

2D – Simulation des Interbandtunnels in inhomogenen elektrischen Feldern

M. STAHL / A. SCHENK / H.-J. WÜNSCHE
Fachbereich Physik der Humboldt-Universität zu Berlin
Invalidenstr. 110, O-1040 Berlin, Germany

Abstract

Das Interbandtunneln ist der feldinduzierte Übergang von Ladungsträgern zwischen verschiedenen Bändern. Dieser Prozeß kann insbesondere bei Bauelementen aus direkten Halbleitern einen wesentlichen Einfluß auf die Kennlinien ausüben. Die bisher in der Literatur für die Abschätzung des Tunnelstromes verwendeten Formeln setzten die Homogenität des elektrischen Feldes voraus. Da diese Annahme jedoch in vielen Fällen nicht gerechtfertigt ist, haben wir eine nichtlokale Theorie für inhomogene Felder entwickelt.

Diese wurde in das 2D-Simulationsprogramm TOSCA implementiert. Dabei kam es darauf an, den Potentialverlauf entlang der jeweiligen Richtung des elektrischen Feldes auch zwischen den Gitterpunkten numerisch möglichst genau zu bestimmen, wozu drei verschiedene Varianten untersucht werden. Es zeigt sich, daß i.a. nur eine zweidimensionale Interpolation dieses Verlaufes hinreichend genaue Ergebnisse liefert.

Für Dioden aus $Hg_{0,8}Cd_{0,2}Te$ wird mit Hilfe dieses Modells gezeigt, daß die Nettotunnelrate und damit auch der Tunnelstrom ganz empfindlich vom Verlauf des elektrischen Feldes abhängen. Darüber hinaus wird demonstriert, daß nicht nur das üblicherweise berücksichtigte Dotierungsprofil in die Tiefe, sondern auch der seitliche Verlauf der Dotierung einen erheblichen Einfluß auf den Tunnelstrom haben kann.

Unter Interbandtunneln versteht man quantenmechanische Übergänge von Ladungsträgern zwischen verschiedenen Bändern im Halbleiter, welche durch das innere elektrische Feld der Bauelemente induziert werden. Diese Prozesse können insbesondere bei Halbleiterdioden aus direktem Material und mit einem schmalen Gap, auf die wir uns hier beschränken wollen, einen starken Einfluß auf die Kennlinien ausüben.

In Abbildung 1 (Seite 30), wo die Energie über dem Ort aufgetragen ist, wurde ein solcher Übergang eines Elektrons vom Valenz- ins Leitungsband schematisch dargestellt. Eine Besonderheit der Interbandtunnelprozesse besteht in ihrer Nichtlokalität, d.h. sie sind immer mit einer Ortsveränderung der beteiligten Ladungsträger um die sogenannte Tunnellänge l_t verbunden. Das resultiert daraus, daß diese Übergänge ohne Energie- und Impulsaustausch mit anderen Teilchen (wie z.B. den Photonen oder den Phononen) stattfinden. Vielmehr entnehmen die Ladungsträger die zur Überwindung des Gaps benötigte Energie aus dem inneren elektrischen Feld des Bauelementes, womit die gesamte potentielle Energie der Ladungsträger im Halbleiter konstant bleibt, wie dies auch anhand von Abbildung 1 (Seite 30) deutlich wird.

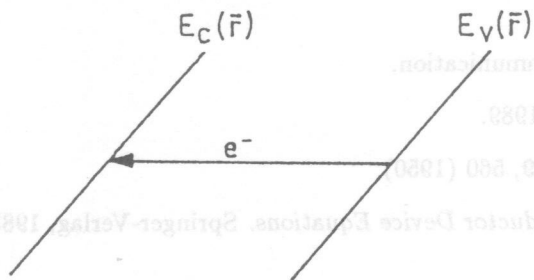


Figure 1: Schematische Darstellung des Interbandtunnels

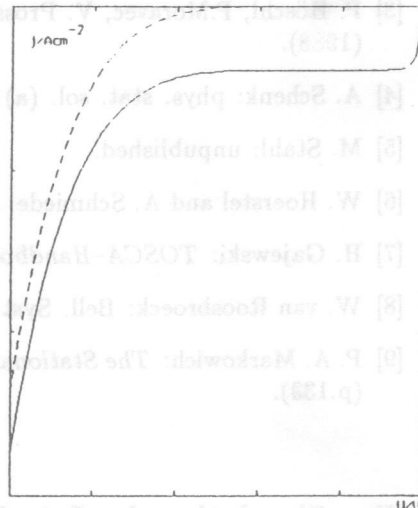


Figure 2: Typische Verläufe der Tunnelkennlinie (---) und der Gesamtkennlinie (—) einer Halbleiterdiode im Sperrfall

In Abbildung 2 (Seite 30) sind typische Verläufe der Tunnelkennlinie und der daraus resultierenden Gesamtkennlinie einer Halbleiterdiode schematisch gezeigt. Man erkennt, daß das typische Sperrverhalten der Diode für große Sperrspannungen durch den Einfluß des Interbandtunnels auf jeden Fall zerstört wird. Jedoch auch für kleinere Sperrspannungen beeinflusst der Tunnelstrom das durch die anderen Stromanteile, wie SHR - und Augerstrom, bestimmte Verhalten der Diode mehr oder weniger stark.

In früheren Arbeiten (z.B. [1], [2]) wurden zur Abschätzung des Tunnelstroms Ausdrücke der Form

$$j_{tu} = j_o(F) \exp\left(\frac{-F_c}{F}\right) \quad (0.1)$$

benutzt, wobei F das elektrische Feld und F_c eine kritische Feldstärke bezeichnen. Diese Modelle setzen jedoch die Homogenität des elektrischen Feldes über der gesamten Tunnellänge voraus. Diese Forderung ist für große Sperrspannungen auch recht gut erfüllt, da der Potentialverlauf in der Raumladungszone hier weitgehend linearisiert ist. Für kleine Sperrspannungen dagegen gilt diese Annahme i.a. jedoch nicht mehr, so daß der Tunnelstrom überschätzt wird. Interessant sind aber gerade der Bereich der Kennlinie, wo die Diode noch ein gutes Sperrverhalten zeigt und insbesondere dessen Grenze, die sogenannte Durchbruchsspannung. Da (1) jedoch gerade in diesem Bereich kleiner und mittlerer Sperrspannungen versagt, haben wir ein Modell entwickelt, welches die Inhomogenitäten des elektrischen Feldes wesentlich besser berücksichtigt.

Ausführlich beschrieben ist dieses Modell der mittleren Inhomogenität in [3]. Die dortige Herleitung einer Nettorate für die Interbandtunnelübergänge (Nettotunnelrate) an einem Ort \vec{r} erfolgt im Rahmen der Effektivmassenäherung. Die Vorteile des Modells beruhen auf der verbesserten näherungsweise Berechnung des Überlappungsintegrals unter der Voraussetzung der mittleren Inhomogenität des elektrischen Feldes

$$\left|1 - \frac{F(x_o + l_i)}{F(x_o)}\right| \ll 1, \quad (0.2)$$

wobei x_o der Punkt der maximalen Überlappung der Wellenfunktionen und l_i ein Maß für die Halbwertsbreite der Überlappungsfunktion sind. Für kleine Sperrspannungen, d.h. große

Feldinhomogenitäten fordert diese Bedingung die Homogenität des elektrischen Feldes nur über einer sehr viel kleineren Strecke als der Tunnellänge, während l_i für große Sperrspannungen in die Größenordnung von l_i kommt, so daß das Modell der mittleren Inhomogenität für große Sperrspannungen automatisch die schon früher benutzten, dort anwendbaren Ausdrücke der Form (1) reproduziert.

Die Ableitung des Modells der mittleren Inhomogenität läßt sich exakt nur für eine parabolische Bandstruktur ausführen. Insbesondere schmallückige Halbleiter, bei denen das Interbandtunneln ja besonders wahrscheinlich ist, weisen jedoch oft eine starke Kopplung der Bänder auf, was zu einer KANEschen Bandstruktur führt, welche durch

$$E_\mu(\vec{r}, \vec{k}) = E_o(\vec{r}) \pm \frac{E_g}{2} \sqrt{1 + 2 \frac{\hbar^2 \vec{k}^2}{m_c E_g}}; \quad \mu = (c) \quad (0.3)$$

$$E_o(\vec{r}) = \frac{E_v + E_c}{2} - e\varphi(\vec{r}) \quad (0.4)$$

beschrieben wird. Dabei bezeichnen E_g die Gapenergie, m_c die effektive Bandkantenmasse der Elektronen und \vec{k} ist der Wellenvektor. Wie in [4] ausführlicher diskutiert wird, ist das Modell der mittleren Inhomogenität jedoch auch hier in vielen Fällen anwendbar.

Man erhält schließlich für die Nettotunnelrate den Ausdruck

$$R_{tu}(\vec{r}) = \frac{3e|e\nabla\varphi(\vec{r})|^4}{32\pi(2\mu E_g)^{5/2}} \frac{m_c m_v}{E_g} [f_c(\vec{r}) - f_v(\vec{r})] S_{red}(\vec{r}) M(\vec{r}). \quad (0.5)$$

Hierbei sind $f_c(\vec{r})$ bzw. $f_v(\vec{r})$ die Besetzungsfunktionen der Valenz- und Leitungsbandzustände

$$f_v(\vec{r}) = [1 + \exp(\frac{E_o(\vec{r}) - F_p(\vec{r})}{kT})]^{-1} \quad (0.6)$$

bzw.

$$f_c(\vec{r}) = [1 + \exp(\frac{E_o(\vec{r}) - F_n(\vec{r})}{kT})]^{-1} \quad (0.7)$$

und $M(\vec{r})$ im wesentlichen die bekannte elektrooptische Funktion

$$M(\vec{r}) = 8\pi z(\vec{r}) \{Ai^2(z(\vec{r})) - z(\vec{r}) [Ai'(z(\vec{r}))]^2\} \quad (0.8)$$

mit

$$z(\vec{r}) = [\frac{3}{2} \Delta S(\vec{r})]^{3/2}, \quad (0.9)$$

$$\Delta S(\vec{r}) = |S_c(\vec{r}) - S_v(\vec{r})| \quad (0.10)$$

und der Airy - Funktion

$$Ai(x) = \frac{1}{\pi} \int_0^\infty dt \cos(\frac{1}{3}t + xt) \quad (0.11)$$

Das reduzierte Wirkungsintegral ist über

$$S_{red}(\vec{r}) = \frac{|S_c(\vec{r})S_v(\vec{r})|}{\Delta S(\vec{r})}. \quad (0.12)$$

definiert. Die schon erwähnte Nichtlokalität des Interbandtunnelns wird nun über die sogenannten Wirkungsintegrale

$$S_\mu(\vec{r}) = \int_0^{\delta r_\mu} dt \kappa_\mu(\vec{r} + \vec{e}t); \quad \mu = c, v \quad (0.13)$$

berücksichtigt, wobei sich der quasiklassische Wellenvektor bei der KANEschen Bandstruktur zu

$$\kappa_{\mu}(\vec{r} + \vec{e}t) = \sqrt{\frac{2em_{\mu}}{\hbar^2}} \sqrt{|(\varphi(\vec{r} + \vec{e}t) - \varphi(\vec{r}_{\mu})) [1 - \frac{e}{E_g}(\varphi(\vec{r} + \vec{e}t) - \varphi(\vec{r}_{\mu}))]|} \quad (0.14)$$

ergibt. Die Integration in (13) erfolgt, beginnend am Aufpunkt \vec{r} , in Richtung des dortigen elektrischen Feldes

$$\vec{e} = \frac{\nabla\varphi(\vec{r})}{|\nabla\varphi(\vec{r})|} \quad (0.15)$$

bis zu dem jeweiligen klassischen Umkehrpunkt

$$\vec{r}_{\mu} = \vec{r} + \vec{e} \delta r_{\mu}, \quad (0.16)$$

wobei sich δr_{μ} aus der Bedingung

$$e[\varphi(\vec{r}_{\mu}) - \varphi(\vec{r})] = \mp \frac{E_g}{2}; \quad \mu = (c) \quad (0.17)$$

ergibt.

Dieses Modell wurde nun in das 2D - Simulationsprogramm TOSCA, welches das Grundgleichungssystem der inneren Elektronik numerisch löst, implementiert. Obwohl der Interbandtunneleffekt auch hier als eindimensionaler Prozeß betrachtet wird, bleibt doch seine grundsätzliche Nichtlokalität erhalten. Es kam also darauf an, den Verlauf des elektrostatischen Potentials in Richtung des elektrischen Feldes am Aufpunkt \vec{r} in einer gewissen Umgebung dieses Aufpunktes möglichst genau zu bestimmen. Hierzu wurden drei verschieden Varianten erstellt und getestet, welche im folgenden kurz beschrieben werden sollen.

Im einfachsten Fall wurde vorausgesetzt, daß die Feldrichtung parallel zu den Gitterlinien des in TOSCA verwendeten Grundgitters verläuft, was insbesondere bei eindimensionalen Modellrechnungen sinnvoll ist. Dann können mit Hilfe eines relativ einfachen Suchalgorithmus prinzipiell alle in Richtung des elektrischen Feldes liegenden Gitterpunkte in eine eindimensionale Interpolation des elektrostatischen Potentials einbezogen werden. Aus Zeitgründen beschränken wir uns bei dieser ersten Variante jedoch auf eine Interpolation 4. Grades.

I.a. weicht die Richtung des elektrischen Feldes jedoch von der Richtung der Gitterlinien aus TOSCA ab. In einer zweiten Variante wurde der Verlauf des elektrostatischen Potentials deshalb mit Hilfe einer Taylorentwicklung 2. Grades in einer gewissen Umgebung des Aufpunktes \vec{r} extrapoliert, was auf

$$\varphi(\vec{r} + t\vec{e}) = \varphi(\vec{r}) + t\nabla\varphi(\vec{r})\vec{e} + \frac{t^2}{2} \left[-\frac{\rho(\vec{r})}{\epsilon_r\epsilon_0} + 2\frac{\partial^2\varphi(\vec{r})}{\partial x\partial y} e_x e_y \right] \quad (0.18)$$

führt, wobei $\rho(\vec{r})$ die Raumladungsdichte bezeichnet. Obwohl sich der Verlauf des elektrostatischen Potentials in diesem Falle prinzipiell aus den lokalen Größen am Aufpunkt \vec{r} bestimmt, weist diese Variante jedoch zwei wesentliche Nachteile auf. Einerseits liefert Gleichung (18) nur in einer vergleichsweise kleinen Umgebung von \vec{r} gute Ergebnisse, so daß sie für große Tunnelängen, d.h. große Feldinhomogenitäten relativ schnell versagt. Andererseits sind in TOSCA weder die partiellen Ableitungen des elektrostatischen Potentials in Richtung der Koordinatenachsen noch dessen gemischte zweite Ableitung, wie sie in (18) benötigt werden, verfügbar, was wiederum zu i.a. nicht gerechtfertigten Einschränkungen führt.

Bei der dritten Variante wurde der Verlauf des elektrostatischen Potentials deshalb mit Hilfe einer zweidimensionalen Interpolation unter Einbeziehung der nächsten Gitterpunkte des Aufpunktes \vec{r} bestimmt. Ein genaueres Vorgehen durch die Einbeziehung weitere Gitterpunkte ist einerseits aus Zeitgründen nicht möglich und andererseits auch nicht sinnvoll, da auch das Programmsystem TOSCA nicht genauer arbeitet. Obwohl der Potentialverlauf hier also insgesamt nur parabolisch interpoliert wird, weist diese Variante den großen Vorteil auf, daß alle Feldrichtungen gleichberechtigt berücksichtigt werden.

Bei Testrechnungen zum Vergleich der drei eben beschriebenen Varianten wurden die diskutierten Vor- und Nachteile bestätigt. Die zweite Variante liefert somit in den meisten Fällen die ungenaueren Ergebnisse. Obwohl auch die mit der dritten Variante erzielten Resultate, insbesondere für sehr kleine Sperrspannungen, von denen der ersten Variante abweichen, bleiben die Fehler im interessierenden Bereich mittlerer Sperrspannungen in vertretbaren Grenzen, womit sich die dritte Variante als die i.a. zuverlässigste erweist.

Abschließend sollen noch kurz einige Ergebnisse vorgestellt werden, wie sie mit Hilfe des beschriebenen Modells bei der Untersuchung von Photodioden aus $Hg_{0.8}Cd_{0.2}Te$ erzielt wurden.

Einerseits zeigte sich schon bei eindimensionalen Modellrechnungen, daß die Nettotunnelrate und damit auch der Tunnelstrom sehr empfindlich vom Verlauf des elektrischen Feldes und somit letztlich vom jeweiligen Dotierungsprofil abhängen. In Abbildung 3 (Seite 34) sind die beiden bei den Rechnungen benutzten Verläufe der Nettodotierung über der Tiefe gezeigt. Es sind dies einerseits ein abruptes Dotierungsprofil und andererseits ein in der Raumladungszone wesentlich flacher verlaufendes n^+n^-p -Profil nach [5], deren asymptotische Werte jedoch ähnlich sind. Der Vergleich der entsprechenden, in Abbildung 4 (Seite 34) dargestellten örtlichen Verläufe der Nettotunnelrate zeigt, daß sich ihr Maximum bei dem flacher verlaufenden Dotierungsprofil sehr stark um ca. 9,5 Größenordnungen verkleinert. Dies kommt natürlich auch in den entsprechenden Kennlinien im Sperrfall, wie sie in Abbildung 5 (Seite 35) gezeigt sind, zum Ausdruck. Während die Struktur mit dem abrupten Dotierungsprofil praktisch kein Sperrverhalten aufweist, wird der Tunnelstrom in der Struktur mit dem flacheren n^+n^-p -Profil erst bei einer Durchbruchspannung von $U = -0,27V$ wesentlich.

Andererseits stellte sich bei der zweidimensionalen Modellierung dieser Strukturen heraus, daß auch der seitliche Verlauf der Dotierungskonzentration ein dem Dotierungsprofil in die Tiefe vergleichbar starken Einfluß auf den Tunnelstrom haben kann. In der Abbildung 6 (Seite 35) sind die Höhenlinienbilder der entsprechenden Dotierungsprofile gezeigt. Während in die mit y bezeichnete Tiefe generell das flache n^+n^-p -Profil benutzt wurde, ist für den seitlichen Abfall der Dotierungskonzentration in x -Richtung einmal ein gaußförmiger und einmal ein dem n^+n^-p -Profil in die Tiefe ähnlicher Verlauf simuliert worden. Die asymptotischen Werte der Dotierungskonzentrationen bleiben dabei, wie auch im eindimensionalen Fall, in etwa vergleichbar. In Abbildung 7 (Seite 36) sind die entsprechenden Oberflächenbilder der Nettotunnelrate gezeigt. Man erkennt, daß im ersten Fall der seitliche n^+p -Übergang so stark dominiert, daß hier praktisch alle für den Tunnelstrom wesentlichen Ladungsträger erzeugt werden. Weitere Rechnungen zeigten darüber hinaus, daß sich auch bei einer Variation der Halbwertsbreite des seitlichen Gaußabfalls der Dotierungskonzentration keine qualitative Änderung dieser Situation ergibt. Dies wird erst im zweiten Fall erreicht, wo auch das seitliche Dotierungsprofil ähnlich flach wie das in die Tiefe verläuft. Einerseits werden hier die Beiträge zum Tunnelstrom aus dem seitlichen np -Übergang und dem in die Tiefe vergleichbar und andererseits wird auch die Gesamtzahl der durch das Interbandtunneln erzeugten Ladungsträger wesentlich kleiner. Das spiegelt sich natürlich wiederum auch an den abschließend in Abbildung 8 (Seite 36) gezeigten Kennlinien wieder. Der Tunneldurchbruch der Struktur mit dem flachen seitlichen Dotierungsprofil tritt wesentlich später auf als bei jener mit dem gaußförmigen seitlichen Abfall der Dotierungskonzentration, obwohl die letztere wenigstens in einem vergleichsweise kleinen Bereich das typische Sperrverhalten von Halbleiterdioden aufweist, da der gaußförmige Abfall doch flacher verläuft als bei dem oben diskutierten abrupten Dotierungsprofil.

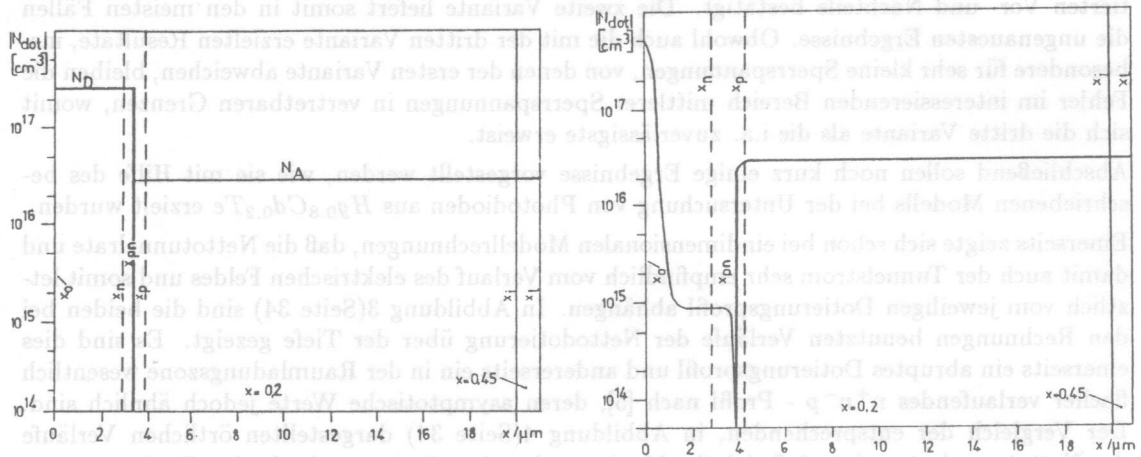


Figure 3: Verläufe der Nettodotierung über der Tiefe für die Struktur mit dem abrupten (a) bzw. dem n^+n^-p -Profil (b)

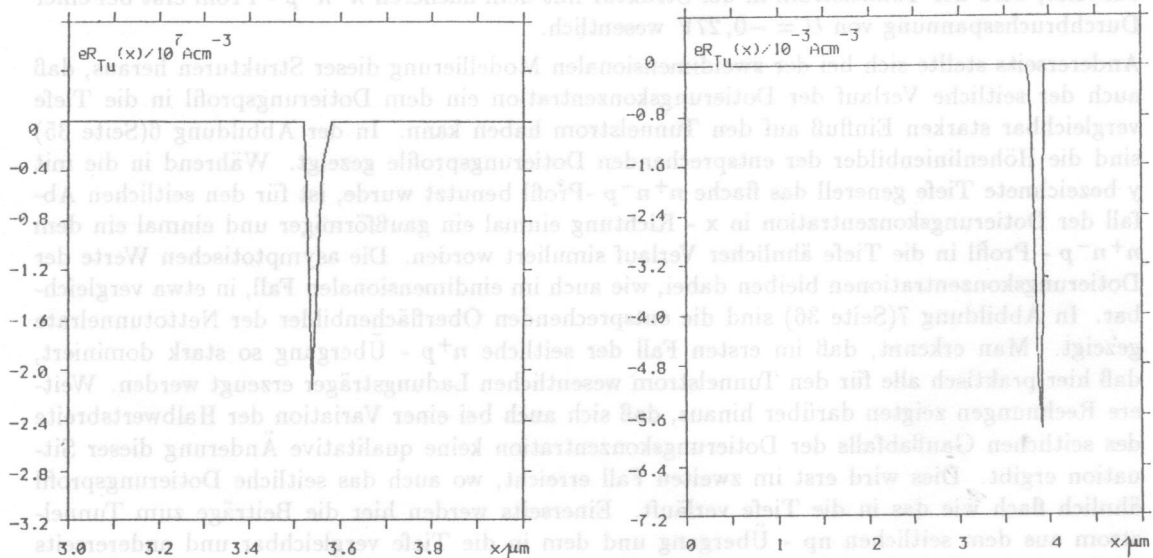


Figure 4: Verläufe der Nettotunnelrate über der Tiefe für die Struktur mit dem abrupten (a) bzw. dem n^+n^-p -Profil (b)

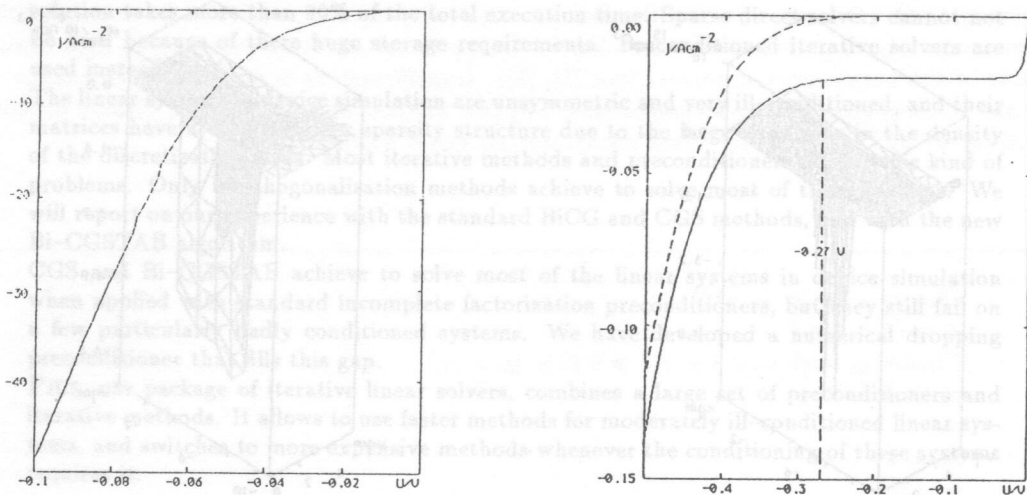


Figure 5: Kennlinien der Struktur mit dem abrupten (a) bzw. dem n^+n^-p -Profil (b)

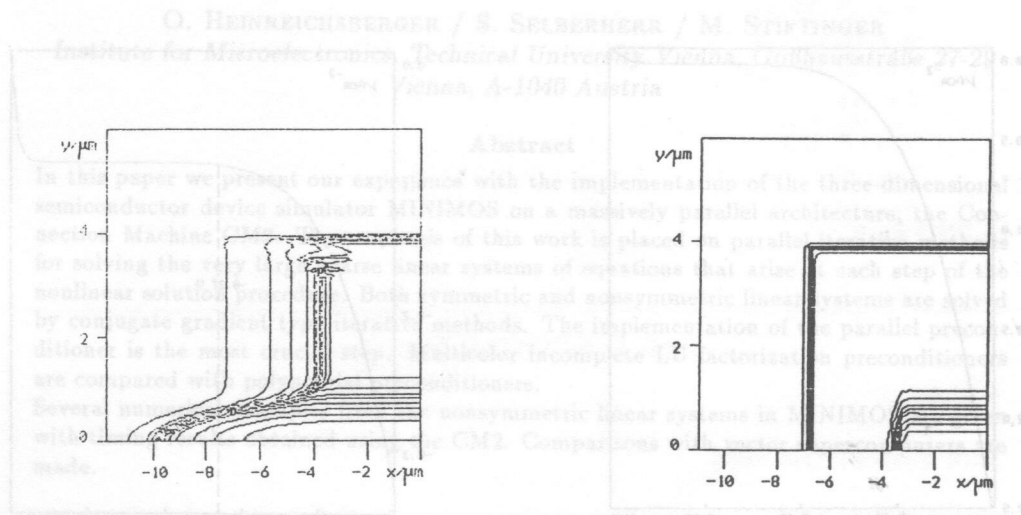


Figure 6: Höhenlinienbilder der Nettodotierung für die Struktur mit dem gaußförmigen (a) bzw. dem n^+n^-p -ähnlichen (b) seitlichen Abfall der Dotierungskonzentration

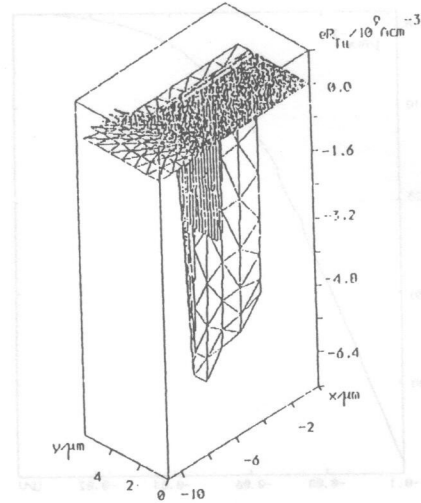
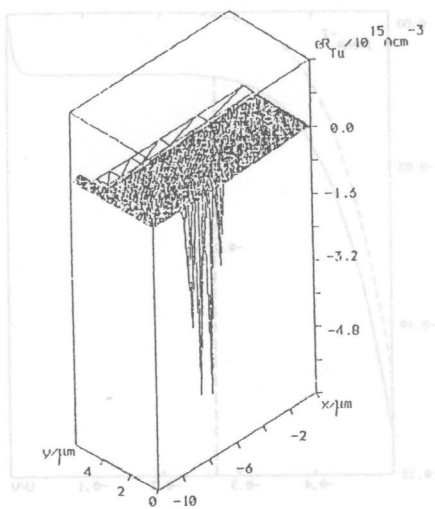


Figure 7: Oberflächenbilder der Nettotunnelrate für die Struktur mit dem gaußförmigen (a) bzw. dem n^+n^-p -ähnlichen (b) seitlichen Abfall der Dotierungskonzentration

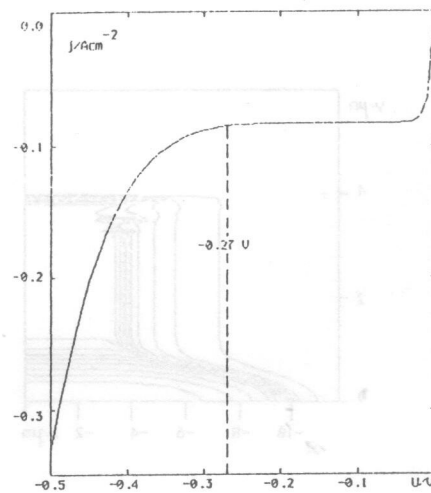
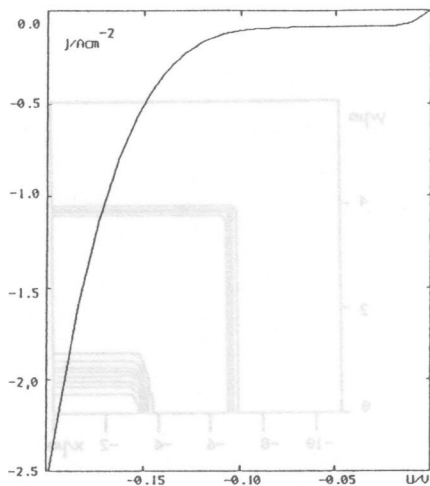


Figure 8: Kennlinien der Struktur mit dem gaußförmigen (a) bzw. dem n^+n^-p -ähnlichen (b) seitlichen Abfall der Dotierungskonzentration

References

- [1] E.O. Kane: J. Phys. Chem. Sol., 12 (1959), 181.
- [2] J.L. Moll: Physics of Semiconductors, New York, 1964.
- [3] M. Stahl: Diplomarbeit, Berlin, 1988.
- [4] M. Stahl: Dissertation, Berlin, 1991.
- [5] L.O. Bubulac et al.: J. Vac. Sci. Technol. A, 5 (1987) 5, 3166.