

WIESŁAW MARCINIAK

PRZYRZĄDY

PÓŁPRZEWODNIKOWE

I UKŁADY

SCALONE

ZASADY DZIAŁANIA

TECHNOLOGIA

I ZASTOSOWANIA



WYDAWNICTWA NAUKOWO-TECHNICZNE
WARSZAWA 1979

SPIS TREŚCI

Od Autora
Wykaz ważniejszych oznaczeń

stronica 11
stronica 13

I PODSTAWOWE WŁAŚCIWOŚCI FIZYCZNE PÓLPRZEWODNIKÓW

stronica 17

- 1.1 Model wiązań walencyjnych w sieci krystalicznej krzemu i germanu/20
 - 1.1.1 Budowa atomu krzemu i germanu/20
 - 1.1.2 Sieć krystaliczna/21
 - 1.1.3 Wiązania międzyatomowe/23
 - 1.2 Energetyczny model pasmowy półprzewodnika/24
 - 1.3 Półprzewodnik samoistny/28
 - 1.4 Półprzewodnik niesamoistny/29
 - 1.5 Koncentracja elektronów i dziur w stanie równowagi termodynamicznej/33
 - 1.5.1 Poziom Fermiego w półprzewodniku samoistnym/35
 - 1.5.2 Koncentracja nośników w półprzewodniku samoistnym/36
 - 1.5.3 Zależności Boltzmanna/36
 - 1.5.4 Bilans koncentracji nośników/36
 - 1.5.5 Równanie neutralności elektrycznej/36
 - 1.5.6 Zależność koncentracji nośników od temperatury/37
 - ✓ 1.6 Transport nośników w półprzewodniku/38
 - 1.6.1 Unoszenie/39
 - 1.6.2 Dyfuzja/43
 - 1.6.3 Unoszenie i dyfuzja/44
 - ✓ 1.7 Półprzewodnik w stanie nierównowagi termodynamicznej/44
 - 1.7.1 Szybkość generacji i rekombinacji/46
 - 1.7.2 Zmiany koncentracji nośników nadmiarowych w funkcji czasu — stan nieustalony/47
 - 1.7.3 Rozkład koncentracji nośników nadmiarowych przy ustalonym poziomie wstrzykiwania lub wyciągania/48
 - 1.7.4 Generacja-rekombinacja pośrednia/50
 - 1.8 Prawo ciągłości ładunku/52
 - 1.9 Półprzewodnik niejednorodny/53
- Zadania/55

2 METODY WYTWARZANIA MATERIAŁÓW I STRUKTUR PÓLPRZEWODNIKOWYCH

stronica 58

- 2.1 Wytwarzanie czystych oraz jednorodnie domieszkowanych materiałów półprzewodnikowych/59
- 2.2 Metody wytwarzania warstw domieszkowanych w (lub na) podłożu półprzewodnikowym/62
 - 2.2.1 Technologia stopowa/63

- 2.2.2 Epitaksja/63
- 2.2.3 Dyfuzja/66
- 2.2.4 Implantacja jonów/74
- 2.3 Kształtowanie lokalne właściwości fizykochemicznych półprzewodnika (domieszkowanie lokalne)/76
 - 2.3.1 Maskowanie/77
 - 2.3.2 Fotolitografia/79
 - 2.3.3 Epitaksja lokalna/81
 - 2.3.4 Dyfuzja lokalna/81
 - 2.3.5 Implantacja lokalna/82
- 2.4 Technologia planarna/82
- 2.5 Inne procesy technologiczne/84
Zadania/85

3 ZJAWISKA KONTAKTOWE I POWIERZCHNIOWE W PÓLPRZEWODNIKACH

stronica 86

- 3.1 Złącze *p-n*/88
 - 3.1.1 Istota złącza *p-n*, opis jakościowy/88
 - 3.1.2 Obszar ładunku przestrzennego/93
 - 3.1.3 Energetyczny model pasmowy złącza *p-n*/104
 - 3.1.4 Charakterystyka prądowo-napięciowa/106
 - 3.1.5 Przebiecie złącza *p-n*/116
 - 3.1.6 Charakterystyka prądowo-napięciowa złącza silnie domieszkowanego/119
 - 3.1.7 Modele złącza *p-n*/123
 - 3.1.8 Wpływ temperatury na charakterystykę prądowo-napięciową złącza *p-n*/140
- 3.2 Heterozłącza/144
- 3.3 Struktura MIS/145
 - 3.3.1 Model fizyczny idealnej struktury MIS/146
 - 3.3.2 Charakterystyka pojemnościowo-napięciowa/154
 - 3.3.3 Rzeczywista struktura MIS/157
- 3.4 Złącza metal-półprzewodnik/162
 - 3.4.1 Uproszczony model złącza *m-s*/163
 - 3.4.2 Model złącza *m-s* z uwzględnieniem stanów powierzchniowych/166
 - 3.4.3 Charakterystyka prądowo-napięciowa złącza prostującego/168
 - 3.4.4 Złącze omowe/173
Zadania/177

4 RODZAJE DIOD PÓLPRZEWODNIKOWYCH I ICH ZASTOSOWANIE

stronica 179

- 4.1 Diody prostownicze/184
 - 4.1.1 Analiza graficzna/186
 - 4.1.2 Schematy zastępcze/189
 - 4.1.3 Przykłady zastosowań/191
- 4.2 Diody uniwersalne/199
 - 4.2.1 Przykłady zastosowań/201
- 4.3 Stabilitrony/202
 - 4.3.1 Analiza graficzna/206
 - 4.3.2 Schematy zastępcze/208
 - 4.3.3 Przykłady zastosowań/208
- 4.4 Diody impulsowe/211
 - 4.4.1 Diody o małym czasie przełączania/213
 - 4.4.2 Diody o małym czasie opadania (diody ładunkowe)/213
 - 4.4.3 Analiza graficzna i schematy zastępcze/216
 - 4.4.4 Zastosowanie diod impulsowych/216

- 4.5 Diody pojemnościowe/217
 - 4.5.1 Schematy zastępcze diod pojemnościowych/218
 - 4.5.2 Przykłady zastosowań/222
- 4.6 Diody tunelowe/225
 - 4.6.1 Schemat zastępczy/227
 - 4.6.2 Przykłady zastosowań/230
- 4.7 Diody mikrofalowe (detekcyjne, mieszające oraz sterowane impedancje)/232
 - 4.7.1 Diody detekcyjne i mieszające/234
 - 4.7.2 Diody o zmiennej impedancji (diody *p-i-n*)/239

5

TRANZYSTOR BIPOLARNY

stronica 240

- 5.1 Podstawowa struktura tranzystora bipolarnego/241
- 5.2 Konstrukcja i technologia tranzystora bipolarnego/245
- 5.3 Podstawowe zasady działania tranzystora bipolarnego/249
- 5.4 Rozkład koncentracji nośników nadmiarowych w bazie/255
- 5.5 Współczynnik wzmocnienia prądowego α_N /260
- 5.6 Zakresy pracy i układy włączenia tranzystora bipolarnego/264
- 5.7 Praca nieliniowa statyczna (modele, charakterystyki, parametry)/268
 - 5.7.1 Modele nieliniowe statyczne/269
 - 5.7.2 Charakterystyki statyczne/272
 - 5.7.3 Parametry statyczne/281
- 5.8 Praca nieliniowa dynamiczna/289
 - 5.8.1 Procesy przejściowe w tranzystorze/292
 - 5.8.2 Czasy przełączania/296
 - 5.8.3 Modele nieliniowe dynamiczne/297
- 5.9 Modele i parametry dla pracy z małym sygnałem/303
 - 5.9.1 Modele czwórnikowe/303
 - 5.9.2 Modele fizyczne (schematy zastępcze)/311
- 5.10 Częstotliwości graniczne tranzystora bipolarnego/320
 - 5.10.1 Częstotliwości graniczne — interpretacja fizyczna/320
 - 5.10.2 Częstotliwości graniczne — definicje techniczne/338
 - 5.10.3 Częstotliwości graniczne wyrażone przez elementy modelu „hybrid π ”/339
- 5.11 Metody wyznaczania parametrów modelu „hybrid π ”/340
- 5.12 Wybrane zagadnienia technologii i konstrukcji tranzystorów bipolarnych/341
 - 5.12.1 Rezystancja rozproszona bazy/342
 - 5.12.2 Tranzystory dla zakresu b.w.cz./343
 - 5.12.3 Tranzystory mocy/346
- Zadania/351

6

TRANZYSTORY POŁOWE

stronica 353

- 6.1 Wstęp/353
- 6.2 Tranzystory połowe ze złączem *p-n* (PNFET)/358
 - 6.2.1 Zasady działania tranzystora PNFET — opis jakościowy/360
 - 6.2.2 Charakterystyki i parametry statyczne — analiza ilościowa/367
 - 6.2.3 Praca dynamiczna nieliniowa/375
 - 6.2.4 Praca z małymi sygnałami/378
- 6.3 Tranzystor MIS/382
 - 6.3.1 Budowa i zasady działania tranzystora MIS — opis jakościowy/383
 - 6.3.2 Charakterystyki i parametry statyczne — analiza ilościowa/392
 - 6.3.3 Praca dynamiczna nieliniowa/398
 - 6.3.4 Praca z małymi sygnałami/402

- 6.4 Tranzystory MESFET/403
- 6.5 Tranzystory cienkowarstwowe/404
- Zadania/406

7

TRANZYSTORY W UKŁADACH ANALOGOWYCH

stronica 408

- 7.1 Sposoby polaryzacji tranzystorów bipolarnych/410
 - 7.1.1 Polaryzacja stałym prądem bazy/413
 - 7.1.2 Polaryzacja stałym prądem emitera/414
 - 7.1.3 Polaryzacja w pętli napięciowego sprzężenia zwrotnego/414
 - 7.1.4 Polaryzacja mieszana/415
- 7.2 Podstawowe układy wzmacniające tranzystora bipolarnego/416
 - 7.2.1 Parametry robocze czwórnika liniowego/418
 - 7.2.2 Wzmacniacz w układzie WE/420
 - 7.2.3 Wzmacniacz w układzie WC (wtórnik emiterowy)/422
 - 7.2.4 Wzmacniacz w układzie WB/424
- 7.3 Sposoby polaryzacji tranzystorów polowych/425
- 7.4 Podstawowe układy wzmacniające z tranzystorami polowymi/426
 - 7.4.1 Układ ze wspólnym źródłem/427
 - 7.4.2 Układ ze wspólnym drenem/428
 - 7.4.3 Układ ze wspólną bramką/430
- 7.5 Tranzystor MOS jako obciążenie rezystancyjne/432
- 7.6 Wzmacniacz z tranzystorem MOS i z tranzystorem bipolarnym (para hybrydowa)/435
 - 7.6.1 Wzmacniacz ze stopniem wyjściowym w układzie WE/436
 - 7.6.2 Wzmacniacz ze stopniem wyjściowym w układzie WB/437
 - 7.6.3 Wzmacniacz ze stopniem wyjściowym w układzie WC/438
- 7.7 Tranzystor MOS jako rezystor sterowany napięciem/440

8

TECHNOLOGIA WYTWARZANIA UKŁADÓW SCALONYCH

stronica 443

- 8.1 Półprzewodnikowe układy scalone/447
 - 8.1.1 Układy scalone bipolarne/447
 - 8.1.2 Układy scalone unipolarne/464
- 8.2 Układy warstwowe/470
 - 8.2.1 Układy cienkowarstwowe/470
 - 8.2.2 Układy grubowarstwowe/472
- 8.3 Porównanie technologii półprzewodnikowej, cienkowarstwowej i grubowarstwowej/474

9

UKŁADY SCALONE ANALOGOWE

stronica 475

- 9.1 Budowa układów elementarnych/477
 - 9.1.1 Układy polaryzacji/478
 - 9.1.2 Układy obróbki sygnału/482
- 9.2 Wzmacniacze operacyjne/490
 - 9.2.1 Podstawowe parametry wzmacniaczy operacyjnych/491
 - 9.2.2 Rodzaje wzmacniaczy operacyjnych/494
 - 9.2.3 Wzmacniacz operacyjny μA 741/494
 - 9.2.4 Zastosowanie wzmacniaczy operacyjnych/499
- 9.3 Komparatory napięcia/501
- 9.4 Stabilizatory napięcia/502

10	UKŁADY SCALONE CYFROWE	stronica 505
10.1	Podstawowe funkcje logiczne/505	
10.2	Bramki logiczne bipolarne/511	
	10.2.1 Bramki logiczne z tranzystorami nasyconymi/511	
	10.2.2 Bramki logiczne z tranzystorami nienasyconymi/524	
10.3	Bramki logiczne MOS/527	
	10.3.1 Bramki logiczne statyczne NMOS/528	
	10.3.2 Bramki logiczne komplementarne/531	
10.4	Porównanie bramek logicznych/534	
10.5	Złożone układy logiczne/535	
10.6	Pamięci półprzewodnikowe/536	
	10.6.1 Komórki dynamiczne rejestrów przesuujących/536	
	10.6.2 Komórki dynamiczne pamięci RAM/538	
	10.6.3 Komórki pamięci stałych (ROM)/540	
10.7	Przyrządy o sprzężeniu ładunkowym/542	
11	ELEMENTY PRZELĄCZNIKOWE	stronica 546
11.1	Tranzystor jednozłączowy/546	
	11.1.1 Budowa i zasada działania/547	
	11.1.2 Podstawowe parametry/549	
	11.1.3 Zastosowania/550	
11.2	Tyrystory/551	
	11.2.1 Tyrystory o działaniu jednokierunkowym (struktury <i>p-n-p-n</i>)/553	
	11.2.2 Tyrystor dwukierunkowy (struktura pięciowarstwowa <i>n-p-n-p-n</i>)/560	
	11.2.3 Parametry tyrystorów/560	
	11.2.4 Zastosowania/561	
12	DIODY MIKROFALOWE Z UJEMNĄ REZYSTANCJĄ DYNAMICZNĄ	stronica 562
12.1	Diody TE/563	
	12.1.1 Ujemna konduktancja dynamiczna/563	
	12.1.2 Powstawanie i unoszenie domeny/566	
	12.1.3 Trzy główne rodzaje pracy (mody) diod TE/569	
12.2	Diody przelotowe/570	
	12.2.1 Diody lawinowo-przelotowe/570	
	12.2.2 Diody iniekcyjno-przelotowe/573	
12.3	Porównanie diod z ujemną rezystancją dynamiczną/573	
	Dodatek A Wpływ temperatury na właściwości tranzystorów/576	
	Dodatek B Szумы w tranzystorach/580	
	Dodatek C Oznaczenia elementów półprzewodnikowych i układów scalonych/585	
	Dodatek D Wybrane stałe fizyczne i jednostki/590	
	Literatura/591	
	Skorowidz/593	

OD AUTORA

Tradycyjny i niejako naturalny podział specjalizacji elektroników na cztery grupy, tj. materiałową, elementową (podzespołową), układową i systemową, w ostatnich latach znacznie się zdezaktualizował. Proces „przenikania się” poszczególnych specjalizacji, zapoczątkowany pojawieniem się układów scalonych, narasta w miarę rozwoju mikroelektroniki. Na przykład od konstruktora układów scalonych wymaga się zarówno znajomości fizyki oraz technologii przyrządów półprzewodnikowych (wiedzy materiałowej i elementowej), jak i dobrej orientacji w zagadnieniach układowych. Z kolei konstruktor systemów elektronicznych, mający zwykle do dyspozycji funkcjonalnie alternatywne podsystemy (układy scalone), wytwarzane różnymi technologiami, dokonuje świadomego wyboru określonego rodzaju układów scalonych, kierując się wiedzą o podstawowych właściwościach poszczególnych technologii. Skoro aktualność tradycyjnego ujęcia specjalizacji w elektronice budzi wątpliwość, to można po prostu podzielić wszystkich elektroników na „producentów” i użytkowników elementów (układów scalonych), chociaż i ten podział nie zawsze jest jednoznaczny. Książka ta jest napisana z myślą o użytkownikach elementów (układów scalonych), lecz może być również przydatna dla kandydatów na „producentów” jako kurs wprowadzający.

Na sposobie prezentacji materiału zaważyło założenie, że powinna to być książka do samodzielnego studiowania, której odbiorcami będą zarówno słuchacze kursów elektronicznych, jak i technicy, studenci oraz inżynierowie reprezentujący specjalizacje pokrewne elektronice. Dlatego za najważniejszy uznano styl zbliżony do konspektu z wykładów. Jest to styl niezbyt zwięzły, świadomie zbliżony do języka „mówionego”, dopuszczający dygresje, a nawet uwagi żartobliwe. Szczególny nacisk położono na słowną i graficzną interpretację zjawisk fizycznych przy jak najoszczędniejszym użyciu dość elementarnego aparatu matematycznego. W książce jest zatem znacznie więcej „sosu fizycznego” niż matematyki. Wszelkie zamieszczone wzory matematyczne starano się należycie objaśnić i wyprowadzić. Zastosowano — w miarę możliwości — jednolity układ treści poszczególnych rozdziałów. Każdy rozdział zaczyna się od wstępu zawierającego wyjaśnienie podstawowych pojęć, klasyfikacje, ciekawostki historyczne, przede wszystkim zaś prezentującego treść rozdziału w ogólnym zarysie. Na początku każdego rozdziału zamieszczono rysunek przedstawiający schematycznie budowę rozdziału. Nie jest to tylko graficzny obraz spisu treści. Chodzi o to, aby Czytelnik mógł zorientować się na wstępie, w jaki sposób poszczególne zagadnienia są powiązane w logicznie spójną jednostkę tematyczną — rozdział. Opis poszczególnych zagadnień w dalszej części rozdziału jest prowadzony dwustopniowo, tj. najpierw przedstawia się elementarne ujęcie jakościowe, po czym następuje w miarę dokładny opis z analizą ilościową.

Zarys tematyczny książki ograniczono do tych przyrządów półprzewodnikowych i układów scalonych, których działanie jest oparte wyłącznie na zjawiskach elektrycznych. Nie uwzględniono więc przyrządów magnetoelektrycznych, fotoelektrycznych itp. Książka składa się z dwunastu rozdziałów, które można pogrupować w cztery części. Rozdziały 1 do 3 stanowią część pierwszą, zawierającą podstawy fizyki i technologii półprzewodników. Znajomość podstawowych właściwości fizycznych półprzewodników (rozdział 1), metod wytwarzania materiałów i struktur półprzewodnikowych (rozdział 2) oraz zjawisk kontaktowych i powierzchniowych w półprzewodnikach (rozdział 3) jest niezbędna do studiowania dalszych rozdziałów książki. Rozdziały 4 do 7 stanowią część drugą książki, w której rozpatrzono najważniejsze rodzaje diod i tranzystorów. Rozdział 4 zawiera przegląd podstawowych rodzajów diod półprzewodnikowych, przy czym dla lepszej ilustracji właściwości funkcjonalnych rozpatruje się nie tylko parametry, charakterystyki i modele, lecz również zastosowania tych elementów. W rozdziałach 5 i 6 omówiono tranzystory bipolarne i unipolarne ze szczególnym wyeksponowaniem zagadnień związanych z modelowaniem. Właściwości elementarnych układów włączenia tych tranzystorów są rozważane w rozdziale 7. Część trzecią książki, dotyczącą układów scalonych, stanowią rozdziały 8 do 10, w których rozpatrzono technologię wytwarzania układów scalonych (rozdział 8) oraz podstawowe rodzaje układów scalonych analogowych (rozdział 9) i cyfrowych (rozdział 10). Ostatnie dwa rozdziały tworzą część czwartą, w której rozpatrzono niekonwencjonalne rodzaje przyrządów półprzewodnikowych, tj. elementy przełącznikowe (rozdział 11) oraz diody mikrofalowe z ujemną rezystancją dynamiczną (rozdział 12). Na końcu książki przedstawiono w dodatkach wybrane zagadnienia szczegółowe oraz podano wykaz książek wydanych w kraju, które mogą służyć do bardziej zaawansowanych studiów poszczególnych zagadnień.

Warto podkreślić, co jest niemal truizmem, że niewiele dziedzin nauki i techniki rozwija się tak szybko jak elektronika. Dlatego unikano podawania w książce szybko dezaktualizujących się szczegółów technicznych, a każde zagadnienie starano się przedstawić z zachowaniem odpowiedniego stopnia ogólności rozważań. Autor ma nadzieję, że dzięki temu książka będzie przydatna nie tylko jako źródło informacji o istniejących przyrządach półprzewodnikowych, lecz również jako podręcznik przygotowujący Czytelnika do samodzielnego studiowania literatury fachowej i poznawania na bieżąco nowych elementów i układów scalonych.

Miłym obowiązkiem autora jest podziękowanie prof. drowi hab. inż. Andrzejowi Ambroziakowi za niezwykle cenne wskazówki przekazane w bardzo wnikliwej opinii. Duży wpływ na ujęcie tematu i ostateczną redakcję książki miały dyskusje z kolegami, m. grem inż. Jerzym Pasierbińskim i drem inż. Mirosławem Ruskiem, którzy przeczytali obszerne fragmenty rękopisu tej pracy i poczynili wiele pomocnych uwag oraz konstruktywnych sugestii. Autor wyraża wdzięczność wszystkim, którzy przyczynili się do powstania tej książki.

WYKAZ WAŻNIEJSZYCH OZNACZEŃ

- a — gradient koncentracji domieszek w złączu liniowym
 A — pole powierzchni
 $A(\omega)$ — charakterystyka amplitudowa
 B — baza
 C — kolektor
 C — pojemność
 C_d, C_{dc}, C_{de} — pojemność dyfuzyjna, dyfuzyjna kolektora, dyfuzyjna emitera
 C_{gb}, C_{gd}, C_{gs} — pojemność bramka-podłoże, bramka-dren, bramka-źródło
 C_i — pojemność dielektryka (na jednostkę powierzchni)
 C_j, C_{jc}, C_{je} — pojemność warstwy zaporowej (złączowa), złączowa kolektora, złączowa emitera
 C_g — pojemność dielektryka pod całkowitą powierzchnią bramki
 C_s — pojemność obszaru ładunku przestrzennego przy powierzchni półprzewodnika
 d — odległość, grubość
 d_0 — odległość między sąsiednimi atomami w kryształce
 D — dren
 D — współczynnik dyfuzji
 D_0 — dyfuzyjność
 D_n, D_p — współczynnik dyfuzji elektronów, dziur
 E — emiter
 E — natężenie pola elektrycznego
 f — częstotliwość
 $f_n(W), f_p(W)$ — funkcje rozkładu prawdopodobieństwa zajęcia poziomu W przez elektron, dziurę
 $f_T, f_{\alpha}, f_{\beta}, f_1, f_{max}$ — częstotliwości graniczne
 F — gęstość strumienia atomów
 F — współczynnik szumów
 g_d — konduktancja dynamiczna
 g_m — transkonduktancja
 g_r — konduktancja różniczkowa złącza (diody)
 G — bramka
 G — szybkość generacji nośników
 G_0 — szybkość generacji w stanie równowagi termodynamicznej
 \hat{h} — macierz mieszana
 H — współczynnik tłumienia sygnału wspólnego
 h_i, h_u — czułość prądowa, napięciowa

- i — wartość chwilowa natężenia prądu¹⁾
 i_{rr} — prąd ustalania charakterystyki wstecznej
 I_A, I_K — prąd anodowy, katodowy
 I_B, I_C, I_E — prąd stały bazy, kolektora, emitera
 I_{CS}, I_{ES} — prądy nasycenia złączy kolektor-baza, emiter-baza
 I_{CO}, I_{EO} — prądy zerowe kolektora, emitera
 I_D, I_S, I_G — prąd stały drenu, źródła, bramki
 I_F — prąd przewodzenia diody
 I_R — prąd wsteczny diody
 J — gęstość prądu
 J_c — gęstość prądu przesunięcia (pojemnościowego)
 J_g — gęstość prądu generacji
 J_{nd}, J_{pd} — gęstość prądu dyfuzji elektronów, dziur
 J_{nu}, J_{pu} — gęstość prądu unoszenia elektronów, dziur
 J_s — gęstość prądu nasycenia
 J_t — gęstość prądu całkowitego
 J_u — gęstość prądu unoszenia
 $k = 1,38 \cdot 10^{-23} \text{ J/K} = 8,62 \cdot 10^{-5} \text{ eV/K}$ — stała Boltzmanna
 k_i, k_p, k_u — wzmożenie prądowe, mocy, napięciowe
 K — katoda
 l_d — szerokość warstwy zaporowej (obszaru zubożonego)
 L — indukcyjność
 L — straty przemiany
 L_C, L — długość kanału
 L_D — droga Debye'a
 L_D, L_{DD} — droga dyfuzji domieszek
 L_n, L_p — długość dyfuzyjna elektronów, dziur
 m_n^*, m_p^* — masa efektywna elektronów, dziur
 M — całkowita liczba atomów domieszki
 M_p — współczynnik powielania lawinowego
 n, n_i, n_n, n_p, n_0 — koncentracja elektronów, samoistnych, w półprzewodniku typu n (większościowych), w półprzewodniku typu p (mniejszościowych), równowagowa
 n' — koncentracja elektronów nadmiarowych
 N, N_A, N_D, N_{AD} — koncentracja domieszek, akceptorowych, donorowych, wypadkowa ($N_A - N_D$)
 N_c, N_v — efektywna gęstość stanów energetycznych, sprowadzona do dna pasma przewodnictwa, wierzchołka pasma walencyjnego
 N_t — koncentracja centrów generacyjno-rekombinacyjnych
 p, p_i, p_n, p_p, p_0 — koncentracja dziur, samoistnych, w półprzewodniku typu n (mniejszościowych), w półprzewodniku typu p (większościowych), równowagowych
 p' — koncentracja dziur nadmiarowych
 P_a — moc admisyjna

¹⁾ Prądy i napięcia są oznaczane w następujący sposób: duża litera z dużą literą w indeksie oznacza wielkość stałą (na przykład I_B, U_{EB}); duża litera z małą literą w indeksie oznacza amplitudę składowej zmiennej (na przykład I_b, U_{eb}); mała litera z dużą literą w indeksie oznacza całkowitą (składowa stała + składowa zmienna) wartość chwilową (na przykład i_B, u_{EB}); mała litera z małą literą w indeksie oznacza wartość chwilową składowej zmiennej (na przykład i_b, u_{eb}).

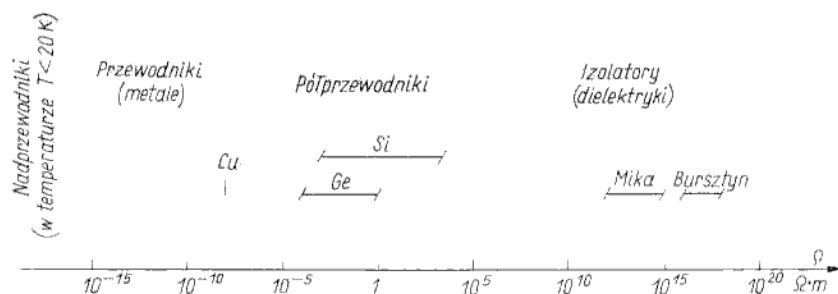
- $q = 1,6 \cdot 10^{-19} \text{C}$ — ładunek jednostkowy
 Q — dobroć
 q_b — ładunek nadmiarowy w bazie
 Q_G — ładunek w bramce struktury MOS
 Q_f — ładunek warstwy zaporowej
 Q_n, Q_p — ładunek elektronów, dziur
 Q_s — ładunek w półprzewodniku
 Q_{sr} — ładunek równoważny stanów powierzchniowych
 r, R — rezystancja
 $r_{bb'}$ — rezystancja rozproszona bazy
 r_z — rezystancja dynamiczna diody Zenera
 R, R_0 — szybkość rekombinacji nośników, w stanie równowagi termodynamicznej
 R_d — rezystancja nieliniowa warstwy zaporowej
 R_L — rezystancja obciążenia
 R_s — rezystancja szeregową złącza
 R_{sn}, R_{sp} — rezystancja szeregową obszaru obojętnego elektrycznie typu n , typu p
 R_u — rezystancja upływu złącza
 R_{th} — opór cieplny
 s — szybkość rekombinacji powierzchniowej
 S — źródło
 S_β, S_i, S_u — współczynniki stabilizacji
 $t, t_d, t_r, t_f, t_{rr}, t_s, t_{on}, t_{off}$ — czas, opóźnienia, narastania, opadania, przełączania, magazynowania, włączania, wyłączania
 T, T_a, T_j — temperatura, otoczenia, złącza
 u — wartość chwilowa napięcia
 U_B, U_C, U_E — napięcie stałe bazy, kolektora, emitera
 U_{BE}, U_{CE} , itp. — napięcie stałe baza-emiter, kolektor-emiter itp.
 $U_{CE(sat)}$ — napięcie nasycenia w tranzystorze bipolarnym
 U_D, U_G, U_S — napięcie stałe drenu, bramki, źródła
 U_{Dsat} — napięcie nasycenia w tranzystorze polowym
 U_{FB} — napięcie wyprostowanych pasm w strukturze MIS
 U_L — napięcie na obciążeniu
 U_p — napięcie przebicia, odcięcia
 U_T — napięcie progowe w tranzystorze MOS
 U_{TO} — napięcie progowe diody (zwykle przyjmuje się $U_{TO} = \varphi_B$)
 U_{WE}, U_{WY} — napięcie wejściowe, wyjściowe
 U_Z — napięcie Zenera (stabilizacji)
 v_{th} — prędkość cieplna nośników
 v_u — prędkość unoszenia nośników
 W — energia
 W_B — szerokość bazy
 W_c, W_v — energia dna pasma przewodnictwa, wierzchołka pasma walencyjnego
 W_F, W_i — energia (poziom) Fermiego, poziom Fermiego w półprzewodniku samoistnym (energia środka pasma zabronionego)
 W_g — szerokość pasma zabronionego
 WB, WC, WE — układ ze wspólną bazą, kolektorem, emiterem
 WD, WG, WS — układ ze wspólnym drenem, bramką, źródłem
 Y — admitancja

- Z — impedancja
 α — współczynnik wzmocnienia prądowego (małosygnalowy) w układzie WB
 α_b — współczynnik transportu nośników przez bazę
 α_e — współczynnik sprawności wstrzykiwania emitera
 α_I, α_N — współczynnik wzmocnienia prądu stałego w układzie WB, w kierunku inwersyjnym, w kierunku normalnym
 β — współczynnik wzmocnienia prądowego (małosygnalowy) w układzie WE
 β_I, β_N — współczynnik wzmocnienia prądu stałego w układzie WE, w kierunku inwersyjnym, w kierunku normalnym
 β^d, β^s — parametr $\beta = W\mu C_i/L_C$ dla tranzystora sterowanego, obciążającego
 $\epsilon_0 = 8,85 \cdot 10^{-12}$ F/m — przenikalność elektryczna próżni
 ϵ_i, ϵ_s — przenikalność elektryczna bezwzględna dielektryka, półprzewodnika
 $\epsilon_r, \epsilon_{ri}, \epsilon_{rs}$ — przenikalność elektryczna, względna dielektryka, półprzewodnika
 $\eta = J_R/J_F$
 η — współczynnik profilu koncentracji domieszek
 η — sprawność transportu ładunku w przyrządach CCD
 $\varphi, \varphi_F, \varphi_s$ — potencjał Fermiego, powierzchniowy
 φ_B — bariera potencjału (napiecie dyfuzyjne)
 φ_i — spadek napięcia w warstwie dielektryka
 φ_{ms} — kontaktowa różnica potencjałów metal-półprzewodnik
 $\varphi_T = kT/q$ — potencjał elektrokinetyczny
 Φ_m, Φ_s — potencjał wyjścia elektronu z metalu, półprzewodnika
 χ_i, χ_s — powinowactwo elektronowe w dielektryku, półprzewodniku
 λ — długość fali
 μ, μ_n, μ_p — ruchliwość, elektronów, dziur
 ρ — rezystywność
 ρ — gęstość ładunku
 σ — konduktywność
 τ, τ_d — stała czasowa, relaksacji dielektrycznej
 τ_b — czas przelotu nośników przez bazę
 τ_r, τ_n, τ_p — czas życia nośników (ogólnie), dziur, elektronów
 ω — pulsacja
 $\omega_x, \omega_\beta, \omega_T, \omega_1, \omega_{max}$ — pulsacje graniczne

W tym rozdziale rozpatrzono niektóre elementarne zagadnienia fizyki półprzewodników, których znajomość jest niezbędna do zrozumienia działania przyrządów półprzewodnikowych.

Słowa *półprzewodnik*, *tranzystor* zrobiły w ostatnim ćwierćwieczu zawrotną karierę. Te jeszcze nie tak dawno specjalistyczne terminy weszły obecnie do języka potocznego, który — kierując się swymi prawami — zniekształcił ich prawidłowe znaczenie, gdyż potocznie półprzewodnikiem jest nazywany tranzystor, a tranzystorem przenośne radio tranzystorowe. Prawdopodobnie uniknie się jakichkolwiek nieporozumień zaznaczając od razu, że półprzewodnikiem będzie nazywany materiał półprzewodnikowy, przyrządem półprzewodnikowym natomiast — element wykonany z tego materiału, tj. dioda, tranzystor itp. By ta uwaga miała pełny sens, należy jeszcze odpowiedzieć na pytanie: „Co to jest półprzewodnik?”

Najprostsza definicja brzmi: „Półprzewodniki są materiałami, których rezystywność jest większa niż rezystywność przewodników (metali) oraz mniejsza niż rezystywność izolatorów (dielektryków)” (rys. 1.1). Na przykład miedź, doskonały



Rys. 1.1
Podział materiałów elektronicznych ze względu na ich rezystywność

przewodnik, ma rezystywność $10^{-8} \Omega \cdot m$; mika, bardzo dobry izolator, ma rezystywność ok. $10^{14} \Omega \cdot m$; czysty krzem, najpowszechniej stosowany materiał półprzewodnikowy, ma rezystywność ok. $2 \cdot 10^3 \Omega \cdot m$.

Takie kryterium klasyfikacji materiałów elektronicznych nie wyczerpuje jednak istoty zagadnienia, gdyż istnieje jakościowa różnica między właściwościami elektrofizycznymi przewodników i półprzewodników, dielektryki natomiast można odróżnić od półprzewodników tylko na podstawie umownego kryterium ilościowego.

wego, a nie różnic jakościowych. Podstawowe właściwości różniące w sposób jakościowy półprzewodniki (i dielektryki) od przewodników są następujące:

- właściwości elektryczne półprzewodników, przede wszystkim ich rezystywność, silnie zależą od znikomo małych ilości zanieczyszczeń materiału (zanieczyszczenia wprowadzane celowo nazywa się domieszkami, a wprowadzanie tych zanieczyszczeń, tj. domieszkowanie, jest podstawowym procesem technologii półprzewodników);
- na rezystywność półprzewodników duży wpływ ma różnego rodzaju promieniowanie zewnętrzne;
- temperaturowy współczynnik rezystancji ma duże ujemne wartości w półprzewodnikach (ze wzrostem temperatury rezystywność maleje ok. 5...10% na 1°C), podczas gdy w przewodnikach ma małe i na ogół dodatnie wartości (ze wzrostem temperatury rezystywność zwiększa się ok. 0,3...0,6% na 1°C).

Te różnice wynikają stąd, że w przewodnikach wszystkie elektrony walencyjne są nośnikami swobodnymi ładunków, nawet w temperaturze 0 K, i mogą uczestniczyć w przepływie prądu, w półprzewodnikach i dielektrykach natomiast elektrony walencyjne są związane z atomami i tylko część z nich po wyrwaniu z orbit walencyjnych bierze udział w przepływie prądu.

A zatem do powstania nośników swobodnych ładunku w półprzewodnikach i dielektrykach (do wyrwania elektronów z orbit walencyjnych atomów) jest konieczne dostarczenie energii z zewnątrz (na przykład energii cieplnej).

A czym różnią się półprzewodniki od dielektryków?

Istnieje tylko ilościowa różnica w wartości energii aktywacji elektronów, tj. energii koniecznej do wyrwania elektronu z orbity walencyjnej atomu do przestrzeni, w której elektron jako nośnik swobodny ładunku może uczestniczyć w przepływie prądu. Energia aktywacji, równoznaczna w czystym materiale z szerokością pasma zabronionego w tzw. energetycznym modelu pasmowym, jest większa dla dielektryków. Za arbitralne kryterium wyznaczające granicę między półprzewodnikami a dielektrykami przyjmuje się na ogół wartość tej energii równą 2 eV¹⁾. Jest oczywiście, że w miarę obniżania temperatury maleje energia cieplna, a więc maleje również prawdopodobieństwo aktywacji elektronów zarówno w półprzewodniku, jak i w dielektryku. Stąd maleje liczba nośników swobodnych i półprzewodnik upodabnia się swymi właściwościami do dielektryka, stając się idealnym izolatorem w temperaturze zera bezwzględnego. Dla zrozumienia fizycznego podłoża wymienionych właściwości półprzewodników należy rozpatrzeć ich budowę oraz mechanizm transportu nośników ładunku.

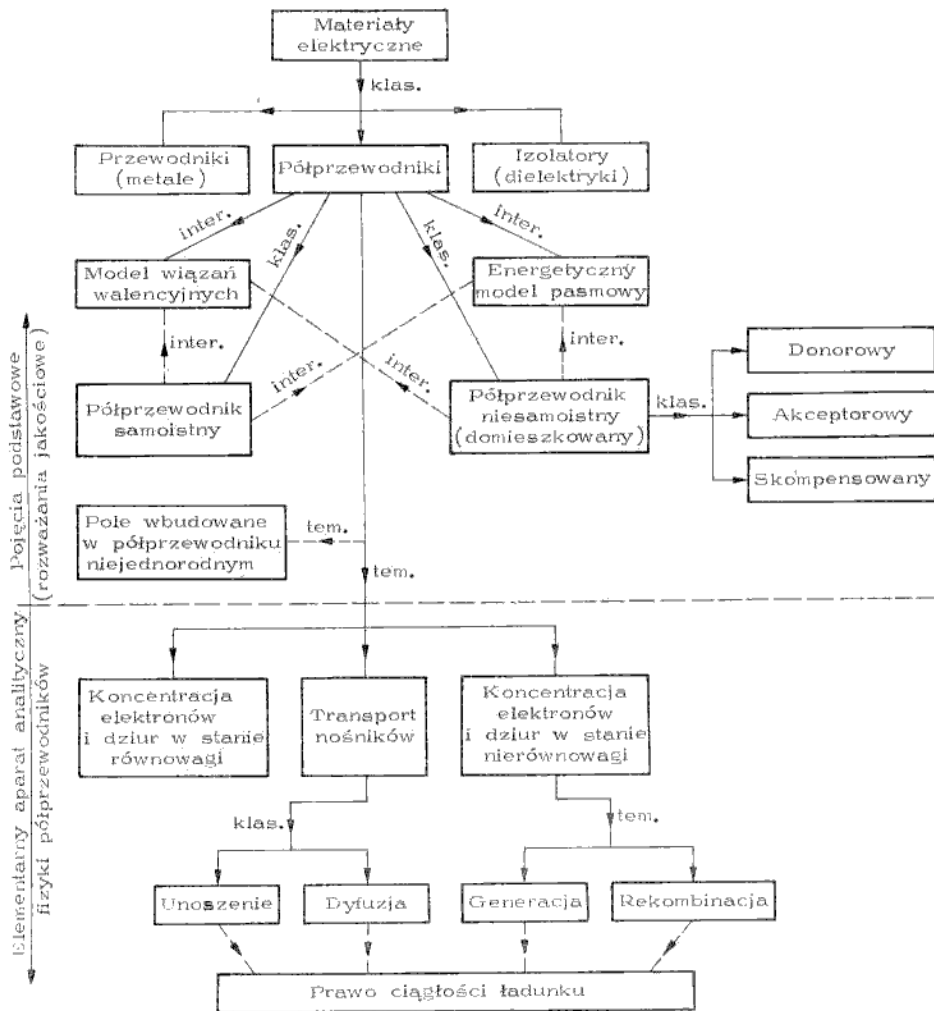
Interpretację zjawisk elektrofizycznych w półprzewodniku można oprzeć na jednym z dwu podstawowych modeli, tj. na jakościowym modelu wiązań walencyjnych lub na energetycznym modelu pasmowym, stosowanym w ilościowym opisie zjawisk. W kolejnych punktach tego rozdziału przedstawiono oba modele oraz wynikające z nich interpretacje poszczególnych zjawisk dla dwu podstawowych materiałów półprzewodnikowych — krzemu i germanu.

Układ pojęciowo-tematyczny rozdziału dla większej przejrzystości przedstawiono schematycznie na rys. 1.2, na którym wyróżniono trzy rodzaje związków między różnymi hasłami. Są to:

- związki typu klasyfikacyjnego (rozbicie pojęcia ogólnego na pojęcia bardziej szczegółowe) — oznaczone skrótem „klas.”;

¹⁾ eV (elektronowolt) jest jednostką energii powszechnie stosowaną w elektronice. Jest to energia uzyskana przez elektron pod wpływem różnicy potencjałów: 1 V (1 eV = 1,602 × 10⁻¹⁹ J).

— związki typu interpretacyjnego, przy czym strzałka jest skierowana od hasła określającego podmiot interpretacji do hasła określającego model (sposób, metodę) interpretacji pewnych właściwości tego podmiotu; takie związki są oznaczone skrótem „inter.”;



Rys. 1.2
Schematyczny układ rozdziału

— związki typu tematycznego (rozbięcie tematu na podtematy); ponieważ jest to podział arbitralny, niepełny i często niejednorodny co do samego kryterium podziału, nie ma on zatem charakteru klasyfikacji, a służy tylko uporządkowaniu materiału w celach dydaktycznych; takie związki oznaczono skrótem „tem.”.

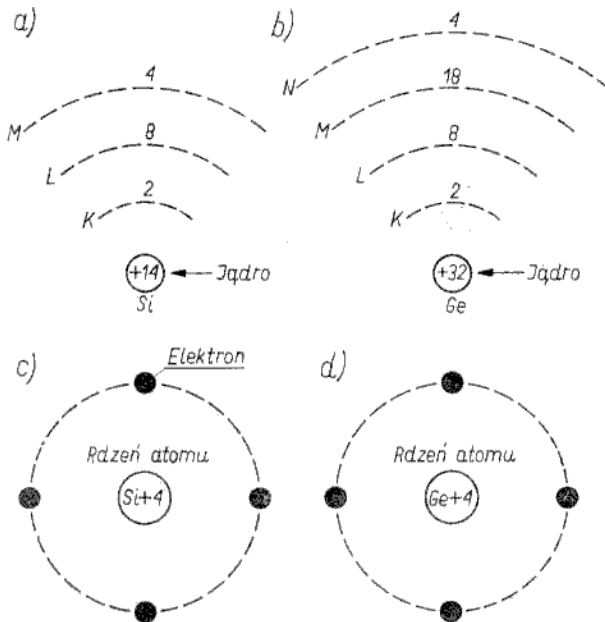
Model wiązań walencyjnych w sieci krystalicznej krzemu i germanu

I.1

Budowa atomu krzemu i germanu

I.1.1

Według najbardziej uproszczonych poglądów atom można sobie przedstawić jako swego rodzaju mikroskopijny odpowiednik systemu słonecznego z ciężkim jądrem (Słońcem) w środku i krążącymi wokół niego po różnych orbitach lekkimi elektronami (planetami). Jądro składa się z neutronów elektrycznie neutralnych oraz protonów dodatnio naładowanych. Każdy elektron ma ładunek ujemny o wartości bezwzględnej dokładnie równej ładunkowi protonu, a ponieważ liczby elektronów i protonów w atomie są sobie równe, więc cały atom jest elektrycznie obojętny. Elektron krążąc wokół jądra doznaje działania siły odśrodkowej, która równoważy siłę przyciągania elektrostatycznego różnoimiennych ładunków protonu i elektronu. Wydawałoby się, że spełnienie warunku równowagi sił dośrodkowej (przyciągania) i odśrodkowej gwarantuje stabilność orbity ruchu elektronu, przy czym może to być dowolna orbita. Jednak tak prosty model stoi w sprzeczności z prawami elektrodynamiki, gdyż elektron jako ładunek elektryczny poruszający się ruchem przyspieszonym (przyspieszenie dośrodkowe) powinien wypromieniowywać energię elektromagnetyczną. Elektron traciłby zatem energię spadając ostatecznie na jądro. Ten paradoks obalili fizyka kwantowa wprowadzając pojęcie tzw. dualizmu korpuskularno-falowego, tj.



Rys. 1.3

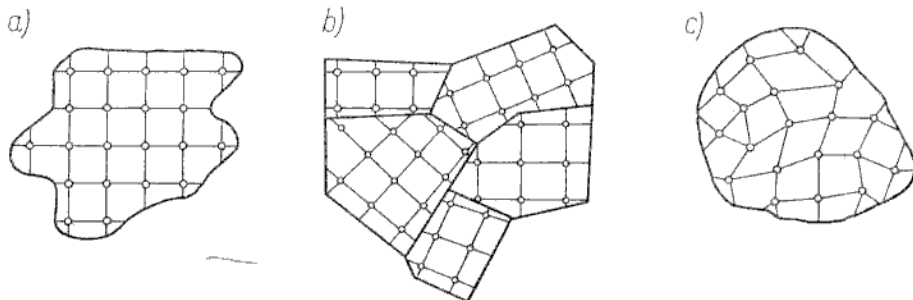
Schematyczna ilustracja struktury atomu krzemu (a, c) i germanu (b, d). Na rysunkach c), d) atom jest przedstawiony w postaci czterech elektronów na orbicie walencyjnej oraz tzw. rdzenia atomu $\text{Si}+4$ lub $\text{Ge}+4$ zawierającego jądro wraz ze wszystkimi elektronami orbit wewnętrznych

utożsamiając elektron z pewną falą i przedstawiając prawa ruchu elektronu w formalizmie falowym. Ponieważ nie ma strat energii w przypadku fali stojącej, dla uniknięcia wspomnianego paradoksu wystarcza postulat, że elektron porusza się tylko po ściśle określonych dozwolonych orbitach, którym odpowiadają fale stojące. Nie zatrzymując się dłużej nad zagadnieniami fizyki atomu przejdziemy do ostatecznej postaci modelu atomu, jaka posłuży nam w dalszych rozważaniach. Otóż elektrony poruszają się po pewnych dozwolonych orbitach, które tworzą tzw. powłoki elektronowe oznaczane literami K, L, M, N, O, P, Q . Liczby elektronów na poszczególnych powłokach podlegają następującemu prawu ogólnemu: $K(1^2 \times 2 = 2)$, $L(2^2 \times 2 = 8)$, $M(3^2 \times 2 = 18)$ itd. Zatem krzem jako czterdziesty (14 protonów i tyleż elektronów) pierwiastek w tablicy periodycznej ma następującą strukturę atomu: na powłoce K — 2 elektrony, na L — 8 oraz na M — 4 (rys. 1.3a). Z kolei german jako 32 pierwiastek ma następującą strukturę atomu: na powłoce K — 2 elektrony, na L — 8, na M — 18, na N — 4 (rys. 1.3b). Elektrony ułożone na zewnętrznej powłoce, tzw. *elektrony walencyjne*, jako najsłabiej związane z jądrem i biorące udział we wszelkich wiązaniach atomu z innymi atomami mają największy wpływ na właściwości określonego pierwiastka. Zarówno atom krzemu, jak i atom germanu mają 4 elektrony walencyjne i to decyduje o dużym podobieństwie właściwości chemicznych, fizycznych i elektrycznych tych dwu materiałów. Ze względu na dominującą rolę elektronów walencyjnych dla uproszczenia rysunków atomy krzemu i germanu będą dalej ilustrowane w identyczny sposób pokazany na rys. 1.3c, d.

Sieć krystaliczna

1.1.2

Stan skupienia materii zależy od sił wzajemnego oddziaływania między atomami bądź cząsteczkami lub jonami. W odróżnieniu od gazów i cieczy w ciałach stałych odległości między atomami są małe, dlatego istnieją duże siły wzajemnego oddziaływania powodujące uporządkowanie atomów w postaci regularnej sieci o periodyczne powtarzalnych w przestrzeni komórkach, nazywanej *siecią krystaliczną*. Jeżeli to periodyczne uporządkowanie budowy jest zachowane w całej bryle materiału, to taki materiał nazywa się *monokrystalicznym* (rys. 1.4a). Jeżeli natomiast bryła składa się jakby z chaotycznie połączonych ziaren, przy czym każde ziarno ma uporządkowaną budowę krystaliczną, to taki materiał nazywa się *polikrystalicznym* (rys. 1.4b). Odrębną grupę ciał stałych tworzą tzw. ciała

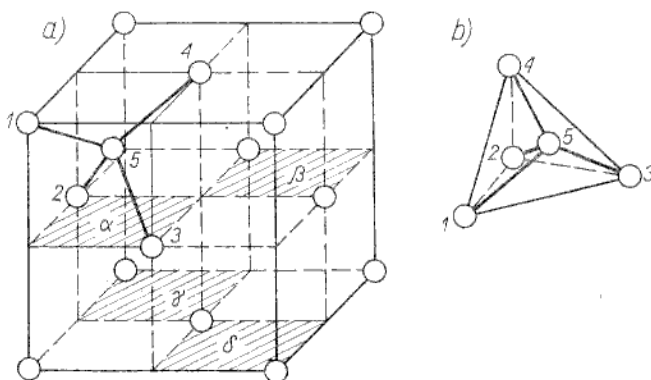


Rys. 1.4

Uproszczone płaski obraz sieci atomów w materiałach: a) monokrystalicznym; b) polikrystalicznym; c) bezpostaciowym

amorficzne (bezpostaciowe), np. szkło, w których istnieje uporządkowanie bliskiego rzędu, co oznacza, że na przestrzeni kilku odległości międzyatomowych występują wyraźne odchylenia od regularnej struktury przestrzennej (rys. 1.4c). Te ciała nie będą w książce rozpatrywane.

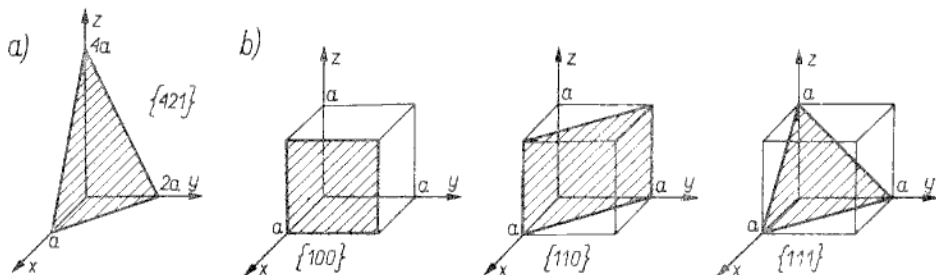
Atomy krzemu oraz germanu tworzą sieć krystaliczną typu diamentu, której komórkę elementarną przedstawiono poglądowo (nie ze wszystkimi szczegółami) na rys. 1.5a. Komórka elementarna jest najmniejszym fragmentem sieci krystalicznej,



Rys. 1.5

Komórka elementarna sieci krystalicznej typu diamentu (a) oraz czworościenny układ sąsiadujących ze sobą atomów (b). W czterech sześcianach, których podstawy oznaczono α , β , γ , δ , są ulokowane atomy centralne, lecz dla większej przejrzystości pokazano tylko atom centralny sześcianu o podstawie α (atom oznaczony cyfrą 5)

powtarzalnym we wszystkich kierunkach przestrzeni tej sieci. Krawędź sześcienniej komórki sieci typu diamentu jest nazywana *stałą sieci krystalicznej*. Oczywiście stała sieci krystalicznej nie jest pojęciem jednoznacznym z najmniejszą odległością między dwoma sąsiednimi atomami. Każdy atom w sieci krzemu i germanu sąsiaduje z czterema innymi atomami tworzącymi układ czworościanu, co dla lepszej ilustracji przedstawiono oddzielnie na rys. 1.5b. Z uwagi na *anizotropię*, to jest zależność właściwości kryształu od kierunku, w wielu przypadkach konieczne jest określenie rozpatrywanej płaszczyzny czy osi w przestrzeni sieci krystalicznej. W tym celu są podawane tzw. *wskaźniki Millera*. W przypadku



Rys. 1.6

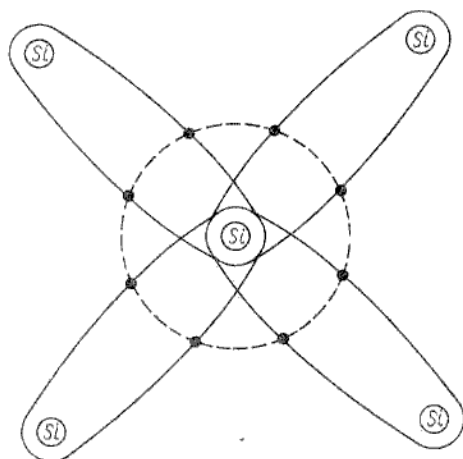
Wskaźniki Millera: a) dla płaszczyzny $\{421\}$ (rozpatrywanej w tekście); b) dla trzech podstawowych płaszczyzn krystalograficznych w sieci krzemu i germanu

rozpatrywanego kryształu sześciennego wskaźniki Millera są to trzy cyfry odnoszące się do prostokątnego układu współrzędnych x, y, z . Aby wyznaczyć wskaźniki Millera, należy określić współrzędne punktów przecięcia opisywanej płaszczyzny z osiami współrzędnych i wyrazić współrzędne tych punktów w jednostkach stałej sieci krystalicznej. Następnie należy wziąć odwrotności tych wielkości i doprowadzić je do wspólnego mianownika, wówczas liczniki otrzymanych ułamków będą wskaźnikami Millera. Na przykład (rys. 1.6a) dla płaszczyzny przecinającej osi x, y, z , we współrzędnych $a, 2a, 4a$ (a — stała sieci krystalicznej) otrzymuje się odwrotności $1, 1/2, 1/4$, skąd po doprowadzeniu do wspólnego mianownika uzyskuje się wskaźniki Millera $\{4\ 2\ 1\}$. Jeśli określana płaszczyzna nie przecina danej osi, to odpowiadający jej wskaźnik Millera ma wartość 0. Na rysunku 1.6b przedstawiono przykładowo wskaźniki Millera dla trzech podstawowych płaszczyzn krystalograficznych sieci krzemu i germanu.

Wiązania międzyatomowe

1.1.3

Zostanie teraz rozpatrzone zagadnienie wiązań międzyatomowych w sieci krystalicznej. W fizyce ciała stałego rozróżnia się następujące rodzaje wiązań: jonowe (inaczej heteropolarne), atomowe (inaczej homopolarne lub kowalencyjne), metaliczne, molekularne, wodorowe. Nas interesuje tylko wiązanie kowalencyjne, gdyż taki typ wiązania istnieje nie tylko w krzemie i germanie, lecz niemal we wszystkich materiałach półprzewodnikowych. Istotę wiązania kowalencyjnego zilustrowano na płaskim modelu atomu krzemu otoczonego czterema atomami sąsiednimi (rys. 1.7). Atomy w kryształe krzemu lub germanu są ułożone



Rys. 1.7
Płaski model wiązań kowalencyjnych atomu krzemu z czterema atomami sąsiednimi

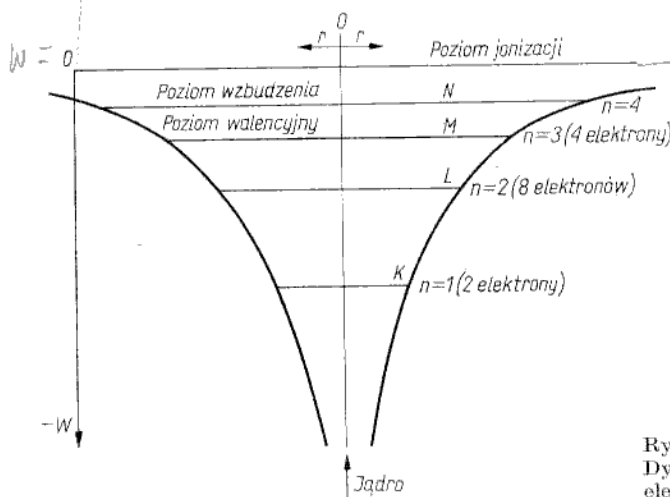
w takich odległościach, że ich powłoki walencyjne zachodzą wzajemnie na siebie, wskutek czego para elektronów walencyjnych sąsiadujących ze sobą dwu atomów staje się jakby wspólna dla obu tych atomów. Niezbyt ściśle, lecz obrazowo można mówić, że para elektronów krąży po wspólnej orbicie okalającej oba atomy. W ten sposób atom wewnętrzny (centralny) na rys. 1.7 uzyskuje trwałą — najbardziej korzystną energetycznie — strukturę ośmioelektronową, gdyż oprócz „własnych” czterech elektronów walencyjnych ma dodatkowe cztery elektrony

wspólne z czterema atomami sąsiednimi. Do wyrwania elektronu walencyjnego z takiego wiązania konieczne jest dostarczenie odpowiedniej energii (1,1 eV dla Si oraz 0,7 eV dla Ge) w postaci ciepła, promieniowania elektromagnetycznego itp. Elektron wyrwany z wiązania kowalencyjnego staje się nośnikiem swobodnym ładunku, który może swobodnie poruszać się w objętości kryształu pod wpływem doprowadzonego z zewnątrz pola elektrycznego. Oderwanie elektronu z wiązania powoduje powstanie luki, czyli częściową jonizację dwu atomów, powstającego przy tym jednostkowego ładunku dodatniego nie należy jednak kojarzyć z którymkolwiek ze związanych w parze atomów, lecz z samym wiązaniem. Taką dodatnio naładowaną lukę w wiązaniu międzyatomowym nazywa się *dziurą*.

Energetyczny model pasmowy półprzewodnika

1.2

Poszczególnym orbitom elektronów w atomie odpowiadają określone wartości energii, rozumianej jako suma energii kinetycznej i potencjalnej tych elektronów. Dlatego mając na myśli orbity elektronów zwykle się mówi o *poziomach* lub *stanach energetycznych*. Elektrony w atomie mogą zajmować tylko ściśle określone dozwolone poziomy energetyczne. Im większy jest promień orbity elektronu, tym większa jest jego energia, gdyż do przejścia elektronu z pewnej orbity na kolejną, bardziej zewnętrzną, konieczne jest dostarczenie określonej porcji (kwantu) energii z zewnątrz. Zatem każdy atom charakteryzuje się pewnym widmem energetycznym. Na rysunku 1.8 przedstawiono przykładowy wykres dyskretnych



Rys. 1.8
Dyskretne poziomy energetyczne elektrony w atomie krzemu

poziomów energetycznych w atomie krzemu wraz z dwiema symetrycznymi krzywymi zależności energii elektronu w funkcji jego odległości od jądra. W tym przypadku dogodnie było przyjąć za bazę odliczania energii poziom energetyczny elektronu „nieskończenie” odległego od jądra (chodzi o taką odległość elektronu od jądra, przy której znika między nimi oddziaływanie elektrostatyczne); ten poziom energetyczny jest nazywany *poziomem jonizacji*. Na rysunku 1.8 pokazano poziomy energetyczne zajęte przez elektrony wraz z dozwolonymi poziomami nie zajętymi. Najwyższy poziom zajęty jest nazywany *poziomem walencyjnym*, a ko-

lejnny nie zajęty poziom o większej energii nosi nazwę *poziomu wzbudzenia*. Zgodnie z ogólnymi prawami materii elektrony w atomie dążą do zajęcia jak najniższych *poziomów energetycznych* (struktura o mniejszej energii jest bardziej stabilna). Jednakże — jak stwierdzono wcześniej — liczba elektronów na poszczególnych poziomach energetycznych (powłokach elektronowych) jest ściśle ograniczona do 2, 8 itd. Jeżeli wszystkie elektrony w atomie zajmują najniższe z możliwych poziomów energetycznych, to mówi się, że atom jest w stanie *normalnym*. Jeżeli pod wpływem jakiegoś bodźca energetycznego elektron lub kilka elektronów zwiększą swą energię i zajmą poziom wzbudzenia, to atom znajdzie się w stanie wzbudzenia.

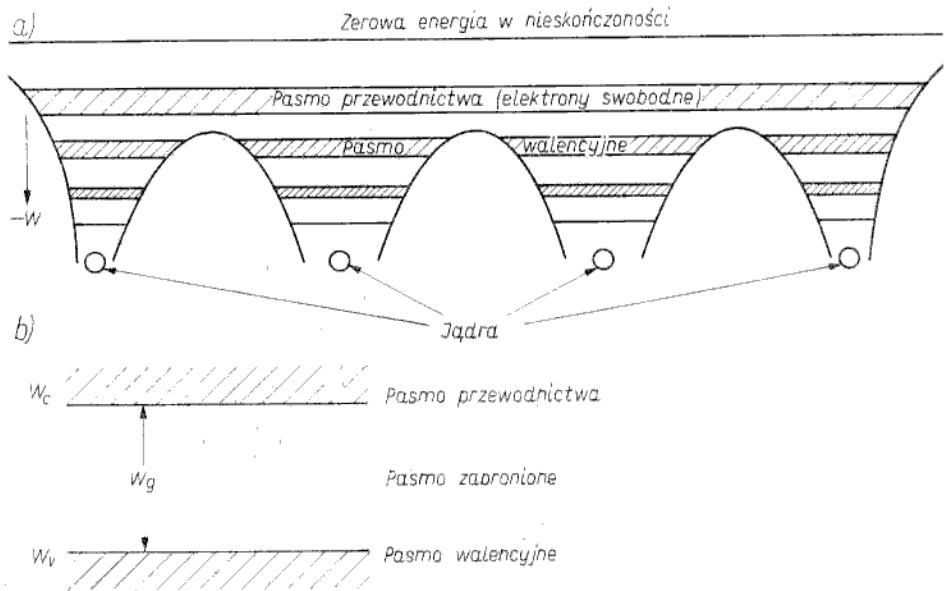
Dla większej jasności dalszych rozważań konieczny jest dodatkowy komentarz do zagadnienia liczby elektronów zajmujących wspólny poziom energetyczny. Jedno z podstawowych praw fizyki atomowej — sformułowana w 1925 r. *zasada Pauliego* — mówi, że w atomie nie ma dwu identycznych elektronów w sensie identyczności ich stanu energetycznego. Bardziej ściśle to prawo brzmi: w atomie nie ma dwu elektronów o jednakowych czterech liczbach kwantowych. Pierwsze trzy liczby kwantowe, oznaczone literami n , l , m , opisują odpowiednio kształt, rozmiary i położenie orbity elektronu w przestrzeni, a czwarta liczba s , nazywana *liczbą spinową*, odnosi się do ruchu obrotowego elektronu¹⁾. Poprzednie zatem stwierdzenie o 2, 8 itd. elektronach na poszczególnych poziomach energetycznych należy rozumieć w ten sposób, że są to poziomy określone z dokładnością do pierwszej liczby kwantowej, nie dające pełnej informacji o stanach energetycznych poszczególnych elektronów. Na przykład 8 elektronów na poziomie odpowiadającym powłoce L to po prostu osiem elektronów o jednakowej liczbie kwantowej n (w tym przypadku $n = 2$), których stany energetyczne jednakże różnią się wzajemnie od siebie kolejnymi liczbami kwantowymi.

Dotychczas omawiano widmo energetyczne pojedynczego atomu. Widma energetyczne atomów jednego pierwiastka są identyczne. Można by stąd wnioskować, że zbiór atomów jednego pierwiastka można opisać takim samym modelem energetycznym jak dla pojedynczego atomu, przy zwiększeniu liczb elektronów na poszczególnych poziomach tyle razy, ile jest atomów w rozpatrywanym zbiorze. Okazuje się, że takie podejście jest słuszne tylko wówczas, gdy odległości między atomami są na tyle duże, że można pominąć ich wzajemne oddziaływanie. W przeciwnym przypadku należy rozszerzyć działanie zasady Pauliego na cały zbiór atomów; wówczas zasadę tę można sformułować w następujący sposób: w zbiorze wzajemnie oddziałujących na siebie atomów nie ma dwu elektronów o identycznych stanach energetycznych. Taki przypadek istnieje w kryształach półprzewodnika, gdyż — jak wcześniej już stwierdzono — każdy atom w sieci krystalicznej jest powiązany z sąsiednimi atomami przez oddziaływania kowalencyjne. Stąd struktura energetyczna kryształu będzie inna niż dla pojedynczego atomu, gdyż z poszczególnych pojedynczych poziomów energetycznych odosobnionego atomu utworzy się tyle różnych poziomów, ile jest atomów w kryształach. W tej sytuacji zwykło się mówić o *rozszczeniu poziomu energetycznego*. Należy zauważyć, że różnice energetyczne między poziomami powstałymi z rozszczepienia pojedynczego poziomu energetycznego są znikomo małe. Na przykład w kryształach o objętości 1 cm^3 jest ok. 10^{23} atomów, czyli w strukturze energetycznej

¹⁾ Taka interpretacja liczby spinowej (ang. *spin* — ruch wirowy, obrotowy) jest daleko idącym uproszczeniem dydaktycznym. Ściśle biorąc, spin jest właściwością pierwotną elektronu, taką jak masa i ładunek, której nie można wytłumaczyć w sposób poglądowy, tj. nie można sprowadzić do pojęć bardziej pierwotnych (na gruncie fizyki kwantowej).

nej tego kryształu każdy poziom energetyczny pojedynczego atomu zostanie rozszczepiony na 10^{23} poziomów¹⁾. Jeśli przyjmie się, że odstęp energetyczny między skrajnymi poziomami w pasmie wynosi 1 eV, to — średnio biorąc — odstęp między dwoma sąsiednimi poziomami wynosi 10^{-23} eV. Jest to tak mała wartość energii, że można mówić o prawie ciągłym pasmie dozwolonych energii.

W ten sposób z modelu dyskretnych poziomów energetycznych w pojedynczym atomie przechodzimy do energetycznego modelu pasmowego w kryształach. Typowy wykres energetycznego modelu pasmowego ciała stałego przedstawiono na rys. 1.9a.



Rys. 1.9
Energetyczny model pasmowy ciała stałego (a) oraz jego uproszczona ilustracja (b)

Należy zauważyć, że bariera energetyczna między sąsiednimi atomami jest mniejsza niż na krawędziach kryształu. Obniżenie bariery energetycznej między poszczególnymi atomami jest jednoznaczne z istnieniem wspólnych elektronów walencyjnych. Ponieważ właściwości elektryczne półprzewodnika są uwarunkowane zjawiskami zachodzącymi w pasmie walencyjnym i kolejnym wyższym pasmie dozwolonych energii (w pasmie przewodnictwa), dalej będziemy się posługiwać bardziej uproszczonym wykresem energetycznego modelu pasmowego, przedstawionym na rys. 1.9b. Odstęp W_g między wierzchołkiem pasma walencyjnego W_v a dnem pasma przewodnictwa W_c jest nazywany *przerwą energetyczną* lub *pasmem zabronionym*²⁾. W temperaturze zera bezwzględne pasmo walencyjne jest całkowicie wypełnione elektronami, co odpowiada efektywnie ośmioelektronowej strukturze poziomu walencyjnego w atomie związanym kowalencyjnie z innymi atomami w sieci krystalicznej. Pasma przewodnictwa

¹⁾ Wyjątkowo posłużono się wymiarem cm^3 , gdyż kryształ o objętości 1 m^3 jest pozbawiony sensu technicznego — nie ma tak dużych kryształów.

²⁾ Indeksy w oznaczeniach W_v , W_c pochodzą z języka angielskiego, tj. *g* — *gap* (przerwa), *v* — *valence* (walencyjny), *c* — *conduction* (przewodnictwo).

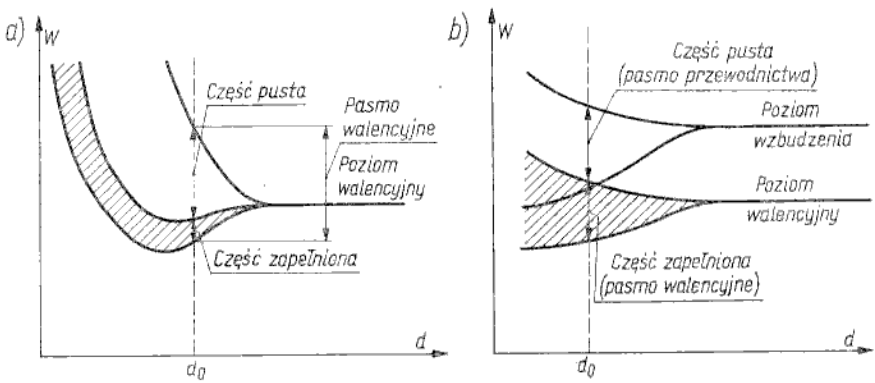
liczba elektronów a tym samym będzie tym większa im wyższe tempo i mniejsza szerokość pasma zabronionego

natomiast jest w tych warunkach całkowicie puste. W miarę wzrostu temperatury część elektronów z pasma walencyjnego „przeskakuje” do pasma przewodnictwa pozostawiając w pasmie walencyjnym wolne miejsca nazywane *dziurami*. Taki proces jest nazywany *generacją par elektron-dziura*. Szerokość pasma zabronionego jest wartością energii, jaką trzeba dostarczyć do sieci krystalicznej dla uwolnienia elektronu z wiązania kowalencyjnego.

Dielektryki różnią się od półprzewodników tylko większą szerokością pasma zabronionego ($W_g > 2 \text{ eV}$), co oczywiście oznacza, że w tych samych warunkach mniejsze jest prawdopodobieństwo uwolnienia elektronu z wiązania, mniejsza jest zatem koncentracja nośników swobodnych ładunku.

W energetycznej strukturze metali nie ma w ogóle pasma zabronionego. Może to wynikać z dwu powodów:

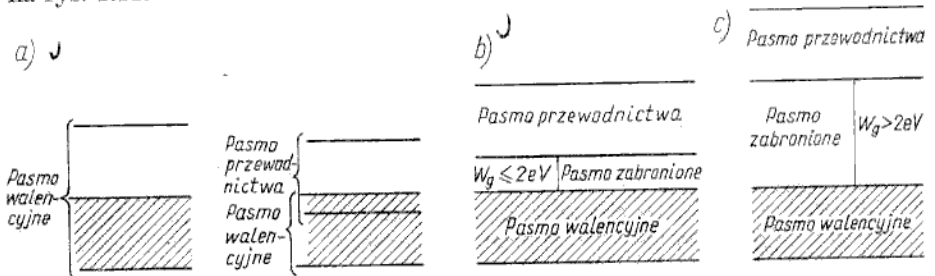
- pasmo walencyjne jest tylko częściowo wypełnione elektronami, nie wypełniona zatem część pasma walencyjnego, nie rozdzielona przerwą energetyczną od części wypełnionej, spełnia funkcję analogiczną do pasma przewodnictwa w półprzewodnikach, co na przykładzie litu przedstawiono na rys. 1.10a;



Rys. 1.10 Rozszczepienie zewnętrznych poziomów energetycznych atomów litu (a) i berylu (b) w ciele stałym (d_0 — odległość między atomami w ciele stałym)

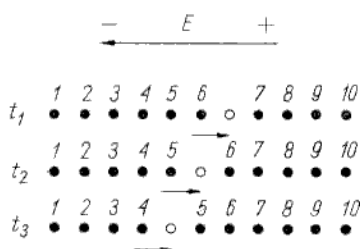
- pasmo walencyjne i pasmo przewodnictwa wzajemnie zachodzą na siebie, co na przykładzie berylu przedstawiono na rys. 1.10b.

Energetyczny model pasmowy dla metalu, półprzewodnika i dielektryka porównano na rys. 1.11.



Rys. 1.11 Porównanie energetycznego modelu pasmowego dla metalu (a), półprzewodnika (b) i dielektryka (c)

Nieco szerszego komentarza wymaga pojęcie dziury. Wiadomo już, że dziura jest zerwanym wiązaniem kowalencyjnym dwu atomów, powstałym wskutek oderwania elektronu z tego wiązania. W modelu pasmowym dziura jest pustym poziomem energetycznym w pasmie walencyjnym. Istnieje duże prawdopodobieństwo przejścia elektronu z jednego z sąsiednich wiązań międzyatomowych do luki utworzonej w zerwanym wiązaniu. Takiemu przejściu elektronu odpowiada przesunięcie dziury w kierunku przeciwnym. W stanie neutralnym, tj. bez zewnętrznego pola elektrycznego, ruch dziury jest chaotyczny ze średnią prędkością równą zeru w sensie wektorowym. W przypadku działania pola elektrycznego następują kolejne przesunięcia elektronów z wiązań kowalencyjnych do sąsiednich luk w wiązaniach zerwanych. Kierunek tego jakby „sztafetowego” ruchu elektronów (rys. 1.12) ma przeważającą składową zgodną z kierunkiem działa-



Rys. 1.12
Ilustracja transportu dziury w rezultacie „sztafetowego” ruchu elektronów (t_1 , t_2 , t_3 — kolejne chwile czasu)

nia pola elektrycznego. Z uwagi na bardzo duże trudności w opisie takiego „sztafetowego” ruchu wielu elektronów zajmujących kolejno puste miejsce korzystniej jest opisywać ruch tego jednego pustego miejsca, czyli dziury. Wówczas dziurze należy przypisać ładunek dodatni, równy co do wartości bezwzględnej ładunkowi elektronu, oraz określoną masę.

Niekiedy dla zilustrowania istoty pojęcia dziury przytacza się analogię do naczynia napełnionego wodą, do którego na dnie jest wpuszczony pęcherzyk powietrza. Pęcherzyk powietrza (dziura) porusza się w kierunku przeciwnym względem działającego pola ciężkości (pola elektrycznego) jako skutek opadania kolejnych cząstek cieczy (transportu kolejnych elektronów) do pęcherzyka (do dziury) zgodnie z kierunkiem działania siły ciężkości (pola elektrycznego). Absolutnie nie wolno mylić ruchu dziury z transportem jonu dodatniego. Ruch dziury jest kolejną jonizacją atomów nieruchomych związanych w węzłach sieci krystalicznej.

Półprzewodnik samoistny

1.3

Dotychczas rozpatrywaliśmy półprzewodnik idealnie czysty, nie mający żadnych domieszek ani defektów sieci krystalicznej. Taki półprzewodnik nazywa się *samoistnym*. Nośniki swobodne ładunku w półprzewodniku samoistnym powstają wskutek zrywania wiązań w sieci krystalicznej, co w energetycznym modelu pasmowym oznacza „przeskok” elektronów z pasma walencyjnego przez całą szerokość pasma zabronionego do pasma przewodnictwa. Z każdym aktem takiego „przeskoku” łączy się generacja pary elektron-dziura. Liczba elektronów w półprzewodniku samoistnym jest zatem równa liczbie dziur

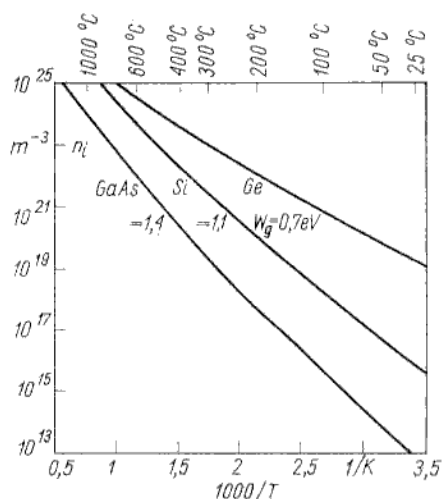
$$n_i = p_i \tag{1.1}$$

przy czym: n — koncentracja elektronów, tj. liczba elektronów w 1 m^3 ; p — koncentracja dziur, tj. liczba dziur w 1 m^3 ; indeks i oznacza półprzewodnik samoistny (ang. *intrinsic*).

Koncentracja par elektron-dziura jest tym większa, im łatwiej występuje zrywanie wiązań w sieci, tj. zwiększa się wraz ze wzrostem energii drgań atomów w węzłach sieci (zatem ze wzrostem temperatury) oraz ze zmniejszaniem się energii wymaganej do zerwania wiązania.

Energia aktywacji dla iloczynu koncentracji elektronów i dziur — jak dalej będzie wykazane — odpowiada szerokości pasma zabronionego W_g . Dalej wykażemy również, że istnieje następująca zależność funkcyjna dla koncentracji elektronów i dziur:

$$p_i = n_i \sim \exp \frac{-W_g}{2kT} \quad (1.2)$$



Rys. 1.13

Wykres zależności $\lg n_i = f\left(\frac{1}{T}\right)$ dla trzech podstawowych materiałów półprzewodnikowych: krzemu, germanu i arsenku galu

Zależność (1.2) przedstawiono w skali logarytmicznej $\left(\lg n_i = f\left(\frac{1}{T}\right)\right)$ na rys.

1.13 dla trzech podstawowych materiałów półprzewodnikowych: krzemu, germanu i arsenku galu.

Półprzewodnik niesamoistny

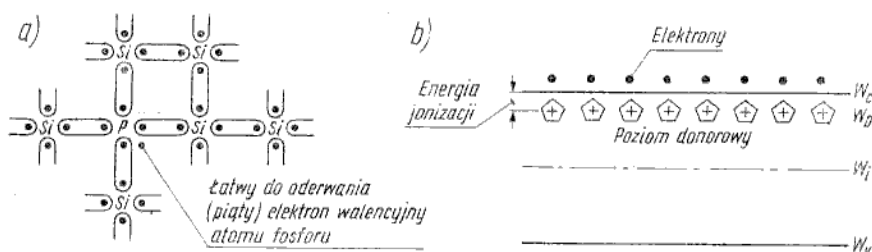
1.4

Rozpatrzony wyżej przypadek półprzewodnika samoistnego jest z punktu widzenia technicznego abstrakcyjny, gdyż materiał półprzewodnikowy nigdy nie ma idealnej struktury krystalicznej. Nawet przy stosowaniu najdoskonalszych procesów wytwarzania materiału półprzewodnikowego pozostają w nim pewne zanieczyszczenia innymi pierwiastkami oraz zakłócenia struktury krystalicznej (tzw. defekty i dyslokacje). Stosowane w praktyce materiały półprzewodnikowe mają zanieczyszczenia wprowadzane umyślnie, nazywane *domieszkami*. W takich materiałach półprzewodnikowych obok zjawiska samoistnego powstawania

swobodnych elektronów i dziur (np. zrywanie wiązań krzem-krzem) istotną rolę odgrywają dodatkowe mechanizmy powstawania nośników swobodnych ładunku, dlatego mówi się o półprzewodnikach *niesamoistnych*. Półprzewodnik umyślnie domieszkowany jest najważniejszym dla praktyki przypadkiem półprzewodnika niesamoistnego. Zostaną rozpatrzone dwa rodzaje domieszek: *donorowe* i *akceptorowe*.

Domieszki donorowe

Są to głównie pierwiastki piątej grupy tabeli okresowej (pierwiastki, których atomy mają pięć elektronów walencyjnych). W krzemie najczęściej jest stosowany fosfor, rzadziej arsen i antymon. Atom pierwiastka piątej grupy, przykładowo atom fosforu, zajmuje położenie w węźle sieci krystalicznej krzemu zamiast atomu krzemu, tzw. położenie zamiennowe (rys. 1.14a). Cztery elek-



Rys. 1.14

Atom domieszki donorowej (fosforu) w węźle sieci krystalicznej krzemu: model wiązań kowalencyjnych (a) oraz energetyczny model pasmowy (b) (półprzewodnik typu *n*). Pięciokątne znaki symbolizują jony pierwiastka piątej grupy

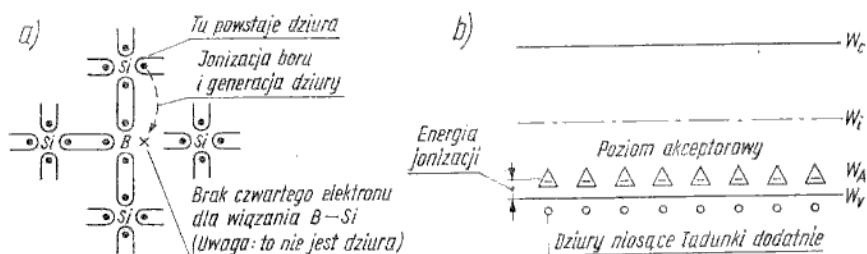
trony walencyjne atomu fosforu „biorą udział” w wiązaniach kowalencyjnych z czterema sąsiadującymi atomami krzemu, piąty elektron natomiast nie zaangażowany w wiązanie może być łatwo oderwany od atomu P. Oderwanie tego elektronu oznacza w modelu pasmowym jego przejście do pasma przewodnictwa (rys. 1.14b), stąd nazwy *domieszka donorowa* (dająca elektrony) oraz *półprzewodnik donorowy*. Natomiast w węźle sieci krystalicznej pozostaje zjonizowany jednododatni atom P. Energia jonizacji jest bardzo mała w porównaniu z szerokością pasma zakazanego¹⁾. Na przykład dla fosforu w krzemie energia jonizacji wynosi 0,044 eV. Dlatego w temperaturze pokojowej można przyjąć, że wszystkie atomy domieszki są zjonizowane, czyli przy $N_D \gg n_i$ koncentracja elektronów w pasmie przewodnictwa jest w przybliżeniu równa koncentracji atomów domieszki donorowej, $n \approx N_D$ (N_D — koncentracja atomów domieszki donorowej). Z uwagi na większą koncentrację elektronów niż dziur półprzewodnik donorowy jest często nazywany półprzewodnikiem typu *n* (negative).

Domieszki akceptorowe

Są to głównie pierwiastki trzeciej grupy tabeli okresowej (pierwiastki, których atomy mają trzy elektrony walencyjne). W krzemie najczęściej jest stosowany

¹⁾ Energię jonizacji można z grubsza oszacować przyjmując model atomu wodoru (jednododatni jon fosforu odpowiada jądro wodoru). Ponieważ energia jonizacji atomu wodoru wynosi 13,6 eV, zatem energię jonizacji atomu fosforu w sieci krystalicznej krzemu można zapisać jako $13,6 \text{ eV} / \epsilon_{rs}^2$, przy czym ϵ_{rs} — przenikalność elektryczna względna. Dla Si $\epsilon_{rs} = 12$, zatem energia jonizacji $\approx 0,1 \text{ eV}$.

bor, rzadziej aluminium, ind, gal. Atom pierwiastka trzeciej grupy, przykładowo atom boru, zajmuje położenie w węźle sieci krystalicznej krzemu zamiast atomu krzemu, tj. położenia zamiennowęzłowe (rys. 1.15a). Do stabilnego wiązania atomu boru z czterema sąsiadującymi atomami krzemu brak jest czwartego



Rys. 1.15

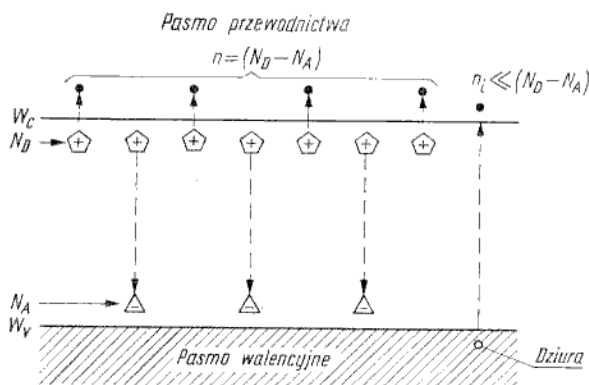
Atom domieszki akceptorowej (boru) w węźle sieci krystalicznej krzemu: model wiązań kowalencyjnych (a) oraz energetyczny model pasmowy (b) (półprzewodnik typu *p*). Trójkątne znaki symbolizują jony pierwiastka trzeciej grupy

elektronu, który może być łatwo uzupełniony po oderwaniu z sąsiadującego wiązania Si—Si. Wyrwanie elektronu z wiązania krzem-krzem i przyłączenie do czwartego brakującego wiązania bor-krzem oznacza w modelu pasmowym zabranie elektronu z pasma walencyjnego (rys. 1.15b), stąd nazwy *domieszka akceptorowa* (przyjmująca elektrony) oraz *półprzewodnik akceptorowy*. Wówczas powstaje dziura w pasmie walencyjnym, atom boru natomiast, ulokowany w węźle sieci krystalicznej, jonizuje się jednoujemnie. Energia jonizacji jest bardzo mała w porównaniu z szerokością pasma zabronionego. Na przykład dla boru w krzemie energia jonizacji wynosi 0,045 eV. Dlatego w temperaturze pokojowej można przyjąć, że wszystkie atomy domieszki są zjonizowane, czyli przy $N_A \gg n_i$ koncentracja dziur w pasmie walencyjnym jest w przybliżeniu równa koncentracji atomów domieszki akceptorowej, $p \approx N_A$ (N_A — koncentracja atomów domieszki akceptorowej). Z uwagi na większą koncentrację dziur niż elektronów półprzewodnik akceptorowy jest często nazywany półprzewodnikiem typu *p* (positive).

Domieszki donorowe i akceptorowe

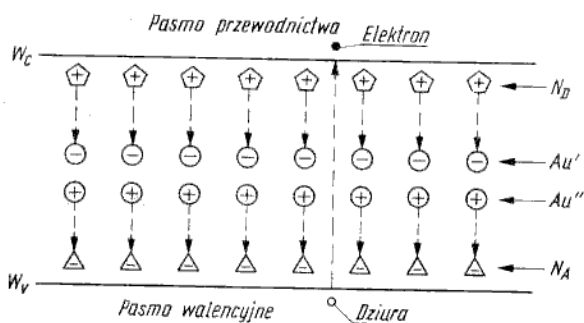
Najbardziej realny jest przypadek występowania w półprzewodniku jednocześnie obu rodzajów domieszek, tj. donorowych i akceptorowych. W tym przypadku rodzaj przewodnictwa w półprzewodniku jest określony tymi domieszkami, których koncentracja jest większa. Jeżeli $N_D > N_A$, to półprzewodnik jest typu *n*, dla $N_A > N_D$ — typu *p*. Można by błędnie przypuszczać, że w półprzewodniku domieszkowanym donorami i akceptorami koncentracje zarówno elektronów jak i dziur są większe niż w półprzewodniku samoistnym. Dalej zostanie wykazane ściśle (wzór (1.16)), że w określonym półprzewodniku (Si, Ge itd.), w określonej temperaturze, iloczyn koncentracji elektronów i dziur jest wielkością stałą, niezależną od rodzaju domieszkowania, równą n_i^2 . Przykładowo dla $N_D \approx N_A$ należy stwierdzić, że $n \neq N_D$ oraz $p \neq N_A$. Natomiast $n \approx (N_D - N_A)$, jeżeli $(N_D - N_A) \gg n_i$ lub też $p \approx (N_A - N_D)$, jeżeli $(N_A - N_D) \gg n_i$. Będzie to wyjaśnione jakościowo na podstawie modelu przedstawionego na rys. 1.16. Elektrony dążą do zajęcia najniższych wakuujących poziomów energetycznych. Dlatego część elektronów z poziomu donorowego zapełni centra akceptorowe, które

energetycznie są ulokowane niżej niż centra donorowe. Pozostanie zatem $N_D - N_A$ atomów donorowych zdolnych do oddania elektronów do pasma przewodnictwa. Stąd $n \approx (N_D - N_A)$ dla $N_D > N_A$ (półprzewodnik typu n) oraz $p \approx (N_A - N_D)$



Rys. 1.16
Model energetyczny półprzewodnika domieszkowanego donorami i akceptorami

dla $N_A > N_D$ (półprzewodnik typu p). Półprzewodnik o koncentracjach $N_D = N_A$ ma taką samą koncentrację nośników jak półprzewodnik samoistny, tj. $n_i = p_i$. Jednakże z uwagi na wpływ zjonizowanych centrów domieszkowych na inne parametry (np. ruchliwość) taki półprzewodnik nie jest nazywany samoistnym, lecz *skompensowanym*. Efekt kompensacji można uzyskać również w inny sposób, przez wprowadzenie domieszek spełniających funkcję tzw. *pułapek*. Taką domieszką jest złoto, którego atomy lokują się w pozycjach międzywęzłowych sieci krystalicznej krzemu i energetycznie zajmują dwa poziomy głębokie (odległe od krawędzi pasma zabronionego) — rys. 1.17. Poziom wyższy, oznaczony



Rys. 1.17
Model energetyczny półprzewodnika skompensowanego z domieszką złota

na rysunku jako Au' , spełnia funkcję pułapek akceptorowych, gdyż przechwytywa elektrony z poziomu donorowego. Poziom niższy, oznaczony na rysunku jako Au'' , spełnia funkcję pułapek donorowych, gdyż oddaje swe elektrony do poziomu akceptorowego. Przy odpowiednio dużej koncentracji atomów złota dochodzi do całkowitego „wyczerpania” domieszek donorowych i akceptorowych, zatem $n_i = p_i$. Taki półprzewodnik również nazywa się *skompensowanym*.

Koncentracja elektronów i dziur w stanie równowagi termodynamicznej

1.5

Rozpatrzmy zależności określające koncentrację elektronów w pasmie przewodnictwa oraz dziur w pasmie walencyjnym dla półprzewodnika w stanie równowagi termodynamicznej. W stanie równowagi półprzewodnik ma ustaloną temperaturę i nie jest poddany żadnym zewnętrznym oddziaływaniom energetycznym w postaci promieniowania jonizującego, pola elektrycznego itp. Koncentracje elektronów i dziur zależą od koncentracji dozwolonych poziomów energetycznych i prawdopodobieństwa ich obsadzenia przez elektron lub dziurę:

$$n = \int_{W_c}^{\infty} N(W) f_n(W) dW \quad (1.3)$$

$$p = \int_{W_v}^0 N(W) f_p(W) dW \quad (1.4)$$

przy czym: $N(W)$ — rozkład koncentracji poziomów energetycznych w funkcji energii; $f_n(W)$ — funkcja rozkładu prawdopodobieństwa zajęcia poziomu o energii W przez elektron; $f_p(W)$ — funkcja rozkładu prawdopodobieństwa zajęcia poziomu o energii W przez dziurę; granice całek są wyznaczone na osi energii skierowanej przeciwnie, niż to było na rys. 1.8 i 1.9a.

Podstawowe parametry elektrofizyczne Si, Ge, GaAs dla $T = 300 \text{ K}$

Tablica 1.1

Parametr	Materiał	Si	Ge	GaAs
Koncentracja atomów lub molekuł [m^{-3}]		$5 \cdot 10^{28}$	$4,42 \cdot 10^{28}$	$2,21 \cdot 10^{28}$
Stała sieci krystalicznej [10^{-1} nm]		5,43	5,66	5,65
Gęstość (masa właściwa) [kg/m^3]		2330	5320	5320
Temperatura topnienia [$^{\circ}\text{C}$]		1420	937	1238
Szerokość pasma zabronionego [eV]		1,1	0,7	1,4
Efektywna gęstość stanów energetycznych $N_c [\text{m}^{-3}]$ $N_v [\text{m}^{-3}]$		$2,8 \cdot 10^{25}$ $1,04 \cdot 10^{25}$	$1,04 \cdot 10^{25}$ $6,0 \cdot 10^{24}$	$4,7 \cdot 10^{23}$ $7,0 \cdot 10^{24}$
Koncentracja nośników samoistnych $n_i [\text{m}^{-3}]$		$1,45 \cdot 10^{16}$	$2,4 \cdot 10^{19}$	10^{13}
Ruchliwość dla półprzewodnika samoistnego [$\text{m}^2/(\text{V} \cdot \text{s})$]				
elektronów dziur	μ_n μ_p	0,13 0,05	0,39 0,19	0,86 0,025
Względna przenikalność elektryczna ϵ_{rs}		11,7	16	12
Wytrzymałość na przebicie [$\text{V}/\mu\text{m}$]		30	8	35

Funkcje $N(W)$ dla pasm przewodnictwa i walencyjnego są wyznaczane w fizyce kwantowej, przy czym zagadnienie można znacznie uprościć przez wprowadzenie pojęcia efektywnych koncentracji stanów energetycznych sprowadzonych do poziomów dna pasma przewodnictwa (N_c) oraz wierzchołka pasma walencyjnego (N_v). Wówczas:

$$n = N_c f_n(W_c) \quad (1.3a)$$

$$p = N_v f_p(W_v) \quad (1.4a)$$

Wielkości N_c , N_v są zależne od temperatury ($\sim T^{3/2}$). Wartości N_c , N_v w temperaturze pokojowej (300 K) dla krzemu, germanu i arsenku galu podano wraz z innymi parametrami w tabl. 1.1. Należy teraz odpowiedzieć na pytanie: jakie jest prawdopodobieństwo zajęcia przez elektron poziomu W_c oraz przez dziurę poziomu W_v ? W fizyce klasycznej dla cząstek gazu doskonałego prawdopodobieństwo posiadania przez cząstkę energii W jest określone rozkładem Boltzmann

$$f(W) = \exp \frac{-W}{kT} \quad (1.5)$$

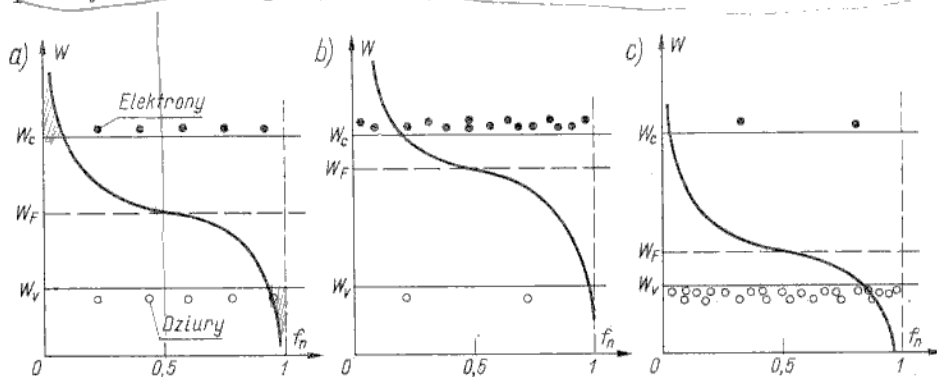
przy czym: k — stała Boltzmann; T — temperatura.

W przypadku elektronów w ciele stałym dodatkowe ograniczenie do rozkładu energetycznego wprowadza zasada Pauliego. Według wzoru (1.5) dla $T = 0$ K wszystkie cząstki przyjmują wartość energii $W = 0$ (w przypadku elektronów w ciele stałym byłaby to pewna wartość energii odniesienia). Taki wniosek jest oczywiście sprzeczny z zasadą Pauliego, według której nie może być dwu elektronów o jednakowych stanach energetycznych. Prawidłowe rozwiązanie zagadnienia rozkładu energetycznego elektronów w ciele stałym daje statystyka Fermiego-Diraca. Rozkład energetyczny, zwany rozkładem Fermiego-Diraca, ma postać:

$$f_n(W) = \frac{1}{1 + \exp((W - W_F)/kT)} \quad (1.6a)$$

$$f_p(W) = 1 - f_n(W) \quad (1.6b)$$

Parametr W_F , nazywany energią Fermiego lub poziomem Fermiego, ma sens potencjału chemicznego, tj. średniej energii swobodnej elektronów liczonej na



Rys. 1.18

Rozkład Fermiego-Diraca dla półprzewodnika samoistnego (a), typu n (b) oraz typu p (c)

(jeden elektron¹⁾). Można go również zdefiniować następująco: poziom Fermiego jest to poziom energetyczny, którego prawdopodobieństwo zajęcia przez elektron jest równe 0,5. Rozkłady Fermiego-Diraca dla półprzewodnika samoistnego, typu n oraz typu p przedstawiono na rys. 1.18. Funkcje rozkładu elektronów i dziur są wzajemnie symetryczne względem poziomu W_F . Dlatego w półprzewodniku samoistnym przy jednakowych koncentracjach elektronów i dziur poziom Fermiego jest ulokowany prawie dokładnie w środku pasma zabronionego. W półprzewodniku typu n z uwagi na większą koncentrację elektronów niż dziur poziom Fermiego jest ulokowany bliżej dna pasma przewodnictwa. W półprzewodniku typu p z uwagi na większą koncentrację dziur niż elektronów poziom Fermiego jest ulokowany bliżej wierzchołka pasma walencyjnego.

Może wydawać się dziwne, że rozkład prawdopodobieństwa przyjmuje wartości różne od zera w pasmie zabronionym, gdzie — jak wiadomo — nie ma elektronów. Jednakże nie ma tu żadnej sprzeczności. Rozkład Fermiego-Diraca jest ogólnym rozwiązaniem statystycznym, na które nakładają się dodatkowe ograniczenia wynikające z teorii pasmowej. Inaczej, rozkład prawdopodobieństwa w obszarze pasma zabronionego określa, jakie byłoby prawdopodobieństwo zajęcia poziomów energetycznych, gdyby były one dozwolone w tym pasmie (dyskretne poziomy dozwolone w pasmie zabronionym istnieją w półprzewodniku domieszkowanym).

Dla poziomów energetycznych odległych o kilka kT od poziomu Fermiego można pominąć jedynekę w mianowniku wyrażenia (1.6) i otrzymać rozkład Boltzmanna:

$$\text{dla elektronów } f_n(W) = \exp[-(W - W_F)/kT] \text{ przy } W > W_F \quad (1.7)$$

$$\text{dla dziur } f_p(W) = \exp[-(W_F - W)/kT] \text{ przy } W < W_F \quad (1.8)$$

Poziom Fermiego spełnia funkcję energii odniesienia. Po podstawieniu zależności (1.7), (1.8) do wzorów (1.3a), (1.4a) otrzymuje się wyrażenia:

$$n = N_c \exp[(W_F - W_c)/kT] \quad (1.9)$$

$$p = N_v \exp[(W_v - W_F)/kT] \quad (1.10)$$

Na ogół przyjmuje się, że rozkład Boltzmanna można stosować w przypadku $|W_F - W_c| \geq 2kT$ (półprzewodnik typu n) oraz $|W_v - W_F| \geq 2kT$ (półprzewodnik typu p). Półprzewodnik nie spełniający tego warunku jest nazywany zdegenerowanym i wymaga stosowania statystyki Fermiego-Diraca. Degeneracja występuje w przypadku dużej koncentracji domieszek (ok. 10^{25} m^{-3} w przypadku krzemu). Poniżej wyprowadzimy kilka ważnych zależności związanych z rozpatrywanym tematem.

Poziom Fermiego w półprzewodniku samoistnym

1.5.1

Z przyrównania (1.9) i (1.10) przy $W_F = W_i$ (W_i — poziom Fermiego w półprzewodniku samoistnym) otrzymuje się

$$\left(W_i = \frac{1}{2}(W_c + W_v) + \frac{1}{2} kT \ln \frac{N_v}{N_c} \right) \quad (1.11)$$

Ponieważ drugi człon wyrażenia (1.11) jest bardzo mały, można stwierdzić, że poziom W_i znajduje się niemal w środku pasma zabronionego.

¹⁾ Chodzi o średnią energię tylko tych nośników ładunku, które mogą poruszać się swobodnie w kryształach, tj. elektronów w pasmie przewodnictwa i dziur w pasmie walencyjnym.

Koncentracja nośników w półprzewodniku samoistnym

1.5.2

Na podstawie zależności (1.9), (1.10) przy $n = n_i$, $p = n_i$ oraz $W_F = W_i$ otrzymuje się

$$n_i^2 = N_c N_v \exp\left(-\frac{W_g}{kT}\right) \quad (1.12)$$

Ponieważ $N_c, N_v \sim T^{3/2}$, zatem

$$n_i = B \left(\frac{T}{T_0}\right)^{3/2} \exp\left(-\frac{W_g}{2kT}\right) \quad (1.13)$$

przy czym: B — stała materiałowa równa $1,7 \cdot 10^{25} \text{ m}^{-3}$ dla krzemu; $T_0 = 300 \text{ K}$.

Zależności Boltzmannna

1.5.3

Na podstawie (1.9), (1.10) można łatwo wyprowadzić następujące zależności:

$$n = n_i \exp\left(-\frac{q\varphi_F}{kT}\right) \quad (1.14)$$

$$p = n_i \exp\left(\frac{q\varphi_F}{kT}\right) \quad (1.15)$$

przy czym $\varphi_F = (W_i - W_F)/q$ — tzw. *potencjał Fermiego* o znaku dodatnim dla półprzewodnika typu p oraz ujemnym dla półprzewodnika typu n . Zależności (1.14), (1.15) są bardzo dogodne, gdyż umożliwiają łatwe wyznaczenie koncentracji nośników na podstawie znajomości położenia poziomu Fermiego lub odwrotnie — poziomu Fermiego przy znanych koncentracjach nośników.

Bilans koncentracji nośników

1.5.4

Po przemnożeniu stronami równań (1.14), (1.15) otrzymuje się podstawową dla dalszych rozważań zależność

$$np = n_i^2 \quad (1.16)$$

Ze wzoru (1.16) wynika bardzo ważny wniosek, że iloczyn koncentracji elektronów i dziur dla danego półprzewodnika w stanie równowagi termodynamicznej i w określonej temperaturze jest wielkością stałą, niezależną od sposobu domieszkowania półprzewodnika.

Równanie neutralności elektrycznej

1.5.5

Sumaryczny ładunek bryły półprzewodnika, w stanie równowagi termodynamicznej, jest równy zeru. Jest to oczywiste, gdyż generacja elektronu lub (i) dziury nie oznacza powstania dodatkowego ładunku, lecz polega na rozdzieleniu dwóch

już istniejących ładunków, różnoimiennych. (Jest to zgodne z podstawowym twierdzeniem elektrostatyki o niezniszczalności i niekreatywności ładunków elektrycznych w przyrodzie). W półprzewodniku samoistnym powstają pary elektron-dziura, w półprzewodniku domieszkowanym natomiast generacja elektronu oznacza dodatnią jonizację jednego atomu domieszki donorowej, a generacja dziury oznacza ujemną jonizację jednego atomu domieszki akceptorowej. Wskutek silnego przyciągania ładunków różnoimiennych półprzewodnik „dąży” do utrzymania neutralności elektrycznej w każdym obszarze lokalnym o wymiarach makroskopowych, tj. o wymiarach umożliwiających traktowanie tego obszaru jako obiektu termodynamicznego. Jednak obszary lokalne półprzewodnika nie zawsze są neutralne elektrycznie. W dalszych rozdziałach będzie się można przekonać, że podział bryły półprzewodnika (będącej zwykle strukturą wielowarstwową) na obszary neutralne i nienaturalne stanowi podstawę wielu interpretacji i analiz działania struktur półprzewodnikowych.

W półprzewodniku domieszkowanym równomiernie (jednorodnym) obowiązuje zasada neutralności elektrycznej w dowolnym obszarze lokalnym o wymiarach makroskopowych. Zatem koncentracja ładunku elektrycznego, utworzonego przez elektrony, dziury, dodatnie jony domieszki donorowej i ujemne jony domieszki akceptorowej, jest równa zeru. Jeśli przyjmie się, że wszystkie centra akceptorowe i donorowe są zjonizowane, to koncentrację ładunku elektrycznego można zapisać w postaci

$$\rho = q(p - n + N_D - N_A) \quad (1.17)$$

Ponieważ $\rho = 0$, zatem

$$p - n = N_A - N_D \quad p + N_D = N \quad (1.18)$$

Korzystając ze wzoru (1.16) można rozwiązać równanie (1.18) ze względu na koncentrację elektronów w półprzewodniku typu n lub dziur w półprzewodniku typu p , dla stanu równowagi:

$$n_n = \frac{1}{2} \left[N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_i^2} \right] \quad (1.19)$$

$$p_p = \frac{1}{2} \left[N_A - N_D + \sqrt{(N_A - N_D)^2 + 4n_i^2} \right] \quad (1.20)$$

Dolny indeks n lub p oznacza typ półprzewodnika. Przykładowo n_n jest koncentracją elektronów w półprzewodniku typu n , p_n jest koncentracją dziur w półprzewodniku typu n . Nośniki o większej koncentracji w danym materiale są nazywane *większościowymi*, a nośniki przeciwnego rodzaju — *mniejszościowymi*. Zatem n_n , p_p są to koncentracje nośników większościowych, zaś n_p , p_n — mniejszościowych. Na podstawie równań (1.19), (1.20) przy $|N_D - N_A| \gg n_i$ otrzymuje się wcześniej postulowane zależności:

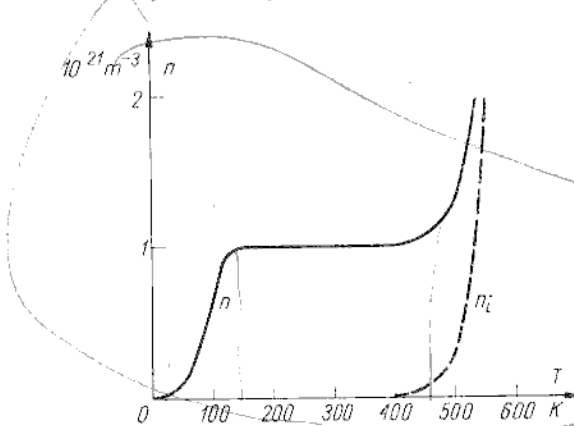
$$\begin{cases} n_n \approx N_D - N_A \\ p_p \approx N_A - N_D \end{cases} \quad (1.21)$$

Zależność koncentracji nośników od temperatury

1.5.6

Wzory (1.19), (1.20) wraz z równaniem (1.13) umożliwiają wyznaczenie zależności koncentracji nośników od temperatury dla przypadku całkowicie zjonizowanych domieszek, tj. dla temperatury wyższej niż 100...150 K. Typową za-

leżność koncentracji elektronów od temperatury dla krzemu typu n przedstawiono na rys. 1.19. W zakresie 0...150 K koncentracja wzrasta w funkcji temperatury, gdyż zwiększa się energia cieplna sieci krystalicznej, wzrasta zatem liczba zjonizowanych atomów domieszki. W szerokim przedziale temperatury



Rys. 1.19
Zależność koncentracji nośników w funkcji temperatury dla krzemu typu n o koncentracji domieszek $N_D = 10^{21} \text{ m}^{-3}$

150...450 K koncentracja jest, praktycznie biorąc, stała, gdyż wszystkie atomy domieszki są już zjonizowane. Powyżej 450 K istotne znaczenie ma wzrost koncentracji samoistnej n_i , wskutek generacji par elektron-dziura (przeskok elektronów przez całą szerokość pasma zabronionego).

Transport nośników w półprzewodniku

1.6

Wiadomo, że gęstość prądu elektrycznego można zapisać ogólnie jako sumę gęstości prądu przewodzenia (ruch nośników ładunku) i prądu przesunięcia

$$J_t = J_c + \epsilon_r \epsilon_0 \frac{\partial E}{\partial t} \quad (1.22)$$

Prąd przesunięcia w metalach i półprzewodnikach nie odgrywa istotnej roli w bardzo szerokim zakresie częstotliwości sygnałów elektrycznych. Można to łatwo wykazać rozpatrując szybkość relaksacji następującej po zakłóceniu lokalnej obciążalności ładunku elektrycznego. Weźmy wycinek materiału o kształcie prostopadłościennym, utworzony przez dwie ściany o powierzchni A znajdujące się w odległości d . Układ zastępczy tego wycinka tworzy rezystor $R = \frac{d}{A} \rho$ połączony równolegle z kondensatorem $C = \epsilon_r \epsilon_0 A/d$. Stąd stała czasowa relaksacji dielektrycznej $\tau_d = RC$ ma postać

$$\tau_d = \epsilon_r \epsilon_0 \rho \quad (1.23)$$

przy czym: ϵ_r — przenikalność elektryczna względna materiału; ϵ_0 — przenikalność elektryczna próżni ($8,85 \cdot 10^{-12} \text{ F/m}$); ρ — rezystywność materiału.

Dla krzemu ($\epsilon_r \approx 12$) w najbardziej typowym przypadku $\rho = 0,01 \Omega \cdot \text{m}$, zatem $\tau_d \approx 10^{-12} \text{ s}$. Stąd wynika, że dla $\omega \ll 10^{12} \text{ Hz}$ składową pojemnościową, czyli

składową prądu przesunięcia można pominąć. Pozostaje więc do rozpatrzenia składową prądu przewodzenia.

W prądzie przewodzenia, utworzonym przez poruszające się nośniki, należy rozróżnić dwa mechanizmy, tj. unoszenie i dyfuzję

$$J_c = J_u + J_d \quad (1.24)$$

Unoszenie

1.6.1

W przypadku równomiernego rozkładu koncentracji nośników w półprzewodniku prąd przewodzenia jest w całości prądem unoszenia, czyli jest utworzony przez strumień nośników poruszających się wskutek działania siły elektrycznej. Gęstość tego prądu można łatwo obliczyć jako iloczyn koncentracji ładunku nośników swobodnych (na przykład ładunku elektronów qn w półprzewodniku typu n) przez średnią prędkość jego unoszenia w polu elektrycznym

$$J_u = qn\bar{v}_u \quad (1.25)$$

Nieco szerzej będzie wyjaśnione pojęcie średniej prędkości unoszenia na przykładzie elektronów w półprzewodniku typu n . Elektron znajduje się w ciągłym ruchu cieplnym, przy czym wskutek wielu zderzeń z węzłami sieci krystalicznej jest to ruch chaotyczny o skomplikowanej trajektorii. Jeśli półprzewodnik jest jednorodnie domieszkowany (nie ma wewnętrznego pola elektrycznego — por. p. 1.9) i nie działa nań zewnętrzne pole elektryczne, to mimo tego ruchu w odpowiednio długim okresie czasu średnie położenie elektronu nie ulega zmianie. Można inaczej powiedzieć, że wartość średnia składowej prędkości ruchu cieplnego wzdłuż dowolnie obranego kierunku jest równa zeru. Jeśli doprowadzi się zewnętrzne pole elektryczne do półprzewodnika, to w ruchu elektronu wystąpi dodatkowa składowa prędkości skierowana przeciwnie względem linii sił pola elektrycznego. Tę dodatkową składową prędkości, nazywaną *prędkością unoszenia*, można łatwo oszacować w wyniku następującego rozumowania. Załóżmy, że w chwili zderzenia elektronu z węzłem sieci krystalicznej prędkość unoszenia zmniejsza się do zera (następuje zmiana kierunku ruchu cieplnego), po czym wzrasta wskutek ruchu jednostajnie przyspieszonego na drodze do kolejnego zderzenia. Wówczas dla średniego czasu między dwoma zderzeniami t_{zd} średnia prędkość unoszenia

$$\bar{v}_u = \frac{at_{zd}}{2}$$

Przyspieszenie $a = F/m$, czyli $a = qE/m_n^*$, przy czym m_n^* — masa efektywna elektronu¹⁾. Zatem

$$\bar{v}_u = \frac{qt_{zd}}{2m_n^*} E \quad (1.26)$$

¹⁾ Pojęcie masy efektywnej pojawia się dlatego, że inne są właściwości transportu elektronu w przestrzeni swobodnej niż w polu potencjalnym sieci krystalicznej półprzewodnika. Przy tej samej sile inne jest przyspieszenie elektronu poruszającego się w przestrzeni swobodnej niż w kryształce półprzewodnika, gdzie doznaje on dodatkowego działania pola potencjalnego, utworzonego przez węzły sieci krystalicznej. Dla celów dydaktycznych może być przydatna analogia mechaniczna ruchu elektronu w polu periodycznym kryształu do ruchu kulki na pofalowanej powierzchni.

lub

$$\bar{v}_u = \mu E \quad (1.26a)$$

przy czym

$$\mu = \frac{\bar{v}_u}{E} = \frac{qt_{zd}}{2m_n^*} \quad (1.27)$$

jest *ruchliwością* elektronów.

Jeżeli uwzględni się rozrzut statystyczny czasów między dwoma kolejnymi zderzeniami, to

$$\mu = \frac{\bar{q}t_{zd}}{m_n^*} \quad (1.27a)$$

przy czym \bar{t}_{zd} — średni czas między zderzeniami.

Ruchliwość

1.6.1.1

Do dalszych rozważań dotyczących przyrządów półprzewodnikowych bardzo przydatna jest znajomość zależności ruchliwości od następujących trzech czynników:

- koncentracji domieszek,
- temperatury,
- natężenia pola elektrycznego.

[Ruchliwość jest wprost proporcjonalna do czasu między dwoma zderzeniami] (1.27).

Można rozpatrywać dwa rodzaje zderzeń, nie rozumianych oczywiście w dosłownym znaczeniu mechanicznym, lecz jako zmiana kierunku ruchu elektronu wskutek oddziaływania elektrostatycznego ze zjonizowanym centrum domieszkowym (rozpraszanie na domieszkach) lub zmiana kierunku ruchu elektronu wskutek zakłócenia periodyczności pola elektrycznego przez drgający (ruch cieplny) atom w węzle sieci krystalicznej (rozpraszanie cieplne lub rozpraszanie na fononach¹⁾). Ponieważ oba rodzaje zderzeń są zjawiskami niezależnymi, liczba zderzeń w czasie 1 s (częstość zderzeń) jest równa sumie częstości zderzeń z domieszkami i fononami, zatem

$$\frac{1}{t_{zd}} = \frac{1}{t_{zd, \text{domieszki}}} + \frac{1}{t_{zd, \text{fonony}}} \quad (1.28)$$

oraz uwzględniając wzór (1.27)

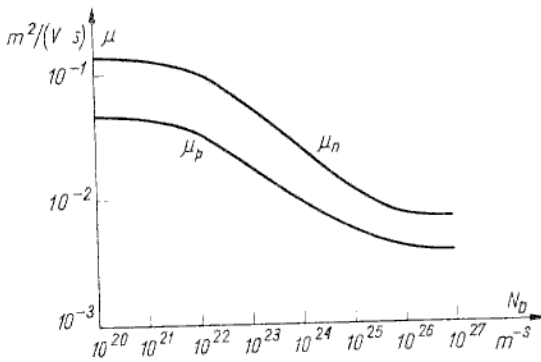
$$\frac{1}{\mu} = \frac{1}{\mu_d} + \frac{1}{\mu_f} \quad (1.29)$$

(Rozpraszanie na jonach domieszek jest tym silniejsze, im większa jest koncentracja domieszek, dlatego ruchliwość maleje w miarę wzrostu poziomu domieszkowania półprzewodnika) (rys. 1.20).

W podstawowym kursie fizyki półprzewodników wyprowadza się wyrażenia na zależność ruchliwości od temperatury dla rozpraszania na jonach domieszek ($\mu_d \sim T^{3/2}$) lub na drgających węzłach sieci krystalicznej ($\mu_f \sim T^{-3/2}$). Ponieważ w niskiej temperaturze (< 150 K) rozpraszanie na fononach jest mało efektywne, w tym

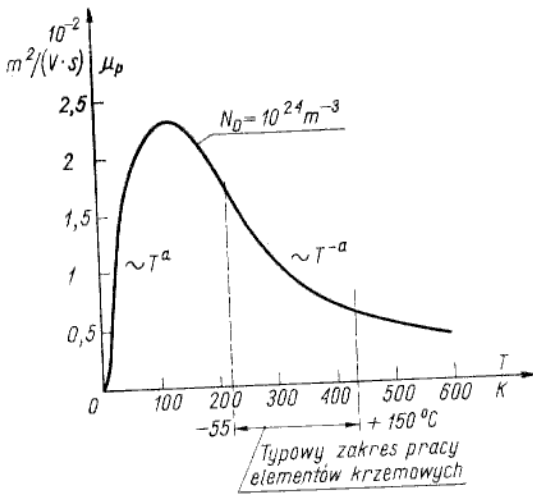
¹⁾ Fonon jest kwantem, czyli najmniejszą niepodzielną porcją energii drgań atomów w kryształce.

Transport nośników w półprzewodniku



Rys. 1.20
Zależność ruchliwości nośników od koncentracji domieszek w krzemie

zakresie temperatury przeważa rozpraszanie na jonach domieszek, czyli $\mu \sim T^{3/2}$. W wyższych temperaturach przeważa rozpraszanie na fononach i wówczas $\mu \sim T^{-3/2}$ (doświadczalnie często obserwuje się zależność $\mu \sim T^{-2.5}$) — rys. 1.21.



Rys. 1.21
Zależność ruchliwości dziur w krzemie od temperatury

W miarę wzrostu wartości natężenia pola elektrycznego powyżej kilkuset kV/m, gdy składowa prędkości unoszenia staje się porównywalna z prędkością ruchu termicznego (ok. 10⁵ m/s), następuje ograniczenie dalszego wzrostu prędkości unoszenia, czemu odpowiada zmniejszenie się ruchliwości (rys. 1.22).

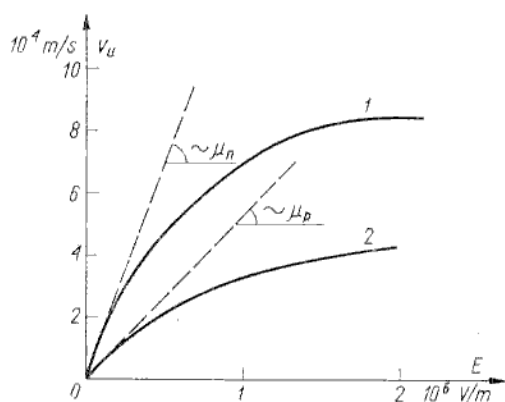
Konduktywność

1.6.1.2

Biorąc pod uwagę wyrażenia (1.25), (1.26a) można napisać

$$J_u = qn\mu E$$

(1.30)



Rys. 1.22
Zależność prędkości unoszenia, a stąd również ruchliwości elektronów i dziur w krzemie od natężenia pola elektrycznego

Po uwzględnieniu prawa Ohma $J_u = \sigma E$, konduktywność można wyrazić w postaci:

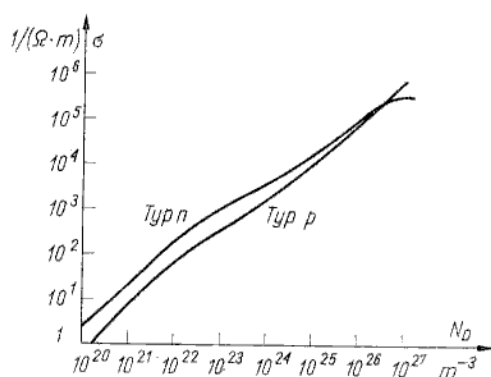
$$\sigma_n = qn\mu_n \text{ — dla półprzewodnika typu } n \quad (1.31)$$

$$\sigma_p = qp\mu_p \text{ — dla półprzewodnika typu } p \quad (1.31a)$$

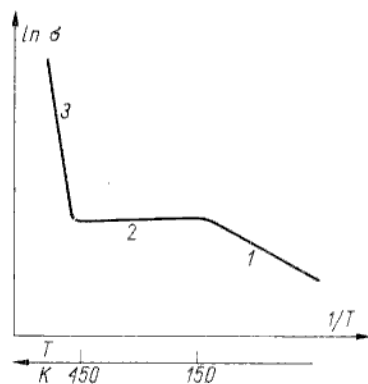
W przypadku udziału obu rodzajów nośników w prądzie unoszenia

$$\sigma = q(n\mu_n + p\mu_p) \quad (1.32)$$

Doświadczalne zależności konduktywności od koncentracji domieszek przedstawiono dla krzemu na rys. 1.23 (są to tzw. krzywe Irvina¹⁾).



Rys. 1.23
Zależność konduktywności krzemu typu n oraz typu p od koncentracji domieszek (krzywa Irvina)



Rys. 1.24
Zależność konduktywności od temperatury
1 — zakres stałej koncentracji nośników (wszystkie domieszki zjonizowane); 2 — zakres generacji samoistnej; 3 — zakres jonizacji domieszek

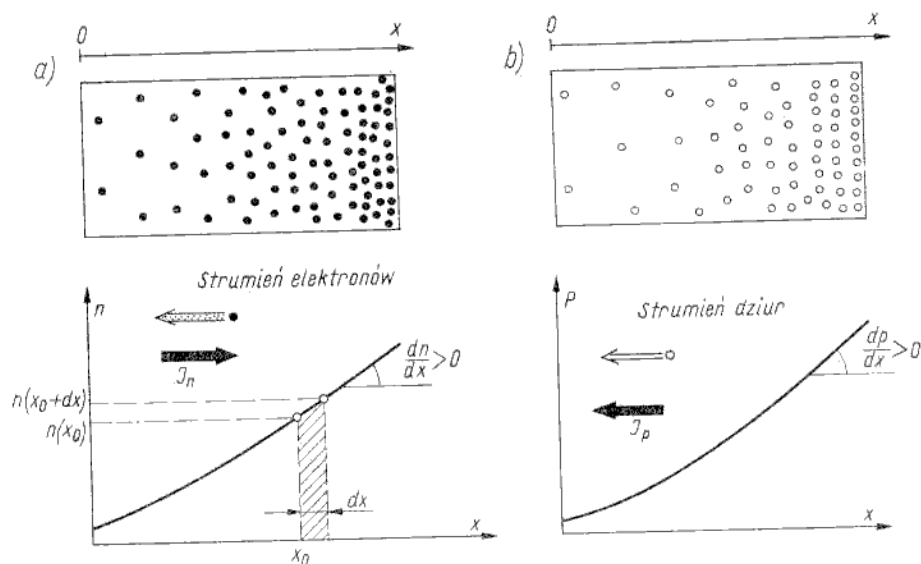
¹⁾ Irvin J. C.: Resistivity of bulk silicon and of diffused layers in silicon. *Bell System Techn. J.*, 1962, Vol. 41, p. 387.

Znając zależność koncentracji nośników i ich ruchliwości od temperatury można wyznaczyć zależność temperaturową konduktywności, którą przedstawiono we współrzędnych $\ln \sigma = f(1/T)$ na rys. 1.24. Interpretacja tej krzywej jest w zasadzie taka jak dla zależności temperaturowej koncentracji nośników, przedstawionej na rys. 1.19, gdyż wpływ zmian temperaturowych ruchliwości jest znacznie słabszy.

Dyfuzja

1.6.2

Składowa prądu dyfuzji występuje wówczas, gdy rozkład koncentracji nośników prądu jest nierównomierny. Wskutek chaotycznego ruchu cieplnego nośników następuje proces wyrównywania się ich koncentracji, tj. przepływ nośników z obszarów o większej koncentracji do obszarów o mniejszej koncentracji. Poglądowo wyjaśnia to rys. 1.25, który ilustruje rozkłady jednowymiarowe kon-



Rys. 1.25

Poglądowa ilustracja prądu dyfuzji: a) elektronów; b) dziur

centracji elektronów (rys. 1.25a) i dziur (rys. 1.25b). Obie funkcje rozkładu koncentracji, tj. $n(x)$ oraz $p(x)$, mają dodatnie nachylenie (koncentracje wzrastają w kierunku osi x). Rozpatrzmy, zgodnie z rys. 1.25a, dwa przekroje oddalone od siebie na odległość dx . Koncentracja dla przekroju x_0 wynosi $n(x_0)$, a dla przekroju $(x_0 + dx)$ jest równa $n(x_0 + dx)$. Wskutek ruchu chaotycznego elektronów znajdujące się w przekroju x_0 po pewnym czasie opuszczają to miejsce. Ponieważ prawdopodobieństwa ruchu w lewo i w prawo (problem jednowymiarowy) są jednakowe, połowa elektronów z przekroju x_0 , tj. $\frac{1}{2} n(x_0)$, znajdzie się z lewej strony współrzędnej x_0 , czyli opuści wycinek dx (obszar zakreskowany na rys. 1.25a). Analogiczny proces zachodzi dla przekroju $(x_0 + dx)$, wskutek

czego połowa elektronów z tego przekroju, tj. $\frac{1}{2}n(x_0+dx)$, znajdzie się z lewej strony współrzędnej x_0+dx , czyli wejdzie do obszaru dx . Ponieważ

$$n(x_0+dx) > n(x_0)$$

zatem wypadkowa liczba elektronów w obszarze dx zwiększy się, czyli skutek ruchów cieplnych przeważa strumień nośników z obszarów o większej koncentracji do obszarów o mniejszej koncentracji. W ten sposób powstaje prąd dyfuzji.

Gęstość prądu dyfuzji jest proporcjonalna do wartości gradientu koncentracji, a współczynnik proporcjonalności nosi nazwę *współczynnika dyfuzji*:

$$J_{dn} = qD_n \text{ grad } n; \quad J_{dp} = -qD_p \text{ grad } p \quad (1.33)$$

Znak prądu elektronów jest dodatni, a prądu dziur — ujemny, gdyż zgodnie z rys. 1.25a, b przy dodatnich gradientach ($dn/dx > 0$, $dp/dx > 0$) konwencjonalny kierunek prądu elektronów (przeciwny względem kierunku strumienia elektronów) jest zgodny z dodatnim kierunkiem osi x , a prąd dziur płynie w ujemnym kierunku osi x (wielkości q , D_n , D_p są podstawiane ze znakami dodatnimi).

Istnieje ścisły związek między ruchliwością a współczynnikiem dyfuzji, określony tzw. *wzorem Einsteina*

$$\frac{D}{\mu} = \frac{kT}{q} \quad (1.34)$$

Unoszenie i dyfuzja

1.6.3

Uwzględniając składową unoszenia i składową dyfuzji prądu elektronów oraz dziur, gęstość prądu w półprzewodniku można zapisać w postaci:

$$\begin{cases} J = J_n + J_p \\ J_n = q\mu_n nE + qD_n \text{ grad } n \\ J_p = q\mu_p pE - qD_p \text{ grad } p \end{cases} \quad (1.35)$$

Półprzewodnik w stanie nierównowagi termodynamicznej

1.7

W punkcie 1.5 rozpatrzono półprzewodnik będący w stanie równowagi termodynamicznej, w którym $np = n_i^2$ oraz ~~szybkość generacji nośników (szybkość powstawania par elektron-dziura wskutek „przeskoku” elektronów z pasma walencyjnego do pasma przewodnictwa) jest równa szybkości ich rekombinacji (szybkości zaniku par elektron-dziura wskutek „powrotu” elektronów z pasma przewodnictwa do pasma walencyjnego):~~

$$G_0 = R_0 \quad (1.36)$$

przy czym G_0 , R_0 — szybkość generacji i rekombinacji w stanie równowagi termodynamicznej (liczba par elektron-dziura powstających (znikających) w 1 m³ objętości kryształu w czasie 1 s).

Obecnie interesują nas zjawiska występujące w półprzewodniku przy zaburzeniu stanu równowagi termodynamicznej. Stan nierównowagi może być spowodowany

napromieniowaniem półprzewodnika (światło, promienie X itp.) lub dostarczeniem czy usunięciem pewnej liczby nośników wskutek przepływu prądu (np. w złączu $p-n$, które będzie rozpatrywane w rozdz. 3, przy odpowiedniej polaryzacji zewnętrznej do obszaru typu n są dostarczane dodatkowe dziury z przylegającego obszaru typu p). W sposób najbardziej ogólny stan nierównowagi można określić jako taki stan, w którym

$$np \neq n_i^2 \quad (1.37)$$

Jeśli do danego obszaru półprzewodnika nośniki są dostarczane, to $np > n_i^2$; wówczas mówi się o wstrzykiwaniu (iniekcji) nośników. W przypadku natomiast $np < n_i^2$ mówi się o wyciąganiu (ekstrakcji) nośników. Koncentracje nośników w stanie nierównowagi można zapisać w postaci:

$$\left. \begin{aligned} n &= n_0 + n' \\ p &= p_0 + p' \end{aligned} \right\} \quad (1.38)$$

przy czym: n_0, p_0 — koncentracje elektronów, dziur w stanie równowagi ($n_0 p_0 = n_i^2$); n', p' — koncentracje dodatkowych elektronów, dziur w stanie nierównowagi; te dodatkowe nośniki nazywa się nadmiarowymi lub nierównowagowymi.

Wiadomo, że w stanie równowagi w dowolnym makroskopowym obszarze półprzewodnika obowiązuje warunek obojętności elektrycznej, tj.

$$q = q(N_D - N_A + p_0 - n_0) = 0$$

A jak jest w stanie nierównowagi? W przypadku naświetlenia półprzewodnika są generowane pary elektron-dziura, zatem $n' = p'$ i warunek obojętności elektrycznej jest spełniony. W przypadku natomiast wstrzykiwania lub wyciągania nośników zagadnienie to komplikuje się. Jeśli na przykład do półprzewodnika typu n są wstrzykiwane dziury, których rozkład koncentracji jest nierównomierny, to wskutek oddziaływania elektrostatycznego ładunek dodatni dziur powoduje bardzo szybkie (stała czasowa ok. 10^{-12} s) przegrupowanie ujemnego ładunku elektronów. Obojętność elektryczna byłaby w pełni przywrócona, gdyby rozkład koncentracji elektronów nadmiarowych był identyczny z rozkładem koncentracji dziur nadmiarowych. Jednakże taka sytuacja oznaczałaby całkowity zanik pola elektrycznego, będącego przyczyną przegrupowania elektronów. Jednocześnie wskutek nierównomiernego rozkładu koncentracji nastąpi dyfuzja elektronów, co doprowadzi do naruszenia obojętności elektrycznej. Stąd wynika, że w stanie ustalonym strumień dyfuzyjny elektronów jest kompensowany przez przeciwnie skierowany strumień unoszeniowy. A to oznacza, że w stanie ustalonym istnieje pewne pole elektryczne spowodowane nierównomiernym rozkładem ładunku. Zasada obojętności elektrycznej, ściśle biorąc, jest zatem naruszona. Przy małych poziomach wstrzykiwania, tj. dla $p' \ll n_n$ (wstrzykiwanie dziur do półprzewodnika typu n) lub $n' \ll p_p$ (wstrzykiwanie elektronów do półprzewodnika typu p), zakłócenie stanu obojętności elektrycznej jest niewielkie i wówczas można przyjąć, że $n' \approx p'$. W tym przypadku mówi się, że półprzewodnik jest quasi-obojętny.

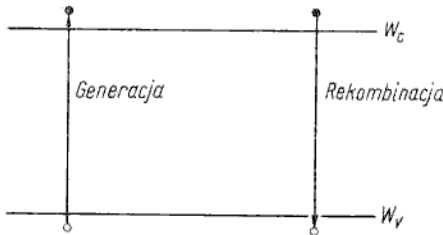
W przypadku dużych poziomów wstrzykiwania, tj. dla p' większego lub porównywalnego z n_n (dla półprzewodnika typu n) oraz n' większego lub porównywalnego z p_p (dla półprzewodnika typu p), należy brać pod uwagę zakłócenie obojętności elektrycznej i wynikające stąd wewnętrzne pole elektryczne.

Poniżej rozważono tylko przypadek małego poziomu wstrzykiwania. Rozpatrzono zjawisko generacji i rekombinacji oraz zagadnienie nieustalonego i ustalonego rozkładu koncentracji nośników w stanie nierównowagi. Wprowadzono trzy nowe pojęcia, tj. czas życia nośników mniejszościowych, drogę dyfuzji, szybkość rekombinacji powierzchniowej.

Szybkość generacji i rekombinacji

1.7.1

Rozpatrzmy najpierw prosty przypadek tzw. generacji i rekombinacji bezpośredniej (rys. 1.26), związanej z bezpośrednimi przejściami elektronów z pasma walencyjnego do pasma przewodnictwa i na odwrót. Do generacji bezpośredniej



Rys. 1.26

Model generacji i rekombinacji bezpośredniej

pary elektron-dziura jest wymagana energia większa niż szerokość pasma zabronionego. Tak duże wartości mają kwanty energii promieniowania świetlnego — fotony (na przykład dla światła czerwonego $h\nu = 1,8$ eV), energia cieplna natomiast, przypadająca na jeden elektron, ma znacznie mniejsze wartości (w temperaturze pokojowej wartość średnia tej energii wynosi $\frac{3}{2} kT \approx 0,04$ eV,

jeżeli elektron potraktuje się na gruncie fizyki klasycznej jako cząstkę gazu doskonałego). Jeśli porówna się wyżej podane wartości energii z szerokością pasma zabronionego (dla Si $W_g = 1,1$ eV), to stanie się jasne, że generacja bezpośrednia pod wpływem energii cieplnej jest bardzo mało prawdopodobna¹⁾, z łatwością natomiast jest powodowana przez energię świetlną. W procesie rekombinacji przejście elektronu z pasma przewodnictwa do pasma walencyjnego łączy się z oddaniem pewnej porcji energii w postaci ciepła lub promieniowania świetlnego (również w zakresie niewidzialnym). Na podstawie rozważań, znacznie wykraczających poza ramy tego rozdziału, udowadnia się, że energia uwalniana przy bezpośrednim przejściu rekombinacyjnym w krzemie i germanie ma postać ciepła, w arsenku galu natomiast — promieniowania świetlnego.

Intuicja podpowiada nam, że szybkość rekombinacji powinna być proporcjonalna do koncentracji elektronów i dziur. Istnieje ogólnie słuszna zasada, że szybkość reakcji dwu składników jest proporcjonalna do ich koncentracji. (Częstość zawierania związków małżeńskich w określonym obszarze geograficznym jest proporcjonalna do liczby kobiet oraz liczby mężczyzn zamieszkujących ten obszar,

¹⁾ Znikome prawdopodobieństwo tej generacji znajduje potwierdzenie w małej wartości koncentracji nośników samoistnych; $n_i = 1,5 \cdot 10^{16} \text{ m}^{-3}$ dla krzemu oznacza, że pod wpływem energii cieplnej w sieci krystalicznej krzemu tylko jedno wiązanie na ok. 10^{12} jest zerwane (koncentracja atomów krzemu wynosi ok. $5 \cdot 10^{28} \text{ m}^{-3}$).

dlatego jest na pewno większa w Warszawie niż na Pustyni Błędowskiej¹⁾. Stąd przykładowo dla półprzewodnika typu n można napisać

$$R = rn_n p_n \quad (1.39)$$

przy czym r — współczynnik proporcjonalności.

Wzór (1.39) jest słuszny zarówno w stanie równowagi, jak i nierównowagi, wskazuje on zatem na wzrost szybkości rekombinacji w czasie wstrzykiwania nośników do określonego obszaru półprzewodnika. Z kolei szybkość generacji nośników jest proporcjonalna do liczby elektronów w pasmie walencyjnym oraz liczby wolnych poziomów energetycznych w pasmie przewodnictwa. Obie te wielkości bardzo mało zmieniają się w stanie nierównowagi (w każdym razie dla półprzewodnika niezdegenerowanego). Dlatego można przyjąć, że szybkość generacji w stanie nierównowagi jest taka jak w stanie równowagi. A ponieważ w stanie równowagi szybkość generacji jest równa szybkości rekombinacji, zatem

$$G = R_0 = rn_{n0} p_{n0} \quad (1.40)$$

Zmiany koncentracji nośników nadmiarowych w funkcji czasu — stan nieustalony

1.7.2

Rozpatruje się zagadnienie powrotu półprzewodnika do stanu równowagi termodynamicznej po usunięciu bodźca zakłócającego równowagę (na przykład po wstrzyknięciu lub wygenerowaniu nośników nadmiarowych pod wpływem oświetlenia).

Szybkość zmian koncentracji nośników mniejszościowych, powodowanych wyłącznie procesami generacyjno-rekombinacyjnymi, można zapisać równaniem (dla półprzewodnika typu n)

$$\frac{dp_n}{dt} = G - R = r(n_{n0} p_{n0} - n_n p_n) \quad (1.41)$$

Dla małego poziomu wstrzykiwania można przyjąć, że $n_n \approx n_{n0}$, zatem

$$\frac{dp_n}{dt} \approx rn_{n0}(p_{n0} - p_n) \quad (1.42)$$

Ponieważ $(p_n - p_{n0}) = p'_n$ jest koncentracją nośników mniejszościowych nadmiarowych, po rozwiązaniu równania różniczkowego (1.42) otrzymuje się

$$p'_n(t) = p'_n(0) \exp\left(-\frac{t}{\tau}\right) \quad (1.43)$$

przy czym wielkość

$$\tau = \frac{1}{rn_{n0}} \quad (1.44)^{2)}$$

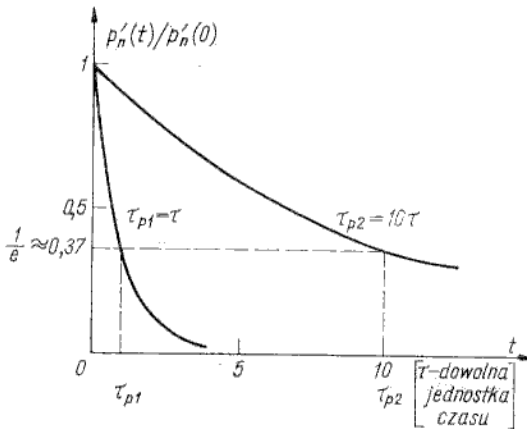
¹⁾ Ta żartobliwa ilustracja tzw. „prawa reakcji mas” nie pretenduje oczywiście do ścisłości demograficznej.

²⁾ Dla półprzewodnika typu p $\tau = 1/rp_{p0}$, a bardziej ściśle i ogólnie dla półprzewodnika typu p lub typu n $\tau = 1/r(n_0 + p_0)$.

jest nazywana czasem życia nośników mniejszościowych. Uwzględniając (1.44), (1.42) można napisać

$$\frac{dp_n}{dt} = \frac{p_{n0} - p_n}{\tau} \quad (1.45)$$

Równania (1.43), (1.45) opisują proces powrotu półprzewodnika do stanu równowagi po wstrzyknięciu nośników mniejszościowych nadmiarowych. W związku z tym czas życia nośników mniejszościowych można zdefiniować jako przedział czasu, w którym koncentracja nośników nadmiarowych maleje e-krotnie (ok. 63%) — rys. 1.27.



Rys. 1.27
Zależność koncentracji nośników nadmiarowych w funkcji czasu dla różnych wartości czasu życia

Wyniki dotychczasowych rozważań można w prosty sposób rozszerzyć dla przypadku stanu ustalonego (ale nierównowagowego) spowodowanego stałym naświetleniem półprzewodnika. Należy wówczas uwzględnić dodatkowy składnik generacji nośników G_L , spowodowanej naświetleniem

$$\frac{dp_n}{dt} = G_L + \frac{p_{n0} - p_n}{\tau} \quad (1.46)$$

skąd dla stanu ustalonego ($dp_n/dt = 0$)

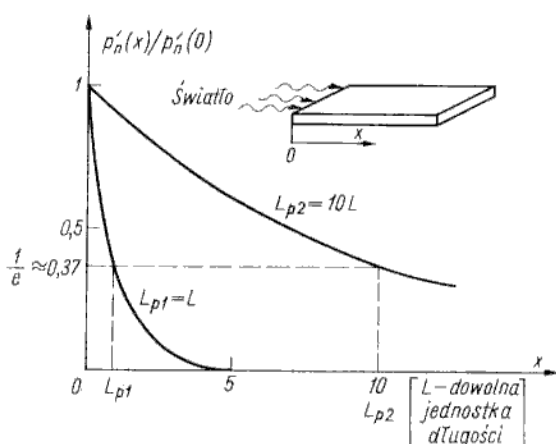
$$p'_n = G_L \tau \quad (1.47)$$

czyli koncentracja nośników nadmiarowych generowanych wskutek naświetlenia jest równa iloczynowi szybkości generacji przez czas życia. Znamy rozwiązanie zagadnienia zmian koncentracji wstrzykniętych nośników nadmiarowych w funkcji czasu. Drugim zagadnieniem, typowym dla pracy wielu przyrządów półprzewodnikowych, jest rozkład koncentracji nośników nadmiarowych w funkcji odległości przy ustalonym poziomie wstrzykiwania lub wyciągania tych nośników.

Rozkład koncentracji nośników nadmiarowych przy ustalonym poziomie wstrzykiwania lub wyciągania

1.7.3

Zgodnie z rys. 1.28 (wstawka) do prostopadłościennej bryły półprzewodnika przez ścianę o współrzędnej $x = 0$ (np. wskutek naświetlania tej powierzchni



Rys. 1.28
Ustalony rozkład koncentracji
nośników nadmiarowych

promieniowaniem pochłanianym w bardzo cienkiej warstwie półprzewodnika) są dostarczane nośniki nadmiarowe. Wskutek nierówności koncentracji nośników przy powierzchni i w głębi półprzewodnika wystąpi przepływ dyfuzyjny nośników nadmiarowych w głąb bryły półprzewodnika.

Jeżeli weźmie się mały wycinek bryły półprzewodnika o powierzchni jednostkowej i długości Δx , to szybkość zmian ładunku nośników mniejszościowych w tym obszarze jest równa różnicy strumieni dyfuzyjnych wpływającego i wypływającego z tego obszaru

$$q \frac{\partial p_n}{\partial t} \Delta x = q D_p \left. \frac{\partial p_n}{\partial x} \right|_{x_1} - q D_p \left. \frac{\partial p_n}{\partial x} \right|_{x_2} \quad (1.48)$$

przy czym: x_1, x_2 — lewa i prawa granica wycinka Δx .

Przy $\Delta x \rightarrow 0$ wzór (1.48) można przepisać w postaci

$$\frac{\partial p_n}{\partial t} = D_p \frac{\partial^2 p_n}{\partial x^2} \quad (1.49)$$

Jeżeli jednocześnie uwzględni się proces rekombinacji, to na podstawie (1.49), (1.45) otrzymuje się

$$\frac{\partial p_n}{\partial t} = D_p \frac{\partial^2 p_n}{\partial x^2} + \frac{p_{n0} - p_n}{\tau} \quad (1.50)$$

W stanie ustalonym $\partial p_n / \partial t = 0$, zatem

$$D_p \frac{\partial^2 p_n}{\partial x^2} = \frac{p_n - p_{n0}}{\tau} \quad (1.50a)$$

Rozwiązanie równania (1.50a) przy warunkach brzegowych:

$$p_n(0) = \text{const}$$

$$p_n(\infty) = p_{n0}$$

ma następującą postać:

$$p_n(x) = p_{n0} + [p_n(0) - p_{n0}] \exp \frac{-x}{L_p} \quad (1.51)$$

lub w zapisie dla nośników nadmiarowych

$$p'_n(x) = p'_n(0) \exp \frac{-x}{L_p} \quad (1.52)$$

przy czym

$$L_p = \sqrt{D_p \tau_p} \quad (1.53)$$

Wielkość L_p jest nazywana *drogą dyfuzyjną* i definiowana jako odległość, na której koncentracja nośników nadmiarowych maleje e -krotnie (ok. 63%) — rys. 1.28.

W sytuacji odwrotnej, gdy nośniki nadmiarowe o równomiernej koncentracji p'_n (np. równomierne naświetlenie całej bryły półprzewodnika) są wyciągane na jednej ze ścian ograniczających tę bryłę (na przykład dla współrzędnej $x = 0$, $p'_n(0) = 0$), w wyniku podobnego rozumowania można otrzymać rozkład koncentracji nośników nadmiarowych w postaci

$$p'_n(x) = p'_n(\infty) \left(1 - \exp \frac{-x}{L_p} \right) \quad (1.54)$$

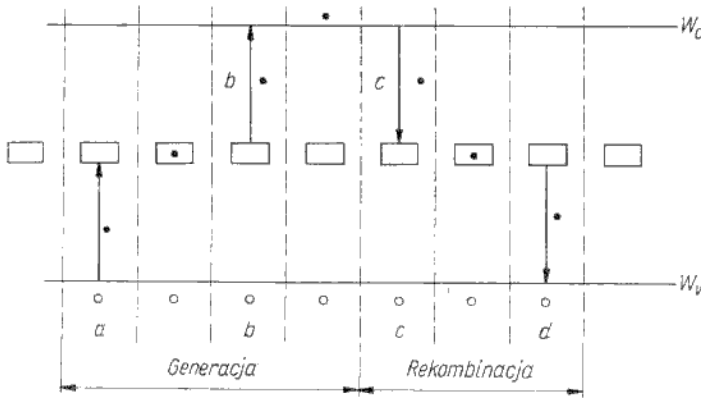
W wyprowadzonych wyrażeniach na zależność koncentracji nośników nadmiarowych od czasu (stan nieustalony — (1.43)) oraz odległości (stan ustalony, wstrzykiwanie — (1.52), wyciąganie — (1.54)) występują dwa nowe parametry materiałowe, tj. czas życia τ i droga dyfuzyjna L . Wyrażenie (1.44) określające czas życia τ wyprowadzono dla uproszczonego modelu generacji-rekombinacji bezpośredniej. Taki model ma znaczenie praktyczne tylko w niektórych półprzewodnikach, np. Ga As. W krzemie i germanie czas życia, a zatem i droga dyfuzyjna są silnie zależne od niektórych zanieczyszczeń. W tym przypadku czas życia jest opisany innym wzorem, wyprowadzonym dla modelu tzw. generacji-rekombinacji pośredniej.

Generacja-rekombinacja pośrednia

1.7.4

Niektóre domieszki, np. miedź w germanie lub złoto w krzemie, powodują powstanie dozwolonych poziomów energetycznych ulokowanych w pobliżu środka pasma zabronionego. Atomy takich domieszek z mniej więcej jednakowym prawdopodobieństwem mogą wymieniać elektrony z pasmem przewodnictwa i pasmem walencyjnym. Domieszki te nie mają zatem charakteru ani donorowego, ani akceptorowego, spełniają natomiast funkcję tzw. *centrów generacyjno-rekombinacyjnych*.

Na rysunku 1.29 przedstawiono możliwe warianty przejść elektronów między pasmami przewodnictwa i walencyjnym a poziomem energetycznym centrów generacyjno-rekombinacyjnych. Widać, że centrum generacyjno-rekombinacyjne spełnia funkcję jakby „stacji przesiadkowej” dla elektronów przechodzących przez pasmo zabronione. Ponieważ prawdopodobieństwo „skoku” elektronu na wyższy poziom energetyczny zwiększa się wykładniczo wraz ze zmniejszaniem się wiel-



Rys. 1.29
Model generacji-rekombinacji pośredniej

kości tego „skoku”, generacja pośrednia, czyli przejście elektronu z pasma walencyjnego do pasma przewodnictwa dwoma „skokami” (fazy *a*, *b* na rys. 1.29), jest wielokrotnie łatwiejsza, a więc i bardziej prawdopodobna niż generacja bezpośrednia. Zatem w obecności centrów generacyjno-rekombinacyjnych zwiększa się szybkość generacji i rekombinacji, czyli maleje czas życia nośników.

Analizę teoretyczną zjawiska generacji-rekombinacji pośredniej wykonali w 1952 r. Shockley, Read oraz Hall (centra generacyjno-rekombinacyjne są często nazywane centrami SRH od pierwszych liter nazwisk tych trzech uczonych). W tym miejscu przytoczymy tylko jeden z najważniejszych wyników tej analizy¹⁾, wzór na czas życia nośników mniejszościowych:

$$\tau_p = \frac{1}{\theta_p v_{th} N_t} \quad \text{dla półprzewodnika typu } n \quad (1.55)$$

$$\tau_n = \frac{1}{\theta_n v_{th} N_t} \quad \text{dla półprzewodnika typu } p \quad (1.56)$$

przy czym: θ_p , θ_n — przekrój skuteczny na wychwyt dziur, elektronów; v_{th} — prędkość cieplna nośników; N_t — koncentracja centrów generacyjno-rekombinacyjnych (indeks *t* od ang. *trap* — pułapka).

Sens fizyczny wzorów (1.55), (1.56) można łatwo zaakceptować, jeśli uwzględni się, że czas życia jest odwrotnością częstości chwywania nośników przez centra generacyjno-rekombinacyjne (g-r). Częstość chwywania jest wprost proporcjonalna do koncentracji centrów g-r, do szybkości ruchu cieplnego nośników (im szybciej się porusza nośnik, tym większą drogę przebywa w jednostce czasu, a większa droga to również większe prawdopodobieństwo trafienia na centrum g-r) oraz przekroju skutecznego na wychwyt, określającego jak blisko centrum g-r musi się znaleźć nośnik, by zostać schwytanym²⁾.

¹⁾ Shockley W., Read W. T., Jr: Statistics of the recombinations of holes and electrons. *Physical Review*, 1952, Vol. 87, p. 835; Hall R. N.: Electron-hole recombination in germanium. *Physical Review*, 1952, Vol. 87, p. 387.

²⁾ Wartość przekroju skutecznego na wychwyt zależy od rozkładu potencjału wokół centrum g-r. Jeżeli centrum g-r jest neutralne elektrycznie, to θ jest powierzchnią takiego rzędu jak dla przekroju atomu, tj. ok. 10^{-19} m². Dla centrum zjonizowanego, czyli przyciągającego ładunek przeciwnego znaku zgodnie z prawem Coulomba, $\theta \approx 10^{-16}$ m².

Należy zauważyć, że oprócz procesów przedstawionych na rys. 1.29 są również możliwe inne kombinacje procesów chwywania i emisji nośników przez centra g-r. Na przykład po schwywaniu elektronu z pasma przewodnictwa (proces *c* na rys. 1.29) może nastąpić emisja tego elektronu z powrotem do pasma przewodnictwa (proces *b* na rys. 1.29). W ten sposób nie dojdzie do rekombinacji pary elektron-dziura, lecz elektron przez pewien czas będzie „przetrzymywany” przez centrum. W podobny sposób może być „przetrzymywana” dziura schwywana z pasma walencyjnego. Centra, dla których prawdopodobieństwo „przetrzymywania” nośników jest większe niż generacji-rekombinacji, są nazywane *pułapkami*.

Dotychczas rozpatrywaliśmy procesy generacyjno-rekombinacyjne w głębi bryły półprzewodnika. Nieco inne warunki istnieją na powierzchni, gdzie funkcję dodatkowych centrów generacyjno-rekombinacyjnych spełniają zerwane wiązania sieci krystalicznej oraz wszelkie zanieczyszczenia powierzchni. Defekty powierzchni tworzą quasi-ciągły zbiór dozwolonych poziomów energetycznych, tzw. *stanów powierzchniowych*. Stany powierzchniowe jako dodatkowe centra g-r zmniejszają czas życia nośników przy powierzchni.

Uwzględniając zjawiska występujące wewnątrz i na powierzchni półprzewodnika czas życia nośników w cienkim obszarze przypowierzchniowym można zapisać w postaci

$$\frac{1}{\tau} = \frac{1}{\tau_b} + \frac{1}{\tau_s} \quad (1.57)$$

przy czym: τ_b — czas życia objętościowy, tj. wewnątrz bryły półprzewodnika (indeks *b* od ang. *bulk*); τ_s — czas życia powierzchniowy (indeks *s* od ang. *surface*).

Zwykle właściwości powierzchni są charakteryzowane nie czasem życia τ_s , lecz tzw. *szybkością rekombinacji powierzchniowej* s_p :

$$s_p = \frac{1 \text{ metr}}{\tau_s} \quad (1.58)$$

$$s_p = \theta_p v_{th} N_{st} \quad (1.59)$$

przy czym N_{st} — gęstość stanów powierzchniowych liczona na 1 m² powierzchni. Szybkość rekombinacji powierzchniowej dla odkrytej powierzchni krzemu wynosi kilka m/s, natomiast w przypadku stosowania tzw. *pasywacji*, tj. pokrycia krzemu warstwą SiO₂, jest bardzo mała, mniej niż 0,10 m/s.

Prawo ciągłości ładunku

1.8

Ładunek nośników w zamkniętym obszarze o powierzchni 1 m² i nieskończenie małej grubości może się zmieniać wskutek różnicy strumieni nośników wpływających i wypływających z tego obszaru oraz różnicy szybkości procesów generacji i rekombinacji. Na przykład w półprzewodniku typu *n*:

$$\frac{\partial p_n}{\partial t} = \frac{1}{q} \frac{\partial J_p}{\partial x} + G - R \quad \text{dla dziur} \quad (1.60)$$

$$\frac{\partial n_n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + G - R \quad \text{dla elektronów} \quad (1.61)$$

Są to tzw. *równania ciągłości ładunku*.

Półprzewodnik niejednorodny

1.9

Półprzewodnikiem niejednorodnym nazywa się półprzewodnik o nierównomiernym rozkładzie koncentracji domieszek. Taka sytuacja jest typowa dla wielu przyrządów półprzewodnikowych. W stanie równowagi termodynamicznej prądy dziurowy i elektronowy są równe zeru. Na przykład dla półprzewodnika typu n :

$$\mu_p p_n E - D_p \frac{dp_n}{dx} = 0 \quad (1.62a)$$

$$\mu_n n_n E + D_n \frac{dn_n}{dx} = 0 \quad (1.62b)$$

Rozwiązując te równania względem E przy uwzględnieniu zależności Einsteina (1.34) otrzymuje się

$$E = \frac{kT}{q} \frac{1}{p_n} \frac{dp_n}{dx} \quad (1.63a)$$

lub

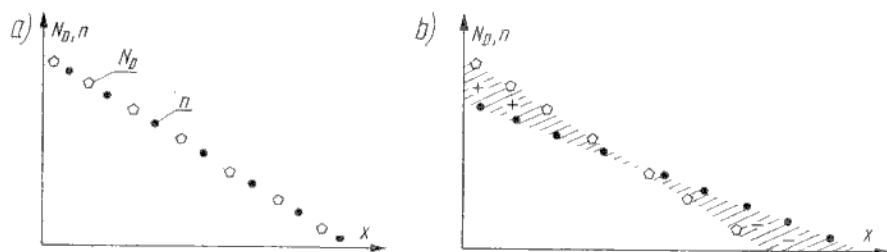
$$E = - \frac{kT}{q} \frac{1}{n_n} \frac{dn_n}{dx} \quad (1.63b)$$

W pierwszym przybliżeniu można przyjąć, że $n_n = N_D$, wówczas

$$E = - \frac{kT}{q} \frac{1}{N_D} \frac{dN_D}{dx} \quad (1.64)$$

Zatem w przypadku nierównomiernego rozkładu domieszek w półprzewodniku istnieje tzw. *wbudowane pole elektryczne*.

Sens fizyczny zjawisk zachodzących w półprzewodniku niejednorodnym ilustruje pogłębowo rys. 1.30, na którym przedstawiono rozkłady koncentracji domieszki i elek-



Rys. 1.30

ilustracja powstawania wbudowanego pola elektrycznego w półprzewodniku niejednorodnym

tronów dla półprzewodnika typu n . Na rysunku 1.30a przedstawiono sytuację „początkową”, charakteryzującą się tym, że jest zachowana neutralność elektryczna w każdym obszarze lokalnym półprzewodnika, przy czym koncentracja domieszek (gęstość ładunku ρ_D) maleje liniowo w funkcji współrzędnej x . Ponieważ można przyjąć, że każdy atom domieszki donorowej „daje” jeden elektron, rozkład koncentracji elektronów (rozkład gęstości ładunku ρ_n) jest identyczny

z rozkładem koncentracji domieszki. Jednakże wskutek gradientu koncentracji elektronów nastąpi ich przepływ dyfuzyjny, powodujący zachwianie lokalnej neutralności elektrycznej. Z jednej strony pozostaną nieskompensowane jony dodatnie domieszki donorowej, z drugiej strony będzie się gromadził nieskompensowany ładunek ujemny elektronów, co przedstawiono na rys. 1.30b.

Wskutek zachwiania neutralności elektrycznej powstaje pole elektryczne przeciwdziałające dyfuzji i unoszące elektrony w kierunku przeciwnym do strumienia dyfuzyjnego. Ostatecznie stan równowagi charakteryzuje się istnieniem wbudowanego pola elektrycznego o wartości zapewniającej równowagę prądów dyfuzji i unoszenia.

Najważniejsze wzory fizyki półprzewodników

Tablica 1.2

Równanie neutralności elektrycznej		$\rho = q(p - n + N_D - N_A) = 0$
Bilans koncentracji nośników		$pm = n_i^2$
Koncentracja nośników samoistnych		$n_i = B \left(\frac{T}{T_0} \right)^{3/2} \exp \left(- \frac{W_g}{2kT} \right)$ $B = 1,7 \cdot 10^{25} \text{ m}^{-3} \text{ dla Si}$ $T_0 = 300 \text{ K}$
Poziom Fermiego w półprzewodniku samoistnym		$W_i = \frac{1}{2}(W_c + W_v) + kT \ln \frac{N_v}{N_c} \approx \frac{W_c + W_v}{2}$
Koncentracja elektronów i dziur	półprzewodnik zdegenerowany	$n = N_c f_n(W_c)$ $p = N_v f_p(W_v)$ <p>przy czym:</p> $f_n(W_c) = \frac{1}{1 + \exp \frac{W_c - W_F}{kT}}$ $f_p(W_v) = \frac{1}{1 + \exp \frac{W_F - W_v}{kT}}$
	półprzewodnik niezdegenerowany	$n = N_c \exp \frac{W_F - W_c}{kT}$ $p = N_v \exp \frac{W_v - W_F}{kT}$ <p>lub w innej postaci</p> $n = n_i \exp \left(- \frac{q\varphi_F}{kT} \right)$ $p = n_i \exp \left(\frac{q\varphi_F}{kT} \right)$ <p>przy czym $\varphi_F = \frac{W_i - W_F}{q}$</p>

Tabl. 1.2 (od.)

Koncentracja nośników większościowych	$n_n = \frac{1}{2} [N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_i^2}]$ $p_p = \frac{1}{2} [N_A - N_D + \sqrt{(N_A - N_D)^2 + 4n_i^2}]$ <p>lub w przybliżeniu dla $N_D - N_A \gg n_i$</p> $n_n \approx N_D - N_A$ $p_p \approx N_A - N_D$
Gęstość prądu unoszenia	$J_{un} = qn\mu_n E$ $J_{up} = qp\mu_p E$
Gęstość prądu dyfuzji	$J_{dn} = qD_n \text{ grad } n$ $J_{dp} = -qD_p \text{ grad } p$
Wzór Einsteina	$\frac{D}{\mu} = \frac{kT}{q}$
Szybkość rekombinacji nośników mniejszościowych	$R_n = \frac{n_{p0} - n_p}{\tau_n}$ $R_p = \frac{p_{n0} - p_n}{\tau_p}$
Czas życia nośników	$\tau_n = \frac{1}{\theta_n v_{th} N_t}$ $\tau_p = \frac{1}{\theta_p v_{th} N_t}$
Szybkość rekombinacji powierzchniowej	$s_p = \theta_p v_{th} N_{st}$ $s_n = \theta_n v_{th} N_{st}$
Prawo ciągłości ładunku	$\frac{\partial p}{\partial t} = \frac{1}{q} \frac{\partial J_p}{\partial x} + G_p - R_p$ $\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + G_n - R_n$
Pole wbudowane w półprzewodniku niejednorodnym	$E = -\frac{kT}{q} \frac{1}{N} \frac{dN}{dx}$

ZADANIA

1.1. Oblicz koncentrację elektronów w krzemie, w temperaturze 300 K, jeżeli koncentracja dziur $p = 10^{21} \text{ m}^{-3}$.

Wzór: (1.16)

Dane liczbowe: n_i według tabl. 1.1

Odpowiedź: $2,25 \cdot 10^{11} \text{ m}^{-3}$

1.2. O ile należy zwiększyć temperaturę powyżej pokojowej (300 K), aby wzrosła 2-krotnie koncentracja nośników samoistnych w krzemie?

Wzór: (1.13)

Dane liczbowe: W_g według tabl. 1.1, $k = 8,625 \cdot 10^{-5}$ eV/K

Odpowiedź: około 10 K

1.3. Oblicz szerokość pasma zabronionego Ge, Si, GaAs, korzystając z wykresów $n_i = f\left(\frac{1}{T}\right)$, wykonanych w skali półlogarytmicznej na rys. 1.13.

Wskazówka: We wzorze (1.13) uwzględniamy w pierwszym przybliżeniu tylko czynnik wykładniczy, czyli

$$n_i \sim \exp\left(-\frac{W_g}{2kT}\right)$$

Stąd

$$W_g = 2k \left(\frac{1}{T_1} - \frac{1}{T_2}\right)^{-1} \ln \frac{n_i(T_2)}{n_i(T_1)}$$

przy czym $n_i(T_1)$ oraz $n_i(T_2)$ odczytujemy z wykresu dla dwóch dowolnie wybranych temperatur T_1 i T_2 (w skali Kelvina).

Odpowiedź: patrz tabl. 1.1

1.4. Porównaj wartości funkcji $f_n(W)$ dla rozkładu Fermiego-Diraca i rozkładu Boltzmanna, jeżeli $W - W_F = 2kT$. Wyznacz względną różnicę tych wartości.

Wzory: (1.6a) i (1.7)

Odpowiedź: różnica wynosi 15%

1.5. Wyznacz koncentrację nośników w krzemie, w temperaturze pokojowej, jeśli jest znana wartość potencjału Fermiego

$$\varphi_F = 0,3 \text{ V}$$

Wzory: (1.14) i (1.15)

Dane liczbowe: n_i według tabl. 1.1, k jak w zadaniu 1.2 lub prościej $\frac{kT}{q} = 0,025 \text{ V}$

Odpowiedź: $n = 2,5 \cdot 10^{21} \text{ m}^{-3}$; $p = 9 \cdot 10^{10} \text{ m}^{-3}$

1.6. Wyznacz położenie poziomu Fermiego w krzemie typu p o koncentracji domieszek $N_A = 1,5 \cdot 10^{21} \text{ m}^{-3}$, w temperaturze pokojowej

Wzory: (1.15) przy uwzględnieniu, że $p \approx N_A$

Dane liczbowe: jak w zadaniu 1.5

Odpowiedź: $\varphi_F = 0,29 \text{ V}$

1.7. Wyprowadź wzory Boltzmanna (1.14) i (1.15)

Rozwiązanie: Zapisujemy wyrażenie (1.9) dla przypadku półprzewodnika samoistnego

$$n_i = N_c \exp \frac{W_i - W_c}{kT}$$

Dzieląc stronami to równanie z równaniem (1.9) otrzymujemy zależność (1.14). Analogicznie wyprowadza się zależność (1.15) dla dziur.

1.8. Oblicz konduktywność krzemu typu n o koncentracji $N_D = 10^{21} \text{ m}^{-3}$, w temperaturze pokojowej

Wzory: (1.32) oraz (1.16) przy uwzględnieniu, że $n \approx N_D$

Dane liczbowe: μ_n , μ_p według tabl. 1.1 oraz $q = 1,6 \cdot 10^{-19} \text{ C}$

Odpowiedź: $21 \text{ I}/\Omega \cdot \text{m}$

1.9. Konduktywność półprzewodnika maleje wraz ze zmniejszaniem koncentracji domieszek, wydaje się zatem, że najmniejszą konduktywność ma półprzewodnik samoistny. Z uwagi jednak na różne wartości ruchliwości dziur i elektronów minimum konduktywności osiąga się dla półprzewodnika lekko domieszkowanego. Wyznacz minimalną wartość konduktywności dla krzemu i porównaj ją z konduktywnością samoistną.

Wskazówka: Znajdź ekstremum funkcji $\sigma(n)$ lub $\sigma(p)$ korzystając z (1.32) oraz (1.16)

Odpowiedź: $\sigma_{min} = 2qn_i \sqrt{\mu_n \mu_p}$

1.10. Wyprowadź wzór Einsteina (1.34)

Rozwiązanie: Rozpatrujemy półprzewodnik niejednorodny (nierównomiernie domieszkowany). W stanie równowagi termodynamicznej (tj. przy braku oświetlenia, gradientów tem-

peratury i zewnętrznego pola elektrycznego) sumaryczne prądy unoszenia i dyfuzji są równe zero zarówno dla elektronów, jak i dziur. Na przykład dla elektronów w półprzewodniku typu n

$$\mu_n n_n E + D_n \frac{dn_n}{dx} = 0$$

Stąd po uwzględnieniu $E = -\frac{d\varphi_F}{dx}$ otrzymujemy

$$\frac{dn_n}{n_n} = \frac{\mu_n}{D_n} d\varphi_F$$

Całkowanie tego równania daje

$$n_n = A \exp \frac{\mu_n}{D_n} \varphi_F$$

Porównując to wyrażenie ze wzorem Boltzmann'a

$$n_n = n_i \exp \frac{q\varphi_F}{kT}$$

stwierdzamy, że

$$\frac{\mu_n}{D_n} = \frac{q}{kT}$$

1.11. Wyznacz drogę dyfuzyjną elektronów w krzemie, jeśli jest znany czas życia $\tau_n = 10^{-6}$ s (temperatura pokojowa).

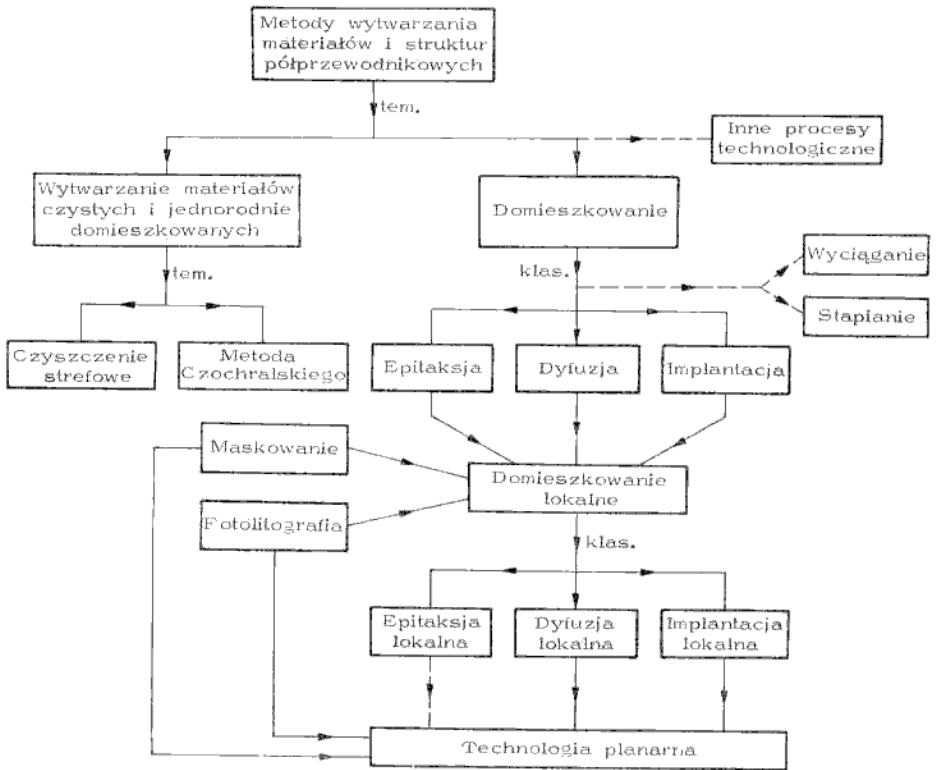
Wzory: (1.53), (1.34)

Dane liczbowe: μ_n według tabl. 1.1 oraz $kT/q = 0,025$ V

Odpowiedź: 57 μm

Znamy już podstawowe właściwości fizyczne półprzewodników samoistnych i domieszkowanych. Przyrząd półprzewodnikowy (dioda, tranzystor) stanowi w sensie technologicznym bryłę materiału półprzewodnikowego, w której wytworzono obszary o różnych koncentracjach i rodzajach domieszek. Inaczej mówiąc, jest to struktura półprzewodnikowa, składająca się z warstw *p*, *n* wytworzonych w monolitycznej bryle półprzewodnika przy użyciu odpowiednich procesów fizykochemicznych.

Zanim rozpatrzymy zjawiska fizyczne w strukturach półprzewodnikowych (rozdziały 3 i dalsze), powinniśmy choćby pobieżnie poznać podstawowe metody kształto-



Rys. 2.1
Schematyczny układ rozdziału

wania struktury fizykochemicznej materiału półprzewodnikowego. Jest to o tyle istotne, że właściwości fizyczne i funkcjonalne przyrządów półprzewodnikowych silnie zależą od ich konstrukcji i technologii.

Przez pojęcie *konstrukcja* elementu półprzewodnikowego rozumie się zespół danych, określających formę geometryczną struktury półprzewodnikowej oraz niektóre parametry fizykochemiczne, a więc: kształt, wielkość i położenie warstw p , n , rozkłady koncentracji domieszek, rezystywności poszczególnych warstw, rozmiary kontaktów itp. W niektórych przypadkach (np. dla elementów mikrofalowych) do istotnych cech konstrukcyjnych elementu należą również kształt i rozmiary jego obudowy.

Przez pojęcie *technologia wytwarzania* elementu półprzewodnikowego rozumie się zespół informacji określających sposoby i kolejność wykonywania określonych operacji w celu zrealizowania zaprojektowanej konstrukcji. Zagadnienia konstrukcyjne będą poruszane w następnych rozdziałach przy okazji omawiania struktur półprzewodnikowych przedstawił w sposób uporządkowany w tym rozdziale. Z uwagi na „nietechnologiczny” profil książki tematyka ta będzie przedstawiona w zakresie elementarnym. Ponieważ większość diod i tranzystorów oraz, praktycznie biorąc, wszystkie układy scalone są obecnie wytwarzane z krzemu, w zasadzie rozważania ograniczono do procesów technologicznych, typowych dla tego materiału. Układ rozdziału przedstawiono schematycznie na rys. 2.1.

Wytwarzanie czystych oraz jednorodnie domieszkowanych materiałów półprzewodnikowych

2.1

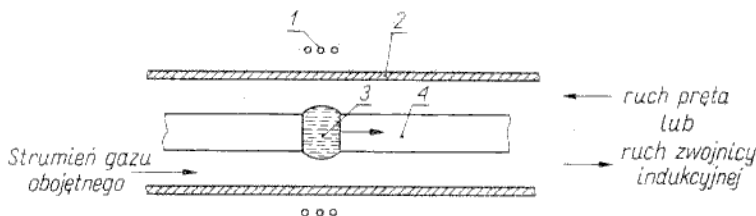
Materiały stosowane w przyrządach półprzewodnikowych muszą spełniać określone wymagania odnośnie do ich właściwości elektrycznych i strukturalnych, przy czym są to właściwości całkowicie współzależne. Podstawowe warunki, jakim powinna odpowiadać struktura materiału półprzewodnikowego, są następujące:

- budowa monokrystaliczna;
- jak najmniejsza liczba defektów;
- precyzyjnie kontrolowana koncentracja domieszek w zakresie $10^{19} \dots 10^{26} \text{ m}^{-3}$.

Podstawowym warunkiem, jaki trzeba spełnić, by możliwe było kontrolowane domieszkowanie materiału półprzewodnikowego na poziomie dolnych wartości koncentracji domieszek (ok. 10^{19} m^{-3}), jest dysponowanie materiałem o koncentracji atomów zanieczyszczeń (domieszek niepożądanych) mniejszej niż 10^{19} m^{-3} . Ponieważ w monokrystalicznym krzemie lub germanie koncentracja atomów wynosi ok. 10^{29} m^{-3} , koncentracja 10^{19} m^{-3} atomów zanieczyszczeń oznacza, że jeden atom pierwiastka obcego przypada na ok. 10 miliardów atomów pierwiastka podstawowego (krzemu lub germanu). Materiał półprzewodnikowy, spełniający to kryterium, nazywa się *technicznie czystym*. W przypadku germanu, w którym koncentracja nośników samoistnych wynosi $2,5 \cdot 10^{19} \text{ m}^{-3}$, czysty oznacza niemal to samo, co samoistny. Natomiast w krzemie koncentracja nośników samoistnych jest 1000 razy mniejsza niż w germanie, dlatego krzem czysty w sensie technicznym nie jest materiałem samoistnym (krzem czysty technicznie ma rezystywność około kilku $\Omega \cdot \text{m}$, podczas gdy samoistny — $2 \cdot 10^3 \Omega \cdot \text{m}$). Obecny stan technologii półprzewodnikowej nie pozwala na otrzymanie krzemu samoistnego, choć czystość materiałów stosowanych w przemyśle półprzewod-

nikowym jest nieporównywalnie większa niż w jakiegokolwiek innej dziedzinie techniki.

Prześledźmy teraz podstawowe etapy wytwarzania czystego krzemu o budowie monokrystalicznej. Krzem jest jednym z najbardziej rozpowszechnionych pierwiastków; występuje w ilości ok. 28% w skorupie ziemskiej w postaci utlenionej jako składnik pospolitego piasku i skał. Mimo powszechnej dostępności surowca wyjściowego krzem czysty jest materiałem drogim ze względu na koszt złożonego procesu oczyszczania. Oczyszczanie chemiczne (reakcja redukcji z węglem w wysokiej temperaturze) umożliwia otrzymanie materiału o zawartości ok. 99,9% krzemu. Ten materiał jest poddawany procesowi oczyszczania strefowego (rys. 2.2). Materiał polikrystaliczny w postaci pręta przesuwany jest powoli w piecu



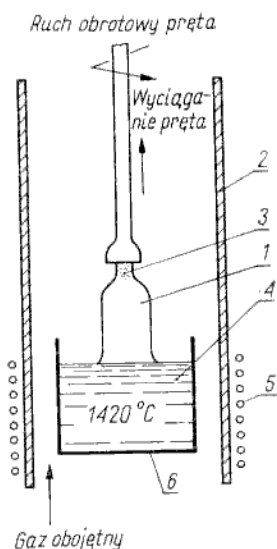
Rys. 2.2

Urządzenie do oczyszczania strefowego

1 — zwojnica indukcyjna; 2 — rura kwarcowa; 3 — strefa roztopionego materiału; 4 — pręt polikrystaliczny

(w atmosferze gazu obojętnego, tj. helu lub argonu, dla uniknięcia utleniania krzemu), przy czym wskutek lokalnego grzania indukcyjnego na krótkim odcinku pręta jest uzyskiwana temperatura wyższa od punktu topnienia krzemu. W ten sposób powstaje lokalna strefa roztopionego materiału, który nie „wycieka” dzięki działaniu sił napięcia powierzchniowego. Ruch pręta w piecu lub ruch zwojnicy grzejnej względem pręta powoduje przesuwanie się strefy roztopionego materiału. Strefa ciekła unosi ze sobą zanieczyszczenia, gdyż mają one większą rozpuszczalność w fazie ciekłej niż w fazie stałej. Po kilku przejściach strefy ciekłej wzdłuż pręta zanieczyszczenia zostają zebrane na jednym końcu pręta. Ten koniec pręta jest następnie odcinany i poddawany powtórnie oczyszczaniu chemicznemu, pozostałą natomiast część pręta można uznać za technicznie czysty krzem o budowie polikrystalicznej. Budowa polikrystaliczna krzemu jest wynikiem losowego charakteru zjawiska rekrytalizacji na obrzeżu przesuwałcej się strefy roztopionego materiału. Zadaniem następnego procesu technologicznego jest otrzymanie krzemu monokrystalicznego. Najczęściej monokryształy krzemu lub germanu są wytwarzane metodą wyciągania z fazy ciekłej, znaną szeroko jako *metoda Czochralskiego*¹⁾. Urządzenie służące do wyciągania monokryształów metodą Czochralskiego przedstawiono schematycznie na rys. 2.3. W tyglu grafitowym lub kwarcowym umieszcza się czysty krzem (german) polikrystaliczny i roztopia go za pomocą nagrzewania indukcyjnego w polu elektrycznym wielkiej częstotliwości (temperatura topnienia krzemu wynosi 1420°C, germanu 937°C). Do roztopionego materiału zanurza się zarodek kryształu o precyzyjnie ustalonej orientacji sieci krystalicznej. Na powierzchni styku za-

¹⁾ Czochralski, polski uczoney, profesor Politechniki Warszawskiej, zastosował tę metodę w 1916 r. i opublikował ją w 1918 r.

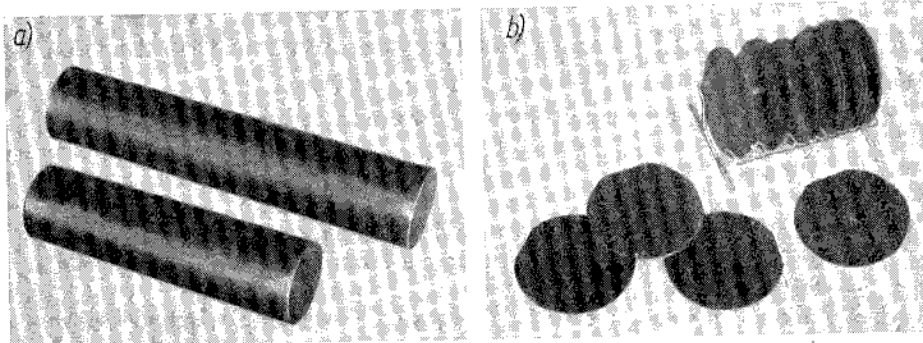


Rys. 2.3

Urządzenie do wyciągania monokryształów z fazy ciekłej metodą Czochralskiego

1 — pręt monokrysztaliczny; 2 — rura kwarcowa; 3 — warstwa kryształu; 4 — roztopiony krzem; 5 — zwojnica indukcyjna; 6 — tygiel grafitowy lub kwarcowy

rodka kryształu z roztopionym materiałem temperatura obniża się o jeden lub dwa stopnie poniżej temperatury topnienia, wskutek czego następuje krystalizacja, tj. atomy poruszające się bezładnie w fazie ciekłej „przyklejają się” w odpowiednich miejscach do powierzchni zarodka krystalizacji. Narastająca warstwa ciała stałego zachowuje dokładnie ciągłość budowy krystalicznej z zarodkiem, przy czym temperatura fazy ciekłej powinna być utrzymywana na poziomie o kilka stopni wyższym niż temperatura topnienia. Przez powolne podnoszenie zarodka (kilka do kilkudziesięciu milimetrów na godzinę) umożliwia się narastanie kolejnych warstw kryształu, czyli powstaje pręt monokrysztaliczny jakby wyciągany z cieczy. Dla zapewnienia jednorodnej budowy kryształu zarodek wraz z wyciąganym prętem obraca się z częstością kilkunastu obrotów na minutę. Zwykle powierzchnia zarodka ma orientację krystalograficzną {111}, charakteryzującą się największą gęstością powierzchniową atomów, co sprzyja narastaniu kryształu o małej liczbie defektów. Typowy pręt monokrysztalicznego



Rys. 2.4

Pręty monokrysztalicznego krzemu, uzyskiwane metodą Czochralskiego (a) oraz płytki otrzymane po cięciu pręta (b)

krzemu uzyskiwany metodą Czochralskiego (rys. 2.4) ma średnicę ok. 5 lub 7,5 cm (2 lub 3 cale) i długość kilkadziesiąt centymetrów, chociaż są również wytwarzane pręty o średnicach większych (do 12 cm). Opisana metoda wyciągania z fazy ciekłej dotyczy zarówno krzemu i germanu, jak również — z pewnymi modyfikacjami — arsenku galu. W przypadku hodowania monokryształów GaAs (temperatura topnienia 1240°C) należy zapobiec rozkładowi GaAs i odparowaniu arsenu. W tym celu atmosfera gazu obojętnego (konieczna również w przypadku krzemu i germanu dla zapobieżenia utlenianiu) powinna zawierać arsen o ciśnieniu cząstkowym 960 hPa.

Warto wspomnieć, że oprócz metody Czochralskiego znanych jest kilka innych metod hodowania monokryształów, między innymi istnieją również metody beztyglowe (na przykład metoda topienia strefowego, analogiczna do przedstawionej na rys. 2.2, lecz z zarodkiem kryształu), interesujące z tego względu, że nie ma tu tygla, który zawsze dostarcza niepożądanych zanieczyszczeń. Jednakże żadna z tych metod nie zyskała tak szerokiego rozpowszechnienia jak metoda Czochralskiego.

Mówiliśmy dotychczas o wytwarzaniu czystych monokryształów, lecz z łatwością można wprowadzać pożądane domieszki do wyciąganego monokryształu półprzewodnika przez dodanie do fazy ciekłej odpowiednich domieszek w ściśle określonych ilościach.

Pręt monokrystaliczny jest cięty wielotarczową piłą diamentową na cienkie płytki o grubości 300...500 μm . Płytką ma charakterystyczny kształt ściętego koła (rys. 2.4), gdyż pręt przed pocięciem na płytki jest poddawany obróbce mechanicznej, nadającej mu kształt ściętego wałka w celu umożliwienia pomiaru orientacji krystalograficznej. Płytki półprzewodnikowe są następnie szlifowane, polerowane mechanicznie i chemicznie (trawienie) do osiągnięcia lustrzanej gładkości, przy czym w wyniku tej obróbki grubość płytki maleje do ok. 200 μm .

Tak przygotowana płytka półprzewodnika monokrystalicznego stanowi podłoże, w którym wytwarza się złożone struktury półprzewodnikowe, czyli diody, tranzystory lub układy scalone. Płytką podłożową jest materiałem półprzewodnikowym o jednorodnych właściwościach fizykochemicznych. Ponieważ struktura półprzewodnikowa zawiera warstwy o różnych właściwościach fizykochemicznych, zatem teraz należy poznać metody kształtowania różnych właściwości fizykochemicznych w lokalnych obszarach półprzewodnika, tj. metody wytwarzania w podłożu półprzewodnikowym obszarów o różnych rodzajach domieszkowania.

Metody wytwarzania warstw domieszkowanych w (lub na) podłożu półprzewodnikowym

2.2

Obecnie są stosowane trzy podstawowe metody:

- epitaksja;
- dyfuzja;
- implantacja.

W przeszłości bardzo dużą rolę odegrały dwie metody:

- domieszkowanie podczas wyciągania monokryształów z fazy ciekłej;
- metoda stopowa.

Wytwarzanie kolejnych warstw P , $N^{1)}$ przez domieszkowanie fazy ciekłej w czasie wyciągania monokryształów powszechnie stosowano na początku lat pięćdziesiątych,

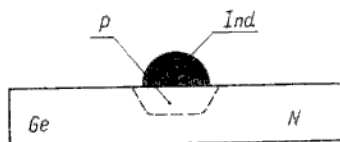
¹⁾ W książce będzie się używać zamiennie zwrotów: „warstwa (półprzewodnik) typu p , typu n ” lub „warstwa (półprzewodnik) P , N ”.

lecz obecnie ta metoda nie ma żadnego znaczenia praktycznego. Metoda stopowa natomiast jest jeszcze niekiedy stosowana przy produkcji elementów germanowych. Dlatego najpierw przedstawimy technologię stopową, po czym omówimy epitaksję, dyfuzję oraz implantację.

Technologia stopowa

2.2.1

Na rysunku 2.5 przedstawiono schematycznie strukturę półprzewodnikową, składającą się z podłoża germanowego typu *n*, w którym wykonano metodą stopową warstwę typu *p*; jest to tzw. złącze *p-n*. Proces wytwarzania takiej



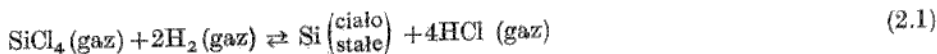
Rys. 2.5
Struktura złącza *p-n* wykonanego metodą stopową

struktury przebiega w następujący sposób. Płytkę germanu typu *n* wraz z kulką indu umieszcza się w kasecie grafitowej i podgrzewa do temperatury ok. 500°C. Jest to temperatura wyższa od punktu topnienia indu, a poniżej punktu topnienia germanu. Roztopiony ind wtapia się do germanu tworząc stop eutektyczny In-Ge. Podczas chłodzenia następuje rekrytalizacja germanu o dużej zawartości indu, tj. domieszki akceptorowej. Zatem warstwa rekrytalizująca germanu jest typu *p*. Istotną wadą tej metody jest jej mała precyzja, zwłaszcza zaś duży rozrzut grubości otrzymywanych warstw typu *p* (rozrzut głębokości złącza *p-n*).

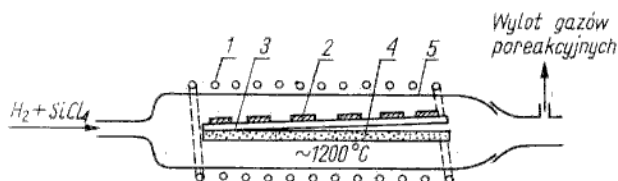
Epitaksja

2.2.2

Słowo *epitaksja* pochodzi z języka greckiego (*epi-*, na; *-taxis*, wytwarzanie) i oznacza proces wytwarzania cienkiej warstwy półprzewodnika monokrystalicznego (w typowych przypadkach o grubości 3...25 μm) na podłożu monokrystalicznym z zachowaniem ciągłości budowy krystalicznej z podłożem. Narastająca warstwa, będąca niejako przedłużeniem podłoża, nosi nazwę *warstwy epitaksjalnej*. W ogólnym przypadku warstwa epitaksjalna może być z innego materiału niż podłoże (na przykład warstwa krzemu na podłożu szafirowym) i mówi się wówczas o *heteroepitaksji*. Największe znaczenie praktyczne ma wytwarzanie warstw epitaksjalnych na podłożu z tego samego materiału (najczęściej warstwa krzemu na podłożu krzemowym). Ściśle biorąc, taki proces nazywa się *homoepitaksją*, lecz zwykle jest nazywany po prostu *epitaksją*. Warstwa epitaksjalna ma identyczną orientację sieci krystalicznej jak podłoże, może jednak znacznie różnić się od podłoża właściwościami elektrofizycznymi, tj. może mieć inny rodzaj przewodnictwa (przykładowo warstwa typu *p* na podłożu typu *n*), inną koncentrację domieszek, a więc również inną rezystywność itp. Istnieje wiele metod otrzymywania warstw epitaksjalnych, spośród których największe znaczenie praktyczne mają metody osadzania chemicznego, zwłaszcza metoda oparta na redukcji czterochlorku krzemu wodorem zgodnie z reakcją



Urządzenie do epitaksji, tzw. reaktor, przedstawiono schematycznie na rys. 2.6. Do rury kwarcowej wpuszcza się gazy reakcyjne (wodór zmieszany z czterochlorkiem krzemu), które opływają powierzchnię płytek krzemowych, ułożonych na podstawie kwarcowej, nagrzanej do temperatury ok. 1200°C. Dotychczas nie ma

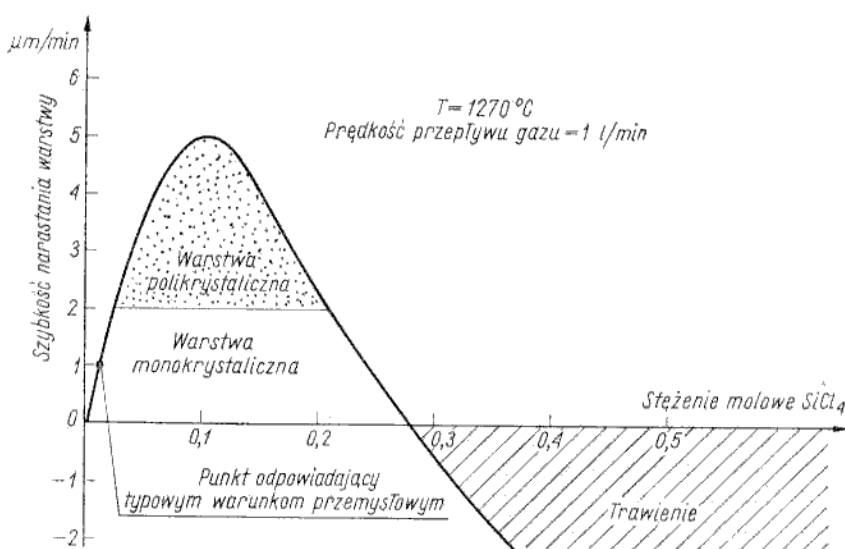


Rys. 2.6

Urządzenie do epitaksji (reaktor)

1 — zwojnica indukcyjna; 2 — płytki podłożowe; 3 — podstawka kwarcowa; 4 — podstawa grafitowa; 5 — rura kwarcowa

ugruntowanej teorii narastania warstwy epitaksjalnej, lecz przypuszcza się, że atomy krzemu, uwalniane wskutek reakcji (2.1), katalizowanej przez powierzchnię płytki krzemowej, wędrują po powierzchni dopóty, dopóki nie zajmą położenia o minimalnych wartościach energii; są to położenia odpowiadające węzłom narastającej sieci krystalicznej. Należy zauważyć, że reakcja (2.1) jest odwracalna, tzn. przy dużym stężeniu HCl może przebiegać z prawa na lewo, co oznacza trawienie płytki krzemowej. Reakcja trawienia jest możliwa również bez udziału HCl przy dużym stężeniu SiCl_4

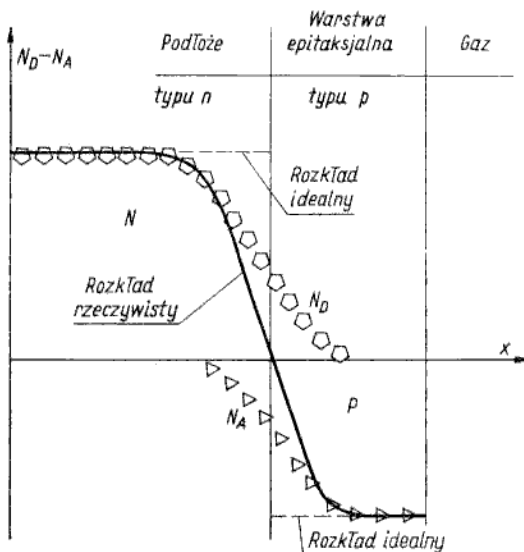


Rys. 2.7

Zależność szybkości narastania (trawienia) warstwy epitaksjalnej od składu gazu reakcyjnego przy ustalonej temperaturze i prędkości przepływu gazu

Szybkości konkurencyjnych reakcji prowadzących z jednej strony do narastania warstwy, z drugiej zaś do trawienia podłoża, zależą od stężenia czterochloru krzemu w wodorze, prędkości przepływu gazu reakcyjnego oraz temperatury. Dlatego temperatura reakcji, skład i prędkość przepływu gazu muszą być dokładnie dobrane. Na rysunku 2.7 przedstawiono wykres zależności szybkości narastania (trawienia) w funkcji stężenia SiCl_4 dla jednej temperatury. W warunkach przemysłowych najczęściej parametry procesu epitaksjalnego są tak dobrane, że szybkość narastania warstwy wynosi $1 \mu\text{m}/\text{min}$. Z rysunku 2.7 można się zorientować, że w miarę wzrostu szybkości narastania pogarsza się jakość warstwy pod względem strukturalnym (więcej defektów), a przy szybkości większej niż $2 \mu\text{m}/\text{min}$ warstwa epitaksjalna ma budowę polikrystaliczną.

Do warstwy półprzewodnika, narastającej w procesie epitaksji, można wprowadzać domieszki donorowe lub akceptorowe o koncentracjach regulowanych w szerokim zakresie. W tym celu najczęściej do reaktora wprowadza się pary PCl_3 (domieszkowanie fosforem w celu wytworzenia warstwy typu n) lub BBr_3 (domieszkowanie borem w celu wytworzenia warstwy typu p), przy czym nośnikiem gazowym par domieszek jest wodór. Bardzo ważne jest uzyskanie ostrego profilu domieszkowania, tj. raptownej zmiany koncentracji domieszek na granicy warstwy epitaksjalnej z podłożem. Ponieważ epitaksja jest procesem wysokotemperaturowym, na granicy warstwy epitaksjalna-podłoże występuje tzw. autodyfuzja, czyli przechodzenie domieszek z podłoża do warstwy epitaksjalnej i w kierunku odwrotnym. Wskutek tego profil domieszkowania jest rozmyty (rys. 2.8). Tę wadę można w znacznym stopniu wyeliminować przez zastosowanie



Rys. 2.8
Rozkład koncentracji domieszek na granicy warstwy epitaksjalnej z podłożem

innej metody osadzania chemicznego, korzystającej z reakcji cieplnego rozkładu silanu (piroliza silanu)



Ten proces przebiega w niższej temperaturze (w temperaturze ok. 1050°C szybkość wzrostu warstwy wynosi 1 $\mu\text{m}/\text{min}$), w której autodyfuzja jest znacznie mniejsza. Jednakże praca z silnym wymaga zachowania szczególnych środków ostrożności, gdyż jest to gaz samozapalający się przy połączeniu z powietrzem.

Dyfuzja

2.2.3

Zjawisko dyfuzji omówiono już w rozdz. I w odniesieniu do ruchomych nośników ładunku (elektronów i dziur). W odpowiednio wysokich temperaturach jest możliwa dyfuzja nie tylko elektronów i dziur, lecz również atomów. Zjawisko to można zatem wykorzystać w procesie domieszkowania półprzewodnika. Jeżeli do powierzchni płytki podłożowej dostarczy się pewną liczbę atomów pierwiastka służącego jako domieszka, to wskutek chaotycznego ruchu cieplnego tych atomów będą one dyfundować w głąb podłoża, tj. przesuwać się z obszarów o większej koncentracji do obszarów o mniejszej koncentracji. Z rozważań zamieszczonych w poprzednim rozdziale są znane dwa podstawowe równania opisujące dyfuzję. Jest to równanie (1.33), które przepisuje się w postaci

$$F = -D \frac{\partial N}{\partial x} \quad \text{I} \quad \text{Fick} \quad \text{Gęstość strumienia} \quad (2.4)$$

oraz równanie (1.49), które przepisuje się w postaci

$$\frac{\partial N}{\partial t} = D \frac{\partial^2 N}{\partial x^2} \quad \text{II} \quad \text{Równ. dyfuzji} \quad (2.5)$$

przy czym: F — gęstość strumienia atomów domieszki, tj. liczba atomów przechodzących w jednostce czasu przez jednostkową powierzchnię prostopadłą do kierunku x ; D — współczynnik dyfuzji; N — koncentracja atomów domieszki, czyli ich liczba w jednostce objętości; x — współrzędna prostopadła do powierzchni płytki półprzewodnikowej; t — czas.

Równanie gęstości strumienia (2.4) oraz równanie dyfuzji (2.5) są znane jako pierwsze i drugie *prawo Ficka*. Równania (2.4) i (2.5) są zapisane w postaci jednowymiarowej (w funkcji jednej współrzędnej przestrzennej x). Nie oznacza to jednak, że dyfuzja przebiega tylko w jednym kierunku. Przeciwnie, dyfuzja w półprzewodniku monokrystalicznym jest zjawiskiem wielokierunkowym, przy czym w przypadku mechanizmu zmiennowęzłowego jest izotropowa, czyli przebiega jednakowo we wszystkich kierunkach. Dla większości zagadnień technicznych istotny jest jednak tylko kierunek w głąb płytki półprzewodnikowej.

Na podstawie rozwiązania równania dyfuzji (2.5) można wyznaczyć rozkład koncentracji domieszek w głąb półprzewodnika (profil domieszkowania), $N(x)$, jeśli są znane tzw. warunki brzegowe procesu dyfuzji. Większość zagadnień występujących w praktyce można opisać następującymi dwoma rodzajami warunków brzegowych:

- *Dyfuzja z nieograniczonego źródła* (ze źródła o stałej wydajności). Chodzi o taki przypadek, gdy nad powierzchnią półprzewodnika jest stała koncentracja atomów domieszki (na przykład ciągły przepływ gazu zawierającego domieszkę).
- *Dyfuzja z ograniczonego źródła*. Chodzi o taki przypadek, gdy całkowita liczba atomów domieszki nie ulega zmianie w procesie dyfuzji (na przykład po krótkotrwałym procesie dyfuzji z fazy gazowej na głębokość δ następuje przerwanie

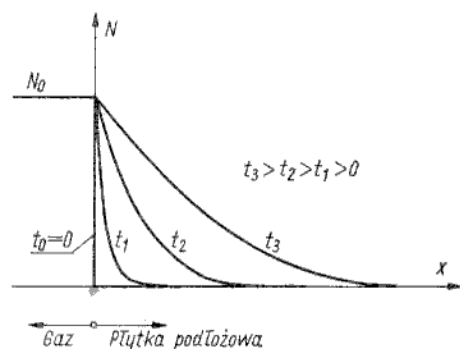
dopływu gazu zawierającego domieszkę, zatem dla dalej zachodzącego procesu dyfuzji źródłem domieszki jest cienki obszar przypowierzchniowy półprzewodnika o grubości δ).

Dyfuzja z nieograniczonego źródła

2.2.3.1

Ten przypadek ilustrują wykresy przedstawione na rys. 2.9. Przy założeniu stałej koncentracji atomów domieszki przy powierzchni (dla $x = 0$)

$$N(0, t) = N_0 = \text{const}$$



Rys. 2.9
Rozkłady koncentracji domieszki dla różnych czasów przy dyfuzji ze źródła nieograniczonego

oraz określonego rozkładu domieszki na początku procesu dyfuzji, tj. rozkładu w postaci schodka od wartości N_0 do

$$N(x, 0) = 0$$

przy trzecim warunku wynikającym z założenia, że grubość płytki podłożowej jest znacznie większa niż głębokość dyfuzji

$$N(\infty, t) = 0$$

otrzymuje się rozwiązanie równania dyfuzji (2.5) w postaci

$$N(x, t) = N_0 \operatorname{erfc} \left(\frac{x}{2L_D} \right) \quad (2.6)$$

przy czym symbol erfc (ang. *error function complement*) oznacza dopełnienie funkcji błędu (jest to funkcja stabelaryzowana w wielu poradnikach matematyki).

Wykresy $N(x, t)$ zgodne z zależnością (2.6) przedstawiono w skali liniowej na rys. 2.9. Całkowitą liczbę M atomów domieszki w półprzewodniku można wyznaczyć całkując funkcję (2.6) od 0 do nieskończoności

$$M(t) = \int_0^{\infty} N(x, t) dx$$

Stąd

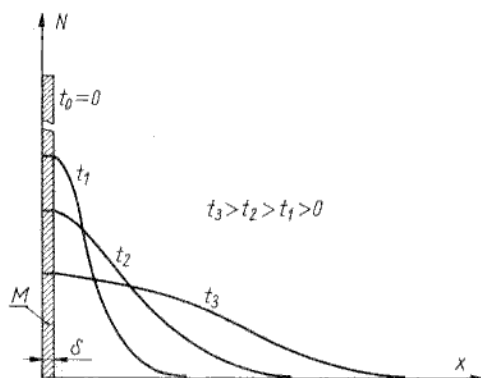
$$M(t) = \frac{2}{\sqrt{\pi}} N_0 L_D = \frac{2}{\sqrt{\pi}} N_0 \sqrt{Dt} \quad (2.7)$$

Dyfuzja z ograniczonego źródła

2.2.3.2

Ten przypadek ilustrują wykresy przedstawione na rys. 2.10. Przy założeniu, że całkowita liczba atomów domieszki jest stała

$$M(t) = \text{const}$$



Rys. 2.10
Rozkłady koncentracji domieszki dla różnych czasów przy dyfuzji ze źródła ograniczonego

(to założenie inaczej oznacza, że nie ma przepływu atomów domieszki przez płaszczyznę $x = 0$, czyli uwzględniając pierwsze prawo Ficka można je sformułować w postaci $\left. \frac{\partial N}{\partial x} \right|_{x=0} = 0$) oraz przyjęciu określonego rozkładu domieszki na początku procesu dyfuzji, tj. w najprostszym przypadku rozkładu w postaci nieskończenie cienkiej warstwy o grubości δ :

$$N(x, 0) = N_0 \quad \text{dla } 0 \leq x \leq \delta$$

$$N(x, 0) = 0 \quad \text{dla } \delta < x$$

przy trzecim warunku wynikającym z założenia, że grubość płytki podłożowej jest znacznie większa niż głębokość dyfuzji

$$N(\infty, t) = 0$$

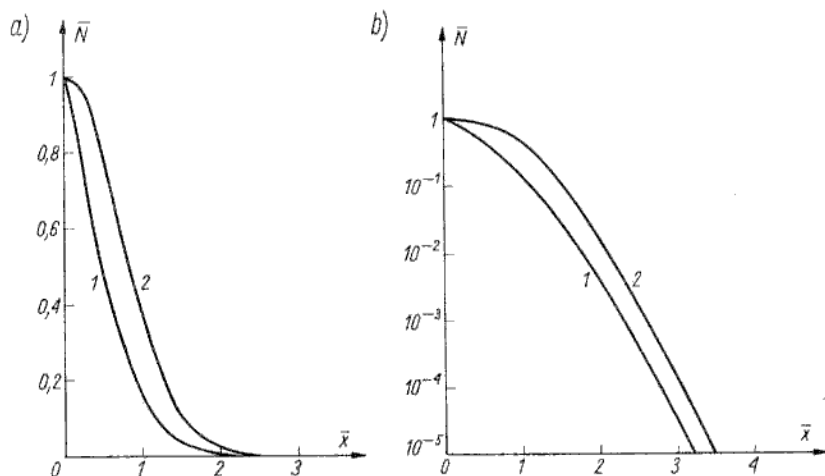
otrzymuje się rozwiązanie równania dyfuzji (2.5) w postaci

$$N(x, t) = \frac{M}{\sqrt{\pi L_D}} \exp \left[- \left(\frac{x}{2L_D} \right)^2 \right] \quad (2.8)$$

Jest to dobrze znany z matematyki i wielu zagadnień fizyki rozkład Gaussa. Powierzchniowa koncentracja domieszki

$$N_0 = \frac{M}{\sqrt{\pi L_D}} = \frac{M}{\sqrt{\pi D t}} \quad (2.9)$$

jest odwrotnie proporcjonalna do pierwiastka z czasu dyfuzji. Wykresy $N(x, t)$ zgodne z zależnością (2.8) przedstawiono na rys. 2.10 w skali liniowej.



Rys. 2.11

Znormalizowana koncentracja \bar{N} w funkcji znormalizowanej współrzędnej \bar{x} dla dwu rozpatrywanych modeli dyfuzji: a) w skali liniowej; b) w skali półlogarytmicznej

1 — rozkład erfc: $N_0 = \text{const}$; $M(t) = \frac{2}{\sqrt{\pi}} N_0 \sqrt{Dt}$; 2 — rozkład Gaussa: $M = \text{const}$;

$$N_0(t) = \frac{M}{\sqrt{\pi Dt}}$$

Dla porównania obu rozwiązań na rys. 2.11 wykreślono funkcje (2.6), (2.8) we współrzędnych znormalizowanych w skali liniowej i półlogarytmicznej. Są to wykresy $\bar{N}(\bar{x})$, przy czym:

$$\bar{N} = \frac{N(x, t)}{N_0}$$

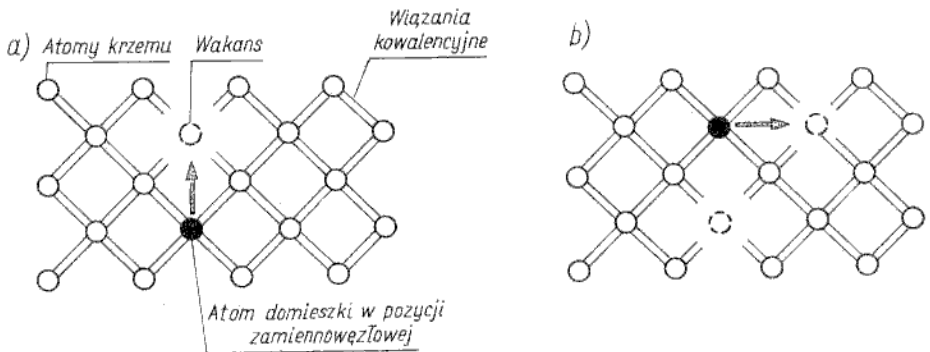
$$\bar{x} = \frac{x}{2\sqrt{Dt}}$$

Wykresy $\bar{N}(\bar{x})$ mogą służyć jako nomogram do wyznaczania rozkładów koncentracji domieszki dla dowolnych wartości czasu t , jeżeli jest znany współczynnik dyfuzji D . Wartości współczynnika dyfuzji zależą od rodzaju domieszki, temperatury i właściwości fizykochemicznych podłoża półprzewodnikowego. Żeby to bliżej wyjaśnić, trzeba rozpatrzeć proces dyfuzji na poziomie zjawisk atomowych. Należy przede wszystkim wyjaśnić dwa mechanizmy dyfuzji: zmiennowęzłowy i międzywęzłowy.

Dyfuzja zmiennowęzłowa

2.2.3.3

Dyfuzja zmiennowęzłowa polega na tym, że atom domieszki zajmuje pozycję w węzle sieci krystalicznej półprzewodnika (na przykład krzemu) i może ją zmienić tylko wówczas, gdy w jego sąsiedztwie pojawi się *wakans*, czyli nie zajęty węzeł sieci krystalicznej (rys. 2.12). Zatem dla przeskoku atomu domieszki z węzła do węzła muszą być spełnione dwa warunki:



Rys. 2.12
Model atomowy dyfuzji zamiennowęzłowej: a), b) kolejne fazy transportu atomu domieszki

- zerwanie wiązań kowalencyjnych atomu domieszki z sąsiadującymi atomami krzemu;
 - brak atomu krzemu w jednym z sąsiednich węzłów.
- Oba te zdarzenia są powodowane drganiami cieplnymi atomów w węzłach sieci krystalicznej, prawdopodobieństwo ich wystąpienia można zatem wyrazić zgodnie ze statystyką Boltzmannna:

$$f_1 \sim \exp \frac{-W_1}{kT}$$

$$f_2 \sim \exp \frac{-W_2}{kT}$$

przy czym: W_1 — energia wiązania kowalencyjnego atomu domieszki z sąsiednimi atomami krzemu; W_2 — energia wiązania kowalencyjnego atomu krzemu z sąsiednimi atomami krzemu; jest to inaczej energia kreacji wakansu.

Prawdopodobieństwo jednoczesnego wystąpienia obu zdarzeń (są to zdarzenia niezależne), czyli prawdopodobieństwo przeskoku atomu domieszki z węzła do węzła:

$$f = f_1 f_2$$

$$f \sim \exp \frac{-W_{az}}{kT}$$

przy czym $W_{az} = W_1 + W_2$ — energia aktywacji (indeks *az* oznacza aktywację przy dyfuzji zamiennowęzłowej).

Ponieważ liczba prób dokonania przeskoku w 1 s jest równa częstotliwości drgań atomu ν_0 , zatem częstotliwość prób udanych, czyli częstotliwość przeskoków

$$\nu = \nu_0 \exp \frac{-W_{az}}{kT}$$

Energia aktywacji W_{az} najczęściej przyjmuje wartości 2...3 eV dla germanu oraz 3...4 eV dla krzemu, a częstotliwość ν_0 wynosi ok. 10^{13} Hz. Przykładowo dla $W_{az} = 3$ eV częstotliwość przeskoków w temperaturze 1200°C (1473 K)

$$\nu = 10^{13} e^{-24} \approx 30 \frac{1}{s}$$

a w temperaturze pokojowej (300 K)

$$\nu = 10^{13} e^{-120} \approx 10^{-38} \frac{1}{s}$$

czyli 1 przeskok na $3 \cdot 10^{31}$ lat!!!

Długi to okres czasu, jeśli porównać choćby z wiekiem Ziemi, który szacuje się na $5 \cdot 10^9$ lat.

Ten przykład wyjaśnia, dlaczego proces dyfuzji trzeba prowadzić w bardzo wysokich temperaturach, a po obniżeniu temperatury można uważać, że rozkład koncentracji domieszek jest zamrożony.

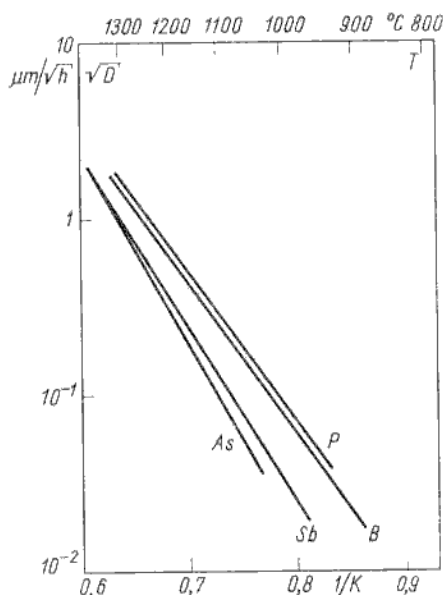
Ponieważ współczynnik dyfuzji jest proporcjonalny do częstotliwości przeskoków

$$D \sim \nu$$

więc

$$D = D_0 \exp \frac{-W_{az}}{kT} \quad (2.10)$$

przy czym parametr D_0 , nazywany niekiedy *dyfuzyjnością*, jest równy ekstrapolowanej wartości współczynnika D przy $T \rightarrow \infty$.



Rys. 2.13

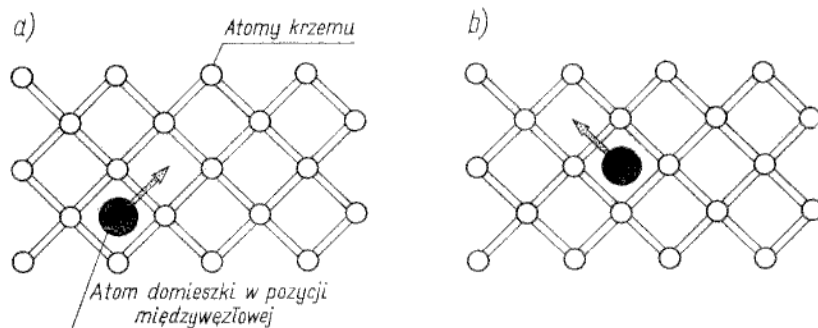
Wykresy $\ln \sqrt{D} = f\left(\frac{10^3}{T}\right)$ dla boru, fosforu, arsenu i antymonu

Wszystkie domieszki akceptorowe i donorowe dyfundują zgodnie z mechanizmem zamiennowęzłowym. Na rysunku 2.13 przedstawiono wykresy $\ln \sqrt{D} = f\left(\frac{10^3}{T}\right)$ dla czterech najczęściej stosowanych domieszek, tj. dla boru (domieszka akceptorowa), fosforu, arsenu i antymonu (domieszki donorowe).

Dyfuzja międzywęzłowa

2.2.3.4

Dyfuzja międzywęzłowa polega na tym, że atom domieszki lokuje się w swobodnej przestrzeni między atomami krzemu, zajmującymi prawidłowe pozycje w węzłach sieci krystalicznej i może zmieniać pozycję w kryształ, przeciskając się między atomami krzemu (rys. 2.14). W tym przypadku wszystkie sąsiadu-



Rys. 2.14

Model atomowy dyfuzji międzywęzłowej: a), b) kolejne fazy transportu atomu domieszki

jące pozycje międzywęzłowe są, praktycznie biorąc, wolne, czyli prawdopodobieństwo zmiany pozycji zależy tylko od energii potrzebnej do „prześciśnięcia się” między atomami krzemu

$$f \sim \exp \frac{-W_{am}}{kT}$$

a więc

$$D = D_0 \exp \frac{-W_{am}}{kT}$$

przy czym W_{am} — energia aktywacji (indeks am oznacza aktywację dla dyfuzji międzywęzłowej).

Ponieważ $W_{am} < W_{az}$, zatem współczynnik dyfuzji dla mechanizmu międzywęzłowego jest znacznie większy niż dla mechanizmu zamiennowęzłowego (oczywiście przy jednakowej temperaturze). Ten rodzaj dyfuzji dotyczy ciężkich atomów metalicznych (złoto, nikiel), które ze względu na duże rozmiary nie mogą zamieniać atomów krzemu w węzłach sieci krystalicznej. Takie domieszki lokują się energetycznie w pobliżu środka pasma zabronionego i spełniają funkcję centrów generacyjno-rekombinacyjnych.

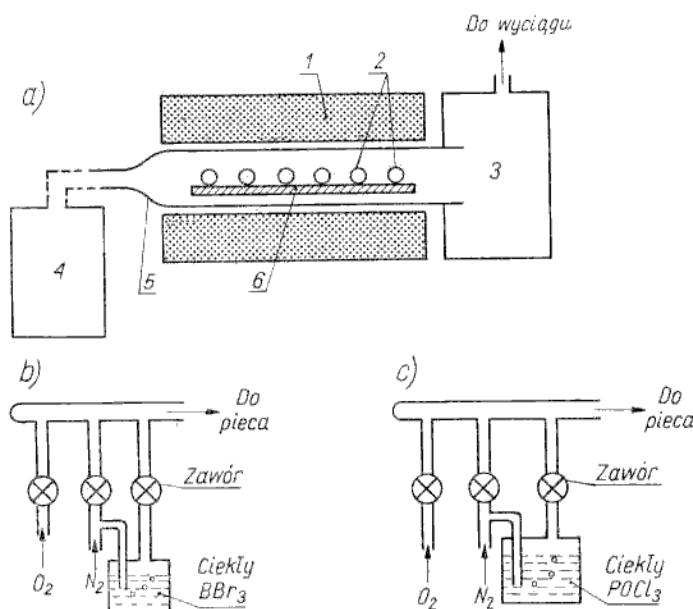
Techniczna realizacja procesu dyfuzji

2.2.3.5

Proces dyfuzji domieszki do podłoża półprzewodnikowego realizuje się w praktyce w urządzeniu nazywanym *piecem dyfuzyjnym*, przedstawionym schematycznie na rys. 2.15.

Domieszkowanie akceptorami

Spośród kilku możliwych domieszek akceptorowych, takich jak gal, aluminium, ind lub bor, do domieszkowania krzemu stosuje się tylko bor, gdyż pozostałe



Rys. 2.15

Schemat układu do dyfuzji: a) piec dyfuzyjny; b) dozownik mieszaniny z domieszką boru; c) dozownik mieszaniny z domieszką fosforu
 1 — piec dyfuzyjny nagrzewany indukcyjnie lub oporowo; 2 — płytki krzemowe; 3 — komora wyciągu; 4 — dozownik mieszaniny domieszkującej (rys. b) lub c); 5 — rura kwarcowa; 6 — kasetka kwarcowa

rodzaje domieszek mają istotne wady. Gal ma duży współczynnik dyfuzji w dwutlenku krzemu, co wyklucza możliwość prowadzenia dyfuzji lokalnej, omawianej w p. 2.3. Aluminium łatwo reaguje z tlenem występującym zwykle w pewnych ilościach w dwutlenku krzemu oraz, tak jak gal, z łatwością przenika przez warstwę dwutlenku krzemu. Ind tworzy poziom akceptorowy zbyt odległy od krawędzi pasma zabronionego krzemu (0,16 eV w porównaniu z 0,045 eV dla boru). Bor jest domieszką akceptorową o najlepszych właściwościach fizykochemicznych; przede wszystkim ma najmniejszą wartość energii jonizacji, największą rozpuszczalność graniczną w krzemie ($5 \cdot 10^{26} \text{ m}^{-3}$) oraz małą wartość współczynnika dyfuzji w dwutlenku krzemu. W celu przeprowadzenia dyfuzji boru do krzemu płytki krzemowe wygrzewa się (temperatura 1000...1200°C) w atmosferze zawierającej pary B₂O₃. Wówczas zachodzi reakcja



W wyniku tej reakcji na powierzchni Si powstaje tzw. szkliwo borokrzemowe (stop SiO₂ · B₂O₃) i uwalnia się bor, który dyfunduje w głąb płytki krzemowej. Jako pierwotne źródło domieszki najczęściej stosuje się ciekły BBr₃, który reaguje z tlenem



a jednym z produktów reakcji jest B₂O₃.

Jak można się zorientować z rys. 2.15b, azot jest nośnikiem gazowym par BBr_3 , które w piecu dyfuzyjnym reagują z tlenem (2.12), a produkt tej reakcji B_2O_3 łączy się z krzemem (2.11) dając dwutlenek krzemu silnie zdomieszkowany borem. Dyfuzję domieszek wykonuje się na ogół dwustopniowo. Wyżej przedstawiony proces stanowi pierwszą fazę — tzw. *predyfuzję*, którą można opisać modelem teoretycznym dyfuzji ze źródła nieograniczonego. Po predyfuzji następuje druga faza nazywana *redyfuzją*, którą można opisać modelem teoretycznym dyfuzji ze źródła ograniczonego, gdyż w tej fazie nie ma dopływu domieszki (do pieca wpuszcza się tylko gaz obojętny lub tlen). Wówczas warstwa szkliwa boro-krzemowego jest źródłem domieszki o skończonej wydajności. Jeżeli redyfuzja jest prowadzona w atmosferze tlenu, to jednocześnie powstaje warstwa dwutlenku krzemu.

Domieszkowanie donorami

Jako domieszki donorowe w krzemie najczęściej stosuje się fosfor, rzadziej arsen lub antymon. Wymienione trzy rodzaje domieszek mają w przybliżeniu jednakowe wartości energii jonizacji (ok. 0,04 eV), jednak fosfor jest częściej stosowany ze względu na większą wartość współczynnika dyfuzji w krzemie przy jednocześnie zadowalająco wolnej dyfuzji w dwutlenku krzemu. Arsen lub antymon stosuje się w tych przypadkach, gdy wykonanie złożonej struktury półprzewodnikowej, na przykład tranzystora w układzie scalonym, wymaga wielokrotnej dyfuzji i ważne jest, by domieszka wdyfundowana na początku nie zmieniła swego rozkładu koncentracji w następujących potem obróbkach wysokotemperaturowych. Dlatego powinna to być domieszka o małym współczynniku dyfuzji. Arsen i antymon spełniają ten warunek, przy czym wadą arsenu jest jego duża toksyczność, niedostatkami antymonu — mała rozpuszczalność graniczna w krzemie ($5 \cdot 10^{23} \text{ m}^{-3}$ w porównaniu z $1,5 \cdot 10^{27} \text{ m}^{-3}$ dla arsenu). Realizacja techniczna procesu dyfuzji fosforu do krzemu jest podobna jak dla boru. Płytki krzemowe, umieszczone w rurze kwarcowej pieca dyfuzyjnego, wygrzewa się (temperatura 1000...1200°C) w atmosferze zawierającej P_2O_5 . Wówczas zachodzi reakcja



W rezultacie na powierzchni Si powstaje tzw. szkliwo fosforo-krzemowe (stop $SiO_2 \cdot P_2O_5$) i uwalnia się fosfor, który dyfunduje w głąb płytki półprzewodnikowej. Jako pierwotne źródło domieszki najczęściej stosuje się ciekły $POCl_3$, który reaguje z tlenem



a jednym z produktów reakcji jest P_2O_5 .

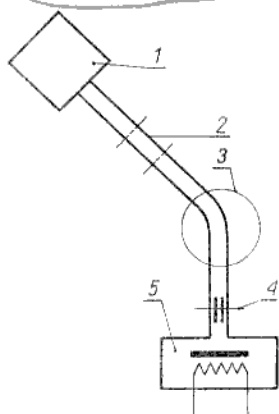
Jak widać z rys. 2.15c, azot jest nośnikiem gazowym par $POCl_3$, które w piecu dyfuzyjnym reagują z tlenem (2.14), a produkt tej reakcji P_2O_5 łączy się z krzemem (2.13) dając dwutlenek krzemu zdomieszkowany fosforem. Tak jak w przypadku boru, dyfuzję fosforu często prowadzi się dwustopniowo (predyfuzja i redyfuzja).

Implantacja jonów

2.2.4

Obok dyfuzji częściowo alternatywną, a częściowo dopełniającą metodą wprowadzania domieszek do półprzewodnika jest implantacja, czyli wszczepianie jonów. Tę metodę — opatentowaną w 1954 r. przez Shockleya — zaczęto stoso-

wać w mikroelektronice w latach 1964/1965. We współczesnej technologii półprzewodnikowej implantacja jonów jest jednym z podstawowych procesów technologicznych, stosowanym niekiedy zamiast dyfuzji, lecz częściej jako proces uzupełniający. Implantacja jonów polega niejako na wbijaniu jonów do kryształu wskutek bombardowania powierzchni płytki półprzewodnikowej jonami rozpędzonymi w silnym polu elektrycznym. Na rys. 2.16 pokazano schematycznie



Rys. 2.16

Schemat implantatora

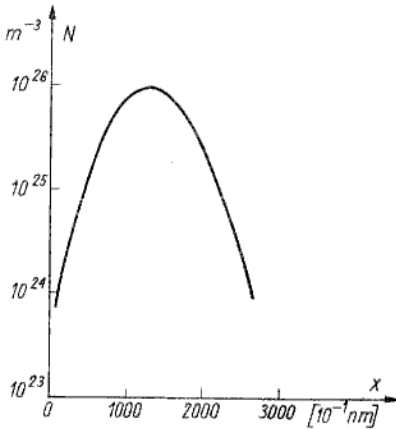
1 — źródło jonów; 2 — układ przyspieszania jonów; 3 — analizator masy (pole magnetyczne); 4 — układ odchylenia; 5 — komora z podgrzewanym podłożem

urządzenie nazywane *implantatorem*, w którym realizuje się ten proces. Implantator składa się z trzech podstawowych części: komory źródła jonów, analizatora masy, komory podłożowej.

W komorze źródła jonów powstają jony domieszki, na przykład metodą rozpylania, tj. wtórnej emisji jonów z ciała stałego pod wpływem bombardujących jonów argonu. Jony domieszki są następnie rozpędzane w polu elektrycznym do bardzo dużych prędkości (przeciętnie uzyskiwane energie kinetyczne jonów wynoszą 10...200 keV). W analizatorze masy wiązka jonów zagina się w polu magnetycznym pod kątem zależnym od stosunku ładunku do masy jonu. Podobnie jak w pryzmacie, wiązka światła białego rozszczepia się na promienie o różnych barwach, tu następuje takie rozdzielenie trajektorii jonów, by w określonym kierunku podążały tylko jony pożądanej domieszki. Po przejściu przez układ ogniskowania i odchylenia jony wpadają do komory, w której znajduje się płytka podłożowa, podgrzewana zwykle do temperatury kilkudziesięciu stopni Celsjusza i bombardując jej powierzchnię wchodzi do kryształu półprzewodnika na określoną głębokość. Wskutek bombardowania jonami w przypowierzchniowej warstwie półprzewodnika powstaje tak dużo defektów, że warstwa implantowana nie ma uporządkowanej struktury krystalicznej; uważa się, że jest to warstwa amorficzna. Dlatego końcową fazą procesu implantacji jest wygrzewanie półprzewodnika zwykle w temperaturze 500...900°C w czasie od kilkunastu minut do kilku godzin, mające na celu uporządkowanie struktury krystalicznej i umożliwienie jonom domieszki zajęcia pozycji zamiennowęzłowych. Tylko w tych pozycjach atomy domieszki są aktywne elektrycznie, czyli wykazują właściwości akceptorowe lub donorowe.

Grubość warstwy implantowanej i koncentracja atomów domieszki zależą od rodzaju jonów, ich energii, dawki i właściwości podłoża półprzewodnikowego. Przy energiach do 200 keV uzyskuje się najczęściej warstwy o grubości do 1 μm. Rozkład

koncentracji implantowanych jonów ma kształt rozkładu Gaussa (rys. 2.17). W przeciwieństwie zatem do dyfuzji, która zawsze daje profil domieszkowania malejący od powierzchni w głąb materiału, w tym przypadku można otrzymać, praktycznie biorąc, dowolny profil koncentracji domieszek. Na przykład zmie-

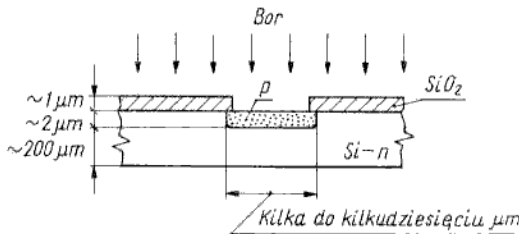


Rys. 2.17
Przykładowy rozkład koncentracji akceptorów w warstwie implantowanej jonami boru o energii 40 keV i dawce 10^{19} m^{-2}

niając odpowiednio energię jonów w czasie prowadzenia procesu implantacji można otrzymać rozkład koncentracji równomierny z gwałtownym spadkiem na określonej głębokości (superpozycja rozkładów Gaussa). Implantacja jonów nie wymaga stosowania wysokich temperatur¹⁾, co jest istotną zaletą tego procesu. Inne ważne aspekty²⁾ stosowania implantacji jonów, związane z możliwością realizacji domieszkowania lokalnego, będą omówione nieco dalej.

Kształtowanie lokalne właściwości fizykochemicznych półprzewodnika (domieszkowanie lokalne) 2.3

W poprzednim punkcie przedstawiliśmy trzy podstawowe metody domieszkowania półprzewodnika, przy czym nie ograniczaliśmy powierzchni poddawanej tym obróbkom. Jest oczywiste, że w typowej płytce półprzewodnikowej o średnicy 5 cm można wykonać wiele tysięcy struktur półprzewodnikowych (diod, tranzystorów). W tym celu trzeba dysponować metodami precyzyjnej kontroli rozmiarów geometrycznych obszarów półprzewodnika, poddawanych procesowi domieszkowania. Dlatego teraz wyjaśnimy, w jaki sposób można stosować



Rys. 2.18
Ilustracja idei domieszkowania lokalnego podłoża krzemowego na przykładzie dyfuzji lokalnej

¹⁾ W obróbce technologicznej krzemu za wysokie temperatury uważa się zakres $T > 900^\circ\text{C}$.

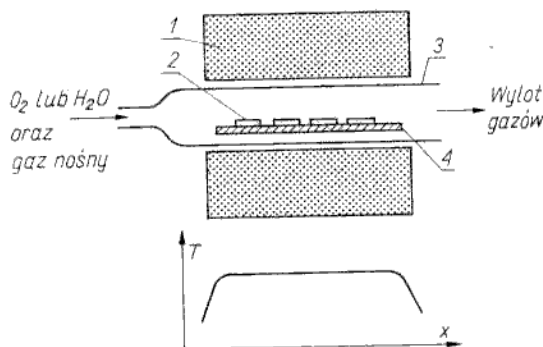
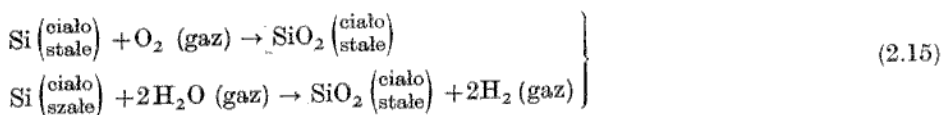
epitaksję, dyfuzję oraz implantację do kształtowania lokalnego właściwości fizykochemicznych półprzewodnika, tj. jak wykonuje się domieszkowanie w precyzyjnie ograniczonych obszarach płytki podłożowej (*domieszkowanie lokalne*). Ideę takiego procesu technologicznego ilustruje rys. 2.18 na przykładzie dyfuzji lokalnej boru do podłoża krzemowego typu *n*. Płytkę krzemową jest pokryta warstwą SiO_2 , w której wykonano otwór (tzw. okno). Atomy domieszki nie są w stanie przeniknąć przez warstwę SiO_2 , gdyż mają bardzo mały współczynnik dyfuzji w dwutlenku krzemu. Warstwa SiO_2 spełnia więc funkcję maski nie przepuszczającej atomów domieszki. W związku z tym krzem jest domieszkowany tylko w obszarze okna. W domieszkowaniu lokalnym pojawiły się dwa nie wyjaśnione dotychczas procesy technologiczne, tj.:

- wytwarzanie warstwy maskującej (maskowanie);
 - wytwarzanie okien w warstwie maskującej (fotolitografia).
- Rozpatrzmy te dwa zagadnienia.

Maskowanie

2.3.1

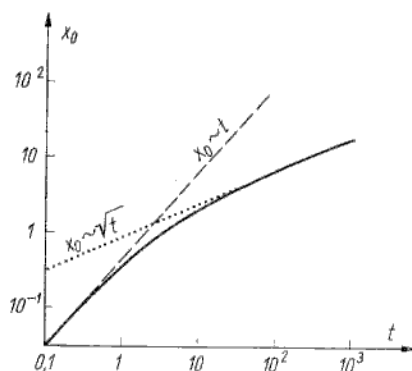
W przykładzie dyfuzji lokalnej, przedstawionym na rys. 2.18, funkcję warstwy maskującej spełnia warstwa dwutlenku krzemu. Jest to w istocie najczęściej stosowany rodzaj warstwy maskującej, choć duże znaczenie praktyczne ma również azotek krzemu, Si_3N_4 . Zajmijmy się najpierw dwutlenkiem krzemu. Najczęściej stosowanym sposobem wytwarzania SiO_2 na powierzchni Si jest utlenianie cieplne płytki krzemowej w atmosferze tlenu, pary wodnej lub tlenu wilgotnego, w temperaturze 900...1200°C. Przebieg odpowiednich reakcji można przedstawić w następujący sposób:



Rys. 2.19
Schemat urządzenia do utleniania cieplnego i wykres rozkładu temperatury
1 — piec grzany oporowo; 2 — płytki Si; 3 — rura kwarcowa; 4 — kasetka kwarcowa

Urządzenie do utleniania (rys. 2.19) stanowi rura kwarcowa, umieszczona w piecu oporowym, do której doprowadza się gaz utleniający. Płytki krzemowe znajdują się w kasecie kwarcowej, umieszczonej w strefie stałej temperatury. Szybkość utleniania w atmosferze tlenu wilgotnego jest kilkakrotnie większa niż w atmosferze tlenu suchego. Kinetyka procesu utleniania jest dość złożona. W początkowej fazie utleniania grubość warstwy wzrasta liniowo w funkcji czasu (szybkość

utleniania jest ograniczona reakcją tlenu z krzemem), po czym w drugiej fazie grubość warstwy wzrasta wolniej według funkcji parabolicznej, tj. proporcjonalnie do pierwiastka kwadratowego z czasu (szybkość utleniania jest ograniczona procesem dyfuzji tlenu przez warstwę SiO_2) — rys. 2.20. Narastanie



Rys. 2.20
Zależność grubości warstwy dwutlenku krzemu x_0 od czasu utleniania t (w dowolnych jednostkach)

warstwy SiO_2 następuje częściowo w głąb płytki podłożowej. Przykładowo warstwa SiO_2 o grubości $1 \mu\text{m}$ wzrasta na głębokość $0,45 \mu\text{m}$ w głąb płytki krzemowej. Jest to bardzo istotny fakt, gdyż w przypadku utleniania lokalnego (w ograniczonym obszarze płytki) i następującego po tym strawienia warstwy SiO_2 powstaje wgłębienie w płytce podłożowej.

Wprowadzie warstwę SiO_2 omawiamy w aspekcie jej właściwości maskujących, jednak warto zwrócić uwagę na inne funkcje spełniane przez warstwy dwutlenku krzemu. Obok maskowania drugą podstawową funkcją warstwy SiO_2 jest pasywowanie powierzchni krzemu. Pasywowanie rozumie się jako ochronę krzemu od zanieczyszczeń i innych wpływów czynników zewnętrznych oraz radykalne zmniejszenie liczby stanów powierzchniowych, a więc poprawę parametrów elektrofizycznych powierzchni Si (por. p. 3.3.3 o strukturze MOS). Należy również pamiętać, że SiO_2 jest doskonałym dielektrykiem, dzięki czemu warstwa dwutlenku krzemu jest stosowana jako warstwa dielektryczna w kondensatorach i tranzystorach MIS. O ile do maskowania i pasywacji stosuje się warstwy SiO_2 o grubości $0,5 \dots 2 \mu\text{m}$ o niezbyt wygórowanych wymaganiach odnośnie do parametrów elektrycznych, o tyle w kondensatorach i tranzystorach MIS są stosowane warstwy SiO_2 o grubości $100 \dots 150 \text{ nm}$ o bardzo dobrych parametrach elektrycznych.

Na koniec należy wspomnieć, że warstwy dwutlenku krzemu mogą być również wytwarzane metodą nanoszenia na powierzchni płytki krzemowej. W tym celu najczęściej stosuje się pirolityczny rozkład związków krzemu z tlenem w temperaturze ok. 800°C . Zaletami tej metody są dość niska temperatura procesu pirolitycznego, przy której nie występuje niepożądana dyfuzja, duża i stała szybkość narastania warstwy (ok. $1 \mu\text{m}$ na 30 min) oraz brak „wohłaniania” podłoża przez narastającą warstwę.

Niekiedy istnieje uzasadniona potrzeba zastosowania warstwy maskującej z bardziej „gęstego” materiału niż SiO_2 . Takie właściwości ma azotek krzemu, który może skutecznie maskować powierzchnię krzemu od wielu domieszek lub zanieczyszczeń swobodnie przenikających przez warstwę dwutlenku krzemu. Między innymi przez warstwę Si_3N_4 nie przechodzą jony Na^+ , H^+ , mające szkodliwy wpływ

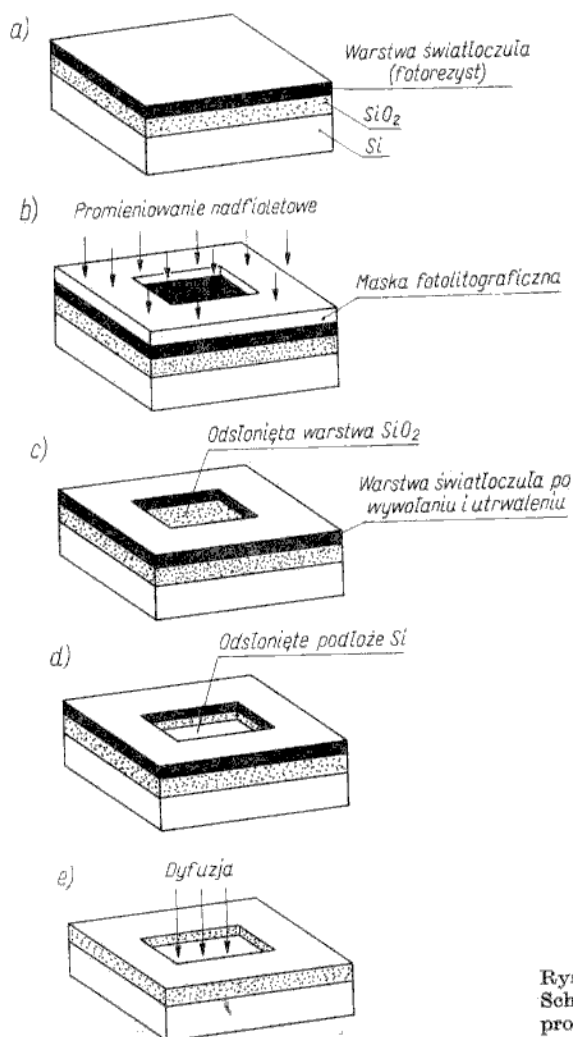
na stabilność przyrządów półprzewodnikowych. Warstwy azotku krzemu są wytwarzane najczęściej metodami chemicznego osadzania z fazy lotnej lub rozpylania katodowego.

Przeszkodą w szerszym stosowaniu warstw azotku krzemu są ich wady: gorsze właściwości elektrofizyczne obszaru granicznego $\text{Si-Si}_3\text{N}_4$ niż w układzie Si-SiO_2 oraz trudności związane z obróbką fotochemiczną tych warstw.

Fotolitografia

2.3.2

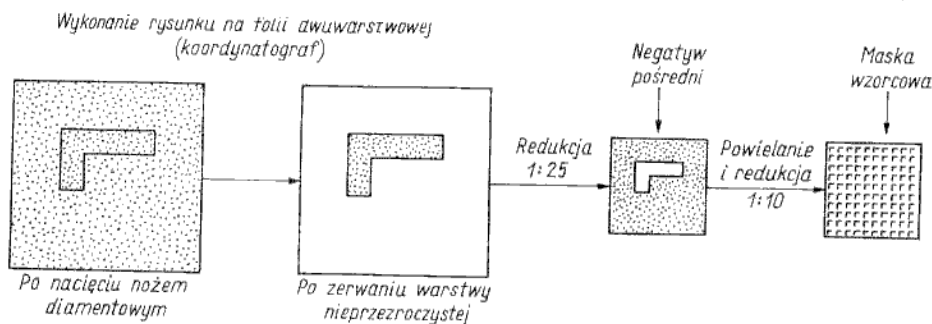
Okna w warstwie maskującej wykonuje się metodą fotolitografii. Polega ona na zastosowaniu warstwy światłoczułej (tzw. fotorezystu) do maskowania i lokalnego wytrawiania odsłoniętych warstw. W omawianym przypadku chodzi o wytrawianie warstwy SiO_2 lub Si_3N_4 , lecz fotolitografia w ogólnym sensie może



Rys. 2.21
Schematyczna ilustracja kolejnych faz procesu fotolitografii

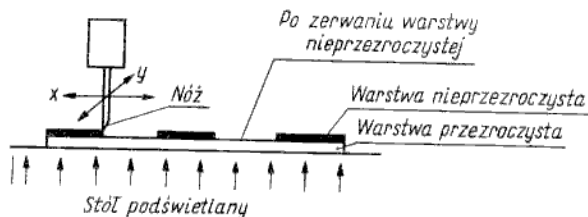
również dotyczyć trawienia lokalnego warstwy metalizacji lub innych warstw. Na rys. 2.21 przedstawiono schematycznie ważniejsze fazy procesu wytwarzania okien w warstwie SiO_2 metodą fotolitografii. Najpierw nakłada się ciekłą emulsję światłoczułą na powierzchni utlenionej płytki podłożowej (rys. 2.21a). Bardzo ważne jest uzyskanie równomiernej warstwy emulsji, dlatego płytka jest przyciśnięta przyssawką próżniową do stolika wykonującego szybki ruch obrotowy (ok. 20 000 obr/min). Po wysuszeniu emulsji w temperaturze ok. 150°C (odparowanie rozpuszczalnika) powierzchnia płytki jest naświetlana przez specjalną maskę kontaktową (tzw. fotomaskę) promieniowaniem nadfioletowym (rys. 2.21b). Do tej chwili płytka z warstwą światłoczułą przebywała w ochronnym świetle żółtym. Naświetlenie ma na celu spowodowanie odpowiednich procesów fotochemicznych w warstwie światłoczułej, tj. polimeryzację powodującą uodpornienie warstwy na działanie roztworu wywołującego. Ponieważ płytka jest naświetlana przez fotomaskę osłaniającą określone obszary powierzchni, zatem po wywołaniu i utrwaleniu warstwa światłoczuła tworzy obraz negatywu fotomaski, czyli osłania tylko tę część powierzchni płytki, która była naświetlona¹⁾. Tę sytuację ilustruje rys. 2.21c, na którym pokazano okno w wywołanej i utrwalonej warstwie emulsji. W następnej fazie płytka jest poddawana działaniu kwasu fluorowodorowego w celu wytrawienia okna w odsłoniętej części warstwy SiO_2 (rys. 2.21d). Potem usuwa się warstwę emulsji (rys. 2.21e), działając w podwyższonej temperaturze takimi odczynnikami nieorganicznymi, jak np. kwas siarkowy lub azotowy. W ten sposób otrzymuje się strukturę gotową do prowadzenia domieszkowania lokalnego.

Bardzo ważną operacją technologiczną jest przygotowanie odpowiednich masek, które z dużą precyzją (z dokładnością ok. $1\ \mu\text{m}$) wyznaczają obrysy obszarów domieszkowanych. Na rysunku 2.22 przedstawiono schematycznie typowy proces wytwarzania masek fotolitograficznych, obejmujący wykonanie rysunku, negatywu



Rys. 2.22

Schemat metody wytwarzania masek fotolitograficznych



Rys. 2.23

Ilustracja działania koordynatografu — wycinanie rysunku w folii dwuwarstwowej

¹⁾ Są również stosowane emulsje światłoczułe o właściwościach pozytywnych.

pośredniego, maski wzorcowej oraz maski roboczej. Rysunek wykonuje się za pomocą urządzenia, nazywanego *koordynatografem*, który najczęściej jest sterowany przez komputer (rys. 2.23).

Obrysy pożądanej figury są wycinane nożem diamentowym w folii dwuwarstwowej, w której warstwa górna jest nieprzezroczysta, a dolna — przezroczysta. Następnie w odpowiednich miejscach warstwa nieprzezroczysta jest zrywana (rys. 2.22). Rysunek jest wykonywany najczęściej w skali 250:1 i zajmuje pole o powierzchni ok. 1 m².

W następnym etapie wykonuje się negatyw pośredni, zmniejszając rysunek zwykle 25-krotnie. Na podstawie tego negatywu wykonuje się maskę wzorcową za pomocą kamery powielającej (ang. *step and repeat camera*), której zadaniem jest zmniejszenie (w typowym przypadku 10-krotne) i powielanie obrazu. Zatem maska wzorcowa o takiej samej powierzchni jak negatyw pośredni zawiera 100 powtarzalnych obrazów negatywu.

Maska fotolitograficzna jest płytką szklaną (zwykle o rozmiarach 5 × 5 cm) z utrwaloną warstwą nieprzezroczystego materiału światłoczułego tworzącego określone figury geometryczne. Ponieważ maska przy wielokrotnym użyciu niszczy się, na podstawie maski wzorcowej wykonuje się maski robocze metodą stykowego przekopiowania.

Przedstawiona w ogólnym zarysie metoda wytwarzania masek fotolitograficznych jest standardowa. Są również znane inne metody, przy czym największe perspektywy rozwoju ma metoda rysowania obrazu wiązką elektronów lub promieniem lasera bezpośrednio na masce wzorcowej (w tym przypadku jest to płytka szklana, pokryta warstwą chromu i specjalną warstwą kopiową) z pominięciem wykonywania dużego rysunku, jego wielokrotnej redukcji i powielania.

Najczęściej domieszkowanie lokalne wykonuje się kilkakrotnie w płycie podłożowej, co wymaga stosowania kilku kolejnych procesów fotolitografii, przy czym następujące procesy fotolitografii muszą dawać obrysy dokładnie scentrowane z poprzednimi. W tym celu na maskach fotolitograficznych znajdują się tzw. znaki ustawcze (krzyżyki itp.), które służą do dokładnego ustawienia maski względem podłoża (tzw. operacja centrowania wykonywana pod mikroskopem).

Zdolność rozdzielcza procesu fotolitografii, uwzględniając efekty podtrawiania tlenku i niedokładność centrowania, wynosi w typowym przypadku $\pm 2 \mu\text{m}$.

Epitaksja lokalna

2.3.3

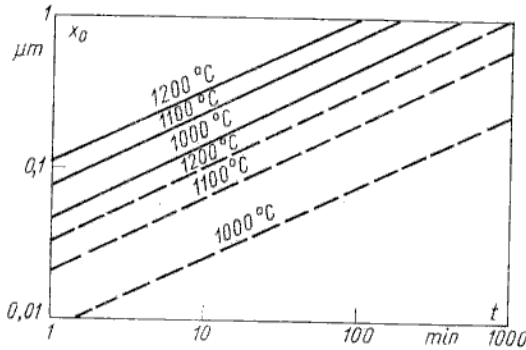
Płytkę krzemową z warstwą SiO₂, w której wykonano okna metodą fotolitografii, można poddać procesowi epitaksji. Wówczas warstwa krzemu monokrystalicznego narasta tylko w obszarze okna, tj. na odsłoniętej powierzchni Si. Taki proces nazywa się *epitaksją lokalną*. Jednym z niepożądanych zjawisk ograniczających możliwości stosowania tej metody jest powstawanie krystalitów krzemu na powierzchni dwutlenku krzemu.

Dyfuzja lokalna

2.3.4

Warstwa dwutlenku krzemu lub azotku krzemu stanowi skuteczną barierę uniemożliwiającą przenikanie domieszek do pokrytej tą warstwą powierzchni półprzewodnika. Dlatego w procesie ilustrowanym na rys. 2.18 domieszki dyfundują tylko w obszarze okna, tj. przez odsłoniętą powierzchnię Si. Taki proces

nazywa się *dyfuzją lokalną*. Wiadomo, że nie wszystkie domieszki są maskowane przez warstwę SiO_2 (np. gal i aluminium przenikają przez tę maskę), a w przypadku boru i fosforu skuteczność maskowania zależy od grubości warstwy SiO_2 , temperatury i czasu trwania procesu dyfuzji. Na rysunku 2.24 przedstawiono



Rys. 2.24

Wykresy minimalnej grubości warstwy SiO_2 , wymaganej do skutecznego maskowania boru i fosforu, w funkcji temperatury i czasu trwania procesu dyfuzji
linie ciągłe — dla fosforu; linie przerywane — dla boru

wykresy minimalnej grubości warstwy SiO_2 , wymaganej dla skutecznego maskowania, w funkcji temperatury i czasu trwania dyfuzji dla boru i fosforu.

Istotną cechą dyfuzji lokalnej, uwidoczną na rys. 2.18, jest rozchodzenie się domieszek we wszystkich kierunkach, a więc również w kierunku poziomym pod maskującą warstwą SiO_2 (tzw. *dyfuzja boczna* lub inaczej *lateralna*). Jest to zjawisko z jednej strony korzystne, gdyż krawędź złącza *p-n* w najbardziej „newralgicznym” miejscu, tj. na powierzchni, jest przykryta przez warstwę SiO_2 , co poprawia stabilność pracy złącza. Z drugiej zaś strony dyfuzja lateralna może powodować niekorzystne skutki, o czym będzie się można przekonać przy omawianiu działania tranzystora MOS.

Implantacja lokalna

2.3.5

Cienka warstwa dwutlenku krzemu (o grubości około kilkuset nm) jest „przezroczysta” dla jonów rozprzeczonych w silnym polu elektrycznym. Ta właściwość jest zaletą metody implantacji jonów, wykorzystywaną w technologii wytwarzania tranzystorów MOS metodą samocentrowania bramki, wymaga jednak zarazem podejścia do zagadnienia implantacji lokalnej w inny sposób niż do epitaksji i dyfuzji lokalnych. Warstwa SiO_2 może spełniać funkcję maski, jeśli ma odpowiednio dużą grubość (większą niż $1 \mu\text{m}$). Jako maski z powodzeniem są stosowane warstwy aluminium, molibdenu, złota, krzemu polikrystalicznego. Są to zwykle warstwy, których obecność w strukturze jest i tak konieczna, gdyż mają spełniać funkcje ścieżek połączeniowych, bramek tranzystorów itp.

Typowy przykład zastosowania implantacji lokalnej podano przy opisie wytwarzania tranzystorów MOS.

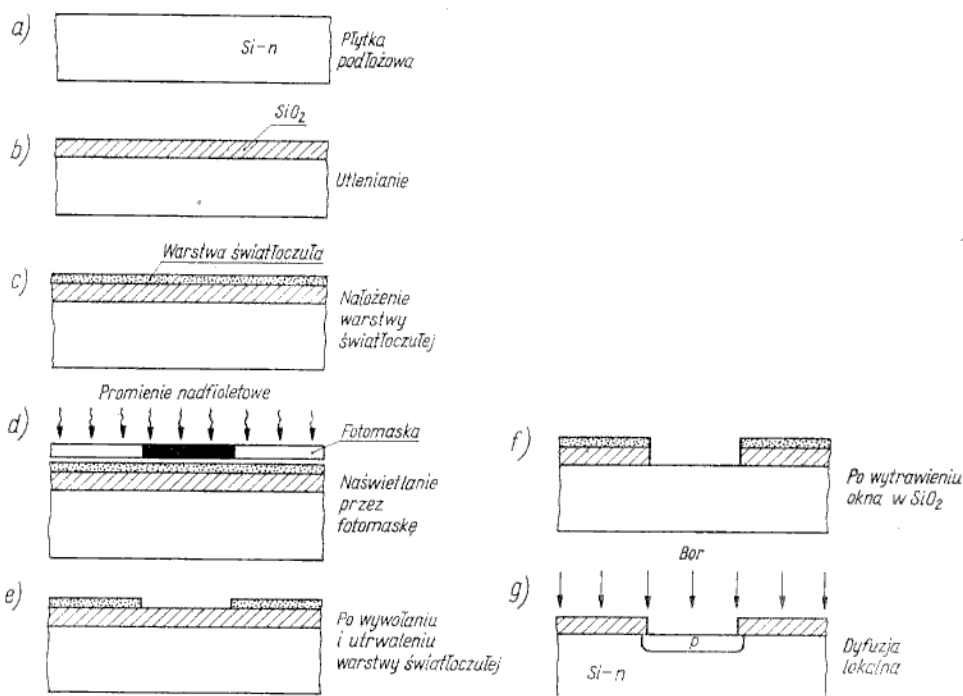
Technologia planarna

2.4

Pojęcie technologia planarna (inaczej płaszczyznowa) obejmuje swym znaczeniem zespół procesów technologicznych, w którym zawsze występują trzy podstawowe operacje:

- maskowanie podłoża krzemowego warstwą SiO_2 ;
- fotolitografia;
- domieszkowanie lokalne.

Najbardziej charakterystyczną cechą półprzewodnikowych struktur planarnych jest to, że krawędzie wszystkich złączy wychodzą na jedną płaszczyznę, którą można pasywować warstwą dielektryka. Przez tę płaszczyznę są prowadzone wszystkie procesy obróbcze. Zasadnicze fazy typowego procesu planarnego przedstawiono na rys. 2.25.



Rys. 2.25. Zasadnicze fazy procesu planarnego

Poszczególne obróbki technologiczne, składające się na proces planarny, są już znane z poprzednich punktów. Jeśli dodatkowo jest stosowana epitaksja, to mówi się o technologii epitaksjalno-planarnej lub krócej — epiplanarnej. Ponieważ epitaksja występuje bardzo często, przyjęto zatem uważać ten proces technologiczny za jeden z podstawowych składników technologii planarnej. Dlatego na ogół przez pojęcie technologia planarna rozumie się zespół procesów technologicznych, w który wchodzi również proces epitaksji.

W ostatnich latach pojęcie technologia planarna uległo jeszcze większemu rozszerzeniu obejmując swym znaczeniem również szereg metod zmodyfikowanych, jak np. technologia izoplanarna, poliplanarna, oraz wchłaniając do zespołu podstawowych operacji technologicznych nowe metody obróbcze — przede wszystkim implantację lokalną jonów¹⁾.

¹⁾ Niektóre warianty technologii planarnej, w zastosowaniu do wytwarzania układów scalonych, będą omawiane w rozdziale 8.

Swego rodzaju fenomenem w historii rozwoju technologii półprzewodnikowej jest długowieczność metody planarnej, gdyż wprowadzona w 1960 r. — wprawdzie ciągle modyfikowana — już bez mała dwadzieścia lat stanowi podstawową metodę wytwarzania struktur półprzewodnikowych (diod, tranzystorów, układów scalonych). Należy wyraźnie podkreślić, że technologia planarna dotyczy tylko jednego materiału półprzewodnikowego — krzemu. Intensywnie prowadzone próby zaadaptowania tej technologii do germanu i arsenku gału na razie nie dały pozytywnych wyników na skalę techniczną, gdyż istnieją duże trudności z wytwarzaniem odpowiednich warstw maskujących.

Inne procesy technologiczne

2.5

Dotychczas przedstawiliśmy procesy technologiczne, związane z wytwarzaniem struktur półprzewodnikowych w płytce podłożowej. Oczywiście na tym nie kończy się proces wytwarzania diod, tranzystorów czy układów scalonych, gdyż pozostaje jeszcze szereg operacji technologicznych, związanych z montażem struktur półprzewodnikowych w obudowach i wykonaniem samych obudów. Najważniejsze spośród tych operacji są:

- wykonanie metalicznych pól kontaktowych;
- testowanie struktur na płytce podłożowej;
- cięcie płytki na poszczególne struktury;
- zamocowanie struktury w odpowiedniej podstawie obudowy;
- mikromontaż, tj. połączenie pól kontaktowych cienkim drutem z wyprowadzeniami zewnętrznymi;
- wykonanie obudowy.

Te operacje technologiczne omówimy w rozdziale dotyczącym technologii wytwarzania układów scalonych.

Ważniejsze wzory

Tablica 2.1

Prawa Ficka		$F = -D \frac{dN}{dx}$ $\frac{dN}{dt} = D \frac{\partial^2 N}{\partial x^2}$
Rozkłady koncentracji domieszek przy dyfuzji	ze źródła nieograniczonego	$N(x, t) = N_0 \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$ $M(t) = \frac{2}{\sqrt{\pi}} N_0 \sqrt{Dt}$
	ze źródła ograniczonego	$N(x, t) = \frac{M}{\sqrt{\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right)$ $N_0 = \frac{M}{\sqrt{\pi Dt}}$

ZADANIA

2.1. Sprawdź, czy funkcje (2.6), (2.8) są rozwiązaniami równania transportu (2.5) oraz spełniają właściwe warunki początkowe i brzegowe.

2.2. Do płytki krzemowej typu n o rezystywności $10^{-2} \Omega \cdot \text{m}$ jest prowadzona dyfuzja boru z par B_2O_3 przy zachowaniu stałej koncentracji boru równej 10^{26} m^{-3} . Oblicz, na jakiej głębokości zrównają się koncentracje N_A i N_D , jeżeli temperatura i czas dyfuzji wynoszą $T = 1000^\circ\text{C}$, $t = 15 \text{ min}$.

Wskazówki:

a) Wyznacz N_D ze wzoru na rezystywność (lub z krzywej Irvina — rys. 1.23) przy założeniu, że $N_D = n$;

b) wyznacz \bar{x} przy $N_A = N_D$, posługując się wykresem na rys. 2.11;

c) wyznacz x biorąc wartość D z wykresu na rys. 2.13.

Odpowiedź: Około $0,25 \mu\text{m}$

2.3. Potraktuj proces technologiczny rozpatrywany w poprzednim zadaniu jako pierwszą fazę dyfuzji (predyfuzję) i wyznacz rozkład koncentracji boru po drugiej fazie dyfuzji (redyfuzji) przebiegającej w temperaturze 1100°C w czasie 1 h.

Wskazówki:

a) Wyznacz według danych z poprzedniego zadania gęstość atomów boru M po predyfuzji i przyjmij tę wartość jako warunek początkowy dla procesu dyfuzji ze źródła ograniczonego;

b) posłuż się wykresami na rys. 2.11, 2.13.

Celem tego rozdziału jest omówienie zjawisk kontaktowych i powierzchniowych w półprzewodnikach, czyli wyjaśnienie podstaw fizycznych działania półprzewodnikowych struktur dwukońcówkowych. Struktury dwukońcówkowe stanowią bezpośrednie odpowiedniki fizyczne diod oraz występują w tranzystorach i innych przyrządach półprzewodnikowych jako ich podstawowe elementy składowe. Dlatego przestudiowanie tego rozdziału jest niezbędne do zrozumienia działania diod i innych przyrządów półprzewodnikowych, omawianych w dalszych rozdziałach. W tradycyjnym układzie wyróżnia się trzy podstawowe rodzaje kontaktów (złączy):

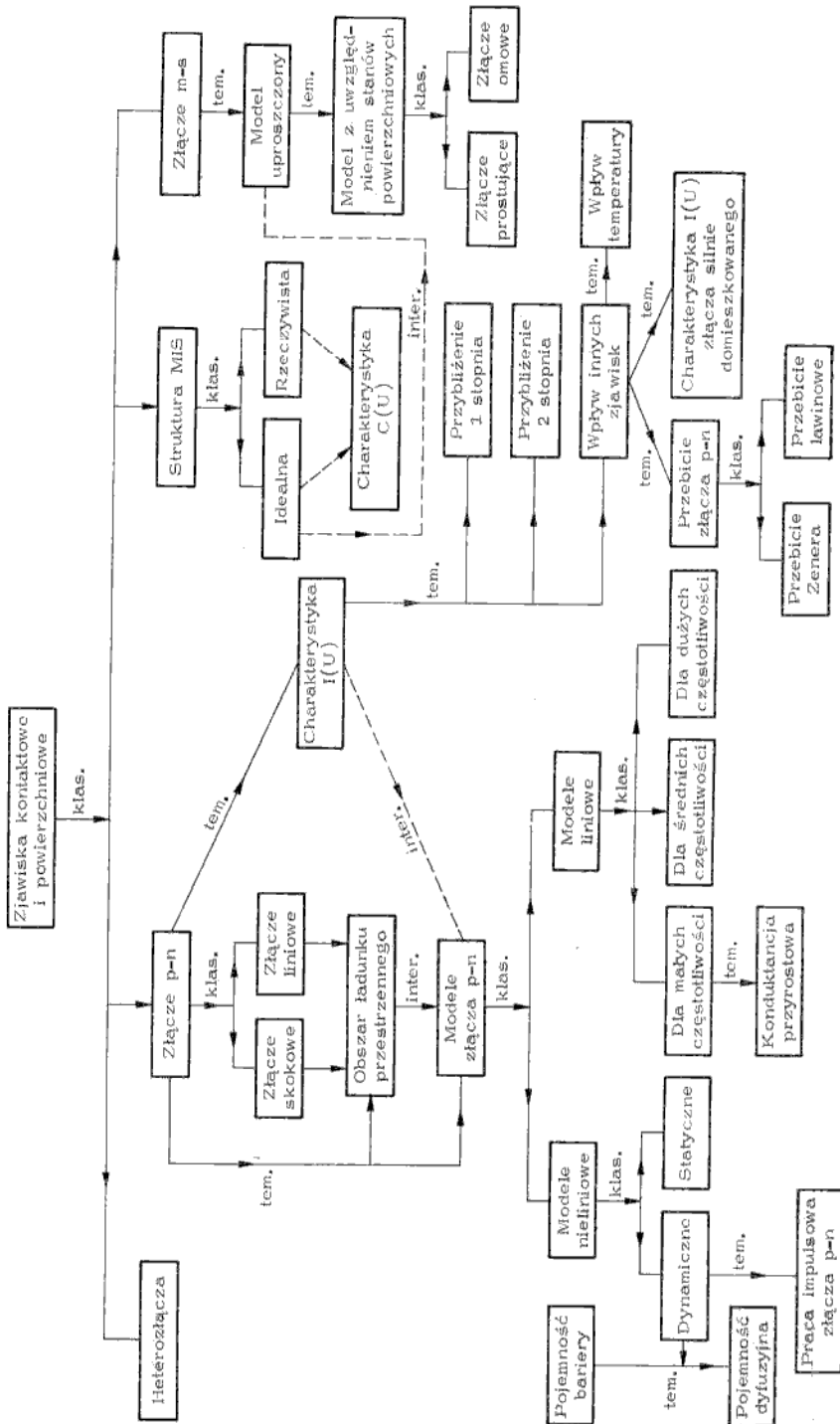
- ⊂ kontakt półprzewodnik-półprzewodnik tego samego rodzaju (np. krzem-krzem, german-german itp.), przy czym dwa obszary będące w kontakcie różnią się rodzajem domieszkania; jest to złącze $p-n$ ¹⁾;
- ⊂ kontakt dwu półprzewodników różnego rodzaju (np. german-krzem); jest to heterozłącze;
- ⊂ kontakt metal-półprzewodnik; jest to złącze $m-s$.

Ten tradycyjny układ został rozszerzony o strukturę metal-izolator-półprzewodnik (MIS). Wprawdzie w tej strukturze, w odróżnieniu od trzech wyżej wymienionych, nie jest istotna charakterystyka prądowo-napięciowa, lecz pojemnościowo-napięciowa, jednak omówienie jej działania w tym rozdziale jest celowe choćby ze względów dydaktycznych. Wyjaśnienie zjawisk zachodzących w obszarze przypowierzchniowym półprzewodnika w strukturze MIS znacznie ułatwia zrozumienie działania złącza $m-s$, traktowanego jako graniczny przypadek struktury MIS przy grubości izolatora dążącej do zera.

Najobszerniej omówiono złącze $p-n$ mające największe znaczenie we współczesnych przyrządach półprzewodnikowych, przy czym interpretację podano w układzie „koncentrycznym”, rozumianym jako stopniowe rozszerzanie uwzględnianych zjawisk i uściślanie analizy ilościowej. Najpierw zatem podaje się prostą interpretację jakościową, po czym następują aproksymacje ilościowe o coraz wyższym stopniu dokładności. Szczególnie wyeksponowano zagadnienie modeli symbolicznych (schematów zastępczych), bardzo ważne dla użytkowników przyrządów półprzewodnikowych. Należy podkreślić, że wstępne informacje o modelowaniu złącza $p-n$ — zamieszczone w p. 3.1.7 — mają charakter ogólny i odnoszą się również do innych przyrządów półprzewodnikowych.

Układ rozdziału przedstawiono schematycznie na rys. 3.1.

¹⁾ Można też wyróżnić kontakt półprzewodnik-półprzewodnik tego samego typu przewodnictwa, lecz o różnych koncentracjach domieszek ($p^+ - p$ lub $n^+ - n$). Jest to tzw. złącze $l-h$ (ang. *lightly doped region* — *heavily doped region*). Złącza $l-h$ są najczęściej stosowane dla uzyskania styku omowego półprzewodnika z metalem. Złącza te nie będą omawiane oddzielnie; o ich funkcji wspomniano w p. 3.4.4.



Rys. 3.1
Schematyczny układ rozdziału

Złącze p-n

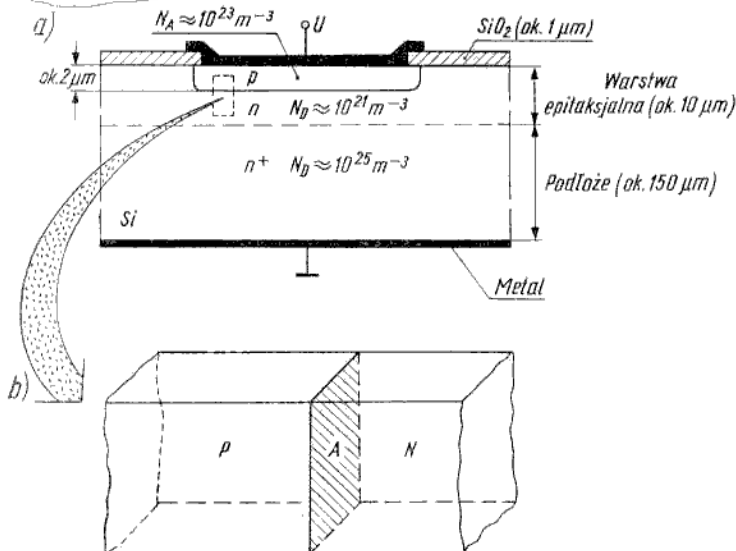
3.1

Złącze p-n, tj. bryła półprzewodnika monokrystalicznego, utworzona przez dwie graniczące ze sobą warstwy typu p i typu n¹⁾, stanowi podstawową część składową wielu elementów półprzewodnikowych (diod, tranzystorów bipolarnych i polowych). Niekiedy złącze p-n jest nazywane diodą. Ściśle biorąc, dioda jest konstrukcyjnie zakończonym elementem (z wyprowadzeniami i często z obudową), zawierającym pojedyncze złącze p-n lub inne złącze, np. metal-półprzewodnik. Technologię wytwarzania złączy p-n omówiono w poprzednim rozdziale, niektóre natomiast zagadnienia konstrukcyjne oraz właściwości funkcjonalne diod opartych na wykorzystaniu złącza p-n będą rozpatrywane w następnym rozdziale. W tym rozdziale ograniczono się do opisu właściwości elektrycznych i fizycznych złączy p-n. Najpierw będzie przedstawiony uproszczony opis jakościowy złącza, po czym dokonamy analizy ilościowej jego podstawowych charakterystyk.

Istota złącza p-n, opis jakościowy

3.1.1

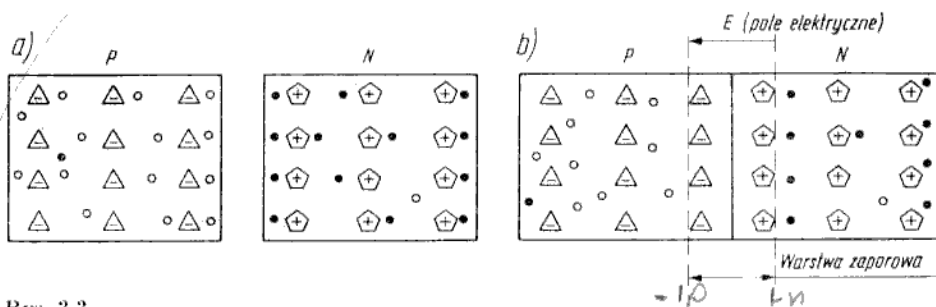
Strukturę fizyczną złącza rzeczywistego (wykonanego technologią epiplanarną) i jej wyidealizowany model jednowymiarowy pokazano na rys. 3.2. Płaszczyzna A dzieli monokryształ półprzewodnika na dwie warstwy P i N. Dla uproszczenia



Rys. 3.2 Schematyczny przekrój struktury fizycznej rzeczywistego złącza p-n (wykonanego technologią epiplanarną) (a) i jej wyidealizowany model jednowymiarowy (b)

przyjęto, że obie warstwy mają równomierny rozkład koncentracji domieszek (akceptorów po stronie P oraz donorów po stronie N). Uproszczony, płaski obraz rozkładu ładunków w złączu przedstawiono na rys. 3.3 dla dwu faz myślowego

¹⁾ Patrz odnośnik na str. 63.



Rys. 3.3

Plaski model rozkładu ładunku w dwu warstwach P , N przed ich połączeniem (a) oraz po utworzeniu złącza p - n (b)

\triangle — jon domieszki akceptorowej; \oplus — jon domieszki donorowej; \bullet — elektron; \circ — dziura

eksperymentu połączenia warstw P , N . W warstwie typu n istnieją dodatnie, nieruchome ładunki zjonizowanych atomów domieszki donorowej oraz — niemalże o takiej samej koncentracji — ujemne ładunki ruchomych elektronów — nośników większościowych. Ponadto jest jeszcze niewielka liczba dziur — nośników mniejszościowych. W warstwie typu p istnieją ujemne nieruchome ładunki zjonizowanych atomów domieszki akceptorowej oraz — niemalże o takiej samej koncentracji — dodatnie ładunki ruchomych dziur — nośników większościowych. Ponadto jest jeszcze niewielka liczba elektronów — nośników mniejszościowych. Obie warstwy przed zetknięciem (rys. 3.3a) zachowują obojętność elektryczną w dowolnym obszarze o rozmiarach makroskopowych. Po zetknięciu¹⁾ warstw N i P , wskutek dużej różnicy koncentracji ruchomych nośników ładunku, nastąpi proces dyfuzji elektronów z warstwy N do warstwy P oraz dziur z warstwy P do warstwy N . Gdyby dziury i elektrony były cząstkami obojętymi elektrycznie, doszłoby do całkowitego wyrównania się ich koncentracji w obu warstwach. Jednakże wskutek dyfuzyjnego przepływu elektronów w obszarze granicznym warstwy N pozostają nieskompensowane ładunki dodatnie nieruchomych centrów donorowych.

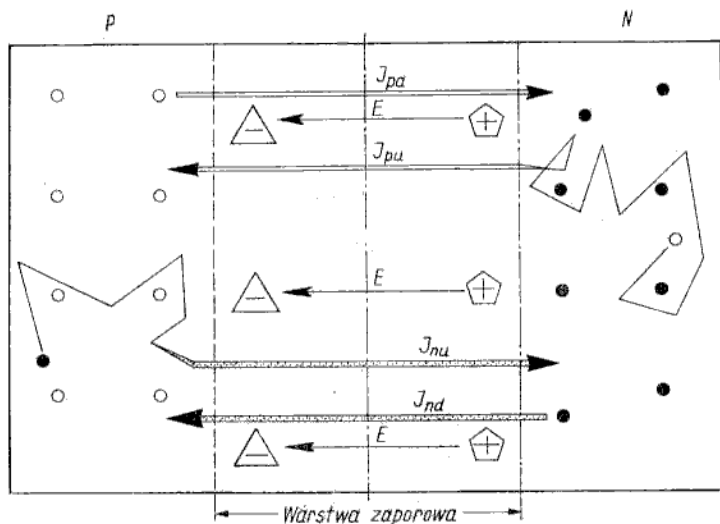
Analogicznie wskutek dyfuzyjnego przepływu dziur w obszarze granicznym warstwy P pozostają nieskompensowane ładunki ujemne nieruchomych centrów akceptorowych.

W obszarze granicznym warstw P , N powstaje zatem warstwa dipolowa ładunku, wytwarzająca pole elektryczne przeciwdziałające dyfuzji nośników większościowych (rys. 3.3b). Tę warstwę dipolową nazywa się warstwą zaporową lub warstwą ładunku przestrzennego, a napięcie wytworzone w obszarze granicznym złącza nosi nazwę bariery potencjału lub napięcia dyfuzyjnego. Napięcie dyfuzyjne powoduje uniesienie elektronów i dziur w kierunkach przeciwnych do ich dyfuzji. Zatem te dziury w warstwie N oraz te elektrony w warstwie P , które wskutek chaotycznego ruchu cieplnego znajdują się przypadkowo na granicy warstwy zaporowej, zostaną przechwycone przez pole elektryczne, istniejące w tej warstwie, i uniesione do obszaru przeciwnego typu zgodnie z kierunkiem działa-

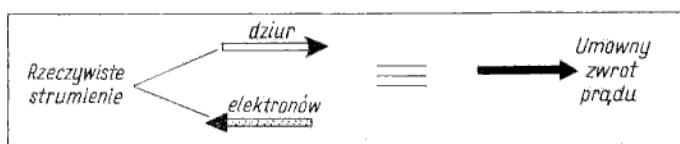
¹⁾ Należy pamiętać, że jest to eksperyment myślowy. W rzeczywistości nie można uzyskać złącza przez mechaniczne zetknięcie dwu warstw, gdyż złącze musi stanowić bryłę półprzewodnika monokrystalicznego z zachowaniem ciągłości budowy krystalicznej w płaszczyźnie styku.

te dziury w warstwie N uniesione do obszaru przeciwnego typu zgodnie z kierunkiem działania pola elektrycznego

jącego pola elektrycznego. W ten sposób powstają dwa strumienie prądu unoszenia nośników mniejszościowych, skierowane przeciwnie do dwu strumieni prądu dyfuzji nośników większościowych. Wymienione cztery składowe prądu płynącego przez złącze p-n przedstawiono schematycznie na rys. 3.4.



Legenda:



Rys. 3.4

Ilustracja czterech składowych prądu płynącego przez złącze p-n w stanie równowagi (bez napięcia zewnętrznego)

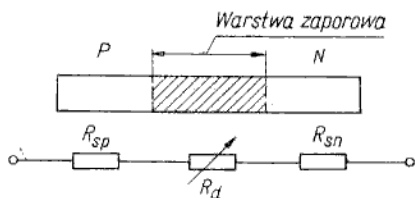
Dla złącza niespolaryzowanego napięciem zewnętrznym, tj. w stanie równowagi termodynamicznej, sumaryczny prąd płynący przez złącze musi być równy zero, a ładunek przestrzenny musi mieć wartość ustaloną. Te dwa warunki są spełnione, gdy prąd dyfuzji jest równy prądowi unoszenia oddzielnie dla dziur i elektronów:

$$\left. \begin{aligned} J_{pd} - J_{pu} &= 0 \\ J_{nd} - J_{nu} &= 0 \end{aligned} \right\} \quad (3.1)$$

przy czym: J_{pd} — gęstość prądu dyfuzji dziur; J_{pu} — gęstość prądu unoszenia dziur; J_{nd} — gęstość prądu dyfuzji elektronów; J_{nu} — gęstość prądu unoszenia elektronów.

W celu rozpatrzenia działania złącza p-n w stanie nierównowagi, tj. przy polaryzacji napięciem zewnętrznym, dogodnie jest podzielić strukturę złącza na trzy obszary o różnych właściwościach elektrofizycznych. Są to dwa obszary (typu p i n) obojętne elektrycznie (a ściślej biorąc quasi-obojętne), leżące poza warstwą zaporową, oraz obszar ładunku przestrzennego, tj. warstwa zaporowa, nazywana również warstwą zubożoną. Ta ostatnia nazwa pochodzi stąd, że w warstwie zaporowej jest mało ruchomych nośników ładunku, które są z niej „wymiatane”

pod wpływem napięcia dyfuzyjnego. Ze względu na małą koncentrację ruchomych nośników ładunku rezystancja warstwy zaporowej w dużym zakresie pracy złącza jest znacznie większa niż rezystancja obszarów obojętnych. Można przedstawić prosty model złącza p-n w postaci trzech rezystorów połączonych szeregowo (rys. 3.5), przy czym dominujący wpływ na kształt charakterystyki pra-

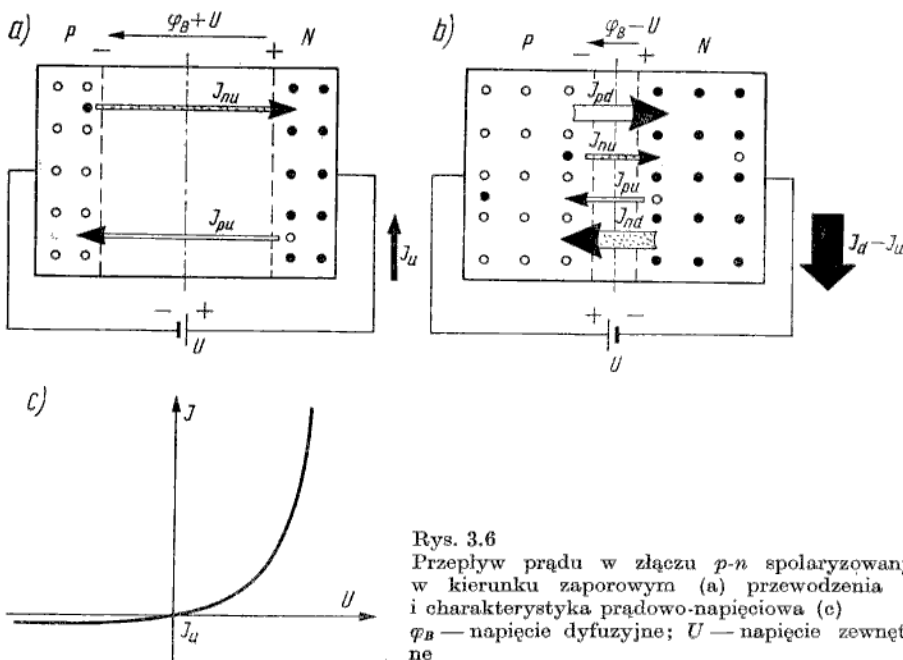


Rys. 3.5

Najprostszy model elektryczny złącza p-n R_{sp} , R_{sn} — rezystancje szeregowe obszarów obojętnych elektrycznie typu p i typu n; R_d — rezystancja warstwy zaporowej

dowo-napięciowej złącza ma nieliniowa rezystancja warstwy zaporowej. W zależności od polaryzacji napięcia zewnętrznego są możliwe dwie różne sytuacje:

1. Polaryzacja w kierunku zaporowym, tj. źródło zewnętrzne połączone biegunem dodatnim z warstwą N, a biegunem ujemnym z warstwą P. W tym przypadku polaryzacja zewnętrzna jest zgodna z biegunowością napięcia dyfuzyjnego, bariera potencjału zwiększa się zatem o wartość napięcia zewnętrznego. Zgodnie z prawami elektrostatyki większej wartości bariery potencjału musi odpowiadać większa wartość ładunku przestrzennego, przeto wzrasta szerokość warstwy zaporowej. Prawdopodobieństwo przejścia nośników większościowych ponad większą barierę potencjału jest mniejsze, czyli maleją składowe dyfuzyjne prądów elektronowego i dziurowego. W pierwszym przybliżeniu można natomiast przyjąć, że nie zmieniają się składowe prądu unoszenia nośników mniejszości-



Rys. 3.6

Przepływ prądu w złączu p-n spolaryzowanym w kierunku zaporowym (a) przewodzenia (b) i charakterystyka prądowo-napięciowa (c) φ_B — napięcie dyfuzyjne; U — napięcie zewnętrzne

wych. Słuszność tego stwierdzenia uzasadniamy w następujący sposób. Ponieważ $R_{sp}, R_{sn} \ll R_d$, zatem napięcie zewnętrzne, praktycznie biorąc, w całości odkłada się na warstwie zaporowej. Stąd w obszarach obojętnych nie ma pola elektrycznego, czyli ruch nośników mniejszościowych jest tak samo chaotyczny i pozbawiony składowej unoszenia, jak to było przy braku polaryzacji. A ponieważ prąd unoszenia nośników mniejszościowych przez warstwę zaporową jest określany prędkością dopływu tych nośników do granic warstwy zaporowej z obszarów obojętnych, jego wartość nie ulegnie zmianie w stosunku do stanu bez polaryzacji. Ostatecznie można stwierdzić, że przy polaryzacji w kierunku zaporowym składowe prądu dyfuzji nośników większościowych maleją do zera, pozostają natomiast składowe prądu unoszenia nośników mniejszościowych, niezależne od napięcia. Mówi się, że w kierunku zaporowym płynie przez złącze niewielki prąd nasycenia (rys. 3.6a).

2. Polaryzacja w kierunku przewodzenia, tj. źródło zewnętrzne jest połączone biegunem dodatnim z warstwą P , a biegunem ujemnym z warstwą N . W tym przypadku polaryzacja zewnętrzna jest przeciwna w stosunku do biegunowości napięcia dyfuzyjnego, bariera potencjału maleje zatem o wartość napięcia zewnętrznego. Mniejszej wartości bariery potencjału musi odpowiadać mniejsza wartość ładunku przestrzennego, przeto maleje również szerokość warstwy zaporowej. Prawdopodobieństwo przejścia nośników większościowych ponad mniejszą barierą potencjału jest większe, czyli wzrastają składowe prądu dyfuzji elektronów i dziur. Składowe prądu unoszenia nośników mniejszościowych pozostają natomiast na niezmiennym poziomie. Przy odpowiednio dużych wartościach napięcia polaryzacji prąd dyfuzji nośników większościowych zdecydowanie przeważa nad prądem unoszenia nośników mniejszościowych. Ostatecznie można stwierdzić, że przy polaryzacji w kierunku przewodzenia płynie prąd dyfuzji nośników większościowych znacznie większy niż prąd unoszenia nośników mniejszościowych (rys. 3.6b).

Na podstawie przedstawionych rozważań jakościowych można łatwo zaproponować wyrażenie analityczne na charakterystykę prądowo-napięciową złącza $p-n$. Ze statystyki Maxwella-Boltzmana wiadomo, że prawdopodobieństwo przejścia cząstki ponad barierą energetyczną W jest równe $\exp(-W/kT)$. Bariera energetyczna dla dyfundujących nośników większościowych jest równa energii pola elektrycznego w warstwie zaporowej, tj.

$$W = q(\varphi_B - U)$$

przy czym: φ_B — napięcie dyfuzyjne, nazywane również *barierą potencjału*; U — napięcie zewnętrzne o znaku dodatnim przy polaryzacji w kierunku przewodzenia oraz ujemnym przy polaryzacji w kierunku zaporowym.

Stąd gęstość prądu dyfuzji, która jest proporcjonalna do liczby nośników o energii większej niż $W = q(\varphi_B - U)$, można zapisać w postaci

$$J_d = a \exp \left[- \frac{q(\varphi_B - U)}{kT} \right]$$

przy czym a — współczynnik proporcjonalności.

Przy $U = 0$ (stan równowagi) składowa prądu dyfuzji jest równa składowej prądu unoszenia

$$J_d = J_u = a \exp \left(- \frac{q\varphi_B}{kT} \right)$$

$$\mu_n = a e^{-\frac{q \phi_0}{kT}}$$

$$a = \mu_n e^{\frac{q \phi_0}{kT}}$$

$$J_d = \mu_n e^{\frac{q \phi_0}{kT}} e^{-\frac{q U + q \phi_0}{kT}}$$

Stąd dla dowolnych wartości U , biorąc pod uwagę, że składowa prądu unoszenia nie zależy od napięcia zewnętrznego, można napisać równanie

$$J_d = J_u \exp \frac{qU}{kT}$$

A ponieważ całkowity prąd jest równy różnicy składowych prądu dyfuzji i unoszenia, zatem:

$$J = J_d - J_u$$

$$J = J_u \left(\exp \frac{U}{\varphi_T} - 1 \right) \quad (3.2)$$

przy czym dla uproszczenia zapisu wprowadzono oznaczenie potencjału elektrokinetycznego $\varphi_T = kT/q$.

Wyrażenie (3.2), znane jako *równanie Shockleya* (rys. 3.6c), będziemy uściślać w kolejnych punktach tego rozdziału, w których dokonamy analizy ilościowej działania złącza p-n.

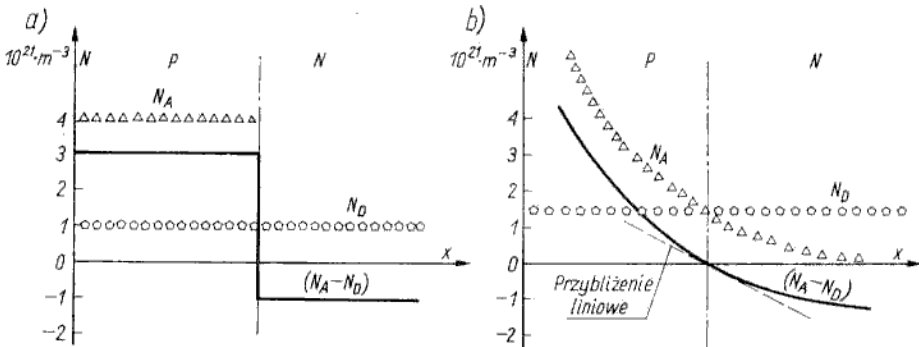
Obszar ładunku przestrzennego

3.1.2

Pojęcie złączy skokowego i liniowego

3.1.2.1

W rzeczywistych złączach p-n przejście od obszaru P do obszaru N może być mniej lub bardziej raptowne w zależności od technologii ich wytwarzania. Do analizy teoretycznej przyjmuje się dwa wyidealizowane modele rozkładu koncentracji domieszek, które są dość dobrymi przybliżeniami rzeczywistych rozkładów w typowych złączach p-n. Jeżeli przejście od obszaru P do obszaru N jest w miarę raptowne, a taka sytuacja istnieje w złączach wytwarzanych metodą stopową oraz w złączach epitaksjalnych i w „płytkich” złączach wytwarzanych metodą dyfuzyjną, to stosuje się przybliżenie skokowego rozkładu koncentracji domieszek (rys. 3.7a). Tę sytuację nazywa się *skokowymi*. W przypadku łagodnego przejścia od obszaru P do obszaru N , co występuje w „głębokich” złączach dyfuzyjnych, stosuje się przybliżenie liniowego rozkładu koncentracji domieszek (rys. 3.7b). Tę sytuację nazywa się *liniowymi*.



Rys. 3.7

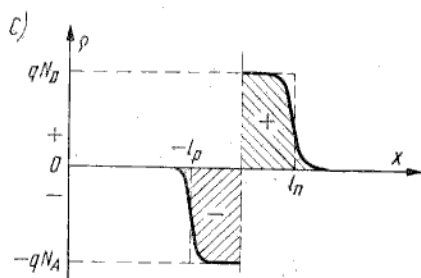
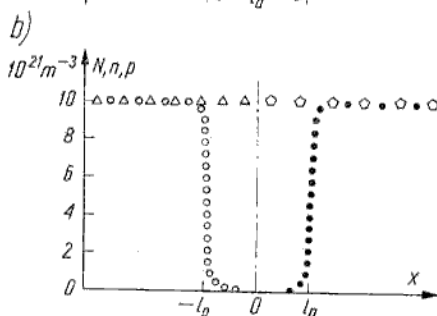
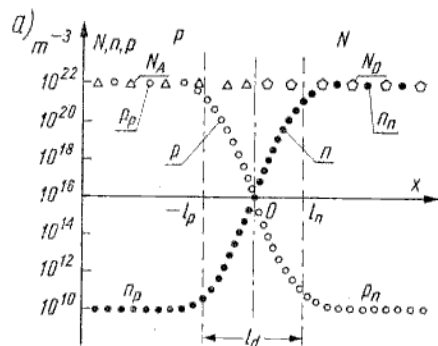
Rozkład koncentracji domieszek w złączach skokowym (a) i liniowym (b)

Złącze skokowe

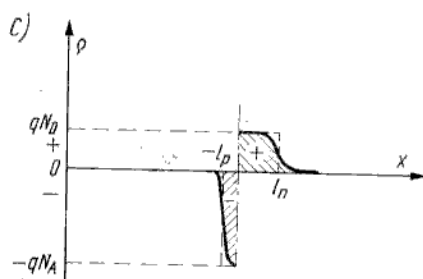
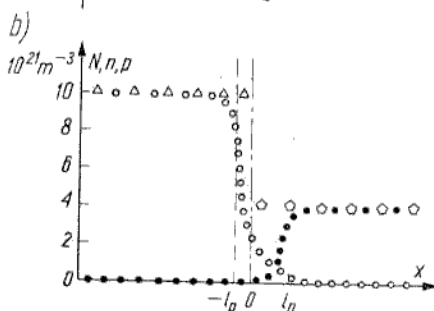
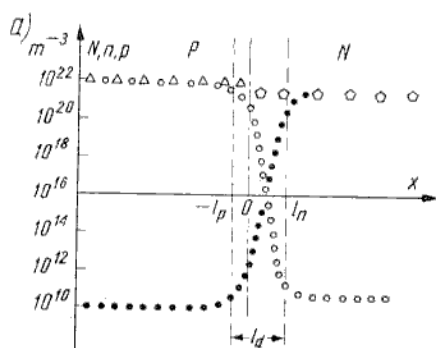
3.1.2.2

Stan równowagi (napięcie polaryzacji $U = 0$)

Na rysunkach 3.8, 3.9, 3.10 przedstawiono rozkłady koncentracji domieszek, koncentracji nośników (w skali logarytmicznej i liniowej) oraz gęstości ładunku



Rys. 3.8
Złącze skokowe symetryczne; rozkłady koncentracji domieszek i nośników w skali: a) logarytmicznej; b) liniowej; c) rozkład gęstości ładunku

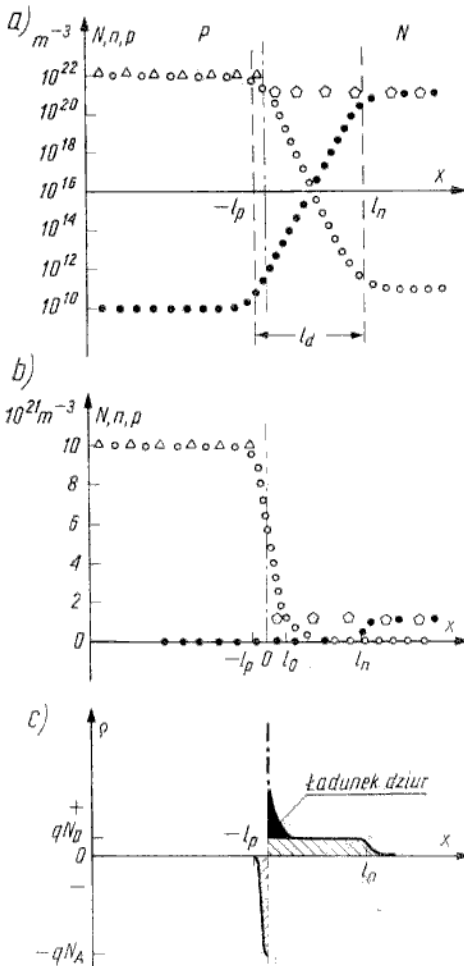


Rys. 3.9
Złącze skokowe niesymetryczne; rozkłady koncentracji domieszek i nośników w skali: a) logarytmicznej; b) liniowej; c) rozkład gęstości ładunku

dla trzech charakterystycznych przypadków złączy skokowych. Są to następujące przypadki:

$N_A = N_D$ — złącze symetryczne¹⁾,

¹⁾ Domieszki donorowe i akceptorowe mogą być po obu stronach złącza, dlatego ogólnie należy zapisać warunek $N_{AD} = N_{DA}$, przy czym $N_{AD} = N_A - N_D$, $N_{DA} = N_D - N_A$. Dla uproszczenia przyjęto, że w warstwie typu *p* są tylko akceptory, a w warstwie typu *n* tylko donory.



Rys. 3.10

Złącze skokowe silnie niesymetryczne; rozkłady koncentracji domieszek i nośników w skali: a) logarytmicznej; b) liniowej; c) rozkład gęstości ładunku

$N_A > N_D$ lub $N_D > N_A$ — złącze niesymetryczne,

$N_A \gg N_D$ lub $N_D \gg N_A$ — złącze silnie niesymetryczne.

Sposób kreślenia rozkładów $n(x)$, $p(x)$ zostanie omówiony na przykładzie rys. 3.8 dla złącza symetrycznego. Przyjęto, że jest rozpatrywane złącze krzemowe o koncentracjach domieszek $N_A = N_D = 10^{22} \text{ m}^{-3}$. Ponieważ w temperaturze pokojowej wszystkie atomy domieszek są zjonizowane, zatem koncentracje nośników większościowych w obszarach obojętnych elektrycznie (poza warstwą zaporową) są, praktycznie biorąc, równe koncentracjom domieszek, czyli $n_n = N_D$ oraz $p_p = N_A$.

Stąd na podstawie zależności $p_p n_p = n_i^2$ oraz $n_n p_n = n_i^2$, przyjmując w zaokrągleniu $n_i \approx 10^{16} \text{ m}^{-3}$, można wyznaczyć koncentracje nośników mniejszościowych:

$$p_n = \frac{n_i^2}{n_n} = 10^{10} \text{ m}^{-3}$$

$$n_p = \frac{n_i^2}{p_p} = 10^{10} \text{ m}^{-3}$$

Wiadomo już z rozważań jakościowych, przeprowadzonych w p. 3.1.1, że w obszarze granicznym złącza nie może być skokowej zmiany koncentracji nośników (w rozpatrywanym przykładzie byłby to skok koncentracji elektronów i dziur z wartości 10^{22} do 10^{10} m^{-3}). Wskutek dyfuzji ustalą się rozkłady $n(x)$, $p(x)$ z łagodnym przejściem od dużych do małych wartości, przy czym oba wykresy przetną się dla współrzędnej $x = 0$ przy wartości $n = p = n_i$. Można łatwo wykazać, że rozkłady $n(x)$, $p(x)$, przedstawione w skali logarytmicznej, są wzajemnie symetryczne względem poziomu n_i . Po zlogarytmowaniu zależności:

$$n_n p_n = n_i^2$$

$$p_p n_p = n_i^2$$

otrzymuje się

$$\lg n_i = \frac{\lg n_n + \lg p_n}{2} = \frac{\lg p_p + \lg n_p}{2}$$

W obszarze ograniczonym współrzędnymi $-l_p$ oraz l_n jest zakłócona obojętność elektryczna. Po stronie warstwy P ładunek zjonizowanych domieszek akceptorowych $-qN_A$ jest znacznie większy niż ładunek nośników większościowych (dziur), a ładunek nośników mniejszościowych (elektronów) jest zgoła pomijalnie mały. Natomiast po stronie warstwy N ładunek zjonizowanych domieszek donorowych qN_D jest znacznie większy niż ładunek nośników większościowych (elektronów), a ładunek nośników mniejszościowych (dziur) jest zgoła pomijalnie mały. Zdecydowaną przewagę ładunku jonów domieszek nad ładunkami nośników ruchomych (dziur i elektronów) szczególnie wyraźnie widać na wykresie rozkładów koncentracji, przedstawionym w skali liniowej (rys. 3.8b). Poczynione spostrzeżenia umożliwiają uproszczenie obliczeń ładunku przestrzennego w obszarze warstwy zaporowej. Mianowicie w pierwszym przybliżeniu można przyjąć, że w warstwie zaporowej istnieją wyłącznie ładunki zjonizowanych domieszek.

Tak uproszczone sformułowanie zagadnienia rozkładu ładunku nosi nazwę *przybliżenia warstwy zubożonej* lub *przybliżenia deplecyjnego* (ang. *depletion* — zubożenie). Trzeba się zastanowić nad słusznością takiego przybliżenia we wszystkich trzech przypadkach zilustrowanych na rys. 3.8, 3.9, 3.10.

W przypadku złącza symetrycznego całkowicie przekonywająca jest słuszność tego uproszczenia w pobliżu środka warstwy zaporowej (dla $x \approx 0$), gdzie koncentracje akceptorów po stronie warstwy P oraz donorów po stronie warstwy N są około milion razy większe niż koncentracje dziur i elektronów (ponadto ładunki dziur i elektronów kompensują się wzajemnie). Natomiast w pobliżu granicy warstwy zaporowej z obszarami obojętnymi elektrycznie koncentracje zjonizowanych atomów domieszek i nośników ruchomych stają się porównywalne. Dlatego rzeczywisty rozkład gęstości ładunku przestrzennego (wykreślony z uwzględnieniem zarówno zjonizowanych domieszek jak i nośników ruchomych), przedstawiony linią ciągłą na rys. 3.8c, ma „rozmyte” brzegi. Przybliżenie warstwy zubożonej oznacza przyjęcie rozkładu gęstości ładunku $\varrho(x)$, przedstawionego na rys. 3.8c linią przerywaną. Widać, że pole powierzchni zawartej pod wykresem rzeczywistej funkcji rozkładu $\varrho(x)$, praktycznie biorąc, nie różni się od pola prostokąta wyznaczonego przez linię przerywaną, reprezentującą aproksymowany rozkład $\varrho(x)$. Stąd należy wnioskować, że obliczenia ładunku przestrzennego na podstawie przybliżenia warstwy zubożonej dają prawidłowe wyniki dla złącza symetrycznego.

1) η Syme



W przypadku złącza niesymetrycznego (rys. 3.9) należy przede wszystkim zauważyć, że punktowi przecięcia się wykresów $n(x)$, $p(x)$ (dla $p = n = n_i$) odpowiada współrzędna $x > 0$, tzn. że punkt ten lokuje się w obszarze warstwy słabiej domieszkowanej. Wówczas w pobliżu współrzędnej $x = 0$ po stronie warstwy słabiej domieszkowanej koncentracja nośników mniejszościowych może nie być pomijalnie mała w porównaniu z koncentracją zjonizowanych atomów domieszek (w przypadku zilustrowanym wykresami na rys. 3.9 dla $x > 0$ na niewielkim odcinku p_n jest porównywalne z N_D). W tej sytuacji dokładność przybliżenia warstwy zubożonej maleje:

Ograniczoną dokładność i ograniczony zakres stosowalności tego przybliżenia jeszcze bardziej drastycznie ilustruje przypadek złącza silnie niesymetrycznego (rys. 3.10). Jak widać na rys. 3.10, w cienkiej warstwie ograniczonej współrzędnymi $(0, l_0)$ koncentracja nośników mniejszościowych jest większa niż koncentracja zjonizowanych atomów domieszek ($p_n > N_D$). W tym przypadku przybliżenie warstwy zubożonej jest niesłuszne i ładunek przestrzenny należy obliczać z uwzględnieniem ładunku nośników ruchomych. Są to wyprowadzenia dość złożone analitycznie, dlatego nie będą podawane. W tej książce ograniczymy się do analizy opartej na przybliżeniu warstwy zubożonej, pamiętając, że w złączach silnie niesymetrycznych jest to przybliżenie mało dokładne (w stanie nierównowagi przy polaryzacji złącza w kierunku wstecznym warstwa zaporowa rozszerza się i wówczas względny udział odcinka $0 \leq x \leq l_0$ jest coraz mniejszy, czyli błąd stosowanego przybliżenia maleje).

Zajmiemy się teraz opisem analitycznym właściwości warstwy ładunku przestrzennego, przy czym powinniśmy uzyskać odpowiedzi na następujące pytania:

- 1) Jaki jest rozkład gęstości ładunku przestrzennego?
- 2) Jaki jest rozkład natężenia pola i potencjału?
- 3) Jaka jest szerokość warstwy zaporowej?

W ramach przyjętego przybliżenia warstwy zubożonej odpowiedź na pierwsze pytanie jest prosta, gdyż rozkład gęstości ładunku przestrzennego jest zgodny z rozkładem koncentracji atomów domieszek. W rozważanych przypadkach złączy skokowych o równomiernych rozkładach koncentracji domieszek gęstość ładunku w warstwie zubożonej jest równa $-qN_A$ po stronie obszaru typu p oraz qN_D po stronie obszaru typu n. Z kolei związek między rozkładami potencjału i natężenia pola a rozkładem gęstości ładunku jest podstawowym zagadnieniem elektrostatyki, którego rozwiązanie jest sformułowane w prawie Gaussa.

$$\frac{\partial E}{\partial x} = \frac{\rho}{\epsilon_s} \quad \text{dla} \quad (3.3a)$$

oraz wyrażone w dogodnej postaci analitycznej równaniem Poissona

$$\frac{\partial^2 \varphi}{\partial x^2} = -\frac{\rho}{\epsilon_s} \quad \text{dla warst. n} \quad \rho = -qN_A \quad \frac{\partial^2 \varphi}{\partial x^2} = \frac{qN_A}{\epsilon_s \epsilon_0} \quad (3.3b)$$

Ponieważ złącze p-n jako całość zachowuje obojętność elektryczną, zatem obszar ładunku przestrzennego jest warstwą dipolową o równych ładunkach ujemnym i dodatnim

$$-qN_A l_p + qN_D l_n = 0 \quad (3.4)$$

Równość (3.4) oznacza inaczej, że nie ma pola elektrycznego w obszarach obojętnych (poza warstwą zaporową), czyli natężenie pola E dla współrzędnych $x < -l_p$

oraz $x \geq l_n$ jest równe zero. Maksymalne natężenie pola, występujące dla współrzędnej $x = 0$, otrzymuje się przez scałkowanie ładunku z lewej ($-l_p \leq x \leq 0$) lub prawej ($0 \leq x \leq l_n$) strony złącza zgodnie z równaniem (3.3a)

$$E_{max} = \frac{1}{\epsilon_s} \int_{-l_p}^0 \rho_A(x) dx = \frac{1}{\epsilon_s} \int_0^{l_n} \rho_D(x) dx$$

a ponieważ:

$$\rho_A(x) = -qN_A$$

$$\rho_D(x) = qN_D$$

więc

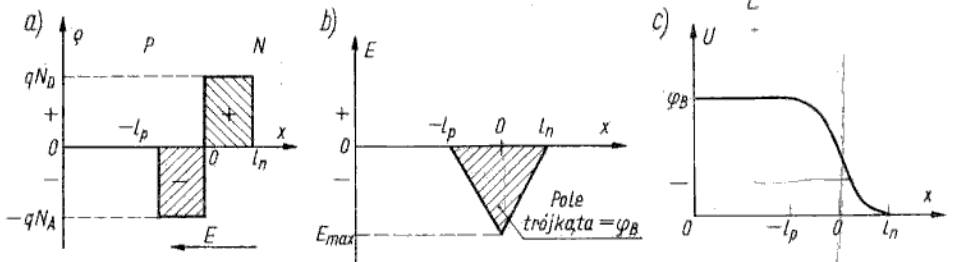
$$E_{max} = \frac{-qN_A l_p}{\epsilon_s} = \frac{-qN_D l_n}{\epsilon_s}$$

(3.5)

obstojnie
koncentracja
ładun. powierzchni
w miejscach typu p

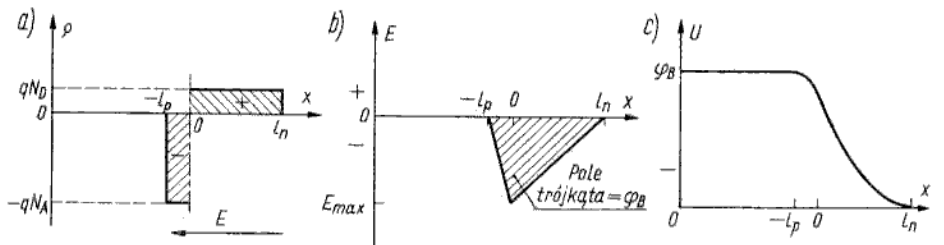
$\frac{1}{2} \rho_A \cdot l_n \cdot E_{max}$ Spr
- $\frac{q N_A l_p}{\epsilon_s}$
 $\frac{-q N_A l_p \cdot (l_p + l_n)}{2 \epsilon_s}$

Wykresy rozkładów gęstości ładunku, natężenia pola i potencjału przedstawiono na rys. 3.11 dla złącza symetrycznego i na rys. 3.12 dla złącza niesymetrycznego. Cal-



Rys. 3.11

Wykresy rozkładów gęstości ładunku elektrycznego (a), natężenia pola (b) i potencjału (c) dla złącza symetrycznego



Rys. 3.12

Wykresy rozkładów gęstości ładunku elektrycznego (a), natężenia pola (b) i potencjału (c) dla złącza niesymetrycznego

kując funkcję $E(x)$, tj. obliczając powierzchnię zawartą pod wykresem $E(x)$, otrzymuje się rozkład potencjału. Różnica potencjałów między krawędziami

aroda to słowo

Złącze p-n

$$\frac{-q N_A x_n}{2 \epsilon_s} \sqrt{2 \epsilon_s \phi}$$

$$-q N_A x_n + q N_D x_p = 0$$

$$N_A l_n + N_D l_p = 0$$

warstwy zaporowej, tj. napięcie dyfuzyjne, jest równe polu powierzchni trójkąta utworzonego przez rozkład $E(x)$

$$\varphi_B = \frac{1}{2} E_{max} l_d \quad \frac{q N_A l_p}{\epsilon_s} (l_p + l_n) \quad \text{Vol } p = N_A l_n \quad (3.6)$$

przy czym $l_d = l_p + l_n$ — szerokość warstwy zaporowej.

Na podstawie równań (3.6), (3.5) z uwzględnieniem (3.4) można wyprowadzić zależność $l_d = f(\varphi_B)$

$$l_d = \sqrt{\frac{2 \epsilon_s}{q} \frac{N_D + N_A}{N_D N_A} \varphi_B} \quad (3.7)$$

Dla złącza niesymetrycznego wyrażenie (3.7) można uprościć do postaci:

$$l_d \approx \sqrt{\frac{2 \epsilon_s}{q} \frac{1}{N_D} \varphi_B} \quad \text{dla } N_A \gg N_D \quad p^+ - n \quad l_n \gg l_p \quad (3.7a)$$

$$l_d \approx \sqrt{\frac{2 \epsilon_s}{q} \frac{1}{N_A} \varphi_B} \quad \text{dla } N_D \gg N_A \quad n^+ - p \quad (3.7b)$$

Ze wzorów (3.7a), (3.7b) widać, że w złączu niesymetrycznym o szerokości warstwy zaporowej decyduje koncentracja domieszek w obszarze słabiej domieszkowanym. Wynika to zresztą w oczywisty sposób z warunku obojętności elektrycznej złącza, gdyż na podstawie (3.4)

$$\text{dla } N_A \gg N_D \quad l_p \ll l_n$$

oraz

$$\text{dla } N_D \gg N_A \quad l_p \gg l_n$$

wysunął ze słupka

Wyrażenia (3.7), (3.7a), (3.7b) umożliwiają wyznaczenie szerokości warstwy zaporowej, jeżeli jest znane napięcie dyfuzyjne φ_B . Z kolei napięcie dyfuzyjne można wyznaczyć z warunku równowagi prądu dyfuzji i prądu unoszenia dla dziur lub elektronów. Na przykład dla dziur

$$q D_p \frac{dp}{dx} = q \mu_p p E \quad (3.8)$$

Równanie (3.8), po uwzględnieniu równania Einsteina $kT/q = D/\mu$, można przepisać w postaci ($\varphi_T = kT/q$)

$$\varphi_T \frac{dp}{p} = E dx \quad (3.9)$$

Równanie (3.9) całkuje się w granicach od $-l_p, p_p$ do l_n, p_n :

$$\varphi_T \int_{p_p}^{p_n} \frac{dp}{p} = \int_{-l_p}^{l_n} E dx = -\varphi_B \quad (3.10)$$

Stąd

$$g_T = \frac{kT}{q}$$

$$\varphi_B = \varphi_T \ln \frac{p_n}{p_p} \quad \text{dla silikonu} \quad (3.11)$$

Gdybyśmy analogiczne rozważania przeprowadzili dla elektronów, wówczas otrzymalibyśmy

$$\varphi_B = \varphi_T \ln \frac{n_n}{n_p} \quad \text{dla germanu} \quad (3.11a)$$

Po uwzględnieniu, że $n_n \approx N_D$ oraz $p_p \approx N_A$, można wyrazić napięcie dyfuzyjne w jeszcze inny sposób (korzysta się z zależności $np = n_i^2$)

$$\varphi_B = \varphi_T \ln \frac{N_A N_D}{n_i^2} \quad (3.11b)$$

Przykładowo w temperaturze pokojowej przy umiarkowanych wartościach koncentracji domieszek, tj. $N_A = N_D = 10^{22} \text{ m}^{-3}$, otrzymuje się:
dla złącza krzemowego

$$\varphi_B = 25 \text{ mV} \ln \frac{10^{44}}{2,25 \cdot 10^{32}} \approx 670 \text{ mV}$$

dla złącza germanowego

$$\varphi_B = 25 \text{ mV} \ln \frac{10^{44}}{6,25 \cdot 10^{38}} \approx 300 \text{ mV}$$

W ten sposób dokonaliśmy analizy ilościowej właściwości obszaru ładunku przestrzennego w złączu p - n skokowym w stanie równowagi, określając:

- rozkład gęstości ładunku,
- rozkład natężenia pola — prawo Gaussa (3.3a),
- rozkład potencjału — równanie Poissona (3.3b),
- napięcie dyfuzyjne — wzory (3.11), (3.11a), (3.11b),
- szerokość warstwy zaporowej — wzory (3.7), (3.7a), (3.7b).

St a n n i e r ó w n o w a g i (napięcie polaryzacji $U \neq 0$)

Przepływ prądu w stanie nierównowagi będzie omówiony nieco dalej. Obecnie interesuje nas tylko wpływ polaryzacji zewnętrznej na obszar ładunku przestrzennego. Bariera potencjału w złączu p - n jest równa sumie napięcia dyfuzyjnego i napięcia zewnętrznej. W związku z tym zmienia się szerokość warstwy zaporowej zgodnie z wzorem

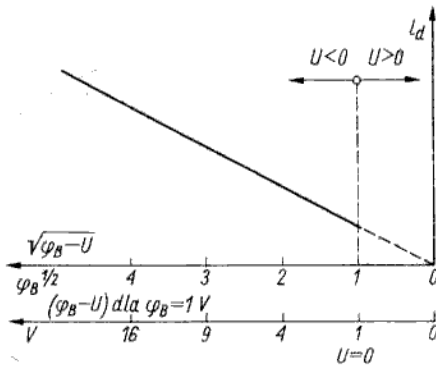
$$l_d = \sqrt{\frac{2\varepsilon_s}{q} \frac{N_A + N_D}{N_A N_D} (\varphi_B - U)} \quad (3.12)$$

przy czym U należy podstawić ze znakiem „-” dla kierunku zaporowego oraz „+” dla kierunku przewodzenia.

Przy polaryzacji w kierunku zaporowym warstwa zaporowa się rozszerza ($l_d \sim \sqrt{|U|}$) przy $|U| \gg |\varphi_B|$ oraz zwiększa się wartość maksymalnego natężenia pola ($E_{max} \sim \sqrt{|U|}$ przy $|U| \gg |\varphi_B|$). Zależność $l_d = f(U)$ dla złącza skokowego

niesymetrycznego, polaryzowanego w kierunku zaporowym przedstawiono na rys. 3.13.

Przy polaryzacji w kierunku przewodzenia warstwa zaporowa zwęża się zgodnie ze wzorem (3.12). Należy jednakże pamiętać, że wyrażenie (3.12) otrzymano dla przy-



Rys. 3.13
Zależność $l_d(U)$ dla złącza skokowego niesymetrycznego

bliżenia warstwy zubożonej. Przy pracy złącza w kierunku przewodzenia w miarę wzrostu napięcia polaryzacji, tj. wzrostu poziomu wstrzykiwania, koncentracja ruchomych nośników ładunku w warstwie zaporowej przyjmuje znaczne wartości. Wówczas dokładność wyrażenia (3.12), wyprowadzonego bez uwzględnienia ładunku nośników ruchomych, jest coraz gorsza.

Złącze liniowe

3.1.2.3

Stan równowagi (napięcie polaryzacji $U = 0$)

Jak już stwierdzono w p. 3.1.2.1, przy głębokiej dyfuzji domieszek rozkład ich efektywnej koncentracji $N_{AD}(N_{AD} = N_A - N_D)$ można na niewielkim odcinku aproksymować linią prostą o nachyleniu $\alpha = dN_{AD}/dx$. Rozkłady gęstości ładunku, potencjału i natężenia pola w złączu liniowym przedstawiono na rys. 3.14. Analogicznie jak dla złącza skokowego przyjmujemy, że jest słuszne przybliżenie warstwy zubożonej, czyli pomijamy ładunki nośników ruchomych w obszarze warstwy zaporowej. Po rozwiązaniu równania Poissona (3.3b) przy $\varrho = qax$ otrzymuje się wyrażenie na szerokość warstwy zaporowej

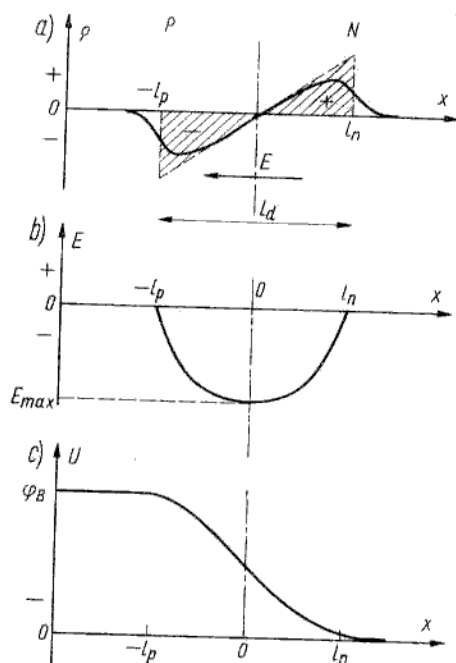
$$l_d = \sqrt[3]{\frac{12\epsilon_s}{qa} \varphi_B} \quad (3.13)$$

oraz maksymalne natężenie pola

$$E_{max} = 1,5 \frac{\varphi_B}{l_d} \quad (3.14)$$

Przez analogię do wzoru (3.11) napięcie dyfuzyjne można zapisać jako proporcjonalne do stosunku koncentracji nośników jednego rodzaju na brzegach warstwy zaporowej. Na przykład dla dziur

$$\varphi_B = \varphi_T \ln \frac{p(-l_p)}{p(l_n)} \quad \text{dla dziur}$$



Rys. 3.14
Wykresy rozkładów gęstości ładunku elektrycznego (a), natężenia pola (b) i potencjału (c) dla złącza liniowego

Ponieważ

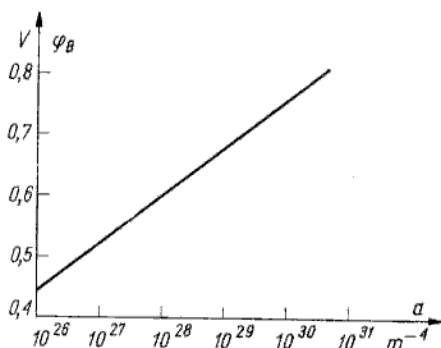
$$p(-l_p) \approx N_{AD}(-l_p) \approx -al_p$$

oraz

$$p(l_n) = \frac{n_i^2}{n(l_n)} = \frac{n_i^2}{-N_{AD}(l_n)} = \frac{n_i^2}{-al_n}$$

zatem

$$\varphi_B = \varphi_T \ln \frac{a^2 l_p l_n}{n_i^2}$$



Rys. 3.15
Zależność napięcia dyfuzyjnego od nachylenia rozkładu koncentracji domieszek w złączu liniowym

Z uwagi na liniowość i warunek obojętności elektrycznej złącza $l_p = l_n = l_d/2$, zatem

$$\varphi_B = 2\varphi_T \ln \frac{al_d}{2n_i} \quad (3.15)$$

Z układu dwu równań (3.13) i (3.15) można wyznaczyć l_d oraz φ_B . Obliczone w ten sposób napięcie dyfuzyjne w funkcji współczynnika a przedstawiono na rys. 3.15.

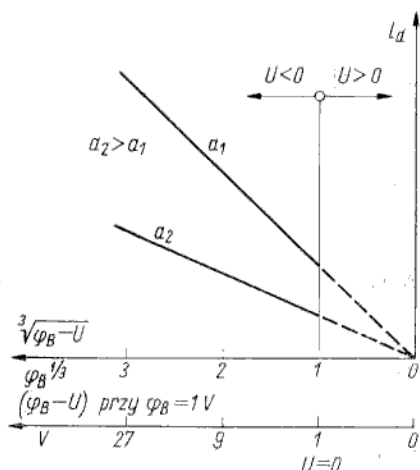
Stan nierównowagi (napięcie polaryzacji $U \neq 0$)

Podobnie jak dla złącza skokowego we wszystkich wyrażeniach, w których występuje napięcie dyfuzyjne, należy zamiast φ_B podstawić wartość bariery potencjału równą sumie napięcia dyfuzyjnego i napięcia zewnętrznego, wziętego z odpowiednim znakiem. Na przykład

$$l_d = \sqrt[3]{\frac{12\varepsilon_s}{qa} (\varphi_B - U)} \quad (3.16)$$

przy czym U należy podstawić ze znakiem „-” dla kierunku zaporowego oraz „+” dla kierunku przewodzenia.

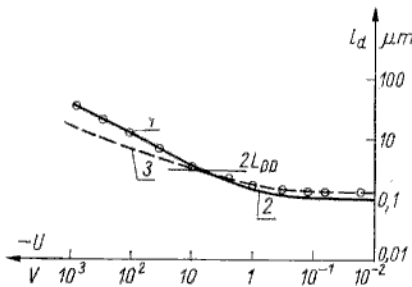
Wykresy $l_d(U)$ dla różnych wartości współczynnika a przedstawiono na rys. 3.16.



Rys. 3.16
Zależność $l_d(U)$ dla złącza liniowego przy różnych wartościach współczynnika a

Złącze rzeczywiste

O ile złącza stopowe, epitaksjalne i „płytkie” dyfuzyjne są dobrze opisane modelem złącza skokowego, o tyle „głębokie” złącza dyfuzyjne są z zadowalającą dokładnością reprezentowane przez model złącza liniowego tylko w ograniczonym zakresie napięcia polaryzacji. Na rysunku 3.17 przedstawiono przykładowo charakterystykę doświadczalną $l_d = f(U)$ oraz odpowiednie krzywe teoretyczne dla dwu modeli złącza. Można zauważyć, że w zakresie małych napięć lepszą zgodność wyników teoretycznych z doświadczalnymi daje model złącza liniowego, w zakresie natomiast dużych napięć bardziej prawidłowy jest model złącza skokowego. Stwierdzono, że punkt wyznaczający granicę między tymi dwoma zakresami odpowiada $l_d = 2L_{DD}$ (L_{DD} — droga dyfuzji domieszek).

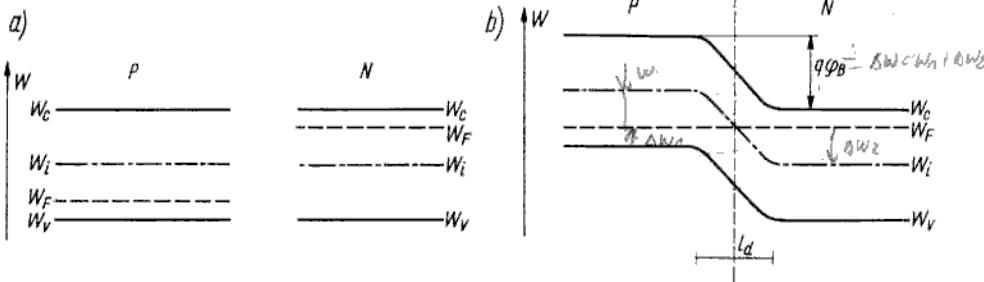


Rys. 3.17
Porównanie doświadczalnej zależności $l_d(U)$ dla złącza rzeczywistego (krzywa 1 — punkty oznaczone kółeczkami) z odpowiednimi zależnościami teoretycznymi dla modelu złącza skokowego (krzywa 2) oraz liniowego (krzywa 3)

Energetyczny model pasmowy złącza p-n

3.1.3

Rozpatrzmy najpierw stan równowagi przy braku polaryzacji zewnętrznej. Na rysunku 3.18 przedstawiono energetyczne modele pasmowe warstw P oraz N przed i po ich połączeniu. Model pasmowy złącza p-n jest kreślony z uwzględnieniem następujących zasad:



Rys. 3.18
Energetyczny model pasmowy warstw P, N przed połączeniem (a) i po utworzeniu złącza p-n (b)

- poziom Fermiego ma jedną stałą wartość w całym obszarze złącza;
- w obszarach obojętnych elektrycznie (leżących poza warstwą zaporową) nie zmieniają się koncentracje elektronów i dziur, tzn. położenie poziomu Fermiego w pasmie zabronionym pozostaje takie, jakie było w oddzielnych warstwach P, N.

Wynikający stąd skok poziomu energetycznego, dna pasma przewodnictwa oraz wierzchołka pasma walencyjnego reprezentuje barierę potencjału w złączu, a płynna zmiana energii potencjalnej elektronów i dziur zachodzi w obszarze warstwy zaporowej.

Stołość poziomu Fermiego w całym obszarze złącza wynika w oczywisty sposób z warunku równowagi termodynamicznej. Wiadomo, że poziom Fermiego ma sens średniej energii swobodnej elektronu, która w stanie równowagi termodynamicznej musi być jednakowa w jednej monolitycznej bryle półprzewodnika, jaką jest złącze p-n.

Energia elektronów zwiększa się w kierunku „do góry” rysunku, dziur natomiast — „do dołu” rysunku. Dlatego krawędzie pasma zabronionego (dno pasma przewodnictwa i wierzchołek pasma walencyjnego) są krzywymi zgodnymi z rozkładami ujemnego potencjału, określającego energię elektronów (rys. 3.11c, 3.12c, 3.14c).

$$\phi_B = (\Delta W_a + \Delta W_z) / q$$

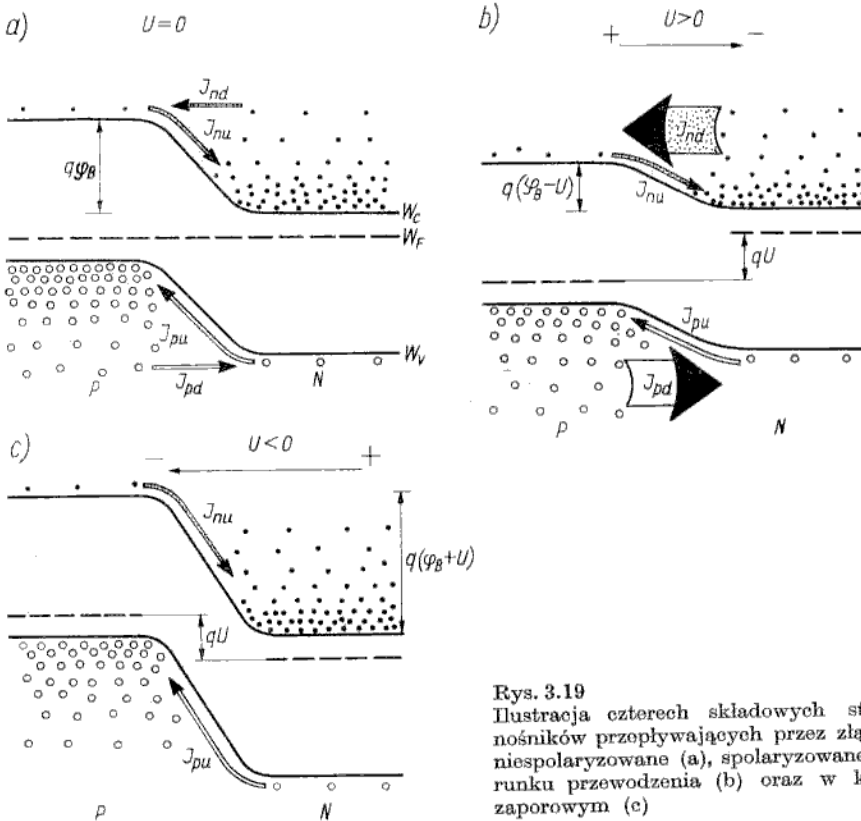
$$n_{p0} = n_i \exp\left(\frac{q\phi_B}{kT}\right)$$

$$p_{n0} = n_i \exp\left(-\frac{q\phi_B}{kT}\right)$$

$$\phi = -U$$

$$\text{dla } U=0 \quad \phi_B = (kT/q) \ln(n_{n0} p_{p0} / n_i^2)$$

Na rysunku 3.19 zilustrowano cztery składowe prądu płynącego przez złącze p-n w stanie równowagi oraz przy polaryzacji w kierunkach przewodzenia i zaporowym. W stanie równowagi składowe prądu dyfuzji i unoszenia kompensują się oddzielnie dla elektronów i dziur. W energetycznym modelu pasmowym oznacza to



Rys. 3.19
Ilustracja czterech składowych strumieni nośników przepływających przez złącze p-n niespolaryzowane (a), spolaryzowane w kierunku przewodzenia (b) oraz w kierunku zaporowym (c)

inaczej, że liczba elektronów w warstwie N (nośniki większościowe), dysponujących energią większą niż energia bariery, jest równa liczbie elektronów w warstwie P (nośników mniejszościowych). Analogicznie liczba dziur w warstwie P (nośników większościowych), dysponujących energią większą niż energia bariery (przypomina się, że energia dziur wzrasta „w dół” rysunku), jest równa liczbie dziur w warstwie N (nośników mniejszościowych). W stanie nierównowagi (przy polaryzacji zewnętrznej) składowe prądu dyfuzji nośników większościowych i unoszenia nośników mniejszościowych nie kompensują się, co bardzo dobrze ilustrują wykresy przedstawione na rys. 3.19b, c. Zauważmy, że na rys. 3.19b, c nie wykreślono poziomu Fermiego w warstwie zaporowej. Przy polaryzacji złącza p-n jest to obszar o zakłóconej równowadze termodynamicznej ($np \neq n_i^2$), dlatego pojęcie poziomu Fermiego nie ma w tym obszarze sensu. Można jedynie w sposób formalny wprowadzić pojęcie tzw. *quasi-poziomu Fermiego*, oddzielnego dla elektronów i dziur. Jest to pojęcie dogodne do analizy stanu nierów-

nowagowego, gdyż umożliwia korzystanie z wyrażań statystycznych, otrzymanych dla półprzewodnika w stanie równowagi. Zagadnienie to wykracza jednak poza ramy naszych zainteresowań.

Charakterystyka prądowo-napięciowa

3.1.4

Przybliżony kształt charakterystyki prądowo-napięciowej złącza $p-n$ jest już znany z rozważań jakościowych, przeprowadzonych w p. 3.1.1. Obecnie zajmemy się systematyczną analizą tej charakterystyki od strony ilościowej. Najpierw rozpatrzmy model wyidealizowany złącza $p-n$ (przybliżenie pierwszego stopnia) spełniający wiele założeń upraszczających, po czym uwzględnimy kolejno wpływ niektórych zjawisk pominiętych w tej uproszczonej analizie.

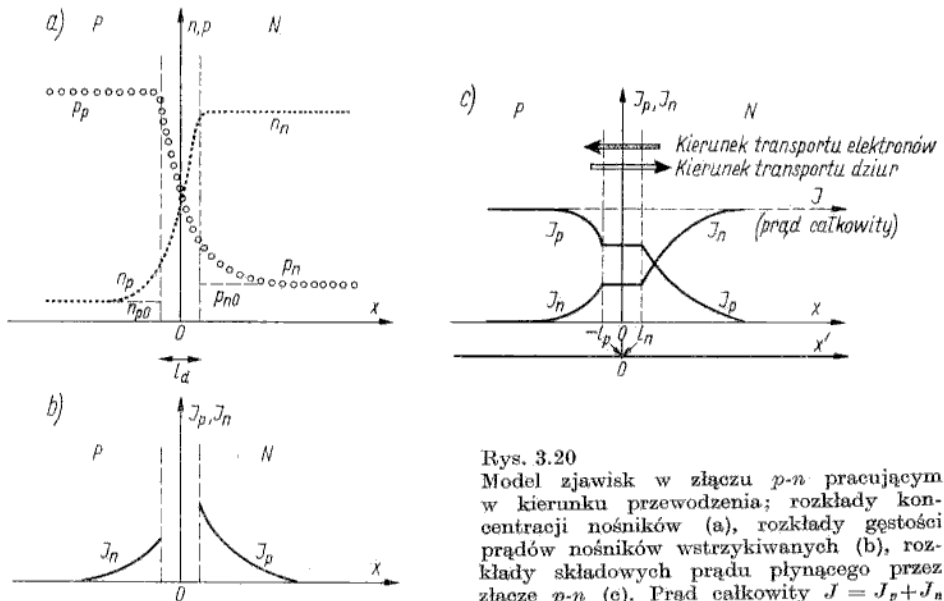
Przybliżenie pierwszego stopnia

3.1.4.1

W najbardziej uproszczonym modelu złącza $p-n$ przyjmuje się przede wszystkim, że:

- pole elektryczne istnieje tylko w warstwie zaporowej, co oznacza, że rezystancje obszarów obojętnych elektrycznie są równe zero, a nośniki wstrzykiwane do tych obszarów są przenoszone tylko na skutek dyfuzji;
- pomija się zjawiska generacji-rekombinacji w warstwie zaporowej;
- poziom wstrzykiwania nośników jest mały;
- pomija się zjawiska przebicia.

Model zjawisk w złączu $p-n$ pracującym w kierunku przewodzenia, przedstawiono na rys. 3.20, na którym pokazano rozkłady koncentracji nośników oraz gęstości prądu.



Rys. 3.20 Model zjawisk w złączu $p-n$ pracującym w kierunku przewodzenia; rozkłady koncentracji nośników (a), rozkłady gęstości prądów nośników wstrzykiwanych (b), rozkłady składowych prądów płynącego przez złącze $p-n$ (c). Prąd całkowity $J = J_p + J_n$

Do obszarów N , P są wstrzykiwane nośniki mniejszościowe nadmiarowe (dziury do warstwy N oraz elektrony do warstwy P), które dyfundują w głąb tych warstw oraz spotykając nośniki przeciwnego typu rekombinują z nimi. Dlatego koncentracja nośników nadmiarowych maleje w miarę oddalania się od krawędzi warstwy zaporowej.

W stanie ustalonym rozkłady ładunków nadmiarowych nie zmieniają się, każdemu zatem aktowi rekombinacji, tj. ubytkowi pary elektron-dziura, musi towarzyszyć proces uzupełnienia tej straty. Na przykład po rekombinacji nadmiarowej dziury z elektronem w obszarze N następuje wstrzyknięcie brakującej dziury z obszaru P oraz dopływ elektronu z głębi warstwy N , co w rezultacie oznacza dopływ elektronu z obwodu zewnętrznego. W ten sposób jest zachowana stałość sumarycznego prądu dziur i elektronów w dowolnym przekroju poprzecznym złącza p - n (rys. 3.20c).

Przy małym poziomie wstrzykiwania (wiadomo z p.1.7, że kryterium małego poziomu wstrzykiwania jest mała wartość koncentracji wstrzykniętych nośników nadmiarowych w porównaniu z koncentracją nośników większościowych) można przyjąć, że poza warstwą zaporową obszary P , N są elektrycznie obojętne, czyli prądy nośników wstrzykiwanych mają charakter dyfuzyjny. Całkowity prąd płynący przez złącze p - n najłatwiej jest wyznaczyć przez dodanie prądów elektronów i dziur na granicy warstwy zaporowej. Z uwagi na pominięcie zjawisk generacji i rekombinacji w warstwie zaporowej oba prądy są stałe w tym obszarze. Dlatego można dodać prąd dziur na prawej krawędzi warstwy zaporowej ($x = l_n$) do prądu elektronów na jej lewej krawędzi ($x = -l_p$). Dla uproszczenia zapisu analitycznego dogodnie będzie pominąć grubość warstwy zaporowej ($l_d = l_n + l_p = 0$), tj. posłużyć się osią x' według rys. 3.20c.

Całkowity prąd

$$J = J_{dn}|_{x'=0} + J_{dp}|_{x'=0} \quad (3.17)$$

Prądy elektronowy i dziurowy mają tylko składowe dyfuzyjne:

$$J_{dn} = qD_n \left. \frac{dn_p}{dx'} \right|_{x'=0} \quad (3.18a)$$

$$J_{dp} = -qD_p \left. \frac{dp_n}{dx'} \right|_{x'=0} \quad (3.18b)$$

przy czym q podstawia się ze znakiem dodatnim zarówno dla elektronów, jak i dziur.

Zagadnienie rozkładu ustalonego koncentracji nośników nadmiarowych w przypadku ich wstrzykiwania na jednej ze ścian bryły półprzewodnika było rozpatrywane w p. 1.7.

Otrzymano wówczas:

$$n_p = n_{p0} + n'_p(0) e^{-x/L_n} \quad (3.19a)$$

$$p_n = p_{n0} + p'_n(0) e^{-x/L_p} \quad (3.19b)$$

przy czym: n_{p0} , p_{n0} — koncentracje nośników mniejszościowych w stanie równowagi; $n'_p(0)$, $p'_n(0)$ — koncentracje nośników mniejszościowych nadmiarowych dla współrzędnej $x = 0$.

Po uwzględnieniu znaków zgodnie z kierunkami dodatnim i ujemnym na osi x' (rys. 3.20c):

$$n_p = n_{p0} + n'_p(0) e^{x'/L_n} \quad (x' \text{ podstawić ze znakiem „-”})$$

$$p_n = p_{n0} + p'_n(0) e^{-x'/L_p} \quad (x' \text{ podstawić ze znakiem „+”})$$

Po podstawieniu tych wyrażeń do (3.18) otrzymuje się:

$$J_{dn} = q \frac{D_n}{L_n} n'_p(0) \quad n'_p(0) = p_n(0) - p_{n0} \quad (3.20a)$$

$$J_{dp} = q \frac{D_p}{L_p} p'_n(0) \quad (3.20b)$$

W celu wyznaczenia koncentracji nośników nadmiarowych $n'_p(0)$, $p'_n(0)$ posłużymy się zależnościami wynikającymi ze statystyki Boltzmann'a. Za punkt odniesienia przyjmujemy stan bez polaryzacji ($U = 0$), w którym koncentracja nośników większościowych o energii większej niż energia bariery, tj. zdolnych do dyfuzji, jest równa koncentracji nośników mniejszościowych po drugiej stronie warstwy zaporowej (można się również posłużyć zależnościami (3.11), (3.11a)):

$$n_{n(dyf)}|_{U=0} = n_{p0} = n_n e^{-\varphi_B/\varphi_T} \quad (3.21a)$$

$$p_{p(dyf)}|_{U=0} = p_{n0} = p_p e^{-\varphi_B/\varphi_T} \quad (3.21b)$$

Przy polaryzacji zewnętrznej bariera potencjału wynosi $\varphi_B - U$, przez analogię zatem do (3.21) koncentracje nośników zdolnych do dyfuzji można wyrazić w postaci:

$$n_{n(dyf)}|_{U \neq 0} = n_n e^{-(\varphi_B - U)/\varphi_T} = n_{p0} e^{U/\varphi_T} \quad (3.22a)$$

$$p_{p(dyf)}|_{U \neq 0} = p_p e^{-(\varphi_B - U)/\varphi_T} = p_{n0} e^{U/\varphi_T} \quad (3.22b)$$

Stąd koncentracje nośników mniejszościowych nadmiarowych mają postać:

$$n'_p(0) = n_{n(dyf)}|_{U \neq 0} - n_{n(dyf)}|_{U=0} = n_{p0} (e^{U/\varphi_T} - 1) \quad (3.23a)$$

$$p'_n(0) = p_{p(dyf)}|_{U \neq 0} - p_{p(dyf)}|_{U=0} = p_{n0} (e^{U/\varphi_T} - 1) \quad (3.23b)$$

Po podstawieniu (3.23) do (3.20) i dodaniu obu składowych prądu zgodnie z (3.17) otrzymuje się

$$J = q \left[\frac{D_n n_{p0}}{L_n} + \frac{D_p p_{n0}}{L_p} \right] (e^{U/\varphi_T} - 1) \quad (3.24a)$$

Po uwzględnieniu, że

$$L = \sqrt{D\tau}$$

$$n_{p0} = \frac{n_i^2}{p_p}; \quad p_{n0} = \frac{n_i^2}{n_n}$$

$$p_p \approx N_A; \quad n_n \approx N_D$$

charakterystykę $J(U)$ można zapisać w innej dogodnej postaci

$$J = \underbrace{qn_i^2 \left(\sqrt{\frac{D_n}{\tau_n}} \frac{1}{N_A} + \sqrt{\frac{D_p}{\tau_p}} \frac{1}{N_D} \right)}_{J_s} (e^{U/\varphi_T} - 1) \quad (3.24b)$$

przy czym U należy podstawić ze znakiem „+” dla kierunku przewodzenia oraz „-” dla kierunku zaporowego.

Przy polaryzacji w kierunku zaporowym składnik $e^{-U/\varphi_T} \ll 1$, wówczas prąd przyjmuje stałą wartość równą tzw. prądowi nasycenia J_s

$$J = -J_s \quad (3.25)$$

W przypadku złącza niesymetrycznego jeden z dwu składników prądu J_s można pominąć. Przykładowo dla $N_A \gg N_D$

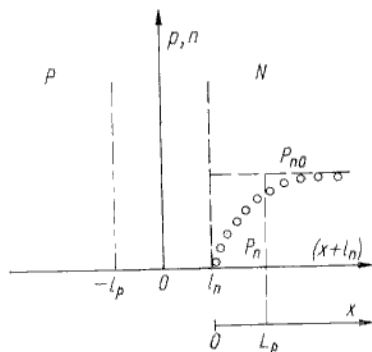
$$J_s \approx qn_i^2 \sqrt{\frac{D_p}{\tau_p}} \frac{1}{N_D} \quad (3.26)$$

skąd widać, że charakterystyka prądowo-napięciowa złącza niesymetrycznego jest określona właściwościami nośników mniejszościowych (D_p, τ_p) w warstwie słabiej domieszkowanej i koncentracją domieszek w tej warstwie. Warstwa słabiej domieszkowana jest nazywana *bazą*, silniej domieszkowana — *emiterem* złącza p-n. Decydującą zatem rolę odgrywają nośniki wstrzykiwane z emitera i dyfundujące w bazie złącza p-n.

Prąd J_s , płynący przy polaryzacji w kierunku wstecznym, ma prostą interpretację fizyczną. Jako przykład posłużmy drugi składnik ze wzoru (3.24)

$$J_s = qD_p \frac{p_{n0}}{L_p} \quad (3.27)$$

Wiadomo, że prąd w kierunku wstecznym jest utworzony przez strumień nośników mniejszościowych, unoszonych w silnym polu elektrycznym warstwy zaporowej. Jednakże wartość tego prądu nie zależy od natężenia pola w warstwie zaporowej, gdyż jest on ograniczony prędkością dopływu nośników mniejszościowych z obszarów obojętnych do krawędzi warstwy zaporowej. Jest to strumień dyfuzyjny, spowodowany gradientem koncentracji od p_{n0} w głębi obszaru N do wartości zerowej na krawędzi warstwy zaporowej (krawędź warstwy zaporowej stanowi płaszczyznę, na której nośniki są wyciągane z nieograniczoną prędkością).



Rys. 3.21
Rozkład koncentracji nośników mniejszościowych przy pracy złącza w kierunku zaporowym

kością). Zagadnienie stanu nierównowagi, spowodowanego wyciąganiem nośników, rozpatrywano w p. 1.7, otrzymując rozkład koncentracji w postaci (rys. 3.21)

$$p_n = p_{no}(1 - e^{-x/L_p}) \quad (3.28)$$

Z kolei prąd dyfuzji tych nośników na krawędzi warstwy zaporowej

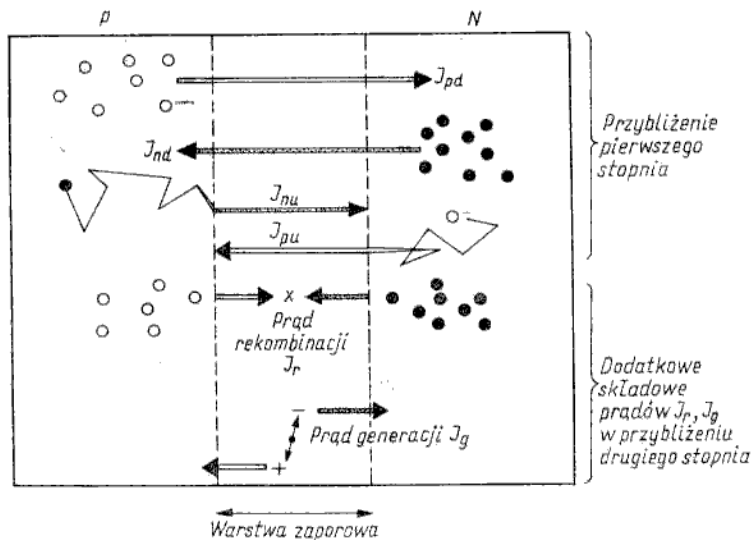
$$J_s = J_d = qD_p \left. \frac{dp_n}{dx} \right|_{x=0} \quad (3.29)$$

Po podstawieniu (3.28) do (3.29) otrzymuje się wyrażenie (3.27).

Przybliżenie drugiego stopnia

3.1.4.2

W najbardziej uproszczonym modelu — rozpatrywanym jako przybliżenie pierwszego stopnia — przyjęliśmy, że generacja i rekombinacja nośników zachodzi wyłącznie w obszarach obojętnych elektrycznie. W tym punkcie uwzględnimy ponadto wpływ zjawisk generacji-rekombinacji w obszarze warstwy zaporowej. Na rysunku 3.22 przedstawiono wszystkie składowe prądu płynącego



Rys. 3.22 Schematyczna ilustracja wszystkich składowych prądu płynącego przez złącze p-n

przez złącze p-n w tej nowej sytuacji. Obserwuje się dwie istotne zmiany w stosunku do „przybliżenia pierwszego stopnia”:

- część nośników większościowych dyfundujących przez warstwę zaporową ulega rekombinacji;
- wzrasta liczba nośników mniejszościowych, unoszonych w polu elektrycznym warstwy zaporowej wskutek generacji par elektron-dziura w obszarze tej warstwy.

Przy pracy złącza w kierunku zaporowym można przyjąć, że — praktycznie biorąc — nie ma nośników swobodnych w warstwie zaporowej, gdyż silne pole elektryczne

usuwa elektrony i dziury na zewnątrz tej warstwy. W tej sytuacji prawdopodobieństwo emisji elektronów i dziur z centrów generacyjno-rekombinacyjnych (procesy *a*, *b* na rys. 1.29) jest znacznie większe niż prawdopodobieństwo przechwytywania (pułapkowania) nośników przez te centra (procesy *c*, *d* na rys. 1.29). Zatem generacja przeważa nad rekombinacją. Można wykazać na podstawie statystyki SRH, że szybkość generacji, tj. liczba par elektron-dziura powstających w 1 m^3 w czasie 1 s jest w tym przypadku określona zależnością

$$G = \frac{n_i}{2\tau} \quad (3.30)$$

Elektron i dziura generowane w warstwie zaporowej są natychmiast z niej usuwane (w przeciwnych kierunkach) wskutek działającego pola elektrycznego. Gęstość prądu generacji jest zatem określona szybkością generacji nośników w warstwie zaporowej o jednostkowej powierzchni

$$J_g = qGl_d \quad (3.31)$$

Po podstawieniu (3.30) oraz l_d w postaci (3.7)

$$J_g = \frac{1}{2} q \frac{n_i}{\tau} \sqrt{\frac{2\epsilon_s}{q} \frac{N_A + N_D}{N_A N_D}} (\varphi_B - U) \quad (3.32)$$

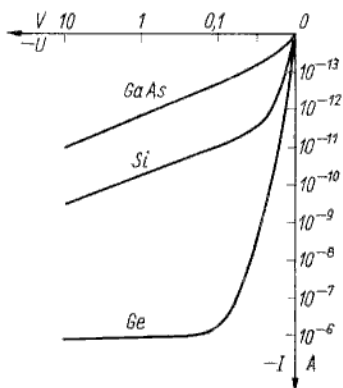
Porównajmy prądy J_s , J_g , stanowiące dwie składowe prądu wstecznego. Dla uproszczenia weźmy złącze niesymetryczne (przykładowo $N_A \gg N_D$), dla którego J_s jest wyrażone wzorem (3.26). Wówczas

$$\frac{J_g}{J_s} = \frac{1}{2n_i} \sqrt{\frac{2\epsilon_s}{q} \frac{N_D}{D_p \tau_p}} (\varphi_B - U) \quad (3.33)$$

lub w prostszej postaci (uwzględniając (3.7a))

$$\frac{J_g}{J_s} = \frac{1}{2} \frac{N_D}{n_i} \frac{l_d}{L_p} \quad (3.34)$$

Widać, że J_g/J_s zależy od koncentracji nośników samoistnych (a więc od szerokości pasma zabronionego), od koncentracji domieszek w bazie złącza oraz od ilorazu szerokości warstwy zaporowej i drogi dyfuzji. W aspekcie porównania



Rys. 3.23
Przykład porównania charakterystyk prądowo-napięciowych w kierunku zaporowym dla złączy z Ge, Si, GaAs (temperatura pokojowa)

złączy wykonanych z różnych materiałów półprzewodnikowych najistotniejszy jest wpływ koncentracji nośników samoistnych. Na przykład w porównywalnych złączach z germanu i krzemu $J_g/J_s \approx 3000$ dla Si oraz ok. 0,1 dla Ge, co wynika przede wszystkim z ok. 1000-krotnej różnicy wartości n_i w tych materiałach

Generalnie można stwierdzić, że im mniejsza jest wartość n_i , a więc im większa szerokość pasma zabronionego, tym większy jest udział składowej prądu generacji. Dlatego pominięcie prądu J_g , tj. przybliżenie pierwszego stopnia, jest usprawiedliwione dla złączy germanowych, lecz jest zupełnie niesłuszne dla złączy krzemowych, czy tym bardziej dla złączy z GaAs. Na rysunku 3.23 porównano charakterystyki $I(U)$ w kierunku zaporowym dla złączy z Ge, Si, GaAs. Tylko dla złącza germanowego prąd ma wyraźne nasycenie, co wynika z dominacji składowej J_s , dla złącza z Si lub GaAs charakterystyka ma kształt paraboliczny ($I \sim \sqrt{U}$) z uwagi na przewagę składowej J_g .

Inna jest sytuacja przy pracy złącza w kierunku przewodzenia. W tym przypadku do warstwy zaporowej są wstrzykiwane nośniki, dlatego koncentracje dziur i elektronów nie są bliskie zeru, lecz przeciwnie — mają większe wartości niż w stanie równowagi. Dlatego proces rekombinacji przeważa nad generacją. Przy $U = 0$ szybkość rekombinacji jest równa szybkości generacji

$$R_0 = \frac{n_i}{2\tau} = \frac{\sqrt{np}}{2\tau} \quad (3.35)$$

przy czym R_0 — szybkość rekombinacji w stanie równowagi.

W stanie równowagi $np = n_i^2$, natomiast w stanie nierównowagi, wywołanym wstrzykiwaniem nośników, $np > n_i^2$. Można wykazać, że w stanie nierównowagi np jest równe kwadratowi koncentracji nośników samoistnych pewnego hipotetycznego półprzewodnika o szerokości pasma zabronionego ($W_g - qU$)

$$np = n_i^2 e^{U/qT} \quad (3.36)$$

Przez analogię do wyrażenia (3.35) dla stanu nierównowagi można napisać

$$R = \frac{n_i}{2\tau} e^{U/2qT} \quad (3.37)$$

Ponieważ gęstość prądu rekombinacji

$$J_r = qRl_d \quad (3.38)$$

Zatem

$$J_r = \frac{1}{2} q \frac{n_i}{\tau} l_d e^{U/2qT} \quad (3.39)$$

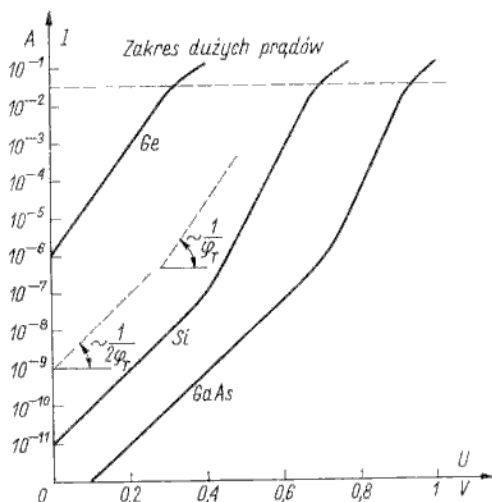
Podobnie, jak to uczyniliśmy dla kierunku zaporowego, porównajmy prąd rekombinacji z prądem dyfuzji dla kierunku przewodzenia (przypadek złącza niesymetrycznego o $N_A \gg N_D$)

$$\frac{J_r}{J_d} = \frac{1}{2} \frac{N_D}{n_i} \frac{l_d}{L_p} e^{-U/2qT} \quad (3.40)$$

Widać, że przy $U \approx 0$

$$\frac{J_r}{J_d} \approx \frac{J_g}{J_s}$$

Prąd rekombinacji jest zatem mniejszy niż prąd dyfuzji w złączu germanowym, jest natomiast zdecydowanie większy w złączu z Si lub GaAs. Ponadto na podstawie (3.40) można zauważyć, że w miarę wzrostu napięcia polaryzacji wpływ składowej dyfuzyjnej staje się coraz większy. Na rysunku 3.24 przedstawiono



Rys. 3.24
Przykład porównania charakterystyk prądowo-napięciowych w kierunku przewodzenia dla złączy z Ge, Si, GaAs (temperatura pokojowa)

trzy charakterystyki $I(U)$ w kierunku przewodzenia dla złączy z Ge, Si oraz GaAs. W przypadku Ge w całym zakresie napięć dominuje składowa prądu dyfuzji ($I \sim e^{qU/kT}$). W przypadku Si dla małych napięć ($< 0,4$ V) przeważa składowa prądu rekombinacji ($I \sim e^{qU/2kT}$), po czym charakterystyka przyjmuje kształt typowy dla prądu dyfuzji ($I \sim e^{qU/kT}$). Dla GaAs jest podobnie przy jeszcze większym zakresie dominacji składowej prądu rekombinacji. W zakresie dużych prądów, oddzielonym linią przerywaną na rys. 3.24, charakterystyka znów zmienia nachylenie. Zagadnienie to będzie omówione odrębnie w następnym punkcie. Należy zauważyć, że na rys. 3.24 uwzględniono tylko składowe prądu w kierunku przewodzenia, dlatego przy $U \rightarrow 0$ prąd nie zeruje się.

Wpływ innych zjawisk

3.1.4.3

W dotychczasowej analizie charakterystyki $I(U)$ złącza p-n uwzględniliśmy tylko cztery zjawiska:

- | | | |
|---------------------------------|---|--|
| przybliżenie pierwszego stopnia | } | — dyfuzja nośników większościowych z jednego obszaru neutralnego do drugiego przez warstwę zaporową, |
| | | — unoszenie nośników mniejszościowych przez warstwę zaporową, |
| przybliżenie drugiego stopnia | } | — jak wyżej oraz rekombinacja nośników w obszarze warstwy zaporowej, |
| | | — generacja nośników w obszarze warstwy zaporowej. |

W zależności od konstrukcji, technologii i zakresu pracy złącza p-n, jego właściwości mogą silnie zależeć od innych zjawisk, nie uwzględnianych w dotychczasowej analizie. Obecnie rozpatrzymy wpływ niektórych z tych zjawisk.

Na rysunku 3.24 pokazano zmianę kształtu charakterystyki $I(U)$ w zakresie dużych prądów. Mamy tu do czynienia z zagadnieniem dużego poziomu wstrzykiwania, gdy koncentracja nośników mniejszościowych nadmiarowych staje się porównywalna lub większa niż koncentracja domieszek w bazie złącza (na przykład dla złącza niesymetrycznego o $N_A \gg N_D$ oznacza to, że $p_n' \gg N_D$). Wówczas z warunku neutralności elektrycznej wynika, że koncentracja nośników większościowych zmienia się, gdyż $q(p_n + N_D - n_n) = 0$. Dla $p_n \gg N_D$ będzie $n_n \approx p_n$. Zatem przy dużym poziomie wstrzykiwania koncentracja nośników większościowych wzrasta zgodnie ze zwiększającą się koncentracją wstrzykiwanych nośników mniejszościowych. Mówi się, że następuje modulacja konduktywności półprzewodnika w obszarze bazy. Należy zwrócić uwagę, że ładunek nośników wstrzykiwanych jest większy niż ładunek centrów zjonizowanych, czyli model warstwy zubożonej przestaje być słuszny. Nie można również pomijać pola elektrycznego, jakie istnieje w tych warunkach w obszarze dotychczas uważanym za obojętny elektrycznie, tj. pozbawiony pola. Analiza ilościowa pracy złącza w zakresie dużych poziomów wstrzykiwania jest bardzo złożona i nie mieści się w zakresie naszych zainteresowań.

Przybliżony wynik końcowy takiej analizy można przewidzieć jakościowo, opierając się na wyrażeniu (3.36). Jeżeli $p_n \approx n_n$, to

$$p_n \approx n_i e^{U/2\varphi_T} \quad (3.41)$$

Jeżeli przyjmie się, że prąd i koncentracja nośników mniejszościowych są podobnymi funkcjami napięcia (analogicznie jak dla małego poziomu wstrzykiwania), to

$$J \sim e^{U/2\varphi_T} \quad (3.42)$$

Sens fizyczny zależności (3.42) jest taki, że przy dużym poziomie wstrzykiwania tylko połowa napięcia zewnętrznego odkłada się na warstwie zaporowej, druga zaś połowa — na quasi-neutralnym obszarze bazy.

Charakterystykę prądowo-napięciową dla całego zakresu przewodzenia można zapisać w postaci ogólnej

$$J \sim e^{U/m\varphi_T} \quad \text{przy } 1 \leq m \leq 2 \quad (3.43)$$

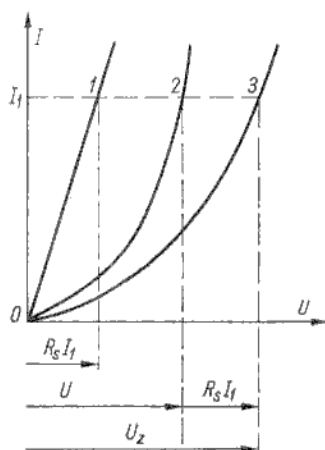
Dla złącza krzemowego (rys. 3.24) w zakresie małych prądów $m = 2$ (prąd rekombinacji), dla średnich prądów $m = 1$ (prąd dyfuzji), dla dużych natomiast prądów $m = 2$ (duży poziom wstrzykiwania).

W miarę wzrostu wartości prądu płynącego przez złącze p - n coraz większy wpływ na kształt charakterystyki może mieć pomijana dotychczas rezystancja szeregową (rezystancja obszarów obojętnych oraz w rzeczywistym złączu rezystancja styków półprzewodnika z wyprowadzeniami metalowymi i rezystancja tych wyprowadzeń). To zjawisko można bardzo łatwo uwzględnić, gdyż napięcie na warstwie zaporowej

$$U = U_z - R_s I \quad (3.44)$$

przy czym: U — napięcie na warstwie zaporowej; U_z — napięcie źródła zewnętrznego; R_s — rezystancja szeregową złącza; I — prąd płynący przez złącze.

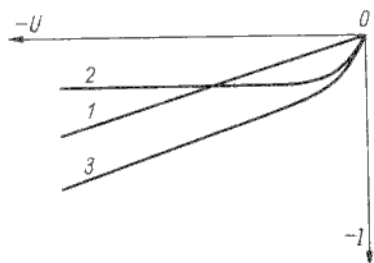
Wpływ rezystancji szeregową na charakterystykę $I(U)$ złącza zilustrowano na rys. 3.25. Przy polaryzacji w kierunku zaporowym istotny wpływ na kształt charakterystyki $I(U)$ może mieć prąd upływu, spowodowany zwykle tzw. stanami powierzchniowymi, tj. zerwanymi wiązaniami sieci krystalicznej w tym obszarze, w którym warstwa zaporowa wychodzi na powierzchnię półprzewodnika. W nowoczesnej technologii



Rys. 3.25

Wpływ rezystancji szeregowej na charakterystykę prądowo-napięciową złącza p-n przy pracy w kierunku przewodzenia
 1 — charakterystyka rezystancji R_s ; 2 — charakterystyka złącza idealnego (wg przybliżenia pierwszego stopnia);
 3 — charakterystyka wypadkowa

planarnej, charakteryzującej się tym, że powierzchnia krzemu jest pokryta (pa-sywowana) ochronną warstwą dwutlenku krzemu, powierzchniowy prąd upływu jest przeważnie pomijalnie mały. Wpływ składowej prądu upływu na charakterystykę złącza w kierunku zaporowym przedstawiono na rys. 3.26.

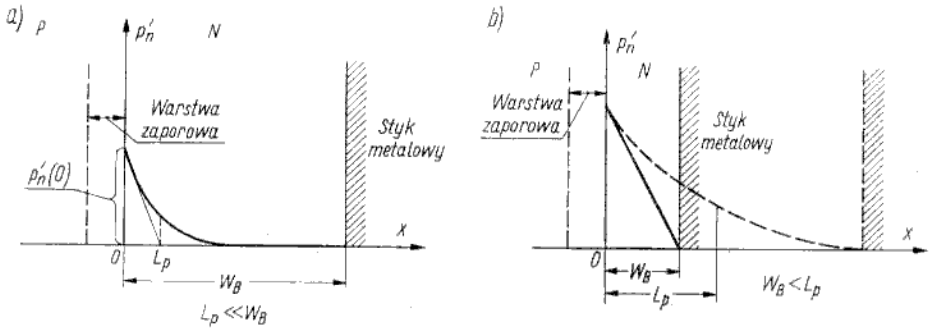


Rys. 3.26

Wpływ rezystancji upływu na charakterystykę prądowo-napięciową złącza p-n w kierunku zaporowym
 1 — charakterystyka rezystancji upływu; 2 — charakterystyka złącza idealnego (wg przybliżenia pierwszego stopnia); 3 — charakterystyka wypadkowa

W przebiegu charakterystyki $I(U)$ w kierunku zaporowym nie braliśmy na razie pod uwagę zakresu przebiecia charakteryzującego się gwałtownym wzrostem prądu przy napięciu przekraczającym pewną wartość. Jest to zagadnienie tak ważne, że będzie omówione odrębnie w następnym punkcie.

Dotychczas rozpatrywaliśmy tzw. złącza z długą bazą, w których obszar bazy miał szerokość znacznie większą niż droga dyfuzji nośników mniejszościowych (na przykład dla złącza niesymetrycznego o $N_A \gg N_D$ przyjmowano w domyśle że, $W_B \gg L_p$, przy czym W_B — szerokość bazy). Ten warunek w rzeczywistości nie zawsze jest spełniony. W przypadku W_B porównywalnego z L_p (tzw. złącze z krótką bazą) zmienia się rozkład koncentracji wstrzykniętych nośników nadmiarowych (rys. 3.27). Ponieważ na końcu bazy, tj. w miejscu styku półprzewodnika z wyprowadzeniem metalicznym, szybkość rekombinacji jest nieskończenie duża, zatem w tym miejscu koncentracja nośników mniejszościowych nadmiarowych jest równa zero. Stąd gradient koncentracji wstrzykniętych nośników nadmiarowych jest większy w złączu z krótką bazą niż w złączu z długą bazą. W granicznym przypadku dla $W_B \ll L_p$ we wzorach na charakterystykę prądowo-napięciową należy podstawić W_B zamiast L_p .



Rys. 3.27

Rozkład koncentracji nośników mniejszościowych nadmiarowych w bazie złącza $p-n$ przy pracy w kierunku przewodzenia: a) złącze z długą bazą ($W_B \geq L_p$); b) złącze z krótką bazą ($W_B < L_p$). Linia przerywaną narysowano rozkład, jaki byłby w przypadku długiej bazy

Przykładowo dla złącza niesymetrycznego o $N_A \gg N_D$

$$J = qp_{n0} \frac{W_B}{\tau_p} (e^{U/\phi_T} - 1) \quad (3.45)$$

Bardzo interesujące zjawiska występują w złączu $p-n$ silnie domieszkowanym. Są to zjawiska tak ważne, powodujące tak istotne zmiany jakościowe w kształcie charakterystyki prądowo-napięciowej, że nie można ich omawiać na zasadzie tylko kolejnego uściślenia dotychczas rozpatrywanego modelu. Dlatego dwa zagadnienia:

- przebicie złącza $p-n$,
 - złącze $p-n$ silnie domieszkowane
- będą omówione odrębnie w dwu następnych punktach.

Przebicie złącza $p-n$

3.1.5

Przebicciem złącza $p-n$ nazywa się zjawisko gwałtownego wzrostu prądu przy polaryzacji złącza w kierunku zaporowym napięciem większym niż pewna charakterystyczna dla danego złącza wartość, nazywana *napięciem przebicia*. Wyróżnia się dwie podstawowe przyczyny gwałtownego wzrostu prądu:

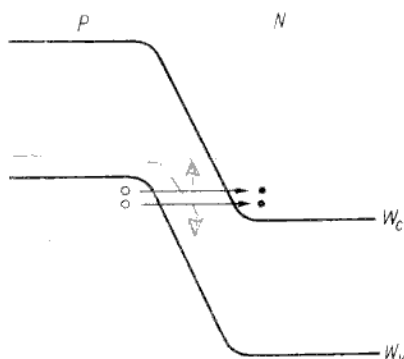
- przebicie Zenera,
- przebicie lawinowe.

Przebicie Zenera

3.1.5.1

Przy polaryzacji złącza w kierunku zaporowym natężenie pola w cienkiej warstwie zaporowej może osiągnąć tak duże wartości, że jest możliwe wyrwanie elektronu z wiązania kowalencyjnego atomów w sieci krystalicznej. Taka jonizacja elektrostatyczna jest nazywana *emisją wewnętrzną* lub *zjawiskiem Zenera*. W rezultacie zerwania wiązania atomów powstaje para nośników elektron-dziura.

W energetycznym modelu pasmowym akt jonizacji elektrostatycznej można zinterpretować jako przejście tunelowe elektronu z pasma walencyjnego przez barierę potencjału (przez pasmo zabronione) do pasma przewodnictwa (rys. 3.28). Tunelowe przenikanie cząstki przez barierę energetyczną, bez konieczności



Rys. 3.28
Prąd tunelowy nośników z pasma walencyjnego do pasma przewodnictwa (przebiecie Zenera)

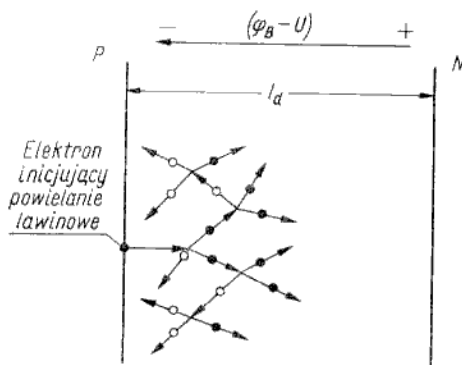
posiadania energii większej niż energia bariery, jest zjawiskiem wyjaśnianym w fizyce kwantowej. Prawdopodobieństwo przejścia tunelowego jest tym większe, im niższa i węższa jest bariera. Szerokość bariery, tj. szerokość warstwy zaporowej, maleje w miarę wzrostu koncentracji domieszek w złączu. Zatem przebiecie Zenera może występować w złączach silnie domieszkowanych. W cienkich warstwach zaporowych takich złączy pole elektryczne osiąga wartości ponad 10^8 V/m, jakie są wymagane do powstania zjawiska Zenera. Dla zjawiska Zenera charakterystyczna jest mała wartość napięcia przebiecia, ograniczona do $U_{pz} < 4 \frac{W_g}{q}$, co w przypadku złącza krzemowego oznacza $U_{pz} < 5$ V. Jeżeli złącze zostanie przebite przy napięciu wyższym niż $6 \frac{W_g}{q}$ (dla złącza krzemowego oznacza to $U_p > 7$ V), to gwałtowny wzrost prądu jest powodowany innym zjawiskiem, nazywanym *lawinową jonizacją zderzeniową*. W zakresie $4 \frac{W_g}{q} \leq U_p \leq 6 \frac{W_g}{q}$ występują jednocześnie przebiecia Zenera i lawinowe.

Przebiecie lawinowe

3.1.5.2

Zjawisko jonizacji zderzeniowej polega na rozerwaniu wiązania atomów w sieci krystalicznej wskutek dostarczenia energii przez swobodny nośnik ładunku, rozpędzony w silnym polu elektrycznym. Wskutek jonizacji zderzeniowej powstaje para nośników elektron-dziura, które poruszając się ruchem przyspieszonym w polu elektrycznym mogą uzyskać energię kinetyczną wystarczającą do jonizacji zderzeniowej. W ten sposób następuje powielenie liczby nośników w obszarze o silnym polu elektrycznym (w warstwie zaporowej). Jeżeli szerokość warstwy zaporowej jest znacznie większa niż droga swobodna nośnika (droga między dwoma kolejnymi zderzeniami), to można mówić o lawinowym powielaniu liczby nośników (rys. 3.29). Wzrost liczby nośników unoszonych w polu elektrycznym powoduje wzrost wartości prądu

$$J_{sp} = M_p J_s \quad (3.46)$$



Rys. 3.29
Ilustracja mechanizmu powielania lawinowego nośników w warstwie zaporowej złącza p-n spolaryzowanego zaporowo

przy czym: J_s — gęstość prądu wstecznego przed przebiciem; M_p — współczynnik powielania, definiowany jako stosunek liczby nośników wychodzących z warstwy zaporowej do liczby nośników doń wchodzącej.

Współczynnik powielania można wyrazić zależnością empiryczną

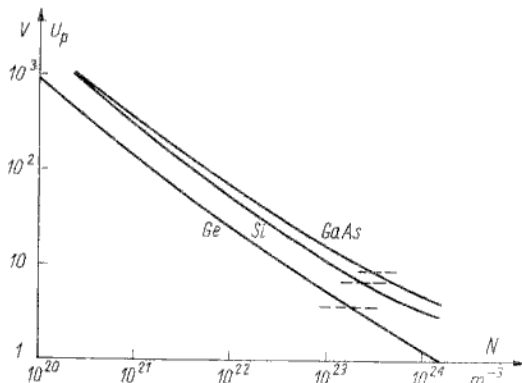
$$M_p = \frac{1}{1 - \left(\frac{|U|}{U_p}\right)^n} \quad (3.47)$$

przy czym n — współczynnik przybierający wartości 3...6.

Napięcie przebicia lawinowego (w V) można oszacować na podstawie wzoru empirycznego

$$U_p = 60 \left(\frac{W_g}{1,1}\right)^{3/2} \left(\frac{N}{10^{22}}\right)^{-3/4} \quad (3.48)$$

w którym N — koncentracja domieszek w bazie złącza [m^{-3}].



Rys. 3.30
Zależność napięcia przebicia od koncentracji domieszek dla złączy z Ge, Si i GaAs. Linia przerywaną zaznaczono granicę między przebiciem lawinowym a przebiciem Zenera

Zależność napięcia przebicia od koncentracji domieszek dla złączy z Ge, Si i GaAs przedstawiono na rys. 3.30, na którym linią przerywaną zaznaczono granicę między przebiciem lawinowym a przebiciem Zenera.

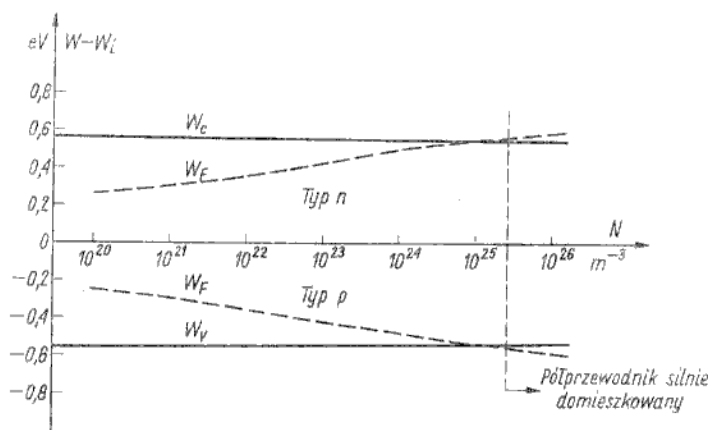
Na zakończenie warto podkreślić, że zarówno przebicie Zenera, jak i przebicie lawinowe nie powodują bezpośrednio uszkodzenia złącza. Jeżeli w obwodzie zewnętrznym

jest odpowiednie ograniczenie prądu, to złącze dowolnie długo może pracować w zakresie przebicia. Uszkodzenie następuje wówczas, gdy wskutek braku ograniczenia prądu w złączu wydziela się zbyt duża moc i nadmiernie wydzielane ciepło spowoduje zniszczenie złącza.

W tym przypadku mówi się o przebiciu cieplnym, rozumianym jako druga i zarazem destruktywna, nieodwracalna faza przebicia, poprzedzona nie niszczącą fazą przebicia Zenera lub przebicia lawinowego.

Charakterystyka prądowo-napięciowa złącza silnie domieszkowanego 3.1.6

Wiadomo, że w miarę wzrostu koncentracji domieszek poziom Fermiego przesuwa się ku krawędzi pasma zabronionego. W półprzewodniku typu *n* zbliża się on do dna pasma przewodnictwa, w półprzewodniku typu *p* — do wierzchołka pasma walencyjnego. Przy odpowiednio dużych koncentracjach domieszek ($N = 2 \times 10^{25} \text{ m}^{-3}$ dla germanu oraz $6 \cdot 10^{25} \text{ m}^{-3}$ dla krzemu) poziom Fermiego lokuje się na krawędzi pasma zabronionego, a jeszcze silniejsze domieszkowanie powoduje przejście poziomu Fermiego do pasma przewodnictwa (typ *n*) lub walencyjnego (typ *p*). Na rysunku 3.31 przedstawiono wykres położenia poziomu Fermiego w funkcji koncentracji domieszek.

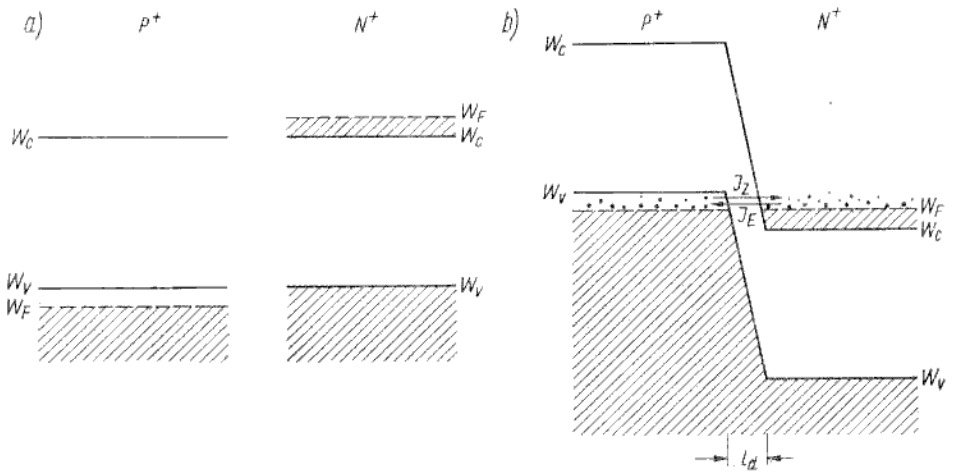


Rys. 3.31

Wykres poziomu Fermiego w funkcji koncentracji domieszek w krzemie

Rozpatrzmy teraz energetyczny model pasmowy złącza $p^+ \cdot n^{+1)}$ utworzonego z półprzewodnika silnie domieszkowanego (dla koncentracji domieszek zaznaczonej na rys. 3.31 linią przerywaną). Na rysunku 3.32 przedstawiono model pasmowy warstw P^+ , N^+ przed i po ich połączeniu. Z uwagi na stałość poziomu Fermiego w całym obszarze złącza okazuje się, że pasmo walencyjne po stronie półprzewodnika typu p^+ oraz pasmo przewodnictwa po stronie półprzewodnika typu n^+ znajdują się częściowo „naprzeciwko siebie”. Dotychczas tego rodzaju sytuacja występowała już raz przy omawianiu przebicia Zenera, które jest spowodowane przepływem prądu tunelowego elektronów z pasma walencyjnego przez barierę do pasma przewodnictwa. Prąd Zenera płynie wtedy, gdy w pasmie przewodnic-

¹⁾ Znak „+” oznacza półprzewodnik silnie domieszkowany.



Rys. 3.32

Model pasmowy warstw P⁺, N⁺ przed połączeniem (a) i po utworzeniu złącza p⁺-n⁺ (b)

twą istnieją wolne poziomy energetyczne „na wysokości” poziomów zajmowanych przez elektrony w pasmie walencyjnym (elektron przy przejściu tunelowym nie zmienia energii). Jednakże w stanie równowagi (bez polaryzacji zewnętrznej) sumaryczny prąd płynący przez złącze p⁺-n⁺ powinien być równy zero. Istnieje zatem jednakowe prawdopodobieństwo przejścia tunelowego elektronu z pasma walencyjnego do pasma przewodnictwa (z P⁺ do N⁺), jak i w kierunku przeciwnym.

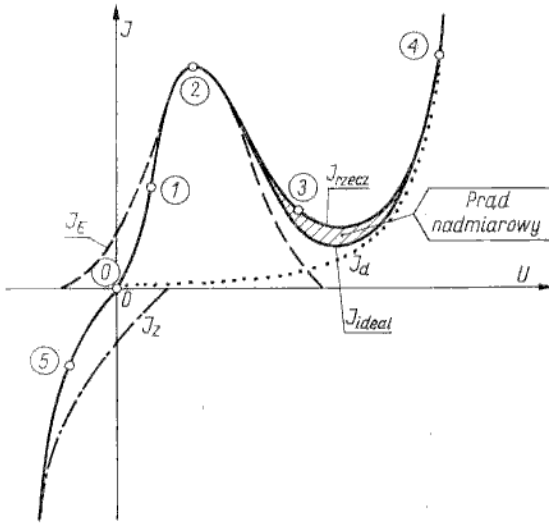
Możliwość przepływu tunelowego elektronów z pasma przewodnictwa do pasma walencyjnego jako pierwszy zauważył Leo Esaki¹⁾ i dlatego taki przepływ elektronów nazwano *prądem Esakiego*. W stanie równowagi prąd Zenera jest równy prądowi Esakiego

$$J_Z + J_E = 0$$

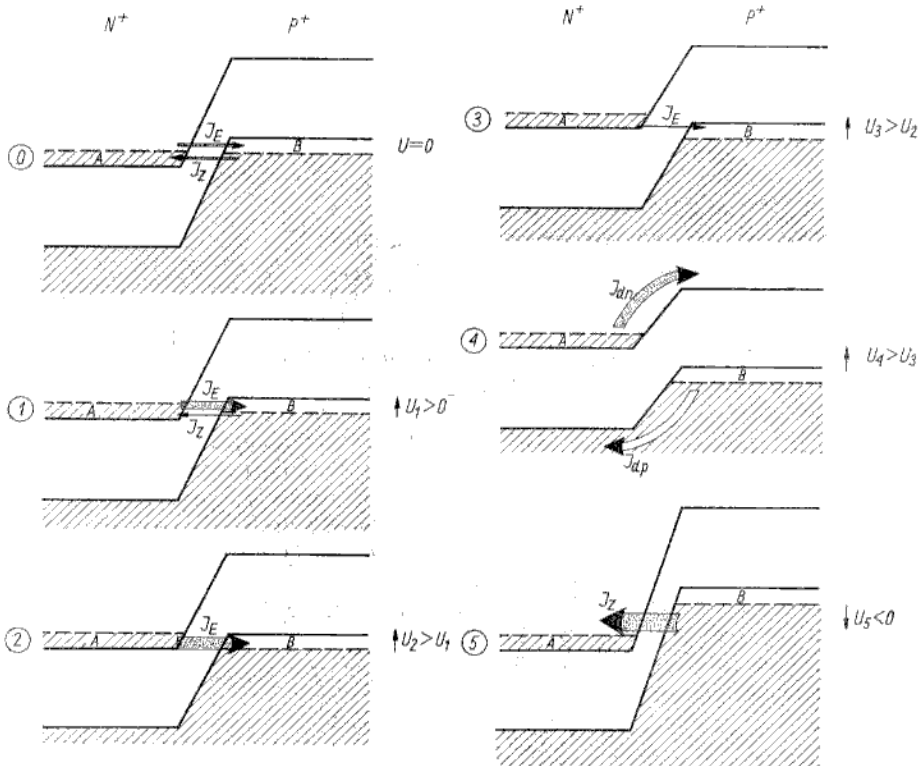
Opisywane zjawiska obrazowo ilustruje model pasmowy na rys. 3.32b. Pola zakreskowane pod poziomem Fermiego oznaczają pasma energetyczne całkowicie zajęte przez elektrony. Poziomy ułożone powyżej energii Fermiego są puste. Wiadomo jednak, że funkcja rozkładu Fermiego-Diraca tylko w temperaturze 0 K ma raptowny skok od wartości 1 do 0 przy przejściu przez poziom Fermiego. W temperaturze pokojowej przejście od wartości 1 do 0 jest bardziej łagodne, czyli tuż pod poziomem Fermiego jest pewna liczba wolnych stanów energetycznych, a tuż nad poziomem Fermiego jest pewna liczba elektronów (kropki na rys. 3.32b). W ten sposób jest spełniony podstawowy warunek wymagany dla wystąpienia przejść tunelowych, tj. istnienie stanów energetycznych, zajętych na jednej stronie złącza, a wolnych na drugiej stronie.

Dodatkowym warunkiem, niezbędnym do przepływu prądu tunelowego, jest mała szerokość bariery, przez którą zachodzi przenikanie tunelowe. Ten warunek jest spełniony, gdyż w złączu silnie domieszkowanym warstwa zaporowa jest bardzo cienka (ok. 10 nm). Natężenie pola elektrycznego w tak cienkiej warstwie zaporowej jest bardzo duże — ok. 10⁸ V/m. W ścisłym opisie działania złącza p⁺-n⁺

¹⁾ Leo Esaki, profesor uniwersytetu w Osace (Japonia) jest laureatem nagrody Nobla z 1973 r.



Rys. 3.33
Charakterystyka prądowo-napięciowa złącza $p^+ \cdot n^+$ i charakterystyki trzech prądów składowych: J_Z — prąd Zenera; J_E — prąd Esakiego; J_d — prąd dyfuzji nośników większościowych



Rys. 3.34
Wykresy modelu pasmowego, odpowiadające wybranym punktom charakterystyki prądowo-napięciowej, przedstawionej na rys. 3.33

należałoby oprócz prądów Zenera i Esakiego uwzględnić znane z wcześniejszych rozważań dwa prądy nośników mniejszościowych i dwa prądy nośników większościowych. Jednak w stanie równowagi (bez polaryzacji zewnętrznej) te cztery składowe prądy są tak małe, że można ich wpływ pominąć.

Rozpatrzmy teraz działanie złącza p - n silnie domieszkowanego przy polaryzacji zewnętrznej. Na rysunku 3.33 przedstawiono charakterystykę prądowo-napięciową, której kształt wyjaśniają modele pasmowe, przedstawione na rys. 3.34. Interpretację charakterystyki prądowo-napięciowej dogodnie jest oprzeć na obserwacji wzajemnego usytuowania pasm oznaczonych na rys. 3.34 literami A i B . Pasma A , ułożone między poziomem Fermiego a dnem pasma przewodnictwa w warstwie typu n^+ , jest zajęte przez elektrony. Pasma B , ułożone między wierzchołkiem pasma walencyjnego a poziomem Fermiego w warstwie typu p^+ , jest puste. W punkcie ① ($U = 0$) prądy tunelowe J_Z, J_E wzajemnie równoważą się. W punkcie ① ($U_1 > 0$) prąd Esakiego przeważa nad prądem Zenera, gdyż duża część pasma A znajduje się „naprzeciwko” pasma B . W punkcie ② ($U_2 > U_1$) płynie tylko prąd Esakiego, który osiąga wartość maksymalną, gdyż pasma A i B są ułożone „naprzeciwko” siebie całymi przekrojami. W punkcie ③ ($U_3 > U_2$) prąd Esakiego jest znacznie mniejszy, gdyż tylko część pasma A jest ułożona „naprzeciwko” pasma B . Z pozostałej części pasma A elektrony nie mogą tunelować, gdyż po przeciwległej stronie warstwy zaporowej nie ma dozwolonych poziomów energetycznych (jest pasmo zabronione). W punkcie ④ ($U_4 > U_3$) prąd Esakiego nie płynie, gdyż pasmo A w całości lokuje się „naprzeciwko” pasma zabronionego. Jest natomiast możliwy przepływ ponadbarierowego prądu dyfuzji nośników większościowych z uwagi na znaczne zmniejszenie wysokości bariery. Przy polaryzacji w kierunku zaporowym, na przykład w punkcie ⑤ ($U_5 < 0$), płynie tylko prąd Zenera, gdyż elektrony z pasma walencyjnego w warstwie typu p^+ mogą tunelować do wolnych poziomów energetycznych w pasmie przewodnictwa warstwy typu n^+ .

Charakterystyka prądowo-napięciowa złącza silnie domieszkowanego powstaje zatem wskutek superpozycji trzech charakterystyk: $J_E(U), J_Z(U), J_d(U)$. Najbardziej interesującą właściwością tej charakterystyki jest istnienie odcinka o ujemnej rezystancji przyrostowej (między punktami ekstremalnymi charakterystyki), tj. odcinka wyróżniającego się wzrostem prądu przy zmniejszaniu napięcia. Maksimum prądu (punkt ②) nazywa się wierzchołkiem charakterystyki, a minimum (tuż za punktem ③) — doliną. W obszarze doliny prąd rzeczywisty jest nieco większy niż prąd wynikający tylko z superpozycji składowych J_E, J_d . Mówi się w tym przypadku o prądzie nadmiarowym (pole zakreskowane na rys. 3.33), który jest spowodowany przez dwa zjawiska:

- rozmycie krawędzi pasma zabronionego w półprzewodniku silnie domieszkowanym wskutek rozszczepienia poziomu energetycznego domieszki w pasmo energetyczne;
- tunelowanie pośrednie przez centra generacyjno-rekombinacyjne, znajdujące się w warstwie zaporowej.

Te zjawiska nie będą omawiane, gdyż mają wpływ drugorzędny i są rozpatrywane w literaturze specjalistycznej. Zrezygnujemy również z wyprowadzenia analitycznej postaci charakterystyki prądowo-napięciowej, gdyż zaprowadziłoby to nas zbyt głęboko w zagadnienia fizyki kwantowej.

Do zjawisk przedstawionych w tym punkcie powrócimy w następnym rozdziale przy omawianiu diody tunelowej — elementu zbudowanego na bazie struktury fizycznej złącza $p^+ - n^+$.

Modele złącza p - n

3.1.7

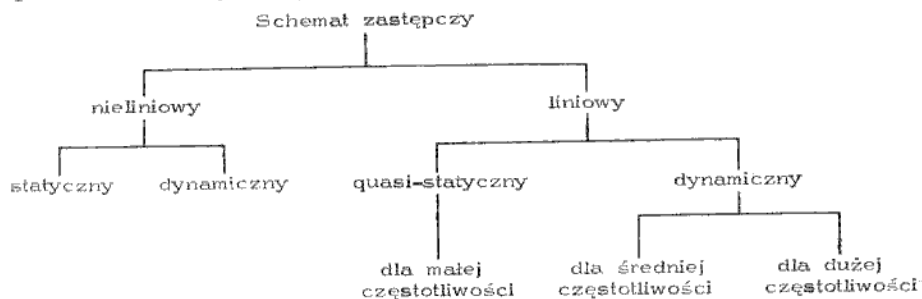
W projektowaniu przyrządów półprzewodnikowych lub układów elektronicznych, w których przyrządy półprzewodnikowe spełniają funkcje elementów składowych, stosuje się modele reprezentujące z lepszą lub gorszą dokładnością fizycznie realne elementy. W zależności od sposobu przedstawienia elementu można wymienić cztery rodzaje modeli:

- analityczne,
- graficzne,
- tablicowe,
- symboliczne (schematy zastępcze).

Modelem analitycznym złącza p - n jest na przykład równanie charakterystyki prądowo-napięciowej, wyprowadzone w poprzednich punktach tego rozdziału. Modelem graficznym jest wykres zależności $I(U)$. Modelem tablicowym są ciągi liczb wyrażających odpowiadające sobie wartości prądu i napięcia, zmierzone „punkt po punkcie”.

Obecnie interesują nas głównie modele symboliczne, tj. schematy zastępcze złącza p - n , przedstawione w postaci symboli elektrycznych odpowiednio połączonych elementów wyidealizowanych (rezystancji, pojemności itd.).

Przedstawienie elementu rzeczywistego w postaci schematu zastępczego ma dużo zalet. Przede wszystkim układ złożony z wielu elementów ma postać sieci, którą można analizować za pomocą metod znanych z teorii obwodów elektrycznych. Ponadto w schemacie zastępczym łatwo jest odzwieriedlić działanie elementu zarówno dla warunków statycznych, jak i dynamicznych. (O warunkach dynamicznych mówi się wówczas, gdy zmienne zależne są uzależnione zarówno od wartości chwilowych, jak i od szybkości zmian zmiennych niezależnych. Na przykład prąd w złączu p - n zależy od czasu, jaki upłynął po natychmiastowej zmianie napięcia). Oprócz rozróżnienia pracy statycznej i pracy dynamicznej w przypadku elementu nieliniowego, jakim jest między innymi złącze p - n , należy rozróżnić pracę przy dużych sygnałach (model nieliniowy) i pracę przy małych sygnałach (model liniowy). Biorąc pod uwagę wyżej wprowadzone pojęcia zostaną rozpatrzone schematy zastępcze według klasyfikacji podanej na rys. 3.35.

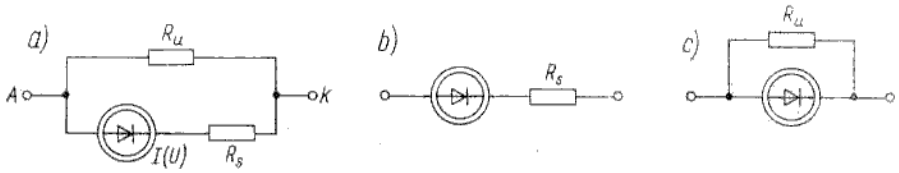


Rys. 3.35
Klasyfikacja schematów zastępczych

Schematy zastępcze nieliniowe

3.1.7.1

Na podstawie rozważań przeprowadzonych w poprzednich punktach tego rozdziału z łatwością można przedstawić schemat zastępczy dla złącza p - n w warunkach pracy statycznej (przy stałych lub bardzo wolno zmieniających się napię-



Rys. 3.36

Schemat zastępczy nieliniowy złącza *p-n* w warunkach pracy statycznej (a) i jego uproszczone warianty dla zakresów przewodzenia (b) oraz zaporowego (c)

ciach i prądach) w postaci pokazanej na rys. 3.36. Elementy R_u , R_s reprezentują rezystancję upływu oraz rezystancję szeregową. Źródło prądowe można zapisać w najbardziej ogólnej postaci:

$$I_F \approx I_R \exp \frac{U}{m\varphi_T} \quad \text{dla kierunku przewodzenia} \quad (3.49a)$$

$$I_R = I_s + I_g(U) \quad \text{dla kierunku zaporowego} \quad (3.49b)$$

$$I = I_R \left(\exp \frac{U}{m\varphi_T} - 1 \right) \quad \text{dla obu kierunków} \quad (3.49c)$$

lub bardziej szczegółowo dla obu kierunków

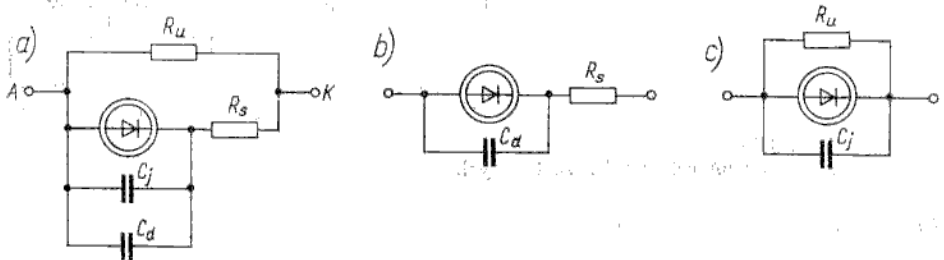
$$I = \underbrace{I_g(U) \left(\exp \frac{U}{2\varphi_T} - 1 \right)}_{\text{prąd rekombinacji}} + \underbrace{I_s \left(\exp \frac{U}{\varphi_T} - 1 \right)}_{\text{prąd dyfuzji}} \quad (3.49d)$$

przy czym: $U = U_z - R_s I$; U_z — napięcie zewnętrzne doprowadzone do złącza.

Wiadomo, że współczynnik m w przypadku złącza krzemowego tylko w zakresie wartości średnich prądu jest równy 1, natomiast dla małych i dużych prądów dąży do 2. Inaczej oznacza to, że w równaniu (3.49d) pierwsza składowa przeważa w zakresie małych napięć (do 0,4 V). Prądy nasycenia I_s oraz generacji I_g są określone zależnościami (3.24), (3.32), przy czym dla złącza krzemowego $I_g \gg I_s$.

Wpływ rezystancji R_s (zwykle rzędu ułamka oma) jest istotny przy pracy w kierunku przewodzenia (rys. 3.25, 3.36b), natomiast rezystancji R_u (rzędu megaoma i więcej) — w kierunku zaporowym (rys. 3.26, 3.36c).

Rozpatrzmy teraz pracę złącza *p-n* w warunkach dynamicznych, tj. przy szybkich zmianach napięcia doprowadzonego do złącza. W tym przypadku oprócz prądu przewodzenia w stanie niestabilnym będzie płynął prąd przesunięcia, związany ze zmianą ładunku magazynowanego w złączu. Wiadomo, że ładunek jest magazy-



Rys. 3.37

Schemat zastępczy nieliniowy złącza *p-n* w warunkach pracy dynamicznej (a) i jego uproszczone warianty dla zakresu przewodzenia (b) oraz zaporowego (c)

nowany w dwu obszarach złącza, tj. w warstwie zaporowej oraz w bazie. W warstwie zaporowej istnieje nieskompensowany dipolowy ładunek zjonizowanych atomów domieszek, w bazie natomiast — ładunek nadmiarowych nośników mniejszościowych. Zmiany obu tych ładunków, spowodowane zmianami napięcia zewnętrznego, można interpretować jako pojemności. Zatem schemat następczy dla pracy dynamicznej złącza p-n można otrzymać na podstawie schematu z rys. 3.36 przez dołączenie równolegle do złącza dwu pojemności (rys. 3.37a). Pojemność C_j , nazywana *pojemnością złączową* lub *pojemnością warstwy zaporowej*, reprezentuje zjawisko zmian ładunku warstwy zaporowej, pojemność C_d nazywana *pojemnością dyfuzyjną*, reprezentuje zjawisko zmian ładunku nadmiarowych nośników mniejszościowych w bazie złącza.

Pojemność warstwy zaporowej

Jeżeli napięcie doprowadzone do złącza zmienia się w funkcji czasu $u = f(t)$, to zmienia się również ładunek warstwy zaporowej, $Q_j = f(t)$. Prąd związany ze zmianą ładunku warstwy zaporowej można zapisać w postaci

$$i = \frac{dQ_j}{dt} = \frac{\partial Q_j}{\partial u} \frac{du}{dt} \quad (3.50)$$

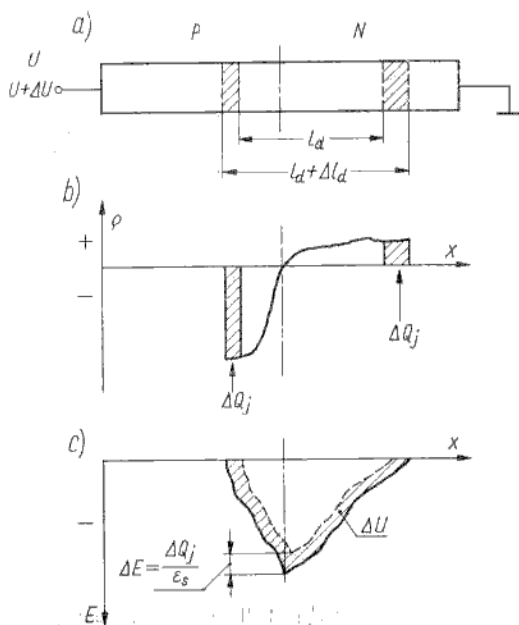
Inaczej

$$i = C_j \frac{du}{dt} \quad (3.50a)$$

przy czym

$$C_j \equiv \frac{\partial Q_j}{\partial u} \quad (3.51)$$

jest definicyjnie określoną *pojemnością złączową*.



Rys. 3.38

Ilustracja zmiany szerokości warstwy zaporowej (a), rozkładu gęstości ładunku (b) oraz natężenia pola (c) w złączu p-n o dowolnym rozkładzie koncentracji domieszek przy zmianie napięcia polaryzacji w kierunku zaporowym o ΔU

W zapisie dla przyrostów skończonych

$$C_j \equiv \frac{\Delta Q_j}{\Delta U} \quad (3.51a)$$

Rozpatrzmy najpierw złącze o dowolnym rozkładzie koncentracji domieszek, czyli o dowolnym rozkładzie gęstości ładunku w warstwie zaporowej. Na rysunku 3.38 przedstawiono szerokość warstwy zaporowej oraz rozkłady gęstości ładunku i natężenia pola elektrycznego dla napięcia U i $(U + \Delta U)$. Zgodnie z prawem Gaussa $\Delta Q_j = \Delta E \varepsilon_s$. Z kolei przy założeniu $\Delta U \ll U$ można przyjąć, że $l_d \approx (l_d + \Delta l_d)$, zatem $\Delta U = \Delta E l_d$. Stąd

$$C_j = \frac{\varepsilon_s}{l_d} \quad \text{dla złącza o powierzchni jednostkowej} \quad (3.52)$$

czyli

$$C_j^* = A \frac{\varepsilon_s}{l_d} \quad (3.52a)$$

przy czym: A — powierzchnia złącza; $\varepsilon_s = \varepsilon_{rs} \varepsilon_0$.

Widać, że pojemność warstwy zaporowej jest po prostu pojemnością kondensatora płaskiego o okładkach oddalonych od siebie na odległość równą szerokości warstwy zaporowej. Należy podkreślić, że *jest to zależność słuszna dla złącza o dowolnym rozkładzie koncentracji domieszek*.

Uwzględniając wcześniej wyprowadzone wyrażenie na szerokość warstwy zaporowej można określić zależność pojemności złączeniowej od napięcia. Dla $|U| \gg |\varphi_B|$ w przybliżeniu

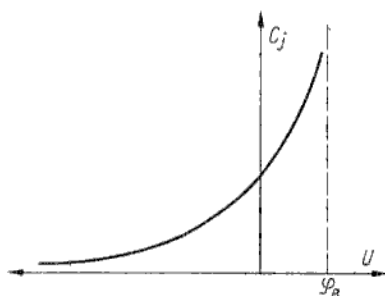
$$C_j \sim \frac{1}{|U|^{1/n}} \quad (3.53)$$

przy czym: $n = 2$ dla złącza skokowego; $n = 3$ dla złącza liniowego.

W przypadku złącza skokowego niesymetrycznego o szerokości warstwy zaporowej decyduje warstwa słabiej domieszkowana (baza). Na przykład dla $N_A \gg N_D$

$$C_j^* \approx A \left[\frac{q \varepsilon_s N_D}{2(\varphi_B - U)} \right]^{1/2} \quad (3.54)$$

Wykres $C_j(U)$ dla złącza skokowego przedstawiono na rys. 3.39.



Rys. 3.39
Zależność $C_j(U)$ dla złącza skokowego

Pojemność dyfuzyjna

Pojemność dyfuzyjna jest znacznie trudniejszym pojęciem niż pojemność złączowa. O ile w warstwie zaporowej występują rozdzielone ładunki nieskompensowane (warstwa dipolowa), co wprost kojarzy się z kondensatorem, o tyle w obszarach neutralnych — jak sama nazwa wskazuje — ładunek nadmiarowych nośników mniejszościowych jest w każdym lokalnym obszarze kompensowany przez przeciwny co do znaku ładunek nadmiarowych nośników większościowych. Jednakże należy zdać sobie sprawę, że każda zmiana ładunku nadmiarowego w bazie złącza wymaga przepływu prądu ładowania, tj. dostarczenia dziur i elektronów w jednakowych ilościach z dwu przeciwnych kierunków. Zatem prąd w stanie nieustalonym można wyrazić jako sumę prądu ustalonego i prądu ładowania bazy

$$i = I_{ust} + \frac{dQ_b}{dt} \quad (3.55)$$

Obie składowe prądu można przedstawić w jednorodnej postaci, stanowiącej sformułowanie tzw. modelu ładunkowego złącza p-n, jeśli przeprowadzi się następujące rozumowanie. Załóżmy, że płynie prąd ustalony, któremu w bazie odpowiada ładunek nadmiarowy Q_b . Jeżeli teraz dokonamy rozwarcia w obwodzie, tj. uczynimy $I_{ust} = 0$, to ładunek Q_b będzie się zmniejszał w funkcji czasu wyłącznie wskutek procesu rekombinacji, tj. z szybkością Q_b/τ (τ — czas życia nośników mniejszościowych). A zatem warunkiem zachowania stanu ustalonego w złączu polaryzowanym jest dostarczanie (wstrzykiwanie) nośników z taką samą szybkością, z jaką zachodzi rekombinacja. Stąd

$$I_{ust} = \frac{Q_b}{\tau}$$

czyli

$$i = \frac{Q_b}{\tau} + \frac{dQ_b}{dt} \quad (3.56)$$

Analogicznie jak w rozważaniach dotyczących pojemności złączowej, zapisujemy prąd ładowania w postaci

$$\begin{aligned} i_t &= \frac{dQ_b}{dt} = \\ &= \frac{\partial Q_b}{\partial u} \frac{du}{dt} = \\ &= C_d \frac{du}{dt} \end{aligned} \quad (3.57)$$

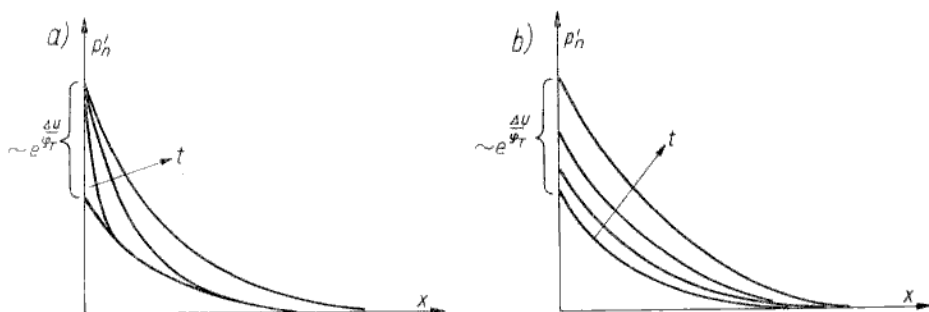
przy czym

$$C_d \equiv \frac{\partial Q_b}{\partial u} \quad (3.58)$$

jest definicyjnie określoną pojemnością dyfuzyjną. W zapisie dla przyrostów skończonych

$$C_d \equiv \frac{\Delta Q_b}{\Delta U} \quad (3.58a)$$

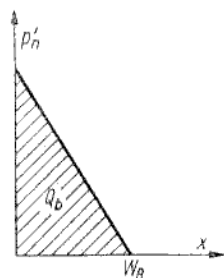
Wyznaczenie pojemności dyfuzyjnej na podstawie definicji (3.58) dla dowolnie szybkich procesów niestabilnych jest zadaniem bardzo trudnym, gdyż ładunek Q_b zależy nie tylko od napięcia, lecz również od czasu. Dla lepszej ilustracji tego zagadnienia na rys. 3.40 przedstawiono rozkłady ładunku nadmiarowego w kolejnych



Rys. 3.40
Rozkłady koncentracji nośników nadmiarowych w bazie złącza spolaryzowanego w kierunku przewodzenia przy raptownej zmianie napięcia (a) oraz przy powolnej (quasi-statycznej) zmianie napięcia (b)

chwilach po raptownym wzroście napięcia oraz przy powolnie wzrastającym napięciu. Tylko dla tego drugiego przypadku można znaleźć proste wyrażenie na pojemność dyfuzyjną. Będzie to rozwiązanie słuszne dla procesu quasi-ustalonego, tj. charakteryzującego się nadążaniem rozkładu ładunku za bieżącą wartością zmieniającego się napięcia. Taki przypadek istnieje wówczas, gdy napięcie zmienia się w czasie dłuższym niż czas życia nośników mniejszościowych, co pozwala przyjąć, że ładunek Q_b jest wyłącznie funkcją napięcia (lub prądu) ustalonego.

Rozpatrzmy najpierw prostszy przypadek złącza z krótką bazą (pojęcie takiego złącza wyjaśniono w p. 3.1.4.3), co umożliwi pominięcie wpływu zjawisk generacji-rekombinacji na proces powstawania przyrostu ładunku w bazie.



Rys. 3.41
Ładunek nośników nadmiarowych w złączu z krótką bazą

Zgodnie z wykresem przedstawionym na rys. 3.41 ładunek Q_b można obliczyć z pola trójkąta

$$Q_b(U) = qp'_n(0, U) \frac{W_B}{2} \quad (3.59)$$

przy czym: $p'_n(0, U)$ — koncentracja nośników nadmiarowych dla współrzędnej $x = 0$ przy napięciu U .

Wiadomo, że

$$p'_n(0, U) = p_{n0} \left(\exp \frac{U}{\varphi_T} - 1 \right) \quad (3.60)$$

Po podstawieniu (3.60) do (3.59) i wykonaniu różniczkowania względem napięcia otrzymuje się

$$C_d = \frac{q}{\varphi_T} p_{n0} \left(\exp \frac{U}{\varphi_T} \right) \frac{W_B}{2} \quad (3.61)$$

Uwzględniając równanie charakterystyki prądowo-napięciowej (3.45), wzór (3.61) można przepisać w postaci

$$C_d = \frac{1}{2\varphi_T} (J + J_s) \tau \quad (3.62)$$

We wzorze (3.61) wystarczy zamienić szerokość bazy W_B na drogę dyfuzyjną L_D , by otrzymać identyczną postać (3.62) dla złącza z długą bazą. Słuszność takiego postępowania nie jest bynajmniej oczywista. Pozornie wydaje się, że w tym przypadku jest słuszne następujące postępowanie (rys. 3.42). Ładunek

$$Q_b = qp'_n(0, U) \int_0^{\infty} e^{-x/L_D} dx$$

Stąd uwzględniając (3.60)

$$Q_b = qp_{n0} \left(\exp \frac{U}{\varphi_T} - 1 \right) L_D$$

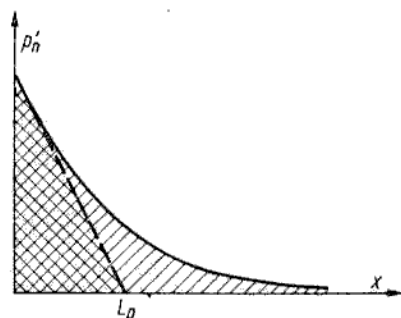
i po zróżniczkowaniu względem napięcia

$$C_d = \frac{q}{\varphi_T} p_{n0} L_D \exp^{U/\varphi_T}$$

czyli

$$C_d = \frac{1}{\varphi_T} (J + J_s) \tau$$

Jednak takie wyrażenie — często podawane w literaturze — jest błędne, gdyż opiera się na fałszywym założeniu, że cały przyrost ładunku ΔQ_b jest związany z przepły-



Rys. 3.42

Interpretacja graficzna istoty fizycznej pojemności dyfuzyjnej w złączu z długą bazą
 ▨ — całkowity ładunek Q_b ; ▩ — część ładunku Q_b występującego w pojemności dyfuzyjnej

wem prądu ładowania. A przecież, jeżeli na przykład napięcie zmaleje o wartość ΔU , to nadmiar ładunku ΔQ_b znika częściowo wskutek usunięcia dziur (z powrotem do obszaru typu p) i elektronów (z bazy typu n do obwodu zewnętrznego) oraz częściowo wskutek rekombinacji „na miejscu” dziur z elektronami. Ten drugi proces nie jest związany z przepływem ładunku w obwodzie zewnętrznym. Można wykazać, że mniej więcej połowa ładunku ΔQ_b jest usuwana wskutek przepływu pojemnościowego prądu rozładowania.

Słuszny jest zatem wzór (3.62), który można łatwo wyprowadzić na podstawie rozkładu prostoliniowego, przedstawionego linią przerywaną na rys. 3.42. Ponieważ pojemność dyfuzyjna jest spowodowana wstrzykiwaniem nośników nadmiarowych, czyli ma sens tylko przy pracy złącza w kierunku przewodzenia, a wówczas zwykle $J \gg J_s$, zatem:

$$C_d \approx \frac{1}{2\varphi_T} J\tau; \quad C_d^* = \frac{1}{2\varphi_T} I\tau \quad (3.63)$$

Do obliczenia pojemności dyfuzyjnej nie jest wymagana znajomość powierzchni złącza. Jednakże błędne byłoby stwierdzenie, że pojemność dyfuzyjna nie zależy od powierzchni złącza, gdyż $I = AJ$ (A — powierzchnia złącza, J — gęstość prądu).

Jeszcze raz należy podkreślić, że schemat zastępczy, przedstawiony na rys. 3.37, przy pojemności dyfuzyjnej, wyrażonej wzorem (3.62), nadaje się do analizy pracy dynamicznej złącza $p-n$ tylko przy niewielkich szybkościach zmian napięcia.

Przebiegi prądów i napięć w złączu $p-n$ pracującym impulsowo (przy wymuszeniach zmiennych skokowo) rozpatrzymy oddzielnie w p. 3.1.7.3.

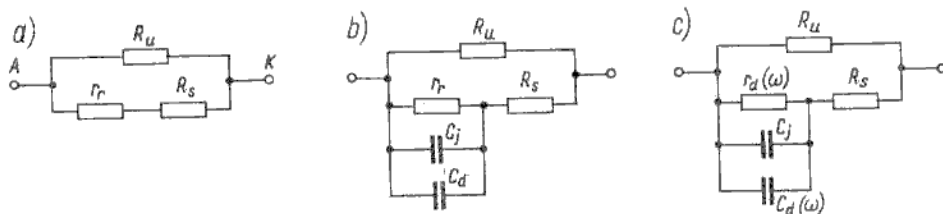
Schematy zastępcze liniowe

3 1.7.2

Jeżeli do złącza $p-n$ spolaryzowanego napięciem stałym jest doprowadzony sygnał napięcia zmiennego o małej amplitudzie, to można uważać, że składowe zmienne prądu i napięcia są związane stałymi parametrami, czyli są to zależności liniowe. Za kryterium małego sygnału przyjmuje się na ogół warunek, by amplituda sygnału była mniejsza niż kT/q , co dla temperatury pokojowej oznacza $U_m < 26$ mV, przy czym

$$u(t) = U_{st} + U_m e^{i\omega t}$$

Na rysunku 3.43 przedstawiono trzy schematy zastępcze liniowe dla małosygnalowej pracy złącza $p-n$ w różnych zakresach częstotliwości.

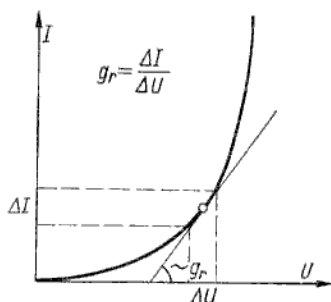


Rys. 3.43

Schematy zastępcze liniowe dla złącza $p-n$ pracującego przy małym sygnale dla częstotliwości małej (a), średniej (b) i dużej (c)

Schemat zastępczy dla przebiegów quasi-statycznych (dla małej częstotliwości)

Przy bardzo wolnych zmianach napięcia ($\omega \approx 0$) można pominąć prądy ładowania pojemności. Wówczas w schemacie zastępczym (rys. 3.43a) wystarczy uwzględnić konduktancję (rezystancję) różniczkową, określoną jako nachylenie



Rys. 3.44
Interpretacja graficzna konduktancji różniczkowej

charakterystyki $I(U)$ w danym punkcie pracy (rys. 3.44). Konduktancja różniczkowa jest definiowana w następujący sposób:

$$g_r = \frac{1}{r_r} = \frac{dI}{dU} \quad (3.64)$$

Ponieważ

$$I = I_R \left(\exp \frac{U}{m\varphi_T} - 1 \right)$$

zatem

$$g_r = \frac{I + I_R}{m\varphi_T} \quad (3.65)$$

Dla pracy w kierunku przewodzenia, praktycznie biorąc, zawsze $I \gg I_R$, zatem

$$g_r \approx \frac{I}{m\varphi_T} \quad (3.65a)$$

W temperaturze pokojowej dla $m = 1$

$$g_r = \frac{I}{26 \text{ mV}}$$

Schemat zastępczy dla średniej częstotliwości

W miarę wzrostu częstotliwości sygnału zmiennego staje się konieczne uwzględnienie pojemności warstwy zaporowej i pojemności dyfuzyjnej (rys. 3.43b). Są to pojemności różniczkowe, omawiane w p. 3.1.7.1. O ile pojemność warstwy zaporowej nie zależy od częstotliwości w bardzo szerokim zakresie (do częstotliwości $10^{11} \dots 10^{13}$ Hz), o tyle pojemność dyfuzyjna i konduktancja dynamiczna mogą być uważane za niezależne od częstotliwości tylko wówczas, gdy $\omega \ll 1/\tau$ (τ — czas życia nośników mniejszościowych, który najczęściej może przyjmować wartości w zakresie $10^{-6} \dots 10^{-9}$ s). W tym zakresie częstotliwości pojemność C_d

i konduktancja g_r są określone wzorami (3.63), (3.65a). Można zauważyć, że kompilacja tych dwóch wzorów daje zależność

$$C_d = \frac{1}{2} g_r \tau \quad (3.66)$$

a więc stała czasowa obwodu C_d , g_r odpowiada połowie czasu życia nośników.

Schemat zastępczy dla dużych częstotliwości

W przypadku $\omega \geq 0,1/\tau$ konduktancję różniczkową oraz pojemność dyfuzyjną różniczkową należy zamienić na konduktancję dynamiczną i pojemność dyfuzyjną dynamiczną, czyli na parametry będące funkcjami częstotliwości (rys. 3.43c). Z rozwiązania równania ciągłości ładunku dla przebiegów zmiennych można otrzymać przybliżone zależności:

$$g_d(\omega) = g_d(0) \sqrt{\frac{\omega\tau}{2}} \quad (3.67)$$

$$C_d(\omega) = C_d(0) \sqrt{\frac{2}{\omega\tau}}$$

przy czym: $g_d(0)$, $C_d(0)$ — parametry dla $\omega \rightarrow 0$; $g_d(0) = g_r$.

Jeżeli zapiszemy admitancję

$$Y = g_d + j\omega C_d$$

to można łatwo zauważyć ciekawą właściwość, że składowe rzeczywista i urojona tej admitancji wzrastają proporcjonalnie do $\sqrt{\omega}$.

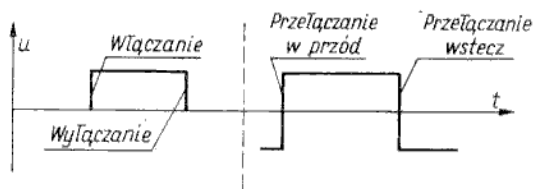
Praca impulsowa złącza p-n

3.1.7.3

W punkcie 3.1.7.1 rozpatrywaliśmy schemat zastępczy złącza p-n dla dynamicznej pracy nieliniowej, przy czym zwróciliśmy uwagę na ograniczoną przydatność tego schematu zastępczego do sygnałów zmieniających się w czasie znacznie dłuższym niż czas życia nośników mniejszościowych (w zależności od rodzaju złącza mogą to być czasy od nanosekund do mikrosekund). To ograniczenie było spowodowane trudnością odpowiedniego zapisu pojemności dyfuzyjnej jako funkcji napięcia i czasu. Dlatego do opisu pracy impulsowej złącza p-n, czyli do opisu reakcji złącza na raptowną zmianę warunków polaryzacji, trzeba stosować inny model. W tym przypadku najwłaściwszy jest model analityczny w postaci uwikłanej. Takim modelem jest równanie ciągłości (p. 1.8), w którym uwzględnia się składową prądu dyfuzji oraz składnik opisujący rekombinację nośników. Jest to równanie różniczkowe drugiego stopnia o pochodnych cząstkowych, którego rozwiązanie wymagałoby użycie aparatu matematycznego, wykraczającego poza ramy przyjęte w tej książce. Dlatego ograniczymy się do jakościowej interpretacji zjawisk fizycznych oraz przytoczenia kilku wzorów przydatnych w przybliżonych obliczeniach inżynierskich.

Procesy przejściowe w złączu p-n zależą od warunków wysterowania. W związku z tym można wyróżnić następujące warianty wysterowania (rys. 3.45):

- włączanie (przejście ze stanu neutralnego, tj. stanu bez polaryzacji, do stanu przewodzenia);
- wyłączenie (przejście ze stanu przewodzenia do stanu neutralnego);



Rys. 3.45
Cztery wariantyysterowania złącza p-n przy pracy impulsowej

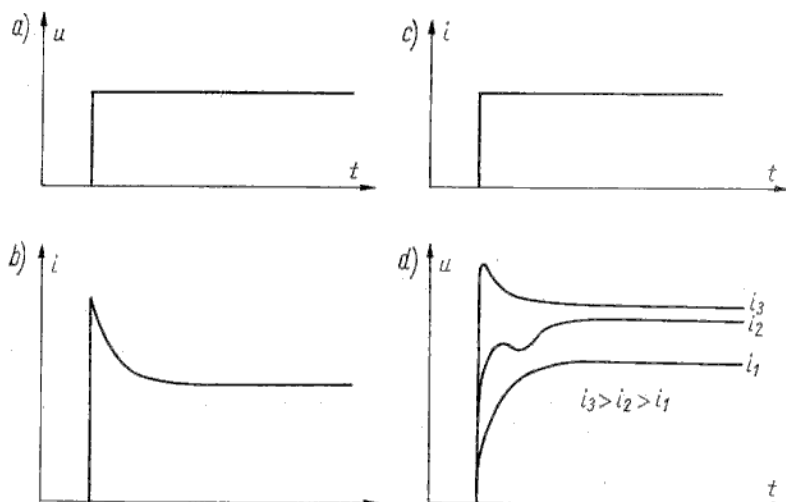
- przełączanie (przejście ze stanu zaporowego do stanu przewodzenia i odwrotnie — ze stanu przewodzenia do stanu zaporowego).

Istotny jest również rodzaj źródła sterującego, przy czym można wyróżnić:

- sterowanie ze źródła napięciowego (o stałej wydajności napięciowej),
- sterowanie ze źródła prądowego (o stałej wydajności prądowej).

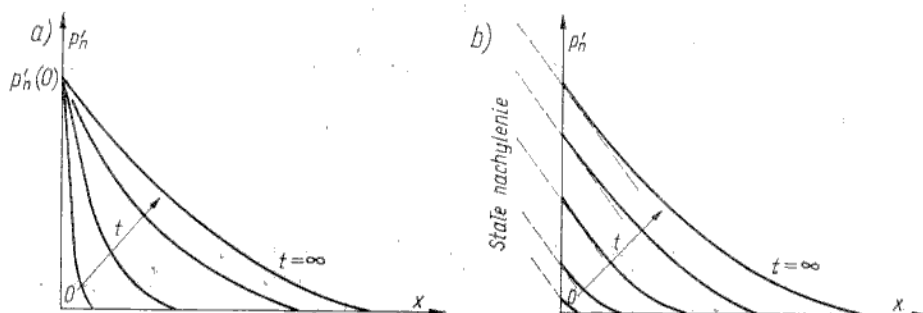
Włączanie

Na rysunku 3.46 przedstawiono przebiegi napięć i prądów przy włączaniu napięciowym lub prądowym złącza p-n. Przebieg prądu $i(t)$ dla przypadku sterowania



Rys. 3.46

Przebiegi napięć i prądów przy włączaniu napięciowym (a, b) lub prądowym (c, d) złącza p-n



Rys. 3.47

Rozkłady koncentracji nośników mniejszościowych nadmiarowych w bazie złącza p-n podczas włączania napięciowego (a) oraz prądowego (b)

napięciowego (rys. 3.46b) wyjaśniają rozkłady koncentracji nośników nadmiarowych w bazie złącza p - n , przedstawione na rys. 3.47a. W chwili doprowadzenia napięcia U , praktycznie biorąc, natychmiast zmniejsza się bariera od wartości φ_B do $(\varphi_B - U)$, wskutek czego na granicy warstwa zaporowa-baza pojawia się dodatkowa koncentracja nośników mniejszościowych (w przypadku bazy typu n są to dziury wstrzykiwane z sąsiadującego obszaru typu p)

$$p'_n(0) = p_n(0) - p_{n0} = p_{n0} \left(\exp \frac{U}{\varphi_T} - 1 \right) \quad (3.68)$$

W pozostałym obszarze bazy koncentracja nośników nadmiarowych w chwili początkowej jest równa zeru, powstaje więc bardzo duży gradient koncentracji tych nośników, powodujący przepływ dużego prądu dyfuzji. Ponieważ prąd płynący przez złącze p - n jest równy prądowi dyfuzji dla współrzędnej $x = 0$, a ten z kolei jest wprost proporcjonalny do gradientu koncentracji nośników nadmiarowych

$$J = qD_p \left. \frac{dp'_n}{dx} \right|_{x=0} \quad (3.69)$$

zatem teoretycznie w chwili $t = 0$ prąd dąży do nieskończoności. W przypadku występującym w praktyce nie ma idealnie prostokątnego skoku napięcia sterującego (choćby wskutek istnienia obwodu R, C utworzonego przez rezystancję szeregową złącza i pojemność warstwy zaporowej), dlatego skok prądu ma wartość skończoną. Zmniejszanie się wartości prądu w funkcji czasu jest związane ze zmniejszaniem się nachylenia rozkładu koncentracji nośników nadmiarowych w miarę ich dyfuzji w głąb bazy, co dla kolejnych chwil czasu przedstawiają wykresy na rys. 3.47a. Po pewnym czasie rozkład koncentracji nośników nadmiarowych, a więc i prąd, ustalają się wskutek całkowitego zrównania się szybkości dostarczania nośników i szybkości ich rekombinacji w każdym wycinku obszaru bazy.

Większe znaczenie praktyczne ma przypadek sterowania złącza p - n ze źródła prądowego (w rzeczywistych układach rezystancja wewnętrzna źródła sterującego jest zwykle większa niż rezystancja diody pracującej w kierunku przewodzenia). Zostanie teraz rozpatrzony ten przypadek reprezentowany przez przebiegi $u(t)$ przedstawione na rys. 3.46d oraz rozkłady koncentracji nośników nadmiarowych, wykreślone na rys. 3.47b.

Stałość prądu oznacza, zgodnie z wyrażeniem (3.69), stały gradient koncentracji nośników nadmiarowych dla współrzędnej $x = 0$. Zatem wykresy $p'_n(x)$ dla kolejnych chwil czasu przesuwają się „w górę” przy zachowaniu stałego nachylenia przy $x = 0$. Wówczas wzrasta koncentracja $p'_n(0)$, a ponieważ na podstawie (3.68)

$$\begin{aligned} u &= \varphi_T \ln \frac{p'_n(0) + p_{n0}}{p_{n0}} = \\ &= \varphi_T \ln \frac{p_n(0)}{p_{n0}} \end{aligned} \quad (3.70)$$

zatem zwiększa się odpowiednio napięcie na złączu p - n .

Napięcie można wyrazić w funkcji czasu, jeśli jest znana zależność $p'_n(0, t)$. Na podstawie rozwiązania równania ciągłości otrzymuje się

$$p'_n(0, t) = p_{n0} \frac{J_F}{J_R} \operatorname{erf} \sqrt{t} \quad (3.71)$$

Po podstawieniu (3.71) do (3.70) uzyskuje się

$$u = \varphi_T \ln \left(1 + \frac{J_F}{J_R} \operatorname{erf} \sqrt{\bar{t}} \right) \quad (3.72)$$

przy czym: $\bar{t} = t/\tau_p$; τ_p — czas życia³ dziur (nośników mniejszościowych); funkcja $\operatorname{erf} x$ — była omawiana w rozdz. 2; w tym miejscu przypomnijmy tylko, że kształt tej funkcji jest podobny jak dla $(1 - e^{-x})$, przy czym $\operatorname{erf} x = 0$ dla $x = 0$ oraz $\operatorname{erf} x = 1$ dla $x = \infty$.

Przebieg określony zależnością (3.72) otrzymuje się doświadczalnie dla małych poziomów wstrzykiwania (wykres dla i_1 na rys. 3.46d). W tym przypadku mówi się, że złącze reaguje pojemnościowo. Przy większych poziomach wstrzykiwania (większe wartości prądu I_F) przebieg $u(t)$ przyjmuje inny kształt. Jest to spowodowane zjawiskiem zmian wartości rezystancji szeregowej R_s w funkcji czasu. Rezystancja szeregowa maleje w funkcji czasu wskutek³ gromadzenia się nośników nadmiarowych w bazie. Pamiętamy, że baza zachowuje obojętność elektryczną w każdym jej obszarze. Dlatego w obszarze bazy sąsiadującym z warstwą zaporową (w pobliżu $x = 0$) jest zwiększona koncentracja zarówno dziur, jak i elektronów, czyli jest³ to obszar o zwiększonej konduktywności. A ponieważ ładunek nośników nadmiarowych wzrasta w funkcji czasu (rys. 3.47b), konduktywność obszaru zajmowanego przez ten ładunek również wzrasta, a więc rezystancja bazy, czyli rezystancja szeregowa złącza, maleje w funkcji czasu.

Przy uwzględnieniu spadku napięcia na rezystancji szeregowej napięcie zewnętrzne na złączu p-n

$$\begin{aligned} u_z &= u + u_b \\ u_z(t) &= u(t) + I_F R_s(t) \end{aligned} \quad (3.73)$$

przy czym u_b — napięcie na rezystancji szeregowej (na bazie).

Zależność $u(t)$ jest dana wzorem (3.72); rezystancję $R_s(t)$ można wyznaczyć przez scałkowanie w całym obszarze bazy rezystywności, zależnej od współrzędnej x i od czasu. W rezultacie otrzymuje się spadek napięcia na bazie

$$\begin{aligned} u_b(t) &= I_F R_s(t) = \\ &= I_F R_s(0) \left[1 - \frac{L_p}{|W_B|} \operatorname{erf} \sqrt{\bar{t}} \ln \left(1 + \frac{J_F}{J_R} \right) \right] \end{aligned} \quad (3.74)$$

Ze wzoru (3.74) widać, że w chwili $t = 0$ napięcie u_b osiąga wartość maksymalną, po czym maleje według zależności zbliżonej do wykładniczej osiągając w stanie ustalonym wartość

$$u_b(\infty) = I_F R_s(0) \left[1 - \frac{L_p}{W_B} \ln \left(1 + \frac{J_F}{J_R} \right) \right] \quad (3.75)$$

Zjawisko modulacji rezystancji bazy ma dominujący wpływ przy dużym poziomie wstrzykiwania, dlatego przebieg dla i_3 na rys. 3.46d jest zgodny z zależnością (3.74).

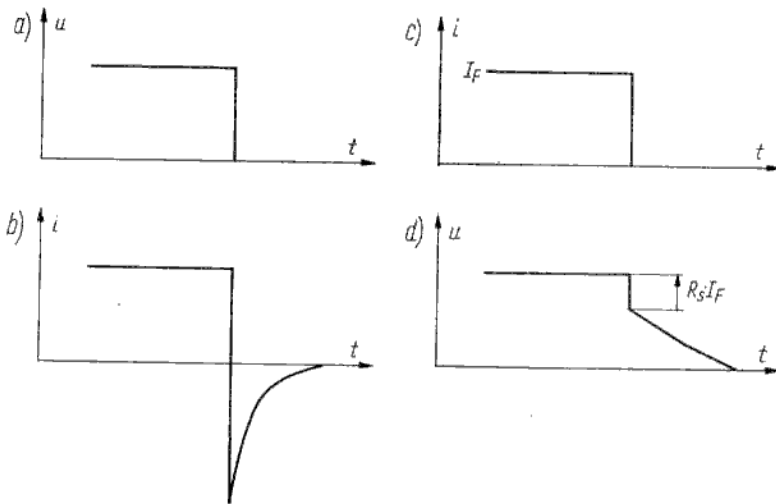
W tym przypadku mówi się, że złącze ma właściwości indukcyjne.

Po podstawieniu (3.74), (3.72) do (3.73) otrzymuje się zależność $u_z(t)$ słuszną dla dowolnych wartości prądu. Wskutek przeciwnych kierunków zmian $u(t)$ oraz $u_b(t)$ w zakresie wartości średnich prądu można obserwować przebieg $u_z(t)$ przedstawiony dla i_2 na rys. 3.46d.

Dla celów praktycznych ważne jest oszacowanie czasu włączania. Ponieważ funkcja $\operatorname{erf} x = 0,85$ dla $x = 1$, zatem, mimo dość złożonych zależności na $u(t)$, można wysnuć prosty wniosek, że *czas włączania jest w przybliżeniu równy czasowi życia nośników mniejszościowych*.

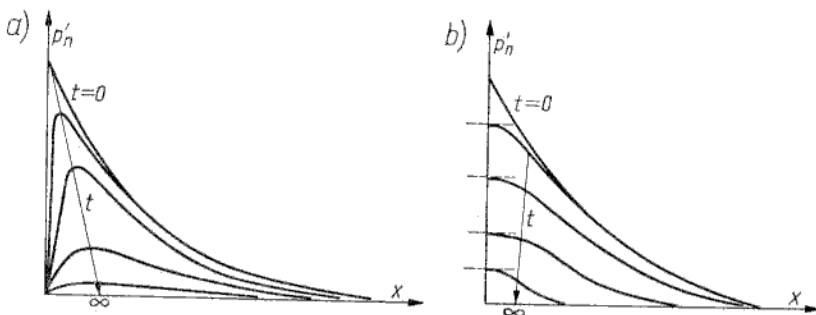
Wyłączenie

Wyłączenie — rozumiane jako przejście ze stanu przewodzenia do stanu bez polaryzacji zewnętrznej — może mieć charakter napięciowy (zwarcie złącza) lub prądowy (rozwarcie złącza). Odpowiednie przebiegi prądów i napięć przedstawiono na rys. 3.48.

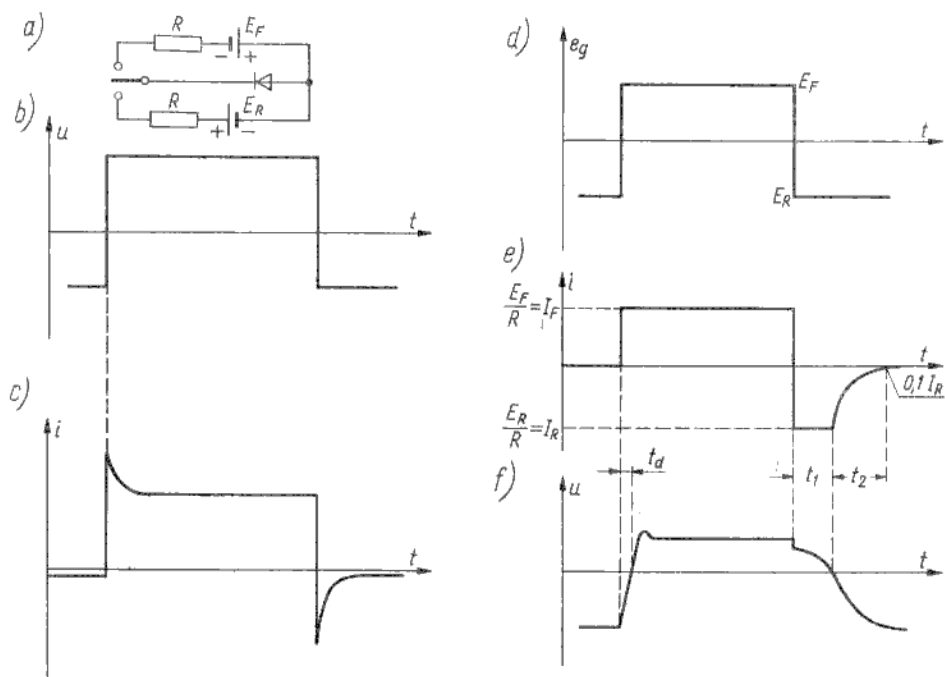


Rys. 3.48
Przebiegi napięć i prądów przy wyłączeniu napięciowym (a, b) lub prądowym (c, d) złącza p-n

Przebieg prądu $i(t)$ dla przypadku wyłączenia napięciowego (rys. 3.48b) wyjaśniają rozkłady koncentracji nośników nadmiarowych w bazie złącza p-n, przedstawione na rys. 3.49a.



Rys. 3.49
Rozkłady koncentracji nośników mniejszościowych nadmiarowych w bazie złącza p-n podczas wyłączenia napięciowego (a) oraz prądowego (b)



Rys. 3.50

Schemat układu, w którym następuje przełączanie złącza (a), oraz przebiegi napięć i prądów podczas przełączania napięciowego (b, c) lub prądowego (d, e, f)

i w czasie przepływu dużego prądu wstecznego; przy pracy w ustalonym stanie zaporowym prąd płynący przez złącze nie jest ograniczony obwodem zewnętrznym. Na rysunku 3.50a pokazano prosty układ, w którym prąd w kierunkach przewodzenia i zaporowym nie może przekroczyć wartości E_F/R , E_R/R (przebieg $i(t)$ na rys. 3.50e).

Podczas przełączania w przód (z kierunku zaporowego do przewodzenia) oprócz zjawisk omówionych dla włączania należy ponadto uwzględnić proces przeładowania pojemności warstwy zaporowej. Wskutek tego zjawiska napięcie na złączu nie osiąga wartości zerowej natychmiast, lecz po upływie czasu t_d , nazywanego *czasem opóźnienia*. Czas opóźnienia jest określony stałą czasową obwodu R, C utworzonego przez rezystancję obwodu zewnętrznego R i pojemność bariery C_j . Obliczanie czasu t_d komplikuje się z uwagi na nieliniową zależność pojemności C_j od napięcia. Można jednak w przybliżeniu przyjąć, że:

$$t_d = 1,6 C_{j0} R \quad \text{dla złącza skokowego}$$

$$t_d = 1,125 C_{j0} R \quad \text{dla złącza liniowego}$$

przy czym C_{j0} — pojemność warstwy zaporowej przy $U = 0$.

Zjawiska występujące podczas przełączania w tył (z kierunku przewodzenia do zaporowego) wyjaśniają rozkłady koncentracji nośników nadmiarowych w bazie złącza $p-n$, przedstawione na rys. 3.51. Od chwili przełączenia ($t = 0$) do chwili t_1 kolejne wykresy rozkładu koncentracji nośników mają stałe nachylenie przy $x = 0$,

Należy nadmienić, że wszystkie wyrażenia analityczne, przedstawione w p. 3.1.7.3, dotyczą bezpośrednio złącza z długą bazą, lecz można je również odnieść do złącza z krótką bazą, jeżeli zamieni się τ_p na W_B^2/D_p (w złączu z długą bazą $\tau_p = L_p^2/D_p$).

Wpływ temperatury na charakterystykę prądowo-napięciową złącza *p-n*

3.1.8

Na podstawie zależności koncentracji nośników oraz konduktywności od temperatury wiadomo, że przyrządy półprzewodnikowe mogą pracować w zakresie tzw. średnich temperatur, w którym — praktycznie biorąc — wszystkie atomy domieszek są już zjonizowane, a zarazem generacja par elektron-dziura („przeskok” elektronów z pasma walencyjnego do pasma przewodnictwa) jest jeszcze mało prawdopodobna. Rozpatrzmy pracę złącza *p-n* w tym przedziale temperatury dla trzech charakterystycznych zakresów charakterystyk prądowo-napięciowych, tj. dla:

- zakresu zaporowego (przed przebicciem),
- zakresu przebiccia,
- zakresu przewodzenia.

Wiadomo, że w zakresie zaporowym przez złącze *p-n* płynie prąd

$$J_R = J_s + J_g$$

przy czym w złączu germanowym przeważa składowa J_s , w złączu krzemowym — składowa J_g .

Zgodnie ze wzorem (3.24a) składowa J_s zależy od temperatury wskutek zależności $n_i^2(T)$, $D(T)$ oraz $\tau(T)$. Zmiany współczynnika dyfuzji oraz czasu życia w funkcji temperatury można uznać za pomijalnie małe w porównaniu z silną zależnością n_i^2 od temperatury. Na podstawie wyrażenia (1.13)

$$n_i^2 \sim T^3 \exp\left(-\frac{W_g}{kT}\right)$$

W rozpatrywanym zakresie temperatury (± 70 K w pobliżu temperatury pokojowej) czynnik T^3 w porównaniu z funkcją wykładniczą ma znikomo mały wpływ na przebieg zależności $n_i^2(T)$. Dlatego można przyjąć z wystarczającą dokładnością, że

$$n_i^2 \sim \exp\left(-\frac{W_g}{kT}\right)$$

Zatem

$$J_s \sim \exp\left(-\frac{W_g}{kT}\right) \quad (3.80)$$

W przypadku złącza krzemowego interesuje nas składowa J_g , która jest proporcjonalna do n_i [wzór (3.32)]. Stąd, przy podobnych uproszczeniach jak w analizie zależności $J_s(T)$, można napisać

$$J_g \sim \exp\left(-\frac{W_g}{2kT}\right) \quad (3.81)$$

Na podstawie wyrażeń (3.80), (3.81) można wyznaczyć współczynnik temperaturowy prądu wstecznego:

dla złącza germanowego ($J_R \approx J_s$)

$$\frac{1}{J_s} \frac{dJ_s}{dT} = \frac{W_g}{kT^2} \quad (3.82)$$

dla złącza krzemowego ($J_R \approx J_g$)

$$\frac{1}{J_g} \frac{dJ_g}{dT} = \frac{W_g}{2kT^2} \quad (3.83)$$

Współczynnik temperaturowy prądu wstecznego, wyznaczony ze wzorów (3.82), (3.83), dla temperatury pokojowej wynosi:

ok. 9% dla złącza germanowego,

ok. 7% dla złącza krzemowego.

Do szacunkowych obliczeń inżynierskich można przyjąć, że zarówno w złączu germanowym, jak i krzemowym prąd wsteczny zwiększa się ok. 2-krotnie przy wzroście temperatury o 10°C.

W zakresie przebiecia zmiany napięcia w funkcji temperatury można zapisać w postaci zależności liniowej

$$U_p = U_p(0) [1 + \beta(T - T_0)] \quad (3.84)$$

przy czym: $U_p(0)$ — napięcie przebiecia w temperaturze T_0 ;

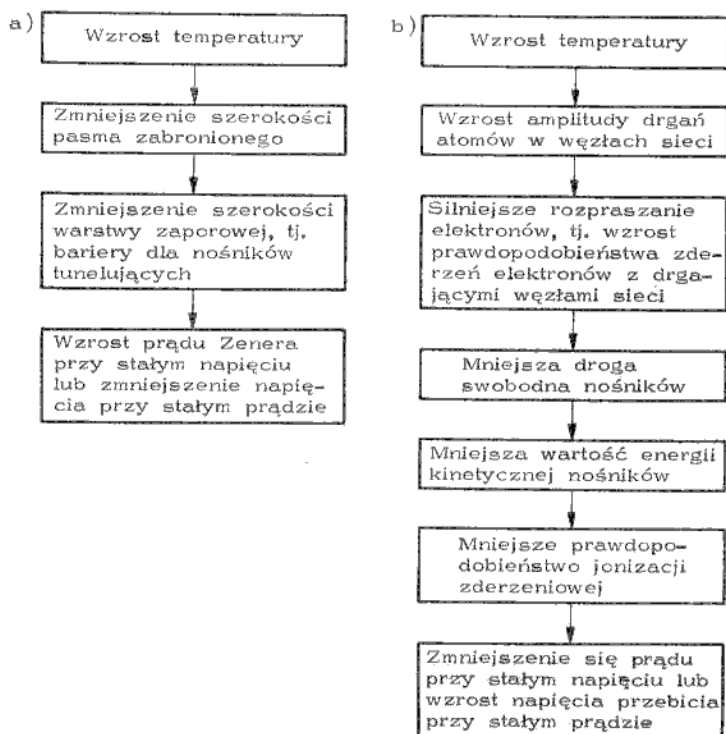
β — temperaturowy współczynnik napięcia przebiecia.

Współczynnik β przyjmuje wartości ujemne, gdy przebiecie jest wywołane zjawiskiem Zenera, lub dodatnie, gdy przebiecie jest wywołane zjawiskiem lawinowym. Mechanizm zjawisk powodujących zmiany temperaturowe napięcia przebiecia jest następujący.

W przypadku przebiecia Zenera wzrost temperatury powoduje nieznaczne zmniejszenie szerokości pasma zabronionego¹⁾. Zmniejsza się wówczas szerokość warstwy zaporowej, a więc również wysokość i szerokość bariery dla nośników tunelujących z pasma walencyjnego do pasma przewodnictwa (rys. 3.28). Z fizyki kwantowej wiadomo, że prawdopodobieństwo przenikania tunelowego cząstki przez barierę energetyczną zależy wykładniczo od szerokości i pierwiastka z wysokości tej bariery. Dlatego nawet nieznaczne zmniejszenie szerokości i wysokości bariery powoduje istotny wzrost prądu tunelowego. Zatem wzrost temperatury powoduje zwiększenie prądu Zenera przy stałym napięciu polaryzacji, a więc zmniejszenie napięcia przebiecia przy stałym prądzie (rys. 3.52a).

W przypadku przebiecia lawinowego wzrost temperatury powoduje zwiększenie amplitudy drgań atomów w węzłach sieci krystalicznej. Atom drgający z większą amplitudą zajmuje efektywnie większy obszar przestrzeni (wzrasta tzw. przekrój skuteczny). Wzrasta zatem prawdopodobieństwo zderzenia się elektronu (dziury) z drgającym atomem, czyli nośniki są silniej rozpraszane na fononach. Częstsze zderzenia to krótsza droga swobodna. Jeśli przyjąć, że ruch elektronu (dziury) na drodze swobodnej jest jednostajnie przyspieszony, to na mniejszej drodze nośnik uzyskuje mniejszą prędkość końcową, czyli ma mniejszą wartość energii kinetycznej w chwili zderzenia. Maleje zatem prawdopodobieństwo

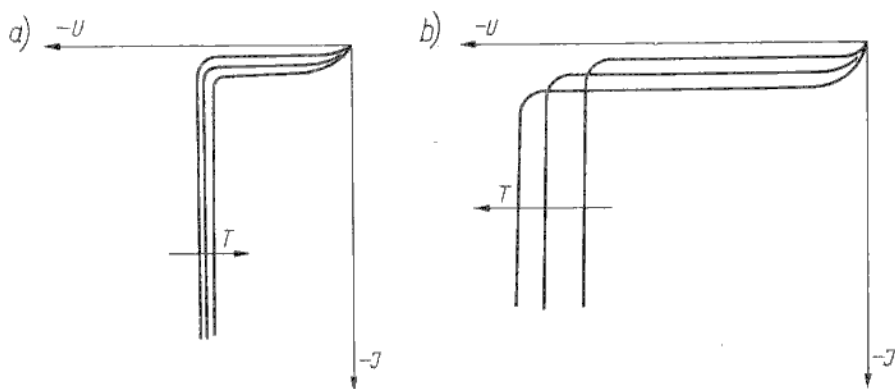
¹⁾ Na przykład w krzemie $W_g = 1,15$ eV w temperaturze 0 K oraz $W_g = 1,1$ eV w temperaturze 300 K.



Rys. 3.52

Mechanizm wpływu temperatury na przebicie Zenera (a) oraz przebicie lawinowe (b)

jonizacji zderzeniowej, co oznacza osłabienie powielania lawinowego. Zatem przy stałym napięciu maleje prąd lawinowy, a więc przy stałym prądzie wzrasta wartość napięcia przebicia lawinowego (rys. 3.52b). Wpływ temperatury na charak-



Rys. 3.53

Charakterystyki prądowo-napięciowe złącza $p-n$ w zakresie zaporowym (wraz z obszarem przebicia) dla różnych temperatur: a) przebicie Zenera, b) przebicie lawinowe

terystyki prądowo-napięciowe złącza p-n w zakresie zaporowym (wraz z obszarem przebiecia) przedstawiono na rys. 3.53.

Jaki jest wpływ temperatury na charakterystykę prądowo-napięciową w zakresie przewodzenia? Wiadomo, że charakterystykę $J(U)$ można zapisać w ogólnej postaci

$$J = J_R \left[\exp \frac{qU}{mkT} - 1 \right] \quad (3.85)$$

Na zależność $J(T)$ mają wpływ dwa czynniki:

- rozważane wcześniej zmiany prądu J_R w funkcji temperatury,
- zmiany argumentu funkcji wykładniczej.

Należy zauważyć, że są to czynniki wzajemnie kompensujące się, gdyż w miarę wzrostu temperatury prąd J_R zwiększa się, funkcja wykładnicza zaś maleje. Dlatego należy oczekiwać, że wpływ temperatury na charakterystykę $J(U)$ dla kierunku przewodzenia będzie słabszy niż dla kierunku zaporowego. Na podstawie równania (3.85) (przy pominięciu jedności w nawiasie kwadratowym) temperatury współczynnik prądu przewodzenia przy stałym napięciu ma postać

$$\frac{1}{J} \frac{dJ}{dT} \Big|_U = \frac{1}{J_R} \frac{dJ_R}{dT} - \frac{qU}{mkT^2} \quad (3.86)$$

Pierwszy człon po prawej stronie równania (3.86) można przedstawić w postaci (3.82) lub (3.83) otrzymując:

$$\frac{1}{J} \frac{dJ}{dT} \Big|_U = \frac{W_g - qU}{2kT^2} \quad \begin{array}{l} \text{złącze krzemowe (zakres małych prądów} \\ U < 0,4 \text{ V, } J_R = J_g, m = 2) \end{array} \quad (3.86a)$$

$$\frac{1}{J} \frac{dJ}{dT} \Big|_U = \frac{W_g - qU}{kT^2} \quad \begin{array}{l} \text{złącze germanowe w całym zakresie oraz} \\ \text{krzemowe w zakresie normalnej pracy} \\ (J_R = J_s, m = 1) \end{array} \quad (3.86b)$$

Z wyprowadzonych wzorów (3.86a, b) można łatwo wyznaczyć współczynnik temperatury prądu przewodzenia, który wynosi ok. 5% dla typowych warunków pracy złączy w temperaturze ok. 300 K:

$U = 0,3 \text{ V}$ dla złącza germanowego [wzór (3.86b)],

$U = 0,3 \text{ V}$ dla złącza krzemowego w zakresie małych prądów [wzór (3.86a)],

$U = 0,7 \text{ V}$ dla złącza krzemowego w zakresie normalnej pracy

— w zakresie średnich prądów [wzór (3.86b)].

Dla wielu zagadnień spotykanych w praktyce jest konieczna znajomość dryfu temperatury napięcia na złączu p-n przy stałym prądzie przewodzenia. Na podstawie równania (3.85) (z pominięciem jedności w nawiasie kwadratowym) można wyrazić napięcie w funkcji prądu

$$U = \frac{mkT}{q} \ln \frac{J}{J_R} \quad (3.87)$$

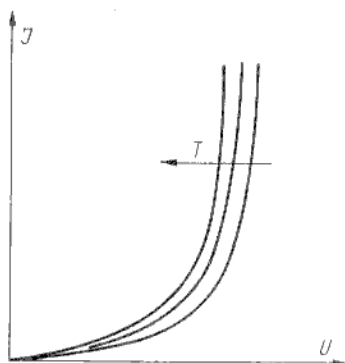
Stąd

$$\frac{dU}{dT} \Big|_J = \frac{U}{T} - \frac{mkT}{q} \frac{1}{J_R} \frac{dJ_R}{dT} \quad (3.88)$$

Uwzględniając zależności (3.82), (3.83) można przedstawić wyrażenie (3.88) w postaci

$$\left. \frac{dU}{dT} \right|_J = \frac{qU - W_g}{qT} \quad (3.88a)$$

Dla typowych warunków pracy złączy germanowych i krzemowych w temperaturze ok. 300 K dryf napięcia przy stałym prądzie przewodzenia wynosi ok. -2 mV/K. Dryf temperaturowy napięcia ma zawsze znak ujemny, gdyż napięcie polaryzacji $U < \varphi_B < W_g/q$. Charakterystykę prądowo-napięciową w zakresie przewodzenia przy różnych temperaturach przedstawiono na rys. 3.54.



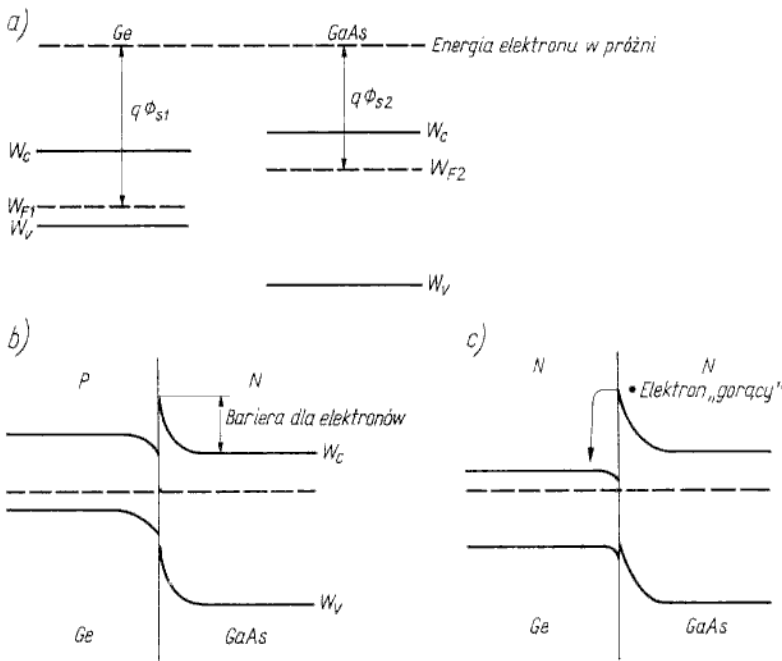
Rys. 3.54
Charakterystyki prądowo-napięciowe złącza $p-n$ w zakresie przewodzenia dla różnych temperatur

Heterozłącza

3.2

Heterozłącze jest złączem wytworzonym z dwu półprzewodników o różnych szerokościach pasma zabronionego (na przykład Ge—Si, Ge—GaAs, GaAs—GaP). Takie złącza znajdują zastosowania między innymi w fotodetektorach i laserach.

W przypadku idealnym jest wymagana bardzo dobra zgodność budowy sieci krystalicznej i współczynników rozszerzalności cieplnej w obu materiałach. Ponieważ te warunki nigdy nie są całkowicie spełnione, na granicy dwu warstw istnieją dyslokacje i defekty, mające istotny wpływ na charakterystyki heterozłącza. Jako przykład będzie rozpatrzony prosty przypadek idealnego złącza $p-n$ (Ge—GaAs). Na rysunku 3.55 przedstawiono model pasmowy obu warstw przed (a) i po ich połączeniu (b). Na rysunku 3.55a widać, że obie warstwy różnią się nie tylko szerokością pasma zabronionego, lecz również pracą wyjścia elektronów, określaną jako energia wymagana do przeniesienia elektronu z poziomu Fermiego na zewnątrz półprzewodnika (do próżni). Po połączeniu dwu warstw następuje wyrównanie poziomów Fermiego, czemu towarzyszy przepływ pewnej liczby elektronów z obszaru N do obszaru P (praca wyjścia elektronów z obszaru N jest mniejsza). W rezultacie powstaje warstwa ładunku przestrzennego o znaku dodatnim po stronie obszaru typu n i ujemnym po stronie obszaru typu p . Dotychczas opisane zjawiska są identyczne jak w „normalnym”, tj. homogennym złączu $p-n$. Istotną różnicą jest ta, że w heterozłączu bariery dla dziur i elektronów nie są jednakowe. W rozpatrywanym przykładzie bariera dla elektronów dyfundujących z obszaru N do obszaru P jest znacznie mniejsza niż bariera dla dziur dyfundujących z ob-



Rys. 3.55

Model pasmowy heterozłącza $p-n$: a) przed; b) po połączeniu obu warstw; c) model pasmowy heterozłącza $n-n$

szaru P do obszaru N . Stąd wynika duża przewaga prądu elektronowego nad prądem dziurowym.

Możliwe są konstrukcje heterozłączy zarówno $p-n$, $n-p$, jak również $n-n$, $p-p$. Przykładowo na rys. 3.55c przedstawiono energetyczny model pasmowy heterozłącza $n-n$. Elektrony dyfundujące ponad barierą z GaAs do Ge mają energię znacznie większą niż elektrony znajdujące się na dnie pasma przewodnictwa w Ge. Mówi się w takim przypadku, że nośniki wstrzykiwane są „gorące”. Ponieważ elektrony „gorące” są wstrzykiwane do obszaru typu n , spowodowany tym przyrost koncentracji elektronów (nośników większościowych) jest pomijalnie mały. Do pominięcia zatem jest zjawisko magazynowania ładunku nadmiarowego, co znakomicie sprzyja uzyskaniu dużych szybkości działania opisanego heterozłącza $n-n$.

Struktura MIS

3.3

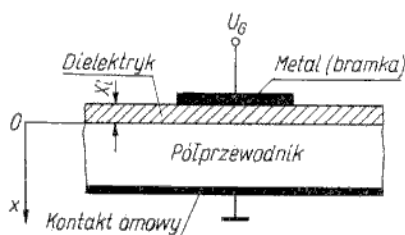
Bardzo ciekawe zjawiska obserwuje się przy powierzchni półprzewodnika poddanego działaniu pola elektrycznego. W celu wytworzenia silnego pola elektrycznego należy doprowadzić napięcie między półprzewodnik a elektrodę metalową, umieszczoną tuż nad powierzchnią półprzewodnika. Taki układ elektrod (metalowej i półprzewodnikowej) realizuje się w praktyce przez rozdzielenie elektrody metalowej od powierzchni półprzewodnika za pomocą cienkiej warstwy dielektryka (izolatora). W ten sposób powstaje struktura metal-izolator-półprzewodnik

(MIS — ang. *Metal-Insulator-Semiconductor*), mająca bardzo duże znaczenie w technologii półprzewodnikowej. Struktura MIS jest zasadniczym elementem składowym tranzystorów MIS i układów scalonych MIS. Z uwagi na duże bogactwo zjawisk w strukturze MIS rozpatrzmy najpierw uproszczony model tzw. idealnej struktury MIS, po czym uzupełnimy ten model opisem dodatkowych zjawisk występujących w strukturze rzeczywistej. Rozpatrzy się również podstawową charakterystykę struktury MIS, tj. zależność pojemności od napięcia polaryzacji.

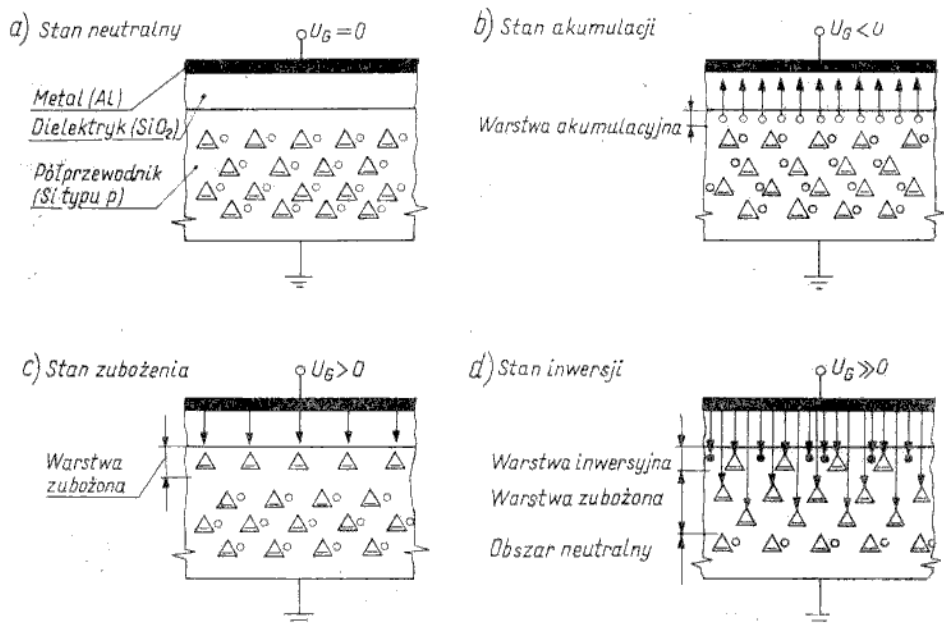
Model fizyczny idealnej struktury MIS

3.3.1

Strukturę MIS zilustrowano schematycznie na rys. 3.56. Przyjmuje się, że bez polaryzacji zewnętrznej zarówno w elektrodzie metalowej, jak również w warstwie dielektryka i w podłożu półprzewodnikowym nie ma żadnych nieskompensowa-



Rys. 3.56
Ilustracja schematyczna struktury MIS
 U_G — potencjał bramki (elektrody metalowej) względem uziemionego podłoża półprzewodnikowego; X_1 — grubość warstwy dielektrycznej



Rys. 3.57
Poglądowe ilustracje struktury MIS dla stanu neutralnego (a) oraz trzech charakterystycznych stanów (b, c, d), wywołanych polaryzacją zewnętrzną

nych (w sensie makroskopowym) ładunków elektrycznych. (W dowolnym obszarze makroskopowym, tj. zawierającym co najmniej kilkadziesiąt atomów, sumaryczny ładunek elektryczny jest — praktycznie biorąc — równy zeru). Jeżeli ponadto przyjmie się, że izolator jest idealny, czyli zupełnie nie przewodzi prądu elektrycznego, to struktura MIS o takich właściwościach będzie nazywana idealną. Rozpatrzmy jakościowo obraz zjawisk w idealnej strukturze MIS dla trzech istotnych przypadków polaryzacji.

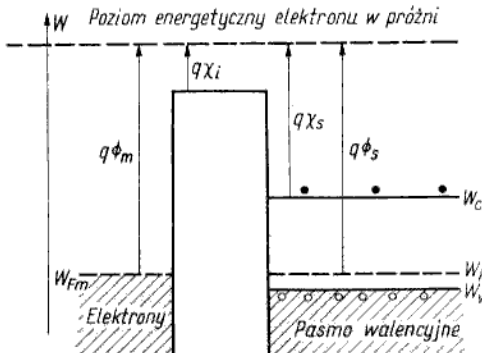
Na rysunku 3.57 przedstawiono ilustracje poglądowe dla stanu neutralnego oraz trzech charakterystycznych stanów wywołanych polaryzacją struktury. Jeżeli do elektrody metalowej jest doprowadzone napięcie ujemne, to dziury (nośniki większościowe) są przyciągane ku powierzchni granicznej półprzewodnik-dielektryk i zachodzi ich *akumulacja* (rys. 3.57b).

W przypadku małych wartości dodatnich napięcia bramki dziury odepchnięte od powierzchni granicznej półprzewodnik-dielektryk pozostawiają nieskompensowany ładunek przestrzenny ujemnie zjonizowanych centrów akceptorowych. Obszar ładunku przestrzennego, ze względu na pomijalnie małą koncentrację nośników prądu, jest nazywany *warstwą zubożoną* (rys. 3.57c). Dalszy wzrost napięcia do dodatniego bramki powoduje przyciąganie elektronów (nośników mniejszościowych) ku powierzchni granicznej półprzewodnik-dielektryk. W warstwie przypowierzchniowej półprzewodnika występuje zatem sumaryczny ładunek ujemny domieszek zjonizowanych oraz nośników mniejszościowych (rys. 3.57d). Im większa jest wartość napięcia dodatniego bramki, tym większą część całkowitego ładunku w półprzewodniku (Q_s) stanowi ładunek nośników mniejszościowych (Q_n).

Obszar przypowierzchniowy półprzewodnika, w którym koncentracja nośników mniejszościowych (w rozpatrywanym przypadku — elektronów) jest większa niż koncentracja nośników większościowych (w rozpatrywanym przypadku — dziur), nazywa się *warstwą inwersyjną*. Przedstawiony opis zjawisk odnosi się bezpośrednio do półprzewodnika typu *p*, lecz jest całkowicie oczywiste, że analogiczne zjawiska występują w przypadku podłoża z półprzewodnika typu *n* przy przeciwnych znakach napięcia polaryzacji.

Rysunki 3.57 dają tylko poglądowy obraz zjawisk, w bardziej zaawansowanych rozważaniach natomiast trzeba posługiwać się energetycznym modelem pasmowym struktury MIS. Spróbujemy skonstruować taki model dla stanu bez polaryzacji i przy dowolnie wybranej polaryzacji, na przykład przy $U_G < 0$. Energetyczny model pasmowy struktury MIS przy $U_G = 0$ przedstawiono na rys. 3.58.

Wprowadzamy tu nowe, bardzo ważne pojęcie pracy wyjścia ($q\Phi_m$ dla metalu, $q\Phi_s$ dla półprzewodnika). *Praca wyjścia* jest różnicą energii Fermiego i energii elektronu



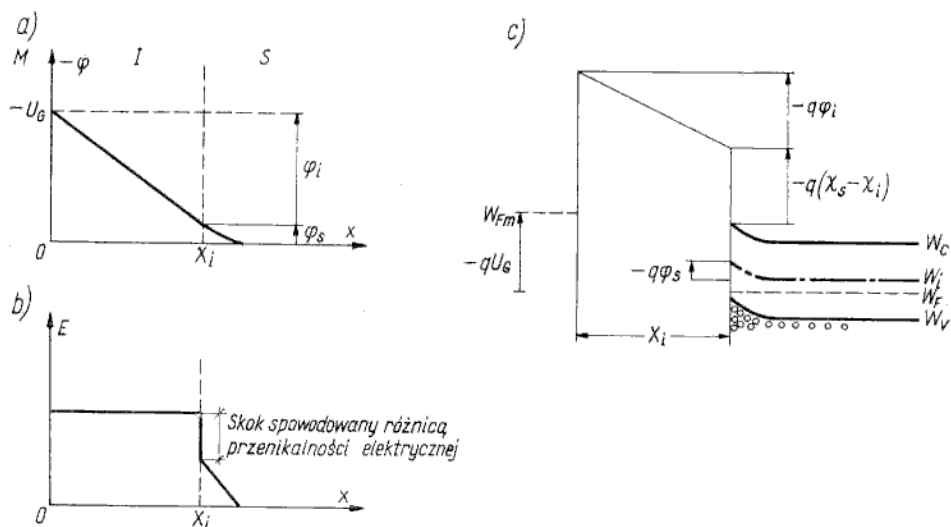
Rys. 3.58

Energetyczny model pasmowy struktury MIS w stanie neutralnym (bez polaryzacji)

w próżni. Jest to inaczej bariera energetyczna, jaką musi pokonać elektron, by opuścić ciało stałe i oddalić się na nieskończenie dużą odległość (chodzi o taką odległość, przy której można pominąć oddziaływanie elektronu z ciałem stałym). Wielkości Φ_m , Φ_s są nazywane *potencjalami wyjścia* z metalu i półprzewodnika. W strukturze idealnej $\Phi_m = \Phi_s$, gdyż w przeciwnym razie nastąpiłby przepływ elektronów z materiału o mniejszej pracy wyjścia i powstałby ładunek nieskompensowany (przypadek ten omówimy w p. 3.3.3). W strukturze MIS elektrony wychodzą z metalu i półprzewodnika nie do próżni, lecz do warstwy dielektryka. Praca wyjścia elektronu z metalu lub półprzewodnika do dielektryka jest nieco mniejsza niż do próżni. Różnica wynosi $q\chi_i$, przy czym χ_i jest tzw. *powinowactwem elektronowym* dielektryka. Jednakże nie zmienia to najistotniejszego faktu, że różnica prac wyjścia metal-izolator, półprzewodnik-izolator pozostaje równa zero. Oznacza to jednakowe prawdopodobieństwo przejścia elektronu z metalu przez dielektryk (ponad barierą $q(\Phi_m - \chi_i)$) do półprzewodnika, jak przejścia w kierunku odwrotnym z półprzewodnika przez dielektryk (ponad barierą $q(\Phi_s - \chi_i)$) do metalu.

W stanie równowagi termodynamicznej nie gromadzą się zatem ładunki nieskompensowane ani w metalu, ani w półprzewodniku. W tej sytuacji obecność elektrody metalowej i warstwy izolatora nie ma żadnego wpływu na energetyczny model pasmowy półprzewodnika (rys. 3.58), a równość prac wyjścia oznacza również jednakowe wartości energii Fermiego w metalu i w półprzewodniku.

Jeżeli spolaryzuje się ujemnie elektrodę metalową, to napięcie U_G odłoży się częściowo na warstwie izolatora, jako spadek napięcia φ_i , oraz częściowo na elektrodzie półprzewodnikowej, jako tzw. potencjał powierzchniowy φ_s (rys. 3.59a). Zauważmy,



Rys. 3.59

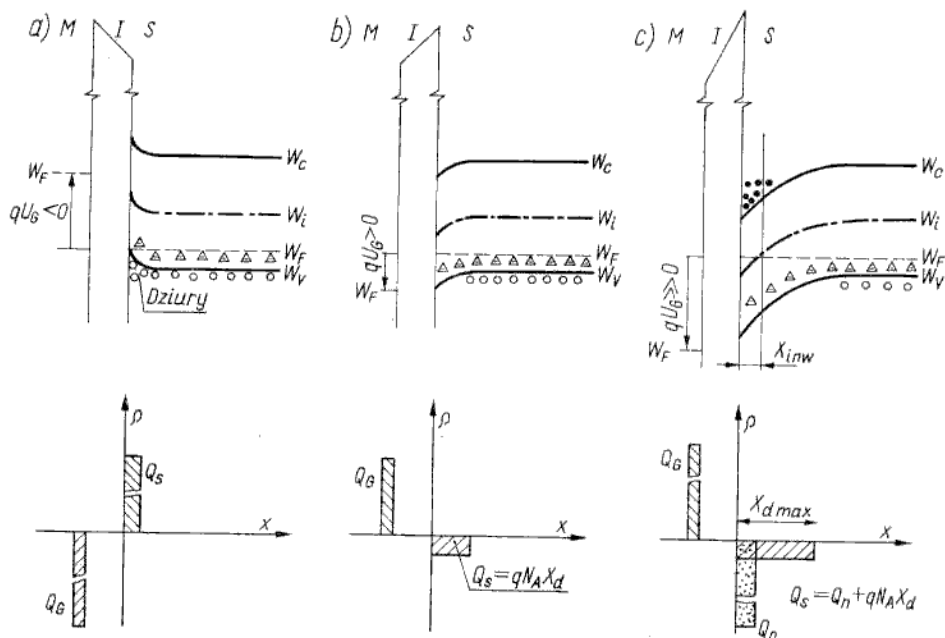
Rozkład potencjału (a), natężenia pola elektrycznego (b) oraz energetyczny model pasmowy dla struktury MIS, spolaryzowanej napięciem ujemnym

że w strukturze metal-izolator-metal (kondensator) całe napięcie odłożyłoby się na warstwie izolatora. W półprzewodniku koncentracja ładunku zaindukowanego przez pole elektryczne jest wiele rzędów wartości mniejsza niż w metalu, dlatego

głębokość wnikania pola elektrycznego jest znacznie większa. Dobrze ilustruje tę sytuację rys. 3.57d, na którym widać, że linie sił pola elektrycznego kończą się na ładunkach rozłożonych w półprzewodniku na znacznej głębokości. Oznacza to inaczej, że natężenie pola elektrycznego nie maleje raptownie do zera przy samej powierzchni (tak jest w metalu), lecz zmniejsza się stopniowo w obszarze przypowierzchniowym półprzewodnika (rys. 3.59b). Ponieważ rozkład ujemnego potencjału $\varphi(x)$ odpowiada rozkładowi energii elektronów, zatem łatwo można przejść z wykresu $\varphi(x)$ do energetycznego modelu struktury MIS (rys. 3.59c). Model przedstawiony na rys. 3.59c (dla $U_G \neq 0$) jest narysowany na podstawie modelu przedstawionego na rys. 3.58 (dla $U_G = 0$) przy zachowaniu następujących reguł:

- poziomy Fermiego w metalu i półprzewodniku różnią się o wartość energii pola elektrycznego qU_G ;
- energia wyjścia elektronów z dna pasma przewodnictwa w półprzewodniku do izolatora, tj. $q(\chi_s - \chi_i)$, pozostaje niezmienną;
- krawędzie pasm energetycznych (W_c , W_v) przyjmują taki sam kształt jak rozkład ujemnej wartości potencjału $\varphi(x)$.

Najbardziej istotną cechą modelu pasmowego półprzewodnika w strukturze MIS spolaryzowanej jest zagięcie krawędzi pasm energetycznych „do góry” odpowiada zmniejszeniu odstepu energetycznego ($W_F - W_v$) przy jednoczesnym zwiększeniu odstepu energetycznego ($W_c - W_F$). Zgodnie ze statystycznym rozkładem prawdopodobieństwa zajęcia poziomów energetycznych przez elektrony wynika stąd, że przy powierzchni półprzewodnika koncentracja dziur jest większa niż w głębi (poziom Fer-



Rys. 3.60

Energetyczny model pasmowy oraz rozkład gęstości ładunku dla trzech charakterystycznych przypadków polaryzacji struktury MIS: a) akumulacja; b) zubożenie; c) inwersja

miego jest bliżej wierzchołka pasma walencyjnego), koncentracja elektronów natomiast jest mniejsza przy powierzchni niż w głębi półprzewodnika (poziom Fermiego przy powierzchni jest bardziej oddalony od dna pasma przewodnictwa). W tym miejscu może nasuwać się pytanie: „Dlaczego nie zagina się poziom Fermiego?” Poziom Fermiego ma fizyczny sens średniej energii swobodnej elektronów, która w warunkach stałej temperatury i bez wpływu jakiegokolwiek innych bodźców energetycznych mogłaby się zmieniać tylko w ciągłym procesie dostarczania lub usuwania elektronów. Taka sytuacja nie występuje w rozważanym przypadku, gdyż nie ma przepływu prądu przez warstwę izolatora.

Obecnie możemy przedstawić energetyczny model pasmowy oraz rozkład gęstości ładunku dla trzech charakterystycznych przypadków polaryzacji struktury MIS (rys. 3.60), adekwatnych przypadkom przedstawianym uprzednio w sposób uproszczony na rys. 3.57b, c, d.

W „języku” modelu pasmowego potencjał powierzchniowy φ_s definiuje się jako różnicę położenia poziomu samoistnego Fermiego (poziom W_i na rys. 3.59c) w głębi i na powierzchni półprzewodnika. Zwrot „do góry” odpowiada wartości ujemnej potencjału. W tabelicy 3.1 podano zakresy potencjału powierzchniowego, odpowia-

Charakterystyczne zakresy potencjału powierzchniowego

Tablica 3.1

Stan powierzchniowy	Typ <i>n</i> ($\varphi_F < 0$)		Typ <i>p</i> ($\varphi_F > 0$)	
	Akumulacja	$\varphi_s > 0$	$Q_s < 0$	$\varphi_s < 0$
Pasma wyprostowane	$\varphi_s = 0$	$Q_s = 0$	$\varphi_s = 0$	$Q_s = 0$
Zubożenie	$\varphi_F < \varphi_s < 0$	$Q_s > 0$	$0 < \varphi_s < \varphi_F$	$Q_s < 0$
Inwersja	$2\varphi_F < \varphi_s \leq \varphi_F$	$Q_s > 0$	$\varphi_F \leq \varphi_s < 2\varphi_F$	$Q_s < 0$
Silna inwersja	$\varphi_s \leq 2\varphi_F$	$Q_s > 0$	$2\varphi_F \leq \varphi_s$	$Q_s < 0$

dające wyróżnianym stanom obszaru przypowierzchniowego półprzewodnika. Scharakteryzujemy krótko poszczególne zakresy pracy na przykładzie struktury MIS z półprzewodnikiem typu *p*.

- Akumulacja.** W stanie akumulacji ($U_G < 0$) pasma energetyczne zaginają się przy powierzchni półprzewodnika „w górę” (rys. 3.60a), co odpowiada zwiększeniu koncentracji nośników większościowych (dziur) i zmniejszeniu koncentracji nośników mniejszościowych (elektronów) w obszarze przypowierzchniowym półprzewodnika. W ten sposób powstaje cienka warstwa akumulacyjna o tym samym typie przewodnictwa, lecz większej konduktywności niż w głębi półprzewodnika.
- Wyprostowane pasma.** W idealnej strukturze MIS przy $U_G = 0$ potencjał powierzchniowy $\varphi_s = 0$. Wówczas pasma energetyczne są wyprostowane przy powierzchni półprzewodnika, czyli w obszarze przypowierzchniowym nie ma ładunku nieskompensowanego.
- Zubożenie.** W stanie zubożenia ($U_G > 0$) pasma energetyczne są zagięte „do dołu” (rys. 3.60b), co odpowiada zmniejszeniu koncentracji nośników większościowych (dziur). Zatem w obszarze przypowierzchniowym półprzewodnika pozostaje nieskompensowany ładunek ujemny zjonizowanych centrów akceptorowych. Powstaje w ten sposób warstwa zubożona o ładunku

$$Q_d = -qN_A X_d \tag{3.89}$$

przy czym X_d — głębokość wnikania pola elektrycznego, tj. szerokość warstwy zubożonej.

- d. *Inwersja*. Przy dalszym wzroście napięcia U_G zagięcie pasm energetycznych „do dołu” jest coraz silniejsze, czemu odpowiada dalsze zmniejszenie koncentracji nośników większościowych (dziur) i zwiększenie koncentracji nośników mniejszościowych (elektronów). Poczynając od stanu, gdy koncentracje obu rodzajów nośników zrównają się przy powierzchni półprzewodnika ($W_i = W_F$, czyli $\varphi_s = \varphi_F$) dalszy przyrost napięcia U_G powoduje, że koncentracja nośników mniejszościowych przy powierzchni zaczyna przeważać nad koncentracją nośników większościowych (rys. 3.60c). Przy powierzchni półprzewodnika typu p powstaje zatem cienka warstwa typu n . Taki stan nazywa się *inwersją* (odwrócenie typu przewodnictwa), a cienką warstwę typu n nazywa się *warstwą inwersyjną*. W tym stanie ładunek przestrzenny przy powierzchni półprzewodnika jest równy sumie ładunku zjonizowanych centrów akceptorowych (ładunek warstwy zubożonej) i ładunku nośników mniejszościowych (ładunek warstwy inwersyjnej)

$$Q_s = Q_d + Q_n \quad (3.90)$$

- e. *Silna inwersja*. Przy jeszcze większym napięciu U_G zagięcie pasm energetycznych „do dołu” jest jeszcze silniejsze i jeszcze bardziej wzrasta koncentracja nośników mniejszościowych (elektronów), czyli rozszerza się warstwa inwersyjna. Stan silnej inwersji jest zapoczątkowany wówczas, gdy koncentracja nośników mniejszościowych przy powierzchni zrówna się z koncentracją nośników większościowych w głębi półprzewodnika. Odpowiada to wartości potencjału powierzchniowego $\varphi_s = 2\varphi_F$. Ponieważ dalsze, nawet nieznaczące przyrosty wartości potencjału powierzchniowego powodują bardzo duże przyrosty ładunku warstwy inwersyjnej, zatem — praktycznie biorąc — można przyjąć, że warstwa inwersyjna ekranuje „głęboką” półprzewodnika od pola elektrycznego, czyli nie ulega zmianie grubość warstwy zubożonej. Mówi się zatem, że w stanie silnej inwersji warstwa zubożona osiąga maksymalną grubość, którą na podstawie analogii do modelu warstwy zaporowej złącza p - n można zapisać w postaci

$$X_{dmax} = \sqrt{\frac{4\epsilon_{rs}\epsilon_0\varphi_F}{qN_A}} \quad (3.91)$$

przy czym $\varphi_s = 2\varphi_F$ odpowiada napięciu dyfuzyjnemu φ_B w złączu p - n .

- f. *Głębokie zubożenie*. Dotychczas rozpatrywaliśmy właściwości struktury MIS w stanie równowagi termodynamicznej. Dla przypadku akumulacji i zubożenia po każdej zmianie napięcia polaryzacji stan równowagi ustala się, praktycznie biorąc, natychmiast, gdyż czas relaksacji nośników większościowych (czas potrzebny na przegrupowanie ładunku nośników większościowych, odpowiadające nowej wartości napięcia U_G), określony wyrażeniem $\tau = \rho\epsilon_r\epsilon_0$, jest bardzo mały (np. dla Si o rezystywności $\rho = 10^{-2}\Omega \cdot m$ czas relaksacji $\tau \approx 10^{-12}$ s). Inna jest sytuacja w stanie inwersji. Na przykład po raptownej zmianie napięcia polaryzacji od zera do dużych wartości dodatnich wskutek pola elektrycznego działającego przy powierzchni półprzewodnika dziury (nośniki większościowe) zostaną, praktycznie biorąc, natychmiast odepchnięte w głąb półprzewodnika. Żeby w obszarze przypowierzchniowym ustalił się stan równowagi ($np = n_i^2$),

konieczny jest odpowiedni wzrost koncentracji nośników mniejszościowych (elektronów). Załóżmy przykładowo, że przy zerowej polaryzacji:

$$p_{p0} = 10^{21} \text{ m}^{-3}$$

$$n_{p0} = \frac{n_i^2}{p_{p0}} \approx \frac{10^{32}}{10^{21}} = 10^{11} \text{ m}^{-3}$$

oraz przyjmijmy, że po odepchnięciu dziur przy powierzchni

$$p_p = 10^{11} \text{ m}^{-3}$$

Wówczas koncentracja nośników mniejszościowych przy powierzchni powinna zwiększyć się do wartości

$$n_p \approx \frac{10^{32}}{10^{11}} = 10^{21} \text{ m}^{-3}$$

Weźmy strukturę MIS o powierzchni $A = 10^{-6} \text{ m}^2$ (tj. 1 mm^2) i przyjmijmy realne wartości grubości płytki półprzewodnikowej $l = 100 \text{ }\mu\text{m}$ oraz grubości warstwy inwersyjnej $X_{inv} = 100 \text{ \AA}$ (tj. 10^{-8} m). Wyznaczymy teraz całkowitą liczbę elektronów przy zerowej polaryzacji oraz przy silnej polaryzacji ujemnej:

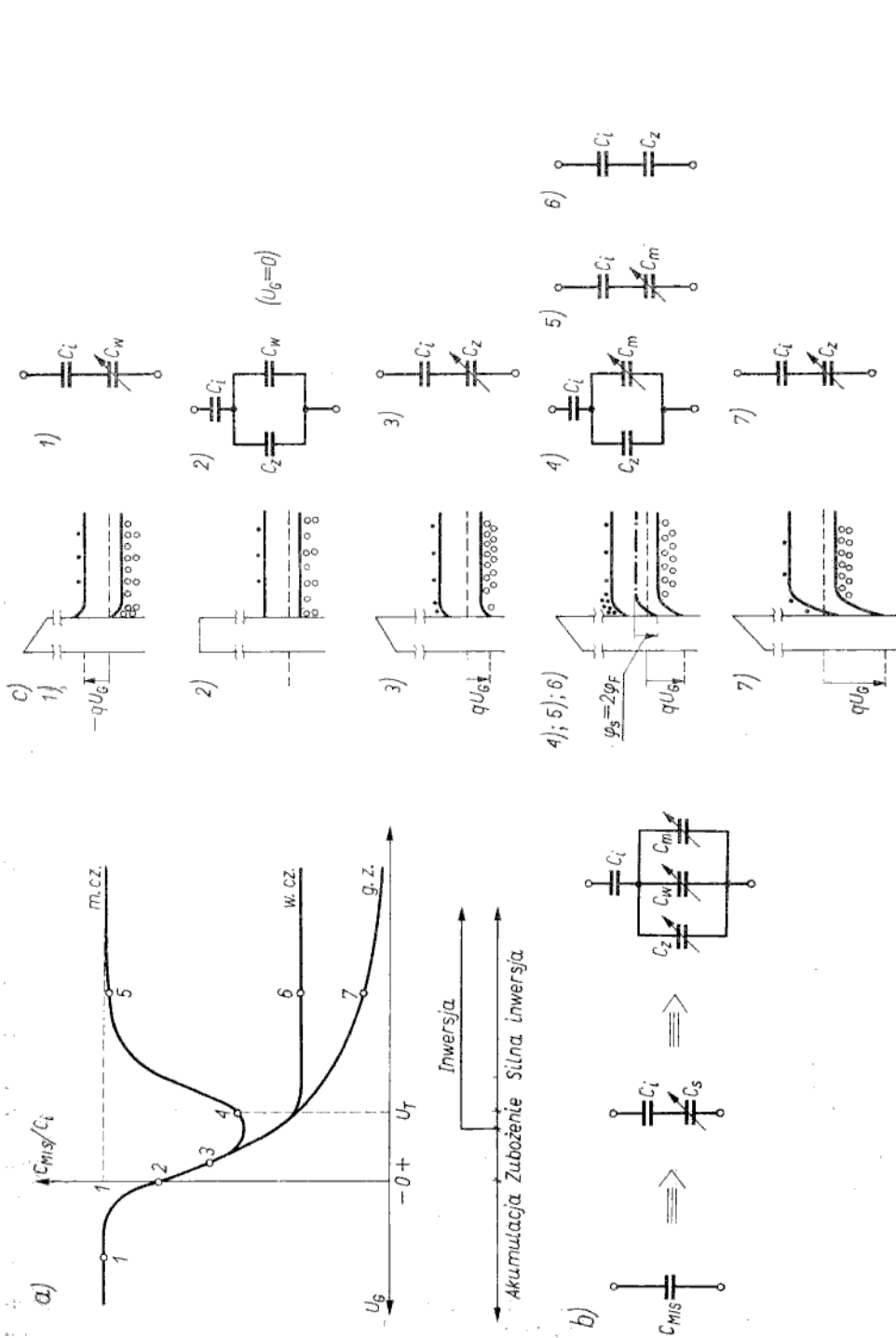
$$\left. \frac{Q_n}{q} \right|_{U_G=0} = n_{p0} A l = 10^{11} \cdot 10^{-6} \cdot 10^{-4} = 10$$

$$\left. \frac{Q_n}{q} \right|_{U_G \gg 0} = n_p A X_{inv} = 10^{21} \cdot 10^{-6} \cdot 10^{-8} = 10^7$$

Przedstawiony przykład liczbowy prawidłowo ilustruje fakt, że warstwa inwersyjna nie może się utworzyć wskutek przyciągnięcia nośników mniejszościowych (elektronów) z głębi półprzewodnika, gdyż w całej płytce półprzewodnikowej jest ich zdecydowanie za mało (10 w rozpatrywanym przykładzie) w stosunku do „potrzeb” warstwy inwersyjnej (10^7 w rozpatrywanym przykładzie). Skąd wziąć brakujące elektrony? Muszą one powstać wskutek generacji cieplnej. Ponieważ w obszarze warstwy zubożonej po odepchnięciu nośników większościowych (dziur) $n_p p_p \ll n_i^2$, zatem proces generacji przeważa nad procesem rekombinacji. A więc w miarę powstawania par elektron-dziura dziury są usuwane w głąb półprzewodnika, a elektrony pozostają przy powierzchni (utrzymywane przez przyciągające pole elektryczne) i stopniowo gromadzi się ładunek warstwy inwersyjnej.

W ten sposób stan równowagi termodynamicznej ustala się powoli. Przy bardzo dobrym stanie powierzchni półprzewodnika, tj. małej liczbie centrów generacyjnych, może to być proces trwający kilka i więcej sekund. Dlatego w pewnym okresie czasu bezpośrednio po doprowadzeniu dużego napięcia dodatniego U_G nie ma warstwy inwersyjnej, czyli pole elektryczne może wnikać głęboko do półprzewodnika. Wówczas potencjał powierzchniowy może przyjmować wartości znacznie większe niż $2\varphi_F$, a szerokość warstwy zubożonej nie jest ograniczona wielkością X_{dmax} . Taki stan nazywa się *głębokim zubożeniem*.

Dotychczas rozpatrywaliśmy stan głębokiego zubożenia jako stan nieustalony, występujący po raptownej zmianie napięcia polaryzacji (taka sytuacja istnieje w układach CCD — por. p. 10.7). Jednakże stan głębokiego zubożenia może mieć charakter ustalony w tych przypadkach, gdy nośniki mniejszościowe nie mogą się



Rys. 3.61 Charakterystyka pojemnościowo-napięciowa struktury MIS (a), pojemnościowy schemat zastępczy (b) oraz interpretacja modelowa stanów pracy struktury MIS, odpowiadających punktom zaznaczonym na charakterystyce $C(U)$ (c) m.c.z. — charakterystyka dla małych częstotliwości, w.c.z. — charakterystyka dla dużych częstotliwości, g.z. — charakterystyka typu głębokiego zubożenia

zgrupować przy powierzchni półprzewodnika wskutek ich usuwania przez warstwę dielektryka (upływ) lub przez złącze *p-n* spolaryzowane zaporowo (taka sytuacja istnieje w tranzystorze MOS — por. p. 6.2).

Charakterystyka pojemnościowo-napięciowa

3.3.2

Struktura MIS jako kondensator nieliniowy (kondensator o pojemności zmiennej w funkcji napięcia polaryzacji) jest doskonałym narzędziem badań właściwości powierzchniowych półprzewodnika. Rozpatrzmy przebieg pojemności różniczkowej (rys. 3.61a) w funkcji napięcia i częstotliwości dla idealnej struktury MIS. Pojemność różniczkową definiuje się jako

$$C = \frac{dQ_G}{dU_G} \quad (3.92)$$

przy czym ładunek bramki Q_G , jak i inne ładunki wprowadzone dalej, jest określony na jednostkę powierzchni.

Ponieważ napięcie bramki można przedstawić w postaci sumy spadku napięcia na warstwie dielektryka i potencjału powierzchniowego

$$U_G = \varphi_i + \varphi_s$$

a ładunek bramki jest równy ładunkowi półprzewodnika z przeciwnym znakiem

$$Q_G = -Q_s$$

Zatem

$$\frac{1}{C} = -\frac{dU_G}{dQ_s} = -\frac{d\varphi_i}{dQ_s} - \frac{d\varphi_s}{dQ_s} = \frac{1}{C_i} + \frac{1}{C_s} \quad (3.93)$$

przy czym

$$C_i = -\frac{dQ_s}{d\varphi_i} = \frac{dQ_G}{d\varphi_i} = \frac{\epsilon_r \epsilon_0}{X_i} \quad (3.94)$$

jest pojemnością warstwy dielektryka,

$$C_s \stackrel{\text{def}}{=} -\frac{dQ_s}{d\varphi_s} \quad (3.95)$$

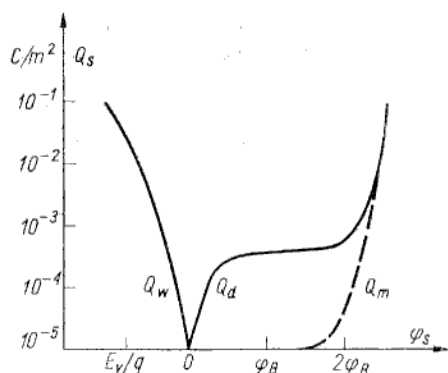
jest pojemnością obszaru ładunku przestrzennego przy powierzchni półprzewodnika.

Elektrycznym schematem zastępczym struktury idealnej MIS jest zatem połączenie szeregowo pojemności C_i dielektryka oraz pojemności C_s obszaru przypowierzchniowego półprzewodnika, zależnej od napięcia (rys. 3.61b).

Charakterystyka $C_s(\varphi_s)$ wykazuje dyspersję częstotliwościową, spowodowaną różnymi czasami relaksacji składowych ładunku Q_s . Dlatego dogodnie jest analizować wpływ poszczególnych składowych uwzględniając, że (rys. 3.62)

$$Q_s = Q_d + Q_w + Q_m \quad (3.96)$$

przy czym: Q_d — ładunek zjonizowanych centrów domieszkowych; Q_w — ładunek nośników większościowych; Q_m — ładunek nośników mniejszościowych.



Rys. 3.62
Gęstość ładunku w obszarze przypowierzchniowym półprzewodnika w funkcji potencjału powierzchniowego

Stąd

$$C_s = -\frac{dQ_s}{d\varphi_s} = -\frac{dQ_d}{d\varphi_s} - \frac{dQ_w}{d\varphi_s} - \frac{dQ_m}{d\varphi_s} = C_z + C_w + C_m \quad (3.97)$$

Po określeniu schematu zastępczego struktury MIS można wyjaśnić kształt charakterystyki $C(U_G)$. W tym celu na rys. 3.61c przedstawiono uproszczone szkice modelu pasmowego i schematy zastępcze, odpowiadające poszczególnym punktom zaznaczonym na charakterystykach $C(U_G)$ — rys. 3.61a.

Punktowi 1 odpowiada stan akumulacji. Wówczas ładunek nośników większościowych $Q_w \gg Q_d + Q_m$. Przyrosty potencjału powierzchniowego $\Delta\varphi_s$ powodują przyrosty ładunku $\Delta Q_s \approx \Delta Q_w$. Zatem $C_w \gg C_z + C_m$, czyli $C_s \approx C_w$. Ponieważ bardzo małym przyrostom potencjału $\Delta\varphi_s$ odpowiadają duże przyrosty ładunku ΔQ_w , więc pojemność C_w przyjmuje wartości znacznie większe niż szeregową pojemność warstwy dielektryka ($C_w \gg C_i$) i — praktycznie biorąc — można przyjąć, że całkowita pojemność struktury MIS

$$C_{\text{MIS}} \approx C_i$$

W uproszczonej interpretacji pojemność półprzewodnika można utożsamiać z pojemnością kondensatora płaskiego, w którym odległość między okładkami jest równa grubości warstwy ładunku przestrzennego (podobnie jak w przypadku pojemności warstwy zaporowej w złączu $p-n$). Warstwa zakumulowanego ładunku nośników większościowych jest bardzo cienka, stąd $C_s \gg C_i$.

Punktowi 2 (napięcie $U_G = 0$) odpowiada stan wyprostowanych pasm energetycznych aż do samej powierzchni. W tym przypadku podział pojemności półprzewodnika C_s na składowe C_z , C_w jest dość sztuczny, lecz został zachowany dla jednolitości interpretacji. W istocie ujemnym przyrostom potencjału $\Delta\varphi_s$ odpowiadają przyrosty ładunku Q_w , czyli należałoby mówić o pojemności C_w ; dodatnim natomiast przyrostom potencjału $\Delta\varphi_s$ odpowiadają przyrosty ładunku Q_d , czyli należałoby mówić o pojemności C_z . Jednakże przyrost ładunku Q_d jest spowodowany zmniejszeniem się ładunku Q_w , zatem pojemności C_z , C_w przejawiają się jako skutek tego samego mechanizmu fizycznego. Dlatego można mówić po prostu o pojemności półprzewodnika.

Jeśli przyjmie się model kondensatora płaskiego o okładkach oddalonych na grubość warstwy ładunku przestrzennego, to pozornie wydaje się, że w stanie płaskich pasm coraz mniejszym przyrostom potencjału powierzchniowego powinien odpo-

wiadać coraz cieńszy obszar ładunku przestrzennego, czyli pojemność półprzewodnika powinna wzrastać do nieskończoności. Jednakże to rozumowanie nie jest słuszne, gdyż istnieje pewna charakterystyczna odległość, na którą rozprzestrzenia się nawet najmniejsze zaburzenie neutralności elektrycznej. Jest to tzw. *droga Debye'a*. Droga Debye'a jest nazywana inaczej *drogą ekranowania*, gdyż jest to odległość, na której — praktycznie biorąc — zanika pole elektryczne (ściśle biorąc, potencjał elektryczny maleje e -krotnie). Zachowując analogię do kondensatora płaskiego można zapisać wyrażenie na pojemność półprzewodnika w stanie płaskich pasm C_{sFB} ¹⁾

$$C_{sFB} = \frac{\epsilon_{rs}\epsilon_0}{L_D} \quad (3.98)$$

przy czym

$$L_D = \sqrt{\frac{kT \epsilon_{rs}\epsilon_0}{q^2 p_p}} \quad (3.99)$$

jest drogą Debye'a dla półprzewodnika typu p . Stąd w punkcie 2 pojemność kondensatora MIS

$$C_{MIS} = \frac{C_i C_{sFB}}{C_i + C_{sFB}} \quad (3.100)$$

Punktowi 3 ($U_G > 0$) odpowiada stan zubożenia, czyli opróżnienia obszaru przypowierzchniowego półprzewodnika z nośników większościowych (dziur) przy jednocześnie małej koncentracji nośników mniejszościowych (elektronów). Zatem ładunek przestrzenny jest, praktycznie biorąc, równy Q_d , a pojemność $C_s \approx C_z$. W miarę wzrostu napięcia U_G warstwa zubożona rozszerza się (okładki kondensatora oddalają się od siebie) i pojemność C_z maleje. (Jest to zjawisko podobne jak dla pojemności warstwy zaporowej w złączu $p-n$). Ponieważ pojemność $C_z < C_i$, maleje również całkowita pojemność struktury MIS.

W dotychczas rozpatrywanym zakresie charakterystyka $C(U_G)$ nie zależy od częstotliwości, gdyż szybkość zmian zarówno ładunku Q_w , jak i Q_d jest określona czasem relaksacji nośników większościowych (ok. 10^{-12} s).

Punkt 4 odpowiada początkowi stanu silnej inwersji ($\varphi_s = 2\varphi_F$). Napięcie polaryzacji w tym punkcie $U_G = U_T$, nazywane *napięciem progowym*, jest podstawowym parametrem tranzystorów MIS, które będą omawiane w rozdz. 6. W warunkach polaryzacji quasi-statycznej w tym punkcie następuje ograniczenie grubości warstwy zubożonej [wzór (3.91)], czyli pojemność C_z uzyskuje wartość najmniejszą i niezależną od dalszych przyrostów napięcia. Jeżeli częstotliwość sygnału pomiarowego (należy pamiętać, że pojemność różniczkową mierzy się za pomocą sygnału sinusoidalnego o małej amplitudzie, nałożonego na stałe napięcie polaryzacji) jest wystarczająco mała, to w takt zmian napięcia zmienia się również ładunek nośników mniejszościowych w warstwie inwersyjnej, czyli występuje pojemność C_m .

Punktowi 5 odpowiada stan silnej inwersji, przy czym jest to gałąź charakterystyki $C(U)$ otrzymywana dla małej częstotliwości sygnału pomiarowego (gałąź m.cz.). Wiadomo, że ładunek nośników mniejszościowych może się zmieniać głównie wskutek procesu generacji-rekombinacji, a jest to proces o dużych czasach re-

¹⁾ FB (ang. *flat-band*) — płaskie pasmo.

laksacji. Dlatego małą częstotliwość należy rozumieć jako częstotliwość mniejszą niż kilkadziesiąt — kilka herców, a niekiedy nawet ułamek herca. W tym zakresie częstotliwości małym przyrostem potencjału powierzchniowego odpowiadają bardzo duże przyrosty ładunku warstwy inwersyjnej, czyli pojemność C_m przyjmuje bardzo duże wartości $C_m \gg C_z$. Ponieważ $C_m \gg C_i$, zatem — praktycznie biorąc — całkowita pojemność struktury MIS $C_{MIS} \approx C_i$ (tak jak w akumulacji).

Punktowi 6 odpowiada taki sam stan przy powierzchni półprzewodnika jak dla punktu 5, lecz jest to gałąź charakterystyki $C(U)$ otrzymywana przy dużej częstotliwości sygnału pomiarowego (według kryteriów podanych wyżej), zatem $C_m \approx 0$. Wówczas gałąź w.cz. charakterystyki $C(U)$ jest określona wzorem

$$C_{MIS} = \frac{C_i C_{z \min}}{C_i + C_{z \min}} \quad (3.101)$$

przy czym

$$C_{z \min} = \frac{\epsilon_{rs} \epsilon_0}{X_{d \max}} \quad (3.102)$$

a $X_{d \max}$ jest dane wzorem (3.91).

Punktowi 7 odpowiada stan głębokiego zubożenia (gałąź g.z.) otrzymywany na przykład przy szybkich zmianach napięcia polaryzacji. W tym przypadku $C_z \gg C_m + C_w$, a ponieważ warstwa zubożona rozszerza się w miarę wzrostu napięcia, pojemność C_z maleje. W związku z tym maleje również

$$C_{MIS} = \frac{C_i C_z}{C_i + C_z}$$

Rzeczywista struktura MIS

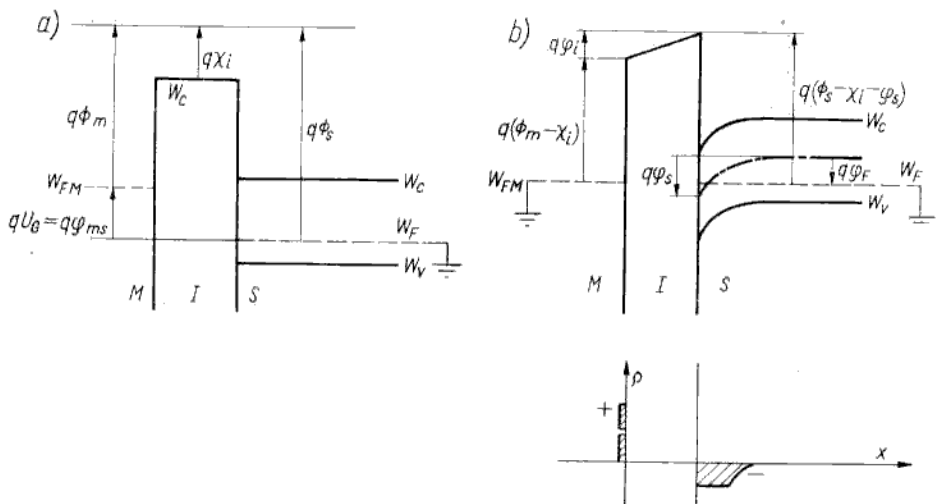
3.3.3

Rzeczywista struktura MIS różni się od idealnej tym, że przy polaryzacji $U_G = 0$ mogą istnieć nieskompensowane przestrzenne ładunki elektryczne. Należy wymienić trzy podstawowe przyczyny takiego stanu:

- prace wyjścia elektronów z metalu i półprzewodnika nie są jednakowe (istnieje kontaktowa różnica potencjałów);
- w warstwie dielektryka istnieje nieskompensowany ładunek elektryczny;
- w obszarze granicznym półprzewodnik-dielektryk występują dodatkowe dozwolone poziomy energetyczne, tzw. stany powierzchniowe.

Na rysunku 3.63 przedstawiono model pasmowy struktury MIS przy różnych potencjałach wyjścia elektronu z metalu i półprzewodnika ($|\Phi_m| < |\Phi_s|$). Na rysunku 3.63a przedstawiono model struktury w stanie nierównowagi (różne poziomy Fermiego w metalu i półprzewodniku). Stan nierównowagi istnieje wówczas, gdy odległość między metalem a półprzewodnikiem jest na tyle duża, że nie zachodzi wzajemny przepływ elektronów, lub gdy do struktury jest doprowadzone napięcie zewnętrzne U_G . W przypadku przedstawionym na rys. 3.63a napięcie zewnętrzne ma taką wartość, że przy powierzchni półprzewodnika pasma energetyczne są wyprostowane.

Na rysunku 3.63b przedstawiono model struktury w stanie równowagi (jeden wspólny poziom Fermiego). Stan równowagi można uzyskać przy $U_G = 0$, tj. przy zwar-



Rys. 3.63
Energetyczny model pasmowy rzeczywistej struktury MIS o kontaktowej różnicy potencjałów różnej od zera: a) stan nierównowagi; b) stan równowagi

ciu bramki z podłożem półprzewodnikowym (na przykład na rys. 3.63b elektrodę metalową i półprzewodnik uziemiono). Przy cienkiej warstwie dielektryka stan równowagi ustala się bez zwierającego obwodu zewnętrznego, gdyż możliwy jest przepływ elektronów między metalem a półprzewodnikiem ponad barierami energetycznymi $q(\Phi_m - \chi_i)$, $q(\Phi_s - \chi_i)$. Praca wyjścia elektronów z metalu jest mniejsza niż z półprzewodnika, a więc w czasie ustalania się równowagi termodynamicznej więcej elektronów będzie przechodzić w kierunku metal \rightarrow \rightarrow półprzewodnik niż w kierunku przeciwnym. Wówczas przy powierzchni półprzewodnika zgromadzi się dodatkowa liczba elektronów, tworząc ujemny ładunek przestrzenny. Należy zauważyć, że zwiększenie koncentracji elektronów oznacza w stanie równowagi jednocześnie zmniejszenie koncentracji dziur ($np = = n_i^2$), zatem ujemny ładunek przestrzenny tworzą nie tylko elektrony, lecz również zjonizowane centra akceptorowe, nieskompensowane przez dziury. W miarę powstawania warstwy ładunku przestrzennego wzrasta efektywna bariera energetyczna dla elektronów przechodzących z metalu do półprzewodnika, gdyż muszą one pokonać dodatkowe działanie odpychające ujemnego ładunku warstwy przypowierzchniowej półprzewodnika. W stanie równowagi ustala się taka wartość ładunku przestrzennego, że prawdopodobieństwa przejść elektronów są w obu kierunkach jednakowe. Wówczas w warstwie dielektryka i w obszarze przypowierzchniowym półprzewodnika istnieje pole elektryczne. Porównując rys. 3.63a z rys. 3.63b widać, że stan wyprostowanych pasm energetycznych w półprzewodniku otrzymuje się przy zewnętrznym napięciu U_G kompensującym wewnętrzne pole elektryczne

$$U_G = -(\varphi_i + \varphi_s)$$

Na podstawie rys. 3.63b można napisać równanie

$$\Phi_m - \chi_i + \varphi_i = \Phi_s - \chi_i - \varphi_s$$

Stąd

$$\varphi_i + \varphi_s = \Phi_s - \Phi_m$$

Zatem

$$U_G = \Phi_m - \Phi_s$$

Tę wartość napięcia zewnętrznego, kompensującą zagięcie pasm energetycznych, spowodowane różnymi pracami wyjścia, nazywa się *kontaktową różnicą potencjałów* φ_{ms}

$$\varphi_{ms} = \Phi_m - \Phi_s \quad (3.103)^{1)}$$

Dla struktury Al—SiO₂—Si (mającej ogromne znaczenie praktyczne w technologii układów scalonych) znaleziono wyrażenie empiryczne

$$\varphi_{ms} = -0,6 \text{ V} - \varphi_F \quad (3.104)$$

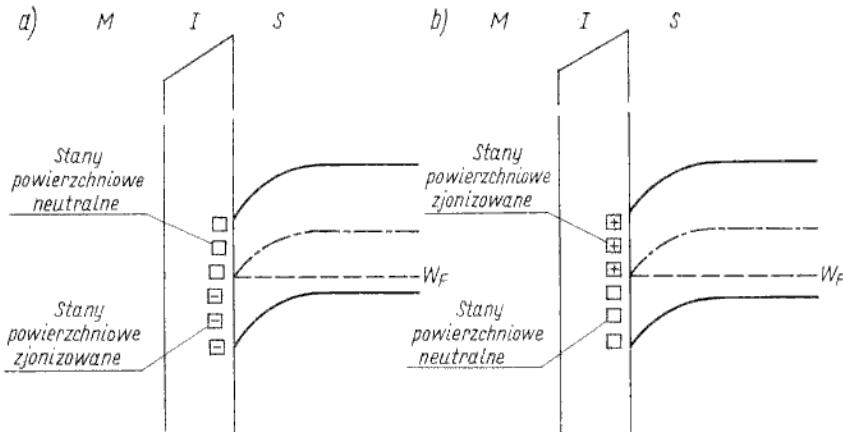
przy czym φ_F należy podstawić ze znakiem „-” dla półprzewodnika typu *n* oraz „+” dla półprzewodnika typu *p*.

Przykładowo dla struktury Al—SiO₂—Si (typu *n*) przy $N_D = 10^{21} \text{ m}^{-3}$ (typowa koncentracja domieszek) $\varphi_F = -0,3 \text{ V}$, czyli $\varphi_{ms} = -0,3 \text{ V}$.

Największe znaczenie praktyczne mają struktury MIS, w których izolatorem jest warstwa SiO₂; są one najczęściej nazywane strukturami MOS (ang. *Metal-Oxide-Semiconductor*).

Rozpatrzmy zatem zagadnienie ładunków nieskompensowanych w dielektryku i stanów powierzchniowych dla tego rodzaju struktur.

Tamm w 1932 r. i Shockley w 1939 r. wykazali teoretycznie, że w pasmie zabronionym ciała stałego istnieją elektronowe stany energetyczne, spowodowane nieciągłością sieci krystalicznej (zerwane wiązania sieci krystalicznej przy powierzchni ciała stałego). Tego rodzaju dozwolonych poziomów energetycznych, nazywanych



Rys. 3.64

Ilustracja stanów powierzchniowych typu akceptorowego (a) i typu donorowego (b)

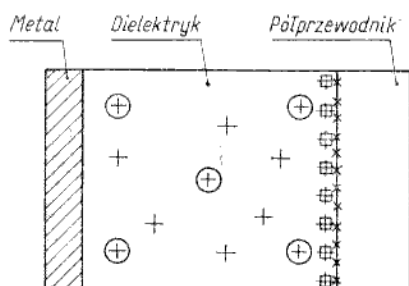
¹⁾ Wzór (3.103) jest słuszny dla prac wyjścia wziętych ze znakiem dodatnim. Ścisłe biorąc Φ_s , Φ_m są potencjałami ujemnymi i przestrzegając tych znaków należałoby napisać $\varphi_{ms} = -\Phi_s - \Phi_m$.

stanami powierzchniowymi T_{ama} , jest tyle, ile atomów przy powierzchni. Zatem gęstość powierzchniowa (określona na jednostkę powierzchni) stanów T_{ama} wynosi ok. 10^{19} m^{-2} . W strukturze metal— SiO_2 —Si gęstość stanów energetycznych w międzypowierzchni Si— SiO_2 jest znacznie mniejsza (mniej niż 10^{15} m^{-2}), gdyż wiązania Si—O w znacznej mierze usuwają nieciągłość powierzchniową sieci krystalicznej krzemu. Stany powierzchniowe (dozwolone poziomy energetyczne w międzypowierzchni Si— SiO_2) są w tym przypadku częściowo spowodowane defektami struktury krystalicznej krzemu oraz częściowo domieszkami. Stany powierzchniowe mogą spełniać funkcje akceptorów lub donorów w zależności od tego, czy w stanie pustym są obojętne elektrycznie, a w stanie zapelnienia elektronami jonizują się ujemnie (akceptory), czy też w stanie zapelnienia elektronami są obojętne elektrycznie, a w stanie pustym jonizują się dodatnio (donory) (rys. 3.64). W jednym i drugim przypadku stany powierzchniowe są źródłem pewnego ładunku nieskompensowanego. W rzeczywistej strukturze MOS oprócz stanów energetycznych, wynikających z defektów budowy międzypowierzchni Si— SiO_2 , należy uwzględnić również dozwolone stany energetyczne i ładunki ułożone przestrzennie w warstwie dielektryka.

Stany energetyczne, istniejące w obszarze granicznym Si— SiO_2 oraz w warstwie dielektryka, można podzielić na wymieniające ładunek z półprzewodnikiem (tj. ułożone energetycznie „na wysokości” pasma zabronionego półprzewodnika) oraz nie wymieniające ładunku z półprzewodnikiem (tj. ułożone energetycznie „poza” pasmem zabronionym półprzewodnika).

Ładunek występujący w dielektryku ze względu na możliwość przesuwania się w warstwie dielektryka można podzielić na ruchomy i nieruchomy.

Zgodnie z tym rozumowaniem przyjmuje się model uwzględniający cztery zasadnicze grupy ładunku nieskompensowanego w strukturze MOS (rys. 3.65):



Rys. 3.65

Cztery rodzaje ładunków nieskompensowanych w rzeczywistej strukturze MOS

⊕ — ładunek Q_f ; + — ładunek Q_i ; ⊖ — ładunek Q_{ss} ; × — ładunek Q_{st}

- Ładunek ruchomy w warstwie dielektryka (Q_f). Jest to ładunek jonów (najczęściej dodatnich) przesuwających się w warstwie dielektryka pod wpływem pola elektrycznego, szczególnie w podwyższonej temperaturze.
- Ładunek nieruchomy w warstwie dielektryka (Q_i). Najczęściej jest to dodatni ładunek pułapek dziurowych, wywołany działaniem promieniowania jonizującego, nie przesuwający się w warstwie dielektryka.
- Ładunek stały stanów powierzchniowych (Q_{ss}). Jest to ładunek dodatni stanów energetycznych, ułożonych w obszarze granicznym Si— SiO_2 , nie wymieniających ładunku z półprzewodnikiem. Jego wartość nie zależy od potencjału

przy czym Q_s' — ładunek w półprzewodniku zaindukowany łącznie przez Q_{ss} , Q_{ss} , Q_j , Q_t .

Ładunek Q_{sr} nie jest sumą algebraiczną czterech ładunków, gdyż na przykład ładunek Q_j , rozłożony przestrzennie w warstwie dielektryka, indukuje częściowo ładunek w półprzewodniku, a częściowo w elektrodzie metalowej. Do potrzeb uproszczonych analiz inżynierskich można przyjąć, że Q_{sr} jest ładunkiem stałym w funkcji napięcia, gdyż w strukturach MOS wytwarzanych w bardzo starannych procesach technologicznych ładunki Q_{ss} , Q_j , Q_t są bardzo małe i można przyjąć, że $Q_{sr} \approx Q_{ss}$.

Na zakończenie tego tematu porównajmy charakterystyki $C(U)$ struktury MIS idealnej z rzeczywistą (rys. 3.67). Widać, że charakterystyka pojemnościowo-napięciowa rzeczywistej struktury MIS jest przesunięta wzdłuż osi napięcia o wartość

$$U_{FB} = \varphi_{ms} - \frac{Q_{sr}}{C_i} \quad (3.106)$$

przy czym U_{FB} — napięcie bramki, któremu odpowiadają wyprostowane pasma energetyczne przy powierzchni półprzewodnika.

Złącza metal-półprzewodnik

3.4

Złącze metal-półprzewodnik będziemy skrótowo nazywać złączem *m-s* (ang. *metal-semiconductor*). Zjawiska występujące na styku warstw metalu i półprzewodnika są bardzo interesujące z dwu powodów:

- złącze *m-s* może mieć niesymetryczną, nieliniową charakterystykę prądowo-napięciową;
- złącze *m-s* może mieć symetryczną, liniową charakterystykę prądowo-napięciową.

W stwierdzeniu, że obie wyżej wymienione właściwości są bardzo cenne, nie ma żadnej przesady, choć brzmi ono nieco przekornie. Z jednej strony, niesymetryczna i nieliniowa charakterystyka $I(U)$ stwarza możliwości zastosowań złącza *m-s* jako diody (diody *m-s*, nazywane często *diodami Schottky'ego*, są niezastąpione w wielu zastosowaniach, o czym będzie mowa w następnym rozdziale). Z drugiej strony, symetryczna i liniowa charakterystyka $I(U)$ złącza *m-s* jest pożądana dla prawidłowej pracy wszystkich elementów półprzewodnikowych, gdyż mają one doprowadzenia metalowe, a kontakty metalu z półprzewodnikiem nie powinny w tym przypadku wpływać na charakterystyki elementu właściwego. Właściwości złącza *m-s* zależą głównie od różnicy prac wyjścia elektronów z metalu i półprzewodnika oraz stanów powierzchniowych półprzewodnika. Najpierw rozpatrzmy uproszczony model fizyczny złącza metal-półprzewodnik przy pominięciu wpływu stanów powierzchniowych. Taki prosty model umożliwia łatwe zrozumienie najbardziej zasadniczych aspektów pracy złącza *m-s*. Przede wszystkim umożliwia on wyjaśnienie fizycznych przyczyn nieliniowości lub liniowości charakterystyk prądowo-napięciowych. Jednakże w pełni prawidłowy obraz zjawisk występujących w złączu *m-s* otrzymamy dopiero po uwzględnieniu wpływu stanów powierzchniowych. Należy nadmienić, że w opisie działania złącza *m-s* występuje wiele pojęć wprowadzonych w poprzednim punkcie dotyczącym struktury MIS.

Uproszczony model złącza *m-s*

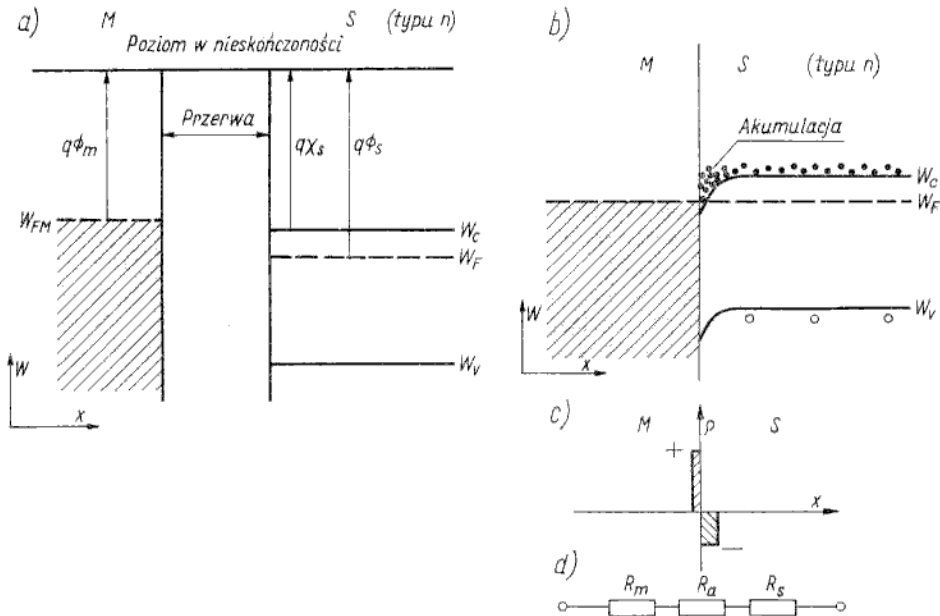
3.4.1

Jeśli pominiemy wpływ stanów powierzchniowych, to właściwości złącza *m-s* zależą od prac wyjścia elektronów z metalu i półprzewodnika oraz od rodzaju przewodnictwa półprzewodnika.

W związku z tym istnieją następujące cztery możliwości:

- | | |
|--|-------------------------------|
| 1 — praca wyjścia elektronów z metalu mniejsza niż z półprzewodnika ($ \Phi_m < \Phi_s $); | } półprzewodnik typu <i>n</i> |
| 2 — praca wyjścia elektronów z metalu większa niż z półprzewodnika ($ \Phi_m > \Phi_s $); | |
| 3 — praca wyjścia elektronów z metalu mniejsza niż z półprzewodnika ($ \Phi_m < \Phi_s $); | } półprzewodnik typu <i>p</i> |
| 4 — praca wyjścia elektronów z metalu większa niż z półprzewodnika ($ \Phi_m > \Phi_s $). | |

Rozpatrzmy układy 1, 2 dla półprzewodnika typu *n*.

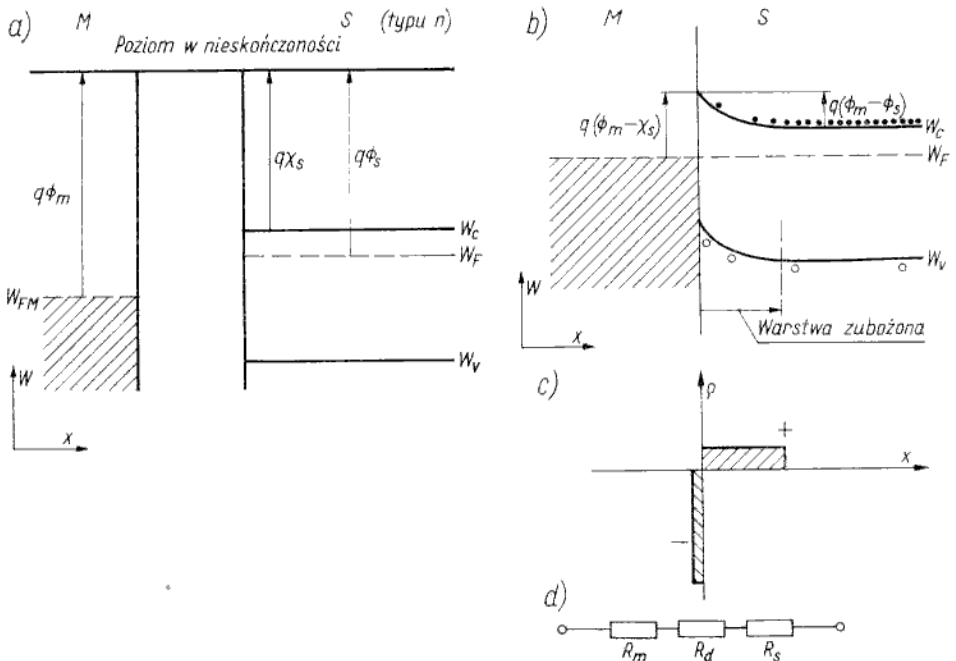


Rys. 3.68

Energetyczny model pasmowy metalu i półprzewodnika ($|\Phi_m| < |\Phi_s|$): a) przed połączeniem; b) po połączeniu; c) rozkład ładunku w złączu; d) rezystancyjny schemat zastępczy struktury *m-s*

Na rysunku 3.68a przedstawiono energetyczny model pasmowy metalu i półprzewodnika przed ich połączeniem. Jest to model analogiczny jak dla struktury MIS w stanie nierównowagi. Poziomy Fermiego w metalu i półprzewodniku mają różne wartości, a ponieważ poziom w nieskończoności jest wspólny, prace wyjścia elektronów różnią się. Jest to przypadek 1, gdy $|\Phi_m| < |\Phi_s|$. Po zetknięciu metalu z półprzewodnikiem (rys. 3.68b) układ metal-półprzewodnik dąży do stanu rów-

nowagi termodynamicznej przez odpowiednie przegrupowanie elektronów. Ponieważ praca wyjścia z metalu jest mniejsza, więcej elektronów „przechodzi” z metalu do półprzewodnika niż w kierunku przeciwnym. Wskutek tego przy powierzchni półprzewodnika gromadzą się elektrony tworząc ujemny ładunek przestrzenny, natomiast przy powierzchni metalu ubytek elektronów powoduje powstanie dodatniego ładunku przestrzennego (rys. 3.68c). Dipolowa warstwa ładunku przestrzennego wytwarza kontaktową różnicę potencjałów, spełniającą funkcję dodatkowej bariery energetycznej dla elektronów wychodzących z metalu do półprzewodnika. W związku z tym liczba elektronów wychodzących z metalu do półprzewodnika maleje, aż ustalą się jednakowe wartości prawdopodobieństw przejścia elektronu z metalu do półprzewodnika i przejścia w kierunku przeciwnym. Po ustaleniu się zatem stanu równowagi termodynamicznej w całej strukturze *m-s* jest jeden wspólny poziom Fermiego, w obszarze granicznym złącza istnieje dipolowa warstwa ładunku nieskompensowanego, będąca źródłem kontaktowej różnicy potencjałów, a strumienie nośników, płynące z metalu do półprzewodnika i w kierunku odwrotnym, wzajemnie się równoważą. Przy stałej wartości poziomu Fermiego większa koncentracja elektronów przy powierzchni półprzewodnika niż w głębi oznacza w „języku” modelu pasmowego zagięcie poziomu dna pasma przewodnictwa „w dół” (większą koncentrację elektronów w pasmie przewodnictwa uzyskuje się przy mniejszej odległości poziomów W_c , W_F). Identycznie zagina się poziom wierzchołka pasma walencyjnego. Omawiane złącze składa się z trzech obszarów o różnych konduktyw-



Rys. 3.69

Energetyczny model pasmowy metalu i półprzewodnika ($|\Phi_m| > |\Phi_s|$): a) przed połączeniem; b) po połączeniu; c) rozkład ładunku w złączu; d) rezystancyjny schemat zastępczy struktury *m-s*

nościach σ , tj. warstwa metalu (σ_m największa), warstwa akumulacyjna przy powierzchni półprzewodnika (σ_a średnia), obszar obojętny półprzewodnika (σ_s najmniejsza). Dlatego można mówić o szeregowym połączeniu trzech rezystancji (rys. 3.68d), tj. R_m, R_a, R_s . Ponieważ $R_s \gg R_a \gg R_m$, zatem rezystancja całkowita złącza jest określona wartością R_s . W połączeniu szeregowym R_m, R_a, R_s przy polaryzacji złącza napięciem zewnętrznym zmienia się tylko wartość R_a , gdyż zmienia się koncentracja elektronów zakumulowanych przy powierzchni półprzewodnika. Jednakże R_a pozostaje mniejsza niż R_s , a ponieważ R_s nie zależy od napięcia zewnętrznego, więc charakterystyka prądowo-napięciowa omawianego złącza $m-s$ jest, praktycznie biorąc, liniowa. Takie złącze nazywa się *liniowym*.

Na rysunku 3.69 przedstawiono przypadek 2, gdy $|\Phi_m| > |\Phi_s|$. Ponieważ praca wyjścia elektronów z półprzewodnika jest mniejsza, zatem po zetknięciu metalu z półprzewodnikiem więcej elektronów „przechodzi” z półprzewodnika do metalu niż w kierunku przeciwnym. Wskutek tego przy powierzchni metalu gromadzą się elektrony, przy powierzchni półprzewodnika natomiast ubytek elektronów powoduje powstanie dodatniego ładunku przestrzennego nieskompensowanych elektrycznie centrów donorowych (rys. 3.69c). Dipolowa warstwa ładunku przestrzennego wytwarza kontaktową różnicę potencjałów, spełniającą funkcję dodatkowej bariery energetycznej dla elektronów wychodzących z półprzewodnika do metalu. W związku z tym liczba elektronów wychodzących z półprzewodnika do metalu maleje, aż ustalą się jednakowe wartości prawdopodobieństw przejścia elektronu z półprzewodnika do metalu i przejścia w kierunku przeciwnym. Po ustaleniu się zatem stanu równowagi termodynamicznej w całej strukturze $m-s$ jest jeden wspólny poziom Fermiego, w obszarze granicznym złącza istnieje dipolowa warstwa ładunku nieskompensowanego, będąca źródłem kontaktowej różnicy potencjałów, a strumienie nośników, płynące z półprzewodnika do metalu i w kierunku odwrotnym, wzajemnie się równoważą. Przy stałej wartości poziomu Fermiego mniejsza koncentracja elektronów przy powierzchni półprzewodnika niż w głębi oznacza w „języku” modelu pasmowego zagięcie poziomu dna pasma przewodnictwa „do góry” (mniejszą koncentrację elektronów w pasmie przewodnictwa uzyskuje się przy większej odległości poziomów W_c, W_F). Omawiane złącze składa się z trzech obszarów o różnych konduktywnościach σ , tj. warstwa metalu (σ_m największa), warstwa zubożona przy powierzchni półprzewodnika (σ_d najmniejsza, gdyż jest to warstwa o najmniejszej koncentracji nośników prądu), obszar obojętny półprzewodnika (σ_s średnia). Dlatego można mówić o szeregowym połączeniu trzech rezystancji (rys. 3.69d), tj. R_m, R_d, R_s . Ponieważ $R_d \gg R_s \gg R_m$, zatem rezystancja całkowita złącza jest określona wartością R_d . W połączeniu szeregowym R_m, R_d, R_s przy polaryzacji złącza napięciem zewnętrznym zmienia się tylko wartość R_d , gdyż zmienia się koncentracja elektronów w warstwie zubożonej. Ponieważ R_d określa, praktycznie biorąc, rezystancję całego złącza, charakterystyka napięciowo-prądowa omawianego złącza $m-s$ jest nieliniowa. Dalej wykazemy że jest to charakterystyka podobna jak dla złącza $p-n$, czyli istnieje kierunek przewodzenia i kierunek zaporowy. Takie złącze nazywa się *prostującymi*.

W podobny sposób można wyjaśnić powstawanie złącza liniowego lub prostującego w przypadku styku metalu z półprzewodnikiem typu p . Wówczas dla $|\Phi_m| > |\Phi_s|$ uzyskuje się złącze liniowe, a dla $|\Phi_m| < |\Phi_s|$ — złącze prostujące.

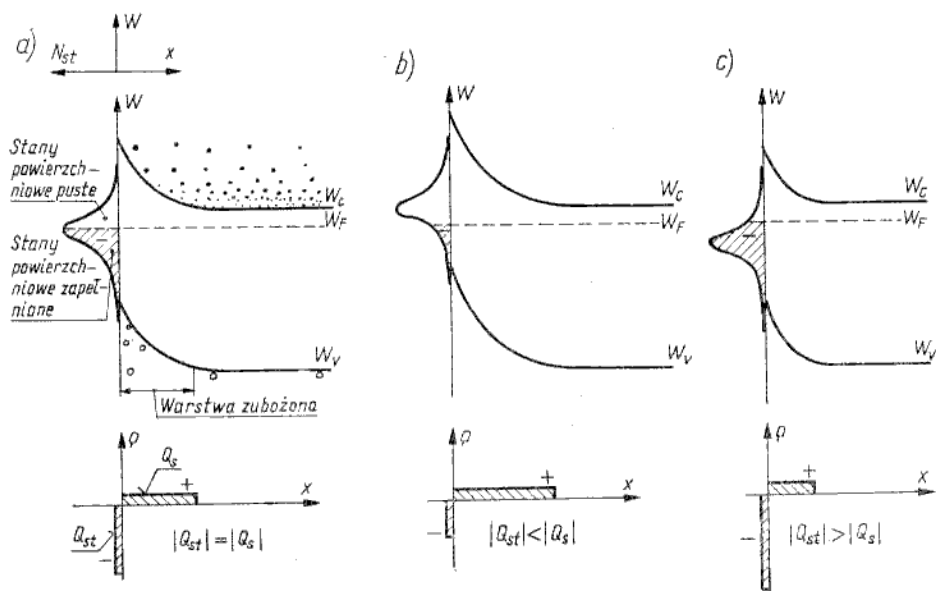
Rozpatrzmy teraz wpływ stanów powierzchniowych na zjawiska występujące w złączu $m-s$.

Model złącza *m-s* z uwzględnieniem stanów powierzchniowych

3.4.2

Na podstawie uproszczonego modelu złącza *m-s* można by sądzić, że w celu uzyskania złącza o pożądanym właściwościach wystarczy dobrać materiały (metal i półprzewodnik) odpowiednie pod względem wartości prac wyjścia elektronów. Jednakże okazuje się, iż bariera energetyczna w rzeczywistym złączu *m-s* w małym stopniu zależy od różnicy prac wyjścia. Złącza metal-przewodnik typu *n* (Ge, Si, GaAs) przeważnie wykazują właściwości prostownicze niezależnie od rodzaju metalu. Nie wystarcza odpowiednia relacja między wartościami prac wyjścia, lecz konieczne są dodatkowe operacje technologiczne, by otrzymać złącze *m-s* o liniowej charakterystyce prądowo-napięciowej.

Dla wyjaśnienia tych właściwości należy uwzględnić wpływ stanów powierzchniowych półprzewodnika. Na rysunku 3.70 przedstawiono model energetyczny półprze-

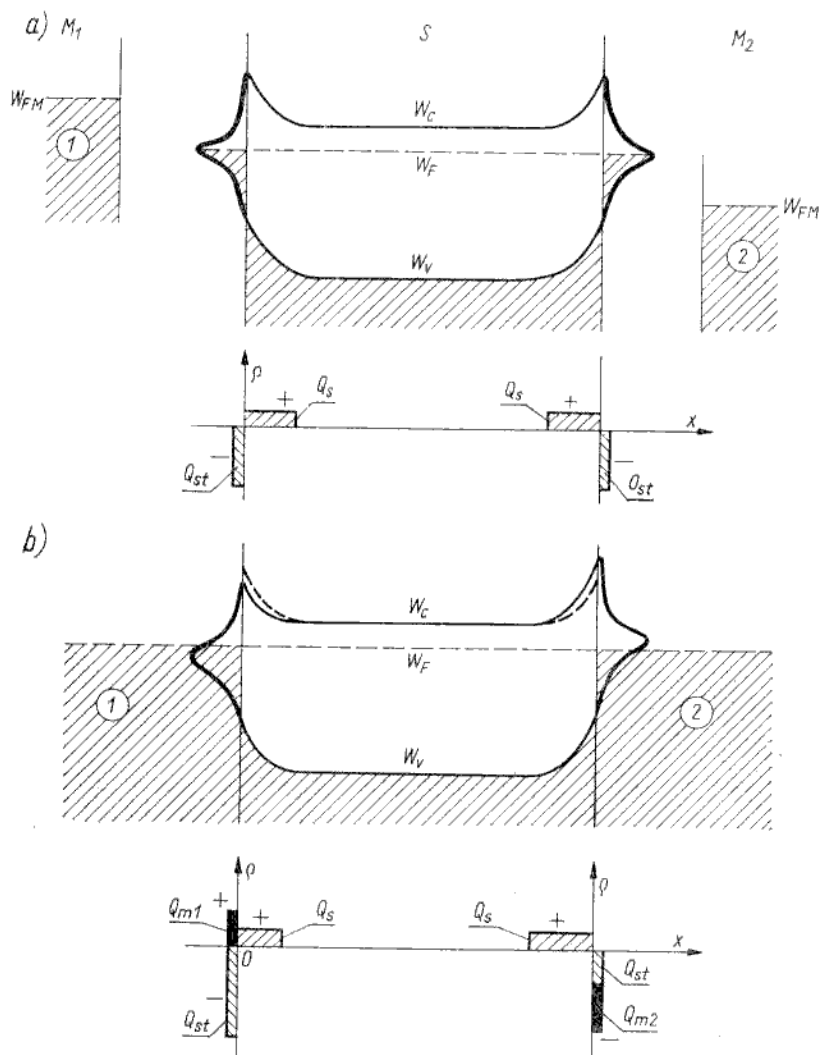


Rys. 3.70
Energetyczny model pasmowy struktury *m-s* ze stanami powierzchniowymi: a) stan realny; b, c) nierealne stany hipotetyczne

wodnika ze stanami powierzchniowymi. Dla germanu, krzemu, arsenku galu stany powierzchniowe mają charakter akceptorowy, tj. przechwytyją elektrony z półprzewodnika jonizując się ujemnie. Zatem przy powierzchni półprzewodnika typu *n* wytwarza się warstwa zubożona, czyli pasma energetyczne zaginają się „do góry”. Przy dużej gęstości stanów powierzchniowych, ulokowanych energetycznie w pobliżu środka pasma zabronionego, poziom Fermiego przy powierzchni półprzewodnika jest jakby „przywiązany” do poziomu stanów powierzchniowych, tj. znajduje się w pobliżu środka pasma zabronionego (rys. 3.70a). Łatwo można to uzasadnić odwołując się do ilustracji pewnych hipotetycznych sytuacji przedstawionych na rys. 3.70b, c. Gdyby poziom Fermiego znajdował się poniżej środka pasma zabronionego (poniżej maksimum gęstości stanów powierzchni-

wych), wówczas ładunek stanów powierzchniowych byłby mniejszy niż ładunek warstwy zubożonej (rys. 3.70b). Taka sytuacja nie jest możliwa w stanie równowagi termodynamicznej, gdyż ładunek dodatni warstwy zubożonej wytworzy w półprzewodniku pole elektryczne przyciągające elektrony ku powierzchni, te zaś częściowo skompensują ładunek centrów donorowych (zmaleje Q_s), a częściowo zapełnią stany powierzchniowe (wzrośnie Q_{st}), co doprowadzi do ustalenia się stanu przedstawionego na rys. 3.70a. Podobnie można wykazać, że nie jest możliwa sytuacja przedstawiona na rys. 3.70c.

Rozpatrzmy teraz zjawiska występujące na styku metalu z półprzewodnikiem, w którym istnieją zagięcie poziomów energetycznych, spowodowane stanami powierzch-



Rys. 3.71

Energetyczny model pasmowy struktury *m-s* i rozkład gęstości ładunku: a) przed zetknięciem metalu z półprzewodnikiem; b) po zetknięciu

niowymi. Na rysunku 3.71 przedstawiono model energetyczny półprzewodnika oraz dwu różnych metali przed i po zetknięciu półprzewodnika z tymi metalami. Metal 1 charakteryzuje się pracą wyjścia mniejszą, metal 2 zaś — większą niż półprzewodnik ($|\Phi_{m1}| < |\Phi_s| < |\Phi_{m2}|$). Zgodnie z uproszczonym modelem złącze metal 1-półprzewodnik powinno mieć właściwości liniowe, natomiast złącze metal 2-półprzewodnik powinno mieć właściwości prostownicze. Jednakże stany powierzchniowe powodują, że oba złącza mają właściwości prostownicze. Dlaczego? Rozważmy najpierw styk m_1 -s. Wiadomo, że wskutek relacji $|\Phi_{m1}| < |\Phi_s|$ po zetknięciu metalu z półprzewodnikiem więcej elektronów będzie wychodził z metalu do półprzewodnika niż w kierunku przeciwnym. Niewielka tylko jednak liczba tych elektronów znajdzie się w półprzewodniku, neutralizując częściowo ładunek warstwy zubożonej, wiele natomiast elektronów przechwyca stany powierzchniowe. Przed zetknięciem metalu z półprzewodnikiem bilans ładunku miał postać

$$Q_s - Q_{st} = 0$$

a po utworzeniu styku

$$Q_s - \Delta Q_s - Q_{st} - \Delta Q_{st} + Q_{m1} = 0$$

przy czym

$$Q_{m1} = \Delta Q_s + \Delta Q_{st}$$

Ponieważ gęstość stanów powierzchniowych jest bardzo duża, więc nieznaczna zmiana zagięcia poziomów energetycznych powoduje duże przyrosty ładunku Q_{st} , czyli $\Delta Q_{st} \gg \Delta Q_s$. Dlatego bariera energetyczna, istniejąca przy powierzchni półprzewodnika, wskutek działania stanów powierzchniowych, tylko nieznacznie maleje przy zetknięciu z metalem m_1 i złącze ma właściwości prostownicze.

Podobne rozumowanie można przeprowadzić dla złącza m_2 -s, przy czym w tym przypadku bariera energetyczna nieznacznie się zwiększa.

W ten sposób przekonaliśmy się, że rzeczywiste złącze metal-półprzewodnik typu n ma właściwości prostownicze niezależne od rodzaju metalu. W następnym punkcie rozpatrzmy nieliniową charakterystykę prądowo-napięciową takiego złącza, a w p. 3.4.4 wyjaśnimy, w jaki sposób można mimo wszystko uzyskać liniowe złącze z półprzewodnikiem typu n .

Charakterystyka prądowo-napięciowa złącza prostującego

3.4.3

Charakterystyka prądowo-napięciowa złącza metalu z półprzewodnikiem typu n ma kształt podobny jak dla złącza p - n , to znaczy istnieje kierunek przewodzenia o wykładniczej zależności prądu od napięcia oraz kierunek zaporowy, charakteryzujący się bardzo małym prądem, praktycznie biorąc, niezależnym od napięcia. Dokładna analiza ilościowa zależności $I(U)$ w złączu m - s prostującym nie jest łatwa z uwagi na konieczność uwzględnienia wielu czynników, spośród których do najistotniejszych należą:

- stany powierzchniowe,
- różnica prac wyjścia,
- cienka warstwa dielektryczna między metalem a półprzewodnikiem,
- emisja termoelektryczna,

- tunelowanie nośników,
- dyfuzja nośników.

Spróbujmy określić zależność $I(U)$ na podstawie rozważań jakościowych. Wiadomo, że nośniki przechodząc z metalu do półprzewodnika lub w kierunku przeciwnym pokonują bariery energetyczne. Wartość prądu jest zatem zależna od wysokości tych barier, a kształt charakterystyki $I(U)$ wynika z określonej zależności barier od napięcia zewnętrznego. Wysokości barier można łatwo wyznaczyć tylko dla uproszczonego modelu złącza $m-s$. Wówczas:

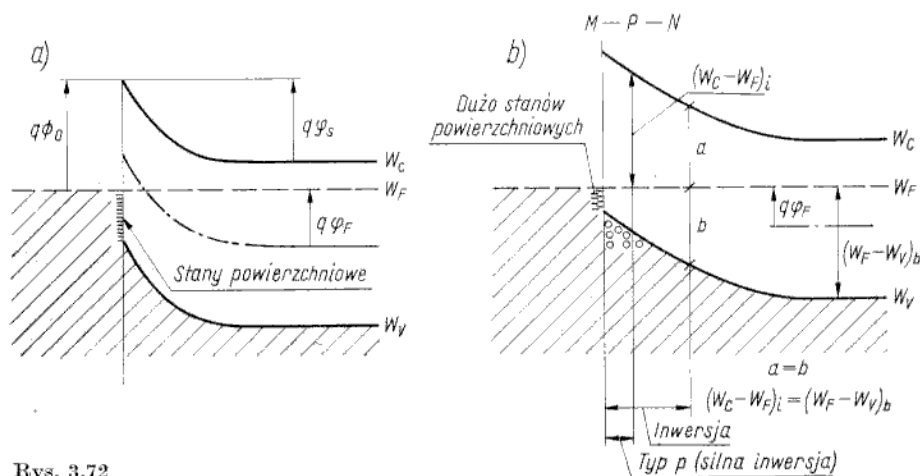
$q\Phi_0 = q(\Phi_m - \chi_s)$ — bariera dla kierunku metal \rightarrow półprzewodnik;

$q\varphi_s = q(\Phi_m - \Phi_s)$ — bariera dla kierunku półprzewodnik \rightarrow metal (potencjał powierzchniowy półprzewodnika).

W rzeczywistym złączu $m-s$ bardzo istotny wpływ na wysokości barier wywierają stany powierzchniowe. Jak wykazano w p. 3.4.2, potencjał powierzchniowy φ_s , czyli bariera dla elektronów przechodzących z półprzewodnika do metalu, jest określony położeniem energetycznym maksimum gęstości stanów powierzchniowych. W p. 3.4.2 przyjęto, że to maksimum znajduje się w środku pasma zabronionego, poziom Fermiego jest zatem przy powierzchni jakby „związany” ze środkiem pasma zabronionego. Ponieważ poziom Fermiego musi mieć jednakową wartość w całej strukturze $m-s$ (warunek równowagi termodynamicznej), następuje odpowiednie zagięcie poziomów W_c , W_v . Doświadczalnie stwierdzono, że w złączach różnych metali (Al, Au, Ag, Pt i inne) z Ge, Si, GaAs istnieje silnie zaznaczone maksimum gęstości stanów powierzchniowych, ułożone energetycznie w odległości około $(0,3 \dots 0,4)W_g$ nad wierzchołkiem pasma walencyjnego. Oznacza to, że poziom Fermiego przy powierzchni półprzewodnika lokuje się w odległości $(0,1 \dots 0,2)W_g$ poniżej środka pasma zabronionego. Stąd łatwo można wyznaczyć obie bariery. Na przykład dla Si o typowej koncentracji domieszek ($10^{21} \dots 10^{22} \text{ m}^{-3}$) potencjał $\varphi_F \approx 0,3 \text{ V}$, zatem (rys. 3.72a):

$$q\Phi_0 = \frac{1}{2}W_g + (0,1 \dots 0,2)W_g = 0,66 \dots 0,77 \text{ eV}$$

$$q\varphi_s = q\varphi_F + (0,1 \dots 0,2)W_g = 0,41 \dots 0,52 \text{ eV}$$



Rys. 3.72

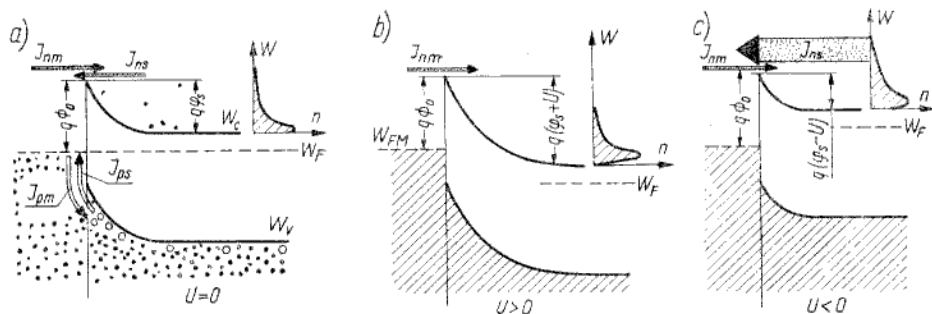
Bariery w $m-s$ złączu ze stanami powierzchniowymi: a) bariera Schottky'ego; b) złącze z inwersją

Warto już w tym miejscu podkreślić, że potencjał φ_s w złączu $m-s$ odpowiada napięciu dyfuzyjnemu φ_B w złączu $p-n$. Wiadomo, że napięcie dyfuzyjne dla typowego złącza krzemowego ma wartość ok. 0,7 V i określa spadek napięcia na złączu $p-n$ przy pracy w kierunku przewodzenia. Ponieważ $\varphi_s < \varphi_B$, spadek napięcia na złączu $m-s$, pracującym w kierunku przewodzenia, będzie mniejszy. Jest to bardzo ważne spostrzeżenie, do którego powrócimy przy omawianiu zastosowań diod Schottky'ego.

Na rysunku 3.72b przedstawiono model pasmowy złącza $m-s$, w którym wskutek bardzo dużej gęstości stanów powierzchniowych, ułożonych w pobliżu wierzchołka pasma walencyjnego, następuje silne zagięcie pasm energetycznych i powstaje warstwa inwersyjna. Taki stan jest typowy dla niektórych złączy ostrzowych (styk igły metalowej z półprzewodnikiem). W tym przypadku występuje swobodny przepływ dziur między metalem a warstwą typu p , a efekt diodowy jest związany ze złączem $p-n$ formującym się w półprzewodniku. Dlatego dalsze rozważania nie dotyczą złączy ostrzowych $m-s$, nie należących do grupy złączy z barierą Schottky'ego.

W rzeczywistych złączach $m-s$ pewien wpływ na wysokość bariery energetycznej ma również cienka warstwa dielektryczna (kilka warstw atomowych), która jest, praktycznie biorąc, „przezroczysta” dla nośników przechodzących przez nią tunelemo.

Przyjmujemy, że mamy określony model energetyczny złącza $m-s$ w stanie bez polaryzacji. Jakie strumienie nośników płyną w tym złączu i jak one zależą od napięcia zewnętrznego (rys. 3.73)?



Rys. 3.73

Strumienie nośników płynące w złączu $m-s$; a) bez polaryzacji; b) polaryzacja w kierunku zaporowym; c) polaryzacja w kierunku przewodzenia

Na rysunku 3.73a zilustrowano obrazowo cztery składowe prądu płynącego przez złącze.

Są to dwa strumienie elektronów:

J_{nm} — strumień elektronów przechodzących ponad barierą $q\varphi_0$ z metalu do półprzewodnika (prąd emisji termoelektrycznej),

J_{ns} — strumień elektronów przechodzących ponad barierą $q\varphi_s$ z półprzewodnika do metalu;

oraz dwa strumienie dziur:

J_{pm} — strumień dziur płynących z metalu do półprzewodnika (jest to inaczej strumień elektronów przechodzących z pasma walencyjnego półprzewodnika do pustych poziomów energetycznych w metalu),

J_{ps} — strumień dziur płynących z półprzewodnika do metalu (jest to inaczej strumień elektronów przechodzących z pasma walencyjnego metalu do pustych poziomów w pasmie walencyjnym półprzewodnika).

W stanie równowagi termodynamicznej ($U = 0$):

$$J_{nm} + J_{ns} = 0$$

$$J_{pm} + J_{ps} = 0$$

Ponieważ strumienie dziur są bardzo małe i prawie nie zmieniają się przy polaryzacji zewnętrznej, można pominąć ich wpływ w dalszych rozważaniach ($J_{pm} + J_{ps} \approx 0$ dla dowolnych U).

Przy polaryzacji w kierunku zaporowym („+” do półprzewodnika, a „-” do metalu, czyli w stosunku do uziemionego metalu $U > 0$) mamy sytuację przedstawioną na rys. 3.73b. Bariera $q\Phi_0$ dla elektronów przechodzących z metalu do półprzewodnika nie zmienia się (praktycznie biorąc, napięcie zewnętrzne w całości odkłada się na warstwie zubożonej półprzewodnika), czyli prąd emisji termoelektrycznej J_{nm} ma taką wartość jak przy $U = 0$. Zwiększa się natomiast bariera potencjału dla elektronów przechodzących z półprzewodnika do metalu, tj. przyjmuje wartość $q(\varphi_s + U)$. A więc praktycznie niemożliwy jest przepływ elektronów w kierunku $s \rightarrow m$, czyli $J_{ns} = 0$. W związku z tym całkowity prąd w złączu $m-s$ spolaryzowanym zaporowo przyjmuje wartość J_{nm} niezależną od napięcia (jest to odpowiednik prądu nasycenia J_s w złączu $p-n$).

Przy polaryzacji w kierunku przewodzenia („+” do metalu, a „-” do półprzewodnika, czyli w stosunku do uziemionego metalu $U < 0$) występuje sytuacja przedstawiona na rys. 3.73c. Bariera $q\Phi_0$ dla elektronów przechodzących z metalu do półprzewodnika nie zmienia się (praktycznie biorąc, napięcie zewnętrzne w całości odkłada się na warstwie zubożonej półprzewodnika), czyli prąd emisji termoelektrycznej J_{nm} ma taką wartość jak przy $U = 0$. Maleje natomiast bariera potencjału dla elektronów przechodzących z półprzewodnika do metalu, tj. przyjmuje ona wartość $q(\varphi_s - U)$. Z wykresu $n(W)$ zamieszczonego na rys. 3.73a, b, c doskonale widać, że nawet nieznaczne zmniejszenie bariery oznacza bardzo duży przyrost koncentracji elektronów ułokowanych energetycznie powyżej bariery, a te elektrony tworzą prąd J_{ns} . Zatem prąd J_{ns} wzrasta gwałtownie. Ponieważ koncentracja elektronów zależy wykładniczo od energii (zależność Boltzmann), więc prąd J_{ns} zwiększa się wykładniczo w funkcji napięcia polaryzacji

$$J_{ns} \sim \exp\left(\frac{U}{\varphi_T}\right)$$

Ponieważ przy $U = 0$

$$J_{ns} = J_{nm}$$

zatem

$$J_{ns} = J_{nm} \exp\left(\frac{U}{\varphi_T}\right)$$

Całkowity prąd

$$J = J_{ns} - J_{nm}$$

czyli

$$J = J_{nm} \left[\exp\left(\frac{U}{\varphi_T}\right) - 1 \right] \quad (3.107)$$

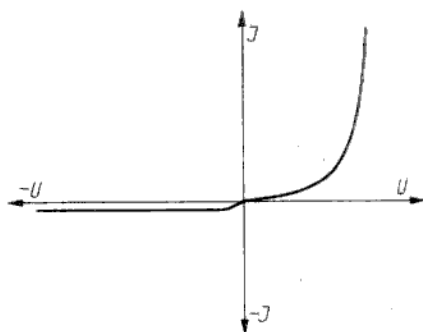
Jest to równanie identyczne jak dla złącza $p-n$.

Wyrażenie na prąd emisji termoelektrycznej J_{nm} stanowi przedmiot rozważań bardziej zaawansowanych kursów fizyki ciała stałego, dlatego w tym miejscu przytoczymy tylko końcową postać takiego wyprowadzenia, w której intuicyjnie oczywista jest obecność czynnika wykładniczej zależności od bariery $q\Phi_0$

$$J_{nm} = A^* T^2 \exp\left(-\frac{\Phi_0}{\varphi_T}\right) \quad (3.108)$$

przy czym A^* — stała Richardsona.

Wyrażenia (3.107) wraz z (3.108) dość dobrze opisują charakterystykę prądowo-napięciową złącza $m-s$ (rys. 3.74). Dla większości rzeczywistych złączy $m-s$ bardzo



Rys. 3.74
Charakterystyka prądowo-napięciowa złącza $m-s$ z barierą Schottky'ego

dobrą zgodność równania teoretycznego z charakterystyką doświadczalną uzyskuje się po wprowadzeniu niewielkiej modyfikacji równań (3.107) i (3.108)

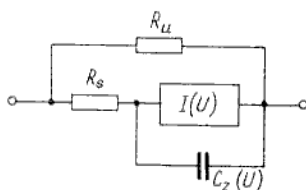
$$J = \underbrace{A^* T^2}_{J_{nm}} \exp\left(-\frac{\Phi_0 - \Delta\Phi_0(U)}{\varphi_T}\right) \left[\exp\left(\frac{U}{m\varphi_T}\right) - 1 \right] \quad (3.109)$$

W równaniu (3.109) wprowadzono współczynnik m , który jest minimalnie większy niż jedność (najczęściej $m = 1,06$), oraz poprawkę $\Delta\Phi_0(U)$, wynikającą z tzw. efektu Schottky'ego. *Efekt Schottky'ego* jest to obniżenie bariery Schottky'ego ($q\Phi_0$) spowodowane zewnętrznym polem elektrycznym, jeżeli uwzględni się oddziaływanie elektronu wychodzącego z metalu z ładunkiem obrazowym powstającym w metalu. Uwzględnienie tego zjawiska ma istotny wpływ na kształt charakterystyki $I(U)$ w kierunku zaporowym. Zwykle $\Delta\Phi_0(U)$ nie przekracza kilkudziesięciu miliwoltów.

W przypadkach złączy $m-s$ wykonanych niestarannie (zła jakość powierzchni półprzewodnika, zbyt gruba warstwa dielektryka między półprzewodnikiem a metalem) rzeczywista charakterystyka $I(U)$ może znacznie różnić się od opisanej równaniem (3.109). Dla uzgodnienia teorii z doświadczeniem w tych przypadkach najczęściej trzeba zwiększyć wartość współczynnika m , niekiedy do $m = 3$.

Spróbujmy teraz przedstawić schemat zastępczy złącza $m-s$. Analogicznie jak dla złącza $p-n$, oprócz nieliniowego źródła prądowego $I(U)$, opisanego równaniem (3.109), należy uwzględnić rezystancję szeregową (w tym przypadku chodzi o rezystancję szeregową obszaru neutralnego półprzewodnika), rezystancję równoległą upływu oraz pojemności złącza. W tym miejscu czeka nas bardzo miła niespodzianka. Otóż złącze $m-s$ nie ma w ogóle pojemności dyfuzyyjnej. Wiadomo, że pojemność dyfuzyjna w złączu $p-n$ była spowodowana magazynowaniem nadmiarowego ładunku nośników mniejszościowych w bazie złącza przy pracy w kierunku przewodzenia. W przypadku złącza $m-s$ nośniki (elektrony) są wstrzykiwane z półprzewodnika do metalu, gdzie i tak jest olbrzymia liczba elektronów. W związku z tym przepływ prądu w kierunku przewodzenia, tj. wstrzykiwanie elektronów z półprzewodnika do metalu, nie powoduje choćby zauważalnych zmian ładunku elektronów w metalu. Elektrony wstrzykiwane to jakby deszczyk majowy spadający do oceanu „własnych” elektronów metalu. To porównanie jest tym trafniejsze, że z półprzewodnika do metalu przechodzą elektrony o energii większej niż energia bariery $q\varphi_s$, są to zatem elektrony o energii $q\Phi_0$ większej niż energia Fermiego w metalu, tj. o energii większej niż średnia energia „własnych” elektronów metalu. Dlatego mówi się często, że do metalu są wstrzykiwane *elektrony gorące*, a złącze z barierą Schottky’ego jest również nazywane *złączem na gorących elektronach*. Oczywiście elektrony gorące po wstrzyknięciu natychmiast się „oziębiają”, oddając swój nadmiar energii do sieci krystalicznej metalu (przez zderzenia z atomami w węzłach sieci krystalicznej metalu). Ponieważ nie ma magazynowania ładunku w metalu, nie ma też pojemności dyfuzyyjnej.

Brak pojemności dyfuzyyjnej jest okolicznością bardzo korzystną, gdyż umożliwia pracę złącza $m-s$ w zakresie bardzo wielkich częstotliwości. Pozostaje natomiast pojemność warstwy zubożonej, która jest określona wzorem (3.54) przy zamianie φ_B na φ_s . Pełny model złącza $m-s$ przedstawiono na rys. 3.75. Jest to model dy-



Rys. 3.75
Schemat zastępczy złącza $m-s$

namiczny separowalny, tj. przez odłączenie elementu C_z , reprezentującego pojemność warstwy zubożonej, otrzymuje się model statyczny, w którym R_u ma istotny wpływ przy pracy złącza w kierunku zaporowym, R_s zaś — w kierunku przewodzenia.

Złącze omowe

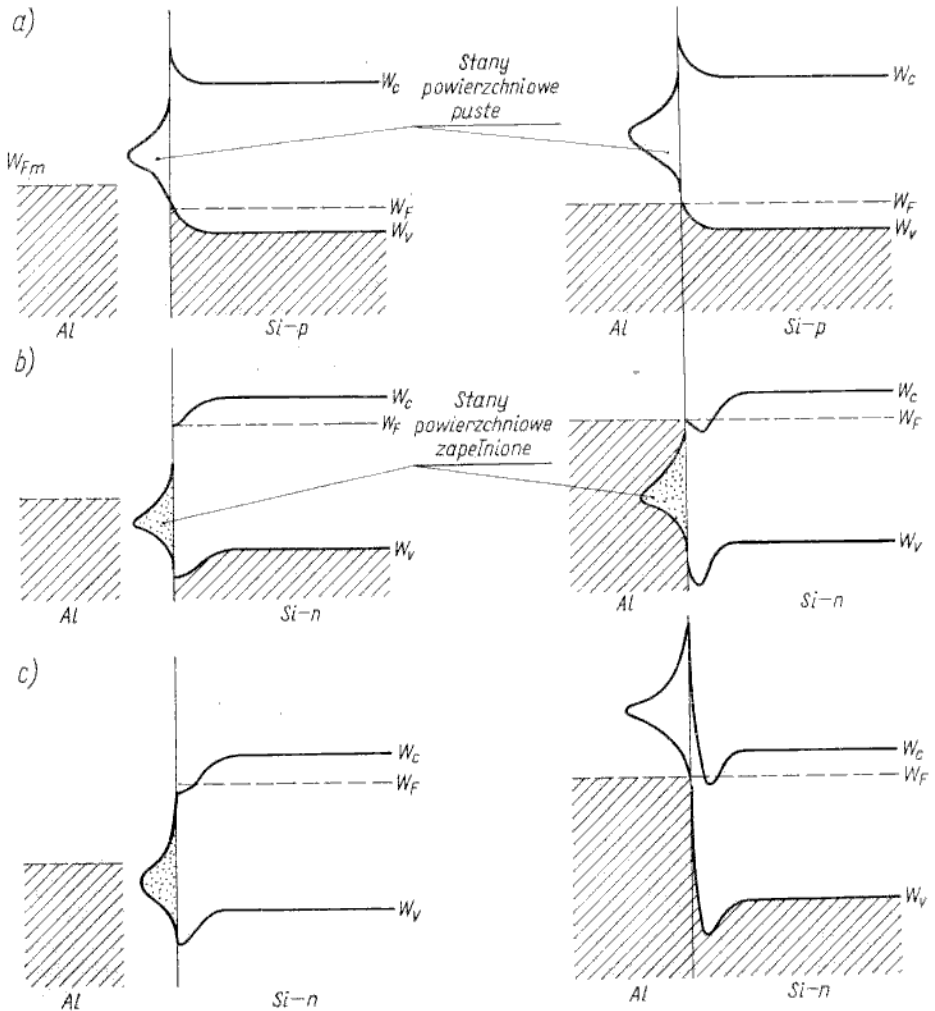
3.4.4

Według uproszczonego modelu złącza $m-s$ wydawało się, że otrzymanie złącza nieprostującego jest bardzo proste. W przypadku styku $m-p$ (metal-półprzewodnik typu p) wystarczy dobrać metal o pracy wyjścia większej niż dla półprzewodnika, a w przypadku styku $m-n$ (metal-półprzewodnik typu n) należy dobrać metal o pracy wyjścia mniejszej niż dla półprzewodnika. Z kolei po uwzględnieniu stanów powierzchniowych okazało się, że niezależnie od rodzaju metalu w złączu

$m-n$ powstaje warstwa zaporowa. Jednak i w tym przypadku możliwe jest uzyskanie złącza (kontaktu) omowego, jeżeli:

- warstwa zaporowa będzie tak cienka, że elektrony mogą przez nią przechodzić tunelowo
- lub
- zostanie wyeliminowany wpływ stanów powierzchniowych przez ich całkowite wypełnienie lub opróżnienie.

Kontakty omowe mają strukturę $m-p^+-p$ lub $m-n^+-n$. Zatem kontakt omowy wykonuje się w dwu fazach. Najpierw półprzewodnik jest bardzo silnie domieszkowany przy powierzchni tak, by powstała cienka warstwa zdegenerowanego półprzewodnika tego samego typu przewodnictwa jak w głębi (powstaje złącze $l-h$).



Rys. 3.76

Energetyczne modele pasmowe, ilustrujące powstawanie kontaktu omowego aluminium z krzemem typu p i typu n

Najczęściej stosowanym metalem jest aluminium, dlatego w drugiej fazie następuje naparowanie cienkiej warstwy aluminium i jej wtopienie (mikrodyfuzja Al w podwyższonej temperaturze). Po tym można łatwo wykonywać połączenia metal-metal z wyprowadzeniami zewnętrznymi. Zanim wykażemy, że tak wykonane struktury $m-p^+-p$ oraz $m-n^+-n$ mają właściwości omowe, uściślijmy najpierw, jakim warunkom powinien odpowiadać kontakt omowy.

Kontaktem omowym nazywa się złącze metalu z półprzewodnikiem, charakteryzujące się następującymi właściwościami:

- Liniowa i symetryczna charakterystyka prądowo-napięciowa (styk nieprostujący).
- Jak najmniejsza rezystancja.

Te właściwości można inaczej sformułować w postaci wymagań, jakie powinien spełniać kontakt omowy ze względu na nośniki mniejszościowe i większościowe:

- Nieskończenie duża szybkość rekombinacji nośników mniejszościowych, czyli czas życia $\tau \approx 0$ (a więc liniowość charakterystyki $I(U)$).
- Brak bariery dla nośników większościowych, czyli metal powinien spełniać funkcję niewyczerpalnego źródła i jednocześnie nienasyconego odbiornika nośników większościowych (a więc symetria charakterystyki $I(U)$).

Spełnieniu pierwszego warunku ($\tau = 0$) sprzyja silne domieszkowanie półprzewodnika i wtopienie aluminium (duża liczba defektów sieci krystalicznej przy powierzchni zwiększa szybkość rekombinacji).

Drugi natomiast warunek (brak bariery) jest spełniony dzięki silnemu domieszkowaniu, gdyż tylko w ten sposób można uzyskać niemal zrównanie się krawędzi pasma nośników większościowych z poziomem Fermiego metalu. (Tuż pod poziomem Fermiego znajduje się niewyczerpalne źródło elektronów, a tuż nad poziomem Fermiego — nienasycony odbiornik elektronów lub inaczej niewyczerpalny zbiornik wolnych stanów energetycznych).

Na rysunku 3.76 przedstawiono energetyczne modele pasmowe, ilustrujące powstawanie kontaktu omowego aluminium z krzemem typu p i typu n . Przedstawiono trzy przypadki:

- a) Kontakt aluminium z krzemem typu p . Przed połączeniem stany powierzchniowe są, praktycznie biorąc, puste wskutek silnego domieszkowania obszaru przypowierzchniowego akceptorami. Wtopienie Al zwiększa jeszcze bardziej koncentrację akceptorów, czyli ułatwia uzyskanie „zszycia” wierzchołka pasma walencyjnego z poziomem Fermiego.
- b) Kontakt aluminium z krzemem typu n dla przypadku mniejszej rozpuszczalności Al niż domieszki donorowej. Wskutek silnego domieszkowania donorami wszystkie stany powierzchniowe są zapełnione. Po wtopieniu Al (domieszka akceptorowa) następuje częściowe skompensowanie donorów, lecz w tak małym stopniu, że dno pasma przewodnictwa pozostaje, praktycznie biorąc, „zszycie” z poziomami Fermiego.
- c) Kontakt Al z krzemem typu n dla przypadku większej rozpuszczalności Al niż domieszki donorowej. W tym przypadku po wtopieniu Al wytwarza się bardzo cienka warstwa przypowierzchniowa typu p , powstaje zatem bariera energetyczna, jednak jest ona tak cienka, że elektrony swobodnie przenikają przez nią tunelowo.

Wykaz ważniejszych wzorów

Tablica 3.2

Szerokość warstwy zaporowej i bariera potencjału w złączu <i>p-n</i>	złącze skokowe	$l_d = \sqrt{\frac{2\epsilon_s}{q} \frac{N_D + N_A}{N_D N_A} (\varphi_B \pm U)}$ przy czym + dla kierunku zaporowego - dla kierunku przewodzenia $\varphi_B = \varphi_T \ln \frac{N_A N_D}{n_i^2}$
	złącze liniowe	$l_d = \sqrt[3]{\frac{12\epsilon_s}{qa} (\varphi_B \pm U)}$ przy czym ± jak wyżej $\varphi_B = 2\varphi_T \ln \frac{al_d}{2n_i}$
Charakterystyka prądowo-napięciowa złącza <i>p-n</i>	przybliżenie pierwszego stopnia	przy czym $J = J_s (e^{\frac{U}{\varphi_T}} - 1)$ $J_s = qn_i^2 \left(\sqrt{\frac{D_n}{\tau_n}} \frac{1}{N_A} + \sqrt{\frac{D_p}{\tau_p}} \frac{1}{N_D} \right)$
	wyrażenia dokładniejsze	przy czym: $1 \leq m \leq 2$ $J_R = J_s + J_g$ $J_g = \frac{1}{2} q \frac{n_i}{\tau} l_d$ J_s jak wyżej lub w innej postaci $J = J_g (e^{\frac{U}{2\varphi_T}} - 1) + J_s (e^{\frac{U}{\varphi_T}} - 1)$
Pojemności złącza <i>p-n</i> (na jednostkę powierzchni)	złączowa	$C_j = \frac{\epsilon_s}{l_d}$ na przykład dla niesymetrycznego złącza <i>p-n</i> o $N_A \gg N_D$ $C_j = \left[\frac{q\epsilon_s N_D}{2(\varphi_B - U)} \right]^{1/2}$ w innej ogólnej postaci $C_j = C_{j0} \left(1 + \frac{ U }{\varphi_B} \right)^{-1/n}$ przy czym C_{j0} oznacza C_j dla $U = 0$ $2 \leq n \leq 3$
	dyfuzyjna	$C_d = \frac{1}{2\varphi_T} (J + J_s) \tau$

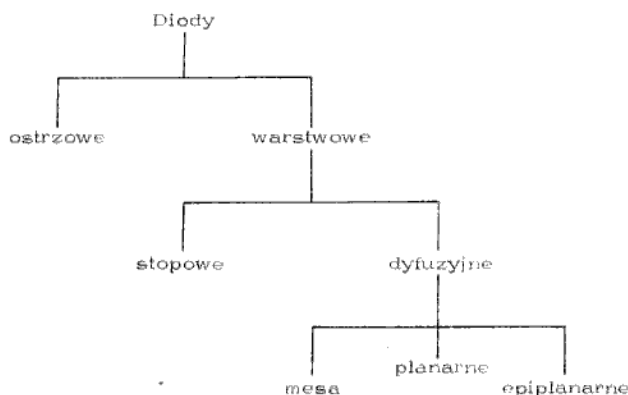
Złącza półprzewodnikowe, omawiane w poprzednim rozdziale pod kątem fizycznych podstaw ich działania, miały postać struktur półprzewodnikowych o bliżej niesprecyzowanych cechach konstrukcyjno-technologicznych i właściwościach użytkowych.

Jeżeli bryle półprzewodnika — będącej złączem — nada się konkretny „kształt” konstrukcyjny, czyli zdefiniuje jej rozmiary, wyposaży ją w doprowadzenia (końcówki), zamknie w określonej obudowie (oprawce) itp., to powstanie element dwukońcówkowy, nazywany *diodą*.

Różnica zatem między złączem półprzewodnikowym a diodą jest mniej więcej taka jak między rezystancją a rezystorem, pojemnością a kondensatorem, czy indukcyjnością a cewką.

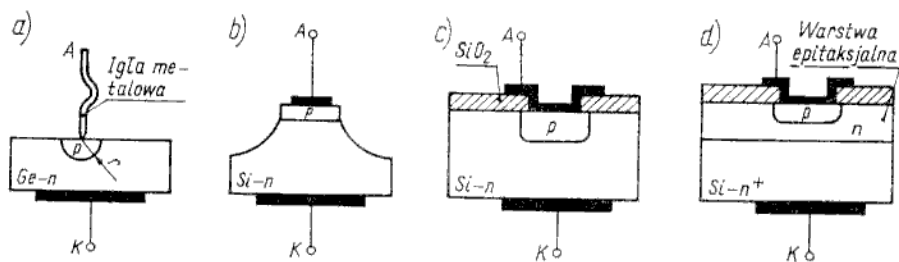
Celem tego rozdziału jest omówienie właściwości funkcjonalnych (częściowo również cech konstrukcyjnych) diod półprzewodnikowych, zbudowanych przy użyciu złączy *p-n* oraz *m-s*.

Diody można różnie klasyfikować w zależności od przyjętego kryterium. Jeżeli przyjmie się za kryterium podziału rodzaj materiału półprzewodnikowego, to można mówić o diodach germanowych, krzemowych i innych (np. z arsenku galu). Dla wytwórcy diod istotne są ich cechy konstrukcyjno-technologiczne. Z tego punktu widzenia diody dzieli się przede wszystkim na ostrzowe i warstwowe



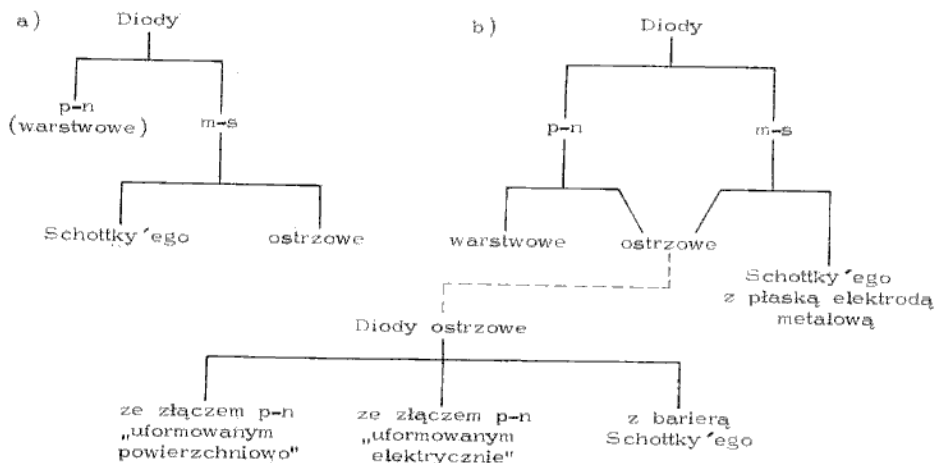
Rys. 4.1
Klasyfikacja diod ze względu na cechy konstrukcyjno-technologiczne

(nazywane również złączowymi) (rys. 4.1). Z kolei diody warstwowe można podzielić na stopowe i dyfuzyjne. Diody dyfuzyjne natomiast można podzielić na diody typu mesa, planarne i epiplanarne (rys. 4.2).



Rys. 4.2
Przekroje struktur diod: a) ostrzowej; b) mesa; c) planarnej; d) epiplanarnej

Można również przedstawić klasyfikację, w której nadrzędnym kryterium jest rodzaj struktury fizycznej złącza. Wydawałoby się, że taka klasyfikacja nie powinna sprawiać żadnych trudności, wiadomo wszak z poprzedniego rozdziału, że rozróżnia się złącza $p-n$, $m-s$ i heterozłącza. Jednak trudności pojawiają się przy próbie sklasyfikowania diod ostrzowych. Tradycyjnie diody ostrzowe zalicza się do grupy diod metal-półprzewodnik (rys. 4.3a). Wówczas diody $p-n$ są rozumiane jako



Rys. 4.3
Klasyfikacja diod ze względu na rodzaj struktury fizycznej złącza: a) podział tradycyjny; b) podział przyjęty w książce

diody warstwowe, natomiast diody $m-s$ dzieli się na diody z barierą Schottky'ego i diody ostrzowe. Jest to klasyfikacja pozornie słuszna, gdyż dioda ostrzowa składa się z igły metalowej, której ostrze jest w kontakcie z kryształem półprzewodnikowym, nie mającym warstw p , n (najczęściej jest to igła wolframowa w kontakcie z germanem typu n). Trudno jednak zaakceptować bez zastrzeżeń taką klasyfikację, jeżeli wnuknie się w istotę zjawisk określających właściwości diod ostrzowych. Z p. 3.4 wiadomo, że kontakt ostrza metalowego z półprzewodnikiem powoduje powstanie dużej liczby stanów powierzchniowych, które z kolei wywołują inwersję typu przewodnictwa przy powierzchni półprzewodnika. W ten sposób pod ostrzem powstaje półsferyczna warstewka inwersyjna na

podłożu typu n . W istocie zatem mamy do czynienia ze złączem p - n uformowanym wskutek zjawisk powierzchniowych — „uformowanym powierzchniowo”.

W celu ustabilizowania właściwości diody ostrzowej najczęściej wykonuje się formowanie elektryczne, tj. przepuszcza się przez diodę krótkotrwały (ok. 100 μ s) impuls prądu o natężeniu kilku amperów. Wówczas pod ostrzem wytwarza się wysoka temperatura i do półprzewodnika dyfundują akceptory z igły metalowej (często pokrytej warstwą indu lub aluminium). W ten sposób powstaje warstewka półsferyczna typu p (rys. 4.2a) i mówi się o złączu p - n „uformowanym elektrycznie”.

W obu przypadkach dioda ostrzowa działa na zasadzie powstawania bariery na złączu p - n , a nie na styku metal-półprzewodnik, jak to jest w diodzie Schottky'ego. Dlatego słuszniejsza wydaje się być klasyfikacja diod przedstawiona na rys. 4.3b. Dla odróżnienia diod p - n ostrzowych od diod p - n warstwowych można by posłużyć się kryterium opartym na porównaniu rozmiarów liniowych powierzchni złącza z szerokością warstwy zaporowej. Według tego kryterium w diodach ostrzowych rozmiary liniowe wyznaczające powierzchnię złącza są mniejsze niż szerokość warstwy zaporowej. Często przyjmuje się, że miarą „punktowości” złącza jest stosunek krzywizny złącza r do drogi dyfuzji nośników mniejszościowych L . Jeżeli $r/L \ll 1$, to złącze p - n jest „punktowe”, czyli mamy do czynienia z diodą ostrzową. To kryterium nie jest wystarczające, gdyż może być również spełnione w złączach planarnych o bardzo małej powierzchni. Aby uniknąć wynikających stąd nieporozumień, przez pojęcie dioda ostrzowa będziemy rozumieć taką diodę, w której uzyskuje się efekt prostowniczy wskutek kontaktu ostrej igły metalowej z półprzewodnikiem (mimo że bariera może być w złączu p - n , a nie na bezpośrednim styku metalu z półprzewodnikiem, to przecież powstanie złącza p - n jest pośrednio spowodowane przez kontakt ostrej igły z kryształem półprzewodnika).

Nas jednak najbardziej interesuje klasyfikacja diod półprzewodnikowych, sporządzona z punktu widzenia użytkownika, tj. uwypuklająca przede wszystkim cechy funkcjonalne elementów. W związku z tym przyjmuje się następujący podział diod:

- prostownicze,
- uniwersalne,
- stabilitrony (diody Zenera)
- impulsowe,
- pojemnościowe (warikapy i waraktory),
- tunelowe,
- mikrofalowe (detekeyjne, mieszające oraz diody o sterowanej impedancji — diody PIN).

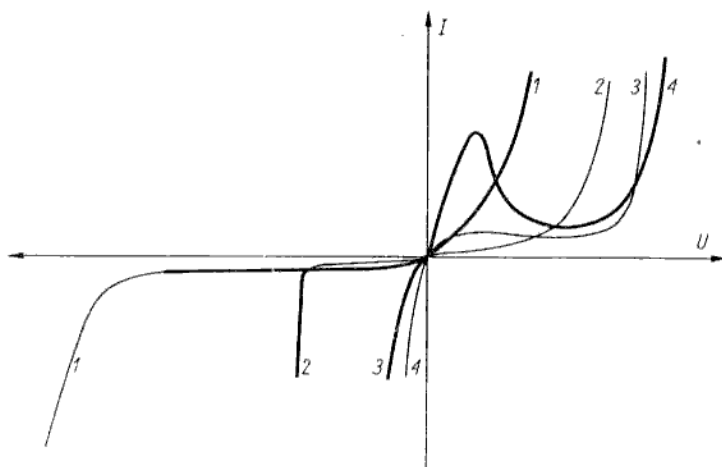
W tej klasyfikacji pomija się diody, których zasada działania jest oparta na zjawiskach nie znanych nam z lektury poprzedniego rozdziału. Chodzi tu o diody działające z wykorzystaniem zjawisk przestrzennych w półprzewodnikach (diody lawinowo-przelotowe oraz Gunna, nazywane często diodami generacyjnymi) oraz diody stosowane w optoelektronice jako detektory i źródła promieniowania świetlnego. Te chwilowo pominięte zagadnienia będą rozpatrywane w innym miejscu (z wyjątkiem detektorów i źródeł promieniowania).

Bardzo ważne jest zrozumienie związków między właściwościami funkcjonalnymi a konstrukcją i technologią diod. Z punktu widzenia użytkownika można reprezentować dla przykładu skrajny pogląd, że każda dioda użyta do prostowania prądu zmiennego jest diodą prostowniczą, lub też można twierdzić, że każda dioda jest diodą

pojemnościową. Jest to pogląd w zasadzie niesłuszny, gdyż przez pojęcie dioda prostownicza, pojemnościowa itp. jest rozumiana klasa diod szczególnie predestynowanych do określonych zastosowań. Jest to inaczej klasa diod o parametrach i charakterystykach optymalizowanych przez wytwórcę pod kątem najlepszej przydatności do określonych zastosowań.

Z drugiej strony ważne jest jednak — szczególnie w aspekcie dydaktycznym — by w zapale klasyfikowania („szufladkowania”) nie stwarzać sztucznych przegród między poszczególnymi rodzajami diod. Przecież wszystkie diody $p-n$ działają na zasadach wynikających z jednolitej teorii złącza $p-n$, przedstawionej w poprzednim rozdziale.

W celu uzyskania określonych właściwości funkcjonalnych diod wytwórca wybiera przede wszystkim odpowiedni poziom i rozkład koncentracji domieszek, odpowiednie rozmiary geometryczne struktury półprzewodnikowej i odpowiedni rodzaj obudowy. Z poprzedniego rozdziału wiadomo, że właściwości złącza $p-n$ silnie zależą na przykład od koncentracji domieszek. Na rysunku 4.4 przedstawiono



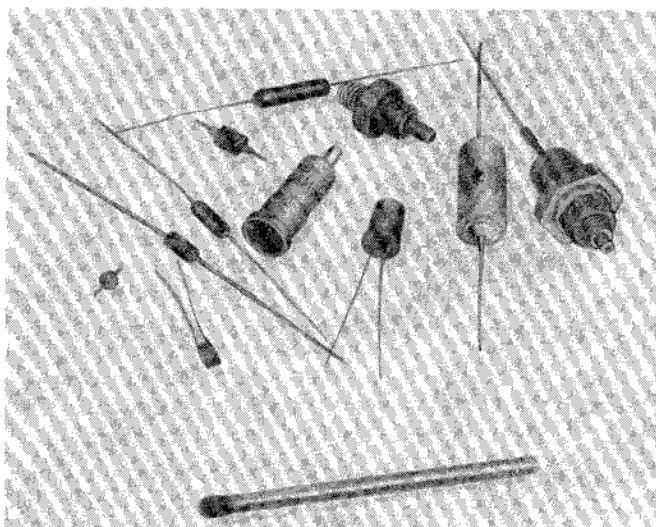
Rys. 4.4

Charakterystyki prądowo-napięciowe diod $p-n$ dla różnych koncentracji domieszek. Koncentracja domieszek zwiększa się zgodnie ze wzrastającą numeracją krzywych 1 — dioda prostownicza; 2 — stabilizator (dioda Zenera); 3 — dioda zwrótowa (dioda mieszająca, detekcyjna); 4 — dioda tunelowa; linią grubą wykreślono odcinki charakterystyk odpowiadające typowym obszarom pracy tych diod

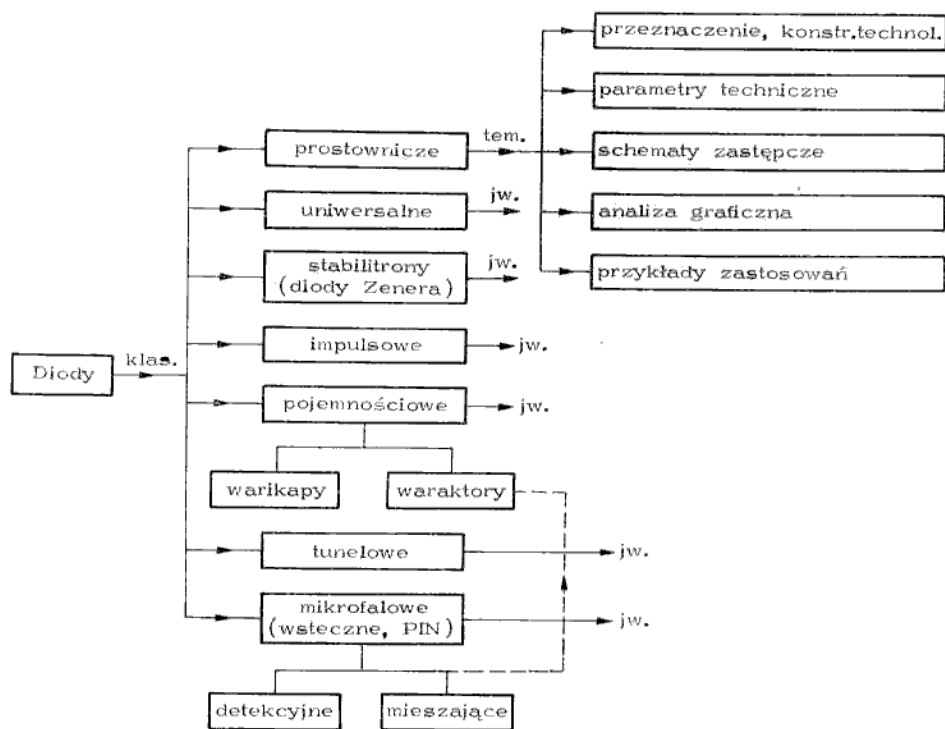
zbiorczy wykres charakterystyk prądowo-napięciowych diod $p-n$ dla różnych koncentracji domieszek i związane poszczególne charakterystyki z odpowiednimi rodzajami diod, wynikającymi z podziału funkcjonalnego.

Rodzaj obudowy (oprawki) ma szczególnie istotne znaczenie w przypadku diod mikrofalowych i diod mocy. Na rysunku 4.5 zaprezentowano asortyment obudów na przykładzie diod produkcji polskiej.

Dla użytkownika elementów półprzewodnikowych obok znajomości parametrów i charakterystyk elementów bardzo ważna jest umiejętność analizy pracy elementów w układzie. Jeżeli element jest zastosowany w typowym układzie, to często wystarcza posługiwanie się tzw. parametrami technicznymi, podanymi przez producenta w katalogu. W dokładniejszych czynnościach projektowych stosuje się metodę analizy graficznej, która jednak ma ograniczoną przydatność w ukła-



Rys. 4.5
Diody produkcji polskiej, prezentujące asortyment stosowanych obudów



Rys. 4.6
Schematyczny układ rozdziału

dach zawierających więcej niż dwa elementy i w analizie dynamicznej. Największe możliwości wykonania dokładnej analizy nawet złożonych układów dają metody analityczne, oparte na teorii obwodów. W tym przypadku element nieliniowy (dioda) jest reprezentowany przez odpowiedni schemat zastępczy.

Poszczególne rodzaje diod będą omawiane w następującym układzie zagadnień (nie zawsze z zachowaniem tej kolejności):

- przeznaczenie i cechy konstrukcyjno-technologiczne,
- parametry techniczne (katalogowe),
- schemat zastępczy,
- analiza graficzna,
- przykłady zastosowań.

Układ całego rozdziału przedstawiono schematycznie na rys. 4.6.

Diody prostownicze

4.1

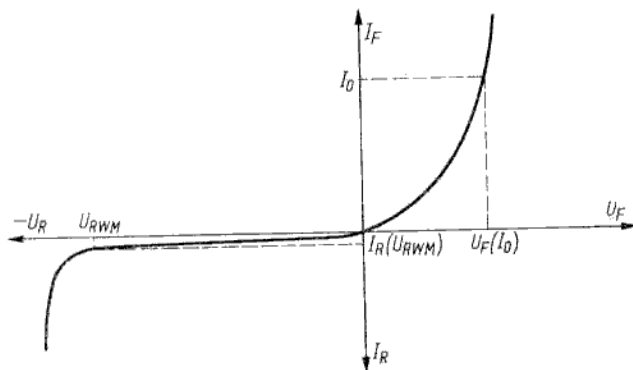
Diodami prostowniczymi nazywa się diody przeznaczone do prostowania prądu przemiennego. W domyśle rozumie się prostowanie prądu o małej częstotliwości (z reguły chodzi o sieć przemysłową 50 Hz) przy znacznych lub zgoła dużych mocach wydzielanych w obciążeniu. Są to zatem diody pracujące przeważnie w układach prostowniczych bloków zasilania różnych urządzeń elektronicznych i elektrycznych. Jak już wspomniano, diody te pracują najczęściej przy częstotliwości 50 Hz, niekiedy przy częstotliwości 400 Hz (systemy pokładowe) lub nawet kilka kiloherców (prostowanie prądu z przetwornicy tranzystorowej, podwyższającej napięcie). Jest to jednak zakres tak małych częstotliwości, że zjawiska dynamiczne nie mają istotnego wpływu na pracę diody w układzie.

Diody prostownicze są to diody warstwowe (dyfuzyjne lub stopowe) krzemowe i germanowe, przy czym te ostatnie stosuje się coraz rzadziej.

Parametry techniczne wszystkich diod można podzielić na dwie grupy, tj.:

- parametry charakterystyczne,
- dopuszczalne parametry graniczne.

Wyróżnia się następujące dwa parametry charakterystyczne diod prostowniczych (rys. 4.7):



Rys. 4.7

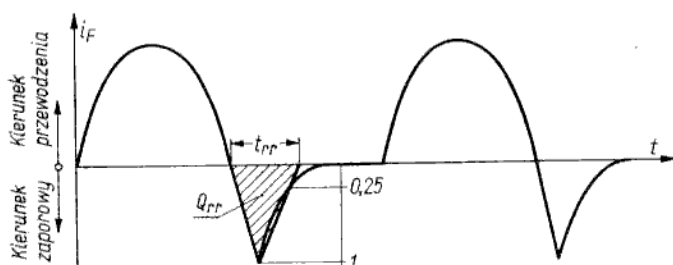
Charakterystyka prądowo-napięciowa diody prostowniczej z zaznaczonymi dwoma punktami charakterystycznymi

- napięcie przewodzenia U_F przy określonym prądzie przewodzenia I_F — najczęściej przy maksymalnym prądzie wyprostowanym I_O (inaczej przy maksymalnej wartości średniej prądu przewodzenia, nazywanej niekiedy prądem znamionowym I_{FN});
- prąd wsteczny I_R przy szczytowym napięciu wstecznym pracy U_{RWM} .

Wyróżnia się następujące dopuszczalne parametry graniczne:

- maksymalny średni prąd przewodzenia I_O lub inaczej prąd znamionowy w kierunku przewodzenia I_{FN} ,
- powtarzalny szczytowy prąd przewodzenia I_{FRM} (na przykład dla impulsów o czasie trwania $< 3,5$ ms i częstotliwości 50 Hz);
- niepowtarzalny szczytowy prąd przewodzenia I_{FSM} (na przykład dla jednokrotnego impulsu o czasie trwania < 10 ms);
- szczytowe napięcie wsteczne pracy U_{RWM} (lub średnie napięcie wsteczne przy pracy diody w prostowniku jednopółkownikowym z obciążeniem rezystancyjnym);
- powtarzalne szczytowe napięcie wsteczne U_{RRM} ;
- niepowtarzalne szczytowe napięcie wsteczne U_{RSM} .

Wszystkie wyżej wymienione parametry są określane dla temperatury otoczenia $T_a = 25^\circ\text{C}$. W najbardziej skrótowych informacjach katalogowych podaje się dwa parametry: I_O , U_R , rozumiane jako maksymalny prąd stały w kierunku przewodzenia i maksymalne napięcie stałe w kierunku zaporowym. Bardziej obszerne dane katalogowe zawierają ponadto dwa parametry charakterystyczne (U_F przy określonym I_F oraz I_R przy określonym U_R) i następujące parametry dopuszczalne: U_{RWM} , U_{RSM} , I_O , I_{FRM} , I_{FSM} oraz dopuszczalną temperaturę złącza T_j . Niekiedy podaje się jeszcze czas wyłączenia t_{rr} lub ładunek wyłączenia Q_{rr} (rys. 4.8) oraz opór cieplny R_{th} . Znajomość oporu cieplnego oraz dopusz-



Rys. 4.8
Sposób określania ładunku Q_{rr} i czasu przełączenia t_{rr} w diodach prostowniczych

czalnej temperatury umożliwia wyznaczenie maksymalnej mocy rozpraszanej w diodzie dla dowolnej temperatury otoczenia

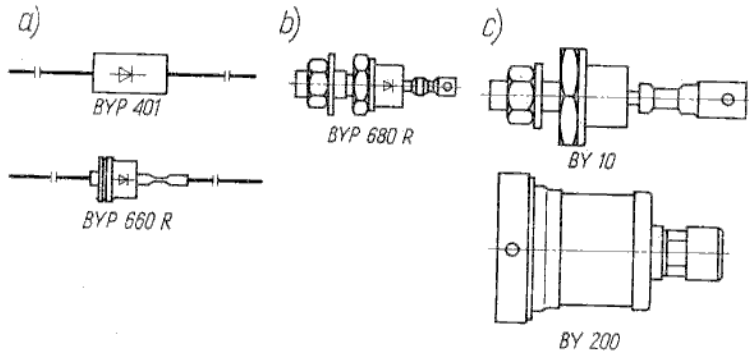
$$P_a = \frac{T_j - T_a}{R_{th}} \quad (4.1)$$

przy czym T_a — temperatura otoczenia.

Ze względu na moc rozpraszaną diody prostownicze dzieli się na:

- małej mocy, $P_a < 1$ W (na przykład diody BYP 401, BYP 660 R¹⁾ — rys. 4.9a);
- średniej mocy, $1 \text{ W} \leq P_a < 10$ W (na przykład BYP 680 R — rys. 4.9b);
- dużej mocy, $P_a \geq 10$ W (na przykład BY 10, BY 200 — rys. 4.9c).

¹⁾ Znaczenie symboli elementów produkcji polskiej opisano w Dodatku.



Rys. 4.9

Typowe obudowy diod prostowniczych produkcji polskiej: a) małej mocy; b) średniej mocy; c) dużej mocy

Na świecie produkuje się diody o mocy do kilku kilowatów (prąd przewodzenia ok. 2 kA) i napięciu wstecznym do kilku kV. Są to oczywiście diody krzemowe. Diody germanowe mają znacznie mniejsze dopuszczalne napięcie wsteczne i mniejszą temperaturę dopuszczalną ($T_j = 75^\circ\text{C}$ dla diod germanowych oraz 150°C dla diod krzemowych). Jedyną zaletą diod germanowych jest mniejsza wartość spadku napięcia na diodzie przy pracy w kierunku przewodzenia ($U_F(I_0) = 0,3...0,5$ V dla diod germanowych oraz $0,8...1,4$ V dla diod krzemowych). Ostatnio zaczęto również wytwarzać diody mocy z barierą Schottky'ego jako diody prostownicze. Ich zalety to mały spadek napięcia w kierunku przewodzenia ($U_F \approx 0,5...0,6$ V) i duża szybkość działania, wadą natomiast jest małe napięcie dopuszczalne w kierunku zaporowym.

Należy zauważyć, że diody na prądy powyżej 10 A pracują z radiatorami (najczęściej aluminiowymi) odprowadzającymi wydzielane ciepło do otoczenia, przy czym im większą moc pobiera dioda, tym większy jest radiator, a w przypadku bardzo dużych mocy stosuje się chłodzenie radiatora wymuszonym opływem powietrza lub cieczy chłodzącej.

Jeżeli nie dysponuje się diodami o wymaganym prądzie znamionowym lub napięciu wstecznym, to można łączyć diody równolegle (w celu zwiększenia prądu przewodzenia) lub szeregowo (w celu zwiększenia napięcia wstecznego). Należy przy tym pamiętać, że z uwagi na rozrzut charakterystyk dla poszczególnych egzemplarzy diod rozptył prądów (połączenie równoległe) lub rozkład napięć (połączenie szeregowo) mogą nie być jednakowe. Dla uzyskania równomiernego rozdziału prądów lub napięć w pierwszym przypadku włącza się szeregowo z każdą diodą rezystory o niewielkiej wartości, a w drugim — równoległe z każdą diodą rezystory o dużej wartości. W tabelicy 4.1 podano przykładowo typowe parametry techniczne różnych diod prostowniczych.

Analiza graficzna

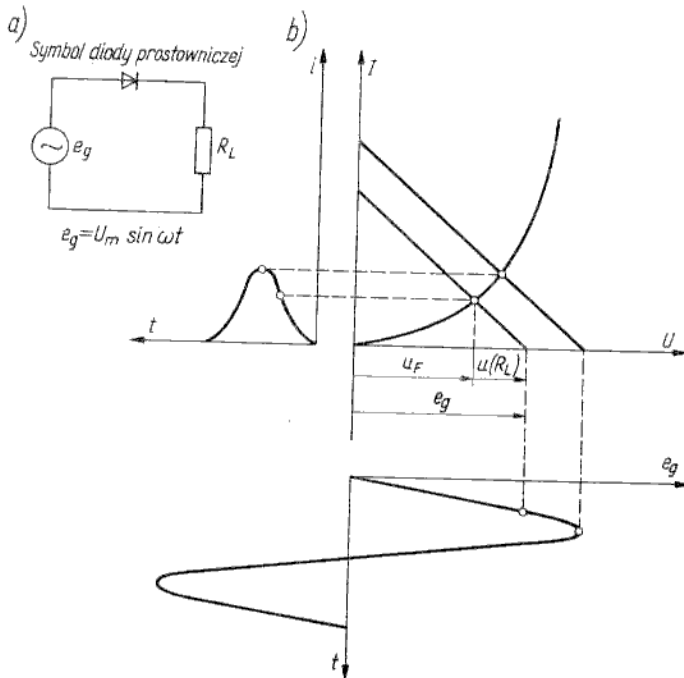
4.1.1

Po omówieniu parametrów technicznych rozpatrzmy podstawowe zasady analizy graficznej układów z diodami prostowniczymi. Na rysunku 4.10 przedstawiono pracę diody w układzie prostownika jednopolówkowego, przy czym w częś-

Tablica 4.1

Typowe parametry techniczne przykładowo wybranych diod prostowniczych

Typ diody	Parametry charakterystyczne			Dopuszczalne parametry graniczne						R_{th}	U w a g i	
	U_F	przy		I_O	I			U				T_J
		I_F	I_R		I_{FRM}	I_{FSM}	U_{RWM}	U_{RRM}	U_{RSM}			
	V	A	mA	A	A	A	V	V	V	$^{\circ}C$		$^{\circ}C/W$
DZG 1	0,5	0,3	0,5	0,3	—	0,9	—	16	—	50	75	Ge Si Si Si Si Si Si Dioda Schottky'ego (czas przełączenia 10 ns) Szeregowo połączone złącza krzemowe we wspólnej obudowie (tzw. stos diodowy)
BYP 660-700 R	1	0,6	0,01	0,6	—	15	—	700	—	—	125	
BYP 401-1000	1,1	1	0,005	1	—	50	—	1 000	—	—	175	
BYP 680-600 R	1,3	5	0,05	5	—	60	—	600	—	1 000	150	
BY 10/16	1,25	30	6	10	30	190	—	—	1 600	1 920	140	
BY 300/12	1,3	1 000	30	300	1 000	5 200	—	—	1 200	1 440	140	
1 N 5834	0,59	40	—	40	—	800	—	—	1 440	—	125	
BYX 29	—	—	—	0,05	0,25	5	—	150 000	—	—	150	



Rys. 4.10
Prosty układ prostownika jednopółkowego (a) i jego analiza graficzna (b)

ci a) pokazano układ włączenia diody, a w części b) — analizę graficzną tego układu. Analiza jest oparta na koncepcji prostej obciążenia. Ponieważ z prawa Kirchhoffa

$$e_g = u_F + u(R_L)$$

czyli

$$e_g = u_F + iR_L$$

przy czym: u_F — wartość chwilowa spadku napięcia na diodzie; R_L — rezystancja obciążenia, zatem

$$i = \frac{e_g}{R_L} - \frac{u_F}{R_L} \quad (4.2)$$

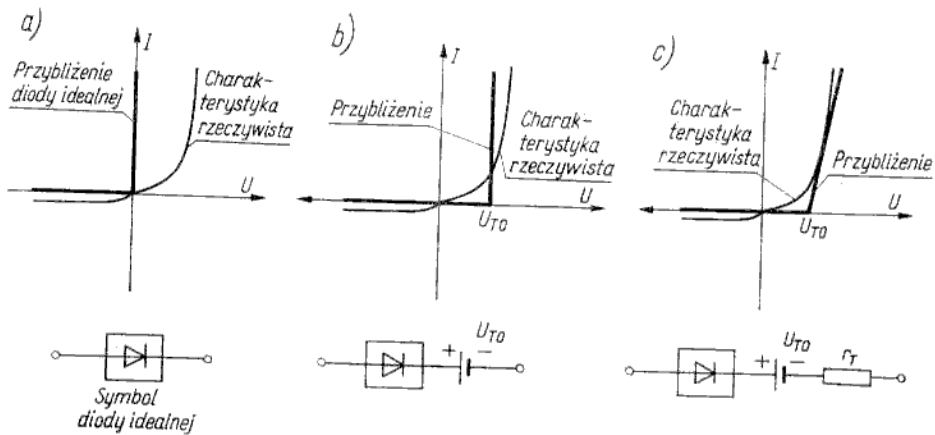
Jest to równanie prostej obciążenia o postaci ogólnej

$$y = b + kx$$

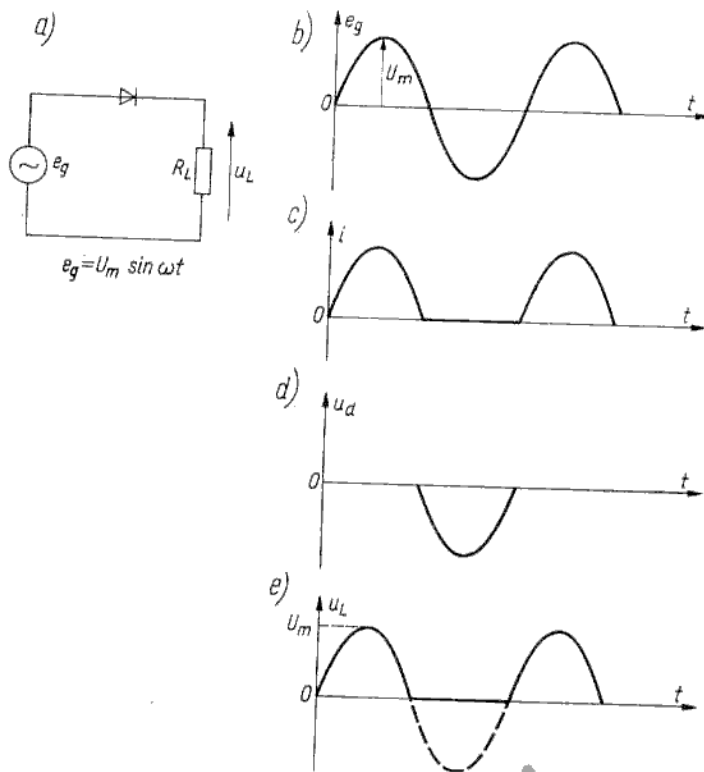
w którym: b — punkt przecięcia tej prostej z osią rzędnych ($b = e_g/R_L$), czyli wartość prądu i dla napięcia $u_F = 0$; k — nachylenie tej prostej ($k = -1/R_L$).

Prosta obciążenia przecina oś odciętych ($i = 0$) dla $u_F = e_g$. Punkt wspólny prostej obciążenia i charakterystyki diody jest rozwiązaniem układu dwu równań, tj. równania (4.2) i równania diody

$$i = I_R \left(\exp \frac{u_F}{m\varphi_T} - 1 \right)$$



Rys. 4.12 Modele diody odcinkami liniowymi: a) tzw. dioda idealna; b) model z uwzględnieniem napięcia progowego; c) model z uwzględnieniem napięcia progowego i skończonej konduktancji przyrostowej



Rys. 4.13 Prostownik jednopółkowy z charakterystycznymi przebiegami dla modelu diody idealnej: a) układ prostownika; b) napięcie wejściowe e_g ; c) przebieg prądu w obwodzie; d) przebieg napięcia na diodzie; e) przebieg napięcia na obciążeniu

Przykłady zastosowań

4.1.3

Rozpatrzmy trzy podstawowe układy prostownicze:

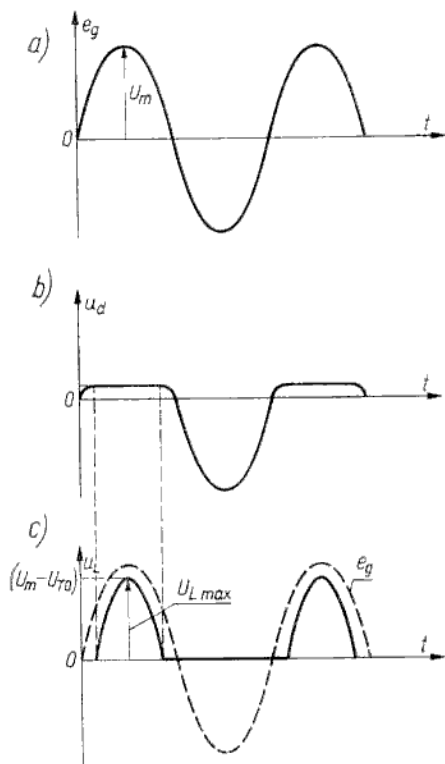
- jednokierunkowy półokresowy (jednopołówkowy);
- jednokierunkowy pełnokresowy (dwupołówkowy z wyprowadzonym środkiem uzwojenia wtórnego transformatora);
- dwukierunkowy (mostkowy).

Interesują nas przede wszystkim warunki pracy i wymagania stawiane diodom prostowniczym, włączonym w tych układach.

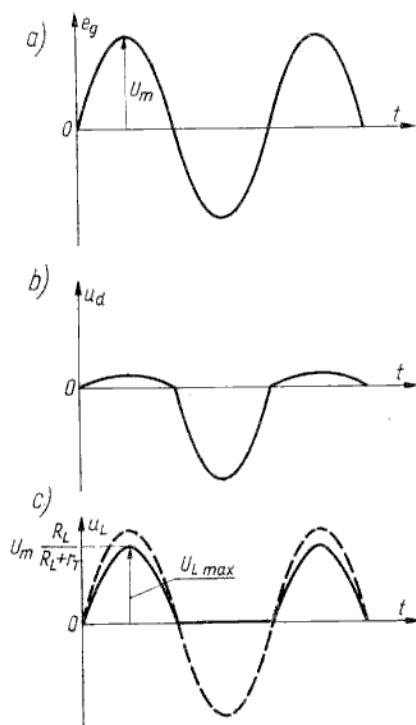
Prostownik jednopołówkowy

4.1.3.1

Na rysunku 4.13 przedstawiono układ prostownika jednopołówkowego wraz z przebiegami napięć i prądów w tym układzie przy zastąpieniu diody rzeczywistym modelem diody idealnej. Prąd płynie w obwodzie tylko dla dodatniej półokreski napięcia sinusoidalnego. Wartość średnia napięcia na obciążeniu



Rys. 4.14
Przebiegi napięć w prostowniku jednopołówkowym dla modelu diody z rys. 4.12b: a) napięcia wejściowe; b) napięcie na diodzie; c) napięcie na obciążeniu



Rys. 4.15
Przebiegi napięć w prostowniku jednopołówkowym, jeśli w modelu diody uwzględnia się tylko rezystancję diody r_T : a) napięcia wejściowe; b) napięcie na diodzie; c) napięcie na obciążeniu

$$U_{sr} = \frac{U_m}{2\pi} \int_0^{\pi} \sin\theta d\theta = \frac{U_m}{\pi} \quad (4.4)$$

Wartość średnia prądu w obciążeniu

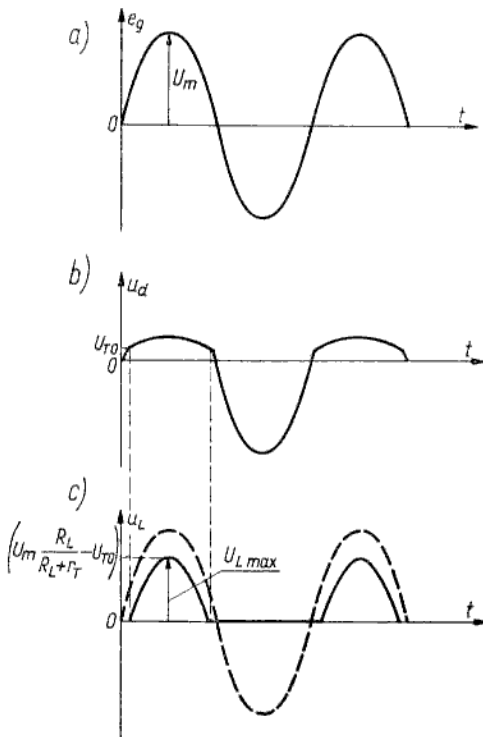
$$I_{sr} = \frac{U_{sr}}{R_L} = \frac{U_m}{\pi R_L} \quad (4.5)$$

Model diody idealnej jest zadowalająco dokładny do analizy układów prostowniczych, w których amplituda napięcia przemiennego jest znacznie większa niż napięcie progowe diody ($U_m \gg U_{TO}$) oraz rezystancja obciążenia jest znacznie większa niż rezystancja diody ($R_L \gg r_T$).

Jeżeli amplituda napięcia przemiennego U_m jest mała, to należy się posłużyć modelem diody ze źródłem szeregowym, reprezentującym napięcie progowe. Wówczas prąd w obwodzie zaczyna płynąć dopiero wówczas, gdy dodatnie napięcie e_g osiągnie wartość większą niż U_{TO} . Napięcie na obciążeniu ma przebieg pokazany na rys. 4.14. Maksymalna wartość chwilowa napięcia na obciążeniu

$$U_{Lmax} = U_m - U_{TO} \quad (4.6)$$

Jeżeli prąd obciążenia jest duży (R_L mała), to trzeba uwzględnić rezystancję diody. Wówczas napięcie na obciążeniu ma przebieg pokazany na rys. 4.15. Maksymalna wartość chwilowa napięcia na obciążeniu



Rys. 4.16

Przebiegi napięć w prostowniku jedno-półówkowym dla modelu diody z rys. 4.12c: a) napięcie wejściowe; b) napięcie na diodzie; c) napięcie na obciążeniu

$$U_{Lmax} = U_m \frac{R_L}{R_L + r_T} \quad (4.7)$$

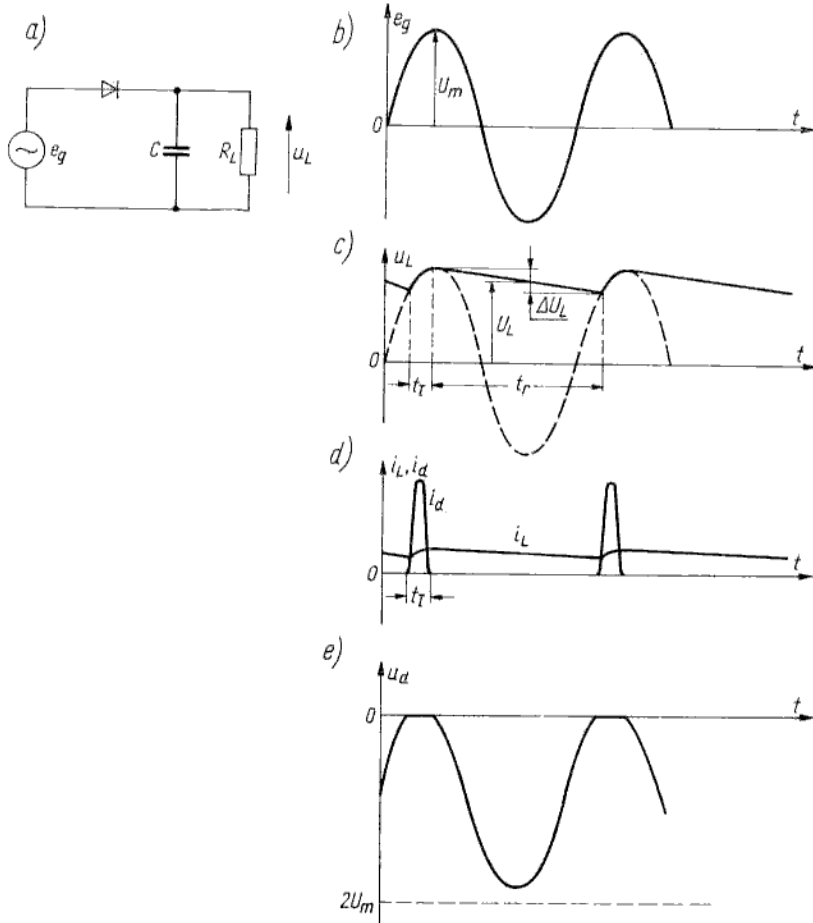
Jeżeli amplituda U_m oraz rezystancja R_L są małe, to należy uwzględnić zarówno napięcie progowe jak i rezystancję diody. Wówczas napięcie na obciążeniu ma przebieg pokazany na rys. 4.16. Maksymalna wartość chwilowa napięcia na obciążeniu

$$U_{Lmax} = U_m \frac{R_L}{r_T + R_L} - U_{TO} \quad (4.8)$$

Dioda pracująca w układzie prostownika jednopółkowego z obciążeniem rezystancyjnym powinna się charakteryzować:

— dopuszczalnym średnim prądem przewodzenia

$$I_o > \frac{U_m}{R_L} \quad (4.9)$$



Rys. 4.17
 Prostownik jednopółkowy z filtrem pojemnościowym i odpowiednie przebiegi: a) układ prostownika; b) napięcie wejściowe; c) napięcie na obciążeniu; d) prąd diody i prąd obciążenia; e) napięcie na diodzie

oraz
— dopuszczalnym szczytowym napięciem wstecznym

$$U_{RWM} > U_m \quad (4.10)$$

Wprawdzie na wyjściu rozpatrywanego prostownika otrzymuje się składowe stałe napięcia i prądu, równe wartościom średnim określonym wzorami (4.4), (4.5), to jednak przebieg napięcia na obciążeniu ma również bardzo dużą składową zmienną. Dlatego w celu zmniejszenia składowej zmiennej w większości układów rzeczywistych stosuje się filtrowanie napięcia wyjściowego przez dołączenie kondensatora równolegle do obciążenia (rys. 4.17). Kondensator ładuje się przez diodę do napięcia równego amplitudzie U_m napięcia przemiennego, po czym rozładowuje się przez obciążenie aż do chwili, gdy dioda zacznie znowu przewodzić, tj. gdy napięcie e_g w dodatniej połowie osiągnie wartość większą niż napięcie na kondensatorze. Składowa zmienna napięcia wyjściowego, charakteryzowana międzyszczytową wartością tętnień ΔU_L jest tym mniejsza, im większa jest stała czasowa obwodu $R_L C$. Zwiększenie stałej czasowej obwodu $R_L C$ prowadzi zarazem do wzrostu składowej stałej napięcia wyjściowego. W granicznym przypadku przy

$$\tau = R_L C \rightarrow \infty, \quad U_L \rightarrow U_m, \quad \Delta U_L \rightarrow 0$$

W przypadku skończonej wartości τ składowa stała

$$U_L = U_{Lmax} - \frac{\Delta U_L}{2} \quad (4.11)$$

Znając stałą czasową $\tau = R_L C$ oraz czas rozładowania t_r można wyznaczyć międzyszczytowe napięcie tętnień. Ponieważ w czasie t_r kondensator rozładowuje się ze stałą czasową $R_L C$ od napięcia U_{Lmax} , zatem

$$\Delta U_L = U_{Lmax} \left(1 - e^{-\frac{t_r}{R_L C}}\right)$$

Uwzględniając, że $R_L C \gg t_r$, można napisać przybliżoną równość

$$\Delta U_L \approx U_{Lmax} \frac{t_r}{R_L C} \quad (4.12)$$

Czas t_r w rzeczywistych układach jest nieznacznie mniejszy niż okres napięcia przemiennego T (najczęściej $t_r \approx 0,8T$), czyli

$$\Delta U_L \approx 0,8 U_{Lmax} \frac{T}{R_L C} \quad (4.13)$$

Należy zauważyć, że im mniejsze jest napięcie tętnień, tym mniejszy jest stosunek czasu ładowania t_l do czasu rozładowania t_r . Ponieważ ładunek dostarczony do kondensatora w czasie ładowania musi być równy ładunkowi odprowadzonemu z kondensatora w czasie rozładowania, zatem prąd ładowania

$$I_l = I_L \frac{t_r}{t_l} \quad (4.14)$$

Im większy jest stosunek t_r/t_l , tym większy jest prąd ładowania kondensatora w porównaniu z prądem obciążenia (prądem rozładowania).

Dlatego dioda pracująca w układzie prostownika jednopółkowego z filtrem pojemnościowym musi spełniać następujące warunki:

— dopuszczalny średni prąd przewodzenia

$$I_o > \frac{U_L}{R_L} \quad (4.15)$$

— dopuszczalny powtarzalny szczytowy prąd przewodzenia

$$I_{FRM} > I_L \frac{t_r}{t_i} \quad (4.16)$$

lub w innej postaci

$$I_{FRM} > I_L \frac{2\pi}{\sqrt{2\Delta U_L/U_{Lmax}}} \quad (4.17)$$

— dopuszczalne szczytowe napięcie wsteczne

$$U_{RWM} > 2U_m \quad (4.18)$$

Ten ostatni warunek wynika stąd, że katoda diody jest na potencjale dodatnim $U_L \approx U_m$, anoda zaś w czasie ujemnej półki napięcia sinusoidalnego uzyskuje potencjał $-U_m$; wówczas na diodzie jest napięcie międzyszczytowe równe $2U_m$.

Prostownik dwupółkowy z transformatorem

4.1.3.2

Na rysunku 4.18 przedstawiono układ prostownika dwupółkowego z wyprowadzonym środkiem uzwojenia wtórnego transformatora. W ten sposób w uzwojeniu wtórnym uzyskuje się dwa napięcia sinusoidalne o jednakowych amplitudach, lecz przesunięte względem siebie w fazie o 180°

$$u_2(t) = -u_1(t)$$

W pierwszym półokresie przewodzi dioda D_1 , a w drugim dioda D_2 . Jeżeli przyjmie się najprostszy model diody idealnej, to w pierwszym półokresie $u_L = u_1(t)$, w drugim zaś $u_L = u_2(t)$. Zatem na rezystancji obciążenia uzyskuje się wyprostowane napięcie dwupółkowe, którego wartość średnia jest dwa razy większa niż dla prostownika jednopółkowego (wzór (4.4))

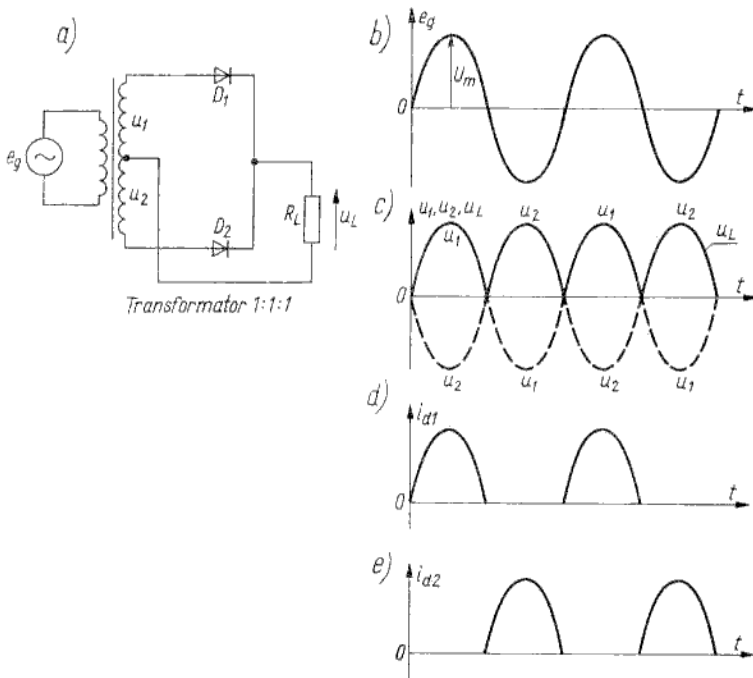
$$U_{sr} = \frac{2U_m}{\pi} \quad (4.19a)$$

Prąd średni natomiast

$$I_{sr} = \frac{2U_m}{\pi R_L} \quad (4.19b)$$

Mimo że prąd wyprostowany jest dwukrotnie większy niż dla prostownika jednopółkowego z obciążeniem rezystancyjnym, to jednak średni prąd płynący przez każdą diodę pozostaje bez zmian. Maksymalne napięcie wsteczne na każdej diodzie osiąga wartość $2U_m$.

Diody pracujące w układzie prostownika dwupółkowego z obciążeniem rezystancyjnym muszą zatem spełniać następujące wymagania:



Rys. 4.18

Prostownik dwupółkowy i charakterystyczne przebiegi: a) układ prostownika dwupółkowego; b) napięcie wejściowe; c) napięcie na obciążeniu; d) prąd diody pierwszej; e) prąd diody drugiej

— dopuszczalny średni prąd przewodzenia

$$I_o > \frac{U_m}{\pi R_L} \tag{4.20}$$

— dopuszczalne szczytowe napięcie wsteczne

$$U_{RWM} > 2U_m \tag{4.21}$$

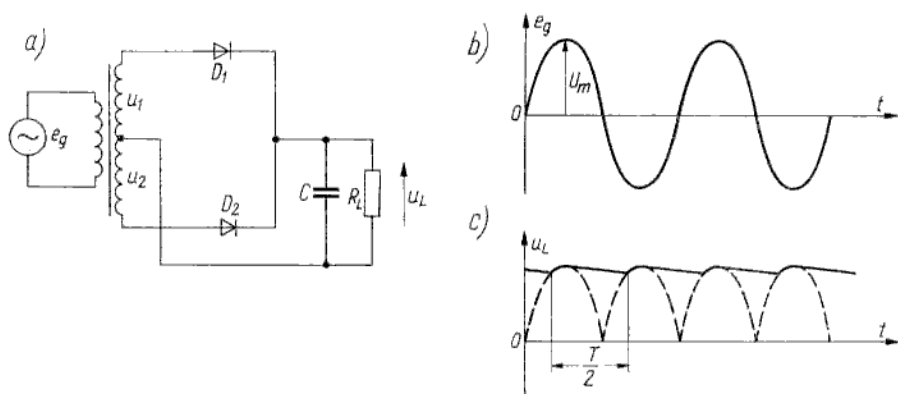
Jeżeli równolegle do obciążenia dołączy się kondensator filtrujący składową zmienną (rys. 4.19), to taki układ pracuje podobnie jak prostownik jednopółkowy z filtrem, przy czym w odpowiednich wzorach trzeba zamienić T na $T/2$. Analogicznie do (4.13) amplituda tętnień

$$\Delta U_L \approx 0,4 U_{Lmax} \frac{T}{R_L C} \tag{4.22}$$

W tym przypadku diody muszą spełniać następujące wymagania:

— dopuszczalny średni prąd przewodzenia

$$I_o > \frac{U_L}{2R_L} \tag{4.23}$$



Rys. 4.19

Prostownik dwupółkowy z filtrem pojemnościowym i odpowiednie przebiegi: a) układ prostownika; b) napięcie wejściowe; c) napięcie na obciążeniu

— dopuszczalny powtarzalny szczytowy prąd przewodzenia

$$I_{FRM} > I_L \frac{t_r}{t_i} \quad (4.24)$$

lub w innej postaci

$$I_{FRM} > I_L \frac{\pi}{\sqrt{2\Delta U_L / U_{Lmax}}} \quad (4.25)$$

— dopuszczalne szczytowe napięcie wsteczne

$$U_{RWM} > 2U_m \quad (4.26)$$

Prostownik dwupółkowy mostkowy

4.1.3.3

Na rysunku 4.20 przedstawiono układ prostownika dwupółkowego mostkowego. W dodatnim półokresie napięcia e_g prąd płynie w obwodzie D_1, R_L, D_2 , natomiast w ujemnym półokresie napięcia e_g prąd płynie w obwodzie D_4, R_L, D_3 . Kierunek przepływu prądu przez obciążenie jest jednakowy w obu półokresach, dlatego na obciążeniu otrzymuje się wyprostowane napięcie dwupółkowe. Napięcie i prąd średni w obciążeniu są określone wzorami (4.19a), (4.19b). Jednak prąd średni, płynący przez każdą diodę, jest taki jak dla prostownika jedno-półkowego. Uwzględniając przy tym, że zawsze dwie diody pracują szeregowo, można określić wymagania stawiane diodom:

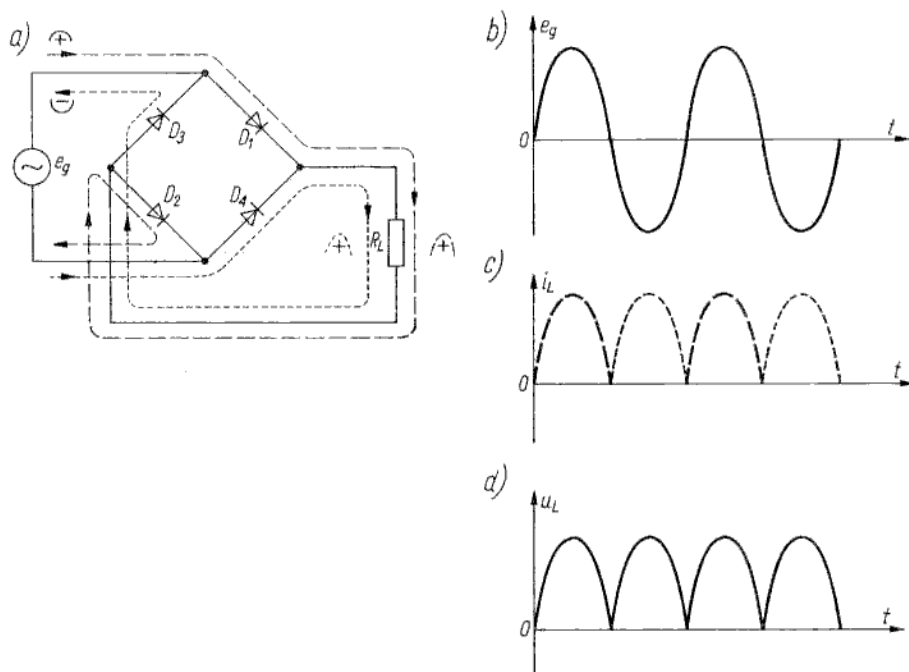
— dopuszczalny średni prąd przewodzenia

$$I_O > \frac{U_m}{\pi R_L} \quad (4.27)$$

— dopuszczalne szczytowe napięcie wsteczne

$$U_{RWM} > U_m \quad (4.28)$$

Jeżeli równoległe do obciążenia dołączy się kondensator filtrujący składową zmienną, to diody muszą spełniać następujące wymagania:



Rys. 4.20

Prostownik dwupółkowy mostkowy i charakterystyczne przebiegi: a) układ prostownika; b) napięcie wejściowe; c) prąd obciążenia; d) napięcie na obciążeniu

— dopuszczalny średni prąd przewodzenia

$$I_O > \frac{U_L}{2R_L} \quad (4.29)$$

— dopuszczalny powtarzalny szczytowy prąd przewodzenia

$$I_{FRM} > I_L \frac{t_r}{t_i} \quad (4.30)$$

lub w innej postaci

$$I_{FRM} > I_L \frac{\pi}{\sqrt{2\Delta U_L / U_{Lmax}}} \quad (4.31)$$

— dopuszczalne szczytowe napięcie wsteczne

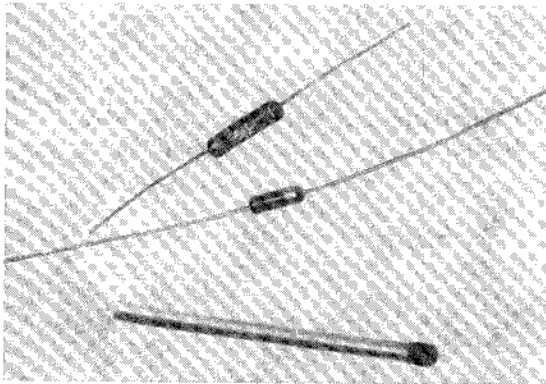
$$U_{RWM} > U_m$$

Z porównania omówionych trzech rodzajów układów prostowniczych wynika, że najłagodniejsze wymagania są stawiane diodom pracującym w układzie mostkowym, który często jest nazywany *układem Graetza*. Cztery diody połączone w mostku Graetza mogą być wytwarzane w jednej wspólnej obudowie jako konstrukcyjnie zwarte podzespół. W Polsce są produkowane mostki Graetza pod nomenklaturą 4 BYP 401, przeznaczone do zastosowań w prostownikach średniej mocy.

Diody uniwersalne

4.2

Diody uniwersalne są to diody przeznaczone głównie do zastosowań w układach detekcyjnych, prostowniczych małej mocy, ogranicznikach i innych układach urządzeń pomiarowo-kontrolnych. Są to na ogół diody ostrzowe germanowe (rys. 4.21), charakteryzujące się niewielkim zakresem napięć (do 100 V) i prą-



Rys. 4.21
Typowe obudowy diod uniwersalnych

dów (do 100 mA) o ograniczonej częstotliwości pracy do kilkudziesięciu megaherców.

Jako diody uniwersalne stosuje się również planarne diody krzemowe małej mocy. W grupie parametrów charakterystycznych diod uniwersalnych wyróżnia się parametry statyczne i dynamiczne.

Parametry statyczne:

- napięcie przewodzenia U_F przy określonym prądzie przewodzenia I_F ;
- prąd wsteczny I_R przy określonym napięciu wstecznym U_R .

Parametry dynamiczne:

- pojemność diody przy określonej częstotliwości i określonym napięciu wstecznym (najczęściej $C < 1$ pF dla $f = 10,7$ MHz, $U_R = 1$ V);
- sprawność detekcji η , tj. stosunek mocy sygnału zdemodulowanego do mocy sygnału wejściowego w cz. (parametr podawany dla diod przeznaczonych do zastosowania w układach detekcyjnych).

Wyróżnia się następujące dopuszczalne parametry graniczne:

- maksymalny stały prąd przewodzenia I_{Fmax} ,
- maksymalny szczytowy prąd przewodzenia I_{FMmax} ,
- maksymalne stałe napięcie wsteczne U_{Rmax} ,
- maksymalne szczytowe napięcie wsteczne U_{RMmax} .

Ponadto podaje się dopuszczalną temperaturę złącza T_j (zwykle 75...90°C dla diod germanowych oraz ok. 150°C dla diod krzemowych).

Analizę graficzną układów z diodami uniwersalnymi wykonuje się w podobny sposób jak dla układów z diodami prostowniczymi. Zachowują również aktualność modele odcinkami liniowe, przedstawione na rys. 4.12 dla diod prostowniczych. Należy tylko zwrócić uwagę, że rezystancja szeregową w diodzie ostrzowej jest większa niż w diodzie warstwowej, dlatego w tym przypadku element r_T ma dużą wartość (ok. 100Ω).

Parametry techniczne dla kilku przykładowo wybranych diod uniwersalnych zestawiono w tabl. 4.2.

Tablica 4.2

Parametry techniczne przykładowo wybranych diod uniwersalnych

Typ diody	Parametry charakterystyczne										Dopuszczalne parametry graniczne				U w a g i		
	statyczne					dynamiczne					I					U	
	U_F V	przy I_F mA		I_R μ A	C pF	przy f MHz		U_R V	η %	przy f_i MHz		I_{Fmax} mA	I_{FMmax} mA	U_{Rmax} V		U_{RMmax} V	T_J $^{\circ}$ C
		U_F	I_F			f	U_R			f_i							
AAP 120	2,2	10	250	0,7	10,7	1	65	10,7	80	25	80	70	100	75	Ge		
AAP 161	1,5	10	40	0,7	10,7	1	55	10,7	50	16	50	10	30	75	Ge		
BA 316	0,85	10	0,2	3	—	—	—	—	225	100	225	10	—	200	Si		

produkcji
polskiej

Przykłady zastosowań

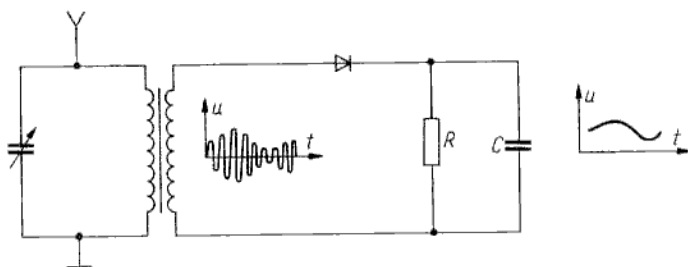
4.2.1

Diody uniwersalne są stosowane głównie w układach detekcji amplitudowej, w układach ograniczników, w detektorach stosunkowych odbiorników FM (dobierane pary diod), w demodulatorach pierścieniowych (cztery połączone diody mogą być wykonane we wspólnej obudowie).

Układ detekcji amplitudowej

4.2.1.1

Zadaniem detektora jest odtworzenie w radioodbiorniku sygnału modulującego z sygnału zmodulowanego amplitudowo. Detektor diodowy (rys. 4.22) różni się od układu prostownika tym, że zawiera strojony obwód rezonansowy, słu-



Rys. 4.22
Układ detektora amplitudy

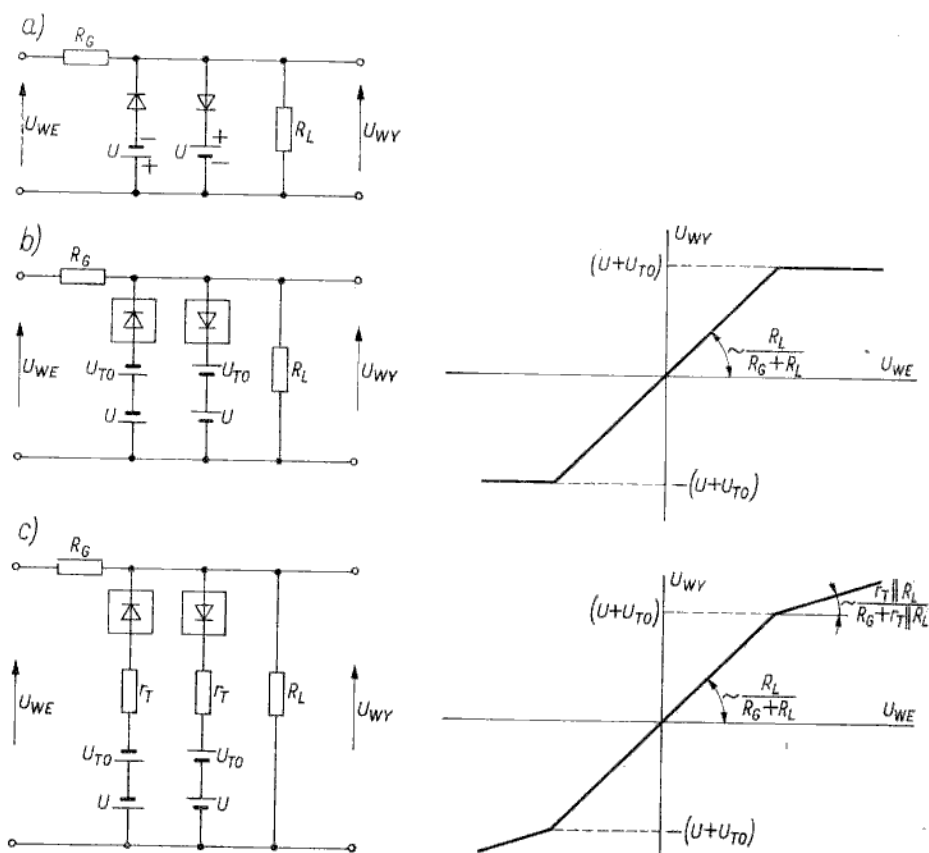
żący do wyboru sygnału pożądanej stacji radiowej, oraz ma filtr spełniający inną funkcję niż w układzie prostowniczym. Zadaniem filtra R, C jest wyeliminowanie składowej nośnej w.cz. bez stłumienia pożądanej składowej sygnału modulującego małej częstotliwości.

Diody ostrzowe germanowe są szczególnie predysponowane do zastosowań w układach detekcji, gdyż mają bardzo małe pojemności oraz ich charakterystyki $I(U)$ można aproksymować odcinkami liniowymi (wskutek dużego wpływu rezystancji szeregowej w zakresie przewodzenia charakterystyka jest zbliżona do prosto-liniowej). Taki detektor ma zatem bardzo małe zniekształcenia nieliniowe, czyli w wyniku detekcji, praktycznie biorąc, nie powstają harmoniczne sygnału podstawowego.

Ograniczniki diodowe

4.2.1.2

Na rysunku 4.23 przedstawiono prosty układ symetrycznego ogranicznika amplitudy i jego dwa modele wraz z charakterystykami przenoszenia. Jeżeli diodę rzeczywistą zastąpi się modelem z rys. 4.12b, to ogranicznik ma charakterystykę $U_{WY}(U_{WE})$ przedstawioną na rys. 4.23b. Jeżeli amplituda napięcia wejściowego przekroczy wartość $U + U_{TO}$, to dioda idealna, spolaryzowana w kierunku przewodzenia, stanowi klucz zwarty i napięcie wyjściowe nie zmienia się, gdyż dalsze przyrosty napięcia wejściowego odkładają się na rezystancji R_G . Jeżeli $U_{WE} < U + U_{TO}$, to dioda idealna, spolaryzowana w kierunku zaporowym, stanowi klucz rozarty i w przypadku braku obciążenia na wyjściu $U_{WY} = U_{WE}$. W schemacie na rys. 4.23c posłużono się dokładniejszym modelem diody, w którym uwzględniono rezystancję szeregową. W tym przypadku charakterystyka



Rys. 4.23

Symetryczny ogranicznik diodowy: a) układ włączenia diod; b) model zastępczy ogranicznika i charakterystyka przejściowa $U_{WY} = f(U_{WE})$; c) dokładniejszy model zastępczy i charakterystyka $U_{WY} = f(U_{WE})$

$U_{WY}(U_{WE})$ nie ma idealnego nasycenia wskutek spadku napięcia na rezystancji r_T . Ogranicznik symetryczny jest przykładem prostego układu przekształcającego. Przez odpowiednią kombinację połączeń diod i rezystorów można otrzymać aproksymację odcinkową bardziej złożonych funkcji $U_{WY}(U_{WE})$, np. funkcji kwadratowej, wykładniczej lub logarytmicznej. Takie układy przekształcające stosuje się często w maszynach analogowych.

Stabilizatory

4.3

Stabilizatory, nazywane również stabilizatorami lub diodami Zenera, są to diody warstwowe $p-n$, przeznaczone do zastosowań w układach stabilizacji napięcia, w układach ograniczników, jako źródła napięć odniesienia itp.

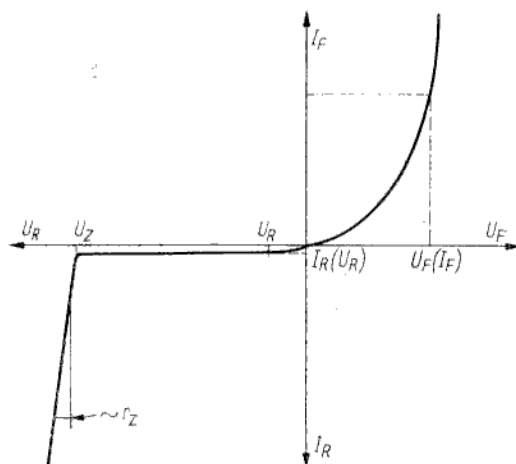
Typowy obszar pracy tych diod znajduje się na odcinku charakterystyki prądowo-napięciowej, odpowiadającym gwałtownemu wzrostowi prądu wstecznego wskutek zjawiska przebicia Zenera lub (i) przebicia lawinowego. Oba wymienione mecha-

nizmy przebicia były rozpatrywane w p. 3.1.5, w którym wyjaśniliśmy przede wszystkim następujące fakty:

- przebicie Zenera występuje w złączach silnie domieszkowanych przy napięciach do 5 V,
- przebicie lawinowe występuje w złączach słabo domieszkowanych przy napięciach powyżej 7 V,
- przebicia Zenera i lawinowe występują w złączach o średniej koncentracji domieszek przy napięciach 5...7 V,
- temperaturowy współczynnik napięcia przy przebiciu Zenera ma znak ujemny,
- temperaturowy współczynnik napięcia przy przebiciu lawinowym ma znak dodatni.

Obecnie na świecie są produkowane stabilitrony na napięcia od 2 do kilkuset woltów, przy czym nazwa dioda Zenera tradycyjnie obejmuje swym znaczeniem zarówno diody o przebiciu Zenera jak i diody o przebiciu lawinowym.

W grupie parametrów charakterystycznych stabilitronów wyróżnia się następujące parametry statyczne i dynamiczne (rys. 4.24).



Rys. 4.24
Charakterystyka prądowo-napięciowa stabilitronu z zaznaczonymi parametrami charakterystycznymi

Parametry statyczne:

- napięcie przewodzenia U_F przy określonym prądzie przewodzenia I_F ;
- prąd wsteczny I_R przy określonym napięciu wstecznym U_R — zwykle przy $U_R = 1$ V;
te dwa parametry mogą być mało istotne, gdy dioda pracuje wyłącznie w zakresie przebicia;
- napięcie stabilizacji U_Z (nazywane również napięciem Zenera), zwykle definiowane jako napięcie odpowiadające dziesiątej części maksymalnego prądu stabilizacji, może być również definiowane jako napięcie odpowiadające umownej wartości prądu stabilizacji; w katalogach podaje się napięcie stabilizacji i rozrzut jego wartości dla określonego typu diod;
- temperaturowy współczynnik napięcia stabilizacji TKU_Z jest to stosunek względnej zmiany napięcia stabilizacji do bezwzględnej zmiany temperatury otoczenia przy określonym prądzie stabilizacji, wyrażony w $1/^\circ\text{C}$ lub $\% / ^\circ\text{C}$

$$TKU_Z = \frac{1}{U_Z} \frac{dU_Z}{dT} \Big|_{I_Z = \text{const}}$$

Parametr dynamiczny:

— rezystancja dynamiczna, definiowana jako

$$r_z = \frac{\Delta U_z}{\Delta I_z}$$

przy określonym prądzie stabilizacji (rezystancję r_z można wyznaczyć z nachylenia charakterystyki statycznej $I(U)$, zatem — ściśle biorąc — jest to rezystancja przyrostowa).

Wyróżnia się następujące dopuszczalne parametry graniczne:

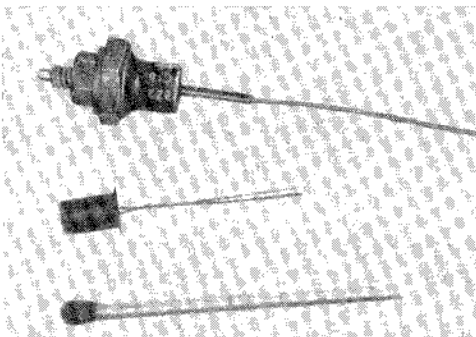
- Maksymalny stały prąd przewodzenia I_{Fmax} (w stabilizatorach średniej i dużej mocy często podaje się dopuszczalny szczytowy prąd przewodzenia I_{FMmax});
- Maksymalny dopuszczalny prąd stabilizacji I_{Zmax} , najczęściej wyznaczany z ilorazu maksymalnej mocy strat i napięcia stabilizacji

$$I_{Zmax} = \frac{P_{max}}{U_z}$$

- Maksymalna moc strat P_{max} — w katalogach podawana dla temperatury otoczenia 25°C . Dla wyższej temperatury otoczenia T_a

$$P_{max} = P_{max(25^\circ\text{C})} \frac{T_j - T_a}{T_j - 25^\circ\text{C}}$$

Ze względu na moc rozpraszaną stabilizator dzieli się na: małej ($P_{max} \leq 1 \text{ W}$), średniej ($1 \text{ W} < P_{max} \leq 10 \text{ W}$), dużej ($10 \text{ W} < P_{max} \leq 100 \text{ W}$) i bardzo dużej mocy ($P_{max} > 100 \text{ W}$). Widok zewnętrzny typowych stabilizatorów małej i średniej mocy przedstawiono na rys. 4.25.



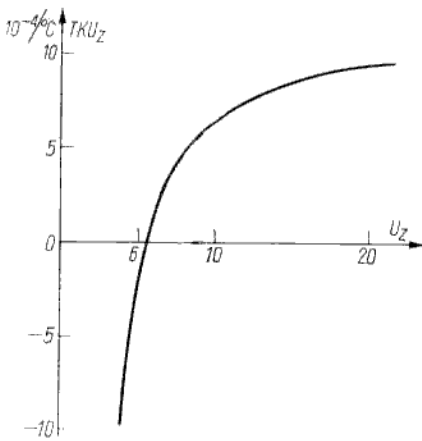
Rys. 4.25
Typowe obudowy stabilizatorów małej i średniej mocy

- Maksymalna temperatura złącza T_j (zwykle 150°C , gdyż stabilizatory są wytwarzane prawie wyłącznie z krzemu¹⁾).

Skomentujmy szerzej dwa bardzo ważne parametry, tj. temperaturowy współczynnik napięcia stabilizacji i rezystancję dynamiczną.

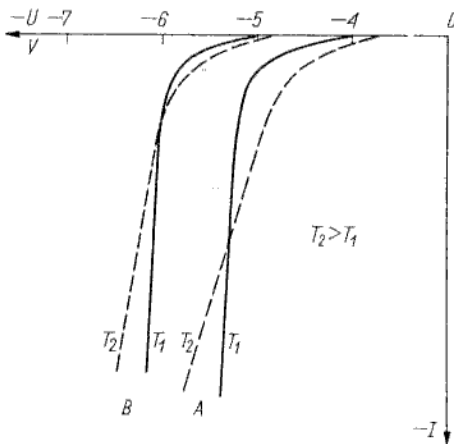
Wiadomo z p. 3.1.8, że temperaturowy współczynnik napięcia stabilizacji może przyjmować wartości ujemne (przebiecie Zenera dla $U_z < 5 \text{ V}$), dodatnie (przebiecie lawinowe dla $U_z > 7 \text{ V}$) lub bliskie zera dla diod o napięciu $U_z = 5 \dots 7 \text{ V}$. Typowy wykres współczynnika TKU_z w funkcji napięcia U_z przedstawiono na rys. 4.26.

¹⁾ W złączach germanowych przebiecie Zenera lub lawinowe łatwo przechodzi w fazę niszczącego przebiecia cieplnego.



Rys. 4.26
Zależność temperaturowego współczynnika napięcia stabilizacji TKU_Z od napięcia stabilizacji U_Z

Zerowy dryf temperaturowy napięcia stabilizacji w diodach o $U_Z = 5...7$ V jest bardzo cenną właściwością tych diod, wynikającą z kompensacji zmian temperaturowych mechanizmu Zenera i mechanizmu lawinowego. Jednak należy zauważyć, że idealna kompensacja tych zmian może nastąpić tylko w jednym, ściśle określonym punkcie pracy, co ilustrują wykresy charakterystyk prądowo-napięciowych dla dwu różnych diod przy dwu różnych temperaturach (rys. 4.27). W za-

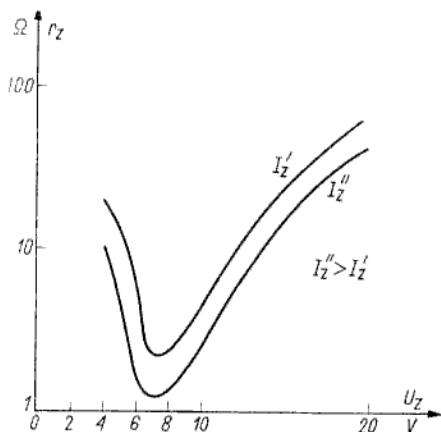


Rys. 4.27
Charakterystyki $I(U)$ dla dwu diod Zenera (A i B) przy dwu różnych temperaturach

kresie małych wartości prądu I_Z przeważa mechanizm Zenera, czyli $TKU_Z < 0$, w miarę natomiast wzrostu prądu I_Z coraz większy jest udział mechanizmu lawinowego i współczynnik TKU_Z przyjmuje wartości większe niż zero. Bardzo dobrą stabilność temperaturową uzyskuje się w tzw. stabilitronach skompensowanych, zbudowanych w ten sposób, że we wspólnej obudowie znajduje się dioda Zenera na napięcie U_Z powyżej 6 V (o przebicciu lawinowym), połączona szeregowo ze złączem pracującym w kierunku przewodzenia. W tym przypadku dodatni TKU_Z diody Zenera o przebicciu lawinowym jest kompensowany przez ujemny dryf temperaturowy charakterystyki diody pracującej w kierunku przewodzenia (ok. -2 mV/°C). Przy odpowiednim doborze obu diod można uzyskać stabili-

tron o współczynniku $TKU_z < 10^{-5}/^{\circ}\text{C}$. Stabilित्रony skompensowane są wytwarzane na napięcie powyżej 6.2 V, przy czym mogą to być połączenia szeregowe więcej niż dwu złączy *p-n*.

Rezystancja dynamiczna silnie zależy od wartości napięcia stabilizacji (tj. od rodzaju diody) i prądu stabilizacji (tj. od punktu pracy) — rys. 4.28. Minimum rezystancji



Rys. 4.28

Zależność rezystancji dynamicznej r_z od napięcia stabilizacji U_z dla dwu wartości prądu stabilizacji I_z

dynamicznej obserwuje się dla napięć 6...8 V, tj. w diodach, które „przebijają się” wskutek łącznego działania mechanizmów Zenera i lawinowego.

Warto zwrócić uwagę, że do stabilizacji małych napięć (poniżej 3 V) stosuje się diody pracujące w zakresie przewodzenia, przy czym są to diody krzemowe dyfuzyjne, charakteryzujące się dość gwałtownym wzrostem prądu dla napięcia polaryzacji, przekraczającego wartość napięcia progowego. Napięcie stabilizacji jest równe napięciu progowemu, czyli wynosi 0,75 V dla pojedynczego złącza, lecz we wspólnej obudowie mogą być połączone szeregowo dwa lub trzy złącza, co umożliwiłoby uzyskanie napięć stabilizacji 1,5 oraz 2,2 V. W Polsce są produkowane takie diody pod nomenklaturą BAP 814 — 816.

Parametry techniczne kilku typowych stabilित्रonów zestawiono w tabl. 4.3.

Na marginesie omawianego tematu warto wspomnieć, że istnieją również diody stabilizujące prąd. Działają one na zupełnie innej zasadzie niż stabilित्रony — są to tzw. polowe ograniczniki prądu, których działanie poznamy w rozdziale o tranzystorach polowych.

Analiza graficzna

4.3.1

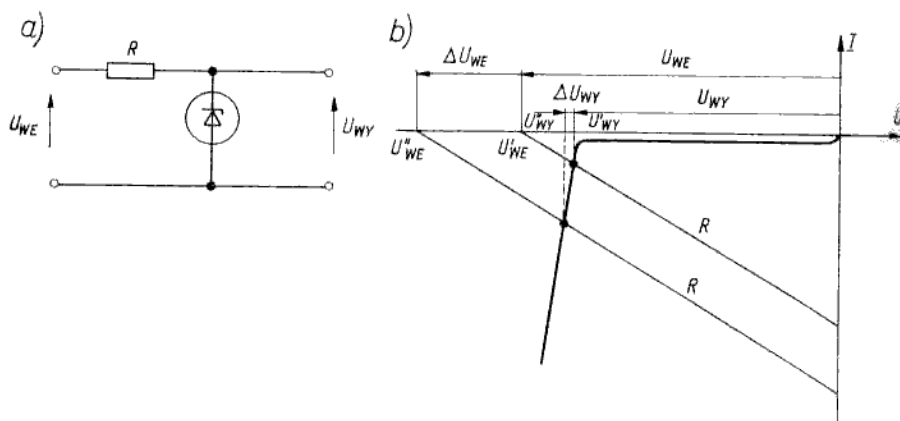
Ogólne zasady analizy graficznej prostych układów z elementami nieliniowymi były omawiane w p. 4.2.1. W tym miejscu rozpatrzmy przykład analizy najprostszego układu stabilizacji napięcia (rys. 4.29).

Punkt przecięcia charakterystyki prądowo-napięciowej z prostą obciążenia wyznacza wartość prądu płynącego w obwodzie i napięcie na stabilित्रonie, czyli napięcie wyjściowe. Jeżeli napięcie wejściowe zmienia się w przedziale $U'_{WE} \leq U_{WE} \leq U''_{WE}$, to napięcie wyjściowe zmienia się w znacznie mniejszym przedziale $U'_{WY} \leq U_{WY} \leq U''_{WY}$. W przypadku idealnego stabilित्रonu (o rezystancji dynamicznej $r_z = 0$) napięcie na wyjściu byłoby stałe.

Tablica 4.3

Parametry techniczne przykładowo wybranych stabilizatorów

Typ diody	Parametry charakterystyczne						Dopuszczalne parametry graniczne				U w a g i			
	I_R μA	przy U_R V	U_F V	przy I_F A		U_Z V	$\frac{\Delta U_Z}{U_Z}$ %	r_Z Ω	TK_{U_Z} $10^{-4}/^{\circ}C$	przy I_F mA		T_J $^{\circ}C$		
				I_F A	U_Z V					I_F A			I_Z A	P W
BAP 814	—	—	—	—	—	0,7	15	—	—	0,04	0,04	—	—	Kierunek przewodze- nia } produkcji polskiej
BZP 611-C4V7	0,5	1	1	0,1	4,7	10	100	-4	5	0,3	P_{max}/U_Z	0,25	150	
BZP 611-C6V8	0,1	1	1	0,1	6,8	10	15	0	5	0,3	P_{max}/U_Z	0,25	150	
BZP 611-C22	0,1	1	1	0,1	22	10	170	+8	5	0,3	P_{max}/U_Z	0,25	150	
BZP 620-D6V8	1	1	1,1	0,5	6,8	20	2	+4	100	3	P_{max}/U_Z	1	150	



Rys. 4.29

Prosty układ stabilizacji napięcia: a) schemat układu; b) analiza graficzna

Schematy zastępcze

4.3.2

W uproszczonej analizie układów ze stabilitronami można stosować modele odcinkami liniowe. Trzy takie modele, o kolejno wzrastającej dokładności, przedstawiono na rys. 4.30, na którym wykreślono również aproksymowane charakterystyki $I(U)$, odpowiadające poszczególnym modelom.

W najprostszym modelu (rys. 4.30a) do gałęzi reprezentującej kierunek przewodzenia (znanej z p. 4.1.2) dołączono gałąź reprezentującą kierunek zaporowy. W obwodzie zawierającym źródło U_Z połączone szeregowo z diodą idealną pojawia się prąd w przypadku spolaryzowania katody względem anody napięciem dodatnim o wartości większej niż napięcie stabilizacji U_Z . W bardziej dokładnym modelu (rys. 4.30b) uwzględniono skończoną rezystancję dynamiczną r_Z w zakresie stabilizacji. W trzecim modelu (rys. 4.30c) dołączono dodatkowo gałąź zawierającą diodę idealną połączoną z rezystorem R_u . W ten sposób w modelu została uwzględniona skończona (aczkolwiek bardzo duża) rezystancja diody w zakresie zaporowym przed przebiciem.

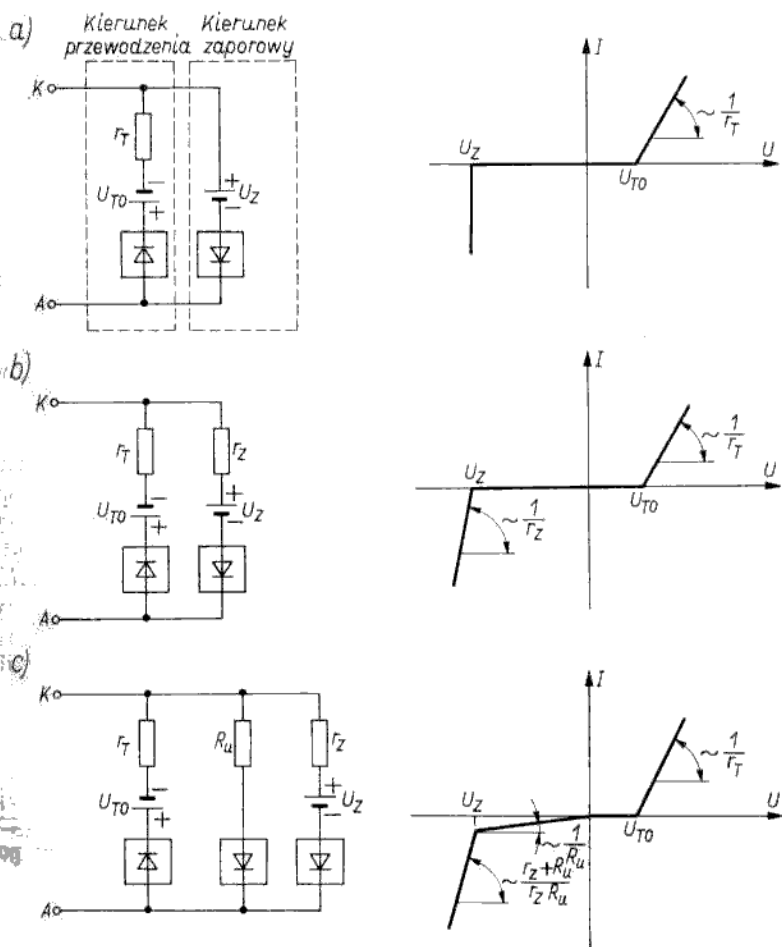
Przykłady zastosowań

4.3.3

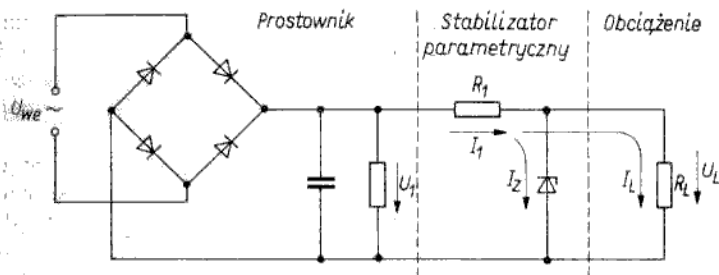
Stabilitrony najczęściej są stosowane w układach zasilających do stabilizacji napięcia. Stabilizatory napięcia stałego można podzielić na dwie grupy:

- układy bez sprzężenia zwrotnego (parametryczne),
- układy ze sprzężeniem zwrotnym.

Ponieważ w drugiej grupie układów jako elementy regulacyjne i wzmacniające są stosowane tranzystory, które poznamy dopiero w następnym rozdziale, w tym miejscu ograniczymy się do omówienia działania prostego zasilacza ze stabilizatorem parametrycznym, którego schemat przedstawiono na rys. 4.31. Układ podzielono na trzy sekcje, tj.: prostownik, stabilizator parametryczny, obciążenie. Stabilizator składa się z diody Zenera i rezystancji szeregowej R_1 . Napięcie stałe U_Z na wyjściu prostownika zmienia się w pewnym przedziale wskutek zmian amplitudy



Rys. 4.30 Odcinkami liniowe modele stabilizatorów od najprostszego a) do najbardziej złożonego c)



Rys. 4.31 Schemat zasilacza ze stabilizatorem parametrycznym

tudy napięcia wejściowego U_{we} oraz zmian rezystancji obciążenia R_L (te ostatnie zmiany mają wpływ na wartość napięcia U_1 wskutek tego, że prostownik ma pewną rezystancję wewnętrzną nie narysowaną na schemacie). Zadaniem stabilizatora jest uniezależnienie napięcia wyjściowego U_L od zmian napięcia U_{we} (a więc również napięcia U_1) oraz od zmian obciążenia.

Przez rezystor R_1 płynie prąd I_1 , który rozgałęzia się na prąd diody Zenera I_Z oraz prąd obciążenia I_L . Z prawa Kirchhoffa

$$U_L = U_1 - R_1(I_L + I_Z) \quad (4.32)$$

Stąd wynika, że $U_L = \text{const}$ przy zmianach napięcia U_1 o ΔU_1 , jeżeli

$$\Delta U_1 - R_1 \Delta I_Z = 0$$

oraz $U_L = \text{const}$ przy zmianach prądu obciążenia I_L o wartość ΔI_L , jeżeli

$$\Delta I_L = -\Delta I_Z$$

Proces stabilizacji napięcia wyjściowego U_L przy zmianach napięcia U_1 można opisać w następujący sposób. Niech U_1 wzrośnie o wartość ΔU_1 i niech to spowoduje przyrost napięcia wyjściowego o $\Delta U'_L$. W wyniku przyrostu napięcia wyjściowego (tj. napięcia na diodzie Zenera) o $\Delta U'_L$ przez diodę popłynie prąd większy o ΔI_Z (przyrost prądu wyjściowego ΔI_L wskutek przyrostu napięcia $\Delta U'_L$ można uznać za pomijalnie mały). Wskutek przyrostu prądu I_Z o ΔI_Z zwiększy się odpowiednio spadek napięcia na rezystancji R_1 , a więc przyrost $\Delta U'_L$ zmniejszy się do pewnej wartości $\Delta U_L \ll \Delta U'_L$. Współczynnik stabilizacji napięciowej układu określa się jako

$$S_u = \frac{\Delta U_L}{\Delta U_1} \quad (4.33)$$

W omawianym układzie można w praktyce uzyskać $S_u \approx 0,01$.

Ponieważ $\Delta U_1 \approx R_1 \Delta I_Z$, współczynnik stabilizacji można wyrazić w innej postaci

$$S_u = \frac{r_Z}{R_1} \quad (4.34)$$

Proces stabilizacji napięcia wyjściowego U_L od zmian prądu obciążenia I_L (przy $U_L = \text{const}$) można opisać w następujący sposób. Niech I_L zmaleje o wartość ΔI_L . Wówczas zmaleje spadek napięcia na rezystancji R_1 , a więc zwiększy się wartość napięcia wyjściowego o $\Delta U'_L$. To z kolei spowoduje przyrost prądu stabilitronu o ΔI_Z , czyli zwiększenie spadku napięcia na rezystancji R_1 , przeto zmniejszenie spadku napięcia $\Delta U'_L$ do pewnej wartości $\Delta U_L \ll \Delta U'_L$. Stabilizacja będzie tym lepsza, im mniejszej zmiany napięcia ΔU_L będzie odpowiadać większa zmiana prądu ΔI_Z , czyli im mniejsza będzie wartość rezystancji dynamicznej stabilitronu ($r_Z = \Delta U_L / \Delta I_Z$).

Przez stabilitron płynie maksymalny prąd w stanie rozwarcia na wyjściu. Stąd można określić prądowy zakres pracy stabilitronu. Jeżeli przy maksymalnym prądzie obciążenia I_{Lmax} przez diodę płynie prąd I_{Zmin} (powinien to być prąd na tyle duży, by dioda pracowała w zakresie stabilizacji), to w stanie rozwarcia ($R_L = \infty$) przez diodę płynie prąd

$$I_{Zmax} = I_{Zmin} + I_{Lmax} \quad (4.35)$$

Należy zatem wybrać diodę o dopuszczalnym prądzie stabilizacji większym niż I_{Zmax} .
 Rezystancję szeregową R_1 dobiera się w taki sposób, aby ograniczyć maksymalny prąd stabilitronu do wartości I_{Zmax} , czyli

$$R_1 = \frac{U_1 - U_L}{I_{Zmax}} \quad (4.36)$$

Diody impulsowe

4.4

Diodami impulsowymi lub przełącznikowymi nazywa się diody przeznaczone do zastosowań w układach impulsowych, w których najczęściej spełniają one funkcje kluczy przepuszczających impulsy tylko w jednym kierunku. Z uwagi na takie zastosowania dioda impulsowa powinna spełniać przede wszystkim następujące wymagania:

- bardzo mała rezystancja w jednym kierunku oraz bardzo duża rezystancja w kierunku przeciwnym;
- bezzwłoczna reakcja na impulsy, czyli brak opóźnień i zniekształceń impulsów.

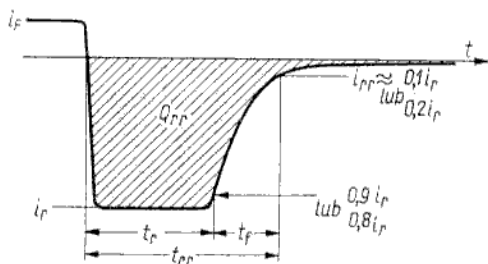
Dlatego w grupie parametrów charakterystycznych diod impulsowych obok dwu standardowych parametrów statycznych:

- napięcie przewodzenia U_F przy określonym prądzie przewodzenia I_F ,
- prąd wsteczny I_R przy określonym napięciu wstecznym U_R , zawsze podaje się parametry dynamiczne,
- czas przełączania t_{rr} (niekiedy zamiast czasu t_{rr} podaje się ładunek przełączania Q_{rr}) przy określonych warunkach wysterowania i obciążenia diody,
- pojemność diody przy określonej częstotliwości i napięciu polaryzacji wstecznej diody.

Ponadto wyróżnia się następujące dopuszczalne parametry graniczne:

- maksymalny prąd stały (lub prąd średni) I_{Fmax} w kierunku przewodzenia;
- maksymalny szczytowy prąd przewodzenia I_{FMmax} ;
- maksymalne stałe napięcie wsteczne U_{Rmax} ;
- maksymalne szczytowe napięcie wsteczne U_{RMmax} ;
- dopuszczalna temperatura złącza T_j .

Dodatkowego komentarza wymaga bardzo ważne pojęcie czasu przełączania t_{rr} . Przypomnijmy z p. 3.1 przebieg prądu przy przełączaniu diody z kierunku przewodzenia



Rys. 4.32
 Definicje czasów charakteryzujących proces przełączania diody

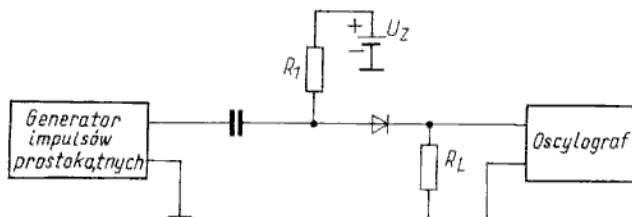
do kierunku zaporowego (rys. 4.32). Czas przełączania t_{rr} jest definiowany jako odstęp czasu od chwili odpowiadającej $i = 0$ do chwili, gdy prąd osiągnie wartość i_{rr} , czyli zmaleje do ok. $0,1i_r$.

Dla niektórych rodzajów diod impulsowych (przede wszystkim dla diody ładunkowej, omawianej nieco dalej) istotna jest informacja o czasie trwania poszczególnych dwu faz procesu przełączania, tj. czasu t_r — „półka” prądu oraz czasu opadania t_f , określanego jako odstęp czasu od chwili, gdy prąd wsteczny przyjmuje wartość $0,9i_r$ do chwili odpowiadającej prądowi wstecznemu $0,1i_r$ (często przyjmuje się od $0,8i_r$ do $0,2i_r$). Zarówno całkowity czas przełączania t_{rr} jak i poszczególne jego składowe zależą od warunkówysterowania i obciążenia diody, dlatego w katalogach podaje się te wielkości dla określonych wartości następujących parametrów:

- prąd przewodzenia i_F (im większy jest prąd i_F , tym większy jest ładunek zmagazynowany w bazie złącza $p-n$, a więc dłużej trwa proces rozładowania bazy — wydłuża się czas t_{rr});
- prąd „półki” w kierunku zaporowym i_r (im większy jest prąd i_r , tym szybciej jest usuwany ładunek zmagazynowany w bazie złącza $p-n$, a więc czas t_{rr} maleje);
- wartość prądu i_{rr} .

Zamiast wartości i_r można podać wartość napięcia wstecznego U_R i rezystancji obciążenia R_L , gdyż $i_r \approx U_R/R_L$. Niekiedy zamiast czasu t_{rr} , a częściej jako informację dodatkową, podaje się wartość ładunku Q_{rr} (rys. 4.32). W ogólnym przypadku nie ma ścisłej zależności między Q_{rr} a t_{rr} . Tylko dla diod, w których $t_r/t_f \gg 1$, można przyjąć, że $Q_{rr} = i_r t_{rr}$. Ponieważ $Q_{rr} = i_F \tau$, zamiast ładunku Q_{rr} można podać czas życia nośników τ .

Metodę pomiaru parametrów dynamicznych diod impulsowych przedstawiono schematycznie na rys. 4.33. Z generatora poprzez kondensator poprzez rezystor R_1 i diodę



Rys. 4.33

Układ do pomiaru parametrów dynamicznych diod

stałej są podawane impulsy polaryzujące diodę w kierunku zaporowym. Dioda poprzez rezystor R_1 jest polaryzowana statycznie w kierunku przewodzenia. Na rezystorze R_L ($R_L \ll R_1$) odkłada się napięcie proporcjonalne do prądu płynącego w obwodzie, przebieg prądu można zatem obserwować na ekranie oscylografu.

Wspomniano już wcześniej, że podstawowym warunkiem, jaki powinna spełniać dioda impulsowa, jest bezzwłoczność jej działania, czyli jak najmniejszy czas przełączania t_{rr} . Niekiedy czas t_{rr} jest mało istotny, pożądana jest natomiast jak największa stromość impulsu prądu otrzymywanego podczas przełączania — zatem istotne jest zmniejszenie czasu opadania t_f . Biorąc pod uwagę te dwie tendencje, diody impulsowe podzielimy na:

- diody o małym czasie przełączania t_{rr} ,
- diody o małym czasie opadania t_f .

Diody o małym czasie przełączania

4.4.1

Jako diody impulsowe o małym czasie przełączania najwcześniej zaczęto stosować diody ostrzowe germanowe o złączu p - n formowanym elektrycznie, przy czym szczególnie duże szybkości przełączania uzyskano w diodach o styku złoto-german. Diody ostrzowe mają dużą szybkość przełączania dzięki małym powierzchniom, a więc i małym pojemnościom złączy.

W nowoczesnych diodach impulsowych wykorzystuje się epiplanarne złącza p - n domieszkowane złotem oraz złącza m - s . Technologia epiplanarna umożliwia wytwarzanie złączy o małych powierzchniach, a domieszkowanie złotem ma na celu zmniejszenie czasu życia nośników mniejszościowych (atomy złota spełniają funkcję centrów generacyjno-rekombinacyjnych). Jednakże nie można stosować zbyt dużych koncentracji atomów złota, gdyż następuje pogorszenie charakterystyki statycznej w kierunku zaporowym, tj. wzrasta prąd generacji. Dlatego domieszkowanie złotem, praktycznie biorąc, umożliwia zmniejszenie czasu przełączania do ok. 1 ns. Krótsze czasy przełączania diod p - n można uzyskać w przypadku zastosowania materiału półprzewodnikowego o szerszym pasmie zabronionym. Na przykład diody z GaAs mają czasy przełączania mniejsze niż 0,1 ns. Jeszcze krótsze czasy przełączania uzyskuje się w diodach Schottky'ego. Duża szybkość działania diod Schottky'ego — jak wiadomo z p. 3.4 — wynika stąd, że są to diody działające na nośnikach większościowych, a więc wolne od inercji spowodowanej magazynowaniem ładunku w bazie.

Diody o małym czasie opadania (diody ładunkowe)

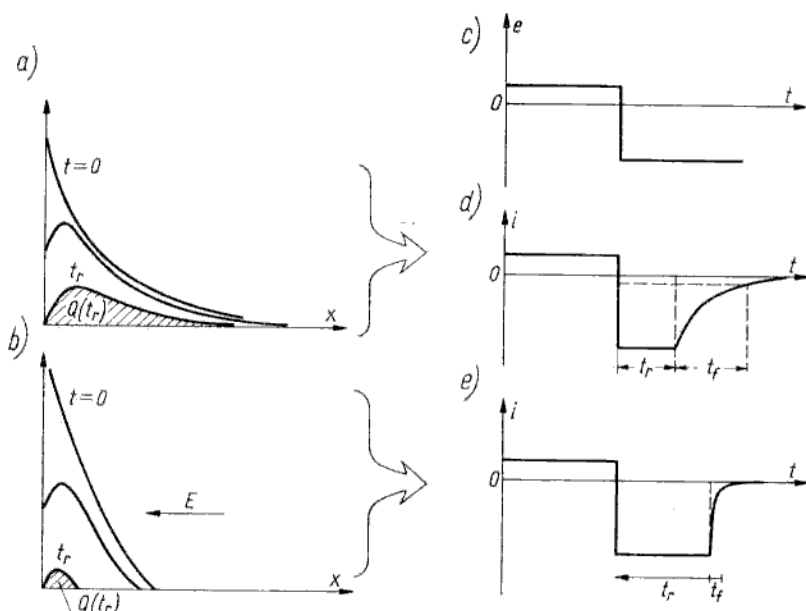
4.4.2

W dotychczas rozpatrywanych diodach obie fazy procesu przełączania mają porównywalne czasy trwania dla $i_F \approx i_r$, przy czym $t_r \approx t_f \approx 0,25\tau$ (τ — czas życia nośników). Dla $i_r > i_F$ czas t_f , praktycznie biorąc, nie zmienia się, a czas t_r maleje, czyli $t_f/t_r > 1$.

Do celów formowania bardzo stromych impulsów prostokątnych byłaby przydatna dioda o bardzo małym czasie t_f , dla której $t_f/t_r \ll 1$. Takie właściwości ma dioda ładunkowa¹⁾. Na rysunku 4.34 porównano rozkłady koncentracji nośników nadmiarowych w bazie diody konwencjonalnej i ładunkowej oraz przedstawiono odpowiednie przebiegi prądów podczas przełączania. Bezpośrednią przyczyną skrócenia fazy opadania prądu w diodzie ładunkowej jest zmniejszenie ładunku $Q(t_r)$, czyli ładunku, jaki pozostaje w bazie w chwili zakończenia „półki” prądu. Ładunek $Q(t_r)$ jest mniejszy w diodzie ładunkowej, gdyż w czasie t_r nośniki nadmiarowe są usuwane z bazy nie tylko wskutek dyfuzji, lecz również wskutek uniesienia w polu elektrycznym E . Na rysunku 4.34b jest widoczne, że kolejne krzywe rozkładu koncentracji dziur przesuwiają się w lewo zgodnie z kierunkiem działania pola elektrycznego.

Należy jeszcze odpowiedzieć na pytanie: skąd się bierze pole elektryczne w diodzie ładunkowej? W pierwszym rozdziale (p. 1.9) rozpatrywaliśmy właściwości tzw. półprzewodnika niejednorodnego, tj. o nierównomiernym rozkładzie koncentracji domieszek. Zauważyliśmy wówczas, że w półprzewodniku niejednorodnym istnieje wbudowane pole elektryczne, spowodowane niepełną neutralizacją ładunków nieruchomych jonów domieszki przez ładunki ruchome elektronów lub dziur.

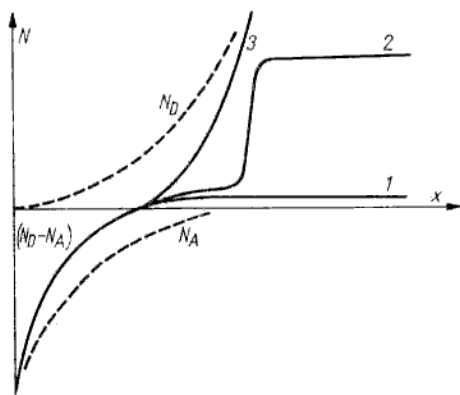
¹⁾ Dioda ładunkowa w literaturze angielskiej ma kilka nazw — *charge storage diode*, *step recovery diode*, *snap-off diode*.



Rys. 4.34

Porównanie rozkładów koncentracji nośników nadmiarowych i przebiegów prądu w czasie przełączania diody konwencjonalnej (a, d) oraz ładunkowej (b, e); c) napięcie sterujące

Z tej właściwości półprzewodnika niejednorodnego korzysta się w diodzie ładunkowej. Typowe rozkłady koncentracji domieszek dla trzech różnych konstrukcji diod ładunkowych przedstawiono na rys. 4.35. Wspólną cechą wszystkich trzech konstrukcji jest wzrastająca koncentracja domieszki donorowej w bazie, co po-



Rys. 4.35

Typowe rozkłady koncentracji domieszek dla trzech różnych konstrukcji diod ładunkowych:

1 — struktura planarna; 2 — struktura epiplanarna; 3 — struktura dwustronnie domieszkowana

woduje powstanie pola elektrycznego, hamującego dziury wstrzykiwane do bazy. W czasie przełączania pole elektryczne spychając dziury w kierunku warstwy zaporowej złącza $p-n$ przyczynia się do wydłużenia pierwszej i skrócenia drugiej fazy procesu przełączania. Diody ładunkowe wytwarza się z krzemu o dość du-

Tablica 1.1

Parametry techniczne przykładowo wybranych diod impulsowych

Typ diody	Parametry charakterystyczne						Dopuszczalne parametry graniczne					U w a g i		
	statyczne			dynamiczne			I_{Fmax}	I_{FMmax}	U_{Rmax}	U_{RMmax}	T_J			
	U_F	przy		t_{rr}	przy									
		I_F	I_R		U_R	R_L	I_F	C_d	U_R					
V	mA	nA	V	Ω	mA	pF	V	mA	V	$^{\circ}C$				
BAYP 94	0,8	20	5	25	6	100	10	10	4	0	200	35	200	Produkcji polskiej t_{rr} mierzone przy $I_F = 10$ mA $t_{rr} =$ $= 1$ mA (prod. Siemens)
BAY 61	1	10	25	20			10	10	4	0	200	100	200	
MBD 702	0,6	10	250	3					1	0	—	—	70	Dioda Schottky'ego (prod. Motorola)

żym czasie życia nośników mniejszościowych, zwykle $\tau \approx 1 \mu\text{s}$ (im większy jest czas życia, tym dłuższa jest droga dyfuzji, a więc większy jest względny wpływ pola elektrycznego na pracę diody, gdyż transport unoszeniowy ma coraz większy udział w porównaniu z dyfuzją).

W tabelicy 4.4 zestawiono parametry techniczne przykładowo wybranych diod impulsowych.

Analiza graficzna i schematy zastępcze

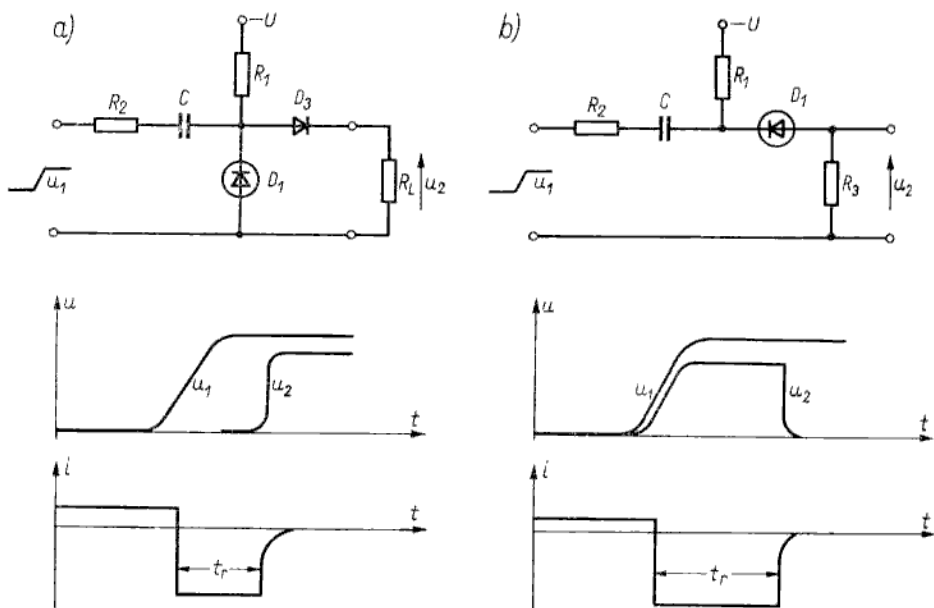
4.4.3

Metody graficzne analizy mają małą przydatność w układach impulsowych i są stosowane tylko do wyznaczania statycznych warunków pracy. Przydatność schematów zastępczych do analizy układów z diodami impulsowymi jest również bardzo ograniczona, gdyż — jak wiadomo z p. 3.1 — model dynamiczny dla pracy wielkosygnałowej złącza *p-n* wymaga wprowadzenia elementów zależnych od punktu pracy i czasu. Niekiedy stosuje się uproszczone schematy zastępcze, w których elementy mają wartości uśrednione dla określonego przedziału pracy.

Zastosowanie diod impulsowych

4.4.4

Z wielu różnych układów z diodami impulsowymi omówimy krótko przykład zastosowania diody ładunkowej w układzie formującym zbocze impulsu prostokątnego.



Rys. 4.36

Dwa proste układy ilustrujące zastosowanie diod do kształtowania zbocza impulsów prostokątnych: a) układ formowania zbocza narastającego z odpowiednimi przebiegami; b) układ formowania zbocza opadającego z odpowiednimi przebiegami

Schematy dwóch prostych układów formowania zbrocza narastającego lub opadającego przedstawiono odpowiednio na rys. 4.36a, b. W pierwszym układzie dioda ładunkowa D_1 jest polaryzowana stałym prądem w kierunku przewodzenia ze źródła $-U$ przez rezystor R_1 . Impuls formowany u_1 przez rezystor R_2 i kondensator C polaryzuje diodę D_1 w kierunku zaporowym. W czasie t_r dioda przewodzi w kierunku zaporowym, czyli stanowi małą rezystancję, na której — praktycznie biorąc — nie odkłada się napięcie. Po czasie t_r dioda raptownie odzyskuje właściwości zaporowe, czyli jej rezystancja zwiększa się wielokrotnie. Napięcie, jakie wówczas pojawia się na diodzie D_1 , jest podawane przez diodę D_3 na rezystancję obciążenia R_L . Dioda D_3 , spełniająca funkcję klucza blokującego napięcie stałe „ $-U$ ” od obciążenia, musi mieć bardzo mały czas przełączania. W układzie przedstawionym na rys. 4.36b dioda ładunkowa jest włączona szeregowo z obciążeniem, dlatego jest formowane zbrocze opadające. W obu układach można regulować w pewnym zakresie czas t_r , zmieniając prąd polaryzacji diody ładunkowej w kierunku przewodzenia (zmieniając „ $-U$ ” lub R_1).

Na zakończenie warto wspomnieć, że diody ładunkowe są również szeroko stosowane w układach nieimpulsowych w zakresie bardzo wielkich częstotliwości (modulatory, powielacze częstotliwości itp.).

Diody pojemnościowe

4.5

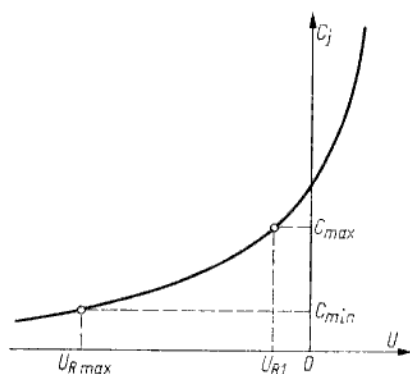
Diody pojemnościowe są to diody $p-n$ przeznaczone do zastosowań, w których wykorzystuje się — znane z poprzedniego rozdziału — zjawisko zmian pojemności warstwy zaporowej złącza $p-n$ pod wpływem polaryzacji w kierunku wstecznym. Pojemność bariery jest właściwością charakterystyczną każdego złącza $p-n$ lub $m-s$. Jednak dioda przeznaczona do zastosowań jako pojemność zmienna ma konstrukcję optymalizowaną pod kątem takich zastosowań.

Na ogół stawia się znak równości między nazwami *dioda pojemnościowa*, *warikap* i *waraktor*, lecz — ściśle biorąc — diody pojemnościowe należy podzielić na warikapy (ang. *VARIABLE CAPACITANCE*) oraz waraktory (ang. *VARIABLE REACTOR*). Warikapy są to diody pojemnościowe przeznaczone do zastosowań jako zmienne pojemności (np. w układach automatycznego przestrajania obwodów rezonansowych), waraktory natomiast są to diody pojemnościowe przeznaczone do zastosowań jako zmienne reaktancje, spełniające funkcję elementów czynnych w układach parametrycznych (np. wzmacniacze parametryczne lub parametryczne powielacze częstotliwości). W zależności od określonych zastosowań jest ustalany zestaw podstawowych parametrów technicznych, dlatego specyfikacje parametrów różnią się nieco dla warikapów i waraktorów.

Właściwości warikapów określają następujące parametry techniczne.

Parametry charakterystyczne:

- pojemność złącza C_j przy określonej częstotliwości i napięciu polaryzacji wstecznej (zwykle przy napięciu bliskim wartości maksymalnej);
- stosunek pojemności C_j przy dwu różnych wartościach napięcia polaryzacji wstecznej (zwykle przy napięciu bliskim zera — pojemność maksymalna oraz przy napięciu bliskim wartości maksymalnej — pojemność minimalna) — rys. 4.37;
- rezystancja szeregowo R_s lub dobroć Q przy określonej częstotliwości i napięciu polaryzacji wstecznej (lub pojemności).



Rys. 4.37
Charakterystyka pojemnościowo-napięciowa warikapu, na której zaznaczono dwa punkty charakterystyczne

Parametry dopuszczalne graniczne:

- maksymalne napięcie stałej polaryzacji wstecznej U_{Rmax} ;
- maksymalne szczytowe napięcie wsteczne U_{RMmax} ;
- maksymalny stały prąd przewodzenia I_{Fmax} ;
- dopuszczalna temperatura złącza T_j .

Właściwości waraktorów określają następujące parametry techniczne.

Parametry charakterystyczne:

- pojemność złącza C_j przy określonej częstotliwości i napięciu polaryzacji wstecznej (zwykle przy $U_R = 0$, czyli jest to pojemność maksymalna);
- stosunek pojemności C_j przy dwu różnych wartościach napięcia polaryzacji wstecznej (zwykle przy $U_R = 0$ — pojemność maksymalna oraz przy $U_R = U_{Rmax}$ — pojemność minimalna);
- prąd wsteczny I_R przy określonym napięciu wstecznym U_R (zwykle przy $U_R = U_{Rmax}$);
- częstotliwość maksymalna przy określonym napięciu polaryzacji wstecznej U_R ;
- wartości parametrów pasożytniczych, tj. indukcyjności szeregowej wyprowadzeń L_s oraz pojemności pasożytniczej oprawki C_p .

Parametry dopuszczalne graniczne:

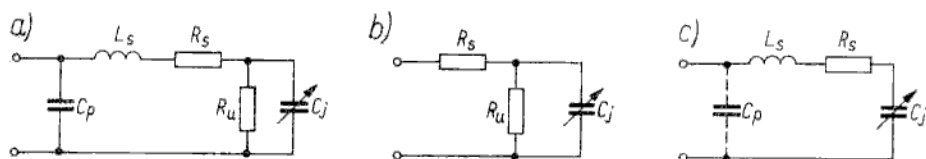
- maksymalne napięcie stałej polaryzacji wstecznej U_{Rmax} ;
- maksymalna moc rozpraszana P_{max} ;
- dopuszczalna temperatura złącza T_j lub dopuszczalna temperatura otoczenia T_a .

Ponieważ w parametrach technicznych występują wielkości określające elementy schematu zastępczego diody pojemnościowej, dodatkowe objaśnienia do ważniejszych parametrów poprzedzimy krótkim omówieniem schematów zastępczych.

Schematy zastępcze diod pojemnościowych

4.5.1

Na rysunku 4.38 przedstawiono ogólny schemat zastępczy diody pojemnościowej (a) oraz jego dwie uproszczone wersje — dla warikapu (b) i waraktora (c). Różnice między dwiema uproszczonymi wersjami schematu zastępczego wynikają stąd, że warikap pracuje w zakresie mniejszych częstotliwości (do kilkuset megaherców), zatem wpływ indukcyjności doprowadzeń L_s i pojemności rozproszonej C_p na jego pracę jest pomijalnie mały. Waraktor jest elementem mikrofalowym (pracuje w zakresie częstotliwości kilkudziesięciu do kilkuset gigaherców), dlatego należy uwzględniać elementy L_s , C_p , lecz nie uwzględnia się rezystancji



Rys. 4.38

Schemat zastępczy diody pojemnościowej (a) i jego dwie uproszczone wersje dla warikapu (b) i waraktora (c)

upływu R_u mającej istotny wpływ na dobroć diody tylko w zakresie małych częstotliwości.

Na podstawie schematu zastępczego z rys. 4.38b można wyrazić dobroć (stosunek energii zmagazynowanej w pojemności do energii rozpraszanej w rezystancjach) w następującej postaci:

$$Q = \frac{\omega C_j R_u}{\frac{R_s}{R_u} + 1 + \omega^2 C_j^2 R_u R_s} \quad (4.37)$$

Ponieważ $R_s/R_u \approx 0$, zatem

$$Q \approx [(\omega C_j R_u)^{-1} + \omega C_j R_s]^{-1} \quad (4.38)$$

Różniczkując wyrażenie (4.38) względem częstotliwości można wyznaczyć maksymalną wartość dobroci i częstotliwość, przy której ona występuje:

$$Q_{max} = \frac{1}{2} \sqrt{\frac{R_u}{R_s}} \quad (4.39)$$

$$f|_{Q_{max}} = \frac{1}{2\pi C_j \sqrt{R_u R_s}} \quad (4.40)$$

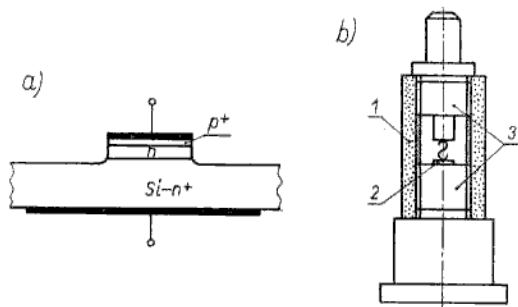
W zakresie małych częstotliwości przeważa pierwszy składnik we wzorze (4.38), czyli minimalną częstotliwość pracy wyznacza rezystancja upływu R_u . W zakresie dużych częstotliwości przeważa drugi składnik, czyli maksymalną częstotliwość pracy wyznacza rezystancja szeregową R_s . Przyjmując za minimalną wartość dobroci $Q = 1$ można wyznaczyć graniczne częstotliwości pracy diody pojemnościowej:

$$f_{min} = \frac{1}{2\pi C_j R_u} \quad (4.41)$$

$$f_{max} = \frac{1}{2\pi C_j R_s} \quad (4.42)$$

W celu uzyskania dużej wartości częstotliwości maksymalnej należy dążyć do zmniejszenia rezystancji szeregowej złącza, tj. stosować półprzewodnik silnie domieszkowany. Dlatego często waraktory są wytwarzane w warstwach epitaksjalnych krzemu na podłożu silnie domieszkowanym.

Na rysunku 4.39 przedstawiono strukturę i szkic konstrukcji typowego waraktora.



Rys. 4.39
Struktura (a) i szkic konstrukcji typowego waraktora (b)
1 — izolator ceramiczny; 2 — płytka Si ze złączem p-n; 3 — elektrody

Jeżeli rezystancja szeregową R_s jest bardzo mała, to ograniczenie maksymalnej częstotliwości pracy waraktora jest spowodowane zjawiskiem rezonansu szeregowego obwodu, jaki tworzy indukcyjność doprowadzeń L_s wraz z pojemnością C_j

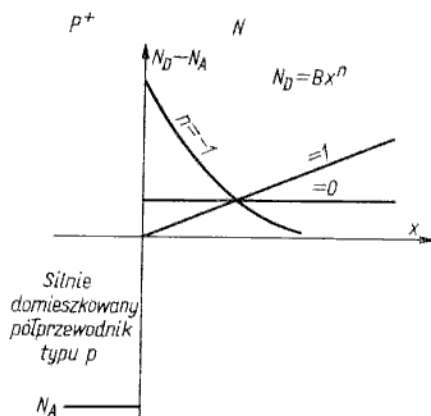
$$f_{gr} = \frac{1}{2\pi \sqrt{L_s C_j}} \quad (4.43)$$

Dodatkowego wyjaśnienia wymaga zagadnienie najważniejszej charakterystyki diody pojemnościowej, jaką jest zależność $C_j(U)$. Wiadomo z teorii złącza p-n, że

$$C_j = C_{j0} \left(\frac{U + \varphi_B}{\varphi_B} \right)^{-m} \quad (4.44)$$

przy czym: $m = 1/2$ dla złącza skokowego lub $1/3$ dla złącza liniowego; φ_B — napięcie dyfuzyjne, równe ok. 0,7 V dla złącza krzemowego.

Większość współcześnie produkowanych diod pojemnościowych są to dyfuzyjne złącza krzemowe lub stopowe złącza z GaAs, można je zatem aproksymować modelem złącza liniowego lub skokowego. Jednak ze względu na efektywne przestrajanie pojemności interesujące jest poszukiwanie takich rozwiązań konstrukcyjno-technologicznych, które umożliwiłyby uzyskanie jak największego nachylenia



Rys. 4.40
Rozkład koncentracji domieszek w złączu p^+-n , w którym $N_D = Bx^n$

charakterystyki $C_j(U)$. Weźmy złącze p^+-n , w którym profil domieszkowania w obszarze N jest opisany ogólną zależnością $N_D = Bx^n$. Na rysunku 4.40 przedstawiono trzy charakterystyczne przypadki, tj. $n = 1$ (złącze liniowe), $n = 0$

Tablica 45

Parametry techniczne kilku przykładowo wybranych warikapów

Typ diody	Parametry charakterystyczne										Dopuszczalne parametry graniczne			U w a g i
	C_j		przy U_R		$C_j(U_{R2})$		przy U_{R1}		Q		U_{RMmax}	I_{Fmax}	T_j	
	pF	MHz	f	MHz	V	V	U_{R1}	U_{R2}	R _s	Ω				
											MHz	V	V	
BS 105 GD	1,8—2,8	1	25	4,5—6	3	25	1,2	—	470	(9)	28	30	100	} Produkcji } polskiej
BBP 624	29—66	30	2	1,4	2	10	—	160	30	2	30	30	150	
BB 110 B	29—33	1	3	2,65	3	30	0,3	—	100	(30)	30	—	100	

Parametry techniczne kilku przykładowo wybranych waraktorów

Typ diody	Parametry charakterystyczne										Dopuszczalne parametry graniczne			U w a g i			
	C_j		przy U_R		I_R		przy U_R		f_r		R_s	przy U_R	C_p		F_{max}	U_{Rmax}	$T_j(T_0)$
	pF	MHz	V	V	μA	V	GHz	V	Ω	GHz							
											GHz	V	V		Ω	GHz	pF
BXYP	10—20	6	2	0/6	10	90	—	—	1	6	3	0,8	0,4	90	(60)		
BXYP 51	0,5—0,9	6	2	0/6	10	12	120	6	—	—	—	3	0,3	12	(70)		

(złącze skokowe), $n < 0$ (złącze hiperboliczno-skokowe). Zależność $C_j(U)$ jest wyrażona wzorem (4.44), przy czym

$$m = \frac{1}{n+2} \quad (4.45)$$

Nachylenie charakterystyki pojemnościowo-napięciowej wzrasta zatem w miarę, jak współczynnik n przybiera coraz większe wartości ujemne. W przypadku $n = -3/2$ wykładnik $m = 2$ i wówczas częstotliwość rezonansowa obwodu L_s, C_j

$$f = \frac{1}{2\pi\sqrt{L_s C_j}} \sim \frac{1}{\sqrt{C_j}} \sim U$$

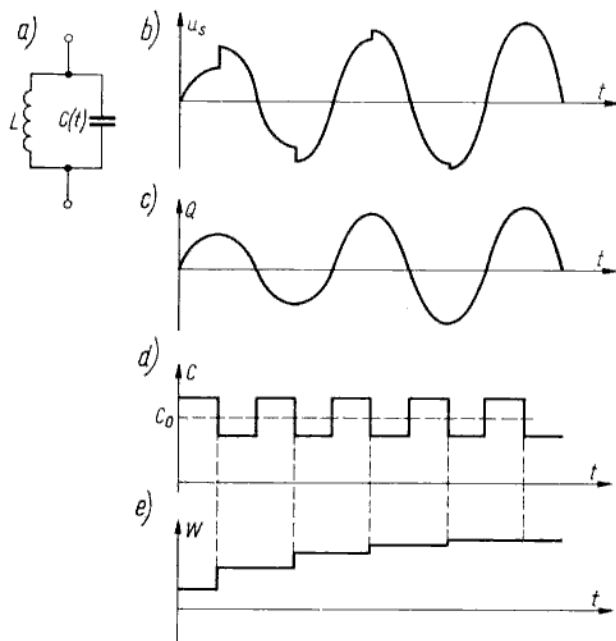
jest liniowo zależna od napięcia. Taka zależność może być bardzo dogodna do wielu zastosowań układowych.

Parametry techniczne kilku typowych warikapów i waraktorów zestawiono w tabl. 4.5.

Przykłady zastosowań

4.5.2

Diody pojemnościowe znajdują zastosowanie przede wszystkim w układach automatycznie strojonych obwodów rezonansowych, wzmacniaczach i generatorach parametrycznych, powielaczach częstotliwości, przełącznikach sygnałów mikrofalowych.



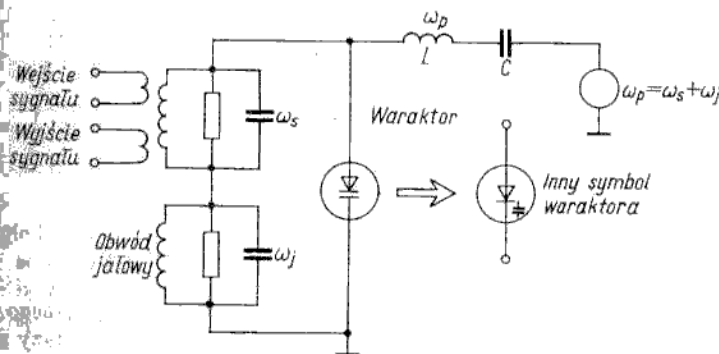
Rys. 4.41

Ilustracja zasady działania wzmacniacza parametrycznego: a) układ; b) przebieg sygnału wzmacnianego; c) przebieg ładunku w kondensatorze; d) zmiany pojemności w funkcji czasu; e) zmiany energii w funkcji czasu

Jako przykład rozpatrzmy podstawowe zasady działania wzmacniaczy parametrycznych, stanowiących bodajże najważniejszą dziedzinę zastosowań waraktorów. Ideę wzmacniacza parametrycznego — znaną już w XIX wieku, a realizowaną praktycznie dopiero w ostatnich dwudziestu latach — można objaśnić poglądowo na prostym przykładzie zilustrowanym przebiegami na rys. 4.41.

Jest dany obwód rezonansowy L, C , w którym kondensator ma zmienną odległość między okładkami. Okładki kondensatora w odpowiednich chwilach będziemy mechanicznie oddalać lub przybliżać do siebie. Zakładamy, że w chwili $t = 0$ obwód pobudzono do drgań o jego częstotliwości własnej. W chwili gdy cała energia obwodu jest zgromadzona w kondensatorze (maksimum wartości chwilowej ładunku w kondensatorze), rozsuwa się raptownie okładki kondensatora. Wówczas zmniejsza się pojemność kondensatora, a ponieważ $U = Q/C$, więc odpowiednio zwiększa się napięcie, a więc zwiększa się energia zmagazynowana w kondensatorze. Wzrost energii elektrycznej w kondensatorze następuje dzięki pracy mechanicznej, wykonanej w czasie rozsuwania przyciągających się okładek. Z kolei zbliżenie okładek, czyli powrót do pierwotnej pojemności kondensatora następuje w chwili, gdy ładunek i napięcie u_s przyjmują wartość zerową. Zatem tej zmianie pojemności nie towarzyszy zmiana energii obwodu. Jeżeli zmiany pojemności następują z częstotliwością 2-krotnie większą niż częstotliwość sygnału u_s , to energia w obwodzie L, C systematycznie zwiększa się kosztem pracy wykonywanej na rozsuwanie okładek kondensatora. Odpowiednio zwiększa się amplituda sygnału u_s . W rzeczywistym obwodzie L, C należy uwzględnić straty energii, które wzrastają w miarę zwiększania amplitudy sygnału i po pewnym czasie ustala się równowaga między energią dostarczaną a energią traconą przez obwód.

Podstawową ideę wzmacniacza parametrycznego realizuje się w praktyce stosując waraktor zamiast kondensatora o zmiennej odległości między okładkami. Pojemność waraktora zmienia się wraz z częstotliwością 2-krotnie większą niż częstotliwość sygnału wzmacnianego. W tym celu do waraktora należy doprowadzić odpowiednio zmienne napięcie polaryzacji ze źródła, które niejako pompuje energię do kondensatora. Stąd zmienne napięcie polaryzacji waraktora jest nazywane *napięciem pompującym*. Zasada działania wzmacniacza z waraktorem jest taka sama jak opisana dla kondensatora o rozsuwanych okładkach, tzn. napięcie pompujące ma częstotliwość 2-krotnie większą niż sygnał wzmacniany i fazy tych dwu przebiegów są dokładnie uzgodnione. W praktyce o wiele częściej są stoso-



Rys. 4.42

Układ dwuobwodowy wzmacniacza parametrycznego

wane bardziej rozbudowane układy wzmacniaczy parametrycznych, które nie narzucają tak krytycznych wymagań co do częstotliwości i fazy przebiegu pompującego. Na przykład w układzie dwuobwodowym (rys. 4.42) odpowiednie relacje fazowe ustalają się automatycznie. Obwód „górny” jest nastrojony na częstotliwość sygnału wzmacnianego f_s i nosi nazwę obwodu sygnałowego. Obwód „dolny”, nazywany jałowym, pracuje przy częstotliwości tzw. sygnału jałowego f_j . Oba obwody są sprzężone pojemnościowo przez waraktor, który jest zasilany przez separujący obwód szeregowy L, C sygnałem ze źródła pompującego. Istnieje przy tym ścisła zależność między częstotliwością sygnału pompującego f_p a dwiema pozostałymi częstotliwościami

$$f_p = f_s + f_j \quad (4.46)$$

Zmienną pojemność waraktora można zatem zapisać w postaci

$$C = [1 + m \sin(\omega_p t + \psi)]^{-1} C_0 \quad (4.47)$$

przy czym $m = \Delta C / C_0$ — wartość maksymalna względnej zmiany pojemności Ładunek w waraktorze

$$Q = Q_1 [\cos \omega_s t + \cos(\omega_j t + \psi)] \quad (4.48)$$

Napięcie na waraktorze $u = Q/C$ jest superpozycją napięcia sygnału u_s o pulsacji ω_s i napięcia jałowego u_j o pulsacji ω_j . Oczywiście jest jeszcze napięcie u_p ze źródła pompującego, które ma znacznie większą amplitudę niż u_s lub u_j i powoduje zmiany pojemności waraktora.

Po podzieleniu (4.48) przez (4.47) przy uwzględnieniu (4.46) otrzymuje się:

$$u_s = \frac{Q_1}{C_0} \left[\cos \omega_s t + \frac{1}{2} m \sin \omega_s t \right] \quad (4.49)$$

$$u_j = \frac{Q_1}{C_0} \left[\cos(\omega_j t + \psi) + \frac{1}{2} m \sin(\omega_j t + \psi) \right] \quad (4.50)$$

Moc dla poszczególnych sygnałów można wyznaczyć jako wartość średnią iloczynu

$$u \cdot dQ/dt:$$

$$P_s = \omega_s (m Q_1^2 / 4 C_0)$$

$$P_j = \omega_j (m Q_1^2 / 4 C_0)$$

oraz przy uwzględnieniu (4.46)

$$P_p = \omega_p \frac{m Q_1^2}{4 C_0}$$

Zasadę zachowania energii można zatem zapisać w postaci

$$\frac{P_p}{\omega_p} = \frac{P_s}{\omega_s} = \frac{P_j}{\omega_j} \quad (4.51)$$

Powyższe równanie określa bilans przekazywania mocy od źródła pompującego do obwodu sygnałowego i obwodu jałowego. W równaniach przyjęto, że fazy napięcia jałowego i napięcia pompującego są jednakowe. Takie postępowanie jest słuszne, gdyż obwód jałowy automatycznie „dostraja się” do fazy napięcia pompującego.

Najważniejszą zaletą wzmacniacza parametrycznego jest niski poziom szumów, gdyż elementy reaktancyjne są bezszumne. Jedynym źródłem szumów są składowe rezystancyjne, reprezentujące straty w rzeczywistych elementach reaktancyjnych. Szumy są szczególnie małe przy pracy wzmacniacza w zakresie niskich temperatur. Na przykład obniżenie temperatury od pokojowej do 77 K (ciekły azot) powoduje kilkukrotne zmniejszenie szumów. Wzmacniacze parametryczne pracują w zakresie mikrofalowym, między innymi są stosowane w systemach łączności sztucznych satelitów Ziemi (3,7...4,2 GHz).

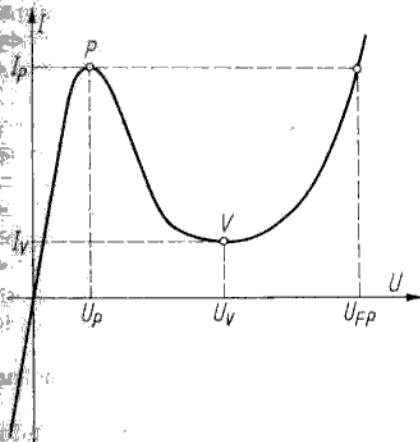
Orientacyjne parametry techniczne jednostopniowego wzmacniacza chłodzonego w ciekłym azocie są następujące:

Częstotliwość sygnałowa	4,17 GHz \pm 30 MHz
Częstotliwość napięcia pompującego	23 GHz
Wzmocnienie	20 dB
Pasmo 3-decybelowe	60 MHz
Temperatura szumów ¹⁾	< 45 K
Stabilność współczynnika wzmocnienia	
w krótkim czasie	< 0,1 dB
w długim czasie	< 0,3 dB
Moc źródła pompującego	20 mW

Diody tunelowe

4.6

Diody tunelowe są to silnie domieszkowane złącza p^+-n^+ , które wyróżniają się tym, że ich charakterystyka prądowo-napięciowa w kierunku przewodzenia ma odcinek o ujemnej rezystancji dynamicznej (rys. 4.43).



Rys. 4.43
Charakterystyka prądowo-napięciowa diody tunelowej

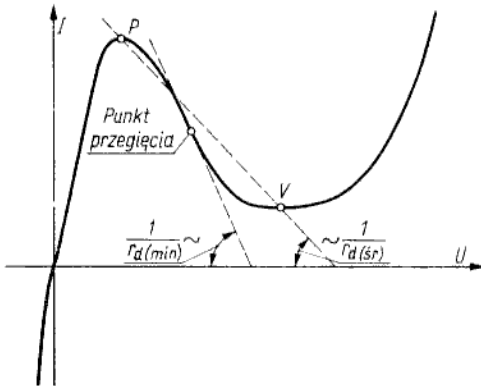
Pod względem funkcjonalnym diody tunelowe można by zaliczyć do grupy diod impulsowych lub (i) mikrofalowych, gdyż są one stosowane głównie w układach przełączających oraz we wzmacniaczach i generatorach mikrofalowych. Jednak specyficzne właściwości tych diod usprawiedliwiają wyodrębnienie grupy diod

¹⁾ Temperatura szumów jest związana ze współczynnikiem szumów F zależnością $T = 290 \text{ K} (F - 1)$.

tunelowych jako elementów o wspólnych cechach fizycznych. Interpretacja fizyczna charakterystyki prądowo-napięciowej diody tunelowej jest znana z p. 3.1, w którym rozpatrywano właściwości fizyczne złącza p^+-n^+ silnie domieszkowanych.

Do podstawowych parametrów charakterystycznych diody tunelowej należy zaliczyć:

- współrzędne punktu wierzchołka (I_P, U_P) oraz punktu doliny (I_V, U_V) charakterystyki $I(U)$; ponieważ różnica napięć ($U_V - U_P$) jest, praktycznie biorąc, stała dla diod wytwarzanych z określonego materiału półprzewodnikowego, często podaje się; współrzędne punktu wierzchołka (I_P, U_P) oraz stosunek prądu wierzchołka do prądu doliny I_P/I_V ;



Rys. 4.44

Interpretacja graficzna minimalnej i średniej ujemnej rezystancji dynamicznej

- ujemna rezystancja dynamiczna r_d (rys. 4.44); podaje się albo minimalną wartość rezystancji dynamicznej, wanikającą z nachylenia charakterystyki $I(U)$ w punkcie przegięcia, przy czym w przybliżeniu

$$r_{d(\min)} \approx -2 \frac{U_P}{I_P} \quad (4.52)$$

albo średnią wartość rezystancji dynamicznej

$$r_{d(sr)} = \frac{U_V - U_P}{I_V - I_P} \quad (4.53)$$

- pojemność warstwy zaporowej C_j ;
- rezystancja szeregową R_s ;
- graniczna częstotliwość odtłumiania f_r , przy czym parametr ten można wyznaczyć na podstawie r_d, C_j, R_s według zależności (4.55);
- inne parametry określające elementy schematu zastępczego, tj. indukcyjność szeregową L_s , pojemność rozproszona oprawki C_p . Ponadto podaje się niekiedy wartość napięcia U_{FP} odpowiadającego zrównaniu się prądu gałęzi wznoszącej z prądem I_P (rys. 4.43).

W grupie granicznych parametrów dopuszczalnych podaje się:

- maksymalny prąd dopuszczalny w kierunku przewodzenia I_F i w kierunku zaporowym I_R ; zwykle $I_F = I_R = kI_P$, przy czym k przyjmuje na ogół wartości 2...10;
- dopuszczalna temperatura złącza T_j .

Część urojona impedancji Z zawiera składową indukcyjną i pojemnościową, czyli dioda tworzy obwód rezonansowy. Częstotliwość rezonansu własnego f_x otrzymuje się z przyrównania do zera części urojonej impedancji Z

$$f_x = \frac{1}{2\pi\sqrt{L_s C_j}} \sqrt{1 - \frac{L_s}{C_j r_d^2}} \quad (4.56)$$

W katalogach podaje się częstotliwości f_{r0} , f_{x0} określone zależnościami (4.55), (4.56) przy podstawieniu minimalnej wartości rezystancji dynamicznej r_d .

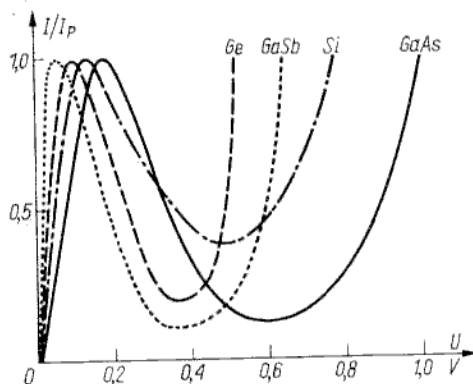
Impedancja diody tunelowej ma charakter pojemnościowy dla $f < f_{x0}$ oraz indukcyjny dla $f > f_{x0}$.

Dla stabilnej pracy diody jest konieczne, by $f_{x0} > f_{r0}$. Zwykle żąda się, by $f_{x0} > f_{r0} > 3f_p$, przy czym f_p — częstotliwość pracy diody tunelowej.

Z wyprowadzonych zależności wynika, że szybkość działania diody tunelowej jest określona przede wszystkim przez stałą czasową $\tau = C_j r_d$. W celu zmniejszenia pojemności C_j wykonuje się złącza p^+-n^+ o jak najmniejszej powierzchni, co jednak prowadzi zarazem do zmniejszenia wartości prądu szczytu I_p , a więc powoduje zwiększenie rezystancji dynamicznej r_d . Dlatego pod tym względem najbardziej miarodajnym wskaźnikiem jakości diody jest wartość ilorazu I_p/C_j . Iloraz I_p/C_j określa szybkość przełączania diody. Zwiększeniu wartości I_p/C_j sprzyjają następujące czynniki:

- zmniejszanie powierzchni złącza,
- zwiększanie koncentracji domieszek (w praktyce stosuje się koncentracje na granicy rozpuszczalności),
- dobór materiału o największej ruchliwości nośników.

Diody tunelowe mogą być wytwarzane z różnych materiałów półprzewodnikowych. Na rysunku 4.46 porównano charakterystyki prądowo-napięciowe (prąd unormo-



Rys. 4.46
Charakterystyki $I(U)$ dla diod tunelowych z GaSb, Ge, Si, GaAs

wany do wartości I_p) dla diod z GaSb, Ge, Si, GaAs. Najczęściej diody tunelowe są wytwarzane z Ge i GaAs, rzadziej z GaSb i bardzo rzadko z Si. Takie preferencje w doborze materiałów są zrozumiałe, jeżeli porówna się podstawowe parametry diod tunelowych, wykonanych z różnych materiałów, zestawione w tabl. 4.6

Diody krzemowe zdecydowanie ustępują pozostałym ze względu na małą wartość I_p/I_p oraz małą szybkość działania. W celu uzyskania dużej szybkości działania należy dobrać materiał o dużej ruchliwości nośników. Pod tym względem bardzo dobre

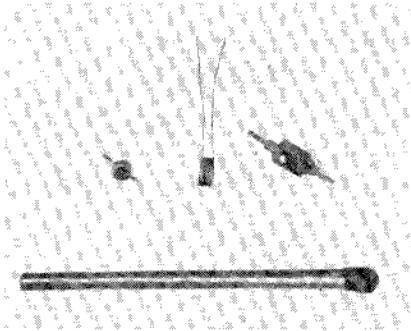
Tablica 4.6

Parametry techniczne kilku przykładowo wybranych diod tunelowych

Typ diody	Parametry charakterystyczne										Dopuszczalne parametry graniczne		U w a g i
	I_P	I_F/I_V	U_P	U_V	U_{FP}	C_J	$-r_d$	R_s	f_{∞}	$I_F = I_R$	$T_J(T_a)$		
	mA		mV	mV	mV	pF	Ω	Ω	GHz		$^{\circ}\text{C}$		
AEY 30 A	1,5	6...9	75	350		0,8 (przy U_V)	90	7	—	10 mA	(100)	Ge (Siemens)	
TD 439	3,7	8	78	400	580	0,13 (przy U_V)	40	7	65	—	—	Ge (General Electric) $I_s = 0,15 \text{ nH}$, $C_p = 0,25 \text{ pF}$	
IN 2931	10	2,5	85	460		400 (przy U_V)	—	—	—	—	—	Si (Motorola)	
IN 3118	10	10	160	600		20	—	—	—	—	—	GaAs (Motorola)	
Wartości przeciętne						$I_F/C_J(U_V)$ [mA/pF]			τ [ns]				
			70	360	500	0,1...3			2...0,07		150	Ge	
		3	80	450	750	0,3			2		300	Si	
		12	150	550	1000	1...10			0,5...0,05		400	GaAs	
		12	50	280	530	2			0,1		200	GaSb	
		10			200				0,005		20	InSb	

właściwości ma InSb. Diody tunelowe wykonane z InSb charakteryzują się ponadto najmniejszymi szumami własnymi, lecz ich wadą jest niska temperatura pracy (poniżej 20°C).

Pod względem technologii i konstrukcji diody tunelowe są to złącza epiplanarne lub stopowe, zamontowane w oprawkach typowych dla przyrządów pracujących w zakresie mikrofalowym (rys. 4.47).



Rys. 4.47
Typowe obudowy diod tunelowych

Przykłady zastosowań

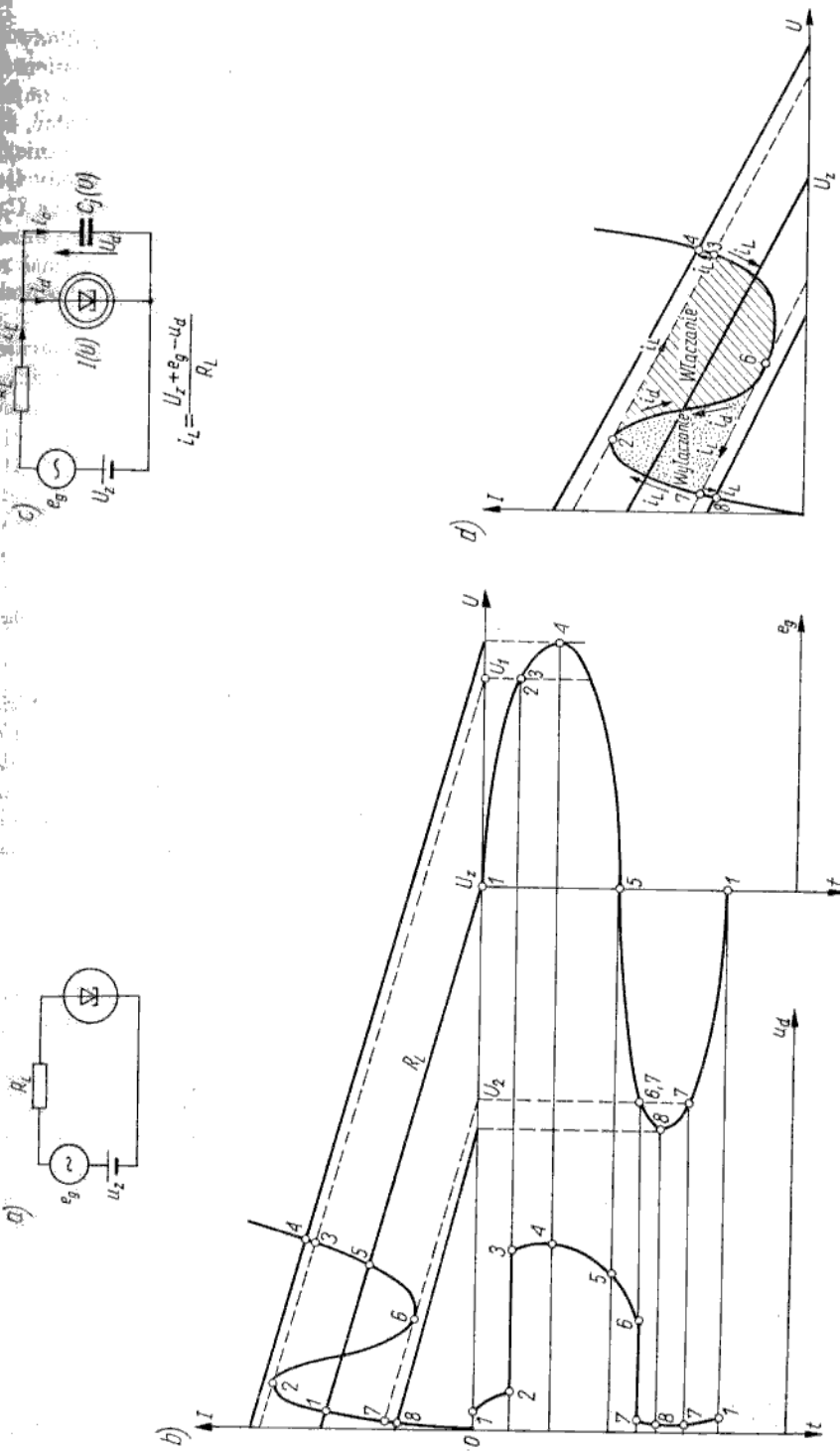
4.6.2

Teraz zostaną omówione dwa proste przykłady ilustrujące możliwości zastosowań diody tunelowej w układach impulsowych i generatorach mikrofalowych.

Układ przerzutnika z diodą tunelową i jego analizę graficzną przedstawiono na rys. 4.48. Statyczny punkt pracy diody jest określony przez źródło polaryzacji U_z i rezystor obciążenia R_L . Źródło sygnału sterującego e_g może dawać przebieg sinusoidalny albo inny przebieg napięcia przemiennego, na przykład trójkątny lub trapezoidalny. Zadaniem diody tunelowej w tym układzie jest uformowanie szybkich skoków napięcia w chwilach, gdy sygnał sterujący przekracza pewne wartości progowe.

Zasadę kształtowania impulsu o stromych zboczach wyjaśnia analiza graficzna, przedstawiona na rys. 4.48b. Wartości progowe napięcia wejściowego oznaczono jako U_1 , U_2 , a odpowiadające im proste obciążenia — narysowane linią przerywaną — przechodzą przez punkty 2, 6 na charakterystyce $I(U)$ diody tunelowej. Prosta obciążenia dla stanu spoczynkowego, czyli dla napięcia U_z , przecina charakterystykę $I(U)$ diody tunelowej w trzech punktach. Jednak tylko jeden z dwóch „zewnętrznych” punktów (1 lub 5) może być stabilnym punktem pracy. (Gdyby rezystancja obciążenia była mniejsza niż moduł ujemnej rezystancji dynamicznej diody, to prosta obciążenia przecinałaby charakterystykę $I(U)$ tylko w jednym punkcie i byłby to stabilny punkt pracy). W chwili gdy napięcie wejściowe przekracza wartość progową U_1 , następuje raptowne przeniesienie punktu pracy z 2 do 3. Natomiast przy obniżeniu napięcia wejściowego poniżej wartości U_2 następuje raptowne przeniesienie punktu pracy z 6 do 7. Z tymi dwoma przeskokami punktu pracy są związane bardzo strome zbocza impulsu napięcia u_d ukształtowanego na diodzie. Przebieg impulsu u_d w pozostałych okresach czasu można z łatwością prześledzić zgodnie z charakterystycznymi punktami oznaczonymi na rys. 4.48b cyframi od 1 do 8.

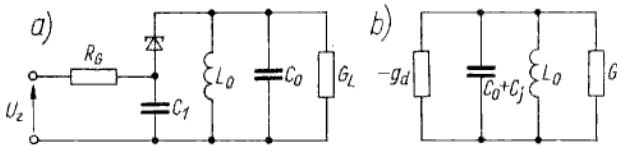
Pelniejszy obraz zjawisk zachodzących w omawianym obwodzie uzyska się, przedstawiając diodę tunelową w postaci schematu zastępczego, zawierającego nieliniowe



Rys. 4.48
 Układ przetrzutnika z diodą tunelową: a) układ włączenia diody; b), d) analiza graficzna
 c) schemat zastępczy; b), d) analiza graficzna

źródło prądowe $I(U)$ — odpowiadające charakterystyce $I(U)$ diody tunelowej, połączone równolegle z pojemnością $C_j(U)$. Układ z uwzględnieniem tego schematu zastępczego diody tunelowej przedstawiono na rys. 4.48c, a najistotniejsze aspekty jego pracy ilustruje analiza graficzna na rys. 4.48d. Widać, że w czasie przejścia punktu pracy z 2 do 3 (włączanie) lub z 6 do 7 (wyłączanie) następuje rozdzielenie prądu obciążenia i_L na dwie składowe, tzn. prąd diody i_d oraz prąd przeladowania pojemności i_c . Ponieważ przeladowanie pojemności C_j jest tym szybsze, im większy jest prąd $i_c = i_L - i_d$, zatem najkorzystniejsze warunki pracy diody tunelowej daje sterowanie prądowe (linia obciążenia niemal równoległa do osi napięcia). Układy przełączające z diodami tunelowymi umożliwiają otrzymywanie impulsów o pikosekundowych zboczach.

Zastosowanie diody tunelowej w generatorach sygnałów sinusoidalnych ilustruje przykład prostego układu generującego, przedstawiony na rys. 4.49. Źródło U_z przez



Rys. 4.49

Generator napięcia sinusoidalnego z diodą tunelową: a) układ generatora; b) schemat zastępczy

rezystor R_G polaryzuje diodę w odpowiednim punkcie pracy. Warunkiem wystąpienia oscylacji jest od tłumienia obwodu L_0 , C_0 , czyli konduktancja obciążenia powinna być mniejsza niż moduł ujemnej konduktancji dynamicznej diody tunelowej, $|G_L| < |-g_d|$. Jak wynika ze schematu zastępczego tego układu, przedstawionego na rys. 4.49b, częstotliwość generacji jest określona częstotliwością rezonansu własnego obwodu L_0 , $C_0 + C_j$ (pojemność C_1 ma bardzo dużą wartość), zatem

$$f = \frac{1}{2\pi \sqrt{L_0(C_0 + C_j)}} \quad (4.57)$$

Diody tunelowe są stosowane w generatorach i wzmacniaczach na zakres częstotliwości do kilkuset gigaherców. Jednak generatory i wzmacniacze z diodami tunelowymi nie zyskały szerokiego rozpowszechnienia ze względu na ich niektóre niedostatki w porównaniu z „konkurencyjnymi” rozwiązaniami.

Wadą generatorów z diodami tunelowymi jest ich bardzo mała moc wyjściowa, nie wystarczająca w większości przypadków do praktycznego wykorzystania tych generatorów. Zdecydowanie lepsze są pod tym względem generatory Gunna lub lawinowo-przelotowe.

Diody mikrofalowe (detekcyjne, mieszające oraz sterowane impedancje)

47

Diody mikrofalowe — jak wskazuje nazwa — są to diody przeznaczone do zastosowań w zakresie mikrofalowym, przy czym według obowiązujących norm za mikrofałe uważa się fale elektromagnetyczne o długości mniejszej niż 1 m, czyli

o częstotliwości większej niż 300 MHz. W praktyce inżynierskiej często przyjmuje się za dolną granicę mikrofal częstotliwość 1 GHz. Za górną granicę mikrofal zwykle przyjmuje się początek zakresu promieniowania podczerwonego, tj. ok. 700 GHz, co odpowiada długości fali 430 μm .

Zakres mikrofalowy można podzielić w systemie dekadowym na następujące pasma:

- pasmo decymetrowe, inaczej U.w.cz. — ultra wielkie częstotliwości ($0,1 \leq \lambda < 1 \text{ m}$, $0,3 < f \leq 3 \text{ GHz}$);
- pasmo centymetrowe, inaczej S.w.cz. — super wielkie częstotliwości ($1 \leq \lambda < 10 \text{ cm}$, $3 < f \leq 30 \text{ GHz}$);
- pasmo milimetrowe, inaczej Skr.w.cz. — skrajnie wielkie częstotliwości ($1 \leq \lambda < 10 \text{ mm}$, $30 < f \leq 300 \text{ GHz}$);
- pasmo submilimetrowe ($\lambda < 1 \text{ mm}$, $f > 300 \text{ GHz}$).

W technice radiolokacyjnej stosuje się powszechnie kod literowy do oznaczenia węższych pasm:

P	(0,225...0,390 GHz)	
L	(0,390...1,55 GHz)	
S	(1,55...5,26 GHz)	
G	(3,95...5,85 GHz)	
C	(5,85...8,2 GHz)	
X	(5,2...10,9 GHz)	← pasma pomocnicze
J	(10,9...17,25 GHz)	
K	(10,9...36 GHz)	
Q	(36...46 GHz)	
V	(46...56 GHz)	

Diody mikrofalowe to bardzo obszerne zagadnienie, któremu poświęcono wiele monografii¹⁾. Najogólniej diody mikrofalowe można podzielić pod względem ich właściwości funkcjonalnych na cztery grupy:

- zmienne rezystancje (warystory),
- zmienne reaktancje (waraktory),
- sterowane impedancje,
- ujemne rezystancje.

Warystory (ang. *VARIABLE resiSTOR*) są to elementy o nieliniowej charakterystyce prądowo-napięciowej, czyli każda dioda jest warystorem. Warystory mikrofalowe muszą odznaczać się dużą szybkością działania, dlatego stosuje się następujące rodzaje diod:

- ostrzowe;
- Schottky'ego;
- zwrotne (inaczej nazywane wstecznymi).

Warystory są stosowane głównie w układach detekcji i mieszaczach, dlatego takie diody nazywa się często detekcyjnymi i mieszającymi.

Waraktory rozpatrywaliśmy oddzielnie jako rodzaj diod pojemnościowych.

Sterowane impedancje są to diody *p-i-n* (półprzewodnik typu *p*-samoistny-typu *n*), stosowane w mikrofalowych przełącznikach, bezpiecznikach, modulatorach i dzielnikach.

¹⁾ Między innymi w języku polskim: J. Klamka: *Diody mikrofalowe półprzewodnikowe*. WNT, Warszawa 1973; H. V. Shurmer: *Mikrofalowe przyrządy półprzewodnikowe*. MON, Warszawa 1974.

Ujemne rezystancje uzyskuje się w diodach tunelowych oraz generacyjnych (diody lawinowo-przelotowe oraz diody Gunna). Diody tunelowe omówiono już, a diody generacyjne będą rozpatrywane odrębnie w rozdziale 12 ze względu na specyficzny i nieomawiany dotychczas rodzaj zjawisk fizycznych w nich występujących.

W tym punkcie zatem z dużej rodziny diod mikrofalowych będą omówione tylko następujące:

- diody detekcyjne i mieszające (ostrzowe, Schottky'ego, zwrotne),
- diody o sterowanej impedancji (diody *p-i-n*).

Diody detekcyjne i mieszające

4.7.1

Diody detekcyjne i mieszające są opisywane następującymi parametrami technicznymi.

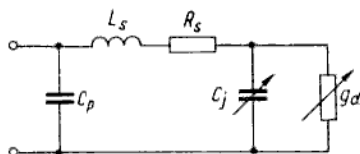
Parametry charakterystyczne:

- napięcie U_F przy określonym prądzie przewodzenia I_F ;
- prąd I_R przy określonym napięciu wstecznym U_R ;
- pasmo częstotliwości pracy;
- czułość (dla diod detekcyjnych), przy czym istnieją trzy określenia czułości: tangencjalne, prądowe i napięciowe; w katalogach najczęściej podaje się czułość tangencjalną (styczna);
- impedancja wejściowa;
- współczynnik fali stojącej;
- straty przemiany (dla diod mieszających);
- impedancja wyjściowa dla częstotliwości pośredniej (dla diod mieszających);
- współczynnik szumów.

Parametry graniczne:

- moc maksymalna,
- energia przepalania,
- temperatura dopuszczalna.

Rozpatrzmy ważniejsze parametry techniczne, biorąc pod uwagę schemat zastępczy diod detekcyjnych i mieszających, przedstawiony na rys. 4.50. W schemacie zastępczym



Rys. 4.50

Schemat zastępczy diod detekcyjnych i mieszających

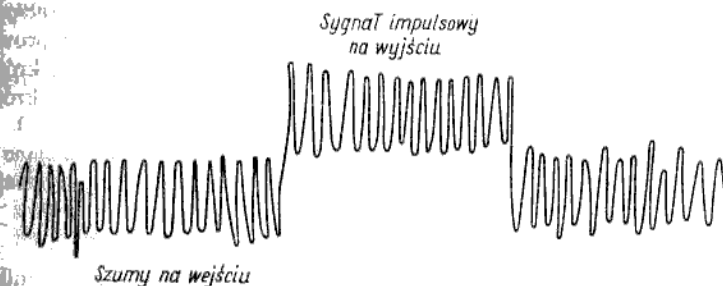
czyż nie uwzględniono pojemności dyfuzyyjnej, co wynika stąd, że „szybkie” diody detekcyjne i mieszające działają na nośnikach większościowych. Wszystkie elementy schematu zastępczego są znane z poprzednich rozważań.

Czułość diod detekcyjnych

Zdolność detektora do różnicznania małych sygnałów na tle szumów własnych jest określana przez *czułość progową*, którą najogólniej definiuje się jako moc dysponowaną sygnału wejściowego, będącą pewną ustaloną krotnością mocy dysponowanej szumów detektora sprowadzonych do wejścia

$$P_{w.cz.} = \gamma P_{sz} \quad (4.58)$$

Najczęściej stosuje się pojęcia tangencjalnej (stycznej) czułości progowej h_t , dla której $\gamma = 6$. Czułość tangencjalna jest określana subiektywnie na podstawie obrazu otrzymywanego na ekranie oscylografu dołączonego do wyjścia detektora przy podaniu na jego wejście sygnału prostokątnego (rys. 4.51). Moc sygnału wejś-



Rys. 4.51

Przebieg obserwowany na ekranie oscylografu podczas wyznaczania czułości tangencjalnej

ciowego reguluje się tak, by dolna granica szumów przy sygnale znalazła się na poziomie górnej granicy szumów bez sygnału. Moc sygnału wejściowego, odpowiadająca tym warunkom i wyrażona w decybelach w stosunku do mocy 1 mW (jednostka dBm), jest czułością tangencjalną.

Czułość prądowa jest określana jako stosunek zwarciovego prądu wyprostowanego na wyjściu detektora do mocy mikrofalowego sygnału wejściowego

$$h_i = \frac{I}{P_{w.cz.}} \quad (4.59)$$

Wartość tego parametru wynosi w typowym przypadku 1...20 A/W.

Czułość napięciowa jest określana jako stosunek napięcia na wyjściu nieobciążonego detektora do mocy mikrofalowego sygnału wejściowego

$$h_u = \frac{U}{P_{w.cz.}} \quad (4.60)$$

Przeciętne wartości tego parametru wynoszą 100...1000 V/W.

Straty przemiany (w dB) jest to parametr wyrażający stosunek mocy wejściowej sygnału mikrofalowego do mocy wyjściowej o częstotliwości pośredniej w warunkach dopasowania na wejściu i wyjściu

$$L = 10 \lg \frac{P_{we(w.cz.)}}{P_{wy(p.cz.)}} \quad (4.61)$$

Na podstawie schematu zastępczego straty przemiany, spowodowane elementami R_s , C_j , można wyrazić w postaci

$$L = 10 \lg \left(1 + \frac{R_s}{r_d} + \omega^2 C_j^2 R_s r_d \right) \quad (4.62)$$

Współczynnik szumów własnych diody jest definiowany jako stosunek mocy dysponowanej szumów, których źródłem jest dioda, do mocy dysponowanej szumów cieplnych

rezystora równoważnego rezystancji różniczkowej diody, który ma tę samą temperaturę co dioda

$$F = \frac{P_{sz}(\text{diody})}{P_{sz}(\text{równoważnego rezystora})} \quad (4.63)$$

Bardzo ważnym parametrem jest *energia dopuszczalna*, nazywana inaczej *energiją przepalania*. Jest to największa energia elektryczna, która może być doprowadzona do diody w przypadku, gdy czas trwania impulsu elektrycznego jest krótszy niż czas ustalania się równowagi cieplnej diody (zwykle chodzi o czasy krótsze niż 10^{-8} s).

Stwierdziliśmy już wcześniej, że jako mikrofalowe diody detekcyjne i mieszające stosuje się diody ostrzowe, Schottky'ego i zwrotne. Porównajmy te trzy rodzaje diod.

Diody ostrzowe

4.7.1.1

Dioda ostrzowa stanowi unikalny przykład elementu półprzewodnikowego, który jest znany od 100 lat¹⁾ i w dalszym ciągu powszechnie stosowany, a mimo to nie ma zadowalająco ugruntowanych podstaw teoretycznych.

Zwróciliśmy na to uwagę już we wstępie do niniejszego rozdziału, gdzie stwierdziliśmy, że dioda ostrzowa może mieć strukturę fizyczną złącza *p-n* albo strukturę złącza *m-s* z barierą Schottky'ego, przy czym jest możliwe powstanie złącza *p-n* warstwy inwersyjnej z podłożem. Niekiedy trudno jest jednoznacznie określić strukturę fizyczną złącza. W przypadku diod germanowych formowanych elektrycznie mamy do czynienia ze złączem *p-n*. Te diody mogą pracować co najwyżej w dolnym pasmie częstotliwości mikrofalowych. Przeważającą większość mikrofalowych diod ostrzowych stanowią diody krzemowe (styk igły wolframowej z krzemem typu *p*), a ostatnio coraz większe znaczenie, szczególnie w zakresie milimetrym, mają diody ostrzowe arsenkowo-galowe (styk igły fosforobrazowej z arsenkiem galu typu *n*).

Diody ostrzowe z uwagi na małą powierzchnię styku metal-półprzewodnik (kilka do kilkadziesiąt μm^2) odznaczają się bardzo małą pojemnością złącza, a więc mają bardzo dużą częstotliwość pracy (do 2000 GHz).

Szczególnie duże możliwości pracy w zakresie bardzo wielkich częstotliwości mają diody arsenkowo-galowe z uwagi na dużą ruchliwość nośników, czyli małą rezystancję szeregową.

Diody Schottky'ego

4.7.1.2

Zasada działania diody Schottky'ego, tj. styku płaskiej elektrody metalowej z półprzewodnikiem, jest znana z poprzedniego rozdziału. Wiadomo, że diody Schottky'ego działają na nośnikach większościowych, odznaczają się zatem bardzo krótkimi czasami przełączania. Z uwagi na tę zaletę były one wyróżniane w grupie diod impulsowych. Teraz należy zauważyć, że mała „bezwładność” działania diod Schottky'ego predysponuje je również do zastosowań w zakresie mikrofalowym jako diody detekcyjne i mieszające. Diody Schottky'ego są powszechnie stosowane w zakresie częstotliwości do 100 GHz (powyżej tej częstotliwości w dalszym ciągu niezastąpione są diody ostrzowe).

Z porównania charakterystyk statycznych diod ostrzowych, Schottky'ego i zwrotnych wynika, że diody Schottky'ego mają najlepsze właściwości w zakresie wstecznym.

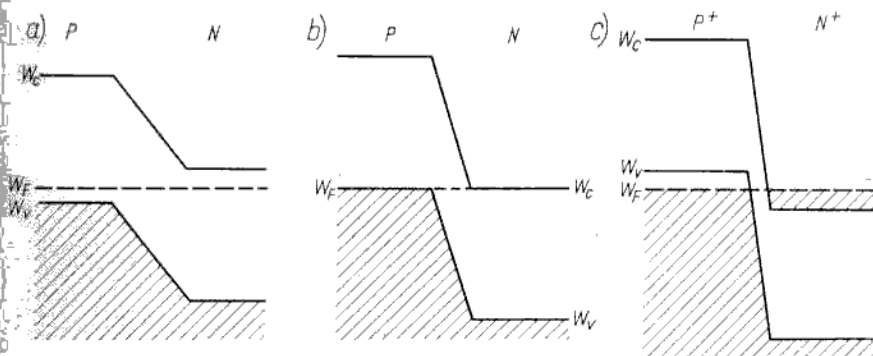
¹⁾ Pierwsze badania właściwości elektrycznych styku igły metalowej z kryształem półprzewodnika wykonał Braun w 1874 r.

lecz ustępują dwu pozostałym w zakresie przewodzenia, gdyż mają największe napięcie progowe. Dlatego odpowiednio dużą czułość prądową detekcji otrzymuje się przy wstępnym spolaryzowaniu diody Schottky'ego w kierunku przewodzenia. Do podstawowych zalet tych diod należy zaliczyć małą wartość rezystancji szeregowej, mały poziom szumów własnych oraz większą stabilność pracy i odporność na wstrząsy lub udary niż dla diod ostrzowych.

Diody zwrotne

4.7.1.3

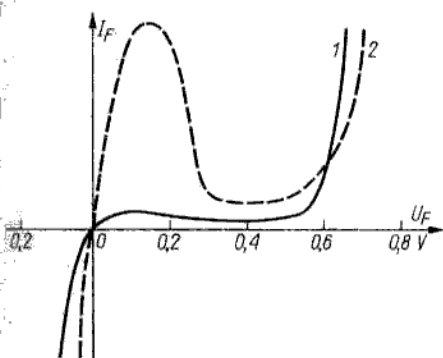
Dioda zwrotna, nazywana również wsteczną (ang. *backward diode*), jest jakby nieudaną diodą tunelową o znikomo małym prądzie szczytu. Jest to po prostu złącze $p-n$ nieco słabiej domieszkowane niż w diodzie tunelowej. Koncentracje domieszek w obszarach P , N są tak dobrane, by poziom Fermiego był ułożony na krawędzi pasma zabronionego. Na rysunku 4.52 porównano model pasmowy



Rys. 4.52

Energetyczny model pasmowy złącza $p-n$ słabo domieszkowanego (a), diody zwrotnej (b), diody tunelowej (c)

złącza $p-n$ słabo domieszkowanego, diody zwrotnej i diody tunelowej. Przy polaryzacji diody zwrotnej w kierunku przewodzenia nie ma tunelowego prądu Esaki'ego, który płynie w diodzie tunelowej. Natomiast przy polaryzacji w kierunku wstecznym — podobnie jak w diodzie tunelowej — raptownie wzrasta tunelowy prąd Zenera. Charakterystyki prądowo-napięciowe diody zwrotnej i diody tunelowej porównano na rys. 4.53.



Rys. 4.53

Porównanie charakterystyki prądowo-napięciowej diody zwrotnej 1 i diody tunelowej 2

Tablica 4.7

Parametry techniczne przykładowo wybranych diod detekcyjnych i mieszających

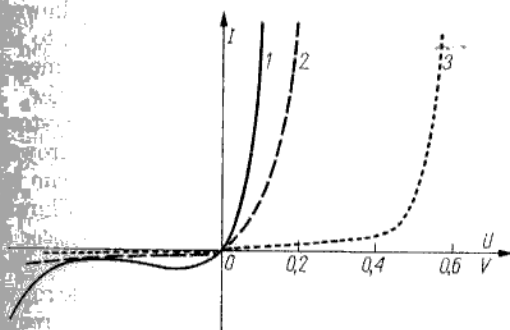
Typ diody	Parametry charakterystyczne										Dopuszczalne parametry graniczne			U w a g i
	przy U_F		przy I_R		f	h_t	L	Z_{we}	$Z_{wz}(p-cz)$	F	P_a	W_s	T_j	
	V	mA	I_F	μA										
AEY 29	0,3	12	100	0,3	J	-43	300						100	Wsteczna germanowa, detekcyjna
AAV 59	0,5	2	2	0,5	Q			5,5	1000	10	0,5	0,03	100	Mieszająca germanowa
BAV 46					do 18 GHz	-52	850			10	1	0,2	150	Detekcyjna krzemowa
BAW 95 D									350	8,2	1	0,2	—	Mieszająca krzemowa

Ponieważ dioda zwrotna bardzo dobrze przewodzi prąd w kierunku wstecznym, a w pewnym zakresie polaryzacji w kierunku przewodzenia odznacza się znikomo małym prądem, istnieje możliwość zastosowania tej diody do detekcji i mieszania sygnałów przy odwróceniu roli obu kierunków polaryzacji. Ten fakt wyjaśnia pochodzenie nazwy tego elementu — dioda zwrotna lub wsteczna.

Podstawowe zalety diody wstecznej są następujące:

- małe napięcie progowe (praktycznie biorąc, prąd wzrasta gwałtownie już od napięcia zerowego), a więc duża czułość prądowa;
- duża szybkość działania, co jest typowe dla elementów działających na nośnikach większościowych;
- duża odporność na wpływ temperatury i promieniowania;
- mały poziom szumów własnych.

Większość diod wstecznych, dostępnych obecnie na rynku, są to diody stopowe germanowe o maksymalnej częstotliwości pracy do 200 GHz.



Rys. 4.54
Porównanie typowych charakterystyk prądowo-napięciowych dla: diody zwrotnej 1, diody ostrzowej 2 i diody Schottky'ego 3

Na rysunku 4.54 porównano charakterystyki prądowo-napięciowe diod ostrzowych, Schottky'ego i zwrotnych.

Wszystkie omawiane diody detekcyjne i mieszające są wytwarzane w oprawkach typowych dla elementów mikrofalowych.

Parametry techniczne wybranych typów diod detekcyjnych i mieszających zestawiono w tabl. 4.7.

Diody o zmiennej impedancji (diody *p-i-n*)

4.7.2

Diody *p-i-n* spełniają funkcję elementów o zmiennej impedancji dla sygnałów b.w.cz., sterowanej napięciem stałym lub napięciem małej częstotliwości. Słabo domieszkowana warstwa wewnętrzna *i* (w idealnym przypadku warstwa półprzewodnika samoistnego) już przy niewielkim napięciu wstecznym jest całkowicie zajęta przez obszar ładunku przestrzennego i opróżniona z elektronów i dziur. Przy dalszym wzroście napięcia wstecznego warstwa zaporowa, praktycznie biorąc, nie rozszerza się, gdyż „napotyka” warstwy silnie domieszkowane *P*, *N*. Zatem w kierunku wstecznym dioda *p-i-n* odznacza się bardzo dużą rezystancją i bardzo małą stałą pojemnością określoną szerokością warstwy *i* (w typowym przypadku kilkadziesiąt mikrometrów). Podczas polaryzacji w kierunku przewodzenia do warstwy *i* są wstrzykiwane dziury i elektrony z sąsiadujących warstw *P*, *N*. Zatem warstwa *i* jest niejako „zalewana” przez elektrony i dziury, a więc dioda przechodzi w stan małej impedancji.

TRANZYSTOR BIPOLARNY

Nazwa tranzystor wywodzi się z języka angielskiego: *TRANS*fer *resISTOR* — co po polsku oznacza element transformujący rezystancję. Tak nazwano pierwszy półprzewodnikowy element wzmacniający, będący funkcjonalnym odpowiednikiem lampy próżniowej (triody). Odkrycie tzw. triody półprzewodnikowej nastąpiło w 1948 r. w Stanach Zjednoczonych, a autorami tego epokowego wynalazku są Bardeen i Brattain¹⁾. Opracowany przez nich element był to tranzystor ostrzowy, który wprawdzie nie miał istotnego znaczenia praktycznego, ale dał początek opracowaniu wielu doskonalszych konstrukcji tranzystorów.

Tranzystory warstwowe, które pojawiły się w kilka lat po odkryciu tranzystora ostrzowego, już w pierwszej połowie lat pięćdziesiątych skutecznie konkurowały z lampami elektronowymi w wielu urządzeniach elektronicznych. Zarówno tranzystor ostrzowy, mający tylko znaczenie historyczne, jak i ciągle modyfikowany tranzystor warstwowy są to tzw. tranzystory bipolarne. Nazwa *tranzystor bipolarny* obejmuje tranzystory o strukturze *p-n-p* lub *n-p-n*, w których działaniu istotną rolę odgrywają jednocześnie oba rodzaje nośników (dziury i elektrony). Zatem nazwa tranzystor, oznaczająca pierwotnie „element transformujący rezystancję”, odnosiła się do elementów znanych obecnie jako tranzystory bipolarne. Jest bardzo trafna jest ta nazwa, będzie się można przekonać nieco dalej przy omówieniu fizycznych podstaw działania tranzystora bipolarnego. W cztery lata po odkryciu tranzystora bipolarnego Shockley opisał triodę półprzewodnikową działającą na zupełnie odmiennym zasadzie²⁾. Ten element nazwano *tranzystorem unipolarnym*, gdyż w jego działaniu istotny udział ma tylko jeden rodzaj nośników (dziury lub elektrony). Tranzystor unipolarny, nazywany również *polowym* lub skrótowo *FET* (ang. *Field-Effect-Transistor*) nie jest „elementem transformującym rezystancję” w takim sensie jak tranzystor bipolarny. Zatem nazwa tranzystor obejmuje obecnie grupę elementów półprzewodnikowych szerszą, niż to wynika ze znaczenia nadanego jej przed laty. Co to jest ostatnie nie tranzystor?

Tranzystor jest to co najmniej (i przeważnie) trójkońcówkowy element półprzewodnikowy zdolny do wzmacniania sygnałów prądu stałego i zmiennego. Każdy tranzystor jest zatem wzmacniaczem. A co to jest wzmacniacz? To — pozornie tylko formalne — pytanie warte jest jednoznacznej odpowiedzi. Często rozumiemy je w ten sposób: „Wzmacniacz — jak wskazuje nazwa — jest przyrządem służącym do wzmacniania, czyli do powiększania czegoś”. Przykładowo lupa

¹⁾ Historyczna publikacja tych dwu autorów ukazała się w 1948 r. (J. Bardeen, W. Brattain: *Phys. Rev.*, 1948, 74, 230). Duży udział w wynalazku miał W. Shockley, który w tym samym roku przedstawił teorię działania tranzystora (W. Shockley: *BSTJ*, 1948, 28, No 3, 435). Dlatego za odkrycie tranzystora przyznano w 1956 r. nagrodę Nobla dwójce: W. Shockley, J. Bardeen, W. H. Brattain.

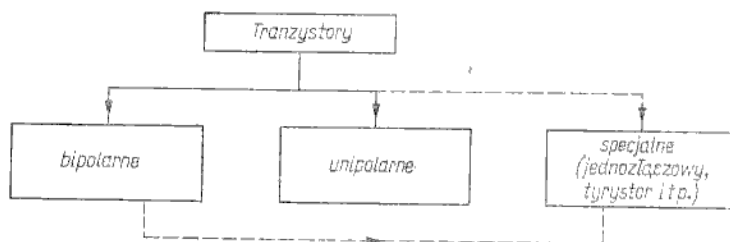
²⁾ W. Shockley: *Proc. IRE*, 1952, 40, 1365.

lornetka teatralna powiększając obraz, dźwignia mechaniczna zwiększa siłę, transformator zwiększa napięcie (lub prąd). Czy te przyrządy są wzmacniaczami? Nie. Nie są to wzmacniacze, gdyż: „Wzmacniacz jest przyrządem umożliwiającym sterowanie większej mocy mniejszą”. Aby nastąpił efekt wzmocnienia, są konieczne dwie rzeczy: źródło energii i przyrząd do sterowania przepływu tej energii — wzmacniacz. Jeżeli na przykład strumień wody z węża ogrodniczego skierujemy na łopatki turbiny, która obracając się będzie wykonywała jakąś pracę, to przekręcając kran wodny (jest to czynność nie wymagająca dużych energii) można powodować znaczne zmiany energii obracającej się turbiny. W tym przykładzie kran jest przyrządem służącym do sterowania dużej mocy za pomocą małej, czyli jest wzmacniaczem. Wzmacniaczem w ogólnym sensie jest również wyłącznik sieci oświetleniowej, wyłącznik radiowy itp. Transformator natomiast nawet 1000-krotnie podwyższający napięcie lub prąd nie jest wzmacniaczem, gdyż moc wydzielana na jego wyjściu może być co najwyżej równa mocy dostarczanej na wejście.

Tranzystor jest wzmacniaczem stosowanym zarówno do liniowego (wprost proporcjonalnego) zwiększania mocy sygnału jak również do nieliniowego, przy czym często dyskretnego (skokowego, kluczującego) sterowania mocy. Stąd generalnie można wymienić dwa rodzaje zastosowań tranzystorów:

- liniowe,
- nieliniowe (analogowe nieliniowe oraz impulsowe — głównie cyfrowe).

Oprócz tranzystorów bipolarnych i unipolarnych, które mają bardzo liczne zastosowania, rozróżnia się wiele typów tranzystorów, np. jednozłączowe, tyrystory itp., przeznaczonych tylko do pewnych zastosowań specjalnych. Formalnie biorąc, są to tranzystory bipolarne, jednak sklasyfikujemy je w odrębnej grupie jako tranzystory specjalne.



Rys. 5.1
Ogólny podział
tranzystorów

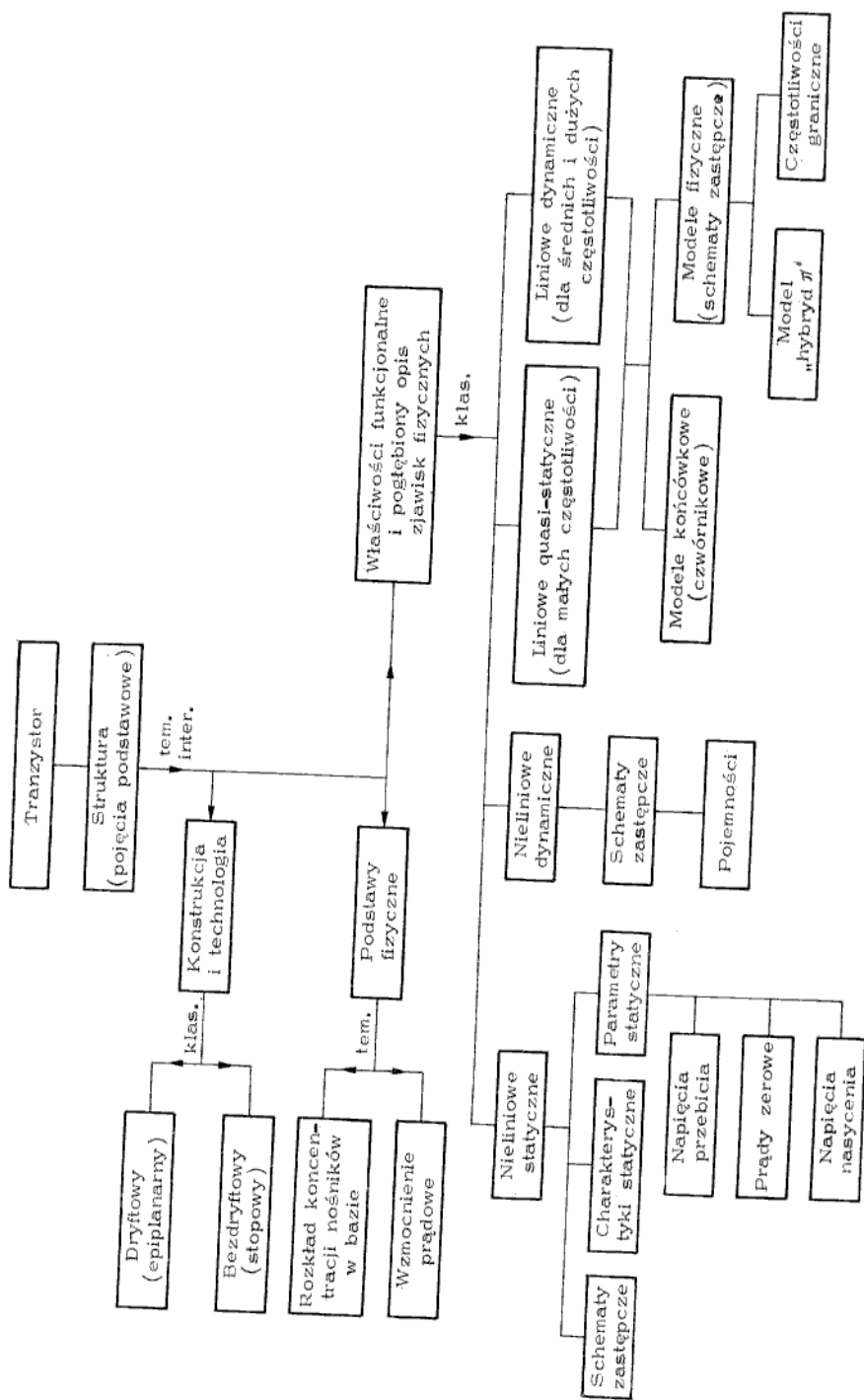
Stąd wynika najbardziej ogólny podział tranzystorów przedstawiony na rys. 5.1. Niektóre tranzystory specjalne będą omawiane w rozdziale II, tranzystory unipolarne są rozpatrywane w następnym rozdziale, w kolejnych natomiast punktach tego rozdziału rozpatrzemy konstrukcję, zasadę działania, charakterystyki, parametry i schematy zastępcze tranzystora bipolarnego. Układ ważniejszych zagadnień rozważanych w tym rozdziale przedstawiono na rys. 5.2.

Podstawowa struktura tranzystora bipolarnego

5.1

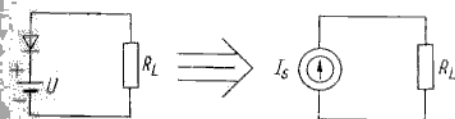
Spróbujmy wymyślić tranzystor. Na ogół odkrywanie rzeczy znanych jest zajęciem mało pożytecznym, ale w tym przypadku może być uzasadnione ze względów dydaktycznych.

Tranzystor powinien mieć wejście i wyjście. Do wejścia doprowadzimy sygnał sterujący, wyjście zaś włączymy w obwód sterowany. Najpierw rozpatrzmy właściwości

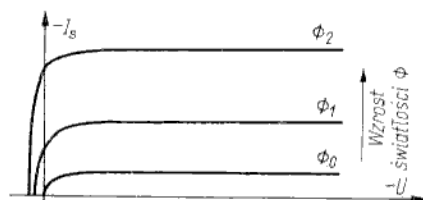


Rys. 5.2 Schematyczny układ rozdzielny

obwodu sterowanego (wyjściowego). Z uwagi na dążenie do uzyskania jak największego wzmocnienia napięciowego obwód wyjściowy powinien mieć właściwości źródła prądowego. Źródło prądowe można obciążyć bardzo dużą rezystancją; wówczas niewielkie przyrosty prądu wywołują bardzo duże przyrosty napięcia na rezystancji obciążenia. Z rozdziału 3 wiadomo, że złącze *p-n*, spolaryzowane w kierunku zaporowym, ma właściwości źródła o stałej wydajności prądowej, gdyż w szerokim zakresie zmian napięcia polaryzacji lub rezystancji obciążenia prąd w obwodzie jest wielkością stałą, określoną prądem nasycenia złącza *p-n* (rys. 5.3). Jeżeli potrafimy sterować prąd I_s , to uzyskamy wzmacniacz o bardzo



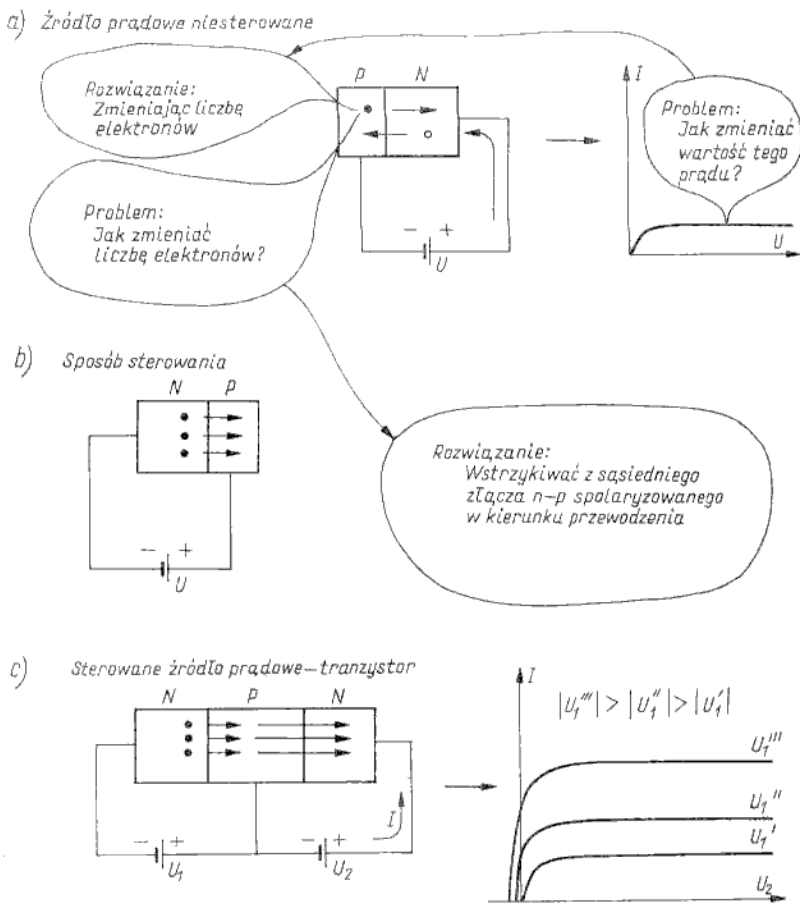
Rys. 5.3
Złącze *p-n* spolaryzowane w kierunku zaporowym jako źródło o stałej wydajności prądowej; R_L — rezystancja obciążenia; I_s — wsteczny prąd nasycenia złącza *p-n*



Rys. 5.4
Zmiany charakterystyki $I_s(U)$ spowodowane zmianami koncentracji nośników mniejszościowych wskutek oświetlenia złącza

dużym wzmocnieniu napięciowym. (Chcielibyśmy, aby wzmacniacz działał tylko w kierunku od wejścia do wyjścia. Prąd I_s powinien więc zależeć tylko od sygnału wejściowego. Zauważmy, że brak zależności prądu I_s od napięcia wyjściowego jest cechą korzystną nie tylko ze względu na możliwość uzyskania dużego wzmacniacza napięciowego, lecz również dlatego, że zostaje wyeliminowane niepożądane oddziaływanie sygnału wyjściowego na wejściowy). Z rozdziału 3 wiadomo, że prąd I_s jest prądem nośników mniejszościowych, zależnym wprost proporcjonalnie od koncentracji tych nośników. Z kolei koncentrację nośników mniejszościowych można zmieniać różnymi sposobami. Między innymi oświetlenie złącza lub wzrost temperatury powodują zwiększenie koncentracji nośników mniejszościowych — przeto zwiększenie prądu I_s (rys. 5.4). Dodatkową liczbę nośników mniejszościowych można również wprowadzić do bazy złącza *p-n* z obwodu zewnętrznego, jeżeli kontakt bazy z obwodem zewnętrznym będzie miał właściwości wstrzykiwania. Takie właściwości ma złącze *p-n* spolaryzowane w kierunku przewodzenia. Zatem kontakt metalu z bazą złącza należy zamienić na złącze *n-p*. W ten sposób powstaje struktura dwuzłączowa *n-p-n*. Polaryzacja złącza pierwszego w kierunku przewodzenia powoduje wstrzykiwanie elektronów z obszaru *N* do obszaru *P* będącego wspólną bazą obu złączy. Elektrony dostarczane do obszaru *P* — jako nośniki mniejszościowe — biorą udział w prądzie I_s drugiego złącza spolaryzowanego w kierunku zaporowym (oczywiście pod warunkiem, że baza będzie cienka). W ten sposób obwód wyjściowy ma właściwości sterowanego źródła prądowego, gdyż wszelkie zmiany prądu płynącego przez pierwsze złącze w kierunku przewodzenia powodują proporcjonalne zmiany prądu I_s drugiego złącza. Kolejne fazy rozumowania prowadzącego do utworzenia tranzystora, czyli struktury mającej pożądaną właściwości wzmacniacza sygnałów elektrycznych, przedstawiono na rys. 5.5.

Analogicznie rozumując można by zaproponować również tranzystor w postaci struktury *p-n-p*, w którym pierwsze złącze spolaryzowane w kierunku przewodzenia wstrzy-

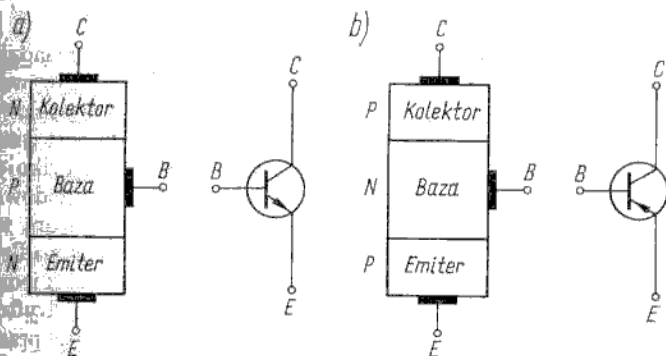


Rys. 5.5
Szkic kolejnych faz rozumowania prowadzącego do utworzenia tranzystora

kuje dziury do obszaru *N*, skąd są one odbierane przez drugie złącze spolaryzowane w kierunku zaporowym. Oczywiście źródła polaryzujące oba złącza miałyby odwrotną biegunowość niż w przypadku struktury *n-p-n*. Trzy kolejne warstwy tranzystora *n-p-n* lub *p-n-p* są nazywane zgodnie z ich funkcjami:

- *emiter* (pierwsza warstwa, która dostarcza nośników mniejszościowych do drugiej warstwy);
- *baza* (druga warstwa);
- *kolector* (warstwa zbierająca nośniki wstrzykiwane z emitera do bazy).

Na rysunku 5.6 przedstawiono obie struktury tranzystora bipolarnego wraz z symbolami tych tranzystorów. Symbole są kreślone z zachowaniem tej zasady, że strzałka oznaczająca emiter ma zwrot zgodny z kierunkiem przepływu prądu według konwencji przyjętej w elektrotechnice (od plusa do minusa). Jest to zwrot zgodny z kierunkiem przepływu ładunków dodatnich (dziur), a przeciwny do kierunku przepływu ładunków ujemnych (elektronów).



Rys. 5.6

Dwie struktury tranzystora bipolarnego i ich symbole elektryczne: a) tranzystor *n-p-n*; b) tranzystor *p-n-p*

Konstrukcja i technologia tranzystora bipolarnego

5.2

W ciągu bez mała 30 lat rozwoju technologii półprzewodnikowej pojawiło się kilkadziesiąt różnych rozwiązań konstrukcyjno-technologicznych tranzystorów bipolarnych. A oto w porządku chronologicznym najważniejsze z nich:

- tranzystor ostrzowy,
- tranzystor wyciągany,
- tranzystor stopowy,
- tranzystor mesa (stopowo-dyfuzyjny),
- tranzystor planarny,
- tranzystor epitaksjalno-planarny (epiplanarny).

Współcześnie wytwarzane tranzystory mają przeważnie strukturę epiplanarną, dlatego omówimy przede wszystkim ten rodzaj konstrukcji. Inne konstrukcje tranzystora mają w zasadzie znaczenie historyczne, a ich opis można znaleźć w wielu wcześniej wydanych podręcznikach. Z dwu powodów warto jednak zwrócić uwagę na tranzystor stopowy.

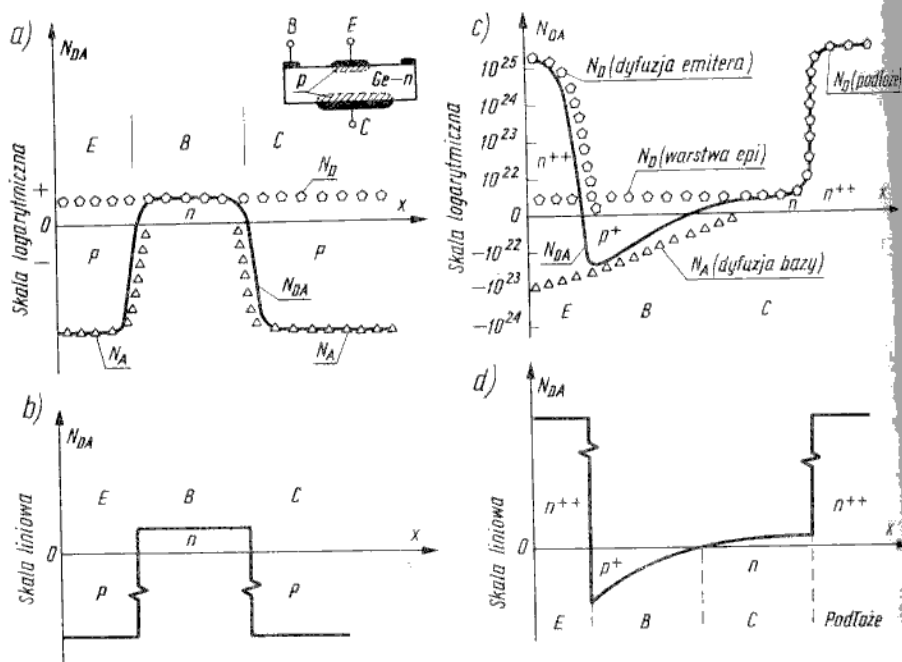
Po pierwsze, ten rodzaj tranzystorów jest jeszcze wytwarzany w niewielkich ilościach z uwagi na jego przydatność w pewnych rodzajach układów (zalety tego tranzystora to: duże wartości napięcia przebicia złącza baza-emiter oraz możliwość realizacji układu klucza symetrycznego).

Po drugie, cała teoria działania tranzystora bipolarnego, opracowana w latach pięćdziesiątych, dotyczyła tranzystora stopowego, który charakteryzuje się równomiernym rozkładem koncentracji domieszek w bazie. Tranzystory planarne i epiplanarne (wytwarzane metodą dwukrotnej dyfuzji) mają nierównomierny rozkład koncentracji domieszek w bazie, co istotnie wpływa na działanie i właściwości tych tranzystorów. W związku z tym przyjęto podział tranzystorów bipolarnych na:

- tranzystory z *jednorodną bazą* (nazywane niekiedy *bezdryftowymi* lub *dyfuzyjnymi*);
- tranzystory z *niejednorodną bazą* (nazywane również *dryftowymi*).

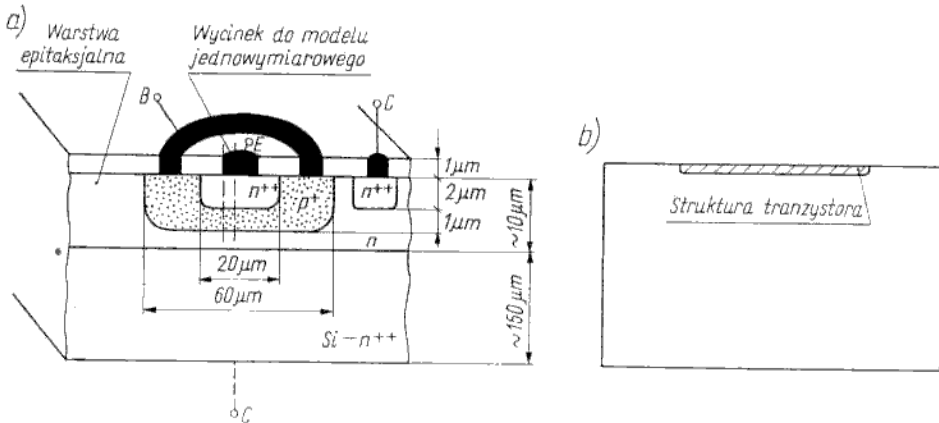
Ponieważ w okresie, gdy pojawiły się pierwsze tranzystory z niejednorodną bazą, była już ugruntowana teoria tranzystora z jednorodną bazą, właściwości tych pierwszych opisywano modyfikując istniejącą teorię. Taka chronologia zdarzeń zaciążyła na układzie treści wielu podręczników, w których najpierw szczegółowo i wyczerpująco jest rozważany tranzystor z jednorodną bazą, po czym mniej wyczerpu-

jąco rozpatruje się tranzystor z niejednorodną bazą — niejako na zasadzie szczególnego przypadku wymagającego niewielkich modyfikacji wyników wcześniejszych rozważań. Zachowanie takiego układu treści nie wydaje się celowe, gdyż współcześnie podstawowym rodzajem tranzystora bipolarnego jest tranzystor z niejednorodną bazą. Można by nawet ograniczyć rozważania tylko do tego rodzaju tranzystorów, co jednakże nie byłoby uzasadnione ze względów dydaktycznych. Spróbujmy to bliżej uzasadnić. Dla działania tranzystora najistotniejsze jest zjawisko transportu nośników w bazie. W ogólnym przypadku transport może być dyfuzyjny i unoszeniowy (dryftowy). W tranzystorze z jednorodną bazą istnieje tylko transport dyfuzyjny (stąd nazwa tranzystor bezdryftowy lub dyfuzyjny, chociaż ta ostatnia nazwa nie jest zalecana, gdyż może nieprawidłowo kojarzyć się z tranzystorem wytwarzanym metodą dyfuzji domieszek). W tranzystorze z niejednorodną bazą zdecydowanie przeważa transport unoszeniowy, ale w niewielkim stopniu istnieje również dyfuzja nośników. Stąd wynika, że najbardziej racjonalne byłoby rozpatrzenie uogólnionej teorii tranzystora z niejednorodną bazą (z uwzględnieniem zarówno unoszenia jak i dyfuzji), przy czym tranzystor z jednorodną bazą należałoby potraktować jako przypadek szczególny. Taki układ mają niektóre specjalistyczne monografie, co jednak wymaga stosowania dość złożonego aparatu matematycznego. Dążąc do uproszczenia rozważań teoretycznych tranzystor z niejednorodną bazą potraktujemy jako dryftowy w dosłownym sensie, tj. pominiemy wpływ dyfuzji nośników. Aby jednak można było jakościowo wyjaśnić pewne zjawiska drugiego rzędu, spowodowane dyfuzją zostanie również rozpatrzony tranzystor z jednorodną bazą.



Rys. 5.7
Rozkłady koncentracji domieszek w tranzystorze stopowym (a, b) i epiplanarnym (c, d) w skali logarymicznej (a, c) i liniowej (b, d)

Na rysunku 5.7 przedstawiono rozkłady koncentracji domieszek w tranzystorze stopowym (bezdryftowym) i tranzystorze epiplanarnym (dryftowym). To, że jeden tranzystor jest typu $p-n-p$, a drugi $n-p-n$, nie ma istotnego znaczenia. Obie technologie umożliwiają wytwarzanie tranzystorów zarówno $p-n-p$ jak i $n-p-n$, lecz najczęściej tranzystor germanowy stopowy jest typu $p-n-p$, a krzemowy planarny — typu $n-p-n$. Na rysunku 5.7a, c przedstawiono rozkłady koncentracji N_A , N_D oraz $N_{DA} = (N_D - N_A)$. Dla lepszej przejrzystości na rys. 5.7b, d pokazano w skali liniowej nieco wyidealizowane rozkłady $N_{DA}(x)$ w obu tranzystorach. Najistotniejszy jest różny rozkład koncentracji domieszek w obszarze bazy obu tranzystorów, tj. stała wartość N_{DA} w bazie tranzystora stopowego i malejąca zależność $N_{DA}(x)$ w bazie tranzystora epiplanarnego (maleje wartość bezwzględna $|N_{DA}|$).

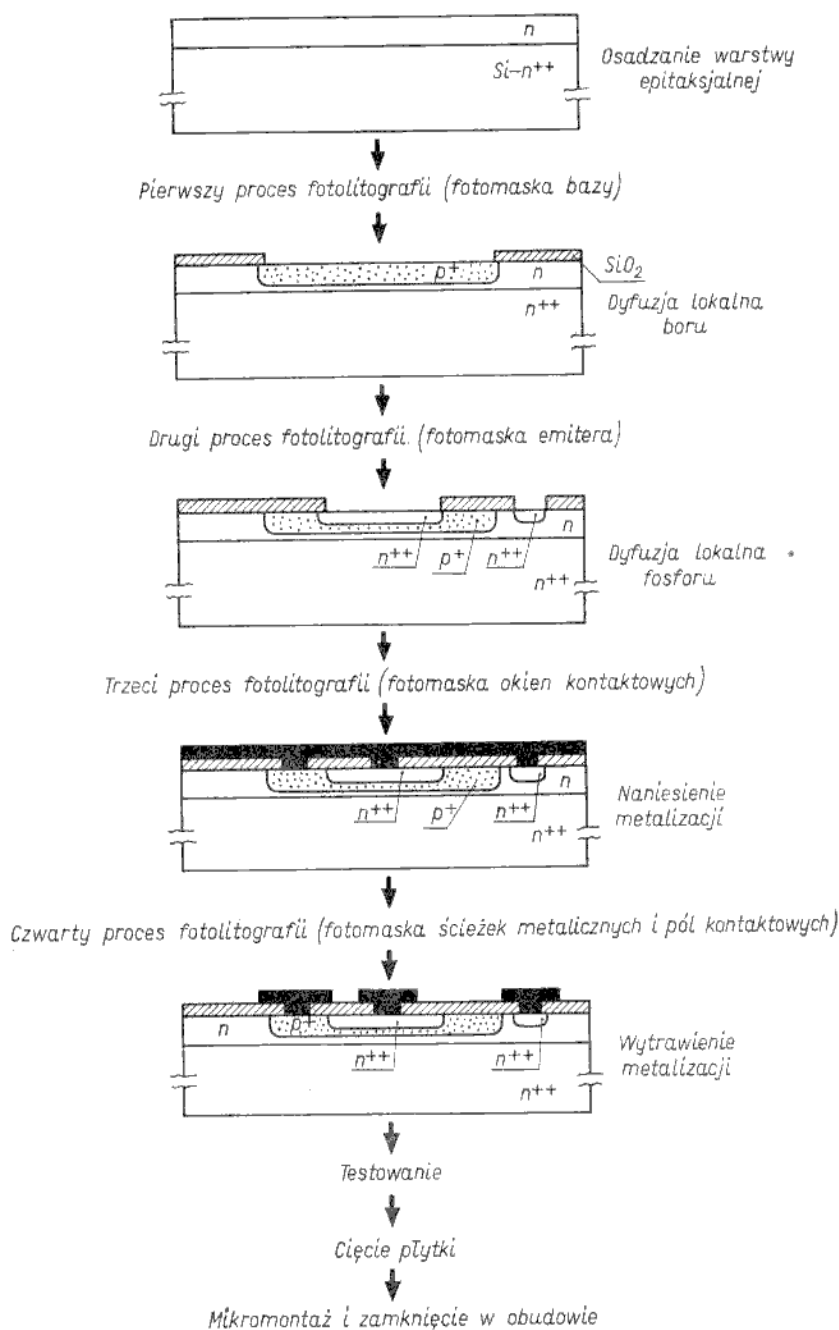


Rys. 5.8

Schematyczny przekrój struktury tranzystora epiplanarnego: a) rysunek szczegółowy bez zachowania skali; b) rysunek w skali

Strukturę tranzystora epiplanarnego przedstawiono na rys. 5.8a, b, kolejne zaś fazy jego wytwarzania — na rys. 5.9. Jest to tranzystor typu $n-p-n$, choć ściśle biorąc, ma on bardziej złożoną strukturę $n^+-p^+-n-n^+$. Podłoże typu n^+ jest to bardzo silnie domieszkowana płytka krzemu o grubości ok. $150 \mu\text{m}$. Podłoże spełnia funkcję nośnika mechanicznego dla właściwej struktury tranzystora, a pod względem właściwości elektrycznych wymaga się, by materiał podłoża miał jak najmniejszą rezystywność (chodzi o to, by rezystancja szeregową kolektora była jak najmniejsza).

Oczywiście w silnie domieszkowanym półprzewodniku typu n^+ nie można wytworzyć tranzystora, gdyż nie ma możliwości przekompensowania bardzo dużej koncentracji domieszek, tj. nie można uzyskać warstwy typu p spełniającej funkcję bazy. Dlatego na powierzchni podłoża osadza się słabo domieszkowaną warstwę epitaksjalną typu n , w której jest wykonywana właściwa struktura tranzystora. Warstwa ta spełnia funkcję obszaru kolektora, przy czym końcówka kolektora może być wyprowadzona „od spodu” lub „na wierzchu” płytki. W procesie dwukrotnej dyfuzji lokalnej wytwarza się najpierw warstwę typu p^+ , stanowiącą obszar bazy, a następnie — warstwę typu n^+ , stanowiącą obszar emitera. Jeżeli końcówka kolektora jest wyprowadzona „na wierzchu” płytki, to jednocześnie z emiterem wytwarza się również warstwę n^+ w obszarze kolek-



Rys. 5.9
Ważniejsze fazy procesu wytwarzania tranzystora epiplanarnego

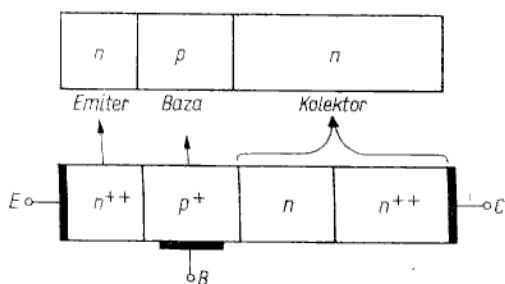
toru w celu uzyskania omowego styku końcówki metalowej z warstwą kolektora. Ponieważ po dyfuzji rozkład koncentracji domieszek jest w przybliżeniu wykładniczo malejący, zatem koncentracja akceptorów w bazie zmienia się od ok. $5 \cdot 10^{23} \text{ m}^{-3}$ przy złączu emiter-baza do ok. $5 \cdot 10^{20} \text{ m}^{-3}$ przy złączu baza-kolektor. W celu wykonania doprowadzeń do struktury półprzewodnikowej tranzystora wykonuje się proces naporowywania metalu (najczęściej Al) na całą powierzchnię płytki, po czym w procesie fotolitografii warstwa metalu jest wytrawiana tak, by pozostały ścieżki metalizacji w obszarach kontaktów z warstwami emitera, bazy i kolektora. Należy zauważyć, że w ten sposób jednocześnie wykonuje się w jednej płytce o średnicy kilku centymetrów kilkaset do kilku tysięcy jednakowych tranzystorów. Po pocięciu płytki na kryształki (tzw. mikroplątki lub struktury) odpowiadające pojedynczym tranzystorom wykonuje się proces mikromontażu. Mikromontaż polega na przylutowaniu mikroplątki do podstawki odpowiedniej obudowy, wykonaniu połączeń cienkim drutem (najczęściej drut złoty o średnicy ok. $25 \mu\text{m}$) pól kontaktowych z tzw. przepustami, czyli z końcówkami wychodzącymi na zewnątrz obudowy i hermetycznym zamknięciu obudowy. Jest to oczywiście bardzo zgrubny opis procesu wytwarzania tranzystora epiplanarnego. Cały proces wytwarzania tranzystora składa się z kilkuset operacji szczegółowych, z których warto jeszcze wspomnieć o testowaniu wykonywanym przed pocięciem płytki na mikroplątki. Chodzi o to, by wadliwie działające struktury nie były poddawane bardzo pracochłonnym i drogim operacjom mikromontażu i obudowywania.

W analizie fizycznej tranzystora będziemy rozważać pewną wyidealizowaną, jednowymiarową strukturę $n^{++}-p^{+}-n-n^{++}$, odpowiadającą wycinkowi rzeczywistej struktury — zaznaczonemu na rys. 5.8a linią przerywaną. Jednowymiarowa idealizacja tranzystora oznacza, że uwzględnia się tylko przepływ nośników w kierunku prostopadłym do powierzchni płytki, czyli pomija się prądy płynące przez boczne powierzchnie złączy. Jest to usprawiedliwione w tym przypadku, gdy rozmiary liniowe, określające powierzchnię struktury tranzystorowej, są znacznie większe niż grubości warstw emitera i bazy. Pierwszy rzut oka na rys. 5.8a każe powątpiewać, czy ten warunek spełnia rozpatrywana struktura. Jednak należy uwzględnić, że schematyczny przekrój struktury tranzystora przedstawiono na rys. 5.8a bez zachowania skali, gdyż chodzi tu o przejrzyste pokazanie poszczególnych warstw. Na rys. 5.8b przedstawiono strukturę identyczną jak na rys. 5.8a, lecz z zachowaniem skali. Ten rysunek uwiadcza, jak bardzo płaska jest struktura tranzystora, co w pełni usprawiedliwia pominięcie zjawisk związanych z krawędziami bocznymi. Tylko niektóre właściwości tranzystora, np. mechanizm przebicia, mogą silnie zależeć od zjawisk krawędziowych i w tych przypadkach analiza oparta na modelu jednowymiarowym może okazać się niesłuszna. W bardziej pogłębionej analizie konieczne będzie również uwzględnienie faktu, że prąd bazy płynie w kierunku równoległym do powierzchni.

Podstawowe zasady działania tranzystora bipolarnego

5.3

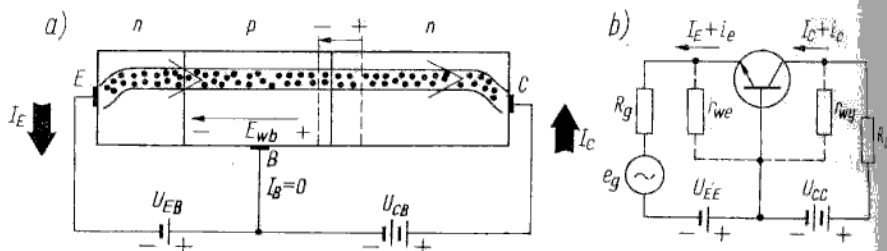
Jak stwierdzono w poprzednim punkcie, działanie tranzystora można analizować korzystając z wyidealizowanej jednowymiarowej struktury $n-p-n$ będącej wycinkiem struktury rzeczywistej (ściśle biorąc jest to struktura $n^{++}-p^{+}-n-n^{++}$ przedstawiona na rys. 5.10). Rozpatrzmy przepływ prądów w tej strukturze przy polaryzacji złącza $E-B$ w kierunku przewodzenia, a złącza $B-C$ w kierunku



Rys. 5.10

Wyidealizowana jednowymiarowa struktura tranzystora epiplanarnego typu n-p-n (wycinek struktury rzeczywistej)

zaporowym. Przy takiej polaryzacji tranzystor spełnia funkcję elementu czynnego, tj. może służyć do liniowego wzmacniania sygnałów elektrycznych. Najbardziej uproszczony obraz zjawisk zachodzących w tranzystorze przedstawiono na rys. 5.11. Podobnie jak w analizie złącza p-n przyjmujemy, że napięcia pola-



Rys. 5.11

Najbardziej uproszczona ilustracja zjawisk zachodzących w tranzystorze: a) obraz fizyczny przepływu prądu; b) układ włączenia tranzystora pracującego jako wzmacniacz

ryzacji U_{EB} , U_{CB} odkładają się wyłącznie na odpowiednich warstwach zaporowych. Wskutek polaryzacji złącza E-B w kierunku przewodzenia z emitera do bazy są wstrzykiwane elektrony. W bazie istnieje tzw. *wbudowane pole elektryczne* E_{wb} , spowodowane nierównomiernym rozkładem koncentracji domieszek. Zjawisko powstawania pola elektrycznego w niejednorodnym materiale półprzewodnikowym jest znane z p. 1.9. W rozważanym przypadku koncentracja domieszki akceptorowej w bazie maleje w kierunku od emitera do kolektora. Dlatego pole elektryczne E_{wb} , przeciwdziałające dyfuzji dziur, jest skierowane od potencjału dodatniego przy kolektorze do potencjału ujemnego przy emiterze. Elektrony wstrzyknięte z emitera do bazy są unoszone przez E_{wb} w kierunku kolektora¹⁾. Po przejściu przez bazę elektrony dostają się do warstwy zaporowej złącza B-C, w której istnieje silne pole elektryczne „wymiatające” te elektrony dalej do obwodu kolektora. Strumień elektronów wstrzykiwanych z emitera do bazy tworzy prąd emitera w obwodzie wejściowym, a strumień elektronów odbieranych przez kolektor tworzy prąd kolektora w obwodzie wyjściowym. Strumień elektronów odbieranych przez kolektor jest równy strumieniowi elektronów wstrzykiwanych przez emiter, czyli prąd kolektora nie zależy od napięcia U_{CB}

¹⁾ Istnienie pola wbudowanego w bazie nie jest warunkiem koniecznym dla pracy tranzystora. W tranzystorze bezdyfuzyjnym nie ma pola E_{wb} , a transport nośników w bazie (od emitera do kolektora) odbywa się wskutek dyfuzji.

lecz jest funkcją napięcia U_{EB} . Zgodnie z konwencją przyjętą w elektrotechnice prąd jest skierowany przeciwnie do strumienia elektronów. W tak uproszczonym modelu tranzystora prąd wyjściowy I_C jest równy prądowi wejściowemu I_E

$$I_C = I_E$$

czyli współczynnik wzmocnienia prądowego definiowany jako I_{wy}/I_{we}

$$\alpha_N = \frac{I_C}{I_E} = 1$$

przy czym α_N jest wzmocnieniem dla prądu stałego. Wzmocnienie dla małych przyrostów prądu jest też równe jedności

$$\alpha = \frac{\Delta I_C}{\Delta I_E} = 1$$

W tym miejscu trudno oprzeć się odruchowi zwątpienia w jakąkolwiek użyteczność elementu, który bez wzmocnienia przenosi prąd z wejścia do wyjścia. Można jednak łatwo wykazać, że moc wydzielana w obwodzie wyjściowym jest większa niż moc dostarczona do wejścia tranzystora.

Zgodnie z układem przedstawionym na rys. 5.11b tranzystor jest polaryzowany z baterii U_{EB} , U_{CC} , a ponadto w obwodzie wejściowym jest włączone źródło e_g małego sygnału sinusoidalnego. Baterie U_{EB} , U_{CC} powodują przepływ prądów stałych I_E , I_C , natomiast ze źródła e_g płynie w obwodzie wejściowym prąd sinusoidalny i_e o amplitudzie I_{em} , który powoduje przepływ prądu sinusoidalnego i_c o amplitudzie I_{cm} w obwodzie wyjściowym. Obliczamy moc sygnału sinusoidalnego na wejściu i wyjściu tranzystora.

Moc dostarczana do tranzystora na wejściu

$$P_{we} = I_{em}^2 r_{we} \quad (5.1)$$

przy czym r_{we} — rezystancja wejściowa tranzystora.

Moc odbierana w obciążeniu R_L

$$P_{wy} = I_{cm}^2 R_L \quad (5.2)$$

Maksimum mocy w obciążeniu uzyskuje się przy spełnieniu warunku dopasowania

$$r_{wy} = R_L \quad (5.3)$$

Uwzględniając (5.1) do (5.3) można wyrazić wzmocnienie mocy w postaci

$$k_p = \frac{I_{cm}^2 R_L}{I_{em}^2 r_{we}} = \frac{I_{cm}^2 r_{wy}}{I_{em}^2 r_{we}} \quad (5.4)$$

Biorąc pod uwagę, że $\alpha = I_{cm}/I_{em}$, to

$$k_p = \alpha^2 \frac{r_{wy}}{r_{we}} \quad (5.4a)$$

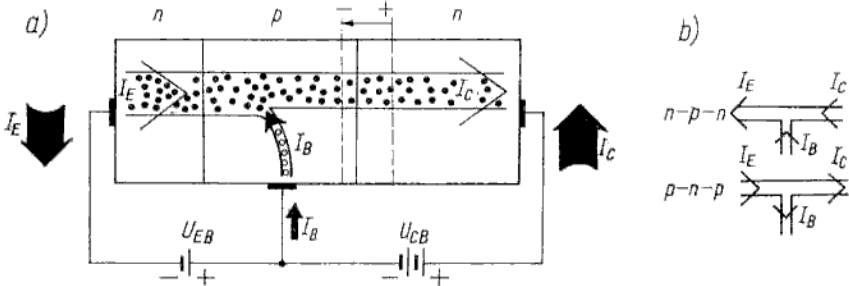
Ponieważ w rozpatrywanym modelu uproszczonym $\alpha = 1$, więc

$$k_p = \frac{r_{wy}}{r_{we}} \quad (5.5)$$

Stosunek r_{wy}/r_{we} wynosi kilka tysięcy, gdyż r_{we} jest małą rezystancją przyrostową złącza $E-B$ spolaryzowanego w kierunku przewodzenia (z rozdziału 3 wiadomo, że rezystancja przyrostowa złącza $p-n$ wynosi φ_T/I , czyli przykładowo w temperaturze pokojowej przy prądzie emitera $I_E = 1 \text{ mA}$, $r_{we} = 25 \Omega$), r_{wy} zaś jest bardzo dużą rezystancją przyrostową złącza $B-C$ spolaryzowanego w kierunku zaporowym (w przypadku idealnego źródła prądowego rezystancja r_{wy} byłaby nieskończenie wielka, w rzeczywistości jest rzędu kilkuset kiloomów).

Tranzystor jest rzeczywiście „elementem transformującym rezystancję” i wzmacniaczem mocy.

Rozpatrzmy teraz nieco dokładniejszy model zjawisk zachodzących w tranzystorze zgodnie z rys. 5.12. Uwzględnimy tu dodatkowo możliwość rekombinacji elektronów



Rys. 5.12

Dokładniejsza ilustracja zjawisk zachodzących w tranzystorze z uwzględnieniem prądu spowodowanego rekombinacją nośników w bazie: a) obraz strumieni nośników; b) schematyczny rozptyw prądów w tranzystorze $n-p-n$ i $p-n-p$

z dziurami w obszarze bazy. Jednym z podstawowych założeń wprowadzanych w analizie tranzystora jest zasada obojętności elektrycznej całego obszaru bazy. Stąd wynika, że liczby elektronów i dziur nadmiarowych w bazie są sobie równe. Jeżeli na przykład z emitera w pewnej chwili wpływa do bazy 100 elektronów, to ładunek ujemny, jaki tworzą te elektrony, przyciąga z najbliższego sąsiedztwa 100 dziur. Niedomiar tych stu dziur „w najbliższym sąsiedztwie” jest uzupełniany przez przyptyw dziur z następnych obszarów bazy, aż ostatecznie wpływa 100 dziur z obwodu zewnętrznego przez elektrodę bazy (jest to oczywiście jednoznaczne z usunięciem 100 elektronów z warstwy bazy do obwodu zewnętrznego, gdyż w obwodzie zewnętrznym płynie tylko prąd elektronowy). Cały proces równoważenia się ładunków nośników nadmiarowych przebiega w czasie $\tau \approx 10^{-11} \dots 10^{-13} \text{ s}$, czyli — praktycznie biorąc — jest to proces natychmiastowy. Dlatego można twierdzić, że zawsze istnieje równowaga ładunku elektronów i dziur nadmiarowych w obszarze bazy.

Skorzystajmy teraz z zasady obojętności elektrycznej bazy w celu wyjaśnienia rozptywu prądów, przedstawionego na rys. 5.12. Szybkość strumienia elektronów wpływających z emitera do bazy niech przykładowo wynosi 100 elektronów na sekundę, a szybkość rekombinacji par elektron-dziura niech będzie 1 para na sekundę. W pierwszej sekundzie z emitera wpływa do bazy 100 elektronów i z uwagi na zasadę obojętności elektrycznej bazy w tym samym czasie przez elektrodę bazy wypływa 100 elektronów do obwodu zewnętrznego, co oznacza inaczej, że do obszaru bazy wpływa 100 dziur. W ten sposób w chwili włączenia tranzystora prąd bazy jest równy prądowi emitera. Jest to stan nieustalony, który będzie

omówiony w dalszej części rozdziału. Obecnie interesuje nas stan ustalony, który w rozpatrywanym przykładzie oznacza, że w jednej sekundzie wpływa do bazy 100 elektronów z emitera, wypływa 99 elektronów do kolektora, a jeden rekombinuje z dziurą. Oznacza to ubytek ładunku jednej dziury, która musi być dostarczona do bazy z obwodu zewnętrznego.

Zatem w stanie ustalonym liczba elektronów odbieranych w jednostce czasu przez kolektor jest mniejsza niż liczba elektronów wstrzykiwanych do bazy z emitera, a więc prąd kolektora jest mniejszy niż prąd emitera. Różnica tych dwu prądów jest spowodowana rekombinacją elektronów z dziurami (nie wszystkie elektrony wstrzyknięte z emitera zdołają dotrzeć do kolektora). A ponieważ baza musi być obojętna elektrycznie, z zewnętrznego obwodu bazy wpływa strumień dziur uzupełniających „straty” ładunku dodatniego, spowodowane rekombinacją. Ten strumień nośników tworzy prąd bazy I_B . Stąd można zapisać podstawowe równanie prądów w tranzystorze

$$I_E = I_B + I_C \quad (5.6)$$

śluszne również dla małych przyrostów

$$\Delta I_E = \Delta I_B + \Delta I_C \quad (5.7)$$

Bilans prądów w tranzystorze wynika konsekwentnie z zasady obojętności elektrycznej bazy i obowiązuje zarówno dla tranzystorów $n-p-n$ jak również $p-n-p$ (rys. 5.12b), zarówno dla stanu ustalonego jak i nieustalonego.

Tranzystor jest tym lepszy (tym większe ma wzmocnienie), im mniej nośników rekombinuje w bazie. W dobrym tranzystorze:

$$I_C \leq I_E$$

$$I_B \ll I_C$$

$$I_B \ll I_E$$

Zatem współczynnik wzmocnienia prądowego

$$\alpha = \frac{\Delta I_C}{\Delta I_E}$$

jest nieco mniejszy niż jedność. Najczęściej

$$\alpha \approx 0,980 \dots 0,995$$

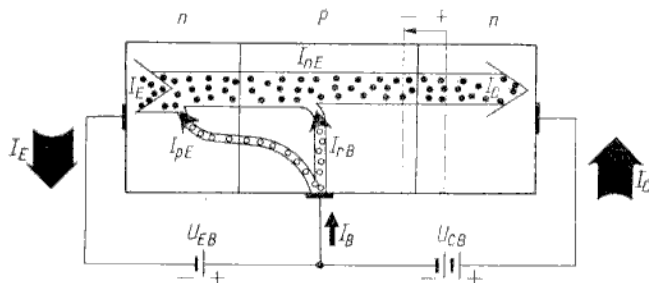
Współczynnik wzmocnienia prądowego tranzystora można również zdefiniować jako stosunek prądu kolektora do prądu bazy:

$$\beta_N = \frac{I_C}{I_B} \quad \beta = \frac{\Delta I_C}{\Delta I_B} \quad (5.8)$$

Ta definicja wzmocnienia prądowego ma zastosowanie w przypadku takiego układu włączenia tranzystora, w którym prądem wejściowym jest prąd bazy. Trzy możliwe warianty układu włączenia tranzystora omówimy nieco dalej, a obecnie zauważmy tylko, że istnieje bezpośredni związek między α i β , gdyż uwzględniając (5.7):

$$\alpha = \frac{\beta}{1 + \beta} \quad \beta = \frac{\alpha}{1 - \alpha} \quad (5.9)$$

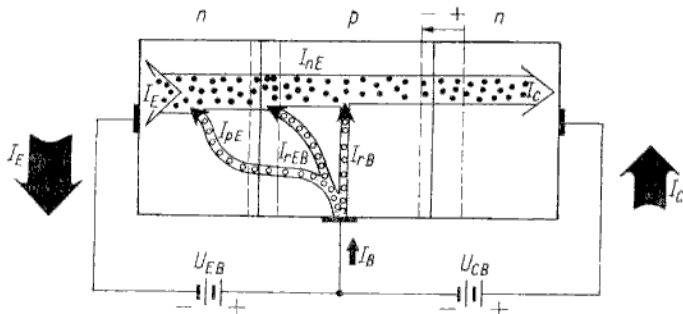
Dalsze uściślenia modelu zjawisk zachodzących w tranzystorze, przedstawione na rys. 5.13, 5.14, wynikają z uwzględnienia dodatkowych składowych prądów emitera i bazy. Na rysunku 5.13 przedstawiono składową I_{pE} prądu dyfuzji dziur z bazy do emitera, gdzie rekombinują one z elektronami. Składowa I_{pE} cyrkuluje w obwo-



Rys. 5.13

Dalsze uściślenie obrazu zjawisk w tranzystorze; uwzględniono prąd dyfuzji dziur z bazy do emitera I_{pE}

dzie wejściowym dając jednakowy wkład do prądu emitera i prądu bazy (w złączu niesymetrycznym $n^{++}-p^+$ strumień dyfuzji dziur z bazy do emitera jest znacznie mniejszy niż strumień dyfuzji elektronów z emitera do bazy). Na rysunku 5.14 uwzględniono również składową prądu rekombinacji w obszarze warstwy



Rys. 5.14

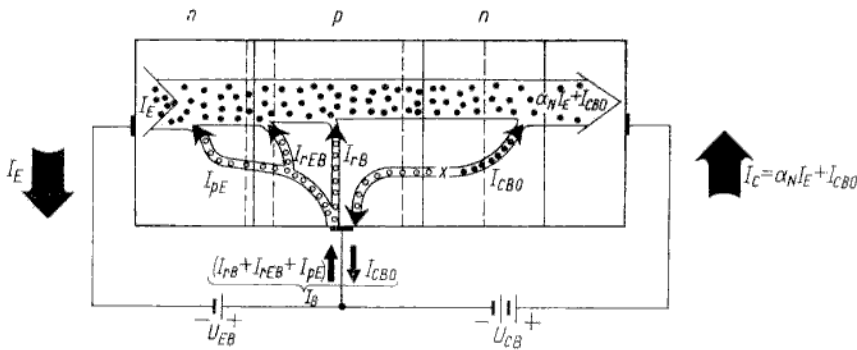
Kolejne uściślenie obrazu zjawisk w tranzystorze; uwzględniono prąd rekombinacji I_{rB} w warstwie zaporowej złącza E-B

zaporowej (por. „przybliżenie drugiego stopnia” w rozdziale 3). Te dokładniejsze modele wykorzystamy w p. 5.5 do analizy zależności wzmocnienia prądowego od parametrów konstrukcyjno-technologicznych tranzystora.

Pełny obraz rozplywu prądów w tranzystorze przedstawiono na rys. 5.15, na którym uwzględniono również tzw. prąd zerowy I_{CBO} .

W tranzystorze krzemowym jest to prąd nośników mniejszościowych generowanych w obszarze warstwy zaporowej złącza C-B spolaryzowanego w kierunku zaporowym. Para elektron-dziura powstająca w warstwie zaporowej jest natychmiast „wymiatana”, przy czym elektron podąża do kolektora, a dziura do bazy. Prąd I_{CBO} dodaje się do prądu kolektora, a odejmuje od prądu bazy. Zatem efektywnie w obwodzie kolektora płynie prąd

$$I_C = \alpha_N I_E + I_{CBO} \quad (5.10)$$



Rys. 5.15

Najbardziej dokładny obraz zjawisk zachodzących w tranzystorze; uwzględniono prąd zerowy kolektora I_{CBO}

a w obwodzie bazy

$$I_B = I_{rB} + I_{rEB} + I_{pE} - I_{CBO} \quad (5.11)$$

przy czym: I_{rB} — prąd rekombinacji w bazie; I_{rEB} — prąd rekombinacji w warstwie zaporowej złącza emiter-baza; I_{pE} — prąd dyfuzji dziur z bazy do emitera.

Uwzględniając zależności (5.10), (5.11) należy skorygować wzory na wzmacnienie dla prądu stałego:

$$\alpha_N = \frac{I_C - I_{CBO}}{I_E} \quad (5.12)$$

$$\beta_N = \frac{I_C - I_{CBO}}{I_B + I_{CBO}} \quad (5.13)$$

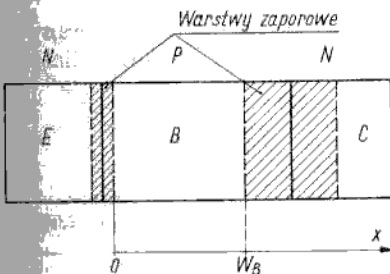
Współczynniki wzmacnienia prądowego α , β dla małych sygnałów pozostają bez zmian.

Rozkład koncentracji nośników nadmiarowych w bazie

5.4

Jest to bardzo ważny temat, który będzie służył za podstawę do wielu dalszych rozważań.

Interesuje nas rozkład koncentracji nośników w obszarze ograniczonym z jednej strony prawą krawędzią warstwy zaporowej złącza E-B, z drugiej zaś — lewą krawędzią



Rys. 5.16

Ilustracja wyjaśniająca pojęcie szerokości efektywnej bazy W_B

warstwy zaporowej złączą C-B. Odległość między tymi krawędziami będziemy nazywać szerokością efektywną bazy i oznaczymy jako W_B (rys. 5.16).

Podstawowym równaniem problemu jest równanie prądu elektronów w bazie, które w postaci ogólnej zawiera składową unoszenia i dyfuzji

$$J_n = q\mu_n n_b E + qD_n \frac{dn_b}{dx} \quad (5.14)$$

Równanie to rozwiążemy najpierw dla dwu skrajnie uproszczonych przypadków, tj. uwzględniając albo tylko składową unoszenia (tak przyjmowaliśmy w dotychczasowym opisie działania tranzystora dryftowego), albo tylko składową dyfuzji (tak jest w tranzystorze bezdryftowym). Następnie rozwiążemy równanie (5.14) w postaci ogólnej.

Przypadek pierwszy (tylko unoszenie)

$$J_n = q\mu_n n_b E \quad (5.15)$$

W stanie ustalonym można przyjąć, że gęstość prądu J_n jest wielkością stałą, niezależną od współrzędnej x (pomija się wpływ rekombinacji w bazie) i prawie dokładnie równą gęstości prądu emitera ($J_n \approx J_E$). Natężenie pola E pochodzi od pola wbudowanego w bazie, określonego wzorem (por. p. 1.9)

$$E_{wb} = \varphi_T \frac{1}{N_A} \frac{dN_A}{dx} \quad (5.16)$$

Profil koncentracji domieszek w bazie można w przybliżeniu zapisać funkcją wykładniczą

$$N_A(x) = N_A(0) \exp \left(-\eta \frac{x}{W_B} \right) \quad (5.17)$$

Stałą η wyznacza się podstawiając $x = W_B$ do (5.17)

$$\eta = \ln \frac{N_A(0)}{N_A(W_B)} \quad (5.18)^{1)}$$

Podstawiając (5.17) do (5.16) otrzymuje się

$$E_{wb} = - \frac{\eta \varphi_T}{W_B} \quad (5.19)$$

Natężenie pola elektrycznego wbudowanego w bazie nie zależy od współrzędnej x , a ponieważ J_n również nie zależy od współrzędnej x , więc po podstawieniu (5.19) do (5.15) otrzymuje się stałą wartość koncentracji n_b w całym obszarze bazy

$$n_b = - \frac{J_n W_B}{q\mu_n \eta \varphi_T} \quad (5.20)$$

Uwzględniając, że prąd J_n ma znak ujemny (jest to prąd skierowany przeciwnie do dodatniego zwrotu osi x) oraz biorąc pod uwagę zależność Einsteina ($\varphi_T = = D_n/\mu_n$)

$$n_b = \frac{|J_n| W_B}{qD_n \eta} \quad (5.21)$$

¹⁾ Często w literaturze wprowadza się stałą $\eta = \frac{1}{2} \ln \frac{N_A(0)}{N_A(W_B)}$, co upraszcza niektóre wzory, a inne komplikuje.

Przypadek drugi (tylko dyfuzja)

$$J_n = qD_n \frac{dn_b}{dx} \quad (5.22)$$

Stąd

$$n_b(x) = \frac{J_n}{qD_n} x + C \quad (5.23)$$

przy czym C — stała całkowania, która jest równa $n_b(0)$, a może być wyznaczona z warunku $n_b(W_B) \approx 0$ (na lewej krawędzi warstwy zaporowej złącza $B-C$ elektrony są silnie unoszone przez pole elektryczne istniejące w tej warstwie, dlatego koncentracja elektronów w tym punkcie jest, praktycznie biorąc, równa zero).

Zatem

$$C = n_b(0) = -\frac{J_n}{qD_n} W_B \quad (5.24)$$

Stąd

$$n_b(x) = \frac{|J_n|}{qD_n} (W_B - x) \quad (5.25)$$

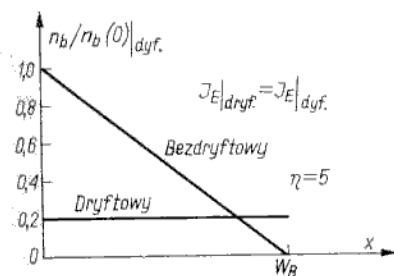
lub

$$n_b(x) = n_b(0) \left(1 - \frac{x}{W_B}\right) \quad (5.26)$$

Z tej zależności wynika, że w przypadku istnienia tylko składowej prądu dyfuzji rozkład koncentracji nośników nadmiarowych w bazie jest liniowy. Koncentracje elektronów dla współrzędnej $x = 0$ można wyznaczyć ze statystyki Boltzmannna

$$n_b(0) = n_{p0}(0) \exp \frac{U_{EB}}{\varphi_T} = \frac{n_i^2}{N_A(0)} \exp \frac{U_{EB}}{\varphi_T} \quad (5.27)$$

Porównajmy teraz rozkłady koncentracji nośników nadmiarowych w bazie tranzystora dla przypadku transportu tylko unoszeniowego (idealizacja tranzystora dryftowego) i tylko dyfuzyjnego (tranzystor z jednorodną bazą — bezdryftowy) — rys. 5.17.



Rys. 5.17

Porównanie wyidealizowanych rozkładów koncentracji nośników nadmiarowych w bazie tranzystorów dryftowego i bezdryftowego przy jednakowych wartościach prądu emitra

Przy jednakowym prądzie J_n

$$n_b(x) \Big|_{dryf} = \eta \left(1 - \frac{x}{W_B} \right) n_b(x) \Big|_{dryft} \quad (5.28)$$

dla $x = 0$

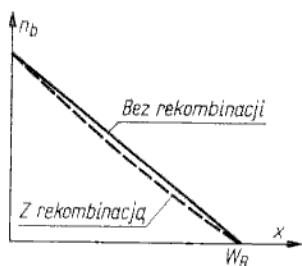
$$n_b(0) \Big|_{dryf} = \eta n_b(0) \Big|_{dryft} \quad (5.29)$$

Przy jednakowych wartościach $n_b(0)$, czyli zgodnie z (5.27) — przy jednakowym napięciu U_{EB}

$$J_n \Big|_{dryft} = \eta J_n \Big|_{dryf} \quad (5.30)$$

Stąd wynika ważny wniosek, że w porównywalnych tranzystorach bezdryftowym i dryftowym (o jednakowych koncentracjach domieszek w emiterze i w bazie dla współrzędnej $x = 0$), przy jednakowych napięciach U_{EB} , prąd emitera jest η razy większy w tranzystorze dryftowym.

Jeżeli chodzi o tranzystor z jednorodną bazą (bezdryftowy), to otrzymany rozkład $n_b(x)$ z dużą dokładnością odpowiada rzeczywistej sytuacji. Liniowa zależność $n_b(x)$ ulega tylko nieznacznej modyfikacji przy uwzględnieniu zjawiska rekombinacji nośników w bazie (rys. 5.18).



Rys. 5.18

Wpływ rekombinacji w bazie na kształt rozkładu $n_b(x)$ w tranzystorze bezdryftowym

Rozkład $n_b(x) = \text{const}$, otrzymany dla uproszczonego modelu tranzystora dryftowego, jest natomiast nieco gorszym przybliżeniem rzeczywistości. Przede wszystkim należy zauważyć, że w rzeczywistości $n_b(W_B) \approx 0$, gdyż na krawędzi warstwy zaporowej złącza $B-C$ nośniki są „wymiatane” do kolektora przez silne pole elektryczne, istniejące w tej warstwie. Jeżeli $n_b(W_B) \approx 0$, to zapewne w części bazy przylegającej do kolektora istnieje pewien gradient koncentracji nośników, a więc w tej części bazy istotne znaczenie ma prąd dyfuzji. Dlatego dokładniejszy przebieg rozkładu koncentracji nośników nadmiarowych w bazie tranzystora dryftowego otrzymuje się rozwiązując równanie prądu (5.14) w ogólnej postaci zawierającej zarówno składową unoszenia jak i dyfuzji.

Przypadek trzeci (unoszenie i dyfuzja). Równanie (5.14) przepisane w postaci

$$\frac{dy}{dx} + Py = Q$$

przy czym:

$$P = \frac{E}{\varphi_T}$$

$$Q = \frac{J_n}{qD_n}$$

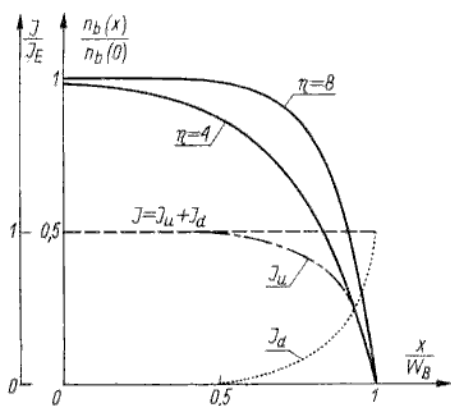
$$y = n_b$$

ma rozwiązanie ogólne (można je znaleźć w każdym poradniku matematyki)

$$y(x) = e^{-Px} \left[\frac{Q}{P} (e^{Px} - 1) + C \right]$$

Po wyznaczeniu stałej całkowania C na podstawie warunku $y(W_B) = 0$ oraz zamianie zmiennych posiłkowych na zmienne pierwotne otrzymuje się

$$n_b(x) \approx \frac{|J_n| W_B}{qD_n \eta} \left[1 - e^{-\eta \left(1 - \frac{x}{W_B}\right)} \right] \quad (5.31)$$



Rys. 5.19
Rozkłady koncentracji nośników i prądów w bazie tranzystora driftowego przy uwzględnieniu składowej unoszenia i dyfuzji (J_u — prąd unoszenia, J_d — prąd dyfuzji)

Na rysunku 5.19 przedstawiono wykresy $n_b(x)$ obliczone na podstawie tej zależności dla typowych wartości współczynnika η , tj. $\eta = 4; 8$. Na tym samym rysunku wykreślono również przebieg całkowitego prądu elektronów w bazie i jego dwie składowe dla $\eta = 8$. Przedstawione wykresy potwierdzają słuszność założenia, że prąd unoszenia przeważa zdecydowanie nad prądem dyfuzji. Tylko w niewielkim obszarze bazy sąsiadującym z kolektorem prąd dyfuzji odgrywa istotną rolę.

Uwzględniając (5.31) można uściślić zależności (5.28 do 5.30):

$$n_b(x) \Big|_{dyf} = \left(1 - \frac{x}{W_B}\right) \frac{\eta}{1 - e^{-\eta \left(1 - \frac{x}{W_B}\right)}} n_b(x) \Big|_{dryft} \quad (5.28a)$$

$$n_b(0) \Big|_{dyf} = \frac{\eta}{1 - e^{-\eta}} n_b(0) \Big|_{dryft} \quad (5.29a)$$

$$J_n \Big|_{dryft} = \frac{\eta}{1 - e^{-\eta}} J_n \Big|_{dyf} \quad (5.30a)$$

Widać, że przybliżone zależności (5.28) do (5.30) są dość dokładne dla dużych wartości współczynnika η ($\eta > 3$).

Współczynnik wzmocnienia prądowego α_N

5.5

Znając rozptyw prądów w tranzystorze (rys. 5.14) można określić zależność współczynnika wzmocnienia prądowego α_N od parametrów materiałowych i punktu pracy.

Z definicji

$$\alpha_N = \frac{I_C}{I_E}$$

Podzielmy licznik i mianownik tego wyrażenia przez I_{nE} (I_{nE} jest prądem elektronów na początku bazy, tzn. na prawej krawędzi warstwy zaporowej złącza $B-B'$)

$$\alpha_N = \frac{I_C}{I_{nE}} \frac{I_{nE}}{I_E} \quad (5.32)$$

Wprowadzamy dwa współczynniki charakteryzujące właściwości wewnętrznej struktury tranzystora:

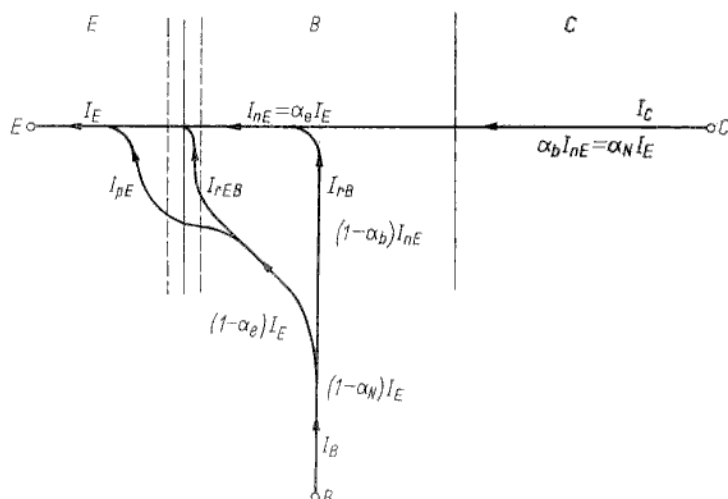
$\alpha_e = I_{nE}/I_E$ — współczynnik sprawności wstrzykiwania emitera, określający, jaką część całkowitego prądu emitera stanowi strumień nośników wstrzykiwanych do obszaru bazy;

$\alpha_b = I_C/I_{nE}$ — współczynnik transportu (nazywany również współczynnikiem rekombinacji w bazie), określający, jaka część strumienia elektronów wstrzykiwanych do bazy jest odbierana przez kolektor.

Zatem

$$\alpha_N = \alpha_e \alpha_b \quad (5.33)$$

Na rysunku 5.20 przedstawiono rozptyw prądów identyczny jak na rys. 5.14, lecz z uwzględnieniem wprowadzonych współczynników. Od jakich parametrów materiałowych zależą współczynniki α_e , α_b ?



Rys. 5.20

Rozptyw prądów w tranzystorze (strzałki wskazują konwencjonalny kierunek przepływu prądu, czyli są zgodne z kierunkiem strumienia dziur, a przeciwne do kierunku strumienia elektronów)

Współczynnik sprawności wstrzykiwania emitera

$$\alpha_e = \frac{I_{nE}}{I_E} = \frac{I_{nE}}{I_{nE} + I_{pE} + I_{rEB}} \quad (5.34)$$

Rozpisałiśmy prąd emitera w postaci sumy prądu dyfuzji elektronów I_{nE} i dziur I_{pE} oraz prądu rekombinacji nośników w warstwie zaporowej złącza emiter-baza I_{rEB} .

Na podstawie teorii złącza $p-n$ (rozdz. 3) oraz uwzględniając związek (5.30a) otrzymuje się zależności na trzy składowe prądu emitera:

$$J_{pE} = qD_p \frac{n_i^2}{N_{DE}W_E} \exp \frac{U_{EB}}{\varphi_T} \quad (5.35)$$

$$J_{nE} = qD_n \frac{n_i^2}{N_{AB}} \frac{1}{W_B} \frac{\eta}{1 - e^{-\eta}} \exp \frac{U_{EB}}{\varphi_T} \quad (5.36)$$

$$J_{rEB} = \frac{1}{2} q \frac{n_i}{\tau} W_{EB} \exp \frac{U_{EB}}{2\varphi_T} \quad (5.37)$$

przy czym: N_{DE} — koncentracja donorów w emiterze na krawędzi warstwy zaporowej złącza $E-B$; N_{AB} — koncentracja akceptorów w bazie na krawędzi warstwy zaporowej złącza $E-B$; W_E — szerokość emitera; przyjęto, że W_E jest znacznie mniejsze niż droga dyfuzji dziur w emiterze; W_{EB} — szerokość warstwy zaporowej złącza $E-B$; τ — czas życia nośników.

Dzielnik w (5.34) licznik i mianownik przez I_{nE} oraz uwzględniając (5.35) do (5.37) otrzymuje się

$$\alpha_e = \frac{1}{1 + K + L} \quad (5.38)$$

przy czym:

$$K = \frac{D_p N_{AB} W_B}{D_n N_{DE} W_E} \frac{1 - e^{-\eta}}{\eta} \quad (5.38a)$$

$$L = \frac{1}{2} \frac{N_{AB} W_B W_{EB}}{n_i D_n \tau} \frac{1 - e^{-\eta}}{\eta} \exp \left(-\frac{U_{EB}}{2\varphi_T} \right) \quad (5.38b)$$

Otrzymane wyrażenia mają dość rozwlekłą postać, są jednak wystarczająco przejrzyste, by można było na ich podstawie wyciągnąć bardzo istotne wnioski natury ogólnej. Przede wszystkim należy stwierdzić, że jest pożądanym, by współczynnik α_e miał wartość jak najbliższą 1 (zawsze $\alpha_e < 1$), tzn. jest pożądanym spełnienie warunku $(K + L) \ll 1$. Składnik L , związany z prądem rekombinacji w warstwie zaporowej złącza $E-B$, ma istotne znaczenie w przypadku małej koncentracji nośników samoistnych n_i , czyli należy go uwzględniać w tranzystorze krzemowym, lecz jest do pominięcia w tranzystorze germanowym. Ponadto należy zauważyć, że ten składnik silnie zależy od napięcia U_{EB} , przy czym maleje on w miarę zwiększania poziomu wstrzykiwania nośników (zwiększanie napięcia U_{EB}). Stąd wynika, że wartość współczynnika α_e jest coraz większa w miarę zwiększania napięcia U_{EB} , czyli w miarę zwiększania prądu emitera. Przy odpowiednio dużym prądzie emitera składnik L można pominąć w porównaniu ze składnikiem K .

Na wartość składnika K decydujący wpływ ma stosunek N_{AB}/N_{DE} , przy czym jest požądane, żeby $N_{AB} \ll N_{DE}$. Dlatego emiter jest bardzo silnie domieszkowany ($N_{DE} \approx 10^{26} \text{ m}^{-3}$), a baza w obszarze sąsiadującym z z emiterem ma koncentrację domieszek kilka rzędów wartości mniejszą ($N_{AB} \approx 5 \cdot 10^{23} \text{ m}^{-3}$).

Składniki K , L są tym mniejsze, im większa jest wartość współczynnika η . Ponieważ w typowym przypadku koncentracja domieszek maleje wzdłuż bazy (w przybliżeniu wykładniczo) od $5 \cdot 10^{23}$ do $5 \cdot 10^{20} \text{ m}^{-3}$, zatem

$$\eta = \ln \frac{5 \cdot 10^{23}}{5 \cdot 10^{20}} \approx 7$$

W tranzystorach z jednorodną bazą $\eta = 0$, a ponieważ

$$\left. \frac{1 - e^{-\eta}}{\eta} \right|_{\eta \rightarrow 0} = 1$$

widać zatem, że brak pola wbudowanego w bazie tych tranzystorów wpływa niekorzystnie na wartość współczynnika α_e .

Dla tranzystorów germanowych z jednorodną bazą $\left(\frac{1 - e^{-\eta}}{\eta} = 1; L \approx 0 \right)$

$$\alpha_e \approx \frac{1}{1 + \frac{D_p N_{AB} W_B}{D_n N_{DE} W_E}} \quad (5.39)$$

Wyrażenie (5.39) jest najczęściej podawane w literaturze jako ogólnie słuszne dla tranzystorów bipolarnych, lecz trzeba pamiętać, że dla tranzystorów krzemowych dryftowych (czyli dla większości współczesnych tranzystorów bipolarnych) sprawność wstrzykiwania emitera jest określona bardziej złożoną zależnością (5.38).

Współczynnik transportu nośników w bazie

$$\alpha_b = \frac{I_C}{I_{nE}} = \frac{I_C}{I_C + I_{rB}} \quad (5.40)$$

Podzielmy licznik i mianownik tego wyrażenia przez Q_b — ładunek nośników nadmiarowych w bazie; a ponieważ iloraz ładunku i prądu ma wymiar czasu, zatem wprowadźmy oznaczenia:

$$t_b = \frac{Q_b}{I_C} \quad (5.41a)$$

$$\tau_r = \frac{Q_b}{I_{rB}} \quad (5.41b)$$

Stąd zależność (5.40) przyjmuje postać

$$\alpha_b = \frac{\tau_r}{\tau_r + t_b} \quad (5.42)$$

Wielkości t_b , τ_r mają określony sens fizyczny. Gdyby nie było rekombinacji nośników w bazie, wówczas prąd I_C spowodowałby w czasie t_b usunięcie całego

ładunku Q_b zmagazynowanego w bazie. Oczywiście w stanie ustalonym na miejsce ładunku usuwanego do kolektora wpływa taki sam ładunek z emitera i stan „naładowania” bazy nie ulega zmianie. Zatem wielkość t_b jest to czas przelotu nośników przez bazę. Wyobraźmy sobie teraz inną sytuację. Niech w bazie będzie pewien ładunek Q_b i zostaje przerwany przepływ prądów I_C , I_E , czyli nie ma wymiany ładunku bazy z warstwami kolektora i emitera. Wówczas ładunek Q_b maleje w funkcji czasu wskutek rekombinacji nośników (w stanie ustalonym przy $I_C = 0$, dla podtrzymania stałej wartości ładunku w bazie, $Q_b = \text{const}$, jest konieczny przepływ odpowiedniego prądu rekombinacji bazy $-I_{rB}$). Wielkość τ_r jest stałą czasową tego procesu, czyli jest to po prostu czas życia nośników (przeważnie $\tau_r = 10^{-6} \dots 10^{-9}$ s). W każdym tranzystorze $t_b \ll \tau_r$, czyli współczynnik α_b przyjmuje wartości bardzo bliskie jedności, lecz zawsze ≤ 1 . Jeżeli przykładowo przelot nośników przez bazę jest sto razy szybszy niż ich rekombinacja ($t_b = 0,01 \tau_r$), to na 100 elektronów wstrzykniętych z emitera do bazy (prąd I_{nE}) 99 dotrze do kolektora (prąd I_C), a jeden zrekombinuje w bazie z dziurą (prąd I_{rB}).

Czas τ_r jest parametrem materiałowym, czas t_b natomiast zależy od konstrukcji tranzystora, co można w przybliżeniu wyrazić prostymi związkami. Można wyjść z zależności (5.41a) lub po prostu obliczać czas jako iloraz drogi i prędkości nośników

$$t_b = \frac{W_B}{v}$$

Jeżeli uwzględnimy tylko unoszenie (idealizacja tranzystora dryftowego), to

$$v = \mu_n E_{wb}$$

a więc na podstawie (5.19) (znak „-” nie ma tu żadnego znaczenia, gdyż rozpatruje się wielkości skalarne)

$$t_b = \frac{W_B^2}{\eta D_n} \quad (5.43)$$

Podstawiając (5.43) do (5.42) oraz biorąc pod uwagę, że droga dyfuzji $L = \sqrt{D_n \tau_r}$, można napisać

$$\alpha_b = \frac{1}{1 + \frac{1}{\eta} \left(\frac{W_B}{L} \right)^2} \approx 1 - \frac{1}{\eta} \left(\frac{W_B}{L} \right)^2 \quad (5.44)$$

Jeżeli uwzględnimy tylko dyfuzję (tranzystor bezdryftowy), to na podstawie (5.26):

$$Q_b = q n_b(0) \frac{W_B}{2} \quad (\text{pole trójkąta})$$

$$I_n = q D_n \frac{dn_b}{dx} = q D_n \frac{n_b(0)}{W_B}$$

Stąd

$$t_b = \frac{Q_b}{I_n} = \frac{W_B^2}{2 D_n} \quad (5.45)$$

Podstawiając (5.45) do (5.42) otrzymuje się

$$\alpha_b = \frac{1}{1 + \frac{1}{2} \left(\frac{W_B}{L} \right)^2} \approx 1 - \frac{1}{2} \left(\frac{W_B}{L} \right)^2 \quad (5.46)$$

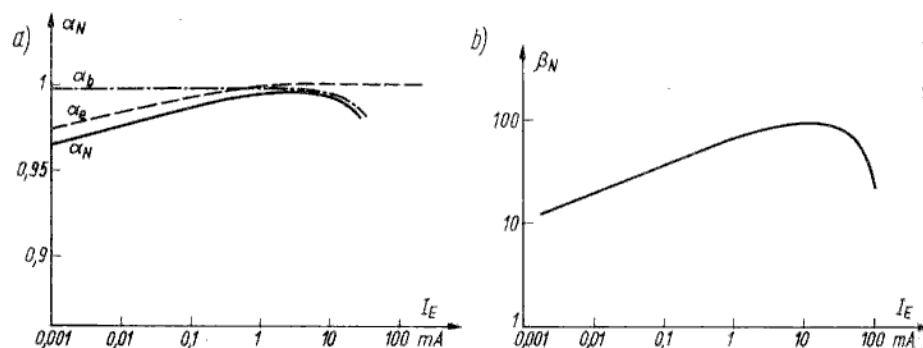
Dokładniejsze wyrażenia na współczynnik α_b w tranzystorze dryftowym (przy uwzględnieniu dryftu i dyfuzji) można otrzymać całkując ładunek nośników nadmiarowych w bazie i obliczając prąd na podstawie rozkładu koncentracji określonego zależnością (5.31):

$$t_b = \frac{W_B^2}{D_n} \frac{\eta - 1 + e^{-\eta}}{\eta^2} \quad (5.47)$$

$$\alpha_b \approx 1 - \left(\frac{W_B}{L} \right)^2 \frac{\eta - 1 + e^{-\eta}}{\eta^2} \quad (5.48)$$

Porównanie wzorów (5.46) z (5.44) lub (5.48) umożliwia stwierdzenie, że współczynnik α_b ma większą wartość w tranzystorze dryftowym.

Na zakończenie tego tematu warto przytoczyć zależność współczynnika wzmocnienia prądowego α_N od prądu emitera (rys. 5.21). Początkowy wzrost współczynnika α_N



Rys. 5.21

Zależność współczynnika wzmocnienia prądowego od prądu emitera: a) $\alpha_N(I_E)$; b) $\beta_N(I_E)$

jest związany z coraz lepszą sprawnością wstrzykiwania emitera w miarę zwiększania prądu emitera, co dyskutowaliśmy już wcześniej. Wzmocnienie maleje w zakresie dużych prądów emitera wskutek zmniejszania się wartości współczynnika α_b , co można wytłumaczyć w następujący sposób. Duży ładunek nośników wstrzykiwanych do bazy wytwarza pole elektryczne kompensujące pole E_{wb} , a więc unoszenie nośników jest coraz słabsze i zwiększa się czas przelotu t_b , przeto maleje α_b (por. wzór 5.42)).

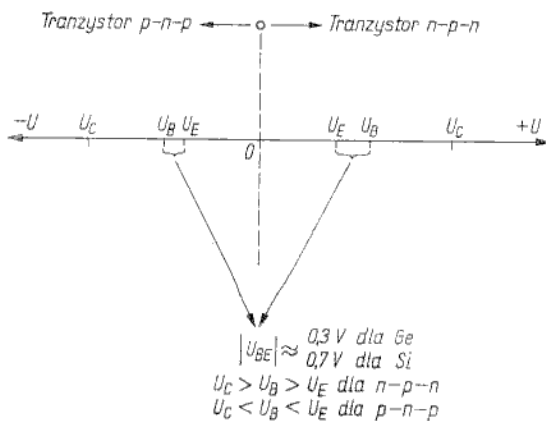
Zakresy pracy i układy włączenia tranzystora bipolarnego

5.6

Dotychczas rozważano zjawiska występujące wewnątrz tranzystora, a w tym i następnych punktach niniejszego rozdziału będziemy rozpatrywać charakterystyki i parametry końcówkowe, czyli tzw. właściwości funkcjonalne tranzy-

stora traktowanego jako element układu. Opis funkcjonalny (zestaw równań wiążących napięcia i prądy na końcówkach elementu) może w zasadzie abstrahować od zjawisk wewnętrznych, zachowując pełną użyteczność do potrzeb analizy układowej. Jednakże taki formalny model funkcjonalny („czarna skrzynka”, której właściwości są wprawdzie znane, lecz nie wiadomo, dlaczego są właśnie takie a nie inne) nie daje możliwości wykonania głębszej analizy pracy układu, czy szybkiego przewidywania „zachowania się” tranzystora w nietypowych warunkach pracy — słowem ogranicza bardzo możliwości świadomego i celowego rozwiązywania zadań projektowych. Dlatego przy opisie właściwości funkcjonalnych tranzystora będziemy dążyć do wykazania związków tych właściwości ze zjawiskami zachodzącymi wewnątrz tranzystora.

Z dotychczasowych rozważań wiadomo, że tranzystor spełnia funkcję elementu wzmacniającego przy polaryzacji złącza $E-B$ w kierunku przewodzenia, a złącza $B-C$ w kierunku zaporowym. Znamy również zasady polaryzacji złącza $p-n$, można



Rys. 5.22

Ogólne relacje między potencjałami poszczególnych elektrod tranzystora

zatem określić ogólne relacje między potencjałami poszczególnych elektrod tranzystora (rys. 5.22):

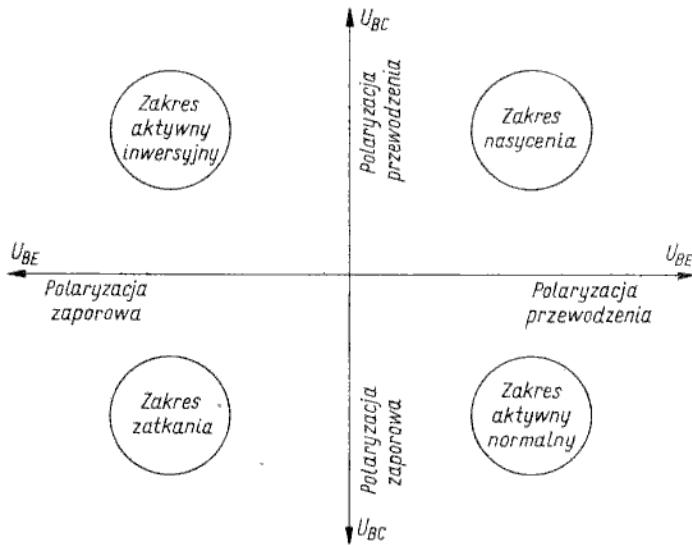
$$U_C > U_B > U_E \quad \text{dla tranzystora } n-p-n \quad (5.49a)$$

$$U_C < U_B < U_E \quad \text{dla tranzystora } p-n-p \quad (5.49b)$$

przy czym U_C , U_B , U_E — potencjały poszczególnych elektrod tranzystora: kolektora, bazy, emitera, względem jakiegoś potencjału wspólnego (najczęściej względem zerowego potencjału masy układu).

Należy przy tym zauważyć, że napięcie baza-emiter (napięcie na złączu $p-n$ spolaryzowanym w kierunku przewodzenia) w typowych warunkach pracy tranzystora przyjmuje wartości 0,2...0,4 V dla tranzystora germanowego oraz 0,6...0,8 V dla tranzystora krzemowego. W wielu układach jednak (przede wszystkim w układach impulsowych) tranzystor może pracować przy innych warunkach polaryzacji. Istnieją cztery warianty polaryzacji, można więc mówić o czterech zakresach pracy tranzystora (rys. 5.23):

- | | | |
|--|---|-------------------------|
| — złącze $E-B$ w kierunku przewodzenia | } | zakres aktywny normalny |
| — złącze $B-C$ w kierunku zaporowym | | |
| — złącze $E-B$ w kierunku przewodzenia | } | zakres nasycenia |
| — złącze $B-C$ w kierunku przewodzenia | | |



Rys. 5.23

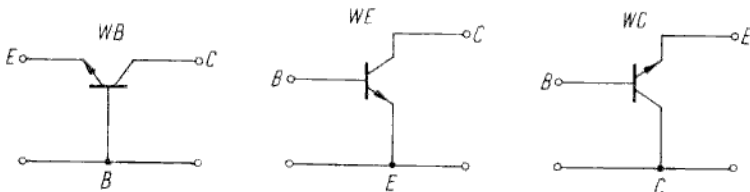
Cztery warianty polaryzacji i związane z nimi cztery zakresy pracy tranzystora

- złącze $E-B$ w kierunku zaporowym
 - złącze $B-C$ w kierunku zaporowym
 - złącze $E-B$ w kierunku zaporowym
 - złącze $B-C$ w kierunku przewodzenia
- } zakres zatkania
- } zakres aktywny inwersyjny (emiter spełnia funkcję kolektora, a kolektor — emitera)

Dla uproszczenia zakres aktywny normalny będziemy również nazywać zakresem normalnym, a zakres aktywny inwersyjny — zakresem inwersyjnym.

Tranzystor jako element trójkońcówkowy, traktowany jako czwórnik, musi mieć jedną z trzech końcówek wspólną dla wejścia i wyjścia. Jeżeli przyjmie się, że obwody wejściowy i wyjściowy mogą stanowić dowolne kombinacje dwu końcówek tranzystora (przykładowo wejście E, B , wyjście B, C — wspólna baza, lub wejście B, C , wyjście C, E — wspólny kolektor itp.), to wszystkich możliwych kombinacji jest $3! = 6$. Jednak dla uzyskania wzmocnienia mocy jest wymagane, by baza była jedną z końcówek wejściowych, a kolektor jedną z końcówek wyjściowych, co ogranicza liczbę racjonalnych konfiguracji tranzystora do trzech. Są to następujące konfiguracje (rys. 5.24):

- wejście E, B , wyjście B, C , czyli układ ze wspólną bazą, skrótowo oznaczany WB;
- wejście B, E , wyjście E, C , czyli układ ze wspólnym emiterym, skrótowo oznaczany WE;



Rys. 5.24

Trzy układy włączenia tranzystora

— wejście B , wyjście C , E , czyli układ ze wspólnym kolektorem, skrótowo oznaczany WC.

Aczkolwiek każdy z tych trzech układów włączenia ma właściwości różne od dwóch pozostałych (między innymi różne wzmocnienia prądu, napięcia i mocy), należy jednak pamiętać, że tranzystor działa zawsze jednakowo. Przypomnijmy najbardziej elementarny opis działania tranzystora $n-p-n$ w zakresie normalnym: „Elektrony są wstrzykiwane z emitera do bazy (prąd emitera) pod wpływem napięcia U_{BE} doprowadzonego w kierunku przewodzenia. Następnie nośniki wstrzykiwane do bazy przesuują się w stronę kolektora (unoszenie pod wpływem wbudowanego pola elektrycznego w bazie tranzystora dryftowego lub dyfuzja w tranzystorze z jednorodną bazą), przy czym znikoma ich część rekombinuje z dziurami tworząc prąd bazy, a reszta (przytłaczająca większość) jest odbierana przez silne pole elektryczne w warstwie zaporowej złącza kolektor-baza tworząc prąd kolektora”. Otóż ten schemat działania tranzystora obowiązuje niezależnie od układu włączenia. Różnice charakterystyk i parametrów w poszczególnych układach włączenia wynikają tylko z przyczyn subiektywnych, tj. są niejako skutkiem różnych „punktów widzenia”.

Prześledźmy to na przykładzie jednego parametru — wzmocnienia prądowego. Korzystamy z równania prądów i dwu wcześniej zdefiniowanych współczynników:

$$I_E = I_C + I_B$$

$$\alpha_N = \frac{I_C}{I_E}$$

$$\beta_N = \frac{I_C}{I_B}$$

W układzie WB prądem wejściowym jest I_E , wyjściowym zaś I_C , a ponieważ wzmocnienie prądowe jest to stosunek prądu wyjściowego do wejściowego, zatem

$$WP = \frac{I_C}{I_E} = \alpha_N$$

(WP-wzmocnienie prądowe).

W układzie WE prądem wejściowym jest I_B , wyjściowym zaś I_C , stąd

$$WP = \frac{I_C}{I_B} = \beta_N$$

W układzie WC prądem wejściowym jest I_B , wyjściowym zaś I_E , stąd

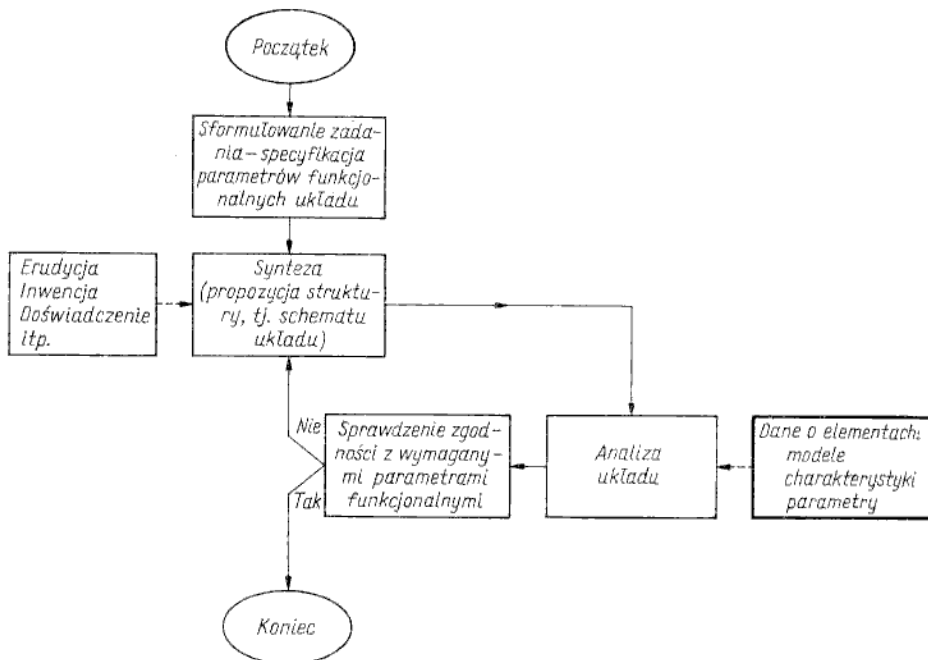
$$WP = \frac{I_E}{I_B} = \frac{I_C + I_B}{I_B} = \beta_N + 1$$

Należałoby jeszcze dodać, że α_N , β_N , $(\beta_N + 1)$ są tzw. zwarciovymi współczynnikami wzmocnienia prądowego, czyli są one mierzone w warunkach bliskich zwarcia na wyjściu tranzystora i rozwarcia na jego wejściu. Chodzi o to, by cały prąd wyjściowy płynął przez obciążenie przy pomijalnie małym wpływie bocznikującej konduktancji wyjściowej tranzystora oraz by cały prąd wejściowy płynął przez wejście tranzystora przy pomijalnie małym wpływie bocznikującej konduktancji wewnętrznej źródła sygnału wejściowego.

Praca nieliniowa statyczna (modele, charakterystyki, parametry)

5.7

W celu lepszego uświadomienia sobie znaczenia problematyki opisu funkcjonalnego (modelowania) elementu dla konstruktora układów rozpatrzmy ogólny schemat rozwiązywania dowolnego zadania projektowego (rys. 5.25). Algorytm



Rys. 5.25

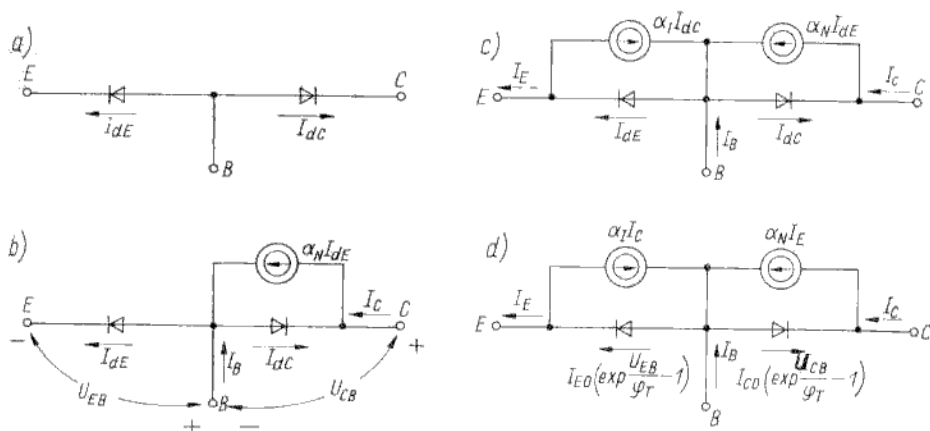
Ogólny schemat (algorytm) rozwiązywania dowolnego zadania projektowego

ma postać pętli zamkniętej. Pierwszym jest etap syntezy, w którym konstruktor decyduje się na określone rozwiązanie układu, spełniającej — jego zdaniem — wymagania postawione w sformułowaniu zadania. Występuje tu element „odgadnięcia” prawidłowego rozwiązania, wyboru jednego spośród wielu możliwych rozwiązań lub nawet twórczej propozycji rozwiązania dotychczas nieznanego, co ma oparcie w erudycji, inwencji, doświadczeniu itp. cechach konstruktora. Następnie wykonuje się analizę układu w celu sprawdzenia prawidłowości przyjętego rozwiązania. Do analizy konieczne są dane o elementach. W zależności od stopnia zaawansowania (dokładności) analizy elementy mogą być opisane modelami symbolicznymi, tj. schematami zastępczymi (dokładna analiza komputerowa), charakterystykami (uproszczona analiza graficzna) lub kilkoma podstawowymi parametrami (proste obliczenia szacunkowe). Dlatego dalej będą opisane te trzy sposoby opisu właściwości funkcjonalnych tranzystora. Podobnie jak dla złącza $p-n$ (por. p. 3.1.7) rozróżnia się pracę tranzystora nieliniową (statyczną i dynamiczną) oraz liniową (dla małych sygnałów małej i dużej częstotliwości). W tym punkcie zajmujemy się pracą nieliniową statyczną tranzystora, czyli interesują nas związki między napięciami a prądami stałymi na końcówkach tranzystora.

Modele nieliniowe statyczne

5.7.1

Ponieważ tranzystor bipolarny składa się z dwu złączy połączonych szeregowo przeciwnie ($n-p$, $p-n$ w tranzystorze $n-p-n$), zatem najprostszym modelem tranzystora wydaje się być połączenie dwu diod, pokazane na rys. 5.26a. Prądy płynące przez te diody są związane z napięciami według znanej zależności nieli-



Rys. 5.26

Modele nieliniowe statyczne: a) prosty model słuszny w pierwszym przybliżeniu tylko dla zakresu zatkania; b) model słuszny dla zakresu zatkania i normalnego; c) model słuszny dla wszystkich czterech zakresów pracy tranzystora (tzw. model Ebersa-Molla); d) inny zapis elementów modelu Ebersa-Molla

niowej (I_{dE} , I_{dC} oznaczają, że chodzi o prądy złączy emiterowego i kolektorowego, traktowanych jako niezależne diody):

$$I_{dE} = I_{ES} \left(\exp \frac{U_{EB}}{\varphi_T} - 1 \right) \quad (5.50)$$

$$I_{dC} = I_{CS} \left(\exp \frac{U_{CB}}{\varphi_T} - 1 \right) \quad (5.51)$$

przy czym: I_{ES} , I_{CS} — prądy nasycenia złączy $E-B$ oraz $B-C$; U_{EB} , U_{CB} — napięcia na złączach $E-B$ oraz $B-C$.

Jednak taki prosty model ma sens tylko wówczas, gdy wzajemne oddziaływanie złączy $p-n$ jest pomijalnie małe. Można w pierwszym przybliżeniu przyjąć, że taka sytuacja istnieje przy polaryzacji obu złączy w kierunku zaporowym, tj. przy polaryzacji tranzystora w zakresie zatkania. Jeżeli tranzystor pracuje w zakresie normalnym (złącze $E-B$ w kierunku przewodzenia, złącze $B-C$ w kierunku zaporowym), to należy uwzględnić, że przez złącze $B-C$ płynie nie tylko prąd wsteczny tego złącza, lecz również prąd nośników wstrzykiwanych przez złącze $E-B$, czyli prąd $\alpha_N I_{dE}$. W schemacie zastępczym uwzględnia się to zjawisko przez włączenie źródła prądowego $\alpha_N I_{dE}$ równolegle do diody $B-C$. Schemat zastępczy, przedstawiony na rys. 5.26b, jest zatem słuszny zarówno dla zakresu zatkania jak i normalnego. W przypadku pracy tranzystora w zakresie nasycenia oraz inwersyjnym kolektor również wstrzykuje nośniki do bazy, które wpływają na

wartość prądu płynącego w złączu $E-B$. W schemacie zastępczym uwzględnia się to zjawisko przez włączenie źródła $\alpha_I I_{dC}$ równoległe do diody $E-B$.

Dla odróżnienia wzmocnienia w kierunku normalnym (od emitera do kolektora) od wzmocnienia w kierunku inwersyjnym (od kolektora do emitera) stosuje się oznaczenia: α_N dla współczynnika wzmocnienia normalnego, α_I dla współczynnika wzmocnienia inwersyjnego. Nie oznacza to bynajmniej, że tranzystor działa całkowicie symetrycznie, a kolektor i emiter są końcówkami w pełni zamienialnymi. Wzmocnienie prądowe w kierunku inwersyjnym jest wiele razy mniejsze niż w kierunku normalnym. Na przykład dla tranzystora BC 107 w typowym przypadku $\beta_N = \alpha_N / (1 - \alpha_N) = 200$, a $\beta_I = \alpha_I / (1 - \alpha_I) = 4$. Schemat zastępczy, przedstawiony na rys. 5.26c, znany jako model Ebersa-Molla¹⁾, jest słuszny dla wszystkich czterech zakresów pracy tranzystora. Na podstawie tego modelu prądy płynące przez końcówki tranzystora są określone następującymi zależnościami:

$$I_E = I_{ES} \left(\exp \frac{U_{EB}}{\varphi_T} - 1 \right) - \alpha_I I_{CS} \left(\exp \frac{U_{CB}}{\varphi_T} - 1 \right) \quad (5.52)$$

$$I_C = -I_{CS} \left(\exp \frac{U_{CB}}{\varphi_T} - 1 \right) + \alpha_N I_{ES} \left(\exp \frac{U_{EB}}{\varphi_T} - 1 \right) \quad (5.53)$$

lub (rys. 5.26d):

$$I_E = I_{E0} \left(\exp \frac{U_{EB}}{\varphi_T} - 1 \right) - \alpha_I I_C \quad (5.54)$$

$$I_C = -I_{C0} \left(\exp \frac{U_{CB}}{\varphi_T} - 1 \right) + \alpha_N I_E \quad (5.55)$$

przy czym:

$$I_B = I_E - I_C$$

$$I_{E0} = I_{ES}(1 - \alpha_I \alpha_N); \quad I_{C0} = I_{CS}(1 - \alpha_I \alpha_N)$$

Teraz zostanie wyjaśniony przyjęty sposób znakowania napięć i prądów. Napięcia U_{EB} , U_{CB} podstawia się ze znakiem dodatnim do równań (5.52) do (5.54), jeżeli złącza są polaryzowane w kierunku przewodzenia. Przykładowo dla pracy w zakresie normalnym U_{EB} ma znak dodatni, a U_{CB} — ujemny. Kierunki prądów oznaczono zgodnie z konwencją przyjętą w elektrotechnice (od plusa do minusa). Zależne źródła prądowe $\alpha_N I_{dE}$, $\alpha_I I_{dC}$ mają kierunki zgodne z odpowiednimi prądami I_{dE} , I_{dC} . Kierunki prądów zaznaczone na rys. 5.26 odnoszą się do tranzystora $n-p-n$.

Szczególnie godne podkreślenia jest to, że w modelu Ebersa-Molla wszystkie zależności prądowo-napięciowe można wyznaczyć znając tylko cztery łatwe do zmierzenia parametry: I_{ES} , I_{CS} , α_N , α_I .

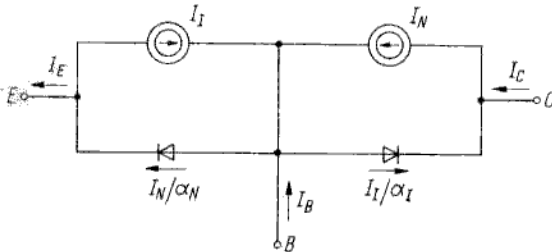
Ponieważ Ebers i Moll udowodnili również, że istnieje zależność

$$\alpha_N I_{ES} = \alpha_I I_{CS} \quad (5.56)$$

¹⁾ Ebers J. J., Moll J. L.: Large-signal behaviour of junction transistors. *Proc. of IRE*, 1954, p. 1761.

przezo liczbę parametrów wyjściowych można nawet zmniejszyć do trzech. Należy jednak pamiętać, że równanie (5.56) nie jest słuszne, jeżeli uwzględni się zależność współczynników α_N , α_I od prądu emitera i kolektora.

Model Ebersa-Molla jest obecnie powszechnie stosowany w dwu wariantach. Pierwszy wariant, wyżej omówiony, charakteryzuje się tym, że jako zmienne niezależne są wybrane prądy wstrzykiwane przez emiter i kolektor. Taki model jest nazywany *iniekeyjnym*. W drugim wariancie za zmienne niezależne przyjmuje się prądy zbierane przez kolektor i emiter. Taki model jest nazywany *transportowym* (rys. 5.27). Nie są to dwa różne modele, lecz tylko dwa warianty zapisu



Rys. 5.27
Model Ebersa-Molla — wariant transportowy

tych samych zależności. Równaniom (5.52), (5.53) dla modelu iniekeyjnego są równoważne następujące równania dla modelu transportowego:

$$I_E = \frac{I_N}{\alpha_N} - I_I \quad (5.57)$$

$$I_C = I_N - \frac{I_I}{\alpha_I} \quad (5.58)$$

przy czym:

$$I_N = \alpha_N I_{ES} \left(\exp \frac{U_{EB}}{\varphi_T} - 1 \right) \quad (5.59)$$

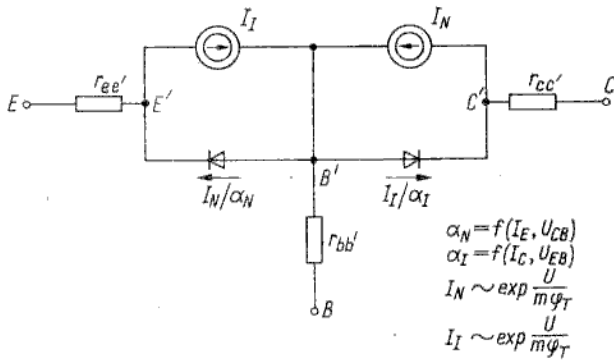
$$I_I = \alpha_I I_{CS} \left(\exp \frac{U_{CB}}{\varphi_T} - 1 \right) \quad (5.60)$$

Jeżeli w pierwszym przybliżeniu przyjmie się, że zależność (5.56) jest słuszna, to oba źródła odniesienia I_N , I_I są określone identyczną charakterystyką prądowo-napięciową. Jest to istotną zaletą modelu transportowego.

Dotychczas rozpatrywaliśmy model uproszczony, którego dokładność można zwiększyć przez uwzględnienie następujących zjawisk:

- zależność współczynników α_N , α_I od prądu emitera i kolektora oraz od napięć polaryzacji obu złączy;
- istnienie rezystancji szeregowych emitera, bazy, kolektora (są to rezystancje doprowadzeń i obszarów leżących poza warstwami zaporowymi);
- korekta zależności wykładniczych przez wprowadzenie współczynnika $m \neq 1$ $\left[I \sim \left(\exp \frac{U}{m\varphi_T} - 1 \right) \right]$.

Na rysunku 5.28 przedstawiono model Ebersa-Molla w wersji transportowej, w którym uwzględniono wymienione czynniki. Największe znaczenie ma rezystancja rozproszona bazy r'_{bb} , pozostałe rezystancje są zwykle pomijalnie małe.



Rys. 5.28
Model Ebersa-Molla o zwiększonej dokładności

Charakterystyki statyczne

5.7.2

Stan statyczny (punkt pracy) tranzystora traktowanego jako czwórnik nieliniowy jest opisany czterema wielkościami, tj. prądem i napięciem wejściowym I_1, U_1 oraz prądem i napięciem wyjściowym I_2, U_2 . W ogólnym przypadku zmiana każdej z tych wielkości powoduje zmiany trzech pozostałych. Zawsze istnieje możliwość wyboru dwóch wielkości jako zmienne niezależne i obserwowania ich wpływu na dwie pozostałe wielkości. Zależność jednej wielkości od dwu innych przy niekontrolowanej wielkości czwartej nazywa się *równaniem czwórnika*. Łatwo można obliczyć, że istnieje 12 różnych równań. Ponieważ do pełnego opisu czwórnika wystarczają 2 równania (para równań czwórnika), można je pogrupować w 6 par. Znaczenie praktyczne mają 3 pary równań: równania impedancyjne

$$\left. \begin{aligned} U_1 &= f(I_1, I_2) \\ U_2 &= f(I_1, I_2) \end{aligned} \right\} \quad (5.61)$$

równania admitancyjne

$$\left. \begin{aligned} I_1 &= f(U_1, U_2) \\ I_2 &= f(U_1, U_2) \end{aligned} \right\} \quad (5.62)$$

równania mieszane

$$\left. \begin{aligned} U_1 &= f(I_1, U_2) \\ I_2 &= f(I_1, U_2) \end{aligned} \right\} \quad (5.63)$$

Najbardziej dogodny jest zestaw równań mieszanych, gdyż warunki ich pomiaru są zbliżone do normalnych warunków pracy tranzystora w układzie (warunki zbliżone do sterowania prądowego na wejściu i sterowania napięciowego na wyjściu). Dlatego rozpatrzy się charakterystyki statyczne, określone na podstawie równań mieszanych. Przez *charakterystykę statyczną* rozumie się związek wielkości zależnej i jednej z dwu niezależnych przy stałej wartości drugiej wielkości niezależnej traktowanej jako parametr. Na podstawie równań (5.63) można określić cztery rodziny charakterystyk statycznych:

charakterystyki wejściowe	$U_1 = f(I_1)_{U_2 = \text{const}}$	(5.64)
charakterystyki zwrotne napięciowe	$U_1 = f(U_2)_{I_2 = \text{const}}$	
charakterystyki przejściowe prądowe	$I_2 = f(I_1)_{U_2 = \text{const}}$	
charakterystyki wyjściowe	$I_2 = f(U_2)_{I_1 = \text{const}}$	

Dowolna kombinacja jednej z dwu pierwszych charakterystyk (wejściowej lub zwrotnej) i jednej z dwu ostatnich (przejściowej lub wyjściowej) jest wystarczająca do wyznaczenia dwu pozostałych charakterystyk. Dlatego w katalogach są najczęściej podawane tylko charakterystyki wyjściowe i wejściowe.

Dla każdego z trzech układów włączenia tranzystora wielkości I_1, I_2, U_1, U_2 oznaczają zupełnie różne prądy i napięcia (na przykład I_1 jest prądem emitera w układzie WB, a prądem bazy w układzie WE). Należałoby zatem rozpatrzyć charakterystyki statyczne oddzielnie dla każdej konfiguracji. Zrobimy to dla konfiguracji WB i WE. Konfigurację WC można pominąć, jeżeli weźmie się pod uwagę znikome różnice (szczególnie w obrazie graficznym) między prądami emitera i kolektora. Interesuje nas głównie praca tranzystora w zakresie normalnym.

Charakterystyki statyczne w układzie WB

5.7.2.1

Zgodnie z układem włączenia tranzystora:

$$I_1 = I_E$$

$$I_2 = I_C$$

$$U_1 = U_{EB}$$

$$U_2 = U_{CB}$$

Interesują nas zatem następujące rodziny charakterystyk:

$$U_{EB} = f(I_E, U_{CB}) \begin{cases} U_{EB} = f(I_E)_{U_{CB}} & \text{— wejściowa} \\ U_{EB} = f(U_{CB})_{I_E} & \text{— zwrotna} \end{cases}$$

$$I_C = f(I_E, U_{CB}) \begin{cases} I_C = f(I_E)_{U_{CB}} & \text{— przejściowa} \\ I_C = f(U_{CB})_{I_E} & \text{— wyjściowa} \end{cases}$$

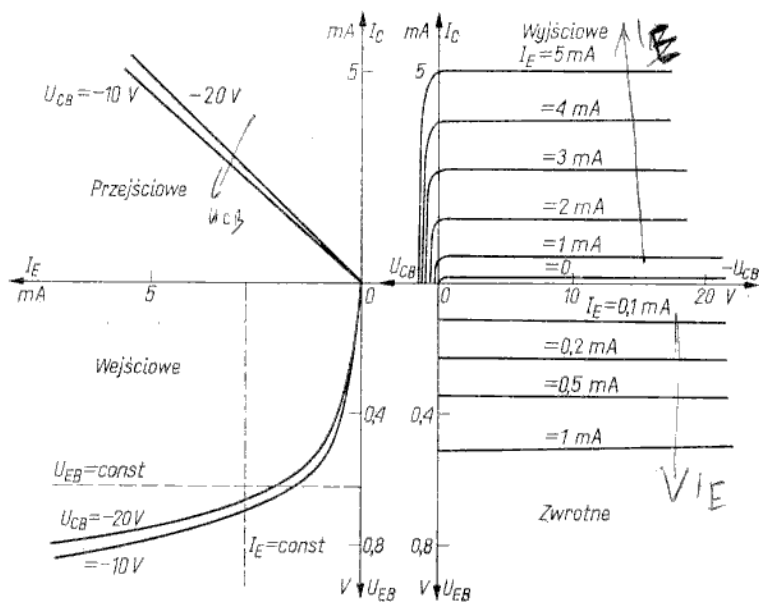
Wiadomo, że złącze E-B jest polaryzowane w kierunku przewodzenia (U_{EB} dodatnie), a złącze B-C w kierunku zaporowym (U_{CB} ujemne).

Wszystkie cztery rodziny charakterystyk przedstawiono na rys. 5.29. Charakterystyki mają sens ogólny, chociaż dla lepszej orientacji wyskalowano osie prądów i napięć w wartościach typowych dla tranzystorów małej mocy.

Charakterystyki wejściowe

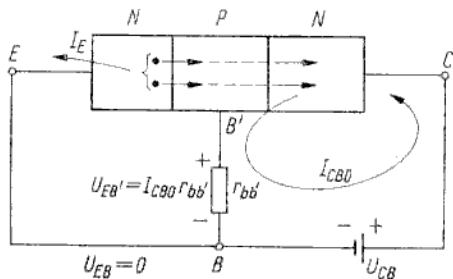
Postać analityczną zależności $U_{EB}(I_E)_{U_{CB}}$ można łatwo otrzymać na podstawie równania (5.52). Jest to zależność logarytmiczna, gdyż w przybliżeniu $I_E \sim \exp \frac{U_{EB}}{\varphi_T}$.

Gdyby wyeksponować niewielki obszar charakterystyk przy początku układu współrzędnych, to okazałoby się, że charakterystyki nie przechodzą dokładnie przez zero



Rys. 5.29
Cztery rodziny charakterystyk statycznych w układzie WB

(dla $U_{EB} = 0$ prąd $I_E > 0$ przy $U_{CB} < 0$). Jest to zjawisko drugorzędne, pominięte na wykresie, które jednakże wynika wprost z równania (5.52). Można je również w prosty sposób wyjaśnić jakościowo, jeżeli rozpatrzy się wpływ rezystancji rozproszonej bazy $r_{bb'}$. Na rysunku 5.30 wykazano, w jaki

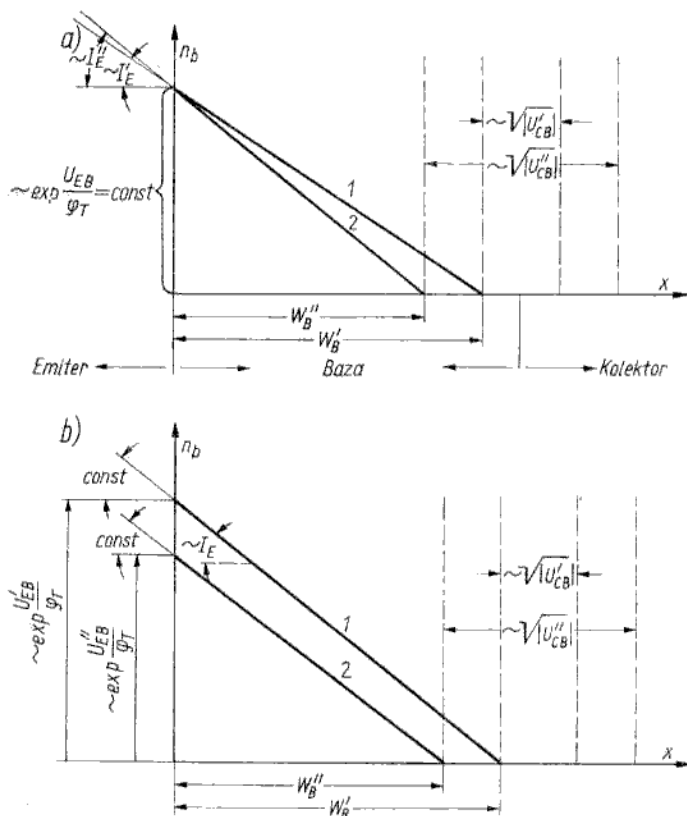


Rys. 5.30
Ilustracja wyjaśniająca wpływ skończonej rezystancji obszaru bazy (rezystancji rozproszonej bazy $r_{bb'}$) na prąd emitra przy $U_{EB} = 0$

sposób prąd I_{CBO} płynący w obwodzie $B-C$ wywołuje spadek napięcia na rezystancji $r_{bb'}$, który z kolei polaryzuje efektywnie złącze $E-B'$ w kierunku przewodzenia i powoduje przepływ pewnego prądu I_E mimo braku zewnętrznego źródła polaryzacji U_{EB} . Zmiany charakterystyk $U_{EB}(I_E)$ dla różnych wartości parametru U_{CB} są spowodowane zjawiskiem modulacji efektywnej szerokości bazy, nazywanym często zjawiskiem *Early'ego*¹⁾. To zjawisko znacznie silniej przejawia się

¹⁾ J. M. Early: Effects of space-charge layer widening in junction transistors. *Proc. IRE*, 1952, Vol. 40, pp. 1401—1406.

w tranzystorze z jednorodną bazą niż w tranzystorze dryftowym, dlatego wyjaśnimy je najpierw na przykładzie tego pierwszego tranzystora. Na rysunku 5.31a przedstawiono rozkłady koncentracji nośników nadmiarowych w bazie tranzystora bezdryftowego dla dwu różnych wartości napięcia baza-kolektor, tj. U'_{CB}



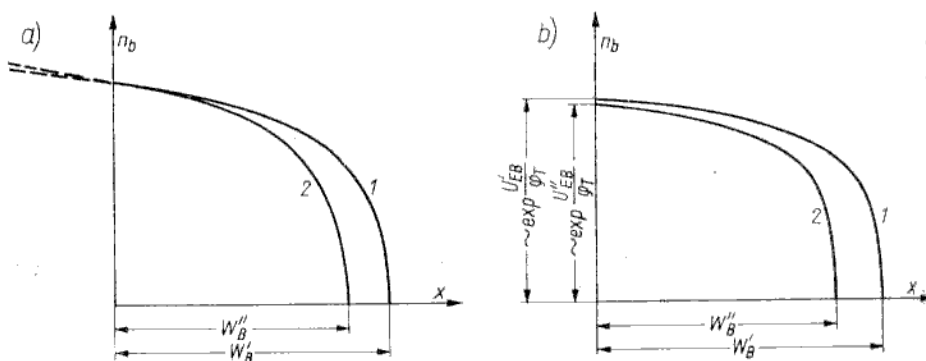
Rys. 5.31

Wpływ modulacji efektywnej szerokości bazy (zjawiska Early'ego) na pracę tranzystora bezdryftowego: a) zmiany prądu I_E przy $U_{EB} = \text{const}$; b) zmiany napięcia U_{EB} przy $I_E = \text{const}$

i U''_{CB} , przy czym $|U''_{CB}| > |U'_{CB}|$. Rozkłady koncentracji elektronów w bazie, oznaczone na rys. 5.31a cyframi 1, 2, wykreślono przy założeniu, że $U_{EB} = \text{const}$, co oznacza stałość koncentracji elektronów na początku bazy. Na końcu bazy (na krawędzi warstwy zaporowej złącza baza-kolektor) koncentracja elektronów jest równa zero (silne „wymiatanie” elektronów). Ponieważ szerokość warstwy zaporowej złącza B-C zmienia się w funkcji napięcia U_{CB} , zmienia się również efektywna szerokość bazy. Powoduje to oczywiście zmiany gradientu koncentracji nośników, a więc zmienia się prąd emitera, który jest prądem dyfuzji proporcjonalnym do gradientu koncentracji nośników. Zgodnie z rys. 5.31a wzrost napięcia $|U_{CB}|$ powoduje zmniejszenie efektywnej szerokości bazy, przeto wzrasta gradient koncentracji elektronów (przy stałym napięciu U_{EB}), czyli wzrasta prąd emitera. Jeżeli zatem rodzinę charakterystyk wejściowych (rys. 5.29)

przetnie się linią stałego napięcia ($U_{EB} = \text{const}$), to większym napięciom $|U_{CB}|$ będą odpowiadały większe wartości prądu I_E . Można by też uczynić odwrotnie, tj. przeciąć rodzinę charakterystyk wejściowych linią stałego prądu ($I_E = \text{const}$). Wówczas należałoby udowodnić, że większym wartościom napięcia $|U_{CB}|$ odpowiadają mniejsze wartości napięcia U_{EB} . Na rysunku 5.31b wykazano, że tak jest w istocie. Interpretacja zjawiska jest identyczna jak na rys. 5.31a. Zmienia się tylko jeden warunek brzegowy, a mianowicie zamiast $U_{EB} = \text{const}$, co oznacza $n_b(0) = \text{const}$, jest $I_E = \text{const}$, co oznacza $\left. \frac{dn_b}{dx} \right|_{x=0} = \text{const}$, czyli stałe nachylenie rozkładu koncentracji elektronów w bazie. Widać, że napięcie U_{EB} maleje w miarę wzrostu wartości napięcia $|U_{CB}|$.

Uogólniając powyższe rozważania można stwierdzić, że zjawisko modulacji efektywnej szerokości bazy powoduje zmiany wielkości wejściowych (I_E, U_{EB}) wskutek zmiany napięcia wyjściowego (U_{CB}), przy czym to oddziaływanie zwrotne jest związane z dyfuzyjnym charakterem transportu nośników w bazie. Stąd wynika wniosek, że w wyidealizowanym tranzystorze dryftowym (transport nośników w bazie tylko pod wpływem unoszenia) nie ma oddziaływania zwrotnego, czyli charakterystyki wejściowe nie zależą od napięcia U_{CB} . W rzeczywistym tranzystorze dryftowym istnieje niewielka składowa prądu dyfuzji obok dominującej składowej prądu unoszenia (chodzi o relację między tymi dwiema składowymi prądu na początku bazy, tzn. dla $x = 0$), dlatego pewne oddziaływanie zwrotne obserwuje się również i w tych tranzystorach. Jest ono jednak około rząd wartości słabsze niż w tranzystorach bezdryftowych, co łatwo można zrozumieć porównując rozkłady koncentracji na rys. 5.31a, b (tranzystor bezdryftowy) i na rys. 5.32a, b (tranzystor dryftowy).



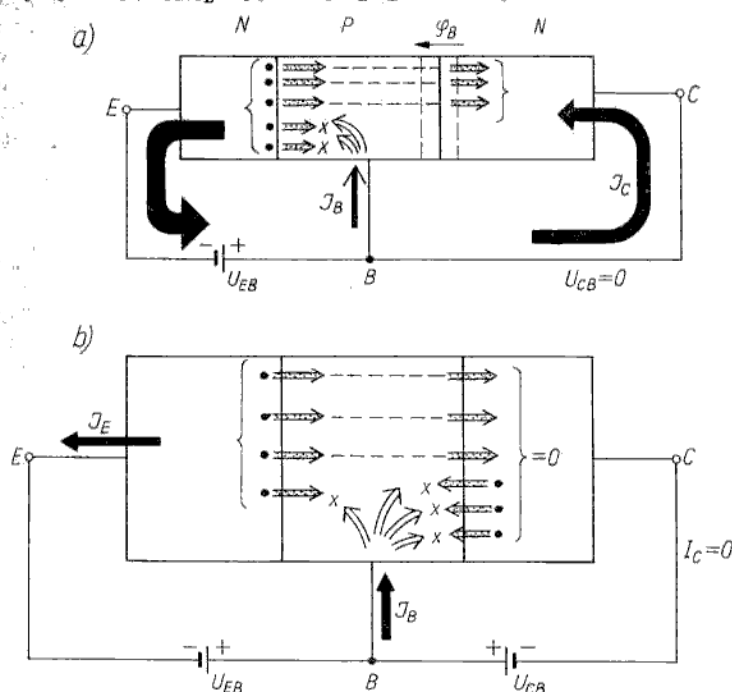
Rys. 5.32

Wpływ modulacji efektywnej szerokości bazy (zjawiska Early'ego) na pracę tranzystora dryftowego: a) zmiany prądu I_E przy $U_{EB} = \text{const}$ (nieznacząca zmiana nachylenia dn_b/dx dla $x = 0$); b) zmiany napięcia U_{EB} przy $I_E = \text{const}$ (nieznacząca zmiana wartości $n_b(0)$)

Charakterystyki wyjściowe

Postać analityczną zależności $I_C(U_{CB})_{I_E}$ można otrzymać na podstawie równania (5.53). Ponieważ parametrem jest prąd emitera, dogodnie jest zatem posłużyć się zależnością (5.55). Jeżeli $I_E \gg I_{CS}$, to w zakresie normalnym ($U_{CB} < 0$) prąd I_C , praktycznie biorąc, nie zależy od napięcia U_{CB} i jest liniowo zależny od prądu I_E (jeżeli $\alpha_N = \text{const}$) $I_C = \alpha_N I_E$.

Dopiero w zakresie nasycenia ($U_{CB} > 0$) pierwszy składnik równania (5.53) kompensuje drugi, czyli prąd I_C maleje w miarę wzrostu $U_{CB} > 0$. Taki przebieg charakterystyki $I_C(U_{CB})_{I_E}$ wyjaśniają poglądowo rys. 5.33a, b.



Rys. 5.33

Ilustracje objaśniające zjawiska przy $U_{CB} = 0$ (a) oraz $I_C = 0$ (b)

Na rysunku 5.33a przedstawiono sytuację dla $U_{CB} = 0$, tj. dla zwarcia kolektora z bazą.

Nośniki wstrzykiwane z emitera do bazy mogą być odbierane przez kolektor, gdyż mimo braku zewnętrznego napięcia U_{CB} istnieje napięcie dyfuzyjne φ_B w warstwie zaporowej złącza B-C, czyli jest pole elektryczne „wymiatające” elektrony z bazy do kolektora. Dlatego prąd $I_C \neq 0$.

Na rysunku 5.33b przedstawiono sytuację dla $I_C = 0$. W tym przypadku złącze C-B jest polaryzowane w kierunku przewodzenia, przy czym warstwa kolektorowa spełnia jednocześnie funkcję kolektora (odbiera część nośników wstrzykiwanych z emitera do bazy) i emitera (wstrzykuje nośniki do bazy). Prąd $I_C = 0$, gdy liczby nośników przechodzących przez warstwę zaporową złącza B-C w obu kierunkach są jednakowe. Aby był spełniony warunek obojętności elektrycznej bazy, prądy emitera i bazy muszą być przy tym jednakowe.

W dokładniejszych rozważaniach charakterystyk wyjściowych okazuje się, że charakterystyka $I_C(U_{CB})$ nie jest idealnie płaska w zakresie normalnym, lecz ma niewielkie nachylenie dodatnie. Oznacza to, że w miarę zwiększania napięcia $|U_{CB}|$ prąd I_C nieznacznie wzrasta. Jest to spowodowane modulacją efektywnej szerokości bazy. W miarę wzrostu $|U_{CB}|$ efektywna szerokość bazy maleje, a im cieńsza jest baza, tym mniej nośników nadmiarowych w niej rekombinuje. Zwiększa się zatem wartość współczynnika transportu α_b , a więc wzrasta α_N , czyli przy stałej wartości prądu I_E zwiększa się nieznacznie prąd I_C .

Charakterystyki przejściowe

Zależność $I_C(I_E)_{U_{CB}}$ jest określona analitycznie równaniem (5.55). Przy $U_{CB} < 0$ pierwszy człon tego wyrażenia jest pomijalnie mały, zatem $I_C = f(I_E)$ jest zależnością liniową o nachyleniu α_N . W dokładniejszych rozważaniach należy uwzględnić niewielką nieliniowość spowodowaną zmianami współczynnika α_N w funkcji prądu emitera (rys. 5.21a). Interpretacja wpływu parametru U_{CB} na nachylenie prostej $I_C = \alpha_N I_E$ jest oparta na zjawisku modulacji efektywnej szerokości bazy i nie różni się od wyżej przedstawionego wyjaśnienia nachylenia charakterystyki $I_C(U_{CB})$.

Charakterystyki zwrotne

Na podstawie uproszczonego modelu tranzystora zależność $U_{EB}(U_{CB})_{I_E}$ powinna być linią prostą równoległą do osi U_{CB} (brak wpływu napięcia U_{CB} na U_{EB}). Wiadomo jednak, że modulacja efektywnej szerokości bazy jest przyczyną pewnego oddziaływania zwrotnego. Zmniejszanie się napięcia U_{EB} w miarę wzrostu U_{CB} przy stałym prądzie I_E wyjaśniono już na rys. 5.31a.

Charakterystyki statyczne w układzie WE

5.7.2.2

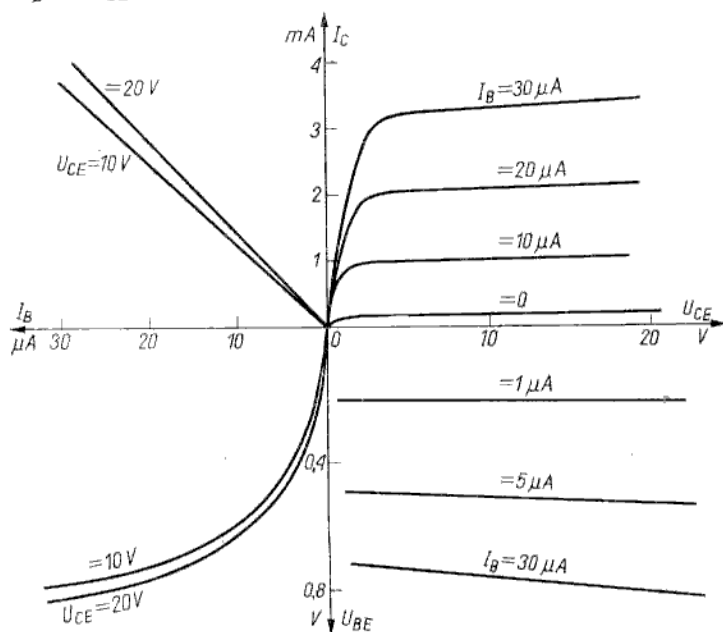
Zgodnie z układem włączenia tranzystora:

$$I_1 = I_B$$

$$I_2 = I_C$$

$$U_1 = U_{BE}$$

$$U_2 = U_{CE}$$



Rys. 5.34

Cztery rodziny charakterystyk statycznych w układzie WE

Interesują nas zatem rodziny charakterystyk:

$$\begin{aligned}
 U_{BE} = f(I_B, U_{CE}) & \begin{cases} U_{BE} = f(I_B)_{U_{CE}} & \text{— wejściowa} \\ U_{BE} = f(U_{CE})_{I_B} & \text{— zwrotna} \end{cases} \\
 I_C = f(I_B, U_{CE}) & \begin{cases} I_C = f(I_B)_{U_{CE}} & \text{— przejściowa} \\ I_C = f(U_{CE})_{I_B} & \text{— wyjściowa} \end{cases}
 \end{aligned}$$

Wszystkie cztery rodziny charakterystyk przedstawiono na rys. 5.34.

Napięcie U_{CE} polaryzuje oba złącza $E-B$ i $B-C$, dlatego nie można w tym przypadku zachować dotychczas stosowanego sposobu znakowania napięć. (Dotychczas przyjmowaliśmy, że napięcie polaryzujące złącze w kierunku przewodzenia jest dodatnie, a w kierunku zaporowym — ujemne). Przyjmijmy w tym przypadku rzeczywiste znaki potencjałów bazy (U_{BE}) i kolektora (U_{CE}) względem potencjału emitera. Dla tranzystora $n-p-n$ pracującego w zakresie normalnym oba napięcia są dodatnie w tej konwencji. Charakterystyki statyczne można obliczać z równań (5.52), (5.53), (5.54), (5.55) uwzględniając, że:

$$\begin{cases} |U_{EB}| = |U_{BE}| \\ |U_{CB}| = |U_{CE} - U_{BE}| \end{cases} \quad (5.65)$$

Do równań podstawia się:

- U_{EB} ze znakiem dodatnim, gdyż złącze $E-B$ jest polaryzowane w kierunku przewodzenia;
- U_{CB} ze znakiem ujemnym, jeżeli $|U_{CE}| > |U_{BE}|$ (wówczas złącze $B-C$ jest polaryzowane w kierunku zaporowym i tranzystor pracuje w zakresie normalnym);
- U_{CB} ze znakiem dodatnim, jeżeli $|U_{CE}| < |U_{BE}|$ (wówczas złącze $B-C$ jest polaryzowane w kierunku przewodzenia i tranzystor pracuje w zakresie nasycenia).

Charakterystyki wejściowe

Zależność $U_{BE}(I_B)_{U_{CE}}$ można znaleźć analitycznie odejmując (5.54) od (5.55). Ponieważ

$$I_B = \frac{I_E}{\beta_N + 1}$$

oraz

$$\beta_N = \frac{\alpha_N}{1 - \alpha_N}$$

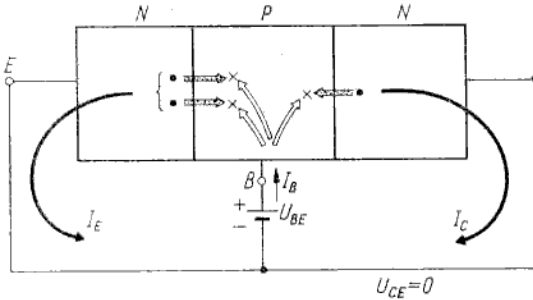
zatem, zakładając $\alpha_N = \text{const}$, można stwierdzić, że charakterystyka wejściowa w układzie WE ma taki sam przebieg logarytmiczny jak w układzie WB (prądy I_E , I_B są liniowo zależne).

Zmiany charakterystyki wejściowej w funkcji napięcia U_{CE} są spowodowane modulacją efektywnej szerokości bazy. W miarę wzrostu U_{CE} (prawie całe napięcie U_{CE} polaryzuje zaporowo złącze $C-B$, a tylko niewielka jego część odkłada się na złączu $E-B$ w kierunku przewodzenia) maleje efektywna szerokość bazy. Jeżeli przyjmie się na przykład, że poziom wstrzykiwania nośników z emitera do bazy jest stały ($U_{BE} = \text{const}$), to przy cieńszej bazie mniej nośników rekombinuje, czyli prąd I_B maleje. Taki kierunek zmian charakterystyk $U_{BE}(I_B)$ przy różnych wartościach U_{CE} reprezentują dwie krzywe przedstawione na rys. 5.34. Gdyby wyeksponować niewielki początkowy obszar charakterystyk, to okazałoby się, że przy $U_{CE} > 0$ charakterystyki nie przechodzą dokładnie przez początek układu współrzędnych, a mianowicie przy $I_B = 0$, $U_{BE} > 0$. Można to wyjaśnić w ten sposób,

że wskutek generacji par elektron-dziura w bazie elektrony są usuwane przez złącze baza-kolektor, a dziury tworzą ładunek przestrzenny w bazie polaryzując nieznacznie złącze $E-B$ w kierunku przewodzenia.

Charakterystyki wyjściowe

W pierwszym przybliżeniu można przyjąć, że charakterystyki $I_C(U_{CE})_{I_B}$ wychodzą z początku układu współrzędnych. W rzeczywistości $I_C = 0$ przy napięciu U_{CE} dodatnim, a przy $U_{CE} = 0$ płynie prąd I_C o kierunku przeciwnym do normalnego i wartości proporcjonalnej do prądu I_B . Taki przebieg charakterystyk „przy zerze” jest zrozumiały, jeśli zauważy się (rys. 5.35), że przy $U_{CE} = 0$



Rys. 5.35

Ilustracja objaśniająca zjawiska zachodzące przy $U_{CE} = 0$

oraz $I_B > 0$ złącza $E-B$, $B-C$ są polaryzowane w kierunku przewodzenia, czyli z emitera i kolektora są wstrzykiwane elektrony, które rekombinują z dziurami dostarczonymi z obwodu bazy. Ponieważ w tym przypadku $I_B = I_E + I_C$, a prąd I_B jest mały, można przyjąć, że prąd I_C przy $U_{CE} = 0$ jest, praktycznie biorąc, równy zeru. Tylko w niektórych zastosowaniach tranzystorów (na przykład tranzystor pracujący jako klucz dwukierunkowy) resztkowy prąd kolektora (kilkadziesiąt μA) przy $U_{CE} = 0$ lub resztkowe napięcie U_{CE} (kilka mV) przy $I_C = 0$ może mieć istotne znaczenie.

W miarę wzrostu napięcia U_{CE} tranzystor „wychodzi” ze stanu nasycenia (złącze $B-C$ jest coraz słabiej polaryzowane w kierunku przewodzenia, gdyż maleje $|U_{CB}| = |U_{CE} - U_{BE}|$) i przy $U_{CE} \geq U_{BE}$ pracuje w obszarze normalnym. Ten zakres pracy charakteryzuje się niemal stałą wartością prądu kolektora, który jest wprost proporcjonalny do prądu bazy ($I_C = \beta_N I_B$). Jednak obserwuje się pewien wzrost prądu I_C w miarę zwiększania napięcia U_{CE} , przy czym nachylenie charakterystyki $I_C(U_{CE})$ jest nawet znacznie większe, niż to było dla charakterystyki $I_C(U_{CB})$ w układzie WB. Przyczyna tego nachylenia jest identyczna jak dla układu WB — modulacja efektywnej szerokości bazy. Ponieważ w tym przypadku niewielka część napięcia U_{CE} odkłada się również na złączu $E-B$ w kierunku przewodzenia, w miarę zwiększania napięcia U_{CE} nie tylko maleje efektywna szerokość bazy, lecz również zwiększa się wstrzykiwanie nośników z emitera. Dlatego nachylenie charakterystyki wyjściowej w układzie WE jest większe niż w układzie WB.

Charakterystyka przejściowa

W pierwszym przybliżeniu jest to liniowa zależność $I_C = \beta_N I_B$. Napięcie U_{CE} wpływa na tę zależność poprzez modulację efektywnej szerokości bazy. W dokładniejszych rozważaniach należy uwzględnić zależność współczynnika β_N od prądu I_C (rys. 5.21b).

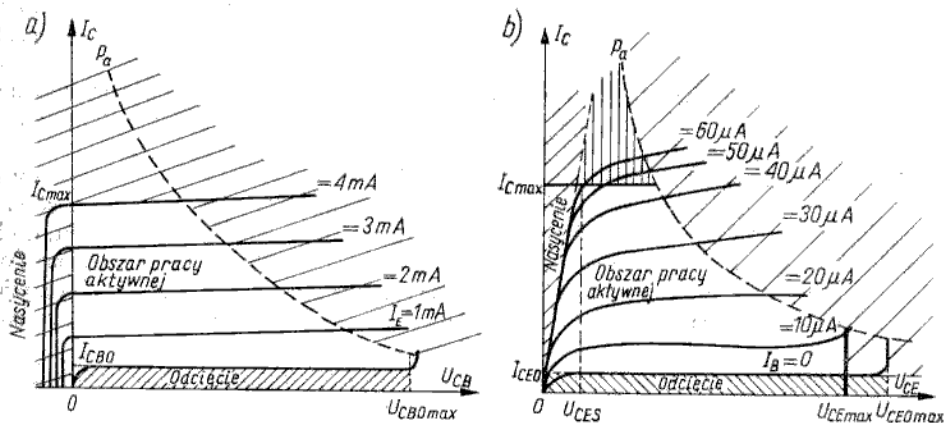
Charakterystyka zwrotna

Oddziaływanie zwrotne w układzie WE jest silniejsze niż w układzie WB, gdyż — jak stwierdzono wyżej — część napięcia U_{CE} polaryzuje złącze E-B.

Parametry statyczne

5.7.3

Większość parametrów statycznych są to wielkości ograniczające dozwolony obszar pracy aktywnej tranzystora. Interesujące jest pokazanie tych ograniczeń na polu charakterystyk wyjściowych, co uczyniono na rys. 5.36 dla układów WB i WE.



Rys. 5.36

Ograniczenia dozwolonego obszaru pracy aktywnej tranzystora, pokazane na polu charakterystyk wyjściowych dla układu WB (a) oraz WE (b)

Wiadomo już z poprzednich rozważań, że zakres aktywny jest ograniczony „od dołu” odcięciem, a „z lewej strony” — nasyceniem. Teraz wprowadzamy trzy nowe ograniczenia (dopuszczalna moc admysyjna, prąd i napięcie maksymalne), które łącznie z dwoma poprzednimi określają dozwolony obszar pracy aktywnej tranzystora.

Mozna zatem wymienić następujące istotne parametry statyczne:

- moc admysyjna P_a ,
- prąd maksymalny I_{Cmax} ,
- prąd zerowy (granica między zakresem aktywnym i odcięciem),
- napięcie maksymalne,
- napięcie nasycenia $U_{CE(sat)}$ (tylko dla układu WE).

Parametrem statycznym jest również współczynnik wzmocnienia prądowego α_N lub β_N , który omówiono wcześniej.

Moc admysyjna P_a określa maksymalną wartość iloczynu prądu przez napięcie stałe, przy którym tranzystor może pracować w sposób długotrwały. Krzywa mocy admysyjnej na polu charakterystyk wyjściowych jest hiperbolą $I = P_a/U$. Zagadnienie mocy rozproszonej w tranzystorze rozpatrzmy dokładniej w p. 5.12.3.

Prąd maksymalny I_{Cmax} jest ograniczeniem wynikającym najczęściej nie z nadmiernej ilości ciepła rozpraszanego w tranzystorze, lecz ze zmian współczynnika wzmocnienia prądowego (α lub β maleje dla dużych prądów emitera, a więc i dla dużych

prądów kolektora — por. rys. 5.21). Dlatego przekroczenie wartości I_{Cmax} najczęściej nie powoduje uszkodzenia cieplnego tranzystora, lecz prowadzi do nieakceptowalnego zmniejszenia współczynnika wzmocnienia prądowego. W przypadku wzmacniania dużych sygnałów przekroczenie wartości I_{Cmax} powoduje duże zniekształcenie nieliniowe wzmacnianego sygnału.

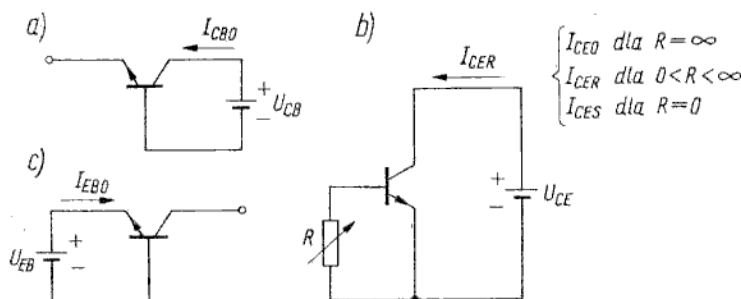
Omówimy teraz bardziej szczegółowo pozostałe parametry statyczne.

Prądy zerowe

5.7.3.1

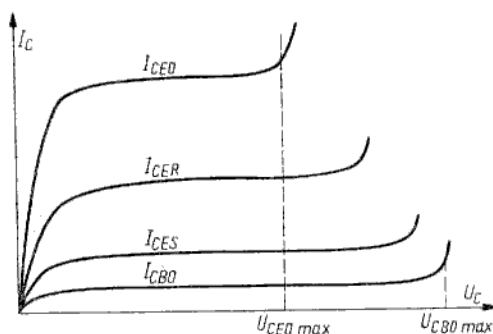
Z pojęciem prądu zerowego zetknęliśmy się już w p. 5.3, w którym stwierdziliśmy, że przez złącze C-B spolaryzowane zaporowo płynie prąd I_{CBO} nie związany ze wstrzykiwaniem nośników przez emiter. Bardziej ogólnie prądem zerowym nazywa się prąd płynący przez tranzystor włączony w układzie dwójnika, tj. przy polaryzacji dwu końcówek bez oddzielnej polaryzacji końcówki trzeciej. W układzie WB jest to prąd I_{CBO} , czyli prąd w spolaryzowanym zaporowo złączu B-C przy $I_E = 0$. W układzie WE jest to prąd I_{CEO} , czyli prąd w obwodzie emiter-kolektor przy $I_B = 0$.

Wyjaśnijmy znaczenie indeksu trójliterowego, który jest powszechnie stosowany w katalogach do oznaczania prądów i napięć w tranzystorze (trzy litery, gdyż tranzystor jest trójnikiem). Pierwsze dwie litery oznaczają dwie końcówki, do których jest doprowadzone napięcie lub przepływa prąd. Druga litera oznacza końcówkę wspólną dla wejścia i wyjścia. Trzecia litera oznacza stan, w jakim znajduje się pozostała trzecia końcówka w odniesieniu do końcówki wspólnej (może to być:



Rys. 5.37

Układy włączenia tranzystora przy pomiarze poszczególnych prądów zerowych: a) I_{CBO} ; b) I_{CEO} ; I_{CER} ; I_{CES} ; c) I_{EBO}



Rys. 5.38

Przebiegi prądów zerowych w funkcji napięcia kolektora

zwarcie — litera *S* od ang. *short*, rozwarcie — litera *O* od ang. *open* lub może być włączony rezystor — litera *R*).

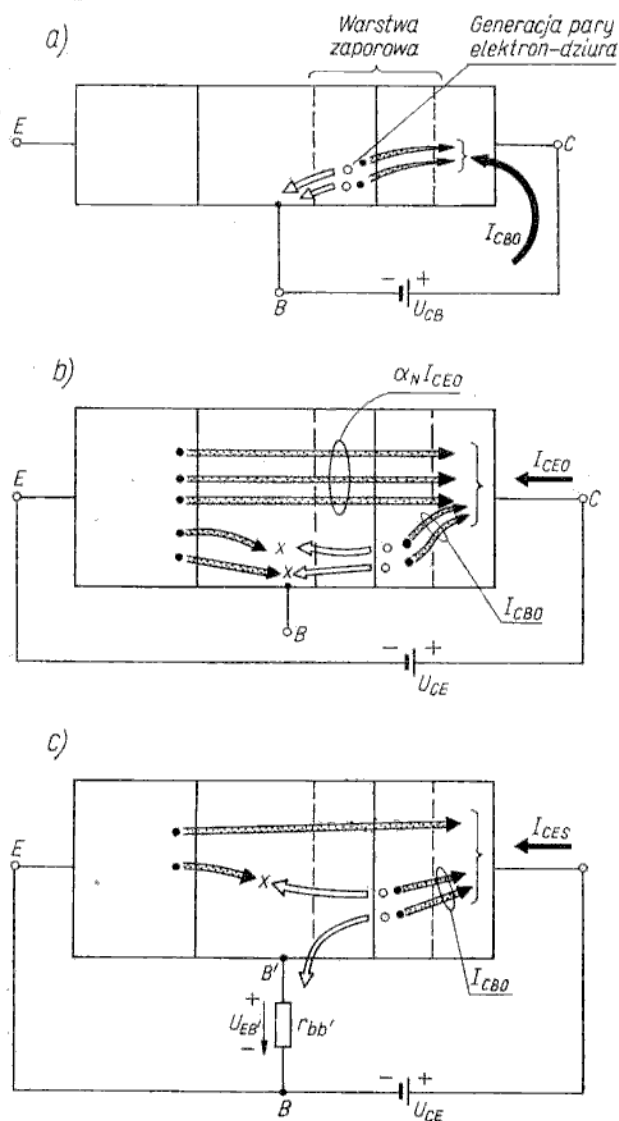
Spśród wielu możliwych prądów zerowych najistotniejsze są następujące:

I_{CBO} — rys. 5.37a

I_{CEO}
 I_{CES} } — rys. 5.37b

I_{CER}

Przy pracy inwersyjnej tranzystora ważny jest również prąd I_{EBO} (rys. 5.37c).



Rys. 5.39

Ilustracje objaśniające sens fizyczny poszczególnych prądów zerowych: a) I_{CBO} ; b) I_{CEO} ;
c) I_{CES}

Na rysunku 5.38 przedstawiono zależność $I_C(U_C)$, przy czym napięcie kolektora U_C jest mierzone względem bazy dla prądu I_{CBO} lub względem emitera dla trzech pozostałych prądów. Wyjaśnijmy dlaczego

$$I_{CEO} > I_{CER} > I_{CES} > I_{CBO}$$

Znamy sens fizyczny prądu I_{CBO} — jest to prąd unoszenia nośników mniejszościowych, typowy dla pracy złącza $p-n$ w kierunku zaporowym. W przypadku tranzystora krzemowego jest to głównie prąd nośników generowanych w warstwie zaporowej złącza $B-C$ (rys. 5.39a).

Prąd I_{CEO} jest mierzony przy napięciu doprowadzonym między kolektor a emiter — U_{CE} . W tym przypadku odkłada się pewien spadek napięcia na złączu $E-B$ w kierunku przewodzenia (na przykład dla $U_{CE} = 10$ V może być $U_{CB} = 9,9$ V oraz $U_{EB} = 0,1$ V). Dlatego przez złącze $B-C$ oprócz prądu I_{CBO} płynie również strumień nośników wstrzykiwanych przez emiter (rys. 5.39b). Ponieważ prąd emitera jest równy prądowi kolektora (rozwarta baza), zatem

$$I_{CEO} = \alpha_N I_{CEO} + I_{CBO} \quad (5.66)$$

Stąd

$$I_{CEO} = \frac{I_{CBO}}{1 - \alpha_N} = \left(\frac{\alpha_N}{1 - \alpha_N} + \frac{1 - \alpha_N}{1 - \alpha_N} \right) I_{CBO} = (\beta_N + 1) I_{CBO} \quad (5.66a)$$

Prąd I_{CER} jest mniejszy niż prąd I_{CEO} , gdyż rezystor R bocznikuje złącze $E-B$. Zmniejsza się zatem napięcie U_{EB} polaryzujące to złącze w kierunku przewodzenia, a więc maleje efekt wstrzykiwania nośników z emitera.

Wydawałoby się, że w przypadku zwarcia emitera z bazą nie ma żadnego napięcia na złączu $E-B$, czyli nie ma wstrzykiwania nośników z emitera do bazy i prąd I_{CES} powinien być równy I_{CBO} . Jednak $I_{CES} > I_{CBO}$, gdyż nie można mówić o idealnym zwarciu emitera z bazą, jeżeli uwzględnimy rezystancję rozproszoną bazy r_{bb} . Spadek napięcia na tej rezystancji, spowodowany przepływem prądu w obwodzie kolektor-baza, powoduje efektywne spolaryzowanie złącza $E-B'$ w kierunku przewodzenia. Istnieje zatem niewielki strumień nośników wstrzykiwanych przez emiter (rys. 5.39c).

Napięcia maksymalne

5.7.3.2

Ograniczenia maksymalnych wartości napięć polaryzacji tranzystora są spowodowane zjawiskiem przebiecia.

W tranzystorach bipolarnych można mówić o trzech rodzajach przebiecia:

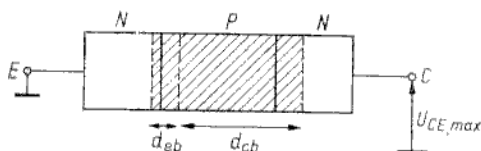
- Zenera (dotyczy złącza $E-B$ przy polaryzacji na końcówkach E, B);
- skrośne (dotyczy w pewnym sensie obszaru bazy przy polaryzacji na końcówkach $E-C$);
- lawinowe (dotyczy złącza $B-C$ przy polaryzacji na końcówkach B, C lub E, C).

Przebiecie Zenera występuje w złączach $E-B$ tranzystorów dryftowych, gdyż są to złącza silnie domieszkowane o napięciu przebiecia mniejszym niż 6 V (por. p. 3.1.5). W tranzystorach stopowych, w których baza jest równomiernie i słabo domieszkowana, napięcie przebiecia złącza $E-B$ ma większe wartości i jest wywołane zjawiskiem lawinowego powielania nośników.

Przebiecie skrośne¹⁾ jest związane z modulacją efektywnej szerokości bazy wskutek roz-

¹⁾ W literaturze angielskiej: *punch-through*.

szereżania się warstwy zaporowej złącza baza-kolektor. Jeżeli przy pewnym napięciu U_{CE} warstwa zaporowa złącza $B-C$ „wchłonie” cały obszar bazy (rys. 5.40), to nośniki wstrzykiwane przez emiter są bez strat odbierane przez kolektor, czyli prąd kolektora gwałtownie wzrasta. To zjawisko nazywa się *przebicciem*



Rys. 5.40

Ilustracja stanu fizycznego w tranzystorze przy przebicciu skośnym

skośnym lub *przekłuciem*. Taki mechanizm przebiccia był dość częsty w tranzystorach stopowych o dużych częstotliwościach granicznych (z uwagi na cieką i słabo domieszkowaną bazę). We współczesnych tranzystorach epiplanarnych warstwa zaporowa złącza $B-C$ rozszerzając się wnika w coraz silniej domieszkowany obszar bazy, a im silniej jest domieszkowany półprzewodnik, tym mniejsze są przyrosty szerokości warstwy zaporowej. Dlatego w tych tranzystorach, zanim dojdzie do całkowitego „wchłonięcia” bazy przez warstwę zaporową złącza $B-C$, zwykle wcześniej wystąpi przebiccie lawinowe tego złącza. Tylko w tranzystorach epiplanarnych o bardzo cienkiej bazie jest możliwe przebiccie skośne. Złącza $B-C$ najczęściej ulegają przebicciu lawinowemu (por. p. 3.1.5). Wartość napięcia przebiccia zależy od układu włączenia tranzystora, przy czym jest największa w układzie WB przy rozwartym emiterze (U_{CBOmax}), a najmniejsza w układzie WE przy rozwartej bazie (U_{CEOmax}). Mechanizm przebiccia lawinowego złącza $B-C$ przy pracy tranzystora w układzie WB nie różni się od przebiccia lawinowego diody. Prąd kolektora w zakresie przebiccia można zapisać w postaci

$$I_C = (\alpha_N I_E + I_{CBO})M \quad (5.67)$$

$$M = \frac{1}{1 - \left(\frac{U_{CB}}{U_{CBOmax}} \right)^n} \quad (5.67a)$$

przy czym: M — współczynnik powielania; n — liczba przyjmująca wartości 3...6.

Prąd $I_C \rightarrow \infty$, jeżeli $M \rightarrow \infty$, czyli $U_{CB} \rightarrow U_{CBOmax}$. Należy zauważyć, że napięcie przebiccia nie zależy od prądu emitera, a więc U_{CBmax} (napięcie przebiccia przy dowolnym $I_E = \text{const}$) jest równe U_{CBOmax} (napięcie przebiccia przy $I_E = 0$).

Przebiccie lawinowe w układzie WE przebiega w nieco inny sposób. O ile w układzie WB wszystkie istotne procesy występują tylko w warstwie zaporowej złącza $B-C$, o tyle w układzie WE duży wpływ na przebieg zjawiska przebiccia ma również złącze $E-B$ wstrzykujące nośniki do bazy.

Równanie (5.66) przepisuje się z uwzględnieniem powielania nośników

$$I_{CEO} = (\alpha_N I_{CEO} + I_{CBO})M$$

stąd

$$I_{CEO} = \frac{MI_{CBO}}{1 - \alpha_N M} \quad (5.68)$$

Wzór (5.68) jest słuszny dla $I_B = 0$. W ogólnym przypadku dla $I_B \neq 0$

$$I_C = (\alpha_N I_E + I_{CBO})M$$

Ponieważ

$$I_E = I_C + I_B$$

zatem

$$I_C = \frac{\alpha_N M I_B}{1 - \alpha_N M} + \frac{M I_{CBO}}{1 - \alpha_N M} \quad (5.69)$$

Z obu zależności (5.68) i (5.69) wynika, że $I_C \rightarrow \infty$ przy $(1 - \alpha_N M) \rightarrow 0$, czyli $M \rightarrow 1/\alpha_N$. Zatem w układzie WE do przebicia wystarczy, żeby współczynnik M nieznacznie wzrósł powyżej jedności (na przykład dla $\alpha_N = 0,99$ wystarczy, aby $M \approx 1,01$), podczas gdy w układzie WB trzeba, aby $M \rightarrow \infty$. Dlatego napięcie $U_{CEO\max}$ jest mniejsze niż $U_{CBO\max}$. Podstawiając do warunku

$$\alpha_N M = 1$$

wartość M (5.67a) przy $U_{CB} = U_{CEO\max}$, można wyprowadzić zależność

$$U_{CEO\max} = \sqrt[n]{1 - \alpha_N} U_{CBO\max} \approx \frac{U_{CBO\max}}{\sqrt[n]{\beta_N}} \quad (5.70)$$

$$\text{Zwykle } U_{CEO\max} = \left(\frac{1}{3} \dots \frac{1}{2} \right) U_{CBO\max}.$$

Napięcie przebicia $U_{CE\max}$ (przy $I_B > 0$) jest mniejsze niż $U_{CEO\max}$, gdyż w miarę wzrostu prądu I_B zwiększa się I_C , a większym prądem kolektora odpowiada większa wartość współczynnika α_N , czyli przy mniejszej wartości M może być spełniony warunek $\alpha_N M = 1$.

Dotychczas mówiliśmy o przebicciu przy następujących warunkach:

$$U_{CB\max} \Big|_{I_E = \text{const.}}, \text{ którego szczególnym przypadkiem jest } U_{CBO\max} \Big|_{I_E = 0},$$

$$U_{CE\max} \Big|_{I_B > 0}, \text{ którego szczególnym przypadkiem jest } U_{CEO\max} \Big|_{I_B = 0}.$$

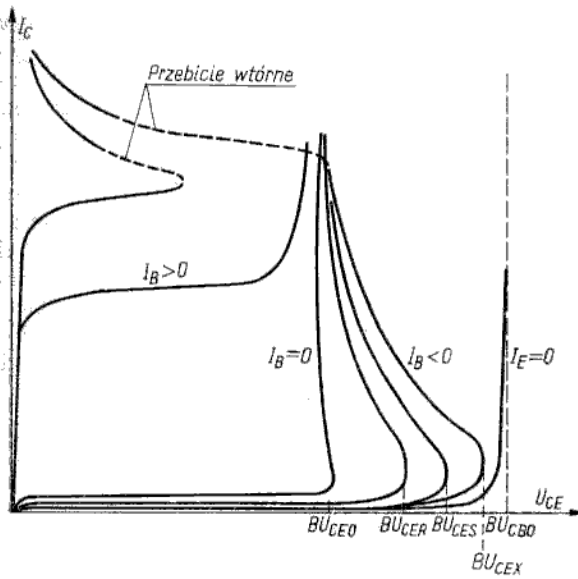
Dla układu WE interesujące jest również przebicie przy $I_B < 0$, mierzone w następujących warunkach:

$$U_{CER\max} - 0 < R_{BE} < \infty$$

$$U_{CES\max} - U_{BE} = 0 (R_{BE} = 0)$$

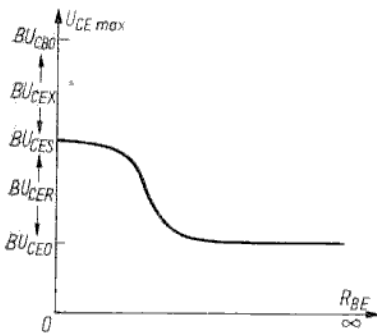
$U_{CEX\max}$ — obwód baza emiter z włączonym źródłem polaryzacji zaporowej. Wszystkie wartości napięć przebicia zaznaczono na typowym wykresie charakterystyk $I_C(U_{CE})$ (rys. 5.41)¹⁾, a na rys. 5.42 przedstawiono zależność $U_{CER\max}$ od R_{BE} , przy czym na osi napięcia zaznaczono również wartości $U_{CEX\max}$ i $U_{CBO\max}$. Wzrost napięcia $U_{CER\max}$ w miarę zmniejszania rezystancji R_{BE} tłumaczy się coraz silniejszym bocznikowaniem złącza $E-B$ przez R_{BE} , czyli coraz mniejszy jest udział nośników wstrzykiwanych z emitera w prądzie kolektora. Napięcie $U_{CES\max}$ nie jest równe $U_{CBO\max}$ tylko dlatego, że istnieje rezystancja rozproszona bazy $r_{bb'}$, która uniemożliwia efektywne zwarcie złącza $E-B$ w obwodzie

¹⁾ Na rysunku podano oznaczenia stosowane w katalogach, tj. BU_{CEO} zamiast $U_{CEO\max}$ itp. (litera B od ang. *breakdown* — przebicie).



Rys. 5.41

Typowa rodzina charakterystyk $I_C(U_{CE})$ z uwzględnieniem obszaru przebicia lawinowego



Rys. 5.42

Zależność napięcia przebicia U_{CERmax} od wartości rezystancji R_{BE}

zewnętrznym. Dlatego przy $R_{BE} = 0$ pozostaje jeszcze pewien wpływ złącza $E-B$ na wartość prądu kolektora. Dopiero przy spolaryzowaniu złącza $E-B$ źródłem zewnętrznym w kierunku zaporowym można doprowadzić do pełnego wyeliminowania wstrzykiwania nośników z emitera i wówczas $U_{CEXmax} = U_{CBOmax}$. Należy zauważyć, że w przypadku zamkniętego obwodu baza-emiter (przez rezystor R_{BE} bądź też zwarcie bazy z emiterem lub włączenie źródła polaryzacji zaporowej) przez rezystancję r_{bb} , oraz zewnętrzny rezystor R_{BE} płynie część prądu MI_{CBO} dając spadek napięcia polaryzujący złącze $E-B$ w kierunku przewodzenia (por. rys. 5.39c). Dlatego w miarę wzrostu MI_{CBO} wzrasta prąd I_E , co powoduje wzrost wartości współczynnika α_N , a więc zmniejszenie wartości napięcia przebicia. To zjawisko tłumaczy istnienie odcinka o ujemnej rezystancji dynamicznej (spadek napięcia U_{CE} przy wzroście prądu I_C) na charakterystykach $I_C(U_{CE})$ zdejmowanych w obszarze przebicia lawinowego przy $I_B < 0$.

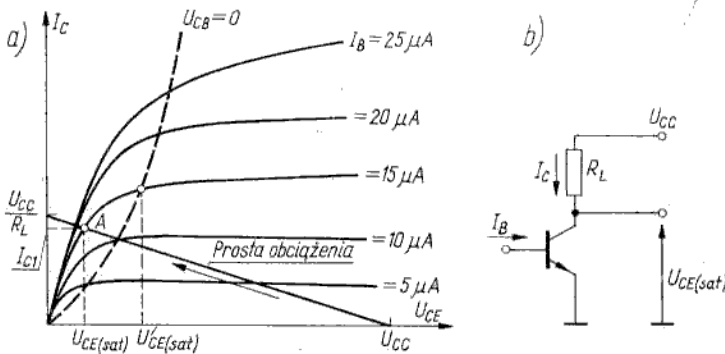
Ponieważ „skok” napięcia przykładowo od U_{CEXmax} do U_{CEOmax} jest bardzo szybki (w czasie ok. 1 ns), pracę tranzystora w obszarze przebicia lawinowego można wykorzystać do generacji impulsów o bardzo stromych zboczach. Tranzystory przeznaczone do takich zastosowań noszą nazwę *tranzystorów lawinowych*.

Należy wyraźnie podkreślić że żaden z omawianych mechanizmów przebicia „sam z siebie” nie niszczy tranzystora. Tylko w przypadku braku odpowiedniego ograniczenia prądowego w obwodzie zewnętrznym może wydzielić się nadmierna ilość ciepła i następuje uszkodzenie tranzystora (zwykle jest to przepalenie cienkich drutów łączących końcówki zewnętrzne z polami kontaktowymi lub inne uszkodzenie cieplne kontaktów). W tranzystorach dużej mocy i dużej częstotliwości występuje niekiedy tzw. *przebicie wtórne*, które powoduje nieodwracalne uszkodzenie tranzystora (rys. 5.41). Istota fizyczna tego przebicia nie jest do końca wyjaśniona. Najprawdopodobniej jest ono spowodowane nierównomiernym rozkładem gęstości prądu emitera i lokalnym zachwianiem równowagi cieplnej. Inaczej mówiąc, w niewielkim obszarze złącza płynie nadmierny prąd i wydziela się zbyt dużo ciepła (tzw. „gorąca plama”), co prowadzi do lokalnej generacji dodatkowej liczby nośników, a więc do dalszego wzrostu prądu. Ostatecznie bardzo wysoka temperatura lokalna powoduje wtopienie aluminium z pola kontaktowego i zwarcie złącza, czyli uszkodzenie tranzystora.

Napięcie nasycenia

5.7.3.3

Napięcie nasycenia jest wielkością ograniczającą minimalne wartości napięcia U_{CE} przy pracy tranzystora w obszarze aktywnym normalnym, czyli wyznacza granicę między zakresem aktywnym a zakresem nasycenia. Stan nasycenia oznacza polaryzację obu złączy w kierunku przewodzenia. W związku z tym nie ma sensu mówić o napięciu nasycenia w układzie WB, gdyż przejście z zakresu aktywnego do zakresu nasycenia odbywa się przy $U_{CB} = 0$. W układzie WE, jeżeli zdefiniuje się napięcie nasycenia $U_{CE(sat)}$ jako takie napięcie U_{CE} , przy którym $U_{CB} = 0$ (rys. 5.43), to $U_{CE(sat)} = U_{BE}$. Na rysunku 5.43 zaznaczono na-



Rys. 5.43

Układ włączenia tranzystora (b) i konstrukcja graficzna (a) wyjaśniająca sens pojęcia napięcia nasycenia

napięcie nasycenia $U'_{CE(sat)}$ wyznaczone według takiej definicji dla prądu $I_B = 15 \mu A$. Parabola wykreślona linią przerywaną łączy punkty pracy spełniające warunek $U_{CB} = 0$ i wyznacza granicę między obszarami liniowych i nieliniowych zmian prądu kolektora w funkcji napięcia kolektor-emiter.

Jeśli chodzi o zastosowanie tranzystora w układach impulsowych, bardziej celowe jest określenie napięcia nasycenia jako niemal najmniejszej wartości spadku napięcia

kolektor-emiter przy przewodzącym tranzystorze (im mniejsze jest to napięcie, tym lepiej tranzystor spełnia funkcję klucza).

Na rysunku 5.43 pokazano układ włączenia tranzystora (rys. 5.43b) i odpowiednią konstrukcję graficzną (rys. 5.43a). W miarę wzrostu prądu bazy punkt pracy przesuwa się po prostej obciążenia w kierunku wskazanym strzałką, czyli zwiększa się prąd I_C i maleje napięcie U_{CE} . Punkt pracy, oznaczony literą A , odpowiada prądowi bazy $I_B = 15 \mu A$, bardzo mało zmienia się przy dalszym wzroście prądu bazy. Liniowa zależność prądu kolektora od prądu bazy jest spełniona przy $I_B \leq 10 \mu A$ ($I_C = \beta_N I_B$), natomiast przy $I_B = 15 \mu A, 20$ itd. $I_C < \beta_N I_B$. Prąd I_{C1} , odpowiadający punktowi A , jest — praktycznie biorąc — równy maksymalnemu prądowi, jaki jest w ogóle możliwy w tym obwodzie

$$I_{C1} \approx \frac{U_{CC}}{R_L}$$

Napięcie $U_{CE(sat)}$ jest, praktycznie biorąc, najmniejszym napięciem kolektor-emiter dla tranzystora pracującego w rozpatrywanym układzie.

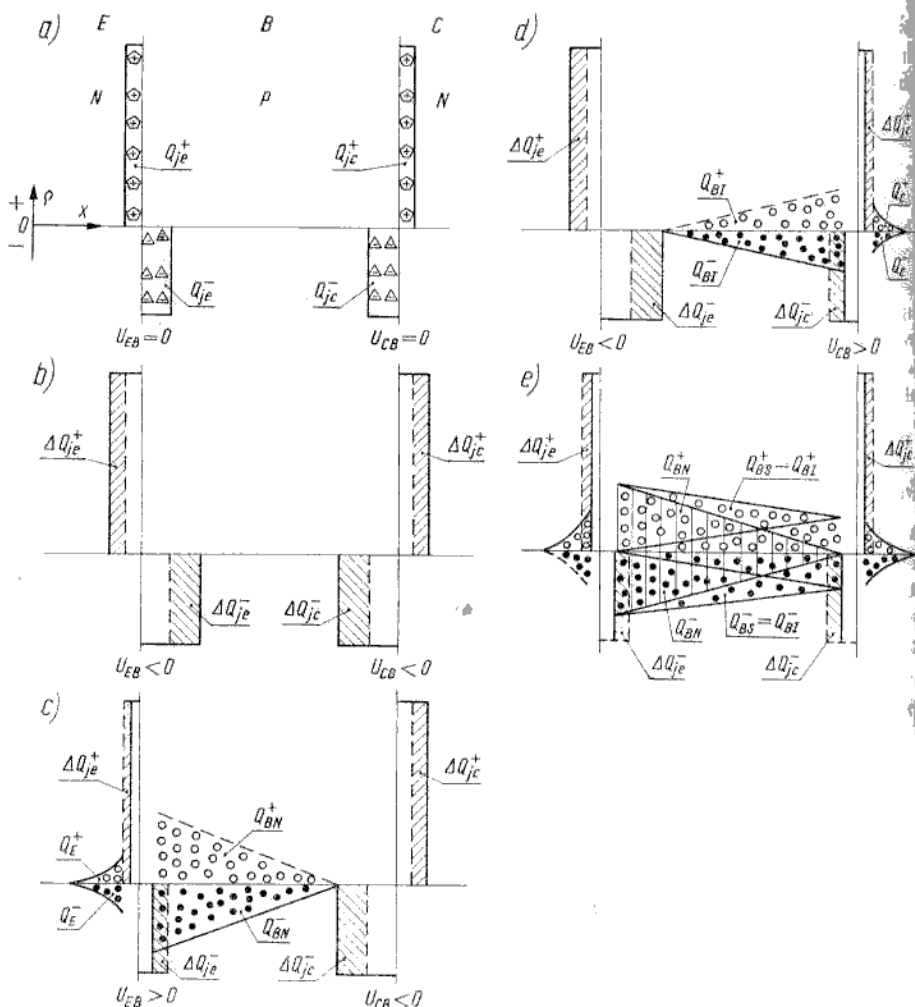
Ogólnie zdefiniujemy napięcie nasycenia $U_{CE(sat)}$ jako napięcie U_{CE} zmierzone przy określonym prądzie kolektora i prądzie bazy spełniającym warunek $I_B > I_C/\beta_N$. Stąd wynika, że w katalogach obok wartości $U_{CE(sat)}$ powinny być podane wartości prądów I_C, I_B , przy jakich to napięcie jest mierzone. W katalogach często podaje się również napięcie U_{BE} odpowiadające tym samym wartościom prądów I_C, I_B , przy czym jest ono oznaczane jako $U_{BE(sat)}$.

Praca nieliniowa dynamiczna

5.8

W tym punkcie rozpatrzmy właściwości dynamiczne tranzystora przy wystęrowaniu dużym sygnałem. Będą nas interesowały w szczególności modele nieliniowe dynamiczne oraz procesy przejściowe przy przełączaniu tranzystora z odcięcia do nasycenia (włączanie) i odwrotnie (wyłączanie). Tranzystor jest elementem inercyjnym, tzn. odpowiedź na szybką zmianę warunków wystęrowania nie jest w nim natychmiastowa. Na przykład skokowy przyrost prądu bazy wywołuje odpowiedni (β razy większy) przyrost prądu kolektora dopiero po pewnym czasie. Inercyjność tranzystora jest spowodowana zjawiskiem gromadzenia ładunków w różnych jego obszarach. Ponieważ wartości tych ładunków zależą od punktu pracy, każda szybka zmiana sygnału sterującego wywołuje proces nieustalony, polegający na przepływie prądów ładowania bądź rozładowania określonych obszarów tranzystora. Wartości tych prądów zależą od przyrostów ładunków i stałych czasowych procesów przeładowania.

W jakich obszarach tranzystora gromadzą się ładunki i jakie są ich przyrosty przy zmianie punktu pracy? Posłużymy się schematyczną ilustracją rozkładów ładunków, przedstawioną na rys. 5.44 dla kilku charakterystycznych stanów pracy tranzystora. Dla uproszczenia jest rozważany tranzystor z jednorodną bazą o budowie symetrycznej, tj. emiter i kolektor są jednakowo domieszkowane, a w bazie jest równomierny rozkład domieszki o koncentracji mniejszej niż w emiterze i kolektorze. Na rysunku 5.44a przedstawiono stan neutralny, tj. bez polaryzacji zewnętrznej ($U_{EB} = 0, U_{CB} = 0$). W tym przypadku istnieją tylko ładunki warstw zaporowych złącza emiter-baza (Q_{je}) oraz baza-kolektor (Q_{jc}). Suma ładunków dodatnich i ujemnych w każdej warstwie zaporowej jest oczywiście równa zeru.

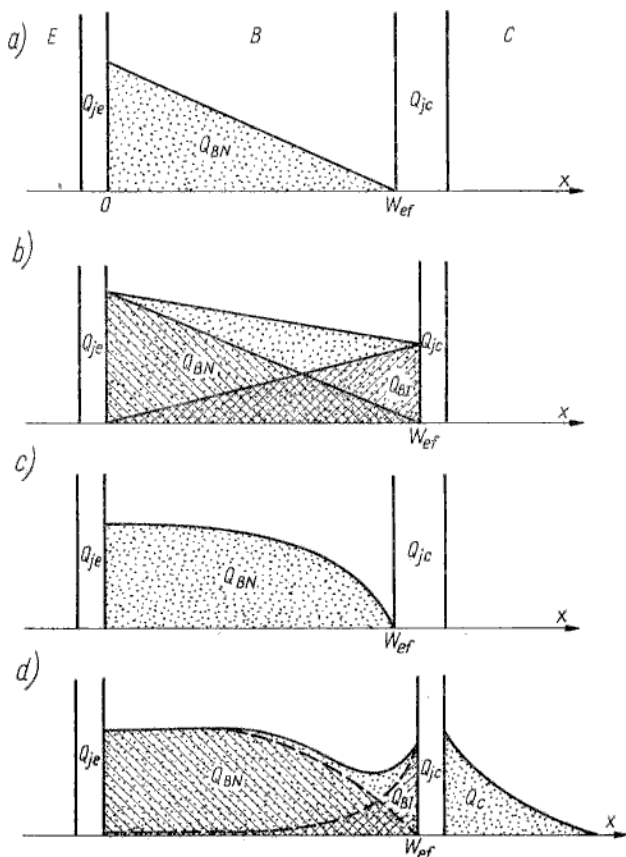


Rys. 5.44

Ilustracje rozkładów ładunków nieskompensowanych w tranzystorze dla kilku charakterystycznych stanów: a) stan neutralny, $U_{EB} = U_{CB} = 0$; b) zakres zatkania, $U_{EB} < 0$, $U_{CB} < 0$; c) zakres aktywny normalny, $U_{EB} > 0$, $U_{CB} < 0$; d) zakres aktywny inwersyjny, $U_{EB} < 0$, $U_{CB} > 0$; e) zakres nasycenia, $U_{EB} > 0$, $U_{CB} > 0$

Kolejne rysunki 5-44b-e przedstawiają przyrosty ładunków w stosunku do stanu neutralnego, występujące przy różnych wariantach polaryzacji tranzystora. W stanie odcięcia (rys. 5.44b) przy obu złączach spolaryzowanych zaporowo ($U_{EB} < 0$, $U_{CB} < 0$) wzrastają ładunki warstw zaporowych (ΔQ_{je} , ΔQ_{jc}). W stanie aktywnym normalnym (rys. 5.44c), tj. przy polaryzacji złącza E-B w kierunku przewodzenia, a złącza B-C w kierunku zaporowym ($U_{EB} > 0$, $U_{CB} < 0$) w stosunku do stanu neutralnego wzrasta ładunek warstwy zaporowej złącza B-C (ΔQ_{jc}), maleje ładunek warstwy zaporowej złącza E-B (ΔQ_{je}) oraz pojawia się w bazie i emiterze ładunek nadmiarowy (Q_{BN} , Q_E). Ładunek Q_{BN} jest utworzony przez

elektrony wstrzykiwane z emitera do bazy. Jest on neutralizowany przez ładunek dziur Q_{BN}^+ , które wskutek przyciągania elektrostatycznego przez wstrzykiwane elektrony w pomijalnie krótkim czasie przegrupowują się tak, by skompensować ładunek elektronów. Analogicznie w emiterze gromadzi się ładunek dziur wstrzykiwanych z bazy (Q_E^+), któremu towarzyszy ładunek elektronów (Q_E^-) przegrupowanych wskutek oddziaływania elektrostatycznego. Pozornie może wydawać się, że ładunki nośników nadmiarowych w bazie i emiterze nie mają znaczenia dla pracy dynamicznej tranzystora, gdyż oba te obszary pozostają obojętne elektrycznie ($Q_{BN}^+ + Q_{BN}^- = 0$; $Q_E^+ + Q_E^- = 0$). Należy jednak zwrócić uwagę, że dziury i elektrony wpływają do bazy oraz emitera z różnych kierunków. Przykładowo ładunek Q_{BN}^- jest utworzony przez elektrony dostarczane z obwodu emitera, a ładunek Q_{BN}^+ — przez dziury dostarczane z obwodu bazy. Ważne jest zatem, że mimo zachowania obojętności elektrycznej obszarów bazy i emitera gromadzenie ładunków Q_{BN} , Q_E wiąże się z przepływem określonych prądów w określonym czasie. Ponieważ zmiany ładunków nośników większościowych (Q_{BN}^+ , Q_E^-) natychmiast nadążają za zmianami ładunków nośników



Rys. 5.45
Porównanie rozkładów ładunków dla tranzystora bezdryftowego (a, b) i dryftowego (c, d).
Rysunki a), c) dotyczą pracy w zakresie normalnym, rysunki b), d) w zakresie nasycenia

mniejszościowych (Q_{BN}^- , Q_E^+), czas ustalania się procesów przejściowych zależy tylko od właściwości nośników mniejszościowych, a nie większościowych. Dlatego rozważając procesy przejściowe w tranzystorze będziemy mówić tylko o nośnikach mniejszościowych, przyjmując za oczywiste, że towarzyszą im zawsze w takiej samej liczbie nośniki większościowe.

W stanie aktywnym inwersyjnym (rys. 5.44d) sytuacja jest symetrycznie analogiczna do przedstawionej na rys. 5.44c, przy czym kolektor spełnia funkcję emitera, a emiter — kolektora. W stanie nasyceńca (rys. 5.44e), tj. przy polaryzacji obu złączy w kierunku przewodzenia, występuje superpozycja stanów przedstawionych na dwu poprzednich rysunkach.

Z przedstawionych ilustracji wynika wniosek, że inercyjność tranzystora jest związana ze zmianami dwu rodzajów ładunków, tj.:

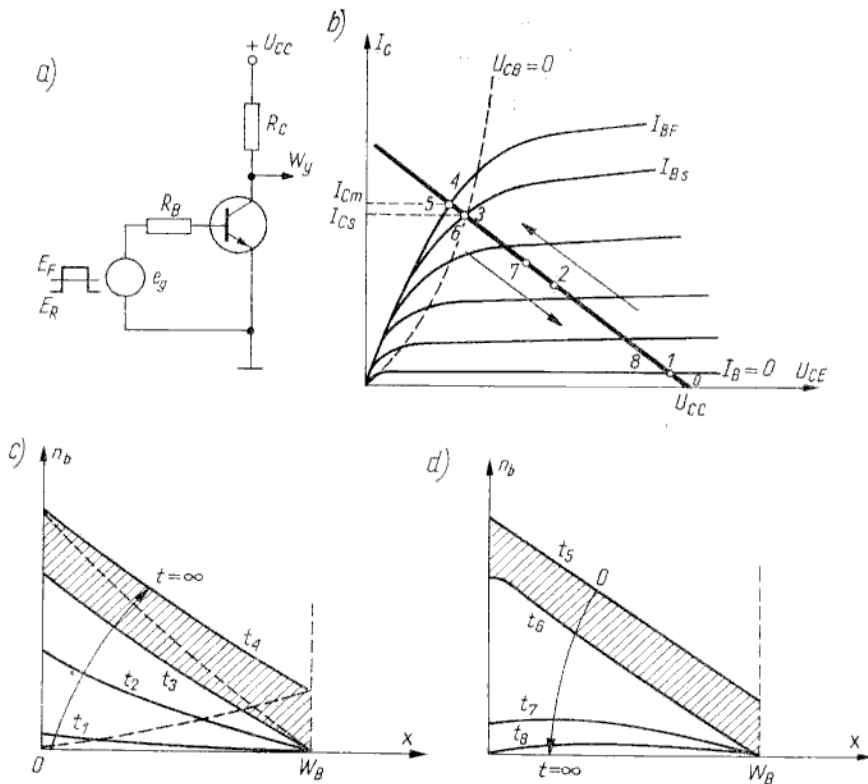
- ładunków warstw zaporowych złączy $E-B$, $C-B$ (Q_{je} , Q_{jc});
- ładunków mniejszościowych nadmiarowych w bazie ($Q_B = Q_{BN} + Q_{BI}$), emiterze (Q_E) i kolektorze (Q_C).

W tranzystorze z jednorodną bazą ładunki Q_E , Q_C są zwykle pomijalnie małe w porównaniu z ładunkiem Q_B , gdyż złącza $E-B$, $B-C$ jako niesymetryczne wstrzykują nośniki praktycznie biorąc, tylko do bazy. Dlatego w tym tranzystorze zwykle rozpatruje się tylko zjawiska związane z ładunkiem Q_B i ładunkami warstw zaporowych. W tranzystorze dryftowym epiplanarnym takie uproszczenie jest na ogół uzasadnione w odniesieniu do ładunku Q_E (można go pominąć), gdyż emiter jest silniej domieszkowany niż baza, czyli liczba nośników wstrzykiwanych z emitera do bazy jest znacznie większa niż z bazy do emitera, nie jest jednak słuszne w odniesieniu do ładunku Q_C — ładunek Q_C może być porównywalny z ładunkiem Q_B , gdyż są porównywalne koncentracje domieszek po obu stronach złącza $B-C$. W tranzystorze dryftowym w odróżnieniu od tranzystora bezdryftowego inny jest również rozkład gęstości ładunku w bazie, co wiadomo już z wcześniejszych rozważań. Uwzględniając te uwagi, na rys. 5.45 przedstawiono uproszczone ilustracje rozkładów ładunku w obu rodzajach tranzystorów dla dwu najciekawszych sytuacji, tj. dla stanu aktywnego normalnego i dla nasyceńca. Przejdźmy teraz do opisu fizycznego procesów przejściowych w tranzystorze.

Procesy przejściowe w tranzystorze

5.8.1

Zostanie teraz rozpatrzony proces przełączania tranzystora włączonego w układzie WE przedstawionym na rys. 5.46a. Tranzystor jest sterowany impulsami prostokątnymi ze źródła e_g . Rezystor R_B ogranicza prąd bazy, który przy polaryzacji dodatniej może osiągnąć co najwyżej wartość E_F/R_B , a przy polaryzacji ujemnej — E_R/R_B . Oczywiście przy polaryzacji ujemnej w stanie ustalonym prąd bazy jest znacznie mniejszy z uwagi na właściwości zaporowe złącza $E-B$. Prosta obciążenia, po której przesuwa się punkt pracy podczas włączania i wyłączenia tranzystora, przedstawiono na rys. 5.46b, przy czym cyframi od 1 do 8 oznaczono charakterystyczne punkty, które umożliwiają prześledzenie zgodności trajektorii punktu pracy z odpowiednimi przebiegami czasowymi i zmianami ładunków wewnątrz tranzystora. Należy zwrócić uwagę, że rys. 5.46b przedstawia charakterystyki statyczne, zatem zgodność wartości parametru (I_B) z wartością prądu kolektora, wynikającą z przecięcia charakterystyki $I_C(U_{CE})|_{I_B}$ z prostą obciążenia, występuje tylko w stanach ustalonych.

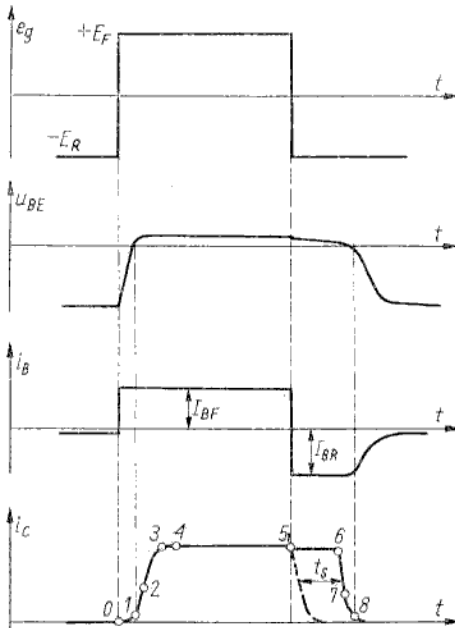


Rys. 5.46

Proces przełączania tranzystora w układzie WE: a) układ włączenia; b) trajektoria punktu pracy na polu charakterystyk wyjściowych; rozkłady koncentracji nośników nadmiarowych w bazie dla procesu włączania (c) i wyłączenia (d)

Interpretację fizyczną przebiegów czasowych przeprowadzimy dla uproszczenia na przykładzie tranzystora bezdryftowego. W tym tranzystorze wystarczy uwzględnić tylko transport dyfuzyjny nośników mniejszościowych w bazie. W tranzystorze dryftowym należałoby natomiast uwzględnić zarówno dyfuzję jak i unoszenie nośników w bazie (w stanie nasycenia warunki transportu nośników w kierunku normalnym i inwersyjnym są całkowicie różne, co stwarza dodatkowe trudności interpretacyjne) oraz nie można by pominąć ładunku nośników magazynowanych w obszarze kolektora. Ponieważ jednak przebiegi prądów i napięć w stanie niestabilnym są jakościowo jednakowe dla obu rodzajów tranzystorów, interpretację fizyczną ograniczymy do prostszego przypadku tranzystora bezdryftowego. Kluczem do zrozumienia procesów przejściowych jest zjawisko przeladowania obszaru bazy, które ilustrują rozkłady koncentracji nośników mniejszościowych nadmiarowych, przedstawione na rys. 5.46c, d dla włączania i wyłączenia. Dla większej przejrzystości na tych rysunkach nie pokazano ładunków warstw zaporowych, jednak wpływ zmian tych ładunków na procesy przejściowe będzie uwzględniany. Teraz zostaną wyjaśnione przebiegi prądów i napięć przedstawione na rys. 5.47.

Niech w chwili t_0 (punkt O na rys. 5.46b oraz 5.47) nastąpi skokowa zmiana napięcia wejściowego e_g od wartości $-E_R$ do E_F . Z uwagi na spełnienie warunku sterowania prądowego podczas włączania tranzystora prąd bazy osiąga natychmiast wartość $I_{BF} = E_F/R_B$. Zmiana natomiast polaryzacji złącza $E-B$ z kierunku



Rys. 5.47
Przebiegi czasowe napięć i prądów podczas włączania i wyłączania tranzystora

zaporowego do kierunku przewodzenia nie jest natychmiastowa, gdyż zmiana napięcia U_{BE} łączy się z koniecznością zmiany ładunku warstwy zaporowej Q_{je} . Aby napięcie U_{BE} zmieniło się od wartości $-E_R$ (w przybliżeniu) do zera, konieczne jest zmniejszenie ładunku Q_{je} o wartość ΔQ_{je} pokazaną na rys. 5.45b, czyli warstwa zaporowa złącza $E-B$ musi być częściowo rozładowana. Ponieważ prąd rozładowania jest ograniczony (jest to prąd bazy I_{BF} , a ściśle biorąc część tego prądu), zatem dopiero po pewnym czasie napięcie U_{BE} osiąga wartość zerową i punkt pracy tranzystora przechodzi z zakresu odcięcia do zakresu normalnego aktywnego (od 0 do 1). Czas ten jest nazywany *czasem opóźnienia*. Wyjaśnijmy teraz, dlaczego stwierdzono wyżej, że — ściśle biorąc — tylko część prądu I_{BF} „idzie” na rozładowanie warstwy zaporowej złącza $E-B$? Otóż należy pamiętać, że przy zmianie potencjału bazy zmienia się napięcie U_{BC} (potencjał kolektora jest w tym czasie stały, gdyż prąd kolektora jest, praktycznie biorąc, równy zero), czyli zmienia się również ładunek Q_{jc} . Właśnie dlatego część prądu I_{BF} przeladunkuje warstwę zaporową złącza $E-B$, a pozostała jego część przeladunkuje warstwę zaporową złącza $B-C$.

Zwróćmy teraz uwagę na wykres rozkładu koncentracji $n_b(x)$ dla chwili t_1 (rys. 5.46e) odpowiadającej punktowi pracy 1. Nachylenie tego rozkładu przy kolektorze (dla współrzędnej $x = W_B$) jest prawie równe zero, co oznacza, że dopiero od tej chwili rozpoczyna się wzrost prądu kolektora. Przypomnijmy, że prąd dyfuzji jest proporcjonalny do gradientu koncentracji nośników. W okresie od t_1 do t_3 , reprezentowanym przez jeden wykres dla chwili t_2 (punkt pracy 2), nachylenie

rozkładu zwiększa się, przeto wzrasta również prąd kolektora. W chwili t_3 (punkt pracy 3) napięcie $U_{CB} = 0$, co oznacza, że tranzystor wchodzi w stan nasycenia. Temu punktowi pracy odpowiada w stanie ustalonym prąd bazy I_{Bs} (rys. 5.46b). Ponieważ jednak prąd wejściowy ma wartość $I_{BF} > I_{Bs}$, stan ustalony jest osiągany dopiero w chwili t_4 dla punktu pracy 4. Należy zauważyć (rys. 5.46c), że rozkład koncentracji nośników w chwili t_4 jest superpozycją dwu rozkładów nośników wstrzykiwanych przez emiter i kolektor (linie przerywane). W okresie od t_3 do t_4 prądy emitera i kolektora, praktycznie biorąc, nie zmieniają się, występuje tylko gromadzenie w bazie pewnego ładunku nadmiarowego, odpowiadającego obszarom zakreskowanym na rys. 5.46c, d. Wartość tego ładunku (nazwijmy go *ładunkiem przesterowania*) zależy od wartości współczynnika przesterowania określonego jako

$$k_f = \frac{I_{BF}}{I_{Bs}} \quad (5.71)$$

Ponieważ

$$I_{Bs} = \frac{I_{Cs}}{\beta_N}$$

a

$$I_{Cs} \approx I_{Cm}$$

zatem

$$k_f \approx \frac{\beta_N I_{BF}}{I_{Cm}} \quad (5.72)$$

przy czym

$$I_{Cm} \approx \frac{U_{CC}}{R_C}$$

Rozpatrzmy teraz proces wyłączenia. Niech w chwili t_5 nastąpi skokowa zmiana napięcia wejściowego e_g od wartości E_F do $-E_R$. W tranzystorze ustali się stan odcięcia po rozładowaniu obszaru bazy z nośników nadmiarowych. Baza jest rozładowywana głównie wskutek usuwania nośników do obwodu bazy (prąd $I_{BR} = -E_R/R_B$). Część ładunku bazy wypływa również do kolektora, a część do emitera (w końcowej fazie procesu wyłączenia). Kolejne wykresy rozkładu koncentracji nośników nadmiarowych podczas rozładowywania bazy przedstawiono na rys. 5.46d. W okresie od t_5 do t_6 jest usuwany ładunek przesterowania (pole zakreskowane). W tym czasie, nazywanym *czasem przeciągania* lub *magazynowania*, prąd kolektora, praktycznie biorąc, nie zmienia się (nachylenie rozkładu koncentracji nośników przy kolektorze jest stałe). W chwili t_6 punkt pracy przechodzi z zakresu nasycenia do zakresu aktywnego normalnego i prąd kolektora zaczyna maleć. Złącze $E-B$ w dalszym ciągu jest polaryzowane w kierunku przewodzenia ($U_{EB} > 0$) wskutek obecności dużego ładunku nośników mniejszościowych na jego prawej krawędzi. Dopiero w chwili t_8 napięcie U_{EB} osiąga wartość zerową i od tej chwili prąd bazy zaczyna maleć, a w złączu $E-B$ ustalają się warunki pracy zaporowej.

Wróćmy teraz do tej fazy wyłączenia, którą nazywaliśmy czasem przeciągania lub magazynowania. Należy zauważyć, że długość tej fazy zależy od stopnia przesterowania.

sowa przelączania τ_N i stała czasowa magazynowania τ_s . Stałe czasowe τ_N , τ_s są podawane w katalogach lub można je wyznaczyć z innych parametrów.

Jeżeli przyjmie się, że wpływ stałej czasowej obwodu obciążenia tranzystora jest do pominięcia (małe wartości rezystancji obciążenia), to można napisać następujące przybliżone związki:

$$t_r \approx \tau \ln \frac{k_f - 0,1}{k_f - 0,9} \quad (5.76)$$

$$t_s \approx \tau_s \ln \frac{k_r + k_f}{k_r + 1} \quad (5.77)$$

$$t_f \approx \tau \ln \frac{k_r + 1}{k_r + 0,1} \quad (5.78)$$

w których τ — czas życia nośników¹⁾.

Wyrażenie na czas t_d będzie podany w następnym punkcie po wprowadzeniu pewnych dodatkowych pojęć.

Modele nieliniowe dynamiczne

5.8.3

Model nieliniowy dynamiczny stanowi najbardziej podstawowy i ogólny opis właściwości tranzystora, gdyż zawiera składniki reprezentujące model statyczny nieliniowy oraz można z niego wyprowadzić modele liniowe (ciągłą zależność nieliniową można aproksymować w niewielkim przedziale odcinkiem linii prostej). Stąd wynika, że modele nieliniowe dynamiczne stosowane w praktyce stanowią mniej lub bardziej uproszczone odbicie ogólnej teorii działania tranzystora. Procesy zachodzące w tranzystorze, jak i w innych elementach półprzewodnikowych, podlegają kilku podstawowym prawom, które matematycznie są formułowane w postaci sześciu równań (jest to najogólniejszy model matematyczny tranzystora):

równania ciągłości (dla półprzewodnika typu n):

$$\text{elektronów } \frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} - \frac{1}{\tau_p} (p - p_{n0})$$

$$\text{dziur } \frac{\partial p}{\partial t} = \frac{1}{q} \frac{\partial J_p}{\partial x} - \frac{1}{\tau_p} (p - p_{n0})$$

równania gęstości prądu:

$$\text{elektronów } J_n = q\mu_n n E + qD_n \frac{\partial n}{\partial x}$$

$$\text{dziur } J_p = q\mu_p p E - qD_p \frac{\partial p}{\partial x}$$

równanie wypadkowej gęstości ładunku:

$$\rho = q[p - n + N_D - N_A]$$

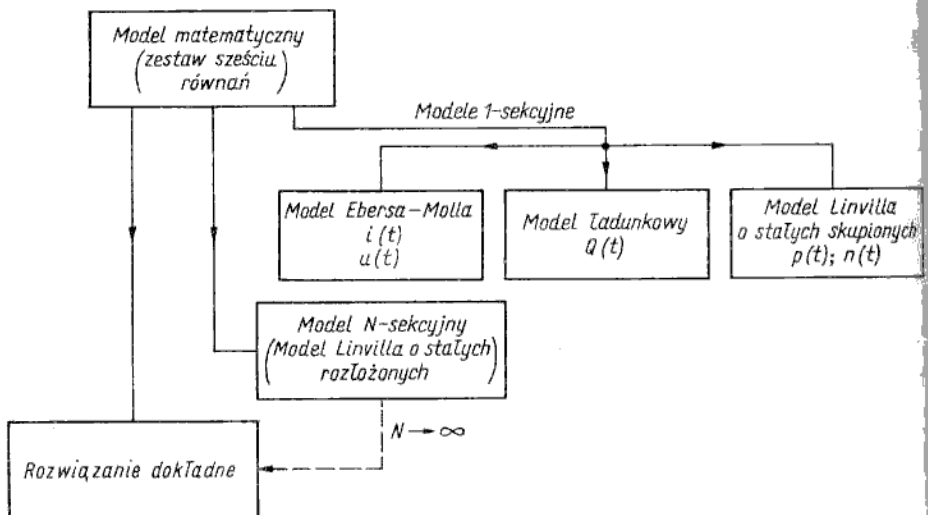
¹⁾ Można go wyznaczyć ze stałej czasowej τ_N , gdyż w pierwszym przybliżeniu $\tau = \beta_N \tau_N$. Wynika to ze wzoru (5.42), jeśli zauważymy, że $\tau \approx \tau_r$ oraz przyjmiemy $\tau_N = t_b$; $\alpha_N = \alpha_b$.

równanie Poissona

$$\frac{\partial^2 \varphi}{\partial x^2} = -\frac{\rho}{\varepsilon}$$

Równania ciągłości uzupełnione pozostałymi czterema równaniami oraz określonymi warunkami granicznymi stanowią pełne sformułowanie zagadnienia. Są to jednakże nieliniowe równania drugiego rzędu z pochodnymi cząstkowymi, które nie mają ogólnego rozwiązania, a po ich linearyzacji rozwiązania mają bardzo złożoną postać szeregów nieskończonych. Tak skomplikowane rozwiązania mają bardzo ograniczoną przydatność w praktyce. Dlatego są opracowywane modele dające bardziej uproszczone, lecz zarazem bardziej przejrzysty opis działania tranzystora. Podstawowa „strategia” modelowania tranzystora polega na usunięciu z równania ciągłości pochodnych po zmiennej x , tj. zastąpieniu równań o pochodnych cząstkowych równaniami różniczkowymi zwyczajnymi.

Na rysunku 5.49 przedstawiono klasyfikację podstawowych modeli tranzystora. Najdokładniejszy (nie licząc bezpośredniego rozwiązania równań różniczkowych meto-



Rys. 5.49

Klasyfikacja podstawowych modeli tranzystora

dami numerycznymi) jest model N -sekcyjny, tzw. *model Linvilla o stałych rozłożonych*. Podział struktury półprzewodnikowej na N sekcji jest równoznaczny z przejściem do przyrostów skończonych, czyli zamianą równań różniczkowych na równania różnicowe (taką procedurę stosuje się przy rozwiązaniu numerycznym równań różnicowych). W ten sposób w każdej sekcji usuwa się zależność od współrzędnej x , czyli równanie o pochodnych cząstkowych jest rozwiązywane w postaci N równań różniczkowych zwyczajnych rzędu pierwszego względem zmiennej t . Linvill zaproponował przedstawienie pojedynczej sekcji półprzewodnika w postaci symbolicznego schematu zastępczego, którego elementy mają sens współczynników występujących w równaniach różnicowych.

Elementy schematu zastępczego Linvilla stanowią również bezpośrednią reprezentację podstawowych procesów fizycznych (dyfuzji, dryftu, generacji itp.). Przy liczbie sekcji dążącej do nieskończoności model N -sekcyjny uzyskuje dokładność identyczną jak bezpośrednio rozwiązanie wyjściowych równań o pochodnych cząstkowych.

Trzy powszechnie znane modele tranzystora bipolarnego, tj.:

- model Linvilla o statych skupionych,
- model ładunkowy (model Beaufoya i Sparkesa),
- model Ebersa-Molla,

stanowią jednakowe przybliżenie, w którym cały obszar bazy jest potraktowany jako jedna sekcja (dwie odrębne sekcje stanowią warstwy zaporowe łączące $E-B$, $B-C$). Wymienione trzy modele różnią się tylko wyborem zależności podstawowych, z których są wyprowadzane wszystkie inne charakterystyki i parametry tranzystora. W modelu Linvilla podstawowymi zależnościami są wartości koncentracji nośników na krawędziach obszaru bazy, wyrażone w funkcji czasu. W modelu ładunkowym podstawową zależnością jest ładunek bazy, wyrażony w funkcji czasu. W modelu Ebersa-Molla podstawowe są zależności prądów i napięć w funkcji czasu.

Ograniczony zakres tej książki nie pozwala szerzej rozwinąć tematyki modelowania tranzystorów bipolarnych. Celem przedstawionego wyżej elementarnego zarysu tego zagadnienia było przede wszystkim zwrócenie uwagi na całkowitą równoważność trzech powszechnie stosowanych modeli. Nie ma zatem sensu pytanie: „który z tych modeli jest dokładniejszy?” (Wszystkie trzy stanowią jednakowo uproszczone jednosekcyjne rozwiązanie równań wyjściowych). Natomiast pytanie: „który z tych modeli jest lepszy?” ma sens, jeżeli „lepszy” będzie się rozumieć jako „dogodniejszy”. Dla fizyka zajmującego się procesami zachodzącymi wewnątrz tranzystora najlepszy może być model Linvilla, w którym występują symboliczne elementy odpowiadające bezpośrednio poszczególnym zjawiskom fizycznym. Dla „układowca” najlepszy jest model Ebersa-Molla, gdyż w postaci bezpośredniej są wyrażone w tym modelu związki między wielkościami występującymi na końcówkach elementu (prądy i napięcia). Model ładunkowy może być dogodny do rozważań na poziomie elementowym jak i układowym, gdyż z jednej strony daje prostą i przejrzystą interpretację działania tranzystora, z drugiej zaś opiera się na łatwo mierzalnych parametrach. Uwzględniając podstawowe przeznaczenie książki skupimy uwagę na modelu Ebersa-Molla, przy czym zgodnie z praktyką stosowaną w modelowaniu dla potrzeb projektowania komputerowego układów posłużymy się częściowo pojęciami wynikającymi z modelu ładunkowego.

Dynamiczny nieliniowy model Ebersa-Molla (zmodyfikowany) powstaje na bazie modelu nieliniowego statycznego (rys. 5.26) przez dołączenie pojemności reprezentujących zjawiska ładowania (rozładowania) obu warstw zaporowych i obszarów neutralnych (głównie obszaru bazy).

Pojemności w tranzystorze bipolarnym

5.8.3.1

Pojemności złączone

Warstwa zaporowa zachowuje się jak kondensator nieliniowy, gdyż zawiera ładunek przestrzenny nieliniowo zależny od napięcia polaryzującego złącze. To zjawisko, jak również związane z nim pojęcie pojemności złączonej, jest znane

z rozdziału 3. W rozdziale 3 wyprowadzono zależność pojemności złączowej od napięcia, którą teraz zapiszemy oddzielnie dla złączy $E-B$ i $B-C$:

$$C_{je} = C_{je0} \left(1 - \frac{U_{BE}}{\varphi_E} \right)^{-n_E} \quad (5.79)$$

$$C_{jc} = C_{jc0} \left(1 - \frac{U_{CE}}{\varphi_C} \right)^{-n_C} \quad (5.80)$$

przy czym: C_{je0} , C_{jc0} — pojemności warstw zaporowych złączy $E-B$ i $B-C$ dla zerowych napięć polaryzacji; φ_E , φ_C — napięcia dyfuzyjne złączy $E-B$ i $B-C$; n_E , n_C — liczby przyjmujące wartości od $1/2$ (złącze skokowe) do $1/3$ (złącze liniowe).

Po wprowadzeniu pojęcia pojemności złączowych można teraz wrócić do określenia czasu opóźnienia t_d związanego z ładowaniem tych pojemności.

W przybliżeniu

$$t_d = R_B (C_{je\bar{s}r} + C_{jc\bar{s}r}) \ln \left(1 + \frac{E_R}{E_F} \right) \quad (5.81)$$

przy czym $C_{je\bar{s}r}$, $C_{jc\bar{s}r}$ — wartości średnie pojemności C_{je} , C_{jc} w określonych przedziałach zmian napięć U_{BE} , U_{CB} .

Pojemności dyfuzyjne

Pojęcie pojemności dyfuzyjnej jest również znane z teorii złącza $p-n$. Pojemność dyfuzyjna jest ogólnie definiowana jako parametr przyrostowy, określający stosunek przyrostu ładunku nośników mniejszościowych nadmiarowych do odpowiedniego przyrostu napięcia

$$C_d = \frac{\Delta Q_{\text{mniej, nadm}}}{\Delta U}$$

Można wymienić dwie przyczyny zmian ładunku nośników mniejszościowych nadmiarowych w tranzystorze:

- wstrzykiwanie nośników (pojemność dyfuzyjną związaną z tym zjawiskiem nazwiemy pojemnością iniekcyjną C_d^I);
- modulacja efektywnej szerokości bazy (pojemność dyfuzyjną związaną z tym zjawiskiem nazwiemy pojemnością modulacyjną C_d^M).

Pojemność iniekcyjna występuje w przypadku polaryzacji złącza w kierunku przewodzenia niezależnie od rodzaju polaryzacji drugiego złącza, pojemność modulacyjna — w przypadku polaryzacji złącza w kierunku zaporowym przy polaryzacji drugiego złącza w kierunku przewodzenia. Zatem pojemność dyfuzyjna emitera

$$C_{de} = \underset{\substack{\text{dla zakresów} \\ \text{normalnego} \\ \text{i nasycenia}}}{C_{de}^I} + \underset{\substack{\text{dla zakresu} \\ \text{inwersyjnego}}}{C_{de}^M} \quad (5.82)$$

i kolektora

$$C_{dc} = \underset{\substack{\text{dla zakresów} \\ \text{inwersyjnego} \\ \text{i nasycenia}}}{C_{dc}^I} + \underset{\substack{\text{dla zakresu} \\ \text{normalnego}}}{C_{dc}^M} \quad (5.83)$$

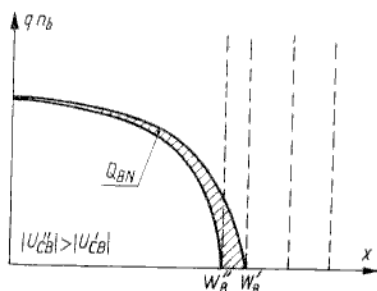
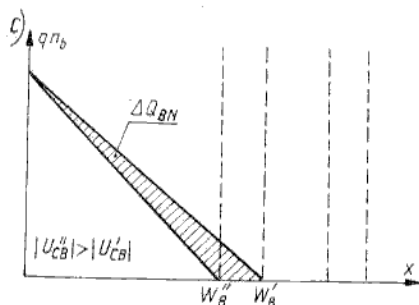
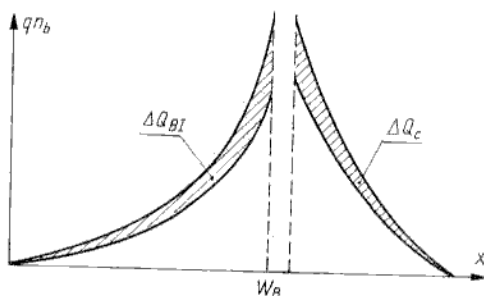
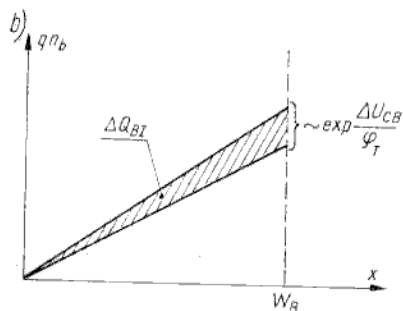
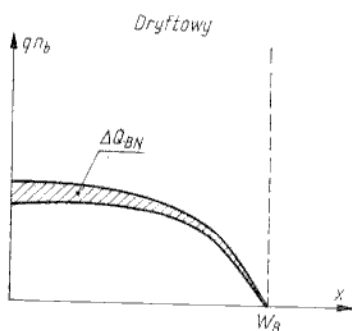
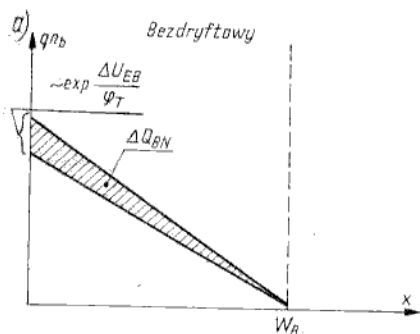
W zakresie zatkania pojemności dyfuzyjne nie występują. Uwzględniając wnioski sformułowane w rozważaniach zawartych na początku p. 5.8 można w następujący spo-

sób zdefiniować pojemności dyfuzyjne dla dwu najważniejszych zakresów: normalnego i nasycenia:

$$C_{de} = C_{de}^I = \frac{\Delta Q_{BN}}{\Delta U_{EB}} \quad (5.84)$$

$$C_{dc} = \frac{\Delta Q_{BI}}{\Delta U_{CB}} + \frac{\Delta Q_C}{\Delta U_{CB}} + \frac{\Delta Q_{BN}}{\Delta U_{CB}} \quad (5.85)$$

$$C_{dc}^I = \frac{\Delta Q_{BI}}{\Delta U_{CB}} + \frac{\Delta Q_{BN}}{\Delta U_{CB}}$$



$$W_B' = W_B(U_{CB}')^2$$

$$W_B'' = W_B(U_{CB}'')$$

Rys. 5.50

Wykresy ilustrujące sens fizyczny pojemności dyfuzyjnych w tranzystorach bezdryftowych i dryftowych: a) pojemność C_{de}^I ; b) pojemność C_{dc}^I ; c) pojemność C_{dc}^M

Wykresy przedstawione na rys. 5.50 ilustrują sens fizyczny tak zdefiniowanych pojemności dla tranzystorów bezdryftowego i dryftowego. Rysunek 5.50a odnosi się do pojemności C_{de}^I , rysunek 5.50b do pojemności C_{dc}^I (w tranzystorze bezdryftowym nie uwzględnia się składowej $\Delta Q_C/\Delta U_{CB}$), a rysunek 5.50c do pojemności C_{dc}^M .

Dla pojemności iniekcyjnych, podobnie jak w przypadku pojemności złącza $p-n$ (rozdz. 3), można napisać następujące zależności:

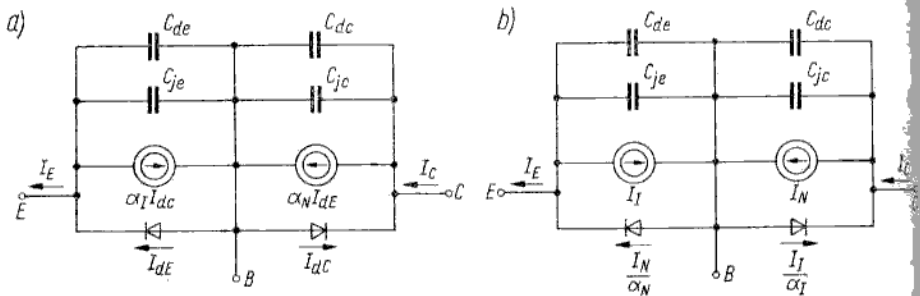
$$C_{de}^I = \frac{\tau_N}{\varphi_T} (I_{dE} + I_{ES}) \quad (5.86)$$

$$C_{dc}^I = \frac{\tau_I}{\varphi_T} (I_{dC} + I_{CS}) \quad (5.87)$$

przy czym τ_N , τ_I — czasy przelotu nośników w kierunkach normalnym i inwersyjnym.

Istnieje ścisła zależność czasów przelotu i częstotliwości granicznych tranzystora. Do tej sprawy powrócimy nieco dalej po omówieniu częstotliwości granicznych.

Pojemność C_{dc}^M jest określona bardziej złożonymi wyrażeniami.



Rys. 5.51

Model Ebersa-Molla dla pracy dynamicznej nieliniowej: a) wariant iniekcyjny; b) wariant transportowy

Model przedstawiony na rys. 5.28 (najprostszy wariant modelu Ebersa-Molla) uzupełniony czterema pojemnościami składowymi (rys. 5.51); równania prądów dla takiego modelu (por. wyrażenia (5.50) do (5.53)) mają postać:

$$i_E = I_{dE} - \alpha_I I_{dC} + (C_{je} + C_{de}) \frac{dU_{BE}}{dt} \quad (5.88)$$

$$i_C = -I_{dC} + \alpha_N I_{dE} + (C_{jc} + C_{dc}) \frac{dU_{BC}}{dt} \quad (5.89)$$

$$i_B = i_E - i_C \quad (5.90)$$

Należy zauważyć, że dokładność modelu przedstawionego na rys. 5.51 jest coraz mniejsza w miarę wzrostu szybkości przełączania tranzystora. Dla procesu przełączania zachodzącego w czasie porównywalnym z τ_N należy w modelu uwzględnić dodatkowe elementy reprezentujące indukcynności doprowadzeń, pojemności rozproszenia itp.

Modele i parametry dla pracy z małym sygnałem

5.9

Zostaną teraz rozpatrzone właściwości tranzystora spolaryzowanego w określonym punkcie pracy i wysterowanego małym sygnałem prądu zmiennego. Mały sygnał jest rozumiany jako sygnał o takiej amplitudzie, przy której można pominąć nieliniowość zależności prądów i napięć na końcówkach tranzystora. Inaczej mówiąc, małe amplitudy prądów i napięć na końcówkach tranzystora można powiązać zależnościami liniowymi, w których współczynniki proporcjonalności nie są funkcjami tych amplitud. Przyjmuje się zatem, że tranzystor sterowany małym sygnałem zachowuje się jak element liniowy. Stąd modele opisujące właściwości małosygnałowe tranzystora nazywa się modelami liniowymi.

Podobnie jak dla modeli nieliniowych można tu wyróżnić pracę statyczną (ściślej biorąc quasi-statyczną w zakresie małych częstotliwości) i dynamiczną (w zakresie dużych częstotliwości). Jednak przede wszystkim należy podzielić modele liniowe tranzystora na dwie grupy, tj.:

- *modele końcówkowe* (sformalizowane obwodowo modele czwórnikowe),
- *modele fizyczne* (schematy zastępcze zbudowane z elementów odpowiadających określonym procesom fizycznym w tranzystorze).

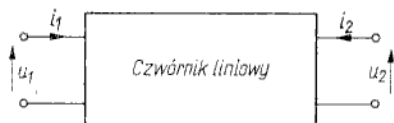
Oba rodzaje modeli omówimy dla przypadku małych i dużych częstotliwości.

Ponieważ sterowanie małym sygnałem występuje przede wszystkim we wzmacniaczach, będzie nas interesować tylko praca tranzystora w zakresie aktywnym normalnym.

Modele czwórnikowe

5.9.1

Tranzystor w dowolnym układzie włączenia może być przedstawiony przy pracy małosygnałowej jako czwórnik liniowy, którego właściwości są opisane przez zależności liniowe prądów i napięć wejściowych i wyjściowych (rys. 5.52). W p.



Rys. 5.52
Czwórnik liniowy

5.7.2 stwierdzono już, że istnieje sześć różnych par równań czwórnika, przy czym każda para równań równorzędnie i jednoznacznie opisuje jego właściwości. Stwierdzono również, że znaczenie praktyczne mają trzy pary równań, które dla czwórnika liniowego przyjmują postać:

$$\left. \begin{aligned} u_1 &= z_{11} i_1 + z_{12} i_2 \\ u_2 &= z_{21} i_1 + z_{22} i_2 \end{aligned} \right\} \text{równania impedancyjne} \quad (5.91)$$

$$\left. \begin{aligned} i_1 &= y_{11} u_1 + y_{12} u_2 \\ i_2 &= y_{21} u_1 + y_{22} u_2 \end{aligned} \right\} \text{równania admitancyjne} \quad (5.92)$$

$$\left. \begin{aligned} u_1 &= h_{11} i_1 + h_{12} u_2 \\ i_2 &= h_{21} i_1 + h_{22} u_2 \end{aligned} \right\} \text{równania mieszane (hybrydowe)} \quad (5.93)$$

przy czym i_1 , i_2 , u_1 , u_2 — wartości chwilowe prądów i napięć małych sygnałów zmiennych na wejściu i wyjściu czwórnika.

Równania (5.91) do (5.93) można zapisać w postaci macierzowej

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = z \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \quad (5.94)$$

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = y \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad (5.95)$$

$$\begin{bmatrix} u_1 \\ i_2 \end{bmatrix} = h \begin{bmatrix} i_1 \\ u_2 \end{bmatrix} \quad (5.96)$$

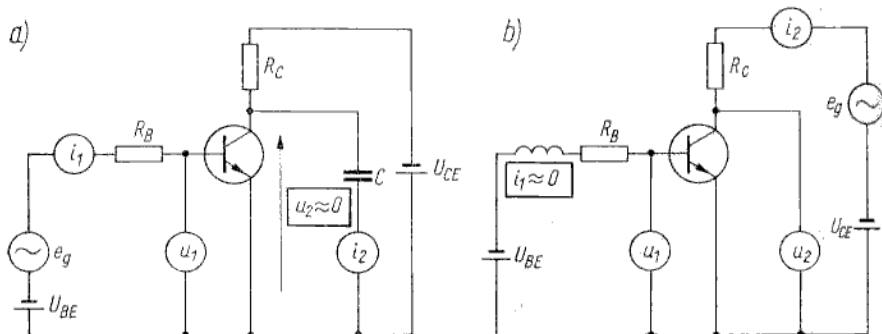
przy czym macierze z , y , h mają postać:

$$z = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix}; \quad y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix}; \quad h = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \quad (5.97)$$

Właściwości czwórnika liniowego są zatem określone macierzą czterech parametrów *impedancyjnych* z bądź *admitancyjnych* y lub *hybrydowych* h . Na podstawie równań (5.91) do (5.93) można zdefiniować poszczególne parametry i określić warunki, w jakich należy je mierzyć.

Parametry impedancyjne są mierzone w warunkach rozwarcia wejścia lub wyjścia czwórnika. Na przykład $z_{11} = u_1/i_1$ przy $i_2 = 0$, czyli przy rozwarciu wyjścia, a $z_{22} = u_2/i_2$ przy $i_1 = 0$, czyli przy rozwarciu wejścia itd.

Parametry admitancyjne są mierzone przy zwarceniu wejścia lub wyjścia, a parametry hybrydowe przy rozwarciu wejścia lub zwarceniu wyjścia czwórnika. Należy podkreślić, że nie chodzi tu o zwarcie i rozwarcie w sensie galwanicznym, gdyż uniemożliwiłoby to spolaryzowanie tranzystora w określonym punkcie pracy. Chodzi tu wyłącznie o zapewnienie specyficznych warunków sterowania i obciążenia tranzystora dla sygnału zmiennego. Przykładowo rozwarcie wejścia oznacza, że z zacisków wejściowych czwórnik „widzi” impedancję znacznie większą niż jego impedancja wejściowa. Zwarcie wyjścia oznacza, że z zacisków wyjściowych czwórnik „widzi” impedancję znacznie mniejszą niż jego impedancja wyjściowa. Dla zilustrowania praktycznych sposobów realizacji tych warunków na rys. 5.53 przedstawiono przykładowe układy włączenia tranzystora, stosowane przy pomiarze parametrów typu h .



Rys. 5.53

Układy włączenia tranzystora, stosowane przy pomiarze parametrów typu h

Parametry macierzy h i y są najczęściej stosowane w praktyce (nieco dalej wyjaśnimy dlaczego), przedstawimy je zatem bardziej szczegółowo. Na podstawie (5.93):

$$h_{11} = \left. \frac{u_1}{i_1} \right|_{u_2=0} \quad \text{— impedancja wejściowa} \quad (5.98)$$

$$h_{12} = \left. \frac{u_1}{u_2} \right|_{i_1=0} \quad \text{— współczynnik oddziaływania zwrotnego} \quad (5.99)$$

$$h_{21} = \left. \frac{i_2}{i_1} \right|_{u_2=0} \quad \text{— współczynnik wzmocnienia prądowego} \quad (5.100)$$

$$h_{22} = \left. \frac{i_2}{u_2} \right|_{i_1=0} \quad \text{— admitancja wyjściowa} \quad (5.101)$$

Na podstawie (5.92):

$$y_{11} = \left. \frac{i_1}{u_1} \right|_{u_2=0} \quad \text{— admitancja wejściowa} \quad (5.102)$$

$$y_{12} = \left. \frac{i_1}{u_2} \right|_{u_1=0} \quad \text{— admitancja zwrotna} \quad (5.103)$$

$$y_{21} = \left. \frac{i_2}{u_1} \right|_{u_2=0} \quad \text{— admitancja przejściowa (transadmitancja)} \quad (5.104)$$

$$y_{22} = \left. \frac{i_2}{u_2} \right|_{u_1=0} \quad \text{— admitancja wyjściowa} \quad (5.105)$$

Należy przy tym zauważyć, że

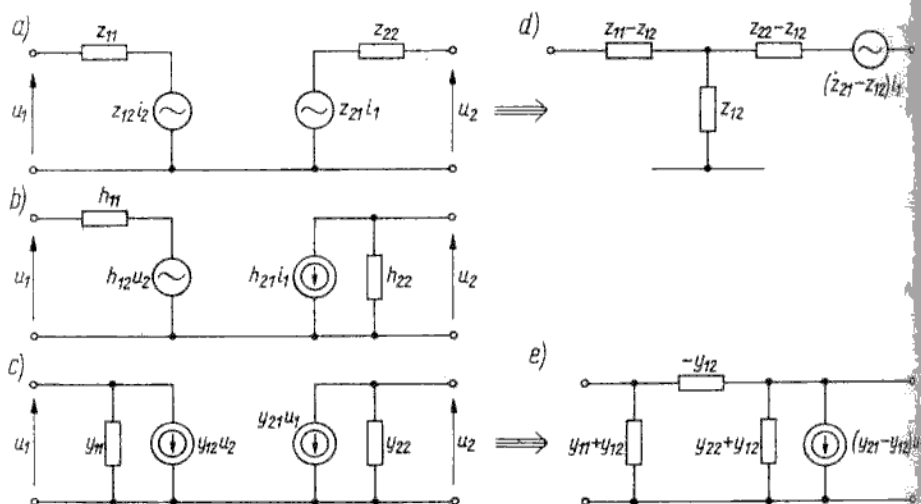
$$h_{11} = \frac{1}{y_{11}}$$

lecz

$$h_{22} \neq y_{22}$$

gdyż h_{22} jest admitancją wyjściową mierzoną przy $i_1 = 0$ (tzw. rozwarciowa admitancja wyjściowa), a y_{22} jest admitancją wyjściową, mierzoną przy $u_1 = 0$ (tzw. zwarcziowa admitancja wyjściowa).

Każdą parę równań czwórnik można przedstawić w postaci równoważnej sieci elektrycznej. Trzy sieci równoważne równaniom impedancyjnym, hybrydowym i admittancyjnym są przedstawione odpowiednio na rys. 5.54a, b, c. Każda z tych sieci zawiera cztery elementy, tj. impedancje (lub admitancje) wejściową i wyjściową oraz dwa źródła sterowane, z których jedno reprezentuje zjawisko wzmocnienia ($z_{21}i_1$, $h_{21}i_1$ lub $y_{21}u_1$), drugie zaś — zjawisko oddziaływania wstecznego ($z_{12}i_2$, $h_{12}u_2$ lub $y_{12}u_2$). Korzystając z metod teorii liniowych obwodów elektrycznych można te sieci przekształcić (w pewnym sensie uprościć) w taki sposób, by zawierały tylko jedno źródło sterowane. Otrzymuje się wówczas sieć typu T (rys. 5.54d) lub typu π (rys. 5.54e). Sieć typu T jest złożona z trzech impedancji i źródła napięciowego, a sieć typu π z trzech admitancji i źródła prą-



Rys. 5.54

Trzy sieci równoważne równaniom: impedancyjnym (a), hybrydowym (b) i admitancyjnym (c)

dowego. Równoważność sieci przedstawionych na rys. 5.54a, d lub 5.54c, e można łatwo sprawdzić dokonując rozwarcia lub zwarcia odpowiednich zacisków i porównując rezultaty. Przykładowo, jeżeli w sieci na rys. 5.54c zewrze się

Związki tożsamościowe między parametrami macierzy z , h , y

Tablica 5.1

Macierz	z	h	y
z	$z_{11} \quad z_{12}$	$\frac{\Delta h}{h_{22}} \quad \frac{h_{12}}{h_{22}}$	$\frac{y_{22}}{\Delta y} \quad \frac{-y_{12}}{\Delta y}$
	$z_{21} \quad z_{22}$	$\frac{-h_{21}}{h_{22}} \quad \frac{1}{h_{22}}$	$\frac{-y_{21}}{\Delta y} \quad \frac{y_{11}}{\Delta y}$
h	$\frac{\Delta z}{z_{22}} \quad \frac{z_{12}}{z_{22}}$	$h_{11} \quad h_{12}$	$\frac{1}{y_{11}} \quad \frac{-y_{12}}{y_{11}}$
	$\frac{-z_{21}}{z_{22}} \quad \frac{1}{z_{22}}$	$h_{21} \quad h_{22}$	$\frac{y_{21}}{y_{11}} \quad \frac{\Delta y}{y_{11}}$
y	$\frac{z_{22}}{\Delta z} \quad \frac{-z_{12}}{\Delta z}$	$\frac{1}{h_{11}} \quad \frac{-h_{12}}{h_{11}}$	$y_{11} \quad y_{12}$
	$\frac{-z_{21}}{\Delta z} \quad \frac{z_{11}}{\Delta z}$	$\frac{h_{21}}{h_{11}} \quad \frac{\Delta h}{h_{11}}$	$y_{21} \quad y_{22}$

$$\Delta z = z_{11}z_{22} - z_{12}z_{21}$$

$$\Delta h = h_{11}h_{22} - h_{12}h_{21}$$

$$y = y_{11}y_{22} - y_{12}y_{21}$$

wyjscie, to pozostanie admitancja wejściowa y_{11} , jeżeli natomiast zewrze się wejście ($u_1 = 0$) to pozostanie admitancja wyjściowa y_{22} itp. Wszystkie pięć sieci przedstawionych na rys. 5.54 są wzajemnie równoważne (wprawdzie strzałki oznaczające równoważność umieszczono tylko między sieciami a), d) oraz c), e), lecz uczyniono tak wyłącznie dla podkreślenia jednakowego rodzaju elementów biernych w tych sieciach, tj. impedancji w pierwszym przypadku i admitancji w drugim). Równoważność poszczególnych sieci wynika z równoważności poszczególnych par równań czwórnika, a więc istnieją również związki tożsamościowe między parametrami macierzy różnych typów (tabl. 5.1).

Wartości parametrów czwórnikowych zależą od układu włączenia tranzystora. Przykładowo parametr h_{21} , wyznaczający wzmocnienie prądowe, przyjmuje następujące wartości:

- w układzie WB — α (ok. 0,99)
- w układzie WE — β (ok. 100)
- w układzie WC — $(\beta + 1)$

Rodzaj układu włączenia jest oznaczany indeksem literowym:

- b — dla układu WB
- e — dla układu WE
- c — dla układu WC

Na przykład h_{11b} jest impedancją wejściową w układzie WB, h_{21e} jest wzmocnieniem prądowym (β) w układzie WE itd.

W literaturze angielskiej i w katalogach często indeks dwucyfrowy zastępuje się jedną literą według następującego klucza (na przykładzie parametrów typu h):

- $h_{11} - h_i$ (input — wejście)
- $h_{12} - h_r$ (reverse — oddziaływanie zwrotne)
- $h_{21} - h_f$ (forward — przenoszenie w przód)
- $h_{22} - h_o$ (output — wyjście).

Zatem:

- $h_{ib} \equiv h_{11b}$
- $h_{fe} \equiv h_{21e}$
- itp.

Zależności między parametrami macierzy h w trzech układach włączenia

Tablica 5.2

Parametr	Układ włączenia		
	WE	WB	WC
h_{11}	h_{11e}	$h_{11b} = \frac{h_{11e}}{1 + h_{21e}}$	$h_{11c} = h_{11e}$
h_{12}	h_{12e}	$h_{12b} = \frac{h_{11e} h_{22e}}{1 + h_{21e}} - h_{12e}$	$h_{12c} = 1 - h_{12e}$
h_{21}	h_{21e}	$h_{21b} = \frac{-h_{21e}}{1 + h_{21e}}$	$h_{21c} = (1 + h_{21e})$
h_{22}	h_{22e}	$h_{22b} = \frac{h_{22e}}{1 + h_{21e}}$	$h_{22c} = h_{22e}$

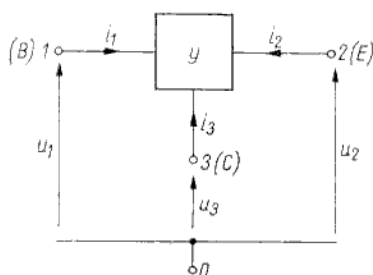
Ponieważ właściwości tranzystora są jednoznacznie opisane jednym zestawem czterech parametrów czwórnikowych, a zmiana układu włączenia nie oznacza zmiany sposobu działania tranzystora, oczywiste jest istnienie związków tożsamościowych między parametrami dla poszczególnych układów włączenia. Zależności te podano w tabl. 5.2 dla macierzy h . Na przykład znając parametry typu h dla układu WE można wyznaczyć parametry tej macierzy dla układów WB i WC, a korzystając ponadto z tabl. 5.1 można nawet wyznaczyć parametry dowolnej macierzy dla dowolnego układu włączenia tranzystora. Przeliczanie parametrów czwórnikowych z jednej konfiguracji na inną jest niezmiernie proste przy posługiwaniu się tzw. *admitancyjną macierzą nieoznaczoną*. Jest to zagadnienie tak interesujące, że wydzielimy je w postaci odrębnego punktu.

Admitancyjna macierz nieoznaczona

5.9.1

W teorii liniowych obwodów elektrycznych udowadnia się, że macierz nieoznaczona admitancji trójkąta (rys. 5.55):

$$y = \begin{matrix} & \begin{matrix} B & E & C \end{matrix} \\ \begin{matrix} B \\ E \\ C \end{matrix} & \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix} \end{matrix} \begin{matrix} B \\ E \\ C \end{matrix} \quad (5.106)$$



Rys. 5.55
Trójkąt liniowy reprezentujący tranzystor

ma tę właściwość, iż suma admitancji w każdym wierszu i w każdej kolumnie jest równa zero. Macierz oznaczoną dla określonej konfiguracji otrzymuje się z macierzy nieoznaczonej przez skreślenie wiersza i kolumny oznaczonych literą określającą wspólną końcówkę w tej konfiguracji. Dla układu WB skreśla się więc pierwszy wiersz i pierwszą kolumnę, dla WE — drugi wiersz i drugą kolumnę itp. Przykładowo dla układu WE

$$y_e = \begin{bmatrix} y_{11} & y_{13} \\ y_{31} & y_{33} \end{bmatrix} = \begin{bmatrix} y_{11e} & y_{12e} \\ y_{21e} & y_{22e} \end{bmatrix}$$

Jeżeli są znane parametry macierzy oznaczonej dla jednej konfiguracji (na przykład y_e dla konfiguracji WE), to łatwo można wyznaczyć macierz nieoznaczoną, czyli znaleźć parametry czwórnikowe dla dwu pozostałych konfiguracji. W tym celu należy wpisać wyrazy macierzy y_e w odpowiednich pozycjach macierzy nieoznaczonej y i uzupełnić puste pozycje w tej macierzy przez wpisanie takich wartości, by suma każdego wiersza i kolumny była równa zero. Otrzymuje się w ten sposób:

$$y = \begin{bmatrix} B & E & C \\ y_{11e} & -(y_{11e} + y_{12e}) & y_{12e} \\ -(y_{11e} + y_{21e}) & y_{11e} + y_{12e} + y_{21e} + y_{22e} & -(y_{12e} + y_{22e}) \\ y_{21e} & -(y_{21e} + y_{22e}) & y_{22e} \end{bmatrix} \begin{matrix} B \\ E \\ C \end{matrix} \quad (5.107)$$

Stąd na przykład dla układu WC można określić następujące relacje (skreśla się trzeci wiersz i trzecią kolumnę):

$$\left. \begin{aligned} y_{11c} &= y_{11e} \\ y_{12c} &= -(y_{11e} + y_{12e}) \\ y_{21c} &= -(y_{11e} + y_{21e}) \\ y_{22c} &= y_{11e} + y_{12e} + y_{21e} + y_{22e} \end{aligned} \right\} \quad (5.108)$$

Parametry macierzy rozproszenia

5.9.1.2

Parametry macierzy z , h , y są wielkościami zespolonymi, zależnymi od częstotliwości. W miarę wzrostu częstotliwości coraz trudniej jest zapewnić warunki rozwarcia lub zwarcia odpowiednich końcówek tranzystora i coraz większy jest wpływ pasożytniczych pojemności obudowy oraz indukcyjności doprowadzeń. Dlatego pomiar tych parametrów w zakresie częstotliwości powyżej 300 MHz jest bardzo utrudniony, co skłania do tego, by w tym zakresie częstotliwości określać parametry tranzystora w stanie roboczym przy określonych wartościach impedancji źródła sterującego i obciążenia. W taki sposób są mierzone *parametry macierzy rozproszenia*. Ponieważ w technice mikrofalowej stosuje się standardowe linie przesyłowe o impedancji falowej 50Ω , zatem parametry rozproszenia są mierzone przy takich wartościach impedancji źródła i obciążenia.

W macierzy rozproszenia zamiast napięć (u_1 , u_2) i prądów (i_1 , i_2) stosuje się ich liniowe kombinacje. Zmiennymi niezależnymi są padające fale napięć (rys. 5.56):



Rys. 5.56

Czwórnik liniowy opisany parametrami s

$$a_1 = \frac{1}{2\sqrt{Z_L}} (u_1 + Z_L i_1) \quad (5.109)$$

$$a_2 = \frac{1}{2\sqrt{Z_L}} (u_2 + Z_L i_2) \quad (5.110)$$

Zmiennymi zależnymi są odbite fale napięć:

$$b_1 = \frac{1}{2\sqrt{Z_L}} (u_1 - Z_L i_1) \quad (5.111)$$

$$b_2 = \frac{1}{2\sqrt{Z_L}} (u_2 - Z_L i_2) \quad (5.112)$$

przy czym Z_L , praktycznie biorąc, zawsze jest równe 50Ω .

Macierz rozproszenia ma postać:

$$\begin{cases} b_1 = s_{11}a_1 + s_{12}a_2 \\ b_2 = s_{21}a_1 + s_{22}a_2 \end{cases} \quad (5.113)$$

Stąd poszczególne parametry są definiowane w następujący sposób:

$$s_{11} = \left. \frac{b_1}{a_1} \right|_{a_2=0} \quad \text{— współczynnik odbicia na wejściu czwórnika przy obciążeniu jego wyjścia impedancją } 50 \Omega \text{ (dopasowanie na wyjściu)} \quad (5.114)$$

$$s_{12} = \left. \frac{b_1}{a_2} \right|_{a_1=0} \quad \text{— współczynnik transmisji wstecznej przy dopasowaniu na wejściu} \quad (5.115)$$

$$s_{21} = \left. \frac{b_2}{a_1} \right|_{a_2=0} \quad \text{— współczynnik przenoszenia przy dopasowaniu na wyjściu} \quad (5.116)$$

$$s_{22} = \left. \frac{b_2}{a_2} \right|_{a_1=0} \quad \text{— współczynnik odbicia na wyjściu czwórnika przy dopasowaniu na wejściu} \quad (5.117)$$

Na podstawie parametrów s_{21} , s_{12} można łatwo wyznaczyć obustronną transmisję mocy, gdyż $|s_{21}|^2$ jest wzmocnieniem skutecznym mocy, a $|s_{12}|^2$ jest wstecznym wzmocnieniem skutecznym mocy. Istnieją tożsamościowe związki między parametrami macierzy rozproszenia a parametrami macierzy z , h , y .

Porównanie parametrów czwórnikowych dla różnych rodzajów macierzy

5.9.13

Przy całkowitej równoważności wszystkich rodzajów macierzy tranzystora traktowanego jako czwórnik liniowy można jednak dokonać wyboru określonego typu macierzy, najbardziej dogodnego w pewnym obszarze zastosowań. Kryteria wyboru najdogodniejszej macierzy są następujące:

- łatwość pomiaru parametrów;
- łatwość analizy układu, w którym pracuje tranzystor opisywany daną macierzą;
- prosta postać związków parametrów macierzy z parametrami modelu fizycznego.

Kryteria te zastosujemy do rozpatrywanych macierzy typu z , h , y i s . Parametry macierzy z w zasadzie nie są obecnie stosowane, gdyż ich pomiar wymaga spełnienia warunków rozwarcia wejścia lub wyjścia czwórnika. Realizacja praktyczna takich warunków pomiaru jest bardzo trudna w odniesieniu do wyjścia z uwagi na bardzo dużą impedancję wyjściową tranzystora, przyjmującą wartości rzędu megomów (rozwarcie wyjścia oznacza, że impedancja obciążenia powinna być wielokrotnie większa niż impedancja wyjściowa tranzystora). Parametry typu z były dość szeroko stosowane w latach pięćdziesiątych ze względu na dużą popularność w tym okresie schematów zastępczych typu T, czyli modeli o strukturze identycznej z siecią elektryczną macierzy z (por. rys. 5.54d).

Najłatwiejszy jest pomiar parametrów macierzy h , gdyż w tym przypadku wymagane jest rozwarcie wejścia lub zwarcie wyjścia. Oba te warunki można spełnić bez trudu, ponieważ tranzystor ma małą impedancję wejściową i dużą impedancję wyjściową. Należy przy tym zauważyć, że w kaskadowym połączeniu tranzystorów (typowa struktura wielostopniowego wzmacniacza małej częstotliwości) warunki sterowania (sterowanie z poprzedniego tranzystora, czyli ze źródła o du-

zej impedancji wewnętrznej) oraz obciążenia (mała impedancja wejściowa następnego tranzystora) są zbliżone do tych, w jakich dokonuje się pomiaru parametrów h . Zatem wartości parametrów h są bardzo zbliżone do wartości parametrów roboczych tranzystora pracującego w układzie kaskadowym. Jeżeli uwzględnimy ponadto, że najważniejszy parametr tranzystora, jakim jest wzmocnienie prądowe, występuje w macierzy h w postaci bezpośredniej jako parametr h_{21} , to uzyskamy w miarę pełny obraz zalet tego zestawu parametrów. Parametry macierzy h są najchętniej stosowane w zakresie małych częstotliwości, w którym można je wyrazić w postaci liczb rzeczywistych. W zakresie dużych częstotliwości, przy zauważalnych przesunięciach fazowych w tranzystorze, najczęściej są stosowane parametry macierzy y , wyrażone w postaci liczb zespolonych. Są to parametry bardzo dogodnie do analizy układów, w których występują gałęzie równoległe. Admitancje tych gałęzi po prostu dodaje się do odpowiednich parametrów admitancyjnych tranzystora. Parametry macierzy y mają prostą interpretację fizyczną, gdyż struktura sieci elektrycznej tej macierzy (struktura typu π — rys. 5.54e) jest identyczna ze strukturą najpowszechniej stosowanego modelu tranzystora, tzw. modelu „hybryd π ”. Istotne znaczenie ma również fakt, że taką samą strukturę mają schematy zastępcze tranzystorów polowych, co ułatwia jednolite podejście do zagadnień analizy układów niezależnie od rodzaju stosowanych tranzystorów.

W zakresie bardzo dużych częstotliwości (powyżej kilkuset megaherców) są stosowane parametry macierzy s ze względu na łatwość ich pomiaru.

Modele fizyczne (schematy zastępcze)

5.9.2

Schematy zastępcze tranzystora są to modele elektryczne alternatywne w stosunku do modeli końcówkowych, tj. macierzy czwórnika liniowego i ich sieci. O ile modele końcówkowe opisują zewnętrzne zachowanie się tranzystora traktowanego jako „czarna skrzynka”, o tyle modele fizyczne (schematy zastępcze) stanowią mniej lub bardziej wierne odbicie zjawisk zachodzących wewnątrz tranzystora. Taka jest podstawowa różnica między tymi dwoma sposobami modelowania tranzystora i z niej wynikają wszelkie dalsze wnioski porównawcze.

Przed przystąpieniem do omawiania modeli fizycznych warto sformułować podstawowe postulaty, jakim powinien odpowiadać prawidłowo skonstruowany schemat zastępczy. Oprócz postulatów dotyczących schematów zastępczych w ogólności, jak np.:

- jak największa dokładność przy jak najmniejszej złożoności modelu, czyli optymalna (nie za duża i nie za mała) dokładność dla danego zakresu zastosowań;
- łatwość pomiaru wartości elementów składowych modelu; wymienimy niektóre ważne dla rozpatrywanej w tym miejscu klasy modeli (tj. modeli małosygnałowych):
- elementy składowe modelu powinny być niezależne od częstotliwości;
- łatwość wzajemnych przekształceń parametrów schematu zastępczego i parametrów macierzy czwórnika liniowego (zgodność struktury schematu zastępczego z siecią czwórnika liniowego).

Większość znanych schematów zastępczych tranzystora bipolarnego ma strukturę typu T lub π , czyli spełniają one ostatni postulat. Modele typu T mają strukturę zgodną z siecią elektryczną macierzy z , natomiast modele typu π — z siecią macierzy y . Wiele lat trwały spory między zwolennikami tych dwu odmiennych koncepcji

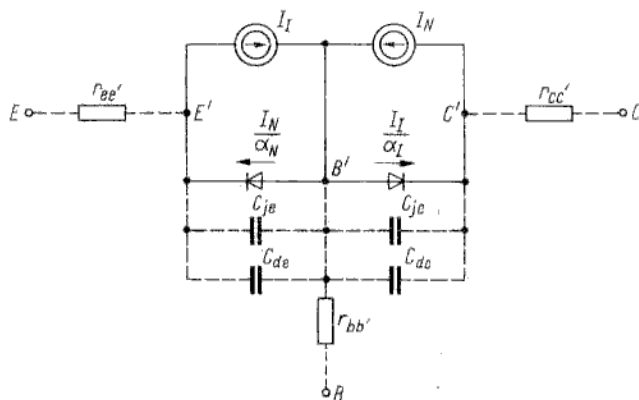
struktury schematu zastępczego. Należy tu zwrócić uwagę na niezasadność wszelkiej argumentacji zmierzającej do wykazania większej dokładności jednej z tych koncepcji. Generalnie dokładność modelu zależy od liczby jego stopni swobody, czyli od liczby elementów składowych (ściślej biorąc, od liczby parametrów niezależnych). Dwa prawidłowo skonstruowane modele o jednakowej liczbie elementów składowych mają identyczną dokładność, jeżeli elementy składowe są opisane z jednakową dokładnością. Wybór jednej z dwu koncepcji schematu zastępczego może być podyktowany tylko względami praktycznymi, jak np. większą łatwością pomiaru parametrów modelu (wartości elementów składowych), prostotą metod analizy obwodu, w skład którego wchodzi tranzystor reprezentowany przez dany schemat zastępczy itp. Szereg tego rodzaju czynników decyduje o tym, że w wielu zastosowaniach bardziej dogodny jest model o strukturze typu π , przy czym istotny wpływ na wybór tej struktury ma jej zgodność z siecią macierzy y (por. porównanie macierzy z , h , y i s w p. 5.9.1.3).

W punkcie 5.8 stwierdzono, że model małosygnałowy (liniowy) można utworzyć na podstawie bardziej ogólnego modelu dynamicznego wielkosygnałowego (nieliniowego). W tym celu należy zastąpić elementy nieliniowe modelu wielkosygnałowego odpowiednimi elementami liniowymi, czyli w sensie matematycznym należy linearyzować funkcje nieliniowe w małych przedziałach.

Przekształcenie modelu nieliniowego w liniowy

5.9.2

Przedstawmy schemat zastępczy z rys. 5.51b w bardziej dogodnej postaci na rys. 5.57. Zadanie rozwiążemy w dwu etapach, tj. najpierw przekształcimy schemat zastępczy pokazany na rys. 5.57 tak, by miał on strukturę typu π , po czym



Rys. 5.57

Schemat zastępczy dla pracy dynamicznej nieliniowej, narysowany w postaci dogodnej do dalszych przekształceń

zamienimy elementy nieliniowe na liniowe. Dla ułatwienia rozważań weźmiemy najpierw pod uwagę tylko gałęzie narysowane linią ciągłą.

Równania prądów końcówkowych mają postać (prądy wpływające do tranzystora mają znak plus, wypływające zaś minus, czyli $I_E + I_C + I_B = 0$):

$$I_E = I_I - \frac{I_N}{\alpha_N} \quad (5.118)$$

$$I_C = I_N - \frac{I_I}{\alpha_I} \quad (5.119)$$

$$I_B = \frac{I_N}{\alpha_N} + \frac{I_I}{\alpha_I} - I_N - I_I \quad (5.120)$$

W równaniu (5.118) do prawej strony dodaje się i odejmuje I_N , w równaniu (5.119) dodaje i odejmuje I_I , w równaniu (5.120) dodaje się i odejmuje $(I_N + I_I)$. W ten sposób otrzymuje się:

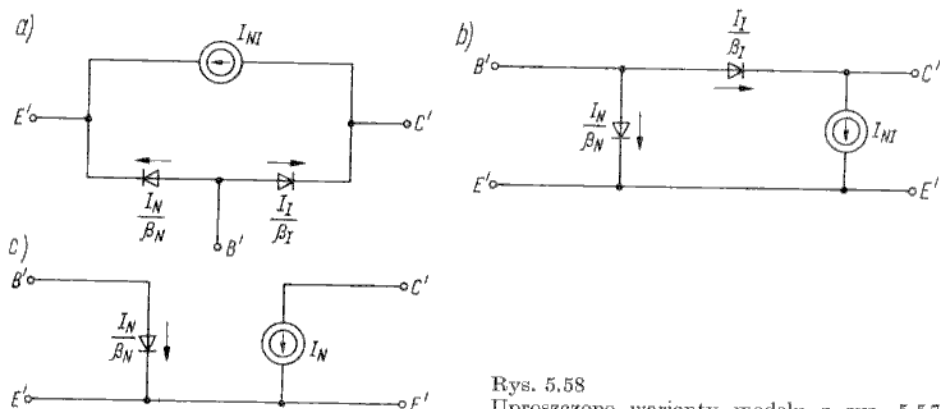
$$I_E = -(I_N - I_I) - \frac{I_N}{\beta_N} \quad (5.121)$$

$$I_C = (I_N - I_I) - \frac{I_I}{\beta_I} \quad (5.122)$$

$$I_B = \frac{I_N}{\beta_N} + \frac{I_I}{\beta_I} \quad (5.123)$$

przy czym:

$$\beta_N = \frac{\alpha_N}{1 - \alpha_N} \quad \beta_I = \frac{\alpha_I}{1 - \alpha_I}$$



Rys. 5.58
Uproszczone warianty modelu z rys. 5.57

Na rysunku 5.58a przedstawiono sieć elektryczną spełniającą równania (5.121) do (5.123), przy czym wprowadzono źródło prądowe

$$I_{NI} = I_N - I_I$$

Na rysunku 5.58b przedstawiono tę samą sieć przerysowaną w postaci tradycyjnie przyjętej dla struktury typu π przy włączeniu tranzystora w układzie WE (przed wszystkim rozpatrzmy konfigurację WE, gdyż ma ona najważniejsze znaczenie praktyczne). Ponieważ interesuje nas praca tranzystora tylko w zakresie aktywnym normalnym, z uwagi na brak wstrzykiwania nośników z kolektora można przyjąć, że:

$$I_I \approx 0$$

$$I_{NI} \approx I_N$$

Przy uwzględnieniu tych uproszczeń schemat zastępczy ma postać przedstawioną na rys. 5.58c. Schemat zastępczy o tak prostej strukturze można było zaproponować intuicyjnie, wychodząc z elementarnych zasad działania tranzystora (pouczająca jest konfrontacja rys. 5.58c z rys. 5.5).

Przejdźmy teraz do drugiego etapu rozważanego zagadnienia, tj. do linearyzacji elementów składowych modelu.

Zapisujemy małe przyrosty prądów bazy i kolektora przyjmując z dużym uproszczeniem że oba te prądy zależą tylko od napięcia wejściowego $U_{B'E'}$:

$$dI_B = dU_{B'E'} \frac{\partial I_B}{\partial U_{B'E'}} \quad (5.124)$$

$$dI_C = dU_{B'E'} \frac{\partial I_C}{\partial U_{B'E'}} \quad (5.125)$$

Po zdefiniowaniu następujących pojęć:

$$g_{b'e'} = \frac{\partial I_B}{\partial U_{B'E'}} \text{ --- konduktacja wejściowa} \quad (5.126)$$

$$g_m = \frac{\partial I_C}{\partial U_{B'E'}} \text{ --- transkonduktancja} \quad (5.127)$$

oraz zamianie małych przyrostów prądów i napięcia (dI_B , dI_C , $dU_{B'E'}$) na amplitudy małych sygnałów zmiennych (i_b , i_c , $u_{b'e'}$) równania (5.124) i (5.125) przyjmują postać:

$$i_b = u_{b'e'} g_{b'e'} \quad (5.128)$$

$$i_c = u_{b'e'} g_m \quad (5.129)$$



Rys. 5.59

Schemat zastępczy liniowy, otrzymany przez linearyzację elementów schematu z rys. 5.58c

Sieć elektryczną spełniającą równania (5.128), (5.129) przedstawiono na rys. 5.59. Jest to schemat zastępczy liniowy, którego elementy można wyznaczyć na podstawie strukturalnie równoważnego mu schematu nieliniowego (rys. 5.58c).

Ponieważ przy $I_I = 0$

$$|I_N| = \alpha_N I_E = \alpha_N I_{ES} \left(\exp \frac{U_{E'B'}}{\varphi_T} - 1 \right) \quad (5.130)$$

$$I_B = \frac{I_N}{\beta_N} \quad (5.131)$$

$$I_C = I_N \quad (5.132)$$

zatem przyjmując dla uproszczenia, że

$$\alpha_N \neq f(I_E, U_{CB})$$

czyli

$$\alpha_N = \alpha = \text{const}$$

po podstawieniu zależności (5.130)—(5.132) do (5.126), (5.127) otrzymuje się

$$g_{b'e'} = (1 - \alpha) \frac{I_E}{\varphi_T} \quad (5.133)$$

$$g_m = \alpha \frac{I_E}{\varphi_T} \quad (5.134)$$

Dzieląc te równania stronami można zauważyć, że istnieje również związek

$$g_{b'e'} = \frac{g_m}{\beta} \quad (5.135)$$

Uwzględnijmy teraz w schemacie zastępczym gałęzie zaznaczone na rys. 5.57 linią przerywaną.

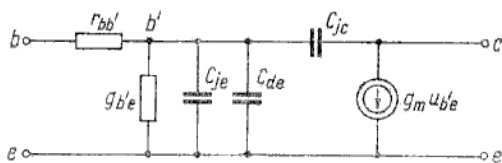
Wpływ rezystancji szeregowych r_{ee}, r_{cc} można uznać za pomijalnie mały oraz, ze względu na pracę tranzystora w obszarze aktywnym normalnym, można pominąć pojemność dyfuzyjną kolektora C_{dc} . (Z p. 5.8.3.1 wiadomo, że $C_{dc} = C_{dc}^I + C_{dc}^M$, przy czym w tym zakresie pracy $C_{dc}^I = 0$, a $C_{dc}^M \ll C_{jc}$). Zatem:

$$r_{ee} \approx 0$$

$$r_{cc} \approx 0$$

$$C_{dc} \approx 0$$

i schemat zastępczy z uwzględnieniem pozostałych elementów ma postać przedstawioną na rys. 5.60. Taki model tranzystora jest nazywany „hybrid π ”, gdyż



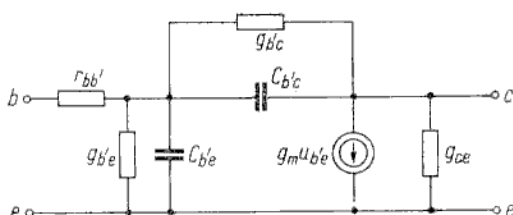
Rys. 5.60
Model „hybrid π ” dla układu WE

stanowi połączenie rezystora $r_{bb'}$ z siecią o strukturze π . Jeżeli uwzględni się ponadto wpływ zjawiska modulacji efektywnej szerokości bazy (w modelu wielkosygnałowym to zjawisko jest uwzględnione w zależności $\alpha_N(U_{CB})$), to model z rys. 5.60 należy uzupełnić pojemnością dyfuzyjną kolektora ($C_{dc} = C_{dc}^M$), konduktancją zwrotną g_{bc} oraz konduktancją wyjściową g_{ce} . Stosując oznaczenia:

$$C_{b'e} = C_{je} + C_{de}$$

$$C_{b'c} = C_{jc} + C_{dc}$$

otrzymuje się model przedstawiony na rys. 5.61. Dokładność modelu „hybrid π ” można zwiększyć rozbudowując go dalszymi elementami reprezentującymi zjawiska drugo- i trzeciorzędne. Z praktycznego punktu widzenia najistotniejsza jest umiejętność dokonania wyboru wariantu modelu o optymalnej złożoności i dokładności. W następnym punkcie rozpatrzmy szereg wariantów modelu „hybrid π ” w ujęciu „ewolucyjnym” od najprostszego do najbardziej złożonego.



Rys. 5.61
Dokładniejszy wariant modelu „hybryd π ” dla układu WE

Model „hybryd π ” dla układu WE

5.9.2.2

Tablica 5.3 zbudowana na zasadzie matrycy o czterech wierszach i czterech kolumnach zawiera kilka wariantów modelu „hybryd π ” dla konfiguracji WE. W pierwszym wierszu są przedstawione cztery warianty modelu dla małosygnałowej quasi-statycznej pracy tranzystora. Pojęcie pracy quasi-statycznej jest rozumiane w ten sposób, że wszystkie wartości chwilowe napięć i prądów są zgodne w fazie, czyli zmiany wielkości zależnych (w przypadku rozpatrywanego modelu są to prądy i_b , i_c) nadążają za zmianami ich argumentów (napięć u_{be} , u_{ce}). Taka sytuacja występuje przy małej częstotliwości sygnału zmiennego i wówczas w modelu nie uwzględnia się elementów reaktancyjnych. Schematy umieszczone w pierwszym wierszu są uszeregowane w kolejności odpowiadającej wzrastającej dokładności. Liniami pogrubionymi narysowano podstawowy (najprostszy) schemat, w którym uwzględniono tylko konduktancję obwodu wejściowego g_{be} oraz źródło prądowe $g_m u_{be}$, dające sygnał na wyjściu tranzystora. Ten podstawowy model jest uzupełniany na kolejnych schematach przez dołączanie elementów w gałęziach narysowanych linią cienką. W ten sposób dołączono najpierw konduktancję wyjściową g_{ce} (jest ona równa tangensowi kąta nachylenia charakterystyki statycznej $I_c(U_{CE})$), następnie konduktancję sprzężenia zwrotnego g_{cb} (reprezentuje ona zjawisko oddziaływania zwrotnego, spowodowanego modulacją efektywnej szerokości bazy) i w końcu rezystancję rozproszoną bazy $r_{bb'}$.

W pierwszej kolumnie tablicy przedstawiono schematy zastępcze, w których uwzględniono elementy reaktancyjne. Są to zatem modele dla pracy dynamicznej tranzystora, tj. dla takiego zakresu częstotliwości, w którym występują przesunięcia fazowe między prądami a napięciami na końcówkach. Modele są uszeregowane w kolejności odpowiadającej wzrastającej dokładności opisu zjawisk dynamicznych. Najpierw uwzględnia się tylko pojemność C_{be} (w najprostszym przypadku jest to tylko pojemność dyfuzyjna, związana ze zjawiskiem ładowania obszaru bazy, w wariancie dokładniejszym uwzględnia się również pojemność złączową C_{je}), następnie pojemność sprzężenia zwrotnego C_{bc} (w wariancie prostszym jest to tylko pojemność złączowa C_{jc} , w dokładniejszym uwzględnia się również pojemność dyfuzyjną $C_{dc} = C_{dc}^M$). Dokładność modelu jeszcze bardziej wzrasta przy rozdzieleniu pojemności C_{bc} na dwie składowe dołączone do węzłów b , b' . Takie postępowanie ma uzasadnienie w rzeczywistej strukturze tranzystora planarnego. Jeszcze większą dokładność dałoby zastąpienie obwodu $r_{bb'}$, $C_{b'e}$ linią długą, co jednakże skomplikowałoby nadmiernie schemat zastępczy.

Kompilując modele przedstawione w pierwszym wierszu z modelami w pierwszej kolumnie można otrzymać warianty modelu „hybryd π ” o różnej dokładności „quasi-statycznej” i „dynamicznej”. Przykładowo — w pozycji wyznaczonej przez czwarty wiersz i czwartą kolumnę przedstawiono wariant „najdokładniejszy”, który z powodzeniem może być stosowany w bardzo szerokim zakresie częstotliwości. Słowo najdokładniejszy wzięto w cudzysłów, gdyż chodzi tylko o naj-

Tablica 5.3

Warianty modelu „hybryd π ” dla konfiguracji WE

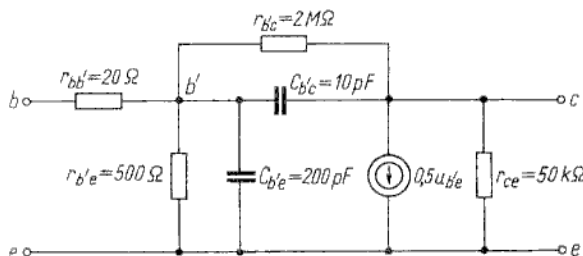
Wzrost dokładności „quasi-statycznej”

	<p>Wzrost dokładności „dynamicznej”</p>	<p>Wariant prosty: $C_{be}=C_{bc}$ Wariant dokładniejszy: $C_{be}=C_{bc}+C_{je}$</p>	<p>Wariant prosty: $C_{bc}=C_{bc}$ Wariant dokładniejszy: $C_{bc}=C_{bc}+C_{dc}$</p>			
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	
g_{be} $g_m u_{be}$	g_{be} C_{be} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$	g_{be} C_{bc} $g_m u_{be}$ g_{cbc}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb}	g_{be} C_{be} C_{bc} $g_m u_{be}$ g_{cbe} g_{cbb} g_{ccb} g_{cce}	
C_{be}	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}$ C_{be} C_{bc}	$g_{ce}; g_{cb}; \Gamma_{bb'}$ C_{be}	

Wzrost dokładności „dynamicznej”

większą dokładność w zbiorze rozpatrywanych wariantów. Nie ogranicza to bynajmniej możliwości dalszego rozwijania tego modelu, co jest szczególnie celowe w przypadku tranzystorów pracujących w zakresie bardzo dużych częstotliwości (powyżej kilkuset megaherców). Zachodzi wówczas konieczność uwzględnienia pojemności oprawki, indukcyjności doprowadzeń, dokładniejszego zapisanie generatora prądowego (z uwzględnieniem tzw. „nadmiaru fazy”, o czym będzie mowa w p. 5.10.1.1) itp.

Istotnym zagadnieniem jest określenie zakresu częstotliwości, w jakim dany model zachowuje zadowalającą dokładność. W związku z tym często można spotkać określenia: model dla małych, średnich, bądź dużych częstotliwości. Takie określenia nie poparte danymi liczbowymi są enigmatyczne, gdyż nie pozwalają się zorientować, jakie są to w istocie zakresy. Dla lepszej ilustracji tego zagadnienia rozpatrzmy przykład liczbowy przedstawiony na rys. 5.62 (wszystkie konduktancje zamieniono na odpowiadające im rezystancje).



Rys. 5.62
Model „hybrid π ” z podanymi wartościami liczbowymi elementów składowych

Rozważmy przy jakich częstotliwościach można pominąć elementy reaktancyjne? Są one określone przez stałe czasowe obwodów $r_{b'e}$, $C_{b'e}$ oraz $r_{b'c}C_{b'c}$

$$f_1 = \frac{1}{2\pi r_{b'e} C_{b'e}} = \frac{1}{2\pi \cdot 500 \cdot 200 \cdot 10^{-12}} \approx 1,6 \text{ MHz}$$

$$f_2 = \frac{1}{2\pi r_{b'c} C_{b'c}} = \frac{1}{2\pi \cdot 2 \cdot 10^6 \cdot 10 \cdot 10^{-12}} \approx 8 \text{ kHz}$$

Dla częstotliwości poniżej 8 kHz można zatem pominąć obie pojemności. W ten sposób otrzymuje się model, który nazwano „quasi-statycznym”, a równie dobrze można go traktować jako model dla zakresu małych częstotliwości, rozumianego jako $0 \leq f < 8 \text{ kHz}$. W zakresie $8 \text{ kHz} \leq f < 1,6 \text{ MHz}$ należy uwzględnić pojemność $C_{b'c}$ (można jej nie brać pod uwagę, jeżeli w ogóle rezygnuje się z uwzględnienia zjawiska sprzężenia zwrotnego; nie miałoby natomiast sensu uwzględnianie $r_{b'c}$ przy pominięciu $C_{b'c}$; uzasadnione jest raczej postępowanie odwrotne, czyli uwzględnianie $C_{b'c}$, przy pominięciu $r_{b'c}$), można natomiast nie brać pod uwagę pojemności $C_{b'e}$. Otrzymany w ten sposób model uproszczony można by nazwać modelem dla zakresu średnich częstotliwości. Dla $f \geq 1,6 \text{ MHz}$ nie można pomijać żadnej pojemności i pełny model można by nazwać modelem dla zakresu dużych częstotliwości. Ten przykład ilustruje, że pojęcia małej, średniej, dużej, czy bardzo dużej częstotliwości mogą mieć różny sens w odniesieniu do różnych modeli.

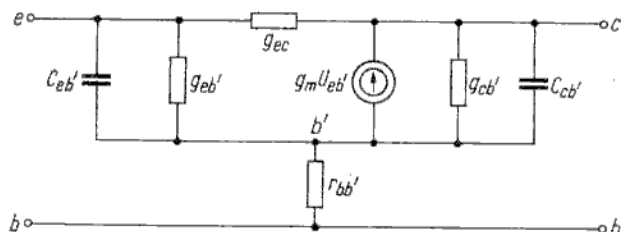
Bardzo ważnym zagadnieniem jest metodyka wyznaczania parametrów modelu „hybrid π ”. Temat ten omówimy po rozpatrzeniu częstotliwości granicznych tranzysto-

ra, gdyż niektóre metody wyznaczania parametrów modelu są oparte na definicjach tych częstotliwości. Przed tym jednak należy uzupełnić rozważania o krótkie omówienie modeli „hybryd π ” dla dwóch rzadziej stosowanych konfiguracji, tj. WB i WC.

Model „hybryd π ” dla układu WB

5.9.2.3

Model „hybryd π ” dla układu WB przedstawiono na rys. 5.63. W stosunku do układu WE zmienia się tylko wartość admittancji wejściowej ($g_{eb'}$, $C_{eb'}$) oraz zwrot źródła prądowego $g_m u_{eb'}$. Ponieważ prądem wejściowym jest w tym przy-



Rys. 5.63
Model „hybryd π ”
dla układu WB

padku prąd emitera, a w układzie WE był prąd bazy, uwzględniając zatem zależność $i_e = (\beta + 1)i_b$ otrzymujemy

$$g_{eb'} = (\beta + 1)g_{b'e} \tag{5.136}$$

Wzory stosowane przy wyznaczaniu parametrów modelu „hybryd π ” w układzie WE

Tablica 5.4

Parametry wyjściowe	$I_c, \beta, h_{11b}, \mu, C_{22b}, k, f_T$
g_m	$ I_c / \varphi_T = \frac{q}{kT} I_c $
$g_{b'e}$	g_m / β
$r_{bb'}$	$h_{11b} - \frac{\beta}{g_m}$
$g_{b'e}$	$\frac{\mu}{\beta} g_m$
g_{ce}	μg_m
C_{bc}	$(1 - k) C_{22b}$
$C_{b'c}$	$k C_{22b}$
$C_{b'e}$	$\frac{g_m}{2\pi f_T} - C_{b'c}$

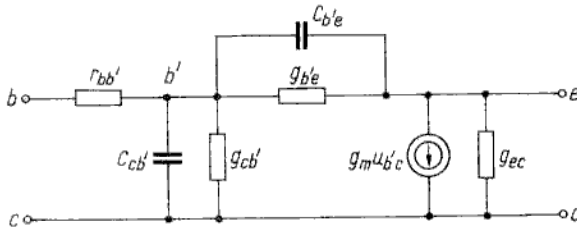
WE

Pojemność C_{cb} składa się z pojemności złączowej C_{je} oraz dyfuzyjnej C_{de} . Ta pierwsza nie ulega zmianie w stosunku do układu WE, druga zaś jest mniejsza niż w układzie WE. Teoretyczne uzasadnienie tego faktu jest dość trudne, dlatego ograniczymy się do podania w tabl. 5.4 zależności umożliwiającej doświadczalne wyznaczenie pojemności C_{cb} .

Model „hybryd π ” dla układu WC

5.9.2.4

Model „hybryd π ” dla układu WC przedstawiono na rys. 5.64. Ten model można otrzymać bezpośrednio ze schematu zastępczego dla układu WE w rezultacie prostej zamiany końcówek. Łatwo można się przekonać, że zwiera-



Rys. 5.64
Model „hybryd π ” dla układu WC

jąc wejście lub wyjście uzyska się admitancje wejściową i wyjściową identyczne jak dla układu WE (w rozdziale 7 zobaczymy, że w warunkach roboczych wymienione admitancje bardzo się różnią dla obu konfiguracji). Porównując dalej modele dla konfiguracji WC, WE można zauważyć tylko nieznaczną różnicę w określeniu źródła prądowego. Dla układu WE

$$g_m u_{b'e} = \beta i_b$$

natomiast dla układu WC

$$g_m u_{b'e} = (\beta + 1) i_b$$

Ponieważ w stanie zwarcia wyjścia $u_{b'e} = u_{b,c}$, zatem

$$g_m \Big|_{\text{dla WC}} = \frac{\beta + 1}{\beta} g_m \Big|_{\text{dla WE}} \tag{5.137}$$

Częstotliwości graniczne tranzystora bipolarnego

5.10

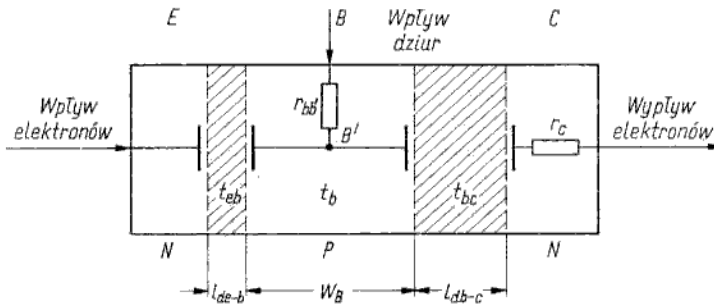
W miarę wzrostu częstotliwości sygnału zmieniają się warunki pracy tranzystora, przede wszystkim maleje jego wzmocnienie prądowe. Dlatego należy określić pasmo częstotliwości, w jakim tranzystor może pracować. Częstotliwość ograniczająca to pasmo „od góry” jest nazywana częstotliwością graniczną. W zależności od przyjętych kryteriów można zdefiniować kilka różnych częstotliwości granicznych. Zanim jednak przedstawimy te kryteria i definicje, podamy najpierw fizyczny opis właściwości częstotliwościowych tranzystora.

Częstotliwości graniczne — interpretacja fizyczna

5.10.1

W punkcie 5.8 stwierdzono, że tranzystor jest elementem inercyjnym, co wynika ze skończonych szybkości przeładowania poszczególnych obszarów (głównie dwu warstw zaporowych i bazy). Te same zjawiska określają właściwości dyna-

miczne tranzystora przy pracy z małym sygnałem. Sygnał prądu zmiennego na drodze od emitera do kolektora ulega osłabieniu i opóźnieniu. W strukturze tranzystora można wyodrębnić trzy warstwy, w których niezależne zjawiska powodują zmniejszenie i opóźnienie sygnału (rys. 5.65). Są to:



Rys. 5.65

Struktura tranzystora podzielona na obszary, w których występują niezależne zjawiska zmniejszenia i opóźnienia sygnału

- warstwa zaporowa złącza E-B;
- warstwa bazy;
- warstwa zaporowa złącza B-C.

Ponieważ transport nośników odbywa się szeregowo przez te warstwy, całkowity czas przejścia sygnału (czas przelotu nośników od emitera do kolektora) jest sumą algebraiczną opóźnień wnoszonych przez każdą warstwę dla sygnału skoku jednostkowego

$$t_{\text{całk}} = t_{eb} + t_b + t_{bc} \quad (5.138)$$

Analogicznie stała czasowa tranzystora jest równa sumie stałych czasowych wnoszonych przez poszczególne warstwy

$$\tau_{\text{całk}} = \tau_{eb} + \tau_b + \tau_{bc} \quad (5.139)$$

Świadomie rozróżnia się czas przejścia sygnału skoku jednostkowego (opóźnienie) i stałą czasową. Czas przejścia sygnału od emitera do kolektora $t_{\text{całk}}$ odpowiada wprowadzonej wcześniej stałej czasowej dla stanu nieustalonego τ_N , określającej zwłokę odpowiedzi prądu kolektora na skok jednostkowy prądu emitera. Natomiast stała czasowa dla sygnału sinusoidalnego jest rozumiana jako odwrotność pulsacji, przy której następuje zmniejszenie amplitudy małego sygnału o 3 dB. Dalej okaże się, że nie zawsze opóźnienie sygnału jest równe stałej czasowej.

Na podstawie (5.139) można napisać równanie pulsacji odpowiadających poszczególnym stałym czasowym

$$\frac{1}{\omega_{\text{całk}}} = \frac{1}{\omega_{eb}} + \frac{1}{\omega_b} + \frac{1}{\omega_{bc}} \quad (5.140)$$

Pierwszoplanowe znaczenie mają zjawiska zachodzące w bazie, dlatego w przybliżeniu pierwszego stopnia przyjmuje się

$$\tau_{\text{całk}} \approx \tau_b$$

Najpierw rozpatrzmy zatem wpływ skończonego czasu przelotu nośników przez bazę t_b (lub odpowiadającej mu stałej czasowej τ_b) na właściwości tranzystora, a następnie uwzględnimy addytywne stałe czasowe τ_{eb} , τ_{bc} .

Częstotliwość graniczna dla układu WB

5.10.1.1

Stała czasowa bazy

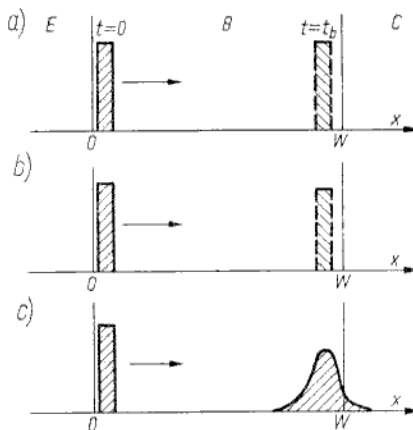
Jaki jest związek między wartością chwilową prądu na końcu i na początku bazy? W punkcie 5.5 stosunek tych dwu prądów nazwano współczynnikiem transportu i oznaczono przez α_b . Dla sygnałów małej częstotliwości można było przyjąć, że faza obu prądów jest identyczna (okres sygnału zmiennego był znacznie większy niż czas przelotu nośników przez bazę), czyli α_b było liczbą rzeczywistą, której wartość była określona zjawiskiem rekombinacji nośników w bazie. Z punktu 5.5 wiadomo, że

$$\alpha_b = \frac{1}{1 + \frac{t_b}{\tau_r}} \quad (5.141)$$

W miarę jak okres sygnału zmiennego staje się porównywalny z czasem przelotu nośników przez bazę, należy uwzględnić dodatkowo dwa zjawiska:

- przesunięcie fazowe związane ze skończonym czasem przejścia sygnału przez bazę;
- osłabienie amplitudy sygnału związane z tzw. rozmyciem dyfuzyjnym, czyli wynikające stąd, że część nośników gromadzonych w bazie w czasie trwania jednej połówki sinusoidalnego prądu emitera jest z niej usuwana w drugiej połowie z powrotem do emitera, a więc nie bierze udziału w prądzie kolektora (jest to inaczej ładowanie i rozładowanie pojemności dyfuzyjnej bazy, związane z przepływem prądu w obwodzie emiter-baza, a nie emiter-kolektor).

Wszystkie trzy zjawiska (osłabienie wskutek rekombinacji, przesunięcie fazowe wskutek skończonego czasu przelotu, osłabienie wskutek rozmycia dyfuzyjnego) obrazowo ilustruje rys. 5.66, na którym pokazano przejście pojedynczego impulsu ładunku od emitera do kolektora. Uwzględniając wszystkie wyżej wymienione



Rys. 5.66

Ilustracja wpływu trzech zjawisk na właściwości przenoszenia sygnału przez tranzystor: a) wpływ skończonego czasu przelotu (tylko opóźnienie); b) wpływ skończonego czasu przelotu i rekombinacji (opóźnienie i osłabienie); c) wpływ dwu poprzednich zjawisk i „rozmycia dyfuzyjnego” (opóźnienie i osłabienie)

zjawiska należy oczekiwać, że dla dużych częstotliwości współczynnik α_b jest liczbą zespoloną, której moduł i faza zależą od częstotliwości. W postaci ogólnej

$$\alpha_b(j\omega) = A(\omega)e^{j\varphi(\omega)} \quad (5.142)$$

Współczynnik $\alpha_b(j\omega)$ można wyrazić w funkcji parametrów fizycznych i konstrukcyjnych tranzystora, rozwiązując równanie transportu (równanie (1.60) w rozdziale 1) nośników w bazie. Rozwiązanie takie wymaga stosowania złożonego aparatu matematycznego, dlatego ograniczymy się do wyprowadzenia przybliżonego wyrażenia przy użyciu uproszczonych wzorów i przekształceń.

Jeżeli na wejście (układ WB) podamy sinusoidalny sygnał prądu

$$i_e = I_{em} e^{j\omega t} \quad (5.143)$$

nałożony na składową stałą I_E , to można przyjąć, że całkowity ładunek nośników nadmiarowych w bazie

$$Q_b = Q_{b0} + Q_{bm} e^{j\omega t} \quad (5.144)$$

przy czym: Q_{b0} — ładunek odpowiadający składowej stałej prądu emitera (określony punktem pracy); Q_{bm} — amplituda przyrostów ładunku spowodowanych składową zmienną prądu emitera.

Składową zmienną ładunku w bazie oznaczamy jako

$$q_b = Q_{bm} e^{j\omega t} \quad (5.145)$$

Szybkość zmian tego ładunku w funkcji czasu jest równa różnicy prądu wpływającego do bazy (i_e) i wypływającego (i_c) oraz szybkości rekombinacji nośników w bazie (q_b/τ_r). Zatem

$$\frac{dq_b}{dt} = \frac{-q_b}{\tau_r} + (i_e - i_c) \quad (5.146)$$

Jest to inaczej równanie transportu dla uproszczonego jednosekcyjnego modelu tranzystora (patrz dyskusja w p. 5.8.3 dotycząca usuwania zmiennej x z równania transportu). Po podstawieniu (5.145) do (5.146) otrzymuje się

$$i_e - i_c = \left(\frac{1 + j\omega\tau_r}{\tau_r} \right) q_b \quad (5.147)$$

Przyjmując, że czas przelotu nośników przez bazę

$$t_b = \frac{Q_{b0}}{I_C}$$

jest równy stałej czasowej dla sygnału sinusoidalnego

$$\tau_b = \frac{q_b}{i_c} \quad \tau_b \approx t_b \quad (5.148)$$

po podzieleniu obu stron równania (5.147) przez i_c oraz przy założeniu, że α_e (współczynnik sprawności wstrzykiwania) nie zależy od częstotliwości i jest równy jedności, czyli

$$\alpha_b(\omega) = \alpha_i(\omega) = \frac{i_c}{i_e} \quad (5.149)$$

otrzymuje się

$$\alpha_i(j\omega) = \frac{1}{1 + \frac{\tau_b}{\tau_r} + j\omega\tau_b} \quad (5.150)$$

przy czym indeks i oznacza parametr (w tym przypadku wzmocnienie prądowe) dla tzw. tranzystora idealnego, w którym uwzględnia się tylko zjawiska zachodzące w bazie. Ponieważ

$$\frac{\tau_b}{\tau_r} \ll 1$$

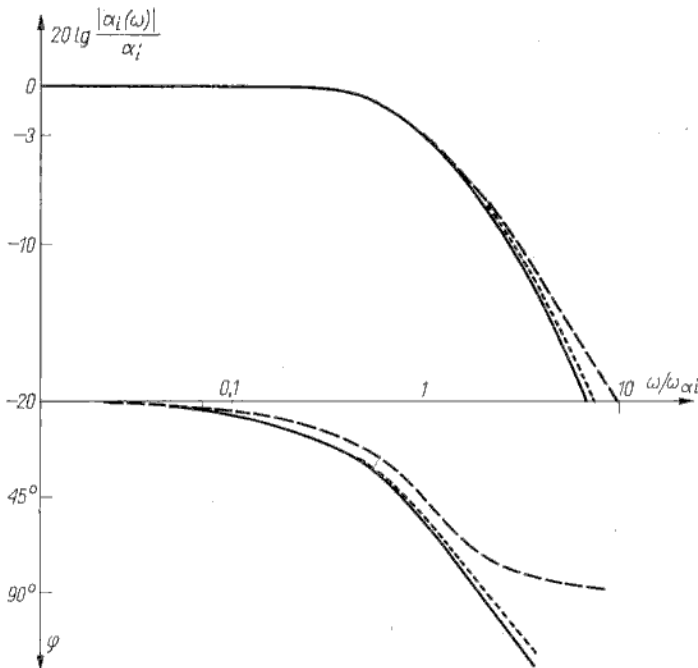
można zatem przepisać (5.150) w postaci

$$\alpha_i(j\omega) \approx \frac{1}{1 + \frac{\tau_b}{\tau_r} + j\omega\tau_b \left(1 + \frac{\tau_b}{\tau_r}\right)}$$

Stąd po uwzględnieniu (5.141) oraz (5.148)

$$\alpha_i(j\omega) = \frac{\alpha_i}{1 + j\omega\tau_b} \quad (5.151)$$

przy czym α_i — wzmocnienie prądowe tranzystora idealnego w układzie WB dla $\omega \rightarrow 0$.



Rys. 5.67

Charakterystyki częstotliwościowe modułu i fazy współczynnika wzmocnienia prądowego (linia ciągła — charakterystyki rzeczywiste, linia przerywana — według (5.152), (5.153), linia kropkowana — dla rozwiązania dokładnego)

Wzmocnienie prądowe (a raczej przenoszenie prądowe) jest zatem opisane funkcją jednobiegunową doskonale znaną z teorii obwodów liniowych, gdyż identyczny zapis ma funkcja przenoszenia prostego obwodu RC (do tej analogii wrócimy jeszcze przy omawianiu interpretacji częstotliwości granicznych w schemacie zastępczym tranzystora).

Na rysunku 5.67 przedstawiono charakterystyki częstotliwościowe modułu (w skali logarytmicznej) i fazy (w skali półlogarytmicznej) współczynnika wzmocnienia prądowego. Na podstawie (5.151) zapis analityczny tych dwu funkcji jest następujący:

$$|\alpha_i(j\omega)| = A(\omega) = \frac{\alpha_i}{\sqrt{1 + \omega^2 \tau_b^2}} \quad (5.152)$$

$$\varphi(\omega) = -\operatorname{arctg} \omega \tau_b \quad (5.153)$$

przy czym charakterystyka amplitudowo-częstotliwościowa w postaci logarytmicznej

$$\lg A(\omega) \approx \lg \alpha_i \quad \text{dla } \omega \tau_b \ll 1$$

$$\lg A(\omega) \approx \lg \alpha_i - \lg(\omega \tau_b) \quad \text{dla } \omega \tau_b \gg 1$$

może być aproksymowana dwoma odcinkami linii prostej, tj.:

- odcinkiem płaskim o wartości $\lg \alpha_i$ dla $\omega \tau_b < 1$;
- odcinkiem o nachyleniu -20 dB na dekadę lub inaczej -6 dB na oktawę (oktawa — odpowiednik dekady w systemie dwójkowym, czyli 6 dB na oktawę oznacza to samo co 2 razy/2 razy) dla $\omega \tau_b > 1$.

Punkt „szczytia” tych dwu odcinków odpowiada

$$\omega \tau_b = 1$$

a częstotliwość spełniająca ten warunek jest nazywana *częstotliwością graniczną* ω_{xi}

$$\omega_{xi} = \frac{1}{\tau_b} \quad (5.154)$$

Na nieaprosymowanej charakterystyce $|\alpha_i(\omega)|$ pulsacji ω_{xi} odpowiada spadek modułu o 3 dB, czyli do wartości $\alpha_i/\sqrt{2}$. Na podstawie (5.153)

$$\varphi(\omega = \omega_{xi}) = -45^\circ = -\frac{\pi}{4} \quad (5.155)$$

Zatem przy częstotliwości sygnału $\omega = \omega_{xi}$ prąd kolektora jest opóźniony w fazie o 45° w stosunku do prądu emitera, a jego amplituda jest $\alpha_i/\sqrt{2}$ razy mniejsza niż amplituda prądu emitera.

Na rysunku 5.67 przedstawiono również charakterystyki częstotliwościowe amplitudy i fazy współczynnika wzmocnienia prądowego dla dokładnego rozwiązania równania transportu. O ile charakterystyka amplitudowo-częstotliwościowa pokrywa się bardzo dobrze z otrzymaną w wyniku bardzo uproszczonej analizy, o tyle charakterystyki fazowo-częstotliwościowe różnią się dość znacznie. Na przykład przy częstotliwości $\omega = \omega_{xi}$ przesunięcie fazy w dokładnym rozwiązaniu wynosi: $57^\circ + 6^\circ \cdot \eta$ dla tranzystora dryftowego

57° dla tranzystora bezdryftowego ($\eta = 0$)

a nie 45° , jak to wynika z uproszczonej analizy.

Mówi się w tym przypadku o „nadmiarze fazy”, który wynosi

$$(57^\circ + 6^\circ \cdot \eta) - 45^\circ = 12^\circ + 6^\circ \cdot \eta$$

Nadmiar fazy oznaczymy przez m_b i zapiszmy go w radianach

$$m_b = 0,2 + 0,1 \eta \quad (5.156)$$

Do wyrażenia (5.151) można łatwo wprowadzić czynnik uwzględniający nadmiar fazy (tzw. poprawka fazowa)

$$\alpha_i(j\omega) = \frac{\alpha_i}{1 + \frac{j\omega}{\omega_{\alpha i}}} e^{-jm_b \omega / \omega_{\alpha i}} \quad (5.157)$$

Sens fizyczny nadmiaru fazy wyjaśnia częściowo rys. 5.66. Należy zauważyć, że w przypadku samego unoszenia nośników w bazie (bez dyfuzji i z pominięciem rekombinacji) prąd kolektora jest opóźniony w stosunku do prądu emitera o czas przelotu t_b (daje to określone przesunięcie fazowe) bez osłabienia amplitudy. Ponieważ częstotliwość graniczna $\omega_{\alpha i}$ jest definiowana nie jako częstotliwość odpowiadająca przesunięciu fazy o $\pi/4$, lecz jako częstotliwość odpowiadająca osłabieniu amplitudy sygnału o 3 dB, jasne jest zatem, że przy braku osłabienia $\omega_{\alpha i} \rightarrow \infty$, czyli stała czasowa $\tau_b = 1/\omega_{\alpha i} \rightarrow 0$, a przesunięcie fazowe $\varphi \rightarrow \infty$ przy $\omega \rightarrow \infty$ (jest to oczywiście bardzo wyidealizowany obraz, w którym pomija się wszelkie inne zjawiska poza unoszeniem).

Zatem

$$\tau_b \ll t_b$$

Stąd należy wnioskować, że przy udziale unoszenia i dyfuzji również stała czasowa będzie mniejsza niż czas przelotu nośników przez bazę, czyli przesunięcie fazowe sygnału będzie większe, niż to wynika z jednobiegunowej funkcji przeniesienia. Okazuje się, że nawet w tranzystorze bezdryftowym dokładne rozwiązanie równania transportu daje $\tau_b < t_b$, czyli konieczne jest wprowadzenie poprawki fazowej. Należy więc zauważyć, że proste wyrażenie na $\alpha_i(j\omega)$ w postaci funkcji jednobiegunowej (5.151) wyprowadzono przy założeniu $\tau_b = t_b$, dlatego nie ma w tym wzorze poprawki fazowej i za τ_b można podstawiać wcześniej wyprowadzone wyrażenia (5.43), (5.45) lub (5.47) w zależności od rodzaju tranzystora. Natomiast do dokładnego wzoru na $\alpha_i(j\omega)$ w postaci (5.157) należy podstawić $1/\omega_{\alpha i} < t_b$, przy czym najczęściej stosuje się następujące zależności:

— dla tranzystora bezdryftowego

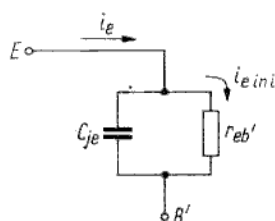
$$\omega_{\alpha i} = 2,43 \frac{D_n}{W_B^2} \quad (5.158)$$

— dla tranzystora dryftowego

$$\omega_{\alpha i} = 2,43 \frac{D_n}{W_B^2} \left[1 + \left(\frac{\eta}{2} \right)^{4/3} \right] \quad (5.159)$$

Stała czasowa warstwy zaporowej złącza emiter-baza

Konduktancja dynamiczna złącza emiter-baza spolaryzowanego w kierunku przewodzenia jest bocznikowana przez pojemność warstwy zaporowej tego złącza C_{je} . Mamy tu do czynienia z prostym obwodem równoległym $r_{eb'}$, C_{je} (rys. 5.68).



Rys. 5.68
Obwód R, C reprezentujący złącze $E-B$

w którym napięcie u_{eb} , a więc i prąd płynący przez $r_{eb'}$ są opóźnione w stosunku do wejściowego prądu emitera. Dla takiego obwodu można napisać prostą zależność

$$i_{e\ ini} = i_e \frac{1}{1 + j\omega r_{eb'} C_{je}} \quad (5.160)$$

W tym przypadku wyrażenie na sprawność wstrzykiwania emitera (5.38) należy uzupełnić o czynnik określający stosunek „użytecznego” prądu $i_{e\ ini}$ (jest to prąd nośników wstrzykiwanych do bazy, stąd indeks *ini* od *iniekcji*) do całkowitego prądu i_e , w który wchodzi również składowa ładowania pojemności C_{je}

$$\alpha_e(j\omega) = \alpha_e \frac{1}{1 + j\omega r_{eb'} C_{je}} \quad (5.161)$$

przy czym α_e — sprawność wstrzykiwania dla małych częstotliwości, określona wzorem (5.38).

Można zdefiniować częstotliwość graniczną emitera

$$\omega_{eb} = \frac{1}{r_{eb'} C_{je}} \quad (5.162)$$

jako taką częstotliwość, przy której moduł współczynnika wstrzykiwania emitera zmniejsza się o 3 dB w stosunku do wartości dla małych częstotliwości

$$\frac{|\alpha_e(\omega = \omega_{eb})|}{\alpha_e} = \frac{1}{\sqrt{2}} \quad (5.163)$$

a przesunięcie fazowe prądów $i_{e\ ini}$, i_e wynosi 45° .

Zatem opóźnienie czasowe, wnoszone przez warstwę zaporową złącza $E-B$, jest spowodowane tym, że proces wstrzykiwania nośników jest zgodny w fazie z napięciem na złączu, a napięcie opóźnia się w stosunku do wejściowego prądu emitera. Ponieważ zjawisko to jest opisane prostą funkcją jednobiegunową, stała czasowa dla sygnału sinusoidalnego jest w tym przypadku równa opóźnieniu odpowiedzi na sygnał skoku jednostkowego

$$\tau_{eb} = \frac{1}{\omega_{eb}} = t_{eb} \quad (5.164)$$

Stała czasowa warstwy zaporowej złącza baza-kolektor
Złącze baza-kolektor wnosi opóźnienie i osłabienie sygnału w zakresie dużych częstotliwości wskutek dwu zjawisk:

- ładowanie pojemności $C_{b'c}$ (dominującym składnikiem tej pojemności jest zwykle C_{jc}) przez szeregową rezystancję kolektora r_c ;
- skończony czas przelotu nośników przez warstwę zaporową tego złącza.

Pierwsze zjawisko ma istotne znaczenie w tranzystorach o dużej wartości rezystancji szeregowej kolektora i można je opisać funkcją jednobiegunową analogiczną jak dla złącza $E-B$. Jeżeli zdefiniuje się stałą czasową

$$\tau'_{bc} = \frac{1}{\omega'_{bc}} = C_{b'c} r_c \quad (5.165)$$

to funkcja przenoszenia prądu przez złącze $B-C$, zdefiniowana jako stosunek prądu wypływającego z warstwy zaporowej tego złącza do prądu wpływającego

$$\alpha'_c(j\omega) = \frac{1}{1 + \frac{j\omega}{\omega'_{bc}}} \quad (5.166)$$

W tranzystorach epiplanarnych rezystancja r_c jest bardzo mała, zatem wpływ tego zjawiska można zwykle pominać. Natomiast szerokość warstwy zaporowej złącza $B-C$ jest w tych tranzystorach bardzo duża (szczególnie w tranzystorach dla zakresu b.w.cz. $l_{d\ b-c} \gg W_B$) i czas przelotu nośników przez tę warstwę (mimo ich bardzo dużej prędkości) może być porównywalny z czasem przelotu przez bazę.

Analiza tego zjawiska jest dość złożona. Funkcja przenoszenia $\alpha''_c(j\omega)$ zdefiniowana identycznie jak $\alpha'_c(j\omega)$ ma postać

$$\alpha''_c(j\omega) = 1 - \frac{j\omega}{\omega''_{bc}} \quad (5.167)$$

przy czym

$$\omega''_{bc} = \frac{1}{\tau''_{bc}} \quad (5.168)$$

a τ''_{bc} jest czasem przelotu nośników przez warstwę zaporową, wyznaczanym z ilorazu szerokości warstwy (drogi przelotu) i szybkości unoszenia

$$\tau''_{bc} = \frac{l_{d\ b-c}}{2v} \quad (5.169)$$

Ponieważ natężenie pola elektrycznego w warstwie zaporowej jest bardzo duże ($E > 10^6$ V/m), zatem szybkość unoszenia nośników jest w przybliżeniu równa szybkości termicznej (ok. 10^5 m/s).

Funkcję przenoszenia $\alpha''_c(j\omega)$ niekiedy przedstawia się w przybliżonej postaci funkcji jednobiegunowej z czynnikiem uwzględniającym poprawkę na „nadmiar fazy”, który wynika z unoszeniowego charakteru transportu nośników

$$\alpha''_c(j\omega) \approx \frac{1}{1 + \frac{j\omega}{\omega''_{bc}}} e^{-jm_{bc}\omega/\omega''_{bc}} \quad (5.170)$$

Sumaryczna stała czasowa złącza baza-kolektor

$$\tau_c = \tau'_c + \tau''_c$$

Wyrażenia opisujące przenoszenie sygnału przez charakterystyczne obszary tranzystora

Tablica 5.5

Rodzaj warstwy i rodzaj zjawiska	Funkcja i parametry	Funkcja przenoszenia	Stala czasowa
Warstwa zaporowa złącza emiter-baza	obwód R, C	$\alpha_e(\omega) = \frac{\alpha_e(0)}{1 + j\omega/\omega_{eb}}$	$\tau_{eb} = \frac{1}{\omega_{eb}} = \tau_{c0} C_{jc}$
Baza	przełot	$\alpha_b(\omega) = \frac{\alpha_b(0)}{1 + j\omega/\omega_b} e^{-j m_b \omega/\omega_b}$ $m_b = 0,2 + 0,1\eta$	$\tau_b = \frac{1}{\omega_b} = \begin{cases} \frac{D_n}{W_B^2} & \text{dla tranzystora bezdriftowego} \\ \frac{D_n}{W_B^2} \left[1 + \left(\frac{\eta}{2} \right)^{4/3} \right] & \text{dla tranzystora driftowego} \end{cases}$
Warstwa zaporowa złącza baza-kolektor	obwód R, C	$\alpha'_c(\omega) = \frac{1}{1 + j\omega/\omega'_{bc}}$	$\tau'_{bc} = \frac{1}{\omega'_{bc}} = \tau_c C_{bc}$
	przełot	$\alpha''_c(\omega) = \frac{1}{1 + j\omega/\omega''_{bc}} e^{-j m_{bc} \omega/\omega''_{bc}}$	$\tau''_{bc} = \frac{1}{\omega''_{bc}} = \frac{l_{db-c}}{2v}$ $\tau_{bc} = \tau'_{bc} + \tau''_{bc}$

a całkowita funkcja przenoszenia tego złącza

$$\alpha_c(j\omega) = \alpha'_c(j\omega)\alpha''_c(j\omega)$$

Wyrażenia związane ze zjawiskami przenoszenia sygnału przez poszczególne warstwy tranzystora zebrano w tabl. 5.5. Całkowita funkcja przenoszenia prądu przez tranzystor w układzie WB

$$\alpha(j\omega) = \alpha_e(j\omega)\alpha_b(j\omega)\alpha'_c(j\omega)\alpha''_c(j\omega) \quad (5.171)$$

Po uwzględnieniu (5.157), (5.161), (5.166), (5.170) otrzymuje się wzór

$$\alpha(j\omega) = K(j\omega)e^{-jm\omega/\omega_\alpha} \quad (5.172)$$

w którym:

$$K(j\omega) = \frac{\alpha}{\left(1 + \frac{j\omega}{\omega_{eb}}\right)\left(1 + \frac{j\omega}{\omega_b}\right)\left(1 + \frac{j\omega}{\omega'_{bc}}\right)\left(1 + \frac{j\omega}{\omega''_{bc}}\right)} \quad (5.173)$$

$$m = \frac{\omega''_{bc}m_b + \omega_b m_{bc}}{\omega_b \omega''_{bc}} \omega_\alpha \quad (5.174)$$

$$\omega_\alpha = \frac{1}{\tau_{eb} + \tau_b + \tau_{bc}} \quad (5.175)$$

przy czym

$$\alpha = \alpha_e \alpha_b \alpha'_c \alpha''_c$$

Jest to zatem funkcja czterobiegunowa z czynnikami opóźniającym, uwzględniającym poprawkę fazową (nadmiar fazy). Tylko w przypadku

$$\omega < \omega_\alpha$$

można zastosować jednobiegunową aproksymację w postaci

$$\alpha(j\omega) \approx \frac{\alpha}{1 + \frac{j\omega}{\omega_\alpha}} e^{-jm\omega/\omega_\alpha} \quad (5.176)$$

Najczęściej przyjmuje się, że

$$\tau_b \gg \tau_{eb} + \tau_{bc} \quad (5.177)$$

i wówczas

$$\omega_\alpha \approx \omega_b \quad m = m_b \quad (5.178)$$

Wyrażenie (5.176) przy uwzględnieniu (5.177), (5.178) jest zwykle przyjmowane za podstawowe przy analizie działania tranzystora. Charakterystyki częstotliwościowe amplitudy i fazy dla tak aproksymowanej funkcji przenoszenia mają przebiegi pokazane na rys. 5.67. Należy jednak pamiętać, że przy porównywalnych wartościach dwu lub więcej stałych czasowych $\alpha(j\omega)$ wyraża się wielobiegunową funkcją przenoszenia (5.172), w której obowiązuje zasada superpozycji faz i logarytmów amplitud. Na przykład w przypadku funkcji dwubiegunowej nachylenie logarytmicznej charakterystyki amplitudowo-częstotliwościowej wynosi nie 6 dB na oktawę, lecz 12 dB na oktawę.

Rozpatrzmy teraz funkcję przenoszenia prądowego w układzie WE.

Częstotliwości graniczne dla układu WE

5.10.1.2

Ponieważ

$$\beta(j\omega) = \frac{\alpha(j\omega)}{1 - \alpha(j\omega)} \quad (5.179)$$

zatem podstawiając $\alpha(j\omega)$ w postaci (5.172) otrzymuje się

$$\beta(j\omega) = \frac{K(j\omega)e^{-jm\omega/\omega_\alpha}}{1 - K(j\omega)e^{-jm\omega/\omega_\alpha}} \quad (5.180)$$

Dalej wykazemy, że wystarczy rozpatrzeć zakres

$$\omega \ll \omega_\alpha \quad (5.181)$$

w którym funkcję $K(j\omega)$ można z dobrym przybliżeniem przedstawić w postaci

$$K(j\omega) = \frac{\alpha}{1 + \frac{j\omega}{\omega_\alpha}} \quad (5.182)$$

Po podstawieniu (5.182) do (5.180) otrzymuje się

$$\beta(j\omega) = \frac{\alpha e^{-jm\omega/\omega_\alpha}}{1 + \frac{j\omega}{\omega_\alpha} - \alpha e^{-jm\omega/\omega_\alpha}} \quad (5.183)$$

Jeżeli przyjmiemy, że

$$\omega \ll \frac{\omega_\alpha}{m}$$

(ponieważ $m \leq 1$, nierówność ta jest na pewno spełniona przy zachowaniu warunku (5.181)), to

$$e^{-jm\omega/\omega_\alpha} \approx -1 \frac{jm\omega}{\omega_\alpha}$$

a więc

$$\beta(j\omega) \approx \frac{\alpha}{1 - \alpha + j(1 + m\alpha) \frac{\omega}{\omega_\alpha}} e^{-jm\omega/\omega_\alpha}$$

Po podzieleniu tego wyrażenia przez $(1 - \alpha)$ otrzymuje się

$$\beta(j\omega) = \frac{\beta}{1 + \frac{j\omega}{\omega_\beta}} e^{-jm\omega/\omega_\alpha} \quad (5.184)$$

przy czym

$$\omega_\beta = \frac{1 - \alpha}{1 + m\alpha} \omega_\alpha \quad (5.185)$$

Pulsacja ω_β (lub częstotliwość $f_\beta = \omega_\beta/2\pi$) jest nazywana *pulsacją (częstotliwością) graniczną* w układzie WE i zgodnie z właściwościami jednobiegunowej funkcji przenoszenia ma sens takiej pulsacji (częstotliwości), przy której moduł wzmocnienia prądowego $|\beta(j\omega)|$ maleje o 3 dB, czyli do wartości $\beta/\sqrt{2}$.

Funkcję (5.184) można uprościć, jeżeli weźmie się pod uwagę realne wartości jej argumentów. W tym celu oszacujemy wpływ członu opóźniającego

$$e^{-jm\omega/\omega_\alpha}$$

czyli ocenimy przede wszystkim, jakie realnie największe wartości może przyjmować współczynnik m . Jeżeli do (5.174) podstawimy dość wygórowane wartości $m_{bc} = m_b = 1$ i przyjmiemy „dla najgorszego przypadku”, że $1/\omega_\alpha = 1/\omega'_{bc} + 1/\omega_b$ (pomija się $1/\omega_{eb}$ oraz $1/\omega'_{bc}$), to $m = 1$. Podstawiając tę wartość do (5.185) otrzymujemy

$$\omega_\beta = \frac{\omega_\alpha}{2\beta}$$

Stąd nadmiar fazy

$$\Delta\varphi = -m \frac{\omega}{\omega_\alpha} = -\frac{\omega}{2\beta\omega_\beta}$$

Przesunięcie fazowe wnoszone przez czynnik inercyjny (funkcję jednobiegunową

$$\varphi = -\arctg \frac{\omega}{\omega_\beta}$$

Przy realnych wartościach β ($\beta \approx 100$) i w realnym zakresie ω ($\omega \leq 10\omega_\beta$)

$$\Delta\varphi \ll \varphi$$

Na przykład dla $\omega = \omega_\beta$, $\beta = 100$

$$\varphi = -\frac{\pi}{4}$$

a nadmiar fazy

$$\Delta\varphi = -\frac{\pi}{628}$$

Można zatem przyjąć, że

$$e^{-jm\omega/\omega_\alpha} \approx 1$$

i funkcję przenoszenia prądu w układzie WE zapisać w postaci

$$\beta(j\omega) \approx -\frac{\beta}{1 + \frac{j\omega}{\omega_\beta}} \quad (5.186)$$

Pominięcie czynnika $e^{-jm\omega/\omega_\alpha}$ nie oznacza, że nadmiar fazy związany z transportem nośników od emitera do kolektora nie jest w ogóle uwzględniany w funkcji (5.186). Należy zwrócić uwagę, że w zależności $\omega_\beta(\omega_\alpha)$ wyrażonej wzorem (5.185) występuje parametr m reprezentujący poprawkę fazową.

Przedstawmy również kilka praktycznie przydatnych przybliżeń związku (5.185). Jeżeli przyjmiemy $m_{bc} = 0$ (pomijają się wpływ nadmiaru fazy, wnoszonego przez prętot nośników w warstwie zaporowej złącza B-C), to na podstawie (5.174)

$$m = m_b \frac{\omega_\alpha}{\omega_b} = m_b \frac{\tau_b}{\tau_{eb} + \tau_b + \tau_{bc}} \quad (5.187)$$

przy czym m_b jest określone wzorem (5.156).

Często można przyjąć, że $\tau_b \gg \tau_{eb} + \tau_{bc}$, zatem $m \approx m_b$ i wówczas:

— dla tranzystora dryftowego

$$\omega_\beta \approx \frac{1}{1,2 + 0,1\eta} \frac{\omega_\alpha}{\beta} \quad (5.188)$$

— dla tranzystora bezdryftowego

$$\omega_\beta \approx 0,82 \frac{\omega_\alpha}{\beta} \quad (5.189)$$

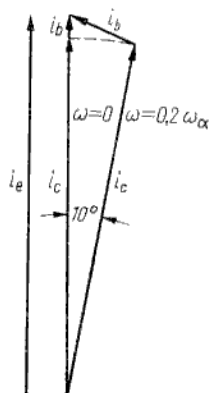
Jeżeli w ogóle pominie się poprawkę fazową ($m = 0$), to

$$\omega_\beta \approx \frac{\omega_\alpha}{\beta} \quad (5.190)$$

Przytoczone wzory (5.188) do (5.190) wyrażają bardzo ważny fakt, że częstotliwość graniczna w układzie WE jest około β razy mniejsza niż w układzie WB. Nie jest to zaskakujący wniosek, jeżeli zauważymy na rys. 5.67, że przesunięcie fazowe między prądem kolektora i emitera pojawia się przy wartościach częstotliwości znacznie mniejszych niż ω_α . Faza funkcji $\alpha(j\omega)$ zmienia się znacznie „wcześniej” (przy mniejszych częstotliwościach) niż jej moduł. Ponieważ suma trzech prądów w tranzystorze jest zawsze równa zero, zatem przesunięcie fazowe między prądem kolektora i emitera prowadzi do wzrostu prądu bazy (w sensie fizycznym jest to prąd ładowania pojemności dyfuzyjnej bazy), czyli do zmniejszenia modułu współczynnika $\beta(j\omega)$. Tę sytuację ilustruje wykres wektorowy trzech prądów (rys. 5.69) przykładowo dla pulsacji $\omega = 0,2\omega_\alpha$, przy której

$$\alpha(j\omega) \approx \alpha$$

$$\varphi = 10^\circ$$



Rys. 5.69

Wykres wektorowy trzech prądów

Z wykresu widać, że mimo braku zmian amplitudy prądu kolektora wzrasta kilkukrotnie amplituda prądu bazy, czyli pulsacja $\omega = 0,2 \omega_\alpha$ jest na pewno znacznie większa niż ω_β (przy $\omega = \omega_\beta$ prąd bazy wzrósłby $\sqrt{2}$ razy), a więc tym bardziej

$$\omega_\alpha \gg \omega_\beta$$

Pulsacja ω_β ma bardzo prosty sens fizyczny dla tranzystora, którego właściwości częstotliwościowe są zależne wyłącznie od stałej czasowej bazy ($\tau_b \gg \tau_{eb} + \tau_{bc}$, czyli $\omega_\alpha \approx 1/\tau_b$). Jeżeli ponadto założymy, że współczynnik sprawności wstrzykiwania $\alpha_e(j\omega) = 1$, to na podstawie (5.141), (5.190)

$$\beta = \frac{\alpha}{1-\alpha} = \frac{\tau_r}{\tau_b} \quad (5.191)$$

oraz

$$\omega_\beta = \frac{1}{\tau_r} \quad (5.192)$$

Pulsacja graniczna ω_β w układzie WE jest zatem odwrotnością czasu życia nośników w bazie. Oznacza to inaczej, że przy tej pulsacji strumień nośników wpływających do bazy z obwodu bazy składa się z dwu strumieni o jednakowych amplitudach, lecz przesuniętych w fazie o $\pi/2$, jest to:

- strumień nośników większościowych kompensujących ładunek nośników mniejszościowych wstrzykiwanych z emitera (prąd ładowania pojemności dyfuzyjnej);
- strumień nośników większościowych uzupełniający ubytek koncentracji nośników wskutek rekombinacji w bazie (prąd rekombinacji typowy dla stanu statycznego i quasi-statycznego).

Z podstawowych właściwości jednobiegunowej funkcji przenoszenia $\beta(j\omega)$ (5.186) wynika, że dla $\omega > \omega_\beta$ każdemu n -krotnemu zwiększeniu pulsacji odpowiada n -krotne zmniejszenie modułu $|\beta(j\omega)|$. Zatem w tym zakresie pulsacji iloczyn $|\beta(j\omega)|\omega$ jest wielkością stałą. Ta właściwość charakterystyki amplitudowo-częstotliwościowej $|\beta(j\omega)| = f(\omega)$ stała się podstawą do zdefiniowania bardzo dogodnego pojęcia pulsacji (częstotliwości) granicznej ω_T (lub f_T), nazywanej często *pojemnością wzmacnienia*, gdyż

$$\omega_T = \omega |\beta(j\omega)| = \text{const przy } \omega > \omega_\beta \quad (5.193)$$

Tak zdefiniowaną częstotliwość graniczną można łatwo zmierzyć, gdyż wystarczy w tym celu zmierzyć moduł $|\beta(j\omega)|$ przy dowolnej częstotliwości w zakresie odpowiadającym nachyleniu charakterystyki amplitudowo-częstotliwościowej 6 dB/na oktawa.

Zanim zaczęto powszechnie stosować pojęcie pulsacji ω_T , wcześniej posługiwano się pojęciem pulsacji granicznej ω_1 , odpowiadającej spadkowi modułu $|\beta(j\omega)|$ do jedności. Jeżeli do (5.193) podstawimy $|\beta(j\omega)| = 1$, to

$$\omega_T = \omega_1 \quad (5.194)$$

dlatego często nie rozróżnia się tych dwu pojęć. Rzeczywista charakterystyka $|\beta(j\omega)| = f(\omega)$ przy $|\beta(j\omega)| \rightarrow 1$ ma nachylenie nieco mniejsze niż 6 dB na oktawa wskutek skrósnego przejścia sygnału przez pojemność rozproszoną kolektor-emiter. Dlatego $\omega_T = \omega_1$, jeżeli korzysta się z dwuodcinkowej aproksymacji charakterystyki amplitudowo-częstotliwościowej, lecz w przypadku pomiaru

rzeczywistej wartości ω_1 może ona być większa niż ω_T . Pulsacja ω_T jest związana z ω_β prostą zależnością wynikającą z dwuodcinkowej aproksymacji charakterystyki $|\beta(j\omega)| = f(\omega)$

$$\omega_T = \beta\omega_\beta \quad (5.195)$$

Jeżeli skorzystamy z (5.190), to

$$\omega_T \approx \omega_\alpha \quad (5.196)$$

Należy jednak zauważyć, że wzór (5.190) jest słuszny przy pominięciu poprawki fazowej, co inaczej odpowiada przyjęciu $\alpha(j\omega)$ w postaci jednobiegunowej funkcji reprezentującej tylko inercję (5.151). Taką postać $\alpha(j\omega)$ uzyskano przy założeniu, że stała czasowa dla sygnału sinusoidalnego jest równa czasowi przejścia sygnału skoku jednostkowego (czasowi przelotu nośników przez bazę w przypadku wzoru (5.151))

$$\tau_b = t_b$$

Przechodząc do sytuacji bardziej ogólnej należy stwierdzić, że o ile ω_α jest odwrotnością sumy stałych czasowych dla sygnału sinusoidalnego, o tyle ω_T jest odwrotnością sumy czasów przejścia skoku jednostkowego

$$\omega_T = \frac{1}{t_{eb} + t_b + t_{bc}} \quad (5.197)$$

czyli

$$\frac{1}{\omega_T} = t_{catk} = \begin{cases} \tau_N & \text{dla pracy normalnej} \\ \tau_i & \text{dla pracy inwersyjnej} \end{cases} \quad (5.197a)$$

jest całkowitym czasem przejścia skoku jednostkowego od emitera do kolektora. Obok łatwości pomiaru jest to druga ważna zaleta parametru ω_T , gdyż zawiera on w postaci bezpośredniej informację o właściwościach dynamicznych tranzystora zarówno w dziedzinie częstotliwości jak i w dziedzinie czasu, czyli umożliwia wyznaczenie parametrów dla pracy impulsowej za pomocą pomiarów małym sygnałem sinusoidalnym.

W ten sposób pulsacja ω_T jest możliwy dzięki temu, że jej odwrotność zawiera sumę stałych czasowych (wpływy zjawisk inercyjnych) i opóźnień (wpływy zjawisk wnoszących nadmiar fazy). Doskonale to ilustruje przekształcenie zależności $\beta(j\omega)$ określonej wzorem (5.186) w $\alpha(j\omega)$ z wykorzystaniem pojęcia pulsacji ω_T , zależności

$$(5.195) \text{ oraz znanego związku } \alpha(j\omega) = \frac{\beta(j\omega)}{\beta(j\omega) + 1}. \text{ W wyniku takiego przekształ-$$

cenia otrzymuje się

$$\alpha(j\omega) \approx \frac{\alpha}{1 + \frac{j\omega}{\omega_T}} \quad (5.198)$$

Zatem jest to funkcja jednobiegunowa bez czynnika opóźnienia (bez poprawki fazowej), odpowiadająca zależności (5.176), w której występuje oddzielnie czynnik inercyjny (funkcja wielobiegunowa) i czynnik opóźnienia ($e^{-j\omega t/\omega_\alpha}$).

Nie twierdźmy bynajmniej, że funkcje (5.198) i (5.176) są całkowicie równoważne, gdyż nie można iloczynem czynnika inercyjnego $\left(\frac{\alpha}{1 + j\omega/\omega_\alpha}\right)$ i opóźniającego ($e^{-j\omega t/\omega_\alpha}$)

zastąpić czynnikiem inercyjnym o innej stałej czasowej. Odpowiedniość tych wyrażeń polega na tym, że oba dają wystarczająco dokładne aproksymacje rzeczywistej charakterystyki fazowo-częstotliwościowej w zakresie $\omega < \omega_T$. Dokładność aproksymacji rzeczywistych charakterystyk częstotliwościowych modułu i fazy $\alpha(j\omega)$ dla $\omega < \omega_T$ można ocenić dla obu funkcji następująco:

$$\begin{aligned} \text{według (5.176)} \quad & \left\{ \begin{aligned} A(\omega) &= \frac{\alpha}{\sqrt{1 + \left(\frac{\omega}{\omega_\alpha}\right)^2}} \text{ — przybliżenie dokładne} \\ \varphi(\omega) &= -\arctg \frac{\omega}{\omega_\alpha} - m \frac{\omega}{\omega_\alpha} \text{ — przybliżenie dokładne} \end{aligned} \right. \\ \text{według (5.198)} \quad & \left\{ \begin{aligned} A(\omega) &= \frac{\alpha}{\sqrt{1 + \left(\frac{\omega}{\omega_T}\right)^2}} \text{ — przybliżenie niedokładne} \\ \varphi(\omega) &= -\arctg \frac{\omega}{\omega_T} \text{ — przybliżenie dokładne} \end{aligned} \right. \end{aligned}$$

Ponieważ parametr ω_T jest stosowany do opisu właściwości tranzystora w układzie WE (wówczas $\omega < \omega_T$) i — jak stwierdzono wcześniej — dominujący wpływ na wartość współczynnika $\beta(j\omega)$ ma przesunięcie fazowe między prądem kolektora i emitera, a nie różnica amplitud tych prądów, istotne zatem znaczenie ma tylko dokładność aproksymacji charakterystyki fazowo-częstotliwościowej. Pod tym względem funkcja (5.198) jest wystarczająco dokładna.

Ponieważ

$$t_{catk} \geq \tau_{catk}$$

zatem ogólnie

$$\omega_T \leq \omega_\alpha$$

Na przykład dla szczególnych przypadków ujętych zależnościami (5.188), (5.189) uwzględniając (5.195) otrzymujemy:

dla tranzystora dryftowego

$$\omega_T = \frac{\omega_\alpha}{1,2 + 0,1\eta} \tag{5.199}$$

oraz dla tranzystora bezdryftowego

$$\omega_T = 0,82\omega_\alpha \tag{5.200}$$

Związek trzech pulsacji granicznych można zapisać w postaci ogólnej (korzystamy z (5.185)):

$$\omega_\alpha \approx (1 + m)\omega_T \tag{5.201}$$

$$\omega_T = \beta\omega_\beta$$

przy czym $0 \leq m \leq 1$. Zatem

$$\omega_\alpha \geq \omega_T \gg \omega_\beta$$

Dotychczas rozpatrywane trzy pulsacje graniczne (ω_β , ω_T , ω_α) nie dają jeszcze pełnej informacji o właściwościach częstotliwościowych tranzystora. Zmniejszenie modułu współczynnika $|\beta(j\omega)|$ poniżej jedności nie wyklucza jeszcze możliwości uzyskania

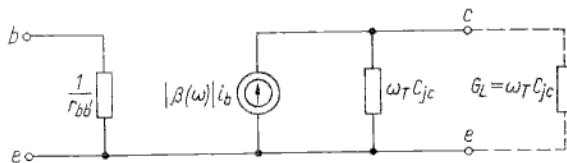
wzmocnienia mocy, gdyż impedancja wyjściowa jest większa niż impedancja wejściowa. Istnieje więc potrzeba zdefiniowania częstotliwości, powyżej której tranzystor nie wzmacnia mocy.

5.10.1.3

Pulsacja graniczna ω_{max}

Pulsacja graniczna ω_{max} jest definiowana jako taka pulsacja, przy której wzmożenie mocy jest równe jedności (ω_{max} nie zależy od układu włączenia tranzystora). Jeżeli wzmożenie mocy jest większe od jedności, to tranzystor może pracować w układzie generatora. Dlatego ω_{max} nazywa się również *maksymalną pulsacją generacji* tranzystora. Warunkiem uzyskania maksymalnego wzmożenia mocy jest dopasowanie na wejściu i wyjściu (impedancja źródła sterującego powinna być równa sprzężonej impedancji wejściowej tranzystora, a impedancja obciążenia — sprzężonej impedancji wyjściowej tranzystora). Należy przy tym spełnić warunek stabilności, tj. zneutralizować wpływ wewnętrznego sprzężenia zwrotnego w tranzystorze przez odpowiedni zewnętrzny obwód bez strat (w przypadku pełnej neutralizacji mówi się, że tranzystor jest zuniliteryzowany). Oszacujemy wzmożenie mocy tranzystora spełniającego te warunki.

W tym celu przedstawimy uproszczony schemat zastępczy tranzystora dla bardzo dużych częstotliwości, w którym będą nas interesowały tylko amplitudy (albo wartości skuteczne) prądów oraz części rzeczywiste admitancji wejściowej i wyjściowej. Taki schemat dla układu WE przedstawiono na rys. 5.70. Przyjęto, że w obwodzie



Rys. 5.70

Uproszczony schemat zastępczy, służący do oszacowania wzmożenia mocy dla bardzo dużych częstotliwości

wejściowym reaktancja pojemnościowa jest pomijalnie mała (oznacza to zwarcie węzłów b' , e w schemacie „hybryd π ”), a część rzeczywistą admitancji wyjściowej wyznaczono na podstawie spostrzeżenia, że dla $\omega = \omega_T$, tj. dla $|\beta(\omega)| = 1$, część rzeczywista i urojona admitancji wyjściowej są sobie równe. Stąd:

$$\text{Re } Y_{wy} = \omega_T C_{jc}$$

$$\text{Re } Y_{we} = \frac{1}{r_{bb'}}$$

Moc na wyjściu w stanie dopasowania (połowa prądu ze źródła $|\beta(\omega)|i_b$ płynie przez konduktancję wyjściową, a połowa przez obciążenie)

$$P_{wy} = \frac{|\beta(\omega)|^2 i_b^2}{4} \frac{1}{\omega_T C_{jc}}$$

Moc na wejściu

$$P_{we} = i_b^2 r_{bb'}$$

Stąd wzmożenie mocy

$$k_p(\omega) = \frac{P_{wy}}{P_{we}} = \frac{|\beta(\omega)|^2}{4\omega_T} \frac{1}{r_{bb'} C_{jc}}$$

Ponieważ $\omega_T = |\beta(\omega)|\omega$, zatem

$$k_p(\omega) = \frac{\omega_T}{4\omega^2} \frac{1}{r_{bb'} C_{jc}}$$

Podstawiając $k_p = 1$ otrzymuje się

$$\omega_{max} = \sqrt{\frac{\omega_T}{4r_{bb'} C_{jc}}} \quad (5.202)$$

Posługując się definicją maksymalnej pulsacji generacji ω_{max} w prosty sposób można przedstawić zależność wzmożenia mocy od częstotliwości

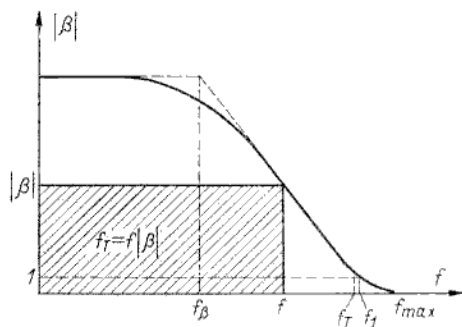
$$k_p(\omega) = \left(\frac{\omega_{max}}{\omega}\right)^2 \quad (5.203)$$

Jest to zależność słuszna tylko w zakresie dużych częstotliwości (dla $\omega \gg \omega_{max}$) $\sqrt{k_p(0)}$, w którym $k_p(\omega)$ maleje 6 dB na oktawę.

Częstotliwości graniczne — definicje techniczne

5.102

Definicje częstotliwości granicznych, przyjęte w technice, wynikają konsekwentnie z rozpatrzonych właściwości fizycznych tranzystora. Dlatego w tym punkcie wymienimy bez komentarza pięć definicji (rys. 5.71):



Rys. 5.71

Charakterystyka częstotliwościowa modułu współczynnika wzmożenia prądowego $|\beta|$ z zaznaczonymi częstotliwościami granicznymi

- f_x — częstotliwość, przy której moduł zwarciovego współczynnika wzmożenia prądowego $|\alpha(f)|$ w układzie WB maleje o 3dB, czyli do poziomu $\alpha/\sqrt{2}$, przy czym α jest wartością $\alpha(f)$ dla $f \rightarrow 0$;
- f_β — częstotliwość, przy której moduł zwarciovego współczynnika wzmożenia prądowego $|\beta(f)|$ w układzie WE maleje o 3 dB, czyli do poziomu $\beta/\sqrt{2}$, przy czym β jest wartością $\beta(f)$ przy $f \rightarrow 0$;
- f_T — iloczyn częstotliwości i modułu współczynnika wzmożenia prądowego $|\beta(f)|$ zmierzony w zakresie $f > f_\beta$ (w zakresie liniowego spadku modułu $|\beta(f)|$ w funkcji częstotliwości);

f_1 — częstotliwość, przy której moduł współczynnika wzmocnienia prądowego $|\beta(f)|$ w układzie WE równa się jedności;

f_{max} — maksymalna częstotliwość generacji, czyli częstotliwość, przy której wzmocnienie mocy równa się jedności.

Ogólnie można ułożyć częstotliwości graniczne w następującym szeregu:

$$f_{max} > f_\alpha \geq f_1 \geq f_T \geq f_\beta \quad (5.204)$$

przy czym: $f_T = \beta f_\beta$; $f_\alpha = (1, 2, \dots, 2)f_T$; oraz często przyjmuje się $f_1 = f_T$.

Częstotliwości graniczne wyrażone przez elementy modelu „hybryd π ”

5.10.3

Rozpatrzmy schemat zastępczy dla układu WE, przedstawiony na rys. 5.61. Wyznaczamy współczynnik wzmocnienia prądowego

$$\beta(j\omega) = \frac{i_{wy}}{i_{we}}$$

w stanie zwarcia na wyjściu:

$$i_{wy} = g_m u_{b'e}$$

$$i_{we} = u_{b'e} [g_{b'e} + g_{b'c} + j\omega(C_{b'e} + C_{b'c})]$$

Stąd, uwzględniając, że $g_{b'c} \ll g_{b'e}$

$$\beta(j\omega) \approx \frac{g_m}{g_{b'e} + j\omega(C_{b'e} + C_{b'c})}$$

lub uwzględniając (5.135)

$$\frac{\beta(j\omega)}{\beta} = \frac{1}{1 + j\omega(C_{b'e} + C_{b'c})r_{b'e}} \quad (5.205)$$

Z tej zależności można wyznaczyć f_β oraz $f_T = f_1$. Częstotliwość f_β wyznacza się przyrównując moduł funkcji (5.205) do $1/\sqrt{2}$ (lub przyrównując część rzeczywistą do części urojonej)

$$f_\beta = \frac{1}{2\pi r_{b'e} (C_{b'e} + C_{b'c})} \quad (5.206)$$

Częstotliwość $f_T = f_1$ wyznacza się przyrównując moduł funkcji (5.205) do $1/\beta$

$$f_T \approx \frac{g_m}{2\pi (C_{b'e} + C_{b'c})} \quad (5.207)$$

Postępując analogicznie można wyznaczyć f_α na podstawie schematu zastępczego dla układu WB (rys. 5.64). Pomijając wpływ $r_{bb'}$ oraz g_{ec} , otrzymuje się

$$f_\alpha \approx \frac{g_{eb'}}{2\pi C_{eb'}} \quad (5.208)$$

Metody wyznaczania parametrów modelu „hybryd π ”

5.11

Parametr g_m można wyznaczyć z zależności fizycznej

$$g_m = \frac{|I_C|}{\varphi_T} \quad (5.209)$$

W temperaturze pokojowej ($T = 290$ K) $\varphi_T = 25$ mV, a więc jeśli I_C wyrazimy w miliamperach, to

$$g_m = 40 |I_C| \quad [\text{mA/V}] \quad (5.210)$$

Parametr $g_{b'e}$ wyznacza się na podstawie pomiaru zwarcioowego współczynnika wzmocnienia prądowego dla małych częstotliwości β (inaczej h_{21e}). Można w tym przypadku skorzystać z danych katalogowych, które jednak podają parametr h_{21e} (dla $f = 1$ kHz) z rozrzutem ponad 100-procentowym

$$g_{b'e} = \frac{g_m}{\beta} \quad (5.211)$$

Parametr $r_{bb'}$ można wyznaczyć na podstawie pomiaru zwarciowej rezystancji wejściowej h_{11e} (pomiar przy małej częstotliwości, na przykład $f = 1$ kHz). Ponieważ

$$h_{11e} = r_{bb'} + r_{b'e} \quad (5.212)$$

zatem

$$r_{bb'} = h_{11e} - \frac{\beta}{g_m} \quad (5.213)$$

Ścisłe biorąc, parametr $r_{bb'}$ zależy od częstotliwości, gdyż rezystancja rozprószonej bazy i pojemność złącza $B-C$ tworzą obwód o stałych rozłożonych. W zakresie bardzo dużych częstotliwości, dla których jest spełniony warunek

$$\omega(C_{b'e} + C_{b'c}) \gg \frac{1}{r_{bb'}}$$

można wyznaczyć $r_{bb'}$ na podstawie pomiaru składowej rzeczywistej zwarciowej admitancji wejściowej

$$\frac{1}{r_{bb'}} = \text{Re } y_{11e} \quad (5.214)$$

Parametr $g_{b'c}$ można wyznaczyć na podstawie pomiaru rozwarciowego współczynnika oddziaływania wstecznego h_{12e} w zakresie małych częstotliwości (najczęściej $f = 1$ kHz). Ponieważ dla rozpatrywanego modelu

$$h_{12e} = \frac{u_1}{u_2} = \frac{r_{b'e}}{r_{b'e} + r_{b'c}} \quad (5.215)$$

stosując zatem tradycyjny zapis skrótowy $h_{12e} \equiv \mu$ oraz biorąc pod uwagę, że zawsze $r_{b'c} \gg r_{b'e}$, jest

$$g_{b'c} = \mu g_{b'e} = \frac{\mu}{\beta} g_m \quad (5.216)$$

Należy zauważyć, że $g_{b'c}$ można również wyznaczyć na podstawie pomiaru rozwarekowej konduktancji wyjściowej w układzie WB

$$g_{b'c} = h_{22b} \quad (5.217)$$

Parametr g_{ce} można wyznaczyć na podstawie pomiaru rozwarekowej konduktancji wyjściowej h_{22e}

$$g_{ce} = h_{22e} \quad (5.218)$$

Ponieważ

$$h_{22e} = (\beta + 1)h_{22b} \approx \beta h_{22b} \quad (5.219)$$

uwzględniając zatem (5.217), (5.216)

$$g_{ce} \approx \beta g_{b'c} \approx \mu g_m \quad (5.220)$$

Pojemność $C_{b'c}$ wyznacza się na podstawie pomiaru pojemności wyjściowej C_{22b} w układzie WB przy rozwartym wejściu. Jeżeli jest spełniony warunek

$$\frac{1}{\omega C_{b'c}} \gg r_{bb'}$$

to

$$C_{b'c} \approx C_{22b} \quad (5.221)$$

W przypadku podziału pojemności $C_{b'c}$ na dwie składowe rozdzielone rezystancją $r_{bb'}$:

$$C_{bc} = (1 - k)C_{22b} \quad (5.222)$$

$$C_{b'c} = kC_{22b} \quad (5.223)$$

przy czym k — współczynnik dobierany eksperymentalnie.

Pojemność $C_{b'e}$ wyznacza się z pulsacji granicznej ω_T . Na podstawie (5.207)

$$C_{b'e} = \frac{g_m}{\omega_T} - C_{b'c} \quad (5.224)$$

W ten sposób określono metody wyznaczania wszystkich parametrów modelu „hybryd π ” dla konfiguracji WE. Otrzymane wyrażenia zestawiono w tabl. 5.4.

Wybrane zagadnienia technologii i konstrukcji tranzystorów bipolarnych

5.12

W rozwoju technologii i konstrukcji tranzystorów bipolarnych najwięcej uwagi poświęcono trzem problemom:

- poprawie niezawodności,
- zwiększeniu częstotliwości granicznej,
- zwiększeniu maksymalnej mocy rozpraszanej.

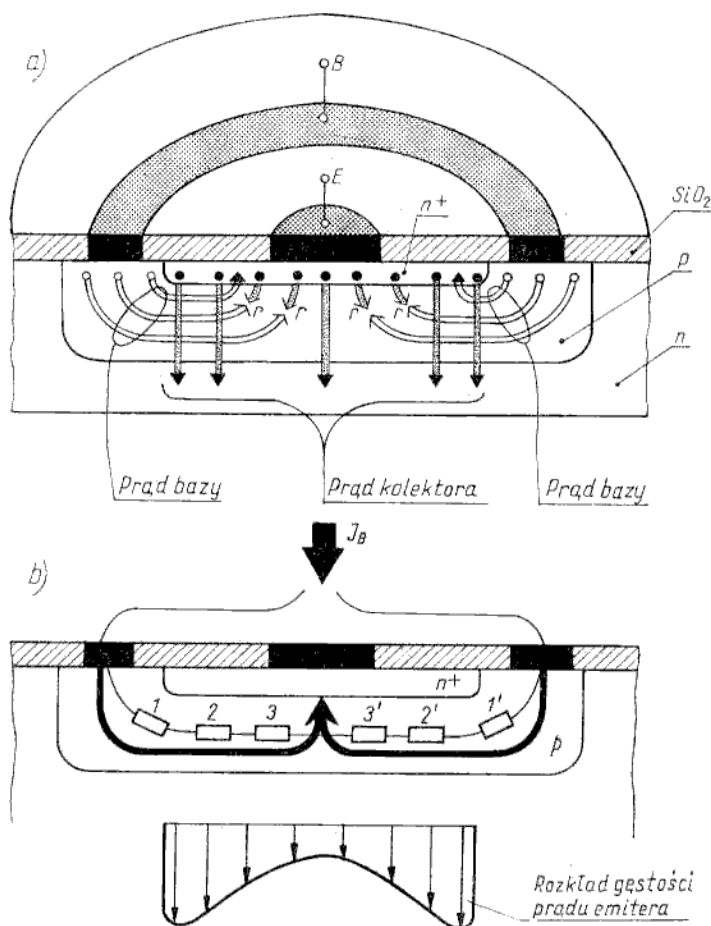
Zagadnienie niezawodności elementów półprzewodnikowych nie będzie rozpatrywane; trzeba tylko zwrócić uwagę, że łączy się ono głównie ze zjawiskami fizycznymi, zachodzącymi na powierzchni półprzewodnika, jakością mikromontażu i obudowy. Niewątpliwie przełomowym wydarzeniem w tym aspekcie było wykorzystanie warstwy SiO_2 do pasywacji powierzchni krzemu (technologia planarna).

Prace nad dwoma kolejnymi problemami (zwiększenie mocy admysyjnej i częstotliwości granicznej) doprowadziły do powstania dwu specyficznych klas tranzystorów tj. tranzystorów mocy i tranzystorów dla zakresu b.w.cz. (mikrofalowych). Omówimy teraz te dwie klasy tranzystorów, przedtem jednak należy nieco szerzej skomentować pojęcie rezystancji rozproszonej bazy $r_{bb'}$.

Rezystancja rozproszona bazy

5.12.1

Z pojęciem rezystancji rozproszonej bazy zetknęliśmy się już wielokrotnie. W modelach tranzystorów uwzględnialiśmy tę rezystancję jako element skupiony podczas gdy obecnie weźmiemy pod uwagę jej rozłożony charakter. Prąd bazy płynie w kierunku prostopadłym do prądu emitera i kolektora. W przypadku struktury o symetrii kołowej, przedstawionej na rys. 5.72a, prąd bazy płynie



Rys. 5.72

Rozpływ prądów w tranzystorze o symetrii kołowej (a) oraz wpływ rezystancji obszaru bazy na nierównomierność gęstości prądu emitera (b)

promieniście od pierścienia elektrody bazy do środka koła utworzonego przez obszar emitera. Jeżeli emiter ma dużą powierzchnię, to droga prądu bazy jest dość długa i trzeba się liczyć ze znacznym spadkiem napięcia na rezystancji obszaru bazy. Warstwę bazy można zastąpić modelem w postaci łańcuchów rezystorów rozłożonych promieniście (rezystory 1, 2, 3 oraz 1', 2', 3' na rys. 5.72b).

Załóżmy, że między elektrodami metalowymi emitera i bazy jest różnica napięć 0,7 V (typowe napięcie polaryzacji dla złącza krzemowego, pracującego w kierunku przewodzenia). Przyjmijmy również, że spadek napięcia na łańcuchu rezystorów 1, 2, 3 (lub 1', 2', 3') od przepływającego prądu bazy wynosi 0,3 V (po 0,1 V na każdy rezystor). Ponieważ silnie domieszkowana warstwa emitera (n^+) tworzy powierzchnię ekwipotencjalną, efektywna różnica napięć między tą warstwą a warstwą bazy maleje promieniście od 0,7 V na okręgu koła utworzonego przez warstwę n^+ (na krawędzi emitera) do 0,4 V w środku tego koła. Wskutek nierównomiernego polaryzacji złącza emiter-baza nierównomierny jest również rozkład gęstości prądu emitera. Gęstość prądu jest największa na krawędzi emitera (napięcie polaryzacji 0,7 V), a najmniejsza w środku koła (napięcie polaryzacji 0,4 V). Prąd emitera jest jakby wypychany na peryferie obszaru emitera. To zjawisko, nazywane *wypychaniem* (lub wyciskaniem) prądu emitera, ma bardzo niekorzystny wpływ na pracę tranzystora mocy, gdyż powoduje zmniejszenie czynnej powierzchni złącza E-B (powierzchni, przez którą płynie przeważająca część prądu emitera). Nie są to wszystkie ujemne skutki istnienia rezystancji rozproszonej bazy $r_{bb'}$. Najistotniejsze z nich są:

- zmniejszenie maksymalnej częstotliwości generacji;
- straty mocy sygnału wejściowego przy pracy w układzie WE;
- wypychanie prądu emitera;
- zwiększenie współczynnika szumów.

Wartość rezystancji $r_{bb'}$ zależy od poziomu domieszkowania bazy i jej wymiarów geometrycznych. Wzory szczegółowe na $r_{bb'}$ odnoszą się do konkretnych struktur geometrycznych tranzystora, dlatego nie będą tu przytaczane. Należy tylko zauważyć, że w przypadku rozpatrywanej struktury kołowej, jeżeli uwzględnimy tylko obszar bazy o średnicy wyznaczonej przez krawędź emitera, to

$$r_{bb'} \sim \frac{\rho}{W_B} \quad (5.225)$$

Natomiast nierównomierność gęstości prądu emitera jest tym mniejsza, im mniejsza jest średnica emitera oraz im bliżej emitera znajduje się elektroda bazy.

Ten wniosek określa najważniejszą tendencję w konstrukcji tranzystorów dużej mocy, polegającą na zastąpieniu jednej warstwy emitera o dużej powierzchni połączeniem równoległym warstw emitera o małych powierzchniach przy takiej samej powierzchni sumarycznej. Najogólniej tę tendencję można sformułować jako dążenie do uzyskania jak największego stosunku długości krawędzi emitera do jego powierzchni.

Tranzystory dla zakresu b.w.cz.

5.12.2

Tranzystorami dla zakresu b.w.cz. (inaczej mikrofalowymi) nazywa się tranzystory, których częstotliwość f_T mieści się w zakresie mikrofalowym, czyli jest

większa niż 1 GHz (niekiedy przyjmuje się za dolną granicę tego zakresu $f = 300$ MHz). Wszystkie tranzystory mikrofalowe dzieli się na dwie grupy:

- tranzystory małej mocy, których najważniejszymi parametrami są: częstotliwość graniczna, wzmocnienie i współczynnik szumów;
- tranzystory dużej mocy, których najważniejszymi parametrami są: częstotliwość graniczna, moc rozpraszana, wzmocnienie mocy.

Tranzystory dużej mocy będą omówione w następnym punkcie. W tranzystorach mikrofalowych małej mocy najważniejszym zadaniem jest uzyskanie jak największej wartości częstotliwości f_T (zgodnie z wymaganiem dużej częstotliwości granicznej i dużego wzmocnienia) oraz jak najmniejszej wartości rezystancji rozproszonej bazy $r_{bb'}$ (im mniejsza jest ta rezystancja, tym mniejsze są szumy tranzystora).

Oszacujmy fizyczne ograniczenia częstotliwości f_T . Na podstawie rozważań przeprowadzonych w p. 5.10 wiadomo, że pulsacja ω_T jest odwrotnością czasu przejścia sygnału od emitera do kolektora, na który składają się:

- czas ładowania warstwy zaporowej złącza $E-B$ t_{eb} ;
- czas przelotu nośników przez bazę t_b ;
- czas ładowania warstwy zaporowej złącza $B-C$ t'_{bc} ;
- czas przelotu nośników przez warstwę zaporową złącza $B-C$ t''_{bc} .

Uwzględniając zatem, że rezystancja przyrostowa złącza $E-B$

$$r_{eb'} = \frac{\varphi_T}{I_E}$$

$$f_T = \frac{1}{2\pi\tau_N} = \left\{ 2\pi \left[\underbrace{\frac{\varphi_T}{I_E} C_{je}}_I + \underbrace{\frac{W_B^2}{kD_n}}_{II} + \underbrace{r_c(C_{jc} + C_{dc})}_{III} + \underbrace{\frac{l_{ab-c}}{2v}}_{IV} \right] \right\}^{-1} \quad (5.226)$$

przy czym k — współczynnik zależny od rozkładu koncentracji domieszek w bazie. Dążąc do minimalizacji poszczególnych składników tego wyrażenia, można sformułować wymagania odnośnie do konstrukcji i technologii tranzystorów mikrofalowych:

- małe powierzchnie złączy (zmniejszenie składnika I i III);
- cienka baza, duży gradient koncentracji domieszek w bazie (duża wartość k) oraz dobór materiału o dużej wartości współczynnika dyfuzji D_n (zmniejszenie składnika II);
- mała wartość rezystancji szeregowej kolektora r_c (zmniejszenie składnika III);
- cienka warstwa zaporowa złącza $B-C$ (zmniejszenie składnika IV).

Wpływ prądu emitera jest bardziej złożony, niż to wynika ze wzoru (5.226). Przy małych wartościach I_E dominuje pierwszy składnik i wówczas f_T wzrasta proporcjonalnie do I_E . W miarę wzrostu prądu emitera składnik I ma coraz mniejsze znaczenie, zwiększa się natomiast wartość składnika II (k zależy od prądu emitera przy dużych poziomach wstrzykiwania) i częstotliwość f_T zmniejsza się.

Na wartość szybkości unoszenia nośników w warstwie zaporowej nie mamy istotnego wpływu, gdyż w silnych polach jest ona w przybliżeniu stała i równa prędkości cieplnej ($v \approx 6 \cdot 10^4$ m/s). Realizacja niektórych wymagań prowadzi do sprzeczności z innymi (szczególnie, jeśli uwzględnimy również postulaty dużej wartości napięcia przebicia i małych szumów) i można im zadośćuczynić tylko w drodze kompromisu. Na przykład szerokość warstwy zaporowej złącza $B-C$ maleje w miarę wzrostu koncentracji domieszek, co jednak powoduje zarazem zmniejszenie wartości współczynnika k oraz zwiększenie pojemności C_{jc} , poza tym maleje wówczas wartość napięcia przebicia. Z kolei zmniejszenie szerokości bazy W_B powoduje zmniejszenie napięcia przebicia i wzrost wartości $r_{bb'}$, czyli wzrost współczynnika

szumów. Składnik III jest pomijalnie mały ze względu na małą wartość rezystancji r_c w tranzystorach epiplanarnych. W tranzystorach mikrofalowych o submikrometrowej szerokości bazy ($W_B \leq 0,2 \mu\text{m}$) dominujące znaczenie mają składniki I i IV. Na przykład w tranzystorze o $f_T = 6 \text{ GHz}$: $\tau_{eb} \approx 10 \text{ ps}$, $\tau_b \approx 2 \text{ ps}$, $\tau_{bc}'' = 12 \text{ ps}$.

Teoretycznie można sobie wyobrazić tranzystor o nieskończenie małych powierzchniach złączy i nieskończenie małej szerokości bazy. W takim tranzystorze f_T byłoby określone tylko składnikiem IV, przy czym zamiast $l_{d\text{-}b\text{-}c}$ należałoby wziąć szerokość warstwy ładunku przestrzennego, rozciągającego się od emitera do kolektora. Wówczas

$$f_T \approx \frac{1}{2\pi} \frac{v}{l_d} \quad (5.227)$$

przy czym $v \approx 6 \cdot 10^4 \text{ m/s}$.

Wielkości l_d nie można uczynić dowolnie małą, ponieważ natężenie pola

$$E = \frac{U_{CE}}{l_d}$$

zwiększałoby się do nieskończoności. Nie jest to możliwe, gdyż istnieje pewne krytyczne natężenie pola, przy którym następuje przebicie półprzewodnika. Pole to wynosi:

$$\text{dla krzemu} \quad E_{kr} \leq 2 \cdot 10^7 \text{ V/m}$$

$$\text{dla germanu} \quad E_{kr} \leq 10^7 \text{ V/m}$$

Dogodnym kryterium jakości tranzystorów mikrofalowych jest wartość iloczynu $f_T U_{CE\text{max}}$. Przyjmując górne wartości v , E_{kr} , można obliczyć maksymalną wartość tego iloczynu (dla krzemu)

$$f_T U_{CE\text{max}} = \frac{1}{2\pi} v E_{kr} = 200 \text{ V} \cdot \text{GHz} \quad (5.228)$$

Jeżeli przyjmiemy, że praktycznie najmniejszą akceptowalną wartością napięcia przebicia jest $U_{CE\text{max}} = 5 \text{ V}$, to

$$f_T = 40 \text{ GHz}$$

To zgrubne oszacowanie daje pogląd o teoretycznym ograniczeniu wartości f_T . W produkowanych współcześnie tranzystorach bardzo małej mocy uzyskuje się maksymalne wartości:

$$f_T U_{CE\text{max}} \approx 100 \text{ V} \cdot \text{GHz}$$

$$f_T \approx 15 \text{ GHz}$$

Są to zatem wartości niewiele odbiegające od teoretycznie możliwych.

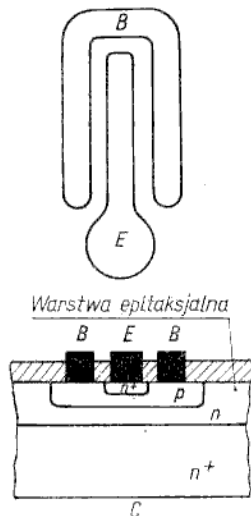
Ponieważ obok częstotliwości granicznej drugim bardzo istotnym parametrem jest współczynnik szumów, należy dodać, że rekordowe pod tym względem osiągnięcia to:

$$F = 1 \dots 2 \text{ dB} \quad \text{przy } f = 1 \text{ GHz}$$

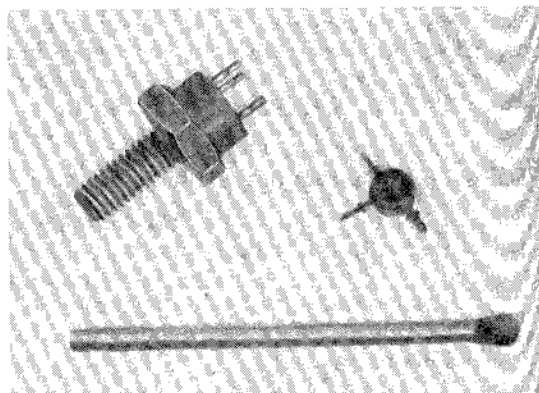
oraz

$$F = 4 \dots 5 \text{ dB} \quad \text{przy } f = 6 \text{ GHz}$$

Współczesne tranzystory mikrofalowe są to prawie wyłącznie krzemowe tranzystory $n-p-n$ o strukturze epiplanarnej. Mają one najczęściej „międzypalczastą” strukturę geometryczną (rys. 5.73), która umożliwia uzyskanie minimalnej powierzchni emitera i małej wartości rezystancji $r_{bb'}$. Parametry tranzystora mikrofalowego



Rys. 5.73
Struktura międzypalczasta tranzystora mikrofalowego małej mocy



Rys. 5.74
Typowe obudowy tranzystorów mikrofalowych

bardzo istotnie zależą od sposobu wykonania obudowy i końcówek. Chodzi o to, by pojemności pasożytnicze i indukcyjności doprowadzeń były jak najmniejsze. Typowe rodzaje obudów pokazano na rys. 5.74.

Tranzystory mocy

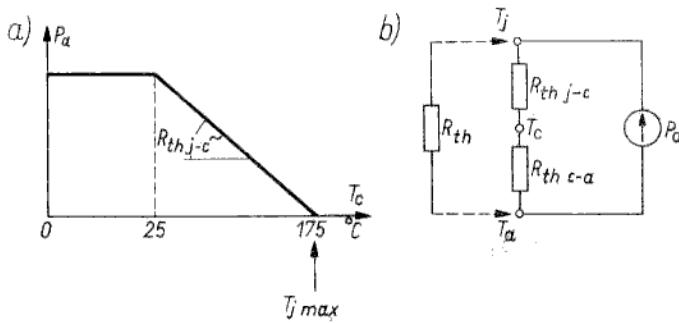
5.12.3

Uzyskanie dużej mocy rozpraszanej przez tranzystor¹⁾ stanowi niemały problem, którego trudność szczególnie wzrasta, jeżeli jest wymagana zarazem duża wartość częstotliwości granicznej. Dlatego można mówić o dwu klasach tranzystorów mocy: dla małej i dla dużej częstotliwości. W zakresie małych częstotliwości za kryterium dużej mocy przyjmuje się $P_a \geq 10$ W lub bardziej ogólnie wartość oporu cieplnego $R_{thj-c} \leq 15^\circ \text{C/W}$. Moc admissyjna (rys. 5.75)

$$P_a = \frac{T_{jmax} - T_c}{R_{thj-c}} \quad (5.229)$$

przy czym: T_{jmax} — maksymalna dopuszczalna temperatura struktury półprzewodnikowej; T_c — temperatura obudowy (indeks c od ang. case — obu-

¹⁾ Uzyskanie dużej mocy rozpraszanej przez tranzystor nie jest celem samym w sobie, gdyż ideałem byłaby taka praca tranzystora, przy której cała moc wydzielalaby się w obciążeniu przy zerowych stratach w samym tranzystorze. Jednak rzeczywiste układy nigdy nie mają 100% sprawności energetycznej, co oznacza, że zawsze część mocy jest tracona w tranzystorze. Przy określonej sprawności układu wartość bezwzględna mocy wydzielanej w obciążeniu jest ograniczona przez dopuszczalną moc rozpraszaną w tranzystorze.



Rys. 5.75

Wykres zależności $P_a(T_c)$ (a); schemat elektryczny równoważny wyrażeniom (5.229) i (5.230) (b)

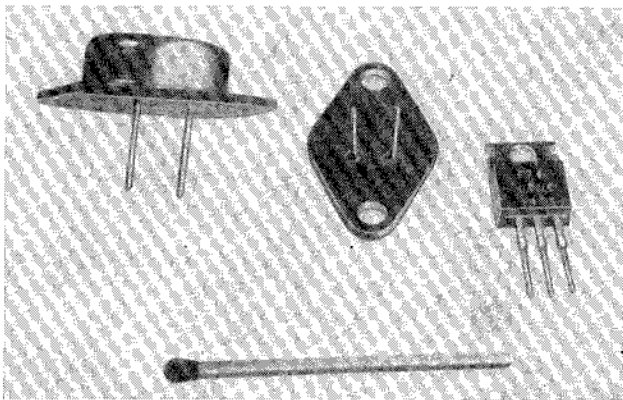
dowa); R_{thj-c} — tzw. własny opór cieplny, czyli opór cieplny na drodze od struktury półprzewodnikowej do zewnętrznych ścianek obudowy.

Temperatura T_{jmax} wynosi dla krzemu 150...200°C (jest to temperatura, przy której koncentracja nośników samoistnych staje się porównywalna z koncentracją domieszek, czyli jest ona tym większa, im większa jest szerokość pasma zabronionego półprzewodnika i koncentracja domieszek). Jeżeli zapewni się temperaturę obudowy równą temperaturze otoczenia (ok. 20°C), to ze wzoru (5.229) uzyskuje się $P_a \approx 10$ W przy $R_{thj-c} \approx 15$ °C/W. Ponieważ istnieje pewien opór cieplny między obudową a otoczeniem (oznaczymy go R_{thc-a}), zatem w rzeczywistości moc admissyjna jest mniejsza

$$P_a = \frac{T_{jmax} - T_a}{R_{th}} \quad (5.230)$$

przy czym: T_a — temperatura otoczenia (indeks a od ang. *ambient*); R_{th} — całkowity opór cieplny równy ($R_{thj-c} + R_{thc-a}$).

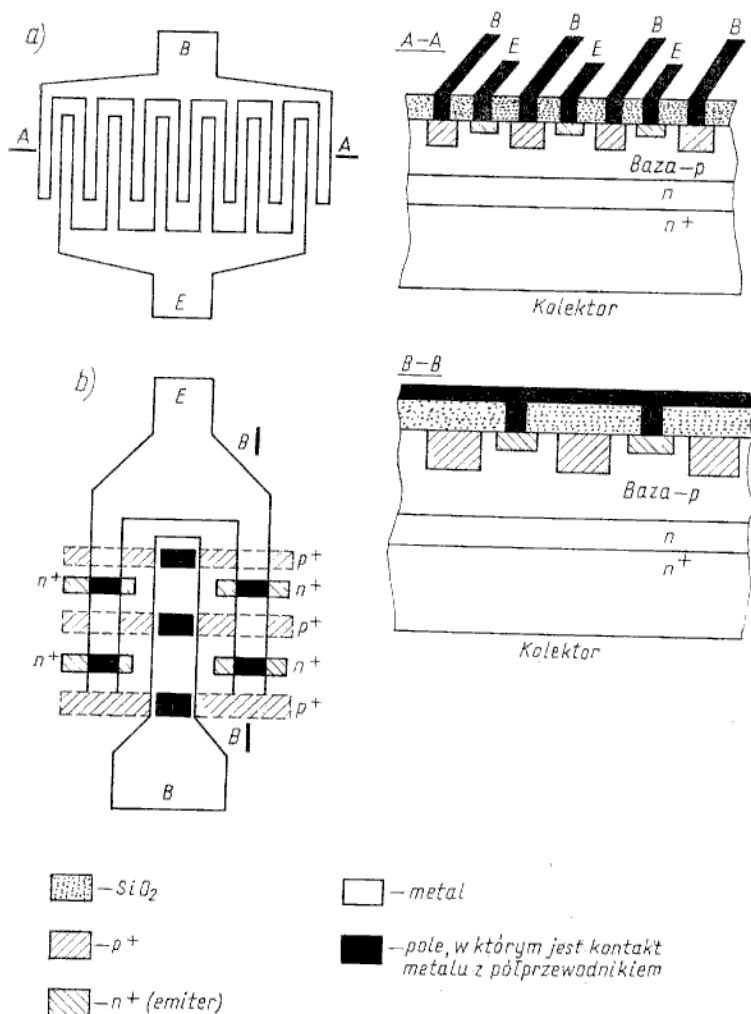
Zwykle jednak korzysta się ze wzoru (5.229) z uwagi na trudności w wyznaczaniu oporu R_{thc-a} . Dąży się przy tym, by $T_c \approx T_a$, co uzyskuje się przez stosowanie radia-



Rys. 5.76

Typowe obudowy tranzystorów mocy

torów, niekiedy nawet chłodzonych wymuszonym obiegiem powietrza lub cieczy chłodzącej. W celu zapewnienia lepszego odprowadzenia ciepła tranzystory mocy mają duże obudowy o jednej płaskiej i gładkiej powierzchni zapewniającej dobry styk z radiatorem (rys. 5.76). Zwykle obudowa spełnia również funkcję elektrody kolektora, co zapewnia dobre odprowadzenie ciepła z obszaru złącza $B-C$, w którym powstaje go najwięcej. Jeżeli obok dużej mocy jest również wymagana duża częstotliwość graniczna, to istotne znaczenie ma nie tylko dobre odprowadzenie ciepła, lecz również równomierny rozkład gęstości prądu emitera. To ostatnie wymaganie spełniają konstrukcje, w których uzyskuje się duży stosunek długości krawędzi (obwodu) emitera do jego powierzchni. Dwa najczęściej stosowane rozwiązania konstrukcyjne przedstawiono na rys. 5.77. Są to struktury



Rys. 5.77
Dwie najczęściej stosowane konstrukcje tranzystorów mocy dla zakresu b.w.cz.: a) grzebieniowa; b) „overlay”

grzebieniowa (rys. 5.77a) oraz „overlay” (rys. 5.77b). Najlepsze rezultaty zapewnia struktura „overlay”, której nazwa pochodzi stąd, że obszary baz typu p są połączone równolegle przez warstwy p^+ , a nad nimi (na powierzchni SiO_2) jest naniesiona warstwa metalizacji, łącząca poszczególne emitery (jest ich zwykle ponad 100)¹⁾.

Przedstawione konstrukcje zapewniają równomierny rozkład gęstości prądu emitera i małe wartości rezystancji $r_{bb'}$. Ten drugi czynnik jest o tyle istotny, że wpływa na maksymalną częstotliwość generacji f_{max} , a w tranzystorach mocy dla zakresu b.w.cz. miarodajna jest raczej wartość f_{max} , a nie f_T , gdyż tranzystor może być użytecznym wzmacniaczem mocy w zakresie $f > f_T$ (por. wzory: (5.203) na wzmożenie mocy i (5.202) na częstotliwość f_{max}).

Wykaz ważniejszych wzorów

Tablica 5.6

<p>Współczynniki wzmożenia prądowego</p>	$\alpha_N = \frac{I_C - I_{CBO}}{I_E}; \quad \alpha = \frac{\Delta I_C}{\Delta I_E}$ $\beta_N = \frac{I_C - I_{CBO}}{I_B + I_{CBO}}; \quad \beta = \frac{\Delta I_C}{\Delta I_B}$ $\alpha_N = \frac{\beta_N}{1 + \beta_N}; \quad \alpha = \frac{\beta}{1 + \beta}$ $\alpha_N = \alpha_e \alpha_b$ <p>przy czym:</p> $\alpha_e = \frac{I_{nE}}{I_E} = \frac{1}{1 + K + L}$ $K = \frac{D_p N_{AB} W_B}{D_n N_{DE} W_E} \frac{1 - e^{-\eta}}{\eta}$ $L = \frac{1}{2} \frac{N_{AB} W_B W_{EB}}{n_i D_n \tau} \frac{1 - e^{-\eta}}{\eta} \exp\left(-\frac{U_{EB}}{2\varphi_T}\right)$ $\alpha_b = \frac{I_C}{I_{nE}} = \frac{\tau_r}{\tau_r + \tau_b} \approx 1 - \left(\frac{W_B}{L}\right)^2 \frac{\eta^{-1+e^{-\eta}}}{\eta^2}$
<p>Prądy zerowe</p>	$I_{CEO} = \frac{M I_{CBO}}{1 - \alpha_N M}$
<p>Napięcia przebicia</p>	$U_{CEO max} \approx \frac{U_{CBO max}}{\sqrt[n]{\beta_N}}$
<p>Pojemności złączowe</p>	$C_{je} = C_{je0} \left(1 - \frac{U_{BE}}{\varphi_E}\right)^{-n_E}$ $C_{jc} = C_{jc0} \left(1 - \frac{U_{CE}}{\varphi_C}\right)^{-n_C}$

¹⁾ Tranzystor o strukturze „overlay” nie ma dotychczas polskiej nazwy. Nazwa tranzystor wieloemiterowy (używana w literaturze radzieckiej) została już nadana innej strukturze stosowanej w układach scalonych TTL. Najbliższa angielskiej byłaby nazwa „tranzystor wia-dukcyjny”, która kojarzy się z najistotniejszą cechą tej konstrukcji, czyli „bezkolizyjnym” krzyżowaniem ścieżek metalizacji łączącej emitery i ścieżek p^+ łączących bazy.

Tabl. 5.6 (cd.)

<p>Pojemności dyfuzyjne (składowe iniekcyjne)</p>	$C_{dc}^I \approx \frac{\tau_N}{\varphi_T} (I_{dE} + I_{ES})$ $C_{dc}^{I'} \approx \frac{\tau_I}{\varphi_T} (I_{dC} + I_{CS})$
<p>Częstotliwości graniczne</p>	$\alpha_i(j\omega) = \frac{\alpha_i}{1 + \frac{j\omega}{\omega_{\alpha i}}} e^{-j m_b \omega / \omega_{\alpha i}}$ <p>przy czym $\omega_{\alpha i} = \frac{1}{\tau_b}$</p> <p>dla tranzystora bezdryftowego</p> $\omega_{\alpha i} = 2,43 \frac{D_n}{W_B^2}$ <p>dla tranzystora dryftowego</p> $\omega_{\alpha i} = 2,43 \frac{D_n}{W_B^2} \left[1 + \left(\frac{\eta}{2} \right)^{4/3} \right]$ $\alpha(j\omega) \approx \frac{\alpha}{1 + \frac{j\omega}{\omega_\alpha}} e^{-j m \omega / \omega_\alpha}$ <p>przy czym $\omega_\alpha = \frac{1}{\tau_{eb} + \tau_b + \tau_{bc}}$</p> $\beta(j\omega) \approx \frac{\beta}{1 + \frac{j\omega}{\omega_\beta}}$ <p>przy czym $\omega_\beta = \frac{1}{\tau_r} \approx \frac{\omega_\alpha}{\beta}$</p> $\omega_T = \omega \beta(j\omega) \text{ przy } \omega > \omega_\beta$ $\omega_T = \omega_\beta \beta$ $\omega_T \approx \omega_\alpha$ $\omega_T = \frac{1}{\tau_N}$ $\tau_N = \frac{\varphi_T}{I_E} C_{Je} + \frac{W_B^2}{k D_n} + r_c (C_{Jc} + C_{dc}) + \frac{l_{d b-c}}{2v}$ $\omega_{max} = \sqrt{\frac{\omega_T}{4 r_{bb'} C_{Jc}}}$

Tabl. 5.6 (cd.)

Parametry modelu „hybryd π ”	$g_m = \frac{I_C}{\varphi_T}$ $g_{b'e} = \frac{g_m}{\beta}$ $g_{eb'} = (\beta + 1)g_{b'e}$ $r_{bb'} = h_{11e} - \frac{\beta}{g_m}$ $g_{b'c} = h_{22b} = h_{12e}g_{b'e}$ $g_{ce} = h_{22c} = \beta g_{b'c} = h_{12e}g_m$ $C_{b'c} = C_{22b}$ $C_{b'e} = \frac{g_m}{\omega_T} - C_{b'c}$
Moc admysyjna	$P_a = \frac{T_{jmax} - T_a}{R_{th}}$

W tranzystorach mocy dla zakresu małych częstotliwości uzyskuje się maksymalną moc admysyjną rzędu kilkuset watów, natomiast w mikrofalowych tranzystorach mocy (dla częstotliwości $f > 1$ GHz moc dopuszczalna maleje w przybliżeniu proporcjonalnie do $1/f^2$) uzyskuje się przykładowo 5 W dla $f = 5$ GHz.

ZADANIA

5.1. Wyznacz wartość współczynnika α dla tranzystora o znanej wartości współczynnika $\beta = 200$.

Wzór: (5.9)

Odpowiedź: 0,995

5.2. Wyznacz wartości współczynników α_N, β_N przy $I_E = 1$ mA, jeśli wiadomo, że przez tranzystor włączony w układzie WB płynął prąd $I_C = 10$ μ A przy $I_E = 0$ (jest to inaczej prąd zerowy I_{CBO}) oraz $I_C = 0,99$ mA przy $I_E = 1$ mA.

Wzory: (5.12), (5.13) oraz (5.6)

Odpowiedź: $\alpha_N = 0,98$; $\beta_N = 49$

5.3. Wyznacz wartość współczynnika sprawności wstrzykiwania emitera dla germanowego tranzystora z jednorodną bazą o następujących parametrach: $N_{AB} = 10^{22}$ m^{-3} ; $N_{DE} = 10^{25}$ m^{-3} ; $W_B = 1$ μ m; $W_E = 3$ μ m.

Wzór: (5.39)

Dane liczbowe: $D_n = 82 \cdot 10^{-4}$ m^2/s ; $D_p = 45 \cdot 10^{-4}$ m^2/s

Odpowiedź: 0,9999

5.4. Wyznacz wartość współczynnika sprawności wstrzykiwania emitera dla krzemowego tranzystora dryftowego, dla którego $\eta = 8$, $\tau = 10^{-7}$ s, a parametry konstrukcyjne, tj. N_{AB}, N_{DE}, W_B, W_E , są identyczne jak w poprzednim zadaniu. Obliczenia należy wykonać dla trzech wartości napięcia $U_{EB} = 50, 250$ oraz 500 mV.

Wzory: (5.38) oraz wzór na szerokość warstwy zaporowej skokowego złącza p-n z rozdziału 3

Dane liczbowe: $\varphi_T = 25$ mV (dla $T = 300$ K); $D_n = 34 \cdot 10^{-4}$ m^2/s ; $n_i = 1,5 \cdot 10^{16}$ m^{-3} , $D_p = 12,5 \cdot 10^{-4}$ m^2/s

Odpowiedź: 0,06 dla $U_{EB} = 50$ mV; 0,78 dla $U_{EB} = 250$ mV oraz 0,998 dla $U_{EB} = 500$ mV

5.5. Wyprowadź wzór (5.47).

5.6. Sprawdź, czy wzór (5.48) jest ogólny w stosunku do (5.46) i (5.44).

5.7. Wyznacz wartość współczynnika transportu nośników w bazie dla tranzystora dryftowego o następujących parametrach: $W_B = 1$ μ m; $\eta = 8$; $\tau_r = 10^{-8}$ s.

Wzory: (5.48) oraz $L = \sqrt{D_n \tau_r}$

Dane liczbowe: $D_n = 34 \cdot 10^{-4}$ m^2/s

Odpowiedź: $\alpha_b = 0,997$

- 5.8.** Napięcia zmierzone na końcówkach tranzystora pracującego w układzie wzmacniacza (w zakresie aktywnym normalnym) wynoszą: $U_C = 5$ V; $U_B = 2$ V; $U_E = 1,3$ V. Jakiego typu jest ten tranzystor: $n-p-n$ czy $p-n-p$?
- 5.9.** Udowodnij identyczność wzorów (5.52), (5.53) ze wzorami (5.54) i (5.55).
- 5.10.** Wykaż równoważność równań (5.52), (5.53) równaniom (5.57), (5.58).
- 5.11.** Wyprowadź wyrażenia analityczne na charakterystyki statyczne w układzie WB oraz WE według wskazówek zawartych w p. 5.7.2.
- 5.12.** Oblicz prąd zerowy I_{CEO} dla tranzystora o prądzie zerowym $I_{CBO} = 100$ nA oraz $\beta_N = 200$.
Wzór: (5.66a)
Odpowiedź: 20,1 μ A
- 5.13.** Oblicz napięcie przebicia BU_{CEO} w układzie WE dla tranzystora, którego napięcie przebicia w układzie WB wynosi $BU_{CBO} = 100$ V oraz współczynnik $\beta_N = 100$.
Wzór: (5.70)
Dane liczbowe: $n = 5$
Odpowiedź: 40 V
- 5.14.** Wyznacz częstotliwość graniczną $f_{\alpha I}$ ($f_{\alpha I} = \omega_{\alpha I}/2\pi$) dla tranzystorów: bezdryftowego germanowego z zadania 5.3 oraz dryftowego krzemowego z zadania 5.4.
Wzory: (5.158), (5.159)
Dane liczbowe: D_n , W_B , η jak w zadaniach 5.3, 5.4
Odpowiedź: 3 GHz dla tranzystora bezdryftowego oraz 9,5 GHz dla tranzystora dryftowego (Uwaga: Otrzymane wartości są nierealnie duże, gdyż częstotliwość $f_{\alpha I}$ dotyczy tranzystora idealnego)
- 5.15.** Wyznacz częstotliwości graniczne f_{β} oraz f_T dla tranzystora dryftowego, dla którego $f_{\alpha} = 1$ GHz, $\beta = 200$, $\eta = 8$.
Wzory: (5.188), (5.199)
Odpowiedź: $f_{\beta} = 2,5$ MHz; $f_T = 500$ MHz
- 5.16.** Oblicz transkonduktancję g_m tranzystora w temperaturze pokojowej dla punktu pracy $I_C = 5$ mA
Wzór: (5.209)
Odpowiedź: 200 mA/V
- 5.17.** Oblicz konduktancję $g_{b'e}$ tranzystora z poprzedniego zadania, jeśli wiadomo, że współczynnik $\beta = 200$.
Wzór: (5.211)
Odpowiedź: 10^{-3} S
- 5.18.** Oblicz rezystancję rozproszoną bazy $r_{bb'}$ dla tranzystora z zadania 5.16 przy znanym parametrze $h_{11e} = 1050$ Ω .
Wzór: (5.213)
Dane liczbowe: $g_m = 200$ mA/V; $\beta = 200$
Odpowiedź: 50 Ω
- 5.19.** Oblicz konduktancje $g_{b'e}$ oraz g_{ce} tranzystora z zadania 5.16 przy znanym parametrze $h_{12e} = 5 \cdot 10^{-4}$.
Wzory: (5.216), (5.220)
Dane liczbowe: $g_m = 200$ mA/V, $\beta = 200$
Odpowiedź: $g_{b'e} = 5 \cdot 10^{-7}$ S, $g_{ce} = 10^{-4}$ S
- 5.20.** Wyznacz moc admisyjną w temperaturze otoczenia 20°C dla tranzystora krzemowego, którego opór cieplny $R_{th} = 5^\circ\text{C/W}$, a $T_{Jmax} = 170^\circ\text{C}$.
Wzór: (5.229)
Odpowiedź: 30 W

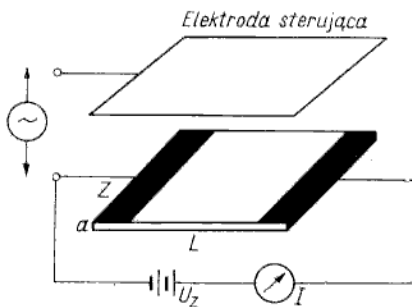
Wstęp

6.1

Z ogólnej klasyfikacji tranzystorów, przedstawionej w poprzednim rozdziale, wynika, że obok tranzystorów bipolarnych drugą nie mniej ważną klasę tych elementów stanowią *tranzystory unipolarne*, inaczej nazywane *polowymi*. Są to tranzystory, których działanie jest oparte na transporcie tylko jednego rodzaju nośników, tj. nośników większościowych (stąd nazwa tranzystory unipolarne), przy czym sterowanie prądu wyjściowego odbywa się za pomocą poprzecznego pola elektrycznego (stąd nazwa tranzystory polowe). W książce będzie stosowana ta ostatnia nazwa. W literaturze światowej tranzystory polowe mają skrótową nazwę FET (ang. *Field-Effect Transistor*), która również dobrze przyjęła się w naszym języku technicznym; nie należy jednak popierać żargonu specjalistycznego w rodzaju: „układ na fetach”.

W pewnym stopniu paradoksalny jest fakt, że ogólna idea działania tranzystora polowego była znana ok. 20 lat wcześniej niż nastąpiło odkrycie tranzystora bipolarnego, a produkcję tranzystorów polowych rozpoczęto niemal 10 lat później niż produkcję tranzystorów bipolarnych. Ponadto zasada działania tranzystora polowego jest znacznie prostsza niż tranzystora bipolarnego, o czym Czytelnik przekona się, porównując materiał zawarty w tym i poprzednim rozdziałach. Różnicę w skali złożoności działania obu rodzajów tranzystorów potwierdza również fakt, że tranzystor polowy wynaleziono, a bipolarny odkryto¹⁾. Istnieje jednak pewna prawidłowość w elektronice (a może i w innych dziedzinach), że realizacja techniczna prostych idei stawia bardzo wysokie wymagania odnośnie do precyzji wykonawstwa (precyzji procesów technologicznych). Tak jest również w przypadku tranzystorów polowych.

W poprzednim rozdziale próbowaliśmy „odkryć” tranzystor bipolarny, teraz zatem (ośmieleni sukcesem) będziemy próbować „wynaleźć” tranzystor polowy. Idea działania



Rys. 6.1
Ilustracja najprostszej koncepcji przyrządu czynnego sterowanego polem elektrycznym

¹⁾ Nagrodę Nobla przyznano za odkrycie tranzystora bipolarnego, a nie za wynalazek tranzystora polowego.

bodajże najprostszego elementu wzmacniającego na ciele stałym nasuwa się sama na podstawie znajomości prawa Ohma i podstawowych praw elektrostatyki. Weźmy płytkę ciała stałego przewodzącego prąd elektronowy (załóżmy chwilowo, że jest to metal lub półprzewodnik typu n), włączoną w obwód pokazany na rys. 6.1.

Przy stałym napięciu U_z można sterować wartość prądu w tym obwodzie przez zmiany rezystancji płytki

$$R = \frac{L}{q\mu nZa}$$

Teoretycznie rezystancję można zmieniać za pośrednictwem pięciu parametrów tj. L , Z , a , μ , n (tylko ładunek elektronu q jest stałą fizyczną). Przyjmuje się że L , Z , μ są wielkościami stałymi. Wówczas

$$R \sim \frac{1}{an}$$

Pozostaje odpowiedzieć na pytanie, jakim oddziaływaniem elektrycznym można zmieniać grubość płytki i koncentrację elektronów? Oczywiście grubości płytki w sensie metalurgicznym nie można zmieniać, lecz można spowodować taki stan, w którym gęstość prądu nie będzie jednakowa w całym przekroju płytki. W przypadku wyidealizowanym można wytworzyć taką sytuację, że płytka będzie się składała z warstw nieprzewodzącej i przewodzącej. Warstwę przewodzącą nazwiemy *kanalem* i będziemy zmieniać jego grubość. W tym celu należy umieścić nad płytką przewodzącą drugą płytkę metalową, spełniającą funkcję elektrody sterującej. Do tych dwu płytek tworzących kondensator doprowadzamy sygnał napięcia sterującego. Pole elektryczne, istniejące między elektrodą sterującą a płytką przewodzącą, powoduje zaindukowanie ładunku (przegrupowanie elektronów) w płycie przewodzącej. Niech na przykład potencjał elektrody sterującej będzie ujemny. Wówczas elektrony są wypychane od powierzchni w głąb płytki przewodzącej. Przy powierzchni powstaje warstwa pozbawiona elektronów, a więc nieprzewodząca. Dodatni ładunek tej warstwy (ładunek jonów nieskompensowanych przez elektrony) spełnia funkcję ekranu nie przepuszczającego linii sił pola elektrycznego w głąb płytki. Pod warstwą nieprzewodzącą znajduje się warstwa przewodząca (kanał), w której koncentracja elektronów nie zmienia się (pozornie może się wydawać, że elektrony spychane w głąb płytki spowodują wzrost koncentracji w kanale, jednak oznaczałoby to ujemne naładowanie kanału, co byłoby sprzeczne z założeniem braku poprzecznego pola elektrycznego w kanale). Im większy jest ujemny potencjał elektrody sterującej, tym głębiej wnika pole elektryczne do płytki, a więc wzrasta grubość warstwy nieprzewodzącej, a maleje grubość kanału, czyli wzrasta całkowita rezystancja płytki. W przypadku dodatniego potencjału na elektrodzie sterującej elektrony są przyciągane ku powierzchni płytki, czyli powstaje warstwa bardzo silnie przewodząca o dużej koncentracji elektronów. Ponieważ ta warstwa bocznikuje niejako pozostałą część płytki, można ją również nazwać kanałem. W tym przypadku wzrost dodatniego potencjału elektrody sterującej powoduje głównie wzrost koncentracji nośników w kanale (grubość kanału zmienia się nieznacznie), czyli zmniejszenie całkowitej rezystancji płytki.

Wartość prądu w obwodzie zasilanym napięciem U_z jest zatem sterowana wskutek oddziaływania elektrostatycznego polem prostopadłym do płytki przewodzącej.

Efektywność tego oddziaływania jest tym silniejsza, im głębiej wnika pole, a więc im mniejsza jest koncentracja elektronów (w stanie bez pola). Dlatego efekt sterowania konduktancji płytki poprzecznym polem elektrycznym jest znacznie silniejszy dla półprzewodnika niż dla metalu.

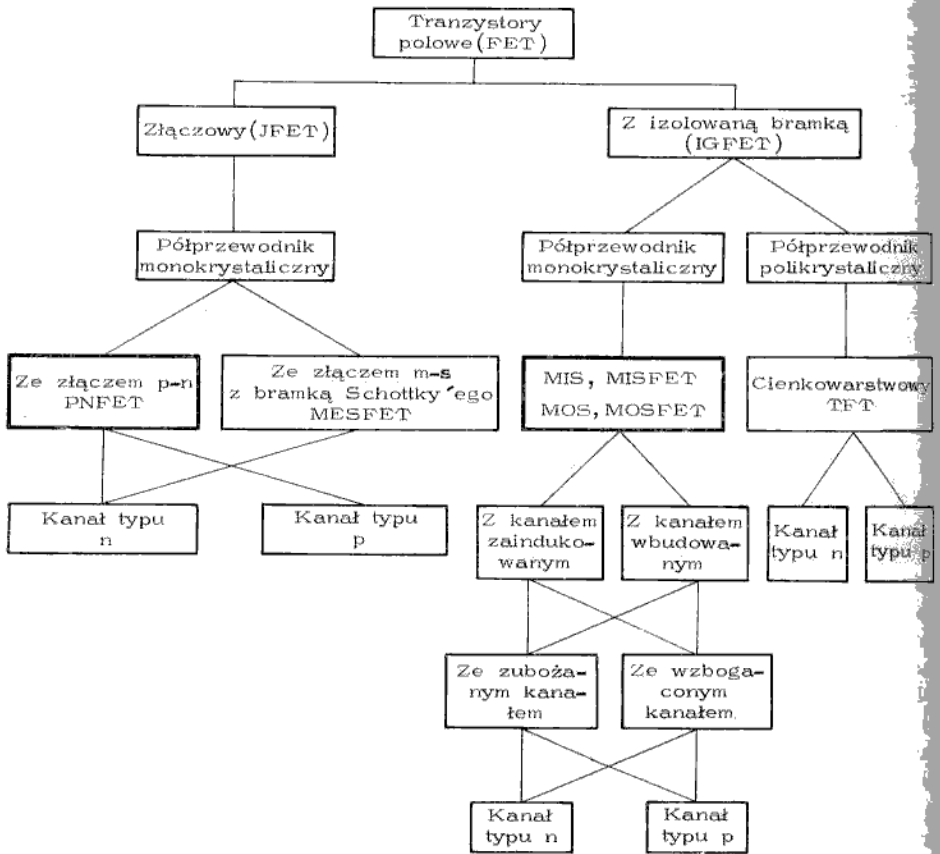
W ten sposób dokonaliśmy „wynalazku” czynnego przyrządu półprzewodnikowego, sterowanego poprzecznym polem elektrycznym. Nie możemy jednak ubiegać się o patent, gdyż przyrząd identyczny z wyżej opisanym stanowił przedmiot patentu uzyskanego przez Heila w 1935 r. (O. Heil — inżynier niemiecki, uzyskał ten patent w Wielkiej Brytanii). Jeszcze wcześniej — w latach 1925—1932 — J. E. Lilienfeld uzyskał kilka patentów w Kanadzie i USA na przyrządy półprzewodnikowe, działające w podobny sposób. Przyrządy półprzewodnikowe, wynalezione przez Lilienfelda i Heila, należy — według obecnych pojęć — uznać za pierwsze tranzystory polowe. Wynalazki te nie miały jednak szans rychłego zastosowania w elektronice, na czym zaważyły dwie podstawowe przyczyny:

- w latach trzydziestych nie myślano jeszcze o zamianie ciągle modyfikowanych i dalekich od szczytu swych możliwości lamp elektronowych;
- poziom wiedzy o właściwościach powierzchniowych półprzewodnika i poziom technologii materiałów półprzewodnikowych były wówczas niewystarczające do opanowania produkcji tranzystorów polowych.

Dopiero w drugiej połowie lat czterdziestych w laboratoriach Bell Telephone podjęto systematyczne badania tranzystora polowego. Jednakże efekt polowy (dla ścisłości sprecyzujmy, że efektem polowym nazywa się zmianę konduktywności ciała stałego wskutek oddziaływania nań polem elektrycznym) obserwowany doświadczalnie był znacznie słabszy niż przewidywany na podstawie teorii. Okazało się, że prawie cały ładunek zaindukowany przy powierzchni płytki półprzewodnikowej przez pole elektryczne, przostopadłe do tej powierzchni, pozostaje unieruchomiony w pułapkach występujących na powierzchni półprzewodnika, czyli nie wpływa na wartość sterowanego prądu. Badania właściwości tych pułapek (ogólniej stanów powierzchniowych) doprowadziły w 1947 r. do przypadkowego odkrycia tranzystora ostrzowego, działającego na zupełnie innej zasadzie niż tranzystor polowy. Dało to początek dużemu zainteresowaniu tym nowo odkrytym przyrządem, z którego wyrosła rodzina tranzystorów bipolarnych. Zmalało zarazem zainteresowanie tranzystorem polowym.

W 1952 r. Shockley przedstawił nową koncepcję tranzystora polowego, w którym funkcję dielektryka rozdzielającego elektrodę sterującą od kanału przewodzącego spełnia warstwa zaporowa złącza $p-n$ spolaryzowanego zaporowo¹⁾. Takie rozwiązanie umożliwia odsunięcie kanału od powierzchni, czyli usunięcie trudności związanych ze stanami powierzchniowymi. Tranzystor polowy ze złączem $p-n$ wykonano laboratoryjnie i przebadano już w 1953 r., jednak dopiero opanowanie technologii planarnej (ok. 1960 r.) stworzyło niezbędne przesłanki techniczne dla produkcji tych elementów na wielką skalę. Nieco później, na początku lat sześćdziesiątych, wykonano laboratoryjnie pierwsze tranzystory polowe z warstwą dielektryka między elektrodą sterującą a płytką półprzewodnikową. W drugiej połowie lat sześćdziesiątych opracowano również tranzystor polowy, w którym elektroda sterująca jest oddzielona od kanału warstwą zubożoną złącza metal-półprzewodnik. A zatem obecnie produkuje się wiele różnych typów tranzystorów polowych. Ich ogólną klasyfikację przedstawiono na rys. 6.2. Wszystkie tranzystory polowe dzieli się na dwie grupy:

¹⁾ W. Shockley: A unipolar field-effect transistor. *Proc. IEEE*, 1952, Vol. 40, p. 1365.



Rys. 6.2
Klasyfikacja tranzystorów polowych

- złączowe (JFET — ang. *Junction Field-Effect Transistor*),
- z izolowaną bramką (IGFET — ang. *Insulated Gate Field-Effect Transistor*).

Tranzystory polowe złączowe mogą być ze złączem $p-n$ (PNFET) lub ze złączem metal-półprzewodnik (MESFET — ang. *Metal-Semiconductor FET*). Warto zwrócić uwagę, że niektóre z wymienionych nazw skrótowych na mocy tradycji utrwały się w nieco innym znaczeniu. Pierwszymi dostępnymi na rynku tranzystorami polowymi były tranzystory ze złączem $p-n$. Z uwagi na brak innych rodzajów tranzystorów polowych nazwa FET oznaczała wówczas to samo co obecnie nazwa PNFET. Wraz z ukazaniem się tranzystorów polowych z izolowaną bramką wynikała konieczność wprowadzenia nazwy JFET oznaczającej tranzystory polowe ze złączem $p-n$ (inne tranzystory złączowe nie były wówczas znane). Ale i ta nazwa okazała się zbyt ogólna, gdy pojawiły się tranzystory ze złączem metal-półprzewodnik. Zrozumiała bezwładność w ewolucji terminologii powoduje, że również obecnie nazwy JFET lub FET są często używane do oznaczania tranzystorów polowych ze złączem $p-n$, które jednak konsekwentnie będziemy nazywać tranzystorami PNFET. Tranzystory PNFET są wytwarzane prawie wyłącznie z krzemu i mogą mieć kanał typu n lub p .

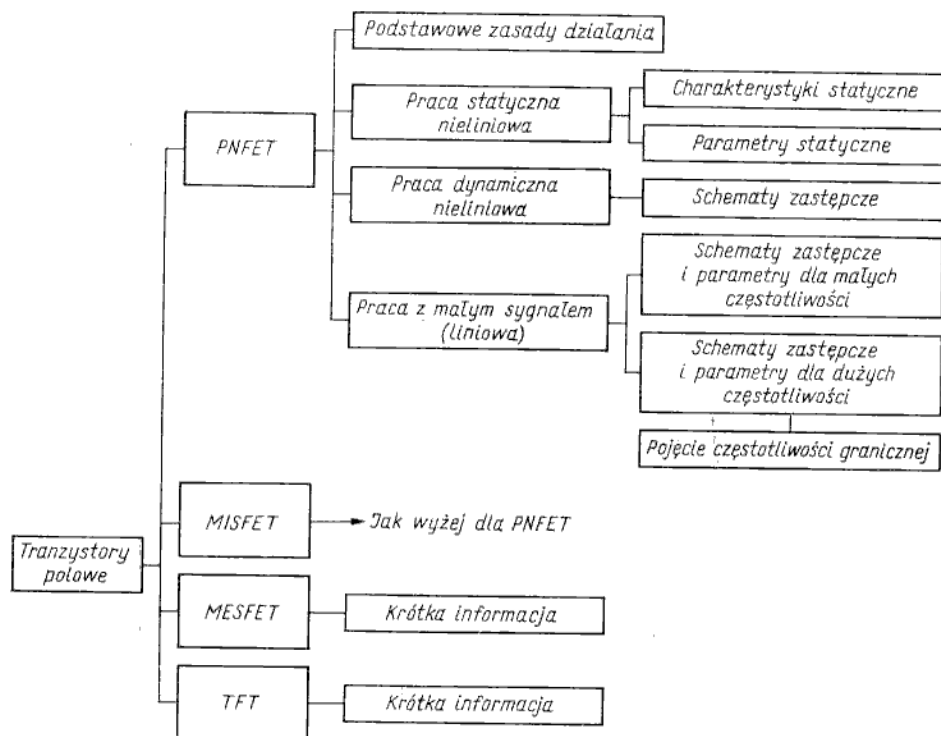
Tranzystory MESFET mają znaczenie wyłącznie jako elementy mikrofalowe, dlatego są wytwarzane z GaAs typu n , tj. materiału o bardzo dużej ruchliwości nośników.

Wszystkie tranzystory JFET są wytwarzane z półprzewodników monokrystalicznych.

Tranzystory z izolowaną bramką mogą być wykonane z półprzewodnika monokrystalicznego (praktycznie biorąc w skali technicznej tylko z krzemu) lub polikrystalicznego (najczęściej są to związki typu $A_{III}B_{VI}$, jak np. CdSe, CdS itp.).

W pierwszym przypadku podstawową strukturę tranzystora tworzą warstwy metal-izolator-półprzewodnik, stąd nazwa *tranzystor MIS* (ang. *Metal-Insulator-Semiconductor*) lub MISFET, przy czym najczęściej funkcję izolatora spełnia warstwa SiO_2 (ang. *Oxide*), stąd nazwa *tranzystor MOS* lub MOSFET. W drugim przypadku przyjęła się nazwa *tranzystor cienkowarstwowy* (skrótowa nazwa TFT, ang. *Thin Film Transistor*) z uwagi na spójność technologii wytwarzania tych tranzystorów z technologią wytwarzania układów scalonych cienkowarstwowych (podstawowa struktura tego tranzystora jest również typu MIS). Oba rodzaje tranzystorów (MISFET i TFT) mogą mieć kanał typu n lub p . Bardziej szczegółowy podział tranzystorów MIS będzie skomentowany w dalszej treści rozdziału.

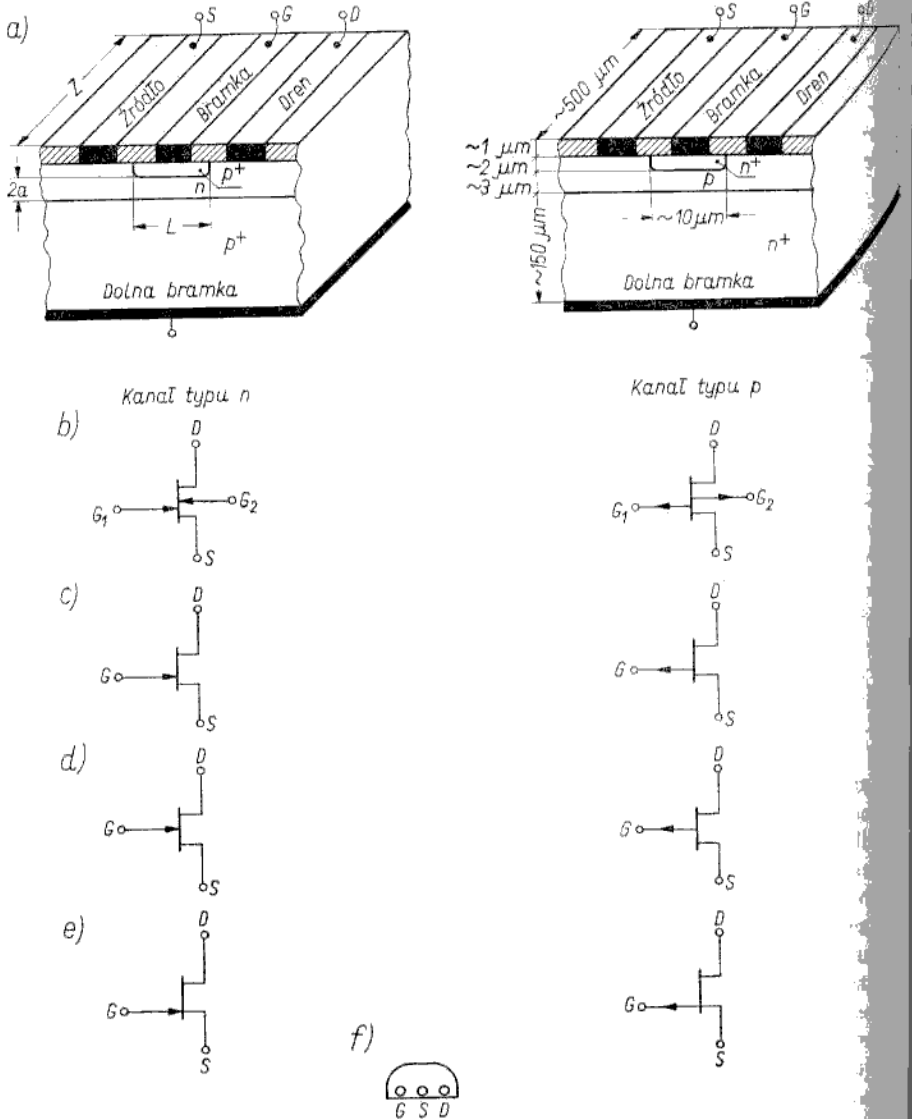
Na rysunku 6.2 liniami grubymi zaznaczono najważniejsze rodzaje tranzystorów, tj. PNFET i MISFET. Tym tranzystorom poświęcimy najwięcej uwagi. Układ rozdziału przedstawiono na rys. 6.3.



Rys. 6.3
Schematyczny układ rozdziału

Tranzystory polowe ze złączem p-n (PNFET)

Podstawowe struktury i symbole graficzne tranzystorów PNFET przedstawione na rys. 6.4. Dla ustalenia uwagi będziemy dalej rozpatrywać tranzystor z kanałem typu *n*, przy czym wszystkie rozważania odnoszą się również do tranzystora z kanałem typu *p* po uwzględnieniu odpowiednich zmian formalnych (jak np. zmiana rodzaju nośników prądu i zmiana polaryzacji).



Rys. 6.4 Podstawowe struktury (a), symbole graficzne (b, c, d, e) i typowy kształt obudowy (f) tranzystorów PNFET

Sposób wykonania tego tranzystora jest następujący. Na podłożu krzemowym typu p^+ (dolna bramka) nanosi się warstwę epitaksjalną typu n (kanał), w której następnie wytwarza się metodą dyfuzji lokalnej warstwę p^+ (górną bramkę). Końcową operacją jest wykonanie kontaktów metalowych do warstw p^+ (elektrody górnej i dolnej bramki) oraz do kanału typu n (elektrody źródła i drenu). Struktury tranzystorów są przedstawione na rys. 6.4a bez zachowania skali, gdyż chodzi tu o przejrzyste pokazanie poszczególnych warstw. Dla lepszej orientacji w rzeczywistych proporcjach na jednym z rysunków podano przybliżone rozmiary. Najistotniejsze są rozmiary kanału, które oznaczono następująco: L — długość (wyznaczona przez krawędzie warstwy dyfuzyjnej górnej bramki), $2a$ — grubość, Z — szerokość. Przedstawiona struktura jest całkowicie symetryczna, a więc różnica między źródłem a drenem jest tylko umowna. O ile w tranzystorach bipolarnych struktura symetryczna, tj. taka, w której funkcje emitera i kolektora byłyby wzajemnie zamienialne, jest rzadko spotykana, o tyle tranzystory PNFET bardzo często mają budowę symetryczną. Zatem nazwy źródło, dren wynikają ze sposobu polaryzacji tych elektrod. Prąd w kanale jest strumieniem nośników większościowych, dostarczanych przez jedną elektrodę, którą nazywa się źródłem, i odbieranych przez drugą elektrodę, którą nazywa się drenem. W przypadku tranzystora z kanałem typu n (prąd kanału jest strumieniem elektronów) drenem jest ta spośród dwu rozpatrywanych elektrod, która ma wyższy potencjał dodatni. Natomiast w przypadku tranzystora z kanałem typu p (prąd kanału jest strumieniem dziur) drenem jest ta spośród dwu rozpatrywanych elektrod, która ma wyższy potencjał ujemny.

Szczególne końcówki tranzystora są oznaczone następującymi literami:

S — źródło (ang. *Source*),

G — bramka (ang. *Gate*),

D — dren (ang. *Drain*).

Potencjały stałe elektrod względem pewnego wspólnego punktu odniesienia (masy) będziemy oznaczać następująco:

U_S — potencjał źródła,

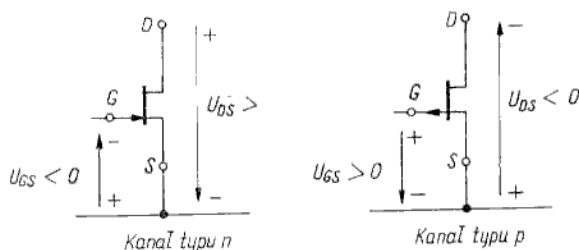
U_G — potencjał bramki,

U_D — potencjał drenu.

Napięcie potencjałów między dwiema elektrodami oznaczmy np. U_{GS} (napięcie bramka-źródło), U_{DS} (napięcie dren-źródło) itp.

Symbol graficzny tranzystorów, w których są wyprowadzone końcówki obu bramek (górną i dolną) przedstawiono na rys. 6.4b. Obowiązuje ogólna zasada strzałkowania od obszaru P do N (zgodnie z konwencjonalnym kierunkiem przepływu prądu, czyli zgodnie z kierunkiem strumienia dziur). Najczęściej obie bramki są zwarte wewnątrz obudowy i tranzystor PNFET jest elementem trójkońcówkowym, którego symbole graficzne (trzy warianty) przedstawiono na rys. 6.4c, d, e. Normy IEC zalecają symbole przedstawione na rys. 6.4c, jednak dwa pozostałe warianty symboli są często stosowane ze względu na łatwość ich kreślenia (w tej książce będzie stosowany wariant e). Typowy kształt obudowy tranzystorów PNFET (widzianej od strony wyprowadzeń) i układ wyprowadzeń elektrod przedstawiono na rys. 6.4f.

Oprócz symetrii tranzystora PNFET uwzględnimy również, że pracuje on tylko przy polaryzacji zaporowej złącza bramka-kanał, to uzyskamy tylko jeden sposób polaryzacji tego tranzystora (rys. 6.5) (o ileż to prościej niż w przypadku tran-



Rys. 6.5
Sposób polaryzacji tranzystora PNFET

zystora bipolarnego, w którym były cztery warianty polaryzacji i wynikające stąd zakresy: odcięcia, nasycenia, normalny i inwersyjny).

W przypadku kanału typu n :

$$U_{GS} < 0$$

$$U_{DS} > 0$$

W przypadku kanału typu p :

$$U_{GS} > 0$$

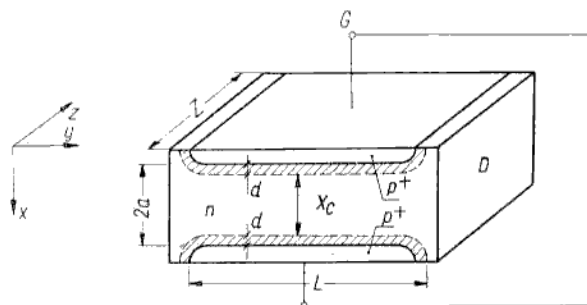
$$U_{DS} < 0$$

Rozpatrzmy teraz jakościowo elementarne zasady działania tranzystora PNFET, po czym dokonamy analizy ilościowej podstawowych charakterystyk, parametrów i modeli tych tranzystorów.

Zasady działania tranzystora PNFET — opis jakościowy

6.2.1

Ponieważ najistotniejsze zjawiska zachodzą w obszarze kanału znajdującym się pod bramką górną, dla uproszczenia będziemy rozpatrywać wyidealizowaną



Rys. 6.6
Wyidealizowana struktura tranzystora PNFET

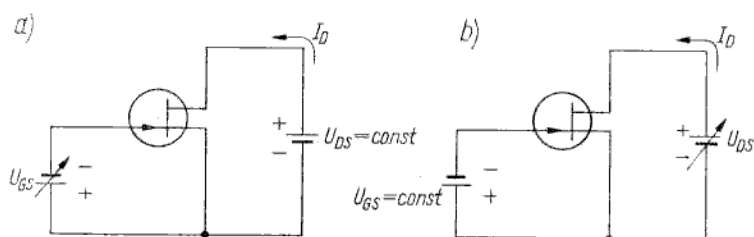
strukturę przedstawioną na rys. 6.6 (stanowi ona jakby wycinek rzeczywistej struktury planarnej). Interesujące są dwie charakterystyki tranzystora, tj.:

— przejściowa $I_D(U_{GS})|_{U_{DS} = \text{const}}$

— wyjściowa $I_D(U_{DS})|_{U_{GS} = \text{const}}$

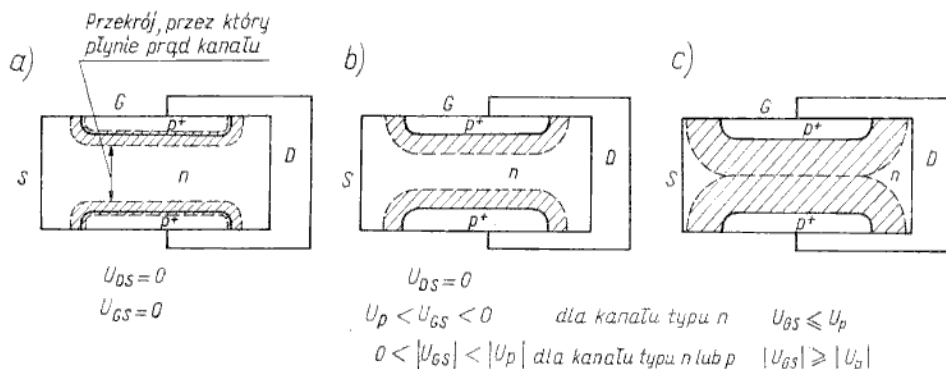
wyznaczane w układach przedstawionych na rys. 6.7.

Zjawiska określające przebieg charakterystyki przejściowej, tj. zależność prądu drenu (prądu wyjściowego) od napięcia bramka-źródło (napięcia wejściowego) ilustrują



Rys. 6.7

Układy włączenia tranzystora przy pomiarze charakterystyki przejściowej (a) oraz wyjściowej (b)



Rys. 6.8

Ilustracje profili kanału, wyjaśniające przebieg charakterystyki przejściowej

przekroje tranzystora przedstawione na rys. 6.8. Dla uproszczenia jest rozważany przypadek bardzo małego napięcia dren-źródło ($U_{DS} \geq 0$). W stanie bez polaryzacji bramki ($U_{GS} = 0$) warstwa zaporowa złącza bramka-kanał wnika na niewielką głębokość do obszarów typu p^+ i n . Ponieważ obszar bramki (p^+) jest o kilka rzędów wartości silniej domieszkowany niż obszar kanału (n), warstwa zaporowa wnika o wiele głębiej do kanału niż do bramki. Dlatego w dalszych rysunkach w ogóle pomija się warstwę zaporową w obszarze bramki. Z rozdziału 3 wiadomo, że warstwa zaporowa jest obszarem o zubożonej koncentracji nośników, a więc o bardzo dużej rezystywności. Stąd wnioskujemy, że prąd kanału płynie tylko w obszarze ograniczonym przez krawędzie warstw zaporowych (przekrój zaznaczony na rys. 6.8a). W miarę jak wzrasta ujemna wartość napięcia U_{GS} , warstwy zaporowe rozszerzają się, powodując zmniejszenie przekroju kanału (rys. 6.8b). Wzrasta zatem rezystancja kanału, czyli przy $U_{DS} = \text{const}$ maleje prąd drenu. W miarę dalszego zwiększania napięcia U_{GS} dochodzi do zetknięcia się obu warstw zaporowych, czyli kanał przewodzący przestaje istnieć i prąd drenu osiąga wartość zerową (rys. 6.8c). Taki stan jest nazywany *odcięciem* lub *zatkaniem*, a wartość napięcia U_{GS} , przy której zostaje on zapoczątkowany, jest nazywana *napięciem odcięcia* lub *zatkania* (ang. *pinch-off voltage*) i oznaczana przez U_p . Dalszy wzrost ujemnej wartości napięcia wejściowego ($|U_{GS}| > |U_p|$) nie zmienia prądu drenu, a może tylko doprowadzić do przebicia złącza bramka-kanał. Rezystancja dren-źródło w stanie zatkania jest bardzo duża i wynosi kilka gigaomów dla tranzystorów małej mocy (jest to

wartość porównywalna z rezystancją złącza $p-n$ spolaryzowanego zaporowej. Dotychczas rozpatrywaliśmy tylko polaryzację zaporową złącza bramka-kanal. Teoretycznie tranzystor może pracować również przy polaryzacji w kierunku przewodzenia, jeżeli napięcie U_{GS} nie jest większe niż napięcie dyfuzyjne. Jednak takie warunki polaryzacji nie mają sensu praktycznego¹⁾, gdyż efekt modulacji przekroju kanału jest bardzo słaby (zwężenie warstwy zaporowej znacznie cieńszej od kanału), a przy tym w obwodzie bramka-źródło płynie duży prąd przewodzenia, czyli tranzystor traci swą największą zaletę — dużą rezystancję wejściową (rezystancja wejściowa przy polaryzacji zaporowej złącza bramka-kanal jest rzędu gigaomów w tranzystorach małej mocy). Choćby rozważanie te mają charakter tylko jakościowy, to jednak bez trudu można zaproponować prosty zapis analityczny charakterystyki przejściowej dla małych wartości napięcia U_{DS} . Przy $U_{DS} = \text{const}$ prąd drenu jest odwrotnie proporcjonalny do rezystancji kanału

$$I_D \sim \frac{1}{R_C}$$

Korzystając z oznaczeń wprowadzonych na rys. 6.6:

$$R_C \sim \frac{1}{X_C}$$

$$X_C = 2(a-d)$$

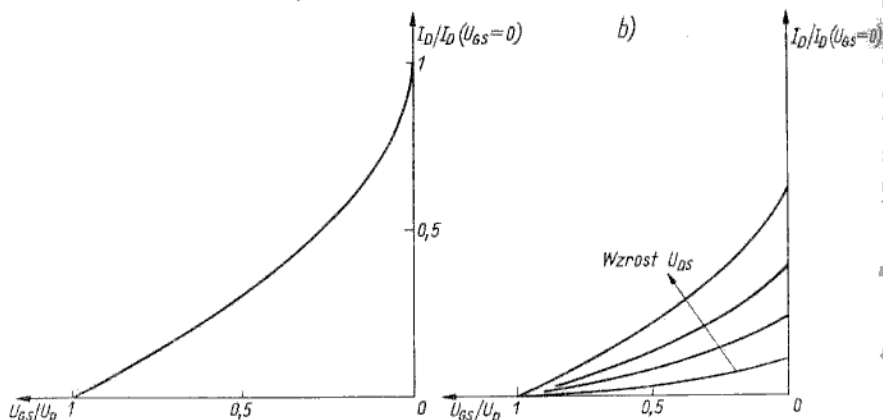
Zatem

$$I_D \sim 2(a-d)$$

Jeżeli przyjmiemy model złącza skokowego (przy założeniach upraszczających: $d=0$ dla $U_{GS} = 0$ oraz napięcie dyfuzyjne złącza bramka-kanal $\varphi_B = 0$), to:

$$d \sim (U_{GS})^{1/2}$$

$$d = a \quad \text{dla} \quad U_{GS} = U_p$$



Rys. 6.9

Charakterystyki przejściowe $I_D(U_{GS})$: a) przy $U_{DS} \approx 0$; b) przy $U_{DS} \neq 0$

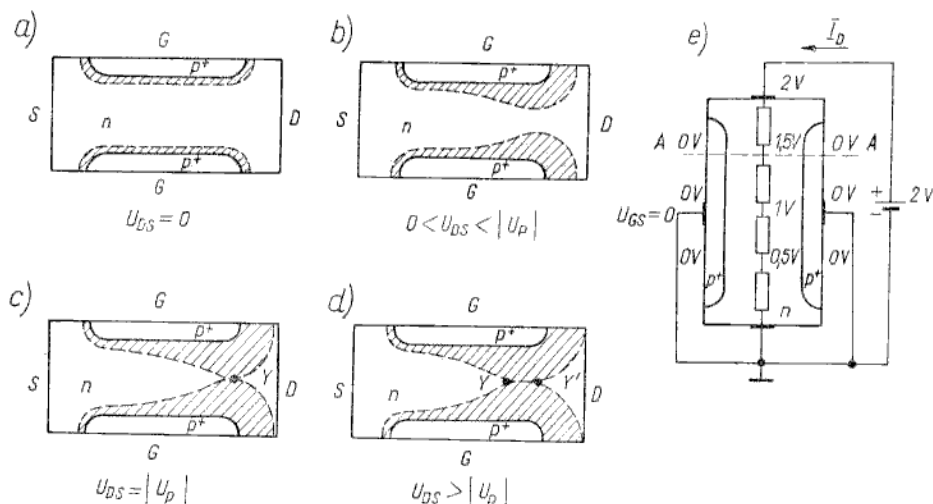
¹⁾ Taki sposób polaryzacji jest stosowany w tranzystorach jednozłączowych (rozdz. 11), które mają budowę podobną do tranzystorów PNFET, lecz działają na zupełnie innej zasadzie.

a więc

$$\frac{I_D(U_{GS})}{I_D(U_{GS} = 0)} \approx \frac{2(a-d)}{2a} = 1 - \left(\frac{U_{GS}}{U_p}\right)^{1/2} \quad (6.1)$$

Na rysunku 6.9a wykreślono charakterystykę przejściową, obliczoną według tej zależności. Należy podkreślić, że jest to charakterystyka dla bardzo małych wartości napięcia U_{DS} , przy których tranzystor można traktować po prostu jak rezystor liniowy sterowany napięciem (sterowanie jest nieliniowe). Na rysunku 6.9b przedstawiono wykresy charakterystyki $I_D(U_{GS})$ dla różnych wartości U_{DS} .

Interpretujemy teraz przebieg charakterystyki wyjściowej. Korzystając z rys. 6.10 rozpatrzmy najpierw najprostszyp przypadk dla $U_{GS} = 0$. Przy małych



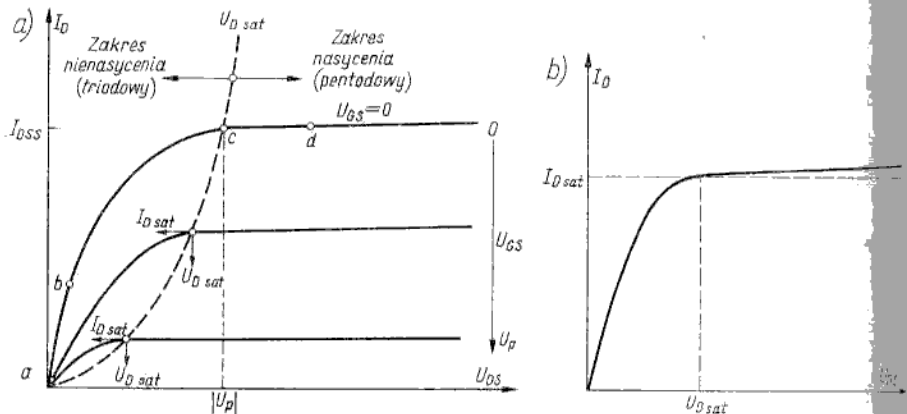
Rys. 6.10

Ilustracje profili kanału, wyjaśniające przebieg charakterystyki wyjściowej dla $U_{GS} = 0$

wartościach napięcia U_{DS} — jak stwierdzono wyżej — tranzystor zachowuje się jak rezystor liniowy, czyli przyrosty prądu drenu są wprost proporcjonalne do przyrostu napięcia U_{DS} . Jednak w miarę zwiększania napięcia U_{DS} należy uwzględnić rozkład potencjału wzdłuż kanału, utworzony wskutek spadku napięcia na rezystancji kanału od przepływającego prądu kanału (prąd kanału jest równy prądowi drenu). Obrazowo ilustruje tę sytuację rys. 6.10e dla konkretnego przypadku $U_{DS} = 2 \text{ V}$. Bramka stanowi powierzchnię ekwipotencjalną o potencjale zerowym. Kanał jest reprezentowany przez łańcuch rezystorów, na których powstają spadki napięć od przepływającego prądu kanału. Wskutek różnicy napięć między bramką a kanałem złącze bramka-kanał jest polaryzowane w kierunku zaporowym, przy czym im bliżej drenu, tym ta polaryzacja jest silniejsza. Na przykład dla przekroju $A-A$ złącze bramka-kanał jest polaryzowane zaporowo napięciem 1,5 V. Odpowiednio do tej polaryzacji rozszerza się warstwa zaporowa złącza bramka-kanał, czyli maleje efektywny przekrój kanału (X_C). Taki stan ilustruje rys. 6.10b, na którym widać, że grubość warstwy zaporowej jest większa w pobliżu drenu. Prąd kanału płynie w obszarze typu

n o kształcie klina, którego profil wyznaczają krawędzie warstw zaporowych. Rezystancja kanału jest w tym przypadku większa niż dla $U_{DS} = 0$. Dalszy wzrost napięcia U_{DS} powoduje dalsze „zaciskanie” kanału, zwiększa się zatem rezystancja kanału i przyrosty prądu drenu nie mogą być wprost proporcjonalne do przyrostów napięcia drenu. Rezystancja dren-źródło R_{DS} jest więc nieliniowa, a przyrosty prądu drenu, odpowiadające kolejnym jednakowym przyrostom napięcia drenu, są coraz mniejsze. Należy również zauważyć, że rozkład potencjału wzdłuż kanału nie może być równomierny, jak to przyjęto na rys. 6.10e. Gdyby podzielić kanał na sekcje o jednakowych długościach, wówczas sekcje położone bliżej drenu miałyby większą rezystancję z uwagi na mniejszy przekrój kanału. Wskutek tego spadek napięcia odkłada się głównie na odcinku znajdującym się w pobliżu drenu, a przekrój kanału przypomina swym kształtem przekrój pucharu spoczywającego dnem na krawędzi, z której jest wyprowadzona końcówka drenu. Jeżeli wreszcie napięcie U_{DS} osiągnie wartość $U_{DS} = |U_p|$, to obie warstwy zaporowe zetkną się w punkcie położonym w pobliżu drenu (punkt Y na rys. 6.10c) i nazywanym *punktem odcięcia* (napięcie U_p jest wcześniej już zdefiniowanym napięciem odcięcia). Dalszy wzrost napięcia U_{DS} nie powoduje prawie żadnych zmian prądu drenu, gdyż cały przyrost napięcia U_{DS} ponad wartość $|U_p|$ odkłada się na rozszerzającym się obszarze zubożonym (odcinek Y, Y' na rys. 6.10d). Mówi się w tym przypadku o nasyceniu prądu drenu.

Całą charakterystykę prądowo-napięciową $I_D(U_{DS})$ dla $U_{GS} = 0$ przedstawiono na rys. 6.11a, na którym zaznaczono charakterystyczne punkty: napięcie odcięcia



Rys. 6.11

Charakterystyki wyjściowe $I_D(U_{DS}/U_{GS})$: a) dla modelu uproszczonego; b) z uwzględnieniem nasylenia w zakresie nasycenia

U_p , prąd nasycenia I_{DSS} oraz punkty a, b, c, d , odpowiadające profilom przedstawionym na rys. 6.10a, b, c, d.

We wstępie do tego rozdziału sugerowaliśmy, że tranzystor polowy jest elementem mało skomplikowanym (w sensie złożoności zjawisk fizycznych). Jeżeli nawet dotychczas prezentowany opis działania tranzystora PN-FET potwierdza tę tezę, to na pewno nie dotyczy ona zjawisk w stanie odcięcia. Interpretacja zjawisk w zakresie $U_{DS} \geq |U_p|$ nie jest prosta i kilka wyżej podanych faktów wymaga dodatkowego komentarza. Niejasne mogą być przede wszystkim dwie sprawy:

- Dlaczego w opisie charakterystyki wyjściowej pojawiła się wartość napięcia odcięcia U_p , która była wprowadzona przy opisie charakterystyki przejściowej?
- Skoro warstwy zaporowe zetknęły się przy drenie, to kanał został przerwany, dlaczego więc prąd drenu jest stały, a nie maleje do zera?

Odpowiedź na pierwsze pytanie jest następująca. Omawiając charakterystykę przejściową wprowadzono pojęcie napięcia odcięcia jako takiej wartości napięcia U_{GS} , przy której następuje zetknięcie krawędzi warstw zaporowych, czyli grubość każdej warstwy zaporowej staje się równa połowie metalurgicznej grubości kanału ($d = a$). O ile w przypadku omawiania charakterystyki przejściowej zmieniano potencjał bramki względem ustalonego potencjału źródła (kanału), o tyle w przypadku charakterystyk wyjściowych występują zmiany potencjału kanału względem ustalonego potencjału bramki. Dla pracy złącza p^+n istotna jest tylko różnica napięć między obu warstwami. Warunek $d = a$ zaistnieje wówczas, gdy ta różnica napięć osiągnie wartość U_p , co stanie się najpierw przy drenie, jeżeli potencjał U_Y w punkcie Y (rys. 6.10c) mierzony względem potencjału bramki (dla $U_{GS} = 0$ jest to potencjał zerowy) osiągnie wartość $|U_p|$. Warunek

$$U_Y = |U_p|$$

można utożsamiać z warunkiem

$$U_{DS} = |U_p|$$

tylko w tym przypadku, jeżeli pominie się spadek napięcia na odcinku od punktu Y do elektrody drenu. Rezystancję tego odcinka nazywa się *rezystancją szeregową drenu* i oznacza przez r_D . Przez U_{Dsat} oznaczymy napięcie drenu, przy którym rozpoczyna się stan odcięcia, czyli prąd drenu wchodzi w zakres nasycenia (indeks *sat* od ang. *saturation* — nasycenie). Zatem dla

$$U_{GS} = 0$$

$$U_{Dsat} = |U_p| \quad \text{przy } r_D = 0 \quad (6.2)$$

lub

$$U_{Dsat} \geq |U_p| \quad \text{przy } r_D \geq 0 \quad (6.3)$$

Odpowiedź na drugie pytanie jest trudniejsza. Istotnie, konsekwencją założenia o braku przepływu prądu w obszarze warstwy zaporowej jest przerwanie drogi prądu w punkcie Y przy połączeniu się obu warstw zaporowych. Oznaczałoby to jednak, że $I_D = 0$, a więc znikłaby przyczyna rozszerzania się warstw zaporowych (brak prądu oznacza brak spadku napięcia w kanale, czyli brak polaryzacji złącza bramka-kanał). Jest to logiczne zaprzeczenie nie tylko możliwości spadku wartości prądu drenu do zera, lecz w ogóle jakiegokolwiek zmniejszenia się prądu drenu. Bo przecież jeśli założyć się, że wzrost napięcia U_{DS} powoduje zmniejszenie prądu I_D , to konsekwentnie:

maleje spadek napięcia w kanale

↓

maleje szerokość warstwy zaporowej

↓

wzrasta przekrój kanału

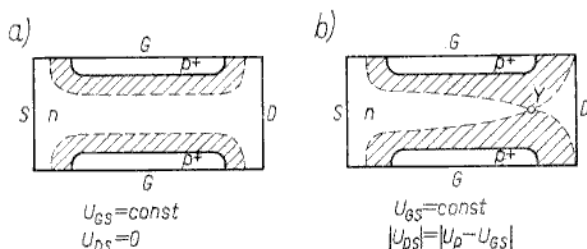
↓

wzrasta prąd drenu

co przeczy postawionej tezie.

Prąd drenu zatem może tylko się zwiększać lub nie zmieniać się. Żeby uzasadnić możliwość przepływu prądu mimo odcięcia kanału, trzeba przyjąć, że istnieje transport nośników również w obszarze zubożonym. W istocie na odcinku Y, Y' (rys. 6.10d) istnieje składowa wzdłużna pola elektrycznego, unosząca elektrony w stronę drenu, podobnie jak dla elektronów wymiatanych z bazy przez pole elektryczne warstwy zaporowej baza-kolektor w tranzystorze bipolarnym. Można również przyjąć, że do całkowitego zetknięcia warstw zaporowych nie dojdzie, gdyż w miarę zężenia się kanału natężenie pola elektrycznego w niewielkim odcinku przylegającym do drenu uzyskuje tak duże wartości, że następuje nasycenie szybkości unoszenia elektronów (szybkość nie zwiększa się przy dalszym wzroście natężenia pola). Powoduje to oczywiście nasycenie prądu, a dalszy przyrost napięcia drenu odkłada się na odcinku kanału o nasyczonej szybkości unoszenia elektronów, przy czym długość tego odcinka zwiększa się. Niezależnie od tego, którą z tych dwu interpretacji przyjmie się, ważne jest, że kanał składa się jakby z dwu obszarów, tj. obszaru od źródła do punktu Y , na którym odkłada się napięcie równe $|U_p|$, oraz obszaru ograniczonego punktami Y, Y' (może to być obszar ładunku przestrzennego albo bardzo cienki kanał, w którym szybkość unoszenia nośników jest nasyciona), na którym odkłada się różnica napięć $(U_{DS} - |U_p|)$. Pomija się w pierwszym przybliżeniu spadek napięcia na odcinku od Y' do elektrody drenu ($r_D = 0$).

Na rysunku 6.11a oprócz charakterystyki $I_D(U_{DS})$ dla $U_{GS} = 0$ przedstawiono również dwie charakterystyki dla $U_{GS} \neq 0$. Wpływ napięcia $U_{GS} \neq 0$ na przebieg charakterystyki wyjściowej można łatwo wyjaśnić na podstawie przekrojów tranzystora, pokazanych na rys. 6.12. W stanie $U_{GS} \neq 0, U_{DS} = 0$ kanał jest



Rys. 6.12

Ilustracje profili kanału, wyjaśniające przebieg charakterystyki wyjściowej dla $U_{GS} \neq 0$

zweżony równomiernie na całej swej długości (rys. 6.12a). W miarę zwiększania napięcia U_{DS} kanał zęża się przy drenie, aż dochodzi do odcięcia przy wartościach napięcia U_{DS} i prądu I_D mniejszych, niż to było dla $U_{GS} = 0$. Zatem porównanie charakterystyk wyjściowych dla $U_{GS} = 0$ i $U_{GS} \neq 0$ jest takie, jak porównanie charakterystyk dwu tranzystorów różniących się tylko grubością kanału. Odcięcie w tranzystorze o cieńszym kanale nastąpi oczywiście przy mniejszych wartościach napięcia U_{DS} i prądu I_D . Ponieważ warunkiem odcięcia jest istnienie różnicy potencjałów między elektrodą bramki a punktem Y (rys. 6.12b) równej U_p , przyjmując, że $U_Y = U_{DS}$ (czyli $r_D = 0$) otrzymuje się

$$U_{GS} - U_{DS} = U_p$$

czyli

$$U_{D \text{ sat}} = U_{GS} - U_p \quad \text{przy } r_D = 0$$

(6.4)

Jeżeli uwzględnimy spadek napięcia na odcinku od punktu odcięcia do elektrody drenu, to

$$U_{Dsat} \geq U_{GS} - U_p \quad \text{przy } r_D \geq 0 \quad (6.5)$$

Wróćmy jeszcze do rys. 6.10d, żeby wyjaśnić rzeczywisty przebieg prądu drenu w stanie po odcięciu kanału (rys. 6.11b). Gdyby odcinek Y, Y' był nieskończenie krótki, wówczas z uwagi na niezależność potencjału $U_Y = U_{Dsat}$ od napięcia U_{DS} (przypomina się, że cały przyrost napięcia U_{DS} ponad wartość U_{Dsat} odkłada się na odcinku Y, Y') oraz brak zmian kształtu kanału w obszarze od źródła do punktu Y , prąd drenu nie zmieniałby się w funkcji napięcia U_{DS} . Ponieważ jednak odcinek Y, Y' nieznacznie wydłuża się w miarę zwiększania napięcia U_{DS} , kanał (od źródła do punktu Y) skraca się, czyli jego rezystancja maleje, co przy stałości potencjału $U_Y = U_{Dsat}$ oznacza zwiększanie się prądu drenu. To zjawisko, nazywane *skracaniem kanału*, wyjaśnia różną od zera nachylenie charakterystyki $I_D(U_{DS})$ w stanie po odcięciu kanału (rys. 6.11b).

Pole charakterystyk wyjściowych $I_D(U_{DS})$ — rys. 6.11a — dzieli się na dwa zakresy (obszary). Jest to *zakres nienasyceń* (często nazywany trydowym lub liniowym) oraz *zakres nasycenia* (często nazywany pentodowym). Punkty o współrzędnych I_{Dsat}, U_{Dsat} (prąd I_D i napięcie U_{DS} odpowiadające początkowi stanu odcięcia) wyznaczają parabolę (linia przerywana na rys. 6.11a) rozdzielająca oba te zakresy.

Reasumując dotychczasowe rozważania stwierdzamy, że tranzystor PNFET może pracować w jednym z trzech zakresów:

- nieprzewodzenia (zatkania) dla $|U_{GS}| > |U_p|$, U_{DS} dowolne;
- nienasyceń dla $|U_{GS}| < |U_p|$ oraz $|U_{DS}| \leq |U_{Dsat}|$;
- nasycenia dla $|U_{GS}| < |U_p|$ oraz $|U_{DS}| > |U_{Dsat}|$.

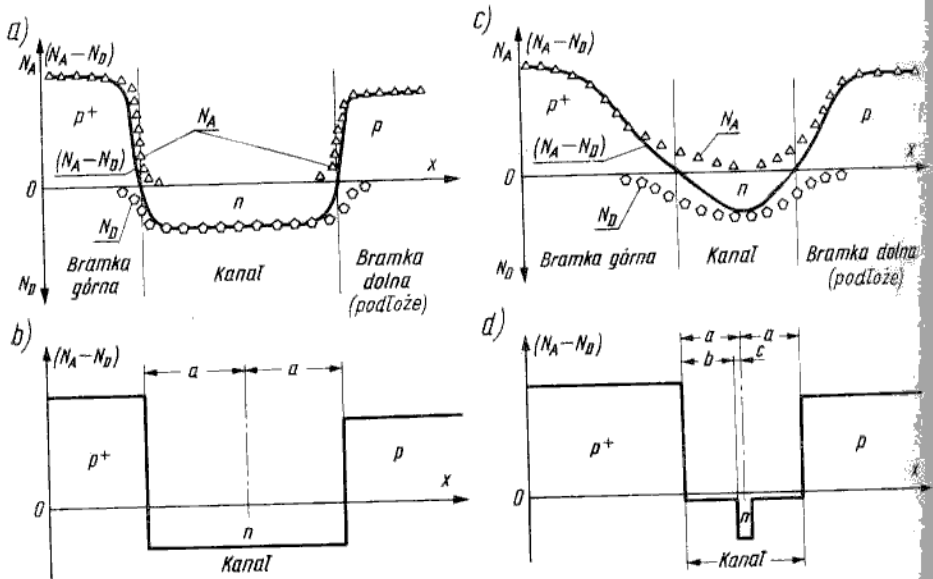
Charakterystyki i parametry statyczne — analiza ilościowa

6.2.2

Statyczne charakterystyki prądowo-napięciowe rozpatrzmy oddzielnie dla zakresów nienasyceń i nasycenia.

Przebieg charakterystyk statycznych w istotny sposób zależy od rozkładu koncentracji domieszek w kanale tranzystora. Bockemuehl¹⁾ wyprowadził równania charakterystyk prądowo-napięciowych tranzystora PNFET, uogólnione dla dowolnego rozkładu domieszek. Wprawdzie wyprowadzenie Bockemuehla jest bardzo błyskotliwe i eleganckie, jest ono jednak zbyt specjalistyczne (wyrażenia końcowe mają dość złożoną postać całkową), by je powtarzać w tej książce. Dlatego przyjmujemy inny sposób postępowania, a mianowicie wyprowadzimy równania charakterystyk dla dwóch skrajnie różnych przypadków rozkładu koncentracji domieszek, przedstawionych na rys. 6.13. Oba przypadki dotyczą tranzystorów epiplanarnych (na podłożu typu p jest nanoszona warstwa epitaksjalna typu n , w której następnie jest wytwarzana warstwa p^+), różniących się tylko grubością kanału (grubością warstwy epitaksjalnej). W przypadku dużej grubości kanału (rys. 6.13a) przejścia od obszaru p^+ do n oraz od n do p są „stromie” w porównaniu z grubością kanału i można przyjąć, że rozkład koncentracji domieszek w kanale jest równomierny. Aproxymowany rozkład koncentracji domieszek w takim tranzystorze przedstawiono na rys. 6.13b.

¹⁾ Bockemuehl R. R.: Analysis of field-effect transistors with arbitrary charge distribution. *IEEE Trans.*, January 1963, Vol. ED-10, pp. 31—34.



Rys. 6.13
Rozkłady koncentracji domieszek dla dwóch przypadków: a), b) duża grubość kanału (rozkład równomierny); c), d) mała grubość kanału (rozkład szpilkowy)

W przypadku cienkiego kanału (rys. 6.13c) przejścia od obszaru p^+ do n oraz od n do p są „łagodne” w porównaniu z grubością kanału i koncentracja domieszek silnie wzrasta w kierunku od krawędzi kanału do jego osi symetrii. Nieco wyolbrzymiając tę sytuację można przyjąć aproksymację rozkładu koncentracji domieszek przedstawioną na rys. 6.13d. Jest to tak zwana *aproksymacja szpilkowa*. Otrzymamy charakterystyki prądowo-napięciowe dla obu przypadków, przedstawiając przy okazji dwie metody wyprowadzenia tych charakterystyk, tj. metodę całkowania potencjału wzdłuż kanału (metoda Shockleya) dla pierwszego przypadku (rozkład równomierny) oraz metodę ładunkową (metoda Middlebrooka) dla drugiego przypadku (rozkład szpilkowy). Należy od razu zwrócić uwagę, że im grubszy jest kanał, tym większa jest wartość napięcia odcięcia, dlatego pierwszy przypadek dotyczy tranzystorów o dużych wartościach U_p (kilkanaście woltów), a drugi — tranzystorów o małych wartościach U_p (kilka woltów).

Zakres nienasylenia

6.2.1

Rozkład równomierny domieszek w kanale
Zanim wyprowadzimy zależności opisujące charakterystyki statyczne w całym zakresie nienasylenia ($0 \leq U_{DS} \leq U_{D sat}$), warto przedstawić prostą analizę dla przypadku bardzo małych napięć drenu ($0 \leq U_{DS} \leq U_{D sat}$), gdyż często tranzystor pracuje w tym zakresie jako liniowy rezystor sterowany napięciem. *Przypadek $0 \leq U_{DS} \leq U_{D sat}$.* Przybliżony przebieg charakterystyki przejściowej dla $U_{DS} \approx 0$ jest już znany z poprzedniego punktu. Teraz będą wyprowadzone dokładniejsze zależności. Ponieważ tranzystor można traktować jak rezystor liniowy, zatem

$$I_D = G_{DS} U_{DS} \tag{6.6}$$

przy czym:

$$G_{DS} = \sigma_c \frac{X_c Z}{L} \text{ — konduktancja kanału} \quad (6.7)$$

$$\sigma_c = q\mu_n n_n \text{ — konduktywność półprzewodnika w obszarze kanału} \\ \text{(przyjmuje się dalej } n_n \approx N_D) \quad (6.8)$$

$$X_c = 2(a-d) \quad (6.9)$$

Można zapisać następujące zależności znane z teorii skokowego złącza p-n:

$$d = \kappa(\varphi_B - U_{GS})^{1/2} \quad (6.10)$$

$$d = a \text{ dla } U_{GS} = U_p, \text{ zatem}$$

$$a = \kappa(\varphi_B - U_p)^{1/2} \quad (6.11)$$

przy czym

$$\kappa = \left| \frac{2\epsilon_s}{qN_D} \right|^{1/2}$$

Na podstawie zależności (6.6) do (6.11) otrzymuje się

$$I_D = \sigma_c \frac{Z}{L} 2\kappa [(\varphi_B - U_p)^{1/2} - (\varphi_B - U_{GS})^{1/2}] U_{DS} \quad (6.12)$$

lub inaczej

$$I_D = G_{DSO} \left[1 - \left(\frac{\varphi_B - U_{GS}}{\varphi_B - U_p} \right)^2 \right] U_{DS} \quad (6.12a)$$

przy czym

$$G_{DSO} = \sigma_c \frac{2aZ}{L} \text{ (konduktancja kanału o grubości metalurgicznej} \\ \text{równej } 2a) \quad (6.13)$$

Ponieważ napięcie odcięcia U_p przyjmuje wartości od kilku do kilkudziesięciu woltów, a $\varphi_B \approx 0,7$ V (dla tranzystorów krzemowych), najczęściej można przyjąć, że:

$$\varphi_B \ll |U_p|$$

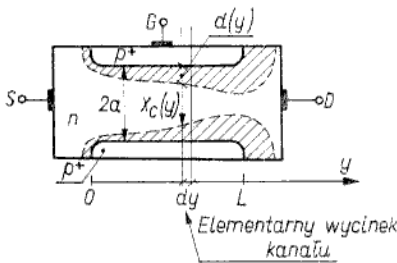
$$\varphi_B \ll |U_{GS}|$$

Stąd wyrażenie (6.12a) można uprościć do postaci znanej już z p. 6.2.1

$$I_D = G_{DSO} \left[1 - \left(\frac{U_{GS}}{U_p} \right)^2 \right] U_{DS} \quad (6.14)$$

Teraz wyprowadzimy równanie charakterystyki prądowo-napięciowej w całym zakresie nienasyceń.

Cały zakres nienasyceń ($0 \leq U_{DS} \leq U_{Dsat}$). Charakterystykę $I_D(U_{DS}, U_{GS})$ wyznaczmy całkując wzdłuż kanału spadki napięć na elementarnych wycinkach kanału (rys. 6.14). Czyni się przy tym dwa założenia:



Rys. 6.14
Przekrój struktury tranzystora, na którym zaznaczono elementarny wycinek kanału.

- prąd kanału dla dowolnej współrzędnej y jest równy prądowi drenu, co wynika z zasady ciągłości prądu w stanie statycznym;
- całe napięcie U_{DS} odkłada się na kanale o długości L , czyli pomija się spadki napięć na obszarach między źródłem a początkiem kanału oraz między końcem kanału a drenem.

Spadek napięcia na elementarnym wycinku kanału dy , zaznaczonym na rys. 6.14

$$dU = I_D dR_C = I_D \frac{dy}{\sigma_C Z \cdot 2[a-d(y)]} \quad (6.15)$$

Dla współrzędnej y napięcie bramka-kanal wynosi $U_{GS} - U(y)$, przy czym $U(y)$ jest potencjałem kanału w punkcie y . Zatem analogicznie do (6.10)

$$d(y) = \kappa[\varphi_B - U_{GS} + U(y)]^{1/2} \quad (6.16)$$

Po podstawieniu (6.16) do (6.15) otrzymuje się równanie różniczkowe pierwszego rzędu, które po rozdzieleniu zmiennych całkuje się od źródła (przy czym $y = 0, U(0) = 0$) do drenu (przy czym $y = L, U(L) = U_{DS}$). Po wykonaniu całkowania i uwzględnieniu (6.13)

$$I_D = G_{DS0} \left\{ U_{DS} - \frac{2}{3} \frac{\kappa}{a} [|\varphi_B - U_{GS} + U_{DS}|^{3/2} - |\varphi_B - U_{GS}|^{3/2}] \right\} \quad (6.17)$$

Ponieważ zwykle $|\varphi_B| \ll |U_{GS}|, |U_{DS}|$, w przybliżeniu zatem

$$I_D \approx G_{DS0} \left\{ U_{DS} - \frac{2}{3} \frac{\kappa}{a} [|U_{DS} - U_{GS}|^{3/2} - |U_{GS}|^{3/2}] \right\} \quad (6.18)$$

Do tych równań podstawia się napięcia z właściwymi znakami, tj. dla tranzystora z kanałem typu n U_{DS} ze znakiem „+”, a U_{GS} ze znakiem „-”.

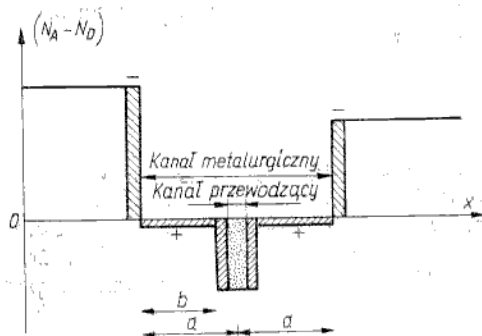
Rozkład szpilkowy domieszek w kanale

Zastosujemy analizę ładunkową, w której podstawową zależnością jest związek między prądem i ładunkiem ruchomym w kanale oraz czasem jego przelotu

$$I_D = \frac{Q_C}{\tau_p} \quad (6.19)$$

W przypadku tranzystora PNFET można przyjąć, że ładunek ruchomy (ładunek kanału Q_C) jest równy różnicy całkowitego ładunku ruchomego, jaki byłby w kanale przy napięciu $U_{GS} = 0$ i ładunku przestrzennego w warstwie zaporowej (rys. 6.15)

$$|Q_C| = |Q_0| - |Q_B| \quad (6.20)$$



Rys. 6.15
 Ilustracja ładunków w tranzystorze o szpilkowym rozkładzie koncentracji domieszek
 Całkowity ładunek w kanale — Q_0 ; powierzchnia zakreskowana w kanale — Q_B ; powierzchnia zakropkowana — Q_C ; $|Q_B| + |Q_C| = |Q_0|$

przy czym: Q_0 — całkowity ładunek ruchomy w kanale przy $U_{GS} = 0$, Q_B — ładunek przestrzenny (nieruchomy) w warstwie zaporowej.

Moduły ładunków wzięto dlatego, że Q_0 jest ujemne (elektrony w kanale typu n), a Q_B — dodatnie. Zachowując znaki należałoby napisać

$$Q_C = Q_0 + Q_B \quad (6.20a)$$

Z kolei czas przelotu

$$\tau_p = \frac{L}{v_n}$$

przy czym v_n — szybkość unoszenia elektronów.

W pierwszym przybliżeniu przyjmuje się, że natężenie pola elektrycznego nie zmienia się wzdłuż kanału, czyli

$$v_n = \mu_n E = \mu_n \frac{U_{DS}}{L}$$

a więc

$$\tau_p = \frac{L^2}{\mu_n U_{DS}} \quad (6.21)$$

Wyrażenie (6.19) można zatem zapisać w postaci

$$I_D = \frac{Q_0 + Q_B}{L^2} \mu_n U_{DS} \quad (6.22)$$

Ponieważ dla $U_{GS} = 0$ ładunek $Q_B = 0$ oraz $I_D = G_{DS0} U_{DS}$, podstawiając zatem do (6.22) $Q_B = 0$ otrzymuje się

$$G_{DS0} = \mu_n \frac{Q_0}{L^2} \quad (6.23)$$

Nie należy przywiązywać żadnej wagi do tego, że podstawienie ujemnej wartości Q_0 daje ujemną wartość G_{DS0} . Konduktancja G_{DS0} jest w sensie fizycznym zawsze dodatnia, co uzyskalibyśmy formalnie uwzględniając konwencję znakowania prądu. Wówczas należałoby do równania (6.22) wprowadzić znak „-”, gdyż prąd drenu powinien być dodatni, odpowiednio (6.23) miałyby postać

$$G_{DS0} = -\mu_n \frac{Q_0}{L^2}$$

Są to jednak uściślenia formalne, więc pominiemy je w dalszych rozważaniach. A więc (6.22) można przepisać w postaci

$$I_D = G_{DS0} \left(1 + \frac{Q_B}{Q_0} \right) U_{DS} \quad (6.24)$$

Pozostaje tylko określić zależność ładunku Q_B od napięć U_{GS} , U_{DS} . Jeżeli pominiemy tę część ładunku Q_B , która jest zawarta w obszarze słabo domieszkowanym (odcinek b na rys. 6.13, 6.15) oraz przyjmiemy, że praktycznie biorąc $b \approx a$, to można zastosować model kondensatora płaskiego. Przyjmuje się więc, że ładunek Q_B jest gromadzony niejako w okładce kondensatora płaskiego, przy czym odległość między okładkami jest równa a . Ten kondensator jest polaryzowany niejednakowo wzdłuż kanału, a mianowicie przy źródle napięcie na okładkach wynosi U_{GS} , ale przy drenie jest ono równe $(U_{GS} - U_{DS})$. Przyjmujemy, że ładunek Q_B jest proporcjonalny do wartości średniej napięcia bramka-kanał

$$Q_B = C_g \left(\frac{U_{DS}}{2} - U_{GS} \right) \quad (6.25)$$

przy czym

$$C_g = 2 \frac{\epsilon_s ZL}{a} \quad (6.26)$$

liczba 2 wynika z symetrii kanału.

Po podstawieniu (6.25) do (6.24) uwzględniając, że

$$U_p = \frac{Q_0}{C_g} \quad (6.27)$$

otrzymuje się

$$I_D = \beta \left[(U_{GS} - U_p) U_{DS} - \frac{U_{DS}^2}{2} \right] \quad (6.28)$$

przy czym $\beta = G_{DS0} / U_p$.

Korzystając z (6.23), (6.27) przepiszemy współczynnik β w innej postaci

$$\beta = \frac{\mu_n}{L^2} C_g$$

Można też wprowadzić pojęcie pojemności C_g^* obliczanej na jednostkę powierzchni ($C_g^* = C_g / 2ZL$) i wówczas uwzględniając (6.26) otrzymuje się

$$\beta = \frac{\mu_n Z C_g^*}{L}$$

Reasumując

$$\beta = \frac{G_{DS0}}{U_p} = \frac{\mu_n C_g}{L^2} = \frac{\mu_n Z C_g^*}{L} \quad (6.29)$$

Zobaczymy dalej, że równanie (6.28) jest słuszne również dla tranzystora MIS, co ma bardzo duże znaczenie, gdyż pozwala ujednoclić metody analizy dla obu rodzajów tranzystorów (PNFET i MISFET).

Zakres nasycenia

6.2.2.2

Rozkład równomierny domieszek w kanale

Równania (6.17), (6.18) mają sens tylko w zakresie $|U_{DS}| \leq |U_{Dsat}|$. W zakresie nasycenia ($|U_{DS}| > |U_{Dsat}|$) można przyjąć w pierwszym przybliżeniu, że prąd drenu nie zmienia się i jest równy $I_D = I_{Dsat}$ dla $U_{DS} = U_{Dsat}$. Po podstawieniu (6.4) do (6.18) otrzymuje się

$$I_{Dsat} = G_{DS0} \left\{ -\frac{U_p}{3} + U_{GS} \left[1 - \frac{2}{3} \left(\frac{U_{GS}}{U_p} \right)^{1/2} \right] \right\} \quad (6.30)$$

Ponieważ $I_{Dsat} = I_{DSS}$, dla $U_{GS} = 0$, zatem

$$\frac{I_{Dsat}}{I_{DSS}} = 1 - 3 \frac{U_{GS}}{U_p} + 2 \left(\frac{U_{GS}}{U_p} \right)^{3/2} \quad (6.31)$$

przy czym

$$I_{DSS} = -G_{DS0} \frac{U_p}{3} \quad (6.32)$$

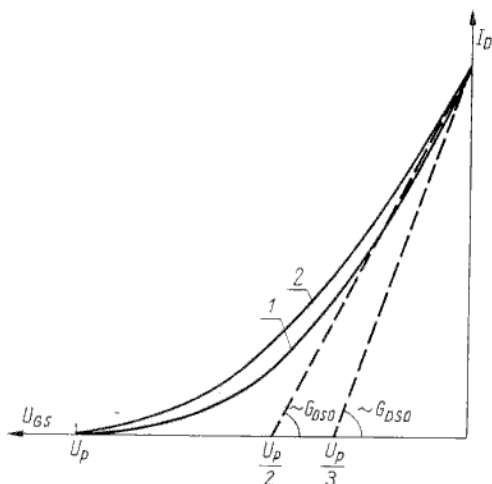
Rozkład szpilkowy domieszek w kanale

Podobnie jak w poprzednim przypadku przyjmuje się, że prąd drenu w zakresie nasycenia nie zależy od napięcia U_{DS} i jest równy wartości I_{Dsat} dla $U_{DS} = U_{Dsat}$. Po podstawieniu (6.4) do (6.28) otrzymuje się

$$\frac{I_{Dsat}}{I_{DSS}} = \left(1 - \frac{U_{GS}}{U_p} \right)^2 \quad (6.33)$$

przy czym

$$I_{DSS} = -G_{DS0} \frac{U_p}{2} \quad (6.34)$$



Rys. 6.16
Charakterystyki przejściowe według (6.31) (krzywa 1) oraz według (6.33) (krzywa 2)

lub w innej postaci

$$I_{D\text{sat}} = \frac{\beta}{2} (U_{GS} - U_p)^2 \quad (6.35)$$

Charakterystyki przejściowe zgodne z wyrażeniami (6.31) i (6.33) porównano na rys. 6.16, na którym pokazano również sposób wyznaczania konduktancji G_{DS0} według zależności (6.32), (6.34).

W dokładniejszej analizie charakterystyki prądowo-napięciowej w zakresie nasycenia należy uwzględnić w obu przypadkach efekt skracania kanału, który powoduje, że charakterystyka $I_D(U_{DS})$ ma nachylenie różne od zera (rys. 6.11b). Wówczas

$$I_D = I_{D\text{sat}} \left(1 - \frac{\delta}{L}\right)^{-1} \quad (6.36)$$

przy czym δ — odcinek Y , Y' na rys. 6.10d.

Oddzielnym i trudnym problemem jest wyznaczenie zależności δ od napięcia U_{DS} , dlatego nie będziemy dalej rozwijać wyrażenia (6.36) traktując je wyłącznie jako wskazówkę informującą, w jaki sposób efekt skracania kanału modyfikuje przebieg charakterystyki prądowo-napięciowej.

Parametry statyczne

6.2.2.3

Dwa zdefiniowane wcześniej parametry statyczne:

— U_p — napięcie odcięcia,

— I_{DSS} — prąd nasycenia dla $U_{GS} = 0$,

wyrazimy w funkcji parametrów fizycznych i konstrukcyjnych tranzystora. Na podstawie (6.11) otrzymujemy dla rozkładu równomiernego domieszek w kanale

$$U_p = -\frac{qN_D a^2}{2\epsilon_s} + \varphi_B \quad (6.37)$$

Na podstawie (6.27), (6.26) otrzymujemy dla rozkładu szpilkowego domieszek w kanale (przyjmuje się $Q_0 = 2qN_D cLZ$)

$$U_p = -\frac{qN_D ca}{\epsilon_s} \quad (6.38)$$

Na podstawie (6.32), (6.37) dla pierwszego przypadku (rozkład równomierny)

$$I_{DSS} = \frac{G_{DS0}}{3} \left(\frac{qN_D a^2}{2\epsilon_s} - \varphi_B \right) \quad (6.39)$$

przy czym

$$G_{DS0} = q\mu_n N_D \frac{2aZ}{L}$$

Na podstawie (6.34), (6.38) dla drugiego przypadku (rozkład szpilkowy)

$$I_{DSS} = \frac{G_{DS0}}{2} \frac{qN_D ca}{\epsilon_s} \quad (6.40)$$

przy czym

$$G_{DSO} = 2 \frac{q\mu_n N_D cZ}{L}$$

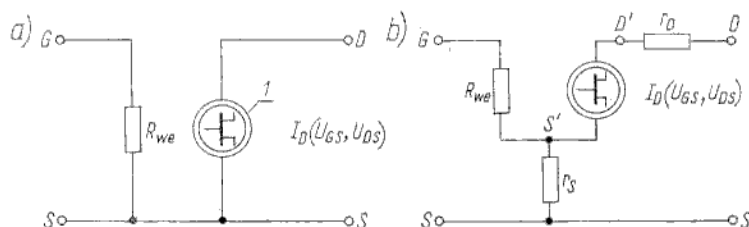
Oprócz tych dwóch parametrów w katalogach zwykle podaje się jeszcze następujące parametry statyczne:

- $r_{DS(O_N)}$ — rezystancja dren-źródło przy $U_{GS} = 0$, $U_{DS} \approx 0$, jest to inaczej $1/G_{DSO}$,
- I_{GSS} — prąd bramki przy dużym napięciu U_{GS} i zwarcie drenu ze źródłem ($U_{DS} = 0$);
- BU_{GSS} — napięcie przebicia bramka-źródło przy $U_{DS} = 0$; przebicie określa się jako taki stan, w którym prąd bramki osiąga pewną dość dużą wartość (np. $I_G = 1 \mu A$).

Model statyczny tranzystora PNFET

6.2.2.4

Najprostszy model statyczny tranzystora PNFET przedstawiono na rys. 6.17a, na którym R_{we} jest rezystancją złącza bramka-kanal spolaryzowanego w kierunku zaporowym (kilka megaomów do kilku gigaomów), a zależne źródło prądowe



Rys. 6.17

Dwa modele statyczne tranzystora PNFET

1 — symbol sterowanego źródła prądowego opisanego równaniami charakterystyk tranzystora PNFET

ma postać odpowiedniego równania charakterystyki prądowo-napięciowej (równania (6.18) lub (6.28) dla zakresu nienasylenia oraz (6.31) lub (6.33) dla zakresu nasylenia).

W dokładniejszym modelu należy uwzględnić rezystancje szeregowe źródła i drenu (rys. 6.17b), których wartości są zwykle mniejsze niż kilkadziesiąt omów.

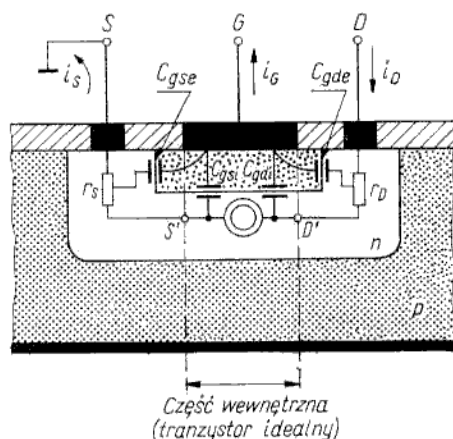
Praca dynamiczna nieliniowa

6.2.3

Reakcja tranzystora PNFET na szybką zmianę warunków polaryzacji nie jest natychmiastowa wskutek dwu zjawisk:

- ładowanie warstwy zaporowej złącza bramka-kanal,
- skończony czas przelotu nośników przez kanał.

Wzważania przeprowadzimy dwustopniowo. Najpierw rozpatrzmy część wewnętrzną tranzystora (tzw. *tranzystor idealny*), zgodnie z podziałem struktury tranzystora, przedstawionym na rys. 6.18. Następnie uwzględnimy elementy reprezentujące



Rys. 6.18
Struktura tranzystora PNFET, ilustrująca sens fizyczny poszczególnych elementów schematu zastępczego. Kierunek prądu bramki odpowiada skokowi napięcia U_{GS} od zera do wartości ujemnej

część zewnętrzną, otrzymując w ten sposób model tzw. *tranzystora rzeczywistego*.

W stanie ustalonym prąd bramki jest, praktycznie biorąc, równy zero, a prądy źródła i drenu są sobie równe. Natomiast w stanie nieustalonym, spowodowanym np. gwałtowną zmianą napięcia U_{GS} , płynie prąd bramki (prąd przesunięcia związany z ładowaniem warstwy zaporowej), a więc zgodnie z prawem Kirchhoffa prądy źródła i drenu różnią się.

Równanie prądów, uwzględniając ich rzeczywiste kierunki (rys. 6.18), dla tranzystora z kanałem typu n ma postać

$$i_G = i_D - i_S \tag{6.41}$$

Ponieważ prądy źródła i drenu są związane z ładunkiem kanału, zatem

$$i_D - i_S = \frac{dQ_C}{dt} \tag{6.42}$$

Zakłada się, że zmiany ładunku kanału są quasi-równowagowe, co oznacza, że bieżącym wartościom napięć u_{GS} , u_{DS} odpowiadają niemal ustalone rozkłady ładunku Q_C . Założenie to oznacza inaczej pominięcie wpływu skończonego czasu przelotu nośników w kanale. W rzeczywistych warunkach jest to założenie usprawiedliwione, gdyż zwykle czasy narastania lub opadania napięć polaryzacji są większe niż czas przelotu.

W granicznym przypadku dla skoku jednostkowego napięcia u_{GS} ze źródła napięciowego stała czasowa procesu przeładowania warstwy zaporowej złącza bramka-kanał jest równa czasowi przelotu nośników przez kanał. Gdyby czas przelotu był dłuższy niż stała czasowa przeładowania warstwy zaporowej, wówczas zmiany ładunku Q_B (w warstwie zaporowej) powodowałyby ładowanie kanału, gdyż elektrony usuwane z warstwy zaporowej do kanału nie byłyby odpowiednio szybko unoszone z kanału do obwodu zewnętrznego. Taki stan byłby fizycznie nierealny, gdyż kanał, tak jak baza w tranzystorze bipolarnym, jest zawsze obojętny elektrycznie. Zatem stała czasowa procesu przeładowania warstwy zaporowej jest równa czasowi przelotu w przypadku sterowania z idealnego źródła napięciowego (ten wniosek potwierdzimy odpowiednimi wzorami przy omawianiu częstotliwości granicznej). Jeżeli jednak weźmiemy pod uwagę

sterowanie ze źródła o skończonej rezystancji wewnętrznej i ponadto uwzględnimy wpływ zjawisk inercyjnych, wnoszonych przez część zewnętrzną tranzystora, to czasu przelotu można nie uwzględniać jako znacznie krótszego niż stała czasowa procesu przeladowywania warstwy zaporowej. Takie jest uzasadnienie założenia o quasi-równowagowych zmianach ładunku w kanale. Przyjmuje się więc, że Q_C zmienia się w funkcji czasu wskutek zależności pośredniej, tj. Q_C jest funkcją napięcia, a napięcie jest funkcją czasu

$$Q_C = f(u_{GS}, u_{GS} - u_{DS}) \quad (6.43)$$

Należy zauważyć, że jako drugą zmienną niezależną wybrano nie u_{DS} , lecz $u_{GS} - u_{DS} = u_{GD}$. Takie podejście umożliwia uzyskanie bezpośredniej reprezentacji modelowej tak ważnego zjawiska, jakim jest sprzężenie bramka-dren. Na podstawie (6.43), (6.42), (6.41)

$$i_G = \frac{\partial Q_C}{\partial u_{GS}} \frac{du_{GS}}{dt} + \frac{\partial Q_C}{\partial u_{GD}} \frac{du_{GD}}{dt} \quad (6.44)$$

Przyrosty ładunku kanału są wywołane odpowiednimi przyrostami ładunku warstwy zaporowej

$$\Delta Q_C = -\Delta Q_B$$

Zdefiniujmy pojemności różniczkowe:

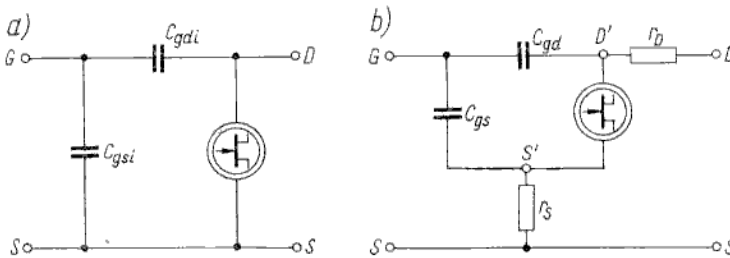
$$C_{gsi} = - \left. \frac{\partial Q_B}{\partial u_{GS}} \right|_{u_{GD}} \text{ — pojemność bramka-źródło dla tranzystora idealnego}$$

$$C_{gdi} = - \left. \frac{\partial Q_B}{\partial u_{GD}} \right|_{u_{GS}} \text{ — pojemność bramka-dren dla tranzystora idealnego}$$

Stąd

$$i_G = C_{gsi} \frac{du_{GS}}{dt} + C_{gdi} \frac{du_{GD}}{dt} \quad (6.45)$$

Pierwsza składowa prądu bramki płynie w obwodzie bramka-źródło, druga składowa — w obwodzie bramka-dren. Jeżeli uwzględnimy, że przez zaciski źródła i drenu płynie ponadto prąd przewodzenia I_D , to można skonstruować schemat zastępczy, przedstawiony na rys. 6.19a. W części zewnętrznej tranzystora (rys. 6.18) są pojemności warstwy zaporowej C_{gse} , C_{gde} oraz rezystancje szere-



Rys. 6.19
Dwa modele dynamiczne nieliniowe tranzystora PNFET

gowe źródła r_S i drenu r_D . Elementy C_{gse} , r_S oraz C_{gde} , r_D mają wartości rozłożone, które jednak dla prostoty zastąpimy obwodami o stałych skupionych, włączając pojemności między bramką a zaciskami struktury wewnętrznej. W ten sposób otrzymuje się schemat zastępczy, przedstawiony na rys. 6.19b, na którym:

$$C_{gs} = C_{gst} + C_{gse}$$

$$C_{gd} = C_{gdi} + C_{gde}$$

Obie pojemności są zależne od napięć U_{GS} , U_{DS} . Charakterystyki pojemnościowo-napięciowe są zwykle podawane w katalogach. Mają one przebieg hiperboliczny, typowy dla pojemności warstwy zaporowej złącza $p-n$. Przeciętne wartości poszczególnych pojemności w tranzystorze małej mocy przy zerowej polaryzacji są następujące:

$C_{gst} \approx 5$ pF, $C_{gdi} \approx 2$ pF, C_{gse} , $C_{gde} \approx 1$ pF. Zmiany tych pojemności w funkcji napięcia są ok. 2-krotne.

Praca z małymi sygnałami

6.2.4

Podobnie jak dla tranzystora bipolarnego, właściwości tranzystora PNFT przy sterowaniu małym sygnałem można opisać modelami czwórnikowymi lub modelami fizycznymi (schematami zastępczymi). Z modeli czwórnikowych najczęściej jest stosowana macierz typu y , gdyż impedancje wyjściowa i wyjściowa tranzystorów polowych są duże, czyli łatwo jest spełnić warunek zwarcia na wejściu i wyjściu przy pomiarze parametrów tej macierzy. Zapis tej macierzy oraz definicje parametrów admitancyjnych, podane w poprzednim rozdziale, odnoszą się również do tranzystora polowego, gdyż modele sformalizowane obwodowo z samej ich istoty nie zależą od właściwości fizycznych elementów. Wystarczy więc rozpatrzeć modele fizyczne. Otrzymamy proste modele dla małych i dużych częstotliwości linearyzując elementy modeli nieliniowych. Małe i duże częstotliwości są rozumiane tak samo jak dla tranzystora bipolarnego, czyli w modelu dla małych częstotliwości nie ma elementów reaktancyjnych.

Modele i parametry dla małych częstotliwości

6.2.4.1

Model dla małych częstotliwości, czyli model quasi-statyczny, otrzymuje się linearyzując elementy modelu statycznego (rys. 6.17). Przyjmuje się, że R_{we} jest rezystancją liniową, która w związku z tym pozostaje bez zmian ($r_{gs} = R_{we}$). Funkcję $I_D(U_{GS}, U_{DS})$ opisującą źródło prądowe należy rozłożyć w szereg Taylora i uwzględnić wyrazy rzędu pierwszego

$$i_d = \frac{\partial I_D}{\partial U_{GS}} u_{gs} + \frac{\partial I_D}{\partial U_{DS}} u_{ds} \quad (6.46)$$

przy czym i_d , u_{gs} , u_{ds} — małe amplitudy sygnałów zmiennych.

Zostaną teraz zdefiniowane następujące dwa pojęcia:

$$\text{— transkonduktancja } g_m = \frac{\partial I_D}{\partial U_{GS}} \quad (6.47)$$

$$\text{— konduktancja wyjściowa } g_{ds} = \frac{\partial I_D}{\partial U_{DS}} \quad (6.48)$$

C lub korzystać z bardziej rozbudowanych modeli o stałych skupionych, w których jednak niektóre parametry muszą być uzależnione od częstotliwości. Jeden z wariantów takiego modelu przedstawiono na rys. 6.21b. Nie będziemy go szerzej komentować, zauważmy tylko, że r_G jest rezystancją doprowadzenia bramki, zaś $r_{g's'}$, $r_{g'd'}$ stanowią określone części rezystancji kanału.

Należy w ogóle stwierdzić, że analiza właściwości tranzystora PNFET w zakresie bardzo dużych częstotliwości jest zagadnieniem bardzo trudnym, rozważanym wyłącznie w specjalistycznych monografiach. Omówimy tylko zagadnienie maksymalnej częstotliwości pracy tego tranzystora. Pojęcie częstotliwości maksymalnej w tranzystorach polowych jest związane ze skończonym czasem przelotu nośników przez kanał i ze stałą czasową ładowania pojemności bramka-kanał.

Czas przelotu

$$\tau_p = \frac{L}{\mu_n E_y} = \frac{L^2}{\mu_n U_{DS}} \quad (6.55)$$

stąd

$$\omega_{max} = \frac{1}{\tau_p} = \frac{\mu_n U_{DS}}{L^2} \quad (6.55a)$$

Stałą czasową ładowania pojemności bramka-kanał w tranzystorze idealnym przyysterowaniu ze źródła napięciowego określa iloczyn pojemności C_g przez rezystancję kanału r_C

$$\omega_{max} = \frac{1}{C_g r_C} \quad (6.56)$$

Ścisłe biorąc, sytuacja nie jest tak prosta, gdyż pojemność bramka-kanał i rezystancja kanału mają wartości rozłożone wzdłuż kanału, jednak można je w pierwszym przybliżeniu zastąpić dwoma elementami, tzw. średnią pojemnością C_g i średnią rezystancją r_C . W stanie nasycenia (dla tranzystora o równomiernym rozkładzie koncentracji domieszek w kanale) przy $U_{GS} = 0$

$$C_g = \underset{\substack{\uparrow \\ \text{uwzględnia się} \\ \text{dwie bramki}}}{2} \frac{\varepsilon_s Z L}{a} \quad (6.57)$$

$$r_C = \frac{L}{\sigma_C a Z} \quad (6.58)$$

Po podstawieniu (6.57) i (6.58) do (6.56) otrzymuje się

$$\omega_{max} = \frac{a^2 q \mu_n N_D}{2 \varepsilon_s L^2} \quad (6.59)$$

Można też zdefiniować częstotliwość graniczną jako taką częstotliwość, przy której prąd wejściowy, płynący przez pojemność C_g , jest równy wydajności wyjściowego źródła prądowego $g_m u_{gs}$

$$\omega_{max} C_g u_{gs} = g_m u_{gs}$$

$$\omega_{max} = \frac{g_m}{C_g} \quad (6.60)$$

Po podstawieniu do tego wzoru (6.50) i (6.29) otrzymuje się

$$\omega_{max} = \frac{\mu_n |U_{GS} - U_p|}{L^2} \quad (6.61)$$

Bardzo interesujące jest, że dla zakresu nasycenia przy $U_{GS} = 0$ (wówczas $U_{Dsat} = |U_p|$) częstotliwości graniczne, określone wzorami (6.55a) i (6.61) są jednakowe. Podstawiając $U_{DS} = U_{Dsat} = |U_p|$ do (6.55a) oraz $U_{GS} = 0$ do (6.61) otrzymujemy

$$\omega_{max} = \frac{\mu_n}{L^2} |U_p| \quad (6.62)$$

Jeżeli do (6.62) podstawimy $U_p = qN_D a^2 / 2\epsilon_s$, to otrzymamy dość zaskakujący wynik, a mianowicie częstotliwości graniczne, określone wzorami (6.62) i (6.59), są jednakowe. Równość trzech różnie zdefiniowanych częstotliwości granicznych wskazuje na dwa fakty:

- czas przelotu nośników przez kanał jest równy stałej czasowej obwodu bramka-źródło (mówiono o tym już w p. 6.2.3); ten wniosek dotyczy oczywiście tylko tranzystora idealnego;
- transkonduktancja g_m jest równa odwrotności rezystancji kanału.

Wyrażenie (6.59) przepiszmy w postaci

$$f_{max} = \frac{\sigma_c}{4\pi\epsilon_s} \left(\frac{a}{L}\right)^2 \quad (6.63)$$

i zwróćmy uwagę, że zwiększeniu częstotliwości maksymalnej sprzyjają następujące czynniki:

- duża konduktywność kanału (tranzystory z kanałem typu n są „szybsze” niż z kanałem typu p , gdyż ruchliwość elektronów jest większa),
- duża grubość kanału (niestety zwiększenie grubości kanału prowadzi do niekorzystnego wzrostu napięcia U_p),
- mała długość kanału (tę wielkość ograniczają możliwości technologii).

Należy zauważyć, że częstotliwość maskymalna nie zależy od szerokości kanału. Trzeba pamiętać, że wyrażenie (6.63) wyprowadzono dla tranzystora idealnego. W tranzystorze rzeczywistym częstotliwość maksymalna jest mniejsza z uwagi na wpływ dodatkowych pojemności.

Tranzystor MIS

6.3

Tranzystor MIS stanowi dojrzałą technicznie realizację idei elementu wzmacniającego, zilustrowanej na rys. 6.1. Zatem tranzystor MIS należy do rodziny tranzystorów z izolowaną bramką (IGFET), w których przewodność kanału półprzewodnikowego jest sterowana poprzecznym polem elektrycznym, działającym poprzez warstwę dielektryka, izolującą kanał od bramki. Do tej samej rodziny tranzystorów IGFET należą jeszcze tranzystory cienkowarstwowe (TFT). Jak już częściowo wyjaśniono w p. 6.1, zasadnicza różnica między tranzystorem MIS a tranzystorem cienkowarstwowym polega na tym, że w pierwszym kanał powstaje w monokrystalicznej warstwie półprzewodnika, stanowiącej zwykle jednolitą całość z płytką podłożową, spełniającą funkcję nośnika mecha-

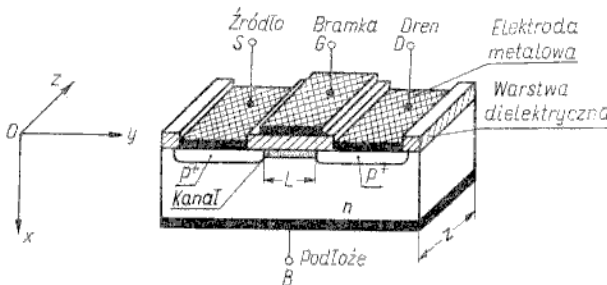
nicznego¹⁾, a w drugim kanał powstaje w cienkiej warstwie półprzewodnika polikrystalicznego, naniesionej na podłoże dielektryczne (płytkę szklaną lub ceramiczną). Znaczenie praktyczne tranzystorów MIS jest nieporównywalnie większe niż tranzystorów cienkowarstwowych, dlatego często nazwy IGFET, MISFET są używane jako synonimy. Istnieje wiele analogii w zasadzie działania i właściwościach tranzystora MIS i PNFET. Żeby maksymalnie wykorzystać to podobieństwo, przyjmujemy układ materiału w niniejszym punkcie niemal identyczny jak w p. 6.2.

Zakłada się, że są znane podstawowe właściwości struktury MIS, opisane w rozdziale 3 (p. 3.3).

Budowa i zasady działania tranzystora MIS — opis jakościowy

6.3.1

Z klasyfikacji podanej na rys. 6.2 wynika, że istnieje kilka typów tranzystorów MIS. Zanim jednak wyjaśnimy tę klasyfikację, rozpatrzmy najpierw elementarne zasady działania tranzystora z kanałem typu *p*, którego budowę przedstawiono schematycznie na rys. 6.22²⁾. W płytce monokrystalicznej krzemu typu *n* o re-



Rys. 6.22

Schematyczna ilustracja budowy tranzystora MIS z kanałem typu *p*

zystywności ok. $0,01 \dots 0,1 \Omega \cdot \text{m}$ wytworzono przez dyfuzję lub implantację dwa obszary silnie domieszkowane o odmiennym od podłoża typie przewodnictwa, tj. p^+ z koncentracją domieszki $10^{24} \dots 10^{26} \text{ m}^{-3}$. Jeden z tych obszarów silnie domieszkowanych, nazywany *źródłem*, znajduje się w odległości kilku do kilkudziesięciu mikrometrów od drugiego, nazywanego *drenem*; oba są pokryte kontaktowymi warstwami metalicznymi. Powierzchnia półprzewodnika między źródłem a drenem jest pokryta warstwą dielektryka (najczęściej SiO_2 o grubości ok. $100 \dots 150 \text{ nm}$). Elektroda metalowa, znajdująca się na powierzchni dielektryka, nosi nazwę *bramki*. Oznaczenia literowe poszczególnych elektrod są takie jak dla tranzystora PNFET, tj. *S* — źródło, *G* — bramka, *D* — dren. Może też być wyprowadzona końcówka podłoża oznaczona literą *B* (ang. *base*). Obwód prądu przepływającego od źródła do drenu zamyka się przez obszar przypowierzchniowy półprzewodnika, leżący pod warstwą dielektryczną. Od konduktancji tego obszaru zależy więc wartość prądu źródło-dren. W zależności od polaryzacji bramki

¹⁾ Wyjątkowo w tzw. technologii SOS (krzem na szafirze — ang. *Silicon on Sapphire*) płytka podłożowa jest izolatorem (por. p. 8.1.1.2).

²⁾ Ten rodzaj tranzystora MIS odegrał największą rolę w rozwoju technologii układów scalonych MIS.

oraz od innych czynników (np. kontaktowa różnica potencjałów, ładunek nieskompensowany w dielektryku i stanach powierzchniowych) w półprzewodniku pod bramką tworzy się warstwa akumulacyjna (dużo elektronów i bardzo mało dziur), zubożona (mało elektronów i mało dziur) lub inwersyjna (mało elektronów i bardzo dużo dziur). W przypadku akumulacji lub zubożenia droga od źródła do drenu odpowiada dwu złączom $p^+ - n$ włączonym szeregowo przeciwnie. Wówczas niezależnie od polaryzacji drenu w stosunku do źródła (dodatniej czy ujemnej) prąd źródło-dren ma znikomo małą wartość, gdyż może to być tylko prąd wsteczny jednego z wymienionych złączy.

Natomiast w przypadku inwersji obszary p^+ źródła i drenu są połączone warstwą inwersyjną, która w sensie fizycznym ma ten sam typ przewodnictwa (dużo dziur i mało elektronów, czyli typ p). Warstwę inwersyjną łączącą źródło z drenem będziemy nazywać *kanalem*. Istnienie kanału typu p umożliwia przepływ dużego prądu dziurowego między źródłem a drenem. Zmiany napięcia bramki modulują konduktancję kanału, sterując w ten sposób wartość prądu źródło-dren. Taka jest elementarna zasada sterowania prądu w obwodzie źródło-dren (w obwodzie wyjściowym) za pomocą napięcia bramki-źródło (napięcia wejściowego).

Mając ogólny obraz sposobu działania tranzystorów MIS, przystąpimy teraz do ich klasyfikacji.

Klasyfikacja tranzystorów MIS

6.3.1.1

Ze względu na typ przewodnictwa kanału tranzystory MIS (podobnie jak PNFET) można podzielić na dwa rodzaje:

- tranzystor MIS z kanałem typu p (na podłożu typu n), w którym występuje przewodnictwo dziurowe;
- tranzystor MIS z kanałem typu n (na podłożu typu p), w którym występuje przewodnictwo elektronowe.

Dalszy podział tranzystorów MIS zależy od przyjętych kryteriów. Pomijamy kryteria technologiczno-konstrukcyjne¹⁾.

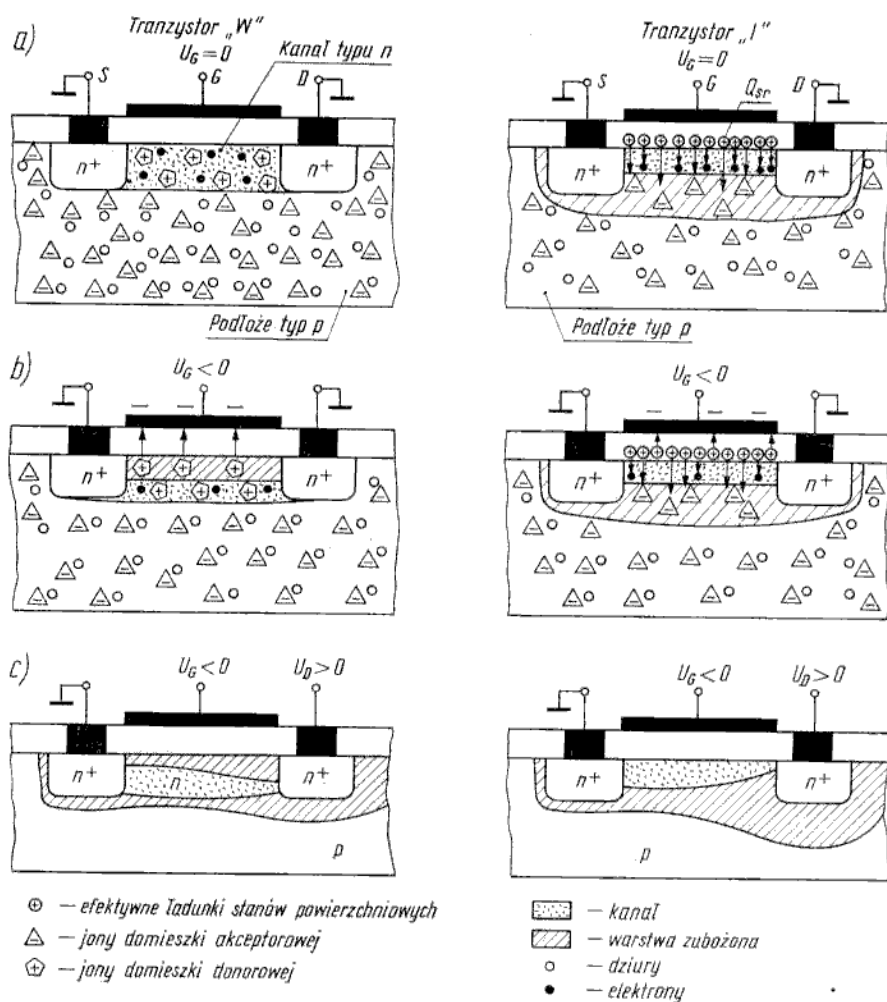
Dokonyamy podziału ze względu na kryteria, które umownie nazwiemy „fizycznymi” i „układowymi”, tj. podzielimy tranzystory biorąc pod uwagę istotne różnice w zjawiskach fizycznych lub w przebiegu podstawowych charakterystyk.

Ze względu na różnice w zjawiskach fizycznych tranzystory MIS dzieli się na dwa rodzaje:

- z *kanalem zaindukowanym*, czyli z kanałem w postaci warstwy inwersyjnej;
- z *kanalem wbudowanym*, czyli z kanałem w postaci warstwy domieszkowanej o przeciwnym typie przewodnictwa niż podłoże.

Występują tu również oczywiste różnice technologiczne, lecz nie są one w tej chwili najważniejsze. Najistotniejsze ze względu na właściwości fizyczne są różnice w sposobie modulacji przewodności kanału, co obrazowo przedstawiono na rys. 6.23 na przykładzie tranzystorów z kanałami typu n . Dogodnie będzie zastąpić zbyt opisowe nazwy tych tranzystorów krótkimi nazwami: tranzystor „I” (z kanałem indukowanym), tranzystor „W” (z kanałem wbudowanym). Żeby łatwiej było porównać zjawiska najistotniejsze w tej fazie rozważań, przyjmujemy, że w obu tranzystorach przewodności kanałów są identyczne przy napięciu $U_{GS} = 0$ (zasady oznaczania napięć i prądów w tranzystorze MIS są identyczne jak dla tranzystora PNFET).

¹⁾ Ponieważ tranzystory MIS mają nieporównywalnie większe znaczenie jako elementy układów scalonych niż jako elementy indywidualne, bardziej szczegółowy opis technologii i konstrukcji tych tranzystorów będzie zamieszczony w rozdziale 8.



Rys. 6.23

Ilustracja zjawisk w tranzystorach typu „W” oraz typu „I” dla trzech stanów polaryzacji: a) stan neutralny, $U_{GS} = U_{DS} = 0$; b) $U_{GS} \neq 0$, $U_{DS} = 0$; c) $U_{GS} \neq 0$, $U_{DS} \neq 0$;

Na rysunku 6.23a przedstawiono sytuację w obu tranzystorach dla stanu neutralnego (wszystkie napięcia równe zero). W tranzystorze „W” kanałem jest warstwa typu n domieszkowana donorami. W tranzystorze „I” kanałem jest warstwa inwersyjna typu n, powstała wskutek przeciągnięcia dużej liczby elektronów do powierzchni przez dodatni ładunek powierzchniowy Q_{sr} . Jednakową przewodność obu kanałów symbolizuje jednakowa liczba elektronów (po sześć). Pod kanałem w tranzystorze „I” jest warstwa zubożona. W stanie neutralnym nie ma jeszcze istotnych różnic między obu tranzystorami poza tą, że kanał w tranzystorze „W” jest obojętny elektrycznie (równowaga ładunku elektronów i jonów dodatnich), a w tranzystorze „I” jest naładowany ujemnie. W przypadku ujemnej polaryzacji bramki (rys. 6.23b) elektrony są odpychane w głąb pół-

przewodnika, co powoduje w tranzystorze „W” powstanie przypowierzchniowej warstwy zubożonej, w tranzystorze „I” natomiast — zmniejszenie ładunku elektronów w kanale. Linie sił pola elektrycznego, narysowane na rys. 6.23b, umożliwiają dokładne prześledzenie, w jaki sposób jest zachowana równowaga ujemnych i dodatnich ładunków w obu tranzystorach.

Przewodności kanałów w obu tranzystorach zmieniły się identycznie, jednak wyraźnie widać podstawową różnicę w ich działaniu, polegającą na tym, że kanał w tranzystorze „W” odsuwa się od powierzchni (duże podobieństwo z tranzystorem PN-FET), a w tranzystorze „I” pozostaje zawsze przy powierzchni.

Rysunek 6.23c już w bardziej uproszczonej postaci przedstawia profile kanałów w obu tranzystorach przy polaryzacji bramki i drenu. Porównanie obu tranzystorów wskazuje na to, że działanie tranzystora „I” jest bardziej zależne od zjawisk powierzchniowych (głównie od rozpraszania nośników na powierzchni). Na szczęście te różnice między obu tranzystorami mają niewielki wpływ na ich charakterystyki i parametry, dlatego w analizie uproszczonej (a taka nas tylko interesuje) można ich nie uwzględniać. Rozpatrzmy zatem tylko jeden typ tranzystora, a mianowicie tranzystor z kanałem indukowanym, traktując wszystkie wyniki i wnioski jako słuszne w pierwszym przybliżeniu również dla tranzystora z kanałem wbudowanym.

Jaki jest podział tranzystorów MIS ze względu na kryteria, które nazwano umownie „układowymi”, tj. ze względu na różnice w przebiegu podstawowych charakterystyk? W tym aspekcie wyróżnia się dwa rodzaje tranzystorów MIS, a mianowicie:

- tranzystor z kanałem zubożonym (ang. *depletion mode*), inaczej nazywany *tranzystorem normalnie włączonym* (ang. *normally on*),
- tranzystor z kanałem wzbogacanym (ang. *enhancement mode*), inaczej nazywany *tranzystorem normalnie wyłączonym* (ang. *normally off*).

Oba tranzystory, przedstawione na rys. 6.23, należą do grupy tranzystorów z kanałem zubożonym. Kanał w tych tranzystorach istnieje bez polaryzacji bramki, czyli przy $U_{GS} = 0$ może płynąć duży prąd drenu (tranzystor jest normalnie włączony), a działając ujemnym napięciem bramki można zmniejszyć konduktancję kanału („zubożyć” go w sensie posiadania mniejszej liczby nośników). W przypadku tranzystora z kanałem indukowanym można sobie wyobrazić inną sytuację niż przedstawiona na rys. 6.23. Na przykład gdyby nie było ładunku dodatniego Q_{sr} , wówczas przy $U_{GS} = 0$ kanał by nie istniał i nie mógłby płynąć prąd drenu (tranzystor byłby normalnie wyłączony). Dopiero działając odpowiednio dużym dodatnim napięciem bramki można by zaindukować kanał (włączyć tranzystor) i dalszy wzrost napięcia bramki powodowałby zwiększenie konduktancji kanału („wzbogacenie” kanału w sensie posiadania coraz większej liczby nośników). Byłby to tranzystor z kanałem wzbogacanym.

W przypadku podłoża typu n dodatni ładunek Q_{sr} wywołuje akumulację elektronów, czyli przy $U_{GS} = 0$ nie ma kanału (nie ma warstwy inwersyjnej typu p). Obecność dodatniego ładunku powierzchniowego Q_{sr} jest niejako naturalną właściwością struktur M-SiO₂-Si (por. p. 3.3.3), co prowadzi do inwersji w półprzewodniku typu p oraz akumulacji w półprzewodniku typu n . Nic więc dziwnego, że często tranzystor z indukowanym kanałem zubożonym kojarzy się z podłożem typu p (kanał typu n), tranzystor zaś z indukowanym kanałem wzbogacanym — z podłożem typu n (kanał typu p). Są jednak możliwe konstrukcje tranzystora z kanałem zubożonym na podłożu typu n oraz tranzystora z kanałem wzbogacanym na podłożu typu p .

Dla ścisłości należy dodać, że w tranzystorach z kanałem zubożonym możliwa jest również praca ze wzbogacaniem. Na przykład w tranzystorach pokazanych na rys. 6.23 można uzyskać zwiększenie konduktancji kanału (zwiększenie prądu drenu) polaryzując bramkę napięciem dodatnim. W związku z tym niekiedy można spotkać nazwę: tranzystory typu zubożanie/wzbogacanie. Będziemy jednak stosować tradycyjną nazwę: tranzystory z kanałem zubożonym.

Reasumując rozważania nad klasyfikacją tranzystorów MIS, przyjmijmy podział na tranzystory z kanałem zubożonym lub wzbogacającym jako nadrzędny, przy czym pierwsze mogą mieć kanał wbudowany lub indukowany, a drugie tylko indukowany. Stwierdzono już wcześniej, że w uproszczonej analizie nie ma istotnych różnic między tranzystorem z kanałem wbudowanym, a tranzystorem z kanałem indukowanym. Należy teraz dodać, że różnice między tranzystorem z kanałem wzbogacającym a tranzystorem z kanałem zubożonym dają się sprowadzić do jednej formalnej różnicy wartości napięcia progowego (napięcie U_{GS} , przy którym następuje przejście ze stanu nieprzewodzenia do przewodzenia). Dlatego szczególne rozważania przeprowadzimy dla jednego rodzaju tranzystora, tj. dla tranzystora z kanałem wzbogacającym typu p . Wyniki tych rozważań można bez trudu odnieść do pozostałych rodzajów tranzystorów MIS po uwzględnieniu odpowiednich zmian formalnych (jak np. zmiana wartości napięcia progowego, zmiana polaryzacji itp.). Tak jak dla tranzystora PNFET rozpatrzmy procesy fizyczne, warunkujące określone przebiegi podstawowych charakterystyk tranzystora, tj.:

— przejściowej $I_D(U_{GS})|_{U_{DS}=\text{const}}$

— wyjściowej $I_D(U_{DS})|_{U_{GS}=\text{const}}$

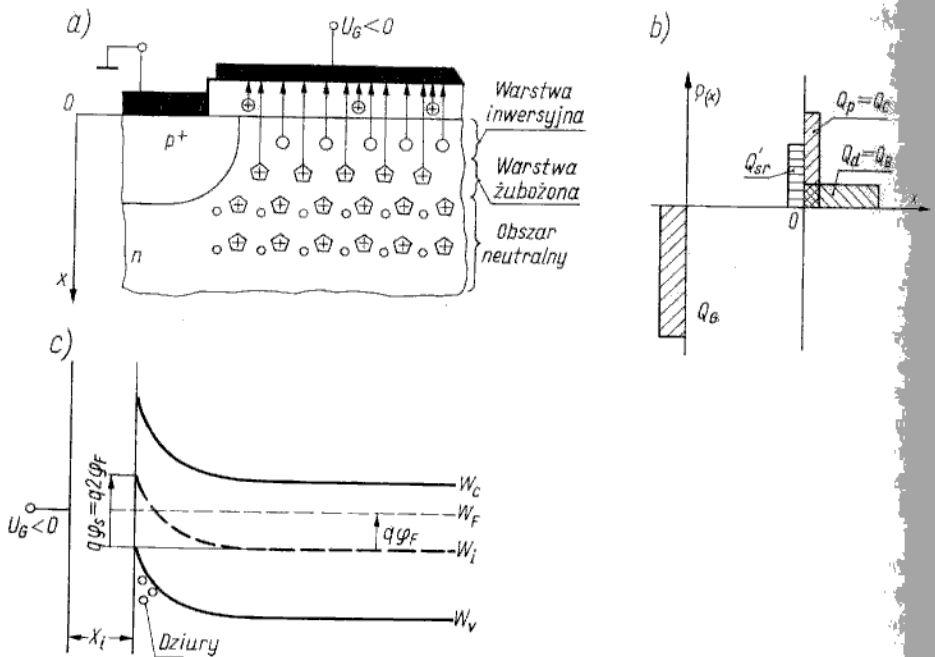
W tranzystorze MIS należy również uwzględniać wpływ polaryzacji podłoża ($U_{BS} \neq 0$) na te charakterystyki.

Charakterystyka przejściowa — pojęcie napięcia progowego

6.3.1.2

Rozpatrzmy zjawiska określające przebieg charakterystyki przejściowej, tj. zależność prądu drenu (prądu wyjściowego) od napięcia bramka-źródło (napięcia wejściowego) dla przypadku bardzo małego napięcia dren-źródło ($U_{DS} \leq 0$). Przyjmuje się również, że źródło jest zwarte z podłożem i uziemione. Zmieniając napięcie U_{GS} od wartości dodatnich do dużych wartości ujemnych uzyskuje się w półprzewodniku stany akumulacji, zubożenia lub inwersji. Ograniczymy się do zilustrowania ładunków oraz poziomów energetycznych w obszarze półprzewodnika znajdującym się pod bramką dla najbardziej interesującego przypadku inwersji (rys. 6.24). Na rysunku 6.24a przedstawiono ładunki i linie sił pola elektrycznego pod bramką tranzystora. Rozkład ładunków ilustruje rys. 6.24b. Ładunek oznaczony przez Q'_{sr} reprezentuje efektywne działanie wszystkich czynników wewnętrznych, powodujących zagięcie pasm energetycznych półprzewodnika, tj. kontaktowej różnicy potencjałów, stanów powierzchniowych i ładunków nieskompensowanych w warstwie dielektryka (czynnikiem zewnętrznym jest napięcie $U_{GS} \neq 0$)¹⁾. Na rysunku 6.24c przedstawiono zagięcie pasm energetycznych przy powierzchni półprzewodnika, odpowiadające wartości potencjału powierzchniowego $\varphi_s = 2\varphi_F$, przy której rozpoczyna się silna inwersja w obszarze przypowierzchniowym półprzewodnika (por. p. 3.3.1). Wartość

¹⁾ W rozdziale 3 przez Q_{sr} oznaczono efektywny ładunek stanów powierzchniowych, natomiast Q'_{sr} uwzględnia ponadto wpływ kontaktowej różnicy potencjałów φ_{ms} , czyli $Q'_{sr} = Q_{sr} - \varphi_{ms} C_t$ (C_t — pojemność dielektryka).



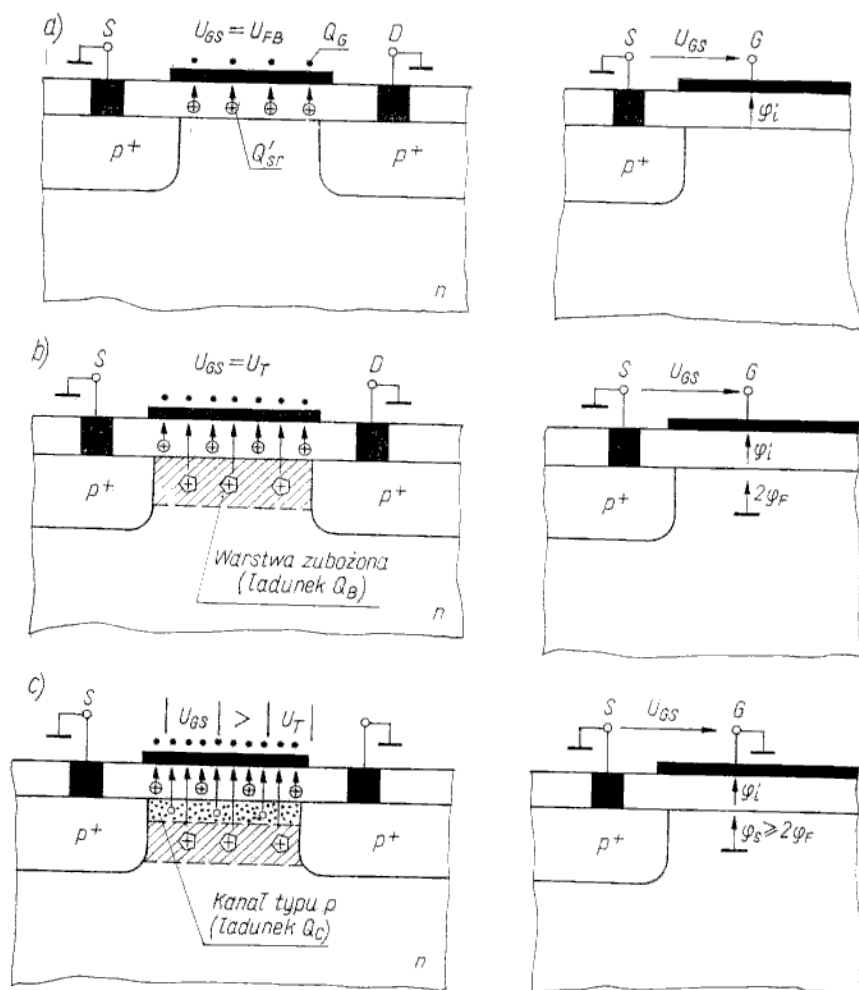
Rys. 6.24

Ilustracja ładunków oraz poziomów energetycznych w obszarze półprzewodnika, znajdującym się pod bramką: a) wycinek tranzystora przedstawiający ładunki i linie sił pola elektrycznego; b) rozkład ładunków w przekroju poprzecznym tranzystora; c) energetyczny model pasmowy przy powierzchni półprzewodnika dla stanu inwersji

napięcia U_{GS} , przy której potencjał powierzchniowy $\varphi_s = 2\varphi_F$, nazywa się *napięciem progowym* i będzie oznaczane przez U_T ¹⁾ (ang. *threshold voltage*). Rola tego parametru w opisie właściwości tranzystorów MIS jest podobna jak rola napięcia U_p w tranzystorach PNFET. Wzrost napięcia ujemnego U_{GS} powyżej wartości U_T odpowiada wzrostowi wartości potencjału powierzchniowego φ_s powyżej wartości $2\varphi_F$, zapoczątkowuje zatem powstanie kanału typu p , łączącego obszary p^+ źródła i drenu. W sensie fizycznym napięcie progowe jest najmniejszą wartością napięcia bramki, niezbędną do wytworzenia ładunku bramki Q_G , kompensującego ładunek Q'_{sr} oraz ładunek domieszek zjonizowanych w półprzewodniku Q_B (tzw. ładunek podłoża). Dowolny przyrost napięcia U_{GS} (ujemnego dla kanału typu p) powyżej wartości U_T tworzy dodatkowy ładunek bramki, który jest kompensowany ładunkiem nośników mniejszościowych (dziur) w powstającym kanale. Na rysunku 6.25 przedstawiono ilustracje objaśniające pojęcie napięcia progowego. Na rysunku 6.25a przedstawiono stan wyprostowanych pasm, tj. $U_{GS} = U_{FB}$. W stanie wyprostowanych pasm ładunek półprzewodnika jest równy zeru, a więc potencjał powierzchniowy również jest równy zeru ($\varphi_s = 0$). Ponieważ napięcie U_{GS} jest równe sumie spadków napięcia na dielektryku i półprzewodniku ($U_{GS} = \varphi_t + \varphi_s$), zatem

$$U_{FB} = \varphi_t = -\frac{Q'_{sr}}{C_i}$$

¹⁾ Napięcie progowe w katalogach jest oznaczane jako $U_{GS(th)}$.



Rys. 6.25

Ilustracje ładunków i potencjałów w strukturze MOS dla trzech stanów: a) $U_{GS} = U_{FB}$; b) $U_{GS} = U_T$; c) $|U_{GS}| > |U_T|$

Uwzględniając przypis 1) na stronie 387 otrzymuje się wyrażenie znane z rozdziału 3 (wzór (3.94))

$$U_{FB} = \varphi_{ms} - \frac{Q_{sr}}{C_i}$$

Na rysunku 6.25b przedstawiono sytuację dla $U_{GS} = U_T$. Wówczas

$$U_{GS} = U_T = \varphi_i + 2\varphi_F$$

co ilustruje rysunek z prawej strony. Napięcie φ_i jest w tym przypadku większe niż dla $U_{GS} = U_{FB}$ o wartość $-Q_B/C_i$. Zatem

$$U_T = U_{FB} - \frac{Q_B}{C_i} + 2\varphi_F$$

Uwzględniając zależność na U_{FB} otrzymuje się napięcie progowe w postaci

$$U_T = \varphi_{ms} - \frac{Q_{sr}}{C_i} + 2\varphi_F - \frac{Q_B}{C_i} \quad (6.64)$$

przy czym ładunek Q_B można wyznaczyć posługując się zależnością znaną z teorii złącza p-n

$$Q_B = (2\epsilon_s q N_D \cdot 2\varphi_F)^{1/2} \quad (6.65)$$

Na rysunku 6.25c przedstawiono sytuację dla $|U_{GS}| > |U_T|$. Ponieważ bardzo małe przyrosty potencjału powierzchniowego powodują duże przyrosty ładunku kanału (zależność wykładnicza), zatem — praktycznie biorąc — cały przyrost napięcia $|U_{GS}|$ powyżej wartości $|U_T|$ odkłada się na dielektryku, a φ_s pozostaje niemal stałe i równe $2\varphi_F$. Można zatem napisać proste wyrażenie na ładunek kanału

$$Q_C = C_g(U_{GS} - U_T) \quad (6.66)$$

przy czym $C_g = C_i ZL$ — pojemność kondensatora, którego okładki stanowią kanał i elektroda bramki.

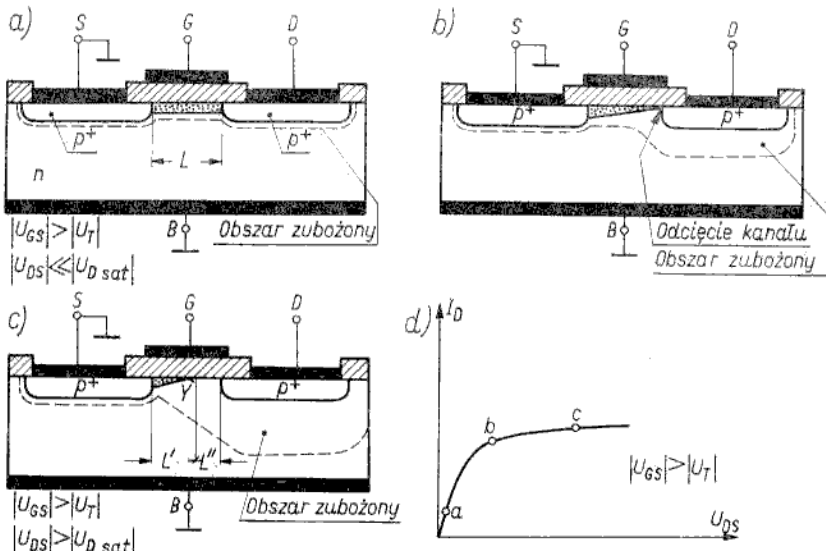
Ponieważ Q_C zależy liniowo od napięcia U_{GS} , prąd drenu jest również liniową funkcją napięcia U_{GS} (przy $U_{DS} \approx 0$).

Wyrażenie na charakterystykę przejściową dla dowolnych wartości napięcia U_{DS} otrzymamy w analizie ilościowej (p. 6.3.2).

Charakterystyka wyjściowa

6.3.13

Jeżeli napięcie drenu U_{DS} jest małe w porównaniu z napięciem bramki U_{GS} , to kanał spełnia funkcję rezystora liniowego, łączącego źródło z drenem (rys. 6.26a i odpowiadający mu punkt a na charakterystyce $I_D(U_{DS})$). W tym zakresie



Rys. 6.26

Schematyczne ilustracje stanów w tranzystorze MIS, objaśniające przebieg charakterystyki wyjściowej: a) zakres nienasyceń; b) odcięcie kanału; c) zakres nasyceń; d) charakterystyka $I_D(U_{DS})$

napięcie drenu zmiany prądu I_D w funkcji napięcia U_{DS} są liniowe. W miarę wzrostu ujemnej wartości U_{DS} zwiększa się wartość prądu I_D i na rezystancji kanału odkłada się znaczny spadek napięcia. Zatem — identycznie jak w tranzystorze PNFET — wzdłuż kanału występuje rozkład potencjału wzrastającego (co do wartości bezwzględnej) od źródła do drenu. Spadek napięcia w kanale powoduje zmniejszenie różnicy potencjałów między bramką a kanałem, tj. zmniejszenie natężenia pola elektrycznego, prostopadłego do powierzchni półprzewodnika. Ponieważ spadek napięcia w kanale zwiększa się w kierunku od źródła do drenu, zatem w tym samym kierunku zmniejsza się natężenie pola elektrycznego bramki, maleje więc również grubość kanału. To zjawisko powoduje modulację rezystancji kanału pod wpływem zmieniającego się napięcia drenu, czyli zostaje naruszona liniowa zależność prądu I_D w funkcji napięcia U_{DS} . Dalszy wzrost napięcia ujemnego U_{DS} prowadzi do całkowitego usunięcia inwersji w części kanału sąsiadującej z drenem. Taki stan nazywa się *odcięciem kanału*. Wartość napięcia U_{DS} , której odpowiada zanik inwersji w punkcie $Y = L$ (rys. 6.26b) i odpowiadający mu punkt b na charakterystyce $I_D(U_{DS})$, nosi nazwę *napięcia nasycenia* U_{Dsat} , przy czym

$$U_{Dsat} = U_{GS} - U_T \quad (6.67)$$

Dla zakresu napięć $|U_{DS}| \geq |U_{Dsat}|$ prąd drenu osiąga, praktycznie biorąc, wartość stałą (punkt c na charakterystyce $I_D(U_{DS})$). Wzrost napięcia ujemnego U_{DS} powyżej wartości U_{Dsat} powoduje rozszerzenie się obszaru zubożonego przy drenie i punkt odcięcia, oznaczony na rys. 6.26c literą Y , przesuwa się w stronę źródła. Jednakże spadek napięcia na części przewodzącej kanału (odcinek od źródła do punktu Y) pozostaje stały i równy wartości U_{Dsat} . W punkcie Y nośniki ładunku są wstrzykiwane do obszaru zubożonego, który spełnia podobną funkcję jak ładunek przestrzenny złącza baza-kolektor w tranzystorze bipolarnym. Dyskusja zjawisk zachodzących po odcięciu kanału jest równie trudna jak dla tranzystora PNFET. Jeżeli przyjmie się, że zwiększenie napięcia U_{DS} powoduje niewielkie zmiany szerokości obszaru zubożonego (odcinek od punktu Y do L), to prąd drenu — praktycznie biorąc — nie zmienia się. Tak jak dla tranzystora PNFET mówi się wówczas, że tranzystor MIS pracuje w zakresie nasycenia.

Wpływ napięcia podłoża, $U_{BS} \neq 0$

6.3.1.4

Spolaryzowanie podłoża w stosunku do źródła uziemionego powoduje modulację konduktancji kanału. Zatem podłoże może być uważane za swego rodzaju drugą bramkę w tranzystorze MIS. Wpływ polaryzacji podłoża na konduktancję kanału tłumaczy się następująco. Podłoże stanowi wraz ze źródłem złącze $n-p^+$ oraz z kanałem złącze $n-p$ (nie w sensie metalurgicznym, lecz fizycznym). Jeżeli to złącze spolaryzuje się zaporowo, to jego warstwa zaporowa (obszar zubożony) rozszerza się w głąb półprzewodnika, tj. do podłoża oraz do kanału. Powoduje to zarazem zmniejszenie ładunku nośników prądu w warstwie inwersyjnej, czyli zmniejszenie grubości kanału, tj. wzrost rezystancji kanału. Przy stałych napięciach U_{GS} , U_{DS} maleje prąd drenu, co odpowiada wzrostowi wartości bezwzględnej napięcia progowego. Można wykazać, że w przybliżeniu

$$U_T(U_{BS}) \approx U_T(U_{BS} = 0) + k \sqrt{|U_{BS}|} \quad (6.68)$$

przy czym współczynnik

$$k = \frac{\sqrt{2\varepsilon_s q N_D}}{C_i}$$

Stąd wynika, że wpływ polaryzacji elektrody podłoża jest tym mniejszy, im jest większa rezystywność podłoża (mniejsza koncentracja domieszek). Wraz ze wzrostem rezystywności podłoża warstwa ładunku przestrzennego złącza kanał-podłoże znacznie głębiej rozszerza się w stronę słabo domieszkowanego podłoża niż w stronę kanału. Wówczas modulacja grubości kanału jest słabsza.

Charakterystyki i parametry statyczne — analiza ilościowa

6.3.2

Rozpatrzmy statyczne charakterystyki prądowo-napięciowe oddzielnie dla zakresów nienasyceń i nasycenia. Przeprowadzimy uproszczoną analizę ładunkową, podobną jak dla tranzystora PNFET o szpilkowym rozkładzie domieszek w kanale.

Zakres nienasyceń ($0 \leq U_{DS} \leq U_{Dsat}$)

6.3.2.1

Punktem wyjścia jest zależność (6.19), którą przepisujemy jako słuszną również dla tranzystora MIS.

$$I_D = \frac{Q_C}{\tau_p} \quad (6.69)$$

W pierwszym przybliżeniu przyjmuje się, że natężenie pola elektrycznego nie zmienia się wzdłuż kanału, czyli nie zmienia się również szybkość unoszenia nośników. Stąd czas przelotu nośników przez kanał

$$\tau_p = \frac{L}{\mu_n E_y} = \frac{L^2}{\mu_n U_{DS}} \quad (6.70)$$

Najpierw rozpatrzmy ogólny przypadek polaryzacji źródła, bramki i drenu w stosunku do uziemionego podłoża ($U_S \neq 0$, $U_G \neq 0$, $U_D \neq 0$, $U_B = 0$). Jeżeli $U_S \neq U_D$ (inaczej $U_{DS} \neq 0$), to istnieje określony rozkład potencjału $U(y)$ wzdłuż kanału. Ładunek $Q_C^*(y)$ zaindukowany w kanale (w jednostce powierzchni) dla współrzędnej y

$$Q_C^*(y) = C_i [U_G - U_T - U(y)] \quad (6.71)$$

Założono, że potencjał $U(y)$ zmienia się liniowo, można zatem wprowadzić pojęcie średniego potencjału w kanale

$$\frac{U_S + U_D}{2}$$

Wówczas całkowity ładunek kanału

$$Q_C = C_g \left(U_G - U_T - \frac{U_S + U_D}{2} \right) \quad (6.72)$$

Po podstawieniu wyrażeń (6.70), (6.72) do podstawowej zależności (6.69) (przy $U_{DS} = U_D - U_S$) otrzymuje się

$$I_D = \beta \left[(U_G - U_T)(U_D - U_S) - \frac{U_D^2 - U_S^2}{2} \right] \quad (6.73)$$

przy czym posłużyliśmy się znanym z analizy tranzystora PNFET parametrem

$$\beta = \frac{\mu_n C_g}{L^2} = \frac{\mu_n Z C_i}{L} \quad (6.74)$$

Równanie (6.73) w przypadku uziemionego źródła ($U_S = 0$, $U_G = U_{GS}$, $U_D = U_{DS}$) przyjmuje postać identyczną jak dla tranzystora PNFET

$$I_D = \beta \left[(U_{GS} - U_T)U_{DS} - \frac{U_{DS}^2}{2} \right] \quad (6.75)$$

Zakres nasycenia ($|U_{DS}| > |U_{DSat}|$)

6.3.2.2

Analizując równanie (6.75) można wyznaczyć granicę między zakresem nienasyceń a zakresem nasycenia. Dla stałej wartości napięcia bramki U_{GS} wzrastające napięcie drenu U_{DS} powoduje coraz większy wkład członu $U_{DS}^2/2$, zatem przyrosty prądu drenu I_D są coraz mniejsze — charakterystyka przestaje być liniową, a prąd I_D dąży do maksymalnej wartości ustalonej. Dyskusja formalna wzoru (6.75) (przy założeniu, że jest on słuszny dla dowolnych wartości U_{DS}) doprowadziłaby do wniosku, że dla wystarczająco dużych napięć U_{DS} prąd I_D powinien się zmniejszać. Jest to oczywiście wniosek nieprawdziwy, gdyż w zakresie $|U_{DS}| > |U_{GS} - U_T|$ jest konieczny inny model tranzystora i inne równanie charakterystyki prądowo-napięciowej. Poszukując wartości napięcia U_{DS} , dla której prąd drenu osiąga maksimum, różniczkuje się równanie (6.75) względem U_{DS} i pierwszą pochodną przyrównuje do zera. Rozwiązanie w postaci wyrażenia

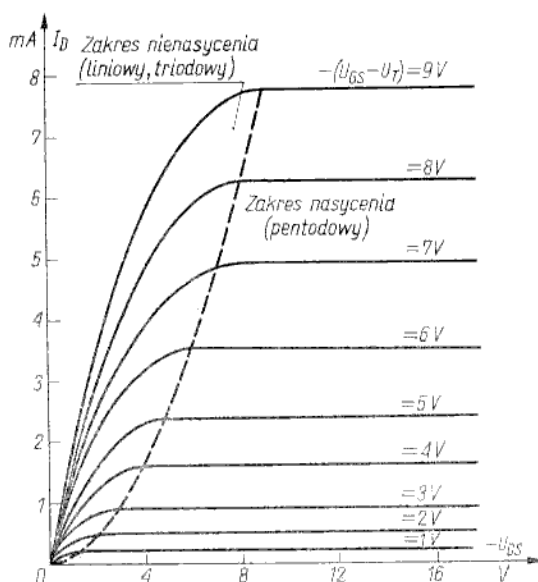
$$U_{DSat} = U_{GS} - U_T$$

jest zgodne ze wzorem (6.67), wyznaczonym na podstawie rozważań jakościowych. Napięcie U_{DSat} jest taką wartością napięcia drenu, przy której prąd I_D wchodzi w zakres nasycenia, czyli wyznacza ono granicę między zakresem nienasyceń a zakresem nasycenia. Zatem dla zakresu nasycenia należy do równania (6.75) podstawić wartość $U_{DS} = U_{DSat} = (U_{GS} - U_T)$. Stąd

$$I_D = \frac{\beta}{2} (U_{GS} - U_T)^2 \quad (6.76)$$

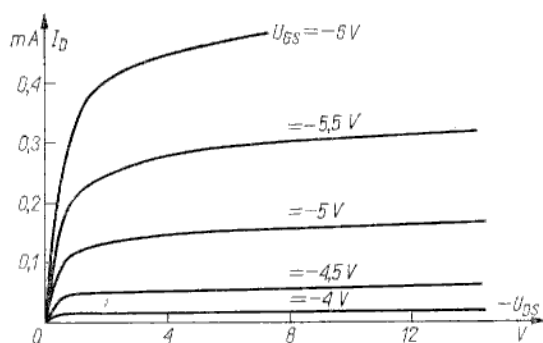
Z równania (6.76) wynika, że prąd drenu I_D w zakresie nasycenia nie zależy od napięcia U_{DS} , jest natomiast funkcją kwadratową napięcia bramki U_{GS} . Jest to wynik identyczny jak dla tranzystora PNFET ze szpilkowym rozkładem koncentracji domieszek w kanale.

Na rysunku 6.27 przedstawiono charakterystyki wyjściowe wyznaczone zgodnie z równaniami (6.75) i (6.76). Należy podkreślić, że te równania opisują tylko w pierwszym przybliżeniu rzeczywiste charakterystyki prądowo-napięciowe tranzystora MIS. W bardziej dokładnej analizie należałoby uwzględnić szereg pominiętych



Rys. 6.27
Teoretyczne charakterystyki wyjściowe tranzystora MIS

zjawisk, a przede wszystkim efekt skracania kanału w zakresie nasycenia, wpływ zmian ładunku podłoża w funkcji długości kanału, zależność ruchliwości od napięcia polaryzacji itd. Na przykład efekt skracania kanału, który można opisać podobną zależnością jak dla tranzystora PNFET (wzór (6.36)), powoduje, że prąd drenu



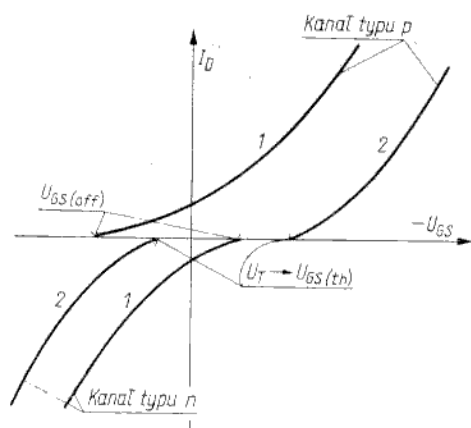
Rys. 6.28
Rzeczywiste charakterystyki wyjściowe dla tranzystora BWP 30

nieznacznie zmienia się w zakresie nasycenia. Jest to zgodne z rzeczywistym przebiegiem charakterystyk wyjściowych, co pokazano na rys. 6.28 na przykładzie charakterystyk tranzystora BWP 30 (produkcji polskiej).

Charakterystyki i oznaczenia czterech rodzajów tranzystora MIS

6.3.2.3

Różnice w przebiegu charakterystyk dla czterech rodzajów tranzystora MIS najbardziej obrazowo przedstawia zbiorczy wykres charakterystyk przejściowych w zakresie nasycenia, przedstawiony na rys. 6.29. W tabelicy 6.1 zestawiono symbole graficzne oraz charakterystyki przejściowe i wyjściowe. Symbole graficzne są kreślone z uwzględnieniem następujących zasad:



Rys. 6.29
 Charakterystyki przejściowe dla czterech rodzajów tranzystorów MIS
 1 — tranzystor z kanałem zubożanym;
 2 — tranzystor z kanałem wzbogacanym

Symbole graficzne i podstawowe charakterystyki czterech typów tranzystorów MIS

Tablica 6.1

Rodzaj tranzystora MIS	Symbol graficzny		Charakterystyka przejściowa	Charakterystyka wyjściowa
	I	II		
Kanał zubożany typu p				
Kanał wzbogacany typu p				
Kanał zubożany typu n				
Kanał wzbogacany typu n				

— w tranzystorach z kanałem wzbogacanym przy zerowym napięciu bramki prąd drenu jest, praktycznie biorąc, równy zero, dlatego odcinek między źródłem a drenem rysuje się linią przerywaną (obwód przerwany), natomiast w tranzystorach z kanałem zubożanym przy zerowym napięciu bramki przepływa prąd w kanale, dlatego odcinek źródło-dren rysuje się linią ciągłą;

- bramka tranzystora jest odizolowana od kanału, dlatego przedstawia się ją w postaci symbolu kąta prostego, którego jedno ramię jest równoległe do kanału, punkt zaś łączący dwa ramiona tego symbolu (wierzchołek kąta prostego) jest skierowany ku elektrodzie źródła;
- podłoże wraz z kanałem stanowi złącze p - n , dlatego przedstawia się je w postaci linii prostopadłej do kanału ze strzałką, której zwrot wskazuje, jaki jest typ kanału, co zarazem jednoznacznie definiuje typ podłoża.

Parametry statyczne

6.3.2.4

Najważniejszym parametrem statycznym jest napięcie progowe U_T , określone wzorem (6.64). W katalogach parametr ten jest podawany jako $U_{GS(th)}$ (niekiedy $U_{GS(RO)}$ dla tranzystorów z kanałem wzbogacanym oraz $U_{GS(off)}$ dla tranzystorów z kanałem zubożonym; napięcie $U_{GS(off)}$ przez analogię do tranzystorów PNFET jest często nazywane napięciem odcięcia). Napięcie U_T , niezależnie od rodzaju tranzystora, jest technicznie definiowane jako taka wartość napięcia U_{GS} , przy której prąd drenu osiąga określoną wartość, np. 1 lub 10 μ A (przy ustalonym napięciu U_{DS}).

Poza tym do ważnych parametrów statycznych podawanych w katalogach należy zaliczyć:

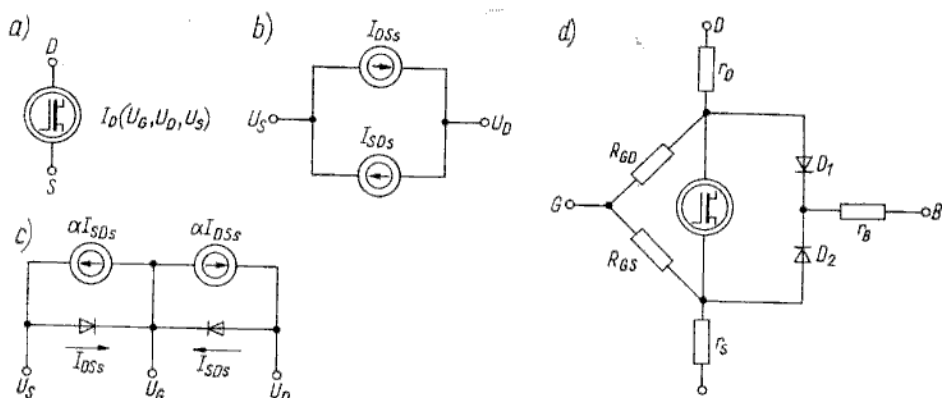
- BU_{GSS} — napięcie przebicia między bramką a podłożem przy zwartych elektrodach drenu i źródła, mierzone przy określonej wartości prądu $I_G^{(1)}$;
- BU_{DSS} — napięcie przebicia między drenem a źródłem przy zwarciu bramki ze źródłem ($U_{GS} = 0$): jest to przebicie lawinowe;
- I_{DSS} — prąd nasycenia tranzystora, gdy elektrody bramki i źródła są zwarte ($U_{GS} = 0$) oraz przy określonym napięciu U_{DS} ; w tranzystorach z kanałem zubożonym prąd I_{DSS} jest mierzony w zakresie nasycenia charakterystyki $I_D(U_{DS})$, natomiast w tranzystorach z kanałem wzbogacanym jest to prąd wsteczny złącza dren-podłoże;
- I_{GSS} — prąd upływu bramki w przypadku zwartych elektrod źródła i drenu ($U_{DS} = 0$) oraz dla określonego napięcia U_{GS} , zawierającego się w granicach 50...80% napięcia przebicia BU_{GSS} ;
- $r_{DS(ON)}$ — rezystancja statyczna między drenem a źródłem w warunkach maksymalnego prądu I_D ;
- $r_{DS(OFF)}$ — rezystancja statyczna między drenem a źródłem w warunkach wyłączonego kanału (dla $|U_{GS}| < |U_{GS(th)}|$ w przypadku tranzystora z kanałem wzbogacanym oraz $|U_{GS}| > |U_{GS(off)}|$ w przypadku tranzystora z kanałem zubożonym).

Modele statyczne

6.3.2.5

Najprostszy model statyczny tranzystora MIS przedstawiono na rys. 6.30a, na którym zależne źródło prądowe ma postać odpowiedniego równania charak-

¹⁾ Interesującą właściwością tranzystorów MIS jest możliwość przebicia dielektryka pod bramką i nieodwracalnego uszkodzenia tranzystora pod wpływem dotyku ręką do końcówki bramki. Jest to powodowane dużą rezystancją dielektryka (ok. $10^{14} \Omega$) i brakiem odpływu ładunku dostarczanego do bramki. Na przykład wystarczy dostarczyć do bramki ładunek 1 nC, by spowodować przebicie w tranzystorze o $C_g = 4$ pF i napięciu przebicia 250 V. Dlatego podczas manipulowania tranzystorem (przy montażu) należy przestrzegać określonych zasad — najczęściej ręka operatora jest uziemiana za pomocą bransolety nakładanej na przegub, a wszystkie końcówki tranzystora są zwierane.



Rys. 6.30

Cztery modele statyczne o różnym stopniu dokładności (zaznaczone kierunki prądów dotyczą przypadku kanału typu p)

terystyki prądowo-napięciowej (równania (6.73) lub (6.75) dla zakresu nienasyceń oraz (6.76) dla zakresu nasycenia). Równanie (6.73) stanowi tu model czterokońcówkowy (trzy końcówki S , G , D są sterowane względem czwartej B), natomiast (6.75) — tzw. model trójkońcówkowy (dwie końcówki G , D są sterowane względem zwartych ze sobą S , B). Prąd bramki jest praktycznie biorąc, równy zeru (zwykle poniżej 1 pA), co umożliwia przyjęcie, że elektroda bramki jest rozwartą.

Jeżeli korzysta się z modelu czterokońcówkowego, to wyrażenie (6.73) można przedstawić w postaci

$$I_D = I_{DSs} + I_{SDs} \quad (6.77)$$

przy czym:

$$I_{DSs} = \beta \left(U_G - U_T - \frac{U_D}{2} \right) U_D \quad (6.78)$$

$$I_{SDs} = -\beta \left(U_G - U_T - \frac{U_S}{2} \right) U_S \quad (6.79)$$

Sens fizyczny równań (6.78), (6.79) jest następujący. Prąd I_{DSs} jest prądem drenu przy źródle zwartym z podłożem, prąd I_{SDs} jest prądem źródła przy drenie zwartym z podłożem (zamiana funkcji źródła i drenu). Na podstawie (6.77), (6.78), (6.79) można przedstawić model w postaci dwu źródeł prądowych (rys. 6.30b). Koncepcja dwu źródeł prądowych umożliwi również zbudowanie modelu tranzystora MIS w pełni analogicznego do modelu Ebersa-Molla dla tranzystora bipolarnego. Taki pseudomodel Ebersa-Molla przedstawiono na rys. 6.30c, przy czym $\alpha = 1$, ponieważ prąd bramki jest równy zeru. Diody przedstawione na rys. 6.30c mają znaczenie abstrakcyjne, gdyż nie istnieją w sensie fizycznym w tranzystorze. Posługiwanie się pseudomodelem Ebersa-Molla ma tę zaletę, że umożliwia korzystanie z typowych programów obliczeń opracowanych dla tranzystorów bipolarnych.

Jeżeli w analizowanym układzie istotne znaczenie mają nawet niewielkie prądy bramki i podłoża, to należy korzystać z modelu przedstawionego na rys. 6.30d, na którym elementy R_{GS} , R_{GD} reprezentują upływności bramka-źródło i bramka-dren,

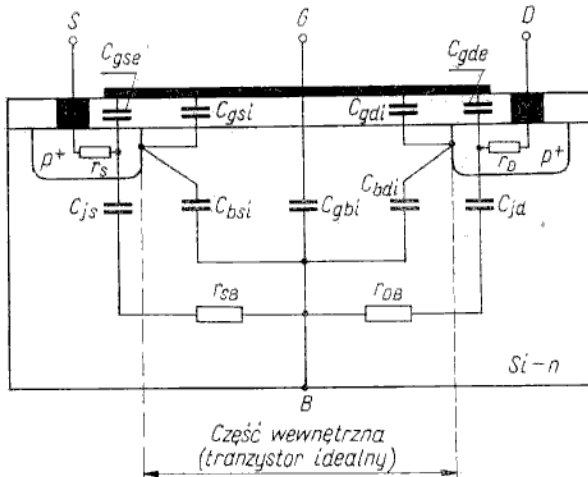
a diody D_1, D_2 reprezentują złącza p^+-n istniejące między źródłem a podłożem oraz między drenem a podłożem. W modelu tym uwzględniono również rezystancje szeregowe źródła r_s , drenu r_D i podłoża r_B .

Praca dynamiczna nieliniowa

6.3.3

Model dynamiczny nieliniowy tranzystora MIS otrzymuje się identycznie jak dla tranzystora PNFET przez dołączenie do modelu statycznego pojemności międzyelektrodowych.

Rozważania przeprowadzimy dwustopniowo. Najpierw rozpatrzmy część wewnętrzną tranzystora (tzw. tranzystor idealny) zgodnie z podziałem struktury tranzystora przedstawionym na rys. 6.31. Następnie uwzględnimy elementy reprezentujące część zewnętrzną, otrzymując w ten sposób model tzw. tranzystora rzeczywistego.

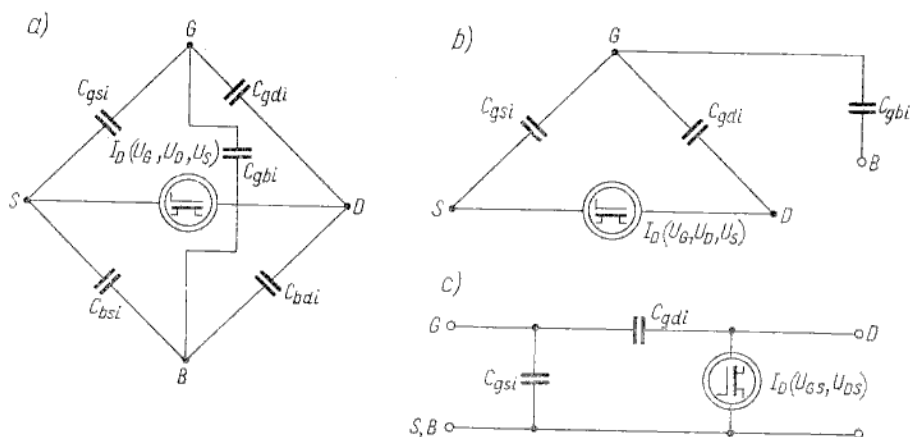


Rys. 6.31
Struktura tranzystora MIS, w której wydzielono część wewnętrzną oraz zaznaczono pojemności wewnętrzne i zewnętrzne

Tak jak dla tranzystora PNFET zakłada się, że zmiany ładunków w tranzystorze MIS (Q_G, Q_C, Q_B) są quasi-równowagowe, czyli wartości tych ładunków są funkcjami wartości chwilowych napięć na elektrodach, nie zależą zaś od szybkości zmian tych napięć. Szybkie zmiany napięć polaryzacji powodują przepływ prądów przesunięcia między poszczególnymi parami elektrod (prądy bramka-źródło, bramka-dren, podłoże-źródło itd.). W dokładnym schemacie zastępczym tranzystora idealnego należałoby zatem uwzględnić pięć pojemności ($C_{gs1}, C_{gd1}, C_{bs1}, C_{bd1}$ i C_{gb1}) — rys. 6.32a. W uproszczonej analizie uwzględnia się tylko pojemności C_{gs}, C_{gd}, C_{gb} , zdefiniowane następująco:

$$C_{gs} = \frac{\partial Q_G}{\partial U_{GS}}$$

(6.30)



Rys. 6.32

Modele dynamiczne nieliniowe idealnego tranzystora MIS: a) pełny model czterokońcówkowy; b) uproszczony model czterokońcówkowy; c) model trójkońcówkowy (źródło zwarte z podłożem)

$$C_{gdi} = \frac{\partial Q_G}{\partial U_{GD}} \quad (6.81)$$

$$C_{gbi} = \frac{\partial Q_G}{\partial U_{GB}} \quad (6.82)$$

Chwilowe prądy bramki i podłoża są zatem prądami przesunięcia:

$$i_G = C_{gsi} \frac{du_{GS}}{dt} + C_{gdi} \frac{du_{GD}}{dt} + C_{gbi} \frac{du_{GB}}{dt} \quad (6.83)$$

$$i_B = -C_{gbi} \frac{du_{GB}}{dt} \quad (6.84)$$

Chwilowe prądy źródła i drenu są równe sumie prądu stałego i prądów przesunięcia:

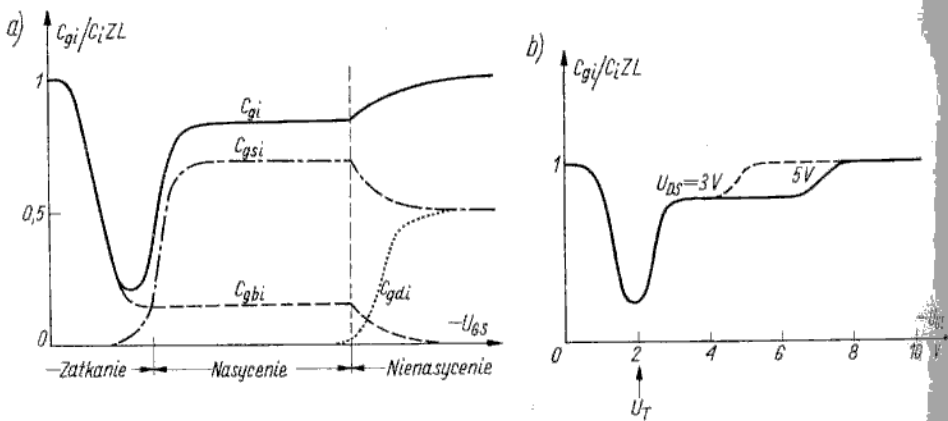
$$i_S = I_D - C_{gsi} \frac{du_{GS}}{dt} \quad (6.85)$$

$$i_D = I_D + C_{gdi} \frac{du_{GD}}{dt} \quad (6.86)$$

przy czym równanie prądów, uwzględniając ich rzeczywiste kierunki, ma postać

$$i_G + i_B = i_D - i_S \quad (6.87)$$

Na podstawie tych zależności można zaproponować model elektryczny tranzystora w postaci przedstawionej na rys. 6.32b lub 6.32c przy zwarceniu źródła z podłożem. Pojemności C_{gsi} , C_{gdi} , C_{gbi} są elementami nieliniowo zależnymi od napięć polaryzacji. Na rysunku 6.33a przedstawiono wykresy pojemności bramki i jej składowych w funkcji napięcia U_{GS} , a na rysunku 6.33b pokazano zmiany



Rys. 6.33
Zależność pojemności bramki od napięcia U_{GS} : a) pojemność C_{gi} i jej składowe; b) wpływ napięcia U_{DS} na przebieg charakterystyki $C_{gi}(U_{GS})$

krzywej $C_{gi}(U_{GS})$ dla różnych wartości napięcia U_{DS} . Pojemność bramki zdefiniowano następująco;

$$C_{gi} = C_{gsi} + C_{gbi} + C_{gdi} \quad (6.88)$$

Na wykresie przedstawionym na rys. 6.33a można wyróżnić trzy obszary odpowiadające pracy tranzystora w zakresie: zatkania, nasylenia, nienasylenia. W zakresie zatkania pojemność bramki tworzy tylko jedna składowa C_{gbi} , która zależy od napięcia w identyczny sposób jak pojemność kondensatora MIS w zakresie akumulacji i zubożenia. Gdy napięcie U_{GS} przekroczy wartość napięcia progowego U_T , pojawia się ładunek kanału, a wraz z nim pojawia się pojemność C_{gsi} , czyli pojemność C_{gi} wzrasta tak jak w kondensatorze MIS dla stanu inwersji. Jednak nie osiąga ona takiej samej wartości jak w stanie akumulacji (tak byłoby w kondensatorze MIS), gdyż dla $|U_{GS} - U_T| \leq |U_{DS}|$ tranzystor pracuje w zakresie nasylenia i kanał zwązając się na drodze od źródła do drenu nierównomiernie ekranuje głębsze warstwy półprzewodnika. W zakresie nienasylenia, tj. dla $|U_{GS}| > |U_{DS} + U_T|$ pojawia się pojemność C_{gdi} , gdyż przy drenie pojawia się ładunek kanału. W miarę dalszego wzrostu napięcia U_{GS} różnice w sterowaniu ładunku kanału napięciem źródła i drenu są coraz mniejsze

i pojemność C_{gi} dzieli się symetrycznie na składowe $C_{gsi} \approx C_{gdi} \approx \frac{1}{2} C_{ZL}$.

natomiast pojemność C_{gbi} znika wskutek silnego ekranowania głębszych warstw półprzewodnika przez kanał.

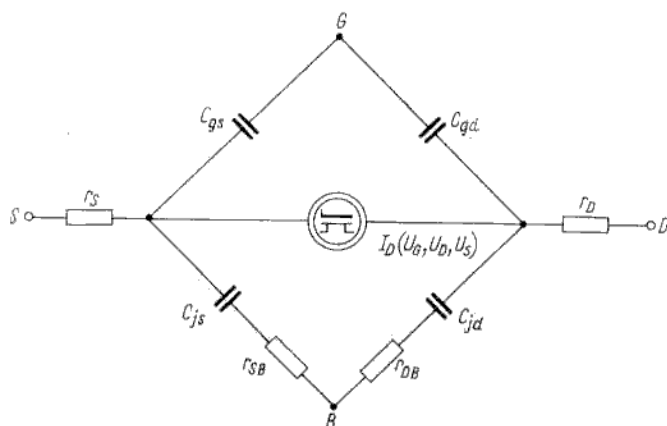
W uproszczonej analizie można zastosować następujące wyrażenia analityczne, opisujące w sposób przybliżony poszczególne pojemności:

$$\text{Zakres zatkania} \quad \left\{ \begin{array}{l} C_{gsi} = 0 \\ C_{gdi} = 0 \\ C_{gbi} = \frac{ZL\epsilon_i}{d_i} \end{array} \right. \quad (6.89)$$

$$\begin{array}{l} \text{Zakres nienasycecia} \\ |U_{GS}| > |U_T| \\ |U_{DS}| \leq |U_{Dsat}| \end{array} \left\{ \begin{array}{l} C_{gsi} = \frac{ZL\epsilon_i}{2d_i} \left(1 + \frac{1}{3} \frac{U_{DS}}{U_{Dsat}} \right) \\ C_{gdi} = \frac{ZL\epsilon_i}{2d_i} \left(1 - \frac{U_{DS}}{U_{Dsat}} \right) \\ C_{gbi} = 0 \end{array} \right. \quad (6.90)$$

$$\begin{array}{l} \text{Zakres nasycenia} \\ |U_{GS}| > |U_T| \\ |U_{DS}| > |U_{Dsat}| \end{array} \left\{ \begin{array}{l} C_{gsi} = \frac{2}{3} \frac{ZL\epsilon_i}{d_i} \\ C_{gdi} = 0 \\ C_{gbi} = 0 \end{array} \right. \quad (6.91)$$

Aby otrzymać schemat zastępczy tranzystora rzeczywistego (rys. 6.34) należy uzupełnić dotychczas rozpatrywany schemat elementami zewnętrznymi, pokazanymi na rys. 6.31. Są to następujące elementy:



Rys. 6.34

Model dynamiczny nieliniowy rzeczywistego tranzystora MIS (pominięto elementy C_{gbi} , C_{bsi} , C_{bdi} , natomiast $C_{gs} = C_{gsi} + C_{gse}$, $C_{gd} = C_{gdi} + C_{gde}$)

- pojemności C_{gse} , C_{gde} , wynikające z częściowego pokrywania obszarów p^+ źródła i drenu przez metalizację bramki;
- pojemności złączy p - n źródła i drenu C_{js} , C_{jd} ;
- rezystancje rozproszone podłoża, tj. rezystancja podłoże-źródło r_{SB} oraz podłoże-dren r_{DB} ;
- rezystancje szeregowo źródła i drenu r_S , r_D .

Pojemności C_{gse} , C_{gde} można łatwo obliczyć, jeśli są znane odpowiednie wymiary geometryczne. Wartości tych pojemności są wyznaczone przez grubość dielektryka, jego przenikalność elektryczną i powierzchnie przekrycia bramki z obszarami p^+ źródła i drenu. Dla typowej grubości warstwy SiO_2 , wynoszącej 120 nm jej pojemność na jednostkę powierzchni wynosi 350 $\mu\text{F}/\text{m}^2$. Zwykle pojemności C_{gse} , C_{gde} mają wartości mniejsze niż 3 pF. Pojemności złączy C_{js} , C_{jd} są zależne od napięcia według wzoru znanego z teorii złącza p - n

$$\frac{C_j}{C_{j0}} = \left(1 + \frac{U}{\varphi_B}\right)^{-1/n} \quad 2 \leq n \leq 3 \quad (6.92)$$

przy czym: C_{j0} — pojemność przy zerowej polaryzacji złącza; U — napięcie polaryzacji złącza; φ_B — napięcie dyfuzyjne (typowe wartości dla krzemu 0,6...1 V).

Rezystory r_S, r_D, r_{SB}, r_{DB} mają wartości rzędu kilkudziesięciu omów.

Praca z małymi sygnałami

6.3.4

Wszystkie uwagi wstępne, podane w p. 6.2.4, w odniesieniu do tranzystora PNFET, zachowują słuszność również dla tranzystora MIS.

Rozpatrzmy modele fizyczne dla małych i dużych częstotliwości, przy czym rozważania ograniczymy do przypadku modeli trójkońcówkowych (źródło zwarte z podłożem).

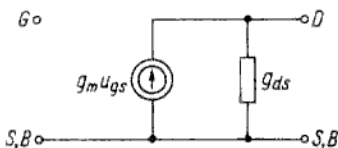
Model dla małych częstotliwości

6.3.4.1

Przechodząc od modelu wielkosygnałowego do modelu małosygnałowego należy rozłożyć w szereg Taylora funkcję $I_D(U_{GS}, U_{DS})$ i uwzględnić człony rzędu pierwszego przy określonym punkcie pracy (określone stałe napięcia polaryzacji)

$$i_d = \frac{\partial I_D}{\partial U_{GS}} u_{gs} + \frac{\partial I_D}{\partial U_{DS}} u_{ds}$$

Jest to równanie identyczne z (6.46), a ponieważ postać analityczna charakterystyki prądowo-napięciowej tranzystora MIS jest również identyczna z charakterystyką tranzystora PNFET o szpilkowym rozkładzie domieszek, pozostają zatem słuszne wszystkie definicje i zależności podane w p. 6.2.4.1. Schemat zastępczy tranzystora MIS dla małych częstotliwości przedstawiono na rys. 6.35.



Rys. 6.35

Schemat zastępczy tranzystora MIS dla małych sygnałów małej częstotliwości

Modele dla dużych częstotliwości

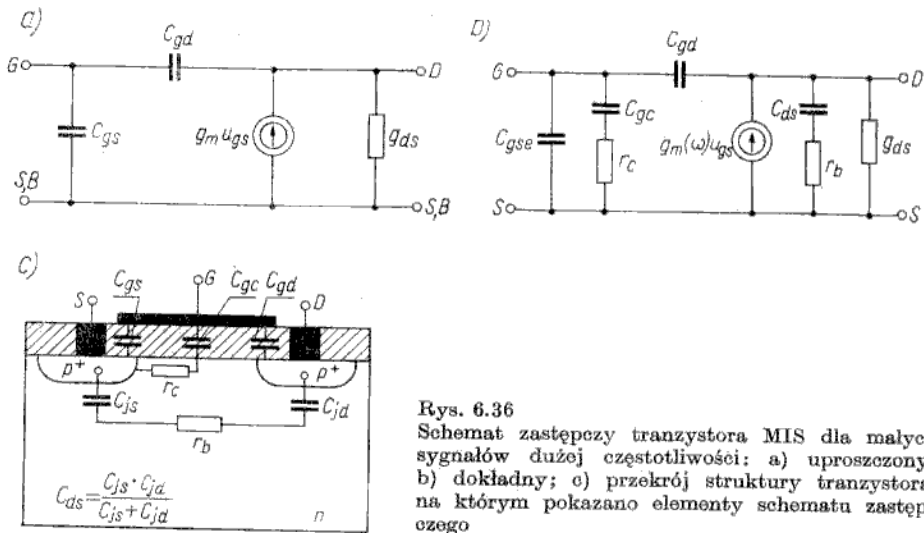
6.3.4.2

Podstawowy model dla dużych częstotliwości przedstawiono na rys. 6.36a, przy czym:

$$C_{gs} = C_{gsi} + C_{gbi} + C_{gse} \quad (6.93)$$

$$C_{gd} = C_{gdi} + C_{gde} \quad (6.94)$$

Komentarz do tego modelu jest identyczny jak w p. 6.2.4.2. Jest to model zadowalająco dokładny w zakresie do kilkudziesięciu magaherców.



Rys. 6.36 Schemat zastępczy tranzystora MIS dla małych sygnałów dużej częstotliwości: a) uproszczony; b) dokładny; c) przekrój struktury tranzystora, na którym pokazano elementy schematu zastępczego

W zakresie bardzo dużych częstotliwości stosuje się model przedstawiony na rys. 6.36b.

W celu fizycznej interpretacji tego modelu zamieszczono ilustrację schematyczną przekroju tranzystora MIS (rys. 6.36c) ze wskazaniem umiejscowienia poszczególnych elementów tego modelu.

Częstotliwość graniczną idealnego tranzystora MIS wyznacza się w taki sam sposób jak dla tranzystora PNFET

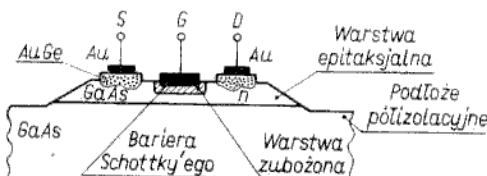
$$f_{max} = \frac{g_m}{2\pi C_{gs}} = \frac{\mu_n |U_{GS} - U_T|}{2\pi L^2} \quad (6.95)$$

W tranzystorze rzeczywistym częstotliwość graniczna jest mniejsza, niż wynika z tej zależności, gdy istotny wpływ mają pojemności zewnętrzne (C_{gse} , C_{gde}).

Tranzystory MESFET

6.4

Tranzystory MESFET z arsenku galu, produkowane na świecie od 1968 r.¹⁾, w ostatnich kilku latach zdobyły trwałą pozycję jako elementy wzmacniające w zakresie częstotliwości kilku do kilkunastu gigaherców. Schematyczny przekrój tranzystora MESFET przedstawiono na rys. 6.37. Na półizolacyjnym podłożu (prawie samoistnym arsenku galu) jest wytwarzana warstwa epitaksjalna



Rys. 6.37 Schematyczna ilustracja budowy tranzystora MESFET z arsenku galu

¹⁾ Pierwszym wytwórcą tych tranzystorów była firma IBM, a obecnie są one wytwarzane głównie w firmach: IBM, Hewlett Packard, Plessey, Watkins-Johnson.

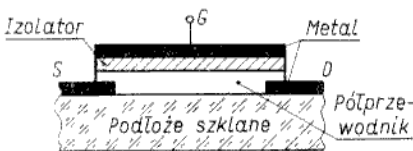
GaAs typu n . Po trawieniu, którego celem jest wydzielenie wysp GaAs typu n na podłożu półizolacyjnym, wykonuje się dwa kontakty omowe (Au-AuGe-GaAs) oraz bramkę (Au). Styk bramki metalowej z arsenkiem galu ma właściwości złącza z barierą Schottky'ego. Sposób działania tego tranzystora polowego jest taki sam jak sposób działania tranzystora PNFET. Warstwa półprzewodnika między bramką a podłożem spełnia funkcję kanału, którego przekrój zależy od szerokości warstwy zubożonej złącza m - s . Przy polaryzacji złącza bramka-kanał w kierunku zaporowym obszar zubożony rozszerza się w głąb warstwy epitaksjalnej typu n , powodując zmniejszanie efektywnego przekroju kanału. Zmiany napięcia bramki powodują zatem zmiany rezystancji kanału, czyli zmiany prądu w obwodzie źródło-dren. Ponieważ długość bramki jest bardzo mała ($0,5 \dots 2 \mu\text{m}$) i ruchliwość elektronów w kanale jest bardzo duża (w litym GaAs $\mu_n \approx 0,8 \text{ m}^2/(\text{V} \cdot \text{s})$, a w warstwie epitaksjalnej jest ok. 2-krotnie mniejsza), częstotliwość maksymalna tranzystorów MESFET z GaAs może być bardzo duża ($f_m \approx 50 \text{ GHz}$). Jest to podstawową zaletą tych tranzystorów. Ponadto odznaczają się one bardzo małymi szumami (ok. 2 do 3 razy mniejszymi niż dla tranzystorów bipolarnych przy identycznych częstotliwościach pracy) i możliwością rozpraszania dość dużej mocy. Iloczyn maksymalnej mocy rozpraszanej przez częstotliwość graniczną wynosi $10 \dots 20 \text{ W} \cdot \text{GHz}$. Typowe parametry tranzystorów MESFET są następujące:

- współczynnik szumów $2 \dots 4 \text{ dB}$ w zakresie częstotliwości $2 \dots 10 \text{ GHz}$;
- moc rozpraszana $1 \dots 2 \text{ W}$ przy częstotliwości $8 \dots 10 \text{ GHz}$ oraz $150 \dots 300 \text{ mW}$ przy częstotliwości 20 GHz .

Tranzystory cienkowarstwowe

6.5

Tranzystory cienkowarstwowe (tzw. tranzystory TFT) wykonał po raz pierwszy Weimer w 1961 r.¹⁾. Budowę tych tranzystorów dla jednego spośród kilku możliwych wariantów technologii przedstawiono schematycznie na rys. 6.38.



Rys. 6.38

Schematyczna ilustracja budowy tranzystora cienkowarstwowego

W pierwszej operacji technologicznej naparowuje się w próżni cienką warstwę metalu (aluminium lub złoto) na podłożu szklanym. Po wytrawieniu lokalnym tej warstwy z zastosowaniem procesu fotolitografii, powstają elektrody źródła i drenu. Drugą operacją jest naniesienie próżniowe cienkiej warstwy półprzewodnika (o grubości kilku do kilkudziesięciu nanometrów), którym jest najczęściej siarczek kadmu (CdS), selenek kadmu (CdSe) lub tellur (Te). Istotne znaczenie — niekorzystne dla właściwości tranzystora — ma fakt, że struktura półprzewodnika jest polikrystaliczna. W następnej operacji nanosi się warstwę dielektryka (najczęściej tlenku krzemu lub tlenku aluminium), a na końcu naparowuje się metalową elektrodę bramki (zwykle Al). Zasada działania i zjawiska fizyczne za-

¹⁾ Weimer, P. K.: An evaporated thin-film triode. *IRE Trans. Electron Dev.*, 1961, Vol. Ed-8, p. 421.

Tablica 6.2

Wykaz ważniejszych wzorów	PNFET		MISFET
	kanal „równomierny”	„kanal szpilkowy”	
Charakterystyki i parametry	<p>Rodzaj tranzystora</p> <p>kanal „równomierny”</p>	<p>„kanal szpilkowy”</p>	<p>MISFET</p>
Zakres nasycenia	$I_D = G_{Dso} \left\{ U_{Ds} - \frac{2}{3} \frac{\alpha}{a} [\varphi_B - U_{Gs} + U_{Ds} ^{3/2} - \varphi_B - U_{Gs} ^{3/2}] \right\}$ <p>przy czym</p> $G_{Dso} = \sigma_c \frac{2aZ}{L}$	$I_D = \beta \left[(U_{Gs} - U_p) U_{Ds} - \frac{U_{Ds}^2}{2} \right]$ <p>przy czym</p> $\beta = \frac{\mu_n C_g}{L^2}$	$I_D = \beta \left[(U_{Gs} - U_T) U_{Ds} - \frac{U_{Ds}^2}{2} \right]$ <p>przy czym</p> $\beta = \frac{\mu_n C_1}{L^2} = \frac{\mu_n Z C_1}{L}$
Charakterystyki statyczne	<p>Zakres nasycenia</p> $I_{Dsat} = I_{DSS} \left[1 - 3 \frac{U_{Gs}}{U_p} + 2 \left(\frac{U_{Gs}}{U_p} \right)^{3/2} \right]$	$I_{Dsat} = I_{DSS} \left(1 - \frac{U_{Gs}}{U_p} \right)^2$ <p>lub w innej postaci</p> $I_{Dsat} = \frac{\beta}{2} (U_{Gs} - U_p)^2$	$I_{Dsat} = \frac{\beta}{2} (U_{Gs} - U_T)^2$
Napięcie odcięcia (zatkania)	$U_p = - \frac{qN_D a^2}{2\epsilon_s} + \varphi_B$	$U_p = - \frac{qN_D c u}{\epsilon_s}$	$U_T = \varphi_{ms} - \frac{Q_{sr}}{C_1} + 2q_F - \frac{Q_B}{C_1}$ <p>przy czym</p> $Q_B = 2\epsilon_s q N_D \varphi_F$
Parametry statyczne	$I_{DSS} = G_{Dso} \frac{ U_p }{3}$	$I_{DSS} = \frac{ U_p }{2} G_{Dso}$	$f_{max} = \frac{g_m}{2\pi C_{gs}} = \frac{\mu_n U_{Gs} - U_T }{2\pi L^2}$
Prąd nasycenia dla $U_{Gs} = 0$	$I_{DSS} = G_{Dso} \frac{ U_p }{3}$	$I_{DSS} = \frac{ U_p }{2} G_{Dso}$	$f_{max} = \frac{g_m}{2\pi C_{gs}} = \frac{\mu_n U_{Gs} - U_T }{2\pi L^2}$
Chętność graniczna	$f_{max} = \frac{g_m}{2\pi C_g} = \frac{\mu_n U_{Gs} - U_p }{2\pi L^2}$	$f_{max} = \frac{g_m}{2\pi C_g} = \frac{\mu_n U_{Gs} - U_p }{2\pi L^2}$	$f_{max} = \frac{g_m}{2\pi C_{gs}} = \frac{\mu_n U_{Gs} - U_T }{2\pi L^2}$

chodzące w tranzystorach cienkowarstwowych są podobne jak w tranzystorach MIS. Ponieważ tranzystory MIS i cienkowarstwowe mają wiele cech wspólnych (podstawowa struktura tranzystora cienkowarstwowego jest również typu metal-izolator-półprzewodnik), odróżnienie tych dwu grup tranzystorów jest niekiedy dość trudne. W typowy sposób ilustruje te trudności przypadek tranzystora MIS na podłożu szafirowym (tzw. technologia SOS rozpatrywana w rozdziale 8). Niekiedy przyjmuje się za kryterium podziału grubość warstwy półprzewodnika, tj. do tranzystorów MIS zalicza się te tranzystory IGFET, w których część aktywna jest wykonana w grubym ($100 \dots 200 \mu\text{m}$) podłożu półprzewodnikowym, natomiast do tranzystorów cienkowarstwowych — te tranzystory IGFET, w których część aktywna jest wykonana w cienkiej warstwie półprzewodnika osadzonej na podłożu izolacyjnym. Kierując się tym kryterium tranzystory MIS na podłożu szafirowym można zaliczyć do cienkowarstwowych. Jednak słusznierzym kryterium podziału jest spójność technologiczna metody wytwarzania tranzystorów z typową technologią wytwarzania układów scalonych półprzewodnikowych lub układów scalonych cienkowarstwowych (patrz rozdział 8). Układy cienkowarstwowe są wytwarzane wyłącznie metodami osadzania próżniowego warstw, w układach półprzewodnikowych natomiast oprócz osadzania próżniowego są stosowane procesy dyfuzji, implantacji, epitaksji i utleniania termicznego. Dlatego tranzystory IGFET, w których cienka warstwa monokrystaliczna półprzewodnika jest wytwarzana metodą epitaksji na izolacyjnym podłożu szafirowym (technologia SOS), zaliczamy do grupy tranzystorów MIS. Należy podkreślić, że tranzystory cienkowarstwowe nie mają żadnego znaczenia jako elementy indywidualne, lecz były opracowywane z myślą o ich zastosowaniu jako elementów aktywnych w układach scalonych cienkowarstwowych, wytwarzanych w jednolitym procesie technologicznym. Jednakże, mimo prac prowadzonych w tym kierunku od kilkunastu lat, tranzystory cienkowarstwowe nie spełniły dotychczas wymagań koniecznych do ich zastosowania w produkcji układów scalonych. Główną przeszkodą w zastosowaniu tych tranzystorów jest duża niestabilność ich parametrów w funkcji czasu.

ZADANIA

6.1. Wyznacz napięcie odcięcia krzemowego tranzystora PNFET, w którym szerokość metalurgiczna kanału $2a = 10 \mu\text{m}$, a koncentracja domieszek w kanale $N_D = 10^{21} \text{ m}^{-3}$.

Wzór: Korzystamy z aproksymacji kanału „równomiernego” — (6.37)

Dane liczbowe: $q = 1,6 \cdot 10^{-19} \text{ C}$; $\epsilon_s = 12 \cdot 8,85 \cdot 10^{-12} \text{ F/m}$; $\varphi_B = 0,7 \text{ V}$

Odpowiedź: 20 V

6.2. Wyznacz napięcie odcięcia krzemowego tranzystora PNFET, w którym $2a = 2 \mu\text{m}$, $N_D = 10^{22} \text{ m}^{-3}$.

Wzór: Korzystamy z aproksymacji kanału „szpilkowego” — (6.38), przy czym przyjmujemy

$$c = \frac{1}{5} a$$

Dane liczbowe: Jak w poprzednim zadaniu

Odpowiedź: $3,2 \text{ V}$

6.3. Oblicz konduktancję kanału G_{DSO} dla tranzystora z zadania 6.1 przy znanych wartościach Z , L .

Wzór: (6.13)

Dane liczbowe: $Z = 100 \mu\text{m}$, $L = 10 \mu\text{m}$, $\mu_n = 0,13 \text{ V/m}^2 \cdot \text{s}$

Odpowiedź: $2 \cdot 10^{-3} \text{ S}$

6.4. Oblicz konduktancję kanału G_{DSO} dla tranzystora z zadania 6.2, przy znanej długości kanału L i pojemności bramki C_g .

Wzór: (6.29)

Dane liczbowe: $L = 10 \mu\text{m}$, $C_g = 5 \text{ pF}$

Odpowiedź: $G_{DS0} = 2 \cdot 10^{-2} \text{ S}$

6.5. Wyznacz prądy I_{DSS} dla dwóch tranzystorów rozpatrywanych w poprzednich zadaniach.

Wzory: (6.32) oraz (6.34)

Odpowiedź: 13 mA oraz 32 mA

6.6. Wyznacz częstotliwość graniczną krzemowego tranzystora MOS o długości kanału $L = 5 \mu\text{m}$, napięciu progowym $U_T = 5 \text{ V}$, przy polaryzacji bramki $U_{GS} = 10 \text{ V}$.

Wzór: (6.95)

Odpowiedź: 4 GHz

W poprzednich dwóch rozdziałach poznaliśmy podstawowe zasady działania, parametry i charakterystyki tranzystorów bipolarnych i unipolarnych. Rozpatrywaliśmy w tych rozdziałach właściwości tranzystorów przy narzuceniu ściśle określonych wyidealizowanych warunków na parametry obwodów, w których tranzystory były włączone (na przykład parametry typu h były wyznaczone przy zwarciu wyjścia i rozwarciu wejścia). W niniejszym rozdziale zajmujemy się opisem właściwości prostych układów analogowych z uwzględnieniem zarówno parametrów tranzystora, jak i dowolnych parametrów elementów połączonych z tranzystorem. Na przykład w przypadku wzmacniacza jednostopniowego interesujący jest zarówno sposób polaryzacji tranzystora jak i wzmocnienie uzyskiwane przy dowolnych wartościach rezystancji źródła sygnału oraz obciążenia. Będziemy zatem wyznaczać parametry robocze tranzystorów pracujących w rzeczywistych układach.

Obszary zastosowań obu klas tranzystorów (bipolarnych i polowych) w układach analogowych nie są jednakowe z uwagi na różnice w ich właściwościach. Jednoznaczne porównanie właściwości tranzystorów bipolarnych i polowych, pracujących jako wzmacniacze małych sygnałów, jest trudne, gdyż różne są sposoby sterowania (prądowe w tranzystorach bipolarnych, a napięciowe w polowych), różne są zestawy podstawowych parametrów (na przykład podstawowym parametrem tranzystora bipolarnego jest wzmocnienie prądowe, które jest pojęciem nieprzydatnym dla tranzystorów polowych). Najbardziej racjonalne wydaje się porównanie transkonduktancji przy podobnych warunkach polaryzacji (na przykład przy jednakowych prądach kolektora i drenu).

Dla tranzystora bipolarnego

$$g_m \approx \frac{I_E}{\varphi_T} \tag{7.1}$$

Na przykład dla $I_E = 25 \text{ mA}$, przy $T = 300 \text{ K}$ ($\varphi_T = 25 \text{ mV}$)

$$g_m = 1000 \text{ mA/V}$$

Z kolei dla tranzystora polowego

$$g_m = \beta |U_{GS} - U_T| = \frac{2I_D}{|U_{GS} - U_T|} \tag{7.2}$$

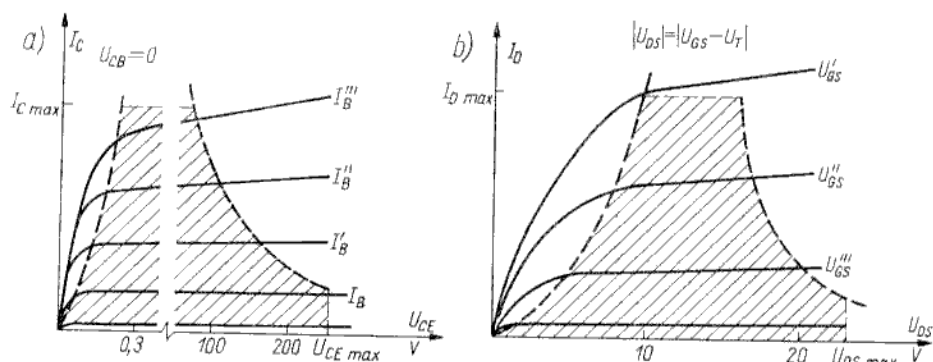
Na przykład dla $\beta = 0,5 \text{ mA/V}^2$ ¹⁾ prąd $I_D = 25 \text{ mA}$ uzyskuje się przy $|U_{GS} - U_T| = 10 \text{ V}$; zatem

$$g_m = 5 \text{ mA/V}$$

¹⁾ Transkonduktancja tranzystora bipolarnego nie zależy od parametrów jego struktury, w przypadku tranzystora polowego natomiast jest zależna od współczynnika β , tj. od roz-

Transkonduktancja tranzystora bipolarnego jest więc kilkaset razy większa niż dla tranzystora polowego.

Interesujące jest porównanie dozwolonych obszarów pracy obu rodzajów tranzystorów we wzmacniaczach liniowych. Na rysunku 7.1 przedstawiono charakterystyki



Rys. 7.1

Porównanie dozwolonych obszarów pracy dla tranzystorów bipolarnych (a) i polowych (b)

wyjściowe tranzystorów bipolarnego i polowego, na których zakreskowano dozwolone obszary pracy. Napięcie nasycenia ograniczające obszar pracy z lewej strony jest kilkanaście razy mniejsze w tranzystorach bipolarnych (dziesiąte części wolta w porównaniu z kilkoma voltami w tranzystorach polowych). Napięcie maksymalne, ograniczające dozwolony obszar pracy z prawej strony, wynosi do 2 kV w tranzystorach bipolarnych oraz kilkaset voltów w tranzystorach polowych. Pod tym względem zatem tranzystory polowe również ustępują bipolarnym. Trzy wyżej przedstawione zalety tranzystorów bipolarnych, tj.:

- duża wartość transkonduktancji,
 - małe napięcie nasycenia,
 - duże napięcie maksymalne,
- sprawiają, że te tranzystory są bezkonkurencyjne w zastosowaniach konwencjonalnych.

Jednakże tranzystory polowe mają również kilka istotnych zalet, jak na przykład:

- bardzo duża impedancja wejściowa;
- małe szумы (szczególnie PNFET);
- kwadratowy przebieg charakterystyki przejściowej $I_D(U_{GS})$, co umożliwia uniknięcie zniekształceń zawierających nieparzyste harmoniczne;
- możliwość zastosowania tranzystora jako obciążenia rezystancyjnego oraz rezystora sterowanego.

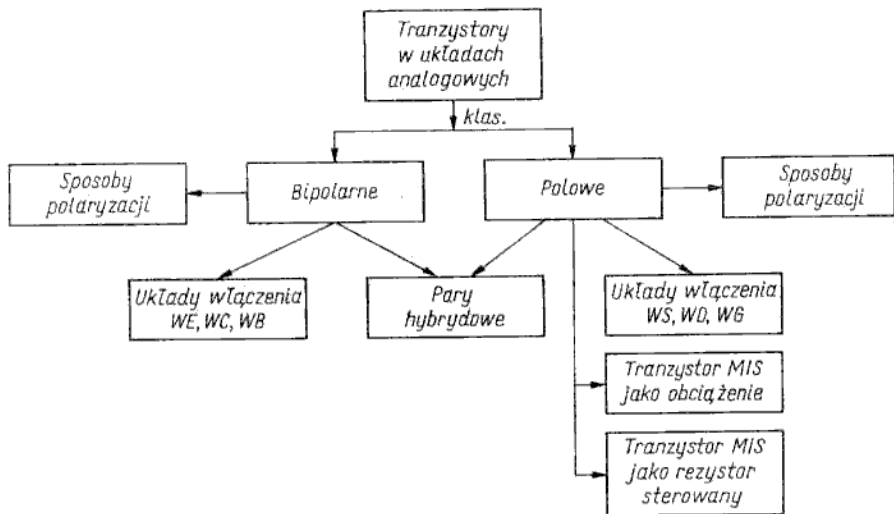
Te i inne zalety sprawiają, że tranzystory polowe znajdują szereg specyficznych zastosowań, między innymi w:

miarów geometrycznych struktury tego tranzystora, przenikalności elektrycznej warstwy dielektryka i ruchliwości nośników. Porównanie transkonduktancji nie miałyby sensu dla tranzystorów o istotnie różnych rozmiarach geometrycznych ich struktur (na przykład porównanie tranzystora bipolarnego małej mocy z tranzystorem polowym dużej mocy). Można przyjąć, że tranzystory są porównywalne, jeżeli mają jednakowe wartości maksymalnych prądów ($I_{Cmax} \approx I_{Dmax}$). Jeżeli przyjmie się, że dla tranzystora bipolarnego $I_E = 25$ mA jest prądem maksymalnym, to tranzystor polowy o maksymalnym prądzie drenu $I_D = 25$ mA powinien mieć współczynnik $\beta \approx 0,5$ mA/V².

- elektrometrach,
- głowicach UKF,
- filtrach i innych układach z tranzystorami polowymi jako rezystorami sterowanymi.

W niektórych przypadkach optymalnym rozwiązaniem układowym jest połączenie tranzystorów bipolarnych z polowymi.

W kolejnych punktach tego rozdziału przedstawimy przede wszystkim sposoby polaryzacji tranzystorów i ich parametry robocze w podstawowych układach włączenia. Rozpatrzmy również niektóre zagadnienia specyficzne dla układów z tranzystorami MIS, tj. zastosowanie tranzystora MIS jako obciążenie oraz jako rezystor sterowany napięciowo. Układ rozdziału przedstawiono schematycznie na rys. 7.2.



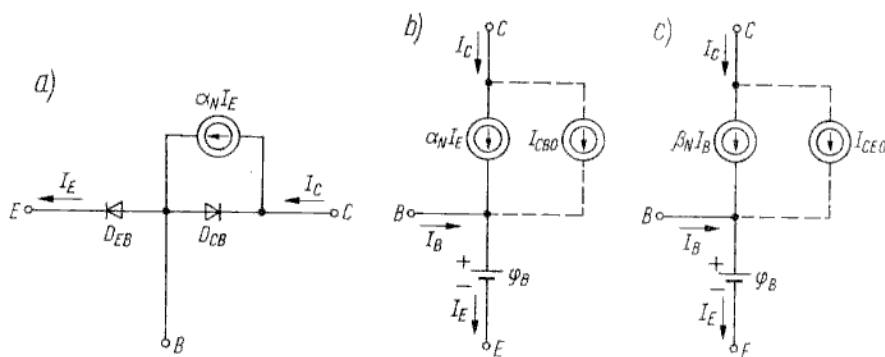
Rys. 7.2
Schematyczny układ rozdziału

Sposoby polaryzacji tranzystorów bipolarnych

7.1

Dla prawidłowej pracy tranzystora w określonym układzie konieczne jest ustalenie punktu pracy, tj. spolaryzowania tranzystora odpowiednimi wartościami prądów i napięć stałych. Konstruktor układu wybiera punkty pracy zapewniające najlepsze wykorzystanie tranzystorów (na przykład największe wzmocnienie, najmniejsze zniekształcenia dla sygnału o określonej amplitudzie itp.).

Interesuje nas tylko praca tranzystora w układzie wzmacniacza małych sygnałów zmiennych, czyli przy polaryzacji złącza emiter-baza w kierunku przewodzenia, a złącza baza-kolektor w kierunku zaporowym. Do analizy stałoprądowych warunków pracy tranzystora w tym zakresie służy model Ebersa-Molla, przedstawiony na rys. 7.3a (identyczny z modelem rozpatrywanym w p. 5.7.1 — rys. 5.26). Do przybliżonych analiz, zamieszczonych w tym punkcie, wystarczająco dokładne są modele uproszczone (rys. 7.3b, c), które otrzymuje się zastępując diodę D_{EB} źródłem napięciowym φ_B (dla tranzystora krzemowego $\varphi_B \approx 0,7$ V, dla germanowego zaś $\approx 0,3$ V), a diodę D_{CB} — źródłem prądowym I_{cno} .



Rys. 7.3

Modele tranzystorów bipolarnych, stosowane w analizie stałoprądowej: a) model Ebersa-Molla dla zakresu aktywnego normalnego; b), c) dwie równoważne wersje uproszczonego modelu Ebersa-Molla

lub I_{CEO} . Źródła prądowe I_{CBO} , I_{CEO} , reprezentujące prądy zerowe tranzystora, narysowano linią przerywaną, gdyż należy je uwzględnić dla tranzystorów germanowych, a można pominąć dla tranzystorów krzemowych.

Zasadniczymi przyczynami zmian punktu pracy są:

- zmiany temperatury,
- rozrzut parametrów poszczególnych egzemplarzy tranzystorów.

Ponieważ obie przyczyny wywołują zmiany tych samych parametrów, zagadnienie stabilności punktu pracy można zatem rozpatrywać łącznie dla obu przypadków. Na podstawie schematu zastępczego z rys. 7.3c

$$I_C = \beta_N I_B + I_{CEO} \quad (7.3)$$

Uwzględniając zależność (5.66a)

$$I_C = \beta_N I_B + (\beta_N + 1) I_{CBO} \quad (7.4)$$

lub w innej postaci

$$I_C = \alpha_N I_E + I_{CBO} \quad (7.5)$$

co jest zgodne ze schematem zastępczym na rys. 7.3b.

Prąd kolektora jest zatem funkcją trzech wielkości

$$I_C = f(\beta_N, I_{CBO}, I_B) \text{ w układzie WE i WC}$$

lub

$$I_C = f(\alpha_N, I_{CBO}, I_E) \text{ w układzie WB}$$

Ponieważ prąd bazy lub prąd emitera w konkretnych układach polaryzacji są zależne od napięcia φ_B , więc

$$I_C = f(\beta_N, I_{CBO}, \varphi_B) \quad (7.6)$$

Przy omawianiu zatem poszczególnych układów polaryzacji interesująca będzie wrażliwość prądu kolektora na zmiany: β_N , I_{CBO} , φ_B .

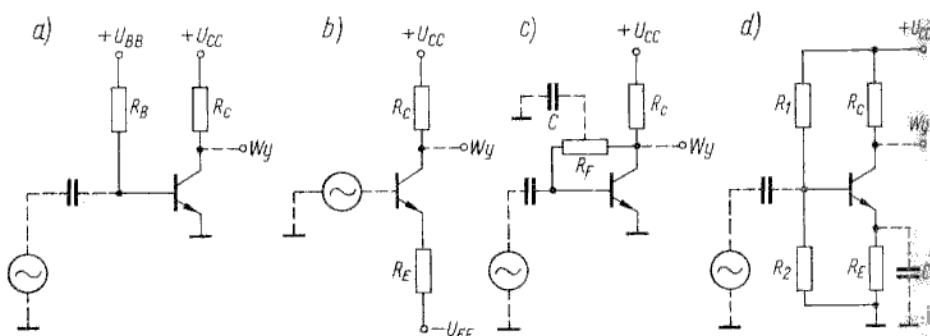
Zależności tych trzech parametrów w funkcji temperatury są rozpatrywane w Dodatku 1. W ogólnym przypadku dla małych przyrostów poszczególnych wielkości przyrost prądu kolektora wyznacza różniczka zupełna

$$dI_C = \frac{\partial I_C}{\partial \beta_N} d\beta_N + \frac{\partial I_C}{\partial I_{CBO}} dI_{CBO} + \frac{\partial I_C}{\partial \varphi_B} d\varphi_B \quad (7.7)$$

Dogodnie jest wprowadzić trzy parametry nazywane *współczynnikami stabilizacji* prądu kolektora:

$$\left. \begin{aligned} S_\beta &= \frac{\partial I_C}{\partial \beta_N} \Big|_{I_{CBO}, \varphi_B} \\ S_i &= \frac{\partial I_C}{\partial I_{CBO}} \Big|_{\beta_N, \varphi_B} \\ S_u &= \frac{\partial I_C}{\partial \varphi_B} \Big|_{\beta_N, I_{CBO}} \end{aligned} \right\} \quad (7.8)$$

Im mniejsze są wartości współczynników S_β , S_i , S_u , tym lepsza jest stałość prądu kolektora. Analizując poszczególne układy polaryzacji będziemy porównywać wartości współczynników stabilizacji, przy czym dla tranzystorów germanowych konieczne jest uwzględnienie wszystkich trzech współczynników, a dla tranzystorów krzemowych można pominąć S_i z uwagi na bardzo małą wartość prądu I_{CBO} , a więc również mały wpływ tego prądu na punkt pracy tranzystorów krzemowych. W celu zachowania ogólności rozważań uwzględnimy wszystkie trzy współczynniki.



Rys. 7.4

Cztery podstawowe układy polaryzacji tranzystorów bipolarnych: a) polaryzacja stałym prądem bazy; b) polaryzacja stałym prądem emitera; c) polaryzacja w pętli napięciowego sprzężenia zwrotnego; d) polaryzacja mieszana

Na rysunku 7.4 przedstawiono cztery podstawowe układy polaryzacji tranzystorów bipolarnych, tj.:

- polaryzację stałym prądem bazy — rys. 7.4a;
- polaryzację stałym prądem emitera — rys. 7.4b;
- polaryzację w pętli napięciowego sprzężenia zwrotnego — rys. 7.4c;
- polaryzację mieszaną — rys. 7.4d.

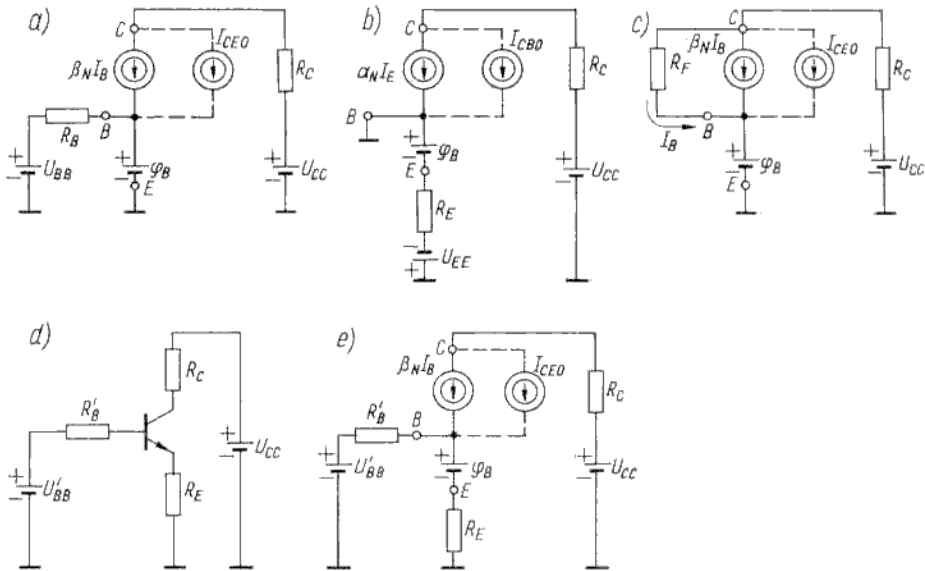
Na wszystkich czterech schematach liniami przerywanymi zaznaczono obwody dla wzmacnianego sygnału zmiennego.

Polaryzacja stałym prądem bazy

7.1.1

W układzie z rys. 7.4a, którego schemat zastępczy przedstawiono na rys. 7.5a, płynie stały prąd bazy ze źródła U_{BB} przez rezystor R_B

$$I_B = \frac{U_{BB} - \varphi_B}{R_B} \quad (7.9)$$



Rys. 7.5

Schematy zastępcze układów polaryzacji przedstawionych na rys. 7.4: a) dla układu polaryzacji stałym prądem bazy; b) dla układu polaryzacji stałym prądem emitera; c) dla układu polaryzacji w pętli napięciowego sprzężenia zwrotnego; d, e) dla układu polaryzacji mieszanej

Ponieważ zwykle $U_{BB} \gg \varphi_B$ (jest zalecane stosowanie jednego źródła polaryzacji, dlatego często $U_{BB} = U_{CC}$), zatem

$$I_B \approx \frac{U_{BB}}{R_B} = \text{const}$$

Po podstawieniu (7.9) do (7.4) otrzymuje się

$$I_C = \beta_N \frac{U_{BB} - \varphi_B}{R_B} + (\beta_N + 1) I_{CBO} \quad (7.10)$$

Stąd:

$$\left. \begin{aligned} S_\beta &= \frac{U_{BB} - \varphi_B}{R_B} + I_{CBO} \\ S_i &= (\beta_N + 1) \approx \beta_N \\ S_u &= -\frac{\beta_N}{R_B} \end{aligned} \right\} \quad (7.11)$$

W tym układzie projektant nie ma prawie żadnych możliwości optymalizacji wartości elementów obwodu polaryzacji (U_{BB} , R_B), gdyż z zadanego punktu pracy (I_C , U_{CE}) wynika określona wartość $I_B \approx U_{BB}/R_B$. Ze wzorów (7.11) można tylko zauważyć, że zwiększenie w jednakowych proporcjach U_{BB} i R_B nie zmienia, praktycznie biorąc, wartości współczynników S_β , S_i , lecz prowadzi do zmniejszenia wartości współczynnika S_u . Jednakże wartości wszystkich współczynników stabilizacji są bardzo duże, czyli punkt pracy tranzystora polaryzowanego stałym prądem bazy jest bardzo niestabilny.

Polaryzacja stałym prądem emitera

7.1.2

W układzie z rys. 7.4b, którego schemat zastępczy przedstawiono na rys. 7.5b, płynie stały prąd emitera ze źródła U_{EE} przez rezystor R_E

$$I_E = \frac{U_{EE} - \varphi_B}{R_E} \quad (7.12)$$

Ponieważ zwykle $U_{EE} \gg \varphi_B$, zatem

$$I_E \approx \frac{U_{EE}}{R_E} = \text{const}$$

Po podstawieniu (7.12) do (7.5) otrzymuje się

$$I_C = \frac{\beta_N}{\beta_N + 1} \frac{U_{EE} - \varphi_B}{R_E} + I_{CBO} \quad (7.13)$$

Stąd:

$$\left. \begin{aligned} S_\beta &= \frac{1}{(1 + \beta_N)^2} \frac{U_{EE} - \varphi_B}{R_E} \approx \frac{1}{\beta_N^2} \frac{U_{EE}}{R_E} \\ S_i &= 1 \\ S_u &= -\frac{\beta_N}{(\beta_N + 1)R_E} \approx -\frac{1}{R_E} \end{aligned} \right\} \quad (7.14)$$

Wszystkie współczynniki stabilizacji, wyrażone wzorami (7.14), mają znacznie mniejsze wartości niż dla przypadku polaryzacji stałym prądem bazy.

Polaryzacja w pętli napięciowego sprzężenia zwrotnego

7.1.3

W układzie z rys. 7.4c, którego schemat zastępczy przedstawiono na rys. 7.5c, prąd bazy płynie przez rezystor R_F włączony w pętli ujemnego sprzężenia zwrotnego napięciowego. Sprzężenie zwrotne działa stabilizująco na wartość napięcia U_{CE} , dlatego ten sposób polaryzacji można również nazwać polaryzacją stałym napięciem U_{CE} . Jakościowa interpretacja działania sprzężenia zwrotnego jest następująca. Jeżeli wskutek pewnej przyczyny zmaleje prąd kolektora, to zwiększy się wartość napięcia U_{CE} (zmaleje spadek napięcia na rezystorze R_C), przeto wzrośnie prąd bazy i odpowiednio zwiększy się prąd kolektora. W rezultacie prąd kolektora zmaleje mniej, niżby to było bez ujemnego sprzężenia zwrot-

nego. Podstawowe równanie Kirchhoffa dla tego obwodu (rys. 7.5e) jest następujące:

$$U_{CC} - (\beta_N I_B + I_{CEO} + I_B) R_C - I_B R_F - \varphi_B = 0 \quad (7.15)$$

Podstawiając do tego równania

$$I_B = \frac{I_C - (\beta_N + 1) I_{CBO}}{\beta_N}$$

otrzymuje się

$$I_C = \frac{\beta_N (U_{CC} - \varphi_B) + I_{CBO} (\beta_N + 1) (R_C + R_F)}{R_F + (\beta_N + 1) R_C} \quad (7.16)$$

Stąd przyjmując $\beta_N / (\beta_N + 1) \approx 1$, jest:

$$\left. \begin{aligned} S_\beta &= \frac{I_C R_F}{(\beta_N + 1) (R_F + R_C)} \\ S_i &= \frac{(\beta_N + 1) (R_C + R_F)}{(\beta_N + 1) R_C + R_F} \\ S_u &= - \frac{\beta_N}{(\beta_N + 1) R_C + R_F} \end{aligned} \right\} \quad (7.17)$$

Sprężenie zwrotne jest tym silniejsze i stabilność tym lepsza, im mniejsza jest wartość rezystora R_F . W granicznym przypadku dla

$$\left. \begin{aligned} R_F &\rightarrow 0: \\ S_\beta &\approx 0 \\ S_i &\approx 1 \\ S_u &\approx - \frac{1}{R_C} \end{aligned} \right\} \quad (7.17a)$$

W celu wyeliminowania ujemnego sprzężenia zwrotnego dla sygnału zmiennego rezystor R_F dzieli się na dwie części i do węzła wewnętrznego dołącza się kondensator odblokowujący składową zmienną (kondensator C na rys. 7.4e).

Polaryzacja mieszana

7.1.4

Układ polaryzacji przedstawiony na rys. 7.4d jest najczęściej stosowany. Korzystając z zasady Thévenina można go przekształcić w postać pokazaną na rys. 7.5d, przy czym:

$$\left. \begin{aligned} U'_{BB} &= \frac{R_1}{R_1 + R_2} U_{CC} \\ R'_B &= \frac{R_1 R_2}{R_1 + R_2} \end{aligned} \right\} \quad (7.18)$$

Łatwo można zauważyć, że układ przedstawiony na rys. 7.5d stanowi kombinację układów z rys. 7.4a, b, czyli łączy dwa sposoby polaryzacji (punkt pracy jest ustalany częściowo prądem bazy, a częściowo również prądem emitera). Źródło U_{BB} jest

zastąpione dzielnikiem napięcia (R_1, R_2) zasilanym ze źródła U_{CC} , a rezystor R_E powoduje ujemne sprzężenie zwrotne prądowe. Jakościowa interpretacja działania tego sprzężenia zwrotnego jest następująca. Jeżeli wskutek pewnej przyczyny zmaleje prąd emitera, to zmniejszy się spadek napięcia na rezystorze R_E , czyli wzrośnie różnica napięć baza-emiter (U_{BE}), a przy większym napięciu U_{BE} płynię większy prąd bazy i odpowiednio większy prąd emitera. W rezultacie prąd emitera zmaleje mniej, niżby to było bez ujemnego sprzężenia zwrotnego. W celu wyeliminowania ujemnego sprzężenia zwrotnego dla sygnału zmiennego rezystor R_E bocznikuje się kondensatorem, którego reaktancja przy najmniejszej częstotliwości sygnału wzmacnianego jest co najmniej kilkakrotnie mniejsza niż rezystancja R_E . Na podstawie schematu zastępczego układu polaryzacji (rys. 7.5e) można napisać równanie napięć dla obwodu wejściowego

$$U'_{BB} - I_B R'_B - \varphi_B - [(\beta_N + 1)I_B + I_{CEO}]R_E = 0$$

Stąd

$$I_B = \frac{U'_{BB} - \varphi_B - I_{CEO}R_E}{R'_B + (\beta_N + 1)R_E} \quad (7.19)$$

Po podstawieniu (7.19) do (7.4) otrzymuje się

$$I_C = \frac{\beta_N(U'_{BB} - \varphi_B) + I_{CEO}(\beta_N + 1)(R'_B + R_E)}{R'_B + (\beta_N + 1)R_E} \quad (7.20)$$

Stąd:

$$\left. \begin{aligned} S_B &= I_C \frac{R'_B}{(\beta_N + 1)(R'_B + R_E)} \\ S_i &= \frac{(\beta_N + 1)(R'_B + R_E)}{(\beta_N + 1)R_E + R'_B} \\ S_u &= - \frac{\beta_N}{(\beta_N + 1)R_E + R'_B} \end{aligned} \right\} \quad (7.21)$$

Stabilność poprawia się w miarę jak $R'_B \rightarrow 0$, a $R_E \rightarrow \infty$.

W granicznym przypadku ($R'_B = 0$, $R_E = \infty$) jest:

$$\left. \begin{aligned} S_B &= 0 \\ S_i &= 1 \\ S_u &= 0 \end{aligned} \right\} \quad (7.22)$$

Te granicznie małe wartości współczynników stabilności nie są w praktyce realizowane, gdyż nie można nieograniczenie zmniejszać rezystancji R'_B (wzrasta pobór prądu dzielnika oraz silnie jest bocznikowany sygnał wejściowy) ani też nie można nieograniczenie zwiększać rezystancji R_E (wzrasta nieograniczenie wymagana wartość napięcia zasilania).

Podstawowe układy wzmacniające tranzystora bipolarnego

7.2

Rozpatrzmy teraz zagadnienie wzmacniania małych sygnałów zmiennych w trzech podstawowych układach włączenia tranzystora bipolarnego: WE, WB, WC. Trzy układy włączenia omawiano już w rozdziale 5, jednakże wówczas

interesowały nas właściwości samego tranzystora, a nie obwodu, w jakim on pracuje, przyjmowano zatem wyidealizowane warunki sterowania i obciążenia (na przykład zwarcie na wyjściu i rozwarcie na wejściu). W ten sposób tranzystor można było opisać za pomocą parametrów czwórnikowych lub przedstawić w postaci schematu zastępczego. Natomiast w tym punkcie interesujące są tzw. parametry robocze tranzystora pracującego w konkretnym układzie włączenia przy realnych wartościach impedancji wewnętrznej źródła sterującego oraz impedancji obciążenia. W zakresie małych i średnich częstotliwości (dla $f \ll f_T$) właściwości układów wzmacniających WE, WB, WC są opisywane pięcioma parametrami roboczymi:

k_u — wzmacnienie napięciowe,

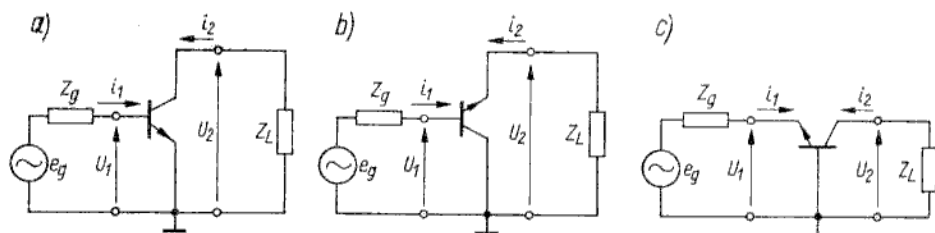
k_i — wzmacnienie prądowe,

k_p — wzmacnienie mocy,

Z_{we} — impedancja wejściowa,

Z_{wy} — impedancja wyjściowa.

Ponieważ $k_p = k_i k_u$, zatem liczbę parametrów pierwotnych można zmniejszyć do czterech.

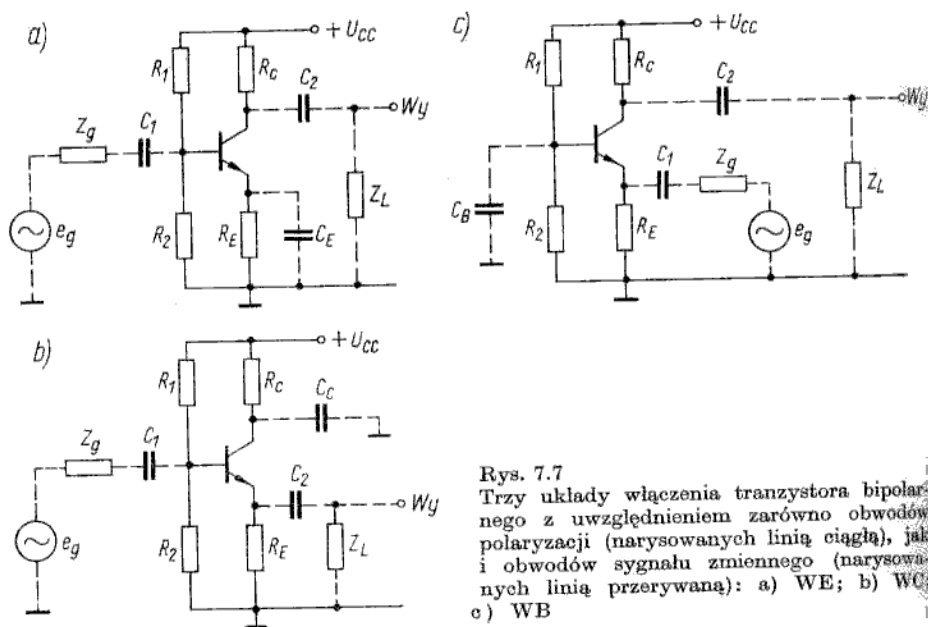


Rys. 7.6

Trzy układy włączenia tranzystora bipolarnego (bez obwodów polaryzacji): a) WE; b) WC; c) WB

Na rysunku 7.6 przedstawiono trzy układy włączenia, uwzględniając tylko sygnał zmienny (bez obwodów polaryzacji stałoprądowej). W rzeczywistości różnice schematowe między poszczególnymi układami nie są tak dobrze widoczne, jak to wynika z rys. 7.6, gdyż nie zawsze wspólna elektroda jest zwarta galwanicznie z masą oraz występują obwody polaryzacji, które dla wszystkich trzech konfiguracji mogą być identyczne. Dobrze ilustruje to rys. 7.7, który przedstawia trzy konfiguracje tranzystora z uwzględnieniem zarówno obwodów polaryzacji (narysowanych linią ciągłą), jak i obwodów sygnału zmiennego (narysowanych linią przerywaną). We wszystkich trzech przypadkach przyjęto umyślnie identyczny układ polaryzacji z dzielnikiem R_1, R_2 i rezystorami R_C, R_E . W układzie na rys. 7.7a sygnał zmienny jest wprowadzany ze źródła e_g , o impedancji wewnętrznej Z_g , przez kondensator C_1 do bazy tranzystora oraz wyprowadzany z kolektora przez kondensator C_2 do obciążenia Z_L . Wejściem jest zatem baza, a wyjściem — kolektor, czyli jest to układ WE. Emiter jest wspólny dla wejścia i wyjścia, gdyż dla sygnału zmiennego kondensator C_E stanowi zwarcie do masy. Jednakże gdyby nie było kondensatora C_E , byłby to w dalszym ciągu układ WE, tylko że z prądowym sprzężeniem zwrotnym na rezystorze R_E (dla składowej stałej i zmiennej).

W układzie na rys. 7.7b sygnał zmienny jest wprowadzany ze źródła e_g o impedancji wewnętrznej Z_g przez kondensator C_1 do bazy tranzystora oraz wyprowadzany



Rys. 7.7
Trzy układy włączenia tranzystora bipolarnego z uwzględnieniem zarówno obwodów polaryzacji (narysowanych linią ciągłą), jak i obwodów sygnału zmiennego (narysowanych linią przerywaną): a) WE; b) WC; c) WB

z emitera przez kondensator C_2 do obciążenia Z_L . Wejściem jest zatem baza, a wyjściem — emiter, czyli jest to układ WC. Kolektor jest wspólny dla wejścia i wyjścia, gdyż dla sygnału zmiennego kondensator C_C stanowi zwarcie do masy.

W układzie na rys. 7.7c sygnał zmienny jest wprowadzany ze źródła e_g o impedancji wewnętrznej Z_g przez kondensator C_1 do emitera tranzystora oraz wyprowadzany z kolektora przez kondensator C_2 do obciążenia Z_L . Wejściem jest zatem emiter, a wyjściem — kolektor, czyli jest to układ WB. Baza jest wspólna dla wejścia i wyjścia, gdyż dla sygnału zmiennego kondensator C_B stanowi zwarcie do masy.

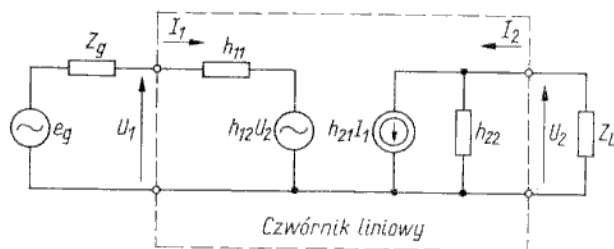
Analizę poszczególnych konfiguracji można wykonać korzystając z parametrów czwórnikowych lub schematów zastępczych tranzystora. Zastosujemy pierwszą metodę analizy (na podstawie parametrów czwórnikowych). Podstawowe układy włączenia tranzystora polowego będziemy analizować na podstawie schematów zastępczych. Ponieważ schematy zastępcze tranzystorów bipolarnych i polowych mają identyczną strukturę typu π , analizę przedstawioną dalej dla tranzystora polowego można również z łatwością odnieść do tranzystora bipolarnego.

Wyprowadzimy najpierw ogólne zależności dla czwórnika liniowego, oparte o macierz parametrów typu h .

Parametry robocze czwórnika liniowego

7.2.1

Wyprowadzimy wyrażenia na cztery parametry robocze (k_i, k_u, Z_{we}, Z_{wy}) czwórnika liniowego sterowanego ze źródła e_g o impedancji wewnętrznej Z_g i obciążonego impedancją Z_L (rys. 7.8).



Rys. 7.8

Czwórnik liniowy reprezentowany parametrami h , sterowany ze źródła e_g o impedancji wewnętrznej Z_g oraz obciążony impedancją Z_L .

a. Wzmocnienie prądowe

$$k_i = -\frac{I_2}{I_1} \quad (7.23)$$

Na podstawie rys. 7.8

$$I_2 = h_{21}I_1 + h_{22}U_2 \quad (7.24)$$

Ponieważ

$$U_2 = -I_2 Z_L$$

zatem

$$I_2 = h_{21}I_1 - I_2 h_{22} Z_L$$

czyli

$$I_2(1 + h_{22} Z_L) = I_1 h_{21}$$

Stąd

$$k_i = -\frac{h_{21}}{1 + h_{22} Z_L} \quad (7.25)$$

b. Impedancja wejściowa

$$Z_{we} = \frac{U_1}{I_1} \quad (7.26)$$

Na podstawie rys. 7.8

$$U_1 = h_{11}I_1 + h_{12}U_2 = h_{11}I_1 - h_{12}I_2 Z_L$$

Stąd

$$Z_{we} = h_{11} + h_{12} Z_L k_i$$

oraz uwzględniając wyrażenie (7.25)

$$Z_{we} = h_{11} - \frac{h_{12} h_{21} Z_L}{1 + h_{22} Z_L}$$

lub inaczej

$$Z_{we} = h_{11} - \frac{h_{12}h_{21}}{h_{22} + Y_L} \quad (7.27)$$

przy czym $Y_L = 1/Z_L$.

c. Wzmocnienie napięciowe

$$k_u = \frac{U_2}{e_g} \quad (7.28)$$

Ponieważ

$$e_g = I_1(Z_g + Z_{we})$$

zatem

$$k_u = k_i \frac{Z_L}{Z_g + Z_{we}} \quad (7.29)$$

d. Admitancja wyjściowa

$$Y_{wy} = \frac{1}{Z_{wy}} = \left. \frac{I_2}{U_2} \right|_{e_g=0} \quad (7.30)$$

Z równania (7.24)

$$Y_{wy} = h_{21} \frac{I_1}{U_2} + h_{22} \quad (7.31)$$

Dla $e_g = 0$ równanie napięć w obwodzie wejściowym na rys. 7.8 jest następujące:

$$Z_g I_1 + h_{11} I_1 + h_{12} U_2 = 0$$

Stąd

$$\frac{I_1}{U_2} = - \frac{h_{12}}{Z_g + h_{11}} \quad (7.32)$$

Po podstawieniu (7.32) do (7.31) otrzymuje się

$$Y_{wy} = h_{22} - \frac{h_{12}h_{21}}{Z_g + h_{11}} \quad (7.33)$$

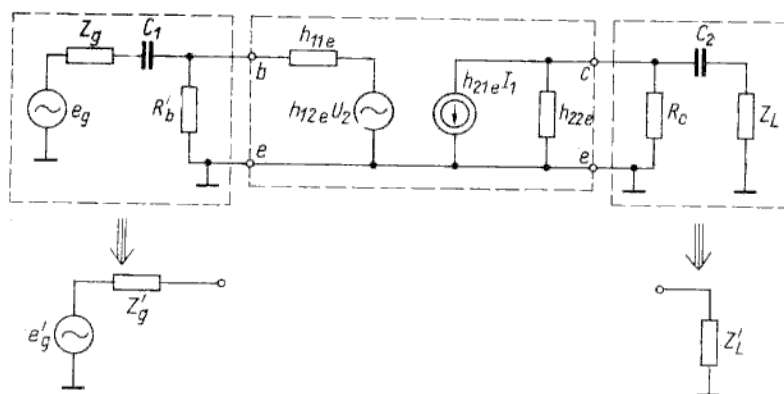
Korzystając z wyprowadzonych zależności ogólnych wyznaczy się parametry robocze dla poszczególnych konfiguracji wzmacniaczy.

Wzmacniacz w układzie WE

7.2.2

Wzmacniacz WE z rys. 7.7a można przedstawić w postaci sieci równoważnej dla małych sygnałów zmiennych (rys. 7.9). Sieć równoważną budowano przy następujących założeniach:

- źródło zasilania U_{CC} stanowi zwarcie dla sygnału zmiennego,
- kondensator C_E zwiiera emiter z masą dla sygnału zmiennego.



Rys. 7.9

Sieć równoważna wzmacniacza w konfiguracji WE, przedstawionego na rys. 7.7a

Jeżeli ponadto przyjmie się, że impedancje kondensatorów C_1 , C_2 są pomijalnie małe, to układ z rys. 7.9 można doprowadzić do postaci zgodnej z ogólnym schematem przedstawionym na rys. 7.8. Należy przy tym uwzględnić wpływ bocznikującej rezystancji

$$R'_B = \frac{R_1 R_2}{R_1 + R_2}$$

na wypadkową impedancję wewnętrzną źródła sygnału oraz rezystancji R_C na wypadkową admitancję obciążenia.

Stosując zasadę Thèvenina otrzymuje się:

$$\left. \begin{aligned} e'_g &= e_g \frac{R'_B}{Z_g + R'_B} \\ Z'_g &= \frac{Z_g R'_B}{Z_g + R'_B} \\ Z'_L &= \frac{Z_L R_C}{Z_L + R_C} \end{aligned} \right\} \quad (7.34)$$

Jeżeli są spełnione warunki:

$$R'_B \gg Z_g$$

$$R_C \gg Z_L$$

to:

$$e'_g \approx e_g$$

$$Z'_g \approx Z_g$$

$$Z'_L \approx Z_L$$

Na podstawie rys. 7.9 oraz korzystając z zależności (7.25), (7.26), (7.29) i (7.33) otrzymuje się:

a. Wzmocnienie prądowe

$$k_i = \frac{-h_{21e}}{1+h_{22e}Z'_L} \quad (7.35)$$

Ponieważ zwykle $(h_{22e}Z'_L) \ll 1$, zatem

$$k_i \approx -h_{21e} \quad (7.35a)$$

b. Wzmocnienie napięciowe

$$k_u = k_i \frac{Z'_L}{Z'_g + Z_{we}} \approx \quad (7.36)$$

$$\approx -h_{21e} \frac{Z'_L}{Z'_g + Z_{we}} \quad (7.36a)$$

c. Impedancja wejściowa

$$Z_{we} = h_{11e} - \frac{h_{12e}h_{21e}}{h_{22e} + Y'_L} \quad (7.37)$$

przy czym

$$Y'_L = \frac{1}{Z'_L} = \frac{Z_L + R_C}{Z_L R_C}$$

Dla $Z'_L \rightarrow 0$, $Z_{we} \rightarrow h_{11e}$.

d. Admitancja wyjściowa

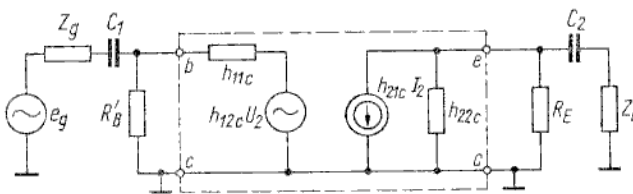
$$Y_{wy} = h_{22e} - \frac{h_{12e}h_{21e}}{Z'_g + h_{11e}} \quad (7.38)$$

Dla $Z'_g \rightarrow \infty$, $Y_{wy} \rightarrow h_{22e}$.

Wzmacniacz w układzie WC (wtórnik emiterowy)

7.2.3

Przy analogicznych założeniach wstępnych jak w poprzednim punkcie można przedstawić wzmacniacz WC z rys. 7.7b w postaci sieci równoważnej (rys. 7.10). Układ z rys. 7.10 doprowadza się do postaci ogólnej, zgodnej z rys. 7.8, za pomocą następujących przekształceń:



Rys. 7.10

Sieć równoważna wzmacniacza w konfiguracji WC, przedstawionego na rys. 7.7b

$$\left. \begin{aligned} e'_g &= e_g \frac{R'_B}{Z_g + R'_B} \\ Z'_g &= \frac{Z_g R'_B}{R_g + R'_B} \\ Z'_L &= \frac{Z_L R_E}{Z_L + R_E} \end{aligned} \right\} \quad (7.39)$$

Parametry robocze układu WC są następujące:

a Wzmocnienie prądowe

$$k_i = - \frac{h_{21c}}{1 + h_{22c} Z'_L} \quad (7.40)$$

Zwykle $(h_{22c} Z'_L) \ll 1$, zatem

$$k_i \approx -h_{21c} \approx (\beta + 1) \quad (7.40a)$$

b. Wzmocnienie napięciowe

$$k_u = k_i \frac{Z'_L}{Z'_g + Z_{we}} \quad (7.41)$$

Ponieważ $Z_{we} \approx (\beta + 1) Z'_L$ oraz zwykle $Z'_g \ll Z_{we}$, więc

$$k_u \leq 1 \quad (7.41a)$$

W układzie WC napięcie wyjściowe jest zatem prawie równe (ale zawsze mniejsze) napięciu wejściowemu. Dlatego układ WC jest nazywany *wtórnikami emiterowym* (sygnał wyjściowy powtarza, praktycznie biorąc, bez zmian sygnał wejściowy).

c. Impedancja wejściowa

$$Z_{we} = h_{11c} + \frac{h_{12c} h_{21c}}{h_{22c} + Y'_L} \quad (7.42)$$

Zwykle $Y'_L \gg h_{22c}$, a ponieważ $h_{12c} \approx 1$, więc

$$Z_{we} \approx h_{11c} + h_{21c} Z'_L$$

Często można pominąć składnik h_{11c} , zatem

$$Z_{we} \approx (\beta + 1) Z'_L \quad (7.42a)$$

d. Admitancja wyjściowa

$$Y_{wy} = h_{22c} + \frac{h_{12c} h_{21c}}{Z'_g + h_{11c}} \quad (7.43)$$

Pomijając bardzo mały składnik h_{22c} oraz uwzględniając $h_{12c} = 1$, otrzymuje się

$$Y_{wy} \approx \frac{\beta + 1}{Z'_g + h_{11c}} \quad (7.43a)$$

Porównanie podstawowych parametrów roboczych dla układów WE, WC, WB

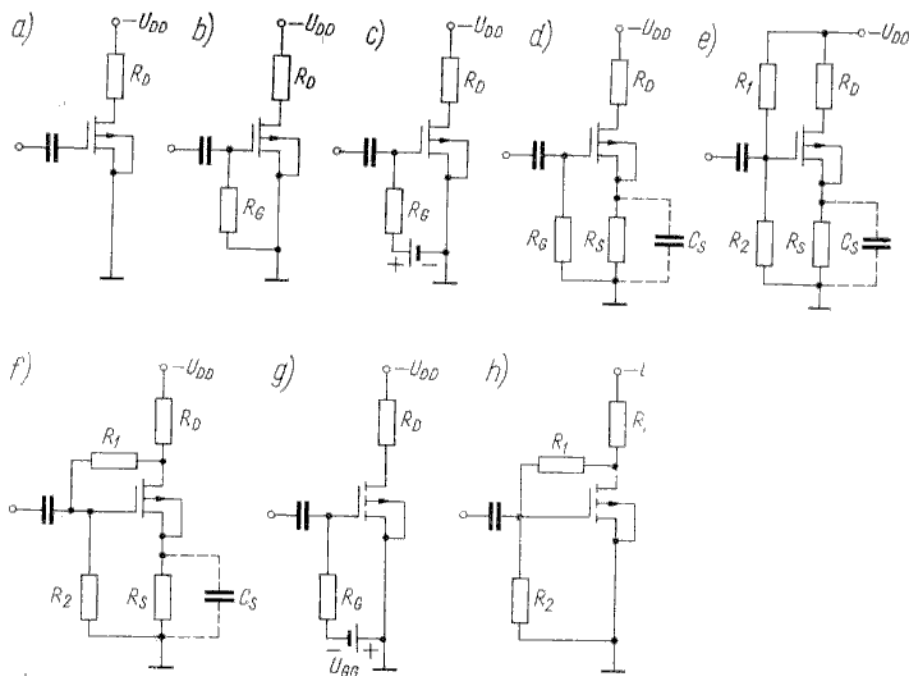
Tablica 7.1

Konfiguracja \ Parametr	k_i	k_u	k_p	Przesunięcie fazy	Z_{we}	Z_{wy}
WE	Duże ($\sim \beta$)	Duże	Największe	180°	Mała	Duża
WC	Największe ($\sim (\beta + 1)$)	Małe ≤ 1	Małe	0°	Największa	Najmniejsza
WB	Małe ($\alpha \leq 1$)	Największe	Duże	0°	Najmniejsza	Największa

Sposoby polaryzacji tranzystorów polowych

7.3

Na rysunku 7.12 pokazano typowe warianty układów polaryzacji tranzystorów polowych. Pierwsze dwa układy (rys. 7.12a, b) odnoszą się do tranzystora MIS z kanałem zubożonym ($U_{GS} = 0$), następane cztery (rys. 7.12c—f) dotyczą w równym stopniu tranzystorów PNFET jak tranzystora MIS z kanałem zubożonym ($U_{GS} > 0$ dla kanału typu p), ostatnie dwa układy (rys. 7.12g, h) odnoszą się do tranzystora MIS z kanałem wzbogacającym ($U_{GS} < 0$). Całkowite wyko-



Rys. 7.12

Typowe warianty polaryzacji tranzystorów polowych

rzystanie bardzo dużej rezystancji wejściowej jest możliwe w tranzystorze z kanałem zubożanym, który nie wymaga wstępnej polaryzacji bramki (rys. 7.12a). W tym przypadku należy się jednak liczyć z możliwością elektrostatycznego magazynowania ładunku w bramce (niebezpieczeństwo przebiecia dielektryka), gdyż brak obwodu polaryzacji uniemożliwia rozładowanie pojemności dielektryka. Ładunek zgromadzony w pojemności bramka-kanał można odprowadzić przez rezystor R_G , włączony między bramkę a źródło (rys. 7.12b). Wówczas tranzystor pracuje przy napięciu $U_{GS} = 0$. Punkt pracy w zakresie zubożenia ($U_{GS} > 0$ dla tranzystora z kanałem typu p) można otrzymać przez włączenie oddzielnego źródła (rys. 7.12c — tzw. polaryzacja stałym napięciem bramki) lub metodą polaryzacji „automatycznej” bramki za pomocą spadku napięcia odkładającego się na rezystorze R_S włączonym szeregowo ze źródłem (rys. 7.12d). Zaletą polaryzacji „automatycznej” jest to, że ujemne prądowe sprzężenie zwrotne w obwodzie źródła w znacznym stopniu uniezależnia punkt pracy od zmian parametrów tranzystora. Jeżeli U_{GS} , I_D są wartościami spoczynkowymi napięcia bramka-źródło i prądu drenu w pożądanym punkcie pracy, to

$$R_S = \frac{U_{GS}}{I_D} \quad (7.49)$$

Rezystor R_S można z bocznikować kondensatorem C_S w celu wyeliminowania ujemnego sprzężenia zwrotnego dla sygnału zmiennego.

Niekiedy duże wartości rezystora R_S , wymagane ze względu na stabilność punktu pracy, dają zbyt duże wartości napięcia U_{GS} . Wówczas stosuje się kombinację polaryzacji „automatycznej”, dającej $U_{GS} > 0$ z dodatkową polaryzacją bramki napięciem $U_{GS} < 0$, przy czym można użyć oddzielne źródło lub dzielnik napięcia włączony w sposób pokazany na rys. 7.12e, f.

W przypadku tranzystora z kanałem wzbogacanym typu p jest wymagana polaryzacja ujemna bramki napięciem $|U_{GS}| > |U_T|$. Polaryzację taką uzyskuje się z oddzielnego źródła (rys. 7.12g) lub dzielnika rezystancyjnego, włączonego na przykład w układzie pokazanym na rys. 7.12h. Ten układ jest często stosowany przy $R_2 = \infty$. Wówczas $U_{GS} = U_{DS}$ i tranzystor pracuje w zakresie nasycenia.

Podstawowe układy wzmacniające z tranzystorami polowymi

74

Rozpatrzmy zagadnienie wzmacniania małych sygnałów zmiennych w trzech podstawowych układach włączenia tranzystora polowego:

- WS (ze wspólnym źródłem),
- WD (ze wspólnym drenem),
- WG (ze wspólną bramką).

Tranzystory MOS potraktujemy jako elementy trójkońcówkowe (podłoże zwarte ze źródłem). Ponieważ schematy zastępcze tranzystorów PNFET i MOS (w układzie trójkońcówkowym) są jednakowe, można wykonać analizę wspólną dla obu rodzajów tranzystorów polowych. Dalsze zatem rozważania dotyczące bezpośrednio tranzystora PNFET z kanałem typu p odnoszą się również do pozostałych rodzajów tranzystorów polowych, przy czym zmiana typu przewodnictwa kanału wymaga odpowiedniej zmiany polaryzacji napięć zasilających. Ponieważ w tranzystorach polowych nie ma sensu pojęcie wzmocnienia prądowego, będziemy analizować tylko trzy parametry robocze, tj.:

W zakresie małych częstotliwości, gdy susceptancje międzyelektrodowe są pomijalnie małe, wyrażenie (7.51) upraszcza się do następującej postaci:

$$k_u(0) = \frac{-g_m R_L}{1 + g_d R_L} \quad (7.52)$$

Wzmocnienie napięciowe $k_u(0)$ w miarę wzrostu rezystancji obciążenia R_L dąży do wartości maksymalnej, równej współczynnikowi wzmocnienia napięciowego tranzystora $\mu_a = -g_m/g_d$. Znak ujemny wzmocnienia napięciowego jest związany ze zjawiskiem odwrócenia fazy sygnału wyjściowego w stosunku do sygnału wejściowego.

b. *Impedancja wejściowa*. Na podstawie schematu zastępczego, przedstawionego na rys. 7.13b, impedancję wejściową można wyznaczyć według następującego wzoru:

$$\begin{aligned} Z_{we} &= \frac{U_{we}}{I_{we}} = \frac{U_{we}}{U_{we} j\omega C_{gs} + (U_{we} - U_{wy}) j\omega C_{gd}} = \\ &= \frac{1}{j\omega \{C_{gs} + C_{gd}[1 - k_u(\omega)]\}} \end{aligned} \quad (7.53)$$

Dla bardzo małych częstotliwości, gdy $k_u(\omega) \approx k_u(0)$ jest liczbą rzeczywistą

$$Z_{we} = \frac{1}{j\omega C_{we}} \quad (7.53a)$$

przy czym

$$C_{we} = C_{gs} + C_{gd}[1 - k_u(0)] \quad (7.53b)$$

Pojemność wejściowa tranzystora składa się z pojemności międzyelektrodowej bramka-źródło oraz zwiększonej $[1 - k_u(0)]$ razy pojemności bramka-dren (k_u podstawia się ze znakiem ujemnym). Ta ostatnia składowa, spowodowana tzw. *zjawiskiem Millera*, znacznie ogranicza maksymalną częstotliwość pracy tranzystora.

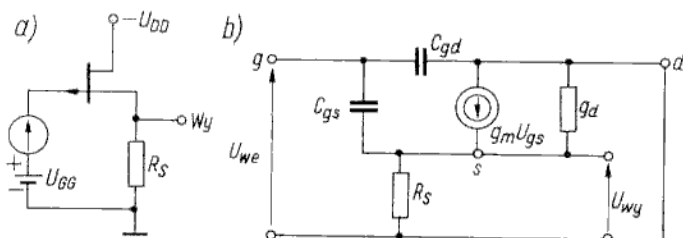
c. *Impedancja wyjściowa*. Na podstawie rys. 7.13b

$$Z_{wy} = \left. \frac{U_{wy}}{I_{wy}} \right|_{U_{we}=0} = \frac{1}{g_d} \quad (7.54)$$

Układ ze wspólnym drenem

7.4.2

Układ ze wspólnym drenem (WD), w którym zacisk wejściowy stanowi bramka, zacisk wyjściowy — źródło, a zacisk wspólny dla wejścia i wyjścia — dren, analogicznie do układu wtórnika emiterowego jest nazywany *wtórnikiem źródłowym*. Wtórnik źródłowy charakteryzuje się bardzo dużą impedancją wejściową, dość małą impedancją wyjściową oraz wzmocnieniem napięciowym k_u mniejszym od jedności (lecz bardzo bliskim jedności). Współczynnik k_u ma znak dodatni, gdyż sygnał wyjściowy jest zgodny co do fazy z sygnałem wejściowym. Konfigurację WD stosuje się w tych przypadkach, w których zachodzi konieczność transformacji impedancji (z dużej na małą), gdy jest pożądana mała pojemność wejściowa układu oraz do pracy przy dużych sygnałach (małe zniekształcenia nieliniowe). Tranzystor włączony w układzie WD przedstawiono na rys. 7.14a, a jego schemat zastępczy — na rys. 7.14b.



Rys. 7.14

Układ ze wspólnym drenem (a) i jego schemat zastępczy (b)

a. *Wzmocnienie napięciowe.* Na podstawie schematu zastępczego (rys. 7.14b) zapisuje się równanie sumy prądów przepływających przez zacisk źródła S

$$-g_m U_{gs} + U_{wy} \left(\frac{1}{R_S} + g_d \right) + (U_{wy} - U_{we}) j\omega C_{gs} = 0$$

Ponieważ

$$U_{gs} = U_{we} - U_{wy}$$

zatem

$$U_{wy} \left(\frac{1}{R_S} + g_d \right) + (U_{wy} - U_{we}) (j\omega C_{gs} + g_m) = 0$$

Stąd otrzymuje się

$$k_u(\omega) = \frac{g_m + j\omega C_{gs}}{\frac{1}{R_S} + g_d + g_m + j\omega C_{gs}} \quad (7.55)$$

W zakresie małych częstotliwości, przy $j\omega C_{gs} \rightarrow 0$, wyrażenie (7.55) upraszcza się do postaci

$$k_u(0) = \frac{g_m}{\frac{1}{R_S} + g_d + g_m} \quad (7.56)$$

Ponieważ w zakresie nasycenia $g_d \ll g_m$, zatem

$$k_u(0) = \frac{g_m R_S}{1 + g_m R_S} \quad (7.56a)$$

Dla $R_S \rightarrow \infty$ wzmocnienie napięciowe dąży do jedności.

b. *Impedancja wejściowa.* Zgodnie ze schematem zastępczym (rys. 7.14b) dodaje się wszystkie prądy płynące przez zacisk wejściowy (zacisk bramki G)

$$I_{we} = U_{we} j\omega C_{gd} + (U_{we} - U_{wy}) j\omega C_{gs}$$

Stąd

$$Z_{we} = \frac{U_{we}}{I_{we}} = \frac{1}{j\omega} \{ C_{gd} + [1 - k_u(\omega)] C_{gs} \}^{-1} \quad (7.57)$$

Dla bardzo małych częstotliwości, gdy $k_u(\omega) = k_u(0)$ jest liczbą rzeczywistą

$$Z_{we} = \frac{1}{j\omega C_{we}}$$

przy czym

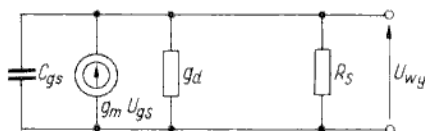
$$C_{we} = C_{gd} + [1 - k_u(0)]C_{gs} \quad (7.58)$$

Ponieważ $k_u(0)$ jest dodatnie i bliskie jedności, zatem

$$C_{we} \approx C_{gd} \quad (7.58a)$$

W układzie WD nie występuje zatem zjawisko zwielokrotnienia pojemności sprzężenia zwrotnego (zjawisko Millera) i graniczna częstotliwość pracy tego układu jest znacznie większa niż dla układu WS.

c. *Impedancja wyjściowa.* W celu wyznaczenia impedancji wyjściowej korzysta się ze schematu zastępczego, przedstawionego na rys. 7.15, który otrzymuje się



Rys. 7.15

Schemat zastępczy do wyznaczenia admitancji wyjściowej układu WD

ze schematu przedstawionego na rys. 7.14b przez zwarcie wejścia. Źródło sygnału jest podłączone do zacisków wyjściowych tranzystora. Zgodnie z tym schematem sumaryczny prąd płynący przez zacisk drenu

$$I_{wy} = U_{wy} \left(\frac{1}{R_S} + g_d + j\omega C_{gs} \right) + g_m U_{gs}$$

Po podstawieniu

$$U_{gs} = U_{wy}$$

otrzymuje się

$$Z_{wy} = \frac{U_{wy}}{I_{wy}} = \left(\frac{1}{R_S} + g_d + j\omega C_{gs} + g_m \right)^{-1} \quad (7.59)$$

W zakresie bardzo małych częstotliwości ($j\omega C_{gs} \rightarrow 0$), biorąc również pod uwagę, że $g_d \ll g_m$, można uprościć wzór (7.59) do postaci

$$Z_{wy} \approx \frac{R_S}{1 + g_m R_S} \quad (7.59a)$$

Dla $R_S \rightarrow \infty$

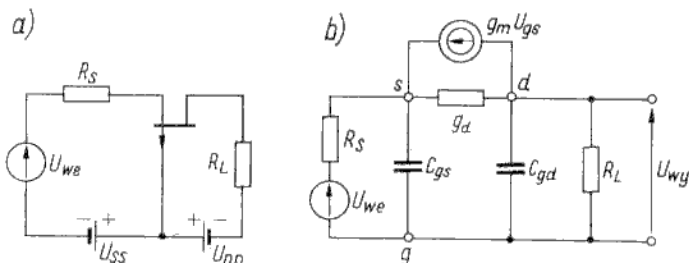
$$Z_{wy} \rightarrow \frac{1}{g_m} \quad (7.59b)$$

Układ ze wspólną bramką

7.4.3

Układ ze wspólną bramką (WG), w którym zacisk wejściowy stanowi źródło, zacisk wyjściowy — dren, a zacisk wspólny dla wejścia i wyjścia — bramka,

charakteryzuje się małą impedancją wejściową i dużą impedancją wyjściową. Tranzystor włączony w układzie WG przedstawiono na rys. 7.16a, a jego schemat zastępczy — na rys. 7.16b.



Rys. 7.16
Układ ze wspólną bramką (a) i jego schemat zastępczy (b)

a. *Wzmocnienie napięciowe.* Równanie prądów płynących przez zacisk drenu

$$-g_m U_{gs} + (U_{wy} - U_{gs})g_d + U_{wy} \left(j\omega C_{gd} + \frac{1}{R_L} \right) = 0$$

Ponieważ

$$U_{gs} = U_{we} \frac{Z_{we}}{R_S + Z_{we}}$$

zatem

$$k_u(\omega) = \frac{g_m + g_d}{g_d + G_L + j\omega C_{gd}} \frac{1}{1 + R_S Y_{we}} \quad (7.60)$$

Dla przypadku $R_S = 0$ oraz $j\omega C_{gd} \rightarrow 0$ (mała częstotliwość)

$$k_u(0) \approx \frac{g_m + g_d}{g_d + G_L} \quad (7.61)$$

b. *Impedancja wejściowa.* Impedancję wejściową, widzianą z zacisku źródła, wyznacza się na podstawie schematu zastępczego (rys. 7.16b)

$$I_{we} = j\omega C_{gs} U_{gs} + (U_{gs} - U_{wy})g_d + g_m U_{gs}$$

Stąd

$$Z_{we} = \frac{U_{gs}}{I_{we}} = \{j\omega C_{gs} + g_m + [1 - k_u(\omega)]g_d\}^{-1} \quad (7.62)$$

Dla $\omega \rightarrow 0$

$$Z_{we} = \frac{1 + g_d R_L}{g_m + g_d} \quad (7.62a)$$

Ponieważ $g_d \ll g_m$, zatem dla $R_L \rightarrow 0$

$$Z_{we} \rightarrow \frac{1}{g_m} \quad (7.62b)$$

c. *Impedancja wyjściowa*. Zwierając wejście układu przedstawionego na rys. 7.16b (zwarcie zacisków s, g) otrzymuje się

$$I_{wy} = U_{wy} (G_L + j\omega C_{gd} + g_d)$$

Stąd

$$Z_{wy} = \frac{U_{wy}}{I_{wy}} = (G_L + g_d + j\omega C_{gd})^{-1} \quad (7.63)$$

Parametry robocze dla wszystkich trzech konfiguracji porównano w tabl. 7.2.

Porównanie parametrów roboczych dla układów WS, WD, WG

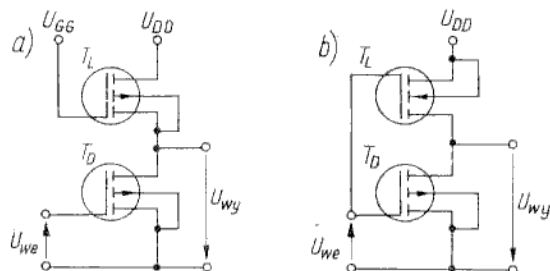
Tablica 7.2

Konfiguracja \ Parametr	k_u	Przesunięcie fazy	Z_{we}	Z_{wy}
WS	Duże (maksymalnie $\frac{g_m}{g_d}$)	180°	Bardzo duża (praktycznie ∞)	Duża (maksymalnie $\frac{1}{g_d}$)
WD	Małe ≤ 1	0°	Bardzo duża (praktycznie ∞)	Mała maksymalnie $\left(\frac{1}{g_m}\right)$
WG	Duże (maksymalnie $\frac{g_m}{g_d}$)	0°	Mała (maksymalnie $\frac{1}{g_m}$)	Duża (maksymalnie $\frac{1}{g_d}$)

Tranzystor MOS jako obciążenie rezystancyjne

7.5

Jedną z podstawowych zalet tranzystorów MOS jest możliwość wytwarzania układów scalonych, zarówno analogowych jak i cyfrowych, składających się wyłącznie z tych tranzystorów, spełniających wszystkie funkcje elementów czynnych i biernych. W tym punkcie rozpatrzmy podstawowe właściwości wzmacniacza liniowego, w którym funkcję obciążenia spełnia tranzystor MOS pracujący jako rezystor quasi-liniowy w zakresie nasycenia lub nienasycenia. Zastosowanie tranzystora MOS jako obciążenia umożliwia uzyskanie dużych wartości rezystancji obciążenia, dobrej liniowości wzmocnienia w dużym zakresie dynamicznym oraz bardzo dobrej stabilności temperaturowej.

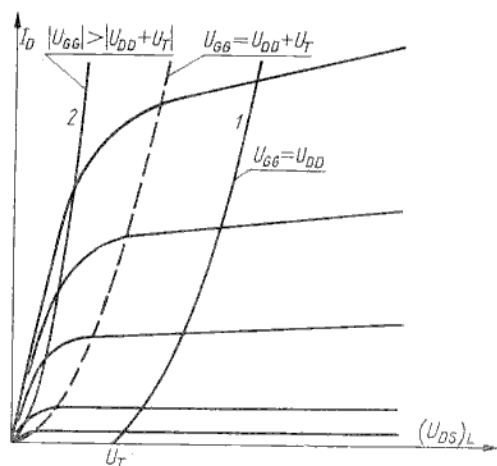


Rys. 7.17

Układ WS z tranzystorem MOS w obciążeniu: a) oba tranzystory z kanałem typu p ; b) tranzystory o przeciwstawnym typie kanału
 T_L — tranzystor obciążający;
 T_D — tranzystor sterowany

Na rysunku 7.17 przedstawiono dwa podstawowe układy, w których tranzystor MOS (dalej nazywany tranzystorem obciążającym) spełnia funkcję obciążenia dla tranzystora MOS sterowanego, włączanego w układzie WS. Istnieje oczywiście możliwość realizacji wielu innych wariantów połączeń tranzystora obciążającego z tranzystorem sterowanym, włączonym w układzie WG lub WD, przy czym są możliwe połączenia tranzystorów z kanałami wzbogacanymi lub zubożanymi albo tranzystora z kanałem wzbogacanym z tranzystorem z kanałem zubożanym. Zatem układy przedstawione na rys. 7.17 należy traktować jako przykładowe.

W układzie przedstawionym na rys. 7.17a charakterystykę prądowo-napięciową obciążenia można otrzymać na podstawie rodziny charakterystyk wyjściowych tranzystora obciążającego. Na rys. 7.18 pokazano charakterystyki wyjściowe tranzystora



Rys. 7.18
Charakterystyki prądowo-napięciowe tranzystora obciążającego, wykreślone na rodzinie jego charakterystyk wyjściowych
1 — charakterystyka w zakresie nasylenia; 2 — charakterystyka w zakresie nienasylenia

obciążającego wraz z dwiema roboczymi charakterystykami prądowo-napięciowymi, typowymi dla pracy tego tranzystora w zakresie nasylenia (krzywa 1) lub nienasylenia (krzywa 2).

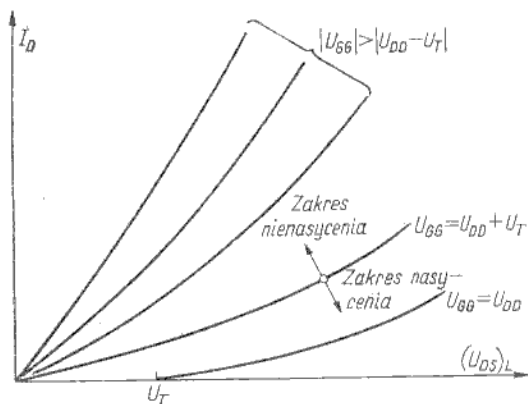
Dla określonych wartości napięć zasilania U_{GG} oraz U_{DD} różnica napięć ($U_{GS} - U_{DS}$) w całym obszarze pracy tranzystora obciążającego jest stała i równa różnicy napięć ($U_{GG} - U_{DD}$). Jak wiadomo, linia $U_{DS} = U_{GS} - U_T$ rozdziela rodzinę charakterystyk na zakresy nienasylenia i nasylenia. Warunki polaryzacji dla odpowiednich zakresów pracy można więc zapisać następująco:

zakres nasylenia: $|U_{DD}| \geq |U_{GG} - U_T|$

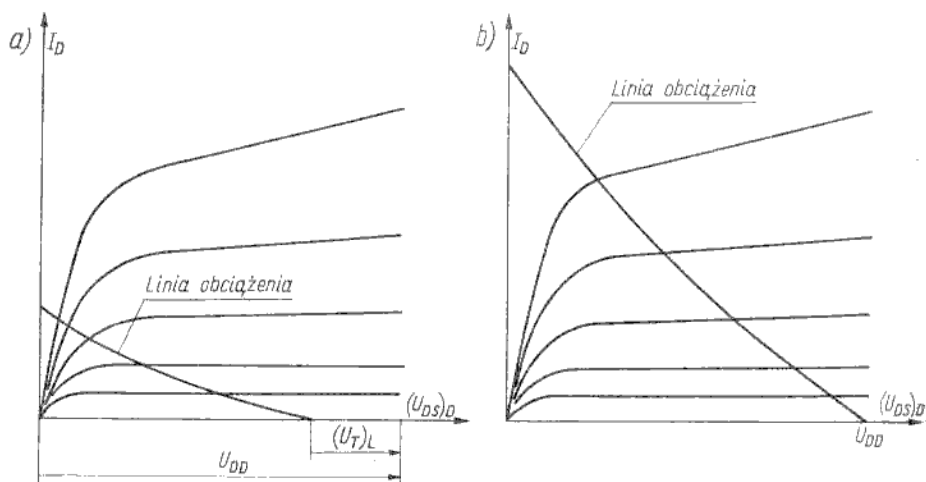
zakres nienasylenia: $|U_{DD}| < |U_{GG} - U_T|$

W szczególnym przypadku zasilania bramki i drenu ze wspólnego źródła ($U_{GG} = U_{DD}$) tranzystor obciążający pracuje w zakresie nasylenia (krzywa 1 na rys. 7.18). Typowy przebieg charakterystyk prądowo-napięciowych obciążenia przedstawiono na rys. 7.19. Dwie spośród tych charakterystyk, reprezentujące odpowiednio zakresy nasylenia i nienasylenia, naniesiono jako linie obciążenia na rodzinie charakterystyk wyjściowych tranzystora sterowanego (rys. 7.20a, b). Punkt przecięcia dowolnych dwu krzywych daje wspólne rozwiązanie układu równań opisujących stany tranzystorów sterowanego i obciążającego.

Wyznamy wzmocnienie napięciowe układu dla przypadku pracy obu tranzystorów (sterowanego i obciążającego) w zakresie nasylenia. Tranzystor obciąż-



Rys. 7.19
Charakterystyki prądowo-napięciowe tranzystora obciążającego



Ry. 7.20
Charakterystyki wyjściowe tranzystora sterowanego z prostymi obciążenia dla pracy tranzystora obciążającego w zakresach: a) nasycenia; b) nienasyceńia

zający pracuje w konfiguracji WG ze zwartym wyjściem, jego admittancej wejściowa zgodnie ze wzorem (7.62b) jest zatem równa $(g_m)_L$. Po podstawieniu do wzoru (7.52) $R_L = 1/(g_m)_L$ oraz przyjęciu $(g_d)_D \ll (g_m)_L$ otrzymuje się

$$k_u(0) = - \frac{(g_m)_D}{(g_m)_L} \tag{7.64}$$

Wzmocnienie napięciowe jest zatem równe ilorazowi transkonduktancji tranzystora sterowanego i obciążającego. Wzór (7.64) przy uwzględnieniu równań prądów drenu obu tranzystorów można przepisać w postaci

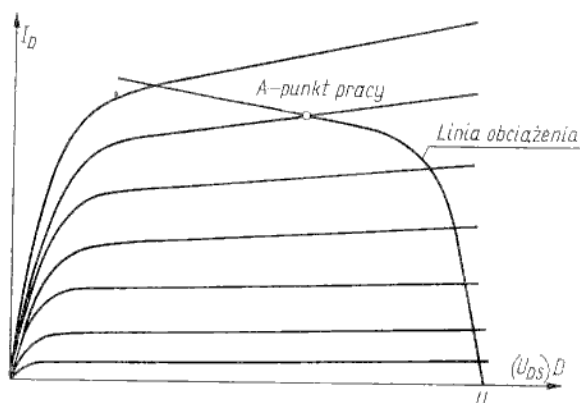
$$k_u(0) = - \sqrt{\frac{(\beta)_D}{(\beta)_L}} \tag{7.65}$$

Po podstawieniu $\beta = \mu C_i W / L_C$ oraz założeniu, że grubość dielektryka (pojemność C_i) i ruchliwości nośników są w obu tranzystorach jednakowe, ostatecznie otrzymuje się

$$k_u(0) = - \sqrt{\frac{\left(\frac{W}{L_C}\right)_D}{\left(\frac{W}{L_C}\right)_L}} \quad (7.66)$$

Wzmocnienie napięciowe zależy wyłącznie od rozmiarów geometrycznych obu tranzystorów. Stąd należy wnioskować o dobrej stabilności temperaturowej takiego wzmacniacza.

W układzie przedstawionym na rys. 7.17b tranzystor sterowany z kanałem typu p jest obciążony tranzystorem z kanałem typu n . Charakterystyki wyjściowe tranzystora sterowanego wraz z linią obciążenia ilustruje rys. 7.21. Linia obciążenia



Rys. 7.21
Charakterystyki prądowo-napięciowe układu z rys. 7.17b

jest jedną z charakterystyk wyjściowych tranzystora obciążającego, pracującego w tym przypadku jako ogranicznik prądu. Jeżeli punkt pracy znajduje się w zakresie nasycenia (punkt A na rys. 7.21), to zaletą takiego układu jest możliwość wykorzystania dużej dynamicznej rezystancji wyjściowej $(r_d)_L$ tranzystora obciążającego przy niewielkiej rezystancji statycznej obciążenia.

Podobne właściwości ma połączenie tranzystora sterowanego z kanałem wzbogacanym z tranzystorem obciążającym z kanałem zubożanym przy jednakowym typie przewodnictwa obu kanałów. Wówczas przy zwarciu bramki ze źródłem w tranzystorze obciążającym prosta obciążenia ma podobny przebieg jak na rys. 7.21.

Wzmacniacz z tranzystorem MOS i z tranzystorem bipolarnym (para hybrydowa)

7.6

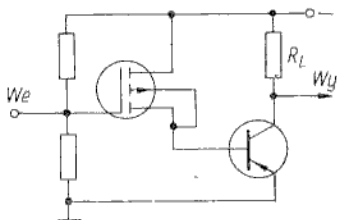
Korzystne właściwości tranzystorów MOS (przede wszystkim duża rezystancja wejściowa) oraz tranzystorów bipolarnych (mała rezystancja wyjściowa i małe napięcie nasycenia) wzajemnie uzupełniają się we wzmacniaczu stanowiącym połączenie hybrydowe dwu stopni, z których pierwszy zawiera tranzystor MOS, drugi zaś tranzystor bipolarny. Tranzystor MOS może pracować w układzie

WS lub WD, stosowanie układu WG jest nieuzasadnione ze względu na małą wartość impedancji wejściowej. Natomiast tranzystor bipolarny może być włączony w dowolnym z trzech układów, tj. WE, WC, WB. W związku z tym wyróżnia się sześć podstawowych konfiguracji: WS-WE, WS-WC, WS-WB, WD-WE, WD-WC, WD-WB. Właściwości pary hybrydowej (tranzystor MOS-tranzystor bipolarny) w znacznej mierze zależą od konfiguracji drugiego stopnia, dlatego omawiane wzmacniacze podzielimy na trzy grupy w zależności od układu włączenia tranzystora bipolarnego. Rozważania w całości dotyczą przypadku, gdy tranzystor MOS pracuje w zakresie nasycenia.

Wzmacniacz ze stopniem wyjściowym w układzie WE

7.6.1

Dobrze znane cechy układu WE, tj. duże wzmocnienie prądowe oraz umiarkowane i stosunkowo bliskie sobie wartości impedancji wejściowej i wyjściowej determinują właściwości konfiguracji WD-WE oraz WS-WE.



Rys. 7.22
Wzmacniacz w konfiguracji WD-WE

Najpierw rozpatrzmy układ WD-WE, którego schemat przedstawiono na rys. 7.22. Transkonduktancję g_{mh} pary hybrydowej wyznacza się w następujący sposób:

$$g_{mh} = \frac{I_{wy}}{U_{we}} = \frac{U_{wy1} h_{21e}}{U_{we} R_{we2}} = k_u \frac{h_{21e}}{R_{we2}} \quad (7.67)$$

przy czym: I_{wy} , U_{we} — składowe zmienne natężenia prądu oraz napięcia odpowiednio na wyjściu i wejściu pary hybrydowej; U_{wy1} — napięcie wyjściowe pierwszego stopnia, tj. na wyjściu tranzystora MOS; h_{21e} — współczynnik wzmocnienia prądowego tranzystora bipolarnego w układzie WE, R_{we2} — rezystancja wejściowa drugiego stopnia (z tranzystorem bipolarnym w układzie WE).

Czynnik U_{wy1}/U_{we} jest wzmocnieniem napięciowym układu WD. Zgodnie ze wzorem (7.56a)

$$k_u \approx \frac{g_m R_{we2}}{1 + g_m R_{we2}} \quad (7.68)$$

Stąd

$$g_{mh} \approx \frac{g_m h_{21e}}{1 + g_m R_{we2}} \quad (7.69)$$

Rezystancję wejściową drugiego stopnia R_{we2} można wyrazić w postaci zawierającej parametry czwórnikowe tranzystora bipolarnego. Zgodnie ze wzorem (7.27)

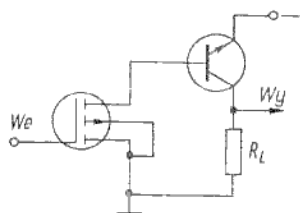
$$R_{we2} = h_{11e} - \frac{h_{21e} h_{12e}}{(G_L + h_{22e})}$$

W przypadku małych wartości g_m , tj. gdy $g_m R_{we2} \ll 1$, transkonduktancja g_{mh} dąży do minimalnej wartości równej $g_m h_{21e}$. Natomiast dla $g_m R_{we2} \gg 1$ transkonduktancja g_{mh} dąży do wartości maksymalnej równej h_{21e}/R_{we2} . Zatem g_{mh} może osiągać wartości w przedziale

$$g_m h_{21e} \leq g_{mh} \leq \frac{h_{21e}}{R_{we2}} \quad (7.70)$$

Na uwagę zasługuje fakt, że wartość maksymalna transkonduktancji pary hybrydowej zależy wyłącznie od parametrów tranzystora bipolarnego. Wzmocnienie napięciowe pary hybrydowej jest określone iloczynem transkonduktancji g_{mh} przez rezystancję wyjściową R_{wy} pary hybrydowej. Uwzględniając, że stopień WE jest sterowany ze źródła o rezystancji wewnętrznej $1/g_m$ (rezystancja wyjściowa stopnia WD), można określić rezystancję wyjściową R_{wy} pary hybrydowej w postaci zawierającej parametry czwórnikowe tranzystora bipolarnego

$$R_{wy} = \frac{h_{11e} + \frac{1}{g_m}}{h_{22e} h_{11e} - h_{12e} h_{21e} + \frac{h_{22e}}{g_m}} \quad (7.71)$$



Rys. 7.23
Wzmacniacz w konfiguracji WS-WE

W przypadku układu WS-WE (rys. 7.23) transkonduktancję g_{mh} wyznacza się w podobny sposób jak dla układu WD-WE. Po podstawieniu do wyrażenia (7.63) wzmocnienia napięciowego k_u układu WS (wzór (7.48)) otrzymuje się

$$g_{mh} = \frac{-g_m R_{we2}}{1 + g_d R_{we2}} \frac{h_{21e}}{R_{we2}} \quad (7.72)$$

Przy założeniu, że $1/R_{we2} \gg g_d$ wzór (7.72) upraszcza się do postaci

$$g_{mh} \approx -g_m h_{21e} \quad (7.72a)$$

Rezystancja wyjściowa pary hybrydowej WS-WE

$$R_{wy} = \frac{h_{11e} + r_d}{h_{22e}(h_{11e} + r_d) - h_{12e} h_{21e}} \quad (7.73)$$

i ze względu na dużą wartość r_d jest w przybliżeniu równa $1/h_{22e}$.

Wzmacniacz ze stopniem wyjściowym w układzie WB

7.6.2

Ze względu na małą wartość rezystancji wejściowej h_{11b} układu WB (h_{11b} jest mniejsze zarówno od $1/g_d$, tj. rezystancji wyjściowej układu WS, jak również

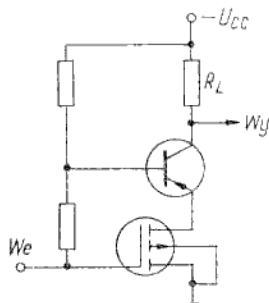
od $1/g_m$, tj. rezystancji wyjściowej układu WB) właściwości par hybrydowych WS-WB¹⁾ i WD-WB są niemal jednakowe.

Transkonduktację w obu przypadkach z dobrym przybliżeniem można zapisać jako

$$g_{mh} \approx g_m h_{21b} \quad (7.74)$$

Współczynnik wzmożenia prądowego h_{21b} układu WB jest bliski jedności, transkonduktancja wzmacniacza z wyjściem w układzie WB jest zatem znacznie mniejsza niż w przypadku wyjścia w układzie WE. Podstawową zaletą wzmacniaczy WS-WB oraz WD-WB jest ich unilateralność, tj. brak transmisji sygnału w kierunku od wyjścia do wejścia. Układ WB działa jako transformator rezystancji (z małej na dużą), a jego mała rezystancja wejściowa obciążając układ WS powoduje zmniejszenie wzmożenia napięciowego, a zatem praktycznie eliminuje zjawisko Millera (sprężenie zwrotne przez pojemność C_{gd}).

Rezystancja wyjściowa wzmacniacza WS-WB lub WD-WB jest w przybliżeniu równa rezystancji wyjściowej $1/h_{22b}$ układu WB. W tego rodzaju wzmacniaczu, mającym umiarkowaną wartość transkonduktancji oraz dużą rezystancję wyjściową, można tylko wykorzystać wzmożenie napięciowe przy odpowiednio dużej impedancji obciążenia. Na przykład wzór na wzmożenie napięciowe typowego



Rys. 7.24
Wzmacniacz w konfiguracji WS-WB (kaskoda)

układu kaskodowego, pokazanego na rys. 7.24, można w przybliżeniu zapisać w postaci

$$k_u \approx g_m \frac{R_L}{h_{22b} R_L + 1} \quad (7.75)$$

Wzmacniacz ze stopniem wyjściowym w układzie WC

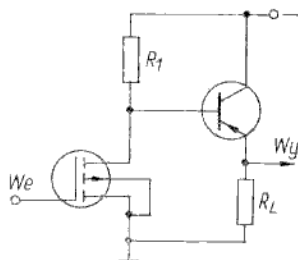
7.6.3

Cechą charakterystyczną wzmacniaczy WS-WC oraz WD-WC jest duża impedancja wejściowa i mała impedancja wyjściowa. Oba układy różnią się wzmożeniem napięciowym, które w przypadku pary WS-WC jest w przybliżeniu równe wzmożeniu pojedynczego stopnia WS, natomiast w układzie WD-WC jest nieco mniejsze od jedności.

¹⁾ Para hybrydowa WS-WB jest również nazywana *kaskodą*.

Para hybrydowa w układzie WS-WC

Wyznamy dwa najistotniejsze parametry układu WS-WC (rys. 7.25), tj. admitancję wyjściową oraz wzmocnienie napięciowe. Admitancję wyjściową



Rys. 7.25
Wzmacniacz w konfiguracji WS-WC

układu WC, obciążonego rezystorem R_L oraz sterowanego ze źródła o rezystancji wewnętrznej R_G , można zapisać w postaci ogólnej

$$Y_{wy} = h_{22c} + G_L - \frac{h_{12c}h_{21c}}{h_{11c} + R_G} \quad (7.76)$$

przy czym

$$h_{22c} = h_{22e}$$

$$h_{12c} = 1$$

$$-h_{21c} = 1 + h_{21e}$$

$$h_{11c} = h_{11e}$$

Rezystancję R_G w przypadku schematu przedstawionego na rys. 7.24 można wyrazić jako

$$R_G = \frac{r_d R_1}{r_d + R_1}$$

Zatem (7.76) przepisuje się w postaci

$$Y_{wy} = h_{22e} + G_L + \frac{h_{21e} + 1}{h_{11e} + \frac{r_d R_1}{r_d + R_1}} \quad (7.76a)$$

W rzeczywistym przypadku $G_L \gg h_{22e}$ oraz $h_{11e} \ll R_1 \ll r_d$. Jeżeli ponadto przyjmie się, że $R_1 \approx R_L$, to wyrażenie (7.76a) można uprościć do postaci

$$Y_{wy} \approx \frac{1 + h_{21e}}{R_1} \quad (7.76b)$$

Układ WC spełnia funkcję transformatora impedancji, zmniejszającego impedancję wyjściową pary hybrydowej WS-WC $1/(1 + h_{21e})$ razy w stosunku do impedancji wyjściowej układu WS.

Wzmocnienie napięciowe pary hybrydowej WS-WC z dobrym przybliżeniem jest równe iloczynowi transkonduktancji g_m układu WS przez rezystancję składającą się

z równolegle połączonych: rezystancji wyjściowej r_d układu WS, rezystancji R_1 oraz rezystancji wejściowej układu WC, równej w przybliżeniu $(1+h_{21e})R_L$. Stąd

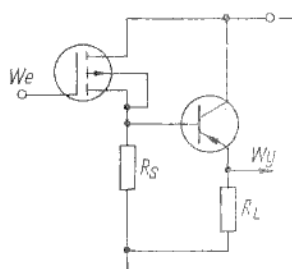
$$k_u = -g_m \left[g_d + \frac{1}{R_1} + \frac{1}{(h_{21e}+1)R_L} \right]^{-1} \quad (7.77)$$

lub w postaci uproszczonej

$$k_u \approx -g_m R_1 \quad (7.77a)$$

Para hybrydowa w układzie WD-WC

Wzmacniacz w układzie WD-WC, pokazany na rys. 7.26 jest również nazywany *wtórnikami hybrydowym*. Wzmocnienie napięciowe tego układu jest nieco mniejsze



Rys. 7.26
Wzmacniacz w konfiguracji WD-WC

od jedności, a jego zasadniczą funkcją jest transformacja impedancji z bardzo dużej na bardzo małą. Admitancję wyjściową można wyrazić w postaci analogicznej do wzoru (7.76a) przy oczywistej zamianie r_d na $1/g_m$.

Jeżeli dla uproszczenia przyjmie się, że $R_S = \infty$, to otrzymuje się

$$Y_{wy} \approx h_{22e} + G_L + \frac{(h_{21e}+1)g_m}{g_m h_{11e} + 1} \quad (7.78)$$

Wzmocnienie napięciowe jest wyrażone wzorem analogicznym do (7.77)

$$k_u \approx \frac{g_m (h_{21e} + 1) R_L}{g_m (h_{21e} + 1) R_L + 1} \quad (7.79)$$

Zadaniem rezystora R_S , występującego w układzie na rys. 7.26, jest zwiększenie prądu drenu baz nadmiernego wzrostu prądu emitera. W ten sposób uzyskuje się większą transkonduktancję, mniejszą rezystancję wyjściową oraz lepszą stabilność temperaturową.

Tranzystor MOS jako rezystor sterowany napięciem

7.7

Bardzo cenną zaletą tranzystorów MOS jest możliwość ich stosowania w różnych układach jako rezystorów liniowych o rezystancji sterowanej napięciem. W najprostszym przypadku korzysta się z niewielkiego odcinka liniowego charakterystyki wyjściowej (zakres nienasyceń) przy napięciu U_{DS} nieznacznie różnym od zera.

Wyrażenie na charakterystykę wyjściową ma postać

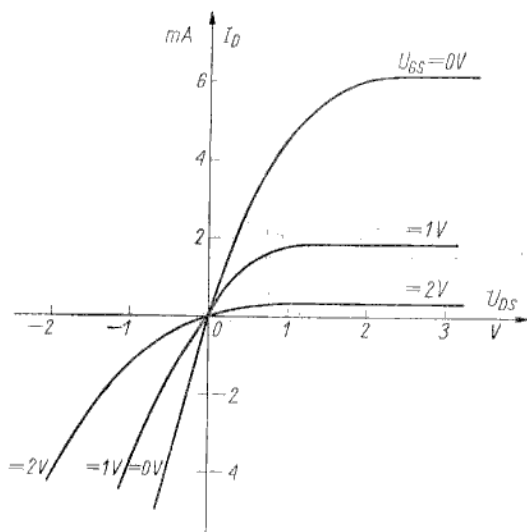
$$I_D = \beta \left[(U_{GS} - U_T)U_{DS} - \frac{U_{DS}^2}{2} \right] \quad (7.80)$$

Dla małych wartości napięcia drenu ($|U_{DS}| \ll |U_{Dsat}|$) drugi człon wyrażenia (7.80) można pominąć. Wówczas przy stałej wartości napięcia bramka-źródło U_{GS} prąd drenu I_D jest liniowo zależny od napięcia drenu U_{DS}

$$I_D \approx \beta(U_{GS} - U_T)U_{DS} \quad (7.80a)$$

Stąd

$$r_{DS} \approx \frac{1}{\beta(U_{GS} - U_T)} \quad (7.81)$$



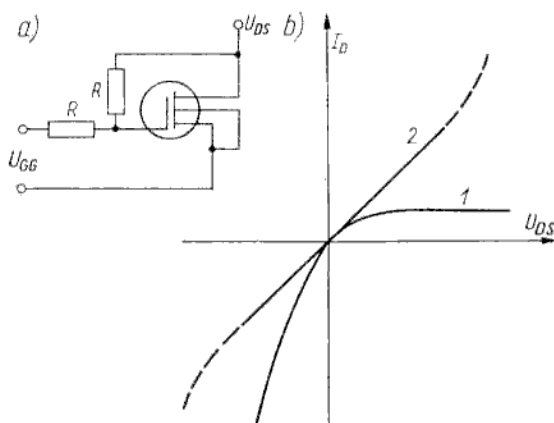
Rys. 7.27
Charakterystyki $I_D(U_{DS})$ wykreślone dla obu polaryzacji napięcia U_{DS}

Charakterystyki $I_D = f(U_{DS})$, wykreślone dla obu polaryzacji napięcia U_{DS} na rys. 7.27 (tranzystor z kanałem zubożonym typu n), wskazują na podstawowe ograniczenia tego przypadku realizacji funkcji rezystora, tj.:

- mały zakres liniowych zmian prądu I_D w funkcji napięcia U_{DS} ,
 - niesymetria charakterystyk względem punktu $U_{DS} = 0$, $I_D(U_{DS}) \neq -I_D(-U_{DS})$.
- W celu usunięcia tych ograniczeń można zastosować odpowiednie stałoprądowe sprzężenie zwrotne dren-bramka.

Wskutek zastosowania stałoprądowego sprzężenia zwrotnego (rys. 7.28a) napięcie bramka-źródło

$$U_{GS} = \frac{U_{GG} + U_{DS}}{2} \quad (7.82)$$



Rys. 7.28

Układ włączenia tranzystora MOS ze sprzężeniem zwrotnym (a) oraz porównanie charakterystyk $I_D = f(U_{DS})$ (b) w układzie bez sprzężenia zwrotnego (krzywa 1) oraz ze sprzężeniem zwrotnym (krzywa 2)

Po podstawieniu (7.82) do (7.80) otrzymuje się

$$I_D = \frac{\beta}{2} (U_{GG} - 2U_T) U_{DS} \quad (7.83)$$

Stąd

$$r_{DS} = \frac{2}{\beta(U_{GG} - 2U_T)} \quad (7.84)$$

Zastosowanie sprzężenia zwrotnego znacznie rozszerza odcinek liniowy charakterystyki $I_D = f(U_{DS})$ oraz wprowadza symetrię tej charakterystyki względem punktu $U_{DS} = 0$. Na rysunku 7.28b porównano kształt charakterystyki $I_D = f(U_{DS})$ dla przypadku układu bez sprzężenia zwrotnego (krzywa 1) oraz ze sprzężeniem zwrotnym (krzywa 2).

W rozdziałach 8, 9, 10 są rozpatrywane układy scalone, przy czym niniejszy rozdział zawiera elementarne informacje z dziedziny technologii wytwarzania układów scalonych. Zanim przejdziemy do szczegółowego omawiania tej tematyki, zostanie przedstawiony najpierw krótki zarys historii rozwoju technologii układów scalonych i zdefiniujemy niektóre pojęcia podstawowe. Trudno jest określić ścisłą datę „narodzin” układów scalonych. Na ogół uważa się, że jedną z pierwszych idei układu scalonego sformułował G. W. A. Dummer w 1952 r. Prognozując rozwój technologii półprzewodnikowej stwierdził on (w wolnym przekładzie): „Wraz z odkryciem tranzystora i ogólnym rozwojem technologii półprzewodnikowej zaistniały przesłanki do rozważenia możliwości realizacji urządzeń elektronicznych w postaci bloków (brył) ciała stałego, pozbawionych połączeń drutowych. Bryła ciała stałego może się składać z warstw spełniających funkcje izolujące, przewodzące, prostownicze i wzmacniające, przy czym występują bezpośrednie połączenia funkcjonalne wydzielonych obszarów ciała stałego, zawierających różne warstwy”.

Pierwszy laboratoryjny układ scalony wykonał J. Kilby (firma „Texas Instruments”) w 1958 r. Był to przerzutnik (dwa tranzystory, dwa kondensatory, osiem rezystorów), wykonany za pomocą technologii mesa w kryształce krzemu o powierzchni $5 \times 2,5$ mm, przy czym poszczególne elementy były połączone drutem złotym za pomocą termokompresji. Niedługo po wynalazku Kilby’ego, Noyce ze współpracownikami (firma „Fairchild”) opracował doskonalszą technologię wytwarzania układów scalonych, w której poszczególne elementy były odizolowane wzajemnie od siebie za pomocą dyfuzyjnych złączy *p-n*, a połączenia międzyelementowe były wykonane w postaci warstwy metalu naparowanej w próżni. Dopiero ta technologia stwarzała realne przesłanki techniczno-ekonomiczne (odpowiednio duży uzysk produkcyjny i małe koszty) do produkcji układów scalonych i z pewnymi modyfikacjami została wdrożona na skalę przemysłową. Trudno zatem odpowiedzieć na pytanie: Kilby czy Noyce jest wynalazcą układu scalonego? Firmy Fairchild i Texas Instruments szukały nawet rozstrzygnięcia tej spornej kwestii w drodze sądowej, przy czym Sąd Najwyższy USA przyznał patent Noyce’owi. Jednakże „era” układów scalonych w sensie ich wielkoseryjnej produkcji przemysłowej rozpoczęła się nieco później, a mianowicie w 1962 r. po odkryciu przez J. A. Hoerniego właściwości maskujących warstwy SiO_2 i wprowadzeniu technologii planarnej¹⁾.

Odkrycie układów scalonych nadało jakościowo nowy kierunek rozwojowi elektroniki. Powstał nowy dział elektroniki, nazywany *mikroelektroniką*. Pojęcie mikroelektronika w ciągu kilkunastu lat przeszło znaczną ewolucję, gdyż początkowo

¹⁾ J. A. Hoerni po raz pierwszy przedstawił zasady technologii planarnej w referacie: „Planar Silicon Transistors and Diodes” — IRE Electron Devices Meeting, Washington, D. C. (1960).

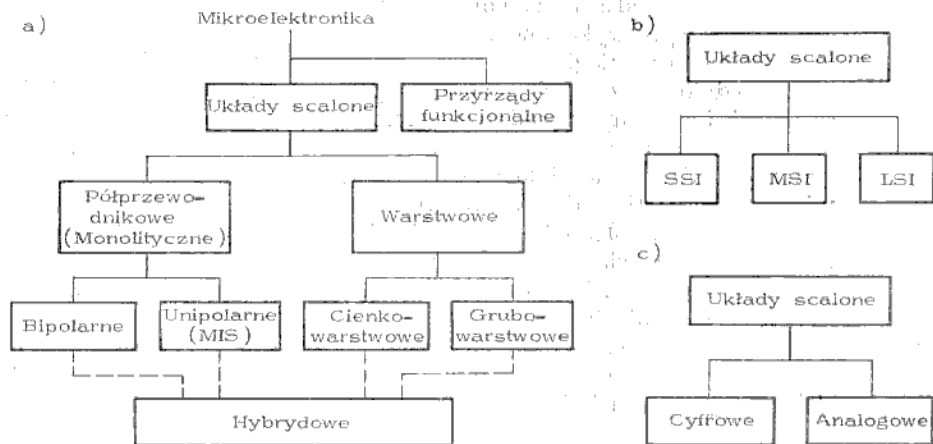
obejmowało swym znaczeniem zarówno tzw. mikroelektronikę dyskretną (miniaturyzacja elementów indywidualnych — minimoduły i mikromoduły produkowane w latach pięćdziesiątych), jak również mikroelektronikę integralną (układy scalone i przyrządy funkcjonalne). Współczesne rozumienie pojęcia mikroelektronika jest ograniczone do zakresu mikroelektroniki integralnej zgodnie z następującą definicją: „Mikroelektronika jest dziedziną nauki i techniki zajmującą się zminiaturyzowanymi układami elektronicznymi, realizowanymi w postaci scalonej przy całkowitym lub częściowym braku elementów indywidualnych (dyskretnych)”. Mikroelektronika obejmuje zatem zespół zagadnień fizycznych, technologiczno-konstrukcyjnych i układowych, związanych z analizą, badaniem właściwości, projektowaniem, wytwarzaniem i zastosowaniami układów scalonych lub (i) przyrządów funkcjonalnych. W terminie mikroelektronika znajdują odbicie dwie podstawowe tendencje rozwojowe współczesnej elektroniki, tj. dążenie do scalenia (elementów lub funkcji układowych) i mikrominiaturyzacji, które służą osiągnięciu następujących celów:

- poprawa niezawodności i jakości parametrów elektrycznych;
- obniżenie kosztów wytwarzania dzięki automatyzacji produkcji i mniejszemu zużyciu materiałów;
- zmniejszenie objętości i ciężaru urządzeń.

Nieprzypadkowo klasyczny cel miniaturyzacji wymieniono na ostatnim miejscu, gdyż małe rozmiary i ciężar układów mikroelektronicznych można uznać za cechy wtórne, podporządkowane dwóm pierwszym celom. Najbardziej ogólnie można stwierdzić, że tendencją dynamizującą rozwój mikroelektroniki jest dążenie do osiągnięcia korzyści ekonomicznych, tj. poprawy stosunku walorów użytkowych do kosztów, przy czym w kosztach należy uwzględnić nie tylko ceny układów mikroelektronicznych, lecz również wydatki związane z projektowaniem, montażem i eksploatacją urządzeń.

Przejdźmy teraz do definicji pojęć szczegółowych, występujących w schematycznej klasyfikacji kierunków mikroelektroniki, przedstawionej na rys. 8.1a.

Układ scalony jest to mikrominiaturowy układ elektroniczny, charakteryzujący się tym, że wszystkie lub część elementów wraz z połączeniami są wykonane nierozłącznie (w sensie możliwości demontażu) w jednym cyklu technologicznym wewnątrz



Rys. 8.1
Klasyfikacja kierunków mikroelektroniki

lub na powierzchni wspólnego podłoża. Układ scalony ma obudowę hermetyczną, chroniącą go przed wpływem czynników mechaniczno-klimatycznych.

Przyrząd funkcjonalny jest to przyrząd realizujący funkcję układową wskutek bezpośredniego wykorzystania określonych właściwości fizycznych materii. Mówi się, że układ scalony ma strukturę topologiczną, a przyrząd funkcjonalny — morfologiczną. Strukturę topologiczną rozumie się w ten sposób, że układ scalony ma określony schemat elektryczny, przy czym poszczególnym elementom tego schematu odpowiadają ściśle określone fragmenty (obszary) bryły ciała stałego. Struktura morfologiczna natomiast charakteryzuje się tym, że nie można w niej wyodrębnić obszarów odpowiadających elementom według klasycznej teorii układów, traci wówczas sens pojęcie schematu ideowego. Bodajże najbardziej dydaktycznym przykładem przyrządu funkcjonalnego o strukturze morfologicznej jest kryształ kwarcu, który spełnia funkcję obwodu rezonansowego L, C , lecz nie ma sensu doszukiwać się w nim wyodrębnionych obszarów spełniających funkcje cewki lub kondensatora. Przyrządy funkcjonalne to przyszłość mikroelektroniki, której zwiastunami są takie przyrządy jak dioda Reada lub generator Gunna. Będziemy je omawiać w rozdziale 12, w tym rozdziale natomiast rozważania ograniczymy do zagadnień dotyczących układów scalonych.

Układy scalone można różnie klasyfikować w zależności od przyjętych kryteriów.

Pod względem cech technologiczno-konstrukcyjnych układy scalone dzieli się ogólnie na półprzewodnikowe (monolityczne) i warstwowe¹⁾ (rys. 8.1a).

W zależności od stopnia rozbudowania układy scalone dzieli się na (rys. 8.1b):

- SSI (ang. *Small Scale Integration*) — o małym stopniu scalenia, tj. do dziesięciu bitów lub innych komórek elementarnych w jednej mikroplycie (inaczej do kilkudziesięciu elementów);
- MSI (ang. *Medium Scale Integration*) — o średnim stopniu scalenia, tj. od dziesięciu do stu bitów (inaczej do kilkuset elementów);
- LSI (ang. *Large Scale Integration*) — o dużym stopniu scalenia, tj. powyżej stu bitów (inaczej powyżej kilkuset elementów).

W zależności od zastosowań układy scalone dzieli się na cyfrowe i analogowe (rys. 8.1c).

W tym rozdziale interesują nas wyłącznie zagadnienia technologiczne i konstrukcyjne, dlatego uwagę skoncentrujemy tylko na klasyfikacji podanej na rys. 8.1a. Najpierw zdefiniujemy układy scalone półprzewodnikowe i warstwowe.

Półprzewodnikowym układem scalonym (inaczej nazywanym monolitycznym) jest układ, w którym zarówno elementy czynne, jak i bierne są ukształtowane w monolitycznej płytce półprzewodnikowej za pomocą odpowiednich procesów fizykochemicznych, przy czym przejście od jednego elementu do drugiego odbywa się z zachowaniem ciągłości mechanicznej kryształu, a zmienia się tylko struktura fizykochemiczna materiału.

Układem scalonym warstwowym nazywa się układ, w którym elementy bierne są wytworzone w postaci warstw różnych materiałów, naniesionych w wyniku odpowiednich procesów fizykochemicznych na biernym podłożu izolacyjnym. Elementy czynne (półprzewodnikowe) są dolutowywane.

Półprzewodnikowe układy scalone dzieli się na *bipolarne* i *unipolarne* w zależności od tego, jaki rodzaj tranzystorów stanowi podstawowe elementy czynne. Układy scalone warstwowe dzieli się na *cienkowarstwowe* i *grubowarstwowe* w zależności od tego, czy grubości warstw są mniejsze, czy też większe niż kilka mikrometrów (zwy-

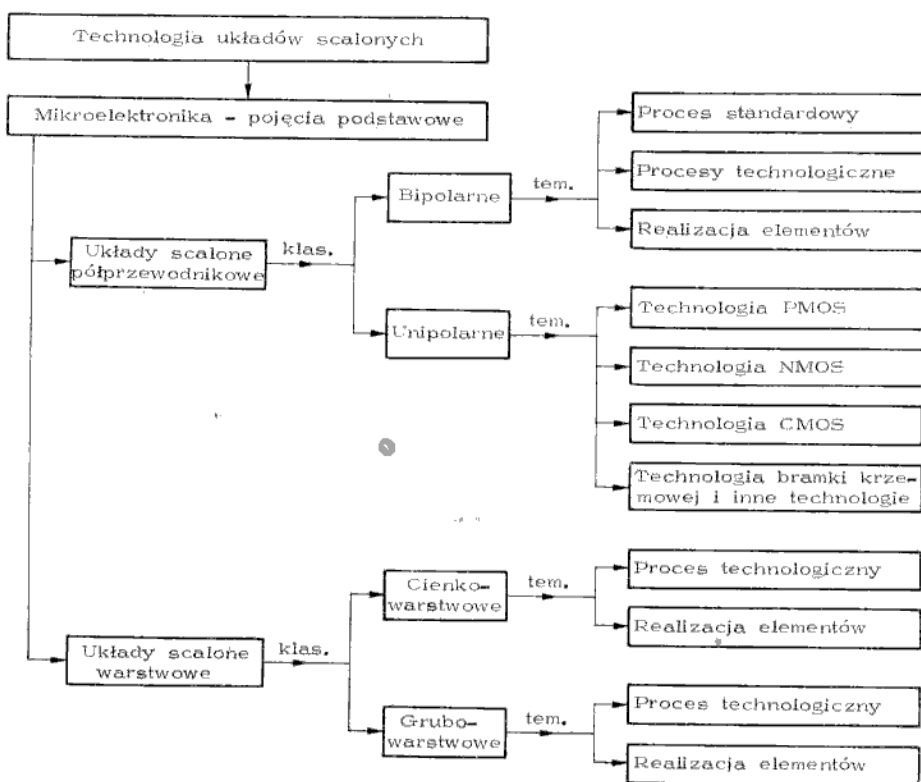
¹⁾ Ściśle biorąc, taki podział nie ma wspólnego kryterium, gdyż pierwsza nazwa dotyczy rodzaju materiału (półprzewodnik), a druga wyszczególnia cechę konstrukcyjną (warstwy). Jest to jednak terminologia utrwalona historycznie.

kle przyjmuje się kryterium grubości $2 \mu\text{m}$). To proste kryterium ilościowe nie oddaje pełnej istoty podziału, gdyż między technologią cienko- i grubowarstwową istnieją jakościowe różnice technologiczne.

Charakteryzując najogólniej różnice między układami scalonymi półprzewodnikowymi a warstwowymi należy stwierdzić, że w pierwszych występują znaczne trudności w uzyskaniu elementów biernych o szerokim zakresie wąsko stolerowanych wartości nominalnych, w drugim zaś nie ma dotychczas możliwości wytwarzania zadowalająco stabilnych elementów czynnych¹⁾.

Pozytywne cechy obu technik łączą w sobie układy hybrydowe, które zawierają wysokiej jakości warstwowe elementy bierne oraz półprzewodnikowe elementy czynne. Ponieważ układy warstwowe „nie mogą się obyć” bez półprzewodnikowych elementów czynnych (z wyjątkiem układów spełniających funkcje pasywne, np. dzielniki, filtry itp.), często zatem pojęcie „układy hybrydowe” jest utożsamiane z pojęciem „układy warstwowe”.

Na podstawie klasyfikacji przedstawionej na rys. 8.1a w kolejnych punktach tego rozdziału omówimy poszczególne technologie wytwarzania układów scalonych. Zakłada się, że Czytelnik zna podstawowe procesy technologiczne, opisane w rozdziale 2. Układ rozdziału przedstawiono na rys. 8.2.



Rys. 8.2
Schematyczny układ rozdziału

¹⁾ Prowadzone od wielu lat prace nad tranzystorem cienkowarstwowym nie dały dotychczas zadowalających rezultatów.

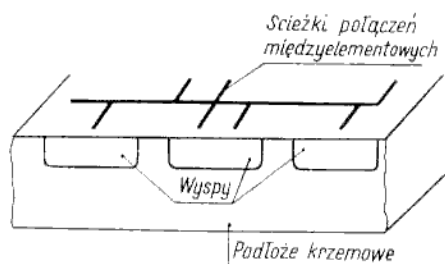
Półprzewodnikowe układy scalone

8.1

Układy scalone bipolarne

8.1.1

Układy scalone bipolarne (USB) są wytwarzane wyłącznie w podłożu krzemowym technologią epiplanarną, której podstawy omówiono w rozdziale 2. Na rysunku 8.3 przedstawiono szkicowo (bez uwzględniania szczegółów) ideę struktury



Rys. 8.3
Szkic struktury topologicznej układu scalonego bipolarnego

topologicznej USB. Każdy element znajduje się w odizolowanej wyspie i jest połączony z innymi elementami (zgodnie ze schematem elektrycznym układu) za pomocą ścieżek metalizacji naniesionej na górnej powierzchni płytki. Przechodząc do rozważań szczegółowych należy wyjaśnić przede wszystkim następujące zagadnienia:

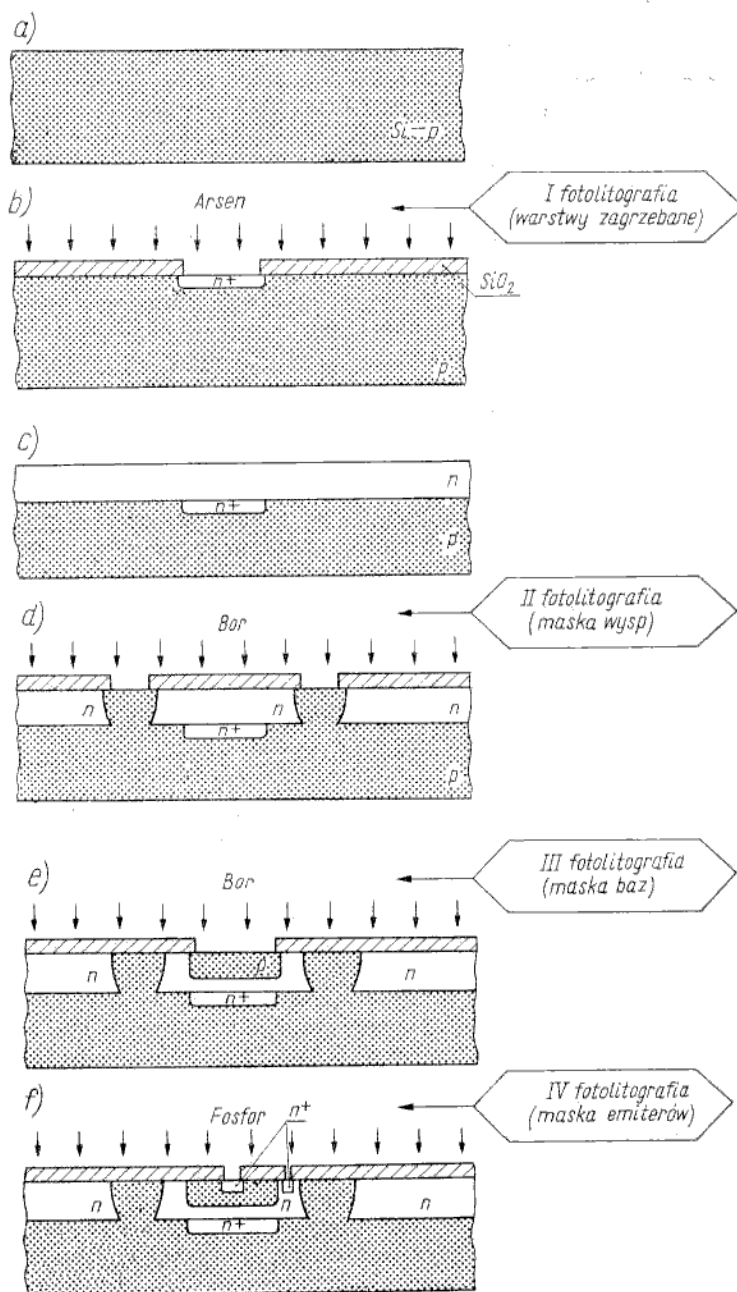
- sposoby izolacji elementów,
- sposoby realizacji poszczególnych elementów.

Te dwa zagadnienia zostaną systematycznie rozpatrzone po wstępnym omówieniu standardowego procesu wytwarzania USB. Ponieważ najbardziej złożonym elementem jest tranzystor, najpierw przedstawimy proces wytwarzania układu scalonego z uwzględnieniem tylko tego elementu.

Standardowy proces wytwarzania USB

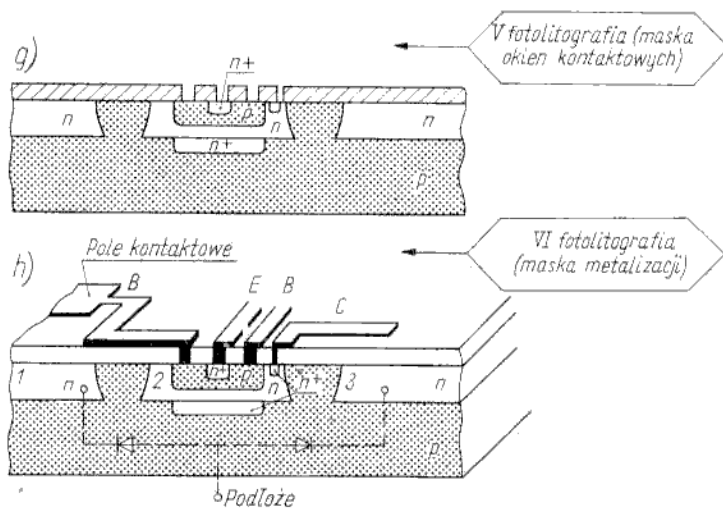
8.1.1.1

Podstawowe fazy standardowego procesu wytwarzania USB przedstawiono na rys. 8.4. Należy najpierw zwrócić uwagę na końcową strukturę pokazaną na rys. 8.4h. Trzy wyspy 1, 2, 3 typu n wykonane w podłożu typu p są odizolowane wzajemnie, gdyż każde dwie wyspy tworzą wraz z podłożem dwa złącza $n-p$ połączone szeregowo przeciwstawnie, jak to pokazano przykładowo dla wysp 1, 3. Przy dowolnej zatem polaryzacji wzajemnej dwóch wysp zawsze jedno ze złączy pracuje w kierunku zaporowym. Najczęściej stosuje się polaryzację podłoża największym napięciem ujemnym, jakim dysponuje się przy określonym zasilaniu układu. Wówczas oba złącza są zawsze spolaryzowane w kierunku zaporowym. Ten sposób izolacji, nazywany *izolacją złączową*, zapewnia rezystancję między wyspami rzędu $10^9 \Omega$, gdyż tyle wynosi rezystancja złącza $p-n$ spolaryzowanego zaporowo. W środkowej wyspie 2 jest wykonany tranzystor. W odróżnieniu od tranzystora indywidualnego (porównaj z rys. 5.9) ma on warstwę n^+ pod obszarem kolektorowym typu n . Ta warstwa nosi nazwę *warstwy zagrzebanej* (ang. *buried layer*), gdyż jest ulokowana „pod spodem” struktury tranzystora. Warstwa zagrzebana spełnia taką funkcję jak podłoże n^+ w tranzystorze indywidualnym, tj. zmniejsza rezystancję szeregową kolektora, czyli ma istotny wpływ na zmniejszenie napięcia nasycenia tranzystora.



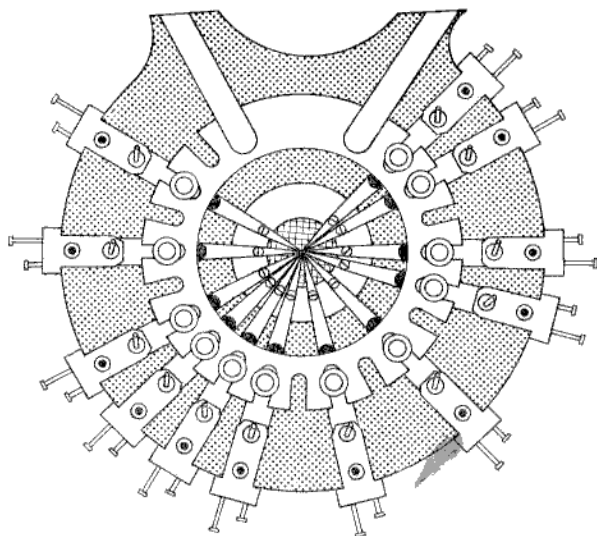
Rys. 8.4
Podstawowe fazy standardowego procesu wytwarzania USB

Prześledźmy teraz ważniejsze fazy cyklu wytwarzania USB zgodnie z rys. 8.4. Podłożem układu scalonego jest płytka krzemu typu *p* (rys. 8.4a), która przede wszystkim spełnia funkcję nośnika mechanicznego, a jej parametry elektrofizyczne nie muszą być ściśle kontrolowane (zwykle $N_A \approx 10^{21} \text{ m}^{-3}$). Górna powierzchnia płytki jest starannie polerowana, gdyż przez tę powierzchnię będą przeprowadzane dalsze procesy obróbcze. Jakość dolnej powierzchni jest mało istotna. Po pierwszym procesie fotolitografii (kolejne fazy procesu fotolitografii były omawiane w rozdziale 2 — por. rys. 2.24) płytka jest pokryta warstwą SiO_2 z wytrawionymi oknami w miejscach odpowiadających lokalizacji tranzystorów. Przez te okna jest przeprowadzana dyfuzja domieszki donorowej w celu wytworzenia warstw zagrzebanych (rys. 8.4b). Jako domieszkę stosuje się arsen lub antymon, gdyż współczynnik dyfuzji tych domieszek ma wartości ok. 10 razy mniejsze niż dla fosforu. Chodzi o to, by warstwa zagrzebana w kolejnych obróbkach wysokotemperaturowych (utleniania i dyfuzji) nie „rozplywała się” w głąb płytki. W następnej fazie (rys. 8.4c) na powierzchni płytki wytwarza się warstwę epitaksjalną typu *n*. Grubość tej warstwy wynosi 5...20 μm . Po drugiej fotolitografii (rys. 8.4d) są otwierane okna, przez które przeprowadza się dyfuzję boru w celu „przebiecia” warstwy epitaksjalnej obszarami typu *p*, czyli w celu utworzenia odizolowanych wysp typu *n*. Po trzeciej fotolitografii są otwierane okna, przez które przeprowadza się dyfuzję boru w celu utworzenia obszaru typu *p* spełniającego funkcję bazy w tranzystorze (rys. 8.4e). Po czwartej fotolitografii są otwierane okna w celu przeprowadzenia dyfuzji fosforu do obszaru emitera i kolektora (rys. 8.4f). W ten sposób powstaje warstwa emiterowa n^+ oraz warstwa n^+ w obszarze kolektora, ułatwiająca uzyskanie styku omowego półprzewodnika z kontaktem metalicznym. W piątym procesie fotolitografii są otwierane okna w tych miejscach, w których będą wykonane kontakty półprzewodnika z metalicznymi ścieżkami połączeniowymi (rys. 8.4g). Są to tzw. okna kontaktowe emitera, bazy i kolektora. W celu zwiększenia szybkości działania tranzystorów zwykle między operacjami przedstawionymi na rys. 8.4f i g przeprowadza się dyfuzję złota do całej płytki (przez dolną powierzchnię). Atomy złota spełniają



Rys. 8.4 (ed.)

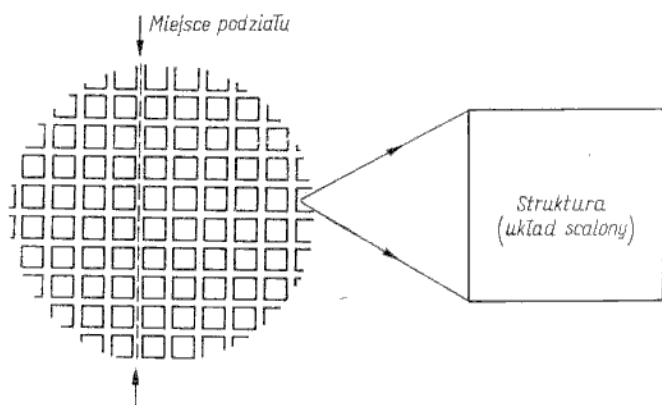
funkcję centrów generacyjno-rekombinacyjnych, tj. powodują zmniejszenie czasu życia nośników, czyli mają korzystny wpływ na zwiększenie szybkości działania układu scalonego. Po etapie przedstawionym na rys. 8.4g na całą powierzchnię płytki naporowuje się warstwę metalu (najczęściej aluminium). W szóstym procesie fotolitografii następuje wytrawienie aluminium tak, by pozostały tylko ścieżki połączeń międzyelementowych i pola kontaktowe służące do wykonania połączeń struktury układu scalonego z wyprowadzeniami zewnętrznymi (rys. 8.4h). Często wykonuje się jeszcze pasywację, tj. pokrywa się powierzchnię płytki warstwą szkliwa; wówczas przeprowadza się siódmy proces fotolitografii w celu otworzenia okien do pól kontaktowych. Zadaniem tej pasywacji jest ochrona powierzchni płytki od uszkodzeń mechanicznych (zadrapań itp.) w kolejnych procesach obróbkowych. Opisany dotychczas cykl technologiczny obejmuje główne etapy obróbki całej płytki, w której jednocześnie wykonuje się kilkadziesiąt do kilkuset struktur układów scalonych. Złożoność technologii określa liczba procesów fotolitografii, których w tym przypadku jest sześć i ewentualny siódmy, związany z pasywacją całego układu. Gdyby wymienić szczegółowo wszystkie operacje technologiczne, to byłoby ich około trzystu. Płytką krzemowa z wytworzonymi układami scalonymi podlega testowaniu. W zautomatyzowanym stanowisku pomiarowym (testerze) są wykonywane pomiary podstawowych parametrów elektrycznych wszystkich układów scalonych na płycie. Tester jest wyposażony w tzw. sondy ostrzowe, które są dociskane do odpowiednich pól kontaktowych (rys. 8.5). Układy wadliwe są oznaczane



Rys. 8.5
Widok głowicy testera wyposażonej w sondy ostrzowe

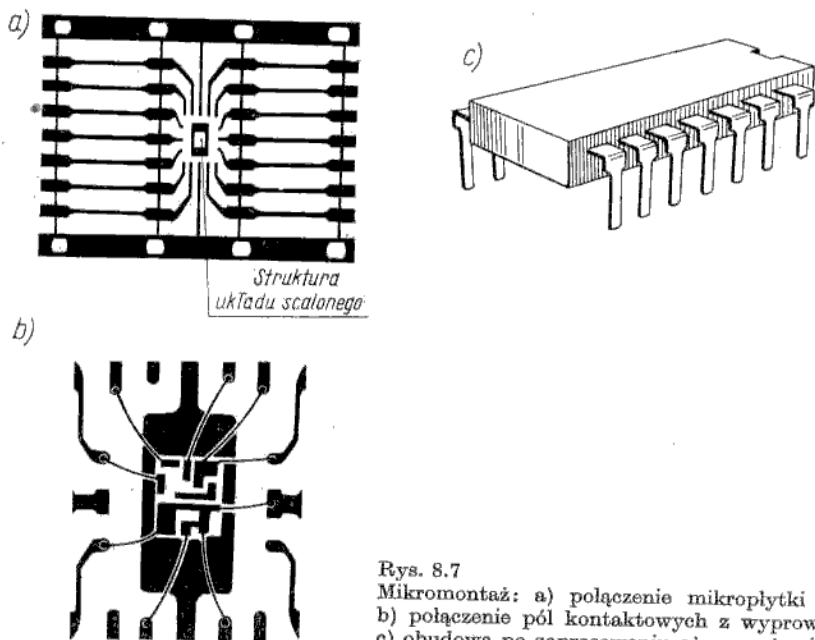
kolorowym tuszem i nie są poddawane dalszym operacjom technologicznym. Jest to bardzo ważne, gdyż końcowe operacje (mikromontaż) są najbardziej pracochłonne, a więc najdroższe.

Po testowaniu płytka krzemowa jest cięta na mikroplątki, przy czym każda mikroplątka zawiera pojedynczą strukturę układu scalonego (rys. 8.6). Najczęściej dzielenie płytki na mikroplątki wykonuje się metodą rysowania po powierzchni płytki ostrzem diamentowym i lupania. Tę operację wykonuje się przy zastosowaniu



Rys. 8.6
Cięcie płytki na mikroplątki (struktury)

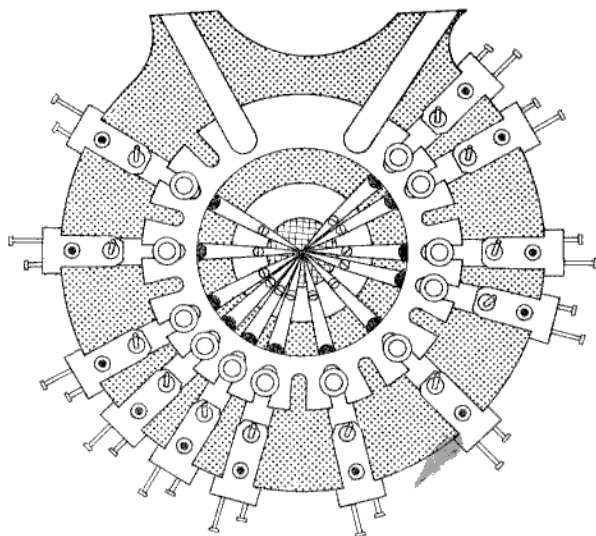
urządzeń automatycznych. Ostatnie operacje technologiczne polegają na umieszczeniu struktury układu scalonego w obudowie. Jest to tzw. mikromontaż, w którym można wyróżnić trzy operacje (rys. 8.7):



Rys. 8.7
Mikromontaż: a) połączenie mikroplątki z ażurem;
b) połączenie pól kontaktowych z wyprowadzeniami;
c) obudowa po zaprasowaniu ażuru w żywicy

- lutowanie albo klejenie struktury układu scalonego (mikroplątki) do obudowy lub ażuru będącego „szkieletem” obudowy plastikowej;
- termokompresja lub zgrzewanie ultradźwiękowe, za pomocą którego łączy się pola kontaktowe struktury układu scalonego z nóżkami (wyprowadzeniami zewnętrznymi) obudowy;

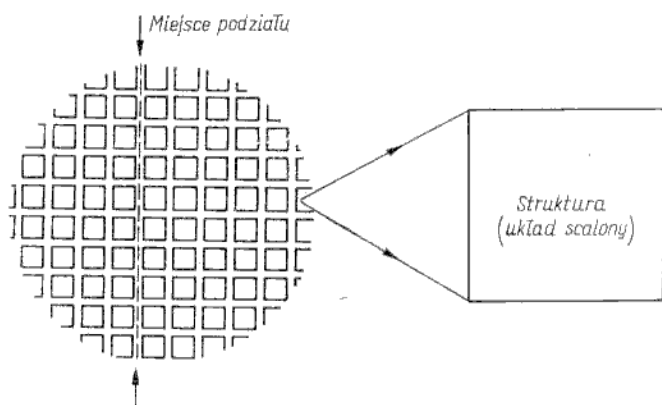
funkcję centrów generacyjno-rekombinacyjnych, tj. powodują zmniejszenie czasu życia nośników, czyli mają korzystny wpływ na zwiększenie szybkości działania układu scalonego. Po etapie przedstawionym na rys. 8.4g na całą powierzchnię płytki naporowuje się warstwę metalu (najczęściej aluminium). W szóstym procesie fotolitografii następuje wytrawienie aluminium tak, by pozostały tylko ścieżki połączeń międzyelementowych i pola kontaktowe służące do wykonania połączeń struktury układu scalonego z wyprowadzeniami zewnętrznymi (rys. 8.4h). Często wykonuje się jeszcze pasywację, tj. pokrywa się powierzchnię płytki warstwą szkliwa; wówczas przeprowadza się siódmy proces fotolitografii w celu otworzenia okien do pól kontaktowych. Zadaniem tej pasywacji jest ochrona powierzchni płytki od uszkodzeń mechanicznych (zadrapań itp.) w kolejnych procesach obróbkowych. Opisany dotychczas cykl technologiczny obejmuje główne etapy obróbki całej płytki, w której jednocześnie wykonuje się kilkadziesiąt do kilkuset struktur układów scalonych. Złożoność technologii określa liczba procesów fotolitografii, których w tym przypadku jest sześć i ewentualny siódmy, związany z pasywacją całego układu. Gdyby wymienić szczegółowo wszystkie operacje technologiczne, to byłoby ich około trzystu. Płytkę krzemową z wytworzonymi układami scalonymi podlega testowaniu. W zautomatyzowanym stanowisku pomiarowym (testerze) są wykonywane pomiary podstawowych parametrów elektrycznych wszystkich układów scalonych na płycie. Tester jest wyposażony w tzw. sondy ostrzowe, które są dociskane do odpowiednich pól kontaktowych (rys. 8.5). Układy wadliwe są oznaczane



Rys. 8.5
Widok głowicy testera wyposażonej w sondy ostrzowe

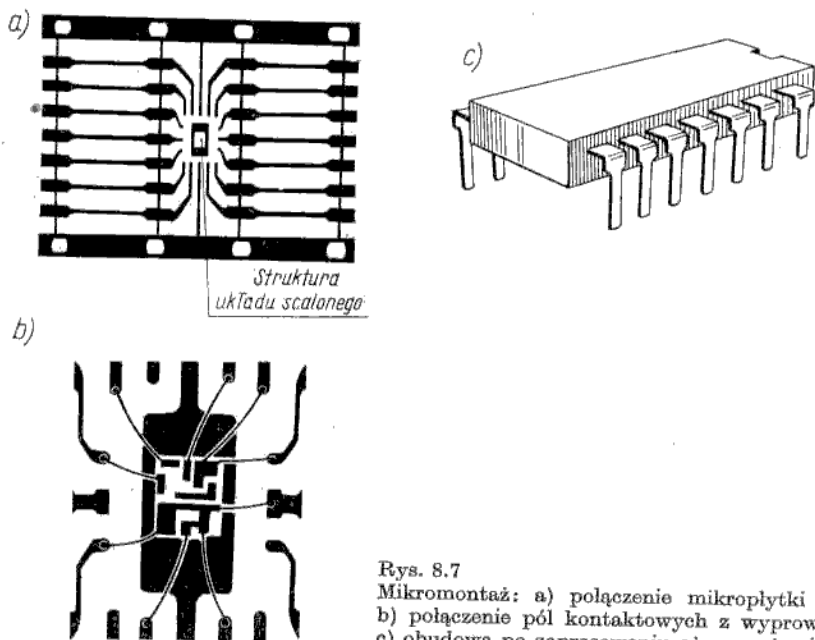
kolorowym tuszem i nie są poddawane dalszym operacjom technologicznym. Jest to bardzo ważne, gdyż końcowe operacje (mikromontaż) są najbardziej pracochłonne, a więc najdroższe.

Po testowaniu płytka krzemowa jest cięta na mikroplątki, przy czym każda mikroplątka zawiera pojedynczą strukturę układu scalonego (rys. 8.6). Najczęściej dzielenie płytki na mikroplątki wykonuje się metodą rysowania po powierzchni płytki ostrzem diamentowym i lupania. Tę operację wykonuje się przy zastosowaniu



Rys. 8.6
Cięcie płytki na mikroplątki (struktury)

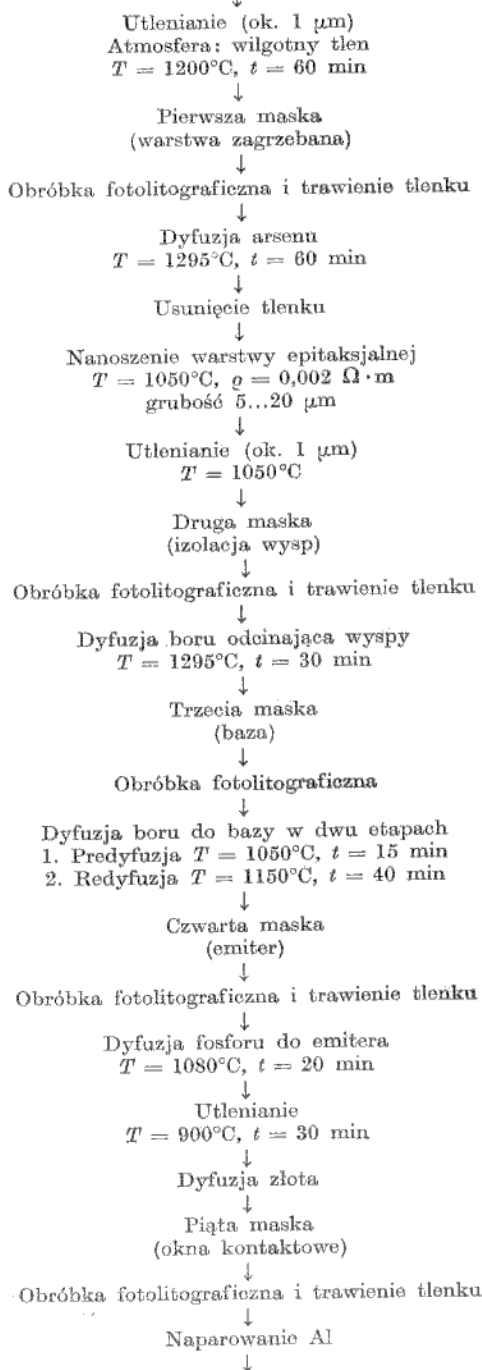
urządzeń automatycznych. Ostatnie operacje technologiczne polegają na umieszczeniu struktury układu scalonego w obudowie. Jest to tzw. mikromontaż, w którym można wyróżnić trzy operacje (rys. 8.7):

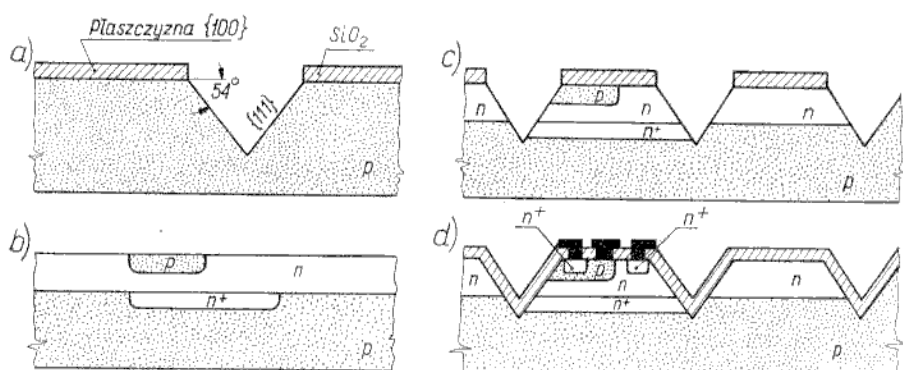


Rys. 8.7
Mikromontaż: a) połączenie mikroplątki z ażurem;
b) połączenie pól kontaktowych z wyprowadzeniami;
c) obudowa po zaprasowaniu ażuru w żywicy

- lutowanie albo klejenie struktury układu scalonego (mikroplątki) do obudowy lub ażuru będącego „szkieletem” obudowy plastikowej;
- termokompresja lub zgrzewanie ultradźwiękowe, za pomocą którego łączy się pola kontaktowe struktury układu scalonego z nóżkami (wyprowadzeniami zewnętrznymi) obudowy;

Płytką krzemową typu *p* o grubości ok. 200 μm i średnicy 50 lub 75 mm
 $\rho = 0,002 \dots 0,15 \Omega \cdot \text{m}$





Rys. 8.13
Podstawowe fazy procesu V-ATE

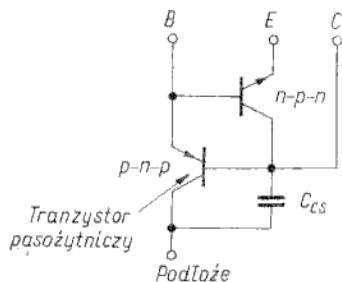
plaszczyna krystalograficzna {111} jest położona pod kątem 54° w stosunku do plaszczyny {100}, a zatem trawiąc powierzchnię o orientacji {100} otrzymuje się rowek o charakterystycznym kształcie litery V (rys. 8.13a). Głębokość tego rowka zależy tylko od rozmiarów okna w warstwie SiO_2 , gdyż z chwilą odsłonięcia plaszczyny {111} trawienie jest, praktycznie biorąc, zatrzymane. Kolejne fazy procesu V-ATE w uproszczeniu przedstawiono na rys. 8.13b, c, d. Gęstość upakowania dla tej technologii jest porównywalna z technologią ISOPLANAR II. Istnieje wiele wariantów zmodyfikowanych technologii V-ATE, jak np. procesy VIP (opracowany w firmie Motorola) lub POLYPLANAR (opracowany w firmie Harris), w których rowki w kształcie litery V są wypełnione krzemem polikrystalicznym w celu uzyskania płaskiej powierzchni płytki, co upraszcza strukturę połączeń wewnątrzukładowych.

Tranzystory

8.1.1.3

Podstawowym rodzajem tranzystora w USB (tranzystor i każdy inny element wykonany w układzie scalonym będzie nazywany monolitycznym) jest tranzystor *n-p-n* o strukturze omawianej w p. 8.1.1.1. Charakterystyczne różnice w budowie tranzystora monolitycznego i dyskretnego są następujące:

- istnienie warstwy zagrzebanej;
- ulokowanie w wyspie odizolowanej od innych elementów;
- wyprowadzenia wszystkich kontaktów znajdują się na górnej powierzchni płytki.



Rys. 8.14
Tranzystor bipolarny w układzie scalonym z uwzględnieniem elementów pasywnych (głównie tranzystora wertykalnego *p-n-p*)

Zadaniem warstwy zagrzebanej jest zmniejszenie rezystancji szeregowej kolektora.

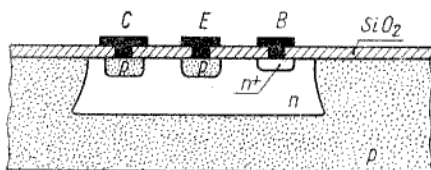
Analiza działania tranzystora monolitycznego jest podobna jak dla tranzystora dyskretnego (rozdz. 5) z tą tylko różnicą, że należy dodatkowo uwzględnić wpływ sprzężeń pasozytniczych tranzystora monolitycznego z podłożem. Przede wszystkim trzeba uwzględnić istnienie pasozytniczego tranzystora $p-n-p$ i pojemności C'_{cs} między kolektorem a podłożem (rys. 8.14). Tranzystor pasozytniczy, którego emiterem jest obszar bazy tranzystora monolitycznego, a kolektorem podłoże, nie występuje w układach z izolacją dielektryczną (pojemność C'_{cs} jest wtedy również bardzo mała). Jego działanie jest słabe w przypadku izolacji złączowo-dielektrycznej (bazą tego tranzystora jest wówczas silnie domieszkowana warstwa zagrzebana), a w istotny sposób może się przejawiać w układach z izolacją złączową. Tranzystor pasozytniczy w połączeniu z tranzystorem właściwym może spowodować wystąpienie efektu tyristorowego. Taki efekt, znany jako „mechanizm zatraskowy” (ang. *latching mechanism*), obserwuje się w niektórych wzmacniaczach operacyjnych (na przykład w popularnym wzmacniaczu $\mu A 709$).

Specyficznym rodzajem tranzystora monolitycznego jest tranzystor wieloemiterowy, stosowany w układach cyfrowych serii TTL (rozdz. 10).

W niektórych układach zachodzi konieczność stosowania połączeń komplementarnych tranzystorów $n-p-n$ i $p-n-p$. Wykonanie tranzystora $p-n-p$ o parametrach porównywalnych z tranzystorem $n-p-n$ jest możliwe, lecz wymaga zastosowania dodatkowych operacji technologicznych. Dlatego najczęściej wykorzystuje się konfiguracje strukturalne, powstające w procesie wytwarzania tranzystora $n-p-n$. W ten sposób mogą być wytworzone dwa rodzaje tranzystorów $p-n-p$:

- tranzystor wertykalny (pionowy),
- tranzystor lateralny (poziomy lub boczny).

Tranzystor wertykalny jest to inaczej tranzystor pasozytniczy z rys. 8.14, w którym kolektorem jest obszar podłoża (typ p), rolę bazy odgrywa warstwa epitaksjalna typu n , a obszar emitera jest wytwarzany w trakcie dyfuzji bazy tranzystora $n-p-n$. W strukturze tego tranzystora nie ma oczywiście warstwy zagrzebanej. Współczynnik wzmocnienia prądowego tranzystora wertykalnego jest bardzo mały ($\beta = 0,5 \dots 5$) z uwagi na dużą szerokość bazy. Możliwości zastosowań tego tranzystora ogranicza fakt, że kolektor (podłoże) jest dołączony do największego potencjału ujemnego w układzie. Może on zatem pracować tylko w układzie wtórnika emiterowego.



Rys. 8.15

Struktura tranzystora lateralnego $p-n-p$

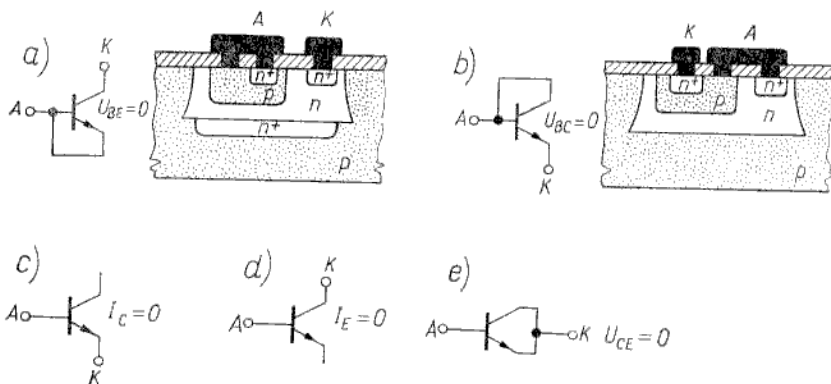
Strukturę tranzystora lateralnego przedstawiono na rys. 8.15. Powstaje on w tym samym ciągu procesów technologicznych co tranzystor $n-p-n$, przy czym obszary kolektora i emitera są tworzone równocześnie z dyfuzją bazy tranzystora $n-p-n$. Akeja tranzystorowa przebiega równoległe do powierzchni struktury i stąd bierze się nazwa lateralny. Wzmocnienie prądowe tranzystora lateralnego jest również bardzo małe ($\beta = 1 \dots 20$) z uwagi na dużą szerokość bazy (ok. $10 \mu m$).

Jeżeli są wymagane większe wzmocnienia, to tranzystor *p-n-p* (wertykalny lub lateralny) łączy się z tranzystorem *n-p-n*, w wyniku czego otrzymuje się układ *Darlingtona* równoważny jednemu tranzystorowi *p-n-p* o dużym wzmocnieniu.

Diody

8.1.1.4

Diody monolityczne realizuje się przez odpowiednie wykorzystanie złączy *p-n* istniejących w strukturze tranzystora. Pięć możliwych konfiguracji tranzystora pracującego jako dioda przedstawiono na rys. 8.16, przy czym dla dwu pier-



Rys. 8.16

Pięć konfiguracji tranzystora pracującego jako dioda

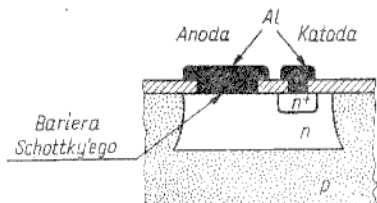
wszych, najważniejszych konfiguracji, pokazano również przekroje struktur. Ogólnie można stwierdzić, że wykorzystuje się złącze emiter-baza lub (i) złącze baza-kolektor. Decydując się na jedną z kilku możliwych konfiguracji bierze się pod uwagę przede wszystkim wartość napięcia przebicia i czas przełączania.

Dioda utworzona ze złącza kolektor-baza przy emiterze zwartym z bazą (rys. 8.16a) ma duże napięcie przebicia (20...60 V) oraz średni czas przełączania (50...70 ns). Analogiczna dioda z emiterem rozwartym (rys. 8.16d) ma identyczne napięcie przebicia oraz większy czas przełączania (ok. 100 ns). Dioda utworzona ze złącza emiter-baza przy kolektorze zwartym z bazą (rys. 8.16b) ma małe napięcie przebicia (6...7 V) oraz najmniejszy czas przełączania (kilka do kilkunastu nanosekund). Analogiczna dioda z kolektorem rozwartym (rys. 8.16c) ma identyczne napięcie przebicia oraz większy czas przełączania (kilkadziesiąt nanosekund). W przypadku polaryzacji złącza *E-B* w kierunku przewodzenia występuje pasożytnicza akcja tranzystorowa w kierunku podłoża (podłoże spolaryzowane napięciem ujemnym spełnia funkcję kolektora). Dlatego w konfiguracjach a), c) pozostawia się warstwę zagrzebaną, która osłabia wzmocnienie prądowe pasożytniczego tranzystora.

Ostatnia konfiguracja (rys. 8.16e), otrzymana przez połączenie równolegle emitera z kolektorem, ma wady obu konfiguracji a), b), tj. małe napięcie przebicia i duży czas przełączania. Z porównania pięciu konfiguracji wynika, że najbardziej celowe jest stosowanie układów a), b). Złącze *E-B* spolaryzowane w kierunku zaporowym może również pracować jako dioda Zenera.

W wielu układach stosuje się diody z barierą Schottky'ego, charakteryzujące się znikomym małym czasem przełączania (mniej niż 1 ns) i małym spadkiem napięcia przy

pracy w kierunku przewodzenia. Bariere Schottky'ego uzyskuje się na styku aluminium z niezbyt silnie domieszkowanym krzemem typu n (na styku aluminium z silnie domieszkowanym krzemem n^+ powstaje kontakt omowy). Strukturę diody Schottky'ego przedstawiono na rys. 8.17.

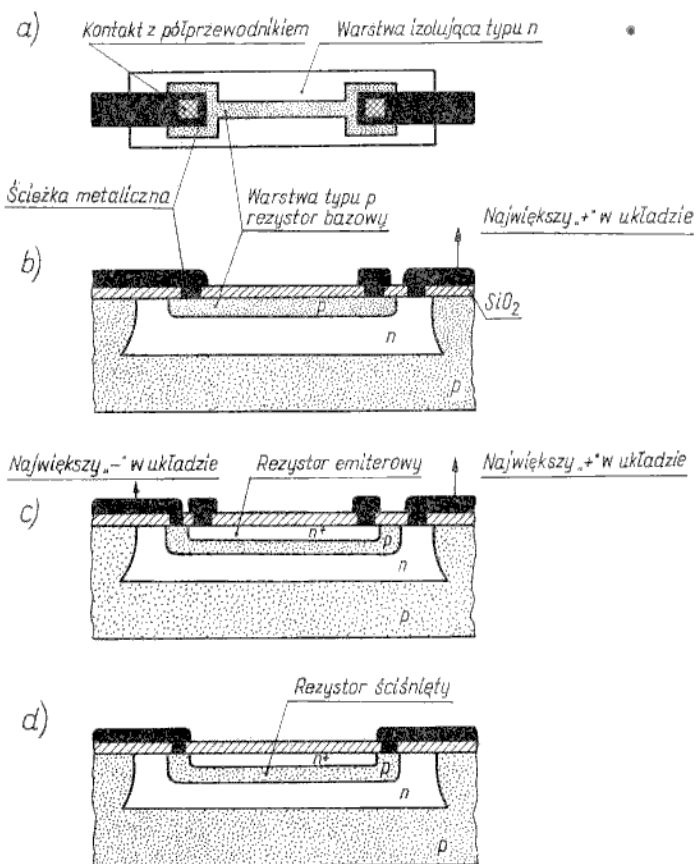


Rys. 8.17
Struktura diody Schottky'ego

Rezystory

8.1.1.5

W USB funkcję rezystorów spełniają warstwy dyfuzyjne, wykonywane równocześnie z poszczególnymi obszarami tranzystora. Teoretycznie istnieje możliwość wykorzystania jednej z czterech warstw: podłoża, warstwy epitaksjalnej (ob-

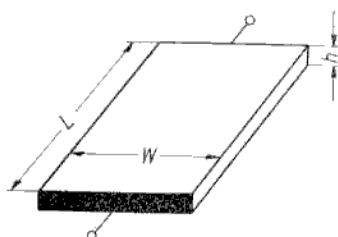


Rys. 8.18
Trzy warianty rezystorów dyfuzyjnych

szaru kolektora), obszaru bazy oraz obszaru emitera. W praktyce najczęściej wykorzystuje się obszar bazy, a w niektórych przypadkach obszar emitera lub warstwę bazy pocienioną przez dyfuzję emitera. Te trzy warianty struktur rezystorów dyfuzyjnych przedstawiono na rys. 8.18.

Przed omówieniem poszczególnych struktur rezystorów należy zdefiniować bardzo dogodny parametr, jakim jest *rezystancja powierzchniowa*. Jeżeli jest warstwa o rezystywności ρ i wymiarach L , W , h (rys. 8.19), to rezystancja tej warstwy

$$R = \rho \frac{L}{hW}$$



Rys. 8.19
Ilustracja do pojęcia rezystancji powierzchniowej

Ponieważ grubości warstw są stałe dla określonej technologii (w układzie scalonym jest optymalizowana konstrukcja tranzystora, dlatego wykonując rezystory dyfuzyjne nie dysponuje się możliwością niezależnej regulacji wymiaru h), można wprowadzić pojęcie rezystancji powierzchniowej

$$R_{\square} = \frac{\rho}{h} \quad (8.1)$$

Wówczas

$$R = R_{\square} \frac{L}{W} \quad (8.2)$$

Rezystancja powierzchniowa jest to inaczej rezystancja warstwy o kształcie kwadratowym, przy czym nie zależy ona od długości boku kwadratu i jest mierzona w omach (często podaje się wymiar Ω/\square — om na kwadrat). Znając R_{\square} łatwo można obliczyć rezystancję warstwy dyfuzyjnej, gdyż wystarczy obliczyć liczbę kwadratów w pasku prostokątnym, jaki tworzy ta warstwa.

Strukturę tzw. *rezystora bazowego* przedstawiono na rys. 8.18a, b (widok z góry i przekrój). Jest to rezystor utworzony przez warstwę typu p wdyfundowaną do wyspy typu n równocześnie z dyfuzją bazy w tranzystorze. Rezystancja powierzchniowa dla tej warstwy wynosi ok. $200 \Omega/\square$. W celu odizolowania warstwy typu p od podłoża i wyeliminowania pasożytniczej akcji tranzystorowej w kierunku pionowym (warstwa dyfuzyjna typu p z wyspą typu n i z podłożem typu p tworzy strukturę $p-n-p$) warstwę typu n (wyspę) polaryzuje się największym potencjałem dodatnim, jaki jest dostępny w układzie. W ten sposób oba złącza $p-n$ są spolaryzowane w kierunku zaporowym przy dowolnych potencjałach na końcówkach rezystora. Rozmiary warstwy dyfuzyjnej są ograniczone zarówno od dołu (z uwagi na tolerancję procesu fotolitografii), jak i od góry (z uwagi na nieakceptowalny wzrost powierzchni zajmowanej przez rezystor). Powoduje to oczywiście ograniczenia wartości nominalnych rezystorów bazowych (praktycznie $100 \Omega \leq R \leq$

$\leq 20 \text{ k}\Omega$). Rezystory o wartościach nominalnych mniejszych niż 100Ω wykonuje się z warstwy emiterowej, są to tzw. *rezystory emiterowe* o strukturze pokazanej na rys. 8.18c. Dla rezystora emiterowego $R_{\square} = 2 \dots 5 \Omega/\square$. W celu dobrego odizolowania warstwy n^+ od podłoża polaryzuje się sąsiadujące z tą warstwą złącze p - n w kierunku zaporowym.

Rezystory o wartościach nominalnych większych niż $10 \text{ k}\Omega$ wykonuje się w postaci warstwy typu p pocienionej przez warstwę n^+ (rys. 8.18d). Są to tzw. *rezystory z dyfuzją odcinającą*, nazywane niekiedy *rezystorami ściśniętymi* (ang. *pinch resistor*). Rezystancja powierzchniowa wynosi w tym przypadku kilka kiloomów na kwadrat i można otrzymać rezystory o wartościach nominalnych rzędu kilkuset kiloomów. Dużą wadą tego rezystora jest nieliniowy przebieg prądu w funkcji napięcia wskutek efektu zaciskania warstwy typu p przez warstwy zaporowe złącza p - n . Jest to zjawisko identyczne jak w tranzystorach PNFET, przy czym warstwa typu p spełnia funkcję kanału.

Wszystkie trzy rodzaje rezystorów mają dodatni współczynnik rezystancji o wartości $0,05 \dots 0,5\%/^{\circ}\text{C}$. Tolerancje rezystorów dyfuzyjnych są bardzo duże, gdyż dla warstw o typowej szerokości $20 \mu\text{m}$ wynoszą ok. $\pm 20\%$ dla rezystorów emiterowych i bazowych oraz $\pm 50\%$ dla rezystorów ściśniętych.

Należy jednak zwrócić uwagę, że odchylenia od wartości nominalnych są współbieżne we wszystkich rezystorach wytwarzanych we wspólnej płytce. Tolerancje stosunku rezystancji poszczególnych rezystorów są małe i zwykle nie przekraczają 3% . Dlatego układy scalone są tak projektowane, że ich parametry są uzależnione silniej od stosunków rezystancji niż od ich wartości bezwzględnych.

Niektóre parametry rezystorów monolitycznych (głównie maksymalna częstotliwość pracy) są ograniczone przez pasożytnicze sprzężenie pojemnościowe z podłożem.

Niekiedy w USB oprócz rezystorów dyfuzyjnych są również wytwarzane rezystory cienkowarstwowe. Rezystory te są zbudowane z cienkich warstw oporowych, naniesionych na podłoże półprzewodnikowe pokryte warstwą SiO_2 . Najczęściej nanosi się ścieżki z nichromu ($80\% \text{ Ni}$, $20\% \text{ Cr}$), tantalu lub cermetu (mieszanka SiO z Cr).

Kondensatory

8.1.1.6

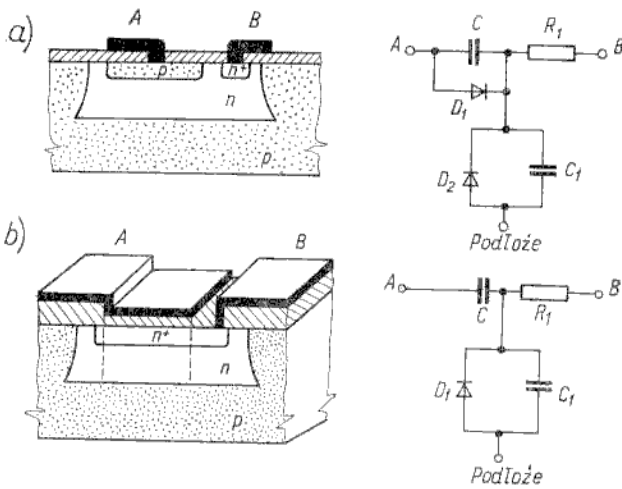
Kondensatory w USB wykonuje się zwykle jednym z dwu sposobów, tj. w postaci:

- złącza p - n ,
- struktury MOS.

Oba sposoby realizacji kondensatora ilustruje rys. 8.20.

Jeżeli stosuje się złącze p - n , to uzyskuje się kondensator nieliniowy, który może pracować tylko przy polaryzacji tego złącza w kierunku zaporowym. Można wykorzystać złącze kolektor-baza (rys. 8.20a) lub baza-emiter. W pierwszym przypadku pojemność jednostkowa jest mała (ok. $100 \text{ pF}/\text{mm}^2$), lecz napięcie przebicia jest duże (kilkadziesiąt woltów). W drugim przypadku pojemność jednostkowa jest duża (ok. $500 \text{ pF}/\text{mm}^2$), lecz napięcie przebicia jest małe ($6 \dots 7 \text{ V}$).

Drugi wariant realizacji kondensatora przedstawiono na rys. 8.20b. Jest to kondensator płaski, którego dolną okładkę stanowi silnie domieszkowana warstwa n^+ , dielektrykiem jest warstwa SiO_2 , górną okładką — warstwa aluminium. Pojemność tego kondensatora nie zależy od napięcia polaryzacji i wynosi ok. $350 \text{ pF}/\text{mm}^2$ przy najmniejszej grubości SiO_2 — ok. 100 nm . Napięcie przebicia wynosi co najmniej 50 V .



Rys. 8.20
Dwa sposoby realizacji kondensatora

Tolerancja pojemności w obu wariantach nie jest lepsza niż $\pm 20\%$. Z uwagi na ograniczenie powierzchni maksymalne pojemności kondensatorów monolitycznych wynoszą do 300 pF.

Niekiedy w USB stosuje się kondensatory cienkowarstwowe typu metal-izolator-metal. Wówczas na podłożu półprzewodnikowe pokryte warstwą SiO_2 nanosi się warstwę aluminium (dolna okładka), następnie warstwę SiO_2 , a na niej warstwę aluminium (górna okładka). Takie rozwiązanie jest rzadko stosowane, gdyż wymaga dodatkowych operacji technologicznych.

Indukcyjności

8.1.1.7

Indukcyjności w zasadzie nie są realizowane w USB. Dla zakresu dużych częstotliwości niekiedy wykonuje się cewki utworzone przez spiralną konfigurację ścieżki metalicznej, jednakże indukcyjności otrzymane w ten sposób wynoszą zaledwie kilka nanohenrów. Czynnione są również próby symulacji działania indukcyjności za pomocą odpowiednich układów czynnych — tzw. żyratorów.

Skrzyżowania ścieżek metalicznych

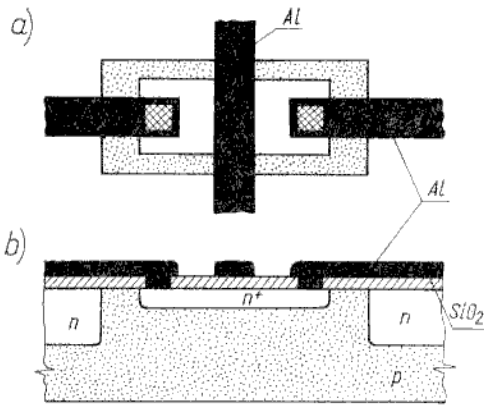
8.1.1.8

Jeżeli zachodzi konieczność krzyżowania ścieżek połączeń między elementami układu scalonego, to stosuje się tzw. *metalizację wielopoziomową* (kilka warstw metalizacji przedzielonych warstwami dielektryka) lub wykorzystuje się warstwę silnie domieszkowanego półprzewodnika dla *skrzyżowania tunelowego* (wiaduktwego). Pierwsze rozwiązanie jest bardzo drogie i znajduje zastosowanie tylko w niektórych układach o dużym stopniu scalenia. Drugie rozwiązanie ilustruje rys. 8.21.

Podsumowanie

8.1.1.9

Reasumując omówienie technologii USB należy zwrócić uwagę na kilka specyficznych cech tych układów:



Rys. 8.21
Skrzyżowanie ścieżek w układzie scalonym

1. Układy scalone bipolarne należy projektować tak, by eliminować elementy bierne i zastępować je elementami czynnymi (odwrotnie niż to jest w układach z elementami indywidualnymi). Ideałem byłby układ złożony z samych tranzystorów. Taka „strategia” projektowania wynika z następującego rozumowania. Koszty wytwarzania układów scalonych zależą od uzysku (stosunku liczby dobrych układów na płytce półprzewodnikowej do liczby wszystkich układów wytwarzanych w płytce). Z kolei uzysk jest odwrotnie proporcjonalny do powierzchni zajmowanej przez układ (im większa jest powierzchnia, tym większe jest prawdopodobieństwo trafienia na defekt). Powierzchnia tranzystora jest około 2-krotnie mniejsza niż powierzchnia rezystora o wartości 5 k Ω lub kondensatora o pojemności 10 pF. Dlatego korzystnie jest zastąpić rezystor bądź kondensator tranzystorem.
2. Parametry tranzystorów i diod monolitycznych są porównywalne z parametrami tranzystorów i diod indywidualnych.
3. Parametry tranzystorów i kondensatorów monolitycznych są znacznie gorsze niż parametry rezystorów i kondensatorów indywidualnych, tj. wartości nominalne są bardzo ograniczone ($10 \Omega \leq R \leq 100 \text{ k}\Omega$, $C \leq 300 \text{ pF}$), a tolerancje bardzo duże (ok. $\pm 20\%$ dla rezystorów i kondensatorów).
4. Tolerancje stosunków rezystancji lub pojemności są małe (ok. $\pm 1\%$ dla rezystorów i kondensatorów).
5. Projektując USB należy wyeliminować indukcyjności, gdyż — praktycznie biorąc — nie można ich zrealizować w postaci scalonej.

Układy scalone unipolarne

8.1.2

Układy scalone unipolarne, podobnie jak układy bipolarne, są wytwarzane wyłącznie w krzemie technologią epiplanarną. Należy od razu wyjaśnić, że z dwu klas tranzystorów unipolarnych (JFET oraz IGFET) w układach scalonych są stosowane tranzystory IGFET, a ściślej biorąc tylko tranzystory MIS¹⁾. Dlatego układy scalone unipolarne można nazywać układami scalonymi MIS lub MOS (biorąc pod uwagę, że izolatorem najczęściej jest tlenek krzemu).

¹⁾ Tranzystory PNFET są stosowane tylko sporadycznie w niektórych układach scalonych bipolarnych jako elementy dodatkowe, a układy scalone z tranzystorami MESFET są w stadium opracowań laboratoryjnych.

Układy scalone MIS w porównaniu z układami bipolarnymi mają szereg zalet. Należy tu wymienić przede wszystkim następujące właściwości:

- prostsza technologia wytwarzania (mniejsza liczba operacji technologicznych);
- możliwość wykonania całego układu wyłącznie z tranzystorów MIS, spełniających wszystkie wymagane elementarne funkcje układowe;
- większa gęstość upakowania, gdyż tranzystor MIS zajmuje powierzchnię mniejszą niż tranzystor bipolarny oraz istnieje naturalna samoizolacja tranzystorów, czyli zbędne są wyspy izolujące, które zajmują dużo miejsca w układach bipolarnych (wyspy izolujące stosuje się tylko w technologii CMOS).

Zalety technologii MIS szczególnie wyraźnie przejawiają się w układach o dużym stopniu scalenia.

Pod względem właściwości elektrycznych, biorąc pod uwagę dwa podstawowe parametry, tj.

- moc pobieraną,
 - szybkość działania,
- układy scalone MIS charakteryzują się znacznie mniejszym poborem mocy niż układy bipolarne (różnica kilku rzędów wartości), ustępują jednak układom bipolarnym pod względem szybkości działania, która jest kilkakrotnie mniejsza.

Właściwości elektryczne obu klas układów będą rozważane w rozdziałach 9, 10, obecnie rozpatrzmy zagadnienia związane z technologią wytwarzania układów scalonych MIS. Zakres rozważanych zagadnień jest inny niż w p. 8.1.1 dla technologii bipolarnej z uwagi na podstawowe różnice między tymi technologiami, tj. brak wysp izolujących i ograniczenie asortymentu elementów tylko do tranzystorów w układach MIS. Omówimy zatem kilka podstawowych procesów technologicznych ze szczególnym uwzględnieniem zagadnień specyficznych tylko dla technologii MIS.

W układach scalonych stosuje się wszystkie cztery rodzaje tranzystorów MIS, tj. z kanałami wzbogacającym i zubożającym, typu n i typu p .

Tranzystor MOS z kanałem wzbogacającym

8.1.2.1

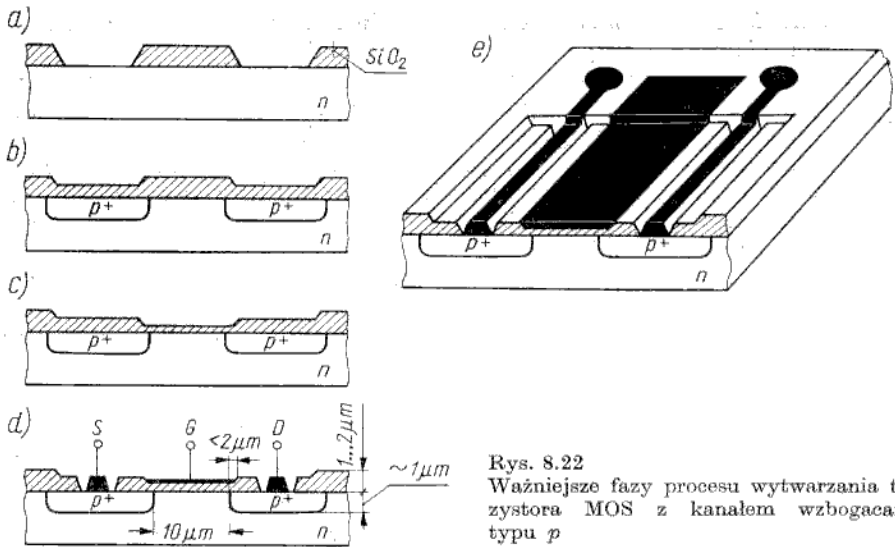
Z rozdziałów 3, 6 wiadomo, że w warstwie SiO_2 istnieje dodatni ładunek nieskompensowany, co powoduje akumulację nośników większościowych w obszarze przypowierzchniowym półprzewodnika typu n oraz inwersję w obszarze przypowierzchniowym półprzewodnika typu p . Trudniej jest zatem uzyskać tranzystor z kanałem wzbogacającym typu n (na podłożu typu p).

Tranzystor MOS z kanałem wzbogacającym typu p

Układy scalone MOS starszej generacji były wykonywane z tranzystorów z kanałem wzbogacającym typu p (tzw. *technologia PMOS*).

Na rysunku 8.22 przedstawiono ważniejsze fazy procesu wytwarzania tranzystora MOS z kanałem wzbogacającym typu p . Na czystej powierzchni płytki krzemowej typu n ($\rho = 0,01 \dots 0,1 \Omega \cdot \text{m}$) jest wytwarzana warstwa SiO_2 o dużej grubości (ok. $1 \mu\text{m}$). W konwencjonalnym procesie fotolitografii w warstwie SiO_2 , spełniającej funkcję maski, są otwierane okna, przez które zachodzi dyfuzja boru do obszarów źródła i drenu (rys. 8.22a).

W wyniku drugiego procesu utleniania obszary źródła i drenu zostają pokryte warstwą SiO_2 oraz jednocześnie zachodzi „rozieganie” (redyfuzja) domieszek w obszarach p^+ (rys. 8.22b). Po drugim procesie fotolitografii, otwierającym okna w obszarze bramki, w trzecim procesie utleniania jest wytwarzana cienka warstwa SiO_2 ($100 \dots 150 \text{ nm}$) — rys. 8.22c. Jest to warstwa tzw. *dielektryka bramkowego*,



Rys. 8.22
Ważniejsze fazy procesu wytwarzania tranzystora MOS z kanałem wzbogacającym typu p

która powinna się charakteryzować szczególnie dobrymi parametrami elektrofizycznymi. W trzecim procesie fotolitografii są otwierane okna nad obszarami p^+ , po czym następuje metalizacja całej powierzchni.

W czwartym procesie fotolitografii jest wytrawiana warstwa metaliczna w obszarach leżących poza źródłem, bramką i drenem; w rezultacie powstają kontakty metaliczne, pokazane na rys. 8.22d, e. Tak jak w układach bipolarnych stosuje się również dodatkowy proces pasywacji całej struktury (naniesienie szkliva w niskiej temperaturze) dla ochrony mechanicznej, głównie przed zadrapaniem. Ta pasywacja wymaga piątego procesu fotolitografii do otworzenia okien do pól kontaktowych.

Cechą charakterystyczną konstrukcji tranzystora z kanałem wzbogacającym jest to, że bramka całkowicie pokrywa kanał przykrywając częściowo obszary p^+ źródła i drenu. Jakakolwiek część kanału, nie przykryta bramką, wnosiliby bardzo dużą rezystancję włączoną szeregowo z kanałem, gdyż w tej części nie zachodziłaby inwersja. Dążąc do całkowitego pokrycia kanału wykonuje się bramkę o powierzchni stolerowanej dodatnio, powstają zatem obszary, w których bramka pokrywa częściowo źródło i dren, tworząc pojemności pasożytnicze C_{gse} , C_{gde} . W standardowej technologii są to pojemności 2...3 pF.

Tranzystor MOS z kanałem wzbogacającym typu n

Istotną zaletą tranzystorów z kanałem typu n w porównaniu z tranzystorami z kanałem typu p jest większa szybkość działania z uwagi na około 3-krotnie większą ruchliwość elektronów niż dziur. Układy scalone MOS młodszej generacji (produkowane od 1974 r.) są przeważnie zbudowane z tranzystorów z kanałem typu n , zarówno wzbogacającym jak i zubożającym (tzw. *technologia NMOS*). Zasadnicze fazy procesu wytwarzania tranzystora z kanałem wzbogacającym typu n są takie same jak dla tranzystora z kanałem typu p przy oczywistej zmianie typu przewodnictwa podłoża z n na p i domieszkowania obszarów dyfuzyjnych źródła i drenu fosforem zamiast borem. Jednakże tranzystor z bramką aluminiową

wą wykonany w ten sposób, przy parametrach fizycznych porównywalnych z typowymi parametrami w technologii PMOS, ma ujemne napięcie progowe, czyli jest tranzystorem z kanałem zubożanym. Żeby uzyskać dodatnią wartość napięcia progowego, czyli tranzystor z kanałem wzbogacanym typu n , trzeba posłużyć się jednym z następujących sposobów:

- zwiększenie poziomu domieszkowania podłoża (na przykład metodą implantacji);
- zastosowanie dodatkowej polaryzacji podłoża;
- zastosowanie bramki krzemowej zamiast aluminiowej, co powoduje zmianę wartości kontaktowej różnicy potencjałów o mniej więcej 1,5 V.

Największe znaczenie praktyczne mają sposoby pierwszy i ostatni, które omówimy w oddzielnym punkcie.

Tranzystor MOS z kanałem zubożanym

8.1.2.2

W przypadku bramki aluminiowej w sposób „naturalny” (wskutek istnienia dodatniego ładunku w warstwie SiO_2 i odpowiedniej wartości kontaktowej różnicy potencjałów φ_{ms}) uzyskuje się tranzystor z kanałem zubożanym typu n . W przypadku bramki krzemowej w sposób „naturalny” uzyskuje się niewielkie dodatnie napięcie progowe dla podłoża typu p (tranzystor z kanałem wzbogacanym typu n) lub napięcie progowe o warstwie bliskiej zeru dla podłoża typu n (sytuacja na pograniczu między kanałem wzbogacanym lub zubożanym typu p).

Żeby otrzymać tranzystor zubożany dla dowolnego typu przewodnictwa kanału, trzeba kanał wbudować postępując się jedną z następujących metod:

- dyfuzja,
- implantacja,
- epitaksja.

Ostatnio najczęściej stosuje się implantację lokalną w wybranych obszarach układu scalonego.

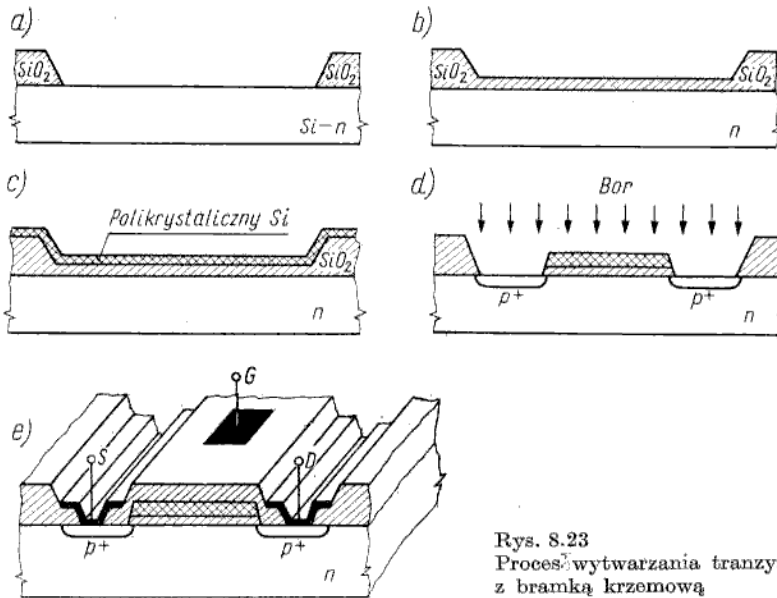
Technologia bramki krzemowej

8.1.2.3

Najważniejszymi czynnikami ograniczającymi szybkość działania tranzystorów MIS są pojemności C_{gs} , C_{gd} , spowodowane częściowym pokryciem obszarów źródła i drenu przez elektrodę bramki. Opracowano wiele sposobów zmniejszania zachodzenia bramki nad obszary źródła i drenu, opartych na różnych koncepcjach samocentrowania obrysów krawędzi bramki i kanału. Można tu wymienić szereg różnych metod technologicznych, jak np.:

- zastosowanie bramki krzemowej;
- zastosowanie bramki molibdenowej lub wolframowej (technologia RMOS — ang. *Refractory Metal Oxide Semiconductor*);
- implantacja jonów w obszarach źródła i drenu;
- technologia LOCOS, planox, COPLAMOS, LOCMOS itp.

Największe rozpowszechnienie uzyskała technologia bramki krzemowej. Idea samocentrowania jest realizowana w tej metodzie przez stosowanie elektrody bramki jako maski podczas domieszkowania obszarów źródła i drenu. Sposób wytwarzania tranzystora MOS z bramką krzemową przedstawiono na rys. 8.23. Najpierw na całej powierzchni płytki krzemowej jest wytwarzana gruba warstwa tlenku (ok. 1 μm) — rys. 8.23a, w której są wytrawiane okna. Celem drugiego procesu utleniania (rys. 8.23b) jest wytworzenie cienkiej warstwy SiO_2 (100...150 nm),



Rys. 8.23
Proces wytwarzania tranzystora MOS
z bramką krzemową

która pozostanie pod bramką tranzystora. Po naniesieniu warstwy krzemu polikrystalicznego na całej powierzchni płytki (rys. 8.23c) warstwa ta zostaje stracona z wyjątkiem obszaru bramki. Następnie wykonuje się dyfuzję (implantację) domieszek do obszarów źródła i drenu (rys. 8.23d). Domieszki nie przechodzą przez warstwę krzemu polikrystalicznego i przez grubą warstwę SiO_2 . Fazą końcową cyklu technologicznego jest pokrycie całej powierzchni płytki warstwą SiO_2 oraz wytrawienie okien dla kontaktów metalicznych źródła, bramki i drenu (rys. 8.23e). Kontakty są wykonywane metodą konwencjonalną.

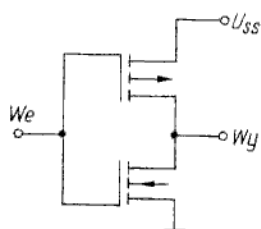
Najważniejsze zalety technologii bramki krzemowej są następujące:

- zwiększenie szybkości działania układów scalonych, spowodowane przede wszystkim zmniejszeniem pojemności pasozytniczych w tranzystorach i możliwością wykonania tranzystorów z krótszymi kanałami ($L_C \approx 5 \mu\text{m}$);
- korzystne wartości napięć progowych (w tranzystorach z kanałem typu p — małe napięcie ujemne, w tranzystorach z kanałem typu n — małe napięcie dodatnie), dzięki czemu jest możliwa bezpośrednia współpraca tranzystorów MOS z tranzystorami bipolarnymi;
- zwiększenie gęstości upakowania wskutek zmniejszenia rozmiarów tranzystora i możliwości stosowania warstwy krzemu polikrystalicznego jako dodatkowy poziom połączeń wewnątrz układu.

Technologia CMOS

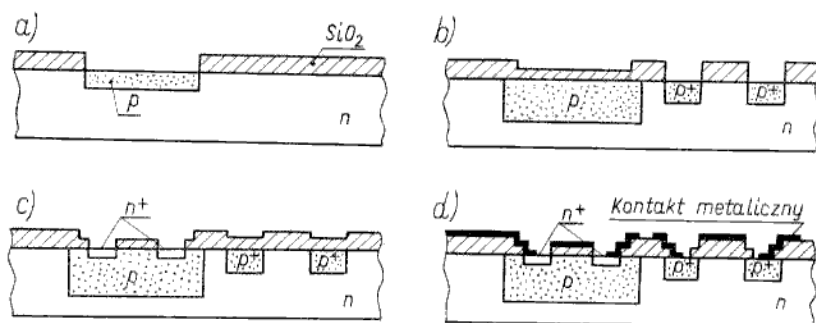
8.1.2.4

Stosując cykl technologiczny podobny jak w układach bipolarnych (z wyspami izolacyjnymi) można wykonać we wspólnym podłożu półprzewodnikowym tranzystory MOS z przeciwstawnymi rodzajami przewodnictwa kanałów. Dwa tranzystory MOS z kanałem typu p i typu n (rys. 8.24) nazywa się parą komplementarną (CMOS lub COS/MOS — *Complementary Symmetry MOS*). Pary komplementarne mają szczególnie korzystne właściwości w zastosowaniach do



Rys. 8.24
Para komplementarna tranzystorów MOS (CMOS)

układów scalonych cyfrowych, a układy CMOS wraz z układami PMOS i NMOS stanowią trzy podstawowe klasy układów scalonych MOS. Sposób wytwarzania komplementarnej pary tranzystorów MOS przedstawiono schematycznie na rys. 8.25. Do płytki krzemowej typu n przez okno w maskującej warstwie tlenkowej



Rys. 8.25
Proces wytwarzania pary komplementarnej tranzystorów MOS

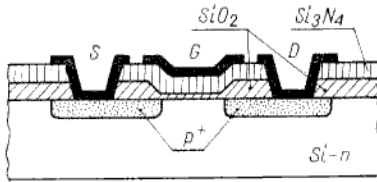
zostają wdyfundowane atomy domieszki akceptorowej, tworząc obszar typu p (tzw. kieszeń) — rys. 8.25a, który będzie podłożem dla tranzystora z kanałem typu n . Po następnym utlenieniu są otwierane okna dla dyfuzji obszarów p^+ , które będą źródłem i drenem tranzystora z kanałem typu p (rys. 8.25b). Po kolejnym utlenieniu są otwierane okna dla dyfuzji obszarów n^+ (rys. 8.25c), które będą źródłem i drenem tranzystora z kanałem typu n . Po otwarciu okien nad obszarami p^+ , n^+ wykonuje się w sposób konwencjonalny kontakty metaliczne (rys. 8.25d).

Tranzystory MIS jako elementy pamięci

8.1.2.5

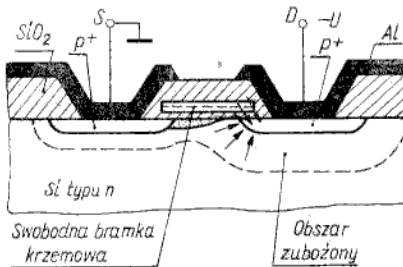
Jeden z najważniejszych rodzajów układów scalonych MIS stanowią tzw. pamięci stałe, w których są wykorzystane zjawiska trwałego magazynowania ładunku w niektórych strukturach MIS. Funkcję elementów pamięci stałej spełniają tranzystory MNOS, MAOS lub FAMOS.

Budowę tranzystora MNOS (ang. *Metal-Nitride-Oxide-Semiconductor*) przedstawiono na rys. 8.26. Pod aluminiową elektrodą bramki znajdują się dwie warstwy dielektryczne, tj. Si_3N_4 o grubości kilkadziesiąt nanometrów oraz SiO_2 o grubości rzędu kilku nanometrów. Przy odpowiednio dużym napięciu U_{GS} ładunek zmagazynowany w stanach pułapkowych na granicy warstw Si_3N_4 — SiO_2 zmienia się wskutek tunelowania nośników do lub z półprzewodnika przez cienką warstwę SiO_2 , zmieniając w ten sposób wartość napięcia progowego. Stan zmagazynowa-



Rys. 8.26
Tranzystor MNOS

nia nośników wychwyconych przez pułapki utrzymuje się, gdy napięcie U_{GS} jest odpowiednio małe, gdyż warstwa SiO_2 działa jak bariera dla prądu tunelowego. Ładunek (informacja w pamięci) może być przechowywany przez wiele lat. Analogicznie działa tranzystor MAOS (ang. *Metal-Alumina-Oxide-Semiconductor*). Elementem pamięci stałej może być również tranzystor z bramką swobodną — FA MOS (ang. *Floating Gate Avalanche-Injection MOS*). Schematyczny przekrój takiej struktury pokazano na rys. 8.27. Bramka swobodna



Rys. 8.27
Struktura FAMOS

z krzemu polikrystalicznego (bramka bez wyprowadzonego kontaktu), ułożona w odległości ok. 100 nm od powierzchni płytki krzemowej, ładuje się wskutek wstrzykiwania nośników podczas przebiega lawinowego dren-podłoże przy napięciu na drenie powyżej -30 V. Stan naładowania bramki jest bardzo trwały, gdyż w temperaturze $125^\circ C$ ładunek maleje do poziomu 70% wartości początkowej po czasie 10 lat. Ładunek można łatwo usunąć przez naświetlenie struktury promieniowaniem X lub nadfioletowym.

Układy warstwowe

8.2

Układy cienkowarstwowe

8.2.1

Układy scalone cienkowarstwowe są wytwarzane metodami nanoszenia w próżni cienkich warstw przewodzących, rezystancyjnych i dielektrycznych na bierne podłoże izolacyjne. W ten sposób otrzymuje się wysokiej jakości rezystory i kondensatory, tranzystory i diody natomiast są wytwarzane w oddzielnym procesie (są to elementy epiplanarne) i montowane indywidualnie w strukturę układu cienkowarstwowego. Układy cienkowarstwowe mają zatem budowę hybrydową. W skrócie omówimy podstawowe procesy technologiczne oraz właściwości podłoża, warstw przewodzących, rezystorów i kondensatorów.

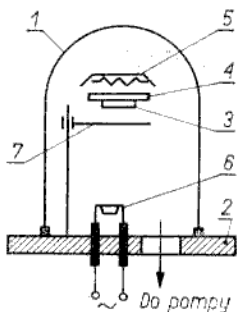
Procesy technologiczne

8.2.1.1

Podstawowym procesem w technologii cienkowarstwowej jest nanoszenie warstw w próżni. Stosuje się głównie dwie metody:

- naparowywanie cieplne,
- rozpylanie katodowe.

Naparowywanie cieplne prowadzi się w urządzeniu nazywanym *napyłarką*, którego podstawową część stanowi komora próżniowa, pokazana na rys. 8.28. Komorę



Rys. 8.28
Komora próżniowa napyłarki

próżniową stanowi klosz metalowy 1 spoczywający na podstawie 2 z podkładką gumową. Wewnątrz komory znajduje się podłoże 3 zamocowane w uchwycie 4, grzejnik podłoża 5 oraz wyparownik z grzejnikiem 6. Między wyparownikiem a podłożem znajduje się przesłona 7. Po odpompowaniu powietrza spod klosza (zwykle jest wymagane ciśnienie ok. 10^{-5} mm Hg) jest przeprowadzany proces naparowywania, który w sensie fizycznym składa się z trzech faz:

- parowanie materiału umieszczonego w wyparowniku (pod wpływem odpowiednio wysokiej temperatury);
- rozprzestrzenianie się par (w próżni cząsteczki rozprzestrzeniają się prostoliniowo, gdyż ich drogą swobodną jest większa niż odległość wyparownik-podłoże);
- kondensacja par na podłożu i powstawanie cienkiej warstwy (dla zapewnienia dobrej przyczepności warstwy do podłoża często podłoże jest podgrzewane do temperatury 200...300°C).

Metodą naparowywania cieplnego można otrzymywać warstwy wielu materiałów stosowanych w technice cienkowarstwowej, takich jak aluminium, nichrom, złoto, tlenek krzemu.

W przypadku materiałów trudno topliwych (o wysokiej temperaturze topnienia), jak np. tantal, stosuje się napylanie w plazmie jonowej (rozpylanie katodowe lub rozpylanie reaktywne, ang. *sputtering*). W tym przypadku „odparowanie” materiału nie jest spowodowane energią cieplną (wysoką temperaturą), lecz następuje wybijanie cząsteczek materiału wskutek bombardowania jonami gazu wypełniającego komorę próżniową.

Po naniesieniu cienkiej warstwy na całą płytkę podłożową jest przeprowadzany proces fotolitografii, w którym jest kształtowana struktura topologiczna warstwy.

Podłoża

8.2.1.2

Materiały, z których są wykonywane podłoża, powinny spełniać przede wszystkim następujące wymagania:

- dobre właściwości izolacyjne,
- dobra gładkość powierzchni,
- dobre przewodnictwo ciepła,
- zgodność temperaturowego współczynnika rozszerzalności liniowej z materiałami nanoszonych warstw,

— duża wytrzymałość mechaniczna.

Najczęściej stosuje się podłoża ze szkła boro-krzemowego. Typowe rozmiary płytek podłożowych są następujące: $20 \times 30 \times 1$ mm.

Warstwy przewodzące

8.2.1.3

Warstwy przewodzące spełniają następujące funkcje w układzie cienkowarstwowym:

- połączenia między elementami;
- pola kontaktowe, służące do przyłączenia podzespołów dyskretnych i wyprowadzeń zewnętrznych;
- okładki kondensatorów;
- uzwojenia cewek indukcyjnych.

Do najważniejszych wymagań, jakie powinny spełniać te warstwy, należy zaliczyć:

- dużą przewodność,
- dobrą przyczepność do podłoża,
- podatność na lutowanie, termokompresję i kształtowanie topologii metodą fotolitografii.

Najczęściej stosowanymi materiałami są złoto i aluminium.

Rezystory

8.2.1.4

Rezystory cienkowarstwowe są wytwarzane przez nanoszenie na podłoże cienkich warstw metalicznych lub mieszanin metali i dielektryka. Najczęściej stosowanym materiałem rezystywnym jest nichrom (zwykle o składzie 80% Ni, 20% Cr), który przy grubości warstwy ok. 10 nm ma rezystancję powierzchniową ok. $200 \Omega/\square$. Niekiedy stosuje się również warstwy z tantalu ($R_{\square} = 50 \dots 600 \Omega/\square$) lub cermetu (mieszanina Cr-SiO). Podstawową zaletą tego ostatniego materiału jest możliwość uzyskiwania warstw o bardzo dużej rezystancji powierzchniowej ($R_{\square} = 200 \dots 20\,000 \Omega/\square$).

Bardzo ważną właściwością technologii warstwowej jest możliwość korekcji rezystancji po naniesieniu warstwy rezystancyjnej na podłoże (takiej możliwości nie ma w układach scalonych półprzewodnikowych). Korekcję najczęściej wykonuje się przez precyzyjne nacinanie warstwy za pomocą strumienia światła laserowego. W ten sposób można uzyskać tolerancje rezystorów $\pm 0,05\%$.

Kondensatory

8.2.1.5

Kondensatory cienkowarstwowe otrzymuje się przez naniesienie na podłoże izolacyjne kolejno warstw: metalicznej, dielektrycznej i metalicznej. Warstwy metaliczne (okładki kondensatora) są najczęściej złote lub aluminiowe. Jako dielektryk najczęściej stosuje się SiO_x . Ze względu na ograniczenia uzyskiwanych wartości nominalnych (10 pF...20 nF) i duże tolerancje ($\pm 10 \dots \pm 20\%$) często zamiast kondensatorów cienkowarstwowych stosuje się kondensatory dyskretnie (tzw. kondensatory monolityczne) dolutowywane do struktury cienkowarstwowej.

Układy grubowarstwowe

8.2.2

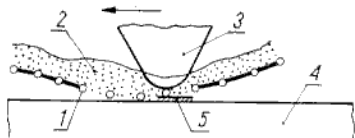
Układy grubowarstwowe są wytwarzane przez nanoszenie metodą sitodruku warstw przewodzących, rezystancyjnych i dielektrycznych na bierne podłoże izolacyjne. Podobnie jak w technologii cienkowarstwowej otrzymuje się w ten

sposób dobrej jakości rezystory i kondensatory, tranzystory i diody są natomiast wytwarzane w oddzielnym procesie (technologią epiplanarną) i montowane indywidualnie w strukturze układu grubowarstwowego.

Procesy technologiczne

8.2.2.1

W technologii grubowarstwowej podstawowym procesem technologicznym jest nanoszenie warstw grubych na podłoże ceramiczne metodą sitodruku (rys. 8.29). W tym celu używa się sita (stalowe lub nylonowe) z maską emulsyjną, naniesioną



Rys. 8.29
Nanoszenie grubych warstw metodą sitodruku
1 — sito; 2 — pasta; 3 — rakla; 4 — podłoże;
5 — nanoszona warstwa

metodą fotolitografii. Pasta o odpowiednim składzie jest przeciskana za pomocą rakli przez oczka sita i osadza się na podłożu tworząc warstwę o kształcie topologicznym, odwzorowującym negatyw maski emulsyjnej. Proces sitodruku przeprowadza się w zautomatyzowanych drukarkach. Następnym procesem jest wypalanie warstwy w temperaturze 900...1000°C w celu zapewnienia odpowiedniej przyczepności do podłoża.

Podłoża

8.2.2.2

W układach grubowarstwowch nie można stosować podłoży szklanych (jak w układach cienkowarstwowch) z uwagi na deformujące działanie wysokiej temperatury podczas wypalania warstw. Dlatego stosuje się podłoża ceramiczne, najczęściej z ceramiki alundowej (o dużej zawartości Al_2O_3). Typowe rozmiary powierzchni płytek podłożowych są następujące: 50 × 50 mm, 50 × 25 mm lub 25 × 25 mm, natomiast grubość płytek wynosi 0,4...1 mm.

Warstwy przewodzące

8.2.2.3

Warstwy przewodzące spełniają podobne funkcje jak w technologii cienkowarstwowej (połączenia między elementami, pola kontaktowe, okładki kondensatorów). Podstawowymi materiałami są kompozycje: Au-Pt, Pd-Au, Pd-Ag.

Rezystory

8.2.2.4

Warstwy rezystancyjne są to pasty zawierające zawiesinę cząsteczek przewodzących (proszku metalicznego) w szkliwie. Podstawowymi składnikami tych past są najczęściej: Pd-PdO, Pd-Au, Pd-Ag. Tak jak w układach cienkowarstwowch istnieje możliwość korekcji wartości rezystancji. W tym celu stosuje się operację „piaskowania”, tj. ścieranie grubej warstwy pod wpływem strumienia drobnego piasku. Jest to proces sterowany automatycznie.

Kondensatory

8.2.2.5

Kondensatory grubowarstwowe otrzymuje się przez naniesienie na podłoże izolacyjne kolejno warstw: przewodzącej, dielektrycznej i przewodzącej. Jako dielektryk najczęściej stosuje się pastę, której podstawowymi składnikami są: $BaTiO_3$ (tytanian baru) lub TiO_2 (rutyl). Często stosuje się kondensatory dyskretne dolutowywane do struktury grubowarstwowej.

Porównanie technologii półprzewodnikowej, cienkowarstwowej i grubowarstwowej

8.3

Podstawowe właściwości układów scalonych, wytwarzanych różnymi technologiami, porównano w tabl. 8.1.

Porównanie podstawowych właściwości układów scalonych półprzewodnikowych, cienkowarstwowych i grubowarstwowych

Tablica 8.1

Właściwość	Układy scalone		
	półprzewodnikowe	cienkowarstwowe	grubowarstwowe
Nakłady na uruchomienie produkcji	duże	średnie	małe
Ekonomicznie opłacalna długość serii	duża (> 100 tys.)	średnia (> 5 tys.)	mała (> 1 tys.)
Cena układów	bardzo mała	duża	średnia
Niezawodność	duża	duża	średnia
Gęstość upakowania	bardzo duża	średnia	mała
Zakresy nominalnych wartości rezystorów i kondensatorów	małe	średnie	duże
Tolerancje elementów	dostateczne	bardzo dobre	dobre
Stalność temperaturowa	dostateczna	bardzo dobra	dobra
Zakres częstotliwości pracy	duży	bardzo duży	duży
Maksymalna moc rozpraszana	średnia	średnia	duża

Z przedstawionego porównania wynikają określone wnioski odnośnie do zakresów za stosowań poszczególnych technologii.

Technologia półprzewodnikowa jest bezkonkurencyjna pod względem gęstości upakowania i stopnia scalenia. Układy scalone półprzewodnikowe są najtańsze pod warunkiem produkcji wielkoseryjnej (rzędu setek tysięcy sztuk). Aby zapewnić zapotrzebowanie na długie serie produkcyjne, konieczna jest daleko posunięta standaryzacja i uniwersalizacja produkowanych układów, co powoduje ograniczenie asortymentu układów scalonych półprzewodnikowych.

Technologie warstwowe nadają się do produkcji niestandardowych układów o niewielkim zapotrzebowaniu (krótkie serie). Ponadto domeną technologii warstwowych są te układy, które wymagają wysokiej jakości elementów biernych, jak np. wzmacniacze o wąskich tolerancjach charakterystyk, wzmacniacze szerokopasmowe itp.

Układy grubowarstwowe znajdują zastosowanie w sprzęcie powszechnego użytku, a układy cienkowarstwowe (droższe i mające lepsze parametry) są stosowane w sprzęcie profesjonalnym. Technologia cienkowarstwowa jest szczególnie predestynowana do wytwarzania układów mikrofalowych.

Należy dodać, że są również produkowane układy hybrydowe o superdużym stopniu scalenia, zbudowane w ten sposób, że do struktury układu warstwowego są dołączone układy scalone półprzewodnikowe (w specjalnym wykonaniu bez obudów i z wyprowadzeniami belkowymi lub sferycznymi) i cała struktura hybrydowa jest zamykana we wspólnej obudowie. Są to bardzo drogie układy przeznaczone głównie do zastosowań w technice wojskowej i kosmicznej.

W poprzednim rozdziale stwierdziliśmy, że użytkownik układów scalonych dzieli je pod względem funkcjonalnym na analogowe i cyfrowe.

W układach analogowych obróbce podlegają sygnały ciągłe, w układach cyfrowych — sygnały dyskretne, charakteryzujące się zwykle dwoma poziomami napięcia, utożsamianymi z „zerem” i „jedyneką” logiczną. Układy analogowe są tradycyjnie nazywane liniowymi, chociaż często służą do nieliniowej obróbki sygnałów ciągłych, jak na przykład do kształtowania przebiegów logarytmicznych, stabilizacji napięcia lub prądu itp. Będziemy stosować nazwę układy analogowe, zdając sobie jednak sprawę z tego, że klasa układów rozpatrywanych w tym rozdziale, w wielu książkach, a zwłaszcza zaś w katalogach, jest przedstawiana pod nazwą układy liniowe.

Układy analogowe scalone są wytwarzane różnymi technologiami (półprzewodnikową, cienkowarstwową lub grubowarstwową). Podstawowe znaczenie ma technologia półprzewodnikowa, która umożliwia uzyskanie długich serii tanich układów. Z tego względu dalsze rozważania będą dotyczyły głównie realizacji funkcji analogowych w postaci półprzewodnikowych układów scalonych.

Dość rozpowszechniony jest pogląd, że produkcję układów scalonych analogowych (u.s.a.) rozpoczęto znacznie później niż produkcję układów scalonych cyfrowych. Nie jest to całkowicie słuszne, gdyż pierwsze opracowania monolitycznych wzmacniaczy szerokopasmowych powstały już w 1963 r. Jednakże w tej początkowej fazie rozwoju technologii układów scalonych dominowało przekonanie o bardzo ograniczonych możliwościach scalania układów analogowych. Małe zakresy wartości nominalnych oraz bardzo duże tolerancje elementów biernych nie stanowią istotnej przeszkody w scalaniu układów cyfrowych, lecz zdają się znacznie ograniczać możliwości scalania układów analogowych. Ten pesymistyczny wniosek jest prawdziwy tylko wówczas, gdy próbuje się wykonać w postaci scalonej układ zaprojektowany do realizacji z elementów dyskretnych. Jednakże postępując w ten sposób akceptuje się niedostatki struktury monolitycznej, nie korzystając z jej zalet. Aby prawidłowo wykorzystać takie specjalne cechy scalenia monolitycznego, jak np. preferowanie elementów czynnych lub współbieżność zmian parametrów wszystkich elementów we wspólnej mikropłytkę, konieczne jest inne podejście do zagadnień projektowania układów.

Układ zaprojektowany do realizacji z elementów dyskretnych nie może być „ślepo” odwzorowywany w postaci scalonej. Równie słuszne jest twierdzenie odwrotne, że nie ma sensu odtwarzanie z elementów dyskretnych układu zaprojektowanego do realizacji w postaci scalonej. Przenoszenie bez zmian nawet najlepszych projektów układów z techniki dyskretniej na scaloną i odwrotnie daje konstrukcje dalekie od optymalnych. Dotyczy to zwłaszcza układów analogowych, które charakteryzują się większą złożonością i elastycznością rozwiązań projektowych niż

układy cyfrowe. Dlatego warunkiem rozwoju produkcji u.s.a. było opracowanie nowych zasad projektowania, nowych rozwiązań schematów elektrycznych, nowych nawyków konstruktorskich, słowem — konieczne było wykształcenie nowych kadr konstruktorów. A oto podstawowe różnice w zasadach konstruowania układów scalonych i układów z elementów dyskretnych.

Konstruktor układów z elementów dyskretnych ma do dyspozycji różnorodne elementy czynne i bierne o, praktycznie biorąc, nieograniczonych zakresach wartości nominalnych (rezystory od ułamka oma do gigaomów, kondensatory od pikofaradów do wielu mikrofaradów), przy czym jego wiedza o elementach może być, praktycznie biorąc, ograniczona do znajomości charakterystyk i parametrów końcówkowych. Dla projektów tych układów jest charakterystyczne dążenie do minimalizacji liczby elementów czynnych (głównie tranzystorów), znacznie droższych niż elementy bierne.

Od konstruktora układów scalonych jest wymagana wiedza nie tylko „układowa”, lecz również „elementowa” w zakresie konstrukcji, technologii i zasad fizycznych działania elementów. Jest to konsekwencją faktu, że wszystkie elementy wraz z ich połączeniami powstają w jednym ciągłym procesie technologicznym i właściwości tego procesu mają bezpośredni wpływ na parametry układu. Inne są również zasady rachunku ekonomicznego, gdyż nie operuje się cenami poszczególnych elementów, lecz podstawowym wskaźnikiem służącym do oceny kosztów wytwarzania jest powierzchnia mikro płytki zawierającej układ scalony. Im mniejsze są powierzchnie układów scalonych, tym więcej można ich wyprodukować w jednym procesie, a więc mniejsza jest również ich cena jednostkowa. Dlatego należy dążyć do zastąpienia elementów biernych czynnymi (tranzystor zajmuje mniejszą powierzchnię niż kondensator o pojemności kilku pikofaradów lub rezystor o wartości powyżej 2 k Ω). Ponadto konstruktor układów scalonych powinien w sposób celowy wykorzystać takie właściwości jak współbieżność zmian parametrów elementów, zgodność temperatury poszczególnych elementów, a nawet istnienie sprzężeń pasożytniczych.

Pierwsze konstrukcje u.s.a., zaprojektowanych już z uwzględnieniem specyficznych właściwości technologii monolitycznej, pojawiły się w 1965 r. Należy tu wymienić przede wszystkim wzmacniacz operacyjny μA 702, opracowany przez J. Widlara w firmie Fairchild. Ten sam konstruktor opracował w następnych latach wzmacniacze operacyjne μA 709 oraz μA 741, które stały się najbardziej popularnymi na świecie układami scalonymi analogowymi. Pouczające jest zestawienie liczby

Porównanie liczby elementów czynnych i biernych w trzech kolejno opracowywanych wzmacniaczach operacyjnych

Tablica 9.1

Typ wzmacniacza	Liczba			Względny udział tranzystorów [%]
	Tranzystorów	Rezystorów	Kondensatorów	
μA 702	9	11	—	45
μA 709	15	15	—	50
μA 741	20	11	1	62,5

elementów czynnych i biernych w tych trzech wzmacniaczach (tabl. 9.1), które ilustruje prawidłową tendencję do zwiększania względnego udziału elementów czynnych.

W niniejszym rozdziale rozpatrzemy przede wszystkim rozwiązania układów elementarnych („podzespołów”) stanowiących typowe i specyficzne fragmenty u.s.a., jak na przykład źródła prądowe, napięciowe, obwody polaryzacji, stopnie wejściowe i wyjściowe. Następnie omówimy zasady działania oraz podstawowe parametry najpowszechniej stosowanych u.s.a., tj.:

- wzmacniaczy operacyjnych,
- komparatorów,
- stabilizatorów napięcia.

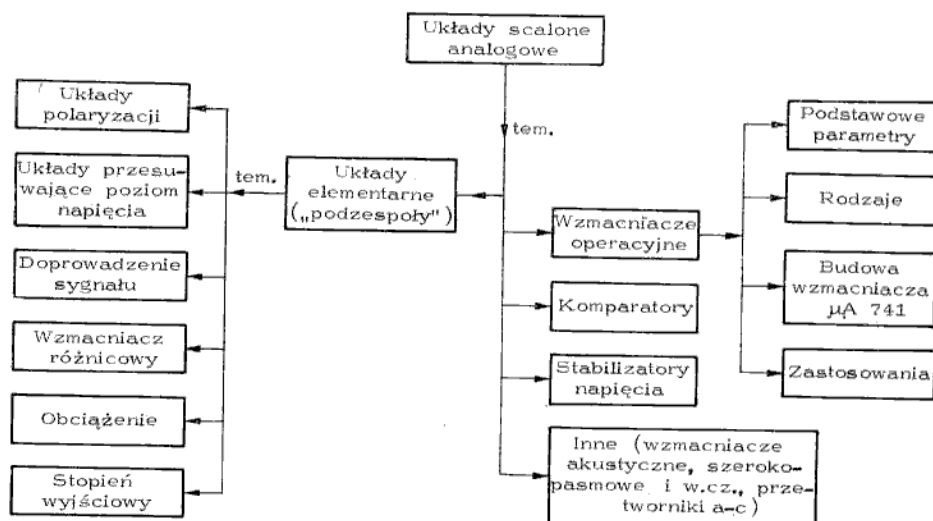
Oprócz tych trzech podstawowych grup u.s.a. należy wymienić również:

- wzmacniacze akustyczne,
- wzmacniacze szerokopasmowe i dla zakresu w.cz.,
- układy analogowo-cyfrowe.

Te układy nie będą rozpatrywane.

W ogóle należy stwierdzić, że niniejszy rozdział dostarcza tylko wstępnych informacji o u.s.a.; wyczerpujące opracowania tej tematyki można znaleźć w odpowiednich monografiach¹⁵.

Układ rozdziału przedstawiono schematycznie na rys. 9.1.



Rys. 9.1
Schematyczny układ rozdziału

Budowa układów elementarnych

9.1

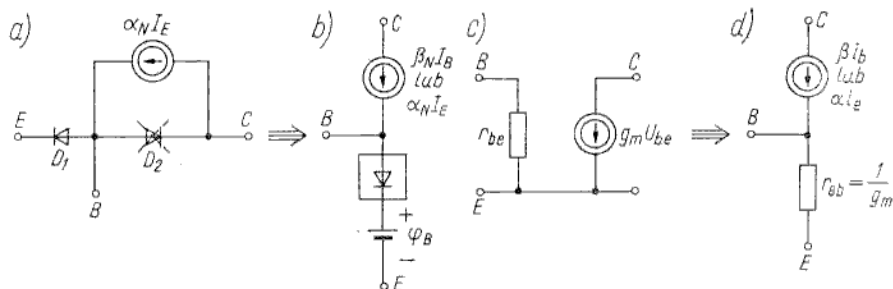
Zostaną rozpatrzone następujące układy elementarne, stanowiące typowe „podzespoły” u.s.a.:

- układy polaryzacji,
- układy przesuwające poziom napięcia,
- wejście sygnału,

¹⁵ Między innymi w literaturze krajowej: Z. Kulka, M. Nadachowski: *Linijowe układy scalone i ich zastosowania*. WKiŁ, Warszawa 1974.

- wzmacniacz różnicowy,
- obciążenie,
- stopień wyjściowy.

W analizie tych układów posłużono się uproszczonymi modelami tranzystora bipolarnego dla prądu stałego oraz dla małych sygnałów prądu zmiennego (rys. 9.2).



Rys. 9.2

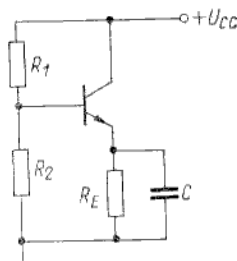
Uproszczone modele tranzystora bipolarnego: a), b) statyczne nieliniowe; c), d) dla małych sygnałów

Na rysunku 9.2a przedstawiono model statyczny nieliniowy dla zakresu normalnego. Ponieważ złącze *C-B* jest spolaryzowane w kierunku zaporowym, można zatem pominąć prąd wsteczny diody *D*₂, a rzeczywistą diodę *D*₁, spolaryzowaną w kierunku przewodzenia, można zastąpić szeregowym połączeniem diody idealnej i źródła napięcia φ_B (ok. 0,7 V). W ten sposób otrzymuje się prosty model pokazany na rys. 9.2b. Najprostszy model dla małych sygnałów prądu zmiennego przedstawiono na rys. 9.2c lub w postaci równoważnej — na rys. 9.2d.

Układy polaryzacji

9.1.1

Układ polaryzacji ustala punkt pracy tranzystora, przy czym najczęściej chodzi o ustalenie określonego prądu kolektora i napięcia na bazie. Na rysunku 9.3 przedstawiono obwód polaryzacji typowy dla układów z elementami dyskretnymi.



Rys. 9.3

Obwód polaryzacji tranzystora bipolarnego, stosowany w układach z elementami dyskretnymi

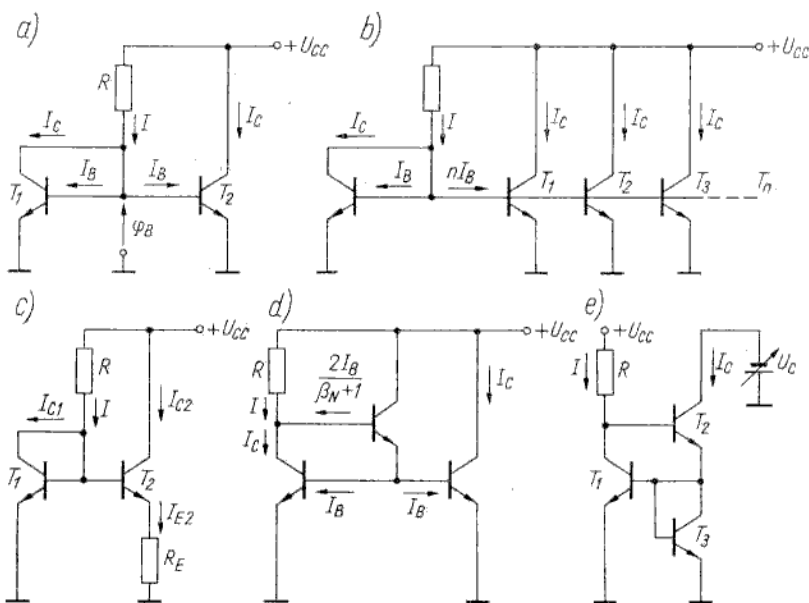
Dzielnik R_1, R_2 wyznacza potencjał bazy, a rezystor R_E stabilizuje prąd emitera. W przypadku zastosowania tego układu jako wzmacniacza do bazy jest doprowadzony sygnał wzmacniany, a w obwodzie kolektora jest włączony rezystor obciążenia, z którego jest zbierany sygnał wzmacniony. W celu wyeliminowania ujemnego sprzężenia zwrotnego dla sygnału prądu zmiennego rezystor R_E jest bocznikowany kondensatorem C .

Taki układ nie może być zrealizowany w postaci scalonej z uwagi na brak możliwości wykonania kondensatora monolitycznego (o wartości rzędu mikrofaradów) oraz rezystorów dzielnika (wartości rezystorów R_1, R_2 wynoszą zwykle powyżej kilkudziesięciu kiloomów). W układach scalonych trzeba stosować inne sposoby polaryzacji, które omówimy w dwu grupach, tj. najpierw układy ustalające prąd kolektora (źródła prądowe), a następnie układy ustalające napięcie bazy (źródła napięciowe).

Układy polaryzacji prądowej (źródła prądowe)

9.1.1.1

Kilka podstawowych układów źródeł prądowych przedstawiono na rys. 9.4. W analizie tych układów zakłada się, że wszystkie tranzystory mają jednakowe parametry, które jednakowo zmieniają się w funkcji temperatury. W układzie



Rys. 9.4
Podstawowe układy źródeł prądowych

pokazanym na rys. 9.4a, nazywanym niekiedy *układem Widlarsa*, tranzystor T_1 jest włączony jako dioda i służy do polaryzacji tranzystora T_2 . Oba tranzystory są jednakowe i mają równolegle włączone złącza baza-emiter, zatem prądy bazy i kolektora są w nich jednakowe.

Ponieważ

$$I = \frac{U_{CC} - \varphi_B}{R}$$

oraz z prawa Kirchhoffa

$$I = I_C + 2I_B$$

Zatem po uwzględnieniu zależności

$$I_C = \beta_N I_B$$

otrzymuje się

$$I_C = I \frac{\beta_N}{\beta_N + 2} = \frac{\beta_N}{\beta_N + 2} \frac{U_{CC} - \varphi_B}{R} \quad (9.1)$$

W normalnych warunkach można przyjąć, że

$$\beta_N \gg 2 \quad \text{oraz} \quad U_{CC} \gg \varphi_B$$

i stąd

$$I_C \approx \frac{U_{CC}}{R} \quad (9.2)$$

czyli wartość prądu kolektora jest, praktycznie biorąc, niezależna od wzmocnienia tranzystorów, jeżeli oczywiście założymy, że współczynniki β_N są jednakowe w obu tranzystorach. Ponieważ przez tranzystor T_2 płynie prąd kolektora, będący niejako repliką prądu wymuszonego w obwodzie tranzystora T_1 , taki układ nazywa się *źródłem prądu zwierciadlanego* (ang. *current mirror*) lub *układem polaryzacji zrównoważonej*. Na rysunku 9.4b przedstawiono układ wielokrotnego źródła prądowego, który działa podobnie jak układ z rys. 9.4a. Kolejne rysunki 9.4c, d, e przedstawiają różne modyfikacje podstawowego układu z rys. 9.4a.

Pierwsza modyfikacja (rys. 9.4c) dotyczy uzyskania małych prądów polaryzacji. W układzie podstawowym z rys. 9.4a uzyskanie małych prądów polaryzacji przy określonym napięciu U_{CC} wymaga stosowania dużych wartości rezystancji. Na przykład dla $U_{CC} = 5 \text{ V}$ prąd $5 \mu\text{A}$ uzyskuje się przy $R \approx 1 \text{ M}\Omega$.

W układzie scalonym półprzewodnikowym nie można wykonać rezystora o tak dużej wartości, dlatego konieczne jest zastosowanie innego rozwiązania, przedstawionego na rys. 9.4c.

Prądy kolektorów zależą w przybliżeniu wykładniczo od napięć emiter-baza

$$I_{C1} = \alpha_{N1} I_{ES1} \exp\left(\frac{U_{EB1}}{\varphi_T}\right) \quad (9.3)$$

$$I_{C2} = \alpha_{N2} I_{ES2} \exp\left(\frac{U_{EB2}}{\varphi_T}\right)$$

nawet niewielkie zatem różnice napięć U_{EB} w tranzystorach T_1 , T_2 powodują bardzo duże różnice prądów kolektora.

Ponieważ

$$U_{EB2} = U_{EB1} - R_E I_{E2}$$

Zatem przy założeniu, że:

$$\alpha_{N1} = \alpha_{N2}$$

$$I_{ES1} = I_{ES2}$$

$$I_{E2} \approx I_{C2}$$

oraz uwzględnieniu (9.3) można oszacować wartość rezystora R_E

$$R_E = \frac{\varphi_T}{I_{C2}} \ln \frac{I_{C1}}{I_{C2}} \quad (9.4)$$

Przykładowo prąd $I_{C2} = 5 \mu\text{A}$ przy $I_{C1} = 0,5 \text{ mA}$ (dla $U_{CC} = 5 \text{ V}$ oznacza to, że $R \approx 10 \text{ k}\Omega$) uzyskuje się dla $R_E \approx 23 \text{ k}\Omega$. Zatem zamiast $1 \text{ M}\Omega$, jak to było w poprzednim przypadku, są dwa rezystory o realizowalnych praktycznie wartościach 10 i $23 \text{ k}\Omega$.

Na rysunku 9.4d przedstawiono zmodyfikowany układ źródła prądowego, w którym wartość prądu bardzo słabo zależy od parametrów tranzystora. W tym układzie

$$I = I_C + \frac{2I_B}{\beta_N + 1}$$

a ponieważ $I_C = \beta_N I_B$, zatem

$$I_C = I \frac{\beta_N(\beta_N + 1)}{\beta_N(\beta_N + 1) + 2} \quad (9.5)$$

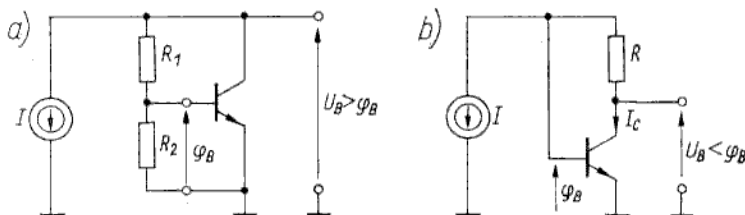
Jak widać z równania (9.5), nawet przy niewielkich wartościach współczynnika β_N prąd I_C jest z dużą dokładnością równy prądowi I . Istotnym problemem jest uniezależnienie wartości prądu od zmian obciążenia lub inaczej od zmian napięcia kolektor-emiter. Pod tym względem bardzo dobre właściwości ma tzw. źródło Wilsona przedstawione na rys. 9.4e. Jest to układ z silnym ujemnym sprzężeniem zwrotnym. Prąd zależny I_C płynie przez tranzystor T_3 , który włączony jako dioda steruje prąd kolektora tranzystora T_1 . Z kolei tranzystor T_1 steruje bazę tranzystora T_2 , a więc wpływa na wartość prądu I_C . Dokładność tego źródła prądowego można jeszcze zwiększyć włączając rezystory szeregowo z emiterami tranzystorów T_1 , T_3 .

Układy polaryzacji napięciowej

9.1.1.2

W niektórych przypadkach istotne jest tylko ustalenie napięcia polaryzacji bazy. Odnosi się to na przykład do stopni wyjściowych wzmacniaczy operacyjnych, których zadaniem jest wzmacnianie prądu. Jeżeli wymagane napięcie polaryzacji jest równe φ_B (ok. $0,7 \text{ V}$), to źródłem odniesienia może być dioda lub tranzystor włączony jako dioda (jak w układzie Widlara na rys. 9.4a). W tych przypadkach, w których jest wymagane napięcie polaryzacji U_B większe niż φ_B , stosuje się szeregowo połączone diody (n diod daje napięcie $U_B = n\varphi_B$) lub też stosuje się układ przedstawiony na rys. 9.5a. Analizę tego układu można przeprowadzić przy użyciu prostego modelu tranzystora pokazanego na rys. 9.2a. Jeżeli pominiemy się prąd bazy, to

$$U_B = \left(1 + \frac{R_1}{R_2}\right) \varphi_B \quad (9.6a)$$



Rys. 9.5
Układy polaryzacji napięciowej

W tych przypadkach, w których jest wymagane napięcie polaryzacji U_B mniejsze niż φ_B , stosuje się układ przedstawiony na rys. 9.5b. Jeżeli tranzystor nie jest zbyt silnie nasycony, to $I_C \approx I$, zatem

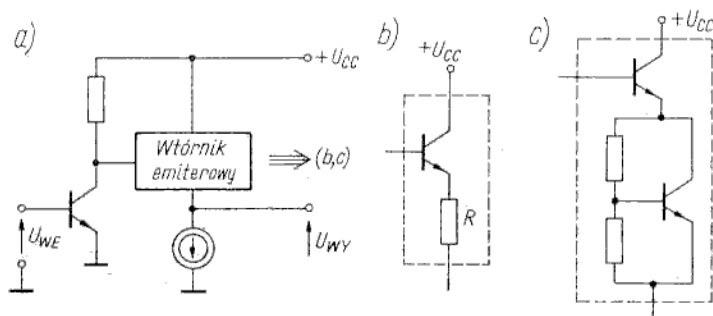
$$U_B = \varphi_B - I_C R \approx \varphi_B - IR \quad (9.6b)$$

Układy przesuwające poziom napięcia stałego

9.1.1.3

W układzie scalonym nie stosuje się kondensatorów sprzęgających kolejne stopnie połączone kaskadowo, gdyż nie ma możliwości realizacji tak dużych pojemności (rzędu mikrofaradów). Stosuje się natomiast sprzężenia bezpośrednie. Ponieważ przy pracy aktywnej tranzystora potencjał kolektora jest wyższy niż potencjał bazy (dla tranzystorów *n-p-n*), kolejne tranzystory są polaryzowane coraz wyższym napięciem. Jeżeli żąda się, żeby napięcia stałe na wyjściu i wejściu układu (wzmacniacza) były jednakowe, to jest konieczne przesunięcie poziomu napięcia, przy czym istotne jest, by układ przesuwający poziom nie powodował zbyt dużej straty wzmocnienia. W najprostszym przypadku można włączyć kilka szeregowo połączonych diod między wyjście poprzedniego stopnia a wejście następnego oraz zapewnić przepływ odpowiedniego prądu polaryzacji tych diod. *N* diod daje przesunięcie napięcia o wartość $N\varphi_B$. Można tu zastosować diodę Zenera (złącze emiter-baza spolaryzowane w kierunku zaporowym), co jednak umożliwia uzyskanie tylko jednej wartości przesunięcia, tj. ok. 6 V.

Dowolnie regulowane przesunięcie poziomu napięcia uzyskuje się w układzie przedstawionym na rys. 9.6a. Wtórnik emiterowy zapewnia odseparowanie dwu kolejnych



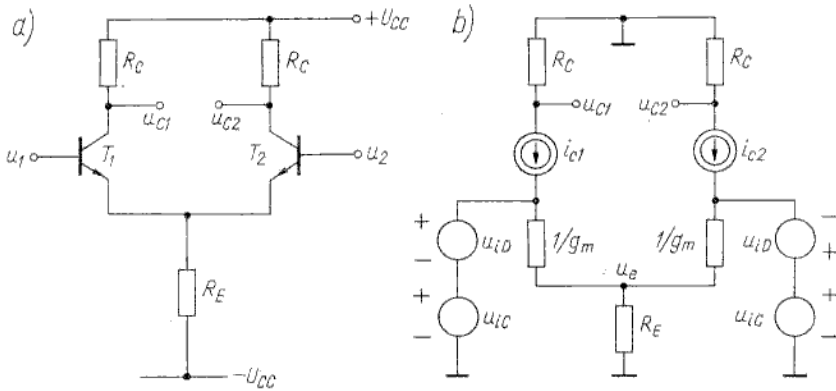
Rys. 9.6
Układ przesuwający poziom napięcia stałego

stopni. Na rysunku 9.6b, c przedstawiono dwa warianty układu wtórnika emiterowego. Wariant c) odznacza się mniejszą rezystancją wyjściową. Źródło prądowe zapewnia stałą wartość przesunięcia poziomu napięcia stałego.

Układy obróbki sygnału

9.1.2

Po omówieniu zagadnień związanych z ustalaniem pożądanego punktu pracy (p. 9.1.1) zostaną przedstawione niektóre zagadnienia związane z obróbką sygnału, tj. wejściem, wzmocnieniem i wyjściem sygnału.



Rys. 9.8
Wzmacniacz różnicowy (a) oraz jego schemat zastępczy (b)

realizacji w postaci scalonej. Podstawową właściwością wzmacniacza różnicowego jest zdolność wzmacniania tzw. sygnałów różnicowych (przychodzących na oba wejścia w fazie przeciwnej) oraz tłumienia tzw. sygnałów wspólnych (przychodzących na oba wejścia w fazie zgodnej). Istnieje więc możliwość wzmocnienia małych sygnałów różnicowych na tle bardzo dużych sygnałów wspólnych (na przykład na tle zakłóceń, które są podawane w zgodnej fazie na oba wejścia).

Przeprowadzimy przybliżoną analizę działania wzmacniacza różnicowego dla małych sygnałów przy następujących założeniach upraszczających:

- układ jest idealnie symetryczny (jednakowe tranzystory oraz jednakowe rezystory R_C);
- współczynnik $\alpha = 1$ w obu tranzystorach.

Przyjmując uproszczony model tranzystora (rys. 9.2d) otrzymuje się schemat zastępczy wzmacniacza różnicowego w postaci pokazanej na rys. 9.8b. Sygnał różnicowy oznaczono indeksem D (ang. *difference*), wspólny zaś — indeksem C (ang. *common*):

$$u_1 = u_{iC} + u_{iD} \quad (9.7)$$

$$u_2 = u_{iC} - u_{iD} \quad (9.8)$$

Najpierw zostaną rozpatrzone właściwości wzmacniacza dla sygnału różnicowego przy $u_{iC} = 0$. Ze względu na symetrię układu napięcie $u_e = 0$, czyli rezystor R_E nie ma żadnego wpływu na pracę układu dla sygnału różnicowego. Można zatem traktować układ jako złożony z dwu niezależnych połówek. Napięcia na kolektorach tranzystorów:

$$u_{C1} = -u_{iD} g_m R_C \quad (9.9)$$

$$u_{C2} = u_{iD} g_m R_C \quad (9.10)$$

Napięcie wyjściowe może być pobierane z jednego kolektora — tzw. *wyjście niesymetryczne*, lub z obu kolektorów — tzw. *wyjście symetryczne* (różnicowe). Wzmocnienie napięciowe:

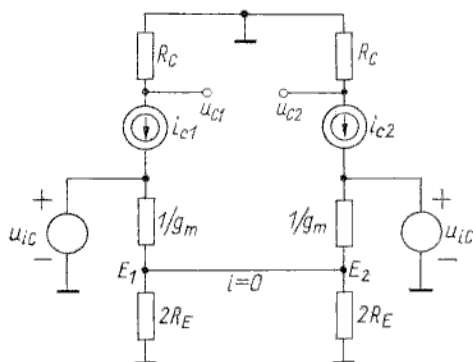
- dla wyjścia niesymetrycznego

$$k_{uD} = \frac{u_{C1}}{2u_{iD}} = -\frac{g_m}{2} R_C \quad (9.11)$$

— dla wyjścia symetrycznego

$$2k_{uD} = \frac{u_{C2} - u_{C1}}{2u_{iD}} = g_m R_C \quad (9.12)$$

Rozpatrzmy teraz właściwości wzmacniacza dla sygnału wspólnego $u_{iD} = 0$. Dla ułatwienia rozważań przerysowano schemat zastępczy w innej postaci (rys. 9.9). Z uwagi



Rys. 9.9
Schemat zastępczy wzmacniacza różnicowego w innej postaci

na symetrię układu można przyjąć, że między węzłami E_1 , E_2 nie płynie prąd, czyli w tym przypadku można również analizować dwie niezależne połowki układu. Napięcia na kolektorach tranzystorów

$$u_{C1} = u_{C2} = -u_{iC} \frac{R_C}{2R_E + \frac{1}{g_m}} \quad (9.13)$$

Wzmocnienie napięciowe:

— dla wyjścia niesymetrycznego

$$k_{uC} = \frac{u_{C1}}{u_{iC}} = - \frac{g_m R_C}{1 + 2g_m R_E} \quad (9.14)$$

— dla wyjścia symetrycznego

$$\frac{u_{C2} - u_{C1}}{u_{iC}} = 0 \quad (9.15)$$

Zwykle $g_m R_E \gg 1$, czyli

$$k_{uC} \approx - \frac{R_C}{2R_E} \quad (9.16)$$

Podstawowym parametrem wzmacniacza różnicowego jest współczynnik tłumienia sygnału wspólnego H (ang. CMRR — *Common Mode Rejection Ratio*), definiowany jako iloraz wzmocnienia sygnału różnicowego i wzmocnienia sygnału wspólnego

$$H = \frac{k_{uD}}{k_{uC}} = \frac{1}{2} + g_m R_E \quad (9.17a)$$

lub w przybliżeniu

$$H \approx g_m R_E \quad (9.17b)$$

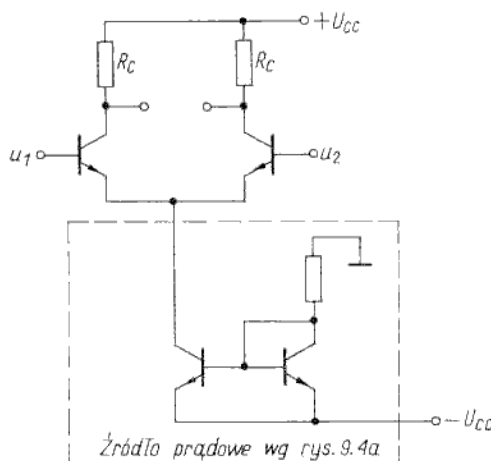
Na przykład dla $I_C = 1 \text{ mA}$ (czyli $g_m = \frac{1}{25} \text{ S}$) oraz $R_E = 5 \text{ k}\Omega$ jest

$$H \approx 200$$

Z zależności (9.17a) wynika, że w celu zwiększenia wartości H należy zwiększać g_m i R_E . Ponieważ

$$g_m = \frac{1}{r_{eb'}} = \frac{I_E}{\varphi_T} \quad (9.18)$$

należy pracować przy jak największych prądach emitera. Z kolei duży prąd emitera i duża wartość rezystora R_E oznaczają konieczność stosowania dużych wartości napięcia zasilania. Korzystne byłoby skonstruowanie takiego układu, w którym wspólna rezystancja w emiterze miałaby małą wartość dla prądu polaryzacji (mały spadek napięcia stałego) i jednocześnie dużą wartość dla sygnału wzmacnianego. Inaczej mówiąc, należy włączyć w obwód emiterów element charakteryzujący się małą rezystancją statyczną i bardzo dużą rezystancją dynamiczną, tj. element o właściwościach źródła prądowego.



Rys. 9.10
Wzmacniacz różnicowy ze źródłem prądowym

Na rysunku 9.10 przedstawiono wzmacniacz różnicowy ze źródłem prądowym zamiast rezystora R_E . To źródło rozważano odrębnie w p. 9.1.1.1 (rys. 9.4a).

Obok wzmocnienia napięciowego i współczynnika tłumienia sygnału wspólnego do najważniejszych parametrów wzmacniacza różnicowego należy jeszcze zaliczyć rezystancję wejściową dla sygnału różnicowego.

Rezystancja wejściowa różnicowa jest definiowana jako iloraz różnicy napięć na obu wejściach i prądu wpływającego przez jeden zacisk wejściowy

$$R_{iD} = \frac{u_1 - u_2}{i_b} = \frac{2u_{iD}}{i_b}$$

Na podstawie schematu zastępczego (rys. 9.8b)

$$i_b = \frac{u_{iD} g_m}{\beta}$$

Stąd uwzględniając (9.18)

$$R_{iD} = \frac{2\beta}{g_m} = 2\varphi_T \frac{\beta}{I_E} \approx \frac{2\varphi_T}{I_B} \quad (9.19).$$

Na przykład dla $\beta = 100$, $I_E = 1$ mA, otrzymuje się $R_{iD} = 5$ k Ω .

Zwykle wymaga się, żeby rezystancja wejściowa była jak największa. Można wymienić kilka sposobów zwiększenia wartości R_{iD} :

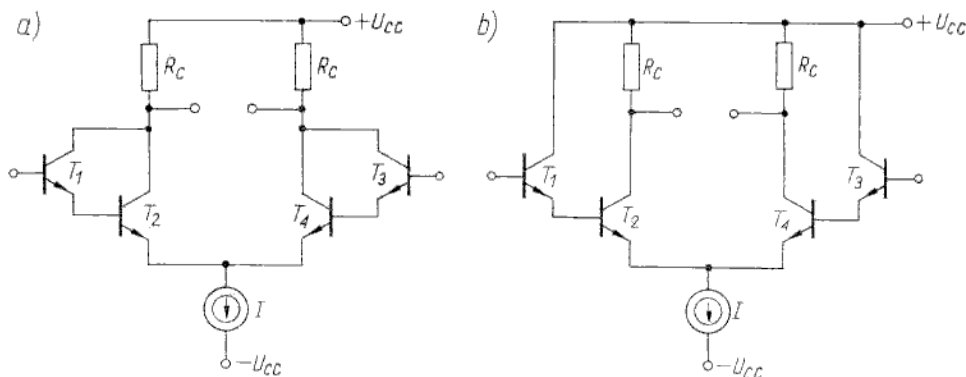
- włączenie rezystorów w obu obwodach emiterów,
- zmniejszenie prądu emitera,
- zwiększenie współczynnika β ,
- zamiana tranzystorów bipolarnych na unipolarne.

Pierwszy sposób jest często stosowany w układach z elementów indywidualnych. Na podstawie schematu zastępczego i wzoru (9.19) łatwo można się zorientować (rys. 9.8b), że włączając jednakowe rezystory R szeregowo z rezystancjami $1/g_m$ otrzymuje się

$$R_{iD} = 2\beta \left(\frac{\varphi_T}{I_E} + R \right) \quad (9.20)$$

Na przykład dla $\beta = 100$, $I_E = 1$ mA, $R = 5$ k Ω , otrzymuje się $R_{iD} \approx 1$ M Ω . Jednakże rezystory włączone w obwodach emiterów powodują ujemne sprzężenia zwrotne dla sygnału różnicowego, co prowadzi do dużego zmniejszenia wzmocnienia.

Drugi sposób (zmniejszenie prądu emitera) również powoduje niekorzystne zmniejszenie wzmocnienia, gdyż maleje wartość g_m . Ponadto konieczne jest w tym przypadku stosowanie specjalnych tranzystorów o dużym wzmocnieniu (dużym β) przy małych prądach emitera. Między innymi w niektórych układach scalonych (na przykład we wzmacniaczu operacyjnym LM 108) stosuje się tzw. tranzystory „super-beta”, które spełniają drugi i trzeci postulat, tj. mają duże wzmocnienie ($\beta \approx 5000$) przy małych prądach emitera (kilka mikroamperów). Są to tranzystory o niezwykle cienkiej bazie (ok. 0,1 μ m), dlatego mają bardzo małe napięcie przebicia (ok. 1 V) i mogą pracować tylko w połączeniu z innymi tranzystorami w konfiguracji zapewniającej małą wartość napięcia U_{CE} .



Rys. 9.11

Wzmacniacz różnicowy o dużej rezystancji wejściowej: a) z parą tranzystorów w układzie Darlingtona; b) z parą tranzystorów w układzie WC-WE

Wartość współczynnika β można efektywnie zwiększyć stosując zamiast jednego tranzystora parę tranzystorów połączonych w układzie Darlingtona (rys. 9.11a). Pary tranzystorów T_1, T_2 oraz T_3, T_4 mają takie właściwości jak pojedyncze tranzystory o efektywnym wzmacnieniu β^2 . Zatem dla $\beta = 100, I_E = 1 \text{ mA}$, otrzymuje się $R_{iD} = 500 \text{ k}\Omega$. Wadą tego układu jest silne zwielokrotnienie pojemności C_{jc} (zjawisko Millera), a więc znaczne ograniczenie górnej częstotliwości pracy. Pod tym względem lepsza jest konfiguracja WC-WE, przedstawiona na rys. 9.11b, w której nie występuje zjawisko Millera, a rezystancja wejściowa jest tylko nieznacznie mniejsza niż dla układu Darlingtona.

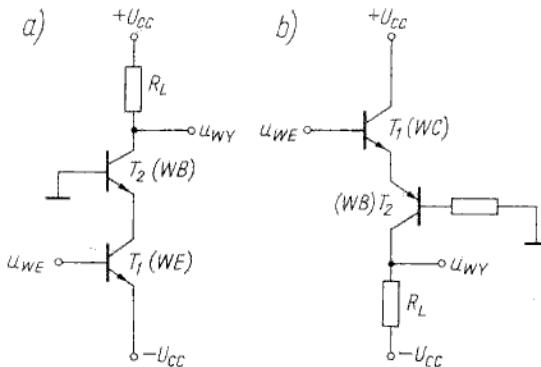
Bardzo dużą rezystancję wejściową (kilkaset megaomów) otrzymuje się zamieniając bipolarnie tranzystory wejściowe na tranzystory polowe PNFET.

Największe wartości rezystancji wejściowej (ok. $10^{12} \Omega$) można otrzymać włączając na wejściach tranzystory MOS. Należy dodać, że stosowanie tranzystorów „super-beta” lub unipolarnych jest związane nie tylko z dążeniem do uzyskania jak największej rezystancji wejściowej, lecz jest często podyktowane dążeniem do zmniejszenia wejściowych prądów polaryzacji.

Obciążenie

9.1.2.3

Pasmo częstotliwościowe układu WE z rezystorem R_L w obciążeniu kolektorowym jest ograniczone przez pojemność C_{jc} zwielokrotnioną $(1 + g_m R_L)$ razy wskutek zjawiska Millera. Żeby zwiększyć częstotliwość graniczną, stosuje się układ kaskodowy, tj. połączenia dwu tranzystorów w konfiguracji WE-WB (rys. 9.12). W tym układzie tranzystor T_1 jest obciążony rezystancją wejściową



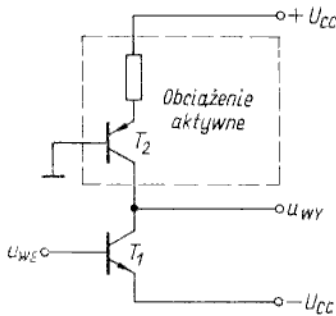
Rys. 9.12 Konfiguracja WE-WB (kaskoda)

tranzystora T_2 , która wynosi $r_{eb'} = 1/g_m$. Zatem wzmacnienie napięciowe tranzystora T_1 , dla którego $R_L = 1/g_m$ wynosi $g_m R_L = 1$ (zakłada się, że oba tranzystory są jednakowe), czyli nie występuje zjawisko zwielokrotnienia pojemności C_{jc} . Ponieważ tranzystor T_2 jest obciążony rezystancją R_L , wzmacnienie napięciowe całej kaskody pozostaje takie jak dla pojedynczego tranzystora włączonego w układzie WE i obciążonego rezystancją R_L . Można powiedzieć, że w kaskodzie pierwszy tranzystor wzmacnia prąd, a drugi napięcie, przy czym zostaje wyeliminowany niekorzystny wpływ zjawiska Millera na częstotliwość graniczną wzmacniacza¹⁾. Analogiczne korzyści daje również połączenie dwu tranzystorów *n-p-n* oraz *p-n-p* w konfiguracji WC-WB (rys. 9.12b).

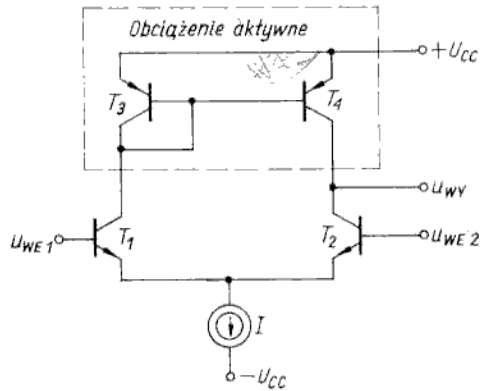
¹⁾ Zjawisko Millera nie występuje w układzie WB, gdyż ten układ nie odwraca fazy napięcia wyjściowego w stosunku do napięcia wejściowego.

Dążąc do uzyskania jak największego wzmocnienia, należy zwiększać wartość rezystancji R_L , co jednakże powoduje konieczność zwiększania napięcia zasilania. Tę trudność można ominąć stosując tzw. obciążenie aktywne, które ma bardzo dużą rezystancję dynamiczną (dla sygnału wzmacnianego) przy jednocześnie małej rezystancji statycznej (dla prądu polaryzacji). Funkcję obciążenia aktywnego o takich właściwościach może spełniać tranzystor (rys. 9.13).

Bardzo interesujące właściwości ma wzmacniacz różnicowy z obciążeniem aktywnym w postaci pary tranzystorów (T_3, T_4 na rys. 9.14). Tranzystory T_3, T_4 (typu



Rys. 9.13
Tranzystor bipolarny jako obciążenie aktywne



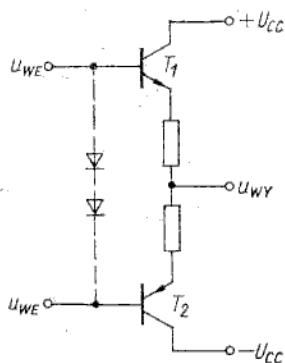
Rys. 9.14
Wzmacniacz różnicowy z obciążeniem aktywnym w postaci pary tranzystorów

$p-n-p$) tworzą zwierciadlane źródło prądowe, podobnie jak na rys. 9.4a. Wzmacniacz ma tylko wyjście niesymetryczne z kolektora tranzystora T_2 , przy czym wzmocnienie napięciowe niesymetryczne jest równe wzmocnieniu symetrycznemu (w układzie z rezystorami w obwodach kolektorów wzmocnienie niesymetryczne jest 2-krotnie mniejsze niż wzmocnienie symetryczne). Dzieje się tak dlatego, że obciążenie w kolektorze tranzystora T_1 dla sygnału zmiennego jest bardzo małe (rezystancja dynamiczna tranzystora T_3 , włączonego jako dioda, jest bardzo mała), a więc tranzystor T_1 pracuje w układzie WC podając wzmocniony sygnał prądowy z emitera na tranzystor T_2 pracujący ze względu na ten sygnał w układzie WB. Napięcie u_{WE2} jest zatem wzmacniane przez tranzystor T_2 pracujący w układzie WE, natomiast napięcie u_{WE1} jest wzmacniane przez tranzystory T_1, T_2 , tworzące konfigurację WC-WB.

Stopień wyjściowy

9.1.2.4

Zadaniem stopnia wyjściowego w u.s.a. jest zwykle dostarczenie dużego prądu do obciążenia, przy czym rezystancja wyjściowa tego stopnia powinna być jak najmniejsza. Wzmocnienie napięciowe jest mało istotne i na ogół nie jest większe niż 10. Ze względu na wymagania dobrej sprawności energetycznej zwykle stopień wyjściowy pracuje w klasie AB lub B. Konfiguracja komplementarna, tj. połączenie tranzystorów $n-p-n$ i $p-n-p$ (rys. 9.15), jest szczególnie predestynowana do takich zastosowań. Tranzystory T_1, T_2 przewodzą prąd na przemian przy sygnałach dodatnich na bazie tranzystora T_1 oraz ujemnych na bazie tranzystora T_2 .



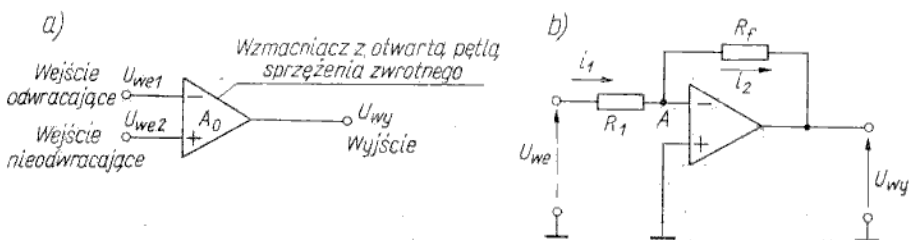
Rys. 9.15
Stopień wyjściowy w konfiguracji komplementarnej

Wzmacniacze operacyjne

9.2

Wzmacniacze operacyjne stanowią najważniejszą grupę analogowych układów scalonych dla zastosowań profesjonalnych. Wzmacniaczem operacyjnym (w.o.) nazywa się układ, który spełnia przede wszystkim następujące trzy warunki:

- bardzo duże wzmocnienie, co najmniej 1000 V/V (60 dB);
- bezpośrednie sprzężenia, tj. możliwość wzmocniania napięć stałych i zmiennych;
- odwrócenie fazy sygnału wyjściowego w stosunku do sygnału wejściowego, tj. możliwość włączenia zewnętrznej pętli ujemnego sprzężenia zwrotnego.



Rys. 9.16
Symbol wzmacniacza operacyjnego (a) oraz jego układ włączenia ze sprzężeniem rezystancyjnym (b)

Symbol wzmacniacza operacyjnego przedstawiono na rys. 9.16a. Ma on wejście symetryczne i wyjście niesymetryczne. Wejście, któremu odpowiada przesunięcie fazy sygnału wyjściowego o 180° , jest nazywane *odwracającym* i oznaczane przez „-”. Sygnał podany na wejście oznaczone „+” jest wzmacniany bez zmiany fazy, dlatego to wejście jest nazywane *nieodwracającym*.

Idealny wzmacniacz operacyjny powinien mieć właściwości źródła napięciowego sterowanego napięciowo i powinien się charakteryzować następującymi parametrami:

- nieskończenie dużym wzmocnieniem przy otwartej pętli sprzężenia zwrotnego ($A_o \rightarrow \infty$);
- impedencją wyjściową równą zero;
- nieskończenie szerokim pasmem przenoszenia częstotliwości;
- napięciem wyjściowym równym zero przy zerowej wartości różnicowego napięcia wejściowego;

- zerowym prądem wejściowym;
- brakiem wrażliwości na wpływ temperatury;
- brakiem zniekształceń.

Właściwości funkcjonalne wzmacniacza operacyjnego, praktycznie biorąc, zależą tylko od parametrów sprzężenia zwrotnego. Wykażemy to na przykładzie wzmacniacza ze sprzężeniem rezystancyjnym (rys. 9.16b). Dla uproszczenia rozważań przyjęto, że wejście nieodwracające jest uziemione, czyli cały sygnał podany na wejście odwracające jest sygnałem różnicowym. Interesuje nas najpierw wartość napięcia bezpośrednio na wejściu wzmacniacza (potencjał u_A w węźle A). To napięcie po wzmożeniu A_o razy daje napięcie wyjściowe u_{WY} . Zatem

$$u_A = \frac{u_{WY}}{A_o} \quad (9.21)$$

Jeżeli $A_o \rightarrow \infty$, to $u_A \rightarrow 0$, czyli w dobrym wzmacniaczu operacyjnym $u_A \approx 0$. Można więc przyjąć, że węzeł A ma potencjał masy, czyli jest wirtualnie (pozornie) uziemiony. W związku z tym prąd i_1 płynący przez rezystor R_1 nie rozgałęzia się do wejścia wzmacniacza, lecz w całości płynie do wyjścia przez rezystor sprzężenia zwrotnego R_f

$$i_1 = i_2$$

Ponieważ:

$$i_1 = \frac{u_{WE} - u_A}{R_1}; \quad i_2 = \frac{u_A - u_{WY}}{R_f}; \quad u_A \approx 0$$

zatem

$$\frac{u_{WE}}{R_1} = - \frac{u_{WY}}{R_f}$$

A więc wzmożenie całego układu z pętlą sprzężenia zwrotnego

$$A_f = \frac{u_{WY}}{u_{WE}} = - \frac{R_f}{R_1} \quad (9.22)$$

Wzrosty rezystorów
sygnału wyjściowego

Widać więc, że wzmożenie zależy wyłącznie od wartości zewnętrznych rezystorów (znak „-” we wzorze (9.22) oznacza odwrócenie fazy sygnału względem wejściowego).

Rozpatrzmy teraz kolejno następujące zagadnienia:

- podstawowe parametry w.o.;
- rodzaje w.o.;
- budowa wzmacniacza μA 741;
- zastosowania w.o.

9.2.1

Podstawowe parametry wzmacniaczy operacyjnych

Wymagają warunkom
dokonać wyboru w.o.
od uwagę rzeczywiste
Definicje, oznaczenia
2.

Parametry rzeczywistego w.o. tylko w przybliżeniu odpowiadają idealizującym, wymienionym w poprzednim punkcie. Aby odpowiednio do określonych zastosowań, trzeba wziąć pod uwagę wartości wielu parametrów specyfikowanych w katalogach. W tabeli 9.1 jednostki podstawowych parametrów zestawiono w tabl. 9.

Tablica 9.2

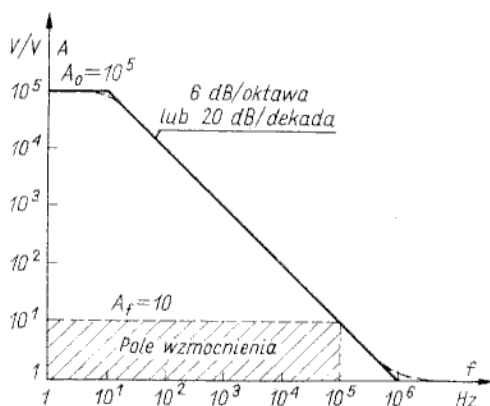
Nazwa i definicja parametru	Ozna- czenie	Jednos- tka	Wartość	
			ide- alna	typowa rzeczywista
1	2	3	4	5
<p>1) Napięciowe wzmocnienie różnicowe z otwartą pętlą. Jest to stosunek przyrostu napięcia wyjściowego do wywołującego ten przyrost napięcia wejściowego</p> $A_o = \frac{\Delta U_o}{\Delta(U_{i1} - U_{i2})} = \frac{\Delta U_o}{\Delta U_{iD}}$	A_o	V/V lub dB	∞	$10^4 \dots 10^5$ V/V
<p>2) Współczynnik tłumienia sygnału wspólnego. Jest to stosunek przyrostu wspólnego napięcia wejściowego do przyrostu różnicowego napięcia wejściowego, który wywołuje taki sam przyrost napięcia wyjściowego</p> $H = \frac{\Delta U_{iC}}{\Delta U_{iD}} \quad \text{przy} \quad \Delta U_o = \text{const}$ <p>inaczej</p> $H = \frac{\Delta U_{iC}}{\Delta U_o} \cdot \frac{\Delta U_o}{\Delta U_{iD}} = \frac{A_o}{A_C}$ <p>przy czym A_C — wzmocnienie napięciowe dla sygnału wspólnego</p>	H	dB	∞	60...100 dB
<p>3) Wejściowe napięcie niezrównoważenia. Jest to stałe napięcie różnicowe wymagane do uzyskania zerowej wartości napięcia wyjściowego przy otwartej pętli</p>	U_{I0}	mV	0	0,5...5 mV
<p>4) Wejściowy prąd niezrównoważenia. Jest to różnica stałych prądów polaryzacji wpływających do obu wejść wzmacniacza, gdy stałe napięcie wyjściowe jest równe zero</p>	I_{I0}	nA	0	$5 \cdot 10^{-4} \dots 5 \cdot 10^2$ nA
<p>5) Wejściowy prąd polaryzacji. Jest to średnia arytmetyczna prądów stałych wpływających do obu wejść wzmacniacza zrównoważonego</p>	I_{IB}	nA	0	$5 \cdot 10^{-3} \dots 5 \cdot 10^3$ nA
<p>6) Rezystancja wejściowa dla różnicowego rodzaju pracy. Jest to rezystancja widziana między dwoma wejściami wzmacniacza</p>	R_{iD}	MΩ	∞	0,05...10 ⁶ MΩ
<p>7) Dryf cieplny wejściowego napięcia niezrównoważenia. Jest to stosunek przyrostu wejściowego napięcia niezrównoważenia do przyrostu temperatury wywołującego ten przyrost napięcia</p> $\alpha_{vI0} = \frac{U_{I0}(T_2) - U_{I0}(T_1)}{T_2 - T_1}$	α_{vI0}	μV/°C	0	1...10 μV/°C
<p>8) Częstotliwość graniczna f_1. Jest to częstotliwość, przy której moduł wzmocnienia jest równy jedności</p>	f_1	MHz	∞	1...100 MHz

Tablica 9.2 (cd.)

1	2	3	4	5
9) Szybkość narastania napięcia wyjściowego. Jest to pochodna napięcia wyjściowego w funkcji czasu po podaniu sygnału skoku jednostkowego na wejście wzmacniacza pracującego ze wzmocnieniem równym jedności (układ wtórnika napięciowego)	Sk (ang. <i>slew rate</i>)	$V/\mu s$	∞	0,3...100 $V/\mu s$
10) Maksymalny prąd wyjściowy	I_{os}	mA	∞	10...100 mA
11) Napięcie zasilania	U_{CC}	V	—	± 15 V
12) Moc pobierana. Jest to moc pobierana przez wzmacniacz przy zerowym napięciu stałym na wyjściu	P_z	mW	0	0,1...200 mW
13) Zakres temperatury pracy	$T_1 \dots T_2$	$^{\circ}C$	—	-55...+125 $^{\circ}C$

Uwaga. Wyjątkowo w tabelicy zastosowano angielskie indeksy dla oznaczenia wielkości wejściowych i wyjściowych (na przykład U_I , U_o zamiast U_{WE} , U_{WY}), gdyż taki system oznaczeń obowiązuje dla parametrów katalogowych.

Dodatkowego komentarza wymaga zagadnienie pasma przenoszonych częstotliwości. Wzmocnienie przy otwartej pętli powyżej pewnej częstotliwości maleje 6 dB na oktawę (inaczej 20 dB na dekadę). Na rysunku 9.17 przedstawiono dla przy-



Rys. 9.17
Zależność wzmocnienia od częstotliwości
dla wzmacniacza μA 741

kładu wykres wzmocnienia w funkcji częstotliwości dla wzmacniacza μA 741. Powyżej częstotliwości 10 Hz wykres ma nachylenie 6 dB/oktawę, czyli odpowiada jednobiegunowej funkcji przenoszenia. Z ogólnych właściwości jednobiegunowej funkcji przenoszenia wiadomo, że iloczyn wzmocnienia przez częstotliwość, wyznaczony w zakresie opadania wykresu $A(f)$, ma stałą wartość. Iloczyn Af jest nazywany *polem wzmocnienia*. W granicznym przypadku dla $A = 1$ pole wzmocnienia jest równe częstotliwości f_1 , która dla rozpatrywanego przykładu wynosi 1 MHz. Stąd wynika bardzo ważny wniosek, że zakres częstotliwości sygnałów przenoszonych z jednakowym wzmocnieniem (pasmo przenoszenia)

zależy od wartości wzmożenia A_f , regulowanej obwodem sprzężenia zwrotnego. Przykładowo na rys. 9.17 pokazano, że przy wzmożeniu $A_f = 10$ pasmo przenoszonych częstotliwości wynosi 100 kHz.

Rodzaje wzmacniaczy operacyjnych

9.2.2

Różnorodność zastosowań wzmacniaczy operacyjnych powoduje, że na świecie produkuje się kilkadziesiąt podstawowych typów w.o. o parametrach optymalizowanych pod kątem określonych zastosowań. Ogólnie w.o. można sklasyfikować w następujących grupach:

- ogólnego przeznaczenia;
- szerokopasmowe (szybkie);
- do zastosowań dokładnych, w których są wymagane: duża rezystancja wejściowa, mały dryf i małe szумы;
- do innych zastosowań specjalnych.

Pierwszym wzmacniaczem operacyjnym ogólnego przeznaczenia, który zyskał szerokie rozpowszechnienie, był układ scalony μA 709 opracowany w firmie Fairchild¹⁾. Wymaga on dołączenia zewnętrznych elementów (dwu kondensatorów) w celu kompensacji częstotliwościowej. Konieczne jest również zewnętrzne ograniczenie prądu wyjściowego, gdyż wzmacniacz może się uszkodzić przy zwarciu wyjścia do masy. Tych wad nie ma wzmacniacz μA 741, stanowiący obecnie podstawowy i najpowszechniej stosowany typ w.o. Wzmacniacz μA 741 ma wewnętrzną kompensację częstotliwościową i wewnętrzne zabezpieczenie od zwarcie wyjścia.

W grupie wzmacniaczy szerokopasmowych podstawowym jest układ μA 715.

W grupie wzmacniaczy dokładnych należy wymienić przede wszystkim układy scalone LM 108 (z tranzystorami „super-beta”), μA 777, μA 740 (z tranzystorami PNFET) oraz CA 3130 (z tranzystorami MOS).

Do zastosowań specjalnych (technika kosmiczna, telemetria biomedyczna) są wytwarzane wzmacniacze o minimalnym poborze mocy (rzędu mikrowatów). Są również znane wzmacniacze o bardzo dużej mocy, na przykład μA 791, którego pobór mocy wynosi 15 W, a maksymalny prąd wyjściowy 1,2 A.

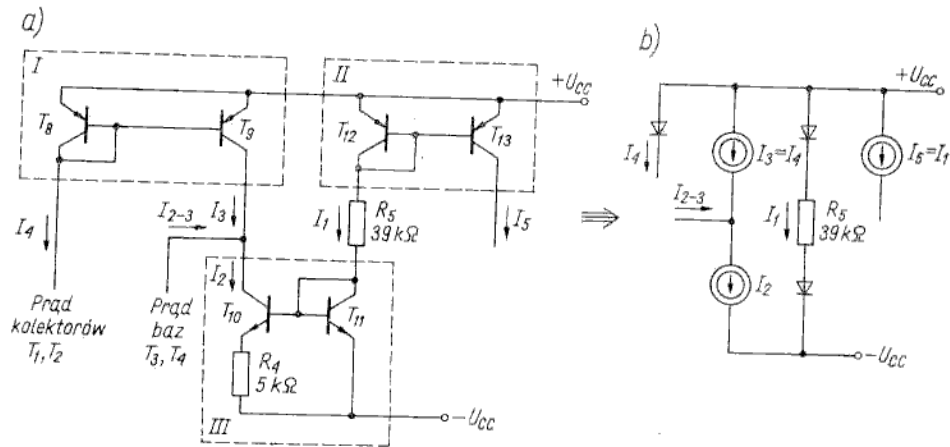
Parametry kilku podstawowych typów wzmacniaczy operacyjnych porównano w tabl. 9.3.

Wzmacniacz operacyjny μA 741

9.2.3

Tytułem przykładu rozpatrzmy budowę najbardziej rozpowszechnionego wzmacniacza typu μA 741. Dokonamy dekompozycji schematu tego wzmacniacza (rys. 9.18) na elementarne obwody rozpatrywane w p. 9.1. Będzie to zarazem ilustracja sposobu „czytania” schematów u.s.a. We wzmacniaczu μA 741, po-

¹⁾ Większość nowych opracowań w.o. należy do trzech firm: Fairchild (oznaczenie literowe μA), National Semiconductor (oznaczenie literowe LM) oraz Motorola (oznaczenie literowe MC). Inne firmy na ogół produkują odpowiedniki w.o. opracowanych przez te trzy firmy, przy czym stosują własne oznaczenia literowe i oznaczenia cyfrowe zgodne z pierwotnymi oznaczeniami, nadanymi przez firmę licencjodawcę. Na przykład wzmacniacz μA 709 jest produkowany przez szereg firm pod następującymi nazwami: LM 709 (National Semiconductor), SN 72 709 (Texas Instruments), SFC 2 709 (Sescosem), MC I 709 (Motorola) — (podkreślenia autora).



Rys. 9.20
Układ polaryzacji we wzmacniaczu μA 741

oraz T_{12} , T_{13} tworzą zwierciadlane źródła prądowe, omawiane w p. 9.1.1.1 (rys. 9.4a), natomiast tranzystory T_{10} , T_{11} wraz z rezystorem R_4 tworzą zmodyfikowane źródło małego prądu, rozważane wcześniej na rys. 9.4c.

W rozpatrywanym układzie podstawowym prądem wymuszającym jest I_1 , czyli prąd płynący od $+U_{CC}$ do $-U_{CC}$ przez rezystor R_5 z dwiema diodami włączonymi szeregowo. Ponieważ spadek napięcia na każdej diodzie wynosi zaledwie ok. 0,7 V, w przybliżeniu

$$I_1 \approx \frac{2U_{CC}}{R_5} \approx 0,7 \text{ mA}$$

Z właściwości zwierciadlanego źródła prądowego wynika, że

$$I_5 = I_1$$

natomiast I_2 można wyznaczyć na podstawie zależności (9.4), która w rozpatrywanym przypadku ma postać

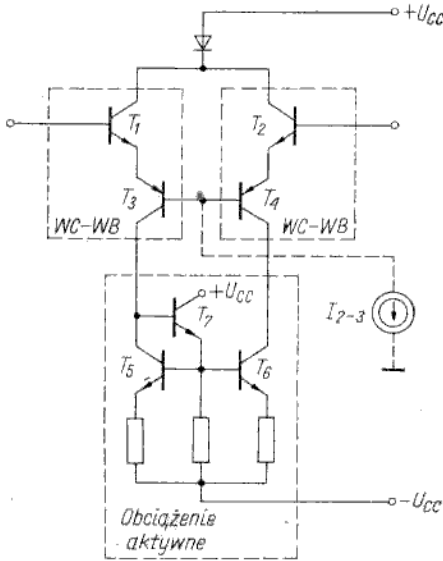
$$I_2 = \frac{\varphi_T}{R_4} \ln \frac{I_1}{I_2}$$

Po podstawieniu poszczególnych wartości otrzymuje się $I_2 \approx 20 \mu A$. Dla źródła I (tranzystory T_8 , T_9) prądem wymuszającym jest I_4 , tj. suma prądów kolektorowych tranzystorów T_1 , T_2 . Z kolei prąd I_3 jest równy I_4 , a prąd baz tranzystorów T_3 , T_4

$$I_{2-3} = I_2 - I_3$$

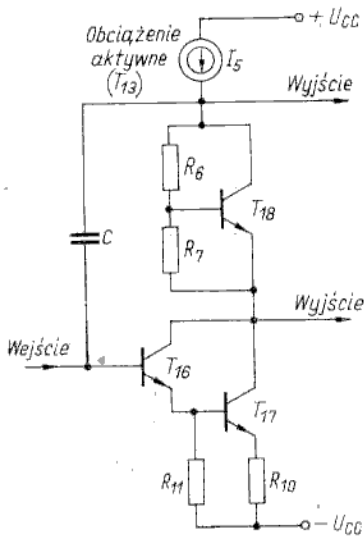
Ponieważ prąd $I_2 = \text{const}$, można zauważyć, że zwiększenie prądu I_3 powoduje zmniejszenie prądu I_{2-3} (prądu baz tranzystorów T_3 , T_4), co z kolei powoduje zmniejszenie sumy prądów emiterowych tranzystorów T_3 , T_4 , czyli zmniejszenie prądu I_4 i zwierciadlanego prądu I_3 . Obwód polaryzacji wzmacniacza różnicowego jest zatem rozwiązany w układzie z ujemnym sprzężeniem zwrotnym, co zapewnia dobrą stałość punktów pracy tranzystorów w stopniu wejściowym. Sieć równoważną układu polaryzacji przedstawiono na rys. 9.20b.

Teraz zostanie rozpatrzony sposób działania stopnia wejściowego (wzmacniacza różnicowego) — rys. 9.21. Pary tranzystorów T_1, T_3 oraz T_2, T_4 tworzą konfiguracje WC-WB (jak na rys. 9.12b). Tranzystory T_5, T_6, T_7 spełniają funkcję obciążenia aktywnego (jak na rys. 9.14 z uwzględnieniem modyfikacji źródła prądowego,



Rys. 9.21
Stopień wejściowy wzmacniacza μA 741

przedstawionej na rys. 9.4d). Konfiguracja WC-WB zapewnia uzyskanie dużej rezystancji wejściowej przy dobrych właściwościach częstotliwościowych (brak zjawiska Millera). Obciążenie aktywne umożliwia otrzymanie dużego wzmocnienia napięciowego. Tranzystory $p-n-p$ (T_3, T_4) są to tranzystory boczne (lateralne) o dużej wartości napięcia przebiecia złącza emiter-baza (ok. 80 V), dzięki temu



Rys. 9.22
Stopień niesymetryczny (separujący, wzmacniający i przesuwający poziom napięcia stałego)

wzmacniacz może pracować przy dużych napięciach na wejściu, w zakresie równym napięciu zasilania.

Stopień niesymetryczny, zbudowany przy użyciu tranzystorów T_{16} , T_{17} , T_{18} (rys. 9.22) spełnia trzy podstawowe funkcje:

- separuje stopień wejściowy od stopnia wyjściowego (układ Darlingtona — tranzystory T_{16} , T_{17});
- przesuwa poziom napięcia w sposób wymagany do wysterowania wyjściowej pary komplementarnej (układ T_{18} , R_6 , R_7 , podobny do układu rozważanego na rys. 9.6d);
- wzmacnia napięcie, gdyż tranzystor T_{13} (źródło I_5) spełnia funkcję obciążenia aktywnego w obwodzie kolektora tranzystora T_{17} .

Stopień wyjściowy tworzy para komplementarna tranzystorów T_{14} , T_{19} (jak na rys. 9.15). Pracują one w klasie AB ze spoczynkowym prądem polaryzacji ok. 200 μA .

Tranzystory T_{15} , T_{20} zabezpieczają wzmacniacz od przeciążeń, tj. ograniczają wartość prądu wyjściowego do ok. ± 30 mA. W przypadku pobierania z wyjścia zbyt dużego prądu dodatniego (płynącego przez tranzystor T_{14}) na rezystorze R_8 odkłada się spadek napięcia, wprowadzający tranzystor T_{15} w stan przewodzenia, co z kolei powoduje zmniejszenie wartości napięcia na bazie tranzystora T_{14} i ograniczenie prądu emitera tego tranzystora. Zbyt duży prąd ujemny (płynący przez tranzystor T_{19}) mógłby płynąć przy odpowiednio dużym prądzie bazy tranzystora T_{19} . Prąd bazy tranzystora T_{19} jest jednocześnie częścią prądu kolektora, a więc i prądu emitera tranzystora T_{17} . Przy zbyt dużej wartości prądu emitera tranzystora T_{17} na rezystorze R_{10} odkłada się spadek napięcia, wprowadzający tranzystor T_{20} w stan przewodzenia, co powoduje zmniejszenie napięcia na bazie tranzystora T_{16} , a więc zmniejszenie prądu emitera tranzystora T_{17} , czyli zmniejszenie prądów bazy i emitera tranzystora T_{19} .

Zastosowanie wzmacniaczy operacyjnych

9.2.4

Nazwa „wzmacniacz operacyjny” pojawiła się około 30 lat temu i dotyczyła klasy wzmacniaczy przeznaczonych do rozwiązywania analogowego równań różniczkowo-całkowych. Zatem wzmacniacze operacyjne sensu stricto są przeznaczone do zastosowań w maszynach analogowych, w których wykonują przede wszystkim następujące operacje:

- dodawanie (odejmowanie);
- mnożenie przez wartość stałą;
- całkowanie;
- różniczkowanie.

W odniesieniu do współcześnie wytwarzanych scalonych wzmacniaczy operacyjnych bardziej właściwą byłaby nazwa: „wzmacniacze uniwersalne”, gdyż zakres realizowanych przez nie funkcji układowych jest, praktycznie biorąc, nieograniczony. Oprócz wyżej wymienionych czterech funkcji podstawowych można zwrócić uwagę na szereg innych zastosowań, jak np.:

- układy mnożące (dzielące);
- wzmacniacze logarytmiczne;
- generatory sygnałów prostokątnych, trójkątnych, sinusoidalnych;
- filtry;
- detektory liniowe;
- detektory wartości szczytowej;

- układy typu „próbuj-trzymaj” (ang. *sample and hold*),
- itd.

Można zaryzykować twierdzenie, że dla konstruktora systemów analogowych w latach siedemdziesiątych wzmacniacz operacyjny jest tak samo podstawowym „elementem”, jakim był tranzystor dla konstruktora układów w latach sześćdziesiątych.

W tabelicy 9.4 podano uproszczone schematy i wzory dla niektórych podstawowych zastosowań wzmacniaczy operacyjnych.

Podstawowe zastosowania wzmacniacza operacyjnego

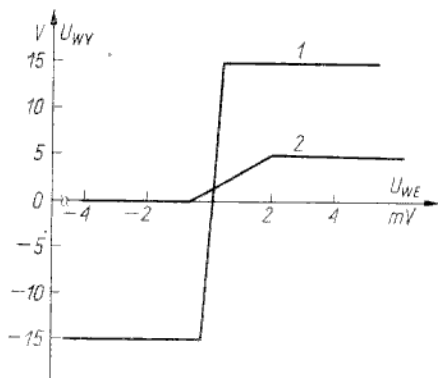
Tablica 9.4

Rodzaj układu	Schemat	Wzór
Wzmacniacz odwracający		$U_{WY} = -\frac{R_F}{R_1} U_{WE}$
Wzmacniacz nieodwracający		$U_{WY} = \frac{R_1 + R_F}{R_1} U_{WE}$ Dla $R_1 \rightarrow \infty$ otrzymuje się wtórnik napięciowy
Wzmacniacz sumujący		$U_{WY} = -R_F \left(\frac{U_{WE1}}{R_1} + \frac{U_{WE2}}{R_2} + \dots \right)$
Układ różniczkujący		$U_{WY} = -R_F C \frac{dU_{WE}}{dt}$
Układ całkujący		$U_{WY} = \frac{1}{R_1 C} \int U_{WE} dt$

Komparatory napięcia

9.3

Komparator napięcia jest układem wytwarzającym na wyjściu napięcia odpowiadające poziomom logicznym 0 lub 1 (na przykład poziomy napięcie 0 lub 5 V) w zależności od znaku różnicy napięć na wejściu. Jest to zatem układ porównujący (ang. *compare* oznacza porównywać) sygnał podany na jedno wejście z napięciem odniesienia na drugim wejściu, przy czym zmiany znaku różnicy tych dwu napięć są sygnalizowane zmianą stanu logicznego na wyjściu. Należy zauważyć, że funkcję komparatora może spełniać wzmacniacz operacyjny z otwartą pętlą sprzężenia zwrotnego, gdyż minimalna różnica napięć na jego wejściach, wzmocniona wiele tysięcy razy, powoduje ustalenie się na jego wyjściu poziomu napięcia zasilania. Na przykład dla w.o. ze wzmocnieniem 100 000 V/V i zasilaniem ± 15 V na wyjściu otrzymuje się $+15$ lub -15 V przy wejściowym napięciu różnicowym większym co do wartości bezwzględnej niż $-0,15$ lub $+0,15$ mV. Jednakże napięcia ± 15 V nie są zgodne z powszechnie stosowanymi poziomami sygnałów logicznych w technice cyfrowej (zero logiczne — 0 V, jedynka logiczna — 5 V). Komparator jest w pewnym sensie szczególnym rodzajem wzmacniacza operacyjnego, w którym minimalne i maksymalne wartości napięcia wyjściowego są dostosowane do współpracy z układami cyfrowymi, czyli wynoszą 0 V oraz 5 V. Jest to więc jednobitowy przetwornik analogowo-cyfrowy.



Rys. 9.23
Charakterystyki przejściowe wzmacniacza operacyjnego (1) i komparatora napięcia (2)

Na rysunku 9.23 porównano charakterystyki przejściowe wzmacniacza operacyjnego i komparatora napięcia. Komparatory są charakteryzowane niemal identycznym zestawem parametrów jak dla wzmacniaczy operacyjnych, przy czym najważniejszymi parametrami są:

- wejściowe napięcie i prąd niezrównoważenia, mające bezpośredni wpływ na zdolność rozdzielczą, czyli dokładność komparatora;
- czas odpowiedzi, określający szybkość działania komparatora.

Wejściowe napięcie niezrównoważenia jest definiowane nieco inaczej niż dla wzmacniacza operacyjnego, tj. jako wartość wejściowego napięcia różnicowego, przy której napięcie wyjściowe jest równe tzw. napięciu progu logicznego (zwykle ok. 1,5 V, co odpowiada stanowi pośredniemu między zerem a jedynką logiczną). W podobnych warunkach na wyjściu mierzy się wejściowy prąd niezrównoważenia. W typowych przypadkach wejściowe napięcie niezrównoważenia wynosi kilka miliwoltów, a wejściowy prąd niezrównoważenia — kilka mikroamperów. Czas odpowiedzi (odpowiednik szybkości narastania napięcia we wzmacniaczu

operacyjnym) jest definiowany jako czas upływający od chwili podania na wejście komparatora określonego skoku napięcia do chwili osiągnięcia na jego wyjściu poziomu progu logicznego. Ponieważ wartość czasu odpowiedzi silnie zależy od wielkości sygnałów różnicowych na wejściu przed i po podaniu napięcia, można porównywać czasy odpowiedzi różnych komparatorów tylko wówczas, gdy są one mierzone w jednakowych warunkach. Standardowe warunki, stosowane przez wiele firm, są następujące: skok różnicowego napięcia wejściowego od 95 do 5 mV ze zmianą znaku (całkowita amplituda impulsu wejściowego wynosi 100 mV). W typowym przypadku czas odpowiedzi wynosi kilkadziesiąt nanosekund.

Pierwszym wprowadzonym do produkcji i dotychczas najbardziej popularnym komparatorem scalonym jest układ μA 710 opracowany w 1966 r. w firmie Fairchild. Jego podstawowe parametry są następujące:

- napięcie niezrównoważenia — 2 mV (dryf temperaturowy tego napięcia — 3,5 mV/°C);
- prąd niezrównoważenia — 3 μA ,
- czas odpowiedzi — 40 ns.

Wadą komparatora μA 710 jest niesymetryczne zasilanie: +12 V, -6 V. Oprócz komparatorów uniwersalnych, których typowym przedstawicielem jest układ μA 710, produkuje się komparatory o bardzo dużej dokładności (małym napięciu i prądzie niezrównoważenia) lub bardzo dużej szybkości (małym czasie odpowiedzi). W pierwszej grupie można wymienić na przykład komparator CMP-02 (firma Precision Monolithics) o napięciu niezrównoważenia 0,8 mV i prądzie niezrównoważenia 3 nA. W drugiej grupie można wyróżnić komparator Am 685 (firma Advanced Micro Devices) o czasie odpowiedzi 7,5 ns.

Stabilizatory napięcia

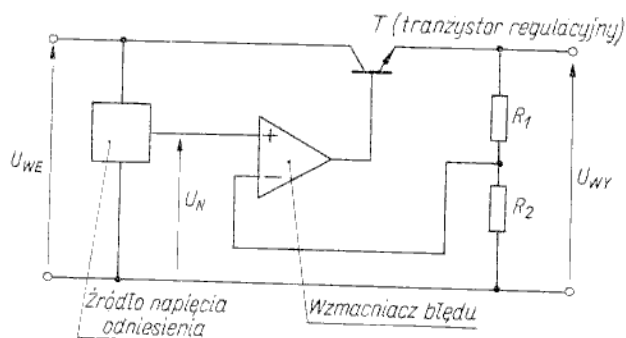
9.4

Stabilizatory napięcia, obok wzmacniaczy operacyjnych i komparatorów napięcia, stanowią trzecią podstawową grupę układów scalonych analogowych. Jak nazwa wskazuje, służą one do stabilizacji napięcia, czyli do utrzymywania na wyjściu stałego napięcia, niezależnie od zmian napięcia wejściowego (będącego jednocześnie napięciem zasilającym) i prądu obciążenia. Bardzo ważna jest również stałość napięcia wyjściowego w funkcji czasu i temperatury. Wszystkie scalone stabilizatory napięcia można podzielić ogólnie na dwie grupy:

- *stabilizatory nastawne* (nazywane również uniwersalnymi) o dowolnie regulowanej wartości napięcia wyjściowego (zwykle w zakresie 2...40 V);
- *stabilizatory nienastawne* (nazywane również lokalnymi) o jednej wartości napięcia wyjściowego, regulowanej tylko w niewielkim zakresie.

Przy dużej różnorodności rozwiązań konstrukcyjnych poszczególnych stabilizatorów zasadnicza idea działania układu stabilizującego jest jednakowa i można ją wyjaśnić na podstawie uproszczonego schematu elektrycznego, przedstawionego na rys. 9.24. Stabilizator składa się z następujących układów (elementów):

- źródła napięcia odniesienia (jest to zwykle skompensowana termicznie dioda Zenera ze wzmacniaczem);
- wzmacniacza błędów z wejściem różnicowym;
- szeregowo włączonego tranzystora regulacyjnego;
- wyjściowego dzielnika rezystancyjnego;
- układów zabezpieczających od przeciążeń (nie pokazanych na rys. 9.24).



Rys. 9.24
Uproszczony schemat elektryczny stabilizatora napięcia

Na wejście nieodwracające wzmacniacza błędu podaje się napięcie odniesienia U_N , do wejścia odwracającego jest doprowadzona część napięcia wyjściowego z dzielnika R_1, R_2 . Różnica tych dwu napięć po wzmocnieniu jest podawana na bazę tranzystora regulacyjnego i powoduje zmianę rezystancji kolektor-emiter, co powoduje zmianę napięcia wyjściowego w takim kierunku, by sygnał różnicowy na wejściu wzmacniacza błędu malał do zera.

Do podstawowych parametrów scalonego stabilizatora napięcia należy zaliczyć:

- zakres napięć wejściowego i wyjściowego;
- napięcie różnicowe wejście-wyjście; jest to najmniejsza wartość różnicy napięć na wejściu i wyjściu stabilizatora, przy której działa on jeszcze poprawnie;
- współczynnik stabilizacji napięcia; jest to stosunek względnej zmiany napięcia wyjściowego do zmiany napięcia wejściowego

$$\frac{\Delta U_{WY}}{U_{WY} \Delta U_{WE}} 100\% \text{ [%/V]}$$

- współczynnik stabilizacji obciążeniowej; jest to względna zmiana napięcia wyjściowego przy zmianie prądu obciążenia od wartości minimalnej do maksymalnej

$$\frac{U_{WY \max} - U_{WY \min}}{U_{WY \max}} 100\%$$

- współczynnik tłumienia tętnień; jest to iloraz międzyszczytowych wartości napięcia tętnień (składowej zmiennej) na wyjściu i wejściu stabilizatora;
- maksymalny prąd wyjściowy.

W grupie scalonych stabilizatorów nastawnych do najbardziej popularnych należy zaliczyć układy LM 105 dla napięć dodatnich i LM 104 dla napięć ujemnych. Do zastosowań o dużej precyzji najczęściej stosuje się układ μA 723. W grupie scalonych stabilizatorów nienastawnych najbardziej rozpowszechniony jest układ LM 109 o nieregulowanym napięciu wyjściowym $+5 \text{ V}$, przystosowanym do współpracy z układami cyfrowymi. Podstawowe parametry wyżej wymienionych układów zamieszczono w tabl. 9.5.

Pewnym ograniczeniem stabilizatorów nastawnych jest niewielka wartość prądu obciążenia, która w typowym przypadku wynosi kilkadziesiąt miliamperów (tylko firma Motorola produkuje układy o prądzie wyjściowym do $0,5 \text{ A}$ — na przykład MC 1560). Jednakże obciążalność prądową można znacznie zwiększyć przez dołączenie tranzystorów zewnętrznych.

Podstawowe parametry wybranych typów scalonych stabilizatorów napięcia

Tablica 9.5

Typ układu		Nastawne			Nienastawny
		LM 105	LM 104	μ A 723	LM 109
Parametry					
Zakres napięcia wejściowego	V	+8,5...+50	-50...-8	+9,5...+40	+7...+35
Zakres napięcia wyjściowego	V	+4,5...+40	-40...0	+2...+37	+5
Napięcie różnicowe	V	3	2	3	2
Współczynnik stabilizacji napięcia	%/V	0,06	0,1	0,02	0,005
Współczynnik stabilizacji obciążeniowej	%	0,02	0,01	0,03	0,6
Współczynnik tłumienia tętnień	%	0,01	0,01	0,02	0,005
Maksymalny prąd wyjściowy	mA	20	20	65	1500

Układy cyfrowe, pracujące przy sygnałach dwójkowych (binarnych) o dwu dyskretnych poziomach napięcia lub prądu (względnie wysokim i względnie niskim), są szczególnie predysponowane do realizacji w postaci scalonych układów półprzewodnikowych. Przemawiają za tym głównie dwa czynniki:

- wymagania odnośnie do zakresów wartości nominalnych i tolerancji elementów są znacznie łagodniejsze niż dla układów analogowych;
- systemy cyfrowe zawierają dużo powtarzalnych podzespołów.

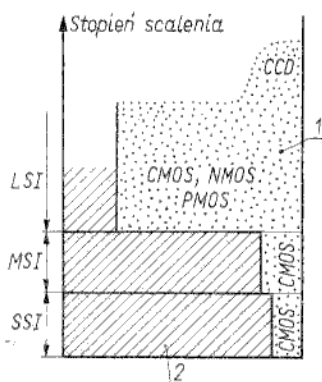
Układy scalone cyfrowe (u.s.c.) można różnie klasyfikować w zależności od przyjętych kryteriów. Pod względem konstrukcyjno-technologicznym dzieli się wszystkie u.s.c. na dwie ogólne klasy:

- bipolarne (zbudowane z tranzystorów bipolarnych),
- unipolarne, nazywane również układami MIS lub MOS (zbudowane z tranzystorów MIS).

Te dwie klasy u.s.c. stanowią tylko częściowo konkurencyjne, a w znacznej mierze dopełniające się rozwiązania. Można je ogólnie porównać, przyjmując za podstawowe kryteria trzy tendencje determinujące kierunki rozwoju mikroelektroniki, tj. dążenie do uzyskania:

- jak największej liczby podzespołów funkcjonalnych w jednostkowej powierzchni płytki (tzw. gęstości funkcji);
- jak najmniejszej energii wymaganej do realizacji funkcji elementarnej, tj. jak najmniejszej wartości iloczynu mocy traconej przez czas opóźnienia;
- jak najmniejszych kosztów wytwarzania.

We wszystkich trzech przypadkach porównanie standardowych układów bipolarnych i MIS wypada na korzyść tych ostatnich. Jednak często w iloczynie mocy przez czas opóźnienia do jednego z tych czynników przywiązuje się większą wagę.



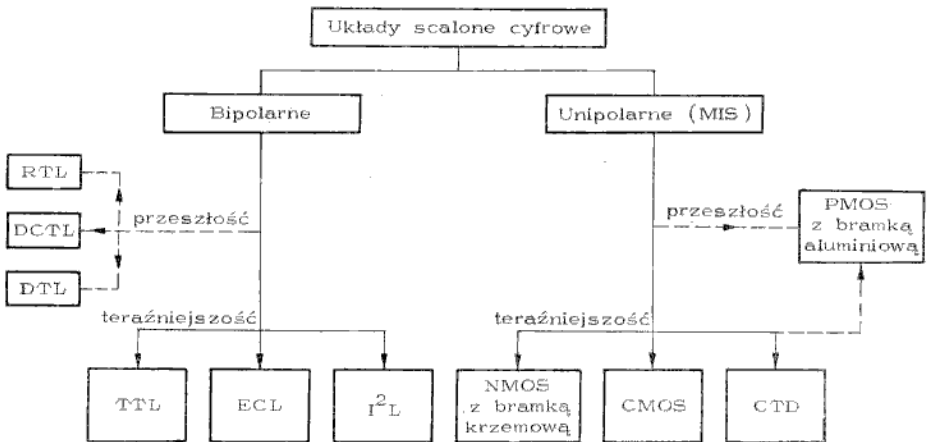
Rys. 10.1

Poglądowa ilustracja udziału układów scalonych cyfrowych bipolarnych i MIS w klasach o małym, średnim i dużym stopniu scalenia
1 — MIS; 2 — bipolarne

Na przykład w kalkulatorach, zegarkach elektronicznych itp. przenośnym sprzęcie powszechnego użytku istotny jest bardzo mały pobór mocy, a duża szybkość działania (mały czas opóźnienia) nie jest konieczna. W tym przypadku lepsze są układy MIS. Z kolei w wielu przypadkach, na przykład w dużych komputerach, najważniejsza jest duża szybkość działania układów cyfrowych i wówczas lepsze są układy bipolarne.

Układy MIS, z uwagi na większą gęstość upakowania w nich uzyskiwaną, są szczególnie predestynowane do scalania wielkostopniowego. Na rysunku 10.1 przedstawiono w sposób poglądowy (bez zachowania ścisłych proporcji) podział układów o małym, średnim i dużym stopniu scalenia na dwie grupy, tj. bipolarne i MIS. Z tego rysunku widać, że domeną układów bipolarnych jest mały i średni stopień scalenia, natomiast mniejszy jest ich udział w klasie układów LSI. Na odwrót, układy MIS są wytwarzane głównie jako układy o dużym stopniu scalenia, przy niewielkim udziale wyspecjalizowanych układów CMOS o średnim i małym stopniu scalenia.

Kontynuując klasyfikację u.s.c. na podstawie kryteriów technologiczno-konstrukcyjnych należy wymienić rodziny układów bipolarnych i unipolarnych (rys. 10.2). Wyjaśnijmy nazwy wyszczególnione na rys. 10.2:



Rys. 10.2
Klasyfikacja układów scalonych cyfrowych

- RTL — (ang. *Resistor-Transistor Logic*, tj. układy rezystorowo-tranzystorowe) — rodzina u.s.c. produkowanych w pierwszej połowie lat sześćdziesiątych, obecnie przestarzałych.
- DCTL — (ang. *Direct-Coupled Transistor Logic*, tj. układy z tranzystorami sprzężonymi bezpośrednio) — jak wyżej.
- DTL — (ang. *Diode-Transistor Logic*, tj. układy diodowo-tranzystorowe) — rodzina u.s.c. bardzo popularnych w latach sześćdziesiątych, obecnie jeszcze produkowanych w niewielkich ilościach (na przykład w Europie Zachodniej w okresie od 1970 do 1975 produkcja tych układów zmalała ok. 5-krotnie).
- TTL — (ang. *Transistor-Transistor Logic*, tj. układy tranzystorowo-tranzystorowe) — rodzina u.s.c. wprowadzonych do produkcji w drugiej połowie lat sześćdziesiątych i do chwili obecnej najbardziej popularnych w zakresie małego i średniego stopnia scalenia.

- ECL — (ang. *Emitter-Coupled Logic*, tj. układy z tranzystorami o sprzężonych emiterach) — rodzina u.s.c. produkowanych już w pierwszej połowie lat sześćdziesiątych i do chwili obecnej niezastąpionych w klasie układów bardzo szybkich o małym i średnim stopniu scalenia.
- I²L — (ang. *Integrated Injection Logic*, tj. układy iniekcyjne) — rodzina u.s.c. opracowanych w 1972 r., o trudnych do określenia perspektywach rozwoju¹⁾.
- PMOS — rodzina u.s.c. zbudowanych z tranzystorów z kanałem typu *p*, bardzo popularna na początku lat siedemdziesiątych, obecnie stosowana tylko w niektórych urządzeniach powszechnego użytku (kalkulatory) ze względu na niskie koszty wytwarzania.
- NMOS — rodzina u.s.c. zbudowanych z tranzystorów z kanałem typu *n* i z bramką krzemową, dominującą w technologii MIS od 1974 r.
- CMOS — rodzina u.s.c. zbudowanych z tranzystorów komplementarnych MOS, produkowana od 1967 r. i mająca w dalszym ciągu bardzo duże znaczenie ze względu na unikalne właściwości elektryczne (wyjątkowo mały pobór mocy, duża odporność na zakłócenia itp.).
- CTD — (ang. *Charge-Transfer-Devices*), przyrządy o transporcie ładunku, których najważniejszą grupę stanowią CCD (ang. *Charge-Coupled Devices*, tj. przyrządy o sprzężeniu ładunkowym), odkryte w 1970 r. Jest to bodajże najbardziej perspektywiczny kierunek rozwoju mikroelektroniki lat siedemdziesiątych; układy CCD są stosowane zarówno w systemach cyfrowych, jak i analogowych (przede wszystkim jako przetworniki obrazu).

Dalej będą opisywane tylko te rodziny u.s.c., które mają obecnie największe znaczenie, tj.:

— w grupie układów bipolarnych: TTL, ECL, I²L,

— w grupie układów MIS: NMOS, CMOS.

Układy CCD ze względu na ich specyficzny obszar zastosowań będą omówione odrębnie.

W tym niewielkim rozdziale nie można oczywiście omówić wszystkich zagadnień dotyczących u.s.c. (poszczególnym rodzinom tych układów poświęca się oddzielne monografie), dlatego trzeba wprowadzić dalsze ograniczenia zakresu rozważań. Aby dokładnie określić zakres interesujących nas zagadnień, sklasyfikuje się

Podział u.s.c. ze względu na rodzaj realizowanych funkcji

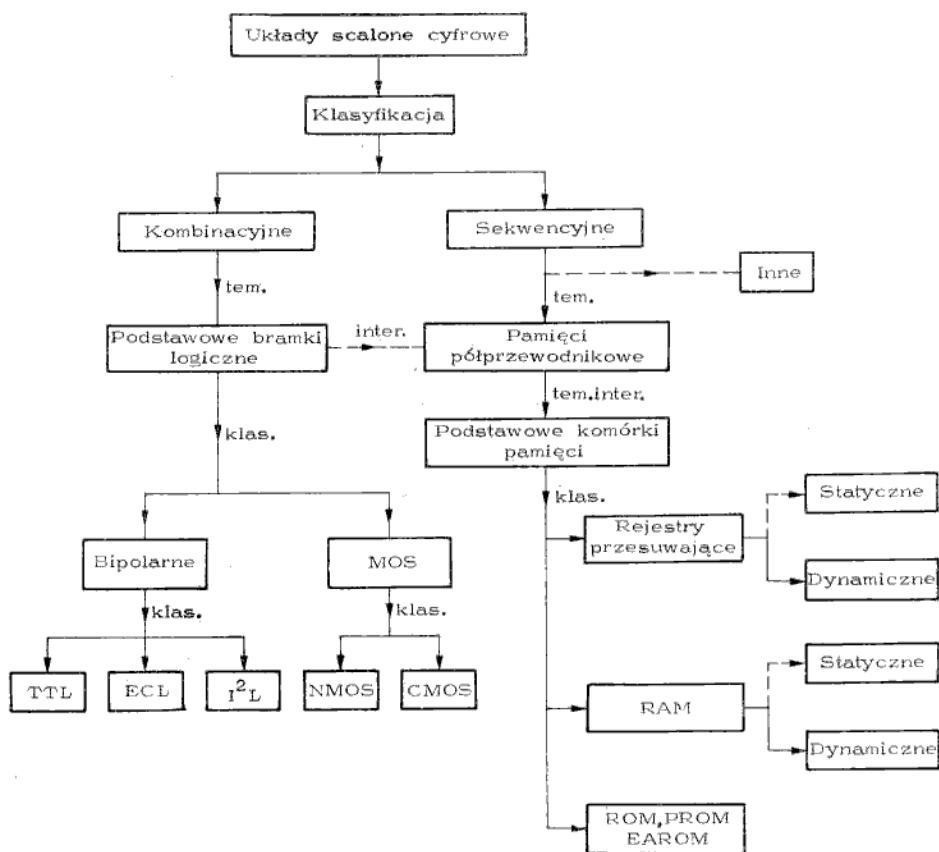
Tablica 10.1

Stopień scalenia \ Rodzaj funkcji	Kombinacyjna	Sekwencyjna
SSI	Podstawowe bramki: NOT, AND, OR, NAND, NOR, EXCLUSIVE OR	Przerzutniki RS, JK, D, T i inne
MSI	Dekodery, selektory (multipleksery) i inne	Rejestry, liczniki, sumatory
LSI	Ekspandery, procesory i inne	Pamięci ROM, RAM

¹⁾ We wcześniejszych porównaniach układów bipolarnych z unipolarnymi nie braliśmy pod uwagę układów I²L. Układy I²L łącząc zalety układów MIS (duża gęstość upakowania i mały pobór mocy) z zaletami układów TTL (duża szybkość działania) w najbliższych latach mogą stać się dominującą rodziną u.s.c. bipolarnych, chociaż entuzjastyczne prognozy rozwoju tych układów (z lat 1975—1977) do roku 1979 nie potwierdziły się.

u.s.c. ze względu na rodzaj realizowanych funkcji logicznych. W tabelicy 10.1 przedstawiono taką klasyfikację, przy czym podano również przykłady konkretnych układów w poszczególnych stopniach scalenia. Zgodnie z profilem tej książki rozważania ograniczymy tylko do „bazy elementowej” u.s.c., tj. do kilku podstawowych bramek logicznych, z których można budować dowolnie złożone układy kombinacyjne i sekwencyjne oraz do kilku podstawowych komórek pamięci RAM i ROM.

Układ rozdziału przedstawiono schematycznie na rys. 10.3.



Rys. 10.3
Schematyczny układ rozdziału

Podstawowe funkcje logiczne

10.1

Sygnały w układach cyfrowych przyjmują dwie wartości dyskretne, z których jedna jest identyfikowana jako stan logiczny „1”, a druga jako stan logiczny „0”. Przykładowo w układzie zasilanym napięciem +5 V mogą wystąpić dwa dyskretne poziomy napięcie +0,4 oraz +3,5 V. Sposób przyporządkowania stanów logicznych tym wartościom napięć jest całkowicie arbitralny. Jeżeli względnie niższe napięcie dodatnie (+0,4 V) nazwie się „0” logicznym, a względnie wyższe

napięcie dodatnie (+3,5 V) — „1” logiczną, to taką konwencję nazywa się *logiką dodatnią*. Konwencja odwrotna, w której przyporządkowuje się względnie niższemu napięciu dodatniemu (+0,4 V) stan logiczny „1”, a względnie wyższemu (+3,5 V) stan logiczny „0”, jest nazywana *logiką ujemną*.

Projektowanie systemu cyfrowego najogólniej składa się z dwu etapów:

- opracowania projektu funkcjonalnego, tj. tzw. struktury logicznej systemu;
- opracowania projektu układowego, tj. przyporządkowania poszczególnym funkcjom logicznym odpowiednich rozwiązań układowych.

Wyboru konwencji odwzorowania napięciowego stanów logicznych (logika dodatnia lub ujemna) dokonuje się w drugim etapie. Rozważania przedstawione w tym punkcie są słuszne dla obu konwencji, gdyż opisuje się podstawowe funkcje logiczne, abstrahując chwilowo od ich realizacji układowej.

George Boole (1815—1864) opracował specjalny dział algebry dla systemu dwójkowego, który stanowi podstawę projektowania funkcjonalnego układów logicznych.

Warto w tym miejscu przytoczyć niektóre podstawowe zasady *algebry Boole’a*.

Dwie wielkości, które mogą przyjmować wartości dwójkowe 0 lub 1, oznaczymy literami A , B . Z definicji:

$$A = 1 \quad \text{jeżeli} \quad A \neq 0$$

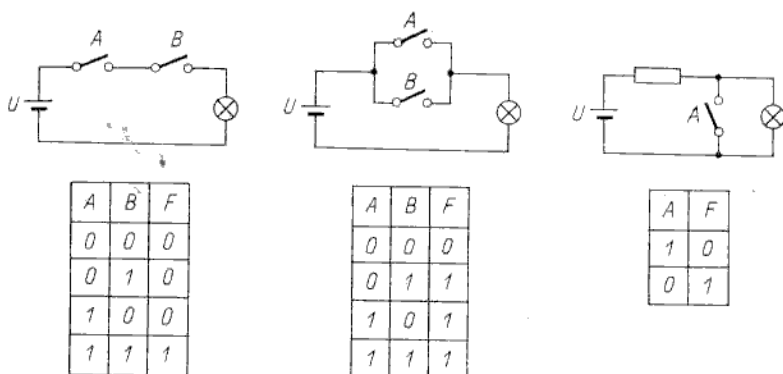
$$A = 0 \quad \text{jeżeli} \quad A \neq 1$$

oraz identycznie dla wielkości B .

W algebrze Boole’a stosuje się trzy operacje elementarne:

- 1 — A lub B , czyli suma logiczna, zapisywana jako $A + B$;
- 2 — A i B , czyli iloczyn logiczny, zapisywany jako $A \cdot B$;
- 3 — nie A , czyli negacja logiczna, zapisywana jako \bar{A} .

Te trzy podstawowe operacje (funkcje) logiczne są powszechnie oznaczane symbolami angielskimi: OR (LUB), AND (I), NOT (NIE).



Rys. 10.4

Przykłady najprostszych obwodów realizujących podstawowe funkcje logiczne AND (a), OR (b) i NOT (c) wraz z tablicami stanów

Najprostsze obwody, realizujące te trzy funkcje logiczne, przedstawiono na rys. 10.4.

Wszystkie możliwe relacje między zmiennymi niezależnymi a zmienną zależną F są podane w tzw. *tablicach stanów*. Zwarty klucz i świecąca się lampka re-

prezentują stany „1”. W pierwszym obwodzie (rys. 10.4a) tylko przy zwarceniu obu kluczy ($A = 1, B = 1$) świeci się lampka (stan 1 na wyjściu), dla pozostałych kombinacji na wyjściu jest stan 0. Zatem jest to obwód iloczynu logicznego AND, dla którego można napisać

$$A \cdot B = F \quad (10.1)$$

W drugim obwodzie (rys. 10.4b) tylko przy rozwarciu obu kluczy ($A = 0, B = 0$) lampka się nie świeci ($F = 0$), dla pozostałych kombinacji włączenia kluczy lampka świeci się, czyli $F = 1$. Zatem jest to układ sumy logicznej OR, dla którego można napisać

$$A + B = F \quad (10.2)$$

W trzecim przypadku (rys. 10.4c) lampka świeci się ($F = 1$) przy kluczu rozwartym ($A = 0$). Natomiast przy kluczu zwartym ($A = 1$) lampka nie świeci się ($F = 0$). Jest to zatem obwód negacji logicznej, dla którego

$$\bar{A} = F$$

Na podstawie trzech operacji elementarnych można wyprowadzić szereg twierdzeń algebry Boole'a, których znajomość jest niezbędna przy projektowaniu struktury logicznej systemu cyfrowego. Najważniejsze twierdzenia wymienimy bez dowodów (o prawdziwości tych twierdzeń można łatwo się przekonać, budując obwody elektryczne z kluczami, analogiczne do przedstawionych na rys. 10.4):

$$1 + A = 1 \quad 0 \cdot A = 0 \quad (10.3)$$

$$0 + A = A \quad 1 \cdot A = A \quad (10.4)$$

$$A + A = A \quad A \cdot A = A \quad (10.5)$$

$$A + \bar{A} = 1 \quad \bar{A} \cdot A = 0 \quad (10.6)$$

$$\bar{\bar{A}} = A \text{ (podwójna negacja)} \quad (10.7)$$

$$A + B = B + A \quad A \cdot B = B \cdot A \quad (10.8)$$

$$A + (B + C) = (A + B) + C \quad A \cdot (BC) = (AB) \cdot C \quad (10.9)$$

$$A(B + C) = AB + AC \quad (A + B)(A + C) = A + BC \quad (10.10)$$

Prawa de Morgana

$$\overline{(A + B)} = \bar{A} \cdot \bar{B} \quad \overline{(AB)} = \bar{A} + \bar{B} \quad (10.11)$$

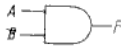






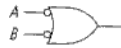

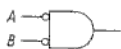
Ostatnie przekształcenia (prawa de Morgana) są szczególnie przydatne, gdyż umożliwiają traktowanie funkcji (układów) OR oraz AND jako wzajemnie zamienialnych.

Ponieważ „naturalną” cechą tranzystora jest odwracanie fazy napięcia na kolektorze w stosunku do napięcia na bazie, w układach tranzystorowych łatwiej można zrealizować funkcje AND, OR łącznie z funkcją NOT niż AND, OR bez negacji. Funkcja powstała wskutek połączenia operacji NOT, AND (negacja iloczynu logicznego) nosi nazwę NAND (NIE — I), a funkcja powstała wskutek połączenia operacji NOT, OR (negacja sumy logicznej) nosi nazwę NOR (NIE — LUB). W tabelicy 10.2 przedstawiono symbole elementarnych funkcji logicznych, kreślone zgodnie z powszechnie przyjętymi normami amerykańskimi.

Podstawowymi układami logicznymi (niejako elementami, z których buduje się bardziej złożone układy i systemy), są układy realizujące funkcje NAND, NOR. Takie podstawowe układy są nazywane *bramkami* lub *funktorami logicznymi*.

Symbole elementarnych funkcji logicznych

Tablica 10.2

Logika dodatnia			Logika ujemna			Tablica stanów															
AND	$F = A \cdot B$		OR	$F = A + B$		<table border="1"> <tr><td>A</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>B</td><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>F</td><td>0</td><td>0</td><td>0</td><td>1</td></tr> </table>	A	0	1	0	1	B	0	0	1	1	F	0	0	0	1
A	0	1	0	1																	
B	0	0	1	1																	
F	0	0	0	1																	
OR	$F = A + B$		AND	$F = A \cdot B$		<table border="1"> <tr><td>A</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>B</td><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>F</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> </table>	A	0	1	0	1	B	0	0	1	1	F	0	1	1	1
A	0	1	0	1																	
B	0	0	1	1																	
F	0	1	1	1																	
NOT	$F = \bar{A}$		NOT	$F = \bar{A}$		$F = \bar{A}$															
NAND	$F = \overline{A \cdot B}$		NOR	$F = \overline{A + B}$		<table border="1"> <tr><td>A</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>B</td><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>F</td><td>1</td><td>1</td><td>1</td><td>0</td></tr> </table>	A	0	1	0	1	B	0	0	1	1	F	1	1	1	0
A	0	1	0	1																	
B	0	0	1	1																	
F	1	1	1	0																	
NOR	$F = \overline{A + B}$		NAND	$F = \overline{A \cdot B}$		<table border="1"> <tr><td>A</td><td>0</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>B</td><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>F</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> </table>	A	0	1	0	1	B	0	0	1	1	F	1	0	0	0
A	0	1	0	1																	
B	0	0	1	1																	
F	1	0	0	0																	

Zostaną teraz rozpatrzone sposoby realizacji bramek logicznych w poszczególnych rodzinach u.s.c. Ponieważ funkcje NAND i NOR są wzajemnie zamienialne (każdy system wykonany z bramek NAND można wykonać również z bramek NOR i na odwrót), dla każdej rodziny u.s.c. wystarczy rozpatrzeć tylko jeden z tych funkcyjów (NAND lub NOR).

Bramki logiczne bipolarne

10.2

Rozpatrzmy podstawowe bramki logiczne, realizowane w trzech rodzinach układów bipolarnych:

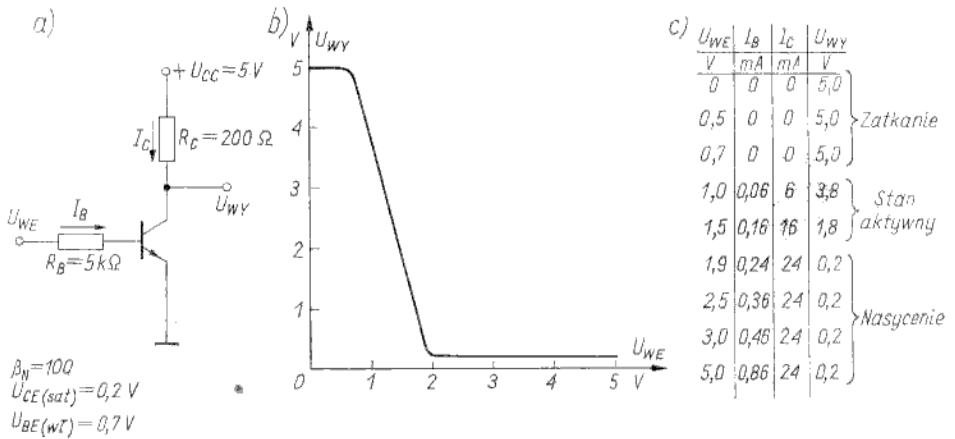
- TTL (z różnymi odmianami),
- ECL,
- I²L.

Ponieważ właściwości układów cyfrowych w istotny sposób zależą od tego, czy tranzystory pracują w zakresie nasycenia, wszystkie u.s.c. bipolarne można podzielić na pracujące z tranzystorami nasycenymi lub nienasycenymi.

Bramki logiczne z tranzystorami nasycenymi

10.2.1

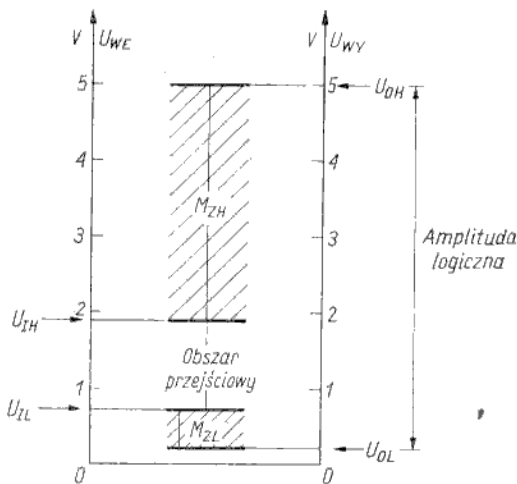
Podstawowym fragmentem bramki logicznej z tranzystorami pracującymi w zakresie nasycenia jest *inverter*, którego typowy przykład wraz z charakterystyką przejściową przedstawiono na rys. 10.5. W warunkach statycznych tran-



Rys. 10.5

Inwerter z tranzystorem bipolarnym pracującym w zakresie nasylenia: a) schemat elektryczny; b) charakterystyka przenoszenia $U_{WY} = f(U_{WE})$; c) tablica wartości napięć i prądów

zystor znajduje się w jednym z dwu stanów, tj. w stanie zatkania lub nasylenia. W czasie przełączania z jednego stanu do drugiego tranzystor pracuje w obszarze aktywnym. Przy założonych parametrach układu tranzystor jest zatkany, jeżeli napięcie wejściowe jest mniejsze niż 0,7 V, a „wchodzi” w zakres nasylenia przy napięciu wejściowym większym niż 1,9 V. Dla napięć wejściowych (rys. 10.6)



Rys. 10.6

Interpretacja marginesów zakłócenowych i amplitudy logicznej inwertera

w przedziale 0,7...1,9 V tranzystor pracuje w obszarze aktywnym. Napięcie 0,7 V jest maksymalnym napięciem wejściowym, interpretowanym jako „0” logiczne (oznaczane przez U_{IL} , indeks I — ang. *input*, L — ang. *low*). Napięcie 1,9 V jest minimalnym napięciem wejściowym, interpretowanym jako „1” logiczna (oznaczane przez U_{IH} , indeks I — ang. *input*, H — ang. *high*). Na wyjściu inwertera napięcia w stanach logicznych „0” i „1”, oznaczane jako U_{OL} , U_{OH}

O — ang. *output*), wynoszą odpowiednio 0,2 i 5 V. Różnica napięć ($U_{OH} - U_{OL}$) jest nazywana *amplitudą logiczną*. Przy pracy kaskadowej inwerterów, tj. przy sterowaniu następnego inwertera z wyjścia identycznego inwertera poprzedniego, dla pewnej pracy układu jest wymagane, żeby $U_{OL} < U_{IL}$ oraz $U_{OH} > U_{IH}$. Dla rozpatrywanego przykładu $U_{OL} = 0,2$ V, a $U_{IL} = 0,7$ V, zatem „zapas” napięcia, przy którym następny inwerter jest sterowany „0” logicznym, wynosi 0,5 V. Ten „zapas” napięcia jest nazywany *marginielem zakłóceńiowym dla stanu niskiego*

$$M_{ZL} = U_{IL} - U_{OL} = 0,7 \text{ V} - 0,2 \text{ V} = 0,5 \text{ V} \quad (10.12)$$

Nazwa margines zakłóceńiowy pochodzi stąd, że przy zakłóceńiach mniejszych niż ($U_{IL} - U_{OL}$) jest gwarantowany stan logiczny „0” na wejściu inwertera.

Analogicznie różnica napięć ($U_{OH} - U_{IH}$) jest nazywana *marginielem zakłóceńiowym dla stanu wysokiego*

$$M_{ZH} = U_{OH} - U_{IH} = 5 \text{ V} - 1,9 \text{ V} = 3,1 \text{ V} \quad (10.13)$$

Przy zakłóceńiach mniejszych niż ($U_{OH} - U_{IH}$) jest zatem gwarantowany stan logiczny „1” na wejściu inwertera.

Do układów cyfrowych, pracujących z tranzystorami nasyconymi, zalicza się następujące rodziny bramek logicznych:

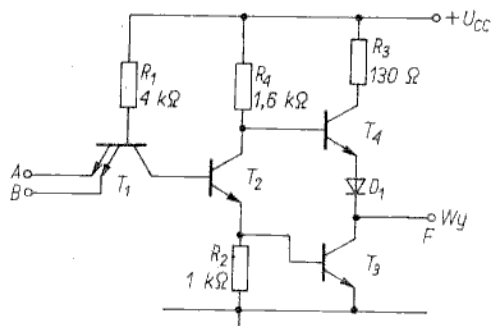
- RTL,
- DCTL,
- DTL,
- TTL,
- I²L.

Zgodnie z wcześniejszymi ustaleniami rozważania ograniczymy tylko do układów TTL oraz I²L.

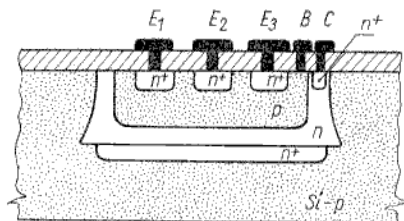
Bramki TTL

10.2.1.1

W rodzinie układów TTL podstawową bramką jest *funktor NAND*, którego schemat ideowy dla przypadku dwu wejść przedstawiono na rys. 10.7. W układzie tym jest stosowany tranzystor wieloemiterowy, wynaleziony w 1961 r. przez P. M. Thompsona (rys. 10.8). Zasadę działania bramki NAND wyjaśnimy naj-

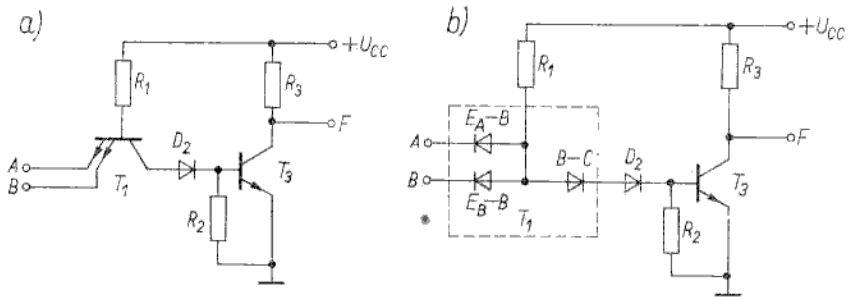


Rys. 10.7
Schemat elektryczny podstawowej bramki NAND zrealizowanej w technice TTL



Rys. 10.8
Struktura tranzystora wieloemiterowego

pierw na podstawie uproszczonego układu (rys. 10.9), który uzyskuje się z układu przedstawionego na rys. 10.7, zamieniając złącze $B-E$ tranzystora T_2 na diodę D_2 oraz zastępując dwutranzystorowy stopień wyjściowy (T_3, T_4) inwerterem jednotranzystorowym (T_3).



Rys. 10.9
Uproszczone schematy bramki NAND/TTL

Układ może pracować w jednym z dwu następujących stanów:

1. Na jedno lub oba wejścia A, B są podane niskie poziomy napięcia (na przykład $0,2\text{ V}$), reprezentujące „0” logiczne. Wówczas ze źródła $+U_{CC}$ (zwykle 5 V) płynie prąd przez rezystor R_1 i złącze (lub oba złącza) baza-emiter tranzystora T_1 do masy. Ponieważ spadek napięcia na złączu krzemowym, pracującym w kierunku przewodzenia, wynosi $0,7\text{ V}$, potencjał bazy tranzystora T_1 jest równy $U_{IL} + 0,7\text{ V}$, tj. $0,9\text{ V}$. Tranzystor T_1 pracuje w stanie nasycenia przy pomijalnie małej wartości prądu kolektora, gdyż dioda D_2 jest włączona w kierunku zaporowym. Zatem tranzystor T_3 jest zatkany, czyli $F = 1$.
2. Na oba wejścia są podane wysokie poziomy napięcia (na przykład 5 V), reprezentujące „1” logiczne. Wówczas złącza emiter-baza tranzystora T_1 są spolaryzowane w kierunku zaporowym. Prąd ze źródła $+U_{CC}$ płynie przez rezystor R_1 , złącze baza-kolektor tranzystora T_1 spolaryzowane w kierunku przewodzenia i diodę D_2 do bazy tranzystora T_3 . Wartość tego prądu jest wystarczająca do wysterowania tranzystora T_3 w zakresie nasycenia, czemu odpowiada $F = 0$.

W pierwszym przybliżeniu można przyjąć, że tranzystor T_1 spełnia funkcję trzech diod połączonych w układzie przedstawionym na rys. 10.9b. Zauważmy, że diody $B-C, D_2$ i złącze $B-E$ tranzystora T_3 są włączone szeregowo. Ponieważ przy pracy w kierunku przewodzenia napięcie na każdym złączu wynosi w przybliżeniu $0,7\text{ V}$, minimalny potencjał na bazie tranzystora T_1 , przy którym popłynie prąd bazy tranzystora T_3 , wynosi $2,1\text{ V}$. Ten warunek będzie spełniony wówczas, gdy na obu wejściach A, B będą napięcia U_A, U_B większe niż $1,4\text{ V}$ (uwzględnia się spadek napięcia $0,7\text{ V}$ na złączach E_A-B oraz E_B-B). Jeżeli na jednym z wejść jest mniejsze napięcie, to potencjał bazy tranzystora T_1 , równy sumie $U_A + 0,7\text{ V}$ lub $U_B + 0,7\text{ V}$ jest mniejszy niż $2,1\text{ V}$, czyli nie wystarcza do spolaryzowania złączy $B-C, D_2$ i $B-E$ tranzystora T_3 w kierunku przewodzenia.

Układ przedstawiony na rys. 10.9b jest w istocie bramką diodowo-tranzystorową (DTL), która nie jest w pełni równoważna układowi z rys. 10.9a, gdyż połączenie trzech diod ($E_A-B, E_B-B, B-C$) nie jest ściśle równoważne tranzystorowi T_1 . W szczególności należy zauważyć, że przy podaniu wysokich poziomów napięć na oba wejścia A, B , złącza emiter-baza tranzystora T_1 są spolaryzowane w kierunku

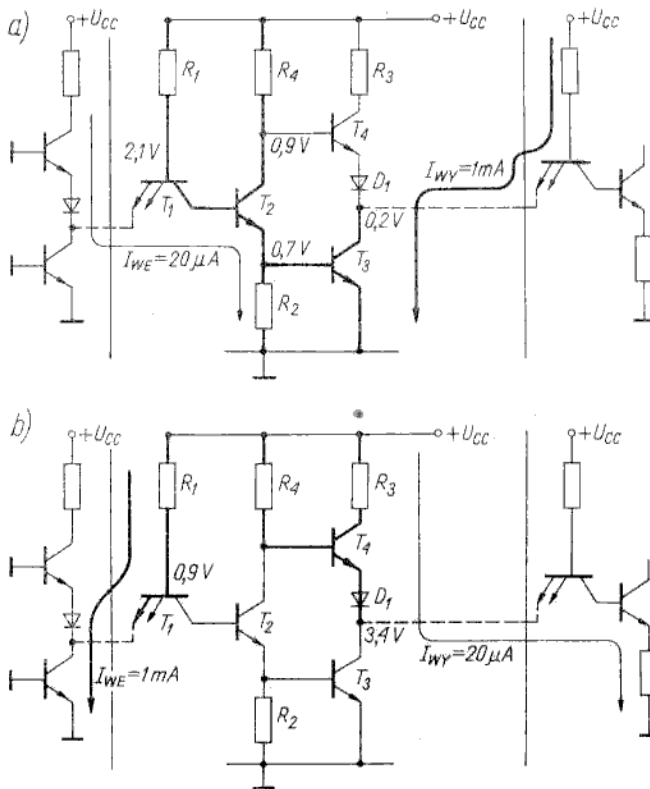
zaporowym, a złącze baza-kolektor tego tranzystora pracuje w kierunku przewodzenia. Taka polaryzacja tranzystora odpowiada pracy aktywnej inwersyjnej, co oznacza, że w obwodzie wejściowym może płynąć prąd emitera β_I razy większy niż prąd bazy płynący przez rezystor R_1 . Prąd wejściowy wywołuje niekorzystne skutki, gdyż powoduje spadek napięcia na rezystancji wewnętrznej źródła sygnału sterującego (sygnał sterujący jest zwykle pobierany z wyjścia F takiej samej bramki), czyli obniżenie poziomu napięcia wejściowego w stanie „1”. Aby osłabić ten niekorzystny efekt, tranzystor T_1 projektuje się tak, żeby współczynnik β_I był jak najmniejszy (ok. 1) oraz dąży się do uzyskania jak najmniejszej rezystancji wyjściowej bramki sterującej (identycznej ze sterowaną).

Omówimy teraz praktyczny układ bramki NAND (rys. 10.7), w którym stopień wyjściowy, zbudowany z dwu tranzystorów (T_3, T_4) połączonych w konfiguracji „totem pole”, ma bardzo małą rezystancję wyjściową. W stanie „0” logicznego na wyjściu tranzystor T_3 jest nasycony, a tranzystor T_4 — zatkany i rezystancja wyjściowa równa rezystancji kolektor-emiter tranzystora T_3 wynosi 10 Ω . W stanie „1” logicznej na wyjściu tranzystor T_3 jest zatkany, a tranzystor T_4 pracuje w układzie wtórnika emiterowego o rezystancji wyjściowej ok. 100 Ω . Tranzystor T_2 wraz z rezystorami R_2, R_4 pracuje w konfiguracji wzmacniacza symetryzującego, sterującego odpowiednie poziomy napięcie baz tranzystorów T_3, T_4 . Sygnały na kolektorze i emiterze tranzystora T_2 są przesunięte w fazie o 180°. Jeżeli tranzystor T_2 jest zatkany, to na jego emiterze jest niski potencjał niewystarczający do wysterowania tranzystora T_3 , a na kolektorze jest potencjał wysoki polaryzujący tranzystor T_4 w stanie przewodzenia. W sytuacji odwrotnej, tj. w stanie nasycenia tranzystora T_2 , na jego emiterze jest względnie duży potencjał (ok. 0,7 V z powodu bocznikowania rezystora R_2 przez złącze $B-E$ tranzystora T_3), powodujący wysterowanie tranzystora T_3 w stanie nasycenia. Na kolektorze tranzystora T_2 jest wówczas potencjał mały (ok. 0,9 V), niewystarczający do wysterowania tranzystora T_4 , który pracuje w stanie zatkania. Gdyby nie było diody D_1 , wówczas tranzystor T_4 pracowałby w stanie przewodzenia, gdyż różnica napięć na kolektorach tranzystorów T_2, T_3 , pracujących w stanie nasycenia, wynosi $0,9 \text{ V} - 0,2 \text{ V} = 0,7 \text{ V}$, czyli jest wystarczająca do spolaryzowania pojedynczego złącza $p-n$ w kierunku przewodzenia. Wówczas stan na wyjściu byłby nieokreślony. Włączenie diody D_1 wyklucza możliwość powstania takiej sytuacji. Zadaniem rezystora R_3 jest ograniczenie wartości prądu w stanie przejściowym. Przy zmianie stanu na wyjściu z „0” do „1” chwilowo oba tranzystory T_3, T_4 przewodzą (jest to taki stan, w którym tranzystor T_3 jeszcze „nie wyszedł” z nasycenia, a tranzystor T_4 już „wchodzi” w nasycenie) i wówczas pobór prądu ze źródła U_{CC} jest ograniczony przez rezystor R_3 do wartości

$$I = \frac{U_{CC} - U_{CE4(sat)} - U_{D1} - U_{CE3(sat)}}{R_3}$$

$$I = \frac{5 \text{ V} - 0,2 \text{ V} - 0,7 \text{ V} - 0,2 \text{ V}}{130} = 30 \text{ mA}$$

Pracę całego układu NAND w dwu różnych stanach ilustrują schematy na rys. 10.10, na którym przedstawiono rozptył prądów oraz wartości prądów i potencjałów w typowych warunkach sterowania i obciążenia identycznymi bramkami. Na rysunku 10.10a przedstawiono sytuację dla sterowania obu wejść „1” logiczną (rozwarcie wejścia odpowiada również „1” logicznej, chociaż pozostawianie



Rys. 10.10
Bramka NAND/TTL w dwu stanach: a) sterowanie obu wejść „1” logiczną; b) sterowanie jednego lub obu wejść „0” logicznym; obwody, w których płyną główne prądy, narysowano linią grubą

rozwartych wejść nie jest zalecane; wejścia nieużywane zwykle łączy się z dodatnim biegunem źródła zasilania U_{CC} , a na rys. 10.10b — dla sterowania jednym lub obu wejść „0” logicznym.

Najbardziej rozpowszechnioną rodziną układów scalonych TTL jest seria układów SN 54/74, opracowana i wprowadzona do produkcji w 1964 r. w firmie Texas Instruments (litery SN pochodzą od słów Semiconductor Networks). Serie SN 54 (tzw. wojskowa) i SN 74 (tzw. przemysłowa) są funkcjonalnie równoważne, a różnią się tylko zakresem dopuszczalnej temperatury pracy ($-55...+125^{\circ}\text{C}$ dla SN 54 oraz $0...70^{\circ}\text{C}$ dla SN 74) oraz tolerancją napięcia zasilania ($5\text{ V} \pm 0,5\text{ V}$ dla SN 54 oraz $5 \pm 0,25\text{ V}$ dla SN 74). Parametry statyczne i dynamiczne układu SN 7400, zawierającego cztery bramki dwuwejściowe (każda bramka jest zbudowana zgodnie ze schematem przedstawionym na rys. 10.7) przedstawiono w tabl. 10.3, a jego podstawową charakterystykę, tj. charakterystykę przenoszenia pokazano na rys. 10.11. Wyszczególnione parametry odnoszą się do tzw. serii standardowej układów TTL. Są również wytwarzane serie zmodyfikowane, w których uzyskuje się bądź mniejsze straty mocy (seria SN 74 L), bądź mniejsze czasy propagacji (seria SN 74 H). W serii SN 74 L (L od ang. *low*, przy czym chodzi o małą moc) bramka NAND ma schemat identyczny jak dla serii standardowej

Podstawowe parametry statyczne i dynamiczne bramki NAND serii SN 54/74

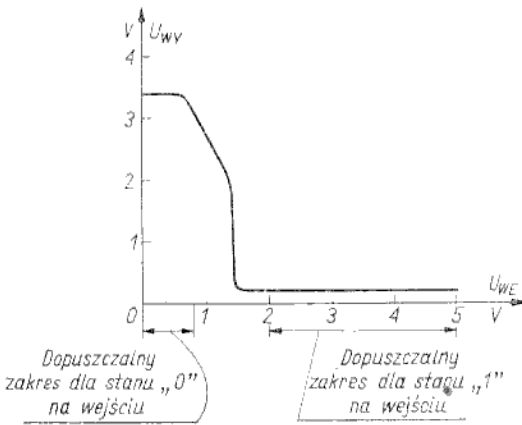
Tablica 10.3

Warunki pracy		SN 54	SN 74
Napięcie zasilania U_{CC}	V	$5 \pm 0,5$	$5 \pm 0,25$
Zakres temperatury T	°C	$-55 \dots +125$	$0 \dots 70$
Maksymalna obciążalność logiczna N		10 bramek	

Parametry		Min.	Typ.	Maks.	Warunki pomiaru		
Stacyjne	U_{IL}	Napięcie wejściowe w stanie „0” na obu końcówkach wejściowych niezbędne do zapewnienia stanu „1” na wyjściu	V		0,8	Minimalne U_{CC}	
	U_{IH}	Napięcie wejściowe w stanie „1” na obu końcówkach niezbędne do zapewnienia stanu „0” na wyjściu	V	2		Minimalne U_{CC}	
	U_{OL}	Napięcie wyjściowe w stanie „0”	V		0,22	0,4	Minimalne U_{CC} $U_I = 2 \text{ V}, I_0 = 16 \text{ mA}$
	U_{OH}	Napięcie wyjściowe w stanie „1”	V	2,4	3,3		Minimalne U_{CC} $U_I = 0,8 \text{ V}, I_0 = -400 \mu\text{A}$
	I_{IL}	Prąd wejściowy w stanie „0” (każde wejście)	mA			1,6	Maksymalne U_{CC} $U_I = 0,4 \text{ V}$
	I_{IH}	Prąd wejściowy w stanie „1” (każde wejście)	μA			40	Maksymalne U_{CC} $U_I = 2,4 \text{ V}$
	I_{CCL}	Prąd zasilania w stanie „0” na wyjściu	mA			22	Maksymalne U_{CC} $U_I = 5 \text{ V}$
	I_{CCH}	Prąd zasilania w stanie „1” na wyjściu	mA			8	Maksymalne U_{CC} $U_I = 5 \text{ V}$
Dynamiczne	t_{pLH}	Czas propagacji sygnału do stanu „1” na wyjściu	ns		11	22	$U_{CC} = 5 \text{ V}, R_L = 400 \Omega$ $C_L = 15 \text{ pF}, N = 10$
	t_{pHL}	Czas propagacji sygnału do stanu „0” na wyjściu	ns		7	15	Jak wyżej

(rys. 10.7), lecz wszystkie rezystory mają 10-krotnie większe wartości nominalne. Straty mocy w pojedynczej bramce SN 74 L wynoszą 1 mW, a czas propagacji jest równy 33 ns. W serii SN 74 H (H od ang. *high*, przy czym chodzi o dużą szybkość — bramka NAND ma schemat niemal identyczny jak dla serii standardowej¹⁾ (rys. 10.7), lecz wszystkie rezystory mają około 2-krotnie mniejsze

¹⁾ W porównaniu z bramką serii standardowej w bramce SN 74 H jest jeden dodatkowy rezystor i tranzystor pracujący w układzie Darlingtona z tranzystorem T_4 .



Rys. 10.11
Charakterystyka przenoszenia $U_{WY}(U_{WE})$ bramki NAND serii SN 54/74

wartości nominalne. W związku z tym straty mocy są około 2-krotnie większe niż dla bramki z serii standardowej i wynoszą 22 mW dla pojedynczej bramki, a czas propagacji jest równy 6 ns.

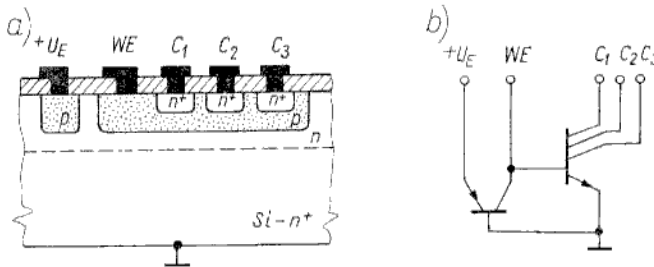
Bardzo duże znaczenie ma jeszcze inna seria układów TTL wytwarzana pod nazwą SN 54/74 S. W tych układach tranzystory nie pracują w nasyceniu, dlatego omówimy je w odrębnym punkcie (p. 10.2.2.1).

Układy I^2L

10.2.1.2

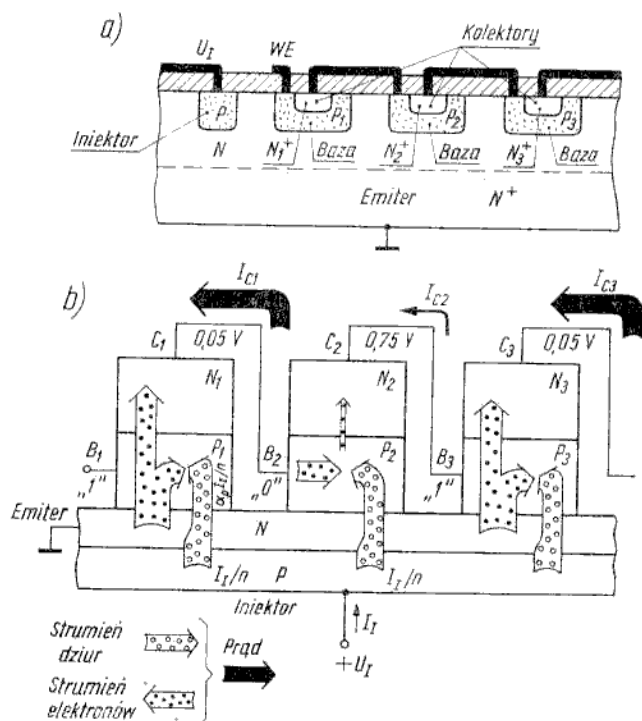
W układach I^2L jest realizowana idea wykorzystania jednego obszaru półprzewodnika jako wspólnej części kilku elementów układu, stąd układy I^2L są również znane pod nazwą MTL (ang. *Merged Transistor Logic* — układy logiczne z tranzystorami złączonymi).

Strukturę oraz schemat elektryczny podstawowej bramki logicznej (inwertera I^2L) przedstawiono na rys. 10.12. Bramka składa się z dwóch tranzystorów typu



Rys. 10.12
Struktura (a) i schemat elektryczny (b) inwertera I^2L

$n-p-n$ oraz $p-n-p$. Baza tranzystora $p-n-p$ jest zarazem emiterym tranzystora $n-p-n$. Wielokolektorowy tranzystor $n-p-n$ o strukturze wertykalnej (pionowej) spełnia funkcję inwertera, natomiast tranzystor $p-n-p$ o strukturze lateralnej (poziomej) spełnia funkcję źródła prądu bazy tranzystora $n-p-n$ oraz prądu kolektora (obciążenia) dla tranzystora $n-p-n$ poprzedniej bramki. Fizyczne

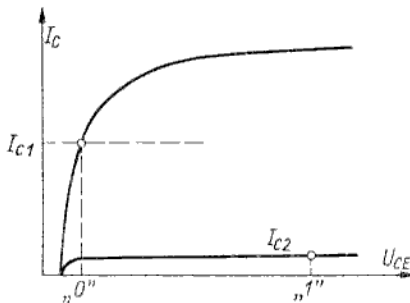


Rys. 10.13

Struktura trzech jednokolektorowych inwerterów I²L połączonych kaskadowo (a) oraz jej ilustracja poglądowa (b)

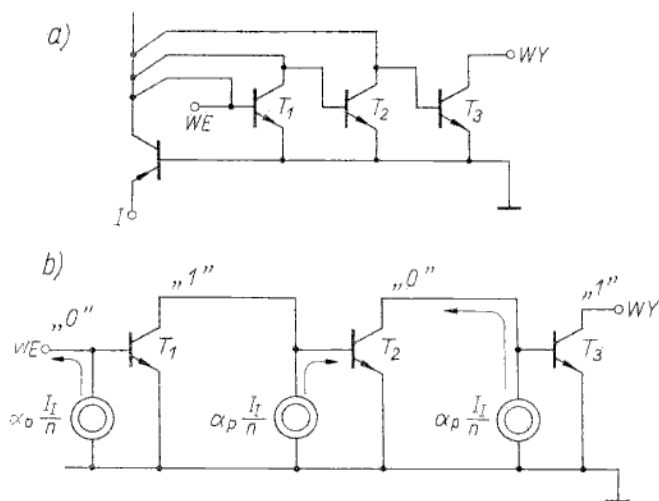
podstawy działania układów I²L zostaną wyjaśnione na przykładzie kaskadowego połączenia trzech inwerterów w postaci struktury pokazanej na rys. 10.13a. Dla uproszczenia są rozpatrywane tranzystory *n-p-n* jednokolektorowe. Wyróżnia się cztery rodzaje elektrod: emiter (obszary *N*⁺, *N* stanowiące wspólny emiter wszystkich tranzystorów *n-p-n* i zarazem bazę tranzystora *p-n-p*), iniektor (emiter wielokolektorowego tranzystora *p-n-p*) oraz bazy i kolektory tranzystorów *n-p-n*. Dla lepszej ilustracji procesów fizycznych zachodzących w tej strukturze, przedstawiono ją w postaci dogodniejszej do rozważań poglądowych na rys. 10.13b. Do elektrody iniektora jest doprowadzone napięcie dodatnie względem uziemionego emitera, o wartości wystarczającej do spolaryzowania złącza *p-n* w kierunku przewodzenia, tj. ok. 0,7 V (w istocie iniektor jest sterowany prądowo). Wskutek tej polaryzacji z iniektora (z warstwy *P* stanowiącej emiter tranzystora *p-n-p*) są wstrzykiwane dziury do emitera (do warstwy *N* stanowiącej bazę tranzystora *p-n-p*). Zatem w warstwie emitera gromadzi się nadmiarowy ładunek dziur. Oczywiście, dla zachowania neutralności elektrycznej, do tej warstwy wpływają również elektrony nadmiarowe z uziemionej końcówki emitera. Jednak dla przejrzystości rysunku nie pokazano na nim zarówno strumienia elektronów wpływających do emitera, jak i wielu innych mniej istotnych zjawisk (na przykład rekombinacji elektronów i dziur w obszarze emitera). Dziury (i elektrony) dyfundują wskutek gradientu koncentracji w kierunku baz tranzystorów *n-p-n*. Nadmiarowe ładunki dziur i elektronów powodują

częściową kompensację ładunków przestrzennych w warstwach zaporowych złącza emiter-bazy, co jest równoważne ze spolaryzowaniem tych złączy w kierunku przewodzenia. Dalszy przebieg zjawisk (idąc „w górę” struktury przedstawionej na rys. 10.13b) zależy od stanu, w jakim znajduje się elektroda bazy tranzystora $n-p-n$. Baza tranzystora T_1 znajduje się w stanie logicznym „1” (rozwarta lub spolaryzowana napięciem ok. $+0,7$ V). W tym przypadku polaryzacja złącza emiter-baza powoduje wstrzykiwanie elektronów z emitera do bazy. Część z tych elektronów rekombinuje w bazie z dziurami, a pozostałe są odbierane przez kolektor. Elektrony i dziury gromadzą się w obszarze złącza baza-kolektor powodując częściową kompensację ładunku przestrzennego w warstwie zaporowej tego złącza (podobnie jak dla złącza emiter-baza), czyli złącze baza-kolektor jest efektywnie spolaryzowane w kierunku przewodzenia. Tranzystor T_1 pracuje w nasyceniu i potencjał na jego kolektorze jest bardzo mały (ok. $+0,05$ V) z uwagi na małą rezystancję kolektor-emiter. W ten sposób złącze emiter-baza następnego tranzystora (T_2) jest, praktycznie biorąc, zwarte. Dlatego w tranzystorze T_2 nie ma wstrzykiwania elektronów z emitera do bazy, a strumień dziur odbierany przez bazę (obszar P_2 stanowiący kolektor tranzystora $p-n-p_2$) rekombinuje ze strumieniem elektronów wpływających do tego obszaru z kolektora poprzedniego tranzystora. Tranzystor T_2 pracuje więc na granicy obszaru aktywnego i odcięcia, czyli ma dużą rezystancję kolektor-emiter. W tych warunkach prąd kolektora tranzystora T_2 jest bardzo mały (na poziomie porównywalnym z prądem zerowym I_{CES}), a potencjał na kolektorze tego tranzystora ustala się na poziomie wysokim (ok. $+0,75$ V), odpowiadającym „1” logicznej. Zatem kolejny tranzystor T_3 pracuje w warunkach analogicznych jak tranzystor T_1 . Na rysunku 10.14 przedstawiono charakterystyk-



Rys. 10.14
Charakterystyki $I_C(U_{CE})$ tranzystorów T_1 , T_2 z zaznaczonymi punktami pracy

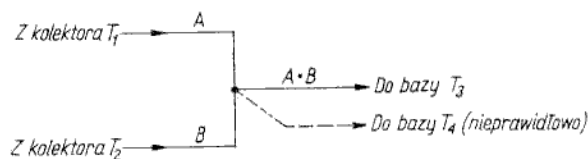
$I_C(U_{CE})$ tranzystorów T_1 , T_2 z zaznaczonymi punktami pracy tych tranzystorów. W rozpatrywanym połączeniu kaskadowym tranzystorów $n-p-n$ każdy tranzystor jest równoważny inwerterowi, stanowiąc jednocześnie element sterowany i obciążenie dla poprzedniego tranzystora. Na wyjściach kolejnych inwerterów ustalają się na przemian stany „0” i „1”. Jest oczywiste, że zwarcie wejścia (bazy tranzystora T_1) spowoduje zmianę wszystkich stanów logicznych na przeciwnie. Na rysunku 10.15a przedstawiono schemat elektryczny struktury n inwerterów iniekcyjnych połączonych kaskadowo. N -kolektorowy tranzystor $p-n-p$ spełnia funkcję n źródeł prądowych, przy czym każde źródło ma wydajność $\alpha_p I_1/n$ (α_p — współczynnik wzmożenia prądowego tranzystora $p-n-p$, I_1 — prąd iniektora). Zatem układ z rys. 10.15a można przedstawić w równoważnej postaci ze źródłami prądowymi (rys. 10.15b).



Rys. 10.15

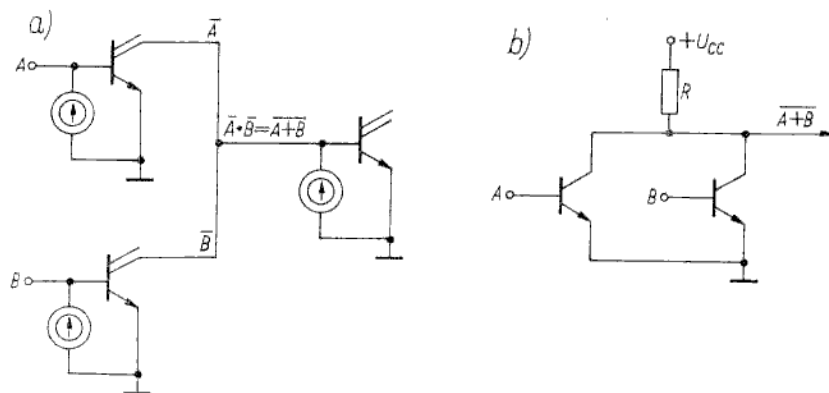
Schemat elektryczny struktury „ (w tym przypadku trzech) inwerterów I²L połączonych kaskadowo (a) oraz jego postać równoważna ze źródłami prądowymi reprezentującymi działanie iniektora (b)

Inwertery, jako najprostsze bramki logiczne, realizują funkcję NOT (negację). W układach I²L stosuje się oprócz funkcyj NOT również funkcyj AND (iloczyn logiczny) oraz NOR (negacja sumy logicznej). Funkcję iloczynu logicznego realizuje się bez tranzystorów przez bezpośrednie połączenie kilku wejść (rys. 10.16). Tylko



Rys. 10.16

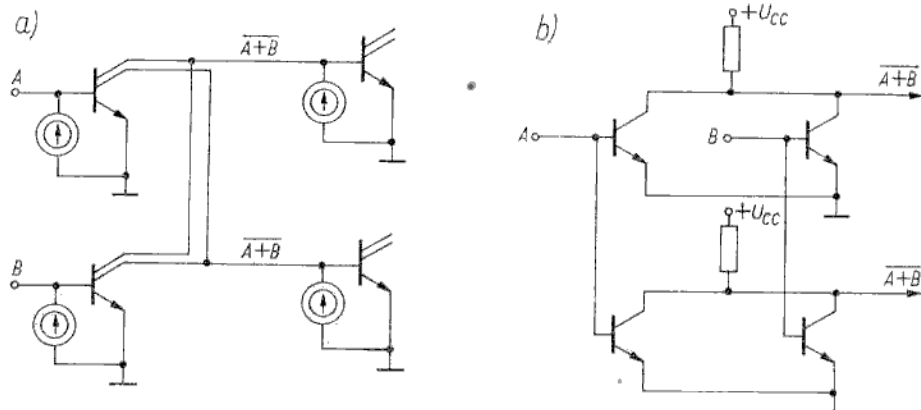
Realizacja funkcyj AND w układach I²L



Rys. 10.17

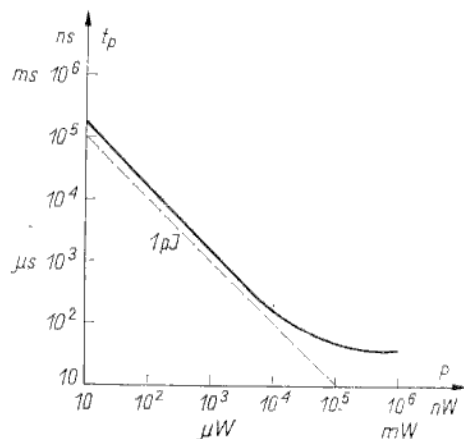
Realizacja funkcyj NOR w technice I²L (a) oraz DCTL (b)

w przypadku stanów „1” na obu wejściach A, B (tj. w przypadku wysokich poziomów napięć na kolektorach „odeiętych” tranzystorów T_1, T_2) na wyjściu dołączonym do bazy tranzystora T_3 otrzymuje się stan „1”. Tak proste rozwiązanie funkcji AND jest możliwe tylko w tym przypadku, jeśli nie ma połączeń między bazami, czyli każdy kolektor jest dołączony tylko do jednej bazy. Dlatego połączenie zaznaczone linią przerywaną na rys. 10.16 jest nieprawidłowe. Aby ten warunek można było spełnić w złożonych sieciach logicznych, wykonuje się inwertery wielokolektorowe o strukturze pokazanej na rys. 10.12. Funkcję logiczną NOR realizuje się podobnie jak w układach DCTL, co ilustruje rys. 10.17. Chociaż układy DCTL — jako przestarzałe — nie są rozpatrywane w tej książce, to jednak czytelnik z łatwością dostrzeże podobieństwo w zasadzie działania i różnice w budowie obu klas układów. Dla lepszej ilustracji sposobu



Rys. 10.18
Funkcjonalnie równoważne fragmenty sieci logicznej zrealizowane w technice I²L (a) oraz DCTL (b)

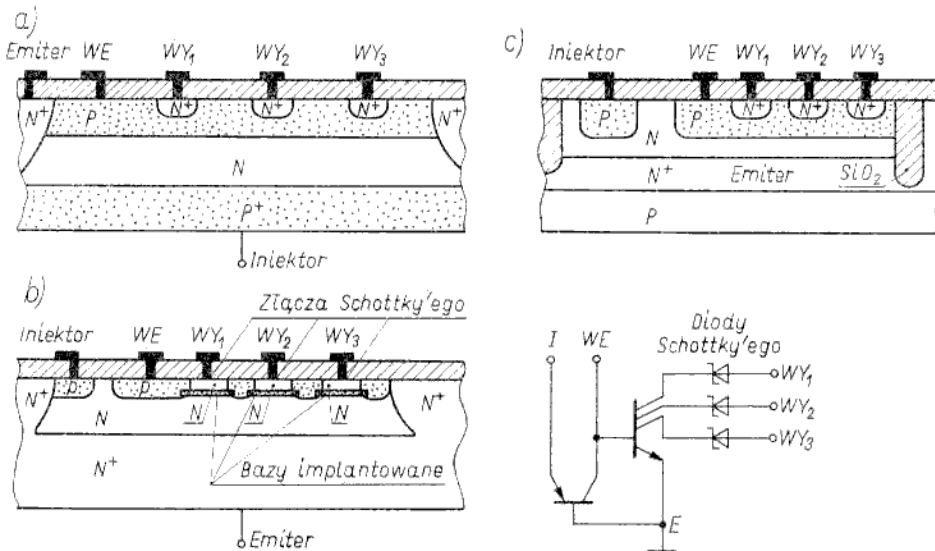
wykorzystania struktur wielokolektorowych na rys. 10.18 porównano funkcyj-
nalnie równoważne fragmenty sieci logicznej, zrealizowane w technice I²L oraz
DCTL.



Rys. 10.19
Zależność czasu propagacji od strat mocy
w pojedynczej bramce I²L

Zostaną teraz krótko omówione podstawowe parametry układów I²L. Amplituda logiczna, tj. różnica napięć odpowiadających poziomom logicznym „1” i „0”, wynosi w typowym przypadku ok. 0,7 V. Czas propagacji jest odwrotnie proporcjonalny do prądu iniektora, który można zmieniać w szerokim zakresie, od kilku nanoamperów do miliampera (w przeliczeniu na jedną bramkę). Ponieważ napięcie iniektora jest, praktycznie biorąc, niezależne od prądu i wynosi ok. 0,7 V (spadek napięcia na złączu *p-n* spolaryzowanym w kierunku przewodzenia), czas propagacji jest odwrotnie proporcjonalny do poboru mocy (rys. 10.19). Cechą charakterystyczną układów I²L jest zatem stała wartość wskaźnika energetycznego (iloczynu mocy pobieranej przez pojedynczą bramkę i czasu propagacji), która wynosi w typowym przypadku ok. 1 pJ. Jest to inaczej energia magazynowana w pojemnościach złączowych, gdyż straty mocy są związane z procesami przeładowania tych pojemności. Dopiero przy dużych wartościach poboru mocy (ponad 10 μW) czas propagacji dąży asymptotycznie do stałej wartości, określonej procesem magazynowania ładunku nośników nadmiarowych w tranzystorze *n-p-n*. Ta graniczna wartość czasu propagacji wynosi w typowym przypadku 20...30 ns. Napięcie zasilania układu I²L może się zmieniać w szerokim zakresie (najczęściej 1...15 V), gdyż źródło zasilania jest łączone z iniektorem przez rezystor ograniczający wartość prądu. Ważnym parametrem nieelektrycznym układów LSI jest gęstość upakowania, która dla układów I²L wynosi 100...300 bramek na 1 mm², czyli jest znacznie większa niż dla pozostałych klas układów bipolarnych i porównywalna z największą gęstością upakowania uzyskiwaną w układach MOS.

Dążenie do poprawy podstawowych parametrów, przede wszystkim wskaźnika energetycznego i gęstości upakowania, doprowadziło do opracowania wielu zmodyfikowanych technologii układów I²L. Trzy zmodyfikowane technologie ilustrują przekroje struktur pokazane na rys. 10.20. Cechą szczególną struktury z rys. 10.20a jest zastosowanie wertykalnej konstrukcji tranzystora *p-n-p*. W tych



Rys. 10.20

Struktury układów I²L zrealizowanych w trzech zmodyfikowanych technologiach; a) struktura z zasilaniem podłożowym; b) struktura z diodami Schottky'ego; c) struktura z izolacją dielektryczną

układach iniektorem jest podłoże (stąd nazwa ang. *Substrate Fed I²L* — układy I²L z zasilaniem podłożowym), dzięki czemu uzyskuje się około 2-krotne zwiększenie gęstości upakowania w porównaniu z technologią standardową (rys. 10.12).

Ponieważ straty energii są związane z przeładowaniem pojemności złączowych (CU^2 , przy czym U — amplituda logiczna), zmniejszenie wartości wskaźnika energetycznego można uzyskać przez zmniejszenie amplitudy logicznej lub pojemności złącza. Zmniejszenie amplitudy logicznej uzyskuje się w układach I²L z diodami Schottky'ego włączonymi szeregowo z kolektorami (rys. 10.20b). Dla uzyskania złącza Schottky'ego jest konieczny styk metalu z półprzewodnikiem słabo domieszkowanym, dlatego w sposób istotny zmienia się cały proces technologiczny. Bazy tranzystorów *n-p-n* są wykonywane metodą implantacji jonów. Amplituda logiczna maleje w tych układach o wartość spadku napięcia na przewodzącej diodzie Schottky'ego, tj. z 700 do 350 mV. Zatem wartość wskaźnika energetycznego maleje 3...4 razy.

W układach I²L z izolacją dielektryczną (rys. 10.20c), nazywanych również I³L (ang. *Isoplanar Integrated Injection Logic*), uzyskuje się radykalne zmniejszenie pojemności złączowych i dyfuzyjnych. Wskutek tego zmniejsza się wskaźnik energetyczny do ok. 0,1 pJ, a minimalna wartość czasu propagacji maleje do 3...4 ns. Uzyskuje się zarazem bardzo dużą gęstość upakowania, wynoszącą ok. 300 bramek/mm².

Warto zaznaczyć, że układy I²L są technologicznie kompatybilne z układami TTL, co umożliwia łączenie obu klas układów we wspólnej strukturze półprzewodnikowej. W podsumowaniu można stwierdzić, że układy I²L stanowią najmlodszą (pierwsze opracowania w 1972 r.) klasę układów scalonych, wzbudzającą wiele kontrowersji co do dalszych perspektyw jej rozwoju. Do 1977 r. był to rozwój dynamiczny, zapowiadający rychłą dominację technologii I²L. W 1977 r. na świecie produkowano układy iniekcyjne w asortymencie obejmującym mikroprocesory, pamięci, układy do zegarków itd. Jednakże oczekiwane (i reklamowane) połączenie zalet układów bipolarnych (duża szybkość) i MOS (mała moc, duża gęstość upakowania) nie nastąpiło dotychczas w układach I²L wytwarzanych ekonomicznie opłacalną technologią. Dlatego w latach 1978-1979 liczba producentów zainteresowanych układami iniekcyjnymi zmalała i dalsze perspektywy rozwoju tych układów są trudne do określenia. Jako ciekawostkę można odnotować fakt, że jest możliwa praca układów I²L bez zasilania elektrycznego, gdyż nośniki nadmiarowe niezbędne dla ich pracy można generować pod wpływem światła lub promieniowania radioaktywnego.

Bramki logiczne z tranzystorami nienasyconymi

10.2.2

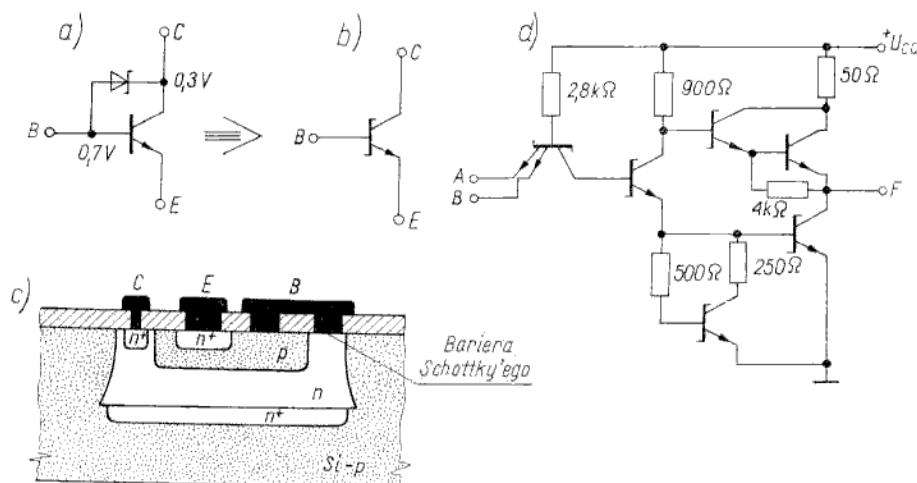
Wadą układów z tranzystorami pracującymi w zakresie nasycenia są duże czasy propagacji (opóźnienie sygnału wyjściowego w stosunku do wejściowego), gdyż — jak wiadomo z rozdziału 5 — w bazie tranzystora nasyconego gromadzi się duży ładunek nośników mniejszościowych, a im większy jest ładunek bazy, tym dłużej trwa proces wyłączenia tranzystora (przejście ze stanu przewodzenia do zatkania). Szybkość działania układów cyfrowych można znacznie zwiększyć przy takim zaprojektowaniu bramek logicznych, żeby tranzystory nie pracowały w nasyceniu. Ta idea jest realizowana w dwóch rodzinach układów logicznych, tj.:

- TTLS (technika tranzystorowa z diodami Schottky'ego),
- ECL (technika ze sprzężeniem emiterowym).

Układy TTLS

10.2.2.1

Przy pracy tranzystora w zakresie nasycenia złącze baza-kolektor jest polaryzowane w kierunku przewodzenia i do bazy są wstrzykiwane nośniki mniejszościowe z kolektora. Typowa wartość napięcia na krzemowym złączu $p-n$ pracującym w kierunku przewodzenia wynosi $0,7\text{ V}$, przy czym — praktycznie biorąc — istotny wzrost prądu obserwuje się dla napięcia większego niż $0,5\text{ V}$. Jeżeli więc napięcie na złączu baza-kolektor (dla kierunku przewodzenia) ograniczymy do wartości mniejszych niż $0,5\text{ V}$, to uniemożliwimy pracę tranzystora w nasyceniu¹⁾. Można to uzyskać bocznikując złącze baza-kolektor innym złączem, które zaczyna dobrze przewodzić przy napięciu polaryzacji mniejszym niż $0,5\text{ V}$. Takie właściwości ma złącze $m-s$ (dioda Schottky'ego). Dla kontaktu Al-Si (typ n) uzyskuje się napięcie pracy w kierunku przewodzenia równe $0,4\text{ V}$. Na rysunku 10.21 przedstawiono sposób włączenia diody Schottky'ego (a), symbol graficzny (b) oraz schemat elektryczny bramki NAND serii SN 54/74 S (d). Na rysunku 10.21 przedstawiono również strukturę planarną tranzystora z diodą Schottky'ego (c).



Rys. 10.21

Bramka NAND serii TTLS: a) sposób włączenia diody Schottky'ego; b) symbol graficzny tranzystora z diodą Schottky'ego; c) struktura planarna tranzystora z diodą Schottky'ego; d) schemat elektryczny bramki NAND serii SN 54/74 S

tranzystora z bocznikującą diodą Schottky'ego (b), przekrój struktury planarnej tego tranzystora (c) oraz schemat typowej bramki TTLS z serii SN 54/74 S (d).

Przy pracy tranzystora w stanie zatkania złącze baza-kolektor wraz z diodą Schottky'ego są polaryzowane zaporowo i w tym stanie dioda Schottky'ego nie ma żadnego wpływu na pracę tranzystora. Przy polaryzacji tranzystora w stanie przewodzenia w miarę „wchodzenia” w stan nasycenia napięcie baza-kolektor może osiągnąć co najwyżej wartość $0,4\text{ V}$, przy której „nadmiar” prądu bazy nie wpływa do obszaru bazy, lecz jest odprowadzany przez diodę Schottky'ego do obwodu kolektora. W diodzie Schottky'ego, jak wiadomo, nie występuje magazynowanie

¹⁾ Teoretycznie tranzystor pracuje w zakresie nasycenia przy dowolnie małej polaryzacji złącza baza-kolektor w kierunku przewodzenia, co wynika z definicji stanu nasycenia. W tym przypadku przyjmuje się jednak kryteria praktyczne.

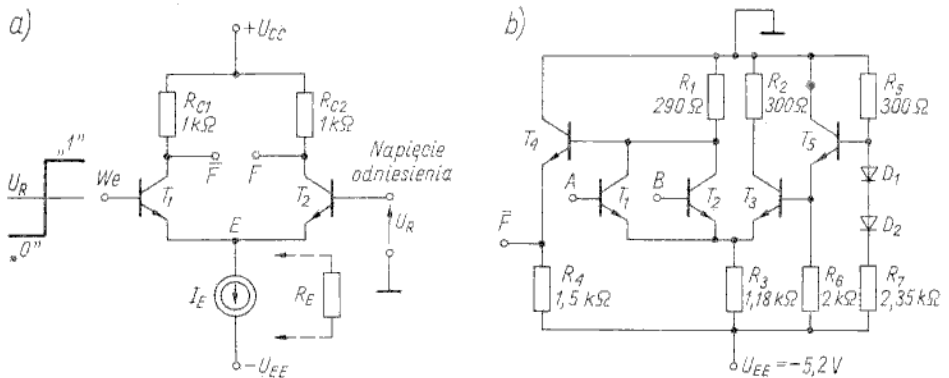
nośników mniejszościowych, zatem podczas wyłączenia tranzystora (powrót do stanu zatkania) dioda Schottky'ego nie wnosi żadnego opóźnienia.

Podstawowa bramka NAND z serii SN 54/74 S ma czas propagacji 3 ns przy stratach mocy 22 mW.

Układy ECL

10.2.2.2

Bramki logiczne ECL są zbudowane z tranzystorów o sprzężonych emiterach i działają na zasadzie przełączania prądu, dlatego niekiedy rodzina tych układów logicznych występuje pod nazwą CML (ang. *Current-Mode Logic*). Na rysunku 10.22 przedstawiono schemat inwertera (a) i podstawowej bramki NOR (b)



Rys. 10.22
Inwerter ECL (a) i podstawowa bramka NOR zrealizowana w technice ECL (b)

zrealizowanej w technice ECL. Inwerter jest zbudowany z dwu tranzystorów o emiterach połączonych razem i zasilanych ze źródła prądowego \$I_E\$. Na bazę tranzystora \$T_2\$ jest podane napięcie odniesienia \$U_R\$. Dla uproszczenia przyjmijmy chwilowo, że napięcie \$U_R = 0V\$, tj. baza tranzystora \$T_2\$ jest połączona z masą. Załóżmy również, że wejściowe sygnały logiczne „0” i „1”, doprowadzane do bazy tranzystora \$T_1\$, wynoszą odpowiednio \$-1\$ i \$1V\$.

Jeżeli na wejście jest podane napięcie \$-1V\$, to złącze emiter-baza tranzystora \$T_2\$ jest spolaryzowane w kierunku przewodzenia i napięcie w punkcie \$E\$ (rys. 10.22a)

$$U_R - \varphi_B = 0 - 0,7V = -0,7V$$

Złącze emiter-baza tranzystora \$T_1\$ jest wówczas spolaryzowane zaporowe napięciem \$-0,7V - (-1V) = 0,3V\$. Tranzystor \$T_1\$ jest więc zatkany, napięcie na jego kolektorze wynosi \$U_{CC}\$, a cały prąd \$I_E\$ płynie przez tranzystor \$T_2\$. Napięcie na kolektorze tranzystora \$T_2\$ jest równe (przy założeniu, że \$I_C \approx I_E\$)

$$U_{CC} - R_{C2} I_E$$

Jeżeli na wejście jest podane napięcie \$1V\$, to złącze emiter-baza tranzystora \$T_1\$ jest spolaryzowane w kierunku przewodzenia, czyli napięcie w punkcie \$E\$

$$1V - \varphi_B = 0,3V$$

Złącze emiter-baza tranzystora \$T_2\$ jest spolaryzowane w kierunku zaporowym napięciem \$0,3V\$. Tranzystor \$T_2\$ jest więc zatkany, napięcie na jego kolektorze

wynosi U_{CC} , a cały prąd I_E płynie przez tranzystor T_1 . Napięcie na kolektorze tranzystora T_1 jest równe (przy założeniu, że $I_C \approx I_E$)

$$U_{CC} - R_{C1} I_E$$

Dla $R_{C1} = R_{C2}$ zmiany napięć na obu kolektorach mają jednakowe wartości, przy czym na kolektorze tranzystora T_1 uzyskuje się sygnał inwersyjny, a na kolektorze tranzystora T_2 — zgodny w fazie z sygnałem wejściowym. Przy odpowiednim doborze wartości rezystorów $R_{C1} = R_{C2}$ i prądu I_E punkty pracy tranzystorów „nie wchodzą” w zakres nasycenia. Na przykład dla $I_E = 2$ mA oraz $R_{C1} = R_{C2} = 1$ k Ω napięcia na kolektorach zmieniają się w zakresie 3...5 V, czyli tranzystory pracują tylko w zakresie aktywnym lub w zakresie zatkania. Dzięki temu szybkość przełączania inwertera ECL jest bardzo duża.

W układach realizowanych w praktyce sterowanie prądowe obwodu emiterów uzyskuje się przez włączenie rezystora o odpowiedniej rezystancji (zamiast źródła I_E na rys. 10.22a). Rezystor ten oznaczamy przez R_E i wyznaczmy wartość napięcia na kolektorze w obu stanach logicznych (wysokim i niskim):

$$U_{OH} = U_{CC}$$

$$U_{OL} = U_{CC} - \frac{R_C}{R_E} (U_R - \varphi_B - U_{EE})$$

Bardzo korzystna jest zależność napięcia U_{OL} od ilorazu R_C/R_E , gdyż dzięki temu dokładność i powtarzalność napięcia U_{OL} zależy od tolerancji względnej rezystorów, która wynosi 1...3% w porównaniu z ok. 20% tolerancją bezwzględną.

Przez połączenie równolegle kilku tranzystorów z tranzystorem T_1 uzyskuje się kilkuwejściową bramkę logiczną, która dla wyjścia z kolektora tranzystora T_1 spełnia funkcję NOR, a dla wyjścia z kolektora tranzystora T_2 — funkcję OR.

Schemat typowej dwuwejściowej bramki NOR, występującej w układach serii MECL II (firma Motorola), przedstawiono na rys. 10.22 b. Rola tranzystorów T_1, T_2, T_3 jest jasna na podstawie wcześniejszych rozważań. Tranzystor T_4 pracuje jako wtórnik emiterowy separujący wyjście kolektorowe od obciążenia. Dzięki temu z wyjścia jednej bramki można sterować wejścia kilkunastu identycznych bramek (wzmocnienie logiczne w typowym przypadku wynosi 15). Istotne znaczenie ma również przesunięcie poziomu napięcia między kolektorami tranzystorów T_1, T_2 a wyjściem \bar{F} o wartość ok. 0,7 V. Tranzystor T_5 wraz z dzielnikiem polaryzującym R_5, D_1, D_2, R_7 służy do ustalenia napięcia odniesienia na bazie tranzystora T_3 (w rozpatrywanym układzie $U_R = -1,2$ V).

Typowy czas propagacji bramki ECL wynosi 1...5 ns przy stratach mocy 30...60 mW (na przykład dla serii MECL III — 1 ns, 55 mW).

Bramki logiczne MOS

10.3

Układy logiczne MOS najogólniej dzieli się na statyczne i dynamiczne. W obu przypadkach podstawowym podzespołem układów logicznych jest inwerter zbudowany z dwu tranzystorów, przy czym jeden z tych tranzystorów jest sterowany, a drugi spełnia funkcję obciążenia. Stosując różne kombinacje połączeń dwu tranzystorów można otrzymać wiele typów inwerterów różniących się parametrami elektrycznymi. We wszystkich inwerterach, stosowanych w praktycznych układach logicznych, elementem sterowanym jest tranzystor z kanałem

wzbogacany, gdyż tylko ten rodzaj tranzystorów umożliwia bezpośrednie łączenie dowolnej liczby inwerterów bez konieczności zmiany poziomu napięcia. Funkcję obciążenia natomiast może spełniać zarówno tranzystor z kanałem wzbogacany, jak i zubożany. W związku z tym można wyszczególnić następujące warianty połączeń mających znaczenie praktyczne:

- 1) tranzystor z kanałem wzbogacany typu p — tranzystor z kanałem wzbogacany typu p (PEMOS — PEMOS)¹⁾;
- 2) tranzystor z kanałem wzbogacany typu p — tranzystor z kanałem zubożany typu p (PEMOS — PDMOS);
- 3) 4) jak w p. 1), 2), lecz kanały typu n , tj. inwertery NEMOS—NEMOS oraz NEMOS—NDMOS;
- 5) tranzystor z kanałem wzbogacany typu p — tranzystor z kanałem wzbogacany typu n (para komplementarna symetryczna CMOS lub COS/MOS).

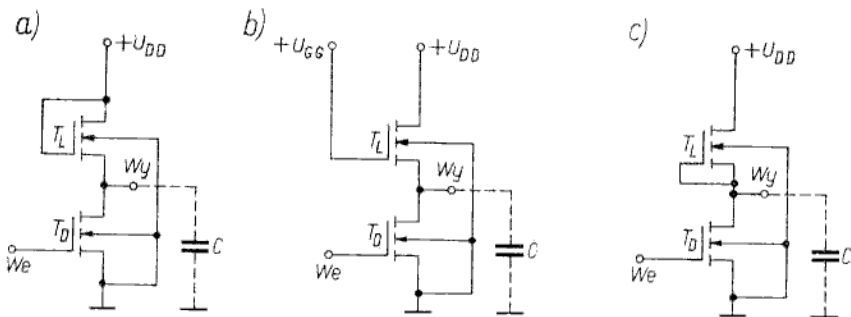
W pierwszych latach rozwoju technologii MOS wytwarzano wyłącznie układy z inwerterami wymienionymi w punkcie 1) (tzw. technologia PMOS). Obecnie dominują układy wytwarzane z inwerterów wymienionych w p. 3), 4) (tzw. technologia NMOS). Główną zaletą układów NMOS w porównaniu z PMOS jest większa szybkość działania z uwagi na większą ruchliwość elektronów niż dziur.

Układy zbudowane z inwerterów CMOS mają od wielu lat „trwałą pozycję” ze względu na ich unikalne właściwości. Uwzględniając powyższe uwagi rozważania ograniczymy do układów logicznych NMOS (statycznych i dynamicznych) oraz CMOS.

Bramki logiczne statyczne NMOS

10.3.1

Na rysunku 10.23 przedstawiono trzy warianty inwerterów statycznych NMOS, różniących się rodzajem obciążenia:



Rys. 10.23

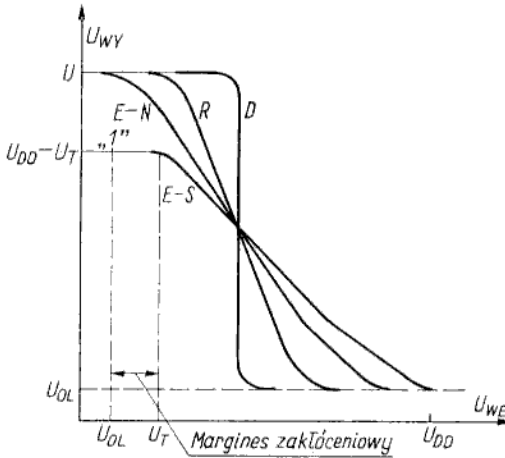
Trzy warianty inwerterów statycznych NMOS, różniące się rodzajem obciążenia: a) ENMOS w zakresie nasycenia; b) ENMOS w zakresie nienasycenia; c) DNMOS

- ENMOS w zakresie nasycenia,
- ENMOS w zakresie nienasycenia,
- DNMOS.

¹⁾W skróconych nazwach tranzystorów pierwsza litera oznacza typ przewodnictwa kanału, druga litera: E — wzbogacany (ang. *enhancement*), D — zubożany (ang. *depletion*). Na przykład PEMOS — tranzystor MOS z kanałem wzbogacany typu p .

Dla tych trzech przypadków obowiązują następujące zasady polaryzacji tranzystora obciążającego¹⁾:

- $U_{GG} < U_{DD} + U_T^l$ (tranzystor obciążający pracuje w zakresie nasycenia). W praktyce stosuje się pojedyncze źródło zasilania $U_{GG} = U_{DD}$.
- $U_{GG} - U_T^l \geq U_{DD}$ (tranzystor obciążający pracuje w zakresie nienasycenia).
- $U_{GS}^l = 0$ (zwarcie bramki ze źródłem w obciążającym tranzystorze z kanałem zubożanym).



Rys. 10.24
Charakterystyki przenoszenia U_{WY} (U_{WE}) inwertera NMOS dla przypadku obciążenia: tranzystorem ENMOS pracującym w zakresie nasycenia (krzywa E-S), tranzystorem ENMOS pracującym w zakresie nienasycenia (krzywa E-N), rezystorem (krzywa R), tranzystorem DNMOS (krzywa D)

Na rysunku 10.24 przedstawiono charakterystyki przenoszenia dla trzech rozpatrywanych wariantów inwertera MOS oraz dla porównania uwzględniono również przypadek obciążenia rezystorem R.

Praca inwertera dla wszystkich trzech przypadków przebiega podobnie. Przy niskim poziomie napięcia wejściowego („0” logiczne), tj. dla

$$U_{IL} < U_T$$

tranzystor sterowany nie przewodzi, jego rezystancja dren-źródło jest zatem bardzo duża i na drenie tego tranzystora odkłada się napięcie, praktycznie biorąc, równe U_{DD} lub $U_{DD} - U_T$ dla przypadku pracy tranzystora obciążającego w zakresie nasycenia. Przy wysokim poziomie napięcia wejściowego („1” logiczna), tj. dla

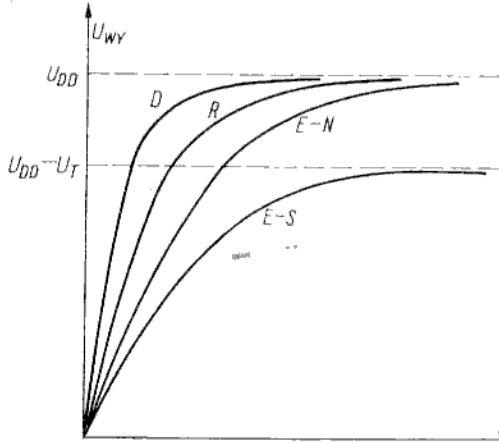
$$U_{IH} \geq U_T$$

tranzystor sterowany jest w stanie przewodzenia, jego rezystancja dren-źródło jest zatem względnie mała. Na drenie tego tranzystora odkłada się wówczas napięcie o wartości wyznaczonej przez iloraz rezystancji źródło-dren tranzystorów obciążającego i sterowanego. Aby uzyskać dużą wartość tego ilorazu, czyli małą wartość napięcia wyjściowego, odpowiadającego stanowi logicznemu „0”, tranzystor obciążający wykonuje się z cieńszym i dłuższym kanałem niż dla tranzystora sterowanego.

Czas przełączania inwertera zależy od stałej czasowej procesu ładowania (rozładowania) pojemności obciążenia. Z uwagi na większą rezystancję źródło-dren tranzystora

¹⁾ Parametry tranzystora obciążającego są oznaczane górnym indeksem l (ang. *load*), a tranzystora sterowanego — górnym indeksem d (ang. *driver*).

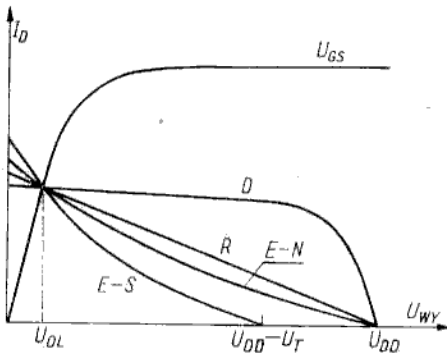
obciążającego stała czasowa ładowania tej pojemności jest większa niż stała czasowa rozładowania. Czas wyłączenia inwertera (wyłączenie — przejście ze stanu „0” do „1” na wyjściu) jest zatem większy niż czas włączania (włączenie — przejście ze stanu „1” do „0” na wyjściu). Na rysunku 10.25 porównano przebiegi



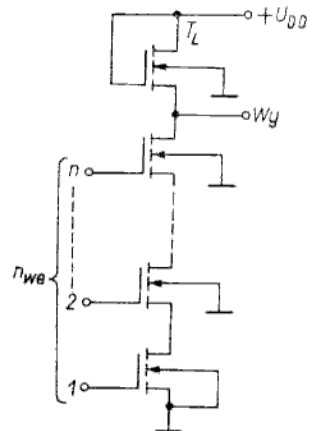
Rys. 10.25
Porównanie przebiegów $U_{wy}(t)$ podczas wyłączenia inwertera NMOS dla różnych wariantów obciążenia

czasowe napięcia wyjściowego podczas wyłączenia dla różnych wariantów inwertera. Z przedstawionych przebiegów widać, że czas wyłączenia jest najkrótszy dla przypadku obciążenia tranzystorem DNMOS, a najdłuższy dla obciążenia tranzystorem ENMOS, pracującym w zakresie nasycenia. Wynika to stąd, że prąd ładowania pojemności obciążającej jest w pierwszym przypadku największy, a w drugim najmniejszy, co ilustrują wykresy linii obciążenia na polu charakterystyk wyjściowych tranzystora sterowanego (rys. 10.26).

Na bazie rozpatrywanych inwerterów można w łatwy sposób uzyskać bramki logiczne NAND lub NOR. Funktor NAND realizuje się przez połączenie szeregowo n tran-

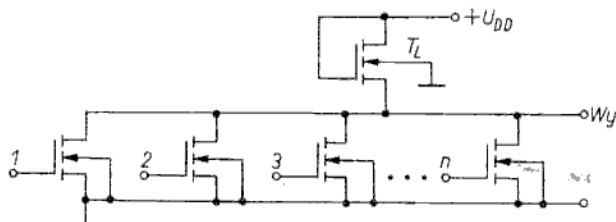


Rys. 10.26
Wykresy linii obciążenia na polu charakterystyk wyjściowych tranzystora sterowanego dla różnych wariantów obciążenia



Rys. 10.27
Funktor NAND zbudowany z tranzystorów NMOS

zystorów sterowanych (n wejść) i tranzystora obciążającego, co na przykładzie obciążenia tranzystorem ENMOS, pracującym w zakresie nasycenia, przedstawia rys. 10.27. Napięcie wyjściowe przyjmuje wartość U_{OL} („0” logiczne) tylko wówczas, gdy wszystkie tranzystory sterowane znajdują się w stanie przewodzenia. Podstawowym problemem w tym układzie jest otrzymanie małej wartości napięcia U_{OL} mimo dodawania się n spadków napięć U_{DS} . Stąd, praktycznie biorąc, liczba wejść jest ograniczona do trzech.



Rys. 10.28
Funktor NOR zbudowany z tranzystorów NMOS

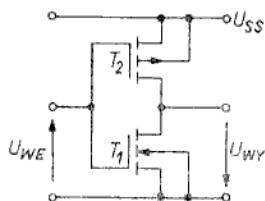
Inna jest sytuacja w funktorze NOR, którego schemat przedstawiono na rys. 10.28. W tym przypadku liczba tranzystorów sterowanych, połączonych równolegle, nie jest ograniczona wartością napięcia wyjściowego, gdyż przy jednoczesnym włączeniu kilku tranzystorów wartość napięcia U_{OL} jest mniejsza niż w przypadku włączenia pojedynczego tranzystora. Natomiast ograniczenie liczby wejść jak i wzmocnienia logicznego (liczby elementów obciążających) wynika ze zwiększania się pojemności obciążającej, a więc zmniejszania szybkości działania bramki NOR.

Wszystkie dotychczas omawiane funktoiry logiczne, zbudowane z tranzystorów MOS, mogą pracować dynamicznie, jeżeli do ich obwodów sterujących lub linii zasilania zostaną doprowadzone napięcia w postaci zsynchronizowanych ciągów periodycznych impulsów taktowych. Jest to możliwe dzięki dużym stałym czasowym rozładowania pojemności pasożytniczych tranzystorów MOS. Funktoiry zasilane dynamicznie są stosowane w rejestrach przesuwających, które rozpatrzmy w p. 10.6.1.

Bramki logiczne komplementarne

10.3.2

Inwerter komplementarny tworzą dwa tranzystory z kanałami wzbogacanymi typu p i typu n , przy czym dreny i bramki tych tranzystorów są połączone ze sobą, natomiast źródła i podłoża są przyłączone do linii zasilania (rys. 10.29). Gdy napięcie wejściowe jest równe napięciu zasilania, $U_{WE} = U_{SS}$, tranzystor T_1 przewodzi, a tranzystor T_2 znajduje się w stanie nieprzewodzenia. Wówczas napięcie wyjściowe jest, praktycznie biorąc, równe zero (typowe rezystancje tran-

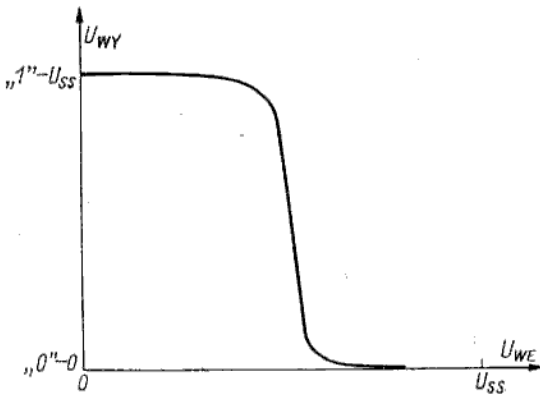


Rys. 10.29
Inwerter komplementarny (CMOS)

zystorów przewodzącego i nieprzewodzącego wynoszą odpowiednio $500 \dots 750 \Omega$ i $5000 \text{ M}\Omega$), czemu odpowiada stan logiczny „0”. Gdy napięcie wejściowe jest równe zeru, tranzystor T_1 nie przewodzi, tranzystor T_2 natomiast znajduje się w stanie przewodzenia. Wówczas napięcie wyjściowe jest, praktycznie biorąc, równe napięciu zasilania U_{SS} , czemu odpowiada stan logiczny „1”. Moc pobierana przez układ w stanie spoczynkowym jest znikomo mała (rzęd nanowatów), gdyż zawsze jeden z tranzystorów nie przewodzi. Natomiast moc tracona w czasie przełączania, związana przede wszystkim ze zjawiskiem przeładowywania pojemności obciążającej inwerter, wzrasta proporcjonalnie do częstotliwości pracy układu

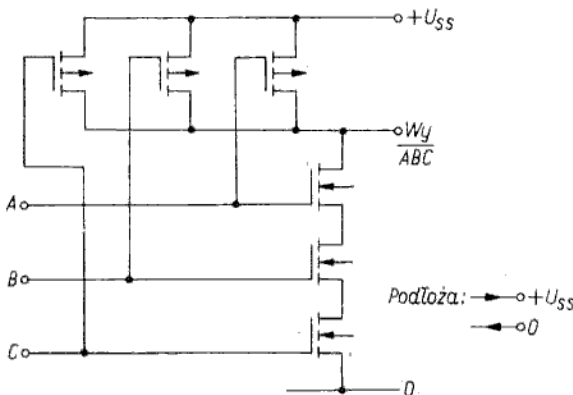
$$P = P_s + CU_{SS}^2 f \tag{10.14}$$

przy czym P_s — znikomo mała moc tracona przy pracy spoczynkowej. Charakterystykę przenoszenia inwertera CMOS przedstawiono na rys. 10.30.



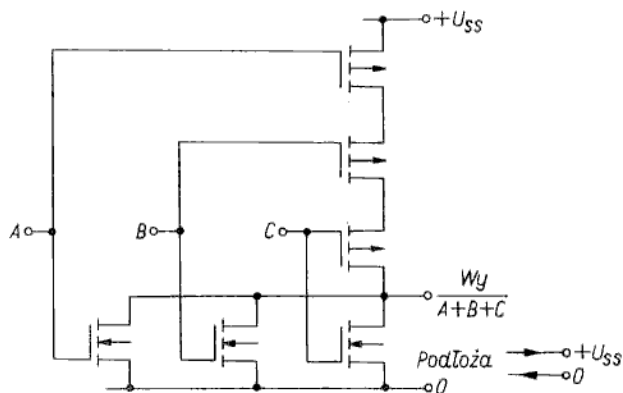
Rys. 10.30
Charakterystyka przenoszenia $U_{WY}(U_{WE})$ inwertera CMOS

Układy logiczne komplementarne wymagają większej liczby tranzystorów niż funkcjonalnie analogiczne układy z tranzystorami o jednakowym typie przewodności kanałów. Ogólnie można stwierdzić, że n -wejściowy funktor z tranzystorami o jednakowym typie przewodności kanałów składa się z $n + 1$ tranzystorów, natomiast n -wejściowy funktor komplementarny składa się z $2n$ tranzystorów,



Rys. 10.31
Trójwejściowy funktor NAND/CMOS

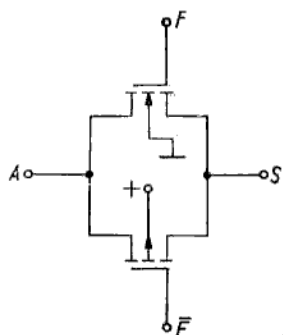
gdyż każde wejście jest przyłączone do oddzielnego inwertera komplementarnego. Stąd oczywisty jest fakt, że układy CMOS zajmują większą powierzchnię płytki. Należy również uwzględnić, że para komplementarna tranzystorów wymaga stosowania wyspy izolującej, dlatego zajmuje większą powierzchnię niż para tranzystorów o jednakowym typie przewodności kanałów. Na rysunku 10.31 przedstawiono funkcyj NAND trójwejściowy, na rys. 10.32 — funkcyj NOR trójwejściowy.



Rys. 10.32
Trójwejściowy funkcyj
NOR/CMOS

Liczba wejść w układzie NAND nie jest ograniczona wzrastającym poziomem sygnału „0”, jak to było w przypadku funkcyj NAND z tranzystorami o jednakowym typie przewodności kanałów, gdyż napięcie „0” w inwerterze komplementarnym jest bardzo małe. W tym przypadku ograniczenie liczby wejść jest spowodowane wzrastającą rezystancją szeregowo połączonych tranzystorów, co powoduje wzrost czasu włączania związanego z większą stałą czasową procesu rozładowania pojemności obciążającej.

W układach scalonych CMOS oprócz funkcyj NAND, NOR bardzo ważnym elementem logicznym jest bramka służąca do dwukierunkowej międzystopniowej transmisji



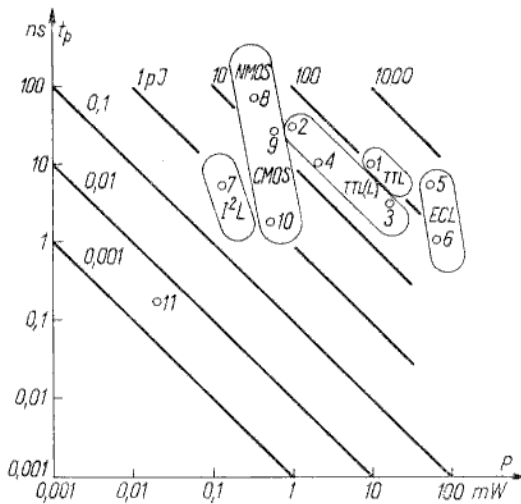
Rys. 10.33
Dwukierunkowa bramka transmisyjna CMOS

sygnałów logicznych (rys. 10.33). Oba tranzystory, sterowane sygnałami F , \bar{F} , są na przemian włączane lub wyłączane, tworząc między zaciskami A , S klucz zwarty lub rozarty.

Porównanie bramek logicznych

10.4

Jak już stwierdzono wcześniej, podstawowym parametrem, według którego przyjęto porównywać różne bramki logiczne, jest iloczyn czasu propagacji przez straty mocy w pojedynczej bramce (wskaźnik energetyczny). Na rysunku 10.34



Rys. 10.34

Orientacyjne obszary zajmowane przez poszczególne rodziny układów logicznych w płaszczyźnie o współrzędnych: czas propagacji t_p , moc strat P

1 — seria standardowa TTL; 2 — T1TL (seria małej mocy SN 54/74 L); 3 — TTLs; 4 — TTLs małej mocy; 5 — ECL (seria MECL-II); 6 — ECL (seria MECL-III); 7 — I²L; 8 — NMOS; 9 — CMOS ($U_{SS} = 10$ V, $f = 1$ MHz, $C_L = 10$ pF); 10 — CMOS w technologii SOS ($U_{SS} = 5$ V, $f = 50$ MHz, obciążenie identyczną bramką); 11 — element logiczny ze złączami Josephsona

przedstawiono orientacyjnie obszary zajmowane przez poszczególne rodziny układów logicznych w płaszczyźnie o współrzędnych: czas propagacji t_p , moc strat P . Z tego wykresu wynika, że pod względem wartości wskaźnika energetycznego współczesne techniki realizacji układów logicznych można uszeregować od najgorszej do najlepszej w następujący sposób:

- ECL,
- TTL,
- MOS,
- I²L.

Na rysunku 10.34 zaznaczono dla porównania punkt odpowiadający rekordowemu osiągnięciu współczesnej nauki, jakim jest wskaźnik energetyczny ok. 10^{-15} J, uzyskany w laboratoryjnych układach ze złączami Josephsona (złącza z nadprzewodników, działające w bardzo niskich temperaturach).

Jest oczywiste, że wartość wskaźnika energetycznego nie wystarcza do pełnego porównania poszczególnych rodzin układów logicznych. W zależności od zakresu zastosowań większa lub mniejsza waga może być przywiązywana do różnych parametrów. Na przykład za najbardziej istotną może być uznana jedna lub kilka następujących właściwości:

- duża szybkość działania,
- mały pobór mocy,
- mała powierzchnia płytki zajmowana przez bramkę logiczną,
- duże marginesy zakłóceń.

Największe szybkości działania uzyskuje się w układach wykonanych techniką ECL.

Najmniejszy pobór mocy mają układy MOS (szczególnie CMOS) oraz I²L.

Najmniejsze powierzchnie struktury bramki logicznej, a więc największe gęstości upakowania i największy stopień scalenia uzyskuje się w technice MOS oraz I²L.

Największe marginesy zakłóceń uzyskuje się w technice MOS (szczególnie CMOS).

Określona hierarchia ważności poszczególnych parametrów wynika z założonych właściwości konkretnego układu cyfrowego.

Układy o małym i średnim stopniu scalenia są wytwarzane głównie w technice TTL zapewniającej uzyskanie dużej szybkości działania przy niezbyt dużych stratach mocy, przy czym największe znaczenie ma rodzina układów TTLS.

Układy o dużym stopniu scalenia są wytwarzane głównie w technice NMOS.

Układy bardzo szybkie (o małym, średnim i dużym stopniu scalenia) są wytwarzane w technice ECL.

Układy bardzo odporne na zakłócenia lub o bardzo małym poborze mocy (o małym, średnim i dużym stopniu scalenia) są wytwarzane w technice CMOS.

Złożone układy logiczne

10.5

We wstępnych uwagach do tego rozdziału stwierdzono, że układy cyfrowe można podzielić na dwie grupy (tabl. 10.1):

- układy kombinacyjne,
- układy sekwencyjne.

W układach kombinacyjnych, nazywanych również *nieregeneracyjnymi*, stany logiczne na końcówkach wyjściowych zależą tylko od bieżących stanów na końcówkach wejściowych.

W układach sekwencyjnych, nazywanych również *regeneracyjnymi*, stany logiczne na końcówkach wyjściowych zależą nie tylko od bieżących stanów na końcówkach wejściowych, lecz także od stanów będących uprzednio.

Wszystkie złożone systemy cyfrowe mogą być w zasadzie tworzone z podstawowych bramek logicznych NAND lub (i) NOR, omawianych w poprzednich punktach. Jednakże asortyment układów scalonych cyfrowych jest znacznie szerszy, gdyż oprócz funkcyj podstawowych wytwarza się układy realizujące bardziej złożone funkcje logiczne. Przykłady układów o małym, średnim i dużym stopniu scalenia, wytwarzanych w postaci scalonej, podano w tabl. 10.1. Struktura logiczna dowolnego układu kombinacyjnego zawiera tylko funkcyj podstawowe. Zagadnienia projektowania i analizy struktury logicznej nie mieszczą się w ramach naszych zainteresowań, dlatego układy te nie będą omawiane. Natomiast układy sekwencyjne są zbudowane z komórek pamięci, które z kolei mogą być utworzone z podstawowych funkcyj logicznych lub też stanowią inne rozwiązania, specyficzne tylko dla układów sekwencyjnych. Nie będą omawiane układy sekwencyjne, poddające się dekompozycji strukturalnej na funkcyj podstawowe (dotyczy to przerzutników RS, JK, D, T itp. oraz układów złożonych z tych przerzutników). Natomiast rozpatrzmy te układy sekwencyjne (pamięci), których komórki elementarne mają specyficzne właściwości pamięciowe i nie dają się sprowadzić do połączeń funkcyj podstawowych.

Pamięci półprzewodnikowe

10.6

Pamięcią półprzewodnikową nazywa się półprzewodnikowy układ scalony, przeznaczony do przechowywania informacji zakodowanej w systemie dwójkowym. Struktura (mikropłytką) scalonej pamięci półprzewodnikowej składa się z matrycy komórek pamięci oraz elementów peryferyjnych i buforowych, umożliwiających współpracę pamięci właściwej (matrycy) z innymi układami wchodzącymi w skład systemu cyfrowego. Pamięci półprzewodnikowe można różnie klasyfikować w zależności od przyjętych kryteriów podziału.

Ze względu na sposób przechowywania i dostępu do informacji można wyszczególnić następujące rodzaje pamięci:

- o dostępie szeregowym (sekwencyjne), zbudowane z rejestrów przesuwających (ang. *shift register*) (np. pamięci FIFO — ang. *First In, First Out*);
- o dostępie swobodnym (RAM — ang. *Random Access Memory*);
- stałe (ROM — ang. *Read Only Memory*).

Ze względu na rodzaj tranzystorów, z jakich są zbudowane pamięci, można je podzielić na bipolarne i unipolarne (MIS, MOS). Często mówi się również o pamięciach ulotnych i nieulotnych, przy czym pamięcią ulotną jest taka pamięć, w której chwilowe odłączenie źródła zasilania powoduje skasowanie informacji w niej zawartej.

Ze względu na sposób zasilania układów pamięciowych (stałe lub impulsowe) dzieli się je na:

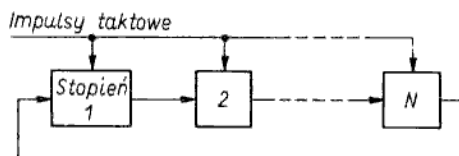
- statyczne,
- dynamiczne.

W pamięciach statycznych funkcję komórki pamięci spełnia przerzutnik zbudowany z tranzystorów bipolarnych lub MOS. Tych układów nie będziemy rozpatrywać, gdyż budowa i zasady działania przerzutników są szeroko opisane w literaturze z zakresu układów elektronicznych. Będą natomiast omówione niektóre komórki dynamiczne rejestrów przesuwających, pamięci RAM oraz elementy pamięci ROM.

Komórki dynamiczne rejestrów przesuwających

10.6.1

Rejestr przesuwający jest w istocie elektroniczną linią opóźniającą. Jak pokazano na rys. 10.35, jest to łańcuch elementarnych komórek, z których każda zawiera 1 bit informacji. Pod wpływem sterowania impulsami taktowymi informacja



Rys. 10.35
Rejestr przesuwający

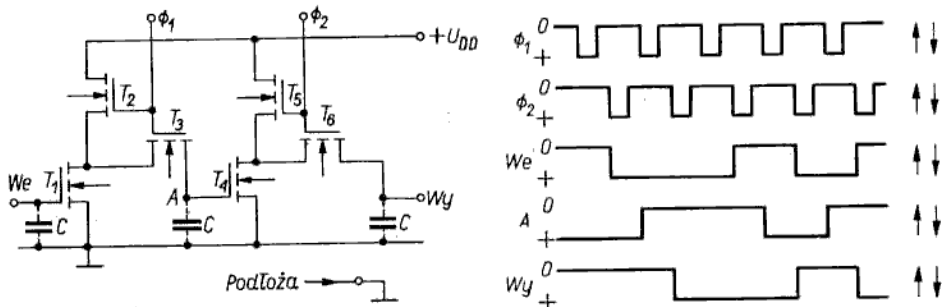
jest szeregowo przenoszona przez wszystkie komórki od wejścia do wyjścia. W przypadku połączenia wyjścia z wejściem informacja zawarta w rejestrze przesuwającym cyrkuluje z szybkością określoną przez częstotliwość powtarzania impulsów taktowych.

Rejestry przesuwające dzieli się na *statyczne* i *dynamiczne*.

W rejestrach statycznych komórkę pamięci, czyli pojedyncze ogniwo rejestru, stanowi przerzutnik tranzystorowy.

Rejestr przesuwający dynamiczny składa się z komórek opóźniających (elementarnych układów logicznych dynamicznych). Znane są dwu-, cztero- i sześciotaktowe układy logiczne dynamiczne, z których dwa pierwsze znalazły szerokie zastosowanie. W zależności od rodzaju stosowanych inwerterów układy logiczne dynamiczne dzieli się na *ilorazowe* i *nieilorazowe*. W pierwszych napięcie „0” logicznego na wyjściu jest wyznaczone ilorazem rezystancji r_{DS} przewodzących tranzystorów obciążającego i sterowanego, dlatego współczynnik $\beta_r = \beta^i / \beta^d$ musi być znacznie większy od jedności. Natomiast w drugich współczynnik $\beta_r = 1$, a napięcie „0” logicznego na wyjściu ustala się w dwu etapach, tj. przykładowo najpierw (przy przewodzących obu tranzystorach) pojemność wyjściowa ładuje się do poziomu 0,5 wartości napięcia zasilania, po czym tranzystor obciążający przechodzi do stanu nieprzewodzenia, a pojemność wyjściowa rozładowuje się przez przewodzący tranzystor sterowany do zera. Układy dwutaktowe mogą być ilorazowe lub nieilorazowe, natomiast układy czterotaktowe są wyłącznie nieilorazowe.

Jako przykład zostanie rozpatrzona praca układów logicznych dwutaktowych ilorazowych. Na rysunku 10.36 przedstawiono elementarny stopień rejestru przesuwającego



Rys. 10.36

Jednobitowa komórka opóźniająca rejestru przesuwającego, pracująca w układzie dwutaktowym ilorazowym

wającego (jednobitowa komórka opóźniająca), składający się z dwu inwerterów (T_1 , T_2 oraz T_4 , T_5) oraz dwu bramek transymisyjnych (tranzystory T_3 , T_6).

Zostanie rozpatrzona praca układu na podstawie wykresów przebiegów napięć (rys. 10.36). Impuls taktowy ϕ_1 powoduje włączenie tranzystorów T_2 , T_3 . Jeśli na bramkę tranzystora T_1 (wejście) jest podany sygnał logiczny „0”, to tranzystor T_1 nie przewodzi i w czasie trwania impulsu taktowego ϕ_1 pojemność wejściowa tranzystora T_4 ładuje się przez tranzystory T_2 , T_3 do wartości napięcia w przybliżeniu równej U_{DD} , co odpowiada poziomowi logicznemu „1”. Jeśli na bramkę tranzystora T_1 jest podany sygnał logiczny „1”, to tranzystor T_1 przewodzi i w czasie trwania impulsu taktowego ϕ_1 pojemność wejściowa tranzystora T_4 rozładowuje się przez tranzystory T_3 , T_1 do zerowej wartości napięcia, tj. w punkcie A ustala się poziom logiczny „0”. W czasie trwania impulsu taktowego ϕ_2 na wyjściu jest zapamiętywany stan logiczny inwersyjny w stosunku do stanu w punkcie A, tj. zgodny z poziomem logicznym na wejściu

układu. Najistotniejszą właściwością układu jest przeniesienie informacji z wejścia do wyjścia po dwu kolejnych impulsach taktowych ϕ_1, ϕ_2 . Dolna granica częstotliwości pracy układu zależy od stałej czasowej procesu rozładowania pojemności wejściowej tranzystora T_4 przez rezystancję wejściową tego tranzystora oraz rezystancję $r_{DS(off)}$ nieprzewodzącego tranzystora sprzęgającego T_3 . Ponieważ $r_{WE} \approx 10^{14} \dots 10^{15} \Omega$, a $r_{DS(off)} \approx 10^9 \dots 10^{10} \Omega$, więc decydujący wpływ ma składowa prądu rozładowania przez tranzystor T_3 ; stała czasowa tego procesu przy $C = 1 \text{ pF}$ wynosi ok. 10^{-3} s . Jeśli przyjmie się, że dopuszczalna jest 10-procentowa zmiana napięcia na pojemności C , to minimalna częstotliwość pracy układu będzie wynosić ok. 10 kHz.

Standardowe rejestry przesuwające, zbudowane z takich komórek opóźniających, pracują w zakresie częstotliwości 10 kHz ... 10 MHz.

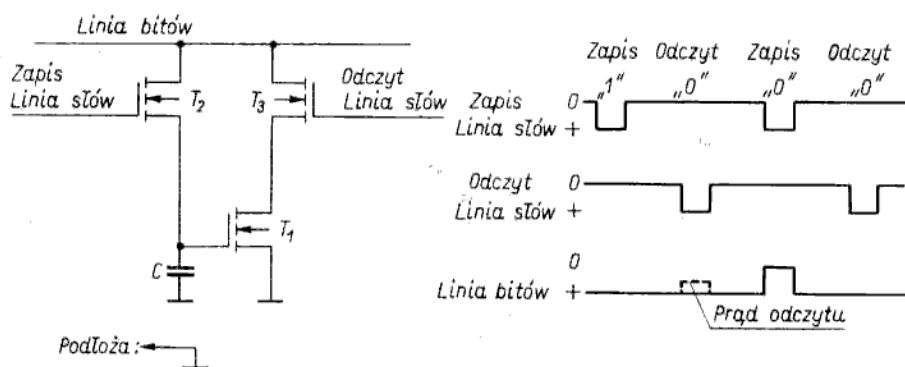
Komórki dynamiczne pamięci RAM

10.6.2

Pamięć półprzewodnikowa o dostępie swobodnym (RAM) jest taką pamięcią, w której dostęp do dowolnej komórki pamięciowej jest jednakowo łatwy i odbywa się w jednakowym czasie. (W pamięci sekwencyjnej czas dostępu do określonego bitu informacji zależy od jego położenia względnego w danej chwili czasu). W tym sensie pamięć RAM jest pojęciem bardzo ogólnym, obejmującym również pamięci stałe (ROM). Przyjęło się jednak stosować nazwę RAM tylko do tych pamięci o dostępie swobodnym, w których czynności zapisu i odczytu informacji są równoprawne, tj. jednakowo łatwe i niemal jednakowo szybkie. Zatem pojęcie pamięci RAM tradycyjnie utrwalilo się w znaczeniu obejmującym tylko pamięci o działaniu odczyt/zapis (niekiedy stosuje się nazwy R/W lub R/W RAM; R — read, W — write).

Pamięci RAM dzieli się na *statyczne* i *dynamiczne* (bipolarne i unipolarne). W pierwszych funkcje komórek pamięciowych spełniają przerzutniki, a w drugich — komórki dynamiczne.

Najczęściej są stosowane trój- lub jednotranzystorowe komórki dynamiczne pamięci RAM. Jeden z kilku możliwych wariantów komórki trójtranzystorowej przedstawiono na rys. 10.37.



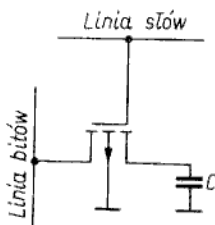
Rys. 10.37

Trójtranzystorowa komórka pamięci dynamicznej RAM/MOS

Informacja jest przechowywana w postaci ładunku zmagazynowanego w pojemności wejściowej tranzystora T_1 . W obecności dużego dodatniego ładunku w bramce, co odpowiada stanowi logicznemu „1”, tranzystor T_1 przewodzi, brak natomiast tego ładunku oznacza stan logiczny „0” i tranzystor T_1 nie przewodzi. Tranzystor T_2 służy do wprowadzenia informacji, tj. przeładowywania pojemności C , tranzystor T_3 zaś do odczytu informacji. Podczas zapisu, np. stanu „1”, do linii bitów jest doprowadzony potencjał dodatni, a impuls dodatni, dochodzący do bramki tranzystora T_2 zapisującą linią słów, powoduje przejście tranzystora T_2 do stanu przewodzenia, tj. podłączenie pojemności C do potencjału dodatniego linii bitów. W czasie odczytu odczytującą linią słów jest podawany impuls dodatni, otwierający tranzystor T_3 i np. dla stanu logicznego „1” przez linię bitów i przewodzące tranzystory T_3 , T_1 płynie prąd odczytu. Odświeżanie informacji odbywa się przez podanie w obwodzie sprzężenia zwrotnego odczytywanej i wzmacnionej informacji. Częstotliwość impulsów odświeżających jest zwykle większa niż kilkaset herców.

Typowa pamięć dynamiczna RAM zbudowana z elementów trójtranzystorowych ma pojemność 4096 bitów (jeden układ scalony), czas dostępu 200...350 ns oraz pobór mocy kilkadziesiąt mikrowatów na bit.

Najprostszym elementem pamięci dynamicznej jest element jednotranzystorowy, w którym informacja jest przechowywana w postaci ładunku zmagazynowanego w kondensatorze (rys. 10.38). Informacja jest zapisywana przez podanie odpowiedniego



Rys. 10.38
Jednotranzystorowa komórka pamięci dynamicznej RAM/MOS

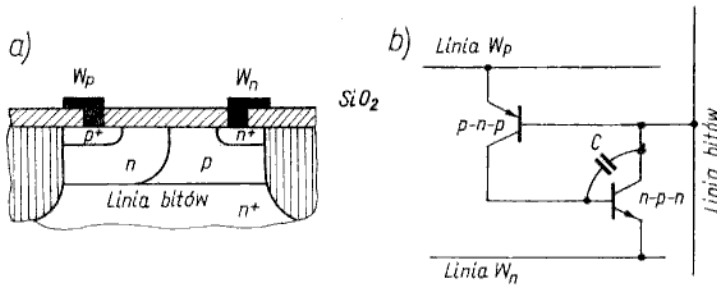
potencjału na linię bitów oraz włączenie tranzystora impulsem dodatnim, doprowadzonym linią słów do bramki tranzystora. Wówczas kondensator C ładuje się do potencjału równego potencjałowi linii bitów. Odczyt jest równoznaczny z przepływem ładunku z kondensatora do linii bitów przy przewodzącym tranzystorze. Wadą tego układu jest „wymazanie” informacji w czasie odczytu. Dlatego informacja jest odświeżana w pętli sprzężenia zwrotnego, zawierającej wzmacniacz.

Pierwszą pamięcią dynamiczną MOS, która zyskała szerokie rozpowszechnienie, była pamięć typu 1103 produkowana przez firmę Intel od początku lat siedemdziesiątych.

Typowe pamięci dynamiczne MOS/RAM, wytwarzane w latach 1976, 1977 miały pojemność 4096 lub 16 384 bitów i były wykonane najczęściej z elementów jednotranzystorowych w technologii NMOS w trzech wariantach:

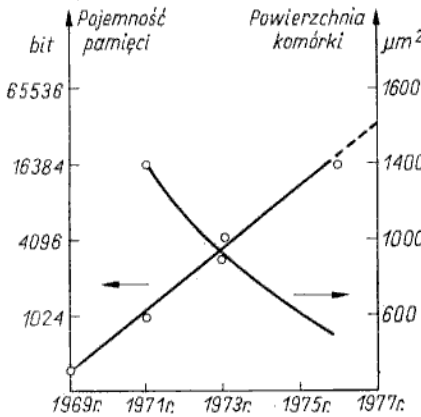
- z pojedynczą lub podwójną warstwą krzemu polikrystalicznego,
- VMOS,
- DMOS (z podwójną dyfuzją).

Duże perspektywy rozwoju mają również pamięci dynamiczne RAM wytwarzane z elementów bipolarnych w technice I²L. Element takiej pamięci przedstawiono na rys. 10.39. Do elementu pamięciowego są doprowadzone dwie linie słów



Rys. 10.39
Element pamięci I²L

(W_n , W_p) oraz jedna linia bitów. Nośnikiem informacji jest ładunek zgromadzony w pojemności złącza baza-kolektor. Jeżeli pojemność jest naładowana, to linia bitów ma potencjał wysoki, interpretowany jako stan logiczny „0”, w przeciwnym przypadku (pojemność rozładowana) niski potencjał linii bitów jest interpretowany jako stan logiczny „1”.



Rys. 10.40
Dynamika rozwoju technologii pamięci półprzewodnikowych RAM w ostatnich kilku latach; na wykresach przedstawiono pojemność pamięci (liczbę bitów w układzie scalonym) i powierzchnię zajmowaną przez pojedynczą komórkę pamięci

Dynamikę rozwoju technologii pamięci półprzewodnikowych RAM ilustruje wykres na rys. 10.40, na którym przedstawiono wzrost stopnia scalenia i zmniejszenie się powierzchni komórki pamięci w ostatnich kilku latach.

Komórki pamięci stałych (ROM)

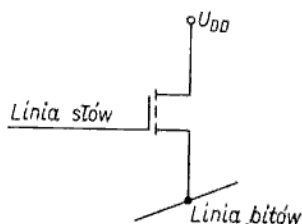
10.6.3

W pamięci stałej raz zapisana informacja dwójkowa jest trwale przechowywana przez bardzo długi czas i może być wielokrotnie odczytywana. Tego rodzaju pamięci są stosowane np. jako tablice funkcji, generatory znaków itp. Pod względem sposobu programowania pamięci stałe można ogólnie podzielić na dwie grupy:

- pamięci programowane w sposób trwały bez możliwości „wymazania” raz zapisanej informacji;

— pamięci programowane w sposób prawie trwały z możliwością „wymazania” i ponownego zapisu informacji (pamięci reprogramowane).

Pamięci stałe mają postać matrycy komórek pamięci. W przypadku pamięci programowanych w sposób trwały informacja dwójkowa jest zapisywana podczas procesu produkcyjnego. Na przykład w pamięci MOS/ROM wykorzystuje się wprost proporcjonalną zależność napięcia progowego od grubości warstwy dielektrycznej, znajdującej się pod powierzchnią elektrody bramki. Matryca takiej pamięci ma warstwę SiO_2 o zmiennej grubości, mianowicie cienki tlenek w tranzystorach „pamiętających” stan „0” oraz gruby tlenek w tranzystorach „pamiętających” stan „1”. Funkcję elementu pamięci spełnia pojedynczy tranzystor (rys. 10.41).



Rys. 10.41
Tranzystor MIS jako element pamięci stałej

W przypadku cienkiej warstwy tlenku impuls napięcia, doprowadzony linią słów, wprowadza tranzystor w stan przewodzenia, czemu odpowiada przepływ prądu odczytu w linii bitów. W przypadku grubej warstwy tlenku amplituda impulsu napięcia doprowadzonego linią słów jest mniejsza niż wartość napięcia progowego i tranzystor nie przewodzi. Zmiana zapisywanej informacji wymaga zmiany maski określającej inną lokalizację obszarów cienkiego i grubego tlenku. Inne rozwiązanie pamięci stałej bez możliwości „wymazania” raz zapisanej informacji polega na odłączeniu poszczególnych tranzystorów za pomocą mechanicznego przecinania lub przepalania prądem odpowiednio delikatnych połączeń.

W pamięciach stałych programowanych w sposób prawie trwały z możliwością zmiany zapisu informacji (tzw. pamięci PROM — ang. *Programmable Read Only Memory*) jest wykorzystywane zjawisko magazynowania ładunku w niejednorodnej warstwie dielektrycznej. Można je podzielić na dwa rodzaje:

- programowane metodą elektryczną, kasowane metodą nieelektryczną przez naświetlenie promieniami X (jeżeli obudowa jest nieprzezroczysta) lub promieniami nadfioletowymi (jeżeli w obudowie jest okienko kwarcowe). Elementami takiej pamięci są struktury FAMOS, omawiane w rozdziale 8;
- programowane i kasowane metodą elektryczną (elementami tych pamięci są tranzystory MNOS lub MAS).

W pamięci stałej MNOS mechanizm zapisu informacji jest następujący. Do bramek tranzystorów wybranych w matrycy zgodnie z programem zapisu informacji jest doprowadzone odpowiednio duże napięcie, które powoduje iniekcję tunelową nośników prądu z półprzewodnika przez cienką warstwę SiO_2 do pułapek zlokalizowanych na granicy warstw SiO_2 - Si_3N_4 . Ładunek zmagazynowany w pułapkach wywołuje odpowiednią zmianę wartości napięcia progowego. Informacja wpisana jest bardzo trwała, gdyż wprowadzany ładunek zmniejsza się do 50% początkowej wartości po czasie przekraczającym 10 lat. Odczyt wielokrotny nie wpływa destruktywnie na zapisaną informację. W każdej chwili natomiast można informację „wymazać”, powodując rekombinację nośników zmagazynowanych w pułapkach przez działanie napięciem bramki o odpowiedniej biegunowości

i wystarczająco dużej amplitudzie. Po wymazaniu pierwotnej informacji matryca jest gotowa do zapisu nowej informacji. Czas zapisu np. przy napięciu 40...50 V wynosi ponad 10 μ s.

Mechanizm zapisu i przechowywania informacji w komórkach pamięci MAS jest podobny.

Pamięci reprogramowane (FAMOS, MNOS, MAS) są często nazywane EAROM (ang. *Electrically Alterable Read-Only Memory*) lub RMM (ang. *Read Mostly Memory*)¹⁾.

Największe pojemności pamięci PROM, produkowanych w 1976 r., wynosiły 16 kilobitów. Jednakże najbardziej rozbudowanymi układami scalonymi MIS nie są pamięci PROM, lecz profesjonalne układy kompleksowe, tzw. *mikroprocesory*, których omówienie nie mieści się w tematycznych ramach tej książki.

Przyrządy o sprzężeniu ładunkowym

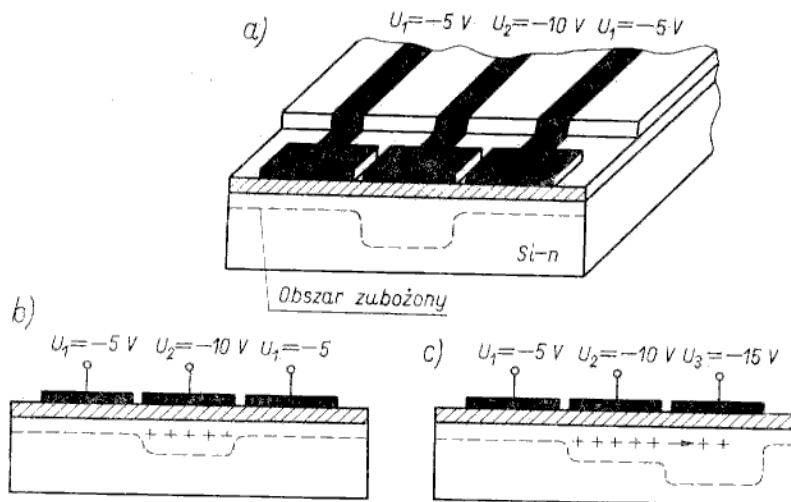
10.7

W ostatnich kilku latach (od 1970 r.) dużą uwagę skupił na sobie nowy typ monolitycznych układów funkcjonalnych, realizowanych w technice MOS, tzw. przyrządy o sprzężeniu ładunkowym (CTD — ang. *Charge Transfer Devices* lub *Charge Transport Devices*). Te niezwykle perspektywiczne przyrządy znajdują zastosowanie zarówno w układach cyfrowych (rejstry przesuwające) jak i analogowych (linie opóźniające, konwertery czasu, filtry, przetworniki obrazu). CTD dzieli się na dwie grupy:

- BBD (ang. *Bucket Brigade Devices*),
- CCD (ang. *Charge-Coupled Devices*),

przy czym ostatnio większe znaczenie przywiązuje się do przyrządów CCD.

CCD jest przyrządem funkcjonalnym, którego nie można zrealizować z elementów indywidualnych. Przyrząd działa na zasadzie magazynowania ładunków nośników



Rys. 10.42

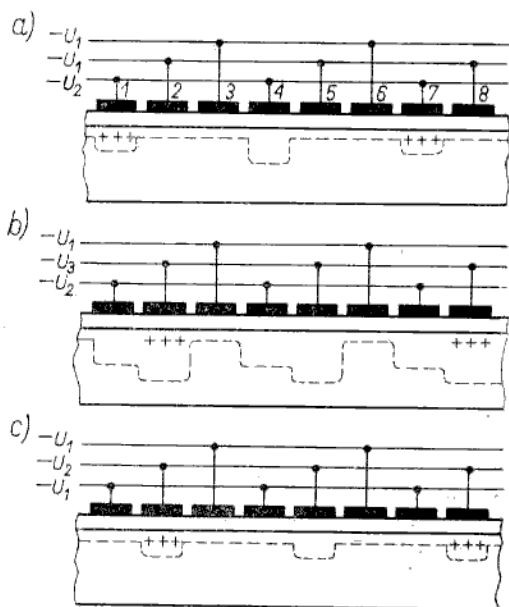
Podstawowa struktura trójelektrodowa CCD (a) oraz ilustracja dwóch faz jej pracy, tj. przechowywanie ładunku (b) i przesunięcie ładunku (c)

¹⁾ Często nazwa PROM jest stosowana w sensie pamięci ROM programowanych jednokrotnie przez użytkownika (pamięci z połączeniami przepalonymi prądem), natomiast nazwy EAROM lub RMM obejmują pamięci reprogramowane wielokrotnie (FAMOS, MNOS, MAS).

mniejszościowych w dołach potencjału, które przemieszczają się wraz z tymi ładunkami. W najprostszej postaci CCD składa się z elektrod metalowych naporowanych na powierzchni warstwy dielektryka osadzonej na jednorodnie domieszkowanym podłożu krzemowym (rys. 10.42).

Na rysunku 10.42 linią przerywaną oznaczono głębokość wnikania obszaru zubożonego w przypadku sterowania elektrod napięciami o różnych wartościach. W normalnych warunkach pracy CCD obszar przypowierzchniowy półprzewodnika znajduje się w stanie nierównowagowym, gdyż czas generacji nośników mniejszościowych jest dłuższy niż okres zmian potencjałów elektrod (współczesny stan technologii materiałów półprzewodnikowych umożliwia uzyskanie krzemu, w którym liczba objętościowych i powierzchniowych centrów rekombinacyjno-generacyjnych jest znikomo mała, a stała czasowa relaksacji nośników mniejszościowych wynosi kilka sekund). W tej sytuacji w obszarze przypowierzchniowym półprzewodnika nie powstaje warstwa inwersyjna (nie ma termicznie generowanych nośników mniejszościowych) mimo istnienia wartości chwilowych potencjałów elektrod, wystarczających do spowodowania inwersji w stanie równowagi termodynamicznej. Jeśli nośniki mniejszościowe zostaną wprowadzone do obszaru przypowierzchniowego półprzewodnika jednym ze sposobów opisanych w dalszej treści, to zgromadzą się one w dole potencjału, tj. pod elektrodą spolaryzowaną napięciem o największej wartości bezwzględnej (na rys. 10.42b — pod elektrodą spolaryzowaną napięciem -10 V). Obecność ładunku nośników mniejszościowych powoduje zmniejszenie głębokości wnikania obszaru zubożonego. Ładunek nośników mniejszościowych można przesunąć pod sąsiednią elektrodę (na rys. 10.42c pod elektrodę 3, jeżeli $|U_3| > |U_2|$).

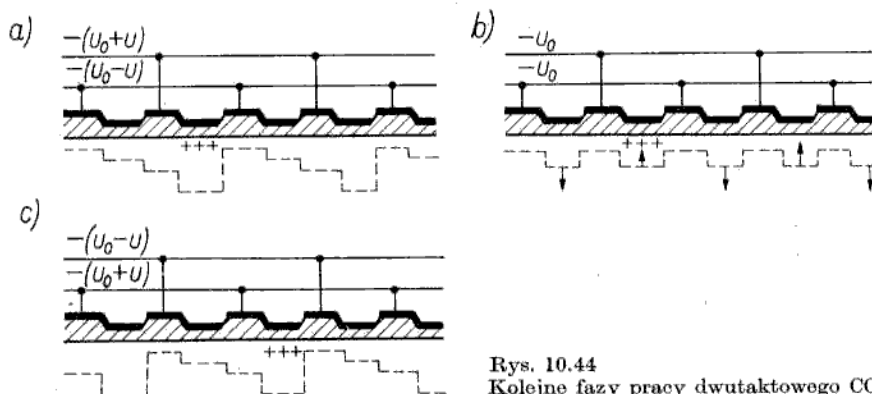
CCD może pracować w sekwencji trójtaktowej lub dwutaktowej. Pierwszy przypadek ilustruje rys. 10.43. Elektrody są sterowane równoległe trzema liniami doprowadzającymi impulsy taktowe o trzech różnych amplitudach napięcia $|U_1| < |U_2| < |U_3|$. Na rys. 10.43a ładunek jest magazynowany pod elektrodami



Rys. 10.43
Kolejne fazy pracy trójtaktowego
CCD

1, 7. W następnej fazie, przedstawionej na rys. 10.43b do elektrod 2, 5, 8 jest doprowadzone napięcie U_3 , wskutek czego ładunki elektrod 1, 7 przesuwają się pod elektrody 2, 8. Po kolejnej zmianie napięć na liniach zasilania (rys. 10.43c) otrzymuje się stan analogiczny do pokazanego na rys. 10.43a, lecz przesunięty w prawo o jedno miejsce.

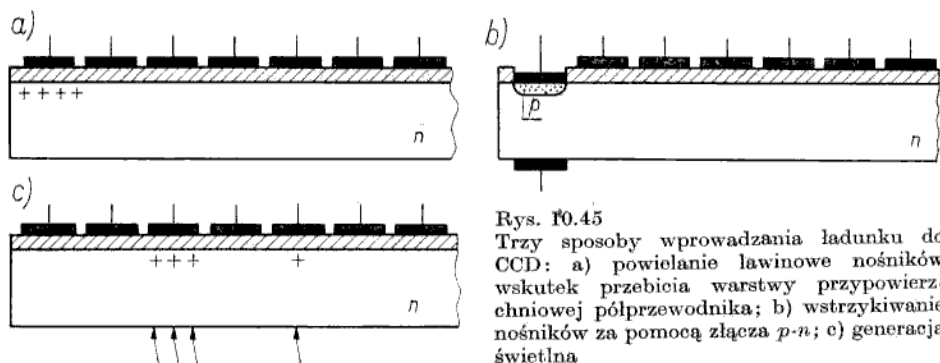
W podobny sposób można rozpatrywać pracę CCD zasilanego dwiema liniami, na których potencjał zmienia się skokowo od wartości $U_0 - U$ do $U_0 + U$ (rys. 10.44). W tym



Rys. 10.44
Kolejne fazy pracy dwutaktowego CCD

przypadku uproszczenie zasilania okupuje się bardziej złożoną technologią, gdyż elektrody muszą znajdować się na dwu różnych poziomach.

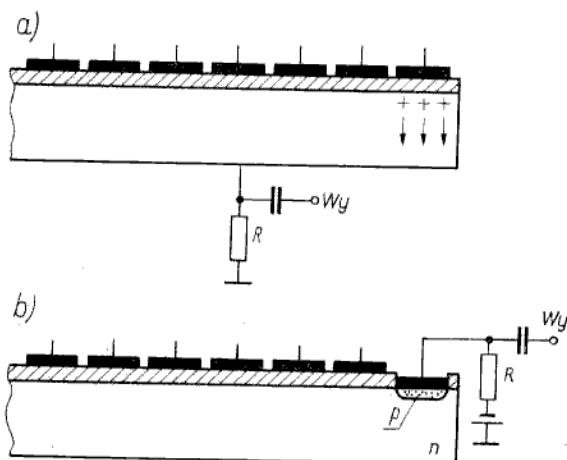
Na rysunku 10.45 pokazano następujące trzy sposoby wprowadzania informacji (ładunku) do CCD:



Rys. 10.45
Trzy sposoby wprowadzania ładunku do CCD: a) powielanie lawinowe nośników wskutek przebicia warstwy przypowierzchniowej półprzewodnika; b) wstrzykiwanie nośników za pomocą złącza p-n; c) generacja świetlna

- Przyłożenie dużego potencjału do pierwszej elektrody powoduje przebicie w obszarze przypowierzchniowym półprzewodnika, zatem generację lawinową ładunku nośników mniejszościowych.
- Obszar wdyfundowany typu p jest źródłem dziur wstrzykiwanych pod pierwszą elektrodę. Sygnał wejściowy może być modulowany napięciem na złączu p-n lub napięciem na pierwszej bramce spełniającej funkcję tzw. bramki wejściowej.
- Oświetlenie jest przyczyną generacji nośników mniejszościowych, magazynowanych następnie pod elektrodą. Ten ostatni sposób wprowadzania informacji

stanowi podstawę działania CCD jako przetworników obrazu. Przyrząd może być oświetlany „od góry” (wówczas światło pada na półprzewodnik w obszarach nie pokrytych przez metalizację lub również przez elektrody metalowe, wykonane jako półprzezroczyste) lub „od dołu”.



Rys. 10.46
Dwa sposoby odbioru ładunku z CCD: a) oddziaływanie pojemnościowe; b) złącze $p-n$ spolaryzowane zaporowo

Rysunek 10.46 ilustruje następujące dwa sposoby odbioru informacji (ładunku) na wyjściu CCD:

- Do ostatniej elektrody jest przykładany potencjał dodatni, powodujący wstrzymanie ładunku dodatniego, magazynowanego pod tą elektrodą, do podłoża. Ładunek ten, przechodząc do masy przez rezystor R , daje odpowiedni sygnał wyjściowy.
- Złącze $p-n$ wyjściowe jest spolaryzowane w kierunku zaporowym. Ładunek doprowadzony pod ostatnią elektrodę przechodzi przez złącze $p-n$, dając impuls prądu w obwodzie wyjściowym.

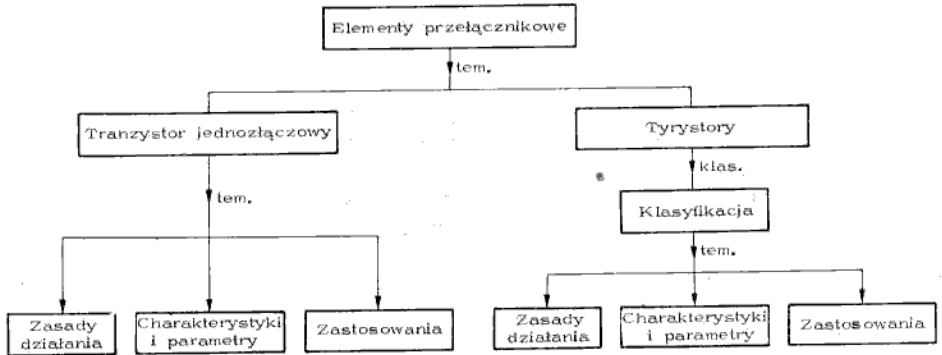
Podstawowym parametrem, charakteryzującym działanie CCD, jest *sprawność transportu ładunku* η_t , rozumiana jako stosunek ilości ładunku odbieranego na wyjściu przyrządu do początkowej ilości tego ładunku, wprowadzonej na wejście. Jest to tzw. całkowita sprawność transportu ładunku, której wartość jest zdeterminowana dwoma zjawiskami:

- skończonym czasem przelotu ładunku; przy określonym okresie zmian napięcia taktowego przeniesienie ładunku nie jest kompletne;
- rekombinacją oraz pułapkowaniem ładunku w stanach powierzchniowych.

W przypadku korzystania z rejestru przesuwającego typu CCD jako pamięci sekwencyjnej, tj. gdy informacja cyrkuluje w zamkniętej pętli (wyjście połączone z wejściem), jest konieczne uzupełnianie strat ładunku (odświeżanie informacji).

Pamięci CCD produkowane w 1977 r. miały maksymalną pojemność 65 kilobitów i wymagały odświeżania informacji po jej przejściu przez 256 komórek.

Jedną z podstawowych funkcji układowych elementów półprzewodnikowych jest przełączanie. Element spełniający funkcję przełącznika (klucza) pracuje na przemian w jednym z dwóch stanów, nazywanych *blokowaniem* (bardzo duża rezystancja elementu — klucz rozarty) i *przewodzeniem* (bardzo mała rezystancja elementu — klucz zwarty). Taką funkcję może spełniać tranzystor bipolarny, przełączany ze stanu odcięcia do nasycenia, lub tranzystor polowy, przełączany ze stanu zatkania do przewodzenia. W tym rozdziale są rozpatrywane elementy predestynowane wyłącznie lub głównie do pracy przełącznikowej. Są to tranzystory jednozłączowe oraz tyrystory. Przedstawiono zasadę działania, podstawowe charakterystyki i parametry oraz przykłady zastosowań tych elementów. Układ rozdziału przedstawiono schematycznie na rys. 11.1.



Rys. 11.1
Schematyczny układ rozdziału

Tranzystor jednozłączowy

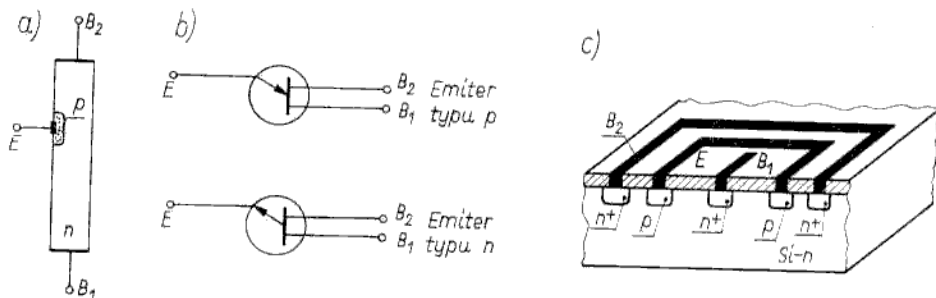
11.1

Tranzystor jednozłączowy (ang. *Unijunction Transistor* — UJT), nazywany również *diodą dwubazową* (ang. *Double-Base Diode*) jest trójkońcówkowym przyrządem przełączającym, działającym na zasadzie modulacji konduktywności półprzewodnika. Jest to element o charakterystyce prądowo-napięciowej typu S, w którym następuje przełączenie ze stanu nieprzewodzenia do stanu przewodzenia przy pewnej krytycznej wartości napięcia, nazywanej *napięciem przełączania*. Istnieje możliwość regulacji wartości napięcia przełączania za pomocą niezależnego źródła napięcia polaryzacji. Rozpatrzmy kolejno budowę i zasadę działania tranzystora jednozłączowego, jego podstawowe charakterystyki i parametry oraz zastosowania.

Budowa i zasada działania

11.1.1

Uproszczoną strukturę fizyczną tranzystora jednozłączowego przedstawiono na rys. 11.2a. Jest to bryła (belka) półprzewodnika typu n , wyposażona w kontakty omowe na obu końcach (bazy B_1, B_2) i mająca złącze p - n , przy czym

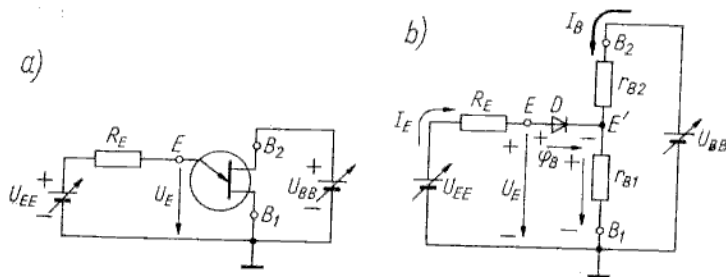


Rys. 11.2

Uproszczone szkice struktury tranzystora jednozłączowego (a), jego symbole graficzne (b), struktura planarna (c)

niewielki obszar typu p spełnia funkcję emitera. Odległość między bazą B_1 a złączem emiterowym wynosi ok. 0,5...0,7 odległości między bazami B_1, B_2 . Na rysunku 11.2b przedstawiono symbole graficzne tranzystorów jednozłączowych dla przypadku emiterów typu p i typu n . Strukturę planarną tranzystora z emiterem typu p przedstawiono na rys. 11.2c.

Podstawowy układ włączenia tranzystora jednozłączowego przedstawia rys. 11.3a. Złącze emiter-baza pierwsza (B_1) jest polaryzowane w kierunku przewodzenia ze źródła

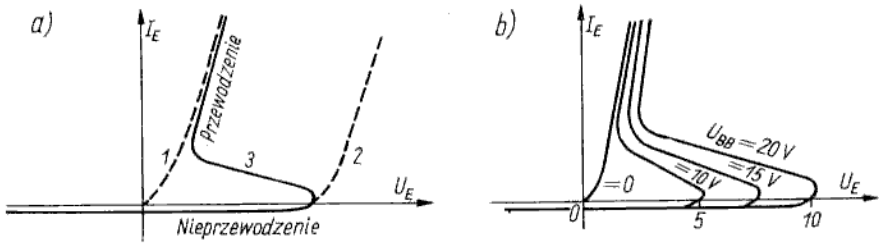


Rys. 11.3

Układ włączenia tranzystora jednozłączowego (a) i jego schemat zastępczy (b)

U_{EE} , przy czym wartość prądu jest ograniczona rezystorem szeregowym R_E . Obwód baza pierwsza (B_1)-baza druga (B_2) jest zasilany ze źródła U_{BB} , przy czym biegun dodatni tego źródła jest dołączony do końcówki B_2 .

Podstawową charakterystyką tranzystora jednozłączowego (rys. 11.4a) jest zależność prądu emitera I_E od napięcia emiter-baza pierwsza, oznaczonego jako U_E . W przypadku $U_{BB} = 0$, tj. przy zwartych obu bazach, zależność $I_E(U_E)$ jest typową charakterystyką prądowo-napięciową złącza p - n (krzywa I na rys.



Rys. 11.4
Charakterystyki tranzystora jednozłączowego

11.4a). Pracę tranzystora przy $U_{BB} \neq 0$ można łatwo wyjaśnić korzystając ze schematu zastępczego, przedstawionego na rys. 11.3b.

Tranzystor jednozłączowy jest reprezentowany przez diodę D i dwa rezystory r_{B1} , r_{B2} , odpowiadające odcinkom półprzewodnika typu n od emitera do bazy pierwszej oraz od emitera do bazy drugiej. Rezystancja półprzewodnika między dwiema bazami ($B_1 - B_2$)

$$r_{BB} = r_{B1} + r_{B2} \quad (11.1)$$

Zwykle r_{BB} ma wartość kilku kiloomów.

Pod wpływem napięcia U_{BB} w obwodzie $B_2 - B_1$ płynie prąd I_B , który powoduje odłożenie się spadku napięcia na rezystorze r_{B1}

$$U_{E'} = I_B r_{B1} = U_{BB} \frac{r_{B1}}{r_{B1} + r_{B2}} = \eta U_{BB} \quad (11.2)$$

przy czym $\eta = r_{B1} / (r_{B1} + r_{B2})$ — tzw. wewnętrzny współczynnik podziału lub wewnętrzny współczynnik blokowania (ang. *intrinsic stand-off ratio*); jego wartość w typowym przypadku wynosi ok. 0,6.

Napięcie $U_{E'}$ (w węźle E') polaryzuje dodatnio katodę diody D , która wskutek tego może przewodzić dla napięcia na anodzie (w węźle E) spełniającego warunek

$$U_E \geq U_{E'} + \varphi_B \quad (11.3)$$

przy czym $\varphi_B \approx 0,5 \dots 0,7$ V dla złącza krzemowego.

Rozpatrzmy teraz zależność $I_E(U_E)$ dla określonej wartości napięcia U_{BB} , przy której istnieje określone napięcie $U_{E'}$. Dla napięcia $U_E < U_{E'} + \varphi_B$ przez diodę D spolaryzowaną w kierunku zaporowym płynie bardzo mały prąd wsteczny, spadek napięcia na zewnętrznym rezystorze R_E jest zatem pomijalnie mały, czyli

$$U_{EE} \approx U_E$$

gdy napięcie U_{EE} osiągnie wartość większą niż $(U_{E'} + \varphi_B)$, przez diodę D zaczyna płynąć duży prąd przewodzenia. Przy stałej wartości rezystancji r_{B1} dalszy wzrost napięcia U_{EE} powodowałby wzrost prądu I_E zgodnie z krzywą 2 na rys. 11.4a. Jednakże rezystor r_{B1} zmienia swą wartość w funkcji prądu I_E wskutek zjawiska nazywanego *modulacją konduktywności*. Jeżeli konduktywność półprzewodnika typu n

$$\sigma_0 = q(\mu_n n_0 + \mu_p p_0) \quad \text{dla } I_E = 0 \quad (11.4)$$

to

$$\sigma = \sigma_0 + q(\mu_n \Delta n + \mu_p \Delta p) \quad \text{dla} \quad I_E > 0 \quad (11.5)$$

Dodatkowa koncentracja dziur Δp wstrzykiwanych z emitera powoduje również zwiększenie koncentracji elektronów o Δn (wskutek przyciągania elektrostatycznego).

Zgodnie ze wzorem (11.5) przyrosty koncentracji dziur i elektronów powodują zwiększenie konduktywności półprzewodnika w obszarze między emiterem a bazą pierwszą, zatem rezystancja r_{B1} maleje. Im większy jest prąd emitera, czyli im więcej nośników nadmiarowych wstrzykuje emiter, tym mniejsza jest rezystancja r_{B1} . Z kolei zmniejszenie rezystancji r_{B1} oznacza zgodnie ze wzorem (11.2) zmniejszenie współczynnika η i napięcia $U_{E'}$. To natomiast prowadzi do zwiększenia różnicy napięć $(U_E - U_{E'})$, czyli do zwiększenia prądu I_E . A większy prąd I_E to mniejsza rezystancja r_{B1} itd. Widać więc, że przy napięciu

$$U_{EE} \geq \eta U_{BB} + \varphi_B$$

rozpoczyna się proces regeneracyjny, przebiegający według schematu

zwiększenie prądu I_E
 \downarrow
 zmniejszenie rezystancji r_{B1}
 \downarrow
 zwiększenie różnicy napięć $(U_E - U_{E'})$ polaryzującej złącze $p-n$
 \downarrow
 zwiększenie prądu I_E
 itd.

Ponieważ zwiększenie prądu I_E przy stałej wartości U_{EE} powoduje zmniejszenie napięcia U_E , gdyż $U_E = U_{EE} - R_E I_E$, w rezultacie otrzymuje się charakterystykę $I_E(U_E)$ mającą odcinek o ujemnej rezystancji przyrostowej (krzywa 3 na rys. 11.4a). W tym obszarze charakterystyki prąd zwiększa się przy malejącym napięciu, przy czym napięcie można w przybliżeniu wyrazić równaniem

$$U_E = U_{BB} \frac{r_{B1}(I_E)}{r_{B2} + r_{B1}(I_E)} + I_E r_{B1}(I_E) + \varphi_B \quad (11.6)$$

Teoretycznie biorąc, najmniejsza wartość napięcia U_E przy przełączaniu tranzystora ze stanu nieprzewodzenia do stanu przewodzenia wynosi $U_E \approx \varphi_B$, jeżeli założy się, że r_{B1} może osiągnąć wartość zerową. Praktycznie napięcie U_E może zmaleć do wartości 1...3 V.

Na rysunku 11.4b przedstawiono rodzinę charakterystyk $I_E(U_E)$ dla różnych wartości napięcia U_{BB} , otrzymaną przy założeniu $\eta = 0,5$.

Podstawowe parametry

11.1.2

Właściwości tranzystora jednozłączowego są opisywane w katalogach następującymi parametrami granicznymi i charakterystycznymi.

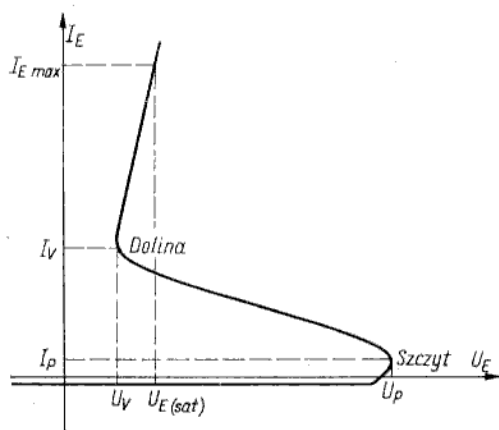
Dopuszczalne parametry graniczne:

- moc rozpraszana P_D ;
- maksymalny prąd emitera I_E ;

- maksymalne napięcie wsteczne złącza emiter-baza druga U_{B2E} ;
- maksymalne napięcie międzybazowe U_{B2B1} (zwykle oznaczane U_{BB});
- zakres temperaturowy T_a .

Parametry charakterystyczne:

- wewnętrzny współczynnik blokowania η , wyznaczany przy określonym napięciu U_{BB} ;
 - rezystancja międzybazowa r_{BB} , wyznaczana przy $U_{BB} = \text{const}$, $I_E = 0$;
 - napięcie nasycenia $U_{E(sat)}$, tj. napięcie emiter-baza pierwsza przy pracy w stanie przewodzenia dla $U_{BB} = \text{const}$, $I_E = \text{const}$ (zwykle dla wartości maksymalnej I_E);
 - prąd w punkcie szczytu I_P ;
 - prąd w punkcie doliny I_V ;
- Trzy ostatnie parametry ilustruje rys. 11.5.

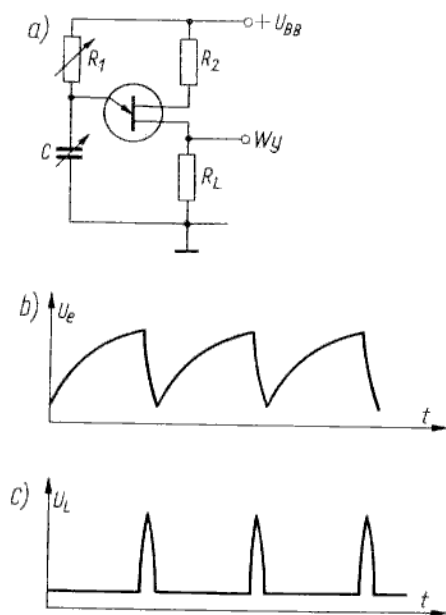


Rys. 11.5
Charakterystyka $I_E(U_E)$ z zaznaczonymi parametrami statycznymi

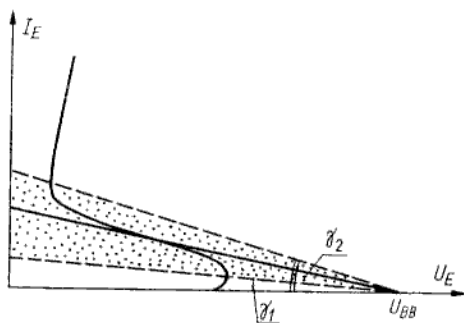
Zastosowania

11.1.3

Istnienie odcinka charakterystyki $I(U)$ o nachyleniu ujemnym umożliwia budowanie z tranzystora jednozłączonego prostych układów astabilnych. Typowy układ generatora relaksacyjnego przedstawiono na rys. 11.6 (jest to zresztą układ stosowany w obwodach sterowania tyrystorów). Jego zasada pracy jest następująca. Kondensator C ładuje się przez rezystor R_1 ze stałą czasową $T = CR_1$. W chwili gdy napięcie na kondensatorze osiągnie wartość równą napięciu przełączania tranzystora jednozłączonego, następuje rozładowanie kondensatora przez rezystor obciążenia R_L . Ponieważ stała czasowa rozładowania jest bardzo mała (rezystor R_L ma małą wartość rzędu kilkudziesięciu omów), na wyjściu otrzymuje się impuls szpilkowy. Po rozładowaniu kondensatora złącze $E-B_1$ uzyskuje polaryzację zaporową i powtarza się cykl ładowania kondensatora. Warunkiem powstania oscylacji jest ulokowanie punktu pracy na odcinku o ujemnej rezystancji przyrostowej. Na rysunku 11.7 zilustrowano graficznie sposób wyznaczania zakresu wartości rezystancji R_1 , przy których jest możliwa praca astabilna (przyjmuje się, że $R_1 \gg R_L$).



Rys. 11.6
Układ generatora z tranzystorem jednozłączowym (a) oraz przebiegi napięcia emitera (b) i napięcia na obciążeniu (c)



Rys. 11.7
Graficzna ilustracja granicznych wartości rezystancji obciążenia dla pracy astabilnej ($\text{tg } \gamma_1 = 1/R_{1\text{max}}$; $\text{tg } \gamma_2 = 1/R_{1\text{min}}$)

Tyristory

11.2

Pierwsze publikacje na temat tyristorów pojawiły się na początku lat pięćdziesiątych, produkcję tych elementów rozpoczęto w 1958 r., a obecnie są to najważniejsze elementy półprzewodnikowe w energoelektronice. Nazwa tyristor, wskazująca na analogię funkcjonalną tego elementu do tyratronu (przełącznikowa lampa gazowana), oznacza półprzewodnikowy element dwustanowy o trzech lub więcej złączach (czterech lub więcej warstwach), który może być przełączany ze stanu blokowania do stanu przewodzenia i odwrotnie. Istnieje wiele rodzajów tyristorów. Najważniejsze z nich sklasyfikowano w tabl. 11.1, przy czym za podstawę klasyfikacji przyjęto liczbę końcówek oraz przebieg charakterystyki prądowo-napięciowej w trzeciej ćwiartce wykresu $I(U)$. Bierze się pod uwagę następujące cechy tego przebiegu:

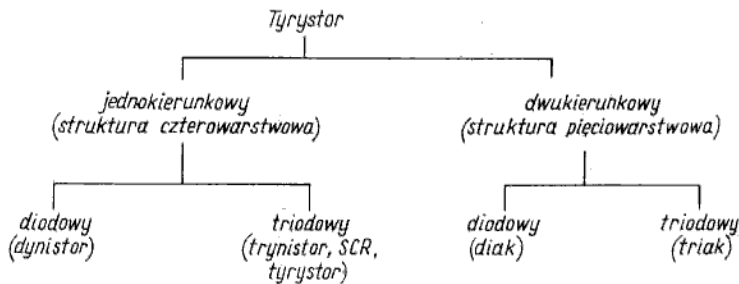
Rodzaje tyrystorów

Tablica 11.1

Liczba końcówek	Rodzaj pracy w III ćwiartce	Liczba warstw półprzewodnikowych	Stosowane nazwy	Symbole graficzne		Charakterystyka
				obowiązujące ¹⁾	spotykane w literaturze	
2	Blokowanie	4	Dynistor			
	Przełączanie	5	Diak			
3	Blokowanie	4	Tyrystor triodowy (ang. SCR)			
	Przełączanie	5	Triak			

¹⁾ Według normy PN-73/E-01206.

- stan zaporowy w III ćwiartce; tj. praca jednokierunkowa z możliwością przełączania tylko w I ćwiartce;
- przełączanie zarówno w I, jak i w III ćwiartce, tj. działanie dwukierunkowe.



Rys. 11.8
Klasyfikacja tyrystorów

Uwzględniając na pierwszym miejscu te właściwości, a na drugim — liczbę końcówek, rozróżnia się następujące cztery rodzaje tyrystorów (rys. 11.8):

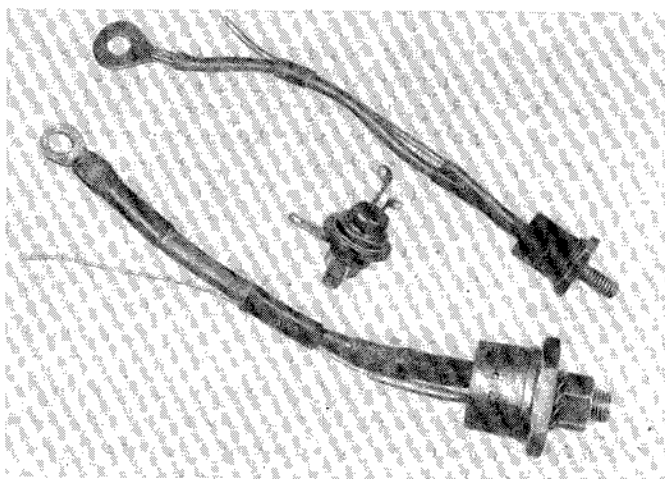
- dynistor, tj. jednokierunkowy element dwukońcówkowy;
- tyrystor triodowy, tj. jednokierunkowy element trójkońcówkowy;
- diak (ang. *diac*), tj. dwukierunkowy element dwukońcówkowy;
- triak (ang. *triac*), tj. dwukierunkowy element trójkońcówkowy.

Z tych czterech rodzajów tyrystorów najbardziej rozpowszechniony jest tyrystor triodowy. W literaturze angielskiej jest on nazywany *półprzewodnikowym zaworem sterowanym* (SCR — *Semiconductor Controlled Rectifier*), a w literaturze polskiej jest po prostu nazywany tyrystorem. Zatem nazwa tyrystor utrwaliła się tradycyjnie w wąskim znaczeniu, obejmującym tylko trójkońcówkowy element przełączający o działaniu jednokierunkowym. Będziemy jednak stosowali terminologię podaną na rys. 11.8, według której nazwa tyrystor ma szersze znaczenie.

Tyrystory o działaniu jednokierunkowym (struktury $p-n-p-n$) 11.2.1

Budowa i zasada działania 11.2.1.1

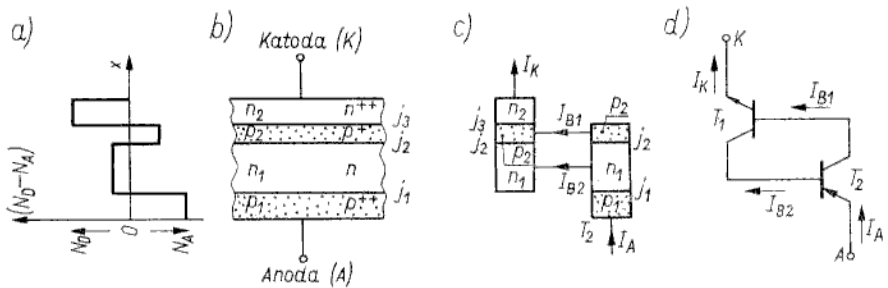
Widok zewnętrzny tyrystorów w typowych obudowach przedstawiono na rys. 11.9. Tyrystory o działaniu jednokierunkowym mają strukturę czterowarstwową (trójzłączową). Podstawową strukturę $p-n-p-n$ wraz z rozkładem koncentracji



Rys. 11.9
Typowe obudowy tyrystorów

domieszek i reprezentacją tranzystorową przedstawiono na rys. 11.10. Poszczególne warstwy mają celowo różne szerokości i różne koncentracje domieszek. Skrajne warstwy P oraz N nazywają się odpowiednio anodą (A) i katodą (K).

Najpierw rozpatrzmy działanie tyrystora jako elementu dwukońcówkowego. Strukturę $p-n-p-n$ można traktować jako połączenie dwóch tranzystorów typu $p-n-p$ oraz $n-p-n$ (rys. 11.10c, d). Wprowadźmy numerację warstw od anody do ka-



Rys. 11.10

Rozkład koncentracji domieszek (a) w strukturze czterowarstwowej (b) oraz reprezentacja tranzystorowa (c, d) tej struktury

tody: $p_1-n_1-p_2-n_2$. Warstwa P_1 jest emiterym tranzystora $p-n-p$, N_1 — bazą tranzystora $p-n-p$ i jednocześnie kolektorem tranzystora $n-p-n$, P_2 — kolektorem tranzystora $p-n-p$ i jednocześnie bazą tranzystora $n-p-n$, N_2 — emiterym tranzystora $n-p-n$. Zatem złącza j_1, j_3 spełniają funkcję złączy emiter-baza, natomiast j_2 jest złączem kolektorowym. W związku z tym dwie skrajne warstwy (P_1, N_2) można nazywać emiterymi, dwie wewnętrzne (N_1, P_2) — bazami, a złącze j_2 — kolektorem.

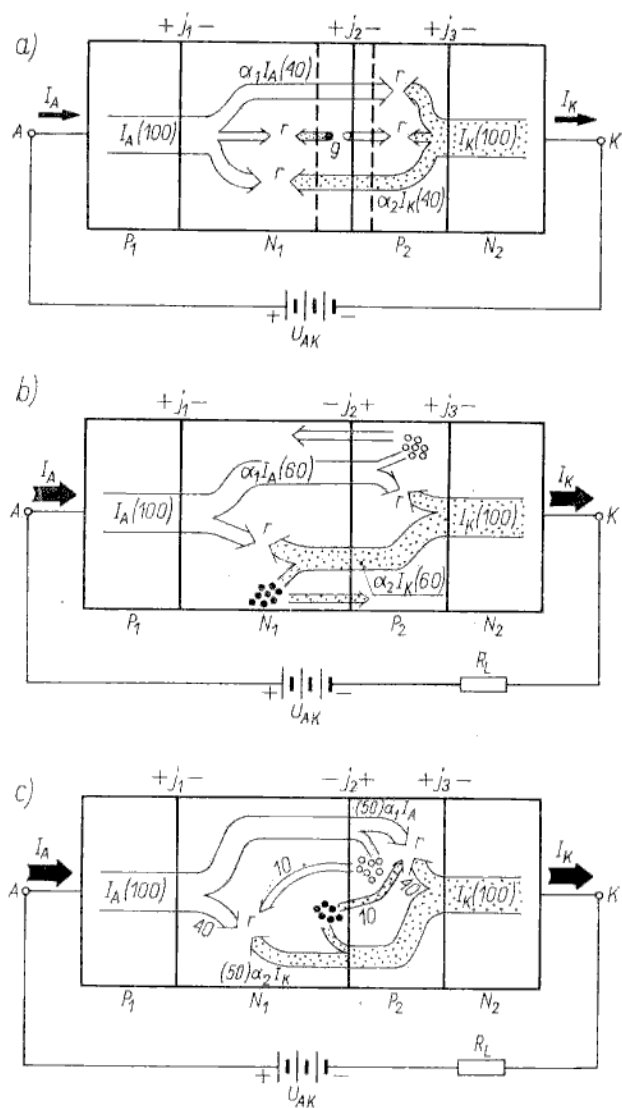
Wyróżnia się następujące trzy zakresy pracy tyrystora:

- zaporowy,
- blokowania,
- przewodzenia.

Stan zaporowy istnieje przy polaryzacji anody napięciem ujemnym względem katody.

Wówczas złącza j_1, j_3 są spolaryzowane w kierunku zaporowym, a złącze j_2 — w kierunku przewodzenia. Z uwagi na niesymetryczne właściwości złączy j_1, j_2 (różne profile domieszkowania) prawie całe napięcie zewnętrzne odkłada się na złączu j_1 , przy czym warstwa zaporowa tego złącza rozszerza się głównie w stronę słabo domieszkowanego obszaru N_1 . W związku z tym warstwa N_1 musi być szeroka, żeby nie nastąpiło przebicie skośne struktury $p_1-n_1-p_2$. Prąd płynący przez tyrystor w tym stanie pracy jest bardzo małym prądem wstecznym złączy j_1, j_3 . W miarę zwiększania napięcia przy pewnej jego wartości następuje przebicie złącza j_1 , po czym przebijają się również złącze j_3 , czyli charakterystyka tyrystora w zakresie zaporowym nie różni się od charakterystyki diody.

Przy polaryzacji anody napięciem dodatnim względem katody są możliwe dwa stany, tj.: blokowanie lub przewodzenie. Pierwszy przypadek (blokowanie) zachodzi wówczas, gdy suma współczynników wzmocnienia prądowego ($\alpha_1 + \alpha_2$) tranzystorów T_1, T_2 (rys. 11.10c, d) jest mniejsza od jedności. Zjawiska fizyczne zachodzące w strukturze $p-n-p-n$ w stanie blokowania wyjaśnia z dużym uproszczeniem rys. 11.11a. Złącza j_1, j_3 są spolaryzowane w kierunku przewodzenia, a złącze j_2 — w kierunku zaporowym. Z obu emiteryw są wstrzykiwane nośniki do baz, tj. z emitera P_1 — dziury do warstwy N_1 , a z emitera N_2 — elektrony do warstwy P_2 . Założmy chwilowo, że $\alpha_1 = \alpha_2$, a ponieważ $(\alpha_1 + \alpha_2) < 1$, więc $\alpha_1 < 0,5, \alpha_2 < 0,5$. Na rysunku 11.11a przedstawiono strumienie nośników dla przykładowo wziętych wartości $\alpha_1 = \alpha_2 = 0,4$. Zgodnie z przykładem liczbowym, podanym na tym rysunku, ze stu dziur wstrzykniętych z emitera P_1 do bazy N_1 czterdzieści przechodzi poza obszar bazy i jest odbierane przez



Rys. 11.11

Ilustracja działania tyristora w stanie blokowania (a) oraz przewodzenia (b, c)

złącze kolektorowe j₂ do warstwy P₂, a pozostałe sześćdziesiąt dziur rekombinuje z elektronami w bazie N₁. Identycznie można opisać strumień elektronów wstrzykiwanych z emitera N₂ do bazy P₂, przy czym ze stu elektronów wstrzykniętych czterdzieści odbiera kolektor do warstwy N₁. Te elektrony rekombinują z czterdziestoma dziurami, a pozostałe dwadzieścia dziur rekombinuje z elektronami generowanymi w warstwie zaporowej złącza j₂. Zatem ze stu dziur wstrzykniętych z emitera P₁ do bazy N₁ czterdzieści odbiera kolektor, inne czterdzieści rekombinuje z elektronami wstrzykiwanymi z emitera N₂ i odbieranymi przez

kolektor, a pozostałe dwadzieścia rekombinuje z elektronami generowanymi w warstwie zaporowej złącza j_2 . Analogicznie przedstawia się bilans strumienia elektronów transportowanych w kierunku przeciwnym (od katody do anody). Ciągłość prądu jest zachowana w dowolnym przekroju struktury, czyli prąd w obwodzie zewnętrznym (anodowy lub katodowy) jest równy sumie prądów składowych płynących przez złącze j_2

$$I_A = \alpha_1 I_A + \alpha_2 I_K + I_g \quad (11.7)$$

przy czym I_g — prąd generacji nośników w warstwie zaporowej złącza j_2 .
Ponieważ $I_A = I_K$, zatem

$$I_A = \frac{I_g}{1 - (\alpha_1 + \alpha_2)} \quad (11.8)$$

A więc w stanie blokowania przez tyrystor płynie prąd anodowy o małej wartości, porównywalnej co do rzędu wartości z prądem generacji (na przykład dla $\alpha_1 + \alpha_2 = 0,8$ prąd $I_A = 5I_g$).

Ze wzoru (11.8) wynika również, że przy $(\alpha_1 + \alpha_2) \rightarrow 1$ prąd anodowy dąży do nieskończoności, czyli tyrystor pracuje w stanie *przewodzenia*. Jakościowy obraz zjawisk w stanie przewodzenia daje rys. 11.11b, na którym przedstawiono strumienie nośników przy założeniu, że $\alpha_1 = \alpha_2 = 0,6$. Według wzoru (11.8) przypadek $(\alpha_1 + \alpha_2) > 1$ nie ma sensu fizycznego, gdyż wynika stąd ujemna wartość prądu anodowego, a to prowadzi do absurdalnego wniosku, że rezystancja statyczna tyrystora jest ujemna. Jednakże przyjmując chwilowo, że $\alpha_1 = \alpha_2 = 0,6$, wykażemy dalej, dlaczego to założenie nie jest słuszne. Zgodnie z rys. 11.11b ze stu dziur wstrzykniętych z emitera P_1 do bazy N_1 sześćdziesiąt odbiera kolektor do warstwy P_2 , a czterdzieści rekombinuje z elektronami w bazie N_1 . Identyczne relacje występują dla strumienia elektronów wstrzykiwanych z emitera N_2 w kierunku przeciwnym. Liczby elektronów i dziur dostarczanych do każdej bazy nie równoważą się zatem wzajemnie. Wskutek tego w bazie N_1 czterdzieści dziur rekombinuje z czterdziestoma elektronami i pozostaje nadmiar dwudziestu elektronów, a w bazie P_2 czterdzieści elektronów rekombinuje z czterdziestoma dziurami i pozostaje nadmiar dwudziestu dziur. W innym ujęciu można stwierdzić, że suma prądów dziurowego i elektronowego, płynących przez warstwę zaporową złącza j_2 ($60 + 60$) jest większa niż prąd zewnętrzny (100). Nie jest więc zachowana ciągłość prądu, co oznacza, że w obu bazach akumulują się ładunki. W bazie N_1 zbiera się ładunek ujemny elektronów, a w bazie P_2 — ładunek dodatni dziur. Są to ładunki przeciwnych znaków w stosunku do ładunków przestrzennych warstwy zaporowej złącza j_2 . W związku z tym elektrony akumulowane w bazie N_1 będą równoważyć ładunek dodatni centrów donorowych w warstwie zaporowej złącza j_2 , a dziury akumulowane w bazie P_2 będą równoważyć ładunek ujemny centrów akceptorowych w warstwie zaporowej tego złącza. W rezultacie szerokość warstwy zaporowej złącza j_2 i spadek napięcia na tym złączu będą się zmniejszać dopóty, dopóki nie zostanie ono spolaryzowane w kierunku przewodzenia. Wówczas dalszy wzrost ładunków nośników nadmiarowych w obu bazach zostanie zahamowany, gdyż złącze kolektorowe (j_2), spolaryzowane w kierunku przewodzenia, będzie nie tylko odbierać nośniki, ale również wstrzykiwać je z powrotem. Na przykład z sześćdziesięciu dziur odebranych z obszaru N_1 do P_2 dwadzieścia dziur występujących w nadmiarze (czterdzieści rekombinuje z elektronami) zostanie wstrzyknięte z powrotem do obszaru N_1).

Nie jest to jeszcze obraz w pełni prawidłowy, gdyż po pierwsze, suma prądów płynących przez złącze kolektora jest w tym przypadku mniejsza niż prąd anodowy

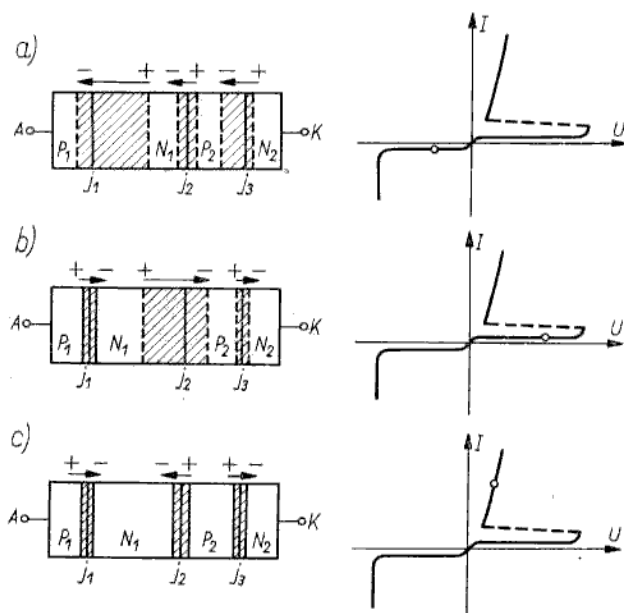
$$(60 - 20) + (60 - 20) < 100$$

a po drugie, nie wiadomo, w jakim miejscu obwodu odkłada się napięcie U_{AK} po utraceniu właściwości zaporowych przez złącze j_2 . Otóż należy wziąć pod uwagę, że nadmiarowe ładunki elektronów i dziur akumulowanych w bazach nie tylko zmniejszają napięcie na złączu j_2 (a w rezultacie końcowym zmieniają polaryzację tego złącza), ale również zwiększają napięcia polaryzujące złącza j_1, j_3 w kierunku przewodzenia. Wskutek tego zwiększają się strumienie nośników wstrzykiwanych przez oba emitery, czyli zwiększa się prąd anodowy (i równy mu prąd katodowy), a więc wzrasta spadek napięcia na rezystorze obciążenia R_L . Zatem zmniejszanie się napięcia na tyristorze oznacza jednocześnie zwiększanie prądu i niejako „przeniesienie” napięcia z tyristora na rezystor R_L .

W stanie ustalonym nie może być nieciągłości prądu, przykład liczbowy rozpatrywany na rys. 11.11b, należy zatem skorygować w sposób przedstawiony na rys. 11.11c. W tym przypadku sumaryczny prąd płynący przez złącze j_2

$$(60 - 10) + (60 - 10) = 100$$

czyli jest równy prądowi w obwodzie zewnętrznym. Efektywnie biorąc, tylko połowa (dokładnie połowa) każdego strumienia nośników wstrzykiwanych jest odbierana przez kolektor. Oznacza to, że współczynniki $\alpha_1 = \alpha_2 = 0,5$. Równość obu współczynników wzmocnienia prądowego założono tylko dla ułatwienia rozważań poglądowych. W rzeczywistości $\alpha_1 < \alpha_2$, gdyż baza N_1 jest bardzo szeroka. Warunek szczególny ($\alpha_1 = \alpha_2 = 0,5$) należy zastąpić warunkiem ogólnym: $(\alpha_1 + \alpha_2) = 1$.



Rys. 11.12

Rozkłady napięć na trzech złączach dla stanu: a) zaporowego; b) blokowania; c) przewodzenia

Na rysunku 11.12 przedstawiono charakterystykę prądowo-napięciową tyrystora z zaznaczeniem punktów odpowiadających trzem wyżej rozpatrywanym stanom. Na tym rysunku pokazano również schematycznie szerokości warstw zaporowych w poszczególnych zakresach pracy. Należy zauważyć, że w stanie przewodzenia spadek napięcia na tyrystorze jest równy sumie napięć na trzech złączach spolaryzowanych w kierunku przewodzenia (po ok. 1 V), przy czym napięcie na złączu wewnętrznym ma przeciwną biegunowość, a więc suma trzech napięć wynosi ok. 1...2 V.

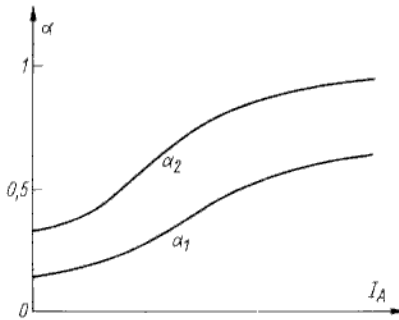
Dotychczas rozpatrywaliśmy działanie tyrystora w każdym z trzech stanów (zaporowy, blokowanie, przewodzenie), nie wiadomo jednak w dalszym ciągu, w jaki sposób następuje przejście ze stanu blokowania do przewodzenia, czyli ze stanu o dużej rezystancji do stanu o małej rezystancji. Wiadomo natomiast, że warunkiem blokowania jest

$$\alpha_1 + \alpha_2 < 1$$

a warunkiem przewodzenia

$$\alpha_1 + \alpha_2 = 1$$

Z teorii tranzystora bipolarnego wiadomo również, że współczynnik wzmocnienia prądowego w zakresie małych prądów emitera zwiększa się wraz ze wzrostem prądu emitera. Podobnie w tyrystorze suma współczynników ($\alpha_1 + \alpha_2$) zwiększa się wraz ze wzrostem prądu anodowego (katodowego), co przedstawiono na rys. 11.13. Jest to związane ze wzrostem współczynnika sprawności wstrzykiwania



Rys. 11.13
Zależność α (I_A)

i kilkoma innymi czynnikami, powodującymi również zwiększenie się współczynnika transportu nośników przez bazę. Zatem przełączanie ze stanu blokowania do przewodzenia następuje wówczas, gdy wskutek jakiejś przyczyny zwiększy się prąd anodowy, co spowoduje wzrost wartości ($\alpha_1 + \alpha_2$) do jedności. W tyrystorze dwukońcówkowym (w dynistorze) można brać pod uwagę dwie przyczyny zwiększenia prądu anodowego, prowadzące w konsekwencji do przełączenia tyrystora ze stanu blokowania do stanu przewodzenia. Są to następujące zjawiska:

- powielanie lawinowe nośników w warstwie zaporowej złącza j_2 przy odpowiednio dużym napięciu U_{AK} ;
- zwiększanie się prądu I_θ wskutek wzrostu temperatury, pod wpływem oświetlenia lub promieniowania innego rodzaju;
- przepływ prądu pojemnościowego przy gwałtownej zmianie napięcia anoda-katoda.

Pierwsze zjawisko polega na tym, że liczba nośników wychodzących z warstwy zaporowej złącza j_2 jest większa niż liczba nośników wchodzących do tej warstwy. Na przykład liczba dziur wychodzących przez prawą krawędź warstwy zaporowej złącza j_2 (rys. 11.11a) jest M_p razy większa niż liczba dziur wchodzących przez lewą krawędź tego złącza. Zjawisko powielania można uwzględnić analitycznie we współczynniku wzmocnienia napięciowego, dając zamiast α iloczyn $M\alpha$. Zakładając w pierwszym przybliżeniu, że współczynnik powielania M jest jednaki dla dziur i elektronów odbieranych przez kolektor oraz generowanych w warstwie zaporowej kolektora, należy przepisać równanie (11.8) w postaci

$$I_A = \frac{MI_g}{1 - M(\alpha_1 + \alpha_2)} \quad (11.9)$$

przy czym:

$$M = \frac{1}{1 - \left(\frac{U_{AK}}{U_p}\right)^n} \quad (11.10)$$

U_p — napięcie przebicia.

Przełączenie następuje przy takiej wartości napięcia U_{AK} , dla której

$$M(\alpha_1 + \alpha_2) = 1$$

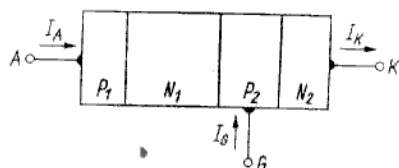
Drugie zjawisko polega na zwiększaniu prądu anodowego wskutek wzrostu prądu generacji I_g .

Trzecie zjawisko polega na ładowaniu pojemności C_j złącza j_2 , co oznacza chwilowy przepływ dużego prądu anodowego

$$i_A \approx C_j \frac{du_{AK}}{dt} \quad (11.11)$$

i prowadzi do spełnienia warunku $(\alpha_1 + \alpha_2) = 1$.

We wszystkich trzech przypadkach przyczyna wywołująca wzrost prądu anodowego i w konsekwencji przełączenie może trwać bardzo krótko, gdyż tyristor po przełączeniu sam utrzymuje się w stanie przewodzenia. Powrót do stanu zablokowania następuje tylko wówczas, gdy prąd anodowy zmaleje poniżej pewnej wartości krytycznej (nazywanej *prądem trzymania*), której odpowiada $(\alpha_1 + \alpha_2) < 1$.



Rys. 11.14
Struktura tyristora triodowego (z bramką)

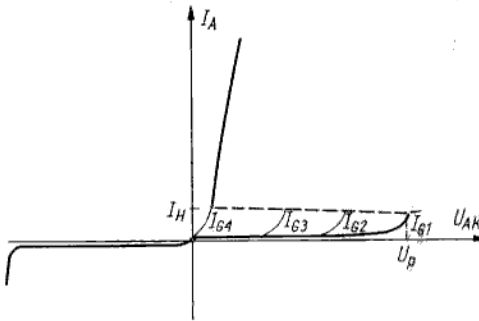
Rozpatrzmy teraz tyristor triodowy, który oprócz anody i katody ma trzecią końcówkę nazywaną *bramką* (G) — rys. 11.14. Ponieważ

$$I_K = I_A + I_G$$

wyrażenie (11.9) należy przepisać w postaci

$$I_A = \frac{M(I_g + \alpha_2 I_G)}{1 - (\alpha_1 + \alpha_2)M} \quad (11.12)$$

W tyrystorze triodowym, oprócz wyżej wymienionych trzech sposobów przełączania ze stanu blokowania do stanu przewodzenia, można wymienić jeszcze jeden sposób (najważniejszy w tym elemencie), tj. zwiększenie prądu anodowego wskutek przepływu prądu bramki. Napięcie przełączania jest funkcją prądu bramki, a więc dla różnych wartości tego prądu otrzymuje się rodzinę charakterystyk prądowo-napięciowych (rys. 11.15).



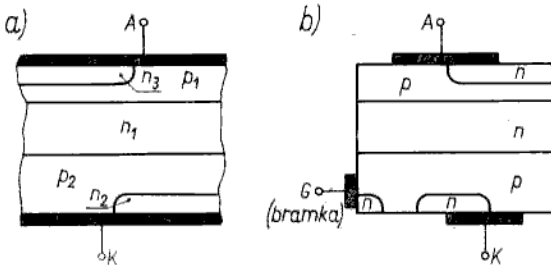
Rys. 11.15

Rodzina charakterystyk $I_A(U_{AK})$ dla różnych wartości prądu bramki ($I_{G4} > I_{G3} > I_{G2} > I_{G1}$)

Tyrystor dwukierunkowy (struktura pięciowarstwowa n-p-n-p-n)

11.2.2

Podstawową strukturę pięciowarstwową tyrystora dwukierunkowego przedstawiono na rys. 11.16 dla wariantu dwukońcówkowego (diak) oraz trójkońcówkowego (triak). Sposób działania tych tyrystorów można wyjaśnić na podstawie



Rys. 11.16

Struktura pięciowarstwowa:
a) diak; b) triak

superpozycji dwóch struktur czterowarstwowych ($p_1-n_1-p_2-n_2$ oraz $p_2-n_1-p_1-n_3$). Taki tyrystor ma symetryczne właściwości dla obu polaryzacji anoda-katoda (charakterystyki $I(U)$ są pokazane w tabl. 11.1). W triaku za pomocą sterowania obwodem bramki można regulować wartość napięcia przełączania.

Parametry tyrystorów

11.2.3

Do podstawowych parametrów granicznych tyrystorów należy zaliczyć następujące:

- dopuszczalny prąd anodowy w stanie przewodzenia $I_{A(AV)}$;
- graniczne napięcie anodowe w stanie zaporowym BU_{AK} ;
- moc admysyjna lub opór cieplny.

Do najważniejszych parametrów charakterystycznych tyrystorów zalicza się następujące:

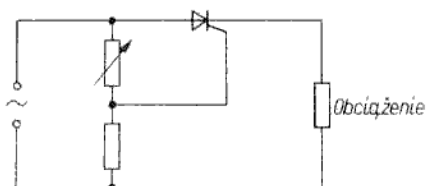
- maksymalne napięcie przełączania U_{A0} , mierzone przy zerowym prądzie bramki;
- prąd trzymania I_H ; jest to minimalna wartość prądu anodowego, przy której tyrystor nie wyłącza się ze stanu przewodzenia do stanu blokowania;
- prąd przełączający bramki I_{GT} , podawany dla określonego napięcia U_{AK} .

Oprócz wymienionych parametrów statycznych bardzo ważne są również dwa parametry dynamiczne, wyznaczające czasy włączania i wyłączania tyrystora. Czas włączania jest wyznaczony przedziałem czasu od czola impulsu przełączającego bramki do chwili, gdy prąd anodowy osiągnie 90% swej wartości ostatecznej. Czas wyłączania jest wyznaczony przedziałem czasu od chwili, gdy prąd anodowy zmaleje poniżej prądu trzymania, do chwili, kiedy bramka odzyska swoje właściwości sterujące. Czas wyłączania jest zwykle kilkanaście razy większy niż czas włączania.

Zastosowania

11.2.4

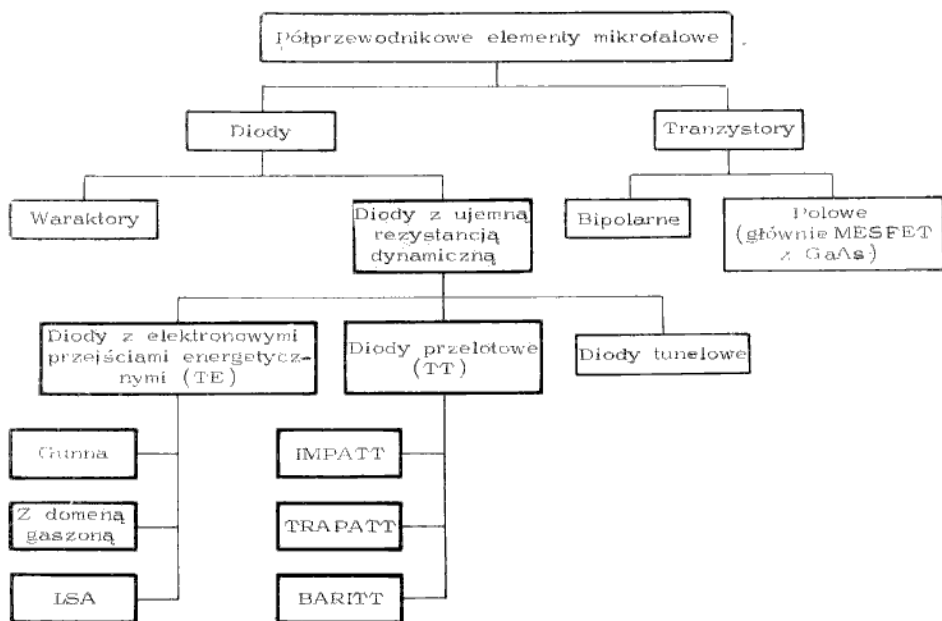
Generalnie tyrystory są stosowane w układach zasilania w celu sterowania dużych mocy małymi. Przykład prostego układu prostownika jednopółkowego, w którym bramka jest sterowana impulsami taktowymi o przesunięciu fazowym, niezależnym od częstotliwości, przedstawiono na rys. 11.17. Zmieniając wartość



Rys. 11.17
Układ tyrystorowego prostownika jednopółkowego z regulacją kąta fazowego zapłonu w niewielkim zakresie

przesunięcia fazowego można płynnie regulować energię oddawaną do obciążenia. W sytuacji, kiedy prostowanie prądu nie jest wymagane (na przykład sieć oświetlenia sali kinowej), stosuje się triaki.

Do wzmacniania i generacji sygnałów mikrofalowych służą elementy półprzewodnikowe, których klasyfikację przedstawiono na rys. 12.1. W rodzinie półprzewodnikowych elementów mikrofalowych bardzo duże znaczenie mają diody



Rys. 12.1

Klasyfikacja półprzewodnikowych elementów mikrofalowych; linią grubą zaznaczono elementy omawiane w tym rozdziale

z ujemną rezystancją dynamiczną (pozostałe elementy, tj. tranzystory mikrofalowe i waraktory, były rozpatrywane w rozdziałach 4, 5, 6). W diodach tych wykorzystuje się ujemną rezystancję dynamiczną w celu bezpośredniego przetwarzania energii prądu stałego w energię prądu zmiennego (sinusoidalnego lub impulsowego). Ujemna rezystancja dynamiczna może być wywołana różnymi zjawiskami fizycznymi, przy czym wyróżnia się trzy rodzaje zjawisk i związane z nimi trzy grupy diod:

- elektronowe przejścia energetyczne — diody TE (ang. *Transferred-Electron*);
- skończony czas przelotu nośników w obszarze ich unoszenia — diody TT (ang. *Transit Time* lub *Triggered Transit*);

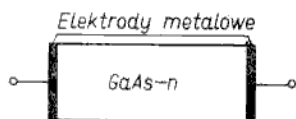
— tunelowanie nośników z pasma przewodnictwa przez warstwę zaporową do pasma walencyjnego — diody tunelowe.

Diody tunelowe mają szereg zastosowań nie związanych z zakresem mikrofalowym, dlatego były już wcześniej rozpatrywane (rozdziały 3, 4). Poza tym mogą one służyć do wzmacniania i generacji sygnałów mikrofalowych małej mocy (zaledwie kilka miliwatów) w odróżnieniu od dwóch pierwszych rodzajów diod, w których uzyskuje się moce rzędu kilkuset miliwatów i większe.

Diody TE

12.1

Działanie diod TE (rys. 12.2) jest oparte na zjawisku elektronowych przejść energetycznych, które powoduje zmniejszanie się ruchliwości elektronów, a więc zmniejszanie się konduktancji diody, w obecności silnego pola elektrycznego.



Rys. 12.2
Ilustracja schematyczna diody TE

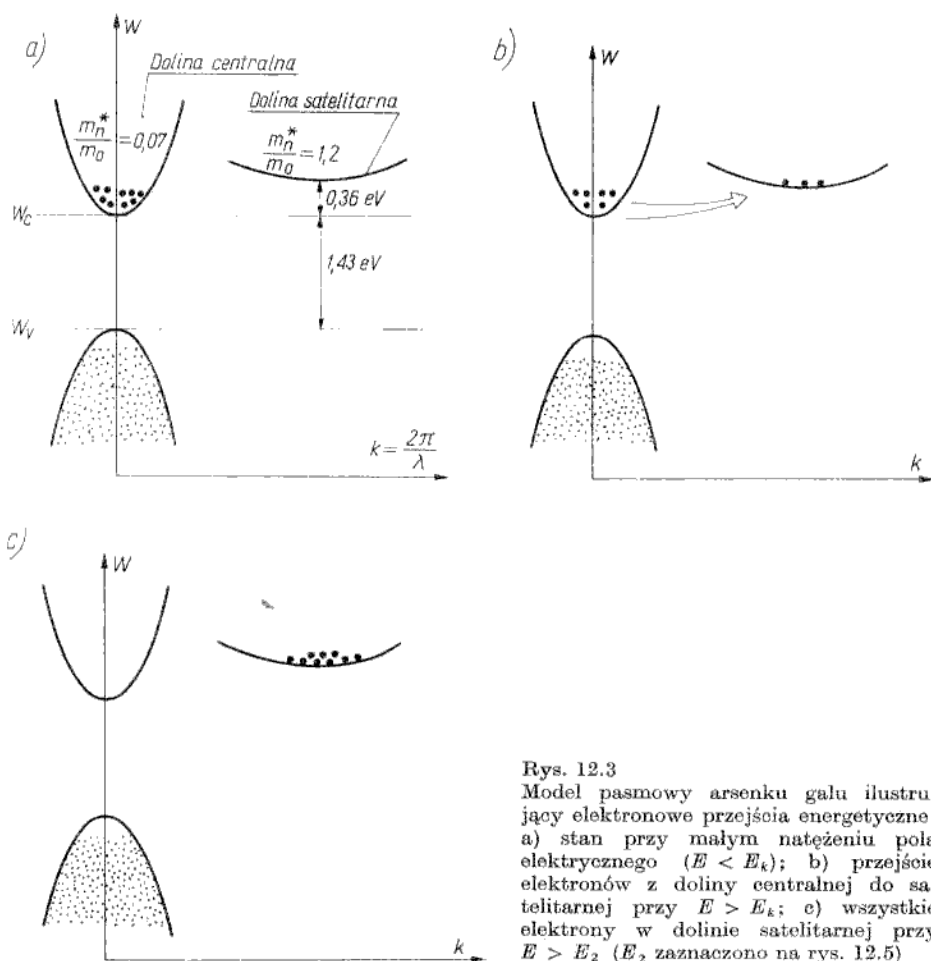
Ponieważ jest to proces regeneracyjny, w obwodzie z diodą TE obserwuje się oscylacje prądu. Zjawisko oscylacji prądu w kryształach GaAs polaryzowanym stałym napięciem po raz pierwszy zaobserwował J. B. Gunn w 1963 r., dlatego jest ono nazywane *zjawiskiem Gunna*. Diody TE są niekiedy nazywane *diodami Gunna*, choć częściej stosuje się tę nazwę tylko w odniesieniu do jednego z trzech możliwych rodzajów diod TE, tj. do tych diod, które działają zgodnie z mechanizmem podanym przez Gunna.

Ujemna konduktancja dynamiczna

12.1.1

Zjawisko Gunna może wystąpić tylko w tych półprzewodnikach, w których są możliwe elektronowe przejścia energetyczne¹⁾ przy dużym natężeniu pola elektrycznego. Do takich półprzewodników należą arsenek galu (GaAs) i fosforek indu (InP). Istotę fizyczną elektronowych przejść energetycznych wyjaśnia rys. 12.3, na którym przedstawiono model pasmowy GaAs. W dotychczasowych rozważaniach model pasmowy półprzewodnika (głównie krzemu) kreśliliśmy z uwzględnieniem tylko energii. Nie interesowaliśmy się natomiast wektorem falowym (pędem) elektronów, gdyż jego wartość w pasmie walencyjnym i w pasmie przewodnictwa była jednakowa. Obecnie w modelu pasmowym GaAs uwzględniamy zarówno energię, jak i pęd elektronów, czyli kreślimy go we współrzędnych W, k . Na takim wykresie jest widoczne istnienie dwóch pasm przewodnictwa, tj. pasma głównego (we współrzędnych W, k ma ono kształt doliny, dlatego jest nazywane *doliną centralną*) oraz pasma dodatkowego (tzw. *doliny satelitarnej*). Odległość od dna doliny centralnej do wierzchołka pasma walencyjnego (pasmo zabronione) wynosi 1,43 eV. Dno doliny satelitarnej ma wartość energii o 0,36 eV większą niż dno doliny centralnej.

¹⁾ Mechanizm elektronowych przejść energetycznych opisali teoretycznie Ridley i Watkins w 1961 r. oraz Hilsun w 1962 r., dlatego niekiedy mówi się o mechanizmie RWH.

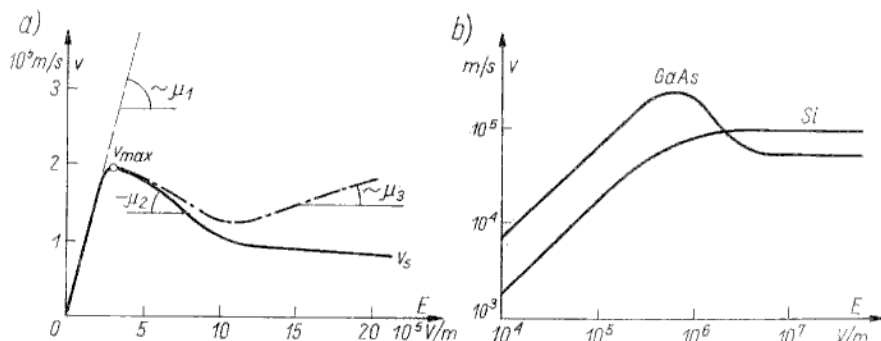


Rys. 12.3

Model pasmowy arsenku galu ilustrujący elektronowe przejścia energetyczne: a) stan przy małym natężeniu pola elektrycznego ($E < E_k$); b) przejście elektronów z doliny centralnej do satelitarnej przy $E > E_k$; c) wszystkie elektrony w dolinie satelitarnej przy $E > E_2$ (E_2 zaznaczono na rys. 12.5)

Najważniejszą właściwością rozpatrywanego modelu jest to, że elektrony w dolinie satelitarnej mają nie tylko inną wartość wektora falowego (pędu) niż w dolinie centralnej, lecz również inną wartość masy efektywnej. Elektrony w dolinie satelitarnej są około 20-krotnie „cięższe” niż w dolinie centralnej, a więc mają znacznie mniejszą ruchliwość. W stanie normalnym (przy małych natężeniach pola elektrycznego) dolina satelitarna jest pusta, gdyż wszystkie elektrony swobodne są ulokowane na najniższych dozwolonych poziomach energetycznych, tj. na dnie doliny centralnej (rys. 12.3a). W miarę zwiększania natężenia pola elektrycznego, przy pewnej wartości krytycznej (E_k) elektrony są „przesiedlane” pod wpływem energii pola elektrycznego z doliny centralnej do doliny satelitarnej (rys. 12.3b), wskutek czego wzrasta ich masa efektywna i maleje ruchliwość. Przy natężeniu pola większym niż krytyczne wszystkie elektrony swobodne przechodzą do doliny satelitarnej (rys. 12.3c). Na rysunku 12.4a przedstawiono wykres prędkości unoszenia elektronów w funkcji natężenia pola elektrycznego. Początkowo prędkość unoszenia elektronów zwiększa się wprost

proporcjonalnie do natężenia, czyli ruchliwość μ_1 jest wielkością stałą. Przy natężeniu krytycznym ($E_k \approx 3 \cdot 10^5$ V/m, czyli 0,3 V na mikrometr) rozpoczynają się przejścia elektronów z doliny centralnej do doliny satelitarnej, gdzie mając większą masę poruszają się one wolniej. Prędkość unoszenia elektronów



Rys. 12.4

Wykres prędkości unoszenia elektronów w funkcji natężenia pola elektrycznego: a) dla arsenku galu; b) porównanie dla arsenku galu i krzemu

maleje zatem w miarę zwiększania natężenia pola ponad wartość krytyczną. Ruchliwość różniczkowa w tym przypadku ma znak ujemny ($\mu_2 = -dv/dE$). Po przejściu wszystkich elektronów do doliny satelitarnej prędkość unoszenia powinna znów zwiększać się wprost proporcjonalnie do natężenia pola (krzywa kreska-kropka na rys. 12.4a) ze współczynnikiem proporcjonalności równym ruchliwości μ_3 znacznie mniejszej niż μ_1 . Jednakże przy tak dużych natężeniach pola elektrycznego w GaAs, tak jak i w innych materiałach półprzewodnikowych, następuje ograniczenie prędkości unoszenia do wartości v_s równej w przybliżeniu 10^5 m/s.

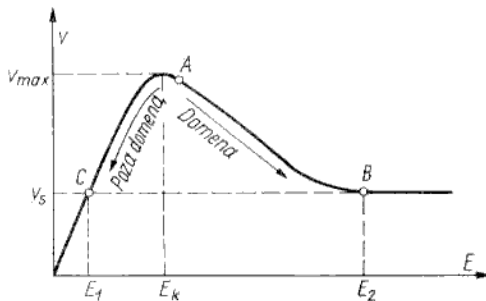
Na rysunku 12.4b porównano krzywe $v(E)$ dla GaAs i Si. W krzemie nie ma elektronowych przejść energetycznych, dlatego obserwuje się tylko nasycenie prędkości unoszenia elektronów przy dużych natężeniach pola elektrycznego. W arsenku galu nasycenie jest poprzedzone odcinkiem o ujemnym nachyleniu. Ponieważ konduktancja bryły półprzewodnika jest wprost proporcjonalnie zależna od ruchliwości nośników, odcinkowi o ujemnej ruchliwości odpowiada odcinek o ujemnej konduktancji, a charakterystyka prądowo-napięciowa $I(U)$ ma przebieg identyczny jak charakterystyka $v(E)$. Podsumowując powyższe rozważania można wymienić trzy następujące warunki niezbędne do wystąpienia ujemnej konduktancji dynamicznej wskutek elektronowych przejść energetycznych;

- różnica energii między dnem doliny centralnej a dnem doliny satelitarnej musi być znacznie większa niż kT , gdyż w przeciwnym przypadku dolina satelitarna byłaby wypełniona w temperaturze pokojowej bez działania zewnętrznym polem elektrycznym;
- wytrzymałość na przebicie musi być większa niż krytyczne natężenie pola elektrycznego;
- ruchliwość elektronów w dolinie satelitarnej musi być znacznie mniejsza niż w dolinie centralnej.

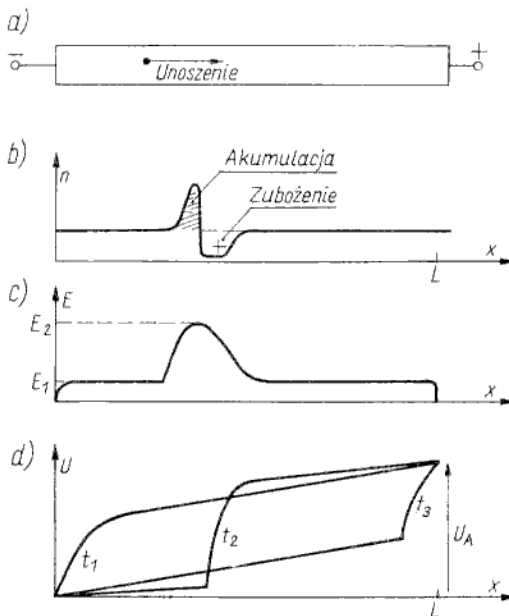
Powstawanie i unoszenie domeny

12.1.2

Charakterystyki $v(E)$, omawianej w poprzednim punkcie, nie można obserwować w warunkach statycznych, gdyż przy natężeniu pola większym niż wartość krytyczna ($E > E_k$) rozkład koncentracji elektronów wzdłuż próbki jest potencjalnie niestabilny i jakiegokolwiek lokalne odchylenie od stanu neutralności powoduje powstanie obszaru ładunku przestrzennego, nazywanego *domeną elektryczną*. Zjawisko powstawania i unoszenia domeny w polu elektrycznym ilustrują rys. 12.5, 12.6. Jeżeli napięcie U_A , doprowadzone do próbki o długości L osiągnie wartość $U_A/L = E_k$, to prędkość unoszenia nośników i prąd płynący przez próbkę uzyskują wartości maksymalne. Rozkład potencjału wzdłuż próbki nie jest idealnie równomierny, gdyż zawsze istnieją pewne lokalne niejedno-

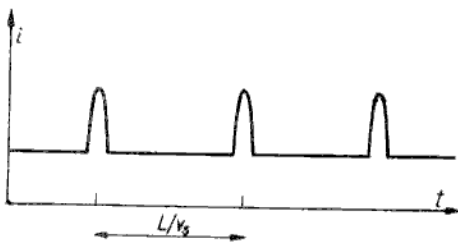


Rys. 12.5
Wykres $v(E)$ ilustrujący zjawisko powstawania domeny



Rys. 12.6
Powstawanie domeny: a) próbka arsenku galu; b) domena-dipolowy ładunek przestrzenny; c) rozkład natężenia pola elektrycznego; d) rozkład potencjału w kolejnych chwilach czasu

rodności (najczęściej przy katodzie), w których natężenie pola osiąga wartość nieco większą niż średnia. Zatem przy średniej wartości $E = U_A/L = E_k$ w obszarze lokalnej niejednorodności pole elektryczne osiąga wartość większą niż krytyczna (punkt A na rys. 12.5). Wówczas w tym obszarze elektrony przechodzą z doliny centralnej do doliny satelitarnej, wskutek czego prędkość ich unoszenia maleje, a więc ruchliwość różniczkowa przyjmuje znak ujemny. Ponieważ elektrony za i przed tym obszarem (za — od strony katody, przed — od strony anody) poruszają się szybciej, z jednej strony (od katody) jest akumulowany ładunek ujemny elektronów, a z drugiej strony (od anody) ubywa elektronów, czyli powstaje obszar zubożenia o ładunku dodatnim (rys. 12.6b). W ten sposób powstaje obszar dipolowego ładunku przestrzennego nazywany *domeną*. Jednocześnie z powstawaniem domeny zmienia się rozkład natężenia pola elektrycznego, gdyż coraz większy jest spadek potencjału w obszarze ładunku przestrzennego, a maleją spadki potencjałów w obszarach znajdujących się poza domeną (na odcinkach od katody do domeny i od domeny do anody). Wskutek tego natężenie pola elektrycznego w domenie wzrasta, co powoduje zmniejszenie się prędkości unoszenia elektronów akumulowanych w domenie, a więc prędkości unoszenia domeny, zgodnie ze strzałką skierowaną od punktu A do punktu B na rys. 12.5. Jednocześnie maleje natężenie pola elektrycznego w obszarach położonych poza domeną, a więc maleje prędkość unoszenia elektronów w tych obszarach zgodnie ze strzałką skierowaną od punktu A do punktu C na rys. 12.5. Proces powstawania domeny zostanie zakończony w chwili, gdy prędkości unoszenia domeny i elektronów w obszarach poza domeną osiągną jednakową wartość równą prędkości nasycenia v_s . Wówczas natężenie pola elektrycznego w domenie (rys. 12.5) przyjmuje wartość E_2 znacznie większą niż natężenie pola E_1 w obszarach przed i za domeną (rys. 12.6c). Podczas powstawania domena jest jednocześnie unoszona w stronę anody, dlatego rozkład potencjału zmienia się w czasie, jak to pokazano na rys. 12.6d. Po dojściu do anody domena znika, wskutek czego zwiększa się konduktancja próbki i prąd przez nią płynący, lecz natychmiast zaczyna powstawać nowa domena (jeżeli $U_A > E_k L$) i wyżej opisane zjawiska powtarzają się. Zatem przy natężeniu pola $E > E_k$, jakiemu odpowiada ujemna ruchliwość różniczkowa, próbka generuje periodycznie impulsy prądu. Przebieg prądu w funkcji czasu przedstawiono na rys. 12.7. Odstęp czasu między dwoma



Rys. 12.7
Przebieg prądu generowany w diodzie
Gunna

kolejnymi impulsami (okres oscylacji) jest równy czasowi przelotu domeny od katody do anody (zakłada się, że domeny powstają przy katodzie); tj.

$$T = \frac{L}{v_s} \quad (12.1)$$

W dotychczasowych rozważaniach przyjmowaliśmy *implicite*, że czas przelotu domeny jest znacznie dłuższy niż czas jej powstawania (formowania). Oszacujmy teraz

w sposób bardzo uproszczony, jakie warunki muszą być spełnione, by to założenie było słuszne. Proces formowania ładunku przestrzennego (domeny) można opisać równaniem

$$Q(x, t) = Q(x - vt, 0) \exp \frac{t}{\tau_d} \quad (12.2)$$

przy czym

$$\tau_d = \frac{\varepsilon_s}{qn|\mu_n|} \text{ — czas relaksacji elektronów}^{1)} \quad (12.3)$$

Przy założeniu, że domena „rodzi się” przy katodzie, jest

$$Q(x, t) = Q(0, 0) \exp \frac{t}{\tau_d}$$

Po przejściu do anody w czasie $t = L/v_s$

$$Q\left(L, \frac{L}{v_s}\right) = Q(0, 0) \exp \frac{L}{v_s \tau_d} \quad (12.4)$$

Jeżeli przyjmiemy, że ładunek akumulowany w domenie wzrasta kilkakrotnie podczas jej formowania, to

$$\frac{Q\left(L, \frac{L}{v_s}\right)}{Q(0,0)} \gg 1$$

Zatem

$$\frac{L}{v_s \tau_d} > 1$$

i uwzględniając (12.3)

$$Ln > \frac{\varepsilon_s v_s}{q|\mu_n|} \quad (12.5)$$

Dla GaAs ($\varepsilon_{rs} = 12$, $v_s = 10^5$ m/s, $\mu_n = -0,01$ m²/(V · s))

$$Ln > 10^{16} \text{ m}^{-2} \quad (12.6)$$

Diody TE, w których jest spełniony ten warunek i okres generowanego sygnału jest równy czasowi przelotu domeny, nazywa się *diodami Gunna*, gdyż takie właściwości miały pierwsze diody badane przez Gunna. Jednakże całkowite uformowanie domeny nie jest warunkiem koniecznym do uzyskania wzmocnienia lub generacji sygnałów mikrofalowych.

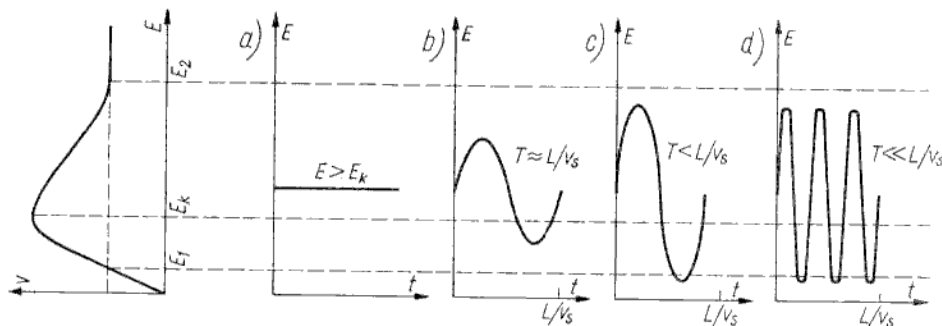
¹⁾ Relaksacja (uspokojenie) jest w istocie procesem powrotu do stanu równowagi po jej zaburzeniu. W rozdziale I wykazano że stała czasowa takiego procesu jest równa iloczynowi przenikalności elektrycznej i rezystywności materiału. Obecnie przyjmuje się, że w identyczny sposób można opisać proces odwrotny do relaksacji, tj. narastanie zaburzenia po zachwianiu równowagi, przy czym do wzoru na τ_d podstawia się ujemną ruchliwość różniczkową.

Trzy główne rodzaje pracy (mody) diod TE

12.1.3

Dotychczas rozpatrywaliśmy przypadek polaryzacji diody TE stałym napięciem, któremu odpowiada wartość średnia natężenia pola elektrycznego, większa niż wartość krytyczna. Wówczas przy spełnionym warunku (12.5) dioda generuje impulsy prądu o okresie równym czasowi przelotu domeny. Taki rodzaj pracy nazywany *modem Gunna*, charakteryzuje się małą sprawnością przemiany energii prądu stałego w energię sygnału mikrofalowego (kilka procent), gdyż czas trwania impulsu prądu jest znacznie mniejszy niż okres oscylacji. Aby zwiększyć sprawność, stosuje się inne rodzaje pracy diod TE, pracujących w układach z obwodami rezonansowymi, przy czym okres oscylacji jest wymuszony przez obwód rezonansowy.

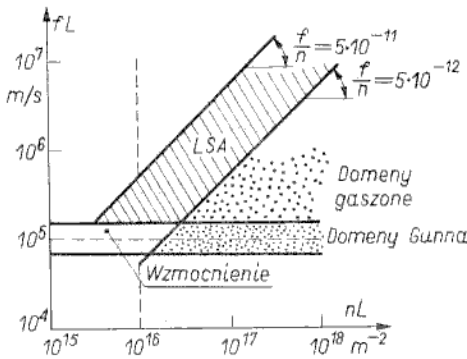
Trzy główne mody ilustrują przebiegi na rys. 12.8.



Rys. 12.8

Ilustracja przebiegów odpowiadających trzem głównym rodzajom pracy (modom) diod TE: a), b) mod Gunna; c) mod domeny gazzonej; d) mod LSA

W przypadku obciążenia rezystancyjnego (rys. 12.8a) lub włączenia diody w obwodzie rezonansowym o stałej czasowej, porównywalnej z czasem przelotu domeny (rys. 12.8b), mamy do czynienia z *modem Gunna*. Jeżeli stała czasowa obwodu rezonansowego jest mniejsza niż czas przelotu domeny (rys. 12.8c) i natężenie pola w połowie ujemnej osiąga wartość mniejszą niż E_1 (według rys. 12.5), to domena przed dojściem do anody jest *gazsona*. Jest to *mod domeny gazzonej*, charakteryzujący się większą sprawnością niż w przypadku modu Gunna. Trzeci i najważniejszy mod nosi nazwę LSA (ang. *Limited Space-Charge Accumulation*) i oznacza taki rodzaj pracy, gdy stała czasowa obwodu rezonansowego jest mniejsza nie tylko niż czas przelotu domeny, lecz również niż czas relaksacji (czas powstawania) domeny. Wówczas niemal w całym okresie z wyjątkiem krótkiego przedziału czasu, w którym natężenie pola jest mniejsze niż E_1 , dioda ma ujemną konduktancję dynamiczną. Dzięki temu diody LSA mają największą sprawność (do 20%). Warunki, w jakich występują trzy mody generacji lub wzmacnianie, ilustruje zbiorczy wykres na rys. 12.9. Pionową linią przerywaną zaznaczono krytyczną wartość iloczynu $nL = 10^{16} \text{ m}^{-2}$, przy której czas przelotu jest równy czasowi relaksacji. Domeny mogą powstawać tylko na prawo od tej linii. Poziomą linią przerywaną zaznaczono krytyczną wartość iloczynu $fL = v_s = 10^5 \text{ V/m}$. Linie ograniczające na tym wykresie mod LSA wynikają z dwóch wymagań: — *okres generowanego sygnału ($1/f$) powinien być mniejszy niż stała czasowa*



Rys. 12.9
Wykres fL w funkcji nL , na którym zaznaczono obszary odpowiadające głównym rodzajom pracy diod TE

procesu powstawania domeny $\left(\tau_d = \frac{\varepsilon_s}{qn|\mu_n|} \right)$, przy czym $\mu_n = -0,01 \text{ m}^2/(\text{V} \cdot \text{s})$, przy czym za kryterium praktyczne przyjmuje się zwykle $T \leq 3\tau_d$,
 — okres generowanego sygnału ($1/f$) powinien być większy (praktycznie biorąc co najmniej trzy razy) niż stała czasowa procesu zanikania domeny $\left(\tau_d = \frac{\varepsilon_s}{qn|\mu_n|} \right)$, przy czym $\mu_n \approx 0,5 \text{ m}^2/(\text{V} \cdot \text{s})$.

Diody przelotowe

12.2

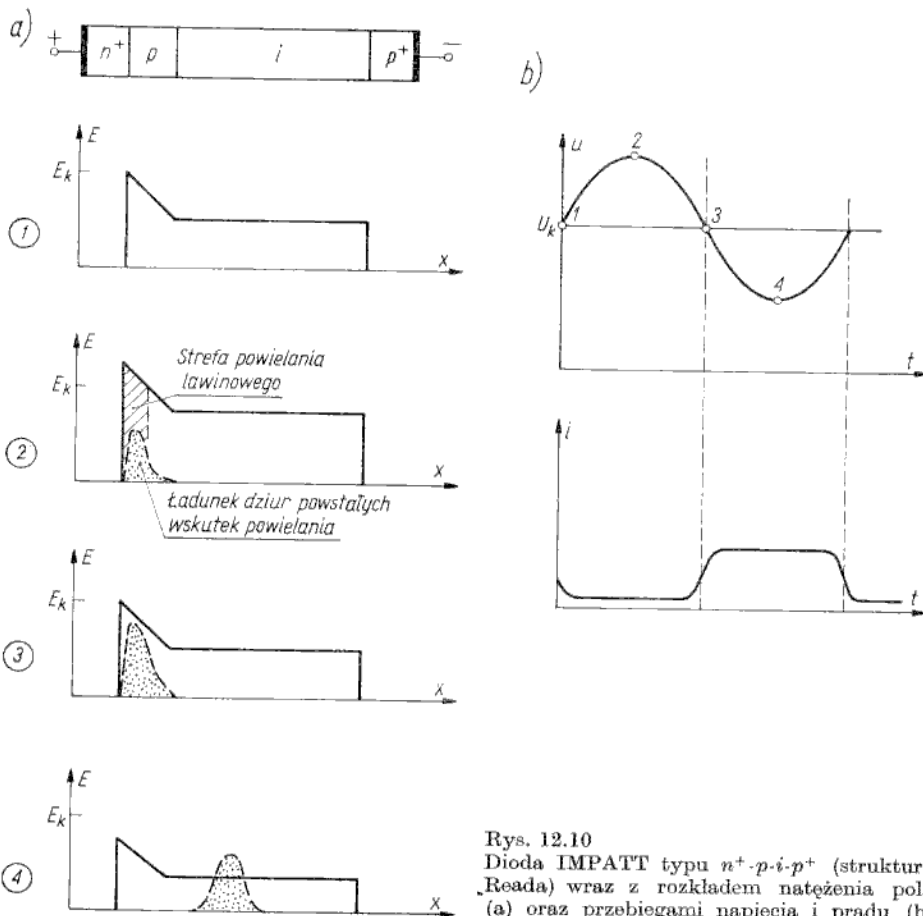
Zgodnie z klasyfikacją przedstawioną na rys. 12.1 rozróżnia się diody lawinowo-przelotowe (IMPATT i TRAPATT) oraz inieekyjno-przelotowe (BARITT).

Diody lawinowo-przelotowe

12.2.1

W diodach lawinowo-przelotowych uzyskuje się ujemną konduktancję dynamiczną wskutek przesunięcia fazowego o 180° między zmiennymi sygnałami prądu i napięcia. Kierunki zmian napięcia i prądu są zawsze przeciwne, tj. wzrastającemu napięciu odpowiada malejący prąd, a malejącemu napięciu — wzrastający prąd. Nazwa diody lawinowo-przelotowe pochodzi stąd, że mechanizm działania tych diod obejmuje dwa zjawiska, tj. powielanie lawinowe nośników w określonym obszarze oraz ich przelot przez warstwę ładunku przestrzennego. Przesunięcie fazowe między napięciem a prądem jest spowodowane tym, że najpierw napięcie wywołuje powielanie lawinowe, a dopiero po pewnym czasie, koniecznym do przelotu nośników przez warstwę ładunku przestrzennego, pojawia się prąd. Opóźnienie prądu względem napięcia jest zatem spowodowane skończonym czasem przelotu nośników przez warstwę ładunku przestrzennego, a w pewnym stopniu może być także spowodowane zwłoką w procesie powielania lawinowego. Bardziej szczegółowo zostanie rozpatrzone działanie diody IMPATT (ang. *Impact Avalanche Transit Time*) w wersji opisanej po raz pierwszy przez

W. T. Reada w 1958 r., tj. w postaci struktury $n^+ - p - i - p^+$ (tzw. dioda Reada). Rozkład natężenia pola elektrycznego w tej strukturze przedstawiono na rys. 12.10a dla polaryzacji w kierunku zaporowym napięciem stałym o wartości U_k , na które jest nałożona składowa zmienna o odpowiednio dobranej częstotli-



Rys. 12.10
 Dioda IMPATT typu $n^+ - p - i - p^+$ (struktura Reada) wraz z rozkładem natężenia pola (a) oraz przebiegami napięcia i prądu (b)

wości. Warstwa zaporowa obejmuje słabo domieszkowane warstwy, tj. warstwę typu p oraz i (warstwę półprzewodnika samoistnego), przy czym natężenie pola elektrycznego jest większe w warstwie typu p . Napięcie krytyczne U_k odpowiada krytycznej wartości natężenia pola E_k , przy której rozpoczyna się proces jonizacji zderzeniowej.

A oto cztery fazy pracy diody, odpowiadające punktom 1 do 4 zaznaczonym na przebiegu $u(t)$ — rys. 12.10b. W punkcie 1 rozpoczyna się proces jonizacji zderzeniowej i następnie narasta w całej dodatniej połowie napięcia. Powielanie lawinowe zachodzi w cienkiej warstwie typu p (strefa powielania), w której natężenie pola elektrycznego jest większe niż wartość krytyczna E_k . Proces powielania kończy się w punkcie 3, gdy napięcie maleje poniżej wartości U_k . Wskutek jonizacji zderzeniowej powstają pary elektron-dziura, przy czym elektrony są unoszone do

warstwy n^+ , a dziury są unoszone ze stałą prędkością (w silnym polu elektrycznym prędkość unoszenia osiąga wartość nasyconą, tj. $v_s = 10^5$ m/s dla krzemu) w warstwie i nazywanej obszarem unoszenia. Jeżeli czas przelotu dziur przez obszar unoszenia (w bardziej ścisłych rozważaniach może to być suma czasu przelotu i stałej czasowej procesu powielania lawinowego) jest równy połowie okresu sygnału sinusoidalnego, to prąd w obwodzie zewnętrznym zaczyna się zwiększać w punkcie 3 i pozostaje na niemal stałym poziomie w ujemnej połowie napięcia. Stąd optymalna wartość częstotliwości

$$f = \frac{v_s}{2L} \quad (12.7)$$

przy czym L — szerokość obszaru unoszenia (warstwy i). Na przykład dla $L = 1 \mu\text{m}$

$$f = \frac{10^5 \text{ m/s}}{2 \cdot 10^{-6} \text{ m}} = 50 \text{ GHz}$$

Przestrzegając obwód rezonansowy można zmieniać częstotliwość oscylacji w niewielkim zakresie (2:1) wokół częstotliwości określonej wzorem (12.7).

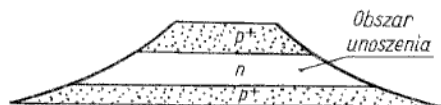
Diody IMPATT mogą mieć strukturę prostszą niż omawiana dioda Reada ($n^+ - p - i - p^+$), a mianowicie mogą być wykonane w postaci złącza $p - n$ lub $p - i - n$. Sprawność tych prostszych struktur jest nawet większa, gdyż oba rodzaje nośników biorą udział w powielaniu lawinowym i unoszeniu, a więc mają wpływ na warstość prądu.

W dotychczasowych rozważaniach pomijano wpływ ładunku generowanego lawinowo na rozkład natężenia pola. Pole elektryczne jest związane z ładunkiem nieskompensowanym zjonizowanych centrów akceptorowych i donorowych. Pary elektron-dziura generowane w warstwie zaporowej kompensują ten ładunek powodując zmniejszanie się natężenia pola. Załóżmy, że napięcie doprowadzone do diody wzrasta gwałtownie powyżej wartości wymaganej do jonizacji zderzeniowej. Wówczas w obszarze o największym natężeniu pola (w strefie powielania lawinowego) są generowane pary elektron-dziura, wskutek czego natężenie pola w tym obszarze maleje. Jednocześnie wzrasta natężenie pola w obszarze sąsiednim, w którym rozpoczyna się powielanie lawinowe itd. Zatem front strefy powielania lawinowego rozprzestrzenia się bardzo szybko wzdłuż warstwy zubożonej, pozostawiając za sobą obszar o bardzo dużej koncentracji nośników wygenerowanych lawinowo (tzw. plazma). Ponieważ natężenie pola elektrycznego w tym obszarze jest małe, nośniki są unoszone z niewielką prędkością (mówi się, że plazma znajduje się w pułapce, tzn. jest niemal nieruchoma w słabym polu elektrycznym). W chwili gdy sfery powielania lawinowego obejmie całą warstwę zubożoną, co stanie się w czasie krótszym niż czas przelotu nośników stanowiących plazmę, napięcie na diodzie gwałtownie zmaleje, a prąd wzrośnie. Następuje druga połowa cyklu, w której prąd spowodowany unoszeniem plazmy jest duży, a napięcie — małe. Po usunięciu plazmy opisany cykl pracy diody powtarza się. Ten rodzaj pracy diody lawinowo-przelotowej nosi nazwę TRAPATT (ang. *TRApped Plasma Avalanche Triggered Transit*). Oscylacja w modzie TRAPATT charakteryzuje się bardzo dużą sprawnością przemiany energii prądu stałego w energię sygnału zmiennego.

Diody iniekcyjno-przelotowe

12.2.2

Strukturę diody iniekcyjno-przelotowej, nazywanej skrótowo BARITT (ang. *BAR*rier *I*njection *T*ransit *T*ime), przedstawiono na rys. 12.11. Dioda składa się z dwu złączy *p-n* rozdzielonych warstwą słabo domieszkowaną (tzw. *obszarem unoszenia*).



Rys. 12.11
Struktura diody BARITT

Ogólna zasada działania tej diody jest podobna do działania diody IMPATT, tj. przesunięcie fazowe między prądem a napięciem jest spowodowane skończonym czasem przelotu przez obszar unoszenia (warstwę typu *n*). Inne jest tylko źródło nośników. O ile w diodzie IMPATT pochodziły one z jonizacji zderzeniowej, o tyle w diodzie BARITT są one wstrzykiwane przez jedno z dwóch złączy spolaryzowane w kierunku przewodzenia.

Porównanie diod z ujemną rezystancją dynamiczną

12.3

Zakresy zastosowań poszczególnych diod ilustruje tabl. 12.1. Dla zastosowań w generatorach najważniejszymi parametrami są:

- moc i częstotliwość; najczęściej za wskaźnik jakości przyjmuje się iloczyn mocy przez kwadrat częstotliwości Pf^2 (rys. 12.12);
- sprawność przemiany energii prądu stałego w energię prądu zmiennego.

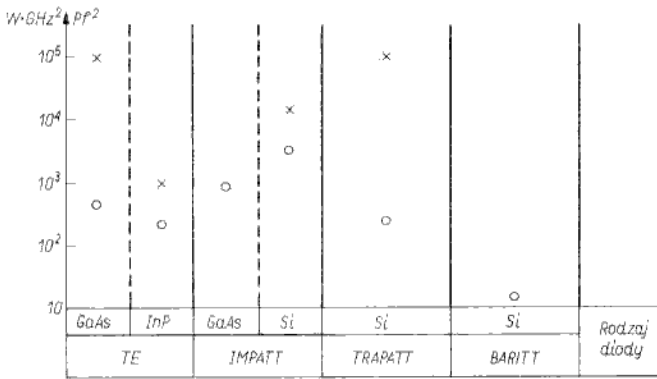
Zakresy mocy i częstotliwości uzyskiwane dla poszczególnych diod przedstawiono na rys. 12.13.

Diody TE umożliwiają uzyskanie dużych mocy przy dużych częstotliwościach (do 100 GHz), a ich sprawność wynosi kilka procent dla pracy ciągłej oraz 20...30%

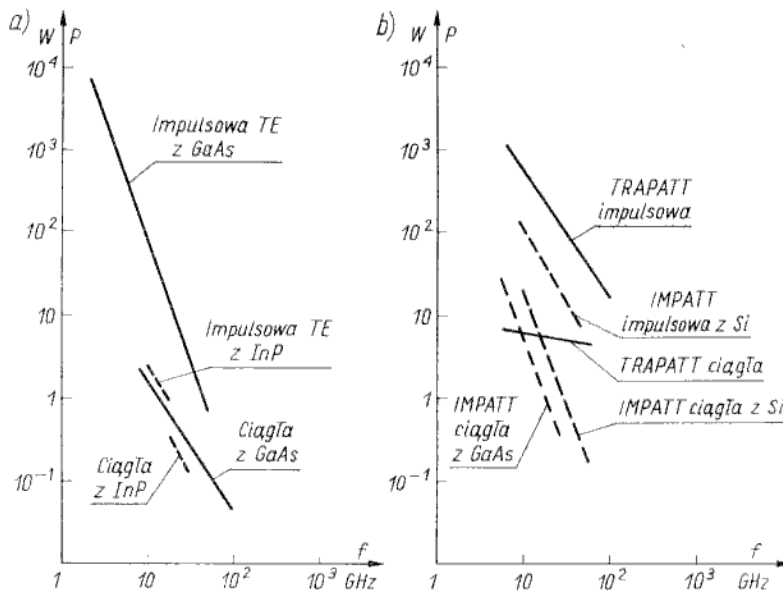
Zakresy zastosowań diod z ujemną rezystancją dynamiczną

Tablica 12.1

Rodzaj diody	Materiał	Generatory			Wzmacniacze				Radary
		lokalne	nadajniki impulsowe	nadajniki fali ciągłej	nisko-szumne	mocy fali ciągłej	mocy impulsowe	parametryczne	
TE	GaAs	+	+	+	+	+	+	+	+
	InP								
IMPATT	GaAs								
	Si	+	+	+		+	+	+	+
TRAPATT	Si		+				+		
BARITT	Si	+		+	+				+



Rys. 12.12
Typowe wartości iloczynu Pf^2 dla pracy impulsowej i ciągłej
x — praca impulsowa; o — praca ciągła



Rys. 12.13
Wykresy $P(f)$ dla diod TE (a) oraz lawinowo-przelotowych (b)

dla pracy impulsowej. Zaletą tych diod jest małe napięcie polaryzacji (ok. 10 V) oraz możliwość przestrajania częstotliwości w szerokim zakresie (dla modu LSA).

Diody IMPATT umożliwiają uzyskanie jeszcze większych mocy, lecz wymagają zasilania większym napięciem (ok. 100 V). Sprawność tych diod wynosi 10...40%.

Największe moce przy pracy impulsowej i zarazem największe sprawności (20...60%) otrzymuje się dla diod TRAPATT.

Diody BARITT służą do generacji fali ciągłej małej mocy (kilkaset miliwoltów) o częstotliwości nie większej niż 12 GHz przy sprawności ok. 5%. Z uwagi na dużą czułość detekcji przy pracy tych diod w układzie oscylatora-mieszacza nadają się one szczególnie do zastosowań w małych radarach (milicyjnych, itp.).

Zalety i wady poszczególnych diod pod kątem ich zastosowań we wzmacniaczach są następujące. Wzmacniacz z diodami TE ma szerokie pasmo przenoszenia (ok. 30% w stosunku do częstotliwości centralnej), dobrą liniowość fazy i amplitudy, małe napięcie zasilania i małe szumy. Wadami tych wzmacniaczy są małe moce na wyjściu i niezbyt dobra stabilność temperaturowa. Z kolei diody IMPATT mają bardzo dobrą stabilność temperaturową i umożliwiają uzyskanie większych mocy wyjściowych. Wadami wzmacniaczy zbudowanych z diod IMPATT są: małe pasmo przenoszenia (kilka procent w stosunku do częstotliwości centralnej), duże napięcie zasilania i złożony układ włączenia tych diod. Diody TRAPATT mają podobne właściwości jak diody IMPATT, przy czym wymagają jeszcze bardziej złożonego układu włączenia. Podstawową zaletą diod BARITT jest ich niski poziom szumów.

Wszystkie charakterystyki i parametry końcówkowe tranzystorów zmieniają się w mniejszym lub większym stopniu w funkcji temperatury. Wyrażenia analityczne, opisujące charakterystyki i parametry tranzystorów, zawierają informacje o zależnościach temperaturowej częściowo w postaci jawnej (jako zależność od potencjału elektrokinetycznego $\varphi_T = kT/q$) oraz częściowo w postaci niejawnej poprzez parametry fizyczne będące funkcjami temperatury. Do najważniejszych parametrów fizycznych podlegających zmianom w funkcji temperatury i mających dominujący wpływ na właściwości końcówkowe tranzystora należy zaliczyć:

- koncentrację nośników samoistnych

$$n_i \sim \exp\left(-\frac{W_g}{2kT}\right)$$

- ruchliwość nośników

$$\mu \sim \frac{1}{T^a}$$

- współczynnik dyfuzji

$$D \sim \frac{1}{T^{1-a}}$$

- czas życia nośników

$$\tau_r \sim T^b$$

- Potencjał Fermiego

$$\varphi_F(T) = \varphi_F(T_0) - cT$$

Wyżej podane wyrażenia, w których współczynniki a , b , c są liczbami większymi niż zero, mogą służyć do szacunkowej oceny wpływu temperatury na poszczególne parametry tranzystorów.

Tranzystory bipolarne

A.1

W tranzystorach bipolarnych największe znaczenie mają zmiany temperaturowe następujących trzech wielkości:

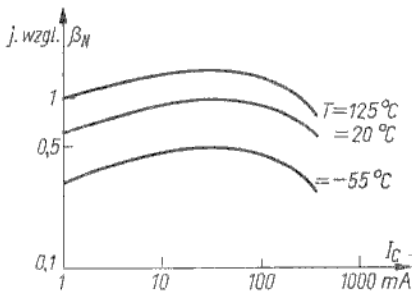
- prądu zerowego I_{CBO} ;
- napięcia U_{EB} przy stałym prądzie I_E ;
- współczynnika wzmocnienia prądowego.

Pierwsze dwie zależności, tj. $I_{CBO}(T)$ oraz $U_{EB}(T)|_{I_E=\text{const}}$ są identyczne jak dla złącza $p-n$ (por. p. 3.1.8).

Prąd I_{CBO} (odpowiednik prądu I_R w złączu $p-n$) zmienia się wykładniczo w funkcji temperatury wskutek zależności temperaturowej koncentracji nośników samoistnych. Dla celów praktycznych można przyjąć w przybliżeniu, że prąd I_{CBO} zwiększa się 2-krotnie przy wzroście temperatury o 10°C .

Dryf napięcia U_{EB} przy $I_E = \text{const}$ wynosi ok. $-2 \text{ mV}/1^\circ\text{C}$.

Zmiany współczynnika wzmocnienia prądowego (α lub β) są spowodowane głównie zależnościami temperaturowymi czasu życia nośników i współczynnika dyfuzji. Czas życia τ_r zwiększa się w funkcji temperatury (im wyższa jest temperatura, tym większe jest prawdopodobieństwo powtórnego wygenerowania nośników przechwyconych przez centra generacyjno-rekombinacyjne), a współczynnik dyfuzji maleje. Zwykle dominujące znaczenie ma zależność $\tau_r(T)$, dlatego współczynnik wzmocnienia prądowego zwiększa się w funkcji temperatury. Ilustruje to typowy wykres $\beta_N(I_C)$ przedstawiony na rys. A.1 dla trzech różnych



Rys. A.1
Typowy wykres $\beta_N(I_C)$ dla trzech różnych temperatur

temperatur. Zwiększenie się współczynnika wzmocnienia prądowego w funkcji temperatury może doprowadzić do uszkodzenia tranzystora wskutek tzw. niestabilności cieplnej (ang. *thermal runaway*), jeżeli nie ma odpowiedniego ograniczenia prądu w układzie. Niestabilność cieplna polega na tym, że przy wydzielaniu dużej mocy tranzystor się nagrzewa, a ze wzrostem temperatury zwiększa się β_N , czyli wzrasta wartość prądu kolektora; zwiększa się więc moc wydzielana w tranzystorze i tranzystor jeszcze bardziej się nagrzewa, aż do uszkodzenia tranzystora.

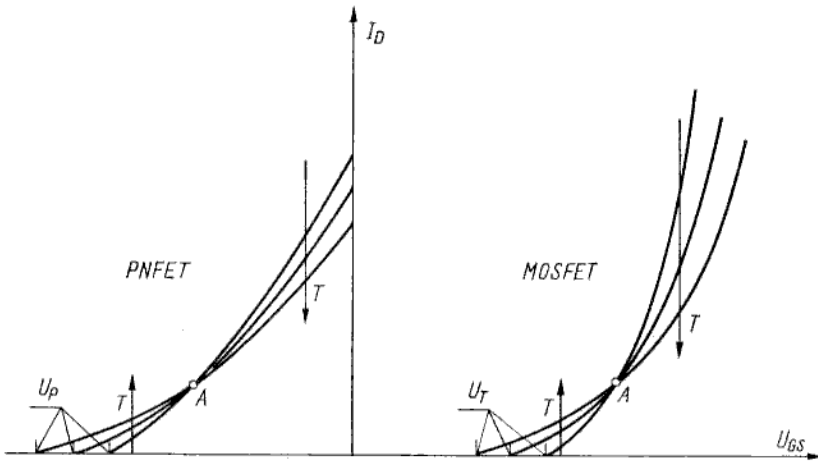
Tranzystory polowe

A.2

W tranzystorach polowych największe znaczenie mają zmiany temperaturowe charakterystyki przejściowej $I_D(U_{GS})$, które przedstawiono na rys. A.2 dla tranzystorów PNFET (z kanałem typu n) i MOSFET (z kanałem wzbogacanym

typu n). Współczynnik temperaturowy prądu drenu ($TKI_D = \left. \frac{1}{I_D} \frac{\partial I_D}{\partial T} \right|_{U_{GS}}$)

we wszystkich tranzystorach polowych może przyjmować wartości zarówno dodatnie jak i ujemne oraz zerowe. Zależność temperaturowa prądu drenu jest określona przez zmiany temperaturowe następujących dwu parametrów fizycznych:

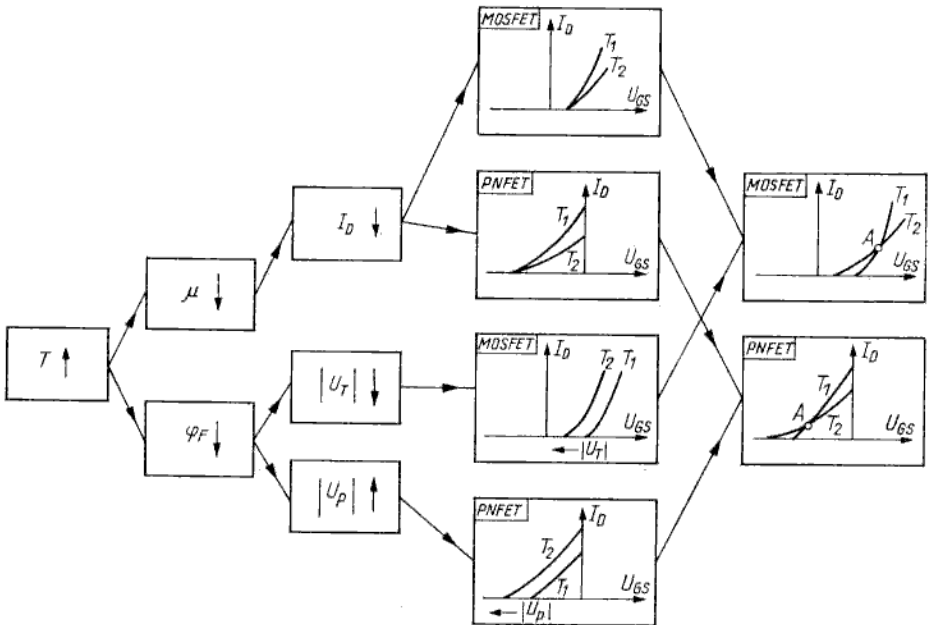


Rys. A.2

Zmiany charakterystyk przejściowych $I_D(U_{GS})$ w funkcji temperatury dla tranzystorów PNFET i MOSFET (zwroty strzałek przy T oznaczają wzrost temperatury)

- ruchliwości nośników w kanale μ ;
- potencjału Fermiego φ_F ;

Wskutek zmniejszania się ruchliwości prąd drenu maleje w miarę wzrostu temperatury. Wpływ drugiego czynnika, tj. potencjału Fermiego na dryf charakterystyki



Rys. A.3

Poglądowy schemat interpretacji wpływu temperatury na charakterystyki $I_D(U_{GS})$ tranzystorów polowych

$I_D(U_{GS})$ jest bardziej złożony. W tranzystorach PNFET i MOSFET z kanałem zubożonym zmniejszanie się potencjału Fermiego w miarę wzrostu temperatury powoduje zmniejszanie bariery potencjału φ_B w złączu $p-n$, a więc zmniejszanie szerokości warstwy zubożonej, przeto zwiększanie szerokości kanału. Z kolei przy większej szerokości kanału jest konieczna większa wartość napięcia bramki do spowodowania odcięcia (zatkania) tranzystora, zwiększa się zatem wartość bezwzględna napięcia zatkania U_P . W tranzystorach MOS z kanałem wzbogaconym zmniejszanie się potencjału Fermiego w miarę wzrostu temperatury powoduje zmniejszanie wartości bezwzględnej napięcia progowego U_T , gdyż przy mniejszej wartości φ_F jest wymagane mniejsze napięcie bramki do spowodowania inwersji (do otwarcia kanału). Reasumując powyższe rozważania można przedstawić skrótowy schemat interpretacji zmian charakterystyk $I_D(U_{GS})$ we wszystkich tranzystorach polowych (rys. A.3). Ponieważ w zakresie dużych wartości prądu drenu współczynnik TKI_D ma znak ujemny, w przeciwieństwie do tranzystorów bipolarnych nie istnieje niebezpieczeństwo zniszczenia tranzystorów polowych wskutek niestabilności cieplnej.

SZUMY W TRANZYSTORACH

W tranzystorach, jak i w innych elementach elektronicznych, powstają szumy własne, które ograniczają możliwości pracy układów tranzystorowych w zakresie małych sygnałów elektrycznych. Najpierw rozpatrzmy zagadnienia ogólne, tj. fizyczne źródła szumów i parametry charakteryzujące właściwości szumowe elementu czynnego, po czym omówimy właściwości szumowe tranzystorów bipolarnych i polowych.

Fizyczne źródła szumów

B.1

W każdym elemencie elektronicznym można wyróżnić trzy składowe szumu, tj.:

- szum cieplny;
- szum śrutowy;
- szum „ $1/f$ ”;

Szum cieplny, nazywany niekiedy *szumem Johnsona*, jest spowodowany chaotycznym ruchem cieplnym nośników ładunku w warunkach równowagi termodynamicznej. Ruch cieplny nośników powoduje powstanie napięcia fluktuacyjnego na brzegach każdego elementu rezystywnego. Wartość średnia tego napięcia jest oczywiście równa zero, a wartość średnia kwadratu napięcia fluktuacyjnego na elemencie o rezystancji R , w przedziale częstotliwości Δf , jest określona wzorem Nyquista

$$\overline{u_n^2} = 4kTR\Delta f \quad (\text{B.1})$$

Wartość średnią kwadratu prądu można zapisać w postaci

$$\overline{i_n^2} = \frac{4kT\Delta f}{R} \quad (\text{B.2})$$

Ze wzorów (B.1), (B.2) wynika, że element rzeczywisty o rezystancji R można zastąpić źródłem napięciowym $\overline{u_n^2}$, połączonym szeregowo z rezystancją „bezszumową” R lub źródłem prądowym $\overline{i_n^2}$, połączonym równolegle z konduktancją „bezszumną” $G = 1/R$.

Moc szumu cieplnego

$$P_n = \frac{\overline{u_n^2}}{4R} = kT\Delta f \quad (\text{B.3})$$

a jej gęstość widmowa

$$\frac{dP_n}{df} = kT \quad (\text{B.4})$$

jest niezależna od częstotliwości. Szum o stałym rozkładzie widmowym mocy jest nazywany *szumem białym*.

Szum śrutowy jest związany z dyskretną postacią ładunków elektrycznych i chaotycznymi procesami ich generacji-rekombinacji, co powoduje fluktuacje liczby nośników prądu. Zmiany statystyczne ładunku przechodzącego przez określoną powierzchnię można przedstawić za pośrednictwem wartości średniej kwadratu prądu fluktuacyjnego

$$\overline{i_n^2} = 2qI\Delta f \quad (B.5)$$

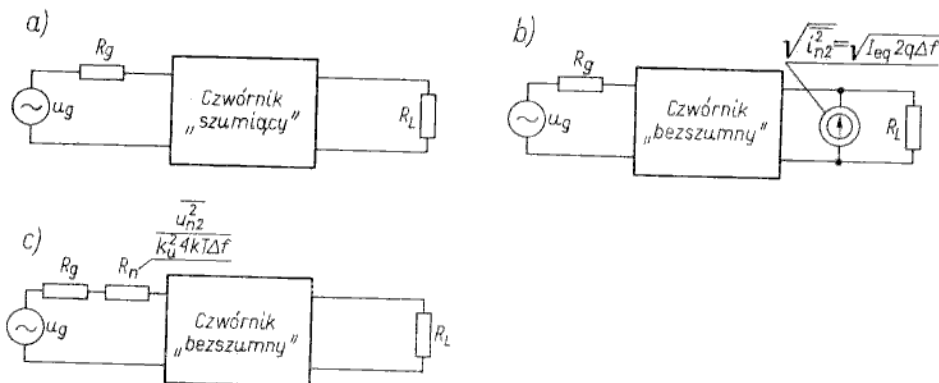
przy czym I — składowa prądu stałego, płynącego przez daną powierzchnię. Szum śrutowy jest również szumem białym, tj. gęstość widmowa jego mocy nie zależy od częstotliwości.

Termin szum „ $1/f$ ” odnosi się do takiego rodzaju szumu, którego gęstość widmowa jest odwrotnie proporcjonalna do częstotliwości. Ten rodzaj szumu dominuje w zakresie małych częstotliwości, przy większych natomiast częstotliwościach jest on maskowany szumem białym (śrutowym i cieplnym). Istota fizyczna szumu „ $1/f$ ” nie jest do końca wyjaśniona, wiadomo jednak, że jest on spowodowany złożonymi zjawiskami pułapkowania nośników na powierzchni półprzewodnika. W miarę postępu technologii, zwłaszcza w zakresie metod obróbki i pasywacji powierzchni półprzewodnika, szum „ $1/f$ ” udaje się znacznie zmniejszyć.

Parametry charakteryzujące właściwości szumowe elementu czynnego

B.2

Dla oceny szumów generowanych przez element czynny rozpatruje się czwórnik liniowy „szumiący” o wzmacnieniu napięciowym k_u , sterowany napięciem u_g ze źródła o rezystancji R_g oraz obciążony rezystorem R_L (rys. B.1a). W takim



Rys. B.1

Element czynny przedstawiony jako czwórnik „szumiący”, sterowany ze źródła o rezystancji R_g oraz obciążony rezystancją R_L (a), b), c) układy równoważne schematowi a)

obwodzie źródłami szumów są rezystory R_g , R_L oraz czwórnik, przy czym szumy czwornika (elementu czynnego) zależą od wartości rezystancji R_g i R_L . Szumy elementu czynnego opisuje się zwykle jednym z następujących parametrów:

- prądem równoważnym diody nasyczonej;
- rezystancją równoważną szumów;
- temperaturą szumów;
- współczynnikiem szumów;

Szumy czwórnika „szumiącego” można przedstawić w postaci generatora prądu i_{n2} , włączonego na wyjściu czwórnika „bezszumowego” (rys. B.1b), równolegle z obciążeniem R_L . Wydajność prądową tego źródła można opisać pośrednio za pomocą równoważnego prądu stałego, zgodnie ze wzorem (B.5)

$$I_{eq} = \frac{\overline{i_{n2}^2}}{2q\Delta f} \quad (\text{B.6})$$

Ponieważ dioda próżniowa, pracująca w zakresie nasycenia, generuje szum śrutowy zgodnie z zależnością (B.5), prąd I_{eq} nazywa się *prądem równoważnym diody nasyczonej*.

Inny sposób opisu właściwości szumowych elementu czynnego polega na tym, że napięcie u_{n2} na wyjściu czwórnika „szumiącego”, pochodzące od szumów elementu czynnego, sprowadza się do wejścia dzieląc je przez wzmocnienie napięciowe czwórnika. Otrzymuje się w ten sposób napięcie sygnału równoważnego u_{n1} , włączonego na wejściu czwórnika „bezszumowego”. Zamiast źródła napięciowego u_{n1} na wejściu czwórnika „bezszumowego” włącza się rezystor o takiej wartości rezystancji R_n , by średnia wartość kwadratu napięcia jego szumu cieplnego w temperaturze pokojowej ($T = 290 \text{ K}$) była równa $\overline{u_{n1}^2}$

$$R_n = \frac{\overline{u_{n1}^2}}{4kT\Delta f} = \frac{\overline{u_{n2}^2}}{k_u^2 4kT\Delta f} \quad (\text{B.7})$$

Wielkość R_n (rys. B.1c) nazywa się *rezystancją równoważną szumów elementu czynnego*.

Przy określonej wartości rezystancji źródła sygnału R_g można zastąpić rezystancję równoważną R_n rezystancją R_g umieszczoną w takiej temperaturze fikcyjnej, że jej szum cieplny jest równy szumowi cieplnemu rezystancji R_n (czyli również elementu czynnego) w temperaturze pokojowej

$$4kT_n R_g \Delta f = 4k(290 \text{ K}) R_n \Delta f$$

Stąd

$$T_n = \frac{R_n}{R_g} 290 \text{ K} \quad (\text{B.8})$$

Wielkość T_n nazywa się *temperaturą szumów*.

Najczęściej stosowanym parametrem charakteryzującym właściwości szumowe elementu czynnego jest *współczynnik szumów F* , definiowany jako stosunek mocy całkowitej szumu na wyjściu czwórnika do tej części mocy szumu, jaka pozostałaby, gdyby jedynym elementem wywołującym szum układu była rezystancja źródła sygnału R_g

$$F = \frac{\text{całkowita moc szumów dostarczona do obciążenia czwórnika}}{\text{moc szumów dostarczona do obciążenia, jeżeli źródłem szumów jest tylko rezystor } R_g \text{ (dla } T = 290 \text{ K)}} \quad (\text{B.9})$$

Ten iloraz jest zwykle wyznaczany dla wąskiego pasma częstotliwości i nazywany *wąskopasmowym* (punktowym, różniczkowym) *współczynnikiem szumów*. Współczynnik szumów wyraża się najczęściej w decybelach, zgodnie z zależnością

$$F_{dB} = 10 \lg F$$

Ponieważ obie składowe mocy szumów (szumy źródła sygnału o rezystancji R_g i szumy elementu czynnego) są rozpraszane na wspólnym rezystorze R_L , można przejść od stosunku mocy do stosunku średnich kwadratów napięcia

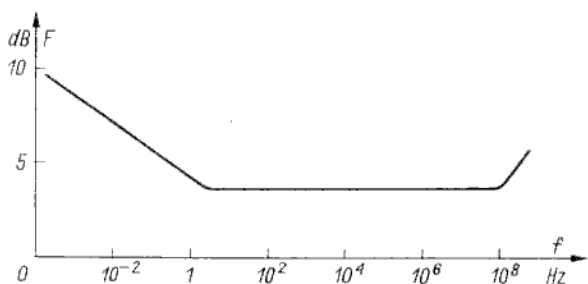
$$F = \frac{\overline{u_{n2}^2} + k_u^2 \cdot 4kTR_g \Delta f}{k_u^2 \cdot 4kTR_g \Delta f} = 1 + \frac{\overline{u_{n2}^2}}{k_u^2 \cdot 4kTR_g \Delta f} = 1 + \frac{R_n}{R_g} \quad (\text{B.10})$$

Teraz scharakteryzujemy właściwości szumowe tranzystorów bipolarnych i polowych przyjmując współczynnik F za podstawowy parametr.

Szumy w tranzystorach bipolarnych

B.3

Podstawową charakterystyką szumową tranzystorów jest zależność współczynnika szumów F od częstotliwości f (rys. B.2). W zakresie małych częstotliwości ($f \leq 100$ Hz dla tranzystorów planarnych) przeważa składowa szumu typu „ $1/f$ ”, spowo-



Rys. B.2
Typowa zależność współczynnika F od częstotliwości f dla tranzystorów bipolarnych

dowanego głównie rekombinacją powierzchniową nośników. W bardzo szerokim zakresie częstotliwości ($100 \text{ Hz} \leq f \leq f_\beta$) charakterystyka $F(f)$ jest płaska, czyli występuje tylko szum biały (ciepły i śrutowy). Szum ciepły jest generowany przez rezystancje rozproszone obszarów neutralnych emitera, bazy i kolektora¹⁾, przy czym dominujące znaczenie ma rezystancja rozproszona bazy $r_{bb'}$, która generuje napięcie fluktuacji

$$\overline{e_{bb'}^2} = 4kTr_{bb'} \Delta f \quad (\text{B.11})$$

Szumy śrutowe pochodzą z fluktuacji wszystkich prądów składowych płynących w tranzystorze, przy czym dominujący wpływ mają szumy dwu składowych, tj.:

— strumienia nośników wstrzykiwanych z emitera i odbieranych przez kolektor

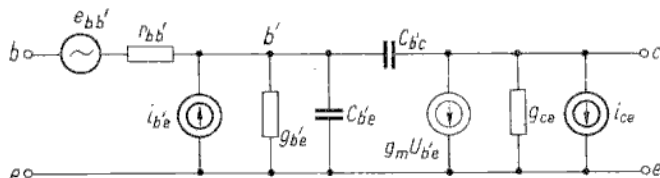
$$\overline{i_{ce}^2} = 2q\alpha_N I_E \Delta f \quad (\text{B.12})$$

¹⁾ Konduktancje (rezystancje) różniczkowe występujące w schemacie zastępczym „hybryd π ” (na przykład g_{be} , g_{ce} itp.), które nie reprezentują rzeczywistych obszarów półprzewodnika, lecz tylko wynikają z określonych zależności między prądami a napięciami, nie są źródłami szumu cieplnego.

— strumienia nośników rekombinujących w obszarze bazy

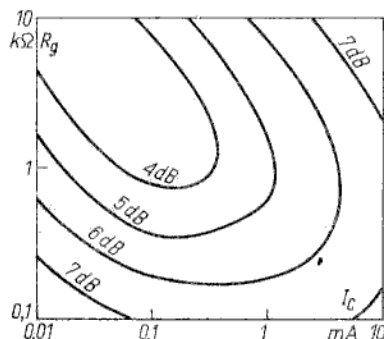
$$\overline{i_{b'e}^2} = 2q(1 - \alpha_N)I_E \Delta f \quad (\text{B.13})$$

Na rysunku B.3 przedstawiono uproszczony schemat zastępczy „hybryd π ”, na którym uwzględniono trzy źródła szumów określone wzorami (B.11), (B.12), (B.13).



Rys. B.3
Schemat zastępczy „hybryd π ” z uwzględnieniem źródeł szumów

W zakresie dużych częstotliwości ($f > f_\beta$) współczynnik szumów F zwiększa się wskutek zmniejszania się wzmacnienia, gdyż moc sygnału użytecznego (wzmacnianego) na wyjściu czwórnika maleje, a moc szumów własnych tranzystora pozostaje niemal bez zmian. Ponieważ współczynnik szumów zależy od prądu kolektora i rezystancji źródła sterującego, często w katalogach podaje się wykresy $F(R_g, I_C)$, które umożliwiają wybranie optymalnych warunków pracy tranzystora. Typowy wykres $F(R_g, I_C)$ przedstawiono na rys. B.4.



Rys. B.4
Typowy wykres zależności współczynnika szumów F od rezystancji źródła sterującego R_g i prądu kolektora I_C przy $U_{CE} = \text{const}$ oraz $f = \text{const}$

Szumy w tranzystorach polowych

B.4

Charakterystyka częstotliwościowa współczynnika szumów $F(f)$ dla tranzystorów polowych ma ogólny przebieg podobny jak dla tranzystorów bipolarnych (rys. B.2). W zakresie małych częstotliwości przeważa szum typu „ $1/f$ ”, który jest szczególnie duży w tranzystorach MOS. Dla większych częstotliwości ($f \geq 100$ Hz w przypadku tranzystorów PNFET oraz $f \geq 10$ kHz w przypadku tranzystorów MOS) podstawowym rodzajem szumu jest szum cieplny kanału. Rezystancja kanału jest równa odwrotności transkonduktancji g_m , zatem średnia wartość kwadratowa fluktuacji prądu drenu

$$\overline{i_d^2} = 4kTg_m \Delta f \quad (\text{B.14})$$

Ponieważ nie ma szumu śrutowego (z wyjątkiem niewielkiej składowej szumu spowodowanej prądem wstecznym złącza $p-n$ w przypadku tranzystorów PNFET), tranzystory polowe mają niższy poziom szumów niż tranzystory bipolarne.

OZNACZENIA ELEMENTÓW PÓLPRZEWODNIKOWYCH I UKŁADÓW SCALONYCH

C

Zasady oznaczania typów elementów półprzewodnikowych i układów scalonych produkcji krajowej ustalają następujące dwie normy branżowe:

- BN-70/3375-13 pt. „Elementy półprzewodnikowe. System oznaczania typów”.
- BN-73/3375-21, pt. „Mikroukłady scalone. System oznaczania typów”.

Najważniejsze ustalenia zawarte w tych normach podano poniżej.

System oznaczania elementów półprzewodnikowych

C.1

Oznaczenie elementu półprzewodnikowego składa się z dwu części, tj.:

- części literowej zawierającej dwie litery;
- numeru seryjnego zawierającego jedną literę i trzy cyfry lub dwie litery i dwie cyfry.

W części literowej pierwsza litera określa rodzaj materiału, z jakiego wykonano element, zgodnie z poniższym zestawieniem:

- A — materiał o szerokości pasma zabronionego 0,6...1,0 eV, taki jak german;
- B — materiał o szerokości pasma zabronionego 1,0...1,3 eV, taki jak krzem;
- C — materiał o szerokości pasma zabronionego większej niż 1,3 eV, taki jak arsenek galu;
- D — materiał o szerokości pasma zabronionego mniejszej niż 0,6 eV, taki jak antymonek indu;
- R — inne materiały.

Druga litera określa rodzaj elementu półprzewodnikowego, zgodnie z poniższym zestawieniem:

- B — diody detekcyjne, mieszające i szybko przełączające;
- A — diody o zmiennej pojemności (warikapy);
- C — tranzystory małej mocy, małej częstotliwości ($R_{thj-c} > 15^{\circ}\text{C/W}$);
- D — tranzystory mocy, małej częstotliwości ($R_{thj-c} \leq 15^{\circ}\text{C/W}$);
- E — diody tunelowe;
- F — tranzystory małej mocy, wielkiej częstotliwości ($R_{thj-c} > 15^{\circ}\text{C/W}$);
- G — elementy powielające złożone z różnych struktur;
- H — sondy do pomiaru natężenia pola (czujnik Halla);
- K — generatory Halla o otwartym obwodzie magnetycznym (np. czujnik sygnałowy, magnetometr);
- L — tranzystory mocy, wielkiej częstotliwości ($R_{thj-c} \leq 15^{\circ}\text{C/W}$);
- M — generatory Halla o zamkniętym obwodzie magnetycznym (np. modulator lub powielacz Halla);
- P — elementy czułe na promieniowanie (np. fotodiody);
- Q — elementy promieniujące (np. dioda luminescencyjna);

- R — tyrystory małej mocy ($R_{thj-c} > 15^{\circ}\text{C/W}$);
- S — tranzystory impulsowe, małej mocy ($R_{thj-c} > 15^{\circ}\text{C/W}$);
- T — tyrystory mocy ($R_{thj-c} \leq 15^{\circ}\text{C/W}$);
- U — tranzystory impulsowe mocy ($R_{thj-c} \leq 15^{\circ}\text{C/W}$);
- Y — diody prostownicze, tłumiąco-usprawniające;
- X — diody powielające;
- Z — stabilistory (diody Zenera).

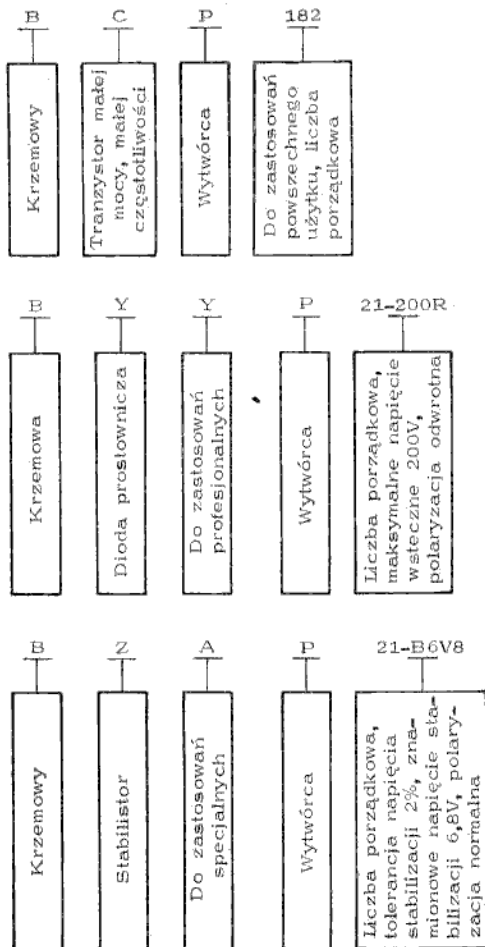
Numer seryjny, oprócz liczby porządkowej określającej grupę oraz konkretny typ elementu w danej grupie, zawiera informacje o przeznaczeniu i wytwórcy wyrobu, w szczególnych przypadkach również o wartościach niektórych parametrów elementu.

W zależności od przeznaczenia numer seryjny ma następującą budowę: P+3 cyfry — dla elementów do zastosowań powszechnego użytku;

YP+2 cyfry — dla elementów do zastosowań profesjonalnych;

AP+2 cyfry — dla elementów do zastosowań specjalnych.

Dla elementów profesjonalnych można zamiast litery Y stosować litery Z, X, W itd., a dla elementów specjalnych zamiast litery A dalsze litery B, C, D itd.



Rys. C.1
Przykłady ilustrujące system oznaczeń elementów półprzewodnikowych

Litera P jest umownym znakiem wytwórcy (Naukowo-Produkcyjnego Centrum Półprzewodników).

Oznaczenia typów stabilizatorów, diod prostowniczych, tyrystorów zawierają ponadto symbole informujące o wartościach niektórych parametrów.

W przypadku stabilizatorów litera poprzedzona znakiem „-” określa tolerancję napięcia stabilizacji według następującego klucza:

A — 1%

B — 2%

C — 5%

D — 10%

E — 15%

Po tej literze następują cyfry określające wartość nominalnego napięcia stabilizacji w woltach. Literę V stosuje się zamiast przecinka, jeżeli napięcie stabilizacji jest liczbą ułamkową. Jeżeli dioda ma odwrotną polaryzację (obudowę połączoną z anodą), to na końcu oznaczenia podaje się literę R. Polaryzacji normalnej (obudowa połączona z katodą) nie oznacza się.

W przypadku diod prostowniczych i tyrystorów cyfry poprzedzone znakiem „-” określają maksymalną wartość impulsowego napięcia wstecznego w woltach. Identycznie jak w oznaczeniach stabilizatorów litera R oznacza odwrotną polaryzację.

Na rysunku C.1 przedstawiono przykłady ilustrujące opisany sposób oznaczeń elementów półprzewodnikowych.

Oznaczenia układów scalonych

C.2

Oznaczenie typu układu scalonego składa się z dwu części, tj.:

— części literowej zawierającej dwie lub trzy litery;

— numeru seryjnego zawierającego cztery lub pięć cyfr i jedną literę.

W części literowej pierwsza litera określa grupę układów ze względu na sposób ich wykonania (podział technologiczny), zgodnie z poniższym zestawieniem:

U — układ półprzewodnikowy, monolityczny bipolarny;

H — układ hybrydowy;

M — układ ze strukturami MOS.

Druga litera określa grupę układów ze względu na spełnianą funkcję (podstawowy podział funkcjonalny), zgodnie z poniższym zestawieniem:

C — układy cyfrowe;

L — układy analogowe;

R — układy inne (np. mieszane biernie).

Trzecia litera (lub jej brak) określa grupę układów ze względu na ich przeznaczenie, zgodnie z poniższym zestawieniem:

brak litery — do zastosowań powszechnego użytku;

Y — do zastosowań profesjonalnych;

A — do zastosowań specjalnych;

T — do zastosowań profesjonalnych o zwiększonej niezawodności;

Q — do zastosowań specjalnych o zwiększonej niezawodności;

X — prototypowe, doświadczalne lub na zamówienie.

W numerze seryjnym pierwsza cyfra określa zakres dopuszczalnej temperatury pracy, według następującego klucza:

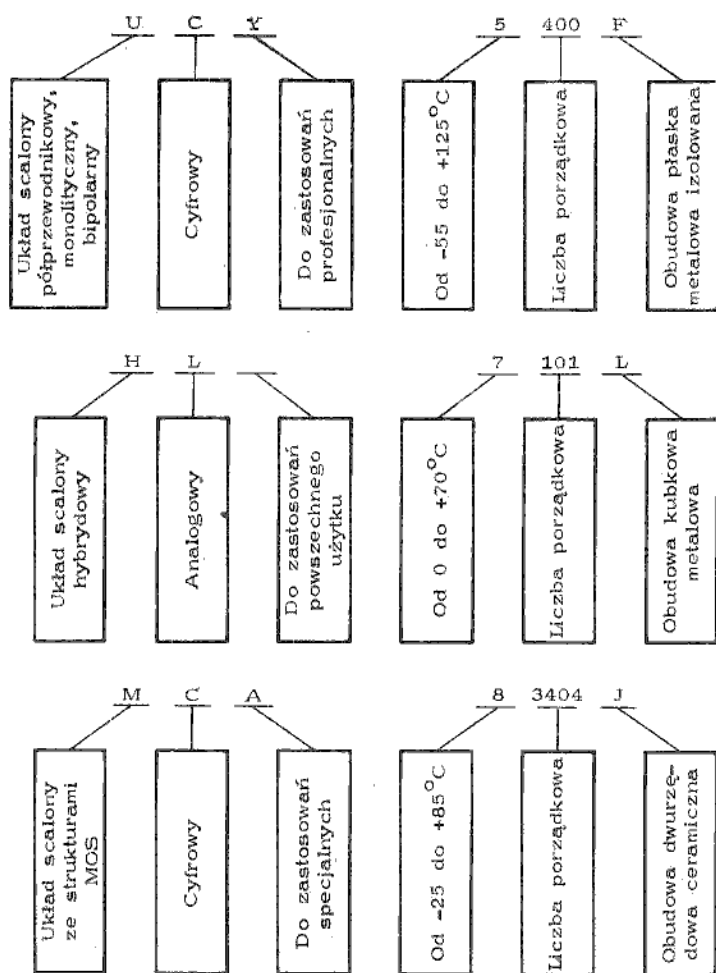
4 — od -55 do $+85^{\circ}\text{C}$;

5 — od -55 do $+125^{\circ}\text{C}$;

- 6 — od -40 do $+85^{\circ}\text{C}$;
- 7 — od 0 do $+70^{\circ}\text{C}$;
- 8 — od -25 do $+85^{\circ}\text{C}$;
- 1 — inny zakres.

Pozostałe trzy lub cztery cyfry stanowią liczbę porządkową określającą grupę oraz konkretny typ elementu w danej grupie. W uzasadnionych przypadkach dopuszcza się wprowadzenie do liczby porządkowej jednej litery dla określenia właściwości charakterystycznych układu, przy czym należy stosować następujące litery dla układów:

- serii szybkiej H;
- serii małej mocy L;
- serii bardzo szybkiej S;



Rys. C. 2
Przykłady ilustrujące system oznaczeń układów scalonych

Litera umieszczona na końcu numeru seryjnego określa rodzaj obudowy, według następującego zestawienia:

F — obudowa płaska, metalowa, izolowana od układu;

S — obudowa płaska, metalowa, mająca kontakt elektryczny z podłożem układu i wyprowadzeniem masy;

H — obudowa płaska z nieprzewodzącego materiału ceramicznego;

J — obudowa dwurzędowa z nieprzewodzącego materiału ceramicznego;

N — obudowa dwurzędowa plastikowa;

L — obudowa kubkowa, metalowa z wyprowadzeniami umieszczonych kołowo;

R — obudowa inna.

Na rysunku C.2 przedstawiono trzy przykłady ilustrujące opisany sposób oznaczeń układów scalonych.

WYBRANE STAŁE FIZYCZNE I JEDNOSTKI

Stała	Oznaczenie	Wartość
Ładunek elektronu	q	$1,602 \cdot 10^{-19} \text{ C}$
Stała Plancka	h	$6,626 \cdot 10^{-34} \text{ J} \cdot \text{s}$
Stała Boltzmannna	k	$1,381 \cdot 10^{-23} \text{ J/K}$
		$8,62 \cdot 10^{-5} \text{ eV/K}$
Stała elektryczna próżni	ϵ_0	$8,849 \cdot 10^{-12} \text{ F/m}$
Masa elektronu w próżni	m_0	$0,911 \cdot 10^{-30} \text{ kg}$

$$1 \text{ eV} = 1,6 \cdot 10^{-12} \text{ erg} = 1,6 \cdot 10^{-19} \text{ J}$$

$$1 \text{ m} = 10^2 \text{ cm} = 10^6 \mu\text{m} = 10^{10} \text{ \AA}$$

$$1 \text{ mil (ang.)} = 10^{-3} \text{ cala} = 25,4 \mu\text{m}$$

$$kT \text{ w temperaturze pokojowej } (T = 300 \text{ K}) = 0,0259 \text{ eV}$$

LITERATURA

1. Alley C. L., Atwood K. W.: *Elementy i układy półprzewodnikowe*. Warszawa 1975, WNT
2. Baranowski J.: *Półprzewodnikowe elementy układów impulsowych*. Warszawa 1969, WNT.
3. Baranowski J.: *Półprzewodnikowe układy impulsowe i cyfrowe*. Warszawa 1976, WNT.
4. Białko M.: *Układy mikroelektroniczne*. Warszawa 1969, WKiŁ.
5. Carr W. N., Mize J. P.: *Projektowanie i zastosowanie układów MOS wielkiej skali integracji*. Warszawa 1976, WNT.
6. Chramiec J., Wojtkiewicz A.: *Mikrofalowe mieszacze diodowe*. Warszawa 1975, WKiŁ.
7. Cobbold R. S. C.: *Teoria i zastosowanie tranzystorów polowych*. Warszawa 1975, WNT.
8. *Elektronika ciała stałego i kwantowa*. Pod red. W. D. Hershbergera. Warszawa 1976, WNT.
9. Filipkowski A.: *Układy elektroniczne analogowe i cyfrowe*. Warszawa 1978, WNT.
10. Golde W.: *Układy elektroniczne*. T. I. Warszawa 1974, WNT.
11. Golde W.: *Układy elektroniczne*. T. II. Warszawa 1976, WNT.
12. Golde W.: *Wzmacniacze tranzystorowe — małej częstotliwości, prądu stałego, szerokopasmowe*. Warszawa 1973, WNT.
13. Gray P. E., Searle C. L.: *Podstawy elektroniki*. Warszawa 1976, PWN.
14. Guziński A.: *Projektowanie i konstrukcja układów warstwowych*. Warszawa 1973, WKiŁ.
15. Guziński A.: *Technologie układów warstwowych*. Warszawa 1971, WKiŁ.
16. Jępijanow G. I.: *Fizyczne podstawy mikroelektroniki*. Warszawa 1976, WNT.
17. Kalisz J.: *Cyfrowe układy scalone w technice systemowej*. Warszawa 1977, MON.
18. Kiriejew P. S.: *Fizyka półprzewodników*. Warszawa 1971, PWN.
19. Kittel C.: *Wstęp do fizyki ciała stałego*. Warszawa 1974, PWN.
20. Korzec Z.: *Tranzystory polowe*. Warszawa 1973, WNT.
21. Kulka Z., Nadachowski M.: *Liniowe układy scalone i ich zastosowanie*. Warszawa 1975, WKiŁ.
22. Linvill J. G.: *Modele tranzystorów i diod*. Warszawa 1967, WNT.
23. Luciński J.: *Układy tyrystorowe*. Warszawa 1978, WNT.
24. Łakomy M., Zabrodzki J.: *Cyfrowe układy scalone TTL*. Warszawa 1975, PWN.
25. Marciniak W.: *Przyrządy półprzewodnikowe typu TIS*. Warszawa 1977, WNT.
26. *Mikrofalowa elektronika ciała stałego*. Praca zbiorowa pod red. A. Jeleńskiego. Warszawa 1976, PWN.
27. Millman J., Halkias C. C.: *Układy scalone analogowe i cyfrowe*. Warszawa 1976, WNT.
28. Nowak E., Sawicki Z.: *Pamięci maszyn cyfrowych*. Warszawa 1972, WNT.
29. Pałycki B., Stefański W.: *Półprzewodnikowe stabilizatory napięcia i prądu stałego*. Warszawa 1970, MON.
30. *Pamięci półprzewodnikowe*. Praca zbiorowa pod red. J. Eimbindera. Warszawa 1975, WNT.
31. Paul R.: *Technika pomiarów tranzystorów*. Warszawa 1973, WKiŁ.
32. Pawłowski J.: *Podstawowe układy elektroniczne*. Cz. I. Warszawa 1975, WKiŁ.
33. *Procesy technologiczne w elektronice półprzewodnikowej*. Praca zbiorowa. Warszawa 1973, WNT.
34. Rosiński W.: *Zasady działania tranzystorów*. Warszawa 1977, WNT.
35. Seely S.: *Układy elektroniczne*. Warszawa 1972, WNT.
36. Sońta S., Kotlewski H.: *Układy scalone liniowe i ich zastosowanie*. Warszawa 1977, WNT.
37. Spiralski L., Kołodziejewski J.: *Miernictwo układów scalonych*. Warszawa 1977, WNT.
38. Streetman B. G.: *Przyrządy półprzewodnikowe*. Warszawa 1976, WNT.
39. Szalimowa K. W.: *Fizyka półprzewodników*. Warszawa 1974, PWN.
40. Szpakowski Z.: *Układy scalone w zastosowaniach*. Warszawa 1977, WKiŁ.

41. Świt A., Pułtorak J.: *Przyrządy półprzewodnikowe*. Warszawa 1976, WNT.
42. *Technika impulsowa*. Praca zbiorowa pod red. S. Sławińskiego. Warszawa 1973, WNT.
43. *Technika warstwowa w mikroelektronice*. Praca zbiorowa. Warszawa 1973, WNT.
44. Tietze U., Schenk Ch.: *Układy półprzewodnikowe*. Warszawa 1976, WNT.
45. Traczyk W.: *Układy cyfrowe automatyki*. Warszawa 1974, WNT.
46. Wolf H.: *Półprzewodniki*. Warszawa 1975, WNT.

Charakterystyki statyczne tranzystora bipolarnego w układzie wspólnego emiteera.

SKOROWIDZ

- Admitancyjna macierz nieoznaczona 308
akumulacja 150
algebra Boole'a 509
amorficzna budowa 22
amplituda logiczna 513
anizotropia 22
ażur 451
- Bariera potencjału 89
bramka 359, 383, 559
— ECL 524
— I²L 518
— logiczna 510
— — komplementarna 531
— — MOS 527
— — statyczna NMOS 528
— — z tranzystorami nasyconymi 511
— — — nienasyconymi 524
— TTL 513
— TTLS 524, 525
- Centra generacyjno-rekombinacyjne 50
charakterystyka przejściowa 360
— statyczna 272
— wyjściowa 360
czas magazynowania 139, 295, 296
— narastania 296
— opadania 139, 296
— opóźnienia 138, 294, 296
— przeciągania 295
— włączania 296
— wyłączenia 296
— życia nośników 48
częstotliwość graniczna 320, 492
czułość napięciowa 235
— prądowa 235
— progowa 234
— tangencjalna (styczna) 235
- Diak 552
dioda 88, 179
— BARIIT 570, 573
— detekcyjna 232
— dwubazowa 546
— Gunna 563
— IMPATT 570
dioda impulsowa 211
— ładunkowa 213
— miesząca 232
— mikrofalowa 232
— monolityczna 459
— o małym czasie opadania 213
— — — — przełączania 213
— — — — zmiennej impedancji (*p-i-n*) 239
— ostrzowa 236
— pojemnościowa 217
— prostownicza 184
— przelotowa 570
— Reada 571
— Schottky'ego 162, 236
— TE 562
— TRAPATT 570, 572
— TT 562
— tunelowa 225
— uniwersalna 199
— zwrotna 237
dolina centralna 563
— satelitarna 563
domena elektryczna 566
domieszki 29
— akceptorowe 30
— donorowe 30
domieszkowanie lokalne 76, 77
dren 359, 383
droga Debye'a 156
— dyfuzyjna 50
— ekranowania 156
dryf cieplny 492
dyfuzja 43, 66
— boczna 82
— lateralna 82
— lokalna 81
— międzywęzłowa 72
— z nieograniczonego źródła 67
— — ograniczonego źródła 68
— zamiennowęzłowa 69
dyfuzyjność 71
dynistor 552
dziura 24, 27, 28
- Efekt Schottky'ego 172
ekstrakcja 45
elektrony walencyjne 21
emisja wewnętrzna 116

- para hybrydowa 435
 parametry admitancyjne 304
 — hybrydowe 304
 — impedancyjne 304
 — rozproszenia 309
 pasmo przewodnictwa 26, 27
 — walencyjne 26, 27
 — zabronione 26
 pasywacja 52
 pasywowanie 78
 piec dyfuzyjny 72
 podłoże 383
 pojemność dyfuzyjna 127, 300
 — warstwy zaporowej 125
 — złączowa 125, 299
 pole elektryczne wbudowane 53, 250
 — wzmocnienia 334, 493
 polikrystaliczna budowa 21
 potencjał Fermiego 36
 — wyjścia 148
 powinowactwo elektronowe 148
 poziom akceptorowy 31
 — donorowy 30
 — energetyczny 24
 — Fermiego 34
 — jonizacji 24
 — wzbudzenia 25
 półprzewodnik 17
 — akceptorowy 31
 — donorowy 30
 — niejednorodny 53
 — niesamoistny 29, 30
 — samoistny 28
 — skompensowany 32
 — zdegenerowany 35
 praca wyjścia 147
 prawo Ficka 66
 prąd Esakiego 120
 — równoważny diody nasyconej 582
 — trzymania 559
 prądy zerowe 282
 predyfuzja 74
 prostownik dwupółkowy 195
 — jednopółkowy 191
 przebiecie lawinowe 117
 — skrośne 285
 — wtórne 288
 — Zenera 116
 przerwa energetyczna 26
 przybliżenie deplecyjne 96
 przyrząd funkcjonalny 445
 — o sprzężeniu ładunkowym (CCD) 542
 pułapka 32, 52
 rezystor emiterowy 462
 — ściśnięty 462
 rozpylanie katodowe 471
 — reaktywne 471
 równanie ciągłości ładunku 52
 ruchliwość 40
 Sieć krystaliczna 21
 sitodruk 473
 skracanie kanału 467
 spin 25
 sprawność transportu ładunku 545
 stabilitron 202
 stabilizator napięcia 502
 — nastawny 502
 — nienastawny 502
 stała czasowa bazy 322
 — — magazynowania 297
 — — przełączania 297
 — sieci krystalicznej 22
 stan energetyczny 24
 — powierzchniowy 52, 160
 straty przemiany 235
 struktura MIS 145
 szerokość efektywna bazy 256
 szum biały 581
 — cieplny 580
 — śrutowy 581
 — „1/f” 581
 szybkość generacji 46
 — narastania napięcia wyjściowego 492
 — rekombinacji 46
 — — powierzchniowej 52
 Tablica stanów 509
 technologia CMOS 468
 — COPLAMOS 467
 — ISOPLANAR 456
 — LOCOS 467
 — LOCOS 467
 — NMOS 466
 — planarna 82
 — PMOS 465
 — POLYPLANAR 457
 — RMOS 467
 — stopowa 63
 — SOS 455
 — V-ATE 456
 — VIP 457
 temperatura szumów 582
 termokompresja 451
 transkonduktancja 314
 tranzystor 17
 — bezdryftowy 245
 — bipolarny 240
 — dryftowy 245
 — epiplanarny 247
 — FAMOS 469
 — jednozłączowy 546
 — lateralny 458, 498
 — lawinowy 287

tranzystor MAOS 469
 — MIS z kanałem wbudowanym 384
 — — — — — wzbogacanym 386, 465
 — — — — — zaindukowanym 384
 — — — — — zubożanym 386, 467
 — MNOS 469
 — polowy (FET) 353
 — — cienkowarstwowy (TFT) 357, 404
 — — MESFET 356, 403
 — — MIS (MOS) 357, 382
 — — z izolowaną bramką (IGFET) 356
 — — ze złączem $p-n$ (PNFET) 356
 — — złączowy (JFET) 356
 — stopowy 247
 — unipolarny 240, 353
 — wertykalny 458
 — z jednorodną bazą 245
 triak 552
 tyrystor 551

Układ Darlingtona 459
 — Graetza 198
 — scalony 444
 — — analogowy 477
 — — bipolarny 445, 447
 — — CCD 507
 — — cienkowarstwowy 445, 464
 — — CML 526
 — — CTD 507
 — — cyfrowy 505
 — — CMOS 507
 — — DTL 506
 — — ECL 507, 526
 — — I²L 507, 518
 — — NMOS 507
 — — PMOS 507
 — — RTL 506
 — — TTL 506, 513
 — — grubowarstwowy 445, 472
 — — LSI 445
 — — półprzewodnikowy 445, 447
 — — SSI 445
 — — TTLS 525
 — — unipolarny 445, 464
 — — warstwowy 445, 470
 — WB 416, 420
 — WC 416, 422
 — WD 426, 428
 — WE 416, 420
 — WG 426, 430
 — WS 426, 427
 — Widlara 479
 układy logiczne dynamiczne 537
 — — — — — dwutaktowe 537
 — — — — — ilorazowe 537
 — — — — — nieilorazowe 537
 — — kombinacyjne 535
 — — nieregeneracyjne 535

układy logiczne regeneracyjne 535
 — — — — — sekwencyjne 535

Wakans 69
 waraktor 217
 warikap 217
 warstwa epitaksjalna 63
 — inwersyjna 147
 — ładunku przestrzennego 89
 — zagrzebana 447
 — zaporowa 89
 — zubożona 90
 wejście nieodwracające 490
 — odwracające 490
 wejściowe napięcie nie zrównoważenia 492, 495
 wejściowy prąd nie zrównoważenia 492, 495
 — — polaryzacji 492, 495
 wewnętrzny współczynnik blokowania 548
 wiązania kowalencyjne 23
 — międzyatomowe 23
 wskaźniki Millera 22
 współczynnik dyfuzji 44
 — stabilizacji napięcia 503
 — — obciążeniowej 503
 — szumów 235, 582
 — tłumienia sygnału wspólnego 485, 492, 495
 — — tętnień 503
 — wzmocnienia prądowego 260
 wstrzykiwanie 45
 wyciąganie 45
 wyprostowane pasma 150
 wzmacniacz operacyjny 490
 — różnicowy 483
 wzory Boltzmanna 36

Zakres nasycenia 367, 393
 — nienasycenia 367, 392
 zasada Pauliego 25
 zaćkanie 361
 zjawisko Early'ego 274
 — Millera 428
 — Zenera 116
 złącze liniowe 93
 — $m-s$ 162
 — na gorących elektronach 173
 — omowe 173
 — $p-n$ 88
 — skokowe 93
 — z długą bazą 115
 — — krótką bazą 115
 zubożenie 150
 — głębokie 151

Źródło 359, 383
 — prądowe 479
 — prądu zwierciadlanego 480
 — Wilsona 481

