## Kurshäfte i Mikroelektronik



## Kjell Jeppson

2012 Chalmers tekniska högskola Institutionen för mikroteknologi och nanovetenskap

Kurshäfte i

# Mikroelektronik

## Kjell Jeppson

2012

Chalmers tekniska högskola Institutionen för mikroteknologi och nanovetenskap

Omslagsbild: Mikrovågstransistor av HEMT-typ, d.v.s. High Electron Mobility Transistor, för höga frekvenser. Transistorn är tillverkad i Indiumfosfid (InP) i renrumslaboratoriet på institutionen för mikroteknologi och nanovetenskap. Den har fyra likadana "fingrar", 50 µm långa med en bredd (kanallängd) på 130 nm. Källa: Mikael Malmkvist, Optimization of Narrow Bandgap HEMTs for Low-Noise and Low-Power Applications, Avdelningen för mikrovågselektronik, Institutionen för mikroteknologi och nanovetenskap, Chalmers, doktorsavhandling, 2008

#### Innehåll

1.	. Intr	roduktion till halvledarmodeller	2
	1.1	Strömmodeller	2
	1.2	Farttriangeln och en hastighetsmodell	2
	1.3	Ledningsförmåga	3
	1.4	Dopning med störatomer	4
	1.5	Förfinad modell för elektron- och hålkoncentrationerna	5
	1.6	Lite mer om styckevis linjära modeller	5
	1.7	Rörlighet och drifthastighet	6
	1.8	Rörlighetens dopningsberoende	6
	1.9	Resistivitetens dopningsberoende	7
	1.10	Hastighetsmättnad	
	1.11	Summering	
	1.12	Övningsexempel	9
2.	Fer	mi-Dirac-statistik	
	2.1	Bohrs atommodell	10
	2.2	Lednings- och valensband	10
	2.3	Statistisk mekanik och Fermi-Dirac-statistik	11
	2.4	Tillståndstäthet	12
	2.5	Fermistatistik för hål och elektroner	12
	2.6	Massverkans lag och egentätheten n <sub>i</sub>	14
	2.7	Med egentätheten n <sub>i</sub> som referens	14
	2.8	Sammanfattning	15
	2.9	Övningsexempel	16
3.	. Ten	nperaturberoende	17
	3.1	Rörlighetens temperaturberoende	17
	3.1.	1 Gitterörlighetens temperaturberoende.	
	2.1.	Z Storröringhetens temperaturberoende	1/
	2.2	L admingatörmågang tomperaturbergande	10
	2.0	Eermionorging, temperaturbergende	10
	5.4 2.5	Samman fattaing	19
	5.5 DN	sammamatumig	
4.	. PN-	-overgangen	
	4.1	pn-overgangen	
	4.2	Parameterextraction frammatningar	
	4.3		
	4.4		
	4.5 4.5	1 Zenergenombrott	
	4.5.	2 Lavingenombrott	
	4.6	Diodens ström/spänningskarakteristik	
	4.7	Avvikelser från ideal diffusionsbegränsning	

4.8	Diodkarakteristikens temperaturberoende	
۳.c ۸ 0	Diodens kanagitans	30
4.9	Diodens transientegenskaper	
4.10	Samman fattning	
4 12	Övningsevennel	34
5 M	Oviningsexemper	
5. IVI	CS-1 Fallsistorn	
5.1	Fallenekteli	
5.2	Mättnadsområdat	
5.5 5.4	Sommonfottning ov MOS transistorne eträmmedell	
5.4	Strakavia liniär MOS-madall	
5.5	Styckevis inijai MOS-moden	
5.6 5.6	Enkla MOS-kretsar	
5.7	Analogförstärkaren	
5.8	Transientförlopp och frekvensgång	
5.8	8.1 Inverterarsnabbhet	
5.8	8.2 Förstärkarens frekvensgång	
5.0	Sommon fottaina	
5.9	Sammaniatining	
5.10	Ovningsexemper	
6. M	OS-kondensatorn	
6.1	Ackumulation, utarmning och inversion	
6.2	Antalet elektroner vid halvledarytan	
6.3	High-kappa MOS-kondensatorer	
6.4	MOS-kondensatorns banddiagram	
6.5	MOS-transistorn i svag inversion	
6.6	Poissons ekvation i svag och inversion	
6.7	Sammanfattning	
6.8	Ovningsexempel	
7. M	OS-transistorn	
7.1	Gradvisa kanalapproximationen.	
7.2	MOS-transistorns banddiagram	
7.3	Parameterextraktion	
7.4	Andra ordningens effekter	
7.4 7.4	4.1 Substrateffekten	
7.4	4.3 Drain-induced barrier-lowering (DIBL).	
7.4	4.4 Hastighetsmättnad	
7.4	4.5 Subtroskelstrom 4.6 Kanallängdsmodulation	
7.5	MOS-transistorns utarmningsområde	
7.6	Samman fattning	
7.0 7.7	Övningsexempel	
Q 1.7		
o. IVI	IKI UCICKU UIIKIIISIUI IA	
9. AI	PPENDIX 1	

9.1	Väteatomens tillåtna energinivåer	. 71	
9.2	Tillståndstätheten vs elektronenergin	71	
10. App	oendix 2: Grundläggande elektroteknik	. 72	
10.1	RC-kretsen	. 72	
10.2	Thevenin och Nortons teorem	. 73	
10.3	Gauss lag lag och Maxwells första ekvation	. 74	
10.4	Ficks lagar	. 76	
10.5	Diffusionsekvationen för majoritets- och minoritetsbärare	. 77	
10.6	Kontinuitetsekvationen	.78	
11. Index			

### Mikroelektronik - komponenter

Kjell Jeppson *(Senior Member, IEEE)* Chalmers University of Technology Department of Microtechnology and Nanoscience Göteborg, Sweden kjell.jeppson@chalmers.se

*Abstract* - Detta kurshäfte innehåller material till kursen i mikroelektronik – en kurs med inriktning mot komponenter. Kursens målsättningar och innehåll beskrivs i kortform nedan.

#### **KURSINTRODUKTION**

Kursens syfte är att ge en introduktion till mikroelektroniken, och en fysikalisk bakgrund till hur mikroelektroniska halvledarkomponenter fungerar. Kursen förutsätter en viss grundläggande kunskap om kvantmekanik och fermistatistik och om betydelsen av begreppet effektiv massa. Vad gäller halvledarkomponenter, dioder och transistorer, resulterar kursen i fysikaliskt baserade modeller i form av ekvivalenta kretsar - d.v.s. kretsscheman där elektriska kretselement som resistorer, kondensatorer, och spänningsstyrda strömkällor används för att beskriva de fysikaliska egenskaperna hos komponenten.

Efter genomgången kurs ska man med hjälp av enkla fysikaliska modeller

- kunna förklara hur halvledares ledningsförmåga varierar med parametrar som dopning, bandgap och temperatur
- kunna föra kvalitativa fysikaliska resonemang kring de mekanismer som begränsar strömmen genom en komponent
- kunna förenkla och renodla fysikaliska fenomen i material och komponenter med hjälp av styckevis linjära modeller
- kunna beräkna strömmen genom en diod eller en transistor som funktion av pålagda spänningar i olika spännings- och temperaturintervall och för olika komponentgeometrier
- kunna mäta på dioder och transistorer och från mätningarna bestämma värdena på viktiga modellparametrar
- kunna illustrera kretsars belastningslinjer, arbetspunkter och överföringskaraktertistika med enkla grafer.
- kunna tillämpa dimensionsanalys vid fysikaliska beräkningar.

Kursen består av fyra huvudsakliga delmoment:

- ledningsförmåga, Fermi-Dirac-statistik
- enkla halvledarkretsar
- pn-övergångens fysik
- MOS-transistorn

Kursen är i princip uppdelad i tre tvåveckors block. De första två kursveckorna är av repetitionskaraktär och avsikten är att hämta upp vissa grundkunskaper i fysik, om elektriska fält, och om enkla diod- och MOS-kretsar.

Nästa två veckor är avsatta för dioden – pn-övergången och dess likriktande egenskaper. Vad betyder det att det finns en inbyggd kontaktpotential och hur ser den strömbegränsande mekanism ut som ger ett exponentiellt spänningsberoende? Diodens spänningsberoende kapacitans studeras också i detalj och paralleller dras med en plattkondensator där "avståndet mellan plattorna" kan styras elektriskt med diodspänningen.

På samma sätt är två veckor avsatta för MOS-transistorn. Vi börjar med att genom gradvisa kanalapproximationen visa vilka förutsättningar den enkla kvadratiska MOS-modellen som Shockley en gång utvecklade bygger på. Därefter behandlas några av de andra ordningens effekter som uppträder i dagens nanometerstora transistorer med extremt korta kanaler.

Kurshäftet spänner upp hela "kursmängden" och alla modeller och formler som ingår i kursen finns med i häftet (och ibland mer därtill). Vissa förklaringar kanske kan uppfattas som knapphändiga eftersom materialet är skrivet ganska kortfattat. I sådana fall hänvisas man till någon utförligare lärobok. Man kan också "googla" på internet – det finns massor av material från olika universitet världen över utlagt på nätet som förklarar olika halvledarfenomen.

Från början var syftet med kurshäftet att ge en snabb överblick över kursinnehållet – varje kursavsnitt var tänkt att rymmas på ett uppslag. Efterhand som materialet kompletterats och utvidgats har den idén dock övergivits. Materialet är numera ganska heltäckande och någon ytterligare kursbok ska egentligen inte behövas. ©

Som framgår av inledningen är alla lärandemål formulerade kring användandet av fysikaliska modeller. Metoder för att formulera och använda modeller är på så vis centrala i kursen. Det är viktigt att förstå att modeller är just - modeller - och att olika antaganden om förutsättningarna resulterar i modeller med olika noggrannhet.

Varför ska man då modellera och vad är en modell bra för? Det finns minst tre svar. Modeller är till för att vi ska

- kunna förstå hur något fungerar och hur en given storhet varierar med olika variabler,
- kunna konstruera elektroniska kretsar med vissa givna prestanda
- lära oss mer om modellering i sig.

#### 1. INTRODUKTION TILL HALVLEDARMODELLER

Syftet med detta kapitel är att ge en introduktion till halvledarmodeller. Efter ett grundläggande resonemang kring ström och flöden av laddningsbärare inför vi begreppen ledningsförmåga och resistivitet, rörlighet och drifthastighet. Målet är att man ska kunna förstå och modellera såväl rörlighet som ledningsförmåga/resistivitet hos en halvledare som funktion av dopningen i halvledaren.

#### 1.1 Strömmodeller

Kapitlet tar sin utgångspunkt i grundläggande begrepp och utifrån dessa begrepp ska vi försöka föra enkla fysikaliska resonemang för att bättre förstå de olika fenomen som vi möter i en halvledare eller i en halvledarkomponent. Förhoppningsvis ska dessa resonemang visa hur man kan kvantifiera fysikaliska fenomen med matematiska formler för att på så sätt förstå vilken inverkan olika storheter, som exempelvis spänning och temperatur, kan ha. Låt oss börja med att se på vad elektrisk ström egentligen är: ström innebär ett flöde av laddningsbärare genom en ledare. Vi kan skriva sambandet mellan ström, laddning och tid, som

$$I = \frac{Q}{t}, \qquad (1.1)$$

ett samband som relaterar en ström I av laddningsbärare till den motsvarande mängd laddning Q som under en viss tid t passerar ett givet tvärsnitt av ledaren. Varje laddningsbärare bär laddningen q motsvarande elektronens laddning, dvs  $q=1,6 \cdot 10^{-19}$  coulomb (eller amperesekund). En enkel dimensionskontroll bekräftar rimligheten i vår modell, ström[A]=laddning[As]/tid[s].

Ström kan uppstå på grund av att ett elektriskt fält får laddningsbärarna att röra sig genom *drift*. Elektriska fält uppstår när vi har potentialskillnader i rummet och dess dimension är således [V/m]. Som vi ska se senare så finns det även andra drivkrafter än elektriska fält (nämligen koncentrationsskillnader som kan få laddningsbärarna att röra sig genom *diffusion*). Laddningsbärarna rör sig i ett elektriskt fält med en genomsnittlig drifthastighet i fältets riktning. Se fig. 1.1. Om vi antar att denna hastighet är v kan vi skriva om ekvation (1.1) på följande sätt:

$$I = \frac{Q}{t} = \frac{Q}{s} \frac{s}{t} = Q' \cdot v, \qquad (1.2)$$



Fig. 1.1. En elektrisk ledare där laddningen Q' per längdenhet rör sig med drifthastigheten v och därmed ger upphov till strömmen I=Q'v.

där s=vt är den sträcka som laddningsbärarna hinner tillryggalägga under tiden t och där Q' är mängden laddning per längdenhet i ledaren. Om vi har n rörliga laddningsbärare per volymsenhet i ledaren kan vi skriva Q'=qAn, där A är ledarens tvärsnittsarea. Det här är en mycket grundläggande formel och vi kommer att bygga många resonemang kring ström i ledare och i komponenter som MOS-transistorer kring denna modell.

#### *1.2 Farttriangeln och en hastighetsmodell*

I våra resonemang kring modeller kommer vi också att använda oss av "farttriangeln". Denna triangel är som ni säkert minns ett sätt som vi använde redan på högstadiet för att att åskådliggöra sambandet s=vt och för att nästan mekaniskt nöta in formeln för olika beräkningar av sträckor, tid och hastighet, se fig 1.2. Figuren visar också att vår välbekanta Ohms lag kan åskådliggöras på samma sätt. Triangeln är helt enkelt ett sätt åskådliggöra linjära samband (som går genom origo).

Vi ska nu använda farttriangeln för att bestämma laddningsbärarnas löptid genom en transistor. Denna löptid avgör hur snabbt en transistor kan arbeta och vilken övre gränsfrekvens den har. När en transistor slår om från ledande (*ON*) till oledande (*OFF*) tillstånd tar det en viss tid för de sist injicerade laddningsbärarna att genomlöpa transistorn. Alla laddningsbärarna måste helt enkelt få tid på sig att hinna ut. Så länge det finns laddningsbärare kvar på väg ut ur transistorn flyter det fortfarande ström. Inte förrän alla laddningsbärare hunnit ut ur transistorn blir denna oledande. Enligt farttriangeln bestäms löptiden av följande samband

$$t_L = \frac{L}{v},\tag{1.3}$$

där L är avståndet mellan transistorns emitter och dess kollektor.



Fig. 1.2. Farttriangeln och Ohms lag.

För att kunna analysera denna formel vidare så behöver vi nu en hastighetsmodell, d.v.s. en modell för hur drifthastigheten v beror på det elektriska fältet E. Den modell vi ska använda oss av är av enklast tänkbara typ, d.v.s. en linjär modell, vilket förstås innebär att vi antar att drifthastigheten är direkt proportionell mot det elektriska fältet. Proportionalitetskonstanten  $\mu$  benämns laddningsbärarna rörlighet, eller mobilitet,

$$v = \mu E . \tag{1.4}$$

Vår linjära hastighetsmodell kan åskådliggöras med en "rörlighetstriangel" som liknar farttriangeln, se fig 1.3. Med denna hastighetsmodell i bagaget kan vi återvända till löptiden som då kan skrivas

$$t_{L} = \frac{L}{v} = \frac{L}{\mu V / L} = \frac{L^{2}}{\mu V}, \qquad (1.5)$$

där vi använt oss av uttrycket E=V/L [V/m] för det elektriska fältet och där V är den pålagda spänningen och L är sträckans längd.

Denna formel visar att löptiden beror kvadratiskt på den sträcka som laddningsbärarna måste tillryggalägga genom transistorn. När vi ser ett så starkt längdberoende kan vi kanske bättre förstå en av de bakomliggande drivkrafterna till varför transistorer görs allt mindre och mindre för varje ny teknikgeneration. Industrin, som hela tiden strävar efter att möta behovet av allt snabbare och kraftfullare hemdatorer och grafikprocessorer, måste helt enkelt göra transistorerna mindre för att uppnå detta. När transistorerna görs mindre får man dessutom plats med fler, varför också allt fler komplexa funktioner kan integreras. Strävan efter att använda material med så bra rörlighet som möjligt blir också uppenbar. Vi ser också att en del av de fördelar som man uppnår genom att krympa dimensionerna upphävs av att man samtidigt måste sänka matningsspänningen för att inte det elektriska fältet ska bli för stort och därmed knäcka transistorerna.

I ett senare avsnitt ska vi återvända till begreppet rörlighet och studera detta i mer detalj eftersom laddningsbärarnas rörlighet är en för en halvledare mycket viktig och grundläggande parameter.

#### 1.3 Ledningsförmåga

Men innan vi gör det ska vi diskutera begreppet elektrisk ledningsförmåga. Material brukar delas in i ledare, halvledare och isolatorer allt efter den ledningsförmåga de har. Om vi nu diskuterar strömtäthet J istället för ström I så slipper vi att blanda in tvärsnittsarean A. Strömtätheten ges då av mängden rörlig laddning per volymsenhet gånger drifthastigheten, dvs J=Q'v [A/m<sup>2</sup>]. I en ledare, typ en koppartråd, med n laddningsbärare per volymsenhet ges strömtätheten då av

$$J = nqv = nq\mu E = \sigma E . \tag{1.6}$$

Proportionalitetskonstanten  $\sigma = nq\mu$ , som anger det linjära sambandet mellan strömtäthet och elektriskt fält, definierar den elektriska ledningsförmågan hos ledaren.

Ett materials ledningsförmåga är en intressant parameter som används för att, som vi nämnde inledningsvis, dela in



Fig. 1.3. Hastighets- och fältmodellen bakom Ohms lag.

material i ledare, halvledare och isolatorer. Är ledningsförmågan större än 1000  $\Omega^{-1}m^{-1}$  anses materialet vara en ledare, är den mindre än 10<sup>-5</sup>  $\Omega^{-1}m^{-1}$  anses materialet vara en isolator. Metaller, typ koppartrådar, är bra ledare. I avancerade snabba integrerade kretsar som mikroprocessorer har man relativt nyss utvecklat ny teknik för att kunna använda just kopparledare för att få så hög ledningsförmåga som möjligt och därmed snabbare kretsar. Kiseldioxid och teflon är bra isolatorer. En del polymerer och plaster är också bra isolatorer, och används som isolatorer kring elektriska kablar.

I intervallet mellan ledare och isolatorer finner vi halvledarna, d.v.s. grupp IV-halvledare som kisel (Si) och germanium (Ge) samt III-V-halvledare som galliumarsenid (GaAs), indiumantimonid (InSb) och indiumfosfid (InP). Det är dessa halvledarmaterial, framför allt kisel, och dess egenskaper som står i fokus i denna bok.

Ledningsförmågan beror som sagt bland annat på hur många rörliga laddningsbärare det finns i materialet. I en bra ledare som en metalltråd bidrar i stort sett varje atom med en ledningselektron. Där finns det således gott om rörliga laddningsbärare. Utifrån kunskap om kristallstrukturen kan vi göra en grov uppskattning av detta antal: om det typiskt är 0,5 nm mellan atomerna i ett material kan vi anta att varje atom upptar volymen av en kub med sidan 0,25 nm - och det motsvarar  $6 \cdot 10^{22}$  atomer per cm<sup>3</sup>. I en ideal isolator finns det inga rörliga laddningsbärare alls. En halvledare, är som namnet antyder ett mellanting, men det mest intressanta med en halvledare är att mängden laddningsbärare kan påverkas genom dopning av halvledaren. I ett egenledande, eller intrinsiskt, halvledarmaterial som kisel finns endast  $n_i = 10^{10}$  laddningsbärare per cm<sup>3</sup> vid rumstemperatur. I en dopad, eller extrinsisk, halvledare kan det däremot finnas från 10<sup>12</sup> upp till 10<sup>20</sup> laddningsbärare per cm<sup>3</sup> beroende på med hur många störatomer halvledarkristallen dopats. Varje störatom bidrar nämligen med en laddningsbärare. Antalet rörliga laddningsbärare, och därmed ledningsförmågan, kan således variera över inte mindre än tio tiopotenser (eller tio storleksordningar - ett uttryck som motsvarar engelskans "orders of magnitude").

En enkel tvådimensionell bild av en kiselhalvledares kristallstruktur visas i fig. 1.4. Varje kiselatom är omgiven av fyra grannar som den delar sina valenselektroner med för att bilda kovalenta bindningar. Eftersom varje kiselatom har fyra valenselektroner får den ett fullt yttersta elektronskal genom att dela ett elektronpar med var och en av sina fyra grannar. Vid varje temperatur är det möjligt för ett litet antal av dessa valenselektroner att bryta sig loss från den kovalenta bindningen. De blir då istället ledningselektroner som kan röra sig fritt i kristallen och därmed bidra till strömtransporten. Det intressanta är att även det hål som valenselektronen lämnar efter sig kan röra sig fritt i kristallen. Hålen rör sig som positivt laddade partiklar, men de bidrar till strömtransporten på samma sätt som ledningselektronerna. Eftersom varje ledningselektron lämnar efter sig ett hål, talar man om termiskt alstrade hål-elektronpar. Den energi som krävs för att alstra ett hål-elektronpar är 1,12 eV (vid rumstemperatur).

Denna energi tillförs elektronen termiskt men den kan också tillföras optiskt genom belysning. Den energi som krävs för att bilda ett hål-elektronpar kan därför mätas upp genom att man mäter fotokonduktiviteten. Genom sådana mätningar kan man från ljusets våglängd bestämma den minsta fotonenergi som krävs för att optiskt alstrade hålelektronpar ska bidra till att öka ledningsförmågan.



Fig. 1.4. Termiskt genererade hål-elektronpar i en intrinsis halvledare. Både hål och elektroner kan röra sig fritt.

#### 1.4 Dopning med störatomer

Antalet termiskt genererade hål-elektronpar i en halvledare är inte särskilt stort, endast ett per tusen miljarder bindningar vid rumstemperatur. Betydligt fler elektroner eller hål kan tillföras om halvledaren dopas med störatomer. Det finns två typer av störatomer i en halvledare, nämligen donatorer och acceptorer. Donatorer bidrar med ledningselektroner, medan acceptorer bidrar med hål. Donatorer är femvärda, typ fosfor och arsenik, medan acceptorer är trevärda, typ bor.

En donator har en elektron "för mycket", dvs en femte elektron som den kan donera till ledningsbandet. Halvledaren får då många fler elektroner än hål och blir av ntyp (efter de negativt laddade elektronerna). En acceptor har bara tre elektroner, vilket är en elektron "för lite". Den behöver därför låna en valenselektron för att dess bindningar ska bli kompletta. För varje valenselektron som binds till en acceptoratom efterlämnas ett fritt rörligt hål. En halvledare dopad med acceptorer får därför många fler hål än elektroner och blir följaktligen av p-typ (efter de positivt laddade hålen).

I fig. 1.5 visas förenklade kristallstrukturer för dopade



*Fig. 1.5. Bor och fosfor som störatomer i dopade halvledar av p- respektive n-typ.* 

halvledare av p- och n-typ. Halvledaren till vänster är dopad med bor och är således av p-typ. Halvledaren till höger är dopad med fosfor och är således av n-typ. Eftersom störatomernas jonisationsenergi är ganska liten, bara 50 meV (jämför t ex med vätes jonisationsenergi som är 13,6 eV), är jonisationen av störatomerna nästan hundraprocentig.

I en n-typ halvledare är således mängden ledningselektroner bestämd av donatorkoncentrationen,  $n=N_D$ , eftersom så gott som samtliga donatorer är joniserade. I en p-typ halvledare är mängden hål på samma sätt bestämd av acceptorkoncentrationen,  $p=N_A$ . Antalet laddningsbärare i en halvledare bestäms således av om halvledaren är *extrinsisk* eller *intrinsisk* (egenledande). I en extrinsisk halvledare är det dopningen, dvs antalet störatomer, som bestämmer antalet laddningsbärare. I en intrinsisk halvledare är det den termiska alstringen av hål-elektronpar som bestämmer antalet laddningsbärare.

Observera att en ledningselektron som donerats av en störatom inte motsvaras av något hål. Ett hål från en acceptor motsvaras inte heller av någon ledningselektron.

Den maximala lösligheten av störämnen motsvarar i grova tal en knapp procent av alla atomerna, d.v.s. maximal dopning är ca  $10^{20}$  störatomer per cm<sup>3</sup>. Minsta elektriskt mätbara dopning bestäms av egentätheten  $n_i$  (och som för kisel vid rumstemperatur är av storleksordningen  $10^{10}$  störatomer per cm<sup>3</sup>).

Det som är unikt med en halvledare är således att både hål och elektroner bidrar till strömledningsförmågan. Ekvation (1.6) som beskriver ledningsförmågan i en metall måste därför generaliseras på följande sätt för att ta hänsyn till strömbidragen från såväl hål som elektroner,

$$\sigma = q \left( n \mu_n + p \mu_p \right), \tag{1.7}$$

där  $\mu_n$  är elektronernas rörlighet och  $\mu_p$  är hålens rörlighet.

Den typ av laddningsbärare som det finns flest av i en halvledare kallas *majoritetsbärare*. Elektronena är således majoritetsbärare i en n-typ halvledare, medan hålen är majoritetsbärare i en p-typ halvledare. Den andra typen av laddningsbärare kallas följaktligen för *minoritetsbärare*. Som vi ska se i ett senare avsnitt är produkten av antalet hål och elektroner, enligt massverkans lag, alltid konstant,

$$np = n_i^2. (1.8)$$

En n-typ halvledare med dopningen  $N_D=10^{16}$  cm<sup>-3</sup> har  $n=10^{16}$  cm<sup>-3</sup> och följaktligen  $p=10^4$  cm<sup>-3</sup> (eftersom  $n_i=10^{10}$  cm<sup>-3</sup>). Ökar vi dopningen till  $N_D=10^{17}$  cm<sup>-3</sup> får vi  $n=10^{17}$  cm<sup>-3</sup> medan hålkoncentrationen minskar till  $p=10^3$  cm<sup>-3</sup>. Som synes skiljer det åtskilliga tiopotenser mellan antalet majoritetsbärare och antalet minoritetsbärare.

En bild av elektronkoncentrationen som funktion av donatordopningen  $N_D$  i en n-typ halvledare visas i fig. 1.7. Sambandet är linjärt,  $n=N_D$ , så länge halvledaren är extrinsisk (dvs så länge  $N_D >> n_i$ ). För försumbara dopningar blir halvledaren intrinsisk (egenledande) med  $n=p=n_i$ . Figuren illustrerar också begreppet styckevis linjär modell som en metod för att förenkla och linjarisera icke-linjära samband.



Fig. 1.7. Styckevis linjär modell för elektronkoncentratione som funktion av antalet donatorer i en n-typ halvledare.

#### 1.5 Förfinad modell för elektron- och hålkoncentrationerna

En mer noggrann modell än den styckevis linjära modellen kan vi få om vi även tar i beaktande att det råder laddningsneutralitet i en halvledare. Den förfinade modellen kan, om man så vill, ses som en "smoothing curve" mellan de två asymptotiska linjerna för extrinsiska och intrinsiska halvledare i fig. 1.7. Antag för fullständighetens skull att vi har både acceptorer,  $N_A$ , och donatorer,  $N_D$ , i halvledaren. Villkoret för laddningsneutralitet ger oss då sambandet att summan av alla negativa laddningar är lika med summan av alla positiva laddningar, d.v.s.

$$n + N_A = p + N_D. \tag{1.9}$$

Om  $N_D > N_A$  så är halvledaren av n-typ med nettodopningen  $N_{D,eff} = N_D - N_A$ , om  $N_A > N_D$  är den av p-typ med nettodopningen  $N_{A,eff} = N_A - N_D$ . Villkoret för laddningsneutralitet tillsammans med massverkans lag (1.8) kan då formuleras i följande två ekvationer

$$\begin{cases} pn = n_i^2 \\ p + N_D = n + N_A \end{cases}$$
(1.10)

Dessa två ekvationer ger oss följande lösning för antalet majoritetsbärare (d.v.s. för antalet elektroner) i en n-typ halvledare

$$n = \frac{N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_i^2}}{2} = \begin{cases} N_D - N_A \\ n_i \end{cases},$$
(1.11)

och hålkoncentrationen i en p-typ halvledare med nettodopningen  $N_A$ - $N_D$ 

$$p = \frac{N_A - N_D + \sqrt{\left(N_A - N_D\right)^2 + 4n_i^2}}{2} = \begin{cases} N_A - N_D \\ n_i \end{cases}.$$
(1.12)

Lösningen i ekvation (1.11) har, som vi försökt antyda i fig. 1.7, två asymptoter:  $n=n_i$  för en intrinsisk halvledare  $(N_D-N_A << n_i)$  och  $n=N_D-N_A$  för en extrinsisk n-typ halvledare  $(N_D-N_A >> n_i)$ . Lösningen ansluter således

naturligt till vår styckevis linjära modell enligt tidigare resonemang om "*smoothing curve*".

En styckevis linjär modell ger en förenklad bild som är lätt att använda när man ska göra överslagsberäkningar för hand i ett litet block, på en lös lapp eller på baksidan av ett använt kuvert (*eng.* 'back-of-the-envelope calculations'). När vi ska lägga in ekvationen i ett plottprogram för att rita kurvan så använder vi dock helst den exakta lösningen så att vi får en jämn "smooth" kurva och att vi slipper att använda if-satser.

#### 1.6 Lite mer om styckevis linjära modeller

Vi kommer att använda styckevis linjära modeller så ofta vi kan i våra resonemang för att få lättbegripliga modeller att "hänga upp" mer detaljerade icke-linjära modeller på. De icke-linjära modellerna fungerar då som "*smoothing curves*" som i ändlägena ansluter asymptotiskt till de förenklade, linjära modellerna. Till exempel kan vi använda styckevis linjära modeller för att beskriva ickelinjära komponenter, som dioder och MOS-transistorer.

I huvudsak använder vi två typer av styckevis linjära modeller; modeller för komponenter som inte börjar leda förrän vid en viss tröskelspänning  $V_T$  och modeller för komponenter där strömmen mättas vid en viss spänning  $V_{DSAT}$ . I fig. 1.8 ser vi en styckevis linjär modell för strömspänningskarakteristiken för en diod med den inbyggda spänningen  $V_{bi}$ . Men det kan också vara ström-spänningskarakteristiken för en MOS-transistor med tröskelspänningen  $V_T$ . För spänningar under en viss kritisk spänning leder komponenten inte ström, I=0 för  $V < V_T$ , medan den för spänningar över den kritiska spänningen uppträder som en konduktans  $G_{ON}$ ,  $I=G_{ON}V$  för  $V>V_T$ .



Fig. 1.8. Styckevis linjär modell för diod eller transistor.

I fig. 1.9 har vi använt en styckevis linjär modell för att illustrera hur strömmen genom en MOS-transistor når ett mättnadsvärde  $I_{DSAT}$ . För små drainspänningar,  $V_{DS} < V_{DSAT}$ , i det så kallade resistansområdet uppvisar MOS-transistorn en resistiv karakteristik som kan beskrivas med hjälp av en konduktans  $G_{ON}$ . För större drainspänningar planar strömkarakteristiken ut för att succesivt nå mättnadsvärdet  $I_{DSAT}$ . Den styckevis linjära modellen är lite av en förenklad "antingen eller" modell; antingen linjär eller mättad. Den icke-linjära modellen är en del av en parabel som i ändpunkterna ansluter "mjukt" till de två asymptoterna

och som visar att MOS-transistorn uppträder som en ickelinjär resistans i resistansområdet.

MOSFET styckevis linjär modell



Fig. 1.9. Styckevis linjär modell för MOSFET med resistans- och mättnadsområde.

#### 1.7 Rörlighet och drifthastighet

Låt oss efter denna utvikning om styckevis linjära modeller återvända till vår hastighetsmodell,  $v=\mu E$ . Denna modell bygger, som vi tidigare nämnt, på att vi har "fria" laddningsbärare, hål eller elektroner, som med hög hastighet far omkring helt slumpmässigt i olika riktningar inne i halvledarkristallen. Laddningsbärarnas hastighet, den s.k. termiska hastigheten  $v_{th}$ , är den medelhastighet som laddningsbärarna har vid en viss temperatur. Med jämna mellanrum kolliderar laddningsbärarna med kristallens gitter för att sedan fara iväg i en ny riktning. Efter en viss tid kolliderar de igen med gittret. Man talar om en medeltid mellan kollisioner, <t>. Se fig. 1.10.



Fig. 1.10. En laddningsbärares slumpmässiga rörelser i er halvledarkristall, med och utan pålagt elektriskt fält.

Laddningsbärarnas rörelseenergier är statistiskt fördelade enligt Fermi-Diracs sannolikhetsfördelning som vi ska återkomma till i nästa kapitel. Ett resultat av laddningsbärarnas statistiska energifördelning är att medelvärdet av deras rörelseenergier ( $mv_{th}^{2}/2$ ) är lika med 3kT/2, där k är Boltzmanns konstant och T är temperaturen i Kelvin. Vid rumstemperatur (290K) är den termiska energin kT=25 meV. Det betyder att laddningsbärarnas termiska hastighet  $v_{th}$ , är ca 10<sup>5</sup> m/s vid rumstemperatur. Elektronens massa är 9,11·10<sup>-31</sup> kg.

I närvaro av ett elektriskt fält kommer laddningsbärarna att få en viss nettohastighet, en viss drifthastighet  $\langle v \rangle$ , i fältets riktning. Om vi utgår från Newtons andra lag,

$$F=m^*a,\tag{1.13}$$

och om vi något förenklat antar att accelerationen ges av drifthastigheten dividerat med medeltiden mellan kollisioner,

$$a = \langle v \rangle / \langle t \rangle, \tag{1.14}$$

så får vi med kraften F=qE pga det elektriska fältet,

$$qE = m^{*} < v > / < t >. \tag{1.15}$$

Man kan också se det som att laddningsbärarna förlorar hela den impuls  $m^{*} < v >$  som de på grund av det elektriska fältet tillförs mellan kollisionerna, dvs qE < t >. I vilket fall som helst betyder det att drifthastigheten kan skrivas

$$\langle v \rangle = \frac{q \langle t \rangle}{m^*} E . \tag{1.16}$$

Som väntat är drifthastigheten proportionell mot det elektriska fältet. Proportionalitetskonstanten

$$\mu = \frac{q < t >}{m^*}.\tag{1.17}$$

benämns *rörlighet*, eller *mobilitet*. Vi ser att rörligheten blir stor om det är lång tid mellan kollisionerna eller om den effektiva massan är liten. En lätt partikel rör sig som väntat snabbare än en tung, och rörligheten blir förstås större med färre kollisioner.

#### 1.8 Rörlighetens dopningsberoende

Vid låga dopningar bestäms kollisionsfrekvensen  $1/\langle t \rangle$ enbart av laddningsbärarnas kollisioner med kristallgittret (*eng.* lattice scattering) och rörligheten antar ett basvärde som vi kallar gittermobiliteten  $\mu_L$  (*eng.* lattice mobility). Gitterrörligheten i kisel vid rumstemperatur är för elektroner 1360 cm<sup>2</sup>/Vs och för hål 460 cm<sup>2</sup>/Vs. Observera att dimensionen för rörligheten är cm<sup>2</sup>/Vs, dvs samma som dimensionen för  $\langle v \rangle / E$ .

Rörligheten minskar med ökande dopning eftersom störatomernas inverkan på laddningsbärarna då inte längre kan försummas. Det tillkommer då ytterligare en kollisionsmekanism när laddningsbärarna kolliderar med joniserade störatomer. Kollisionerna med störatomerna är lite annorlunda till sin karaktär än kollisionerna med gittret, eftersom störatomerna är laddade. Det är på grund av laddningarna som laddningsbärarna böjs av i en ny bana i en annan riktning.

Vi får då *en* grundläggande kollisionsfrekvens för kollisionerna med gittret och *en annan* frekvens för kollisionerna med störatomerna. Enligt Mathiessens regel kan frekvenserna för olika kollisionstyper adderas för att ge den totala (sammanlagda) kollisionsfrekvensen. Det betyder att den effektiva rörligheten, med god approximation, kan skrivas

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I}, \qquad (1.18)$$

där  $\mu_l$  är "störrörligheten" (*eng.* impurity mobility) på grund av kollisionserna med störatomerna (*eng.* impurity scattering). Om vi helt enkelt antar att stömobiliteten är omvänt proportionell mot dopningen (den ska ju minska med ökande dopning) enligt

$$\mu_I = \mu_L \frac{N_0}{N},$$
 (1.19)

så kan den totala rörligheten skrivas

$$\mu = \frac{\mu_L}{1 + N/N_0}, \qquad (1.20)$$

där  $N_0$  betecknar den dopning där båda kollisionsfrekvenserna är lika stora, dvs  $\mu_I = \mu_L$ , och den totala rörligheten därmed halverats på grund av störkollisionerna. Ett typiskt riktvärde är  $N_0=1,3\,10^{17}$  cm<sup>-3</sup>. Hål- och elektronrörligheterna i kisel som funktion av dopningen vid rumstemperatur visas i fig. 1.11. Vi ser att rörligheten är relativt opåverkad av störkollisionerna vid låga dopnivåer (<10<sup>16</sup> cm<sup>-3</sup>). Vi ser också att rörligheten antar ett minimivärde  $\mu_{min}$  vid höga dopnivåer (> 10<sup>18</sup> cm<sup>-3</sup>), något som inte förutspås av modellen i (1.20)<sup>1</sup>. I kisel är minimirörligheten lika med 90 respektive 50 cm<sup>2</sup>/Vs för elektroner och hål. I dopningsintervallet upp till 10<sup>18</sup> cm<sup>-3</sup>



silicon electron and hole mobility

Fig. 1.11. Rörligheten i kisel som funktion av dopningen.

#### 1.9 Resistivitetens dopningsberoende

Vi har tidigare i inledningsavsnittet definierat ledningsförmågan  $\sigma$  hos ett ledande material med rörligheten  $\mu$  på följande sätt

$$J = nqv = nq\mu E = \sigma E , \qquad (1.21)$$

där n är laddningsbärare per volymsenhet.

<sup>*l*</sup> I figur 1.11 har vi använt en mer noggrann empirisk modell för rörligheten enligt  $\mu = \mu_{\min} + \frac{\mu_L - \mu_{\min}}{1 + (N/N_0)^{\alpha}}$ , där  $N_0 = 1,3.10^{17}$  cm<sup>-3</sup> och  $\alpha = 0,9$ . Jfr (1.20). Eftersom vi fann att det i en halvledare finns två typer av laddningsbärare, hål och elektroner, så modifierade vi uttrycket för ledningsförmågan så att det tar hänsyn inte bara till elektrontransport urtan även till håltransport

$$\sigma = q \left( n \mu_n + p \mu_p \right). \tag{1.22}$$

Vi kan också definiera materialets resistivitet  $\rho = 1/\sigma$ [ $\Omega$ cm]. I figur 1.12 har vi plottat resistiviteten för n- och p-typ kisel med hjälp av våra modeller för rörligheten och för antalet rörliga laddningsbärare som bidrar till strömtransporten. Vi ser att resistiviteten varierar mellan 1000  $\Omega$ cm vid dopningen 10<sup>13</sup> cm<sup>-3</sup> och 0,001  $\Omega$ cm vid löslighetsgränsen, d.v.s. vid maximalt möjliga dopningen 10<sup>20</sup> cm<sup>-3</sup>. Det är ett intervall som spänner över inte mindre än sex storleksordningar (tiopotenser).



Fig. 1.12. Resistiviteten [ $\Omega$ cm] i kisel vs dopningen.

Man kan lätt visa att resistansen hos en ledare med längden L och tvärsnittsarean A i ett material med resistiviteten  $\rho$  ges av

$$R = \rho \frac{L}{A}.$$
 (1.23)

Tvärsnittsarean ges av  $A=W^*d$ , där W är ledarens bredd och d är det ledarens tjocklek. I en integrerad krets är ofta alla resistanser formade i ett visst skikt med en viss tjocklek. Det betyder att alla resistanser i detta skikt har samma tjocklek. Det kan då vara smidigt att baka samman resistiviteten och skikttjockleken i en enda parameter, nämligen i skiktresistiviteten RS= $\rho/d$ . Denna anges i  $\Omega$ per ruta, eller egentligen  $\Omega$  per kvadrat (eng. ohms per square). Skiktresistiviteten anger nämligen resistansen hos en kvadratisk (W=L) resistor av godtycklig storlek. För att beräkna resistansen hos en viss resistor eller ledare i ett visst skikt är det därmed bara att beräkna antalet seriekopplade rutor, eller kvadrater, som ledaren rymmer längs hela sin längd, d.v.s. L/W, och multiplicera detta med skiktresistiviteten. (En dimensionskontroll ger  $[\Omega/\Box]^*[\Box]=[\Omega].$ 

I fig. 1.13 ser vi en förenklad skiss av en metalledare i en integrerad krets. Ledaren vi ser kan vara en del av en ledare mellan en processorkärna och cache-minnet. Ledaren är isolerad från kislet med ett tunt oxidskikt med tjockleken  $d_{ox}$ . Ledarens resistans *R* ges av

$$R = R_s \frac{L}{W} . \tag{1.24}$$

Ledarens kapacitans *C* ges av kapacitansen per ytenhet  $C_{ox} = \varepsilon_{ox}/d_{ox}$  där  $\varepsilon_{ox} = \kappa_{ox}\varepsilon_0$  är isolatorns dielektricitetskonstant och  $d_{ox}$  dess tjocklek ( $\kappa_{ox}$ =4 är isolatorns relativa permittivitet) samt av dess bottenarea A = WL enligt

$$C = WLC_{ox} . \tag{1.25}$$

Det betyder att ledarens *RC*-produkt, som är ett mått på den tid det tar att skicka en signal genom denna ledare, är proportionell mot ledarens längd i kvadrat,

$$RC \sim L^2 \,. \tag{1.26}$$

Vi kan nu dra samma slutsats som tidigare när det gällde en transistors löptid att fördröjningen ökar kvadratiskt med ledarlängden och att det därför gäller att hålla alla ledare så korta som möjligt. Långa ledare måste delas i kortare delar som drivs av var sitt drivsteg, s.k repeaters.



*Fig. 1.13*. *En ledares resistans och kapacitans som funktion av dess längd L och bredd W.* 

#### 1.10 Hastighetsmättnad

I det här kapitlet har vi talat om laddningsbärarnas drifthastighet i ett elektriskt fält och vi har byggt våra resonemang på en enkel modell där drifthastigheten antagits vara proportionell det elektriska fältet. Vi har också diskuterat egenskaperna hos denna proportionalitetskonstant, som benämns rörlighet, och dennas beroende av effektiva massan hos laddningsbärarna och dessas kollisionsfrekvens med gitter och störatomer. I ett senare kapitel ska vi också diskutera rörlighetens temperaturberoende.

Nu har förstås denna hastighetsmodell, som så många andra modeller, sina begränsningar. Modellen är en utmärkt modell för låga och måttliga elektriska fält, och är ju själva grunden för Ohms lag, men den fungerar inte för de höga elektriska fält som vi ibland ser i halvledarkomponenter. Även med så låga spänningar som typ en volt får vi höga elektriska fält i dagens moderna komponenter med dimensioner på 1  $\mu$ m och mindre. För höga elektriska fält mättas hastigheten, vilket innebär att den når ett maximalt värde vid ett visst kritiskt elektriskt fält. Vi har då nått hastighetsmättnad (*eng.* velocity saturation). Även om det elektriska fältet ökar över det kritiska värdet så ökar inte hastigheten mer eftersom laddningsbärarna förlorar sin energi till kristallgittret. Mättnadshastigheten  $v_{sat}$  är typiskt av samma storleksordning som den termiska hastigheten  $v_{th}$ , d.v.s. av storleksordningen  $10^5$  m/s (eller  $10^7$  cm/s som det står i amerikanska läroböcker).

En styckevis linjär modell för drifthastigheten för elektroner och hål i kisel visas i fig. 1.14 tillsammans med kontinuerliga modeller för hål- och elektronhastigheterna. En enkel sådan icke-linjär modell kan formuleras

$$v = \frac{\mu E}{1 + \frac{\mu E}{v_{sat}}} \Longrightarrow \begin{cases} \mu E, & E << E_{krit} \\ v_{sat}, & E >> E_{krit} \end{cases}.$$
 (1.27)

Storleksordningen på det kritiska fältet när hastighetsmättnad inträffar är för elektroner ungefär

$$E_{krit} = \frac{v_{sat}}{\mu} \approx \frac{10^7}{1000} = 10^4 \,\mathrm{V/cm}$$
 (1.28)

vilket motsvarar 1 V/ $\mu$ m. Det kritiska fältet för hål är en faktor två större.

#### DRIFT VELOCITY



Fig. 1.14. Drifthastigheten för hål och elektroner i kisel som funktion av det elektriska fältet.

#### 1.11 Summering

I detta kapitel betraktade vi först ström som transport av laddning per tidsenhet. Begreppet elektrisk ledningsförmåga definierades och material delades in i ledare, isolatorer och halvledare. I en ledare är ledningsförmågan hög på grund av det stora antalet fria laddningsbärare, medan den är låg i en isolator eftersom där knappt finns några fria laddningsbärare alls. Vi studerade sedan halvledare och konstaterade att både hål och elektroner bidrar till ledningsförmågan. I en intrinsisk halvledare alstras hålelektronpar termiskt. I en extrinsisk halvledare bestäms antalet laddningsbärare av störatomerna. Genom att välja hur hårt man dopar en halvledare kan ledningsförmågan styras över flera storleksordningar. En extrinsisk halvledare är antingen n-typ eller p-typ eftersom vissa störämnen är n-dopande och andra är p-dopande. Störatomerna är antingen donatorer eller acceptorer. I en ntyp halvledare är elektronerna majoritetsbärare, dvs i majoritet, medan hål är majoritetsbärare i p-typ halvledare. Nästa begrepp som diskuterades var laddningsbäranas rörlighet. Vi fann att rörligheten är proportionell mot medeltiden mellan laddningsbärarnas kollisioner med såväl kristallens gitteratomer som dess störatomer.

#### 1.12 Övningsexempel

- 1. Illustrera följande linjära samband med hjälp av "farttrianglar"!
  - a. Ohms lag
  - b. Kapacitansdefinitionen
  - c. Sambandet mellan strömstyrka och elektriskt fält
  - d. Sambandet mellan drifthastighet och fält
  - e. Sambandet mellan ström, laddningstäthet och drifthastighet!
- 2. Termisk energi
  - a. Hur stor är elektronens laddning?
  - b. Hur definieras energienheten elektronvolt?
  - c. Hur stor är den termiska energin kT vid rumstemperatur (290 K)? 25 meV
  - d. Hur stor är motsvarande termiska spänning V<sub>t</sub>=kT/q?
  - e. Beräkna elektronernas termiska hastighet om deras rörelseenergi är 3kT/2!
  - f. Elektronens massa är  $m_0=9,11\cdot 10^{-31}$  kg.
- 3. Dopning
  - a. För att få en uppfattning om antalet kiselatomer per volymsenhet tänker vi oss ett enkelt kubiskt gitter med avståndet 5 Å mellan atomerna. Hur många atomer per cm<sup>-3</sup> motsvarar det?
  - b. Vilken är maximala lösligheten av dopämnen i kisel?
  - c. Nämn några störämnen som i kisel är p-dopande respektive n-dopande!
  - d. Vad menas med donatorer resp. acceptorer?
  - e. Vilka två typer av laddningsbärare finns det i en halvledare?
  - f. Hur stor är egentätheten av laddningsbärare i en odopad kiselkristall?
  - g. Vad menas med majoritetsbärare respektive minoritetsbärare?
  - h. Hur många majoritetsbärare respektive minoritetsbärare finns det i en dopad kiselkristall?
- 4. Rörlighet
  - a. Resonera dig fram till ett samband mellan en grupp elektroners drifthastighet  $\langle v \rangle$  och det elektriska fältet, *E*, som ger upphov till drifthastigheten. Utgå från att kraften på laddningsbärarna F=qE i det elektriska fältet ger upphov till en acceleration *a*, bestämd av F=am. Antag att denna acceleration ges av det förenklade, men inte orimliga sambandet  $\langle v \rangle = a \langle t \rangle$ , där  $\langle t \rangle$  är medeltiden mellan kollisioner.
  - b. Vilken modell för laddningsbärarnas rörlighet leder detta resonemang fram till?
  - vilka är de två dominerande kollisionsmekanismerna och hur beror deras kollisionsfrekvenser 1/<t>
     på dopningen i kristallen?
  - d. Vid vilken dopning är de båda kollisionsfrekvenserna lika stora i kisel?
  - e. Hur lyder Mathiessens regel?

f. Hur stor är hålens drifthastighet i m/s om det elektriska fältet är 5 V/cm och om hålrörligheten i kisel antas vara 500 cm<sup>2</sup>/Vs?

#### 5. Resistivitet

- a. Antag som en första approximation att hål- och elektronrörligheterna är konstanta i dopnings-intervallet  $10^{10} < N < 10^{18}$  cm<sup>-3</sup> med värdena  $\mu_n$ =1300 cm<sup>2</sup>/Vs respektive  $\mu_p$ =500 cm<sup>2</sup>/Vs. Skissa resistiviteten som funktion av dopningen i detta intervall!
- b. Vilken n-dopning motsvaras av resistiviteten 5 Ωcm?
- c. För vilken dopning har resistiviteten i p-typ kisel ett maximum?
- 6. Majoritets- och minoritetsbärarkoncentrationer

Den intrinsiska koncentrationen i kisel har ett temperaturberoende enligt tabellen nedan<sup>2</sup>. Hur stora är elektron- och hålkoncentrationerna  $n_0$  och  $p_0$  i termisk jämvikt i följande homogent dopade kiselprover! Avgör om proven är av n- eller p-typ!

- a. 300K:  $N_A \ll N_D$ ,  $N_D = 10^{15} \text{ cm}^{-3}$
- b. 300K:  $N_D << N_A$ ,  $N_A = 10^{16}$  cm<sup>-3</sup>
- c. 300K:  $N_A = 9.10^{15} \text{ cm}^{-3}$ ,  $N_D = 10^{16} \text{ cm}^{-3}$
- d. 480K:  $N_A << N_D$ ,  $N_D = 10^{14}$  cm<sup>-3</sup>
- e. 670K:  $N_A << N_D$ ,  $N_D = 10^{14}$  cm<sup>-3</sup>

TEMP [K]	$n_i [cm^{-3}]$
300	$10^{10}$
370	10 <sup>12</sup>
480	$10^{14}$
670	10 <sup>16</sup>

- 7. Vilket *temperaturberoende* får gitterrörligheten om vi antar att medelvägen mellan kollisioner minskar med ökande temperatur enligt  $\langle s \rangle \sim 1/T$  och att rörelseenergin  $E \sim kT$ ? Antag att medeltiden  $\langle t \rangle$  mellan kollisioner ges av  $\langle s \rangle = \langle v_{th} \rangle \langle t \rangle$ .
- 8. *Hastighetsmättnad*. Om den maximala elektronhastigheten är 10<sup>5</sup> m/s, vid vilken elektrisk fältstyrka inträffar då denna hastighetsmättnad?

<sup>2</sup> Följande uttryck har använts för bandgapets temperaturberoende [se Pierret]:

$$E_G = E_{G0} - \frac{0,0005T^2}{T+636}$$
, där  $E_{G0} = 1,17$  eV.

#### 2. FERMI-DIRAC-STATISTIK

I förra kapitlet diskuterade vi hål och elektroner som laddningsbärare i en halvledare med hjälp av en mycket enkel kovalent bindningsmodell. Modellen var konceptuellt enkel och en bra grund för vidare studier, men den är inte tillräcklig för förståelsen av halvledares och halvledarkomponenters funktion. Många fenomen i halvledare förklaras bäst av energibandmodellen och vi ska därför kortfattat ge en bakgrund till denna energibandmodell.

#### 2.1 Bohrs atommodell

Låt oss därför gå tillbaka ända till Bohrs atommodell. Enligt denna modell kretsar elektronerna i en atom runt en atomkärna i väl bestämda banor, se fig. 2.1. Man talar om olika elektronskal där de innersta elektronerna är hårt bundna till kärnan medan de yttersta elektronerna, som kretsar längre från kärnan, är lösare bundna. Varje bana representerar en tillåten energinivå för elektronen och är enligt Schrödinger associerat med en vågfunktion som anger banans (mest sannolika) radie. Rent matematiskt är de tillåtna energinivåerna egenvärden till Schrödingerekvationen, medan vågfunktionen är lösningen till samma ekvation. Denna vågfunktion anger sannolikheten för att en elektron på en tillåten energinivå ska befinna sig på ett visst avstånd från kärnan. Ur fig. 2.1 framgår också sambandet mellan elektronbanans radie (dvs avståndet till kärnan) och elektronens energi.



Fig. 2.1. Enkelt energidiagram för väteatomen.

#### 2.2 Lednings- och valensband

I en kristall där flera atomer befinner sig ganska nära varandra kommer de enskilda atomernas diskreta energinivåer att splittras upp i band av tätt belägna energitillstånd. Detta har att göra med Paulis uteslutningsprincip som säger att två elektroner inte kan vara i samma kvanttillstånd samtidigt. De två energiband som är av intresse för en halvledare är valensbandet, som är i det närmaste helt fyllt med elektroner, och *ledningsbandet* som är i det närmaste helt tomt på elektroner. Ledningsbandet och valensbandet är i en halvledare åtskilda av ett förbjudet bandgap där inga tillåtna energinivåer finns. Se fig. 2.2.

Banddiagrammet kan användas för att förklara skillnaden mellan halvledare, isolator och ledare, se fig. 2.3. En halvledare har ett bandgap,  $E_G = E_C - E_V$ , där  $E_C$  är ledningsbandets undre kant och  $E_V$  är valensbandets övre kant, ett bandgap som är av storleksordningen 1 eV. Ett bandgap av den storleken är tillräckligt litet för att vi ska få tillgång till termiskt exciterade ledningselektroner i ledningsbandet och hål i valensbandet. I en isolator däremot är bandgapet så stort att vi varken har några ledningselektroner i ledningsbandet eller några hål i valensbandet. Ledningsförmågan blir därför väldigt låg. I ett fullt valensband, som i isolatorn, finns det nämligen inga lediga tillstånd för elektronerna att flytta till. Därmed får vi inte heller någon nettotransport.



Fig. 2.2. När flera atomer befinner sig nära varandra i en kristall splittras de yttre elektronbanornas diskreta energinivåer upp i band av tätt belägna energinivåer.

I en ledare är ledningsbandet endast delvis fyllt. Därmed får vi tillgång till en stor mängd ledningselektroner som kan röra sig fritt mellan ett stort antal lediga tillstånd. Varje energiband har plats för två elektroner per atom. Grundämnen med ett udda antal elektroner, typ guld, aluminium och silver, är därför bra ledare för då blir hälften av tillstånden lediga i ledningsbandet. Även grundämnen med ett jämnt antal elektroner kan vara ledare om valens- och ledningsbanden överlappar varandra (s.k. semimetaller). Se fig. 2.3.



*Fig. 2.3. Energibanddiagram för halvledare, isolator och ledare.* 

I fig. 2.4 visas ett mer detaljerat energibanddiagram för en halvledare. Ökande elektronenergi har markerats med en pil uppåt medan ökande hålenergi har markerats med en pil nedåt. Figuren illustrerar också ett antal termiskt alstrade hål-elektronpar med ledningselektronerna i botten av ledningsbandet och hålen överst i valensbandet. Laddningsbärarnas genomsnittliga rörelseenergi är 3kT/2, där k är Boltzmanns konstant och T är temperaturen i Kelvin. Denna energi är ungefär 37,5 meV vid rumstemperatur så ledningselektronerna befinner sig verkligen i botten av ledningsbandet mycket nära ledningsbandkanten.



Fig. 2.4. Energibanddiagram för halvledare som kisel med ledningselektroner längst ner i ledningsbandet och hål längst upp i valensbandet.

Ett detaljerat banddiagram för en dopad halvledare visas i fig. 2.5. I denna figur har därför även donatornivån och acceptornivån lagts in. En donator lämnar mycket lätt ifrån sig sin femte elektron till ledningsbandet. Jonisationsenergin för denna femte elektron är därför ganska liten, typiskt av storleksordningen 50 meV. Därför kan vi rita in donatornivån 50 meV under ledningsbandkanten. Det är på den energinivån donatorns femte elektron befinner sig när den är bunden till donatorn, dvs när donatorn inte är joniserad. Donatorns jonisationsenergi är således  $E_C$ - $E_D$ . På samma sätt är acceptorns jonisationsenergi  $E_A$ - $E_V$ , dvs det är den energi som krävs för att jonisera acceptorn genom att ta emot en elektron från valensbandet och därvid lämna kvar ett hål.



Fig. 2.5. Enkelt banddiagram med störnivåer.

#### 2.3 Statistisk mekanik och Fermi-Dirac-statistik

I detta avsnitt utgår vi ifrån att läsaren någorlunda känner till egenskaperna hos rörliga partiklar som slumpmässigt far omkring bland många andra likadana partiklar. Exempel på sådana partiklar är molekyler i en gas, och - som vi ska intressera oss för här - elektroner och hål i en halvledare. Om vi för en stund studerar en klassisk ideal gas så vet vi att molekylerna rör sig olika fort och i olika riktningar och därmed har de också olika energier,  $E=mv^2/2$ , där *m* är molekylens massa och *v* dess hastighet. Enligt Boltzmann är sannolikheten *f*(*E*) att hitta en molekyl med energin *E* exponentiellt avklingande med ökande energi,

$$f(E) = e^{-E/kT}, \qquad (2.1)$$

där k är Boltzmanns konstant och T är temperaturen i Kelvin. Med värdet  $k=8,62\cdot10^{-5}$  eV/K på Boltzmanns

konstant får vi värdet på den termiska energin kT=25 meV vid 290K, ett värde som är mycket lättare att komma ihåg än värdet på Boltzmanns konstant.

Enligt Boltzmanns sannolikhetsfördelning är sannolikheten lika med ett att hitta en molekyl på den lägsta energinivån, och sedan avtar sannolikheten exponentiellt med ökande energi. Medelenergin för alla molekyler är som vi tidigare nämnt  $E_{th}=3kT/2$ . Om vi nu återvänder till ledningselektroner och hål i en halvledare och betraktar dessa laddningsbärare som moln av oskiljbara partiklar så gäller liknande regler. Dock tillkommer Paulis uteslutningsprincip som säger att två elektroner inte kan befinna sig i samma kvanttillstånd. Elektroner och hål följer därför istället Fermi-Diracs sannolikhetsfördelning,

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}},$$
(2.2)

där  $E_F$  är den så kallade ferminivån. På denna referensnivå  $E_F$  är sannolikheten att hitta en elektron lika med ½ och inte lika med ett som i Boltzmanns sannolikhetsfördelning. Fermi-Diracs sannolikhetsfördelning vid tre olika temperaturer visas i fig. 2.6. Sannolikheten för att hitta ett hål på energinivån E i valensbandet är lika med sannolikheten för att inte hitta en elektron, och ges således av  $f_p(E)=1-f(E)$ .



Fig. 2.6. Fermi-Diracs fördelningsfunktion.



Fig. 2.7. Boltzmann-approximationen.

Fermi-Diracs sannolikhetsfördelning är ganska snarlik Boltzmannfördelningen utom allra närmast ferminivån. Så snart  $\exp((E-EF)/kT) >> 1$ , d.v.s. i praktiken så snart  $E-E_F > 3kT$ , kan den approximeras med en Boltzmannfördelning

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}} \approx e^{-(E - E_F)/kT}.$$
 (2.3)

Se fig. 2.7. Villkoret för Boltzmannapproximationen är uppfylld för de flesta "normala" dopningar, exempelvis för kisel vid rumstemperatur är villkoret uppfyllt så länge dopningen är mindre än 10<sup>18</sup> cm<sup>-3</sup>. En halvledare som är så hårt dopad att villkoret för Boltzmannapproximationen inte längre är uppfyllt kallas degenererad.

#### 2.4 Tillståndstäthet

I förra avsnittet betraktade vi sannolikheten för att ett tillstånd med energin E ska vara besatt av en elektron. Vi kom fram till att ledningselektroner och hål följer Fermi-Diracs sannolikhetsfördelning men att denna för nästan alla dopningar kan approximeras med Boltzmanns sannolikhetsfördelning. I detta avsnitt ska vi gå vidare och ta reda på hur många tillstånd det finns att besätta på varje energinivå.

Låt oss först tillfälligt återvända till en gas av molekyler. Antalet molekyler som har en viss hastighet v i någon riktning, och därmed en viss energi E, kan man i en enkel hastighetsmodell tänka sig ges av arean av en sfär med radien v. Vi tänker oss då att vi kan dra en hastighetsvektor från sfärens centrum till varje punkt på sfärens yta. Molekylernas hastighetsfördelning kan då beskrivas med Maxwell-Boltzmanns sannolikhetsfördelning genom följande normaliserade uttryck:

$$f'(v) = \left(\frac{m}{2\pi kT}\right)^{3/2} 4\pi v^2 e^{-mv^2/(2kT)}, \qquad (2.4)$$

där normaliseringen gjorts så att sannolikheten för att en molekyl ska ha någon hastighet är ett. Molekylernas hastighetsfördelning i fyra olika ädelsgaser visas i fig. 2.8.



Fig. 2.8. Gasmolekylernas hastighetsfördelning i några olika ädelgaser.

Nu ska vi genomföra ett liknande resonemang för elektroner för att ta reda på deras energifördelning i ledningsbandet. Vi tänker oss då ledningselektronerna som partiklar med den effektiva massan  $m_e^*$  och så byter vi variabel från hastighet till impuls, eller rörelsemängd, enligt sambandet

$$p = m_e * v \,. \tag{2.5}$$

Volymen mellan två sfäriska skal med radierna p och p+dp i impulsrymden ges då av

$$dV = 4\pi p^2 dp. \tag{2.6}$$

Enligt den av de Broglie formulerade våg-partikeldualiteten besitter elektronen också elektromagnetiska vågegenskaper. Elektronens våglängd  $\lambda$  ges enligt de Broglies relation av uttrycket,

$$\lambda = \frac{h}{p},\tag{2.7}$$

där  $h=6,625 \cdot 10^{-34}$  J·s är Plancks konstant. Enligt villkoret för stående vågor är denna våglängd kvantiserad och i en kubisk halvledarkristall med sidan *L* gäller att,  $L=n\lambda$ , där *n* är ett heltal. Således är även rörelsemängden kvantiserad,

$$p = \pm \frac{nh}{L}.$$
(2.8)

De tillåtna rörelsemängderna är således åtskilda med h/L i impulsrymden och med två tillåtna spinn för varje rörelsemängd upptar varje tillstånd volymen  $h^3/2L^3$ . Antalet tillstånd (per volymsenhet) i skalvolymen  $4\pi p^2 dp$  ges då av

$$dn(p) = \frac{2}{h^3} 4\pi p^2 dp.$$
 (2.9)

Om vi sedan går över från rörelsemängd till energi enligt sambandet för rörelseenergi

$$E - E_c = \frac{m_e * v^2}{2} = \frac{p^2}{2m_e *},$$
(2.10)

får vi följande uttryck för tillståndstätheten i ledningsbandet

$$N_C(E) = \frac{4\pi}{h^3} \left(2m_e^*\right)^{3/2} \sqrt{E - E_C} . \qquad (2.11)$$

Tillståndstätheten ökar således med energin som  $\sqrt{(E-E_C)}$ . Elektronernas energifördelning i ledningsbandet ges nu av produkten av tillståndstätheten och sannolikheten för att tillståndet ska vara besatt av en elektron

$$dn(E) = N_c(E)f(E)dE , \qquad (2.12)$$

dvs av uttrycket

$$dn(E) \sim \sqrt{E - E_C} e^{-(E - E_C)/kT} dE$$
 (2.13)

Det här uttrycket har ett maximum för  $E-E_C=kT/2$ , dvs den mest sannolika elektronenergin är lika med halva den termiska energin. Medelvärdet av samtliga elektroners energi är dock tre gånger större, 3/2kT.

De tre funktionerna  $N_C(E)$ , f(E) och  $N_C(E)f(E)$  som funktion av energin E visas i fig. 2.9.

#### 2.5 Fermistatistik för hål och elektroner

Vi kan nu beräkna antalet elektroner i ledningsbandet genom att summera över alla energinivåer, d.v.s. genom att integrera över ledningsbandet



Fig. 2.9. Laddningsbärarnas energifördelning.

$$n = \int_{E_C}^{\infty} f(E) N_C(E) dE . \qquad (2.14)$$

Här är den övre integrationsgränsens exakta värde av mindre betydelse eftersom exponentialfunktionen snabbt går mot noll och den har därför, för enkelhets skull, satts till oändligheten. Om vi i nästa steg approximerar Fermi-Dirac-fördelningen med en Boltzmann-fördelning enligt (2.3) blir integrationen relativt enkel att genomföra. Villkoret för approximationens giltighet,  $E_C-E_F$ >kT, har diskuterats tidigare och det innebär att ferminivån måste ligga åtminstone 3-4kT under ledningsbandkanten. För att genomföra integrationen har vi använt oss av definitionen av gammafunktionen och att

$$\Gamma\left(\frac{3}{2}\right) = \int_{0}^{\infty} \sqrt{x} e^{-x} dx = \frac{\sqrt{\pi}}{2}.$$
 (2.15)

Resultatet av integrationen kan skrivas

$$n = N_C e^{-(E_C - E_F)/kT}, \qquad (2.16)$$

där

$$N_{C} = 2 \left( \frac{2\pi m_{e} * kT}{h^{2}} \right)^{3/2}.$$
 (2.17)

 $N_C$  kallas för den effektiva tillståndstätheten. Uttrycket (2.17) är lätt att komma ihåg om man noterar att det faktiskt utgör produkten av denna effektiva tillståndstäthet och sannolikheten enligt Boltzmann för att hitta en elektron i ett tillstånd vid ledningsbandkanten,

$$n = N_C f(E_C) . (2.18)$$

Den effektiva tillståndstätheten är helt enkelt det effektiva antalet tillstånd vi får om vi tänker oss att alla tillstånden vore flyttade till ledningsbandkanten. Rimligheten i detta betraktelsesätt stöds av att ledningsbektronerna faktiskt befinner sig väldigt nära ledningsbandkanten. Medelenergin är ju endast 3/2kT, en ganska liten energi jämfört med kisels energibandgap  $E_G$  som motsvarar 44 kT.

Motsvarande samband mellan hålkoncentrationen och ferminivån kan beräknas på liknande sätt. Tillståndstätheten för hål ges analogt som för elektroner av uttrycket

$$N_{V}(E) = \frac{4\pi}{h^{3}} (2m_{h} *)^{3/2} \sqrt{E_{V} - E} , \qquad (2.19)$$

där  $m_h^*$  är hålens effektiva massa i valensbandet. Hålens rörelseenergi i valensbandet ges av

$$E_{\nu} - E = \frac{m_h * \nu^2}{2} \,. \tag{2.20}$$

Sannolikheten för att hitta ett hål är nu lika med sannolikheten för att vi inte ska hitta en elektron, dvs

$$1 - f(E) = \frac{1}{1 + e^{(E_F - E)/kT}} \approx e^{-(E_F - E)/kT}, \qquad (2.21)$$

där vi återigen utnyttjar Boltzmannapproximationen. Efter integration av uttrycket

$$p = \int_{-\infty}^{E_V} (1 - f(E)) N_V(E) dE . \qquad (2.22)$$

får vi följande samband mellan hålkoncentrationen och ferminivåns läge

$$p = N_V e^{-(E_F - E_V)/kT}, \qquad (2.23)$$

där den effektiva tillståndstätheten vid valensbandkanten ges av

$$N_V = 2\left(\frac{2\pi m_h * kT}{h^2}\right). \tag{2.24}$$

Ett riktvärde för den effektiva tillståndstätheten vid lednings- och/eller valensbandkanten är 2,5  $\cdot 10^{19}$  cm<sup>-3</sup> (om vi för enkelhets skull använder den fria elektronens massa vid beräkningen, d.v.s. antar att  $m^*=m_0$ ). För noggrannare beräkningar bör vi för kisel använda  $m_e^*/m_0=1,18$  för elektroner och  $m_h^*/m_0=0,81$  för hål (för Ge är motsvarande värden 0,55 respektive 0,36).

**Exempel**: Beräkna sannolikheten för att hitta en elektron vid ledningsbandkanten och ferminivåns läge i en n-typ halvledare med dopningen  $N_D=10^{16}$  cm<sup>-3</sup>.

**Lösning**: Elektronkoncentrationen ges av  $n=N_D$ , dvs av dopningen. Sannolikheten för att hitta en elektron vid ledningsbandkanten ges då enligt (2.18) av

$$f(E_C) = \frac{n}{N_C} = \frac{N_D}{N_C} = 0,0004$$
.

Sannolikheten är således bara 0,4 promille, men å andra sidan är den effektiva tillståndstätheten mycket stor. Antalet ledningselektroner blir därför ändå ganska stort (eftersom det bestäms av dopningen).

Ferminivåns läge vid dopningen  $N_D=10^{16}$  cm<sup>-3</sup> ges av

$$E_C - E_F = kT \ln \frac{N_C}{N_D} = 0,20 \text{ eV}$$

dvs ferminivån ligger 0,2 eV under ledningsbandkanten (vilket också innebär att den ligger 0,35 eV över mitten av bandgapet). ■

Ferminivåns läge som funktion av dopningen för n-typ och p-typ kisel visas i fig. 2.10. Observera att modellen är styckevis linjär. Linjerna startar på mittnivån vid en dopning som motsvarar egentätheten  $n_i$  som vi ska diskutera i nästa avsnitt. Linjernas lutning,  $kT \ln 10 \approx 60$ meV per dekad, antyder att ferminivån ökar med cirka 60 meV för varje tiofaldig ökning av dopningen.



Fig. 2.10. Ferminivåns läge vs dopningen.

#### 2.6 Massverkans lag och egentätheten n<sub>i</sub>

I en intrinsisk halvledare benämns ferminivån  $E_{i_2}$  och denna intrinsiska nivå ligger nästan mitt i bandgapet. Det innebär att sannolikheten för att hitta en elektron vid ledningsbandkanten är lika stor som sannolikheten för att hitta ett hål vid valensbandkanten. Det är också precis vad vi förväntar oss av en intrinsisk halvledare där hål och elektroner alstras termiskt i form av hål-elektronpar. Med hjälp av (2.16) och (2.23) kan produkten av elektron- och hålkoncentrationerna i termisk jämvikt skrivas

$$np = N_V N_C e^{-(E_C - E_V)/kT} = N_V N_C e^{-E_G/kT}.$$
 (2.25)

Detta uttryck utgör *massverkans lag* som säger att *np*produkten i termisk jämvikt är konstant oavsett halvledarens dopning (ferminivån ingår ju inte i uttrycket). Massverkans lag kan skrivas på formen,

$$np = n_i^2, \qquad (2.26)$$

där  $n_i$  är den intrinsiska koncentrationen (egenkoncentrationen). Således kan vi genom att jämföra ekvationerna (2.25) och (2.26) identifiera egenkoncentrationen som

$$n_{i} = \sqrt{N_{V} N_{C}} e^{-E_{G}/2kT} .$$
 (2.27)

Den intrinsiska koncentrationen vid rumstemperatur är ungefär  $10^{10}$  cm<sup>-3</sup> i kisel (2·10<sup>13</sup> cm<sup>-3</sup> i Ge, och 3·10<sup>6</sup> cm<sup>-3</sup> i GaAs). Sannolikheten för att hitta en elektron i ett tillstånd vid ledningsbandet i en intrinsisk halvledare av kisel är således endast  $10^{10}/2,5\cdot10^{19}$ =4·10<sup>-10</sup>, d.v.s. en sannolikhet som är mindre än en på miljarden (<1 ppb).

Vi ser i (2.27) att egentätheten  $n_i$  påverkas av bandgapets storlek på så sätt att ju större bandgap, desto mindre egentäthet. Detta förefaller naturligt eftersom sannolikheten för att ett hål-elektronpar ska skapas genom termisk excitation blir allt mindre ju större energigapet är.

Ett generellt uttryck för egentätheten ges av

$$n_i = 2,5 \cdot 10^{19} \left(\frac{T}{300}\right)^{3/2} \left(\frac{\sqrt{m_e * m_h *}}{m_0}\right)^{3/2} e^{-E_G/2kT}$$
(2.28)

där  $m_e^*$  och  $m_h^*$  är ledningselektronernas och hålens respektive effektiva massor. Som vi nämnt tidigare är riktvärdet för den effektiva tillståndstätheten vid 300 K och  $m_0=m_e^*=m_h^*$ .

#### 2.7 Med egentätheten n<sub>i</sub> som referens

Vi har nu konstaterat att i en intrinsisk halvledare är  $n=p=n_i$  och att den intrinsiska ferminivån  $E_i$  ligger mitt i bandgapet. Ett exakt uttryck för  $E_i$  kan härledas ur följande två ekvationer:

$$\begin{cases} n_i = N_C e^{-(E_C - E_i)/kT} \\ n_i = N_V e^{-(E_i - E_V)/kT} \end{cases}$$
(2.29)

Det ger

$$E_{i} = \frac{E_{C} + E_{V}}{2} - kT \ln \sqrt{\frac{N_{C}}{N_{V}}} .$$
 (2.30)

Vi ser att den intrinsiska ferminivån ligger mitt i bandgapet om  $N_C=N_V$ . Vi kan nu använda den intrinsiska ferminivån som referensnivå istället för att använda bandkanterna  $E_C$  och  $E_V$ . Uttrycken för elektron- och hålkoncentrationerna i (2.16) respektive (2.23) med  $N_C$ ,  $N_V$ ,  $E_C$ och  $E_V$  som parametrar, d.v.s. uttrycken

$$\begin{cases} n = N_C e^{-(E_C - E_F)/kT} \\ p = N_V e^{-(E_F - E_V)/kT} \end{cases}$$
(2.31)

kan nu enkelt skrivas om med egentätheten  $n_i$  och den intrinsiska ferminivån  $E_i$  som parametrar. Vi får

$$\begin{cases} n = n_i e^{(E_F - E_i)/kT} \\ p = n_i e^{-(E_F - E_i)/kT} \end{cases}$$
(2.32)

Vi ser att för en n-typ halvledare  $(n>n_i)$  så ligger ferminivån ovanför intrinsiska nivån  $E_i$  och för en p-typ halvledare ligger den under, se fig. 2.9. Ju fler elektroner vi har i halvledaren, desto mer skjuts ferminivån uppåt. Ju fler hål vi har i halvledaren, desto mer skjuts ferminivån nedåt. Ferminivåns läge i förhållande till intrinsiska nivån definierar den s.k. fermienergin,  $E_F$ - $E_i$ . Fermienergin i en n-typ halvledare med dopningen  $N_D=10^{16}$  cm<sup>-3</sup> och  $n=N_D$ ges enligt (2.32) av uttrycket

$$E_F - E_i = kT \ln \frac{N_D}{n_i} = 0,35 \text{ eV}.$$
 (2.33)

Ferminivån ligger alltså 0,35 eV över mittnivån, och därmed också 0,20 eV under ledningsbandkanten precis som vi också redan beräknat i ett tidigare exempel. Ett enkelt banddiagram för en n-typ halvledare visas i fig. 2.10.



Fig. 2.10. Enkelt banddiagram för n-typ halvledare.

Fermienergin är också indirekt ett mått på elektronens utträdesarbete, dvs ett mått på den minsta energi som det krävs för att frigöra en elektron från ett material. I vårt energidiagram är utträdesarbetet den energi som krävs för att lyfta en elektron från ferminivån till den s.k vakuumnivån<sup>3</sup> för fria elektroner. Vakuumnivån används ofta som referensnivå när man ska jämföra egenskaper hos olika material. Exempelvis har en p-typ halvledare större utträdesarbete än en n-typ halvledare eftersom dess ferminivå ligger på en lägre energinivå. Man kan emellertid lika gärna, som vi gör här, använda den intrinsiska nivån som referensnivå. Den intrinsiska nivån i kisel ligger 4,6 eV under vakuumnivån. Vi ska strax se att skillnaden i fermienergier mellan p-typ och n-typ kisel ger upphov till en inbyggd kontaktpotential i en pn-övergång.

I samband med denna diskussion ska vi också passa på att definiera en halvledares fermipotential  $\phi_F$ . Denna potential har samma värde i [V] som fermienergin  $E_F$ - $E_i$  har i [eV] men den har motsatt tecken. Fermipotentialen fås genom att fermienergin divideras med elektronladdningen

$$\phi_F = \frac{E_F - E_i}{-q} = \frac{E_i - E_F}{q} = V_t \ln \frac{p}{n_i}, \qquad (2.34)$$

där  $V_t = kT/q$  är den termiska spänningen (d.v.s. 25 mV vid 290K).

Fermipotentialen kan ses som en kontaktpotential mellan en extrinsisk halvledare och dess intrinsiska motsvarighet. Fermipotentialen i en p-typ halvledare är positiv eftersom potentialen på p-sidan måste lyftas med en positiv spänning för att komma i nivå med den intrinsiska potentialen. Fermipotentialen i en n-typ halvledare däremot är negativ eftersom potentialen på n-sidan måste sänkas med en negativ spänning för att komma i nivå med den intrinsiska potentialen. Se fig. 2.11.

Även fermipotentialen avspeglar på sätt och vis ett materials utträdesarbete. Den positiva fermipotentialen hos p-typ kisel innebär ett större utträdesarbete, medan den negativa fermipotentialen hos n-typ kisel innebär ett mindre utträdesarbete än det för intrinsiskt kisel<sup>4</sup>.

Den inbyggda kontaktpotentialen  $V_{bi}$  i en pn-övergång som ett resultat av skillnaden mellan p- och n-sidornas olika fermipotentialer visas i fig. 2.12. Vi ser att potentialen på p-sidan är lägre än den på n-sidan och att potentialen på p-sidan måste öka med  $V_{bi}$  för att skillnaden ska elimineras. Om båda sidorna är dopade med  $10^{16}$ störatomer per cm<sup>3</sup> blir kontaktpotentialen enligt tidigare

<sup>4</sup> Utträdesarbetet för en halvledare som kisel ges av uttrycket  $q\Phi_S = q\Phi_i + q\Phi_F$ , där  $q\Phi_i = 4,6$  eV är utträdesarbetet för den intrinsiska halvledaren och där  $\Phi_F$  är den dopade halvledarens fermipotential. exempel  $V_{bi}$ =0,7V. Det är också ett typiskt värde att man måste lägga en spänning på ca 0,7V över en pn-övergång (diod) för att denna ska kunna leda ström.



Fig. 2.11. Fermipotentialskillnaden mellan p- och n-ty halvledare ger upphov till en inbyggd kontaktpotential en pn-övergång.

#### 2.8 Sammanfattning

I detta kapitel har vi infört en enkel energibandmodell för att kunna till fullo förstå vissa egenskaper hos halvledare. Vi har konstaterat att laddningsbärare i en halvledare är Fermi-Dirac-fördelade, men att denna sannolikhetsfördelning i de flesta fall kan approximeras med den matematiskt mycket enklare Boltzmannfördelningen. Vi har vidare konstaterat att ledningselektronerna rör sig i botten av ledningsbandet och att hålen rör sig överst i valensbandet. Laddningsbärarnas medelenergi är endast 3/2kT. Som en konsekvens av detta införde vi de effektiva tillståndstätheterna  $N_C$  och  $N_V$  för ledningsbandet respektive för valensbandet. Storleksordningen för dessa tillståndstätheter är i kisel vid rumstemperatur 2,5 10<sup>19</sup> cm<sup>-3</sup>.

Vi har även studerat hur dopningen bestämmer ferminivåns läge, och att denna ligger över den intrinsiska mittnivån om halvledaren är av n-typ och att den ligger under den intrinsiska mittnivån om halvledaren är av ptyp. Till sist studerade vi också kort hur ferminivåns läge påverkar ett materials utträdesarbete och hur skillnaden i utträdesarbete mellan två material ger upphov till en tentialskillnad, en s.k. kontaktpotential, mellan de två materialen.

<sup>&</sup>lt;sup>3</sup> Vakuumnivån är energinivån för en fri elektron utanför kristallen, dvs för en elektron i perfekt vakuum.

#### 2.9 Övningsexempel

Fermi-Dirac statistiken introducerades 1929 av Enrico Fermi och Paul Dirac. Det finns en animering som visar F-D statistikens temperaturberoende [ $^{5}$ ].

- 1. Vilken sannolkhhet anges genom Fermi-Diracs sannolikhetsfördelning?
  - a. Hur definieras ferminivån?
  - b. Under vilka förutsättningar kan F-D statistiken approximeras med en Boltzmannstatistik?
  - c. Hur stor är den termiska energin E=kT, där k är Boltzmanns konstant och T är den absoluta temperaturen? Hur stor är motsvarande termiska spänning  $V_t=kT/q$ ?
  - d. Hur stort är kisels bandgap vid 300 K och hur många *kT* motsvarar detta bandgap? Hur stor är den termiska energin i procent av bandgapet?
  - e. Vilket är medelvärdet av elektronernas rörelseenergi?
  - f. På vilken energinivå finns det flest elektroner i ledningsbandet och på vilken energinivå finns det flest hål i valensbandet?
- 2. Var i bandgapet ligger
  - a. Den intrinsiska nivån
  - b. Donatornivån?
  - c. Acceptornivån?
- 3. Vad menas med degenererat kisel? Hur markerar man att en halvledare är degenererad? Är det något "fel" på en degenererad halvledare?
- 4. Hur avgör man om en halvledare är intrinsisk, extrinsisk, eller degenererat extrinsisk?
- 5. Hur ser sambandet mellan elektronkoncentration, egentäthet och ferminivåns läge ut?
  - a. Hur stor är den effektiva tillståndstätheten  $N_C$  vid bandkanten om elektronens effektiva massa är lika med den fria elektronmassan?

- b. Vad menas med effektiva tillståndstätheten, d.v.s. hur har  $N_C$  respektive  $N_V$  fått sina namn?
- 6. Under vilka förutsättningar kan F-D statistiken approximeras med en Boltzmannstatistik?
- Skissa i ett semilogaritmiskt diagram ferminivåns läge som funktion av dopningen i intervallet 10<sup>8</sup><N<10<sup>18</sup> cm<sup>-3</sup> vid temperaturerna 300K! Utnyttja följande "hållpunkter" för att förtydliga skissen!
  - a. För vilka dopningar är halvledaren att betrakta som intrinsisk, dvs  $E_F = E_i$ ?
  - b. Hur mycket ökar ferminivån om dopningen tiofaldigas i en extrinsisk halvledare?
  - c. Vilken är den övre begränsningen för den enkla linjära modellen?
  - d. Hur ser motsvarande diagram ut vid 450 K?
- 8. Om vi istället skulle plotta ferminivåns läge som funktion av temperaturen [K] vilka "hållpunkter" skulle vi då utnyttjaför vår linjariserade modell?
- 9. Väteatomens jonisationsenergi, dvs utträdesarbete, är 13,6 eV. Hur stort är utträdesarbetet för intrinsiskt kisel?
- 10. Vilket har störst utträdesarbete, p-typ eller n-typ kisel?
- 11. Är fermipotentialen för p-typ kisel positiv eller negativ?
- 12. Material med stora bandgap är intressanta i många tillämpningar för sin höga spänningstålighet. Ett aktuellt halvledarmaterial med stort bandgap är GaN. Vid mätningar på provtillverkad GaN mäter man en resistivitet på ungefär 10  $\Omega$ cm vid rumstemperatur och man ser att resistiviteten ökar med temperaturen upp till åtminstone 1400 K. Gör en grov uppskattning av storleken på bandgapet för GaN om vi använder oss av den fria elektronens massa!

<sup>&</sup>lt;sup>5</sup> <u>http://nanohub.org/resources/8822</u>

#### **3. TEMPERATURBEROENDE**

I det här kapitlet ska vi studera temperaturberoendet hos mobilitet, ledningsförmåga, egentäthet, och fermipotential. Temperaturberoendet är en viktig parameter som ofta kan användas för att avgöra vilken av två konkurrerande mekanismer som dominerar i en halvledare. Sålunda kan man avgöra om en halvledare är extrinsisk eller intrinsisk genom att studera om resistiviteten ökar eller minskar med stigande temperatur. År halvledaren extrinsisk så ökar resistiviteten med ökande temperatur (positiv temperaturkoefficient, PTC), medan resistiviteten minskar om halvledaren är intrinsisk (negativ temperaturkoefficient, NTC). I en extrinsisk halvledare är det rörlighetens temperaturberoende som bestämmer resistivitetens temperaturberoende. I en intrinsisk halvledare är det egentäthetens temperaturberoende som bestämmer resistivitetens temperaturberoende. Rörlighetens temperaturberoende är en mycket viktig parameter för prestanda även hos integrerade kretsar. En integrerad CMOS-krets, typ processorkretsen i en hemdator, är inte alls lika snabb vid 100°C som den är vid rumstemperatur på grund av den längre RC-tidskonstanten.

#### 3.1 Rörlighetens temperaturberoende

Låt oss börja med att betrakta rörlighetens temperaturberoende mer i detalj. Som vi nämnt i ett tidigare avsnitt så dominerar gitterrörligheten i en lågdopad halvledare, medan inverkan av störrörligheten är försumbar. Således dominerar gitterrörlighetens temperaturberoende i en lågdopad halvledare medan störrörlighetens temperaturberoende inverkar först vid högre dopningsnivåer.

#### 3.1.1 Gitterörlighetens temperaturberoende

En mycket enkel temperaturmodell för gitterrörligheten  $\mu_L$ (eng. impurity mobility) innebär att  $\mu_L \sim T^{3/2}$ , vilket motsvarar temperaturberoendet hos medeltiden mellan kollisioner med kristallgittret enligt följande förenklade resonemang: vi utgår från att laddningsbärarna far omkring slumpmässigt i kristallen med en termisk hastighet som ges av  $mv_{th}^2/2=3/2kT$ . Den termiska hastigheten ökar enligt denna modell som roten ur temperaturen, dvs  $v_{th} \sim T^{1/2}$ . Den medelsträcka som en laddningsbärare hinner tillryggalägga mellan två gitterkollisioner betecknar vi med  $\langle s \rangle$  och det är väl inte orimligt att anta att denna sträcka blir kortare och kortare ju högre temperaturen är. Låt oss som en enkel ansats anta att <s>~1/T. Medeltiden mellan kollisioner ges då av farttriangeln till  $\langle t \rangle = \langle s \rangle / \langle v \rangle$ , d.v.s. medeltiden  $\langle t \rangle$  mellan kollisoner blir proportionell mot  $T^{3/2}$ . Eftersom rörligheten är proportionell mot medeltiden mellan kollisioner så får även rörligheten detta temperaturberoende.

#### 3.1.2 Störrörlighetens temperaturberoende

Vid högre dopningsnivåer kommer rörligheten att påverkas också av kollisioner med störatomerna. Störrörligheten  $\mu_l$  (*eng.* impurity mobility) är annorlunda till sin karaktär än gitterrörligheten eftersom den styrs av coulombkraften mellan två laddade partiklar. I Fig. 3.1 visas hur en rörlig laddningsbärare, i det här fallet en elektron, sprids i närheten av en laddad störatom i kristallgittret. Störrörlighetens temperaturberoende är motsatt mot gitterrörlighetens, d.v.s.  $\mu_{l} \sim T^{3/2}$ . Vid högre temperaturer har laddningsbärarna större rörelseenergi (3/2kT) och därmed högre hastighet ( $\sqrt{3kT/m}$ ). Därför påverkas de inte lika mycket av coulombkrafterna och sprids mindre.

Rörlighetens totala temperaturberoende visas i fig. 3.2. Vi ser ett ganska entydigt temperaturberoende för låga dopningar ( $10^{14}$  cm<sup>-3</sup>), men ju högre dopningen är desto större blir inslaget av störkollisioner. Därmed blir också temperaturberoendet ett annat när två olika mekanismer med motsatta temperaturberoenden tenderar att "balansera ut" varandra. Vid dopningen  $10^{18}$  cm<sup>-3</sup> är rörligheten nästan helt temperaturoberoende i det aktuella temperatur-intervallet.



Fig. 3.1. En elektron "kolliderar" med acceptorer och donatorer på ungefär samma sätt trots motsatta laddningar.



*Fig. 3.2. Elektronrörligheten i kisel vs temperaturen.* 

I en fotnot i kapitel 1 gavs följande modell för rörligheten

$$\mu = \mu_{\min} + \frac{\mu_L - \mu_{\min}}{1 + (N/N_0)^{\alpha}}.$$
(3.1)

Varje parameter i denna formel kan antas ha ett temperaturberoende enligt

$$P = P_{300} \left(\frac{T}{300}\right)^{\eta}, \qquad (3.2)$$

där  $P_{300}$  är parametervärdet vid 300 K. De olika parametrarnas (något förenklade) temperaturberoende ges för kisel i tabell 3.1.

Parameter	Elektroner	Hål	Temperatur-
	P <sub>300</sub>	P <sub>300</sub>	exponent η
$N_0[cm^{-3}]$	$1,3^{-}10^{17}$	$2,35\cdot 10^{17}$	2,4
$\mu_{\min}$	90	50	-0,6
$\mu_L$ - $\mu_{min}$	1260	400	-2,33
α	≈1 (0,91)	≈1 (0,88)	-0,15

Tabell. 3.1 Rörlighetsparametrarnas temperaturkoefficienter.

För ett temperaturberoende som beskrivs av  $\mu \sim T^{\eta}$  kan temperaturkoefficienten elegant beräknas om vi logaritmerar uttrycket innan vi deriverar. Vi får då

$$\frac{d\mu/dT}{\mu} = \frac{\eta}{T} \,. \tag{3.3}$$

För  $\eta$ =-1,5 och T=300K får vi därmed en temperaturkoefficient på -0,5%/°C.

#### 3.2 Egentäthetens temperaturberoende

Egentätheten har enligt tidigare diskussioner ett exponentiella temperaturberoende eftersom den ges av uttrycket

$$n_{i} = \sqrt{N'_{C} N'_{V}} \left(\frac{T}{300}\right)^{3/2} e^{-E_{G}/2kT} .$$
(3.4)

där  $N'_C$  och  $N'_V$  är de effektiva tillståndstätheterna vid 300 K. Detta temperaturberoende visas för kisel, germanium, galliumarsenid och kiselkarbid i fig. 3.3. Ur diagrammet kan vi läsa av egentätheten vid rumstemperatur för de fyra halvledarmaterialen och dessa värden listas i tabell 3.2. I tabell 3.3 visas de effektiva massor och bandgap som har använts vid beräkningarna i fig. 3.3.

kisel	germanium	galliumarsenid	kiselkarbid
Si	Ge	GaAs	SiC
$10^{10}  \mathrm{cm}^{-3}$	$2^{\cdot}10^{13} \text{ cm}^{-3}$	$3.10^6 \text{ cm}^{-3}$	3.10-5 cm-3

Tabell. 3.2. Egentätheten vid rumstemperatur.

	$m_{e}^{*}/m_{0}$	$m_{h}^{*}/m_{0}$	E <sub>G</sub>
Si	1,18	0,81	1,1
Ge	0,55	0,36	0,66
GaAs	0,066	0,52	1,4

Tabell. 3.3. Bandgap (300 K) och effektiva massor.

Om vi logaritmerar och deriverar egentäthetens temperaturberoende på motsvarande sätt som nyss för rörligheten fås ett temperaturberoende som ges av

$$\frac{dn_i/dT}{n_i} = \frac{3/2}{T} + \frac{E_G/2kT}{T}.$$
 (3.5)

För kisel betyder det att egentäthetens temperaturkoefficient vid 300K är 7,5%/°C. Här bidrar  $T^{3/2}$  med 0,5 %/°C, medan exponentialuttrycket bidrar med 7 %/°C. Intrinsic Concentration vs Temperature



Fig. 3.3. Egentätheten n<sub>i</sub> i kisel, Ge, GaAs och SiC vs T.

#### 3.3 Ledningsförmågans temperaturberoende

När vi nu vet både rörlighetens och egentäthetens temperaturberoende kan vi också beräkna ledningsförmågans, eller alternativt resistivitetens, temperaturberoende. Resultatet av en sådan analys visas i fig. 3.4.

#### Conductivity vs Temperature



*Fig. 3.4. Ledningsförmågan i n-typ kisel som funktion av temperaturen för olika dopningar.* 

I diagrammet syns tydligt skillnaden mellan extrinsisk och intrinsisk halvledare. Ledningsförmågan hos en extrinsisk halvledare minskar med temperaturen eftersom rörligheten minskar, medan ledningsförmågan hos en intrinsisk halvledare ökar med temperaturen (röda linjen) trots att rörligheten minskar eftersom antalet laddningsbärare ökar snabbare än rörligheten avtar. Det framgår också tydligt att en extrinsisk halvledare blir intrinsisk bara temperaturen är tillräckligt hög. Ju hårdare dopad halvledaren är, desto högre temperatur krävs för att halvledaren ska bli intrinsisk, och därmed förlora sina extrinsiska egenskaper. Övergången mellan extrinsisk och intrinsisk halvledare ges av det ungefärliga villkoret att  $n_i=N_D$ .

Vi ser också i fig. 3.4 att ledningsförmågans temperaturkoefficient i extrinsiskt kisel är störst för lågdopat kisel och att denna sedan minskar med ökad dopning när störrörligheten får större inverkan.

#### 3.4 Fermienergins temperaturberoende

Eftersom Boltzmannssannolikheten är kraftigt temperaturberoende är också en halvledares fermienergi temperaturberoende. Ferminivåns läge kan som tidigare uttryckas med hjälp av effektiva tillståndstätheten. För n-typ halvledare skriver vi så länge  $n=N_D$  att

$$E_{C} - E_{F} = kT \ln \frac{N_{C}}{N_{D}} = kT \left( \ln N_{C} - \ln N_{D} \right).$$
(3.5)

Eftersom Boltzmannssannolikheten är kraftigt temperaturberoende är också en halvledares fermienergi kraftigt temperaturberoende. Ferminivåns läge kan som tidigare uttryckas med hjälp av effektiva tillståndstätheten. För en n-typ halvledare skriver vi så länge halvledaren är extrinsisk, dvs så länge  $n=N_D$ , att

$$E_{C} - E_{F} = kT \ln \frac{N_{C}}{N_{D}} = kT \left( \ln N_{C} - \ln N_{D} \right).$$
(3.5)

Med hjälp av detta uttryck kan ferminivåns läge plottas som funktion av dopningen med temperaturen som parameter i ett semilogaritmiskt diagram som det i fig. 3.5. Vi har redan i ett tidigare kapitel gjort detta vid 300 K. I ett sådant diagram ser vi ferminivåns läge beskrivas av en rät linje som går från den intrinsiska mittnivån  $E_F=E_i$  för  $N_D=n_i$  till bandkanten  $E_C=E_F$  för  $N_D=N_C$ . Linjens lutning är kTln10 meV/dekad, ett mått som anger hur mycket högre ferminivån hamnar för varje tiofaldig ökning i dopningen. Vid 300 K motsvarar detta cirka 60 meV/dekad,  $n_i=10$  cm<sup>-3</sup> och  $N_C=2,5\cdot10^{19}$  cm<sup>-3</sup>.

När vi plottar sådana här förenklade diagram som bygger på styckevis linjära modeller så får vi inte glömma modellens begränsningar. Sålunda förutsätts att Boltzmannapproximationen gäller dvs att  $E_C - E_F >> kT$ .

Vi kan också plotta ferminivåns läge som funktion av temperaturen med dopningen som parameter. Ett sådant diagram visas i fig. 3.6.

Något förenklat ser vi i figuren en samling i det närmaste räta linjer som startar vid lednings- respektive valensbandkanten vid absoluta nollpunkten T=0 och som når den intrinsiska mittnivån  $E_F=E_i$  vid en temperatur då egentätheten är lika med den aktuella dopningen,  $n_i=N_D$ .



*Fig. 3.5. Fermienergin som funktion av dopningen i n- och p-typ kisel vid olika temperaturer.* 

#### 3.5 Sammanfattning.

I detta kapitel har vi studerat temperaturberoendet hos några viktiga halvledarparametrar som rörlighet, egentäthet, ledningsförmåga och fermienergi. Vi har tagit fram en enkel modell för gitterrörlighetens temperaturberoende, och sedan konstaterat att störrörligheten har ett lika stort, men motsatt temperaturberoende. Det betyder att en lågdopad halvledare har ett kraftigare temperaturberoende än en högdopad halvledare. Det teoretiska temperaturberoendet är  $T^{-3/2}$ , men i praktiken ligger det i kisel nog när-mare  $T^{-7/3}$ . Vid normala temperaturer runt rumstemperatur är rörlighetens temperaturberoende sådant att rörligheten minskar med ökande temperatur. Det betyder att ledningsförmågan i en extrinsisk halvledare också avtar med ökande temperatur. Det ger resistansen en positiv temperaturkoefficient (PTC). I en intrinsisk halvledare ökar däremot ledningsförmågan med temperaturen på grund av egentäthetens kraftiga (exponentiella) temperaturberoende. Det ger resistansen hos en intrinsisk, egenledande halvledare en negativ temperaturkoefficient (NTC).



*Fig. 3.6. Fermienergin som funktion av temperaturen med dopningen som parameter.* 

#### 4. PN-ÖVERGÅNGEN

I detta kapitel ska vi betrakta pn-övergången och dess likriktande egenskaper ur ett halvledarperspektiv. Vi förutsätter då att läsaren är väl förtrogen med en diods likriktande egenskaper och hur man med papper och penna kan analysera enkla diodkretsar med hjälp av enkla styckevis linjära diodmodeller. Beroende på spänningens storlek kan man antingen använda modellen för en ideal diod som leder utan spänningsfall, eller en något mer detaljerad modell med inbyggd kontaktpotential och dynamisk resistans. I en sådan modell kan dioden vid analys av en diodkrets ersättas med en spänningskälla med inbyggd källresistans. Se fig. 4.1.



*Fig. 4.1. Definition av ideal diod (t.v.) och styckevis linjär modell för icke-ideal diod (t.h.)* 

Vid simulering av diodkretsar med datorhjälpmedel används dock mer komplicerade icke-linjära modeller i de simuleringsverktyg typ Spice (eng. Simulation Program with Integrated Circuit Emphasis) som finns för analys av integrerade kretsar.

I detta kapitel ska vi jobba vidare med den ickelinjära diodmodell som går under namnet ideala diodekvationen. Denna modell är en ickelinjär modell med tre modellparametrar och vi kommer att visa hur man extraherar modellparametrar till denna modell från experimentella mätdata. Ideala diodekvationen har ett spänningsberoende som är exponentiellt - ett helt nytt fenomen för de flesta av oss. Vi kommer att se att detta exponentiella spänningsberoende kommer från Boltzmanns fördelningsfunktion. När vi studerar de fysikaliska förhållandena kring den inbyggda potentialbarriären kommer vi att finna att ideala diodekvationen är en modell som innehåller en för oss helt ny ickelinjär strömbegränsningsmekanism, nämligen diffusion på grund av koncentrationsskillnader. Strömmen begränsas inte som tidigare av drift i ett elektriskt fält utan av andra fysikaliska fenomen som diffusion och rekombination.

Vi kommer också att studera diodens genombrott i backriktningen samt dess kapacitiva egenskaper. Vi kommer därvid att introducera ett nytt begrepp i form av en ickelinjär kapacitans på grund av diodens spärrskikt. Slutligen kommer vi att runda av kapitlet med laddningskontrollmodellen – en modell med vilken man kan studera hur snabbt en diod kan switchas mellan ledande och oledande tillstånd.

#### 4.1 pn-övergången

Pn-övergången är den enklaste och vanligaste halvledarkomponenten. Den förekommer både som enskilda dioder, s k diskreta komponenter och som komponent eller del av komponent i integrerade kretsar med hundratal miljoner transistorer. Som vi ska se senare används pnövergångar även för att isolera komponenter som transistorer ifrån varandra i integrerade kretsar.

Pn-övergångar tillverkas genom att störatomer av motsatt typ införs i ett begränsat område av en dopad kiselskiva. Pn-övergångar kan tillverkas i n-typ halvledarskivor genom att ett begränsat diodområde p-dopas med trevärt bor, eller i en p-typ halvledare genom att ett område n-dopas med fosfor eller arsenik. Dopningen sker vanligen genom s.k. jonimplantation, dvs genom att högenergetiska störjoner skjuts in i halvledarkristallen. Efter värmebehandling tar dessa störatomer sedan plats i kristallgittret.

I fig. 4.2 visas en pn-övergång där ett begränsat område av en n-typ kiselskiva dopats till p-typ. Till vänster visas en kiselskiva i vilken man tillverkat en stor mängd likadana dioder. Figuren försöker också illustrera hur kiselskivan sågas upp i individuella dioder som plockas upp från skivan en och en för kapsling som diskreta komponenter. Tvärsnittsfiguren är tänkt att illustrera att en diod i verkligheten är en tredimensionell komponent, men att vi här i vår teoribildning enbart kommer att betrakta endimensionella modeller av pn-övergången (se den streckade linjen). För att få enkla modeller som kan öka förståelsen för komponentens funktion och för att få modeller som som är lämpliga för handberäkningar bortser vi således från alla två- och tredimensionella kanteffekter. För enkelhets skull förutsätts dessutom alla pnövergångar vara abrupta, dvs dopningen förutsätts ändras



Fig. 4.2. pn-övergången – principskiss och kretssymbol.

abrupt från p-typ till n-typ (dopningen  $N_A$  för x < 0 och  $N_D$  för x > 0. Även om en verklig dopningsprofil inte är helt abrupt är denna förenklade modell förvånansvärt bra. Vi ser också i figuren en endimensionell principfigur för pn-övergången, samt kretssymbolen för en diod.

#### 4.2 Parameterextraktion från mätningar

Den första övning vi ska genomföra är s. k. parameterextraktion, dvs vi ska bestämma värden för diodens modellparametrar genom att passa modellen till experimentella mätdata.

Typiska resultat från experimentella mätningar på några verkliga dioder visas i fig. 4.3. Fyra av de fem dioderna börjar leda för spänningar runt 0,7V och de är kiseldioder. Den femte dioden som börjar leda redan vid 0,3V är en germaniumdiod med ganska stor serieresistans. Att germaniumdioden börjar leda för lägre spänningar än kiseldioder beror på att germanium har ett mindre bandgap än kisel (och därmed en lägre potentialbarriär).



Fig. 4.3. Diodmätningar i framriktningen.

**Exempel 4.1**. Bestäm modellparametrarna  $V_{bi}$  och  $r_D$  i den styckevis linjära diodmodellen för diod 1 i fig. 4.3.



Fig. 4.4. Styckevis linjär diodmodell passad till mätdata i framriktningen.

**Lösning**: Från mätdata (uppmätt 100 mA vid 1,0 V och 45 mA vid 0,90 V) kan resistansen  $r_D$  bestämmas till 1,8  $\Omega$  och den extrapolerade kontaktpotentialen till  $V_{bi}$ =0,82V. Passning mellan modell och mätdata visas i fig. 4.4.

**Exempel 4.2.** Bestäm modellparametrarna  $I_0$ , n, och  $R_S$  i ideala diodmodellen för diod 1 med hjälp av samma mätdata.

**Lösning**. För att lösa detta problem måste vi plotta om ström/spänningskarakteristiken i semilogaritmisk skala, se fig. 4.5. Idealitetskonstanten n kan nu bestämmas från lutningen hos den räta linje som beskriver det exponentiella spänningsberoendet i ideala diodekvationen.



Fig. 4.5. Exponentiell diodmodell och mätdata i framriktningen.

Ideala diodekvationen ges av

$$I = I_0 \left( e^{(V_A - R_S I) / nV_t} - 1 \right).$$
(4.1)

där idealitetsfaktorn n helst ska vara n=1 för att dioden ska betraktas som riktigt ideal. Med hjälp av två lämpligt valda mätpunkter i diodens exponentiella spänningsintervall får vi följande uttryck

$$n = \frac{\left(V_{GS2} - V_{GS1}\right)}{V_{th} \left(\ln I_{DS2} / I_{DS1}\right)} = \frac{(0, 8 - 0, 27)}{0,025 \cdot 8 \cdot \ln 10} = 1, 1.$$
(4.2)

Vi ser att dioden är ganska nära ideal. Med obetydligt sämre passning till mätdata skulle man kunna välja n=1. På samma sätt kan  $I_0$  relativt en referensnivå på 1 A bestämmas ur följande ekvation med hjälp av samma två mätpunkter

$$\ln I_{OFF} = \frac{V_{GS1} \ln I_{DS2} - V_{GS2} \ln I_{DS1}}{V_{CS1} - V_{CS2}}$$
(4.3)

Vi får med insatta mätvärden

$$\ln I_{OFF} = \frac{0.8 \cdot \ln 10^{-10} - 0.27 \cdot \ln 10^{-2}}{0.8 - 0.27} = -33.5 \quad (4.4)$$

Således får vi  $I_0$ =e<sup>-33,5</sup> A = 3 fA. Diodens teoretiska läckström är således 3 fA, ett värde som ligger långt under mätsystemets noggrannhet, men ett värde som genererar en modell med god överensstämmelse med mätdata i spänningsintervallet 0.4 < V < 0.7 V. Se fig. 4.6.



Fig. 4.6. Passning mellan modell och mätdata för diod.

Återstår att bestämma serieresistansen  $R_s$ . Ur diagrammet kan vi utläsa att serieresistansen åstadkommer ett extra spänningsfall på 0,15 V på strömnivån 100 mA. Det motsvarar ett resistansvärde på  $R_s=1,5 \Omega$ .

I den styckevis linjära diodmodellen hade vi bestämt resistansen till 1,8  $\Omega$ . Går det ihop med en serieresistans på 1,5  $\Omega$ ? Jo, diodens dynamiska resistans utan serieresistans ges av

$$r = \frac{nV_{th}}{I} = \frac{1,1\cdot25}{100} = 0,275.$$
 (4.5)

Den totala dynamiska resistansen ges således av

$$r_D = R_S + r = 1,5 + 0,275 = 1,775 \approx 1,8 \ \Omega,$$
 (4.6)

vilket stämmer bra överens med det tidigare extraherade värdet. Överensstämmelsen mellan modell och mätdata framgår nu av fig. 4.7. ■

#### 4.3 Inbyggd kontaktpotential.

Existensen av den inbyggda kontaktpotentialen i en diod kan förklaras med att pn-övergångens båda sidor har olika fermipotentialer, dvs olika utträdesarbeten. Materialen är helt enkelt inte i elektrisk balans med varandra på grund av de olika dopningarna. En inbyggd spänning  $V_{bi}$  uppstår således på grund av denna skillnad mellan p- och nsidornas fermipotentialer och vi kan teckna uttrycket

$$V_{bi} = \phi_p - \phi_n = V_t \left( \ln \frac{N_A}{n_i} - \ln \frac{n_i}{N_D} \right), \qquad (4.7)$$

där  $N_A$  och  $N_D$  är dopningarna på p- respektive n-sidan. Ett banddiagram som visar hur en intern kontaktpotential uppstår i övergången mellan två material med olika fermipotential visas i fig. 4.8.



Fig. 4.7. Passning mellan modell och mätdata för diod.



Fig. 4.8. Banddiagram för pn-övergången i termisk jämvikt.

När man ska rita sådana banddiagram är det lämpligt att man först ritar ut ferminivån som den horisontella referenslinje den är. I termisk jämvikt, alltså när halvledaren är i jämvikt med den omgivande temperaturen utan belysning och utan pålagd spänning, uppträder ferminivån som en konstant jämviktsnivå genom materialet. Lednings- och valensbanden på p- och n-sidorna ritas sedan in på varje sida i förhållande till ferminivån. Till sist förenas de båda sidornas banddiagram genom spänningsfallet över pnövergången. I fig. 4.8 är övergången ritad linjär men i ett senare avsnitt ska vi i detalj diskutera hur denna övergång sker. Vi ska också diskutera vad som sker med ferminivån när vi lägger på en spänning.

**Exempel 4.3:** Beräkna den inbyggda kontaktpotentialen i en pn-övergång där dopningen på p-sidan är  $N_A=10^{17}$  cm<sup>-3</sup> och på n-sidan  $N_D=10^{14}$  cm<sup>-3</sup>.

Lösning: Vi får kontaktpotentialen med hjälp av (4.7):

$$V_{bi} = \phi_p - \phi_n = V_t \left( \ln \frac{N_A}{n_i} - \ln \frac{n_i}{N_D} \right) =$$
  
= 0, 4 + 0, 23 = 0, 63V

Om vi nu förspänner pn-övergången kommer potentialbarriären att variera med spänningen. Den kommer att minska eller öka beroende på hur vi förspänner pn-övergången, dvs om vi framspänner eller backspänner. Detta illustreras i fig. 4.9.



Fig. 4.9. Potentialbarriären  $V_j$  i en pn-övergång som är a) framspänd, b) utan förspänning, och c) backspänd.

Vi ser att potentialbarriären minskar så att ström kan flyta när pn-övergången framspänns och att potentialbarriären ökar i storlek när vi backspänner pn-övergången så att dioden spärrar. Potentialbarriärens storlek ges av

$$V_i = V_{bi} - V, \tag{4.8}$$

där V är den pålagda spänningen och  $V_{bi}$  som förut den inbyggda kontaktpotentialen. Först när  $V_i \approx 0$ , och barriären i stort sett är eliminerad, kan dioden leda "full" ström, en ström som då enbart begränsas av serieresistansen. I översta figuren visas hur strömmen kan ta sig över den låga barriären, medan den i nedersta figuren stöter på patrull i form av en snudd på "oöverstiglig" barriär.

#### 4.4 Utarmningsområdet

I banddiagrammet i fig. 4.9 ser vi att det elektriska fältet är ritat försumbart på p- och n-sidorna medan det inte är försumbart i övergångsområdet mellan de båda sidorna. Detta förhållande har Gauss synpunkter på, för enligt Gauss lag kan det elektriska fältet inte ändra sig utan närvaro av laddningar.

Om man då analyserar en pn-övergång närmare så finner man att området närmast pn-övergången är utarmat på rörliga laddningsbärare. Kvar finns finns endast okompenserade störatomer; dvs negativt laddad acceptorer på p-sidan sedan hålen närmast övergången försvunnit och positivt laddade donatorer på n-sidan sedan elektronerna närmast övergången försvunnit. En störatom benämns vara okompenserad om den har blivit av med sin neutraliserande laddningsbärare och istället blivit en laddad störjon. Den laddningsfördelning som uppstår i pn-övergången visas i fig. 4.10. Vi ser en negativ konstant rymdladdning  $-qN_A$  med utbredningen  $W_p$  på p-sidan, och en positiv rymdladdning  $qN_D$  med utbredningen  $W_n$  på n-sidan.

Totalt sett råder laddningsneutralitet, dvs

$$qN_D W_n = qN_A W_p. aga{4.9}$$



Fig. 4.10. Laddningsfördelning i en pn-övergång.

Gauss lag säger att fältändringen mellan två punkter  $x_1$  och  $x_2$  ges av den laddning som finns mellan punkterna dividerat med dielektricitetskonstanten. Det kan formuleras på följande sätt:

$$\varepsilon (E_2 - E_1) = \int_{x_1}^{x_2} \rho dx$$
 (4.10)

Det betyder således att fältet i punkten x=0 på andra sidan den negativa laddningen -Q' är  $Q'/\varepsilon$  mindre än fältet i punkten  $x=-W_p$ . På samma sätt är fältet i punkten  $x=W_n$  $Q'/\varepsilon$  större än fältet i punkten x=0. Konstanta rymdladdningar ger fält som varierar linjärt med avståndet. Det resulterar i den fältbild som visas i fig. 4.11, där det maximala fältet vid x=0 ges av

$$E_{max} = \frac{qN_D W_n}{\varepsilon_{Si}} = \frac{qN_A W_p}{\varepsilon_{Si}}.$$
(4.11)



Fig. 4.11. Elektrisk fältbild i en pn-övergång.

Sambandet mellan elektriskt fält och spänning ger då sambandet mellanpotentialbarriärens storlek, det elektriska fälte och utarmningsområdets utbredning

$$V_j = E_{\max} \frac{\left(W_p + W_n\right)}{2}.$$
(4.12)

Potentialbarriärens storlek bestäms helt enkelt av maxfältet multiplicerat med fältets utbredning dividerat med 2 – dvs av fälttriangelns area.

**Återigen:** att fältet varierar linjärt med x och potentialen varierar kvadratiskt med x är en *följd av de konstanta rymdladdningarna -qN*<sub>A</sub> respektive *qN*<sub>D</sub>.

Gauss lag på differentiell form i en dimension lyder

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon} \,. \tag{4.13}$$

Genom att integrera Gauss lag med konstant rymdladdning  $-qN_A$  på p-sidan och  $+qN_D$  på n-sidan kan vi beräkna det elektriska fältet som funktion av x. Fältet kommer då att ges av uttrycket

$$E(x) = \begin{cases} -E_{max} \left( 1 + \frac{x}{W_p} \right) & -W_p \le x \le 0\\ -E_{max} \left( 1 - \frac{x}{W_n} \right) & 0 \le x \le W_n \end{cases}$$
(4.14)

När nu det elektriska fältet är känt kan vi enkelt beräkna potentialen genom att använda definitionen på elektriskt fält,

$$\frac{dV}{dx} = -E . ag{4.15}$$

och integrera detta uttryck över utarmningsområdet. Vi får

$$V(x) = \begin{cases} -E_{\max}\left(1 + \frac{x}{2W_p}\right)x & -W_p \le x \le 0\\ +E_{\max}\left(1 - \frac{x}{2W_n}\right)x & 0 \le x \le W_n \end{cases}$$
(4.16)

där vi godtyckligt satt V(0)=0. Oavsett alla formeldetaljer här så ges den totala potentialbarriären  $V_i$  som tidigare av

$$V_j = \frac{E_{\max}W}{2}, \qquad (4.17)$$

där  $W=W_p+W_n$ . Och det sa vi ju direkt i (4.12) att det totala spänningsfallet ges av fälttriangelns area. Det nya som vi tillfört med den detaljerade härledningen är att vi nu vet att potentialkurvan är sammansatt av två skarvade parabler. Se fig. 4.12.



Fig. 4.12. Potentialen i en pn-övergångs utarmningsområde.

Speciellt enkelt att hantera är specialfallet med en enkelsidigt abrupt pn-övergång, dvs en pn-övergång där  $N_A \gg N_D$ , och följaktligen  $W_n \gg W_p$ , dvs  $W = W_n$ . Se fig. 4.13. I detta fall reduceras våra två grundläggande ekvationer till

$$\begin{cases} V_{j} = \frac{E_{max}W}{2} \\ E_{max} = \frac{qN_{D}W}{\varepsilon_{si}} \end{cases}$$
(4.18)

Dessa två uttryck ger nu utarmningsområdets utbredning som funktion av pålagd spänning genom att  $E_{max}$  elimineras. Vi får

$$W = X_D \sqrt{V_j} , \qquad (4.19)$$



Fig. 4.13. Förenklad laddnings, fält- och potentialbild i en enkelsidig abrupt  $p^+n$ -övergång.

där potentialbarriärens storlek  $V_j$  bestäms av den pålagda spänningen V enligt

$$V_j = V_{bi} - V,$$
 (4.20)

och där konstanten  $X_D$  ges av<sup>6</sup>

$$X_D = \sqrt{\frac{2\varepsilon}{qN_D}} \,. \tag{4.21}$$

**Exempel 4.4**: Beräkna utarmningsområdets utbredning i en p<sup>+</sup>n-diod med  $N_D=10^{14}$  cm<sup>-3</sup> och  $V_{bi}=0.81$  V dels utan pålagd spänning och dels vid en pålagd backspänning på 10 V. Hur stort är det maximala elektriska fältet vid 10 V backspänning?

**Lösning:** Med kisels relativa permitivitet  $\varepsilon_{Si}$ =12 och  $\varepsilon_0$ =8,85·10<sup>-12</sup> F/m får vi för dopningen  $N_D$ =10<sup>14</sup> cm<sup>-3</sup> att  $X_D$ =3,6 µm/V<sup>1/2</sup>. Det ger W=3,6· $\sqrt{0}$ ,81=3,24 µm utan pålagd spänning och W=3,6· $\sqrt{10}$ ,8=12 µm vid V=-10V.

På grund av att fältbilden är "triangulär" får vi inte glömma att det maximala fältet för en viss spänning är dubbelt så stort som det hade varit om det hade varit konstant, dvs  $E_{max}=2V_{f}/W=1,8 V/\mu m$ , vilket är en ganska hög fältstyrka (18 kV/cm) men ändå mycket mindre än det maximala fält 24 V/ $\mu$ m som dioden tål utan genombrott. **■ Summering**: I detta avsnitt har vi lärt oss att rita potentialbarriären i en pn-övergång genom att "skarva" två parabler, se fig. 4.12. Vi vet att potentialbarriären minskar när dioden framspänns och att den ökar när dioden backspänns. Vi vet också att utarmningsområdet växer i utbredning vid ökande backspänning.

<sup>6</sup> Det generella uttrycket för  $X_D$  i en tvåsidig pn-övergång

#### 4.5 Genombrott

Om vi eliminerar W istället för  $E_{max}$  med hjälp av ekvationerna i (4.18) får vi följande samband mellan potentialbarriären och det elektriska fältet:

$$V_j = \frac{\varepsilon E_{\max}^2}{2qN_D}.$$
(4.22)

För varje dopningsnivå i en p<sup>+</sup>n- eller n<sup>+</sup>p-diod finns det ett maximalt fält, eller kritiskt fält, vid vilket det sker strömgenombrott i dioden. Detta kritiska genombrottsfältet är av storleksordningen 400 kV/cm, vilket motsvarar 40 V/ $\mu$ m. En diod som ska tåla 200 V måste således ha utarmningsområde på minst 10  $\mu$ m vid denna spänning för att fältet inte ska bli för stort.



Fig. 4.14. Genombrottsspänning som funktion av dopningen i en enkelsidig abrupt pn-övergång i kisel.

Uttrycket (4.22) kan nu användas för att beräkna pn-övergångens genombrottsspänning BV som funktion av dopningen på den lågdopade sidan när vi känner det kritiska fältet för genombrott  $E_{BD}$ , dvs

$$BV = \frac{\varepsilon E_{BD}^2}{2qN_D} - V_{bi} \approx \frac{\varepsilon E_{BD}^2}{2qN_D}, \qquad (4.23)$$

För genombrottsspänningar större än cirka tio volt, dvs för BV>10 V, kan vi förenkla uttrycket genom att försumma  $V_{bi}$ . Genombrottsspänningen som funktion av bakgrundsdopningen visas i fig. 4.14. Som synes är den teoretiska formeln inte helt "rättvisande" utan dopningsberoendet är snarare  $1/N^{0.75}$  än 1/N. Detta beror på att det kritiska fältet för genombrott är något dopningsberoende<sup>7</sup>.

**Exempel 4.5:** Beräkna genombrottsspänningen i en pnövergång där dopningen på p-sidan är  $N_A=10^{17}$  cm<sup>-3</sup> och på n-sidan  $N_D=10^{14}$  cm<sup>-3</sup>.

**Lösning:** Genombrottsfältet vid dopningen 10<sup>14</sup> cm<sup>-3</sup> på den lågdopade sidan kan beräknas till 240 kV/cm (10.lltså något lägre än nominella 400 kV/cm) och därmed kan vi beräkna genombrottsspänningen till

<sup>7</sup> En något bättre dopningsberoende modell för genombrottsfältet ges av  $E_{BD} = \frac{400}{1 - 0,33 \cdot \log \frac{N_D}{10^{16}}} [\text{kV/cm}].$ 

$$BV = \frac{\varepsilon E_{BD}^2}{2qN_D} = \frac{12 \cdot 8,85 \cdot 10^{-14} \cdot (240 \cdot 10^3)^2}{2 \cdot 1,6 \cdot 10^{-19} \cdot 10^{14}} = 1900V. \quad \blacksquare$$

Det finns dioder att köpa med genombrottsspänningar i hela intervallet - från några få volt till tusentals volt. Vissa dioder är avsedda att tåla höga spänningar utan genombrott, andra är avsedda att arbeta vid genombrottsspänningen, t ex som spänningsregulatorer. Sådana dioder kallas zenerdioder efter en av de två mekanismer som ligger bakom genombrottet, nämligen zenergenombrott.

Geombrottskarakteristika för tre av våra experimentella dioder visas i fig. 4.15. Vi ser att tre av dioderna har genombrott för relativt låga backspänningar, -2,3 V, -12V, och -44 V. En av dioderna tål riktigt höga spänningar, typ -600 V. Andra dioder som 1N4007 tål minst 1000 V.

Genombrottsspänningen definieras oftast som den backspänning,  $V_{BD}$  (*eng.* breakdown voltage) eller  $V_Z$  (*eng.* Zener voltage) som resulterar i en viss ström  $I_Z$ , t ex  $I_Z=5$ mA. I databladet anges t. ex.  $V_Z=6,3V$  @  $I_Z=5$  mA. Genombrottets "skarphet" definieras genom zenerresistansen  $R_Z$  (som även den specificeras vid strömnivån  $I_Z$ ).

Det finns två olika fysikaliska genombrottsmekanismer i dioder, nämligen zenergenombrott och lavingenombrott.



Fig. 4.15. Genombrott i diodernas backriktning.

#### 4.5.1 Zenergenombrott

Zenergenombrott, som är den mekanism som dominerar för genombrottsspänningar mindre än 6 V, innebär att det höga fältet i spärrskiktet kan få valenselektroner att tunnla genom det tunna spärrskiktet till ledningsbandet på nsidan. För att tunnling ska kunna ske måste spärrskiktet vara tunt, d.v.s. pn-övergången måste vara högdopad. Vid zenergenombrott minskar genombrottsspänningen vid ökande temperatur eftersom mekanismen är bandgapsberoende och bandgapet minskar med ökande temperatur. Temperaturkoefficienten är typiskt -0,1 %/°C i storleksordning. Se fig. 4.16.

Genombrott i backriktningen



*Fig. 4.16. Zenergenombrott genom tunnling av valenselektroner.* 

#### 4.5.2 Lavingenombrott

I lågdopade pn-övergångar sker genombrott redan vid ett kritiskt fält som är lägre än det som krävs för tunnling, d.v.s. för zenergenombrott. Det fysikaliska fenomenet bakom denna typ av genombrott kallas lavinmultiplikation och den illustreras i fig. 4.17.



Fig. 4.17. Genombrott genom lavinmultiplikation.

Elektroner och hål som befinner sig i området mellan noch p-sidornautsätts för så högt elektriskt fält att den energi som de tillförs mellan kollisionerna är tillräcklig för att skapa nya hål-elektronpar när de kolliderar med kristallgittret. Dessa nya hål-elektronpar får så småningom även de så hög energi att de kan skapa nya hål-elektronpar, och på det viset är lavinen snart i rullning.

Lavingenombrottsspänningen ökar med ökande temperatur. Det är rörlighetens temperaturberoende som slår igenom och rörligheten minskar med ökande temperatur som vi diskuterat tidigare. Med minskande rörlighet krävs ökande fält och därmed större spänningar för att uppnå en viss hastighet (energi). Därmed ökar också genombrottsspänningen. Temperaturkoefficienten är typiskt av storleksordningen 0,1 %/°C.

#### 4.6 Diodens ström/spänningskarakteristik

Nu har vi etablerat kunskap om det utarmningsområde som finns mellan p- och n-sidorna i en pn-övergång. Vi har tagit fram modeller med hjälp av vilka vi kan räkna ut utarmningsområdets utbredning som funktion av den pålagda spänningen. Vi har också konstaterat att för små strömstyrkor sådana att spänningsfallet över serieresistansen kan försummas, så faller hela den pålagda spänningen över utarmningsområdet. Pn-övergångens band- och potentialdiagram i förhållande till dess utarmningsområde illustreras i fig. 4.18.



Fig. 4.18.Band och potentialdiagram i förhållande till utarmningsområdet i en pn-övergång.

Vi vet att vid framspänningar minskar utarmningsområdets utbredning, medan vid backspänningar så ökar dess utbredning. Då kan vi illustrera pn-övergången och dess utarmningsområde på följande sätt, se fig. 4.19.



Fig. 4.19. Utarmningsområdets tjocklek vid olika förspän-

Eftersom vi har en inbyggd potentialbarriär mellan p- och n-sidorna har vi också ett elektriskt fält, precis som vi nyss beräknade när vi behandlade utarmningsområdet i ett föregående avsnitt. Detta gäller även när dioden inte är spänningssatt. I detta fält flyter förstås driftströmmar, bestående av både hål- och elektronströmmar. Dessa strömmar kan beskrivas med våra transportekvationer

ningar.

$$\begin{cases} J_p = qp\mu_p E\\ J_n = qn\mu_n E \end{cases}$$
(4.24)

Även utan pålagd spänning flyter det således driftströmmar i utarmningsområdet, men ändå flyter det ingen nettoström. Det måste således finns motbalanserande strömmar som gör att nettoflödet av laddningsbärare genom utarmningsområdet är noll. Dessa motbalanserande strömmar utgörs av diffusionsströmmar.

Enligt Ficks första lag är en diffusionsström proportionell mot koncentrationsgradienten och kan därmed beskrivas med följande uttryck:

$$\begin{cases} J_p = -qD_p \frac{dp}{dx} \\ J_n = +qD_n \frac{dn}{dx} \end{cases}$$
(4.25)

där  $D_n$  och  $D_p$  är elektronernas respektive hålens diffusionskonstanter. En negativ koncentrationsgradient ger upphov till ett flöde av partiklar i positiv riktning - något som innebär positiv hålström och negativ elektronström. Enligt Einsteins relation är diffusionskonstanten proportionell mot rörligheten enligt

$$\begin{cases} D_p = V_t \mu_p \\ D_n = V_t \mu_n \end{cases}$$
(4.26)

I närvaro av diffusion måste våra transportekvationer således kompletteras med diffusionsströmmen och således skrivas om på formen

$$\begin{cases} J_p = qp\mu_p E - qD_p \frac{dp}{dx} \\ J_n = qn\mu_n E + qD_n \frac{dn}{dx} \end{cases}$$
(4.27)

#### Balans mellan drift och diffusion i utarmningsområdet

Med dessa transportekvationer kan vi nu beräkna den inbyggda kontaktpotential som får drift- och diffusionsströmmarna i utarmningsområdet att balansera ut varandra. Om vi sätter  $J_p=0$  och  $J_n=0$  i (4.27) får vi

$$\begin{cases} qp\mu_p E = qD_p \frac{dp}{dx} \\ qn\mu_n E = -qD_n \frac{dn}{dx} \end{cases} \Rightarrow \begin{cases} Edx = V_t \frac{dp}{p} \\ Edx = -V_t \frac{dn}{n} \end{cases}$$
(4.28)

Om man separerar variabler och integrerar dessa ekvationer mellan utarmningsområdets båda kanter finner man att den inbyggda potentialbarriären ges av uttrycket

$$V_{bi} = V_{t} \ln \frac{N_{A} N_{D}}{n_{i}^{2}}, \qquad (4.29)$$

vilket är exakt samma uttryck för kontaktpotentialen som vi fick i ett tidigare avsnitt när vi beräknade den som skillnaden mellan p- och n-sidans fermipotentialer. Således har vi fått en sorts bekräftelse på att den inbyggda kontaktpotentialen blir precis så stor som krävs för att driftströmmarna ska motbalansera de diffusionsströmmar som blir följden av koncentrationsskillnaderna mellan poch n-sidorna.

#### Minoritetsbärarinjektion i framspänd diod

Om vi nu framspänner dioden kommer potentialbarriären att sänkas, och därmed kommer diffusionsströmmarna att ta överhanden när de motriktade driftströmmarna minskar när fältet minskar. Vi får då ett nettoströmflöde av hål från p-sidan genom utarmningsområdet till n-sidan och på motsvarande sätt ett nettoflöde av elektroner från n-sidan till p-sidan. I den framspända pn-övergången får vi således på båda sidor en injektion av minoritetsbärare. Dessa minoritetsbärare fortsätter sedan att diffundera djupare drivna av den genom injektionen uppkomna koncentrationsgradienten.

Dessa flöden av minoritetsbärare kommer emellertid succesivt att minska när minoritetsbäraröverskottet tunnas ut allt eftersom minoritetsbärarna försvinner genom rekombination med de väntande majoritetsbärarna. Vi ska nu visa att det injicerade minoritetsbäraröverskottet dels beror exponentiellt av den pålagda spänningen, och dels att det avklingar exponentiellt med den tillryggalagda sträckan på ett sätt som kan beskrivas med den genomsnittliga diffusionslängden *L*.

Vi ska således visa att

$$\begin{cases} p_n(x) - p_{n0} = p_{n0} \left( e^{V/V_t} - 1 \right) e^{-x/L_p} & \text{på n-sidan} \\ n_p(x) - n_{p0} = n_{p0} \left( e^{V/V_t} - 1 \right) e^{x/L_n} & \text{på p-sidan} \end{cases}$$
(4.30)

När vi väl vet minoritetsbäraröverskotten kan vi enkelt beräkna diffusionsströmmarna med hjälp av (4.25). Vi får

$$\begin{cases} J_{p}(x \ge 0^{+}) = qD_{p} \frac{p_{n0}}{L_{p}} \left( e^{V/V_{t}} - 1 \right) e^{-x/L_{p}} \\ J_{n}(x \le 0^{-}) = qD_{n} \frac{n_{p0}}{L_{n}} \left( e^{V/V_{t}} - 1 \right) e^{x/L_{n}} \end{cases},$$
(4.31)

där  $x=0^{-}$  och  $x=0^{+}$  betecknar utarmningsområdets båda kanter. Hålkoncentrationen och hålströmmen på n-sidan visas i fig. 4.20. Genom att anta att utarmningsområdet är så tunt att  $J_p$  och  $J_n$  inte ändras nämnvärt i utarmningsområdet kan vi beräkna den totala strömmen genom att addera hål- och elektronströmmarna vid x=0. Vi får då den totala diodströmmen

$$I = A\left(J_n(0^+) + J_p(0^-)\right) = I_0\left(e^{V/V_t} - 1\right), \qquad (4.32)$$

vilket är den ideala diodekvationen där diodmättnadsströmmen ges av

$$I_{0} = qA\left(\frac{D_{p}}{L_{p}}p_{n0} + \frac{D_{n}}{L_{n}}n_{p0}\right).$$
 (4.33)

De flesta dioder är nu enkelsidiga  $p^+n$ - eller  $n^+p$ -övergångar vilket leder till en avsevärd förenkling vad gäller uttrycket för diodmättnadsströmmen,

$$I_{0} = \begin{cases} qA \frac{D_{p}}{L_{p}} p_{n0} & \text{för } p^{+}n\text{-dioder} \\ qA \frac{D_{n}}{L_{n}} n_{p0} & \text{för } n^{+}p\text{-dioder} \end{cases}$$
(4.34)



Fig. 4.20. Exponentiellt avklingande hålöverskott enligt diffusionsekvationen och motsvarande hålströmstäthet.

Den formella gången vid härledning av uttrycken för minoritetsbäraröverskotten i (4.30) omfattar flera steg. Först måste vi ta hjälp av kontinuitetsekvationerna för hål och elektroner. Dessa ekvationer motsvarar Ficks andra lag kompletterade med en term som beskriver minoritetsbärarnas rekombination. Enklast tänkbara modell innebär att rekombinationstakten antas vara proportionell mot minoritetsbäraröverskottet. Med en sådan modell luktar det redan exponentiellt avklingande minoritetsbärarkoncentrationer lång väg. Kontinuitetsekvationerna kan, efter införande av denna matematiska sänka som rekombinationen innebär, beskrivas med uttrycken

$$\begin{cases} \frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} - r_p & r_p = \frac{p - p_{n0}}{\tau_p}, \\ \frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} - r_n & r_n = \frac{n - n_{p0}}{\tau_n}, \end{cases}$$
(4.35)

där  $r_p$  och  $r_n$  är rekombinationstakterna för hål respektive elektroner och där  $\tau_p$  och  $\tau_n$  är minoritetsbärarnas livslängder. I stationärtillstånd får vi

$$-\frac{1}{q}\frac{J_p}{dx} = \frac{p - p_{n0}}{\tau_p}$$

$$\frac{1}{q}\frac{dJ_n}{dx} = \frac{n - n_{p0}}{\tau_n}$$
(4.36)

Nästa steg utgår från att minoritetsbärarströmmarna kan antas vara diffusionsströmmar. Det verkar inte orimligt att försumma minoritetsbärarnas driftströmmar eftersom de är just minoritetsbärare, dvs så få till antalet. Minoritetsbärarnas koncentrationsgradienter är dock inte försumbara. Kontinuitetsekvationerna omvandlas nu till diffusionsekvationerna för minoritetsbärare enligt

$$\begin{cases} \frac{\partial^2 p}{dx^2} - \frac{p - p_{n0}}{D_p \tau_p} = 0\\ \frac{\partial^2 n}{dx^2} - \frac{n - n_{p0}}{D_p \tau_n} = 0 \end{cases}$$
(4.37)

Lösningarna till dessa andra ordningens linjära differentialekvationer utgörs av följande exponentiellt avklingande minoritetsbäraröverskott som karakteriseras av diffusionslängderna  $L_p = \sqrt{D_p \tau_p}$  och  $L_n = \sqrt{D_n \tau_n}$  enligt

$$\begin{cases} p_n(x) - p_{n0} = (p_n(0^+) - p_{n0})e^{-x/L_p} \text{ på n-sidan} \\ n_p(x) - n_{p0} = (n_p(0^-) - n_{p0})e^{x/L_n} \text{ på p-sidan} \end{cases}. (4.38)$$

Återstår nu att visa att randvillkoren ges av

$$\begin{cases} p_n(0^+) = p_{n0} e^{V/V_t} \\ n_p(0^-) = n_{p0} e^{V/V_t} \end{cases}$$
(4.39)

Dessa samband brukar betecknas som "the law of the junction" och bygger förstås på Boltzmanns exponentiella fördelningsfunktion. Som vi tidigare visat splittrar den pålagda spänningen upp ferminivån i två kvasiferminivåer  $E_{Fp}$  och  $E_{Fn}$  för hål och elektroner där energiskillnaden bestäms av den pålagda spänningen,  $qV=E_{Fn}-E_{Fp}$ . Se fig 4.21. Massverkans lag gäller därför inte längre i utarmningsområdet utan vi får istället

$$pn = n_i^2 e^{V/V_i}.$$
 (4.40)

Om minoritetsbärarinjektionen är att betrakta som liten, dvs om för hål  $p_n(0) \ll N_D$ , s.k. *low-level injection*, kan majoritetsbärarkoncentrationen vid utarmningsområdets kant anses vara opåverkad av spänningen. Med  $n=N_D$  i (4.40) får vi då följande randvärde för antalet hål vid  $x=0^+$ 

$$p_n(0^+)N_D = n_i^2 e^{V/V_i}$$
, dvs  $p_n(0^+) = p_{n0}e^{V/V_i}$ . (4.41)

På samma sätt får vi randvärdet för elektroner på p-sidan



Fig. 4.21. Banddiagram för framspänd pn-övergång med illustration av kvasiferminivå.

**Exempel 4.6:** Hur stor är hålens diffusionslängd i en p<sup>+</sup>ndiod med tvärsnittsarean 0,1 mm<sup>2</sup>? Vilken minoritetsbärarlivslängd motsvarar det? Antag  $I_0=3$  fA och BV=12 V.

**Lösning:** Eftersom genombrottsspänningen för dioden är ca 12 V så är dopningen på n-sidan av storleksordningen  $N_D=10^{17}$  cm<sup>-3</sup>, se fig. 4.14. Då vet vi också minoritetsbärarkoncentrationen  $p_{n0}=n_i^2/N_D=10^3$  cm<sup>-3</sup>. Med  $I_0=3$  fA får vi för denna p<sup>+</sup>n-diod att hålens diffusionslängd blir

$$L_p = qAD_p \frac{p_{n0}}{I_0} = 6,7 \mu m.$$

Vid denna beräkning har vi också utnyttjat vetskapen att  $\mu_p$ =500 cm<sup>2</sup>/Vs vilket enligt Einsteins relation medför att  $D_p=V_t\mu_p=12,5$  cm<sup>2</sup>/s. Den beräknade diffusionslängden motsvarar minoritetsbärarlivslängden  $\tau_p=L_p^2/D_p=35$  ns.

Sammanfattningsvis visas hål- och elektronströmmarna i tre olika typer av pn-övergångar i fig. 4.22.


Fig. 4.22. Hålströmmen i tre pn-övergångar med olika dopningsförhållanden.

## 4.7 Avvikelser från ideal diffusionsbegränsning

Denna diodmodell som leder fram till den ideala diodekvationen med idealitetsfaktorn *n*=1 bygger på ren diffusionsbegränsning av strömmen. Utarmningsområdet har förutsatts vara försumbart tunt. Eftersom vi har beräknat den totala strömmen genom att summera hål- och elektronströmmarna vid utarmningsområdets respektive kant så har vi helt enkelt förutsatt att såväl hål- som elektronströmmarna är konstanta i utarmningsområdet, se fig. 4.2. All rekombination av hål- och elektronpar i utarmningsområdet anses försumbar. Därför kan vi beräkna den totala strömmen genom att addera minoritetsbärarströmmarna på utarmningsområdets respektive kanter.

Detta är givetvis ett inte oacceptabelt resonemang när dioden är så framspänd att vi kan försumma utarmningsområdets utbredning. Men om man sänker framspänningen ökar utarmningsområdets utbredning och rekombinationen i utarmningsområdet kan i vissa dioder inte längre försummas. Det leder till att idealitetsfaktorn ökar.

Vissa dioder som är avsedda att tåla höga backspänningar är så lågdopade att deras utarmningsområden inte är försumbara ens vid stor framspänning. I sådana dioder kan den tidigare diffusionsbegränsningen vara försumbar och istället sker all rekombination i utarmningsområdet. I sådana dioder är idealitetsfaktorn n=2. Se fig. 4.23.

I vissa andra dioder kan man urskilja att idealitetsfaktorn för små framspänningar är lika med två, för att sedan för större framspänningar anta det ideala värdet n=1. Man ser alltså tydligt den framspänning där utarmningsområdets inverkan blir försumbar, se fig. 4.24.

En strömmodell för en sådan diod skulle kunna skrivas

$$I = I_{01} \left( e^{(V_A - R_S I)/V_t} - 1 \right) + I_{02} \left( e^{(V_A - R_S I)/2V_t} - 1 \right) (4.43)$$



Fig. 4.23. pn-övergång där större delen av hål- och elektronströmmarna rekombinerar i utarmningsområdet.

där  $I_{02}$  är den rekombinationsbegränsade diodmättnadsströmmen och  $V_A$  den pålagda spänningen. I backriktningen,  $V_A < 0$ , flyter strömmen  $-I_{02}$  genom en sådan diod. Kretsekvivalenten för en sådan diod visas i fig. 4.25.



Fig. 4.24. Jämförelse mellan två dioder med idealitetsfaktorerna n=1 (diod 1) och n=2 (diod 3).



Fig. 4.25. Kretsmodell för diod med såväl diffusions- som rekombinationsbegränsad ström.

## 4.8 Diodkarakteristikens temperaturberoende

Eftersom dioder inte alltid arbetar vid rumstemperatur utan ofta utsätts för stora temperaturvariationer, framför allt i avancerade processorchips för t ex persondatorer, ska vi också kort beröra diodens temperaturberoende.

Låt oss börja med att betrakta temperaturberoendet hos diodmättnadsströmmen  $I_0$ . Från ideala diodekvationen vet vi att  $I_0$  för en diffusionsbegränsad diod är proportionell mot  $n_i^2$  eftersom den är proportionell mot minoritetsbärarkoncentrationen som i sin tur är proportionell mot  $n_i^2$ . Men även diffusionskonstanten är temperaturberoende såväl genom termiska spänningen  $V_t$  som genom rörligheten. Temperaturberoendet för såväl ideala som ickeideala diodmättnadsströmmar i backriktningen kan sammanfattas med följande generella modell

$$I_0 \sim T^{-m} e^{-E_G / nkT} \,. \tag{4.44}$$

där m och n är två parametrar som oftast ges empiriskt grundade värden.

Om vi logaritmerar (4.39) och sedan deriverar med avseende på temperaturen så får vi följande uttryck för temperaturberoendet uttryckt i %/°C:

$$\frac{dI_0 / dT}{I_0} = \frac{m}{T} + \frac{E_G / kT}{nT}.$$
(4.45)

För en kiseldiod med  $E_G$ =44kT vid 300K och typiska värdena *m*=3/2 och *n*=2 får vi temperaturkänsligheten till 7,5%/°C. En praktisk tumregel som man ofta ser är att läckströmmen  $I_0$  fördubblas för var tionde grads temperaturökning (eftersom 1,07<sup>10</sup>≈2).

I diodens framriktning kan vi på motsvarande sätt sammanfatta diodspänningen på följande sätt

$$V = n \frac{kT}{q} \ln \frac{I}{I_0}.$$
(4.46)

Om vi deriverar med avseende på temperaturen vid konstant ström så får vi följande uttryck för diodspänningens temperaturberoende

$$\frac{dV}{dT} = \frac{V}{T} - nV_t \frac{dI_0 / dT}{I_0} \approx \frac{V - E_G / q}{T} .$$
(4.47)

Vid framspänningen V=0,5 V och T=300 K får vi i grova tal att den spänning som krävs för att hålla diodströmmen konstant minskar med  $dV/dT\approx-2mV/^{\circ}C$ . En konservativ tumregel som man ofta stöter på anger som typiskt riktvärde en temperaturkoefficient på -2,5 mV/°C. Diodens temperaturberoende illustreras i fig. 4.26.

# 4.8.1 Temperaturkompensering

Det är intressant att notera att zenerdioder med en genombrottsspänning på cirka 6 V har en temperaturkoefficient som är nästan noll. I detta spänningsintervall har vi ungefär lika mycket zener- som lavingenombrott, och de två fysikaliska mekanismernas temperaturberoenden tar ut varandra. Önskar man få en låg temperaturkoefficient vid en högre genombrottsspänning måste temperaturkompenserade dioder användas. Dessa består av en zenerdiod i serie med en eller flera framspända dioder. Dessa har en negativ temperaturkoefficient som motverkar lavingenombrottets positiva temperaturkoefficient. Den framspända diodens temperaturkoefficient är, som vi noterade i förra avsnittet, -2 mV/°C. Temperaturkompenserade zenerdioder kan fås som en komponent i samma kapsel, ofta med utmärkt temperaturstabilitet över stora temperaturområden.



Fig. 4.26. Diodströmmens temperaturberoende.

# 4.9 Diodens kapacitans

I detta avsnitt ska vi kort beröra diodens kapacitiva egenskaper. En diod i backriktningen leder ingen ström, men den har ett stort internt elektriskt fält och med detta en icke-försumbar elektrisk laddning. Nu vet vi att denna laddning utgörs av okompenserade störatomer som bildar ett utarmningsområde. Ju mer vi backspänner dioden ju mer växer utarmningsområdets utbredning.

En p<sup>+</sup>n-övergång som är backspänd med en viss spänning V har ett utarmningsområde med en viss utbredning W, huvudsakligen på den lågdopade n-sidan. Utarmningsladdningen ges av  $Q=qN_DW$ . Vi vet att utarmningsområdets utbredning ges av

$$W = X_D \sqrt{V_{bi} - V}, \tag{4.48}$$

och utarmningsladdningen har således via utarmningsområdets utbredning ett icke-linjärt beroende av backspänningen. Småsignalkapacitansen i en godtycklig arbetspunkt definieras genom

$$C_j(V) = -\frac{dQ}{dV},\tag{4.49}$$

en definition som resulterar i den icke-linjära kapacitansen

$$C_j(V) = \frac{\varepsilon}{W_0}.$$
(4.50)

Detta uttryck är inget annat än den traditionella plattkondensatorformeln! Den enda skillnaden är att vi har ett avstånd W mellan plattorna som är elektriskt styrbart genom backspänningen. Vanligtvis beskrivs spärrskiktskapacitansens spänningsberoende på följande sätt genom "nollkapacitansen"  $C_i(0)$  genom uttrycket

$$C_{j}(V) = \frac{C_{j}(0)}{\sqrt{1 - V/V_{bi}}}.$$
(4.51)

Utarmningsladdningens och spärrskiktskapacitansens spänningsberoenden illustreras i fig. 4.27.



Fig. 4.27. Laddning och småsignalkapacitans.

Fig. 4.28 visar utarmningsområdet i en p<sup>+</sup>n-diod som är backspänd med en viss likspänning. Ovanpå denna likspänning finns överlagrat en sinusformad småsignalspänning dV som ger upphov till en (rödmarkerad) variation i utarmningsområdets utbredning, dW, huvudsakligen på den lågdopade n-sidan.



# Fig. 4.28. Variationer i utarmningsområdets utbredning vid viss backspänning pga en överlagrad småsignalspänning.

Denna variation dW i utarmningsområdets utbredning motsvarar en positiv laddning dQ på n-sidan och en lika stor negativ laddning -dQ på p-sidan. Avståndet mellan dessa två "plattladdningar" utgörs av utarmningsområdets utbredning W. Enligt Gauss lag ger dessa laddningar upphov till en fältändrning dE i utarmningsområdet,

$$dE = \frac{-dQ}{\varepsilon_{Si}}.$$
(4.52)

Denna fältändring orsakas förstås av småsignalspänningen dV och ges definitionsmässigt av

$$dV = dE \cdot W. \tag{4.53}$$

Om vi eliminerar fältändringen *dE* mellan dessa två ekvationer får vi småsignalkapacitansen

$$C_j(V) = -\frac{dQ}{dV} = \frac{\varepsilon_{Si}}{W}.$$
(4.54)

Notera att vi inte har gjort några som helst antaganden om dopningarna i dioden, dessa behöver inte nödvändigtvis vara konstanta eller abrupta.



Fig. 4.29. Uppmätt diodkapacitans vs backspänning.

**Exempel 4.9.** Beräkna spärrskiktskapacitansen utan pålagd backspänning för en p<sup>+</sup>n-övergång med dopningen  $10^{18}$  cm<sup>-3</sup> på p-sidan och  $10^{15}$  cm<sup>-3</sup> på n-sidan.

Lösning. Diodens inbyggda potentialbarriär är

$$V_{bi} = V_t \ln(10^{33}/10^{20}) = 0,75 \text{ V}.$$

Utarmningsområdets utbredning utan pålagd spänning kan beräknas till  $W=1 \mu m$ . Spärrskiktskapacitansen blir då

 $C_i(0) = \varepsilon_{Si}/W = 12.8,85 \cdot 10^{-12}/10^{-6} \approx 0.1 \text{ mF/m}^2$ 

vilket motsvarar 100 pF/mm<sup>2</sup> eller 0,1 fF/µm<sup>2</sup>.

Ett exempel på uppmätt diodkapacitans visas i fig. 4.29 medan ett ekvivalent småsignalschema för dioden i back-riktningen visas i fig. 4.30.



Fig. 4.30. Ekvivalentschema för diod i backriktningen.

Vid kapacitansmätningar kan dioden tolkas som en seriekombination  $C_S/R_S$  eller som en parallellkombination  $C_P/R_P$ , där  $R_S$  är serieresistansen och  $R_P$  parallellresistansen. För en icke-läckande diod kan  $R_P$  anses oändlig.

#### 4.10 Diodens transientegenskaper

Avslutningsvis ska betrakta diodens transientegenskaper. Hittills har vi enbart betraktat diodens statiska egenskaper, dvs dess egenskaper vid pålagda likspänningar eller enbart långsamt varierande spänningar. I detta avsnitt ska vi betrakta diodens omslagsegenskaper, dvs hur snabbt vi kan få en diod att slå om från framspänt till backspänt läge, kort sagt betrakta diodens egenskaper som switch. Vi har tidigare använt dioden som likriktare för 50 Hz växelspänning, men hur höga frekvenser kan en diod egentligen likrikta? Hur högt upp i frekvens hänger dioden med?

I de här avsnittet ska vi visa att en diod inte kan slå om från ledande till spärrande förrän den har tömts på sitt laddningsinnehåll. Vi ska visa att den fortsätter att leda så länge minoritetsbäraröverskottet finns kvar i dioden. Först när minoritetsbäraröverskottet är borta börjar dioden att backspännas och då tar det ändå ytterligare tid innan spärrskiktskapacitansen laddats upp så att dioden tar hela backspänningsfallet.



Fig. 4.31. Kretskoppling för studier av en switchdiods omslagstid.

Låt oss nu studera diodens omslagstider med hjälp av den enkla kretsen i fig. 4.31. Genom att mäta spänningen över motståndet kan vi få en uppfattning om strömmen genom dioden. Tidigare har vi studerat RC-kretsen, en liknade krets men med en kondensator istället för dioden. I RCkretsen ges strömmen genom kondensatorn av ekvationen

$$I = \frac{dQ}{dt} = C\frac{dV}{dt},\tag{4.55}$$

där Q är laddningen i kondensatorn och C är kapacitansen. Utan att vi kanske tänkt på det så är detta kontinuitetsekvationen i integrerad form, en ekvation som beskriver laddningens oförstörbarhet.

Motsvarande ekvation för dioden lyder

$$\frac{dQ}{dt} = I - \frac{Q}{\tau_F} \implies I = \frac{Q}{\tau_F} + \frac{dQ}{dt}, \qquad (4.56)$$

där  $\tau_F$ är en för dioden karakteristisk tidskonstant. I en p<sup>+</sup>ndiod är  $\tau_F = \tau_p$  (hålens livslängd) och i en n<sup>+</sup>p-diod är  $\tau_F = \tau_n$ (elektronens livslängd). Den extra termen  $Q/\tau_F$  i (4.56) motsvarar rekombinationstermen *r* som vi la till när vi beskrev kontinuitetsekvationen i differentiell form för en pn-övergång i (4.xx). Termen beskriver att likströmmen  $I=Q/\tau_F$  genom en diod är precis så stor att den vidmakthåller ett visst minoritetsbäraröverskott Q, ett överskott som annars skulle försvinna med rekombinationstakten  $Q/\tau_F$ . Kontinuitetsekvationen i (4.56) kan illustreras med det ekvivalenta kretsschemat för dioden i fig. 4.32 där vi dock även har lagt till spärrskiktskapacitansen.

Låt oss nu anta att dioden i kretskopplingen i fig. 4.31 är framspänd och att strömmen  $I_F$  flyter genom dioden och att minoritetsbäraröverskottet därmed ges av $Q = \tau_F I_F$ . Låt sedan den pålagda spänningen slå om från  $+V_F$  till  $-V_R$  för att backspänna dioden. Diodens omslagsförlopp kommer nu att beskrivas av (4.56) med en viss ström  $I=-I_R$  i backriktningen,



Fig. 4.32. Ekvivalent kretsschema för diod med hänsyn tagen även till kapacitiva egenskaper.

Dioden backspänns således inte omedelbart utan den kommer fortsatt att vara framspänd ( $V_{diod} >0$ ) så länge minoritetsbäraröverskottet finns kvar, dvs så länge Q>0. Backströmmen genom kretsen ges därför av

$$I_R = \frac{V_R + V_{diod}}{R_L},\tag{4.58}$$

där  $V_{diod}$  är diodspänningsfallet. Det intressanta är nu att om vi antar att  $I_R$  är konstant under denna efterledningstid så kan vi faktiskt lösa differentialekvationen i (4.57) och därmed också beräkna den efterledningstid som krävs för att minoritetsbärarladdningen ska fösvinna. Vi får

$$\frac{Q}{\tau_F} = -I_R + (I_R + I_F)e^{-t/\tau_F}, \qquad (4.59)$$

vilket ger *Q*=0 vid tiden

$$t_s = \tau_F \ln\left(1 + \frac{I_F}{I_R}\right). \tag{4.60}$$

Det är först efter denna efterledningstid som dioden börjar backspännas. Det fortsatta omladdningsförloppet medan dioden backspänns bestäms nu av spärrskiktskapacitansen. Kretskopplingen i fig. 4.31 reduceras då till en vanlig RC-krets. Den enda komplikationen är att spärrskiktskapacitansen är komplicerat spänningsberoende. Men om vi ersätter kapacitansen med en konstant kapacitans lika med det effektiva medelvärdet under backspänningsförloppet, till exempel med värdet vid halva backspänningen, dvs  $C_j = C_j(-V_R/2)$ , så får vi en approximativ beskrivning av backspänningsförloppet som inte är så tokig.

Diodens återhämtningstid (*eng.* reverse recovery rime,  $t_{rr}$ ) från omslaget till dess att diodströmmen sjunkit till något väldefinierat värde  $I_{RR}$ , t ex 10% av toppvärdet  $I_R$ , ges då av

$$t_{rr} = t_s + R_L C_1 \ln 10.$$
 (4.61)

Hela omslagsförloppet illustreras i fig. 4.33 där det framgår att den teoretiska modellen ganska väl beskriver det verkliga förloppet.

Slutligen kan vi visa hur minoritetsbäraröverskottet klingar av under efterledningstiden, se fig. 4.34. Kurvorna

är plottade med hjälp av analytiska uttryck som efter en noggrann men inte helt okomplicerad analys utgör lösningen till den tidsberoende kontinuitetsekvationen.



Fig. 4.33. Diodens omslagsförlopp. (Mätdata= symbol, teori=heldragen linje).



Fig. 4.34. Minoritetsbäraröverskottets avklingning med tiden som parameter när dioden backspänns.

Slutligen kan vi visa hur minoritetsbäraröverskottet avklingar under efterledningstiden, se fig. 4.34. Kurvorna är plottade med hjälp av analytiska uttryck som erhålls som lösningen till den tidsberoende kontinuitetsekvationen efter en noggrann men inte helt okomplicerad analys. Speciellt intressant är att notera att derivatan i punkten x=0 är konstant under efterledningstiden, precis som vi antog då vi utnyttjade att  $I_R$  var konstant under efterledningstiden. Vi ser också att derivatan är negativ, vilket också är något vi förväntade oss med tanke på att backströmmen är negativ.

# 4.11 Sammanfattning

I detta kapitel började vi med att relatera ideala diodekvationen till en enkel styckevis linjär diodmodell lämplig för analys av enkla diodkretsar med papper och penna. Vi visade sedan hur man kan extrahera modellparametrar från mätdata.

Sedan fortsatte vi med att analysera potentialbarriären som finns mellan mellan diodens p- och n-sidor och hur balans mellan drift och diffusion kräver en viss inbyggd kontaktpotential. I samband med detta studerade vi laddningsfördelning, elektriskt fält och potentialbild i det utarmningsområde som bildas mellan p- och n-sidorna, och hur ett sådant område breder ut sig som funktion av pålagda spänningar. I samband med utarmningsområdena studerade vi även genombrott i backriktningen på grund av zener- och lavineffekter.

Nästa viktiga avsnitt behandlade bakgrunden till ideala diodekvationens exponentiella spänningsberoende och de fysikaliska mekanismer som ligger till grund för detta. Vi visade att diffusion utgör den dominerande strömbegränsningen i en ideal pn-övergång. Injektion och diffusion av minoritetsbärare behandlades kvantitativt med hjälp av diffusionsekvationen för minoritetsbärare och vi införde därvid en enkel rekombinationsmodell.

Därefter behandlade vi avvikelser från ideal diffusionsbegränsning i form av rekombination i utarmningsområdet, samt hur idealitetsfaktorn då ändras från n=1 vid diffusionsbegränsning till n=2 vid rekombinationsbegränsning. Avsnitten om ideala diodekvationen avslutades genom en studie av dess temperaturberoende.

Slutligen gjorde vi en kortfattad analys av diodens transienta switchegenskaper och med hjälp av laddningskontrollmodellen kunde vi ta fram ett uttryck för diodens återhämtningstid.

# 4.12 Övningsexempel

- Beräkna kontaktpotentialen i en pn-övergång där psidan har fermipotentialen 0,43 V och n-sidan har fermipotentialen -0,27 V! Denna uppgift löses enklast med ingenjörsmässiga approximationer och med Gauss lag.
- 2. Beräkna kontaktpotentialen i en pn-övergång där psidan har dopningen 10<sup>18</sup> cm<sup>-3</sup> och n-sidan har dopningen 10<sup>14</sup> cm<sup>-3</sup>!
- 3. Beräkna utarmningsområdets utbredning i ovanstående pn-övergång utan pålagd spänning! Hur mycket större är utarmningsområdet vid 10 V backspänning?
- Hur stor är potentialbarriären i en pn-övergång om utarmningsområdet är 5 μm och maxfältet är 10 V/μm?
- 5. Hur stor är en pn-övergångs kapacitans om utarmningsområdet är 2 μm?
- 6. I en backspänd pn-övergång är spänningsfallet i psidans 0,1 μm tjocka utarmningsområde 1V.
  - a. Hur stort är spänningsfallet i n-sidans 1,2 μm tjocka utarmningsområde?
  - b. Rita potentialdiagram för det aktuella fallet!
- Beräkna genombrottsspänningen i en pn-övergång med dopningen 10<sup>18</sup> cm<sup>-3</sup> och 10<sup>15</sup> cm<sup>-3</sup> på n-sidan!
- 8. Antag att vi har en pn-övergång med lika dopning på både p- och n-sidan. Antag vidare att  $D_n/D_p = L_n/L_p=2$ .
  - a. Skissa hålströmmen genom dioden om vi antar att  $L_n=0,2 \ \mu m!$
  - b. Hur ändras skissen om n-sidan istället är dubbelt så hårt dopad som p-sidan?
- 9. En pn-övergång har dopningarna  $10^{18}$  cm<sup>-3</sup> på psidan och  $10^{15}$  cm<sup>-3</sup> på n-sidan. Dioden är framspänd med 0,5 V? Antag att  $L_p=0,1$  µm och att  $L_n=0,2$  µm!
  - a. Hur många gånger ökar minoritetsbärarkoncentrationen vid kanten till utarmningsområdet som en följd av den pålagda spänningen?
  - b. Hur påverkas majoritetsbärarkoncentrationen?
  - c. Skissa minoritetsbärarkoncentrationerna på poch n-sidorna i pn-övergången!

- 10. Bredvidstående figur visar en p<sup>+</sup>nn<sup>+</sup> diod där n-skiktet är 10 μm tjockt. Dopningen i p<sup>+</sup>-skiktet är tusen gånger högre än i n-skiktet, och dopningen i n<sup>+</sup>-substratet är hundra gånger högre än i n-skiktet. Vid 10 V backspänning når utarmningsområdet precis fram till n<sup>+</sup>-skiktet varvid max-fältet vid p<sup>+</sup>n-övergången är 2 V/μm.
  - a. Hur stort är maxfältet vid 20 V backspänning?
  - b. Hur stort är fältet vid nn<sup>+</sup>-övergången?
  - c. Hur långt in i n<sup>+</sup>-skiktet tränger utarmningsområdet?
  - d. Hur långt in i p<sup>+</sup>-skiktet tränger utarmningsområdet?
  - e. Hur stor är dopningen i n-skiktet?

Denna uppgift löses enklast med ingenjörsmässiga approximationer och med Gauss lag! Det finns inga "färdiga" formler!



11. I mättade MOS-transistorer uppträder ett tunt utarmningsområde i slutet av kanalen vid substrat/ drainövergången. Man kan beräkna detta utarmningsområdes utbredning som funktion av den "överskjutande" spänningen  $V_j=V_{DS}-V_{DSAT}$ , där  $V_{DS}$ är drainspänningen och  $V_{DSAT}$  är MOS-transistorns mättnadsspänning. Man måste då anta att fältets randvärde vid utarmningsområdets kant mot kanalen är  $E_p$  (och inte noll som vi vanligen antar).

Ta fram detta uttryck för utarmningsområdets utbredning  $\Delta L$  och uttryck svaret mha parametrarna  $E_p$ definierad ovan och  $X_D$ , där  $X_D$  i sin vanliga betydelse innehåller substratdopningen  $N_A$ .



## 5. MOS-TRANSISTORN

Syftet med detta kapitel är att visa hur man med hjälp av en enkel strömmodell kan analysera enkla MOS-kretsar. Först ges en enkel bakgrund till Shockleys strömmodell och hur MOS-transistorn i sitt TILL-läge (stark inversion) kan betraktas som en icke-linjär resistans i resistansområdet och som en konstant-strömkälla i mättnadsområdet. Vi presenterar också en förenklad styckevis linjär modell som ofta är fullt tillräcklig för kretsanalys. I sitt FRÅN-läge antas MOS-transistorn i en första approximation inte leda ström, men vi ska i ett senare kapitel ta fram en strömmodell även för detta subtröskelområde. Dessa läckströmmar i subtröskelområdet spelar en allt större roll för effektförbrukningen i komplexa integrerade kretsar allt eftersom tekniken utvecklas. Moderna kretsar kan innehålla miljarder MOS-transistorer och summan av miljarder läckströmmar är knappast längre försumbar.

Kapitlet inleds med en kort sammanfattning av fälteffekten och visar hur strömmen genom en MOStransistor är proportionell mot såväl den styrspänning som alstrar rörliga laddningar som mot den spänning över komponenten som ger dessa laddningar en drifthastighet och således orsakar strömmen. MOS-transistorn har således i sitt ledande tillstånd ett kvadratiskt spänningsberoende.

Eftersom MOS-transistorn är en funktion av två variabler kan den lämpligen beskrivas som en så kallad tvåport. Strömmen som är en funktion av två variabler presenteras antingen med hjälp av överföringskarakteristiken där utspänningen betraktas som parameter, eller med hjälp av utgångskarakteristiken där styrspänningen får vara parameter.

Kapitlet avslutas med att en enkel MOS-krets analyseras ur såväl digitalkonstruktörens som ur analogkonstruktörens perspektiv. Såväl kretsens statiska egenskaper vid konstanta spänningar som dess dynamiska egenskaper vid tidsberoende insignaler analyseras. När det gäller kretsens egenskaper vid tidsberoende signaler försöker vi så långt möjligt återföra dessa till egenskaperna hos enkla RC-kretsar.

#### 5.1 Fälteffekten

Som namnet fälteffekttransistor, FET, antyder är MOStransistorn en komponent där man kan styra strömstyrkan genom komponenten med ett elektriskt fält. Komponenten är således kapacitiv och den för transistorn centrala MOSstrukturen (metal-oxide-semiconductor) utgör en kondensator med en viss kapacitans. Via denna plattkondensator med en "metallplatta" och en "halvledarplatta" bildas rörliga laddningar i en kanal som förbinder source och drain i halvledarmaterialet.

MOS-transistorn bygger alltså på fälteffekten som innebär att det elektriska fältet inverterar halvledarytan från p-typ till n-typ så att en ledande kanal bildas mellan source och drain. Mängden rörliga laddningar varierar med styrspänningen och kan beräknas med vanliga samband mellan laddning, kapacitans och spänning.

En spänning mellan source och drain skapar sedan ett lateralt elektriskt fält längs kanalen; ett fält som ger dessa laddningar en viss drifthastighet så att ström flyter genom MOS-transistorn. Det betyder att två mot varandra vinkelräta fält är inblandade på det sätt som antyds i fig. 5.1. Figuren visar ett fält  $E_y$  över MOS-kondensatorn och ett lateralt fält  $E_x$  mellan source och drain. Eftersom isolatorns tjocklek  $t_{ox}$  är mycket mindre än kanallängden L mellan source och drain,  $t_{ox} \ll L$ , är det vertikala fältet mycket större än det laterala fältet längs kanalen, dvs  $E_y \gg E_x$ . Strömmen genom MOS-transistorn kommer därigenom att bli proportionell mot produkten av dessa båda elektriska fält, dvs vi får ett kvadratiskt spänningsberoende.



Fig. 5.1. MOS-transistor i genomskärning

#### 5.2 Kapacitans och icke-linjär konduktans

Så länge drainspänningen är försumbart liten och det laterala elektriska fältet är konstant kan använda enkel endimensionel analys för att beräkna mängden rörlig laddning i kanalen. Ett flerdimensionellt problem formulerat enligt Gauss sats kan nämligen förenklas till ett endimensionellt problem om fältet är konstant i de andra riktningarna. Detta villkor är uppfyllt så länge  $E_x = -V_{DS}/L$  och därmed  $dE_x/dx=0$ .

Mängden rörlig laddning [per längdenhet] i kanalen vid source respektive drain ges då enligt kondensatorformeln av uttrycken

$$\begin{cases} Q'_{s} = -WC_{ox}V_{GST} \\ Q'_{D} = -WC_{ox}(V_{GST} - V_{DS}) \end{cases}$$
(5.1)

där  $V_{GST}=V_{GS}-V_T$  är den effektiva styrspänningen vid source och där W är MOS-transistorns bredd. Här måste vi använda den effektiva styrspänningen eftersom all laddning i MOS-kondensatorn inte är rörlig. Medelvärdet av den rörliga laddningen vid source och vid drain ges av

$$Q'_{inv} = -WC_{ox}\left(V_{GST} - \frac{V_{DS}}{2}\right),$$
(5.2)

och multiplicerar vi denna laddning med drifthastighetens medelvärde  $v=\mu V_{DS}/L$ , får vi följande uttryck för strömmen genom MOS-transistorn:

$$I_{DS} = \frac{W}{L} \mu C_{ox} \left( V_{GST} - \frac{V_{DS}}{2} \right) V_{DS},$$
 (5.3)

där  $\mu$  är laddningsbärarnas rörlighet i kanalen. Detta är Shockleys modell för strömmen genom en MOS-transistor. Eftersom strömmen är proportionell mot  $V_{DS}$  kan MOS-transistorn enligt denna strömmodell betraktas som en konduktans  $G_{DS}$  eftersom

$$I_{DS} = G_{DS} V_{DS}, \tag{5.4}$$

där då  $G_{DS}$  är ickelinjär eftersom den beror på  $V_{DS}$ 

$$G_{DS} = G_{ON} \left( 1 - \frac{V_{DS}}{2V_{GST}} \right), \tag{5.5}$$

där den så kallade TILL-konduktansen ges av

$$G_{ON} = k V_{GST}.$$
 (5.6)

Transistorkonstanten k ges av definitionen

$$k = \frac{W}{L} \mu C_{ox}.$$
 (5.7)

Ström/spänningskarakteristiken i detta MOS-transistorns resistansområde illustreras i fig. 5.2. De streckade linjerna illustrerar den förenklade linjariseringen med  $G_{DS}=G_{ON}$ .



Fig. 5.2. Strömkarakteristiken i resistansområdet.

Om vi nu återvänder till uttrycket för MOS-kondensatorns spänningsberoende laddning i (5.2) och multiplicerar med kanallängden så får vi den totala mängden rörlig laddning i kanalen

$$Q_{inv} = -C_G \left( V_{GST} - \frac{V_{DS}}{2} \right), \tag{5.8}$$

där styrekapacitansen  $C_G = WLC_{ox}$ . Denna styrekapacitans kan delas lika mellan source och drain eftersom

$$C_{GS} = \frac{\partial Q_G}{\partial V_S} = -\frac{\partial Q_{inv}}{\partial V_S} = C_G / 2, \qquad (5.9)$$

och

$$C_{GD} = \frac{\partial Q_G}{\partial V_D} = -\frac{\partial Q_{inv}}{\partial V_D} = C_G / 2.$$
 (5.10)

Denna modell av MOS-transistorn kan representeras av den ekvivalenta krets som visas i fig. 5.3. Modellen innehåller såväl en styrekapacitans som delats lika mellan source och drain som en styrbar icke-linjär konduktans.



Fig. 5.3. Modell för MOS-transistorn som styrbar resistans.

Kapacitansen  $C_{GD}$  mellan in- och utgång utgör här en komplikation vid enkla handberäkningar, varför man oftast använder den förenklade modellen i fig. 5.4 vid analys av enkla kretsar.



Fig. 5.4. Modell för MOS-transistorn som styrbar resistans.

#### 5.3 Mättnadsområdet

Från Shockleys modell för MOS-transistorns strömspänningskarakteristik som den illustreras i fig. 5.2 framgår att strömmen genom MOS-transistorn ökar allt långsammare när  $V_{DS}$  ökar. Slutligen når strömmen sitt mättnadsvärde  $I_{DSAT}$ . Enligt modellen sker detta vid drainspänningen  $V_{DS}=V_{GST}$ , den så kallade mättnadsspänningen  $V_{DSAT}$ , varvid mättnadsströmmen ges av uttrycket

$$I_{DSAT} = \frac{k}{2} V_{GST}^{2}.$$
 (5.11)

MOS-transistorn lämnar då resistansområdet och övergår till mättnadsområdet. Se fig. 5.5.



Fig. 5.5. Strömkarakteristiken i mättnadsområdet.

Anledningen till att strömmen mättas är att kanalladdningen vid drain går mot noll, dvs blir försumbar (s.k. pinch-off), när drainspänningen närmar sig  $V_{DSAT}$ . Den effektiva styrspänningen vid drain blir noll när  $V_{DS}=V_{DSAT}$ . Det betyder att större drainspänningar inte kan påverka laddningsförhållandena i kanalen och därmed inte heller påverka det elektriska fältet i kanalen som bestäms av laddningsfördelningen. Därför mättas strömmen och blir oberoende av drainspänningen.

Laddningen i kanalen är således ganska ojämnt fördelad när MOS-transistorn är i mättnadsområdet. Laddningen vid source ges fortfarande av

$$Q'_{S} = -WC_{ox}V_{GST}, \qquad (5.12)$$

medan den vid drain ges av

$$Q'_{D} = -WC_{ox} \left( V_{GST} - V_{DS} \right).$$
 (5.13)

Vi ser att  $Q'_D \rightarrow 0$  när  $V_{DS} \rightarrow V_{DSAT}$ . I motsats till förhållandena i resistansområdet, där vi hade nästan lika mycket laddning vid source och drain, är laddningen i mättnadsområdet förskjuten mot source och försumbar vid drain. I resistansområdet delade vi styrekapacitansen lika mellan source och drain, men i mättnadsområdet lägger vi hela kapacitansen till source eftersom kapacitansen  $C_{GD}$ vid drain är försumbar, se fig. 5.5.



Fig. 5.6. Kretsmodell av MOS-transistorn som styrbar strömkälla i mättnadsområdet.

Om man känner laddningsfördelningen i kanalen så känner man också det elektriska fältet eftersom produkten av laddning och fält ska ge en konstant ström. I nästa avsnitt ska vi använda den s.k. gradvisa kanalapproximationen för att beräkna såväl denna laddningsfördelning som det elektriska fältet i kanalen.

Från vad vi hittills vet om strömmodellen i (5.11) och om laddningen vid source och drain enligt (5.12) och (5.13) kan vi enbart dra slutsatsen att fältet vid source ges av  $V_{GST}/2L$  medan det vid drain rent teoretiskt går mot "oändligheten". Någonstans i kanalen borde det dessutom finnas en punkt där laddningen är lika med medelvärdet av laddningarna vid source och drain och där fältet är lika med medelvärdet  $V_{GST}/L$  av fältet. I denna punkt är laddningen hälften av sourceladdningen medan fältet och därmed drifthastigheten är den dubbla mot vid source.

# 5.4 Sammanfattning av MOS-transistorns strömmodell

För att sammanfatta strömmodellen för MOS-transistorn kan vi konstatera att den är en funktion av två spänningar, styrspänningen  $V_{GS}$  och utgångsspänningen  $V_{DS}$ . Det betyder att MOS-transistorn levererar en unik ström  $I_{DS}$  för varje punkt ( $V_{GS}$ ,  $V_{DS}$ ) i  $V_{GS}$ - $V_{DS}$ -planet, se fig. 5.7.

Eftersom detta kan vara svårt att illustrera i två dimensioner betraktar man denna funktion av två variabler som en funktion av en variabel i taget och där den andra variablen betraktas som parameter. Man får då en överföringskarakteristik och en utgångskarakteristik. MOS-transistorns överföringskarakteristik visas i Fig. 5.8 och dess utgångskarakteristik visas i Fig. 5.9.

En komponent som är en funktion av två variabler kan lämpligen representeras av en tvåport, se fig. 5.10. Detta i motsats till resistorer, kondensatorer och induktorer som är att betrakta som enportar.

$$I_{DS} = \begin{cases} \ddot{o}verf\ddot{o}rings-\\ karakteristik: \\ utgångs-\\ karakteristik: \end{cases} f(V_{GS}) \quad parameter V_{DS} \quad . (5.14)$$



Fig. 5.7.  $I_{DS}$  som funktion av  $V_{DS}$  och  $V_{GS}$ .



Fig. 5.8. Utgångskarakteristik med V<sub>GS</sub> som parameter.



Fig. 5.9. Överföringskarakteristik med V<sub>DS</sub> som parameter.



Fig. 5.10. MOS-transistorn som tvåport och dess symbol.

I en tvåport kan strömmarna genom de två portarna uttryckas som funktion av spänningarna över de båda portarna, styrporten och utgångsporten. Första ordningens förenklade tvåportar som representerar MOS-transistorn har redan illustrerats för resistansområdet i fig. 5.3 och för mättnadsområdet i fig. 5.6.

Om vi nu tar på oss tvåportsglasögonen och kikar in i tvåporten så ser vi en kapacitans på ingångssidan, se fig. 5.11.



Fig. 5.11. Tvåportens inimpedans är en kapacitans.

Om vi däremot kikar in i utgångsporten ser vi en styrbar resistans i resistansområdet och en styrbar strömkälla i mättnadsområdet, se fig. 5.12 respektive 5.13.



Fig. 5.12. MOS-transistorn i resistansområdet.



Fig. 5.13. MOS-transistorn i mättnadsområdet.

Dessa modeller förutsätter förstås att MOS-transistorn är TILL-slagen, dvs att  $V_{GS}$ > $V_T$  Det finns som vi ska diskutera närmare sedan även ett subtröskelområde då  $V_{GS}$ < $V_T$  då MOS-transistorn är frånslagen och att betrakta som en oändlig resistans  $R_{OFF}$ . MOS-transistorns arbetsområden i  $V_{GS}$ - $V_{DS}$ -planet illustreras i fig. 5.14.



Fig. 5.14.. MOS-transistorns arbetsområden.

# 5.5 Styckevis linjär MOS-modell

Vid förenklad analys av MOS-kretsar med papper och penna kan man ofta förenkla MOS-modellen och ersätta den med en styckevis linjär modell där  $G_{DS}$  i resistansområdet ersätts med  $G_{ON}$ . En sådan modell illustreras i fig 5.15. I detta perspektivet kan man se Shockleys modell som en "smoothing" kurva som ger en mjuk övergång mellan de båda linjariserade asymptoterna  $I_{DS}=G_{ON}V_{DS}$  och  $I_{DS}=I_{DSAT}$ .



Fig. 5.15. En styckevis linjär MOS-modell.

# 5.6 Enkla MOS-kretsar

Med den MOS-modell som vi nu utvecklat har vi ett verktyg som är tillräckligt bra för att kunna analysera enkla MOS-kretsar som digitala inverterare och analoga gemensam-source-förstärkare (GS) och få den grundläggande förståelse som krävs för analys av mer komplicerade kretsar.

Den enklaste krets man kan tänka sig är en MOStransistor i serie med ett lastmotstånd  $R_L$ . Vi ska nu kortfattat analysera dess digitala och analoga egenskaper.

#### 5.6.1 Inverteraren

MOS-kretsen som digital inverterare visas i fig. 5.16. En hög insignal ger en låg utsignal,  $V_{UT,låg}$  och en låg insignal ger en hög utsignal  $V_{DD}$ . MOS-transistorn uppträder som en styrbar resistans,  $R_{OFF}$  i FRÅN-läget och  $R_{ON}$  i tilläget. Digitalkonstruktörens bild av kretsen visas i fig. 5.17.

Utspänningen i de två arbetspunkterna bestäms genom spänningsdelning mellan de två resistanserna, se fig. 5.18. Endast under själva omslaget när utsignalen är nånstans mittemellan hög och låg passerar MOS-transistorn mättnadsområdet, se fig. 5.19.



Fig. 5.16. MOS-förstärkarsteg med resistiv last.



Fig. 5.17. Digitalkonstruktörens syn på MOS-kretsen.



Fig. 5.18 MOS-transistorn som styrbar resistans definierar inverterarens arbetspunkter



Fig. 5.19. MOS-transistorn som strömkälla under omslaget.

Inverterarens överföringskarakteristik,  $V_{UT}$  som funktion av  $V_{IN}$  kan nu enkelt beräknas genom att vi sätter strömmen genom MOS-transistorn enligt vår strömmodell lika med strömmen genom lastresistorn. Här har vi stor nytta av arbetsområdesdiagrammet i fig. 5.14 för att hålla ordning på vilken modell vi ska välja för MOS-transistorn.

Överföringskarakteristiken för tre olika värden på  $R_L/R_{ON}$  visas i fig. 5.20. Vi ser i figuren att det krävs minst ett resistansförhållande  $R_L/R_{ON}$ =10 för att inverteraren ska få tillräckligt lågt värde på  $V_{UT,låg}$  och ett tillräckligt distinkt omslag. Denna typ av logisk MOS-krets kallas på grund av den resistiva spänningsdelningen för *ratioed logic* eftersom den kräver ett visst minsta förhållande mellan två resistanser för korrekt funktion. Nackdelen med denna kretstyp är att den i sitt låga utläge drar ström, vilket t ex CMOS-logik inte gör.

#### 5.7 Analogförstärkaren

För analogkonstruktören är kretsen i fig. 5.21 ingen inverterare utan en förstärkare. Små variationer i inspänningen blir förstärkta och resulterar i stora spänningsvariationer på utgången. Analogkonstruktörens bild av MOS-kretsen med en styrbar strömkälla visas i fig. 5.22.



Fig. 5.20. MOS-kretsens överföringskarakteristik för tre olika värden på  $R_L/R_{ON}$ .



Fig. 5.21. MOS-förstärkarsteg med resistiv last.



Fig. 5.22. Analogkonstruktörens syn på MOS-kretsen.



Fig. 5.23. Illustration av småsignallinjarisering av överföringskarakteristiken.

Analogkonstruktören börjar alltid med att biasera sin krets, dvs med att välja den inspänning som i förhållande till den valda transistorstorleken ger den önskade förstärkningen, se fig. 5.23. Det gäller då att biasera kretsen i överföringsfunktionens mest linjära del där MOS-transistorn är mättad. I denna biaspunkt har kretsen spänningsförstärkningen

$$A_{v} = \frac{dV_{UT}}{dV_{in}} = \frac{\partial V_{UT}}{\partial I_{DS}} \frac{\partial I_{DS}}{\partial V_{IN}} = -g_{m}R_{L}, \qquad (5.15)$$

där

$$g_m = \frac{\partial I_{DS}}{\partial V_{IN}},\tag{5.16}$$

är MOS-transistorns transkonduktans, dvs överföringskarakteristikens branthet. Att en spänningsvariation  $v_{gs}$  på ingången ger upphov till en strömvariation  $i_{ds}=g_m v_{gs}$  och en spänningsförstärkning  $A_v=-g_m R_L$  illustreras i fig. 5.24.



Fig. 5.24. MOS-transistorn som styrbar strömkälla.

När analogkonstruktören valt sin biaspunkt fortsätter analysen med en småsignalanalys runt denna arbetspunkt. MOS-transistorns storsignal strömmodell linjariseras därvid runt biaspunkten och alla spänningsberoende parametrar ersätts med motsvarande konstanta värde i arbetspunkten, se fig. 5.25. Med vår enkla modell där  $C_G$  är konstant är ju  $C_g=C_G$ , men i en mer detaljerad modell där  $C_G$  är spänningsberoende ersätts  $C_G$  i småsignalmodellen med sitt värde  $C_g$  i biaspunkten. Analogkonstruktörens linjariserade bild av MOS-kretsen visas i fig. 5.26.



Fig. 5.25. MOS-transistorns ekvivalenta kretsscheman.



Fig. 5.26. Analogkonstruktörens småsignalmodell.

Småsignalanalys gäller förstås enbart för små insignaler eftersom överföringskarakteristikens ickelinjaritet ger upphov till distorsion vid för stora insignaler, se fig. 5.27. Här ser vi hur den verkliga utsignalen (streckad) avviker från den linjariserade sinustonen (blå heldragen linje). Denna distorsion av signalen motsvarar att utsignalen nu inte längre är en ren sinussignal utan att den innehåller övertoner.



Fig. 5.27. Distorsion av utsignalen vid stora insignaler.

# 5.8 Transientförlopp och frekvensgång

Minst lika viktig som analysen av MOS-kretsens statiska egenskaper i föregående avsnitt är analysen av de dynamiska egenskaperna, dvs hur kretsen reagerar på insignalen vid höga frekvenser.

En digitalkonstruktör är förstås intresserad av att kunna räkna ut hur snabbt hans logikkretsar slår om för att veta hur många logisk operationer som kretsen kan utföra per tidsenhet. Analogkonstruktören vill på samma sätt veta vilken gränsfrekvens förstärkaren har, dvs hur högt i frekvens kretsen fungerar med oförminskad förstärkning.

MOS-kretsens snabbhet begränsas av den RC-tidskonstant som uppstår på grund av den kondensator som belastar utgången. Denna kapacitans kan vara en kombination av ledningskapacitans och kapacitans från ingången av en MOS-transistor i nästa steg.

# 5.8.1 Inverterarsnabbhet

Låt oss börja med digitalkonstruktörens syn på MOSkretsen och analysera den tidsfördröjning som RCtidskonstanten ger upphov till från det insignalen slår om tll dess att utsignalen också slår om.



Fig. 5.26. Modell för fördröjningsberäkningar för MOSinverterare belastad av annan inverterare.

Kretsmodellen för fördröjningsberäkningarna visas i fig. 5.28. Tidsfördröjningen från det insignalen går låg till dess att utsignalen går hög kan beräknas ur den ekvivalenta RC-krets vi får genom att sätta  $R_{DS}=R_{OFF}=\infty$ , se fig. 5.27. Vanligen mäts fördröjningen vid halva matningsspänningen, dvs vid  $v_{UT}=V_{DD}/2$ . Det ger stigfördröjningen

$$t_{dr} = R_L C_G \ln 2 \approx 0,7 R_L C_G.$$
(5.17)



Fig. 5.27. Ekvivalent RC-krets för beräkning av stigfördröjning.

Fallfördröjningen när utsignalen går låg som svar på att insignalen gått hög är lite mer komplicerad att beräkna på grund av att MOS-transistorns resistans  $R_{DS}$  inte är konstant under urladdningsförloppet. En exakt analys av stegsvaret är möjlig men något tidskrävande. För en approximativ modell är det tillräckligt noggrant att ersätta  $R_{DS}$  med dess effektiva värde under urladdningsförloppet,  $R_{DS}=V_{DD}/I_{DSAT}$ , se fig. 5.28. Utsignalen ges då av uttrycket

$$v_{ut}(t) = V_{DD}\left(\frac{R_{ON}}{R_L + R_{ON}} + \frac{R_L}{R_L + R_{ON}}e^{-t/R_{eff}C_G}\right), (5.18)$$

där vi ser att exponentialförloppets startvärde är  $V_{DD}$  och att dess slutvärde ges av spänningsdelningen mellan  $R_{ON}$  och  $R_L$ . Fallfördröjningen blir då

$$t_{df} = R_{eff} C_G \ln \frac{2}{1 - R_{ON} / R_L} 2 \approx 0,7 R_{eff} C_G.$$
(5.19)



Fig. 5.28. Ekvivalent RC-krets för beräkning av fallfördröjning.



Fig. 5.29. MOS-kretsens svar på en pulsad insignal.

MOS-kretsens utsignal som svar på en pulsad insignal visas i fig. 5.29. Vi ser att stigfördröjningen är mycket längre än fallfördröjningen eftersom  $R_L \gg R_{eff}$ . Vi ser också att utsignalen inte har fullt spänningssving *rail-to-rail*. Detta är de två nackdelarna med ratioed logic förutom det faktum att inverteraren drar ström inte bara i omslagen utan även i sitt statiska läge när utspänningen är låg.

Av speciellt intresse är inverterarfördröjningen när inverteraren är belastad av fyra identiska inverterare, den så kallade *fanout-of-four* (FO4) fördröjningen. Ofta anges fördröjningen hos logikkretsar och processorkärnor i form av antalet FO4-fördröjningar för den aktuella kretsen i den aktuella MOS-teknologien.

# 5.8.2 Förstärkarens frekvensgång

Analogkonstruktören är intresserad av få veta vid hur höga frekvenser som förstärkaren klarar av att leverera en förstärkt utsignal. Det finns en gränsfrekvens som definieras som den frekvens vid vilken spänningsförstärkningen sjunkit till ett. Låt oss studera förstärkarens utspänning vid olika frekvenser. Förstärkaren, kapacitivt belastad av en så kallad sourceföljare, visas i fig. 5.30. Analogkonstruktörens ekvivalenta småsignal RC-krets visas i fig. 5.31. Nortonkretsen med en strömkälla kan ritas om till en Theveninekvivalent med en spänningskälla och ett inre motstånd. En helt vanlig RC-krets med andra ord.



Fig. 5.30. MOS-förstärkare belastad av sourceföljarsteg.



Fig. 5.31. Småsignalekvivalent för MOS-förstärkaren.

Om vi låter insignalen vara en sinussignal med vinkelfrekvensen  $\omega=2\pi f$ ,

$$v_{in} = V_1 \sin \omega t, \tag{5.20}$$

så blir utsignalen en dämpad och fördröjd sinussignal som ges av

$$v_{ut}(t) = \frac{A_v}{\sqrt{\left(1 + \omega^2 R^2 C_G^2\right)}} v_{\text{in}} \sin\left(\omega t + \phi\right), \qquad (5.21)$$

där fasvinkeln  $\phi$  ges av  $\phi$ =-arctan( $\omega R_L C_G$ ).

Utsignalen som funktion av tiden vid tre olika frekvenser visas i Fig. 5.32.



Fig. 5.32. MOS-kretsens utsignal vid tre olika frekvenser.

Vi ser tydligt att amplituden minskar med ökande frekvens och att fasvinkeln samtidigt ökar precis som den gör i alla RC-kretsar av denna typ (med en pol). Av uttrycket i (5.21) kan vi dra slutsatsen att det finns en vinkelfrekvens  $\omega$ =1/RC då förstärkningen sjunkit till 1/ $\sqrt{2}$ , och att det finns en högre gränsfrekvens

$$\omega_T = \frac{g_m}{C_G},\tag{5.22}$$

där förstärkningen sjunkit till ett. Vid denna frekvens orkar MOS-transistorn endast leverera samma ström som krävs för att driva den, dvs  $i_{in}=i_{ut}$ . Se Fig. 5.33 där förstärkningens amplitud och fasvinkel visas som funktion av frekvensen.



Fig. 5.33. Bode-diagram för RC-kretsen

## 5.8.3 CMOS-kretsar

Med samma metoder som vi illustrerat i föregående två avsnitt kan man analysera CMOS-kretsar. CMOS-inverteraren, dess överföringskarakteristik och dess stegsvar visas i fig. 5.34, 5.35 och 5.36. Vi ser att CMOSinverteraren har fullt spänningssving *rail-to-rail*, att den har en tydligt definierad omslagsspänning (med hög förstärkning) och att stig- och fallfördröjningarna kan göras lika stora utan att det nämnvärt påverkar överföringskarakteristiken. Vi ser att vi återigen har stor nytta av arbetsområdesdiagrammet från fig. 5.14 för att beräkna överföringskarakteristiken. I fig. 5.39 har detta diagram kompletterats med en markering av p-kanaltransistorns subtröskelområde och med dess begränsningslinje mellan resistans- och mättnadsområdena

$$v_{UT} = v_{IN} - V_{TP},$$
 (5.22)

där V<sub>TP</sub> är p-kanaltransistorns (negativa) tröskelspänning.



Fig. 5.34. CMOS-inverterare med överföringskarakteristik.



Fig. 5.35. CMOS-inverterarens överföringskarakteristik och strömförbrukning.



Fig. 5.36. CMOS-inverterarens stegsvar.

En CMOS-förstärkarkrets biaserad med hjälp av en referensströmkälla och en strömspegel M2/M3 visas i fig. 5.37. Här utgörs förstärkarens lastresistor av en p-kanal MOS-transistor M2 eftersom resistorer inte finns att tillgå i integrerade CMOS-kretsar. Belastningslinjen för M2 visas i fig. 5.38. Som jämförelse visas belastningslinjen för en resistor med resistansen  $R_L = V_{DD}/I_{SS}$  (streckad linje).

Med hjälp av storsignalmodellens strömekvationer kan man sedan beräkna biaseringspunkten, och därefter i denna beräkna spänningsförstärkningen med hjälp av småsignalmodellen. Detta illustreras i fig. 5.39. Skillnaden mellan de båda modellerna är att man i småsignalmodellen enkelt kan ta hänsyn till MOS-transistorns utgångskonduktans

$$g_d = \frac{\partial I_{DS}}{\partial V_{DS}},\tag{5.16}$$

något som man för enkelhets skull försummar vid storsignalanalysen.



Fig. 5.37. Strömmatad MOS-förstärkarkrets med kapacitiv last på utgången.



*Fig. 5.38. Belastningslinjen för en passiv pMOS-last jämförd med motsvarande belastningslinje för en resistor.* 



Fig. 5.39. Förstoring av överföringsfunktionen kring biaseringspunkten. Småsignalmodell (röd linje).

# 5.9 Sammanfattning

I detta kapitel har vi beskrivit fälteffekten som en kapacitiv effekt med vilken man kan styra strömmen genom en MOS-transistor. MOS-transistorn är som en styrbar switch, eller dimmer, den har ett frånläge då den inte leder ström (bortsett från en försumbar subtröskelström) och ett tilläge då den leder ström. MOS-transistorn kan beskrivas med en enkel tvåportsmodell med en kapacitans på ingången och en styrbar resistans eller strömkälla på utgången. Renodlar man denna modell får man en enkel styckevis linjär transistormodell. Shockleys kvadratiska MOS-modell kan om man så vill, ses som en smoothing-kurva som förenar de linjära approximationerna med en mjuk övergång.

Med dessa enkla transistormodeller analyserade vi sedan några enkla MOS-kretsar genom att se på dessa ur såväl ett digitalt perspektiv som ett analogt perspektiv. Överföringskurvor, belastningslinjer, småsignalmodeller och förstärkning diskuterades. Kapitlet avslutades med en analys av inverterarsnabbhet och frekvensgång genom att beskriva kretsarna med enkla RC-kretsar.

# 5.10 Övningsexempel

- 1. Rita en enkel storsignal tvåportsmodell för en MOStransistor i
  - a. resistansområdet
  - b. mättnadsområdet

Det är tillräckligt om modellen tar hänsyn till MOStransistorns ingångskapacitans och till dess transkonduktanskonduktans på utgången! Hur skulle modellen förändras om MOS-transistorn dessutom har en icke-försumbar utgångsresistans?



- 2. Hur stor är den effektiva styrspänningen vid
  - a. source
  - b. drain?
- 3. Hur stor är mängden rörlig kanalladdning per längdenhet vid source respektive drain?
- 4. Teckna ekvationerna i Shockleys strömmodell för MOS-transistorn i resistans- och mättnadsområdena genom att multiplicera ihop kanalens medelladdning med dessa laddningars medelhastighet!
- 5. Vid vilken drainspänning mättas strömmen genom en MOS-transistor ( $\partial I_{DS}/\partial V_{DS}=0$ ) och hur stor är laddningen vid drain vid denna drainspänning?
- 6. Hur stort är det elektriska fältet längs kanalen vid source i en mättad MOS-transistor?
- Skissa MOS-transistorns överförings- och utgångskarakteristikor enligt ovanstående strömmodeller. Rita även in den förenklade styckevis linjära modellen i utgångskarakteristiken.
- 8. Markera MOS-transistorns arbetsområden, dvs subtröskelområdet, resistansområdet och mättnadområdet, i bifogade V<sub>DS</sub>-V<sub>GS</sub>-diagram.
- 9. Hur beror MOS-transistorns transkonduktans i mättnadsområdet på den effektiva styrspänningen resp. strömnivån i arbetspunkten?
- 10. Bifogade figur visar en enkel MOS-krets med en aktiv MOS-transistor och en passiv resistiv last.
  - a. Skissa kretsens överföringsfunktion i bifogade  $V_{DS}$ - $V_{GS}$ -diagram.

- b. Ta fram den exakta ekvationen för överföringsfunktionen i det aktiva området där MOS-transistorn är mättad! Resultatet borde bli en del av en parabel, eller?
- c. Ta fram ekvationen för överföringsfunktionen i det område där MOS-transistorn är resistiv!
- d. För  $V_{IN}=V_{DD}$  borde  $V_{UT}$  gå mot  $V_{DD}R_{ON}/R_L$  om  $R_L >> R_{ON}$  enligt principen för vanlig resistiv spänningsdelning. Stämmer det med dina ekvationer?





STYRSPÄNNING

- 11. Småsignalegenskaper hos kretsen.
  - a. Hur stor är kretsens spänningsförstärkning i det aktiva området?
  - b. Uttryck spänningsförstärkningen med hjälp av transkonduktansen!
  - c. Vilken inspänning krävs för att biasera förstärkaren i en arbetspunkt  $V_{UT}=0,6V_{DD}$ ? Antag att  $V_T=0,2V_{DD}$  och att  $R_L/R_{ON}=64$ .
- 12. Bifogade figur visar en enkel digital CMOS-krets i form av en inverterare med två aktiva switch-transistorer; en nMOS-transistor till jord och en pMOS-transistor som last till matningsspänningen  $V_{DD}$ .
  - a. Identifiera pMOS-transistorns arbetsområden i samma typ av diagram som ovan, dvs vilka begränsningslinjer anger när pMOS-transistorn är avstängd respektive mättad?
  - b. Beräkna CMOS-inverterarens omslagsspänning genom att använda lämpligt "lika-ström" uttryck för de två transistorerna.



Inverteraren – en digital CMOS-krets.



STYRSPÄNNING

#### 6. MOS-KONDENSATORN

I det här kapitlet ska vi studera MOS-kondensatorn där förkortningen MOS, som vi redan nämnt, står för metaloxide-semiconductor. Denna förkortning beskriver MOSkondensatorns uppbyggnad med en oxid som isolator mellan kondensatorns båda elektroder. Elektroderna eller "kondensatorplattorna" utgörs av metallplattan (styret) och av halvledarsubstratet. Förståelsen av MOS-strukturen och den kapacitiva fälteffekten får anses fundamental eftersom den utgör grunden för MOS fälteffekttransistorns (MOSFET) funktion. En endimensionell modell av en tvådimensionell MOS-kondensator på p-typ substrat visas i fig. 6.1 (streckad linje).



Fig. 6.1. Intrinsisk MOS-kondensator på p-typ substrat.

MOS-kondensatorn kan ses som en speciell form av plattkondensator. När man lägger en spänning på *styret (eng.* gate), dvs på "metallplattan", händer mycket intressanta saker vid halvledarmaterialets gränsyta mot isolatorn, alltså på "halvledarplattan,. I detta kapitel ska vi närmare studera hur man med styrspänningen kan försätta en MOS-kondensator i tre olika "tillstånd", nämligen i *ackumulation, utarmning och inversion*.

## 6.1 Ackumulation, utarmning och inversion.

I ackumulation lägger man en sådan spänning på MOSkondensatorns metallplatta att majoritetsbärare ansamlas, ackumuleras, på halvledarplattan. I fallet med en MOSkapacitans på p-typ halvledarsubstrat betyder det att en negativ spänning på styret attraherar hål till halvledarens gränsyta mot isolatorn. Dessa majoritetsbärare befinner sig alla i ett ytterst tunt ytskikt utan egentlig tjocklek. Begreppet "halvledarplatta" är således högst relevant. MOS-kondensatorn kan därför betraktas som en plattkondensator med kapacitansen  $C_{ox}$ , där avståndet mellan "plattorna" bestäms av isolatorns tjocklek  $t_{ox}$ . Oxidkapacitansen  $C_{ox}$  [per ytenhet] ges enligt plattkondensatorformeln av uttrycket

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3,9\varepsilon_0}{t_{ox}},\tag{6.1}$$

där  $\varepsilon_{ox}$  är isolatorns dielektricitetskonstant, eller permittivitet  $\kappa$ . I fallet med kiseldioxid är kappa för isolatorn lika med 3,9, ett värde som kan jämföras med 12 för kisel. I de senaste MOS-teknologierna har kiseldioxid som isolator i stor utsträckning kommit att ersättas av andra så kallade high-kappa material, som typ hafniumoxid med  $\kappa$ =25. MOS-kondensatorn i ackumulation visas i fig 6.2a. Med positiv styrspänning får MOS-kondensatorn positiv laddning på styret och negativ laddning i halvledaren. Det som sker vid positiv styrspänning är att *majoritetsbärare repelleras* bort från halvledarytan och lämnar kvar okompenserade störatomer. I fallet med en MOS-kondensator på ett p-typ substrat innebär det en negativ acceptorladdning i halvledarsubstratet. Halvledarmaterialet *utarmas* således på majoritetsbärare och ett utarmningsområde bildas i halvledaren. MOS-kondensatorn sägs nu befinna sig i *utarmning*. Precis som i en pnövergång växer utarmningsområdets tjocklek med ökande styrspänning. MOS-kondensatorn i utarmning visas i fig. 6.2b.



*Fig. 6.2. MOS-kondensatorn i ackumulation, utarmning och inversion.* 

Eftersom kapacitansmätningar är ac-mätningar ligger också den laddning som svarar på mätsignalen vid utarmningsområdets kant. Kantens läge bestäms av styrspänningen vilket något förenklat betyder att "halvledarplattan" flyttas djupare och djupare in i halvledaren med ökande styrspänning. Det effektiva avståndet mellan kondensatorplattorna ökar således med ökande styrspänning varvid MOS-kapacitansen sjunker. Att kapacitansen sjunker innebär att vi i ett elektriskt ekvivalentschema har två kondensatorer i serie, se fig 6.3b, dvs en spärrskiktskapacitans  $C_{ax}$ .



Fig. 6.3. MOS-kondensatorn i ackumulation, utarmning och inversion.

Spärrskiktskapacitansen ges enligt plattkondensatorformeln av uttrycket

$$C_B = \frac{\varepsilon_{Si}}{W} = \frac{12\varepsilon_0}{W},\tag{6.2}$$

där W är utarmningsområdets spänningsberoende utbredning. Anledningen till att vi ser MOS-kondensatorn i utarmning som två kondensatorer i serie är ju huvudsakligen för att vi har två isolatorer med olika kappa som ligger i serie. Dessutom är  $C_B$  spänningsberoende medan  $C_{ox}$  är konstant. Den ekvivalenta oxidtjockleken EOT för dessa två kapacitanser i serie ges av

$$EOT = t_{ox} + \frac{\varepsilon_{ox}}{\varepsilon_{si}} W.$$
(6.3)

Nu är det dags att fundera på om det inte finns några elektroner vid halvledarytan? Dessa är ju också negativt laddade och borde attraheras till halvledarplattan av den positiva styrspänningen. Detta är en riktig anmärkning. Det finns en viss mängd elektroner vid halvledarytan, men så länge MOS-kondensatorn befinner sig i utarmning så är denna elektronladdning försumbar jämfört med utarmningsladdningen. Denna approximation är rimlig upp till en viss kritisk spänning över MOS-kondensatorn – en spänning som definierar MOS-kondensatorns tröskelspänning,  $V_T$ .

MOS-kondensatorn går in i *inversion* så fort spänningen över MOS-kondensatorn överstiger tröskelspänningen. I inversion anses elektronkoncentrationen vid halvledarytan inte längre vara försumbar. Man har definitionsmässigt ansett att övergången mellan försumbar och ickeförsumbar mängd elektroner ligger vid den styrspänning då elektronkoncentrationen vid halvledarytan når upp till dopningskoncentrationen i substratet. I det läget kan i alla fall inte elektronladdningen inte längre anses försumbar jämfört med utarmningsladdningen.

När styrspänningen överstiger tröskelspänningen växer elektronkoncentrationen snabbt med ökande styrspänning. Vi får då ett tunt skikt av negativt laddade elektroner vid ytan. Detta skikt kallas för inversionsskikt (*eng.* inversion layer) eftersom laddningarna är av motsatt typ mot majoritetsbärarna och eftersom ytan inverteras från p-typ till n-typ.

Inversionsskiktet kommer att lägga sig som ett lock över utarmningsområdet och avskärma detta från att ytterligare påverkas av styrspänningen. Vid styrspänningar större än tröskelspänningen är det nämligen enbart inversionsladdningen som ökar. Spärrskiktskapacitansen "kortsluts" på detta sätt och den mätbara MOSkapacitansen är återigen lika med oxidkapacitansen  $C_{ox}$ . Se fig. 6.2c och 6.3c.

Detta kvalitativa resonemang leder nu fram till en enkel modell för att beräkna inversionsladdningen i MOSkondensatorn. Denna modell innebär att inversionsladdningen för styrspänningar större än tröskelspänningen kan beräknas ur uttrycket

$$Q_{inv}' = -C_{ox}(V_{GB} - V_T).$$
(6.4)

För styrspänningar mindre än tröskelspänningen,  $V_G < V_T$ , anser vi att inversionsladdningen är försumbar,  $Q_{inv}'=0$ och för styrspänningar större än tröskelspänningen anser vi att inversionsladdningen växer linjärt med styrspänningen. Vi har således resonerat oss fram till en styckevis linjär laddningsmodell. Med denna modell blir också benämningen tröskelspänning fullt begriplig. Denna styckevis linjära laddningsmodell illustreras i fig. 6.4.



Fig. 6.4. Styckevis linjär laddningsmodell for inversionsladdningen.

**Exempel 6.1**: Beräkna  $C_{ox}$  för en MOS-struktur med kiseldioxid som isolator och med en oxidtjocklek på 10 nm.

**Lösning**: Isolatorn utgörs av kiseldioxid SiO<sub>2</sub> med  $\kappa$ =3,9. Kapacitansen blir då

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{3,9 \cdot 8,85 \cdot 10^{-12}}{10 \cdot 10^{-9}} = 3,45 \text{mF/m}^2$$

vilket motsvarar 3,45 fF/µm<sup>2</sup>. Jämför med kapacitansen hos ett spärrskikt på 1 µm som är av 0,1 fF/µm<sup>2</sup>. Dagens MOS-transistorer i digitala integrerade kretsar har en area långt mindre än en kvadratmikrometer så vi rör oss alltså med kapacitanser i storleksordningen från några attofarad till någon eller några femtofarad.

Laddningsfördelningen i en MOS-struktur i *ackumulation*, *utarmning* och *inversion* visas i fig. 6.5.



Fig. 6.5. Laddningsfördelningen i en MOS-kapacitans i ackumulation, utarmning, och inversion.

Vi ser att i *ackumulation* motsvaras den negativa laddningen på styret av en lika stor mängd positiv hålladdning vid halvledarytan. I *utarmning* motsvaras den positiva laddningen  $Q'_G$  på styret av en lika stor negativ utarmningsladdning  $Q'_B$  i halvledaren. I *inversion* ser vi att en negativ laddning bestående av rörliga elektroner har börjat växa fram vid halvledarytan. Varje ökning av  $Q'_G$ motsvaras av en lika stor ökning av inversionsladdningen  $Q'_{inv}$ , medan utarmningsladdningen  $Q'_B$  förblir konstant.

När MOS-strukturen befinner sig i utarmning kan vi använda vår utarmningsapproximation från pn-övergången för att beräkna sambandet mellan utarmningsladdningen  $Q'_B$  (*eng.* bulk charge) och den del av styrspänningen som faller över utarmningsområdet, den s.k. ytpotentialen  $\Psi_S$ . Utarmningsområdets djup ges därvid som tidigare för den enkelsidiga pn-övergången av uttrycket

$$W = X_D \sqrt{\Psi_S}, \tag{6.5}$$

där som förut

$$X_D = \sqrt{\frac{2\varepsilon_{Si}}{qN_A}}.$$
(6.6)

Utarmningsladdningen ges då av

$$Q'_{B} = -qN_{A}W = -\sqrt{2\varepsilon_{Si}qN_{A}\Psi_{S}}.$$
(6.7)

För MOS-kondensatorn gäller den vanliga kondensatorformeln för laddningen på metallplattan,

$$Q'_G = C_{ox} V_{ox}. \tag{6.8}$$

Eftersom inversionsladdningen är försumbar när MOSkondensatorn befiiner sig i utarmning och vi har lika stor positiv som negativ laddning så blir  $Q'_G = -Q'_B$ . Vi får då följande relation mellan styrspänning och ytpotential:

$$V_{GB} = V_{ox} + \Psi_S = -\frac{Q'_B}{C_{ox}} + \Psi_S, \qquad (6.9)$$

Sambandet mellan styrspänning och ytpotential ges då av det enkelt formulerade uttrycket

$$V_{GB} = \gamma \sqrt{\Psi_s} + \Psi_s, \qquad (6.10)$$

där

$$\gamma = \frac{\sqrt{2\varepsilon_{st}qN_A}}{C_{ex}} \,. \tag{6.11}$$

**Exempel 6.2.** Beräkna gamma för en MOS-struktur med dopningen  $N_A$ =1,25·10<sup>16</sup> cm<sup>-3</sup> och oxidtjockleken 10 nm!

**Lösning**. Ur (6.1) fås  $C_{ox}$ =3,45 mF/m<sup>2</sup>. Insättning i (6.11) ger  $\gamma$ =0,19 V<sup>1/2</sup>.

**Exempel 6.3.** Beräkna hur stor del av styrspänningen som hamnar över oxiden på gränsen till inversion om ytpotentialen då är 0,7 V!

**Lösning.** Med ovanstående värde på gamma får vi  $V_{ox} = \gamma \sqrt{\Psi_{\rm S}} = 0,16$  V. MOS-strukturens tröskelspänning är således 0,86 V och av denna spänning hamnar 0,7 V över utarmningsområdet och 0,16 V över oxiden.

**Exempel 6.4**. Beräkna utarmningsområdet djup med ovanstående värden på  $N_A$  och  $\Psi_S=0,7$  V.

**Lösning.** Utarmningsområdets djup ges av (6.5) till W= 270 nm där enligt (6.6)  $X_D=0,33 \ \mu \text{m/V}^{1/2}$ .

**Exempel 6.5**. Beräkna utarmningskapacitansen  $C_B$  under ovanstående förutsättningar!

Lösning. Utarmningskapacitansen ges av (6.2) till

$$C_B = \frac{\varepsilon_{Si}}{W} = \frac{12\varepsilon_0}{W} = 0,4 \text{ mF/m}^2.$$

Eftersom utarmningsområdet är mycket tjockare än vad oxiden i isolatorn är (W=270 nm jämfört med  $t_{ox}$ =10 nm i exempel 6.1) så är också utarmningskapacitansen endast en tiondel av oxidkapacitansen.

En styckevis ickelinjär modell för MOS-kapacitansen som funktion av styrspänningen visas i fig. 6.6.



Fig. 6.6. Normaliserad MOS-kapacitans  $C_G/C_{ox}$  per ytenhet som funktion av styrspänningen . Styckevis modell.

Vi ser att i ackumulation och i inversion så är styrekapacitansen lika med oxidkapacitansen, medan i utarmning så hamnar oxidkapacitansen i serie med utarmningskapacitansen  $C_B$ . Det betyder att den totala MOS-kapacitansen som ges av

$$C_G = \frac{C_B C_{ox}}{C_B + C_{ox}},\tag{6.12}$$

minskar vid ökande styrspänning tills inversion inträffar och kapacitansen återgår till att bestämmas av  $C_{ox}$  enbart. När man plottar denna kapacitansmodell har man inget direkt samband mellan styrspänningen  $V_G$  och kapacitansen  $C_G$ . Istället får man göra en parametrisk plot där såväl styrspänningen  $V_G$  som MOS-kapacitansen (genom uttrycket för  $C_B$ ) är funktioner av ytpotentialen.

Sambandet mellan  $V_G$  och ytpotentialen  $\Psi_S$  gavs tidigare i (6.10) som

$$V_{GB} = \gamma \sqrt{\Psi_s} + \Psi_s. \tag{6.13}$$

Eftersom  $\gamma$  oftast är ganska liten är denna relation mellan styrspänningen  $V_G$  och ytpotentialen  $\Psi_S$  i stort sett att betrakta som linjär. Det är därför lockande att linjarisera detta samband för att få enklare ekvationer i våra senare resonemang. Man brukar då välja att linjarisera detta uttryck kring  $\Psi_S=2\phi_B$ , dvs på gränsen till inversion där  $V_{GB}=V_T$ . Se fig. 6.7. Vi får genom Taylorutveckling

$$V_{GB} = V_T + n \left( \Psi_S - 2\phi_B \right), \tag{6.14}$$

där idealitetsfaktorn n ges av derivatan

$$n = \frac{dV_{GB}}{d\Psi_S}\Big|_{\Psi_S = 2\phi_B} = 1 + \frac{\gamma}{2\sqrt{2\phi_B}}.$$
(6.15)

**Exempel 6.6**: Beräkna idealitetsfaktorn för MOS-strukturen i de föregående exemplen!

Lösning: Genom insättning av siffervärden i (6.15) får vi

$$n = 1 + \frac{\gamma}{2\sqrt{\Psi_s}} \approx 1 + \frac{\gamma}{2\sqrt{2\phi_B}} = 1 + \frac{0.19}{2\sqrt{0.7}} = 1.11.$$



Fig. 6.7. Styckevis linjär modell för ytpotentialen.

Vi har därmed utvecklat en styckevis linjär modell för sambandet mellan styrspänningen och ytpotentialen. Denna modell kan sammanfattas på följande sätt:

$$\begin{cases} \Psi_{S} - 2\phi_{B} = \frac{V_{GB} - V_{T}}{n} & V_{GB} < V_{T} \\ \Psi_{S} = 2\phi_{B} & V_{GB} > V_{T} \end{cases}$$
(6.16)

MOS-strukturens idealitetsfaktor har ett värde som vanligtvis ligger mellan 1 och 1,5. Se Fig. 6.7. Vi ser att man i denna modell har valt att låsa ytpotentialen vid det värde  $2\phi_B$  då koncentrationen av elektroner vid ytan är lika med dopningskoncentrationen, dvs i brytpunkten mellan utarmning och inversion.

#### 6.2 Antalet elektroner vid halvledarytan

I detta avsnitt ska vi se närmare på hur många elektroner vi egentligen har vid halvledarytan och vid vilken ytpotential dessa elektroner inte längre kan försummas utan att vi får en inversionsladdning vid ytan.

Vi vet sedan tidigare att minoritetsbärarkoncentrationen i en halvledare i termisk jämvikt är försumbar jämfört med majoritetsbärarkoncentrationen som i sin tur bestäms av dopningen,  $p_{p0}=N_A$ . Minoritetsbärarkoncentrationen ges av massverkans lag,  $n_{p0}=n_i^2/N_A$ . När vi lägger en spänning över MOS-strukturen så kommer ytpotentialen  $\Psi_s$  att öka som vi nyss visat och därmed kommer också mängden elektroner vid ytan att öka. Enligt Boltzmanns fördelningsfunktion och enligt "the-law-ofthe-junction" kommer mängden elektroner att öka

$$n_{s}(\Psi_{s}) = n_{n0} e^{\Psi_{s}/V_{t}} .$$
(6.17)

Med hjälp av massverkans lag och med hjälp av definitionen på fermipotentialen  $\phi_B$  kan vi skriva om elektronkoncentartionen på formen

$$n_{s}(\Psi_{s}) = N_{A} e^{(\Psi_{s} - 2\phi_{B})/V_{t}}.$$
(6.18)

Ur dessa samband kan vi se att elektronkoncentrationen ökar exponentiellt med ytpotentialen  $\Psi_s$ . Eftersom den emellertid startar på en mycket låg nivå kan vi ändå betrakta den som försumbar så länge MOS-strukturen är i utarmning. Något mer korrekt uttryckt kanske man tvärtom ska säga att MOS-kapacitansen befinner sig i utarmning så länge den rörliga laddningen kan anses försumbar i förhållande till utarmningsladdningen.

Vi noterar i förbigående att hål- och elektronkoncentrationerna vid halvledarytan är lika stora,  $n_s(\phi_B)=p_s(\phi_B)=n_i$ när ytpotentialen är lika med substratets fermipotential,  $\Psi_s=\phi_B$ . Ytan är då intrinsisk. Vid ytpotentialer större än fermipotentialen,  $\Psi_s > \phi_B$ , växer sig elektronkoncentrationen vid ytan större än hålkoncentrationen. Vi går nu in i ett område då MOS-strukturen fortfarande befinner sig i utarmning, men ett område då vi samtidigt ser en begynnande *svag inversion*. Var finns då den brytpunkt då MOS-kapacitansen lämnar utarmning och svag inversion och går över i stark inversion? En sådan brytpunkt som visat sig vara både lättbegriplig och praktiskt lämplig är vid den ytpotential som är lika med dubbla fermipotentialen,  $\Psi_s=2\phi_B$ . Vid denna ytpotential är nämligen elektronkoncentrationen vid ytan lika stor som dopningen,

$$n_s(2\phi_B) = N_A. \tag{6.19}$$

Inversion anses således definitionsmässigt inträffa vid denna ytpotential,  $\Psi_S=2\phi_B$ . Ibland förtydligar man denna brytpunkt ytterligare genom att markera att man inte bara går från utarmning till inversion utan från svag inversion till stark inversion. Den styrspänning vid vilken inversionsskiktet börjar växa till sig ordentligt definieras följaktligen som MOS-strukturens tröskelspänning,

$$V_T = \gamma \sqrt{2\phi_B} + 2\phi_B. \tag{6.20}$$

**Exempel 6.7**: Beräkna tröskelspänningen  $V_T$  för vår MOSstruktur från de tidigare exemplen samt spänningsfördelningen mellan oxid och utarmningsområde vid styrspänningen 2 V!

**Lösning**: Med data för MOS-strukturen enligt tidigare exempel så låser ytpotentialen vid  $\Psi_S=2\phi_B=0,7$  V. Tröskelspänningen ges av  $V_T=0,7+0,16=0,86$  V, varav 0,7 V hamnar över utarmningsområdet och 0,16 V över oxiden. Vid  $V_G=2$  V hamnar fortfarande 0,7 V av styrspänningen över utarmningsområdet medan resten hamnar över oxiden, dvs 1,3 V. Den effektiva styrspänningen som är den del av styrspänningen som skapar inversionsladdning ges av styrspänningen minus tröskelspänningen, dvs i vårt fall ges den av  $V_{GST}=V_{GS}-V_T=2-0,86=1,14$ V.

# 6.3 High-kappa MOS-kondensatorer

Metallplattan i en MOS-struktur utgjordes ursprungligen av ett tunt aluminiumskikt, därav namnet metallstyre. Men från 70-talets början tills helt nyligen har styret utgjorts av ett hårddopat polykristallint kiselskikt, i dagligt tal polykisel, på grund av sina goda och stabila egenskaper. Isolatorn, utgörs oftast av kiseldioxid SiO<sub>2</sub> utom i de allra senaste generationernas MOS-transistorer; poly-Si och SiO<sub>2</sub> har varit ett väletablerat koncept med stabila egenskaper.

I de allra mest moderna MOS-teknologierna utgörs isolatorn av ett s.k. high-kappa material med hög dielektricitetskonstant, typ hafniumoxid. Med ett  $\kappa$ =24, d.v.s. med en permittivitet som är sex gånger större än kiseldioxidens, kan en isolator i HfO<sub>2</sub> kan göras sex gånger tjockare än motsvarande i skikt i kiseldioxid för samma kapacitans. Fördelen med en tjockare isolator är de lägre läckströmmarna och den lägre effektförbrukningen, en mycket viktig aspekt i dagens många batteridrivna elektroniska apparater som lap-tops och smarta mobiler. I samband med bytet av isolatormaterial runt 2008 har man också återinfört styren i metall för att skapa nya stabila metall/isolator-koncept.

## 6.4 MOS-kondensatorns banddiagram

Ytterligare hjälp med att förstå funktionen hos en MOSkapacitans kan man få genom att rita dess banddiagram. Banddiagrammet för en ideal MOS-struktur utan pålagda spänningar visas i Fig. 6.8. Oxiden är som synes en bra isolator med ett stort bandgap på 9 eV för SiO<sub>2</sub>, ett bandgap som är mycket större än kislets 1,1 eV. Ferminivån i substratet av p-typ är markerad  $E_{Fp}$  och i metallen  $E_{Fm}$ . Vi ser att ferminivån är konstant eftersom MOS-kapacitansen är i termisk jämvikt utan pålagda spänningar. I banddiagrammet har vi även markerat referensnivån för intrinsiskt kisel.



Fig. 6.8. Banddiagram för ideal MOS-kapacitans utan pålagda spänningar.

Eftersom en ideal MOS-kondensator i termisk jämvikt är både laddnings- och fältfri är såväl lednings- som valensbanden i halvledaren helt flata, det s.k. flatbandsvillkoret. (*eng.* flatband).

I nästa figur, fig. 6.9, ska vi nu visa banddiagrammet med en positiv spänning på styret. Styrspänningen antas vara mindre än tröskelspänningen, så MOS-kapacitansen befinner sig i utarmning.

Eftersom elektroner har negativ laddning så innebär som vi ser en positiv styrspänning  $V_G$  att vi sänker ferminivån i metallen med motsvarande energi  $qV_G$ .

I figuren ser vi också laddningsfördelningen i MOSkapacitansen med en positiv skiktladdning på metallplattan och en utarmningsladdning  $Q'_B$  i halvledaren. Dessutom finns en försumbar mängd elektroner vid halvledarytan,  $Q'_{imv} \ll Q'_B$ . Om vi nu följer den intrinsiska referensnivån från halvledarens inre mot isolatorn så ser vi att elektronenergin sjunker när vi passerar genom utarmningsområdet.



Fig. 6.9. Banddiagram för MOS-kapacitans i utarmning (svag inversion).

Vid halvledarytan har vi således en ytpotential  $\Psi_{S}$ >0. Vi vet sedan tidigare att energibandens kurvform i utarmningsområdet är parabolisk. Fortsätter vi genom isolatiorn mot metallen så ser vi att intrinsiska referensnivån sjunker ytterligare  $qV_{ox}$ , där  $V_{ox}$  är spänningsfallet över oxiden. Eftersom oxiden är laddningsfri är det elektriska fältet konstant i isolatorn och därmed sjunker energin linjärt med avståndet. Vi ser till sist att när vi är framme vid metallplattan så har den intrinsiska referensnivån sjunkit sammanlagt med energin

$$qV_{GR} = qV_{\alpha r} + q\Psi_s, \qquad (6.21)$$

där  $V_{GB}$  är spänningen över MOS-kondensatorn (*eng.* gate-bulk voltage  $V_{GB}$ ). Motsvarande spänningsfall  $V_{GB}$  illustreras ytterligar i fig. 6.10.

Ökar vi styrspänningen når vi så småningom gränsen för inversion. I det läget är ytpotentialen  $\Psi_S=2\phi_B$ , ett värde vid vilket ytpotentialen låses. All överskjutande spänning hamnar därefter över isolatorn, se fig. 6.11.



Fig. 6.10. Potentialfallet i en MOS-struktur i utarmning.



Fig. 6.11. Potentialen i en MOS-struktur i stark inversion.



Fig. 6.12. Banddiagram för MOS-kapacitans i stark inversion.

I fig. 6.12 ser vi motsvarande banddiagram för MOSkapacitansen i inversion. Vi kan se att vi nu har en ickeförsumbar mängd elektroner vid halvledarytan. Detta skikt av elektroner, *inversionskiktet*, fungerar som ett lock som skärmar av utarmningsområdet från att ytterligare påverkas av styrspänningen sedan inversionsvillkoret passerats. Vi ser vidare att med högre spänning på styret så ökar det elektriska fältet i oxiden.

Vi ser att båda figurerna är ritade så att det mesta av spänningen över MOS-kondensatorn ligger över oxiden. Figuren skulle kunna motsvara en äldre teknik med en matningsspänning typ 3,3 V. I så fall skulle ca 0,7 V ligga över utarmningsområdet, medan resten, 2,6 V, skulle ligga över oxiden. Det är naturligtvis önskvärt att så mycket som möjligt av styrspänningen hamnar över oxiden så att vi får så stor inversionsladdning som möjligt. Inverkan av utarmningsladdningen,

$$V_T = 2\phi_B + \gamma \sqrt{2\phi_B}, \qquad (6.22)$$

ska helst vara så liten som möjligt, men den kan så klart inte bli mindre än  $2\phi_B$  ens om  $\gamma=0$ . Det betyder att i modern MOS-teknik med låga matningsspänningar  $V_{DD}$  på bara cirka 1 V så blir spelrummet mindre. Om man justerar tröskelspänningen så att effektiva styrspänningen  $V_{DD}-V_T$  blir större, så blir det samtidigt svårare att stänga av transistorn ordentligt eftersom spänningsintervallet från  $V_T$  till  $V_{SS}$  (jord) blir mindre. Den kommer att vara kvar i utarmning eller i svag inversion även utan spänning på styret. Det betyder att subtröskelområdet, dvs området med styrspänningar strax under tröskelspänningen blir allt viktigare i moderna MOS-transistorer. På den tiden som Shockley utvecklade sin strömmodell för stark inversion var subtröskelområdet inte så intressant som det är idag. Därför ska vi ägna nästa avsnitt åt subtröskelområdet.

#### 6.5 MOS-transistorn i svag inversion

I detta avsnitt ska vi studera övergången mellan svag och stark inversion mer i detalj. För att kunna göra det behöver vi en modell för inversionsladdningen som är giltig i detta område. Vad vi har med oss sedan tidigare är en modell för inversionsladdningen i stark inversion (6.1). Vi har också en modell (6.18) för elektronkoncentrationen vid halvledarytan, en modell som är giltig i både svag och stark inversion. Däremot har vi ingen modell för den totala mängden elektroner i utarmningsområdet som är giltig i svag inversion.

Man borde dock kunna skapa enkel modell för denna inversionsladdning i svag inversion genom att multiplicera ytkoncentrationen med laddningens effektiva djup,  $d_{eff}$ . Det skulle ge följande modell för inversionsladdningen

$$Q'_{inv} = -qN_A d_{eff} e^{(\Psi_s - 2\phi_B)/V_t} .$$
(6.23)

Vill vi göra det enkelt för oss så antar vi helt enkelt att djupet är en konstant vars värde vi kan bestämma genom att passa modellen för svag inversion mot modellen för stark inversion. Ska vi emellertid vara riktigt fysikaliska, vilket man förstås bör vara, så måste man lösa Poissons ekvation under MOS-kapacitansens förutsättningar. En sådan analys är dock inte helt okomplicerad. Resultatet av en sådan analys ger dock efter en del beräkningar och en del approximationer vid handen att djupet kan skrivas

$$d_{eff} = V_t \sqrt{\frac{\varepsilon_{Si}}{2qN_A \Psi_S}}.$$
 (6.24)

Om vi nu skriver detta uttryck på formen

$$d_{eff} = \frac{V_t}{2\Psi_s} \sqrt{\frac{2\varepsilon_{si}\Psi_s}{qN_A}} = \frac{V_t}{2\Psi_s} W, \qquad (6.25)$$

där W är utarmningsområdets djup,

$$W = \sqrt{\frac{2\varepsilon_{si}\Psi_s}{qN_A}},$$
 (6.26)

så ser vi att inversionsladdningens djup endast är en bråkdel av utarmningsområdets. Inversionsladdningen är

således starkt koncentrerad till halvledarytan, och vi ser dessutom att inversionsladdningen koncentreras allt närmare halvledarytan ju mer vi närmar oss stark inversion. Just på gränsen till stark inversion motsvarar det effektiva djupet cirka 5 nm för ett p-typ substrat med dopningen  $N_A$ =1,25 10<sup>16</sup> cm<sup>-3</sup>; ett djup att jämföra med motsvarande utarmningsområdes djup på 270 nm.

Jämfört med ytkoncentrationens exponentiella spänningsberoende är det effektiva djupets potentialberoende i praktiken försumbart. Vår ursprungliga tanke var därför inte helt tokig. Därför är det inte heller helt orimligt att skriva inversionsladdningen i (6.23) som

$$Q'_{inv} = -Q'_{0} e^{(\Psi_{s} - 2\phi_{B})/V_{t}}, \qquad (6.27)$$

där  $Q_0$  är elektronladdningen vid tröskelspänningen. I svag inversion har vi tidigare linjariserat sambandet mellan styrspänning och ytpotential enligt

$$\Psi_s - 2\phi_B = \frac{V_{GST}}{n}.$$
(6.14)

Det ger laddningen i svag inversion följande spänningsberoende:

$$Q'_{inv} = -Q'_{0} e^{V_{GST} / nV_{t}}.$$
 (6.28)

Eftersom det är viktigt att få en modell som inte bara är kontinuerlig utan också har kontinuerlig derivata vid övergången från svag till stark inversion, så väljer vi att skarva modellen i svag inversion med modellen i stark inversion i den punkt då derivatorna är lika, dvs när  $V_{GST}=nV_t$ . Det ger

$$Q'_{inv} = -\delta C_{ox} V_t e^{V_{GST} / nV_t}, \qquad (6.29)$$

där  $\delta = n/e$  är en dimensionslös passningsparameter. Vi noterar att laddningens dimension är korrekt då  $Q'_{inv} \sim C_{ox} V_t$ .

Sammanfattningsvis har vi nu tagit fram en modell för inversionsladdningen som har ett exponentiellt spänningsberoende i svag inversion och ett linjärt spänningsberoende i stark inversion. Har vi kanske stött på sådana modeller förut? Någon som minns pn-övergångens strömkurvor? Vi sammanfattar:

$$Q'_{inv} = \begin{cases} -\delta C_{ox} V_t e^{(V_{GS} - V_T)/nV_t} & V_{GS} < V_T \\ -C_{ox} (V_{GS} - V_T) & V_{GS} > V_T \end{cases}.$$
(6.30)

Vår modell för inversionsladdningen enligt (6.27) blev ju begränsad till svag inversion när vi ersatte ytpotentialen med styrspänningen genom den linjariserade relationen i (6.14). Denna approximation är bara giltig i svag inversion. Om vi däremot backar ett steg så kan vi använda modellen i (6.27) för att studera det fysikaliskt "korrekta" sambandet mellan styrspänning och ytpotential. Detta samband kan skrivas

$$V_{GS} = \frac{Q'_G}{C_{ox}} + \Psi_S = -\frac{Q'_{inv}}{C_{ox}} + \gamma \sqrt{\Psi_S} + \Psi_S.$$
(6.31)

Efter insättning av (6.27) i (6.31) kan vi plotta styrspänningen  $V_{GS}$  som funktion av ytpotentialen  $\Psi_S$ . Resultatet visas i fig. 6.13. I figuren bekräftas tydligt våra två tidigare förenklingar: det linjära sambandet i svag inversion och ytpotentialens låsning till ett konstant värde i stark inversion.



Fig. 6.13. Ytpotentialen som funktion av  $V_{GS}$  i såväl svag som stark inversion.

Värt att notera är dock att ytpotentialen inte låser vid  $2\phi_B$ som vi väntat oss utan vid ett något större värde, cirka  $2\phi_B+6V_t$ . Det är dock inte något man i allmänhet bekymrar sig om utan man betraktar helt enkelt denna låsningsnivå som en "effektiv"  $2\phi_B$ -nivå, typ  $2\phi_{B,eff}$ .

## 6.6 Poissons ekvation i svag och stark inversion

För att härleda en fysikaliskt korrekt modell för laddningens effektiva djup i (6.24) måste vi lösa Poissons ekvation med hänsyn tagen inte bara till utarmningsladdningen som tidigare utan även till inversionsladdningen. Vi ska nu ägna detta avsnitt åt att göra just detta, vilket inte är helt okomplicerat utan snudd på överkurs. ©

I detta avsnitt kommer vi därmed också att ge prov på den lösningsmetodik och den typ av approximationer som är vanlig vid komponentmodellering. Syftet med avsnittet är således att på fysikalisk grund motivera uttrycket (6.27) för inversionsladdningen i svag inversion.

Laddningstätheten i halvledaren med hänsyn tagen till såväl utarmnings- som inversionsladdning kan skrivas

$$\rho(y) = -q \left( N_A + N_A e^{(\Psi(y) - 2\phi_B)/V_t} \right), \tag{6.32}$$

där såväl rymdladdningstätheten som potentialen i utarmningsområdet varierar med avståndet y till isolatorgränssnittet. Poissons ekvation antar då formen

$$\frac{\partial^2 \Psi}{\partial y^2} = \frac{q}{\varepsilon} \left( N_A + N_A e^{(\Psi - 2\phi_B)/V_t} \right).$$
(6.33)

Genom att multiplicera båda sidor med d $\Psi$ /dy och sedan separera variabler får vi

$$\frac{\partial^2 \Psi}{\partial y^2} \frac{\partial \Psi}{\partial y} \partial y = \frac{q}{\varepsilon} \Big( N_A + N_A e^{(\Psi - 2\phi_B)/V_t} \Big) \partial \Psi .$$
 (6.34)

Nu kan vi utföra integrationen från ytan y=0 till y= $\infty$  och får därvid

$$\left(\frac{\partial\Psi}{\partial y}\right)^{2}\Big|_{\Psi=\Psi_{S}} = \frac{2qN_{A}}{\varepsilon} \left(\Psi_{S} + V_{t}e^{(\Psi_{S}-2\phi_{B})/V_{t}}\right). \quad (6.35)$$

Enligt Gauss lag är nu det elektriska fältet vid ytan,  $d\Psi/dy$ , lika med den i utarmningsområdet inneslutna laddningen dividerad med  $\varepsilon_{Si}$ . Det ger totala laddningen i halvledaren, dvs summan av utarmnings- och inversionsladdningarna,

$$Q'_{s} = -\sqrt{2q\varepsilon N_{A}}\sqrt{\Psi_{s} + V_{t}e^{(\Psi_{s} - 2\phi_{B})/V_{t}}} . \qquad (6.36)$$

I denna ekvation som beskriver den totala laddningen i halvledaren, dvs summan av utarmnings- och inversionsladdningarna kan vi identifiera först utarmningsladdningen genom den första termen under rottecknet och sedan inversionsladdningen genom den andra termen under rottecknet. Om vi något förenklat subtraherar bort utarmningsladdningen från den totala laddningen kan inversionsladdningen skrivas

$$Q'_{inv} = -\sqrt{2q\varepsilon N_A \Psi_S} \left( \sqrt{1 + \frac{V_t}{\Psi_S} e^{(\Psi_S - 2\phi_B)/V_t}} - 1 \right).$$
(6.37)

I svag inversion när inversionsladdningen är liten jämfört med utarmningsladdningen kan vi sedan serieutveckla det stora rotuttrycket och därmed skriva inversionsladdningen som

$$Q'_{s} = -\sqrt{2q\varepsilon N_{A}\Psi_{s}} \frac{V_{t}}{2\Psi_{s}} e^{(\Psi_{s} - 2\phi_{B})/V_{t}}, \qquad (6.38)$$

vilket verifierar vår tidigare modell i (6.24) och vårt resonemang kring det effektiva laddningsdjupet  $d_{eff}$ .

Den härledning som vi nu genomfört är som synes inte helt okomplicerad och vi har behövt ta till en hel del approximationer. Låt oss dock inte haka upp oss på enskilda detaljer utan istället fokusera på det viktiga, nämligen att inversionsladdningen är exponentiellt beroende av styrspänningen  $V_{GS}$  i svag inversion. Detta i motsats till det linjära spänningsberoendet i stark inversion.

Inversionsladdningens linjära beroende av styrspänningen i stark inversion verifieras i fig. 6.14 där den kompletta laddningsmodellen i (6.37) plottats som funktion av  $V_{GS}$ , båda parametriska funktioner av ytpotentialen, tillsammans med den ursprungliga styckevis linjära modellen i (6.4).



*Fig. 6.14. Verifiering av den linjära laddningsmodellen i stark inversion.* 

# 6.7 Sammanfattning

I detta kapitel har vi introducerat MOS-strukturen och den därmed förknippade fälteffekten. Vi har visat hur MOSstrukturen kan fås i ackumulation, utarmning och inversion. I inversion skapas ett från substratet isolerat ytskikt av rörliga laddningar vars antal kan styras genom den på styreelektroden anbringade styrspänningen. I enlighet med plattkondensatorformeln antogs ett linjärt samband mellan inversionsladdningen och styrspänningen. Denna modell vilar på en styckevis linjär modell för sambandet mellan ytpotentialen och styrspänningen. I utarmning och svag inversion kan sambandet linjariseras med hjälp av en idealitetsfaktor. I stark inversion låser ytpotentialen vid dubbla värdet av substratets fermipotential. MOS-strukturens tröskelspänning definierades som den styrspänning då inversion inträffar. Banddiagram har använts för att ytterligare förtydliga MOS-strukturens funktion. En modell utvecklades för beräkning av tröskelspänningen.

Slutligen kompletterade vi den linjära laddningsmodellen i stark inversion med en exponentiell laddningsmodell i svag inversion.

## 6.8 Övningsexempel

- 1. En ideal MOS-struktur har en oxidkapacitans  $C_{ox}=3,5 \text{ mF/m}^2$ . Hur stor kapacitans motsvarar det per kvadratmikrometer?
- 2. En ideal MOS-struktur har en oxidtjocklek på 10 nm och i utarmning ett utarmningsområde på 100 nm. Hur stor är då den effektiva oxidtjockleken EOT? Använd för enkelhets skull värdena  $\varepsilon_{ox}$ =4 and  $\varepsilon_{Si}$ =12.
- 3. En ideal MOS-struktur är tillverkad på ett p-typ substrat med fermipotentialen 0,4 V. Vid vilken ytpotential  $\Psi_{\rm S}$  inträffar inversion enligt villkoret för stark inversion?
- 4. Vid gränsen till stark inversion finns laddningen -1,6 mAs/m<sup>2</sup> i utarmningsområdet. Hur stor är spänningen över oxiden om oxidkapacitansen enligt ovan är  $C_{ox}$ =3,5 mF/m<sup>2</sup>?
- 5. Hur stor är MOS-strukturens tröskelspänning enligt ovanstående beräkningar?
- 6. Inversionsladdningen:
  - a. Hur ser den styckevis linjära modellen för inversionsladdningen ut för MOS-strukturer i stark inversion?
  - b. Vid vilken styrspänning är inversionsladdningen lika stor som utarmningsladdningen?
- I detta exempel har vi en annan ideal MOS-struktur med 8 nm tjock oxid på ett p-typ substrat med dopningen 5<sup>-</sup>10<sup>16</sup> cm<sup>-3</sup>. Beräkna tröskelspänningen för denna MOS-struktur!

Som delmoment i denna uppgift ingår att beräkna

- a. oxidkapacitansen Cox [per ytenhet]!
- b. ytpotentialen  $\Psi_{s}$  vid gränsen till inversion!
- c. parametern X<sub>D</sub> och utarmningsområdets djup vid gränsen till inversion!
- d. parametern  $\gamma$  och spänningsfallet över oxiden vid gränsen till inversion!

Lös även följande relaterade problem:

- e. Hur stor är MOS-strukturens kapacitans [per ytenhet] i utarmning vid gränsen till inversion?
- f. Hur stor är MOS-strukturens kapacitans i inversion?
- g. Beräkna idealitetsfaktorn *n* genom att linjarisera relationen mellan styrspänning och ytpotential när MOS-strukturen är i utarmning (dvs i sub-tröskelområdet)!
- 8. Hur stor är elektronkoncentrationen vid halvledarytan i en MOS-struktur vid gränsen till inversion enligt "the-law-of-the-junction"?
- 9. Vilka linjariserade samband gäller mellan styrspänning och ytpotential i subtröskelområdet respektive i stark inversion?
- 10. Visa att idealitetsfaktorn också kan skrivas

 $n=1+C_B/C_{ox}$ , där  $C_B$  är utarmningskapcitansen.

- 11. Hur ser den "smoothande" laddningsmodellen för subtröskelområdet (i svag inversion och utarmning) ut och i vilken punkt ansluter denna modell till modellen för stark inversion?
- 12. Vilket är inversionsladdningens effektiva djup  $d_{eff}$ enligt följande definition om vi har skarvat modellerna enligt förra uppgiften:

 $Q_{inv} = -qN_A d_{eff} \exp(V_{GST}/n/V_t)?$ 

- 13. Rita banddiagram för en MOS-kondensator vid följande ytpotentialer
  - a.  $\Psi_{s}=0$  (flatband)
  - b.  $\Psi_{\rm S} = \phi_{\rm B}$  (onset of weak inversion)
  - c.  $\Psi_{\rm S}=2\phi_{\rm B}$  (onset of strong inversion

#### 7. MOS-TRANSISTORN

Syftet med detta kapitel är att gå igenom MOS-transistorns grunder utifrån ett fysikaliskt perspektiv. Det innebär bl a att vi ska härleda Shockleys transistormodell utifrån de fysikaliska approximationer som går under namnet "gradvisa kanalapproximationen" (eng. gradual channel approximation). Nu är Shockleys modell framtagen för transistorer med relativt lång kanallängd (typ 3-10 µm) och som varande en "långkanal" modell (eng. long channel model) är den inte utan vidare tillämpbar på moderna MOS-transistorer med extremt korta kanallängder (<100 nm). Modellen måste därför modifieras för att ta hänsyn till de andra ordningens effekter som uppträder vid mycket korta kanallängder (eng. short channel effects, SCE). Några sådana andra ordningens effekter är hastighetsmättnad, drain-induced barrier-lowering DIBL, subtröskelströmmar, mobility roll-off och kanallängdsmodulation. Inte desto mindre utgör Shockleys MOStransistormodell fortfarande grunden även för mycket moderna modeller där Shockleys ursprungliga modell byggts på med andra ordningens effekter.

Vi kommer också att se på metoder för att ur mätdata extrahera fram värden på de viktigaste modellparametrarna. I Shockleys strömmodell som gäller stark inversion är transistorkonstanten k och tröskelspänningen  $V_T$  de viktigaste parametrarna och medan det i subtröskelområdet är idealitetsfaktorn n och OFF-strömmen  $I_{OFF}$  som är de viktigaste parametrarna. De metoder vi ska använda bygger alla på "räta linjens fysik", vilket innebär att vi är ute efter lutning och x/y-intercept hos de räta linjer som utgör våra modeller.

Vi ska också se på en kapacitansmodell för MOS-transistorn samt använda energibanddiagram för att tydligare förklara MOS-transistorns funktion.

#### 7.1 Gradvisa kanalapproximationen.

För att lösa detta tvådimensionella fält- och laddningsproblem som MOS-transistorn utgör brukar man använda den gradvisa kanalapproximationen (*eng.* gradual channel approximation). Denna approximation innebär att man delar upp kanalen i segment, dx, se fig. 7.1 och att man antar att kanalpotentialen  $V_X$  är konstant i varje segment. Eftersom kanalfältet antas vara konstant i varje segment kan mängden rörlig laddning beräknas genom en endimensionell analys av varje segment. Namnet gradvis antyder då att  $V_X$  antas öka i steg  $dV_X$  mellan segmenten.



Fig. 7.1. Uppdelning av kanalen i endimensionella segment.

Om kanalpotentialen i ett segment ges av  $V_X$  kan kanalladdningen för segmentet skrivas

$$Q'_{inv}(V_x) = -WC_{ox}(V_{GST} - V_x).$$
(7.1)

Därmed har vi ett uttryck för kanalladdningen som funktion av kanalpotentialen. Nu glömmer vi den gradvisa potentialbilden och antar att segmenten är så smala att potentialen är kontinuerlig. Enligt definitionen på elektriskt fält kan drifthastigheten i varje segment då skrivas

$$v(x) = \mu \frac{dV_x}{dx} \,. \tag{7.2}$$

Strömmen I(x) genom varje segment kan då skrivas

$$W(x) = -WC_{ox}(V_{GST} - V_x)\mu \frac{dV_x}{dx}.$$
(7.3)

Eftersom strömmen är konstant längs kanalen,  $I(x)=-I_{DS}$ , kan vi nu relativt enkelt summera alla spänningsfall  $dV_X$ från source till drain genom att separera variabler och integrera över alla segment dx längs kanalen, från x=0 vid source till x=L vid drain, dvs

$$\int_{0}^{L} I_{DS} dx = W \,\mu C_{ox} \int_{0}^{V_{DS}} \left( V_{GST} - V_X \right) dV_X \,. \tag{7.4}$$

Vi får efter integration

$$I_{DS} = \frac{W}{L} \mu C_{ox} (V_{GST} - \frac{V_{DS}}{2}) V_{DS} .$$
 (7.5)

Detta är således den fysikaliska grund på vilken Shockleys MOS-transistormodell vilar. Modellens svaghet är förstås att förutsättningarna för modellen blir mindre giltiga ju närmare strömmättnad vi kommer, men å andra sidan har modellen visat sig stämma bra överens med mätdata åtminstone för MOS-transistorer med långa kanallängder.

Vi kan också beräkna hur potentialen  $V_X$  varierar med *x* genom att avbryta integrationen i en godtycklig punkt *x*, dvs

$$\int_{0}^{x} I_{DS} dx = W \mu C_{ox} \int_{0}^{V_{DS}} (V_{GST} - V_X) dV_X.$$
(7.6)

Vi får då

$$I_{DS} = \frac{W}{L} \mu C_{ox} (V_{GST} - \frac{V_X}{2}) V_X , \qquad (7.7)$$

Genom att lösa ut V(x) får vi följande uttryck som funktion av x:

$$V_X = V_{GST} \left( 1 - \sqrt{1 - \frac{x}{L}} \right). \tag{7.8}$$

Det ger följande uttryck för det elektriska fältet längs kanalen:

$$E_x = -\frac{V_{GST}}{2L\sqrt{1-\frac{x}{L}}}.$$
(7.9)

Det elektriska fältet  $E_X$  ökar tydligen enligt denna modell från  $V_{GST}/2L$  vid source till, om inte oändliga värden så i alla fall, mycket höga värden vid drain.

Med denna potentialfördelning längs kanalen ges den rörliga laddningen som funktion av x ges i en mättad transistor efter insättning i (7.5) av uttrycket

$$Q'_{inv} = -C_{ox}V_{GST}\sqrt{1-\frac{x}{L}}.$$
 (7.10)

Om man integrerar laddningen längs kanalen får man den totala inversionsladdningen i kanalen till

$$Q'_{inv} = -\frac{2}{3}C_{ox}V_{GST}.$$
(7.11)

Det betyder att i MOS-transistorns mättnadsområde gäller att

$$C_{GS} = -\frac{\partial Q_G}{\partial V_S} = \frac{\partial Q_{inv}}{\partial V_S} = 2/3C_G, \qquad (7.12)$$

och

$$C_{GD} = -\frac{\partial Q_G}{\partial V_D} = \frac{\partial Q_{inv}}{\partial V_D} = 0.$$
(7.13)

Sourcekapacitansen är således större i mättnadsområdet än vad den är i resistansområdet, men å andra sidan är drainkapacitansen nu försumbar. På grund av den ojämna laddningsfördelningen i kanalen i en mättad MOS-transistor är den totala styrekapacitansen endast två tredjedelar av styrekapacitansen i resistansområdet. I förenklade överslagsberäkningar brukar man dock i allmänhet bortse från denna faktor 2/3, något som kan berättigas genom att kanteffekter ger upphov till en kapacitans som motsvarar just  $C_G/3$ . Tyvärr finns dessa kanteffekter också mellan styre och drain vilket komplicerar analys med papper och penna. CGD försummas därför i överslagsberäkningar eller betraktas som inbakade i  $C_{GS}$  och  $C_{DB}$ . Vi får då den kapacitansmodell som visas i fig. 7.2 där vi också har lagt till parasitkapacitansen CDB mellan drain och source/ substrat på grund av den backspända drain/substratdioden.



Fig. 7.2. MOS kapacitansmodell i mättnadsområdet med kant- och parasitkapacitanser.

#### 7.2 MOS-transistorns banddiagram

Viss hjälp i den fysikaliska förståelsen av MOS-transistorn kan man få genom att rita dess banddiagram. I Fig. 7.3 ser vi banddiagrammet i längsled, från source till drain, utan och med pålagd drainspänning.

Vi ser att en pålagd drainspänning inte är tillräckligt som villkor för strömtransport eftersom potentialbarriären



Fig. 7.3. Banddiagram i längsled för MOS-transistorn, a) utan och b) med pålagd drainspänning.

mellan source och substrat effektivt blockerar all injektion av laddningsbärare. På något sätt behöver vi således sänka potentialbarriären så att laddningsbärarna kan komma in i kanalen. Denna sänkning sker för MOS-transistorn kapacitivt via en spänning på styret, en effektiv styrspänning  $V_{GST}$ . En positiv styrspänning sänker potentialbarriären i kislet så att elektroner kan injiceras till drain.

Med ett banddiagram ritat i tvärsled från styre genom isolatorn till substrat kan vi nu i Fig. 7.4 illustrera hur en positiv styrspänning sänker energibanden i substratet så att de kommer i nivå med banden i source. Därigenom sänks också potentialbarriären. Översta bilden visar det s.k. flatbandsvillkoret, d.v.s. att energibanden i kislet är helt flata (horisontella). Det innebär att kislet är oladdat och i en ideal MOS-struktur har vi då inte heller någon laddning på styret. Detta är förstås den förväntade situationen när MOS-strukturen är utan pålagd styrspänning. Undre bilden visar att en positiv spänning  $V_{GS}$  på styret har sänkt styrets ferminivå. En del av den pålagda styrspänningen hamnar över oxiden, och en del hamnar över kislet från ytan och inåt till utarmningsområdets kant. Den spänning som hamnar över kislet motsvarar ytpotentialen  $\psi_S$  och som vi tidigare diskuterat är denna potential vid stark inversion dubbelt så stor som substratet fermipotential  $\phi_B$ , d.v.s.  $\psi_S = 2\phi_B$ . Ett typiskt värde på ytpotentialen vid inversion är 0,6-0,7V.

Banddiagrammet i längsled vid svag och stark inversion visas i Fig. 7.5. Jämfört med utgångsläget i fig. 7.4 har vi nu lagt på en styrspänning som sänkt potentialbarriären något. I den övre bilden är vi fortfarande kvar i området med svag inversion, dvs i subtröskelområdet,  $V_{GS}$  $< V_T$ . Vi ser att kanalområdet i princip är fältfritt. Subtröskelströmmen måste därför vara en diffusionsström. I den undre bilden har vi ökat styrspänning så mycket att MOS-transistorn nu befinner sig i stark inversion. Potentialbarriären mellan source och kanal är i det närmaste eliminerad.



Fig. 7.4. Banddiagram i tvärsled för ideal MOStransistor. a) Utan pålagd styrspänning med flata band, b) med pålagd styrspänning i stark inversion.



*Fig. 7.5. Banddiagram för MOS-transistorn i längsled, a) i svag inversion och b) i stark inversion.* 

Vi ser att vi nu har ett elektriskt fält i kanalen och att strömmen nu i stark inversion måste vara en driftström i enlighet med våra tidigare diskussioner. Potentialen i kanalen ges av uttrycket i (7.8). Vi ska nu avsluta detta avsnitt om banddiagram genom att visa en tvådimensionell figur där man samtidigt kan se banddiagrammen i både längs- och tvärsled. Vi måste då först tänka oss en tvådimensionell bild av MOS-transistorn så som visas i Fig. 7.6. Med MOS-transistorn i denna figur som referens kan vi nu rita det tvådimensionella banddiagrammet i fig. 7.7.



Fig. 7.6. MOS-transistor i genomskärning.



Fig. 7.7. Tvådimensionellt banddiagram för MOSFET.

## 7.3 Parameterextraktion

En mycket viktig aspekt när man bygger modeller för komponenter i allmänhet och för MOS-transistorer i synnerhet är förstås att modellen måste passa till mätdata. Att passa en modell till mätdata innebär att man bestämmer eller extraherar modellparametrarna så att felet mellan modell och mätdata blir så litet som möjligt. Det finns i huvudsak två metoder för att göra detta på, lokala och globala metoder.

Den lokala metoden innebär att man bestämmer ett fåtal parametrar åt gången. Om en modell är linjär kan man bestämma de två parametrarna från lutning och xeller y-interceptet för den räta linje som utgör modellen. Detta kan man göra genom att prova sig fram, eller mer metodiskt genom att göra en minsta-kvadratanpassning (lineär regression) mellan mätdata och modell. Den globala metoden innebär att man använder något optimeringsprogram som bestämmer alla parametrarna samtidigt genom att minimera felet. Även för den globala metoden kan det vara bra att göra en lokal extraktion först för att ge optimeringsprogrammet lämpliga startvärden.

Ett typiskt exempel på lokal parameterextraktion är den metod som man ofta använder för att bestämma tröskelspänningen  $V_T$  och parametern k i resistansområdet. Modellparametrarna bestäms från lutning och x-intercept för den streckade trendlinjen som får representera modellen i det mest relevanta mätdataintervallet. Se Fig. 7.8. Här har man bestämt ekvationen för den trendlinje som passar bäst till mätdata i det förväntade intervallet där överensstämmelsen är som bäst och där andra ordningens effekter kan försummas. Från trendlinjens ekvation kan vi beräkna transistorkonstanten k ur sambandet

$$k = \frac{5}{V_{DS}} = 100 \ \mu \text{A/V}^2. \tag{7.14}$$

och x-interceptet till 1,725/5=0,345 V. Därifrån kan vi bestämma tröskelspänningen till  $V_T$ =0,345-0,05/2=0,32V. Samtidigt ser vi att verklighetens MOS-transistor inte är helt linjär utom för små  $V_{GST}$ . Det ser ut som om modellen borde ha åtminstone en parameter till för att beskriva att transistorkonstanten k verkar avta med ökande styrspänning  $V_{GS}$ . Detta icke-linjära spänningsberoende kommer betecknande nog från en andra ordningens effekt som går under namnet *mobility roll-off*.

På liknande sätt borde man kunna bestämma k och  $V_T$  i mättnadsområdet genom att plotta  $\sqrt{I_{DS}}$  vs  $V_{GS}$ . Se Fig. 7.9.



Fig. 7.8. Extraktion av k och  $V_T$  från den streckade räta linjen.  $V_{DS}$ =50 mV. Resultat: k=100  $\mu A/V^2$  och  $V_T$ =0,32 V.

Eftersom mättnadsströmmen har ett kvadratiskt spänningsberoende förväntar vi oss att en sådan plott resulterar i en rät linje så att k och  $V_T$  kan bestämmas från trendlinjens lutning och x-intercept på liknande sätt som i resistansområdet.

Från trendlinjens skärning med x-axeln får vi  $V_T=0,29$  V och från dess lutning får vi transistorkonstanten ur sambandet

$$k = 2 \cdot (lutningen)^2 = 68 \ \mu A/V^2.$$
 (7.15)



Fig. 7.9. Extraktion av k och  $V_T$  från den streckade räta linjen. Resultat:  $k=66 \ \mu A/V^2$  och  $V_T=0,29 \ V$ .

Här kan vi nu konstatera att vi inte får samma parametervärden i resistans- och mättnadsområdena. Uppenbarligen är det något som inte är helt hundra med modellen. Detta ser vi ännu tydligare om vi plottar utgångskarakteristiken med parametervärden från såväl resistansområdet, se Fig. 7.10, som med parametervärden från mättnadsområdet, se Fig. 7.11.

Om man som i fig. 7.10 använder parametervärden extraherade i resistansområdet så stämmer modellen följaktligen bra i resistansområdet, men den stämmer inte lika bra i mättnadsområdet. Det vill säga *så* dålig är inte modellen; felet är mindre än 50%. Storleksordningen är således korrekt. I Fig. 7.11 ser vi på motsvarande sätt att om modellparametrarna bestäms från mätdata i mättnadsområdet så stämmer modellen inte så bra i resistansområdet. TILL-resistansen  $R_{ON}$  får inte alls korrrekta värden. Det är uppenbart att vi för större noggrannhet behöver en modell med fler modellparametrar. Detta leder oss helt osökt in på andra ordningens effekter. Vi ser i fig. 7.12 att med en modell med fler modellparametrar så kan man uppnå ganska god passningsnoggrannhet.



Fig. 7.10. Utgångskarakteristiken med modellparametrarna bestämda i resistansområdet. Modell (linje) och mätdata (symbol).



Fig. 7.11. Utgångskarakteristiken med modellparametrarna bestämda i mättnadsområdet.



Fig. 7.12. Utgångskarakteristiken med fler modellparametrar för att ta hänsyn till andra ordningens effekter. Modell (linje) och mätdata (symbol).

## 7.4 Andra ordningens effekter

Som vi nämnde i inledningen till kapitlet är Shockleys MOS-transistormodell ursprungligen framtagen för transistorer med relativt lång kanallängd (typ 3-10  $\mu$ m). Den är därför inte utan vidare tillämpbar på transistorer med dagens extremt korta kanallängder (typ 22-90 nm). Dock kan modellen ges viss giltighet även för transistorer med kort kanallängd om man korrigerar den för ett antal andra ordningens effekter. Några sådana andra ordningens effekter är t ex

- substrateffekten
- mobility roll-off,
- hastighetsmättnad,
- drain-induced barrier-lowering (DIBL)
- subtröskelström,
- kanallängdsmodulation

Vi ska nu kortfattat behandla dessa andra ordningens effekter. Det primära målet med kapitlen om MOS-transistorn i denna kurs är visserligen att man bekvämt ska kunna behärska och använda sig av den MOS-modell som bygger på första ordningens effekter, men det kan vara viktigt att känna till att man måste ta hänsyn även till andra ordningens effekter för att få god överensstämmelse mellan modell och mätdata för dagens moderna transistorer med extremt korta kanaler.

## 7.4.1 Substrateffekten

Den första andra ordningens effekt som vi ska ta upp är substrateffekten. Detta effekt tar hänsyn till att utarmningsområdet mellan kanal och substrat ökar i djup längs kanalen från source till drain, se fig. 7.13. Kanalpotentialen  $V_X$ , dvs kanalens backspänning mot substratet, ökar längs kanalen eftersom drain är mer backspänd än source. Det innebär in sin tur att utarmningsområdet som är ganska tunt vid source växer i djup mot drain. Det innebär förstås också att utarmningsladdningen  $Q'_B$  ökar med ökande  $V_X$  och att vi i motsvarande grad får en mindre mängd rörlig inversionsladdning. Detta kan man hantera i modellen för inversionsladdningen genom att öka inverkan av  $V_X$ . Vanligtvis linjariserar man sambandet mellan  $Q'_B$  och  $V_X$  vilket sammanfattningsvis ger en modell för inversionsladdningen enligt

$$Q'(V_X) = -WC_{\alpha X}(V_{GST} - \alpha V_X).$$
(7.16)



Fig. 7.13. Utarmningsområdet ökar i djup längs kanalen eftersom backspänningen  $V_X$  till substratet är större vid drain än vid source.

Denna laddningsmodell resulterar i följande strömmodell:

$$\begin{cases} I_{DS} = k(V_{GST} - \alpha \frac{V_{DS}}{2})V_{DS} & V_{DS} \leq \frac{V_{GST}}{\alpha} \\ I_{DSAT} = \frac{k}{2\alpha}V_{GST}^{2} & V_{DS} \geq \frac{V_{GST}}{\alpha} \end{cases}.$$
(7.17)

I denna modifiering av Shockleys modell har vi således infört en ny parameter  $\alpha$  för att hantera substrateffekten. I Shockleys modell var  $\alpha$ =1, men nu blir  $\alpha$ >1.

Eftersom denna strömmodell har tre modellparametrar kan nu k och  $V_T$  bestämmas för att ge bästa passning i resistansområdet, medan  $\alpha$  därefter, med k och  $V_T$  givna, kan bestämmas för bästa passning i mättnadsområdet. Vi observerar samtidigt att mättnadsvillkoret modifierats något, och att MOS-transistorn nu med hänsyn tagen till substrateffekten går i mättnad vid en något lägre drainspänning. Sambandet mellan mättnadsspänningen och styrspänning är fortfarande linjärt men modifierat med parametern  $\alpha$  enligt

$$V_{DSAT} = \frac{V_{GST}}{\alpha}.$$
 (7.18)

## 7.4.2 Mobility roll-off

Nästa andra ordningens effekt som vi ska behandla är mobility roll-off. Denna effekt innebär att rörligheten i kanalen minskar när styrspänningen ökar. Detta beror på att laddningsbärarnas kollisionsfrekvens mot kiselytan påverkas av det vertikala fältet mot kiselytan, d.v.s. av  $V_{GST}$ . Enkelt uttryckt drar fältet laddningsbärarna närmare ytan så att de kolliderar oftare. När kollisionsfrekvensen ökar så sjunker rörligheten; som vi diskuterat tidigare så adderas kollisionsfrekvenserna för olika kollisionsmekanismer enligt Mathiessens regel och rörligheten är omvänt proportionell mot kollisionsfrekvensen. För att ta hänsyn till denna effekt i transistormodellen modelleras rörligheten ofta med ett uttryck av typen

$$\mu = \frac{\mu_0}{1 + \theta_1 V_{GST} + \theta_2 V_{GST}^2},$$
(7.19)

där  $\theta_1$  och  $\theta_2$  är modellparametrar. Ett enkelt sätt att bestämma dessa modellparametrar är att plotta

$$\frac{kV_{GST}V_{DS}}{I_{DS}} = 1 + \theta V_{GST} + \theta_2 V_{GST}^2 , \qquad (7.20)$$

som funktion av  $V_{GST}$  och sen bestämma parametrarna  $\theta_1$ och  $\theta_2$  med lineär regression utgående från att man redan vet *k* och  $V_T$ . För att få modellen att passa till mätdata i fig. 7.12 har vi använt oss av parametervärdet  $\theta$ =0,1 V<sup>-1</sup>.

#### 7.4.3 Drain-induced barrier-lowering (DIBL).

Nästa effekt, drain-induced barrier-lowering (DIBL), förklarar varför vi får olika tröskelspänning i resistans- och mättnadsområdena. Som ett exempel bestämde vi tidigare tröskelspänningen till 0,32 V i resistansområdet och till 0,29 V i mättnadsområdet. Detta fenomen att tröskelspänningen minskar när man ökar drainspänningen beror på att i dagens kortkanaliga MOS-transistorer befinner sig drain så nära source att potentialbarriären vid source faktiskt påverkas av drainspänningen  $V_{DS}$ , se energibanddiagrammet i Fig. 7.14. I allmänhet modelleras DIBLeffekten med ett enkelt linjärt uttryck

$$V_{T} = V_{T0} - \sigma V_{DS}, \qquad (7.21)$$

där  $\sigma$  är en modellparameter för DIBL.



Fig. 7.14. Banddiagram för MOS-transistor med DIBL.

## 7.4.4 Hastighetsmättnad

En av de mest betydelsefulla andra ordningens effekter i moderna kortkanaltransistorer är fenomenet hastighetsmättnad. Det innebär att laddningsbärarna nära drain uppnår mättnadshastigheten  $v_{sat}$  vilket leder till strömmättnad. Tidigare i avsnitt 1.10 utvecklade vi den styckevis linjära hastighetsmodell som för vår bekvämlighet visas igen i Fig. 7.15. I denna hastighetsmodell finns ett kritiskt fält  $E_{krit}=v_{sat}/\mu$  vid vilket laddningsbärarna uppnår sin maximala hastighet  $v_{sat}$ , och där således den linjära relationen  $v=\mu E$  inte längre är giltig. I en MOS-transistor motsvarar detta kritiska fält en drain/sourcespänning  $V_C=2Lv_{sat}/\mu$ .

Hastighetsmättnaden påverkar mättnadsspänningen  $V_{DSAT}$  som då inte längre bestäms enbart av det sk "pinchoff"-villkoret (försumbar drainladdning). Man kan visa att strömmättnaden vid hastighetsmättnad påverkas lika mycket av båda effekterna, d.v.s. både av att drainladdningen blir försumbar och att hastighetsmättnad inträffar.



Fig. 7.15. Rörlighetsmodell med hastighetsmättnad.

Detta kan vi visa med följande starkt förenklade resonemang: vid strömmättnad,  $V_{DS}=V_{DSAT}$ , rör sig laddningen vid drain med mättnadshastigheten och kan således skrivas

$$I_{DSAT} = WC_{ox} \left( V_{GST} - \alpha V_{DSAT} \right) v_{sat} \,. \tag{7.22}$$

Om vi sätter denna ström lika med strömmen enligt den linjära strömmodellen i resistansområdet och inser att denna når mättnadsvärdet vid halva mättnadsspänningen får vi mättnadsspänningen

$$V_{DSAT} = \frac{V_{GST}V_C}{V_{GST} + \alpha V_C}.$$
(7.23)

En analys av det nya uttrycket för mättnadsspänningen visar att för små  $V_{GST}$  ( $V_{GST} << V_C$ ) så är hastighetsmättnaden försumbar och MOS-transistorn som tidigare kvadratisk i  $V_{GST}$  eftersom  $V_{DSAT} \rightarrow V_{GST}/\alpha$ . För stora  $V_{GST}$  blir inverkan av hastighetsmättnaden alltmer märkbar eftersom  $V_{DSAT} \rightarrow V_C$  varvid MOS-transistorn blir alltmer linjär som funktion av  $V_{GST}$  istället för som tidigare kvadratisk. Detta framgår också av fig. 7.12 där vi använt  $V_C$ =4 V för bästa passning mot mätdata.

Mättnadsströmmen kan då eftyer insättning av (7.23) i (7.22) skrivas

$$I_{DSAT} = \frac{k}{2} V_{GST} V_{DSAT} \,. \tag{7.24}$$

Vi ser att detta uttryck också motsvarar att den linjära modellen för strömmen i resistansområdet når mättnad redan för halva mättnadsspänningen. En styckevis linjär modell har alltid sin brytpunkt vid halva mättnadsspänningen jämfört med en parabolisk "square-law" modell.

En enkel "smoothing" kurva som förenar resistansoch mättnadsområdena ges av parabeln

$$I_{DS} = k V_{GST} \left( 1 - \frac{V_{DS}}{2V_{DSAT}} \right) V_{DS} \,. \tag{7.25}$$

Denna "smoothing" modell är en något förenklad variant av Berkeleyuniversitetets välkända och i industrin ofta använda simuleringsmodell BSIM. Enligt BSIM-modellen ges strömmen i resistansområdet av uttrycket

$$I_{DS} = k \frac{V_{GST} V_{DS} - \frac{V_{DS}^2}{2}}{1 + \frac{V_{DS}}{V_C}},$$
(7.26)

en modell som har samma mättnadsspänning och samma mättnadsström som vår förenklade modell. Resultatet av en passning av en komplett modell till mätdata för en modern MOS-transistor med 60 nm kanallängd visas i Fig. 7.16.



Fig. 7.16. MOSFET-modell passad till mätdata.

# 7.4.5 Subtröskelström

I en verklig MOS-transistor är inte övergången mellan tilloch frånlägena, mellan *ON* och *OFF*, så abrupt som en styckevis linjär modell antyder. Istället sker en gradvis avstängning av MOS-transistorn med en ström som avtar exponentiellt med minskande styrspänning. Detta beteende är mycket likt diodens avstängningsområde då även diodkarakteristiken går från linjär till exponentiell vid minskande diodspänning.

Den ström som flyter genom en MOS-transistor för styrspänningar under tröskelspänningen kallas inte helt omotiverat för subtröskelström. Eftersom MOS-transistorn "borde" kunna betraktas som avstängd, *OFF*, för styrspänningar under tröskelspänningen betraktas subtröskelströmmen i många sammanhang som en läckström. Till den strömmodell som vi tidigare tagit fram för stark inversion ( $V_{GS} > V_T$ ) måste vi därför lägga en modell för subtröskelströmmen ( $V_{GS} < V_T$ ). En sådan exponentiell strömmodell kan formuleras på följande sätt

$$I_{sub} = I_0 e^{V_{GST} / nV_t} \left( 1 - e^{-V_{DS} / V_t} \right),$$
(7.27)

där *n* är subtröskelområdets idealitetsfaktor. En komplett strömmodell för både svag inversion (i subtröskelområdet) och stark inversion visas i Fig. 7.17.

Läckströmmen I<sub>OFF</sub> vid styrspänningen V<sub>GS</sub>=0, d.v.s.

$$I_{OFF} = I_0 e^{-V_T / nV_t} , (7.28)$$

är en särskilt viktig storhet när det gäller moderna kretsar eftersom den talar om hur mycket läckström som flyter genom en avstängd MOS-transistor som är spänningssatt mellan source och drain.



Fig. 7.17. MOS-modell med subtröskelström. Streckad linje motsvarar Shockleys modell i stark inversion.

I ett SRAM-minne finns det två sådana läckande transistorer per minnescell och med kanske en miljard minnesceller per krets kan det bli en avsevärd effektutveckling. Det är inte ovanligt att man bland en MOS-transistors prestanda speciellt anger förhållandet  $I_{ON}/I_{OFF}$  för att framhäva dess prestanda. Vi talar alltså om förhållandet mellan tillströmmen  $I_{ON}$  (för  $V_{GS} = V_{DD}$ ) och frånströmmen  $I_{OFF}$  (för  $V_{GS}=0$ ) (i båda fallen är  $V_{DS}=V_{DD}$  där  $V_{DD}$  är matningsspänningen). Idealitetsfaktorn ligger typiskt mellan ett och två. Med hjälp av idealitetsfaktorn kan man också beräkna subtröskelsvinget  $S = nV_t ln10$ , d.v.s. den ändring av styrspänningsen som krävs för att öka eller minska subtröskelströmmen med en faktor tio.

Precis som för dioden är den exponentiella subtröskelströmmen en diffusionsström. Om vi återvänder till Fig. 7.5a där vi illustrerat MOS-transistorns banddiagram i kanalen med horisontella energiband, förstår vi att det elektriska fältet i kanalen är försumbart vid svag inversion. Driftströmmen kan således försummas i subtröskelområdet. Om vi tecknar subtröskelströmmen som en ren diffusionsström kan vi skriva den som

$$I_{sub} = -WD_n \frac{dQ'}{dx} = W \mu V_t \frac{Q'_s - Q'_D}{L}.$$
 (7.29)

där  $Q'_{S}$  och  $Q'_{D}$  är kanalladdningen per ytenhet vid source respektive vid drain [As/m<sup>2</sup>]. För elektroner som rör sig genom diffusion enbart kan kontinuitetsekvationen i stationärtillstånd skrivas

$$\frac{1}{q}\frac{dJ_n}{dx} = D_n \frac{d^2n}{dx^2} = 0.$$
 (7.30)

Denna differentialekvation har en lösning där elektronladdningen Q'(x)=qn(x) i den svagt inverterade kanalen avtar linjärt från värdet  $Q'_S$  vid source till  $Q'_D$  vid drain enligt följande:

$$Q'(x) = Q'_{s}\left(1 - \frac{x}{L}\right) + Q'_{D}\frac{x}{L}.$$
 (7.31)

Att elektronladdningen i kanalen avtar linjärt betyder förstås att derivatan, dvs strömmen, är konstant mellan source och drain. Det innebär enkelt uttryckt att det inte "försvinner" någon ström på vägen längs kanalen.

Vi har tidigare tagit fram en modell för kanalladdningen vid svag inversion och försumbar drainspänning,

$$Q'_{inv} = -\delta C_{ox} V_{l} e^{(\Psi_{s} - 2\phi_{B})/V_{l}} .$$
 (6.29)

Vid icke-försumbar drainspänning minskar kanalladdningen vid drain på grund av den minskande spänningen  $V_{GD}$  över MOS-kapacitansen vid drain. Vi kan då teckna elektronladdningen i svag inversion vid source respektive drain som

$$Q'_{s} = -\delta C_{ox} V_{t} e^{(\Psi_{s} - 2\phi_{B})/V_{t}}.$$
(7.32a)

$$Q'_{D} = -\delta C_{ox} V_{t} e^{(\Psi_{s} - 2\phi_{B} - V_{DS})/V_{t}} = Q'_{S} e^{-V_{DS}/V_{t}}.$$
 (7.32b)

Subtröskelströmmen kan efter insättning i (7.29) av dessa båda uttryck för source- och drainladdningarna skrivas som

$$I_{sub} = \frac{W}{L} \mu V_t Q'_s \left( 1 - e^{-V_{DS}/V_t} \right).$$
(7.33)

eller på formen

$$I_{sub} = \delta k V_t^2 e^{(\Psi_s - 2\phi_B)/nV_t} \left( 1 - e^{-V_{DS}/V_t} \right).$$
(7.34)

Vidare antog vi följande approximativa relation mellan ytpotential och styrspänning i svag inversion:

$$\Psi_s - 2\phi_B = \frac{V_{GST}}{n} \,. \tag{6.14}$$

Subtröskelströmmen kan således efter insättning skrivas

$$I_{sub} = I_{OFF} e^{V_{GS} / nV_t} \left( 1 - e^{-V_{DS} / V_t} \right),$$
(7.35)

där

$$I_{OFF} = \delta k V_t^2 e^{-V_T / n V_t}.$$
 (7.36)

Alternativt kan subtröskelströmmen uttryckas mha subtröskelsvinget *S*,

$$I_{sub} = I_{OFF} 10^{V_{GS}/S} \left( 1 - e^{-V_{DS}/V_t} \right),$$
(7.37)

där  $S=nV_t ln 10$ . Subtröskelsvinget motsvarar den styrspänningsökning som krävs för en tiofaldig ökning av subtröskelströmmen.

Kvar att bestämma är nu enbart den styrspänning vid vilken strömmen i en förenklad modell övergår från att vara en diffusionsström till att vara en driftström.

**Exempel 7.1**. Visa att skarvpunkten mellan modellerna för svag och stark inversion ges av  $V_{GST}=2nV_t$ , samt bestäm passningsparametern  $\delta$  för strömkontinuitet!

**Lösning**. Eftersom det är önskvärt att transkonduktansen, dvs derivatan, är kontinuerlig så väljer vi detta villkor för att bestämma skarvpunkten. Transkonduktansen i stark inversion ges av

$$g_m = \frac{I_{DS}}{V_{GST}/2},$$

medan den i subtröskelområdet ges av

$$g_m = \frac{I_{DS}}{nV_t}$$

Sätter vi båda uttryckena lika så får vi  $V_{GST} = 2nV_t$ . Nu återstår att bestämma passningsparametern  $\delta$  så att strömmodellen blir kontinuerlig. Lika ström i skarvpunkten ger

$$\delta = \frac{2n^2}{e^2}.$$

En komplett strömmodell för såväl resistans- som mättnadsområdena passad till mätdata visas i Fig. 7.18.



Fig. 7.18. Subtröskelströmsmodell för MOS-transistor.

#### 7.4.6 Kanallängdsmodulation

Den sista andra ordningens effekt som vi ska se på här i detta kapitel är kanallängdsmodulation (*eng.* channel length modulation, CLM). Som namnet antyder innebär detta att den effektiva kanallängden varierar med drainspänningen. Det betyder att bredd/längdförhållandet W/L och därmed också strömkonstanten k påverkas av drainspänningen. Detta innebär i sin tur att MOS-transistorns utgångskonduktans ökar, dvs att strömmen inte riktigt mättar i mättnadsområdet utan att den fortsätter att öka med ökande drainspänning. Det är samma inverkan på utgångskarakteristiken som DIBL-effekten har.

Vi har redan tidigare konsterat att en MOS-transistor lämnar resistansområdet vid en viss drainspänning  $V_{DSAT}$ . Vi har lärt oss hur man beräknar  $V_{DSAT}$  under inverkan av såväl pinch-off effekten och hastighetsmättnad. I båda fallen har vi konstaterat att kanalladdningen blir försumbar vid drain vid denna drainspänning. Spänningsfallet över den inverterade kanalen är således begränsat till V<sub>DSAT</sub>. Om man då ökar drainspänningen ytterligare, var hamnar då detta spänningsfall om det inte kan hamna över den inverterade kanalen? Den effektiva kanallängden  $L_{eff}$ som ingår i MOS-modellens bredd/längdförhållande är helt enkelt definierad av den punkt där kanalladdningen blir försumbar. Resten av drainspänningen,  $V_{DS}-V_{DSAT}$ , hamnar över ett litet område  $\Delta L$  mellan denna pinch-off punkt och drain. Se fig. 7.19. Ju större drainspänning desto större behöver detta område vara för att kunna ta upp det överskjutande spänningsfallet, V<sub>DS</sub>-V<sub>DSAT</sub>.

Detta innebär att "pinch-off"-punkten förskjuts i riktning mot source när drainspänningen ökar. Vid beräkning av detta områdes utbredning betraktas det som utarmat på laddningsbärare – dvs det behandlas som ett utarmningsområde. Som vi minns från tidigare kapitel så har vi särskilda metoder för att beräkna utbredningen hos utarmningsområden. Fast i detta sammanhang är det mer komplicerat eftersom problemet är tvådimensionellt och det är svårt att förenkla även om ska visa ett sätt nedan.



Fig. 7.19. Kanallängdsmodulation.

Kanalen består alltså av en starkt inverterad del med längden L- L, och en utarmad del med längden L. Rent matematiskt och modellmässigt behandlar vi dessa båda delar av kanalen på olika sätt. Den starkt inverterade kanalen kan behandlas som förut om vi ersätter den nominella kanallängden med den effektiva kanallängden  $L_{eff} = L - \Delta L$ . Som vi just nämnde har vissa försök gjorts att beräkna  $\Delta L$  med hjälp av utarmningsområdesmodellen men oftast nöjer man sig med en enkel linjär modell,

$$L_{eff} = L(1 - \lambda V_{DS}), \qquad (7.38)$$

där  $\lambda$  är en modellparameter. Med approximationen

$$1 - \lambda V_{DS} \approx \frac{1}{1 + \lambda V_{DS}}, \qquad (7.39)$$

som är en god approximation för små  $\lambda V_{DS}$  så resulterar denna modell för den effektiva kanallängden i följande uttryck för strömmen i resistans- respektive mättnads-områdena:

$$I_{DS} = \begin{cases} k V_{GST} V_{DSAT} \left( 1 - \frac{V_{DS}}{2V_{DSAT}} \right) (1 + \lambda V_{DS}) \\ \frac{k}{2} V_{GST} V_{DSAT} \left( 1 + \lambda V_{DS} \right) \end{cases}.$$
(7.40)

Inversen av den nya modellparametern,  $-1/\lambda$ , är den så kallade Earlyspänningen (den spänning vid vilken den extrapolerade strömmen blir lika med noll), se Fig. 7.20.



Fig. 7.20. MOS-transistor med kanallängdsmodulation.



Fig. 7.21. Ekvivalentschema för MOS-transistor med kanallängdsmodulation.

I fig. 7.21 illustreras att kanallängdsmodulationen ger upphov till en utgångskonduktans,

$$g_d = \lambda \frac{k}{2} V_{GST} V_{DSAT} = \lambda I_{DSAT} .$$
(7.41)

Ett potential- och fältdiagram för en MOS-transistor med kanallängsmodulation visas i fig. 7.22. I denna figur följer kanalpotentialen och det elektriska fältet i den inverterade delen av kanalen de tidigare modellerna från (7.8) och (7.9). I den utarmade delen av kanalen har vi använt utarmningsmodellen med fältrandvärdet

$$E\left(L_{eff}\right) = E_{P}.\tag{7.42}$$

Fältbilden blir då enligt Gauss sats en parallelltrapets med

$$E(L_{eff}) = E_P$$
 respective  $E(L) = E_P + \frac{qN_A\Delta L}{\varepsilon_{Si}}$ . (7.43)

Man kan visa att  $\Delta L$  under dessa förutsättningar kommer att ges av uttrycket

$$\Delta L = X_D \left( \sqrt{V_{bi} + V_{DS} - V_{DSAT}} - \sqrt{V_{bi}} \right), \tag{7.44}$$

där  $V_{bi} = \left(\frac{X_D E_P}{2}\right)$ .



Vi kan i förbigående notera att uttrycket för  $\Delta L$  har formen av en skillnad mellan två utarmningsområden.

Fig. 7.22. Fält och potential vid kanallängdsmodulation.

# 7.5 MOS-transistorns utarmningsområde

Vi har redan tidigare i samband med bestämning av en MOS-transistors tröskelspänning studerat inverkan av utarmningsområdet under kanalen. I en första ansats antog vi något förenklat att utarmningsområdets djup under kanalen är konstant. Men efter att man insett att utarmningsområdet på grund av drainspänningen måste vara djupare under drain än under source, inser man förstås att utarmningsområdet även växer i djup med den ökande kanalpotentialen längs kanalen. Se fig 7.23. Tar man hänsyn till denna andra ordningens substrateffekt inverkan på mängden rörlig laddning i kanalen kan Shockleys MOS-transistormodell modifieras med ytterligare en parameter.

Ibland kallas utarmningsområdet i en backspänd pnövergång för spärrskikt eftersom den potentialbarriär som är förknippad med utarmningsområdet hindrar ström från att flyta. I det här sammanhanget är begreppet spärrskikt ganska beskrivande eftersom det antyder att utarmningsområdet uppträder som en "spärr" som isolerar MOStransistorn från substratet. MOS-transistorn är med andra ord "självisolerande", dvs isolerad från substratet genom sin egen struktur och genom hur den är spänningssatt.

Det betyder att man kan bygga integrerade kretsar med hundratals miljoner MOS-transistorer placerade i samma substrat eftersom spärrskikten isolerar transistorerna från varandra. Alla MOS-transistorer kan därför ha olika source- och drainspänningar – inga problem; allt som sker är att de isolerande spärrskikten anpassar sig i tjocklek till de aktuella backspänningarna. Detta är ett av skälen till att man med MOS-transistorer har kunnat bygga integrerade kretsar med så oerhört hög packningstäthet.

Observera att i en MOS-transistor är source och drain hårddopade medan substratet är lågdopat. Därför har endast spärrskiktets utbredning i substratet ritats ut i figurerna. Eftersom ett spärrskikt är förknippat med en spärrskiktskapacitans uppstår i MOS-transistorn oönskade kapacitanser mellan source och substrat, och mellan drain och substrat, sk parasitkapacitanser. Oftast är sourcepotentialen densamma som substratpotentialen och då är kapacitansen  $C_{SB}$  kortsluten och utan inverkan. Däremot har drainkapacitansen  $C_{DB}$  stor inverkan på MOS-transistorns prestanda så den är viktig att ta med i modellen för MOS-transistorn.



Fig. 7.23. Utarmningsområdet i en MOS-transistor och dess parasitkapacitanser.

## 7.6 Sammanfattning

I detta kapitel har vi förklarat de grundläggande principerna bakom fälteffekten och hur MOS-transistorn är elektriskt styrbar så att den kan slås till och från. Vi har förklarat varför strömmen genom en MOS-transistor mättas på grund av en kombination av "pinch-off" och hastighetsmättnad. Gradvisa kanalapproximationen har använts för att härleda en enkel ideal MOS-modell, Shockleys MOS-modell. Denna modell har sedan kompletterats steg för steg för att ta hänsyn även till andra ordningens effekter så att vi får en bra överensstämmelse mellan modell och mätdata.
#### 7.7 Övningsexempel

#### Gradvisa kanalapproximationen (teori)

- 1. Vad menas med gradvisa kanalapproximationen?
- 2. Inför en kanalpotential V(x) och använd sedan gradvisa kanalapproximationen för att steg för steg härleda Shockleys MOS-transistormodell genom att teckna ett uttryck för följande storheter som funktion av kanalpotentialen V(x):
  - a. kanalladdningen [per längdenhet]!
  - b. kanalladdningens drifthastighet!
  - c. strömmen genom kanalen!
  - d. Separera sedan variabler och integrera längs kanalen från kanalpotentialen V<sub>S</sub> vid source (x=0) till V<sub>D</sub> vid drain (x=L)!
- 3. Gör om härledningen i uppgift 2, men avbryt integrationen i en godtycklig punkt x < L i kanalen. Om man gör det får man uttrycket

$$I_{DS} \frac{x}{L} = k(V_{GST}V(x) - \frac{V(x)^2}{2})$$

- a. Lös ut V(x) ur detta uttryck och använd därvid mättnadsgraden  $\eta = I_{DS}/I_{DSAT}$ , där  $I_{DSAT} = kV_{GST}^2/2$ , som parameter.
- b. Förenkla sedan uttrycket för det fall att MOStransistorn är mättad, dvs η=1!
- 4. Använd det härledda utrycket för V(x) för att beräkna det elektriska fältet i en mättad MOS-transistor.
  - a. Hur stort är fältet enligt denna modell vid source respektive vid drain?
  - b. I vilken punkt är fältet lika med medelfältet  $E_{av}$ =- $V_{GST}/L$ ?
- 5. Använd laddningsmodellen i stark inversion,

$$Q_{inv} = -C_{ox} \left( V_{GST} - V(x) \right),$$

och det ovan härledda uttrycket för V(x) för att beräkna MOS-transistorns totala kanalladdning!

6. Eftersom drain är mer backspänd än source i en MOS-transistor är också utarmingsområdet mellan drain och substrat djupare än utarmingsområdet mellan source och substrat. Det betyder att utarmningsladdningen  $Q_B$  ökar med V(x) längs kanalen från source till drain. Låt oss därför modifiera  $Q_B$ -modellen så att den får ett V(x)-beroende. Ett enkelt linjärt samband (typ Taylorutveckling) enligt

$$Q_B = -C_{ox} \left( \gamma \sqrt{2\phi_B} + (\alpha - 1)V(x) \right),$$

kan då vara en enkel första ansats.

Använd denna modifierade laddningsmodell för att härleda en MOS-modell som tar hänsyn till denna andra ordningens substrateffekt.

7. [Överkurs] En annan andra ordningens effekt att ta hänsyn till i moderna MOS-transistorer är hastighetsmättnad, dvs att drifthastigheten "slår i taket" vid sitt mättnadsvärde  $v_{sat}$ . En vanlig hastighetsmodell som tar hänsyn till hastighetsmättnad ges av uttrycket

$$v = \frac{\mu \frac{dV}{dx}}{1 + \mu \frac{dV/dx}{v_{sat}}}.$$

ı

- a. Genomför härledningen av en ny modell för I<sub>DS</sub> med hjälp av gradvisa kanalapproximationen och denna hastighetsmodell! Använd gärna parametern V<sub>B</sub>=Lv<sub>sal</sub>/μ för att förenkla härledningen!
- b. Vilken är modellens mättnadsspänning  $V_{DSAT}$ ?
- c. Validera din modell genom att kontrollera om mättnadsspänningen  $V_{DSAT} \rightarrow V_{GST}$  när  $V_B \rightarrow \infty$ ?

#### Hastighetsmättnad och drain-induced barrier-lowering

 En styckevis linjär MOS-modell där mättnadsströmmen ges av hastighetsmättnad kan skrivas

$$I_{DS} = \begin{cases} kV_{GST}V_{DS} & \text{resistansområdet} \\ WC_{ax}(V_{GST} - V_{DS})v_{sat} & \text{mättnadsområdet}. \end{cases}$$

Vad har denna modell för brytpunkt mellan resistans- och mättnadsområde? Verkar brytpunkten rimlig med tanke på dess värde i fallet när hastighetsmättnaden blir försumbar ( $V_B >> V_{GST}$ )?

 För att få samma värde på mättnadsströmmen som den allmänt vedertagna Berkeleymodellen BSIM4 måste man modifiera den styckevis linjära modellen enligt följande:

$$I_{DS} = \begin{cases} kV_{GST}V_{DS} & \text{resistansområdet} \\ WC_{ax} \left(V_{GST} - 2V_{DS}\right)v_{sat} & \text{mättnadsområdet}. \end{cases}$$

Vad har denna modell för brytpunkt? Verkar brytpunkten rimlig?

- Teckna ekvationen för den parabel som "smoothar" ut övergången mellan resistansområdet och mättnadsområdet i uppgiften ovan! Vid vilken drainspänning mättar denna parabel jämfört med brytpunkten?
- 11. En långkanalig MOS-transistor har TILL-resistansen  $R_{ON}$ =50  $\Omega$  och levererar mättnadsströmmen 20 mA vid den effektiva styrspänningen  $V_{GST}$ =2 V. En kortkanalig MOS-transistor med samma TILLresistans levererar endast 10 mA mättnads-ström vid samma effektiva styrspänning. Vilken kanallängd har denna MOS-transistor? Använd följande värden:  $v_{sat}$ =10<sup>7</sup> cm/s och  $\mu$ =400 cm<sup>2</sup>/Vs!
- 12. Vad innebär drain-induced barrier-lowering (DIBL) och i vilken typ av MOS-transistorer uppträder denna effekt?
- 13. Illustrera DIBL-effekten med hjälp av banddiagram!
- 14. Teckna en första ordningens linjär modell för DIBLeffekten!



15. Para ihop nedanstående diagram med a) DIBL, b) hastighetsmättnad, och c) substrateffekten  $\alpha \neq 1!$ 

#### Subtröskelströmmar

- 16. Rita ett banddiagram för en MOS-transistor i subtröskelområdet och jämför detta banddiagram med motsvarande banddiagram för en MOS-transistor i stark inversion!
- 17. Vilket är sambandet mellan styrspänningen  $V_{GS}$  och ytpotentialen  $\Psi_S$  i utarmning/svag inversion? Hur kan vi linjarisera detta samband?
- 18. Kanalladdningen  $Q_S$  vid source i subtröskelområdet nära övergången till stark inversion,  $\Psi_S=2\phi_B$ , skrivs ofta  $Q_S=\delta C_{ox}V_t$ , där  $V_t$  är den termiska spänningen  $V_t=kT/q$  och där  $\delta$  är en dimensionslös passningsparameter.
  - c. Teckna ett uttryck för  $Q_s$  som funktion av ytpotentialen i subtröskelområdet!
  - d. Teckna motsvarande uttryck för drainladdningen  $Q_D!$
  - e. Gör om uttrycken för  $Q_S$  och  $Q_D$  till funktioner av  $V_{GST}$  mha det linjariserade sambandet mellan  $V_{GST}$  och ytpotentialen  $\Psi_S$  i uppgift 18.
- 19. Hur lyder Ficks första lag?
- 20. Utgå från laddningsutrycken i föregående uppgift och använd Ficks första lag för att teckna ett uttryck för subtröskelströmmen.

- a. Hur beror subtröskelströmmen på drainspänningen V<sub>DS</sub>?
- b. Hur beror subtröskelströmmen på styrspänningen?
- c. Hur stor är OFF-strömmen I<sub>OFF</sub> för styrspänningen V<sub>GS</sub>=0?
- d. Hur stort är subtröskelsvinget om idealitetsfaktorn *n*=1,5?
- 21. Hur bestämmer man passningsparametern  $\delta$  för kontinuitet i ström och transkonduktans vid övergången mellan subtröskelområdet och mättnadsområdet?
- 22. En MOS-transistor levererar TILL-strömmen  $I_{ON}$ =810 mA vid matningsspänningen 1,2 V.
  - a. Hur stor är FRÅN-strömmen I<sub>OFF</sub> om subtröskelsvinget är 80 mV/dekad? V<sub>T</sub>=0,3 V. V<sub>DS</sub>=V<sub>DD</sub>.
  - b. I vilken punkt är det lämpligt att skarva modellerna för stark och svag inversion?
  - 23. Bestäm  $I_{ON}/I_{OFF}$  och subtröskelsvinget S ur diagrammet nedan!



#### 8. MIKROELEKTRONIKHISTORIA

1. Vad heter de tre nobelprisbelönade uppfinnarna av transistorn?



- 2. Vilka var de företag som var tidigast ute med att försöka tjäna pengar på den nyuppfunna transistorn? Vilken var den produkt där transistorn först fick sitt stora genombrott?
- 3. Vad heter de två uppfinnare som är förknippade med nedanstående två bilder och vilka är de fundamentala skillnaderna mellan de två olika lösningarna på ett problem av epokgörande betydelse för teknikutvecklingen?



4. Nedanstående diagram över utvecklingen av integrerade kretsar och möjlig chipkomplexitet är ganska berömt och har ett speciellt namn. Vilket? Vilken är den aktuella ökningstakten under 70-talets första hälft (uttryckt som en fördubbling per X månader) och vilken ökningstakt har vi sett från 1975 och fram till idag?



5. Nämn några av de "åtta förrädarna" (*eng.* "the traitorous eight") som 1957 lämnade Shockley för att istället grunda Fairchild Semiconductor och där utveckla den integrerade kretsen.



6. Kisel i all ära, men innan kisel blev gemene mans favorithalvledare fanns andra halvledarmaterial i rampljuset. Nämn en annan halvledare som användes för komponenttillverkning innan kisel togs i bruk! En berömd demonstration som illustrerar ögonblicket när kisel slog ut detta halvledarmaterial visas nedan i ett foto från Smithsonian Institute, Washington D.C.



7. Nedanstående figur visar första tillämpningen av en då helt ny typ av integrerad krets och starten för ett stort halvledarföretags senare nästan totala världsdominans i branschen. Vad kallas kretsen som sitter på insidan av denna kalkylator och vad heter företaget som tillverkade den?



"This is the engineering prototype of the BUSICOM calculator, the first device in the world to use a <xxxx> circuit and the device for which the world's first <xxxx> was designed. Three <company name> engineers -- Federico Faggin, Ted Hoff, and Stan Mazor -- along with Masatoshi Shima of BUSICOM created the <xxxx> circuit

on-a-chip as a standard product alternative to a set of custom circuits. Hoff, with the help of Mazor, developed the architecture. Faggin was the project leader and with the help of Shima, designed and built the circuit." http://www.computerhistory.org

- 8. Vem anses vara uppfinnare av principen för logiska kretsar i CMOS-teknik, kretsar som drar mycket låg effekt? På vilket företag jobbade han vid det tillfället när arbetet gjordes och patentet söktes? Vem var hans chef och vad är denne chef mest känd för? Se bifogade artikel för mer information!
- 9. Nedanstående foton visar Robert Noyce och Gordon Moore. Vilken roll har de spelat i mikroelektronikens historia?



# Wanlass's CMOS circuit



June 18, 1963: Frank Waniass applies for a patent on complementary fieldeffect transistor circultry that reduces the standby power consumption of digital logic by six orders of magnitude

Most people remember the '60s as the era of the Beatles, the Vietnam War, Woodstock, and 35-cent-a-gallon gasoline. But electronics engineers of a certain age remember them as the days of free-wheeling experimentation when—at some companies, at least—bright young Ph.D.s were given latitude to see what they could create without interference from corporate managers.

One bright young Ph.D. of that time was Frank Wanlass, who this February received the 1991 IEEE Solid-State Circuits Award for his invention of complementary-MOS (CMOS) logic circuitry.

Wanlass's interest in MOS technology dates to 1962, when he was still studying for his doctorate in solid-state physics at the University of Utah in Salt Lake City. Upon reading about the Radio Corporation of America's work with thin-film cadmium sulfide field-effect transistors (FETs), he became intrigued by the simple structure of the devices, which he thought would make it easy to design fairly complex integrated circuits. But the devices were unstable. When left on a shelf for just a few hours, their electrical parameters changed dramatically.

Wanlass believed that making the FETs in silicon would solve the problem. After all, Fairchild Semiconductor Research and Development had made some very stable and reliable bipolar transistors in silicon using its planar process, so why not use the same material to make stable MOSFETs?

So he went to work for the Fairchild Semiconductor subsidiary of Fairchild Camera and Instrument Co. in Palo Alto, Calif., in August 1962, unaware that other researchers were already working on silicon MOS-FETs at both RCA and Bell Laboratories. A few months later, he made his first p-channel silicon MOSFETs, which, like all early MOSFETs, were a great disappointment. At 10–20 volts, their threshold voltages were very high and very unstable. They were no better than RCA's CdS devices.

Wanlass speculated that the aluminum gate electrode was diffusing into the gate oxide. If that were the case, the use of a more inert metal would solve the problem.

Michael J. Riezenman Contributing Editor



Wanlass's patent portrayed an integrated CMOS inverter.

While investigating that possibility, he found to his surprise that aluminum electrodes deposited by an electron beam evaporation machine yielded quite stable devices. The problem, he began to suspect, was not aluminum diffusion after all, but contamination.

The usual way to deposit aluminum in those days was to evaporate it by placing an aluminum wire in contact with a heated tungsten filament. Wanlass reasoned that the process was introducing positive ions into the aluminum and thence the gate oxide.

Further investigation inculpated sodium contamination from both the tungsten and the aluminum. Electron beam evaporation solved the problem because the electronbeam apparatus had a shutter mechanism that protected the silicon wafers from a carbon crucible of molten aluminum until the aluminum was at evaporation temperature. The sodium, having a much lower boiling point, boiled away before the shutter opened.

Wanlass next had the idea for CMOS. "It occurred to me," he told *IEEE Spectrum* in a recent interview, "that a complementary circuit of NMOS and PMOS devices, if it could be made, would use very little power. In standby, it would draw practically nothing—just the leakage current."

His boss Gordon Moore, now the chairman of Intel Corp., gave him a free hand to pursue his idea. At first, Wanlass tried to build a CMOS circuit monolithically, but that was so difficult he decided to prove the concept with discrete p-channel and n-channel MOSFETs instead. Because only p-channel devices were available, he had to start by building an n-channel silicon MOSFET.

The CMOS concept requires that both of its transistors be enhancement-mode devices. But whereas PMOS transistors were inherently of that type, the n-channel MOSFET was not. It would be years before MOS surface physics was well enough understood to permit the fabrication of such devices. Consequently, Wanlass made depletion-mode n-channel MOSFETs and back-biased their bodies negative with respect to their sources to turn them into enhancement-mode units.

The concept worked. The first demonstration circuit, a two-transistor inverter, consumed just a few nanowatts of standby power. Equivalent bipolar and PMOS gates consumed milliwatts of power even in standby. CMOS shrank the standby power consumption by six orders of magnitude!

The speed was impressive enough, too. Propagation delay times were on the order of 100 nanoseconds—about half the speed of bipolar, but almost an order of magnitude faster than PMOS.

On June 18, 1963, Frank Wanlass applied for patent protection for his CMOS concept, and in due course, was granted U.S. patent no. 3 356 858, the rights to which became part of Fairchild's patent portfolio. That patent described the overall concept and three specific circuits-an inverter, a NOR gate, and a NAND gate-from which any digital function can be built. In addition to the discrete implementations that were actually built, the patent includes the representation of an integrated CMOS inverter shown here. (Wanlass's laboratory notebooks, with his original drawings, were lost, probably when Fairchild Semiconductor was taken over by National Semiconductor Corp. in 1987.)

Neither Wanlass nor Fairchild Semiconductor grew rich from the invention. In those days companies traded the rights to their patent portfolios. Still, the integrated CMOS inverter shown, although never built, is the progenitor of all CMOS ICs today.

#### 9. APPENDIX 1

#### 9.1 Väteatomens tillåtna energinivåer

Den enklaste atomen att studera är väteatomen eftersom den endast har en elektron som snurrar runt en atomkärna. Elektronen är negativt laddad (-q) och atomkärnan är positivt laddad (+q). Elektronens potentiella energi som funktion av avståndet r från atomkärnan ges då av

$$U = -\frac{q^2}{4\pi\varepsilon r}.$$
(9.1)

Coulombkraften mellan elektronen och kärnan ges av

$$F = -\frac{dU}{dr} = \frac{q^2}{4\pi\varepsilon r^2}.$$
(9.2)

Denna kraft är utgör den centripetalkraft som håller elektronen snurrande i sin bana runt atomkärnan, dvs

$$F = \frac{mv^2}{r} \,. \tag{9.3}$$

Dessa två samband (9.2) och (9.3) ger elektronens rörelseenergi

$$\frac{mv^2}{2} = \frac{q^2}{8\pi\varepsilon r}.$$
(9.4)

Det ger att elektronens totala energi är hälften av den potentiella energin, dvs E=U/2,

$$E = \frac{mv^2}{2} + U = -\frac{q^2}{8\pi\varepsilon r}.$$
(9.5)

Ur detta samband framgår också att mot varje energinivå svarar en viss radie, Bohrradien, en radie som enligt lösningen till Schrödingerekvationen är den mest sannolika elektronradien.

Enligt den av de Broglie formulerade våg-partikeldualiteten är varje partikels rörelsemängd (impuls) associerad med en elektromagnetisk våglängd

$$\lambda = \frac{h}{p},\tag{9.6}$$

där *h* är Plancks konstant och p=mv är rörelsemängden (impulsen). Enligt de Broglie måste vidare varje banas omkrets vara ett helt antal våglängder

$$2\pi r = n\lambda, \qquad (9.7)$$

där n är ett heltal. Om vi nu eliminerar radien ur (9.5) med hjälp av (9.6) och (9.7) får vi

$$E = -\frac{q^2 p}{8\pi\varepsilon n\hbar},\tag{9.8}$$

där  $\hbar = h/2\pi$ . Eftersom

$$\frac{mv^2}{2} = \frac{p^2}{2m} = -E,$$
(9.9)

kan vi eliminera rörelsemängden och får därmed följande uttryck för de tillåtna energinivåerna i väteatomen

$$E = \frac{mv^2}{2} + U = -\frac{mq^4}{2(2nh\varepsilon)^2}.$$
 (9.10)

För n=1 får vi den lägsta tilltna energinivån E=-13,6 eV, dvs väteelektronens normalläge. Vätets utträdesarbete, dvs det är den energi som krävs för att jonisera väteatomen, är därmed 13,6 eV. För denna lägsta tillåtna energinivå får vi Bohrradien r=0,05 nm.

#### 9.2 Tillståndstätheten vs elektronenergin

I fig. Al visas impulsrymden och den volym som ryms mellan två skal med radierna p och p+dp. Denna volym ges av

$$volym = 4\pi p^2 dp. \tag{9.11}$$

Om varje tillstånd i en halvledarkub med sidan L upptar volymen  $h^3/2L^3$  får vi följande tillståndstäthet

$$dn(p) = \frac{2L^3}{h^3} 4\pi p^2 dp.$$
(9.12)

Om vi sedan går över från rörelsemängd till energi enligt sambandet för rörelseenergi

$$E - E_c = \frac{p^2}{2m_e^*},$$
 (9.13)

får vi följande uttryck för tillståndstätheten i ledningsbandet

$$N_{C}(E) = \frac{4\pi}{h^{3}} \left(2m_{e}^{*}\right)^{3/2} \sqrt{E - E_{C}} . \qquad (9.14)$$



Fig. 10.1. Volymen i impulsrymden mellan p och p+dp.

1



Gustaf Robert Kirchhoff 1824-1887



Leonhard Euler 1707-1783



Pierre Simon de Laplace 1749-1827

#### 10. APPENDIX 2: GRUNDLÄGGANDE ELEKTROTEKNIK

I detta kapitel ska vi repetera några viktiga grundläggande elektrotekniska samband.

#### 10.1 RC-kretsen

I figuren visas en enkel RC-krets.



Fig. 10.1. RC-krets.

Med hjälp av Kirchhoffs nodekvation kan man teckna den första ordningens linjära differentialekvation som bestämmer villkoren för utspänningen V över kondensatorn C. Ekvationen kan skrivas

$$C\frac{dV}{dt} + \frac{V}{R} = V_0. \tag{10.1}$$

Lösningen till denna differentialekvation är en funktion som är lika med sin egen derivata, nämligen exponentialfunktionen. Talet e utgör basen för den naturliga logaritmen, ln. Det fick sin nuvarande beteckning av Leonhard Euler och kallas efter honom ibland *Eulers tal*, och är ungefär lika med 2,718. Talet kan definieras som gränsvärdet

$$e = \lim_{n \to \infty} \left( 1 + \frac{1}{n} \right)^n$$
, eller genom serien  $e = \sum_{n=0}^{\infty} \frac{1}{n!}$ , där  $n! = 1 \cdot 2 \cdot 3 \cdot ... \cdot n$ .

Talet *e* är ett irrationellt tal. Adderar man några av de första termerna i serien ovan får man en hygglig approximation [Wikipedia], typ

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} \approx 2,717$$
(10.2)

Man kan också teckna spänningen över kondensatorn med hjälp av Laplace. Utspänningen blir då

$$V(s) = \frac{V_0(s)}{1 + sRC},$$
(10.3)

där V(s) och V<sub>0</sub>(s) är Laplacetransformerna av in- och utspänningarna. Lösningen för uppladdningsförloppet från begynnelsevärdet V=0 till slutvärdet  $V_0=V_{DD}$  blir

$$V(t) = V_{DD} \left( 1 - e^{-t/RC} \right), \tag{10.4}$$

medan urladdningsförloppet från  $V(0)=V_{DD}$  till slutvärdet  $V_0=0$  beskrivs av

$$V(t) = V_0 e^{-t/RC}.$$
 (10.5)

Den tid det tar för utspänningen att nå halva nivån,  $V_{DD}/2$ , ges av  $t_d=RC\ln 2=0,7RC$ , medan stig- och falltiderna mellan 10 och 90% nivåerna ges av  $RC\ln 9\approx 2,2RC$ .

Om vi däremot låter insignalen vara en sinussignal med vinkelfrekvensen  $\omega = 2\pi f$ ,

$$V_0 = V_1 \sin \omega t, \tag{10.6}$$

så blir utsignalen en dämpad och fördröjd sinussignal som ges av

$$V(t) = \frac{V_1 \sin\left(\omega t + \phi\right)}{\sqrt{\left(1 + \omega^2 R^2 C^2\right)}},$$
(10.7)

där fasvinkeln  $\phi$  ges av  $\phi$ =-arctan( $\omega RC$ ).

RC-kretsar, som den i fig. 11.2, med två tidskonstanter är lite mer komplicerade att hantera, men de resulterar förstås i andra ordningens linjära differentialekvationer.



Fig. 10.2 RC-krets med två RC-tidskonstanter.

Lösningen kan skrivas

$$V(t) = V_{21}e^{-t/t_1} + V_{22}e^{-t/t_2},$$
(10.8)

där  $t_1$  och  $t_2$  är de två tidskonstanterna och där konstanterna  $V_{21}$  och  $V_{22}$  bestäms av nodspänningarnas begynnelsevärden. Ofta är den ena tidskonstanten dominerande, typ  $t_1 >> t_2$ , och då ges tidskonstanterna av uttrycken

$$t_1 = (R_1 + R_2)C_1 + R_2C_2$$
, respective  $t_2 = R_1R_2C_1C_2/t_1$ . (10.9)

W.C. Elmore generaliserade detta resonemang att även omfatta RC-nät med n steg (n > 2) för vilka det inte finns analytiska lösningar. Han antog att en tidskonstant var dominerande och att denna tidskonstant kunde approximeras med

$$t_{1} = (R_{1} + R_{2} + \dots + R_{n-1} + R_{n})C_{1} + (R_{1} + R_{2} + \dots + R_{n-1})C_{2} + \dots + R_{n}C_{n}.$$
(10.10)

Detta är Elmores formel som är mycket användbar när man ska beräkna fördröjningar i RC-ledningar.

#### 10.2 Thevenin och Nortons teorem

Thevenins teorem för linjära elektriska kretsar med två anslutningar säger att varje kombination av spänningskällor, strömkällor och resistorer är ekvivalent med *en* spänningskälla i serie med *en* källresistans (som då kan betraktas som spänningskällans inre resistans). Detta teorem upptäcktes först av den tyske vetenskapsmannen Hermann von Helmholtz år 1853, och återupptäcktes 1883 av den franske telegrafingenjören Léon Charles Thévenin (1857–1926). [Wikipedia]



Fig. 10.3. Thevenins teorem.



Hermann von Helmholtz 1821-1894



Léon Charles Thévenin 1857-1926



Edward Lawry Norton 1898–1983



Carl Friedrich Gauss 1777–1855



James Clerk Maxwell 1831–1879

På samma sätt säger Nortons teorem att varje kombination av spänningskällor, strömkällor och resistorer är ekvivalent med *en* strömskälla i parallell med *en* källresistans (som då kan betraktas som strömkällans inre resistans). Nortons teorem är en utvidgning av Thevenins teorem och introducerades 1926 samtidigt av den tyske fysikern Hans Ferdinand Mayer (1895–1980) och den amerikanske ingenjören Edward Lawry Norton (1898–1983).

black box



Fig. 10.4. Nortons teorem.

#### 10.3 Gauss lag lag och Maxwells första ekvation

Inom elektrostaiken åberopas ofta Gauss lag, en lag som också är en av Maxwells ekvationer. I **differentiell form** kan den skrivas

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon},\tag{10.11}$$

där  $\nabla \cdot \mathbf{E}$  är divergensen av det elektriska fältet och där  $\rho$  är den elektriska laddningstätheten. Gauss sats kan också skrivas på **integrerad form** enligt

$$\Phi_E = \oint_S \mathbf{E} \cdot d\mathbf{A} = \frac{Q_S}{\epsilon_0}$$

där  $\Phi_E$  är det elektriska flödet,  $Q_S$  är nettoladdningen innesluten av ytan S och  $\varepsilon_0$  är den dielektriska konstanten. Gauss sats är också en av Maxwells ekvationer.



Fig. 10.5 Gauss lag.

Låt oss nu reducera Gauss lag till en dimension i syfte att senare kunna analysera plattkondensatorer, pn-övergångar och MOS-transistorer. Eftersom divergensen av det elektriska fältet avser

$$\nabla \cdot \mathbf{E} = \frac{\partial \mathbf{E}_{\mathbf{x}}}{\partial \mathbf{x}} + \frac{\partial \mathbf{E}_{\mathbf{y}}}{\partial \mathbf{y}} + \frac{\partial \mathbf{E}_{\mathbf{z}}}{\partial \mathbf{z}},$$
(10.12)

kan problemet reduceras till en dimension om fältet är konstant i y- och z-riktningarna.

$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x}.$$
(10.13)

Ett tredimensionellt problem kan således under dessa förutsättningar förenklas till ett endimensionellt problem. Denna förenkling används exempelvis i den gradvisa kanalapproximationen (*eng.* gradual channel approximation) vid härledning av MOS-transistorns strömmodell.

I integrerad form får vi då att den laddning [per ytenhet] som är innesluten mellan två parallella ytor ges av



Fig. 10.6. Gauss lag i en dimension.

Med det elektriska fältet **E** i en punkt (x,y,z) avses den negativa gradienten av den elektriska potentialen V(x,y,z) i punkten, dvs

$$\mathbf{E} = -\nabla \mathbf{V},\tag{10.15}$$

där

$$\nabla V = \left(-\frac{dV}{dx}, -\frac{dV}{dy}, -\frac{dV}{dz}\right).$$
(10.16)

För endimensionella fält mellan två parallella elektroder förenklas sambandet till

$$E = -\frac{dV}{dx}.$$
(10.17)

Sambandet mellan fält och potential för ett enkelt endimensionellt fall med ett konstant (positivt) elektriskt fält visas i fig. 11.6. Om det elektriska fältet är konstant så är fältet helt enkelt lika med pålagd spänning,  $V=V_1-V_2$ , dividerad med avståndet  $d=x_2-x_1$  mellan ytorna,

$$E = \frac{V_1 - V_2}{L}.$$
 (10.18)

I en plattkondensator där spänningen mellan plattorna ges av  $V=V_1-V_2$  får vi således genom våra två uttryck för det elektriska fältet

$$\frac{Q_s}{\varepsilon} = \frac{V_1 - V_2}{d} \rightarrow Q_s = \frac{\varepsilon}{d} V \rightarrow C = \frac{\varepsilon}{d}.$$
 (10.19)

$$\begin{array}{c|c}
E=0 \\
\hline V_1 \\
\hline \end{array} \\
\hline \end{array} \\
\begin{array}{c}
E=Q_{S} \\
\hline \\ V_2 \\
\hline \end{array} \\
\hline \end{array}$$

Fig. 10.6. Gauss lag tillämpad på plattkondensatorn.



Adolf Fick 1829–1901

Antoine Lavoisier (1743–1794)

#### 10.4 Ficks lagar

Ficks lagar beskriver fenomenet diffusion och utvecklades av Adolf Fick 1855.

Ficks första lag relaterar flödet av diffunderande partiklar till motsvarande koncentrationsgradient i stationärtillstånd. Det postulerar hur flödet går från områden med hög koncentration till områden med låg koncentration med en storlek som är proportionell mot koncentrationsgradienten. I en dimension lyder Ficks första lag

$$J = -D\frac{d\phi}{dx} . (10.20)$$

där *J* är flödet [antal påer volymsenhet], *D* är diffusionskonstanten [area/tidsenhet,  $m^2/s$ ], och  $\phi$  är koncentrationen [antal per volymsenhet]. I flera dimensioner skrivs Ficks första lag på formen

$$J = -D\nabla\phi \ . \tag{10.21}$$

Det är väl känt att gaser sprider sig genom diffusion, men i halvledare där stora koncentrationsgradienter förekommer på grund av dopning, kan även hål och elekt-roner diffundera.

Eftersom flödet av laddade partiklar utgör ström kan vi för hål och elektroner skriva Ficks lag påformen,

$$J_p = -qD_p \frac{dp}{dx}$$
 respective  $J_n = qD_n \frac{dn}{dx}$ . (10.22)

Ficks andra lag beskriver hur diffusion får koncentrationen  $\phi$  att ändra sig med tiden,

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2} . \tag{10.23}$$

Den kan härledas från Ficks första lag och lagen om massans bevarande (i frånvaro av kemiska reaktioner)

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x}J = 0 \implies \frac{\partial \phi}{\partial t} - \frac{\partial}{\partial x}\left(D\frac{\partial \phi}{\partial x}\right) = 0.$$
(10.24)

Lagen om massans bevarande beskrevs första gången entydigt av Antoine Lavoisier. Under förutsättning att diffusionskonstanten är konstant kan deriveringsordningen ändras

$$\frac{\partial}{\partial x} \left( D \frac{\partial \phi}{\partial x} \right) = D \frac{\partial^2 \phi}{\partial x^2} , \qquad (10.25)$$

vilket leder till Ficks andra lag som den formulerats ovan. I flera dimensioner skrivs Ficks andra lag på formen

$$\frac{\partial \phi}{\partial t} = -D\nabla^2 \phi , \qquad (10.26)$$

vilket är ekvivalent med värmeledningsekvationen där  $\phi$  då representerar temperaturen. Den är också ekvivalent med telegrafekvationen som för en distribuerad RCledning kan skrivas

$$\frac{\partial \phi}{\partial t} = -\frac{1}{RC} \nabla^2 \phi , \qquad (10.27)$$

där R och C är ledningsresistansen respektive ledningskapacitansen per längdenhet

#### 10.5 Diffusionsekvationen för majoritets- och minoritetsbärare

Ficks andra lag är mycket intressant för halvledare. Till exempel kan den användas för att beskriva subtröskelströmmen i en MOS-transistor, en ström som utgörs av diffunderande laddningsbärare i kanalen mellan source och drain. Den beskriver då hur antalet rörliga laddningsbärare i stationärtillstånd avtar linjärt från värdet  $Q_S$  vid source till  $Q_D$  vid drain, något som ger upphov till den konstanta strömmen

$$I_{DS} = -WD \frac{Q_S - Q_D}{L} = -\frac{W}{L} DQ_S \left(1 - \frac{Q_D}{Q_S}\right). \quad (10.28)$$

Ficks andra lag är också intressant när man studerar balansen mellan drift och diffusion vid potentialbarrären i en pn-övergång. Men när det gäller minoritetsbärare i en halvledare måste man lägga till en källa g för att beskriva den termiska alstringen (*eng.* generation) av hål-elektronpar och en sänka *r* som beskriver rekombinationstakten, dvs den takt med vilken ledningselektroner faller tillbaka ner i valensbandet när den hittar en ledig plats (minoritetsbärare).

Ficks andra lag kan då skrivas

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2} + g - r . \qquad (10.29)$$

Rekombinationstakten begränsas i första hand av tillgången på minoritetsbärare varför man i en första ordningens modell antar att

$$r = \frac{\phi}{\tau} \quad , \tag{10.30}$$

där tidskonstanten  $\tau$  beskriver minoritetsbärarnas livslängd. Dynamisk jämvikt mellan alstring och rekombination i stationärtillstånd ger då följande diffusionsekvation för minoritetsbärare

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2} - \frac{\phi - \phi_0}{\tau} . \qquad (10.31)$$

I stationärtillstånd ser vi en annan variant på andra ordningens linjära differentialekvation

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{\phi - \phi_0}{D\tau} = 0 , \qquad (10.32)$$

med lösningen

$$\phi - \phi_0 = A e^{x/L} + B e^{-x/L} , \qquad (10.33)$$

där  $L=\sqrt{D\tau}$  är den så kallade diffusionslängden. I motsats till tidigare andra ordningens differentialekvation för *RC*-kretsar med två dämpande tidskonstanter ser vi här hur den spatiala lösningen i rummet innehåller två termer, en som tyder på utbredning i x-axelns riktning och en som tyder på rörelse i motsatt riktning. I fallet med utbredning enbart i ena riktningen (B=0) ser vi ett exponetiellt avklingande minoritetsbäraröverskott

$$\phi - \phi_0 = (\phi(0) - \phi_0) e^{-x/L} , \qquad (10.34)$$

där  $\phi(0)$  är randvärdet vid injektionspunkten x=0.

#### 10.6 Kontinuitetsekvationen

Ficks andra lag är också ett specialfall av kontinuitetsekvationen, en ekvation som beskriver laddningens bevarande:

$$\frac{\partial \phi}{\partial t} + \frac{1}{\pm q} \frac{\partial J}{\partial x} = -\frac{\phi - \phi_0}{\tau}, \qquad (10.35)$$

där  $\pm q$  är hålens respektive elektronens laddning. I integrerad form kan kontinuitetsekvationen användas för att beskriva laddningens bevarande

$$\frac{dQ}{dt} + \frac{Q}{\tau} = I, \qquad (10.36)$$

vilket är en utmärkt modell för att beskriva transienta förlopp i framspända dioder. I en framspänd diod fylls dioden med minoritetsbärarladdning Q tills stationärtillstånd inträffar, och strömmen  $I_F$  in i dioden exakt motsvarar minoritetsbärarnas rekombinationstakt  $Q/\tau$ . En negativ ström  $-I_R$  tömmer sedan dioden på minoritetsbärare enligt ett förlopp som beskrivs av

$$Q = \tau \left( I_F + I_R \right) e^{-t/\tau} - \tau I_R, \qquad (10.37)$$

vilket innebär att det tar tiden

$$t = \tau \ln\left(1 + \frac{I_F}{I_R}\right),\tag{10.38}$$

att helt tömma dioden på minoritetsbärare.



Fig. 10.7. Diodens omslagsförlopp. (Mätdata= symbol, teori=heldragen linje).

Utan källa och sänka beskriver kontinuitetsekvationen i integrerad form förskjutningsströmmen, dvs omladdningsströmmen till en kondensator med kapacitansen C=dQ/dV,

$$I = \frac{dQ}{dt} = C \frac{dV}{dt}.$$
(10.39)

#### 11. INDEX

# Α

acceptor, 4 -koncentration,  $N_A$ , 4

# В

Boltzmannfördelning, 12

# С

CMOS-kretsar, 42-43

# D

degenererad halvledare, 12 diod genombrott genombrottsspänning, 25 lavin-, 26 zener-, 25 ideala diodekvationen temperaturberoende, 30 donator, 4 -koncentration, N<sub>D</sub>, 4 dopning, 4 drifthastighet, 2

# Ε

Earlyspänning MOS-transistor, 63 effektiv tillståndstäthet, N<sub>C</sub>, 13 egentäthet, n<sub>i</sub>, 4, 14 elektriskt fält, 2 *extrinsisk*, 4

# F

farttriangeln, 2 Fermi -nivå, 11 -potential, φ<sub>F</sub>, 15 temperaturberoende, 17 Fermi-Dirac statistik, 6, 10–14 fördelningsfunktion Boltzmanns, 12 Fermi-Diracs, 12, 10–14

#### G

gittermobilitet. *Se gitterrörlighet* gitterrörlighet, 6

Н

hastighetsmättnad, 8, 60 styckevis linjär modell, 60 hastighetsmodell, 2

# I

III-V-halvledare, 3 intrinsisk, 4 isolator, 3

# Κ

kollisionsfrekvens, 6 kontaktpotential, 22 kristallstruktur, 3

# L

laddningsneutralitet, 5 ledare, 3 ledningselektroner, 4 ledningsförmåga, 3, 7 temperaturberoende, 17 löptid, 2

#### Μ

majoritetsbärare, 4 massverkans lag, 4, 14 Mathiessens regel, 6, 60 Maxwell-Boltzmanns fördelningsfunktion, 12 minoritetsbärare, 4 mobilitet. Se rörlighet MOS-kapacitans, 46–53 MOS-kretsar, 38–43 MOS-transistor andra ordningens effekter, 59-64 drain-induced barrier-lowering, 60 hastighetsmättnad, 60 kanallängdsmodulation Early-spänning, 63 mobility roll-off, 60 substrateffekten, 59 subtröskelström, 61 banddiagram, 57 stark inversion, 56 svag inversion, 56 första ordningens modell, 35-38 inversion

stark, 60 mättnadsspänning, 36 vid hastighetsmättnad, 60 vid pinch-off med substrateffekt, 59 överföringskarakteristik, 37 styckevis linjär modell, 38, 61 subtröskelström, 61 subtröskelström, 62 transkonduktans, 40 utgångskarakteristik, 37

### Ν

n-typ, 4

## Ρ

Plancks konstant, 12

# R

resistivitet, 7 rörlighet, 2 temperaturberoende, 17

# S

skiktresistivitet, 7 smoothing curve, 5, 38, 43, 61 störatom, 4 störmobilitet. *Se störrörlighet* störrörlighet, 7 ström, 2 strömmodell, 2 strömtäthet, 3 styckevis linjär modell, 5–6 drifthastighet vs *E*, 8

# Т

termisk energi, 6 termisk hastighet, 6

# U

utträdesarbete, 15

# Ζ

Zenereffekt, 25