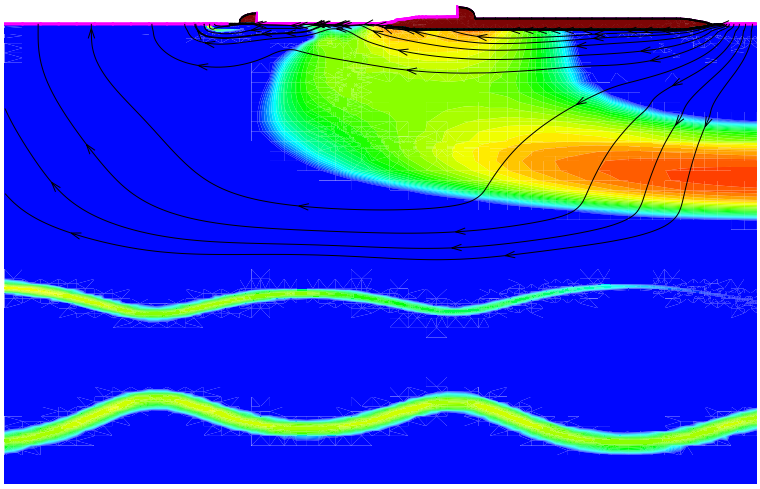
Universiteit Gent
Faculteit Toegepaste Wetenschappen
Vakgroep Elektronica en Informatiesystemen

Vermogenstransistoren in een Submicron Digitale CMOS-Technologie

Power Transistors in a Submicron Digital CMOS Technology

Benoit Bakeroot



Proefschrift tot het verkrijgen van de graad van
Doctor in de Toegepaste Wetenschappen
Elektrotechniek
Academiejaar 2003–2004

Promotoren: Prof. André Van Calster en Prof. Jan Doutreloigne

Universiteit Gent
Vakgroep Elektronica en Informatiesystemen
TFCG/IMEC
Sint-Pietersnieuwstraat 41
B-9000 Gent
België

Picture on the front cover shows a lateral IGBT device with double buried layer structure (containing a patterned BLN) at on-state breakdown (red represents the highest electric fields). The black lines show how the hole current is flowing at that stage.

Dankwoord

Dit manuscript mag dan wel door mezelf geschreven zijn, het zou niet tot stand gekomen zijn zonder de liefde en de steun van Vanessa, zonder de zorg van mijn ouders of zonder het vertrouwen van prof. Van Calster.

Het werk dat hier beschreven wordt, kon enkel verwezenlijkt worden met hulp van de eerste zorgen van Miguel, van het immer luisterend oor van prof. Jan Doutreloigne, van het wetenschappelijk aanstekelijk enthousiasme van Peter Moens. Dank aan AMIS en de technologiegroep van Marnix Tack voor het realiseren van de chips, in het bijzonder aan Davide Bolognesi en zijn (ex-)collega's Davy Villanueva, Dominique Wojciechowski en Henri-Xavier Delecourt. Ook Ronny Blomme, die steeds paraat stond bij het oplossen van de software- en hardwareproblemen, Peter Sebrechts voor de vele pc-hulp en de ganse TFCG-groep kon ik niet missen. Dank aan mijn lotgenoten uit Lausanne, Costin Anghel en Nasser Hefyene; aan Guido Groeseneken, Geert Van den Bosch en Brahim Elattari uit IMEC, Leuven voor de verhelderende discussies.

Niet in het minst zou ik mijn vrienden willen danken, vooral de eetmakkers die sinds het begin van mijn verblijf in de TFCG-groep elk middagmaal tot een rustpunt konden scheppen; Youri voor de vele levenswijsheden, Jürgen de oeverloze verteller; Michiel, Ewout, Piere en Johannes. Bedankt !

Benoit Bakeroot

Gent, 13 april 2004.

Inhoudsopgave — Contents

Nederlandstalige Samenvatting (*Dutch Summary*)

1 Inleiding

Een beetje geschiedenis

Alhoewel de eerste patenten op de veldeffecttransistor (of MOSFET) reeds in de jaren 30 van de vorige eeuw verschenen, toch was het pas kort na de tweede wereldoorlog dat de eerste bipolaire siliciumtransistor ook echt werd gemaakt. Sommigen noemen deze gebeurtenis de eerste elektronische revolutie, daarbij veronderstellen ze stilzwijgend dat er nog een tweede is. Dit is misschien wat overdreven en de term evolutie is meer gepast. De grootste stappen tijdens deze evolutie waren—van ons perspectief uit gezien—de uitvinding van de thyristor in 1956, de ontwikkeling van de eerste MOS transistor in de jaren 60 (dus het duurde ongeveer 30 jaar om de technische problemen te overwinnen) en van de eerste vermogen-MOS in de jaren 70 en de uitvinding van de IGBT in de jaren 80.

Deze elektronische evolutie heeft twee wegen bewandeld: langs de ene kant de informatieverwerkende technologie met haar typische, constant afnemende transistordimensies. Als gevolg is deze technologie op dit moment in staat om miljarden (!) transistoren op een chip van enkele vierkante centimeter groot te integreren. De snelheid van deze evolutie is verbazingwekkend, de eerste transistor (gefabriceerd in 1947) was enkele vierkante centimeters groot op zich. Maar wat meer is, wie zou nu nog een wereld kunnen voorstellen zonder computers, zonder GSM, zonder satellieten...

Zoals gezegd, de elektronische evolutie volgt ook nog een tweede

spoor die misschien minder zichtbaar is, maar daarom niet minder belangrijk. Het is het ontstaan van de vermogenelektronica die het mogelijk maakt de elektrische energie te controleren en te converteren. Het begon in de jaren 50 met de ontwikkeling van de bipolaire vermogenstransistor. Alhoewel deze vermogentechnologie bijna gelijktijdig met de informatieverwerkende technologie startte, is ze sindsdien er door overschaduwd. Niettemin heeft ze haar eigen weg gevolgd en is ze of zal ze in de nabije toekomst uit de schaduw treden.

Dit is wat sommigen de tweede elektronische revolutie noemen, het ontstaan van de intelligente vermogentechnologie. Het is in feite het samenkomen van beide wegen—de controle van energie met vermogenelektronica en de informatieverwerkende elektronica met de CMOS-technologie. Deze intelligente vermogentechnologieën bestaan reeds en worden gebruikt in talloze toepassingen waar controle op elektrische motoren belangrijk is (van printers tot auto's). De verwachting is dat deze intelligente vermogentechnologieën een even grote maatschappelijke impact zullen hebben als de informatieverwerkende technologieën. Om nog maar te zwijgen over de impact op het milieu, aangezien ongeveer 70 % van alle elektriciteit door 1 of meerdere vermogenstransistoren stroomt. Welk een besparing van elektrisch vermogen is er niet mogelijk indien deze gigantische hoeveelheid energie op een efficiëntere manier kan worden beheerst?

Hoeveel vermogen? Welke technologie? Welke bouwstenen?

De vorige paragraaf werpt een blik op wat er bestaat in de elektronica: van de CMOS-technologie met de kleine, extreem vlugge transistor tot de vermogenelektronica met de thyristor, traag en extreem groot, maar in staat om duizenden volts te blokkeren en duizenden ampères te controleren. Het spreekt voor zich dat de vermogentechnologie een brede waaier van toepassingen heeft.

Net als de meeste doctoraten, situeert ook dit werk zich in een klein deel van deze brede waaier, namelijk op de grens tussen de CMOS-technologie en de vermogentechnologie; de intelligente vermogentechnologie. In deze vermogentechnologie zullen we ons richten op 1 van de twee klassen van vermogensbouwstenen: de verschillende schakelaars en niet op de klasse van de gelijkrichters (zie Hoofdstuk 2). Natuurlijk is het absurd te spreken over de integratie op chip van bv. de extreem grote thyristor zoals eerder vermeld, die op zich een volledige wafer in

beslag neemt (het is onvermijdelijk een discrete bouwsteen, d.w.z. een bouwsteen die niet met andere bouwstenen op een chip is geïntegreerd). Het is duidelijk dat die bouwstenen die geïntegreerd worden op chip een beperkende blokkeerspanning en stroomniveau hebben en dat niet alle bestaande bouwstenen in aanmerking komen voor integratie omdat ze voor welbepaalde, extreme toepassingen werden bedacht.

Dit brengt ons bij het probleem van de vergelijkbaarheid van de verschillende vermogenstransistoren. Dit is belangrijk omdat er een maatstaf nodig is om de efficiëntste transistor te bepalen. Deze maatstaf zal niet enkel afhangen van de spannings- en stroomniveaus, maar ook andere criteria zijn mogelijk. De kwaliteit van de MOS-vermogenstransistoren wordt meestal bepaald door de specifieke aan-weerstand versus de doorslagspanning. Maar wanneer MOS-vermogenstransistoren vergeleken worden met IGBTs dan zullen andere parameters genomen moeten worden aangezien de IGBT een exponentiële stijging van de stroom kent nadat een drempelspanning wordt overschreden. Bovendien zijn er soms andere criteria van tel zoals het bereik van de transistor in de aan-toestand, de afhankelijkheid van de temperatuur, het verval (*degradation* in het Engels) van de transistor... Deze bouwstenen kunnen ook gebruikt worden om speciale redenen (bv. als zekeringen tegen elektrostatische ontlading) met specifieke eigenschappen en kwaliteiten als gevolg.

Soms bepalen de toepassingen de eigenschappen van de technologie; bv. wanneer chips in een omgeving met hoge temperaturen dienen te werken, zal de silicium op isolator (SOI) technologie verkozen worden. Dit heeft een belangrijk gevolg voor het ontwerp van de vermogenstransistoren aangezien de isolatie nu diëlektrisch i.p.v. met behulp van sperlagen gebeurt. In een junctiegeïsoleerde technologie (zoals in dit werk) gebeurt de isolatie van de verschillende transistoren van elkaar namelijk door het handig gebruik van deze sperlagen. Sommige van de transistoren moeten op een potentiaal staan die hoger is dan die in de omgeving (op de chip). Deze zogenaamde zwevende transistoren hebben extra isolatie lagen nodig die soms moeilijk te realiseren zijn. Dit is trouwens een van de redenen waarom IGBT-transistoren moeilijk te integreren zijn in een junctiegeïsoleerde CMOS-technologie, maar daarover later meer.

Het is dus duidelijk dat een volledig overzicht nodig is van wat er precies nodig is voordat de CMOS-technologie wordt uitgebreid met extra processtappen om de vermogenstransistoren te maken. Niet alleen de verschillende beoogde toepassingen moeten gekend zijn, ook de factor kost kan een rol spelen daar voor sommige applicaties verschillende

realisaties mogelijk zijn. Dit is echter iets dat hier niet zal worden bestudeerd. We beperken ons tot het vermelden van het feit dat de basistechnologie waarop verder gewerkt moet worden, een 0.35 µm standaard-CMOS, junctiegeïsoleerde technologie is, dat het toepassingsgebied voornamelijk de auto industrie is, dat het spanningsbereik ruwweg tussen de 10 en de 100 V ligt, dat het stroombereik alles onder de 1 A is en dat de schakeltijden zelden sneller dan 10 ns zijn. Dit beperkt het aantal vermogenstransistoren die in aanmerking komen voor integratie tot de MOS-vermogenstransistoren (zie onder andere [Bal96], Hoofdstuk 10). Niettemin zal ook de integratie van de IGBT worden onderzocht.

Waarom TCAD?

Het grootste voordeel van Technologie CAD (TCAD) is dat men een bouwsteen kan creëren zonder het ook effectief te moeten maken. Men kan verschillende concepten uitproberen en TCAD voorspelt welke ideeën haalbaar zijn en welke niet. Een ander groot voordeel van TCAD is dat het inzicht geeft in de 2D-distributie (zelfs in 3D) van fysische grootheden. Men kan effectief *in* een bouwsteen kijken en zien wat er gebeurt wanneer deze of gene spanning aangelegd wordt. Dit heeft reeds meer dan eens geholpen bij het oplossen van problemen in bestaande transistoren. TCAD wordt vaak gebruikt in de literatuur om problemen van allerlei aard uit te leggen, te analyseren en te begrijpen. Een ander voordeel van TCAD is kost. Eens het standaardproces gekalibreerd is in TCAD, kan men met grote betrouwbaarheid nieuwe transistoren ontwikkelen, zelfs indien één of meerdere nieuwe procesmodules gedefinieerd moeten worden. Transistoren kunnen worden ontwikkeld zonder een dure tweede of derde poging, wat de ontwikkelingskosten sterk reduceert. Zelfs de extractie van SPICE-parameters kan al gebeuren voordat de echte transistoren er zijn (wat effectief ook gebeurd is voor de pLDE-MOS, die besproken wordt in dit werk).

Een conditie sine qua non is dat de TCAD-simulaties betrouwbaar zijn. Het nog bestaande scepticisme in de industrie jegens TCAD is hoofdzakelijk tweeledig. Het eerste probleem is de simulatie van 2D-doperingsprofielen wegens het gebrek aan 2D-ijkmateriaal. Het tweede probleem is de constante evolutie van de transistoren naar kleinere dimensies, waardoor de fysische modellen ook dienen mee te evolueren, wat niet altijd het geval is. Beide bezwaren zijn echter niet van toepassing op het werk dat in dit boek wordt gepresenteerd. Eerst en vooral gebeurt de integratie van vermogenstransistoren in een technologie die

reeds goed gekend is (de intelligente vermogentechnologie hinkt verschillende generaties achter op de informatieverwerkende technologie). Ten tweede zijn de vermogenstransistoren van nature groter in dimensie dan hun digitale tegenhangers, waardoor de nood aan fijne 2D-ijking niet zo hoog is. Niettemin blijven ijking en de verwante numerieke problemen een belangrijk punt. Daarom wordt er ook een volledig hoofdstuk aan gewijd (Hoofdstuk 3).

Doelstelling

Het doel van dit werk is vermogenstransistoren te ontwerpen en te integreren in een bestaande standaard-CMOS-technologie en nieuwe concepten te bedenken om de efficiëntie van deze transistoren te optimaliseren. De criteria die gebruikt worden, zullen te zijner tijd worden verklaard en zijn hoofdzakelijk heel eenvoudig: de doorslagspanning versus de specifieke aan-weerstand of versus gedissipeerd vermogen. Een belangrijk gegeven dat telkens voorkomt in deze parameters is oppervlakte, welke opnieuw de factor kost is die meespeelt. Hoe kleiner een bouwsteen, hoe minder silicium er wordt gebruikt, hoe kleiner en goedkoper de chip. Om dit te bekomen, moet de vermogensdissipatie tot een minimum worden herleid anders zou te veel warmte op een te kleine oppervlakte gegenereerd worden, wat onherroepelijk tot schade leidt. Indien we de vermogensdissipatie verminderen, wil dit ook zeggen dat we het verlies van energie inperken. Aangezien 60 tot 70 % van alle energie door één of meerdere vermogenstransistoren stroomt, betekent dit een efficiëntere controle over de elektrische energie. Of hoe een hoofdzakelijk door kost gedreven motivatie kan leiden tot een ecologisch verantwoorde trend...

Overzicht

Aangezien de focus van dit werk is het begrijpen, analyseren en ontwerpen van vermogenstransistoren met behulp van TCAD, zullen we ons het meest concentreren op de simulatie van de werking van de verschillende bouwstenen. Daarom geeft Hoofdstuk 2 een overzicht van de fysica nodig voor deze simulaties. Voor een behandeling van de proces- en halfgeleidersfysica wordt verwezen naar standaardwerken. Er wordt in Hoofdstuk 2 echter een overzicht gegeven van deze fenomenen die zo belangrijk zijn dat ze niet kunnen ontbreken in een werk over vermogenstransistoren. Over deze vermogenstransistoren bestaan enkele uitstekende werken die de belangrijkste werkingsprincipes en eigenschappen

verklaren aan de hand van analytische modellen. Maar, zoals is geschreven in [Bal96, Voorwoord, p. viii] (eigen vertaling):

> "Voor een volledig karakterisering van de elektrische eigenschappen van transistoren zijn numerieke technieken die gebruik maken van computerprogramma's broodnodig. Deze programma's kunnen de fundamentele transportvergelijkingen oplossen in 2 dimensies (en soms in 3 dimensies), met tijdsafhankelijkheid indien nodig."

Dit is in een notendop wat gedaan zal worden in de overige hoofdstukken. Maar vooraleer we daarmee beginnen, moeten we het belangrijk punt van de TCAD-simulatie en -ijking behandelen (Hoofdstuk 3). Hoofdstuk 4 en 5 bestuderen dan respectievelijk de vermogen-MOS en de IGBT. Conclusies worden genomen in Hoofdstuk 6 door het vergelijken van de verschillende transistoren met elkaar en met voorbeelden uit de literatuur.

2 Fundamentele Beschouwingen

2.1 Inleiding

Het is niet de bedoeling van dit boek om een overzicht te geven van alle proces-, halfgeleiders- en transistorfysica nodig voor TCAD-simulaties. We beperken ons tot een verwijzing naar de belangrijkste standaardwerken. Niettemin wordt er een korte schets gegeven van de fysica nodig voor de simulatie van de werking van de bouwstenen. Dit wordt niet gedaan voor de procesfysica daar de klemtoon in dit boek op de werking van de transitoren wordt gelegd.

Vooraleer we beginnen te werken op vermogenstransistoren, wordt er een definitie gegeven van de vermogensbouwsteen, die in 2 klassen wordt opgedeeld: de gelijkrichters en de schakelaars. De gelijkrichters worden slechts kort behandeld met de nadruk op hun doorslagspanning versus doperingsniveau. Dit o.w.v. het feit dat deze eigenschappen gebruikt zullen worden bij het ontwerp van vermogenstransistoren. Dan worden de schakelaars besproken, bestaande uit een classificatie op basis van het type gate. De laatste paragraaf somt een aantal minder bekende fenomenen op die voorkomen bij de studie van geïntegreerde schakelingen. De meeste van deze onderwerpen komen uitgebreid aan bod in de verschillende hoofdstukken over de vermogenstransistoren.

2.2 De fysica

Voor de procesfysica refereren we naar [Sze88] en [CS96], voor de halfgeleidersfysica naar [Wan66] en [Sze81] en voor de transistorfysica in het algemeen naar dit laatste werk en voor de fysica van de vermogenstransistoren in het bijzonder naar [Gha77], [Bal92], [Bal96] en [BGG99].

De meeste simulaties van transistoren in dit boek maken gebruik van wat algemeen bekend staat als het drift-diffusiemodel. De continuïteitsvergelijkingen

$$\frac{\partial n}{\partial t} = G_n - R_n + \frac{1}{q}\nabla \mathbf{J}_n \tag{1}$$

$$\frac{\partial p}{\partial t} = G_p - R_p + \frac{1}{q}\nabla \mathbf{J}_p \tag{2}$$

met n en p de elektronen- en gatendichtheid, G_n en G_p de elektronen- en gatengeneratiesnelheid en R_n en R_p de elektronen- en gatenrecombinatiesnelheid, worden opgelost met behulp van de transportvergelijkingen voor respectievelijk de elektronendichtheidsstroom $\mathbf{J}_n$ en de gatendichtheidsstroom $\mathbf{J}_p$:

$$\mathbf{J}_n = q\mu_n n\mathbf{E} + qD_n\nabla n \tag{3}$$

$$\mathbf{J}_p = q\mu_p p\,\mathbf{E} - qD_p\nabla p, \tag{4}$$

waarin q de elementaire ladingseenheid is, $\mathbf{E}$ het elektrisch veld, μ_n en μ_p de elektronen en gatenmobiliteit en D_n en D_p de diffusieconstanten van respectievelijk de elektronen en de gaten.

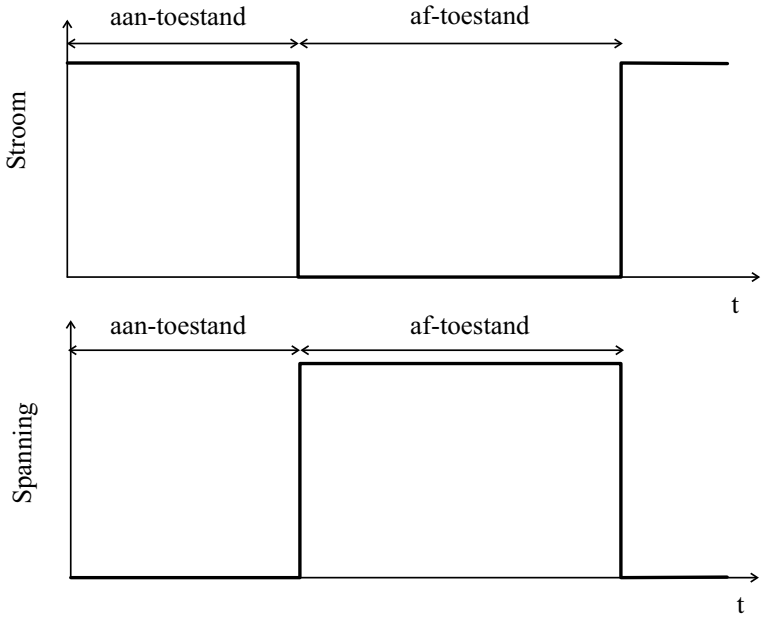
Deze vergelijkingen geven samen met de vergelijkingen van Maxwell een volledige beschrijving van de dynamica van elektronen en gaten in een halfgeleider onder invloed van externe velden. Voor de bouwstenen bestudeerd in dit boek, wordt enkel de vergelijking van Poisson gebruikt.

Soms is het nodig de temperatuur van het rooster in rekening te brengen, met als gevolg dat de uitdrukkingen voor de elektronen- en gatendichtheidsstroom wijzigen, en tevens de warmtetransportvergelijking wordt opgelost. Dit wordt het thermodynamisch model genoemd.

Voor heel kleine bouwstenen volstaat zelfs deze benadering niet en worden ook nog de temperaturen van elektronen en gaten afzonderlijk in rekening gebracht. Dit model staat bekend als het hydrodynamisch model, omdat het stelsel van vergelijkingen analoog is aan dit uit de vloeistoffysica. Dit tijdrovend model wordt echter maar zelden gebruikt bij het ontwerpen van vermogenstransistoren.

2.3 Wat is een vermogensbouwsteen ?

Een vermogensbouwsteen controleert het vermogen dat aan een last wordt gegeven. Dit wordt meestal gedaan door de bouwsteen periodiek te schakelen zodoende stroompulsen te genereren. Het ideale stroom- en spanningsverloop worden getoond in figuur 1. Over de ideale vermo-



Figuur 1 Stroom- en spanningsverloop van een ideale vermogensbouwsteen.

gensbouwsteen staat geen spanning wanneer stroom wordt geleid (geen vermogensdissipatie tijdens de aan-toestand), vloeit er geen stroom in de af-toestand (geen vermogensdissipatie tijdens de af-toestand) en is de schakeltijd tussen af- en aantoestand en omgekeerd oneindig vlug (geen vermogensdissipatie tijdens het schakelen). De ideale vermogensbouwsteen verbruikt dus zelf geen vermogen en geeft alle vermogen aan de last (actief of reactief), die capacitief, resistief of inductief kan zijn. De vermogensbouwstenen worden in 2 klassen onderverdeeld: de gelijkrichters (diodes) en de schakelaars. Deze laatsten zijn in staat de hoeveelheid vermogen naar de last te controleren, de eersten zijn dat niet.

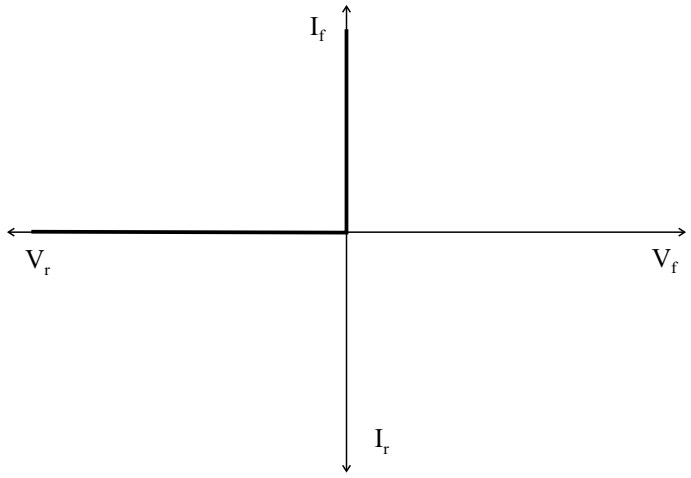
2.4 Gelijkrichters

Aangezien de gelijkrichters niet bestudeerd worden in dit boek, wordt maar heel kort een overzicht gegeven van de bestaande types. Niettemin

komen diodes in alle mogelijke vormen voor in de structuren van de schakelaars (en dit zeker in junctiegeïsoleerde technologieën, zoals in dit boek). Daarom worden de voor ons zo belangrijke karakteristieke doorslagmechanismen van de elementaire pn- en PT- (uit het Engels *punch-through*) diodes in een grafiek weergegeven.

De ideale gelijkrichter

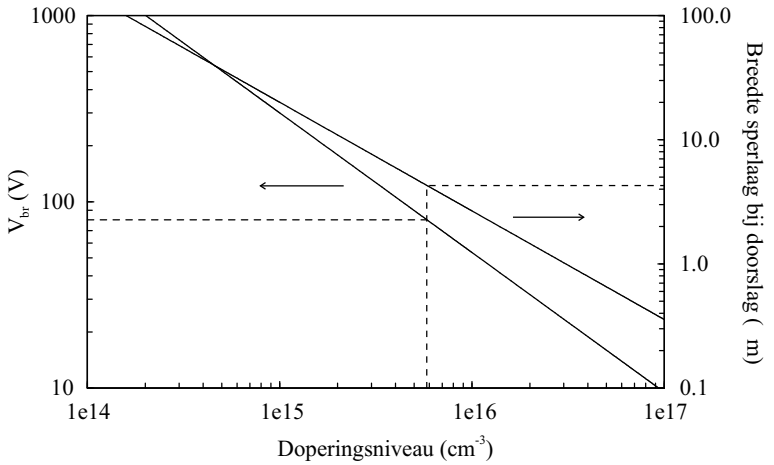
De ideale gelijkrichter blokkeert alle spanning zonder lekstromen in de af-toestand en geleidt zonder weerstand in de aan-toestand, hij is bovendien in staat tussen beide toestanden oneindig vlug te schakelen (figuur 2).



Figuur 2 Uitgangskarakteristiek van een ideale gelijkrichter.

De pn-diode

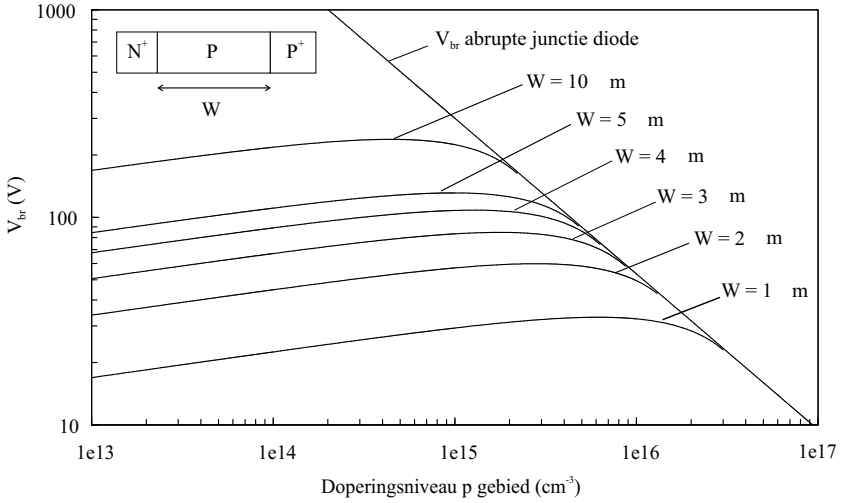
Een echte diode heeft eindige blokkerings-, geleidings- en schakeleigenschappen. Deze worden aangeduid met de volgende parameters: blokkeringsspanning V_{br}, voorwaartse spanningsval tijdens de aan-toestand V_{on}, schakeltijd van de af- naar de aan-toestand t_{on} en omgekeerd t_{off}. Bovendien is er een lekstroom tijdens de af-toestand, waardoor er ook vermogensdissipatie is tijdens de af-toestand. Als voorbeeld van een abrupte n^+p-diode wordt op figuur 3 een 80 V diode gegeven. Uit de figuur leidt men af dat het maximale doperingsniveau $5.8e15\,cm^{-3}$ is en dat de breedte van de sperlaag in dit geval 4.3 µm is.



Figuur 3 Doorslagspanning V_{br} en breedte van de sperlaag bij doorslag als functie van het doperingsniveau van het laag gedopeerd gebied van n^+p-diode.

De PT-diode

Indien in het vorige voorbeeld de sperlaag in het p-gebied bij een bepaalde spanning een p^+-gebied tegenkomt, dan stopt de sperlaag virtueel en groeit het elektrisch veld verder in het volledige p-gebied. Het gevolg is een lagere doorslagspanning (figuur 4).

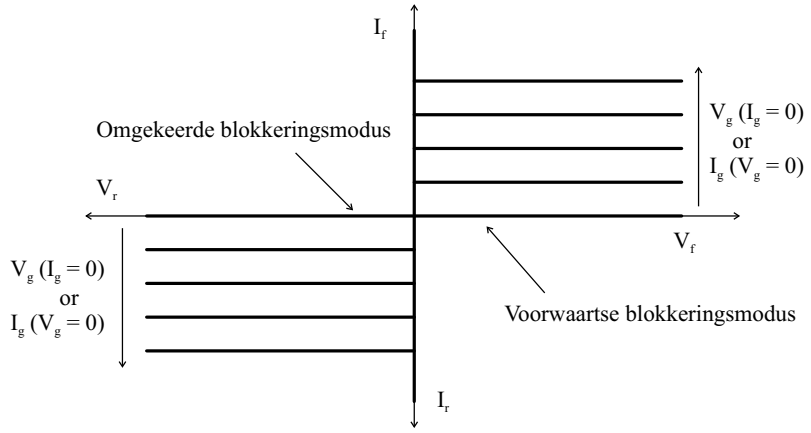


Figuur 4 V_{br} voor een abrupte junctie diode en voor PT-diodes met verschillende breedtes. De inzet toont een PT-diode.

Het belang van deze PT-diodes ligt in het feit dat men in een gebied met een kleinere dimensie als de breedte van de sperlaag bij een abrupte n^+p-diode, dezelfde spanning kan bekomen *indien men het doperingsniveau van het laag gedopeerd gebied verlaagt*. Als voorbeeld nemen we opnieuw een 80 V diode, waar bv. een breedte van 3 µm voldoende is indien men het doperingsniveau verlaagt tot $1.8e15\,\mathrm{cm}^{-3}$.

2.5 Schakelaars

De schakelaar heeft één contact meer als de diode, namelijk de poort of gate die de uitgangsstroom controleert. De ideale uitgangskarakteristieken van een schakelaar worden geschetst in figuur 5.



Figuur 5 Uitgangskarakteristieken van een ideale schakelaar.

Stroomgecontroleerde schakelaars

De bipolaire vermogenstransistor is de allereerste vermogensschakelaar ooit gemaakt. Zijn gebruik nam echter vanaf de jaren 70 af door de concurrentie van de MOS-vermogenstransistoren. De bipolaire vermogenstransistor wordt echter wel nog gebruikt in welbepaalde toepassingen.

De tweede oudste vermogensschakelaar die stroomgestuuurd is, is de thyristor. Deze schakelaar bestaat uit 4 lagen verschillend gedopeerd silicium en is de allerbeste geleider uitgedrukt in hoeveelheid stroom per oppervlakte-eenheid. Het grootste nadeel is dat de controle over de stroom door de gate wordt verloren bij grote stroomdichtheden. De thyristor wordt vooral gebruikt in omstandigheden waar extreem grote spanningen en stromen voorkomen.

Spanningsgecontroleerde schakelaars

De sperlaag-veldeffecttransistor (JFET of SIT) is een schakelaar die stroom geleidt wanneer er geen signaal op de gate is, wat maakt dat deze schakelaar maar in welbepaalde toepassingen wordt gebruikt.

De MOS-vermogenstransistor kwam er als gevolg van de successen van de digitale CMOS-transistoren, de nMOS en de pMOS. Deze MOS-vermogenstransistoren hebben hun bipolaire tegenhangers in vele toepassingsgebieden uit de markt geconcurreerd. Dit komt omdat de MOS-vermogenstransistor een hogere ingangsimpedantie heeft, sneller is wegens de unipolariteit, de aan-weerstand een negatieve temperatuursafhankelijkheid heeft (waardoor het veilig wordt om deze transistoren in parallel te plaatsen) en het een groot spanningsbereik (SOA) heeft.

De bipolaire transistor met geïsoleerde gate (IGBT) is een schakelaar die de beste eigenschappen van de bipolaire en van de MOS-vermogensschakelaars probeert te combineren: een lage voorwaartse spanningsval bij geleiding en een hoge ingangsimpedantie. Jammer genoeg erft heeft hij een kleinere SOA, is hij trager dan de MOS en heeft hij een positieve temperatuursafhankelijkheid.

Een laatste type van spanningsgecontroleerde schakelaars is de MOS-gecontroleerde thyristor, die net als de thyristor meestal voorkomt in toepassingen voor extreme omstandigheden.

2.6 Fundamentele concepten omtrent geïntegreerde vermogensschakelaars

De bedoeling van deze paragraaf is een overzicht te geven van de belangrijkste concepten die gebruikt worden bij het ontwerpen van vermogenstransistoren in geïntegreerde schakelingen. De meeste van deze concepten komen uitgebreid aan bod in de verschillende hoofdstukken over de vermogenstransistoren.

De siliciumlimiet

De siliciumlimiet is de analytische uitdrukking voor de specifieke aanweerstand $R_{on,sp}$ in functie van de doorslagspanning V_{br} voor een ideale MOS-vermogenstransistor. Indien beide parameters voor een bestaande transistor worden uitgezet in een grafiek samen met de siliciumlimiet en samen met de parameters van andere bestaande transitoren van concurrenten, dan kan men de beste transistor voor een bepaalde spanning in één oogopslag eruit halen (nl. deze die het dichtst bij de limiet ligt).

Het dient vermeld te worden dat er verschillende siliciumlimieten bestaan, nl. voor de verschillende vormen van de MOS-vermogenstransistor (verticaal, lateraal, RESURF, SOI, superjunctie of COOLMOS^{TM}...).

RESURF-effect

Het RESURF-effect (uit het Engels *RE*duced *SUR*face *F*ield) is een 2D-techniek waarbij de distributie van de elektrische velden op een optimale manier gespreid worden. Het is één van de meest gebruikte technieken bij het ontwerp van vermogensbouwstenen. Aanvankelijk werd het gebruikt in diodes, maar tegenwoordig wordt het in bijna elke vermogenstransistor en op verschillende manieren (met verschillende lagen in de zogenaamde superjunctietransistoren, op SOI, in 3D...) toegepast.

PT- en RT-doorslag

PT- en RT- (uit het Engels *Reach-Through*) doorslag zijn verschillende fenomenen. PT-doorslag komt voor in een PT-diode wanneer het kritische elektrische veld wordt bereikt aan de pn-junctie. RT-doorslag daarentegen komt voor in een pnp- of npn-structuur, waarbij de sperlaag komende van één van de juncties de andere raakt. Dan wordt een pad gecreëerd die geleiding veroorzaakt. Dit mechanisme kan dus voorkomen lang vòòr dat het kritische elektrische veld wordt bereikt. Niettemin worden beide termen in de literatuur vaak door elkaar gebruikt.

Tweede doorslag, terugslag en thermische instabiliteit

Vroeger gebruikte men de term tweede doorslag wanneer men thermische instabiliteit beschreef (bv. in [Gha77]), maar tegenwoordig wordt de term meestal gebruikt om een plotse reductie van de doorslag in de aan-toestand (i.v.m. de doorslag in de af-toestand) aan te duiden. Deze tweede doorslag kan gepaard gaan met een terugval in de spanning, met een negatieve weerstand als gevolg. Dit is wat men in het Engels *snap-back* (terugslag) noemt. De term thermische instabiliteit wordt enkel nog gebruikt bij temperatuursfenomenen van destructieve aard.

Kirk-effect en adaptieve RESURF

Het Kirk-effect wordt normaal beschreven in een bipolaire transistor [Sze81, p.145]. Het komt echter ook voor in MOS-vermogenstransistoren, en in het bijzonder in laterale structuren die gebruik maken van het RESURF-effect. Het gevolg is een afname van het spanningsbereik in

de aan-toestand. Een mogelijke oplossing is wat men noemt "adaptieve RESURF", waarmee de verschuiving van het elektrisch veld naar de drain toe wordt verholpen door een verhoging van het doperingsniveau in de buurt van het draincontact.

SOA

De SOA (staat voor *Safe Operating Area*) is gedefinieerd als het spanningsgebied waarbinnen de schakelaar op een veilige manier kan gebruikt worden. Dit is een vage definitie en men onderscheidt dan ook drie verschillende SOAs:

- *Elektrische SOA en ESD*

  Een schakelaar kan gedurende een heel korte tijd (ns tot µs) gebruikt worden in het bereik van de uitgangskarakteristiek na de doorslag en na de terugslag. Bepaalde schakelaars worden zelfs speciaal ontworpen om deze eigenschap te benadrukken. Deze worden bijvoorbeeld gebruikt in ESD (E*lectro*S*tatic* D*ischarge*) protectiecircuits. Deze bouwstenen worden hier niet besproken.

- *Thermische SOA en energetisch opvangvermogen*

  Wanneer de stroompulsen langer duren (µs tot ms) dan worden temperatuurseffecten belangrijk, waardoor er gevaar op thermische instabiliteit ontstaat. Sommige bouwstenen worden speciaal ontworpen om een zo groot mogelijke energiestoot op te vangen, bv. bij het uitschakelen van een inductieve last. Ook dit soort van bouwstenen wordt niet besproken in dit boek.

- *Hetelading-SOA en degradatie*

  Wanneer een schakelaar niet wordt gebruikt in de extreme omstandigheden zoals in beide bovenstaande voorbeelden, dan is de SOA meestal nog kleiner dan de elektrische en thermische SOA. Zeker in schakelaars met een MOS-gate is dit het geval, waar de ladingen die door het kanaal stromen na verloop van tijd het oxidelaagje vervuilen. Deze vervuiling veroorzaakt verschuivingen in de elektrische eigenschappen van de transistor. Dit is wat men noemt degradatie, die gekarakteriseerd wordt op basis van stressmetingen. Het toegelaten spanningsbereik wordt dan bepaald aan de hand van extrapolatie van de meest verslechterende parameter (en dit kan verschillend zijn naargelang de spanningen op de verschillende contacten).

Hoge injectie en geleidingsmodulatie

Injectie van minoritairen in een laag gedopeerd p- of n-gebied kan zo hoog zijn dat de elektronen en gaten in een hogere concentratie aanwezig zijn dan het oorspronkelijk doperingsniveau. De weerstand in dit gebied vermindert dan aanzienlijk, i.e. geleidingsmodulatie. Dit fenomeen zorgt ervoor dat in vele vermogensbouwstenen een lage voorwaartse spanningsval in de aan-toestand leidt tot hoge stroomniveaus.

Isolatie

Het is goed mogelijk dat men een geïntegreerde vermogensbouwsteen ontwerpt met een interne doorslagspanning van bijvoorbeeld meer dan 80 V, maar dat er toch problemen optreden bij lagere spanningen. Dit komt omdat men ervoor moet zorgen dat de elektrische isolatie van deze geïntegreerde bouwsteen op zijn minst de doorslagspanning ervan moet aankunnen. Verder moeten sommigen geïntegreerde bouwstenen in hun geheel (d.w.z. op alle contacten tegelijkertijd) op een spanning staan die hoger is dan de spanningen in de directe omgeving. Dit noemt men de zwevende bouwstenen en deze hebben vaak een isolatiestructuur die heel wat uitgebreider is dan die van hun niet-zwevende tegenhangers.

3 TCAD-Simulatie en -IJking

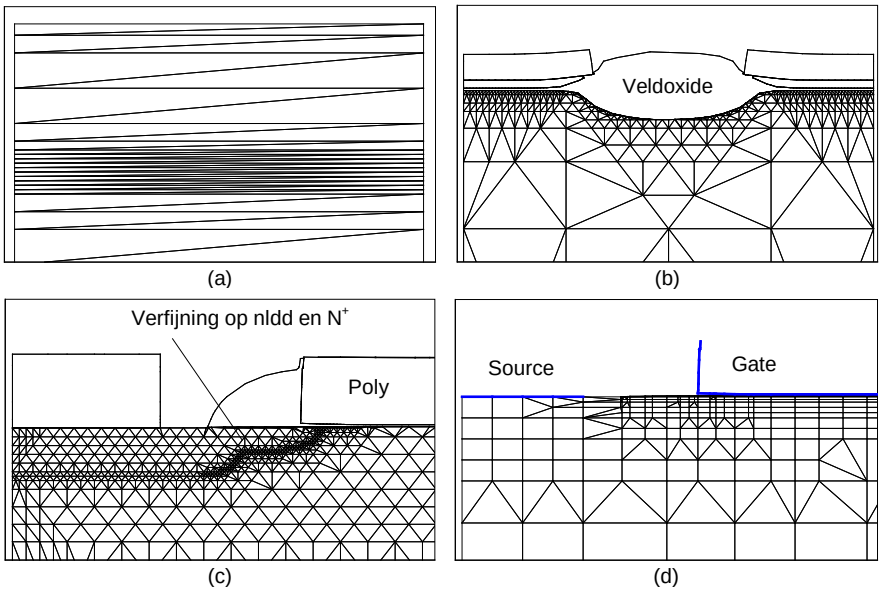
3.1 Inleiding

Technologie-CAD (TCAD) gebruikt fysische modellen ten einde een volledig productieproces stap voor stap, en daarna de werking van de gesimuleerde bouwstenen, te simuleren. Deze proces- en bouwsteensimulatoren brengen de eigenschappen van deze bouwstenen op een niet-uniform, discreet raster van punten aan (in 1, 2 of zelfs 3 dimensies) om de differentiaalvergelijkingen die de verschillende fysische processen beschrijven numeriek op te lossen. Dit hoofdstuk bestaat uit twee delen, het eerste beschrijft de ijking en rasterproblematiek in de processimulator, het tweede doet dit voor de bouwsteensimulator.

3.2 Processimulatie en -ijking

Het raster

Het raster zorgt voor het grootste deel van de problemen gedurende TCAD-werk. De gebruiker wil een raster dat zo ruw mogelijk is om

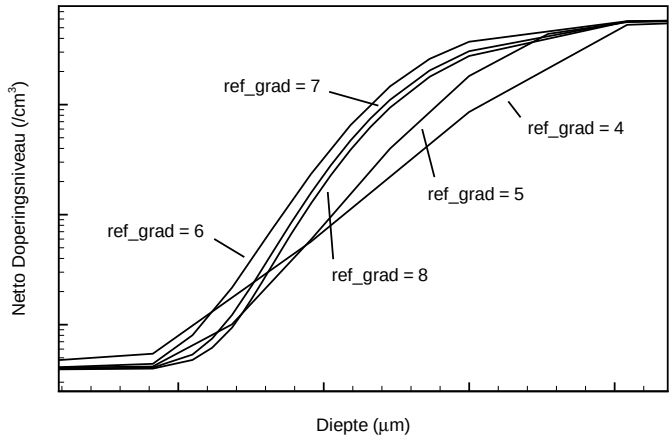


Figuur 6 Evolutie van het raster doorheen de simulatie: (a) 1D in proces simulatie, (b) eerste proces stap in 2D, (c) op het einde van het proces en (d) nieuw raster voor simulatie van de werking van de transistor.

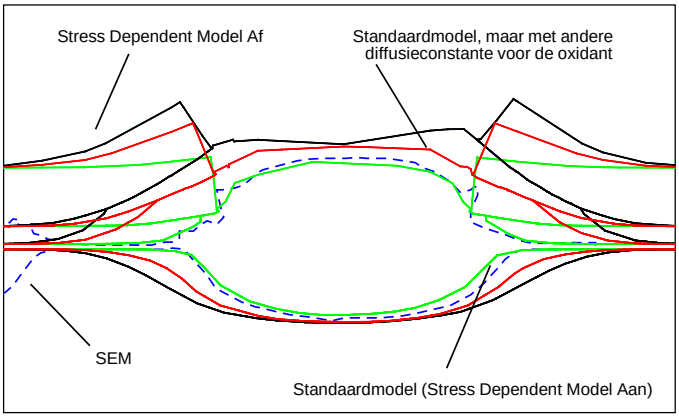
de simulatietijd te beperken, maar het moet wel fijn genoeg blijven om correcte resultaten af te leveren.

Ter illustratie van deze problematiek zal een simpele nMOS gesimuleerd worden als leidraad doorheen dit hoofdstuk. Omdat de simulaties gebaseerd zijn op het bestaande 0.35 µm CMOS proces van AMIS, worden de assen van vele figuren weggelaten door de confidentiële aard ervan. Deze werkwijze ondermijnt het doel van deze discussie geenszins.

In het begin van de simulatie is het raster nog in 1 dimensie (d.w.z. een verzameling lijnen in een rechthoek). Wanneer het nodig wordt de tweede dimensie mee in rekening te brengen (bij de definitie van een masker), dan schakelt de simulator automatisch over op 2 dimensies. Het raster wordt enkel verfijnd daar waar nodig (aan een junctie, gradiënt in doperingsniveau, oneffenheden op de Si/SiO_2 grens...) met behulp van welbepaalde parameters (`Refinejunction`, `RefineGradient`, `RefineBoundary`...) uit de ISE-software. Deze verfijning gebeurt door de definitie van rechthoeken in het simulatiedomein gedurende de simulatie (bijvoorbeeld net voor een implantatie), zie figuur 6 voor enkele voorbeelden. Een voorbeeld van de invloed van een dergelijke parameter op het simulatieresultaat wordt gegeven in figuur 7.



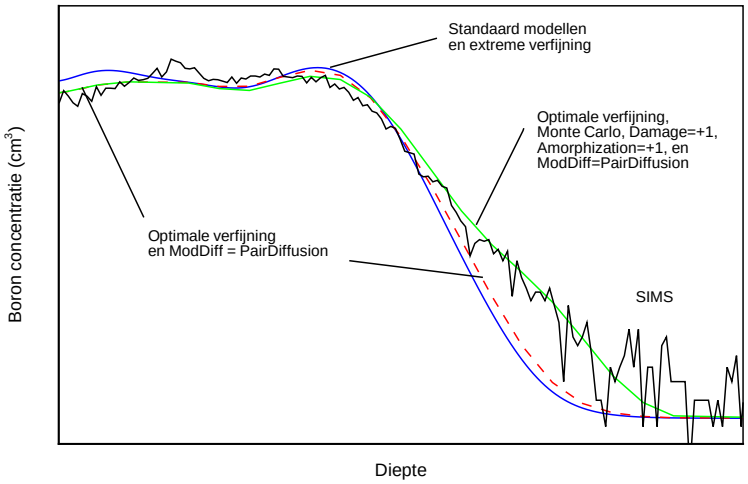
Figuur 7 Invloed van de verfijning van het raster (Refgrad) op een doperingsprofiel.



Figuur 8 Vergelijking van verschillende oxidatiemodellen met een SEM-foto.

Simulatie en ijking

Het zou ideaal zijn indien we zouden beschikken over ijkmateriaal (SIMS, SEM...) na elke gesimuleerde processtap. Jammer genoeg is dit niet haalbaar vanwege de kost en moeten we ons beperken tot enkele belangrijke stappen. Zoals gezegd in het inleidend hoofdstuk, beschikken we niet over 2D-doperingsprofielen en werken we dus steeds met SIMS-profielen in 1D. De veldoxidatie kunnen we wel kalibreren met behulp van SEM-foto's in 2D. Daartoe zijn verschillende oxidatiemodellen in de software opgenomen (met elk hun eigen verzameling modelparameters). In figuur 8 kan men zien dat het standaardmodel het veldoxide nagenoeg correct simuleert.



Figuur 9 Vergelijking van het SIMS-profiel van de pwell na gate-oxidatie met verschillende gesimuleerde profielen.

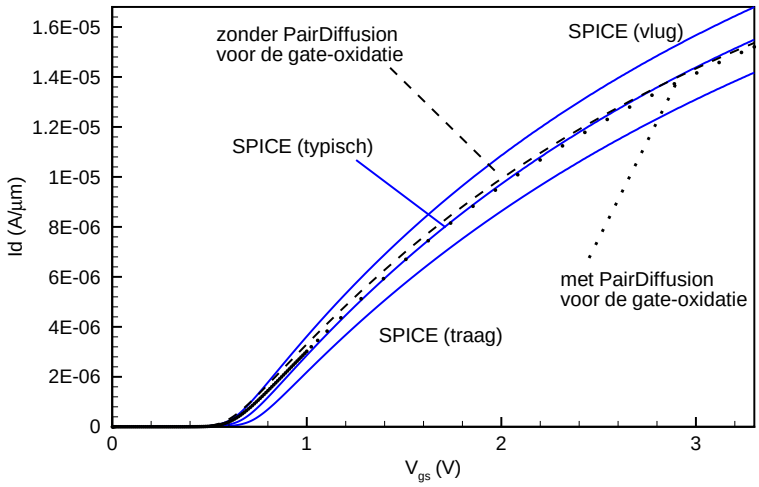
De SIMS worden genomen na de eerste temperatuurstap die volgt op de implantatie of op het einde van het proces (d.w.z. na de laatste hoge temperatuurstap, na dewelke de profielen niet meer veranderen). Ook hier zien we de invloed van de verschillende implantatiemodellen (standaard, i.e., m.b.v. analytische uitdrukkingen; ofwel met Monte Carlo simulatie) in combinatie met de verschillende diffusiemodellen (indien bv. schade aan het rooster in rekening wordt gebracht met het model `Damage = +1`, zie figuur 9).

Alle overige processtappen (maskers, etsen, deposities) gebeuren geometrisch en hun commando's zijn dan ook veel minder uitgebreid dan deze voor de implantatie-, de oxidatie- en de diffusieprocessen. Het einde van de processimulatie wordt bereikt bij de laatste hoge temperatuurstap in het productieproces.

3.3 Bouwsteensimulatie en -ijking

Van proces- naar transistorsimulatie

Vooraleer we de werking van de nMOS-transistor simuleren, moeten we de structuur een nieuw raster geven. Dit is broodnodig omdat het raster dat gebruikt wordt voor de processimulatie niet fijn genoeg is in het ene gebied en dan weer te fijn in een ander voor de transistorsimulatie. Een voorbeeld daarvan zien we in figuur 6. Onzichtbaar in deze figuur is het ultra fijne rooster ($\sim 2\,\mathrm{nm}$) in het kanaal, dat nodig is voor de transistor



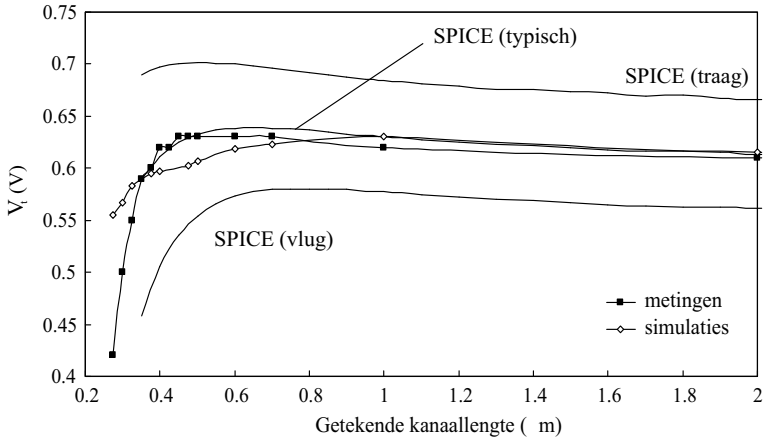
Figuur 10 $I_d(V_{gs})$ voor een nMOS met getekende kanaallengte van 2 µm: SPICE-simulaties versus TCAD-simulaties.

simulatie (figuur 6, (d)), maar dat overbodig is voor de simulatie van het pwell-profiel gedurende de processimulatie (figuur 6, (c)).

Transistorsimulatie en -ijking

De transistorsimulatie eist veel minder ijkwerk aangezien de werking van een transistor idealiter fabrieksonafhankelijk is. De belangrijkste uitzondering hierop is verbonden aan het productieproces, nl. de werkfunctie van de gate. Deze dient gekalibreerd te worden m.b.v. de gemeten V_t waarden.

Indien de gemeten elektrische karakteristieken van een transistor verschillend zijn dan de gesimuleerde, dan dient in de eerste plaats gekeken te worden naar eventuele fouten tijdens de processimulatie. Een voorbeeld hiervan wordt gegeven in figuur 10, waar SPICE-simulaties worden vergeleken met TCAD-simulaties. Het is duidelijk te zien dat door gebruik te maken van een ander model voor de gate-oxidatie, er een invloed is op de transferkarakteristiek. Niettemin is deze invloed erg klein en dient het complexere, meer tijdrovend oxidatiemodel enkel gebruikt te worden indien dit vereist is. Dit is bijvoorbeeld het geval indien men korte-kanaalseffecten wil bestuderen. In figuur 11 ziet men dat men de correcte V_t waarde niet kan simuleren voor kleine kanaallengtes door enkel gebruik te maken van de standaardmodellen, zoals werd gedaan voor de TCAD-simulaties op deze figuur.



Figuur 11 V_t in functie van getekende kanaallengte: vergelijking tussen metingen, SPICE- en TCAD-simulaties.

3.4 Besluit

In dit hoofdstuk werd in de processimulatie en in de transistorsimulatie enkel gebruik gemaakt van de standaardmodellen met een minimum aan ijkwerk. Indien men problemen als de korte-kanaalseffecten wenst te simuleren, dan zal men ook complexere modellen tijdens de processimulatie dienen te gebruiken, met een navenante stijging van de hoeveelheid ijkwerk als gevolg. Voor het werk dat gepresenteerd wordt in dit boek, zal dit echter nooit het geval zijn aangezien we in de eerste plaats eerder grote transistoren simuleren en in de tweede plaats TCAD zullen gebruiken als hulpmiddel bij het onderzoeken van trends in proces- en lay-outvariaties en bij het uitproberen van nieuwe ideeën en concepten. Net zoals in de literatuur, zullen we ook hier TCAD gebruiken bij het analyseren, begrijpen en voorspellen van problemen en trends, maar niet voor het voorspellen van absolute, exacte waarden.

4 Vermogen-MOS

4.1 Inleiding

Eén van de eerste referentiewerken over vermogensbouwstenen, geschreven in 1977 door S.K. Ghandhi [Gha77], behandelt de MOS-vermogenstransistor niet en spreekt enkel over de bipolaire vermogenstransistor en de thyristor. Daartegenover staat dat één van de meest recente standaardwerken, geschreven in 1996 door B.J. Baliga [Bal96], de bipolaire

vermogenstransistor enkel behandelt als inleiding tot een bepaald type MOS-vermogenstransistor, nl. de IGBT. Dit illustreert de evolutie van de vermogenstransistoren over een tijdspanne van 20 jaar en toont het belang aan van de introductie van de MOS-gate in deze bouwstenen. De stroomgestuurde gate en bijhorende complexe ingangscircuiten voor de bipolaire transistoren werd vervangen door een spanningsgestuurde gate met een veel eenvoudiger ingangscircuit. De MOS-transistor heeft bovendien een veel vluggere schakelsnelheid (wegens zijn unipolariteit), heeft een negatieve temperatuursafhankelijkheid (hoe warmer, hoe minder stroom wordt geleid, wat ideaal is om deze transistoren in parallel te plaatsen), en is veel minder onderhevig aan tweede doorslag.

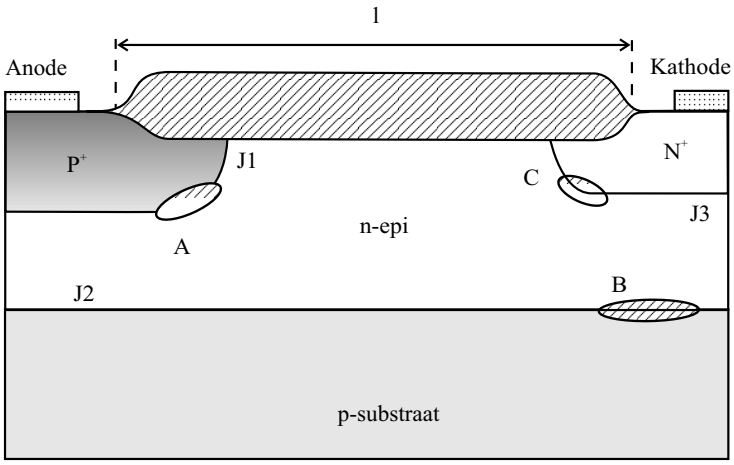
Omdat de MOS-gate zijn oorsprong kent in de digitale CMOS-technologieën, is de eerste MOS-vermogenstransistor, een soort van uitgebreide MOS of DEMOS (*Drain Extended* MOS). Daarna werd de dubbelediffusie-MOS (DMOS) uitgevonden, zo genoemd omdat het kanaal samen met de source-gebieden worden geïmplanteerd en gediffundeerd na de depositie van de poly-gate. De volgende belangrijke stap was de introductie van het RESURF-effect, dat behandeld wordt in de volgende paragraaf. Het is één van de meest belangrijke technieken voor het ontwerpen van vermogensbouwstenen geworden (niet enkel MOS-transistoren).

De zeer belangrijk siliciumlimiet, die dient als een waardemeter voor de MOS-vermogenstransistoren wordt besproken in het volgende luik. Hierop volgt een korte uiteenzetting over de verschillende vormen en types van de MOS-vermogenstransistoren, die daarna elk op hun beurt bestudeerd worden. De conclusie vergelijkt de bekomen MOS-vermogenstransistoren met deze gevonden in de literatuur.

4.2 RESURF-effect

Het RESURF-effect werd bij toeval ontdekt in 1979 [AV79] bij de studie van diodes (figuur 12).

Deze basisstructuur bestaat uit 2 diodes: een verticale diode (n^+–n-epi–p-substraat) en een laterale diode (p^+–n-epi–n^+). De doorslagspanning in een dergelijke structuur is een samenspel tussen verschillende 2D-effecten. Samenvattend kunnen we stellen dat het RESURF-effect erop neerkomt dat de doorslag (i.e., bij een negatief spanningsverschil tussen anode en kathode) niet gebeurt op plaats A in figuur 12, wat normaal gezien gebeurt wanneer de epilaag dik genoeg is in vergelijking met de lengte l; maar dat het elektrisch veld vanaf een bepaalde spanning terzelfder tijd op de plaatsen A, B en C zal groeien, waar-



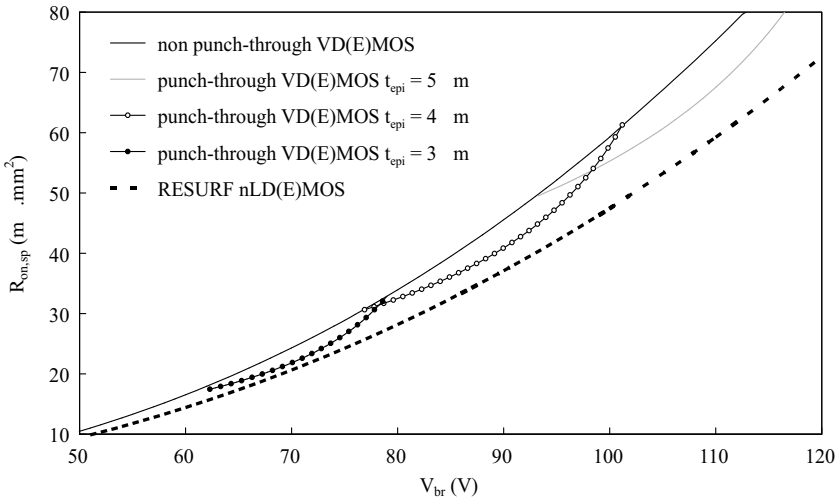
Figuur 12 Een RESURF-diode.

door een doorslagspanning wordt verkregen die veel groter is dan verwacht. Deze doorslagspanning zal dan bepaald worden door het laagst gedopeerd gebied in de structuur, wat dus het p-substraat kan zijn, en die dus een veel hogere spanning aan kan als wat men van de hoger gedopeerde epilaag kan verwachten. De voorwaarde om een optimale RESURF-werking te bekomen, is dat de epidosis een welbepaalde waarde ($D_{opt} = t_{epi,opt} \times N_{epi}$ met t_{epi} en N_{epi} respectievelijk de dikte en het doperingsniveau van de epilaag) heeft die afhankelijk is van het doperingsniveau van het substraat. De werking van het RESURF-effect zal verder geïllustreerd worden bij de behandeling van de verschillende laterale RESURF-vermogenstransistoren.

4.3 De Siliciumlimiet

De siliciumlimiet is de analytische uitdrukking van de specifieke aanweerstand als functie van de doorslagspanning voor een geïdealiseerde MOS-bouwsteen. Op deze manier kunnen beide belangrijkste elektrische parameters van de MOS-vermogenstransistor in één grafiek tegenover elkaar uitgezet worden en vergeleken worden met de ideale siliciumlimiet. Deze grafiek doet dus dienst als maatstaf voor alle bestaande MOS-vermogenstransistoren: in één oogopslag kan men zien voor welke doorslagspanning welke transistor het best presteert, nl. deze die zich het dichtst bij de siliciumlimiet bevindt.

Er bestaan verschillende siliciumlimieten naargelang de vorm (lateraal of verticaal, RESURF of niet), de aard (n- of p-type), en zelfs de



Figuur 13 De siliciumlimieten voor NPT en PT nVD(E)MOS en voor de laterale RESURF nD(E)MOS.

technologie (SOI). In figuur 13 worden de verschillende limieten getoond die voor ons van belang zijn.

Het moet echter gezegd dat deze theoretische limieten steunen op een aantal benaderingen, waardoor de schijnbare conclusie dat de verticale PT-transistoren niet veel onder moeten doen voor de laterale RESURF-transistoren (en zelfs beter doen voor dunnere epilagen) met de nodige voorzichtigheid dient bejegend te worden. Zoals zal blijken uit de paragrafen die deze verschillende transistoren bestuderen, zullen andere fenomenen die niet in rekening werden gebracht bij deze theoretische beschouwingen bepalen welke transistor nu de betere is. Tevens zullen we zien dat er ook andere criteria als de doorslagspanning en de specifieke aan-weerstand een doorslaggevende rol kunnen spelen.

4.4 Welke DMOS: n of p, lateraal of verticaal, RESURF of niet ?

N- of p-type ?

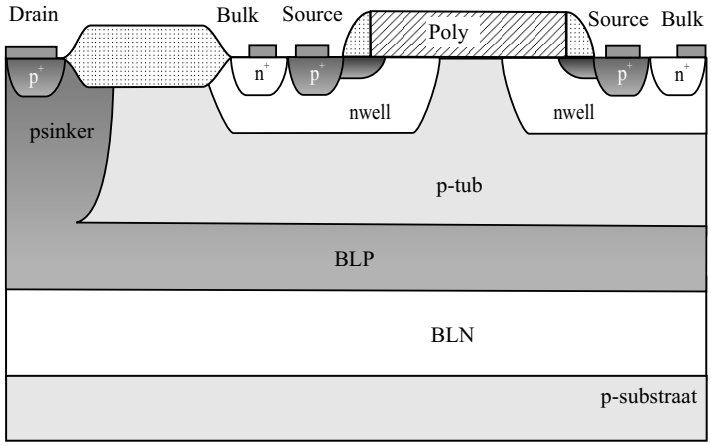
Aangezien de mobiliteit van elektronen ongeveer drie maal hoger is dan die van gaten in silicium, hebben de n-type transistoren een aanweerstand die drie maal beter is dan die van de p-types. Of, in andere woorden, om eenzelfde hoeveelheid stroom te genereren moeten de p-type transistoren ongeveer driemaal groter zijn dan de n-types. Het is duidelijk dat de n-type transistoren verkozen worden boven de p-types.

Niettemin zijn de p-type transistoren belangrijk voor circuitontwerpers daar ze een oplossing bieden voor sommige circuitproblemen waar anders 2, 3, of meerdere n-type transistoren voor nodig zijn (dit komt door het feit dat p-type transistoren zich in de aan-toestand bevinden wanneer $V_{gs} < V_t < 0\,\mathrm{V}$). Het verlies aan siliciumoppervlakte wordt dan grotendeels (zo niet volledig) gecompenseerd. De pDMOS wordt normaal gezien als zwevende transistor gebruikt, wat betekent dat het volledige gebied waarin de transistor zich bevindt op een hogere potentiaal staat dan de omliggende gebieden.

Lateraal of verticaal ?

Men spreekt van verticale bouwstenen in een geïntegreerde schakeling wanneer (een deel van) de drain zich onder de transistor bevindt. De stroom wordt daarna wel gerecupereerd aan de oppervlakte m.b.v. begraven lagen (*buried layers*) en pluggen die deze lagen opnieuw met de oppervlakte verbinden via een laag resistief pad.

Aangezien we werken op een p-substraat, betekent dit dat voor een nVDMOS een begraven laag van het n-type (BLN), een n-plug en een n-epi nodig is (zie figuur 20). Een pVDMOS is moeilijker te realiseren omdat de drain geïsoleerd moet worden van het p-substraat zoals geschetst in figuur 14. Er zullen ook nog andere redenen aangehaald worden waarom een nVDMOS verkozen wordt boven een pVDMOS (in de paragraaf over de pLDMOS).



Figuur 14 Geïntegreerde verticale pDMOS.

Verder kan dezelfde technologie ook nog gekozen worden voor een laterale (RESURF) nDMOS, die al dan niet-zwevend kan zijn (zie ver-

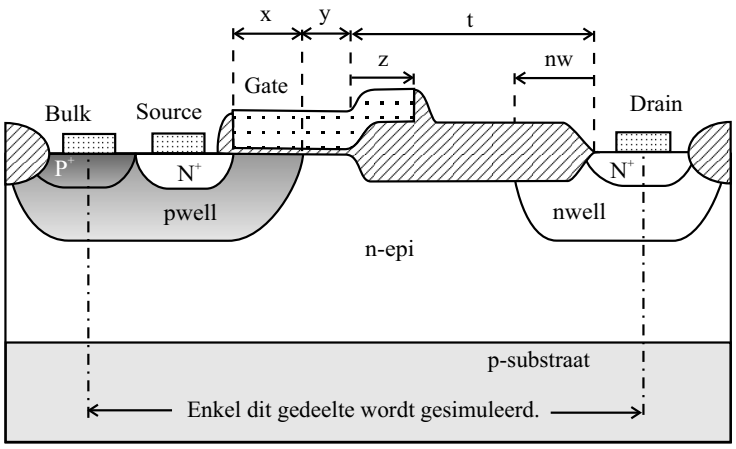
der). Merk op dat de nVDMOS van nature uit zwevend is, wat een groot voordeel is. Een interessante vraag—die in de loop van dit hoofdstuk wordt beantwoord—is of de laterale (RESURF) nDMOS nu werkelijk beter presteert dan de verticale nDMOS.

RESURF of niet ?

Het is duidelijk dat de nVDMOS geen gebruik maakt van het RESURF-effect. De pLDMOS daarentegen zal dit wel doen (zie onder). We kunnen ook nog een laterale RESURF nDMOS ontwerpen die zowel zwevend als niet-zwevend kan zijn (zie onder).

4.5 Niet-zwevende, laterale, RESURF nDEMOS zonder begraven lagen

De zwevende, laterale RESURF nDEMOS wordt getoond in figuur 15 samen met de belangrijkste lay-outparameters die deze transistor beschrijven. Het p-substraat is lager gedopeerd dan de n-epilaag en dus kan hier het "ware" RESURF-effect optreden. Daarmee wordt bedoeld dat bij optimale RESURF-condities (in de eerste plaats een optimale n-epidosis en in de tweede plaats een $t > t_{epi}$) het p-substraat de doorslagspanning zal bepalen. Deze ligt ver boven de 80 V die door de I3T80-technologie wordt opgelegd vanwege het lage doperingsniveau van het substraat en dit voor een groot bereik van epidiktes en -concentraties.



Figuur 15 De niet-zwevende, laterale RESURF nDEMOS met belangrijkste lay-outparameters.

Wanneer deze transistor geschaald moet worden naar 80 V, dan is de enige oplossing een verkleining van de lay-outparameter t. Wanneer we

dit doen voor verschillende epilagen, dan zien we dat er een optimum is. Dit komt omdat hoe hoger het doperingsniveau van de epilaag is, hoe dunner de epilaag moet zijn om aan de optimale RESURF-condities te blijven voldoen. Nu heeft een dunner wordende epi in eerste instantie weinig invloed op de aan-weerstand daar een groot deel van de stroom toch net onder de SiO_2-grensoppervlak stroomt. De hogere concentratie aan ladingsdragers in de epilaag zal dus een sterkere invloed hebben dan de dunnere epilaag. Bij een bepaalde concentratie wordt de epilaag echter zo dun, dat de aan-weerstand opnieuw stijgt (zie tabel 1).

Tabel 1 nLDEMOS op laag gedopeerd substraat met optimale RESURF condities voor de n-epi en geschaald naar 80 V

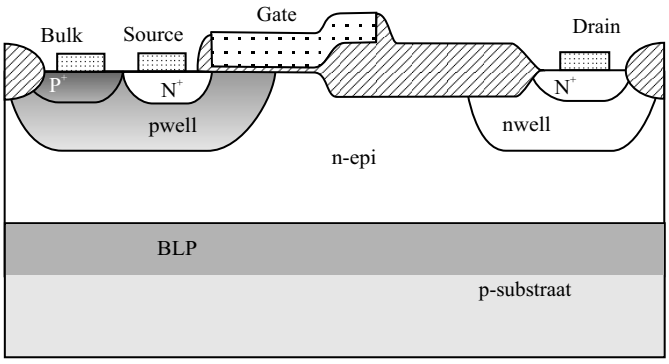
N_{sub} (cm^{-3})	N_{epi} (cm^{-3})	$t_{epi,opt}$ [a] (µm)	t^b (µm)	z	nw	V_{br} (V)	$R_{on,sp}$ ($m\Omega.mm^2$)	SOA[c] (V)
1e15	4e15	3.4	2.8	$t/3$	$t/3$	82	140	32
1e15	6e15	2.6	2.8	$t/3$	$t/3$	84	118	26
1e15	8e15	2.2	2.8	$t/3$	$t/3$	82	103	26
1e15	1e16	1.8	2.8	$t/3$	$t/3$	81	98	26
1e15	1.2e16	1.4	2.8	$t/3$	$t/3$	80	101	23
1e15	1.4e16	1.4	3.2	$t/3$	$t/3$	87	104	26
1e15	1.6e16	1.0	3.2	$t/3$	$t/3$	82	124	23

[a] Initiële waarde.
[b] Kleinste waarde waarvoor $V_{br} > 80\,V$.
[c] Gedefinieerd als de V_{ds} waarde waarvoor $I_{sub}/I_d = 0.001$ (bij $V_{gs} = 3.3\,V$).

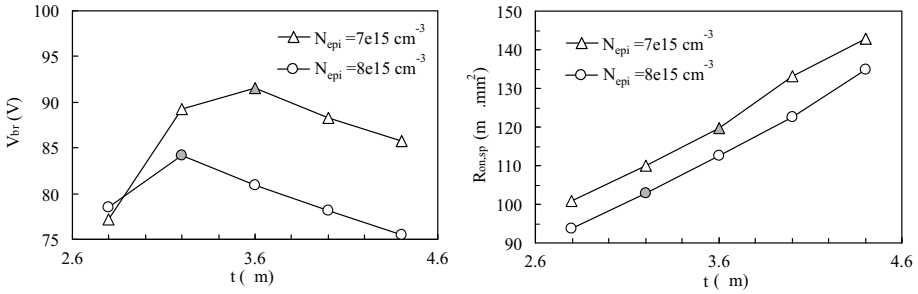
4.6 Niet-zwevende, laterale RESURF nDEMOS op een p-type begraven laag

Figuur 16 toont een niet-zwevende, laterale RESURF nDEMOS op een p-type begraven laag (BLP). Het grootste verschil met de vorige structuur is dat de n-epi nu het laagst gedopeerd gebied is, en deze dus ook de doorslagspanning zal bepalen. Toch blijven we hier spreken van een RESURF-transistor, aangezien de distributie van de elektrische velden in deze bouwsteen ook op een ideale manier wordt gespreid, net zoals in een "ware" RESURF-transistor. Het resultaat daarvan is dat men voor het doperingsniveau van de n-epilaag toch waarden kan halen die hoger zijn dan men zou verwachten. Figuur 17 toont dat 80 V nog net haalbaar is in deze RESURF-transistor met een $N_{epi} = 8e15\,cm^{-3}$, terwijl dat een

dergelijk hoge concentratie in een abrupte diode reeds bij 63 V zou doorslaan. Deze nLDEMOS met $V_{br} = 84\,\mathrm{V}$ en $R_{on,sp} = 103\,\mathrm{m\Omega.mm^2}$ doet het net iets slechter dan de beste nLDEMOS op een laag gedopeerd substraat.



Figuur 16 De niet-zwevende, laterale RESURF nDEMOS op een BLP.

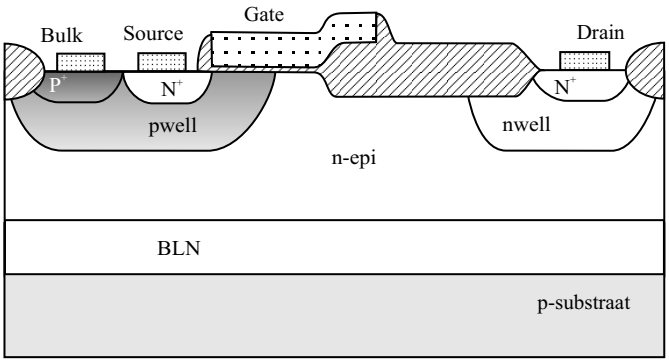


Figuur 17 Doorslagspanning (links) en specifieke aan-weerstand (rechts) voor de niet-zwevende RESURF nLDEMOS met BLP in functie van de lengte van het veldoxide t en dit voor de hoogst mogelijke doperingsniveaus voor de n-epi. De gevulde datapunten tonen het optimum voor elk van deze doperingsniveaus.

4.7 Zwevende, laterale, niet-RESURF nDEMOS op een n-type begraven laag

De zwevende, laterale RESURF nDEMOS die geschetst wordt in figuur 18, maakt duidelijk geen gebruik van het RESURF-effect aangezien de n-type begraven laag de vorming van een depletielaag als gevolg van het potentiaalverschil tussen substraat en drain in de af-toestand volledig voor zijn rekening zal nemen. Er is m.a.w. geen 2D-effect, waardoor de doorslagspanning enkel en alleen bepaald wordt door pwell–nepijunctie. Het doperingsniveau van de n-epi moet dan ook drastisch dalen i.v.m. de waarden bij de niet-zwevende laterale transistoren. Verder schenken

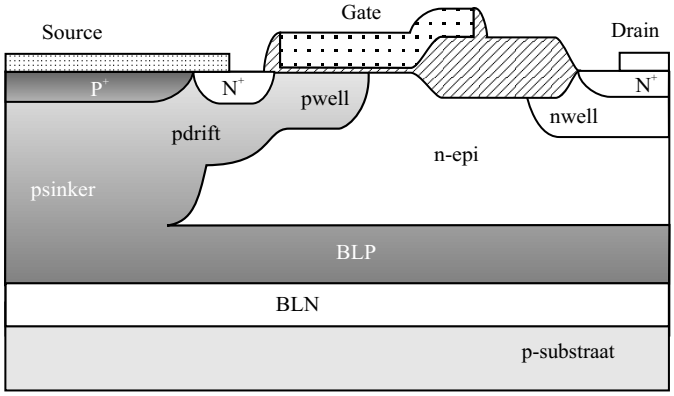
we geen aandacht aan deze transistor aangezien zelfs de verticale DMOS beter doet dan dit type bouwsteen.



Figuur 18 De zwevende, laterale niet-RESURF nDEMOS op een n-type begraven laag (BLN).

4.8 Zwevende, laterale, RESURF nDEMOS op twee begraven lagen

Het is ook mogelijk om een structuur te maken op twee begraven lagen (figuur 19). Het is in principe mogelijk om deze transistor net zo goed te maken als zijn niet-zwevende tegenhanger, het enige verschil is immers de aanwezigheid van de BLN. Technologisch is deze struc-



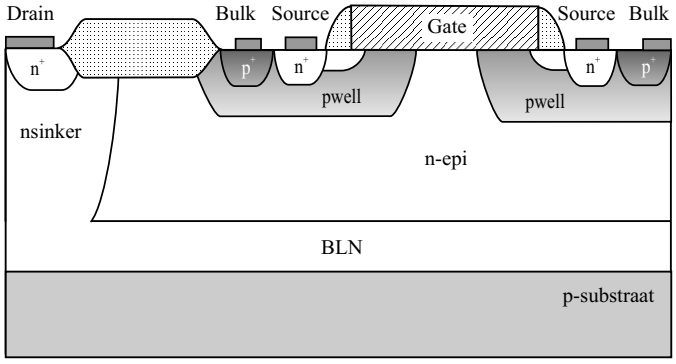
Figuur 19 Een zwevende, laterale RESURF nDEMOS op twee begraven lagen.

tuur echter moeilijk te verwezenlijken in zijn ideale vorm (BLN en BLP streng gescheiden van elkaar). In de I3T80-technologie bijvoorbeeld bedekt de BLN een groot gedeelte van de BLP omdat beide lagen voor de n-epigroei worden geïmplanteerd. Toch levert een dergelijke structuur

in deze technologie een transistor op met een hoge doorslagspanning (nl. 103 V), maar wel met een specifieke aan-weerstand die ongeveer tweemaal zo groot is als deze van de verticale nDEMOS.

4.9 Zwevende, geïntegreerde, verticale nDEMOS

De zwevende, geïntegreerde, verticale nDEMOS heeft in wezen een eenvoudige structuur (figuur 20).



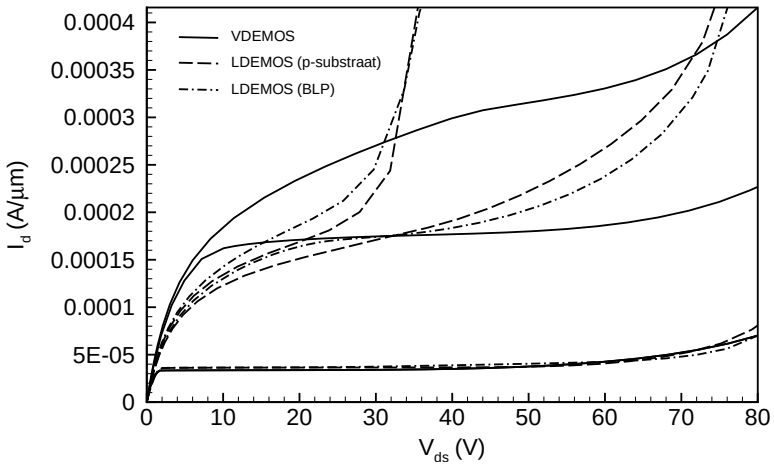
Figuur 20 De zwevende, geïntegreerde, verticale nDEMOS.

Een variatie van de n-epidikte voor verschillende doperingsniveaus van de epilaag, leert ons dat $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ het maximum niveau haalbaar is, met de beste resultaten wat betreft $R_{on,sp}$. Deze weerstandswaarden zijn afhankelijk van de afstand tussen beide pwells, aangezien een te kleine afstand de stroom zal afsnijden. Er is echter een optimum daar een te breed pad de lengte van de transistor te veel doet toenemen, en daardoor $R_{on,sp}$ opnieuw stijgt. Dit optimum wordt samen met de optima voor de niet-zwevende, laterale RESURF-transistoren op p-substraat enerzijds en op BLP anderzijds, in tabel 2 opgenomen.

Het valt onmiddellijk op in deze tabel dat de nVDEMOS weliswaar slechtere V_{br} en $R_{on,sp}$ waarden heeft dan de laterale RESURF-transistoren, maar dat de SOA voor deze transistor veel groter is dan voor de andere. Dit komt omdat de laterale RESURF-transistoren te kampen hebben met het Kirk-effect, die heel sterk aanwezig is in laterale RESURF-structuren van het n-type. Dit is niet het geval in de verticale DMOS, die daardoor een veel groter spanningsbereik heeft (zie ook figuur 21).

Tabel 2 Best presterende, niet zwevende RESURF nLDEMOS op een laag gedopeerd substraat en op een BLP vergeleken met de best presterende, zwevende nVDEMOS. Ter vergelijking zijn ook de niet zwevende transistoren met eenzelfde doperingsniveau voor de n-epi als de nVDEMOS weergegeven.

N_{sub} (cm^{-3})	N_{epi} (cm^{-3})	$t_{epi,opt}$ (µm)	t (µm)	z	nw	V_{br} (V)	$R_{on,sp}$ ($m\Omega.mm^2$)	SOA (V)
$6e17$	$8e15$	4.4	3.2	$t/3$	$t/3$	84	103	33
$1e15$	$1e16$	1.8	2.8	$t/3$	$t/3$	81	98	26
nVDEMOS	$4e15$	6.8	–	–	–	80	122	80
$6e17$	$4e15$	6.8	2.8	$t/3$	$t/3$	83	140	26
$1e15$	$4e15$	3.4	2.8	$t/3$	$t/3$	82	140	32



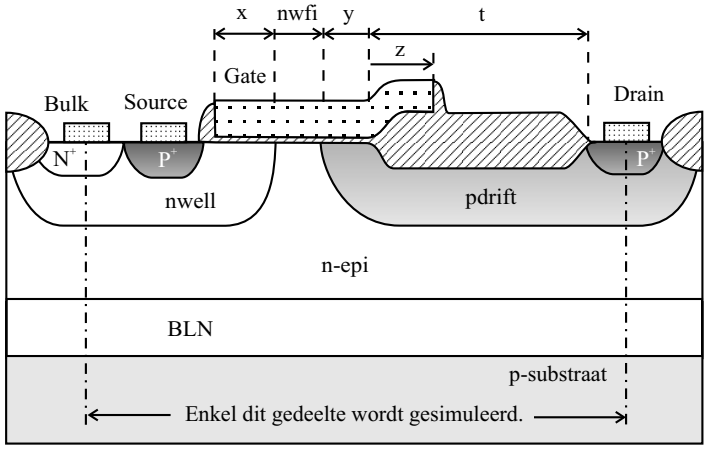
Figuur 21 $I_d(V_{ds})$ karakteristieken voor $V_{gs} = 1.1, 2.2$ and $3.3\,\mathrm{V}$ voor de best presterende nVDEMOS en voor de RESURF nLDEMOS structuren op een n-epi met $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ (zie tabel 2).

4.10 Zwevende, laterale RESURF pDEMOS

Waarom lateraal ?

Aangezien we werken op een n-epi, moet er ofwel een ptub ofwel een pdrift voor respectievelijk de verticale of de laterale pDMOS, gedefinieerd worden (selectieve epi groei is geen optie in de I3T80-technologie). Daardoor moeten zowel de ptub als de pdrift een doperingsniveau heb-

ben dat hoger is dan dat van de n-epi. Dit sluit de verticale pDMOS uit aangezien alle n-epi lagen voor de beste nDMOS-transistoren te zwaar gedopeerd zijn om een combinatie met een pVDMOS te verwezenlijken. Dit wordt verklaard door het feit dat voor een 80 V pVDMOS ongeveer hetzelfde doperingsniveau als voor de 80 V nVDMOS nodig is (d.w.z. $N_{ptub} \approx N_{epi}$), wat betekent dat zelfs in combinatie met de n-epi voor de beste nVDMOS, de n-epi nog zou moeten verminderen in doperingsniveau. Aangezien de pDMOS voorrang moet geven aan de nDMOS (vanwege de hogere mobiliteit van elektronen), rest enkel nog de laterale pDMOS als alternatief (figuur 22).

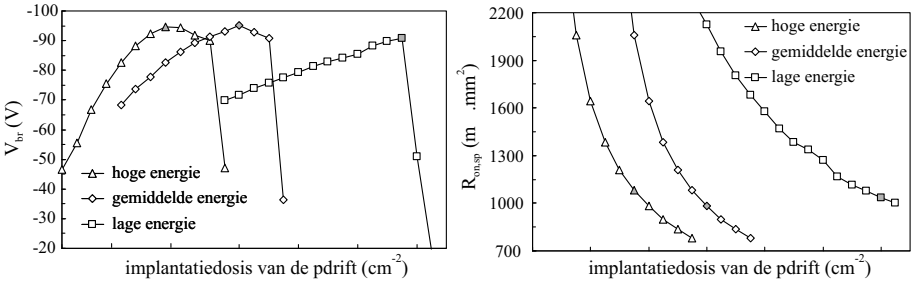


Figuur 22 De zwevende, laterale RESURF pLDEMOS met de meest belangrijke lay-out parameters.

Omdat de pLDMOS gebruik maakt van het RESURF-effect, kan het doperingsniveau van de pdrift een stuk hoger zijn dan dit van de verticale pDMOS (analoog aan de n-type transistoren). Nu blijkt het echter wel zo te zijn dat de combinatie van de pLDMOS met één van de beste laterale nDMOS (tabel 2) quasi onhaalbaar is. De n-epi moet een lager doperingsniveau hebben ($N_{epi} < 8e15\,\mathrm{cm}^{-3}$), willen we een pLDMOS in deze technologie integreren. Omdat voor deze lagere doperingsniveaus de prestatie van de laterale transistoren en de beste verticale nDMOS vergelijkbaar worden, kiezen we voor het doperingsniveau van de n-epidie gebruikt wordt voor de beste verticale nDMOS ($N_{epi} = 4e15\,\mathrm{cm}^{-3}$). Daarbij komt nog dat de verticale nDMOS een grote SOA heeft zonder toevoeging van extra lagen, wat een extra motivering is om voor deze transistor te kiezen.

De pdriftlaag wordt geïmplanteerd voor de veldoxidatie, wat ervoor zorgt dat er geen extra temperatuurstap nodig is. De implantatiedosis

is de belangrijkste parameter aangezien deze rechtstreeks de dosis van het pdrift gebied bepaald en zodoende zorgt voor het RESURF-effect. Figuur 23 toont dat de optimale implantatiedosis stijgt naarmate de implantatie-energie afneemt, en dit omdat de daaropvolgende veldoxidatie meer boor opneemt indien de implantatie-energie daalt (het boor zit minder diep). Alhoewel op de figuur blijkt dat de $V_{br} - R_{on,sp}$ getallen sterk gelijkmatig zijn voor een breed bereik van implantatie-energieën (de hoge energie is driemaal de lage) kiezen we voor de hoge energie aangezien deze het minst onderhevig zal zijn aan de procesvariaties op de veldoxide dikte en aangezien deze resulteert in een doperingsniveau onder het actief gebied en onder het veldoxide dat het meest gelijk blijft.

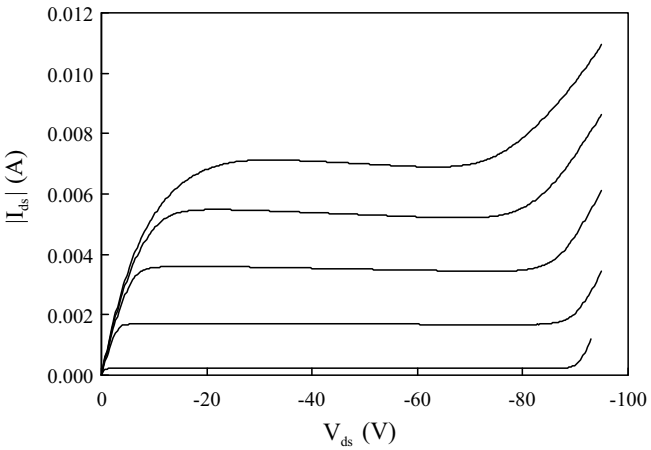


Figuur 23 V_{br} (links) en $R_{on,sp}$ i.f.v. de implantatiedosis voor verschillende implantatie-energieën. Gevulde datapunten tonen de optima.

De lay-out die gebruikt werd in figuur 23 is niet ideaal en kan verder geoptimaliseerd worden—vooral naar $R_{on,sp}$ toe, zonder V_{br} aan te tasten. Verschillende frames met lay-outvariaties werden op testchip geplaatst en uitgemeten. Eén voorbeeld van uitgemeten karakteristieken wordt getoond in figuur 24. De pLDEMOS—in tegenstelling tot de nLDEMOS—heeft geen last van het Kirk-effect. Enkel bij V_{gs} waarden buiten het normale bereik ($V_{gs} = -4.4$ en $-5.5\,\mathrm{V}$) wordt het Kirk-effect uitgesproken zichtbaar. Uiteindelijk werd er één transistor gekozen die commercieel wordt aangeboden door AMIS.

4.11 Besluit

We hebben gezien dat in een 80 V junctiegeïsoleerde technologie de laterale DMOS beter is als de verticale DMOS indien enkel naar V_{br} en $R_{on,sp}$ wordt gekeken. Zonder extra lagen heeft de laterale DMOS een minder groot spanningsbereik als de verticale DMOS. Daarbij komt nog dat de beste laterale DMOS-transistoren niet of nauwelijks te combine-



Figuur 24 *Gemeten* $I_d(V_{ds})$ voor $V_{gs} = -1.1, -2.2 \ldots -5.5\,\mathrm{V}$ van een pLDEMOS (breedte $W = 40\,\mu\mathrm{m}$) onder optimale RESURF-condities.

ren vallen met een pDMOS. Daardoor moet het doperingsniveau van de n-epi omlaag en komen we tot de conclusie dat de beste combinatie een verticale DMOS en een laterale pDMOS is. De door AMIS commercieel aangeboden nVD(E)MOS- en pLDEMOS-transistoren worden in het afsluitend hoofdstuk vergeleken met de transistoren van de concurrentie.

5 IGBT

5.1 Inleiding

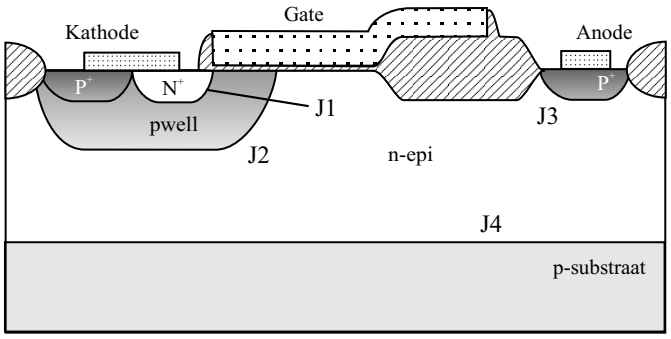
De IGBT (I*nsulated* G*ate* B*ipolar* T*ranstistor*) werd pas in de begin jaren 80 uitgevonden door twee onafhankelijke groepen ([RGGN83] en [BAG$^+$82]). Ze kwamen tot dezelfde synthese terwijl ze op zoek waren naar een bouwsteen die het beste van de MOS-vermogenstransistor en van de bipolaire vermogenstransistor combineert. Startend van een thyristor, ontwierpen ze een transistor die niet zoals de thyristor in regeneratieve toestand komt, maar waar de MOS-gate controle op elk moment bewaard blijft. Het is in wezen een MOS-transistor die een bipolaire transistor aanstuurt, maar dan geïntegreerd in één enkele transistor.

Sinds de jaren 80 wordt de IGBT vooral gebruikt als discrete bouwsteen met spannings- en stroombereiken die ver boven die van de geïntegreerde schakelingen liggen. Niettemin zullen we onderzoeken of deze transistor ook in de I3T80-technologie kan geïntegreerd worden. We beginnen met een structuur die sterk gelijkt op een LDMOS (enkel de n^+-drain wordt vervangen) en naarmate het hoofdstuk vordert, intro-

duceren we al dan niet bestaande lagen ten einde de werking van de geïntegreerde IGBT te verbeteren.

5.2 Doorslag in een nLIGBT zonder begraven lagen

Figuur 25 schetst een laterale IGBT zonder begraven lagen. Het enige verschil met de laterale RESURF nDEMOS is het p^+-draincontact, dat we vanaf nu de anode noemen. Deze schijnbaar kleine aanpassing maakt een wereld van verschil aangezien we nu een bipolaire bouwsteen hebben die pas stroom zal leveren wanneer de p^+–nepi-junctie voorwaarts gepolariseerd wordt. Merk op dat de n-epi zwevend is.

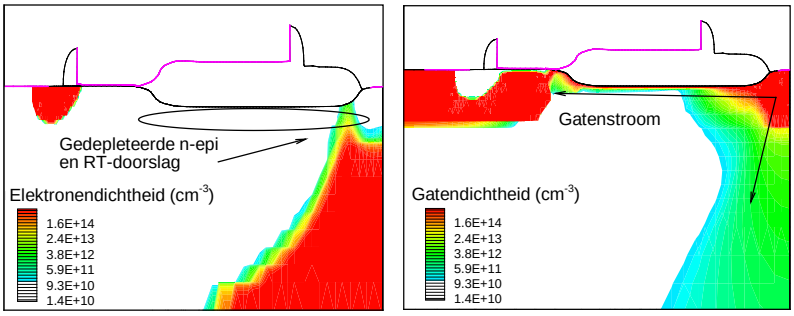


Figuur 25 Een laterale IGBT zonder begraven lagen.

Wanneer er een positieve spanning staat op de anode i.v.m. de andere contacten, dan is het meteen duidelijk dat deze transistor een lage doorslag zal hebben doordat de depletielagen komende van beide juncties J2 en J4, de junctie J3 zal bereiken. Dit is wat we RT-doorslag genoemd hebben en wordt geïllustreerd in figuur 26. Dit fenomeen gaat niet gepaard met impactionisatie en wordt enkel veroorzaakt door het feit dat er een stroompad ontstaat voor de gaten. Het is duidelijk dat deze vorm niet kan worden behouden en dat verhinderd moet worden dat er een RT-doorslag ontstaat. Dit kan op een eenvoudige manier gebeuren door de introductie van bestaande lagen, zoals in de volgende paragraaf wordt verduidelijkt.

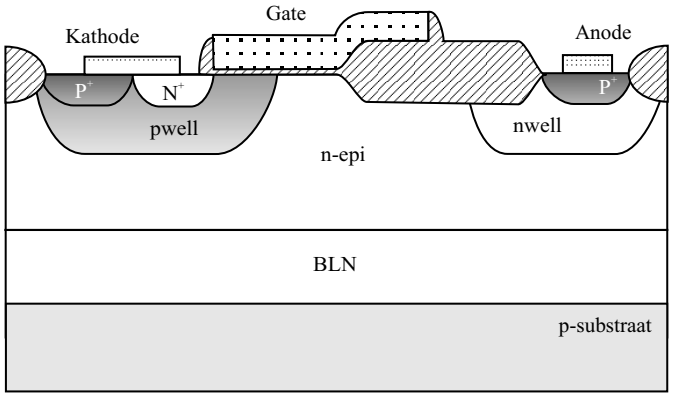
5.3 Zwevende nLIGBT met BLN

De introductie van de BLN, zoals in figuur 27, verslechtert aanzienlijk de parasitaire bipolaire transistor naar het substraat toe, maar verhoogt de doorslagspanning niet aangezien de RT tussen pwell en p^+ blijft bestaan. Daartoe is een n-type buffer nodig, waarvoor we in eerste in-



Figuur 26 Elektronen- en gatendichtheid bij doorslag (voorwaarts, i.e., $V_{ak} > 0\,\mathrm{V}$) voor een LIGBT zonder begraven lagen.

stantie de standaard-nwell kiezen. Deze is echter niet sterk genoeg daar de doorslagspanning slechts $\sim 52\,\mathrm{V}$ is. We moeten dus een nieuwe laag introduceren willen we de doorslagspanning opkrikken.



Figuur 27 Een LIGBT met BLN en nwell.

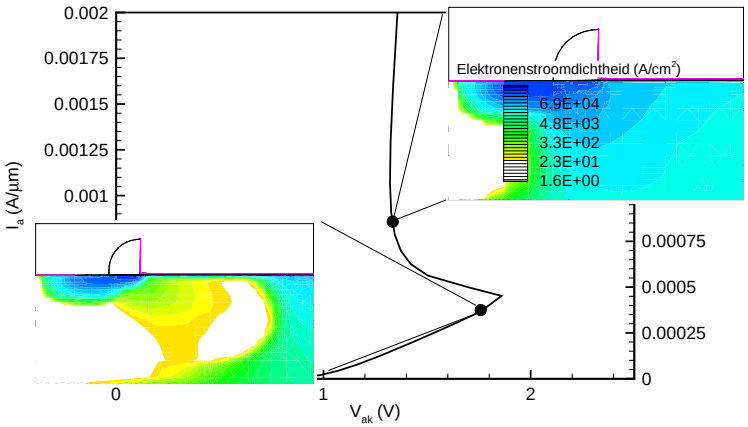
Zwevende nLIGBT met BLN en een toegevoegde nbuffer

Tabel 3 geeft enkele simulatieresultaten weer voor verschillende implantatiedosissen van de nbuffer en toont aan dat hogere doorslagspanningen te bereiken zijn voor een transistor met een redelijke grootte (lengte driftgebied $\sim 5\,\mu\mathrm{m}$). In deze tabel zijn ook de voorwaartse spanningsvallen opgenomen bij een arbitrair gekozen stroomniveau. Dit bewijst dat bij te hoge dosissen (te zware nbuffer) de bipolaire transistor (p^+–nbuffer + n-epi–pwell) niet meer naar behoren werkt en de prestatie van de LIGBT sterk afneemt.

Tabel 3 V_{br} en voorwaartse spanningsval V_{fwd} voor verscheidene nbuffer implantatiedosissen bij $500\,keV$.

Dose (cm^{-2})	V_{br} (V)	V_{fwd} (V)[a]
2.5e14	64	1.5[b]
5e14	66	1.6
1e15	66	1.8
2e15	68	2.2
4e15	70	9.6

[a]Bij $I_a = 0.0004\,\mathrm{A/\ m}$ voor $V_{gk} = 3.3\,\mathrm{V}$
[b]Slaat door bij $0.00038\,\mathrm{A/\ m}$



Figuur 28 Terugslag bij $V_{gk} = 3.3\,\mathrm{V}$ voor de LIGBT uit tabel 3 met een nbufferdosis gelijk aan $1e15\,\mathrm{cm}^{-2}$. De linkse doorsnede toont de LIGBT net voor de terugslag, de rechtse toont de LIGBT in thyristormode.

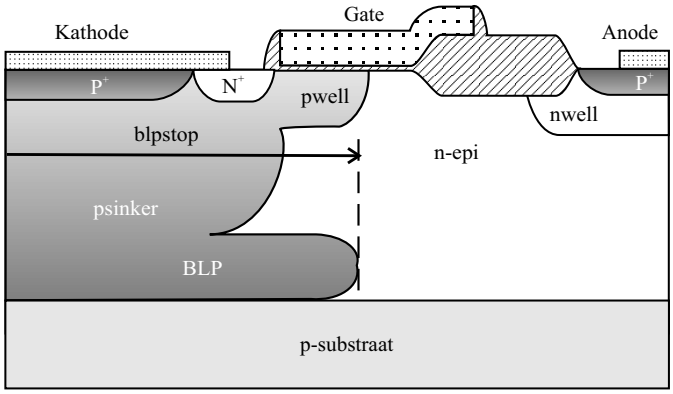
Voor al deze transistoren wordt de thyristormode bereikt voordat er sprake is van saturatie (bij $V_{gk} = 2.2$ en $3.3\,\mathrm{V}$), zie figuur 28: de gatenstroom die geïnjecteerd wordt door de p^+ in de n-epi en naar de kathode stroomt via een pad onder de pwell, veroorzaakt een voorwaarts gepolariseerde n^+–pwell junctie. Dit brengt een elektronenstroom op gang die gecollecteerd wordt door de basis van de werkende pnp. Deze positieve terugkoppeling zorgt voor een terugslag in de spanning en een aanzienlijke toename van de stroom aangezien een thyristor in werking is waarop alle gatecontrole verloren is.

Deze thyristorwerking kan uitgesteld worden door het doperingsniveau van de basis van de bipolaire npn te verhogen, door deze basis breder te maken, door het pad onder de pwell te verkorten, of door de winst van de werkende bipolaire transistor te verkleinen. Deze laat-

ste maatregel is echter geen optie daar de prestatie van de LIGBT op die manier nog wordt verlaagd. Het pad onder de pwell kunnen we niet kleiner maken aangezien we reeds het minimum—opgelegd door de ontwerp-regels—hanteren. Dit laat enkel als alternatief de p-basis van de parasitaire npn te verslechteren. Dit kan echter gemakkelijk m.b.v. twee bestaande lagen: de pdrift en de psinker, zoals zal blijken in de volgende paragrafen.

5.4 Niet-zwevende nLIGBT met BLP

De vorige paragraaf behandelde een nieuwe vorm van de LIGBT. De standaard-nLIGBT in een junctiegeïsoleerde technologie neemt de vorm aan zoals geschetst in figuur 29 [AU01]. Merk op dat deze transistor niet langer zwevend is, maar wel opnieuw gebruik maakt van het RESURF-effect (indien de BLP voldoende onder het drift gebied aanwezig is).

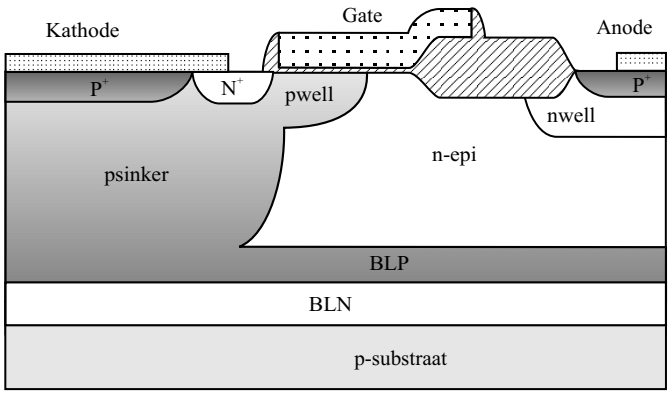


Figuur 29 De niet-zwevende nLIGBT met BLP en psinker.

De lay-outparameter blpstop die wordt aangegeven op figuur 29 is van cruciaal belang voor het bepalen van de doorslagspanning. Uit simulaties en metingen blijkt dat er een ideale lengte (blpstop = 8 µm) is die een doorslagspanning van 85 V garandeert. De BLP diffundeert voor deze waarde van blpstop tot net onder de anode. De reden waarom een blanco implantatie geen betere resultaten haalt, is dat de BLP niet ontworpen werd met deze transistor in het achterhoofd en het dus geen ideale RESURF-condities voor de n-epidosis teweeg brengt. De substraatstroom van de standaard-nLIGBT wordt in figuur 33 vergeleken met een nieuw ontwerp (zie volgende paragraaf).

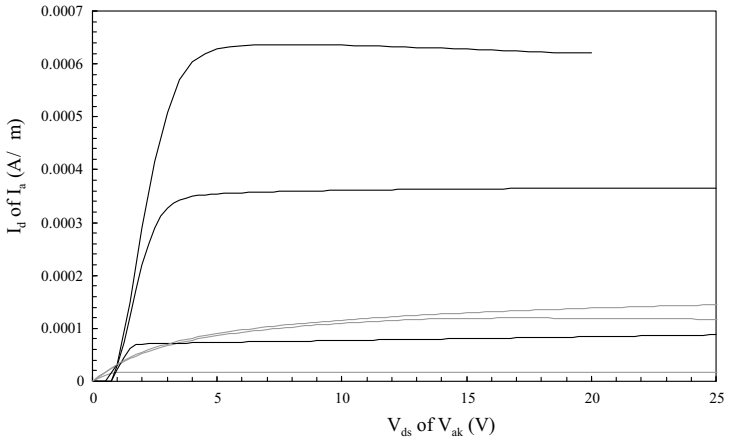
5.5 Zwevende nLIGBT met BLN en BLP

Indien we beide vorige ontwerpen combineren, dan komen we tot de synthese zoals getekend in figuur 30 met BLN en BLP. Dit wordt opnieuw een zwevende transistor en toch gebruikt het de RESURF-techniek. Dankzij dit RESURF-effect is het opnieuw mogelijk de nwell te gebruiken als nbuffer, wat betekent dat we een nLIGBT creëren die enkel bestaat uit bestaande lagen. Zowel simulaties als metingen geven aan dat de doorslagspanning van deze transistor competitief is: $V_{br} = 73 - 76\,\mathrm{V}$.



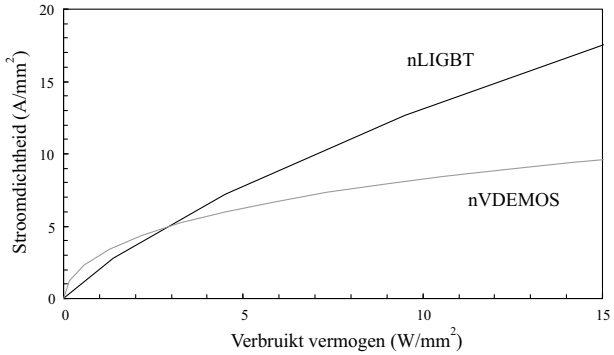
Figuur 30 De zwevende nLIGBT met BLN en BLP.

De *gemeten* uitgangskarakteristieken van de nLIGBT met BLN en BLP worden in figuur 31 vergeleken met deze van de nVDEMOS, waaruit het grote verschil van stroomniveau blijkt. Dit geeft echter geen volledig



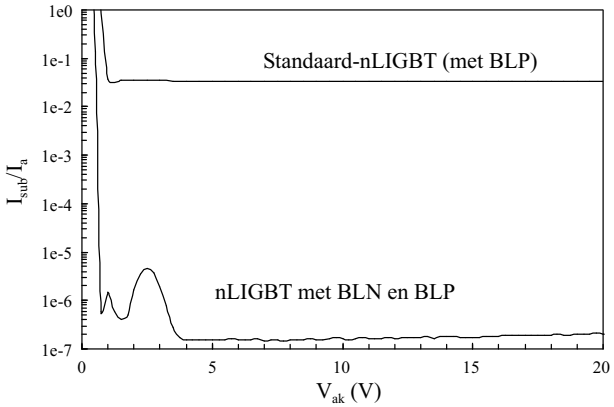
Figuur 31 Gemeten uitgangskarakteristieken voor $V_{gk} = 1.1$, 2.2 en 3.3 V voor een LIGBT met BLN en BLP vergeleken met een typische VDEMOS (grijs, met $V_{gs} = 1$, 2 en 3.3 V).

beeld daar de VDEMOS bij lage V_{ds} reeds in de aan-toestand is, en de LIGBT niet. Correcter is een vergelijking van de stroomdichtheden in functie van de verbruikte vermogens in de aan-toestand (50 % duty cycle) per oppervlakte-eenheid (figuur 32). Zo zien we dat de LIGBT het haalt van de nVDMOS vanaf $3\,\mathrm{W/mm^2}$. AMIS maakt reeds gebruik van DMOS-transistoren die $6.7\,\mathrm{W/mm^2}$ in DC verbruiken en er kunnen vermoedelijk nog hogere waarden gehaald worden. Er zijn dus voldoende aanwijzingen om de LIGBT in een 80 V technologie te overwegen.



Figuur 32 *Gemeten* stroomdichtheid versus verbruikt vermogen per oppervlakte-eenheid bij $V_{gk} = 3.3\,\mathrm{V}$ voor de LIGBT van figuur 31, vergeleken met een typische VDEMOS ($V_{gs} = 3.3\,\mathrm{V}$).

Bovendien slagen we erin met dit ontwerp de substraatstromen verder te onderdrukken (figuur 33). Uit metingen op andere lay-outvariaties van de nLIGBT met BLN en BLP blijkt dat de piek in de substraat-

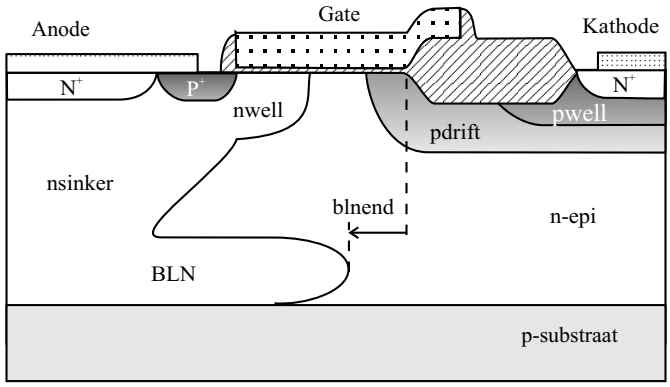


Figuur 33 *Gemeten* verhouding van de substraatstroom tot de anodestroom voor $V_{gk} = 3.3\,\mathrm{V}$ van een zwevende nLIGBT met BLN en BLP, vergeleken met een niet-zwevende nLIGBT met BLP.

stroom vermeden kan worden. Tot slot kunnen we stellen dat we erin geslaagd zijn een junctiegeïsoleerde LIGBT te ontwerpen met verwaarloosbare substraatstromen, een competitieve doorslagspanning, een grote SOA en saturatiestromen die 4 tot 6 keer deze van een verticale DMOS zijn.

5.6 P-type LIGBTs

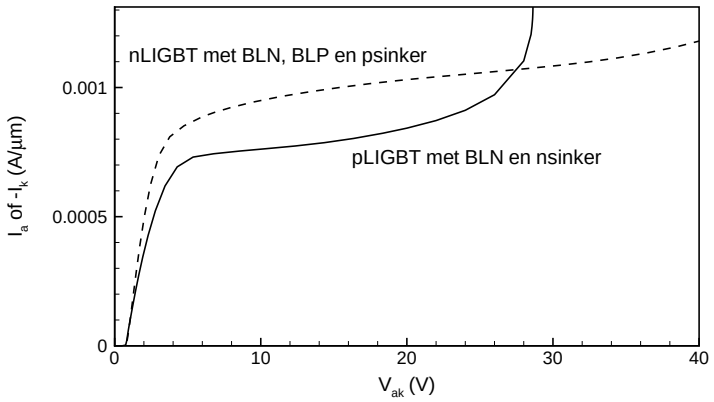
De pLIGBT kan men vormen door de standaard-nLIGBT in zijn duale vorm om te zetten m.b.v. de bestaande lagen (figuur 34). Het grote verschil met de standaard-nLIGBT (figuren 29) is dat er nu een vierlagenstructuur verschijnt aan de kathode kant van de transistor met de n-epi die deel uitmaakt van de anode, in tegenstelling met de pLIGBT waar de n-epi deel uitmaakt van het zwevende driftgebied. Het gevolg is dat in de af-toestand het potentiaalverschil tussen anode en kathode dient gedragen te worden door de bipolaire n^+–pwell + pdrift–nepi-transistor, die een veel lagere RT-doorslag heeft dan het geval was voor de nLIGBT. Bij een $blnend = 0\,\mu m$ bereikt de pLIGBT een $V_{br} = 43\,V$.



Figuur 34 De zwevende pLIGBT met nsinker en BLN.

De meest interessante eigenschap van de pLIGBT is dat de aan-weerstand niet langer bepaald wordt door de mobiliteit van de gaten alleen. In tegenstelling tot de pLDEMOS, die inherent een aan-weerstand heeft die drie keer slechter is dan die van de nLDEMOS, wordt het driftgebied van de p-type LIGBT tijdens de aan-toestand overspoeld met elektronen en gaten die in dergelijke concentraties voorkomen dat de geleiding in dit gebied sterk toeneemt (i.e., geleidingsmodulatie). Het resultaat is een saturatiestroom die maar 20 % moet onderdoen voor die van de nLIGBT (figuur 35). Indien we de vermogensdissipatie van deze

pLIGBT vergelijken met die van de pLDEMOS komt het niveau waarvoor de pLIGBT het wint van de pLDEMOS dan ook een stuk lager dan bij de n-type transistoren, nl. op $\sim 1\,\mathrm{W/mm^2}$. Bovendien is er in deze transistor geen enkel probleem met substraat stromen zolang de BLN onder de kathode aanwezig is ($blnend \leq 0\,\mu\mathrm{m}$).



Figuur 35 Uitgangskarakteristiek van een pLIGBT (voor $V_{ga} = -3.3\,\mathrm{V}$) met $blnend = 0\,\mu\mathrm{m}$ vergeleken met een nLIGBT (voor $V_{gk} = 3.3\,\mathrm{V}$) met een zwevende BLN, BLP en psinker.

5.7 Besluit

We hebben verschillende nieuwe IGBT structuren geïntegreerd in de I3T80-technologie van AMIS. Simulaties en metingen tonen aan dat de beste benadering een nLIGBT met BLN en BLP is. Zo zijn we erin geslaagd een nLIGBT te ontwerpen met verwaarloosbare substraatstromen, een competitieve doorslagspanning, een grote SOA, vlugge schakeltijden en saturatiestromen die 4 tot 6 maal hoger liggen dan bij de nVDMOS. Een gelijkaardige pLIGBT werd ontworpen die in vergelijking nog beter doet vanwege de geleidingsmodulatie in het pdrift gebied tegenover de geleiding door gaten in de pLDEMOS. De pLIGBT heeft een brede SOA, opmerkelijk lage substraatstromen en vlugge schakeltijden. Jammer genoeg is de doorslagspanning van dit device slechts 42 V, maar dit is enkel te wijten aan het feit dat we ook voor deze transistor enkel gebruik maken van de reeds bestaande lagen.

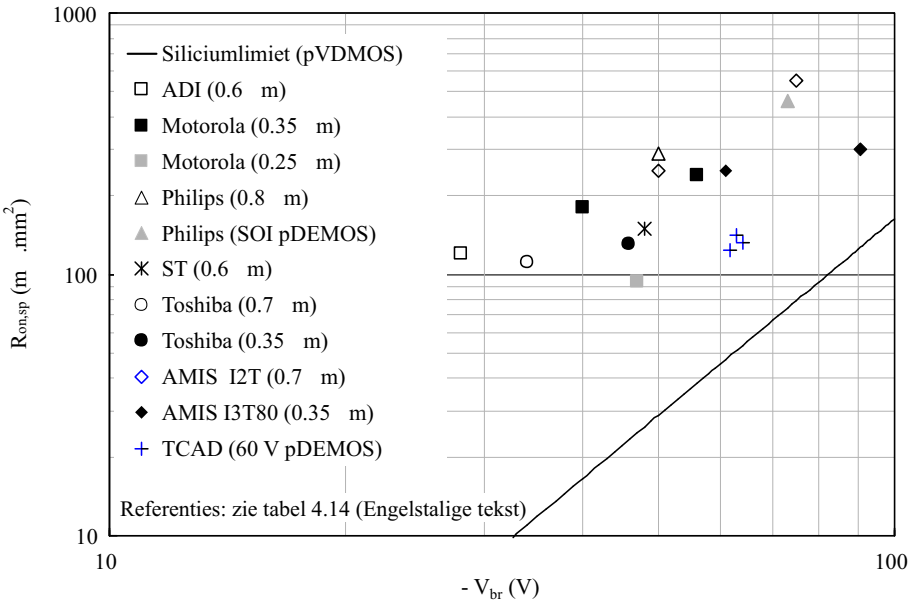
6 Synopsis

Na twee korte, algemeen inleidende hoofdstukken en één hoofdstuk ter introductie van het TCAD-simulatiewerk, hebben we zowel de integratie van de MOS-vermogenstransistor als van de IGBT in een junctiegeïsoleerde standaard-CMOS-technologie onderzocht.

We zijn tot de conclusie gekomen dat in de junctiegeïsoleerde, 80 V technologie van AMIS, de beste combinatie van complementaire MOS-vermogenstransistoren de nVDMOS en de pLDMOS is. Alhoewel de laterale nDMOS beter presteert wat betreft de V_{br} versus $R_{on,sp}$ dan de verticale transistor, wordt toch gekozen voor de laatste en dit om drie redenen. Ten eerste heeft deze transistor een groot stroom- en spanningsbereik en dat zonder gebruik te maken van extra (dure) lagen, die wel nodig zijn indien we met de laterale RESURF nDMOS een even groot bereik zouden willen bereiken. Ten tweede is deze transistor zwevend van nature uit. De nLDMOS kan in theorie ook zwevend gemaakt worden zonder verlies van prestatie van de transistor, maar dit is technologisch moeilijk realiseerbaar. Ten derde is de verticale nDMOS gemakkelijk te combineren met een laterale pDMOS, wat voor de beste nLDMOS-transistoren moeilijker is.

Omdat de ontwikkeling en het ontwerp van de pLDEMOS gebeurde in de TFCG-groep en die van de nVD(E)MOS bij AMIS, wordt in deze samenvatting enkel de grafiek met de siliciumlimiet en de transistoren van de concurrenten voor de pLDEMOS opgenomen (figuur 36). Het analogon voor de nVDEMOS vindt men terug in de Engelstalige versie. Uit deze figuur leiden we af dat de I3T80 pLDEMOS één van de beste huidige pLDEMOS-transistoren is.

Wat betreft IGBT-transistoren, zijn we erin geslaagd zowel een n- als een pLIGBT te ontwerpen die—naar wij weten—nooit eerder in de literatuur werden beschreven. Voor beide transistoren zijn de substraatstromen onderdrukt, is er een groot spannings- en stroombereik, en zijn er vlugge schakeltijden. Bovendien zijn beide zwevend, maken ze enkel gebruik van bestaande lagen en hebben we aangetoond dat de stroomdichtheid (per oppervlakte-eenheid) de concurrentie aan kan met de DMOS-transistoren.



Figuur 36 De I3T80 pLDEMOS i.v.m. de concurrentie en de siliciumlimiet.

Bibliografie

[AU01] G. Amaratunga and F. Udrea. Power Devices for High Voltage Integrated Circuits: New Device and Technology Concepts. In *Semiconductor Conference*, volume 2, pages 441–448, 2001.

[AV79] J.A. Appels and H.M.J. Vaes. High Voltage Thin Layer Devices (RESURF devices). In *Electron Devices Meeting*, pages 238–241, 1979.

[BAG+82] B.J. Baliga, M.S. Adler, P.V. Gray, R. Love, and N. Zommer. The Insulated Gate Rectifier (IGR): A New Power Switching Device. In *IEEE International Electron Devices Meeting Digest*, pages 264–267, 1982.

[Bal92] B.J. Baliga. *Modern Power Devices*. Krieger, 1992.

[Bal96] B.J. Baliga. *Power Semiconductor Devices*. PWS, 1996.

[BGG99] V. Benda, J. Gowar, and D.A. Grant. *Power Semiconductor Devices*. John Wiley & Sons, 1999.

[CS96] C.Y. Chang and S.M. Sze, editors. *ULSI Technology*. McGraw-Hill, 1996.

[Gha77] S.K. Ghandhi. *Semiconductor Power Devices*. John Wiley & Sons, 1977.

[RGGN83] J.P. Russell, A.M. Goodman, L.A. Goodman, and J.M. Neilson. The COMFET — A New High Conductance MOS-Gated Device. *IEEE Electron Device Letters*, EDL-4(3):63–65, March 1983.

[Sze81] S.M. Sze. *Physics of Semiconductor Devices*. John Wiley & Sons, 1981.

[Sze88] S.M. Sze, editor. *VLSI Technology*. McGraw-Hill, 1988.

[Wan66] S. Wang, editor. *Solid-State Electronics*. McGraw-Hill, 1966.

1 Introduction

Some History

Although the first patents on the principle of the field effect transistor (FET) appeared in the early 30s of the previous century, it was not until shortly after the second world war that the first silicon bipolar transistor was actually made. Some people refer to this event as the start of the first electronic revolution, thereby tacitly assuming that there is a second one too. This might be somewhat exaggerated, and the term evolution might be more appropriate. The major steps during this evolution are—from our point of view—the announcement of the thyristor in 1956, the fabrication of the first metal-oxide-semiconductor FET (MOSFET) in 1960 (so it took about 30 years to overcome technical problems), the development of the power MOSFET in the 70s and of the insulated gate bipolar transistor (IGBT) in the 80s.

This electronic evolution has two faces: on the one hand there is the information processing electronics with its typical constantly decreasing transistor dimensions (also referred to as the complementary MOS or CMOS technology). As a result, this technology is now capable of integrating billions (yes, billions) of transistors in one chip of only a few square centimeters large. The sheer pace of this electronic evolution is the more striking if we realize that the first bipolar transistor ever made (in 1947) was actually a few square centimeters large all by its own. But what is more, after half a century of industrial evolution, the face of the world is still inevitably changing: who could nowadays imagine a world without computers, mobile phones, satellites...

On the other hand, there is the power electronics, albeit less apparent at first sight, therefore not less important. The emergence of the power semiconductor technology enables the efficient control and conversion of electrical energy. The power semiconductor technology started in the early 50s with the development of the power bipolar transistor. Even

though this is almost simultaneously with the information processing technology, the power semiconductor technology has been outshined by the impressive CMOS technology ever since. Nevertheless, it has somewhat followed its own way, by the development of the various power devices and has recently, or will in the near future (depending on the viewpoints) come out of the shadows.

This is what some call the second electronic revolution: the advent of smart power technology. It is, in fact, the merging of control of energy—the power semiconductor technology—with the information processing capabilities—the CMOS technology. These smart power technologies do already exist and are used in countless applications where motor control is important (from printers to car engines). It is expected that this smart power technology will have the same social impact as CMOS technology. Moreover, its environmental impact can be huge, as it has been documented that over 70 % of all electricity used in the world flows through one or more power semiconductor devices. Imagine the enormous savings in power losses when this huge flow of power could be controlled more efficiently.

How much power ? Which technology ? Which devices ?

The previous section took a quick glance at what exists in electronics: from the CMOS technology with its tiny, extremely fast transistor to the power technology with its thyristor, slow and extremely big, but capable of blocking thousands of volts and conducting thousands of amperes. The power technology itself again consists of a vast domain of applications (see figure 1.1).

Like most Ph. D.s, this work will be situated in a small part of this broad spectrum, at the frontier of CMOS technology with the power technology: the smart power technology. In this smart power technology, we will only focus on one of the two classes of power devices: the various switches—and we will not focus on the rectifiers (see Chapter 2). Of course, one realizes that it is absurd to speak about integration on chip of the example of the extremely big thyristor given above, which on itself sometimes takes up a whole wafer (it is then inevitably a discrete device; that is, a device that is not integrated on chip with other devices). It is obvious that the devices qualifying for integration in a CMOS technology are limited in blocking voltage and current capability. It is also clear

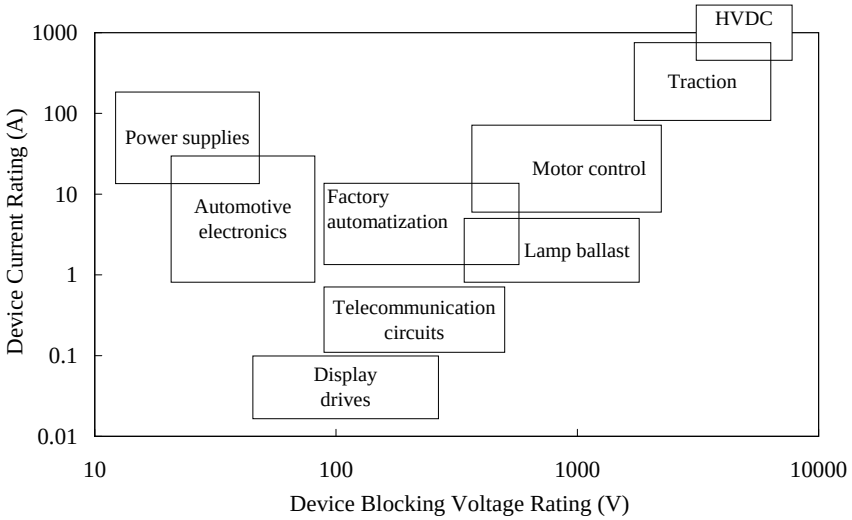


Figure 1.1 Applications for power devices (after Baliga [Bal96b]).

that it is redundant to try to integrate all existing types of power devices on chip, as some of them are especially designed for extreme situations.

This brings us to the problem of the comparison between the different power devices. This is important, as a method is needed to define the most efficient device. The outcome of this will not only depend on voltage and current ranges; but also on the various criteria possible. In power MOSFETs, the figure of merit will mostly be the specific on-resistance versus the breakdown voltage. But when comparing power MOS devices to IGBTs, other trade-offs will have to be made, as the IGBT does not behave as a resistor in its on-state, but has an exponential turn-on, with an extra threshold. Moreover, sometimes other criteria are used, such as the safe operating area (SOA), the device's temperature dependence, the device's degradation behaviour... Once in a while these devices are even used to serve other purposes—e.g., electrostatic discharge (ESD) protection—and have to be designed as such.

Sometimes the final application will determine the nature of the technology, as is the case for chips operating in high temperature environments, where silicon on insulator (SOI) technology is preferred, which has its important impact on the design of the power devices. This is also linked to the problem of the isolation of the power devices to the digital CMOS part of the chip. The reason being that in an SOI technology the electrical isolation is ideal because of the presence of the buried oxide, which isolates the devices from the substrate, and of the trenches, wich

isolate devices from each other. However, other problems, mostly temperature related, do arise. In a junction isolation (JI) technology (as in this work), the electrical isolation is provided through the use of reverse biased pn junctions, and is a possible cause of latch-up and cross-talk. It is often necessary that the entire power device is lifted in potential in comparison with the immediate surroundings. These floating devices need extra isolation when compared to their non-floating counterparts. These extra necessities are sometimes difficult to fullfill, and is one of the reasons why IGBT devices are seldom integrated in CMOS JI technologies.

One understands that a complete picture is needed before the CMOS technology is extended with process modules for power devices. One needs to know the different applications that will be targeted for, and the various needs for each of these applications. Not to mention the issue cost, as some applications have several possible realizations, but then the cost will be the decisive factor. This is however not a matter that will be discussed here, and we will limit ourselves to the statement that the technology is based on a 0.35 µm standard, junction isolated CMOS technology, that the field of applications is the automotive and consumer electronics, that the voltage range is roughly between 10 and 100 V, that the current range is anything below 10 A and that the switching times are seldom faster than 10 ns. This limits the number of power switches available, and based on what is given in [Bal96a, Chapter 10] (see also conclusion of this work) the power MOS devices are the most efficient devices in these ranges. Nevertheless, the possibility of integrating the IGBT will be studied as well.

Why TCAD ?

The major advantage of technology computer aided design (TCAD) is that one can create single devices without the costly need to actually make the devices. One can try out different concepts and TCAD anticipates on which options are feasible and which are not before any real silicon is out. Another major advantage of TCAD is that it gives insight in the distribution of the physical quantities on a 2D plot. One can actually *see* what is happening *in* the device when such and such biases are applied. This helped more than once to solve problems encountered in existing devices. TCAD is often used in literature to discuss, understand and analyze problems of all kind. Yet another advantage of TCAD is again cost driven. Once the core process flow is known and calibrated in

TCAD, one can develop new devices with great reliability, even if one or two new process modules have to be defined. Devices can be developed without the need for a second or third try; in other words, the chance of being *first time right* greatly increases. Even the extraction of SPICE parameters can be done before any real device is measured (which actually happened on the pLDEMOS in I3T80 presented in this work). All of this results in a drastic decrease of the time to market, and thus of the cost.

A condition sine qua non is that TCAD simulations are reliable. The remaining scepticism in industry is due to this calibration issue, the reason being twofold. The first problem is the adequate determination of 2D dopant profiles because of the lack of 2D calibration material (SIMS). The second problem is the constant evolution of technology (new materials and new physical phenomena as the scaling towards smaller dimensions goes on). However, both objections do not longer hold in our case. The integration of power devices happens in a technology that is already well-known, and—due to the nature of power devices (their larger dimensions)—the need for very fine calibration of 2D dopant profiles hardly ever occurs. Nevertheless, calibration, and the related meshing problems, need to be addressed. Therefore, one chapter in this work (Chapter 3) will be entirely dedicated to this.

Objective

The ojective of this work is to design and develop power devices as an extension to a standard CMOS technology and to think up new concepts ameliorating the performance of these devices. The criteria used will be explained in due course and are mainly very simple: the breakdown voltage versus specific on-resistance or versus dissipated power. One important figure that appears in all these trade-offs is area—which is again the factor cost showing up. The smaller the device, the less silicon is used, the cheaper the chip. To achieve this, the power dissipation has to be decreased as much as possible. Otherwise, with decreasing area, one would generate too much power on a too small area, which inevitably results in breakdown. Decreasing the power dissipated means that we decrease the power loss in the switch. As 60 to 70 % of all electricity passes through one or more power devices, this means a more efficient control of electrical energy. Or how a basically cost driven motive can result in an ecologically favourable trend...

Outline

Since the focus in this work is on the understanding and the design of the power switch using TCAD simulations, most of the time we will concentrate on device simulation. Therefore, Chapter 2 gives an overview of the general physical framework for numerical TCAD device simulations. For a treatment of the process, the semiconductor physics, and the basic device physics, we refer to textbooks. Chapter 2 gives, however, a short reminder of some topics that are so important for power devices, that they can not be omitted; as well as an overview of some phenomena that might be lesser known (punch-through, reach-through, second breakdown, conductivity modulation, RESURF effect, Kirk effect...). Concerning the power devices, many excellent textbooks exist that explain their operation principles through analytical expressions (see Chapter 2: the books on power devices by S.K. Ghandi and B.J. Baliga). But, as is written in [Bal96a, Preface, p. viii]:

> "For a complete characterization of the electrical properties of devices, it is necessary to resort to numerical techniques using computer programs that solve the fundamental semiconductor transport equation in two dimensions (and sometimes in three dimensions), with time dependence included for the transient case."

This is in a nutshell what will be done in the core chapters of this book. But before we embark on this, the important matter of TCAD simulation and calibration needs to be treated (Chapter 3). The following chapters deal with the actual simulation, design, and measurement of power devices: Chapter 4 on the power MOS and Chapter 5 on the IGBT. Finally conclusions are made in Chapter 6 by comparing the different devices with each other and with devices described in literature.

References

[Bal96a] B.J. Baliga. *Power Semiconductor Devices*. PWS, 1996.

[Bal96b] B.J. Baliga. Trends in Power Semiconductor Devices. *IEEE Transactions on Electron Devices*, 43(10):1717–1731, October 1996.

2 Fundamental Considerations

2.1 Introduction

It is not the intention of this book to serve as an overview for all process, semiconductor and device physics needed for TCAD simulations. Therefore, the next section refers to books where all subjects related to this work are treated in various ways. We will, however, give a short overview on the general framework of device simulation as it is important to have an idea about the physical framework we are working in while performing device simulations. It is considered less suitable for this work to do the same for the process simulations, since most of our attention will go to device operation and not to process related physics. As a simple and maybe a little odd example of device simulation, a silicon bar is treated. At first sight this seems a trivial case, but no such thing. It introduces some of the major basic topics when dealing with power devices: velocity saturation, avalanche generation, conductivity modulation...

After this non-device, a definition of the power device and the characteristics of the ideal power device is given. Then, the two classes of power devices are presented: the rectifiers and the switches. Less thought is given to the rectifiers, as this is not the subject of this work. Although they are important because they show up as integrated in power switches. Therefore, only some very important plots with basic characteristics as breakdown voltage versus doping level are given; for details about their working, we refer to the textbooks on power devices. Then, the power switches are discussed, consisting of a classification on the type of gate control. For each device, a brief summary on when, why, and where it is used is given. The last section gives an overview of the basic concepts encountered when dealing with integrated power devices. It is solely intended as an introduction to some topics that might be lesser known, a sort of synopsis of designing concepts and of fundamental (power) device physics. Most of these issues will be treated in detail in the different chapters on power devices.

2.2 The Physics

2.2.1 Process, semiconductor and device physics

As has been mentioned above, we refer to the different books that served us well during the course of the work presented in this book. For the process related physics, these are *VLSI Technology* by S.M. Sze [Sze88] and *ULSI Technology* by C.Y. Chang and S.M. Sze [CS96]. For the semiconductor physics in general and the physics of the semiconductor devices, *Solid-State Electronics* by S. Wang [Wan66] and *Physics of Semiconductor Devices* by S.M. Sze [Sze81] are the reference works. The entire Modular Series on Solid-State Devices edited by R.F. Pierret and G.W. Neudeck is also recommended, as it treats some of the subjects in great detail, e.g., *Volume X* on the *Fundamentals of Carrier Transport* by M. Lundstrom [Lun92], which served as a basis for the next section. Also Kano's *Semiconductor Devices* [Kan98], as well as some books in the Modular Series on Solid-State Devices ([Neu89b], [Neu89a], and [Pie90]) can be used as an introduction to device physics. For details about the various models used in process and device simulation (e.g., for the mobility models), we refer to the manuals of the ISE software [ISE00].

For the physics of semiconductor power devices in particular, we refer to such excellent textbooks as *Semiconductor Power Devices* by S.K. Ghandhi [Gha77], *Modern Power Devices* [Bal92] and *Power Semiconductor Devices* [Bal96] by B.J. Baliga, and *Power Semiconductor Devices* by V. Benda, J. Gowar and D.A. Grant [BGG99].

2.2.2 General framework for device simulation

A fundamental way of treating carrier transport in a semiconductor material is by solving the Schrödinger Equation

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m_0}\nabla^2\Psi + \Big[E_{C0}(\mathbf{r}) + U_C(\mathbf{r}) + U_s(\mathbf{r},t)\Big]\Psi(\mathbf{r},t) \qquad (2.1)$$

which describes the electron by its wave function $\Psi(\mathbf{r},t)$ propagating through the material under the influence of the built in or applied (E_{C0}), the crystal (U_C), and scattering potentials (U_s). The quantity $\Psi^*(\mathbf{r},t)\Psi(\mathbf{r},t)d\mathbf{r}$ then gives the probability of finding the electron between $\mathbf{r}$ and $\mathbf{r}+d\mathbf{r}$. Solving equation (2.1), which is already a "simplification" of the many-body problem, is not an easy task and therefore further assumptions are needed for practical device simulation.

When the size of the device is large enough (as will certainly be the case for the devices described in this book), the various electron waves don't interfere and a classical approach is appropriate. The quantum mechanics is treated indirectly by the use of an effective mass or an energy band structure; carrier scattering is treated quantum mechanically. This is known as the semiclassical treatment, where carriers behave as particles, moving under influence of a force $\mathbf{F}$ (a combination of applied or built in forces and random forces due to impurities and lattice vibrations); and obey Newton's laws

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}(\mathbf{r}, \mathbf{p}, t) \tag{2.2}$$

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i(t). \tag{2.3}$$

Each of the $i = 1, \ldots, N$ carriers has a position $\mathbf{r}_i$ and a momentum $\mathbf{p}_i$, which are known by solving the previous equations. Alternatively, we could ask: what is the probability of finding an electron with crystal momentum $\mathbf{p}$ at position $\mathbf{r}$, at time t ? The answer is given by the distribution function $f(\mathbf{r}, \mathbf{p}, t)$, determined by the famous *Boltzmann Transport Equation*:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \mathbf{F} \cdot \nabla_p f = s(\mathbf{r}, \mathbf{p}, t) + \left.\frac{\partial f}{\partial t}\right|_{coll}. \tag{2.4}$$

The quantum mechanics enters via the evaluation of the $E(\mathbf{p})$ relation, which is used to calculate the carrier velocity $\mathbf{v} = \nabla_p E(\mathbf{p})$; and in the scattering rates $s(\mathbf{r}, \mathbf{p}, t)$, which represents the actual carrier generation-recombination processes such as impact ionization and recombination through defects (Shockley-Read-Hall recombination) and $\partial f/\partial t|_{coll}$, which represents collisions sending carriers from one momentum state to another.

Solving the Boltzmann Transport Equation directly is still extremely difficult and further assumptions are needed to move on. One way to solve the Boltzmann Transport Equation is by using the Monte Carlo numerical technique. This technique directly mimics the physical processes that occur during carrier transport by simulating individual particle trajectories, determined by choosing random numbers (properly distributed to reflect the probabilities of the various scattering effects). If the number of simulated trajectories is large enough, then the average results provide a good approximation of the behaviour of the carriers within a real device. Although this method has its limitations (of which

statistical noise is the most important one), it is most often the standard against which the validity of simpler approaches to solving the Boltzmann Transport Equation is gauged.

The simplest way for solving the Boltzmann Transport Equation is by generating balance or conservation equations, which are directly derived from it. These balance or conservation equations are derived by introducing one unknown at a time, beginning with the electron density:

$$n(\mathbf{r},t) \equiv \frac{1}{\Omega}\sum_{\mathbf{p}} \phi(\mathbf{p}) f(\mathbf{r},\mathbf{p},t) \text{ with } \phi(\mathbf{p}) = 1. \tag{2.5}$$

Using (2.5), one can derive the first balance equations (two equations if the same is done for holes), better known as the continuity equations:

$$\frac{\partial n}{\partial t} = G_n - R_n + \frac{1}{q}\nabla \mathbf{J}_n \tag{2.6}$$

$$\frac{\partial p}{\partial t} = G_p - R_p + \frac{1}{q}\nabla \mathbf{J}_p \tag{2.7}$$

where G_n and G_p are the electron and hole generation rates and R_n and R_p are the electron and hole recombination rates for n-type and p-type semiconductors, respectively. Each of these balance equations introduces another unknown, the electron and hole current densities $\mathbf{J}_n$ and $\mathbf{J}_p$, for which a new balance equation can be sought. However, the balance equation for the current density (the momentum balance equation) introduces on its turn another unknown, the tensor $\overleftrightarrow{W}$, whose trace can be interpreted as the kinetic energy density. The balance equation for the carrier energy (the energy balance equation) introduces the energy flux, and so forth. No matter how many balance equations are derived, they always contain one more unknown than the number of equations. The only way out is by introducing simplifications that truncate the set of equations. The best known example is an extremely rough truncation of the momentum balance equations, the well-known drift-diffusion equations (or current density equations or transport equations), see e.g., [Sze81, chapter 1]:

$$\mathbf{J}_n = q\mu_n n\mathbf{E} + qD_n\nabla n \tag{2.8}$$

$$\mathbf{J}_p = q\mu_p p\,\mathbf{E} - qD_p\nabla p \tag{2.9}$$

for electrons and holes, respectively. The first term accounts for the drift, caused by the electric field $\mathbf{E}$ and the second term accounts for the diffusion caused by the carrier concentration gradient (∇n and ∇p). The

electron and hole mobilities (μ_n and μ_p) values are given by means of various models (depending on the situation under study), and determine the diffusion constants (D_n and D_p) for nondegenerate semiconductors through the Einstein relationship ($D_n = (kT/q)\mu_n$, etc.).

These balance equations (the continuity and current density equations), together with the Maxwell equations, give a complete description of the dynamics of electrons and holes in a semiconductor under the influence of external fields. For the devices presented in this book, only one Maxwell equation is needed, the Poisson equation:

$$\nabla\big(\epsilon_s \cdot \mathbf{E}\big) = \nabla\big(\epsilon_s \cdot (-\nabla\psi)\big) = \rho = q\,(p - n + N_{D^+} - N_{A^-}) \qquad (2.10)$$

where ϵ_s is the semiconductor permittivity, ψ is the electrical potential, ρ is the space charge density, and N_{D^+} and N_{A^-} are the densities of ionized donors and acceptors, respectively.

A complete description of the problem is obtained by using only 3 independent variables (the electron concentration n, the hole concentration p and the potential ψ), by introducing "secondary" models for the mobility, generation and recombination; and by defining the necessary boundary conditions. These "secondary" models, based on quantum mechanical calculations or experiments, only use the 3 independent variables.

The major part of the device simulations takes place within this framework, wich will be referred to as the drift-diffusion model. Smaller or larger variations to the set of equations as discussed above can be introduced by truncating the sequence of balance equations in a different way.

A simple variation to the current density equations (2.8) and (2.9) is easily presented through the use of the electron and hole quasi-Fermi potentials, ϕ_n and ϕ_p, which are linked to the carrier concentrations and the potential by the Boltzmann statistics:

$$n = n_{i,\mathit{eff}} \exp\left[\frac{q(\psi - \phi_n)}{kT_L}\right] \qquad (2.11)$$

$$p = n_{i,\mathit{eff}} \exp\left[\frac{-q(\psi - \phi_p)}{kT_L}\right] \qquad (2.12)$$

where $n_{i,\mathit{eff}}$ is the effective intrinsic density and T_L is the lattice temperature. The current densities are then given by:

$$\mathbf{J}_n = -q\mu_n n \nabla\phi_n \qquad (2.13)$$

$$\mathbf{J}_p = -q\mu_p p\, \nabla\phi_p \qquad (2.14)$$

as one can easily verify by substituting ϕ_n and ϕ_p, using the equations (2.11) and (2.12) in (2.13) and (2.14), respectively. If $\nabla n_{i,\mathit{eff}} \neq 0$, an extra term in the current densities is introduced that accounts for the gradient in the effective intrinsic carrier concentration (bandgap narrowing effects). Still further, in a more general consideration, the Fermi-Dirac distribution function can be used instead of the Boltzmann function. This introduces important changes to the current density equations (and the energy flux equations, see below).

Yet another approach, which is a less rough truncation of the momentum balance equation, has to be discussed as it will be used in a small part of the device simulations presented in this book. In the truncation as presented above, it has been assumed that the carrier temperatures, T_n and T_p, are given by T_L *and* that this temperature is constant. A straightforward extension is to assume that the carriers are still in thermal equilibrium with the lattice, but the latter's temperature is no longer constant. Hence the name *thermodynamic* or *non-isothermal* model, which extents the drift-diffusion approach to account for electrothermal effects. The result of this different truncation of the current density equations is an extra term, which accounts for the flow of current due to the gradient of temperature:

$$\mathbf{J}_n = -q\mu_n n\big(\nabla\phi_n + P_n\nabla T_L\big) \tag{2.15}$$

$$\mathbf{J}_p = -q\mu_p p\,\big(\nabla\phi_p + P_p\nabla T_L\big) \tag{2.16}$$

with P_n and P_p as the absolute thermoelectric powers and the lattice temperature given by the lattice heat flow equation:

$$c\frac{\partial T_L}{\partial t} = \nabla\big(\kappa\nabla T_L\big) + H \tag{2.17}$$

where c is the heat capacitance per unit volume, κ is the thermal conductivity and H is the heat generation given by [Wac90] (steady-state case, without electromagnetic radiation):

$$\begin{aligned} H =& \frac{\left|\mathbf{J_n}\right|^2}{q\mu_n n} + \frac{\left|\mathbf{J_p}\right|^2}{q\mu_p p} \\ &+ q(R - G)\big[\phi_p - \phi_n + T_L(P_p - P_n)\big] \\ &- T_L\big(\mathbf{J_n}\cdot\nabla P_n + \mathbf{J_p}\cdot\nabla P_p\big) \end{aligned} \tag{2.18}$$

where the first two terms are the Joule heat of electron and holes, the third term is the recombination and generation heating and cooling, and the last term accounts for the Peltier and Thomson effects.

The extra driving terms for the current densities in equations (2.15) and (2.16) have to be included when accurate simulation for self-heating effects is desired. A simplified model of the thermodynamic model is defined when both the thermoelectric powers are set to zero. Then, the lattice heat flow equation is still solved and the lattice temperature is still used in the various "secondary" models.

With ever decreasing device dimensions, even this thermodynamic model does no longer suffice for deep submicron device simulations. As the thermodynamic model already makes fully use of the momentum balance equation, the third moment of the Boltzmann Transport equation, the energy balance equation, has to be introduced. As this set of equations is similar to that which describes the flow of fluids, this approach is often referred to as the *hydrodynamic* model. We do not enter into the details of this approach, as this model will be rarely used in this book. We simply mention that the current density equations are no longer truncated and thus each of them (for electrons and holes) introduces a new unknown, the tensor $\overleftrightarrow{W}$, for which we write down two new balance equations (the energy balance equations). These two new balance equations on their turn introduce two new unknowns and the system has to be truncated as has been done for the momentum balance equations.

One understands that many variations of this truncation (and thus of the hydrodynamic model) exist. The version included in the device simulator used in this book exists of 6 partial differential equations [Str62]. In this case, the new unknown tensor in the current density equations is reduced to an isotropic tensor, represented by the carrier temperature as unknown (for both electrons and holes). These carrier temperatures (T_n and T_p) are directly related to the carrier kinetic energy densities, for which two new energy balance equations are written down, both containing a new unknown, the electron and hole energy (or heat) flux. To close the system, these energy fluxes are approximated in terms already known (J_n, T_n, J_p and T_p). The energy balance or conservation equations for the electron and hole temperature are accompanied by the energy conservation equation for the lattice temperature (actually the lattice heat flow equation (2.17)). These 3 equations together with the Poisson equation and the electron and hole continuity equations make up the whole set of 6 partial differential equations.

Example: a silicon bar

The different models as described above are now applied to the most simple of structures conceivable, the silicon bar. This example might seem to be too obvious at first sight. However, the non-linear behaviour of figure 2.1 does call for some more explanation. This simple example introduces already quite some subjects that are vital for the understanding of the operation of power devices.

Consider a silicon bar, doped with phosphor ($N_{D^+} = 1e15\,\mathrm{cm}^{-3}$), with a length $l = 10\,\mu\mathrm{m}$. A potential difference V_{ak} is applied between the top (anode) and bottom (cathode) contact, both of which are ohmic. The current density versus applied voltage is simulated using the drift-diffusion model together with the doping-dependent, high-field saturation model for the mobility, the Shockley-Read-Hall and Auger recombination models, and the avalanche generation model according to van Overstraeten - de Man. These are the so-called "secondary" models mentioned in the previous section. Figure 2.1 shows the current density versus V_{ak} and this may seem odd at first sight. However, some basic observations clarify a great deal of it. In the beginning, the current flow has to be ohmic, as given by (2.8) (no diffusion and $\|\mathbf{E}\| = E = V_{ak}/l$ as there is no net space charge, since $n = N_{D^+}$):

$$\|\mathbf{J}_n\| = J = q\mu_n N_{D^+} V_{ak}/l = 224 V_{ak}\,\mathrm{A/cm}^2. \tag{2.19}$$

With increasing V_{ak}, E increases and the mobility becomes inversionally proportional to the electric field. This is known as the velocity saturation (since $v_n = \mu_n E$) and occurs at $v_{n,sat} = 10^7\,\mathrm{cm/s}$, or at a current density

$$J_{sat} = q v_{n,sat} N_{D^+} = 1600\,\mathrm{A/cm}^2. \tag{2.20}$$

Eventually, ohmic behaviour changes to velocity saturation behaviour when $J = J_{sat}$; that is, at $V_{ak} = 7\,\mathrm{V}$. This is confirmed by Figure 2.1, which indicates saturation over a large voltage range. At some point, the electric field is so strong in the silicon bar that impact ionization occurs. This onset of avalanche generation can be calculated quite easily when the following assumptions are made: the total current density is only due to drift, which happens at the saturation velocity for both carriers (diffusion is negligible), the recombination rate is negligible in comparison with the generation rate and the space charge is also negligible. Using the expressions for the electron current density (2.8) without the diffusion term and with the saturation velocity, and the continuity equation (2.6) with the assumptions made above, results in a linear

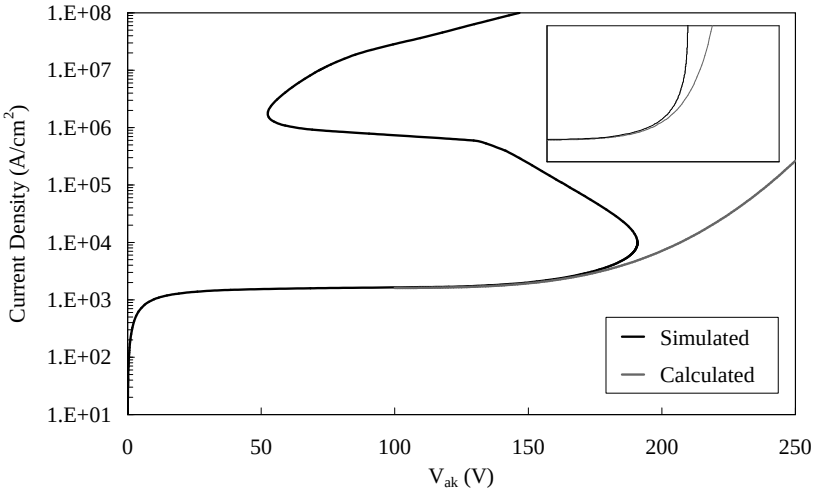


Figure 2.1 Current density versus applied voltage for a silicon bar, simulated with the drift-diffusion model and calculated as explained in the text. The insert is a detail around the first breakdown on a lin-lin scale.

differential equation of the first order in n:

$$G_n = -v_{n,sat}\frac{dn}{dx} \tag{2.21}$$

with

$$G_n = n\alpha_n v_{n,sat} \tag{2.22}$$

and

$$\alpha_n = a_n e^{-b_n/E} = a_n e^{-lb_n/V} \tag{2.23}$$

where a_n and b_n are two constants determined by the Van Overstraeten - De Man model. Solving this differential equation with the boundary condition that the concentration of electrons at the cathode is equal to the background doping level and the assumption that all current at the anode is due to electrons only ($J_{tot} = J_n(x = 0)$), results in a characteristic as is plotted in figure 2.1.

The calculated characteristic can not track the snapback with the assumptions made above. Indeed, when the current density continues to increase, the generated hole density exceeds the background doping level and the space charge can no longer be neglected. This makes it more difficult to solve the problem analytically, as now the poisson equation needs to be taken into account with both continuity equations. However, based on the numerical simulations, we can understand what

is going on. With increasing current density, holes travel to the cathode and electrons to the anode, resulting in a negative space charge towards the anode and a positive space charge towards the cathode. The mid-region of the bar is flooded with electrons and holes, and the resistivity of the bar drops. This is what is known under the term conductivity modulation, a phenomenon that is normally described in power bipolar transistors under high-level double injection conditions. Consequently, the electric field drops in this mid-region, and the impact ionization process shifts towards the anode and cathode. This process explains the negative resistance as is seen in figure 2.1. With further increasing current density, the electric field continues to drop in an increasing part of the silicon bar, now confining the avalanche generation to fine strips besides the cathode and the anode. The electron and hole densities, now both several orders of magnitude above the background doping level, are equal along a large part of the silicon bar, and start to recombine at increasing rates. At a certain moment the recombination rate competes with the impact ionization rate, and a positive resistance results as the current density reaches high levels.

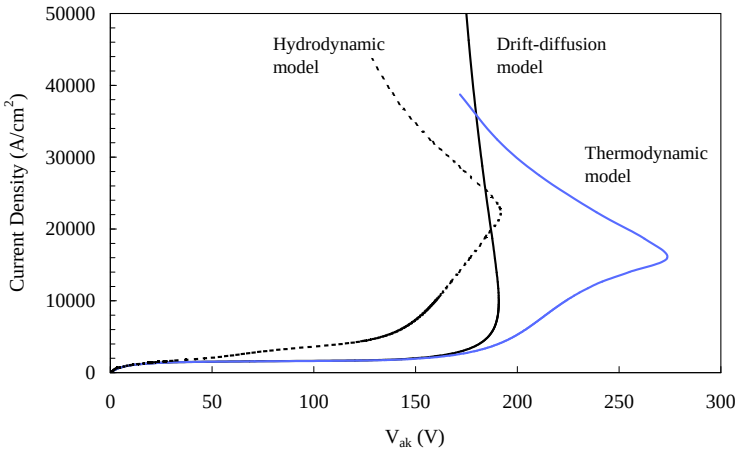


Figure 2.2 Current density versus applied voltage for a silicon bar, simulated with the drift-diffusion model, the thermodynamic model and the hydrodynamic model.

These extremely high current densities seem to be unrealistic, and call for a more realistic simulation. Taking into account the lattice temperature (and eventually the carriers temperatures), and telling the simulator to stop at the melting point of silicon, is a more realistic way of treating the high current density region above breakdown. The result

is shown in figure 2.2, where the output characteristic of the silicon bar, simulated with the three different models based on the moments of the Boltzmann transport equation are plotted against each other. Unfortunately, the real output characteristic is not at our disposal, but this will be the case when the same models will be applied on several power devices. Here, it is only our intention to illustrate the influence of the temperature on the output characteristics.

2.3 What is a power device ?

A power device controls the flow of power to a load. This is most often done by switching the device on a periodic basis to generate pulses of current through the device. The current and voltage waveforms are shown in figure 2.3 for the ideal case.

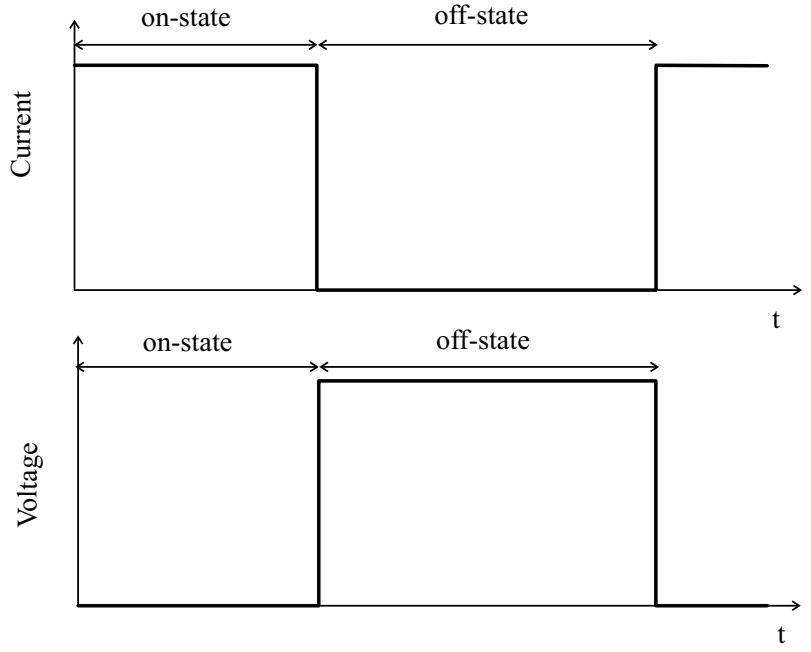


Figure 2.3 Current and voltage waveforms of an ideal power device.

The ideal power device thus has no voltage drop when the device is conducting current (no on-state power dissipation), has no current flowing through the device when it is supporting voltage (no off-state power dissipation), and can switch between the on- and off-state instantaneously (no power dissipation during switching). The ideal power

device is therefore a device that does not dissipate any power itself and passes all power to the load. The load may be resistive, capacitive, or inductive. The power devices are divided in two main categories: the power rectifier (or diode) and the power switch. The latter is able to control the amount of power that is going through, the former is not.

2.4 Rectifiers

Although this work does not treat the rectifiers, it will become clear (in the next chapters) that power switches contain a lot of diodes. Especially in the junction isolated technologies as used in this book, a thorough knowledge of the diode is of vital importance. Therefore, a short note on the very basic pn diode (when to denote it as a "power" diode is a matter of taste), on the very important punch-through diode, and on the existing power rectifiers is given here. For analytical expressions and details about working principles, we refer once again to the text books on power devices (see above).

2.4.1 The ideal rectifier

The ideal rectifier blocks all current in the off-state and conducts with zero resistance in the on-state (figure 2.4). As an ideal device, it supports an infinite voltage, and it conducts infinite current; while being capable of switching between those two states at infinite speed.

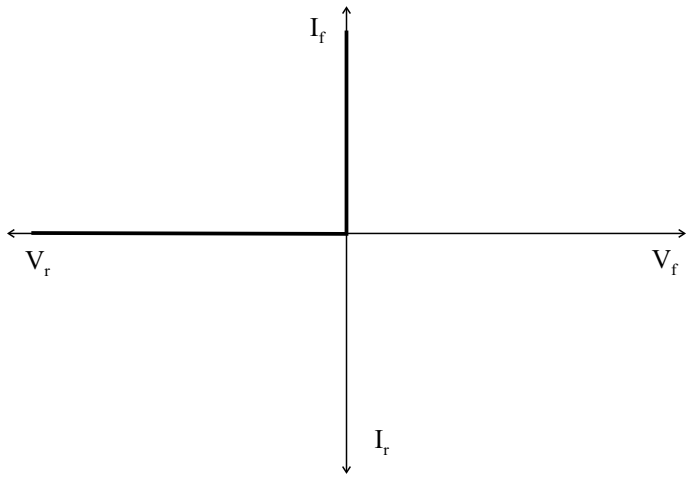


Figure 2.4 Output characteristic (forward and reverse) of an ideal diode.

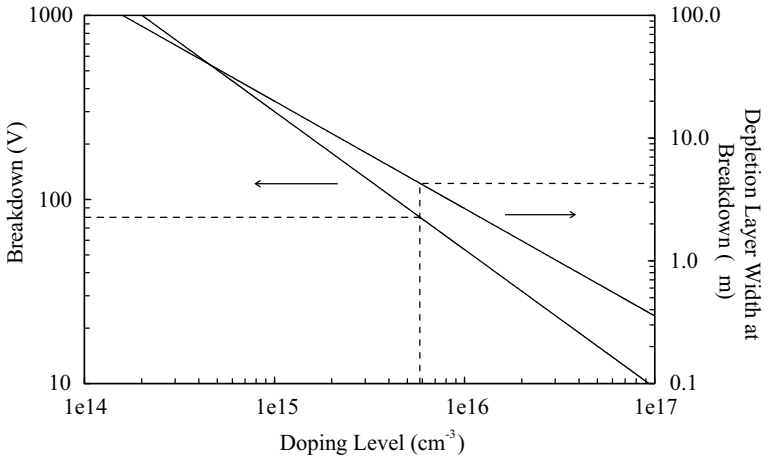


Figure 2.5 Breakdown voltage and depletion layer width for an abrupt parallel-plane n^+/p junction diode versus the doping level of the p side.

2.4.2 Junction diode

A real diode has finite blocking, conducting and switching capabilities. These are denoted by the parameters breakdown voltage (V_{br}), on-state voltage drop (V_{on}), turn-on (t_{on}), and turn-off (t_{off}) times. In addition, it has a non-zero leakage current, causing power dissipation in the off-state.

As an example, we choose a n^+/p abrupt junction diode with a breakdown voltage of 80 V. Figure 2.5, which plots breakdown voltage and depletion layer width at breakdown versus the doping level of the p side, shows that a doping level of $5.8e15\,\mathrm{cm}^{-3}$ is the maximum level possible (higher doping levels give lower breakdown voltages) if 80 V needs to be reached. It also shows that the depletion layer width at breakdown is around 4.3 µm. If this needs to be used in a device— say in the vertical direction—a thickness of at least 4.3 µm for the p layer needs to be foreseen, otherwise the formation of the depletion layer is altered by the next layer (most of the times this means that the depletion layer is stopped). If this next layer is of the opposite type of the p layer, then a npn transistor has been created and reach-through occurs (see further). If the next layer is of the same type of the p layer and higher doped (if lower doped then the depletion layer just keeps on increasing with increasing voltage), then a punch-through diode is created.

2.4.3 Punch-through diode

The very important plots for the breakdown voltage of the abrupt pn junction diode and of punch-through diodes of various widths are shown in figure 2.6. A punch-through diode is a junction diode where the depletion layer at reverse bias is stopped at one side (might also be stopped at both sides) by a higher doped region of the same type (p^+ or n^+) of that side of the diode, with a decrease of the breakdown voltage as a consequence.

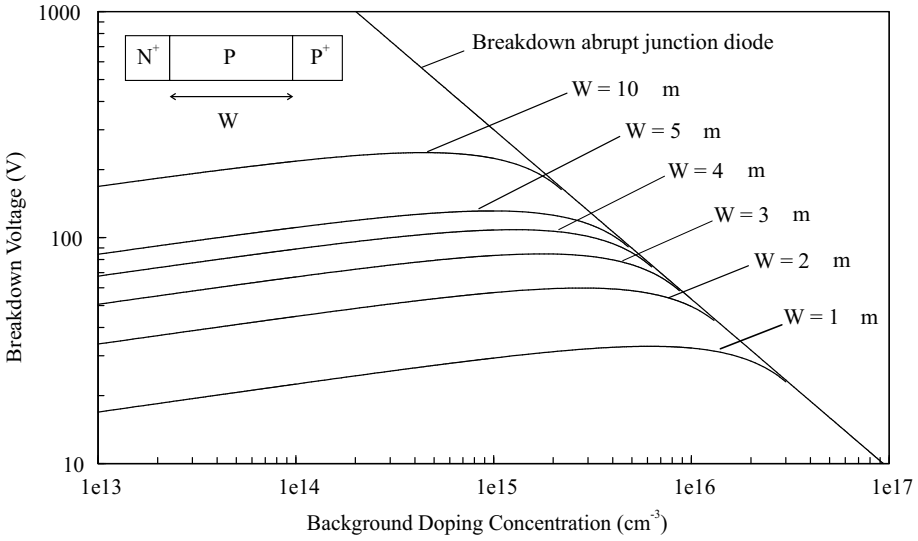


Figure 2.6 Breakdown voltage for an abrupt junction diode and for a punch-through diode of several widths. The insert shows a punch-through diode.

The importance being that if one aims for a blocking voltage, different punch-through diodes can be made with different widths, depending on the background doping concentration. This is a principle that will be used many times while designing power devices. In the example of 80 V, a p layer thickness of 3 µm can be used if the doping level is decreased to $1.8e15\,\mathrm{cm}^{-3}$ in comparison with the doping level needed ($5.8e15\,\mathrm{cm}^{-3}$) when a non-punch-through diode (with a minimum thickness of 4.3 µm) is used.

2.4.4 P-i-N rectifier

The p-i-n rectifier was one of the first power devices created. Its major working principle is conductivity modulation during on-state (see also the silicon bar): the i-region (n or p type) is flooded with minority carriers during current conduction. This yields a device with a very low resistance. In the off-state, the p-i-n diode's blocking capability is determined by the doping level and width of the i-region. Since the lower this doping level is, the lower the on-state voltage drop is, the p-i-n rectifier is normally a punch-through diode. The major disadvantage of this rectifier comes with its working principle: when the device is switched off, the minority carriers in the i-region need time to flow away. This switching off process is referred to as reverse recovery, and, together with the on-state voltage drop, forms the major trade-off for this device.

2.4.5 Other power rectifiers

A further decrease of the on-state forward drop is possible when a pn junction grid is integrated in the drift region. Such devices are called *pinch* rectifiers or *junction barrier Schottky* (JBS) rectifiers.

Another way to improve the on-state forward drop of a rectifier is by using a metal semiconductor contact. Such a contact has a similar non-linear current transport behaviour as a pn diode, which was already described in 1938 by Schottky. In recent years, *Schottky barrier rectifiers* have improved a lot and are likely to replace the p-i-n rectifier in high-voltage power electronic circuits.

Combination of both types of rectifiers results in *the merged p-i-n/Schottky* (MPS) rectifier. Furthermore, power switches (like the power JFET and the power MOSFET, see next section) can also be used as power rectifiers as long as the gate receives a synchronous gate signal. These *synchronous* rectifiers are rarely used and only occur in high-performance systems. Still other classes of rectifiers exist, like the *gallium arsenide* rectifier, which exhibit high-speed swithing charactersitics and rectifiers based upon *silicon carbide* with their high blocking voltage capabilities.

2.5 Switches

2.5.1 The ideal switch

As has been said, the ideal switch has the ability to control the output power. To do this, the switch has one more terminal than the rectifier—the gate, which regulates the output current. See figure 2.7 for the current saturation characteristics. Under this gate control the ideal switch is capable of conducting infinite current in both directions of current flow, and of supporting infinite voltage in both directions of applied bias, see figure 2.7. The gate control signal must be a current with zero voltage drop in the control circuit or a voltage signal with zero current flow in the control circuit in order to maintain zero power losses in the input circuit. Once again, the ideal switch must be able to switch between on- and off-states instantanously.

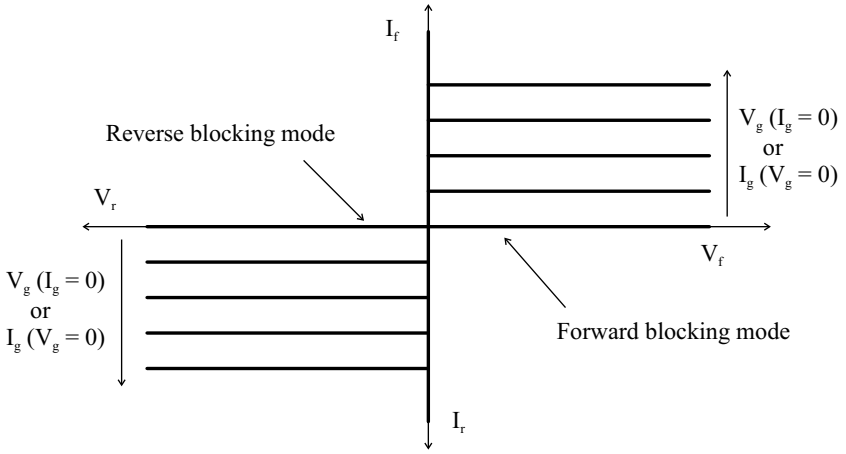


Figure 2.7 Output characteristics (forward and reverse) of an ideal switch.

The switch shown in figure 2.7 is *normally-off*, which means that no current is flowing when there is no gate signal. It is also possible to design *normally-on* switches that conduct current when there is no gate signal.

2.5.2 Current controlled switches

Power bipolar transistor

The power bipolar transistor is the first power switch made, and has long been the only choice in many applications. Witness the fact that in the

book by S.K. Ghandhi [Gha77], written in 1977, only two power switches are described: the power bipolar and the power thyristor. However, from the 70s on, the power bipolar transistor has been gradually replaced by the power MOSFET, the IGBT and the GTO (see further). The main reason for this decline in popularity is the low current gain of the power bipolar transistor due to the large base drive current needed and the complex input circuitry. Other drawbacks are its small safe operating area due to second breakdown, the decrease of the forward voltage drop with increasing temperature (which makes it difficult to parallel these devices), and its slow speed due to its bipolar nature. Therefore, the book on power devices written by B.J. Baliga in 1996 [Bal96] treats the power bipolar only as an introduction to the IGBT.

The only ratings where the power bipolar still holds an advantage is in applications working from (relatively) low voltages (240 V) and requiring switching frequencies in excess of 75 kHz. A recent article on power bipolar transistors claims that the arrival of wide bandgap materials (SiC, GaN...) will reintroduce the bipolar as a power device [HZ01].

Power thyristors

The second oldest power switch is the power thyristor, developed in the 50s and consisting of 4 layers of semiconductor of alternating dopant type (npnp). These devices are unmatched when it comes to current carrying capability per unit area. The price to pay is the loss of gate control when current conduction is at its best—the major difference with the transistors. Many variants of the basic thyristor structure (also called *the silicon controlled rectifier* (SCR)) exist and a lot of them are designed to improve the regain of the gate control when the device is switched off: *the gate-assisted turn-off thyristor* (GATT), *the gate turn-off thyristor* (GTO)... Another popular, related device is the triac, consisting of a 5 layer structure (npnpn, actually two anti-parallel thyristors). It is a bidirectional switch, meaning that is not only capable of blocking current flow in both directions, it is also capable of conducting current in both directions, depending on the gate signal.

Thyristors are mainly used in applications demanding extreme high current conduction and voltage supporting capabilities. They actually determine the outer limits of what is possible with solid-state power electronics.

2.5.3 Voltage controlled switches

Power junction field-effect devices

Although the principle of the JFET (*junction field effect transistor*, also known as the *static induction transistor* (SIT)) was described in 1952 by Shockley, it took until the 70s before the first power JFET was made, mainly due to technological problems. The main advantages of this device over the power bipolar transistor are its higher input impedance, its negative temperature coefficent for the drain current that prevents thermal runaway and a higher switching speed because of the absence of minority carrier recombination.

The major disadvantage of the JFET is that it is a normally-on device. This is the reason why it is relegated to only a few application where its unique characteristics (e.g., the very high dV/dt) are necessary. The same holds for the *static induction thyristor* (SITh) or *field controlled diode* (FCD) or *field controlled thyristor* (FCT), a device derived from the JFET.

Power MOSFET

Built on the successes of the digital technology, the nMOS and the pMOS, the power MOS devices have pushed aside the power bipolar transistors in many applications from the 70s on. Their major advantages are the MOS controlled gate, the unipolarity, the negative temperature dependence and the excellent safe operating area. The power MOS does not need the large base drive current and the complex gate drive circuitry of the power bipolar, due to its voltage controlled MOS gate. This actually means that voltage controlled switches are found to be closer to the ideal situation than current controlled switches. The power MOSFET is not as vulnerable to a second breakdown mode and it has a switching speed that is orders of magnitude faster than for power bipolar devices (because of the unipolarity). Furthermore, power MOSFETs can be easily paralleled as their forward voltage drop increases with increasing temperature.

Due to these 4 main advantages, the power MOSFET has replaced the power bipolar in all applications with low operating voltages (computer peripherals, automotive electronics...). This has not occurred for systems above 200 volts, mainly because of the increasing on-resistance of the power MOSFET with increasing blocking voltage.

Insulated gate bipolar transistor

In search for the ideal power device, the IGBT was invented some 20 years ago and has ever since been used in an increasing number of applications. It combines the best of both the power bipolar and the power MOS devices: high on-state current conduction with low on-state voltage drop and a voltage controlled gate. Unfortunately, this device also inherits some of the drawbacks: its safe operating area is relatively small due to the presence of an inherent parasitic thyristor, it has a slower switching speed than the power MOS (it is a bipolar device !) and its on-state forward voltage drop decreases with increasing temperature.

The IGBT has replaced the power bipolar device in many applications (adjustable speed motor control for air conditioning, numerical controls for factory automation and robotics, and appliance controls) and is recently threatening the GTO thyristor, as the voltage and current ratings of the IGBT keep scaling up.

MOS-controlled thyristors

Yet another class of power switches that tries to combine the advantages of two of the most successful devices: the power MOS and the thyristor. Since the IGBT's on-state forward voltage increases with increasing blocking voltage, these MOS-gated thyristors try to make use of the lower on-state voltage observed in thyristors. They are thus mainly used in very high voltage power switching applications.

One of the main difficulties in creating this device, is the MOS-controlled turn-off of the thyristor once it has entered into its regenerative current conduction mode. Therefore, the most important figure for these devices is the *maximum controllable current density*, above which the gate can no longer turn-off the thyristor. The MOS-controlled turn-on does not seem to be a problem and it has even been demonstrated that it was superior when compared to a conventional thyristor.

Many different MOS controlled devices have been created to improve the MOS-gated turn-off of the thyristor. The first one was the *MOS-controlled Thyristor* (MCT) or the *MOS-controlled Gate Turn-Off Thyristor* (MOS-GTO). Another structure—the *Base Resistance Controlled Thyristor* (BRT) uses a diverter contact that creates an alternative path for the holes from the base. These two devices offer a lower on-state voltage drop and a higher surge current capability than the IGBT, but they lack the controlled turn-on and current saturation capability. But there is another gate-controlled structure that does offer these last

two features—the *emitter switched thyristor* (EST). However, it remains difficult to match the fully gate-controlled output of the IGBT with its wide safe operating area. Therefore, the EST, as the other MOS-gated thyristors, are likely to replace the GTO rather than the IGBT.

2.6 Fundamental Concepts Concerning Integrated Power Switches

The purpose of this section is to provide some kind of overview of the important concepts concerning power switches in integrated circuits. Most of these topics are found in the text books on power devices. Nevertheless, they are introduced and summarized here, as they are key features for understanding the work presented in this book. Most of these subjects will be treated in more detail in the chapters on the different devices.

2.6.1 The silicon limit

In the previous section, we introduced some of the different trade-offs and figures of merit for the different devices. This is what the silicon limit is about, but then for the power MOS devices. As has been pointed out, the most important trade-off for these devices is the blocking voltage (V_{br}) versus the (specific) on-resistance (for a definition, see the chapter on power MOS devices), which actually gives an idea of the power loss during the on-state (compared to the forward voltage drop of the other power devices). If those two figures are plotted against each other, then a universal tool is at hand to compare all power MOSFET devices. For an example, see figure 4.12. One understands that there is a theoretical limit that cannot be crossed. This limit, however, depends on the form of the power MOSFET (vertical DMOS, LDMOS, RESURF LDMOS, DMOS on SOI, super junction DMOS or COOLMOS^{TM}...). Not to mention the difference if another material (e.g., SiC) comes into play. For silicon, these limits are therefore referred to as the *silicon limit*.

2.6.2 RESURF effect

The *RE*duced *SUR*face *F*ield effect is one of the most important designing techniques for power devices. The effect was discovered by accident while experimenting on power diode structures [AV79]. The abstract of that article defines the effect as follows

> *"The application of a somewhat unusual diode structure opens the possibility to make novel kinds of high voltage devices even with very thin epitaxial or implanted layers. In the new structures crucial changes in the electric field distribution take place at or at least near the surface."*

Since then, it has been used in virtually all power devices (RESURF bipolar, JFET, DMOS, IGBT...), as well as in vertical, discrete devices as in lateral, integrated devices. It has been extended with double, triple and multi layer acting RESURF structures (e.g., the COOLMOS structure); with 3D acting RESURF; and towards other technologies (e.g., SOI). An excellent overview of the RESURF technology is given in [Lud00]. The RESURF effect and its drawbacks will be treated in detail in the chapter on power MOS devices.

2.6.3 Punch-through and reach-through

In literature, it is not always obvious at first sight which phenomenon is referred to when one of the above terms is used. In the book by Sze [Sze81], for instance, the terms *punch-through* and *reach-through* are used as synonyms. In the book by Benda [BGG99], only the term punch-through occurs; and in one of the books by Baliga [Bal96] reach-through is defined as open base bipolar transistor breakdown and punch-through is only used when referring to punch-through diodes (see above). This last approach seems to be the most correct one as now punch-through refers to a structure with two different layers of dopants, and reach-through to a structure with three different layers of dopants. Moreover, the terms refer to two distinct physical phenomena as punch-through is breakdown caused by impact ionization (see the section on the punch-through diode); where reach-through is breakdown caused by a depletion layer reaching a region of opposite dopant type, thereby creating a current path for one type of carriers (electrons or holes). Both phenomena will be encountered many times during this work.

2.6.4 Second breakdown, snapback and thermal runaway

The term *second breakdown* refers to breakdown during current conduction. At high current levels and high voltage, a positive feedback loop can cause breakdown at much lower voltages than in the off-state. This can happen in a power bipolar device, but also in any power device containing a parasitic bipolar (so, virtually every power device) with a

decrease of the blocking voltage as a consequence (when compared to the breakdown voltage in the off-state).

This second breakdown effect can also be accompanied with *snapback*, which denotes a sudden reduction in blocking voltage when the current is further increased during conduction. For instance, suppose that at first impact ionization is occurring, with consequently a large increase of the current level, when suddenly the blocking voltage is no longer hold due to a second breakdown effect.

When a pn junction has reached locally a critical temperature, the local current density increases and the current is drawn to the region where the temperature is the highest. The pn junction is then effectively shunted by a small filament of highly conducting intrinsic semiconductor known as *mesoplasma*. This phenomenon used to be called second breakdown (e.g., in the book by Ghandhi [Gha77]), but is more and more called *thermal runaway*. This phenomenon is always destructive, in contradiction with second breakdown and snapback, which can be sustained by a device during a short period of time.

The different forms of second breakdown and snapback will be treated for each device in the corresponding chapter.

2.6.5 Kirk effect and adaptive RESURF

The *Kirk effect* is normally described when treating high-injection conditions in a bipolar transistor (e.g., [Sze81, p.145]). Under these high current conditions, the electric field peak is relocated from the the base-collector junction towards the collector contact region, which greatly affects the transistor's performance. An analogue relocation of the electric field peak can occur in other power devices, especially in RESURF type structures. It is actually one of the drawbacks of the RESURF effect, and is the cause for introducing an extra layer in the process flow that needs to increase the doping level towards the place where the electric field is shifting. This technique is called *adaptive RESURF*. Once again, these subjects will be dealt with in detail in the chapter on the power MOS.

2.6.6 Safe operating area

The *safe operating area* of a device is the working region beyond which the transistor should not be used. This is a very rough definition, and how the limit is determined depends on the criteria used, classified in 3 groups:

Electrical safe operating area and ESD

We have already introduced terms as breakdown (in the off-state) and second breakdown (in the on-state). These are voltages that—at first sight—define the absolute outer limits of the device's operating region, above which a device can no longer be used. However, when very short pulses are used (ns to µs), the devices *can* be used beyond second breakdown and snapback. A method called *T*ransmission *L*ine *P*ulsing can determine how far the device can go under these short pulses. Devices can be specially made to handle such short, heavy pulses to serve special purposes. The most common purpose being ESD (*E*lectro*S*tatic *D*ischarge) protection. ESD protection devices and ESD measurements on power devices are not treated in this work.

Thermal safe operating area and energy capability

When the pulses are longer (µs to ms), temperature related phenomena come into play and reduce the safe operating area further when *thermal breakdown* or thermal runaway is reached. These type of pulses are used in *energy capability* measurements in order to know the amount of energy a device can take, during, e.g., an inductive turn-off. It is outside the scope of this book to perform energy capability measurements on power devices.

Hot carrier safe operating area and degradation

When a device is not used for very demanding purposes (ESD protection, inductive turn-off) and is used as a switch—that is, it is mostly on or off, and switches very fast between those two states—it still can have a safe operating area that is smaller than the thermal safe operating area. Especially in MOS-gated devices during the on-state, when current is flowing, some of the carriers have enough energy to surmount the Si/SiO_2 barrier (hence *hot carriers*) and cause damage to the oxide. This damage induces shifts in the important electrical parameters of the device. This is called device *degradation* and is a very complex phenomenon. The limits of the safe operating area are normally defined by the most degrading electrical parameter (measured by the fab) for a specific set of biasing conditions. It is based on a simple extrapolation of the most degrading parameter after some initial stress measurements. The stronger the criterion (e.g., 5 % degradation after 25 years instead of 10 %), the smaller is the safe operating area.

Things get even more complicated when the switching of the device is taken into account. Indeed, a device can switch *very fast* between the on- and off-state, but not *infinitely* fast. This means that when a device is meant to have a long life (e.g., 25 years) and it is switched on and off a lot during its life, it has spent a long net time between on- and off-state. Not to mention the type of load the device has to control ! Depending on whether it is a resistive, a capacitive or an inductive load, the device degrades faster or not.

ESD, as well as energy capability can not be simulated with the simple drift-diffusion model and they actually take TCAD to its utter possibilities. Furthermore, the very complex issue of degradation is not always possible to simulate. Although we will mention some hot carrier measurements on power devices, it is not the purpose of this work to go into great detail about the various degradation phenomena.

2.6.7 High level injection and conductivity modulation

Conductivity modulation is one of the reasons why a p-i-n rectifier can conduct current with a very small forward voltage drop, while blocking high voltage in the off-state. Most books on power devices therefore introduce the concept conductivity modulation while treating the p-i-n rectifier under *high level injection* conditions. Namely, during on-state at high current density values, the amount of minority carriers injected in the i-base exceeds the background doping concentration. Charge neutrality requires that the amount of minority and majority carriers are equal. These concentrations can become several orders of magnitude larger than the background doping concentration, with a considerable decrease of the resistance of the i-base as a result. One can prove that this low on-state voltage drop is maintained even at very high current densities because the voltage drop across the i-base region is independent of the current through it (e.g., [Gha77, p. 110]). This conductivity modulation can take place in any device where a junction is forward biased during the on-state (bipolars, IGBTs, thyristors. . .).

2.6.8 Isolation

Low side and high side (or floating) devices

One can design a device with an internal breakdown voltage of e.g., 80 V, but still have a device that is not able to work at this voltage. The reason is—contrary to a discrete device—that the integrated device needs to

be isolated vertically towards the substrate and laterally towards other devices on the chip. Furthermore, some devices have stronger needs for isolation: the high-side or so-called floating devices.

Latch-up, cross-talk and substrate currents

Some devices need to be able to float above (and in a very few occasions even below) the substrate potential during on- and off-state. To achieve this in a junction isolated technology, they are put in a well or tub surrounded by buried layers and plugs (or sinkers) to provide electrical isolation from the substrate and the other devices on the chip. In a SOI technology, the devices are dielectrically isolated by the use of the buried oxide layer and oxide trenches. These major differences have a severe impact on the design of power devices in a SOI technology. Using junction isolation, as in this book, there is always the potential danger of latch-up between power devices or cross-talk of the power device with nearby CMOS circuitry caused by substrate currents.

References

[AV79] J.A. Appels and H.M.J. Vaes. High Voltage Thin Layer Devices (RESURF devices). In *Electron Devices Meeting*, pages 238–241, 1979.

[Bal92] B.J. Baliga. *Modern Power Devices*. Krieger, 1992.

[Bal96] B.J. Baliga. *Power Semiconductor Devices*. PWS, 1996.

[BGG99] V. Benda, J. Gowar, and D.A. Grant. *Power Semiconductor Devices*. John Wiley & Sons, 1999.

[CS96] C.Y. Chang and S.M. Sze, editors. *ULSI Technology*. McGraw-Hill, 1996.

[Gha77] S.K. Ghandhi. *Semiconductor Power Devices*. John Wiley & Sons, 1977.

[HZ01] A.Q. Huang and B. Zhang. The Future of Bipolar Power Transistors. *IEEE Transactions on Electron Devices*, 48(11):2535–2543, November 2001.

[ISE00] ISE Integrated Systems Engineering. *Manual, Release 7.0*, 2000.

[Kan98] K. Kano. *Semiconductor Devices*. Prentice Hall, 1998.

[Lud00] A.W. Ludikhuize. A Review of RESURF Technology. In *Symposium on Power Semiconductor Devices & ICs*, pages 11–18, 2000.

[Lun92] M. Lundstrom. *Volume X Fundamentals of Carrier Transport*. Modular Series on Solid State Devices. Addison-Wesley, 1992.

[Neu89a] G.W. Neudeck. *Volume II The PN Junction Diode*. Modular Series on Solid State Devices. Addison-Wesley, 1989.

[Neu89b] G.W. Neudeck. *Volume III The Bipolar Junction Transistor*. Modular Series on Solid State Devices. Addison-Wesley, 1989.

[Pie90] R.F. Pierret. *Volume IV Field Effect Devices*. Modular Series on Solid State Devices. Addison-Wesley, 1990.

[Str62] R. Stratton. Diffusion of Hot and Cold Electrons in Semiconductor Barriers. *Physical Review*, 126(6):2002–2014, June 1962.

[Sze81] S.M. Sze. *Physics of Semiconductor Devices*. John Wiley & Sons, 1981.

[Sze88] S.M. Sze, editor. *VLSI Technology*. McGraw-Hill, 1988.

[Wac90] G.K. Wachutka. Rigorous Thermodynamic Treatment of Heat Generation and Conduction in Semonductor Device Modeling. *IEEE Transactions on Computer-Aided Design*, 9(11):1141–1149, November 1990.

[Wan66] S. Wang, editor. *Solid-State Electronics*. McGraw-Hill, 1966.

3 TCAD Simulation and Calibration

3.1 Introduction

*T*echnology *C*omputer *A*ided *D*esign (TCAD) uses physical models to simulate an entire process flow, process step by process step, and to simulate the working of the device. The so-called process and device simulators discretize the properties of a real structure onto a non-uniform grid or 'mesh' (in 1, 2 or even 3 space dimensions) to solve the differential equations describing the various physical phenomena. Thus, when simulating device structures using TCAD, one needs a translation of the real process flow into the software code used by the process simulator. Ideally, this translation would be trivial. The process simulator would then interpret each process step and use the relevant models: simple models when applicable and complex models when necessary. Ideally, the grid would be generated by the process simulator: refining where necessary and relaxing where possible. Once the structure has gone through the process simulator, it passes to the device simulator that—ideally—would generate its own grid (or several grids depending on the desired simulation) and that on its turn would use the correct models.

Unfortunately (or fortunately, depending on your politics), this ideal simulator does not exist. Although the simulators do get more intelligent over the years, the user still has to intervene at each step described above and this for several reasons. First of all, there is the physics: some process related physics and some device physics are still research topics. Secondly, some model parameters are inherently fab related. The third reason is CPU time related: the simulator can not know what degree of accuracy is desired and therefore the user has to define what models have to be used and how fine the mesh must be. These problems call for calibration. The process calibration is treated in the next section, while the device calibration is treated in the last one. Both sections deal with the problems and difficulties associated with the models and with the meshing.

3.2 Process Simulation and Calibration

One needs to calibrate the process flow (doping profiles, oxide thicknesses, bird's beaks...). It is obvious that the more measured material is available, the better the simulations. Ideally one should have a SEM picture and/or SIMS or SRP measurements after each relevant process step. But this is not always possible: there is the problem of cost and sometimes the measurement techniques are just not available yet (e.g., 2D SIMS). So, on the one hand one is bounded by the limited measurement data, and, on the other hand by the limitations inherent to TCAD (the meshing and the models). Depending on the structure and the technology one wants to simulate, the available TCAD tools can or can not be sufficient. For the work presented in this book, process simulation is adequate and even more physical models are at one's disposal than actually necessary. An example is the simulation of the ldd implantations and the subsequent anneal. Present TCAD tools make it possible to simulate the damage and amorphization associated with these implantations and to take into account these damage distributions in order to simulate transient enhanced diffusion (TED). This is necessary for deep submicron CMOS simulation and demands a lot of calibration work. But is it relevant for large structures like DMOS devices ? One understands that TCAD simulation translates the process flow (the so-called 'input deck') depending on the device structure under study. But before we embark on the translation of the process flow, the important matter of meshing needs to be dealt with.

3.2.1 The mesh

The great majority of problems encountered during TCAD work is mesh related. The user wants a mesh that is as coarse as possible to reduce the CPU time. This is a possible dangerous situation, as one easily introduces faults during simulation when the mesh is not adequate enough.

The purpose of this section is to introduce the reader into the TCAD work, with its typical meshing and calibration problems, using the ISE software. We will not go into too much details about the ISE software language itself, but focus on *how* we procede. Some extracts of the input deck will be given in `typed` text style. The reader who is interested in doing simulations can use this information together with the ISE manuals [ISE00] to get started. The correct use of the TCAD tools is mainly a process of a lot of trial and error and of gaining practical experience. To illustrate this process of meshing and calibration, an existing stan-

dard 0.35 µm nMOS device of AMI Semiconductor will be simulated. Doping concentrations will not be indicated, as this is confidential information. However, this does not compromise the purpose of the following discussion.

In a previous work by our group, a great deal of effort goes into mesh related problems [Ver01]. The available TCAD tools in this work required the definition of the mesh *before* process simulation. Therefore the user had to define beforehand where the mesh should be fine. This is not a very effective way of working since the fine mesh is used during the entire process simulation, even when it is not necessary. This time consuming method can be avoided with a more appropriate meshing strategy, called "adaptive" meshing. It refines and relaxes the grid depending on certain criteria set by the user.

This adaptive meshing tool determines our meshing strategy in ISE. The ISE inputdeck normally begins with a header containing parameter definitions, refinement parameters and possibly specifications on what models and model parameters have to be used. For a definition of all refinement parameters, we refer to the ISE manuals [ISE00]. Important to note here is that the parameter `MaxTrl` defines the maximum number of refinement levels and holds for the rest of the simulation (all other refinement parameters can be redefined later on in the input deck). One refinement level consists of dividing a triangle into 4 congruent subtriangles by splitting the edges in the midpoint. For the simulation discussed here the header is:

```
#header
Title("c035_00_dio_cmd", maxv=50000)

! begin geometrical data
set end=1.0000
! end geometrical data

! Refinement Control Parameters

Replace(Control(Sidiff=0, Newdiff=1, Ngraphic=1000))
Replace(Control(MaxTrl=9, RefineBoundary=-2,
RefineGradient=-2, RefineJunction=-2, RefineMaximum=-2,
RefineACInterface=-2, RefineBeforeFront=-2, RefineGreen=0,
RefinePoints=0))
Replace(Control(si(MaxTrl=11, RefineBoundary=-2,
RefineGradient=-2, RefineJunction=-2, RefineMaximum=-2,
```

```
RefineACInterface=-2, RefineBeforeFront=-2)))
#endheader
```

All refinement parameters are set to -2, which means that 1 subdivision of the starting grid is allowed (as -1 means no refinement at all). Note that the `MaxTrl` parameter has been set to -11, which means that the *possibility* of refining 10 times (in silicon) has been given !

The next step to take is to define a rectangle (i.e., the 2D simulation domain) and the number of initial grid triangles, which tessellate this domain:

```
!    PROCESS FLOW - C035M-D

!1  LOT START
grid(X(0.0, $end), Y (-20.0,0), nx=1)
```

Since `nx` has been set to one, and since the length of the simulation domain is one (see the parameter `end` in the header), the length of the edges of the triangles at the beginning of this simulation is 1 µm. With all refinement parameters set to -2 in the header, this means that the smallest edge possible is 0.5 µm. In other words, we start with a very coarse grid, and a very coarse grid will be used throughout the simulation as long as we do not redefine the refinement parameters. This is our meshing strategy using the adaptive meshing tool: refinement will only be used when and where it is necessary by means of boxes (see further).

After the definition of the grid, the user has to the define the starting material:

```
! 1A Select starting material
substrate(Element=B, Orientation=100, Conc=$sub_conc)
```

with the necessary specifications. Now the simulation of the process flow can begin. The first important process step is the epi growth.

Here, the importance of the mesh is illustrated with this first high temperature step. Therefore, we have defined a box just before the n-epi growth, that tells the simulator to refine the grid when the gradient of the doping profile is larger than a specified value (default value taken):

```
# postheader
Replace(Control(Rec1(Refinegradient=@<-1*ref_grad>@,
Xleft=0, Xright=$end, Ytop=1, Ybot=-1)))
```

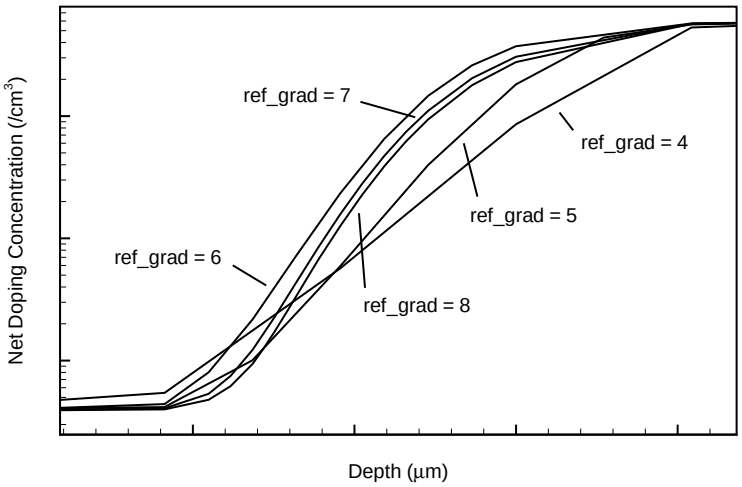


Figure 3.1 Influence of the refinement gradient parameter on a doping profile.

```
# endpostheader
adapt()
Diffusion(Atmo=EPI, Time=$time, Temperature=$temp,
Thickness=$epi_thicknessum, Elem=B, Conc=$epi_conc)
```

Since 1 µm is the initial length of the edges, the smallest possible edge is now easily calculated:

$$\text{Smallest edge length} = 1/2^{(\text{ref\_grad - 1})}. \tag{3.1}$$

This is illustrated in figure 3.1, where the epi to substrate doping profile is plotted in detail for several refinement criteria. Once again, the user has to choose what degree of refinement is wanted for the subsequent process and device simulations. Since we want to simulate a nMOS device in this section, `refgrad = 4` is chosen. If later on (even during device simulation), this would turn out to be inadequate, the simulation would have to be redone with a higher value.

During the subsequent process simulation, the definition of boxes at appropriate times and places is extremely important. We have to learn when, where and what refinement parameters to use and this will be treated in the next section where relevant. This is an illustration of TCAD being a trial and error process and a matter of practical experience. Some examples of the evolution of the grid during an entire simulation (that is, from 1D process simulation over 2D process simulation toward the grid used for device simulation) are shown in figure

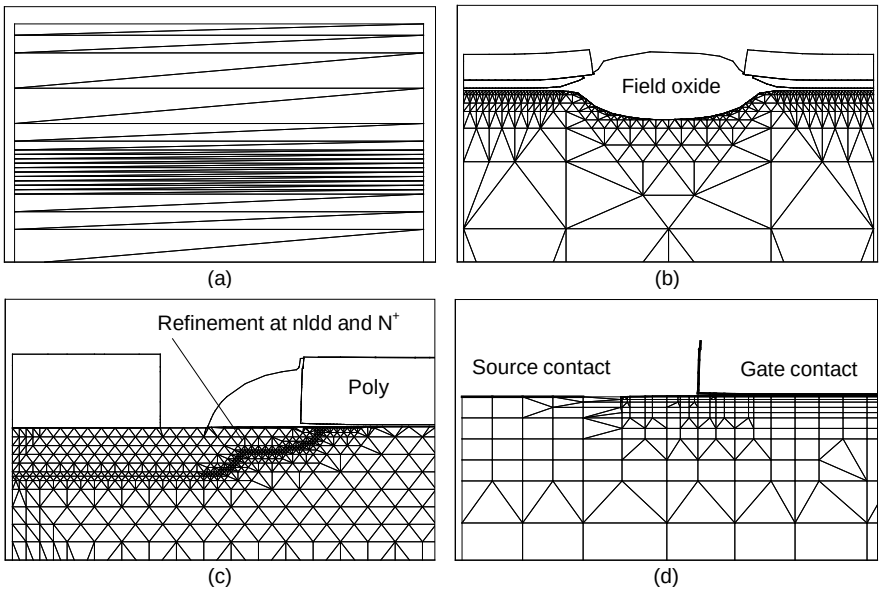


Figure 3.2 Evolution of the mesh during simulation: (a) 1D in process simulation, (b) first process step in 2D, (c) at the end of the process flow, and (d) regridding before device simulation.

3.2 (only the mesh in the silicon is shown): (a) shows the grid at the 1D stage of the process simulation, (b) is just after the first process step that was simulated in 2D, (c) is at the end of the process flow, where one clearly can see the refinements in the grid around the nldd and N^+ region (i.e., where the gradient of the net doping profile is the largest), and (d) gives an example of the regridding that happens just before device simulation (discussed below), where the very fine grid following the net doping profile is no longer necessary, but where a very fine grid in the channel is present (not visible), which was not the case after process simulation.

Now, we will move on to process calibration, as the epi to substrate profile could already have been subject to calibration.

3.2.2 Simulation and calibration

The process simulation remains a 1D problem as long as there is no mask definition. This means that the simulations are carried out in one dimension (refining only in the y direction; that is, the direction perpindicular to the wafer). The process simulator switches automatically

to 2D simulation when necessary, which depends on the intended device structure.

The epi growth

We have stated in the previous section that "the first important process step is the epi growth". But is this true ? What is important ? To answer these questions, we take a look at the real process flow and ask ourselves what would be really important to simulate and especially *when* would that be.

The actual process flow begins with an RCA clean. It is obvious that it is not necessary to simulate this in our virtual process flow. This is the first and most trivial example of how the translation of the actual process flow into the TCAD input deck omits or simplifies process steps.

The next process step is the epi growth. At first sight, this has to be simulated, as it is a high temperature step and thus induces a gradient in the doping profiles. But what if we do not simulate this process step and define the epi as starting material ? Would this be correct ? The answer depends on the desired simulation, and, especially on the subsequent device simulation. If one only wants to simulate on-state DC characteristics of a nMOS, it will probably not be necessary to simulate the extra doping profile well underneath the silicon surface. However, for power devices, it will be of utmost importance to simulate this process step as often buried layers are defined before the epi growth. This epi growth changes the profile of the buried layer and plays a major role in most of the subsequent device simulations (breakdown !). This is a first important example of how the translation of the input deck depends on the intended device and its subsequent device simulation. Note, however, that at all times the simulation domain has to be deep enough if one wants to exclude errors due to boundary effects. One can for example define an arbitrary epi depth, disregarding the actual epi depth and epi to substrate junction.

There is yet another way to translate the epi growth: define the substrate as is, but then deposit the epi ! As simulation of deposition is based on purely geometrical models in ISE, this demands less CPU time than several diffusion steps. Although this is negligible in the case considered here, as the simulation is still a 1D problem at this point. Therefore we choose to simulate it with the diffusion steps, since it is probably more correct. But what is correct ? Don't we have to calibrate this profile ? The answer is yes... but we have a limited choice as to

when in the process flow and where on the wafer a SIMS is taken, mainly because of cost. Since it is believed that other profiles (the wells, the ldds...) are of greater importance, it is chosen not to take a SIMS of the epi to substrate profile.

1D Oxidation

The following process step is a pad oxidation. This step is easy to calibrate, as it still is a 1D simulation and there is a lot of information coming from the fab: ramp up and down times, oxidation time and gas flows, and, of course, the statistical data on the thickness. This step is translated in the following way (process data are left out for confidential reasons):

```
!2  PAD OXIDATION

! 2A RCA clean
! nothing simulated here

! 2B Pad oxidation
Diffusion(Time=($t1, $t2, $t3), Temperature=($te1, $te1,
           $te2,$te2), Flow(N2=$flow1, O2=$flow2))
Diffusion(Time=$t_ox, Temperature=$te2, Flow(O2=$flow_ox))
Diffusion(Time=($t4, $t5), Temperature=($te2,
           $te2, $te1), Flow(N2=$flow3))
```

This literal translation gives an oxide thickness that is 11 % higher than the actual one. In ISE, refining at the boundary (`RefineBoundary`), which results in a finer grid in both the silicon and the oxide, does *not* have an impact on the simulated thickness. This was verified with very coarse and fine grids, both giving the same 11 % discrepancy.

So calibration is needed for this step. We remark here that the `diffusion` command is by far physically the most elaborate command, compared to the simulation of other process steps (deposition, etching, implantation). This is comprehendable, as it serves for all temperature steps (oxidations, anneals, epi growth...).

During such a step, dopant redistribution has to be simulated and therefore the correct model needs to be chosen. In ISE, 5 different models can be chosen. Actually, they are basically the same model with different levels of simplification. The most complex model fully couples

the dopant and point defect diffusions. Additional modelling of clustering and trapping reactions is also possible. This most complex model is needed when one wants to simulate transient diffusion effects, causing e.g., the reverse short channel effect. The default model can not simulate such effects as it does not solve point defect equations and as it assumes no point defect assisted diffusion. For a first calibration cycle, this default model is chosen for every temperature step. If it turns out to be inadequate, complex models can be used for the important temperature steps. However, for the simulation of a simple pad oxidation, the use of the default model is sufficient.

And this is only half of the story for the simulation of a thermal process that changes the layer structure (oxidation, but also silicidation). In addition to the diffusion, the simulator has to take into account the chemical reactions and the segregation at interfaces, the diffusion and convection of dissolved particles, the screening property of some interfaces or layers for particle fluxes, mechanical deformation of the entire layer structure... Several oxidation models are possible, with different degrees of complexity and coupling of the physical models.

We choose the default model for almost all oxidation steps and will verify when doing device simulations whether we should redo some oxidation steps with a more complex oxidation model or not. As in the case of the epi, we could even choose to deposit the oxide as this does not induce great differences here. However, since CPU time is still not an issue here, we stick to the simulation of the diffusion steps.

There still is the issue of the calibration, the correct way would be to look for a model parameter that can be tuned and to use this to simulate the correct thickness. But since we are dealing with a simple pad oxidation and since we don't want to loose too much time on this relatively unimportant process step, we choose a more arbitrary way of calibrating this oxidation. It is possible to specify the intended oxide thickness instead of the time in the diffusion command. This has been done for this process step (in the diffusion command where the actual oxidation takes place).

Deposition

Simulation of deposition is not physically based and is modelled in a geometrical way with the aid of local depositon rates that mimic the geometry observed in reality.

For the simulation of a 1D deposition it is trivial that an isotropic

deposition suffices (default). The next 2 process steps are such depositions:

```
! 3A Amorphous silicon deposition
Deposit(Material=Po, Thickness=$poly_buffer)

! 4B Nitride deposition
Deposit(Material=Ni, Thickness=$nitr_thick)
```

Lithography process

The same as for the simulation of deposition holds here: the lithography process is not simulated physically, but purely geometrical, using the command `Mask`.

The simulator changes from one dimension to two dimensions with the introduction of a mask that is not covering the entire simulation domain length.

Since no field oxide is grown during the simulation of a nMOS device, our simulation domain is entirely covered with the photoresist:

```
5  ACTIVE AREA MASK
Mask(Material=Resist, Thickness=2.0)
```

If one wants to open a window where the field oxide should be grown, the command would look like:

```
Mask(X(0,$aa_begin+$aa_bias,$aa_begin+$t-$aa_bias,$end))
```

This brings us to the issue of the mask biases. It is not the case for the simple example of a nMOS device, but it is vital to take into account the mask bias values if one wants to compare measured data with simulated ones.

Take e.g., the simple case of a field oxide length t. If one chooses not to simulate with these mask biases, a field oxide length of 1 µm in the simulation has to be compared with a *drawn* field oxide length of $1\,\mu\text{m} + 2 * \text{mask bias}$. Because, in reality, one can only *draw* this gap in the active area mask, after which the photoresist is *generated* using the mask biases. Therefore we choose to take into account these biases as in the previous command, in order to ease the process of comparison.

Etching

Etching is the last example of a process step that is not physically, but geometrically (by the means of local etch rates) simulated in ISE.

In our example, the following process steps (3 consecutive etches: the nitride, the poly and a tiny part of the pad oxide) need not to be simulated as the simulation domain is completely covered with a photoresist. But suppose we have left a gap in the photoresist, then the commands look like:

```
!6  ACTIVE AREA NITRIDE ETCH

! 6A Nitride etch
Etch(Material=Ni, Remove=$R1, Rate(Aniso=280))

! 6B Amorphous silicon etch
Etch(Material=Po, Remove=$R2, Rate(Aniso=275))
Etch(Material=Ox, Remove=$R3, Rate(Aniso=100))
```

Note that all 3 etches are supposed to be purely directional etching in the direction of the incident "etching beam".

Before field oxidation, whether or not there is a window in the aa mask, the photoresist has to be removed. This is done with an isotropic etch:

```
! 6C Resist strip
Etch(Material=Resist)
```

For the field oxidation simulation, we could just simulate the thermal budgets as long as the entire simulation domain is covered with nitride (as in the case of a nMOS).

We will however simulate the field oxidation, because it is important for the power devices presented in this work.

2D Oxidation

Not only the thickness of the field oxide needs to be calibrated, but it is of utmost importance in power devices with a field oxide in the device structure to simulate the so-called bird's beaks correctly. In order to be able to verify the shape of the bird's beaks in reality, SEM pictures have been taken.

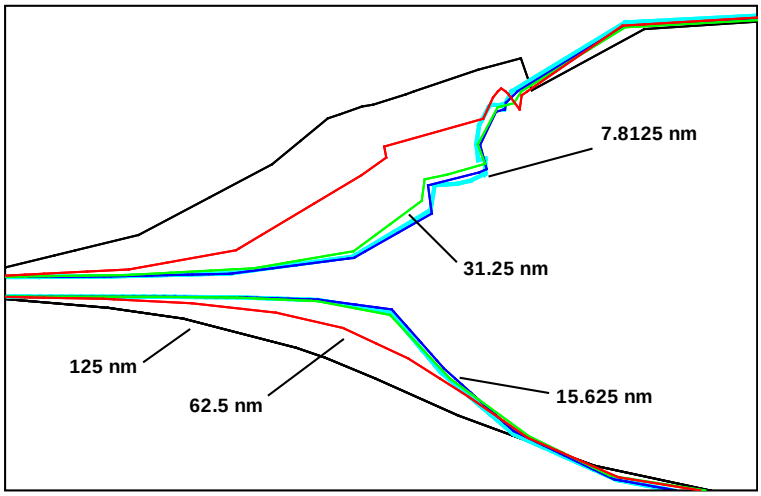


Figure 3.3 Dependence of the bird's beak's shape on the edge length of the triangles.

First of all a refinement box is defined just before the field oxidation with the refinement level of the `RefineBoundary` as a parameter. Figure 3.3 shows that a refinement level of 6 (3.1) is sufficient. The actual length of the simulation domain for these simulations was 2 µm, but we have redefined `nx` to a value of 2. This is also what has to be done when layout variations are carried out that change the length of the simulation domain: one has to make sure that the basic edge length does not vary too much between the different layout variations. Otherwise the simulations are prone to variations that are mesh related (see below).

Now we are ready to calibrate this bird's beak's shape. First of all, the field oxide thickness is calibrated in 1D in the same way as it was done for the pad oxide. Then, a 2D simulation is carried out where we compare a SEM picture with the result of the default simulation. Note that the default model in the ISE version used here is a viscoelastic model with the stress dependent model on. Figure 3.4 shows the importance of this stress dependent model when simulating the bird's beaks. It also shows that the default model with the default model parameters is very accurate and needs little or no calibration at all.

For the simulation of a nMOS, the default models are used (although this is actually an overkill since the stress dependent model is not necessary during a 1D simulation). No attention has been given to the doping profiles under the field oxidation neither. In power devices, e.g., the pdrift in a pDEMOS, this can become important.

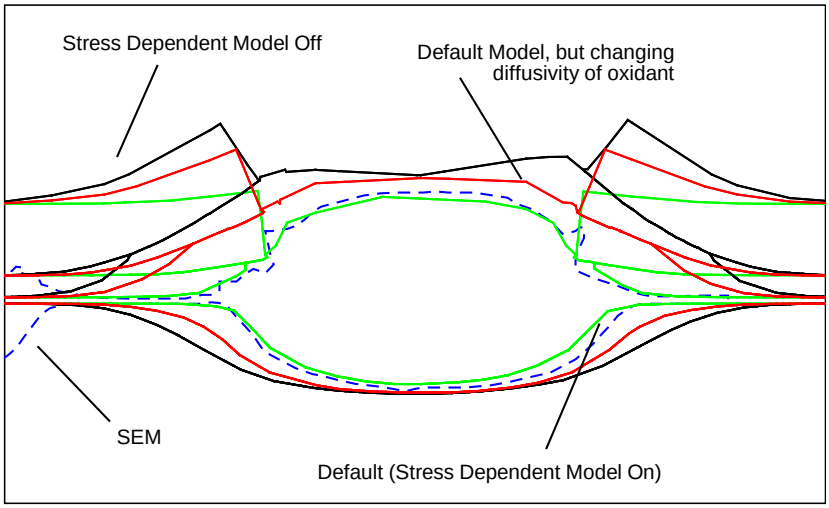


Figure 3.4 Comparison of different oxidation models with SEM picture.

After the field oxidation, the nitride and the polysilicon are removed, a pre-implant oxidation is grown and the nwell mask is defined. During nitride and polysilicon removal, some oxide is etched as well. This could have an impact on the shape of the bird's beak. Since we don't have a SEM picture after each etching step, we just performed a 1D calibration on the thickness of the active area oxide and field oxide after each etching step. The pre-implant oxide is calibrated in the same way. After these process steps, the first implantation in the standard CMOS process flow is executed.

Implantation, anneal and oxidation

It is obvious that the simulation of an implantation is very important, as the doping profiles at the end of the process flow (for a definition, see the following subsection) depend in the first place on the initial implantation profile. Ideally, one should have several SIMSs: one right after the implantation, one after the first temperature step following the implantation and one after each important temperature step up to the end of the process. Unfortunately, we have to stress once again the problem of cost and we have to choose where and when a SIMS is taken. Indeed, *where* as well, since the profile depends also on where it is taken (e.g., in active area or not). Because of the great importance of the wells (they define the channel regions), a SIMS profile is taken after the gate oxidation and at the end of the process, both in active area.

Simulation of the distribution of the implanted ions and of the implantation damage is done by either using analytical distribution functions or Monte Carlo computations. The moments used for the analytical distribution functions are looked up in tables, which are based on experimental data. If the subsequent anneal or oxidation step does not take into account point defect assisted diffusion, then the damage profile should not be simulated during the implantation. If transient diffusion effects are likely to occur, then the damage profile should be simulated, for which several models are available. This is an illustration that the simulation and calibration of process steps can not be treated independently of each other, hence the title of this subsection.

Depending on whether a nMOS or a pMOS is envisaged, a pwell or a nwell has to be simulated and calibrated. Since we are focusing on a nMOS, the pwell is treated here. First of all, a refinement box has to be defined before the pwell is implanted. In order to have an idea about the level of refinement needed, one reference simulation is carried out where `RefineAll = -8` is specified. This means that all triangles are subdivided 7 times before the pwell is implanted. Then, simulations are carried out with different refinement levels for several refinement parameters. It turns out that the following box gives satisfying results without taking too much CPU time:

```
# postheader
Replace(Control(Rec3(RefineGradient=-5, RefineMax=-8,
Xleft=0, Xright=$end, Ytop=$surface, Ybot=$surface-1.2)))
# endpostheader
adapt()
```

This box is sufficient for both the simulation of the implantation and the subsequent anneal. The pwell is annealed with the gate oxidation, which has been calibrated without the presence of the pwell. For the sake of completeness, the refinement box needed for nwell implantation and anneal is

```
# postheader
Replace(Control(Rec3(RefineGradient=-4, RefineJunction=-8,
Xleft=0, Xright=$end, Ytop=$surface, Ybot=$surface-1.2)))
# endpostheader
adapt()
```

Note that here a refinement was carried out on the junction, which is justified by the fact that the background dopant is boron.

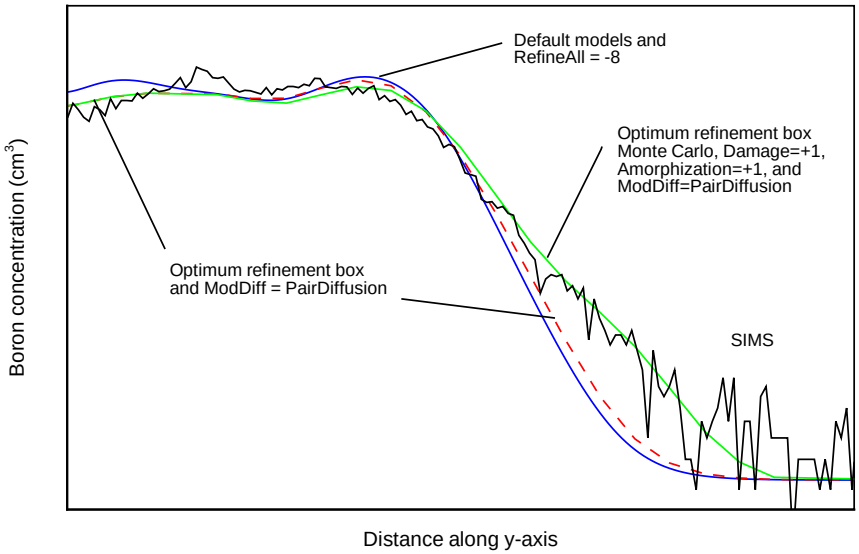


Figure 3.5 Comparison of SIMS profile of the pwell after gate oxidation in active area with several simulation results.

For the calibration of the pwell profile, a SIMS profile is taken after the gate oxidation in active area (figure 3.5). Several models have been applied in a trial and error way. For a correct simulation of the tail of the profile, it seems that a Monte Carlo simulation of the implantation profile is necessary. For a correct simulation of the surface concentration, calculation of the damage during implantation and the use of the most complex defect coupled diffusion model is needed. Once again, one could think of several ways to obtain the same result as plotted in figure 3.5. First of all, we don't know whether the Monte Carlo simulation is closer to reality or not, since we don't dispose of a SIMS taken after the implantation. We could use the default implantation tables for the implantation, alter the damage profiles and try to get to the same result. Or, more arbitrarily, neglect all damage and alter segregation and diffusion model parameters of boron. In this way, using the most simple models, the same result might be obtained in a much shorter simulation time. However, since the sheer application of more complex models give good results and since we don't have to worry too much about simulation times (still 1D !); this option has been chosen.

As to the calibration of the nwell profile, some odd features were noticed in the SIMS profile. It turned out that energy contamination in the implanter was the cause (about 10 % of the nwell dose was implanted at

1/3 or 2/3 of the intented energy). This illustrates that a close cooperation with the fab and a full access to all process flow details is necessary for the calibration of the TCAD input deck.

By now, the simulation and calibration of all possible process steps (except for the salicidation) have been illustrated. Since we are in the first place interested in a correct simulation of the basic DC electrical parameters (V_t, $\beta \ldots$), the rest of the process flow is simulated in a rather arbitrary, almost generic way: no calibration is carried out whatsoever and some process steps are literally left out. The salicidation, for example, which will only have a negligible effect on DC characteristics is not simulated. The damage introduced by nldd and nplus implants, and their influence on the pwell profile, is not simulated. This will only be important if one wants to simulate short channel or reverse short channel effects. But since we will use the input deck to simulate power devices, it will not be necessary to simulate such effects.

3.2.3 The end of the process flow

For the work presented in this book, the end of the process flow is defined as the last important thermal step. From this point on, doping profiles and the device itself will no longer change. For most simulations, it will suffice to define the electrodes in a rather arbitrary way (by deposition of aluminium). These aluminium regions are replaced by equipotential lines where they touch another material in the subsequent device simulations anyway. The poly too will be treated in this way, since the poly is identified as a metal by us. One can however simulate for example poly depletion effects by defining an aluminium electrode on top of the poly. The device simulator will then recognize the poly as a semiconductor region with its own physical properties. The following text gives an example of the end of an input deck. An example of a definition of a thermode is given in comment (sometimes needed for specific device simulations of power devices):

```
!43 ILD THERMIC BUDGET
comment('dio_02: ILD Thermic Budget')

Diffusion(Time=$time, Temperature=$temp, Flow(N2=15))

! END OF FRONT-END SIMULATION
```

```
! etching possible rests of oxide on poly
Mask (X(0, $ref01+0.5, $ref02-0.5, $ref04))
Etch(Material=Ox, Remove=100nm)
Etch(re)

! Metal Contact Definition
Mask(Material=Al, Thick=0.01, x(0, $ref01, $ref02, $end))

! Thermal resistance oxide
! Deposit(material=Ox,thickness=3um)

! Thermode
! Deposit(material=Al, thickness=0.5um)

#set Y_SI 1.0
Print(x=$ref01+0.250, layers)
Measure(Template=msr1.tmpl, LabelAndName("y_si","Y_SI"))


Save(File=n@node@,type=Mdraw,synonyms(Po=metal,Al=metal),
contacts(
        contact1(name='source', 0, 8.005)
        contact2(name='gate', $end/2, 8.005)
        contact3(name='drain', $end, 8.005)
        contact4(name='substrate', location=bottom))
        ! contact5(name='thermo_top', 1, 20)
)

end
```

The extraction of the parameter `Y_SI` gives the exact value of the thickness of the silicon located under the channel (due to oxidations this value is less than the initial thickness of the epi and substrate together). This value will be used later on while remeshing between process and device simulations (see below).

3.2.4 Performing layout variations

TCAD is often used to investigate the influence of layout parameters on the device's performance. For a correct simulation, one must realize that

the device length (or 'pitch') can alter by performing layout variations. So, if the pitch gets larger, one must ensure that the used grid does not get to coarse and alter the grid specifications accordingly to pitch variations.

One way to do this is by changing the initial number of triangles of the grid (`nx` in the command `grid`). An important disadvantage of such an approach is that for large devices, a large number of gridpoints can already be present at the start of the process simulation.

Another way to solve the problem of the varying pitch is by changing the refinement level in each box used before an important process step. The big advantages are that even for large structures one starts with an equal number of mesh points, reaching the finest mesh only there where needed. Thus, the same header as given above is used, together with `nx = 2`. But now, a `refinement_level` parameter is defined, which will determine the level of refinement depending on the pitch of the device (sw, x, y, t and dw being the layout parameters defining the pitch):

```
#postheader
#if (@<sw+x+y+t+dw>@ < 4.6)
set refinement_level=-8
#elif (@<sw+x+y+t+dw>@ < 9.2)
set refinement_level=-9
#elif (@<sw+x+y+t+dw>@ < 18.4)
set refinement_level=-10
#elif (@<sw+x+y+t+dw>@ < 36.8)
set refinement_level=-11
#elif (@<sw+x+y+t+dw>@ < 73.6)
set refinement_level=-12
#endif
#endpostheader
```

Then, this `refinement_level` parameter is used in each box to set the needed level for that specific box, i.e., by *adding* 1, 2, 3... (since the `refinement_level` parameter defines the finest level needed for the entire simulation, determined by trial and error for the largest pitch using this level):

```
#postheader
Replace(Control(Rec3(RefineGradient=$refinement_level+4,
Xleft=0, Xright=$ref01, Ytop=8, Ybot=6.5)))
# endpostheader
adapt()
```

3.2.5 Simulation of power devices

One important issue, concerning the mesh, which will prove to be handy for the simulation of power devices, is further discussed here. Since these power devices are integrated in a CMOS process, one does need to simulate the typical power modules (buried layers, sinkers...) together with the entire CMOS process flow. These so-called smart power technologies therefore tend to have a large process flow, and, what is important for TCAD, these power modules are mostly situated *before* the CMOS process flow. This means that, if one of the first masks is used (e.g., a device that is only partly situated on top of a buried layer), the rest of the simulation is carried out in 2D. This in contrast with the nMOS example given above, where 2D simulation only starts with the definition of the poly gate.

In order to save CPU time, it is therefore vital that an efficient meshing is used throughout the simulation of power devices. To that end it is important to know that the refinement boxes are numbered and that one can *redefine* each box at any given time. Take for example the simulation of a n-type buried layer, which typically is formed with antimony (Sb). This heavy atom is concentrated in a very shallow region under the surface right after implantation. This means that one needs a very fine grid in a small region in order to be able to track the sharp profile right after implantation (the same can be said for ldd implantations, where the dopants are lighter but implanted with a very low energy). The subsequent anneal smoothens the profile and diffuse the dopant outside this refinement box. Therefore a second box is needed that is bigger than the first one, but with refinement specifications that might be looser than the ones needed for the implantation. After the anneal, the first box with its stringent refinement might not be needed anymore. Therefore it is vital that one can *redefine* the box that has been used for the implantation; i.e., one can *coarsen* the grid again.

3.3 Device Simulation and Calibration

Device simulation differs from process simulation in that the physical models are less fab dependent and much better known (at least for 'normal' device working). Device simulation usually demands much less calibration effort and device calibration is therefore normally only used as a final 'fine tuning'. Important exception is the work function of the poly, which is fab related and needs to be calibrated. The same holds for

the fixed gate oxide charges, but is less important for DC characteristics.

In order to have an idea about the correctness of the simulations performed this far, we will compare the TCAD simulation results with SPICE simulations. The used SPICE models are provided by the fab and are confidential. Important here is to realize that this way of working is actually putting things on their heads. Indeed, TCAD simulation results are compared with existing SPICE models, which means that the devices are already characterized and understood (at least to the extent that is important for designers). Normally, TCAD is used to anticipate problems, to understand and to create optimized devices before any actual silicon has been processed. TCAD simulation results can even be used as a basis for extraction of SPICE model parameters *before* any device has been measured (this was actually the case for the pDEMOS presented in this book). The reader has thus to be warned that the comparison made here is not the usual way of working, but it will give an idea about how close (or how far) TCAD simulations are from reality.

But before this is done, we do need to think about the mesh once again.

3.3.1 From process to device simulation

The mesh that has been used for process simulation is no longer appropriate for device simulation. The best example is the channel region, where in the process simulation it suffices to keep track of the pwell profile, which results in a spacing of 50 nm between the mesh lines under the gate oxide (at the end of the process simulation). During device simulation, however, one must take into account the thickness of the inversion layer to assure correct simulation of e.g., an $I_d(V_{gs})$ characteristic.

The remeshing of the structure is done through an extra tool available in TCAD software. One defines several regions (rectangular boxes), and specifies for each region the spacing between mesh lines in the x and y directions (and possibly a dopant serving as a refinement criterion). An example of such a remeshing 'input deck' is given below. Note that the parameter `Y_SI`, extracted from process simulation, is used to define, among others, the channel region. This channel refinement box has a thickness of 4 nm (2 nm above and 2 nm below the oxide/silicon interface) and the spacing between the horizontal mesh lines is fixed by the parameter `spacing`. The result of varying this parameter on a $I_d(V_{gs})$ characteristic is shown in figure 3.6.

A spacing of 2 Å is adequate and has been used in all simulations

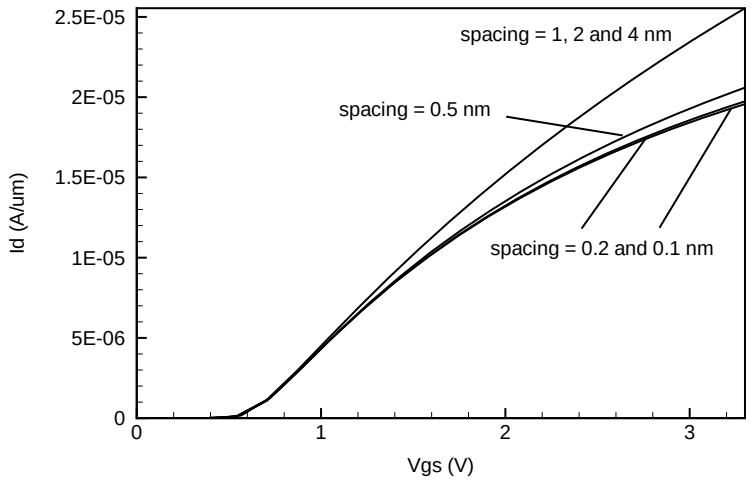


Figure 3.6 $I_d(V_{gs})$ at $V_{ds} = 0.1\,\mathrm{V}$ for a nMOS with drawn channel length of 2 µm for several distances between the mesh lines in the y direction in the channel (i.e., the parameter 'spacing').

presented in this book. Remark that other regions need to be defined as well and that one can define several different remeshing input decks, depending on the subsequent device simulation (a breakdown simulation requires other regions of refinement than an $I_d(V_{gs})$ characteristic). These remeshing input decks are defined through a trial and error process, where device simulation results for coarser grids are compared with a reference simulation with a very fine grid. A simple example of such an optimized remeshing input deck (meaning that the grid is as coarse as possible, but the device simulation results remain the same), used for the simulations presented in the next subsection, is given here:

```
Title "C035/nMOS"

Definitions {

  # Refinement regions
    Refinement "Default Region"
    {
      MaxElementSize = (2.5 2.5)
      MinElementSize = (0.5 0.5)
      RefineFunction = MaxTransDiff(Variable =
      "DopingConcentration", Value = 1)
    }
```

```
    Refinement "active area"
    {
      MaxElementSize = (0.500 0.500)
      MinElementSize = (0.010 0.010)
      RefineFunction = MaxTransDiff(Variable =
      "DopingConcentration", Value = 1)
    }
    Refinement "channel"
    {
      MaxElementSize = (0.100 @spacing@)
      MinElementSize = (0.020 @spacing@)
    }
    Multibox "under channel"
    {
      MaxElementSize = (0.500 0.500)
      MinElementSize = (0.100 0.020)
      Ratio = (0 1.5 )
    }
    Refinement "source"
    {
      MaxElementSize = (0.100 0.050)
      MinElementSize = (0.100 0.050)
    }

    Refinement "drain"
    {
      MaxElementSize = (0.100 0.050)
      MinElementSize = (0.100 0.050)
    }

  # Profiles
    SubMesh "SubMesh_0"
    {
      Geofile = "@grid@"
      Datafile = "@doping@"
    }
}


Placements {
```

```
# Refinement regions
  Refinement "Default Region"
  {
    Reference = "Default Region"
    # Default region
  }
  Refinement "active area"
  {
    Reference = "active area"
    RefineWindow = rectangle [(0 0),(@<2*sw+l>@ 1.2)]
  }
  Refinement "channel"
  {
    Reference = "channel"
    RefineWindow = rectangle [
    (@<sw>@   @<-1*(Y_SI-20000.0)/1000-0.002>@),
    (@<sw+l>@  @<-1*(Y_SI-20000.0)/1000+0.002>@)]
  }
  Multibox "under channel"
  {
    Reference = "under channel"
    RefineWindow = rectangle [
    (@<sw-0.1>@ @<-1*(Y_SI-20000.0)/1000+0.002>@),
    (@<sw+l+0.1>@ @<-1*(Y_SI-20000.0)/1000+0.202>@)]
  }
  Refinement "source"
  {
    Reference = "source"
    RefineWindow = rectangle [
    (0 @<-1*(Y_SI-20000.0)/1000>@),
    (@<sw+0.15>@  @<-1*(Y_SI-20000.0)/1000+0.3>@)]
  }
  Refinement "drain"
  {
    Reference = "drain"
    RefineWindow = rectangle [
    (@<sw+l-0.15>@ @<-1*(Y_SI-20000.0)/1000>@),
    (@<2*sw+l>@ @<-1*(Y_SI-20000.0)/1000+0.3>@)]
  }
```

```
  # Profiles
    SubMesh "SubMesh_0"
    {
      Reference = "SubMesh_0"
      Replace
    }
}
```

3.3.2 Device simulation and calibration

As has been pointed out in the beginning of this section, SPICE simulations are compared with TCAD simulations, although this is a rather unusual way of working. It will, however, demonstrate what degree of precision TCAD can yield. We have chosen to simulate an $I_d(V_{gs})$ characteristic for the nMOS presented this far with a drawn gate length of 2µm. The device input deck needed for this simulation is given below and uses only default physical models: the drift-diffusion model in combination with Boltzmann statistics, a basic mobility model including doping depence, high field saturation and transverse field dependence and a definition of the band gap (and, therefore, the intrinsic carrier density). The work function of the poly has been calibrated to a value of 4.26 eV.

Figure 3.7 compares the slow, typical and fast SPICE models with 2 TCAD simulations. The only difference between both TCAD simulations is that one uses the pair diffusion model for the gate oxidation, and the other one does not. Note how small the differences are between the TCAD and the typical SPICE simulations and both TCAD characteristics. Although the one with the pair diffusion model takes more CPU time; especially in power devices where the simulation of the gate oxidation is already in 2D mode. An important conclusion is that for this device and for this device simulation, the use of the pair diffusion model is redundant. It will, however, sometimes be necessary to use this model, as is shown in simulations where the drawn gate length is varied. For these simulations, the pair diffusion model has not been used and the work function has been recalibrated to a value of 4.28 eV. Otherwise the same device input deck has been used as for the simulations shown in figure 3.7:

```
* des.cmd    dessis input deck for Id-Vg characteristic
```

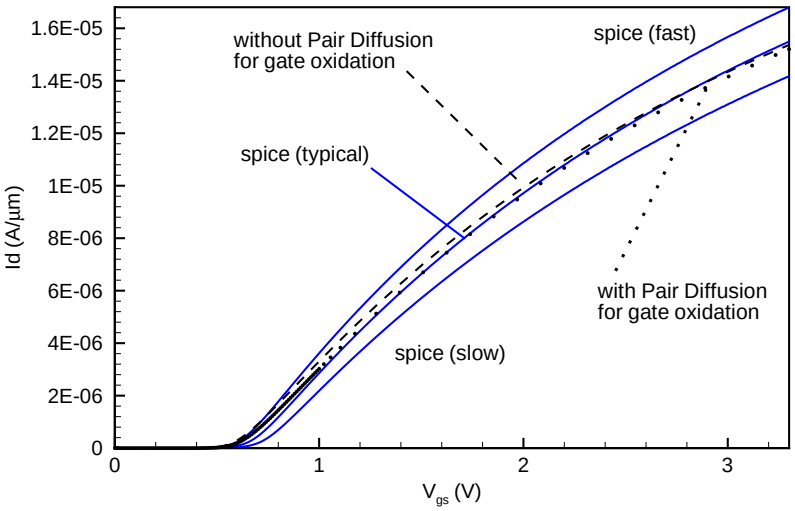


Figure 3.7 $I_d(V_{gs})$ for a nMOS with drawn channel length of 2 µm: SPICE simulations versus TCAD simulations.

```
*            of a nMOS using default models

File {
    Grid      = "@grid@"
    Doping    = "@doping@"
    Current   = "@plot@"
    Plot      = "@dat@"
}

Electrode {
        { name = "drain"     voltage=0.0 }
        { name = "substrate" voltage=0.0 }
        { name = "source"    voltage=0.0 }
        { name = "gate"      voltage=0.0 Workfunction=4.28}
}

Physics {
        Mobility ( DopingDependence
                   HighFieldSaturation
                   Enormal )
        EffectiveIntrinsicDensity ( BandGapNarrowing (
                   OldSlotboom ) )
}
```

```
Physics ( MaterialInterface="Silicon/Oxide" ) {
        Charge ( Conc=2e10 )
}

Plot {
        Doping DonorConcentration AcceptorConcentration
        ElectricField/Vector eDensity hDensity SpaceCharge
        eGradQuasiFermi/Vector hGradQuasiFermi/Vector
}

Math {
        Extrapolate
        Derivatives
        RelErrControl
        NewDiscretization
}

* Id-Vg characteristic

Solve {
        Poisson
        Coupled { poisson electron hole }
        QuasiStationary ( InitialStep=0.1
                Goal { name="drain" voltage=0.1 } )
                { Coupled { poisson electron hole } }
        NewCurrentFile="idvgvd0.1_default"
        QuasiStationary ( MaxStep=0.02
                Goal { name="gate" voltage=3.3 } )
                { Coupled { poisson electron hole } }
}
```

Spice and TCAD simulations for these layout variations (changing drawn gate length) are compared by means of two important device parameters: the threshold voltage (V_t) and the transconductance per µm width (the SPICE parameters have been divided by the width used for the simulations). The results are plotted in figures 3.8 and 3.9. Measurements on a so-called 'golden' wafer (i.e., a wafer containing representative devices) have been added. Once again, TCAD simulations differ only negligibly from the typical SPICE model and the measurements.

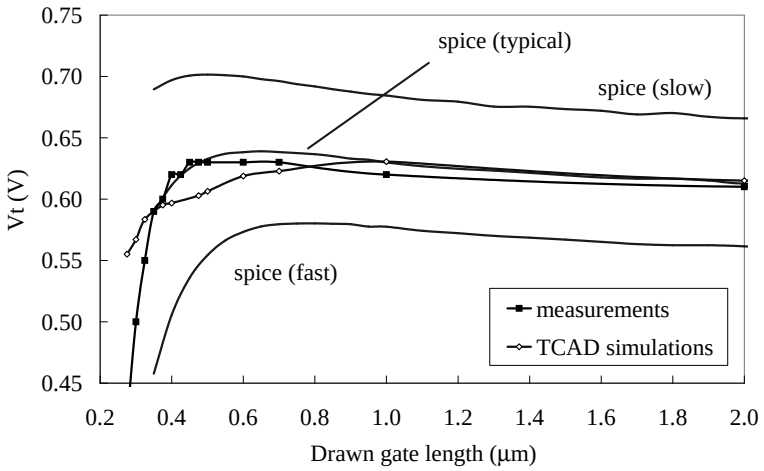


Figure 3.8 Threshold voltage versus drawn channel length for nMOS transistors: comparison between measurements, TCAD and SPICE simulations.

The difference becomes more pronounced for smaller gate lengths, showing that more complex TCAD models are needed to simulate, e.g., the reverse short channel effect (the increase of the V_t with decreasing channel lengths just before the V_t drops drastically).

3.4 Conclusion

Despite the use of default models for both the process and the device simulation, figures 3.8 and 3.9 demonstrate the reliability of TCAD thanks to a correct meshing strategy and a limited amount of calibration work (actually only the oxide thicknesses and the poly work function were calibrated). Of course, one must stay alert for the shortcomings of TCAD and of the calibration work and be aware of the fact that more effort is needed if the complexity of the problem raises (cf. the reverse short channel effect). However, for most of the work presented in this book, the default models are mostly adequate since we are interested in the electrical behaviour of rather large devices.

Furthermore, one has to realize that TCAD is mostly used as a qualitative tool to try out and to analyze new ideas before any silicon is actually processed, to anticipate possible problems and to understand trends when changing process or layout conditions. This is also the case in most of the articles in literature that make use of TCAD. TCAD is seldom used as a pure quantitative, predictive tool (predictive in the

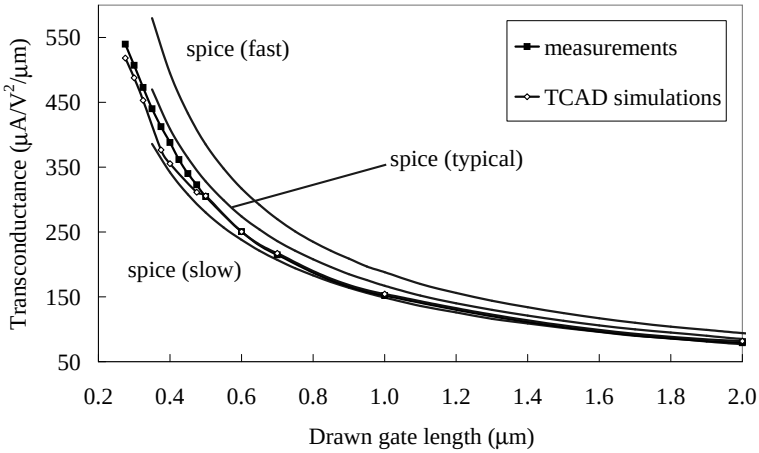


Figure 3.9 Transconductance per μm width versus drawn channel length for nMOS transistors: comparison between measurements, TCAD and SPICE simulations.

sense that such and such process condition together with such and such layout parameter will give such and such electrical parameter). This is understandable, since it takes a lot of effort to calibrate all TCAD results quantitatively and it will only be certain that the correct parameter set has been found (the sheer amount of process and device model parameters makes the calibration work an extremely difficult task) if one disposes of a huge (unrealistic ?) amount of calibration material. And, what's more, there will be no gain in the understanding of the devices, rather only a gain in the percentage of deviation between TCAD simulations and real measurements. A job, actually, that has to be done for SPICE simulators rather than for TCAD simulators (the only and maybe important difference being that process conditions could be included in TCAD predictions). Therefore, like in literature, in this book, TCAD is used as a tool to understand rather than to predict (in the sense described above).

References

[ISE00] ISE Integrated Systems Engineering. *Manual, Release 7.0*, 2000.

[Ver01] M. Vermandel. *Integration of p-type High-Voltage Devices in a Sub-*μm *CMOS Technology using TCAD*. PhD thesis, University of Gent, 2001.

4 Power MOS

4.1 Introduction

One of the first reference works on power devices, written in 1977 by S.K. Ghandhi [Gha77] does not mention the power MOS devices, and only treats the bipolar transistor and thyristor as power switches. On the other hand, one of the latest works on power devices, written in 1996 by B.J. Baliga [Bal96], only treats the bipolar transistor as an introduction to the IGBT. This illustrates the evolution of power devices over a period of 20 years, and proves the importance of the introduction of the MOS gate in power devices. From the 70s on, the MOS gate became available in power devices. The current controlled gate used in power bipolar transistors and thyristors, and demanding complex gate drive circuitry, was now being replaced with a voltage controlled gate with a much simpler input circuitry. Except for some specific applications and for the very high end of power electronics, the voltage controlled gate proved to be much closer to the ideal case than the current controlled gate. Furthermore, the power MOS device has a faster switching speed (due to its unipolarity), a forward voltage drop that increases with increasing temperature (which makes it much easier to parallel these devices), and it is less vulnerable to second breakdown.

Since the MOS gate originates from the digital CMOS technology, it is obvious that the first "power" (a better denomination here is *high voltage*) MOS device is some kind of an extended MOS transistor, with the ability to sustain higher voltages than a normal n or pMOS transistor. In literature, these structures are referred to as drain extended MOS transistors: DEMOS (nDEMOS or pDEMOS) or simply as EMOS (EnMOS or EpMOS). The next step was the introduction of the *double-diffusion process*, where the p-base region and the n^+ source regions are diffused through a common window defined by the edge of the polysilicon gate, which enables to define channel lenths that are much smaller than

the limitation imposed by the lithography. Unfortunately, these devices have the acronym DMOS, which might be a little confusing. The next major improvement of the DMOS' performance came with the introduction of the RESURF technique. The next section treats this matter, as it has become one of the most important techniques for designing power devices (not only DMOS devices).

The third section deals with the ideal silicon limit and the most important plot used when comparing power MOS devices with each other: specific on-resistance versus breakdown voltage. The silicon limit also shows that the on-resistance of the power MOS increases fast with increasing breakdown voltage, which explains the use of the power MOS device in the lower power ratings.

The fourth section sketches all possible forms of the DMOS device: the n versus the p type device, the low-side versus the high-side device, the lateral versus the integrated vertical device, and the RESURF versus the non-RESURF device. It gives an overview of which devices are most feasible in the I3T0 technology.

These devices are treated in the following sections. First of all the low-side or non-floating RESURF n-type devices are discussed: one on a lowly doped substrate and one on a highly doped substrate, representing a BLP. Secondly the high-side or floating devices are studied: one lateral non-RESURF n-type device on a BLN, a lateral RESURF nDMOS on a double buried layer structure, the vertically integrated nDMOS, and finally, the lateral RESURF pDMOS. The conclusions compare the I3T80 devices with the competitor's devices in the form of tables, whereas the $V_{br} - R_{on,sp}$ plot of the silicon limit together with all these DMOS transistors, both for the n-type and the p-type devices, can be found in the concluding chapter of this book.

4.2 Reduced Surface Field Effect (RESURF)

Analytical expressions

The RESURF effect was discovered by coincidence in 1979 [AV79]. When measuring a high voltage diode with an expected breakdown voltage of 400 V, the actual value of more than 1000 V could not be measured until new measuring equipment arrived.

Since then, the RESURF technique has been used in virtually all power devices (e.g., in bipolars [CSN00], in DMOS devices [MBC$^+$99], [HLMP00] and [ZPB$^+$00]), an excellent overview can be found in [Lud00a].

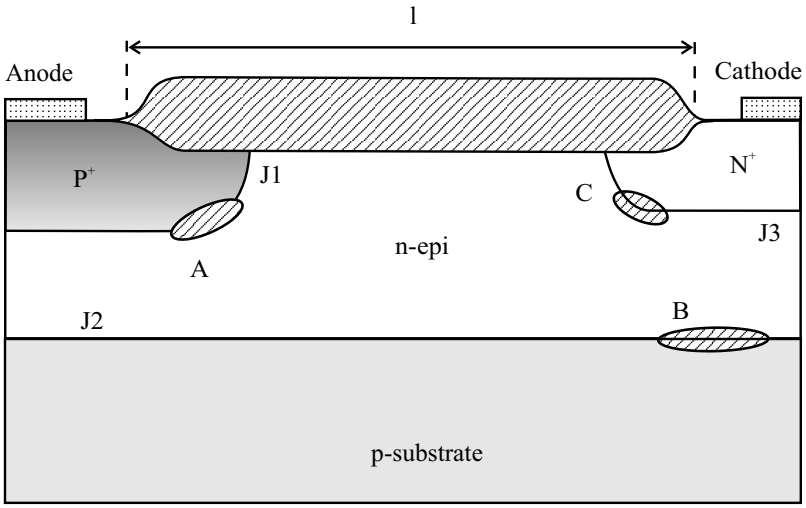


Figure 4.1 A RESURF diode.

It has been extended to other technologies (mainly SOI, e.g., [HB91], [MAB+91] and [vdPLH+00]), towards multiple layers of RESURF layers (e.g., [KPZB02], [HIFT02]), and towards 3D RESURF (the so-called COOLMOS^{TM} or superjunction devices [LDKM99], for an overview see [Udr02]).

Analytical expressions for the breakdown voltage in RESURF structures can be found in [KCA96] (2D analytical model for RESURF structures as in figure 4.1), [MAB+91] and [Chu00] (SOI), and [Fuj97] (superjunction devices).

The basic RESURF diode structure (figure 4.1) consists of two diodes: a vertical (n^+–n-epi–p-substrate) and a lateral (p^+–n-epi–n^+) one. Breakdown can occur at three different places in such a structure (see figure 4.1):

1. Region A: Corner breakdown at the curvature of the p^+–n-epi junction. This can be punch-through (laterally towards n^+) or non punch-through breakdown: $V_{br,lat|npt}$ or $V_{br,lat|pt}$, respectively.

2. Region B: Breakdown at the p-substrate–n-epi junction. This can only be punch-through (vertically towards n^+) breakdown: $V_{br,vert|pt}$. It will be explained in due course why non punch-through can not happen at this place.

3. Region C: Corner breakdown at the curvature of the n^+–n-epi junction. This is corner breakdown due to a depletion layer either

coming from the lateral p^+–n-epi or the vertical p-substrate–n-epi junction, or coming from both junctions at the same time. This breakdown voltage strongly depends on the degree of curvature.

Important to explain is what is meant by the expression *breakdown at a place*. It is defined as the region where the highest impact ionization rate G is observed. Since $G = n\alpha_n v_n + p\alpha_p v_p$ and since $\alpha_n \approx \alpha_p \approx \alpha = A\,|E|^7$ (see equation (C.8)), the impact ionization rate can be written as:

$$G \approx A\,|E|^7\,(nv_n + pv_p) \approx A/q\,|E|^7\,J, \tag{4.1}$$

with J being the current density. So, when only leakage currents are present in a device (as is the case in the RESURF diode prior to breakdown), the place with the highest impact ionization rate is also the place with the highest electrical fields. Therefore, when plotting 2D cross-sections in order to illustrate where breakdown occurs, most of the times the electric field is plotted. However, in some occasions, breakdown can occur (i.e., highest impact ionization rate) at a different place than where the highest electrical fields are located. This can happen, for example, at on-state breakdown in a power device. If, along the currentpath, there is a place with high electric fields (but therefore not necessarily the highest electric fields in the whole structure), the breakdown is likely to occur along this current path as the impact ionization rate not only depends on the local electric field strength, but also on the local current density.

Breakdown in the RESURF diode can happen at three different locations depending on 4 vital parameters: l (the length of the diode, see figure 4.1), t_{epi} (the thickness of the n-epi), N_{epi} (the doping level of the n-epi), and N_{sub} (the doping level of the p-substrate). Secondary parameters are, for example, the curvature of the n^+–n-epi junction. To get an idea of the dominant breakdown mechanism for the different ranges of these 4 parameters, a rough classification based on comparison between l and t_{epi} on the one hand and between N_{epi} and N_{sub} on the other hand is made.

1. $l << t_{epi}$: breakdown at A
   (a) $N_{sub} << N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$
   When there is no punch-through (that is, both l and t_{epi} are large enough to sustain at breakdown both depletion layers coming from the lateral and the vertical diode respectively), the lateral diode breaks down first because it has the highest

lowly doped region (n-epi in comparison with p-substrate in the vertical diode). Breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, it firstly happens in the lateral diode. Breakdown occurs at A: $V_{br,lat|pt}$.

(b) $N_{sub} \approx N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$
When there is no punch-through, the lateral diode breaks down first because it has the highest highly doped region (p^+ in comparison with p-substrate in the vertical diode, where an equal part of the potential drop in the vertical diode is held by the p-substrate as is held by the n-epi). Breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, it firstly happens in the lateral diode. Breakdown occurs at A: $V_{br,lat|pt}$.

(c) $N_{sub} >> N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$
When there is no punch-through, the lateral diode breaks down first because it has the highest curvature in its junction (the p-substrate and the p^+ doping level are assumed to be approximately equal). Breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, it firstly happens in the lateral diode. Breakdown occurs at A: $V_{br,lat|pt}$.

2. $l \approx t_{epi}$: breakdown at A or C

(a) $N_{sub} << N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$
The same reasoning holds as in case 1 (a).

(b) $N_{sub} \approx N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$
The same reasoning holds as in case 1 (b).

(c) $N_{sub} >> N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$, or at C
When there is no punch-through, the same reasoning holds as in case 1 (c): breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, both depletion layer coming from the lateral and the vertical diode reach the n^+–n-epi junction at the same time. Depending on the curvature of the n^+–n-epi junction, breakdown can occur at A: $V_{br,lat|pt}$, or C.

3. $l >> t_{epi}$: breakdown at A, B or C

(a) $N_{sub} << N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$, at B: $V_{br,vert|pt}$, or at C
When there is no punch-through, the same reasoning holds as in case 1 (a): breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, it either occurs first in the lateral diode or in

the vertical diode depending on which of both depletion layers hits the n^+–n-epi junction first. Breakdown at A: $V_{br,lat|pt}$, at B: $V_{br,vert|pt}$. Breakdown occurs at C when both depletion layers reach the n^+–n-epi junction at the same moment and when the curvature of this junction is strong enough.

(b) $N_{sub} \approx N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or $V_{br,lat|pt}$, at B: $V_{br,vert|pt}$, or at C
When there is no punch-through, the same reasoning holds as in case 1 (b): breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, the same reasoning holds as in case 3 (a): breakdown at A: $V_{br,lat|pt}$, at B: $V_{br,vert|pt}$, or at C.

(c) $N_{sub} >> N_{epi}$: breakdown at A: $V_{br,lat|npt}$ or at B: $V_{br,vert|pt}$
When there is no punch-through, the same reasoning holds as in case 1 (c): breakdown occurs at A: $V_{br,lat|npt}$. When punch-through occurs, it firstly happens in the vertical diode: breakdown at B: $V_{br,vert|pt}$.

So, non punch-through at place B can not occur, as mentioned above. Furthermore, when there is no punch-through (in both the vertical and the lateral diode), then breakdown always happens at A. Only in the case where $l \approx t_{epi}$ and $N_{sub} >> N_{epi}$ and the cases with $l >> t_{epi}$, breakdown can happen at another place than A when there is punch-through.

When several breakdown places are possible for the same case, it is likely that electric fields build up at all places. Breakdown thus can happen at two or even at all three places nearly at the same time. It is obvious that in such a case the electrostatic potential is spread over the entire device and that the breakdown voltage can be higher than the lateral diode breakdown. The RESURF case as was first described in [AV79] where the breakdown voltage is *much* higher than the lateral diode breakdown can only occur when $l >> t_{epi}$ and $N_{sub} << N_{epi}$.

Then and only then can the larger part of the voltage be sustained by the substrate and can the breakdown voltage be determined by the lower doping level of the substrate, and it can thus be *much* higher than lateral diode breakdown. The condition being that the depletion layer in the n-epi, coming from the vertical junction, reaches the silicon surface just before lateral diode breakdown occurs. The electric field is then almost at its critical value at A, but stops building up as the vertical diode punches through and the electric field now builds up at B and C. It is clear that the number of charges present in the n-epi

($= D_{opt} = N_{epi} \times t_{epi}$) is critical in order to obtain this ideal RESURF case. When the dose is too low, punch-through occurs too soon and breakdown can even happen at place C due to the curvature of the n^+–n-epi junction. This will be referred to as the *over resurfed* case. When the dose is too high, the depletion layer in the n-epi coming from the vertical diode does not reach the silicon surface soon enough and lateral diode breakdown occurs. This will be referred to as the *under resurfed* case.

It is important to note that in a lot of so-called RESURF devices, the lowly doped substrate is no longer present directly under the epi, but is covered with a highly doped buried layer. This yields a device where the breakdown is not determined by the substrate, but by the lowest doped region in the device, in our example being the n-epi. Therefore, the breakdown voltage remains determined by the doping level of the n-epi and can not be higher than the theoretical limit for punch-through diodes with a lowly doped region equal to the n-epi. Still this is referred to as RESURF, since in such a structure the electric fields can also build up at all three places, and the breakdown voltage, though determined by punch-through (either in the lateral or in the vertical diode), can be the result of an optimal spreading of the electrostatic potential, just like in the ideal RESURF case.

In order to have an idea of the optimal RESURF dose D_{opt}, some simple calculations, based on what is presented in Appendix C, are performed. The breakdown of the lateral diode (p^+–n-epi–n^+) in the RESURF structure is approximated by the expression of the breakdown voltage of an abrupt p^+–n diode. The non punch-through case is considered here (see (C.11)):

$$V_{br,lat|npt} = \left(\frac{\epsilon_s^3}{2q^3\, 1.8 \times 10^{-35}} \right)^{1/4} \left(N_{epi} \right)^{-3/4}. \tag{4.2}$$

The ideal RESURF is defined by the condition that the vertical depletion layer must reach the silicon surface just before lateral diode breakdown occurs ($W_{n-epi,vert} = t_{epi,opt}$). Therefore, in the ideal RESURF case, the vertical diode (n^+–n-epi–p-substrate) in the RESURF structure is approximated by a punch-through abrupt n^+–n–p diode with the extension of the depletion layer in the n-epi at lateral diode breakdown (see equation (C.22)) equal to the epi thickness:

$$W_{n-epi,vert} = t_{epi,opt} = \sqrt{\frac{2\epsilon_s N_{sub} V_{br,lat|npt}}{qN_{epi}(N_{sub} + N_{epi})}}. \tag{4.3}$$

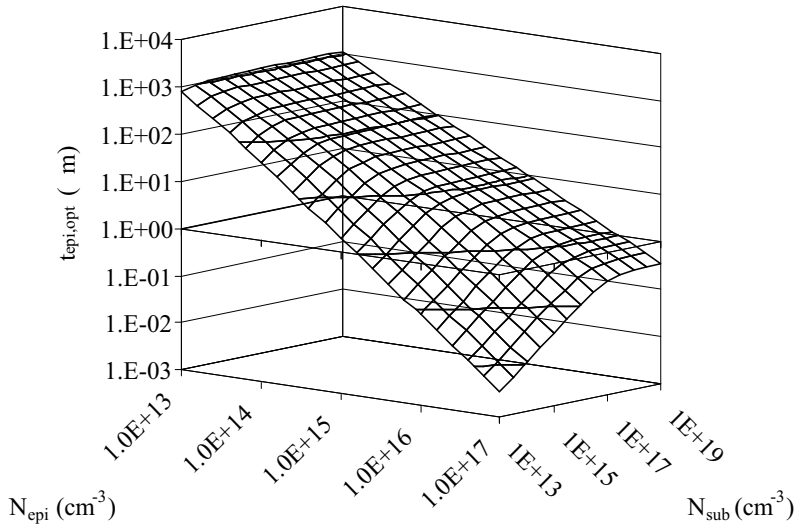


Figure 4.2 Optimal RESURF epi thicknesses for the entire range of realistic n-epi and p-substrate doping levels.

The expression for D_{opt} is obtained by substitution of (4.2) in (4.3):

$$D_{opt} = N_{epi} \times t_{epi,opt} = \left(\frac{\epsilon_s}{q}\right)^{7/8} \left(\frac{8N_{epi}}{1.8 \times 10^{-35}}\right)^{1/8} \sqrt{\frac{N_{sub}}{(N_{sub} + N_{epi})}}. \tag{4.4}$$

When this expression for D_{opt} is divided by N_{epi}, then the theoretically optimal epi thickness $t_{epi,opt}$ is obtained as a function of the n-epi and p-substrate doping level. This has been done for the entire range of realistic doping levels for both the epi and the substrate and is shown in figure 4.2.

The range of optimal epi thicknesses is spread over a large number of orders of magnitude, with the outer limits being unrealistic (especially the lower ranges, the higher ranges are realistic for discrete devices, but not for a CMOS based technology). In order too narrow down the range of optimal epi thicknesses, the value of the optimal epi thickness, together with its corresponding n-epi and p-substrate doping level is used in the expression (C.30) for the punch-through breakdown voltage of a n^+–n–p diode, which determines the breakdown voltage of an ideal RESURF diode. These punch-through breakdown voltages are plotted in figure 4.3.

When all values of $t_{epi,opt}$ corresponding to a breakdown voltage lower than 80 V are disregarded (they are set to 10^{-4}) and when all values of

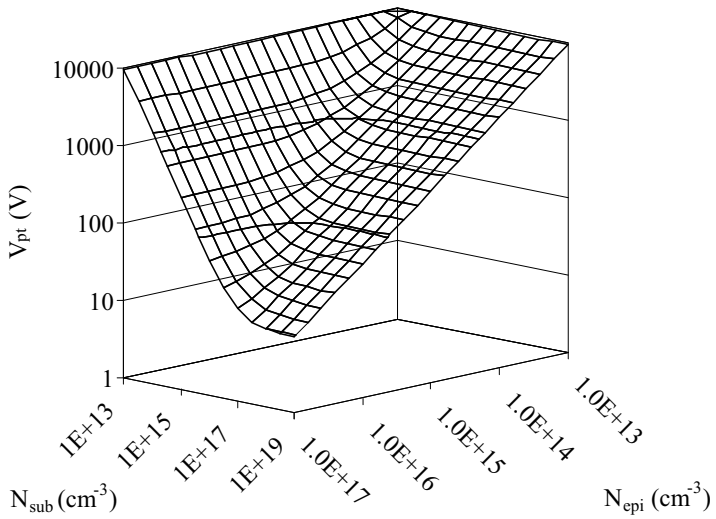


Figure 4.3 Breakdown in an ideal RESURF diode is given by the vertical diode punch-through breakdown voltage and is a function of the n-epi and p-substrate doping levels.

$t_{epi,opt}$ smaller than 1 µm are not plotted, then we get an idea of the lower realistic boundaries for the epi thickness for the breakdown voltage we are aiming at (figure 4.4). Of course, the extremely high values of epi thickness (see figure 4.4) correspond with extremely high breakdown voltages (see figure 4.3). Therefore, the upper limit for breakdown voltages is set to 120 V; that is, all optimal epi thicknesses corresponding to a voltage larger than 120 V are set to 10^{-4}. The result is shown in figure 4.5, where it can be seen that the ranges for the epi thickness and the doping levels are narrowed down seriously for the voltage ranges we are interested in. Figure 4.6 shows that the value of the corresponding optimal RESURF dose lays between $1e12$ and $2.5e12\,\mathrm{cm}^{-2}$, which is in agreement with the values that are given in literature [Lud00a].

Figure 4.5 shows that when a breakdown voltage between 80 V and 120 V is desired, the n-epi doping level has to be between $3e15$ and $2e16\,\mathrm{cm}^{-3}$, and the substrate doping level has either to be high (between $5e16$ and $1e19\,\mathrm{cm}^{-3}$) for the lower n-epi doping levels, or has to be low (between $5e15$ and $5e16\,\mathrm{cm}^{-3}$) for the higher n-epi doping levels. This, however, does not mean that the doping levels can not be lower. Only, when they are, the optimal RESURF epi thickness is higher than what is shown in figure 4.5 and the corresponding ideal RESURF diode has a breakdown voltage higher than 120 V. The upper limit on the n-epi

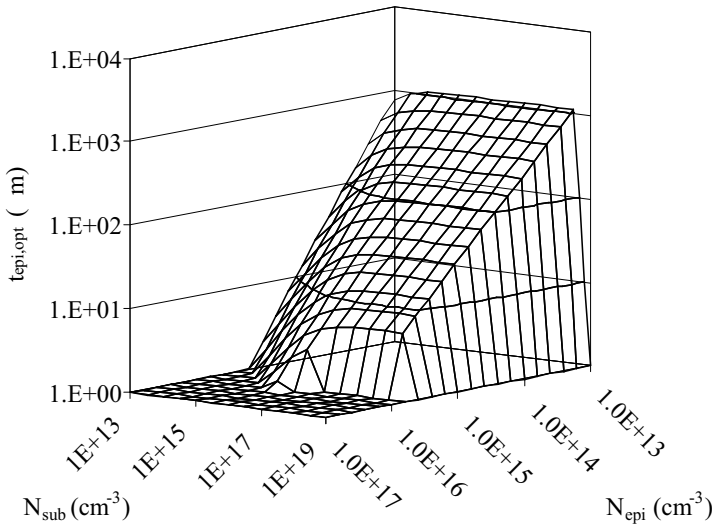


Figure 4.4 Optimal RESURF epi thicknesses larger than 1 µm and corresponding to an ideal RESURF breakdown larger than 80 V for the entire range of realistic n-epi and p-substrate doping levels.

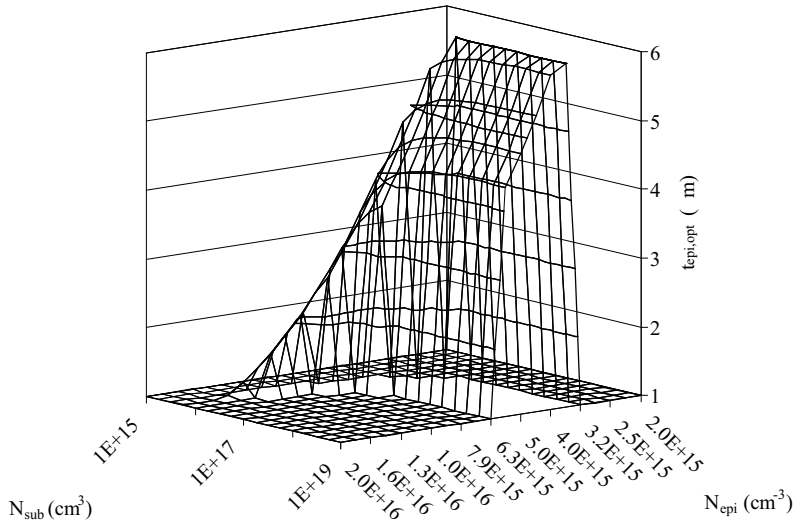


Figure 4.5 Optimal RESURF epi thicknesses larger than 1 µm and corresponding to an ideal RESURF breakdown between 80 V and 120 V.

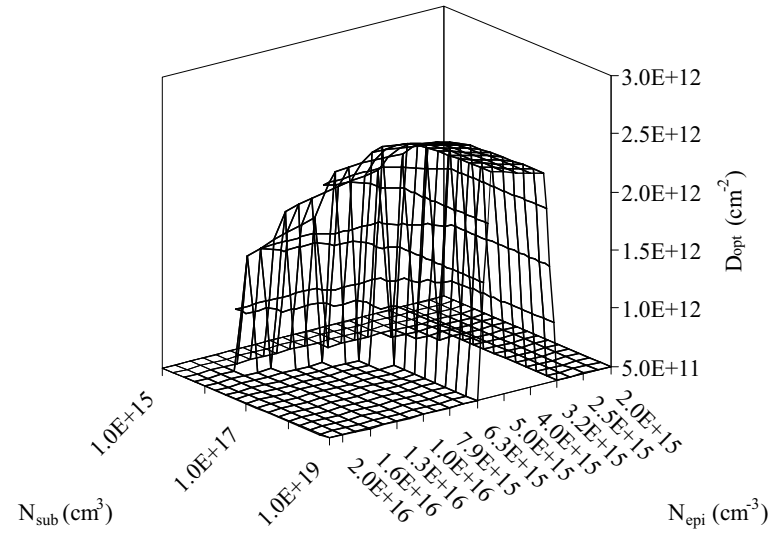


Figure 4.6 Optimal RESURF epi dose D_{opt} for a corresponding ideal RESURF breakdown between 80 V and 120 V, and $t_{epi,opt} > 1\,\mu\text{m}$.

doping level ($2e16\,\text{cm}^{-3}$), on the other hand, can not be crossed, as then the breakdown voltage drops below 80 V or the optimal RESURF epi thickness has an unrealistic low value ($< 1\,\mu\text{m}$, see figure 4.4).

Another observation in figure 4.5 is that for a punch-through breakdown voltage between 80 V and 120 V, the p-substrate doping level is higher than the n-epi doping level in virtually all cases. In view of what has been said before, this means that lateral breakdown is likely to occur first in such a structure. When the p^+–n-epi junction curvature (place A in figure 4.1) is very gentle, the punch-through almost happens at the same time at place A and place C when $l \approx t_{epi}$. This leads to the very important conclusion that increasing l to values larger than t_{epi} for RESURF structures that have a punch-through breakdown voltage between 80 V and 120 V makes no sense (as it does not increase the breakdown voltage, and thus only increases the distance between cathode and anode; that is, the on-state resistance in a power device using such a RESURF structure as drift region).

Let us consider two examples of different substrate doping level in more detail: the first case having a $N_{sub} = 1e15\,\text{cm}^{-3}$, which is typically a p-substrate doping level, and the second case having a $N_{sub} = 6e17\,\text{cm}^{-3}$, which is typically the doping level of a p-type buried layer (BLP). The optimal RESURF punch-through breakdown voltage, as well as the optimal epi thickness are plotted against the entire range of realis-

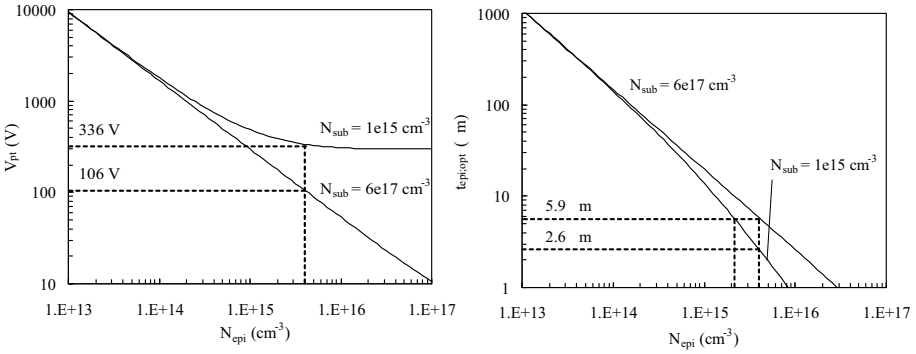


Figure 4.7 The optimal RESURF punch-through breakdown voltage (left), and the optimal epi thickness as a function of the entire range of realistic n-epi doping levels for the cases with $N_{sub} = 1e15\,\mathrm{cm}^{-3}$ and $N_{sub} = 6e17\,\mathrm{cm}^{-3}$.

tic n-epi doping levels for both cases in figure 4.7. For $N_{sub} = 1e15\,\mathrm{cm}^{-3}$, the optimal RESURF breakdown voltage is higher than 300 V *for the entire range of n-epi doping levels*. This is understandable because of the very low substrate doping level. This substrate *can* be used to design an 80 V RESURF structure, if the length of the structure is decreased in orde to obtain premature punch-through along the length. Figure 4.7 also shows that the optimal epi thickness varies for the p-substrate and the BLP if the same n-epi doping level is present, which is inevitable at first sight. Yet it could be possible to use both the p-substrate and the BLP as a basis of RESURF structures in the same technology, e.g., by introducing an extra n-type implantation to increase the epi doping level. One could choose an epi thickness and doping level for the n-epi that gives an optimal RESURF structure upon a $N_{sub} = 1e15\,\mathrm{cm}^{-3}$ (e.g., 5.9 µm and $\sim 2e15\,\mathrm{cm}^{-3}$, see figure 4.7) and increase the net epi doping level (to $4e15\,\mathrm{cm}^{-3}$, determined by the optimal epi thickness) for an optimal RESURF structure upon the BLP. This would result in two RESURF structures: one capable of blocking more than 300 V (p-substrate based) and one capable of blocking 100 V (BLP based).

So, the second example with the higher substrate doping level, representing a BLP, can make an *optimal* RESURF diode in the desired voltage range. If $N_{epi} = 4e15\,\mathrm{cm}^{-3}$, the optimal epi thickness is 5.9 µm with a corresponding punch-through breakdown equal to 106 V (see figure 4.7). Note again that in this case $N_{sub} > N_{epi}$, which means that when a BLP is chosen to create an 80 V RESURF structure, the length of the drift region (l) is ideally equal to the epi thickness (same reasoning as above).

Another extremely important conclusion that can be drawn from figure 4.7 is that the n-epi doping level must not be higher than $5.6e15\,\mathrm{cm}^{-3}$ when using a highly doped buried layer under a RESURF structure and when aiming at 80 V (and when an extra 5 V is taken as possible variation in the fab, then $N_{epi} < 5.3e15\,\mathrm{cm}^{-3}$).

Numerical verification

Both examples given in the previous section, are now verified through numerical 2D TCAD simulations. The process flow used to make the RESURF diodes is arbitrary in order to speed up process simulation. However, they give a very realistic end result as the junctions are still defined by implantation and anneal (no abrupt junctions, contrary to what was assumed in the previous section).

The first example, with a p-substrate doping level of $N_{sub} = 1e15\,\mathrm{cm}^{-3}$ and a n-epi doping level of $N_{epi} = 4e15\,\mathrm{cm}^{-3}$, should have an ideal RESURF epi thickness around 2.6 µm with a breakdown voltage as high as 336 V (see figure 4.7). As has been said above, this is also a *true* RESURF case as $N_{sub} < N_{epi}$, and thus the breakdown voltage in this structure can be much higher than the lateral diode breakdown voltage determined by the n-epi doping level and equal to 106 V (that is, for a p^+–n-epi parallel plane abrupt junction). The condition for the ideal RESURF case being that the length of the RESURF diode l must be much larger than the epi thickness. The results are shown in figure 4.8 with different large l values ranging from 10 µm to 60 µm.

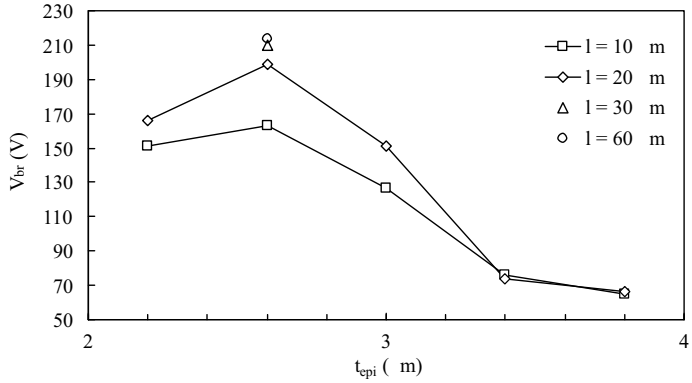


Figure 4.8 Simulated breakdown voltage in a RESURF diode as a function of the epi thickness for the case where $N_{sub} = 1e15\,\mathrm{cm}^{-3} < N_{epi} = 4e15\,\mathrm{cm}^{-3}$.

The simulation confirms that the optimal epi thickness is 2.6 µm, exactly the value that is predicted via simple calculations. The breakdown

voltage, unfortunately, is lower that what is predicted (around 200 V instead of 336 V). However, the breakdown voltage is almost double that of the lateral diode and thus proves that the RESURF effect is acting. The reason why the breakdown voltage is lower that the ideal RESURF value is twofold. First of all, there is the depth of the n^+ region, which was not taken into account in the previous section. As the optimal epi thickness is as low is 2.6 µm in this example, the thickness of the n^+ region, which is about 0.5 µm, can not be neglected. This has a serious effect on the punch-through voltage of the vertical diode, which determines the breakdown of the ideal RESURF structure. Secondly, the curvature of the n^+–n-epi junction also lowers the theoretically calculated breakdown voltage. Furthermore, the simulation also confirms that when the epi thickness is lower than the optimal value, premature punch-through happens in the vertical diode (the so-called over-resurfed case). When the epi thickness is higher than the optimal value, lateral diode breakdown occurs, which is —due to the curvature in the p^+–n-epi junction— lower than the 106 V calculated for the case of an abrupt parallel plane junction.

The second example treats the case of a highly doped p-substrate $N_{sub} = 6e17\,\mathrm{cm}^{-3}$, with the same n-epi doping level as above ($N_{epi} = 4e15\,\mathrm{cm}^{-3}$). Since $N_{sub} > N_{epi}$, lateral and vertical diode punch-through breakdown happen virtually at the same moment (depending on the curvatures in the junctions). As is explained above, there is no sense in taking a RESURF diode length much larger than the optimal epi thickness, which is calculated to be 5.9 µm. The results of the simulations are shown in figure 4.9.

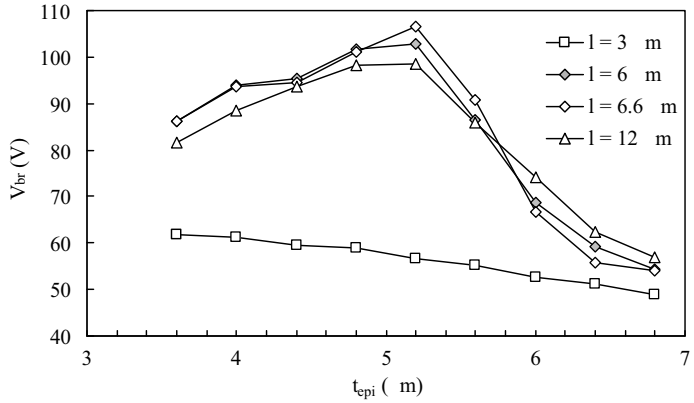


Figure 4.9 Simulated breakdown voltage in a RESURF diode as a function of the epi thickness for the case where $N_{sub} = 6e17\,\mathrm{cm}^{-3} > N_{epi} = 4e15\,\mathrm{cm}^{-3}$.

The simulated optimal epi thickness is smaller than the calculated one, while the breakdown voltage is suspiciously close to the theoretical value. The reason being that although the curvature in the junctions tends to lower the breakdown voltage, other effects such as the finite doping levels of the p^+, n^+ and p-substrate, and the non-abrupt junctions increase the breakdown voltage. The net result is that the simulated optimal breakdown voltage is exactly equal to the calculated one (106 V).

Because of the lateral out-diffusions of both the p^+ and n^+ regions the net or "pure" RESURF diode length l is about 5.2 µm for the case with the maximum breakdown voltage ($l = 6.6\,\mu m$). This means that $l \approx t_{epi}$ really gives the best results, as previously explained. The reason why the optimum drops again for higher l values is that the 2D RESURF effect is no longer performing at its best (see figure 4.10). Nonetheless, a rather large variation around the optimal l is allowed without degrading the breakdown voltage too much (see cases for $l = 6\,\mu m$ and $l = 12\,\mu m$). It is obvious that smaller values of l make it possible to scale the breakdown voltage towards smaller values, as then the length of the lateral diode determines the lateral punch-through. This means that in an 80 V technology any RESURF structure can be scaled down toward smaller breakdown voltage and that the on-state resistance of a DMOS with such a structure as its drift region scales down correspondingly as the length of the current path diminishes.

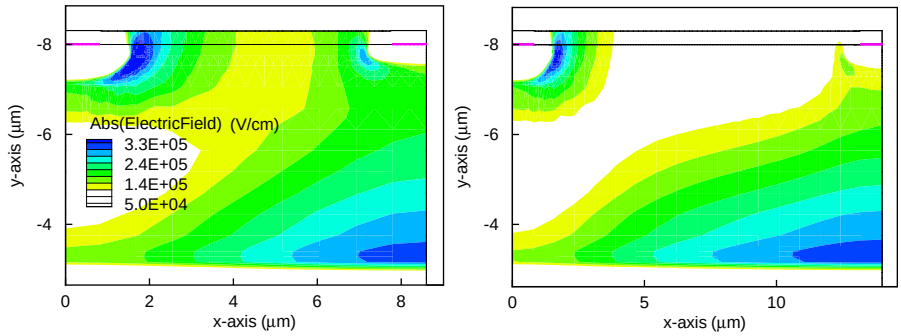


Figure 4.10 Electric field (in absolute value) distribution for two RESURF diodes with optimal epi thickness ($t_{epi} = 5.2\mu m$) and $N_{sub} = 6e17\,cm^{-3} > N_{epi} = 4e15\,cm^{-3}$. The diode on the left has a $l = 6.6\,\mu m$ and the diode on the right $l = 12\,\mu m$.

4.3 The Silicon Limit

The silicon limit is the analytical expression of the specific on-resistance as a function of the breakdown voltage for an idealized DMOS device. Hence the trade-off between those two most important electrical parameters for power DMOS devices can be visualized in one single plot. It is used as a bench-mark, since all power DMOS devices can be represented by one point on this graph. The closer this point is to the ideal silicon limit, the better the performance of the device. Of course, depending on the form of the power DMOS, the silicon limit—i.e., the analytical expression $R_{on,sp}(V_{br})$—changes as well. Here, we will present the silicon limits that are important for the I3T80 technology and desired voltage range.

The *specific on-resistance* ($R_{on,sp}$) is defined as the on-state (gate fully open, e.g., $V_{gs} = 3.3\,\mathrm{V}$ and typically $V_{ds} = 0.1\,\mathrm{V}$, where the I_d characteristic is linearly dependent on V_{ds}) resistance multiplied by the area (factor cost actually enters into this figure of merit). The area is the width times the pitch. The width is the extension of the device in the third space dimension (that is, the dimension absent in the typical 2D cross-sections, e.g., figure 4.11). The pitch is the total device's length. Note that for instance in figure 4.11 the device is not entirely shown, as this would include several more gates (there is actually an optimal number of gates, but this will be disregarded in the present discussion), and a drain region at the right hand side as well. The pitch is then the total distance between both drain contacts (measured from half drain to half drain contact). The *breakdown voltage* is defined as the V_{ds} voltage at which current starts to flow (measured at a certain current level) when the gate is closed ($V_{gs} = 0\,\mathrm{V}$).

The most cited limit is the one of the vertical DMOS (VDMOS) device. Since we are not working on discrete devices, a VDMOS is integrated on chip as shown in figure 4.11. As it is not a real vertical device (the current is flowing through the buried layer and the sinker towards the surface again), this device is sometimes called the quasi VDMOS or QVDMOS. When calculating the theoretical, ideal relationship between the breakdown voltage and the specific on-resistance of this device, the following assumptions are made:

1. The breakdown between the source and drain of this device is modelled as a parallel-plane, abrupt, p^+/n junction breakdown. In other words, the finite doping level of the pbody and the curvature of the pbody/n-epi junction are neglected. The punch-through

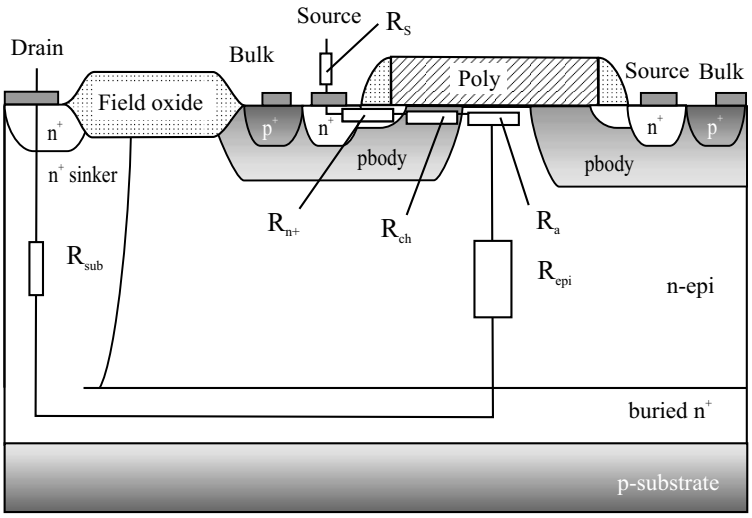


Figure 4.11 Integrated vertical nDMOS.

effect (towards the buried layer) can be included if necessary.

2. The contribution of the drift region to the on-state resistance (in figure 4.11 this is R_{epi}) is considered to be the most important one. Therefore, the contributions of the other regions (see figure 4.11) are neglected. This assumption is inaccurate for low voltage devices (i.e., devices with a small drift region length), but is reliable for higher voltage ranges (table 4.1).

3. Because of the last assumption, the on-resistance is proportional to the drift region's length and inversionally proportional to the drift region's area that is perpendicular to the current flow. Note that in the vertical case the current flow in the device is perpendicular to the axis on which the pitch is defined, whereas in a lateral device the current flow is parallel to the axis on which the pitch is defined. In order to simplify the calculations, in the (integrated) vertical DMOS, the length of the drift region is set equal to the epi thickness and the area through which the current flows is set equal to the pitch times the width of the device. In the lateral DMOS, the drift region's length is set equal to the pitch and the area through which the current flows is set equal to the width times the epi thickness.

Using these assumptions, together with Poisson's equation in the n-epi region, and the resistivity of the n-epi region, a relationship between

Table 4.1 Percentage of the on-resistance that each region of the vertical DMOS takes up (taken from [LDKM99]).

Part	$V_{br} \approx 30\,\mathrm{V}$	$V_{br} \approx 600\,\mathrm{V}$
R_S	7%	0.5%
R_{n^+}	6%	0.5%
R_{ch}	28%	1.5%
R_a	23%	0.5%
R_{epi}	29%	96.5%
R_{sub}	7%	0.5%

the breakdown voltage and the specific on-resistance can be derived.

A general expression of the specific on-resistance, using the second assumption made above, is thus

$$\begin{aligned} R_{on,sp} &= R_{on} \times \text{Area} \\ &= \frac{\text{length}_{\text{drift}}}{\text{area}_{\text{drift}}} \frac{1}{q\mu_{\text{drift}} N_{\text{drift}}} \times \text{pitch} \times \text{width} \end{aligned} \tag{4.5}$$

For a vertical device with the drift region being the n-epi and with the third assumption, this results in

$$\begin{aligned} R_{on,sp} &= \frac{t_{epi}}{\text{pitch} \times \text{width}} \frac{1}{q\mu_n N_{epi}} \times \text{pitch} \times \text{width} \\ &= \frac{t_{epi}}{q\mu_n N_{epi}} \end{aligned} \tag{4.6}$$

where t_{epi} is the epi thickness and N_{epi} is the epi doping level. Important to note is that the pitch of the device disappears in (4.6) (a result of the third assumption). For an ideal, non-punch-through device the epi thickness is equal to the depletion layer width at breakdown. Thus, the specific on-resistance (4.6) can be written as a function of V_{br}, eliminating $t_{epi} = W_{br}$ and N_{epi} by using the expressions (C.10) and (C.11):

$$\begin{aligned} R_{on,sp} &= \frac{2\sqrt{1.8 \times 10^{-35}}}{\epsilon_s \mu_n} V_{br}^{5/2} \\ &= 5.93 \times 10^{-9}\, V_{br}^{5/2}. \end{aligned} \tag{4.7}$$

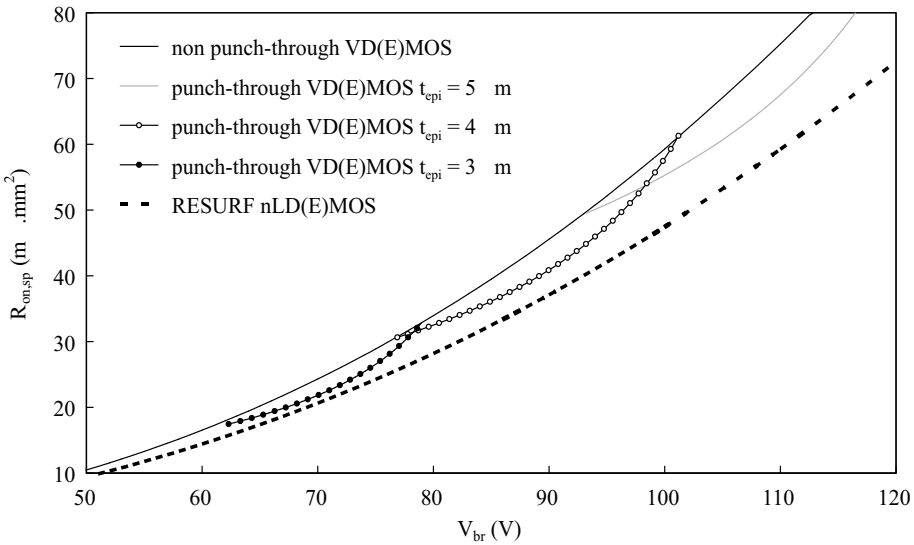


Figure 4.12 Silicon limits for non punch-through and punch-through vertical nD(E)MOS devices, and for a lateral RESURF nD(E)MOS.

This is the well-known ideal silicon limit for non punch-through vertical nD(E)MOS devices, shown in figure 4.12, where the range of breakdown voltages is plotted, which is aimed for in the technology we are interested in. It is also clearly seen that the specific on-resistance increases rapidly with growing breakdown voltage. This proves our previously mentioned statement that the power MOS devices become worse with increasing breakdown voltage, and thus are likely to be replaced by other devices when high breakdown voltages are targeted for ($> 200\,\mathrm{V}$).

Figure 4.12 also shows several other silicon limits, all for different types of D(E)MOS devices. First of all, there are the punch-through VD(E)MOS devices, where the electric field distribution (C.2) is stopped at the buried layer before breakdown occurs.

The relationship (C.15) is used in figure 4.12 to show the small gain a punch-through VD(E)MOS has over a non punch-through device. The n-epi thicknesses are choses arbitrarily—in the range that is interesting for us. The specific on-resistance (4.7), which is determined by an arbitrarily chosen epi thickness and doping level N_{epi} (which, of course, must be such that the depletion layer width at breakdown never becomes smaller than the chosen epi thickness) is plotted against the punch-through breakdown voltage, which on its turn is determined by the chosen epi thickness and doping level N_{epi} through (C.15). Furthermore, the values of the punch-through breakdown that are higher than

in the non punch-through case are not plotted in figure 4.12.

One more silicon limit—that is important for us—is plotted in figure 4.12: the one for lateral RESURF D(E)MOS devices. From the previous section, it is easily understood that a first approximation of the electrical field in the drift region of a RESURF device is given by

$$|E_{\text{const}}| = \frac{V_{br}}{\text{pitch}}. \tag{4.8}$$

Using (C.7) and (C.8) together with (4.8) yields

$$\begin{aligned} 1.8 \times 10^{-35} \int_0^{\text{pitch}} |E_{\text{const}}|^7 \, dx &= 1 \\ 1.8 \times 10^{-35} \, \frac{V_{br}^7}{\text{pitch}^6} &= 1 \\ \text{pitch} &= \left(1.8 \times 10^{-35}\right)^{1/6} V_{br}^{7/6}. \end{aligned} \tag{4.9}$$

This result is used to eliminate the pitch in the expression of the specific on-resistance, all by introducing the breakdown voltage:

$$\begin{aligned} R_{on,sp} &= R_{on} \times \text{Area} \\ &= \frac{\text{pitch}}{t_{epi} \times \text{width}} \frac{1}{q\mu_n N_{epi}} \times \text{pitch} \times \text{width} \\ &= \frac{1}{q\mu_n N_{epi}} \times \frac{\text{pitch}^2}{t_{epi}} \\ &= \left(1.8 \times 10^{-35}\right)^{1/3} \frac{1}{q\mu_n} \frac{1}{N_{epi} t_{epi}} V_{br}^{7/3} \\ &= \left(1.8 \times 10^{-35}\right)^{1/3} \frac{1}{q\mu_n} \frac{1}{D_{opt}} V_{br}^{7/3}, \end{aligned} \tag{4.10}$$

with D_{opt} the optimal RESURF dose as presented in the previous section, for which a constant value of $1e12\,\text{cm}^{-2}$ was chosen. As in the case for vertical devices, the pitch disappears in (4.10). And, once again, this is the result of the third assumption made above.

Interesting to see in figure 4.12 is that the minimum relative difference between the punch-through vertical and the non punch-through vertical devices remains constant, while the minimum relative difference between the punch-through vertical and lateral RESURF devices increases with increasing breakdown voltage (figure 4.13). Of course, the lower the

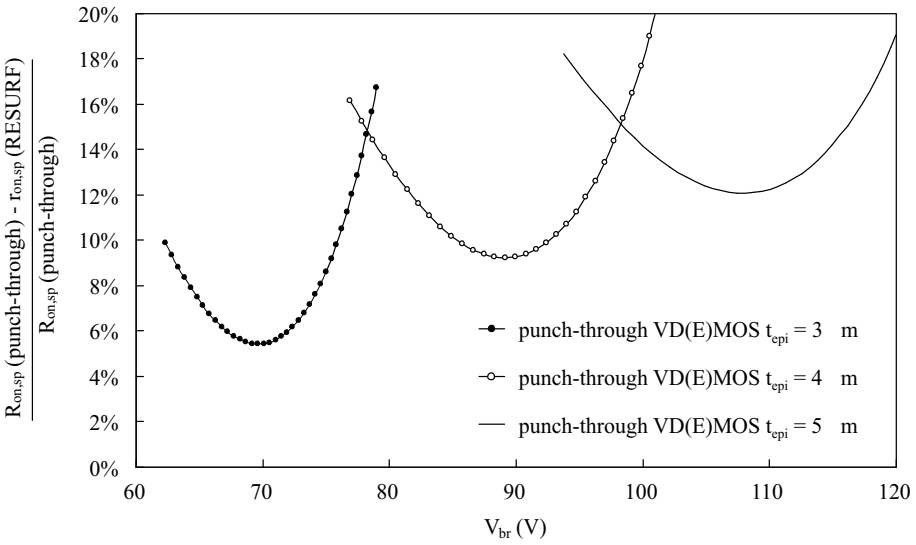


Figure 4.13 Relative difference (in %) between the silicon limit for punch-through vertical nD(E)MOS devices and lateral RESURF nD(E)MOS devices.

breakdown voltage of the device, the less accurate the above made calculations are, and thus the question remains if the vertical punch-through devices are really capable of beating the lateral RESURF devices (which is theoretically the case for smaller epi thicknesses). However, there are several indications that do point out that the vertical punch-through devices could be the best choice:

- If the inaccuracy resulting from the second assumption made above is considered to run parallel for both devices, then the first and third assumption would be the cause for greater differences between both silicon limits as predicted by the above formulas. The most important inaccuracy for the breakdown voltage is caused by the fact that in the vertical device the curvature of the pwell or pbody junction deteriorates the breakdown voltage more in the vertical non-RESURF case than in the lateral RESURF case, as will be seen in the following sections on these devices. The most important inaccuracy for the specific on-resistance is caused by the fact that for the vertical devices one can easily parallel 6, 8 or more channels, while keeping two drain contacts at each end of the device. In the lateral DMOS, each drain is limited to serve only two channels at the same time. The choice thus also depends on the actual use of these devices: as a huge driver (large current

control) or a small switch (smaller current control).

- Moreover, since the theoretical differences are so small, the choice could also be based on other considerations such as the device's robustness or reliability. These are likely to be better in the vertical device, where the current is flowing well under the silicon surface; whereas the current is flowing near the surface for the lateral devices.

So, although the theoretical silicon limit yields the lateral RESURF device as the best choice (at least above 70 V), the practical development of DMOS devices should also consider the vertical device. This discussion solely serves as an illustration of the difficulty when choosing a device in a certain voltage and current range and will be continued in the next sections on the different types of DMOS devices.

We have only discussed the silicon limits of the types of devices in the technology, and the voltage and current ratings we are interested in. Of course, other silicon limits—for other technologies (SOI, see e.g., [HB91] and [Chu00]) and for other types of DMOS devices (e.g., superjuncion [Fuj97])—exist (for a comparison, see [Zin01]).

4.4 Which DMOS: n or p, lateral or vertical, RESURF or not?

N or p-type?

Since the mobility of electrons in silicon is about three times higher than the one of holes in the drift region and since this drift region determines the larger part of the on-state resistance, the n-type DMOS devices have an on-state current level that is three times higher than the p-type devices. Or, in other words, for the same silicon area, the p-type device generates three times less current than the n-type devices. It is clear that the n-type DMOS is preferred to the p-type DMOS.

Yet p-type devices are important for designers as they are the answer to some circuit related problems where otherwise two, three or even more n-type devices are needed. This is because the p-type transistors are in the on-state when $V_{gs} < V_t < 0\,\mathrm{V}$. The fact that the pDMOS needs more area to produce the same current as the nDMOS is then largely compensated for. The pDMOS is normally used as a high-side switch, which means that the source (and bulk) of the device are tight to the supply voltage, while the gate and drain are below source potential,

but above substrate potential. Therefore these devices are referred to as floating devices, since the entire device is lifted above substrate potential. The non-floating devices (used as low-side switch) are then devices where source (and bulk) and substrate can not be put at a potential difference equal to the supply voltage.

Lateral or vertical?

When the term vertical is used for devices that are designed in a VLSI technology, what is really meant is that these devices are vertically integrated (or "quasi" vertical). The current path is eventually flowing perpendicular to the silicon in a vertical direction but is redirected and collected at the surface afterwards. For the DMOS devices, this means that the drain is situated under the silicon surface in the form of a buried layer and that this buried layer is contacted with the metal via a sinker or plug (a highly doped region in the epi).

If we start from a p-substrate, the n-type device needs a highly doped nsinker, a BLN (buried layer of n-type), and a n-epi (see figure 4.11), with a doping level and thickness that are determined by the blocking voltage requirements, as explained above. For a p-type device a psinker, BLP and p-epi is required. The different types of epi can be obtained by selective epi growth, which is technologically difficult. A simpler solution is the introduction of a p-tub layer, which can be implanted and annealed. But this is not all for the p-type device: since we are working on p-substrate and the pDMOS should be a floating device, the drain (BLP) needs to be isolated from the p-substrate. The only possible way to do this in a junction isolated technology is to make use of the BLN (figure 4.14). This results in a rather difficult to realize p-type device, while the lateral pDMOS is much simpler to realize, as only the definition of a pdrift region is needed (see below).

Furthermore, let us recall why these integrated vertical devices should be considered in the first place. It has been proven in the previous section that the breakdown versus on-state trade-off is approximately the same for both the lateral and the vertical device. Therefore, other criteria come into play. First of all, these vertical devices can be parallelled in such a way that the silicon area is used in a very efficient manner: the poly (with two channels) can be repeated without the need to repeat the drain region as well. It suffices to provide the structure with a drain at the left hand side and at the right hand side. The number of gates that can serve the same drain is then determined by the current

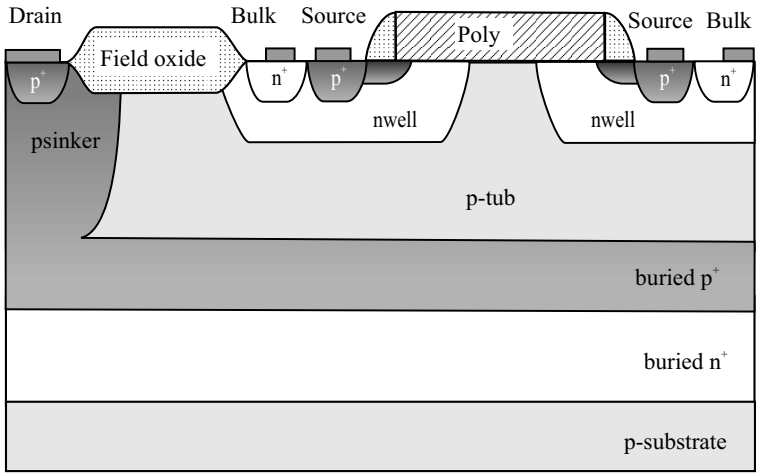


Figure 4.14 Integrated vertical pDMOS.

handling capability of the buried layer. These devices will thus mainly be used as huge drivers capable of controlling large currents. In view of what has been said on n and p-type devices, the first choice for such devices will always be the n-type one. Secondly, as will be seen later on, the vertical devices are robust and have a large safe operating area, which is an important asset in a harsh environment like the automotive application targeted for by the I3T80 technology.

As a conclusion, we can state that the most logical choice for a vertical device is the n-type one. Eventually a p-type vertical device can be made, but there is no much use in such a device, and the lateral form is much easier and safer to integrate. Besides a vertical nDMOS and a lateral pDMOS, there can still be chosen to develop a lateral nDMOS as well. An interesting question is whether or not the RESURF nLDMOS is beating the vertical nDMOS in terms of breakdown voltage versus specific on-resistance in a real—albeit simulated—situation (as in theory, based on very simple calculations, the RESURF device *is* better). An answer will be given in the following sections treating these devices.

RESURF or not?

It is clear that the vertical nDMOS is not a RESURF device, but yet will be considered as well. The lateral pDMOS, on the other hand, is a RESURF device as a pdrift region needs to be defined with the n-epi (and eventually the BLN) acting as "substrate" in the RESURF diode

made this way. A nLDMOS can be made as well, with the p-substrate or the BLP acting as "substrate" and a n-epi, ndrift and/or nwell as RESURF layer. Note, however, that this device is non-floating as it has a parasitic pnp between source and substrate (see figures 4.15 and 4.19).

To make the lateral nDMOS floating a n-type buried layer has to be introduced in this structure, which kills completely the RESURF effect, and leaves the possibity of making a punch-through diode in the drift region of this device. The question is if this drift region is therefore less performant than a drift region of a RESURF device on top of a highly doped p-substrate; as it has been seen that such a RESURF structure has a breakdown voltage determined by punch-through of the lowly doped region as well (in casu the n-epi).

Yet another solution is possible with two buried layers on top of each other. The first buried layer is needed to provide for the RESURF effect (so, a BLP), and the second one is needed to isolate the device from the substrate (so, a BLN). This solution was first proposed in [KPB03], but has been studied simultaneously in the work presented in this book. Here it was first applied on the nLIGBT, and only after noticing [KPB03], this concept was also applied to the DMOS. However, in the DMOS it is used as a method to make the device floating and to achieve a higher breakdown voltage; a technique that actually had already been used in [Lud00b] (with a BLN shorted to the source and bulk of the nLDMOS) and [TYH00]. In the nLIGBT, on the other hand, it is mainly used as a method to suppress substrate currents (see Chapter 5).

4.5 Low-Side, RESURF, n-Type Lateral Drain Extended MOS (nLDEMOS) without Buried Layers

4.5.1 The "true" RESURF effect at work

In a technology based on a n-epi on top of a lowly doped p-substrate, RESURF devices can be made without the use of an extra BLP layer. Hence the name "free" nLDEMOS used hereunder, as they can be made without the need for an extra mask. The standard nwell can eventually be used in the drift region of this device, as depicted in figure 4.15. For the time being, it is only used to smoothen the curvature of the n^+–n-epi junction: $nw = 1.5\,\mu m$. Other important layout parameters are (in µm): $x = 1$, $y = 1$, $t = 10$ and $z = 3$ (for a complete description of the structure, see table 4.2).

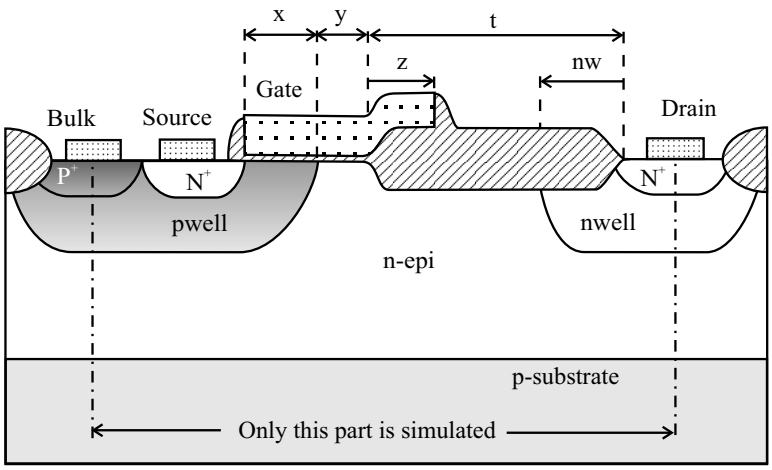


Figure 4.15 The low-side "free" RESURF nLDEMOS device and its most important layout parameters.

For the time being, we assume that the n-epi thickness and doping level are not defined yet and can be freely chosen. As will be discussed in due course, this will not be true if other devices (vertical nDMOS, lateral pDMOS) have to be taken into account as well. Yet from an academic point of view, it is interesting to see the "true" RESURF effect at work, as well as to have an idea about the range of doping levels and thicknesses. These values are mentioned in the following discussion as they do not represent the existing I3T80 technology. For other devices,

Table 4.2 Layout parameters of the "free" RESURF nLDEMOS

Parameter	Description	Value (µm)
$c0^*$	p^+ aa width as bulk region	1.1
$c1^*$	n^+ aa width as source region	1.3
x^*	channel length	varied
y^*	distance between end of channel and start field oxide	varied
t^*	length of field oxide on drift region	varied
$c2^*$	n^+ aa width as drain region	0.8
z	poly overlap on field oxide	varied
nw	nwell overlap on field oxide	varied

Parameters indicated with (*) define the pitch of the device (distance from half-bulk to half-drain): $pitch = (1.1/2) + 1.3 + x + y + t + (0.8/2) = 2.25 + x + y + t$.

the real I3T80 processing conditions (e.g., the n-epi for the nVDMOS, the pdrift for the pLDMOS) need to be used and will not be mentioned because of confidentiality.

In the section on the RESURF effect, it was shown that the optimal epi thickness is around 2.6µm for $N_{epi} = 4e15\,\mathrm{cm}^{-3}$. As this is a realistic epi thickness, a variation around this epi doping level is carried out. Since $N_{sub} < N_{epi}$, a true RESURF effect can occur here, provided that the length of the drift region ($\approx y + t - nw$) is larger than the net n-epi thickness, as explained above. Therefore t is rather large to start with and variations on this parameter are given below.

Figure 4.16 shows that the breakdown voltage clearly has an optimum as a function of the epi thickness for each N_{epi}. Furthermore, this optimum is constant ($V_{br} = 175\,\mathrm{V}$) for all n-epi doping levels (only the highest doping level has a slightly lower breakdown voltage). As this optimum is 175 V, it is comparable to what was simulated for the RESURF diode with a comparable length ($l \approx y + t - nw = 9.5\,\mu\mathrm{m}$). The optimal epi thickness is higher than what was simulated for the RESURF diode. This is caused by the change in the process flow between the simulation of the RESURF diode (arbitrary flow) and the simulation of the nLDMOS (the I3T80 process flow of AMI Semiconductor), as well as by the change in the RESURF structure (the poly field plate).

Figure 4.17 plots the specific on-resistance versus the epi-thickness for the cases considered above. The filled data points show the optima that are reached for the breakdown voltage. Interesting to observe is that

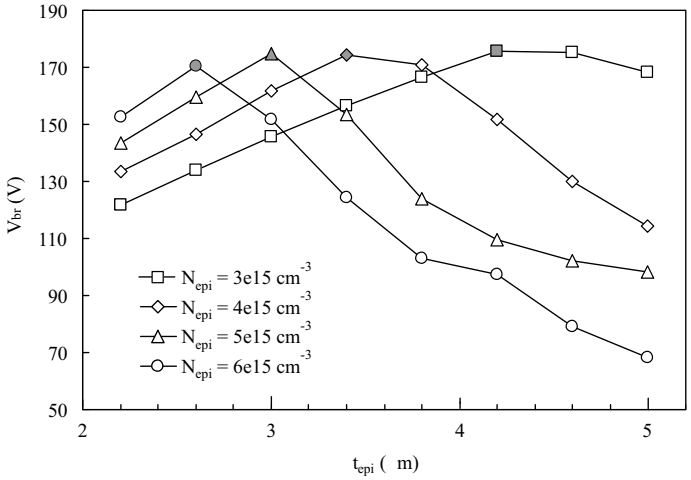


Figure 4.16 Breakdown voltage of the "free" RESURF nLDEMOS device as a function of epi thickness for several n-epi doping levels. Filled data points are the optima for each doping level.

the on-resistance decreases with increasing n-epi doping level, although at the same time the optimal epi thickness decreases. This is explained through the fact that the increasing doping level has a stronger effect on the on-resistance than the decreasing epi thickness, since most of the current in the drift region flows right under the Si/SiO_2 interface.

As the drift region of this DMOS is very much behaving like the RESURF diode in forward blocking mode, the question arises if the breakdown voltage can be further increased by increasing the drift region length, as was the case for the RESURF diode up to a maximum of 210 V (for the case with $N_{epi} = 4e15\,\mathrm{cm}^{-3}$). Figure 4.18 shows that this can be easily done, even up to 220 V. The small mismatch is again accounted for by the changes in the process flow and in the design of the structure. Especially the extension of the poly field plate on top of the field oxide plays an important role (examples will be given for other devices). Of course, the maximum achievable breakdown voltage decreases with increasing n-epi doping level. For $N_{epi} = 6e15\,\mathrm{cm}^{-3}$, for example, $V_{br} \approx 170\,\mathrm{V}$ is already the maximum. Figure 4.18 also demonstrates that the specific on-resistance increases strongly for higher breakdown voltages, an important issue that has been stressed before.

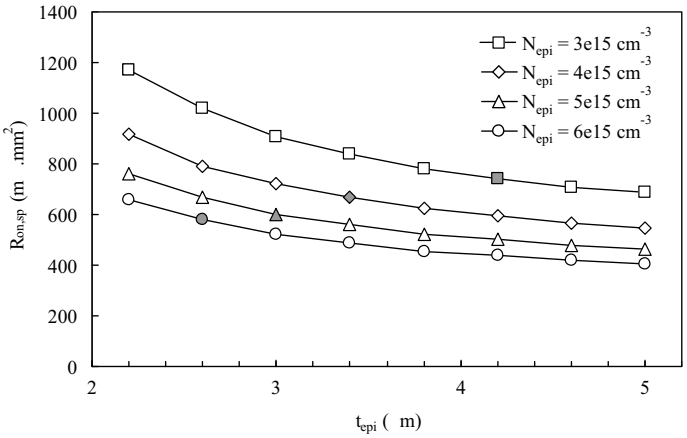


Figure 4.17 Specific on-resistance of the "free" RESURF nLDEMOS device as a function of epi thickness for several n-epi doping levels. Filled data points correspond to the optima in breakdown voltage (figure 4.16).

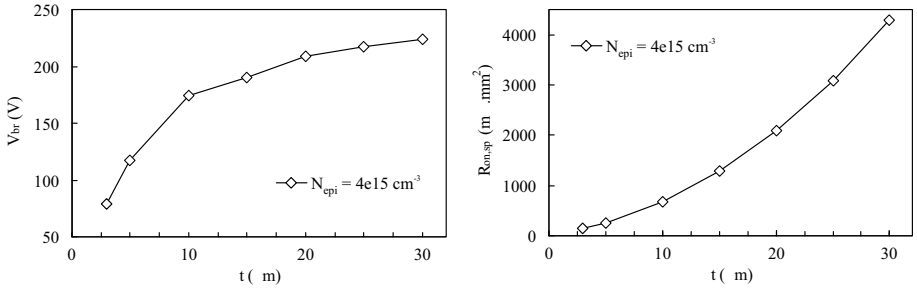


Figure 4.18 Breakdown voltage (left) and specific on-resistance (right) of the "free" RESURF nLDEMOS device as a function of the field oxide length t for one chosen optimal RESURF case ($t_{epi} = 3.4\,\mu\text{m}$ and $N_{epi} = 4e15\,\text{cm}^{-3}$).

4.5.2 Is there a limit to the n-epi doping level ?

The breakdown voltage for a RESURF nLDEMOS with a n-epi doping level as high as $6e15\,\mathrm{cm}^{-3}$ is 170 V. The question is how much further this n-epi doping level can be increased while still assuring a blocking capability of 80 V.

The optimal RESURF breakdown voltages (i.e., the highest possible V_{br} with such and such N_{sub} and N_{epi}) are all higher than 80 V for the nLDEMOS devices designed on a p-substrate with a doping level equal to $1e15\,\mathrm{cm}^{-3}$ (see the left plot of figure 4.7). The only way to optimize this device towards 80 V breakdown is by decreasing its drift region length. As a consequence, the lateral diode shrinks and punches through sooner than the vertical diode. This can be seen in figure 4.24, where clearly the high electrical fields are situated along the lateral diode and no longer along the vertical diode (the high electric field at the p-substrate–n-epi junction has completely disappeared). At the same time, the on-state resistance drops as the length of the current path decreases (table 4.3).

Table 4.3 Free nLDEMOS device with optimal RESURF conditions and shrinked towards 80 V

N_{sub} (cm^{-3})	N_{epi} (cm^{-3})	$t_{epi,opt}$ [a] (µm)	t^b (µm)	z	nw	V_{br} (V)	$R_{on,sp}$ (mΩ.mm^2)	SOA[c] (V)
1e15	4e15	3.4	2.8	$t/3$	$t/3$	82	140	32
1e15	6e15	2.6	2.8	$t/3$	$t/3$	84	118	26
1e15	8e15	2.2	2.8	$t/3$	$t/3$	82	103	26
1e15	1e16	1.8	2.8	$t/3$	$t/3$	81	98	26
1e15	1.2e16	1.4	2.8	$t/3$	$t/3$	80	101	23
1e15	1.4e16	1.4	3.2	$t/3$	$t/3$	87	104	26
1e15	1.6e16	1.0	3.2	$t/3$	$t/3$	82	124	23

[a] Initial value.
[b] Smallest value for which $V_{br} > 80$ V.
[c] Defined as the V_{ds} at the point where $I_{sub}/I_d = 0.001$ (at $V_{gs} = 3.3$ V).

Table 4.3 reveals that the nLDEMOS with a $N_{epi} = 1e16\,\mathrm{cm}^{-3}$ gives the best result. A further increase of the doping level results in a epi-thickness that is so thin that the on-state resistance starts to increase again. So it is not the breakdown voltage but the ever decreasing optimal RESURF epi thickness that imposes an upper limit on the epi doping level. One simulation with a $N_{epi} = 8e16\,\mathrm{cm}^{-3}$ was carried out

to demonstrate this: the result is an 81 V device on a 0.2 µm thick epi with an $Ron, sp = 513\,\mathrm{m\Omega.mm^2}$ (the length of the field oxide has to be 4.4 µm to guarantee a $V_{br} > 80\,\mathrm{V}$).

Table 4.3 also contains a somewhat arbitrarily chosen point on the safe operating area. Yet it gives an idea of the shrinking of the safe operating area once the device is on. This is the major drawback of RESURF devices and will be discussed in the following section on the RESURF nLDEMOS on a highly doped p-substrate.

4.5.3 Using a ndrift ?

The ultra thin epi layers from the previous section might seem unrealistic, especially when a floating vertical DMOS has to made in the same technology (see below). For the case with $N_{epi} = 1e16\,\mathrm{cm^{-3}}$, the optimal RESURF epi thickness is 1.8 µm, which means that the pwell almost touches the p-substrate. So, if a non-floating RESURF lateral nLDEMOS would have to be designed together with a vertical nDMOS in the same technology, it should either be designed with the use of a BLP (see next section) or a new layer should be defined: a ndrift. Provided that the n-epi needed for the vertical DMOS does not yield a RESURF dose that is higher than the dose as can be calculated from table 4.3: for the case $N_{epi} = 1e16\,\mathrm{cm^{-3}}$, $D_{opt} = t_{epi} \times N_{epi} = 1.8e12\,\mathrm{cm^{-2}}$. However, as it will be seen in the section on the VDEMOS (see figure 4.31), the n-epi needed to make a VDEMOS has a dose that is much higher than this optimal RESURF dose.

Therefore, the non-floating RESURF nLDEMOS on top of the lowly doped p-substrate can not be combined with a floating vertical nDEMOS for an 80 V technology as obviously the introduction of a ndrift layer can only increase the already present n-epi dose. The only way to combine both a non-floating RESURF nLDEMOS and a floating vertical nDEMOS is by the use of the BLP as this layer can be designed as such that it eats up the surplus of epi dose. The epi dose thus obtained can be equal to the optimal RESURF dose in the case of a highly doped substrate (representing the BLP on top of a lowly doped p-substrate). In fact, the next section uses a highly doped p-substrate, so without the definition of a BLP, not only to speed up the simulation time, but also because of reasons of confidentiality. This, however, does not comprise the discussion and will still allow to draw conclusions on which device is performing the best at the end of this chapter.

4.6 Low-Side RESURF nLDEMOS with BLP

4.6.1 $V_{br} - R_{on,sp}$ trade-off

Optimizing the n-epi

In the section on the RESURF effect, it has been shown that a RESURF diode on a highly doped p-substrate has an optimal breakdown voltage when punch-through occurs almost simultaneously between the lateral and vertical diode present in the RESURF structure. This punch-through is determined mainly by the n-epi doping level and thickness, and the length of the diode, which should be the same as the epi thickness. The example of $N_{sub} = 6e17\,\mathrm{cm}^{-3}$ and $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ has been given, with an optimal epi thickness of $5.2\,\mu\mathrm{m}$ and an optimal length $l = 6.6\,\mu\mathrm{m}$, which yielded a punch-through equal to $106\,\mathrm{V}$.

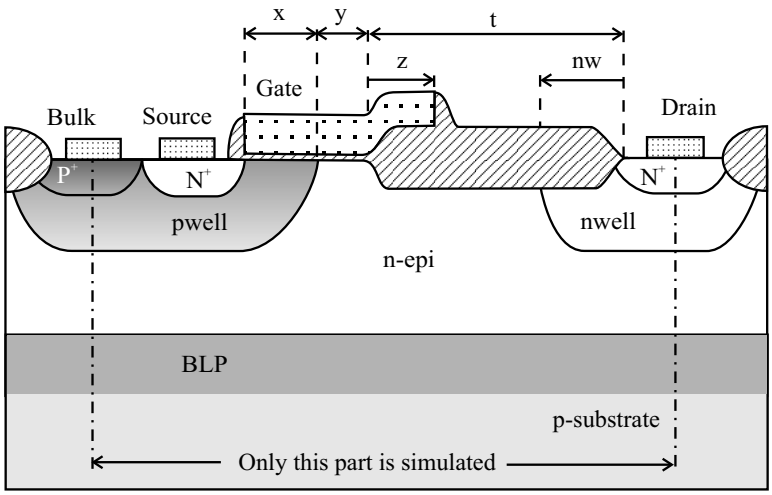


Figure 4.19 The low-side RESURF nLDEMOS device on top of a BLP and its most important layout parameters.

In this section, the same RESURF diode is integrated as the drift region of a nLDEMOS (figure 4.19). The breakdown voltage for these devices are shown in figure 4.20 for the same range of n-epi doping levels and thicknesses as in the example; that is, targeting $80\,\mathrm{V}$. The length of the field oxide t has been set equal to the epi-thickness and will be optimized in simulations treated below. This explains why the specific on-resistance (figure 4.21) increases with increasing epi thicknesss for each n-epi doping level (and because of the previously explained fact that the current flows mainly near the Si/SiO_2 interface, therefore the

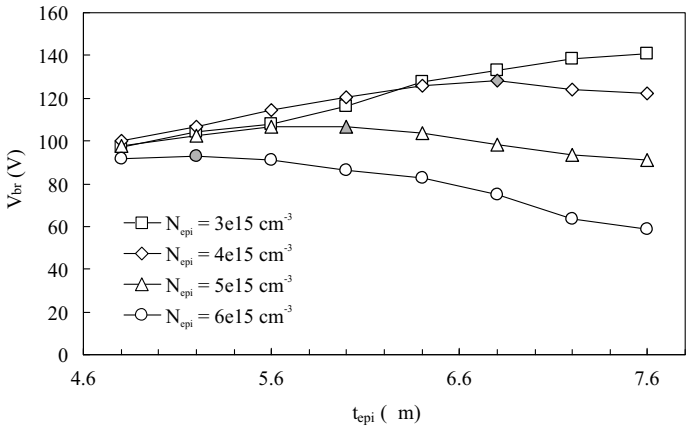


Figure 4.20 Breakdown voltage of the RESURF nLDEMOS device with BLP as a function of epi thickness for several n-epi doping levels. Filled data points are the optima for each doping level.

increasing t has a stronger impact than the increasing t_{epi}).

What is most remarkable in figure 4.20 is that for the n-epi doping level as high as $6e15\,\mathrm{cm}^{-3}$, the breakdown voltage reaches 93 V for the optimal device. This contravenes the conclusions of the section on the RESURF devices, where it was stated that a n-epi doping level of $5.6e15\,\mathrm{cm}^{-3}$ is the absolute maximum if one wants to reach 80 V in the case of $N_{sub} > N_{epi}$. The reason is that the field plate (the extension of the poly on the field oxide, layout parameter z in figure 4.19) acts as an extra boost to the breakdown voltage. Figure 4.22 shows that the n-epi doping can even be increased up to $7e15\,\mathrm{cm}^{-3}$ and the question comes to mind whether some layout variations can further optimize this breakdown voltage.

Optimizing the layout

Only an optimization of the breakdown voltage for the higher n-epi doping levels by changing the drift region length is presented here. It is clear that the extra field plate effect has an influence on the optimal drift region length. Therefore the optimal drift region length might not be equal to the optimal epi thickness, as is the case in the simple RESURF diode structure. Figure 4.23 shows that smaller field oxide lengths ameliorate the performance of the device considerably. Not only does the breakdown voltage increase towards an optimum, the specific on-resistance decreases due to the decreasing drift region length. As a result, a breakdown voltage of 84 V with a specific on-resistance equal to $103\mathrm{m}\Omega.\mathrm{mm}^2$

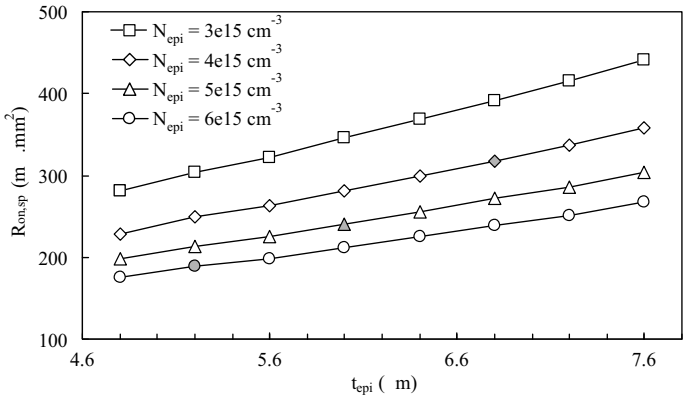


Figure 4.21 Specific on-resistance of the RESURF nLDEMOS device with BLP as a function of epi thickness for several n-epi doping levels. Filled data points correspond to the optima in breakdown voltage (figure 4.20).

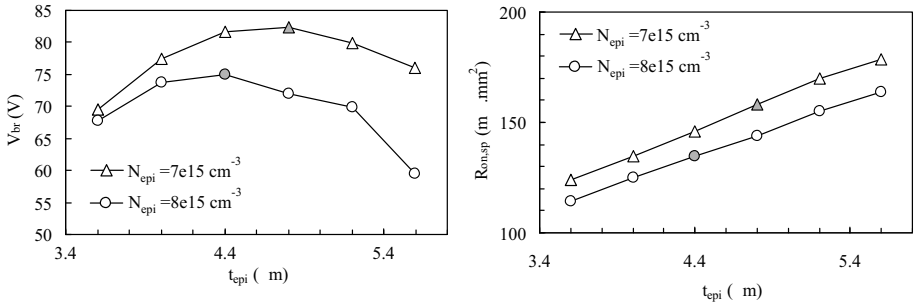


Figure 4.22 Breakdown voltage (left) and specific on-resistance of the RESURF nLDEMOS device with BLP as a function of epi thickness for the highest possible n-epi doping levels. Filled data points are the optima for each doping level.

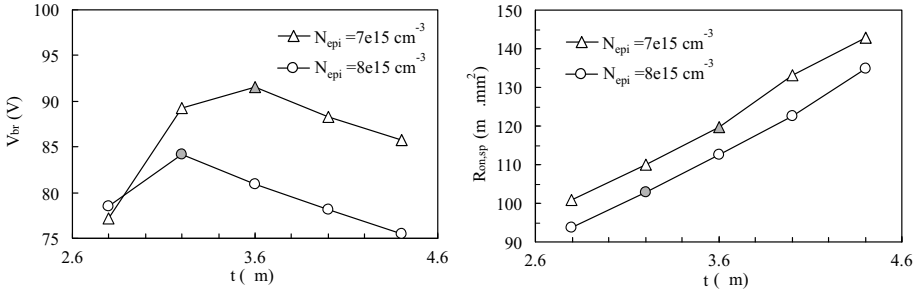


Figure 4.23 Breakdown voltage (left) and specific on-resistance of the RESURF nLDEMOS device with BLP as a function of the field oxide length t for the highest possible n-epi doping levels. Filled data points are the optima for each doping level.

is found for the best device. Moreover, this has been obtained with $N_{epi} = 8e15\,\mathrm{cm}^{-3}$, a doping level that is unexpectedly high, as a parallel plane abrupt p^+–n-epi diode with such a background n-epi doping level breaks down at 63 V ! For these simulations, the optimal epi thickness as plotted in figure 4.22 was used (4.4 µm for $N_{epi} = 8e15\,\mathrm{cm}^{-3}$). Although this value is larger than the optimal field oxide length ($t = 3.2\,\mu\mathrm{m}$), the total drift region's length ($\approx 3.7\,\mu\mathrm{m}$, with the part under the gate oxide included) is approximately equal to the net epi thickness at the end of the process flow ($\approx 3.5\,\mu\mathrm{m}$). This proves once again that the optimal RESURF condition for a structure with a highly doped p-substrate is determined by simultaneous punch-through in both the vertical and the lateral diode present in the RESURF structure.

4.6.2 nLDEMOS with BLP versus Free nLDEMOS

Optimal RESURF doses

When the optimal field oxide length as obtained in figure 4.23 is fed back to simulations with varying epi thickness, the simulations yield the same optimal epi thickness as previously obtained. This proves that the n-epi dose ($= N_{epi} \times t_{epi}$) is dominant when determining the optimal RESURF conditions. The optimal RESURF doses for the nLDEMOS with a lowly doped p-substrate and with a highly doped substrate are summarized in table 4.4. The values are somewhat higher than what was predicted by the simple calculations presented in the section 4.2 of this chapter. Nevertheless, the doses for the lowly doped p-substrate correspond with the figures obtained in [KCA96] through a 2D analytical

model for breakdown voltages in a RESURF diode. The case with a substrate as highly doped as a BLP has—to the best of our knowledge—never been discussed in literature.

Table 4.4 Optimal RESURF doses for the nLDEMOS with a lowly and a highly doped substrate

N_{sub} (cm^{-3})	N_{epi} (cm^{-3})	$t_{epi,opt}$ [a] (µm)	D_{opt} (cm^{-2})
1e15	3e15	4.0	1.2e12
1e15	4e15	3.3	1.3e12
1e15	5e15	2.9	1.5e12
1e15	6e15	2.6	1.6e12
1e15	7e15	2.2	1.5e12
1e15	8e15	2.2	1.8e12
1e15	1e16	2.0	2.0e12
1e15	1.2e16	1.6	1.9e12
1e15	1.4e16	1.6	2.2e12
1e15	1.6e16	1.0	1.6e12
6e17	4e15	6.0	2.4e12
6e17	5e15	5.2	2.6e12
6e17	6e15	4.3	2.6e12
6e17	7e15	3.9	2.7e12
6e17	8e15	3.5	2.8e12

[a] Measured under the field oxide, at the end of the process flow.

Safe operating area

One might wonder if there is any difference between the optimal device of the previous section (with low substrate doping level and small field oxide length) and the best device as obtained with a BLP, as both $V_{br} - R_{on,sp}$ trade-offs are very close to each other (see table 4.5). Layout variations around z and nw have been performed and the values as shown in this table are the best choice. These devices can be further optimized towards lower $R_{on,sp}$ by decreasing the layout parameters x and y, but these changes would run parallel for both devices and are not performed here.

The breakdown voltages of both nLDEMOS devices with the best trade-offs (table 4.5) are very close to each other. Yet we know that the

Table 4.5 nLDEMOS with BLP versus free nLDEMOS

N_{sub} (cm^{-3})	N_{epi} (cm^{-3})	$t_{epi,opt}$[a] (µm)	t[b] (µm)	z	nw	V_{br} (V)	$R_{on,sp}$ (mΩ.mm^2)	SOA[c] (V)
$6e17$[d]	$4e15$	6.8	2.8	$t/3$	$t/3$	83	140	26
$6e17$[d]	$6e15$	5.2	2.8	$t/3$	$t/3$	82	110	26
$6e17$[d]	$8e15$	4.4	3.2	$t/3$	$t/3$	84	103	33
$6e17$[d]	$8e15$	4.8	3.6	$t/3$	$t/3$	81	111	38
$1e15$	$4e15$	3.4	2.8	$t/3$	$t/3$	82	140	32
$1e15$	$6e15$	2.6	2.8	$t/3$	$t/3$	84	118	26
$1e15$	$6e15$	2.6	3.6	$t/3$	1.73 µm	85	133	34
$1e15$	$8e15$	2.2	2.8	$t/3$	$t/3$	82	103	26
$1e15$	$1e16$	1.8	2.8	$t/3$	$t/3$	81	98	26

[a]Initial value.
[b]Smallest value for which $V_{br} > 80\,V$.
[c]Defined as the V_{ds} at the point where $I_{sub}/I_d = 0.001$ (at $V_{gs} = 3.3\,V$).
[d]Represents the device with BLP.

device with the lowly doped substrate has lateral breakdown and the device with the highly doped substrate has lateral and vertical breakdown at the same time. This is shown in figure 4.24, where it is seen that the device with the lowly doped substrate has high electric fields at the pwell–n-epi junction, at the bird's beak under the poly, at the end of the poly plate, and at the nwell–n^+ junction. It has no high electric field at the n-epi–p-substrate junction, as expected.

The device with the highly doped substrate, on the other hand, has high electric fields at the pwell–n-epi junction, at the bird's beak under the poly, at the end of the poly plate, and at the n-epi–p-substrate junction. It has no high electric fields near the nwell–n-epi junction. This proves to be important if the on-state breakdown is taken into account as well. It is clearly seen in figure 4.25 that the high electric fields shift towards the nwell–nepi junction for both devices. This is an example of the Kirk effect working in the nLDEMOS device. Due to the high current levels in the on-state ($V_{gs} = 3.3\,V$), the charges responsible for these currents push the potential lines towards the drain, which results in the high electric fields near the drain.

Because there is already a high electric field at the nwell–n-epi junction at off-state breakdown for the device with the lowly doped p-substrate, the on-state breakdown is reached at a much lower voltage

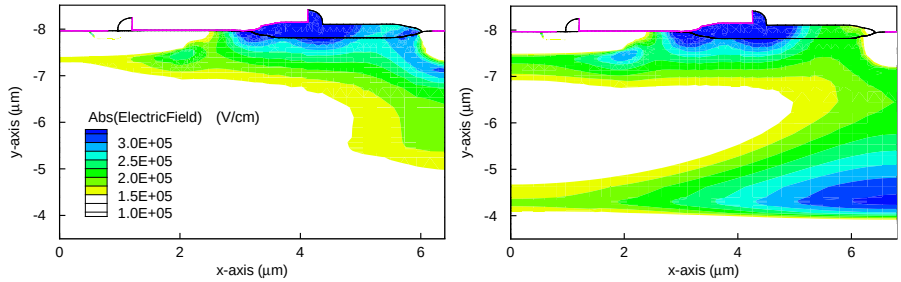


Figure 4.24 Absolute value of the electric field at breakdown for a device with a lowly doped substrate ($= 1e15\,\mathrm{cm}^{-3}$ and $N_{epi} = 6e15\,\mathrm{cm}^{-3}$, left) and a highly doped substrate ($= 6e17\,\mathrm{cm}^{-3}$ and $N_{epi} = 8e15\,\mathrm{cm}^{-3}$).

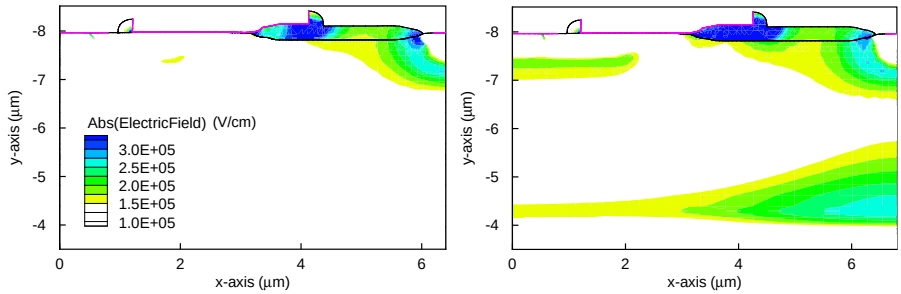


Figure 4.25 Absolute value of the electric field at breakdown in the on-state ($V_{gs} = 3.3\,\mathrm{V}$) for a device with a lowly doped substrate ($= 1e15\,\mathrm{cm}^{-3}$ and $N_{epi} = 6e15\,\mathrm{cm}^{-3}$, left) and a highly doped substrate ($= 6e17\,\mathrm{cm}^{-3}$ and $N_{epi} = 8e15\,\mathrm{cm}^{-3}$).

($55\,\mathrm{V}$, for the device with $N_{epi} = 6e15\,\mathrm{cm}^{-3}$) due to the Kirk effect. The nLDEMOS with the highly doped substrate and with $N_{epi} = 8e15\,\mathrm{cm}^{-3}$ has more defense against the Kirk effect and breaks down at $68\,\mathrm{V}$ in the on-state. The safe operating area as given in table 4.5 is much lower due to its stringent definition. The values for these SOAs do not change much between the lateral RESURF device on p-substrate and the one on BLP. This indicates that the Kirk effect and the corresponding reduction of the SOA is the major drawback of the lateral RESURF devices. One solution to this problem is referred to as the "adaptive RESURF" technique, consisting of the introduction of an extra layer which increases the n-type doping level in the vicinity of the drain (e.g., [HLMP00]). Note that not only the electrical SOA (that is, breakdown in the on-state) causes a problem, but also the lifetime SOA (degradation during switching, e.g., [PHD$^+$02]). One example of such a device is given in table 4.5, using the nwell as adaptive RESURF layer. The

layout parameter nw has been increased to 1.73 µm for the device with $N_{sub} = 1e15\,\mathrm{cm}^{-3}$ and $N_{epi} = 6e15\,\mathrm{cm}^{-3}$, while the length of the field oxide has been increased in order to guarantee a breakdown voltage of 85 V. The result is a device with a larger SOA, but with a lower specific on-resistance, thereby illustrating the extra trade-off that has to be made. The poor increase of the SOA is due to the use of the standard nwell as adaptive RESURF layer. Better results can be obtained if a dedicated implant is defined. This is not discussed here, because the vertical DMOS will prove to have a larger SOA (see the corresponding section).

Another, more drastic, less elegant, but also less expensive way out, is to increase the optimal RESURF dose, thereby creating a slightly "under-resurfed" device in the off-state, but a larger SOA in the on-state (one example of such a device is given in table 4.5 for the case with $N_{sub} = 6e17\,\mathrm{cm}^{-3}$ and $N_{epi} = 8e15\,\mathrm{cm}^{-3}$). A drift region length of 3.6 µm has to be foreseen to guarantee 80 V, with a higher $R_{on,sp}$ as a consequence, but also with a larger SOA. The lower the n-epi doping level is made, the more "under-resurfed" the device can be made while still assuring 80 V, and thus the larger the SOA, the worser the specific on-resistance. The question is whether these RESURF devices can beat the vertical devices when this triple trade-off is taken into account. The answer will be given in the section on the vertical nDMOS.

4.7 High-Side Non-RESURF nLDEMOS with BLN

A nLDEMOS conceived upon a buried layer of n-type, as in figure 4.26, has the advantage that it can be used as a floating (i.e., high-side) device. That is, the drain can be at supply voltage, while the substrate is at ground (potential difference between both is then carried by the lowly doped p-substrate), and while the source, bulk and gate voltages are floating between ground and supply voltage.

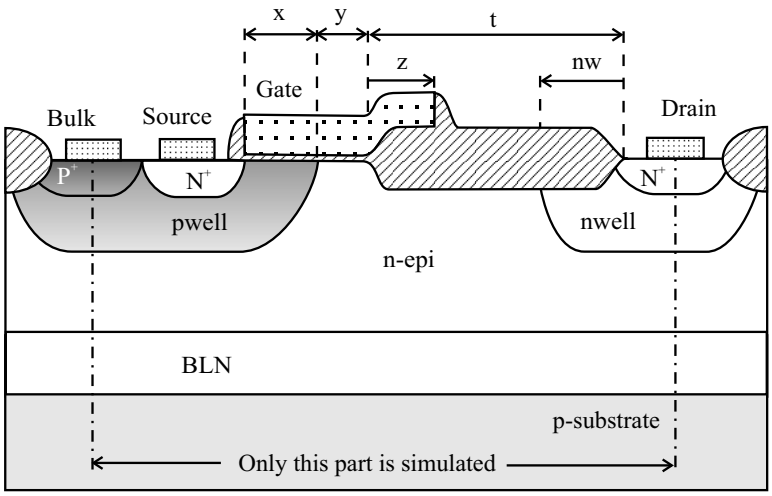


Figure 4.26 A floating non-RESURF nLDEMOS device with BLN and its most important layout parameters.

The major disadvantage of this device is that the RESURF effect is completely killed as the depletion layer caused by the potential difference between drain and substrate only extends into the lowly doped p-substrate and is virtually non-existent in the highly doped BLN. Therefore, the blocking capability between source and drain is now completely provided by the lateral (punch-through) diode with the n-epi as lowly doped layer, which is the dominant factor in determining breakdown. Moreover, since the BLN is at drain potential, the blocking capability has also to be provided by the vertical (punch-through) diode *under the bulk region* of the nLDEMOS, consisting of the pwell, the n-epi, and the BLN.

The breakdown of this vertical diode is determined by the doping level and net epi thickness (between pwell and BLN). There is no chance of creating an extra field peak between the peaks at the pwell–n-epi

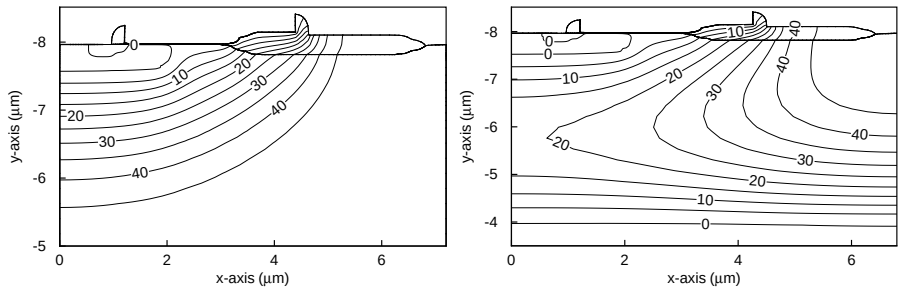


Figure 4.27 Potential lines (5 V steps) at breakdown (V_{br} = 48 V) for a floating nLDEMOS on top of a BLN (left) with a $N_{epi} = 8e15\,\mathrm{cm}^{-3}$. The figure on the right shows the optimal nLDEMOS on top of a BLP with the same n-epi doping level at the same biasing conditions (far below breakdown for this RESURF structure).

junction and at the n-epi–BLN junction, as was the case in the lateral diode of an RESURF nLDEMOS (at the end of the poly field plate, see figure 4.24). A n-epi doping level as high as $8e15\,\mathrm{cm}^{-3}$ is no longer feasible here, as is shown in figure 4.27. The breakdown voltage is as low as 48 V for this n-epi doping level with two regions of high electrical fields: one at the curvature of the pwell–n-epi junction and one at the bird's beak under the gate. In figure 4.27, the optimal device of the previous section with this n-epi doping level is shown as well—at the breakdown voltage of the floating nLDEMOS with BLN. The RESURF effect in this device is already acting at this voltage, pushing the higher potential lines away from under the gate and away from under the pwell–n-epi junction.

It might be interesting to note that although the floating nLDEMOS has an off-state breakdown of 48 V, the on-state breakdown (that is, at V_{gs} = 3.3 V) is as high as 70 V. The increase is due to the Kirk effect, which deteriorates the breakdown from off to on-state in a RESURF nLDEMOS, but the redistribution of the potential lines in the on-state (i.e., the Kirk effect) has a positive impact on the SOA in the floating nLDEMOS of figure 4.27.

It is clear that the n-epi doping level needs to drop if we want to make an 80 V floating nLDEMOS on top of a BLN. Hence, the floating nLDEMOS has a worse $V_{br} - R_{on,sp}$ trade-off than its non-floating counterpart. Figure 4.28 demonstrates that even with $N_{epi} = 4e15\,\mathrm{cm}^{-3}$, the lateral floating device hardly reaches 80 V.

There is little or no variation in the breakdown voltage for large variations of t, which proves that the breakdown is not determined by

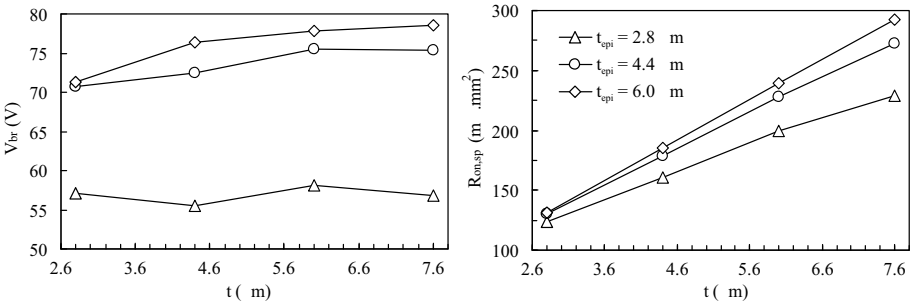


Figure 4.28 Breakdown (left) and specific on-resistance of a floating nLDEMOS on top of a BLN with $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ as a function of the field oxide length t for several variations of epi thickness.

the drift region and that the electric field peaks do not appear in this region. Large variations on the n-epi thickness are also allowed and do not alter the breakdown voltage, nor the specific on-resistance. Only when the epi thickness becomes very small ($t_{epi} = 2.8\,\mu\mathrm{m}$), then both parameters deteriorate. The breakdown voltage decreases because the punch-through breakdown voltage is always lower than the non punch-through breakdown voltage when the n-epi doping level is kept constant (see figure 2.6). From this figure it is also clear that if we want to achieve an 80 V punch-through nLDEMOS with a n-epi on top of a BLN, the n-epi doping level has to be decreased even further. The best device with a $N_{epi} = 3e15\mathrm{cm}^{-3}$ has an $V_{br} = 81\,\mathrm{V}$ and an $R_{on,sp} = 279\,m\Omega.mm^2$ (for $t = t_{epi} = 5.2\,\mu\mathrm{m}$).

The performance of this device can be increased at the expense of one extra mask, a ndrift. The n-epi can then be chosen low enough to insure the desired blocking capability, while the ndrift layer can be introduced in the right part of the drift region to decrease the on-resistance. It might even cause an extra field peak along the drift region as it introduces an extra n-epi–ndrift junction, thereby allowing a reduction of the drift region's length while obtaining the desired breakdown voltage. This option is not studied here as it happens at the (costly) expense of one extra mask. Furthermore, as will be seen in the section on the vertical nDEMOS, the best trade-off for these floating LDMOS devices without the ndrift is already almost two times worse than the best trade-off for the vertical devices, and, it is virtually impossible that the introduction of an extra ndrift can improve the specific on-resistance this much.

4.8 High-Side RESURF nLDEMOS with BLN and BLP

In section 4.4 it has been said that this device is first proposed in [KPB03], where the authors use the denomination FRESURF (Floating RESURF), to stress the fact that a floating layer is introduced in the device (i.e., the BLN in figure 4.29 with $blptofox = -(t + aw)$). The main objective of this device is to keep the RESURF effect, while having a floating (high-side) switch. Exactly the same concept has been developed while performing the work presented in this book, but then in a nLIGBT device, while in search for a method to suppress substrate currents (see chapter 5).

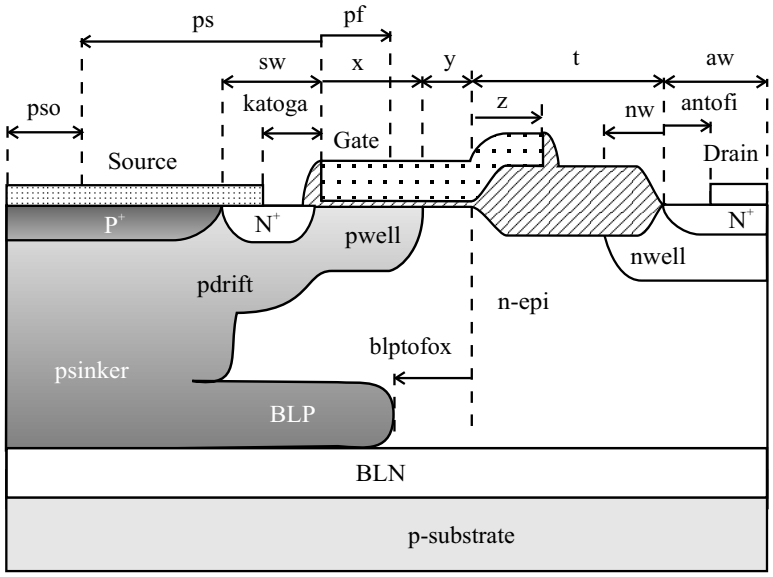


Figure 4.29 A floating RESURF nLDEMOS device with BLN and BLP and all layout parameters defining the device.

In view of what has been said previously in this chapter, it is obvious that a BLP needs to be used to obtain a RESURF effect. And, as a consequence, a BLN has to isolate the device in a junction isolated technology on a p-substrate. If the BLN and the BLP can be made separately from each other (e.g., by implanting and annealing the BLN first, then growing part of the n-epi, implanting and annealing the BLP, and finally growing the rest of the n-epi), a device with the same $V_{br} - R_{on,sp}$ trade-off as the non-floating lateral RESURF nDMOS on a highly

doped substrate can be made. This is illustrated in [KPB03] with $V_{br} = 62\,V - R_{on,sp} = 53\,\mathrm{m\Omega.mm^2}$. Unfortunately, the BLN and BLP in the I3T80 technology have been designed while targeting other requirements than the ones needed for this device.

When the BLN and the BLP are superimposed on each other in the I3T80 technology, the BLN covers a large part of the BLP, which results in a structure that is not ideal for this device. Therefore, only a few examples are given in table4.6.

Table 4.6 V_{br} and $R_{on,sp}$ for several values of the layout parameter *blptofox*

blptofox (µm)	V_{br} (V)	$R_{on,sp}(\mathrm{m\Omega.mm^2})$
-6[a]	98	411
-5[b]	103	410
-4[b]	103	406
-3[c]	30	402
-2[c]	30	395

[a]Blank implant.
[b]BLP diffuses up to under the drain.
[c]BLN is shorted to n-epi, breakdown occurs at BLN–BLP junction.

Due to the properties of the n-epi and the BLP in the real I3T80 process flow, this device results in an "over-resurfed" case which means that the actual present n-epi dose is lower than the optimal RESURF dose. This results in premature breakdown at the nwell–n-epi junction, but is still higher than the 85 V blocking capability of the nVDEMOS in this technology. Unfortunately, the specific on-resistance is also two to three times higher than the value for the VDEMOS. The only use for this device is when 100 V is really a must, which is unlikely in an 80 V technology.

4.9 High-Side Integrated Vertical nDEMOS

4.9.1 $V_{br} - R_{on,sp}$ trade-off

Optimizing the n-epi

The integrated vertical nDEMOS is a very simple device: it has no field oxide along its current path, it has no field plate in the drift region and no nwell nor any other layer in the drift region as well. It solely consists of the nMOS with a BLN as drain situated under the n-epi (figure 4.30). The field oxide near the drain is only there to guarantee a breakdown of the nsinker–n-epi–pwell lateral diode that is higher than the blocking capability of the core device: the vertical BLN–n-epi–pwell (punch-through) diode. Despite its simple layout, it will be shown that the nVDEMOS has a good $V_{br} - R_{on,sp}$ trade-off, has a large SOA, and has a some very robust qualities.

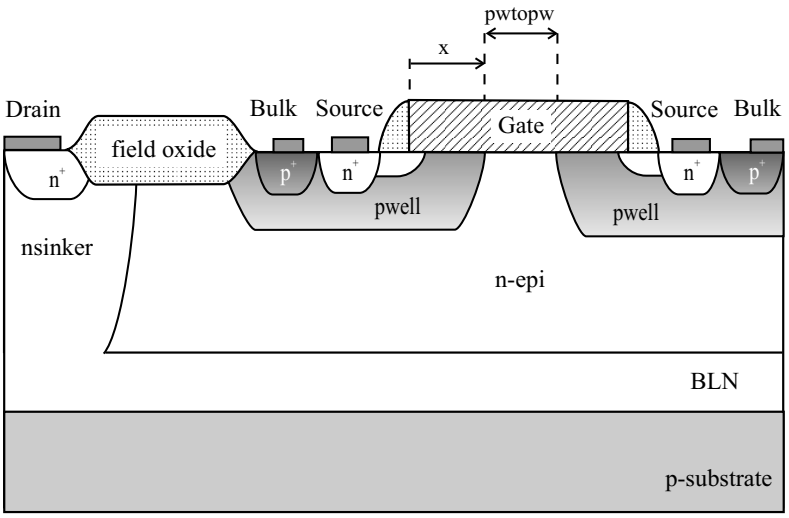


Figure 4.30 A floating integrated vertical nDEMOS device with BLN and its most important layout parameters.

First of all, variations on the n-epi thickness and doping level are presented for a device with an optimal layout (treated below). If we compare figure 4.26 with figure 4.30, it is observed that the same vertical diode is present in both devices. However, this vertical diode is not the only limiting factor in the floating nLDEMOS, as the lateral diode has the same lowly doped region (the n-epi). The weakest (i.e., the one that

breaks down first) obviously determines the breakdown voltage of the device. As it has been seen, the bird's beak under the gate plays an important role and causes breakdown to occur at 75 V even when the n-epi doping level is as low as $4e15\,\mathrm{cm}^{-3}$.

In the nVDEMOS, on the other hand, this vertical diode determines the breakdown voltage, provided that the lateral diode between drain and bulk breaks down at higher values. This is the case due to the large out-diffusion of the nsinker and due to the presence of the pdrift (not yet mentioned here and therefore not drawn in figure 4.30), which has been designed for the pLDEMOS (see below), but proves to be handy to guarantee a high breakdown voltage (as it is used to smoothen the pwell–n-epi junction) in this lateral isolation structure. The influence of the layout parameter *pwtopw* on V_{br} is discussed below, but it is expected that the smaller this value is, the larger the breakdown. This can be understood if one imagines the limit of $pwtopw = 0\,\mu\mathrm{m}$, which results in a parallel-plane junction, and thus in a higher breakdown than in the other limit ($pwtopw = \infty$) where the curvature effect is acting the strongest. We have seen in chapter 2 that a doping level of $5.8e15\mathrm{cm}^{-3}$ is the absolute maximum in an abrupt parallel-plane diode with a breakdown voltage of 80 V. Due to the curvature effect, the doping level of the n-epi in the nVDEMOS has to be lower, as can be seen in figure 4.31: a doping level as high as $5e15\,\mathrm{cm}^{-3}$ is not feasible when 80 V is desired.

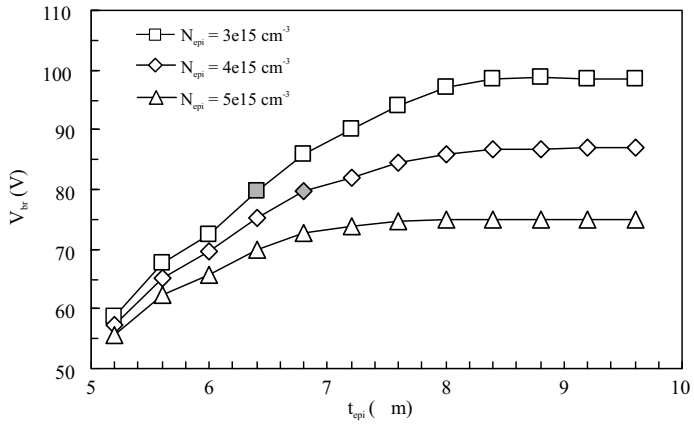


Figure 4.31 Breakdown of a floating nVDEMOS with an optimal pwtopw as a function of the n-epi thickness for the highest possible n-epi doping levels. Filled data points indicate the devices with a breakdown voltage over 80 V with the lowest specific on-resistance for each doping level.

Lower n-epi doping levels guarantee higher breakdown voltages, but result in devices with higher specific on-resistances (figure 4.32). Split lots in the fab have to determine which doping level and thickness give the safest results (i.e., yield always an 80 V device, despite process variations). For reasons of confidentiality, these values are not published here.

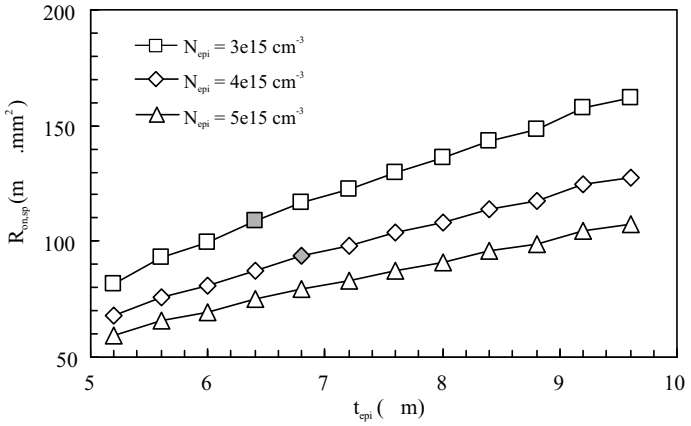


Figure 4.32 Specific on-resistance of a floating nVDEMOS with an optimal pwtopw as a function of the n-epi thickness for the highest possible n-epi doping levels. Filled data points indicate the device with a breakdown voltage over 80 V with the lowest specific on-resistance.

Optimizing the layout

Figure 4.33 plots V_{br} and $R_{on,sp}$ as a function of the key layout parameter *pwtopw* for an optimal n-epi thickness and doping level. The smaller this value, the higher V_{br}, although the degree of the increase is rather small (+8% from the lowest to the highest value). $R_{on,sp}$, on the other hand, increases drastically with decreasing *pwtopw* (+112% from the lowest to the second highest value). It can also be seen that the $R_{on,sp}$ increases again if *pwtopw* gets too large. This is because the growing pitch becomes a stronger factor than the widening of the current path.

In the vertical DMOS as presented here the pitch is the distance between the middle of the two bulk contacts in figure 4.30. Contrary to the simulations of the lateral devices, two parallel devices are simulated here without the drain overhead (the drain is defined at the bottom of the simulation domain). Therefore the simulated specific on-resistances are too low and need to be adjusted. This is difficult as the drain overhead depends on the number of parallel gates. If we assume that the maximum number of gates is applied (this depends on the BLN),

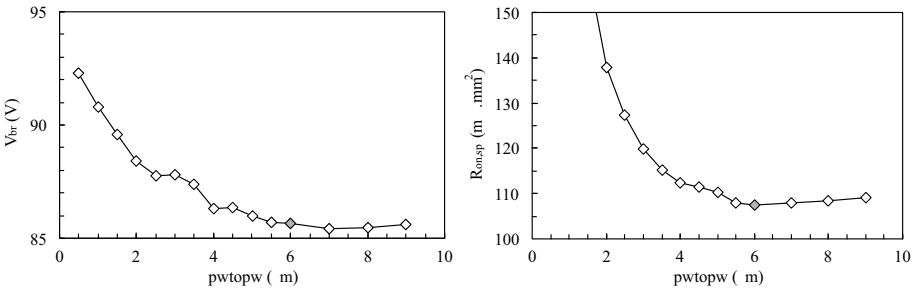


Figure 4.33 Breakdown (left) and specific on-resistance of a floating nVDEMOS with an optimal epi-thickness and doping level as a function of the layout parameter *pwtopw*. Filled data points show the optimum.

then the drain overhead per pair of channels is approximately 3 µm. $R_{on,sp}$ as simulated in figure 4.33 need to be adjusted by adding $\sim 30\%$.

This results in a $V_{br} = 80\,V - R_{on,sp} = 122\,\mathrm{m\Omega.mm^2}$ trade-off for the best vertical nDEMOS, which leads to the very important conclusion (see table 4.7) that the lateral RESURF devices perform better than the vertical DMOS; that is, if only the $V_{br} - R_{on,sp}$ trade-off is considered. Although the floating capability of the vertical DMOS is an important natural asset, the lateral RESURF DMOS can be made floating as well, as has been explained in the previous section. Yet the conclusion that the lateral RESURF DMOS is better than the vertical DMOS turns around if one more trade-off parameter is included: the safe operating area.

Table 4.7 Best non-floating RESURF nLDEMOS on a lowly doped p-substrate and on a BLP, and best floating nVDEMOS compared with each other. The last two non-floating RESURF devices have a n-epi doping level equal to the highest one possible for the nVDEMOS device.

N_{sub} ($\mathrm{cm^{-3}}$)	N_{epi} ($\mathrm{cm^{-3}}$)	$t_{epi,opt}$ (µm)	t (µm)	z	nw	V_{br} (V)	$R_{on,sp}$ ($\mathrm{m\Omega.mm^2}$)	SOA (V)
6e17	8e15	4.4	3.2	$t/3$	$t/3$	84	103	33
1e15	1e16	1.8	2.8	$t/3$	$t/3$	81	98	26
nVDEMOS	4e15	6.8	–	–	–	80	122	80
6e17	4e15	6.8	2.8	$t/3$	$t/3$	83	140	26
1e15	4e15	3.4	2.8	$t/3$	$t/3$	82	140	32

4.9.2 Safe operating area

When the output characteristics for the best DMOS devices with a comparable $V_{br} - R_{on,sp}$ trade-off are compared with each other, it is clear that the nVDEMOS has the largest SOA of them all (figure 4.34). These devices (the non-floating lateral RESURF nDEMOS on a lowly doped p-substrate, the non-floating lateral RESURF nDEMOS on a BLP and the floating vertical nDEMOS, respectively) are all conceived upon the same n-epi doping level, but with a different n-epi thickness. Note that the drain overhead is not simulated, which results in an over estimated $I_{d,sat}$ in figure 4.34 for the vertical device.

As a final conclusion, it can be stated that the RESURF devices are the best devices when the SOA is of no importance (this is actually the case for some applications). When the SOA does have to be large enough, then the nVDEMOS device is the best choice. Yet two important remarks have to be made. First of all, these conclusions have been drawn when only the n-type DMOS is taken into account. When other devices, like the pLDEMOS, have to be included in the same technology, another DMOS device could be the best choice, as now new requirements on the technology could be imposed. Secondly, the fact that the vertical nDMOS devices are inherently floating in the I3T80 technology and the lateral RESURF devices are not, does play a major role as has been argumented before.

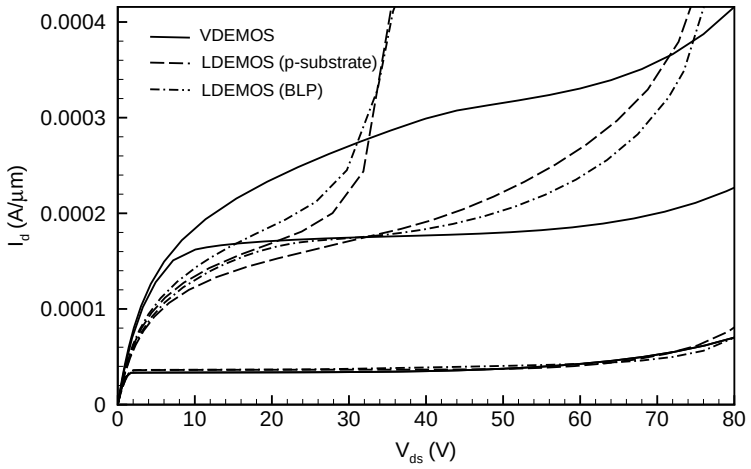


Figure 4.34 $I_d(V_{ds})$ characteristics at $V_{gs} = 1.1, 2.2$ and $3.3\,\mathrm{V}$ for the best floating nVDEMOS and non-floating nLDEMOS devices on a n-epi with $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ (as given in table 4.7). The nVDEMOS has clearly a larger SOA than the lateral RESURF devices.

4.10 High-Side RESURF pLDEMOS

4.10.1 Why lateral ?

Since the n-epi is the top silicon layer, either a ptub or a pdrift has to be defined in the process flow in order to be able to make a vertical pDMOS or a lateral pDMOS, respectively (selective epi growth is not considered here). Therefore, the ptub and pdrift have a doping level that are inevitably higher than the n-epi doping level. If a vertical pDMOS has to be made in combination with one of the best nLDMOS devices (see table 4.7), this would be impossible as the n-epi doping levels are much too high to start with. Moreover, the n-epi thickness would be to thin to realize a pVDMOS, certainly on the lowly doped p-substrate. The n-epi has to be lower doped and thicker, which leaves as only options the vertical nDMOS of table 4.7 and the lateral RESURF nDMOS on a BLP with $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ of table 4.5 — which has an optimal epi thickness that might just be thick enough.

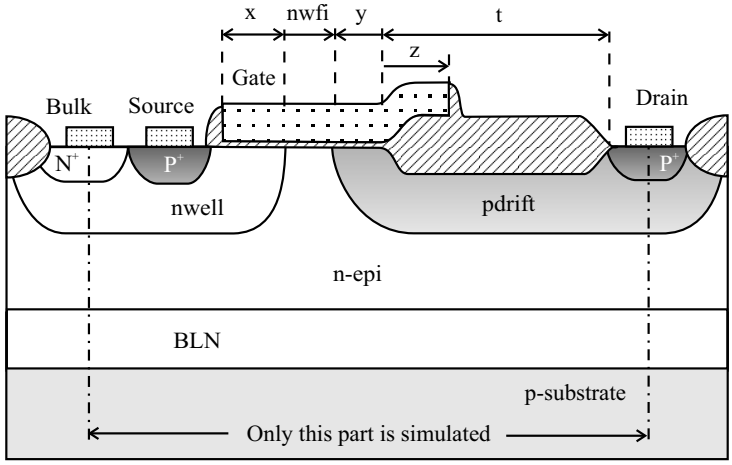


Figure 4.35 A high-side RESURF pLDEMOS device with its most important layout parameters (described in table 4.8).

However, the ptub is higher doped than the n-epi and $N_{epi} = 4e15\,\mathrm{cm}^{-3}$ is the maximum in a nVDMOS, which means that it is also approximately the maximum for a ptub doping level in a pVDMOS. An 80 V pVDMOS is not possible unless the doping level of the n-epi is decreased further. This deteriorates the $V_{br} - R_{on,sp}$ trade-off for the n-type devices, and, as explained in section 4.4, the p-type has to yield to the n-type device. The pVDMOS is thus not an option in this technology, and the only possible p-type device is a pLDMOS (figure 4.35).

Table 4.8 Layout parameters of the pLDEMOS with pdrift

Parameter	Description	Value (μm)
$c0^*$	n^+ aa width as bulk region	1.1
$c1^*$	p^+ aa width as source region	1.3
x^*	channel length	varied
$nwfi^*$	nwell spacing to pdrift region	varied
y^*	pdrift overlap on active area	varied
t^*	length of field oxide on drift region	varied
$c2^*$	p^+ aa width as drain region	0.8
z	poly overlap on field oxide	varied

Parameters indicated with (*) define the pitch of the device (distance from half-bulk to half-drain): $pitch = (1.1/2) + 1.3 + x + y + t + (0.8/2) = 2.25 + x + y + t$.

4.10.2 Conditions sine qua non on the n-epi

What holds for the ptub also holds for the pdrift: it has inevitably a higher doping level than the n-epi. Yet in the lateral case this is not a problem, since the blocking capability is also determined by the n-epi. A "true" RESURF effect is now at hand, and, as was the case for the non-floating nLDEMOS on a lowly doped p-substrate, the pdrift doping level can be a lot higher than the n-epi doping level, as long as the pdrift *dose* fulfils the optimal RESURF conditions. A translation of the expression for the optimal RESURF dose for a n-type structure can be obtained if the N_{sub} and the N_{epi} in equation (4.4) are replaced by N_{epi} and N_{pdrift}, respectively:

$$\begin{aligned} D_{opt,pdrift} &= N_{pdrift} \times t_{pdrift,opt} \\ &= \left(\frac{\epsilon_s}{q}\right)^{7/8} \left(\frac{8N_{pdrift}}{1.8 \times 10^{-35}}\right)^{1/8} \sqrt{\frac{N_{epi}}{(N_{epi} + N_{pdrift})}}. \end{aligned} \tag{4.11}$$

This expression serves as a tool to investigate the possible conflicts that can arise if the pLDMOS is combined with each of the nDMOS devices as given in table 4.7. Each of the n-epi doping levels in this table are used in the expression (C.30) for the punch-through breakdown voltage of a n^+–n–p diode, which also determines the breakdown voltage of the ideal RESURF diode present in the pLDMOS (it is actually a p^+–p–n diode, but this leads to the same expression). These values are plotted in figure 4.36 for all $N_{pdrift} > N_{epi}$. It is observed that the

values of N_{pdrift} corresponding to a $V_{pt} > 80\,\mathrm{V}$ are situated in a very narrow range for the case with a n-epi doping level equal to $1e16\,\mathrm{cm}^{-3}$. Moreover, as can be seen in the plot at the right hand side in figure 4.36 (using equation (4.11)), the optimal pdrift thicknesses resulting from this narrow range of allowed pdrift doping levels are almost equal to the optimal n-epi thickness as given in table 4.7. This rules out the possibility of combining the RESURF nLDMOS on a lowly doped substrate with a RESURF pLDMOS. The RESURF nLDMOS on top of a BLP and with a corresponding ideal n-epi doping level of $8e15\,\mathrm{cm}^{-3}$ results in a larger range of possible pdrift doping levels (roughly between $8e15$ and $1.6e16\,\mathrm{cm}^{-3}$). Yet it still is a very narrow window to work in, which is almost impossible to handle in a fab with all possible process variations on the n-epi and pdrift. So it is safer to work in a lower doped n-epi, say $6e15\,\mathrm{cm}^{-3}$. But, as can be seen in table 4.5, a nLDEMOS with this n-epi doping level has a $V_{br} - R_{on,sp}$ trade-off with a $R_{on,sp}$ that is 10% better than the $R_{on,sp}$ of the nVDEMOS as given in table 4.7 with a n-epi doping level equal to $4e15\,\mathrm{cm}^{-3}$. As this nVDEMOS guarantees a large SOA, it is chosen that we work on this n-epi, which also yields a large range of allowable pdrift doses and corresponding thicknesses (figure 4.36).

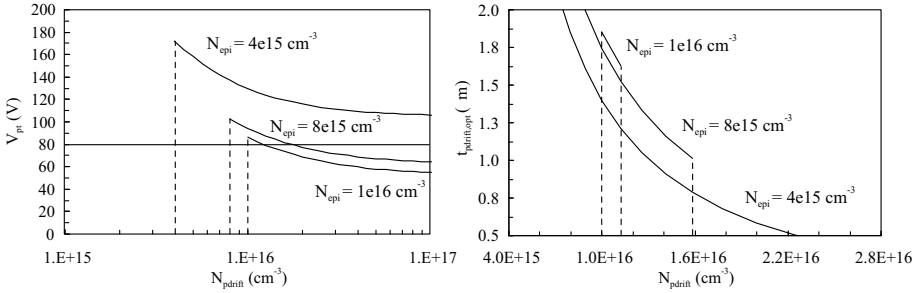


Figure 4.36 Punch-through breakdown of the vertical diode in a pLDEMOS as a function of the pdrift doping level at several n-epi doping levels (left). The optimal RESURF pdrift thickness for the pdrift doping levels corresponding to a $V_{pt} > 80\,\mathrm{V}$ and satisfying the condition $N_{pdrift} > N_{epi}$ (right).

The V_{pt} as plotted in figure 4.36 is based on very simply 1D calculations, and, like is the case in the RESURF nLDEMOS, the actual value is determined by many other factors, as the smoothness of the junctions (the pdrift is implanted and annealed, and there is the n-epi, so the nwell-pdrift junction can be made very smooth), the poly field plate, the thickness of the p^+...

Yet another factor influences the breakdown voltage in the pLDMOS:

the presence of the BLN, which makes the vertical punch-through diode (p^+–pdrift–n-epi) into a double punch-through vertical diode structure: p^+–pdrift–n-epi–BLN. The presence of the BLN is justified by the harsh applications this technology targets for. In an automotive environment the drain can be forward biased with respect to the bulk and source during a very short period (during the turn-off of the pLDMOS with an inductive load). This condition turns on the parasitic bipolar (p^+–n-epi–p-substrate) with high substrate currents as a consequence. When the BLN is introduced, these currents are collected by the BLN as part of a forward biased diode.

4.10.3 pLDEMOS with pwell as pdrift layer

Feasibility of the pwell as pdrift layer

If the pwell could be used as the pdrift layer, a dedicated pdrift layer would not be necessary (figure 4.37).

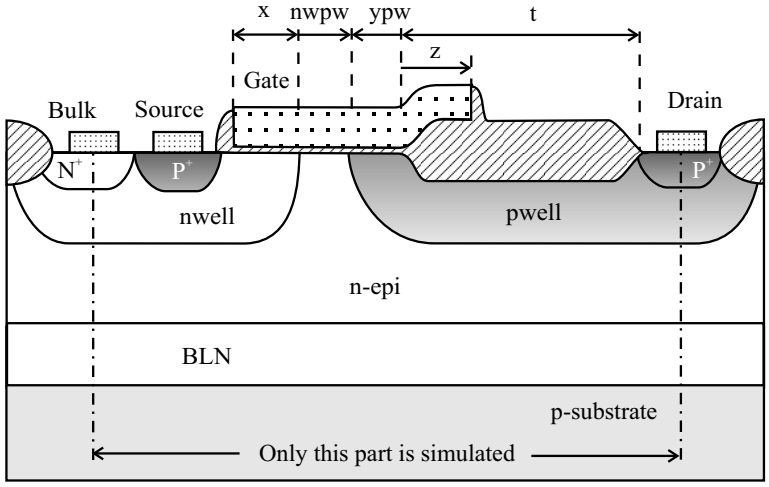


Figure 4.37 A high-side RESURF pLDEMOS device with the standard CMOS pwell as pdrift region and with its most important layout parameters.

If a long drift region under the gate oxide is used (e.g., $ypw \sim 1\,\mu\text{m}$), the total dose present in this part of the drift region is approximately one order of magnitude higher than the dose as calculated from equation (4.11). This results in an "under resurfed" device: at $V_{gs} = 0\,\text{V}$ breakdown occurs near the bird's beak under the gate, while at lower V_{gs} values, the Kirk effect helps to improve the $I_d(V_{ds})$ characteristics. An easy way out, is to layout the devices in such a way that high doses

do not occur in the drift region under the gate, i.e., by decreasing the layout parameter *ypw*.

Layout variations and main measurement results

Figure 4.38 demonstrates the problems encountered when trying to use the pwell as pdrift layer. For the first time, measured data are presented as the development of the p-type DMOS devices was entrusted to the author of this work. The shown measured devices are designed as such that $nwpw + ypw = constant = 1\,\mu m$. Thus, the pitch remains constant as well, i.e., $10.25\,\mu m$, since $t = 6\,\mu m$ is rather large. This means that a larger pdrift region ($+2\,\mu m$) is needed to obtain the same breakdown voltage for devices that use a pwell instead of a pdrift (see below). This larger pitch results in $R_{on,sp}$ values that are roughly two times higher than in the case of a pdrift as pdrift region, but this is not the only reason why the specific on-resistance is too high for pLDEMOS devices with a pwell as pdrift region.

The breakdown voltage clearly drops when *ypw* is becoming too large ($> 0.1\,\mu m$). On the other hand, if *ypw* is negative, $R_{on,sp}$ increases and eventually becomes infinitely high, since the p-type drift region is no longer present between the channel and the actual beginning of the pwell region. It seems that $ypw = 0\,\mu m$ is the only acceptable value. But, if variations in the misalignments of the masks are accounted for, a larger *ypw* value should be taken. As a consequence, the specific on-resistance would be larger than $1\,\Omega.mm^2$, which is unacceptably high and thus eliminates the pwell as possible pdrift region.

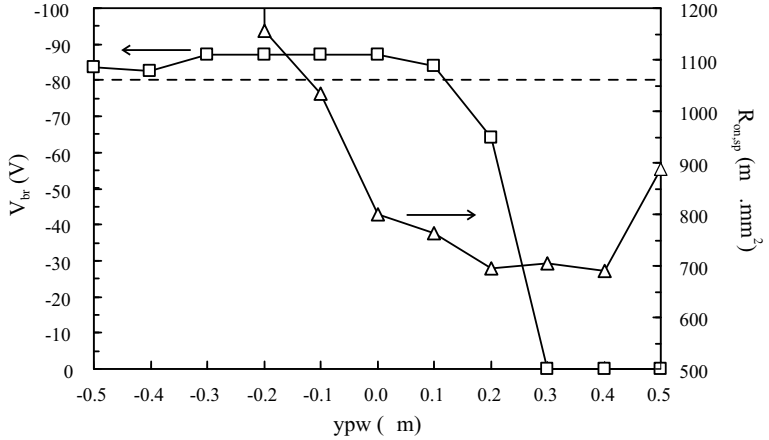


Figure 4.38 *Measured* V_{br} and $R_{on,sp}$ as a function of *ypw*.

4.10.4 pLDEMOS with a dedicated pdrift layer

Introduction of the pdrift layer in the I3T80H process flow

The implant conditions for the pdrift region are in the first place determined by the RESURF condition for the pLDMOS, with consequently the pdrift implant dose as most important parameter. The introduction of the pdrift implant in the process flow is a matter of elimination based on the following criteria:

- The pdrift layer must be introduced before gate oxidation.
- The impact on the CMOS process flow of the pdrift layer must be negligible.
- The pdrift dose must be as stable as possible and must be determined as much as possible by the pdrift implant conditions alone.

This means that ideally the pdrift layer consists of one mask, one implant and a drive that already exists. Since, in general, a graded junction offers higher breakdown voltage than an abrupt junction for the same range of doping levels, a drive with a high thermal budget is preferred. Yet the sinker anneal has too much thermal budget, and the other anneals appear after field oxidation. This would imply a change in the pdrift profile between the active area and the drift region under the field oxide (not only because of the different Si/SiO_2 surface but also due to process variations of the field oxide thickness which would induce variations between different lots). This leaves actually only one option: just before field oxidation.

Determining the pdrift implant conditions

Figure 4.39 and 4.40 demonstrate that the pdrift implant dose is a vital parameter. It shows the breakdown voltage V_{br} and specific on-resistance $R_{on,sp}$ as a function of the pdrift implant dose for three different pdrift implant energies. All devices have the following layout parameters (in µm): *nwfi* = 0, $t = 9$, $z = 3$, $x = 1$ and $y = 1$. As in the section on the RESURF diodes and on the nLDEMOS on a lowly doped substrate, the length of the drift region has to be larger than the thickness of the RESURF layer. Since the double punch-through structure (p^+–pdrift–n-epi–BLN) is a new feature that has not been studied yet, the field oxide has been chosen larger than the total thickness of this double punch-through diode. Optimization of this parameter is discussed below.

The optimal pdrift dose *as implanted* shifts to higher values for decreasing energies, because lower energies result in less deep implantations, and more boron segregates into the field oxide. The optimal pdrift dose (equation (4.11)) is approximately given by integrating the netto doping versus depth (up to the pdrift–n-epi junction) and remains almost constant.

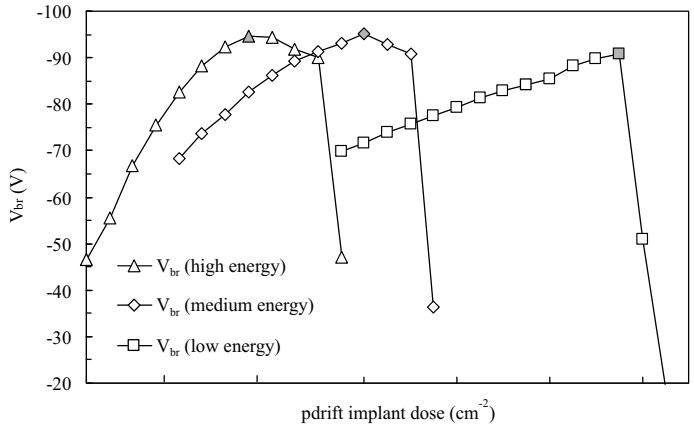


Figure 4.39 Breakdown voltage V_{br} as a function of the prift implant dose for several implant energies. Filled data points show the maxima.

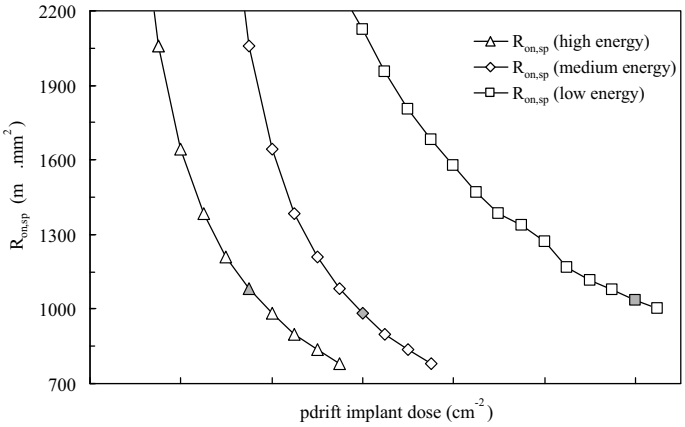


Figure 4.40 Specific on-resistance $R_{on,sp}$ as a function of the prift implant dose for several implant energies. Filled data points correspond to the maxima in V_{br} (figure 4.39).

Figures 4.39 and 4.40 show that over a broad range of pdrift implant energies (the highest energy is three times the lowest), the optimal V_{br}

and $R_{on,sp}$ are roughly the same. Yet there is one important reason to choose for the highest energy. Namely that it has a broader range of safe pdrift implant doses around which the V_{br} remains above target. This is because the higher the energy is, the less boron segregates into the field oxide and the more control over the pdrift region's dose there is.

Layout variations and main measurement results

The layout as used in the previous paragraph can be optimized. First of all, the length of the drift region, with the foremost parameter t, has to be optimized towards 85 V. Figure 4.41 shows that the smallest value for this parameter is around 4 µm.

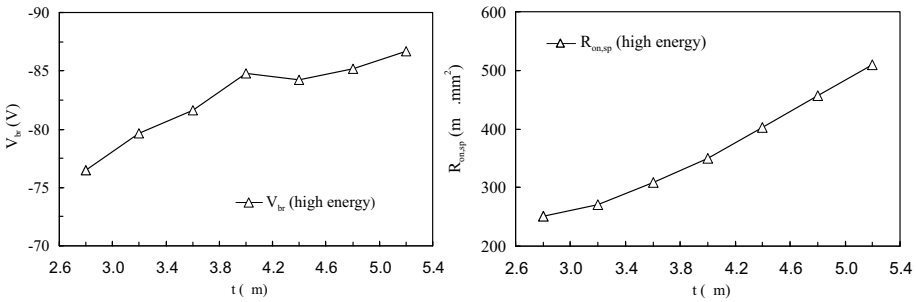


Figure 4.41 Breakdown voltage V_{br} and specific on-resistance $R_{on,sp}$ as a function of the layout parameter t for the highest pdrift implant energy with its corresponding optimal pdrift implant dose (simulated results).

Secondly, the following considerations are taken into account while defining the layout variations to be put on the test chip, for which the frames with all possible layout variations are given in tables 4.9 to 4.12. The channel length x can still be optimized towards a smaller value. Short channel effects can be simulated, but it is hard to simulate the exact x value for which those effects start to play an important role. Eventually, the layout variations as in table 4.10 and table 4.11 can be used to calibrate the lateral out-diffusions of nwell and pdrift.

The drift region under the gate can be taken smaller too, i.e., *nwfi* and y need to be optimized (e.g., in table 4.9). And last but not least, layout variations on t and z have to be put on the test chip, since a correct simulation of the absolute value for the breakdown voltage is unlikely with an uncalibrated n-epi and pdrift doping profile (see all frames given below). Indeed, the reader has to be aware of the fact that all simulations were performed without calibration of the newly defined layers (BLN, n-epi & pdrift). The SIMS profiles only became

available after the process conditions and the test chip were defined. Note that some devices with pwell *and* pdrift were put on the test chip as well. These devices are not discussed here as it turns out in both the simulations and the measurements that the extra pwell (added to the drain side of the drift region) deteriorates the $V_{br} - R_{on,sp}$ trade-off. Furthermore, as will be seen in the next paragraph, the SOA is already large enough for the pLDEMOS (so adaptive RESURF is not an issue like it was in the RESURF nLDEMOS).

Table 4.9 pLDEMOS with pdrift: frame 1

Pad	Connection	poly on nwell **x**	nwell to pdrift **nwfi**	pdrift on active poly **y**	fox length **t**	poly on fox **z**	Width **W**
1	Common S						
2	D	1	1	1.2	4	2	40
3	D	1	1	1	4	2	40
4	D	1	1	0.8	4	2	40
5	D	1	1	0.6	4	2	40
6	D	1	1	0.4	4	2	40
7	D	1	1	0.2	4	2	40
8	D	1	1	0	4	2	40
9	D	1	1	-0.2	4	2	40
10	p-substrate						
11	D	1	1.2	1	4	2	40
12	D	1	1	1	4	2	40
13	D	1	0.8	1	4	2	40
14	D	1	0.6	1	4	2	40
15	D	1	0.4	1	4	2	40
16	D	1	0.2	1	4	2	40
17	D	1	0	1	4	2	40
18	D	1	-0.2	1	4	2	40
19	Common B						
20	Common G						

Table 4.10 pLDEMOS with pdrift: frame 2

Pad	Connection	poly on nwell **x**	nwell to pdrift **nwfi**	pdrift on active poly **y**	fox length **t**	poly on fox **z**	Width **W**
1	Common S						
2	D	0.8	1	1	4	2	40
3	D	0.6	1	1	4	2	40
4	D	0.4	1	1	4	2	40
5	D	0.8	1	0.5	4	2	40
6	D	0.6	1	0.5	4	2	40
7	D	0.4	1	0.5	4	2	40
8	D	0.8	0.5	1	4	2	40
9	D	0.6	0.5	1	4	2	40
10	p-substrate						
11	D	0.4	0.5	1	4	2	40
12	D	1	0.5	0.5	4	2	40
13	D	0.8	0.5	0.5	4	2	40
14	D	0.6	0.5	0.5	4	2	40
15	D	0.4	0.5	0.5	4	2	40
16	D	0.8	0	1	4	2	40
17	D	0.6	0	1	4	2	40
18	D	0.4	0	1	4	2	40
19	Common B						
20	Common G						

Table 4.11 pLDEMOS with pdrift: frame 3

Pad	Connection	poly on nwell **x**	nwell to pdrift **nwfi**	pdrift on active poly **y**	fox length **t**	poly on fox **z**	Width **W**
1	Common S						
2	D	1	0	0.5	4	2	40
3	D	0.8	0	0.5	4	2	40
4	D	0.6	0	0.5	4	2	40
5	D	0.4	0	0.5	4	2	40
6	D	1	1	1	2	1	40
7	D	1	1	1	3	1.5	40
8	D	1	1	1	3	2	40
9	D	1	0.5	0.5	3	1.5	40
10	p-substrate						
11	D	1	0	0.5	3	1.5	40
12	D	1	1	1	3.5	1.2	40
13	D	1	1	1	3.5	1.75	40
14	D	1	0.5	0.5	3.5	1.75	40
15	D	1	0	0.5	3.5	1.75	40
16	D	1	1	1	3.5	2.3	40
17	D	1	0.5	0.5	3.5	2.3	40
18	D	1	0	0.5	3.5	2.3	40
19	Common B						
20	Common G						

Table 4.12 pLDEMOS with pdrift: frame 4

Pad	Connection	poly on nwell **x**	nwell to pdrift **nwfi**	pdrift on active poly **y**	fox length **t**	poly on fox **z**	Width **W**
1	Common S						
2	D	1	1	1	4	1.3	40
3	D	1	1	1	4	2.7	40
4	D	1	0.5	0.5	4	2.7	40
5	D	1	0	0.5	4	2.7	40
6	D	1	1	1	4.5	1.5	40
7	D	1	1	1	4.5	2.25	40
8	D	1	0.5	0.5	4.5	2.25	40
9	D	1	0	0.5	4.5	2.25	40
10	p-substrate						
11	D	1	1	1	4.5	3	40
12	D	1	1	1	5	1.7	40
13	D	1	1	1	5	2.5	40
14	D	1	0.5	0.5	5	2.5	40
15	D	1	0	0.5	5	2.5	40
16	D	1	1	1	5	3.3	40
17	D	1	1	1	6	3	40
18	D	1	1	1	6	4	40
19	Common B						
20	Common G						

Some of the measurements on these frames are presented in the following figures. Figure 4.42 shows that the layout variations on y (from table 4.9) result in a constant breakdown voltage that is above 85 V for the optimal RESURF dose. The $R_{on,sp}$ remains constant between $y = -0.2\,\mu\text{m}$ and $y = 0.6\,\mu\text{m}$, although the pitch increases with 10 %. The same effect as previously explained for the pLDEMOS with the pwell as pdrift layer is acting here. The layout parameter y can be taken smaller, but too small a value results in higher $R_{on,sp}$, since a

bottleneck through which the current has to flow is created.

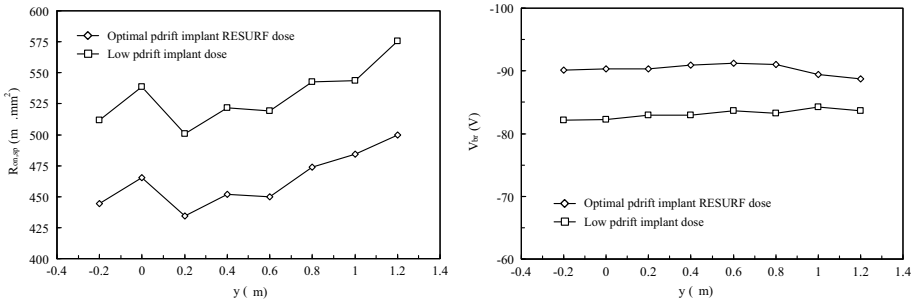


Figure 4.42 *Measured* V_{br} and $R_{on,sp}$ as a function of y for two pdrift implant doses.

The influence of the layout parameter *nwfi* is shown in figure 4.43, where it can be seen that $R_{on,sp}$ is decreasing with decreasing *nwfi*, while V_{br} remains constant (apart from some bad measurements). Since y remains constant during these variations, there is no "bottleneck effect" here. Positive values of *nwfi* not only increase the total device's length, they also enlarge the channel length of the device with n-epi. Negative values decrease the device's total pitch, but do not decrease the channel length any further as the nwell has a doping level that is approximately one order of magnitude larger than the pdrift's doping level. When the overlap is too large, other device's characteristics (like the V_t) can become affected. A value of $nwfi = 0\,\mu\text{m}$ is a good choice.

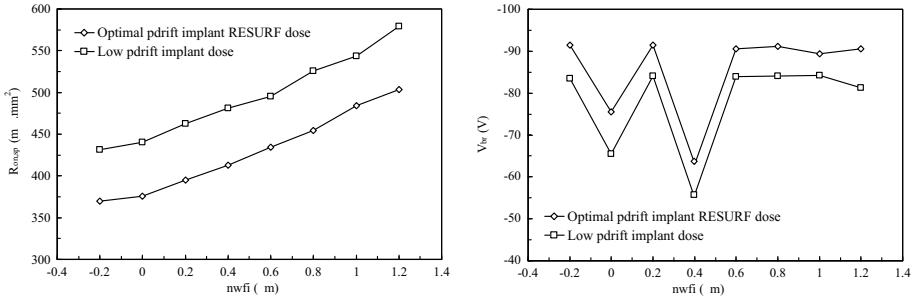


Figure 4.43 *Measured* V_{br} and $R_{on,sp}$ as a function of *nwfi* for two pdrift implant doses.

Eventually one device has been picked out to serve as the 80 V pLDEMOS in the I3T80 technology, its trade-off is compared with some competitor's devices at the end of this chapter. The famous plot of these values —with the ideal silicon limit for pDMOS devices— is given in the conclusions of this book (chapter 6).

Safe operating area

The safe operating area has not been discussed yet and the question arises whether or not it is a critical parameter for the RESURF pLDEMOS, as it was in the RESURF nLDEMOS. The measured $I_d(V_{ds})$ characteristics for a pLDEMOS are plotted in figure 4.44. The potential difference between gate and source has to be stressed beyond the maximum working value in real applications (i.e., $V_{gs} = 3.3\,\mathrm{V}$) to observe the Kirk effect. Unlike the RESURF nLDMOS devices, the RESURF pLDMOS device does not suffer from reduced on-state breakdown.

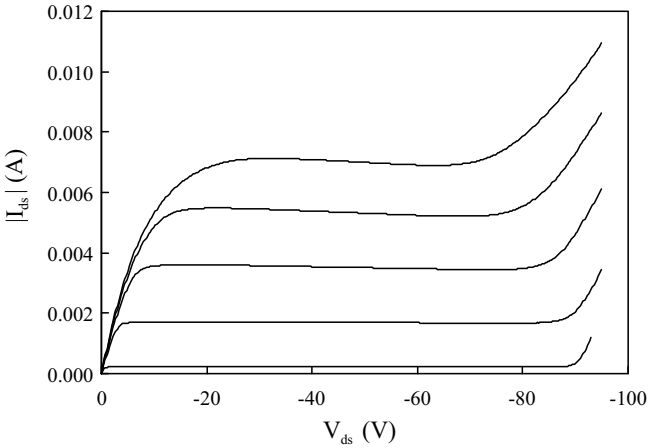


Figure 4.44 *Measured* $I_d(V_{ds})$ characteristics at $V_{gs} = -1.1, -2.2 \ldots -5.5\,\mathrm{V}$ of a pLDEMOS (width $W = 40\,\mu\mathrm{m}$) under optimal RESURF conditions.

The Kirk effect *can* occur in a RESURF pLDEMOS, as is illustrated in figure 4.45, where the optimal RESURF conditions are not fulfilled for the device under study. This is nice example of a so-called "underresurfed" device, in which the pdrift's dose is higher than the optimal RESURF dose. As a consequence, off-state breakdown is approximately $30\,\mathrm{V}$ and remains this low at low current densities ($V_{gs} = -1.5\,\mathrm{V}$). But, at higher current densities (e.g., $V_{gs} = -3.0\,\mathrm{V}$), the Kirk effect is acting, and the blocking capability increases ! The corresponding shift of the electric field peaks is shown in figure 4.46. Note, however, that this example is atypical and that in devices with an optimized RESURF dose, the Kirk effect always causes a drop in the blocking voltage at high current densities.

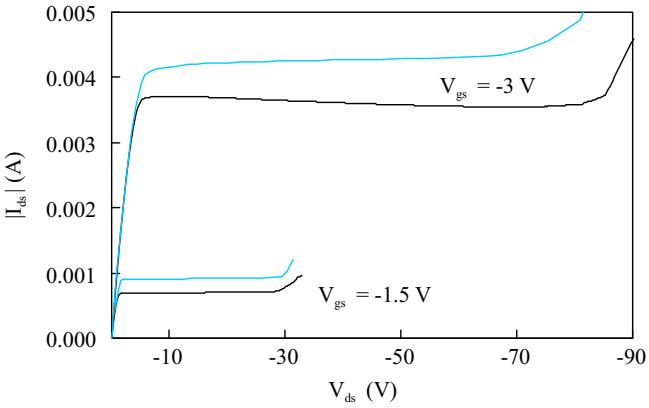


Figure 4.45 $I_d(V_{ds})$ characteristics at $V_{gs} = -1.5$ and $-3.0\,\mathrm{V}$ of an "under resurfed" device (width $W = 40\,\mu\mathrm{m}$): *measured* (black) versus simulated (blue).

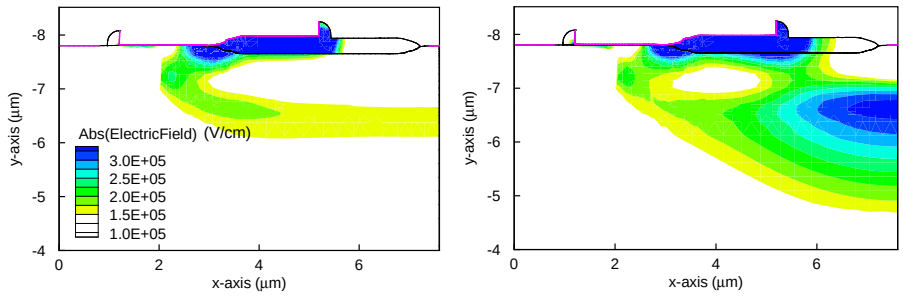


Figure 4.46 Electrical field at breakdown in an "under resurfed" structure at $V_{gs} = -1.5\,\mathrm{V}$ and $V_{gs} = -3.0\,\mathrm{V}$.

Degradation

The gate current is a direct measure for the carriers that are injected in the gate oxide and thus for the damage caused to the gate oxide. Therefore the maximum gate current at a real life bias ($V_{gs} = -1.5\,\mathrm{V}$ and $V_{ds} = -70\,\mathrm{V}$) was chosen as stress condition. It has been observed that R_{on} is the most degrading parameter, and is therefore chosen as monitor for degradation (figure 4.47). Figure 4.47 shows that the degradation of R_{on} of the pLDEMOS obeys the conventional power law [TS83]: $\Delta R_{on}/R_{on,t=0} = At^n$ with $A = 0.44$ and $n = 0.15$. Although the stress measurements were carried out up to $1.6e5\,\mathrm{s}$, no saturation was observed. The very low value for n guarantees a full lifetime of 25 years (whether or not saturation would occur at longer stress times) at the most severe stress condition.

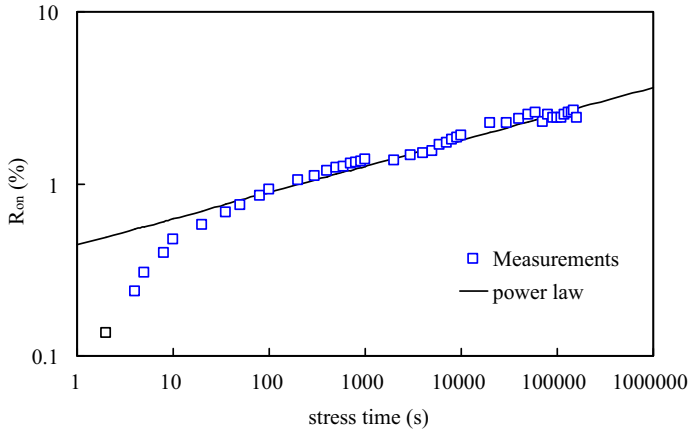


Figure 4.47 Degradation of the pLDEMOS: measurements (squares) and fitted power law.

4.11 Conclusions

It has been shown that for an 80 V junction isolated technology, the lateral RESURF nDMOS beats the vertical nDMOS. That is, if only the $V_{br} - R_{on,sp}$ trade-off is considered. The SOA of the vertical DMOS is larger than the lateral RESURF device's SOA, as the latter one suffers from the Kirk effect which seriously diminishes the blocking capability from off to on-state. "Adaptive RESURF" techniques in which the doping level towards the drain is gradually increased, ameliorate the SOA of the RESURF device, but at the costly expense of one extra mask. Furthermore, the vertical device is inherently floating, which makes it possible to use this device as a high-side switch. A solution with a double buried layer renders the lateral RESURF device floating as well, but this is technically more difficult to realize (e.g., epi growth in several steps) and the problem of the small SOA remains.

Moreover, the best vertical nDMOS is easy to combine with the best pDMOS, being a lateral RESURF structure. Due to the lower current levels of the p-type compared to the n-type device, the lateral RESURF pLDMOS does not suffer from the Kirk effect and thus has a large SOA.

In tables 4.13 and 4.14, the commercially available nVD(E)MOS and pLDEMOS devices of AMIS' I3T80 technology are compared with different technologies from the most important competitors. The nVDEMOS as treated in this chapter can be improved by the introduction of the pbody, thereby creating a DMOS device (i.e., double diffused MOS). The development of these nVD(E)MOS and nLD(E)MOS devices has

been done at AMIS. As can be seen from table 4.13, the choice for the VDMOS is exceptional, but is nevertheless justified by the harsh requirements (ESD, energy capability...) imposed by the automotive applications this technology aims for. An overview of these trade-offs is

Table 4.13 Benchmarking of the nLD(E)MOS and nVD(E)MOS.

Competitor	Technology	Device Type	V_{br} (V)	$R_{on,sp}$ ($m\Omega.mm^2$)	Reference
ADI	0.6 µm	LDMOS	22	23	[WBD+00]
ADI	0.6 µm	LDMOS	25	28	[WBD+00]
ADI	0.6 µm	LDMOS	36	43	[WBD+00]
Mitsubishi	0.5 µm	LDMOS	33	60	[TYH00]
Mitsubishi	0.5 µm	LDMOS	74	127	[TYH00]
Mitsubishi	0.5 µm	LDMOS	94	230	[TYH00]
Mitsubishi	0.35 µm	Trench Gate	32	18	[NMK+00]
Motorola	0.8 µm	LDMOS	60	70	[MBC+99]
Motorola	0.8 µm	LDMOS	70	110	[MBC+99]
Motorola	0.8 µm	LDMOS	80	130	[MBC+99]
Motorola	0.35 µm	LDMOS	65	56	[ZPB+00]
Motorola	0.35 µm	LDMOS[a]	62	53	[KPB03]
Motorola	0.25 µm	LDEMOS	47	36	[dFSM+02]
Philips	0.8 µm	LDMOS	50	110	[NLBN97]
Philips	SOI, 1.2 µm	LDMOS	29	43	[vdPLH+00]
Philips	SOI, 1.2 µm	LDMOS	79	135	[vdPLH+00]
Philips	SOI, 1.2 µm	LDMOS	86	170	[vdPLH+00]
ST	BCD6, 0.35 µm	LDMOS	27	20	[MMC+00]
ST	BCD6, 0.35 µm	LDMOS	38	45	[MMC+00]
ST	BCD6, 0.35 µm	LDMOS	48	62	[MMC+00]
ST	BCD6, 0.35 µm	LDMOS	70	92	[MMC+00]
TI	SOI	LDMOS	55	90	[MEH+02]
TI	SOI	LDMOS	68	100	[MEH+02]
TI	SOI	LDMOS	80	120	[MEH+02]
Toshiba	0.35 µm	Trench Gate	25	13	[NK00]
Toshiba	0.35 µm	LDMOS	49	54	[KWK+03]
Toshiba	0.35 µm	LDMOS	59	70	[KWK+03]
AMIS	I3T80, 0.35 µm	VDMOS	89	165	[MBD+02a]
AMIS	I3T80, 0.35 µm	VDMOS	84	155	[MBD+02a]
AMIS	I3T80, 0.35 µm	LDMOS	84	183	[MBD+02a]
AMIS	I3T80, 0.35 µm	LDMOS	65	100	[MBD+02a]
AMIS	I3T80, 0.35 µm	LDEMOS	17	36	[MBD+02a]

[a]FRESURF device with floating capabilities

Table 4.14 Benchmarking of the pLDEMOS.

Competitor	Technology	V_{br} (V)	$R_{on,sp}$ (mΩ.mm^2)	Reference
ADI	0.6 µm	−28	120	[WBD+00]
Motorola	0.35 µm	−40	180	[PZK+00]
Motorola	0.35 µm	−56	240	[PZK+00]
Motorola	0.25 µm	−47	94	[dFSM+02]
Philips	0.8 µm	−50	290	[NLBN97]
Philips	SOI, 1.2 µm	−73	460	[vdPLH+00]
ST	0.6 µm	−48	150	[CGP+98]
Toshiba	0.7 µm	−34	111	[KNK+01]
Toshiba	0.35 µm	−46	131	[KWK+03]
AMIS	I2T, 0.7 µm	−75	553	[Ver01]
AMIS	I2T, 0.7 µm	−50	250	[Ver01]
AMIS	I3T80, 0.35 µm	−90	300	[BVD+02]
AMIS	I3T80, 0.35 µm	−61	250	[MBD+02b]

given in a $V_{br}(R_{on,sp})$ plot, where all devices can be compared to each other and to the ideal silicon limits at a single glance (see chapter 6, figures 6.1 and 6.2).

References

[AV79] J.A. Appels and H.M.J. Vaes. High Voltage Thin Layer Devices (RESURF devices). In *Electron Devices Meeting*, pages 238–241, 1979.

[Bal96] B.J. Baliga. *Power Semiconductor Devices*. PWS, 1996.

[BVD+02] B. Bakeroot, M. Vermandel, J. Doutreloigne, P. Moens, and D. Bolognesi. Cost Effective Implementation of a 90 V RESURF p-Type Drain Extended MOS in a 0.35 µm Based Smart Power Technology. In *European Solid-State Device Research Conference*, pages 291–294, 2002.

[CGP+98] C. Contiero, P. Galbiati, M. Palmieri, G. Ricotti, and R. Stella. Smart Power Approaches VLSI complexity. In *Symposium on Power Semiconductor Devices & ICs*, pages 11–16, 1998.

[Chu00] S.-K. Chung. An Analytical Model for Breakdown Voltage of Surface Implanted SOI RESURF LDMOS. *IEEE Transactions on electron devices*, 47(5):1006–1009, May 2000.

[CSN00] G.J. Cao, M.M. De Souza, and E.M.S. Narayanan. Resurfed Lateral Bipolar Transistors for High-Voltage, High-Frequency Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 185–188, 2000.

[dFSM+02] E. de Fresart, R. De Souza, J. Morrison, P. Parris, and J. Heddleson an V. Venkatesan et al. Integration of Multi-Voltage Analog and Power Devices in a 0.25µm CMOS + Flash Memory Process. In *Symposium on Power Semiconductor Devices & ICs*, pages 305–308, 2002.

[Fuj97] T. Fujihira. Theory of Semiconductor Superjunction Devices. *Japanese Journal of Applied Physics*, 36(10):6254–6262, October 1997.

[Gha77] S.K. Ghandhi. *Semiconductor Power Devices*. John Wiley & Sons, 1977.

[HB91] Y.S. Huang and B.J. Baliga. Extension of RESURF Principle to Dielectrically Isolated Power Devices. In *Symposium on Power Semiconductor Devices & ICs*, pages 27–30, 1991.

[HIFT02] Z. Hossain, M. Imam, J. Fulton, and M. Tanaka. Double-Resurf 700 V N-Channel LDMOS with Best-in-Class On-Resistance. In *Symposium on Power Semiconductor Devices & ICs*, pages 137–140, 2002.

[HLMP00] P. Hower, J. Lin, S. Merchant, and S. Paiva. Using "Adaptive Resurf" to Improve the SOA of LDMOS Transistors. In *Symposium on Power Semiconductor Devices & ICs*, pages 345–348, 2000.

[KCA96] D. Križaj, G. Charitat, and S. Amon. A New Analytical Model for Determination of Breakdown Voltage of RESURF Structures. *Solid-State Electronics*, 39(9):1353–1358, 1996.

[KNK+01] Y. Kawaguchi, K. Nakamura, K. Karouji, K. Watanabe, Y. Yamaguchi, and A. Nakagawa. 0.6 µm BiCMOS Based

15 and 25 V LDMOS for Analog Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 169–172, 2001.

[KPB03] V. Khemka, V. Parthasarathy, and A. Bose. A Floating RESURF (FRESURF) LD-MOSFET Device Concept. *IEEE Electron Device Letters*, 24(10):664–666, October 2003.

[KPZB02] V. Khemka, V. Parthasarathy, R. Zhu, and A. Bose. Correlation Between Static and Dynamic SOA (Energy Capability) of RESURF LDMOS Devices in Smart Power Technologies. In *Symposium on Power Semiconductor Devices & ICs*, pages 125–128, 2002.

[KWK+03] T. Kubota, K. Watanabe, K. Karouji, M. Ueno, Y. Anai, and Y. Kawaguchi et al. Cost-Effective Approach in LDMOS with Partial 0.35 µm Design into Convetional 0.6 µm Process. In *Symposium on Power Semiconductor Devices & ICs*, pages 245–248, 2003.

[LDKM99] L. Lorenz, G. Deboy, A. Knapp, and M. März. COOLMOS^{TM} - A New Milestone in High Voltage Power MOS. In *Symposium on Power Semiconductor Devices & ICs*, pages 3–10, 1999.

[Lud00a] A.W. Ludikhuize. A Review of RESURF Technology. In *Symposium on Power Semiconductor Devices & ICs*, pages 11–18, 2000.

[Lud00b] A.W. Ludikhuize. Self-aligned and Shielded-Resurf LDMOS for dense 20V Power IC's. In *Symposium on Power Semiconductor Devices & ICs*, pages 81–84, 2000.

[MAB+91] S. Merchant, E. Arnold, H. Baumgart, S. Mukherjee, H. Pein, and R. Pinker. Realization of High Breakdown Voltage (> 700 V) in thin SOI Devices. In *Symposium on Power Semiconductor Devices & ICs*, pages 31–35, 1991.

[MBC+99] S. Merchant, R. Baird, S. Chang, P. Hui, V. Macary, and M.G. Neaves. High-Performance 13 – 65 V Rated LDMOS Transistors in an Advanced Smart Power Technology. In *Symposium on Power Semiconductor Devices & ICs*, pages 225–228, 1999.

[MBD+02a] P. Moens, D. Bolognesi, L. Delobel, D. Villanueva, H. Hakim, and S.C. Trinh et al. I3T80: A 0.35 µm Based System-on-Chip Technology for 42 V Battery Automotive Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 225–228, 2002.

[MBD+02b] P. Moens, D. Bolognesi, L. Delobel, D. Villanueva, K. Reynders, A. Lowe, G. Van Herzeele, M. Tack, and B. Bakeroot. Future Trends in Intelligent Interface Technologies for 42 V Battery Automotive Applications. In *European Solid-State Device Research Conference*, pages 287–290, 2002.

[MEH+02] S. Merchant, T. Efland, S. Haynie, W.Headen, K. Kajiyama, and S. Paiva et al. Robust 80 V LDMOS and 100 V DECMOS in a Streamlined SOI Technology for Analog Power Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 185–188, 2002.

[MMC+00] A. Moscatelli, A. Merlini, G. Croce, P. Galbiati, and C. Contiero. LDMOS Implementation in a 0.35 µm BCD Technology (BCD6). In *Symposium on Power Semiconductor Devices & ICs*, pages 323–326, 2000.

[NK00] A. Nakagawa and Y. Kawaguchi. Improved 20 V Lateral Trench Gate Power MOSFETs with Very Low On-Resistance of 7.8 mΩ.mm^2. In *Symposium on Power Semiconductor Devices & ICs*, pages 47–50, 2000.

[NLBN97] A. Nezar, A.W. Ludikhuize, R. Brock, and N. Nowlin. A Submicron Bi-CMOS-DMOS Process for 20-30 and 50V Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 333–336, 1997.

[NMK+00] A. Narazaki, J. Maruyama, T. Kayumi, H. Hamachi, J. Moritani, and S. Hine. A 0.35 µm Trench Gate MOSFET with an Ultra Low On-State Resistance and a High Destruction Immunity During the Inductive Switching. In *Symposium on Power Semiconductor Devices & ICs*, pages 377–380, 2000.

[PHD+02] S. Pendharkar, R. Higgins, T. Debolske, T. Efland, and B. Nehrer. Optimization of Low Voltage N-Channel LD-

MOS Devices to Achieve Required Electrical and Lifetime SOA. In *Symposium on Power Semiconductor Devices & ICs*, 2002.

[PZK+00] V. Parthasarathy, R. Zhu, V. Khemka, M.L. Ger, T. Bettinger, S. Chang, P. Hui, and A. Bose. Integration of High-Voltage Bipolars into a 0.35 µm CMOS based smart power platform. In *Bipolar/BiCMOS Circuits and Technology Meeting*, pages 32–35, 2000.

[TS83] E. Takeda and N. Suzuki. An Empirical Model for Device Degradation Due to Hot-Carrier Injection. *IEEE Electron Device Letters*, EDL-4(4):111–113, April 1983.

[TYH00] T. Terashima, F. Yamamoto, and K. Hatasako. Multi-Voltage Device Integration Technique for 0.5 µm BiCMOS & DMOS Process. In *Symposium on Power Semiconductor Devices & ICs*, pages 331–334, 2000.

[Udr02] F. Udrea. Advanced 3D RESURF Devices for Power Integrated Circuits. In *Semiconductor Conference*, pages 229–238, 2002.

[vdPLH+00] J.A. van der Pol, A.W. Ludikhuize, H.G.A. Huizing, B. van Velzen, R.J.E. Hueting, and J.F. Mom et al. A-BCD: An Economic 100V RESURF Silicon-On-Insulator BCD Technology for Consumer and Automotive Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 327–330, 2000.

[Ver01] M. Vermandel. *Integration of p-type High-Voltage Devices in a Sub-µm CMOS Technology using TCAD*. PhD thesis, University of Gent, 2001.

[WBD+00] S. Whiston, D. Bain, A. Deignan, J. Pollard, C. Ni Chleirigh, and C. Musgrave et al. Complementary LDMOS Transistors for a CMOS/BICMOS Process. In *Symposium on Power Semiconductor Devices & ICs*, pages 51–54, 2000.

[Zin01] R.P. Zingg. New Benchmark for RESURF, SOI and Super-Junction Power Devices. In *Symposium on Power Semiconductor Devices & ICs*, pages 343–346, 2001.

[ZPB+00] R. Zhu, V. Parthasarathy, A. Bose, R. Baird, V. Khemka, and T. Roggenbauer et al. A 65V, $0.56 m\Omega.cm^2$ Resurf LDMOS in a 0.35 µm CMOS Process. In *Symposium on Power Semiconductor Devices & ICs*, pages 335–338, 2000.

5 The IGBT

5.1 Introduction

In december 1982, two independent groups proposed a device capable of handling large currents in the on-state, with low on-state voltage drop (low power loss), while keeping high input impedance and a voltage controlled gate. They came to this synthesis while in search of a device combining the best features of two existing power transistors. On the one hand, the bipolar power transistor, capable of handling large currents in the on-state with low forward voltage drop. The major drawbacks being the current controlled gate (complex gate control circuit requires a lot of space), the relatively small current gain due to the large base width needed for these power devices and the slow switching speed. On the other hand, the power MOSFET, which is a voltage controlled device (simple gate control circuit), is fast and has a large safe operating area. Unfortunately, this device has an increasing on-resistance when its breakdown voltage is increased. Therefore, this device is not suitable for applications with high DC supply voltages (over 200 V).

One way to combine the best of both devices is in a Darlington configuration where the DMOS device drives the power bipolar transistor. This circuit solution gives good results but an even more elegant way out was to look for one single device combining the physics of both the bipolar and the MOS power transistors. The first breakthrough came with the study of the thyristors. These devices are unmatched when it comes to current-carrying capability per unit area because of the conductivity modulation of lightly doped regions due to minority injection. But then again, these devices are current controlled, induce substrate currents and are difficult to switch off. In search for an improved version of these devices, different researchers ([Bal79], [Tih80] and [PS80]) proposed a planar thyristor with an insulated gate. This device had improved isolation, had a voltage controlled gate, but still behaved like a thyristor once it was triggered, requiring current interruption to shut it off. The final step towards the currently called IGBT (*I*nsulated *G*ate *B*ipolar *T*ranstistor) was taken simultaneously by the two groups men-

tioned above when they realized a device that remains gate controlled over a wide range of anode current and voltage. The first group called it the COMFET (COnductivity Modulated FET [RGGN83]), the second group called it the IGR (Insulated Gate Rectifier [BAG$^+$82]). These devices look exactly the same as the insulated gate thyristor, but are different in a fundamental way: it is designed *not* to go into regenerative action, or, in other words, not to latch like a conventional thyristor; thereby avoiding the positive feedback and the loss of the voltage controlled gate. Inevitably, the device inherited some of the properties of the thyristor, it is prone to latch and it is difficult to switch off. To date, these topics remain research topics.

Since its invention in the 80s, the IGBT is used in a growing number of applications, especially in the higher voltage ranges (up to several thousands of volts). Thus, the main occurrence of the IGBT is as a discrete device, and rarely as an integrated device. The reason being twofold: first of all, devices that are integrated in a technology based on a standard CMOS platform have inherently limited voltage and current ranges. In these limited voltage ranges, they compete with power DMOS devices, which have superior characteristics in the low voltage range ($< 200\,\mathrm{V}$). Secondly, lateral IGBTs are difficult to integrate as they produce large substrate currents. A problem that is normally solved with the introduction of huge isolation structures, which kills the eventual advantage the device had over a DMOS. This is also the reason why the LIGBT shows up in SOI technologies, where problems with substrate currents are avoided due to the dielectric isolation of the devices.

Yet the possibility of integrating an IGBT in the I3T80H technology of AMIS is investigated. First of all, a straightforward adaptation of an DMOS device is carried out. The n^+ drain region is simply replaced by a p^+ anode region. Using this structure, the next section discusses the device's operation. Albeit not an optimized example, it gives an overview of the basic operation of the lateral IGBT and the problems encountered. In the following sections, we gradually introduce more layers so as to improve the device's performance.

Using the existing, calibrated TCAD input decks, several ideas are tried out and the most appealing ones have been put on testchip. A lot of effort went into the suppression of the substrate current, although a lot of other issues need to be studied as well: breakdown, on-state forward voltage drop, SOA, turn-off, latch-up.... At the end of the chapter, a comparison is made between the different nLIGBT devices and a nVDEMOS based on measurements.

5.2 Breakdown in a nLIGBT without Buried Layers

Figure 5.1 shows a simple realization of an LIGBT in the I3T80H technology. There is one major difference with the LDMOS transistor: the drain side of the nLIGBT (hereafter called the anode) is formed with an p^+ region instead of a n^+. This seemingly small adaptation makes a world of difference, since now the device is only conducting current when the p^+/n-epi junction is forward biased. Therefore this is a truly bipolar device with the *floating* n-epi now serving as the base of the inherent lateral pnp (pwell/n-epi/p^+) transistor. First of all, breakdown is discussed in this section. The forward conduction state will be treated in the next section.

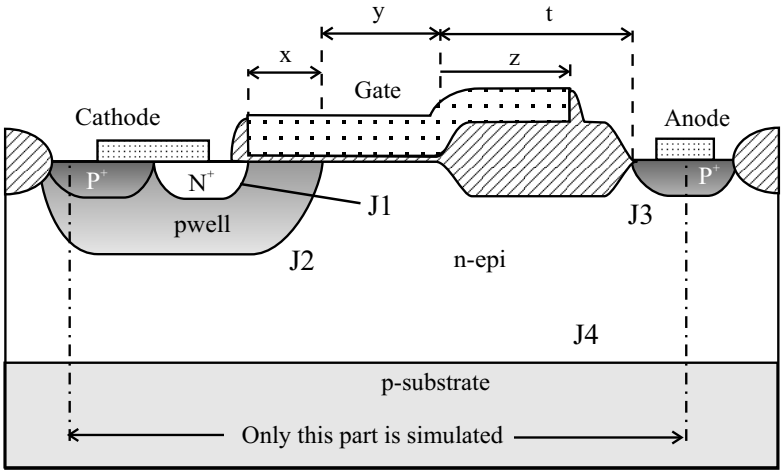


Figure 5.1 The LIGBT device and its most important layout parameters.

When the anode is positively biased (cathode, gate and substrate at ground), the forward blocking capability is provided by both the pwell/n-epi junction (J2), and the n-epi/p-substrate junction (J4), which renders this a RESURF type of device. But, this device is susceptible to a few more breakdown phenomena than the DMOS. First of all, there is the possible reach-through of the junction J2 to the cathode/pwell junction (J1). This is also possible in the DMOS and is avoided by designing a pwell that is wide enough (layout parameter x), so as to keep the depletion layer coming from junction J2 away from junction J1. Furthermore, there is the same threat of reach-through at the anode side of the device. The depletion layer extending from junction J2 must not

reach junction J3 *and*, at the same time, the depletion layer extending from junction J4 must not reach this same junction J3.

This has been simulated with an arbitrarily chosen LIGBT device (actually an exact copy of an existing DMOS transistor, where only the drain has been changed: $x = 1$, $y = 0.3$, $t = 3.6$ and $z = 2.4\,\mu\text{m}$).

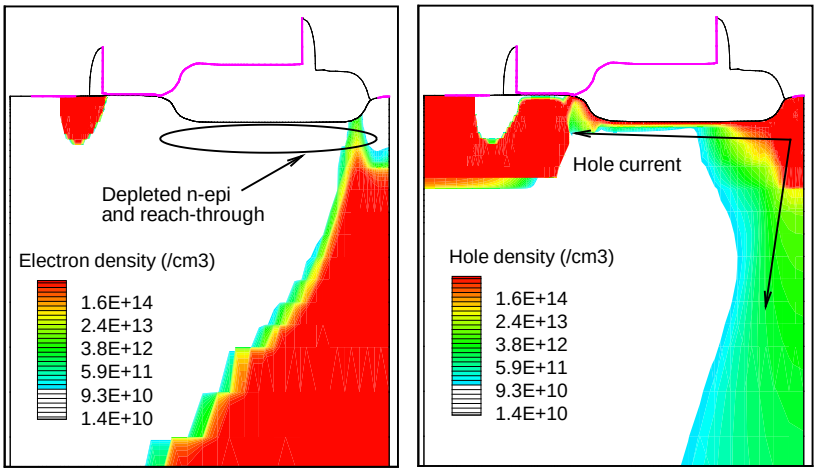


Figure 5.2 Electron and hole densities at breakdown (forward, i.e., $V_{ak} > 0\,\text{V}$) for an LIGBT with $x = 1$, $y = 0.3$, $t = 3.6$ and $z = 2.4\,\mu\text{m}$.

Figure 5.2 clearly shows that with a small n-epi width ($\approx y + t$), reach-through occurs between the junctions J2 and J3. The depletion layer coming from the junction J2 extends mostly in the n-epi (because the n-epi is approximately 10 times lower doped than the pwell). Reach-through does not occur between junctions J4 and J3 because the depletion layer extends into the p-substrate, due to its low doping level (note the substrate current in figure 5.3). Another proof that the breakdown is initiated by a pure reach-through effect and not by impact ionization is given in figure 5.3. With or without the impact ionization model, the breakdown of the device remains under 20 V.

With the base of the pnp transistor taken larger, this pure reach-through effect is killed and now breakdown is an interplay between impact ionization and bipolar working.Without the impact ionization model (figure 5.4) breakdown clearly does not take place at the same voltage as with the avalanche current taken into account. This current (mainly starting in the vicinity of the bird's beak under the gate) forward biases the anode p^+/n-epi junction and results in a hole current flowing to the substrate (figure 5.4). The holes generated at the bird's

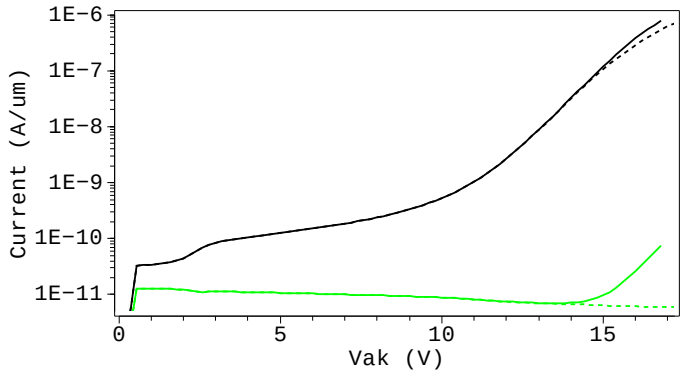


Figure 5.3 Current through the anode (black) and substrate (green) contacts of a small LIGBT (figure 5.2) before and at breakdown. Solid lines are simulated with impact ionization, dashed ones without.

beak flow to the cathode contact, resulting in high current levels at all three contacts at breakdown.

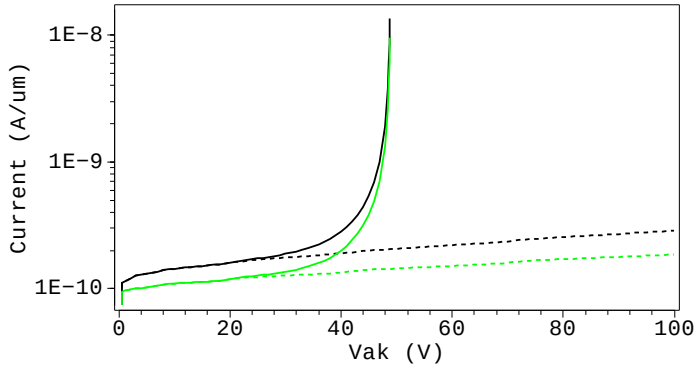


Figure 5.4 Current through the anode (black) and substrate (green) contacts of a large LIGBT before and at breakdown. Solid lines are simulated with impact ionization, dashed ones without.

It is obvious that the substrate needs to be protected by means of a buried layer. Several possible solutions are worked out in the following sections.

5.3 High-Side nLIGBTs with BLN

The use of the BLN in the LIGBT (figure 5.5) is first treated. A device with a BLP and a psinker contact shorted to the cathode (the 'standard' approach) will be treated later. The BLN serves as a buffer between the anode and substrate, thereby killing the parasitic bipolar between those two contacts. Introducing the BLN also kills the RESURF effect because the depleted region between BLN and p-substrate mainly extends in the p-substrate and is very thin in the BLN.

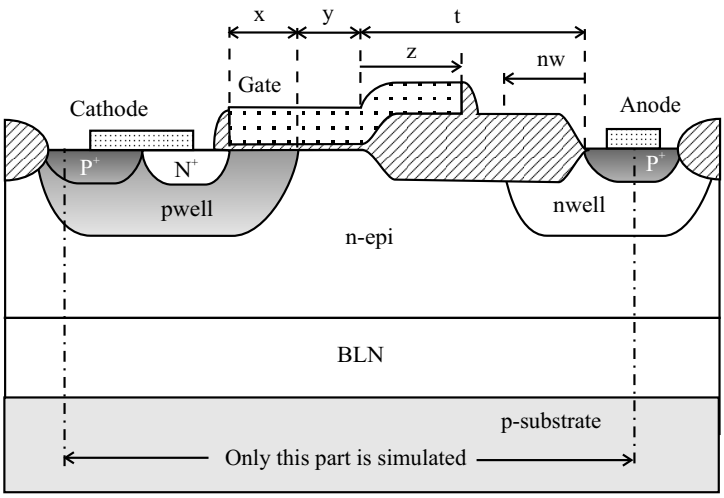


Figure 5.5 The LIGBT with standard I3T80 layers.

Figure 5.5 also depicts the nwell, used as a buffer in the pwell/n-epi/p^+ bipolar. As mentioned before, this bipolar needs to work as good as possible. However, the buffer is unavoidable, as will be shown.

First of all, a device with BLN and without nwell was simulated. The breakdown voltage was only 48 V, although the field oxide length was 10 µm. This is comprehensible, as this is the BV_{CEO} of a lateral bipolar with the n-epi as base. Increasing the base width would increase the blocking capability, but the pitch of the device is already large, so other solutions are tried out. This is done by introducing a buffer around the emitter (i.e., the anode), thereby trying to avoid that the electrons coming from the impact ionization zone (i.e., around the bird's beak under the poly) forward bias the n-epi/p^+ junction too soon.

This was simulated using a device with the same pitch as the previous one, and with two nw values: $nw = 1.5\,\mu m$ and $nw = 3\,\mu m$. The

breakdown voltages were 51 V and 53 V, respectively. This means that the nwell is not highly enough doped to prevent early breakdown due to a forward biased p^+/n-epi junction. Thus a heavier buffer is necessary if the breakdown voltage needs to be higher.

A heavier buffer could also be realized by another standard I3T80 layer, the nsinker. One simulation has been carried out using the nsinker around the anode contact. The breakdown voltage of the device was 68 V. However, due to the large lateral out-diffusion of the nsinker, devices using the nsinker have a large pitch and have a bad pwell–n-epi+nsinker–p^+ bipolar, which deteriorates the device's performance. The use of a dedicated nbuffer is investigated in more detail, as these devices have a superior performance.

5.3.1 A nLIGBT with a Dedicated nbuffer

Breakdown

In order to keep the out-diffusion under control, the nbuffer is introduced at the same place in the process flow as the nwell. The layout parameter nb describing the nbuffer is defined in the same way as layout parameter nw in figure 5.5. Table 5.1 gives an overview of simulation results using a device with the following layout parameters: $x = 1$, $y = 0.8$, $t = 4.5$, $z = 1.5$ and $nb = 1.5\,\mu\text{m}$. Notice the reduction of the field oxide length in comparison with the devices in the previous section, as well as the high breakdown voltages.

Table 5.1 Breakdown voltage (V_{br}) and forward drop voltage (V_{fwd}) for several nbuffer implantation doses at 500 keV.

Dose (cm^{-2})	V_{br} (V)	V_{fwd} (V)[a]
2.5e14	64	1.5[b]
5e14	66	1.6
1e15	66	1.8
2e15	68	2.2
4e15	70	9.6

[a] at a current level of $I_a = 0.0004\,\text{A/ m}$ for $V_{gk} = 3.3\,\text{V}$
[b] only reached 0.00038 A/ m at latch-up

Table 5.1 also mentions another electrical parameter: the forward voltage drop at an arbitrarily chosen current level ($I_a = 0.0004\,\text{A}/\mu\text{m}$)

at $V_{gk} = 3.3\,\mathrm{V}$. It is important to know at what percentage of the latch-up current this current level is chosen. The maximum feasible current level is not only determined by this static latch-up, but also by the dynamic latch-up, which is a more severe condition (see further). For the time being, the current level was chosen very close to the latch-up current ($\sim 90\,\%$). More details about the forward conduction state are given hereunder.

Forward conduction state

Figure 5.6 shows the output characteristics of the LIGBT. In order to have a good understanding of the static behaviour of this device, several phenomena leading to this forward conduction behaviour will be discussed in more detail.

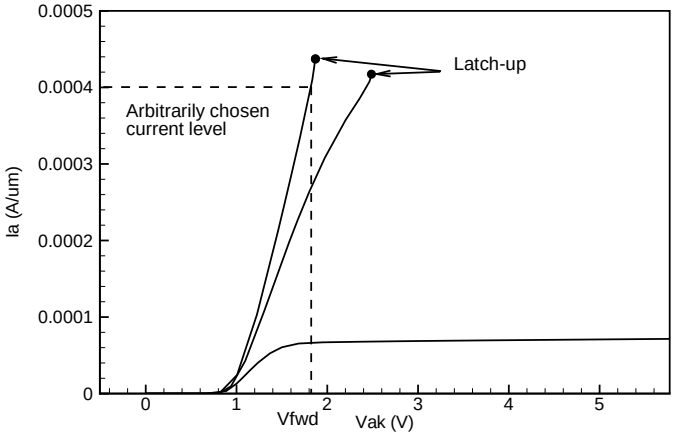


Figure 5.6 Output characteristics at $V_{gk} = 1.1$, 2.2 and 3.3 V of an LIGBT with a nbuffer dose of $1e15\,\mathrm{cm}^{-2}$ of table 5.1.

The first difference when comparing the characteristics of figure 5.6 to those of a normal DMOS device, is that the current only starts to flow if a high enough potential is applied to the anode. This is the bipolar working in the LIGBT, where the nbuffer/p^+ junction becomes forward biased if V_{ak} increases (see figure 5.7 at $V_{ak} = 0.5\,\mathrm{V}$). With increasing anode bias, at the same moment electrons are entering the n-epi via the MOS channel, holes are injected into the n-epi through the forward biased nbuffer/p^+ junction.

In the forward conduction state, the electrons and holes are abundant in the drift region. To such an extent, that their concentrations become greater than the background doping level—and actually equal to one

another as to preserve charge neutrality, see figure 5.7 at $V_{ak} = 1.5\,\mathrm{V}$—giving rise to conductivity modulation, where a low potential drop is observed in the drift region which is approximately independent of the current through it.

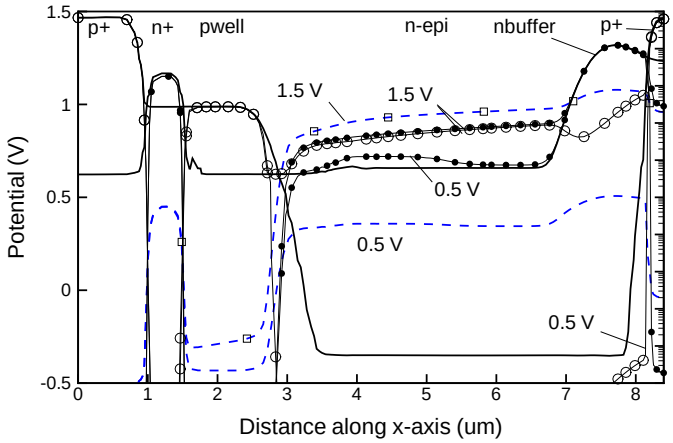


Figure 5.7 Electron density (•), hole density (○), donor and acceptor concentrations (solid lines) are plotted against the right y-axis (log scale), the potential (dashed lines) against the left y-axis for the LIGBT of figure 5.6 at $V_{gk} = 1.1\,\mathrm{V}$ with $V_{ak} = 0.5\,\mathrm{V}$ and $V_{ak} = 1.5\,\mathrm{V}$. The cutline runs parallel with the silicon surface, along the entire device length, just underneath the field oxide.

Another question that might arise is what happens at saturation (that is, when a constant current level is reached). The potential at the end of the channel is plotted for several anode to cathode (V_{ak}) values in figure 5.8, revealing that for this device the MOSFET is causing the saturation at $V_{gk} = 1.1\,\mathrm{V}$. When V_{ak} increases, the surface potential at the end of the channel increases until pinch-off is reached.

Latch-up

In the case of $V_{gk} = 2.2$ and $3.3\,\mathrm{V}$, latch-up occurs before saturation. This means that the current level is too high for the device to deal with. The holes that are injected into the base (n-epi) by the emitter (p^+ anode) are collected by the cathode contact. The path they follow goes through the pwell, thereby creating a potential difference along it. If this hole current is large enough, the potential difference becomes large enough to forward bias the n^+/pwell junction. This results in an injection of electrons into the pwell, thus the turn-on of the parasitic

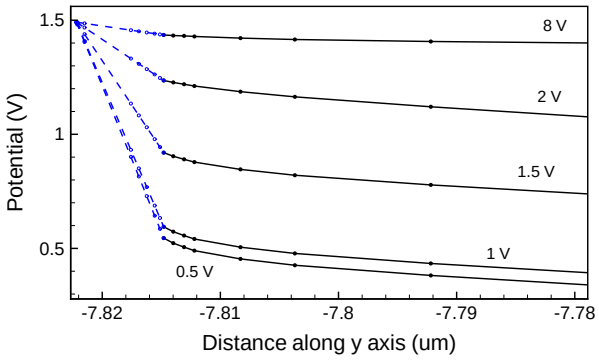


Figure 5.8 Potential at the end of the channel (y-axis is perpendicular to the silicon surface) for $V_{gk} = 1.1\,\mathrm{V}$ and $V_{ak} = 0.5, 1, 1.5\ldots\,\mathrm{V}$. The dashed lines show the potential in the gate oxide, the solid ones the potential in the silicon.

npn bipolar and the loss of gate control. The device is in thyristor mode, where two bipolars are working in a positive feedback loop. The device has been simulated with a load of $10\,\mathrm{k}\Omega$ to illustrate this thyristor behaviour (figure 5.9). Note that all simulations are carried out without thermal effects.

The question arises if this latch-up can be avoided or at least be postponed. The latch-up effect observed here has the same origin as in a thyristor structure with shunted emitters and an approximation for this latch-up current is [BGG99, p. 323]

$$I_L \approx \frac{g}{\alpha_{pnp} l_k \rho_p} \tag{5.1}$$

where l_k is the length of the n^+ cathode region, g the thickness of the p layer (here the pwell), ρ_p the resistivity of that layer, and α_{pnp} the common-base current gain. This gives us immediately some clues as to how to increase the latch-up current:

- increasing the doping level of the p layer (e.g., by adding the existing pdrift)
- increasing the depth of the p layer (e.g., using the pdrift together with the psinker !?)
- decreasing the cathode contact length (limited by design rules)
- reducing the gain of the pnp (pwell/n-epi/p^+) transistor

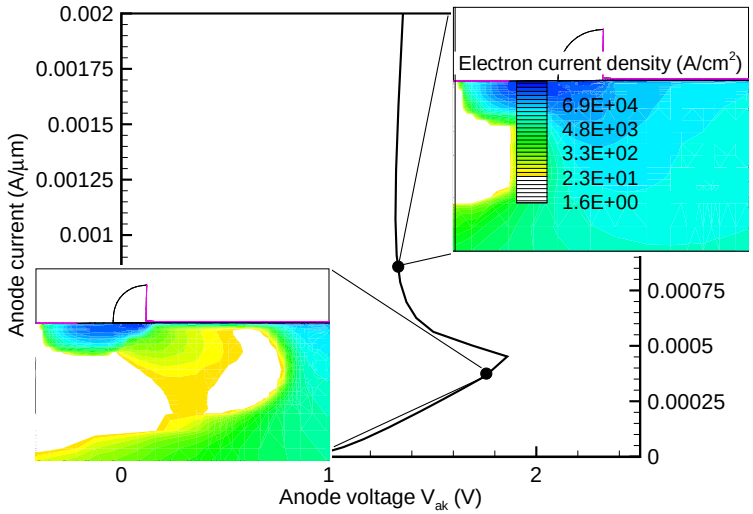


Figure 5.9 Snapback at $V_{gk} = 3.3\,\mathrm{V}$ for the device which output characteristics are shown in figure 5.6. The lower inset shows the electron current density at the onset of latch-up, the upper inset the electron current density in thyristor mode.

The first suggestion (adding the pdrift) is straightforward and will be illustrated at the end of this section. The second suggestion (adding the psinker) seriously changes the device's performance and will be studied in a next subsection. The third suggestion (decreasing the cathode length) has been done to the limit of the design rules. The forth suggestion is not the ideal one as this affects the quality of the device as the hole current is essential for a good operation of the nLIGBT. The gain of the pnp is approximated using device simulations to demonstrate this. The *DC common-base current gain* α_{pnp} is defined as [Sze81, p. 140]:

$$\alpha_{pnp} = \frac{\partial I_C}{\partial I_E} = \frac{\partial I_{pE}}{\partial I_E}\frac{\partial I_{pC}}{\partial I_{pE}}\frac{\partial I_C}{\partial I_{pC}} \tag{5.2}$$

with the *emitter injection efficiency* $\gamma = \partial I_{pE}/\partial I_E$, the *base transport factor* $\alpha_T = \partial I_{pC}/\partial I_{pE}$ and the collector multiplication factor $M = \partial I_C/\partial I_{pC}$:

$$\alpha_{pnp} = \gamma \alpha_T M. \tag{5.3}$$

In the case of the pnp in the LIGBT at low V_{ak} values, the multiplication factor is negligible and set to 1. For the time being, the substrate current is neglected as well and we assume a linear relationship between

the hole current reaching the collector (in the LIGBT the cathode) and the total current from the emitter (in the LIGBT the p^+ anode):

$$\alpha_{pnp} = \frac{\partial I_{pE}}{\partial I_E}\frac{\partial I_{pC}}{\partial I_{pE}} = \frac{\partial I_{pC}}{\partial I_E} = \frac{I_{pK}}{I_A}. \tag{5.4}$$

This has been plotted in figure 5.10, where one can see that $\alpha_{pnp} \approx 0.4$ for $V_{ak} > 0.7\,\mathrm{V}$ (which has to be valid otherwise the bipolar is in cut-off). This is about the value that has been reported as typical (≈ 0.5) for vertical IGBT devices [Bal92, p. 363], where the current gain α_{pnp} is determined primarily by the base transport factor. Here, on the contrary, simulations show that both the injection efficiency *and* the transport factor are about $0.6 - 0.7$. For the device presented here, the injection efficiency is lower because of the higher doping of the base (nbuffer) relative to the emitter when compared to typical vertical IGBTs. On the other hand, the base transport factor is higher, since the base length is much shorter than the typical lengths in vertical devices.

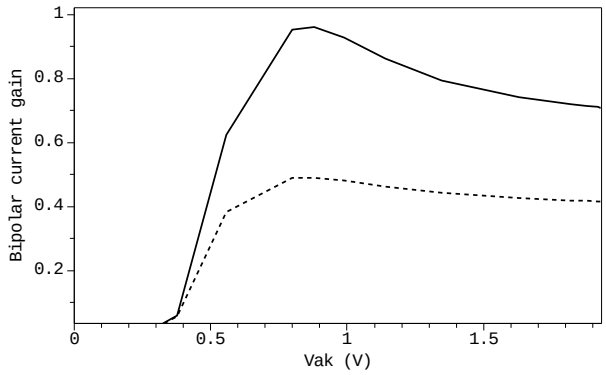


Figure 5.10 The common-emitter current gain β (solid) and the common-base current gain α (dashed) at $V_{gk} = 3.3\,\mathrm{V}$ of the intrinsic bipolar of the LIGBT with a buffer dose of $1e15\,\mathrm{cm}^{-2}$ of table 5.1.

Since the α_{pnp} is smaller than 0.5 in the forward conduction state, the β of the intrinsic bipolar is smaller than 1. This means that the LIGBT—once its intrinsic bipolar is on—is conducting a current that is almost twice as large as that of a comparable LDMOS. Indeed, in the approximation as made above:

$$I_E = I_B + I_C \quad \textit{and} \quad I_C = \beta I_B \tag{5.5}$$

$$I_E = (1+\beta) I_B \quad \textit{or} \quad I_A \approx 1.7 I_{eK}. \tag{5.6}$$

From equation (5.6) and since $\beta = \alpha_{pnp}/(1-\alpha_{pnp})$, the closer α_{pnp} is to unity, the more current is conducted. This is a demonstration of one

of the trade-offs within the IGBT: the higher the α_{pnp}, the better the device; but with high gain comes high levels of hole current, and thus faster latch-up (as all holes need to pass under the n^+ region).

One way to ameliorate this device is by introducing the pdrift together with the pwell, as to increase the doping in the p-layer under the n^+ cathode contact. This is shown in figure 5.11 where the latching current is higher, but there is still no saturation observed before latch-up at $V_{gk} = 3.3\,\mathrm{V}$. If a higher current level together with a wider SOA is wanted, then other approaches for the integration of IGBT devices have to be looked for. Note that neither the V_t nor the α_{pnp} is influenced by the pdrift. Yet one important matter, that has been neglected this far, needs to be studied: the substrate current.

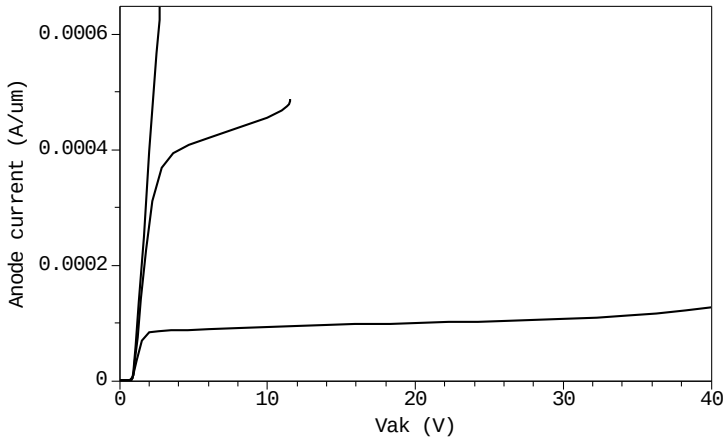


Figure 5.11 Output characteristics at $V_{gk} = 1.1$, 2.2 and 3.3 V of an LIGBT with a nbuffer dose of $1e15\,\mathrm{cm}^{-2}$ of table 5.1, and with pdrift over pwell.

Substrate current

Figure 5.12 shows the percentage of substrate current compared to the anode current of the simulated LIGBT (the pdrift has no influence on the substrate current level). This percentage (0.15%) seems to be acceptable, but improvement is still wanted, as these devices will be used as huge drivers (up to 10 A). The difference between the devices without and with BLN, is that those with BLN have a parasitic bipolar to the substrate with a larger and heavier doped base, which results in substrate currents that are several orders of magnitude smaller than the devices without BLN.

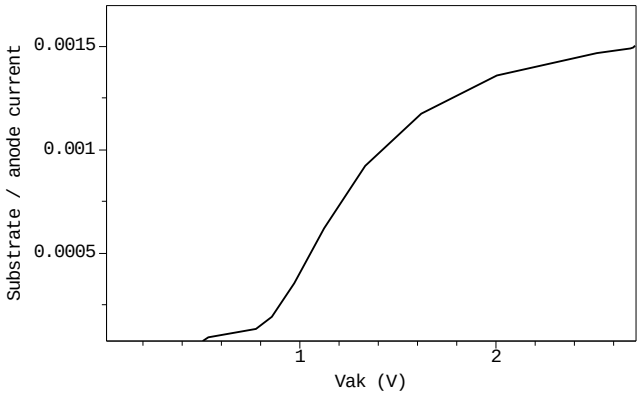


Figure 5.12 The substrate current at $V_{gk} = 3.3\,\mathrm{V}$ is about 0.1% of the anode current in the forward conduction mode for the LIGBT of figure 5.11.

5.3.2 A nLIGBT with BLN, psinker, pdrift and nwell

The LIGBT as designed this far is still subject to improvement, and this for several reasons:

- Latch-up occurs before saturation for $V_{gk} = 3.3\,\mathrm{V}$.

- Although the device has higher current levels in the on-state than the VDEMOS, higher common-emitter current gains have to be aimed for. The device with pdrift has a β that is smaller than 1, which is probably too low to compete with the VDEMOS. Further on, the power dissipation in the on-state will be discussed, and it will be seen that in order to have a competitive device, the on-state current of the LIGBT needs to be increased.

- The nbuffer, needed to prevent early breakdown in the off- and the on-state, asks for one extra mask. It would be ideal if an LIGBT could be developed without the need for an extra mask. Therefore, the LIGBT discussed in this subsection reintroduces the standard nwell layer. This will also facilitate comparison with designs studied further on. Thus, the LIGBT proposed in this section takes up the form as shown in figure 5.13.

- Furthermore, it is desirable that the substrate current level could be further decreased.

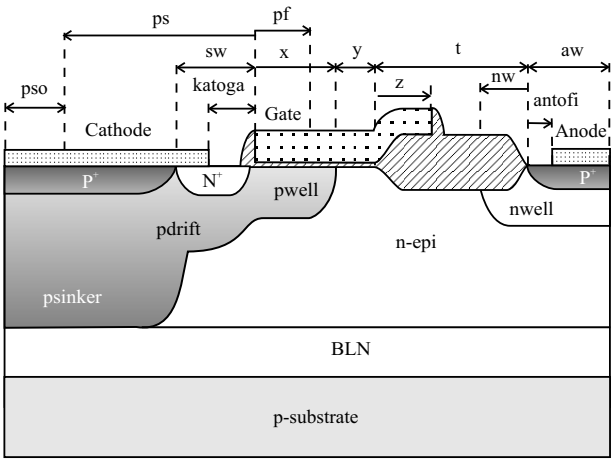


Figure 5.13 The LIGBT device with BLN, psinker, and pdrift. Note the introduction of a set of new parameters, of which the definition should be clear from the figure. Except maybe the opening in the psinker mask, *pso*.

Before discussing the results of the simulations, some important remarks regarding the layout have to be made first. The opening in the psinker mask (= *pso*) has to be large enough in order to introduce enough boron to reach the BLP. There is also a critical distance between the end of the opening (at the right hand side) in the psinker mask and the beginning of the channel. If the psinker is too close to this point, it influences the V_t. By taking this into account and by performing a limited number of layout variations, the following device parameters were used (in µm): $pso = 3$, $ps = 3$, $katoga = 0.4$, $sw = 0.8$, $pf = 0$, $x = 1$, $y = 0.8$, $t = 4.5$, $z = 1$, $nw = 2$, $aw = 1$, and $antofi = 0.2$. The reintroduction of the standard nwell yields a $45 - 50\,\mathrm{V}$ device, as previously discussed.

Figure 5.14 shows that the psinker helps to suppress the latch-up, as it increases the doping level and depth of the p-layer under the n^+ cathode. Indeed, the factor g/ρ_p plays the most important role in equation (5.1) when compared to l_k, which is already at its minimum, and α_{pnp}, which—as previously explained—may not decrease too much. Moreover, the psinker introduces an alternative path for the holes to flow. The holes that are injected in the n-epi are partly collected by the psinker without flowing near the surface via the pwell, which also helps to increase the latch-up current. When compared with a typical VDEMOS (figure 5.14), it is clearly seen that *once the pnp is on*, the nLIGBT has an on-state current level that is 4 to 6 times higher. This, however, does not mean that the device's *performance* is 4 to 6 times

better as will be explained in the paragraph on power dissipation.

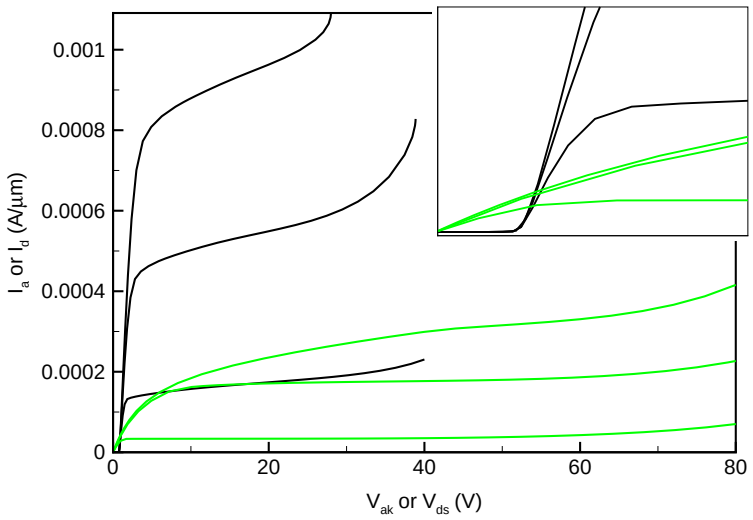


Figure 5.14 Output characteristics at $V_{gk} = 1.1$, 2.2 and 3.3 V of an LIGBT as shown in figure 5.13 (black) compared with a typical VDEMOS (grey). The inset shows the important region at low V_{ak} (or V_{ds}) values.

The reintroduction of the nwell, and the psinker helping to collect holes have their positive impact on the performance of this device: the β is slightly higher than 1, whereas the β of the device of the previous section was lower than 1 (figure 5.15). The substrate current level is almost unaffected by the introduction of the psinker. (figure 5.16).

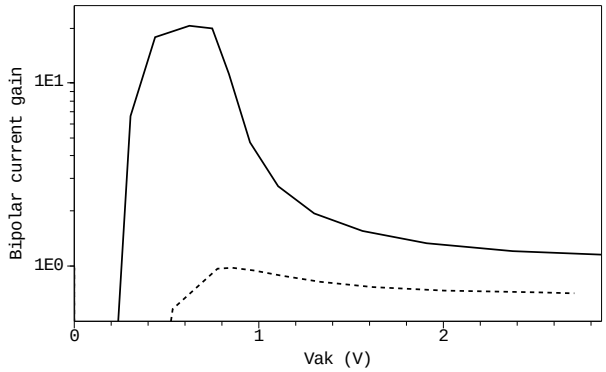


Figure 5.15 The common-emitter current gain β at $V_{gk} = 3.3$ V of an LIGBT as shown in figure 5.13 (solid line) compared with an LIGBT of the previous section (with pdrift).

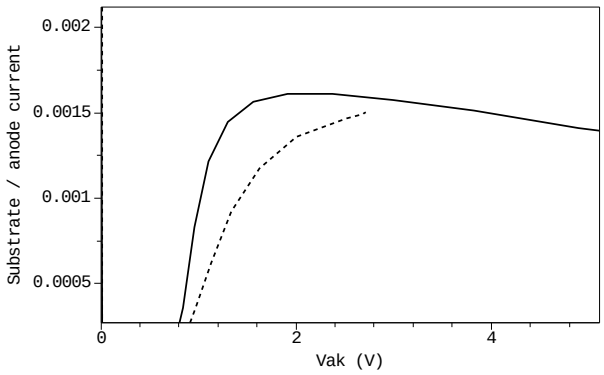


Figure 5.16 Substrate to anode current ratio at $V_{gk} = 3.3\,\mathrm{V}$ of an LIGBT as shown in figure 5.13 (solid line) compared with an LIGBT of the previous section (with pdrift).

Power dissipation

We have created a floating, $45 - 50\,\mathrm{V}$ nLIGBT without extra layers, with an acceptable substrate current, and with a wide SOA. Yet two other topics need to be addressed: the power dissipation and the turn-off. First of all, there is the competition with the DMOS device in terms of specific on-resistance versus breakdown voltage. From the beginning of this chapter, it has been stated that the DMOS devices are hard to beat under $200\,\mathrm{V}$. Yet we have developed a $50\,\mathrm{V}$ (!) LIGBT and hope that it outshines the DMOS device. In order to be able to compare the power MOS with the LIGBT device, a new method has to be defined. The forward conduction characteristic of an IGBT before saturation depends exponentially on the forward voltage drop (i.e., V_{ak}), instead of linearly, as is the case for a DMOS. Therefore, the specific on-resistance of the DMOS is no longer suitable, and a new figure comes into play—the power dissipation. The power dissipation in both the DMOS and the LIGBT are compared for devices with the same breakdown voltages. The power dissipation in a power device can arise during steady-state and switching conditions. For the LIGBTs under study with fast turn-on, turn-off and very low leakage currents, only the power loss during steady-state on-state is considered:

$$P_d = V_{fwd} I_{fwd} \frac{t_{on}}{T} \tag{5.7}$$

where V_{fwd} is the forward voltage drop at a current I_{fwd}, assumed to be constant during a fraction t_{on} of the total period T. A duty cycle of 50%

is arbitrarily chosen, which means that $t_{on}/T = 1/2$. If (5.7) is divided by the area (= width × pitch) the transistor takes up, then the factor cost enters the figure of merit, like in the specific on-resistance. The power dissipation per unit area thus obtained is actually the forward voltage drop times the "current density" times the duty cycle. Note that in the case of lateral devices, the "current density" is defined as the total current the device conducts divided by the area the transistor takes up, and is therefore not a "normal" current density.

If this current density is plotted against the power dissipation per unit area, then a tool is created to compare different types of devices (provided that they have the same breakdown voltage, which is not exactly the case since the VDMOS is an 80 V device). This is shown in figure 5.17, where the pitch of the VDMOS and the LIGBT are 7.5 µm and 13.3 µm, respectively.

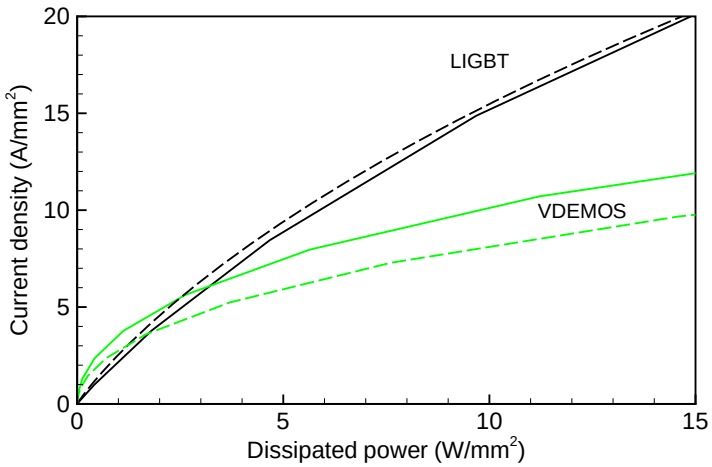


Figure 5.17 Current density versus power dissipation at $V_{gk} = 3.3\,\mathrm{V}$ for an LIGBT as shown in figure 5.13 (solid black) compared with a typical VDEMOS (solid grey) at 27° C. The dashed curves are at 125° C.

Figure 5.17 shows there is a break-even around $6.1\,\mathrm{A/mm^2}$ at a power dissipation per unit area of $3.2\,\mathrm{W/mm^2}$. At lower values of power dissipation, the DMOS device—although it has a saturated current level of 4 to 6 times lower than that of the LIGBT—is better than the LIGBT, since it has a higher current density for the same amount of dissipated power per unit area. This is because the DMOS does not know the 0.7 V forward voltage drop, which is typically the case for an IGBT (forward biasing of the p^+/nwell junction). This break-even point is at so high a value ($320\,\mathrm{W/cm^2}$!) that the feasibility of the LIGBT as an integrated

medium voltage device becomes questionable. However, it was reported by AMIS that in a similar technology a $0.15\,\mathrm{mm}^2$ DMOS driver dissipated approximately $1\,W$ in DC (i.e., $6.7\,\mathrm{W/mm}^2$), and no problems were observed whatsoever. It is even assumed that higher power dissipation values are possible. So with a duty cycle of 50 % in AC mode much higher power dissipation per unit area than the break-even point at $3.2\,\mathrm{W/mm}^2$ is possible.

The power dissipation in a power device is limited due to the temperature rise it causes:

$$\Delta T = T_j - T_a = P_d R_\theta \tag{5.8}$$

where R_θ is the thermal resistance, and T_j and T_a are the junction and ambient temperatures, respectively. The power device should actually be simulated at the maximum allowable temperature (generally $T_j = 125^\circ\,\mathrm{C}$). Therefore, isothermal simulations (non-isothermal effects are neglected since the on-state voltage is rather low) at this temperature give a better idea of the power dissipation per unit area. This has been done in figure 5.17, where it can be seen that due to the drop of the mobility in the DMOS device (which is not the case for the IGBT because of the conductivity modulation) the break-even point drops to $3\,\mathrm{A/mm}^2$ at a power dissipation per unit area of $1.2\,\mathrm{W/mm}^2$. This leads to the conclusion that the LIGBT devices as drivers are likely to beat the DMOS devices, even at medium voltage ranges ($50 - 80\,\mathrm{V}$).

Turn-off

It has been assumed in the previous paragraph that the turn-off of the LIGBT is fast enough to assure negligible power loss during turn-off. In literature, the turn-off of the IGBT is treated as one of the critical trade-off parameters. The holes that are injected in the drift region need time to recombine after gate turn-off, and this becomes critical in large, discrete IGBT devices. However, in $50 - 80\,\mathrm{V}$ lateral integrated IGBT devices this is not a problem, as is shown below.

The gate has been turned off from $V_{gk} = 3.3\,\mathrm{V}$ to $V_{gk} = 0\,\mathrm{V}$ in $10\,\mathrm{ns}$ with a resistive load of $10\,\mathrm{k}\Omega$ and an off-state voltage of $40\,\mathrm{V}$. The result of this set-up is that the initial steady on-state current is equal to the static latch-up onset current. Figure 5.18 shows that the anode current decreases fast as the gate voltage drops from $3.3\,\mathrm{V}$ to the threshold voltage because of the reduction of the channel current to zero (demonstrated by the electron cathode current). After this point, the

hole current (on cathode and anode) decreases more gradually. Although the initial steady on-state current is the static latch-up onset current, the anode current drops one order of magnitude in 40 ns. Note that in large, discrete devices the initial current is likely not to be equal to the static latch-up current as the dynamic latch-up is most of the times smaller than the static latch-up.

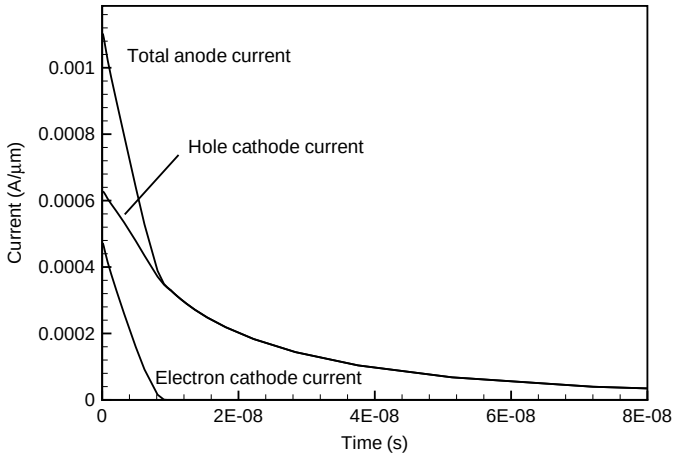


Figure 5.18 Gate controlled turn-off from $V_{gk} = 3.3\,\mathrm{V}$ to $V_{gk} = 0\,\mathrm{V}$ in 10 ns with a resistive load of 10 kΩ and an off-state voltage of 40 V. The initial current level is equal to the static latch-up onset current.

5.4 A Low-Side nLIGBT with BLP (standard nLIGBT)

The previous section treated a new form of the lateral IGBT. The standard implementation of the LIGBT in a junction isolated technology takes up the form as sketched in figure 5.19 [AU01]. This device is non-floating (low-side) and profits from the RESURF effect.

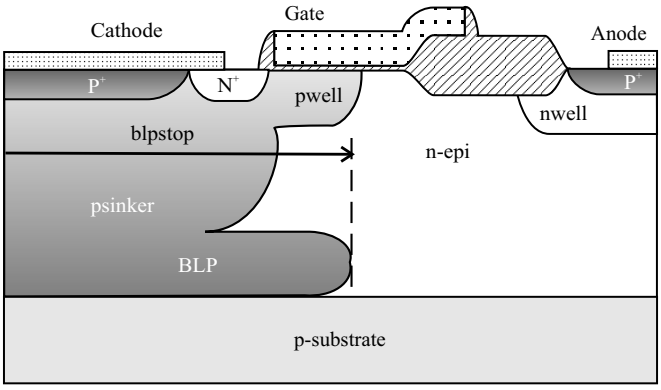


Figure 5.19 The non-floating LIGBT device with psinker and BLP. One extra layout parameter in comparison with figure 5.13 is needed here: blpstop.

Table 5.2 Breakdown voltage (V_{br}) for several values of the layout parameter blpstop.

blpstop (μm)	V_{br} (V)
0[a]	58
2	60
4	65
6	80
8	90
10	85
12	76
14[b]	72

[a]no blp present
[b]blp uniformly present (i.e., blank implant)

The fact that the nbuffer layer was omitted (as mentioned in the previous section), is justified by the high breakdown voltages obtained with the standard CMOS nwell layer for this form of the LIGBT (table 5.2), all by using the same layout parameters as for the device with BLN

and psinker (resulting in a pitch of 13.3 µm). The breakdown voltage reaches a maximum at a certain value of blpstop because the BLP was primarily not designed to serve as a RESURF layer. The lateral diffusion of the BLP under the anode contact yields better RESURF states (figure 5.20) for values of blpstop around 8 µm.

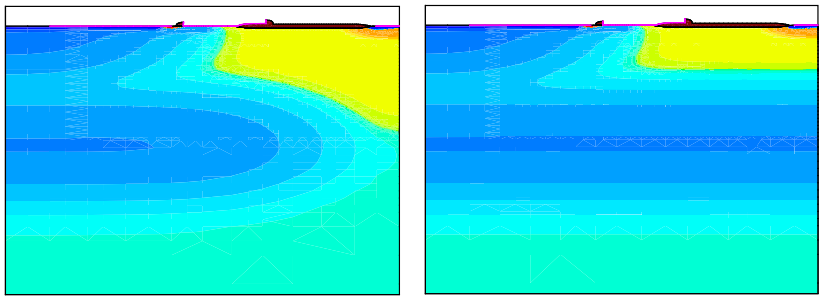


Figure 5.20 The non-floating nLIGBT with blpstop = 8 µm has a better RESURF state (left) than the nLIGBT with blank BLP implant (right) due to the bevelled edge of the blp/n-epi junction.

The output characteristics are plotted in figure 5.21. As expected, there is no problem with latch-up (wide SOA) and the saturation currents of the non-floating devices are higher than those seen until now.

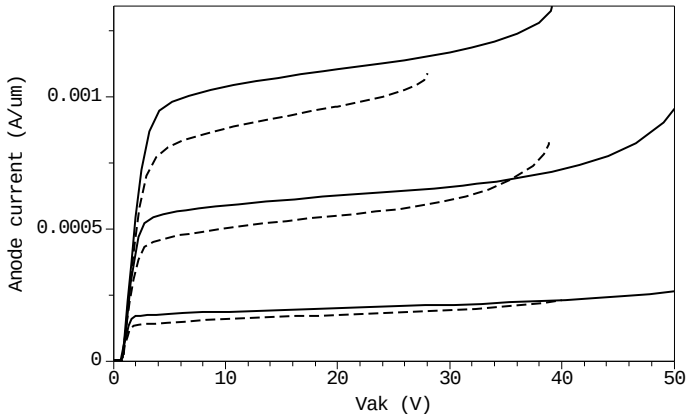


Figure 5.21 Output characteristics at V_{gk} = 1.1, 2.2 and 3.3 V of a non-floating nLIGBT (figure 5.19 with blpstop = 8 µm, solid lines) compared with a floating nLIGBT with BLN and psinker (figure 5.13).

The common-emitter current gain β of these devices is thus higher than the β of the floating LIGBT with BLN and psinker. The turn-off is as good as the previous device and poses no problem.

The potential distribution at forward conduction is such that it is slightly unfavourable for holes to flow towards the substrate contact. This condition is not so severe at $V_{ak} = 0.7\,\mathrm{V}$, but grows with increasing V_{ak} until it saturates at $V_{ak} \approx 2\,\mathrm{V}$. As a result, a behaviour as plotted in figure 5.22 is observed. This is compared with the substrate to anode current ratio for a device with BLN and psinker.

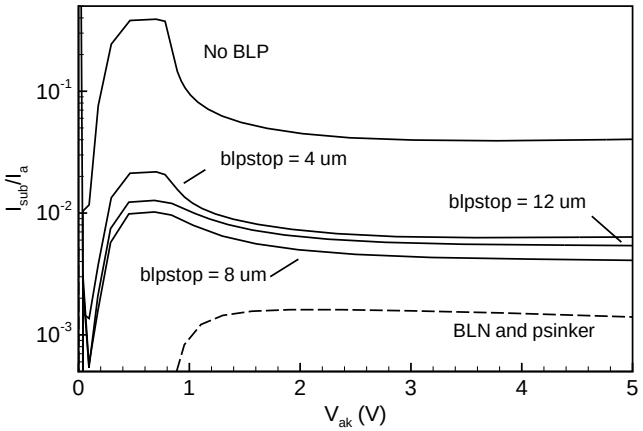


Figure 5.22 Substrate to anode current ratio at $V_{gk} = 3.3\,\mathrm{V}$ of a non-floating nLIGBT (figure 5.19 with several blpstop values) compared with a floating nLIGBT with BLN and psinker (figure 5.13).

As a conclusion, the standard non-floating device performs better than the one described in the previous section (higher V_{br}, wider SOA, higher β), except for the one important matter of the substrate current. The non-floating LIGBT without protection towards the substrate is harder to integrate on chip as it has a substrate current level in the order of 0.5 % to 1 % of the anode current. In search for a device that combines the best of both the option with the BLN and with the BLP, we designed a device with BLN *and* BLP, as described in the following section.

5.5 A High-Side nLIGBT with BLP *and* BLN

5.5.1 The standard process flow

The approach discussed in the previous section could no longer be used as a floating device since the p-substrate and the cathode are shorted. It is possible, however, to use the BLN together with the BLP in the I3T80 technology, which results in a situation as shown in figure 5.23. Again, the same layout has been used for this device as for the devices with BLN and psinker, and BLP and psinker (in µm): $pso = 3$, $ps = 3$, $katoga = 0.4$, $sw = 0.8$, $pf = 0$, $x = 1$, $y = 0.8$, $t = 4.5$, $z = 1$, $nw = 1.5$, $aw = 1$, and $antofi = 0.2$ (only nw is 0.5 µm smaller, but most important is that the pitch remains the same: pitch = 13.3 µm).

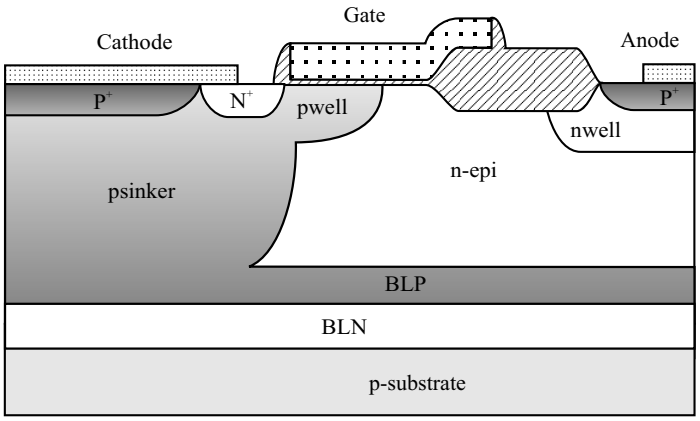


Figure 5.23 The LIGBT device with psinker, BLP and BLN.

This is a floating device, as the eventual potential difference between cathode and substrate is held by the BLN/p-substrate junction. This is the case whether the BLN is contacted and shorted to the cathode, or not. When it is shorted to the cathode, the potential difference is clearly sustained by the substrate for a large part as the BLN is much higher doped than the substrate. When the BLN is not contacted via a nsinker, then it floats between the cathode and substrate, and behaves as the open base of a pnp transistor. If the cathode potential becomes higher than the substrate potential, the BLN/substrate junction sustains the potential difference. Yet there *is* an important difference between an LIGBT with contacted BLN and an LIGBT with floating BLN, as will be explained later.

This device is designed as an alternative for the non-floating standard

device of the previous section, and it is hoped that the substrate current level decreases as in the case of the LIGBT with BLN and psinker. Concerning the off-state and the forward conduction state, this device is competitive: $V_{br} = 73\,\mathrm{V}$ (reintroduction of the RESURF technique !) and the common-emitter current gain at e.g., $V_{fwd} = 1.8\,\mathrm{V}$ is 1.6 (where it was 1.8 for the device with BLP and psinker and 1.4 for the device with BLN and psinker). Figure 5.24 shows the output characteristics compared with those of the previous devices.

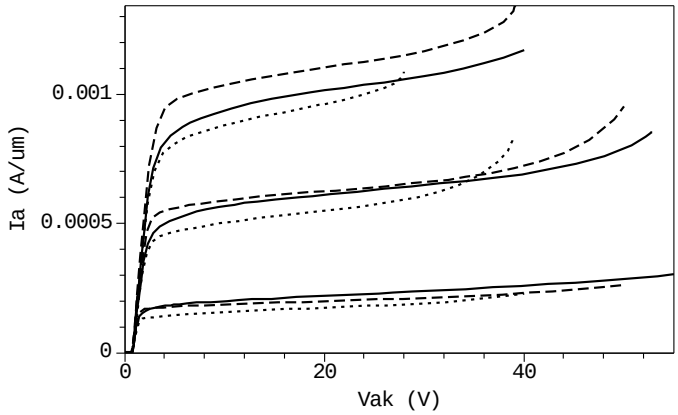


Figure 5.24 Output characteristics at $V_{gk} = 1.1,\ 2.2$ and $3.3\,\mathrm{V}$ of a floating nLIGBT (figure 5.23, solid lines) compared with a non-floating nLIGBT (figure 5.19 with blpstop = $8\,\mu\mathrm{m}$, long dashed) and with a floating nLIGBT with BLN and psinker (figure 5.13, short dashed).

The 30 V difference between the off-state and the on-state breakdown voltage is explained through figure 5.25. The blocking voltage in the off-state is held by the BLP/n-epi junction. As the n-epi is much lower doped than the BLP, the n-epi depletes completely up to the gate oxide and the nwell. This yields a good spreading of the potential lines, with several electric field peaks (RESURF effect).

In the on-state, on the other hand, the high current levels in the ndrift (= nwell + n-epi) region result in a small potential drop in this region (conductivity modulation). Since the BLP is weakly linked (the standard BLP on top of the BLN is not very wide, nor very highly doped) to the cathode contact, it becomes completed depleted under the anode. As a result, the BLN is now at a high potential. Since the BLN is wide and very highly doped, it acts as an equipotential surface, which means that almost the entire potential difference between the anode and the cathode is held by the BLN/BLP + psinker junction at the cathode side.

However, impact ionisation mostly occurs under the bird's beak situated under the gate, as this is the region where both high current levels and high electric fields occur (although the electric fields at the BLN/BLP junction are higher, almost no current is passing, so almost no impact ionization is taking place there).

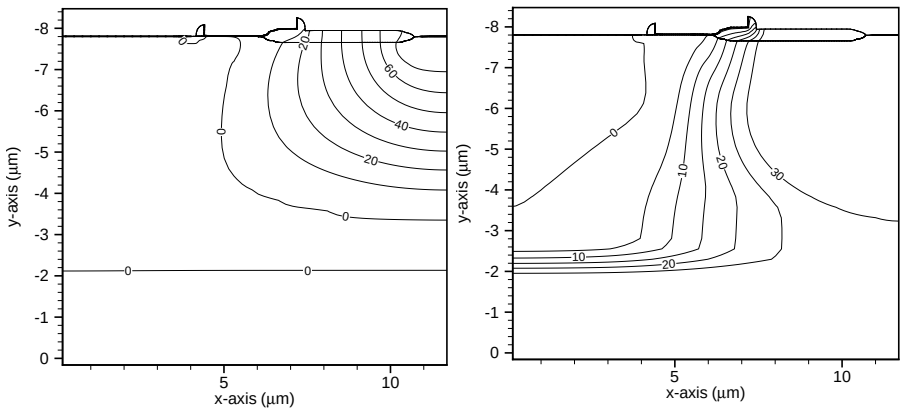


Figure 5.25 Electrostatic potential distribution (V) in the off-state (left) and the on-state (right) for a nLIGBT with psinker, BLP and BLN.

Substrate current behaviour

Figure 5.26 plots the ratio of substrate to anode current for the device with a floating BLN, compared with the LIGBTs discussed this far. The ratio is the smallest for this new device and remarkable is the decreasing ratio with increasing V_{ak}. These results open new perspectives as to whether or not the lateral IGBT is feasible for an 80 V junction isolated technology. The question arises if the behaviour as seen in figure 5.26 can be further improved by altering some process or layout conditions.

To answer this question, the particular behaviour of the substrate current needs to be explained. First of all, 2D plots at several V_{ak} values help to gain insight into the physical processes at stake. And secondly, an equivalent circuit will help to understand which parameters have to be varied in order to reduce the substrate current.

Figure 5.27 shows that at $V_{ak} = 2\,\mathrm{V}$ the potential differences *inside* the device between BLP and BLN are not yet pronounced, while the

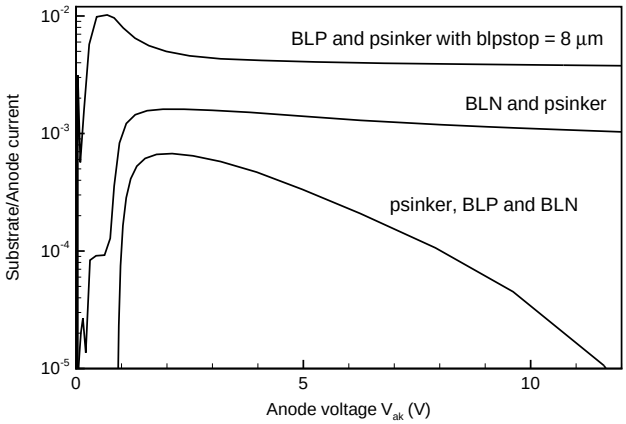


Figure 5.26 Substrate to anode current ratio at $V_{gk} = 3.3\,\mathrm{V}$ for an LIGBT with psinker, BLP and floating BLN compared with previous designs.

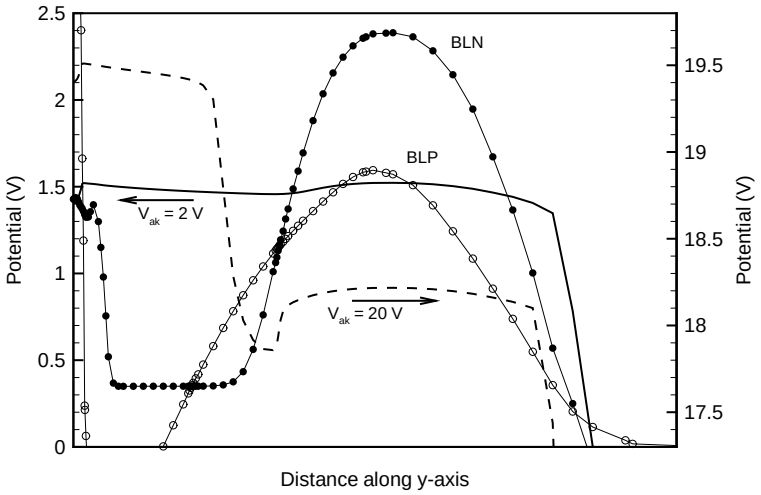


Figure 5.27 Donor (•) and acceptor concentration (○), potential distribution for $V_{gk} = 3.3\,\mathrm{V}$, and at $V_{ak} = 2\,\mathrm{V}$ (solid line) and $V_{ak} = 20\,\mathrm{V}$ (dashed line) as a function of depth taken at the end of the device (x = pitch). Note that both the left and the right axis use the same scale for the potential, which shows that the potential bump at $V_{ak} = 2\,\mathrm{V}$ becomes a serious barrier at $V_{ak} = 20\,\mathrm{V}$.

hole current has already reached its maximum (saturation) at this anode voltage. At higher V_{ak} values, the initial potential bump between the BLP and the BLN becomes a high barrier for the holes to cross. This results in a decrease of the substrate current. This is illustrated in figure 5.27, where the potential distribution is plotted versus depth, taken at the end of the device (anode side) at $V_{ak} = 2\,\mathrm{V}$ and $V_{ak} = 20\,\mathrm{V}$.

The use of an equivalent circuit allows a further analysis of the phenomena encountered. As can be seen in figure 5.23, the anode side of the present LIGBT becomes a potentially dangerous parasitic structure as it joins 3 bipolars (the fourth, pnp_3, is the primal one). The first bipolar—an intended one (pnp_1)—consists of the p^+ anode, nwell (+ n-epi) and BLP, the second—parasitic—one consists of the nwell (+ n-epi), BLP and BLN (npn) and the third—parasitic—one consists of the BLP, BLN and p-substrate (pnp_2).

This situation is drawn in figure 5.28, together with the MOS structure, the primal bipolar pnp_3 and the important resistive elements. Using this equivalent circuit, the substrate current behaviour can be explained. At low anode voltage ($\sim 2\,\mathrm{V}$), the hole current through pnp_1 is already at its saturation value, causing V_2 to rise above V_1, switching on the parasitic npn, which on its turn provides the pnp_2 with base current, resulting in a substrate current. If V_a increases, V_1 follows V_a more than V_2, because V_2 is tight to the cathode via the psinker. This results in a cut-off of the npn, and – at the same time – of the pnp_2 (its base current is vanishing). Furthermore, the potential in the floating BLN follows V_1 (see also figure 5.27, where the potential in the BLN is still above 17 V at $V_{ak} = 20\,\mathrm{V}$). This puts the BLP at a lower potential than the BLN, which also cuts off the parasitic pnp_2.

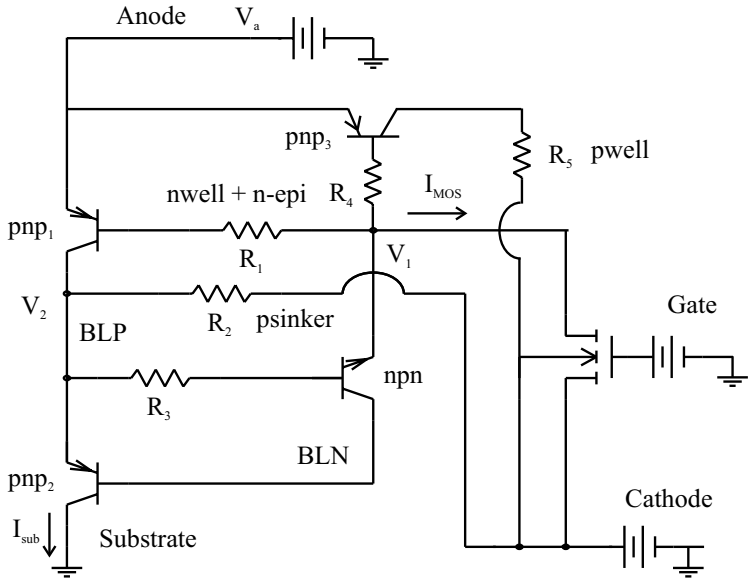


Figure 5.28 Equivalent circuit for a nLIGBT with floating BLN.

Using figure 5.28, assuming a constant MOS current I_{MOS} (controlled by the gate voltage) and assuming ideal bipolar transistors ($I_C = \beta I_B$ and $V_{BEpnp/npn} = -/+ 0.7\,\mathrm{V}$), one can calculate the substrate current I_{sub} as a function of the anode voltage V_a, I_{MOS}, the common emitter current gains of the bipolar transistors (β_1 for pnp_1, β_2 for pnp_2, β_3 for pnp_3, and β_n for the npn), and the resistors R_1, R_2, R_3 and R_4:

$$I_{sub} = \frac{1}{\delta}\left(\gamma I_{MOS} - V_a\right) \tag{5.9}$$

with

$$\gamma = \frac{1}{\kappa}\left(R_1 + \beta_1 R_2\right) \quad \text{and} \tag{5.10}$$

$$\delta = \left(\frac{1}{\kappa\beta_2} + \frac{1}{\kappa\beta_2\beta_n}\right) R_1 + \left(1 + \frac{1}{\beta_2} + \frac{\beta_1}{\kappa\beta_2} + \frac{1}{\beta_2\beta_n} + \frac{\beta_1}{\kappa\beta_2\beta_n}\right) R_2 + \frac{1}{\beta_2\beta_n} R_3 \quad \text{and} \tag{5.11}$$

$$\kappa = 1 + \frac{R_1}{R_4}. \tag{5.12}$$

This means that $I_{sub} = 0$ if $V_a > \gamma I_{MOS}$ ($V_a > 0\,\mathrm{V}$!), that I_{sub} decreases with increasing V_a/δ, and thus decreases faster for smaller values of δ. Furthermore, the maximum I_{sub} (i.e., for the smallest V_a for which the assumptions are still valid) decreases with decreasing γ/δ. These results will be demonstrated in the section discussing the process and layout variations of the BLN.

Contacting the BLN

Using the equivalent circuit approach also explains what happens when the BLN is contacted via the nsinker. The equivalent circuit for this situation is shown in figure 5.29. The npn is no longer conducting current towards V_1, but *away from* V_1; that is, if the nsinker is at a low enough voltage. The npn is then part of a thyristor (together with pnp_1), which latches as soon as the npn is on. Gate control is lost as a huge electron current flows towards the nsinker contact. This current provides base current for pnp_2, yielding substantial substrate currents. This behaviour has been confirmed by TCAD simulation, see figures 5.30 and 5.31 for e.g., a nsinker to cathode voltage $V_{ns} = 0\,\mathrm{V}$.

If the nsinker is put on a higher voltage, the current changes sign and flows away from the nsinker. This means that the collector and

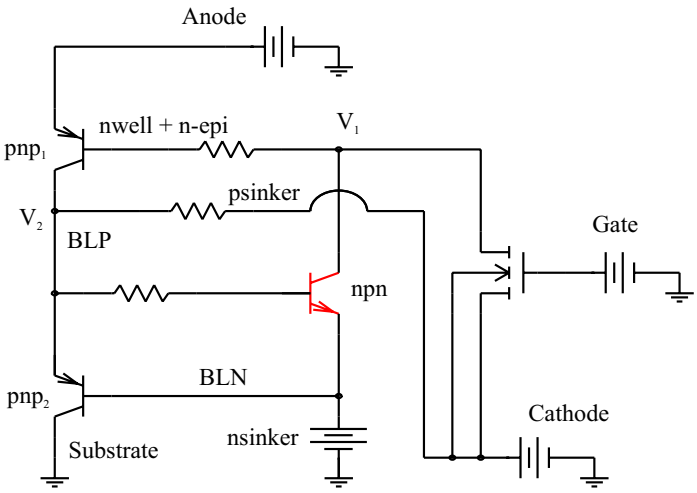


Figure 5.29 Equivalent circuit for a nLIGBT with contacted BLN, in the case where the nsinker contact is at low voltage compared to the anode voltage.

emitter of the npn have changed place again. In this condition, V_2 is slightly higher than V_1 (enough to forward bias the BLP/n-epi emitter junction), but lower than the potential at the nsinker contact. Therefore *electrons* flow from the nwell + n-epi to the nsinker (the current has an opposite sign) *without* generating *any* substrate current (V_2 is lower than the potential at the nsinker contact). But, as soon as the anode voltage increases above a certain potential (depending on the potential on the nsinker contact), the thyristor mode is attained again, yielding loss of gate control and large substrate currents. This is confirmed by simulations and shown in figures 5.30 and 5.31.

When the nsinker is biased, a substantial current flows through the nsinker contact (about 20 % of the device's anode current). Furthermore, the nsinker can not be biased too high relative to the cathode, because of possible BLP/BLN breakdown. These disadvantages, together with the potentially dangerous thyristor working, let us conclude that the floating BLN approach is the safest and easiest one *when the standard process conditions are used*.

5.5.2 Changing the BLN layer

Until now, the standard process flow was considered, which has as a huge advantage that the devices discussed come "for free"; that is, without any changes to the process flow, nor any new defined layers. However, from a more academic standpoint, it is interesting to know to what de-

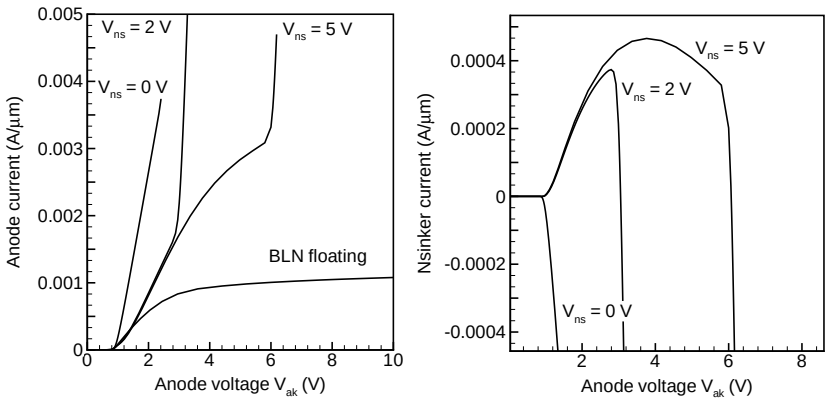


Figure 5.30 Anode currents at $V_{gk} = 3.3\,\mathrm{V}$ in the case of floating and contacted BLN (left). Currents at $V_{gk} = 3.3\,\mathrm{V}$ through the nsinker contact in the case of a contacted BLN (right).

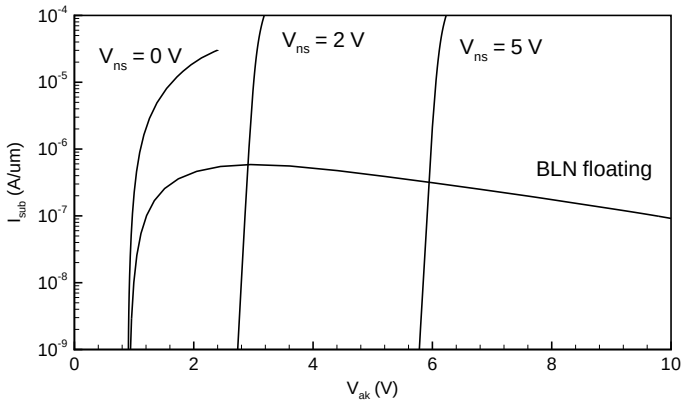


Figure 5.31 Substrate currents at $V_{gk} = 3.3\,\mathrm{V}$ in the case of floating and contacted BLN.

gree the substrate current can be suppressed. Variations on the BLN implant conditions demonstrate how the substrate current can be suppressed in this LIGBT. The equivalent circuit helps to understand the mechanisms at stake.

Breakdown and on-state

As can be seen in figure 5.25, the BLP/n-epi junction, together with the RESURF effect, determines the breakdown voltage. Therefore, when the BLN is altered, the influence on the blocking capability of the nLIGBT is small (from 68 V for a very highly doped BLN to 72 V when no net BLN is present).

The influence on the on-state and SOA, on the other hand, is substantial. When the BLN dose is lowered (when compared to the standard BLN), the BLP comes out more pronounced between the BLN and the n-epi. As a consequence, when the BLP is wide enough, it is able to electrically isolate the BLN from the n-epi in the on-state, which is not the case for the standard BLN (figure 5.25). The potential distribution in the on-state is then similar to the one in the off-state.

Another important consequence of this more pronounced BLP is that the holes coming from the anode and passing through the BLP towards the psinker, have a less resistive path to follow. Therefore, more holes follow this path in comparison with the standard nLIGBT. As a result, both the on-state and SOA improve (figure 5.32).

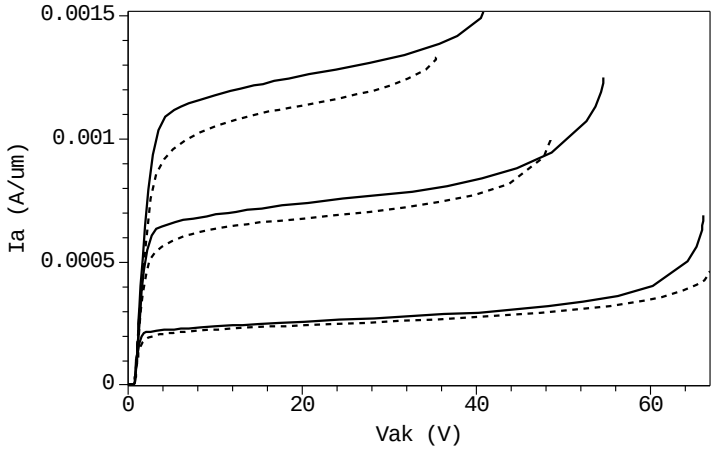


Figure 5.32 Output characteristics at V_{gk} = 1.1, 2.2 and 3.3 V of a floating nLIGBT with non-standard BLN (solid lines) compared with a floating nLIGBT with standard BLN (dashed lines). The non-standard BLN dose is 20 times lower than the standard dose.

Substrate current

First of all, a reference dose (*Refdose*) is chosen that is so low that no net BLN is present. This yields a device without the extra BLN buffer, and serves as a reference (this corresponds to a device with a blank BLP implant, compare figure 5.22 with figure 5.33).

For the lowest BLN dose (2 × *Refdose*) where the BLN *is* present, β_2 (i.e., the β of the pnp_2 towards the substrate, see figure 5.28) is at its highest value. But, at the same time, the BLP is strongly present for this BLN dose, which means that β_n (i.e., the β of the npn) is at

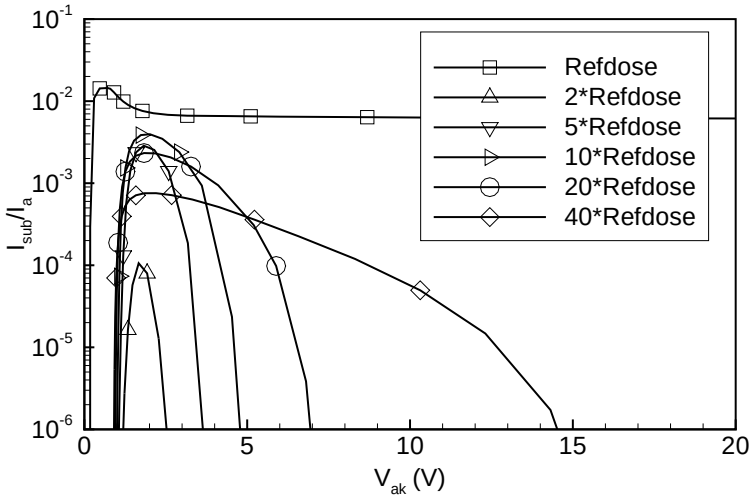


Figure 5.33 Substrate to anode current ratio at $V_{gk} = 3.3\,\mathrm{V}$ for several LIGBTs with different BLN doses. Note that the lowest dose yields a device without BLN and that the highest dose is the standard dose.

its lowest value, as well as R_2 and R_3. This means that although the pnp towards the substrate is most effective for this low BLN dose, the substrate to anode current ratio is the lowest (see figure 5.33) because of the highly conductive path towards the psinker and because of the very bad npn that provides the base current for the pnp towards the substrate.

When the BLN dose is increased, β_2 decreases as its base (= BLN) not only increases in dose but also widens. As a result of this growing BLN, the BLP shrinks, which on its turn yields a increasing β_n, R_2 and R_3. This explains the growing substrate to anode current ratio for the BLN doses = 2 ×, 5 × and 10 × *Refdose*.

When the BLN dose is yet further increased, the β_2 becomes so small that it is the dominant factor. As a consequence, the substrate current decreases again. For the highest BLN dose (= 40 × *Refdose*), it is also observed that the decrease of the substrate to anode current ratio with increasing V_{ak} is the slowest. As the BLP is the smallest for this BLN dose (highest R_2 value), the npn stays on even with high V_{ak}, because V_2 remains larger than V_1 even for the higher V_{ak} values. Whereas for the smaller BLN doses, the BLP is more conductive (smaller R_2), which shuts off the npn faster because V_2 becomes smaller than V_1 at lower V_{ak} values. This also explains why for all BLN doses the substrate to anode current ratio vanishes completely when V_{ak} is large enough.

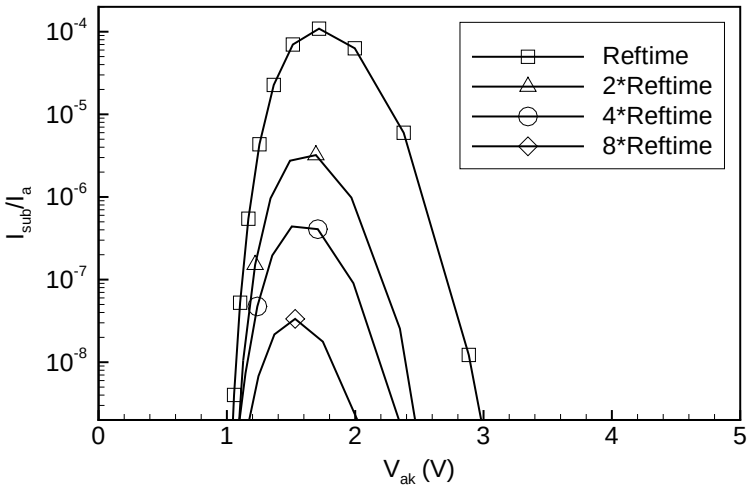


Figure 5.34 Substrate to anode current ratio versus V_{ak} (at $V_{gk} = 3.3\,\mathrm{V}$) for different BLN anneal times (BLN dose $= 2 * \mathit{Refdose}$). *Reftime* is the standard anneal time as used in figure 5.33.

Of course, when the BLN and the BLP are made as such that both the pnp towards the substrate *and* the npn have a very small *beta* (and, consequently, R_2 and R_3 are small as well), then the substrate current can be completely killed for all V_{ak} values. This is demonstrated in figure 5.34 where the anneal time is increased for the low BLN dose ($2 \times$ *Refdose*). Due to processing conditions, this results in the situation just described. In this way, the substrate to anode current ratio is reduced to 4×10^{-8}.

Substrate current with a biased BLN

Contacting the standard BLN is out of the question because of the parasitic thyristor latch-up. When the BLN is changed, it is possible to obtain extremely good results with a biased BLN via a nsinker contact. The anode current remains the same when compared to a device with the same —but floating— BLN (in the present example $2 \times$ *Refdose* and *Reftime*), provided that the voltage on the nsinker contact is higher than the voltage on the cathode contact (e.g., $V_{ns} > 3.3\,\mathrm{V}$, for the nLIGBT at stake, see figure 5.35). When V_{ns} is too low, the same happens as described when contacting the standard BLN (high currents at all contacts, see figures 5.35 and 5.36).

When V_{ns} is high enough, it keeps the BLN at a potential that keeps

all parasitic bipolars in cut-off, even at breakdown (in the off-sate as well as in the on-state). As can be seen in figures 5.35 and 5.36, the nsinker and substrate current for $V_{ns} = 5.5\,\mathrm{V}$ are reduced to negligible values.

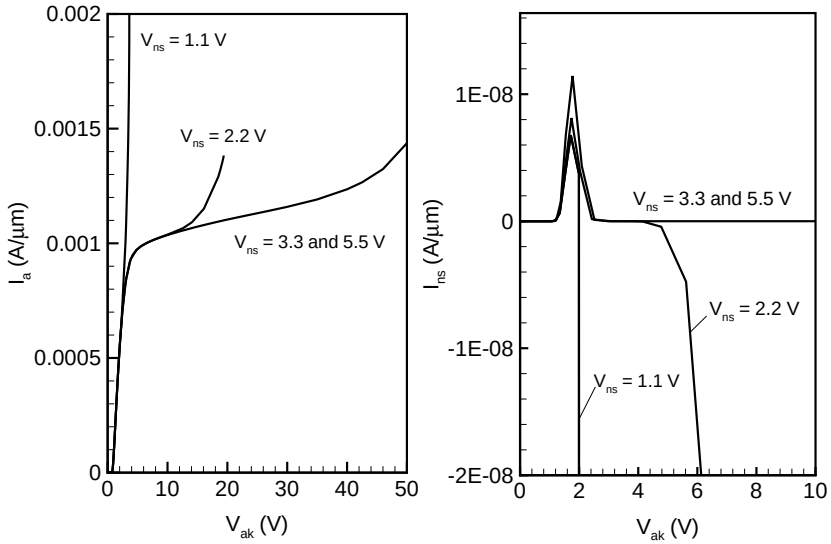


Figure 5.35 Anode currents at $V_{gk} = 3.3\,\mathrm{V}$ in the case of a contacted non-standard BLN (left). Currents at $V_{gk} = 3.3\,\mathrm{V}$ through the nsinker contact in the case of a contacted non-standard BLN (right).

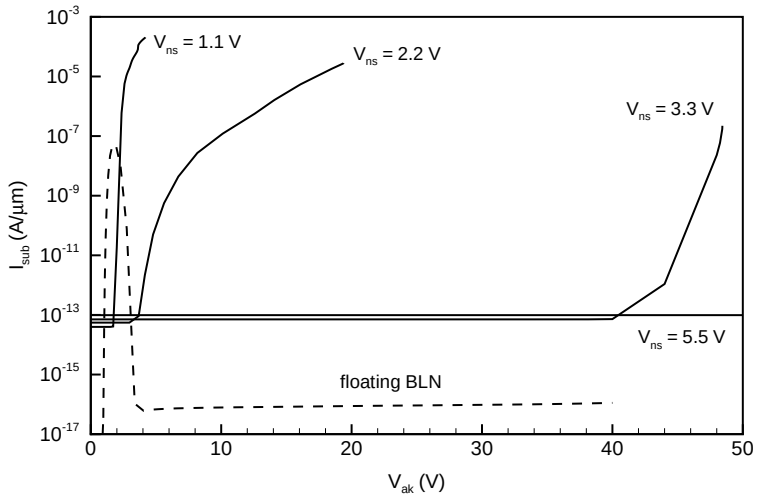


Figure 5.36 Substrate currents at $V_{gk} = 3.3\,\mathrm{V}$ in the case of a contacted non-standard BLN.

5.5.3 Patterning of the BLN masker

Changing the process conditions of a layer that already serves other purposes, is a difficult and hazardous undertaking. Since it has been shown that lower BLN doses improve the device's performance, another approach is possible: the patterning of the BLN masker. The BLN implant is no longer blank, but masked in a regular way in order to keep a doping profile that is almost constant, but lower than the blank implant. Normally the openings are squares, and the ratio of the side of such a square to the distance between two squares determines the doping level that is obtained. Since we are working in 2D, the openings are infinitely long stripes. Both mask types can be related to one another.

The lowering of the BLN doping profile has the same effect as explained above. Of course, the best results —when a full control over the BLN layer is available— are not obtained through a simple patterning of the BLN. Nevertheless, the substrate to anode ratio could be further diminished to a value of 5×10^{-5} (figure 5.37).

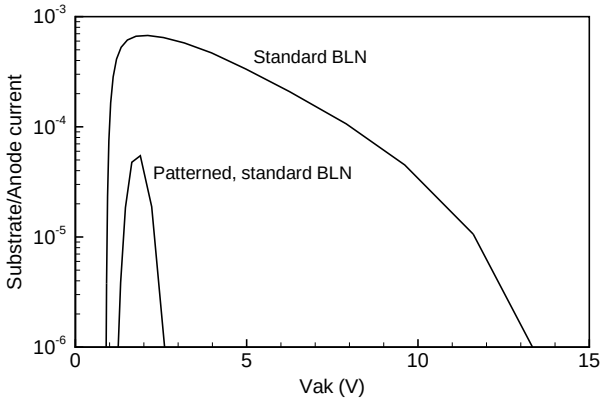


Figure 5.37 Substrate to anode current ratio at $V_{gk} = 3.3\,\mathrm{V}$ for the nLIGBT with a patterned BLN, compared to the standard case (no BLN mask).

The price to pay when patterning the BLN is the lost of the floating capability of the device. It is a sheer consequence of the nature of the I3T80 technology, where the BLP peeks out from behind the BLN when the BLN is patterned. As a result the floating capability is no longer determined by the BLN/p-substrate junction, but by the BLN/BLP + p-substrate junction. For the example given above (figure 5.37), this junction breaks down at 25 V.

5.6 P-type LIGBTs

5.6.1 Blocking capability

P-type devices need to have a p-type drift region, which means that in the I3T80 technology the pdrift and the pwell can be used. Since this technology is n-epi based, the pdrift/n-epi/p-substrate parasitic bipolar needs to be killed. This is done by the BLN shorted to the anode by the nsinker (figure 5.38), which results in the mirror image of the standard nLIGBT (figure 5.19). Except for the one important difference that now a 4 layer structure is appearing at the cathode side of the device (n^+/pwell + pdrift/n-epi + BLN/p-substrate).

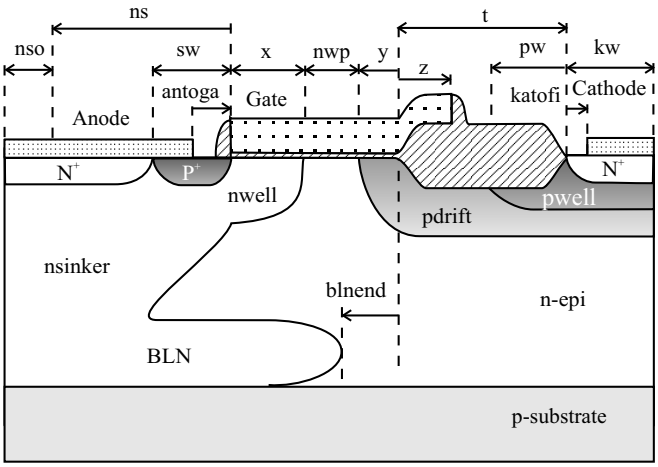


Figure 5.38 The floating pLIGBT with nsinker and BLN. All the layout parameters defining this device are shown.

The values of the layout parameters are approximately the same as for the nLIGBT (in μm): $nso = 3$, $ns = 7$, $antoga = 0.4$, $sw = 0.8$, $x = 1$, $nwp = 0$, $y = 0.8$, $t = 4.5$, $z = 1$, $pw = 2$, $kw = 1$, and $katofi = 0.2$ (the larger out-diffusion of the nsinker results in a larger ns value compared to the ps in the nLIGBT).

Important difference between the pLIGBT and the nLIGBT is that the n-epi is part of the anode (or "source") side of the pLIGBT, whereas in the nLIGBT the n-epi was part of the conductivity modulated (in the on-state) drift region. The potential difference between anode and cathode has to be supported by the n^+/pdrift/n-epi bipolar; thus by the pdrift/n-epi junction. However, the pdrift/n-epi junction only partly determines the breakdown voltage, as both in the pdrift and the n-epi, the depletion layer is stopped: by the pwell in the pdrift and by the BLN

in the n-epi. This explains the influence of the layout parameter *blnend* on the breakdown voltage (table 5.3). Furthermore, the breakdown is caused by a reach-through of the depletion layer in the pdrift + pwell to the n^+/pwell junction, creating a current path for the electrons from cathode to anode. Hence the occurrence of the pwell, without which the reach-through happens at even lower voltages. One could try to increase this reach-through voltage by introducing a new pbuffer layer.

Table 5.3 Breakdown voltage (V_{br}) for several values of the layout parameter blnend.

blnend (μm)	V_{br} (V)
-1.5[a]	31
0	43
1.5	40
3	37

[a]BLN diffuses up to under the cathode

5.6.2 On-state

The most interesting characteristic of the pLIGBT is that the on-state resistance is no longer determined by the mobility of the holes alone. Contrary to pDMOS devices —which have inherently an on-state resistance that is three times lower than that of nDMOS devices with the same breakdown voltage— pLIGBTs have a conductivity modulated pdrift region during forward conduction. This means that the resistivity of the pdrift region is no longer determined by the doping level of the holes. As a consequence, the on-state resistance is comparable to that of a nLIGBT since the total resistance of a high-voltage device is for a great part determined by the resistivity of the drift region.

Figure 5.39 plots the output characteristics of both a nLIGBT and an pLIGBT, showing that the current level of the pLIGBT is approximately 20 – 25 % lower than that of the nLIGBT for the entire V_{ak} range. Note that for the pLIGBT, the simulation is carried out with $V_{suba} = -80\,\mathrm{V}$ (substrate to anode voltage), $V_{ga} = -3.3\,\mathrm{V}$ (gate to anode voltage), and a cathode voltage V_{ka} that is swept between 0 and $-28\,\mathrm{V}$. This corresponds to a real situation where $V_{asub} = 80\,\mathrm{V}$, $V_{gsub} = 76.7\,\mathrm{V}$, and V_{ksub} is swept between 80 and 52 V.

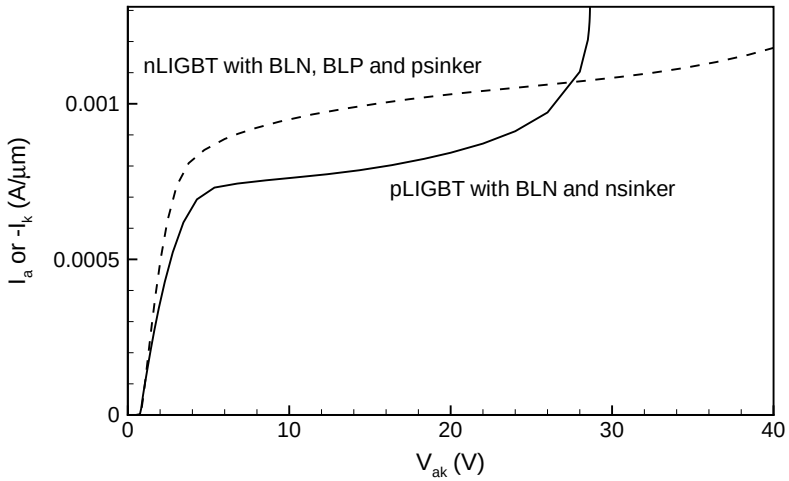


Figure 5.39 The output characteristic of a pLIGBT (at $V_{ga} = -3.3\,\mathrm{V}$) with $blnend = 0\,\mu\mathrm{m}$ compared with a nLIGBT (at $V_{gk} = 3.3\,\mathrm{V}$) with a floating BLN, BLP and psinker.

Power dissipation

Since the on-state characteristic of the pLIGBT has a much higher current level than that of a pLDEMOS, both characteristics are compared. However, this has to be done by applying the same method as described for the comparison of a nLIGBT with a nVDEMOS, since both devices have a different pitch and since the pLIGBT has a forward voltage drop($V_{ka} \sim -0.7\,\mathrm{V}$).

The current density per unit area is plotted against the power dissipation per unit area in figure 5.40 and shows that the break-even point is at a relatively low value: when the power dissipation per unit area is allowed to go beyond $1\,\mathrm{W/mm^2}$, the pLIGBT conducts more current per unit area than the pLDEMOS (from $2.4\,\mathrm{A/mm^2}$ on).

Therefore it is believed that the pLIGBT is likely to beat the pLDEMOS as this break-even point is at about a three times lower value than that of the n-type devices. And when the temperature increases, the break-even point is at an even lower value.

Yet the pLDEMOS device is an 80 V device—whereas the p-type LIGBT is a 40 V device. However, it is believed that the reach-through of the pLIGBT can be increased without deteriorating the on-state characteristics.

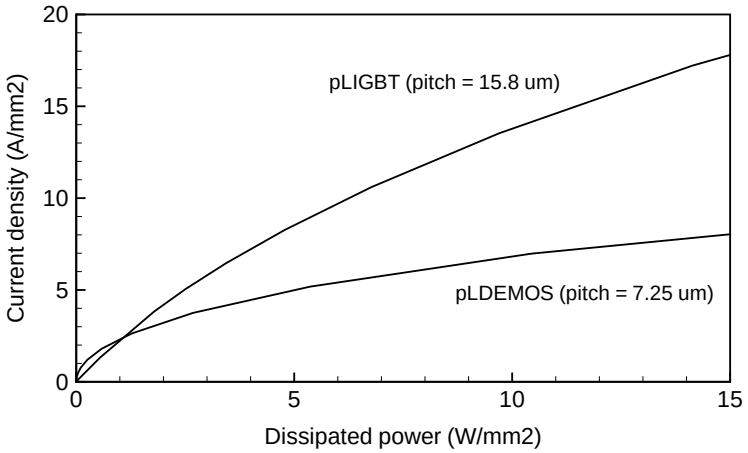


Figure 5.40 Current density versus power dissipation (at $V_{ga} = -3.3\,\mathrm{V}$) for an pLIGBT with $blnend = 0\,\mu\mathrm{m}$ compared with a typical pLDEMOS (at $V_{gs} = -3.3\,\mathrm{V}$).

5.6.3 Substrate current

Another distinguishing feature of the pLIGBT is the extremely low substrate to anode current ratio. Figure 5.41 plots this ratio at $V_{ga} = -3.3\,\mathrm{V}$ as a function of V_{ak} for several values of *blnend*. As long as the BLN is present under the cathode drift region ($blnend = 1.5$ and $0\,\mu\mathrm{m}$), there is no problem with substrate current whatsoever as the BLN is still big enough to kill the parasitic pnp towards the psubstrate. When the BLN is pushed away from under the pdrift region, the doping level of the base of this parasitic bipolar decreases drastically (from the BLN doping level to the n-epi), resulting in a rapid increase of the substrate to anode current ratio.

The substrate current behaviour of the pLIGBT can also be explained through the use of an equivalent circuit. In figure 5.42, it is clearly seen that the four layer structure at the cathode side can not work as a thyristor, since the psubstrate is always at the lowest potential. The pnp can however conduct current as soon as $V_{xy} \approx -0.7\,\mathrm{V}$. This happens in the case of large *blnend* values, as then R_1 increases, with a potential drop at x as a consequence. In the other cases (small and negative *blnend* values), the pnp remains in cut-off, which explains the extremely low substrate currents in figure 5.41.

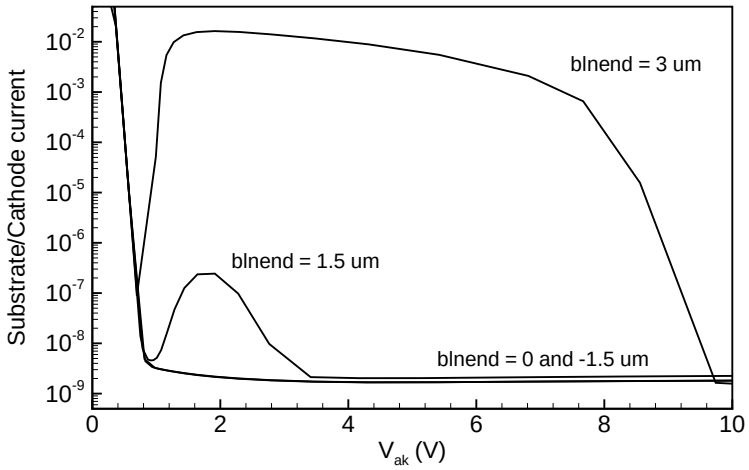


Figure 5.41 Substrate to cathode current ratio (at $V_{ga} = -3.3\,\mathrm{V}$) as as function of V_{ak} for pLIGBTs with different *blnend* values.

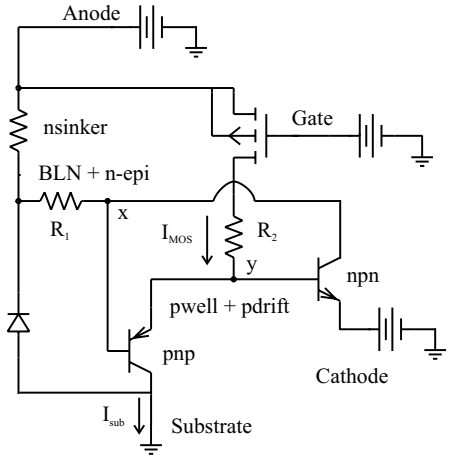


Figure 5.42 Equivalent circuit for an pLIGBT.

5.7 Integrated Vertical IGBTs

This section studies the possibility of integrating vertical IGBTs. To that end, a highly doped buried layer—serving as the anode contact situated under the device instead of next to it—and a highly doped sinker linking this "buried" contact back to the surface, are needed. For the n-type IGBT, this means a BLP with psinker; for the p-type IGBT, a BLN with nsinker. Since we are working on a p-type substrate, isolation is needed between the BLP and the substrate for the n-type IGBT. Fortunately, the BLN has been designed in such a way that it can serve this purpose. Keeping this in mind, 2 of the most common approaches used for the vertical IGBT—the punch-through (PT) and the non punch-through (NPT) IGBT—are presented in this section.

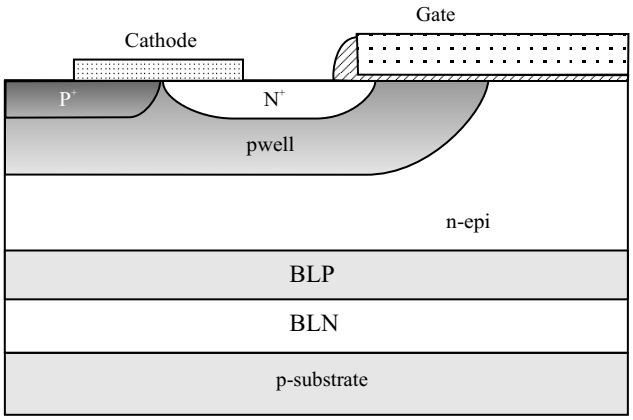


Figure 5.43 An integrated non punch-through IGBT.

Figure 5.43 depicts the non punch-through IGBT device. A non-punch-through vertical device should be designed as such that at forward blocking, the pwell/n-epi depletion region does not punch through to the n-epi/BLP junction. For discretes, the n-epi thickness can be several hundred micrometers. In the I3T80 technology, however, the free epi thickness (i.e., the net n-epi thickness between the pwell and BLP junctions) is in the order of a few micrometers. This means that this type of IGBT almost immediately reaches through (at $V_{ak} \approx 10 - 20\,\mathrm{V}$).

Since the n-epi thickness can not be changed in the I3T80 technology, other approaches are needed. A solution might be the so-called punch-through devices, as shown in figure 5.44, where the depletion region coming from the n-epi/pwell junction at forward blocking is stopped at the nbuffer. A new layer would be needed, implanted after the BLP

anneal and after 4 to 5 µm of epi growth, in order to ensure a clean n-epi/nbuffer transition without a part of the out-diffusion of the BLP between them. This would leave 1 to 2 µm free epi. One easily calculates that breakdown voltages of 60 V and more are impossible (with a maximum electrical field of 40 V/µm).

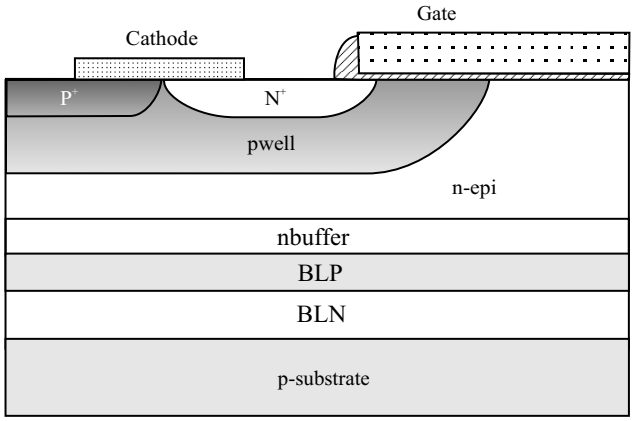


Figure 5.44 A vertical punch-through IGBT.

Other, more exotic, solutions to the quasi-vertical approach might be thought of. See e.g., figure 5.45, a 3D approach, where the BLN acts as a buffer between substrate and pwell/n-epi junction and the BLP as a current path along the width of the device. This could also be a so-called anode-shorted device, where the n-epi is no longer floating but shorted to the anode (i.e., BLN and BLP contacts shorted).

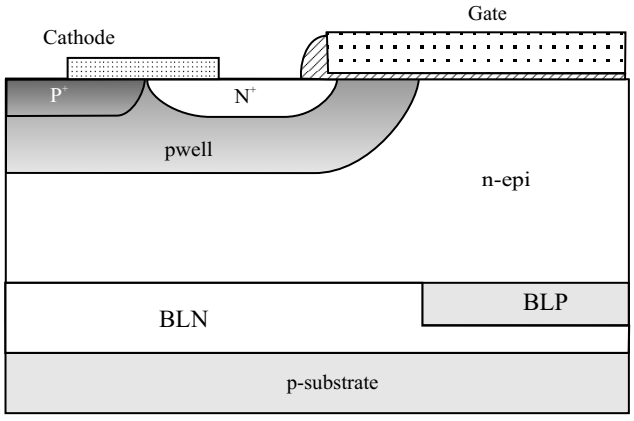


Figure 5.45 A 3D integrated anode-shorted IGBT.

Absolutely impossible to integrate is the integrated p-type IGBT. Not only should there be a p-type analogue of the n-epi; there should also be an alternative for the p-substrate. As can be seen in figure 5.46), when the BLN/p-type junction is forward biased (needed for conduction), then the BLN/p-substrate junction is also forward biased. This means that integrating this device in I3T80 is simply not possible.

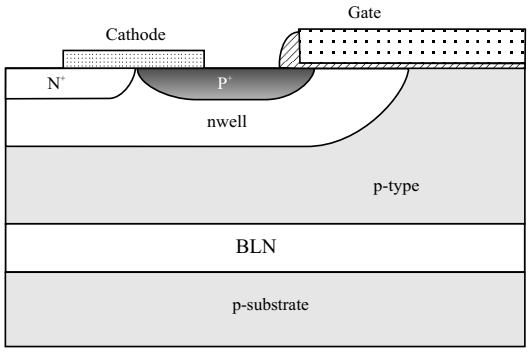


Figure 5.46 A p-type quasi-vertical non punch-through IGBT.

There were no simulations carried out on these integrated quasi-vertical IGBTs because of the above explained limitations, the restrictions concerning changing the process flow, and the more straightforward and easier to integrate lateral alternative.

5.8 nVDEMOS versus standard nLIGBT versus nLIGBT with BLN and BLP: measurements

All three different types of the nLIGBT (with BLN, with BLP, and with both buried layers) have been put on testchip, which consists of 7 frames, each containing 18 devices. For each type of nLIGBT several layout variations around the 'standard' layout (see above (in μm): $pso = 3$, $ps = 3$, $katoga = 0.4$, $sw = 0.8$, $pf = 0$, $x = 1$, $y = 0.8$, $t = 4.5$, $z = 1$, $nw = 1.5$, $aw = 1$, and $antofi = 0.2$) have been processed. We give a short overview of the main results in this section.

5.8.1 Breakdown

The measurements on the blocking capabilities of the nLIGBT devices are very much in line with the TCAD simulations. The standard nLIGBT

with BLP yields a $V_{br} = 80 - 85\,\mathrm{V}$ for the device with *blpstop* $= 8\,\mu\mathrm{m}$, whereas the simulation gave 90 V. The peak in breakdown voltage as function of *blpstop* (see table 5.2) is also confirmed by measurements.

The nLIGBT with BLN indeed yields a device with lower breakdown voltages, but unfortunately, due to a design fault on the frames, further results on these devices are not at our disposal. Finally, the nLIGBT with BLN and BLP, and with the standard layout, has a $V_{br} = 75\,\mathrm{V}$, which also corresponds with TCAD simulations.

5.8.2 On-state, saturation and latch-up

The pdrift and the psinker play an important role in these devices. An example is given in figure 5.47, where the layout parameter *ps* has been varied. For the smallest value ($ps = 2\,\mu\mathrm{m}$), the latch-up occurs at 38 V when the gate is fully open. Unfortunately, the psinker influences the channel for this value, which deteriorates the on-state. For the larger values ($ps = 3$ and $3.5\,\mu\mathrm{m}$), the latch-up decreases, but the on-state gets better. The optimum as has been predicted by TCAD simulations (i.e., the 'standard' layout; that is, $ps = 3\,\mu\mathrm{m}$), is confirmed by measurements. Yet there are important discrepancies between simulation

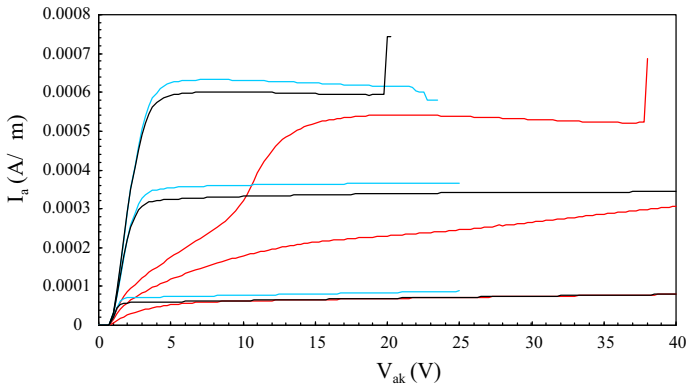


Figure 5.47 *Measured* anode currents at V_{gk} =1.1, 2.2 and 3.3 V for the nLIGBT with BLN and BLP with three different values for *ps* (in μm): 2.0 (red), 3.0 (blue) and 3.5 (black).

and measurement (e.g., between figures 5.24 and 5.47), which indicate that further calibration of the input deck is needed. Note as well that all simulations have been done without taking into account thermal effects and thus are not able to track the self-heating effects that have been measured.

Figure 5.48 plots the *measured* output characteristics of the nVDEMOS and the nLIGBT with BLN and BLP (and with the standard layout). The large values for the saturation currents of the nLIGBT compared to the nVDEMOS are striking, yet do not give a complete picture. As has been explained above, the comparison of the power

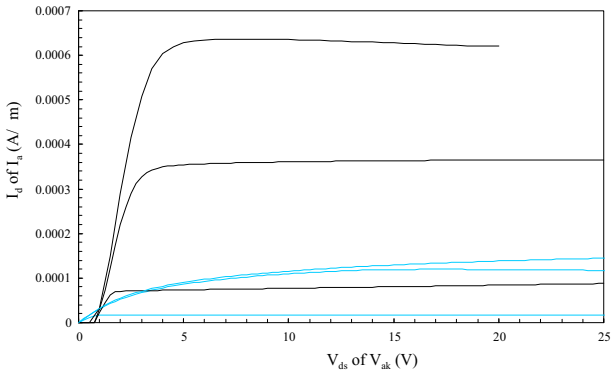


Figure 5.48 *Measured* output characteristics at V_{gk} =1.1, 2.2 and 3.3 V for the nLIGBT with BLN and BLP compared with a typical nVDEMOS (blue, at V_{gk} =1, 2 and 3.3 V).

dissipation per unit area at a certain current density is a more correct criterion (figure 5.49). Again the measurements are very much in line with the simulations and reveal that the nLIGBT can beat the nVDEMOS if a power dissipation of 3 W/mm^2 is allowed.

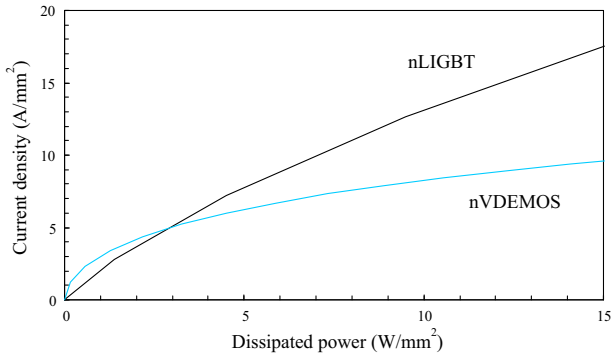


Figure 5.49 *Measured* current density versus power dissipation at V_{gk} =3.3 V for the nLIGBT with BLN and BLP compared with a typical nVDEMOS (at V_{gs} =3.3 V).

5.8.3 Substrate current behaviour

The most important asset of the nLIGBT with BLN and BLP is the suppression of the substrate current. Figure 5.50 shows that the actual devices perform better than what was predicted by TCAD simulations. It has even been observed that the peak in the substrate current can be avoided when the pdrift is drawn under the channel as well ($pf > 0\,\mu m$, see figure 5.13).

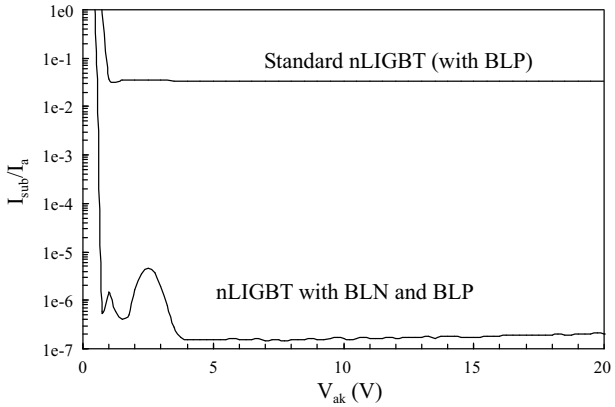


Figure 5.50 *Measured* substrate to anode current ratio at $V_{gk} = 3.3\,V$ for the nLIGBT with BLN and BLP, compared to the standard nLIGBT (only BLP present with *blpstop* $= 8\,\mu m$).

5.9 Conclusions

We have integrated several new IGBT structures into the I3T80 technology. Simulations and measurements reveal that the best approach for the n-type is a lateral IGBT with BLP and psinker on top of a floating BLN. This device has an acceptable breakdown voltage ($70-75\,V$), has a large SOA, has a saturation current 4 to 6 times the saturation current of a nVDEMOS, has the lowest substrate current when compared to other LIGBT layouts ($I_{sub}/I_a < 10^{-6}$!), and has a fast turn-off. We believe that hereby a nLIGBT is developed with acceptable characteristics.

Concerning the p-type LIGBTs, a similar device is created with an on-state behaviour comparable (in quality and quantity !) to the nLIGBT, with good SOA, and with remarkably low substrate currents. Unfortunately, the breakdown voltage of this device is $\sim 42\,V$, but this is entirely due to the limited capability of the pwell + pdrift as pbuffer on the cathode side of this device.

Any comparison with LIGBTs presented in literature is difficult, for most recent articles on LIGBTs work in an SOI technology with forward blocking voltages of several hundred volts. However, we have compared the LIGBTs with the existing I3T80 integrated nVDEMOS devices. The LIGBT can beat the nVDEMOS in terms of current per unit area if the power dissipation per unit area is allowed to reach $\sim 1 - 3\,\mathrm{W/mm^2}$ for the n-type devices and $\sim 1\,\mathrm{W/mm^2}$ for the p-type devices.

References

[AU01] G. Amaratunga and F. Udrea. Power Devices for High Voltage Integrated Circuits: New Device and Technology Concepts. In *Semiconductor Conference*, volume 2, pages 441–448, 2001.

[BAG+82] B.J. Baliga, M.S. Adler, P.V. Gray, R. Love, and N. Zommer. The Insulated Gate Rectifier (IGR): A New Power Switching Device. In *IEEE International Electron Devices Meeting Digest*, pages 264–267, 1982.

[Bal79] B.J. Baliga. Enhancement- and Depletion Mode Vertical-Channel MOS Gated Thyristors. *Electronics Letters*, 15(20):645–647, September 1979.

[Bal92] B.J. Baliga. *Modern Power Devices*. Krieger, 1992.

[BGG99] V. Benda, J. Gowar, and D.A. Grant. *Power Semiconductor Devices*. John Wiley & Sons, 1999.

[PS80] J. D. Plummer and B.W. Scharf. Insulated-Gate Planar Thyristors: I—Structure and Basic Operation. *IEEE Transactions on Electron Devices*, 27(2):380–387, February 1980.

[RGGN83] J.P. Russell, A.M. Goodman, L.A. Goodman, and J.M. Neilson. The COMFET — A New High Conductance MOS-Gated Device. *IEEE Electron Device Letters*, EDL-4(3):63–65, March 1983.

[Sze81] S.M. Sze. *Physics of Semiconductor Devices*. John Wiley & Sons, 1981.

[Tih80] J. Tihaji. Functional Integration of Power MOS and Bipolar Devices. In *IEEE International Electron Devices Meeting*, pages 75–78, 1980.

6 Synopsis

6.1 Overview Main Results

After two short general introductory chapters and a quick survey of the TCAD simulation and calibration work, we have integrated both the power MOS and IGBT into an existing, standard CMOS, junction isolated technology.

We concluded that the best combination of a n-type and a p-type power MOS in an $80 - 100\,\mathrm{V}$ technology (e.g., the I3T80 technology of AMIS) is a nVDMOS and a pLDMOS. Although the nLDMOS has a better $V_{br} - R_{on,sp}$ trade-off than the nVDMOS, we chose the latter, and this for three reasons. First of all, the vertical device has a large SOA without using (costly) new layers, which are needed when the same SOA should have to be obtained in the nLDMOS. Secondly, the nVDEMOS is by nature a floating device. In theory, the nLDEMOS can be made floating without loss of performance as well, but this is technically harder to realize. Thirdly, the nVDMOS is easy to combine with the pLDMOS, which is not possible for the best RESURF nLDMOS devices.

Figures 6.1 and 6.2 show the existing nVD(E)MOS and pLDMOS devices together with the competitor's transistors and with the silicon limits. The nVD(E)MOS was developed at AMIS, while the pLDEMOS was developed at TFCG. Although the n-type device in the I3T80 technology is a vertical device, whereas all the competitor's devices are lateral, it has a competitive $V_{br} - R_{on,sp}$. Note that the best devices on figure 6.1 are non-floating, with the one important exception of the 63 V "FRESURF" device by Motorola ($0.35\,\mu\mathrm{m}$). Figure 6.2 shows that the I3T80 pLDEMOS is amongst the best p-type devices ever made.

For the IGBT devices, we have succeeded in integrating both a n-type and a p-type LIGBT, which have—to the best of our knowledge—never been described in literature before. We managed to suppress the substrate currents in both devices, to create a large SOA, and to guarantee

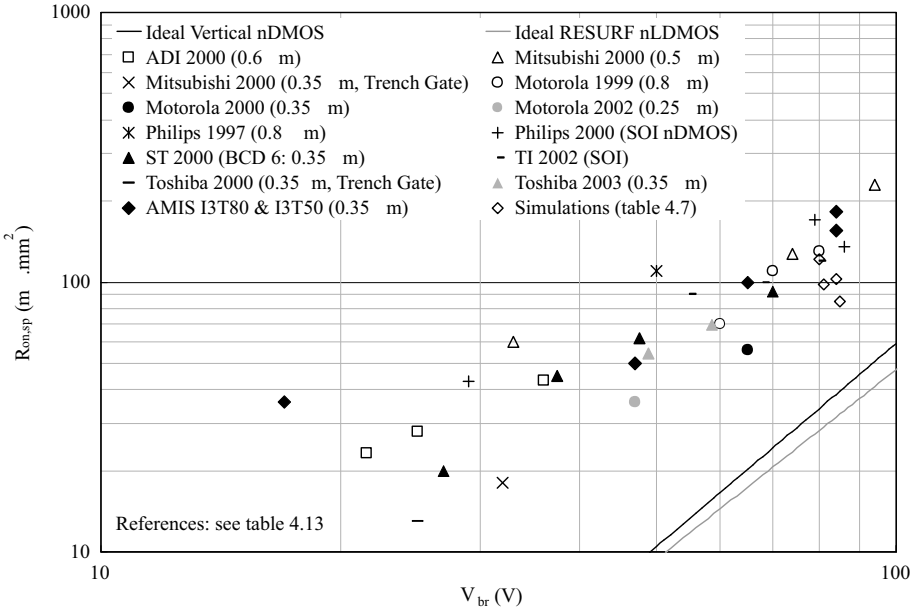


Figure 6.1 Benchmark and competitors for the nDMOS devices.

fast switching. Moreover, both devices are floating, make use of existing layers only, and have a current carrying capability per unit silicon area that is competitive with the DMOS devices.

To conclude, the author of this work was also author or co-author of the following articles: [BMVD01], [AHI$^+$01a], [HAI$^+$01], [AHI$^+$01b], [BVD$^+$02], [MBD$^+$02], [HVB$^+$02], [AHV$^+$03], and [BDM04].

6.2 Future Research

The nVDMOS and pLDEMOS are commercially available, yet a further characterization of the thermal behaviour of these devices is ongoing (especially the nVDEMOS as huge driver). The nLIGBT thus has to be further optimized and characterized as well. Furthermore, a thorough study of the feasible power dissipation per unit area or per package is necessary.

A further calibration of the TCAD input deck is also a necessity, especially for the unusual BLP on BLN structure and for the psinker (also in lateral direction !), for it is now used as a working part of a device. This second calibration round could then be used to simulate these LIGBTs in SPICE-like circuits, simulating e.g., inductive turn-off of a nLIGBT or a nVDMOS. These so-called mixed mode simulations

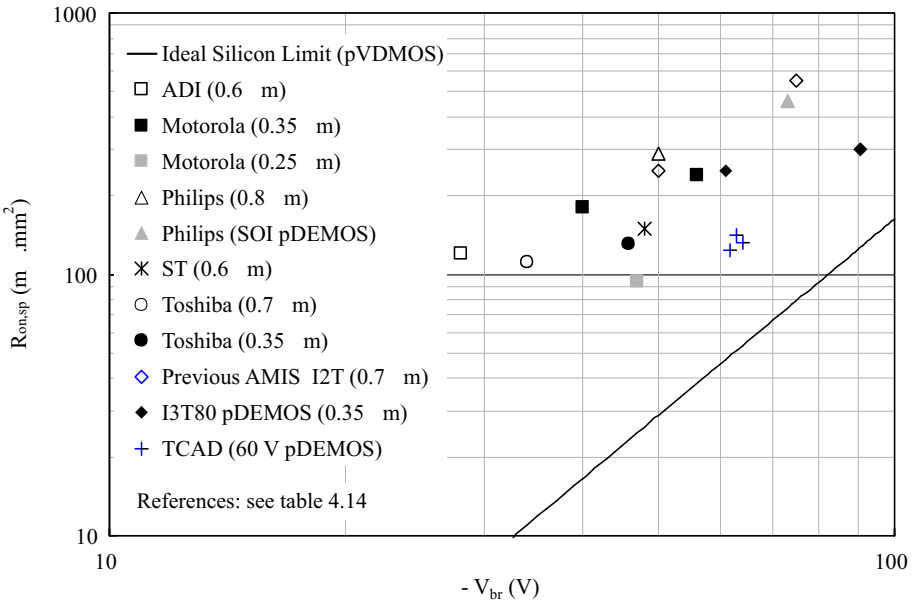


Figure 6.2 Benchmark and competitors for the pDMOS devices.

could also be applied to energy capability simulations or ESD simulations, thereby trying to design robust devices. These simulations should use the thermodynamical or hydrodynamical models, or even be performed in 3D (which is still in the future), in order to be able to track the self-heating effects. Recent device simulators make it even possible to simulate hot carrier degradation, which could be linked to stress measurements, something that has not yet been done for LIGBTs.

On the device side, new types of power devices could be thought up and tried out (e.g., the *I*nsulated *B*ase *T*ransistor [PMS86], the *L*ateral *I*nsulated *G*ate *p*-i-n *T*ransistor [Hua96], the *L*ateral *I*nversion *L*ayer *E*mitter *T*ransistor [UAH+96], ...), one could try to integrate thyristors (e.g., the *L*ateral *C*onductivity *M*odulated *T*hyristor [LKO+02]), superjunctions and/or trench gates, and power devices in smaller technologies (to date, the smallest smart power technology is 0.25 µm [ZPK+03]) or devices on SOI could be considered. In the distant future, new materials (SiC, GaP...) could be tried out on smart power technologies.

References

[AHI+01a] C. Anghel, N. Hefyene, A.M. Ionescu, M. Vermandel, B. Bakeroot, and J. Doutreloigne. Investigations and Phys-

ical Modelling of Saturation Effects in Lateral DMOS Transistor Architectures Based on the Concept of Intrinsic Drain Voltage. In *European Solid-State Device Research Conference*, pages 399–402, 2001.

[AHI+01b] C. Anghel, N. Hefyene, A.M. Ionescu, M. Vermandel, B. Bakeroot, and J. Doutreloigne et al. Physical Modelling Strategy for (Quasi-)Saturation Effects in Lateral DMOS Transistor Based on the Concept of Intrinsic Drain Voltage. In *Semiconductor Conference*, pages 417–420, 2001.

[AHV+03] C. Anghel, N. Hefyene, M. Vermandel, B. Bakeroot, J. Doutreloigne, and R. Gillon et al. Electrical Characterization of High Voltage MOSFETs Using MESDRIFT. In *Semiconductor Conference*, pages 257–260, 2003.

[BDM04] B. Bakeroot, J. Doutreloigne, and P. Moens. A New Substrate Current Free nLIGBT for Junction Isolated Technologies. In *European Solid-State Device Research Conference*, 2004, accepted for publication.

[BMVD01] B. Bakeroot, P. Moens, M. Vermandel, and J. Doutreloigne. Using Adaptive RESURF Technique and Field Plate Working to Improve the Safe Operating Area of n-type Drain Extended MOS. In *Conference on Modeling and Simulation of Microsystems*, pages 468–501, 2001.

[BVD+02] B. Bakeroot, M. Vermandel, J. Doutreloigne, P. Moens, and D. Bolognesi. Cost Effective Implementation of a 90 V RESURF p-Type Drain Extended MOS in a 0.35 µm Based Smart Power Technology. In *European Solid-State Device Research Conference*, pages 291–294, 2002.

[HAI+01] N. Hefyene, C. Anghel, A.M. Ionescu, M. Vermandel, B. Bakeroot, and J. Doutreloigne et al. An Experimental Approach for Bias-Dependent Drain Series Resistances Evaluation in Asymmetric HV MOSFETs. In *European Solid-State Device Research Conference*, pages 403–406, 2001.

[Hua96] A.Q. Huang. Lateral Insulated Gate p-i-n Transistor (LIGPT)—A New MOS Gate Lateral Power Device. *IEEE Electron Device Letters*, 17(6):297–299, June 1996.

[HVB+02] N. Hefyene, E. Vestiel, B. Bakeroot, C. Anghel, A.M. Ionescu, and R. Gillon. Bias-Dependent Drift Resistance Modelling for Accurate DC and AC Simulation of Asymmetric HV-MOSFET. In *Simulation of Semiconductor Processes and Devices*, pages 203–206, 2002.

[LKO+02] Y.-S. Lee, S.-S. Kim, J.-K. Oh, Y.I. Choi, and M.-K. Han. A New Lateral Conductivity Modulated Thyristor With Current Saturation and Low Turn-off Time. In *Symposium on Power Semiconductor Devices & ICs*, pages 113–116, 2002.

[MBD+02] P. Moens, D. Bolognesi, L. Delobel, D. Villanueva, K. Reynders, A. Lowe, G. Van Herzeele, M. Tack, and B. Bakeroot. Future Trends in Intelligent Interface Technologies for 42 V Battery Automotive Applications. In *European Solid-State Device Research Conference*, pages 287–290, 2002.

[PMS86] Z. Parpia, J.G. Mena, and C.A.T. Salama. A Novel CMOS-Compatible High-Voltage Transistor Structure. *IEEE Transactions on Electron Devices*, 33(12):1948–1952, December 1986.

[UAH+96] F. Udrea, G.A.J. Amarantunga, J. Humphrey, J. Clark, and A.G.R. Evans. The MOS Inversion Layer as a Minority Carrier Injector. *IEEE Electron Device Letters*, 17(9):425–427, September 1996.

[ZPK+03] R. Zhu, V. Parthasarathy, V. Khemka, A. Bose, T. Roggenbauer, and G. Lee et al. A 0.25 Micron Smart Power Technology Optimized For Wireless and Consumer Applications. In *Symposium on Power Semiconductor Devices & ICs*, pages 178–181, 2003.

A List of Basic Symbols

Symbol	Despcription	Unit
c	Heat capacity	$\mathrm{J/(K\,cm^3)}$
C	Capacitance	F
D	Diffusion constant	$\mathrm{cm^2/s}$
D_n	Electron diffusion constant	$\mathrm{cm^2/s}$
D_p	Hole diffusion constant	$\mathrm{cm^2/s}$
$\mathbf{E}$	Electric field	V/cm
$\mathbf{F}$	Force	N
$\hbar$	Planck constant	J s
G_n	Electron generation rate per unit volume	$\mathrm{cm^{-3}/s}$
G_p	Hole generation rate per unit volume	$\mathrm{cm^{-3}/s}$
H	Heat generation per unit volume	$\mathrm{J/(s\,cm^3)}$
I	Current	A
J	Current density	$\mathrm{A/cm^2}$
J_n	Electron current density	$\mathrm{A/cm^2}$
J_p	Hole current density	$\mathrm{A/cm^2}$
k	Boltzmann constant	J/K
l	Length	cm
m_0	Electron rest mass	kg
n	Density of free electrons	$\mathrm{cm^{-3}}$
$n_{i,eff}$	Effective intrinsic density	$\mathrm{cm^{-3}}$
N_{A^-}	Acceptor impurity density	$\mathrm{cm^{-3}}$
N_{D^+}	Donor impurity density	$\mathrm{cm^{-3}}$
p	Density of free holes	$\mathrm{cm^{-3}}$
$\mathbf{p}$	Carrier momentum	kg cm/s
P_n	Absolute electron thermoelectric power	V/K
P_p	Absolute hole thermoelectric power	V/K
q	Magnitude of electronic charge	C
$\mathbf{r}$	Carrier position	cm

Symbol	Despcription	Unit
R_n	Electron recombination rate per unit volume	$\mathrm{cm}^{-3}/\mathrm{s}$
R_p	Hole recombination rate per unit volume	$\mathrm{cm}^{-3}/\mathrm{s}$
t	Time	s
T	Absolute temperature	K
T_L	Absolute lattice temperature	K
T_n	Absolute electron temperature	K
T_p	Absolute hole temperature	K
$\mathbf{v}$	Carrier velocity	cm/s
$v_{n,sat}$	Electron saturation velocity	cm/s
V	Voltage	V
ϵ_s	Semiconductor permittivity	F/cm
κ	Thermal conductivity	$\mathrm{W}/(\mathrm{m\,K})$
μ_n	Electron mobility	$\mathrm{cm}^2/(\mathrm{V\,s})$
μ_p	Hole mobility	$\mathrm{cm}^2/(\mathrm{V\,s})$
ρ	Space charge density	C/cm^3
ϕ_n	Electron quasi-Fermi potential	V
ϕ_p	Hole quasi-Fermi potential	V
ψ	Potential	V
Ψ	Wave function	
Ω	Ohm	Ω

B List of Abbreviations

1D, 2D or 3D One, Two or Three Dimension(s)

AMIS American MIcroelectronics Semiconductor

BLN Buried Layer of N-type

BLP Buried Layer of P-type

BRT Base Resistance controlled Thyristor

CMOS Complementary MOS (see MOSFET)

CPU Central Processing Unit

DEMOS Drain Extended MOS (see MOSFET)

DMOS Double diffused MOS (see MOSFET)

ESD ElectroStatic Discharge

EST Emitter Switched Thyristor

FCD Field Controlled Diode

FCT Field Controlled Thyristor

GATT Gate-Assisted Turn-off Thyristor

GTO Gate Turn-Off thyristor

IBT Insulated Base Transistor

IGBT Insulated Gate Bipolar Transistor

ISE Integrated Systems Engineering

JBS Junction Barrier Schottky rectifier

JFET Junction Field Effect Transistor

LCMT Lateral Conductivity Modulated Thyristor

LDD Lightly Doped Drain

LDEMOS Lateral Drain Extended MOS (see MOSFET)

LDMOS Lateral Double diffused MOS (see MOSFET)

LIGBT Lateral IGBT (see IGBT)

LILET Lateral Inversion Layer Emitter Transistor

MOSFET Metal-Oxide-Semiconductor Field Effect Transistor

MPS Merged p-i-n/Schottky rectifier

QVDEMOS Quasi-Vertical Drain Extended MOS (see MOSFET)

QVDMOS Quasi-Vertical Double diffused MOS (see MOSFET)

RESURF REduced SURface Field

SCR Silicon Controlled Rectifier

SEM Scanning Electron Microscopy

SIMS Secondary-Ion Mass Spectroscopy

SIT Static Induction Transistor

SITh Static Induction Thyristor

SOA Safe Operating Area

SOI Silicon On Insulator

SPICE Simulation Program with Integrated Circuit Emphasis

SRP Spreading Resistance Profiling

TCAD Technology Computer Aided Design

TED Transient Enhanced Diffusion

TLP Transmission Line Pulsing

VDEMOS Vertical Drain Extended MOS (see MOSFET)

VDMOS Vertical Double diffused MOS (see MOSFET)

C Calculating Breakdown and Punch-through

C.1 Breakdown of an abrupt n–p$^+$ diode

Consider an abrupt, parallel-plane n–p$^+$ junction in which the n-type doping level (N_{epi}) is very low compared to the p-type doping level (N_{sub}). When a reverse bias voltage is applied to such a diode, the depletion layer extends only into the n side as a result of the very high doping level on the p side. Therefore, Poisson's equation needs to be solved only for the n side:

$$\frac{dE}{dx} = \frac{\rho(x)}{\epsilon_s} = \frac{qN_{epi}}{\epsilon_s} \quad for \quad 0 \leq x \leq W_n \tag{C.1}$$

where $\rho(x)$ is the charge in the depletion layer on the n side due to ionized donors, ϵ_s is the dielectric constant of silicon, q is the electronic charge, and W_n is the end of the depletion layer in the n side region. Integration of (C.1) and use of the boundary conditon that the electric field at the end of the depletion layer is zero ($E(x = W_n) = 0$) results in the electric field distribution:

$$\int_0^{E(x)} dE = \int_{W_n}^{x} \frac{qN_{epi}}{\epsilon_s} dx'$$

$$E(x) = \frac{qN_{epi}}{\epsilon_s}(x - W_n). \tag{C.2}$$

Integration of (C.2) with the boundary condition that $V(x = W_n) = V_a$ gives the voltage distribution:

$$\begin{aligned} -\frac{dV}{dx} = E(x) &= \frac{qN_{epi}}{\epsilon_s}(x - W) \\ -\int_{V(x)}^{V_a} dV &= \frac{qN_{epi}}{\epsilon_s}\int_x^{W_n}(x' - W_n)dx' \\ V(x) &= V_a - \frac{qN_{epi}}{2\epsilon_s}\,(W_n - x)^2\,. \end{aligned} \tag{C.3}$$

Substituting the values at breakdown $V_a = V_{br}$ and $W_n = W_{br}$ in (C.3) and solving for $x = 0$ yields

$$V_{br} = \frac{qN_{epi}}{2\epsilon_s}\,W_{br}^2. \tag{C.4}$$

From (C.2), it is clear that the maximum absolute value of the electric field (E_m) occurs at $x = 0$. At breakdown, this value is therefore called the critical electric field E_{crit}:

$$E_{crit} = |E(x = 0)| = \frac{qN_{epi}}{\epsilon_s}\,W_{br}. \tag{C.5}$$

Combining (C.5) and (C.4) gives an expression of the breakdown voltage as a function of the critical electric field:

$$V_{br} = \frac{\epsilon_s}{2qN_{epi}}\,E_{crit}^2. \tag{C.6}$$

From the theory of impact ionization, it is known that breakdown occurs as soon as the *ionization integral* is equal to 1 (e.g., [Bal92, pp. 63–66]):

$$\int_0^{W_{br}} \alpha\; dx = 1 \tag{C.7}$$

where W_{br} is the depletion layer width at breakdown, and α an approximation of the impact ionization coefficient of both the electrons and holes (see also (2.23)):

$$\alpha_n \approx \alpha_p \approx \alpha = A\;|E|^7 \tag{C.8}$$

where A is Fulop's ionization rate [Ful67] equal to $1.8 \times 10^{-35}\,\mathrm{cm^6V^{-7}}$. Substituting (C.8) in (C.7) gives:

$$1.8 \times 10^{-35}\;\int_0^{W_{br}} |E(x)|^7\;dx = 1 \tag{C.9}$$

from which, using (C.2), the following expression for the depletion layer width at breakdown W_{br} can be found:

$$W_{br} = \left(\frac{8\epsilon_s^7}{1.8 \times 10^{-35} q^7}\right)^{1/8} \left(N_{epi}\right)^{-7/8}. \tag{C.10}$$

A similar expression for the depletion layer width at breakdown for a parallel-plane n$^+$/p junction has been used in figure 2.5. In the same figure, the breakdown voltage is also given as a function of the doping level, which can be easily obtained by substituting (C.10) in (C.4):

$$V_{br} = \left(\frac{\epsilon_s^3}{2q^3\, 1.8 \times 10^{-35}}\right)^{1/4} \left(N_{epi}\right)^{-3/4}. \tag{C.11}$$

C.2 Punch-through of an abrupt n$^+$–n–p$^+$ diode

When the doping level in the n side of the abrupt n–p$^+$ diode suddenly increases to a very high value, then the extension of the depletion layer with increasing reverse bias is stopped at the n–n$^+$ junction. Of course, provided that the depletion layer width at breakdown of the n–p$^+$ diode is larger than the thickness of the low doped n region (t_{epi}). For this so-called punch-through diode, Poisson's equation is only solved in the lightly doped n region.

When the critical electric field is reached at $x = 0$, breakdown occurs in the punch-through diode. At that moment, the electric field at the n–n$^+$ junction is not zero, but given by (C.2) with $x = t_{epi}$:

$$E_{t_{epi}} = \frac{qN_{epi}}{\epsilon_s}(t_{epi} - W_{br}), \tag{C.12}$$

where W_{br} is the depletion layer thickness as defined in the previous section and given by (C.10). The punch-through breakdown voltage can be calculated using (C.12) and the fact that the potential drops in both the p$^+$ and the n^+ regions are neglected:

$$V_{pt} = \frac{1}{2}\left(|E_{t_{epi}}| + E_{crit}\right) \times t_{epi}. \tag{C.13}$$

Using (C.12) and (C.5) in (C.13) results in

$$\begin{aligned} V_{pt} &= \frac{1}{2}\left(\frac{qN_{epi}}{\epsilon_s} W_{br} + \frac{qN_{epi}}{\epsilon_s} W_{br} - \frac{qN_{epi}}{\epsilon_s} t_{epi}\right) t_{epi} \\ &= \frac{qN_{epi}}{\epsilon_s} W_{br} t_{epi} - \frac{qN_{epi}}{2\epsilon_s} t_{epi}^2. \end{aligned} \tag{C.14}$$

Using (C.10) gives an expression of V_{pt} as a function of the n-epi doping level and thickness:

$$V_{pt} = \left(\frac{8q}{1.8 \times 10^{-35}\epsilon_s}\right)^{1/8} \left(N_{epi}\right)^{1/8} t_{epi} - \frac{q}{2\epsilon_s} N_{epi} t_{epi}^2. \tag{C.15}$$

A similar expression has been used in figure 2.6, where the punch-through breakdown of a n^+–p–p^+ diode is plotted as a function of both the doping level and the thickness of the lightly doped region.

C.3 Breakdown of an abrupt n–p diode

When the doping levels of each side of a diode are comparable to one another, then the depletion layers extend into both the n and p side, and therefore Poisson's equation should be solved in both regions:

$$\frac{dE}{dx} = \frac{-qN_{sub}}{\epsilon_s} \qquad \text{for} \quad W_p \leq x \leq 0 \tag{C.16}$$

$$\frac{dE}{dx} = \frac{qN_{epi}}{\epsilon_s} \qquad \text{for} \quad 0 \leq x \leq W_n. \tag{C.17}$$

Equation (C.16) has been solved in the first section and with the boundary conditions that $E(x = W_p) = 0$ (W_p is the end of the depletion layer in the p side) and $V(x = W_n) = 0$, equation (C.17) yields the following electric field and voltage distributions:

$$E(x) = \frac{-qN_{sub}}{\epsilon_s}(x - W_p) \tag{C.18}$$

$$V(x) = \frac{qN_{sub}}{2\epsilon_s}\ (x - W_p)^2\,. \tag{C.19}$$

As both distributions should be continuous at $x = 0$, (C.2) and (C.18), and (C.3) and (C.19) have to be equal at $x = 0$. This results in the following relationships:

$$\frac{W_n}{W_p} = -\frac{N_{sub}}{N_{epi}} \tag{C.20}$$

$$V_a = \frac{q}{2\epsilon_s}\ \left(N_{sub}W_p^2 + N_{epi}W_n^2\right). \tag{C.21}$$

Substituting (C.20) in (C.21) and solving for W_n and W_p results in:

$$W_n = \sqrt{\frac{2\epsilon_s N_{sub} V_a}{qN_{epi}(N_{sub} + N_{epi})}} \tag{C.22}$$

$$W_p = -\sqrt{\frac{2\epsilon_s N_{epi} V_a}{qN_{sub}(N_{sub} + N_{epi})}}. \tag{C.23}$$

The depletion layer widths at breakdown can be written as function of N_{epi} and N_{sub} alone when solving the ionization integral with Fulop's ionization rate, analogous to what has been done for an abrupt n–p$^+$ diode:

$$\int_{W_{br,p}}^{0} A\left(\frac{qN_{sub}}{\epsilon_s}\right)^7 (x - W_{br,p})^7 dx + \int_{0}^{W_{br,n}} A\left(\frac{qN_{epi}}{\epsilon_s}\right)^7 (W_{br,n} - x)^7 dx = 1$$

with $W_{br,p}$ and $W_{br,n}$ the depletion layer widths at breakdown for the p side and the n side of the junction, respectively. Using (C.20) and solving for $W_{br,p}$ and $W_{br,n}$ yields:

$$W_{br,n} = \left(\frac{N_{sub}}{N_{epi} + N_{sub}} \frac{8}{A}\right)^{1/8} \left(\frac{\epsilon_s}{qN_{epi}}\right)^{7/8} \tag{C.25}$$

$$W_{br,p} = -\left(\frac{N_{epi}}{N_{epi} + N_{sub}} \frac{8}{A}\right)^{1/8} \left(\frac{\epsilon_s}{qN_{sub}}\right)^{7/8}. \tag{C.26}$$

An expression of the breakdown voltage for this non punch-through abrupt pn junction is obtained by substituting equations (C.25) and (C.26) in (C.21):

$$V_{br} = \left(\frac{\epsilon_s^3}{2q^3\, 1.8 \times 10^{-35}}\right)^{1/4} \left(\frac{N_{epi} + N_{sub}}{N_{epi} N_{sub}}\right)^{3/4}. \tag{C.27}$$

C.4 Punch-through of an abrupt n^+–n–p diode

The last example is important for calculating the breakdown voltage in an *ideal RESURF* diode, determined by the vertical punch-through diode present in the RESURF structure (see section 4.2). The thickness of the lowly doped region is defined as the thickness of the total n-epi layer (t_{epi}) thickness minus the thickness of the n^+ region (t_{n^+}). Analogous with the case of the abrupt punch-through n^+–n–p^+ diode, the electric field at the n^+–n junction E_{n^+} is given by

$$E_{t_{n^+}} = \frac{qN_{epi}}{\epsilon_s}\left((t_{epi} - t_{n^+}) - W_{br,n}\right), \tag{C.28}$$

where $W_{br,n}$ is the depletion layer thickness as defined in the previous section. The punch-through breakdown voltage is now given by the sum of the potential drop in both the n side and the p side region:

$$V_{pt} = \frac{1}{2}\left(|E_{t_{n^+}}| + E_{crit}\right) \times \left(t_{epi} - t_{n^+}\right) + \frac{1}{2}E_{crit}|W_{br,p}|. \tag{C.29}$$

Eliminating $E_{crit} = \frac{qN_{sub}}{\epsilon_s}|W_{br,p}|$ and $E_{t_{n^+}}$ in this equation, and assuming charge equality ($W_{br,n}N_{epi} = |W_{br,p}|N_{sub}$) yields:

$$V_{pt} = \frac{qN_{sub}}{\epsilon_s}\left((t_{epi} - t_{n^+})|W_{br,p}| + \frac{|W_{br,p}|^2}{2}\right) - \frac{qN_{epi}}{2\epsilon_s}(t_{epi} - t_{n^+})^2. \tag{C.30}$$

V_{pt} can thus be written as a function of the net epi thickness ($t_{epi} - t_{n^+}$), N_{epi} and N_{sub}, by using this equation together with the expression (C.26) for $W_{br,p}$.

References

[Bal92] B.J. Baliga. *Modern Power Devices*. Krieger, 1992.

[Ful67] W. Fulop. Calculation of Avalanche Breakdown Voltages of Silicon p–n Junctions. *Solid-State Electronics*, 10:39–43, 1967.