

Elektronické součástky

Autoři textu:

Prof. Ing. Jaromír Brzobohatý, CSc.

Prof. Ing. Vladislav Musil, CSc.

Doc. Ing. Arnošt Bajer, CSc.

Ing. Jaroslav Boušek, CSc.

Obsah (Styl Nadpis seznamu)

1	ZÁKLADY FYZIKY POLOVODIČŮ	5
1.1	ZÁKLADY KVANTOVÉ TEORIE PEVNÝCH LÁTEK, PÁSOVÝ MODEL PEVNÝCH LÁTEK	5
1.1.1	<i>Elektron v izolovaném atomu</i>	5
1.1.2	<i>Elektron v krystalu - základy pásové teorie.....</i>	6
1.2	ROZDĚLENÍ LÁTEK PODLE JEJICH ELEKTRICKÝCH VLASTNOSTÍ	7
1.3	ELEKTRICKÁ VODIVOST POLOVODIČŮ	9
1.3.1	<i>Vlastní polovodiče</i>	9
1.3.2	<i>Nevlastní polovodiče</i>	10
1.4	ELEKTRONY V PEVNÝCH LÁTKÁCH A JEJICH ENERGETICKÉ ROZDĚLENÍ.....	13
1.4.1	<i>Hustota stavů.....</i>	13
1.4.2	<i>Fermi - Diracova rozdělovací funkce.....</i>	14
1.4.3	<i>Rovnovážné rozložení nosičů.....</i>	15
1.5	ROVNOVÁŽNÁ KONCENTRACE NOSIČŮ	16
1.5.1	<i>Vztahy pro výpočet koncentrace elektronů a děr</i>	16
1.5.2	<i>Nábojová neutralita.....</i>	17
1.5.3	<i>Výpočet polohy Fermiho hladiny</i>	19
1.6	VEDENÍ PROUDU V POLOVODIČÍCH	20
1.6.1	<i>Drift nosičů náboje.....</i>	21
1.6.2	<i>Difúze nosičů náboje</i>	23
1.6.3	<i>Generace a rekombinace.....</i>	25
1.6.4	<i>Stavové rovnice polovodiče</i>	27
2	POLOVODIČOVÉ PŘECHODY	28
2.1	KLASIFIKACE PŘECHODŮ	28
2.1.1	<i>Homogenní přechody</i>	28
2.1.2	<i>Heterogenní přechody.....</i>	29
2.2	PŘECHOD PN V ROVNOVÁŽNÉM STAVU	29
2.2.1	<i>Kvalitativní popis PN přechodu v rovnovážném stavu.....</i>	30
2.2.2	<i>Kvantitativní popis PN přechodu v rovnovážném stavu.....</i>	31
2.3	PŘECHOD PN S PŘILOŽENÝM VNĚJŠÍM NAPĚTÍM	35
2.3.1	<i>Propustně a závěrně pólovaný PN přechod</i>	35
2.3.2	<i>Ampérovoltová charakteristika ideálního PN přechodu.....</i>	36
2.3.3	<i>Kapacita PN přechodu</i>	41
2.4	PŘECHOD PN V DYNAMICKÉM REŽIMU	43
2.4.1	<i>Přechod PN v impulsovém režimu</i>	43
2.4.2	<i>Přechod PN při sinusovém napětí</i>	45
2.4.3	<i>Přechod PN s přiloženým stejnosměrným a střídavým napětím</i>	46
2.5	PRŮRAZ PŘECHODU PN	46
2.5.1	<i>Tunelový (Zenerův) průraz</i>	46
2.5.2	<i>Lavinový průraz.....</i>	48
2.5.3	<i>Tepelný průraz.....</i>	49
2.6	PŘECHOD KOV-POLOVODIČ	51
2.6.1	<i>Schottkyho bariéra.....</i>	51
2.6.2	<i>Usměrňující kontakt kov-polovodič.....</i>	52
2.6.3	<i>Neusměrňující kontakt kov-polovodič</i>	53
3	POLOVODIČOVÉ DIODY	54
3.1	HROTOVÉ DIODY	57

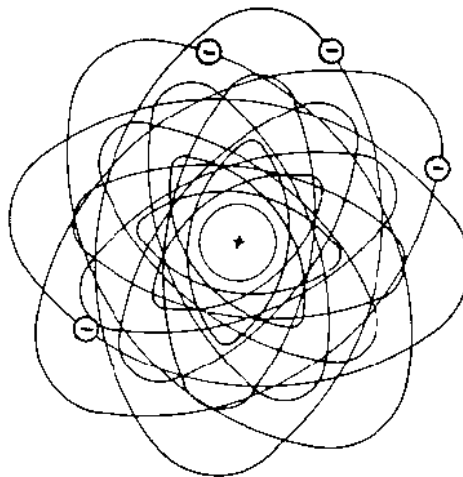
3.2	PLOŠNÉ DIODY	59
3.2.1	<i>Detekční a spínací diody</i>	63
3.2.2	<i>Stabilizační a referenční diody</i>	64
3.2.3	<i>Kapacitní diody</i>	68
3.2.4	<i>Tunelová a inverzní dioda</i>	69
3.2.5	<i>Usměrňovací diody</i>	69
3.2.6	<i>Rychlé usměrňovací diody</i>	74
3.2.7	<i>Lavinové usměrňovací diody</i>	76
3.2.8	<i>Schottkyho usměrňovací diody</i>	76
3.3	OSTATNÍ PRVKY DIODOVÉHO CHARAKTERU	77
3.4	NĚKTERÉ OBVODOVÉ APLIKACE DIOD	77
3.5	VLASTNOSTI A MODEL Y POLOVODIČOVÝCH DIOD	78
4	POLOVODIČOVÉ SOUČÁSTKY BEZ AKTIVNÍHO PŘECHODU PN	84
4.1	SOUČÁSTKY Z MONOKRYSTALICKÝCH POLOVODIČŮ BEZ AKTIVNÍHO PŘECHODU PN 84	
4.2	SOUČÁSTKY Z POLYKRISTALICKÝCH POLOVODIČŮ	84
5	BIPOLÁRNÍ TRANZISTORY	85
5.1	PRINCIP ČINNOSTI BIPOLÁRNÍHO TRANZISTORU	86
5.2	KVANTITATIVNÍ ANALÝZA FUNKCE BIPOLÁRNÍHO TRANZISTORU	94
5.2.1	<i>Proudové zesílení a vztahy mezi stejnosměrnými proudy tranzist. v základních zapojeních</i>	94
5.2.2	<i>Ebersův-Mollův model bipolárního tranzistoru NPN</i>	98
5.3	STATICKÉ VLASTNOSTI BIPOLÁRNÍHO TRANZISTORU	100
5.3.1	<i>Statické charakteristiky tranzistoru NPN v zapojení se společnou bází (SB)</i> 102	
5.3.2	<i>Statické charakteristiky tranzistoru NPN v zapojení se společným emitemorem (SE)</i>	103
5.3.3	<i>Průrazy tranzistorové struktury</i>	107
5.3.4	<i>Earlyho jev</i>	109
5.4	MODEL Y BIPOLÁRNÍHO TRANZISTORU	109
5.4.1	<i>Nelineární modely bipolárního tranzistoru</i>	109
5.4.2	<i>Linearizované modely bipolárního tranzistoru (bipolární tranzistor jako impedanční čtyřpól)</i>	114
5.4.3	<i>Mezní kmitočty bipolárního tranzistoru</i>	118
5.5	PRINCIPY OBVODOVÝCH APLIKACÍ BIPOLÁRNÍHO TRANZISTORU	120
5.5.1	<i>Bipolární tranzistor jako spínač</i>	120
5.5.2	<i>Bipolární tranzistor jako zesilovač</i>	123
5.6	VÝKONOVÉ A VYSOKOFREKVENČNÍ TRANZISTORY	131
6	UNIPOLÁRNÍ TRANZISTORY	134
6.1	UNIPOLÁRNÍ TRANZISTOR S PŘECHODEM PN (TRANZISTORY JFET)	136
6.1.1	<i>Kvalitativní popis činnosti tranzistoru JFET</i>	136
6.1.2	<i>Kvantitativní analýza tranzistoru JFET</i>	138
6.1.3	<i>Statické charakteristiky tranzistoru JFET</i>	140
6.1.4	<i>Linearizovaný malosignálový model tranzistoru JFET</i>	143
6.2	UNIPOLÁRNÍ TRANZISTOR S IZOLOVANOU ŘÍDICÍ ELEKTRODOU	146

6.2.1	<i>Ideální struktura MOS</i>	146
6.2.2	<i>Kvalitativní popis činnosti tranzistoru MOS</i>	150
6.2.3	<i>Statické charakteristiky tranzistoru MOS</i>	154
6.2.4	<i>Linearizovaný malosignálový model tranzistoru MOS</i>	155
6.2.5	<i>Nastavení pracovního bodu tranzistoru MOS</i>	155
6.2.6	<i>Setrvačné vlastnosti tranzistorů MOS</i>	158
6.2.7	<i>Model tranzistoru MOS v programu SPICE</i>	160
6.3	SPECIALIZOVANÉ TYPY UNIPOLÁRNÍCH TRANZISTORŮ	161
6.3.1	<i>Tranzistor MOS se dvěma hradly</i>	161
6.3.2	<i>Tranzistory MESFET a HEMT</i>	161
6.3.3	<i>Výkonové unipolární tranzistory</i>	163
6.3.4	<i>Výkonové tranzistory JFET (SIT)</i>	163
6.3.5	<i>Výkonové tranzistory MOS</i>	163
6.3.6	<i>Tranzistory IGBT</i>	166
6.4	STRUKTURY CCD	167
6.4.1	<i>Využití struktur CCD</i>	175

1 ZÁKLADY FYZIKY POLOVODIČŮ

1.1 Základy kvantové teorie pevných látek, pásový model pevných látek

Dříve než přistoupíme ke studiu vlastností elektronů v pevných látkách, musíme se seznámit s určitými základními pravidly kvantové teorie nejprve na nejjednodušším ze všech atomových systémů, na izolovaném vodíkovém atomu, potom se budeme zabývat základy pásové teorie.



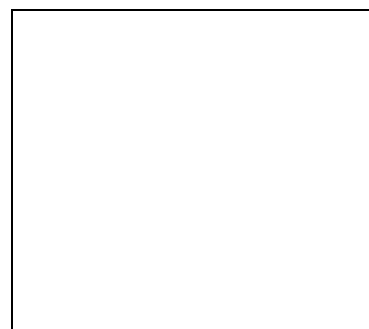
1.1.1 Elektron v izolovaném atomu

V roce 1913 stanovil Niels Bohr tři postuláty, vysvětlující některé do té doby nepochopitelné vlastnosti elektronů:

1. Elektron může trvale kroužit kolem jádra jen v některé z určitých drah (kvantových drah, orbitů).
2. Pokud elektron obíhá v některé z kvantových drah, atom nevyzařuje žádnou energii, jeho energie je konstantní.
3. Při přechodu elektronu z jedné kvantové dráhy na druhou atom energii buď absorbuje nebo emituje.

Tato "kvantizace" drah pohybu elektronu vede ke *kvantizaci energie* elektronu. Dá se odvodit [1], že pro elektron s hlavním kvantovým číslem n je tato energie:

Obr. 1.1: Jádro a elektronový obal.



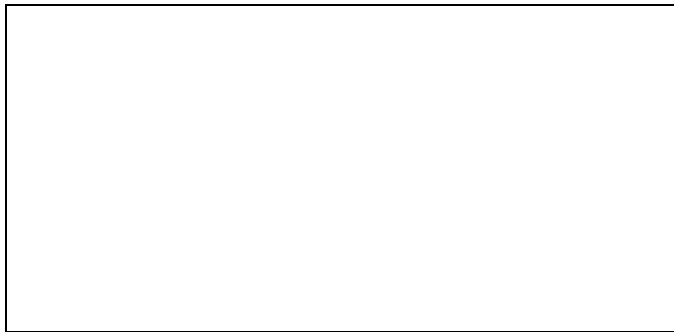
Obr. 1.2: Vodíkový atom - první tři dovolené kvantové dráhy a jejich energie

$$E = -\frac{m_0 q^4}{2(4\pi\epsilon_0 h n)^2} = -\frac{13,6}{n^2} \text{ eV}, \quad n = 1, 2, 3, \quad (1.1)$$

kde E je vazebná energie elektronu v atomu vodíku, m_0 je hmotnost volného elektronu ($9,1 \cdot 10^{-31}$ kg), q je absolutní hodnota náboje elektronu ($1,60210^{-19}$ C), ϵ_0 je permitivita vakua ($8,854 \cdot 10^{-12}$ Fm⁻¹), h je Planckova konstanta ($6,626 \cdot 10^{-34}$ Js), $\hbar = h/2\pi$ a n je hlavní kvantové číslo (identifikátor orbitu). Jednotka *elektronvolt* (eV) je jednotka energie ($1 \text{ eV} = 1,6 \cdot 10^{-19}$ J). Vazebná energie elektronu je záporná, vazebnou energii v nekonečno (tj. energii volného elektronu ve vakuu) klademe rovnu nule. Energie elektronu je tedy uvnitř vodíkového atomu omezena pouze na určité hodnoty (viz **Obr. 1.2.**) a z Bohrova modelu vyplývá, že přechod elektronu z orbitu s vyšším kvantovým číslem n na orbit s nižším kvantovým číslem n je provázen vyzařením energie o hodnotě rovné rozdílu energií těchto dvou orbitů. Tato energie

je zpravidla světelná ($\Delta E = h\nu$) a proto můžeme pozorovat emitované světlo pouze o diskretních vlnových délkách.

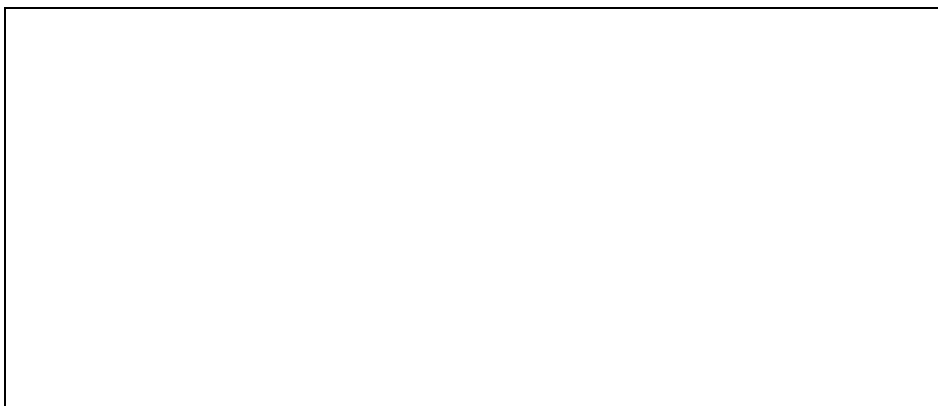
Podobně jako vodíkový atom můžeme znázornit i komplexnější 14 elektronový atom křemíku (**Obr. 1.3**). Zde 10 ze 14 elektronů obsazují energetické úrovně o velmi nízké energii a jsou velmi těsně svázány s jádrem atomu. Tato vazebná energie je tak silná, že během chemických reakcí zůstává prakticky neporušena. Tyto elektrony nazýváme *vnitřní elektrony*. Zbývající 4 elektrony jsou vázány relativně slabě, účastní se chemických reakcí a jsou nazývány *valenčními elektrony*.



Obr. 1.3: Izolovaný atom křemíku

1.1.2 Elektron v krystalu - základy pásové teorie

Pro izolovaný atom tedy můžeme uvažovat čárový energetický diagram sestávající z diskretních energetických hladin. Při přibližování dvou atomů dochází k projevům přitažlivých (vazebních, atraktivních) a odpudivých (repulsivních) sil. Odpudivé síly převládají při malých vzdálenostech mezi atomy a přitažlivé ve větších vzdálenostech. V určité vzdálenosti a_0 , kterou nazýváme rovnovážná *mřížková konstanta*, jsou tyto síly v rovnováze, soustava atomů má minimum potenciální energie a je proto ve stabilním stavu. Při vzájemném působení dvou atomů dochází také ke *štěpení energetických hladin*, které byly původně v obou atomech na stejné úrovni (na tento jev se můžeme dívat jako na rozšíření Pauliho vylučovacího principu na více atomů nebo na celý krystal).

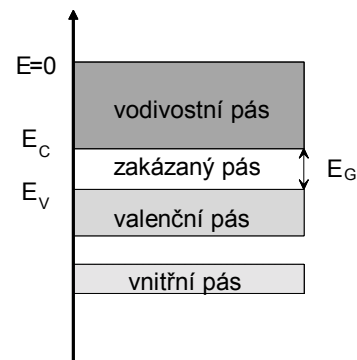


Obr. 1.4: Teoretický energetický diagram prvků IV. skupiny v závislosti na mřížkové konstantě. Pásový model napravo vznikne „řezem“ v místě příslušné mřížkové konstanty

Na **Obr. 1.4** je znázorněna část energetického diagramu prvků IV. skupiny periodické tabulky (které mají podobnou elektronovou strukturu se čtyřmi valenčními elektrony) v závislosti na jejich mřížkové konstantě. V případě velké vzdálenosti mezi jednotlivými atomy je energetické spektrum čárové, protože jednotlivé atomy spolu neinteragují. Při vzájemném přibližování atomů v krystalu, obsahujícím N atomů, se hladina s rozštěpí na $2N$ dovolených stavů a hladina p na $6N$ dovolených stavů. Představíme-li si, že 1 cm³ křemíku obsahuje asi $5 \cdot 10^{22}$ atomů, můžeme hovořit o spojitých pásech dovolených energií.

Pásové diagramy dovolených energií elektronů jsou důležitým modelem pro studium elektrických vlastností pevných látek. Hovoříme o tzv. *pásové teorii* nebo *pásovém modelu pevných látek*. Pásy dovolených hodnot energie elektronu označíme jako *dovolené pásy*, oblasti mezi těmito pásy jsou pásy zakázaných hodnot energie elektronu - *zakázané pásy*. Jednoduchý pásový model pevné látky je na **Obr. 1.5**. Energetické pásy dělíme na tyto skupiny:

1. *Vnitřní pásy*: Tyto pásy patří elektronům pevně vázaným k jádrům. Jsou poměrně úzké a pro přenos elektrického náboje nemají význam.
2. *Valenční pás*: V tomto pásu jsou hladiny elektronů vytvářejících chemickou vazbu.
3. *Vodivostní pás*: V tomto pásu jsou hladiny elektronů uvolněných z chemických vazeb. Tyto elektrony se mohou pohybovat v meziatomovém prostoru a způsobovat vodivost látky. Je to nejvyšší dovolený energetický pás.
4. *Zakázané pásy*: Oddělují od sebe pásy dovolených energií.



Obr. 1.5: Jednoduchý pásový model pevné látky

Poslední tři pásy - vodivostní, zakázaný a valenční - svojí polohou a energetickou šířkou určují základní vlastnosti pevných látek z hlediska vedení elektrického proudu (přenosu náboje). Běžně se dolní okraj vodivostního pásu označuje E_C (C = conductivity), E_V je horní kraj valenčního pásu (V = valence). Zakázaný pás má šířku E_G (nebo ΔE ; G = gap).

Elektrické vlastnosti látky charakterizuje zejména šířka zakázaného pásu E_G , která mimo jiné závisí na meziatomové vzdálenosti a_0 atomů v krystalu. Na **Obr. 1.4** vidíme, že Ge s mřížkovou konstantou 0,565 nm má menší E_G ($E_G = 0,66$ eV) než Si ($a_0 = 0,543$ nm, $E_G = 1,12$ eV) a mnohem menší než diamant C ($a_0 = 0,357$ nm, $E_G = 5,47$ eV). Hodnoty jsou uvedeny pro teplotu 300 K.

1.2 Rozdělení látek podle jejich elektrických vlastností

Objev dovolených a zakázaných pásů energie v krystalických látkách vedl k jednoduchému a názornému výkladu elektrické vodivosti. Rozdělení látek na *vodiče*, *polovodiče* a *izolanty* se provádí podle hodnot měrné vodivosti a její teplotní závislosti a jako výrazné kritérium se uvádí šířka zakázaného pásu.



Obr. 1.6: Pásový model a) izolantu; b) polovodiče; c) monovalentního kovu; d) bivalentního kovu

Na **Obr. 1.6** jsou vedle sebe znázorněny pásové modely pro izolant, polovodič a vodič (kov). Předpokládáme, že látky zde uvažované jsou na teplotě absolutní nuly (0 K), kdy elektrony zaujímají nejnížší možné energetické úrovně. Výpočty při teplotě absolutní nuly

jsou jenom teoretickou abstrakcí, která nám umožňuje jednodušší postupy a snazší pochopení problému, než při uvažování reálných teplot.

U *izolantů* je valenční pás zaplněn elektrony, nad ním ležící zakázaný pás je širší než 3 eV (např. diamant 5,48 eV při 0 K). Izolanty téměř nevedou elektrický proud, protože příliš široký zakázaný pás nedovoluje elektronům přejít z valenčního pásu do pásu vodivostního (to platí za všech teplot - až do tepelného nebo elektrického průrazu).

Pásový diagram *polovodičů* je shodný s pásovým diagramem izolantů, ovšem šířka zakázaného pásu je 0,1 až 2 eV (u Ge 0,66 eV a u Si 1,12 eV při 0 K). Za teploty 0 K je vodivostní pás prázdný, tj. v polovodiči nejsou žádné volné elektrony, které by mohly přenášet náboj (vést proud), a proto se polovodič chová jako izolant. Při zvyšování teploty mohou některé elektrony získat dostatečnou energii a přejít do vodivostního pásu - polovodič začíná vést proud.

U *kovů* se vyskytují dva případy. Monovalentní kovy (např. sodík, rubidium) mají velmi úzký zakázaný pás ($E_G < 0,1$ eV) a vodivostní pás je i za teploty 0 K částečně zaplněn. Bivalentní kovy (např. měď) nemají zakázaný pás a vodivostní pás se překrývá s valenčním. Kovy jsou proto dobré (elektrické i tepelné) vodiče, a to i za velmi nízkých teplot. Mrak volných elektronů v kovu se též nazývá elektronovým plynem.

Tab. 1.1: Měrný odpor kovů, polovodičů a izolantů

LÁTKA	KOVY (VODIČE)	POLOVODIČE	IZOLANTY (NEVODIČE)
měrný odpor [Ωcm]	se stoupající teplotou se zvětšuje 10 ⁻⁵ až 10 ⁻⁶	se stoupající teplotou se zmenšuje 10 ⁻² až 10 ⁸	se stoupající teplotou se zmenšuje 10 ⁸ až 10 ¹⁸

Polovodiče netvoří samostatnou ostře ohraničenou skupinu látek. V podstatě jsou to izolanty, které se působením vnitřních nebo vnějších činitelů (teplo, světlo, tlak, elektromagnetické pole, radioaktivní záření, poruchy krystalové mřížky) mění na částečné vodiče elektrického proudu.

Některé charakteristické vlastnosti kovů, polovodičů a izolantů jsou shrnuty v **Tab. 1.1.** a na **Obr. 1.7.**



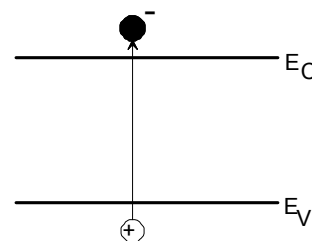
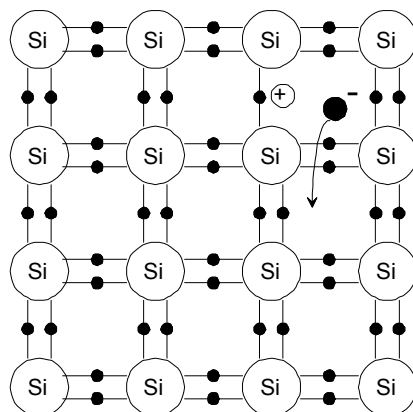
Obr. 1.7: Vliv teploty a osvětlení na vodič a polovodič

1.3 Elektrická vodivost polovodičů

U polovodičů existují dva typy vodivosti, tzv. vlastní a nevlastní vodivost. *Vlastní* (intrinseckou) *vodivostí* se vyznačují všechny polovodiče, *nevlastní vodivost* existuje jen u tzv. *příměsových* (dotovaných, extrinzeckých, nevlastních, legovaných) *polovodičů*.

1.3.1 Vlastní polovodiče

Dokonalý vlastní polovodič (dokonalý krystal bez poruch a příměsí) se v přírodě nevyskytuje, při pečlivé výrobě se mu však můžeme přiblížit. Vlastní (intrinzecký) polovodič se podobá izolantu. Za teploty 0 K je vodivostní pás prázdný, tj. v polovodiči nejsou žádné volné elektrony, které by mohly vést proud. Vlivem teploty nebo jiného vnějšího vlivu však může elektron získat



Obr. 1.8: Vznik páru elektron - díra u vlastního polovodiče; pokud nepůsobí elektrické pole, může se volný elektron pohybovat libovolným směrem

dostatečnou energii a "přeskočit" do vodivostního pásu. Hovoříme o tzv. *ionizaci*, k níž je nutné určité množství tzv. *ionizační energie*. Ionizační energie (užívá se termín aktivační energie) je nutná k překonání zakázaného pásu a uvolnění elektronu z vazby, její velikost musí tedy být nejméně E_G .

Elektron, který se uvolní, zanechá po sobě ve valenčním pásu atom postrádající jeden elektron (chová se potom jako kladný iont). Tento kladný iont se může zneutralizovat přijetím elektronu, který k němu přejde ze sousedního atomu. Tím se však z tohoto sousedního atomu stane kladný iont a může být opět zneutralizován příchodem dalšího elektronu. Vidíme, že atomy se jako pevná součást mřížky nepohybují, ale kladný náboj se může přemísťovat. Tento kladný náboj nazýváme *díra*. Přestože díra jako reálná částice neexistuje, je užitečné jí přiřadit některé vlastnosti materiální částice (podobné, jako má elektron, ale opačný náboj $+q$).

Ve vlastních polovodičích mohou tedy přenášet elektrický proud za teplot nad 0 K jednak volné elektrony ve vodivostním pásu (každý elektron přenáší náboj $-q$), jednak pohyblivé díry ve valenčním pásu (náboj $+q$). Pro obě tyto částice se užívá název *nosiče* (nebo *nositelé*) *náboje*. Protože současně vzniká volný elektron a pohyblivá díra, hustota elektronů se rovná hustotě děr

$$n = p = n_i \quad (1.2)$$

kde n je počet elektronů, p je počet děr a n_i je intrinzická koncentrace nosičů na jednotku objemu; podle soustavy SI je to 1 m^3 , v praxi i v literatuře se však běžně používá 1 cm^3 .

Je užitečné si zapamatovat, že intrinzická koncentrace nosičů je při 300 K pro křemík 10^{16} m^{-3} (10^{10} cm^{-3}) a pro germanium 10^{19} m^{-3} (10^{13} cm^{-3}), intrinzická koncentrace však silně závisí na teplotě. Proces uvolňování elektronů při ionizaci atomů nezpůsobuje změnu výsledného náboje daného objemu polovodiče (po uvolnění elektronu zůstává v krystalové mřížce ionizovaný atom s kladným nábojem).

Vlastní polovodiče jsou tedy při teplotách kolem 300 K vodiče elektrického proudu, vlivem závislosti intrinzické koncentrace na teplotě i jejich vodivost silně závisí na teplotě. Rozsah jejich měrných odporů je přibližně 10^{-2} až $10^9 \Omega\text{cm}$ (dobré vodiče mají měrný odpor $< 10^{-6} \Omega\text{cm}$, izolanty $> 10^{14} \Omega\text{cm}$).

1.3.2 Nevlastní polovodiče

Na elektrickou vodivost polovodičů mají výrazný vliv cizí atomy zabudované do krystalové mřížky (poruchy), které jsou v reálném krystalu vždy přítomny. Příměsové atomy jsou do co nejčistšího materiálu zaváděny i uměle, abychom získali materiál s vhodnými elektrickými vlastnostmi. Tento technologický postup, kdy se do základního materiálu přidávají atomy cizích prvků, nazýváme *dotování* (nebo též dopování či legování; přidávaná látka se nazývá dopant či legura). Nejčastěji přítom vstupují atomy příměsí do krystalové mřížky jako náhrada některých původních atomů (tzv. substituce).

Polovodič typu N

Je-li atom krystalové mřížky čtyřmocného prvku (prvku IV. skupiny - např. Si, Ge) nahrazen atomem pětimocného prvku (prvku V. skupiny - např. P, As, Sb), pak čtyři z jeho valenčních elektronů se účastní vytvoření vazby se sousedními atomy, pátý elektron je nadbytečný a je vázán k atomu velmi slabě. Tento elektron může být uvolněn dodáním velmi malé (aktivační) aktivační energie a účastnit se vedení proudu. Typická hodnota aktivační energie je asi 0,05 eV pro běžné příměsi v Si; každá příměs má svoji typickou hodnotu aktivační energie. Tuto energii elektron získá už při nízkých teplotách, takže při pokojové teplotě jsou téměř všechny atomy příměsí ionizovány. Vznikají tak kladné ionty příměsí, které zůstávají v mřížkových polohách, a volné elektrony. Proto se pětimocné příměsi nazývají

donory (dodávají elektrony do vodivostního pásu).

V pásovém diagramu vytvářejí donorové atomy tzv. *donorovou hladinu* (s energií E_D), která se nachází v zakázaném pásu těsně pod dnem vodivostního pásu (E_C) - viz **Obr. 1.9**.

V polovodiči typu N (v polovodiči s elektronovou vodivostí) převažují jako nosiče proudu elektrony, nazýváme je proto *majoritní nosiče*.

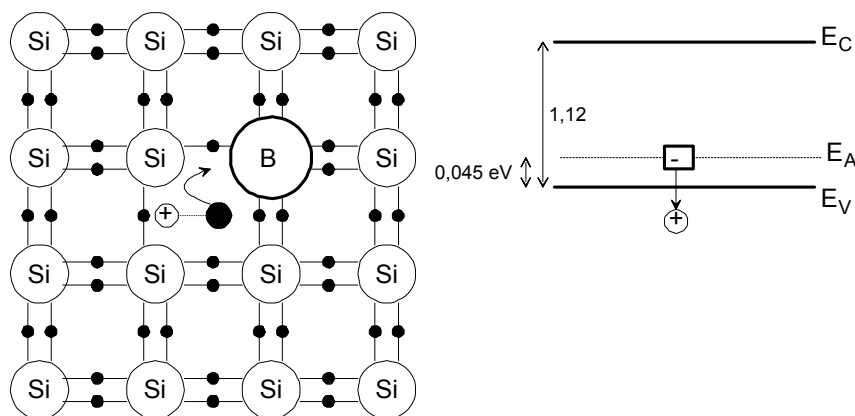


Obr. 1.9: Krystalová mřížka Si s jedním donorovým atomem a pásový model polovodiče typu N; pokud nepůsobí elektrické pole, může se volný elektron pohybovat libovolným směrem

Polovodič typu P

Podobně nahradíme-li atom krystalové mřížky čtyřmocného prvku atomem trojmocného prvku (prvku III. skupiny - např. B, In, Ga, Al), pak se všechny tři jeho valenční elektrony účastní vazeb se sousedními atomy a čtvrtá zůstane neúplná. Stačí malá energie k tomu (opět typická hodnota je asi 0,05 eV pro Si), aby se některý z elektronů ze sousedních vazeb uvolnil a zaplnil neúplnou vazbu trojmocného atomu - tj. trojmocný atom se ionizuje záporně a vytvoří v sousedství "kladnou" díru (tj. nedostatek jednoho elektronu). Tato díra se může pohybovat a umožnit tak vedení proudu (působí-li na ni vnější elektrické pole). Trojmocné příměsi se nazývají *akceptory* (přijímají, lépe řečeno zachycují elektrony z valenčního pásu). V pásovém diagramu vytvářejí akceptorové příměsi tzv. *akceptorovou hladinu* (s energií E_A),

ležící těsně nad stropem valenčního pásu (E_V) - viz **Obr. 1.10**. Již za malé teploty jsou téměř všechny akceptorové příměsi ionizovány a ve valenčním pásu jsou vytvořeny díry, které v polovodiči typu P (v polovodiči s děrovou vodivostí) převažují jako nosiče proudu - majoritními nosiči jsou zde díry.



Obr. 1.10: Krystalová mřížka Si s jedním akceptorovým atomem a pásový model polovodiče typu P; pokud nepůsobí elektrické pole, může se volná díra pohybovat libovolným směrem

Vodivost

příměsových polovodičů (typu N a P) se nazývá souborně *nevlastní* (extrinzičnou) *vodivostí*. Při ionizaci (aktivaci) příměsi vzniká jeden pohyblivý nosič a jeden nepohyblivý iont. Při

ionizaci atomů ve vlastních (intrinzických) polovodičích vzniká pohyblivý elektron a pohyblivá díra.

Případ vlastního polovodiče bez příměsí i případy čistě donorového nebo čistě akceptorového polovodiče (bez příměsí opačného typu) jsou případy pouze teoretické, protože reálné technologické možnosti nám neumožňují takovéto ideální polovodiče vyrobit.

Při zvyšování koncentrace příměsí se donorové nebo akceptorové hladiny štěpí - vznikají pásy příměsí. Při velké koncentraci příměsí splývá akceptorová hladina s valenčním pásem a donorová hladina s vodivostním pásem - polovodič se chová jako kov (*degenerovaný polovodič*).

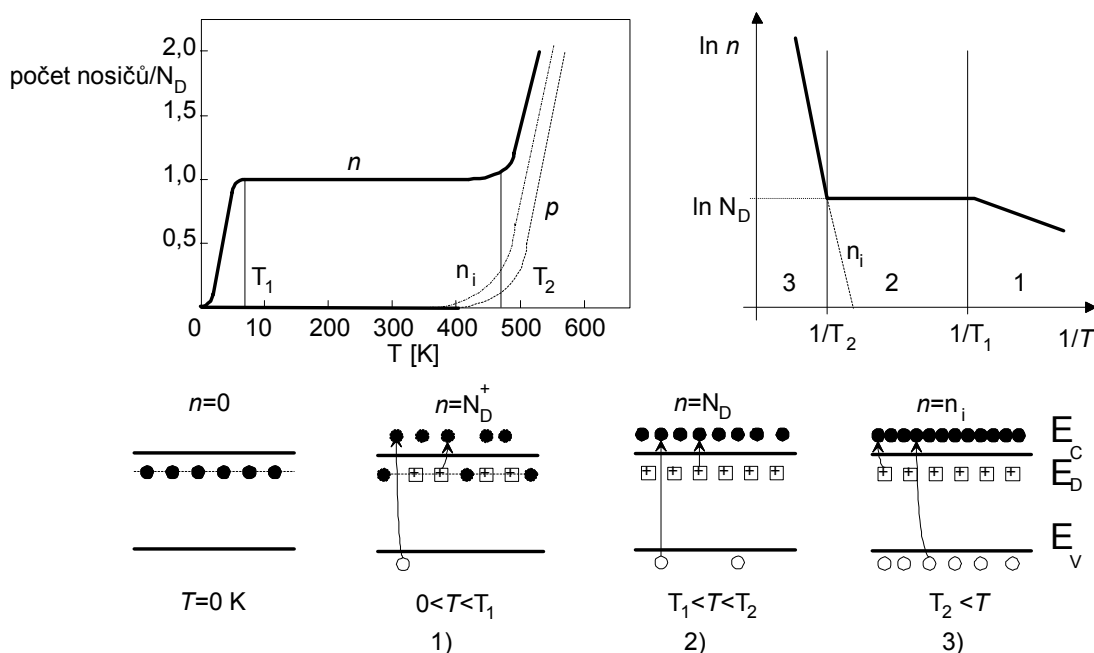
Teplotní závislost koncentrace nosičů náboje v polovodičích

Kromě nevlastní vodivosti má i nevlastní polovodič vlastní vodivost. Je možné odvodit, že za *tepelné rovnováhy* platí mezi koncentrací děr a volných elektronů důležitý vztah:

$$np = n_i^2 \quad (1.3)$$

Koncentrace volných nosičů náboje závisí na koncentraci donorů a akceptorů, teplotě a pásové struktuře polovodiče. Pro pochopení principů vlastní a nevlastní vodivosti nám dobře poslouží závislost koncentrace volných nosičů na teplotě např. pro polovodič typu N - viz **Obr. 1.11**.

Při nulové teplotě ($T = 0 \text{ K}$) je koncentrace obou typů nosičů rovna nule. Při zvyšování teploty dochází nejprve k aktivaci nevlastních příměsí, protože lokální hladiny donorů jsou v těsné blízkosti vodivostního pásu. Koncentrace volných elektronů je přibližně rovna koncentraci těchto aktivovaných donorů (N_D^+). Při tzv. *první aktivační teplotě* T_1 (pro Si asi 80 K) dochází téměř k plné aktivaci příměsí a při dalším zvyšování teploty až do teploty T_2 je koncentrace nosičů téměř konstantní a rovná koncentraci donorů (N_D). Při teplotě T_2 , která se nazývá *druhou aktivační teplotou*, dochází k aktivaci mechanismu vlastní generace a z nevlastního polovodiče se stává polovodič vlastní. Oblast mezi první a druhou aktivační teplotou (označenou jako oblast 2 na **Obr. 1.11**), tedy přibližně interval 100 až 450 K, vymezuje normální pracovní teplotní rozsah většiny polovodičových součástek (tzv. *extrinzickou teplotní oblast*), kdy se polovodič chová jako nevlastní s poměrně malou závislostí vodivosti na teplotě. Pod první aktivační teplotou se polovodič nachází v *oblasti zamrznutí* (oblast 1) a nad druhou aktivační teplotou (oblast 3) v *intrinzické teplotní oblasti*. Sklon přímky $\ln n = f(1/T)$ v oblasti 3 odpovídá aktivační energii, respektive ΔE_G .



Obr. 1.11: Závislost koncentrace nosičů na teplotě u nevlastního polovodiče typu N v lineárním měřítku (pro $N_D = 10^{15} \text{ cm}^{-3}$) a přehlednější závislost $\ln n$ na $1/T$. Ve spodní části obrázku je kvalitativní vysvětlení teplotní závislosti koncentrace nosičů na teplotě.

Podrobněji se budeme problematikou teplotní závislosti koncentrace nosičů zabývat později, až se seznámíme s dalšími důležitými pojmy z fyziky polovodičů.

1.4 Elektrony v pevných látkách a jejich energetické rozdělení

Dosud byly naše poznatky o vlastnostech polovodičů spíše kvalitativní. Pro praktické výpočty však potřebujeme znát například číselnou hodnotu koncentrace nosičů náboje nebo jejich energetické rozložení v příslušném energetickém pásu. Proto se teď budeme elektrony v pevných látkách zabývat podrobněji.

1.4.1 Hustota stavů

Jak můžeme vyvodit z našich předchozích poznatků, celkový počet dovolených stavů ve vodivostním a valenčním pásu Si je roven čtyřnásobku počtu atomů v krystalu (každý atom Si má čtyři valenční elektrony). Nyní si položíme otázku, jak jsou tyto dovolené stavy energeticky rozloženy, tj. kolik dovolených stavů můžeme nalézt pro nějakou danou energii ve valenčním nebo vodivostním pásu. Tomuto energetickému rozložení dovolených stavů říkáme *hustota stavů*. Výpočet hustoty stavů je založen na kvantově mechanické analýze problému a přesahuje rámec těchto skript. Uvedeme zde proto jenom souhrn výsledků. Pro energie nepříliš vzdálené od hranice pásu můžeme odvodit [2, 3]:

$$g_C(E) = \frac{m_n^* \sqrt{2m_n^* (E - E_C)}}{\pi^2 \hbar^3} = \text{konst.} \sqrt{E - E_C}, \quad E \geq E_C \quad (1.4)$$

$$g_V(E) = \frac{m_p^* \sqrt{2m_p^* (E_V - E)}}{\pi^2 \hbar^3} = \text{konst.} \sqrt{E_V - E}, \quad E \leq E_V \quad (1.5)$$

kde $g_C(E)$ a $g_V(E)$ jsou hustoty stavů v závislosti na energii ve vodivostním a valenčním pásu a m_n^* a m_p^* jsou efektivní hmotnosti elektronu a díry.

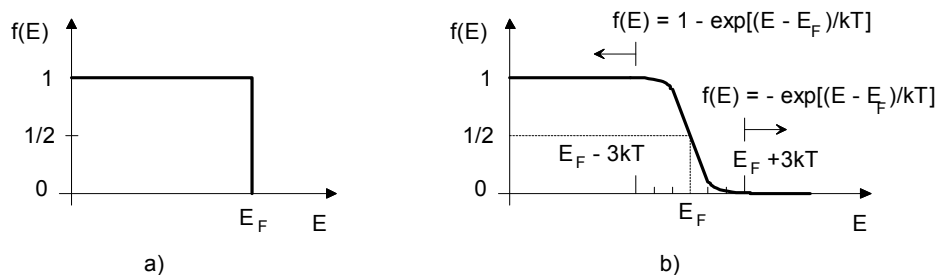
Na **Obr. 1.13** a, b, c. vidíme obecnou závislost hustoty stavů na energii. Jestliže budeme předpokládat dvě velmi blízké energie E a $E + dE$, pak $g_C(E)dE$ je počet dovolených stavů v jednotce objemu, které leží ve vodivostním pásu mezi energiemi E a $E + dE$ (podobnou úvahu můžeme provést i pro valenční pás). Jednotkou hustoty stavů je tedy "počet.cm⁻³.eV⁻¹".

1.4.2 Fermi - Diracova rozdělovací funkce

Zatímco hustota stavů udává, kolik dovolených stavů existuje pro danou energii E , Fermi-Diracova rozdělovací funkce $f(E)$ určuje, kolik existujících dovolených stavů pro energii E bude zaplněno elektrony, nebo jinými slovy, určuje pravděpodobnost, že dovolený stav bude zaplněn elektronem. Fermi-Diracova funkce je tedy funkce hustoty pravděpodobnosti:

$$f(E) = \frac{1}{1 + \exp\left[\frac{E - E_F}{kT}\right]} \quad (1.6)$$

kde E_F je Fermiho energie (Fermiho hladina), k je Boltzmannova konstanta ($k = 1,38 \cdot 10^{-23}$ JK⁻¹) a T je teplota v Kelvinech.



Obr. 1.12: Energetická závislost Fermi-Diracovy funkce. **a)** $T \rightarrow 0$ K; **b)** zobecněná závislost (s jednotkou na ose energie v kT) pro $T > 0$ K.

Podívejme se na tuto funkci blíže. Jestliže $T \rightarrow 0$ K, pak $(E - E_F)/kT \rightarrow -\infty$ pro všechny energie $E < E_F$ a $(E - E_F)/kT \rightarrow +\infty$ pro všechny energie $E > E_F$. Tedy $f(E) = 1$ pro $E < E_F$ a $f(E) = 0$ pro $E > E_F$. Tento výsledek je zakreslen na **Obr. 1.13a**. Vidíme, že všechny stavy pro energie pod E_F budou obsazeny elektrony a všechny stavy pro energie nad E_F budou prázdné.

Nyní předpokládejme teplotu $T > 0$ K. Potom pro $E = E_F$ nalezneme $f(E_F) = 1/2$. Protože pro běžné výpočty je použití Fermi-Diracovy funkce (1.6) nepraktické, používáme její zjednodušený tvar. Dá se dokázat, že pro $E \geq E_F + 3kT$ platí $f(E) \approx \exp[-(E - E_F)/kT]$ a pro $E \leq E_F - 3kT$ platí $f(E) \approx 1 - \exp[(E - E_F)/kT]$. Při pokojové teplotě ($T \approx 300$ K) je $kT \approx 0,026$ eV a $3kT \approx 0,078$ eV $\ll E_G$ (pro Si je E_G asi 1,12 eV), což znamená, že se nedopouštíme příliš velké chyby, jestliže používáme tyto zjednodušené vztahy pro výpočet $f(E)$. Tyto výsledky jsou shrnuty na **Obr. 1.13b**. Je zřejmé, že se vzrůstající teplotou se zvětšuje hodnota $3kT$ a křivka se "roztahuje".

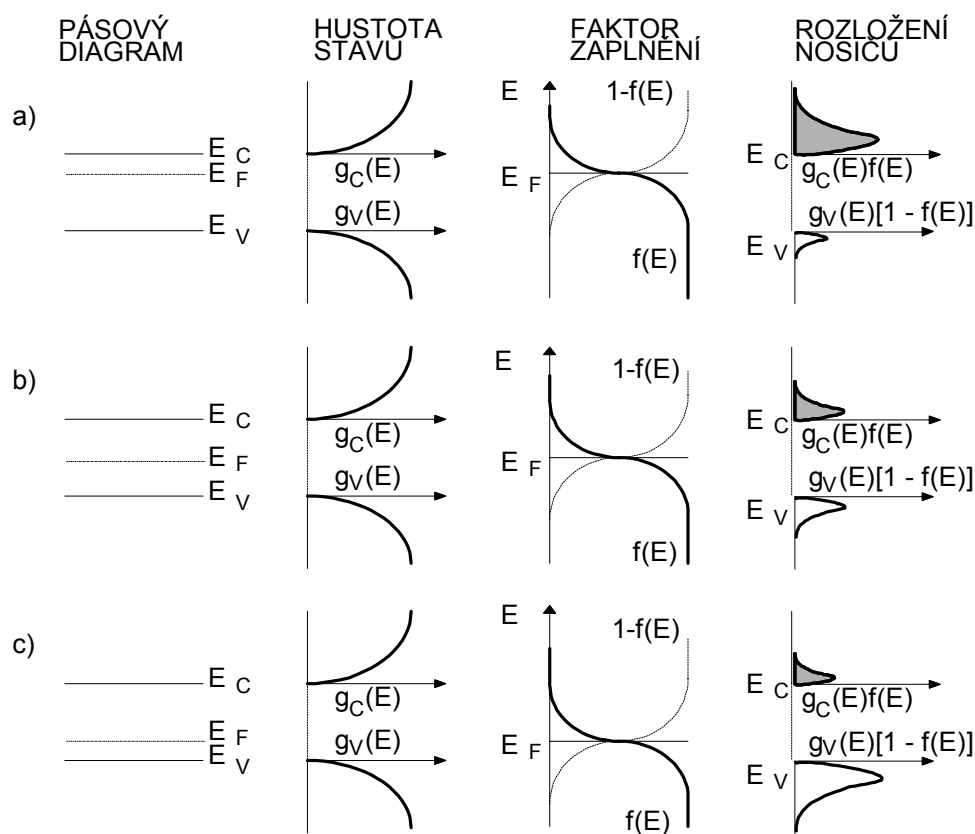
Z předchozího rozboru plynou dvě možné **definice Fermiho energie**:

1. Při teplotě $T = 0$ K odděluje Fermiho energie obsazené stavy od neobsazených.
2. Při teplotě $T > 0$ K je Fermiho energie definována jako hladina, která je obsazena s pravděpodobností $1/2$.

Ještě musíme zdůraznit, že Fermi-Diracova rozdělovací funkce platí v polovodičích pouze za rovnovážných podmínek. I když jsme ji zavedli ve spojitosti s fyzikou polovodičů, je to statistická funkce, která má v kvantové mechanice obecnější platnost.

1.4.3 Rovnovážné rozložení nosičů

Stanovili jsme již hustotu dovolených stavů a pravděpodobnost zaplnění těchto stavů za rovnovážných podmínek. Nyní můžeme jednoduše stanovit rozložení nosičů v jednotlivých energetických páslech. Součin hustoty dovolených stavů a faktoru jejich zaplnění, tedy $f(E) \cdot g_C(E)$, udává rozložení elektronů ve vodivostním pásu. Podobně součin hustoty stavů a faktoru jejich nezaplnění, $[1-f(E)] \cdot g_V(E)$, dává rozložení děr (nezaplněných stavů) ve valenčním pásu. Souborně je rovnovážné rozložení nosičů znázorněno na **Obr. 1.13**.



Obr. 1.13: Rozložení nosičů při poloze Fermiho hladiny: **a)** v horní polovině zakázaného pásu (polo-vodič N), **b)** uprostřed zakázaného pásu (vlastní polovodič), **c)** v dolní polovině zakázaného pásu (polovodič P)

Ve vlastním (intrinzičím) polovodiči leží Fermiho hladina uprostřed zakázaného pásu ($E_F = E_i$), v polovodiči typu N v horní polovině zakázaného pásu a v polovodiči typu P v dolní polovině zakázaného pásu.

1.5 Rovnovážná koncentrace nosičů

1.5.1 Vztahy pro výpočet koncentrace elektronů a děr

Počet elektronů na jednotku objemu ve vodivostním pásu, ležících mezi energiemi E a $E + dE$, můžeme spočítat jako součin počtu dovolených stavů v tomto energetickém intervalu a pravděpodobnosti jejich obsazení elektronem, tedy $f(E)g_C(E)dE$. Je zřejmé, že chceme-li spočítat celkový počet elektronů ve vodivostním pásu, musíme tento vztah integrovat přes všechny energie vodivostního pásu, tedy od dna vodivostního pásu E_C až po jeho strop E_{strop}

$$n = \int_{E_C}^{E_{strop}} f(E)g_C(E)dE \quad (1.7)$$

Podobný vztah můžeme napsat i pro celkový počet děr ve valenčním pásu, integrujeme-li od dna valenčního pásu E_{dno} do jeho stropu E_V

$$p = \int_{E_{dno}}^{E_V} [1 - f(E)]g_V(E)dE \quad (1.8)$$

Dosadíme-li do tohoto vztahu z rovnic (1.4) a (1.6), pak dostaneme pro koncentraci elektronů ve vodivostním pásu a děr ve valenčním pásu

$$n = \frac{m_n^* \sqrt{2m_n^*}}{\pi^2 \hbar^3} \int_{E_C}^{E_{strop}} \frac{\sqrt{E - E_C} dE}{1 + \exp[(E - E_F)/kT]} \quad (1.9)$$

$$p = \frac{m_p^* \sqrt{2m_p^*}}{\pi^2 \hbar^3} \int_{E_{dno}}^{E_V} \frac{\exp[(E - E_F)/kT] \sqrt{E_V - E} dE}{1 + \exp[(E - E_F)/kT]} \quad (1.10)$$

Po integraci a dalších úpravách dostaneme důležité vztahy pro výpočet koncentrace elektronů a děr

$$n = N_C \exp [-(E_C - E_F)/kT] \quad (1.11)$$

$$p = N_V \exp [-(E_F - E_C)/kT] \quad (1.12)$$

kde N_C je efektivní hustota stavů ve vodivostním pásu a N_V je efektivní hustota stavů ve valenčním pásu

$$N_C = 2 \left[\frac{2\pi m_n^* kT}{h^2} \right]^{3/2} = \text{konst} \cdot T^{3/2} \quad (1.13)$$

$$N_V = 2 \left[\frac{2\pi m_p^* kT}{h^2} \right]^{3/2} = \text{konst} \cdot T^{3/2} \quad (1.14)$$

Vztahy (1.11) a (1.12) však nejsou příliš pohodlné pro běžné výpočty. Protože už víme, že Fermiho hladina u vlastního polovodiče leží ve středu zakázaného pásu ($E_F = E_i$) a dále pro vlastní polovodič platí $n = p = n_i$, můžeme vztahy (1.11) a (1.12) dále zjednodušit:

$$n_i = N_C \exp [-(E_C - E_i)/kT] \quad (1.15)$$

$$n_i = N_V \exp [-(E_i - E_C)/kT] \quad (1.16)$$

Odtud

$$N_C \exp(-E_C/kT) = n_i \exp(-E_i/kT) \quad (1.17)$$

$$N_V \exp(E_V/kT) = n_i \exp(E_i/kT) \quad (1.18)$$

Nyní po dosažení ze vztahů (1.15) a (1.16) do původních vztahů (1.11) a (1.12) získáme velmi důležité vztahy

$$n = n_i \exp [(E_F - E_i)/kT] \quad (1.19)$$

$$p = n_i \exp [(E_i - E_F)/kT] \quad (1.20)$$

Stejně jako rovnice (1.15) a (1.16) platí rovnice (1.19) a (1.20) **pro libovolný polovodič za rovnovážných podmínek.**

Nyní si už můžeme odvodit vztah pro podmínku tepelné rovnováhy (1.3). Vynásobíme-li rovnice (1.19) a (1.20), dostáváme skutečně vztah (1.3):

Vynásobením rovnic (1.11) a (1.12) můžeme vypočítat hodnotu intrinzecké koncentrace n_i :

$$n_i^2 = N_C N_V \exp [-(E_C - E_V)/kT] = N_C N_V \exp(-E_G/kT) \quad (1.21)$$

a po odmocnění

$$n_i = \sqrt{N_C N_V} \exp(-E_G / 2kT) = \text{konst} \cdot T^{3/2} \cdot \exp(-E_G/2kT) \quad (1.22)$$

Z rovnice (1.22) je zřejmé, že intrinzecká koncentrace n_i s rostoucí teplotou exponenciálně vzrůstá. Při pokojové teplotě (300 K) je pro křemík n_i přibližně rovno $1,45 \cdot 10^{10} \text{ cm}^{-3}$ ($= 1,45 \cdot 10^{16} \text{ m}^{-3}$).

1.5.2 Nábojová neutralita

Další důležitou vlastností polovodičů je zachovávat za rovnovážných podmínek *nábojovou neutralitu*. Předpokládáme stejnoměrně dotovaný polovodič s konstantním počtem atomů dopantu v celém objemu. Za rovnovážných podmínek musí být takovýto polovodič nábojově neutrální, součet všech nábojů v jeho objemu musí být roven nule. Předpokládáme-li dále, že všechny příměsi jsou ionizované a uvědomíme-li si, že náboj ionizovaných donorů je kladný a náboj ionizovaných akceptorů záporný, pak musí platit:

$$\text{náboj/objem} = qp - qn + qN_D - qN_A = 0 \quad (1.23)$$

nebo

$$p - n + N_D - N_A = 0 \quad (1.24)$$

Vztah (1.24) nám umožňuje provádět další výpočty koncentrace nosičů pro vlastní i nevlastní polovodič.

Vlastní polovodič

Vlastní polovodič neobsahuje žádné příměsi a proto $N_D = 0$, $N_A = 0$. Platí proto

$$p - n + N_D - N_A = 0 \quad \Rightarrow \quad p = n \quad (1.25)$$

$$np = n^2 = p^2 = n_i^2 \Rightarrow \quad n = p = n_i \quad (1.26)$$

Z podmínky nábojové neutrality jsme tedy dostali stejné výsledky, které jsme na základě našich předchozích úvah pro vlastní polovodič předpokládali.

Nevlastní polovodič

Předpokládejme nyní nedegenerovaný polovodič typu N s plně ionizovanými příměsemi. Reálný polovodič typu N obsahuje vždy určité množství akceptorových příměsí (takový polovodič je nazýván *kompensovaný polovodič*), ale platí pro něj $N_D \gg N_A$ nebo $N_D - N_A \approx N_D$. Z podmínky nábojové neutrality plyne

$$p - n + N_D = 0 \quad (1.27)$$

a z podmínky tepelné rovnováhy (1.3)

$$p = n_i^2/n \quad (1.28)$$

Po dosazení vztahu (1.28) do (1.27) dostaneme

$$n_i^2/n - n + N_D = 0 \quad (1.29)$$

$$n^2 - nN_D - n_i^2 = 0 \quad (1.30)$$

což je kvadratická rovnice s neznámou n , jejíž řešení je

$$n = \frac{N_D}{2} + \left[\left(\frac{N_D}{2} \right)^2 + n_i^2 \right]^{1/2} \quad (1.31)$$

Odtud můžeme pro daný polovodič a teplotu spočítat koncentraci elektronů n a z (1.28) pak koncentraci děr p . Pro většinu běžných výpočtů nám však postačí následující úvaha: Při pokojové teplotě je pro křemík n_i asi 10^{10} cm^3

Koncentrace dopantů bývá nejméně 10^{14} cm^{-3} , tedy $N_D \gg n_i$. Z (1.31) potom plyne $n \approx N_D$ a $p \approx n_i^2/N_D$. Analogickou úvahu můžeme provést i pro polovodič typu P. Shrňme-li tyto výsledky, pak pro nedegenerovaný polovodič typu N s plně ionizovanými příměsemi a při splnění podmínek $N_D \gg N_A$ a $N_D \gg n_i$ platí:

$$n = N_D, \quad p = n_i^2/N_D \quad (1.32)$$

a pro nedegenerovaný polovodič typu P s plně ionizovanými příměsemi a při splnění podmínek $N_A \gg N_D$ a $N_A \gg n_i$

$$p = N_A, \quad n = n_i^2/N_A \quad (1.33)$$

Při vzrůstající teplotě se intrinzičká koncentrace n_i zvětšuje (viz vztah (1.22) a **Obr. 1.13**) a při dostatečně vysoké teplotě může dokonce přesáhnout koncentraci příměsí N_A nebo N_D . Jestliže bude teplota dále vzrůstat, $n_i \gg N_D$ a z rovnice (1.31) vyplyne, že $n \approx n_i$. Z (1.32)

potom dostaneme také $p \approx n_i$. Při dostatečně vysokých teplotách se každý polovodič stává vlastním polovodičem (viz také **Obr. 1.10**).

1.5.3 Výpočet polohy Fermiho hladiny

Z **Obr. 1.13.** a ze vztahů (1.11) a (1.12) nebo (1.19) a (1.20) je zřejmá závislost koncentrace nosičů na poloze Fermiho hladiny. Jestliže za rovnovážných podmínek známe n nebo p , můžeme vypočítat E_F nebo naopak.

Vlastní polovodič

Protože pro vlastní polovodič platí $n = p$, pak ze vztahů (1.11) a (1.12) plyne

$$N_C \exp[(E_F - E_C)/kT] = N_V \exp[(E_V - E_F)/kT] \quad (1.34)$$

Řešením této rovnice pro E_F a dosazením za N_V a N_C z (1.13) a (1.14) dostaneme

$$E_F = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln\left(\frac{m_p^*}{m_n^*}\right)^{3/2} \quad (1.35)$$

Protože efektivní hmotnost elektronů a děr není při teplotách větších než 0 K stejná, můžeme vypočítat, že Fermiho hladina se s teplotou posunuje v zakázaném pásu nahoru a leží asi 0,012 eV nad středem zakázaného pásu při teplotě 300 K. Pro běžné výpočty však můžeme tuto odchylku zanedbat.

Nevlastní polovodič

Pro nevlastní polovodič typu N platí přibližně $n = N_D$ a po dosazení do (1.19)

$$N_D = n_i \exp[(E_F - E_i)/kT] \quad (1.36)$$

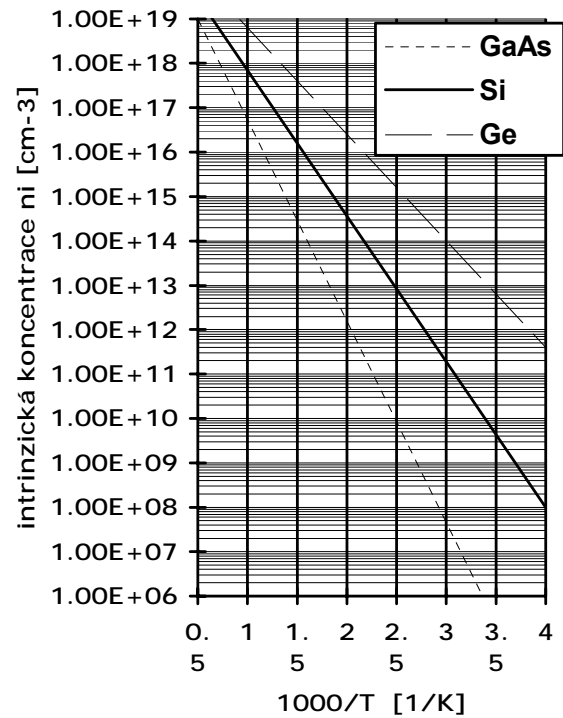
Odtud plyne

$$E_F - E_i = kT \ln(N_D/n_i) \quad (1.37)$$

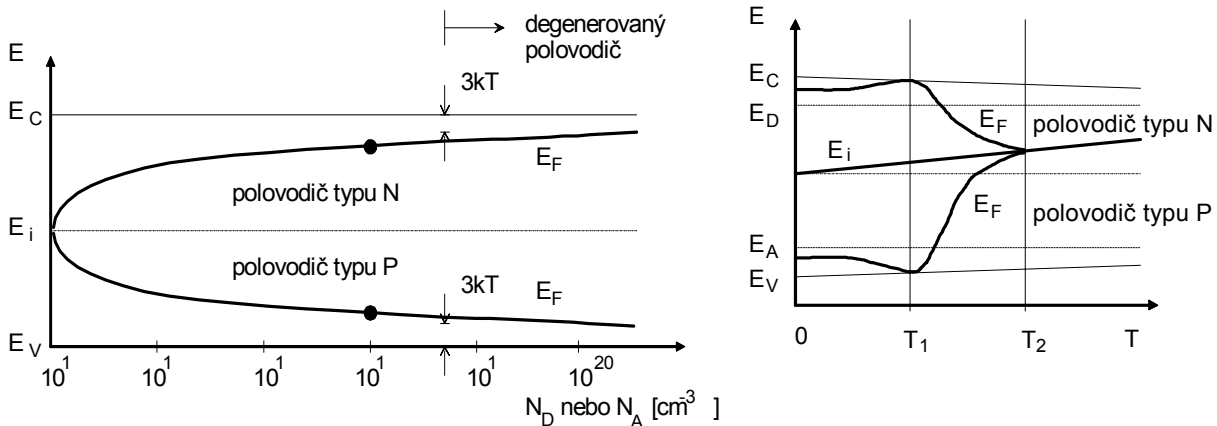
Analogicky pro polovodič typu P dostaneme

$$E_i - E_F = kT \ln(N_A/n_i) \quad (1.38)$$

Z rovnic (1.29) a (1.30) vyplývá, že Fermiho energie je závislá na koncentraci příměsí v nevlastním polovodiči a na teplotě. Fermiho hladina se posunuje se vzrůstající teplotou i se vzrůstající koncentrací příměsí nahoru v polovodiči typu N a dolů v polovodiči typu P. Tato závislost je schématicky naznačena na **Obr. 1.15**. Jestliže se Fermiho hladina přiblíží k E_C nebo E_V na méně než $3kT$, dostaneme *degenerovaný* nebo *vysoce dotovaný* polovodič, který označujeme jako N^+ nebo P^+ materiál. Maximální koncentrace příměsí pro nedegenerovaný polovodič jsou při 300 K pro Si $N_D \approx 1,6 \cdot 10^{18} \text{ cm}^{-3}$ a $N_A \approx 7,7 \cdot 10^{17} \text{ cm}^{-3}$.



Obr. 1.14: Teplotní závislost intrinzičké koncentrace nosičů n_i [cm⁻³] pro GaAs, Si a Ge



Obr. 1.15: a) Poloha Fermiho hladiny v křemíku v závislosti na koncentraci příměsí. V označeném bodě je nakreslena b) závislost polohy Fermiho hladiny na teplotě (pro degenerovaný polovodič). Teploty T_1 a T_2 jsou první a druhá aktivační teplota. Poloha Fermiho hladiny u intrinzičkého polovodiče je označena E_i . Se stoupající teplotou se také zužuje šířka zakázaného pásu se strmostí řádu $10^{-4} \text{ eV K}^{-1}$

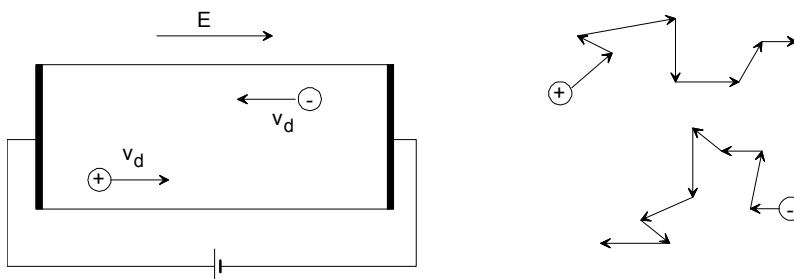
1.6 Vedení proudu v polovodičích

Dosud jsme se zabývali vlastnostmi polovodičů za rovnovážných podmínek. Tyto podmínky nastanou v normální polovodičové součástce jen v případě, že jí neteče žádný proud. Běžné operační podmínky polovodičové součástky se však od rovnovážných liší.

Můžeme rozlišit tři základní mechanizmy podílející se na vedení proudu v polovodičích: *drift, difúzi a generaci - rekombinaci*.

1.6.1 Drift nosičů náboje

Drift je definován jako pohyb nabitých částic způsobený přiloženým elektrickým polem. Můžeme jej popsat následovně: Jestliže na polovodič přiložíme elektrické pole E , pak na díry s nábojem $+q$ působí síla urychlující jejich pohyb ve směru elektrického pole a na elektrony s nábojem $-q$ působí síla urychlující jejich pohyb proti směru elektrického pole. Z mikroskopického hlediska není sice pohyb nosičů náboje přímočarý, protože jsou rozptylovány kolizemi - tepelné kmity mřížky a ionizované atomy příměsí. Sledujeme-li však jejich pohyb z makroskopického hlediska, zjistíme, že se všechny nosiče stejného typu pohybují s konstantní *driftovou rychlostí* v_d . Na Obr. 1.16. je driftový pohyb nosičů náboje znázorněn.



Obr. 1.16: Znázornění driftového pohybu nosičů náboje, a) v makroskopickém a b) v mikroskopickém měřítku

Nyní určíme *driftový proud*, který teče polovodičem vlivem přiloženého elektrického pole. Uvažujme hranol z polovodiče typu P s průřezem o ploše A (viz Obr. 1.17.). Protože proud je definován jako náboj, který za jednotku času projde libovolnou plochou kolmou ke směru toku nosičů, můžeme například driftový proud děr spočítat podle vztahu

$$I_{p,drift} = qp v_d A \quad (1.39)$$

Protože proud je většinou uvažován jako skalární veličina, je někdy užitečné pracovat s *proudovou hustotou* J , jejíž velikost je rovna proudu procházejícím jednotkou plochy $J = I/A$. Vztah (1.39) tedy můžeme přepsat ve vektorovém tvaru

$$\mathbf{J}_{p,drift} = qp \mathbf{v}_d \quad (1.40)$$

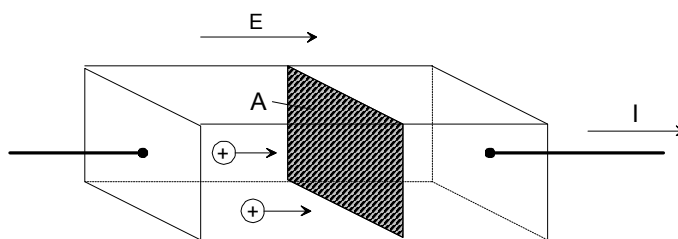
Protože driftový proud vzniká jako odezva na přiložené elektrické pole, zajímá nás přímý vztah mezi těmito veličinami. Experimentálně bylo zjištěno, že driftová rychlost je přímo úměrná přiloženému elektrickému poli

$$\mathbf{v}_d = \mu_p \mathbf{E} \quad (1.41)$$

kde μ_p je *pohyblivost děr*, konstanta úměrnosti mezi driftovou rychlostí a elektrickým polem. Tento vztah neplatí pouze při velkých hodnotách elektrického pole, se kterými se však při běžné činnosti polovodičových součástek nesetkáme (viz Obr. 1.18). Vztah (1.40) tedy můžeme napsat ve tvaru:

$$\mathbf{J}_{p,drift} = q\mu_p p \mathbf{E} \quad (1.42)$$

Podobný vztah můžeme psát i pro elektrony



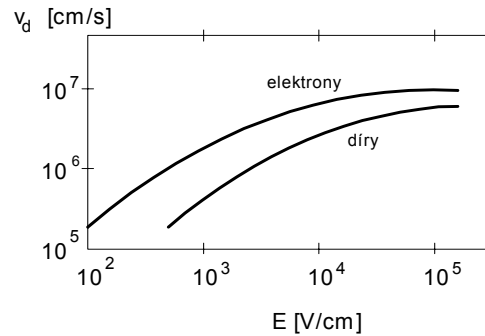
Obr. 1.17: Driftový proud děr polovodičem typu P

$$\mathbf{J}_{n,drift} = q\mu_n n \mathbf{E} \quad (1.43)$$

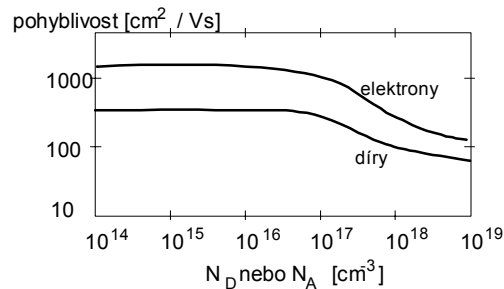
kde μ_n je *pohyblivost elektronů*. Musíme si však uvědomit, že i když drift elektronů jde proti směru přiloženého elektrického pole, tedy $\mathbf{v}_d = -\mu_n \mathbf{E}$, je proud přenášen záporně nabitými nosiči ($\mathbf{J}_{n,drift} = -qn\mathbf{v}_d$) a výsledkem je opět proud ve směru přiloženého elektrického pole (1.43).

Pohyblivost nosičů μ je významná veličina popisující působení elektrického pole na pohyb elektronů a děr. Zmíníme se zde proto stručně alespoň o jejich nejdůležitějších vlastnostech:

1. *Jednotkou* pohyblivosti je v jednotkách SI $\text{m}^2\text{V}^{-1}\text{s}^{-1}$. Běžně je používána také jednotka $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$.
2. Pohyblivost nosičů je *závislá na koncentraci příměsí* v polovodiči a na typu polovodiče. Obecně platí, že se *vzrůstající koncentrací příměsí* pohyblivost klesá (viz **Obr. 1.19**). Jako numerický příklad si uveďme, že $\mu_n = 1300 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ pro $N_D = 10^{14} \text{ cm}^{-3}$ a $\mu_p = 490 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ pro $N_A = 10^{14} \text{ cm}^{-3}$. Obě hodnoty platí pro křemík při 300 K.
3. Pro všechny důležité polovodiče platí, že $\mu_n > \mu_p$.
4. *Teplotní závislost pohyblivosti nosičů* je malá, pro nedegenerované polovodiče (polovodiče s poměrně nízkou koncentrací příměsí) platí, že pohyblivost se *vzrůstající teplotou* klesá. Při vyšších teplotách zpomalují pohyb nosičů náboje tepelné kmity mřížky.



Obr. 1.18: Závislost driftové rychlosti elektronů a děr v křemíku na intenzitě elektrického pole (při teplotě 300 K)



Obr. 1.19: Závislost pohyblivosti elektronů a děr v Si na koncentraci příměsí (při teplotě 300 K)

Vodivost σ je dalším důležitým parametrem, popisujícím vlastnosti polovodičů. Pokud je polovodič stejnoměrně dotován, může jím protékat pouze driftový proud

$$\mathbf{J}_{drift} = \mathbf{J}_{p,drift} + \mathbf{J}_{n,drift} = q(\mu_n n + \mu_p p) \mathbf{E} \quad (1.44)$$

Vztah (1.44) je vlastně Ohmův zákon v diferenciálním tvaru:

$$\mathbf{J}_{drift} = \sigma \mathbf{E} = 1/\wp \mathbf{E} \quad (1.45)$$

kde σ je vodivost a \wp měrný odpor polovodiče (pro měrný odpor budeme používat symbol \wp , protože řecké ρ je snadno změnitelné s p). Pro vodivost polovodiče tedy obecně platí

$$\sigma = q(\mu_n n + \mu_p p) \quad (1.46)$$

Protože už víme, že pro polovodič typu N platí $n = N_D$ a $p \ll n$ a pro polovodič typu P platí $p = N_A$ a $n \ll p$, dostáváme v praxi často používané vztahy

pro polovodič typu N

$$\sigma = q\mu_n N_D \quad (1.47)$$

pro polovodič typu P

$$\sigma = q\mu_p N_A \quad (1.48)$$

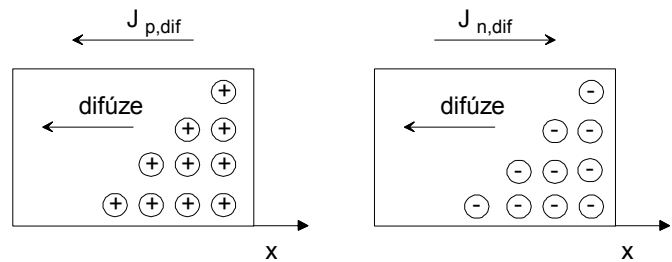
Ze vztahů (1.47) a (1.48) vyplývá, že vodivost polovodiče je přibližně přímo úměrná koncentraci příměsí. Závislost pohyblivosti na koncentraci, o které jsme se zmínili v předchozím odstavci, není tak významná.

1.6.2 Difúze nosičů náboje

Difúze je fyzikální jev, který můžeme pozorovat u jakýchkoliv částic, tedy i u částic bez elektrického náboje. Je to výsledek jejich tepelného pohybu, který směřuje k neuspořádanému rozložení částic. Z makroskopického hlediska pozorujeme, že částice migrují z oblasti s vysokou koncentrací do oblasti s nízkou koncentrací až do doby, kdy se koncentrace částic v celém objemu vyrovná.

V případě částic nesoucích elektrický náboj, jako jsou elektrony a díry v polovodiči, je výsledkem difúze *difúzní proud* (viz **Obr. 1.20**).

Difúzní tok J libovolných částic s koncentrací c je úměrný zápornému gradientu koncentrace těchto částic a lze jej vyjádřit 1. Fickovým zákonem:



Obr. 1.20: Znázornění difúze elektronů a děr v makroskopickém měřítku

$$\mathbf{J} = -D \left(\frac{\partial c}{\partial x}, \frac{\partial c}{\partial y}, \frac{\partial c}{\partial z} \right) = -D \text{ grad } c \quad (1.49)$$

kde D je *difúzní koeficient*, konstanta úměrnosti mezi difúzním tokem a gradientem koncentrace. Její jednotkou je m^2s^{-1} , používá se i cm^2s^{-1} .

Jestliže tedy v polovodiči existuje nenulový gradient nosičů náboje ve směru osy x , způsobený např. rozdílnou koncentrací příměsí, pak z 1. Fickova zákona dostaneme vztahy pro difúzní proud elektronů a děr

$$J_{n,dif} = qD_n \frac{dn}{dx} \quad (1.50)$$

$$J_{p,dif} = -qD_p \frac{dp}{dx} \quad (1.51)$$

kde D_n a D_p jsou difúzní koeficienty elektronů a děr.

Je důležité si uvědomit rozdíly mezi driftovým a difúzním proudem. Jestliže koncentrační gradient elektronů i děr je kladný ve směru osy x (viz **Obr. 1.20**), pak elektrony i díry difundují ve směru $-x$. Také $J_{p,dif}$ teče ve směru $-x$, ovšem $J_{n,dif}$ teče ve směru $+x$. **Takže na rozdíl od driftového proudu, kdy proud elektronů a děr teče stejným směrem a**

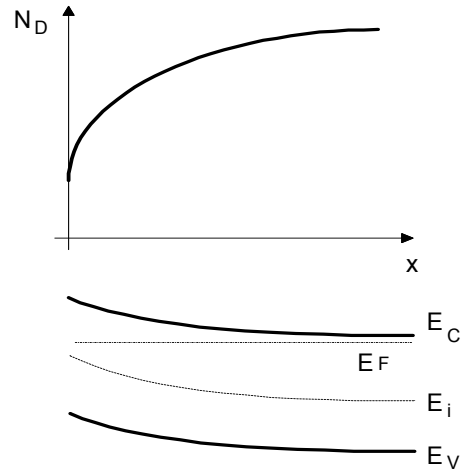
celkový driftový proud je součtem obou složek, v případě difúzního proudu je směr toku elektronů a děr opačný a celkový difúzní proud je jejich rozdílem. Celkový proud ve směru osy x je potom součtem difúzního i driftového proudu, tedy součtem čtyřech složek vyjádřených vztahy (1.42), (1.43) a (1.50), (1.51)

$$J_x = qE_x(\mu_p p + \mu_n n) + q(D_n \frac{dn}{dx} - D_p \frac{dp}{dx}) \quad (1.52)$$

Vztah (1.52) pro celkový proud je velmi důležitý; přímo nebo nepřímo jej používáme při analýze činnosti jakékoliv součástky.

Ve vztahu (1.52) se vyskytuje difúzní koeficient pro elektrony i pro díry. Ten můžeme nahradit pohyblivostí elektronů a děr použitím tzv. *Einsteinova vztahu*, který nyní odvodíme.

Jeden ze základních fyzikálních zákonů stanoví, že za rovnovážných podmínek není Fermiho energie uvnitř materiálu funkcí polohy. Jak vidíme na **Obr. 1.21**, E_F má pro všechna x stejnou hodnotu, tedy $dE_F/dx = 0$. Zároveň však z odstavce 1.5.3 víme, že vzdálenost Fermiho hladiny od E_C nebo E_V závisí na koncentraci nosičů. V polovodiči typu N se tedy Fermiho hladina bude se vzrůstající koncentrací N_D přibližovat ke dnu vodivostního pásu E_C - viz **Obr. 1.21**. Zakřivení energetických pásů je tedy přirozeným důsledkem měnící se koncentrace příměsí a dá se odvodit [3], že toto zakřivení vytvoří elektrické pole velikosti



Obr. 1.21: Nehomogenně dotovaný polovodič, koncentrace příměsí a jí odpovídající pásový diagram

$$E_x = \frac{1}{q} \frac{dE_C}{dx} = \frac{1}{q} \frac{dE_V}{dx} = \frac{1}{q} \frac{dE_i}{dx} \quad (1.53)$$

Vyjdeme ze vztahu pro rovnovážnou koncentraci nosičů (1.19), ze kterého po derivaci a uvědomíme-li si, že $dE_F/dx = 0$, dostaneme

$$\frac{dn}{dx} = -\frac{n_i}{kT} \exp[(E_F - E_i)/kT] \frac{dE_i}{dx} = -\frac{q}{kT} n E_x \quad (1.54)$$

Změna koncentrace nosičů tedy vytvořila elektrické pole, ale zároveň víme, že za rovnovážných podmínek polovodičem bez přiložení vnějšího zdroje neprotéká žádný proud, tedy že toto elektrické pole vytvoří driftový proud velikosti právě kompenzující difúzní proud

$$J_n = q\mu_n n E_x + qD_n \frac{dn}{dx} = 0 \quad (1.55)$$

Dosazením (1.54) do (1.55) dostaneme

$$qnE_x)\mu_n - (qnE_x) D_n q/kT = 0 \quad (1.56)$$

a protože $E_x \neq 0$

$$D_n/\mu_n = kT/q \quad (1.57)$$

To je *Einsteinův vztah* pro elektrony. Analogicky můžeme postupovat při odvození Einsteinova vztahu pro díry, takže dostáváme

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{q} = U_T \quad (1.58)$$

V předcházejících rovnicích tedy můžeme difúzní koeficient elektronů a děr nahradit jejich pohyblivostí, což je lépe měřitelná fyzikální veličina. I když jsme Einsteinův vztah odvodili za rovnovážných podmínek, platí i v podmínkách nerovnovážných. Výraz kT/q definuje tzv. *teplotní napětí* U_T . Jeho hodnota je pro 300 K přibližně rovna 0,026 V.

1.6.3 Generace a rekombinace

Generace je proces, při kterém je vytvořen *pár* elektron a díra. *Rekombinace* je proces zániku *páru* elektron-díra. Generace páru elektron-díra vyžaduje určité množství energie (tepelné, světelné apod.), při rekombinaci se určité množství energie uvolňuje. Protože všechny volné elektrony a díry podléhají možnosti rekombinace, existují jako volné nosiče jenom po určité časové intervaly. Definujeme proto *dobu života* elektronů τ_n a dobu života děr τ_p . U vlastního polovodiče $\tau_n = \tau_p$. U nevlastního polovodiče je doba života majoritních nosičů podstatně delší než doba života minoritních nosičů. Ve velmi čistém křemíku může být doba života nosičů až 1 ms, zvýšíme-li záměrně koncentraci rekombinačních center (např. legováním zlatem), pak může klesnout až na 1 ns, což bývá využíváno u vysokofrekvenčních součástek. Typické hodnoty doby života se pohybují kolem 1 μ s.

Nejčastěji se setkáváme s *generací tepelnou* a *světelnou*, existují však i jiné generační děje. Rozeznáváme tři druhy *rekombinace*:

1. *přímou* (mezipásovou) *rekombinaci*, tj. přímý přestup elektronu z vodivostního do valenčního pásu, elektron ztrácí ionizační energii náraz; podle typu uvolněné energie rozlišujeme
 - zářivou (fotonovou) rekombinaci,
 - nezářivou (fononovou) rekombinaci a
 - nárazovou neboli Augerovu [o:žerovu] rekombinaci - to je srážka tří nosičů, z nichž dva rekombinují a současně odevzdají energii třetímu, který přejde na vyšší energetickou hladinu;
2. *nepřímou* (postupnou) *rekombinaci*, která se uskutečňuje přes hladiny defektů nebo příměsí (nazýváme je rekombinační centra nebo pasti) uvnitř zakázaného pásu; opět může nastat rekombinace zářivá, nezářivá a nárazová;
3. *povrchovou rekombinaci*, což je nepřímá rekombinace, která nastane u povrchu polovodiče.

Označme nyní G a R množství nosičů, které je za jednotku času generuje a rekombinuje v jednotce objemu. V tepelné rovnováze je generace kompenzována rekombinací, takže střední koncentrace volných nosičů náboje je konstantní.

Pro rovnovážný stav, který označujeme indexem "0", platí tedy podmínka tepelné rovnováhy (1.3)

$$p_0 n_0 = n_i^2 \quad (1.59)$$

Působením vnějších vlivů (nemusí se jednat jenom o generaci) může být tato rovnováha porušena, takže platí

$$n = n_0 + \Delta n \quad (1.60)$$

$$p = p_0 + \Delta p \quad (1.61)$$

Přírůstek Δn a Δp , o který se skutečná koncentrace liší od rovnovážné koncentrace, se nazývá *koncentrace nerovnovážných nosičů*. Protože může současně působit více vnějších vlivů, může a nemusí platit $\Delta n = \Delta p$. Podle znaménka přírůstku hovoříme o *injekci nosičů* - nosiče jsou do polovodiče dodávány (vstříkovány) a platí $np > n_i^2$, nebo o *extrakci nosičů* - nosiče jsou z polovodiče odsávány a platí $np < n_i^2$. Změnu koncentrace nosičů vzhledem k rovnovážnému stavu (tj. koncentraci nerovnovážných nosičů) můžeme porovnat s rovnovážnými koncentracemi, takže například v případě injekce můžeme hovořit o *nízké, střední a vysoké úrovni injekce* (tj. např. $\Delta n \ll n_0$, $\Delta n \approx n_0$, $\Delta n \gg n_0$).

Je velmi důležité si uvědomit, že při malých úrovních injekce se podstatně zvyšuje pouze koncentrace minoritních nosičů, zatímco změna majoritních nosičů je zanedbatelná.

Příklad: Předpokládejme například křemík s koncentrací donorů $N_D = 10^{14} \text{ cm}^{-3}$ a injekci, při které $\Delta n = \Delta p = 10^9 \text{ cm}^{-3}$. Pro tento materiál $n_0 \approx N_D = 10^{14} \text{ cm}^{-3}$ a $p_0 \approx n_i^2/N_D \approx 10^6 \text{ cm}^{-3}$. Potom

$$n = n_0 + \Delta n \approx n_0 \approx 10^{14} \text{ cm}^{-3}$$

$$p = p_0 + \Delta p \approx \Delta p = 10^9 \text{ cm}^{-3}$$

Srovnáme-li součin np v rovnovážném a nerovnovážném stavu, vidíme, že

$$n_0 p_0 = n_i^2 = 10^{14} 10^6 \text{ cm}^{-3} = 10^{20} \text{ cm}^{-3}$$

$$np = 10^{14} 10^9 \text{ cm}^{-3} = 10^{23} \text{ cm}^{-3}$$

Závěr: Tepelná rovnováha je proto výrazně porušena minoritními nosiči.

Definujme $\left. \frac{\partial n}{\partial t} \right|_{G-R}$ a $\left. \frac{\partial p}{\partial t} \right|_{G-R}$ jako časovou změnu koncentrace nosičů vlivem *obou* dějů

- generace i rekombinace nosičů, je to tedy rozdíl $G - R$. Označme dále n_n a p_n koncentraci nosičů v polovodiči typu N a n_p a p_p koncentraci nosičů v polovodiči typu P. Potom pro výsledek *tepelné* generace a rekombinace minoritních nosičů platí [2]

$$\left. \frac{\partial n_p}{\partial t} \right|_{G-R} = -\frac{\Delta n_p}{\tau_n} \quad \text{pro elektrony v polovodiči typu P,} \quad (1.62)$$

$$\left. \frac{\partial p_n}{\partial t} \right|_{G-R} = -\frac{\Delta p_n}{\tau_p} \quad \text{pro díry v polovodiči typu N.} \quad (1.63)$$

Jestliže například $\Delta p_n < 0$, pak $\partial p_n / \partial t|_{G-R} > 0$. To znamená, že jestliže v polovodiči existuje "deficit" minoritních nosičů (např. vlivem jejich extrakce), pak to bude mít za následek zvýšenou rychlost generace nosičů, která převýší rychlost jejich rekombinace.

1.6.4 Stavové rovnice polovodiče

Dosud jsme se zabývali jednotlivými mechanismy pohybu nosičů proudu jednotlivě. Ve skutečnosti však téměř vždy nastává několik dějů současně, což popisují *stavové rovnice polovodiče* (rovnice kontinuity, dvě difúzní rovnice a Poissonova rovnice).

Rovnice kontinuity

Celková změna koncentrace nosičů náboje za jednotku času je vyjádřena vztahy

$$\frac{\partial n}{\partial t} = \left. \frac{\partial n}{\partial t} \right|_{drift} + \left. \frac{\partial n}{\partial t} \right|_{dif} + \left. \frac{\partial n}{\partial t} \right|_{tepelná\ G-R} + \left. \frac{\partial n}{\partial t} \right|_{ostatní\ procesy} \quad (1.64)$$

$$\frac{\partial p}{\partial t} = \left. \frac{\partial p}{\partial t} \right|_{drift} + \left. \frac{\partial p}{\partial t} \right|_{dif} + \left. \frac{\partial p}{\partial t} \right|_{tepelná\ G-R} + \left. \frac{\partial p}{\partial t} \right|_{ostatní\ procesy} \quad (1.65)$$

Rovnice kontinuity vyjadřuje podmínku dynamické rovnováhy pohyblivých nábojů v elementárním objemu a má obecný tvar [50]

$$\frac{\partial \varrho}{\partial t} = -\operatorname{div} \mathbf{J} = -\left(\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} \right) \quad (1.66)$$

Divergence má význam výtoku vektoru \mathbf{J} z elementárního objemu vztažený na tento elementární objem. Uvážíme-li dále, že hustota pohyblivého náboje $\varrho = q(p - n)$ a použijeme-li obecnou rovnici kontinuity (1.66) na difúzní a driftovou složku proudu, dostaneme

$$\frac{\partial n}{\partial t} = \frac{1}{q} \operatorname{div} J_n + (G - R) \quad (1.67)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \operatorname{div} J_p + (G - R) \quad (1.68)$$

kde

$$\mathbf{J}_n = q\mu_n n \mathbf{E} + qD_n \operatorname{grad} n \quad (1.69)$$

$$\mathbf{J}_p = q\mu_p p \mathbf{E} - qD_p \operatorname{grad} p \quad (1.70)$$

a $(G - R)$ je výsledná změna náboje způsobená všemi generačními a rekombinačními procesy.

Difúzní rovnice pro minoritní nosiče

Zavedením určitých zjednodušujících podmínek dostaneme tuto rovnici z rovnice kontinuity. Tyto zjednodušující podmínky jsou následující: omezíme naši analýzu pouze na minoritní nosiče, předpokládáme jednorozměrný systém, neexistenci "ostatních procesů" v (1.64) a (1.65), rovnovážnou koncentraci nosičů nezávislou na poloze ($n_0 = \text{konst.}$, $p_0 = \text{konst.}$) a neexistenci elektrického pole ($\mathbf{E} = \mathbf{0}$). Potom z (1.69) dostaneme

$$\mathbf{J}_n = qD_n \partial n / \partial x \quad \text{a} \quad \partial n / \partial x = \partial n_0 / \partial x + \partial \Delta n / \partial x = \partial \Delta n / \partial x.$$

Použijeme-li dále rovnici (1.62) pro vyjádření tepelných generačně-rekombinačních dějů a uvědomíme-li si, že rovnovážná koncentrace nosičů není *nikdy* funkcí času, tedy

$$\partial n / \partial t = \partial n_0 / \partial t + \partial \Delta n / \partial t = \partial \Delta n / \partial t,$$

pak pro minoritní nosiče z rovnic kontinuity (1.67) a (1.68) dostaneme *difúzní rovnice*, které jsou variantou 2. Fickova zákona

$$\frac{\partial \Delta n_p}{\partial t} = D_n \frac{\partial^2 \Delta n_p}{\partial x^2} - \frac{\Delta n_p}{\tau_n} \quad (1.71)$$

$$\frac{\partial \Delta p_n}{\partial t} = D_p \frac{\partial^2 \Delta p_n}{\partial x^2} - \frac{\Delta p_n}{\tau_p} \quad (1.72)$$

Poissonova rovnice

Poslední stavovou rovnicí polovodiče je *Poissonova rovnice*, jejíž obecný tvar je

$$\operatorname{div} \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \frac{\rho}{\varepsilon} = \frac{q}{\varepsilon} (p - n + N_D - N_A) \quad (1.73)$$

Tuto rovnici je možné použít i ke stanovení průběhu potenciálu v polovodiči. Pro jednorozměrný případ můžeme psát $\mathbf{E} = -\operatorname{grad} V = -dV/dx$ a Poissonovu rovnici dostaneme ve tvaru

$$\frac{dE}{dx} = -\frac{d^2 V}{dx^2} = \frac{\rho}{\varepsilon} \quad (1.74)$$

2 POLOVODIČOVÉ PŘECHODY

Pro pochopení činnosti moderních polovodičových součástek je nezbytné poznat vlastnosti jejich základní stavební jednotky - polovodičového přechodu.

2.1 Klasifikace přechodů

Ve struktuře polovodičových součástek můžeme najít celou řadu různých rozhraní mezi jednotlivými materiály, z nichž je součástka vyrobena. Z hlediska činnosti a vlastností součástky jsou nejvýznamnější *polovodičové přechody*, tj. rozhraní mezi různě dotovanými polovodiči, mezi různými typy polovodičů nebo mezi polovodičem a jiným materiálem. V teorii polovodičů se tato problematika zahrnuje mezi *kontaktní jevy*.

2.1.1 Homogenní pøechody

Homogenní pøechody jsou pøechody uvnitř stejnorodých materiálů (např. Si, GaAs, Ge apod.). Jsou vytvořeny různou dotací stejného materiálu příměsími. V energetickém pásovém diagramu je šířka zakázaného pásu na obou stranách tohoto pøechodu stejná. Můžeme je rozdělit na:

1. *Pøechody PN, NP* - pøechody mezi stejnorodými materiály z různým typem vodivosti. Tyto pøechody můžeme podle poměru koncentrace příměsí na obou

stranách přechodu dále rozdělit na *souměrné*, kde koncentrace donorů na straně N a akceptorů na straně P je přibližně stejná ($N_D \approx N_A$), a *nesouměrné*, kde je dotace příměsí na jedné straně přechodu výrazně vyšší.

2. *Přechody N^+N , P^+P* - přechody mezi stejnorodými materiály stejného typu vodivosti s různou koncentrací příměsí.
3. *Přechody PI , NI a jejich kombinace PIN* - přechody mezi stejnorodými materiály s nevlastní a vlastní vodivostí.

Podle průběhu koncentrace příměsí můžeme tyto přechody rozdělit na *strmé* (stupňovité), kde ke změně koncentrace dochází skokově ($dN/dx \rightarrow \infty$) a přechody *plynulé* s konečným spádem koncentrace, jejichž zvláštním případem jsou přechody *lineární*.

2.1.2 Heterogenní přechody

Heterogenní přechody (heteropřechody) jsou přechody mezi dvěma nestejnorodými materiály s různou krystalovou strukturou. V energetickém pásovém diagramu je šířka zakázaného pásu na obou stranách přechodu různá. Rozeznáváme tyto typy heteropřechodů:

1. *Přechody mezi dvěma nestejnorodými polovodiči* (např. moderní materiály na bázi různých kompozitů GaAs, GaP apod.). Podle dalších vlastností daných zejména koncentrací rekombinačních center v přechodové vrstvě, která souvisí s rozdíly ve velikosti mřížkové konstanty obou materiálů, můžeme tyto přechody dále rozdělit na přechody
 - s malými rozdíly v krystalové struktuře,
 - s velkými rozdíly v krystalové struktuře.
2. *Přechod kov-polovodič (MS , MP , MN)* - podle velikosti výstupních prací obou materiálů může mít tento přechod jak usměrňující, tak i neusměrňující vlastnosti.
3. *Přechody kov-izolant-polovodič (MIS)* [metal-insulator-semiconductor] - nacházíme u unipolárních tranzistorů, jeho vlastnosti jsou dosti odlišné od všech výše uvedených.

2.2 Přechod PN v rovnovážném stavu

Vlastnosti PN přechodu budeme zkoumat za následujících zjednodušujících předpokladů:

1. *Struktura PN přechodu je jednorozměrná*, tj. koncentrace příměsí se mění pouze ve směru osy x , v ostatních směrech je stále konstantní; zanedbáme také veškeré okrajové efekty.
2. *Metalurgický přechod leží v $x = 0$* . Metalurgickým přechodem nazýváme plochu, na které jsou si efektivní koncentrace příměsí ($N = |N_A - N_D|$) na obou stranách přechodu rovny. U strmého PN přechodu je to plocha, na které dochází ke skokové změně koncentrace.
3. *Přechod je strmý s homogenně dotovanou P a N oblastí*.
4. *Kontakty jsou dokonale ohmické, dostatečně vzdálené od metalurgického přechodu*, takže neovlivňují vlastnosti PN přechodu.

2.2.1 Kvalitativní popis PN přechodu v rovnovážném stavu

Mějme nyní krystal křemíku, jehož jedna část je dotovaná donory (např. As) a druhá akceptory (např. Ga). Na obr. 2.1.a) jsou schématicky znázorněny nepohyblivé ionty akceptorů a donorů a pohyblivé volné elektrony a díry. Na obr. 2.1.b) je nakreslen průběh koncentrací donorů a akceptorů takovéto struktury a průběh koncentrací volných nosičů

(všimněme si jejich značení n_n ,

n_p, p_p, p_n). V bodě x_i platí, že $n = p = n_i$ a $E_F = E_i$. Protože v místě

metalurgického přechodu ($x = 0$) dochází k prudké změně koncentrace příměsí a tím i k prudké změně koncentrace volných nosičů, difundují volné

díry na straně P k místu metalurgického přechodu a zanechávají za sebou záporný prostorový náboj a volné elektrony na straně N difundují k metalurgickému přechodu a zanechávají za sebou kladný

prostorový náboj (viz obr. 2.1.c). V oblasti přechodu volné

elektrony

a volné díry rekombinují.

Zároveň kladný prostorový náboj na straně N odpuzuje kladné volné díry difundující ze strany P a záporný prostorový náboj ze strany P odpuzuje záporné elektrony difundující ze strany N.

Driftové síly elektrického pole vytvořeného

prostorovým nábojem (viz obr. 2.1.d), které je kolmé na plochu metalurgického přechodu, tak působí proti

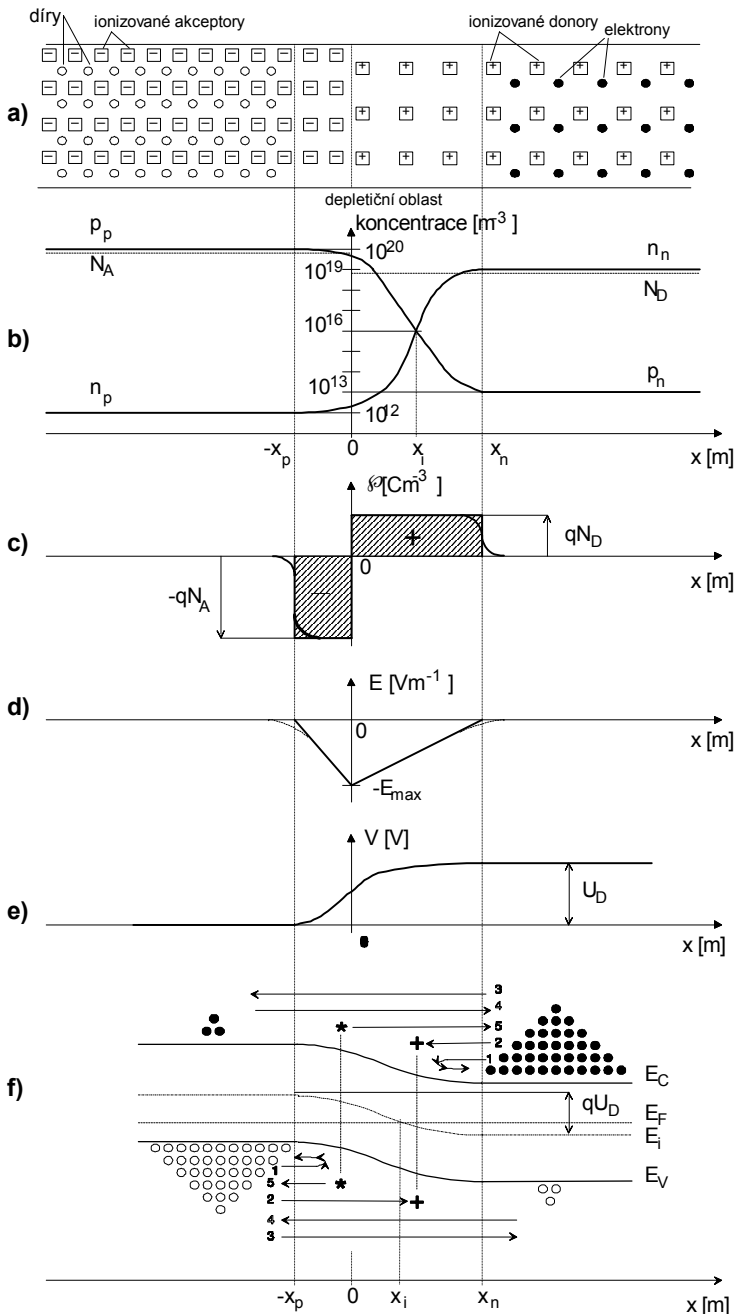
difúznímu pohybu a v oblasti přechodu je nastolena rovnováha

difúzních a driftových sil.

V důsledku existence elektrického pole vznikne v oblasti přechodu rozdíl potenciálů - tzv. *kontaktní potenciál* (viz obr. 2.1.e), často nazývaný také

difúzní napětí [built-in voltage], protože příčinou jeho vzniku je

difúzní proud. Značíme jej U_D . Opatříme-li oblast typu N i P



Obr. 2.1. Nesouměrný stupňovitý PN přechod: **a)** schématické znázornění; **b)** průběh koncentrací; **c)** prostorový náboj; **d)** intenzita elektrického pole; **e)** potenciál; **f)** pásový diagram. Skutečné průběhy prostorového náboje a intenzity elektrického pole jsou nakresleny přerušovanou čarou, zjednodušení vyplývající z depleční aproximace plnou čarou.

polovodiče např. kovovými kontakty, vznikne vlivem rozdílné koncentrace elektronů v kovu a v polovodiči difúzní napětí i na přechodech mezi kovovými kontakty a oběma oblastmi polovodiče. Protože se tato difúzní napětí na jednotlivých přechodech vzájemně ruší, mezi kovovými kontakty nenaměříme žádné napětí a PN přechod nebude dodávat do obvodu žádný proud.

V oblasti přechodu PN (mezi souřadnicemi $-x_p$ a x_n na obr. 2.1.) je ve srovnání s oblastmi typu P a N velmi malá koncentrace volných nosičů a existuje v ní prostorový náboj. Proto je také nazývána *depletiční* (ochuzená, vyprázdněná) *vrstva* [depletion region] nebo *oblast prostorového náboje* [space charge region]. V oblastech typu P a N je koncentrace volných nosičů

a prostorového náboje téměř stejná, hovoříme proto o *kvazineutrálních oblastech*.

Energetický pásový diagram PN přechodu vidíme na obr. 2.1.f). Všimněme si, že rozdíl mezi dnem vodivostního pásu (resp. stropem valenčního pásu) v oblastech P a N je roven qU_D . Na obr. 2.1f) jsou také znázorněny možné přechody nosičů z kvazineutrálních oblastí:

1. Majoritní nosiče nemají dostatečnou energii k překonání potenciálové bariéry qU_D a vrátí se zpět.
2. Majoritní nosiče vstoupí do závěrné vrstvy a rekombinují v ní.
3. Majoritní nosiče mají dostatečnou energii pro překonání potenciálové bariéry qU_D , přejdou přes závěrnou vrstvu do opačné oblasti, kde se z nich stanou minoritní nosiče (*injekce minoritních nosičů*).
4. Minoritní nosiče jsou urychleny potenciálovou bariérou qU_D a jsou injikovány do opačné oblasti jako majoritní nosiče (*extrakce minoritních nosičů*).
5. V závěrné vrstvě se generuje pár elektron-díra, který je elektrickým polem odstraněn z depletiční vrstvy a injikován jako majoritní nosič do kvazineutrálních oblastí.

Elektrické pole v depletiční vrstvě (a spád potenciálu) brání přechodu majoritních nosičů a naopak umožňuje přechod minoritních nosičů. Naopak difúzní síly (spád koncentrace) brání přechodu minoritních nosičů a umožňují přechod majoritních nosičů. Z rozložení nosičů, naznačeného na obr. 2.1.f) je zřejmé, že **proud majoritních nosičů je závislý na výšce potenciálové bariéry, kdežto proud minoritních nosičů na ní nezávisí**. Kolik nosičů jednoho typu přejde přechodem na jednu stranu, tolik nosičů opačného znaménka přejde přechodem na stranu druhou. Přechod je v termodynamické rovnováze, takže Fermiho hladina leží v kvazineutrálních oblastech i v depletiční oblasti na stejné úrovni.

2.2.2 Kvantitativní popis PN přechodu v rovnovážném stavu

Uvažujme strukturu PN přechodu zapojenou do jednoduchého obvodu podle obr. 2.2. V rovnovážném stavu není na PN přechod přiloženo žádné napětí ($U = 0$ V). Je zřejmé, že obvodem nebude protékat žádný proud, takže pro proudovou hustotu platí

$$\mathbf{J} = \mathbf{J}_n = \mathbf{J}_p = 0 \quad (2.1)$$

Pro hustotu proudu elektronů a děr pak z rovnice (1.41) dostaneme

$$\mathbf{J}_n = \mathbf{J}_{n,drift} + \mathbf{J}_{n,dif} = q\mu_n n \mathbf{E} + qD_n \frac{dn}{dx} = 0 \quad (2.2a)$$

$$\mathbf{J}_p = \mathbf{J}_{p,drift} + \mathbf{J}_{p,dif} = q\mu_p p \mathbf{E} - qD_p \frac{dp}{dx} = 0 \quad (2.2b)$$

Všimněme si, že driftový proud musí být s opačným znaménkem rovný proudu difúznímu, abychom dostali celkovou hustotu proudu rovnou nule v rovnicích (2.2). Řešením rovnice (2.2a) pro elektrické pole po dosazení Einsteinova vztahu (1.47) dostaneme

$$\mathbf{E} = \left(-\frac{qD_n}{q\mu_n} \right) \left(\frac{dn}{dx} \right) = -\left(\frac{D_n}{\mu_n} \right) \left(\frac{1}{n} \right) \left(\frac{dn}{dx} \right) = -\left(\frac{kT}{q} \right) \left(\frac{1}{n} \right) \left(\frac{dn}{dx} \right) \quad (2.3)$$

Protože napětí mezi dvěma body získáme jako integrál elektrického pole mezi těmito dvěma body, můžeme napětí mezi oběma konci (ohmickými kontakty, ležícími podle našich zjednodušujících předpokladů v nekonečnu) PN přechodu vypočítat jako

$$U_D = - \int_{-\infty}^{+\infty} \mathbf{E} dx = \frac{kT}{q} \int_{-\infty}^{+\infty} \left(\frac{1}{n} \right) \left(\frac{dn}{dx} \right) dx = \frac{kT}{q} \int_{n(-\infty)}^{n(+\infty)} \frac{dn}{n} = \frac{kT}{q} [\ln n]_{n(-\infty)}^{n(+\infty)} \quad (2.4)$$

Zbývá už jenom dosadit meze integrálu do vztahu (2.4). Jak víme, v dostatečné vzdálenosti od přechodu platí

$$n(-\infty) = n_p = n_i^2 / N_A \quad (2.5a)$$

$$n(+\infty) = n_n = N_D \quad (2.5b)$$

Pro **difúzní napětí** (kontaktní potenciál) U_D tak dostaneme vztah

$$U_D = \frac{kT}{q} (\ln n_n - \ln n_p) = \frac{kT}{q} \ln \left(\frac{n_n}{n_p} \right) \quad (2.6)$$

který můžeme s použitím (2.5) přepsat do často užívaného tvaru

$$U_D = \frac{kT}{q} \ln \left(\frac{N_D N_A}{n_i^2} \right) \quad (2.7)$$

Příklad: Při pokojové teplotě ($T = 300 \text{ K}$) je teplotní napětí $U_T = kT/q \cong 0,026 \text{ V}$. Mějme PN přechod vytvořený v křemíku ($n_i = 10^{16} \text{ m}^{-3} = 10^{10} \text{ cm}^{-3}$) tak, že oblast P je dotovaná akceptory o koncentraci

$N_A = 10^{21} \text{ m}^{-3} = 10^{15} \text{ cm}^{-3}$ a oblast N donory o koncentraci $N_D = 10^{21} \text{ m}^{-3} = 10^{15} \text{ cm}^{-3}$. Pak po dosazení do (2.7) dostaneme pro difúzní napětí

$$U_D = 0,026 \cdot \ln \left(\frac{10^{15} \cdot 10^{15}}{10^{20}} \right) = 0,599 \text{ V}$$

Abychom mohli kvantitativně popsat vlastnosti PN přechodu v rovnovážném stavu, to znamená spočítat hustotu prostorového náboje \wp , intenzitu elektrického pole $\mathbf{E}(x)$ a elektrický potenciál $V(x)$, musíme kromě výše uvedených zjednodušujících předpokladů zavést ještě další, který nazýváme **depletiční aproximace**:

1. pro $-x_p \leq x \leq 0$ $N_A \gg n_p$ nebo $p_p \Rightarrow \wp = -qN_A$
2. pro $0 \leq x \leq x_n$ $N_D \gg n_n$ nebo $p_n \Rightarrow \wp = qN_D$
3. pro $x < -x_p$ a $x > x_n$ $\wp = 0$

Depletiční aproximace tedy předpokládá, že v oblasti mezi body $-x_p$ a x_n nejsou volné nosiče náboje, zatímco mimo tuto oblast (v kvazineutrálních oblastech) je prostorový náboj plně kompenzován volnými nosiči.

Depletiční aproximace nám nyní umožní vyřešit Poissonovu rovnici a spočítat **intenzitu elektrického pole** v PN přechodu. Vyjdeme-li z Poissonovy rovnice pro jednorozměrný případ (1.57)

$$\frac{dE}{dx} = \frac{\rho}{\epsilon}$$

pak po dosazení podmínek depletiční aproximace dostaneme

$$\frac{dE}{dx} = -\frac{q}{\epsilon} N_A \quad \text{pro } -x_p \leq x \leq 0 \quad (2.8a)$$

$$\frac{dE}{dx} = \frac{q}{\epsilon} N_D \quad \text{pro } 0 \leq x \leq x_n \quad (2.8b)$$

$$E = 0 \quad \text{pro } x \leq -x_p \text{ a } x \geq x_n \quad (2.8c)$$

Integrujeme-li rovnice (2.8), pak

$$E(x) = \int \frac{-qN_A}{\epsilon} dx = \frac{-qN_A}{\epsilon} x + C_1 \quad \text{pro } -x_p \leq x \leq 0 \quad (2.9a)$$

$$E(x) = \int \frac{qN_D}{\epsilon} dx = \frac{qN_D}{\epsilon} x + C_2 \quad \text{pro } 0 \leq x \leq x_n \quad (2.9b)$$

Po dosazení okrajových podmínek $E(-x_p) = E(x_n) = 0$, které plynou z rovnice (2.8c), dostaneme pro integrační konstanty

$$C_1 = x_p(-qN_A)/\epsilon \quad (2.10a)$$

$$C_2 = -x_n(qN_D)/\epsilon \quad (2.10b)$$

a pro intenzitu elektrického pole v depletiční oblasti

$$E(x) = -\frac{qN_A}{\epsilon} (x_p + x) \quad \text{pro } -x_p \leq x \leq 0 \quad (2.11a)$$

$$E(x) = -\frac{qN_D}{\epsilon} (x_n - x) \quad \text{pro } 0 \leq x \leq x_n \quad (2.11b)$$

Z podmínky spojitosti elektrického pole pro $x = 0$ plyne

$$-\frac{qN_A}{\epsilon} x_p = -\frac{qN_D}{\epsilon} x_n \quad (2.12)$$

To je vlastně *podmínka nábojové neutrality*, která stanovuje, že velikost prostorového náboje na straně P se musí v rovnovážném stavu s opačným znaménkem rovnat velikosti prostorového náboje na straně N, tedy $-\rho_p = \rho_n$. Rovnici (2.12) můžeme přepsat do tvaru

$$\boxed{\frac{x_p}{x_n} = \frac{N_D}{N_A}} \quad (2.13)$$

Z podmínky nábojové neutrality (2.13) je zřejmé, že **depletiční vrstva se** u nesouměrného přechodu **rozšíří více do méně dotované oblasti**. Pro tuto oblast se užívá termín *vysokoohmová oblast* nebo v případě diody *báze diody*.

Nyní můžeme stanovit **průběh potenciálu** v PN přechodu. Vyjdeme ze vztahu mezi potenciálem a elektrickým polem

$$\frac{dV}{dx} = -E \quad (2.14)$$

a z Poissonovy rovnice (1.57)

$$-\frac{d^2V}{dx^2} = \frac{dE}{dx} = \frac{q\varphi}{\varepsilon}$$

Integrací vztahů (2.11) dostaneme

$$V(x) = -\int_{-x}^x E(x)dx = \frac{qN_A}{\varepsilon} \int_{-x_p}^x (x_p + x)dx = \frac{qN_A}{2\varepsilon} (x_p + x)^2 + C_3 \quad \text{pro } -x_p \leq x \leq 0 \quad (2.15a)$$

$$V(x) = -\int_{x_n}^x E(x)dx = \frac{qN_D}{\varepsilon} \int_{x_n}^x (x_n - x)dx = -\frac{qN_D}{2\varepsilon} (x_n - x)^2 + C_4 \quad \text{pro } 0 \leq x \leq x_n \quad (2.15b)$$

Pro určení integračních konstant C_3 a C_4 je nutné stanovit okrajové podmínky. Odvodili jsme, že rozdíl potenciálu mezi oběma konci PN přechodu je roven difúznímu napětí U_D (2.7). Dále víme, že v kvazineutrálních oblastech je $E = 0$, takže tam nedochází k žádné změně potenciálu. Celé difúzní napětí U_D je proto mezi body $-x_p$ a x_n . Jestliže tedy zvolíme $V(-x_p) = 0$, pak $V(x_n) = U_D$ a z (2.15) dostáváme vztahy pro průběh potenciálu v depletiční oblasti

$$\begin{array}{ll} V(x) = \frac{qN_A}{2\varepsilon} (x_p + x)^2 & \text{pro } -x_p \leq x \\ \leq 0 & \\ V(x) = -\frac{qN_D}{2\varepsilon} (x_n - x)^2 + U_D & \text{pro } 0 \leq x \\ \leq x_n & \end{array} \quad \begin{array}{l} (2.16a) \\ (2.16b) \end{array}$$

Nyní stanovíme **šířku depletiční**

oblasti. V bodě $x = 0$ musí platit (spojitost potenciálu)

$$\frac{qN_A}{2\varepsilon} (x_p)^2 = \frac{qN_D}{2\varepsilon} (x_n)^2 + U_D \quad (2.17)$$

Po dosazení podmínky nábojové neutrality (2.13) ve tvaru

$$x_p = \frac{N_D}{N_A} x_n$$

do rovnice (2.17) dostaneme kvadratickou rovnici s neznámou x_n

$$\frac{qN_A}{2\varepsilon} \left(\frac{N_D}{N_A} x_n \right)^2 = \frac{qN_D^2 x_n^2}{2\varepsilon N_A} = \frac{qN_D}{2\varepsilon} (x_n)^2 + U_D \quad (2.18)$$

jejíž řešení je

$$x_n = \left[\frac{2\varepsilon U_D}{q} \cdot \frac{N_A}{N_D(N_A + N_D)} \right]^{1/2} \quad (2.19a)$$

Dosadíme-li x_n zpět do (2.13), dostáváme obdobný vztah pro x_p

$$x_p = \left[\frac{2\varepsilon U_D}{q} \cdot \frac{N_D}{N_A(N_A + N_D)} \right]^{1/2} \quad (2.19b)$$

Celková šířka depletiční oblasti $w = x_p + x_n$. Použitím vztahů (2.19) dostaneme

$$w = \left[\frac{2\varepsilon U_D}{q} \cdot \frac{(N_A + N_D)}{N_A N_D} \right]^{1/2} \quad (2.20)$$

V případě nesouměrného přechodu P⁺N, kde $N_A \gg N_D$ můžeme vztah (2.20) zjednodušit

$$w \approx x_n \approx \left[\frac{2\varepsilon U_D}{q} \cdot \frac{1}{N_D} \right]^{1/2} \quad (2.21)$$

2.3 Přechod PN s přiloženým vnějším napětím

Nyní budeme předpokládat, že na PN přechod je přiloženo nějaké nenulové napětí U . Kromě výše použitých zjednodušujících předpokladů zavedeme ještě jeden: zanedbáme úbytek napětí v kvazineutrálních oblastech, takže celé napětí zdroje U bude mezi hranicemi depletiční oblasti $-x_p$ a x_n .

2.3.1 Propustně a závěrně pólovaný PN přechod

Na obr. 2.2. je nakreslený vliv přiloženého napětí na PN přechod. Vzhledem k polaritě vnějšího napětí mohou nastat dva případy. Pokud napětí potenciálovou bariéru snižuje, označujeme je jako *napětí v propustném směru* (napětí v přímém směru). Jeho znaménko je kladné, někdy se značí U_F [forward voltage]. Pokud vnější napětí potenciálovou bariéru zvyšuje, označujeme je jako *napětí v závěrném směru* (napětí ve zpětném směru). Jeho znaménko je záporné, někdy se značí U_R [reverse voltage]. V propustném směru je polarita vnějšího napětí taková, že na oblasti P je kladné napětí, na oblasti N záporné.

Potenciálový rozdíl mezi hranicemi depletiční oblasti $-x_p$ a x_n je nyní dán rozdílem difúzního napětí U_D a vnějšího napětí U , $U_D - U$ (při respektování výše uvedené znaménkové konvence). Pro šířku depletiční oblasti, průběh potenciálu a intenzitu elektrického pole budou proto platit rovnice (2.19), (2.16) a (2.11), ve kterých budeme místo U_D psát $U_D - U$.

Pro depletiční oblast na straně P, tedy pro $-x_p \leq x \leq 0$, bude platit

$$x_p = \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_D}{N_A(N_A + N_D)} \right]^{1/2} \quad (2.22a)$$

$$V(x) = \frac{qN_A}{2\varepsilon} (x_p + x)^2 \quad (2.23a)$$

$$E(x) = -\frac{qN_A}{\varepsilon} (x_p + x) \quad (2.24a)$$

Pro depletiční oblast na straně N, tedy pro $0 \leq x \leq x_n$, bude platit

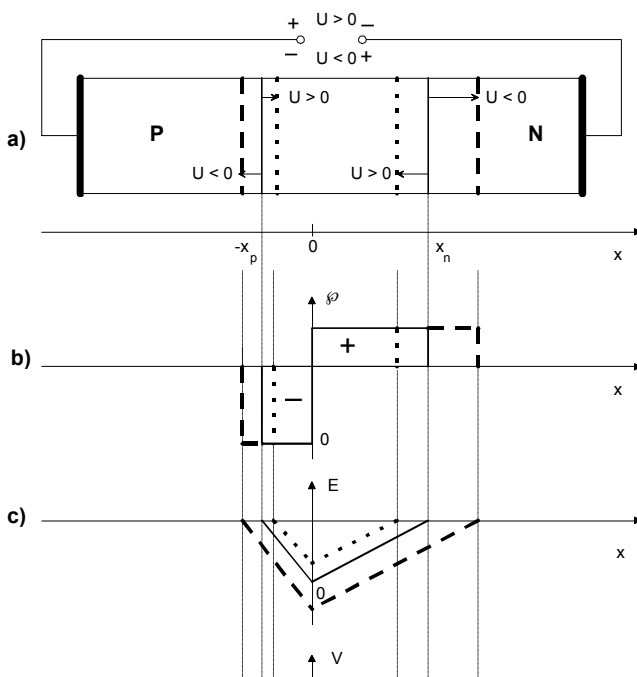
$$x_n = \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_A}{N_D(N_A + N_D)} \right]^{1/2} \quad (2.22b)$$

$$V(x) = (U_D - U) - \frac{qN_D}{2\varepsilon} (x_n - x)^2 \quad (2.23b)$$

$$E(x) = -\frac{qN_D}{\varepsilon} (x_n - x) \quad (2.24b)$$

Celková šířka depletiční oblasti je

$$w = \left[\frac{2\varepsilon}{q} (U_D - U) \frac{(N_A + N_D)}{N_A N_D} \right]^{1/2} \quad (2.25)$$



Obr. 2.2. Nesymetrický stupňovitý PN přechod s přiloženým vnějším napětím; **a)** schématické znázornění; **b)** prostorový náboj; **c)** intenzita elektrického pole; **d)** potenciál.

Ilustrace vztahů (2.22) až (2.25) je na obr. 2.2. Vidíme, že při přiložení závěrného napětí se zvyšuje výška potenciálové bariéry, šířka depletiční oblasti

a velikost elektrického pole v depletiční oblasti. Při přiložení kladného napětí se snižuje výška potenciálové bariéry, šířka depletiční oblasti a velikost elektrického pole v depletiční oblasti.

2.3.2 Ampérvoltová charakteristika ideálního PN přechodu

Všimněme si teď pohybu nosičů náboje přes přechod v případě přiloženého vnějšího propustného nebo závěrného napětí. V obou těchto případech není přechod PN v termodynamické rovnováze. **Změna výšky potenciálové bariéry bude ovlivňovat velikost proudu**

majoritních nosičů přechodem, zatímco opačný proud minoritních nosičů zůstává konstantní.

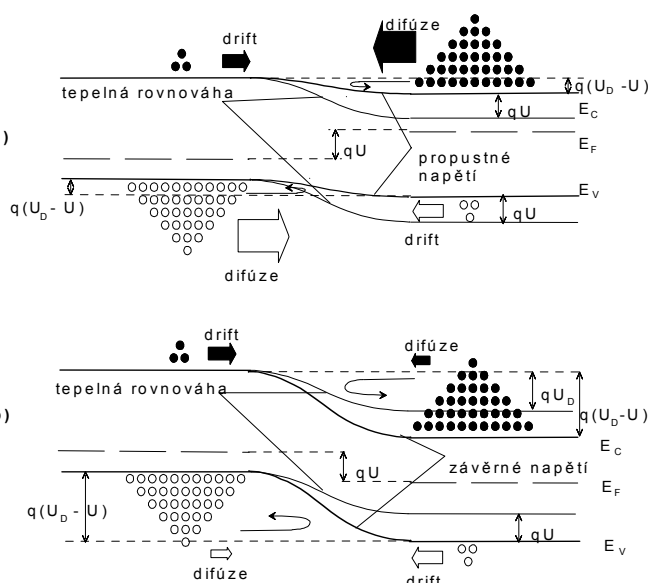
Propustně pólovaný PN přechod:

Energetický pásový diagram propustně pólovaného PN přechodu je na obr. 2.3.a). Při snížení potenciálové bariéry má více majoritních elektronů a děr větší energii než je výška potenciálové bariéry $q(U_D - U)$, takže více elektronů a děr může difundovat přes přechod, i když gradient koncentrace nosičů zůstává téměř stejný. Tím se zvětší difúzní proud majoritních nosičů, zatímco opačný driftový proud minoritních nosičů, který nezávisí na výšce potenciálové bariéry, ale pouze na množství minoritních nosičů n_p a p_n , zůstává stejný. Majoritní nosiče, které prošly přechodem, jsou vstřikovány (injektovány) do kvazineutrálních oblastí na obou stranách depleční vrstvy, kde zvyšují koncentraci minoritních nosičů (*injekce minoritních nosičů*). Protože rozložení nosičů je se zvyšující se energií téměř exponenciální (viz odstavec 1.4.), můžeme předpokládat, že se difúzní proud majoritních nosičů bude exponenciálně zvyšovat se snižující se potenciálovou bariérou. Propustný proud bude tedy záviset exponenciálně na přiloženém napětí U .

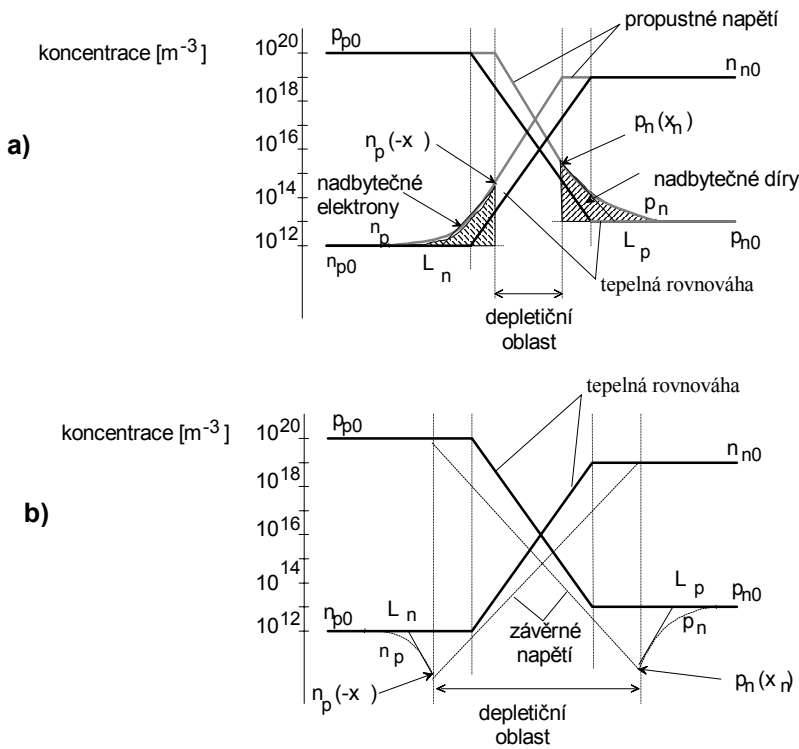
Závěrně pólovaný PN přechod:

Energetický pásový diagram závěrně pólovaného PN přechodu je na obr. 2.3.b). Při zvýšení potenciálové bariéry se sníží počet majoritních elektronů a děr, které mají větší energii než je výška potenciálové bariéry $q(U_D - U)$, což má za následek snížení difúzního proudu majoritních nosičů. Protože driftový proud minoritních nosičů je nezávislý na výšce potenciálové bariéry a zůstává stejný (a difúzní proud majoritních nosičů je velmi malý), výsledkem bude malý proud opačného směru, než v případě propustně pólovaného přechodu. V kvazineutrálních oblastech dojde ke snížení koncentrace minoritních nosičů (*extrakce minoritních nosičů*). Velikost závěrného proudu je v podstatě dána množstvím tepelně generovaných minoritních nosičů n_p a p_n , které je nezávislé na velikosti přiloženého napětí. Tento proud zůstává konstantní pro přiložené napětí U v rozsahu několika desítek voltů.

Průběh koncentrací minoritních a majoritních nosičů při propustně a závěrně pólovaném přechodu je na obr. 2.4.a) a 2.4.b).



Obr. 2.3. Energetický pásový diagram a) propustně a b) závěrně pólovaného PN přechodu



Obr. 2.4. Průběh koncentrací minoritních a majoritních nosičů při **a)** propustně a **b)** závěrně pólovaném PN přechodu

Za určitých obecných předpokladů je možné odvodit **rovnici ideální diody** [2], [3]. Tyto obecné předpoklady jsou:

1. neexistuje žádný externí zdroj generace nosičů (např. světlo),
2. uvažujeme strmý přechod a platnost depletiční aproximace,
3. neuvažujeme generaci a rekombinaci v depletiční oblasti,
4. uvažujeme režim nízké injekce,
5. elektrické pole je nulové mimo depletiční oblast,
6. N_A a N_D jsou konstantní.

Podrobné odvození rovnice ideální diody zde nebudeme rozebírat, uvedme jenom, že postupujeme podle následujícího schématu:

1. řešíme stavové rovnice polovodiče pro minoritní nosiče,
2. spočítáme okrajové podmínky pro minoritní nosiče na hranicích depletiční oblasti,
3. určíme proud elektronů a proud děr - celkový proud je jejich součtem.

Výsledek vyjadřuje rovnice

$$I = q A \left(\frac{D_n}{L_n} n_{p0} + \frac{D_p}{L_p} p_{n0} \right) \left[\exp\left(\frac{qU}{kT}\right) - 1 \right] \quad (2.26)$$

kde A ... plocha přechodu [m^2],

D_n, D_p ... difúzní koeficient, definovaný v (1.47) jako $D_n = \mu_n kT/q$, $D_p = \mu_p kT/q$ [m^2s^{-1}],

L_n, L_p ... střední difúzní délka minoritních nosičů [m].

Střední difúzní délka minoritních elektronů L_n je průměrná dráha, kterou minoritní elektrony s dobou

života τ_n difundují v polovodiči typu P než zaniknou rekombinací (viz obr. 2.4.). Je definována jako

$$L_n = \sqrt{D_n \tau_n} \quad (2.27a)$$

Podobně je definována i *střední difúzní délka minoritních děr*

$$L_p = \sqrt{D_p \tau_p} \quad (2.27b)$$

Pro výrazy D_n/L_n a D_p/L_p plyne z (1.47) a (2.27)

$$\frac{D_n}{L_n} = \sqrt{\frac{D_n}{\tau_n}} = \sqrt{\frac{\mu_n k T}{q} \frac{1}{\tau_n}} \quad (2.28a)$$

$$\frac{D_p}{L_p} = \sqrt{\frac{D_p}{\tau_p}} = \sqrt{\frac{\mu_p k T}{q} \frac{1}{\tau_p}} \quad (2.28b)$$

můžeme je tedy považovat za materiálovou konstantu (závislost na teplotě je malá). Koncentrace n_{p0} a p_{n0} jsou *rovnovážné koncentrace minoritních nosičů*, pro které z (1.24) a (1.25) plyne

$$n_{p0} = n_i^2 / N_A \quad , \quad (2.29a)$$

$$p_{n0} = n_i^2 / N_D \quad . \quad (2.29b)$$

Nyní můžeme definovat **saturační proud** I_0 jako

$$I_0 = q A \left(\frac{D_n}{L_n} \frac{1}{N_A} + \frac{D_p}{L_p} \frac{1}{N_D} \right) n_i^2 \quad (2.30)$$

V případě nesouměrného přechodu P⁺N, kde $N_A \gg N_D$ můžeme vztah (2.30) zjednodušit na

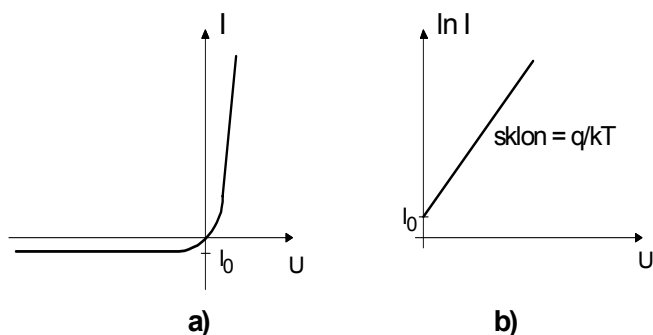
$$I_0 \approx q A \left(\frac{D_p}{L_p} \frac{1}{N_D} \right) n_i^2 \quad (2.31)$$

Jak vyplývá z (2.31), saturační proud

- roste s plochou přechodu A,
- klesá s rostoucí koncentrací v bázi diody N_D ,
- roste s teplotou T (n_i roste exponenciálně s teplotou T - viz (1.14)),
- klesá se šířkou zakázaného pásu E_G (n_i klesá exponenciálně s E_G - viz (1.14)),
- klesá se zvětšující se dobou života minoritních nosičů (τ_n , τ_p).

U křemíkových diod bývá saturační proud 10^{-12} až 10^{-8} A a zdvojnásobí se při zvýšení teploty o 6 K,

u germaniových diod bývá 10^{-5} až 10^{-3} A a zdvojnásobí se při zvýšení teploty o 10 K.



Obr. 2.5. VA-charakteristika PN přechodu: a) v lineárním
a b) semilogaritmickém měřítku

Použijeme-li definici teplotního napětí (1.47) $U_T = kT/q$, I_0 saturačního proudu (2.30) a dosadíme-li (2.29) do (2.26), dostaneme **Shockleyho rovnici ideální diody**

$$I = I_0 \left(\exp \frac{U}{U_T} - 1 \right)$$

(2.32)

Ampérovoltová charakteristika ideálního PN přechodu je na obr. 2.5. Vidíme, že v závěrném směru pro

$U \ll |U_T|$ je $\exp(U/U_T) \ll 1$ a (2.32) přechází do tvaru

$$I \approx I_0 \quad (2.33a)$$

a v propustném směru pro $U \gg U_T$ je $\exp(U/U_T) \gg 1$ a potom

$$I \approx I_0 \left(\exp \frac{U}{U_T} \right) \quad (2.33b)$$

Jako u jiných nelineárních prvků můžeme i u přechodu PN definovat **statický** (prostý) a **dynamický** (diferenciální) **odpor** jako funkci přiloženého napětí. Ze vztahu (2.32) dostaneme

$$U = U_T \ln \left(1 + \frac{I}{I_0} \right) \approx U_T \ln \left(\frac{I}{I_0} \right) \quad (2.34)$$

přičemž přibližný výraz platí pro $I \gg I_0$.

Statický odpor je potom definován

$$R = \frac{U}{I} = \frac{U}{I_0 \left(\exp \frac{U}{U_T} - 1 \right)} = \frac{U_T \ln \left(1 + \frac{I}{I_0} \right)}{I} \quad (2.35)$$

a **dynamický odpor**

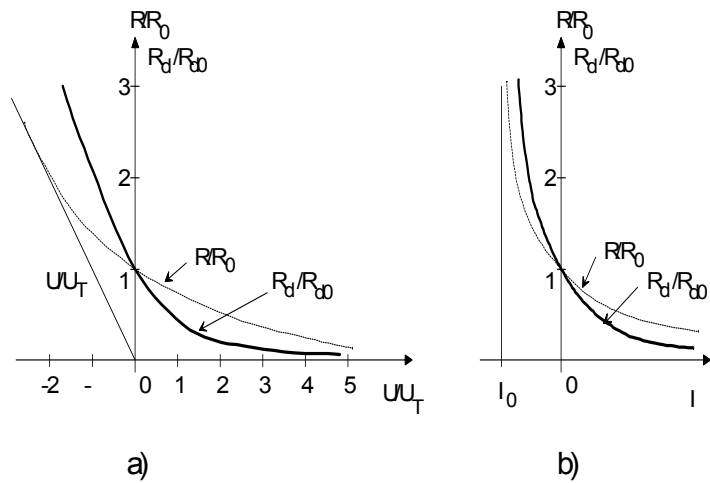
$$R_d = \frac{dU}{dI} = \frac{U_T}{I_0 \exp \frac{U}{U_T}} = \frac{U_T}{I + I_0} \approx \frac{U_T}{I} \quad (2.36)$$

Přibližný výraz v (2.36) platí pro $I \gg I_0$.

V počátku souřadnic ($U = I = 0$) jsou oba odpory stejné

$$R_0 = R_{d0} = \frac{U_T}{I_0} \quad (2.37)$$

Velikost odporu R_0 je pro křemíkové diody 10^6 až $10^{10} \Omega$, pro germaniové diody 10 až $10^3 \Omega$. Závislost statického a dynamického odporu PN přechodu na napětí a na proudu je na obr. 2.6. S dynamickým odporem počítáme při malých změnách napětí a proudu, tj. především při superponovaném střídavém signálu. V propustném směru se dynamický odpor teoreticky blíží k nule. Jak uvidíme později, musíme k němu ve skutečnosti přičíst ještě sériový odpor R_s .



Obr.2.6. Statický a dynamický odpor PN přechodu: a) napěťová závislost, b) proudová závislost

2.3.3 Kapacita PN přechodu

PN přechod vytváří v polovodičové struktuře dva typy kapacity:

1. *bariérovou* (přechodovou) kapacitu C_j [junction capacitance, transition region capacitance, depletion layer capacitance] vytvořenou dvojvrstvou prostorového náboje v depletiční vrstvě a
2. *difúzní kapacitu* [diffusion capacitance, charge storage capacitance] tvořenou akumulací náboje v důsledku průchodu proudu.

Obě tyto kapacity musíme brát v úvahu při návrhu polovodičových součástek a při jejich použití při časově proměnném signálu. Při některých aplikacích omezuje kapacita součástky její použití, při jiných naopak kapacitu PN přechodu využíváme.

Bariérová kapacita převažuje v závěrném směru. Při jejím odvození vyjdeme z obecné definice kapacity

$$C_j = \left| \frac{dQ}{dU} \right| \quad (2.38)$$

Prostorový náboj na obou stranách přechodu můžeme spočítat z (2.12)

$$|Q| = qAx_nN_D = qAx_pN_A \quad (2.39)$$

přičemž šířku depletiční oblasti určíme z (2.22)

$$x_n = \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_A}{N_D(N_A + N_D)} \right]^{1/2}, \quad x_p = \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_D}{N_A(N_A + N_D)} \right]^{1/2}$$

Po dosazení (2.22) do (2.39) dostaneme pro velikost náboje $|Q|$

$$\begin{aligned}
|Q| &= qAN_D \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_A}{N_D(N_A + N_D)} \right]^{1/2} = \\
&= qAN_A \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_D}{N_A(N_A + N_D)} \right]^{1/2} = \\
&= qA \left[\frac{2\varepsilon}{q} (U_D - U) \frac{N_A N_D}{(N_A + N_D)} \right]^{1/2} \quad (2.40)
\end{aligned}$$

Bariérovou kapacitu určíme podle (2.38) derivací (2.40) podle napětí

$$C_j = \left| \frac{dQ}{d(U_D - U)} \right| = \frac{qA}{2} \left[\frac{2\varepsilon}{q} \frac{1}{(U_D - U)} \frac{N_A N_D}{(N_A + N_D)} \right]^{1/2} \quad (2.41)$$

Označíme-li kapacitu pro $U = 0$ V jako C_{j0} , pak můžeme (2.41) zjednodušit do tvaru

$$C_j = \varepsilon A \left[\frac{q}{2\varepsilon} \frac{1}{(U_D - U)} \frac{N_A N_D}{(N_A + N_D)} \right]^{1/2} = \frac{C_{j0}}{\sqrt{1 - \frac{U}{U_D}}} \quad (2.42)$$

Vidíme, že na závěrně pólovaný PN přechod můžeme také pohlízet jako na deskový kondenzátor, ve kterém se oblast prostorového náboje chová jako dielektrikum o tloušťce w a relativní permitivitě ε_r (pro Si je $\varepsilon_r \approx 12$) mezi vodivými oblastmi N a P. Pro strmý přechod můžeme proto bariérovou kapacitu spočítat jako

$$C_j = \varepsilon \frac{A}{w} \quad (2.43)$$

kde w určíme podle (2.25)

$$w = \left[\frac{2\varepsilon}{q} (U_D - U) \frac{(N_A + N_D)}{N_A N_D} \right]^{1/2}$$

Jak vidíme, rovnice (2.43) je po dosazení (2.25) totožná s (2.42). Závislost bariérové kapacity na přiloženém napětí vidíme na obr. 2.7. Bariérová kapacita je závislá:

- na ploše přechodu A ,
- na koncentracích N_A a N_D ,
- na velikosti přiloženého napětí $\sim U^{-1/2}$.

Difúzní kapacita převažuje v propustném směru. Je způsobena změnou akumulovaného náboje minoritních nosičů injikovaných do kvazineutrálních oblastí diody při změně napětí. Při odvození její velikosti uvažujeme např. nesouměrný P⁺N přechod, kterým protéká proud I . V tomto případě je možno počítat pouze s akumulovaným nábojem v N oblasti (báze diody, vysokohmová oblast). Akumulovaný náboj injikovaných minoritních děr je pro $U \gg U_T$ viz vzorce 2.26 a 2.27):

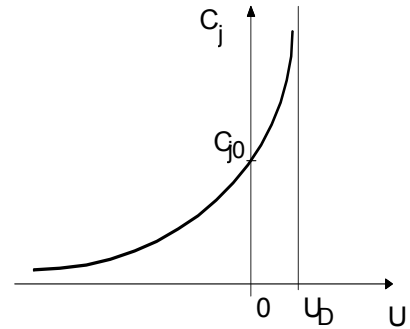
$$Q_{ap} = I\tau_p = qA\Delta p_n L_p = qAL_p p_n \exp(qU/kT) \quad (2.44)$$

Difúzní kapacitu potom spočítáme pomocí obecné definice kapacity (2.38)

$$C_D = \frac{dQ_{ap}}{dU} = \frac{q^2}{kT} AL_p p_n \exp\left(\frac{qU}{kT}\right) = \frac{q}{kT} I_p = \frac{I}{U_T} \tau_p \quad (2.45)$$

Celková kapacita PN přechodu je dána součtem bariérové a difúzní kapacity, tedy

$$C = C_j + C_D \quad (2.46)$$



Obr. 2.7. Závislost bariérové kapacity PN přechodu na přiloženém napětí

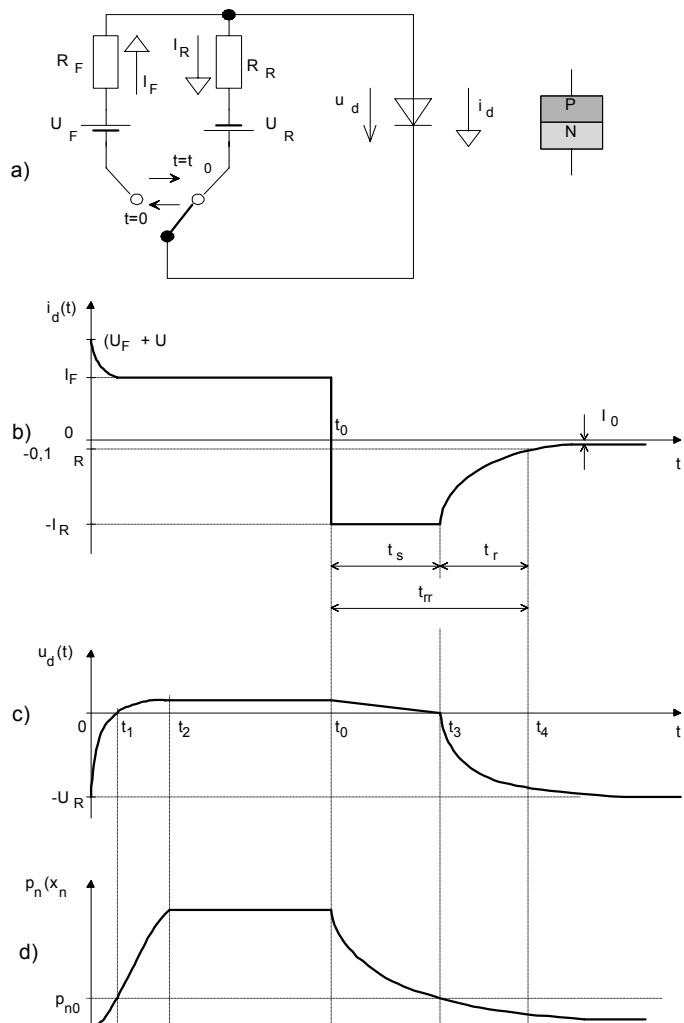
2.4 Přechod PN v dynamickém režimu

Pro různé aplikace je třeba znát chování přechodu PN při různých velkých a různě rychlých změnách napětí

a proudu. Souvislost mezi napětím a proudem daná Shockleyho rovnicí ideální diody (2.32) platí jen pro dostatečně pomalé změny napětí a proudu. Při rychlých změnách platí až po odeznění přechodného děje, který ovlivňuje jak difúzní, tak i bariérová kapacita. Rozebereme zde tři typické případy.

2.4.1 Přechod PN v impulsovém režimu

Nejnázornější je sledovat chování PN přechodu při buzení skokovým signálem, tj. při přepnutí přechodu z propustného do závěrného směru a naopak. Uvažujme obvod principiálně podle obr. 2.8a) s předpokladem, že polarita napětí se v časech $t = 0$ a $t = t_0$ mění přepínačem S skokově. V praxi bychom samozřejmě museli použít jiné experimentální uspořádání.

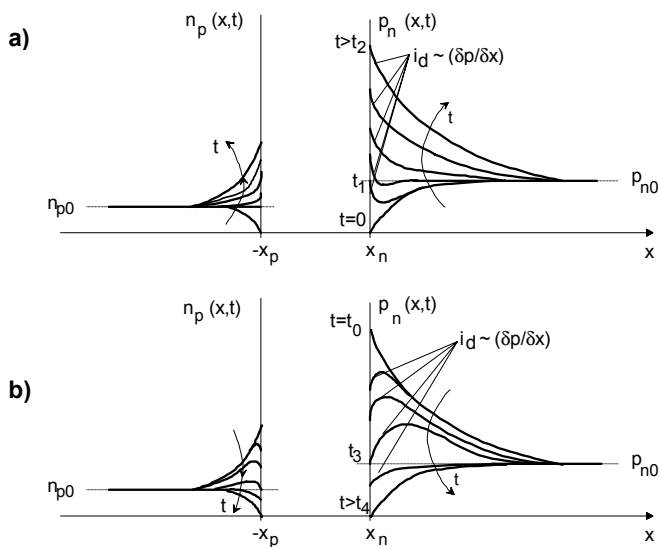


Obr. 2.8. a) Obvod pro vyšetřování impulsových vlastností PN přechodu. Časové průběhy b) proudu, c) napětí a d) koncentrace minoritních děr při impulsových změnách napětí.

Nejprve vyšetříme chování PN přechodu při **přepnutí ze závěrného do propustného směru**. Předpokládejme, že v čase $t = 0$ je na přechod přiloženo kladné napětí. Protože při závěrné polarizaci má PN přechod velký odpor, je napěťový úbytek na rezistoru R_R zanedbatelný a celé závěrné napětí U_R spočívá na depletiční vrstvě, tj. na kapacitoru C_j ($u_d \approx U_R$). Potom bezprostředně po přepnutí poteče přechodem proud $i_d \approx (U_F + U_R)/R_F$, protože napětí na depletiční vrstvě se nemůže vzhledem k její konečné kapacitě okamžitě měnit. V čase t_1 se kapacitor C_j vybije na napětí $u_d(t_1) \approx 0$ V, proud přechodem přestane klesat a dosáhne hodnoty $i_d(t_1) = (U_F - u_d)/R_F \approx U_F/R_F = I_F$ (protože $U_F \gg u_d$). Rozhodující časová konstanta pro tento průběh je $\tau_I = R_F C_j$. Průběh napětí $u_d(t)$ lze použít k určení doby t_1 , protože $u_d(t_1) \approx 0$ V. Od tohoto okamžiku napětí na diodě $u_d(t)$ stoupá jen velmi málo až do doby t_2 , kdy koncentrace injikovaných minoritních nosičů v kvazineutrálních oblastech N a P dosáhnou ustálené hodnoty (viz obr. 2.8d). Akumulovaný náboj Q_a těchto minoritních nosičů vytvoří difúzní kapacitu C_D . Pro $t > t_2$ protéká obvodem ustálený proud I_F . Průběh koncentrací injikovaných minoritních nosičů v kvazineutrálních oblastech N a P při přepínání ze závěrného do propustného stavu je na obr. 2.9a). Všimněme si, že ustálené koncentrace minoritních nosičů na hranicích depletiční vrstvy $p_n(x_n)$ a $n_p(-x_p)$ se v propustném stavu liší od koncentrací rovnovážných p_{n0} a n_{p0} .

Dále se budeme zabývat chováním PN přechodu při **přepnutí z propustného do závěrného směru**. Předpokládejme, že v čase $t = t_0 > t_2$, kdy již bylo dosaženo na přechodu ustálených podmínek, změníme přepínačem polaritu přiloženého napětí.

V okamžiku t_0 klesne proud obvodem na hodnotu $i_d(t_0) = -(U_R + u_d)/R_R \approx -U_R/R_R = -I_R$, protože v kvazineutrálních oblastech je stále ještě akumulován náboj Q_a . V kvazineutrálních oblastech poblíž depletiční vrstvy nastane prudký spád koncentrace minoritních nosičů, což způsobí velký zpětný proud I_R , který je prakticky omezen jen odporem R_R , který je zapojen v sérii s PN přechodem.



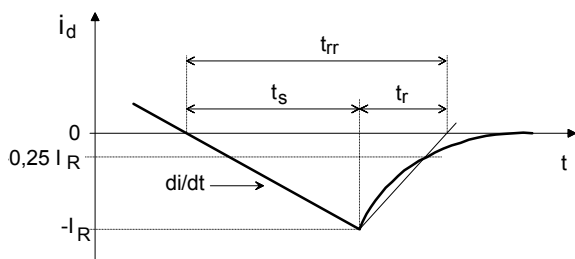
Obr. 2.9. Časový vývoj průběhu koncentrace minoritních nosičů v PN přechodu **a)** při přechodu z nevodivého stavu do stavu vodivého a **b)** při přechodu z vodivého stavu do stavu nevodivého.

Aby přechod mohl přejít do závěrného stavu s napětím $u_d(\infty) \approx -U_R$, musí injikované minoritní nosiče odtéct a musí se ustálit na koncentraci odpovídající závěrnému proudu. To znamená, že v oblastech hraničících s depletiční vrstvou dojde k extrakci minoritních nosičů a jejich ustálený průběh koncentrace bude nižší než jejich rovnovážná koncentrace p_{n0} a n_{p0} . Po celou dobu t_s , po kterou minoritní nosiče odtékají a rekombinují, teče obvodem konstantní proud $i_d = -I_R \approx -U_R/R_R$. Jakmile jsou nadbytečné minoritní nosiče odstraněny, přechází přechod do závěrného stavu, objeví se na něm závěrné napětí a proud přechodem začne klesat, až za dobu t_r dosáhne hodnoty $-0,1 I_R$. Záporné napětí na

diodě roste a za dobu t_r dosáhne přibližně hodnoty U_R . V čase $t \rightarrow \infty$ se závěrný proud ustálí na hodnotě saturačního proudu $-I_0$ a koncentrace minoritních nosičů v kvazineutrálních oblastech dosáhnou ustálené hodnoty. Dobu t_s nazýváme *dobou zpoždění* nebo *dobou akumulace* [storage time], dobu t_r *dobou poklesu* nebo *dobou doběhu* [recovery time]. Jejich

součet $t_{rr} = t_s + t_r$ nazýváme *dobou zpětného zotavení* [reverse recovery time]. Tato doba je závislá:

- přímo úměrně na době života τ minoritních nosičů ve vysokoohmové oblasti (v bázi diody),
- přímo úměrně na velikosti proudu I_F (množství akumulovaného náboje Q_a),
- nepřímo úměrně na velikosti proudu I_R (doba poklesu),
- přímo úměrně na hloubce báze.



Obr. 2.10. Časový průběh proudu PN přechodem při zpětném zotavení s lineárním poklesem proudu i_F

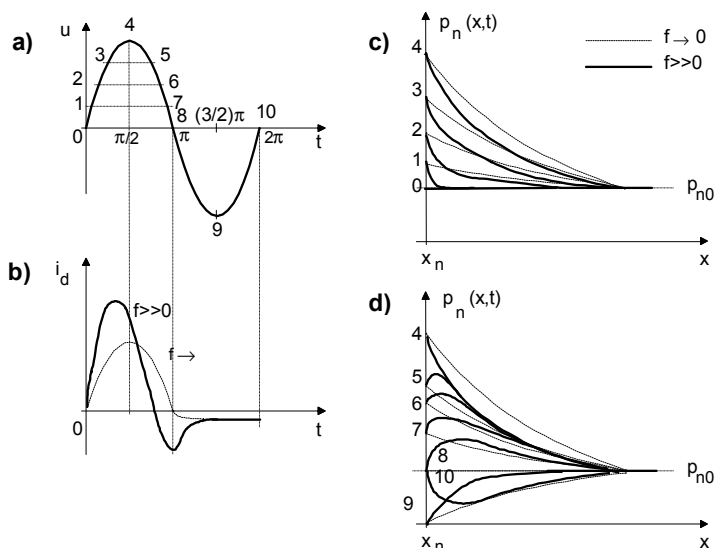
V reálném obvodu nedochází ke skokové změně proudu. Rychlost změny proudu je limitována indukčností obvodu. V takovém případě je možno předpokládat lineární pokles proudu s konstantní rychlostí $-di/dt$, tedy $i_d(t) = I_F - (di/dt)t$. Situace je znázorněna na obr. 2.10. Doba zpoždění t_s závisí na rychlosti poklesu proudu di/dt . Protože pro $t = t_0 + t_s$ se prudce mění strmost proudu, dochází na indukčnostech obvodu ke vzniku překmitu napětí.

2.4.2 Přechod PN při sinusovém napětí

Dalším důležitým případem je proces zpětného zotavení PN přechodu při sinusovém průběhu přiloženého napětí. Průběh vytváření akumulovaného náboje v N oblasti P⁺N přechodu je znázorněn na obr. 2.11.

Při **nízkém kmitočtu** ($f \rightarrow 0$) koncentrace nosičů stačí dosáhnout ustáleného průběhu pro každou okamžitou hodnotu střídavého napětí. Tomu odpovídají čárkované křivky jednotlivých průběhů koncentrace nosičů na obr. 2.11c), d). Proud sleduje napětí, je zkrácen pouze nelinearitou charakteristiky a v záporné půlperiodě teče PN přechodem pouze saturační proud (obr. 2.11b).

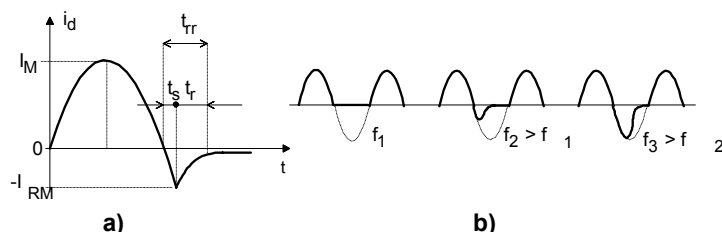
Při **vysokém kmitočtu** ($f \gg 0$) přiloženého napětí nestačí injekce nosičů vytvořit celý ustálený náboj. Průběhy koncentrace minoritních nosičů odpovídající jednotlivým napětím na diodě (body 1 až 4) mají na začátku kvazineutrální oblasti větší gradient, čemuž odpovídá i větší proud diody v první čtvrt-periodě napětí, který dosáhne maxima dříve než při kvazistacionárním režimu ($f \rightarrow 0$). Při poklesu napětí ve druhé čtvrtperiodě má v důsledku setrvačnosti akumulovaného náboje průběh koncentrace snahu udržet se na původní hodnotě a koncentrace se začíná zmenšovat nejdříve na začátku kvazineutrální oblasti (body 4 až 8). Gradient koncentrace na začátku kvazineutrální oblasti se zmenšuje rychleji než v kvazistacionárním režimu, prochází již před koncem periody nulou a stává se dokonce záporným. Proto i proud PN přechodu prochází nulou pro $\omega t < \pi$ a za $\omega t = \pi$ má záporné maximum, z něhož klesá na hodnotu ustáleného



Obr. 2.11. PN přechod s přiloženým střídavým napětím při nízkém a vysokém kmitočtu: a) průběh přiloženého napětí, b) průběh proudu, c), d) koncentrace minoritních nosičů v N oblasti

závěrného (saturačního) proudu. Vidíme že průběh proudu je zkreslen oproti průběhu napětí a jeho střední hodnota klesá při zvyšování kmitočtu

Průběh proudu PN přechodem při sinusovém napětí je na obr. 2.12a). Při zvyšování kmitočtu se začíná uplatňovat výše popsany zotavovací proces, který způsobuje, že přechod začíná být vodivý i v závěrném směru, až přestává vůbec usměrňovat (tj. být jednosměrně vodivý). Tento jev je naznačen na obr. 2.12b).



Obr. 2.12. a) Zpětné zotavení PN přechodu při sinusovém průběhu napětí b) snižování jeho usměrňovací schopnosti s rostoucím kmitočtem

2.4.3 Přechod PN s přiloženým stejnosměrným a střídavým napětím

V praxi je častý případ, že je na PN přechod přiloženo kromě stejnosměrného napětí (tzv. předpětí) ještě "malé" střídavé

napětí. Při závěrné polarizaci se bude uplatňovat diferenciální vodivost a bariérová kapacita (závislá na přiloženém ss napětí, příp. i na st napětí, není-li amplituda $U_{st} \gg U_R$). Při propustné polarizaci se bude uplatňovat opět diferenciální vodivost a difúzní kapacita. S rostoucím kmitočtem se kapacita zmenšuje a vodivost zvětšuje. V obou případech vyhovuje náhradní paralelní RC zapojení PN přechodu.

2.5 Průraz přechodu PN

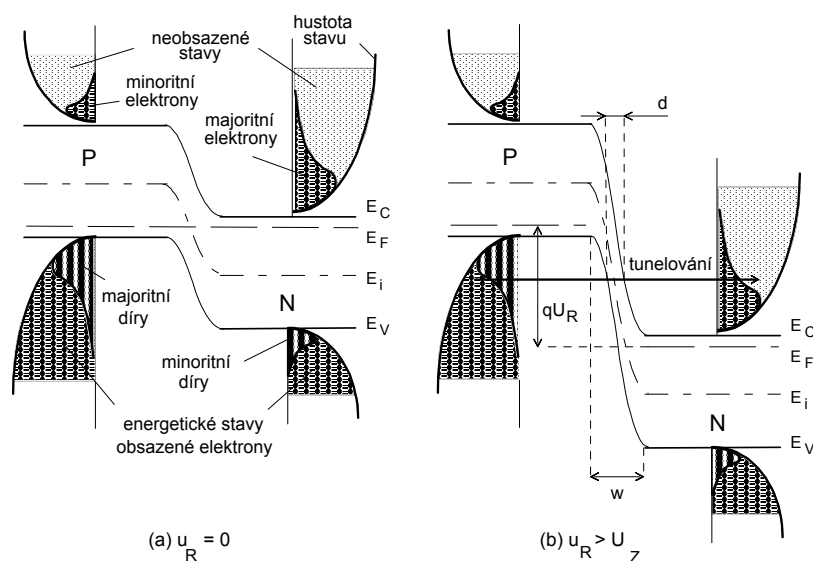
Překročí-li závěrné napětí u_R jistou kritickou hodnotu $U_{(BR)}$, dochází k rychlému vzrůstu závěrného proudu, což nazýváme *průrazem přechodu*. Velikost průrazného napětí je určována oblastí s menší koncentrací příměsí. Mechanismus průrazu závisí na typu přechodu (strmý, lineární), materiálu polovodiče, tvaru přiloženého napětí a na teplotě. Podle fyzikálního mechanismu rozeznáváme několik různých typů průrazu:

1. průraz elektrickým polem (tunelový, Zenerův),
2. průraz lavinový,
3. průraz tepelný,
4. průraz povrchový.

2.5.1 Tunelový (Zenerův) průraz

V klasické fyzice musí mít částice energii větší než je potenciálová bariéra, aby se mohla dostat na druhou stranu. V kvantové mechanice však stačí, aby bariéra byla dostatečně tenká a částice se může dostat na druhou stranu tzv. *tunelováním*. Dva základní předpoklady pro tunelování jsou:

1. Tenká potenciálová bariéra; čím je bariéra tenčí, tím větší je pravděpodobnost tunelování.
2. Dostatečné množství elektronů pro tunelování na jedné straně bariéry a na stejné energetické úrovni dostatečné množství volných stavů na druhé straně bariéry.



Obr. 2.13. Princip tunelového (Zenerova) průrazu: **a)** PN přechod v termodynamické rovnováze a **b)** při přiložení závěrného napětí většího než průrazné napětí U_Z

obsazené stavy ve valenčním pásu P oblasti posunou proti neobsazeným stavům ve vodivostním pásu N oblasti a může docházet k tunelování elektronů z oblasti P do N (viz obr. 2.13.b). Přivedení závěrného napětí má ještě dva další důsledky příznivé pro vznik tunelového průrazu. I když poněkud vzroste šířka depletiční oblasti w ($w \sim u_R^{1/2}$), šířka "tunelové" bariéry (d na obr. 2.13.b) se zmenšuje a pravděpodobnost tunelového průchodu vzrůstá. Druhým důsledkem přiložení závěrného napětí je mírné zvětšení intenzity elektrického pole v závěrné oblasti ($E \sim w \sim u_R^{1/2}$), což opět zvětšuje pravděpodobnost tunelového průchodu. Jestliže tedy na PN přechod přiložíme takové závěrné napětí u_R , že dojde ke splnění obou kritérií pro tunelování, nastává průraz přechodu a dojde k prudkému zvýšení proudu přechodem. Toto průrazné napětí nazýváme *Zenerovo napětí* a značíme U_Z . Zvětšujeme-li dále závěrné napětí, ocitá se na stejné energetické úrovni více elektrony obsazených stavů na straně P a více neobsazených stavů na straně N (viz obr. 2.13.), zmenšuje se více šířka "tunelové" bariéry d a proud přechodem dále silně vzrůstá.

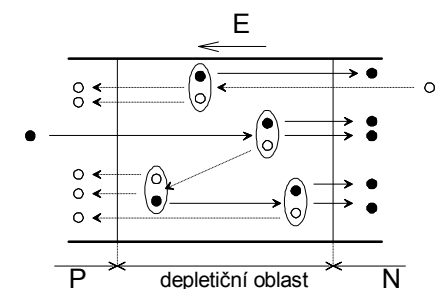
Tunelový průraz lze pozorovat jen pro napětí $u_R < 4E_G/q$, (tj. pro napětí $u_R < 4,5$ V pro Si a $u_R < 2,8$ V pro Ge). Při napětí v intervalu $4E_G/q < u_R < 6E_G/q$ se kromě tunelového průrazu uplatní i průraz lavinový a pro napětí $u_R > 6E_G/q$ (tj. pro napětí $u_R > 6,7$ V pro Si a $u_R > 4,2$ V pro Ge) se uplatní pouze lavinový průraz. Který z těchto průrazů u konkrétního přechodu nastane, to závisí na koncentraci příměsí v obou částech přechodu, obecně s rostoucí koncentrací příměsí velikost průrazného napětí klesá.

Protože se šířka zakázaného pásu s rostoucí teplotou zmenšuje (pro Si o $2,8 \cdot 10^{-4}$ eV/K), je také *teplotní součinitel průrazného napětí* způsobeného tunelovým jevem záporný, tj. průrazné napětí se s rostoucí teplotou zmenšuje. Tunelový průraz je *nedestruktivní*.

První kritérium pro tunelování je u PN přechodu splněno tehdy, jestliže je šířka depletiční oblasti velmi malá (menší než je střední volná dráha elektronu, aby nedocházelo k ionizačním srážkám) a intenzita elektrického pole dosáhne určité hodnoty (pro Si $E_{krit} = 1,2 \div 1,4 \cdot 10^8$ Vm⁻¹, pro Ge $E_{krit} = 2,2 \div 3 \cdot 10^7$ Vm⁻¹). Toho dosáhneme použitím vysoce dotovaného polovodiče na obou stranách přechodu, tj. pro koncentrace N_D a $N_A > 10^{17}$ cm⁻³ = 10^{23} m⁻³ (viz rovn. 2.25). Na obr. 2.13.a) vidíme takový přechod v termodynamické rovnováze. Na obrázku je znázorněna hustota stavů ve valenčním i vodivostním pásu a energetické rozložení majoritních i minoritních nosičů (srovnej s obr. 1.13.). Splnění druhého kritéria dosáhneme přiložením závěrného napětí na tento PN přechod. Tím se elektrony

2.5.2 Lavinový průraz

Nejčastěji se vyskytující *lavinový průraz* je charakteristický pro vysokohodnotný (s malou koncentrací příměsí) PN přechod, který má dostatečně širokou depletiční vrstvu (mnohem širší než je střední volná dráha elektronu). Pak při dostatečné intenzitě elektrického pole, kterou vyvolá přiložené závěrné napětí, získají minoritní nosiče procházející depletiční vrstvou velkou kinetickou energii a může dojít k nárazové ionizaci atomů krystalové mřížky, tj. ke generaci párů elektron-díra (viz obr. 2.14.). Touto ionizací lavinovitě narůstá koncentrace párů elektron-díra, tj. počet nosičů, který vystupuje z depletiční vrstvy (r) je mnohonásobně vyšší než počet nosičů, který vstupuje (s). Toto zvětšení se charakterizuje *multiplikačním činitelem*



Obr. 2.14. Princip nárazové ionizace v depletiční oblasti při závěrné polarizaci PN přechodu

multiplikačním činitelem

$$M = \frac{s}{r} = \frac{1}{1 - \left(\frac{U_R}{U_{(BR)}} \right)^n} \quad (2.47)$$

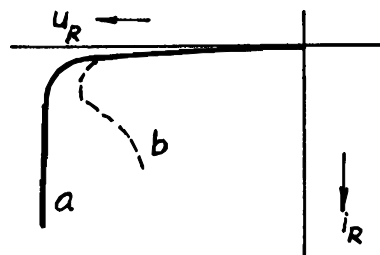
kde n je empirický exponent. Napětí $U_{(BR)}$ nazýváme průrazné napětí.

Průrazné napětí strmého přechodu závisí na stupni dotování materiálu (tj. také na měrném odporu) a při zvyšování koncentrace příměsí se snižuje (obr. 3.5. a 2.16.). Rozhodující je méně dotovaná strana přechodu. Průrazné napětí lineárního přechodu závisí na gradientu koncentrace. Průrazné napětí může být v jednotkách až stovkách voltů (zdířka více než 3000 V).

Z experimentů dále plyne, že průraz nastává při nižších hodnotách napětí, než vyplývá z teorie. Je to způsobeno tím, že průraz (násobení nosičů) začíná v povrchových vrstvách a ne uvnitř přechodu (viz později).

Také pokud existují v oblasti prostorového náboje přechodu PN nehomogenity a to jak v objemu, tak na povrchu - a obvykle tomu tak je, může být podmínka průrazu splněna lokálně a k průrazu může dojít v malé lokalizované oblasti přechodu PN. V takovém případě vzniká tzv. mikroplasma. Tento jev brání dosažení vysokých průrazných napětí. Bráníme se používáním co nejhomogennějších materiálů a minimalizací strukturních defektů (viz stat' o lavinových usměrňovacích diodách).

Na rozdíl od tunelového průrazu, který má záporný teplotní součinitel průrazného napětí, má lavinový průraz teplotní součinitel průrazného napětí kladný, průrazné napětí se s rostoucí teplotou zvyšuje (obr. 2.17.). Při průrazném napětí kolem 5,6 V, kde se oba jevy překrývají, je vliv teploty minimální. Toho se využívá u referenčních stabilizačních diod.



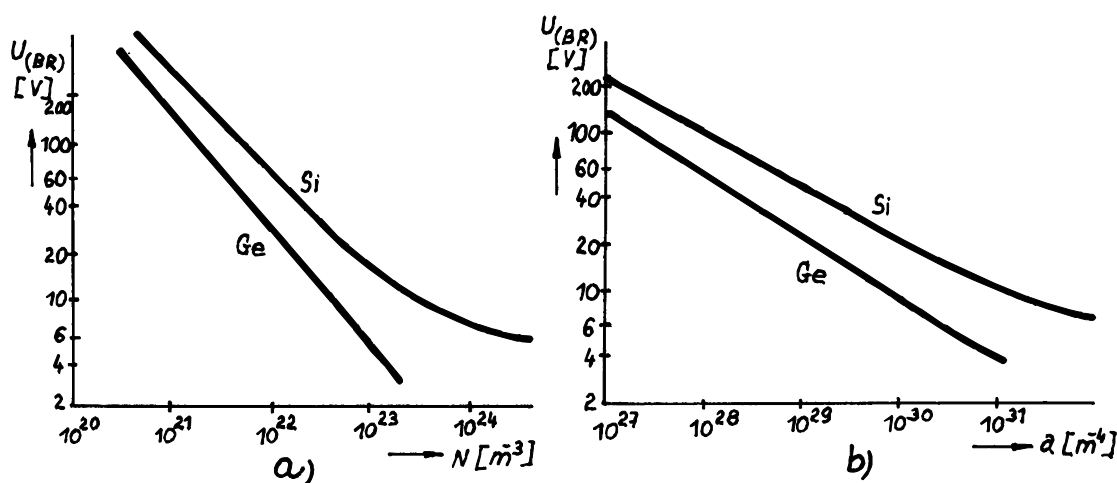
Obr. 2.15. Závěrná větev AV charakteristiky při a) lavinovém nebo b) tepelném průrazu

Srovnání vlastností tunelového a lavinového průrazu je přehledně uvedeno v tabulce 2.1. a na obr. 2.17 a 3.6.

S rostoucí koncentrací příměsí velikost průrazného napětí klesá !

Tab. 2.1. Srovnání vlastností tunelového a lavinového průrazu

VLASTNOST	TUNELOVÝ PRŮRAZ	LAVINOVÝ PRŮRAZ
průrazné napětí	< 6,7 V (Si) < 4,2 V (Ge)	> 4,5 V (Si) > 2,8 V (Ge)
teplotní koeficient	záporný	kladný
koncentrace příměsí	velká (> 10 ¹⁷ cm ⁻³)	střední a malá (rozhodující je méně dotovaná strana přechodu)



Obr. 2.16. Průrazné napětí při lavinovém průrazu v závislosti na koncentraci příměsí a) u strmého přechodu a na gradientu koncentrace příměsí b) u lineárního přechodu

2.5.3 Tepelný průraz

Tepelný průraz přechodu PN souvisí s průchodem proudu závěrně polarizovaným přechodem. Při průchodu proudu vzniká na přechodu PN ztrátový výkon, který se mění na teplo. Množství tepla vznikající na přechodu za jednotku času je

$$P_j = u_R i_R \quad (2.48)$$

Zároveň je za jednotku času odváděno do okolí o teplotě T_a množství tepla

$$P_a = \frac{T_j - T_a}{R_{th}} \quad (2.49)$$

kde T_j je teplota přechodu PN a R_{th} je tepelný odpor mezi přechodem PN a okolím.

Principem tepelného průrazu je vznik kladné zpětné vazby. Jestliže množství vzniklého tepla je větší než je množství tepla odváděného, teplota přechodu se počne zvyšovat, v důsledku toho vzroste závěrný proud přechodem (při konstantním závěrném napětí), množství vznikajícího tepla se opět zvýší, atd. Pokud proud není omezen vnějším obvodem, dojde k "nedovolenému" ohřevu přechodu a k jeho tepelnému průrazu.

Pro kvantitativní studium tepelného průrazu je třeba především vyšetřovat podmínky tepelné rovnováhy přechodu. Tepelná rovnováha nastává v případě, kdy je množství tepla

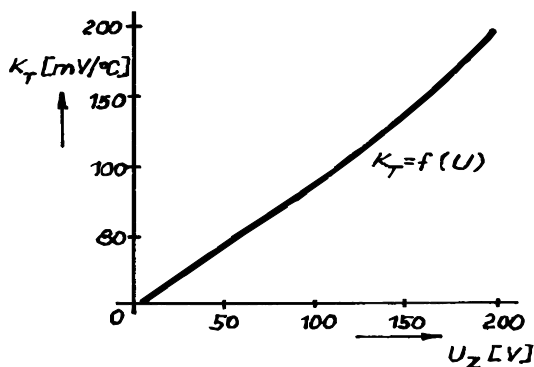
vytvářeného na přechodu PN výkonovou ztrátou rovno množství tepla odváděného od přechodu. Z rozboru plyne, že závěrný proud a tedy i ztrátový výkon P_j roste s teplotou přibližně exponenciálně, kdežto odváděný výkon P_a závisí na teplotě lineárně. Situaci ilustruje obr. 2.18. Ve stabilním stavu se závislosti $P_j(T_j)$ a $P_a(T_j)$ protínají ve dvou bodech. Po zapnutí proudu se tepelným přechodovým dějem ustaví $T_j = T_{op}$ (pokud je to fyzikálně možné). Tento pracovní bod (průsečík) je stabilní, protože při zvýšení teploty nad T_{op} množství odváděného tepla převyšuje množství tepla vznikajícího (a přechod se ochladí na T_{op}). Druhý průsečík T_{kr} odpovídá nestabilnímu stavu, protože zvýšení teploty nad T_{kr} vede k rychlému zvýšení generovaného tepla a následně k dalšímu zvýšení teploty přechodu. Vzniká tak *tepelný průraz*. Podmínkou stabilního stavu je:

$$\frac{\partial P_j}{\partial T_j} < \frac{\partial P_a}{\partial T_j} \quad (2.50)$$

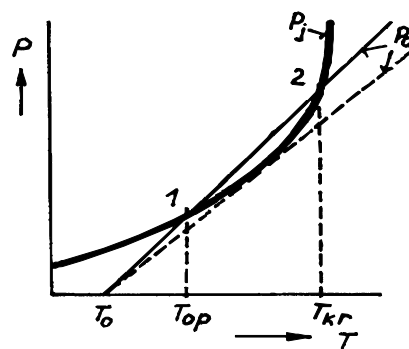
Při překročení teploty T_{kr} dochází k rychlému ohřevu struktury tak, že teplota může lokálně dosáhnout intrinsické teploty T_i . Pro $T > T_i$ se podstatně zvyšuje rychlost generace nosičů (i v oblasti mimo přechod PN) nezávisle na velikosti elektrického pole. Vzniká tak místo s rychlým růstem teploty doprovázeným lokálním zvýšením hustoty proudu o několik řádů, vzniká tzv. *mesoplasma*.

Vznik mesoplasmy má obvykle za následek nevratný proces degradace přechodu PN. Tento jev je nazýván *druhý průraz* [second breakdown], na rozdíl od lavinového průrazu, který nevede ke zničení nebo degradaci charakteristik přechodu PN. V případě druhého průrazu (resp. tepelného průrazu) dochází k poškození struktury jedním z možných mechanismů:

- v průběhu průrazu překročí teplota mesoplasmy teplotu tání křemíku,
- teplota v oblasti mesoplasmy překročí eutektickou teplotu slitiny křemíku s kovovým kontaktem,
- vlivem tepelného šoku dochází k porušení krystalické mřížky a vytvoření prasklin.

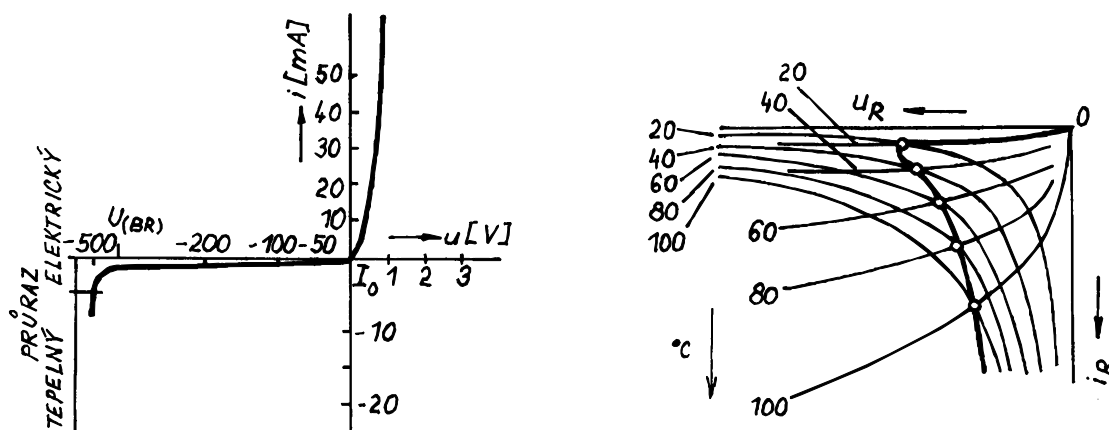


Obr. 2.17. Graf závislosti teplotního koeficientu průrazného napětí na jmenovitém Zenerově napětí



Obr. 2.18. Tepelná stabilita přechodu PN

Možnost vzniku druhého průrazu je limitujícím faktorem pro provoz polovodičových součástek s přechodem PN.



Obr. 2.19. Vznik tepelného průrazu

Na obr. 2.19. je znázorněna závislost závěrného proudu na teplotě a jsou nakresleny hyperboly konstantního ztrátového výkonu při týchž teplotách. Křivka, která prochází průsečíky odpovídajících čar, ukazuje hranici, jejíž překročení vede k tepelnému průrazu.

Poznámka: Tepelný průraz vzniká v přechodech PN s velkými zpětnými proudy. Běžné přechody PN mají natolik malé zpětné proudy, že tepelný průraz je u nich málo pravděpodobný. Může však nastat, je-li zpětný proud nějakým způsobem zvětšen, např. v tranzistorech, nebo v důsledku elektrického (lavinového nebo Zenerova) průrazu.

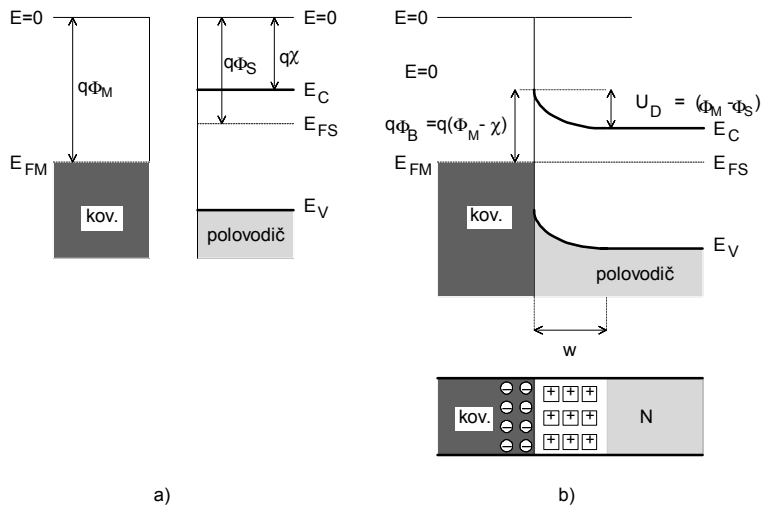
2.6 Přechod kov-polovodič

Kontakt (přechod) kov-polovodič se musí vytvořit při každé konstrukci polovodičových součástek i při každém měření polovodičových materiálů. Podle konkrétního provedení existuje přechod kov-polovodič ve dvou variantách: *usměrňující* (Schottkyho) a *neusměrňující* (ohmický).

2.6.1 Schottkyho bariéra

O tom, zda se při vytváření přechodu kov-polovodič realizuje usměrňující či neusměrňující kontakt, rozhoduje především *výstupní práce* elektronů z jednotlivých materiálů $W = q\Phi$. Připomeňme si, že výstupní práce je energie potřebná k přenosu elektronu z Fermiho hladiny na hladinu o nulové energii ve vakuu. Φ je *výstupní potenciál*. Typické hodnoty $q\Phi_M$ jsou 4,3 eV pro Al a 4,8 eV pro Au. Hodnoty $q\Phi_S$ silně závisí na typu a dotaci polovodiče, pro Si typu N se pohybují kolem 4 eV.

Vytvoříme-li kontakt mezi kovem s výstupní prací $q\Phi_M$ a polovodičem s výstupní prací $q\Phi_S$, dojde k přenosu náboje a ustaví se termodynamická rovnováha, tj. Fermiho energie v kovu a v polovodiči se vyrovná. Jestliže jako příklad použijeme polovodič typu N a $\Phi_M > \Phi_S$, pak je Fermiho hladina v polovodiči před vytvořením kontaktu výše než Fermiho hladina v kovu. Situace je znázorněna na obr.2.20a. Po vytvoření



Obr.2.20. Vznik Schottkyho bariéry: **a)** pásový diagram kovu a polovodiče před vytvořením kontaktu, **b)** pásový diagram přechodu kov-polovodič v rovnovážném stavu

potenciály kovu a polovodiče

$$U_D = \Phi_M - \Phi_S \quad (2.51)$$

Výška potenciálové *Schottkyho bariéry* pro injekci elektronů z kovu do vodivostního pásu polovodiče je

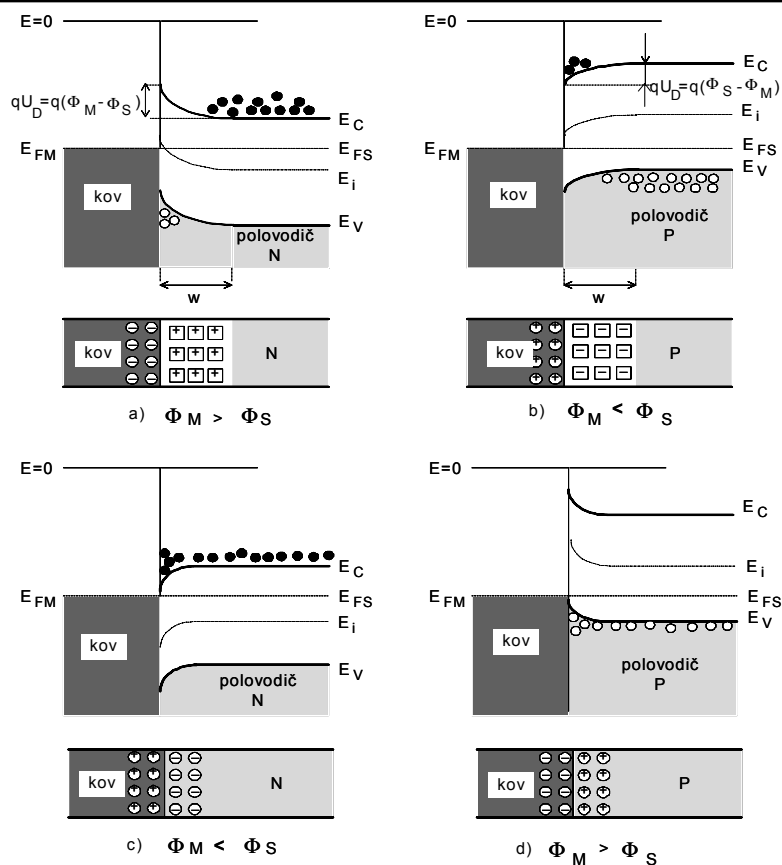
$$\Phi_B = \Phi_M - \chi \quad (2.52)$$

kde χ je *elektronová afinita* (rozdíl mezi hladinou vakua a dnem vodivostního pásu). Potenciálový rozdíl, který je v rovnovážném stavu roven U_D , můžeme zvyšovat nebo snižovat přiloženým závěrným nebo propustným napětím. Výška Schottkyho bariéry se vlivem přiloženého napětí nemění.

Podobná situace jako u přechodu kov-polovodič typu N a $\Phi_M > \Phi_S$ nastane i v případě přechodu kov-polovodič typu P a $\Phi_M < \Phi_S$. Také v tomto případě vznikne v polovodiči v blízkosti přechodu ochuzená až inverzní vrstva. Přehledně je situace znázorněna na obr.2.21a), b).

2.6.2 Usměrnující kontakt kov-polovodič

Jestliže na přechod z obr. 2.20a přiložíme propustné napětí U_F (tj. na kov bude přiloženo kladné napětí a na polovodič záporné), pak se kontaktní potenciál sníží z U_D na $U_D - U_F$ (obr. 2.22a). To způsobí, že elektrony mohou difundovat z vodivostního pásu polovodiče přes depletiční oblast do kovu a přechodem začne protékat kladný proud (z kovu do polovodiče). Naopak, jestliže je na přechod přiloženo závěrné napětí U_R , kontaktní potenciál se zvýší na $U_D + U_R$ (obr. 2.22b) a elektronový tok z polovodiče do kovu bude zanedbatelný. V obou případech toku elektronů z kovu do polovodiče brání bariéra Φ_B , která se s přiloženým napětím nemění, proto na její výšce závisí i velikost závěrného saturačního proudu. Typická VA-charakteristika usměrnujícího přechodu kov-polovodič je na obr.2.22c.



Obr.2.21. Pásové modely přechodů kov-polovodič při různých poměrech výstupních prací

2.6.3 Neusměrňující kontakt kov-polovodič

Při vytváření kontaktů na polovodičových součástkách potřebujeme naopak vytvořit přechod neusměrňující (ohmický) s lineární VA-charakteristikou pro obě polarity napětí. Ohmický kontakt dostaneme, jestliže je náboj potřebný k vyrovnání Fermiho hladiny v obou materiálech tvořen majoritními nosiči, tj. jestliže v blízkosti přechodu dochází v polovodiči k jejich akumulaci. Mějme například opět polovodič typu N, ale $\Phi_M < \Phi_S$ (obr.2.21c).

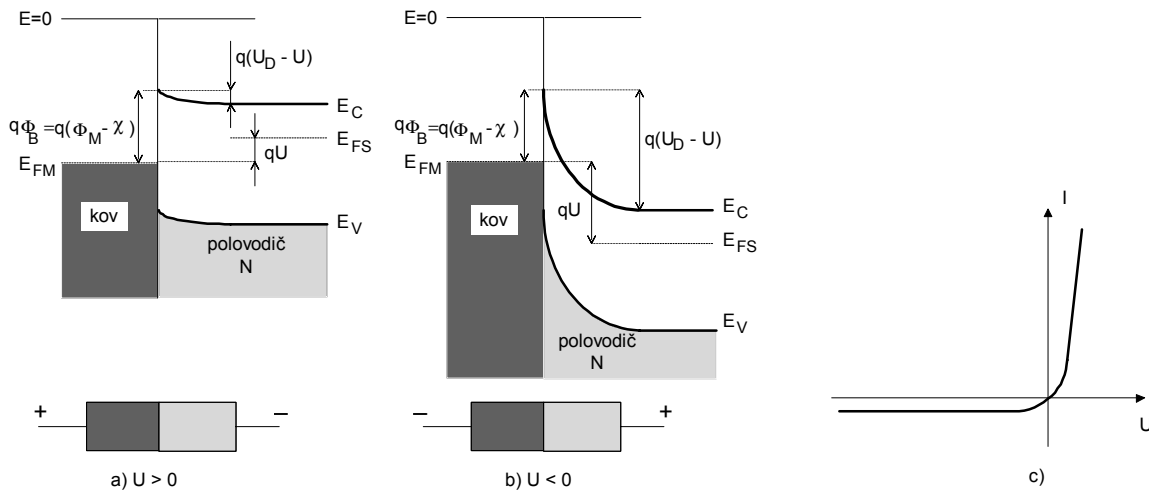
V rovnováze je Fermiho hladina vyrovnána přitečením elektronů z kovu do polovodiče, vznikne obohacená vrstva s vyšší vodivostí a potenciálová bariéra pro tok elektronů z kovu do polovodiče je velmi nízká a snadno překonatelná i při nízkých napětích. Podobná je situace pro přechod kov-polovodič typu P a $\Phi_M > \Phi_S$ (obr.2.21d).

1. Usměrnující přechod:

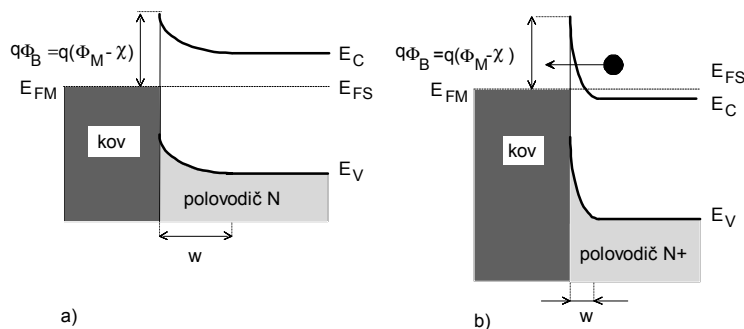
- a) $\Phi_M > \Phi_S$ kov - polovodič typu N: vznik ochuzené (až inverzní) vrstvy
- b) $\Phi_M < \Phi_S$ kov - polovodič typu P: vznik ochuzené (až inverzní) vrstvy

2. Neusměrňující přechod:

- c) $\Phi_M < \Phi_S$ kov - polovodič typu N: vznik obohacené vrstvy
- d) $\Phi_M > \Phi_S$ kov - polovodič typu P: vznik obohacené vrstvy



Obr.2.22. Pásový diagram přechodu kov-polovodič po přiložení **a)** propustného a **b)** závěrného napětí;
c) AV-charakteristika tohoto přechodu



Obr.2.23. Přechod kov-polovodič **a)** usměrňující, **b)** neusměrňující (tunelový)

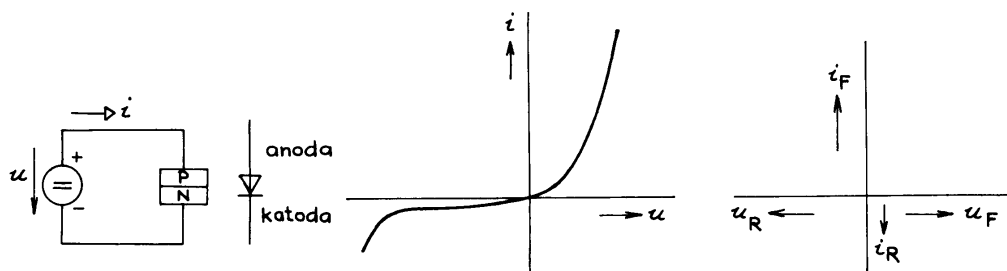
Praktickou metodou používanou při vytváření ohmických kontaktů je využití tunelového efektu při kontaktu kovu se silně dotovaným polovodičem. Šířka potenciálové bariéry se s rostoucí koncentrací příměsí v polovodiči zmenšuje a jakmile se přiblíží hodnotě 10^{19} cm^{-3} (10^{25} m^{-3}), je šířka

bariéry natolik malá, že elektrony mohou bariérou pronikat tunelovým jevem. V takovém případě na kovu a jeho výstupní práci nezáleží, odpor kontaktů je malý s lineární napětíovou závislostí, teplotně málo závislý. Zatímco tedy usměrňujícího kontaktu kov-polovodič dosáhneme použitím málo dotovaného polovodiče se širokou oblastí prostorového náboje a relativně vysokou potenciálovou bariérou (obr.2.23a), ohmický kontakt lze spolehlivě realizovat na silně dotovaném (degenerovaném) polovodiči (obr.2.23b). V praxi se proto běžně pro vytvoření kontaktu používá struktury MN^+N a MP^+P , tj. pod kovovým kontaktem se vytvoří oblast silně dotovaného polovodiče téhož typu vodivosti.

3 POLOVODIČOVÉ DIODY

Polovodičové diody jsou elektronické součástky, které obsahují zpravidla jeden přechod PN nebo MS. Proto pro jednotlivé typy polovodičových diod platí vše co bylo řečeno v kapitole 2. V dalším textu budeme polovodičové diody označovat zkráceně jako diody (existují ještě vakuové diody a diodový charakter mají i dnes již zapomenuté elektrolytické usměrňovače a tzv. rtuťové ventily).

Někdy obsahují diody více přechodů, ale pro určení vlastností součástky je rozhodující přechod jeden (je nositelem nelinearity) a ostatní mají pomocný význam, na straně druhé tzv. Gunnova dioda neobsahuje přechod žádaný.

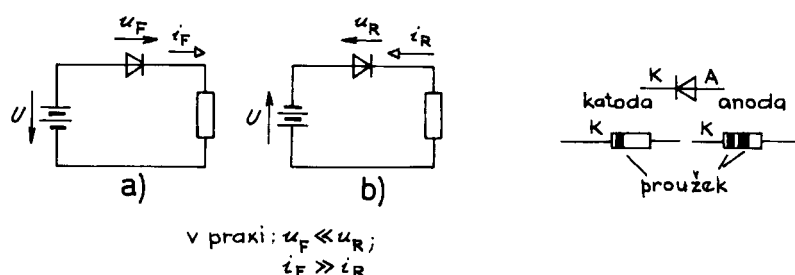


Obr. 3.1. Polovodičová dioda: a) orientace obvodových veličin, b) ampérovoltová charakteristika, c) indexy proudů a napětí pro propustný a závěrný směr (F = forward, R = reverse)

Na obr. 3.1 je nakreslena v řezu plošná dioda a její schematická značka. Stykem polovodiče typu N a P vznikne již

zmíněný polovodičový přechod. Bez přivedeného napětí je vnitřní potenciál styku obou polovodičů vyrovnán a dioda se navenek chová jako kondenzátor s konečným svodovým odporem (jakostí). Podstatou činnosti diod je usměrňovací jev. Tento jev je způsoben větší vodivostí v jednom směru oproti směru druhému.

Zopakujeme: Připojí-li se k diodě vnější stejnosměrné napětí tak, že kladný pól je připojen k polovodiči typu N, zvětší se potenciální rozdíl na přechodu a diodou proud téměř neprochází. Říkáme, že dioda je polarizována v závěrném směru. V opačném případě dioda vede (říkáme, že je polarizována v propustném směru nebo také v předním směru). Této vlastnosti diody se využívá k usměrňování.



Obr. 3.2. Základní zapojení diody ve směru propustném (vlevo) a závěrném (vpravo), označení vývodů diody

několik desetin voltu, proud diodou prudce roste. Je-li dioda polarizována v závěrném směru, její závěrný proud je o několik řádů menší než v propustném směru a se zvyšujícím se napětím roste. Pokud se nepřekročí dovolené závěrné napětí diody, je závěrný proud vůči propustnému zanedbatelný.

Na obr. 3.2 je uvedena polarita přiloženého napětí pro zapojení diody ve směru propustném a závěrném. Podle analogie s elektronkami se vývody diody (a také vrstvy, ke kterým jsou připojeny) nazývají **anoda** a **katoda**. Anodou je vývod diody připojený k vrstvě P, tj. po přiložení kladného napětí na anodu je dioda polarizována v propustném směru.

Na obr. 3.3 je opět nakreslena AV charakteristika diody, kde je vyznačena propustná a závěrná část a obvodová zapojení, ve kterých se obě části dají změřit. Obě části AV charakteristiky jsou kresleny v různých měřítkách se zřetelem k tomu, že jak závěrný proud k proudu propustnému, tak i propustné napětí k napětí závěrnému bývají ve vzájemném poměru více jak 1:100. Při stejném měřítku pro obě části by byl průběh charakteristiky v oblasti nuly (průsečíku os) plynulý.

Základní vlastnosti diod se dají nejlépe znázornit na jejich voltampérových charakteristikách. Typický průběh statické charakteristiky diody je na obr. 3.1. Charakteristika má dvě části: **závěrnou** část a **propustnou** část.

Zvyšuje-li se kladné napětí na anodě diody nad

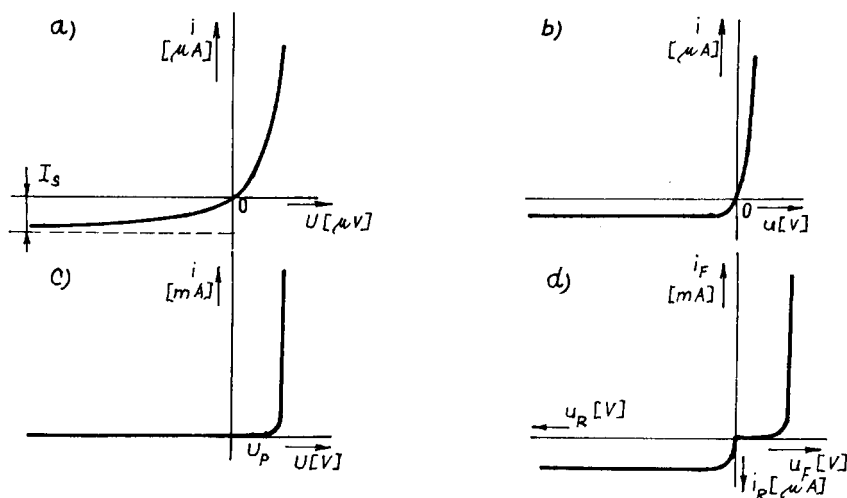


Obr. 3.3. Ampérvoltová charakteristika polovodičové diody. Obě části charakteristiky jsou kresleny v různých měřítkách se zřetelem k tomu, že jak závěrný proud k proudu propustnému, tak i propustné napětí k napětí závěrnému bývají ve vzájemném poměru 1:100 i více. Při stejném měřítku pro obě části by byl průběh v oblasti počátku plynulý (viz také obr. 3.4)

Nahoře: Polarita napětí pro závěrný a propustný směr (mnemotechnik: **P**lus na **P** [na anodě])

Dole: Měření závěrné a propustné části (větve) ampérvoltové charakteristiky

Pro lepší porozumění AV charakteristikám diod uváděných v různé literatuře si prohlédněte obr. 3.4, kde jsou na osách použita různá měřítka. Pro praxi je účelná charakteristika dle obr. 3.4.c, na níž je vidět, že dioda se prakticky otvírá až při prahovém napětí U_P (pro Si asi 0,6V a pro Ge asi 0,2V). Charakteristika dle obr. 3.4.d se uvádí v katalozích součástek. Na obr. 3.4.e je vidět, že závěrný proud se "nasytí" při poměrně malém závěrném napětí (asi 0,1 V). Další zajímavou vlastnost exponenciální funkce ukazuje obr. 3.5, kde uvažujeme pouze propustnou část charakteristiky. Exponenciální charakter závislosti způsobí, že při různých měřítkách proudu na svislé ose budou charakteristiky „podobné“.



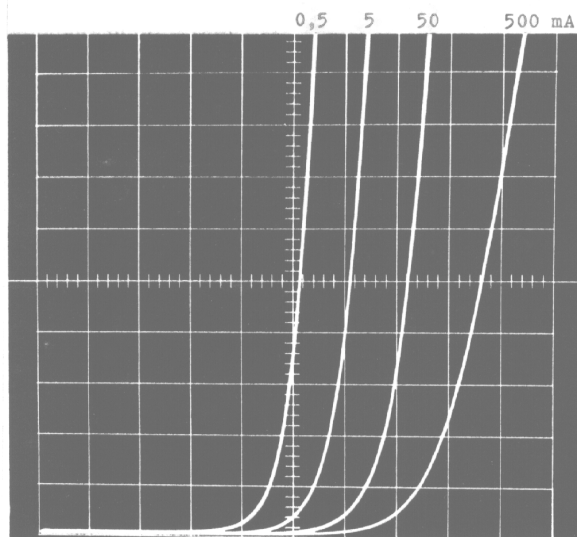
Obr. 3.4. AV charakteristika přechodu PN: **a)** oblast kolem počátku, **b)** hrubší měřítko napětí, **c)** hrubší měřítko pro proud i napětí, **d)** oddělená charakteristika pro propustný a závěrný stav

V praxi dělíme diody podle různých hledisek, která se vzájemně překrývají. Z hlediska **konstrukce** rozeznáváme diody

- hrotové,
- plošné (slitinové, difúzní, planární a Schottkyho) a
- na rozhraní jsou tzv. mikroplošné diody.

Důležitější je kombinované hledisko **funkce** a možné **použití**, podle kterého rozeznáváme diody

- usměrňovací (nízkovýkonové a výkonové),
- detekční a spínací,
- stabilizační a referenční (tzv. Zenerovy),
- tunelové a inverzní,
- kapacitní (varikapy a varaktory),
- luminiscenční (svítivé a laserové),
- fotodiody,
- speciální diody (mikrovlnné diody apod.)

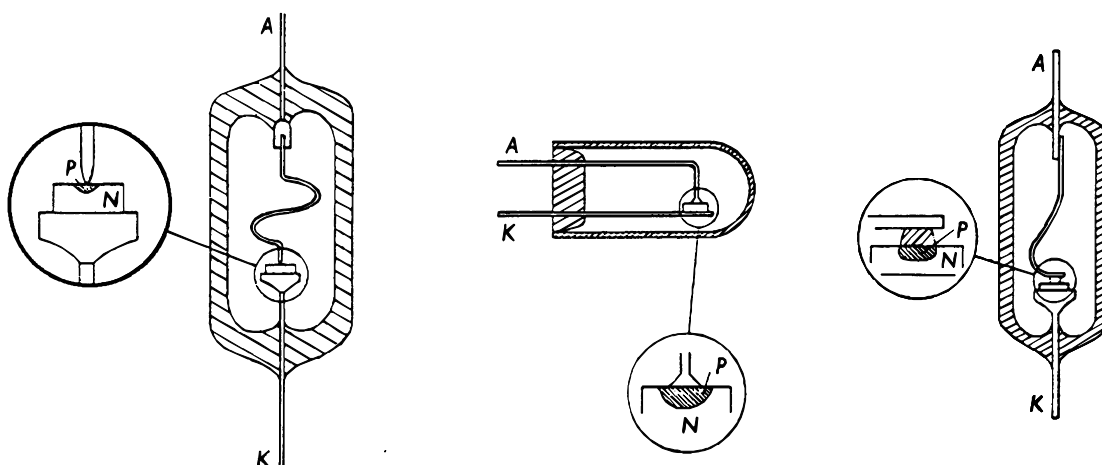


3.1 Hrotové diody

Hrotové diody zdánlivě využívají přechodu kov-polovodič. Skutečnost je však složitější, většinou jde o miniaturní přechod PN, pouze někdy o skutečný přechod kov-

Obr. 3.5. Grafy znázorňující charakteristiku diody ve čtyřech různých měřících proudů (viz čísla nahoře)

- je vidět posun kolena charakteristiky (relativnost prahového napětí diody)

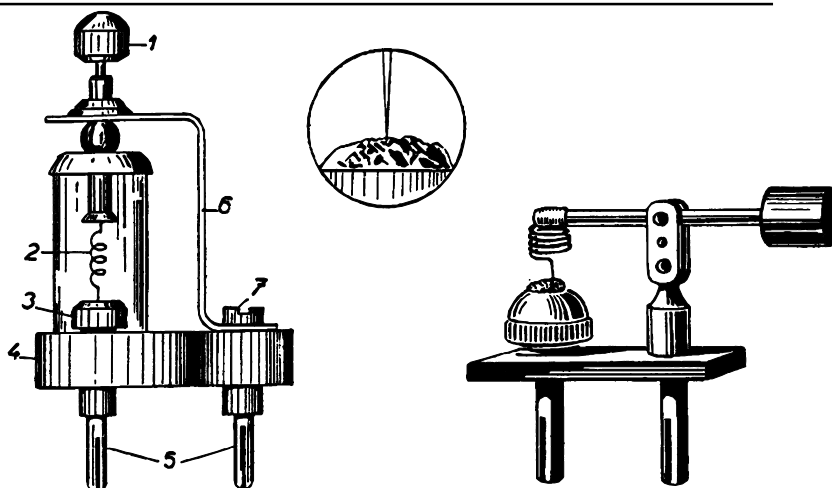


Obr. 3.6. Polovodičové diody: a) hrotová, b) s přivařeným zlatým hrotem, c) plošná polovodič. Hrotové diody je možno v podstatě rozdělit na:

- a) vysokozávěrné vf diody ($U_{(BR)} > 10 \text{ V}$),
- b) mikrovlnné diody ($U_{(BR)} \cong 2 \text{ až } 10 \text{ V}$),
- c) diody s přivařeným zlatým hrotem,
- d) z historie známý galenitový detektor (lidově zvaný "krystal") používaný v prvních krystalových přijímačích (krystalkách) obr. 3.7.

Hrotové germaniové diody se vyrábějí z germania typu N. Čtvercová, opracovaná destička Ge rozměru asi $1 \times 1 \times 0,1 \text{ mm}$ se připájí k držáku, který se umístí do skleněné trubičky. Proti ní se posunuje držák s esovitě zahnutým wolframovým drátkem s hrotem. V okamžiku, kdy dojde k přitlačení hrotu do germánia (dojde k elektrickému kontaktu), posun drátku se zastaví a skleněná trubička se na obou koncích zataví. Takto vzniklou diodou je třeba formovat. Dioda se formuje elektrickým impulsem v propustném směru. Vlivem značné teploty se pod hrotem pozmění krystalická struktura. V krystalu vzniknou dislokace, které dají malé oblasti vlastnosti polovodiče typu P. Jde tedy o diodu s přechodem PN, přičemž wolframový hrot

tvoří jen mechanický kontakt s oblastí P. To bylo dokázáno tím, že po odstranění wolframového drátku z diody a jeho nahrazení jiným drátkem měla dioda původní vlastnosti, aniž byla znovu formována. Hrotové diody obdobných vlastností se vyráběly i z křemíku typu N s hliníkovým drátkem, jejich propustná charakteristika je posunuta doprava.



Obr. 3.7. Galenitový detektor: uzavřený (vlevo) a otevřený (vpravo)

Konstrukce hrotové diody je znázorněna na obr. 3.6a. V

důsledku malé plochy přechodu se hrotová dioda vyznačuje

- malou parazitní kapacitou (≤ 1 pF),
- malým ztrátovým výkonem (cca 10 mW),
- malou hodnotou maximálního proudu (do 10 až 20 mA),
- velkým sériovým odporem,
- krátkou zotavovací dobou (< 1 μ s).

Protože přechod PN má kapacitu přibližně 1 pF, mohou tyto diody usměrňovat napětí asi do 100 MHz. Je tu však ještě jedna charakteristická protichůdná závislost, a to vztah mezi mezním kmitočtem a závěrným napětím. Čím je závěrné napětí větší, tím je menší mezní kmitočet. Závěrné napětí těchto diod se pohybuje od -10 do -200 V. Rozsah pracovních teplot -40 až +80 °C.

Křemíkové hrotové diody pro cm vlny (tzv. mikrovlnné diody) se vyrábějí pro usměrňování malých střídavých napětí do kmitočtu 10 GHz. Od běžných Si hrotových diod se liší v geometrickém uspořádání pouzdra (bezvývodové patronové nebo koaxiální), aby bylo dosaženo minimálních hodnot C a L a dále ve volbě základního polovodičového materiálu. Křemíková destička na kterou je přitlačován wolframový hrot, má typ P s velmi malým měrným odporem, tím se zvýší jejich mezní kmitočet, ale současně poklesne závěrné průrazné napětí. Průběh propustné části charakteristiky je přibližně *kvadratický* a této vlastnosti se využívá u kvadratických detektorů např. pro měření vysokofrekvenčních výkonů z usměrněného proudu, který je úměrný druhé mocnině intenzity elektrického pole.

K usměrňování dochází na styku kovového hrotu s polovodičem. Nejde tu tedy o přechod PN jako u germaniové diody. Dioda se formuje mechanickým nárazem. Mikrovlnné diody lze zatížit nárazem řádu 10^{-8} až 10^{-7} Ws. Protože mají malé závěrné napětí, musí být chráněny před indukovaným napětím, proto se balí do hliníkových fólií. Nehodí se například do rozhlasových přijímačů, které pracují s větší amplitudou vf signálu.

Poznámka: Mikrovlnná technika byla v podstatě první oblastí elektroniky, do které polovodiče pronikly. Z této doby pochází název radarové diody.

Poznámka: Přestože bychom právě u mikrovlnných diod předpokládali použití germania, které má vyšší pohyblivost nosičů, používá se převážně křemíku typu P, a to čistě z technologických důvodů. Podobně i u dalších součástek se užívá materiálů nebo typů vodivosti vrstev, pro které nejsou fyzikální, ale technologické důvody.

Germániové diody s přivařeným zlatým hrotem mají lepší mechanické i elektrické vlastnosti než diody s hrotem wolframovým. Zlatý hrot s přísadou galia (asi 3%) se přitlačí ke krystalu germánia a definovaným proudovým impulsem se přivaří (wolframový hrot se nepřivařoval). V místě sváru proniknou galiové atomy do Ge a po ochlazení vznikne miniaturní slitinový přechod PN. Vlastnosti v propustném i závěrném směru jsou výrazně lepší než u diody s wolframovým hrotem. Kapacita přechodu těchto diod bývá asi 2 pF.

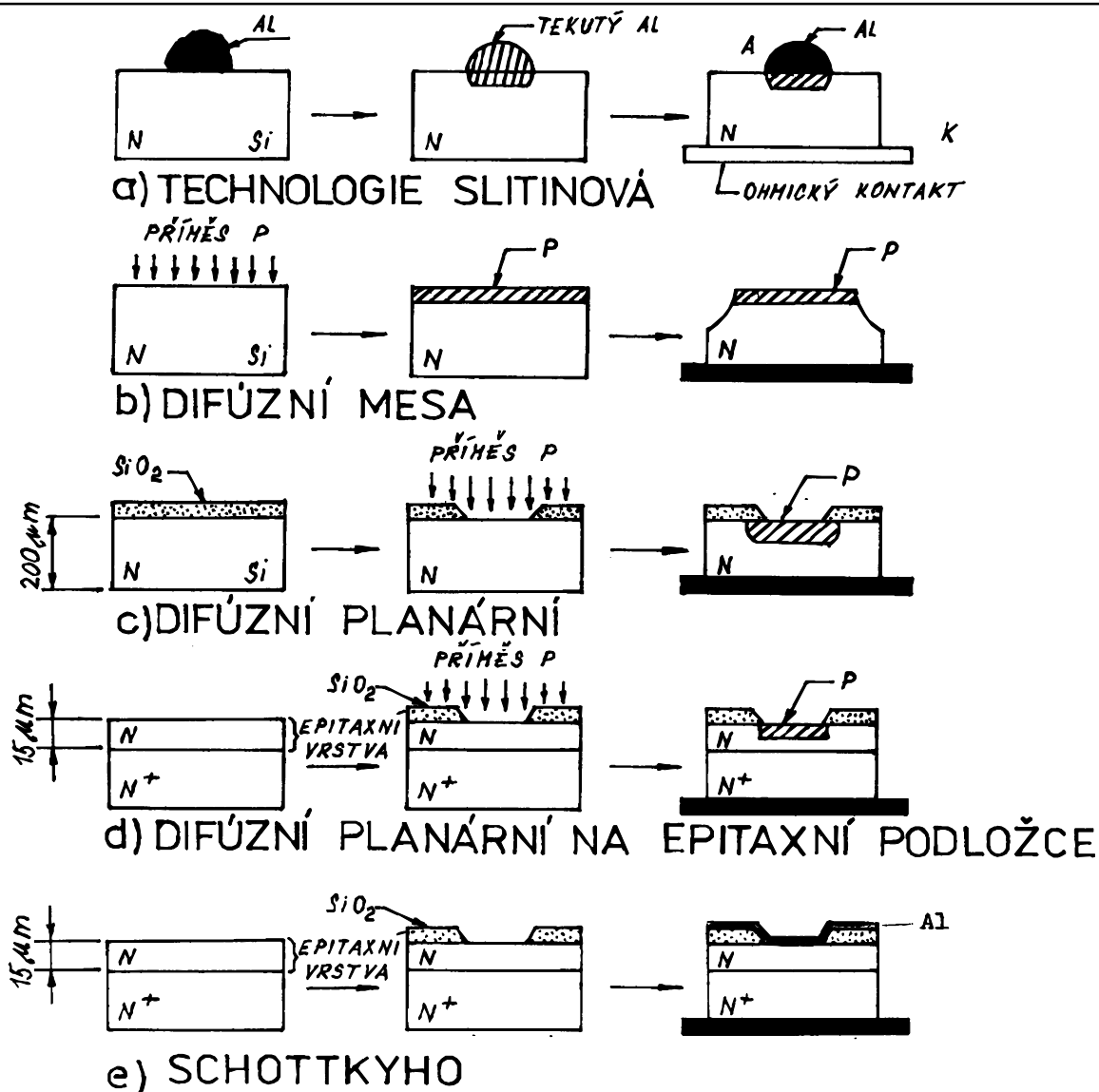
Poznámka: Hrotové diody kromě mikrovlnných, jsou dnes již neperspektivní a nevyrábí se.

3.2 Plošné diody

První technologií výroby plošných diod byla technologie slitinová. Na základní destičku monokrystalu např. typu N se položí legující materiál (indium pro typ P) ve tvaru kuličky, válečku apod. Destička se uloží do žíhací pece s inertní atmosférou a zahřívá se na požadovanou teplotu. U Ge je to 550 °C, u Si 1400 °C. Legující látka se roztaví a slije se s částí základního materiálu, po vychladnutí a rekrytalizaci zůstane část legujícího materiálu v destičce, vytvoří inverzní (obrácenou) oblast vůči původní a vznikne přechod PN. Přechod akceptorů z oblasti P do oblasti N je strmý. Touto technologií je obtížné v sériové výrobě zhotovit přechody PN se stejnými vlastnostmi, protože různé rozdíly vah kuliček, míra nečistoty, teplotní rozdíly a některé další faktory silně ovlivňují elektrické parametry přechodu.

Nejčastěji se přechod PN zhotovuje difúzí příměsí z plynné nebo kapalně (někdy i pevné) fáze do základní destičky při vyšších teplotách. Např. Si destička s požadovanou vodivostí N se vloží do plynné atmosféry s teplotou 1200 °C a nechá se působit dotující plyn (sloučenina fosforu pro P) tak dlouho, až dotující molekuly proniknou do požadované hloubky, která bývá 1 až 15 μm. Na rozhraní obou prostředí vznikne přechod PN, který je pozvolný. Touto technologií lze zhotovit reprodukovatelné přechody PN s menším rozptylem parametrů, protože lze řídit přesně hloubku difúze v delším časovém intervalu. Příměsí z plynné fáze pronikají při teplotě 1200 °C u Si do hloubky 1 μm přibližně za hodinu a tak lze při dostatečné čistotě látky, dodržení teploty ($\pm 0,5$ K) a času dosáhnout při difúzním přechodu malého rozptylu parametrů.

V případě, že požadujeme malou kapacitu přechodu, můžeme difúzní přechod odleptat a tím zmenšit rozměry ochuzené oblasti. Tak vznikne systém s typickým výstupkem, který připomíná obrys Stolové hory, nazývané Mesa (původem španělský název). Uvidíme později, že tento tvar přechodu dovoluje docílit velké průrazné napětí, a proto se používá u výkonových diod.



Obr. 3.8. Různé technologie výroby plošných diod (na obr. b až d není pro přehlednost nakreslen napařený hliník, který tvoří vývod horní elektrody, na obr e) musí být zakreslen, protože jde o funkční vrstvu)

Slitinová a difúzní technologie mají společnou nevýhodu, a to tu, že okraj přechodu PN je obnažen a vlivem jeho vnitřního elektrického pole, pronikajícího na povrch krystalu, zachycuje atmosférické nečistoty, což vyvolává nestabilitu a snižuje průrazné napětí. Proto se přešlo k dokonalejšímu způsobu vytváření přechodu, a to *selektivní difúzí s oxidovými maskami*. Křemíková destička v atmosféře vodních par zoxiduje (narůstá vrstva SiO₂). Oxid křemíku SiO₂ má maskovací vlastnosti vzhledem k některým plynům tj. zabrání difúznímu průchodu těchto plynů do Si destičky. Do krycí oxidové vrstvy se vyleptá otvor a tímto otvorem se potom provede na požadovaném místě (tj. selektivně) příslušná difúze. Před leptáním se pokryje oxid fotocitlivou želatinovou emulzí (fotorezistem), ultrafialovou výbojkou se osvětlí přes masku místo, které se nemá leptat. Světlo vytvrdí lak, na neexponovaném místě se rozpouštědlem odstraní fotorezist a leptadlem se vyleptá v kysličníku otvor. Přechod PN je přikrytý oxidem a tím chráněn (pasivován), čímž se zvětší stabilita parametrů. Otvor se potom překryje napařeným hliníkem a přiboduje se vývodní elektroda (obvykle Au drát). Pro vznik přechodu PN bylo nutné tedy použít litografické techniky, maskování oxidem a selektivní difúze. Vzniklý přechod PN je tedy na povrchu chráněný, označuje se jako *planární přechod*, a příslušná technologie výroby se nazývá planární.



Obr. 3.9. Výkonová křemíková difúzní dioda s odleptanými boky (mesa)

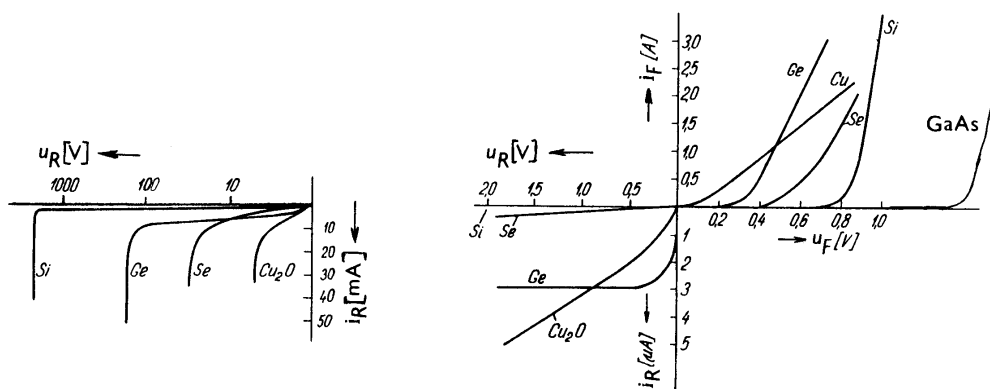
Poznámka: Důležitou výhodou difúzní a planární technologie je, že při jednom výrobním cyklu lze vyrábět velké množství - několik desítek až set diodových systémů najednou.



Obr. 3.10. Křemíková planární dioda s epitaxní vrstvou

Systémy zhotovené v jednom cyklu mají pak velmi podobné vlastnosti (malý rozptyl parametrů). To u slitinové technologie nebylo možné.

Nosiče proudu musí mezi elektrodami (obr. 3.8c) urazit krátkou vzdálenost oblastí P a velmi dlouhou dráhu oblastí N. Tato dlouhá dráha se projevuje především ve formě parazitního odporu, který je nežádoucí. Proto je vhodné zhotovit tuto oblast buď krátkou, což nelze jednoduše z důvodů mechanické pevnosti materiálu, nebo silně dotovanou, aby stoupla vodivost materiálu. Druhé řešení je nevyhovující pro případy větších závěrných napětí (obr. 3.14) a proto se zavádí další technologické opatření, a to *epitaxní růst*. Na základní destičku polovodiče o dostatečné vodivosti s koncentrací N^+ se nechá narůst se stejnou krystalografickou strukturou polovodič o požadované koncentraci N a tloušťce d a do této sestavy se potom vytvoří přechod PN. Na obr. 3.8d je znázorněn difúzní planární přechod na epitaxní podložce, zkráceně zvaný *epitaxně - planární přechod PN*, vytvořený cestami fotolitografie, maskování oxidem, leptání a selektivní difúze na epitaxní podložce.



Obr. 3.11. Porovnání charakteristik polovodičových diod z různých materiálů



Obr. 3.12. Porovnání propustných charakteristik různých typů diod

Dalším typem plošné diody je **Schottkyho dioda**, která využívá přechodu **kov - polovodič**. Je-li výstupní práce elektronu z kovu $q_e\phi_m$ větší než výstupní práce $q_e\phi_s$ elektronu z polovodiče typu N, takže platí nerovnost $\phi_m > \phi_s$, elektrony při styku obou látek pronikají z polovodiče do kovu. V kovu vzhledem k jeho vodivosti se rozptýlí a nevytvoří oblast prostorového náboje. V polovodiči vznikne v blízkosti kovu tak silné odčerpání elektronů, že převládnu co do počtu minoritní nosiče, dochází ke změně vodivosti z N na P, k tzv. *inverzi vodivosti*. V polovodiči tedy vznikne přechod s jednosměrným ventilovým účinkem. Po přiložení napětí v propustném směru (kladné napětí na kovu) přecházejí elektrony jako majoritní nosiče přechodem do

kovu, kde jsou zase majoritními nosiči a vlivem jeho vodivosti se ihned rozptýlí a nevytvoří nadbytečné nosiče. Po změně polarit neexistuje prostorový náboj, který by mohl přecházet z kovu do polovodiče, teoreticky by měla být zotavovací doba $t_{rr} = 0$. V praxi se ovšem projevují některé parazitní jevy, které způsobují hromadění náboje na rozhraní kov - polovodič a tak se objeví náboj s pikosekundovým zotavovacím časem.

Schottkyho diody se buď zhotovují z křemíku nebo GaAs. Diody z GaAs mají lepší vf i spínací vlastnosti, jsou však dražší, a proto se používají pro náročnější aplikace. V případě, že se kovová elektroda nahradí zlatou elektrodou, AV charakteristika v propustném směru má menší úbytek napětí, je lepší než dioda se zlatým hrotem a o něco horší než inverzní dioda. Tato dioda je označována jako dioda ZBS (Zero - Bias - Schottkyho).

Schottkyho diodu můžeme považovat za unipolární prvek, u kterého se na transportu nosičů podílejí jen elektrony jako majoritní nosiče. Do křemíku typu N nepřecházejí z kovu nosiče a tím nedochází k jejich skladování v polovodiči. Mají malý dynamický odpor v propustném směru a jako vf diody mají velké závěrné napětí. Používají se především jako rychlé spínače, detektory v oblasti cm vln a usměrňovače.

Plošné diody jsou různé a v mnohém se od sebe liší. Není lhostejné z jakého materiálu a jakou technologií je taková dioda vyrobena. Dnes se užívají diody z křemíku, někdy z GaAs nebo germania, dříve i diody selenové nebo kuproxidové.

Kuproxidové a selenové diody jsou polykrystalického typu, mají poměrně velký vnitřní odpor a proto velký úbytek napětí; charakteristiky uvedené na obr. 3.11 se u nich vztahují na plochu 10 cm^2 . Křemíkové a germaniové diody jsou naopak monokrystalické, mají malý vnitřní odpor a tím malý úbytek napětí v propustném směru a uvedené charakteristiky jsou vztaženy na plochu 1 mm^2 , což znamená že hustota proudu je (může být) u křemíkových a germaniových diod až tisíckrát větší.

V závěrném směru jsou charakteristiky na obr. 3.11 vztaženy na plochu 1 mm^2 . Pro přímé porovnání s měřítkem v propustném směru je zapotřebí u selenových a kuproxidových usměrňovačů přečtené údaje násobit tisícem, tj. číst na měřítku proudu údaje v mikroampérech jako údaje v miliampérech.

Kuproxidové diody mají malé prahové napětí a celkem plynulý přechod mezi závěrným a propustným směrem. Maximální závěrné napětí je však nejnížší ze všech uvedených diod. Pro

výhodný průběh charakteristiky se jejich v minulosti používalo v měřicích přístrojích.

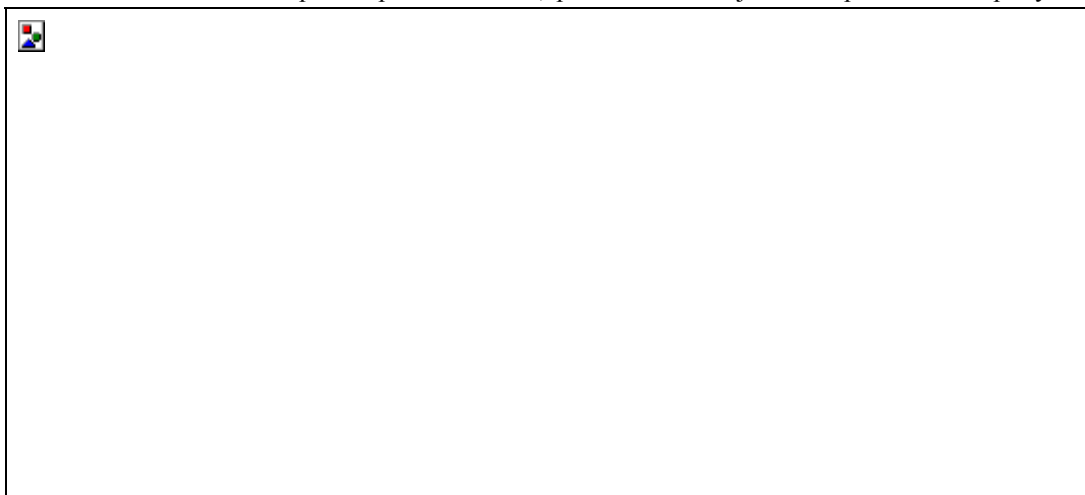
Selenové diody mají větší prahové a závěrné napětí. Jak kuproxidové, tak i selenové diody v provozu stárnou, tj. jejich vnitřní odpor se zvětšuje. Protože jejich stárnutí závisí na provozní teplotě, která pak stoupá následkem zvyšování ztrát při větším úbytku napětí, stárnutí se stále zrychluje.

U germaniových a křemíkových diod je stárnutí nevýznamné. Nevýhodou germaniových a křemíkových diod je však značná citlivost na přepětí. Selenové diody měly totiž mnohem větší plochu a hmotnost a tedy i tepelnou setrvačnost a odolávaly krátkodobému proudovému i napětíovému přetížení.

Dnes se používají v podstatě jen křemíkové diody. Germaniové diody se používaly ještě nedávno v oblastech, pokud bylo usměrněné napětí malé a bylo výhodné malé prahové napětí, bylo to hlavně v napětíových měřicích obvodech. Dnesse v těchto oblastech užívá Schottkyho diod. Křemíkové diody mají větší úbytek napětí v pro-pustném směru (větší prahové napětí), avšak závěrné napětí je několikanásobné v porovnání s diodami germaniovými a proud v závěrném směru je je mnohem menší. Také jejich pracovní teplota může být až 150 °C.

V různých obvodech (např. měřicích) má velký význam, jak se projevuje změna teploty na vlastnosti diod. Vliv teploty na průběh typické germaniové a křemíkové diody je na obr. 3.13. Vidíme především, že nasycený proud

v závěrném směru s teplotou prudce vzrůstá, přibližně na dvojnásobek při vzrůstu teploty o 10 °C (v různé



Obr. 3.13. Vliv teploty na průběh charakteristik germaniové diody GAZ51 (vlevo) a křemíkové diody KA262 (vpravo)

literatuře se objevují různé hodnoty, obecně je tento růst více patrný u germaniových diod, protože je řádově větší než u diod germaniových).

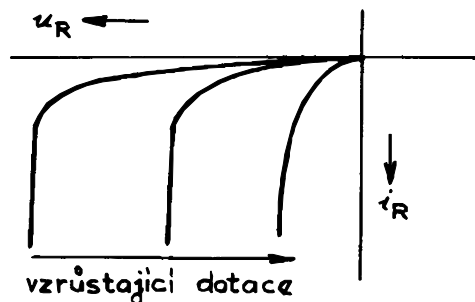
Prahové napětí v propustném směru se při stoupající teplotě zmenšuje tak, že se úbytek napětí pro určitý propouštěný proud zmenší asi o 0,8 % na 1 °C u germaniové diody, o 0,2 % na 1 °C u křemíkové diody, absolutně je to však hodnota přibližně stejná 2,2 mV/°C.

3.2.1 Detekční a spínací diody

Pro usměrňování elektrických signálů vyšších kmitočtů s malou amplitudou napětí se používá diod s malou plochou přechodu PN (nebo přechodu kov - polovodič), navrhovaných pro malá závěrná napětí (do 200 V) a malé proudy (stovky mA) (Jako vyšší kmitočty rozumíme kmitočty vyšší než tzv. průmyslové kmitočty 50, 60, příp. 400 Hz; pro odlišení hovoříme o jejich detekci či demodulaci). Tyto diody označujeme nejčastěji jako detekční (někdy signálové). Zhotovují se buď jako diody hrotové, s přivařeným hrotem nebo plošné. Za perspektivní považujeme pouze Schottkyho diody (a mikrovlákné diody, jak již bylo uvedeno) a užívají se také křemíkové plošné detekční diody (planárně epitaxní) označené jako diody pro všeobecné použití. V některých případech se používají tzv. inverzní diody (viz později).

Spínací diody se používají v obvodech pro spínání elektrických signálů, pracují většinou se signály obdélníkového průběhu a přecházejí ze stavu vodivého do nevodivého.

Použitelnost diod pro spínání je omezena setrvačností diody při přepnutí z vodivého stavu do nevodivého. Příčinou je nahromadění nadbytečných nosičů náboje v blízkosti přechodu vlivem injekce (storage effect - „skladovací jev“). Pokud toto nahromadění trvá, teče diodou větší proud v závěrném směru než odpovídá rovnovážnému stavu a dioda je tedy vodivá.



Obr. 3.14. Vliv vzrůstající dotace na průřezném napětí přechodu PN

Určitými technologickými opatřeními při výrobě lze dosáhnout snížení doby života minoritních nosičů náboje. Také zmenšením bariérové kapacity přechodu je možné snížit zotavovací dobu, a tím rovněž zvýšit rychlost přepínání diod. Typickým znakem spínacích diod je tedy malá plocha přechodu PN a krátká doba života nadbytečných nosičů náboje.

Z funkce spínacích obvodů vyplývá, že parametry, kladenými na spínací diody budou závěrná zotavovací doba t_{TR} , charakterizující rozepnutí diody, dále největší propustný proud I_F (buď trvalý nebo časově omezený) a největší závěrné napětí U_R . Za spínací diody obvykle považujeme ty druhy, u nichž je

$t_{TR} < 500 \text{ ns}$. V minulosti se pro spínání používaly především germaniové diody s přivařeným hrotem. V současnosti se užívají křemíkové plošné diody. Podle požadavků praxe se dělí do několika skupin dle tab. 3.1.

3.2.2 Stabilizační a referenční diody

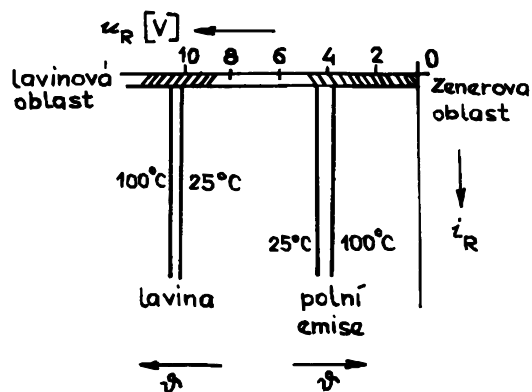
Stabilizační (Zenerova) dioda je křemíková plošná dioda s ostrým zlomem (průrazem) závěrné části AV charakteristik. Pracovní oblast těchto diod leží v oblasti elektrického nedestruktivního průrazu. Při (téměř) stálém napětí narůstá proud. Jeho velikost je omezena prakticky pouze odporem vnějšího obvodu. Jak již bylo uvedeno, průraz může být tunelový (Zenerův) nebo lavinový (v obou případech hovoříme o Zenerových diodách, název lavinová dioda se užívá pro jiné prvky).

Dovolенý ztrátový výkon P_Z spolu s průrazným napětím udává maximální dovolený proud diodou. Diferenciální odpor Zenerových diod v oblasti průrazu závisí na U_Z (je o něco menší v ohmech než je průrazné napětí ve voltech) i na teplotě.

Vzhledem k tomu, že se Zenerovy diody nejčastěji používají při závěrné polarizaci přechodu PN, může se uplatnit bariérová kapacita přechodu. Tato kapacita je značná ($C_0 = 200 \text{ pF}$ až 10 nF).

Kromě uvedených základních vlastností je jednou z ne-příjemných vlastností Zenerových diod šum v oblasti přechodu charakteristiky ze závěrné do průrazné oblasti. Tento šum se uplatňuje především u Zenerových diod s $U_Z > 6 \text{ V}$, tedy u diod, které využívají pro svou činnost lavinového průrazu. Šum je způsoben nespojitostí charakteristiky v oblasti nasazení průrazu (při proudu pod $100 \mu\text{A}$) v důsledku postupných místních mikropřerazů. Při vyšších hodnotách proudu se průraz stává stabilním a šum mizí. Amplituda šumu může dosáhnout úrovně několika milivoltů. Kmitočtové spektrum tohoto šumu začíná od velmi nízkých kmitočtů a je do kmitočtů řádově 1 MHz . Uvedeného jevu je možné využít ke konstrukci velmi jednoduchých a spolehlivých generátorů šumu.

Stabilizační diody používáme ke stabilizaci napětí při kolísání vstupního napětí nebo při proměnné zátěži (nebo obou). Vliv teploty zanedbáváme.



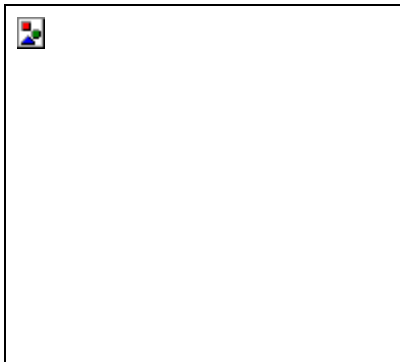
Obr. 3.15. Vliv teploty na závěrnou charakteristiku při různých mechanismech průrazu

Tab. 3.1. Druhy křemíkových spínacích diod

	$t_{TR} [\text{ns}]$	$I_F [\text{mA}]$	$U_R [\text{V}]$
Pomalé	$> 1 \mu\text{s}$		
Rychlé	4	10 - 100	20 - 50
Velmi rychlé	0,7	10	10
Pro velké proudy	4 - 10	100 - 400	30 - 50
Vysokonapět'ové	50 - 100	10 - 100	500 - 250

Referenční diody mají vykompenzovaný teplotní koeficient průrazného napětí - v principu sériovým zapojením diody v propustném směru (s teplotou napětí na stabilizační diodě stoupá, na kompenzační klesá). Touto kombinací ovšem stoupá diferenciální odpor. Protože však referenční diody nebývají zatěžovány proměnným proudem, tato vlastnost není vždy na závadu (vstupní napětí bývá předstabilizováno a zátěž je konstantní), nebo má malý proudový odběr.

Shrňme tedy:



Obr. 3.16. Zapojení referenčních diod TESLA KZ 745 až 747 (nahore) a KZ 781 až 783 (dole)

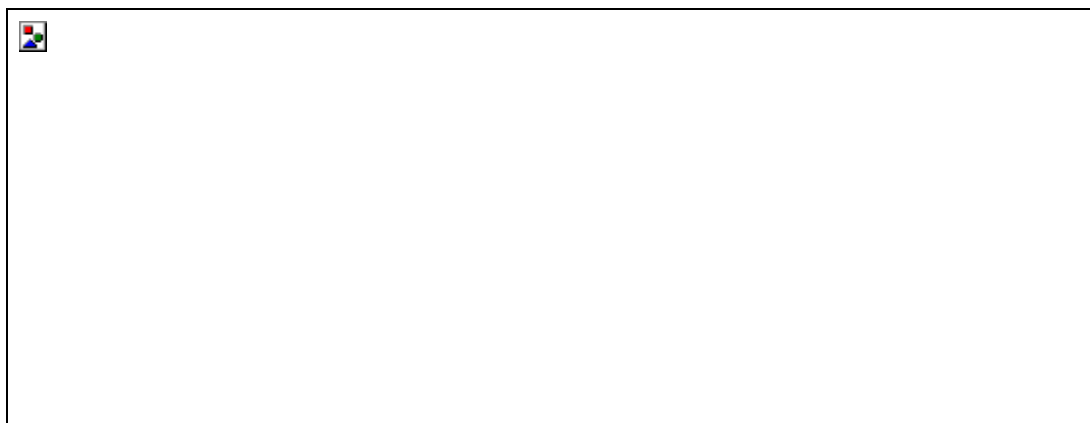
a) *stabilizační diody* jsou ty diody, které vytvoří na svých vývodech konstantní napětí, jestliže proud procházející diodou je ve stanoveném rozsahu;

b) *referenční diody* jsou ty diody, které vytvoří na svých vývodech referenční napětí stanovené přesnosti, jestliže proud procházející diodou a teplota okolí je ve stanoveném rozsahu.

Na obr. 3.17 je uvedeno zapojení Zenerovy diody jako paralelního stabilizátoru. Zamezíme-li předřadným rezistorem vzrůst závěrného proudu Zenerovy diody nad dovolenou mez (I_{\max}), můžeme tuto diodu použít jako stabilizátor napětí (nebo také omezovač nebo zdroj referenčního napětí). Zenerovy diody se vyrábějí pro stabilizaci napětí zhruba od tří voltů do dvou set voltů a pro zatížení od zlomků do desítek wattů. Pracovní oblast Zenerovy diody je vyznačena na obr. 3.18. Požadavek je, aby

dynamický r_D odpor byl co nejmenší (přímka za ohybem co nejstrmější).

Pracovní oblast je vymezena maximálním proudem I_{\min} a mezním proudem I_{\max} . Pro proudy $i_D < I_{\min}$ již nelze hovořit o stabilizaci napětí. Pro proudy $i_D > I_{\max}$ je překročena mezní výkonová ztráta diody. V pracovní oblasti $I_{\min} < i_D < I_{\max}$ lze definovat dynamický odpor diody r_D (obr. 3.18b). V ideálním případě se napětí u_D se změnou i_D nemění, tj. $\Delta u_D = 0$, $r_D = 0$. Jednoduchý statický model je na obr. 3.17.



Obr. 3.17. Zenerova dioda: a) zapojení používané ke stabilizaci (paralelní stabilizátor), ampérvoltová charakteristika, c) statické náhradní zapojení

Platí

$$u_D = U_Z + i_D r_D \quad (3.2)$$

pro $I_{\min} < i_D < I_{\max}$. Pokud není stabilizátor zatížen, tj. při $R_z = \infty$, je situace velmi jednoduchá, platí $i_1 = i_D$. Pro $u_1 < U_Z$ je $u_D \approx u_1$. Pro $u_1 > U_Z$ platí vztah (3.2), přičemž

$$i_1 = i_D = (u_1 - u_D)/R_1 = u_1/R_1 - (U_Z + r_D i_D)/R_1$$

Po úpravě dostaneme

$$i_1 = i_D = \frac{u_1 - U_Z}{R_1} \frac{1}{1 + r_D / R_1} \quad (3.3)$$

Dosadíme-li z (3.3) do (3.2) dostaneme po úpravách

$$u_D = U_Z + (u_1 - U_Z) \frac{r_D}{r_D + R_1} \quad (3.4)$$

Změně napětí Δu_1 proto odpovídá změna napětí

$$\Delta u_D = \Delta u_1 \frac{r_D}{r_D + R_1} \quad (3.5)$$



Obr. 3.18. Pracovní oblast Zenerovy diody a definice dynamického odporu

Je-li obvod zatížen ($R_z \neq \infty$), je vhodné upravit zapojení na obr. 3.19a podle obr. 3.19b a použít Théveniovu větu. Ekvivalentní svorkové napětí naprázdno je

$$u_{th} = u_1 \frac{R_z}{R_1 + R_z}$$



Obr. 3.19. Napětíové a proudové poměry ve stabilizátoru se zátěží

Ekvivalentní odpor R_{th} je dán paralelní kombinací R_1 a R_z

$$R_{th} = R_1 \frac{R_z}{R_1 + R_z}$$

Situace na obr. 3.19c je již shodná s dříve uvažovanou situací pro $R_z = \infty$. Musíme však udělat záměnu pro $u_1 \rightarrow u_{th}$

a $R_1 \rightarrow R_{th}$. Dioda nestabilizuje pro napětí $U_{th} = u_1 R_z / (R_1 + R_z) < U_Z$, tedy pro

$$u_1 < U_Z \left(1 + \frac{R_1}{R_z} \right) \quad (3.6)$$

Pro napětí větší dioda stabilizuje a platí

$$i_D = \frac{u_{th} - U_Z}{R_{th}} \frac{1}{1 + r_D / R_{th}} \quad \text{a} \quad u_D = U_Z + (u_{th} - U_Z) \frac{r_D}{r_D + R_{th}}$$

Po dosazení a úpravách dostaneme

$$i_D = \left(\frac{u_1 - U_Z}{R_1} - \frac{U_Z}{R_z} \right) \frac{1}{1 + r_D/R_1 + r_D/R_z}$$

$$u_D = U_Z + \left(\frac{u_1}{1 + R_1/R_z} - U_Z \right) \frac{r_D}{r_D + \frac{R_1}{1 + R_1/R_z}}$$

Platí potom, že změna napětí Δu_D je se změnou Δu_1 svázána vztahem



Obr. 3.21. Zenerovy diody KZZ7x pro zatížení 280 mW, je vyznačené jmenovité Zenerovo napětí a jeho tolerance

Obr. 3.20. Stabilizátor se Zenerovou diodou při změně vstupního napětí (vlevo) a při změně zatěžovacího odporu (vpravo)

$$\Delta u_D = \Delta u_1 \frac{r_D}{R_1 + r_D(1 + R_1/R_z)} \quad (3.7)$$

Je zřejmé, že v praxi platí téměř vždy $r_D \ll R_1/(1 + R_1/R_z)$ a $r_D/R_1 \ll 1$ a také $r_D/R_z \ll 1$. Potom

$$i_D \approx \frac{u_1 - U_Z}{R_1} - \frac{U_Z}{R_z}$$

$$u_D \approx U_Z + \left(\frac{u_1}{1 + R_1/R_z} - U_Z \right) r_D \left(\frac{1}{R_1} + \frac{1}{R_z} \right)$$

Důležité je určit napětí U_{1H} , při němž je proud diodou maximální, tj. $i_D = I_{\max}$. Dosadíme tedy do rovnice pro i_D veličiny U_{1H} a I_{\max} a dostaneme

$$I_{\max} \approx \frac{U_{1H} - U_Z}{R_1} - \frac{U_Z}{R_z}$$

a po úpravě dostaneme

$$U_{1H} \approx R_1 I_{\max} + U_Z \left(1 + \frac{R_1}{R_z} \right)$$

Je-li zátěž R_z připojena, potom stabilizátor podle obr. 3.20 pracuje v rozsahu vstupních napětí

$$U_{1D} \subset u_1 \subset U_{1H}$$

$$\text{kde } U_{1D} \approx U_z(1 + R_1/R_z)$$

Vstupní napětí by mělo být stále v tomto intervalu (obvykle to bude napětí na filtračním kondenzátoru usměrňovače). Pro přenos zvlnění ze vstupu na výstup platí v tomto intervalu vztah (3.7), pokud bude $R_1 \gg r_D(1+R_1/R_z)$ zjednoduší se vztah pro přenos zvlnění na

$$\frac{\Delta u_D}{\Delta u_1} \approx \frac{r_D}{R_1}$$

$$\text{Zmenší-li se napětí } u_1 \text{ pod velikost } U_{1D}, \text{ platí } u_D = u_1 \frac{R_z}{R_1 + R_z}$$

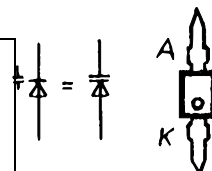
3.2.3 Kapacitní diody

Varikap (variable capacitor) a *varaktor* (variable reactor) označované také jako kapacitní diody, využívají kapacitního charakteru přechodu PN, který je navržen tak, aby činitel jakosti tohoto polovodičového kondenzátoru byl co největší, tj. R_S bylo malé, a aby závěrné napětí bylo veliké, dále aby průběh kapacity na napětí měl požadovaný tvar. Vyrábějí se z křemíku a jejich VA charakteristika je shodná s běžnou VA charakteristikou Si diody. Pracovní oblastí je závěrná část VA charakteristiky. Přivedením závěrného napětí se náboje v polovodiči přemístí a ochuzená vrstva se rozšíří přivedením kladného napětí se zúží. Změna kapacity na napětí lze aproximovat vztahem $C = C_0(1 - u/U_D)^{-1}$. Hodnota exponentu a je dána tvarem dotačního profilu. Je-li profil koncentrace nečistot lineární (difúzní), exponent má hodnotu $1/3$ a závislost $C = f(u_R)$ v logaritmickém měřítku je lineární. Je-li koncentrační profil strmý, odpovídající slitinové technologii, exponent je $n = 1/2$ a $C = f(u_R)$ je stále v logaritmickém měřítku lineární. Zhotovíme-li však hyperstrmý přechod (tj. přechod s inverzním gradientem příměsí), potom jeho závislost kapacity na napětí $C = f(u_R)$ v logaritmickém měřítku je nelineární.

Lineární průběh $C = f(u_R)$ má výhodu při hledání dvojic, trojic se stejnými vlastnostmi (souběh), protože stačí změřit 1 bod a podle něho vybrat příslušné varikapy. Při třídění hyperstrmých přechodů se musí měřit souběh ve 4 bodech.



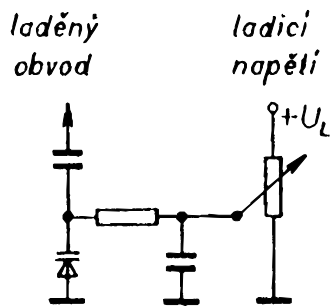
Obr. 3.24. Varikap: a) závislost kapacity na napětí, b) závislost činitele jakosti na napětí



Obr. 3.22. Schématická značka kapacitní diody varikapu a její provedení

Kapacitní diody jsou charakterizovány napětovým zdvihem U_{Rmax}/U_{Rmin} , který

bývá 30 V : 3 V a menší, a odpovídající kapacitní zdvih $C_{max} : C_{min}$, který se pohybuje u povlnných přechodů kolem hodnoty 2,5, u strmých přechodů asi 5 a u hyperstrmých kolem 20. Dále se uvažuje střídavé náhradní zapojení ve tvaru sériového článku $R_S C_S$. Jakost varikapu (stejně jako u kondenzátorů s pevným dielektrikem) se posuzuje činitelem jakosti $Q = (\omega C_S R_S)^{-1}$, a je závislá na kmitočtu.



Obr. 3.23. Zapojení varikapu v laděném obvodu LC

Shrňme nyní základní rozdíl mezi varikapem a varaktorem:

Varikap je kapacitní dioda, měnící hodnotu kapacity v závislosti na ss napětí. Přivedené vf napětí a jeho změny jsou relativně malé, takže nezpůsobí změnu kapacity. Je to tedy lineární kondenzátor, jehož kapacita se dá měnit pomocným (konstantním nebo pomalu se měnícím) napětím.

Varaktor je kapacitní dioda, pracující s tak velkým ω signálem, že během jeho periody dochází k výrazné změně kapacity. Varaktor se chová vůči ω signálu jako nelineární kondenzátor (vznikají vyšší harmonické složky proudu). Stejnoseměrná předpětí se zpravidla nepřivádí (vytváří se usměrněním ω signálu). Požadavkem je co nejvyšší nelinearita závislosti $C = f(u_R)$.

3.2.4 Tunelová a inverzní dioda

Tunelová dioda je založena na tunelovém přechodu elektronu z valenčního do vodivostního pásu (kap. 2). Na propustné větvi jeví AV charakteristiky je oblast s negativním diferenciálním odporem. Tunelová dioda má velmi úzký přechod PN (kolem 1 μm) s vysokou koncentrací příměsí (10^{19} cm^{-3}). Tunelové diody se považují za neperspektivní pro aplikace.

Inverzní dioda (obrácená dioda, tunelový usměrňovač, backward diode) je zvláštním druhem tunelové diody s nevyjádřenou (nebo jen slabě vyjádřenou) oblastí záporného diferenciálního odporu na AV charakteristice. V okolí počátku je proud ve zpětném směru větší než proud v přímém směru při stejné absolutní hodnotě přiloženého napětí. V přímém směru vykazuje velký odpor až do napětí asi $U_p = 450 \text{ mV}$ (Ge). Používá se k usměrňování malých napětí asi do 300 mV v měřicích přístrojích.



Obr. 3.25. Ampérvoltová charakteristika tunelové diody



Obr. 3.26. Charakteristika typu N (dynatronová, tetrodová) a typu S (oblouková, tyristorová) s oblastí záporného diferenciálního odporu

Poznámka: Připomeňme si, že u polovodičových (i vakuových) součástek se vyskytují dva typy charakteristik s oblastí záporného diferenciálního odporu dle obr. 3.26. Záporný diferenciální odpor $R_D^{(ef)}$ v určité oblasti AV charakteristiky znamená, že přírůstek napětí $+\Delta u$ vyvolá buď pokles proudu $-\Delta i$ (např. u tunelových diod); takovou charakteristiku nazýváme charakteristikou typu N podle jejího tvaru, nebo přírůstek proudu $+\Delta i$ vyvolá pokles napětí $-\Delta u$, takovou charakteristiku nazýváme charakteristikou typu S.

3.2.5 Usměrňovací diody

Usměrňovací diody jsou konstruovány pro usměrňování střídavého proudu, dříve pouze průmyslových kmitočtů, nyní i kmitočtů značně vyšších (např. u měničů a střídačů).

U usměrňovacích diod se vyžaduje malý úbytek napětí v propustném směru, velký propustný proud a velké závěrné napětí (u detekčních diod požadujeme také malý úbytek napětí v propustném směru, ale především malé prahové napětí a malou parazitní kapacitu).

Usměrňovací diody se obvykle dělí na **nízkovýkonové** (do proudu 20 A) a na **výkonové** (silové, silnoproudé) usměrňovací diody. V současnosti se používají diody křemíkové, výjimečně germániové (v minulosti se používaly tzv. kovové usměrňovače - kuproxové a selenové; neužívalo se termínu dioda, ale usměrňovač nebo ventil).

Podívejme se, jaké nároky musí splňovat dioda v obvodu střídavého proudu. V jedné půlperiodě je dioda pólována v propustném směru a vede proud. Následuje druhá půlperioda, kdy se směr střídavého proudu obrací a dioda se tak dostane do závěrného směru. Polovodičový přechod se uzavírá a roste intenzita elektrického pole, zatěžující polovodičový přechod. Při určité intenzitě pole (tj. určitém závěrném napětí) dojde k průrazu, který může zničit diodu (u usměrňovacích diod ji zničí jistě, protože v obvodu není odpor, který by růst proudu omezil). Proto se při volbě diody musí respektovat průrazné napětí $U_{(BR)}$, resp. max. závěrné napětí U_{Rmax} , protože musíme počítat s nahořilým přepětím (špičkami), viz obr. 3.28. Bude-li následovat filtr se vstupním kondenzátorem, je třeba zabezpečit

$U_{Rmax} \geq 2\sqrt{2} U_{ef}$, kde U_{ef} je

efektivní hodnota usměrněného sinusového napětí. Dále diodu vybíráme podle maximální velikosti proudu I_F .

Výkonové diody se zpravidla dělí do čtyř skupin: *obyčejné* (vysokonapětové), *rychlé*, *lavinové* a *Schottkyho*. Od nízkovýkonových (a dalších plošných) se liší především průměrem přechodů (až 20 mm) a požadovanou čistotou monokrystalu.

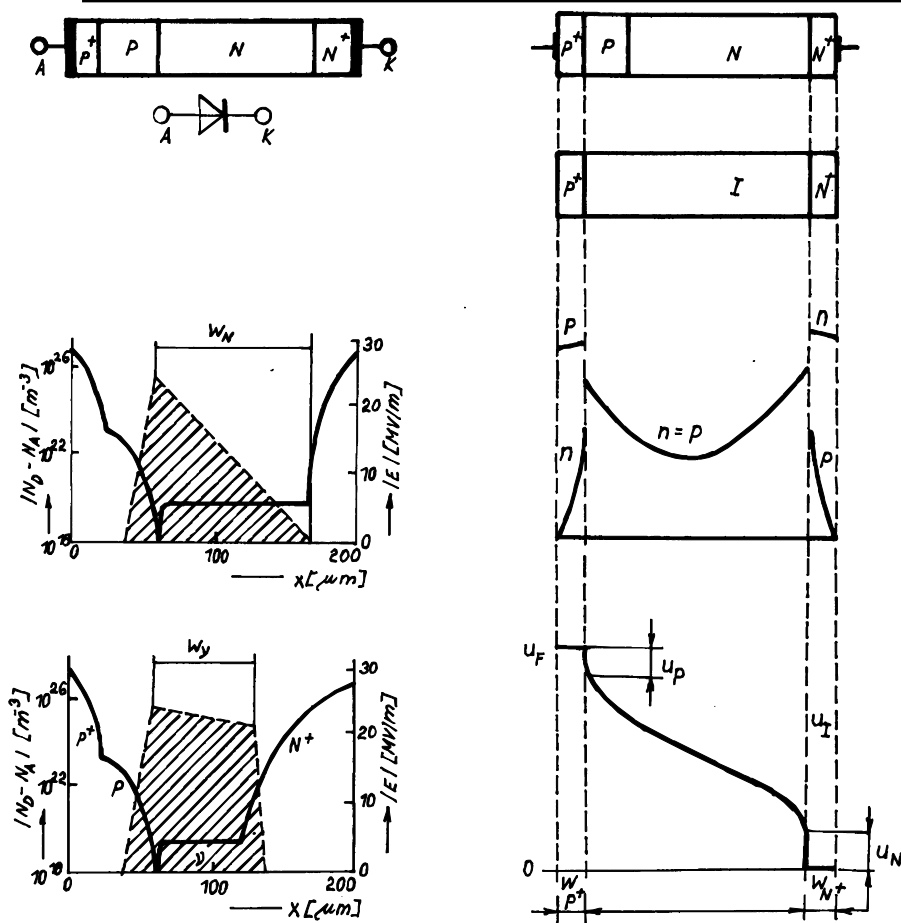
Výkonové diody se vyrábí difúzní technologií (někdy v kombinaci se sléváním, nebo iontovou implantací či epitaxním růstem) obvykle ve tvaru komolého kužele získaného broušením nebo leptáním (s ohledem na povrchový průraz - nejlepší rozložení pole). Pro dosažení vysokých napětových i proudových hodnot se používá uspořádání P^+PNN^+ . Znamená to, že výchozím materiálem je slabě dotovaný (vysokoohmový) křemík např. typu N. Z jedné strany se vytvoří vrstva P a z druhé strany vrstva N^+ . V nově vytvořené vrstvě P se ještě zesílí dotace příměsí, takže vzniká další tenká vrstva P^+ (obr. 3.24).

Přechod PN na vysokoohmovém výchozím křemíku zajišťuje pak požadované (dobré) závěrné vlastnosti diody, kdežto přídavné vrstvy N^+ a P^+ zabezpečují při průchodu proudu bohatý vstřík nosičů do vysokoohmové vrstvy, a te-dy dobré propustné vlastnosti. Tak vzniká dioda s vysokým závěrným napětím (až 5000 V) a současně s velkou proudovou zatížitelností 10^6 A/m^2 .

Propustný stav výkonové diody. Při polarizaci přechodu PN v přímém směru (+ na oblasti typu P) jsou injektovány díry do polovodiče typu N a elektrony do polovodiče typu P. Nerovnovážné nosiče přitom zvyšují celkovou koncentraci nosičů náboje a modifikují vodivost vrstev vně vlastního přechodu PN. V případě struktury P^+PNN^+ výkonové diody dochází ještě k injekci děr z oblasti P^+ do oblasti P a dále do oblasti typu N, stejně pak k injekci elektronů z oblasti N^+ do oblasti N a dále pak do oblasti typu P. Vzhledem k tomu, že koncentrace nosičů

SOUČÁSTKA	USPOŘÁDÁNÍ DESTIČKY Si	PRŮBĚH AV CHARAKTERISTIKY
VÝKONOVÁ USMĚRŇOVACÍ DIODA		
RYCHLÁ USMĚRŇOVACÍ DIODA		
LAVINOVÁ USMĚRŇOVACÍ DIODA		
SCHOTTKYHO DIODA		

Tab. 3.2. Přehled usměrňovacích diod (propustné a závěrné charakteristiky jsou kresleny v různých měřítkách)



Obr. 3.27.: Základní struktura výkonové diody (a), rozložení nosičů a průběhu potenciálu na výkonové diodě v propustném stavu (b) a průběh elektrického pole při závěrném stavu na základní struktuře (c) a struktuře se stlačeným polem (d) výchozího materiálu a provedení technologického procesu.

v oblastech P⁺ a N⁺ o několik řádů převyšuje koncentraci nosičů v oblastech P a N, dochází při vyšších proudových hustotách (řádově jednotky A/mm²) k tomu, že koncentrace nosičů nainjektovaných z krajních vysokodotovaných oblastí mnohonásobně pře-vyšuje koncentraci majo-ritních nosičů ve vysoko-odporové oblasti. Výkono-vou diodu lze proto modelovat jedním přecho-dem PN pouze při malých proudových hustotách; při vysokých proudových hustotách je nejvýhodnějším modelem model PIN diody. Struktura PIN diody je znázorněna na obr. 3.27.

Z rozboru vyplývá, že úbytek napětí na diodě silně závisí na době života nerovnovážných nosičů a tloušťce střední vysoko-odporové oblasti. Pro dosažení nízkých úbytků napětí (malých propustných ztrát) je nutné buď minimalizovat vzdálenost krajních vrstev N⁺ a P⁺ nebo zajistit dlouhou dobu života nosičů, což je velice náročné na kvalitu

Propustná AV charakteristika výkonových diod je teplotně závislá. Na jedné straně s rostoucí teplotou klesá napětí na krajních injektujících přechodech ($u_P + u_N$), na druhé straně s rostoucí teplotou roste (vzhledem k poklesu difúzního koeficientu) úbytek napětí na střední oblasti (u_I). Při malých přímých proudech je tedy u_F klesající funkcí teploty, při velkých proudech u_F s teplotou roste. Porovnání voltampérových charakteristik výkonové diody při různých teplotách je znázorněno na obr. 3.29. Proudová zatížitelnost výkonových diod závisí na schopnosti odvádět vznikající ztrátový výkon, protože přípustná teplota křemíku s ohledem na závěrné vlastnosti a možnost tepelného půrazu je omezena shora nejvyšší provozní teplotou ϑ_{jmax} . Zvětšení proudové zatížitelnosti diod spočívá jednak ve zvětšování průměru destičky výchozího Si materiálu, dále pak v optimalizaci vnitřní konstrukce a technologie.

Závěrný stav výkonové diody. Nejdůležitější funkcí z hlediska dosažení vysokého závěrného napětí U_{RRM} má přechod PN, který je v propustném stavu obvykle zaplaven nerovnovážnými nosiči náboje, a proto nebyl v předcházejícím bodu uvažován. Přechod PN (tzv. „napěťový“) bývá obvykle realizován difúzní technologií ve vzdálenosti x_j od povrchu obvykle větší než 50 μm (do této hloubky již nedosahují poruchy krystalu, způsobené mechanickým opracováním výchozího monokrystalického materiálu).

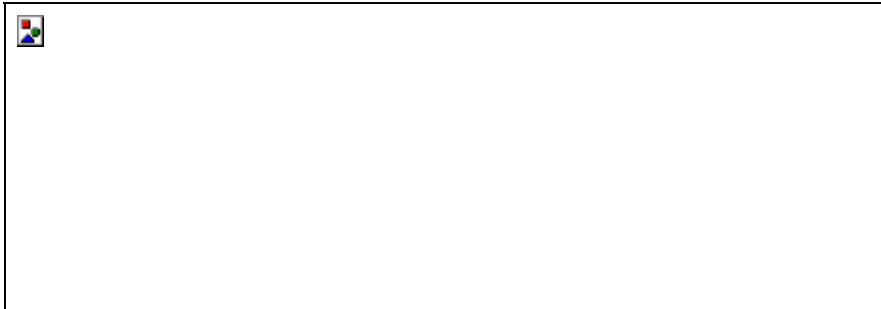
Přesto, že u difúzních přechodů roste koncentrace aktivních příměsí směrem od přechodu PN, je možno pro napětí větší než několik desítek voltů považovat přechod za strmý a použít pro výpočet tloušťky depletiční vrstvy, rozložení elektrického pole a průrazného napětí vzorců platných pro strmý přechod PN.

Velikost průrazného napětí $U_{(BR)}$ je v podstatě dána koncentrací donorů v oblasti typu N. Pro vysoké průrazné napětí je zapotřebí jednak materiálu o vysokém měrném odporu (nízké koncentraci donorů), dále pak velká šířka oblasti prostorového náboje a tedy i velká x_n . Je zřejmé, že s velkou šířkou x_n je spojena velká vzdálenost mezi emitery N⁺ a P⁺ a dosažení nízkých úbytků napětí u_F je spojeno s požadavkem dlouhé doby

života nerovnovážných nosičů. Dlouhá doba života τ_0 vede ke zmenšení generačně-rekombinační složky závěrného proudu.

Dosažení dlouhé doby života v objemu výkonových součástek klade značné nároky na technologii, zejména na čistotu používaných chemi-kálií a pomalé chlazení po vysokoteplotních operacích, atd.

Jak již bylo uvedeno, dosažitelné průrazné napětí $U_{R(BR)}$ je limitováno rozložením elektrického pole na povrchu součástky. U diod je obvykle používáno buď pozitivní zkosení vytvořené sbrou-šením povrchu, nebo struktura MESA vytvořená chemickým leptáním.



Obr. 3.28. Závěrná část AV charakteristiky s vyznačenými mezními napětími

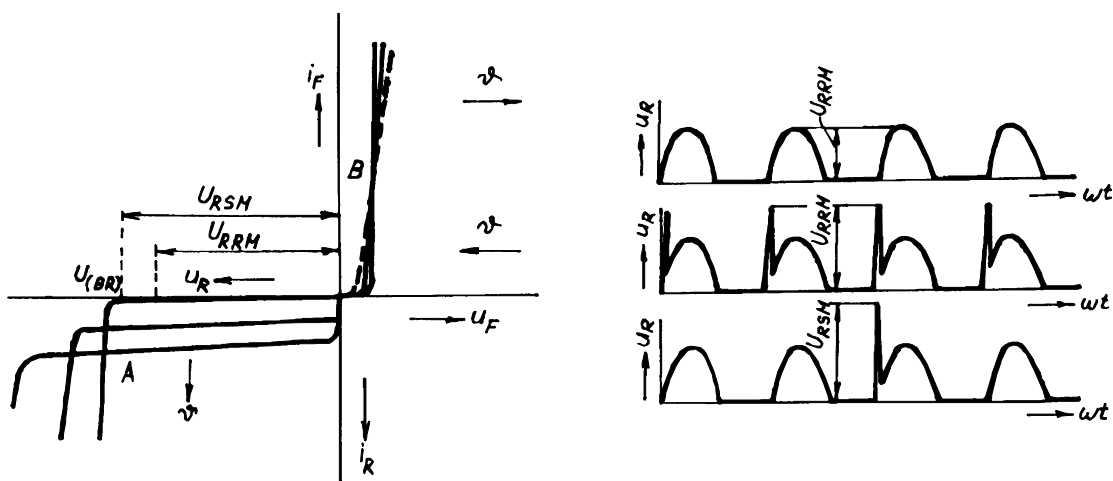
Vážným technologickým problémem je homogenita výchozího monokrystalického materiálu. Pokud je v objemu polovodiče nerovnoměrně rozložena koncentrace aktivních příměsí, můžeme si výkonovou diodu představit jako paralelně zapojené dílčí diodové struktury s různou koncentrací příměsí $|N_D - N_A|$. Může proto docházet k lokálnímu průrazu v malých oblastech, kde pak vlivem vysoké proudové hustoty může dojít k tepelnému průrazu. Proto se pro výrobu součástek s průrazným napětím přesahujícím 3 kV velmi často používá křemík, dotovaný ozářením tepelnými neutrony (tzv. jadernou transmutací), který je vysoce homogenní.

Z provozního hlediska je nejdůležitější část závěrné charakteristiky do napětí u kterého se ještě neprojevuje výrazněji multiplikace nosičů nárazovou ionizací. Pro potřeby aplikací jsou udávány následující parametry, charakterizující provoz diody v závěrném stavu.

Opakovatelné špičkové napětí U_{RRM} , tj. nejvyšší přípustná hodnota závěrného napětí, kterou lze periodicky zatížit diodu v celém rozsahu pracovních teplot.

Mezní závěrný opakovatelný proud I_{RRM} , tj. nejvyšší přípustná hodnota závěrného proudu diody při nejvyšší přípustné provozní teplotě ϑ_{jmax} a napětí U_{RRM} .

Neopakovatelné špičkové závěrné napětí U_{RSM} , tj. nejvyšší přípustná (okamžitá) hodnota závěrného napětí, která náhodně vznikne při provozu. Z definice je zřejmé, že pro tato napětí platí vztah $U_{RRM} \leq U_{RSM} < U_{R(BR)}$.



Obr. 3.29. Průběh AV charakteristiky silové diody při různých teplotách přechodu (a) a možné průběhy přiloženého (závěrného) napětí s vyznačenými parametry závěrné charakteristiky (b)

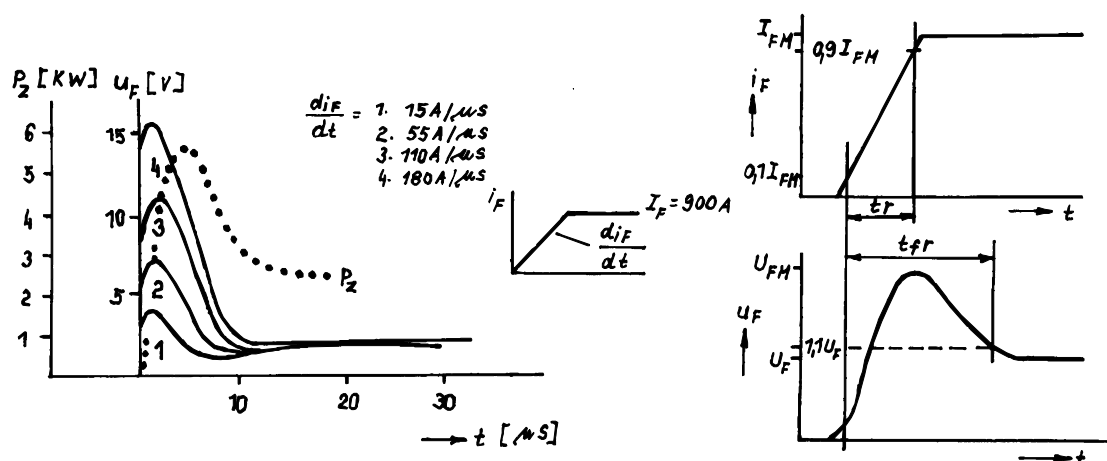
Pro praktické potřeby se diody podle naměřených hodnot U_{RRM} rozdělují do *napětíových tříd*.

Dynamické procesy ve výkonových diodách. Pokud jsou časové změny proudu a napětí v obvodech s výkonovými diodami pomalé, okamžité hodnoty napětí na diodě a proudu diodou odpovídají statickým AV charakteristikám. Při prudkých změnách již není možné nahradit okamžité rozdělení nosičů náboje rozdělením odpovídajícím ustálenému stavu a dochází k výrazným změnám. Teoretické řešení přechodových procesů předpokládá řešení časového průběhu rozložení koncentrace nerovnovážných nosičů v jednotlivých vrstvách polovodičové struktury, tj. řešení rovnic kontinuity s odpovídajícími počátečními a okrajovými podmínkami.

Přechod ze závěrného do propustného stavu. Teoreticky jsou přechodové jevy řešeny obvykle buď za podmínek skokové změny napětí na diodě nebo skokové změny proudu diodou. Z hlediska výkonové polovodičové techniky je podstatně důležitější případ, kdy vnějším zdrojem je v obvodu vynucen nárůst proudu v
přímém
směru
s vysokým di_F/dt .

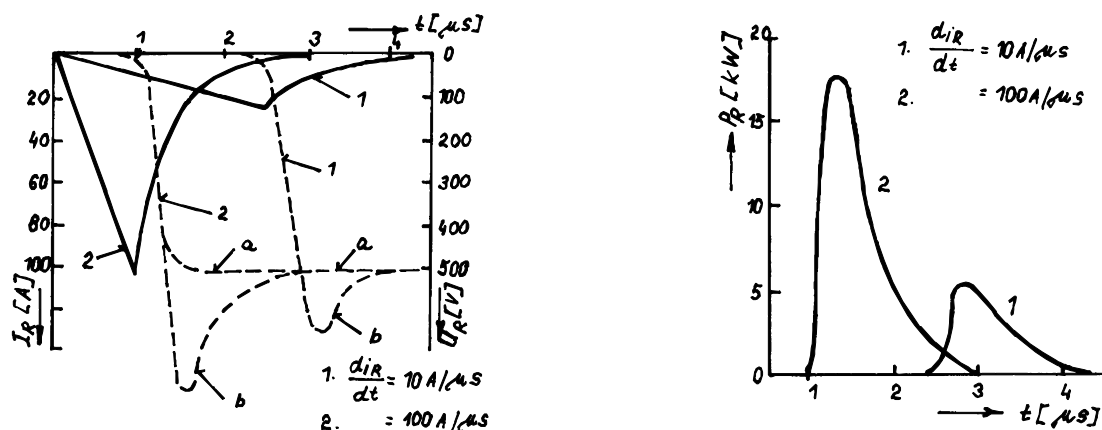
Na počátku přechodového procesu je koncentrace nerovnovážných nosičů různá od nuly pouze v bezprostřední blízkosti krajních přechodů a prakticky celý úbytek napětí na diodě je dán odporem základního materiálu na střední oblasti. S postupným zapláváním střední oblasti nerovnovážnými nosiči napětí na diodě klesá a blíží se po průchodu nevýrazným minimem hodnot odpovídající AV charakteristice. Přechodový proces závisí na proudové hustotě, injekční účinnosti přechodů, tloušťce vrstev a difúzní délce a také na tvaru impulsu proudu. Příklad průběhu napětí na diodě pro různé strmosti di_F/dt je znázorněn na obr.3.30.

K porovnání rychlosti přechodu ze závěrného do propustného směru je používán parametr t_{fr} , což je doba od přiložení impulsu propustného proudu (se strmostí nárůstu di_F/dt) do okamžiku, kdy poklesne napětí na diodě na 1,1 hodnoty stejnosměrného úbytku napětí U_F . průběh napětí a proudu je znázorněn na obr. 3.30.



Obr 3.30. Rychlý přechod diody do propustného stavu: a) průběh napětí na diodě (čárkovaně je znázorněn možný průběh ztrátového výkonu P_Z), b) způsob stanovení doby propustného zotavení t_{fr}

Zvýšení úbytku napětí na diodě v počátečních fázích přechodového procesu má za následek prudké zvýšení ztrátového výkonu, který roste rychle s di_F/dt a nelze jej při vysokých di_F/dt a vyšších frekvencích neuvažovat. Příklad experimentálně zjištěného ztrátového výkonu při přechodu diody do propustného stavu je znázorněn na obr. 3.30.



Obr. 3.31. Průběh proudu, napětí a ztrátového výkonu při zpětném zotavení výkonové diody

Přechod z propustného do závěrného stavu. Na počátku přechodového procesu je oblast mezi krajními vrstvami N^+ a P^+ zaplavena nerovnovážnými nosiči náboje, jejichž rozložení odpovídá hustotě propustného proudu. Po přepnutí směru proudu (změně polarity napětí ve vnějším obvodu) zabraňuje vysoká koncentrace nerovnovážných nosičů náboje v místě přechodu vytvoření oblasti prostorového náboje.

V první fázi procesu je proud ve zpětném směru omezen pouze parametry obvodu. Teprve až klesne vlivem odvedení a rekombinace koncentrace nosičů na rovnovážnou hodnotu, dochází k zotavení závěrných vlastností diody. Situace byla již rozebírána ve 2. kapitole. Zde bylo také uvedeno, že při zpětném zotavení vzniká na přechodu PN přepětí, jak je znázorněno na obr. 3.31 křivkou b.

Velikost přepětí roste se strmostí poklesu zpětného proudu. Z těchto důvodů se výrobci snaží o relativně pomalý pokles zpětného proudu, tzv. „měkké zotavení“ (soft recovery). K porovnání se používá parametr $s = t_f/t_s$; snahou je dosáhnout co největších hodnot tohoto parametru.

Jelikož závěrné napětí roste rychle v okamžiku (intervalu), kdy diodou protéká velký zpětný proud, je se zpětným zotavením spojen velký ztrátový výkon, který roste s dI/dt v obvodu.

3.2.6 Rychlé usměrňovací diody

Rychlá usměrňovací dioda, často nazývaná dioda s malým komutačním nábojem, frekvenční dioda nebo jen rychlá dioda, má strukturu P^+PNN^+ konstruovanou tak, aby umožňovala velmi rychlý přechod z propustného do závěrného směru nebo jinak řečeno, aby měla co nejmenší komutační náboj. Toho se zpravidla dosahuje zavedením rekombinačních center v podobě atomů Au nebo Pt (zkrácení doby života minoritních nosičů u výše uvedené struktury ve vrstvě N).

Tvar AV charakteristiky u rychlých diod je stejný jako u obyčejných usměrňovacích diod. Ovšem při daném závěrném napětí mají rychlé diody větší závěrné proudy. Zvětšení závěrného proudu je tím větší, čím kratší doby života (menšího komutačního náboje) chceme dosáhnout. Se zkrácením doby života se zvětšuje i úbytek napětí na diodě. Tedy proti obyčejným diodám mají rychlé diody menší napěťovou zatížitelnost a větší úbytek napětí v propustném směru. V provozu je nelze zatěžovat napětím $U_R \geq U_{(BR)}$.

Jak vyplývá z předchozího výkladu, požadavek současného dosažení vysokého závěrného napětí, nízkých propustných ztrát a malého komutačního náboje (krátkých zotavovacích časů) je konfliktní, protože jednotlivé konstrukční faktory působí proti sobě.

Protichůdně působí zejména požadavek současného dosažení vysokého závěrného napětí a vysoké proudové zatížitelnosti. Vysoké závěrné napětí vyžaduje velkou tloušťku vrstvy vysokoodporového materiálu pro rozšiřování depletiční vrstvy přechodu PN, vysoká proudová zatížitelnost je spojena s malým úbytkem napětí v propustném stavu, a vyžaduje tedy minimální vzdálenost krajních vysokodotovaných vrstev.

Určitý kompromis je možno dosáhnout konstrukcí s tzv. „stlačeným polem“, znázorněné na obr. 3.27d. V tomto případě je dioda vytvořena strukturou P^+PvN^+ , u které dochází k rozšíření oblasti prostorového náboje až do oblasti N^+ . Intenzita elektrického pole je ve vrstvě v téměř konstantní.

Stejně propustné charakteristiky obyčejné diody a diody se stlačeným polem odpovídá zhruba stejná tloušťka nízkodotované oblasti, tj. $w_N \approx w_V$. Pak z výrazů



$$a \quad U_{(BR)v} = E_{BR} w_v - \frac{e N_{Dv}}{2\epsilon} w_v^2$$

je zřejmé, že vzhledem k $N_{Dv} < N_D$ je vyšší závěrné napětí dosažitelné na struktuře s tzv. „stlačeným polem“, nebo je možno stejného závěrného napětí dosáhnout při menší vzdálenosti w krajních N^+ a P^+ vrstev, čemuž odpovídá nižší propustné napětí při stejné proudové hustotě a stejné době života nosičů, nebo je možné dosáhnout stejného závěrného napětí a stejné propustné charakteristiky při kratší době života nerovnovážných nosičů. Je tedy možno pomocí této konstrukce realizovat diody s vysokým závěrným napětím, vysokou proudovou zatížitelností a malým komutačním nábojem pro použití ve středofrekvenčních aplikacích. Tento typ diody se obvykle nazývá velmi rychlá usměrňovací dioda nebo *výkonová PIN dioda*.

V současné době jsou pro středofrekvenční aplikace vyvíjeny diody s parametry v úrovni $U_{R(BR)} = 1000$ V a $t_{rr} = 0,5$ μ s, pro nižší napětí je možno dosáhnout kratší doby zpětného zotavení.

V literatuře jsou popsány diody s napětím $U_{RRM} = 200$ V, $I_{FAV} = 70$ A, u kterých vzhledem k tloušťce vysokoodporové oblasti okolo 25 μ m je možno zkrácením doby života na úroveň $\tau \approx 50$ ns zkrátit dobu zpětného zotavení $t_{rr} \leq 60$ ns. Tyto typy diod se svými komutačními vlastnostmi blíží Schottkyho diodám, mají však vyšší napětí $U_{R(BR)}$.

Určitou nevýhodou diody se stlačeným polem je strmý pokles zpětného proudu v okamžiku zotavení závěrného napětí, což má za následek vznik velkých přepětí v obvodech s indukčností.

3.2.7 Lavinové usměrňovací diody

Lavinová dioda je zvláštním druhem usměrňovací diody, která má přetížitelnost v závěrném směru přibližně stejně velkou jako v propustném směru. U obyčejných výkonových diod s přechodem PN dosahuje přetížitelnost v propustném směru relativně vysokých hodnot (např. u diody na 200 A je ztrátový výkon trvale asi 200 W, avšak po dobu 10 ms je 14 kW), naproti tomu v závěrném směru při přepětíové špičce stačí již ztrátový výkon jednotek wattů ke zničení diody.

Zcela jinak je tomu u lavinové usměrňovací diody. Pro přetížitelnost lavinových diod v závěrném směru platí obdobné křivky přetížitelnosti jako pro výkonové diody v propustném směru (křivky proudu či výkonu v závislosti na době přetížení). Lavinová dioda odpovídající výše uvedené obyčejné diodě snese po 10 ms přetížení 4 kW a po 10 μ s již 90 kW. Pracovní bod se tedy může na závěrné charakteristice pohybovat i v oblasti za kolenem (napětím $U_{(BR)}$), aniž by došlo ke zničení součástky. Vlivem speciální technologie výroby nastává průraz rovnoměrně po celé ploše přechodu PN uvnitř krystalu a nikoli jen místně (zpravidla na povrchu krystalu, jako je tomu u obyčejné výkonové diody). K výrobě se užívají strukturálně velmi dokonalé krystaly s velkou homogeností měrného odporu (Si dotovaný neutronovou přeměnou). Dvojí difúzí se vytvoří obvodový ochranný prstenec, aby průrazné napětí při povrchu krystalu bylo větší než uvnitř krystalu. Pro průraz nejsou pak rozhodující jevy na povrchu (ten je samozřejmě v místech vyústění přechodu vhodně pasivován), ale celá plocha přechodu PN uvnitř krystalu.

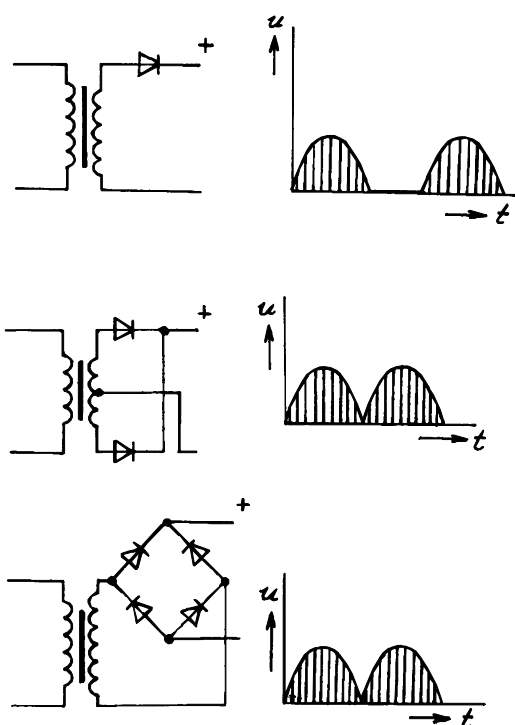
Výhody lavinových usměrňovacích diod jsou zřejmé:

- a) Jsou odolné proti přepětím, protože převedou bez poruch velký špičkový závěrný proud procházející při přepětí. Lze je tedy zapojovat do obvodů bez přepětíových ochran.
- b) Mohou být více napětíově využívány než obyčejné napětíové diody.
- c) Jsou zvlášť vhodné pro sériové řazení bez přidavných děličů napětí.

Nevýhodou lavinových diod je jejich obtížnější výroba (a vyšší cena), zejména pro závěrná napětí větší než 1500 až 2000 V. Praktické uplatnění našly pro napětí do 1000 až 1200 V. Používají se buď jako usměrňovací diody, nebo velmi často jako ochranné prvky v obvodech s obyčejnými výkonovými diodami.

3.2.8 Schottkyho usměrňovací diody

Schottkyho usměrňovací dioda je založena na přechodu kov - polovodič. Vyrábějí se většinou epitaxní technologií (obr. 3.8). Význačnou vlastností oproti jiným usměrňovacím diodám je



Obr. 3.32. Zapojení usměrňovačů: a) jednocestné, b) dvojcestné, c) dvojcestné můstkové

přesahuje rámec těchto skript.

Třívrstvá struktura P^+NP^+ se nazývá *diak* a má v 1. a 3. kvadrantu souměrnou AV charakteristiku s oblastí záporného diferenciálního odporu (obloukového typu).

Čtyřvrstvá dioda, označovaná někdy po svém původu jako *Shockleyova*, má strukturu PNPN. AV charakteristika vykazuje v 1. kvadrantu oblast záporného diferenciálního odporu. Jde vlastně o tyristor bez řídicí elektrody, proto se také někdy nazývá tyristorová dioda. Podobnou charakteristiku má dioda s dvojíou bází (nazývaná také jednopřechodový tranzistor, zkratkou UJT, je to trojpól). Tento prvek existoval také v řadě modifikací.

Pětivrstvá dioda, používaná v počátcích tyristorové techniky má charakteristiku jako diak, ale s menším úbytkem napětí v propustném stavu. O těchto spínacích diodách se ještě zmíníme.

Velmi významné jsou diody používané jako zdroje záření (světelného a IR), tzv. luminiscenční (svítivé, LED) diody a laserové diody (polovodičové lasery); nebo přijímače záření (nejrůznější fotodiody). Magnetodiody jsou prvky citlivé na magnetické pole.

3.4 Některé obvodové aplikace diod

Nejčastější použití diod je v usměrňovačích střídavého napětí

(v napájecích zdrojích). Využívá-li se k usměrnění jedna nebo obě půlperiody střídavého proudu, hovoříme o jednocestném nebo dvoucestném usměrnění.

Na obr. 3.32a je zapojení jednocestného usměrňovače. Na obr. 3.32b je zapojení dvoucestného usměrňovače, které používá dvě usměrňovací diody a transformátor se zdvojeným vinutím. Toto zapojení je obdobou zapojení usměrňovače s dvojíou vakuovou diodou, což bylo nejobvyklejší zapojení usměrňovače v éře elektronek. Na obr. 3.32c je zapojení

- malý úbytek napětí v propustném směru (zhruba poloviční)
- velký závěrný proud a jeho značná teplotní závislost,
- malý komutační náboj; protože propustný proud je přenášen pouze většinovými nosiči, dynamické vlastnosti nejsou ovlivňovány chováním prostorových nábojů menšinových nosičů.

Schottkyho diody mají proto vysoké rychlosti zapínání a vypínání. Mohou usměrňovat proudy s kmitočtem až 1 MHz. V provozu lze Schottkyho usměrňovací diody zatěžovat napětím $u_R < U_{(BR)}$. Vyrábí se pro proudy do 100 A a úbytek napětí menší než 0,5 V.

3.3 Ostatní prvky diodového charakteru

Diodou rozumíme prvek s jedním přechodem PN nebo MS. Dodejme, že bývá dvou až čtyřvrstvá (P^+PNN^+). V širším smyslu používáme název dioda pro polovodičové prvky se dvěma vývody, i když ve struktuře mají libovolný počet přechodů PN (0 až 4). Postupně uvedeme nejdůležitější z nich.

Gunnova dioda neobsahuje žádný přechod, je to homogenní polovodič se dvěma kontakty. Princip činnosti



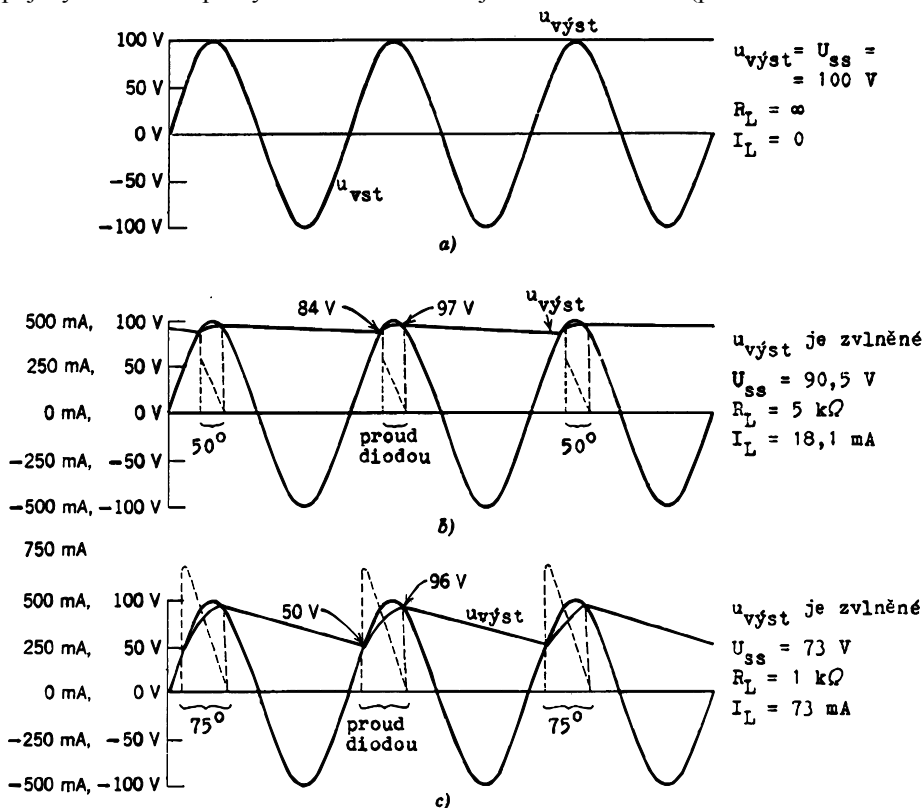
Obr. 3.33. Graetzův můstkový usměrňovač (full-wave bridge) jako hybridní integrovaný obvod (dole) a jeho historický předchůdce - selenový můstek (nahore)

dvojcestného můstkového usměrňovače, které obvykle označujeme jako Graetzovo.

Na obr. 3.32 je také zakreslen tvar výstupního napětí, vidíme, že je sice stejnosměrné, ale zvlněné (pulzující). Pulzující stejnosměrné napětí se dá vyhladit kondenzátorem nebo (složitějším) filtrem. Je tedy třeba zkoumat vlastnosti usměrňovače spolu s připojenými dalšími prvky - konkrétně může jít o kondenzátor (paralelně k němu připojenou zátěž) na vstupu filtru nebo tlumivku na výstupu filtru.

Nejjednodušší a nej názornější na pochopení je první případ, tedy **usměrňovač s RC zátěží** (v literatuře se označuje také jako usměrňovač se sběrným kondenzátorem). Situaci ilustruje obr. 3.34 (jde o ustálený stav). Během části půlvlny pulzujícího stejno-směrného napětí se kondenzátor nabije a při poklesu napětí předává svůj náboj do zátěže. V tomto případě musí být závěrné napětí použitých diod větší než dvojnásobek amplitudy (špičkové hodnoty) usměrňovaného napětí.

3.5 Vlastnosti a modely polovodičových diod



Obr. 3.34. Průběhy napětí a proudu v jednocestném usměrňovači s RC zátěží ($C = 20 \mu\text{F}$, $R_L =$ a) nekonečno b) $5 \text{ k}\Omega$ a c) $1 \text{ k}\Omega$), vstupní střídavé napětí u_{vst} má efektivní hodnotu $70,7 \text{ V}$

Vlastnosti polovodičových diod plynou z již uvedených vlastností přechodu PN. Zde uvedeme několik poznámek, které vycházejí z potřeb inženýrské praxe. Vžitý způsob zakreslování voltampérových charakteristik diod je naznačen na obr. 3.35a), odpovídající popis modelu diody na obr. 3.35b). Obojí se vztahuje nejen k diodám, ale k jakémukoliv přístupnému přechodu PN, tedy k emitorovému či kolektorovému přechodu



bipolárního tranzistoru nebo k „utajeným“ přechodům v MOSFETech nebo v monolitických integrovaných obvodech.

Často je užitečné modelovat diodu pomocí spí-nače (ideální diody) a příp. zdroje prahového napětí a dynamického odporu propustné větve nebo závěrné větve při Zenerově nebo lavinovém průrazu, viz obr. 3.36. Dynamický odpor (strmost) propustné větve silně závisí na poloze pracovního bodu, tj. na procházejícím proudu (obr. 3. 37), příslušné teoretické vztahy byly již odvozeny

Obr. 3.35. Ampérvoltová charakteristika polovodičové diody (a) a model diody pro statický režim (b)

v kap. 2. U reálné diody je dynamický odpor roven teoretickému ($r_d \approx r_j$) je v oblasti kolem prahového napětí, potom převládá odpor polovodičového materiálu (charakteristika diody je více lineární než exponenciální) - obr. 3.38.

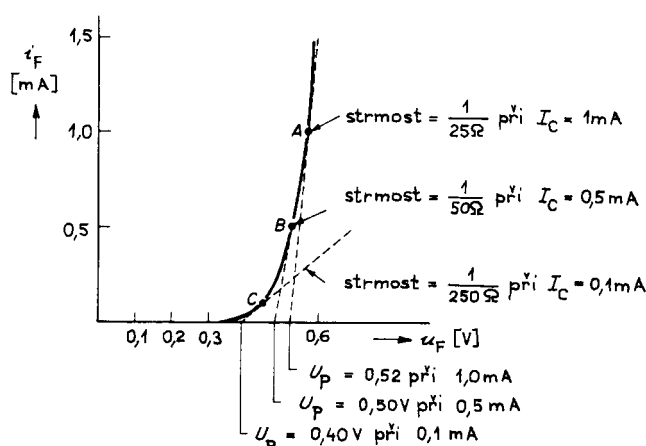


Obr. 3.36. Různé způsoby modelování reálné diody pomocí spínače (ideální diody)

V katalogových údajích diod je zvykem uvádět maximální přípustnou hodnotu U_F pro zvolenou hodnotu I_F (při dané okolní teplotě) a maximální přípustnou hodnotu U_R při zvolené hodnotě I_R . Aby měl výrobce minimální odpad, jsou zvolené hodnoty I_R a U_F značně pesimistické. Skutečný závěrný proud diody (resp. zbytkový nasycený proud I_O) je často ve skutečnosti i o několik řádů menší než I_R , takže o skutečném průběhu voltampérové charakteristiky diod se od výrobce nic bližšího nedovíme.

Model diody podle obr. 3.35b) sestavujeme obvykle ze dvou lineárních rezistorů R_S a R_P a z „ideálního“ přechodu PN, popsaného známým exponenciálním vztahem. Tato náhrada je zcela oprávněná u křemíkových součástí. Jejich skutečné charakteristiky se k popisovanému exponenciálnímu průběhu velice blíží od hodnoty I_O při závěrném napětí cca -0,5 V přes nulu do oblasti vodivé a zde v rozsahu několika řádů hodnot proudu. Pro diody je běžná shoda v rozsahu 4 až 5 řádů, pro PN přechody tranzistorů do 7 řádů. Existují speciální diody a tranzistory (např. UL1111 z Polska) se zbytkovým proudem v řádu desítek pA a s charakteristikou blízkou se k ideální v rozsahu až 9 řádů. Při statickém proměřování průběhu voltampérové charakteristiky (tj. bod po bodu, stejnosměrně) zjistíme průběh odlišný od průběhu podle modelu z obr. 3.35b). Rozdíl spočívá v tom, že u modelu počítáme se stálou vnitřní teplotou přechodu ϑ_j , zatímco při měření můžeme zaručit jen stálou teplotu okolí ϑ_a . Věřitelnou voltampérovou charakteristiku bychom tedy měli proměřovat dynamicky při známé okolní teplotě a vnitřní výkonové ztrátě.

Přirozenou a významnou vlastností všech přechodů PN je jejich teplotní závislost. Ve vztahu popisujícím ideální přechod PN závisí na teplotě přechodu dva členy: člen a (převrácená hodnota teplotního napětí) a zbytkový proud I_O . Dominantním vlivem je změna zbytkového proudu (více než 20 x silnější). Pro odhad teplotních změn průběhu voltampérové charakteristiky vyjdeme z exponenciální závislosti $I_O(\vartheta)$, takže vzniknou charakteristiky podle obr. 3.39.



Obr. 3.37. Strmost (dynamický odpor) propustné větve závisí na poloze pracovního bodu



V praxi lze vystačit se zjednodušeným pohledem. Je-li dioda využívána jen v závěrné oblasti při napětích mezi U_R a 0,5 V, stačí brát v úvahu jen posuv ve směru osy proudu, přičemž je výhodné počítat se vzrůstem I_0 na dvojnásobek při změně teploty o 10 °C, jak je naznačeno na obr. 3.40 vlevo. Je-li využívána jen ve vodivé oblasti, je možné příslušný úsek v prvním přiblížení nahradit přímkou nebo přesněji exponenciálou a ta bude lineárně posouvána s rostoucí teplotou vlevo. Velikost tohoto posuvu ve skutečnosti závisí na mnoha činitelích a mění se v rozmezí od 2 do 4 mV/°C. Pro odhady však stačí uvažovat nejčastěji se vyskytující posuv - 2,5 mV/°C (v integrovaných obvodech se často uvažuje jen posun - 2 mV/°C).

Obr. 3.40. Zjednodušený pohled na teplotní závislost charakteristik diody



Obr. 3.38. Dynamický odpor polovodičové diody

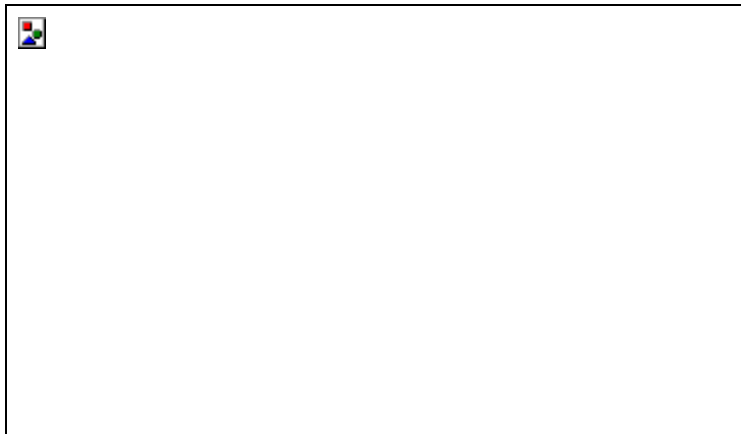
Na obr. 3.41 jsou naznačeny typické průběhy AV charakteristik některých nízkovýkonových polovodičových diod ve vodivé oblasti. Jsou zde vyznačena prahová napětí po úsecích linearizovaného průběhu, tj. 0,3 V pro germaniové přechody, 0,6 V pro křemíkové přechody, 1,5 V pro červené galium-arsenidové LED a 1,8 V pro zelené galium-fosfidové LED. Prahové napětí Schottkyho diod KAS se pohybuje kolem 0,25 V. Výsledky měření (kromě KAS34) ukazují na poměrně malý rozptyl teplotních napětí U_T asi 40 až

Obr. 3.39. Změna ampérvoltových charakteristik diody s teplotou

50 mV (tedy konstanta a je cca 20 až 25 V⁻¹) u všech kontrolovaných přechodů. U diod KAS21 stojí za povšimnutí, že Schottkyho přechod se uplatní při pracovních proudech pod 5 mA, při větších proudech přebírá větší část celkového proudu přechod PN, který působí paralelně k Schottkyho přechodu.

Obr. 3.42 znázorňuje průběhy proudu běžných diod a přechodů PN v závěrné oblasti. Lavinový průraz při napětí větším než 20 V se uplatní nejen u všech usměrňovacích i detekčních diod, ale i u kolektorových přechodů bipolárních tranzistorů, všech přechodů PN vznikajících v monolitických integrovaných obvodech vůči substrátu. S menším průrazným napětím (při napětí nižším než 20 V je lavinový průraz kombinován se Zenerovým průrazem, od cca 6 V převládá lavinový průraz) se setkáváme u pravých Zenerových diod, diod LED a u emitorových přechodů planárních a planárně epitaxních bipolárních tranzistorů, tedy i u všech tranzistorů v integrovaných obvodech. Průrazná napětí emitorových přechodů se pohybují mezi 7 a 12 V, a to podle typu a výrobce součásti. Rozptyl hodnot tohoto průrazného napětí u určitého typu tranzistoru (a u určitého výrobce) zřejmě závisí na dodržování technologické kázně a především na přesnosti dodržení koncentrace příměsí v prostoru báze tranzistoru. Zatímco průběhy charakteristik Zenerových diod jsou obvykle značně zaoblené, průraz emitorového přechodu je charakterizován zcela ostrým zlomem a zakřivení je patrné jen v oblasti proudů pod 1 uA. (Typický emitorový přechod u integrovaného obvodu má průrazné napětí 6,5 V a teplotní koeficient + 3 mV/°C).

Charakteristiky na obr. 3.43 zachycují vliv osvětlení přechodu PN. Takto probíhají nejen charakteristiky fotodiod a kolektorových přechodů fototranzistorů, ale též charakteristiky nechtěně osvětlených přechodů PN. Rušivé osvětlení PN se může projevit i u systémů v kovových pouzdrech prosvítáním skleněných závitů, a to především v případech, kdy má součástka pracovat v oblasti velmi malých proudů (např. zesilovač v mikrobežím, spínač nA proudů). Průběhy z obr. 3.43 jsou podobné též výstupním charakteristikám diodových optronů.



Obr. 3.41. Typické průběhy AV charakteristik některých diod

napětíového rozdílu

$$\Delta U_d = U_{d2} - U_{d1} \text{ vypočítat teplotní napětí } U_T, \text{ resp. konstantu } a.$$

Vztah pro proud diodou (2.32) byl na obr. 3.21 upraven zavedením emisního součinitele m , který u křemíkových diod je přibližně roven dvěma. O rozsahu platnosti tohoto vztahu pro určitou diodu se můžeme přesvědčit na AV charakteristice změřené v co nejširším rozsahu proudu a zobrazené v semilogaritmických souřadnicích. Vztah (2.32) lze v semilogaritmických souřadnicích zobrazit přímkou pro rozsah $i_d \gg I_0$. Přímkovou náhradu můžeme použít i k zajištění konstant I_0 a na obr. 3.21. V průsečíku náhradní přímky s osou proudu ($u_d = 0$) zjistíme I_0 . Zvolíme-li na přímce vhodně dva proudy I_{d1}, I_{d2} , můžeme z odečteného



Obr. 3.42. Charakteristiky různých diod v oblasti průrazu (je též naznačen vliv teploty na průrazné napětí)

Základní údaje v přechodu PN v dynamickém režimu jsou v kapitole 2.5. Je tedy známo, že každou polovodičovou diodu (a každý přechod uvnitř libovolného tranzistoru) lze nahradit nelineárním modelem podle obr. 3.44b. Vliv prvků R_S a L_S lze obvykle proti vlivu vnějších sériových prvků zanedbat, takže nám postačí model z obr. 3.44c.

V případě malých změn signálu můžeme model linearizovat, tj. vlastní nelineární nesetřvačný přechod PN („ideální“ diodu) nahradit rezistorem s odporem rovným dynamickému odporu (2.36) v daném pracovním bodě. Kapacita C_p reprezentuje souhrn lineárních konstrukčních kapacit, C_T a C_D jsou dva nelineární, hypoteticky oddělené kapacitory, lišící se značně fyzikálním původem



Obr. 3.43. Vliv osvětlení přechodu PN

a funkcí. Pro jejich diferenciální kapacity byly již odvozeny vztahy (2.42) a (2.45). Bariérovou kapacitu v praxi obvykle označujeme jako C_T , v teorii potom jako C_j (viz odstavec 2.3.3.).

Difúzní kapacitor C_D má původ v difúzi nosičů náboje přes otevřený přechod, takže se projeví při vodivém stavu přechodu. Jeho chování se zcela vymyká běžným představám o chování kondenzátorů. Difúzní kapacita C_D tak vyjadřuje náboje hromaděné v přechodu a jeho blízkém okolí a jejich změny s protékajícím proudem. Pro difúzní diferenciální kapacitu platí (2.45). Vidíme, že je přibližně přímo úměrná protékajícímu proudu.



Obr. 3.44. Dynamický model polovodičové diody

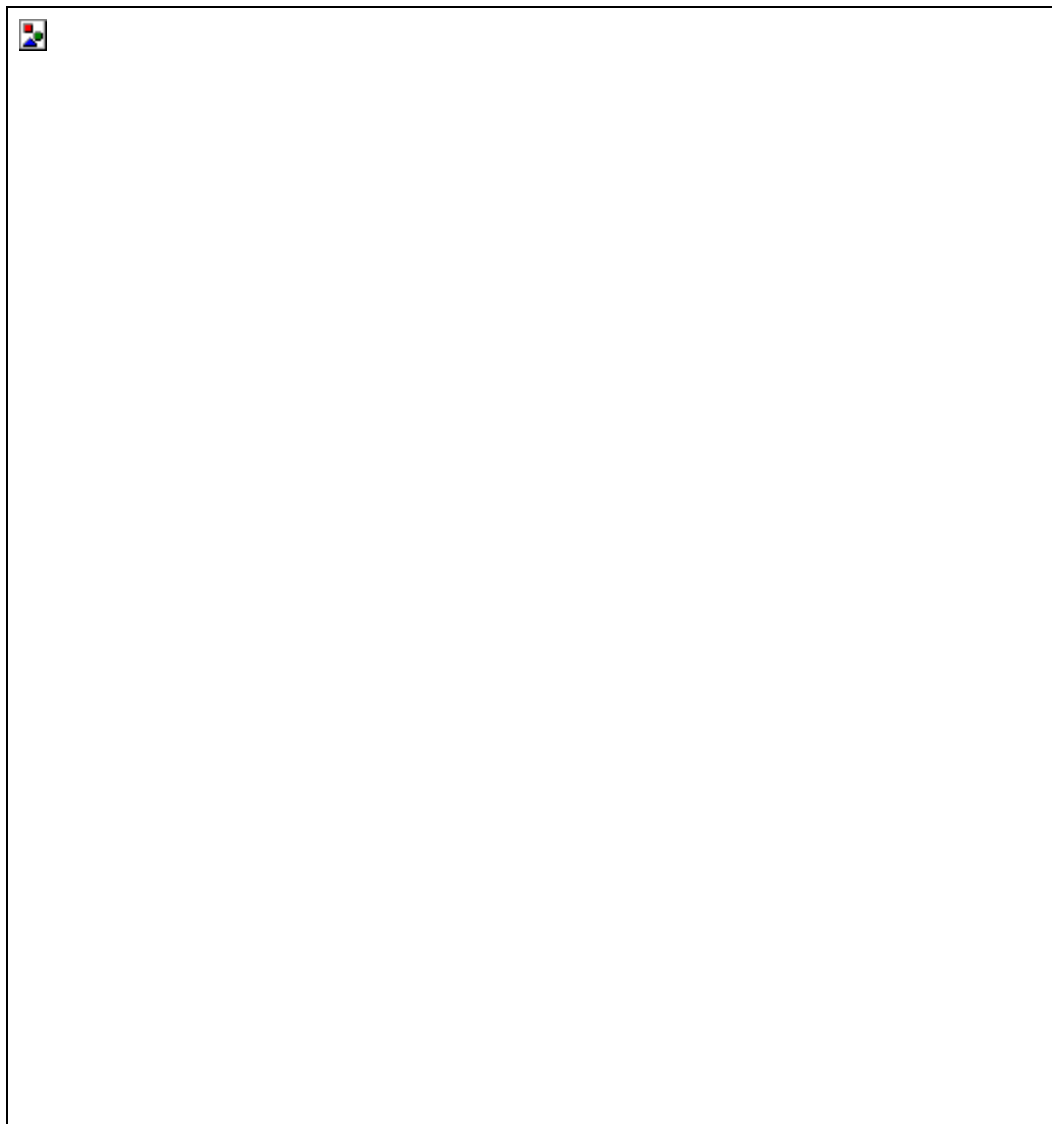
Hypotetické průběhy C_T a C_D umožňují vysvětlit skutečný průběh napěťové závislosti diferenciální kapacity diody C_d (a jakýchkoli dalších přechodů PN) a technologům dávají možnosti do určité míry pro daný účel tento průběh ovlivnit. Příklad napěťové závislosti C_d a jejich složek je uveden na obr. 3.45. Závislosti na obr. 3.45 nám slouží především k posouzení diferenciálních kapacit přechodu při malých signálech a konstantním nebo pomalu se měnícím U_d (bez omezení na relativně malé změny), tedy k zjištění diferenciálních kapacit v linearizovaných přenosových obvodech nebo v obvodech pro fázovou nebo frekvenční modulaci (dioda ve funkci varikapu).



Při velkých změnách u_d , při nichž přechod přechází z vodivého do nevodivého stavu (spínací nebo impulsový režim) však změny C_d neumožní vytvořit jednoznačnou představu výsledného chování. Zde se totiž uplatní doba generace a rekombinace a postupné hromadění a vyprazdňování nosičů náboje z oblasti přechodu, jehož časový průběh nemůže být z průběhu $C_d = f(u_d)$ patrný. V takovém případě dáváme přednost empirickému zjištění časového průběhu odezvy jednoduchého obvodu s příslušným elektronickým prvkem na veliký skok vstupního napětí.

Obr. 3.45. Závislost bariérové a difúzní kapacity na napětí diody

Základní poznatky o chování diody při velkých změnách přiloženého napětí byly již uvedeny v odstavci 2.4.1. Na obr. 3.46 jsou uvedeny průběhy napětí a proudu v impulsním obvodu pro běžnou a rychlou spínací diodu. Na obr. b) je naznačen přechod z nevodivého do vodivého stavu. Časový průběh odezvy záleží na velikosti R . Při relativně velikých hodnotách R se odezva blíží průběhu, který můžeme očekávat při obvyklém spojení prvků R - C . Zřejmě se zde především uplatní kapacitory C_p a C_T . Při malých hodnotách R a relativně velikých proudech i_d se však objeví odezva zcela neočekávaná, jakoby se jednalo o obvod R - L . Projevuje se především kapacitor C_D a to tak, že je zřejmě nutné změnou elektrického pole nejprve pohyblivé nosiče náboje v dostatečně velkém množství do přechodu vtáhnout. Po ustálení přechodného děje se udržuje v rovnovážném stavu generování nových nosičů a jejich postupná rekombinace v polovodiči opačného typu vodivosti. Je-li opačným skokem napětí vtahování nosičů do



Obr. 3.46. Průběhy napětí a proudu v impulsním nebo spínacím obvodu s diodou I_F a na proudu zpětném I_Z , kterým je přebytečný náboj odváděn. Pro její výpočet se uvádí vztah

$$t_s = \tau_s \ln \left(1 + \frac{I_F}{I_Z} \right)$$

kdy τ_s je časová konstanta zotavení, která závisí na konstrukci i technologických parametrech přechodu; její hodnota není výrobcí uváděna, pro určitý rozsah pracovních podmínek její hodnotu zjišťujeme měřením.

Z průběhu napětí u_d na obr. 3.46c) můžeme usoudit, že přibližně v časovém intervalu doby zotavení zůstává přechod vodivě pólován.

U speciálních rychlých spínacích diod se nedá z časového průběhu dost věrohodně zjistit doba zotavení t_s . Pro ně je zotavení charakterizováno časem t_{tr} podle obr. 3.46d). Doba t_{tr} se zde definuje pro pokles zpětného proudu na vypočítanou hodnotu $0,1 \cdot U/R$. Pro rychlé impulsní diody nejmenších výkonů se uvádí t_{tr} v rozmezí 3 až 15 ns při I_F kolem 10 mA.

Vliv zotavení přechodů se při návrzích některých obvodů dá účelně posuzovat, známe-li celkový náboj Q_C , který má být během zotavení odveden. U rychlých diod se hodnota tohoto náboje pohybuje kolem

$$Q_C \approx 10^{-8} I_F \quad [\text{C}, \text{A}].$$

Schottkyho diody mají tento náboj o 2 až 3 řády menší.

přechodu
zastaveno, nastal v
okolí přechodu
okamžitý přebytek
nahromaděných
menšinových
nosičů náboje,
které je zapotřebí
odvést. To se děje
vnějším zpětným
proudem I_Z ,
přičemž vnější
napětí na
přechodu z
původního stavu
 U_{d+} jen nepatrně
pokleslo na
počáteční
zotavovací napětí
 U_Z , jak je patrné
na obr. c) nebo d).
Časové průběhy
napětí u_d
a proudu i_d na
obr. 3.46c)
odpovídají
průběhům u
většiny planárních
přechodů PN. V
průběhu proudu je
poměrně přesně
ohraničena doba
zotavení t_s
přechodu. Doba
zotavení závisí na
předním proudu

Evropské značení polovodičových diod.

Typové označení obvykle nese znaky: výchozí materiál a použití. Materiál polovodičové součásti je uveden vždy v prvním znaku (G nebo A = germanium, K nebo B nebo S = křemík, KS = Schottkyho dioda.) Druhý písmenový znak řadí diody podle použití:

A - všeobecně slaboproudé,
 B - varikapy a varaktory,
 Y - všeobecně silnoproudé,
 Z - Zenerovy diody.

Mezní údaje podle katalogu je nutné ve všech aplikacích dodržet. Charakteristické údaje slouží jen pro prvotní orientaci při výběru součástí. Bližší údaje je nutné hledat v konstrukčním katalogu nebo v katalogových listech, které podávají konstruktérům obvodů bližší údaje o vlastnostech součástí. Ani v nich však často nenajdeme dost přesné informace o vlastnostech použité součásti. Při prvním návrhu obvodu je použijeme pro odhady parametrů použité součásti. Nenajdeme-li potřebný údaj v katalogovém listu použitého typu součástky, můžeme si vypomoci vyhledáním podobného typu v katalogu jiného výrobce. Pro porovnání použijeme mezní údaje, u diod především závěrné napětí (souvisí s tloušťkou přechodu), maximální proud (souvisí s plochou přechodu) a výkonová ztráta (souvisí s odvodem tepla do okolí).

Pro výběr diod v konstrukcích slaboproudých obvodů často potřebujeme alespoň informativní údaj o jejich kapacitě. Pro hrubou informaci proto uvedeme „pořadník“ diod podle průchozích kapacit při nulovém napětí.

Typ	KA 136	KA 207	KA 225	KA 261	KZ 260
C[pF]	1,5	2	3	10	50 až 400

4 POLOVODIČOVÉ SOUČÁSTKY BEZ AKTIVNÍHO PŘECHODU PN

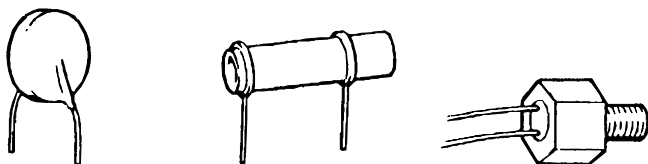
4.1 Součástky z monokrystalických polovodičů bez aktivního přechodu PN

Mezi součástky z monokrystalických polovodičů bez aktivního přechodu PN můžeme zařadit Hallovy sondy, magnetorezistory, Gunnovy diody a tenzoelektrické polovodičové součástky.

4.2 Součástky z polykrystalických polovodičů

Součástky vyrobené na bázi polykrystalických polovodičů jsou z hlediska svých elektrických vlastností blízké rezistorům. Tyto prvky neobsahují přechod PN, jejich voltampérová charakteristika je souměrná vzhledem k počátku, nevykazují usměrňovací účinek. Protože vodivost polovodičových materiálů je závislá na vnějších vlivech (teplota, světlo, elektrické pole), mohou být vlastnosti součástek vyrobených z těchto materiálů závislé na uvedených vlivech.

Podle dominantního vlivu na elektrické parametry rozlišujeme termistory, tj. rezistory jejichž odpor závisí na teplotě. Odpor termistoru může při růstu teploty monotónně klesat, tj. teplotní součinitel odporu je záporný, termistory tohoto druhu jsou označovány jako termistory NTC (Negative Temperature Coefficient), někdy jen jako termistory. V případě, že odpor termistoru v obvykle využívané pracovní oblasti při růstu teploty roste, je teplotní součinitel odporu kladný. Termistory tohoto druhu jsou označovány jako termistory PTC (Positive Temperature Coefficient), někdy též jako pozistory.



Obr. 4.1.: Různá provedení termistorů

Varistory jsou polovodičové rezistory, jejichž odpor je závislý na přiloženém napětí, varistory jsou též označovány zkratkou VDR (Voltage Dependent Resistor)

Fotorezistory jsou polovodičové rezistory, jejichž odpor závisí na osvětlení, nikdy nejsou jako fotonky apod. zdrojem elektrické energie, s osvětlením pouze klesá jejich elektrický odpor.

5 BIPOLÁRNÍ TRANZISTORY

Objev a realizace bipolárního tranzistoru [transfer resistor] způsobily převrat v elektronice. Tranzistor (bipolární a později také unipolární, který byl zkonstruován v padesátých letech) se stal základní součástkou a z většiny aplikací postupně vytlačil vakuové elektronky. Tranzistorová struktura je rovněž základem (bipolárních a unipolárních) monolitických integrovaných obvodů. Také výkonové tranzistory hrají v současné době důležitou úlohu, zejména ve středofrekvenční oblasti.

Na obrázku 5.1 je schematický řez strukturou bipolárního tranzistoru spolu s vhodným polarizačním napětím a schematickou značkou. Ta připomíná původní hrotový tranzistor, takže se ve značce dochovalo kousek historie. Podstatné je, že šipka u emitoru směřuje z kruhu, tzn. směrem od báze. Občas používaný slogan jednoznačně praví: "NPN šipka ven". Jiným umístěním nebo spíš směrem šipky se liší značka tranzistoru opačné vodivosti PNP. Střední elektroda se označuje jako báze B. Krajní elektroda, polarizovaná vůči bázi v propustném směru, se nazývá emitor E, elektroda, polarizovaná vůči bázi v závěrném směru, je kolektor C. Nemá-li báze polarizována vnějším napětím a mezi kolektor a emitor je zapojen zdroj napětí, prochází mezi kolektorem a emitorem tranzistoru pouze nepatrný, tzv. zbytkový proud.

Polarizací báze vůči emitru v propustném směru se proud mezi emitorem a kolektorem zvětšuje. Tento proud je několikanásobně větší než proud, který prochází bází. Poměr těchto dvou proudů je nejzákladnějším parametrem tranzistoru a označuje se jako proudový zesilovací činitel h_{21} . Polarita tranzistoru PNP nebo NPN nemá na jeho činnost téměř vliv a je lhostejné, jaký druh tranzistoru pro výklad použijeme.

V obvodech se však tranzistory NPN a PNP liší připojením napájecího napětí k emitru a kolektoru. Rovněž na bázi bude potřeba odlišit polaritu napětí. To ostatně vyplývá z poznatků o závěrném a propustném směru polovodičového přechodu. Samotné polovodičové přechody tranzistoru se chovají jako diody, můžeme se o tom přesvědčit měřením ohmetrem.

Obr. 5.1: Zjednodušený model tranzistoru NPN a PNP

Obr. 5.3. Základní zapojení bipolárního tranzistoru: a) po připojení napětí k emitoru a kolektoru prochází obvodem C-E jen nepatrný (zbytkový) proud, báze není připojena, b) když bázi připojíme k malému kladnému napětí vzhledem k emitoru, začne obvodem C-E procházet proud, obvodem B-E přitom prochází jen malý proud I_B

Zjednodušeně bychom tedy mohli tranzistor znázornit dvojicí diod (obr. 5.1). Obě vodivosti, tedy NPN i PNP mají dvojice diod postaveny vzájemně opačně. Obvod C-E je u obou vodivostí pólován v závěrném směru, ať připojíme zdroj jakkoliv. Je to pochopitelné, protože když jedna dioda z dvojice bude pólována v propustném směru, druhá dioda bude zároveň pólována obráceně, tedy v nepropustném směru. Měřením odporu mezi elektrodami C-E můžeme získat základní informaci o stavu tranzistoru. Bez ohledu na způsob připojení

Obr. 5.2. K orientační kontrole bipolárního tranzistoru lze použít ohmetru (s měřicím napětím cca 1,5 V) ohmetru, přesněji řečeno na polaritu baterie v ohmetru, musí vykázat obvod C-E poměrně velký odpor. Dalším měřením ověříme stav emitorového a kolektorového přechodu (obr. 5.2).

To co bylo řečeno o dvojici diod zjednodušeně nahrazující tranzistor, obecně platí a hodí se při zběžné zkoušce tranzistoru. Neobjasňuje to však funkci tranzistoru. Ta je složitější a spolupracují tu všechny tři oblasti. Jak tranzistor pracuje, anebo jak vzniká tranzistorový jev? Můžeme se zeptat i jinak: "Co musí nastat, aby obvodem E-C, o kterém víme, že je pólován vždy v závěrném směru, aby jím začal procházet proud? A dále, jak zařídit, abychom takto vzniklý proud dokázali využít?"

Uvažujeme obvod s tranzistorem NPN podle obr. 5.3a, kde pro názornost měříme proud každé elektrody ampérmetrem. Podle očekávání zjistíme, že obvodem teče jen nepatrný proud. Naznačují to měřidla, jejichž výchylka ručky je velmi malá. Nyní se podíváme, jak se projeví vliv báze, když ji připojíme k malému napětí vzhledem k napětí emitoru. Bude to kladné napětí, ale velice malé, protože víme, že obvod B-E je pólován v propustném směru. Znamená to, že v obvodu se uplatňuje jen malý odpor, který nebrání průchodu proudu. Měřidla na obr. 5.3b znázorňují, že nyní prochází emitorem i kolektorem velký proud. Zdůrazněme, že tento emitorový nebo častěji řečený kolektorový proud prochází jen tehdy, když současně prochází proud bází, třebaže nepatrný. To ostatně znázorňuje měřidlo zapojené v bázi. Můžeme říci, že malým proudem v obvodu báze-emitor vyvoláme velký proud v obvodu emitor-kolektor. Žádným jiným způsobem nelze (za normálních okolností) závěrně polarizovaný obvod C-E otevřít.

5.1 Princip činnosti bipolárního tranzistoru

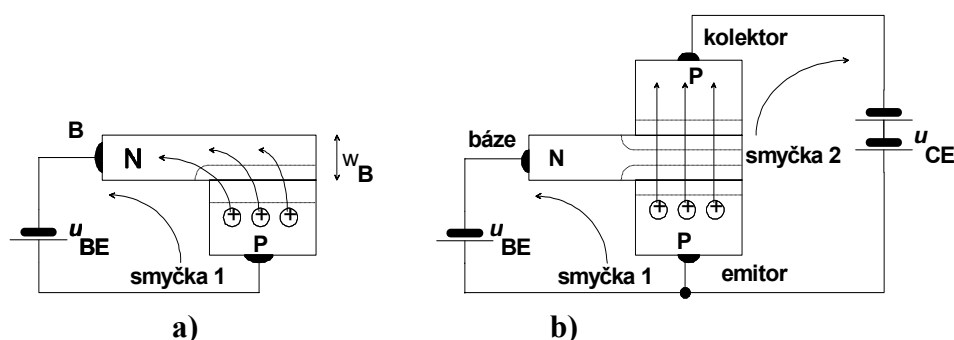
Činnost tranzistoru je založena na *interakci* mezi dvěma přechody PN. Pojem interakce dvou přechodů můžeme vysvětlit, uvažujeme-li dva přechody PN, z nichž jeden působí jako *injektor* nosičů do polovodiče, ve kterém je v určité vzdálenosti umístěn další přechod PN. Není obtížné dokázat, že při vzdálenosti obou přechodů větší než je pětinasobek difuzní délky injektovaných nosičů, se může druhému přechodu přiblížit necelé jedno procento těchto nosičů. Je zřejmé, že pokud má k nějakému vzájemnému ovlivňování obou přechodů dojít, musí být jejich vzdálenost podstatně menší.

Základní myšlenka činnosti tranzistoru může být ilustrována pomocí obr. 5.4. Na obr. 5.4.a) je zobrazen přechod PN s velmi úzkou oblastí N. Je-li oblast P polovodiče silně legována a zdroj u_{BE} je zapojen tak, že přechod PN je v propustném směru, dochází k silné injekci děr do oblasti N. Tato injekce je řízena napětím u_{BE} . Protože oblast N je velmi úzká, je za přechodem přeplněna nadbytečnými injektovanými děrami. Proud je určen difúzí těchto nadbytečných děr ke kontaktu B a rychlostí rekombinace.

Obr. 5.4.b) ukazuje další oblast P umístěnou za oblastí N blízko již existujícího přechodu PN. Napětí u_{CE} je takové, že druhý přechod je polarizován v závěrném směru. Za této situace převážná většina děr emitovaných do oblasti N původním přechodem opustí tuto oblast tak, že se bude difúzním pohybem přibližovat k nové oblasti P. Protože tato oblast je velmi blízko (zatímco kontakt B je velmi daleko) a elektrické pole depletiční vrstvy dalšího přechodu působí jako "nora" pro injektované díry, přejdou tyto díry do oblasti P. Vlivem druhého přechodu tak byl proud ze "smyčky 1" přemístěn do "smyčky 2", ale tento proud je stále řízen napětím zdroje u_{BE} .

Struktura znázorněná na obr. 5.4.b) tvoří tranzistor. Jak je zřejmé, je tranzistor tvořen monokrystalem polovodiče se třemi oblastmi různé vodivosti. Může být typu PNP nebo NPN podle vodivosti jednotlivých oblastí. Oblast, která *emituje* nosiče náboje do mezilehlé oblasti se nazývá **emitor** (se zkratkou **E**), prostřední oblast se nazývá **báze (B)**; její potenciál *řídí emisi*. Třetí vrstva, *odčerpávající* minoritní nosiče báze, které tam byly injektovány emitorem, se nazývá **kolektor (C)**. Jednotlivé oblasti s různým typem vodivosti jsou opatřeny ohmickými kontakty a přívodními vodiči. Přechod PN mezi bází a emitorem se nazývá *emitorový*, přechod mezi kolektorem a bází se nazývá *kolektorový*.

Abychom zajistili, že co největší počet děr vstříknutých do oblasti N dosáhne závěrně polarizovaného přechodu PN a nikoliv ohmického kontaktu oblasti N, umístíme tento kontakt do vzdálenosti několika difúzních délek od místa vstříku nosičů, zatímco druhý přechod ve vzdálenosti menší než difúzní délka vstříknutých nosičů. Kromě toho je plocha závěrně polarizovaného kolektorového přechodu větší než plocha vstříkujícího emitorového přechodu, aby byly zachyceny i nosiče rozptýlené do stran.



Obr. 5.4. Princip činnosti bipolárního tranzistoru: a) proud polovodičové diody se silně dotovanou oblastí P může být "přemístěn" ze smyčky 1 do smyčky 2, b) vytvořením dalšího PN přechodu v blízkosti původního přechodu; nový přechod je polarizován v závěrném směru.

činnost tranzistoru, lze shrnout následovně:

1. Šířka báze je menší než střední difúzní délka minoritních nosičů v bázi

$$w_B \ll L_{pB}$$

(5.1a)

2. Kontakt báze je několik středních difúzních délek minoritních nosičů vzdálen od emitorového přechodu

$$|B - J_C| \gg L_{pB} \quad (5.1b)$$

3. Koncentrace příměsí v emitru je mnohem vyšší (10^2 až 10^4 krát) než koncentrace příměsí v bázi

$$N_{AE} \gg N_{DB} \quad (5.1c)$$

4. Aby tranzistor dobře zesiloval, musí být plocha kolektoru větší než plocha emitoru
 $S_C > S_E$ (5.1d)

Struktura tranzistoru znázorněná na obr. 5.4.b) je geometricky symetrická. Oblast emitoru a kolektoru, stejně jako plochy emitorového a kolektorového přechodu jsou stejné. Je-li navíc stejná koncentrace příměsí v oblasti kolektoru a emitoru, budou i elektrické vlastnosti takového tranzistoru stejné, změníme-li emitor za kolektor. Takový tranzistor se nazývá symetrický, jeho vlastnosti v normálním zapojení (emitor pracuje jako emitor, kolektor pracuje jako kolektor) i v inverzním zapojení (kolektor pracuje jako emitor, emitor pracuje jako kolektor) jsou stejné.

Převážná většina dnes vyráběných tranzistorů je nesymetrická (vodivostí kolektoru a emitoru i geometricky), jejich vlastnosti se v normálním a inverzním zapojení od sebe výrazně liší.

Každý ze dvou přechodů tranzistoru může být zapojen buď v propustném nebo v závěrném směru. Podle toho rozlišujeme čtyři (příp. pět, uvažujeme-li i průraz jako režim) základní pracovní režimy tranzistoru - jeho provoz v různých oblastech statických charakteristik:

1. OBLAST I: Emitorový i kolektorový přechod jsou polarizovány v závěrném směru. Tranzistorem prochází jen nepatrný proud a tento případ nazýváme *závěrným režimem*.
2. OBLAST IIa: Emitorový přechod je zapojen v propustném směru, kolektorový přechod v závěrném směru. Toto je *normální aktivní režim* tranzistoru.
3. OBLAST IIb: Emitorový přechod je zapojen v závěrném směru, kolektorový přechod v propustném směru (funkce emitoru a kolektoru jsou zaměněny). Toto je *inverzní aktivní režim* tranzistoru.
4. OBLAST III: Emitorový i kolektorový přechod jsou polarizovány v propustném směru. Tranzistorem prochází poměrně velký proud, tranzistor je v *saturačním* (nasyceném) *režimu*.
5. OBLAST IV: *Průraz* je obvykle nežádoucím režimem činnosti tranzistoru.

Režim I a III nacházíme při spínacím provozu. Režim I představuje rozpojený tranzistor, tj. který nevede. Režim III představuje sepnutý tranzistor, tj. který vede. V režimu II pracuje tranzistor jako zesilovač signálu, a to ve směru buďto od emitoru ke kolektoru (normální), nebo od kolektoru k emitoru (inverzní). Přehledně jsou pracovní režimy tranzistoru shrnuty v tabulce 5.1 a na obr. 5.34a.

Z uvedených pracovních režimů tranzistoru je nejdůležitější normální aktivní režim. Proto nyní výklad zaměříme na tento režim u tranzistoru PNP.

Jak již bylo uvedeno, emitorový přechod vstříkne díry do oblasti báze, kde jsou minoritními nosiči (báze je typu N). Vstříknuté nosiče částečně rekombinují v objemu báze nebo na jejím povrchu, některé mohou dosáhnout vývodu báze a rekombinovat na něm. Většina vstříknutých děr difunduje bázi až ke kolektorovému přechodu a zvětší jeho závěrný proud.

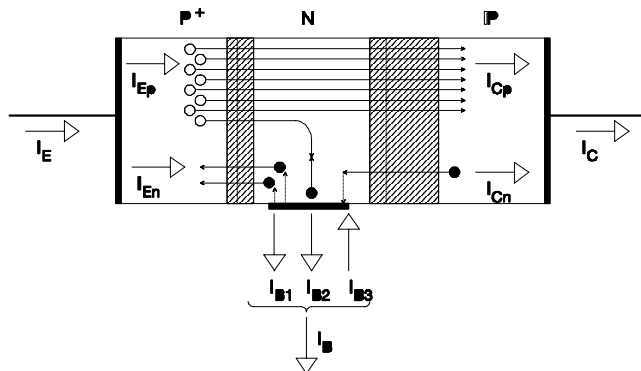
Tab. 5.1.: Pracovní režimy bipolárního tranzistoru

oblast	režim	přechod BE	přechod BC
I	závěrný	zavřen	zavřen
IIa	normální aktivní	otevřen	zavřen
IIb	inverzní aktivní	zavřen	otevřen
III	saturační	otevřen	otevřen
IV	průraz		

Na obr. 5.5. jsou schématicky znázorněny jednotlivé složky proudu v normálním aktivním režimu v tranzistoru PNP. **Emitorový proud i_E** se skládá z elektronového a děrového proudu i_{En} a i_{Ep} a má tyto složky:

a) *Proud děr vstříknutý z emitoru do báze (i_{Ep})*. Tento proud je pro činnost tranzistoru rozhodující, protože jeho část se dostává až ke kolektoru (i_{Cp}) a tvoří nejpodstatnější složku kolektorového proudu.

b) *Proud elektronů vstříknutých z báze do emitoru (i_{En})*. Tento proud nesouvisí s funkcí kolektoru a měl by proto být pokud možno co nejmenší (to je umožněno nesouměrností emitorového přechodu $N_{AE} \gg N_{DB}$).



Obr. 5.5. Složky proudu v tranzistoru PNP

c) *Rekombinační proud emitorového PN přechodu spolu s rekombinačním proudem v objemu emitoru*.

Kolektorový proud i_C tvořený elektronovým a děrovým proudem i_{Cn} a i_{Cp} má tyto složky:

a) *Proud děr vstříknutý emitorem do báze (i_{Cp})*, zmenšený o díry, které rekombinovaly v bázi během difúze ke kolektoru nebo na vývodu báze či jejím povrchu (i_{B2}). Protože jsou splněny podmínky (5.1a) a (5.1b) pro rozměry báze, dosáhne většina injektovaných děr kolektorového přechodu a je vtažena el. polem do kvazineutrální oblasti kolektoru, kde se tím zvýší koncentrace majoritních nosičů a dochází k jejich transportu (difúzí i driftem - vzniklo el. pole, viz poznámka) ke kontaktu kolektoru (i_{Cp}), kde rekombinují.

b) *Závěrným proudem kolektorového PN přechodu (i_{Cn})*, který je tvořen závěrným proudem vyvolaným driftovým mechanismem a proudem vzniklým tepelnou generací nosičů náboje v závěrně polarizovaném kolektorovém přechodu.

Celkový proud báze i_B je tedy tvořen:

a) *Proudem elektronů vstříknutých z báze do emitoru (i_{B1})* propustně pólovaným emitorovým přechodem.

b) *Rekombinačním proudem (i_{B2})*, tvořeným elektrony využitými při rekombinaci s děrami vstříknutými z emitoru.

c) *Závěrným proudem kolektorového přechodu (i_{B3})*.

Poznámka: Pokud koncentrace příměsí v bázi nebude konstantní, vznikne zde urychlující pole a minoritní nosiče se budou pohybovat vlivem difúze i driftu, pokud se oba vlivy sčítají, hovoříme o tzv. *driftových tranzistorech*.

Přehledně se důsledky splnění podmínek (5.1) dají vyjádřit takto:

$$N_{AE} \gg N_{DB} \quad \Rightarrow \quad i_{B1} \ll i_{Ep} \approx i_E$$

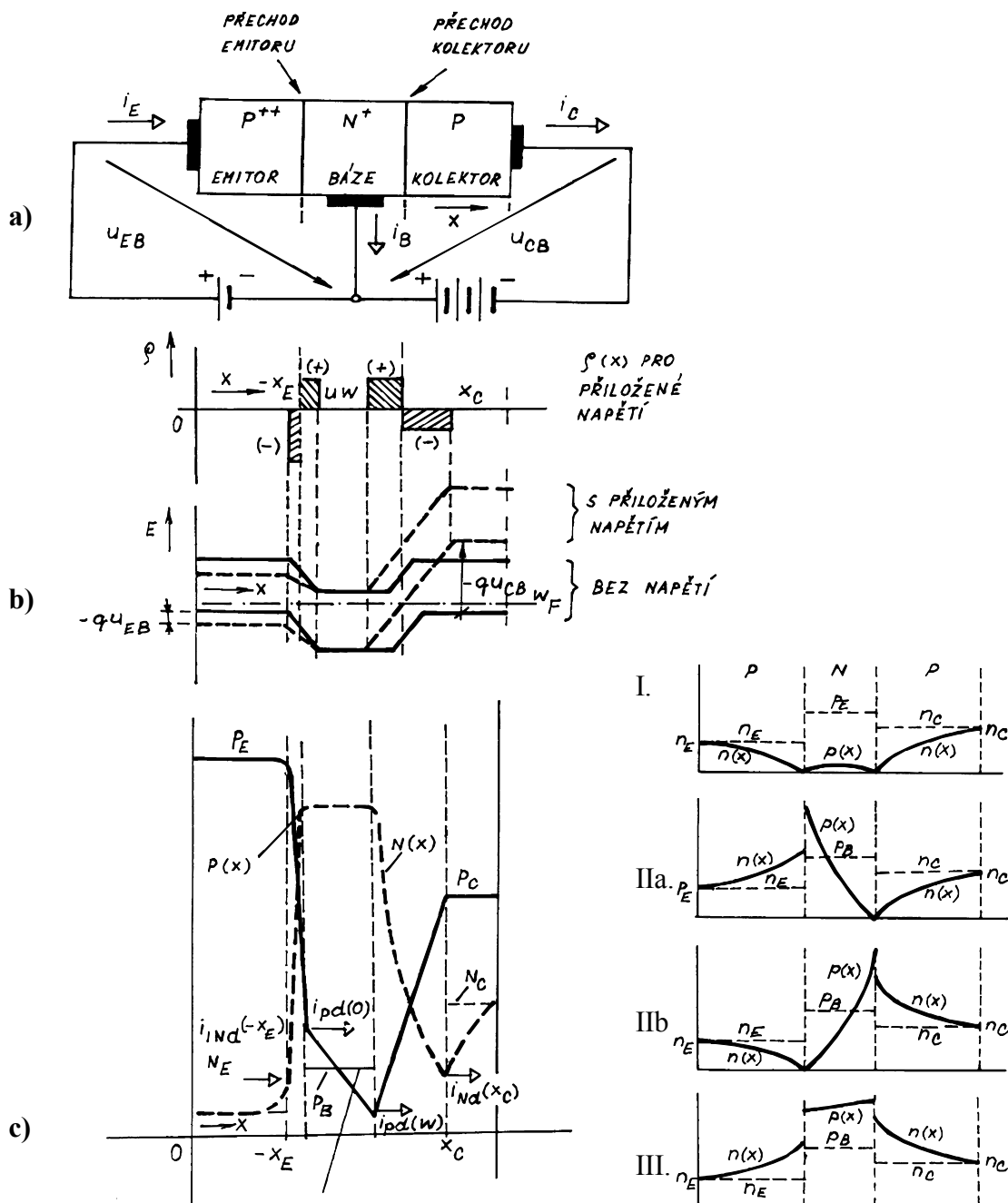
$$w_B \ll L_{pB} \ll |B - J_C| \quad \Rightarrow \quad i_{B2} \ll i_{Ep} \approx i_E$$

Proud děr injektovaný z emitoru i_{Ep} je tedy přibližně roven celkovému proudu emitoru i_E . Celkový proud báze $i_B \ll i_E$ a proud kolektoru i_C je proto jen o málo menší než proud emitoru i_E . Stačí tedy pouze malý bázevý proud, aby bylo přes propustně pólovaný emitor. přechod vstříknuto do báze (a do kolektoru) ohromné množství elektronů.

Při ostatních režimech (I, IIb, III, IV) lze udělat obdobný rozbor. Režim průrazu (IV) se od popsaného režimu IIa liší výraznou nárazovou ionizací v kolektorové depletiční vrstvě. Na obr. 5.6.a) je schématicky znázorněno uspořádání tranzistoru PNP s polaritami napájecích napětí odpovídajícími normálnímu aktivnímu režimu tranzistoru. Na obr. 5.6.d) vidíme, že při závěrném režimu je koncentrace minoritních nosičů v celé bázi tranzistoru nižší než rovnovážná, a proto převládá generace nosičů. Při aktivním režimu je koncentrace minoritních nosičů téměř v celé bázi vyšší než rovnovážná, a proto převládá rekombinace. Při saturačním režimu oba přechody injektují nosiče do báze.

Podstatou činnosti bipolárního tranzistoru je řízené vstřikování minoritních nosičů z emitoru do oblasti báze a jejich extrakce kolektorem (dříve, než stačí v bázi rekombinovat). Z podmínky aktivní funkce tranzistoru vyplývá, že napětí mezi emitorem a bází budou malá, řádově desetiny voltu. Proto bude malý i odpor propustně polarizovaného emitorového přechodu. Napětí mezi kolektorem a bází budou jednotky až desítky voltů a odpor kolektorového pře-

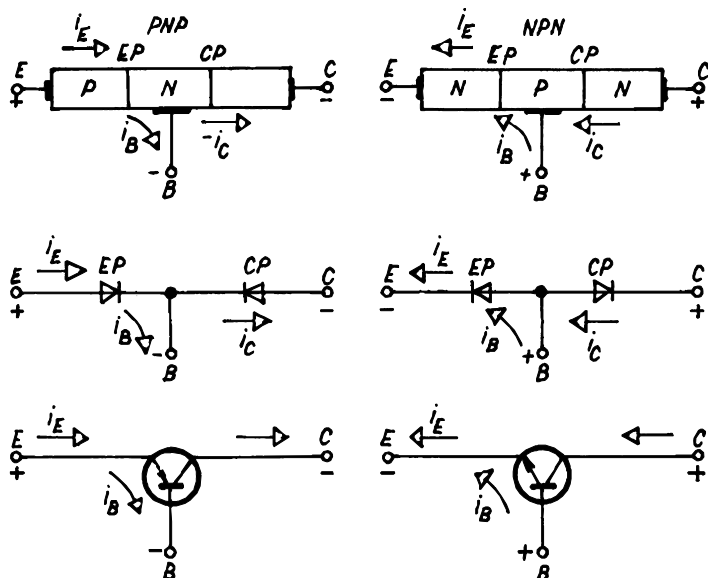
chodu polarizovaného v závěrném směru bude vždy o několik řádů větší než je odpor emitorového přechodu. Emitorový proud je přitom stovnatelný s proudem kolektorovým.



Obr. 5.6.: Princip činnosti bipolárního tranzistoru PNP: **a)** uspořádání tranzistoru s polaritami napájecích napětí odpovídajícími normálnímu aktivnímu režimu; **b)** rozložení hustoty prostorového náboje při přiloženém napětí podle **a)**; **c)** energetický pásový diagram tranzistoru v rovnovážném stavu a s přiloženým napětím; **d)** průběh koncentrace minoritních nosičů v režimech I, IIa, IIb, III

Poznámka: Pro dynamický odpor emitoru můžeme odvodit výraz $r_E = kT/qi_E$, platný na nízkých kmitočtech, kde se neuplatňuje vliv kapacit, a za podmínek malé injekce. Při pokojové teplotě je $r_E \cong 26 \text{ mV}/i_E$. Dynamický odpor emitoru je tedy nepřímě úměrný stejnosměrnému proudu i_E emitoru.

V předchozím výkladu jsme poznali, že bipolární tranzistor je založen na principu *injekce a extrakce* nosičů náboje, přičemž **vedení proudu v bipolárním tranzistoru se uskutečňuje oběma typy nosičů** (proto *bipolární*), tj. elektrony i dírami (z jiného pohledu: v emitoru a kolektoru majoritními, v bázi minoritními nosiči). Později uvidíme, že v unipolárním tranzistoru se na vedení proudu podílejí pouze nosiče jednoho typu (proto *unipolární*). Podle typu vodivosti jednotlivých oblastí rozeznáváme bipolární tranzistory NPN nebo PNP (obr.



Obr. 5.7.: Struktury bipolárního tranzistoru, zjednodušená diodová náhradní schémata a schématické značky tranzistoru PNP a NPN, vyznačení konvenčních směrů proudů

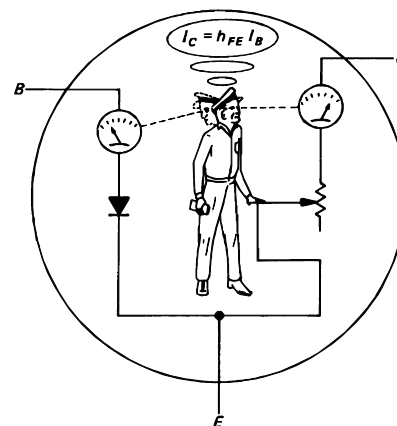
přechodu kolektoru. V podstatě jde o to, že proud vyvolaný vnějším zdrojem na nízké impedanční úrovni (propustně polarizovaný emitorový přechod) protéká jen nepatrně zmenšený i kolektorovým obvodem, ve kterém je zapojen "velký" odpor (odpor kolektorového přechodu polarizovaného v nepropustném směru). Výstupní výkon je mnohem větší než vstupní. Úbytek napětí na výstupním odporu bývá mnohokrát větší než napětí přiložené na vstup tranzistoru, tj. bázi nebo emitor, viz také obr. 5.8. (Z této skutečnosti je odvozený i název tranzistor jako

"TRANSfer resISTOR, tj. přenesený rezistor", původní anglický pravopis "transistor" byl počestěn na "tranzistor".)

Tranzistorový jev vzniká tehdy, jsou-li dva přechody PN v tak těsné blízkosti, že nosiče jednoho přechodu polarizovaného v propustném směru ovlivňují proud druhým přechodem PN, který je polarizován v závěrném směru (tj. báze musí být velmi tenká). Nestačí tedy pouze spojit dva přechody PN.

5.7.). Co do principu činnosti jsou si rovny. Liší se v polaritě napájecích napětí a směrech proudů, případně v hodnotách některých parametrů (např. v důsledku odlišných vlastností elektronů a děr).

Při vhodném připojení napětí mezi jednotlivé elektrody lze procházející proud řídit. Nejčastěji se tranzistor používá ke spínání nebo zesilování proudu (napětí, výkonu). Schopnost zesilovat, označovaná jako *tranzistorový jev*, vzniká při propustně polarizovaném přechodu emitoru a závěrně polarizovaném



Obr. 5.8. Princip funkce tranzistoru podle Horowitz [116]

Poznámka: Z historického hlediska byly první tranzistory hrotové a by-ly typu

PNP nebo spíše PNPN. Již v 50. letech se přestaly vyrábět a nyní se používají výhradně plošné tranzistory, a to v mnoha druzích pro různé účely a ty ještě v mnoha typech. Za zmínku ještě stojí, že schematická značka tranzistoru je vlastně značkou hrotového tranzistoru (šipka v přívodu emitoru vyjadřuje vyjadřuje směr proudu v normálním aktivním režimu), byla navržena zvláštní značka pro plošný tranzistor, ta se však neujala. Kružnice ve schematické značce tranzistoru představuje jeho pouzdro (je to zvyk z doby elektronek, u jiných polovodičových součástek a také u integrovaných tranzistorů pouzdro nekreslíme).

Poznámka: Pro výrobu tranzistorů i při vyspělé technologii je charakteristický určitý rozptyl parametrů. Proto vytříděním podle vhodných parametrů rozčleňují výrobci vyrobené tranzistory do tolerančně užších skupin, které jsou označeny jako *typ* tranzistoru. Jednotlivé typy mohou tvořit ucelené typové řady [ST 5/62:175]. Např. KC 507, KC 508 a KC 509 jsou

Tranzistor v kovovém pouzdře: a) detailpřipojení přívodních drátků k emitoru a bázi, b) přívody emitoru a báze, kolektor je

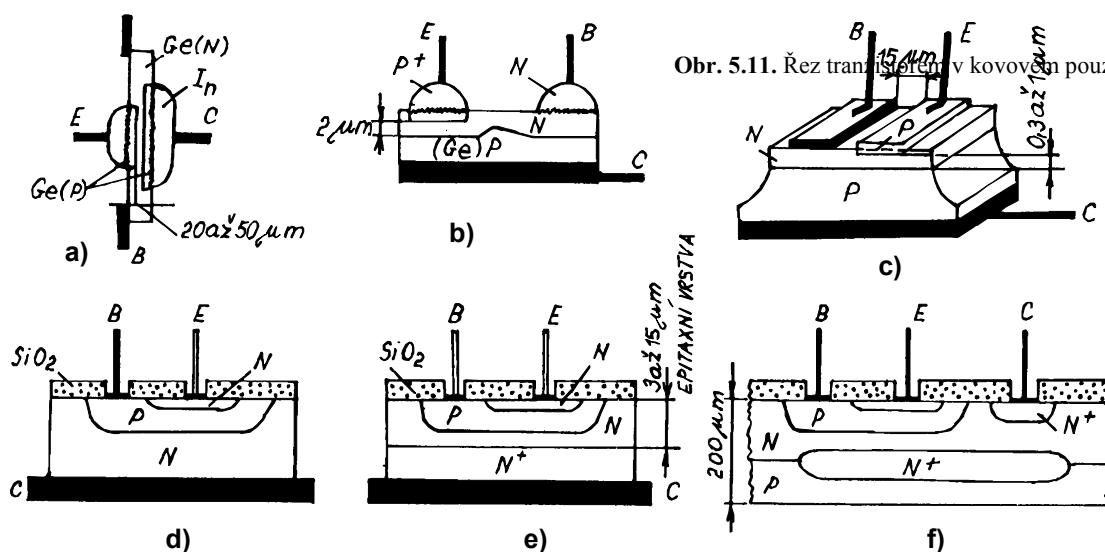
vodivě připojen k podložce

technologicky shodné tranzistory, ale KC 507 je vybrán na větší závěrné napětí a typ KC 509 má největší proudový zesilovací činitel.

Tranzistory se dříve vyráběly z germania, nyní nejčastěji z křemíku, někdy z arsenidu galia (spíše unipolární než bipolární). Podle použití je dělíme na *zesilovací* a *spínací*, podle kmitočtových vlastností na *nízkofrekvenční*, *vysokofrekvenční* a *mikrovlnné*, podle výkonu se od 3 W kolektorové ztráty považují za *výkonové* (někdy od 100 W za vysokovýkonové). Existuje řada technologií výroby tranzistorů (obr. 5.9.).

Pouzdrění a značení tranzistorů. Vlastní polovodičový systém tranzistoru má nepatrné rozměry a je umístěn v různých pouzdrech. Pouzdra se liší nejen tvarem, ale i materiálem. Vyrábějí se z kovu nebo plastu. Kovová pouzdra

Obr. 5.12. Průřez planárním



Obr. 5.9. Tranzistory vyráběné různou technologií: a) slitinový, b) slitinově difúzní, c) mesa, d) planární difúzní, e) planární epitaxní, f) planární epitaxní pro integrované obvody (s ponořenou vrstvou N^+ , která plní funkci ponořeného kolektoru a současně izoluje tranzistorovou strukturu od substrátu vodivosti P)

Obr. 5.10. Tvary pouzder tranzistorů: a) pro malé výkony (napětíové zesilovače), b) vysokofrekvenční tranzistory, c) spínací tranzistory, d) výkonové tranzistory, e) starší germaniové tranzistory

Obr. 5.11. Řez tranzistorem v kovovém pouzdře TO5, kolektor je vodivě připojen

jsou dražší než plastová, zato lépe odvádějí ztrátové teplo z tranzistorového systému. Často se setkáme s kombinací dvou materiálů, zvláště u výkonových tranzistorů. Potom je součástí plastového pouzdra kovová destička, určená k přišroubování na chladič. Jejím prostřednictvím se ztrátové teplo snadno odvádí z tranzistoru. Tvary a velikosti pouzder se u mnohých výrobců prakticky neliší. Proto řada pouzder má již mezinárodní označení, což přispívá k zjednodušení a rychlejší informovanosti.

Nejčastěji budeme pracovat s tranzistory určenými pro napěťové zesilovače (obr. 5.10a). Elektrody mají většinou drátové vývody a jejich rozmístění je standardní. Uprostřed je báze, po stranách emitor a kolektor. Tyto tranzistory snesou jen nepatrné zatížení a jsou rozměrově nejmenší. Jinou skupinu tvoří spínací tranzistory (obr. 5.10b). Vyrábějí se v kovových nebo plastových pouzdrech. Svými vlastnostmi se trochu liší od tranzistorů pro napěťové zesilovače, protože jsou konstruovány pro funkci: proud vede - nevede. Většinou mohou pracovat s většími kolektorovými proudy než tranzistory předchozí skupiny. Vysokofrekvenční tranzistory mají tvar někdy nezvyklý a nezřídka mají čtyři vývody (obr. 5.10c). Čtvrtý vývod si vyžádalo připojení kovového pouzdra (stínění). Tím se zlepšuje činnost tranzistoru při vysokých kmitočtech. Rozmístění elektrod bývá různé.

Zvláštní skupinu tvoří výkonové tranzistory, jejichž starší pouzdro je celokovové (obr. 5.10d) a zdánlivě jim chybí třetí elektroda - kolektor. Ten je spojen s kovovým pouzdrem, a proto musí být tranzistor upevněn izolovaně na chladiči. Novější pouzdra se tvarově liší. Jsou výrobně méně náročná a snadněji se instalují. Z plastového pouzdra vyčnívá na jedné straně kovová destička, na druhé straně tři úzké páskové vývody.

5.2 Kvantitativní analýza funkce bipolárního tranzistoru

5.2.1 Proudové zesílení a vztahy mezi stejnosměrnými proudy tranzist. v základních zapojeních

Základním parametrem tranzistoru je tzv. **proudový zesilovací činitel α** , vyjadřující proudový přenos tranzistoru. Je definován jako poměr změny proudu kolektoru Δi_C k odpovídající změně proudu emitoru Δi_E při určitém napětí kolektoru u_{CE}

$$\alpha = \left. \frac{\Delta i_C}{\Delta i_E} \right|_{u_{CE} = konst.} \quad (5.2)$$

Je zvykem vyjadřovat proudový zesilovací činitel α jako součin třech koeficientů, z nichž každý vyjadřuje určitý mechanismus činnosti tranzistoru:

$$\alpha = \frac{\Delta i_C}{\Delta i_E} = \frac{\Delta i_{Cn} + \Delta i_{Cp}}{\Delta i_{En} + \Delta i_{Ep}} = \frac{\Delta i_{Cn} + \Delta i_{Cp}}{\Delta i_{En} + \Delta i_{Ep}} \cdot \frac{\Delta i_{Ep}}{\Delta i_{Ep}} \cdot \frac{\Delta i_{Cp}}{\Delta i_{Cp}} = \frac{\Delta i_{Ep}}{\Delta i_{En} + \Delta i_{Ep}} \cdot \frac{\Delta i_{Cp}}{\Delta i_{Ep}} \cdot \frac{\Delta i_{Cn} + \Delta i_{Cp}}{\Delta i_{Cp}} = \gamma_E \alpha_T M_C \quad (5.3)$$

První člen je *injekční účinnost emitoru* γ_E . Je žádoucí, aby elektronová složka emitorového proudu Δi_{En} byla co nejmenší a injekční účinnost emitoru γ co největší, tj. blízká jedné. Toho lze dosáhnout především dodržením podmínky $N_{AE} \gg N_{DB}$. Druhý člen je *bázový přenosový* (transportní) *součinitel* α_T , který vyjadřuje, jak velký podíl děrového proudu vstříknutého emitorem do báze dosáhne kolektorového přechodu. Abychom dosáhli co největší bázový

přenosový součinitel (tj. co nejbližší jedné), snažíme se, aby rekombinace vstříknutých děr byla v bázi co nejmenší (šířka báze co nejmenší ve srovnání s difúzní délkou minoritních nosičů, snížení povrchové rekombinace, vývod báze se umísťuje do vzdálenosti mimo dosahu děr pohybujících se difúzí, velká plocha kolektorového přechodu). Třetí člen je *kolektorový multiplikační součinitel* M_C , který charakterizuje vzrůst elektronové složky kolektorového proudu lavinovým jevem. Zvýšení účinnosti kolektoru nárazovou ionizací se nevyužívá, může vést k nestabilní činnosti tranzistoru.

Stejným způsobem, jako byl zaveden proudový zesilovací činitel α z oblasti emitoru do oblasti kolektoru, můžeme zavést **proudový zesilovací činitel** β z oblasti báze do oblasti kolektoru. Je definován jako poměr změny proudu kolektoru Δi_C k odpovídající změně proudu emitoru Δi_B při konstantním napětí mezi kolektorem a emitorem u_{CE}

$$\beta = \left. \frac{\Delta i_C}{\Delta i_B} \right|_{u_{CE} = \text{konst.}} \quad (5.4)$$

Podle Kirchhoffova zákona platí pro proudy emitoru i_E , kolektoru i_C a báze i_B vztah

$$i_E = i_C + i_B \quad (5.5)$$

Vyjádříme-li tedy změnu proudu báze Δi_B jako rozdíl změn $\Delta i_E - \Delta i_C$ a dosadíme do rovnice (5.4), dostaneme po úpravě vztah mezi proudovými zesilovacími činiteli α a β (napětí $u_{CE} = \text{konst.}$)

$$\beta = \frac{\Delta i_C}{\Delta i_B} = \frac{\Delta i_C}{\Delta i_E - \Delta i_C} = \frac{\Delta i_C / \Delta i_E}{(\Delta i_E - \Delta i_C) / \Delta i_E} = \frac{\alpha}{1 - \alpha} \quad (5.6a)$$

$$\boxed{\alpha = \frac{\beta}{1 + \beta}} \quad (5.6b)$$

Činitel α nazýváme také *proudový zesilovací činitel v zapojení se společnou bází* (SB) a činitel β *proudový zesilovací činitel v zapojení se společným emitorem* (SE). Činitel α dosahuje zpravidla hodnot 0,9 až 0,9975, činitel β odpovídající těmto hodnotám je 9 až 399.

Podle počtu svých svorek je tranzistor *trojpólem*. V aplikacích je vždy jedna z jeho svorek určena pro vstup i pro výstup, takže lze tranzistor uvažovat také jako *čtyřpól*. Tranzistor v obvodech se často považuje za čtyřpól již také proto, že čtyřpólová teorie je dobře

Obr. 5.13. Základní zapojení tranzistoru

propracována (je propracována pro lineární prvky, tranzistor je ovšem nelineární, proto při malých signálech a v daném pracovním bodě jeho vlastnosti linearizujeme).

Podle toho, která z elektrod bude společně využita pro vstup a výstup (tj. bude "společná"), mluvíme o *zapojení*

se **společným emitorem** (SE),

se **společným kolektorem** (SK, SC),

se **společnou bází** (SB).

Každé z těchto zapojení má své výhody a nevýhody. Zapojení se **společným emitorem** (obr. 5.13a) se používá nejčastěji. Vstupní signál se přivádí mezi bázi a společnou elektrodu - emitor. Výstupní signál se pak odebírá

z kolektoru, který je připojen k napájecímu napětí U přes zatěžovací rezistor R_Z . Průchodem poměrně velkého kolektorového proudu zatěžovacím rezistorem vzniká na něm velké napětí (úbytek napětí) v porovnání se vstupním napětím.

Protože přechod B-E je pólován v propustném směru, vstupní odpor R_{VST} zapojení je malý. Ve výstupním obvodu je naopak přechod C-E pólován v závěrném směru. Výstupní odpor R_{VYST} v zapojení SE je proto velký. Konkrétně vstupní bývá $500\ \Omega$ až $10\ k\Omega$ a výstupní odpor $10\ k\Omega$ až $20\ k\Omega$. Proudový zesilovací činitel bývá v rozmezí 10 až 500 (i více) podle druhu tranzistoru. Výstupní signál je vzhledem k vstupnímu fázově natočen o 180° (je v protifázi).

Charakteristickou vlastností tranzistoru v zapojení se společným emitorem je jeho výkonové zesílení (je dáno součinem napětového a proudového zesílení), největší ze všech uvedených zapojení, dosahující hodnoty až 20 000.

Zapojení se **společným kolektorem** (obr. 5.13b) má velký vstupní odpor (až $1\ M\Omega$ i více), který vypočteme násobením zatěžovacího odporu R_Z proudovým zesilovacím činitelem. Výstupní odpor je malý (v rozmezí několika desítek až stovek ohmů). Výstupní napětí je se vstupním ve fázi.

Protože výstupní napětí tu má přibližně stejnou velikost jako vstupní, nazývá se toto zapojení též emitorový sledovač. Používá se k dosažení nezávislosti výstupního napětí na zatěžovacím odporu R_Z (tj. tam, kde vyžadujeme poměrně velký vstupní a malý výstupní odpor). Výkonové zesílení bývá až 500.

Zapojení se **společnou bází** (SB) má proudové zesílení menší než 1, napětové až 1000 a výkonové až 200. Vstupní odpor dosahuje pouze $100\ \Omega$ až $1000\ \Omega$, výstupní naopak až $1\ M\Omega$. Ani zde nedochází k fázovému natočení procházejícího signálu. Dnes se používá spíše výjimečně, v minulosti se běžně zapojení SB používalo ve vysoko-frekvenčních obvodech kvůli vyššímu meznímu kmitočtu (uvidíme později, že mezní kmitočet f_α je řádově větší než mezní kmitočet f_β) a umožnilo dosáhnout zvláště malý vstupní odpor, který byl podmínkou ke stabilní činnosti vysokofrekvenčního zesilovače.

Porovnáme-li výše uvedená zapojení tranzistoru podle jejich hlavních vlastností, zjistíme, že nejvýhodnější vlastnosti má zapojení se společným emitorem (SE), a to se také nejčastěji používá. Jedině u tohoto zapojení se setkáváme s proudovým, napětovým i výkonovým zesílením. Rozdíl mezi velikostí vstupního i výstupního odporu je u zapojení SE nejmenší (to je výhodné při spojování několika tranzistorových zesilovacích stupňů), zapojení SC má vstupní odpor mnohem menší než výstupní, zapojení SB má naopak vstupní odpor mnohem menší.

Rozdělení proudů a napětí na tranzistoru.

Hovoříme-li o činnosti tranzistoru, užíváme předem domluvených pojmů, ať již v souvislosti s napětím či proudem. Pojmenujme si proudy a napětí na tranzistoru, s nimiž se často setkáváme a které je třeba rozlišovat. Poslouží k tomu zapojení tranzistoru SE (obr. 5.14 a 5.15).

Obr. 5.14. Napětí na tranzistoru: a) stejnosměrné v ustáleném stavu (bez buzení střídavým signálem), c) časově proměnné veličiny, c) komplexní amplitudy (při buzení harmonickým signálem)

Ve všech případech se jedné o stejnosměrné proudy a stejnosměrná napětí. Proto v indexu u veličin píšeme velká písmena. Hodnoty zapsané u těchto veličin budou platit za podmínek, že na bázi nedochází ke změně napětí a proudu. Tím je zaručeno, že ani ostatní hodnoty se nebudou měnit. Jde vlastně klidový stav tranzistoru.

Obr. 5.15. Proudů tranzistoru (elektrod tranzistoru): a) stejnosměrné v ustáleném stavu (bez buzení střídavým signálem), c) časově proměnné veličiny, c) komplexní amplitudy (při buzení harmonickým signálem)

Stejnoseměrné hodnoty musí být vždy správně nastaveny. Jen tak je splněn předpoklad pro správnou činnost tranzistoru.

Obr. 5.16. Měření stejnosměrného napětí

Z předcházejícího výkladu víme, že emitorem prochází společný proud kolektoru a báze. Vyjádřeno rovnicí $I_E = I_C + I_B$. Napětí mezi kolektorem a emitorem je složeno ze dvou napětí menších. Sčítá se tu napětí U_{CB} a U_{BE} . Vyjádřeno rovnicí $U_{CE} = U_{CB} + U_{BE}$. Poznali jsme, že U_{BE} je velmi malé a u zesilovače bývá v rozmezí 0,6 až 0,7 V. Překročíme-li uvedené napětí, prochází obvodem B-E velký proud a polovodičový přechod mezi bází a emitorem se může zničit.

Napětí U_B , U_E a U_C jsou napětí proti nulovému (zápornému) potenciálu. Názorněji to ukazuje obr. 5.16. Z obrázku je patrné, že U_E je napětí na emitoru měřené proti zápornému pólu zdroje (platí pro NPN). Je to malé kladné napětí, řádově desetiny voltu, které se vytvoří průchodem emitorového proudu na rezistoru R_E . Napětí na bázi U_B je opět měřeno proti zápornému pólu zdroje. Rozhodně bude vyšší než napětí U_{BE} . Musí být vyšší o úbytek napětí na rezistoru R_E , takže je-li napětí $U_E = 0,3$ V, pak napětí $U_B = 1$ V. Obvyklou hodnotu napětí mezi bází a emitorem ($U_{BE} = 0,7$ V) je třeba zachovat. Napětí na kolektoru U_C se může lišit jen málo od napětí U_{CE} . Rozhodně to platí pro běžná zapojení, kde rezistor R_E má malou hodnotu odporu. V ojedinělých případech bude nutno vzít v úvahu (odečíst) napětí U_E , které se vytvořilo na velkém (výjimečně) odporu R_E . V této souvislosti se nabízí otázka, jaké napětí bychom naměřili mezi kolektorem a emitorem, tedy U_{CE} u tranzistoru, který je v činnosti.

I_B - proud báze
 I_C - proud kolektoru
 I_E - proud emitoru
 U_{BE} - napětí mezi bází a emitorem
 U_{CB} - napětí mezi kolektorem a bází
 U_{CE} - napětí mezi kolektorem a emitorem
 U_S - napětí báze proti nulovému potenciálu
 U_C - napětí kolektoru proti nulovému potenciálu
 U_E - napětí emitoru proti nulovému potenciálu

Na otázku nelze jednoznačně odpovědět, protože napětí tu mohou být velmi rozdílná. Záleží totiž na tom, k jakému účelu se tranzistor používá, neboli jak je zapojen. Tak např. zastává-li tranzistor funkci zesilovače napětí, pak stejnosměrné napětí U_{CE} nastavujeme předem na poloviční napětí zdroje (obr. 5.17). Není-li báze připojena, nebo neprochází-li z jiných důvodů proud I_C , pak napětí U_{CE} dostoupí až na velikost napětí zdroje. Naopak mezi kolektorem a emitorem můžeme jindy naměřit nepatrné napětí, např. $U_{CE} = 0,5$ V. Je to v případě, že tranzistor pracuje ve spínacím režimu. Takovému napětí říkáme zbytkové - saturační a u běžných tranzistorů

bývá v rozmezí 0,4 V až 1 V. Je to nejnižší možné napětí mezi kolektorem a emitorem.

Nyní najdeme vzájemné vztahy mezi stejnosměrnými proudy tranzistoru. Nejprve vyjádříme proud kolektoru pomocí proudu emitoru v **zapojení SB**. Víme už, že kolektorový proud tranzistoru i_C obsahuje dvě složky, proud minoritních nosičů vstříknutý emitorem do báze $i_{Cp} = \alpha i_{Ep} \approx \alpha i_E$ a závěrný (zbytkový) proud kolektorového přechodu $i_{Cn} = I_{CB0}$, potom

$$i_C = \alpha i_E + I_{CE0} \quad (5.7)$$

Proud I_{CB0} nezávisí na proudu emitoru a při konstantní teplotě téměř nezávisí na napětí u_{CE} .

Vztah mezi proudem kolektoru a proudem báze v **zapojení SE** dostaneme tak, že vypočteme proud i_E z i_C a dosadíme do (5.5)

$$(5.7)$$

$$i_C = \frac{\alpha}{1 - \alpha} i_B + \frac{1}{1 - \alpha} I_{CB0}$$

Obr. 5.17. Na pracovním rezistoru RC v kolektoru vzniká průchodem klidového proudu napětí. Jeho velikost nastavujeme nejčastěji tak, aby klidové napětí na kolektoru odpovídalo polovočinnému napětí zdroje

(5.8)

který s použitím vztahu (5.6) a rovnice $1/(1 - \alpha) = 1 + \beta$ upravíme na tvar

$$i_C = \beta i_B + (1 + \beta) I_{CB0} = \beta i_B + I_{CE0} \quad (5.9)$$

kde

$$I_{CE0} = (1 + \beta) I_{CB0} = \frac{1}{1 - \alpha} I_{CB0} \approx \beta I_{CB0} \quad (5.10)$$

je zbytkový proud tranzistoru při $i_B = 0$.

Použitím vztahu (5.5) můžeme vyjádřit proud emitoru pomocí proudu báze

v **zapojení SC**

$$i_E = i_C + i_B = (1 + \beta) (i_B + I_{CB0}) = (1 + \beta) i_B + I_{CE0} \quad (5.11)$$

Poznamenejme, že pro moderní křemíkové tranzistory lze zbytkové proudy zanedbat a dostáváme v praxi používaný tvar rovnic (5.7) a (5.9)

$$i_C = \alpha i_E \quad (5.12)$$

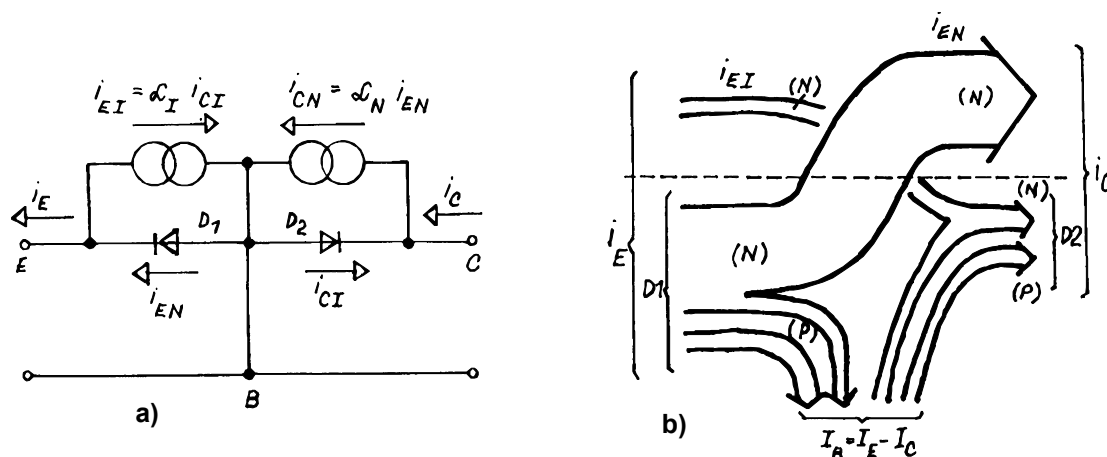
$$i_C = \beta i_B \quad (5.13)$$

5.2.2 Ebersův-Mollův model bipolárního tranzistoru NPN

Při odvozování vztahů mezi proudy a napětími na tranzistoru je možné v prvním přiblížení použít jednorozměrného modelu podle obr. 5.18a). Tento model je v literatuře znám jako *Ebersův-Mollův model* (EM model). Vlastní vztahy mezi proudy a napětími a jejich souvislost s fyzikálními vlastnostmi a geometrií použitého polovodiče lze teoreticky odvodit z rovnice kontinuity.

EM model je založen na superpozici normálního a inverzního režimu tranzistoru za předpokladu lineárního gradientu nosičů náboje v bázi. Všeobecně se pro vystižení principu a popis EM modelu užívá náhradní zapojení dle obr. 5.18a (platí pro tranzistor NPN, pro PNP se "obrátil" diody a orientace obvodových veličin). Vychází se ze skutečnosti, že každý přechod tranzistoru se může chovat jako samostatná dioda. Tranzistorový jev je do modelu zaveden pomocí proudových zdrojů, jejichž proudy jsou řízeny proudy diodami druhého přechodu (tj. "do kříže"). Model pracuje ve všech čtyřech oblastech činnosti tranzistoru.

Kvalitativní představu o proudových poměrech bipolárního tranzistoru nám poskytne obr.



Obr. 5.18. a) Ebersův-Mollův model bipolárního tranzistoru NPN (pro tranzistor PNP je třeba obrátit orientaci diod, orientaci proudů řízených zdrojů a orientaci svorkových napětí a proudů), **b)** kvantitativní znázornění proudů

5.18b). Proud emitoru máme rozdělený na dvě části: na proud náhradní diody $D1$ (ve smyslu obr. 5.18a) a na proud řízeného zdroje. Obdobně na dvě části je rozdělený proud kolektoru. Oprávněnost zvolených proporcí by ukázal hlubší rozbor.

Podrobné řešení je však příliš složité a přesahovalo by rámec těchto skript. Poznamenejme však, že všechna základní řešení vycházejí z předpokladů:

- báze tranzistoru je zhotovena z nedegenerovaného polovodiče;
- úroveň injekce minoritních nosičů je tak nízká, že neovlivňuje podstatně vodivost báze;
- velikost stejnosměrného napětí kolektoru je mnohem větší než hodnota U_T ;
- zanedbatelný úbytek napětí na odporu báze, což znamená, že proud procházející bází do kolektoru (závěrný proud kolektorového přechodu) je velmi malý;
- zanedbatelná rekombinace na povrchu tranzistoru;
- homogenní rozložení proudové hustoty v celé ploše přechodů PN.

Znění rovnic Ebersova-Mollova modelu si můžeme pro jednoduchost intuitivně "odvodit" z modelu na obr. 5.18a). Index N značí tzv. *normální alfa*, tj. proudový zesilovací činitel v normálním aktivním režimu, někdy se užívá ve stejném významu index F [forward], obdobně I značí tzv. *inverzní alfa*, někdy se užívá index R [reverse]. U běžných tranzistorů $\alpha_I < \alpha_N$. Proud emitoru je podle obr. 5.18a)

$$i_E = i_{EN} - \alpha_I i_{CI} \quad (5.14)$$

proud kolektoru

$$i_C = \alpha_N i_{EN} - i_{CI} \quad (5.15)$$

Proud emitorovou diodou se rovná

$$i_{EN} = I_{ES} \left(\exp \frac{-u_{EB}}{U_T} - 1 \right) \quad (5.16)$$

a proud kolektorovou diodou

$$i_{CI} = I_{CS} \left(\exp \frac{-u_{CB}}{U_T} - 1 \right) \quad (5.17)$$

I_{ES} a I_{CS} jsou saturační proudy obou diod. Dosazením rovnic (5.16) a (5.17) do rovnic (5.14) a (5.15) dostaneme

$$i_E = I_{ES} \left(\exp \frac{-u_{EB}}{U_T} - 1 \right) - \alpha_I I_{CS} \left(\exp \frac{-u_{CB}}{U_T} - 1 \right) \quad (5.18)$$

a

$$i_C = \alpha_N I_{ES} \left(\exp \frac{-u_{EB}}{U_T} - 1 \right) - I_{CS} \left(\exp \frac{-u_{CB}}{U_T} - 1 \right) \quad (5.19)$$

Doplníme-li ještě rovnici pro proud báze

$$i_B = i_E - i_C \quad (5.20)$$

pak (5.18), (5.19) a (5.20) tvoří soustavu *Ebersových-Mollových rovnic* pro ideální tranzistor typu NPN.

Saturační proudy I_{ES} a I_{CS} nejsou přímo měřitelné, proto je běžnější udávat tzv. zbytkové proudy I_{EB0} a I_{CB0} . (Symbol "0" v indexu udává stav naprázdno, tj. nulový proud zbývajícím vývodem). Zbytkové proudy zavedeme za podmínky:

$$\begin{aligned} i_E &= 0, & u_{CB} &> 0, & u_{CB} &\gg U_T \Rightarrow i_C = I_{CB0} \\ i_C &= 0, & u_{EB} &> 0, & u_{EB} &\gg U_T \Rightarrow i_E = -I_{EB0} \end{aligned}$$

Po dosazení těchto podmínek do (5.12) a (5.13) dostaneme

$$I_{CS} = \frac{I_{CB0}}{1 - \alpha_N \alpha_I} \quad (5.21)$$

$$I_{ES} = \frac{I_{EB0}}{1 - \alpha_N \alpha_I} \quad (5.22)$$

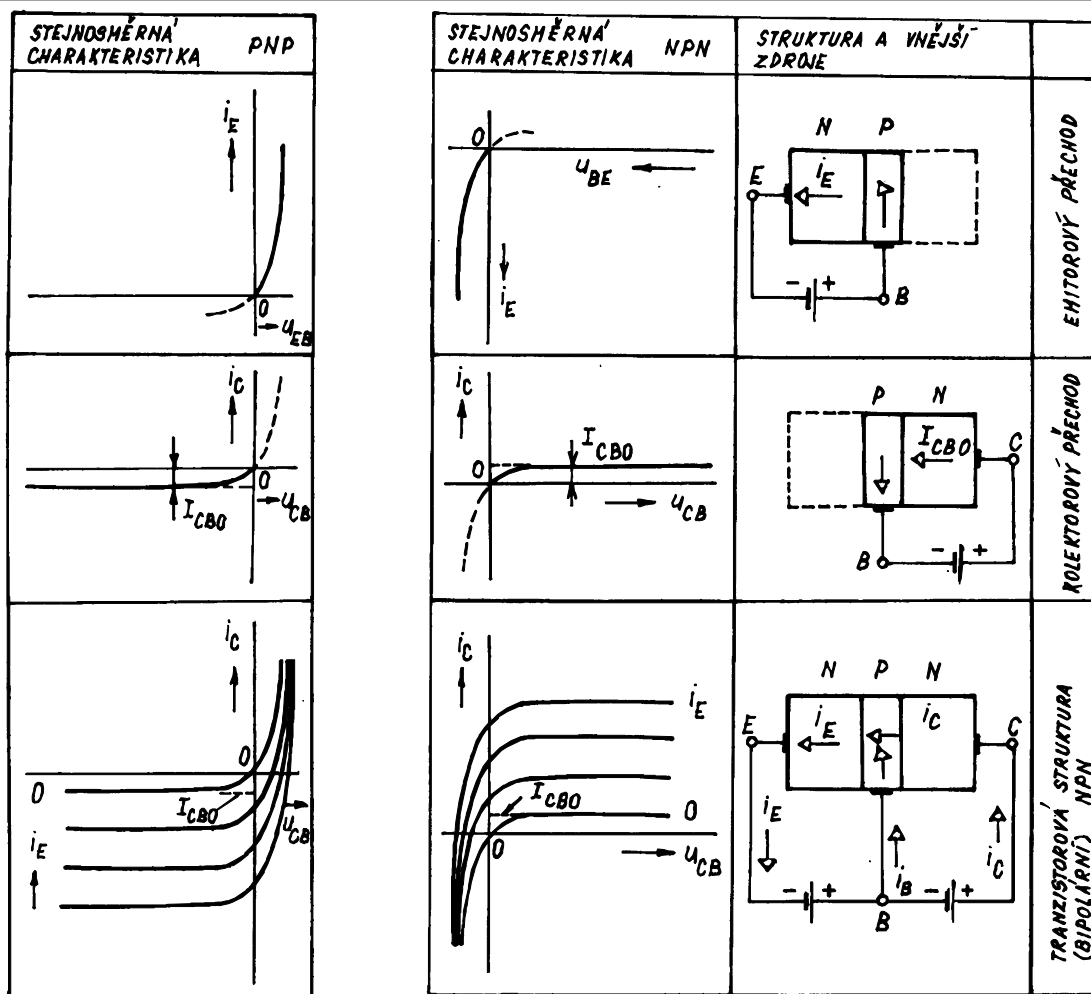
S pomocí (5.21) a (5.22) bychom mohli napsat další tvar rovnic (5.18) a (5.19). Pro praxi je důležité si uvědomit, že v normálním aktivním režimu je možné rovnici pro kolektorový proud uvažovat ve tvaru (5.7), což odpovídá zapojení SB, nebo (5.9), což odpovídá zapojení SE.

5.3 Statické vlastnosti bipolárního tranzistoru

Chování bipolárních tranzistorů se popisuje stejnosměrnými charakteristikami, které se znázorňují graficky a vyjadřují vždy závislost dvou veličin (proudů a napětí), přičemž parametrem je veličina třetí. (Lze ukázat, že existuje celkem 54 různých soustav charakteristik. V praxi jsou důležité pouze některé z nich.)

Charakteristiky tranzistorů nám obecně umožňují:

- a) přehledným způsobem zjistit statické (klidové) hodnoty proměnných veličin v závislosti na pracovních podmínkách (poloze pracovního bodu) a vhodně nastavit vnější napájecí obvody;
 - b) stanovit velikosti parametrů (např. h , y) kvazilineárního režimu malých signálů a odhadnout jejich průběh
v závislosti na poloze pracovního bodu;
 - c) graficky řešit statické poměry při vybuzení do režimu velkých signálů, jak je to potřeba při navrhování výkonových zesilovačů a použití tranzistoru jako spínacího prvku;
 - d) společně se zakreslením mezních čar (I_{Cmax} , U_{CEmax} , P_{Cmax}) zvolit polohu pracovního bodu tak, aby nebyla snížena spolehlivost tranzistoru;
 - e) odhadnout i některé dynamické vlastnosti tranzistoru v kvazistatickém režimu, jako je velikost zkreslení, průběh výstupního signálu při daném průběhu vstupního napětí apod.
- Charakteristiky jsou zakreslovány pro jednotlivá zapojení tranzistoru, nejčastěji pro SB a SE.



Obr. 5.19. Charakteristiky, struktury a vnější zdroje emitorového přechodu, kolektorového přechodu a tranzistorové struktury NPN (v zapojení SB); pro porovnání jsou v levém sloupci nakresleny tytéž charakteristiky pro případ tranzistoru PNP (pro názornost si můžete vedle nakreslit příslušné struktury a vnější zdroje)

5.3.1 Statické charakteristiky tranzistoru NPN v zapojení se společnouází (SB)

Úpravou Ebersových-Mollových rovnic (5.18) až (5.20) je možné odvodit funkční závislosti jednotlivých charakteristik tranzistorů. V zapojení SB se nejčastěji setkáváme s těmito charakteristikami:

1. Výstupní charakteristika naprázdno: $i_C = f(u_{CB}), i_E = \text{konst.}$
2. Vstupní charakteristika nakrátko: $i_E = f(u_{EB}), u_{CB} = \text{konst.}$
3. Proudová převodní charakteristika nakrátko: $i_C = f(i_E), u_{CB} = \text{konst.}$
4. Zpětná napěťová převodní charakteristika naprázdno: $u_{EB} = f(u_{CB}), i_E = \text{konst.}$

Výstupní charakteristiky získáme vyloučením u_{EB} z (5.19) pomocí (5.18) při použití (5.21) a (5.22)

$$i_C = \alpha_N i_E - I_{CB0} \left(\exp \frac{-u_{CB}}{U_T} - 1 \right) \quad (5.23)$$

Z výstupních charakteristik na obr. 5.19. vidíme, že pro $i_E = 0$ charakteristika prochází počátkem a pro $u_{CB} \gg U_T$ je $i_C = I_{CB0}$. Pro $i_E > 0$ se charakteristiky posouvají nahoru o $\Delta i_C = \alpha_N \Delta i_E$. Normální aktivní režim je v 1. kvadrantu nad charakteristikou pro $u_{EB} = 0$ (probíhá mírně nad charakteristikou pro $i_E = 0$; pod touto charakteristikou je závěrný režim), saturační režim je ve 2. a 3. kvadrantu.

Vstupní charakteristiky získáme přímo z (5.18). V normálním aktivním režimu ($u_{CB} > 0$) se charakteristiky se změnou kolektorového napětí příliš nemění.

5.3.2 Statické charakteristiky tranzistoru NPN v zapojení se společným emitorem (SE)

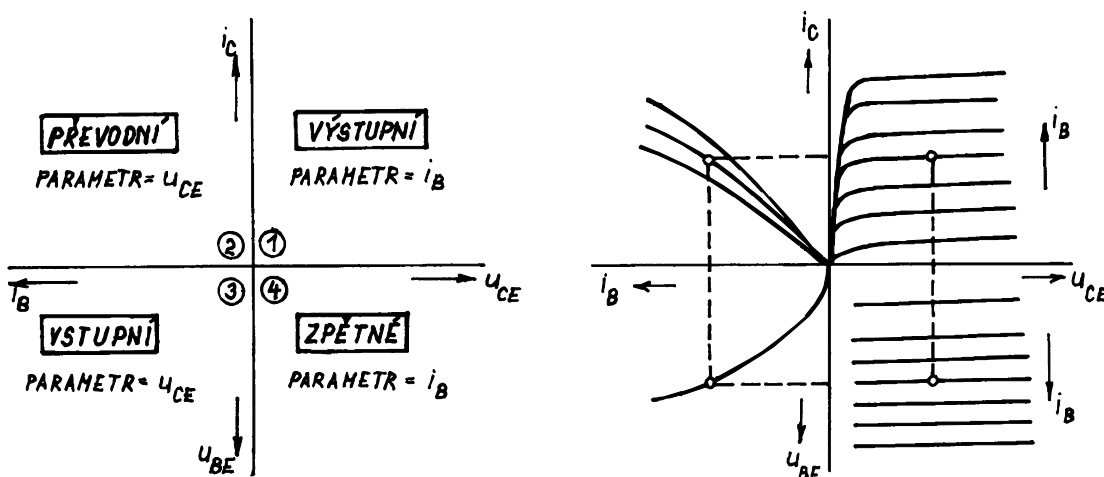
Čtveřice charakteristik pro zapojení SE je:

1. Výstupní charakteristika naprázdno:

$$i_C = f(u_{CE}), i_B = \text{konst.}$$

2. Vstupní charakteristika nakrátko:

$$i_B = f(u_{BE}), u_{CE} = \text{konst.}$$



Obr. 5.20. Soustava charakteristik bipolárního tranzistoru (NPN) v zapojení SE: pojmenování charakteristik (vlevo) a obvyklý tvar (vpravo)

3. Proudová převodní charakteristika nakrátko:

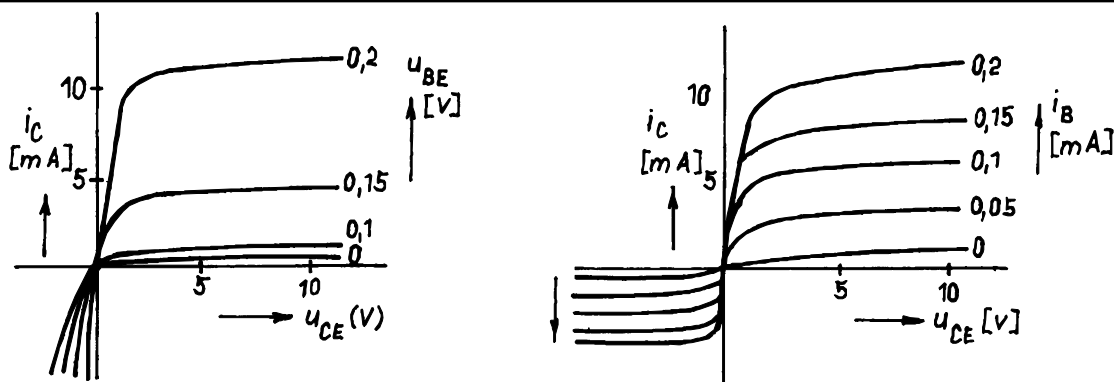
$$i_C = f(i_B), u_{CE} = \text{konst.}$$

4. Zpětná napěťová převodní charakteristika naprázdno:

$$u_{BE} = f(u_{CE}), i_B = \text{konst.}$$

Možný tvar charakteristik pro zapojení SE udává obr. 5.20. Pro různé účely můžeme potřebovat tyto charakteristiky i s jinými parametry. Dále si musíme uvědomit, že charakteristiky na obr. 5.20. i v katalogích jsou umístěny do jednoho kvadrantu (a do společného osového kříže), ale často mají pokračování i v jiných kvadrantech (obr. 5.22). V praxi jsou nejpotřebnější výstupní charakteristiky.

Soustava podle obr. 5.20 se také nazývá hybridní soustava ampérvoltových charakteristik - je to v souvislosti se čtyřpólovými rovnicemi typu h se s odečítáním diferenciálních parametrů jako směrnic jednotlivých křivek. V praxi se někdy namísto převodní charakteristiky zobrazuje dynamická převodní charakteristika (pro konstantní zatěžovací odpor) a zpětné



Obr. 5.22. Výstupní charakteristiky tranzistoru NPN v zapojení SE včetně jejich pokračování ve 3. kvadrantu

výstupní charakteristiky se nezobrazují vůbec (odečítání v nich je nepřesné a vede k chybným výsledkům).

Protože jsou charakteristiky zachyceny jen v jediném kvadrantu, odpovídajícím normálnímu, tedy zesilovacímu režimu tranzistoru, mohou soužit především k odhadu linearizovatelné

$\Delta\vartheta[^\circ\text{C}]$	1	10	20	100
$i(\vartheta) / I(\vartheta_0)$	1,05	1,64	2,7	148

oblasti, do které by měl být umístěn klidový pracovní bod a ve které se při vybuzení může pracovní bod pohybovat, aniž by došlo ke znatelnému omezování signálu při přenosu. V mnoha případech můžeme pro zjednodušení křivky z obr. 5.20 (5.22) nahradit přímkami podle obr. 5.43.

U bipolárního tranzistoru se výrazně projevuje **vliv teploty na průběh charakteristik**. Můžeme jej tedy považovat za součástku řízenou teplotou.

Obr. 5.21. Model k posouzení teplotních změn pracovního

V principu se při zvýšení teploty zvýší kolektorový proud. Tato závislost je velmi nepříjemná, mění se poloha klidového pracovního bodu. V aplikacích je proto nutné zabývat se stabilizací pracovního bodu, zvláště pokud se okolní teplota bude měnit ve velkém rozsahu. To platí zejména pro germaniové tranzistory, u křemíkových tranzistorů je nutné spíše "stabilizovat" rozpětí parametrů.

Model z obr. 5.21 pak snadno použijeme k posouzení teplotních změn klidového pracovního bodu **P** tranzistoru. Praxe ukázala, že na teplotní změny **P** u tranzistorů germaniových má vždy dominantní vliv zdroj proudu $I_{CB0}(\vartheta)$, zatímco u křemíkových tranzistorů má obvykle (výjimkou je provoz v mikrorežimu) dominantní vliv zdroj napětí $\Delta u_{BE}(\vartheta)$. Pokud by tranzistor byl připojen na stálé pracovní předpětí U_{BEP} , mohli bychom snadno spočítat změny klidového proudu kolektoru I_{CP} s teplotou. Bude platit

$$i_{CP}(\vartheta) = I_{CP}(\vartheta_0) \cdot \exp[a \cdot \Delta u_E(\vartheta)]$$

takže můžeme určit např. pro

Změnu proudu o 5% při změně teploty o 1°C by bylo možné připustit, tranzistorová zařízení však podléhají větším teplotním změnám, takže teplotní posuv klidového proudu je nutné redukovat. K jeho zmenšení můžeme použít buď zápornou zpětnou vazbu nebo pomocný kompenzační prvek.

Teplota přechodů tranzistoru je ovlivňována teplotou okolí, množstvím elektrického výkonu, který se v tranzistoru mění na teplo a vydatností ochlazování tranzistoru. Při provozu tranzistoru jako aktivního prvku protéká emitorovým i kolektorovým přechodem přibližně stejný proud. Na závěrně polarizovaném kolektorovém přechodu je mnohem větší úbytek napětí než na propustně polarizovaném emitorovém přechodu, proto i vznik tepla na kolektorovém přechodu je daleko větší. K celkovému oteplení tranzistoru přispívá tedy rozhodující měrou vývin tepla na kolektoru. Jeho velikost udává tzv. *kolektorová ztráta*.

Pracovní oblast charakteristik tranzistoru, kterou lze skutečně využít, je omezena třemi faktory:

1. *maximální kolektorovou ztrátou* P_{Cmax} (je určena maximální dovolenou teplotou přechodu a chlazením tranzistoru);
2. *maximálním kolektorovým proudem* I_{Cmax} (je určen konstrukcí tranzistoru a dovoleným oteplením);
3. *maximálním kolektorovým napětím* U_{CEmax} (je omezeno průrazem kolektoru).

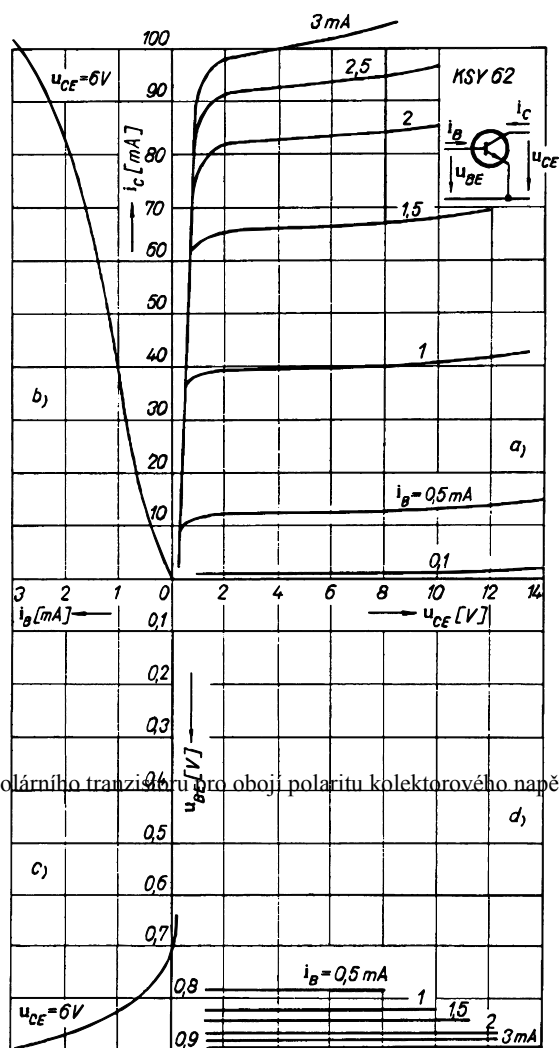
Omezení pracovní oblasti výstupních charakteristik tranzistoru vlivem těchto tří faktorů je znázorněno graficky na obr. 5.30., a to pro zapojení SE. Jsou zde ještě další dvě omezení, a to *mezní přímkou* a *nulovým proudem báze*.

V katalozích se oblast omezená třemi výše uvedenými faktory nazývá **SOA** [Safe Operating Area], proud bývá v logaritmickém měřítku (abychom nemuseli kreslit hyperbolu) a uvažuje se možnost "dovoleného přetížení" tranzistoru v impulsním režimu (viz obr. 5.35).

Obr. 5.24. Výstupní charakteristiky bipolárního tranzistoru pro obojí polaritu kolektorového napětí

Tranzistory často používáme též ve funkci omezovačů, spínačů nebo převodníků úrovní signálu. I zde používáme tranzistory především v zapojení se společným emitorem (SE) a zajímá nás jeho chování v okolí nuly výstupního napětí, jakým řídicím napětím můžeme manipulovat ve stavu zániku kolektorového proudu a jak se tranzistor chová při obrácení polarity napětí u_{CE} mezi kolektorem a emitorem. Chování tranzistorů při obrácené polaritě u_{CE} se výrazně liší při řízení proudem v bázi i_B od buzení napětím u_{BE} .

Na obr. 5.24 jsou znázorněny výstupní charakteristiky tranzistoru pro obojí polaritu u_{CE} při buzení proudem do báze (obr. 5.22b). Oblast charakteristik A je "normální", která je výrobcí dostatečně popisována. Oblast B je oblastí inverzní, kdy připojené napětí u_{CE} nutí otevírat kolektorový přechod víc než emitorový a směr difúze převážné části nosičů náboje v bázi se



Obr. 5.23. Soustava změřených stejnosměrných charakteristik (křemíkový tranzistor NPN, zapojení SE)

obrací. V této oblasti jsou výstupní charakteristiky mnohem hustší, proudový zesilovací činitel h_{21e} v linearizovatelné části se pohybuje mezi 0,3 a 5. Oblast B je ostře ohraničena částí C, která je dána Zenerovým průrazem emitorového přechodu.

Proud do báze lze zavádět jen v jedné polaritě, která vede k otevírání buď emitorového anebo kolektorového přechodu. Při obrácení směru i_B přejde emitorový přechod do Zenerovy

Obr. 5.25. Detail charakteristik v okolí počátku (oblast D)

oblasti, což je jev nežádoucí. Je proto prospěšné při obrácení polarity u_{BE} napětí mezi bází a emitorem omezit dostatečně pod úroveň průrazu.

Detail charakteristik (oblast D) je rozveden na obr. 5.25a. Z něho je zřejmé, že při nulovém proudu i_C není diferenciální odpor tranzistoru nejmenší, jak se na první pohled zdá z průběhu charakteristik na obr. 5.24 při zmenšování u_{CE} k nule. Kromě toho je zbytkové napětí na tranzistoru kolem 100 mV. Zbytkové napětí U_{CEO} (tj. napětí průsečíku charakteristiky s osou napětí) lze patrně zmenšovat zvětšováním proudu báze i_B . Další detail na obr. 5.25b ukazuje, že U_{CEO} lze zmenšit na určité minimum, dané pro určitý typ tranzistoru optimální hodnotou proudu báze. Při dalším zvětšování proudu báze zbytkové napětí opět stoupá. U spínacích tranzistorů bývá minimální zbytkové napětí mezi 20 a 50 mV. Takové tranzistory se ovšem nehodí pro funkci analogových spínačů napětí. K tomuto účelu lépe poslouží některé typy nízkofrekvenčních tranzistorů s velkým proudovým zesilovacím činitelem, jak ukazuje obr. 5.25c. U nich se v normálním zapojení dosahuje zbytkového napětí pod 10 mV. Dalšího výrazného snížení zbytkového napětí lze dosáhnout výběrem vhodného typu tranzistoru a jeho zapojením jako spínač v inverzním režimu (obr. 5.26).

Obr. 5.26. Inverzně zapojený tranzistor a jeho charakteristiky

Při optimálním proudu báze (např. pro KC 509 je $I_{Bopt} \sim 0,1$ mA) lze dosáhnout zbytkové napětí U_{ECO} pod 0,2 mV. Při použití uvedené úpravy je však třeba si uvědomit, že průrazné napětí rozepnutného spínače výrazně kleslo, jak je naznačeno na obr. 5.26c.

Tranzistorové spínací obvody bývají buzeny zdrojem chovajícím se jako zdroj proudu nebo zdrojem napětí v sérii s rezistorem, omezujícím maximální proud. Pokud je tímto buzením otevřen emitorový nebo kolektorový přechod, lze i takové buzení považovat za proudové. Jsou-li oba přechody zavřené, takže bází teče jen nepatrný zbytkový proud, je i úbytek napětí na sériovém rezistoru zanedbatelný a buzení se uplatní jako napěťové.

Obr. 5.27. Napěťové buzení bipolárního tranzistoru v normálním a inverzním režimu

Na obr. 5.27a je naznačeno, že při záporném pólování

u_{BE} i u_{CE} se snadno může otevřít kolektorový přechod. Při změně polarity napětí u_{CE} můžeme tranzistor udržet v uzavřeném stavu, jen pokud udržíme postačující předpětí, tedy u NPN tranzistorů $u_{BE} < u_{CE} - 1$ V. Jinak tranzistor přechází do inverzního režimu a otvírá se.

Zenerův průraz emitorového přechodu. Tento průraz ovlivní závěrné předpětí u_{BC} v inverzním zapojení především s ohledem na maximální přípustné u_{CE} , jak je naznačeno na obr. 5.27b.

Některé typy bipolárních tranzistorů nejsou vhodné pro použití v diodovém zapojení nebo jako spínače při proudech větších než 1 mA. Patří mezi ně velká část tzv. "stíněných" vf tranzistorů, např. naše KF 167, KF 173, KF 272 a některé tranzistory vysokonapěťové, např.

Obr. 5.28. Výstupní charakteristiky tranzistoru, který není vhodný pro spínací účely

Obr. 5.29. Způsob odstínění kapacity mezi kontaktem báze a kolektorem

KF 504 a další. Jejich výstupní charakteristiky jsou znázorněny na obr. 5.28.

Za účelem zmenšení kapacity mezi bází a kolektorem (zpětnovazební kapacita C_{12e}), která je převážně tvořena plochou expandovaného kontaktu báze, je použito principu jejího rozdělení mezi vstup a výstup tranzistoru pomocí **stínící diody**. Podstatou je nadifundování vodivosti typu P pod kontakt báze. Tak vzniká mezi základním materiálem typu N a vytvořenou vrstvou P, která se propojí s emitorem, dioda vykonávající funkci stínění spojeného se společným bodem.

5.3.3 Průrazy tranzistorové struktury

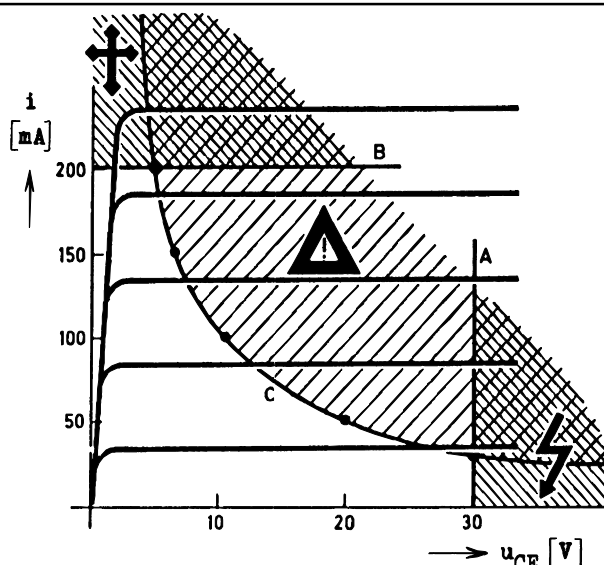
Při zatížení tranzistoru "vysokým" napětím dochází průchodem proudu oblastí prostorového náboje přechodu C-B k lavinové nárazové ionizaci. Dochází k průrazu tranzistoru, který označujeme jako tzv. **první průraz**. Podle obr. 5.33 budeme rozlišovat tato napětí:

- u_{CEU} napětí při záporném předpětí přechodu E-B,
- u_{CES} napětí při $R_{BE} = 0$,
- u_{CER} napětí při $R_{BE} \neq 0$,
- u_{CE0} napětí při $R_{BE} \rightarrow \infty$.

Průrazné napětí $U_{(BR)CEO}$ je vždy menší než $U_{(BR)CBO}$, které odpovídá průrazu izolovaného kolektorového přechodu. Jejich rozdíl se zvětšuje s rostoucím činitelem α (a tedy i rostoucím β). Z rozboru plyne, že snížíme-li vliv zbytkového proudu I_{CB0} , např. paralelním rezistorem mezi bází a emitorem R_{BE} , dojde ke zvýšení napětí $U_{(BR)CEO}$. K dalšímu zvýšení $U_{(BR)CEO}$ dojde zkratováním emitoru a báze (napětí u_{CES}) nebo přivedením záporného napětí mezi emitor a bází (napětí u_{CEU}). Závěrné charakteristiky za výše uvedených podmínek jsou znázorněny na obr. 5.33.

Jak již bylo uvedeno, při překročení maximálního ztrátového výkonu P_{max} (u tranzistorů P_{Cmax}) může vzrůst teplota součástky natolik, že dojde k **tepelnému průrazu**.

V případě tranzistorů však může dojít k destrukci součástky i při podstatně nižším ztrátovém výkonu než P_{Cmax} , k tzv. **druhému průrazu** [second breakdown]. Druhý průraz se projeví prudkým poklesem napětí u_{CE} , ztrátou schopnosti řízení proudem báze, což má za následek zničení součástky. Je důsledkem lokálního extrémního zvýšení hustoty kolektorového proudu a projevuje se obvykle lokálním protavením tranzistorové struktury. Druhý průraz může vzniknout dvěma způsoby, rozlišujeme **tepelný** a **proudový průraz**.



Obr. 5.30. Omezení pracovní oblasti tranzistoru v zapojení SE s vyznačením maximálních veličin ($U_{CEmax} = 30$ V, $I_{Cmax} = 200$ mA a $P_{Cmax} = 1$ W); v provozu nesmí pracovní bod překročit čáry A a B, hyperbolu maximální kolektorové ztráty smí překročit jen ve spínacím režimu

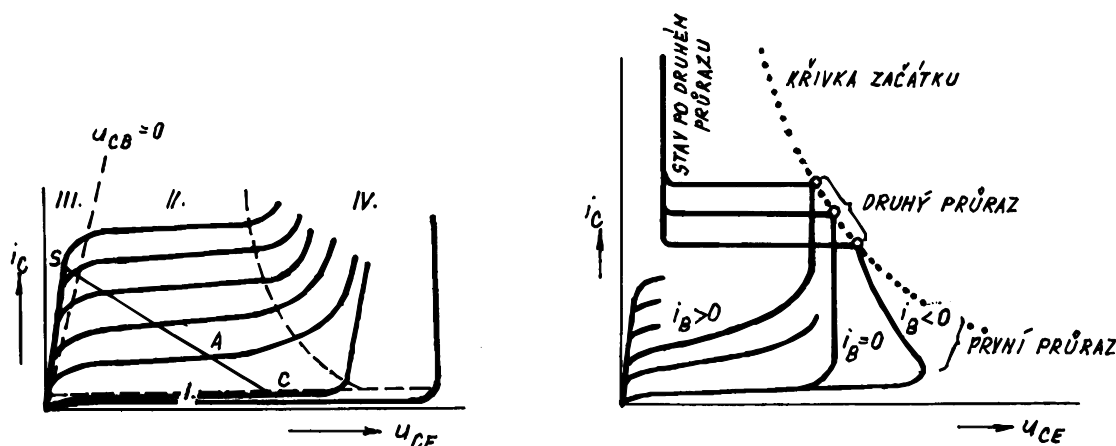
Obr. 5.31. Napěťová závislost zbytkových proudů a první

Obr. 5.32. Model tranzistoru s vyznačenými podmínkami ex jednotlivých zbytkových proudů

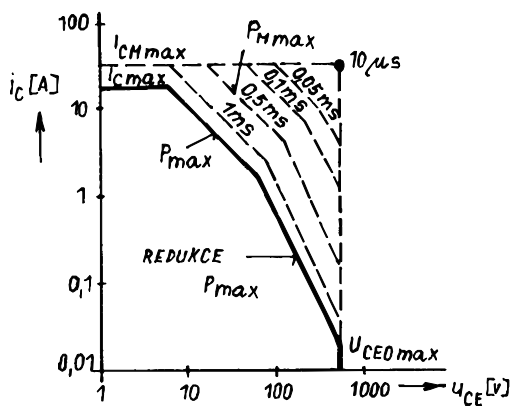
Obr. 5.33. Schématické znázornění prvního průrazu uzavřeného tranzistoru a definice jednotlivých průrazných napětí

Tepelný druhý průraz je vlastně lokální tepelný průraz. Může k němu dojít, pokud z nějakého důvodu (např. vzhledem k nehomogenitě struktury) je v některém místě vyšší proudová hustota. Pokud je tranzistor v aktivní pracovní oblasti, zvýšení teploty vede ke zvýšení proudové hustoty. Vzniká tak horké místo, do kterého se koncentruje proud. Kladná zpětná vazba vede ke vzniku tepelného průrazu s následným zničením součástky. Jiný způsob vzniku lokálního tepelného průrazu, který může být spojen i s impulsním režimem, souvisí s nerovnoměrným rozložením proudové hustoty pod emitorem.

V tranzistorové struktuře je nejvyšší hustota proudu u okraje emitorové oblasti. Teplo v této oblasti je však odváděno do okolních oblastí křemíkové destičky. Může přitom vzniknout situace, kdy je teplota uprostřed emitorové oblasti vyšší, než na okrajích. Pokud teplota překročí určitou kritickou hodnotu, tepelně generovaný proud nerovnovážných nosičů v kolektorové oblasti může nahradit proud báze a vyvolat prudké zvýšení hustoty proudu



Obr. 5.34. Režimy bipolárního tranzistoru (vlevo), první a druhý průraz (vpravo)



Obr. 5.35. Dovolená pracovní oblast tranzistoru (SOA)

uprostřed emitorové oblasti. Tomuto mechanismu vzniku druhého průrazu čelí výrobci buď zmenšením tloušťky emitoru (zvětšení šířky báze) uprostřed emitorové oblasti nebo zlepšením podmínek pro odvod tepla. U výkonových tranzistorů se používá oboustranný odvod ztrátového tepla. Příklad mezních provozních podmínek, vedoucích ke vzniku druhého průrazu je znázorněn na obr. 5.35 a vymezuje tzv. *bezpečnou pracovní oblast* (safe operating area). S ro-stoucím kolektorovým napětím roste možnost vzniku druhého průrazu, takže závislost je podstatně strmější, než odpovídá podmínce konstantního mezního ztrátového výkonu.

Proudový druhý průraz vzniká při vypínání tranzistoru záporným impulsem proudu báze v obvodu s indukční zátěží. Jde o mechanismus lavinové ionizace. Při vypínání tranzistoru kolektorové napětí u_{CE} rychle roste při velkém proudu i_C . Dochází k nárazové ionizaci (přestože $u_{CE} < U_{(BR)CEO}$). Díry vzniklé při nárazové ionizaci jsou urychleny směrem k emitoru, působí jako velký lokální proud báze a vyvolají lokální prudké zvýšení proudové hustoty (plasma). Výsledkem je velmi rychlá destrukce tranzistoru v příslušné oblasti.

Proudový druhý průraz způsobuje velmi vážné omezení činnosti tranzistorů ve spínacím režimu, neboť v tomto případě pracují téměř vždy s induktivní zátěží. Proto je třeba dodržovat tzv. bezpečnou pracovní oblast (trajektorie pracovního bodu při vypínání nesmí protnout hranici vzniku druhého průrazu). Rozšíření bezpečné pracovní oblasti je možné dosáhnout zpomalením nárůstu kolektorového napětí v průběhu vypínacího procesu odlehčovacím obvodem sestávajícím z diody, kondenzátoru a rezistoru (paralelně připojeným k tranzistoru).

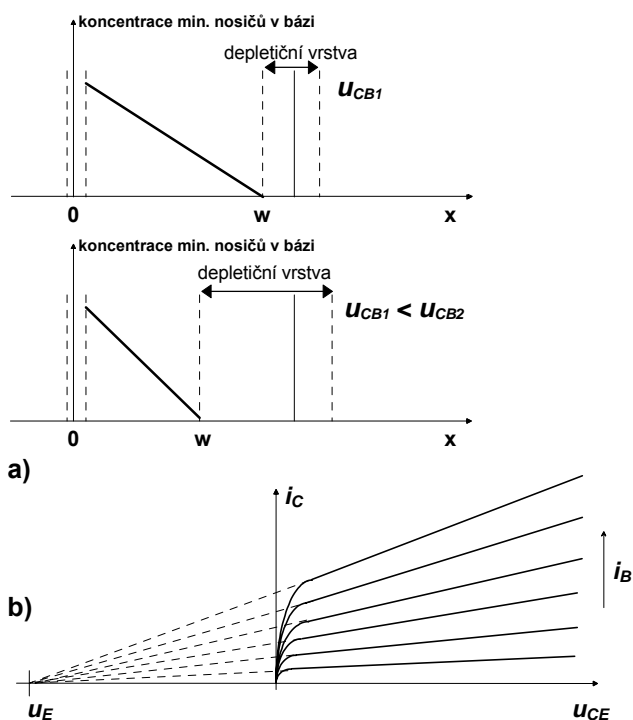
5.3.4 Earlyho jev

Dosud jsme předpokládali, že efektivní šířka báze w nezávisí na velikosti napětí přivedených na emitorový a kolektorový přechod. Tento předpoklad však není vždycky správný. Jestliže je koncentrace příměsí v bázi relativně nízká (srovnatelná s koncentrací příměsí v kolektoru), bude se depletiční oblast závěrně polarizovaného kolektorového přechodu rozšiřovat i do báze a zmenšovat tak její efektivní šířku. Při zvýšení závěrného napětí kolektorového přechodu dojde proto ke zmenšení efektivní šířky báze a tím i ke vzrůstu gradientu koncentrace minoritních nosičů vstříknutých do báze z emitoru (obr. 5.36a). To vede ke vzrůstu kolektorového proudu i_C , který se tak stává závislým na napětí u_{CB} .

Tento jev, nazývaný podle svého objevitele *Earlyho jev*, ovlivňuje tvar charakteristik tranzistoru. Na obr. 5.36b) vidíme jeho vliv na výstupní charakteristiky tranzistoru v zapojení SE. Se zvyšujícím se napětím u_{CE} (a zároveň zvyšujícím se závěrným napětím u_{CB}) vzrůstá proud i_C . Proložíme-li charakteristikami přímky, protnou se při napětí U_E , které nazýváme *Earlyho napětím*.

V praxi Earlyho jev způsobí, že proudový zesilovací činitel β je také napěťově závislý a pro daný pracovní bod musíme použít jeho upravenou hodnotu

$$\beta_{ef} = \beta \left(1 + \frac{u_{CE}}{U_E} \right) \quad (5.24)$$



Obr. 5.36.: a) Vliv napětí u_{CB} na šířku báze a gradient koncentrace minoritních nosičů v bázi; b) vliv Earlyho jevu na výstupní charakteristiky tranzistoru v zapojení SE

5.4 Modely bipolárního tranzistoru

5.4.1 Nelineární modely bipolárního tranzistoru

V praxi (při rozboru činnosti elektronických obvodů a při jejich návrhu) se běžně používají modely, které jsou označovány jako *Ebersovy-Molloyovy modely*, protože první model tohoto

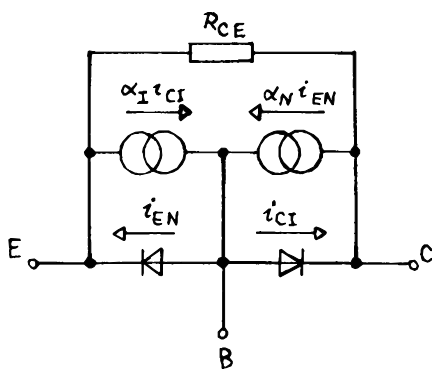
typu včetně fyzikální interpretace jeho parametrů (popsaný v kap. 5.2.2.) navrhli J.J. Ebers a J.L. Moll v roce 1954. V důsledku malé přesnosti byl původní model od té doby několikrát modifikován (takže původní model má dnes již jen historický význam).

Teplotní závislosti parametrů tohoto modelu tranzistoru odpovídají teplotním závislostem parametrů modelu diody.

Základní EM model lze v případě nutnosti přesnějšího modelování dále rozšiřovat. První takové rozšíření spočívá v připojení sériových rezistorů do přívodů jednotlivých elektrod, což je znázorněno na obr. 5.37. Chceme-li modelovat Earlyho jev, přidáme rezistor mezi kolektor a emitor (obr. 5.38). Velikost tohoto odporu určíme ze sklonu charakteristik jako

$$R_{CE} = \frac{U_E + u_{CE}}{i_C} = \frac{U_E + u_{CE}}{I_{CB0} + \alpha_N i_E} \quad (5.25)$$

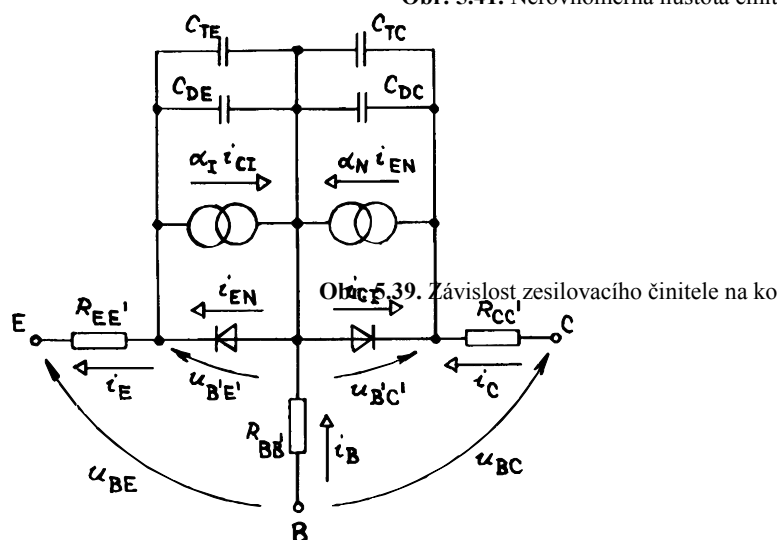
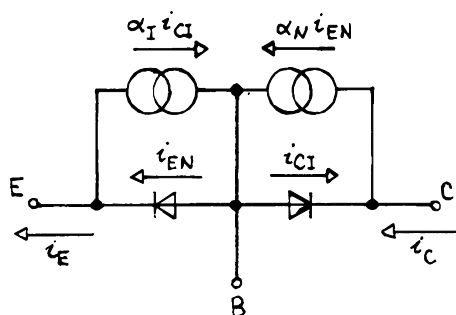
Pro praktické výpočty s EM modelem je výhodné zesilovací činitele α_N a α_I nahradit činiteli β_N a β_I podle známých vztahů.



Obr. 5.38.: Modelování Earlyho jevu rezistorem R_{CE}

Významné zvýšení přesnosti EM modelu docílíme zavedením závislosti proudového zesilovacího činitele β_N na poloze pracovního bodu, konkrétně na proudu kolektoru nebo emitoru. Užívá se aproximace polynomem třetího stupně apod. Poznamenejme ještě, že zavedení proudové závislosti parametru β_N přináší určitou komplikaci pro počítačovou analýzu, neboť na začátku výpočtu požadujeme znalost pracovního bodu, který je až výsledkem výpočtu (řeší se to vhodným uživatelským podprogramem). Důvod je v tom, že s hustotou nosičů náboje v bázi se poněkud mění rekombinační schopnost báze.

Obvykle lze zanedbat vlivy odporů přívodu emitoru R_E a přívodu kolektoru R_C . U diskretních tranzistorů je zanedbatelnost R_C dána konstrukcí - oblast kolektoru je (pro optimální odvod tepla) po celé ploše polovodičové destičky připájena ke kovovému pouzdru, případně k masivnímu přípojnému vodiči. Naproti tomu oblast emitoru je k vývodu pouzdra propojena jen několik desítek mikrometrů silným drátkem (obvykle zlatým), jehož odpor bývá několik ohmů. Tento drátek často funguje při proudovém přetížení tranzistoru jako tavná pojistka - při jeho přetavení je tranzistor neodvratitelně znehodnocen.



Obr. 5.37.: Statický (vlevo) a dynamický (vpravo) Ebersův-Mollův model bipolárního tranzistoru včetně sériových rezistorů

Odpor r_{BB}' představuje nejen odpor přívodního drátku, ale především jakýsi integrální odpor, vytvořený bázovým přívodem nosičů náboje do centra báze. Ve skutečnosti by podle konstrukce přívodu báze měl být tento odpor nahrazen obvodem s rozloženými parametry (viz obr. 5.40a). Velikost r_{BB}' je nepřímo úměrná tloušťce báze a je ovlivněna uspořádáním kontaktních plošek (tj. výrobní masky kovových vývodů báze a emitoru).

Nejjednodušší uspořádání vývodů podle obr. 5.40b mají tranzistory starších typů a běžné tranzistory integrované.

U tranzistorů spínacích a vysokofrekvenčních je žádoucí co nejmenší r_{BB}' . Dosahuje asi 10 až 30 Ω , a to použitím tlusté báze (proudový zesilovací činitel $h_{21e} \approx 10$ až 50) a hřebínkovým (tzv. interdigitálním) uspořádáním vývodů podle obr. 5.40c. Naproti tomu tranzistory nízkofrekvenční mají bázi velmi tenkou a dosahují h_{21} v rozmezí 100 až 500. Jejich odpor báze (u konstrukcí pro nejmenší výkony) se proto pohybuje mezi 300 a 1000 ohmy. Protože ostré zlomy ve struktuře kontaktních plošek přívodů vedou k nerovnoměrnostem toku nosičů

Obr. 5.40. Odpor báze u různých konstrukcí bipolárního tranzistoru je závislý na konstrukčním provedení náboje a tím i ke zvětšení šumu, vyrábí se velká část nízkofrekvenčních tranzistorů s kruhovou strukturou přívodů podle obr. 5.40d. Toto nízkošumové provedení umožňuje snížit úroveň šumového pozadí až o 20 dB.

EM model ovšem nevystihuje všechny jevy v bipolárních tranzistorech a neumožňuje odhadnout mezní stavy v pracovní činnosti tranzistorů v obvodech. Mezní stavy v činnosti tranzistorů obvykle dimenzujeme podle mezních údajů výrobců s postačující rezervou. Zpravidla využíváme mezní údaje nejvýše z 50 %.

Zjednodušené EM modely

V mnoha obvodech pracuje tranzistor v (normální) aktivní oblasti. V těchto případech je možno EM model dále zjednodušit. Dioda báze-kolektor je pólována v závěrném směru, tj. $u_{BC} < 0$, potom

$$i_{CI} \approx -I_{CS} \quad (5.26)$$

Dioda báze-emitor je pólována v propustném směru. Jestliže $i_{EN} \approx I_{ES} \exp\left(\frac{u_{BE}}{M_E U_T}\right)$

(5.27)

kde M_E je obdoba konstanty zavedené v rovnici (3.2), pak pro $u_{BE} > M_E U_T$ platí

$$i_{EN} > \alpha_I I_{CS} \quad (5.28)$$

protože u kvalitních tranzistorů je $\alpha_I < 1$.

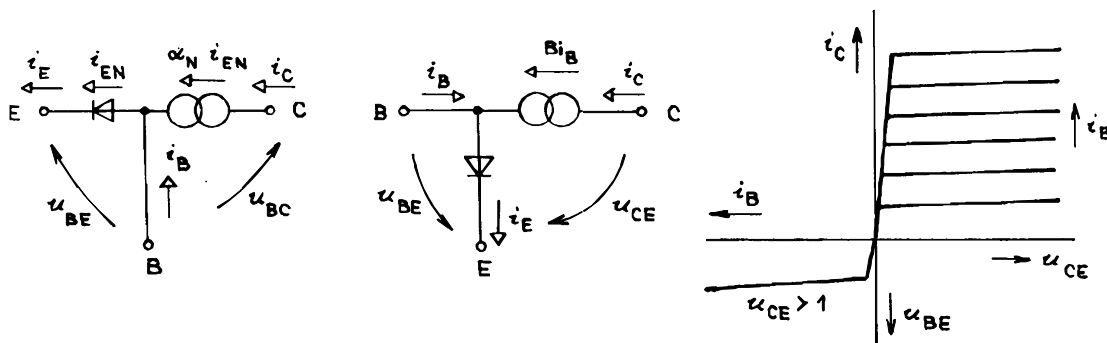
Zjednodušený model pro aktivní oblast je zobrazen na obr. 5.42 a je popsán vztahy

$$i_C = \alpha_N i_{EN}, \quad i_E = -i_{EN}, \quad i_B = (1 - \alpha_N) i_{EN} \quad (5.29)$$

kde i_{EN} je dáno rovnicí (5.27). V tomto modelu samozřejmě neuvažujeme sériové rezistory v přívodech, proto $u_{B'E'} = u_{BE}$.

Na obr. 5.42 vidíme variantu modelu, ve kterém jsme zavedli stejnosměrný proudový zesilovací činitel $B = i_C/i_B = \alpha_N/(1 - \alpha_N)$. Tento model platí za předpokladu, že $u_{CE} > u_{BE}$.

Rovnice (5.29) jsou základní a vedle Ohmova zákona prakticky jediný matematický vztah, který je nutné znát pro porozumění funkce bipolárních tranzistorů. Při malých proudech, tj.

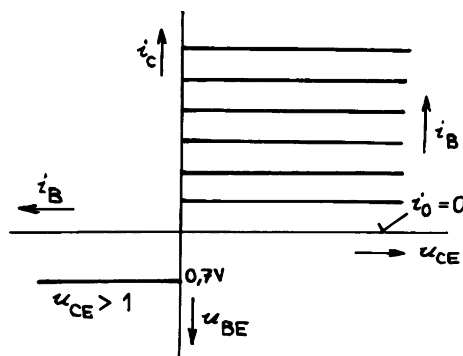
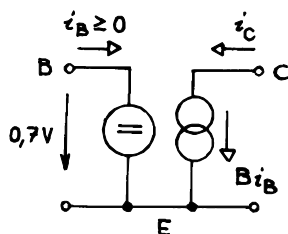


Obr. 5.42. Dvě varianty EM modelu pro aktivní oblast a aproximace vstupní a výstupních charakteristik

jednotky až tisíce mikroampérů, platí prakticky úplně přesně. Uvádí se v praktickém tvaru $i_C = I_S \exp(u_{BE}/U_T)$, případně $u_{BE} = U_T \ln(i_C/I_S)$.

Při použití vztahu (5.29) narazíme na jeden problém. Vypočteme, jak závisí napětí u_{BE} na procházejícím proudu.

Uvažujme $I_S = 10^{-16} A$ a necht' $i_{C1} = 1 \mu A$, potom odpovídající $u_{BE1} = U_T \ln(10^{-6}/10^{-16}) = 599 mV$. Při proudu $i_{C2} = 500 \mu A$ je $u_{BE} = 760 mV$.



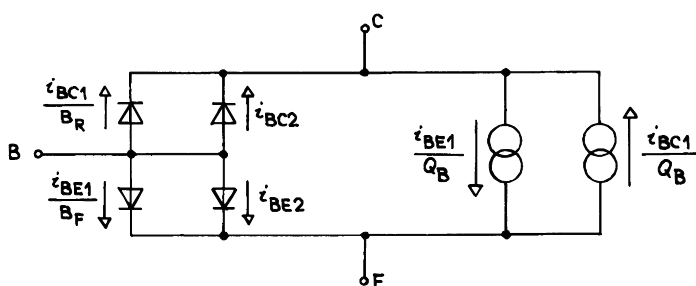
Obr. 5.43. Nejjednodušší EM model pro aktivní oblast a tvar jeho vstupní a výstupních charakteristik

Vidíme, že při velmi podstatném zvětšení proudu i_C (500krát) se napětí u_{BE} změní jen málo. Pokud víme, že hodnota proudu i_C je v rozsahu jednotek μA až jednotek mA , tak řekneme, že u_{BE} je asi 0,7 V. Chyba, které se

tímto způsobem dopustíme je malá a tato přesnost výpočtu nám pro základní orientaci stačí. Takto dojdeme ke snad nejvíce zjednodušenému EM modelu, který je zobrazen na obr. 5.43.

Pro úvahy v obvodové technice je velmi užitečné znát kvalitativní i kvantitativní souvislosti vztahů

$$i_C = I_S \exp(u_{BE}/U_T) \quad \text{a} \quad i_C = \beta i_B$$



Obr. 5.44.: Obvodová struktura Gummelova-

Platí: a) Teče-li do báze tranzistoru proud i_B , vyvolá to v kolektorovém obvodu proud $i_C = \beta i_B$ a ten vyvolá

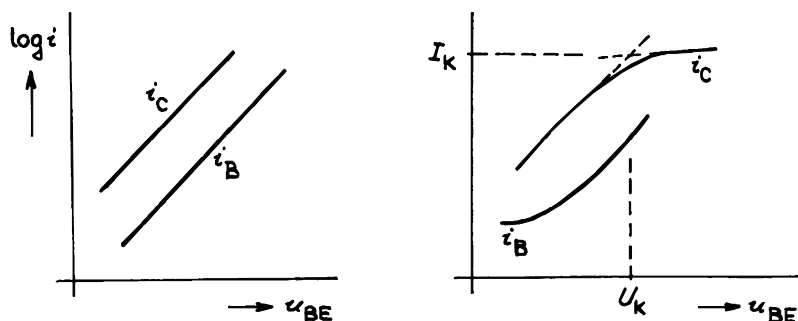
úbytek napětí $u_{BE} = U_T \ln(i_C/I_S)$.

b) Je-li mezi bází a emitorem napětí u_{BE} , protéká kolektorem proud $i_C = I_S \exp(u_{BE}/U_T)$ a bází protéká proud $i_B = i_C/\beta$.

c) Připojíme-li k emitoru proudový zdroj $i_E \approx i_C$ (a tranzistor má dostatečně velkou hodnotu činitele β), pak je $u_{BE} = U_T \ln(i_C/I_S)$ a $i_B = i_C/\beta$.

Gummelův-Poonův model

Gummelův-Poonův model bipolárního tranzistoru je modelem nábojovým. Je odvozen bez zavedení aproximace na nízkou úroveň injekčních proudů. Tento model je velmi přesný, bere v úvahu mnoho fyzikálních jevů, ale nevýhodou je značný počet parametrů (až 25), které je potřeba znát při jeho použití. Gummelův-Poonův model (dále GP model) existuje v několika verzích. Pro praxi je důležité, že je součástí programu SPICE, což je světově nejrozšířenější program pro obvodovou analýzu elektrických obvodů.



Obr. 5.45. Vstupní a převodní charakteristiky pro Ebersův-Mollův (vlevo) a Gummelův-Poonův model (vpravo)

modelu je saturační proud závislý na okamžité hodnotě tzv. *normovaného náboje báze* (tedy také na okamžité hloubce báze). Tímto mechanismem je možno věrně modelovat Earlyho a další jevy. Pomocí normovaného náboje báze je ovládána také velikost průletových dob minoritních nosičů bází a u některých verzí i hodnota bázevého odporu.

Velmi názorně lze demonstrovat rozdíl mezi oběma typy modelů na charakteristikách $i_B = f(u_{BE})$ a $i_C = f(u_{BE})$ při $u_{CB} = 0$. Z obr. 5.45 je zřejmé, že GP model je schopen velmi přesně vystihnout průběhy proudových zesilovacích činitelů i obou charakteristik. EM model to neumožňuje ani po dodatečném zavedení proudové závislosti zesilovacího činitele, dojde totiž k významnému zkreslení vstupních charakteristik.

Svorkové proudy GP modelu jsou popsány dvěma základními rovnicemi

$$i_C = \frac{i_{BE1}}{Q_B} - \frac{i_{BC1}}{Q_B} - \frac{i_{BC1}}{B_R} - i_{BC2} \quad (5.30)$$

$$i_B = \frac{i_{BE1}}{B_F} - i_{BE2} + \frac{i_{BC1}}{B_R} + i_{BC2} \quad (5.31)$$

kde

$$\begin{aligned} i_{BE1} &= I_S \left(\exp \frac{u_{BE}}{N_F U_T} - 1 \right) & i_{BE2} &= I_{SE} \left(\exp \frac{u_{BE}}{N_E U_T} - 1 \right) \\ i_{BC1} &= I_S \left(\exp \frac{u_{BC}}{N_R U_T} - 1 \right) & i_{BC2} &= I_{SC} \left(\exp \frac{u_{BC}}{N_C U_T} - 1 \right) \end{aligned}$$

kde $I_{SE} = C_2 I_S$ a $I_{SC} = C_4 I_S$; C_2 a C_4 jsou násobící konstanty. Proud i_{BE1} je proud ideálního přechodu B-E, i_{BC1} je proud ideálního přechodu B-C, proudy i_{BE2} a i_{BC2} jsou svodové (rekombinační) složky. Rozdíl $(i_{BE1}/Q_B - i_{BC2}/Q_B)$ představuje dominantní složku kolektorového proudu. Veličina Q_B je dána podílem náboje majoritních nosičů v bázi při daném a nulovém předpětí. Platí

$$Q_B = \frac{Q_1}{2} (1 + \sqrt{1 + 4Q_2}) \quad (5.32)$$

$$Q_1 = \frac{1}{1 - \frac{u_{BC}}{U_{AF}} - \frac{u_{BE}}{U_{AR}}} \quad (5.33)$$

$$Q_2 = \frac{i_{BE1}}{I_{KF}} + \frac{i_{BC1}}{I_{KR}} \quad (5.34)$$

I_{KF} a I_{KR} jsou proudy v "koleně" (z angl. "knee") charakteristiky $i_C = f(u_{BE})$ v normálním a inverzním režimu - viz I_K na obr. 5.45. U_{AF} a U_{AR} je obdobně Earlyho napětí v normálním a inverzním režimu.

GP model je, podobně jako jiné modely, v počítačových programech implementován tak, že ty parametry, které v popisu nejsou specifikovány, nabývají svých implicitních hodnot. Implementace GP modelu je řešena tak, že nejsou-li zadány parametry definující normovaný náboj báze Q_B , GP model automaticky degeneruje na EM model.

5.4.2 Linearizované modely bipolárního tranzistoru (bipolární tranzistor jako impedanční čtyřpól)

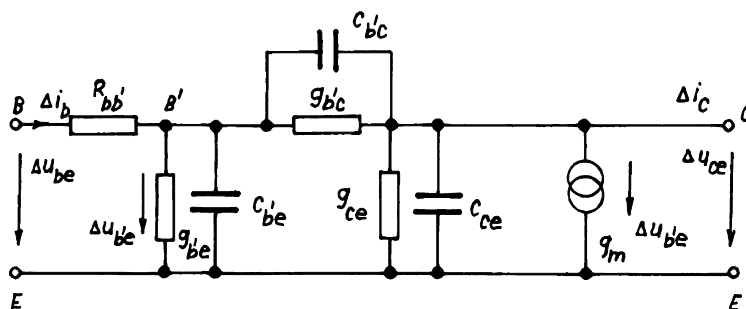
Při použití tranzistoru pro zpracování střídavých signálů o malé amplitudě lze tento prvek *linearizovat*, tj. využít k jeho popisu *linearizované modely* - *lineární čtyřpóly* (dvojbrany). Vztahy mezi "malými" střídavými proudy a napětími, které jsou superponovány na stejnosměrná napětí a proudy (určujícími polohu stejnosměrného pracovního bodu v charakteristikách) mohou být popsány dvěma způsoby:

1. pomocí linearizovaného EM modelu;
2. pomocí diferenciálních čtyřpólových parametrů.

Do linearizovaných modelů lze též zahrnout setrvačné vlastnosti tranzistorů, které se projeví při vyšších kmitočtech zpracovávaných signálů. Každý linearizovaný model (náhradní zapojení) je navržen pro určité pracovní podmínky a modeluje funkci tranzistoru jen přibližně. Z tohoto důvodu se používá celá řada linearizovaných modelů.

Linearizované modely vycházející z EM modelu

Uvažujme EM model podle obr. 5.18a. Při zpracování malých signálů můžeme diody nahradit jejich diferenciálním odporem v daném pracovním bodě. Předpokládáme-li, že pracovní bod (ve kterém $i_{EN} = I_{EN}$) leží v aktivní oblasti, můžeme vypustit kolektorovou diodu (je polarizována v závěrném směru) a zdroj proudu paralelně k emitorové diodě. Tak dojdeme k modelu na obr. 5.46 (tzv. *T-model*). Derivací i_{EN} podle u_{BE} lze dojít ke vztahu



Obr. 5.48. Giaciolettovo náhradní schéma bipolárního tranzistoru

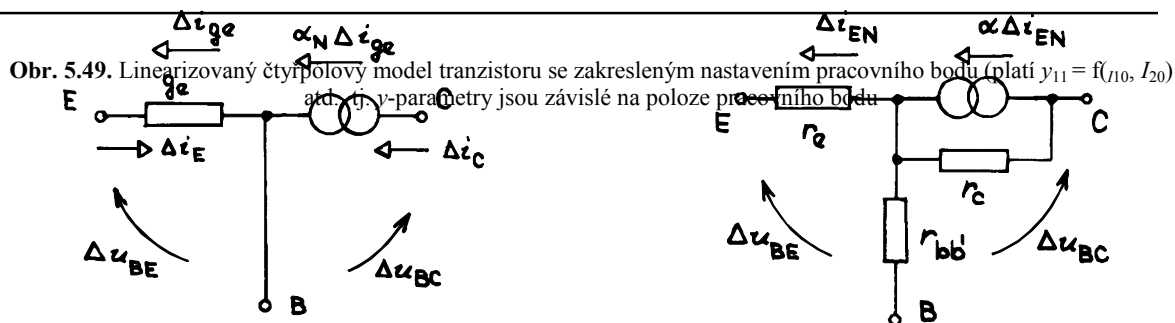
$$\Delta i_{EN} = I_{EN} \frac{\Delta u_{BE}}{M_E U_T} \quad (5.35)$$

Na obr. 5.47 je varianta T-modelu doplněná rezistory $r_{bb'}$ a r_c . Vodivost $g_e = 1/r_e$ má význam strmosti v daném pracovním bodě

$$g_e = \frac{\Delta i_{EN}}{\Delta u_{BE}} = \frac{I_{EN}}{M_E U_T} \quad (5.36)$$

Pokud k rezistorům r_e a r_c připojíme paralelní kapacitory, může tento T-model vystihnout i vysokofrekvenční chování tranzistoru.

Obdobně můžeme vytvořit i *π-model* bipolárního tranzistoru. Vysokofrekvenční verze tohoto modelu se nazývá *Giaciolettovo náhradní schéma* (obr. 5.48.). V praxi se s ním setkáváme při výpočtu vysokofrekvenčních zesilovačů.



Obr. 5.46. Linearizovaný EM model tvaru T

Obr. 5.47. Model tvaru T doplněný rezistory $r_{bb'}$ a r_c

Linearizované modely vycházející ze čtyřpólových parametrů

Druhou možností je využít klasické teorie čtyřpólů a popsat tranzistor pomocí *hybridních*, *admitančních* a někdy i *rozptylových rovnic*. Z teorie čtyřpólů víme, že pro **admitanční parametry** platí

$$I_1 = y_{11} U_1 + y_{12} U_2 \quad (5.37a)$$

$$I_2 = y_{21} U_1 + y_{22} U_2 \quad (5.37b)$$

a pro **hybridní** (smíšené) **parametry** platí

$$U_1 = h_{11} I_1 + h_{12} U_2 \quad (5.38a)$$

$$I_2 = h_{21} I_1 + h_{22} U_2 \quad (5.38b)$$

Pro kmitočty v mikrovlnném pásmu je výhodnější použít soustavy **rozptylových parametrů**, neboť to lépe odpovídá obvodové technice a měřicím metodám v tomto pásmu:

$$b_1 = s_{11} a_1 + s_{12} a_2 \quad (5.39a)$$

$$b_2 = s_{21} a_1 + s_{22} a_2 \quad (5.39b)$$

kde a_1, b_1 je dopadající a odražená vlna na vstupních svorkách; a_2, b_2 na výstupních svorkách.

V rovnicích (5.37) a (5.38) jsme změny napětí Δu a Δi označili velkými písmeny U a I , které představují amplitudy malého střídavého (harmonického) signálu. K těmto rovnicím můžeme také nakreslit jejich obvodové vyjádření (náhradní zapojení) pomocí impedancí a řízených zdrojů (obr. 5.49 a 5.50). Admitanční parametry (y -parametry) obvykle měříme při vyšších kmitočtech, proto jednotlivé parametry jsou komplexní a U a I jsou komplexní amplitudy. Hybridní parametry (h -parametry) se běžně užívají pro nízké kmitočty, uvažujeme je jako skalární veličiny.

Poznamenejme, že k těmto rovnicím bychom došli pomocí Taylorova rozvoje daných funkcí dvou proměnných v pracovním bodě P (tj. aproximací tečnou rovinou). Parametry y_{ij} a h_{ij} nazýváme *diferenciální parametry čtyřpólu*, mají význam derivací. Určujeme je graficky z charakteristik nebo přímo měřením. Jsou funkcí pracovního bodu, kmitočtu a příp. teploty.

Jednotlivé h -parametry jsou definovány takto:

$$\text{vstupní impedance při výstupu nakrátko:} \quad h_{11} = \left. \frac{U_1}{I_1} \right|_{U_2 = \text{konst.}} \quad (5.40a)$$

$$\text{zpětný napěťový činitel při vstupu naprázdno:} \quad h_{12} = \left. \frac{U_1}{U_2} \right|_{I_1 = \text{konst.}} \quad (5.40b)$$

proudový zesilovací činitel při výstupu nakrátko:
$$h_{21} = \left. \frac{I_2}{I_1} \right|_{U_2 = \text{konst.}} \quad (5.40c)$$

výstupní admitance při vstupu naprázdno:
$$h_{22} = \left. \frac{I_2}{U_2} \right|_{I_1 = \text{konst.}} \quad (5.40d)$$

Z podmínky $U_2 = \text{konst.}$, $I_1 = \text{konst.}$,... je zřejmý význam podmínky "výstup nakrátko", "vstup naprázdno",...

Vrátíme-li se ještě k *T-modelu* a rovnici (5.36) pro vodivost g_e , lze pro dostatečně velký činitel β ($\beta \geq 40$, tj. $\alpha_N \geq 0,98$) odvodit přibližně vztahy

$$y_{21e} \approx g_e \quad y_{11e} \approx \frac{g_e}{\beta} \quad h_{21e} \approx \beta \quad (5.41)$$

Bude-li $M_E \approx 1$ a $\alpha_N \rightarrow 1$, potom lze napsat důležitý přibližný vztah

$$(y_{21e} \approx g_e) \approx 40 I_{C0} \quad , \quad \text{resp.} \quad r_e \approx \frac{25 \text{ mV}}{I_{C0}} = \frac{U_T}{I_{C0}} \quad (5.42)$$

kde I_{C0} je klidový kolektorový proud tranzistoru.

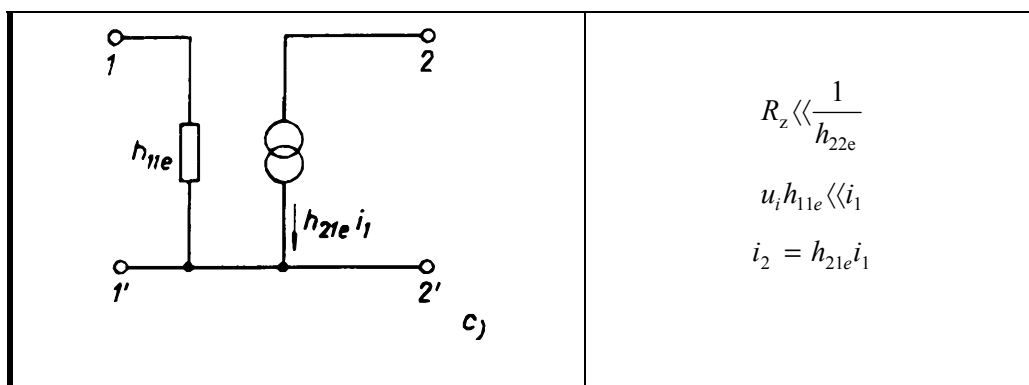
Poznámka: Parametr h_{11} čteme "h jedna - jedna" atd. Index ještě doplňujeme písmeny B, E, C, b, e nebo c podle toho, jedná-li se o zapojení SB, SE nebo SC a stejnosměrný nebo střídavý parametr.

Poznámka: Z hlediska vysokofrekvenčních aplikací je třeba uvést, že Giacolettův náhradní obvod má oproti

y -parametrům tu výhodu, že je "širokopásmový" (platí v daném pracovním bodě v širokém rozsahu kmitočtů), kdežto y -parametry jsou "úzkopásmové".

náhradní schéma	podmínky zjednodušení čtyřpólové rovnice
<p>$i_1 = 4,59 \mu\text{A}$ $i_2 = 137,6 \mu\text{A}$ $i_2^* = 135,1 \mu\text{A}$ $u_1 = 10 \text{ mV}$ $u_2 = -97 \mu\text{V}$ $h_{11e} = 2200 \Omega$ $h_{21e} = 43,6 \text{ k}\Omega$ $h_{22e} = 1/43,6 \text{ k}\Omega$ $R_2 = 796 \Omega$ $a)$</p>	<p>platí obecně</p> $u_1 = h_{11e}i_1 + h_{12e}u_2$ $i_2 = h_{21e}i_1 + h_{22e}u_2$
<p>$b)$</p>	$h_{12e}u_2 \ll (u_1, h_{11e}i_1)$ $u_1 = h_{11e}i_1$ $i_2 = h_{21e}i_1 + h_{22e}u_2$

Obr. 5.52. Napěťová převodní charakteristika tranzistoru NPN, její 1
 y_{21e} (strmost S)



5.4.3 Mezní kmitočty bipolárního tranzistoru

U linearizovaného tranzistoru můžeme definovat tzv. *mezní kmitočty*. Se vzrůstajícím kmitočtem klesá schopnost tranzistoru zesilovat. Způsobuje to konečná doba průchodu nosičů proudu oblastí báze. Se vzrůstajícím kmitočtem se také proud kolektoru zpožďuje za proudem emitoru. Situaci názorně ukazuje obr. 5.55, kde jsou definovány jednotlivé mezní kmitočty. Z něj je patrné, že až do určitých kmitočtů jsou proudové zesilovací činitele kmitočtově nezávislé a rovny $h_{21b} = \alpha_0 = \text{konst.}$ a $h_{21e} = \beta_0 = \text{konst.}$ Dále se pak jejich velikost s rostoucím kmitočtem zmenšuje.

Z obr. 5.55. je zřejmý význam mezních kmitočtů f_α a f_β pro zapojení tranzistoru SB a SE: jsou definovány poklesem na hodnotu $1/\sqrt{2}$ (tj. o 3 dB). Platí

$$f_\alpha = (1 + \beta_0) f_\beta, \quad f_\beta = (1 - \alpha_0) f_\alpha \quad (5.43)$$

Obr. 5.54. Odvození proudového zesilovacího činitele z výstupní

Obr. 5.51. Zjednodušené hybridní náhradní zapojení

Obr. 5.53. Odvození proudového zesilovacího činitele z převodní charakteristiky. Je patrný rozdíl mezi statickým činitelem B (h_{21E}) a dynamickým činitelem β (h_{21e})

Protože kmitočtový průběh modulu a fáze zesilovacího činitele $\alpha(f)$ a $\beta(f)$ je velmi podobný průběhu amplitudové a fázové charakteristiky obvodu RC , zavádíme tzv. *Rc aproximaci* vztahem

$$\alpha(f) = \frac{\alpha_0}{1 + j \frac{f}{f_\alpha}} \quad (5.44a)$$

a podobně pro $\beta(f)$

$$\beta(f) \frac{\beta_0}{1 + j \frac{f}{f_\beta}} = \frac{g_m r_{be}}{1 + j \omega C_{be} r_{be}} \quad (5.44b)$$

kde jsme využili zjednodušeného Giaciolettova modelu tranzistoru, kde platí

$$i_C = \frac{g_m r_{be}}{1 + j \omega C_{be} r_{be}}$$

tj. paralelní kombinace r_{be} C_{be} tvoří kmitočtově závislý dělič - s rostoucím kmitočtem se zmenšuje napětí na vnitřní bázi. Potom

$$f_\beta = \frac{1}{2\pi C_{be} r_{be}}$$

Poznamenejme ještě, že kapacita C_{be} obsahuje jen difúzní složku.

Dále se používají tyto mezní kmitočty:

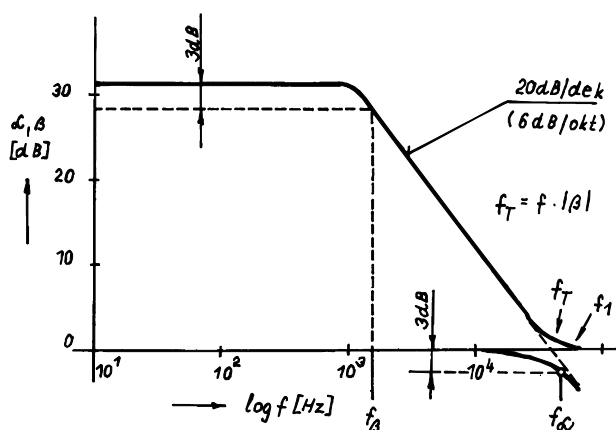
Mezní kmitočet f_T (tzv. *tranzitní kmitočet*) je definován jako $f_T = f \cdot |\beta|$ v klesající části (pro $2 < |\beta| < \beta_0 / 2$)

kmitočtové závislosti $\beta(f)$, tj. proudový zesilovací činitel β je při f_T roven jedné (ideálně ovšem). Lze nalézt vztah mezi mezními kmitočty f_α a f_T . Při kmitočtu $f = f_\alpha$ je $\alpha = 0,707 \cdot \alpha_0$, potom $\beta = 0,7$ ($1/\alpha_0 - 0,7$) = 2,3. Proto

$$f_T \approx 2,3 f_\alpha \quad (5.45)$$

Dalším mezním kmitočtem je tzv. *maximální kmitočet* f_{max} , při kterém je jednotkový přenos výkonu. Z praxe je totiž známo, že i když $|h_{21e}| < 1$, lze ještě dosáhnout výkonového zesílení tranzistoru. Je to způsobeno rozdílem mezi vstupní a výstupní impedancí tranzistoru. Kmitočet f_{max} představuje též maximální kmitočet oscilací tranzistoru. Platí

$$f_{max} = \sqrt{\frac{f_T}{8\pi r_{be} C_{cb}}} = \sqrt{\frac{g_m}{16\pi^2 r_{be} C_{bc} C_{be}}} \quad (5.46)$$



Obr. 5.55. Mezní kmitočty tranzistoru

kde R_{be} je celkový sériový odpor báze a emitoru. Maximální kmitočet je tedy přímo úměrný kmitočtu meznímu a nepřímo úměrný parazitnímu sériovému odporu v přívodu báze $R_{bb'}$ (což je rozhodující část R_{be}) a kapacitě kolektorového přechodu C_{bc} .

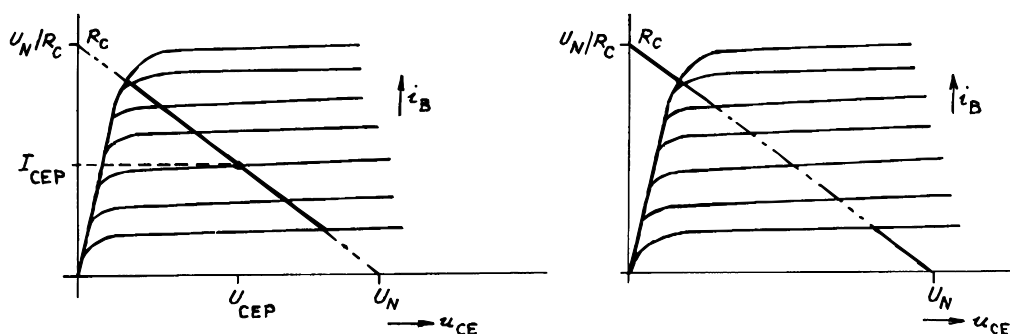
Mezní kmitočet f_I při němž $|\beta| = 1$, je poněkud vyšší než f_T (přibližně ovšem $f_I \approx f_T$), protože vlivem kapacity C_{bc} a rovněž parazitních prvků pouzdra se v oblasti kolem f_T zpomaluje zmenšování činitele $|\beta|$ s kmitočtem. Z rovnice (5.44) potom (pro $f_T \gg f_\beta$)

$$f_1 = \frac{g_m}{2\pi C_{bc}} = f_\beta \beta_0$$

Tranzitní kmitočet f_T je důležitým obecným měřítkem jakosti vf tranzistorů, má být co největší. Je také označován jako “*jakostní součin*” nebo “*součin zesílení x šířka pásma*” (viz definice). Toto označení však není přesné, neboť skutečná hodnota jakostního součinu je vlivem mezistupňových vazebních členů vždy menší než kmitočet f_T .

5.5 Principy obvodových aplikací bipolárního tranzistoru

V této kapitole si všimneme základních obvodů, které využívají bipolární tranzistory jako zesilovače a spínače. Na obr. 5.56 je ve výstupních charakteristikách nakreslena zatěžovací



Obr. 5.56. Zatěžovací přímka s vyznačenými oblastmi, ve kterých se pohybuje pracovní bod zesilovače a spínače

přímka s vyznačenými oblastmi, ve kterých se pohybuje pracovní bod zesilovače a spínače.

5.5.1 Bipolární tranzistor jako spínač

Často používáme bipolární tranzistor jako spínač - má malou spotřebu výkonu potřebného ke spínání, krátkou dobu sepnutí a vysoký opakovací kmitočet. Tranzistor jako spínač může pracovat v oblasti nasycení, v oblasti aktivní a v oblasti průrazu (obr. 5.57). Nejužívanější je spínací režim v oblasti nasycení. Jeho výhodou je odstranění vlivu rozptylu parametrů tranzistoru. Čím menší je poměr kolektorového a bázevého proudu, tím je tranzistor více uveden do nasyceného stavu. Výhodou je také malý ztrátový výkon rozptýlený na tranzistoru.

Ve spínacích obvodech má tranzistor pouze dva pracovní stavy sepnuto anebo rozepnuto. Dá se proto srovnat s mechanickým spínačem (viz obr. 5.58 a 5.59). Tranzistor bývá téměř vždy v zapojení se společným emitorem.

Je-li spínač rozepnut, odpovídá to nevodivému stavu tranzistoru, kdy zatěžovacím odporem R_Z prochází pouze zbytkový proud tranzistoru, prochází pouze zbytkový proud I_{CE0} při odpojené bázi, popřípadě (ještě menší) proud I_{CB0} při bázi spojené s emitorem. Tranzistor se chová jako velký odpor R_N . Ve vodivém stavu, který odpovídá sepnutému spínači, prochází bází tak velký proud I_B , že tranzistor je v nasyceném stavu. Jeho odpor ve vodivém stavu R_V je malý, v řádu jednotek ohmů.

Výrazným znakem je to, že tranzistorem bez budicího signálu trvale prochází kolektorový, tzv. klidový proud. Tím se zásadně liší zapojení tranzistoru - zesilovače od zapojení tranzistoru - spínače. Má-li tranzistor pracovat jako spínač, pak využíváme výlučně jeho krajní stavy: tranzistor je naplno otevřen, sepnut, anebo zavřen, rozepnut.

Obr. 5.58. Tranzistor jako spínač: tranzistor v uzavřeném stavu představuje proud I_{CE0} . Ten se ještě zmenší, je-li báze připojena k nulovému potenciálu.

I rozepnutým tranzistorem prochází určitý, malý, kolektorový proud. Tím se poněkud liší rozepnutý tranzistor od rozepnutého mechanického snímače. U spínače v rozepnutém stavu proud vůbec neprochází. U tranzistoru prochází malý tzv. zbytkový proud, bez ohledu na to, že k bázi není připojeno napětí. Značí se I_{CE0} . Nula na konci indexu značí, že třetí elektroda tedy báze, není připojena. Pro úplnost je třeba dodat, že báze bývá často spojena přes rezistor s emitorem, tedy se záporným pólem napětí zdroje. Tím se zmenší zbytkový proud, který se pak označuje I_{CER} (viz obr. 5.31). Písmeno R v indexu značí, že je zapojen rezistor mezi bázi a emitorem. Pomocí tohoto rezistoru je báze připojena k nulovému potenciálu. Má tedy stejný potenciál jako emitor. Rezistor nijak nebrání tomu, abychom přivedli na bázi kladné napětí. Pak se tranzistor otevře (sepne), tj. přejde do vodivého stavu (obr. 5.58). Tranzistor NPN potřebuje na bázi kladné napětí přibližně 1 V. U PNP tranzistoru přivedeme stejně velké, ale záporné napětí. Potom tranzistorem prochází proud I_C .

Obr. 5.59. Tranzistor jako spínač: tranzistor v otevřeném stavu představuje úbytek napětí (saturační napětí).

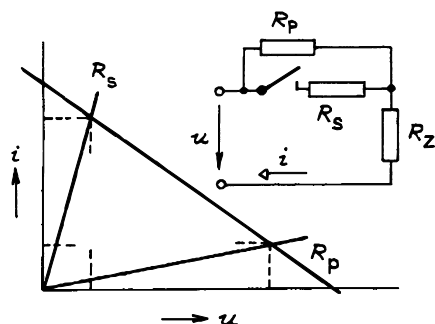
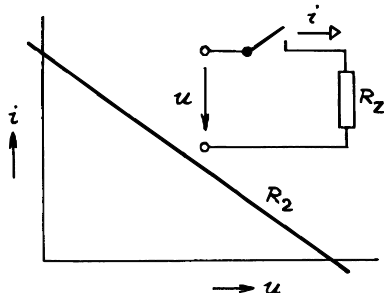
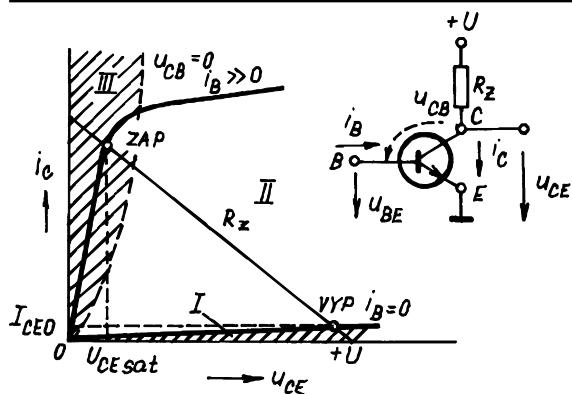
Bylo by chybou domnívat se, že sepnutý tranzistor představuje zanedbatelný odpor. I v tomto případě se liší od mechanického spínače (obr. 5.59). V obvodu C-E zůstává sice malý odpor, avšak průchodem proudu se na něm přece jen vytvoří jistý úbytek napětí, kterému říkáme saturační napětí. Podle druhu tranzistoru se pohybuje v rozmezí od 0,2 V do 1 V. Ve většině použití tento nedostatek nevadí. Jistou obtíž způsobuje skutečnost, že prochází-li proud rezistorem, spotřebovává se výkon. V tomto případě je to ztrátový výkon tranzistoru, který se přeměňuje v teplo. Ztrátový výkon vzniká i při zapojení tranzistoru jako zesilovač. I tady vzniká ztrátové teplo. Odpor, který klade procházejícímu proudu přechod E-C je v tomto případě větší, protože tranzistor je méně otevřen. Mohli bychom tedy předpokládat větší ztráty. Avšak není tomu tak, protože prochází jen (relativně) malý proud. Je výrazně menší než tzv. nasycený - saturační proud u spínacího tranzistoru.

Tranzistor pro spínací účely je konstruován tak, aby jeho odpor v nevodivém stavu byl co největší a ve vodivém co nejmenší s pokud možno velkým proudovým zesilovacím činitelem. Dalším důležitým požadavkem, kladeným na tranzistor, je jeho minimální spínací a rozpínací doba. To znamená, že skoková změna napětí báze vůči emitoru musí vyvolat v co nejkratší době odpovídající změnu kolektorového proudu.

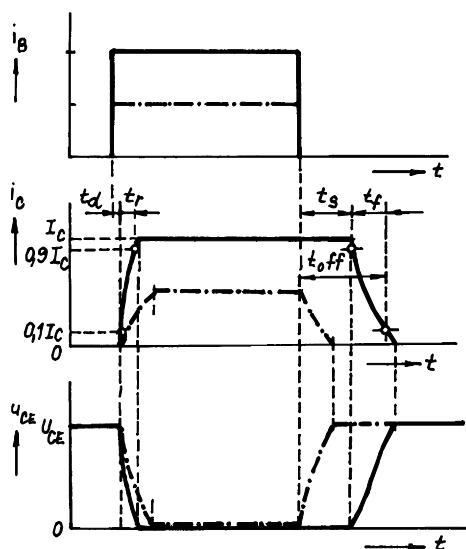
Princip použití bipolárního tranzistoru jako spínače vidíme na obr. 5.60. Přivedeme-li na bázi pravoúhlý impuls, vlivem transportu nosičů náboje tranzistorem dojde ke zkreslení impulsu kolektorového proudu. Lze určit následující zkreslení, vyjádřené časy:

t_d doba zpoždění (zpoždění náběhu, prodleva odezvy, d = delay)

t_r doba čela (náběhu, r = rise)



Obr. 5.60. a) Bipolární tranzistor v zapojení SE jako spínač (v oblasti nasycení); b) jeho náhradní zapojení



Obr. 5.61. Průběhy napětí a proudů při spínání (plnou čarou je vyznačeno spínání do saturace, čerchovanou čarou spínání do aktivní oblasti)

t_s doba přesahu (saturační zpoždění, s = saturation)

t_f doba týlu (doba sestupu, doběhu, f = fall)

Definice jednotlivých časů jsou zřejmé z obr. 5.61.

Celková doba sepnutí

$$t_{on} = t_d + t_r \approx t_r \quad (5.47)$$

protože doba t_d je obvykle zanedbatelná.

Celková doba vypnutí

$$t_{off} = t_s + t_f \quad (5.48)$$

Pro pochopení funkce spínače je třeba si uvědomit vlastnosti jednotlivých režimů (viz tab. 5.1 a obr. 5.15A) a jim odpovídajících oblastí ve výstupních charakteristikách (obr. 5.34).

I. *Oblast nevodivého stavu* tranzistoru (tj. závěrná). Je ohraničena čarou pro $i_B = 0$. Tranzistor je nevodivý a chová se jako rezistor s velkým odporem (R_p na obr. 5.60B).

II. *Oblast aktivního zesílení*. Je ohraničena čarou pro $i_B = 0$ a $u_{CB} = 0$. V této oblasti pracuje tranzistor jako lineární zesilovač. Ve spínacím režimu se tato oblast využívá jen v okolí křivky

$u_{CB} = 0$, kdy spínací tranzistor nepracuje v "nasycení", aby se dosáhlo krátkých spínacích dob.

III. *Oblast nasycení* (saturace). Je ohraničena čarou pro $u_{CB} = 0$. V této oblasti je tranzistor v zapnutém (vodivém) stavu a chová se jako rezistor s velmi malým odporem R_s na obr. 5.60B). Diferenciální vstupní odpor tranzistoru je přitom velice malý, řádově ohmy až desítky ohmů. Název "nasycení" tranzistoru vystihuje stav, kdy v bázi je nadbytek volných nosičů náboje. Tento nadbytečný náboj už nemůže vyvolat další proudové zesílení vstupního signálu, ale vyvolá přepolarizaci kolektorového přechodu PN (do propustného směru). Při vypínání, tj. při přechodu z oblasti nasycení do oblasti nevodivého stavu, je třeba odčerpat nadbytečný náboj báze, aby se kolektorový přechod polarizoval do závěrného stavu. K tomu je třeba určit doby (t_s), která prodlužuje dobu vypnutí t_{off} .

Náboj báze je odváděn změnou polarity proudu báze i_B

Obr. 5.62. Spínání tranzistoru při zátěži R (vlevo), RL (uprostřed) a RC (vpravo): pohyb pracovního bodu a výkonové poměry

(podobně jako u vypínání diod). Saturační zpoždění t_s je úměrné poměru proudů báze v přímém a závěrném směru. Běžné nízkopříkonové křemíkové tranzistory mohou mít saturační zpoždění až v řádu mikrosekund. U velmi rychlých spínacích tranzistorů je technologicky upravena oblast báze tak, aby docházelo k minimálnímu hromadění nadbytečného náboje, takže je saturační zpoždění zmenšeno řádově na desítky nanosekund (a průběhy se mění podobně jako u rychlých spínacích diod).

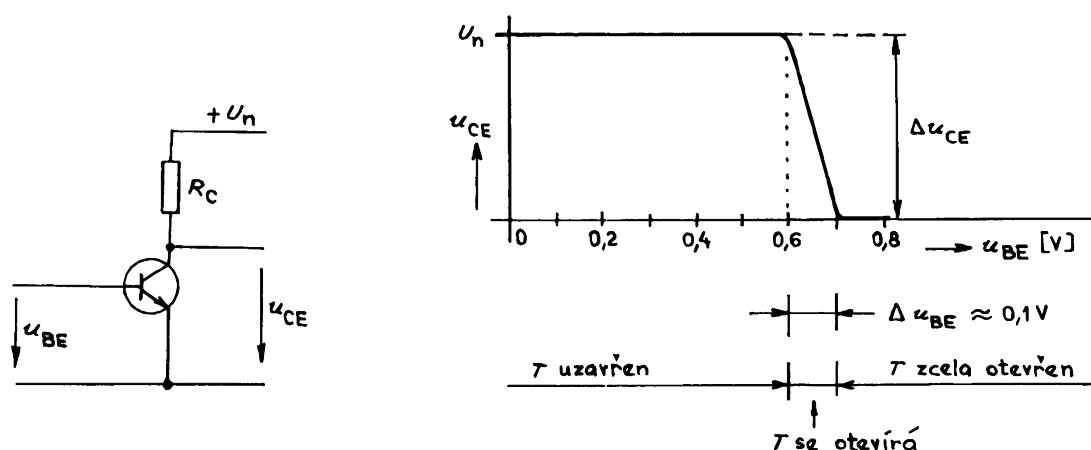
Všimněte si, že pro nenasyčený spínací režim (čerchovaně) je doba vypnutí $t_{off} = t_f$, tj. $t_s = 0$. Naopak doba zapnutí t_{on} je pro nasycený režim kratší než pro nenasyčený.

Při spínání vzniká v tranzistoru *ztrátový výkon*, a to především při "přechodu" pracovního bodu přes aktivní oblast, kdy okamžitý ztrátový výkon může dosáhnout relativně značných hodnot.

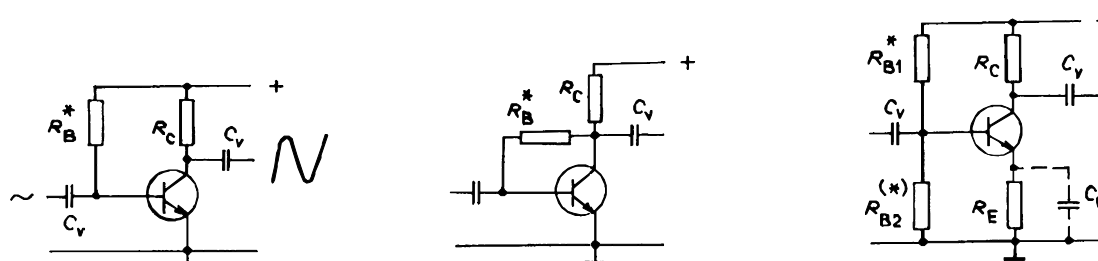
5.5.2 Bipolární tranzistor jako zesilovač

Na obr. 5.63. vidíme převodní charakteristiku jednotranzistorového zesilovače bez předpětí. Relativně malá změna vstupního napětí u_{BE} (cca 0,1 V) vyvolá velkou změnu výstupního napětí (přechod tranzistoru z uzavřeného do otevřeného stavu). Pro praktické aplikace je třeba vhodně nastavit klidový pracovní bod (zavést určité předpětí báze). U *jednostupňového zesilovače* využijeme obvodů podle obr. 5.64. Možné napětíové poměry při správně voleném pracovním bodě ukazuje obr. 5.65.

Celková situace je znázorněná na obr. 5.66. Klidový proud tranzistoru se v určitých mezích mění (kolísá), jestliže na bázi působíme malými změnami napětí. Změny kolektorového



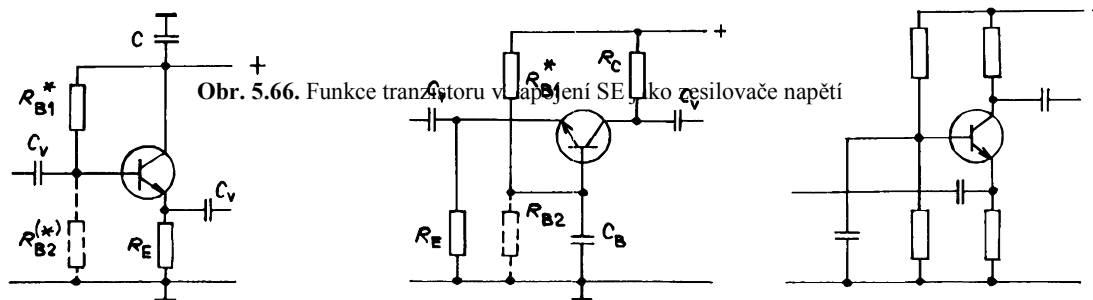
Obr. 5.63. Pevodní charakteristika jednotranzistorového zesilovače bez předpětí



Obr. 5.64. Jednostupňové zesilovače s nastavením klidového pracovního bodu

Obr. 5.65. Napětíové poměry v jednos

proudu svým průběhem odpovídají zmenám proudu báze. Proto přiváděné napětí na bázi, stejně jako napětí na kolektoru, se průběhem (tvarem) neliší. Nikoliv však velikostí. Na



Obr. 5.68. Zesilovač s tranzistorem v zapojení SC a SB (dvě varianty kreslení)

kolektoru je střídavé napětí větší, tranzistor tedy zesiluje (pokud je R_C přiměřeně velký).

Obr. 5.67. Ukázka grafického řešení jednostupňového zesilovače

Zesilovaný (vstupní) signál přivedeme na bázi tak, aby se pouze přičetl (superponoval) k nastavenému předpětí U_{BEP} (obr. 5.67.). Vazební kondenzátor zaručuje, že zesilovaný signál může mít jinou stejnosměrnou složku než činí hodnota nastaveného předpětí.

Dosud uvažované obvody jsou v zapojení se společným emitorem (SE). Jeho vlastnosti: zesiluje napětí, proud i výkon; střídavé napětí na kolektoru má opačnou fázi než napětí vstupní; výstupní odpor (20 k Ω až 100 k Ω) je větší než vstupní (500 Ω až 2 k Ω).

Zapojení se společným kolektorem (SC) se užívá v *emitorovém sledovači* (obr. 5.68a). Výstupní signál se odebrá z emitoru a má stejnou fázi jako napětí vstupní. Napěťové zesílení je menší než 1 (výstupní napětí je skoro stejné jako vstupní). Proudové a výkonové zesílení je větší než jedna. Předností je velký vstupní odpor (3 k Ω až 1 M Ω) a velmi malý výstupní odpor (asi 30 Ω až 20 k Ω). Na obr. 5.68a) je kondenzátorem C naznačeno, že kolektor je signálově na potenciálu země.

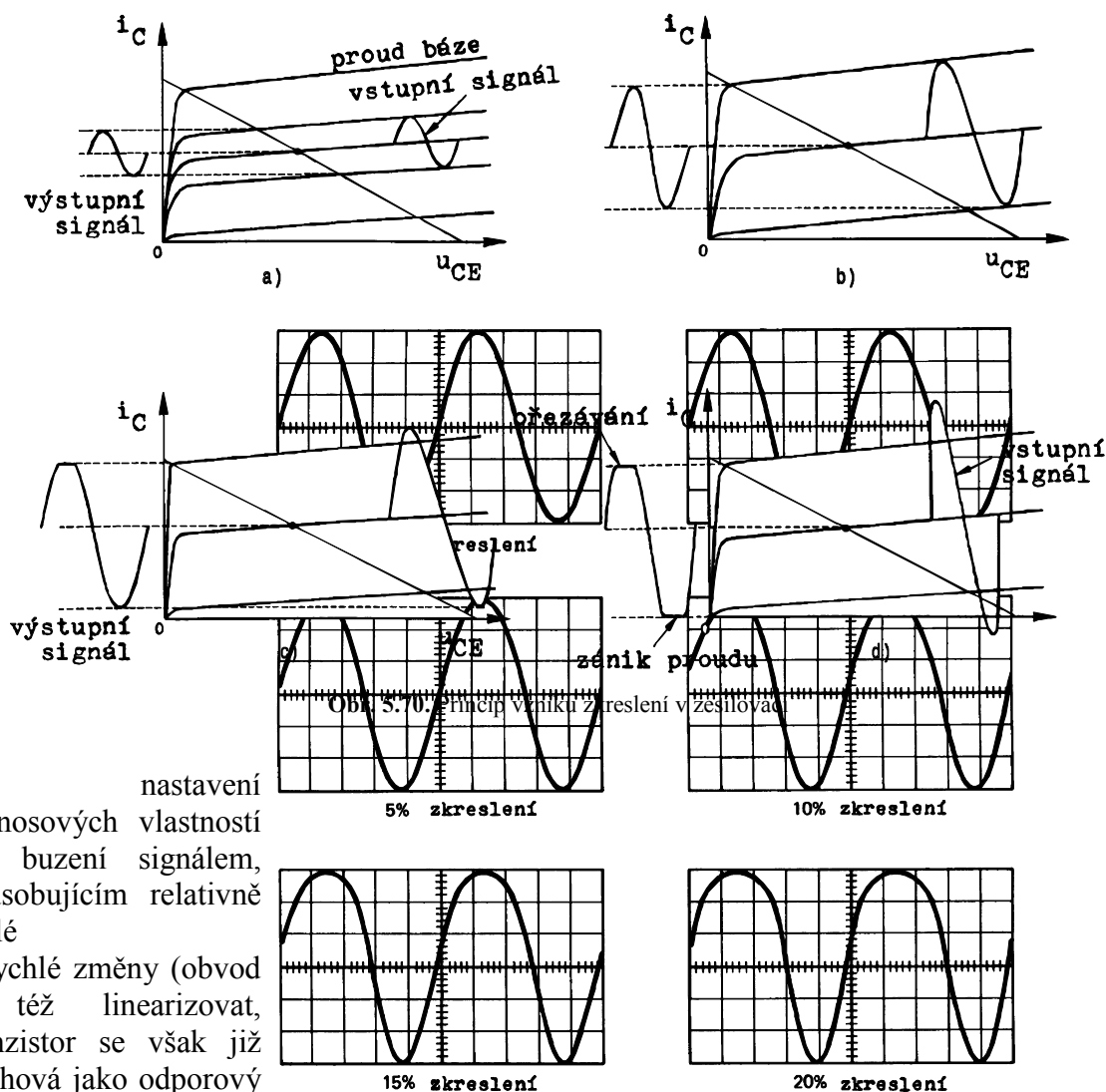
Zapojení se společnou bází (SB) má bázi pro střídavý signál uzemněnou (obr. 5.68b). Proudové zesílení je menší než jedna, napěťové zesílení je velké. Výstupní napětí je ve fázi s napětím vstupním. Vstupní odpor je velmi malý (25 až 500 Ω), výstupní odpor je naopak velký (100 k Ω až 1 M Ω).

Je zřejmé, že zapojení SE, SC i SB užívají stejné principy pro nastavení pracovního bodu. Nebylo tomu tak vždy (v 50. letech se u hrotových tranzistorů užívaly zvláštní zdroje v oblasti emitoru a kolektoru, přímo uzemněná báze a transformátorová vazba).

Návrh obvodů s bipolárními tranzistory

Při návrhu linearizovaných i nelineárních tranzistorových obvodů potřebujeme nejprve nastavit klidový (tj. stejnosměrný) pracovní bod (viz obr. 5.64 a 5.67), a potom podle charakteru zpracovávaného signálu máme čtyři možnosti:

- a) nastavení přenosových vlastností při buzení signálem, způsobujícím relativně malé a pomalé změny (obvod lze tedy linearizovat a tranzistor lze považovat za nesetrvačný, tedy odporový prvek); to odpovídá i činnosti tranzistoru v tzv. nízkofrekvenčních zesilovačích malých signálů,
- b) nastavení přenosových vlastností pro buzení relativně velkým a pomalým signálem; celý obvod je přitom nelineární, tranzistor se chová stále jako odporový prvek,



Obr. 5.69. Tvary různě zkresleného signálu na výstupu

zesilovače

- d) nastavení přenosových vlastností při buzení relativně velkým a rychlým signálem (tranzistor se v obvodu uplatňuje jako nelineární setrvačný prvek).

Klidový pracovní bod (značený P nebo také Q) závisí na zamyšleném použití tranzistoru. Při použití v impulsových a logických obvodech tranzistor používáme jako nelineární element a nastavujeme ho do stavu "sepnuto" nebo "vypnuto". V analogových linearizovaných nebo nelineárních obvodech nastavujeme klidový stav v aktivní oblasti voltampérových

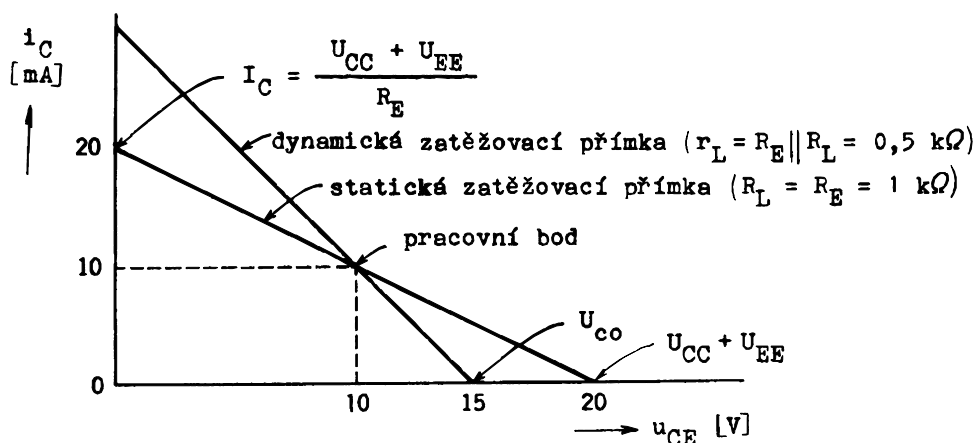
charakteristik tranzistoru. K tomu potřebujeme znát alespoň informativně průběh výstupních charakteristik, případně i charakteristiky vstupní. V linearizovaných obvodech (zesilovačích) lze zvolenou polohu pracovního bodu v aktivní oblasti velmi účinně stabilizovat s využitím silné záporné zpětné vazby na emitorovém rezistoru R_E (obr. 5.64c).

Pro dosažení dostačující teplotní stability pracovního bodu i malého rozptylu pracovních proudů při výměně tranzistoru v zapojení na obr. 5.64c postačí splnit dvě podmínky: zvolit R_E tak, aby byl na něm úbytek napětí větší než asi 2 V a zvolit vnitřní odpor děliče $R_{B1} \parallel R_{B2}$ dostatečně malý.

Je-li stejnosměrný proudový zesilovací činitel tranzistoru B (resp. h_{21E}) ≥ 50 , pak stačí volit $R_{B2} = 10 R_E$. Při výběru hodnoty R_{B1} je nutné pamatovat na to, že u všech křemíkových tranzistorů se pracovní napětí mezi bází a emitorem v aktivní oblasti pohybuje v rozmezí $U_{BE} = 0,6$ až $0,7$ V při teplotě $t = 25^\circ\text{C}$ a při vzrůstu teploty o každých 10°C klesne přibližně o 25 mV. Pro přesné nastavení pracovního bodu vždy potřebujeme soubor výstupních a vstupních charakteristik tranzistoru. Navrhne-li však obvod se stabilizací rezistorem R_E podle výše uvedených zásad, nepotřebujeme ani charakteristiky; stačí zajistit dostatečně velké U_{CE} , a to buď podle požadované amplitudy výstupního napětí anebo větší než U_{BE} (např. $U_{CE} \geq 2$ V). Nejsou-li kladeny na obvod žádné další požadavky, postačí např. postupovat takto:

1. U_n rozdělíme na přibližně 3 stejná napětí U_{RC} , U_{CE} , U_{RE} ;
2. Pro daný kolektorový proud I_C (tj. vlastně I_{CP}) určíme R_E za předpokladu, že $I_E \approx I_C$;
3. Vypočítáme odpovídající R_{B2} a R_{B1}
4. Zvolíme R_C ve vztahu k R_E dle požadovaného zesílení (viz později, např. $R_C = R_E$ pro zesílení jedna).

Pracovní bod ve stavu "rozepruto" snadno navodíme vynecháním rezistoru R_{B1} . Stav "sepruto" lze zavést vynecháním R_{B2} a splněním podmínky $R_{B1} < \beta \cdot R_C$ nebo



Obr. 5.71. Statická a dynamická zatěžovací přímka tranzistorového zesilovače (emitorového sledovače)

přesněji $R_C > \frac{R_{B1}}{\beta} + \frac{U_{BE}}{I_C}$, takže i rezistor R_C musí mít dostatečně velkou hodnotu.

a) Pro **relativně malé a pomalé změny** obvodových veličin lze obvod linearizovat. Bipolární tranzistory se nahrazují čtyřpólovými parametry. Na tomto místě je nutné upozornit na nespolehlivost některých měřičů. Jejich konstrukcí je dáno, že při měření h_{11e} a h_{21e} budí vstup tranzistoru střídavým proudem tak velké amplitudy, že při klidovém proudu $I_C < 1 \text{ mA}$ pracuje tranzistor nelineárně a změřené hodnoty jsou pak zatíženy neúnosně velikou chybou. Podobná situace nastává při měření h_{12e} a h_{22e} , je-li $U_{CE} < 6 \text{ V}$ (výstup je buzen napětím s amplitudou 5 V). Podobné vlastnosti má měřič BM 429 bývalé firmy Tesla Brno. Pamatujte, že při každém měření parametrů lineárních modelů musíte zachovat podmínky linearizace, tj. dostatečně malý budicí signál. Při buzení mezi bázi a emitor bipolárních tranzistorů můžeme tak připustit amplitudu napětí 5 mV, nejvýše 10 mV, při buzení do kolektoru nejvýše 1 V. Pak teprve můžeme očekávat dostatečnou věrohodnost naměřených parametrů.

Admitanční model se hodí mnohem lépe, a proto se téměř výhradně používá pro výpočty linealizovaných obvodů s tranzistorem. Je slučitelný také s obvyklým modelem pro tranzistory FET (viz později). Dá se zjistit, že parametry tohoto modelu pro jakýkoliv bipolární tranzistor při pracovním bodu v aktivní oblasti lze seřadit do "žebříčku" podle velikosti:

$$g_{21e} \gg g_{11e} \gg g_{22e} \gg |g_{12e}|$$

První tři parametry jsou vždy kladné, poslední je vždy záporný. Dále se dá zajistit, že hodnoty všech čtyř parametrů rostou s rostoucí hodnotou klidového pracovního proudu. Dokonce pro g_{21e} u bipolárního tranzistoru v určité pracovní oblasti platí téměř lineární závislost, kterou můžeme při pokojové teplotě vyjádřit přibližným vztahem

$$g_{21e} \approx I_C / U_T \approx 40 \cdot I_C \quad [\text{mS, mA}] \quad (5.49)$$

To bylo již ukázáno ve vztahu (5.42). O správnosti tohoto vztahu můžeme přesvědčit, známe-li závislost

$i_C = f(u_{BE})$. Vztah platí pro oblast, v níž můžeme uvedenou závislost v semilogaritmických souřadnicích nahradit přímkou (i tato voltampérová charakteristika se blíží k exponenciále; v tomto případě platí pro germanium i křemík $m = 1$). Ve skutečnosti je g_{21e} vždy poněkud menší než hodnota spočítaná z (5.49). Je to způsobeno zmenšením přenosu vlivem odporu přívodu báze. Zpravidla tato chyba není větší než 10 %.

Mezi parametry h a $y(g)$ platí následující převodní vztahy

$$g_{21e} = h_{21e} / h_{11e} \quad (5.50)$$

$$g_{11e} = 1 / h_{11e} \quad (5.51)$$

$$g_{22e} = h_{22e} - (h_{21e} \cdot h_{12e}) / h_{11e} \quad (5.52)$$

$$g_{12e} = -h_{12e} / h_{11e} \quad (5.53)$$

Při přepočtu parametrů z katalogových údajů (nebo i z vlastních měření) se stává, že vyjde g_{22e} záporné. Tento výsledek je neklamnou známkou toho, že h parametry byly nepřesně změřeny. Ve skutečnosti je g_{22e} vždy kladné a bývá až o 2 řády menší než h_{22e} . Abychom získali správný výsledek, musely by všechny h parametry být změřeny s chybou menší než 1 %. Naštěstí při většině aplikací je zatěžovací odpor na výstupu zesilovačů většinou tak malý, že lze vliv g_{22e} a g_{12e} na parametry zesilovačů zanedbat. Proto i nepřesný údaj h_{22e} a h_{12e} je nadbytečný a nepoužíváme jej. Tyto údaje by měly smysl jen v těch případech, kdy je vnitřní odpor zdroje buzení velmi velký, takže by se skutečné buzení blížilo buzení proudovému. V

moderních linearizovaných obvodech se však již s takovým buzením setkáváme velmi zřídka, protože je tímto buzením značně zmenšována rychlost odezvy obvodu.

Z těchto důvodů nejčastěji charakterizujeme lineární chování bipolárních tranzistorů pouze parametry h_{21e} a g_{21e} (h_{11e} nebo g_{11e} lze z (5.50) vypočítat). Pro určení h_{21e} použijeme katalogové údaje g_{21e} vypočítáme ze vztahu (5.49).

b) Při **buzení velkým a pomalým signálem** se tranzistor chová jako odporová nelineární součást. Zde mohou nastat dvě alternativy podle relace velikost signálu - nelinearita součásti.

- pracovní bod se během buzení pohybuje v aktivní oblasti charakteristik (např. zesilovač třídy A)

- pracovní bod se během buzení pohybuje až za zánik proudu nebo do oblasti saturace (např. zesilovač třídy AB nebo C).

Nejlepší je počítačové řešení (program PSpice).

c) Pro **buzení malým ale rychlým signálem** lze tranzistor linearizovat, je však nutné zachytit vlivy jeho setrvačnosti. Má-li pracovat s harmonickým signálem při jednom kmitočtu nebo jen ve velmi úzkém pásmu kmitočtů (např. v rezonančních mezifrekvenčních zesilovačích rozhlasových a televizních přijímačů) je výhodné jej charakterizovat admitačním modelem s komplexními parametry y . V takovém případě je nutné pro každý pracovní bod a každý kmitočet zjistit 8 složek parametrů. S těmito údaji se setkáváme v katalogových listech tranzistorů, které byly původně odzkoušeny a určeny pro rezonanční zesilovače. Tyto údaje proto platí jen pro kmitočet, který je v katalogu uveden. Některé z nich můžeme také použít pro odhad širokopásmových vlastností tranzistorů v rychlých analogových obvodech. Obvykle je

$$C_{22b} \rightarrow C_{12e} \rightarrow C_{b'e} \quad (5.54)$$

je-li kmitočet měření dostatečně malý proti kmitočtu f_T tranzistoru.

Pro výpočet rychlosti odezvy linearizovaných obvodů na přechodové signály a pro odhad horního mezního kmitočtu analogových obvodů používáme linearizovaný model tranzistoru se soustředěnými parametry, tzv. Giacolettův (obr. 5.27) a jeho zjednodušením vzniklý model "hybridní π " popsany parametry $g_{b'e}$, $C_{b'e}$, $C_{b'c}$, g_m ($b \equiv b'$). Parametry zjednodušeného modelu

$$g_m \approx 40 \cdot I_C \quad (5.55)$$

(g_m je vázáno na I_C přesněji než parametr g_{21e}),

$$g_{b'e} \approx g_{11e} = 1/h_{11e} \quad (5.56)$$

$$C_{b'e} \approx g_m / (2\pi f_T) \quad (5.57)$$

$C_{b'c}$ zjistíme z katalogu nebo z (5.54). V aktivní oblasti je $C_{b'e}$ alespoň o řád větší než $C_{b'c}$.

d) Pro **buzení velkým a rychlým signálem** nemůžeme získat nikdy dostatečně výstižný model. Tuto činnost můžeme jen velmi zhruba odhadovat. Při odhadech obvykle kombinujeme podklady ad b) a c). Linearizaci pak provádíme pro několik vybraných poloh pracovních bodů.

V současné době k řešení všech výše uvedených úloh používáme programy typu PSpice. Správnost výsledku závisí na správnosti použitých modelů.

Jednotranzistorový nízkofrekvenční zesilovač

Nyní uvedeme krátký příklad výpočtu zesilovače typu a), tj. pro relativně malé a pomalé změny obvodových veličin.

Uvažujeme obvod dle obr. 5.66 doplněný kapacitorem C_E . Napěťové zesílení je definováno vztahem

$$A_U = du_2 / du_1 \quad (5.49)$$

pro nekonečně velký zatěžovací odpor. Zanedbáme-li napětí u_{BE} , platí $u_1 = u_E u_E = i_E R_E$. Pro výstupní napětí platí $u_2 = +U_{CC} - R_C i_C = +U_{CC} - R_C h_{21b} i_E = +U_{CC} - R_C h_{21b} (u_1 / R_E)$

Dosazením do rovnice (5.49) dostaneme s uvažováním $h_{21b} \cong 1$ a např. $R_C = 1,6 \text{ k}\Omega$ a $R_E = 1 \text{ k}\Omega$

$$A_U = du_2 / du_1 = d[+U_{CC} - R_C h_{21b} (u_1 / R_E)] / du_1 = -R_C / R_E = -1,6 / 1 = -1,6 \quad (5.50)$$

Záporné znaménko znamená otočení fáze o 180° (invertor).

Poznámka: Výše uvedený přibližný vzorec platí s dobrou přesností pro případy, kdy $(h_{21e} + 1)R_E \gg R_S + h_{11e}$ (R_S je sériový odpor zdroje napětí na vstupu). To lze odvodit s pomocí náhradního lineárního obvodu s h -parametry. Uvážíme-li, že h_{11e} nabývá obvykle hodnot kolem $10 \text{ }\Omega$ a h_{21e} hodnoty kolem 100, platí odvozený vzorec s dobrou přesností pro hodnoty $R_E > 100 \text{ }\Omega$. Uvedeným způsobem lze získat zesilovač napětí nezávislý na parametrech tranzistoru (a současně je zvýšena hodnota vstupního odporu $R_{vst} = (h_{21e} + 1)R_E$).

Je-li v obvodu zapojen kapacitor C_E , chová se tento pro střídavý signál ($f = 1 \text{ kHz}$) jako zkrat rezistoru R_E , protože platí $Z_E \gg Z_{CE}$. Vzorec (5.50) aplikovaný na tento případ lze psát ve tvaru:

$$A_U = -R_C / r_E \quad (5.51)$$

kde r_E je diferenciální odpor emitoru (řádově $10 \text{ }\Omega$), který lze nalézt ve známém vzorci

$$h_{11e} \cong r_B = h_{21e} r_E \quad (5.52)$$

kde r_B je diferenciální odpor báze (při zapojeném kapacitoru C_E odpovídá vstupnímu odporu R_{vst}).

Dosazením r_E ze vzorce (5.52) do (5.51) dostaneme vzorec

$$A_U \cong - (R_C h_{21e}) / h_{11e} = - (1600 \cdot 100) / 10^3 = -160$$

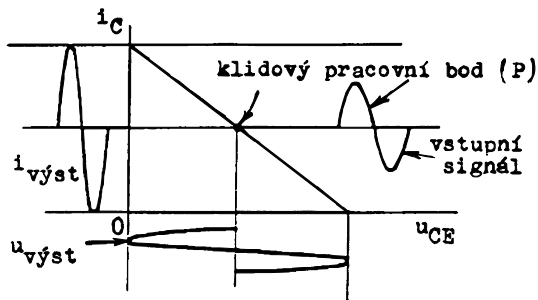
Poznámka 1: Z výsledku je patrné, že odstraněním rezistoru R_E , tj. záporné zpětné vazby, zesílení stupně výrazně vzrostlo. Se zařazeným kapacitorem C_E působí rezistor jako záporná zpětná vazba jen pro stejnosměrné napětí (teplotní stabilizace pracovního bodu).

Poznámka 2: Hodnotu r_E lze spočítat ze Shockleyho rovnice pro PN přechod:

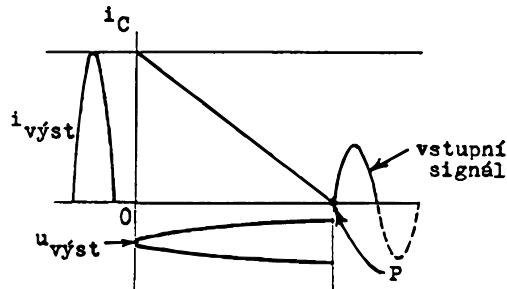
$$I_E = I_{E0} (e^{\frac{eU_{BE}}{kT}} - 1), \quad \frac{dI_E}{dU_{BE}} = \frac{1}{r_E} = \frac{eI_E}{kT} \quad \Rightarrow \quad r_E = \frac{kT}{eI_E} = 26 \text{ mV} / I_E$$

Pro $I_E = 1 \text{ mA}$ dostaneme $r_E = 26 \text{ }\Omega$. Je-li zapojen rezistor R_E , lze r_E zanedbat.

Poznámka 3: Vztah (5.52) lze odvodit následujícím způsobem:



Třída A: Pracovní bod je nastaven přibližně ve středu dynamické zatěžovací přímky. Proud teče po celých 360° periody vstupního signálu. Maximální účinnost 50%



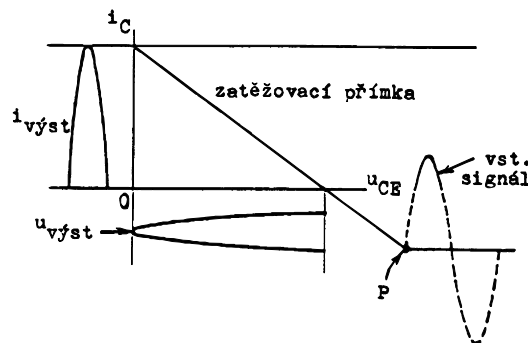
Třída B: Pracovní bod je nastaven na rozhraní aktivní a závěrné oblasti.

Proud teče po 180° periody vstupního signálu.

Maximální účinnost 50%

Třída C: Pracovní bod je nastaven v závěrné oblasti. Proud teče po méně než 180° periody vstupního signálu.

Maximální účinnost větší než 78,5% a závisí na úhlu otevření



Obr. 5.72. Nastavení klidového pracovního bodu v zesilovačích třídy A, B a C a jejich základní vlastnosti

$$r_B = \frac{\Delta u_{vst}}{\Delta I_b} \cong \frac{r_E \Delta I_e}{\Delta I_b} = \frac{r_E (h_{21e} + 1) \Delta I_b}{\Delta I_b} = r_E (h_{21e} + 1) \cong r_E h_{21e}.$$

Zpětná vazba je základním prostředkem pro konstrukci tranzistorových obvodů. Z faktu, že parametry tranzistorů se liší u jednotlivých kusů téhož typu nejméně +100 / -50 % by se totiž mohlo zdát, že asi vůbec nelze sestavit přístroj, který by spolehlivě fungoval, pracují-li v něm součástky s takovým rozptylem parametrů. Zpětná vazba záměrně natolik "zhorší" vlastnosti zesilovače, že se vyloučí působení rozdílné velikosti zesilovacího činitele (v katalogu bývá uvedeno rozmezí "od ... do ..."). Dále není vhodné zatěžovat tranzistor v oblasti mezních proudů a napětí, tj. je třeba volit přiměřenou rezervu, např. polovinu nebo čtvrtinu mezních hodnot těchto výrobcem zaručovaných parametrů.

Praktici tvrdí, že 50 až 60 vhodně vybraných typů tranzistorů nahradí téměř všechny tranzistory vyráběné v tisících typových označeních. Všechny běžné tranzistory lze totiž zařadit do několika málo typových skupin, u nichž je skutečně důležitý jen malý počet provozních vlastností, především průrazné napětí kolektoru, jeho největší přípustný proud, zesilovací čísel β (h_{21E}) je většinou nedůležitý

Poznámka 4: Výstupní odpor emitorového sledovače (zapojení SC) je $R_o = \frac{1}{y_{21e}} + \frac{R_i}{h_{21e}}$, kde

R_i je výstupní odpor zdroje signálu. První člen výrazu se uplatňuje při nejmenších pracovních proudech, druhý člen při větších proudech. Na obr. 5.73 vidíme sledovač s dvojitým napájením. Tento sledovač bude bez omezování přenášet u_i až do hodnoty $+U_{CC}$, při zatížení rezistorem R_Z však bude omezovat záporné odchylky u_i pod hodnotou $-U_{EE} \cdot R_Z / (R_Z + R_E) + U_{BE}$. Je-li na závalu nesymetrické omezování, je nutné použít komplementární sledovač.

Obr. 5.73. Poměry v emitorovém sledovači

Zatímco sledovač s tranzistorem řízeným elektrickým polem má poměrně velký výstupní odpor (přes 100 Ω), u sledovačů s bipolárními tranzistory lze snadno získat výstupní odpor poměrně malý (pod 10 Ω). Má-li zátěž takového sledovače kapacitní charakter, sledovač se s parazitní indukčností přívodu k bázi může parazitně rozkmitat na kmitočtu srovnatelném s mezním kmitočtem f_α tranzistoru. Kmity vzniknou tím snadněji, čím je hodnota C_Z bližší k C_{be} . Kmitům můžeme zabránit, jestliže zatlumíme parazitní rezonanční obvod v bázi nebo jestliže oddělíme kapacitní zátěž od emitoru. V prvním případě použijeme buď rezistor R_b podle obr. b) nebo rezistor R_c podle obr. d), ve druhém rezistor R_E podle obr. c). Při použití R_b nebo R_E je zřejmé, že přidaný rezistor zvětší výstupní odpor sledovače. Rezistor R_E má u stejnosměrných sledovačů výhodu v tom, že poněkud přispěje k teplotní stabilizaci výstupního stejnosměrného proudu. Rezistor R_c přispívá k omezení maximálního výstupního proudu - hodí se k použití v omezovačích a v zesilovačích logických členů (viz standardní výstup obvodů DTL, TTL).

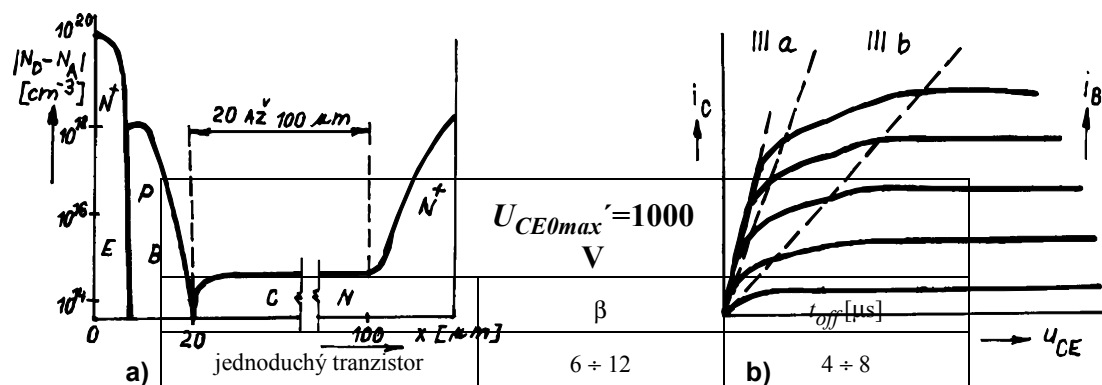
5.6 Výkonové a vysokofrekvenční tranzistory

Pro použití ve výkonové technice je třeba, aby bylo možné ovládat pomocí výkonových tranzistorů výkon na zátěži v oblasti síťových napětí. To vyžaduje, aby průrazné napětí přechodu báze-kolektor výrazně převyšovalo předpokládané pracovní napětí. Proto kolektorová oblast musí být realizována s vysokým měrným odporem (malou dotací) a velkou tloušťkou (obr. 5.74a). Relativně široká vysokoodporová oblast, nutná pro dosažení vysokého průrazného napětí, ovlivňuje výstupní charakteristiky tranzistoru. Oblast saturace se dělí na *oblast kvazisaturace* (IIIa) a *oblast plné saturace* (IIIb).

Vysokovýkonové tranzistory se realizují ve struktuře N^+PNN^+ . Dosahuje se $U_{CEOmax} = 400$ V až 1200 V. Protože požadavek vysokého proudového zesilovacího činitele α a β a nízkého saturačního napětí U_{CEsat} a U_{BEsat} je konfliktní

- s požadavkem vysokého napětí U_{CEOmax} , U_{CBOmax}
- s požadavkem krátké vypínací doby t_{off}

jsou vyvíjeny různé skupiny tranzistorů (např. z hlediska vypínací doby jde o dvě skupiny: 1. tranzistory s vysokým proudovým zesilovacím činitelem a delší vypínací dobou, 2. rychlé tranzistory s krátkou vypínací dobou a nižším proudovým zesilovacím činitelem.)



Obr. 5.74. Výkonové tranzistory: a) rozložení příměsí, b) výstupní charakteristiky s oblastí plné saturace a kvazisaturace

Bipolární výkonový tranzistor potřebuje vzhledem k nízkým hodnotám β relativně velký proud báze. Snížení proudu báze a zvýšení parametru β je možné dosáhnout integrací dvou nebo tří výkonových tranzistorových struktur v *Darlingtonově zapojení* (obr. 5.75.). Dosažený stav ilustruje tab. 5.3.

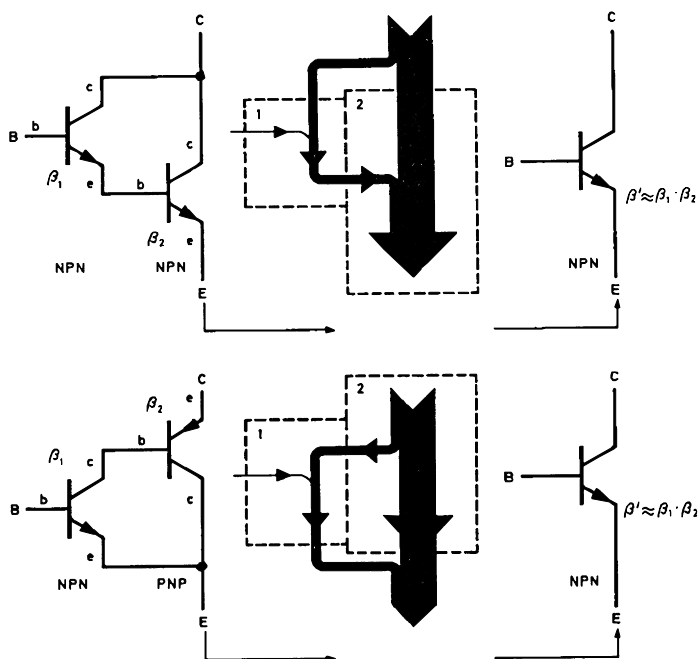
Tab 5.3.: Srovnání jednoduchého tranzistoru a Darlingtonovy dvojice

Nevýhodou Darlingtonova zapojení jsou vyšší napětí U_{CEsat} (a tedy i vyšší výkonová ztráta) a proudový zesilovací činitel β při vysokých proudových hustotách velmi rychle klesá s kolektorovým proudem ($\beta \sim i_C^{-3}$), u jednoduchého tranzistoru je pokles pomalejší ($\beta \sim i_C^{-2}$). Také se zvětšuje doba vypnutí t_{off} .

Poznámky k výběru tranzistorů pro aplikace

V katalogu jsou tranzistory hrubě rozříděny do skupin:

- pro nf zesilovače
- vysokofrekvenční a spínací



Obr. 5.75. Darlingtonovy dvojice tranzistorů NPN-NPN a NPN-PNP

- pro kanálové anténní zesilovače UHF
- VHF
- pro malé vysílače
- pro nf zesilovače výkonu
- výkonové
- výkonové spínací
- pro diferenční obvody

Toto třídění pouze naznačuje původní určení jednotlivých typů tranzistorů, neusnadňuje však jejich výběr pro nová, všeobecnější použití, naopak v určitém směru jej znesnadňuje. Např. z údajů o f_T zjistíme, že téměř všechny tzv. nízkofrekvenční tranzistory mají vyšší tranzitní kmitočet než většina tzv. vysokofrekvenčních tranzistorů. Pro výběr tranzistorů pro rychlé analogové obvody je důležité znát průchozí kapacity (tedy kapacity

kolektorového přechodu). Údaj o nich však v katalogových tabulkách chybí. Tyto údaje najdeme v různých katalogových listech. Následující přehled přináší informativní roztřídění některých dostupných bipolárních tranzistorů podle kapacit kolektorového přechodu.

C_{bc}	C_{ce}	Typy
0,5 pF	2pF	BF 167 BF 173, SF 240, SF 245
1 pF	2pF	KF 124, KF 125, KF 524, KF 525
3 pF		KF 503, KF 504
4 pF		KC 147-149, KC 507-509, BC 157-159, BC 177-179, KSY 71, KSY 72, BF 257-259
6 pF		KF 621, KF 630, KS 500, KSY 21, KSY 34, KSY 62
8 pF		2N 2905, 2N 2907
10 pF		BSXP 59-61
20 pF		KF 506-508, KF 517, KFY 16, KFY 18, KFY 34, KFY 46
30 pF		BC 211, BC 313
70 pF		BD 354, BD 355
140 pF		SU 161
250 pF		KU 601, KU 602
750 pF		KU 605, KU 606

Provedeme-li užší výběr tranzistorů podle tvaru výstupních charakteristik (některé v tranzistory mají nevhodně veliké zbytkové napětí v sepnutém stavu), podle tranzitních kmitočtů přepočítaných na stejný pracovní proud (např. $I_C = 2$ mA) a podle uvedených kapacit, zjistíme, že pro všeobecné aplikace v analogových i spínacích obvodech bude možné vhodné typy seřadit podle rychlostí od nejrychlejšího po nejpomalejší typ:

SF 245, KF 124, KC 147-149, KC 507-509,

BC 157-159, BC 177-179

Tranzistory řady KC se neosvědčují jen v takových obvodech, kde se vyžaduje velmi malý odpor přívodu báze (mají odpor báze až několik set ohmů). Naopak pro použití v emitorových sledovačích se pro velký odpor báze ideálně hodí - sledovače s nimi jsou spolehlivě stabilní i při kapacitní zátěži.

Označení tranzistoru je kódováno kombinací písmenného a číselného znaku. V evropském značení se znak tranzistoru skládá buď ze dvou písmen a tří číslic nebo ze tří písmen a dvou číslic. Za číselným znakem obvykle následuje další písmeno.

První písmeno (v obou jmenovaných skupinách):

- A - germaniové tranzistory
- B - křemíkové tranzistory
- G - germaniové tranzistory (značení TESLA)
- K - křemíkové tranzistory (značení TESLA)

Druhé písmeno (v obou skupinách):

- C - nízkofrekvenční tranzistory
- D - nízkofrekvenční výkonové tranzistory
- F - vysokofrekvenční tranzistory
- L - vysokofrekvenční výkonové tranzistory
- S - spínací tranzistory
- U - spínací výkonové tranzistory

Třetí písmeno ve druhé skupině: Neobsahuje zpravidla žádné technické údaje, výrobce jím sděluje jiné informace.

Příklad: uvažujme tranzistory označené:

	BC 168 A		BC
108 A			
	BC	168	B
BC 108 B			
	BC	168	C
BC 108 C			

Ve všech případech se jedná o křemíkový nízkofrekvenční tranzistor, určený pro napěťové zesilovače. Liší se pouzdem, což vyjadřuje rozdílný číselný znak. První trojice označuje tranzistory v plastovém pouzdru, druhá trojice v kovovém pouzdru. Písmena za číselným znakem udávají zesilovací činitel tranzistoru, konkrétně: A = 170, B = 290 a C = 500. Všechny ostatní parametry těchto tranzistorů jsou stejné.

Mimoevropští výrobci používají i jiná značení, takže bez srovnávacích tabulek se neobejdeme. Americké značení tranzistorů začíná dvojicí znaků 2N, např. 2N2218A. Japonské tranzistory mají označení začínající 2CA..., 2SB..., podobně 2SC, 2SD, 2SJ, 2SK, 3SK atd.

6 UNIPOLÁRNÍ TRANZISTORY

Unipolární tranzistor je založen na principu řízení pohybu nosičů náboje elektrickým polem, přičemž vedení proudu se uskutečňuje v tzv. *kanále* pouze jedním typem nosičů náboje (proto *unipolární*). Unipolární tranzistory také nazýváme *tranzistory řízené elektrickým polem* a užíváme užíváme pro ně zkratku FET [Field - Effect Transistor].

Vodivost tohoto typu tranzistoru je - na rozdíl od bipolárního tranzistoru, kde vodivost je řízena injekcí minoritních nosičů do báze - řízena dvojím způsobem:

1. *změnou průřezu vodivého kanálu rozšiřováním depletiční vrstvy* přechodu PN nebo MS pólovaného v závěrném směru rozšiřováním závěrné vrstvy, tj. účinkem prostorového náboje přechodu PN, nebo
2. *změnou koncentrace majoritních nosičů v inverzním kanálu* (inverzní povrchové vrstvě) pod izolační vrstvou dielektrika, tj. účinkem indukovaného náboje pod vrstvičkou izolantu.

V současné době můžeme unipolární tranzistory rozdělit na tři základní typy:

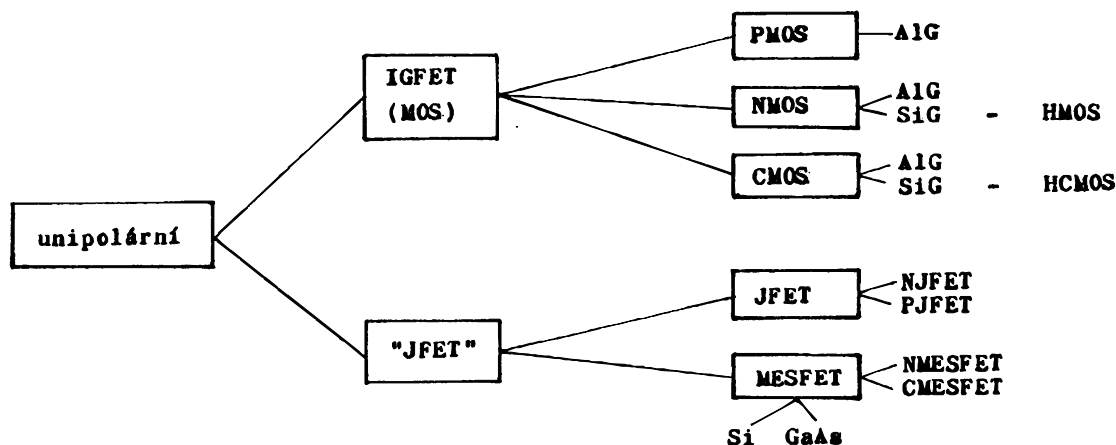
- **tranzistory s přechodovým hradlem** označované zkratkou **JFET** [Junction FET]
- **tranzistory s izolovaným hradlem** označované **IGFET** [Insulated Gate FET]
- **tenkovrstvé tranzistory s izolovaným hradlem** označované **TFT** [Thin Film Transistors].

Uvidíme později, že tranzistory JFET jsou obvykle s přechodem PN, tedy J_{pN} FET a v praxi označované jednoduše JFET, ve speciálních případech se Schottkyho přechodem, tedy J_{MS} FET a v praxi označované jako MESFET. Tranzistory IGFET mají strukturu kov-izolant-polovodič, které odpovídá zkratka MISFET (zkráceně MIS). Ta se v praxi neujala a užívá se zkratka MOSFET a především její zkrácený tvar MOS (které odpovídají konstrukcím se strukturou kov-oxid-polovodič), a to i pro tranzistory se strukturou MNOS (-nitrid-oxid-) nebo dokonce SNOS, kde hradlo je z polykrystalického křemíku.

K tranzistoru TFT poznamenejme jen, že nachází uplatnění až v poslední době, např. v ploché televizní obrazovce a je zhotoven z amorfního křemíku.

Dále v rámci právě uvedeného dělení rozlišujeme unipolární tranzistory s

- kanálem typu N (nosiči proudu jsou elektrony), označované zkratkou **NMOS**,
- kanálem typu P (nosiči proudu jsou díry), označované zkratkou **PMOS**.



Obr. 6.1. Rozdělení unipolárních tranzistorů s použitím obvyklých zkratk

Uvedené zkratky je výhodné používat ve stále stejném grafickém tvaru, tj. např. ...tranzistor FET ... a ne ... tranzistor FE ...

Protože proud v unipolárních tranzistorech je přenášen majoritními nosiči, jsou tyto prvky odolnější vůči změnám teploty a dopadajícímu ionizujícímu záření než bipolární tranzistory.

Pro nepřítomnost minoritních nosičů jsou zapínací a vypínací doby unipolárních struktur dány především parazitními kapacitami, které musí být nabity a vybity při každém sepnutí a vypnutí. Tyto kapacity jsou teplotně nezávislé, a proto ani časy t_{on} a t_{off} nejsou teplotně závislé, což je velká výhoda oproti bipolárním součástkám. Nesetkáváme se tedy s jevy akumulace (hromadění) minoritních nosičů a jejich postupnou rekombinací.

Obr. 6.2. Principiální uspořádání unipolárních tranzistorů. Schematicky je znázorněn vliv elektrického pole pod elektrodou G: vyvolává změnu průřezu nebo koncentrace

6.1 Unipolární tranzistor s přechodem PN (tranzistory JFET)

Na obr. 6.2 vidíme princip technologického provedení tranzistoru JFET a jeho schématickou značku. Vlastní pracovní oblast se nazývá *kanál* a zde má vodivost N. Kanál ústí do dvou oblastí N^+ , které jsou opatřeny neusměrňujícími kontakty. Jeden z nich je *emitor* (je označován písmenem S jako [source], tj. zdroj), druhý je *kolektor* (je označován písmenem D jako [drain], tj. odtok, nora, sběrná elektroda). Kontakty na horní a dolní straně destičky jsou přívody k řídicí elektrodě G (hradlo [gate]), kterou tvoří vrstva polovodiče opačné vodivosti než kanál, zde P^+ .

Přechod P^+N mezi hradlem a emitorem je polarizován v závěrném směru a potenciál (napětí) hradla ovládá proud i_D tekoucí kanálem mezi emitorem a kolektorem. Víme z dřívějšího výkladu, že jeli přechod polarizován vnějším napětím (zde např. U_{GS}) v závěrném směru, rozšiřuje se oblast prostorového náboje do méně dotovaného polovodiče, tj. N, tj. kanálu. Tím dochází ke změně tloušťky uvažovaného vodivého kanálu (oblast prostorového náboje má odpor mnohem větší než odpor polovodiče N).

6.1.1 Kvalitativní popis činnosti tranzistoru JFET

Pro pochopení principu činnosti tranzistoru JFET předpokládejme zjednodušenou

Obr. 6.3. Tranzistor JFET: princip a schématická značka

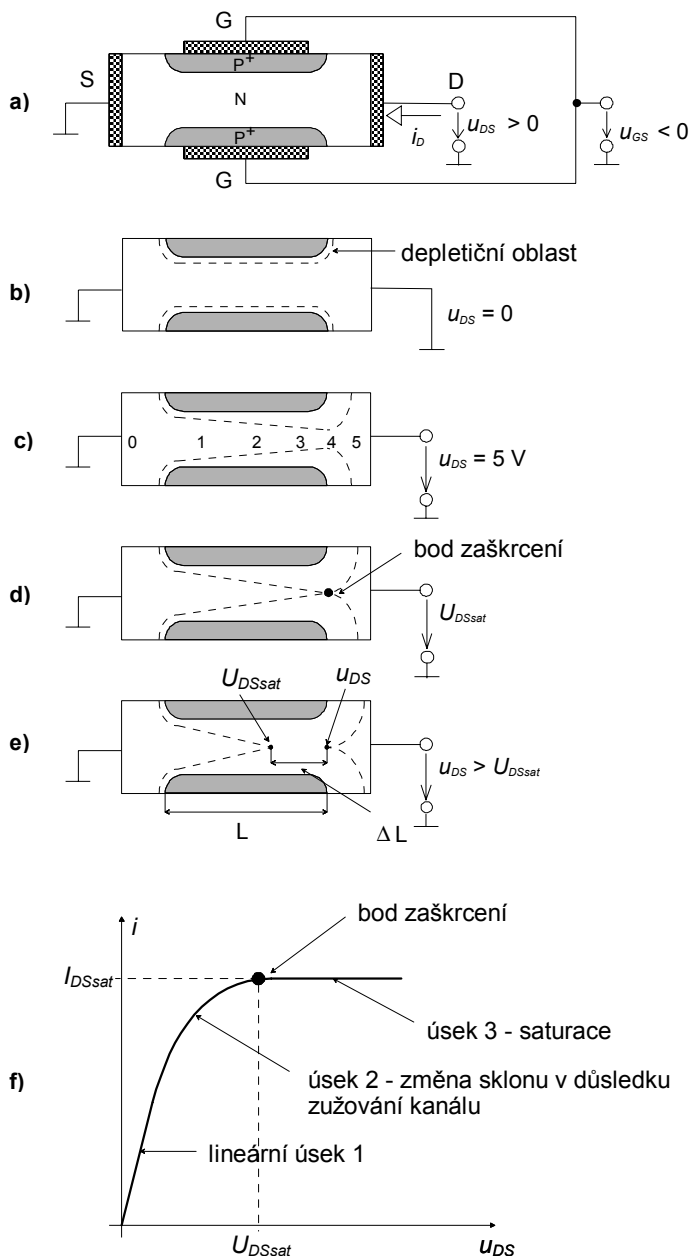
symetrickou strukturu na obr. 6.3a). Protože oba P^+N přechody ohraničující kanál musí být pro správnou činnost tranzistoru závěrně pólovány, je $u_{GS} \leq 0$. Dále předpokládejme, že $u_{DS} \geq 0$.

Nejprve si vysvětlíme činnost tranzistoru při uzeměné hradlové elektrodě, tj. $u_{GS} = 0$, a napětí u_{DS} se pomalu zvětšuje z hodnoty $u_{DS} = 0$. Pro $u_{DS} = 0$ je tranzistor v termodynamické rovnováze a v okolí P^+N přechodu se vytvoří depletiční oblast (viz obr. 6.5b), která se rozšíří

Obr. 6.4. Příklady uspořádání unipolárních tranzistorů: a)
JFET (tranzistor
s hradlem odděleným přechodem PN a s kanálem N, b)
MESFET (tranzistor
s hradlem odděleným přechodem kov-polovodič a s kanálem
P)

převážně do méně dotované N oblasti. Kvazineutrální oblast typu N mezi oběma depletičními oblastmi nazýváme *kanál*. Zvýšíme-li napětí u_{DS} nad nulovou hodnotu, začne kanálem protékat proud i_D . Pro malá napětí u_{DS} (typicky do několika desetin voltů) se součástka chová jako rezistor a proud i_D vzrůstá lineárně se změnou napětí u_{DS} (viz obr. 6.5f, úsek 1). Při dalším vzrůstu napětí u_{DS} dojde vlivem konečného odporu kanálu k rozdělení potenciálu, které je pro $u_{DS} = 5 \text{ V}$ naznačeno na obr. 6.5d). Protože $u_{GS} = 0$, bude přechod P^+N pólován "závěrněji" v blízkosti elektrody D (např. $u_R = 4 \text{ V}$) než v blízkosti elektrody S (např. $u_R = 1 \text{ V}$). Proto také depletiční vrstva bude širší u elektrody D než u elektrody S (viz obr. 6.5c). Součástka se už nechová jako jednoduchý rezistor, musíme brát v úvahu zužování vodivého kanálu a tím i vzrůst odporu se vzrůstajícím napětím. Dochází proto ke zpomalení vzrůstu proudu i_D se vzrůstajícím napětím u_{DS} , jak je naznačeno na obr. 6.5f, úsek 2; charakteristika má "koleno". Jestliže dále zvyšujeme napětí u_{DS} , zužování kanálu pokračuje, až při napětí u_{DSsat} dojde k úplnému zaškrcení vodivého kanálu [pinch-off] u elektrody D (obr. 6.5d).

Pro napětí $u_{DS} \geq U_{DSsat}$ je velikost proudu i_D konstantní a je rovna I_{Dsat} , dochází pouze ke zkracování délky vodivého kanálu, jak je naznačeno na obr. 6.5e). Hovoříme o *saturačním napětí* U_{DSsat} a *saturačním proudu* I_{Dsat} . To, že elektrony mohou vůbec projít ochuzenou vrstvou vzniklou před kolektorem, je způsobeno tím, že hradící účinky působí od kanálu N k hradlu (P^+), ale nevznikají od kanálu N ke kolektoru (N^+). Můžeme si to představit tak, že **emitorový přechod N^+N injektuje elektrony do kanálu a kolektorový přechod NN^+ je polarizován v závěrném směru a tyto nosiče extrahuje.**



Obr. 6.3. a) - e) Znáznornění jednotlivých fází činnosti tranzistoru JFET při $u_{GS} = 0$; f) ampérovoltová charakteristika tranzistoru JFET při $u_{GS} = 0$

Zbývá nám ještě vysvětlit, proč pro napětí $u_{DS} \geq U_{DSsat}$ zůstává proud i_D konstantní.

Předpokládejme zkrácení délky vodivého kanálu o ΔL (obr. 6.5e). Jak vidíme na obr. 6.5e), napětí na kolektorové straně úseku ΔL je u_{DS} , zatímco na emitorové straně je U_{DSsat} . Přiložené napětí, které převyšuje U_{DSsat} , tedy $u_{DS} - U_{DSsat}$ je úbytek napětí na depletiční části kanálu ΔL . Jestliže je tedy $\Delta L \ll L$ (což je obvyklý případ), pak má vodivá část kanálu stejný tvar a také napětí se na obou jeho koncích nemění, zůstává rovné nule resp. U_{DSsat} . Proto také proud i_D zůstává po dosažení saturace konstantní

a rovný I_{Dsat} . Bude-li se napětí u_{DS} dále zvyšovat, dojde při určité hodnotě tohoto napětí k průrazu mezi hradlem a kolektorem. Nyní předpokládejme, že na elektrodu G je přivedeno záporné napětí, tedy $u_{GS} < 0$. Toto napětí způsobí, že P^+N přechod je polarizován závěrně i při $u_{DS} = 0$ a depletiční vrstva je širší (a vodivý kanál užší) než v předchozím případě.

Počáteční vodivost kanálu je tedy menší a při zvyšování napětí u_{DS} proud i_D vzrůstá pomaleji (obr. 6.6b)

a dříve dochází k zaškrcení kanálu. Hodnoty U_{DSsat} a I_{Dsat} jsou menší než při $u_{GS} = 0$. Při dostatečně velkém závěrném napětí u_{GS} je depletiční vrstva rozšířena přes celou šířku kanálu už při $u_{DS} = 0$ (obr. 6.6a) a strukturou neprotéká žádný proud. Toto napětí $u_{GS} = U_P$ nazýváme *prahové napětí* (označuje se i U_T , [threshold voltage]).

Pro $u_{GS} \leq U_P$ zůstává kolektorový proud i_D nulový pro všechna napětí $u_{DS} \geq 0$.

6.1.2 Kvantitativní analýza tranzistoru JFET

Při odvození velikosti proudu i_D protékajícího tranzistorem JFET vyjdeme z jeho zjednodušené symetrické struktury a zavedeme soustavu souřadnic s počátky naznačenými na obr. 6.7. Předpokládáme homogenní koncentraci donorů v kanálu rovnou N_D a zanedbáme úbytek napětí mezi elektrodou S a bodem $y = 0$ a elektrodou D a bodem $y = L$.

Při výpočtu proudu i_D před zaškrcením kanálu, tedy pro $0 \leq u_{DS} \leq U_{DSsat}$ a $0 \geq u_{GS} \geq U_P$, vyjdeme z obecného vztahu pro proudovou hustotu v kanále

$$\mathbf{J}_n = q\mu_n n \mathbf{E} + qD_n \nabla n \quad (6.1)$$

Za zjednodušujících podmínek ($n \approx N_D$, proud tekoucí pouze ve směru y a zanedbatelně malá difúzní složka) můžeme pro proudovou hustotu ve vodivém kanále psát

$$J_n = J_{ny} = q\mu_n N_D E_y = -q\mu_n N_D \frac{dV}{dy} \quad (6.2)$$

Protože proud tekoucí kanálem musí být v kterémkoli místě kanálu roven i_D (nikde nedochází k úbytku nebo nárůstu počtu nosičů, pak integrací proudové hustoty přes průřez kanálu v libovolném bodě y dostaneme

$$i_D = \int \int J_{ny} dx dz = Z \int_{w(y)}^{2a-w(y)} J_{ny} dx = -2Z \int_{w(y)}^a q\mu_n N_D \frac{dV}{dy} dx = -2Zq\mu_n N_D \frac{dV}{dy} a \left(1 - \frac{w(y)}{a}\right) \quad (6.3)$$

V (6.3) jsme vzali v úvahu fakt, že struktura je symetrická podle roviny $x = a$. Protože proud i_D nezávisí na y , můžeme jej integrovat podle délky kanálu

$$\int_0^L i_D dy = i_D L = -2Zq\mu_n N_D a \int_{V(0)=0}^{V(L)=u_{DS}} \left[1 - \frac{w(V)}{a} \right] dV \quad (6.4)$$

Přejdeme-li od potenciálu k napětí, pak

$$i_D = -\frac{2Zq\mu_n N_D a}{L} \int_0^{u_{DS}} \left[1 - \frac{w(u)}{a} \right] du \quad (6.5)$$

Řešení této rovnice, které je ve skutečnosti dvojrozměrné, můžeme zjednodušit, jestliže předpokládáme, že $L \gg a$ (JFET s "dlouhým kanálem"). Potom se elektrostatické veličiny ve směru y mění pomalu ve srovnání s jejich změnou ve směru x a pro šířku depletiční oblasti uvažovaného P⁺N přechodu platí rovnice (2.25)

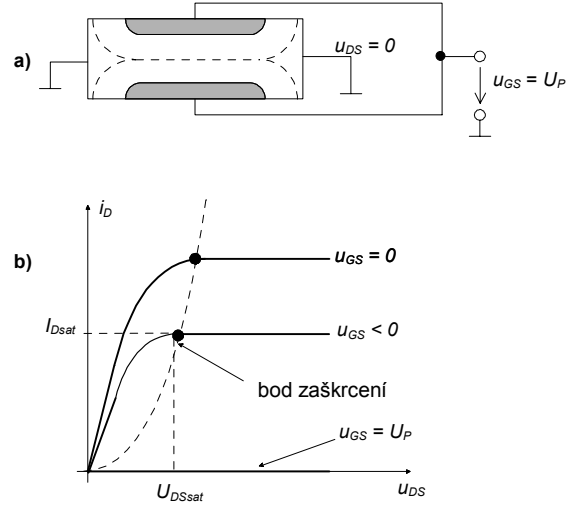
$$w(u) = \left[\frac{2\epsilon}{q} (U_D - u) \right]^{1/2} = \left[\frac{2\epsilon}{q} (U_D + u_k(y) - u_{GS}) \right]^{1/2} \quad (6.6)$$

kde u_D je difúzní napětí, $u_k(y)$ je napětí ve středu kanálu (jeho "původcem" je napětí u_{DS} - viz např. obr. 6.5c) a $u = u_{GS} - u_k(y)$ je napětí přiložené na P⁺N přechod v daném bodě y (viz obr. 6.7.). Víme, že jestliže je $u_{DS} = 0$ (tedy $u_k(y) = 0$) a $u_{GS} = U_P$, rozšíří se depletiční vrstva přes celou šířku kanálu (obr. 6.6.a), $w(y) = a$. Rovnice (6.6) v tomto případě přechází do tvaru

$$a = \left[\frac{2\epsilon}{qN_D} (U_D - U_P) \right]^{1/2} \quad (6.7)$$

a výraz

$$\frac{w(y)}{a} = \left(\frac{U_D + u_k(y) - u_{GS}}{U_D - U_P} \right)^{1/2} \quad (6.8)$$



Obr.6.4. a) Zaškrvení kanálu při $u_{GS} = U_P$; b) průběh AV-charakteristiky pro $u_{GS} < 0$

Po dosazení (6.8) do (6.5) a po integraci dostaneme pro $0 \leq u_{DS} \leq U_{DSsat}$ a $0 \geq u_{GS} \geq U_P$

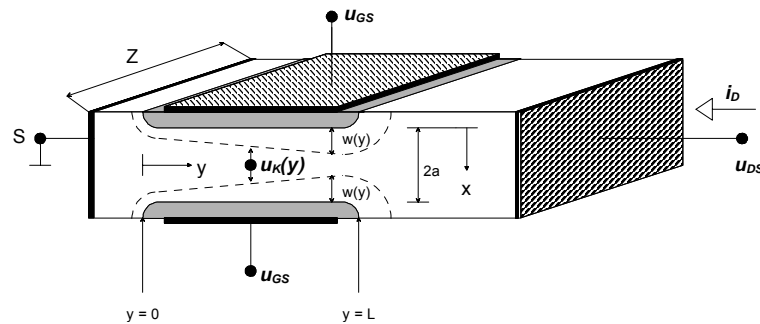
$$i_D = -\frac{2Zq\mu_n N_D a}{L} \left\{ u_{DS} - \frac{2}{3} (U_D - U_P) \left[\left(\frac{u_{DS} + U_D - u_{GS}}{U_D - U_P} \right)^{3/2} - \left(\frac{U_D - u_{GS}}{U_D - U_P} \right)^{3/2} \right] \right\} \quad (6.9)$$

Zanedbáme-li ve výrazu (6.9) u_D a označíme-li

$$I_{Dsat0} = -\frac{2Zq\mu_n N_D a U_P}{3L} \quad (6.10)$$

získáme po úpravě výraz

$$i_D = I_{Dsat0} \left\{ 3 \frac{u_{DS}}{U_P} - 2 \left[\left(\frac{u_{DS} - u_{GS}}{-U_P} \right)^{3/2} - \left(\frac{-u_{GS}}{-U_P} \right)^{3/2} \right] \right\} \quad (6.11)$$



obr. 6.5. Struktura tranzistoru JFET, její rozměry a orientace souřadnic pro kvantitativní analýzu

nebo v normalizovaném tvaru

$$\frac{i_D}{I_{Dsat0}} = 3 \frac{u_{DS}}{U_P} - 2 \left[\left(\frac{u_{GS} - u_{DS}}{U_P} \right)^{3/2} - \left(\frac{u_{GS}}{U_P} \right)^{3/2} \right] \quad (6.12)$$

Vztah (6.12) je již hledaná rovnice udávající závislost kolektorového proudu i_D na napětí u_{DS} kolektoru a napětí u_{GS} hradla vzhledem k emitoru. Proud i_{Dsat0} je saturační proud pro $u_{GS} = 0$, U_P je prahové napětí, jehož hodnota je daná strukturou tranzistoru JFET a jeho technologií.

6.1.3 Statické charakteristiky tranzistoru JFET

S ohledem na skutečnost, že proud hradlem je velmi malý (pokud $u_{GS} < 0$ V), jsou i voltampérové charakteristiky tranzistorů JFET jednodušší než u bipolárních tranzistorů. Chybí ty charakteristiky, které obsahují vstupní proud (tj. vstupní a zpětná převodní). Kreslí se sít' převodních a výstupních charakteristik nebo jen sít' výstupních charakteristik (převodní jdou z ní odvodit). Tranzistor JFET může pracovat v různých režimech. **Soustava normovaných výstupních charakteristik** je pro jednotlivé pracovní režimy znázorněna na obr. 6.8. Omezení činnosti tranzistoru JFET je dáno zejména podmínkou, že přechod P⁺N mezi hradlem a kanálem nesmí mít v žádném místě kanálu kladné napětí (resp. musí být menší než cca 0,5 V u Si tranzistorů).

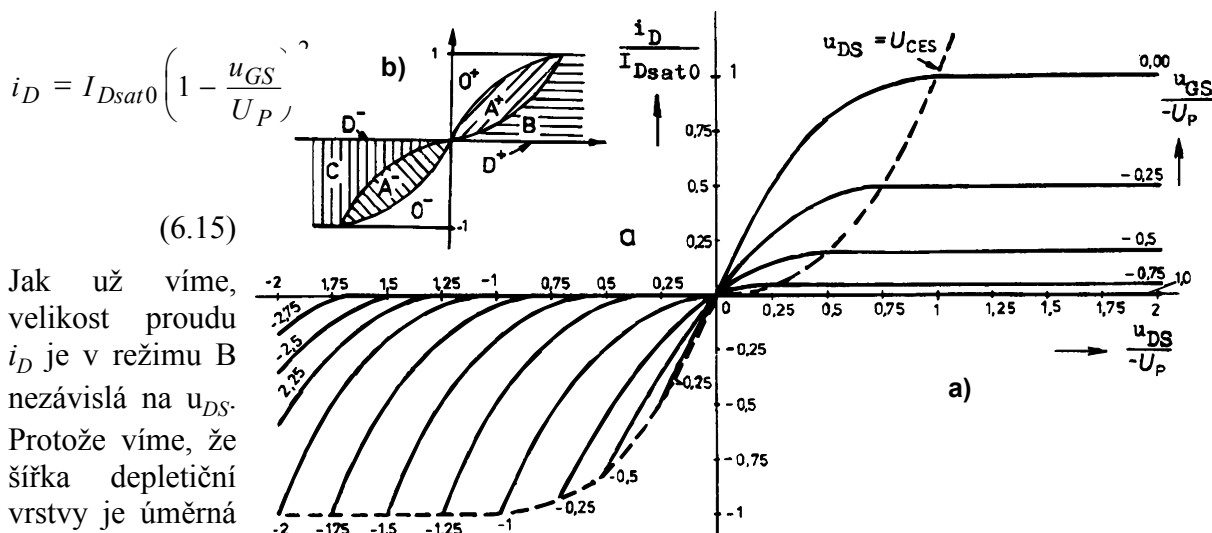
Základní pracovní režim tranzistoru JFET je **aktivní režim A**, kdy je velikost proudu i_D dána rovnicí (6.12). Tento režim se dělí na režim A⁺ při kladném napětí kolektoru (v I. kvadrantu) a režim A⁻ při záporném napětí kolektoru (ve III. kvadrantu). **Režim A⁺** končí nasycením (saturací) kolektorového proudu při

$$u_{DS} = U_{DSsat} = u_{GS} - U_P \quad (6.13)$$

kdy dojde k zaškrcení kanálu na jeho kolektorovém konci. Vztah pro velikost proudu v **saturačním režimu B** dostaneme dosazením (6.13) do (6.12)

$$\frac{i_D}{I_{Dsat0}} = 1 - 3 \frac{-u_{GS}}{-U_P} + 2 \left(\frac{u_{GS}}{U_P} \right)^{3/2} \quad (6.14)$$

Tento vztah, který zároveň vyjadřuje převodní charakteristiku tranzistoru v režimu B, bývá při praktických výpočtech s dobrou přesností nahrazen mnohem jednodušším vztahem



Obr. 6.8. a) Normované výstupní charakteristiky a b) pracovní oblasti tranzistoru JFET

Jak už víme, velikost proudu i_D je v režimu B nezávislá na u_{DS} . Protože víme, že šířka depletiční vrstvy je úměrná odmocnině ze závěr-ného

napětí, musí být

funkční závislost $i_D = f(u_{GS})$, $u_{DS} = \text{konst.}$, rovněž nelineární (parabolická, přibližně kvadratická). To potvrzují i rovnice (6.14) a (6.15). V soustavě výstupních charakteristik se jejich vzdálenost při klesajícím u_{GS} zmenšuje. Soustava charakteristik je v režimu A i B omezena charakteristikou pro $u_{GS} = 0$ V. Při překročení této hodnoty (kladné u_{GS}) by tranzistor přešel do **nulového režimu 0^+** , při němž by měl PN přechod u emitorového konce kanálu kladné předpětí, hradlem by tekla proud a tranzistor by nepracoval (prakticky od + 0,5 V).

Aktivní inverzní režim A^- je omezen charakteristikou pro $u_{GS} = U_P$ a křivkou $u_{DS} = u_{GS} < 0$.

Při překročení druhé podmínky by tranzistor přešel do **nulového režimu 0^-** , při němž by měl PN přechod u kolektorového konce kanálu kladné předpětí. Pro hodnoty $u_{GS} < U_P$ přechází tranzistor do **triódového režimu C**. Každá z charakteristik končí u hodnoty $u_{DS} = u_{GS} < 0$, při níž tranzistor přechází do režimu 0^- . **Režim D** představuje stav tranzistoru, kdy je celý kanál uzavřen a neteče jím žádný proud.

Soustava normovaných převodních charakteristik je nakreslena na obr. 6.9. Zmenšuje-li se záporné předpětí hradla z hodnoty $-\infty$ k nule při kladném napětí kolektoru, nachází se tranzistor JFET nejdříve v **režimu D^+** , kanál je po celé délce uzavřen a neprotéká jím proud. Jakmile záporné předpětí hradla klesne pod $-U_P$, přejde tranzistor do **aktivního saturačního režimu B** a velikost proudu se řídí vztahem (6.14) až do okamžiku, kdy přestává platit podmínka pro režim B: $u_{GS} > u_{DS} + U_P$. Pro $u_{GS} < u_{DS} + U_P$ pracuje tranzistor v **aktivním režimu A^+** a proud se řídí vztahem (6.12). Charakteristiky končí na ose $u_{GS} = 0$ V, kdy tranzistor přejde do **režimu 0^+** (kladné předpětí přechodu PN mezi hradlem a kanálem u jeho emitorového konce).

Při nulovém napětí kolektoru je proud tranzistorem v celém rozsahu pracovních napětí hradla nulový.

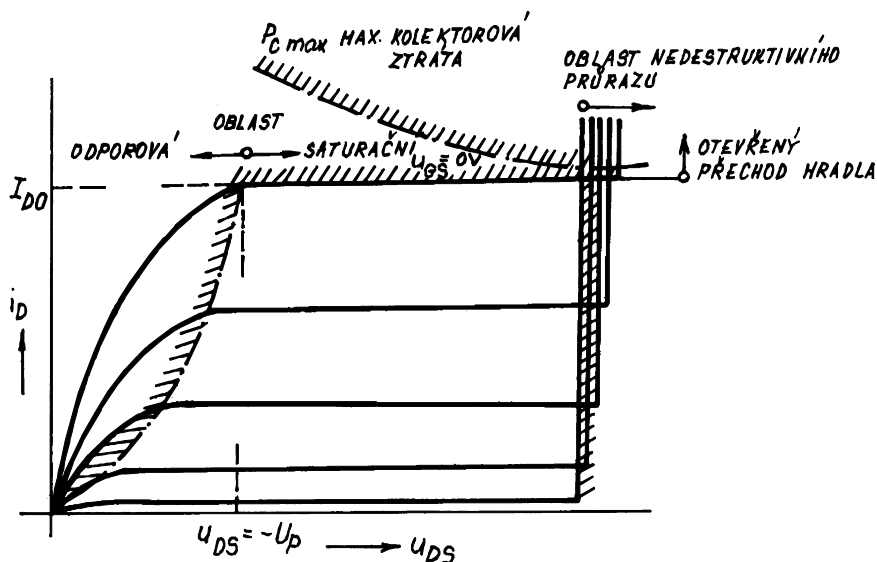
Při záporném napětí kolektoru pracuje tranzistor při snižování záporného předpětí hradla nejdříve v **režimu D⁻**, kanál je po celé délce uzavřen a neprotéká jím proud. Při dosažení stavu, kdy $u_{DS} = u_{GS} - U_P$ přejde tranzistor do **triodového režimu C**. Při dosažení podmínky $u_{GS} < U_P$ přejde tranzistor do **aktivního režimu A⁻**. Při dosažení podmínky $u_{GS} = u_{DS}$ a to jak v režimu C, tak i v režimu A⁻, dosáhne tranzistor **režimu 0⁻** (kladné předpětí přechodu PN mezi hradlem a kanálem u jeho kolektorového konce). Maximální záporný proud i_D je v režimu C vždy roven hodnotě $-i_{DSat}$, v režimu A⁻ je pro dané $|u_{DS}|$ roven stejné (ale záporné) hodnotě jako v režimu A⁺ při $u_{GS} = 0$ V.

Podobně jako u bipolárního tranzistoru i u tranzistoru JFET uvažujeme tři základní zapojení: SS, SG, SD, nejužívanější je opět zapojení se společným emitorem, tedy SS. V zásadě je zřejmé, že u tranzistoru JFET můžeme řídit velký kolektorový proud napětím hradla. Je tedy s vhodným zatěžovacím rezistorem v kolektoru použitelný jako zesilovač napětí (viz obr. 6.30). V odporové oblasti jej lze použít jako proměnný rezistor. Lze jej použít i jako spínač. Teplotní závislost je menší než u bipolárního tranzistoru. Základní výhodou je velký vstupní odpor až $10^{11} \Omega$.

Dosud jsme uvažovali tranzistor JFET s kanálem N. Uvažujeme-li JFET s kanálem P, je v jeho symbolu šipka v opačném směru. Dále jsou opačné polaritty všech proudů a napětí, tj. u_{DS} je záporné, u_{GS} je kladné a i_D vytéká z kolektoru D. V praxi se především z technologických důvodů prosadil tranzistor JFET s

kanálem N označovaný NJFET.

Poznámka: Všimněte si "umístění" oblasti saturace ve výstupních charakteristikách bipolárního a unipolárního tranzistoru.



Obr. 6.8.: Obvyklý tvar výstupních charakteristik tranzistoru JFET

Šířka kanálu, zde značená Z se v literatuře též často označuje symbolem W (width).

6.1.4 lineární malosignálový model tranzistoru JFET

Charakteristické veličiny pro sestavení náhradního obvodu tranzistoru JFET je možné stanovit ze vztahů pro ampérovoltové závislosti.

Kolektorová (výstupní) vodivost je definovaná jako

$$g_d = \left. \frac{\partial i_D}{\partial u_{DS}} \right|_{u_{GS} = \text{konst.}} \quad (6.16)$$

Strmost hradla, často nazývaná také *přenosová vodivost* [transconductance, mutual conductance], závisí na poloze pracovního bodu a vzrůstá při $u_{GS} \rightarrow 0$. Dosahuje hodnot až 3,5 mA/V. Je definovaná jako

$$g_m = \left. \frac{\partial i_D}{\partial u_{GS}} \right|_{u_{DS} = \text{konst.}} \quad (6.17)$$

Poslední veličinou, která bývá někdy v malosignálovém modelu používána, je **napětový zesilovací činitel**

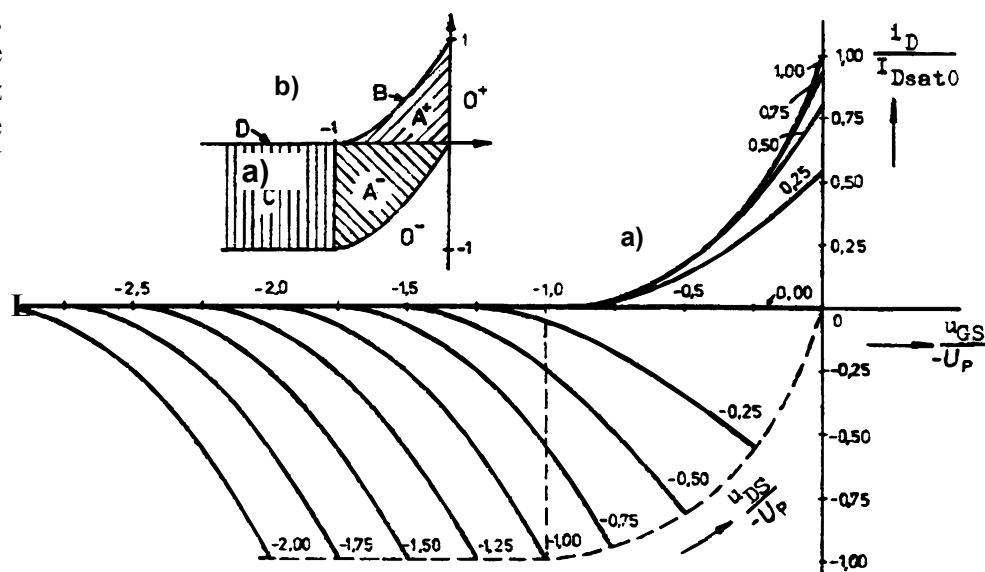
$$\mu_g = - \left. \frac{\partial u_{DS}}{\partial u_{GS}} \right|_{i_D = \text{konst.}} = \frac{g_m}{g_d} \quad (6.18)$$

Výrazy pro jednotlivé parametry můžeme odvodit z rovnic pro proud i_D (6.12) a (6.14). Přehledně jsou shrnuty v tabulce 6.1.

Použijeme-li při odvození strmosti hradla v režimu B jednodušší rovnici (6.15) místo (6.14), dostaneme v praxi velmi užitečný výraz

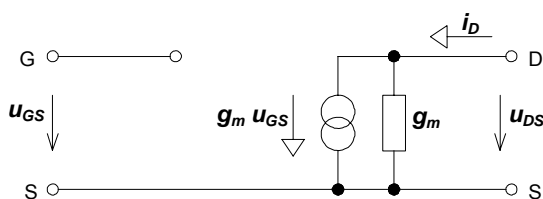
$$g_m = \frac{2I_{Dsat0}}{U_P} \left(\frac{u_{GS}}{u_P} - 1 \right) \quad (6.19)$$

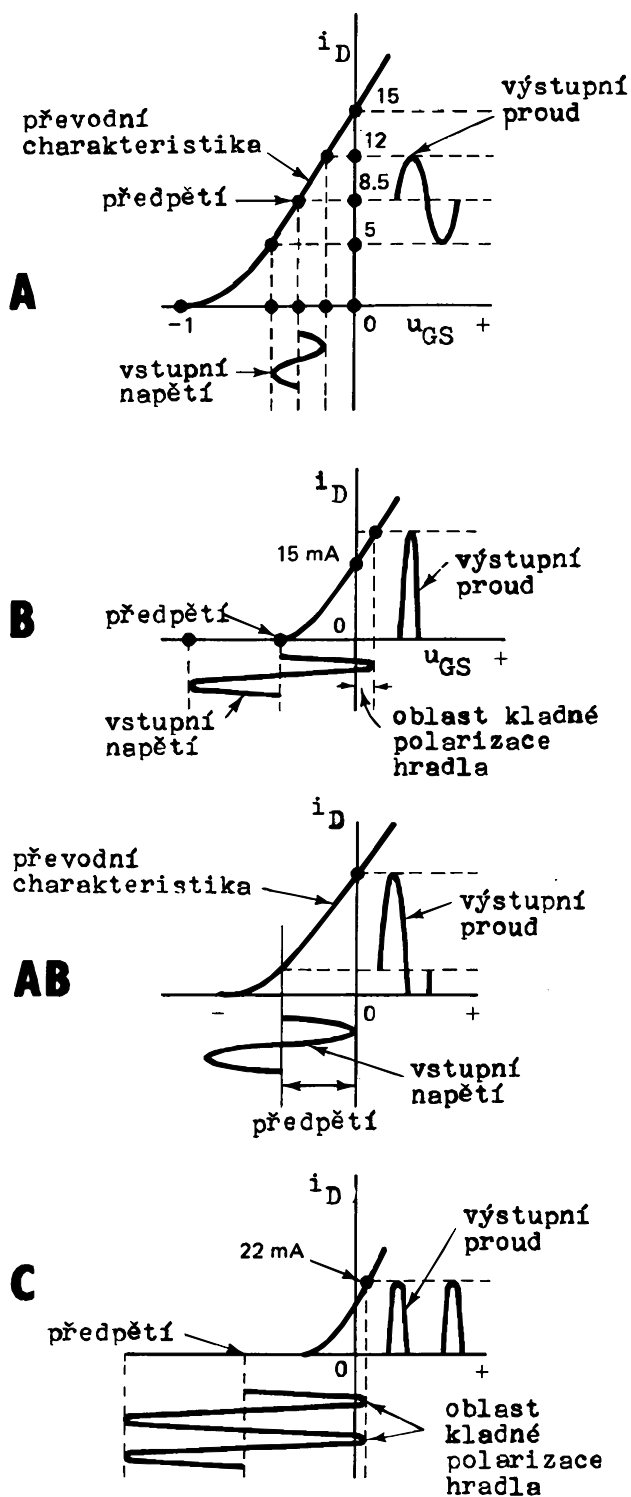
Nejjednodušší lineární náhradní



Obr. 6.10. a) Normované převodní charakteristiky a b) pracovní oblasti tranzistoru JFET

r. 6.11. Linearizovaný malosignálový náhradní obvod tranzistoru JFET pro nízké frekvence





6.12.: Nastavení pracovního bodu zesilovacích stupňů s tranzistorem JFET (třída A, AB, B a C)

nám také umožňují vysvětlit, proč lze tranzistory JFET (a i MOSFET) při větším výkonovém zatížení řídit paralelně, a to v libovolném množství. Vzroste-li v jednom systému jeho vnitřní teplota, klesne proud a tím jeho vlastní výkonová ztráta, takže nemůže dojít (jako u bipolárních tranzistorů) ke kumulativnímu nárůstu proudu a výkonové ztráty (a tím k samozničení).

obvod tranzistoru JFET pro nízké frekvence je na obr. 6.11. Všimněme si, že strmost hradla g_m má v tomto modelu podobný význam jako parametr β v modelu bipolárního tranzistoru.

Kmitočtové vlastnosti tranzistoru JFET při vyšších frekvencích jsou ovlivněny skutečností, že při změně potenciálu hradla se musí kapacita hradla nabít nebo vybit přes odpor kanálu. Proto je i náhradní obvod tranzistoru o něco složitější.

O teplotní závislosti FETů nebývá v literatuře mnoho bližšího a jednoznačného uváděno. Pouze u některých JFETů se dá zjistit teplotní závislost převodních charakteristik podle obr. 6.13. Na těchto charakteristikách je pozoruhodné, že se charakteristiky pro různé teploty křížují na velice malé ploše, což umožňuje (z hlediska minimálních teplotních změn pracovního proudu) nalézt optimální nastavení pracovního bodu. Stačí tedy nalézt optimální hodnotu emitorového rezistoru R_S ($\equiv R_E$) k zajištění stálého, téměř teplotně nezávislého, pracovního proudu (R_{Sopt}). Při větší hodnotě klidového proudu než je optimální, zřejmě podle obr. 6.13 proud s rostoucí teplotou klesá a při menší hodnotě proud s rostoucí teplotou stoupá. Z tohoto důvodu se JFETy používají jako proudové omezovače pro napájení referenčních diod nebo buzení proudových zrcadel, kde se dá nastavit klidový proud I_0 tak, aby v širokém teplotním rozsahu byla výstupní klidová veličina (napětí nebo proud) konstantní. Z obr. 6.13 je patrné, že se tato optimalizace nevztahuje na strmost převodní charakteristiky, která v kterémkoliv bodě charakteristiky při rostoucí teplotě klesá. Proto bude i zesílení zesilovače s JFETem s rostoucí teplotou klesat.

Teplotní změny charakteristik z obr. 6.13

Tabulka 6.1.: Malosignálové parametry tranzistoru JFET. Strmost hradla pro $u_{GS} = 0$ V je $G_0 = \frac{2Zq\mu_n N_D a}{L}$

Poznámka: Sledujeme-li řez tranzistoru JFET (obr. 6.4), můžeme si povšimnout toho, že struktura JFETu je nápadně podobná struktuře bipolárního planárního tranzistoru. To umožňuje vyrábět tyto tranzistory jednou společnou "bipolární" technologií a proto se vyrábí monolitické integrované obvody označované BiFET, ve kterých se plně dají využít výhody obou typů součástí.

režim A ($u_{DS} \leq u_{DSsat}$)	režim B ($u_{DS} \geq u_{DSsat}$)
$g_d = G_0 \left[1 - \left(\frac{u_{DS} - u_{GS}}{-U_P} \right)^{1/2} \right]$	$g_d = 0$
$g_m = G_0 \left[\left(\frac{u_{DS} - u_{GS}}{-U_P} \right)^{1/2} - \left(\frac{-u_{GS}}{-U_P} \right)^{1/2} \right]$	$g_m = G_0 \left[1 - \left(\frac{-u_{GS}}{-U_P} \right)^{1/2} \right]$

Obr. 6.13. Teplotní závislost tranzistorů JFET

Obr. 6.14. Nastavení pracovního bodu zdrojem v řídící elektrodě

Obr. 6.15. Nastavení pracovního bodu úbytkem napětí na emitorovém rezistoru

Obr. 6.16. Nastavení pracovního bodu děličem napětí

Obr.6.17. Srovnání tří výše uvedených metod nastavení pracovního bodu při rozptylu převodních charakteristik (prahového napětí)

6.2 Unipolární tranzistor s izolovanou řídicí elektrodou

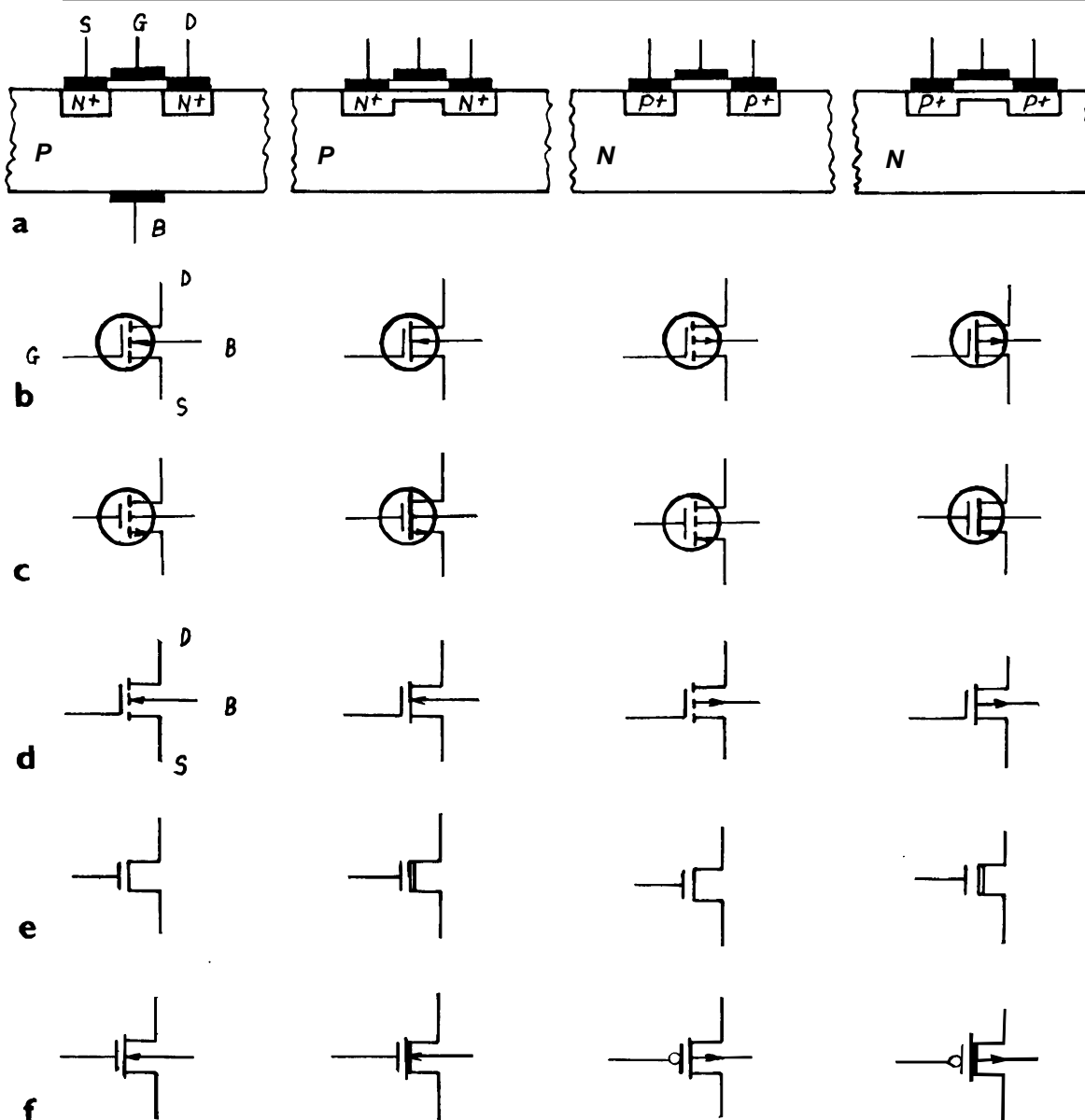
Tranzistory IGFET, jak již bylo řečeno, využívají pro svou funkci elektrické pole, které ovládá přes tenkou vrstvu izolantu vodivost polovodiče pod touto vrstvou. Bylo také řečeno, že se pro ně užívá zkratka MOS nebo MOSFET, především proto, že se lépe vyslovuje, této zkratky se budeme držet zde.

Tranzistor MOS existuje ve čtyřech variantách, jejich princip a schematické značky vidíme na obr. 6.18. Pod vlivem anglosaské literatury existuje řada pojmenování pro jednotlivé varianty tranzistoru MOS a jejich pracovní režimy. Vodivý kanál může být vodivosti N nebo P a buď existuje již při nulovém napětí na hradle (*tranzistor s trvalým kanálem, s automatickým otevřením, ochuzovací modifikace tranzistoru* [depletion type]) nebo vznikne až při určité velikosti kladného nebo záporného napětí na hradle (*tranzistor s indukovaným kanálem, s automatickým uzavřením, obohacovací modifikace tranzistoru* [enhancement type]), které nazýváme *prahovým napětím* a značíme u_P (nebo u_T , T = threshold = práh). Slovní i písmenové označení elektrod je shodné jako u tranzistoru JFET.

Dříve, než přistoupíme k vysvětlení činnosti tranzistoru MOSFET, musíme pochopit vlastnosti sendvičové struktury MOS.

6.2.1 Ideální struktura MOS

Ideální struktura MOS je vytvořena z kovové elektrody, z ideálního izolantu, přes který nemůže protékat žádný proud a který neobsahuje volné nosiče náboje, a z homogenně dotovaného polovodiče, opatřeného na spodní straně ohmickým kontaktem. V oxidu a na rozhraní oxid-polovodič nejsou žádná nabitá centra (povrchové stavy apod.).

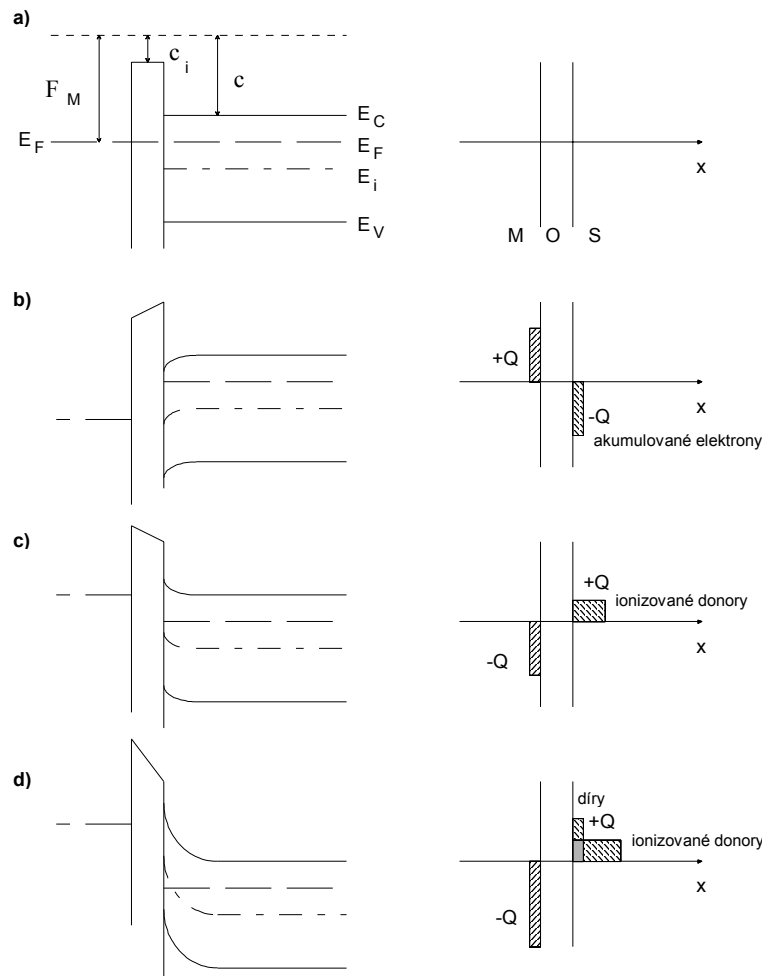


Průřez strukturou a schématické značky čtyř variant tranzistoru MOS (od leva): NMOS s indukovaným kanálem, NMOS s trvalým kanálem, PMOS s indukovaným kanálem, PMOS s trvalým kanálem

Energetický pásový diagram ideální MOS struktury se substrátem typu N v rovnovážném stavu vidíme na obr. 6.13a). Energetický rozdíl mezi úrovní vakua a Fermiho energií v kovu je výstupní práce elektronu z kovu, označovaná $q\Phi_M$, χ je elektronová afinita (energetický rozdíl mezi hladinou vakua a dnem vodivostního pásu v polovodiči). Izolant může být modelován jako intrinzický polovodič s velkou šířkou zakázaného pásu a elektronovou afinitou χ_i . V **rovnovážném stavu** ($u_G = 0$) není na strukturu přivedeno žádné napětí a Fermiho energie v kovu i v polovodiči leží ve stejné úrovni.

V tomto zidealizovaném případě předpokládáme, že výstupní práce elektronu z kovu i z polovodiče jsou si rovny a proto v rovnovážném stavu nedochází k zakřivení energetických pásů v polovodiči [flat band]. Ve skutečnosti vlivem rozdílných výstupních prací z kovu a z polovodiče dojde pod povrchem polovodiče k vytvoření akumulované, depleční nebo dokonce inverzní vrstvy a k tomu odpovídajícímu zakřivení energetických pásů. S tímto jevem jsme se již setkali při rozboru přechodu kov-polovodič.

Nyní předpokládejme, že na kovovou řídící elektrodu (hradlo G) přiložíme napětí u_G . Jestliže $u_G > 0$ (obr. 6.20.b), sníží se Fermiho hladina v kovu vzhledem k Fermiho hladině v polovodiči. To vede ke kladnému sklonu v zakřivení energetických pásů v izolantu i v polovodiči a ke vzniku **akumulované vrstvy** pod povrchem polovodiče, ve které je koncentrace elektronů vyšší než v substrátu (vyšší než N_D). Situaci si můžeme znázornit i z hlediska elektrického náboje. Přiložením kladného napětí na kovovou elektrodu v ní dojde k vytvoření kladného náboje $+Q$, který musí být kompenzován vznikem záporného náboje volných elektronů $-Q$ v polovodiči (elektrony jsou přitahovány z objemu polovodiče).



20. Energetické pásové diagramy ideální struktury MOS se substrátem typu N a odpovídající rozložení náboje pro různá napětí u_G :

- | | | |
|----|-----------|--------------------|
| a) | $u_G = 0$ | rovnovážný stav |
| b) | $u_G > 0$ | akumulovaná vrstva |
| c) | $u_G < 0$ | depleční vrstva |

Obr. 6.19. Princip uspořádání tranzistorů MOS: a) E-NMOSFET [tranzistor s izolovaným hradlem, kanálem N a obohacováním (enhancement)], b) D-PMOSFET [tranzistor s izolovaným hradlem, kanálem P a ochuzováním (depletion)]

Jestliže na hradlo přiložíme malé záporné napětí $u_G < 0$ (obr. 6.20.c), Fermiho hladina v kovu se zvýší vůči Fermiho hladině v polovodiči, zakřivení energetických pásů má záporný sklon a pod povrchem polovodiče vznikne **depletiční vrstva**. Záporný náboj kovového hradla je kompenzován kladným nábojem ionizovaných donorů (volné elektrony jsou odpuzeny do substrátu).

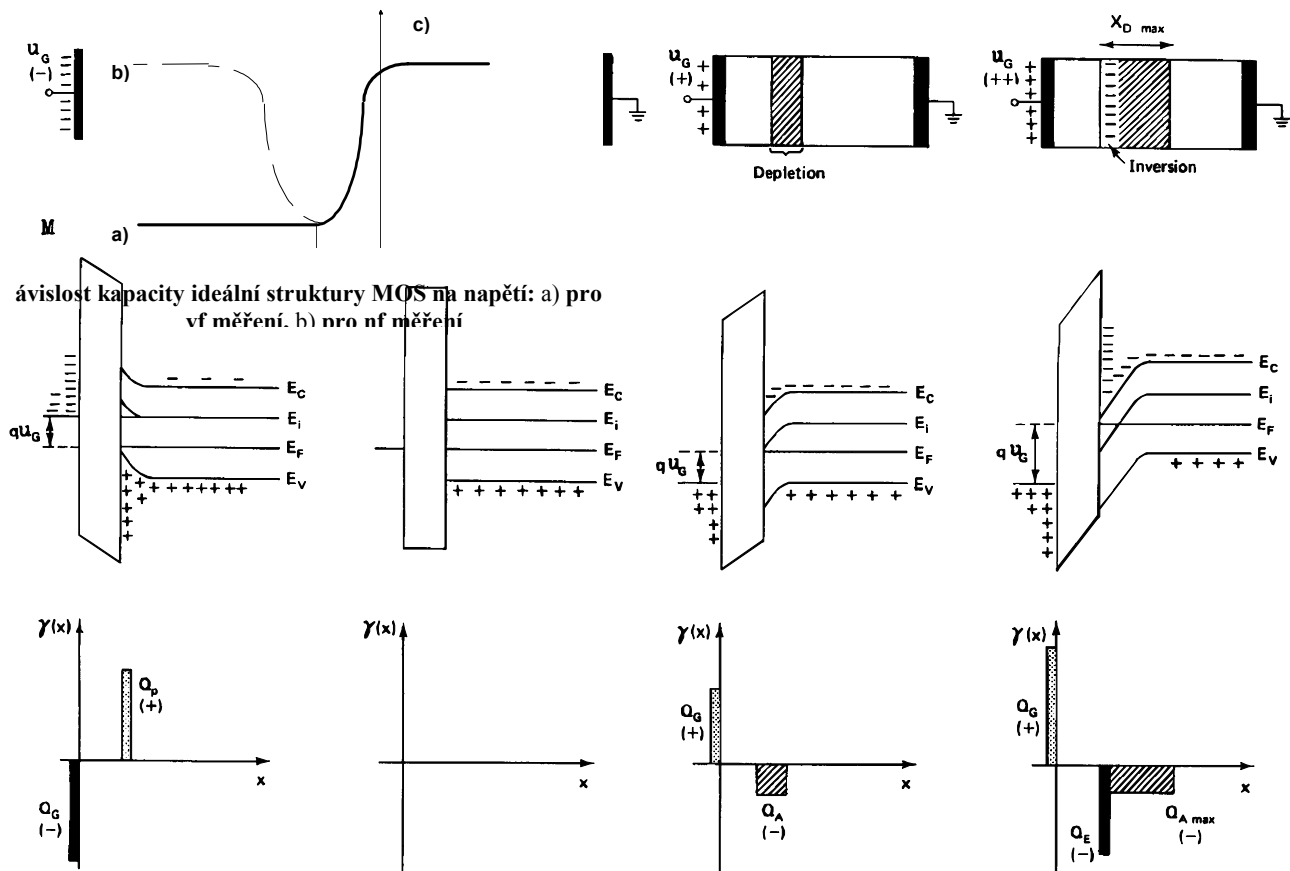
Při dalším zvyšování záporného napětí dojde při $u_G < U_P < 0$ (obr. 6.20.d) k tak velkému zakřivení energetických pásů, že pod povrchem polovodiče klesne Fermiho hladina pod střed zakázaného pásu a pod hradlem se začne vytvářet **inverzní vrstva** s opačným typem vodivosti (v našem případě P) než je substrát. Záporný náboj hradla je kompenzován kladným nábojem ionizovaných donorů a kladným nábojem pohyblivých děr, které jsou přitahovány ze substrátu.

Podobně můžeme popsat vznik akumulované ($u_G < 0$), depletiční ($u_G > 0$) a inverzní ($u_G > U_P > 0$) vrstvy v případě ideální MOS struktury se substrátem typu P.

Velmi důležitou vlastností je **kapacita struktury MOS**, která je tvořena sériovou kombinací kapacity polovodičové vrstvy pod rozhraním izolant-polovodič C_s a kapacitou izolační vrstvy C_i . Celková kapacita struktury MOS je

$$C = \frac{C_i C_s}{C_i + C_s} \quad (6.20)$$

Přivedeme-li mezi kovové hradlo a polovodičový substrát (typu N) napětí $u_G > 0$, vytvoří se pod hradlem akumulovaná vrstva a $C \approx C_i$. Při malém záporném napětí $u_G < 0$ se pod hradlem vytvoří depletiční vrstva s kapacitou C_s a celková kapacita je menší, daná vztahem (6.20). Při dalším zvyšování záporného napětí se při $u_G < U_P < 0$ vytváří inverzní vrstva, ale celková šířka depletiční vrstvy zůstává konstantní (viz obr. 6.20.d) a velikost celkové kapacity se nemění (plná čára na obr. 6.22). To však platí pouze pro vysoké hodnoty měřicího kmitočtu. Při dostatečně nízkém měřicím kmitočtu se během jedné periody ustaví rovnováha mezi koncentrací elektronů v inverzní oblasti a střídavým elektrickým polem a nemůže se uplatnit vliv depletiční vrstvy. Celková kapacita je tedy opět daná pouze kapacitou izolační vrstvy $C \approx C_i$ (čárkovaná čára na obr. 6.22.).



Obr. 6.21. Energetické pásové diagramy ideální struktury MOS se substrátem typu P

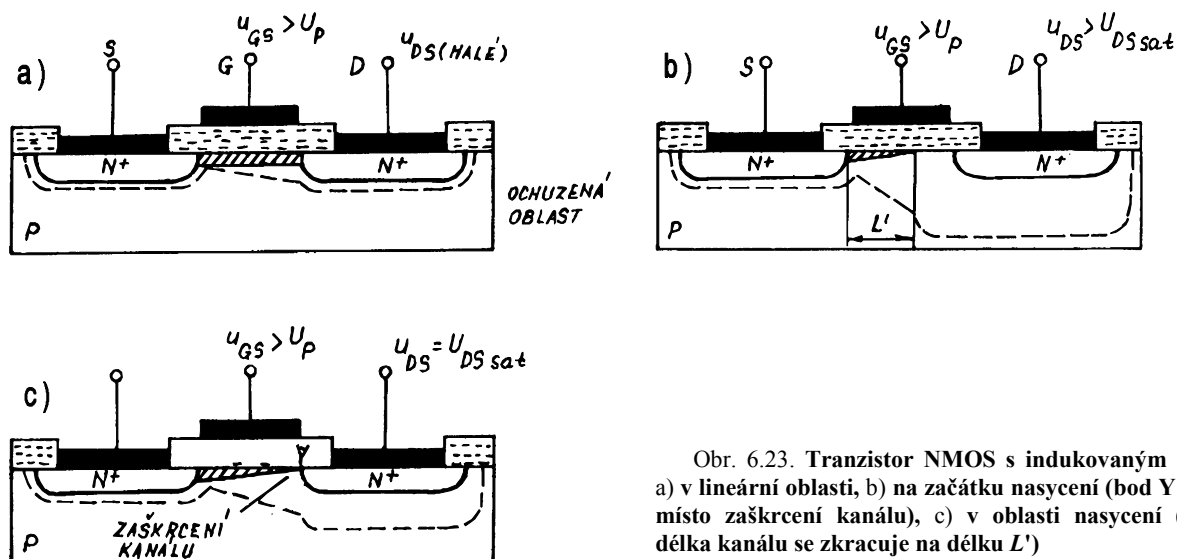
6.2.2 Kvalitativní popis činnosti tranzistoru MOS

Uvažujeme strukturu podle obr. 6.23. Pokud je mezi kolektor a emitor připojeno napětí $u_{DS} \neq 0$ a napětí $u_{GS} = 0$, proud mezi D a S bude velmi malý, řádově stovky nA (uvažujeme ideální případ bez povrchových stavů). Jeho velikost bude dána pouze měrným odporem ϱ základního materiálu - substrátu (zde Si o vodivosti P). Při přiložení kladného napětí u_{GS} na řídicí elektrodu G se pod rozhraním dielektrikum-polovodič začíná vytvářet depletiční a při $u_{GS} > U_P$ inverzní vrstva s vodivostí typu N. Při $u_{DS} \neq 0$ začne kanálem protékat proud i_D . Kanál je tím vodivější a i_D tím větší, čím větší je kladný potenciál řídicí elektrody. Podobně jako u tranzistoru JFET, i u tohoto typu tranzistoru pro $u_{DS} > U_{DSsat}$, kde $U_{DSsat} = f(u_{GS}) = u_{GS} - U_P$, přestane proud i_D narůstat.

Při zvyšování u_{DS} totiž zároveň klesá rozdíl potenciálů mezi D a G a zužuje se kanál. Při nulovém rozdílu dojde na straně kolektoru k uzavření kanálu a k saturaci proudu. Tj. procházející proud způsobí úbytek napětí "podél" kanálu od D k S, proto D je kladnější vůči S. Tento kladný úbytek napětí působí proti napětí hradla tak, že se kanál ve směru od S k D zúží a při dostatečně velkém i_D (tj. u_{DS}) se u D uzavře ("zaškrtní"), procházející proud ovšem nezanikne a zůstává konstantní při dalším zvyšování u_{DS} .

Popsaná struktura, shodná s obr. 6.18a), je tranzistor MOS s indukovaným kanálem N. Název pochází od toho, že kanál (inverzní vrstva N) se indukuje vlivem elektrického pole.

Charakteristická je ta vlastnost, že při nulovém napětí na řídicí elektrodě G neprotéká tranzistorem proud. Zvolíme-li substrát s vodivostí N a oblast emitoru a kolektoru P, dostaneme tranzistor MOS s indukovaným kanálem P (obr. 6.18c), který vyžaduje obrácenou polaritu obvodových veličin.



Obr. 6.23. Tranzistor NMOS s indukovaným kanálem: a) v lineární oblasti, b) na začátku nasycení (bod Y označuje místo zaškrčení kanálu), c) v oblasti nasycení (efektivní délka kanálu se zkracuje na délku L')

Jestliže zabudujeme (difúzí nebo iontovou implantací) mezi oblast emitoru a kolektoru vodivý kanál, bude tranzistorem protékat proud i při $u_{GS} = 0$. V tomto případě dostáváme tranzistor MOS s trvalým kanálem N nebo P. I u tohoto tranzistoru je možno definovat prahové napětí U_P (napětí, při kterém dojde k zániku kolektorového proudu).

Tranzistor ochuzovacího typu (s trvalým kanálem) může pracovat jenom v ochuzovacím režimu (depletion mode), kdežto tranzistor obohacovacího typu (s indukovaným kanálem) může pracovat v ochuzovacím režimu i v obohacovacím režimu (enhancement mode). (Dříve zmíněné tranzistory JFET pracují jen v režimu ochuzování kanálu).

Přehledné srovnání tranzistorů JFET a MOSFET a jejich typů je na obr. 6.24.

Obtížně realizovatelným typem tranzistoru MOS byl typ s indukovaným kanálem N, a to proto, že několik fyzikálních mechanismů vytváří automaticky vodivý kanál. V SiO_2 se nacházejí sodíkové atomy, které jsou při běžných teplotách ionizovány s kladným nábojem. Dále na rozhraní oxid-polovodič v důsledku ne zcela stejné struktury vznikají kladné náboje právě tak, jako mezi kovovou elektrodou a oxidem. Tyto tři základní nehomogenity ve struktuře, projevující se vznikem kladného náboje, způsobí přitažení elektronů bez přiloženého napětí, a tak vniká samovolně vodivý kanál. Nepříjemné je i to, že množství nábojů se časem mění - odtud pramení časová nestálost prahového napětí. Technologie se musí s těmito vlivy vypořádat.

V integrovaných obvodech se hradlo vyrábí nejen kovové (Al), ale i z polykrystalického křemíku.

Principiálně můžeme konstatovat, že tranzistor MOS může pracovat ve třech různých zapojeních: SS, SG, a SD. Protože jde o symetrickou součástku, lze elektrody S a G zaměnit. Přitom má smysl rozlišovat zapojení SS a SD, např. zesilovač a emitorový sledovač. Tranzistory s indukovaným kanálem mají velké použití v logických obvodech (nyní pouze v integrované podobě a jako NMOS a CMOS [complementar MOS]).

Tranzistory MOS mají obdobné vlastnosti jako tranzistory JFET. MOS mají větší hodnotu vstupního odporu (vstupní proud nabíjí parazitní kapacitor, kdežto u tranzistoru JFET je to proud závěrně polarizovaného přechodu PN). Vstupní kapacita bývá menší. Tranzistory MOS mívají v nízkofrekvenční oblasti větší šum než JFET. Hlavní použití tranzistorů MOS je v integrovaných obvodech.

Na obr. 6.18 vidíme u tranzistoru ještě čtvrtý vývod - substrát B [bulk]. Obvykle jej spojujeme s emitorem. Pokud je vyveden z pouzdra, bývá označen jako druhé hradlo G_2 a lze jej používat k řízení vodivosti kanálu, ovšem s asi řádově menší citlivostí (strmostí) než z hradla G_1 .

Může být také připojen na pevné předpětí. (U tranzistorů JFET byl substrát připojen na hradlo).

Existují také speciální tranzistory MOS se dvěma rovnocennými hradly, které používáme ke konstrukci řízených zesilovačů, modulátorů, směšovačů apod. Nazývají se dvoubázové tranzistory MOS [dual - gate MOS] nebo tetrody MOS (viz později, kap 6.3.1).

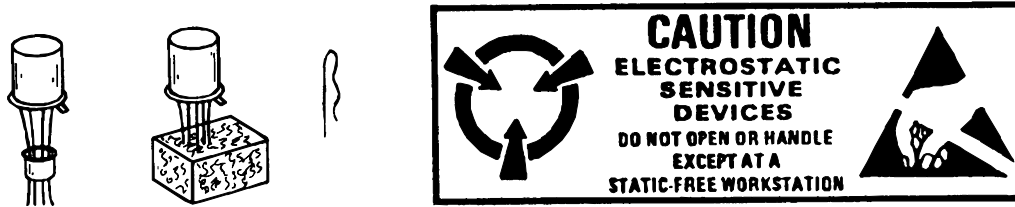
Je patrné, že všechny druhy FETů mají uvnitř zabudovány přechody PN (obr. 6.4 a 6.19). Jen u JFETů je přechod PN využit k řízení tloušťky kanálu elektrickým polem, u ostatních nejsou přechody k vlastní funkci využity. Aby

	SCHEMATICKÁ ZNAČKA	KANÁL	VÝSTUPNÍ CHARAKTERISTIKY	i_D u_{DS}	PŘEVODNÍ CHARAKTERISTIKY	u_{GS}
PŘECHODOVÝ FET		N		> 0		< 0
		P		< 0		> 0
MOSFET S TRVALÝM KANÁLEM		N		> 0		≥ 0
		P		< 0		≤ 0
MOSFET S INDUKOVANÝM KANÁLEM		N		> 0		> 0
		P		< 0		< 0

Obr. 6.24. Výstupní a převodní charakteristiky tranzistorů JFET a MOSFET

nebyla narušena správná činnost (tj. řízení tloušťky kanálu elektrickým polem), musejí být všechny přechody PN v nevodivém stavu. (Z tohoto důvodu je mezní napětí JFETů $u_{GS} = 0$ V. Při tomto napětí je kanál plně vodivý. Je-li zvětšováno závěrné napětí hradla, je kanál vzniklým elektrickým polem zaškrcován - rozšiřuje se vyprázdněná vrstva přechodu PN.)

Při velkém zvýšení napětí u_{DS} může dojít k průrazu kanálu lavinovou ionizací. Velmi vážné je nebezpečí elektrického průrazu dielektrické vrstvy pod hradlem (její tloušťka se pohybuje kolem 500 nm) při manipulaci, a to statickou elektřinou (např. dotykem ruky). Proto výrobci



Obr. 6.25. Elektrostaticky citlivé součástky vyžadují speciální zacházení (zleva):

- vývody bývají zkratovány kroužkem nebo pružinkou, může být použita vodivá pryž nebo vhodná porézní hmota;
- štítek označující elektrostaticky citlivé součástky (kombinace černé a žluté barvy)

opatřují elektrody zkratovací pružinkou, která se odstraňuje až po montáži. Vícevývodové součástky přechováváme ve vodivé pryži, v hliníkové fólii nebo ve speciálních tyčových zásobnících (obr. 6.25). Praktici někdy na součástku před manipulací dýchnou. Dále výrobci do přívodů hradel zapojují ochranné rezistory a Zenerovy diody. Hovoříme o tzv. chráněných či "protektorovaných" [protected] tranzistorech (integrováných obvodech).

6.2.3 Statické charakteristiky tranzistoru MOS

Podobně jako u tranzistoru JFET můžeme i u tranzistoru MOSFET odvodit vztahy pro výpočet velikosti kolektrového proudu v jednotlivých režimech činnosti. Výsledky, které zde uvedeme, platí pro MOS tranzistor s indukovaným kanálem typu N. Pro **aktivní režim A** dostaneme

$$\frac{i_D}{I_{Dsat0}} = 2 \frac{u_{DS}}{U_P} \left(\frac{u_{GS}}{U_P} - 1 \right) - \left(\frac{u_{DS}}{U_P} \right)^2 \quad (6.21)$$

kde I_{Dsat0} je podobně jako u tranzistoru JFET saturační proud pro $U_{GS} = 0$ V

$$I_{Dsat0} = \frac{Z\mu_n C_i U_P^2}{2L} \quad (6.22)$$

Význam veličin je stejný jako u JFET tranzistoru, tj. L je délka kanálu a Z je jeho šířka. C_i je kapacita izolantu na jednotku plochy.

Tranzistor se v této oblasti chová jako nelineární rezistor řízený napětím u_{GS} . V tabulce 6.2 je vztah pro diferenciální vodivost v oblasti v okolí počátku charakteristiky. Poznamenejme, že vztah pro i_D i pro G_{DS} platí pro obojí polaritu napětí u_{DS} a proudu i_D . Zajímavou veličinou je odpor sepnutého kanálu $r_{DS(on)}$, u nízkovýkonových tranzistorů FET bývá v rozmezí od 30 Ω do 1 k Ω , tedy více než v bipolárního tranzistoru. Je také třeba počítat se značným výrobním rozptylem tohoto odporu (rozsah až 1 : 3).

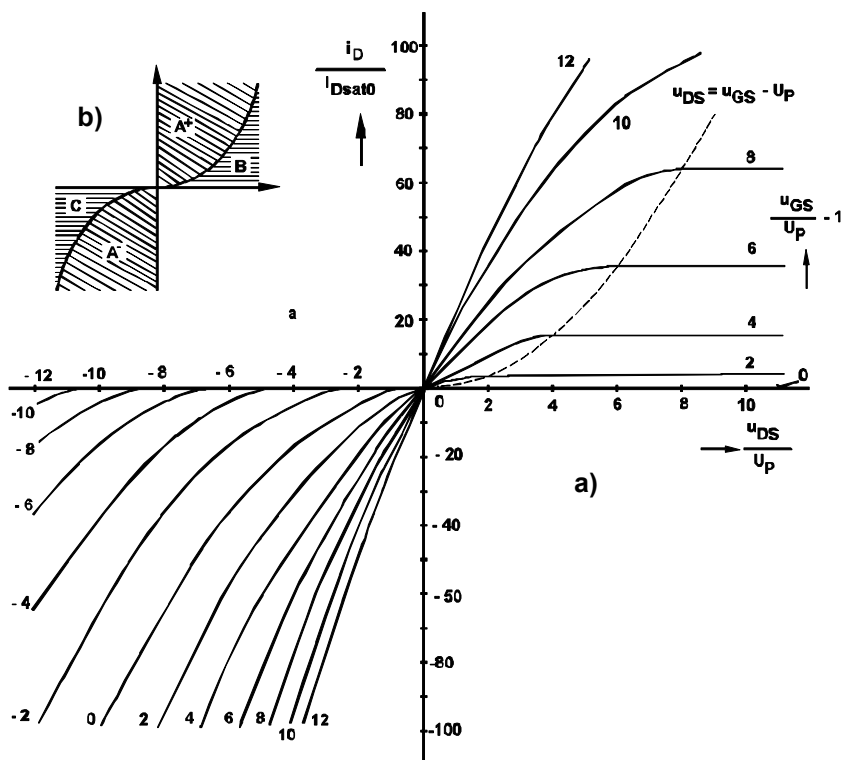
Hranicí mezi aktivním a **saturačním režimem B** je podobně jako u tranzistoru JFET křivka

$$u_{DS} = U_{DSsat} = u_{GS} - U_P \quad (6.23)$$

a rovnice pro proud I_D má v saturačním režimu tvar

$$i_D = I_{Dsat0} \left(\frac{u_{GS}}{U_P} - 1 \right)^2 \quad (6.24)$$

Tato rovnice je shodná se zjednodušenou rovnicí (6.15), která platí pro tranzistor JFET v režimu B.



Obr. 6.26. a) Normované výstupní charakteristiky a b) pracovní oblasti tranzistoru MOS

MOS můžeme definovat obdobně jako u tranzistoru JFET. Nenajdeme zde však nulový režim (kladné předpětí PN přechodu mezi hradlem a kanálem), který výrazně omezoval pracovní oblast tranzistoru JFET. Výstupní a převodní charakteristiky tranzistoru MOS jsou znázorněny na obr. 6.26 a 6.27.

6.2.4 Linearizovaný malosignálový model tranzistoru MOS

Nejjednodušší náhradní obvod tranzistoru MOS je shodný s náhradním obvodem tranzistoru JFET, popsaným v odstavci 6.1.4. Shodné jsou i definice jeho parametrů. Výrazy pro jednotlivé parametry můžeme odvodit z rovnic pro proud i_D (6.21) a (6.24). Přehledně jsou shrnuty v tabulce 6.2.

Strmosti hradla v režimu B je shodná s výrazem (6.19).

6.2.5 Nastavení pracovního bodu tranzistoru MOS

V saturační oblasti se výstupní proud se změnou napětí u_{DS} příliš nemění. Tato oblast bývá obvykle využívána pro zesilování signálů.

Naznačené výstupní charakteristiky s výraznou saturační oblastí jsou typické pro všechny FETy. Jen průběhy charakteristik některých spínacích FETů se od tohoto pravidla výrazněji odchyľují vzhledem k tomu, že vodivost kanálu i v očekávané saturační oblasti je značně ovlivněna napětím u_{DS} . Takové tranzistory nejsou vhodné k použití ani v zesilovačích ani jako řízené rezistory. (Z domácích tranzistorů to byl spínací JFET KS 4391).

Režimy činnosti tranzistoru

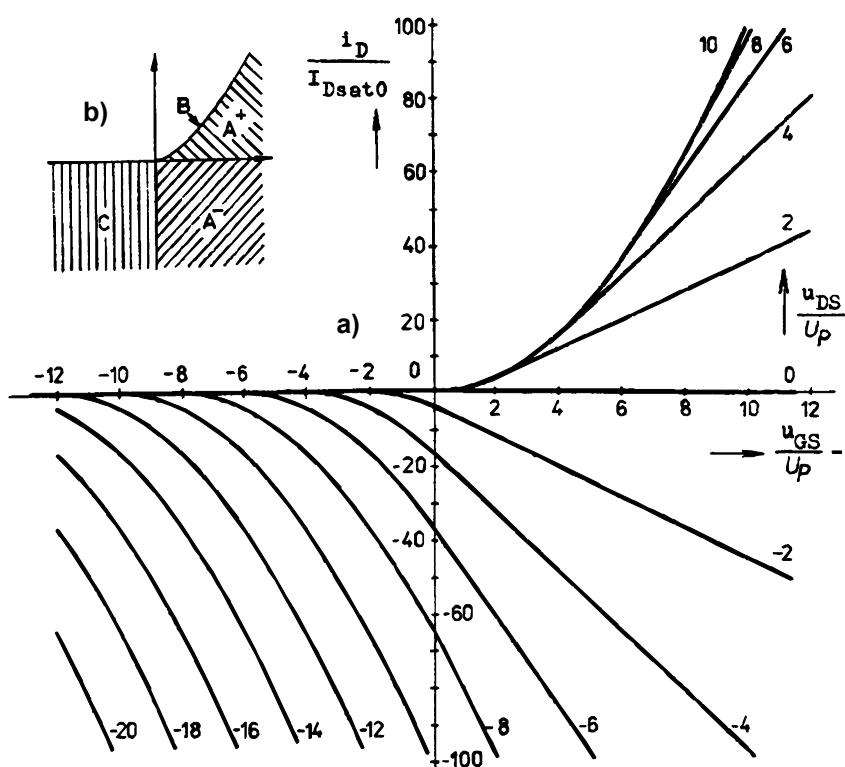
režim A ($u_{DS} \leq U_{DSSat}$)	režim B ($u_{DS} \geq U_{DSSat}$)
$g_d = G_0 \left[\left(\frac{u_{GS}}{U_P} - 1 \right) - \frac{u_{DS}}{U_P} \right]$	$g_d = 0$
$g_m = G_0 \frac{u_{DS}}{U_P}$	$g_m = G_0 \left(\frac{u_{GS}}{U_P} - 1 \right)$

Tabulka 6.2. Malosignálové parametry tranzistoru MOS. Strmost hradla

U tranzistoru MOS s vodivým kanálem je možné pracovat bez předpětí hradla. V praxi obvykle připojujeme rezistor mezi hradlo a zem a připojujeme také emitorový rezistor (který zavádí záporné předpětí). Pracovní bod se tedy nastavuje podobně jako u elektronek. U tranzistorů s indukovaným kanálem je nutné nastavit pracovní bod přiložením napětí většího než je prahové napětí (tj. uvést kanál do vodivého stavu). Zapojení je obdobou můstkového nastavení pracovního bodu bipolárního tranzistoru.

Abychom mohli spolehlivě nastavit pracovní bod obvodů s FETy (obr. 6.30), musíme znát průběh výstupních charakteristik těchto tranzistorů. Hodnotu R_S , případně R_D (není-li dána), pak zjišťujeme graficko-početním řešením. Nejjednodušší postup řešení můžeme uplatnit, je-li dáno u_{DS} , i_D , U_n . Potom musí platit

$$R_D + R_S = (U_n - U_{DS})/I_D$$



Pro dané U_{DS} a I_D odečteme z výstupních charakteristik potřebné předpětí hradla U_{GS} . Pro hodnotu R_S v zapojení a) pak platí

$$R_S = |U_{GS}| / I_D$$

V zapojení podle obr. 6.30b) závisí hodnota R_S i na napětí na děliči R_{G1}/R_{G2} . Označíme-li toto

a) Normované převodní charakteristiky a b) pracovní oblasti tranzistoru MOS

napětí U_{de} , musí pro tento obvod platit

$$U_{de} = U_{GS} + I_D R_S.$$

V obou případech se však dá předpokládat (na rozdíl od obvodu s bipolárním tranzistorem), že stabilizační účinek záporné zpětné vazby na R_S bude poměrně slabý. Musíme tedy očekávat poměrně velký rozptyl pracovních bodů při výměně tranzistorů i větší drift pracovního proudu s teplotou.

Obr. 6.28. Převodní a výstupní charakteristika tranzistorů MOS s indukovaným kanálem

Stav "rozepnuto" v obvodech podle obr. 6.30b) snadno navodíme tím, že v obvodu vynecháme rezistor R_{G1} .

V obvodu podle obr. a) bychom museli přes R_G přivést na hradlo záporné předpětí U_G , převyšující prahové předpětí tranzistoru. Pro uvedený typ stačí např. $U_G = -4$ V. Není-li k dispozici zdroj pomocného předpětí, lze tranzistor zavřít úpravou napájecího obvodu podle obr. 6.30c. Hodnota R_p musí být zvolena tak, aby na R_S vznikl i při uzavření tranzistoru potřebný úbytek napětí. Při obvodových aplikacích zjišťujeme, že některé typy MOSFETů mají vyveden substrát (např. KF521, KF522, MH2009), ostatní mají jen 3 vývody, takže se dá očekávat, že mají substrát spojen s emitorem. Třívývodové FETy bývají určeny především pro aplikace v zesilovačích, případně jako spínače napětí jedné polarity. Je-li substrát vyveden, dají se tyto MOSFETy použít pro spínání napětí obojí polarity, bude-li substrát stejnosměrně tak předeprnut, aby se nemohly otevřít přechody PN mezi emitorem a substrátem a mezi kolektorem a substrátem. Pro takováto použití není vhodný např. typ KF521, protože se u něho vyskytuje parazitní, celkem nedefinovatelný, poměrně malý odpor substrátu proti kanálu. Moderní MOSFETy s velkou strmostí mají kanál velmi krátký, takže vzdálenost mezi emitorem a kolektorem je menší než střední volná dráha minoritních nosičů náboje. Následkem toho se otevřením jednoho přechodu může uplatnit tranzistorový jev, přičemž se část substrátu mezi emitorem a kolektorem chová jaké báze laterálního tranzistoru (zesilovací činitel takového parazitního bipolárního tranzistoru se pohybuje v rozmezí $h_{21e} \approx 2$ až 5).

Obr. 6.28. Nastavení pracovního bodu tranzistoru MOS: a) pro tranzistor s trvalým kanálem a všechny tranzistory JFET, b) pro tranzistor s indukovaným kanálem, c) úprava pro stav rozepruto pro tranzistor s trvalým kanálem

Jak již bylo uvedeno, všechny tranzistory MOS mají řídicí elektrodu od polovodičové destičky odizolovanou velmi tenkou vrstvou SiO_2 . Tato vrstva musí být tím tenčí, čím účinněji má být vodivost kanálu řízena. Účinnost řízení je charakterizována strmostí S tranzistoru v zesilovacím režimu. Proto starší typy MOSFETů, které snášely řídicí napětí do 70 až 100 V a tudíž byly poměrně odolné proti průrazu isolační vrstvy, měly strmost jen několik desetin milisiemensů. Moderní MOSFETy s velkou strmostí dosahují strmosti $S = 2 - 15 \text{ mS}$ a jejich isolační vrstva snáší napětí od 80 do 15 V. Proto vyžadují účinnou ochranu proti průrazu, který může nastat i přivedením velmi malého elektrostatického náboje (kapacita řídicí elektrody proti ostatním částem systému bývá u malých systémů pod 10 pF). Výborná izolace řídicí elektrody je považována za nenahraditelnou principiální přednost MOSFETů. Celkem běžné typy (např. KF522) mají svodové proudy v řádu 10^{-14} A a lze je použít pro elektrometrické účely. Řada typů MOSFETů je vyráběna pro účely, při kterých nejsou vyžadovány tak malé vstupní proudy. Pro zmenšení nebezpečí jejich destrukce výrobci do takového systému obvykle zabudují Zenerovy diody trvale připojené mezi vývod hradla a substrátu. Životnost takového FETu je zaručena, jestliže proud Zenerovou diodu při výboji nepřekročí dovolenou hodnotu.

6.2.6 Setrvačné vlastnosti tranzistorů MOS

Setrvačnost FETů záleží na době průchodu nosičů náboje kanálem a na téměř konstantních mezielektrodových kapacitách. Kapacita hradla proti kanálu je rovnoměrně podél kanálu rozložena, obvykle ji však nahrazujeme dvojicí soustředěných kapacit C_{GS} a C_{GD} , přičemž se výrobci snaží konstrukcí tranzistoru docílit co nejmenší hodnotu C_{GD} . Rozdělení těchto kapacit se při uzavírání kanálu poněkud mění, velikost změn však vesměs není podstatná. Tam, kde záleží na malé setrvačnosti FETů, jsou tyto konstruovány s co nejkratším kanálem (délka řádově kolem desítky mikrometrů). Nízkopříkonové FETy mají kapacity C_{GS} a C_{GD} mezi 3 a 15 pF. Výkonnost FETů se zvětšuje rozšiřováním kanálu (parametr Z), respektive paralelním řazením dílčích systémů. Je-li výkonová ztráta zvětšena např. 100 krát, jsou ve stejné míře zpravidla zvětšeny i kapacity.

Setrvačnost obvodů s FETy je obvykle určován právě působením mezielektrodových kapacit. Obvykle platí $C_{GS} > C_{GD} > C_{DS}$. Pokusíme se alespoň v prvním přiblížení stanovit vlivy na průběh výstupní odezvy jednoduchého obvodu s MOSFETem z obr. 6.31a). Chování FETu vystihneme detašovanými kapacitami a "odporovým FETem" podle obr. 6.32b, přičemž odporový FET budeme linearizovat pro 3 pracovní oblasti:

A - oblast, kdy FET je nevodivý, $u_1 \leq U_p$, kde U_p je prahové napětí hradla,

B - aktivní oblast, kde $U_p < u_1 < U_M$ (při U_M přechází FET do stavu sepnutí),

C - stav sepnutí $u_1 \geq U_M$, kdy diferenciální výstupní odpor FETu výrazně klesl na hodnotu R_S a při zvětšování u_1 se již mnoho nemění.

Obr. 6.31. Tranzistor MOS a jeho mezelektrodové kapacity

Pro funkci FETu bude rozhodující časový průběh u_1 na řídicí elektrodě. Pro odhad časového

Obr. 6.32. Modely kmitočtového chování ve třech pracovních oblastech
průběhu $u_1(t)$ v uvedených oblastech sestavíme pro tyto oblasti lineární modely na obr. 6.20-3. Budeme-li $R_G \approx R_D \gg R_S$, lze odhadovat

$$\tau_A \rightarrow R_i \cdot C_{GS}, \quad \tau_B \approx R_i [C_{GS} + C_{GD}(1 + SR_Z)], \quad \tau_C \rightarrow R_i (C_{GS} + C_{GD})$$

Bude tedy nejmenší τ_A , poněkud větší τ_C a několikanásobně větší τ_B . Odtud lze již odhadnout dílčí průběhy $u_1(t)$ na obr. 6.33 a z nich odvodit i průběh $u_2(t)$. Je zřejmé, že změny u_2 probíhají při aktivní činnosti tranzistoru a podléhají působení největší časové konstanty τ_B , která je vázána na Millerovu kapacitu úměrnou zesílení SR_Z . Bude-li zvětšováno zesílení např. zvětšením R_Z , prodlouží se intervaly B, ostatní přechodné děje však zůstanou nezměněny. Za povšimnutí stojí poměrně malá zpoždění začátků přeběhů t_1 a t_2 a to, že tyto doby závisejí na napětí U_p a $U_{iM} - U_M$. Dá se usoudit, že bude obtížné zajistit $t_1 = t_2$.

Kromě toho někdy nastává situace, kdy $R_G \ll R_D$, přičemž SR_D je nepříliš veliké. Potom probíhá zapínání i rozepínání FETu poměrně rychle a doba přeběhu výstupního napětí při vypnutí je značně ovlivněna časovou konstantou $R_D(C_{DS} + C_{GD})$ (na obr. 6.33 čerchovaně naznačené exponenciální doznívání $u_2(t)$). Zpoždění začátku přeběhů a vlastní doby přeběhu se u nízkopříkonových FETů pohybují od 5 do 100 ns, případně až do několika mikrosekund při velkých hodnotách R_D .

Obr. 6.33. Časové průběhy při sepnutí a vypnutí tranzistoru MOS

6.2.7 Model tranzistoru MOS v programu SPICE

Model tranzistoru MOS v programu SPICE je při použití tam zavedené symboliky určen v normálním pracovním režimu relacemi

$$i_D = \begin{cases} 0 & \text{pro } V_{GS} - V_{TO} \leq 0 \\ -\beta(V_{GS} - V_{TO})^2 \cdot (1 - \lambda V_{DS}) & \text{pro } 0 < V_{GS} - V_{TO} \leq V_{DS} \\ \beta V_{DS} [2(V_{GS} - V_{TO}) + V_{DS}] \cdot (1 - \lambda V_{DS}) & \text{pro } 0 < -V_{DS} < V_{GD} - V_{TO} \end{cases}$$

kde

$$\beta = \frac{I_{Dsat0}}{U_P^2}, \quad V_{TO} = U_P, \quad \lambda = (r_D \cdot I_{Dsat0})^{-1}$$

Parametry β , V_{TO} a případně i I_{Dsat0} se snadno určí změřením a zobrazením charakteristik podle obr. 6.34 a následujícími jednoduchými grafickými konstrukcemi.

Obr. 6.34. Odvození parametrů modelu tranzistoru MOS

6.3 Specializované typy unipolárních tranzistorů

6.3.1 Tranzistor MOS se dvěma hradly

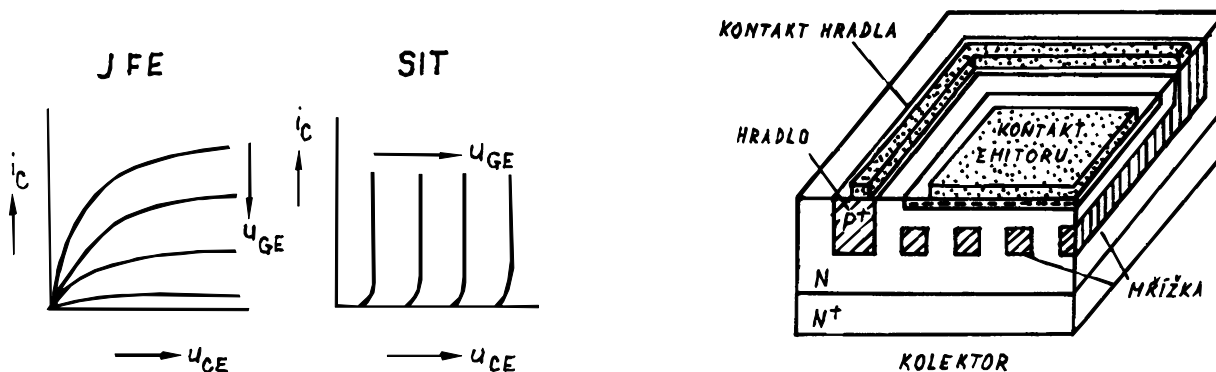
Od samotného vzniku tranzistorů hledají konstruktéři takovou konstrukci tranzistoru, která by zajistila nepatrnou průchozí kapacitu mezi bází a kolektorem. U pentod byla tato kapacita snížena až na 0,01 pF, zatímco stíněním báze v bipolárních tranzistorech nešlo tuto kapacitu zmešit pod 0,3 pF. Jakmile byl nalezen způsob výroby MOSFETů s malými kapacitami, vysokým mezním kmitočtem a dostatečně velkou strmostí ($S > 5 \text{ mS}$), vznikly opět pokusy odstínit kolektor od hradla. Vznikl tak tranzistor MOS se dvěma hradly (double-gate FET), jehož schématická značka (na obr. 6.35a) napovídá, že se jedná o typ s vodivým kanálem vodivosti N. Na obr. 6.35b) je naznačen řez jeho strukturou, ze kterého je patrné, že se nejedná o tranzistorovou "tetrodu", ale o kaskódové spojení dvou jednoduchých systémů MOSFET, jak naznačuje obrázek c). Na obr. 6.35d) jsou uvedeny průběhy výstupních charakteristik změřených na tranzistoru KF907. Je zde patrné, že řízení prvním hradlem je účinné pouze v ochuzovaném režimu. Čerchované křivky zde naznačují hranice linearizovatelné oblasti. Pouze při $U_{G2S} = 0 \text{ V}$ je tato hranice normální jako u jednoduchých FETů. Při kladném předpětí druhého hradla se tato hranice posouvá doprava prakticky o toto napětí. Citlivost kanálu na řízení prvním hradlem výrazně stoupá a posouvá se i prahové napětí prvního kanálu. Tyto tranzistory jsou použitelné jako multiplikativní směřovače a nízkošumové zesilovače až do kmitočtu 1 GHz.

6.3.2 Tranzistory MESFET a HEMT

Tranzistory JFET nebo MOSFET jsou použitelné jen do frekvencí kolem jednoho gigahertzu. K dalšímu zvýšení této hranice je třeba zkrátit kanál a zvětšit pohyblivost nosičů náboje, což však uvedená struktura nedovoluje. Výrazný pokrok v tomto směru proto představuje tranzistor FET s hradlem vytvořeným Schottkyho bariérovou diodou, označovaný symbolem MESFET (tj. Metal Semiconductor FET). Hradlo zde má podobu tenkého (napařeného) kovového pásu. Délka kanálu zde může být zmenšena pod $1 \mu\text{m}$, což při použití křemíku umožňuje dosáhnout mezní frekvence f_{max} až 10 GHz. Při použití arzenidu galia, vlivem několikrát větší pohyblivosti elektronů tohoto materiálu

Obr. 6.35. Tranzistor MOS se dvěma hradly

se může zvýšit mezní frekvence až na několik desítek gigahertzů. To jsou ovšem hodnoty mnohem vyšší, než jaké se vyskytují i u těch nejlepších bipolárních tranzistorů. Uvážíme-li, že i šumové vlastnosti jsou v gigahertzové oblasti znatelně lepší než u bipolárních tranzistorů, je zřejmé, že tranzistory MESFET mohou najít v mikrovlnné radiotechnice velmi úspěšné



Obr. 6.36. Charakteristiky a prostorové uspořádání tranzistoru SIT

uplatnění.

V roce 1978 byla objevena a záhy nato zavedena do praxe další varianta mikrovlnných tranzistorů FET, označovaná symbolem HEMT (angl. High Electron Mobility Transistor). U těchto tranzistorů je kanál tvořen jen velmi tenkou (150 nm) vrstvičkou "elektronového plynu", která může být z hlediska mechaniky transportu nosičů náboje - elektronů - považována za dvojrozměrnou. Pohyblivost elektronů je zde podstatně větší než u tranzistorů MESFET

s trojrozměrným kanálem, což je jednou z hlavních příčin jejich příznivějších frekvenčních i šumových vlastností. Dalšími výhodami tranzistorů HEMT je i možnost dosažení větší strmosti, a to při podstatně menším klidovém proudu kolektoru, než u tranzistoru MESFET. Menší technologická náročnost a nižší cena potom umožňuje používat tranzistory HEMT

téměř ve všech aplikacích dosud vyhrazených tranzistorům MESFET (vstupní jednotky přijímačů družicové televize a pod.)

6.3.3 Výkonové unipolární tranzistory

Dosud diskutované tranzistory MOS i JFET bývají nazývány **horizontální** FET vzhledem k horizontální poloze kanálu, a tím i k průběhu proudu. Kvůli odvodu tepla (vyvíjí se v kanále, tedy na povrchu) a relativně velkému odporu otevřeného kanálu (je větší než odpor sepnutého bipolárního tranzistoru) tyto horizontální tranzistory nejsou vhodné k použití asi nad 100 mW kolektorové ztráty. Nevýhodné je uložení všech tří elektrod na jedné straně křemíkové destičky (při potřebě přiložit mezi kolektor a emitor vyšší napětí i z hlediska využití plochy). Pro vyšší proudy a napětí používáme **vertikální** tranzistory FET, označované **VFET**, které řeší uvedené problémy.

6.3.4 Výkonové tranzistory JFET (SIT)

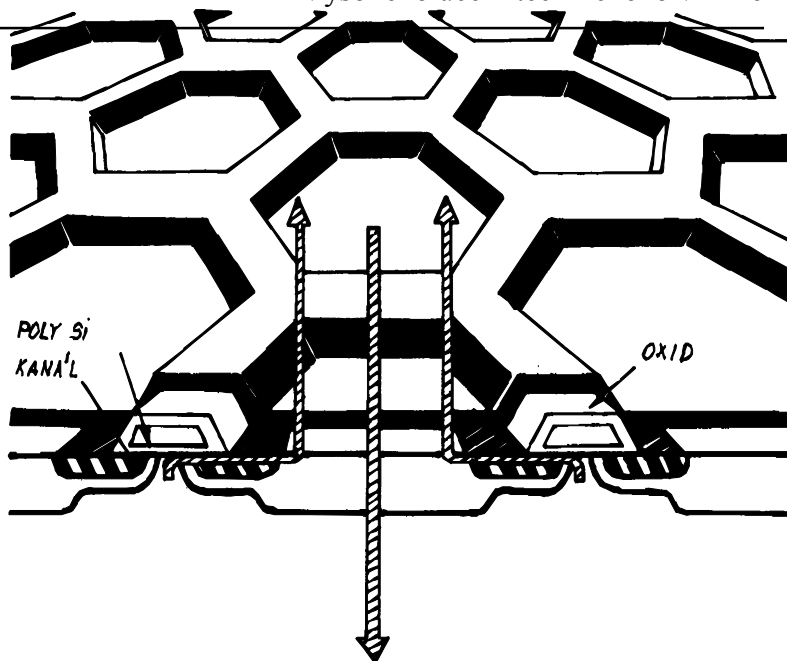
Výkonové struktury SIT [Static Induction Transistor] jsou určitou analogií tranzistoru JFET. Vlastní konstrukce může být různá, např. obr.6.36. Jde o mnoho paralelně spojených struktur s velmi krátkým kanálem. Podobně jako u ostatních výkonových součástek i zde slabě dotovaná vrstva N je podmínkou pro vytvoření vysokonapěťové součástky.

Poznámka: Podobně lze konstruovat tyristory řízené elektrickým polem známé pod stejnou zkratkou SIT [Static Induction Thyristor] nebo FCT [Field Controlled Thyristor]. Patří mezi perspektivní součástky, lze je vypínat. Z důvodu velmi náročné technologie a vysoké ceny se struktury SIT dosud málo vyrábějí a používají.

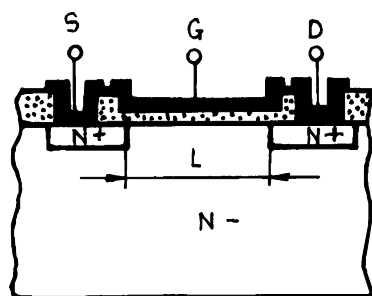
6.3.5 Výkonové tranzistory MOS

Jak již bylo uvedeno, u horizontálního tranzistoru MOS je pro výkonové účely na závalu především velký úbytek napětí v sepnutém stavu (způsobený délkou kanálu, kterou nelze z výrobních důvodů příliš zkrátit).

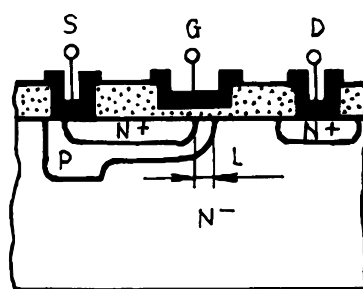
Prvním pokusem o zkrácení délky kanálu byla struktura **LDMOS** [Lateral Double Diffused]. Výhodou této struktury je možnost přesného nastavení délky kanálu pomocí boční difúze. Přenesením elektrody D na druhou stranu křemíkové destičky vznikla vertikální struktura **VDMOS** [Vertical Double diffused MOS]. Tím byla lépe využita plocha křemíkové destičky u součástek pro větší proudy, ale i zajištěna dostatečná izolační pevnost mezi kolektorem a emitorem u vysokonapěťových součástek a zároveň možnost dobrého odvodu



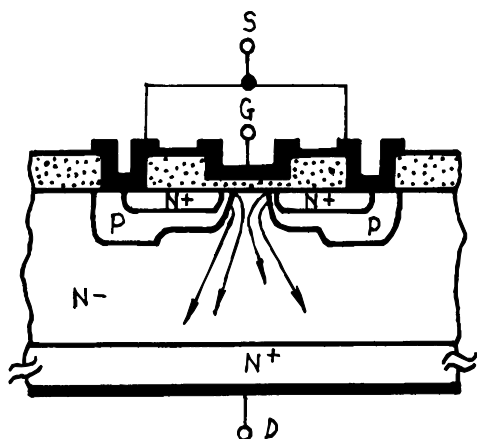
Obr. 6.38. Struktura tranzistoru HEXFET



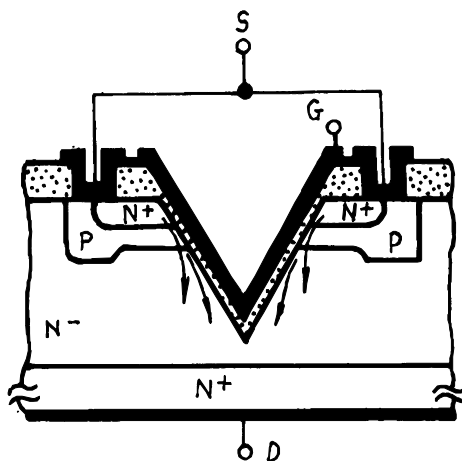
LMOS



LDMOS



VDMOS



VMOS (VVMOS)

Obr. 6.37. Struktury výkonových tranzistorů MOS

tepla.

Dalším způsobem, jak zajistit dobrou reprodukovatelnost délky velmi krátkého kanálu, je využití technologie vertikálního selektivního leptání. Vzniká charakteristický tvar tzv.

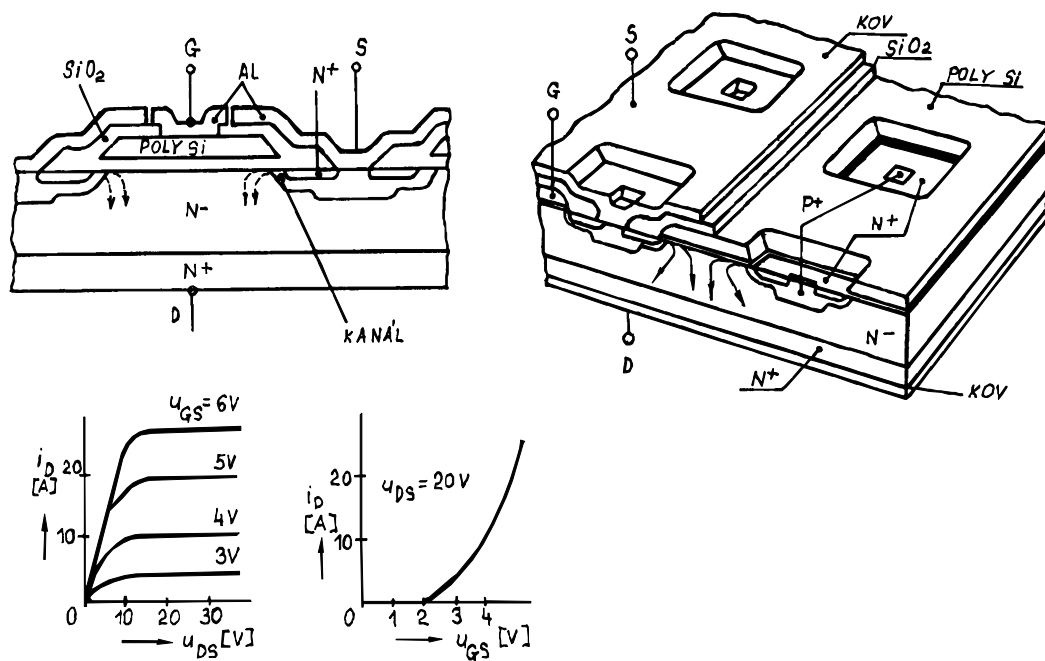
V leptu nebo U leptu. Výslednou strukturu označujeme jako VMOS [V lept MOS] nebo správněji VVMOS [Vertical V lept MOS].

Velmi významné je zjištění, že vlivem ohřevu dochází při konstantním napětí k mírnému snížení proudu kanálem, na rozdíl od bipolárních součástek. Výkonové tranzistory MOS jsou tedy relativně teplotně stabilní, což usnadňuje jejich paralelní řazení.

Pro výkonové aplikace jsou v současné době k dispozici tranzistory MOS s kanálem N i P (včetně komplementárních dvojic). Jsou převážně s indukovaným kanálem.

Tranzistor s paralelními strukturami VMOS se vyrábí jen pro menší výkony. Pro větší výkony jsou užívány tranzistory VDMOS s kanálem typu N (nebo dvojice N a P), které jsou vytvořeny velkým množstvím (až 10 000) dílčích struktur malých rozměrů (délka kanálu okolo 5 μm), a to paralelně spojených. Technologie výkonových tranzistorů MOS je blíže technologii integrovaných obvodů než technologii bipolárních výkonových součástek. Součástky různých výrobců se při stejném principu od sebe liší geometrickým uspořádáním jednotlivých dílčích struktur a jsou známy pod svými obchodními názvy:

- a) trojúhelníkové oblasti - TRIMOS (RCT a Texet),
- b) šestiúhelníkové oblasti - HEXFET (International Rectifier),



39. Tranzistor SIPMOS: a) řez strukturou, b) prostorové uspořádání, c) charakteristiky jednoho vyráběného typu

- c) čtvercové emitorové oblasti - SIPMOS (Siemens), TMOS (Motorola).

Dosažitelné maximální napětí mezi kolektorem a emitorem bývá až 1000 V při říditelném proudu asi 5 A, při napětích asi 400 V bývá říditelný proud 20 A až 50 A.

6.3.6 Tranzistory IGBT

Struktura tranzistorů IGBT vznikla ze struktury tranzistoru VD MOS nahražením substrátu N^+

Obr. 6.40. Struktura tranzistoru IGBT

substrátem P^+ . Struktura je schematicky znázorněna na obr. 6.40. V období po vynálezu struktury v r. 1982 byla pro tuto strukturu různými firmami používáno různé označení např. IGT (Insulated - Gate Transistor) používala firma General Electric, COMFET (Conductivity Modulated FET) používala firma RCA, GEMFET (Gain - Enhanced MOSFET) používala firma Motorola. V současné době (zhruba od r. 1988) se pro strukturu všeobecně používá označení IGBT.

Přestože struktura IGBT se tranzistoru MOS podobá, její funkce je značně odlišná. K pochopení funkce struktury jsou nutné znalosti funkce tranzistoru MOS, bipolárního tranzistoru a výkonové diody. Náhradní schéma struktury tranzistoru IGBT je znázorněno na obr. 6.41.

Funkce tranzistoru IGBT je následující. Při přiložení kladného napětí u_G na řídicí elektrodu tranzistoru MOS se vytvoří inverzní vrstva pod elektrodou, spojující vodivým kanálem emitor N^+ s oblastí báze N. Pokud řídicí napětí je dostatečně vysoké, je odpor kanálu malý. Vodivé propojení emitoru E s bází typu N vyvolává injekci děr z přechodu P^+N . Injekce nosičů má za následek výrazné snížení sériového odporu R_D struktury MOS. Tato vlastnost umožňuje vysokou proudovou hustotu při provozu IGBT.

Tranzistory IGBT jsou vhodné v aplikacích vyžadujících vysoká blokovací napětí a velké proudy. Špičkové součástky dosahují kombinace parametrů $U_{CE(BR)} = 1400 \text{ V}$, $I_{CEmax} = 300 \text{ A}$. Na druhou stranu, diodová část struktury má velký úbytek napětí i pro relativně malé kolektorové proudy. IGBT nejsou proto vhodné v aplikacích, kde se požaduje úbytek napětí u_{CE} menší než 0,7 V.

Obr. 6.41. Model a používané schématické značky tranzistoru IGBT

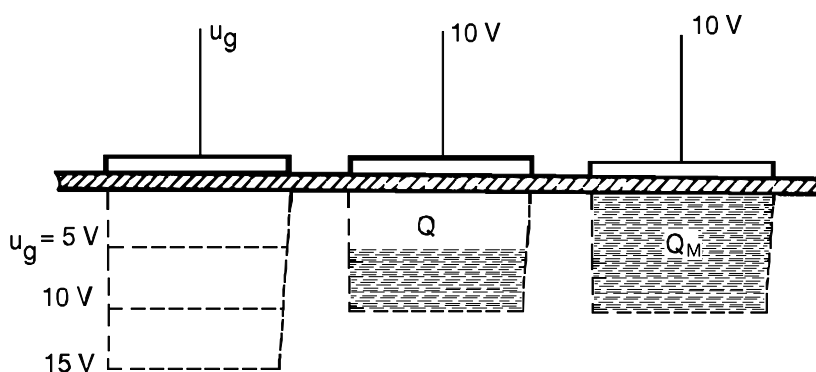
6.4 Struktury CCD

Struktury CCD (charge coupled devices) jsou v podstatě složeny z tranzistorů MOS, vyrobených na jednom substrátu a umístěných blízko sebe. Tato skutečnost je spojená s vlastnostmi které nejsou dosažitelné u zapojení s diskrétními součástkami. Je třeba zdůraznit, že tyto struktury nemají zesilovací schopnosti. Mohou pouze uchovat v čase a přemístit v prostoru signál v podobě náboje do nich zavedeného. Jedná se přitom o stále stejné nosiče náboje. Ve všech ostatních existujících typech polovodičových prvků nastává v každém okamžiku zpracování informace změna "informačního náboje" jiným nábojem dodávaným napájecím zdrojem.

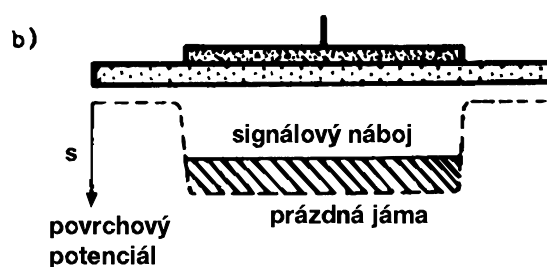
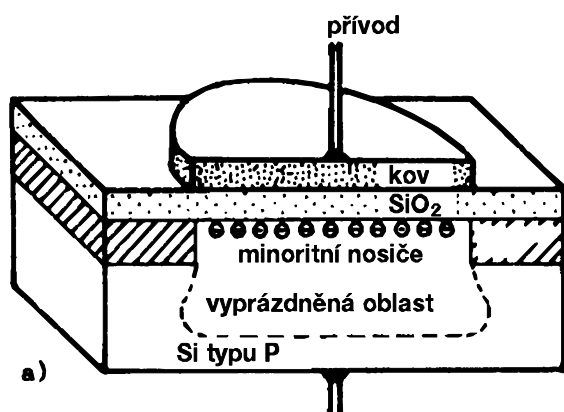
Charakteristickou zvláštností struktur CCD je možnost uchování náboje pohyblivých nosičů v tzv. potenciálové jámě, lokalizované určitým způsobem v polovodiči pod kovovou elektrodou, oddělenou od polovodiče vrstvou dielektrika, a přenosu tohoto informačního náboje do sousední, hlubší potenciálové jámy. Pro všechny nábojově vázané struktury je typický přenos informačního náboje v polovodiči posloupností potenciálových jam.

Předpokládejme křemíkovou destičku s vodivostí typu P, na které je kovová elektroda oddělená od polovodiče vrstvou izolantu (oxidu), obr.6.26 . Na takto vzniklou strukturu MOS budeme přikládat rostoucí napětí u_g tak, že kovová elektroda je vzhledem k polovodiči kladná. Za tohoto stavu bude stále větší počet děr odpuzován z oblasti polovodiče pod kovovou elektrodou, takže se zde vytvoří tzv. ochuzená (vyprázdněná) oblast. Nejvýrazněji poklesne koncentrace děr těsně u rozhraní mezi polovodičem a oxidovou vrstvou, směrem dovnitř polovodiče vliv vnějšího elektrického pole slábne. Při dalším zvyšování napětí u_g klesne koncentrace děr na rozhraní polovodič - oxid pod hodnotu koncentrace elektronů (ta je mnohem menší než intrinsická), typ vodivosti se obrátí. Napětí na kovové elektrodě struktury MOS, při které dochází ke změně typu vodivosti, se nazývá prahové napětí u_p .

Vzrůst napětí u_g nad hodnotu prahového napětí u_p však nemusí mít za následek okamžitý vznik inverzní vrstvy. U tranzistoru MOS tato inverzní vrstva vznikne okamžitě, protože se propojí oblast polovodiče zcela ochuzená o elektrony s nadifundovanými oblastmi typu N (kolektor a emitor tranzistoru).



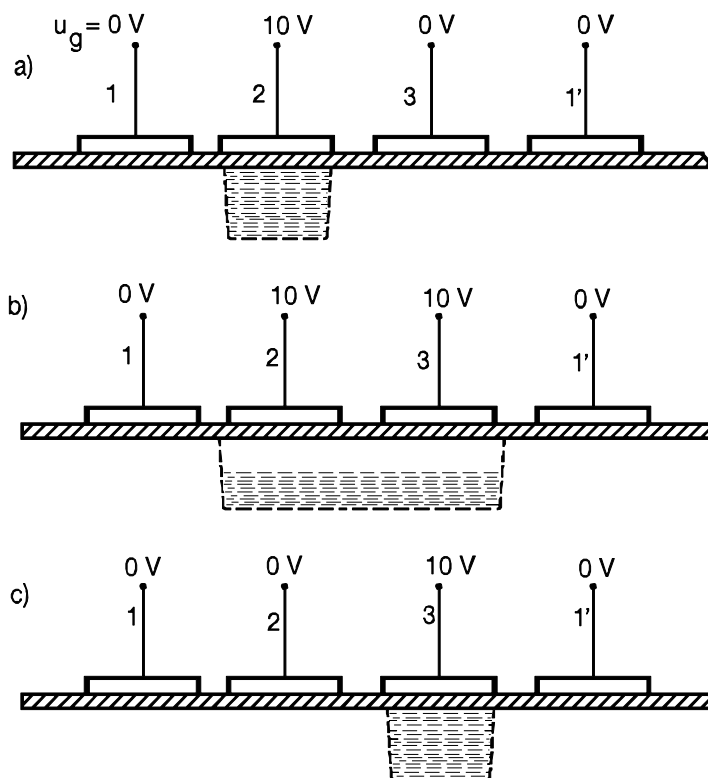
Obr. 6.27.: Pojem potenciálové jámy: a) hloubka prázdné potenciálové jámy je přibližně úměrná napětí u_g mezi kovovou elektrodou a polovodičem, b) a c) při daném napětí u_g hloubka jámy (měřená k povrchu kapaliny) klesá se vzrůstajícím nábojem



Obr. 6.26.: Struktura MOS: a) řez strukturou a vytvoření potenciálové jámy, b) potenciálová jáma se signálovým nábojem znázorněným kapalinou na dně potenciálové jámy

Tyto oblasti jsou schopny téměř okamžitě dodat do ochuzené oblasti polovodiče libovolné množství elektronů a vytvořit tak stacionární inverzní vrstvu. U struktur CCD (které jsou v podstatě složeny ze struktur MOS, vyrobených na jednom substrátu a umístěných blízko sebe) je situace zcela odlišná. Polovodičový materiál má malou vodivost, takže koncentrace nosičů

je zde velmi malá. Navíc v okolí vnější kovové elektrody nejsou žádné zdroje minoritních nosičů, takže se i při napětí u_g kovové elektrody vyšším, než je prahové napětí u_p struktury, nevytvoří u povrchu polovodiče inverzní vrstva, ale pouze oblast zcela ochuzená o volné nosiče náboje. Vytvoří se tak "prázdná oblast" o hloubce úměrné napětí u_g a průřezu závislém na tvaru vnější kovové elektrody. Tato oblast se chová jako jáma schopná přijmout do sebe elektrický náboj minoritních nosičů úměrný napětí mezi kovovou elektrodou a polovodičem, a samozřejmě i ploše elektrody.



Obr. 6.28.: Znárodnění přenosu náboje v polovodiči pomocí soustavy kovových elektrod izolovaných od polovodiče tenkou vrstvou dielektrika (SiO_2)

Je možné si představit, že napětí u_g vytvoří potenciálovou jámu o hloubce úměrné u_g . Zavedení náboje Q do jámy, které způsobí snížení povrchového potenciálu ϕ_s , je analogické nalití kapaliny do jámy, jejíž hloubka (měřená od okraje jámy k povrchu kapaliny) klesá. Jáma má tedy svoji maximální nábojovou kapacitu Q_M odpovídající velmi malé hodnotě povrchového potenciálu ϕ_s (< 1 V). Na obr. 6.27 je znázorněna potenciálová jáma při různých napětích u_g a při různých nábojích uchovávaných v této jámě.

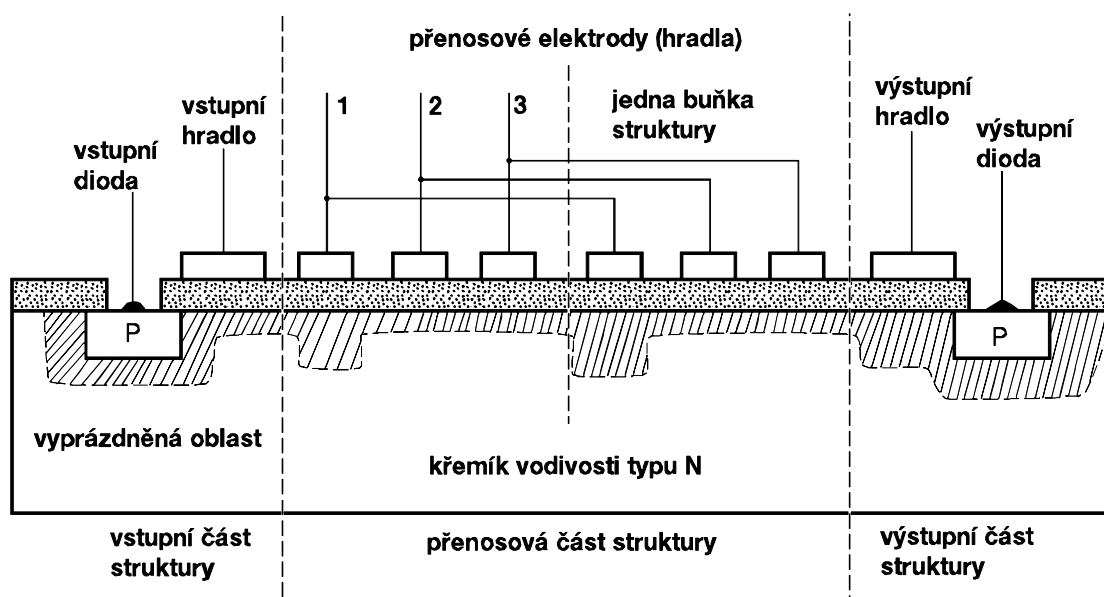
Je třeba připomenout, že potenciálová jáma vytvořená v polovodiči na struktuře MOS, nezůstane prázdná neomezeně dlouho. Vlivem tepelných procesů - generace (ale i rekombinace) párů elektron - díra - shromažďují se v potenciálové jámě elektrony generované v oblasti pod kovovou elektrodou a jejím okolím do té doby, než se obnoví rovnovážný stav. Doba, za kterou se potenciálová jáma naplní nábojem vzniklým tepelnou generací, je řádu desítek milisekund. Chlazením polovodiče lze dosáhnout podstatného zlepšení a prodloužení doby, po kterou zůstane potenciálová jáma bez tepelně generovaného náboje, v zásadě však tepelná generace nosičů ovlivňuje dobu, po kterou zůstane potenciálová jáma prázdná nebo je-li v ní uchován nějaký náboj - dobu po kterou uchovaný náboj není výrazně změněn nábojem vzniklým tepelnou generací nosičů. Aby k popisované situaci nedošlo, je vhodnou konstrukcí struktury zajištěn zdánlivý posun potenciálové jámy v polovodiči a v případě, že je v ní uchován nějaký náboj, posun tohoto náboje.

Přenos vázaného náboje strukturou lze vysvětlit pomocí části struktury, složené ze čtyř kovových elektrod blízko sebe, obr.6.28. Předpokládejme, že v potenciálové jámě pod druhou elektrodou s napětím $u_g = 10$ V ($u_g > u_p$) je uložen určitý náboj Q , který tam byl nějakým způsobem zaveden.

Napětí ostatních elektrod jsou nulová. Připojíme-li nyní na třetí elektrodu napětí $u_g = 10$ V nebo vyšší, vytvoří se i pod touto elektrodou potenciálová jáma a při dostatečně malé vzdálenosti mezi elektrodami obě potenciálové jámy splynou v jámu jedinou. Při stejném napětí u_g na druhé a třetí elektrodě se náboj rovnoměrně rozdělí mezi obě oblasti. Odpojíme-li nyní napětí z elektrody 2, přeteče do potenciálové jámy pod třetí elektrodou i zbývající část náboje, náboj byl posunut o jednu elektrodu vpravo. Stejným způsobem je možné posunout náboj pod elektrodu 1'.

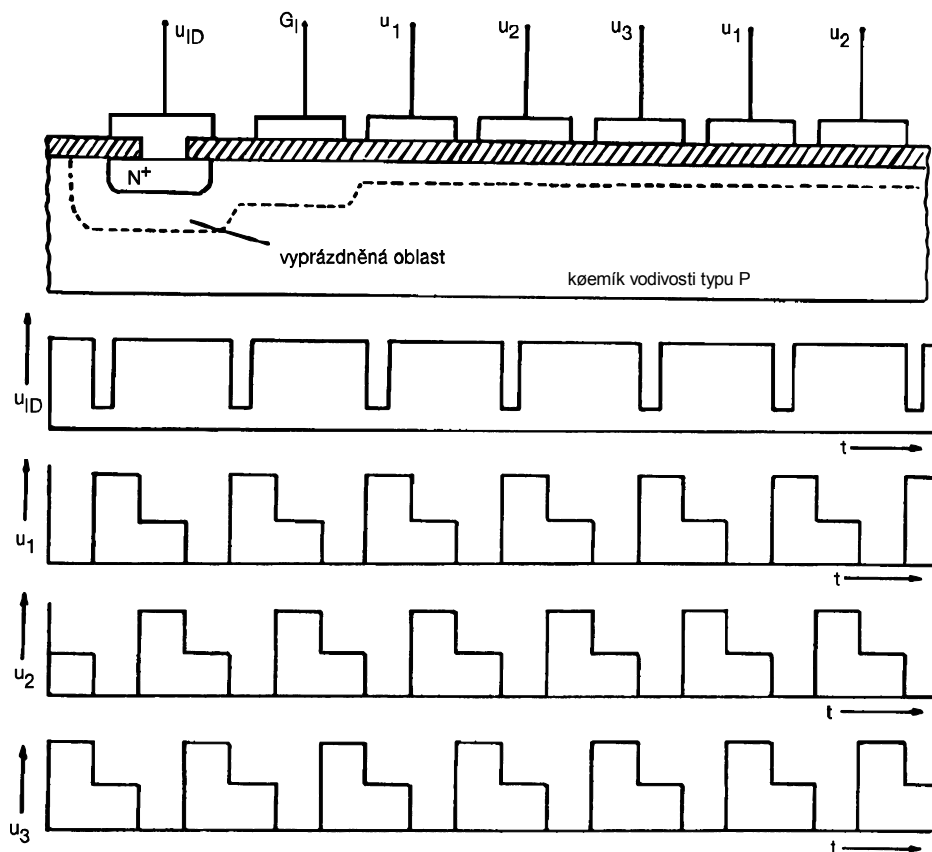
Velikost náboje Q přenášeného strukturou je úměrná okamžité hodnotě vstupního signálu v době jeho převodu na náboj. Má-li být tento náboj přenesen beze zbytku, musí být splněny následující požadavky:

- náboj musí být přenesen dříve, než se jeho velikost změní vlivem tepelné generace minoritních nosičů uvnitř dané potenciálové jámy a v jejím okolí, tedy napětí u_g na vnějších kovových elektrodách musí být dostatečně rychle měněno tak, aby se potenciálová jáma neustále přesouvala,
- napětí na dvou sousedních elektrodách se musí alespoň částečně časově překrývat, aby vznikla jediná potenciálová jáma a náboj se mohl "rozlít". Profil společné potenciálové jámy musí být takový, aby v prostoru pod dielektrikem mezi kovovými elektrodami nevznikala potenciálová bariéra a náboj se přenesl beze zbytku.



Obr. 6.29.: Zjednodušený obrázek typické struktury CCD - dvoubitového třífázového posuvného registru

Uvedený výklad vzniku potenciálové jámy a přenosu náboje minoritních nosičů (elektronů) v polovodiči typu P je možné aplikovati na polovodiči typu N. Polarity napětí kovových elektrod musí být ovšem opačné a minoritními nosiči přenášejícími signálový náboj jsou díry. Na základě analogie s tranzistory MOS můžeme pak o struktuře CCD s kanálem typu P, je-li náboj přenášen děrami nebo o struktuře CCD s kanálem typu N, pokud je signálový náboj tvořen náboji elektronů, pohybujících se v potenciálových jamách vytvořených v polovodiči typu P.



Obr. 6.30.: Zapojení vstupní části struktury CCD určené pro zpracování číslicových signálů

V praxi se používá řada různých druhů struktur CCD, lišících se mimo jiné způsobem jakým jsou potenciálové jámy vytvářeny a jak je prakticky zajištěno "přelití" náboje do sousední potenciálové jámy. Bez ohledu na konstrukční a technologické varianty jednotlivých struktur můžeme každou strukturu rozdělit na tři základní části: vstupní, přenosovou a výstupní.

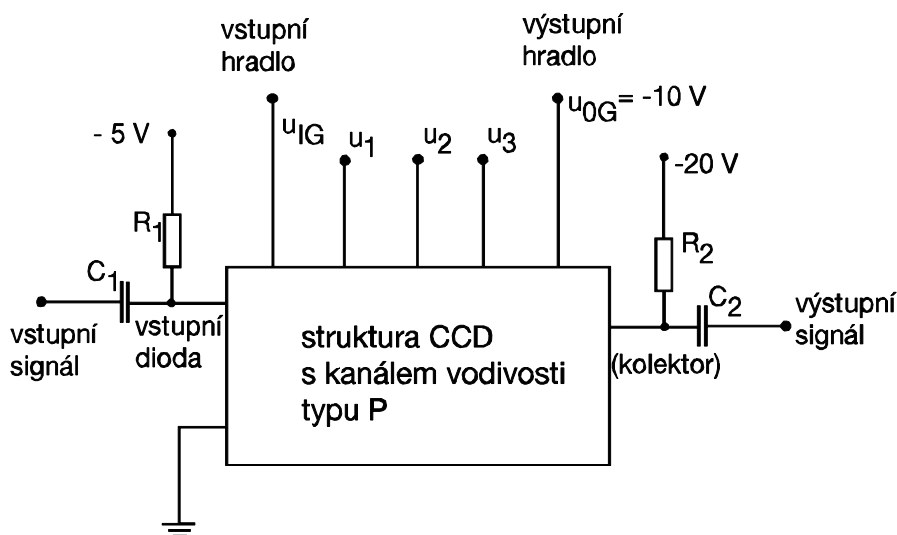
Vstupní část struktury CCD.

Úkolem vstupní části nábojově vázané struktury je přeměna vstupního signálu na sled nábojů s velikostí úměrnou okamžité hodnotě vstupního signálu a zavedení těchto nábojů do přenosové části struktury. Protože vstupním signálem nábojově vázaných struktur může být elektrický i optický signál spojitý nebo diskrétní, závisí uspořádání vstupní části struktury na tom, pro jaký signál je struktura určena.

Vstupní část struktury nejčastěji tvoří přechod PN - vstupní dioda, vytvořená difúzí do základního polovodičového materiálu, a hradlo. Toto hradlo se obvykle nazývá vstupní, přesto, že je v některých zapojeních připojeno pouze na stejnosměrné napětí a vstupní signál je přiváděn na vstupní diodu. Na obr.6.29 je schématicky znázorněna třífázová struktura CCD - třífázový dvoubitový posuvný registr. Jsou zde vyznačeny jednotlivé části struktury.

Při použití struktury CCD ke zpracování číslicových signálů se používá např. uspořádání vstupní části podle obr.6.30. Každá buňka struktury je schopna uchovávat informaci jednoho bitu. Stav logické "0" odpovídá situaci, kdy v dané buňce není žádný náboj. Stav logické "1" odpovídá maximálnímu náboji v příslušné buňce. Při tomto uspořádání vstupní části struktury

podle obr. 6.30 se napětím u_s mezi polovodičem (substrátem) a přenosovými elektrodami udržuje na povrchu polovodiče vyprázdněná oblast. Přechod PN na vstupu struktury je napětím u_{ID} (řádově asi 10 V) polarizován v závěrném směru. Jestliže je napětí na přechodu PN (vstupní diodě) blízké nule, pak při příchodu hodinového impulsu fáze j_1 (na první přenosovou elektrodu), jsou do struktury po dobu trvání tohoto impulsu injektovány minoritní nosiče náboje. Tyto náboje projdou vyprázdněnou oblastí pod vstupní elektrodou G_I do



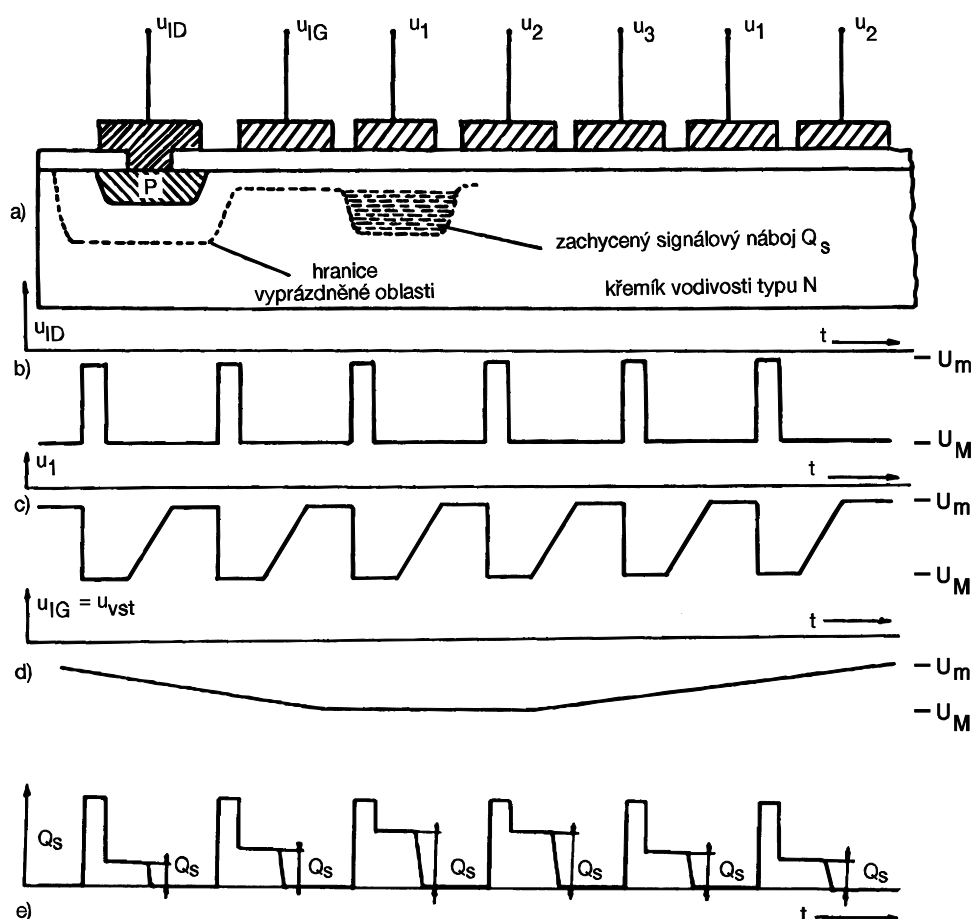
Obr. 6.31.: Zapojení struktury CCD pro zpracování spojitého signálu. Znáznorněná struktura má kanál vodivosti typu P. U struktur s kanálem vodivosti typu N jsou polarity napětí na vstupu a výstupu opačné, t.j. jsou kladné (včetně polarit napětí hodinových impulsů)

potenciálové jámy pod první přenosovou elektrodou, odtud jsou pak přenášeny dále.

Zapojení struktury CCD určené pro spojitý vstupní signál je na obr.6.31. Vstupní dioda (emitor) je polarizována v závěrném směru napětím kolem -5 V, připojeným přes odpor R_1 vstupní hradlo je připojeno na napětí -10 V, vstupní střídavý signál je přes kondenzátor C_1 (řádu 10 pF) přiváděn na vstupní diodu. V

okamžiku působení napěťového impulsu na vodiči fáze j_1 se potenciálová jáma pod první přenosovou elektrodou naplní nábojem úměrným okamžité hodnotě vstupního napětí. Toto zapojení je možné obměnit tak, že vstupní hradlo je zapojeno jako první přenosová elektroda (u_{IG} - časový průběh napětí na vstupním hradle, je stejný jako u_1 - časový průběh napětí první fáze). Průběh napětí u_2 , u_3 fází druhé a třetí jsou shodné s průběhem napětí u_1 , ale jsou fázově posunuty tak, že se vzájemně překrývají.

Při zpracování analogových signálů se musí struktura CCD chovat jako lineární obvod. Zajištění linearity ve vstupní části struktury má mimořádný význam. Pro udržení dobrého poměru signálu k šumu je dále třeba, aby úroveň vstupního signálu byla několik voltů. Požadavkům lineární závislosti mezi přivedeným nábojem a vstupním signálem i dobrému poměru signál-šum vyhovuje



Obr. 6.32.: Struktura CCD: a) zapojení vstupní části struktury pro lineární injekci náboje, b) průběh napětí na vstupní diodě, c) časové průběhy napětí u_1 fáze j_1 , d) vstupní napětí, e) zachycený náboj

uspořádání vstupní části struktury podle obr.6.32. Vstupní část struktury se v tomto případě skládá ze vstupní diody (emitoru) a vstupního hradla. Vstupní signál u_{IG} je převáděn na vstupní hradlo. Jeho amplituda se může pohybovat v rozmezí od u_m do u_M , což je rozsah napětí, v jakém se mění napětí hodinových impulzů každé fáze. Je-li na první přenosové elektrodě napětí u_M , je napětí u_{ID} na vstupní diodě takové, že tato dioda je jen částečně uzavřena a náboje minoritních nosičů protékají pod vstupním hradlem do potenciálové jámy pod první přenosovou elektrodou. Pak je závěrné napětí vstupní diody zvýšeno a náboj minoritních nosičů se odčerpává, dokud se povrchový potenciál pod první přenosovou elektrodou nerovná povrchovému potenciálu pod vstupním hradlem. Náboj zachycený v potenciálové jámě pod přenosovou elektrodou je pak úměrný rozdílu napětí hodinových impulzů u_{ID} a napětí vstupního. Protože amplitudy napětí hodinových impulzů jsou konstantní, je zachycený náboj přímo úměrný vstupnímu napětí. Tento náboj se pak přenáší dále.

Při zpracování **optických signálů** dochází ke vzniku párů elektron-díra vlivem elektromagnetického záření, které dopadá na příslušnou část struktury. V zásadě je možné ozařovat povrch struktury ze strany elektrod i ze strany opačné. V prvním případě dopadá optický signál na polovodič mezerami mezi netransparentními kovovými elektrodami, nebo proniká tenkými transparentními elektrodami vyrobenými z polykrystalického křemíku. Této konstrukce se používá nejčastěji, především pro televizní obrazové snímáče. Je-li optický signál přiváděn na opačnou stranu než jsou umístěny přenosové elektrody, je třeba tuto část

struktury zeslabit, aby byly páry elektron-díra vytvářeny v oblasti, ve které se uplatňuje vliv napětí na přenosových elektrodách.

Po ozáření struktury a vzniku dvojic elektron-díra jsou majoritní nosiče napětím přiloženým na elektrodách odpuzovány, náboj minoritních nosičů úměrný ozáření je pod elektrodami akumulován, a přenášen k dalšímu zpracování.

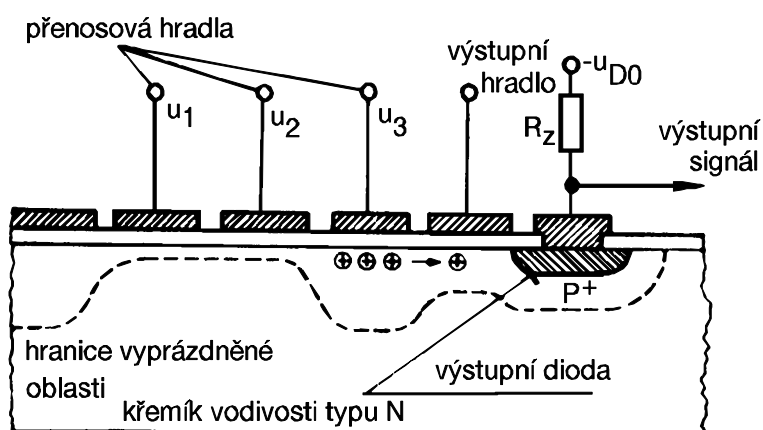
Přenosová část struktury CCD.

Úkolem této struktury je přenést jednotlivé "nábojové balíky", úměrné vstupnímu signálu na výstup struktury. Od přenosové části struktury požadujeme, aby přenášela tyto "nábojové balíky" s co nejmenším zkreslením a s požadovanou rychlostí. Pro zajištění přenosu náboje pouze jedním směrem je nutné, aby přenosová část struktury nebyla elektricky symetrická. Podle toho, jakým způsobem je vytvořena nesymetrie potenciálové jámy v polovodiči a podle toho kolik fází musí mít napětí hodinových impulzů pro zdánlivý pohyb potenciálových jam v polovodiči, rozeznáváme struktury jednofázové, dvoufázové, třífázové a čtyřfázové. Počet fází hodinových impulzů je obvykle shodný s počtem přenosových elektrod tvořících elementární buňku struktury. Podrobný popis činnosti je možné nalézt např. v [36].

Výstupní část struktury CCD.

Slouží k přeměně "nábojových balíků", přenášených strukturou, na napětí. Existuje mnoho obvodů umožňujících tuto přeměnu; pro ilustraci uvedeme alespoň dva.

Nejjednodušší uspořádání výstupní části struktury je na obr.6.33. Na přechod PN - výstupní diodu - se přivádí napětí u_{D0} v závěrném směru. V okolí přechodu se tak vytvoří vyprázdňená oblast, působící jako nora pro minoritní nosiče přenášené strukturou. Mezi poslední přenosovou elektrodou a výstupní diodou je umístěna výstupní elektroda (výstupní hradlo) na kterém je takové stejnosměrné napětí, aby se po přivedení hodinového impulzu na poslední přenosovou elektrodu spojily



Obr. 6.33.: Výstupní obvod struktury s přenosem vázaného náboje

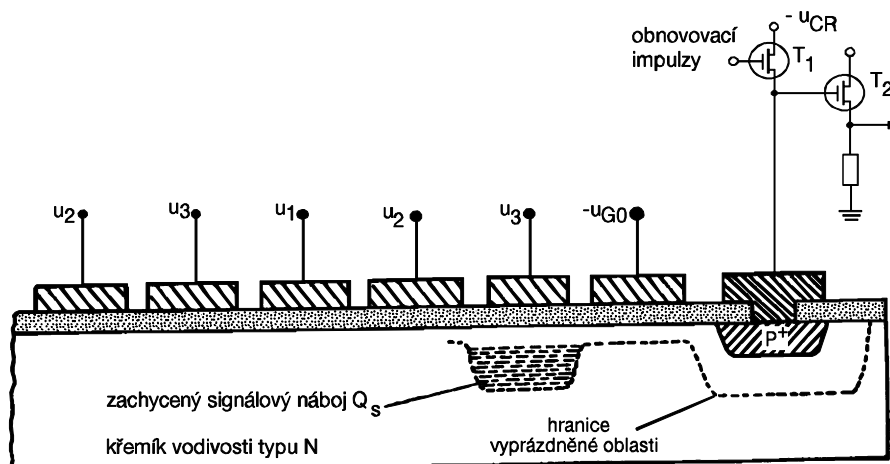
vyprázdňené oblasti pod těmito elektrodami s vyprázdňenou vrstvou v okolí přechodu PN a přenášený náboj byl vtažen do přechodu. Při průchodu náboje vhodným zatěžovacím rezistorem s odporem R_Z vznikne napěťový impulz. Jeho amplituda je úměrná velikosti náboje, t.j. velikosti vstupního napětí v okamžiku jeho přeměny na "nábojový balík".

Jiné uspořádání výstupní části struktury je na obr.6.34 je tvořeno:

- výstupním hradlem (napětí u_{G0}),
- tranzistorem T_1 zvaným obnovovací tranzistor (reset transistor) a
- výstupním tranzistorem T_2 .

Výstupní hradlo je umístěno mezi poslední přenosovou elektrodou a nadifundovanou oblastí P^+ , tvořící s materiálem substrátu typu N přechod P^+N výstupní diody (nábojového detektoru) a emitoru tranzistoru T_1 .

Na výstupní hradlo je připojeno stejnosměrné napětí u_{G0} (asi -15 V). Výstupní dioda je tak stíněna od přenosových elektrod a nepřenášejí se na ni poruchy způsobené hodinovými impulzy. Na výstupní diodu se přes obnovovací tranzistor T_1 přivádí napětí u_{CR} takové polarity, aby dioda byla uzavřena. Za předpokladu, že zpětný proud diody je zanedbatelný, zůstane dioda na potenciálu u_{CR} i po rozepnutí tranzistoru T_1 . Po příchodu nábojového balíku do potenciálové jámy pod poslední přenosovou elektrodou se při snižování napětí na této elektrodě nábojový balík přenesou na "plovoucí" oblast P^+ nábojového detektoru (výstupní diody) a způsobí změnu jejího



Obr. 6.34.: Jiné zapojení výstupní struktury CCD ("zapojení s plovoucí elektrodou")

potenciálu. Tato změna, úměrná přenesenému nábojovému balíku je přiváděna na hradlo tranzistoru T_2 a může být dále zesilována. Před příchodem dalšího nábojového "balíku" je třeba znovu nabít oblast P^+ nábojového detektoru na potenciál u_{CR} . Zápornými obnovovacími impulzy přiváděnými na hradlo tranzistoru T_1 je proto tento tranzistor periodicky spínán se stejnou frekvencí, jako je frekvence hodinových impulzů, někdy bývá hradlo obnovovacího tranzistoru spojeno s některou z přenosových elektrod.

6.4.1 Využití struktur CCD

Řádkové obrazové snímače CCD obsahují dnes maximálně asi 1800, plošné asi 150 000 až 200 000 snímacích prvků, které umožňují převádět optický signál na signál elektrický (největší dostupný obvod vyrábí firma KODAK. Jde o snímač 2048 x 2048 bodů s plochou 4 cm²). Pomocné obvody pro zpracování nábojů představují nejsložitější část snímačů a jsou tvořeny počtem prvků, který 2,2 až 3-krát převyšuje počet snímacích prvků.

Jako řádkových obrazových snímačů lze použít běžných struktur CCD, konstrukčně uspořádaných tak, aby záření dopadající na strukturu mohlo generovat páry elektron-díra v polovodiči substrátu. Řádkový obrazový snímač je tak prakticky identický s posuvným registrem CCD; jediný rozdíl spočívá v tom, že u obrazového snímače lze vynechat vstupní diodu, protože není třeba zavádět do struktury náboje elektricky.

Náboje jsou generovány zářením dopadajícím na snímač v křemíkovém substrátu. K akumulaci nábojů je třeba, aby přenosové elektrody jedné fáze měly stejné napětí, jako mají registry CCD při běžném použití (asi -20 až -30 V pro struktury CCD s kanálem P, nebo asi +10 V s kanálem N). Toto napětí je přivedeno na elektrody po dobu integrace náboje (a jeho akumulace pod elektrodami), která je řádu milisekund. Během této doby se generované náboje akumulují pod příslušnými elektrodami. Náboj pod každou elektrodou je úměrný intenzitě záření v mezerách mezi elektrodami, je-li užito elektrod netransparentních pro dopadající záření, nebo intenzitě záření dopadající na transparentní elektrody. Ve snímači se

tak vytvoří nábojový "obraz" odpovídající dopadajícímu záření. Ten je pak "přečten" obvyklým způsobem, t.j. přivedením pulzního napětí na přenosové elektrody. Integrace a čtení náboje se pak periodicky opakují, výstupní signál je vyveden přes vhodný zesilovač. Čtecí doba u tohoto druhu snímače musí být mnohem kratší (asi 50 μs) než je doba akumulace náboje. Jinak by mohlo dojít k "rozmazání" informace zářením dopadajícím na snímač během čtecí doby.

Dvourozměrné obrazy lze snímat pomocí plošných snímacích matic. Matice se skládá z části nazývané zápisová část, ze čtecí části a z výstupního registru. Během integrační doby dopadá záření na zápisovou část. V potenciálových jamách se vytvoří náboje, úměrné intenzitě dopadajícího záření (a době ozařování). Po uplynutí doby integrace je celá informace přesunuta do čtecí části snímače. Zápisová a čtecí část musí být stejné, nemá-li docházet ke ztrátě informace. Náboje úměrné ozáření jsou pak po řádcích posouvány do výstupního registru a dále na výstup struktury. Protože všechny náboje jsou detekovány pouze jednou malou výstupní diodou, je výstupní kapacita struktury malá a lze dosáhnout dobrého poměru signálu k šumu.

Uvedená konstrukce obrazového snímače není jediná. Existují ještě obrazové snímače s tzv. meziřádkovým přenosem (interline transfer technique), u kterých se jednotlivé řádky zápisové a čtecí části střídají. Výhodou takové konstrukce je, že náboje nahromaděné v zápisových řádcích jsou jen s velmi malými ztrátami přeneseny do sousedících řádek, takže lze dosáhnout většího poměru signálu k šumu. U struktur s různou tloušťkou dielektrika pod přenosovými elektrodami se mohou zápisové a čtecí řádky vzájemně překrývat. Výhodou této konstrukce je možnost dosažení větší rozlišovací schopnosti, nevýhodou je náročnější výroba.

Použití struktur CCD v optoelektronice jako obrazových snímačů je omezeno proudem snímače za tmy. Křemík použitý k výrobě obrazových snímačů musí proto mít malý obsah rychle difundujících příměsí (Cu, Fe), které jsou hlavními centry generace a rekombinace nosičů náboje a musí si zachovávat i po tepelném zpracování velkou (150 až 300 μs) dobu života minoritních nosičů náboje. Proud za tmy je totiž nepřímo úměrný době života minoritních nosičů náboje, pro dobu života $t = 150 \mu\text{s}$ je proud za tmy asi 5 nA/cm^2 , pro $t = 10 \mu\text{s}$ je asi 100 nA/cm^2 .

Zpracování signálů. Přímou aplikací principu činnosti struktur CCD je jejich použití jako zpožďovacích linek pro spojitě video a audio signály. Signál je vzorkován a náboje úměrné jeho okamžité velikosti jsou přenášeny strukturou. Na výstupu jsou tyto náboje detekovány a původní signál je získán průchodem signálu dolní propustí.

Časové zpoždění t_d může být buď konstantní nebo elektricky říditelné. Zpoždění lze řídit kmitočtem hodinových impulzů. Typické hodnoty zpoždění t_d jsou stovky mikrosekund až stovky milisekund, horní mez může být až 1 sekunda při pokojové teplotě. Maximální zpoždění je omezeno vlastním proudem struktury. Zpožďovací linky pro spojitě signály, realizované pomocí struktur CCD, se používají jako samostatné jednotky (v televizní technice, v radiolokaci, elektroakustice atd.) nebo jako stavební prvky složitějších funkčních bloků.