal spezifizierten statischen Nennspannung.

Im Zusammenhang mit der vorgestellten Gateansteuerung ergibt sich nun ein Problem: Mit der Spannungserfassung in Bild 6.8 wird als Referenzspannung U_{ges}/n gemessen. Kann ein IGCT nicht mehr blockieren, benötigt man aber für die restlichen $n-1$ IGCTs $U_{ges}/n-1$ als Referenzwert. Somit muß die Schaltung zur Spannungsdetektion (Bild 6.8) erweitert werden, um eine Fehlfunktion beim Auftreten eines Ausfalls zu vermeiden. Bild 6.23 zeigt eine Schaltung, die ein Umschalten vom Modus n IGCTs (Normalfall) auf einen Modus $n-1$ IGCTs (Fehlerfall) ermöglicht.

Betrachtet wird wieder der GCT_i, dessen Spannungsabweichung ermittelt wird, wobei die Serienschaltung aus insgesamt n IGCTs besteht. Im Normalfall steht der Schalter S auf $s1$ und die Detektion funktioniert wie die Grundschaltung (Bild 6.8), d.h. die Summe aus R_3^* und R_4 entspricht genau R_3 in Bild 6.8. Fällt nun einer der oberen IGCTs (GCT₁ bis GCT_{i-1}) bzw. einer der unteren IGCTs (GCT_{i+1} bis GCT_n) aus, benötigt man ein Widerstandsverhältnis $R_{1,2} : R_3 = n - 1 : 1$ statt $R_{1,2} : R_3 = n : 1$, um richtig zu messen.

Hier wird alternativ weiterhin der Strom I_3 mit dem „alten“ Widerstandsverhältnis $n : 1$ gemessen, dieser aber dann so aufgeteilt, daß nur der „richtige“ Strom $I_{3S2} = I_3 \cdot n-1/n$ den „Stromsubtrahierer“ erreicht, während der Strom $I_6 = I_3/n$ direkt daran vorbeigeleitet wird und der Kathode vom GCT_i zufließt (Stromteiler). Dies wird erreicht, indem der Schalter S im Fehlerfall nach $s2$ umgeschaltet wird. Dabei entspricht der Gesamtwiderstand $R_3^* + (R_5 || R_6)$ genau dem Widerstandswert $R_3 = R_1/n$ aus Bild 6.8. Weiterhin muß für das Widerstandsverhältnis beim Stromteiler $R_5 : R_6 = n - 1 : 1$ gelten.

Es ist leicht einzusehen und kann auch durch Simulationen, deren Ergebnisse hier nicht abgebildet sind, bestätigt werden, daß dieses Prinzip funktioniert. Allerdings kann nur der Ausfall eines IGCTs verarbeitet werden, . Liegt ein Ausfall eines IGCTs vor, sollte dieser schnellstmöglich durch einen neuen IGCT ersetzt werden.

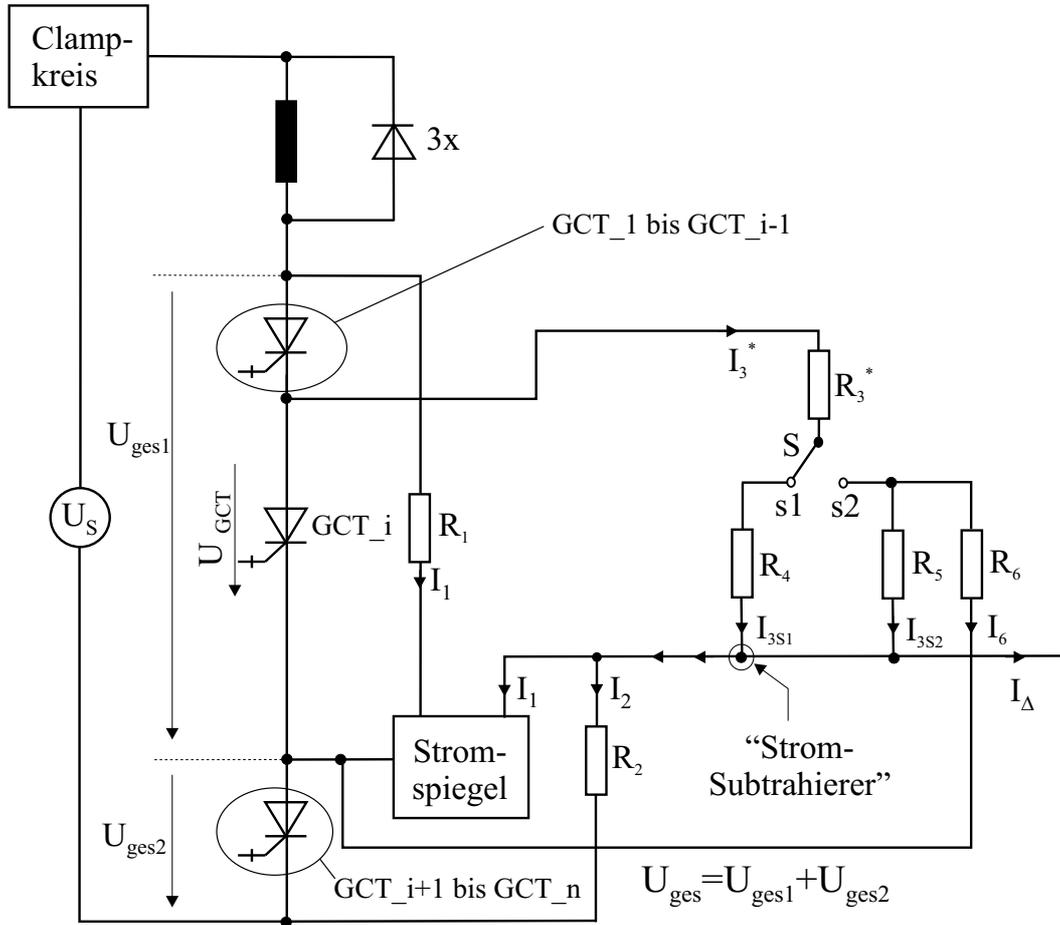


Bild 6.23: Detektion der Spannungsabweichung eines GCT_i bei Berücksichtigung von Redundanz

Die technische Umsetzung des Schalters kann mittels MOSFETs erfolgen. Die Schwierigkeit besteht nun darin, den Ausfall eines IGCTs zu detektieren und den richtigen Umschaltzeitpunkt des Schalters S zu bestimmen. Überlegungen dahingehend wurden u.a. aus Zeitgründen nicht weiterverfolgt.

6.4.3 Aktive Ansteuerung und Pufferschicht

Bei den Ergebnissen in Abschnitt 6.3 (Bild 6.14) fällt auf, daß der statische Spannungsanteil eines IGCT mit $2kV$ niedriger ist als mit einem $4,5kV$ -IGCT erlaubt ist ($U_{dc,max} = 2,8kV$). Das ist mit Blick auf die oben genannte Redundanz durchaus realistisch. Fällt nun ein IGCT aus, müßte jeder der beiden restlichen IGCTs $3kV$ blockieren. Schon für einen Spannungsanteil ab etwa $U_{GCT} = 2,5kV$ kommt

es aber zu Problemen, die von der Pufferschicht (s. dazu Kapitel 3) verursacht werden. Bekanntlich findet während des Abschaltvorgangs ein Übergang von der dreiecksförmigen zur trapezförmigen Felddaufnahme am Puffer statt, was in einer zweiten Spannungsspitze resultiert (siehe beispielsweise Bild 4.3). Trotz der Regelung wird nicht verhindert, daß einer der IGCTs mit diesem Übergang und der besagten Spannungsspitze etwas früher dran ist als die restlichen Elemente. Folglich wird zu diesem Zeitpunkt eine starke und vor allem schnell wachsende Spannungsabweichung an den restlichen IGCTs detektiert, wie dies Bild 6.24 für den idealen P-Regler, der diese Situation noch gut ausregeln kann, zeigt. Zwischen $t \approx 4\mu\text{s}$ und $t \approx 8\mu\text{s}$ findet der eigentliche Abschaltvorgang mit relativ kleinen ΔU bzw. $d(\Delta U)/dt$ statt, während für $t \approx 9\mu\text{s}$ (RLZ erreicht den Puffer) diese beiden Größen auffallend groß werden. Daß nachfolgend ΔU ebenso recht heftig negativ wird hängt mit einer Überreaktion des Reglers, d.h. zu starkes Bremsen, zusammen.

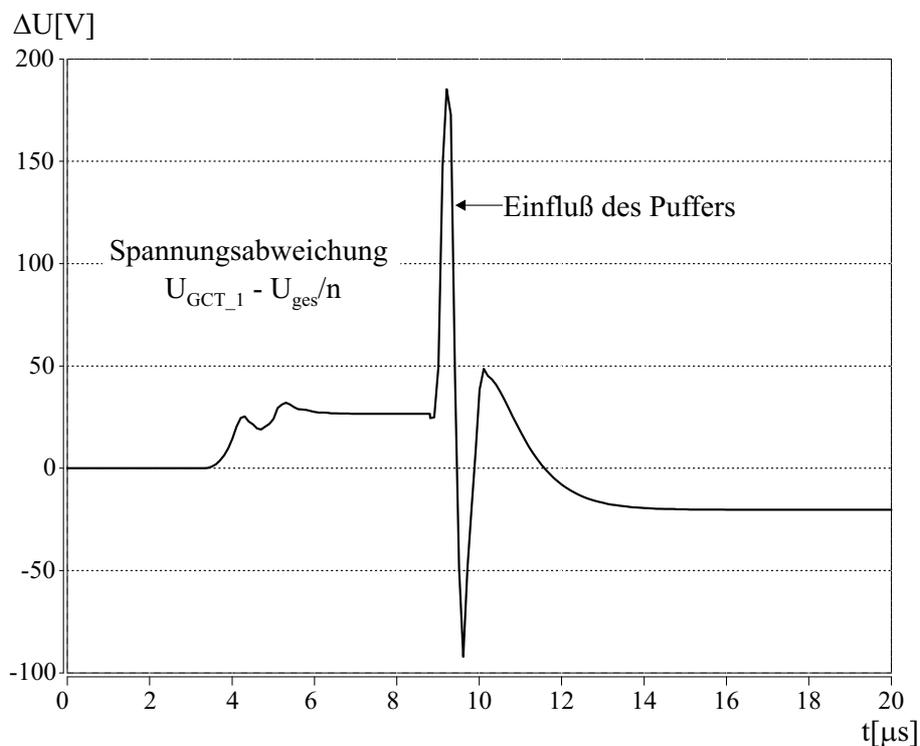


Bild 6.24: *SABER-Simulation: Verlauf der Spannungsabweichung*

Bei der vorgestellten Schaltung mit einer Grenzfrequenz von etwa 3MHz funktioniert das Ausregeln dieser Spannungsabweichungen nicht mehr. Zu diesem Zeitpunkt ist der Gatestrom I_G schon sehr klein und folglich die Eingriffsmöglichkeiten begrenzt. Wählt man K_P groß, kann es zu einem Wiedereinschalten einzelner IGCTs kommen. Für kleinere K_P zeigen die Simulationen ein stark schwingendes Verhal-

ten, was ebenfalls inakzeptabel ist.

Insgesamt zeigt sich, daß die Ansteuerung eine Grenzfrequenz von mehr als 100MHz haben müßte, um auch für $U = 2,8\text{kV}$ je IGCT, bei einem Laststrom von $I = 3\text{kA}$, symmetrisch abschalten zu können. Bauelemente, mit denen man derart hohe Grenzfrequenzen bei gleichzeitiger Einhaltung aller anderen Anforderungen verwirklichen könnte, sind nicht auf dem Markt erhältlich. Somit bleiben zwei Möglichkeiten, um dieses Problem zu umgehen. Zum einen kann man die Serienschaltung so dimensionieren, daß nach Ausfall eines IGCTs der statische Spannungsanteil der verbleibenden Elemente unter $2,5\text{kV}$ gehalten wird. Eine andere Möglichkeit ist die Verwendung eines RC-Snubbers, dessen Kapazität erheblich kleiner sein darf als dies bei Verzicht auf die aktive Ansteuerung der Fall ist. Bild 6.25 zeigt die simulierten Spannungsverläufe für das oben bereits öfters zitierte „Worst Case Szenario“, wobei jeder IGCT einen Spannungsanteil von 3kV blockiert.

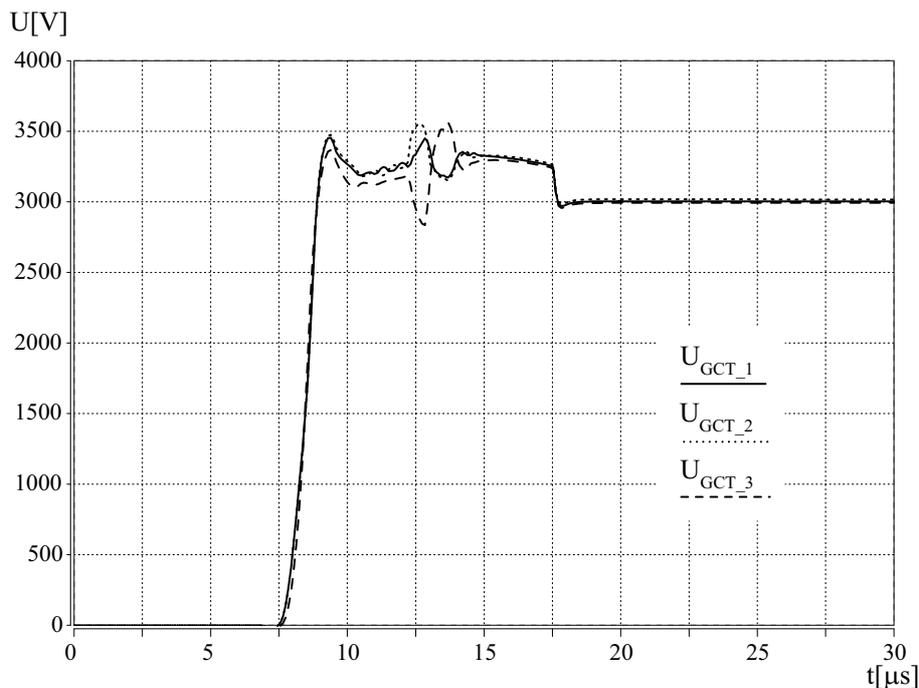


Bild 6.25: SABER-Simulation: Spannungsverläufe beim harten unentlasteten Abschalten von 3kA gegen 9kV (!) mit aktiver Ansteuerung; „Worst Case Szenario“, RC-Snubber mit $C_S = 30\text{nF}$

Der Betrag der Snubberkapazität kann immerhin von $3\mu\text{F}$ bei konventioneller Gatansteuerung (vgl. Abschnitt 5.2.2) auf 30nF , also um den Faktor 100 gesenkt werden, wobei dann die verbleibende statische Spannungsdifferenz praktisch null ist (vgl Bild 6.25 mit Bild 5.13).

6.5 Diskussion und Bewertung

Die vorgestellte Schaltung zur aktiven Ansteuerung von seriengeschalteten IGCTs ist unter bestimmten Randbedingungen sehr gut geeignet, die Asymmetrie in der Spannungsaufteilung beim Abschalten wirksam zu verhindern. Die erwähnten Randbedingungen hängen zunächst vom eingesetzten IGCT-Typ ab. Bei symmetrischen IGCTs (s. z.B. [75]), d.h. rückwärtsperrfähige Elemente ohne Pufferschicht, sind keine Einschränkungen hinsichtlich des maximalen Spannungsanteils eines IGCTs zu erwarten. Bei den weitaus mehr verbreiteten asymmetrischen IGCTs mit Pufferschicht ist dagegen der maximale Spannungsanteil eines IGCTs kleiner als die spezifizierte statische Nennspannung (vgl. Abschnitt 6.4.3).

In der Regel werden bei Anwendung der Serienschaltung, zusätzliche redundante Elemente eingesetzt, um die Funktionsfähigkeit der leistungselektronischen Schaltung auch bei Ausfall eines IGCTs zu gewährleisten. Damit sind die Spannungsanteile je IGCT ohnehin kleiner als die Nennspannung. Ein Beispiel aus der Praxis, soll dies verdeutlichen [66]: In einer Schaltung, die zur Kopplung zweier Netze unterschiedlicher Frequenz dient, besteht in ihren kleinsten Einheiten aus Phasenmodulen mit 4+2 hartgeschalteten 4,5V/3,5kA-GTOs in Serie, d.h. zwei GTOs sind redundant. Die Symmetrierung der Blockierspannungen geschieht mit Hilfe von RC-Gliedern. Die Zwischenkreisspannung beträgt dabei 10kV. Verwendet man statt den GTOs die asymmetrischen 4,5kV/3,5kV-IGCTs, muß jeder IGCT mit einer statischen Nennspannung in Höhe von 2,8kV, nur 1,67kV blockieren. Fällt ein IGCT aus, verbleiben 2kV für jeden IGCT. Dies wäre für die aktive Ansteuerung kein Problem. Fällt der zweite redundante IGCT aus, blockieren die restlichen vier IGCTs jeweils 2,5kV und die aktive Ansteuerung bekommt die oben genannten Schwierigkeiten. Somit wäre es bei Applikation der aktiven Ansteuerung sinnvoll, einen ausgefallenen IGCTs sofort auszuwechseln.

Insgesamt ist die praktische Umsetzung des hier vorgestellten und unter realistischen Vorgaben per Simulation untersuchten Ansteuerkonzepts, relativ aufwendig und erfordert entsprechendes Know-how. Trotzdem sieht der Autor einiges Potential für dieses Konzept, auch wenn letztlich die Kosten entscheidend sind, ob es zur Anwendung kommt oder nicht.

7 Zusammenfassung

7.1 Inhalt und Ergebnisse der Arbeit

Im Mittelpunkt der Arbeit steht die Modellierung des Integrated Gate Commutated Thyristor (IGCT) für die Schaltungssimulation. Beim IGCT handelt es sich um einen modifizierten GTO, der hart und unenlastet abgeschaltet werden kann und damit im Bereich hoher Leistung dem IGBT mindestens ebenbürtig ist. Als Maßstab für die Komplexität der Modellbeschreibung dient im allgemeinen die Ergebniserwartung, die man sich auf Basis von Simulationen mit dem jeweiligen Modell erhofft. Möchte man lediglich wissen, ob eine leistungselektronische Schaltung funktioniert, oder ein Steuerungsverfahren (z.B. PWM) testen, genügen bereits ideale Schalter. Je genauer man dagegen transiente Vorgänge, die auf Ladungsträgervorgänge beruhen, erfassen will, desto genauer müssen die physikalischen bauelementeinternen Vorgänge modelliert werden. Gerade bipolare und hybride Leistungsbauelemente, wie die Leistungsdiode, der GTO, der IGCT und der IGBT, besitzen eine, im Vergleich zu allen anderen Zonen, sehr lange schwach dotierte Driftzone, die beim Abschalten von Überschlußladungsträgern leergeräumt und beim Einschalten mit Überschlußladungsträgern gefüllt wird, was in beiden Fällen zu dynamischen Verlusten führt. Dabei ist die Beschreibung des Ladungsträgerverhaltens in dieser Driftzone keineswegs trivial.

Sogenannte Verhaltensmodelle mit meist vereinfachten empirischen Gleichungen, anstelle der komplizierten physikalischen Beschreibung, sind dabei ein erster Schritt. Jedoch gelangt man mit diesen Modellen sehr schnell an unüberwindliche Grenzen, was einerseits die Rechengenauigkeit betrifft und andererseits sich auf die Simulation bestimmter physikalischer Phänomene überhaupt bezieht. Oft läßt sich gerade durch die richtige Auswahl von Bauelementen die Performance einer bekannten leistungselektronischen Schaltungsform optimieren, wobei die genaue Kenntnis der Schalttransienten dabei eine ganz wesentliche Rolle spielt. Mit einem Modell, das dies leistet, können kritische Schaltzustände, Worst Case-Situationen, Bauelementegrenzen, unerwartete Oszillationen, und vieles andere mehr, schon am Rechner prognostiziert und analysiert werden.

Für diese Art von Fragestellungen benötigt man ein Modell, das die wesentlichen

halbleiterphysikalischen Gleichungen, besonders der Driftzone, enthält. Dieser Weg wird in der vorliegenden Arbeit beschritten, wobei zur Aufwandsbegrenzung ein eindimensionaler mathematischer Ansatz gewählt wird. Konkret wird die Kontinuitätsgleichung (Diffusionsgleichung), die bei Berücksichtigung der Elektronen-Löcher-Streuung nach Mnatsakanov (EHS), eine nichtlineare partielle Differentialgleichung darstellt, für den Fall der Hochinjektion numerisch gelöst. Die besagte EHS-Theorie hat sich vor allem bei tiefen Temperaturen bewährt. Alle anderen Gleichungen, wie beispielsweise zur Beschreibung der Randzonen, sind kompakt als analytische Ausdrücke beschrieben. Als Plattform für diese Form der Modellierung eignet sich SABER mit seiner analogen Modellierungssprache MAST besonders gut, da Gleichungen direkt eingegeben werden können. Alleine der Numerikteil mit der notwendigen Ortsdiskretisierung, wird in eine C-Subroutine ausgelagert. Damit wird klar, daß die große Genauigkeit mit etwas langsameren Rechenläufen erkauft wird, wobei die Rechenzeit im Vergleich zur Bauelementesimulation immer noch um Welten kürzer ist.

Der Vergleich mit Meßergebnissen zeigt, daß sich das vorliegende IGCT-Modell hervorragend an das reale Verhalten adaptieren läßt und zwar über einen großen Strombereich (0,5–3kA), ohne daß die Parameter des Modells geändert werden müssen. Weiterhin stimmen auch das Einschaltverhalten und die Kennlinie mit der Realität gut überein. Es liegt also ein IGCT-Modell vor, das sehr gut zur Beantwortung der im vorhergehenden Absatz genannten Fragen, eingesetzt werden kann. Den Grad an Genauigkeit, den man im Vergleich mit den Verhaltensmodellen gewinnt, kompensiert die etwas längere Rechenzeit bei weitem.

Das transiente Auseinanderlaufen von Blockierspannungen beim Abschalten zweier oder mehrerer seriengeschalteter IGCTs ohne Entlastung, ist ein Beispiel für das Auftreten physikalischer Phänomene, die mit diesem Modell analysiert werden können. Zurückzuführen ist diese Asymmetrie auf zum Zeil schon sehr kleine Abweichungen in den Daten der einzelnen IGCTs oder den zugehörigen Gatekreisen. Simulationen mit dem Modell liefern eine Idee davon, welche Parameterschwankungen besonderen Einfluß auf diese Spannungsunterschiede haben. Weiterhin kann festgestellt werden, daß im Vergleich zur dU/dt -Begrenzung bei GTOs, relativ kleine Kapazitäten für ein paralleles RC-Glied genügen, um Symmetrie sicherzustellen. Somit kann das Modell genutzt werden, die passive Beschaltung zu optimieren und damit die Kosten zu senken.

Des weiteren kann man das Modell zu Synthesezwecken einsetzen. Da das Modell auch den Einfluß der Gateansteuerung auf den IGCT richtig wiedergibt, kann auch die Gateansteuerung optimiert werden. In dieser Arbeit steht dabei wieder die Reihenschaltung von IGCTs im Vordergrund, die immer dann zum Einsatz kommt, wenn die Zwischenkreisspannung höher ist, als für einen einzelnen IGCT vorgesehen ist. Genauer wird der Frage nachgegangen, ob es möglich ist, mit einer aktiven

Gateansteuerung und unter Verzicht paralleler RC-Glieder, die Blockierspannung zu symmetrieren. Vorbild ist hierbei die aktive Ansteuerung beim IGBT, die längst Stand der Technik ist. Beim IGCT, bei dem im Gegensatz zum IGBT hohe Gateströme auftreten, fällt eine solche Ansteuerung damit zwangsläufig komplexer aus. Die hier entwickelte, mit diskreten Bauteilen virtuell aufgebaute Schaltung, die auf einer proportionalen Regelung basiert, wird mit adäquaten Modellen und realistischen Parametern für alle Komponenten, per Simulation untersucht.

Die erhaltenen Simulationsergebnisse lassen darauf schließen, daß die aktive Ansteuerung eine vielversprechende Alternative zu RC-Gliedern darstellt. Allerdings gilt diese Aussage bei den gängigen IGCTs mit Pufferschicht nur, wenn die maximale mögliche Spannung nicht ausgereizt wird. Gibt es beim Abschalten eine ausgeprägte zweite Spannungsspitze an der Pufferschicht, versagt die Schaltung. Abhilfe schafft dann ein RC-Glied mit einer nochmals stark verkleinerten Kapazität.

7.2 Ausblick

Im letzten Kapitel dieser Arbeit wird beschrieben, daß sich die Blockierspannung seriengeschalteter IGCTs beim harten unentlasteten Abschalten, mittels aktiver Ansteuerung symmetrieren läßt. Damit ist der Aufbau und der anschließende Test dieser Schaltung der nächste logische Schritt. Vorher sind aber noch einige Vorarbeiten zu leisten. Einerseits kann die Einstellung und Adaption des Proportionalitätsfaktors K_P sicher noch optimiert werden, indem Untersuchungen über eine optimale Strategie besonders auch im Hinblick auf die Pufferschichtproblematik, durchgeführt werden. Andererseits fehlt auch noch eine Auswerteelektronik, die den Ausfall eines defekten IGCTs erkennt und im richtigen Moment auf den Modus „ein IGCT weniger“ schaltet. Letztlich erfordert der Aufbau der vorgestellten Schaltung, nicht zuletzt aufgrund der hohen Gateströme und der geforderten niedrigen Gateinduktivität als Voraussetzung für hartes Schalten, viel Know-how und Erfahrung. Eine Zusammenarbeit mit entsprechenden Partnern, z.B. aus der Industrie, wäre dabei sicherlich vorteilhaft, wenn nicht sogar Voraussetzung.

Literaturverzeichnis

- [1] ABB SEMICONDUCTORS AG: *The GTO/GCT-Data Book* 1997
- [2] ANALOGY, INC.: *The SaberGuide and SaberScope Manual, Release 4.0*. Registered in England No. 2372734, Registered Office: Aldwych House, Aldwych, London WC2B 4JP, 1995.
- [3] ALONSO, C., MEYNARD, T., FOCH, C., BATARD, C. und PIQUET, H.: *A Model of GTO Compatible with Power Circuit Simulation*. In: Proc. of 5th EPE, vol. 2, 232–437, Brighton, Großbritannien, 1997.
- [4] ARORA, NARAIN D., JOHN R. HAUSER und DAVID J. ROULSTON: *Electron and hole mobilities as a function of concentration and temperature*. IEEE Transactions on Electron Devices, 29(2):292–95, Febr. 1982.
- [5] ASKEROV, B.M.: *Electron Transport Phenomena in Semiconductors*. World Scientific Publishing, Singapore, New Jersey, London, Hongkong, 1. Auflage, 1994.
- [6] BERNET, S., TEICHMANN, R., ZUCKERBERGER, A. und STEIMER, P.: *Comparison of High Power IGBTs and Hard Driven GTOs for High Power Inverters*. In: Proc. of APEC, Anaheim, USA, 1998.
- [7] BERNET, S., *Recent Developments of High Power Converters for Industry and traction Applications*. In: IEEE Transactions on Power Electronics, vol. 15, no. 6, 1102–1117, November 2000.
- [8] BIELA, J.: *Erstellung, Untersuchung und Implementierung von Konzepten zur aktiven Ansteuerung von seriengeschalteten IGCT-Bauelemente*. Diplomarbeit, Universität Erlangen–Nürnberg, angefertigt an der Technische Universität München, Lst. für Elektrische Antriebssysteme, 2000.
- [9] BLUDAU, W., A. ONTON und W. HEINKE: *Temperature dependence of the band gap in silicon*. J. Appl. Phys, 45(4):1846–48, April 1974.
- [10] BRONSTEIN, I.N. und SEMENDJAJEW, K.A.: *Taschenbuch der Mathematik*. Harri Deutsch, Thun und Frankfurt, 19. Auflage, 1981.

-
- [11] BROOKS, H.: *Scattering by Ionized Impurities in Semiconductors*. Physical Review 83, Seiten 879, 1951.
- [12] BURDEN, R.L., FAIRES, J.D. und RYNOLDS, A.C.: *Numerical Analysis*. Prindle, Weber & Schmidt, Boston, 2. Auflage, 1981.
- [13] CANALI, C., MAJNI, G., MINDER, R. und OTTAVIANI, G.: *Electron and Hole Drift Velocity Measurements in Silicon and Their Empirical Relation to Electric Field and Temperature*. IEEE Trans. on Electron Devices ED-22, Seiten 1045–47, 1975.
- [14] CAUGHEY, D. und THOMAS, R.: *Carrier mobilities in silicon empirically related to doping and field*. Phys Rev., vol. 109, no. 5, Seiten 1537–40, 1958.
- [15] CHYNOWETH, A.: *Ionization rates for electrons and holes in Silicon*. IEEE Proc., Seiten 2192–93, 1967
- [16] CONWELL, E. und WEISSKOPF, V.F.: *Theory of Impurity Scattering in Semiconductors*. Physical Review 77, no. 3, Seiten 388–390, 1950.
- [17] DANNHAUSER, F.: *Die Abhängigkeit der Trägerbeweglichkeit in Silizium von der Konzentration der freien Ladungsträger-I*. Solid-State Electron., 15:1371, September 1972.
- [18] DEBOY, GERALD: *Charakterisierung von Leistungshalbleiterbauelementen durch Interne Laserdeflektion*. Doktorarbeit, TU München, 1995.
- [19] DEBYE, P.P. und CONWELL, E.M.: *Electrical Properties of N-Type Germanium*. Physical Review 93, Seiten 693–706, 1950.
- [20] DORKEL, J.M. und LETURCQ, PH.: *Carrier mobilities in silicon semi-empirically related to temperature, doping and injection level*. Solid-State Electron., 24(9), Seiten 821–25, 1981.
- [21] EICHER, S.: *The Transparent Anode GTO (TGTO): A New Low-Loss Power Switch*. Doktorarbeit, ETH Zürich, 1996.
- [22] EICHER S., BERNET S., STEIMER P. und WEBER, A.: *The 10kV IGCT – A New Device for Medium Voltage Drives*. In: Proc. of IAS, Rom, Italien, 2000.
- [23] GERLACH, W.: *Thyristoren*. Springer-Verlag, Berlin, 1979.
- [24] GERSTER, C.: *Fast High-Power/High-voltage Switch Using Series-connected IGBTs with Active Gate-controlled Voltage-balancing*. In: Proc. of APEC, Orlando, USA, 1994.

- [25] GOEBEL, H. und HOFFMANN, K.: *Full Dynamic Power Diode Model Including Temperature Behavior for Use in Circuit Simulators*. Proc. of ISPSD, Tokio, Japan, 1992.
- [26] GREEN, M.A.: *Intrinsic concentration, effective densities of states, and effective mass in silicon*. J. Appl. Phys., 67(6):2944–2954, March 1990.
- [27] GRESSEROV, B.N. und T.T. MNATSAKANOV: *Estimate of the role of the electron–hole scattering in the transport of carriers in multilayer gallium arsenide structures*. Sov. Phys. Semicond., 24(9):1042–43, September 1990.
- [28] GRÜNING, H., ØDEGÅRD, B. und REES, B.: *High-Power-Driven GTO Module for 4.5kV/3kA Snubberless Operation*. In: Proc. of PCIM, Nürnberg, 1996.
- [29] HÖPFEL, R.A., J. SHAH, P.A. WOLFF und C.G. GOSSARD: *Negative absolute mobility of electrons GaAs quantum wells*. Phys. Rev. Lett., 56(25):2736–39, Juni 1986.
- [30] ISE, INTEGRATED SYSTEMS ENGINEERING AG: *Dessis Reference Manual, Version 1.3*. ISE, ETH–Zentrum, Gloriastraße 35, CH–8092 Zürich, Switzerland, 1994.
- [31] KITTEL, CH.: *Einführung in die Festkörperphysik*. Oldenbourg–Verlag, Berlin, 9. Auflage, 1991.
- [32] KLAKA, S., FRECKER, M. und GRÜNING, H.: *The Integrated Gate-Commutated Thyristor: A New High-Efficiency, High-Power Switch for Series or Snubberless Operation*. In: Proc. of PCIM, Nürnberg, 1997.
- [33] KRAFFT, E., STEIMEL, A. und STEINKE, J.: *Three-Level High-Power Inverters with IGCT and IGBT Elements Compared on the Basis of Measurements of the Device Losses*. In: Proc. of EPE, Lausanne, Schweiz, 1999.
- [34] KRAUSSE, J.: *Die Abhängigkeit der Trägerbeweglichkeit in Silizium von der Konzentration der freien Ladungsträger–II*. Solid–State Electron., 15:1377, September 1972.
- [35] KUHN, H.: *Modellierung von lateralen Effekten am Beispiel der SPEED-Struktur*. Diplomarbeit, Technische Universität München, Lst. für Elektrische Antriebssysteme, 1996.
- [36] KUHN, H. und SCHRÖDER, D.: *Physikalische Modellierung von GTO–Thyristoren mit Pufferschicht für hartschaltende Anwendungen*. Beitrag zum 26. Kolloquium Halbleiter–Leistungsbaulemente und Materialgüte von Silizium, Freiburg im Breisgau, 1997

- [37] KUHN H., und SCHRÖDER, D.: *Circuit Simulation of Hard-Driven IGCT for Snubberless Operation using a Physically Based Model*. In: Proc. of IPEC, Tokio, Japan, 2000.
- [38] KUHN H., und SCHRÖDER, D.: *A New Validated Physically Based IGCT Model for Circuit Simulation of Snubberless and Series Operation*. In: Proc. of IAS, Rom, Italien, 2000.
- [39] KUHN H., und SCHRÖDER, D.: *Investigation of an Active Gate Drive Concept for Series Connected GCTs*. In: Angenommen für die Proc. of PCC, Osaka, Japan, April 2002.
- [40] LANG, J.E., F.L. MADARASZ und P.M. HEMEGER: *Temperature dependent density of states effective mass in non-parabolic p-type silicon*. J. Appl. Phys., 54(6):3612, 1983.
- [41] MA, C.L., LAURITZEN, P.O. und SIGG, J.: *A Physics-based GTO Model For Circuit Simulation*. In: Proc. of PESC, Atlanta, USA, 1995.
- [42] MA, C.L., LAURITZEN, P.O. und SIGG, J.: *Modelling of Power Diodes with Lumped-Charge Modeling Technique*. In: IEEE Transactions on Power Electronics, vol. 12, no. 3, 398–405, May 1997.
- [43] MACFARLANE, G.G., T.P. MCCLEAN, J.E. QUARRINGTON und V. ROBERTS: *Fine structure in the absorption-edge spectrum of silicon*. Phys.Rev., 111(5):1245–54, September 1958.
- [44] METZNER, D., T. VOGLER und D. SCHRÖDER: *A modular concept for the circuit simulation of bipolar power semiconductors*. IEEE Trans. on Power Electronics, 9:506–13, 1994.
- [45] METZNER, DIETER: *Netzwerkmodelle abschaltbarer Leistungshalbleiter-Bauelemente*. Doktorarbeit, TU München, 1994.
- [46] MNATSAKANOV, T.T., I.L. ROSTOVTSEV und N.I PHILATOV: *Investigation of the effect of nonlinear physical phenomena on charge carrier transport in semiconductor devices*. Solid-State Electron., 30(6):579–585, 1987.
- [47] MNATSAKANOV, T.T.: *Transport coefficients and Einstein relation in a high density plasma of solids*. Phys. Status Solidi B, 143:225–33, 1987.
- [48] MNATSAKANOV, T.T., SCHRÖDER, D. und A.E. SCHLÖGL: *Effect of high injection level phenomena on the feasibility of diffusive approximation in semiconductor device modeling*. Solid-State Electron., 42(1):153–63, 1998.

- [49] MÜLLER, R.: *Grundlagen der Halbleiter-Elektronik*. Springer-Verlag, Berlin, 4. Auflage, 1983.
- [50] NAGEL A., BERNET S., STEIMER P. und APeldoorn, O.: *A 24MVA Inverter using IGCT Series Connection for Medium Voltage Application*. In: Proc. of IAS, Chicago, USA, 2001.
- [51] OGURA, T., KITAGAWA M., NAKAGAWA, A. und OHASHI, H.: *6000-V Gate Turn-Off Thyristors GTO's with n-Buffer and New Anode Short Structure*. IEEE Transactions on Electron Device, vol. 38, no. 6, 1491–96, June 1991.
- [52] OVERSTRAETEN R. und MAN, D.: *Measurement on the ionization rates in diffused Silicon p-n junctions*. Solid-State Electronics, vol. 13, Seiten 583–608, 1970
- [53] PALMER, P. und GITHIARI, A.: *The Series Connection of IGBTs with optimised Voltage Sharing in the Switching Transient*. Proc. of PESC, Seiten 44–50. Atlanta, USA, 1995.
- [54] PIRES, R.G., DICKSTEIN, R.M., TITCOMB, S.L. und R.L. ANDERSON: *Carrier freezeout in silicon*. Cryogenics, 30:1064–68, December 1990.
- [55] REISCH, M.: *Elektronische Bauelemente*. Springer-Verlag, Berlin, 1. Auflage, 1997.
- [56] SCHLANGENOTTO, H., SERAFIN, J., SAWITZKI, F. und MAEDER, H.: *Improved Recovery of Fast Power Diodes with Self-Adjusting p Emitter Efficiency*. In: IEEE Electron Device Letters, vol. 10, no. 7, July 1989.
- [57] SCHLÖGL, A.E., MNATSAKANOV, T.T., KUHN, H. und SCHRÖDER, D.: *Temperature dependent behaviour of silicon power semiconductors — a new physical model validated by device-internal probing between 400K and 100K*. In: Proc. of PESC, Seiten 1720–25. Kyoto, Japan, 1998.
- [58] SCHLÖGL, A.E.: *Theorie und Validierung der Modellbildung bipolarer Leistungshalbleiter im Temperaturbereich von 100K bis 400K*. Doktorarbeit, TU München, 1999.
- [59] SCHLÖGL, A.E., MNATSAKANOV, T.T., KUHN, H. und SCHRÖDER, D.: *Temperature Dependent Characterization of Bipolar Silicon Power Semiconductors — A New Physical Model Validated by Device-Internal Probing between 400K–100K*. In: IEEE Transactions on Power Electronics, vol. 15, no. 6, 1267–74, November 2000.
- [60] SCHRODER, D.K.: *Carrier lifetimes in silicon*. IEEE Trans. on Electron Devices, 44(1):160–70, January 1997.

- [61] SCHRÖDER, D.: *Elektrische Antriebe 3: Leistungselektronische Bauelemente*. Springer-Verlag, Berlin, 1. Auflage, 1996.
- [62] SCHRÖDER, D.: *Elektrische Antriebe 4: Leistungselektronische Schaltungen*. Springer-Verlag, Berlin, 1. Auflage, 1997.
- [63] SHENG, A.G.M. und WILLIAMS, B.W.: *A Review of IGBT Models*. In: IEEE Transactions on Power Electronics, vol. 15, no. 6, 1250–1266, November 2000.
- [64] SELBERHERR, S.: *Analysis and Simulation of Semiconductor Devices*. Springer-Verlag, Berlin, 1984.
- [65] SLOTBOOM, J.W. und H.C. DEGRAAF: *Bandgap narrowing in silicon bipolar transistors*. IEEE Trans. on Electron Devices, 24(4):1123–25, 1977.
- [66] STEIMER, P., GRÜNING, H., WERNINGER, J. und SCHRÖDER, D.: *State-of-the-Art Verification of the Haed Driven GTO Inverter Development for a 100 MVA Intertie*. In: Proc. of PESC, Baveno, Italien, 1996.
- [67] STROLLO, A.G.M.: *A New SPICE Model of Power P-I-N Diode Based on Asymptotic Waveform Evaluation*. In: IEEE Transactions on Power Electronics, vol. 12, no. 1, 12–20, January 1997.
- [68] SZE, S.M.: *Physics of Semiconductor Devices*. John Wiley & Sons, New York, 2. Auflage, 1981.
- [69] SZE, S.M.: *Semiconductor Devices – Physics and Technology*. John Wiley & Sons, New York, 2. Auflage, 1985.
- [70] TAN, C.M. und TSENG, K.: *Using Power Diode Models for Circuit Simulation—A Comprehensive Review*. In: IEEE Transactions on Industrial Applications, vol. 46, no. 3, 637–645, June 1999.
- [71] TIETZE, U. und SCHENK, CH.: *Halbleiter-Schaltungstechnik*. Springer-Verlag, Berlin, 11. Auflage, 1999.
- [72] VOGLER, T., A. SCHLÖGL, D. SCHRÖDER und N. KASAHARA: *Modeling and characterizing power semiconductor devices at low temperatures*. In: Proc. of IPEC, Seiten 1232–1238, Yokohama, Japan, 1995.
- [73] VOGLER, T.: *Physikalische Netzwerkmodelle von Leistungshalbleiter-Bauelementen unter Berücksichtigung von Modularität und Temperatur*. Doktorarbeit, TU München, 1996.
- [74] XU, CHIHAO: *Netzwerkmodelle von Leistungshalbleiter-Bauelementen (Diode, BJT und MOSFET)*. Doktorarbeit, TU München, 1990.

- [75] YAMAMOTO M., KURACHI K., SATOH K. und IWAMOTO H.: *Novel characteristics of SGCT (Symmetrical GCT)*. In: Proc. of IPEC, Tokio, Japan, 2000.

A Anhang

Der Modellkopf

```

template igct anode gate kathode = model

electrical anode, gate, kathode

struc {

    enum {yes, no} message = no    # if messages desired for debug
    enum {yes, no} plot      = yes  # if carrier plot desired (only
                                    # once in the whole design !!)
    enum {yes, no} carrier = no    # carrier densities at borders

    string plotfile = "plasma_igct"

    number in0=250p, ip0=10p, Na=5e19, Nb=2e13, Np=5e17, Nd=1e19,
           Nbuf=5e16, NO_l=1.0e19, NO_r=1.0e19, wa=0.0003, w=0.06,
           wp=0.005, wc=0.002, wbuf=0.003, xj_l=0, xj_r=0, A=58,
           vk=0.5, tauh0=15u, taunp=1.0u, tau_law=1, alpha_n=1.5,
           alpha_p=1.5, zeta_n=3, zeta_p=3, T=300, fc=.5, m=0.5,
           freeze=1, dop_d=1, dop_a=1, ip=30, pow=3, aug=1, ehs=1,
           bgn=0, cj0=10n, ugk_av=21, Ecrit=2e5, avf=3e8, num_tau=4,
           x_tau[20]=[0,0.049,0.05,1], tau_rel[20]=[0.3,0.3,1,1]

    } model = ()

{
Modellkörper
}

```

Das Wort „template“ steht allgemein für Modell, das hier im besonderen igct heißt. Danach folgen die Anschlußpins und nach dem „=“ die Modellparameter. Wegen

der großen Anzahl an Parametern werden diese zu „model“ zusammengefaßt und weiter unten aufgelistet. Nach der Deklaration folgt der Modellkörper, der Gleichungen und die Aufrufe von Submodulen und der C-Routinen enthält. Zu den Parametern gehören Defaultwerte, die immer dann gelten, wenn der Anwender diese nicht in der Netzliste überschreibt. Diese Defaultwerte entsprechen nicht den Parametrisierungen eines realen IGCTs.

Auflistung der Parameter

<code>in0</code>	Sperrsättigungsstrom der Anode in A
<code>ip0</code>	Sperrsättigungsstrom der Kathode in A
<code>Na</code>	Anodendotierung in $1/cm^3$
<code>Nb</code>	Driftzonendotierung in $1/cm^3$
<code>Np</code>	Dotierung der p -Basis in $1/cm^3$
<code>Nd</code>	Dotierung der Kathode in $1/cm^3$
<code>Nbuf</code>	Dotierung der Pufferschicht in $1/cm^3$
<code>NO_l</code>	max. Dotierung eines exponentiellen p -Profils (vgl. Abschnitt 2.4.1)
<code>NO_r</code>	max. Dotierung eines exponentiellen n -Profils (vgl. Abschnitt 2.4.1)
<code>wa</code>	Anodenweite in cm
<code>w</code>	Gesamtweite Driftzone plus Pufferschicht in cm
<code>wp</code>	Weite der p -Basis in cm
<code>wc</code>	Kathodenweite in cm
<code>wbuf</code>	Weite der Pufferzone in cm
<code>xj_l</code>	Ausdehnung des exponentiellen p -Profils in cm
<code>xj_r</code>	Ausdehnung des exponentiellen n -Profils in cm
<code>A</code>	aktive Anodenfläche in cm^2
<code>vk</code>	Anteil der Kathodenfläche, d.h. $vk = A_K/A$
<code>tauh0</code>	Hochinjektionslebensdauer nach Shockley-Read-Hall (SRH) in s
<code>taunp</code>	τ_n in der p -Basis in s
<code>tau_law</code>	Umschaltmöglichkeit zwischen zwei Formeln für $\tau(T)$
<code>alpha_n</code>	Exponent in einer der beiden Formeln für $\tau_n(T)$
<code>alpha_p</code>	Exponent in einer der beiden Formeln für $\tau_p(T)$
<code>zeta_n</code>	Exponent in der anderen Formel für $\tau_n(T)$
<code>zeta_p</code>	Exponent in der anderen Formel für $\tau_p(T)$
<code>T</code>	Umgebungstemperatur in K
<code>fc</code>	Parameter in Formel zur Berechnung der Verarmungskapazität
<code>m</code>	Parameter in Formel zur Berechnung der Verarmungskapazität
<code>freeze</code>	=1 unvollständige Ionisation; =0 vollständige Ionisation

dop_d	Art der Donatoren: =1 Phosphor; =2 Arsen; =3 Antimon
dop_a	Art der Akzeptoren: =1 Bor
ip	Anzahl der ortsdiskreten Stützstellen
pow	Legt Lage der Stützstellen fest: Exponent α_S von (2.70)
aug	=1 Auger ein; =0 Auger aus
ehs	=1 EHS nach Mnatsakanov ein; =0 EHS aus
bgn	Bandgap Narrowing: =0 kein BGN; =1,2 BGN ein (s. Abschnitt 1.2.3)
cj0	Verarmungskapazität für $U=0$
ugk_av	Durchbruchsspannung des Gate/Kathoden-Übergangs
Ecrit	kritische Feldstärke für Lawineneffekt
avf	Parameter, der Stärke des Lawineneffekts bestimmt (vgl. (2.95))
num_tau	Anzahl der Stützstellen eines Lebensdauerprofils (=0: kein Profil)
x_tau	Lage der Stützstellen normiert auf w
tau_rel	Zu x_tau gehörende Lebensdauern normiert auf τ_{uh0}

Symbolverzeichnis

Formelgrößen

Symbol	Beschreibung	Einheit
A	(Anoden)Fläche	m^2
$\alpha_{n,p}$	Ionisationskoeffizient bzgl. Lawineneffekt	$1/m$
\vec{B}	magnetische Flußdichte	Vs/m
b	Beweglichkeitsverhältnis μ_n/μ_p	
C	Kapazität	F
C_j	Speerschichtkapazität	F
c	Verhältnis der Diffusionskoeffizienten D_n/D_p	
\vec{D}	dielektrische Verschiebungsdichte	As/m^2
D	Diffusionskoeffizient allgemein	m^2/s
$D_{n,p}$	Diffusionskoeffizient für Elektronen, Löcher	m^2/s
D_A	ambipolare Diffusionskonstante	m^2/s
E	Elektrische Energie	J oder eV
E_C	Energieniveau der Leitungsbandkante	J oder eV
E_F	Ferminiveau	J oder eV
$E_{F_{n,p}}$	Quasiferminiveaus	J oder eV
E_G	Bandlücke E_C-E_V	J oder eV
E_{F_i}	Fermienergie bei Eigenleitung	J oder eV
E_T	Energieniveau der Fangstellen (Traps)	J oder eV
E_V	Energieniveau der Valenzbandkante	J oder eV
E_A^*	Ionisationsenergie für Akzeptoren	J oder eV
E_D^*	Ionisationsenergie für Donatoren	J oder eV
\vec{E}, E	Elektrisches Feld	V/m
E_{crit}	kritische Feldstärke bzgl. Lawineneffekt	V/m
ϵ_r	materialabhängige Dielektrizitätszahl	
$f(E)$	Besetzungswahrscheinlichkeit	
$\phi_{n,p}$	Quasifermpotentiale	V
$G(T)$	Generationsrate	$1/(cm^3s)$
g	Degenerationsfaktor	
\vec{H}	magnetische Feldstärke	A/m
I	Strom	A
\vec{j}, j	Stromdichte	A/m^2
j_{dis}	Maxwellsche Verschiebestromdichte	A/m^2
j_{ges}	Gesamtstromdichte	A/m^2
$\vec{j}_{n,p}, j_{n,p}$	Elektronen- bzw. Löcherstromdichte	A/m^2
$j_{n0,p0}$	Sperrsättigungsstromdichte in der Anode, Kathode	A/m^2

Symbol	Beschreibung	Einheit
\vec{k}, k	Wellenzahlvektor, Wellenzahl	$1/m$
L	Induktivität	H
L_A	ambipolare Diffusionslänge	m
$L_{n,p}$	Diffusionslänge von Elektronen bzw. Löchern	m
M_{AV}	Multiplikationsfaktor beim Lawineneffekt	
$\mu_{n,p}$	effektive Elektronen-, bzw. Löcherbeweglichkeit	$m^2/(Vs)$
$m_{n,p}^*$	Effektive Masse für Elektronen bzw. Löcher	kg
$N(E)$	Zustandsdichte freier Ladungsträger	$1/(Jcm^3)$
N_C	äquivalente Zustandsdichte des Leitungsbandes	$1/(cm^3)$
N_V	äquivalente Zustandsdichte des Valenzbandes	$1/(cm^3)$
N_A	Dotierdichte mit Akzeptoren	$1/(cm^3)$
N_D	Dotierdichte mit Donatoren	$1/(cm^3)$
N_{eff}	effektive Dotierung während des Schaltens	$1/cm^3$
N_ν	Donatordichte der schwach dotierten Driftzone	$1/(cm^3)$
n	Dichte der Leitungselektronen	$1/cm^3$
n_i	Eigenleitungsträgerdichte	$1/cm^3$
n_0	Elektronendichte im thermischen Gleichgewicht	$1/(cm^3)$
P	Elektrische Leistung	W
p	Dichte freier Löcher	$1/cm^3$
p_0	Löcherdichte im thermischen Gleichgewicht	$1/(cm^3)$
Q	Elektrische Ladung	As
$R(T)$	Rekombinationsrate	$1/(cm^3s)$
$r(T)$	Rekombinationskoeffizient	$1/(cm^9s)$
ρ	elektrische Raumladungsdichte	As/m^2
σ	spezifischer Leitwert	$A/(Vm)$
T	Temperatur	K
t	Zeit	s
$\tau_{n,p}$	Lebensdauer von Elektronen, bzw. Löchern	s
τ_A	Auger-Lebensdauer	s
τ_c	mittlere Flugzeit	s
τ_d	Relaxationszeit	s
τ_h	Hochinjektionslebensdauer	s
τ_{SRH}	Lebensdauer aufgrund indirekter Rekombination	s
U	Spannung	V
U_D	Diffusionsspannung	V
U_j	Übergangsspannung	V
V	Potential	V
$\vec{v}_{n,p} v_{n,p}$	Driftgeschwindigkeit für Elektronen, Löcher	m/s
$v_{sat,n,p}$	Sättigungsgeschwindigkeit für Elektronen, Löcher	m/s

Symbol	Beschreibung	Einheit
w	Zonenweite	m
x	Ortskoordinate im eindimensionalen Modellen	m
$\vec{\nabla}$	Nabla-Operator	

Physikalische Konstanten

Symbol	Bezeichnung	Wert
q_0	Elementarladung	$1.6021892 \cdot 10^{-19} \text{As}$
k_B	Boltzmannkonstante	$1.380662 \cdot 10^{-23} \text{J/K}$
m_0	Elektronenruhemasse	$9.109543 \cdot 10^{-31} \text{kg}$
ϵ_0	Elektrische Feldkonstante	$8.85418782 \cdot 10^{-12} \text{As/(Vm)}$