

ANDRZEJ OPOLSKI

ELEKTRONIKA DLA ELEKTRYKÓW

**WYDAWNICTWO
POLITECHNIKI GDAŃSKIEJ**

ELEKTRONIKA DLA ELEKTRYKÓW

ANDRZEJ OPOLSKI

GDAŃSK 2008

PRZEWODNICZĄCY KOMITETU REDAKCYJNEGO
WYDAWNICTWA POLITECHNIKI GDAŃSKIEJ
Romuald Szymkiewicz

REDAKTOR
Wiesław Pudlik

RECENZENCI
Przemysław Pazdro
Michał Polowczyk

Wydanie I – 1997
Wydanie II – 2002

Wydano za zgodą
Rektora Politechniki Gdańskiej

© Copyright by Wydawnictwo Politechniki Gdańskiej
Gdańsk 2008

Wydawnictwa PG można nabywać w Księgarni PG (Gmach Główny, I piętro)
bądź zamówić pocztą elektroniczną (ksiegarnia@pg.gda.pl),
faksem (058 347 16 18) lub listownie (Wydawnictwo Politechniki Gdańskiej,
Księgarnia PG, ul. G. Narutowicza 11/12, 80-952 Gdańsk)

ISBN 978–83–7348–248–7

SPIS TREŚCI

Przedmowa	5
1. MATERIAŁY I TECHNOLOGIE PÓLPRZEWODNIKÓW	7
1.1. Krzem i inne materiały półprzewodnikowe	7
1.2. Zarys technologii krzemu	8
1.3. Ładunki elektryczne w półprzewodniku	10
1.4. Domieszkowania	16
1.5. Prąd w półprzewodniku	21
2. STRUKTURY NIEJEDNORODNE	25
2.1. Złącze pn	25
2.1.1. Złącze pn w stanie równowagi	25
2.1.2. Złącze spolaryzowane	28
2.1.3. Pojemność złącza pn	30
2.1.4. Przebieg złącza pn	32
2.1.5. Wpływ temperatury na złącze pn	33
2.2. Złącze ms	33
2.3. Struktura MIS (MOS)	37
2.4. Realizacja struktur niejednorodnych	39
3. ELEMENTY ELEKTRONICZNE	42
3.1. Rezystory	43
3.1.1. Rezystory dyskretne liniowe	43
3.1.2. Rezystory w układach scalonych	44
3.1.3. Rezystory nieliniowe półprzewodnikowe	45
3.2. Kondensatory	47
3.3. Diody	48
3.3.1. Dioda prostownicza	49
3.3.2. Dioda przełączająca (uniwersalna)	51
3.3.3. Dioda Zenera	52
3.3.4. Dioda pojemnościowa	53
3.4. Tranzystory	54
3.4.1. Tranzystory unipolarne	55
3.4.2. Tranzystor bipolarny	61
3.4.3. Podobieństwa i różnice między tranzystorami	64
3.5. Tyrystory	68
3.5.1. Tyrystor triodowy (tyrystor)	69
3.5.2. Odmiany tyrystorów	71
3.6. Elementy optoelektroniczne	72
3.6.1. Fotodioda	73
3.6.2. Fototranzystor	74
3.6.3. Diody świecące	75
3.6.4. Transoptory	76
4. MIKROELEKTRONIKA, ENERGEOELEKTRONIKA	78
4.1. Sygnał elektryczny, energia elektryczna	78
4.2. Sygnał analogowy a sygnał cyfrowy	79
4.3. Zakłócenia, szumy, dryft	81
4.4. Przekształcanie sygnałów elektrycznych	83
4.4.1. Widmo sygnału, filtracja	84
4.4.2. Próbkowanie, kwantowanie	86
4.4.3. Modułacja, demodulacja	87
4.4.4. Kodowanie i transmisja sygnałów cyfrowych	89
4.5. Przekształcanie energii elektrycznej	91

5. WZMACNIACZE	94
5.1. Schemat blokowy	94
5.2. Opis wzmacniacza	96
5.2.1. Parametry czwórnikowe wzmacniacza	96
5.2.2. Sprzężenie zwrotne	98
5.2.3. Charakterystyki wzmacniaczy	102
5.3. Wzmacniacz różnicowy	102
5.3.1. Budowa i działanie wzmacniacza różnicowego	102
5.3.2. Parametry wzmacniacza różnicowego	110
5.3.3. Model nieliniowy wzmacniacza różnicowego	111
5.4. Wzmacniacz operacyjny (WO)	115
5.4.1. Parametry wzmacniacza operacyjnego	115
5.4.2. Zastosowania liniowe WO	117
5.5. Wzmacniacze mocy	122
6. GENERATORY ELEKTRONICZNE	125
6.1. Sprzężenie zwrotne dodatnie	125
6.2. Generacja sygnałów sinusoidalnych	128
6.2.1. Generator sygnału sinusoidalnego LC	129
6.2.2. Generator sygnału sinusoidalnego RC	130
6.3. Generacja sygnałów impulsowych	134
6.3.1. Przerzutnik Schmitta	134
6.3.2. Multiwibrator astabilny	135
6.3.3. Multiwibrator monostabilny	138
6.3.4. Generator sygnału trójkątnego	140
6.4. Stabilizacja częstotliwości – "generator kwarcowy"	142
6.5. Generator mocy	143
7. ELEKTRONIKA CYFROWA	147
7.1. Podstawowe pojęcia elektroniki cyfrowej	147
7.2. Funktory TTL, CMOS i ECL	150
7.3. Cyfrowe bloki funkcjonalne	154
7.3.1. Kodery	155
7.3.2. Multipleksery, demultipleksery	155
7.3.3. Sumator	156
7.3.4. Komparator	157
7.3.5. Jednostka arytmetyczno-logiczna	158
7.4. Przerzutniki, rejestry	159
7.5. Liczniki	161
7.6. Mikroprocesor	163
7.7. Przetworniki c/a i a/c	166
8. ZASILACZE	170
8.1. Struktury zasilaczy	170
8.2. Prostowniki	171
8.2.1. Prostownik jednopulsowy	172
8.2.2. Prostowniki dwupulsowe	175
8.2.3. Wygładzanie napięcia wyprostowanego	176
8.3. Stabilizatory napięcia stałego	179
8.3.1. Stabilizator równoległy, parametryczny	181
8.3.2. Ciągły stabilizator szeregowy ze sprzężeniem zwrotnym	182
8.4. Przetwornice napięcia stałego	184
8.4.1. Przetwornica dwutaktowa, przeciwbieżna	185
8.4.2. Przetwornica dwutaktowa współbieżna	189
8.4.3. Stabilizacja, zabezpieczenia, zakłócenia	191
LITERATURA	193

PRZEDMOWA

W programie studiów, zarówno magisterskich jak i inżynierskich, tak stacjonarnych jak i zaocznych, na Wydziale Elektrotechniki i Automatyki Politechniki Gdańskiej jest umieszczony przedmiot o nazwie Elektronika. Przedmioty o podobnych nazwach występują też w programach nauczania elektryków na wszystkich wyższych uczelniach, tak w Polsce jak i poza jej granicami. Oznacza to, że wszędzie przeszłych inżynierów elektryków uczy się też elektroniki.

Cel studiowania elektroniki przez elektryków jest dwojaki. Pierwszym zadaniem nauczania elektroniki jest przygotowanie elektryków do studiowania energoelektroniki, która dla części z nich będzie przedmiotem profesjonalnym, czyli jednym z elementów zawodu inżyniera-elektryka. Drugim zadaniem nauczania elektroniki jest uzupełnienie ogólnego wykształcenia technicznego, niezbędnego dziś każdemu inżynierowi, o wybrane pojęcia wywodzące się z elektroniki a także o informacje na temat jej możliwości i ograniczeń. To drugie zadanie nabiera szczególnej wagi ze względu zarówno na szeroki zakres zastosowań elektroniki, jak i na powszechność używania pojęć z tej dziedziny.

Dwoistość celów nauczania pociąga za sobą niejednorodność programu przedmiotu i konsekwentnie także skryptu z elektroniki. W treści skryptu znalazły się więc informacje, których znajomość jest niezbędna inżynierowi elektrykowi (rozdział 3 oraz rozdziały 5 do 8), obok wiadomości poszerzających wiedzę z fizyki, technologii materiałów czy elektrotechniki teoretycznej (rozdziały 1, 2 i 4), których zadaniem jest uzupełnienie wiedzy ogólnotechnicznej.

Niniejszy skrypt jest adresowany głównie do studentów trzeciego semestru studium magisterskiego, ale może być użyteczny i dla innych studentów elektrotechniki. Zamiarem autora było stworzenie skryptu w klasycznym sensie, czyli pomocy dydaktycznej, która pozwoli studentom skupić się na aktywnym słuchaniu wykładu, bez konieczności ustawicznego notowania przekazywanych informacji.

Zakres skryptu, zgodnie z dobrymi tradycjami w tej dziedzinie, wychodzi nieco poza przyjęte minimum programowe, ale bardziej ambitni studenci studiów magisterskich zdecydowanie powinni poszerzać swoją wiedzę z elektroniki korzystając z literatury, której podstawowy wykaz zamieszczono na końcu skryptu. Dla studentów poziomu inżynierskiego porcja teorii zawarta w tym skrypcie jest w większości przypadków całkowicie wystarczająca, niekiedy nawet z nadmiarem, ale powinni oni aktualizować swoją wiedzę praktyczną przez śledzenie prasy technicznej. Niektóre partie skryptu, dotyczące trudniejszych tematów, zostały napisane w postaci bardziej szczegółowych komentarzy, mając na względzie studentów zaocznych, samodzielnie wchodzących w nową dla nich tematykę techniczną.

Racjonalne ułożenie programu przedmiotu i zakresu skryptu z elektroniki dla elektryków jest zadaniem bardzo trudnym, być może nierozwiązalnym. Szybki rozwój elektroniki w ostatnich 30–40 latach spowodował znaczne poszerzenie i pogłębienie wiedzy w tej dziedzinie, potwierdza to fakt prowadzenia kilku, różnych kierunków studiów w dziedzinie

elektroniki. Ograniczona ilość czasu przeznaczanego na elektronikę w programie nauczania elektryków na PG wymusza ostrą selekcję tematyki i zakresu prezentacji elektroniki, a kryteria selekcji zawsze są dyskusyjne.

Niejednolite przygotowanie studentów, wywodzących się z różnych typów szkół, pociąga za sobą celowość zachowania tematów podstawowych, wprowadzających w elektronikę, wiążących elektronikę techniczną z fizyką i technologią półprzewodników. Taka tematyka, dobrze znana niektórym studentom dla innych stanowi nowość. Przygotowanie zawodowe elektryka wymaga z kolei wprowadzenia tematyki ściśle technicznej, prezentacji układów scalonych, elementów logicznych i cyfrowych. Współcześnie następujący szybki postęp techniki, a szczególnie elektroniki, nakazuje uzupełniać tematykę wykładu i skryptu o nowości, co jednak zagraża wzrostem objętości, czyli powszechnie krytykowanym "przeładowaniem" programu.

Niniejszy skrypt jest próbą pogodzenia wyżej zarysowanych sprzeczności i ograniczeń. O tym, w jakim stopniu udało się te sprzeczności pogodzić zadecyduje opinia czytelników – studentów.

Autor wyraża gorące podziękowanie recenzentom skryptu, profesorowi Przemysławowi Pazdro z Wydziału Elektrycznego i profesorowi Michałowi Polowczykowi z Wydziału Elektroniki, za trud włożony w ocenę pracy oraz cenne uwagi, które przyczyniły się do istotnego poprawienia pierwotnej wersji tekstu skryptu.

1. MATERIAŁY I TECHNOLOGIE PÓLPRZEWODNIKÓW

Współczesna elektronika jest oparta na elementach półprzewodnikowych, nieznaczne ilościowo grupy stanowią jeszcze lampy elektronowe (kineskopy w telewizorach, lampy dużej mocy w nadajnikach radiowych i telewizyjnych), wskaźniki ciekłokrystaliczne (w kalkulatorach, zegarkach itp.) i inne. Cechą charakterystyczną elementów półprzewodnikowych (zwanymi popularnie półprzewodnikami) jest bardzo bliskie powiązanie konstrukcji, zapewniającej żądane właściwości elektryczne elementu, z technologią jego wytwarzania. Zrozumienie zasad działania, oszacowanie możliwości i ograniczeń w zastosowaniach urządzeń elektronicznych jest trudne bez znajomości, choćby pobieżnej, materiałów i technologii półprzewodników oraz zachodzących w nich zjawisk fizycznych.

1.1. Krzem i inne materiały półprzewodnikowe

Obecnie podstawowym materiałem półprzewodnikowym jest krzem, pierwiastek stanowiący ok. 25% masy kuli ziemskiej. Krzem, używany w elektronice w postaci krystalicznej, ma kolor ciemnoszary, jest twardy i wyglądem przypomina metal. Właściwości elektryczne krzemu są jednak całkowicie odmienne od właściwości metali, co wykrywano stopniowo, rozpoczynając badania różnych materiałów, w tym i półprzewodników, w drugiej połowie XIX wieku. Ważniejsze dane fizyczne krzemu zestawiono w tablicy 1.1. Dane zawarte w tablicy pozwalają ocenić właściwości mechaniczne krzemu (dla porównania w tablicy przytoczono też wybrane parametry stali konstrukcyjnej) – właściwości elektryczne krzemu będą przedmiotem dalszych objaśnień.

Podstawową cechą elektryczną materiału jest jego rezystywność – oporność właściwa. W temperaturze pokojowej ($300\text{K} \approx +27^\circ\text{C}$) rezystancja kostki o boku 1 cm, wykonanej z czystego krzemu, mierzona między przeciwległymi ścianami, wynosi 0,4 megaoma (1 megaom = 10^6 omów). Rezystancja podobnej kostki miedzianej wynosi 1,7 mikrooma (1 mikroom = 10^{-6} oma), kostki ze stopu oporowego (konstantan) 48 mikroomów, a kostki porcelanowej ok. miliona megaomów. Rezystywność krzemu, jako półprzewodnika, jest zdecydowanie różna od rezystywności dobrych przewodników prądu elektrycznego (metali) jak też od rezystywności izolatorów. Ponadto rezystywność krzemu, jak i innych półprzewodników, jest silnie zależna od oddziaływania czynników fizycznych, takich jak nagrzewanie, oświetlenie, naprężenia mechaniczne i in.

Na przykład podgrzanie krzemu o 10 K powoduje blisko dwukrotne zmniejszenie jego rezystywności (podgrzanie miedzi o 10 K powoduje wzrost rezystywności około 4%). Podobnie silnie działają na rezystywność krzemu, nawet bardzo niewielkie, zanieczyszczenia chemiczne innymi substancjami. Dlatego krzem stosowany do wytwarzania elementów elektronicznych ma zwykle znacznie mniejszą, aniżeli przytoczona w tablicy, wartość rezystywności.

Wrażliwość rezystywności półprzewodników na oddziaływanie czynników fizycznych niekiedy stanowi ich zaletę, ale częściej musi być traktowana jako wada. Zaletą jest przy

wytwarzaniu półprzewodnikowych przetworników pomiarowych (takich jak termometry, fotometry, tensometry elektryczne) do pomiarów wybranych wielkości fizycznych, bo silna zmienność rezystywności w funkcji tych wielkości pozwala uzyskiwać przetworniki o dużej czułości. W innych przypadkach zależność prądu w obwodzie elektrycznym zawierającym elementy półprzewodnikowe, od np. temperatury otoczenia, jest zjawiskiem niepożądanym, zakłócającym planowany sposób działania tego obwodu. Przeciwdziałanie skutkom wrażliwości półprzewodnika, szczególnie na temperaturę, jest jednym z podstawowych problemów elektroniki, rozwiązywanym zarówno przy wytwarzaniu elementów, jak i przy konstruowaniu układów i urządzeń.

Tablica 1.1

Niektóre właściwości krzemu oraz stali konstrukcyjnej

Właściwość	Krzem	Stal konstrukcyjna
liczba atomowa	14	
masa atomowa	28,1	
temperatura topnienia [°C]	1420	1480 – 1530
temperatura wrzenia [°C]	2600	2800
gęstość [g/cm ³]	2,33	7,8
wsp. rozszerzalności liniowej [1/K]	$4,68 \times 10^{-6}$	$11,5 \times 10^{-6}$
moduł sprężystości [N/mm ²]	106 830	211 000
twardość w/g Brinella [N/mm ²]	2350	1200–3500
wytrzymałość na ściskanie [N/mm ²]	93,5	90 – 400
rezystywność przy 300 K [Ω cm]	4×10^5	$(5 - 10) \times 10^{-5}$
wzgl. stała dielektryczna	12	
wytrzymałość na przebicie [V/cm]	2×10^5	
szerokość pasma zabronionego [eV]	1,2	
koncentracja nośników w 300 K [1/cm ³]	$1,45 \times 10^{10}$	
gęstość atomów w kryształach [1/cm ³]	5×10^{22}	

Wrażliwość rezystywności półprzewodników na zanieczyszczenia chemiczne jest poważnym utrudnieniem w przygotowywaniu surowców i produkcji elementów półprzewodnikowych. Do usunięcia zanieczyszczeń stosuje się specjalne zabiegi technologiczne, a gotowe elementy muszą być zamykane w szczelne obudowy metalowe, szklane lub plastikowe. Zarys technologii półprzewodników podaje się na przykładzie krzemu.

1.2. Zarys technologii krzemu

Produkcja krzemu dla celów elektroniki jest dość złożona i wieloetapowa ze względu na konieczność uzyskania materiału o bardzo dużej czystości. Stosuje się różne techniki, tytułem przykładu w dalszym ciągu wymienia się tylko wybrane metody. Pierwszym etapem jest np. topienie dwutlenku krzemu SiO₂ wraz z węglem w łuku elektrycznym bez

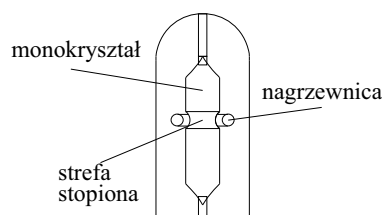
dostępu powietrza. Następuje wówczas wydzielanie tlenku węgla CO i pozostaje krzem Si. Tak otrzymuje się materiał zawierający do 98% krzemu (krzem hutniczy zawierający ok. 2% zanieczyszczeń). Następnie rozpuszcza się uzyskany materiał w silnych kwasach uzyskując sole krzemu. Kolejnym etapem jest destylacja roztworu dla uzyskania chemicznie czystej frakcji zawierającej tylko określoną sól (np. czterochlorek krzemu SiCl_4). Poddając otrzymaną sól redukcji wodorem (oczywiście bez dostępu powietrza) w temperaturze bliskiej 1000°C wytwarza się chemicznie czysty krzem i produkty lotne, które odparowują.

Wyprodukowany krzem, aczkolwiek nazywany chemicznie czystym, zawiera zbyt wiele zbędnych dodatków aby mógł być użyty do produkcji elementów elektronicznych. Właściwy poziom czystości, dopuszczający mniej niż 10^{-9} zanieczyszczeń (mniej niż jedna miliardowa część innych substancji w krzemie) otrzymuje się stosując metodę strefowego czyszczenia monokryształu.

Monokryształ krzemu wytwarza się metodą Czochralskiego. Do powierzchni płynnego krzemu, umieszczonego w odpowiednim tyglu wewnątrz komory próżniowej, zbliża się tzw. zarodek krystalizacji – mały kawałek odpowiednio zorientowanego kryształu krzemu. Po zetknięciu zarodka z powierzchnią cieczy i częściowym nadtopieniu go, następuje powolne podnoszenie zarodka połączone z ruchem obrotowym. Za podnoszonym zarodkiem narasta kryształ krzemu, zachowujący układ krystaliczny zarodka. Proces jest prowadzony w ściśle określonej i stabilizowanej temperaturze, zapewniającej płynność krzemu w tyglu i umożliwiającą zastyganie wyciąganego monokryształu. Szybkość wyciągania wynosi kilka do kilkunastu milimetrów na godzinę co pozwala uzyskiwać monokryształy o długości do jednego metra po wielodniowych procesach krystalizacji. Średnica kryształu zależy od jakości urządzeń technologicznych, obecnie produkuje się kryształy o średnicach od ok. 25 do 150 mm (1 do 6 cali).

Monokryształ zostaje umieszczony w ogrzewanej komorze próżniowej i dodatkowo, lokalnie nagrzany prądem wysokiej częstotliwości aż do utworzenia strefy stopionej, szkic pokazano na rys. 1.1. Wewnątrz strefy stopionej następuje rozdzielanie zanieczyszczeń, cięższe opadają na dno strefy a lżejsze podpływają do góry. Strefę stopioną przemieszcza się wielokrotnie wzdłuż monokryształu w górę i w dół, powodując topienie materiału wchodzącego do strefy i jednocześnie rekrytalizację materiału wychodzącego ze strefy stopionej. Przy przesuwaniu strefy stopionej zanieczyszczenia są "przeprowadzane" do końcowych odcinków pręta krzemowego, a jego część środkowa nabiera pożądanej czystości.

Po oczyszczeniu metodą topienia strefowego i odrzuceniu końców, pręt monokryształu zostaje pocięty na plastry przy pomocy piły diamentowej. Plastry są następnie szlifowane i polerowane do uzyskania połysku zwierciadlanego. Wytworzone płytki, o grubości ok. 0,5–1 mm i średnicy takiej, jak średnica monokryształu, są surowcem do produkcji elementów elektronicznych i układów scalonych.



Rys. 1.1. Schemat czyszczenia strefowego monokryształu

Przedstawiony zarys technologii krzemu wskazuje na znaczne trudności w prowadzeniu tych procesów, szczególnie ze względu na konieczność zachowywania wyjątkowej czystości i prace przy temperaturach sięgających 1400°C.

Otrzymywany tymi sposobami surowiec nie może być tani, co było jednym z powodów wprowadzania miniaturyzacji elementów, tak aby z jednej płytki można było wytworzyć setki, tysiące i dziesiątki tysięcy pojedynczych elementów (a z jednego monokryształu odpowiednio kilkadziesiąt do kilkuset razy więcej). Przy takim postępowaniu, prowadzącym do produkcji wielkoseryjnej, pomimo dużego kosztu surowca, koszt pojedynczego elementu został znacznie obniżony.

Krzem nie jest jedynym półprzewodnikiem stosowanym w elektronice. Wcześniej powszechnie był stosowany german (Ge), o nieco łatwiejszej technologii wytwarzania, a obecnie prowadzi się prace zmierzające do szerokiego zastosowania arsenku galu (GaAs). Innym interesującym materiałem, o potencjalnie dużym zastosowaniu w elektronice, jest węglík krzemu (SiC). Korzystanie z różnych materiałów półprzewodnikowych wynika z takich powodów jak: ograniczone zakresy temperatur pracy elementów wykonywanych z różnych materiałów; możliwość wykorzystywania właściwości specyficznych dla danego materiału (np. widzialne promieniowanie arsenku galu) oraz ze stopnia trudności technologicznych ich obróbki.

Jednym ze skutków, wzmiankowanej powyżej, wrażliwości półprzewodników na temperaturę jest ograniczony przedział dopuszczalnych temperatur pracy elementów. Na przykład elementy germanowe mogą prawidłowo funkcjonować w temperaturach nieprzekraczających 80°C, podczas gdy elementy krzemowe dopuszczają pracę w temperaturach do 180°C. To ograniczenie dotyczy temperatury wewnątrz obudowy elementu półprzewodnikowego, który jest z reguły zainstalowany w szczelnie zamkniętym aparacie, podgrzewanym w środku ciepłem nieuniknionych strat elektrycznych. W takich warunkach może wystąpić znaczny wzrost temperatury wewnątrz elementu ponad temperaturę otoczenia. Dlatego, pomimo pozornie wysokich wartości dopuszczalnych temperatur pracy, utrzymanie ich nie zawsze jest łatwe, a celem prac rozwojowych jest wytworzenie elementów odpornych na jeszcze wyższe temperatury. Arsenek galu i węglík krzemu mają cechy pozwalające przewidywać dopuszczalne temperatury pracy elementów do 300–400°C.

Technologia wytwarzania i obróbki GaAs i SiC, które są związkami chemicznymi, jest znacznie trudniejsza od technologii germanu czy krzemu. Z tego powodu stale prowadzi się prace rozwojowe, m.in. w warunkach nieważkości na pokładzie statków kosmicznych, mając na celu opracowanie technologii obróbki tych, a także innych, materiałów.

1.3. Ładunki elektryczne w półprzewodniku

W fizyce jest znany korpuskularny model atomu, stworzony przez Nielsa Bohra na początku XX stulecia. Ten model, wraz z kolejnymi modyfikacjami, jest wystarczający dla wstępnego objaśnienia zjawisk elektrycznych w półprzewodnikach.

Atom składa się z ciężkiego jądra, zawierającego protony i neutrony oraz znacznie lżejszych elektronów, które obiegając jądro tworzą tzw. powłoki elektronowe. Jądro atomu

krzemu, o liczbie atomowej 14, zawiera 14 protonów, a ponieważ masa atomowa podstawowego izotopu wynosi 28, to liczba neutronów wynosi też 14. Każdy proton ma elementarny ładunek elektryczny dodatni, neutrony są elektrycznie obojętne, zatem całe jądro ma ładunek elektryczny dodatni i wytwarza wokół siebie pole elektryczne. Liczba elektronów w atomie krzemu też jest 14, każdy elektron ma elementarny ładunek ujemny, cały atom jest elektrycznie obojętny.

Przyciąganie różnoimiennych ładunków jądra i elektronów jest równoważone siłą odśrodkową powstającą w wyniku ruchu elektronów wokół jądra. Każdemu elektronowi można przypisać pewną energię całkowitą, zawierającą energię kinetyczną (masa, prędkość) i energię potencjalną (pole elektryczne, ładunek). Energie całkowite elektronów są skwantowane, tj. każdy elektron ma energię o wartości równej jednej ze ściśle określonych wartości dozwolonych energii całkowitej.

Zmiany energii elektronów odbywają się w sposób skokowy, o wartości równe różnicom wartości dozwolonych. Dostarczając do atomu odpowiedni kwant energii, można zwiększyć energię elektronu, czemu odpowiada zwiększenie odległości między tym elektronem a jądrem. Przy dostatecznie dużym kwancie dostarczonej energii można oderwać elektron od atomu, tj. odsunąć go na taką odległość, przy której przyciąganie jądra jest pomijalnie małe. To zjawisko nazywa się jonizacją atomu, ponieważ atom pozbawiony elektronu nabiera wypadkowego ładunku elektrycznego dodatniego, czyli staje się jodem dodatnim.

Energię można dostarczać do atomów pod różnymi postaciami, np. jako światło lub ciepło. Mając na myśli atom w temperaturze pokojowej można przyjąć, że został on doprowadzony do tego stanu od temperatury zera bezwzględnego; oznacza to, że w temperaturze pokojowej co najmniej jeden elektron w atomie jest już odsunięty od jądra dalej, aniżeli znajdowałby się w temperaturze bliskiej 0 K.

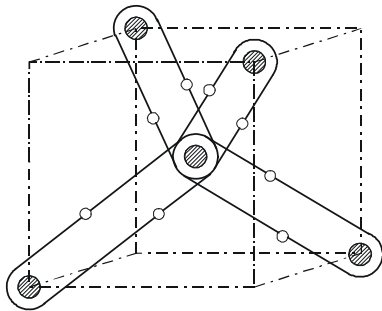
Rozmieszczenie elektronów wokół jądra jest uporządkowane w taki sposób, że każdy elektron ma inny komplet liczb kwantowych, charakteryzujących jego stan energetyczny. Stan energetyczny to opis zarówno energii elektronu, jak i jego momentu magnetycznego i spinu. W temperaturze 0 K (minimum energii atomu) elektrony zajmowałyby położenia możliwie najbliższe jądra atomu. W atomie krzemu oznacza to, że dwa elektrony wypełniają całkowicie powłokę 1, kolejnych osiem elektronów wypełnia całkowicie powłokę 2, a pozostałe cztery elektrony lokują się w powłoce 3, której całkowite wypełnienie nastąpiłoby, gdyby znalazło się tam 18 elektronów. Elektrony z najbardziej zewnętrznej powłoki nazywane są elektronami walencyjnymi.

Elektrony walencyjne są wyróżnione oddzielną nazwą dlatego, że właśnie one umożliwiają powiązanie jednego atomu z innymi atomami. Jeżeli są połączone identyczne atomy, to mamy do czynienia z pierwiastkiem, którego właściwości fizyczne i chemiczne zależą od elektronów walencyjnych. Jeżeli kontaktują się ze sobą różne atomy (różnych pierwiastków), to rodzaj powstających związków chemicznych także zależy od elektronów walencyjnych. Ważną cechą elektronów walencyjnych jest również to, że mogą być one dość łatwo oderwane od atomu i poruszać się samodzielnie. Elektron przenoszący w przestrzeni swój ładunek elektryczny, to jest właśnie najmniejszy element płynącego prądu elektrycznego.

Atomy mogą być różnie rozmieszczone w przestrzeni. Jeżeli odległości między atomami są bardzo duże w porównaniu do ich wymiarów, a rozmieszczenie jest nieuporządkowane i zmienne w czasie, to mamy do czynienia z gazem. Przypomniany powyżej model atomu odpowiada dość wiernie właśnie atomowi gazu. Jeżeli odległości między atomami są małe, a atomy są rozmieszczone w przestrzeni w sposób uporządkowany, tworząc regularną, powtarzającą się strukturę, to mamy do czynienia z ciałem stałym o budowie krystalicznej. Półprzewodniki w elektronice stosuje się najczęściej w postaci krystalicznej.

Fakt powiązania atomów w kryształ modyfikuje model atomu. Przyjmuje się, że jądro i elektrony powłok wewnętrznych zachowują się tak, jak w modelu Bohra, natomiast zmienia się obraz elektronów walencyjnych. Dla uproszczenia rozważań traktuje się jądro atomowe z elektronami wewnętrznymi, czyli tzw. rdzeń atomu, jako pewną całość, o dużej masie i dodatnim ładunku elektrycznym, unieruchomioną w określonym punkcie przestrzeni. Elektrony walencyjne obiegają rdzeń atomu, tworząc wiązania międzyatomowe (wiązania kowalენტne), utrzymujące strukturę kryształu. Wypadkowy ładunek elektryczny jądra i elektronów wewnętrznych (dodatni) jest równoważony ładunkami ujemnymi elektronów walencyjnych, cały atom pozostaje elektrycznie obojętny.

Cztery elektrony walencyjne atomu krzemu poruszają się po orbitach eliptycznych, o różnym ułożeniu w przestrzeni. Na rys. 1.2 pokazano w sposób uproszczony, jeden z atomów kryształu krzemu, umieszczony wewnątrz umownego sześciangu i orbity jego elektronów walencyjnych. Ten atom, podobnie jak każdy inny w kryształ, znajduje się w sąsied-



Rys. 1.2. Model wiązań atomu w kryształ

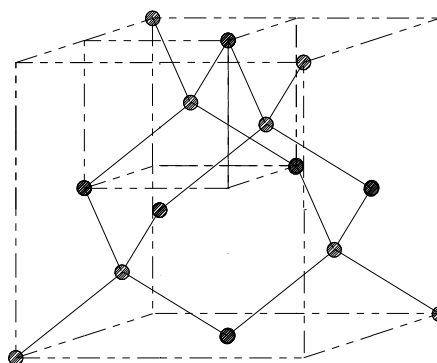
twie innych atomów. Rdzenie najbliższych czterech atomów ulokowane są w narożach umownego sześciangu, w odległościach mniejszych od długości orbit elektronów walencyjnych. Każdy elektron walencyjny obiega więc rdzenie dwu atomów, zapewniając powiązanie między tymi atomami. Orbita jednego z elektronów walencyjnych atomu umieszczonego w narożniku sześciangu, pokrywa się z orbitą jednego z elektronów atomu umieszczonego w środku sześciangu. W ten sposób wiązanie między dwoma atomami składa się z dwu elektronów walencyjnych, krążących po wspólnej orbicie obejmującej rdzenie obydwu tych atomów.

Atom krzemu, umieszczony wewnątrz umownego sześciangu, jest powiązany z czterema najbliższymi atomami, ulokowanymi w narożach tego sześciangu, czterema wiązaniami, z których każde zawiera dwa elektrony walencyjne. Warto zauważyć, że w odniesieniu do elektronów walencyjnych zanika przyporządkowanie do określonego atomu. Dla dwu, sąsiadujących ze sobą, atomów można określić, że dwa elektrony, tworzące wiązanie, należą do tych dwu atomów, natomiast rozróżnienie, który elektron należy do którego atomu jest niemożliwe. Jeżeli atomów jest więcej, to uwzględniając możliwość zamiany orbit między elektronami walencyjnymi atomu, określa się, że wszystkie elektrony walencyjne,

wiążące między sobą atomy, należą do tych atomów. Ta okoliczność, między innymi, prowadzi do podejścia statystycznego do zagadnień fizyki ciała stałego.

Kryształ krzemu zbudowany jest z większej ilości atomów; na rys. 1.3 pokazano jedną komórkę takiego kryształu zaznaczając umowny (nieistniejący) sześcian z rys. 1.2. Długość krawędzi sześcianu obejmującego całą komórkę nazywa się stałą sieci krystalicznej i wynosi ona w krzemie 0,543 nm.

Podobnie jak w pojedynczym, izolowanym atomie każdy elektron ma inny stan energetyczny, to w kryształe każdy elektron walencyjny musi mieć inną wartość energii. Poziom energii dozwolonej, określony dokładnie dla izolowanego atomu, w przypadku kryształu rozszczepia się na tyle bliskich, ale różnych wartości dozwolonych, ile atomów tworzy kryształ. Ponieważ w jednym centymetrze sześciennym kryształu krzemu znajduje się 5×10^{22} atomów, to w kryształe o takiej objętości pojawia się $4 \times 5 \times 10^{22}$ różnych, dozwolonych wartości energii elektronów walencyjnych. Pomimo, że różnice wartości tych energii są bardzo małe, to jednak zajmują one pewien przedział wartości, zwany pasmem energii walencyjnych, lub wprost pasmem walencyjnym.



Rys. 1.3. Elementarny kryształ krzemu

Wyższy poziom energii, który w izolowanym atomie zajmuje elektron po dostarczeniu mu kwantu energii, w kryształe ulega też rozszczepieniu na wiele dozwolonych wartości energii. Przedział energii dozwolonych, większych od energii elektronów walencyjnych uczestniczących w wiązaniach międzyatomowych, nazywa się pasmem energii elektronów przewodnictwa lub pasmem przewodnictwa. Elektron walencyjny, po dostarczeniu mu kwantu energii, zwiększa swoją energię do wartości mieszczącej się w paśmie przewodnictwa.

Największa dozwolona wartość energii pasma walencyjnego (wierzchołek pasma walencyjnego – W_v) w półprzewodnikach jest mniejsza od najmniejszej dozwolonej energii pasma przewodnictwa (dno pasma przewodnictwa – W_c), czyli występuje przedział energii niedozwolonych dla elektronów, zwany pasmem energii zabronionych lub wprost pasmem zabronionym. Szerokość pasma zabronionego:

$$W_G = W_c - W_v \quad (1.1)$$

jest ważną cechą półprzewodnika, określającą wiele jego właściwości, w tym i wrażliwość na temperaturę. Obecnie przyjmuje się szerokość pasma zabronionego jako kryterium podziału na półprzewodniki ($W_G < 2\text{eV}$) i izolatory ($W_G > 2\text{eV}$). W metalach pasmo zabronione zanika, wierzchołek pasma walencyjnego sięga wyżej aniżeli dno pasma przewodnictwa.

Jeżeli kryształ półprzewodnika byłby w temperaturze zera bezwzględnego, to wszystkie elektrony walencyjne miałyby minimalne możliwe energie, odpowiadające pasmu wa-

lencyjnemu i poruszałyby się po orbitach walencyjnych, z których każda obejmuje dwa atomy, tworząc między nimi wiązanie. Jeżeli kryształ półprzewodnika jest w temperaturze wyższej, np. pokojowej, to znaczy, że już pobrał pewną ilość energii cieplnej, ogrzewając się do tej temperatury. Tę energię przejęły niektóre elektrony walencyjne, zwiększając swoje energie od wartości odpowiadających pasmu walencyjnemu do wartości odpowiadających pasmu przewodnictwa. Orbita, po której porusza się elektron o energii pasma przewodnictwa, ma skomplikowany kształt i obejmuje wszystkie atomy tworzące kryształ. Dzięki temu elektron o takiej energii może poruszać się w całym kryształcie bez zmiany swojej energii.

Zwiększenie energii elektronu do wartości leżących w paśmie przewodnictwa oznacza też, że ten elektron został usunięty z wiązania międzyatomowego, przez co to wiązanie zostało uszkodzone, powstała tam "dziura" w wiązaniu. Ilość uszkodzeń zależy od temperatury półprzewodnika; im jest ona wyższa tym więcej energii cieplnej zostało dostarczone do półprzewodnika i tym samym więcej elektronów walencyjnych jest uwolnionych z wiązań. W kryształie krzemu, w temperaturze pokojowej (300 K, ok. +27°C) uszkodzonych jest 1.45×10^{10} wiązań w każdym cm^3 . Wydaje się to dużo, ale porównując ilość uszkodzeń z ilością atomów w tej objętości, można oszacować, że zostaje uszkodzone jedno wiązanie na 10^{13} istniejących.

Podobnie jak jest niemożliwe przyporządkowanie elektronów walencyjnych konkretnym atomom kryształu, to tym bardziej takie przyporządkowanie nie może dotyczyć elektronów o energiach mieszczących się w paśmie przewodnictwa. Przyjmuje się, że wszystkie elektrony wyzwolone z wiązań międzyatomowych i poruszające się w przestrzeni pomiędzy atomami przynależą do tych atomów, czyli do całego kryształu. Ruch tych elektronów odbywa się w całej objętości kryształu, wyprowadzenie elektronu poza kryształ wymagałoby dostarczenia dodatkowej energii. Tory elektronów są trudne do opisanego, już z tego choćby powodu, że pole elektryczne wytworzone w przestrzeni przez ładunki elektryczne jąder wszystkich atomów kryształu ma bardzo złożoną postać (w odróżnieniu od prostych form tworzonych przez jeden czy dwa ładunki punktowe). Przyjmuje się, że elektrony o energiach mieszczących się w paśmie przewodnictwa poruszają się w przestrzeni międzyatomowej półprzewodnika w sposób przypadkowy ("gaz elektronowy"), zmieniając kierunki i prędkości, ruch ten można porównać do obserwowalnych ruchów Browna cząstek materialnych.

Uszkodzenie niektórych wiązań międzyatomowych stwarza też możliwość przemieszczania się w przestrzeni elektronów walencyjnych o energiach zawartych w paśmie walencyjnym. Elektron wchodzący w skład wiązania pełnego, obok którego znajduje się wiązanie uszkodzone, może przemieścić się tak, aby uzupełnić to uszkodzenie. Oczywiście wiązanie, w którym ten elektron był poprzednio, zostanie teraz zdekompletowane. Takie przesunięcie elektronu odbywa się przy minimalnej zmianie jego energii, a jego skutek można zinterpretować jako przesunięcie w przestrzeni "dziury" w wiązaniu.

Jak wynika z powyższego, jeżeli kryształ półprzewodnika jest w temperaturze powyżej zera bezwzględnego, to pojawiają się dwa sposoby poruszania się elektronów walencyjnych w przestrzeni wewnątrz kryształu. Elektrony o energiach wyższych od W_c poruszają się jako tzw. elektrony swobodne lub swobodne nośniki ładunku ujemnego w całej objęto-

ści kryształu, natomiast elektrony o energiach niższych od W_v mogą wykonywać lokalne przesunięcia z jednego wiązania do sąsiedniego. Aby unikać długich opisów i możliwych niejednoznaczności wprowadzono pojęcie dziury, jako uszkodzenia w wiązaniu, które może być uzupełnione elektronem.

Dziurze, która jest pojęciem abstrakcyjnym, można przypisać cechy fizyczne, pozwalające traktować ją jako cząstkę materialną. Na przykład brak elektronu powoduje, że atom staje się dodatnim jonem, to znaczy, że pojawienie się dziury jest równoznaczne z pojawieniem się ładunku elektrycznego dodatniego, co pozwala potraktować dziurę jako cząstkę o ładunku elektrycznym dodatnim, równym co do wartości ładunkowi elektronu. Wymuszając zewnętrznym polem elektrycznym częściowe uporządkowanie i ukierunkowanie przemieszczeń elektronów można określić prędkość poruszania się dziur, w innych doświadczeniach można wyznaczyć odpowiednik masy dziury itd., co prowadzi do nadania dziurze atrybutów cząstki materialnej. Tak więc opisane powyżej dwa sposoby poruszania się elektronów walencyjnych wewnątrz kryształu, przedstawia się jako ruch swobodnych elektronów, czyli swobodnych nośników ładunku ujemnego i ruch swobodnych dziur, czyli swobodnych nośników ładunku dodatniego.

Ze względu na różne sposoby poruszania się elektronów i dziur mają one różne parametry fizyczne. W tabelicy 1.2 zestawiono niektóre dane nośników ładunku elektrycznego w czystym krzemie (określenie masy nośnika jest dość złożone, w tabelicy przytoczono masy efektywne przewodnictwa wg [31]).

Zerwanie każdego wiązania międzycząsteczkowego w kryształce jest równoznaczne z pojawieniem się dwu nośników ładunku elektrycznego: swobodnego elektronu i swobodnej dziury. Ten proces nazywa się generacją pary nośników. Ponieważ obydwa typy nośników poruszają się wewnątrz kryształu, to może nastąpić ich spotkanie w przestrzeni. Jeżeli przy tym elektron swobodny uzupełni uszkodzone wiązanie (oddając odpowiedni kwant energii), to zanikną obydwa nośniki. Takie zjawisko nazywa się rekombinacją pary nośników.

Tablica 1.2

Wybrane parametry nośników ładunku elektrycznego

Cecha nośnika	elektron	dziura
ładunek elektryczny ($q = 1,602 \times 10^{-19} \text{ C}$)	$-q$	$+q$
masa efektywna ($m_0 = 0,91 \times 10^{-27} \text{ g}$ – masa spoczynkowa elektronu)	$1,1 m_0$	$0,59 m_0$
ruchliwość [$\text{cm}^2/(\text{Vs})$]	1350	480
stała dyfuzji w 300K [cm^2/s]	35	12,5

Ilość nośników występująca w jednostce objętości półprzewodnika nazywa się koncentracją nośników. Koncentrację swobodnych elektronów oznacza się literą n , a koncentrację swobodnych dziur literą p . Z opisanego powyżej mechanizmu generacji nośników wynika, że w czystym krzemie ilość swobodnych elektronów jest równa ilości swobodnych dziur, czyli ich koncentracje są równe:

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

a ponadto obowiązuje zależność bardziej ogólna o postaci:

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

Symbolem n_i oznacza się tzw. koncentrację nośników samoistnych. Koncentracja nośników zależy od ilości energii wprowadzonej do półprzewodnika, jest więc funkcją jego temperatury. W fizyce ciała stałego wyprowadza się wzór:

$$n_i^2 = 4 \left(\frac{2\pi kT}{h^2} \right)^3 (m_e m_h)^3 \exp\left(\frac{W_G}{kT} \right) \quad (1.4)$$

gdzie: $k = 8,63 \times 10^{-5}$ eV – stała Boltzmanna,
 $h = 6,63 \times 10^{-34}$ Js – stała Plancka,
 m_e, m_h – masy efektywne: elektronu, dziury,
 W_G – szerokość pasma zabronionego,

pozwalający wyliczyć koncentrację nośników samoistnych w idealnym kryształie półprzewodnika w temperaturze T . Obliczenia przeprowadzone dla krzemu w temperaturze około 300 K wykazują, że koncentracja nośników wzrasta o około 9% przy wzroście temperatury o jeden kelwin. Wzrost ilości nośników ładunku elektrycznego ułatwia przepływ prądu przez półprzewodnik, czyli zmniejsza jego rezystywność.

Koncentrację swobodnych nośników ładunku elektrycznego można traktować jako miernik energii zgromadzonej w kryształie półprzewodnika. Jeżeli, oprócz ciepła, doprowadzi się jeszcze inną postać energii (np. przez oświetlenie półprzewodnika), to koncentracja nośników wzrasta powyżej wartości n_i . W warunkach równowagi energetycznej, tj. przy stałej temperaturze i bez dostarczania energii w jakiegokolwiek postaci, ilość nośników musi być stała.

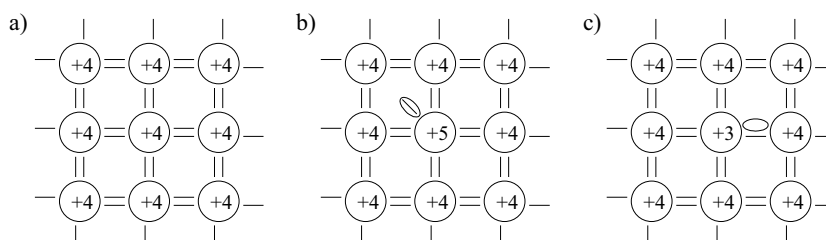
Uwzględniając możliwą rekombinację (zanikanie) nośników i konieczność utrzymania ich stałej liczby przyjmuje się, że w półprzewodniku ciągle przebiegają obydwie zjawiska: generacja par nośników i ich rekombinacja. W warunkach równowagi energetycznej ilość nośników generowanych w jednostce czasu jest równa ilości nośników rekombinujących w tym samym czasie. Średnia wartość czasu od momentu generacji do momentu rekombinacji nazywa się średnim czasem życia nośników. W ten sposób tworzy się dynamiczny obraz półprzewodnika, w którym ciągle powstają nowe nośniki, poruszają się one ruchami przypadkowymi wewnątrz kryształu i, po pewnym czasie, zanikają, ale ich ilość w kryształie (koncentracja), w stałej temperaturze, pozostaje stała.

1.4. Domieszkowania

Przedstawiony w poprzednim punkcie obraz półprzewodnika dotyczy struktury idealnie jednorodnej, tzn. idealnie czystego monokryształu, np. krzemu. W rzeczywistości każdy półprzewodnik zawiera zanieczyszczenie innymi substancjami. Szczególne oddziaływanie

mają pierwiastki III grupy układu okresowego (np. bor – B), zwane akceptorami i pierwiastki grupy V (np. fosfor – P), zwane donorami. Te pierwiastki mają odpowiednio 3 i 5 elektronów walencyjnych. Jeżeli znajdują się one przypadkowo w półprzewodniku, to są nazywane zanieczyszczeniami, ale jeżeli zostaną wprowadzone celowo nazywa się je domieszkami.

Atom domieszki zajmuje miejsce atomu krzemu w sieci krystalicznej półprzewodnika. Atom donora lokuje swoje cztery elektrony walencyjne w wiązaniach z sąsiednimi atomami krzemu, a jego piąty elektron walencyjny pozostaje poza wiązaniami; trzy elektrony walencyjne atomu akceptora wchodzi w wiązania, a czwarte wiązanie pozostaje niekompletne. Na rys. 1.4 pokazano, w sposób uproszczony, atomy domieszek w sieci krystalicznej półprzewodnika. Kółko oznacza rdzeń atomu, liczba przedstawia ładunek elektryczny, a kreski symbolizują elektrony walencyjne, elipsami oznaczono miejsca, gdzie powstaje nadwyżka lub brak elektronu.



Rys.1.4. Modele półprzewodnika: samoistnego (a), domieszkowanego donorowo (b) oraz domieszkowanego akceptorowo (c)

Energia niezwiązanego elektronu donora ma inną wartość aniżeli energie elektronów krzemu, Ta wartość mieści się w paśmie energii zabronionych dla elektronów krzemu, w odległości ok. 0,03 eV od dna pasma przewodnictwa. Przy dostarczeniu niewielkiej porcji energii "piąty" elektron zostaje oderwany od atomu donora i jego energia przyjmuje wartość mieszczącą się w paśmie przewodnictwa. Ten elektron staje się więc swobodnym nośnikiem ładunku ujemnego i może poruszać się w całym kryształ, podczas gdy atom donora przekształca się w jon dodatni, unieruchomiony w sieci krystalicznej.

Przyłączenie innego elektronu do jonu donora jest mało prawdopodobne, ponieważ taki elektron musiałby mieć energię różną od energii dozwolonych w pasmach walencyjnym czy przewodnictwa półprzewodnika. Jonizacja cieplna donorów ma miejsce w temperaturze kilkudziesięciu kelwinów, zatem w temperaturze pokojowej każdy atom donora jest już jodem, a ilość nośników elektronowych wzrasta o tyle, ile atomów donorów znajduje się w kryształ.

Atom domieszki akceptorowej ma jedno wiązanie krystaliczne niekompletne, to wiązanie jest łatwo uzupełniane. Energia elektronu, uzupełniającego niepełne wiązanie atomu akceptora, też mieści się w paśmie zabronionym, ok. 0.05 eV powyżej wierzchołka pasma walencyjnego. Przy niewielkiej dawce energii, jeden z elektronów walencyjnych krzemu może uzupełnić zerwane wiązanie atomu akceptora, powodując powstanie dziury w sąsied-

nim wiązaniu. Tak więc, atom akceptora przekształca się w nieruchomy jon ujemny, a powstająca dziura staje się swobodnym nośnikiem ładunku dodatniego. W temperaturze pokojowej wszystkie atomy akceptorów są już zjonizowane.

Jeżeli ilość donorów jest równa ilości akceptorów, to ilość swobodnych elektronów pochodzących od atomów donorowych jest równa ilości swobodnych dziur pochodzących od atomów akceptorowych i właściwości elektryczne takiego półprzewodnika są takie same jak półprzewodnika czystego, samoistnego. Jeżeli bilans nie jest zrównoważony, to w półprzewodniku występują różne ilości (koncentracje) elektronów i dziur. W przypadku $n > p$ półprzewodnik nazywamy typu n (o przewodnictwie elektronowym), swobodne elektrony nazywamy nośnikami większościowymi, a swobodne dziury nośnikami mniejszościowymi. Jeżeli jest odwrotnie, $p > n$, to mówi się o półprzewodniku typu p lub półprzewodniku o przewodnictwie dziurowym, w którym nośnikami większościowymi są dziury a mniejszościowymi elektrony.

W rzeczywistym półprzewodniku część nośników pochodzi z jonizacji domieszek, a część z generacji samoistnej (wyżej opisana generacja par nośników). Cały kryształ półprzewodnika pozostaje elektrycznie obojętny, każdemu elektronowi odpowiada dziura lub jon dodatni, każdej dziurze odpowiada elektron lub jon ujemny. Nadal łączna ilość nośników obydwu typów jest miernikiem energii zawartej w kryształach i wzór (1.3) – $n \cdot p = n_i^2$ – nadal obowiązuje.

Gdyby ilość atomów zanieczyszczeń i domieszek była duża, to ilość pochodzących od nich elektronów i dziur mogłaby być tak duża, że iloczyn $n \cdot p$ mógłby być większy od n_i^2 . Taki półprzewodnik nazywa się zdegenerowanym.

Domieszkowanie półprzewodników jest ważnym zabiegiem technologicznym, umożliwiającym produkcję materiałów półprzewodnikowych typu p i typu n. Waga tego procesu wynika z dwu czynników:

- 1) domieszkowanie zmniejsza wrażliwość półprzewodnika na zmiany temperatury i, co ważniejsze
- 2) domieszkowanie umożliwia tworzenie struktur niejednorodnych w półprzewodniku, w szczególności złączy pn.

Pierwsze oddziaływanie rozpatrzmy tu na przykładzie liczbowym, drugie będzie opisane w następnym rozdziale.

Dla krzemu w temperaturze 300 K wartość n_i wynosi ok. $1,5 \times 10^{10}$ [1/cm³], a w temperaturze 310 K ok. $2,8 \times 10^{10}$ [1/cm³]. W półprzewodniku samoistnym są to jednocześnie koncentracje elektronów i dziur, decydujące o jego rezystywności. Blisko dwukrotny wzrost ilości nośników, wywołany podgrzaniem krzemu o 10 kelwinów, powoduje blisko dwukrotne zmniejszenie rezystywności.

Jeżeli krzem jest domieszkowany np. donorowo i ilość domieszki wynosi np. 10^{13} [1/cm³], tj. 10^{13} atomów donora w każdym centymetrze sześciennym kryształu, to koncentracja nośników elektronowych pochodzących z jonizacji donorów wynosi też 10^{13} [1/cm³]. Koncentracja dziur w temperaturze 300 K musi być taka, aby iloczyn np. wynosił $\{(1,5 \times 10^{10})[1/\text{cm}^3]\}^2 = 2,25 \times 10^{20}$ [1/cm⁶], zatem p wynosi ok. $2,25 \times 10^7$ [1/cm³]. Ponieważ koncentracja mniejszościowych dziur jest blisko pół miliona razy mniejsza od koncentracji większościowych elektronów, to wpływ dziur na przewodzenie prądu jest zni-

komy. Oznacza to, że w tym przypadku o rezystywności materiału decyduje koncentracja elektronów.

W temperaturze zwiększonej o 10 kelwinów iloczyn np. musi wynosić ok. $7,8 \times 10^{20}$ [1/cm⁶]. Ilość atomów domieszki nie zmienia się od podgrzania, czyli ilość elektronów z jonizacji donorów pozostanie niezmienna 10^{13} [1/cm³], zatem koncentracja dziur wzrośnie do wartości ok. $7,8 \times 10^7$ [1/cm³]. Pomimo trzykrotnego wzrostu nadal koncentracja dziur pozostaje znikoma wobec koncentracji elektronów i jej wpływ na rezystywność jest pomijalny. Decydująca o rezystywności koncentracja elektronów nie zmieniła się.

Przedstawiony rachunek jest niedokładny, dziury w półprzewodniku typu n pochodzą z procesu generacji par nośników, wzrostowi ilości dziur towarzyszy wzrost ilości elektronów, ale w rachunku sprowadzałyby się to do sumowania ilości elektronów od domieszki (10^{13}) i elektronów generowanych cieplnie (10^7), czyli dodawania znacznie różniących się składników (ponadto całe zagadnienie ma charakter raczej statystyczny, podane obliczenia są bardzo uproszczonym modelem). Pomimo niedokładności obliczeń można sformułować następujące wnioski:

- w półprzewodniku domieszkowanym wpływ temperatury na rezystywność jest znacznie mniejszy niż w półprzewodniku samoistnym,
- przy znacznym podwyższeniu temperatury (w przykładzie: ilość nośników generowanych termicznie byłaby współmierna z ilością nośników od domieszki) powraca wrazliwość półprzewodnika na temperaturę,
- koncentracja domieszki powinna być większa od n_i , ale znacznie mniejsza od wartości n_i^2 .

Ostatni wniosek potwierdza konieczność stosowania bardzo czystych materiałów. Przyjęta w przykładzie ilość domieszki 10^{13} [1/cm³] (przy wartości n_i rzędu 10^{10} [1/cm³]), odpowiadająca stosowanym w praktyce, oznacza dodatek 1 atomu domieszki na 5×10^9 atomów krzemu, jest to taka relacja jak dodanie 1 grama domieszki do 5 tysięcy ton materiału.

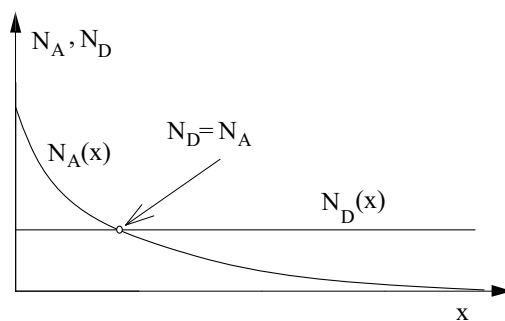
Monokryształ krzemu, pomimo starannego oczyszczania, zawiera zwykle zanieczyszczenia nadające mu przewodnictwo typu p lub n. Zależy to, między innymi, od procesów technologicznych zarówno we wstępnych, jak i finalnych etapach produkcji. Domieszkowanie może polegać zarówno na zwiększaniu koncentracji nośników większościowych – np. wprowadzanie donorów do krzemu n, jak też na zwiększaniu koncentracji nośników mniejszościowych – np. wprowadzanie akceptorów do krzemu typu n. W tym drugim przypadku można doprowadzić do wystąpienia jednakowych koncentracji elektronów i dziur, czyli uzyskać materiał o właściwościach półprzewodnika samoistnego albo, wprowadzając jeszcze większą ilość domieszki, uzyskać materiał o odwrotnym typie przewodnictwa w stosunku do materiału wyjściowego.

Stosowane są trzy metody wprowadzania domieszek do półprzewodnika. Są to: dyfuzja, epitaksja i implantacja jonów.

Dyfuzja, znane w fizyce zjawisko, polega na samorzutnym przemieszczaniu się cząstek z obszaru gdzie jest ich wiele do rejonów, gdzie jest ich mniej czyli w kierunku spadku koncentracji. Dyfuzja przebiega stosunkowo szybko w gazach (np. rozprzestrzenianie się zapachów w spokojnym powietrzu) i nieco wolniej w cieczach. Dyfuzja w ciałach stałych

też występuje, aczkolwiek jej przebieg jest trudny do zaobserwowania w warunkach naturalnych ze względu na bardzo powolny przebieg (tysiące lat). Szybkość procesu dyfuzji w ciele stałym można zwiększyć podnosząc jego temperaturę.

Dyfuzję domieszek do półprzewodnika przeprowadza się umieszczając płytki, wycięte z monokryształu, w ogrzewanej komorze próżniowej. W tejże komorze umieszcza się wyparownik, zawierający materiał domieszki. Doprowadza się do parowania materiału domieszki i osiadania pary na powierzchni płytek krzemowych. W odpowiednio wysokiej temperaturze atomy domieszki wnikają w kryształ krzemu, wypierając niektóre jego atomy z sieci krystalicznej i zajmując ich miejsca. Głębokość wnikania jest różna, najwięcej atomów domieszki lokuje się tuż przy powierzchni półprzewodnika a nieliczne wnikają w głąb kryształu. Po dyfuzji prowadzonej w temperaturach do 1250°C (stabilizowanej z dokładnością do 0,25 K), przez czas do kilku godzin, domieszki wnikają na głębokość kilku do kilkudziesięciu mikrometrów. Uzyskiwane efekty dyfuzji oczywiście zależą od parametrów prowadzenia tego procesu, szczegółowe recepty technologiczne są niechętnie ujawniane przez producentów elementów i układów elektronicznych.



Rys. 1.5. Profil koncentracji wdyfundowanej domieszki $N_A(x)$ w monokryształach ze stałą koncentracją donorów N_D

Na rys. 1.5 pokazano profil koncentracji domieszki akceptorowej, wdyfundowanej do półprzewodnika typu n. Proces poprowadzono tak, aby warstwa przypowierzchniowa uzyskała przewodnictwo typu p, a obszar położony w głębi zachował przewodnictwo typu n. W ten sposób wytworzono złącze pn, zlokalizowane w niewielkiej odległości od powierzchni kryształu.

Epitaksja (z greckiego, dosłownie: układanie powyżej) oznacza wzrost kryształu z fazy gazowej (niekiedy ciekłej), prowadzony na podłożu krystalicznym. Płytkę krzemową umieszcza się w strumieniu gazu zawierającym SiCl_4 , wodór i inne związki chemiczne. Dzięki reakcjom przebiegającym w temperaturze 800 do 1100°C zostaje uwolniona część atomów Si dostarczonych w SiCl_4 , a także atomy domieszek, wprowadzone w innych związkach chemicznych. Te atomy osiadają na płytce w takich miejscach, aby następowała kontynuacja istniejącej sieci krystalicznej. Na płytce narasta warstwa epitaksjalna krzemu z odpowiednią domieszką, zachowująca strukturę krystaliczną podłoża. Grubości wytwa-

rzanych warstw epitaksjalnych wynoszą od 1 do 25 mikrometrów. Dzięki regulacji składu chemicznego gazu kierowanego na płytkę krzemową i innych parametrów technologicznych, istnieje możliwość dokładnego formowania profilu wprowadzonej domieszki.

Implantacja (z łaciny, dosłownie: wsadzanie) jonów polega na "wstrzeliwaniu" jonów dodatnich odpowiedniej domieszki do płytki krzemowej. Atomy domieszki jonizuje się (przez pozbawienie elektronu), aby było możliwe rozpędzanie ich silnym polem elektrycznym, formowanie wiązki i skierowanie jej na powierzchnię płytki. Implantację prowadzi się w komorze próżniowej, w niezbyt wysokich temperaturach. Jony są implantowane na głębokość do 1 mikrometra w krzemie. Po implantacji przeprowadza się aktywizację jonów przez wygrzewanie płytki w temperaturze 550 do 1000°C. W trakcie aktywizacji następuje rekrytalizacja ewentualnych uszkodzeń powierzchni płytki i wprowadzenie jonów domieszki we właściwe miejsca, tj. w sieć krystaliczną półprzewodnika.

Przedstawione powyżej zarysy technologii domieszkowania półprzewodników wskazują na możliwości nadawania półprzewodnikom określonego typu przewodnictwa elektrycznego, wytwarzania elementów o określonej rezystywności, tworzenie struktur niejednorodnych. Warto zauważyć, że najczęściej obróbka technologiczna, (jak i zjawiska elektryczne przebiegające w gotowych elementach) dotyczy bardzo cienkich warstw przypowierzchniowych na płytkach półprzewodnika. Z tego powodu, w procesach produkcyjnych, dąży się zwykle do stosowania możliwie cienkich płytek, aby uzyskać ich możliwie dużo z monokryształu. Właściwości mechaniczne kryształów, jak też i możliwości ich cięcia czy szlifowania, powodują, że używane są znacznie grubsze płytki (o grubości około 1 mm), aniżeli byłoby to niezbędne dla przebiegu procesów elektrycznych.

1.5. Prąd w półprzewodniku

Dla uzyskania przepływu prądu przez półprzewodnik trzeba doprowadzić do ukierunkowania ruchu strumienia nośników ładunku elektrycznego. Idealne uformowanie strumienia elektronów w kryształach jest niemożliwe, ponieważ jego wnętrze jest zajęte regularnie rozmieszczonymi atomami. Kształt wspólnej powłoki elektronowej, po której mogą poruszać się nośniki elektronowe jest skomplikowany, bo wytworzony dzięki superpozycji pól elektrycznych wszystkich rdzeni atomowych. Ten kształt uniemożliwia prostoliniowy ruch nośników. W półprzewodniku można tylko doprowadzić do częściowego uporządkowania ruchu nośników.

Częściowe uporządkowanie należy rozumieć w taki sposób, że nośniki, poruszając się przypadkowo, zarówno co do kierunku jak i prędkości, jednak w miarę upływu czasu przemieszczają się w określoną stronę (podobnie jak liczni odwiedzający rozległy teren wystawy czy targów przemieszczają się generalnie od głównego wejścia do głównego wyjścia, pomimo indywidualnych wyborów różnych tras zwiedzania i różnych ilości czasu przeznaczonego na oglądanie eksponatów). W takich warunkach nie można mówić o prędkości czy drodze określonego, pojedynczego nośnika, można natomiast określać wartości średnie prędkości czy przemieszczenia.

Ważną rolę w przepływie prądu przez półprzewodnik odgrywają dwa zjawiska fizyczne:

- 1) dyfuzja nośników,
- 2) unoszenie nośników.

Dyfuzja nośników, to przemieszczanie się ich w kierunku malejącej koncentracji, czyli od miejsca gdzie jest ich dużo do miejsca gdzie jest ich mniej. Unoszenie nośników następuje, jeżeli półprzewodnik znajdzie się w polu elektrycznym, wtedy na każdy nośnik ładunku elektrycznego działa siła zgodna z kierunkiem pola, modyfikująca ruch nośnika.

Jeżeli w półprzewodniku wytworzy się różne koncentracje nośników, czy to przez wprowadzenie różnych ilości domieszek, czy przez dodatkowe, lokalne podgrzanie lub oświetlenie, to nastąpi ruch dyfuzyjny nośników. Ilość dyfundujących nośników jest proporcjonalna do zmienności ich koncentracji dn/dx lub dp/dx , powierzchni A , na której to zjawisko występuje i czasu jego trwania t , współczynnik proporcjonalności nazywa się stałą dyfuzji D . Ilość dyfundujących np. elektronów można zapisać jako:

$$n = \left| AtD_n \frac{dn}{dx} \right| \quad (1.5)$$

Na przykład, jeżeli w płytce o powierzchni $A = 1 \text{ cm}^2$ i grubości 1 mm wytworzyć (i utrzymywać trwale, co nie jest zadaniem prostym) liniowo zmienną koncentrację swobodnych elektronów, wynoszącą z jednej strony płytki $10^{12} [1/\text{cm}^3]$, a na jej przeciwległej powierzchni $10^{14} [1/\text{cm}^3]$, to gradient koncentracji wyniesie:

$$dn/dx = (100 \times 10^{12} - 10^{12})[1/\text{cm}^3] / 0,1[\text{cm}] = 990 \times 10^{12} [1/\text{cm}^4]$$

Uwzględniając stałą dyfuzji ($35 \text{ cm}^2/\text{s}$, jak przytoczono w tablicy 1.2), można obliczyć ilość nośników przemieszczających się w czasie każdej sekundy od jednej powierzchni płytki do drugiej:

$$n = 1[\text{cm}^2] 1[\text{s}] 35[\text{cm}^2/\text{s}] 0,99 \times 10^{15} [1/\text{cm}^4] = 35 \times 10^{15}$$

Te elektrony przeniosą łącznie w czasie jednej sekundy ładunek elektryczny o wartości:

$$q = 1,602 \times 10^{-19}[\text{C}] \times 35 \times 10^{15} = 5,6 \times 10^{-3}[\text{C}]$$

a ponieważ ładunek przepływający w jednostce czasu jest liczbowo równy natężeniu prądu, to w przedstawionych warunkach przez płytkę będzie przepływał prąd dyfuzyjny o natężeniu $5,6 \text{ mA}$. Ogólnie wartość prądu elektronowego, dyfuzyjnego podaje wzór:

$$I_{\text{ndyf}} = A q D_n \frac{dn(x)}{dx} \quad (1.6)$$

a prądu dziurowego:

$$I_{\text{pdyf}} = -A q D_p \frac{dp(x)}{dx} \quad (1.7)$$

gdzie: q – ładunek elementarny (elektronu),

A – powierzchnia,

D – stała dyfuzji odpowiednio dziur (p) i elektronów (n);

ujemny znak prądu dziurowego wynika z przyjętych konwencji znaków – dodatniej pochodnej koncentracji dziur odpowiada ich przesunięcie w kierunku ujemnym, a z kolei za kierunek techniczny prądu przyjęto kierunek ruchu ładunków dodatnich.

Pole elektryczne w półprzewodniku powoduje unoszenie nośników. Na każdy ładunek elektryczny w polu elektrycznym działa siła proporcjonalna do natężenia tego pola. Jeżeli odbywa się to w próżni, to pod wpływem tej siły ładunek porusza się prostoliniowo, ze stałym przyspieszeniem, ale wewnątrz kryształu taki sposób ruchu nie jest możliwy. W tym przypadku pole elektryczne tylko częściowo porządkuje przemieszczanie się nośników, nadając im średnią prędkość w kierunku działania sił pola. Prędkość v nośników jest wprost proporcjonalna do natężenia pola elektrycznego E , a współczynnik proporcjonalności μ nazywa się ruchliwością nośników. Uwzględniając kierunki przemieszczania się ładunków dodatnich i ujemnych w polu elektrycznym, zapisuje się:

$$v_p = \mu_p E \quad \text{oraz} \quad v_n = -\mu_n E \quad (1.8)$$

Korzystając z danych w tabl. 1.2 można obliczyć, że przyłożenie napięcia, na przykład 10V, z dwu stron płytki półprzewodnikowej o grubości 1 mm (natężenie pola elektrycznego 100 V/cm) powoduje unoszenie elektronów ze średnią prędkością 1350 m/s, tj. poprzez płytkę swobodne elektrony przesuną się w czasie mniejszym od 1 milionowej sekundy. Średnia prędkość przesuwania się dziur jest mniejsza (480 m/s), przejście przez płytkę wykonają w czasie ok. 2 μ s.

Przesunięcie ładunków elektrycznych jest inaczej przepływem prądu elektrycznego. Aby wyznaczyć natężenie tego prądu trzeba obliczyć ładunek elektryczny, który przesuwa się przez określoną płaszczyznę w jednostce czasu (np. w czasie jednej sekundy). Pamiętając, że prędkość v liczbowo jest równa drodze przebywanej w jednostce czasu, można zapisać, że przez powierzchnię A (prostopadłą do wektora prędkości) w tejże jednostce czasu przesuną się nośniki zawarte w objętości Av . Ilość tych nośników wylicza się z ich koncentracji w objętości Av , zatem ładunek elektronów przesuujących się przez powierzchnię A w jednostce czasu wynosi:

$$Q_n = q A \mu_n E n \quad (1.9)$$

a ładunek dziur:

$$Q_p = q A \mu_p E p \quad (1.10)$$

gdzie: q – oznacza ładunek elementarny,
 E – natężenie pola elektrycznego,
 n, p – koncentracje nośników,
 μ – ruchliwość nośników, odpowiednio elektronów i dziur.

Ładunek elektryczny można też wyrazić jako iloczyn natężenia prądu I oraz czasu Δt przez jaki on przepływa $Q = I \Delta t$, w czasie jednostkowym ($\Delta t = 1$) wartość liczbową natężenia prądu jest równa ładunkowi. W ten sposób można zapisać dwa prądy unoszenia w półprzewodniku, to jest prąd elektronowy:

$$I_n = A q (\mu_n n) E \quad (1.11)$$

i prąd dziurowy:

$$I_p = A q(\mu_p p)E \quad (1.12)$$

Obydwa prądy unoszeniowe, elektronów i dziur, płyną zgodnie z kierunkiem pola elektrycznego, ponieważ przeciwne są jednocześnie i znaki ładunków i kierunki ruchu tych dwu typów nośników. Całkowity prąd unoszeniowy, wywołany działaniem pola elektrycznego o natężeniu E wynosi:

$$I_{\text{unosz}} = A q(\mu_p p + \mu_n n)E \quad (1.13)$$

Dzieląc obydwie strony wzoru (1.13) przez powierzchnię A otrzymuje się związek pomiędzy gęstością prądu j a natężeniem pola E , czyli prawo Ohma w postaci:

$$j = \sigma E \quad (1.14)$$

gdzie: σ – przewodność elektryczna półprzewodnika

$$\sigma = q(\mu_p p + \mu_n n) \quad (1.15)$$

Przewodność elektryczna półprzewodnika (inaczej konduktywność albo odwrotność rezystywności) zależy od koncentracji i ruchliwości nośników ładunku elektrycznego. Wzmiankowane uprzednio metody wprowadzania domieszek do półprzewodnika pozwalają na uzyskiwanie określonych wartości koncentracji. Ruchliwość nośników jest wielkością prawie stałą w dość szerokim zakresie warunków fizycznych, dzięki temu można kształtować potrzebne wartości rezystywności elementów półprzewodnikowych. Należy jednak odnotować, że wartość ruchliwości nośników ulega silnym zmianom przy bardzo dużych zmianach: temperatury, naprężeń mechanicznych czy natężenia pola elektrycznego i dlatego w warunkach ekstremalnych mogą wystąpić odmienne zjawiska fizyczne.

2. STRUKTURY NIEJEDNORODNE

Elementy elektroniczne używane są w obwodach elektrycznych do odpowiedniego sterowania płynącym prądem. Element wykonany z jednorodnego półprzewodnika samostannego (czystego), lub domieszkowanego jest po prostu rezystorem i jako taki bywa używany w obwodach elektrycznych, ale nie może tam spełniać wszystkich niezbędnych funkcji sterowniczych czy regulacyjnych.

Znaczna większość elementów elektronicznych zawiera półprzewodnikowe struktury niejednorodne, o odpowiednich właściwościach elektrycznych. Do takich struktur należą:

- złącze pn – półprzewodnik w części o przewodnictwie typu p, a w części o przewodnictwie typu n,
- złącze ms – połączenie metalu z półprzewodnikiem,
- struktura MIS – półprzewodnik pokryty warstwą izolacyjną, na której jest umieszczona warstwa metalu.

Aby uzyskać potrzebne dla praktyki elementy elektroniczne prawie zawsze konieczne jest zastosowanie odpowiedniej kombinacji struktur niejednorodnych. Np. włączenie dowolnego półprzewodnika do obwodu elektrycznego wymaga zaopatrzenia go w elektrody metalowe, umożliwiające dalsze połączenia elektryczne, dlatego prawie każdy element półprzewodnikowy jest wyposażony w złącza ms. W niniejszym rozdziale przedstawia się ważniejsze właściwości struktur niejednorodnych.

2.1. Złącze pn

2.1.1. Złącze pn w stanie równowagi

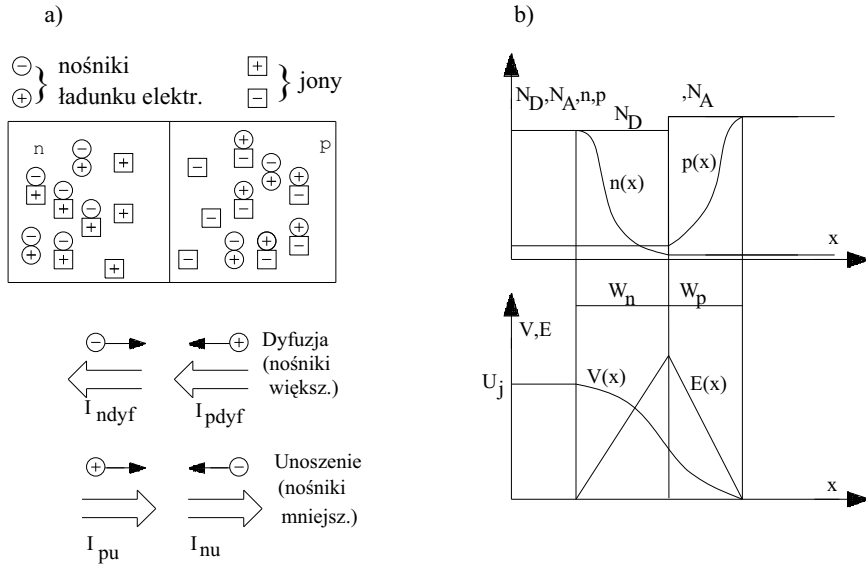
Zarys możliwości technicznych wykonania złącza pn w półprzewodniku podano w p. 1.3. Właściwości elektryczne złącza rozpatruje się na uproszczonym modelu, zakładając skokowy profil domieszek w półprzewodniku, jak pokazano na rys. 2.1. Działanie złącza pn tłumaczy się rozważając dwa zjawiska:

- 1) dyfuzję nośników większościowych,
- 2) unoszenie nośników mniejszościowych.

Dyfuzja nośników, czyli przemieszczanie się ich w kierunku malejącej koncentracji, dotyczy zarówno swobodnych elektronów jak i swobodnych dziur. Elektrony, występujące w dużej ilości w obszarze n, dyfundują do obszaru p, gdzie jest ich niewiele i podobnie dziury dyfundują z obszaru p do obszaru n.

Ponieważ średnie czasy życia nośników są skończone, podobnie jak i ich średnie prędkości, to dyfuzja może wystąpić tylko w ograniczonym obszarze wokół złącza. W tym obszarze, skutkiem dyfuzji, zmniejsza się ilość (koncentracja) nośników większościowych. Obszary położone dalej od złącza traktuje się jako niewyczerpany rezerwuuar nośników większościowych, z którego stale dopływają nośniki do rejonu przyzłączeniowego, dyfundują

przez złącze i następnie zanikają w procesie rekombinacji po drugiej stronie złącza. Ustalony rozkład koncentracji nośników w złączu pn pokazano na rys. 2.1.



Rys. 2.1. Złącze pn: przemieszczenie nośników ładunku elektrycznego (a); rozkład koncentracji domieszek $N_D(x)$ i $N_A(x)$, koncentracji nośników $p(x)$ i $n(x)$, rozkład potencjału $V(x)$ oraz natężenia $E(x)$ pola elektrycznego (b)

Ciągłe przemieszczanie się nośników ładunku elektrycznego, dyfundujących przez złącze, oznacza przepływ dyfuzyjnego prądu elektrycznego przez to złącze. Na rys. 2.1 pokazano strzałkami kierunki dyfuzyjnego przemieszczania się elektronów i dziur oraz umowne kierunki prądów dyfuzyjnych: elektronowego I_{ndyf} oraz dziurowego I_{pdyf} . Sumaryczny prąd dyfuzyjny wynosi:

$$I_{dyf} = I_{pdyf} + I_{ndyf} = -qA \left[D_p \frac{dp(x)}{dx} - D_n \frac{dn(x)}{dx} \right] \quad (2.1)$$

gdzie: q – ładunek elementarny,
 A – powierzchnia złącza,
 D – stała dyfuzji odpowiednio: dziur (p) i elektronów (n).

Innym skutkiem dyfuzji, oprócz przepływu prądu, jest lokalne naruszenie stanu równowagi ładunków elektrycznych w obszarze złącza. W części typu n ubywa swobodnych elektronów, ale ilość dodatnich jonów donorów pozostaje niezmienna, to znaczy, że pojawia się wypadkowy ładunek elektryczny dodatni (jonowy), występujący lokalnie w warstwie przyzłączowej półprzewodnika typu n. Upraszczając rozważania można przyjąć, że od płaszczyzny x_j złącza pn, na odległość W_n w głąb półprzewodnika typu n, sięga ładunek

dodatni. Podobnie, w warstwie przyłączeniowej o grubości W_p , półprzewodnika typu p, pojawia się niezrównoważony ładunek ujemnych jonów akceptorów.

Ponieważ rozpatrywane zjawisko nie jest związane z dostarczaniem lub odbieraniem ładunków elektrycznych poza kryształ, to wyodrębniający się ładunek ujemny musi być równy ładunkowi dodatniemu. Bilans tych ładunków można zapisać, pamiętając o całkowitej jonizacji domieszek w temperaturze pokojowej, jako:

$$-qN_A AW_p + qN_D AW_n = 0 \quad (2.2)$$

gdzie: N – koncentracja domieszek, odpowiednio: akceptorowej (A) i donorowej (D),

A – powierzchnia złącza,

W – szerokość stref niezrównoważonych ładunków jonowych w obszarze p i n.

Powstający w ten sposób układ ładunków elektrycznych przypomina sytuację w naładowanym kondensatorze płaskim. Podobnie, jak w naładowanym kondensatorze istnieje pole elektryczne między okładzinami, a okładziny mają różne potencjały, tak i w obszarze złącza, pojawia się pole elektryczne o natężeniu $E(x)$, obejmujące swym zasięgiem warstwy przyłączeniowe o grubości W_p i W_n oraz występuje różnica potencjałów pomiędzy obszarem n a obszarem p. Ta różnica potencjałów, zwana napięciem bariery potencjału, zależy od koncentracji wprowadzonych domieszek i wynosi:

$$U_j = V_T \ln \frac{N_A N_D}{n_i^2} \quad (2.3)$$

Wartość $V_T = kT/q$ jest to potencjał elektrokinetyczny wynoszący ok. 26 mV przy 300 K. Przy stosowanych w praktyce poziomach domieszkowania napięcie bariery potencjału krzemowych złączy pn wynosi poniżej 1 wolta.

Pole elektryczne niezrównoważonych ładunków jonowych złącza powoduje, że w złączu następuje unoszenie nośników, swobodne elektrony w obszarze złącza są unoszone w kierunku dodatniego ładunku zlokalizowanego w obszarze n, a dziury w kierunku obszaru p. Przemieszczanie nośników ładunku jest równoznaczne z przepływem prądu elektrycznego. Można więc, w obszarze złącza, określić dwa prądy unoszeniowe: elektronowy I_{nu} i dziurowy I_{pu} . Umowne kierunki tych prądów pokazano na rys. 2.1. Można zauważyć, że prądy unoszeniowe mają kierunek odwrotny do kierunku prądów dyfuzyjnych, a uwzględniając ich źródło pochodzenia stwierdza się, że są stworzone strumieniem nośników mniejszościowych. Wartość prądu unoszenia wynosi:

$$I_u = I_{pu} + I_{nu} = qA[p(x)\mu_p + n(x)\mu_n] E(x) \quad (2.4)$$

gdzie: μ – ruchliwość nośników, odpowiednio: dziur i elektronów.

Wewnętrzne pole w złączu daje więc efekt przeciwny do efektu dyfuzji i w tym sensie można mówić o przeciwdziałaniu dyfuzji przez wewnętrzne pole elektryczne. Rozkład koncentracji nośników $p(x)$ i $n(x)$ oraz natężenie pola w różnych płaszczyznach złącza $E(x)$ są wyznaczone przez łączne oddziaływanie zjawisk: dyfuzji i unoszenia, prowadzące do powstania stanu ustalonego w złączu.

W warunkach równowagi energetycznej tj. w ustalonej temperaturze, bez dostarczania innych postaci energii, w szczególności energii elektrycznej (bez włączenia złącza w obwód elektryczny) wypadkowy prąd złącza wynosi zero. Prąd dyfuzyjny ma taką samą wartość jak prąd unoszenia, ale przeciwny kierunek. Wzajemne zrównoważenie dotyczy też poszczególnych składowych prądu, tyle samo elektronów przemieszcza się dyfuzyjnie z n do p ile jest unoszonych z p do n , i podobnie z nośnikami dziurowymi. Pytanie o celowość rozważania różnych zjawisk, skoro wynikiem jest banalne stwierdzenie o zerowym prądzie, znajduje odpowiedź przy analizie warunków nierównowagi energetycznej złącza.

2.1.2. Złącze spolaryzowane

Polaryzacją nazywa się połączenie złącza pn , lub innego elementu, ze źródłem napięcia stałego. Złącze pn w obwodzie elektrycznym można spolaryzować na dwa sposoby:

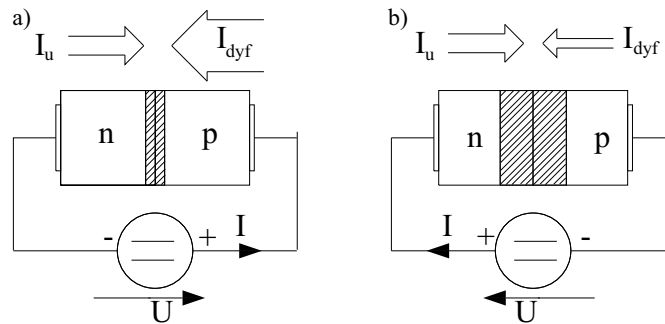
- 1) biegun dodatni zewnętrznego źródła napięcia połączyć z obszarem p a biegun ujemny z obszarem n , jest to kierunek dobrego przewodzenia złącza,
- 2) biegun dodatni z obszarem n a biegun ujemny z obszarem p , jest to kierunek złego przewodzenia złącza, zwany również kierunkiem wstecznym, możliwe połączenia pokazano na rys. 2.2.

Rozkład potencjału wzdłuż półprzewodnika zawierającego złącze zależy od jego rezystywności i wartości przepływających prądów. W obszarze występowania niezrównoważonych ładunków jonów domieszek koncentracje nośników $n(x)$ i $p(x)$ są bardzo małe (rzędu n_i i mniejsze), a rezystywność tego obszaru jest stosunkowo duża, natomiast koncentracja nośników w rejonach oddalonych od złącza jest równa koncentracji domieszek. Ponieważ koncentracja domieszek jest zwykle o kilka rzędów wyższa od wartości n_i , to przewodność półprzewodnika w częściach oddalonych od złącza jest o kilka rzędów większa od przewodności półprzewodnika w obszarze złącza. Tak więc niespolaryzowany półprzewodnik ze złączem pn można uważać za szeregowe połączenie dwu obszarów skrajnych, o małych rezystywnościach, przedzielonych obszarem o dużej rezystywności. Jeżeli do takiego półprzewodnika doprowadzi się zewnętrzne napięcie, to podział tego napięcia następuje stosownie do rezystywności, czyli praktycznie całe napięcie będzie odkładać się na stosunkowo wąskiej strefie wokół złącza.

Dołączenie zewnętrznego źródła napięcia powoduje, że w obszarze złącza pojawia się drugie pole elektryczne, o natężeniu zależnym do przyłożonego napięcia, które osłabia pole wywołane ładunkami niezrównoważonych jonów domieszek (kompensacja częściowa lub całkowita pól o przeciwnych zwrotach) albo wzmacnia je (sumowanie się dwu pól o zgodnych zwrotach). Osłabienie pola (plus do p , minus do n) powoduje zmniejszenie wartości ładunku jonowego i zwięźenie stref ładunków aż do ich zaniku, co nastąpi przy napięciu zewnętrznego źródła o wartości bliskiej wartości wewnętrznej bariery potencjału. Wzmocnienie pola (plus do n , minus do p) powoduje zwiększenie wartości ładunków jonowych i pogrubienie obszaru jaki one zajmują.

Zmiana wypadkowego pola elektrycznego w złączu zakłóca zerowy bilans prądów. Osłabienie pola zwiększa efekty dyfuzyjne, co powoduje, że pojawia się wypadkowy prąd

złącza, o charakterze prądu dyfuzyjnego. Wzmocnienie pola osłabia proces dyfuzji nośników, co powoduje, że w złączu pn pojawia się wypadkowy prąd o charakterze prądu unoszenia.



Rys. 2.2. Polaryzacja złącza pn: w kierunku przewodzenia (a), w kierunku wstecznym (b)

Prąd unoszenia powstaje dzięki ruchowi nośników mniejszościowych, docierających do obszaru złącza z rejonów oddalonych od niego. Ilość tych nośników zależy od ich koncentracji w obszarach oddalonych i średniej prędkości tzw. ruchu cieplnego w kierunku bariery potencjału – przyłożone do złącza napięcie nie ma wpływu na te zjawiska. Skutkiem tego jest niezależność natężenia prądu unoszenia od wartości napięcia przyłożonego do złącza (ale wartość tego prądu zależy np. od temperatury półprzewodnika).

Inaczej jest z prądem dyfuzyjnym, którego wartość silnie zależy od wysokości bariery potencjału, czyli też silnie zależy od wartości napięcia przyłożonego do złącza.

Znaczne osłabienie pola wewnętrznego w złączu następuje przy napięciu doprowadzonym z zewnątrz o wartości takiej, jak napięcie bariery potencjału, dla złącza krzemowego wartość ta wynosi 0,6 do 1,0 V. Przy takiej wartości napięcia zewnętrznego przepływający przez złącze prąd dyfuzyjny jest znacznie większy od prądu unoszenia. Płynący prąd jest prądem nośników większościowych, o dużym natężeniu odpowiadającym koncentracjom tych nośników. Ten prąd nazywa się prądem przewodzenia złącza.

Przy wzmacnianiu pola wewnętrznego oddziaływaniem zewnętrznym (polaryzacja w kierunku wstecznym), następuje zahamowanie dyfuzji i zmniejszenie prądu dyfuzyjnego do wartości mniejszych od prądu unoszenia. Wypadkowy prąd ma charakter prądu unoszenia czyli prądu nośników mniejszościowych i małe natężenie, odpowiadające koncentracjom tych nośników. Ten prąd nazywa się prądem wstecznym złącza.

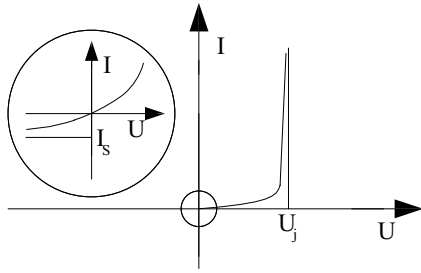
Ponieważ koncentracja nośników większościowych jest setki tysięcy razy większa od koncentracji nośników mniejszościowych (zob. p. 1.3), to prąd przewodzenia może być też setki tysięcy razy większy od prądu wstecznego. Związek pomiędzy wypadkowym prądem złącza I , a napięciem przyłożonym z zewnątrz do złącza U podano we wzorze:

$$I = I_S \left(\exp \frac{U}{V_T} - 1 \right) \quad (2.5)$$

gdzie: I_S – tzw. prąd nasycenia złącza,

V_T – potencjał elektrokinetyczny (26 mV przy 300 K).

Prąd nasycenia złącza I_S jest to prąd nośników mniejszościowych dopływających do obszaru złącza i unoszonych przez ten obszar, czyli jest to prąd unoszenia I_u . Jego wartość zależy od konstrukcji złącza, a można go interpretować jako maksymalny prąd wsteczny jaki płynąłby przy nieskończenie wielkim napięciu wstecznym.



Rys. 2.3. Charakterystyka złącza pn

Dodatnim wartościom napięcia U odpowiada osłabienie pola wewnętrznego w złączu, jest to kierunek przewodzenia złącza, ujemnym wartościom U odpowiada kierunek wsteczny. Wykres zależności $I(U)$, czyli charakterystykę prądowo-napięciową złącza pokazano na rys. 2.3.

Prąd wsteczny jest zwykle tak mały w porównaniu do prądu przewodzenia, że w praktyce jego wykres pokrywa się z osią odciętych. W kółku pokazano powiększony przebieg charakterystyki przy małych wartościach napięcia.

Widoczna niesymetria przebiegu charakterystyki prądowo-napięciowej $I(U)$ jest główną cechą złącza pn, wykorzystywaną w diodach prostowniczych, inne właściwości złącza też są użyteczne w różnych elementach elektronicznych.

2.1.3. Pojemność złącza pn

Jak wspomniano, ładunki jonowe zgromadzone po obydwu stronach złącza można porównać do ładunków zgromadzonych w naładowanym kondensatorze. Ponieważ ten ładunek, opisany we wzorze (2.2), zależy od szerokości stref ładunków jonowych W , które z kolei zmieniają się przy zmianach napięcia to można zdefiniować pojemność elektryczną $C = dQ/dU$. Jest to tzw. pojemność warstwy zaporowej C_j , jej wartość wylicza się ze wzoru podobnego do opisu pojemności kondensatora płaskiego:

$$C_j = \frac{\epsilon \epsilon_0 A}{W_p + W_n} \quad (2.6)$$

gdzie: ϵ – względna stała dielektryczna półprzewodnika,

ϵ_0 – przenikalność dielektryczna próżni.

Wartość tej pojemności jest stosunkowo niewielka, wynosi od ułamka pikofarada w złączach o małej powierzchni do kilku nanofaradów w dużych diodach.

Analizując zależność szerokości stref ładunków W od przyłożonego do złącza napięcia U , a ponadto, uwzględniając jaki to ma wpływ na wartość zgromadzonych tam ładunków elektrycznych, wyprowadza się wzór:

$$C_j(U) = \frac{C_j(0)}{\sqrt{1 - \frac{U}{U_j}}} \quad (2.7)$$

wskazujący na to, że zmiana napięcia U przyłożonego do złącza powoduje zmianę pojemności warstwy zaporowej. Przy napięciu wstecznym ($U < 0$) następuje poszerzenie warstwy ładunków jonowych, czemu towarzyszy zwiększanie zgromadzonego tam ładunku elektrycznego. Przyrost tego ładunku nie jest jednak proporcjonalny do przyrostu napięcia, bo poszerzenie strefy ładunków jednocześnie działa podobnie jak rozsuwanie okładzin kondensatora płaskiego, a zatem wzrost napięcia wstecznego powoduje zmniejszenie wartości pojemności warstwy zaporowej C_j , względem wartości $C_j(0)$, występującej przy braku napięcia polaryzującego. Złącze przy polaryzacji wstecznej zachowuje się jak kondensator o zmiennej pojemności, której wartość zależy od wartości przyłożonego napięcia.

Pojemność warstwy zaporowej występuje tylko wtedy, gdy istnieją ładunki jonowe, jak w stanie równowagi złącza lub przy polaryzacji wstecznej. Przy polaryzacji w kierunku przewodzenia ($U > 0$), szczególnie gdy zewnętrzne napięcie sięga do wartości napięcia bariery potencjału U_j , ładunki jonowe zanikają a wzór (2.6), jak i definiowanie pojemności C_j , tracą sens.

W stanie przewodzenia zwiększają się strumienie nośników ładunku elektrycznego dyfundujących przez złącze. Do obszaru typu n wpływa strumień dziur, a do obszaru typu p strumień elektronów. Zwiększający się w obszarze n ładunek dodatni dziur jest neutralizowany elektronami, dopływającymi z części obszaru n oddalonej od złącza, podobnie w obszarze p wdyfundowany ładunek elektronów jest neutralizowany dopływającymi dziurami. Ta neutralizacja, polegająca na rekombinacji nośników wstrzykiwanych przez złącze (wstrzykiwaniem nazywa się efekt dyfuzyjny wzmocniony zewnętrznym napięciem) z nośnikami większościowymi, dostarczonymi z głębi półprzewodnika, nie przebiega bezwzględnie. W związku z tym pojawia się wypadkowy ładunek elektryczny nośników wstrzykiwanych przez złącze.

Wartość tego ładunku zależy od intensywności dyfuzji, czyli pośrednio zależy od wartości napięcia przyłożonego do złącza. Ponieważ przy zmianach wartości napięcia zmienia się też ładunek wstrzykiwanych nośników, to można zdefiniować pojemność elektryczną $C = dQ/dU$. Jest to tzw. pojemność dyfuzyjna złącza C_{dyf} . Pojemnością dyfuzyjną opisuje się zjawisko występujące przy polaryzacji w kierunku przewodzenia, kiedy to przyłożone napięcie wywołuje przepływ prądu przewodzenia złącza. Analizując te zjawiska łącznie można wyznaczyć przybliżoną wartość pojemności dyfuzyjnej jako:

$$C_{dyf} \cong \frac{I}{V_T} \tau \quad (2.8)$$

gdzie: τ – czas życia nośników,
 V_T – potencjał elektrokinetyczny,
 I – prąd przewodzenia złącza.

Wartość pojemności dyfuzyjnej jest stosunkowo duża w porównaniu do pojemności np. kondensatorów używanych w technice, ale wykorzystanie jej jest praktycznie bardzo trudne, ponieważ występuje ona przy prądzie przewodzenia, co oznacza pozorne zaobcowanie tej pojemności bardzo małą rezystancją dobrze przewodzącego złącza. Istnienie tej pojemności jest łatwo obserwowane przy gwałtownej zmianie napięcia na złączu od wartości przewodzenia do polaryzacji wstecznej. Pojawia się wówczas stosunkowo duży prąd rozładowania pojemności dyfuzyjnej płynący w kierunku wstecznym złącza. Po rozładowaniu pojemności C_{dyf} prąd zmniejsza się do wartości wynikających z charakterystyki statycznej.

Warto zapamiętać, że pojemność warstwy zaporowej i pojemność dyfuzyjna przejawiają się "wymienne". Przy polaryzacji wstecznej w złączu dominuje pojemność C_j , przy polaryzacji przewodzenia pojemność C_{dyf} , nadto obydwie te pojemności są zmienne, pojemność warstwy zaporowej zależy od napięcia złącza a pojemność dyfuzyjna od jego prądu.

2.1.4. Przebiecie złącza pn

Opisane powyżej główne zjawiska występujące w złączu pn, tj. nieliniowość charakterystyki prądowo-napięciowej i nieliniowe pojemności elektryczne wynikają już z prostego modelu złącza. Istnieje jeszcze jedno ważne zjawisko, z tego modelu nie wynikające – jest to zjawisko przebiecia elektrycznego złącza. Przebiecie występuje przy polaryzacji wstecznej złącza, przy dostatecznie dużej wartości napięcia i objawia się znacznym wzrostem prądu wstecznego, wielokrotnie przewyższającym wartości wynikające z prostego modelu złącza.

Towarzyszące zazwyczaj przebieciu znaczne nagrzanie półprzewodnika w rejonie złącza może powodować bądź pęknięcie płytki półprzewodnikowej (lokalne rozszerzenie cieplne), bądź nawet lokalne stopienie kryształu. Skutkiem przebiecia jest więc często uszkodzenie kryształu zawierającego złącze, chyba, że podejmie się środki zabezpieczające kryształ przed przegrzaniem.

Wyróżnia się dwa mechanizmy przebiecia: przebiecie lawinowe oraz przebiecie Zenera. Pierwsze z nich tłumaczy się podobnie jak jonizację lawinową gazów rozrzedzonych. W silnym polu elektrycznym poruszający się z dużą prędkością elektron może zderzyć się z atomem półprzewodnika i spowodować jego jonizację, czyli uwolnić kolejną parę nośników ładunku, w tym kolejny elektron. Jeżeli następuje lawinowe powielanie takiego zjawiska to wzrasta koncentracja nośników obydwu typów, zatem i mniejszościowych, co powoduje zwiększenie natężenia prądu unoszenia. Warunkiem wystąpienia przebiecia lawinowego jest nadanie elektronom dostatecznie dużych energii i zapewnienie im dostatecznie długiej drogi aby mogły napotkać atom. Ponieważ prędkość, a zatem i energia elektronów zależy od natężenia pola elektrycznego, to określa się krytyczną wartość tego pola, przy której nastąpi jonizacja atomów. Dla krzemu ta wartość jest rzędu 10^5 V/cm. Ponieważ jednocześnie grubość warstwy zaporowej (ładunków jonowych) powinna być dość duża, aby umożliwić rozwój lawiny nośników, to przebiecie tego typu występuje przy większych wartościach napięcia, dla złączy krzemowych powyżej 7 V.

Przebiecie Zenera interpretuje się jako wyrywanie elektronów z wiązań międzyatomowych, jeżeli siły pola elektrycznego są dostatecznie duże. Skutkiem tego zjawiska jest

wzrost koncentracji nośników obydwu typów, a zatem i wzrost prądu wstecznego ponad wartości wynikające z modelu złącza pn. Podobnie jak poprzednio, warunkiem wystąpienia tego przebiccia jest dostatecznie duże natężenie pola elektrycznego, ale w tym przypadku przy małej grubości warstwy ładunków jonowych, uniemożliwiającej wystąpienie efektu lawinowego. Przebiccie Zenera w złączach krzemowych występuje przy małych napięciach, poniżej 5 V.

Możliwe jest też jednoczesne działanie obydwu mechanizmów, co ma miejsce jeżeli przebiccie występuje przy napięciu około 6 V.

Ponieważ grubość strefy ładunków jonowych zależy od koncentracji oraz profilu (rozkładu) domieszek w złączu, a wielkości te zostają jednoznacznie ustalone w procesie produkcji złącza, to typ przebiccia i wartość napięcia przebiccia są praktycznie określone przez producenta.

2.1.5. Wpływ temperatury na złącze pn

Temperatura oddziałuje na złącze pn zmieniając jego parametry. Przy przeciętnych temperaturach użytkowych, około 250 K – 350 K, co odpowiada zakresowi -20°C do $+80^{\circ}\text{C}$, te oddziaływania są następujące:

- bariera potencjału i wartość spadku napięcia w stanie przewodzenia złącza maleją o ok. 20 mV przy wzroście temperatury o 10 kelwinów,
- wsteczny prąd nasycenia wzrasta około dwukrotnie przy wzroście temperatury o 10 kelwinów,
- napięcie przebiccia Zenera maleje a napięcie przebiccia lawinowego wzrasta przy wzroście temperatury, intensywność zmian zależy od wartości napięć przebiccia.

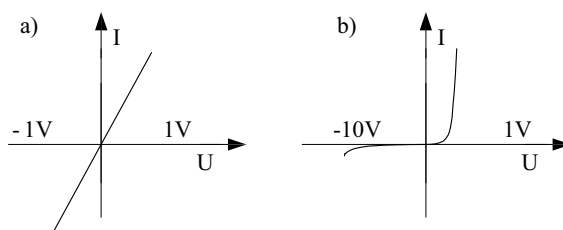
2.2. Złącze ms

Złącza metalu z półprzewodnikiem stosowano w elektrotechnice już w latach dwudziestych stulecia jako tzw. prostowniki selenowe, kuprytowe i in. W odróżnieniu od stosunkowo prostej technologii wytwarzania, teoria działania tych elementów była na tyle trudna, że właściwie w pierwszej połowie wieku nie wypracowano zadawalająco przekonujących modeli zjawisk. Obecna znajomość tych zjawisk jest znacznie lepiej rozwinięta, aczkolwiek dokładne analizy procesów elektrycznych występujących na granicy półprzewodnika i metalu nadal są dość trudne.

Złącza ms powstają przy pokryciu powierzchni np. krzemu warstwą metalu, najczęściej aluminium, srebra, złota, niklu, tantalu, wolframu a także różnych stopów, metodami napyłania, naporowywania itp. Procesy technologiczne prowadzone są w komorze próżniowej, w której wytwarza się pary metalu. Znaczne obniżanie ciśnienia (do wartości około 0,01 paskala) umożliwia parowanie metali przy temperaturach znacznie niższych od temperatur wrzenia pod normalnym ciśnieniem. Nagrzewanie metali (w formie cienkich drutów, nitok, płytek itp.) prowadzi się za pomocą prądów wysokiej częstotliwości, bombardowania strumieniem elektronów lub jonów gazów szlachetnych i podobnymi metodami. Pary metalu osiadają na we-

wewnętrznych ścianach komory i wszystkich elementach w niej umieszczonych, w tym także na płytkach półprzewodnikowych. Stosuje się też ukierunkowywanie strumienia zjonizowanych par metalu za pomocą pól elektrycznych, tak aby większość jonów metalu kierować na półprzewodnik. Grubość warstw metalowych nanoszonych na wypolerowaną, oczyszczoną i odgazowaną powierzchnię półprzewodnika jest rzędu 1 mikrometra.

Właściwości elektryczne złącza metalu z półprzewodnikiem są różne, zależą od doboru złączonych materiałów. W niektórych przypadkach złącze zachowuje się podobnie jak rezystor o małej rezystancji, tj. przewodzi dobrze prąd w obydwu kierunkach przy stosunkowo małych napięciach. Takie złącze nazywa się kontaktem (niekiedy dodając "omowy" dla podkreślenia symetrycznej charakterystyki elektrycznej). W innych przypadkach złącze ms przewodzi prąd podobnie jak złącze pn, tj. w jednym kierunku przewodzi duże prądy przy małych spadkach napięcia, a dla odwrotnego kierunku nawet dość duże wartości napięć powodują przepływ znikomo małych prądów. Takie złącze nazywa się złączem prostującym lub złączem Schottky'ego. Na rys. 2.4 pokazano charakterystyki prądowo-napięciowe dla obu przypadków.



Rys. 2.4. Charakterystyki kontaktu omowego (a), złącza prostującego (b)

Wy tłumaczenia tak różnych efektów poszukuje się, opierając się na pojęciu pracy wyjścia i różnicy tej wielkości w różnych materiałach. Praca wyjścia jest to ilość energii jaką trzeba zużyć, aby odłączyć elektron od atomu i przenieść go do poziomu energetycznego próżni, czyli w takie miejsce, gdzie nie będą działały na ten elektron jakiegokolwiek pola sił. Zakłada się, że jest tylko jeden poziom energetyczny próżni i wszystkie elektrony znajdujące się tam mają jednakową energię, niezależnie od tego, gdzie były poprzednio. Pracę wyjścia można więc traktować jako wskaźnik energii elektronów w materiale, jeżeli praca wyjścia jest duża, to oznacza, że elektrony w materiale mają małe energie i odwrotnie.

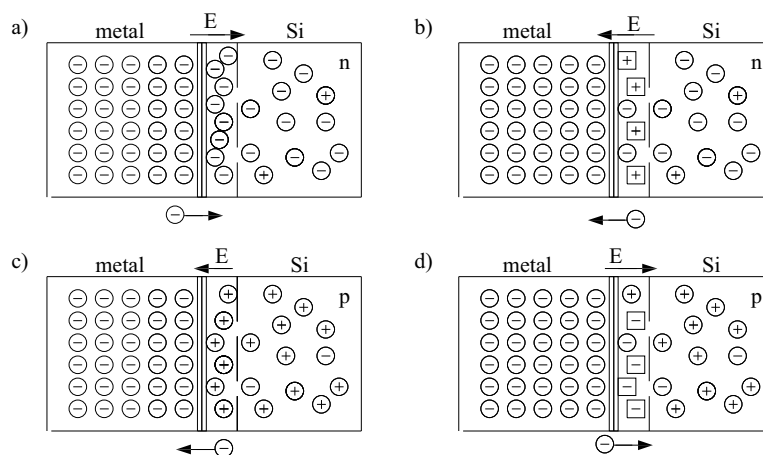
Przy złączeniu materiałów o różnych pracach wyjścia następuje przemieszczanie elektronów z materiału, w którym mają one większe energie do tego materiału, w którym energie elektronów są mniejsze, czyli następuje przepływ elektronów z materiału o mniejszej pracy wyjścia do materiału o większej pracy wyjścia. Dokładniej należy przyjąć, że obydwa złączone materiały wymieniają się wzajemnie elektronami, ale bilans nie jest zrównoważony, czyli następuje wypadkowy przepływ elektronów. Przyjmuje się, że w metalu jest nieograniczona ilość swobodnych elektronów ("gaz elektronowy"), natomiast w półprzewodniku ilość swobodnych elektronów zależy od typu przewodnictwa, tj. typu n czy p. Konsekwentnie, zmiana ilości elektronów w metalu praktycznie nie wpływa na jego rezy-

stywność, natomiast zmiana ilości swobodnych elektronów w półprzewodniku może spowodować istotne zmiany jego właściwości elektrycznych.

W złączu ms są możliwe różne sytuacje:

- metal o mniejszej pracy wyjścia niż półprzewodnik typu n,
 - metal o większej pracy wyjścia niż półprzewodnik typu n
- i podobne dwie kombinacje z półprzewodnikiem typu p. Na rys. 2.5 pokazano te możliwości i skutki przesunięcia elektronów.

W pierwszym przypadku – rys. 2.5a – elektrony przechodzą z metalu do półprzewodnika. Skutkiem tego przemieszczenia jest wzrost koncentracji elektronów w warstwie przyłączeniowej półprzewodnika, której rezystywność maleje. Jednocześnie dodatkowy, nieskompensowany, ładunek elektronów wprowadzonych do przyłączeniowej warstwy półprzewodnika wytwarza pole elektryczne powstrzymujące dopływ dalszych elektronów z metalu. W ten sposób powstaje stan stabilny złącza.



Rys. 2.5. Kontakt omowy (a) oraz (c) i złącze prostujące (b) oraz (d)

Jeżeli do takiego złącza przyłoży się z zewnątrz napięcie, to nastąpi podział tego napięcia stosownie do rezystancji poszczególnych obszarów. Rezystancja warstwy metalowej oraz rezystancja przyłączeniowej warstwy półprzewodnika, o zwiększonej koncentracji elektronów, są znacznie mniejsze aniżeli rezystancja półprzewodnika w części oddalonej od złącza. Wobec tego główny spadek przyłożonego napięcia wystąpi na warstwie o dużej rezystancji, czyli na warstwie półprzewodnika. Spadek napięcia na warstwach okołozłączowych, oraz pole elektryczne wywołane zewnętrznym napięciem, będą bardzo małe i prawie nie będą oddziaływać na właściwości elektryczne tego obszaru. Rezystywność tych warstw będzie mała, niezależnie od wartości czy kierunku polaryzującego napięcia. Tak wykonane złącze jest kontaktem omowym.

Jeżeli półprzewodnik typu n ma mniejszą pracę wyjścia aniżeli metal, to nastąpi przemieszczenie elektronów z warstwy przyłączeniowej półprzewodnika do metalu – rys. 2.5b.

W przyłączonej warstwie półprzewodnika wyodrębni się nieskompensowany ładunek dodatnich jonów donorowych, wytwarzający pole elektryczne hamujące dalszy przepływ elektronów z półprzewodnika n do metalu, co umożliwi powstanie stanu równowagi. W tym stanie przyłączona warstwa półprzewodnika n, o zmniejszonej koncentracji elektronów, ma dużą rezystywność, znacznie większą aniżeli rezystywność sąsiadujących warstw metalu i półprzewodnika. Różnica potencjałów metalu i półprzewodnika, odpowiadająca wytworzonemu polu elektrycznemu, czyli bariera potencjału Schottky'ego, zależy od różnicy wartości pracy wyjścia złączonych materiałów.

Polaryzacja złącza ms zewnętrznym źródłem napięcia stwarza dodatkowe pole elektryczne, głównie w warstwie przyłączonej o dużej rezystancji. Jeżeli to pole skompensuje pole wewnętrzne (metal dodatni, półprzewodnik n ujemny), to nastąpi przesunięcie nośników elektronowych z półprzewodnika n w kierunku metalu, co spowoduje, że rezystancja warstwy przyłączonej zmaleje. Jeżeli jest odwrotnie (metal ujemny, półprzewodnik n dodatni), to wzmocnione pole elektryczne wymusza pogrubienie warstwy pozbawionej nośników, czyli zwiększa jej rezystancję.

Zmiana rezystancji warstwy przyłączonej powoduje, że charakterystyka prądowo-napięciowa takiego złącza jest nieliniowa. Przy zwiększaniu pola wewnętrznego (metal ujemny, półprzewodnik n dodatni) przyrostowi napięcia towarzyszy przyrost rezystancji, dzięki czemu prąd, o małej wartości, pozostaje prawie stały, jest to kierunek wsteczny złącza. Przy odwrotnej polaryzacji (metal dodatni, półprzewodnik n ujemny) przyrostowi napięcia towarzyszy zmniejszanie rezystancji, dzięki czemu przyrost prądu jest znacznie większy aniżeli przyrost napięcia, jest to kierunek przewodzenia złącza ms. Charakterystyka prądowo-napięciowa prostującego złącza ms, pokazana na rys. 2.4b, ma przebieg wykładniczy, podobnie jak charakterystyka złącza pn.

Wytwarza się też złącza metalu z półprzewodnikiem typu p, podobnie jak w opisanych powyżej złączach z półprzewodnikiem n, można tam uformować kontakt omowy lub złącze prostujące jak pokazano na rys. 2.5c i d.

Przedstawiony opis złączy metal-półprzewodnik jest silnie uproszczony, głównie przez pominięcie stanów powierzchniowych półprzewodnika. Powierzchnia kryształu stanowi bardzo silne zakłócenie regularności układu atomów, co pociąga za sobą inne dopuszczalne stany energetyczne elektronów aniżeli wewnątrz kryształu. Te stany, zwane powierzchniowymi, modyfikują znacznie właściwości elektryczne powierzchni półprzewodnika. Dokładniejszą analizą tych zjawisk zajmuje się fizyka ciała stałego.

Prostujące złącza ms wykorzystywane są w różnych elementach elektronicznych, zarówno przy małych wartościach prądów (np. w funktozach logicznych) jak i przy dużych prądach jako prostownicze diody Schottky'ego. Warto odnotować, że dla typowych prostujących złączy ms, wykonywanych na krzemie, spadek napięcia w kierunku przewodzenia jest mniejszy aniżeli w złączach pn. W prostującym złączu ms także może wystąpić przebiecie przy nadmiernych napięciach w kierunku wstecznym. Ponieważ warstwa podwyższonej rezystywności występuje tylko po jednej stronie złącza, a także skutkiem trudności technologicznych w wykonaniu jednorodnych pokryć metalowych, napięcie przebiccia złączy ms jest zwykle niewysokie i nie przekracza 100 V.

Cenną zaletą prostujących złączy ms jest to, że wprowadzenie ich w stan przewodzenia nie wymaga wstrzyknięcia nośników mniejszościowych, jak w złączu pn, a tylko odtworzenia stanu pierwotnego półprzewodnika w obszarze przyłączowym. Konsekwencją tego jest brak pojemności dyfuzyjnej złącza ms, pozwala to na stosowanie złączy ms do prostowania prądów o wysokich częstotliwościach.

2.3. Struktura MIS (MOS)

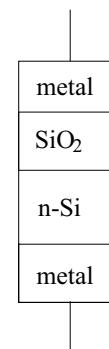
Struktura złożona z półprzewodnika (S), na którym znajduje się warstwa izolacyjna (I), pokryta warstwą metalu (M) jest ważnym elementem nowoczesnych, scalonych układów elektronicznych. W przypadku krzemu jako warstwę izolacyjną stosuje się dwutlenek krzemu (SiO_2 – szkło krzemowe) pokrywany warstwą aluminium, a od angielskiej nazwy oxide (tlenek) strukturę i elementy ją zawierające nazywa się MOS.

Na rys. 2.6 pokazano szkic budowy tej struktury, rysunek nie zachowuje proporcji struktur realnych, grubość płytki półprzewodnika wynosi około 1 milimetra natomiast grubości warstw izolacyjnej i metalowej są około 1 mikrometra. Warstwę tlenku krzemu wytwarza się wystawiając płytkę krzemową na działanie strumienia tlenu. Dobierając czas, ciśnienie i temperaturę uzyskuje się właściwą grubość utlenionej powłoki na powierzchni krzemu. Warstwę metalową nakłada się na warstwę tlenku podobnie jak przy wytwarzaniu złączy ms.

Struktura MOS przypomina kondensator o jednej okładzinie metalowej a drugiej półprzewodnikowej. Podobnie jak przez idealny kondensator nie może przepływać prąd stały, tak i w idealnej strukturze MOS przepływ prądu stałego na drodze metal–półprzewodnik jest niemożliwy, ale mogą występować zjawiska ładowania i rozładowywania pojemności struktury MOS albo przepływ prądu przemiennego przez metal, izolator i półprzewodnik.

Pojemność struktury MOS jest nieliniowa, to znaczy jej wartość zmienia się zależnie od przyłożonego napięcia. Zewnętrzne napięcie zmienia nie tylko pojemność struktury ale także modyfikuje inne cechy półprzewodnika, a ściślej jego warstwy stykającej się z izolatorem. Właśnie zmienność rezystywności półprzewodnika w strukturze MOS jest powodem jej szerokiego zastosowania w elementach elektronicznych.

Polaryzacja, pokazanej na rys. 2.6 struktury MOS, zewnętrznym napięciem powoduje różne skutki, zależnie od wartości i kierunku wytworzonego pola elektrycznego. Siły pola, działające na ładunki elektryczne, powodują przesunięcie swobodnych nośników ładunku elektrycznego. W półprzewodniku typu n swobodne elektrony są gromadzone pod warstwą izolatora, jeżeli do metalu przyłożono biegun plus a do półprzewodnika n biegun minus źródła napięcia, lub są odpychane od izolatora w głąb półprzewodnika, jeżeli do metalu przyłożono biegun ujemny.



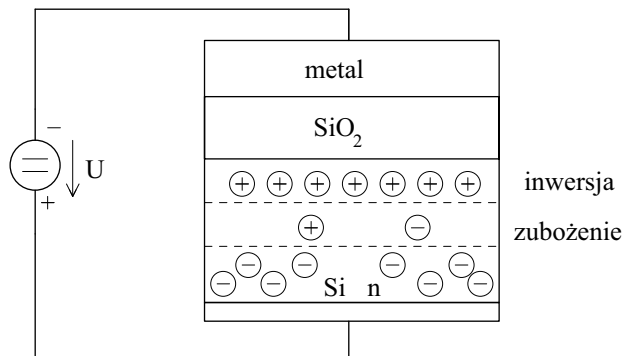
Rys. 2.6.
Struktura MOS

Zwiększenie koncentracji elektronów w przylegającej do izolatora warstwie półprzewodnika typu n powoduje zmniejszenie rezystywności tej warstwy, którą nazywa się wówczas warstwą akumulacyjną.

Zmniejszenie koncentracji nośników większościowych w części stykającej się z izolatorem, czyli wytworzenie warstwy zubożonej, powoduje wzrost rezystywności tej warstwy, ale to zjawisko występuje tylko w pewnym przedziale stosowanych napięć.

Przy ujemnej elektrodzie metalowej swobodne, większościowe, elektrony są odsuwane z obszaru półprzewodnika n znajdującego się pod izolatorem, natomiast są tam przyciągane mniejszościowe dziury. Przy określonej wartości napięcia koncentracje nośników obydwu typów mogą stać się jednakowe, taki stan odpowiada półprzewodnikowi samoistnemu, o dużej rezystywności. Warstwę o dużej rezystywności nazywa się warstwą zubożoną.

Przy dalszym zwiększeniu napięcia ujemnego, w części półprzewodnika stykającej się z izolatorem, koncentracja mniejszościowych dziur przewyższy koncentrację większościowych elektronów. W ten sposób, w półprzewodniku typu n, pod warstwą izolatora pojawi się warstwa półprzewodnika typu p. Taką warstwę nazywa się inwersyjną (odwrotną) względem technologicznie uformowanego typu przewodnictwa półprzewodnika, a ze względu na sposób jej wywołania nazywa się ją warstwą zaindukowaną (od zjawiska indukcji elektrycznej, nie elektromagnetycznej!). Warstwa inwersyjna jest oddzielona warstwą zubożoną od głębiej położonych części półprzewodnika, czyli warstwa inwersyjna jest odizolowana od podłoża warstwą o dużej rezystywności jak pokazano na rys. 2.7.



Rys. 2.7. Warstwa inwersyjna w strukturze MOS

Wartość napięcia, niezbędnego do wywołania warstwy inwersyjnej nazywa się napięciem progowym. W miarę wzrostu wartości napięcia, ponad wartość progową, rośnie koncentracja nośników (większościowych dla warstwy inwersyjnej ale mniejszościowych dla pozostałej części półprzewodnika) i wzrasta grubość warstwy inwersyjnej, obydwa te zjawiska powodują zmniejszanie rezystancji tej warstwy.

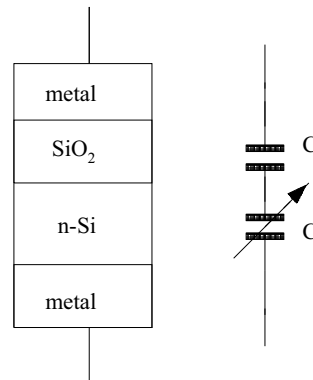
Pojemność elektryczna struktury MOS jest kształtowana przez dwa zjawiska: pojemność C_i wynikającą z obecności dielektryka (izolatora) i pojemność C_s związaną z gromadzeniem ładunków w warstwie półprzewodnika pod izolatorem. Pojemność struktury moż-

na więc przedstawić jako szeregowe połączenie dwu kondensatorów, stałego o pojemności C_i wynikającej z powierzchni, grubości i stałej dielektrycznej izolatora oraz zmiennego C_s , o pojemności zależnej od napięcia.

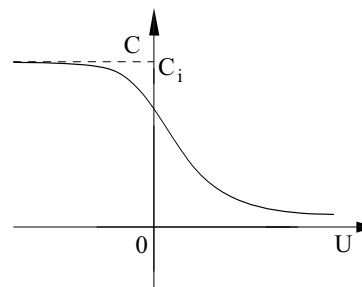
Jeżeli struktura zostaje spolaryzowana, tak aby powstała warstwa akumulacyjna (metal dodatni, półprzewodnik n ujemny) to przy stosunkowo małym napięciu zostaje tam zgromadzony stosunkowo duży ładunek nośników większościowych, w cienkiej warstwie półprzewodnika i dlatego pojemność C_s przyjmuje znaczne wartości, dużo większe od pojemności C_i . Ze względu na połączenie szeregowe wypadkowa pojemność struktury MOS jest wtedy bliska wartości pojemności warstwy izolacyjnej C_i i praktycznie nie zmienia się przy zwiększaniu napięcia polaryzacji struktury MOS. Obniżanie napięcia polaryzacji do zera powoduje zmniejszanie zakumulowanego ładunku oraz pojemności z nim związanej, wypadkowa pojemność struktury MOS zmniejsza się.

Przy odwrotnej polaryzacji (metal ujemny, półprzewodnik n dodatni) następuje zmniejszanie koncentracji elektronów w warstwie przy izolatorze i jednocześnie pogrubianie tej warstwy – jest to odpowiednik rozsuwania okładzin kondensatora przy równoczesnym odprowadzaniu ładunku elektrycznego. Oznacza to znaczne zmniejszenie pojemności C_s , wypadkowa pojemność struktury maleje. Taki stan ma miejsce aż do pojawienia się warstwy inwersyjnej, w której gromadzi się stosunkowo duży ładunek nośników mniejszościowych, w cienkiej warstwie inwersyjnej półprzewodnika, ale nadal pozostaje, głębiej położona, warstwa zubożona, której mała pojemność decyduje o wypadkowej pojemności struktury MOS. Ponieważ grubości warstwy zubożonej i jej ładunek słabo zależy od zmian napięcia to i wypadkowa pojemność struktury jest mało zależna od napięcia, jak pokazano na rys. 2.9

Podobne rozważania można przeprowadzić dla struktury MOS wykonanej na podłożu z półprzewodnika typu p, wyciągając podobne wnioski. Różnice dotyczą tylko znaków napięcia i ładunku swobodnych nośników.



Rys. 2.8. Pojemności struktury MOS



Rys. 2.9. Pojemność struktury MOS

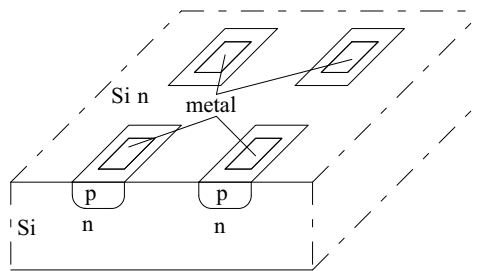
2.4. Realizacja struktur niejednorodnych

Tylko w nielicznych przypadkach całą powierzchnię płytki półprzewodnikowej, wyciętej z monokryształu, poddaje się obróbce technologicznej jednakowo, wprowadzając jednolicie domieszkę czy tworząc pojedynczą elektrodę metalową. Dotyczy to przypadku

wykonywania elementów energoelektronicznych, przeznaczonych do przewodzenia bardzo dużych prądów. Częściej domieszkowanie i inne procesy technologiczne są wykonywane tylko w określonych częściach powierzchni płytki. Aby uzyskać takie efekty, niezbędne przy wytwarzaniu układów scalonych, wykorzystuje się technikę fotolitografii.

Powierzchnię płytki krzemowej poddaje się wstępnemu utlenieniu oraz pokrywa się ją cienką (0,5 do 1,5 mikrometra) warstwą substancji światłoczułej (podobnej w działaniu do emulsji fotograficznych).

Oddzielnie przygotowuje się tzw. maskę – odpowiednik negatywu fotograficznego. Wykonuje się czarno–biały rysunek, w powiększeniu 200 do 1000-krotnym, wyróżniający miejsca, które mają być poddane procesowi technologicznemu. Po zmniejszeniu rysunku metodami fotograficznymi otrzymuje się maskę, w której odpowiednie pola są przezroczyste, a pozostałe nieprzezroczyste. Przygotowaną płytkę krzemową naświetla się, zazwyczaj ultrafioletem, poprzez maskę i poddaje wywołaniu. Przy kąpeli wywołującej naświetlone części warstwy światłoczułej (lub nienaświetlone – przy innym składzie chemicznym warstwy światłoczułej) ulegają wypłukaniu, odsłaniając lokalnie tlenek krzemu. W następnym etapie stosuje się kąpiel trawiącą, która rozpuszcza tlenek krzemu, ale nie reaguje z pozostałą częścią warstwy światłoczułej. W ten sposób zostaje odsłonięta powierzchnia krzemu w tych miejscach, gdzie należy przeprowadzić np. dyfuzję domieszki. Po właściwym zabiegu technologicznym usuwa się pozostałości fotowarstwy oraz tlenku krzemu.



Rys. 2.10. Przykład obróbki półprzewodnika

Z reguły produkcja elementów elektronicznych wymaga kilku do kilkunastu podobnych operacji, przeprowadzanych przez różne maski. Na rys. 2.10 pokazano przykład mozaiki wykonanej na płytce z krzemu typu n, gdzie w jednej operacji przeprowadzono dyfuzję domieszki akceptorowej, tworząc sieć wysepek krzemu typu p, a w drugiej operacji, przez maskę o mniejszych okienkach, naniesiono warstwę metalową, przygotowując kontakty omowe. Tak przygotowana płytka krzemowa zawiera kilkaset do kilkudziesięciu tysięcy oddzielnych złączy pn.

W technologii fotolitograficznej minimalne okno "otwierane" na powierzchni krzemu może mieć wymiary około 5×5 mikrometrów, minimalne odległości między polami są też tego rzędu. Przy stosowaniu doskonalszych technologii wymiary te mogą być jeszcze dziesięciokrotnie zmniejszone. Te wymiary wskazują na niezbędną dokładność zarówno w przygotowaniu masek, jak i w naprowadzaniu kolejnych masek na zaplanowane miejsca na

płytkę półprzewodnikowej o średnicy do 150 mm. Wysokie wymagania narzucają stosowanie złożonych urządzeń technologicznych, z szerokim udziałem sterowania komputerowego i wykorzystaniem laserów.

Kolejnym etapem technologicznym jest sprawdzanie elektryczne wykonanych elementów przy pomocy elektrod igłowych połączonych z aparaturą pomiarową i oznakowanie lakierem elementów niezgodnych z wymaganiami.

Następnie płytka jest dzielona na części przez rozłamywanie wzdłuż zarysowanych linii (podobnie jak przy cięciu szkła). Pojedyncze części, o wymiarach od około 1×1 mm dla pojedynczych elementów do ok. 10×10 mm dla złożonych układów, nazywane są czipami (ang. chip – odprysk). Nazwa "czip elektroniczny" uzyskała nieoczekiwaną popularność, a jej zakres znaczeniowy w języku potocznym został rozszerzony na wszystkie elementy scalone.

Produkcję kończy montaż czipów w obudowach, kontrola i pakowanie.

3. ELEMENTY ELEKTRONICZNE

Każde urządzenie elektroniczne (np. telewizor, komputer, regulator przemysłowy lub przekształtnik) złożone jest z wielu podzespołów i elementów mechanicznych, elektrycznych i elektronicznych. Widocznymi z zewnątrz są części mechaniczne, takie jak obudowa, wsporniki, uchwyty i in., oraz niektóre części elektryczne, jak wyłączniki, przewody itp. Wewnątrz obudowy można dostrzec dalsze elementy konstrukcji mechanicznej, połączenia elektryczne i elementy elektroniczne, zapewniające założone funkcjonowanie całego urządzenia.

Element elektroniczny to samodzielna, konstrukcyjnie nierozłączna, część urządzenia. Cechą charakterystyczną elementu elektronicznego jest więc "nienaprawialność" – element uszkodzony musi być wymieniony na inny, sprawny. Elementy są tak konstruowane i produkowane, aby było możliwe ich magazynowanie i transport bez szczególnych wymogów, a także ich instalowanie w urządzeniach przy użyciu określonych technik. Dla umożliwienia przechowywania i transportu w przeciętnych warunkach elementy są zamykane w indywidualne obudowy metalowe, lub zaprasowywane w tworzywa sztuczne, czy szkło lub, co najmniej, zabezpieczane powłokami lakierniczymi. Ponieważ do instalowania (montażu) stosuje się lutowanie, zaciskanie, owijanie, mocowanie wkrętami i podobne techniki, to elementy muszą być zaopatrzone w elektrody metalowe w postaci drutów, linek, pasków z cienkiej blachy itp.

Elementy elektroniczne, ze względu na zastosowaną konstrukcję, dzieli się na dwie grupy: elementy dyskretne i elementy scalone. Pierwsze określenie dotyczy elementów o jednolitym przeznaczeniu i funkcji, jak np. rezystor, kondensator, dioda, tranzystor itd., wykonanych w postaci kompletnego elementu z elektrodami i indywidualną obudową. Drugie określenie stosuje się do elementów, których podstawowy schemat elektryczny zawiera więcej niż jeden symbol schematowy oraz odpowiednie połączenia. Może to być np. kilka rezystorów, o odpowiednich rezystancjach, połączonych tak, jak potrzeba w określonym układzie elektronicznym, a których konstrukcja i technologia wykonania narzuca umieszczenie ich w jednej obudowie. Może to być też zespół setek tysięcy tranzystorów, diod, rezystorów i niezbędnych połączeń między nimi, tworzących razem skomplikowany układ procesora komputerowego, ale wykonanych w toku jednego ciągu operacji technologicznych i zamkniętych w jednej obudowie. Różnorodność stopnia komplikacji elementów scalonych powoduje, że producenci często nazywają je przyrządami, przyborami lub układami, aczkolwiek powinny być one nazywane elementami, ze względu na niemożność dokonywania jakichkolwiek operacji technicznych wewnątrz obudowy tzw. układu scalonego.

Na koniec, należy zauważyć, że skomplikowany wytwór techniczny nie może być opisany jedną wielkością znamionową. Dlatego opis techniczny elementu zawiera wiele parametrów, precyzujących zarówno jego podstawowe właściwości, jak i cechy drugorzędne, które też muszą być uwzględniane przez użytkownika.

3.1. Rezystory

Rezystory (oporniki) wykonywane są z drutów oporowych, mas oporowych i półprzewodników. Pierwsze dwa materiały stosuje się do produkcji elementów dyskretnych liniowych, rezystory półprzewodnikowe stanowią część układów scalonych lub są wytwarzane jako nieliniowe sensory, czyli czujniki lub przetworniki pomiarowe.

Podstawową cechą elektryczną rezystora jest jego rezystancja. Przy masowej produkcji elementów dyskretnych wytwarza się rezystory o wartościach ze znormalizowanych szeregów, przy określonych dopuszczalnych odchyłkach. Najmniej dokładne rezystory, o tolerancji 20%, wytwarza się o wartościach: 10; 15; 22; 33; 47; 68 omów i ich dziesiętnych krotnościach. Dokładniejsze rezystory (10%, 5%, 2%, ...0,1%) mają wartości z odpowiednio zagęszczonych szeregów.

Rezystory są wytwarzane jako stałe, tj. o ustalonej wartości rezystancji, i zmienne (w tym tzw. potencjometry). Zmienność rezystancji uzyskuje się wprowadzając ruchome elementy mechaniczne, umożliwiające zmianę długości czynnej części rezystora. Szczególnym przypadkiem zmienności, wykorzystywanym w sensorach, jest stosowanie materiałów, których rezystywność zmienia się przy zmianach oddziaływań zewnętrznych.

W rezystorach następuje zamiana energii elektrycznej na ciepłą, co powoduje ich nagrzewanie. Dla ograniczenia temperatury podaje się moc znamionową rezystora, której przekroczenie powoduje przegrzanie (i często uszkodzenie) tego elementu. Znamionowe moce rezystorów najczęściej stosowanych w układach elektronicznych wynoszą: 0,05; 0,125; 0,25; 0,5; 1 i 2 waty, wytwarzane są też seryjnie rezystory o większych mocach, do około 100 watów. Dopuszczalna moc zależy od powierzchni, z której jest odprowadzane ciepło, a więc i od wymiarów rezystora. Na przykład rezystor o mocy 0,125 W ma średnicę 2,5 mm i długość ok. 7 mm, a rezystor 100 W ma średnicę ok. 30 mm przy długości ok. 170 mm.

3.1.1. Rezystory dyskretnie liniowe

Rezystor drutowy powstaje przez nawinięcie drutu z odpowiedniego stopu oporowego na tzw. karkasie – rurce lub pręcie z materiału ceramicznego. Końce drutu są unieruchamiane zaciskami metalowymi, do których są dołączone przewody miedziane stanowiące wyprowadzenia montażowe. Dla unieruchomienia drutu oporowego na karkasie może być zastosowany lakier lub pokrycie warstwą ceramiczną (rezystory cementowane). Rezystancja przewodu wynosi:

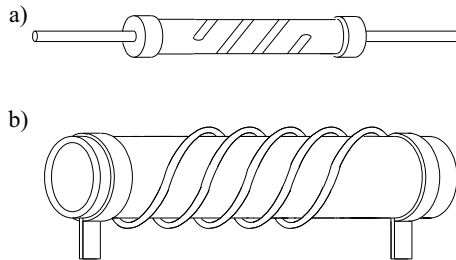
$$R = \rho \frac{l}{A} \quad (3.1)$$

gdzie: ρ – rezystywność materiału przewodu,

l – długość przewodu,

A – pole przekroju przewodu.

Dobierając parametry przewodu można uzyskać rezystor o potrzebnej rezystancji. Praktyczne ograniczenia wnoszą długość przewodu (do kilku metrów) i przekrój (nie mniej niż setne części milimetra kwadratowego). Zakres rezystancji wynosi od 0,1 oma do kilkadziesiąt kiloomów, moc zwykle nie mniej niż 2 waty.



Rys. 3.1. Rezystory: masowy (a) i drutowy (b)

Masy oporowe to mieszaniny sproszkowanych materiałów przewodzących (m.in. węgiel), izolacyjnych i półprzewodnikowych oraz lepiszcza. Z takich mas formowane są kształtki (tzw. rezystory masowe) ale częściej masy oporowe są nakładane na walcowe, ceramiczne karkasy metodą napyłania, co pozwala uzyskiwać bardzo cienkie warstwy przewodzące. Po napyleniu często stosuje się nacięcie spiralne (jak przy gwintowaniu), co daje efekt "nawinięcia" bardzo cienkiej i wąskiej wstęgi materiału przewodzącego na izolacyjnym karkasie. Na końcach karkasu zaciska się metalowe elementy z wyprowadzeniami montażowymi. Podobnie jak w rezystorach drutowych dobór materiału, grubości warstwy i sposobu jej nacięcia pozwala wytworzyć potrzebny rezystor. W taki sposób wykonywane są rezystory o rezystancjach od 10 omów do kilkudziesięciu megomów, o mocach zwykle do 2 watów. Przykłady rozwiązań konstrukcyjnych rezystorów pokazano na rys. 3.1. Warto zauważyć, że pokazane konstrukcje są podobne do zwojnicy (solenoidu) i dlatego w takich rezystorach występuje też pewna wartość indukcyjności pasywności.

3.1.2. Rezystory w układach scalonych

Rezystory w układach scalonych wytwarza się jako fragment płytki półprzewodnikowej o odpowiedniej rezystywności i wymiarach. Dodatkowym zadaniem jest oddzielenie elektryczne tej części półprzewodnika, która ma spełniać rolę rezystora od pozostałej części płytki (podłoża). Zwykle wykorzystuje się do tego celu właściwości złącza pn, tworząc w np. krzemie typu n część typu p o odpowiednim kształcie i dbając o to, aby powstałe złącze było spolaryzowane wstecznie przy normalnym włączeniu rezystora do pracy. W ten sposób zapobiega się przepływowi prądu od rezystora do podłoża, co jest równoważne odizolowaniu go od pozostałej części półprzewodnika.

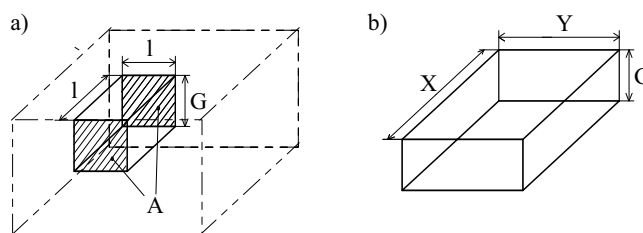
Rezystory wykonywane w półprzewodniku opisuje się charakterystycznym parametrem rezystancja/kwadrat. Ten parametr wyprowadza się z podstawowego wzoru $R = \rho l/A$ adaptowanego do metod projektowania i technologii półprzewodników. Na rys. 3.2a pokazano prostopadłościan z materiału o rezystywności ρ , jaki można wytworzyć np. metodą dyfuzji.

Rezystancja tego elementu (o długości l takiej samej jak szerokość, czyli kwadratowego w widoku z góry) pomiędzy ścianami bocznymi o powierzchni A wynosi:

$$R_k = \rho \frac{l}{A} = \rho \frac{l}{lG} = \frac{\rho}{G} \quad (3.2)$$

czyli jest zależna tylko od głębokości wprowadzenia domieszki G i rezystywności ρ , wynikającej z ilości wprowadzonej domieszki, a więc zależy tylko od sposobu prowadzenia domieszkowania. W praktyce uzyskiwane wartości wynoszą do kilkuset omów/kwadrat. Rezystancja podobnego prostopadłościanu o wymiarach X i Y , jak pokazano na rys. 3.2b, wynosi:

$$R_{XY} = \rho \frac{X}{GY} = R_k \frac{X}{Y} \quad (3.3)$$



Rys. 3.2. Rezystor półprzewodnikowy jednostkowy i dowolny

Projektując rezystor o określonej rezystancji i mocy podaje się wymiary X i Y okna w masce (por. p. 2.4) i parametry procesu domieszkowania zapewniające wymagane R_k . Rezystory zajmują stosunkowo duże powierzchnie płytek półprzewodnikowych (np. 0,3 mm × 0,05 mm, wielokrotnie więcej aniżeli tranzystor) i dlatego w układach scalonych stosuje się możliwie mało rezystorów, o małych mocach i rezystancjach z przedziału kilkaset omów do kilkudziesięciu kiloomów. Dokładność wykonywanych rezystorów jest mała (20%), ale ważną zaletą jest prawie jednakowy uchyb wszystkich rezystorów wykonanych na jednej płytce półprzewodnikowej. Dzięki temu można wytwarzać dokładne oporowe dzielniki napięcia i podobne układy, w których ważniejszy jest stosunek rezystancji dwu, lub więcej, rezystorów aniżeli bezwzględne wartości ich rezystancji.

3.1.3. Rezystory nieliniowe półprzewodnikowe

Rezystory półprzewodnikowe wykonuje się także jako elementy dyskretne, stosowane jako czujniki (przetworniki pomiarowe, sensory) różnych wielkości fizycznych. Aby maksymalnie wykorzystać wrażliwość różnych półprzewodników stosuje się różne materiały, takie jak:

- tlenek manganu, tytanian baru, tlenek wanadu w termistorach,
- siarczek kadmu, siarczek ołowiu w fotorezystorach,
- antymonek indu, arsenek galu w gaussotronach

i inne, w postaciach monokrystalicznych lub polikrystalicznych, tj. mieszaniny małych kryształków połączonych odpowiednim lepiszczem.

Szczegółowe analizy działania i sposoby stosowania tych elementów są omawiane w literaturze specjalistycznej, w niniejszym skrypcie ograniczamy się do wymienienia i krótkiego skomentowania ważniejszych odmian rezystorów półprzewodnikowych.

Termistor jest to rezystor półprzewodnikowy zmieniający rezystancję pod wpływem temperatury. Zależnie od materiału półprzewodnikowego, a także od technologii produkcji, wytwarza się termistory o rezystancji malejącej przy wzroście temperatury (NTC – Negative Temperature Coefficient), o rezystancji rosnącej z wzrostem temperatury (PTC = Positive Temperature Coefficient), i o rezystancji zmieniającej się prawie skokowo po nagraniu termistora do temperatury krytycznej (CTR – Critical Temperature Resistor). We wszystkich termistorach występuje silna zmienność rezystancji przy zmianach temperatury, sięgająca kilku do kilkudziesięciu procent wartości odniesienia przy zmianie temperatury o jeden kelwin.

Najczęściej używane są termistory NTC, w których zależność rezystancji R od temperatury bezwzględnej T ma postać:

$$R = A \exp\left(\frac{B}{T}\right) \quad (3.4)$$

gdzie: A i B są stałe materiałowe.

Przy stosowaniu termistorów NTC trzeba zwracać uwagę na to, że włączenie termistora do obwodu elektrycznego powoduje wydzielanie w nim mocy elektrycznej pod postacią ciepła. Skutki tego zjawiska mogą być różne, albo wydzielane ciepło tylko nieznacznie podgrzeje termistor, albo jest możliwy lawinowy "samopodgrzew" termistora, którego malejąca rezystancja wywoła zwiększenie mocy grzejącej. W tym drugim przypadku trzeba przewidzieć odpowiednie środki ustalające warunki termiczne pracy.

Termistory stosowane do pomiarów temperatury są wykonywane jako miniaturowe perełki o wymiarach rzędu jednego milimetra, termistory do innych zastosowań mogą być i znacznie większych rozmiarów. Termistory CTR stosuje się jako czujniki przekroczeń określonych temperatur w maszynach elektrycznych.

Fotorezystor jest to rezystor półprzewodnikowy zmieniający rezystancję pod wpływem oświetlenia. Mechanizmy fizyczne wpływu promieniowania na półprzewodnik przedstawiono nieco dokładniej w p. 3.6, tutaj ograniczamy się do stwierdzenia, że pod wpływem oświetlenia odpowiednim światłem rezystancja fotorezystora maleje. Czułość fotorezystorów jest duża, zmiana rezystancji może wynosić kilka tysięcy razy przy zmianie oświetlenia od pełnej ciemności do wartości rzędu setek luksów (oświetlenie pomieszczeń do pracy). Na rezystancję fotorezystorów ma wpływ także napięcie pracy oraz temperatura. Szybkość reakcji fotorezystorów na zmianę oświetlenia jest niezbyt duża, czasy reakcji wynoszące zwykle setne części sekundy, mogą w niektórych typach wzrastać do pojedynczych sekund. Fotorezystory wytwarzane z różnych materiałów półprzewodnikowych wykazują maksima czułości dla różnych barw światła. Typowe zastosowania fotorezystorów to pomiary natężenia oświetlenia.

Tensometr półprzewodnikowy to rezystor wykonany najczęściej z monolitycznego, domieszkowanego krzemu w postaci cienkiego pręcika o długości kilku milimetrów i przekroju mniejszym od 1 mm^2 . Poddając ten pręcik naprężeniom mechanicznym powoduje się

jego deformację, co prowadzi do zmiany rezystancji proporcjonalnej do wielkości naprężeń. Zaletą tensometrów półprzewodnikowych względem tensometrów metalowych jest to, że zmiana rezystancji jest dziesiątki lub setki razy większa aniżeli wynikałoby ze zmiany wymiarów, która zwykle nie przekracza 1 promila.

Gaussotron jest to rezystor półprzewodnikowy zmieniający rezystancję pod wpływem pola magnetycznego. Zjawisko Gaussa występuje w płytce lub folii przewodzącej, przez którą przepuszcza się prąd a jednocześnie działa na nią pole magnetyczne, o kierunku nierównoległym względem kierunku ruchu nośników ładunku elektrycznego. W takich warunkach następuje odchylenie torów nośników poruszających się w polu magnetycznym, prowadzące do przedłużenia ich drogi, co działa jak pozorne wydłużenie płytki czy folii.

Analiza zjawiska Gaussa, oczywiście bardziej precyzyjna aniżeli podane wyżej uwagi, pozwala znaleźć związek pomiędzy względną zmianą rezystancji gaussotronu $\Delta R/R$ a indukcją pola magnetycznego B w postaci:

$$\frac{\Delta R}{R} = kB^m \quad (3.5)$$

Zmiana rezystancji jest nieliniową funkcją indukcji B , wykładnik m jest bliski 2 dla słabych pól magnetycznych i przyjmuje wartość około jedności dla dużych wartości B . Stała materiałowa k dla półprzewodników przyjmuje wartości znacznie większe aniżeli dla materiałów przewodzących. Zastosowanie gaussotronów to pomiary indukcji pól magnetycznych.

3.2. Kondensatory

Podstawową cechą elektryczną kondensatora jest jego pojemność elektryczna, która dla kondensatora płaskiego wynosi:

$$C = \varepsilon \frac{A}{d} \quad (3.6)$$

gdzie: ε – przenikalność dielektryczna izolatora,

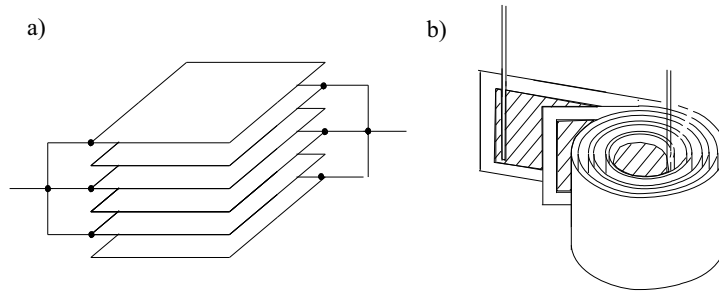
A – powierzchnia płyt,

d – odległość między płytami.

Z tego wzoru wynikają wnioski konstrukcyjne takie, że dla otrzymania znacznych wartości pojemności należy stosować płyty o dużej powierzchni, oddzielone izolatorem o małej grubości, a mającym dużą wartość ε . Kondensator w postaci płyt o dużej powierzchni byłby niewygodny w montażu i dlatego najczęściej stosowane są inne konstrukcje, wielowarstwowe, płaskie lub zwijane. Jako dielektryki stosuje się mikę, ceramikę, papier lub tworzywa sztuczne. Przykłady konstrukcji pokazano na rys. 3.3.

Jeżeli izolator jest kruchy, jak np. mika, konieczna jest wielopłytkowa konstrukcja płaska, warstwa izolacyjna nie może być bardzo cienka i kondensatory o takiej konstrukcji mają pojemności od pojedynczych pikofaradów do kilkudziesięciu nanofaradów. Jeżeli izolator jest elastyczny, jak np. folie polipropylenowe o grubości kilku mikrometrów, to

kondensator może mieć konstrukcję zwijaną. W takich przypadkach elektrody metalowe wykonywane są z folii aluminiowej lub wprost aluminium jest napyłane na folię izolacyjną. Wstęgi z takiej folii są zwijane w ścisły rulon i zaopatrywane w wyprowadzenia drutowe. Pojemności kondensatorów zwijanych sięgają 10 mikrofaradów. Jeżeli warstwy izolacyjne otrzymuje się na drodze elektrolizy (kondensatory tzw. elektrolityczne), to ich grubość wynosi ułamki mikrometra, a pojemności sięgają 10 milifaradów.



Rys. 3.3. Kondensator płaski (a) oraz zwijany (b)

Wytwarzanie kondensatorów, o konwencjonalnej konstrukcji elektroda–izolator–elektroda, w technologiach półprzewodnikowych jest możliwe, ale ze względu na duże powierzchnie zajmowane przez kondensatory, nawet o niewielkich pojemnościach, stosowane jest bardzo rzadko. Jeżeli zastosowanie pojemności w układzie scalonym jest niezbędne, to raczej stosuje się złącze pn i wykorzystuje się pojemność warstwy zaporowej lub tworzy się strukturę MOS. Tak więc najczęściej kondensatory występują jako elementy dyskretne.

Mała grubość warstw izolacyjnych powoduje, że napięcie występujące między okładzinami kondensatorów musi być ograniczone, aby nie powodowało przebicia elektrycznego izolacji. Dlatego, napięcie znamionowe kondensatora, jest równie ważnym parametrem jak pojemność. Ze względu na procesy elektryczne i elektrochemiczne, przebiegające głównie w warstwie dielektrycznej kondensatorów, określenie dopuszczalnych napięć dotyczy nie tylko wartości, ale także rodzaju napięcia – przemiennie (zakres częstotliwości), stałe, stałe ze składową zmienną itp.

3.3. Diody

Diody wytwarzane są z półprzewodnika zawierającego złącze pn lub ms, odpowiednio obudowane i zaopatrzone w elektrody. Diody są produkowane zarówno jako elementy dyskretne, jak też są one bardzo często wytwarzane jako elementy układów scalonych. Jako elementy dyskretne diody są zaopatrzone w obudowy szklane lub plastikowe, duże diody prostownicze zamykane są w obudowy metalowo-ceramiczne. Ważną cechą użytkową diod jest nieliniowość charakterystyki prądowo-napięciowej, ale także inne zjawiska elektryczne w złączach są często wykorzystywane. Z tego powodu produkuje się dziesiątki odmian

elementów, o różnorodnym zastosowaniu, zawierających w nazwie wyraz "dioda". W tym skrypcie ograniczamy się do przedstawienia tylko kilku, ważniejszych dla zastosowań przemysłowych, diod.

3.3.1. Dioda prostownicza

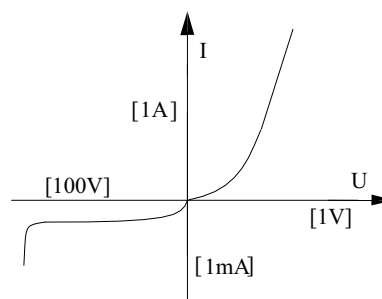
Dioda prostownicza jest przeznaczona do budowy układów przekształcających prąd przemienny, zwykle o częstotliwości 50 (60) Hz, na prąd jednokierunkowy, zawierający składową stałą. Najważniejszą cechą diody prostowniczej jest więc zdolność przewodzenia prądu tylko w jednym kierunku.

Charakterystyka statyczna, prądowo-napięciowa diody prostowniczej, pokazana na rys. 3.4, jest bardzo podobna do charakterystyk złączy pn czy ms. Niewielkie różnice pojawiają się przy dużych wartościach prądów przewodzenia, gdzie spadek napięcia na diodzie jest większy aniżeli na teoretycznym złączu.

Jest to wywołane rezystancją doprowadzeń metalowych i półprzewodnikowych, na których spadek napięcia może być dość duży, do kilku dziesiątych części wolta. Większe różnice pomiędzy diodą a złączem są przy polaryzacji w kierunku wstecznym, przy napięciach bliskich wartości napięcia przebicia, kiedy następuje znaczący wzrost prądu wstecznego.

Podstawowymi parametrami granicznymi (maksymalne, dopuszczalne wartości) diod prostowniczych są: prąd w kierunku przewodzenia, napięcie w kierunku wstecznym i temperatura złącza. Ze względu na odmienne mechanizmy uszkodzeń przy przepięciach i przetężeniach, parametry graniczne opisują wartości chwilowe dopuszczalnych napięć wstecznych, ale wartości średnie lub skuteczne dopuszczalnych prądów przewodzenia. Przekroczenie tych parametrów w trakcie użytkowania diody powoduje uszkodzenie tego elementu lub, co najmniej, sprowadza poważne zagrożenie uszkodzenia.

Obecnie są produkowane diody o dopuszczalnych prądach od 1 A (powierzchnia złącza ok. 1 mm²) do ok. 10 kA (powierzchnia złącza około 5000 mm²), dopuszczalnych napięciach od 50 V do 5 kV, dopuszczalne temperatury złącza diod krzemowych wynoszą od 110 do 180°C. Diody prostownicze, szczególnie te o wyższych parametrach, produkowane jako elementy dyskretne osiągają, wraz z radiatorami do odprowadzania ciepła, spore wymiary (np. 500×300×200 mm). Produkowane są też zespoły diod (zwykle o mniejszych prądach granicznych) w odpowiednim połączeniu, np. mostka prostowniczego lub szeregowego stosu kilkunastu diod o sumarycznej dużej wytrzymałości napięciowej, zamknięte w jednej obudowie. Innym rozwiązaniem konstrukcyjnym są tzw. moduły – dobrane przez producenta i zmontowane w odpowiednich uchwytych izolacyjnych zespoły diodowe.



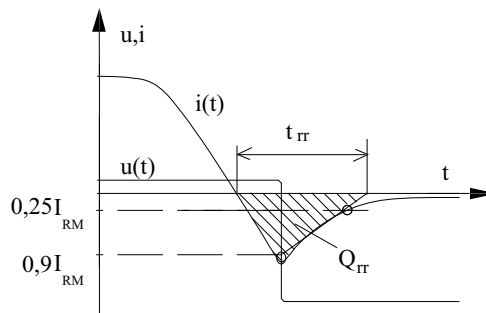
Rys. 3.4. Charakterystyka statyczna diody prostowniczej

Maksymalny, dopuszczalny prąd diody zależy głównie od powierzchni złącza, powierzchni elektrod metalowych i technologii połączenia elektrod z kryształem. Przekroczenie prądu granicznego powoduje zwykle nadmierne nagrzanie półprzewodnika i elektrody co może spowodować ich lokalne stopienie, pęknięcie lub inne uszkodzenie mechaniczne. Uszkodzenia, szczególnie w większych diodach, występują lokalnie, ze względu na technologicznie uwarunkowaną niejednorodność kryształu i elektrod na całej, stosunkowo dużej, powierzchni.

Maksymalne, dopuszczalne napięcie wsteczne zależy od wytrzymałości elektrycznej warstwy ładunków jonowych złącza, która jest funkcją konstrukcji (wymiary) i technologii (profil koncentracji domieszek) złącza. Szczególnie wrażliwe na przebicie są miejsca, gdzie złącze dochodzi do powierzchni kryształu, miejsca te są dodatkowo obrabiane i zabezpieczane. Przebicie, następujące po przekroczeniu napięcia granicznego diod prostowniczych, ma zwykle charakter lawinowy.

Dopuszczalna temperatura jest podawana jako wartość maksymalna temperatury złącza T_{jmax} . Ponieważ pomiar tej temperatury jest kłopotliwy, dla celów projektowych często podaje się dopuszczalną temperaturę obudowy T_{cmax} oraz tzw. rezystancję termiczną złącze-obudowa R_{thj-c} , zdefiniowaną jako stosunek różnicy temperatur złącze-obudowa do mocy strat w diodzie.

Właściwości dynamiczne diod prostowniczych objawiają się przy zmianach prądów i napięć. Na rys. 3.5 pokazano przebieg prądu diody i spadku napięcia na niej, w układzie zawierającym źródło napięcia o zmieniającej się biegunowości oraz rezystor ograniczający wartość prądu. Pokazane na rysunku przebiegi odpowiadają sytuacji kiedy początkowo dioda przewodzi prąd stały, od pewnego momentu napięcie źródła szybko zaczyna się zmniejszać, co powoduje zmniejszanie prądu w obwodzie, a po zmniejszeniu do zera napięcie źródła zaczyna narastać, ale w kierunku przeciwnym, polaryzującym diodę wstecznie. Przy takim działaniu obserwuje się krótkotrwały przepływ prądu wstecznego, o wartości wielokrotnie przewyższającej dane wynikające z charakterystyki statycznej.



Rys. 3.5. Wyłączanie diody

Fizyczne objaśnienie tego efektu opiera się na modelu złącza: w stanie przewodzenia prądu w obszarach przyłączowych pojawia się duży ładunek dyfuzyjny, który musi być

wyprowadzony z obszaru złącza, aby wystąpił stan polaryzacji wstecznej. Właśnie wyprowadzanie nośników wymaga przemieszczania ich w odwrotnym kierunku, aniżeli przy prądzie przewodzenia, co powoduje impuls prądu wstecznego, który można też zinterpretować jako rozładowania pojemności dyfuzyjnej diody. Ponadto konieczne jest odtworzenie stref ładunków jonowych, lub inaczej naładowanie pojemności warstwy zaporowej złącza. Oba te zjawiska opisuje się łącznie podając tzw. czas odzyskiwania zdolności zaporowej t_{tr} lub mierząc powierzchnię zawartą pod wykresem, która we współrzędnych czas-prąd ma sens ładunku elektrycznego zwanego ładunkiem przejściowym Q_{tr} .

Przy powolnej zmianie napięcia polaryzującego diodę, z kierunku przewodzenia do nieprzewodzenia, zanik zgromadzonych w niej nośników może nastąpić przez rekombinację, bez znacznego przepływu prądu wstecznego.

Wartości czasu odzyskiwania zdolności zaporowej t_{tr} diod prostowniczych w przeciętnych warunkach pracy, wynoszą od kilkunastu do kilkudziesięciu mikrosekund. Przy częstotliwości pracy 50 Hz, (okres 20 milisekund) proces dynamiczny odzyskiwania zdolności zaporowej diody prostowniczej trwa tak małą część okresu, że nie ma to istotnego wpływu na sposób jej pracy. Inaczej jest przy podwyższonych częstotliwościach pracy, gdy czas procesów dynamicznych jest współmierny z okresem powtarzania cykli działania diody. W takim przypadku dioda traci praktycznie swoją podstawową zdolność przewodzenia prądu tylko w jednym kierunku, czyli traci swoje właściwości prostownicze.

Dla prostowania prądów o podwyższonych częstotliwościach i w innych układach o dużej szybkości zmian prądów, konieczne jest stosowanie tzw. diod szybkich, o zmniejszonym czasie t_{tr} . Takie diody są wykonywane przy użyciu bardziej złożonych technologii (np. kombinowana technologia dyfuzyjno-epitaksjalna) lub przy wykorzystaniu złącza ms (diody Schottky'ego), ale ich inne parametry techniczne zwykle ulegają pogorszeniu, a cena wzrasta.

3.3.2. Dioda przełączająca (uniwersalna)

W wielu układach elektronicznych wykorzystywane są właściwości kierunkowe diod do celów innych, aniżeli prostowanie prądu. Są to zadania takie jak: zabezpieczenie innego elementu przed pojawieniem się napięcia o niewłaściwej biegunowości, narzucenie odmiennych właściwości układu dla sygnałów o różnej biegunowości, połączenie wyjść kilku układów do wspólnego punktu bez zwarcia ich itd. W takich zadaniach wymaga się często, aby dioda spełniała rolę podobną do zadań przełącznika, tj. zwierała lub rozwierała odpowiedni obwód i dlatego diody nazywa się przełączającymi, lub ze względu na szeroki i różnorodny zakres zastosowań uniwersalnymi.

Diody przełączające wykorzystują złącze pn lub ms, ich charakterystyka prądowo-napięciowa jest podobna do charakterystyki diody prostowniczej.

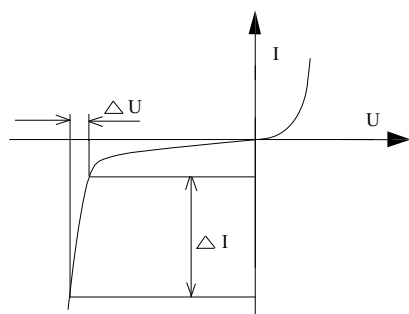
Prądy graniczne diod przełączających wynoszą od 50 do 400 miliamperów, napięcia graniczne 15 do 250 woltów, ale nie są to wielkości najważniejsze dla tych zastosowań. Ważniejszymi są parametry charakterystyczne, takie jak prąd wsteczny i spadek napięcia w kierunku przewodzenia, które powinny być możliwie małe (np. 5 nanoamperów, 0,6 wolta). Często za najważniejszą uważa się szybkość działania, tj. czasy przejścia od stanu

przewodzenia do nieprzewodzenia i odwrotnie. Czas odzyskiwania zdolności zaporowych tych diod wynosi od 2 do 200 nanosekund dla złączy pn i maleje do wartości około 50 pikosekund dla złączy ms. Niekiedy podaje się wartość pojemności diody, co umożliwia oszacowanie efektów dynamicznych diody (0,5 do 5 pF).

Diody przełączające wykonane jako elementy dyskretne zwykle są zaprasowywane w szkło lub tworzywa sztuczne, z wyprowadzeniami z drutu. Kształt obudów jest walcowy lub prostokątny, a wymiary niewielkie np. średnica 2,5 a długość 5,4 mm czy $2,5 \times 2,5 \times 4$ mm. Diody przełączające są też często wykonywane w układach scalonych przy użyciu typowych technologii półprzewodnikowych.

3.3.3. Dioda Zenera

Dioda Zenera to element przeznaczony do stabilizacji napięcia stałego. Dioda, zawierająca złącze pn o prawie skokowym profilu domieszek, ulega przebiciu w kierunku wstecznym przy niewielkich wartościach napięcia, np. kilku woltów. Jeżeli wartość prądu w stanie przebicia jest tak ograniczona, aby nie nastąpiło przegrzanie diody, to stan przebicia może być podtrzymywany dowolnie długo.



Rys. 3.6. Charakterystyka diody Zenera

W stanie przebicia napięcie na diodzie jest prawie stałe, tj. prawie niezależne od prądu płynącego przez diodę, jak pokazano na rys. 3.6 przedstawiającym charakterystykę prądowo-napięciową diody Zenera.

Dioda Zenera w stanie przebicia zachowuje się podobnie jak źródło napięcia – utrzymujące na zaciskach prawie stałe napięcie, nieznacznie zależne od prądu obciążenia. Dlatego ta dioda w stanie przebicia, może być użyta jako odpowiednik źródła napięcia stałego czy to do zasilania układu elektronicznego, czy to jako wzorzec napięcia.

Podstawowymi parametrami diody są: napięcie Zenera – wartość napięcia utrzymywana na diodzie w stanie przebicia (od ok. 3 do 30 woltów), graniczna (maksymalna, dopuszczalna) moc strat wyznaczająca maksymalny prąd diody (od 0,2 do 5 watów). Ponadto określa się parametry charakteryzujące nieidealność diody jako wzorca napięcia stałego, należą tu takie wielkości, jak: termiczny współczynnik napięcia Zenera – opisujący zmiany napięcia stabilizowanego przy zmianach temperatury, i rezystancja dynamiczna – przedstawiająca wpływ prądu diody na napięcie stabilizowane.

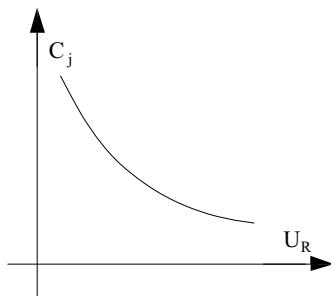
W diodach stabilizujących napięcie, zależnie od wartości napięcia przebicia, występuje albo przebicie lawinowe albo zenerowskie albo jednocześnie mogą działać obydwie mechanizmy przebicia. Niezależnie od tego który z nich działa, diody stabilizujące napięcie nazywa się diodami Zenera.

3.3.4. Dioda pojemnościowa

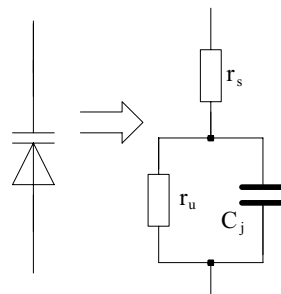
W diodach pojemnościowych wykorzystuje się pojemność warstwy zaporowej złącza pn. Te diody są używane jak kondensatory o zmiennej pojemności, sterowanej napięciem polaryzującym wstecznie diodę. Wyróżnia się dwie odmiany diod pojemnościowych – tzw. warikapów i waraktory.

Pierwsza z nich stosowana jest ściśle jako zmienna pojemność (variable capacitance), najczęściej do przestrajania, czyli zmieniania częstotliwości rezonansowej, obwodu LC. Praktyczny zakres częstotliwości pracy warikapów obejmuje przedział od setek kiloherców do setek megaherców.

Druga odmiana, waraktory, jest stosowana jako zmienna reaktancja pojemnościowa (variable reactance) w pewnej klasie układów elektronicznych, tzw. parametrycznych, stosowanych przy bardzo wysokich częstotliwościach – do setek gigaherców.



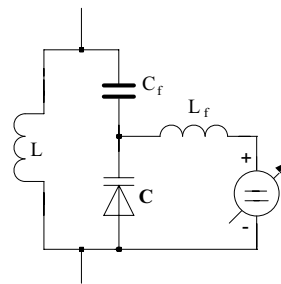
Rys. 3.7. Charakterystyka warikapu



Rys. 3.8. Schemat zastępczy warikapu

Podstawową charakterystyką warikapu jest zależność jego pojemności od napięcia, przykład charakterystyki podano na rys. 3.7. Wartości pojemności wynoszą od pojedynczych pikofaradów do dziesiątków pikofaradów. Sterowanie pojemnością odbywa się w określonym zakresie napięć wstecznych, ograniczonym od góry wytrzymałością napięciową złącza (zwykle kilkadziesiąt woltów), a od dołu minimalną wartością gwarantującą dostateczną polaryzację wsteczną (zwykle od 1 do 2 woltów). W zakresie pracy, czyli przy sterowaniu w całym, dopuszczalnym zakresie napięć, otrzymuje się kilkakrotną (2 do 5 razy) zmianę wartości pojemności.

Przy dokładniejszych analizach układów zawierających diodę pojemnościową trzeba uwzględnić obok pojemności i inne jej cechy. Na rys. 3.8 pokazano przykład schematu zastępczego, na którym przepływ prądu wstecznego złącza przedstawia rezystancja upływu r_u , a rezystywność doprowadzeń i, przede wszystkim, obszarów półprzewodnika reprezentuje rezystancja r_s .



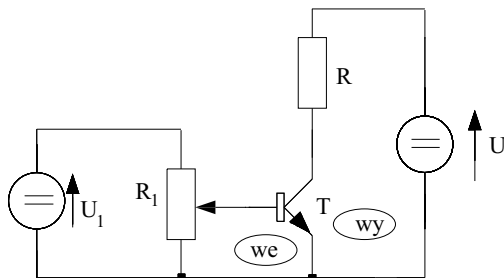
Rys. 3.9. Schemat aplikacji warikapu

Połączenie warikapu do obwodu rezonansowego wymaga rozdzielania obwodu prądu stałego od obwodu prądu przemiennego, przykład schematu pokazano na rys. 3.9. Kondensator C_f , o pojemności znacznie większej od pojemności warikapu, chroni sterujące źródło napięcia stałego przed zwarcie przez indukcyjność rezonansową L i podobnie dławik L_f , o indukcyjności znacznie większej od L , zabezpiecza obwód rezonansowy przed zwarcie przez źródło napięcia stałego.

3.4. Tranzystory

Tranzystory są podstawowymi elementami w prawie wszystkich układach elektronicznych. Główną właściwością tranzystorów jest możliwość regulowania prądu płynącego przez tranzystor włączony do obwodu elektrycznego, za pomocą sygnału elektrycznego (napięcie, prąd) działającego w drugim obwodzie elektrycznym.

Na rys. 3.10 pokazano schemat funkcjonalny włączenia tranzystora. Rezystor R symbolizuje odbiornik, w którym ma być regulowany prąd, wymuszany przez źródło napięcia stałego U . Obwód zawierający R , U oraz tranzystor jest obwodem wyjściowym dla tranzystora. Wartość prądu w tym obwodzie zależy od napięcia źródła U , rezystancji R oraz właściwości prądowo-napięciowych tranzystora T , występujących między jego elektrodami wyjściowymi. Te właściwości, z kolei, zależą od wysterowania tranzystora w jego obwodzie wejściowym, tj. od napięcia i/lub prądu doprowadzonego do elektrod tranzystora włączonych w obwód wejściowy. Na rysunku podano jako przykład "szkolny", że wysterowanie tranzystora można regulować nastawą potencjometru R_1 , zasilanego ze źródła U_1 . Zmiana nastawy potencjometru R_1 będzie powodować zmianę wartości prądu pobieranego ze źródła U i płynącego przez obciążenie, reprezentowane przez rezystor R .



Rys. 3.10. Schemat funkcjonalny zastosowania tranzystora

Ważną cechą tranzystora jest, że moc elektryczna w obwodzie wejściowym jest znacznie mniejsza od mocy w obwodzie wyjściowym; w uproszczeniu oznacza to, że małym prądem i/lub napięciem wejściowym steruje się znacznie większy prąd i/lub napięcie w obwodzie wyjściowym, ze względu na tę właściwość tranzystor nazywa się elementem wzmacniającym sygnał elektryczny.

Sterowanie prądu odbiornika w obwodzie pokazanym na rys. 3.10 można realizować na dwa sposoby. Pierwszy z nich polega na ciągłych, płynnych zmianach wartości prądu i napięcia w obwodzie wejściowym wymuszających podobne ciągłe i płynne zmiany wartości prądu obciążenia. Taki sposób pracy nazywa się analogowym. Drugi sposób polega na wykorzystywaniu sytuacji skrajnych – zmniejszanie prądu praktycznie do zera przez wprowadzenie tranzystora w stan zatkania lub odcięcia czy zablokowania albo wprowadzania tranzystora w stan możliwie dobrego przewodzenia, tak aby wartość prądu była bliska do U/R . Sterowanie polega wówczas na skokowym przełączaniu tranzystora z jednego stanu do drugiego, taki sposób pracy nazywa się impulsowym i jest stosowany w technice cyfrowej i energoelektronice. Nieco więcej informacji na ten temat przedstawiono w rozdziale 4.

Istniejące różne typy i odmiany tranzystorów można dzielić według różnych kryteriów. Podstawowym jest podział według budowy, i związanego z nią sposobu działania, na tranzystory unipolarne, bipolarne oraz, stosowane coraz częściej, struktury mieszane (np. IGBT – Insulated Gate Bipolar Transistor). Inne stosowane podziały tranzystorów dotyczą właściwości użytkowych np. podział na tranzystory małej częstotliwości lub dużej częstotliwości, impulsowe, wysokonapięciowe, dużej mocy itd. Tranzystory wytwarzane są jako elementy dyskretnie albo w strukturach scalonych (obecnie ilość tranzystorów stosowanych w układach scalonych sięga wielu setek tysięcy w jednej strukturze).

3.4.1. Tranzystory unipolarne

Tranzystory unipolarne, inaczej zwane polowymi, wykorzystują sterowanie ruchu jednego typu nośników ładunku za pomocą pola elektrycznego. Na rys. 3.11 pokazano klasyfikację tranzystorów unipolarnych według budowy, podając jednocześnie ich symbole graficzne.

TRANZYSTORY UNIPOLARNE - FET					
BRAMKA IZOLOWANA -MOS				ZŁĄCZE PN	
N	P	N	P	N	P
WZBOGAC.		ZUBOŻANIE			

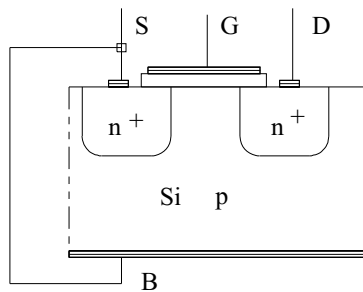
Rys. 3.11. Klasyfikacja tranzystorów unipolarnych

Tranzystory unipolarne, czyli polowe (Field Effect Transistor – FET), dzieli się się na tranzystory z izolowaną bramką, popularnie nazywane tranzystorami MOS (Metal Oxide Semiconductor), i tranzystory z bramką oddzielną złączem pn, nazywane tranzystorami PNFET lub JFET (Junction FET). Jeżeli uzyskanie prądu w obwodzie wyjściowym tranzy-

stora wymaga doprowadzenia odpowiedniego napięcia do obwodu wejściowego, to takie tranzystory nazywa się tranzystorami typu wzbogacanego (enhancement mode), jeżeli prąd wyjściowy przepływa także bez sterowania napięciowego w obwodzie wejściowym, to tranzystor nazywa się tranzystorem typu zubożanego (depletion mode). Przepływ prądu przez tranzystor odbywa się dzięki ruchowi nośników jednego typu, jeżeli są to elektrony to kanał przewodzący jest typu n, jeżeli dziury to typu p. Typ kanału pokazuje zwrot strzałki na symbolu graficznym.

Działanie tranzystorów przedstawia się na przykładzie tranzystora unipolarnego, z izolowaną bramką, z kanałem n, typu wzbogacanego, w skrócie nazywanego NMOS (lub E – NMOS dla podkreślenia także typu Enhancement).

Tranzystor NMOS wytwarza się na powierzchni płytki półprzewodnikowej, zwanej podłożem B (body). Na rys. 3.12 pokazano uproszczony przekrój tranzystora wykonanego na płytce krzemowej, o przewodnictwie typu p. W płytce wytwarza się dwa obszary n^+



Rys. 3.12. Tranzystor NMOS

(silnie domieszkowane donorowo), ułożone w odległości np. 10 μm od siebie, jeden z nich nazywa się źródłem S (Source), a drugi drenem D (Drain). Strefa pomiędzy obszarami jest przeznaczona na zaindukowany kanał. Ta strefa jest przykryta warstwą izolacyjną SiO_2 , na której wytwarza się warstwę metalową zwaną bramką G (Gate). W ten sposób, między obszarami S i D powstaje struktura MOS (zob. p. 2.3).

Obwód wyjściowy tranzystora stanowią elektrody przyłączone do obszarów źródła i drenu, obwód wejściowy czyli sterujący tranzystorem, stanowią elektrody podłoża i bramki. Warto zauwa-

żyć, że obszary S i D tworzą jednocześnie złącza pn^+ z podłożem, co stwarza niepożądane możliwości przepływu prądu od źródła lub drenu do podłoża. Aby uniemożliwić przepływ takich prądów trzeba zapewnić polaryzację wsteczną, lub co najmniej brak polaryzacji przewodzenia, obydwu tych złączy. Osiąga się to, dla rozpatrywanego tranzystora, zwierając elektrody B i S (napięcie na złączu pn wynosi zero, zatem i prąd tego złącza jest zero), oraz tak dołączając zewnętrzne źródła napięcia aby D było dodatnie względem B (przy zwarciu B–S dodatnie względem S).

Bez wystereowania tranzystora, pomiędzy elektrodą S a elektrodą D są dwa złącza pn, o przeciwnych kierunkach przewodzenia, z których każde otoczone jest strefą ładunków jonowych pozbawioną nośników, czyli praktycznie stanowiącą izolator. Ten układ uniemożliwia przepływ prądu wyjściowego od D do S; tranzystor jest w stanie zatkania (nieprzewodzenia, blokowania).

Doprowadzając napięcie z zewnętrznego źródła do elektrod bramki G i podłoża B (lub źródła S przy zwarciu B–S) tak, aby bramka była dodatnia względem podłoża, a wartość napięcia była większa od progowej ($U_{GS(th)}$) – napięcie progowe bramki wynosi zwykle 2 do 6 woltów) indukuje się kanał inwersyjny, tj. w tym przypadku typu n, zapewniający połączenie obszarów drenu i źródła bez złączy pn między nimi. Powstaje wówczas układ

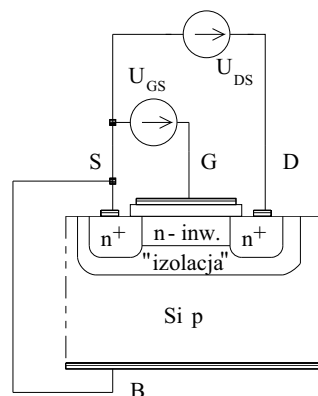
podobny do trzech rezystorów połączonych szeregowo: mała rezystancja obszaru źródła (n^+), rezystancja zaindukowanego kanału n , mała rezystancja obszaru drenu (n^+). Te trzy "rezystory" są oddzielone od podłoża izolującą strefą ładunków jonowych, jak pokazano na rys. 3.13.

Rezystancja kanału zależy od jego wymiarów i rezystywności półprzewodnika. Długość i szerokość kanału są ustalone w procesie produkcji tranzystora, natomiast jego "wysokość" (tj. jak głęboko kanał wnika w podłoże) oraz rezystywność (odwrotnie proporcjonalna do koncentracji zgromadzonych w kanale elektronów) są zależne od wartości napięcia bramki. Tak więc, zmieniając wartość napięcia U_{GS} (w zakresie przekraczającym wartość progową $U_{GS(T0)}$) zmienia się rezystancję kanału pomiędzy S a D, co prowadzi do sterowania prądem w obwodzie wyjściowym, do którego włączane są elektrody źródła i drenu.

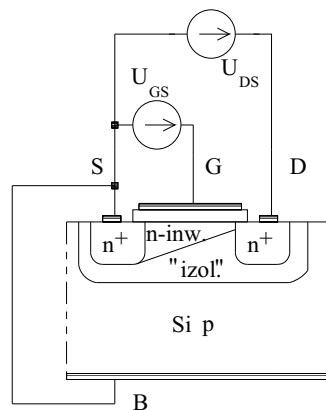
Zamknięcie obwodu wyjściowego, zawierającego także źródło napięcia, i przepływ prądu w tym obwodzie, wywołuje spadek napięcia na rezystancji kanału i stwarza różnicę potencjałów pomiędzy obszarami drenu i źródła. Jeżeli napięcie, oznaczane U_{DS} , jest małe, to nie ma istotnego wpływu na kształt i rezystancję kanału. Prąd w obwodzie wyjściowym – prąd drenu I_D – zależy od napięcia U_{DS} (prawo Ohma) i rezystancji kanału, sterowanej napięciem U_{GS} . Przy większych wartościach U_{DS} trzeba uwzględnić zmianę rozkładu pola elektrycznego w obszarze między źródłem a drenem.

W tym obszarze działają dwa pola elektryczne, o kierunkach prawie prostopadłych, jedno kształtowane w obwodzie bramka-podłoże, a drugie w obwodzie źródło-dren. Linie sił wypadkowego pola, jak i linie ekwipotencjalne, decydujące o kształcie kanału, ulegają ukosowaniu, jak pokazano na rys. 3.14.

Założmy na przykład, że napięcie sterujące U_{GS} jest większe od progowego $U_{GS(T0)}$ o 3 wolt, a napięcie U_{DS} wynosi 2 wolt. Tylko nadwyżka napięcia bramki ponad wartość progową powoduje powstanie kanału, przy obszarze źródła ta nadwyżka wynosi założone 3 wolt i powstanie tam kanał o głębokości odpowiadającej napięciu trzech woltów. Obszar drenu ma potencjał +2 V, a bramka potencjał $U_{GS(T0)} + 3$ V, zatem napięcie między obszarem drenu, wraz z przyległą częścią kanału, a bramką wynosi tylko $U_{GS(T0)} + 1$ V, w tym miejscu nadwyżka napięcia bramki ponad progowym wynosi tylko 1 wolt, czemu odpowiada znacznie mniejsza głębokość kanału, jak pokazano na rys. 3.14



Rys. 3.13. Tranzystor NMOS przy małym napięciu U_{DS}



Rys. 3.14. Ukosowanie kanału przy większym napięciu U_{DS}

W takich warunkach przewodzący kanał ma kształt klina, którego wymiary zależą od obydwu napięć U_{GS} oraz U_{DS} . Przez to prąd tranzystora jest funkcją tych dwu napięć:

$$I_D = \beta \left[(U_{GS} - U_{GS(T0)})U_{DS} - \frac{U_{DS}^2}{2} \right] \quad (3.7)$$

współczynnik β zawiera parametry konstrukcyjno-technologiczne tranzystora. Współczynnik β może być określony na drodze pomiarowej jako:

$$\beta = \frac{2I_{DSX}}{U_{GS(T0)}^2} \quad (3.8)$$

gdzie: I_{DSX} – prąd drenu przy $U_{GS} = 2U_{GS(T0)}$ i $U_{DS} = U_{GS(T0)}$.

Przedstawiony na rys. 3.14, klinowy kształt przewodzącego kanału występuje, jeżeli w rejonie drenu jest nadwyżka napięcia bramki ponad progowym, ma to miejsce dla

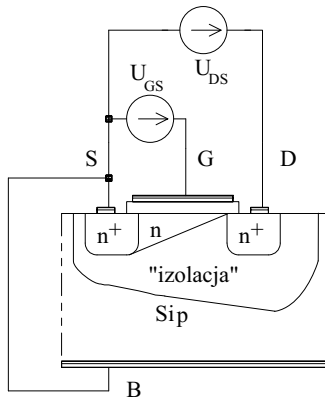
$$U_{DS} < U_{GS} - U_{GS(T0)} \quad (3.9)$$

i tak należy rozumieć określenie "małe napięcie drenu". Jeżeli zwiększa się napięcie U_{DS} tak, że nierówność (3.9) zamieni się w równość, to nastąpi zmniejszenie przekroju kanału do zera, czyli zaciśnięcie kanału w pobliżu obszaru drenu.

To zaciśnięcie nie prowadzi jednak do przerwania połączenia elektrycznego drenu ze źródłem, ponieważ w wąskiej strefie zaciskanej następuje efekt podobny do przebicia elektrycznego i prąd przepływa nadal. Takie zjawisko ma miejsce ponieważ zaciskanie kanału występuje prawie punktowo, a cała nadwyżka napięcia U_{DS} ponad wartość $U_{GS} - U_{GS(T0)}$ odkłada się na tej właśnie, bardzo krótkiej, zaciśniętej części kanału.

Prawie punktowe zaciśnięcie kanału powoduje, że występuje tam bardzo duże natężenie pola elektrycznego, umożliwiające unoszenie nośników przez zaciśniętą część kanału, co obserwuje się jako przepływ prądu. Jednocześnie kształt kanału od strony obszaru źródła pozostaje prawie niezmienny, ponieważ na części niezaciśniętej kanału występuje niezmienna część napięcia U_{DS} , o wartości $U_{GS} - U_{GS(T0)}$, nadwyżka napięcia U_{DS} ponad tę wartość odkłada się na zaciśniętej części kanału.

Na drodze od źródła do drenu są teraz dwa różne obszary: do obszaru źródła przylega zaindukowany kanał o kształcie klina, a od strony drenu wąski "kanalik przebity elektrycznie" w części zaciśniętej. Przy zwiększaniu napięcia U_{DS} część klinowa nie zmienia się, jej rezystancja pozostaje stała, co powoduje przepływ prawie stałego prądu, niezależnego od napięcia U_{DS} , ale zależnego od rezystancji części klinowej, zatem i od kształtującego ją napięcia U_{GS} , jak pokazano na rys. 3.15. W takich warunkach wartość prądu drenu opisuje wzór:

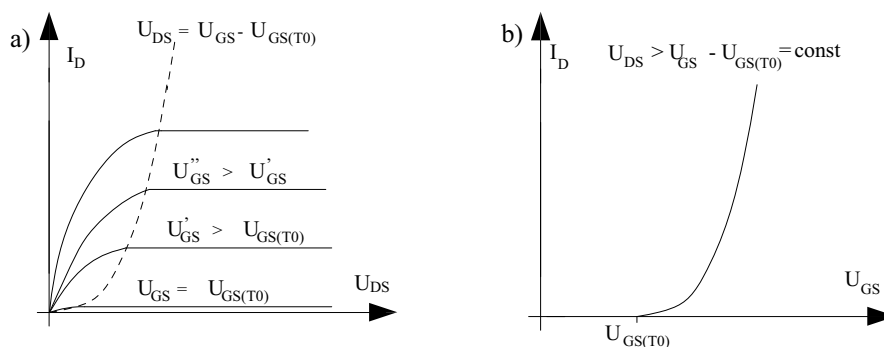


Rys. 3.15. Tranzystor NMOS przy dużym napięciu U_{DS}

$$I_D = \frac{\beta}{2}(U_{GS} - U_{GS(T0)})^2 \quad (3.10)$$

Na rys. 3.16a, pokazano charakterystyki obwodu wyjściowego tranzystora unipolarnego NMOS. Każda z charakterystyk prądowo-napięciowych $I_D(U_{DS})$, przy ustalonych wartościach napięcia sterującego U_{GS} , wykazuje dwa przedziały, odpowiadające zależnościom (3.7) i (3.10), rozgraniczenie opisuje zależność (3.9).

Przy ustalonej wartości napięcia U_{DS} prąd drenu zależy od napięcia bramka-źródło, przykład zależności pokazano na rys. 3.16b, jest to charakterystyka przejściowa lub charakterystyka sterowania prądu drenu. Na tej charakterystyce znajduje się potwierdzenie możliwości przerywania prądu drenu, jeżeli napięciu sterującemu nada się wartość mniejszą od progowej, lub powiększania prądu przez zwiększanie napięcia sterującego.



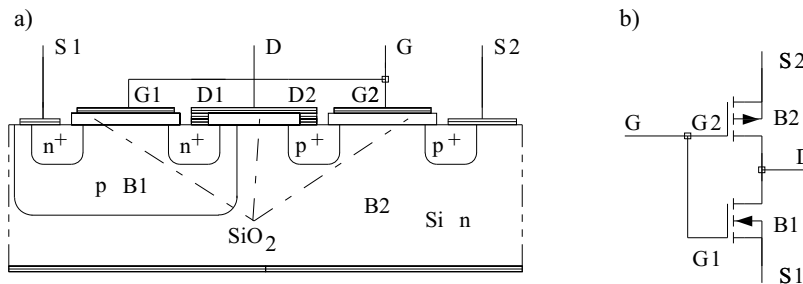
Rys. 3.16. Charakterystyki: wyjściowe (a) i przejściowa (b) tranzystora NMOS

Ważną cechą tranzystora unipolarnego jest sterowanie "napięciowe". Przy ustalonej wartości napięcia U_{GS} prąd bramki jest pomijalnie mały, ze względu na izolującą warstwę między elektrodą bramki a resztą struktury tranzystora. Jednak jeżeli do bramki doprowadzi się napięcie zmienne w czasie, to popłynie też prąd bramki, która zachowuje się wówczas jak okładzina kondensatora. Pojemność elektryczna pomiędzy bramką a źródłem jest nieliniową pojemnością struktury MOS, i dlatego podaje się jej wartość przy określonym napięciu, to samo dotyczy pojemności bramka-dren. Wartości tych pojemności są rzędu pikofarada.

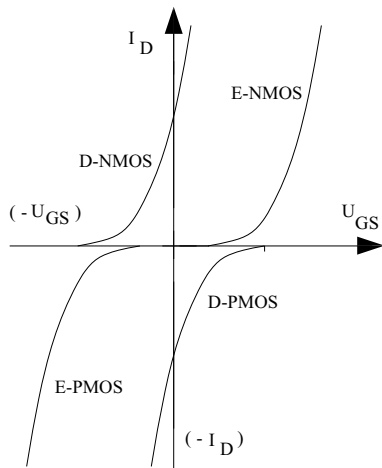
Oprócz, powyżej opisanego, tranzystora unipolarnego z izolowaną bramką, typu wzbogacanego (z kanałem indukowanym), z kanałem typu n; wytwarzane są też tranzystory o odwrotnym typie przewodnictwa. Są to tranzystory unipolarne, z izolowaną bramką, typu wzbogacanego (z kanałem indukowanym), z kanałem typu p czyli PMOS. W tym tranzystorze wszystkie procesy przebiegają tak samo jak w NMOS, ale ze zmianą znaków ładunków, prądów i napięć. Jeżeli tranzystory mają takie same parametry techniczne ale odwrotny typ przewodnictwa nazywa się je komplementarnymi.

Na rys. 3.17 pokazano przekrój płytki typu n, w której wykonano tzw. kieszeń typu p, co pozwala na wykonanie pary tranzystorów komplementarnych. Dwa tranzystory, jeden z kanałem p a drugi z kanałem n, zostają odpowiednio połączone warstwami metalizacji

dzięki czemu powstaje układ zwany inwertorem CMOS, którego schemat pokazano na rys. 3.17b. Ten układ znalazł bardzo szerokie zastosowanie w technice cyfrowej w budowie elementów logicznych CMOS. Na rysunku pokazano też sposób wykonania połączenia źródła z podłożem przez metalizację odpowiedniego okna.



Rys. 3.17. Struktura CMOS: budowa (a) i schemat (b)



Rys. 3.18. Charakterystyki przejściowe tranzystorów MOS (E – wzbogacane, D – zubożane)

Inne odmiany tranzystorów unipolarnych bardziej różnią się szczegółami budowy od opisanego tranzystora NMOS. Na przykład w tranzystorze unipolarnym typu zubożanego pomiędzy obszarami źródła i drenu zostaje technologicznie wytworzony kanał o tym samym typie domieszkowania co te obszary. Taki kanał umożliwia przepływ prądu w obwodzie dren-źródło bez sterowania bramką. Napięcie bramki, dodatnie lub ujemne względem podłoża, powoduje zwiększanie lub zmniejszanie przekroju kanału czyli sterowanie prądem drenu; aż do zatkania tranzystora. Na rys. 3.18 zestawiono charakterystyki przejściowe czterech odmian tranzystorów z izolowaną bramką.

W innym rozwiązaniu – tranzystor unipolarny złączowy – zamiast warstwy izolującej bramkę stosuje się w tym miejscu wstecznie spolaryzowane złącze pn. Taki tranzystor może być tylko typu zubożanego. Zwiększanie wstecznego

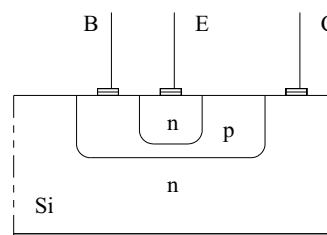
napięcia złącza bramkowego doprowadza do powstania beznóskowej strefy ładunków jonowych zajmującej cały przekrój kanału, czyli do zatkania tranzystora. Przy zmianie biegunowości napięcia sterującego, można wprowadzić złącze bramkowe w stan przewodzenia, co likwiduje możliwość sterowania rezystancją kanału i prądem drenu czyli uniemożliwia działanie tranzystora. Podobnie jak w tranzystorach z izolowaną bramką także w tranzystorach unipolarnych złączowych mogą być wytwarzane kanały n lub p.

Niezależnie od szczegółów budowy działanie wszystkich tranzystorów unipolarnych opiera się na takich samych podstawach fizycznych, opisujących wpływ pola elektrycznego na półprzewodnik.

3.4.2. Tranzystor bipolarny

Tranzystor bipolarny powstaje zwykle w procesie dwukrotnej dyfuzji domieszek do półprzewodnika. Na rys. 3.19 pokazano przekrój płytki krzemowej typu n po wykonaniu na niej tranzystora bipolarnego typu npn metodą dyfuzji. Oczywiście stosując inną kolejność operacji technologicznych można wytworzyć tranzystor typu pnp. W takim tranzystorze wszystkie procesy przebiegają podobnie jak w tranzystorze npn z uwzględnieniem odwrotnych znaków nośników, napięć i prądów. W dalszym ciągu będzie rozważany tranzystor npn.

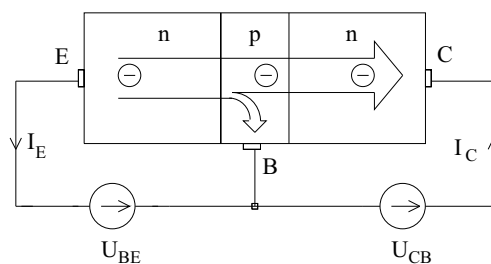
Elektrody tranzystora są dołączone do trzech jego części i nazywają się emiter E, kolektor C i baza B. Efekt tranzystorowy występuje wtedy, gdy odległość między kolektorem a emitrem czyli grubość bazy jest mniejsza od średniej drogi dyfuzji nośników ładunku elektrycznego. Elektrodamy wejściowymi, na które oddziałuje sygnał sterujący są baza-emiter, elektrodami wyjściowymi zwykle są kolektor-emiter; jest to tzw. układ wspólnego emitera.



Rys. 3.19. Budowa tranzystora bipolarnego

Dwa złącza pn tranzystora bipolarnego, tj. złącze emiterowe E–B oraz złącze kolektorowe C–B mogą być niezależnie polaryzowane zewnętrznymi źródłami napięcia. Z czterech możliwych kombinacji biegunowości napięć polaryzujących tylko jedna zapewnia tzw. stan aktywny tranzystora, umożliwiając sterowanie prądem wyjściowym.

W stanie aktywnym złącze B–E powinno być spolaryzowane w kierunku przewodzenia, a złącze B–C w kierunku wstecznym. Na rys. 3.20 pokazano model tranzystora bipolarnego, o strukturze n–p–n, połączony z dwoma źródłami napięć zapewniającymi stan aktywny.



Rys. 3.20. Polaryzacja tranzystora bipolarnego w stanie aktywnym

Polaryzacja złącza baza-emiter w kierunku przewodzenia powoduje przepływ prądu przez to złącze, wartość prądu emitera wyznacza się na podstawie zależności pomiędzy napięciem a prądem złącza, podanej w p. 2.1.2. W odniesieniu do złącza emiterowego ta zależność ma postać:

$$I_E = I_{E0} \left(\exp \frac{U_{BE}}{V_T} - 1 \right) \quad (3.11)$$

gdzie: I_{E0} – prąd nasycenia złącza emiterowego.

Przewodzenie prądu przez złącze B–E tranzystora n–p–n polega na wprowadzaniu elektronów z obszaru emitera do obszaru bazy oraz dziur z obszaru bazy do obszaru emitera. Dobierając koncentracje domieszek w tych obszarach dąży się do tego, aby składowa elektronowa była wielokrotnie większa od dziurowej, tj. aby prawie cały prąd złącza emiterowego przepływał dzięki przemieszczaniu, "wstrzykiwaniu", elektronów do obszaru bazy.

Elektrony wprowadzone do obszaru bazy p stają się tam nośnikami mniejszościowymi i, jeżeli znajdują się w pobliżu wstecznie spolaryzowanego złącza B–C, zostają unoszone do obszaru kolektora. Część elektronów przesuujących się od złącza B–E do złącza B–C ulega rekombinacji, co obserwuje się w postaci prądu bazy, na który składają się też inne składniki, ale podstawowa ilość elektronów, wstrzykniętych do cienkiej warstwy bazy, dociera do obszaru kolektora. Dzięki temu z obszaru kolektora wypływa prąd I'_C , wynoszący ok. 90% do 99% prądu emitera:

$$I'_C = \alpha_F I_E \quad (3.12)$$

współczynnik α_F nazywa się współczynnikiem wzmocnienia prądowego w układzie wspólnej bazy, jego wartość wynosi 0,9 do 0,99. Pozostała część prądu emitera – 1 do 10% – wypływa jako prąd bazy.

Także nośniki mniejszościowe w obszarze kolektora, tj. dziury, są unoszone polem elektrycznym złącza B–C do obszaru bazy, ale ich ilość jest znacznie mniejsza od ilości elektronów wstrzykniętych do obszaru bazy. Prąd tych nośników I''_C można opisać wychodząc z równania złącza w odniesieniu do złącza B–C i zakładając ponadto, że napięcie wsteczne tego złącza jest znacznie większe od V_T (26mV):

$$I''_C = I_{C0} \left(\exp \frac{-U_{BC}}{V_T} - 1 \right) \approx -I_{C0} \quad (3.13)$$

Uwzględniając kierunek zastrzałkowania prądu kolektora i obydwie jego składowe można prąd I_C zapisać jako:

$$I_C = \alpha_F I_{E0} \left(\exp \frac{U_{BE}}{V_T} - 1 \right) + I_{C0} \approx \alpha_F I_{E0} \exp \frac{U_{BE}}{V_T} + I_{C0} \quad (3.14)$$

Ze wzoru (3.14) wynika, że prąd kolektora ma dwie składowe: sterowaną napięciem baza-emiter oraz niesterowaną, zwaną prądem wstecznym.

Proporcja podziału prądu emitera na prądy kolektora i bazy, czyli współczynnik α_F , jest wielkością prawie stałą dla danego tranzystora. Oznacza to, że można np. zmieniać

wartość prądu emitera wywołując proporcjonalne zmiany prądów kolektora i bazy, albo zmieniać wartość znacznie mniejszego prądu bazy, powodując proporcjonalne zmiany znacznie większych prądów emitera i kolektora. Ten ostatni sposób sterowania tranzystorem bipolarnym, zapewniający wzmocnienie prądowe, uzyskał bardzo duże znaczenie w wielu układach elektronicznych.

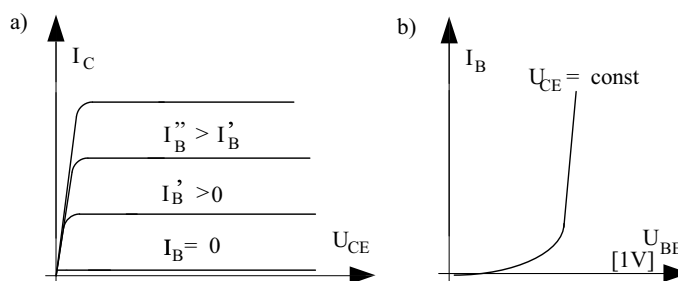
Zapewnienie stanu aktywnego tranzystora i wywołanie przepływu prądów wymaga doprowadzenia napięć nie tylko o odpowiednich zwrotach ale i wartościach. Złącze baza-emiter, spolaryzowane w kierunku przewodzenia, zachowuje się podobnie do przewodzącej diody, tj. po przekroczeniu wartości progowej (dla krzemu 0,6 V do 0,7 V, por. punkt 2.1), dalsze, nawet nieznaczne, przyrosty wartości napięcia U_{BE} (miliwoltowe) powodują silne zmiany prądu bazy I_B , który może zwiększać się kilkadziesiąt czy kilkaset razy (w zakresie mikro- i miliamperów). Z tego powodu określa się, że tranzystor bipolarny jest sterowany prądowo – prądem bazy – często przyjmując w uproszczeniu, że napięcie baza-emiter jest stałe.

Polaryzację złącza baza-kolektor wprowadza się zwykle pośrednio, przez elektrody emitera i kolektora. Ponieważ:

$$U_{CB} + U_{BE} = U_{CE} \quad (3.15)$$

to przy prawie stałej i niewielkiej wartości U_{BE} jest $U_{CB} \approx U_{CE}$ zatem napięcie kolektor-emiter powinno mieć taki zwrot, aby spolaryzować złącze B–C w kierunku wstecznym, a wartość tego napięcia powinna być większa od ok. 0,6 V. Maksymalną wartość tego napięcia ogranicza możliwość wystąpienia przebiccia.

Tranzystor bipolarny można też wprowadzić w stany inne aniżeli aktywny. Na przykład zmniejszając napięcie baza-emiter można sprowadzić sterujący prąd bazy do zera, co powoduje przerwanie prądu wyjściowego kolektora, ale nie całkowite, bo pozostaje niesterowalny prąd zerowy. Wartość prądu zerowego jest zwykle pomijalnie mała wobec prądów roboczych, a taki stan tranzystora nazywa się stanem zatkania, odcięcia lub nieprzewodzenia.



Rys. 3.21. Charakterystyki: wyjściowe (a) i wejściowa (b) tranzystora bipolarnego

Zmniejszenie wstecznej polaryzacji złącza B–C do zera, przy zachowaniu polaryzacji przewodzenia złącza B–E, wprowadza tranzystor w stan nasycenia, kiedy obydwa złącza są zdolne do przewodzenia znacznych prądów przy małych napięciach.

Można też wprowadzić tranzystor w stan inwersyjnie aktywny, zamieniając rolami emiter z kolektorem. Pomimo że uproszczony rysunek, jak np. 3.20, sugeruje symetrię budowy tranzystora, to jednak rzeczywista budowa (rys. 3.19) nie jest symetryczna. Skutkiem tego działanie tranzystora w stanie inwersyjnie aktywnym jest podobne do działania w stanie aktywnym, ale jego parametry techniczne ulegają znacznym zmianom.

Na rys. 3.21a pokazano charakterystyki prądowo-napięciowe $I_C(U_{CE})$ dla ustalonych wartości prądu bazy I_B , określające właściwości obwodu wyjściowego tranzystora bipolarnego, głównie w stanie aktywnym. W tych samych współrzędnych można też przedstawić charakterystyki odpowiadające innym stanom. Na przykład charakterystyka pokazana na rys. 3.21a dla $I_B = 0$ odpowiada stanowi zatkania tranzystora bipolarnego, a prawie pionowe odcinki charakterystyk pokazanych na rys. 3.21a odpowiadają stanowi nasycenia. Na rys. 3.21b pokazano charakterystykę obwodu wejściowego $I_B(U_{BE})$ przy ustalonej wartości napięcia kolektor-emiter.

Każde złącze ma określoną pojemność, tak więc i w tranzystorze bipolarnym można określić nie mniej niż dwie pojemności. Ładunki zgromadzone w tych pojemnościach są związane z napięciami i prądami tranzystora. Można wykazać, że prąd kolektora I_C jest proporcjonalny do ładunku elektrycznego Q_B zgromadzonego w obszarze bazy:

$$I_C = \frac{Q_B}{t_B} \quad (3.16)$$

gdzie: t_B – czas przelotu wstrzykniętych nośników przez obszar bazy.

Wzór (3.16) można zinterpretować następująco: każdej wartości prądu kolektora odpowiada określony ładunek zgromadzony w bazie, zatem każda zmiana prądu I_C wymaga zmiany ładunku Q_B . Wymuszając, nawet skokową, zmianę prądu emitera i bazy do nowej wartości, powodujemy liniowe zmiany ładunku zgromadzonego w bazie ($dQ = i dt$), a zatem i liniowe zmiany prądu kolektora, trwające aż do ustalenia wartości tego prądu odpowiadającej nowej wartości prądu emitera i bazy.

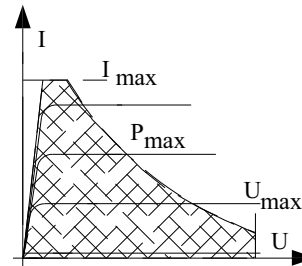
3.4.3. Podobieństwa i różnice między tranzystorami

Opisane powyżej odmiany tranzystorów, o różnej budowie i zjawiskach wewnętrznych, wykazują, obok swoistych właściwości, wiele cech podobnych, szczególnie z punktu widzenia użytkowników.

Każdy tranzystor ma określoną odporność na dołączane napięcia i prądy, zbyt duże napięcia powodują przebicia elektryczne prowadzące często do zniszczenia elementu, zbyt duże prądy powodują pewne uszkodzenia. Ponadto może występować przegrzanie tranzystora (przekroczenie maksymalnej temperatury pracy) powodujące bądź całkowite zniszczenie, bądź istotne pogorszenie innych parametrów tranzystora. Powodem przegrzania jest wydzielenie nadmiernej mocy elektrycznej, zamieniającej się w tranzystorze na ciepło (strat z punktu widzenia użytkownika). Zatem moc, określona w uproszczeniu przez iloczyn prądu i napięcia w obwodzie wyjściowym, jest też ograniczona. Z reguły ograniczenia są takie, że jednoczesne wystąpienie maksymalnych, dopuszczalnych wartości prądu

i napięcia powoduje wielokrotne przekroczenie dopuszczalnej mocy strat tranzystora. Na rys. 3.22 pokazano typowe rozmieszczenie ograniczeń, określone jako tzw. obszar bezpiecznej pracy – SOA (Safe Operating Area), na charakterystykach prądowo-napięciowych tranzystora.

Każdy tranzystor, w pewnym zakresie napięć polaryzujących jego elektrody, zachowuje się jak sterowane źródło prądowe, tj. wartość prądu wyjściowego I nie zależy (lub prawie nie zależy) od napięcia w obwodzie wyjściowym U (charakterystyki $I(U)$ obwodu wyjściowego równoległe, lub prawie równoległe do osi U). W takich warunkach, na schemacie funkcjonalnym, pokazanym na rys.

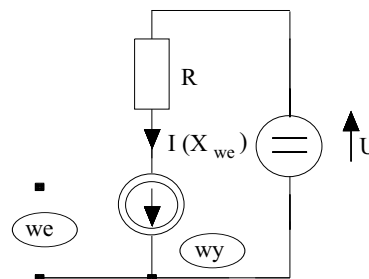


Rys. 3.22. Obszar bezpiecznej pracy

3.10, można zamiast tranzystora wprowadzić sterowane źródło prądowe, jak pokazano na rys. 3.23, gdzie wielkością X_{we} , sterującą tym źródłem jest: napięcie U_{GS} jeżeli źródło reprezentuje tranzystor unipolarny, lub prąd bazy I_B w przypadku tranzystora bipolarnego.

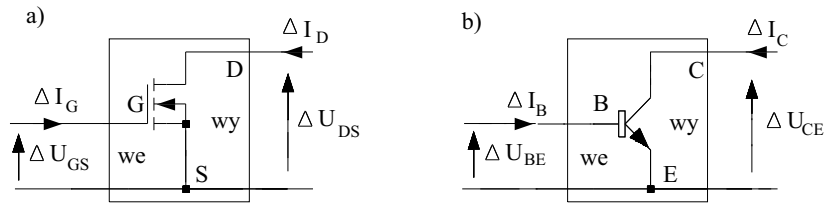
Przedstawione na rys. 3.23 zastąpienie tranzystora elementem, pozwalającym na proste obliczanie obwodu elektrycznego, jest przykładem wprowadzania modeli zastępczych, umożliwiających analizę układów elektronicznych zawierających elementy nieliniowe, takie jak tranzystory.

Model zastępczy tranzystora może być utworzony na podstawie teoretycznej analizy odpowiednich zjawisk fizycznych, w postaci wzorów analitycznych, jak np. powyżej przytoczone wzory (3.7), (3.14) i im podobne. W takich przypadkach trzeba tworzyć model możliwie kompletny, zawierający np. opis matematyczny pojemności, efektów przebieciowych i in., a także poprawić zbieżność modelu matematycznego z wynikami pomiarowymi, wprowadzając dodatkowe uzależnienia parametrów modelu od temperatury czy napięć w postaci dodatkowych współczynników i funkcji. Modele budowane na tej drodze dość wiernie odpowiadają rzeczywistym właściwościom tranzystorów, ale ich stopień komplikacji rachunkowej jest znaczny. Takie modele, używane w programach symulacyjnych jak PSPICE, zawierają po kilkanaście do kilkudziesięciu parametrów, powiązanych układem wielu równań, oczywiście posługiwanie się tym modelem bez komputera nie ma sensu.



Rys. 3.23. Sterowane źródło prądowe jako ekwiwalent tranzystora

Inne, znacznie prostsze, modele budowane są na bazie czwórnikowego, zlinearyzowanego opisu tranzystora. Na rys. 3.24 pokazano tranzystory potraktowane jako czwórniki, ponieważ jedną elektrodę tranzystora wykorzystuje się zarówno w obwodzie wejściowym jak i wyjściowym, to istnieją trzy różne układy przetworzenia tranzystora w czwórnik. Na rysunku pokazano najczęściej używane w elektronice przemysłowej połączenia w układzie OS i OE (wspólne źródło i wspólny emiter).



Rys. 3.24. Tranzystor jako czwórnik: unipolarny w układzie wspólnego źródła (a) i bipolarny w układzie wspólnego emitera (b)

Linearyzację tranzystora uzyskuje się zakładając, że zmiany prądów i napięć w tranzystorze będą na tyle małe, że związki między nimi można traktować jako liniowe. Inaczej oznacza to, że z rzeczywistych, nieliniowych charakterystyk tranzystorów wybiera się tylko odcinki na tyle krótkie, aby można było traktować je jako prostoliniowe. Taki typ linearyzacji zaznaczymy, używając w opisie symboli Δ , dla oznaczenia tylko zmiany (przyrostu) odpowiedniej wielkości. Powszechnie przyjęto opis tranzystora unipolarnego macierzą parametrów konduktancyjnych y_{ij} , a tranzystora bipolarnego macierzą parametrów hybrydowych h_{ij} , uzasadnienie wyboru takich macierzy można znaleźć zarówno w analizie działania jak i możliwościach pomiarowych tranzystorów.

Tranzystor unipolarny opisuje się równaniami:

$$\begin{aligned}\Delta I_G &= y_{11s}\Delta U_{GS} + y_{12s}\Delta U_{DS} \\ \Delta I_D &= y_{21s}\Delta U_{GS} + y_{22s}\Delta U_{DS}\end{aligned}\quad (3.17)$$

a tranzystor bipolarny równaniami:

$$\begin{aligned}\Delta U_{BE} &= h_{11e}\Delta I_B + h_{12e}\Delta U_{CE} \\ \Delta I_C &= h_{21e}\Delta I_B + h_{22e}\Delta U_{CE}\end{aligned}\quad (3.18)$$

Współczynniki występujące w tych równaniach mają określony sens fizyczny. Współczynnik o indeksie 11 wiąże prąd z napięciem w obwodzie wejściowym i dlatego jest to rezystancja lub konduktancja wejściowa, a współczynnik indeksowany 22 przedstawia konduktancję wyjściową. Współczynnik z indeksem 12 wskazuje na oddziaływanie napięcia w obwodzie wyjściowym na obwód wejściowy, jest to oddziaływanie zwrotne w tranzystorze, często pomijalnie małe. Współczynnik z indeksem 21 opisuje podstawową cechę tranzystora, to jest sterowalność prądu wyjściowego, współczynnik y_{21s} nazywa się transkonduktancją tranzystora unipolarnego, a współczynnik h_{21e} wzmocnieniem prądowym tranzystora bipolarnego.

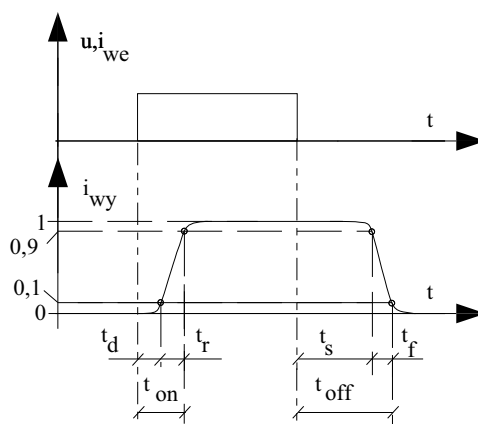
Zmiany " Δ " prądów czy napięć mogą być zrealizowane jako zmiana z jednej wartości stałej w czasie, na inną wartość stałą w czasie, albo też można do np. napięcia stałego dodać napięcie przemiennie, którego podwójną amplitudę można wtedy traktować jako wartość " Δ ". W takim przypadku, skutkiem istnienia pojemności w tranzystorach, pojawi się zależność parametrów czwórnikowych od częstotliwości wprowadzanego napięcia przemiennego. W zakresie małych częstotliwości ten wpływ jest zwykle pomijalnie mały,

parametry czwórnikowe są liczbami rzeczywistymi, niezależnymi od częstotliwości. Inaczej jest przy dużych częstotliwościach, kiedy wartości parametrów ulegają znacznym zmianom. Szczególnie istotne jest obserwowane zmniejszenie wartości parametru o indeksie 21, opisującego wzmocnienie tranzystora. Zmniejszenie wzmocnienia jest podstawą do zdefiniowania maksymalnej częstotliwości pracy tranzystora.

Równania (3.17) czy (3.18), nazywane modelami małosygnałowymi, mogą być modyfikowane, czy to w kierunku uściślenia opisu np. przez wprowadzenie zespolonych współczynników y_{ij} lub h_{ij} , czy też mogą być upraszczane przez odrzucanie składników o mniejszych wagach np. przez przyjęcie, że współczynniki o indeksie 12 mają wartość zero. Równaniom (3.17) i (3.18) odpowiadają schematy elektryczne, nazywane schematami zastępczymi tranzystorów. Na przykład skrajnie uproszczonym schematem zastępczym tranzystora unipolarnego jest źródło prądowe pokazane na rys. 3.23. Taki schemat uzyskuje się zakładając, że oprócz transkonduktancji y_{21} , wszystkie inne parametry y mają wartość zero.

Jak uprzednio wspomiano, tranzystory stosuje się albo w układach analogowych albo impulsowych. W układach analogowych wykorzystuje się możliwości płynnego zmieniania wartości prądu za pomocą też płynnie zmienianego sygnału sterującego. W tych przypadkach powyżej przedstawione modele są dogodne do wykonywania obliczeń projektowych.

W układach pracujących impulsowo, logicznych lub cyfrowych tranzystor powinien spełniać rolę "wyłącznika", częściej zwanego kluczem, pozwalającego otwierać lub zamykać obwód wyjściowy tak, aby prąd w tym obwodzie został przerwany (lub prawie przerwany) albo przeciwnie, aby prąd wyjściowy przyjął wartość maksymalną, określoną w przybliżeniu przez iloraz napięcia źródła występującego w obwodzie wyjściowym przez rezystancję obciążenia (por. rys. 3.10).



Rys. 3.25. Określenie czasów przełączania tranzystorów

Do analizy układów, w których tranzystor pracuje dwustanowo, bardziej dogodne są inne opisy i parametry tranzystorów. Przy maksymalnym prądzie napięcie między elektrodami wyjściowymi tranzystora silnie maleje, zwykle do wartości mniejszej od 1 wolta, jed-

nakże nie może przyjąć wartości zero. W takich warunkach spadek napięcia dren–źródło tranzystora unipolarnego jest zależny prawie liniowo od prądu i dlatego jednym z ważnych parametrów tych tranzystorów przy pracy dwustanowej jest rezystancja w stanie włączenia $R_{DS(on)}$. Inaczej w tranzystorze bipolarnym, spadek napięcia kolektor–emiter nieznacznie zależy od prądu, dlatego parametrem tych tranzystorów jest napięcie kolektor–emiter w stanie nasycenia $U_{CE(sat)}$. Dla tranzystorów w stanie nieprzewodzenia w danych katalogowych określa się wartość prądu zerowego (lub prądu upływu) definiując warunki jego pomiaru.

Proces przełączanie od stanu nieprzewodzenia do przewodzenia i odwrotnie jest opisywany wartościami czasowymi zmian prądu tranzystora przy skokowym sygnale sterującym, jak pokazano na rys. 3.25. Przy włączeniu tranzystora określa się czas opóźnienia (t_d) i czas narastania (t_r) prądu, których łączna wartość jest nazywana czasem załączenia tranzystora t_{on} , przy wyłączeniu prądu określa się czas magazynowania (t_s , o wartości zerowej w tranzystorze unipolarnym) i czas opadania (t_f) wyznaczające czas wyłączenia t_{off} .

3.5. Tyrystory

Tyrystory są elementami półprzewodnikowymi, których główną cechą jest skokowa zmienność stanu elektrycznego w tzw. obwodzie głównym. W tym obwodzie tyrystory albo mogą przewodzić stosunkowo duży prąd (dla największych tyrystorów do 5000 A) przy małym spadku napięcia rzędu 1–2 woltów, albo przy znacznych napięciach (dla największych tyrystorów do 6000 V) przepuszczają prąd o znikomo małej wartości. Jednoczesne wystąpienie stosunkowo dużych prądów przy większych napięciach między elektrodami głównymi jest niemożliwe (może to wystąpić tylko przejściowo, w trakcie szybkich procesów zmiany stanu tyrystora). Tyrystory są półprzewodnikowymi elementami elektronicznymi ale, jak wynika z przytoczonych danych, są zwykle stosowane przy dużych wartościach mocy elektrycznej i dlatego zalicza się je do elementów energoelektronicznych.

Dwustanowa praca tyrystorów upodabnia je do łączników elektrycznych, przełączanie tyrystora z jednego stanu do drugiego następuje pod wpływem elektrycznego sygnału sterującego doprowadzanego do elektrody sterującej zwanej bramką. Sygnał sterujący ma zwykle kształt krótkiego impulsu prostokątnego, ponieważ jego zadaniem jest tylko zainicjowanie lawinowych procesów przełączania tyrystora. Procesy przełączania przebiegają zwykle szybko, w czasie kilku do kilkudziesięciu mikrosekund, co pozwala realizować zamykanie i otwieranie obwodu głównego np. 50 lub więcej razy w czasie jednej sekundy, to znaczy z częstotliwością energetyczną 50 Hz, lub też z jeszcze większą częstotliwością, ale zwykle nie przekraczającą kilkunastu kiloherców.

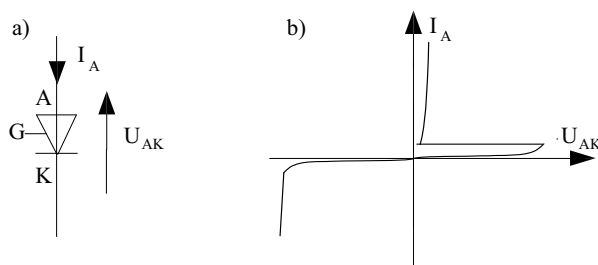
Najstarsza odmiana tyrystora (opracowana w roku 1956) nazywa się tyrystorem triodowym, ale często jest nazywana skrótowo "tyrystor", natomiast do innych odmian stosuje się w praktyce nazwy poszerzone o opis ich charakterystycznej cechy jak np. "tyrystor wyłączalny prądem bramki". Wewnętrzna budowa tyrystorów jest dość złożona, w najprostszych konstrukcjach są to cztery warstwy półprzewodnika, na przemian typu p oraz n, odpowiednio rozmieszczone przestrzennie.

3.5.1. Tyrystor triodowy (tyrystor)

Symbol schematowy tyrystora i charakterystykę obwodu głównego pokazano na rys. 3.26. Elektrodamy głównymi są anoda (A) i katoda (K), elektrodą sterującą jest bramka (G).

Przy polaryzacji wstecznej (anoda ujemna – katoda dodatnia) tyrystor zachowuje się podobnie jak dioda polaryzowana wstecznie, tj. w obwodzie głównym przewodzi bardzo mały prąd wsteczny, o wartości często pomijalnej w praktyce projektowej. W sytuacji przeciwnej, tj. przy dodatniej anodzie a ujemnej katodzie, tyrystor jest w jednym z dwu możliwych stanów:

- stan blokowania – podobnie jak przy polaryzacji wstecznej, tj. w obwodzie głównym mały prąd przy dużych napięciach,
- stan przewodzenia – w obwodzie głównym duży prąd, przy spadku napięcia około 1 V, podobnie jak w przewodzącej diodzie.



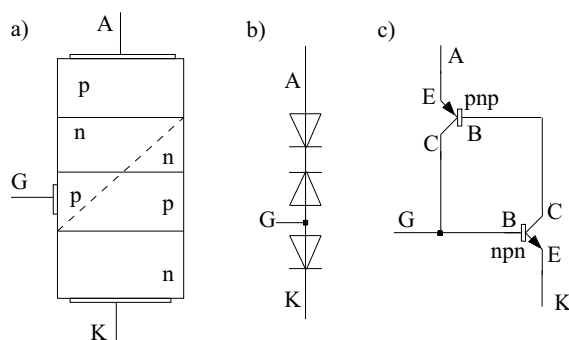
Rys. 3.26. Symbol schematowy (a) i charakterystyka główna (b) tyrystora

Tyrystor znajdujący się w stanie blokowania można przełączyć do stanu przewodzenia krótkim impulsem bramkowym w obwodzie bramka-katoda. Tyrystor pozostaje w stanie przewodzenia tak długo, jak długo będzie podtrzymywany prąd główny, wyłączenie tyrystora elektrodą sterującą jest niemożliwe. Aby nastąpiło wyłączenie tyrystora, tj. wyjście ze stanu przewodzenia, konieczne jest zmniejszenie jego prądu głównego, przez odpowiednią obniżkę napięcia źródła wymuszającego ten prąd. Często przyjmuje się, że dla wyłączenia tyrystora należy zmniejszyć jego prąd główny do zera, ale uściślenie brzmi: poniżej pewnej, małej wartości, zwanej prądem wyłączenia tyrystora.

Uproszczony model budowy tyrystora pokazano na rys. 3.27a. Traktując każde złącze pn jako diodę, można tyrystor przedstawić jako szeregowe połączenie trzech diod, jak pokazano na rys. 3.27b. Taki model elektryczny tłumaczy stany blokowania i polaryzacji wstecznej jako stany nieprzewodzenia odpowiednich diod.

Aby wyjaśnić proces załączania tyrystora należy odwołać się do modelu dwutranzystorowego, pokazanego na rys. 3.27c. Traktując umownie rozcięte dwie części tyrystora jako dwa tranzystory bipolarne, jeden typu npn a drugi typu pnp, można prześledzić następujący proces. Włączenie prądu bramki odpowiada włączeniu prądu bazy tranzystora npn, co powoduje przepływ prądu kolektora tego tranzystora. Prąd kolektora tranzystora npn jest jednocześnie prądem bazy tranzystora pnp, wysterowanie tranzystora npn prowadzi więc do wysterowania także tranzystora pnp. Prąd kolektora tego tranzystora, większy od

jego prądu bazy, płynie z kolei do bazy tranzystora npn, dodając się do prądu bramki, który zapoczątkował cały proces. Dzięki dwukrotnemu wzmocnieniu, prąd z układu zastępczych tranzystorów, dodający się do prądu bramki wytworzonego przez obwód zewnętrzny, jest od niego wielokrotnie większy. To powoduje, że w "drugim cyklu" wszystkie prądy tranzystora npn, a także tranzystora pnp, stają się znacznie większe niż na początku procesu załączania.



Rys. 3.27. Uproszczona budowa tyrystora (a), model diodowy (b) i model dwutranzystorowy (c)

Opisane zjawisko przebiega ciągle w sposób lawinowy, tj. ciągle następuje narastanie prądów obydwu tranzystorów, a więc i prądu tyrystora. Zakończenie procesu załączania ma miejsce dopiero wtedy, gdy inne, poza tyrystorem, elementy obwodu głównego (napiecie źródła i rezystancja) uniemożliwią dalszy wzrost prądu.

Czas trwania procesu załączania wynika z prędkości nośników elektronowych, wstrzykiwanych z obszaru n (katodowego) i nośników dziurowych, wstrzykiwanych z obszaru p (anodowego) do wnętrza struktury tyrystora oraz z grubości jego warstw.

Jak wynika z modelu dwutranzystorowego, w stanie przewodzenia tyrystora obydwie zastępcze (umowne) tranzystory wysterowują się prądowo nawzajem, wobec czego inicjujący prąd bramki staje się niepotrzebny. To zjawisko obserwuje się jako utratę sterowności po załączeniu tyrystora. Powrót sterowności następuje dopiero po wyłączeniu tyrystora i polega na możliwości powtórnego załączenia tego elementu.

Wyłączenie tyrystora, jak wspomniano wyżej, wymaga przerwania procesu wzajemnego wysterowywania się jego wewnętrznych struktur przez odpowiednie ograniczenie prądu obwodu głównego. Jeżeli nawet spowoduje się zmniejszenie prądu do zera, to jeszcze wewnątrz tyrystora pozostaje duża ilość uprzednio wstrzykniętych nośników ładunku elektrycznego. Pełne przywrócenie stanu, jaki był przed okresem przewodzenia, nastąpi dopiero po zaniku tych nośników. Proces likwidacji nośników odbywa się przez ich rekombinację i może być zintensyfikowany przez wyprowadzanie ich na zewnątrz. Niezależnie od mechanizmu ten proces trwa przez pewien czas, zwany czasem wyłączenia tyrystora lub czasem odzyskiwania właściwości blokowania.

Wartości czasów załączania i wyłączania konwencjonalnych tyrystorów triodowych wynoszą od kilkunastu do ponad stu mikrosekund, co pozwala na stosowanie ich przy niezbyt wysokich częstotliwościach pracy.

3.5.2. Odmiany tyrystorów

Podstawowy typ tyrystora dobrze funkcjonuje w ograniczonej klasie układów energoelektronicznych. Jednak w innych klasach objawiają się jego niedoskonałości, co pociągnęło za sobą intensywny rozwój podobnych elementów, stanowiących ulepszone modyfikacje podstawowej struktury tyrystora. Zwykle poprawienie jakiejś cechy tyrystora powoduje zmianę i innych parametrów, dlatego stosuje się liczne odmiany tyrystorów, odpowiednio do wymagań stawianych przez określone układy energoelektroniczne. Poniżej wymienia się tylko odmiany tyrystorów, ich bardziej szczegółowy opis stanowi domenę energoelektroniki.

Tyrystory szybkie – skrócenie czasu załączania i wyłączania przez modyfikacje grubości, domieszkowania i konfiguracji warstw półprzewodnika kosztem zmniejszenia dopuszczalnych wartości prądu i napięcia. Szczególne konstrukcje, praktycznie pozbawione możliwości zaworowych, są nazywane tyrystorami impulsowymi.

GATT – (Gate Assisted Turn off Thyristor) – skrócenie czasu wyłączania i ułatwienie tego procesu przez wspomaganie wyłączania sygnałem bramkowym. W niektórych układach wyłączanie tyrystora odbywa się za pomocą dodatkowych obwodów elektrycznych (komutacja wymuszona). W tych przypadkach można uzyskać zmniejszenie gabarytów i masy elementów w dodatkowych obwodach oraz energii tam zużywanej, jeżeli skraca się czas procesu wyłączania tyrystora. Zmniejszenie tego czasu (o kilkanaście procent) uzyskuje się wprowadzając, jednocześnie z początkiem procesu wyłączania w obwodzie anodowym, także ujemny (odwrotny niż przy załączaniu) sygnał bramkowy. Taki sposób może być stosowany w niektórych tyrystorach konwencjonalnych lub w tyrystorach o zmodyfikowanej konstrukcji, nazywanych wtedy tyrystorami z wyłączaniem wspomaganym bramką lub tyrystorami typu GATT.

ASCR, RCT – (Asymmetric Silicon Controlled Rectifier, Reverse Conducting Thyristor) – skrócenie czasu wyłączania kosztem wytrzymałości napięciowej w kierunku wstecznym. Dla zwiększenia szybkości przełączeń tyrystora modyfikuje się grubości jego warstw i profile domieszek w warstwach. Osiągane skrócenia czasu wyłączania o 30–40% w stosunku do struktur konwencjonalnych, ale zmniejsza się odporność tyrystora na działanie napięcia wstecznego. Napięcie przebicia wstecznego maleje do 30–50 V podczas gdy wytrzymałość napięciowa w stanie blokowania może wynosić do 2 kV. Takie tyrystory nazywane są asymetrycznymi – ASCR. Niektórzy wytwórcy produkują strukturę asymetryczną scaloną z dodatkową diodą, połączoną równolegle do tyrystora ale o przeciwnym, niż tyrystor, kierunku przewodzenia. Tak włączona dioda powoduje, że cała struktura przewodzi prąd przy napięciu wstecznym tyrystora ok. 1V. Element nazywa się tyrystorem wstecznie przewodzącym – RCT.

GTO – (Gate Turn Off) tyrystor wyłączalny prądem bramki. Specjalna konstrukcja, często asymetryczna, tj. bez wytrzymałości napięciowej w kierunku wstecznym, załączanie impulsem bramkowym jak tyrystora konwencjonalnego oraz wyłączanie też impulsem bramkowym, ale o ujemnej polaryzacji i o znacznie większej wartości prądu aniżeli przy załączaniu. Wymaga rozbudowanych układów sterowania i zabezpieczeń.

TRIAC – (TRIode Alternating Current switch) – tyrystor dwukierunkowy, może przewodzić prąd w obydwu kierunkach. Struktura pięciowarstwowa p–n–p–n–p ma dwie elektrody główne i elektrodę sterującą, czyli bramkę. Przy dowolnej polaryzacji obwodu głównego może być przełączona ze stanu nieprzewodzenia do stanu przewodzenia impulsem bramkowym. Wyłączenie jak w tyrystorze konwencjonalnym, tj. przez obniżenie prądu głównego do zera. Tyrystor dwukierunkowy zachowuje się podobnie jak dwa tyrystory zwykle, połączone równolegle, ale o przeciwnych kierunkach przewodzenia. Element skonstruowany do zastosowań przy prądzie przemiennym.

MCT – (MOS Controlled Thyristor) – obwód główny jest strukturą czterowarstwową, podobnie jak w innych tyrystorach, sterowanie za pomocą dwu tranzystorów unipolarnych MOS scalonych w strukturze tyrystora. Często element asymetryczny, o dużej szybkości przełączeń, z możliwością zarówno załączenia jak i wyłączenia bramkowego. Jest to grupa tyrystorów intensywnie rozwijana od drugiej połowy lat osiemdziesiątych, rokująca nadzieję na znaczne usprawnienie nowoczesnych układów energoelektroniki.

3.6. Elementy optoelektroniczne

Światło padające na półprzewodnik może być: odbite, pochłonięte lub przepuszczone (przezroczystość półprzewodnika). Tylko światło pochłonięte wywołuje efekty elektryczne w półprzewodniku, polegające na zwiększeniu koncentracji swobodnych nośników prądu. Pochłonięty w półprzewodniku foton przekazuje swoją energię elektronowi walencyjnemu, wyzwala go z wiązania, powoduje to powstanie swobodnego elektronu i swobodnej dziury.

Aby nastąpiło uwolnienie elektronu z wiązań międzycząsteczkowych energia kwantu promieniowania musi być większa od energii pasma zabronionego W_G .

$$h\nu = \frac{hc}{\lambda} > W_G \quad (3.19)$$

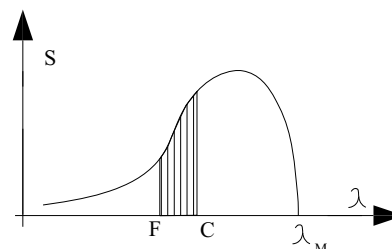
gdzie: h – stała Plancka,
 v – częstotliwość fali świetlnej,
 λ – długość fali świetlnej,
 c – prędkość światła.

Z tego warunku można wyznaczyć maksymalną długość fali, która jest w stanie spowodować generację pary nośników. Fale o większej długości są przepuszczone przez półprzewodnik, czyli dla strumienia fotonów o zbyt małych energiach półprzewodnik jest przezroczysty. Dla krzemu próg długofalowy mieści się w zakresie podczerwieni i dlatego dla światła widzialnego krzem jest nieprzezroczysty.

Pochłonięte fotony, o wystarczających energiach, wywołują dodatkową generację par nośników dziura-elektron, powodując wzrost koncentracji nośników obydwu typów ponad wartość równowagową, właściwą dla danej temperatury. Ilość par nośników generowanych fotoelektrycznie w jednostce objętości półprzewodnika i w jednostce czasu jest proporcjonalna do mocy strumienia świetlnego padającego na powierzchnię półprzewodnika, a odwrotnie proporcjonalna do energii fotonu $h\nu$.

Stan powierzchni półprzewodnika, kąt padania oraz energia fotonu decydują o tym, czy zderzenie fotonu z powierzchnią wywoła wnikięcie do wnętrza kryształu, czy odbicie od powierzchni. Dla fotonów o dużych energiach wzrasta prawdopodobieństwo odbicia od powierzchni półprzewodnika, a maleje prawdopodobieństwo pochłonięcia.

W analizie zjawisk fotoelektrycznych uwzględnia się jeszcze głębokość absorpcji, czyli skutki zróżnicowania głębokości, na której następuje absorpcja fotonu w półprzewodniku. Niektóre fotony przekazują swoją energię elektronom blisko powierzchni, przez którą wniknęły do półprzewodnika, inne przesuwają się wewnątrz kryształu nawet na głębokość kilku milimetrów. W związku z tym, że praktyczna grubość płytki półprzewodnika jest około jednego milimetra, to oprócz tego, że na różnych głębokościach powstają pary nośników, może także nastąpić przelot fotonu przez całą grubość płytki. Dla tego fotonu płytka jest przezroczysta. Przezroczystość półprzewodnika jest bardziej prawdopodobna dla fotonów o większych energiach.



Rys. 3.28. Charakterystyka widmowa czułości S krzemu

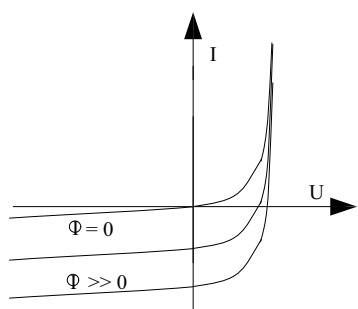
Powyżej wzmiankowane zjawiska kształtują tzw. charakterystykę widmową półprzewodnika, przykład pokazano na rys. 3.28, wskazująca na skuteczność oddziaływania światła o różnych długościach fali λ (i odpowiadających im barwach) na półprzewodnik. Dla krzemu maksimum wrażliwości na promieniowanie występuje przy podczerwieni, dla światła widzialnego (zaznaczonego na rysunku, jako pasmo od fioletu F do czerwieni C) krzem jest mniej wrażliwy.

Fotoelektryczny przyrost koncentracji nośników, ponad równowagową dla danej temperatury, dotyczy zarówno nośników większościowych, jak i mniejszościowych; ale względny (procentowy) przyrost koncentracji nośników większościowych zwykle jest nieznaczny, podczas gdy względny przyrost koncentracji nośników mniejszościowych może być bardzo duży, koncentracja tych nośników może zwiększyć się wielokrotnie, zależnie od oświetlenia półprzewodnika.

3.6.1. Fotodiody

Złącze pn umieszczone w obudowie umożliwiającej dostęp światła nazywa się fotodiody. Przy oświetleniu złącza następuje wzrost koncentracji nośników ponad równowagową dla danej temperatury, co zakłóca stan równowagi elektrycznej złącza. Względny przyrost koncentracji nośników większościowych jest nieznaczny, więc i względny wzrost prądów dyfuzyjnych jest nieznaczny. Względny wzrost koncentracji nośników mniejszościowych może być bardzo duży, co powoduje duży przyrost prądów unoszenia, tworzących prąd wsteczny złącza.

Z tego powodu, w wielu zastosowaniach technicznych, wykorzystuje się właściwości fotoelektryczne złącza przy polaryzacji wstecznej. Prąd wsteczny fotodiody nieoświetlonej, zwany prądem ciemnym, po oświetleniu zwiększa się o wartość wprost proporcjonalną do



Rys. 3.29. Charakterystyki fotodiody

mocy strumienia świetlnego. Przy intensywnym oświetleniu prąd wsteczny złącza może zwiększyć się wielokrotnie.

Na rys. 3.29 pokazano statyczne charakterystyki prądowo-napięciowe fotodiody, na których można zauważyć, że w stanie przewodzenia wpływ oświetlenia na przebieg charakterystyki jest stosunkowo niewielki, podczas gdy przy polaryzacji wstecznej charakterystyki dla różnych strumieni światła różnią się między sobą znacznie.

W stanach dynamicznych, przy zmianach oświetlenia występuje opóźnianie zmian prądu fotoelektrycznego względem zmian strumienia świetlnego. Spowodowane to jest skończoną szybkością przemieszczania się nośników przez obszar złącza, a także ich czasem życia, powodującym że np. po skokowym zaciemnieniu prąd zanika stopniowo, odpowiednio do stopniowego zmniejszania się koncentracji rekombinujących nośników.

Warto dostrzec, że charakterystyki oświetlonej fotodiody przebiegają także przez IV ćwiartkę układu współrzędnych, co świadczy o wystąpieniu także zakresu generatorowego.

Jak wspomniano powyżej oświetlenie złącza zakłóca zerowy bilans zjawisk dyfuzji i unoszenia powodując wzrost wartości prądów unoszenia ponad dyfuzyjne. Jeżeli elektrody oświetlonej fotodiody są zwarte, to obserwuje się wypadkowy prąd płynący w takim kierunku jak prąd wsteczny złącza, wartość tego prądu jest proporcjonalna do natężenia oświetlenia. Jeżeli oświetloną fotodiodę pozostawi się w stanie rozwarcia to następuje ładowanie obszarów p oraz n tak, aby nastąpiło zwiększenie prądów dyfuzyjnych do wartości kompensującej zwiększone prądy unoszenia. Temu odpowiada wytworzenie napięcia, pomiędzy obszarami p oraz n, o wartości nie większej od napięcia bariery potencjału. Stanom pośrednim pomiędzy zwarciem a rozwarciem odpowiadają napięcia i prądy pokazane na charakterystykach statycznych w IV ćwiartce układu współrzędnych.

W ten sposób oświetlona fotodioda może stać się źródłem energii elektrycznej, jeżeli dołączy się do niej jakiegokolwiek obciążenie. Dobór wartości optymalnego obciążenia jest dość trudny, wymaga to zazwyczaj odpowiednich układów regulacji. Napięcie pojedynczej fotodiody jest mniejsze od jednego wolta, prąd, zależnie od powierzchni złącza i oświetlenia, przyjmuje wartości tysięcznych lub setnych części ampera, dla otrzymania znaczących mocy trzeba więc łączyć wiele fotodiod szeregowo i równolegle w tzw. baterie, umożliwiające przetworzenie energii np. słonecznej w elektryczną.

3.6.2. Fototranzystor

Fototranzystor ma najczęściej strukturę bipolarną, oświetlany jest obszar bazy. Fotoelektryczna generacja par nośników wywołuje dwa skutki:

- 1) nośniki mniejszościowe, generowane w bazie, są unoszone do obszaru kolektora i stanowią część prądu kolektora;

2) nośniki większościowe, generowane w bazie, częściowo kompensują ładunki jonowe i obniżają w ten sposób barierę potencjału na złączu emiterowym.

To drugie zjawisko umożliwia wstrzykiwanie nośników z emitera do bazy, jak przy konwencjonalnym działaniu tranzystora bipolarnego. Strumień tych nośników stanowi znaczną część prądu kolektora.

Dzięki akcji tranzystorowej fotoelektryczny prąd kolektorowy jest znacznie większy, aniżeli prąd podobnej fotodiody. Inaczej określając, czułość fototranzystorów jest kilkadziesiąt razy większa aniżeli czułość fotodiod. Ze względu na pojemność obydwu złączy szybkość zmian fotoprądu w tranzystorach jest znacznie mniejsza aniżeli w fotodiodach. Parametry fototranzystorów określa się podobnie jak tranzystorów konwencjonalnych, podobne są też ich charakterystyki wyjściowe.

Fototranzystory wytwarzane są jako elementy dwuelektrodowe (emiter-kolektor) lub trójelektrodowe, z elektrodą bazy, co umożliwia kombinowane, optyczno-elektryczne, sterowanie prądem kolektora.

3.6.3. Diody świecące

Przepływ prądu przewodzenia przez złącze pn polega na przemieszczaniu się nośników większościowych do obszarów o przeciwnym typie przewodnictwa (zob. p. 2.1). Te nośniki, po przejściu przez złącze, ulegają rekombinacji, z czym związany jest zwrot energii, uprzednio pobranej przy uwalnianiu elektronów z wiązań. Zwrot energii następuje pod postacią ciepła i pod postacią promieniowania. Tak więc w każdej diodzie, przez którą przepływa prąd przewodzenia, następuje wydzielanie energii.

Proporcje pomiędzy ilością energii przemienianej na ciepło wewnątrz złącza pn a ilością energii wypromieniowanej na zewnątrz złącza są różne. Na przykład w krzemie większość energii odzyskiwanej w procesie rekombinacji (około 99%) powoduje podgrzanie diody, a tylko niewielka część jest wypromieniowana jako promienie podczerwone. W złączach ze związków galu (arsenek, fosforek) energia promieniowana stanowi dość dużą część (do kilkudziesięciu procent) energii odzyskiwanej w procesie rekombinacji.

Wyróżnia się dwa mechanizmy promieniowania: spontaniczny i wymuszony. Promieniowanie spontaniczne wykorzystuje się w diodach elektroluminescencyjnych, natomiast promieniowanie wymuszone w diodach laserowych.

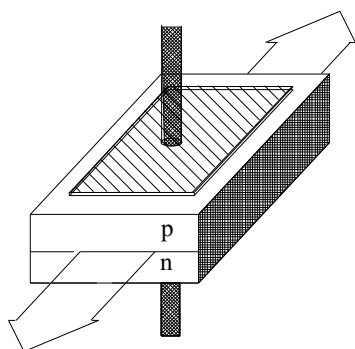
Dioda elektroluminescencyjna (LED – Light Emitting Diode), zasilana prądem przewodzenia, promieniuje światło o częstotliwości zależnej od szerokości pasma zabronionego $\nu = W_G/h$ (por. wzór (3.19)). Oprócz promieniowania o tej częstotliwości pojawiają się fale i o innych częstotliwościach, ponieważ rekombinujące nośniki mogą mieć energie z różnych poziomów pasma przewodnictwa i przeskakiwać na różne poziomy pasma walencyjnego. Przez dobór odpowiednich materiałów (np. kompozycja arsenku i fosforu galu) i technologii wykonania złącza pn można modyfikować promieniowany kolor od zielonego przez żółty, czerwony aż do podczerwieni.

Podstawowymi parametrami diod elektroluminescencyjnych są parametry optyczne, jak pasmo promieniowanych długości fal (z przedziału 550 do 950 nm) i moc promienio-

wana (0,2 do 2 mW) oraz parametry elektryczne, jak dopuszczalny prąd przewodzenia (10 do 300 mA) i spadek napięcia przy przewodzeniu (1,5 do 3,5 V).

Dioda laserowa (LASER – Light Amplification by Stimulated Emission of Radiation) działa na podstawie zjawiska promieniowania wymuszonego. Warunkiem promieniowania wymuszonego jest inwersja obsadzenia stanów czyli wywołanie takiej sytuacji, że w paśmie przewodnictwa jest więcej elektronów aniżeli pozostało ich w paśmie walencyjnym. Jeżeli w takiej sytuacji w półprzewodniku pojawi się foton o właściwej energii, to nastąpi jednoczesny powrót elektronów do pasma walencyjnego, czemu towarzyszy emisja promieniowania o dużym natężeniu i praktycznie jednej długości fali (promieniowanie monochromatyczne). Ponieważ promieniowanie ma miejsce dopiero po pobudzeniu półprzewodnika fotonem, to jest ono nazywane wymuszonym.

W złączu pn inwersję obsadzenia uzyskuje się stosując bardzo silne domieszkowanie oraz dużą wartość prądu przewodzenia (gęstość ponad 10 A/mm^2 , tj. 10–20 razy więcej aniżeli w innych diodach). Fotony wzbudzające promieniowanie wymuszone pochodzą z emisji spontanicznej, która w ograniczonym zakresie zawsze występuje.



Rys. 3.30. Dioda laserowa

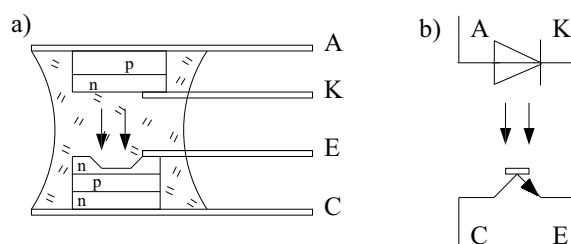
Dioda laserowa ma specyficzną konstrukcję, pokazaną na rys. 3.30. Dwie z bocznych powierzchni prostopadłościennej diody są gładkie, a dwie matowane. Fotony emitowane spontanicznie i trafiające w boczne powierzchnie matowane są usuwane na zewnątrz diody, a fotony padające na powierzchnie gładkie są odbijane między nimi. Te fotony stanowią pobudzenie promieniowania wymuszonego, które wydostaje się przez gładkie ścianki w postaci wiązki światła prawie monochromatycznego o małym kącie rozwartości.

Akcja laserowa w diodzie występuje jeżeli prąd przewodzenia przekroczy wartość progową, przy mniejszych prądach dioda emituje promieniowanie spontaniczne, o szerszym paśmie. Ponieważ

wartość progowa prądu akcji laserowej jest znaczna, to często dla ograniczenia nagrzewania stosuje się pracę impulsową, impulsom prądu odpowiadają impulsy świetlne, częstotliwość impulsowania może być dość znaczna, rzędu setek megaherców.

3.6.4. Transoptory

Transoptorem nazywa się parę złożoną z diody świecącej jako nadajnika światła i elementu czułego na światło jako odbiornika, zamkniętych w jednej obudowie. Nadajnikiem jest najczęściej dioda elektroluminescencyjna a odbiornikiem fotodiody, fototranzystor lub inny element fotoelektryczny. Elementy lokuje się tak, aby światło wydzielane przez diodę emitującą padało na odbiornik. Dla zapewnienia prawidłowego poprowadzenia strumienia świetlnego stosuje się światłowód, np. w postaci przezroczystego kleju epoksydowego, jak pokazano na rys. 3.31.



Rys. 3.31. Transoptor zawierający LED i fototranzystor, budowa (a), schemat (b)

Wejściem transoptora jest dioda emitująca, a wyjściem np. fototranzystor. Ilość światła wydzielanego przez diodę świecąca jest proporcjonalna do prądu wejściowego, a prąd fototranzystora jest proporcjonalny do ilości odbieranego światła. W ten sposób sterując prądem diody emitującej powoduje się zmianę prądu fototranzystora – można więc przekazywać np. impulsy prądu elektrycznego za pośrednictwem światła. W takiej sytuacji obwody elektryczne są odseparowane galwanicznie, tj. nie ma połączeń elektrycznych między nimi, ale jest zapewnione przekazywanie sygnałów prądowych.

Parametry elektryczne transoptora to parametry obwodu wejściowego (diody świecącej), parametry obwodu wyjściowego (fototranzystora) i parametry przejściowe, jak np. stosunek prądu wyjściowego do wejściowego, maksymalna częstotliwość impulsowania, wytrzymałość elektryczna izolacji pomiędzy obwodem wejściowym a wyjściowym i in.

Podobnie może być skonstruowane łącze optoelektroniczne, w którym odległość pomiędzy nadajnikiem światła o odbiornikiem jest duża – do dziesiątków kilometrów. W takim przypadku dla prowadzenia światła stosuje się cienkie włókna szklane, objęte ochronnym płaszczem i zwane kablem światłowodowym.

4. MIKROELEKTRONIKA, ENERGIELEKTRONIKA

W odległej przeszłości kultury i cywilizacji można odnaleźć początki dylematu pojęciowego: materia – idea. To rozróżnienie pojawiało się w filozofii i innych dziedzinach wiedzy, a także w życiu codziennym, jako przeciwstawienia: ciało – dusza, siła – rozum i temu podobnych. W epoce nowożytnej materia stała się przedmiotem zainteresowania nauk ścisłych jak fizyki, chemii, biologii i in. dzięki czemu szereg pojęć uzyskało dość ścisłe definicje, wykryto liczne zależności i prawa po einsteinowskie stwierdzenie równoważności masy i energii. Równoległe z opracowaniami naukowymi, szczególnie fizyków, liczne zagadnienia z tej dziedziny stały się też przedmiotem zainteresowania techniki dla praktycznego ich wykorzystania.

Inaczej toczyła się historia stosunku ludzi do zjawisk z gatunku "idea". Prawdopodobnie od zarania kultury zdawano sobie sprawę z tego, że istnieje coś takiego jak wiadomość lub wiedza, lub – jak dzisiaj to się określa – informacja, ale jako coś niematerialnego traktowano to zupełnie inaczej niżeli zjawiska materialne. Taki stan utrzymywał się bardzo długo i dopiero w ostatnich kilkudziesięciu latach zaczęto wprowadzać tu zmiany. Zmiany te zostały zapewne wywołane rozwojem technik przesyłu wiadomości, co pociągnęło za sobą konieczność podejścia technicznego do tej dziedziny. Brak wyczerpujących badań podstawowych powoduje, że problemy dotyczące informacji nie są tak uściśnione i jednoznacznie zdefiniowane jak problemy fizyczne. Jednak silny związek elektroniki z technicznym podejściem do informacji nakazuje zwrócenie uwagi na tę problematykę w tym skrypcie.

Techniczne podejście do informacji charakteryzuje się całkowitym pominięciem treści czy sensu informacji, natomiast zainteresowaniem się formą informacji i procesami, jakim może ona podlegać. Do procesów takich należą:

- przechowywanie informacji,
- przesył informacji,
- przetwarzanie informacji.

W każdym przypadku informacja, jako abstrakt, musi zostać związana z materialnym nośnikiem, który może być poddawany zabiegom technicznym. Istnieje bardzo wiele sposobów związania informacji z nośnikami, poczynając od kucia napisów na kamieniu do odpowiedniego namagnesowania drobnych fragmentów powierzchni komputerowego twardego dysku. Techniki przesyłu, przetwarzania i przechowywania informacji muszą być nie tylko dostosowane do typu nośnika, ale także uwzględniać niektóre "wrodzone" cechy samej informacji.

4.1. Sygnał elektryczny, energia elektryczna

W procesach przesyłu i przetwarzania bardzo dogodnie jest korzystać z napięcia i prądu elektrycznego jako nośnika informacji. W takim przypadku używa się nazwy "sygnał elektryczny" dla podkreślenia faktu, że nadrzędnym czynnikiem jest informacja, a prąd czy napięcie są tylko jej nośnikiem. Z doświadczenia wynika, że zwykle przesył czy przetwa-

rzanie sygnałów elektrycznych może odbywać się przy stosunkowo małych prądach i napięciach, czyli przy małym poziomie mocy elektrycznej, ale wymaga dość często skomplikowanych operacji technicznych na sygnale. Do takich celów zostały rozbudowane układy elektroniczne, niekiedy o dużym stopniu złożoności, ale pracujące przy małej mocy. Do tych układów można stosować ogólne pojęcie "mikroelektronika", obejmujące zarówno ich zastosowanie do obróbki sygnałów, jak i technologię pozwalającą na budowę skomplikowanych urządzeń o małych wymiarach.

Napięcie i prąd elektryczny są też powszechnie używanymi nośnikami energii. Obecnie uważa się, że energia elektryczna jest najbardziej uniwersalną i najczystsza formą energii, umożliwiającą łatwy przesył od miejsca uzyskania do miejsca użytkowania, a także łatwe przetwarzanie na inne, dogodne dla określonych celów technicznych postaci (mechaniczna, cieplna i in.). Jediną poważną wadą energii elektrycznej jest niemożność jej magazynowania. W procesach przesyłu i użytkowania energii elektrycznej często występuje konieczność przekształcenia jej parametrów technicznych (napięcia i prądu).

Jedno z takich przekształceń realizują konwencjonalne transformatory energetyczne, zmieniające wartość napięcia przy prądzie przemiennym. Inne przekształcenia, jak na przykład zmiana prądu przemiennego na stały czy zmiana częstotliwości prądu, wymagają odpowiednich urządzeń przekształcających. Obecnie takie przekształtniki budowane są na bazie elektronicznych elementów półprzewodnikowych. Przekształtniki energii elektrycznej projektowane są zazwyczaj tak, aby umożliwiały przetwarzanie znacznych mocy. Do tego rodzaju układów oraz elementów stosuje się nazwę "energoelektroniczne", dla podkreślenia ich przeznaczenia oraz zwykle, relatywnie dużej mocy.

Tak więc wielkości elektryczne (prąd, napięcie) poddawane przetworzeniom w układach elektronicznych, mogą mieć dwojaki charakter: sygnału elektrycznego niosącego informację lub nośnika energii elektrycznej. To rozróżnienie, w niektórych sytuacjach całkowicie oczywiste, w innych przypadkach może być rozmyte. Tak jak każdy sygnał ma pewne parametry energetyczne (np. moc), tak i każdą formę energii elektrycznej można powiązać z pewną informacją (np. obniżona wartość napięcia w gniazdku wtykowym niesie informację o przeciążeniu systemu energetycznego). Ważniejszym od prób ustalania granic ilościowych jest ustalenie priorytetów technicznych.

Jeżeli, obok uzyskania założonych parametrów technicznych, podstawowym zadaniem jest zachowanie niezniekształconej informacji, to mamy do czynienia z przetwarzaniem sygnałów, jeżeli natomiast nadrzędnym wymaganiem jest wysoka sprawność, to dotyczy to przekształcania energii. Tak więc różnice między mikroelektroniką a energoelektroniką sprowadzają się nie tylko do różnic w parametrach ilościowych, ale też różnic jakościowych, polegających na odmiennym ustaleniu priorytetowych zadań technicznych.

4.2. Sygnał analogowy a sygnał cyfrowy

Wielkości, szczególnie dotyczące procesów technicznych, można podzielić na ciągłe i nieciągłe. Wielkościami ciągłymi są takie, których możliwym wiernym modelem jest matematyczna funkcja ciągła; do tej grupy należą: prędkość, temperatura, napięcie itp. Wielko-

ściami nieciągłymi są takie, których możliwie wiernym modelem jest matematyczna funkcja nieciągła; do tej grupy należą takie wielkości jak stan ustawienia zaworów dwupołożeniowych (zamknięte – otwarte), stan nagrzewnic elektrycznych sterowanych stycznikami (wyłączone – załączona 1 sekcja – załączone 2 sekcje – ...), liczba wyprodukowanych sztuk określonego produktu itp. Sygnały, przynoszące informacje o wielkościach ciągłych lub nieciągłych, powinny być dostosowane do ich charakteru.

Prostą formę sygnałów otrzymuje się np. w metrologii elektrycznej wielkości nieelektrycznych, stosując przetworniki wytwarzające napięcie lub prąd proporcjonalne do innej wielkości fizycznej. W takim przypadku wielkość elektryczna jest odzwierciedleniem czyli analogiem innej wielkości, a sygnał nazywamy analogowym. Sygnał analogowy (napięciowy lub prądowy) jest ciągłą funkcją czasu: $u(t)$ lub $i(t)$.

Na przykład napięcie na końcówkach termoelementu (termopary) zmienia się tak, jak zmienia się temperatura spoiny termoelementu. Mierząc to napięcie, odpowiednio wskazanym woltomierzem, możemy odczytywać wprost temperaturę, czyli uzyskiwać informację o temperaturze obiektu z zainstalowanym termoelementem. Dla prawidłowego odczytu konieczne jest właściwe dobranie zakresu woltomierza, poprowadzenie prawidłowej instalacji ewentualnie dobór innych elementów obwodu pomiarowego. Warto zauważyć, że w tym przykładzie każdy uchyb pomiaru napięcia powoduje zniekształcenie informacji o temperaturze obiektu.

Prawidłowy dobór aparatury do obróbki sygnału analogowego wymaga znajomości cech tego sygnału, czyli znajomości funkcji $u(t)$ lub $i(t)$. Jednak zazwyczaj dokładne, matematyczne postacie tych funkcji nie są znane a priori, dlatego stosuje się różne sposoby opisu sygnału, umożliwiające wyciąganie wniosków projektowych i konstrukcyjnych. Ważną cechą sygnałów analogowych jest konieczność zachowania dużej dokładności w ich obróbce.

Sygnały odpowiadające wielkościom nieciągłym nazywane są dyskretnymi. Szczególnym, acz bardzo szeroko rozpowszechnionym przypadkiem sygnałów dyskretnych, są sygnały cyfrowe.

Na przykład sygnał elektryczny, przekazujący informację z taśmy produkcyjnej o liczbie wyprodukowanych konserw, zwykle ma postać impulsów wytwarzanych wg zasady: jedna sztuka – jeden impuls. Tak więc jest to dyskretny sygnał dwustanowy, zmieniający się w pewnych momentach (czasu) ze stanu "brak impulsu" na stan "impuls" lub odwrotnie. Bardzo ważną zaletą takiego sygnału są minimalne wymagania co do dokładności. Wartość napięcia odpowiadającego impulsowi musi tylko zdecydowanie różnić się od wartości odpowiadającej brakowi impulsu. Pozwala to na przyjęcie dwu przedziałów dopuszczalnych wartości sygnału, rozdzielonych przedziałem wartości niedozwolonych. Na przykład stanowi "impuls" może odpowiadać wartość napięcia od 5 do 15 woltów, a stanowi "brak impulsu" wartość od 0 do 2 woltów. Z tego wynika, że bezwzględna wartość sygnału dyskretnego nie jest ważna dla wierności przesyłanej informacji, co łagodzi wymagania odnośnie dokładności obróbki tych sygnałów.

Aby uzyskać informację o liczbie wyprodukowanych sztuk, np. od początku zmiany, trzeba zastosować licznik impulsów, którego stan będzie wzrastał o jeden pod wpływem każdego doprowadzonego impulsu. Stan licznika elektronicznego może być uzewnętrzniony dwojako: wyświetlenie odpowiednich cyfr w sposób czytelny dla obsługi oraz uformo-

wanie sygnałów elektrycznych, prezentujących zawartą w nim liczbę. Takie właśnie sygnały dyskretne, niosące informację o liczbie, przedstawionej w sposób umowny (zakodowany), nazywa się sygnałami cyfrowymi.

Sygnały cyfrowe zachowują informację, nawet przy niewielkiej dokładności ich parametrów napięciowych czy prądowych. To pozwala dość łatwo dokonywać dalszych operacji na takich sygnałach. W powyższym przykładzie, zliczania sztuk, można przewidywać np. także obliczanie ilości opakowań zbiorczych, obliczanie wydajności (liczba sztuk/godz.) i innych danych.

Zalety techniczne sygnałów cyfrowych wywołały ich coraz szersze zastosowania. Oczywistym przykładem jest obserwowany rozwój technik komputerowych. Te techniki wykroczyły poza progi specjalizowanych ośrodków obliczeniowych i wypierają skutecznie, w postaci układów mikroprocesorowych, inne rozwiązania uprzednio stosowane w automatyce, metrologii i podobnych dziedzinach. Wiele urządzeń pomiarowych, regulacyjnych, sterujących, które poprzednio przetwarzały informacje przekazywane sygnałami analogowymi, obecnie realizuje się korzystając z sygnałów cyfrowych. Pojawiają się także nowe dziedziny, np. robotyka, których rozwój jest niemożliwy bez wprowadzenia cyfrowych technik obróbki informacji.

Zastosowanie sygnałów cyfrowych rozszerza się także na przenoszenie informacji o wielkościach ciągłych, dla których bardziej "naturalną" formą są sygnały analogowe. Dla takich celów opracowano układy przetwarzania sygnałów analogowych na cyfrowe i odwrotnie. Tego rodzaju układy znane są pod nazwą przetworników A/C (analogowo-cyfrowych) lub A/D (Analog to Digital) i przetworników C/A lub D/A.

4.3. Zakłócenia, szumy, dryft

W rzeczywistych warunkach pracy urządzeń przesyłających czy przetwarzających sygnały elektryczne obserwuje się dodatkowe, zmienne w czasie, napięcia i prądy, których źródła nie są zaprojektowane przez konstruktorów. Te dodatkowe napięcia nazywa się zakłóceniami lub szumami lub dryftem, odpowiednio do tego, co jest ich rzeczywistą przyczyną i jaki mają one charakter.

Powodem wystąpienia zakłóceń są zmienne w czasie pola elektromagnetyczne, wytwarzane przez inne obiekty aniżeli ten, w którym są obserwowane. Znane zjawisko indukcji elektromagnetycznej działa także przy oddaleniu obwodów elektrycznych. Np. przepływ zmieniającego się prądu w przewodzie trakcyjnym, zasilającym elektrowóz pociągu, powoduje powstanie napięć we wszystkich przewodach elektrycznych, zarówno połączonych w układy jak i w niepołączonych, znajdujących się w pobliżu, lub w oddaleniu, od torowiska. Wartości tych napięć są stosunkowo niewielkie i maleją ze wzrostem odległości od obiektu, który je wywołuje.

W urządzeniach zaprojektowanych do przetwarzania sygnałów elektrycznych o małych mocach, nawet takie małe napięcia mogą być obserwowalne właśnie jako zakłócenia. Cechami zakłóceń są: wielość ich źródeł i przypadkowy (z punktu widzenia projektanta lub użytkownika urządzenia elektronicznego) sposób ich pojawiania się. Źródłami zakłóceń są

właściwie wszystkie urządzenia i instalacje elektryczne, szczególnie takie, w których występują szybkie zmiany dużych prądów (spawanie elektryczne, iskrzenie komutatorów maszyn elektrycznych itp.), a także wyładowania atmosferyczne. Chwile powstawania zakłóceń oraz ich przebieg w czasie jest praktycznie niemożliwy do wyznaczenia a priori, dlatego opisuje się te zjawiska przy użyciu statystyk lub wybranych wskaźników syntetycznych.

Oprócz zakłóceń wprowadzanych za pośrednictwem zjawiska indukcji elektromagnetycznej (sprężenia indukcyjne), występują też zakłócenia wprowadzane przez zmienne pola elektryczne na zasadzie indukcji elektrycznej (sprężenia pojemnościowe). Ogólnym sposobem przeciwdziałania zakłóceniom jest ich tłumienie w źródłach oraz zmniejszanie wrażliwości urządzeń, na które mogą one oddziaływać przez np. ekranowanie metalowymi obudowami. W związku z coraz większym rozpowszechnieniem urządzeń elektrycznych (potencjalne źródła zakłóceń) i elektronicznych (wrażliwych na zakłócenia) problem ich bezkonfliktowej współpracy, czyli kompatybilności elektromagnetycznej, nabiera coraz większej wagi.

Niezależnie od zakłóceń pochodzących z zewnątrz, w urządzeniach elektronicznych występują dodatkowe napięcia i prądy, powstające przypadkowo wewnątrz tych urządzeń. Takie napięcia czy prądy nazywa się szumami, jeżeli zmieniają się one szybko w czasie (doprowadzone do głośnika dają słyszalny szum) lub dryftem, jeżeli zmiany są powolne.

Powodem szumów są zjawiska występujące w rezystorach, diodach, tranzystorach i innych elementach elektronicznych, a związane z fizyczną nieciągłością prądu elektrycznego. Prąd elektryczny, przy bardziej szczegółowych analizach, trzeba traktować jako ruch strumienia pojedynczych elektronów, na dodatek poruszających się dość bezładnie. Oznacza to, że wartości chwilowe prądu ulegają przypadkowym zmianom, przypadkowość dotyczy zarówno wartości, jak i szybkości zwiększania się lub zmniejszania prądu. Wartość średnia tych zmian wynosi zero, ale wartość średnia kwadratu prądu, czyli kwadrat wartości skutecznej jest niezerowy. Stopień nieuporządkowania ruchu elektronów, a więc i stopień zmienności prądu, charakteryzowany wartością skuteczną składowej przypadkowej, są zależne od temperatury.

Prąd płynący nierównomiernie przez rezystor powoduje, podobnie nierównomierne, przypadkowo zmienne napięcie, zwane szumem cieplnym. Wartość skuteczną szumu cieplnego wyznacza się jako napięcie na rezystorze o rezystancji R , mierzone woltmierzem reagującym na wartość skuteczną napięć sinusoidalnych o częstotliwościach nie mniejszych niż f i nie większych niż $(f + \Delta f)$. Wartość skuteczną U_c tego napięcia określa wzór Nyquista:

$$U_c = \sqrt{4kTR\Delta f} \quad (4.1)$$

gdzie: k – stała Boltzmanna,

T – temperatura bezwzględna.

Na przykład wartość skuteczna szumu cieplnego rezystora $10 \text{ k}\Omega$, w temperaturze 300 K , mierzona w paśmie 10 kHz wynosi $1,3 \mu\text{V}$. Wartość napięcia szumu cieplnego nie zależy od częstotliwości, a jedynie od szerokości pasma pomiarowego Δf , szum niezależny od częstotliwości nazywa się szumem białym.

Podobne zjawisko w elementach półprzewodnikowych wywołane przypadkowością procesów generacji i rekombinacji, powodujących przypadkowo zmienną ilość nośników

prądu, nazywa się szumem śrutowym, szum śrutowy też jest biały. Ponadto w półprzewodnikach obserwuje się szum nazywany "1/f", powodowany prawdopodobnie przez efekty powierzchniowe. Wartość tego szumu maleje wraz ze wzrostem częstotliwości f , przy częstotliwościach rzędu pojedynczych herców lub dziesiątków herców szum 1/f może być większy od szumu śrutowego i cieplnego, przy wyższych częstotliwościach jest niezauważalny. Wartości napięcia szumów są zwykle rzędu mikrowoltów, a częstotliwości na jakich są obserwowane obejmują zakres od zera do setek gigaherców.

Wytwarzanie szumów w czwórnikach, do których dają się sprowadzić tranzystory, opisuje się współczynnikiem szumów F , definiowanym przez porównanie mocy szumów. Zakłada się, że do wejścia czwórnika zostaje doprowadzony szum, który na wyjściu czwórnika pojawi się wzmocniony, lub osłabiony, stosownie do właściwości tego czwórnika. Gdyby sam czwórnik był bezszumny, to na jego wyjściu byłby wyłącznie ten przeniesiony szum wejściowy. W rzeczywistości wartość mocy szumu wyjściowego jest większa o moc szumu powstającego w czwórniku. Stosunek rzeczywistej mocy szumu wyjściowego do, przeliczonej na wyjście, mocy szumu wejściowego jest właśnie współczynnikiem szumów czwórnika.

Zmiany temperatury elementów, zarówno półprzewodnikowych jak i rezystorów oraz kondensatorów, powodują zmiany ich parametrów elektrycznych, co z kolei wywołuje zmiany prądów i napięć w układach elektronicznych. Takie zmiany, też o charakterze przypadkowym dla obserwatora, nazywa się dryftem cieplnym. Zwykle pierwotną przyczyną są tu zmiany temperatury otoczenia, a obserwowane efekty prądowe i napięciowe następują stosunkowo powoli, w ciągu minut, godzin, dni lub miesięcy. Jeszcze wolniej przebiegają dryfty starzeniowe, powodowane zmianami chemicznymi w materiałach. Napięcia dryftu mogą przyjmować większe wartości, znacznie przekraczające poziom miliwoltów, szczególnie w trudnych warunkach przemysłowych.

Minimalizacja szumów i dryftów, polegająca na odpowiednim doborze konstrukcji i technologii elementów oraz układów, jest jednym z trudniejszych zadań konstruktorów.

Skutkiem występowania zakłóceń, szumów i dryftów jest to, że każdy sygnał elektryczny, poddawany obróbce w rzeczywistych warunkach, należy traktować jako sumę dwu składników. Jednym z nich jest "czysty" sygnał, przebieg elektryczny niosący zaplanowaną informację, a drugim "szum", przebieg elektryczny przypadkowy, wynikający z łącznego oddziaływania powyżej wyliczonych zjawisk. Obecność szumu może zmniejszyć informację przenoszoną przez sygnał. O wielkości takiego zagrożenia decyduje stosunek mocy sygnału do mocy szumu. Dla zapewnienia najmniejszych zmniejszeń informacji należy zwiększać sygnał a minimalizować szum. Taki sposób łatwo można stosować przy sygnałach dyskretnych (cyfrowych) a znacznie trudniej przy sygnałach analogowych, co jest jednym z powodów rozwoju technik cyfrowych.

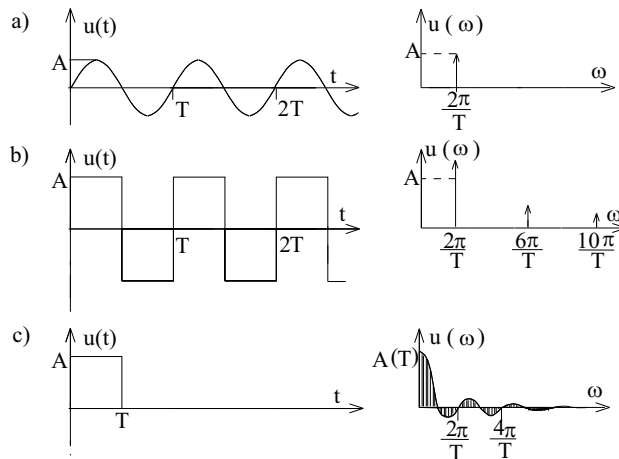
4.4. Przekształcanie sygnałów elektrycznych

Procesy przechowywania (pamiętania), przesyłania lub przetwarzania informacji z reguły wymagają dokonywania przekształceń sygnałów niosących te informacje. W tym punkcie przedstawia się pojęcia, opisujące wybrane operacje techniczne dokonywane na sygnałach.

4.4.1. Widmo sygnału, filtracja

Sygnał analogowy jest to napięcie lub prąd zmieniające się w czasie. Dobrze zaprojektowana aparatura przesyłająca czy przetwarzająca sygnał powinna być dostosowana do jego cech. Zazwyczaj funkcja $u(t)$ lub $i(t)$ jest trudna do opisanego a priori – jak np. matematycznie opisać, w postaci funkcji czasu, napięcie jakie będzie wytworzone w mikrofonie przy rejestracji koncertu? Z tego powodu stosuje się inne formy opisu, przenosząc zagadnienie z dziedziny czasu do dziedziny częstotliwości. Podstawą teoretyczną są przekształcenia Fouriera, według których funkcję czasu można przekształcić w funkcję innej zmiennej, mającej sens częstotliwości.

Na rys. 4.1 pokazano niektóre sygnały jako funkcję czasu i funkcje częstotliwości. Sygnałowi sinusoidalnemu o częstotliwości f odpowiada, w układzie współrzędnych (u, ω) , jeden prążek o wysokości równej amplitudzie sinusoidy, umieszczony przy wartości zmiennej ω równej częstości $2\pi f$.



Rys. 4.1. Funkcje czasu i odpowiadające im funkcje częstotliwości: sinusoida (a), ciąg impulsów prostokątnych (b), pojedynczy impuls prostokątny (c)

Sygnałowi w postaci nieskończonego, okresowego ciągu impulsów prostokątnych odpowiada nieskończony ciąg malejących prążków, umieszczonych przy określonych częstotliwościach. Równoważność sygnałów w dziedzinie czasu i częstotliwości w tym przypadku należy rozumieć tak, że sygnał $u(t)$, będący funkcją czasu, jest też nieskończoną sumą sygnałów, z których każdy jest sinusoidalnie zmienny w czasie. Amplitudy A_i , częstotliwości $i\omega$ oraz fazy ϕ_i tych sygnałów sinusoidalnych są odpowiednio, opisane we wzorach przedstawiających zasady budowy szeregu Fouriera.

$$u(t) = \sum_{i=1}^{\infty} A_i \sin(i\omega t + \phi_i) \quad (4.2)$$

Podobnie sygnałowi nieokresowemu w postaci pojedynczego impulsu, pokazanemu na rys. 4.1c), odpowiada sumowanie nieskończenie wielu sinusoid, o nieskończenie małych różniących się częstotliwościach i odpowiednich amplitudach i fazach, czyli całka Fouriera. Zależności amplitud i faz od częstotliwości nazywa się widmami, amplitudowym i fazowym, sygnału $u(t)$. Widma amplitudowe sygnałów okresowych są nieciągłe (prążkowe) jak pokazano na rys. 4.1a i b, widmo sygnału nieokresowego jest ciągłe – rys. 4.1c.

Często niektóre parametry widma sygnału łatwiej można oszacować aniżeli parametry tego sygnału jako funkcji czasu. Wiadomo np., że sinusoidalne zmiany ciśnienia powietrza, aby były słyszalne przez człowieka jako tony, powinny mieć częstotliwości zawarte w przedziale akustycznym (w fizyce przyjmuje się 20 Hz do 20 kHz, w elektroakustyce zwykle ten przedział jest mniejszy, np. w telefonii 300 Hz do 3 kHz). Można zatem przyjąć założenie projektowe, że aparatura dobrze działająca dla elektrycznych sygnałów sinusoidalnych, o częstotliwościach z pasma akustycznego, powinna dobrze przenosić sygnał $u(t)$ odwzorowujący koncert.

Podobnie jak widma sygnałów można określić widma zakłóceń i szumów. Jeżeli widmo sygnału występuje w innym zakresie częstotliwości aniżeli widmo zakłóceń i szumów, to pojawia się możliwość oddzielenia sygnału od zakłóceń i szumów metodą filtracji.

Filtrem elektrycznym nazywa się czwórnik, który przepuszcza sygnały sinusoidalne o pewnej częstotliwości, lub z pewnego przedziału częstotliwości, oraz tłumi (nie przepuszcza) sygnały sinusoidalne o innych częstotliwościach. Wprowadzając na wejście filtru liniowego sygnał sinusoidalny, zapisany w rachunku symbolicznym jako:

$$u_{we} = U_{we\max} e^{j\omega t} \quad (4.3)$$

otrzymujemy na wyjściu filtru sygnał o takiej samej częstotliwości, który może mieć inną amplitudę oraz być przesunięty fazowo względem sygnału wejściowego:

$$u_{wy} = U_{wy\max} e^{j(\omega t + \varphi)} \quad (4.4)$$

Stosunek sygnału wyjściowego do wejściowego jest to transmitancja filtru, liczba zespolona, której moduł i argument są funkcjami częstotliwości:

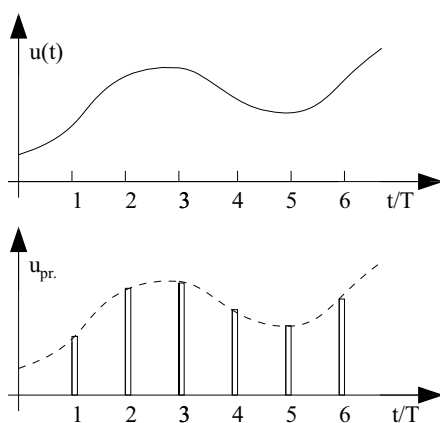
$$K(j\omega) = |K(\omega)| e^{j\varphi(\omega)} \quad (4.5)$$

Zależność modułu transmitancji filtru od częstotliwości nazywa się amplitudową charakterystyką częstotliwościową, a zależność argumentu od częstotliwości nazywa się fazową charakterystyką częstotliwościową.

Stosując filtrację sygnałów analogowych, dobiera się taki filtr, który przenosi dobrze widmo sygnału, tj. w paśmie sygnału transmitancja filtru K jest duża, natomiast częstotliwości spoza widma sygnału są tłumione; poza widmem sygnału wartość K powinna być mała. Dzięki temu procesowi zostaje stłumiona część zakłóceń i szumów, co poprawia stosunek sygnał/szum i sprzyja zachowaniu wierności przenoszonej informacji. Niekiedy dopuszcza się nawet utratę części widma sygnału, jeżeli w tej części występuje zwiększony poziom widma zakłóceń.

4.4.2. Próbkowanie, kwantowanie

Próbkowanie polega na zastąpieniu oryginalnego sygnału analogowego ciągiem wartości, ustalanych zwykle w regularnych odstępach czasu. Na rys. 4.2 pokazano przykład przebiegu sygnału $u(t)$ oraz próbek tego sygnału pobierane w odstępach czasu T .



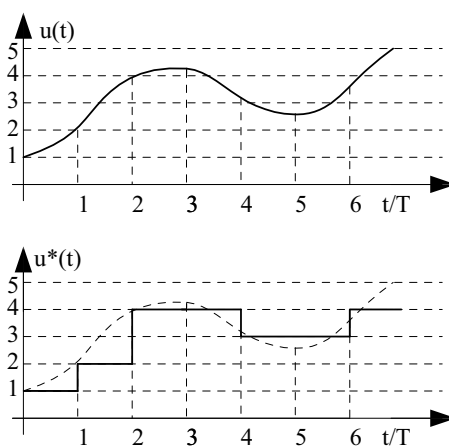
Rys. 4.2. Przebieg oryginalny $u(t)$ i ciąg próbek u_{pr}

Próbkowanie stosuje się w różnych sytuacjach. Np. przy występowaniu okresowych zakłóceń można skorzystać z próbkowania tak, aby moment pobierania próbki "minął się" z momentem wystąpienia zakłóceń. Próbkowanie jest też stosowane przy jednoczesnej (ściślej: cyklicznie powtarzalnej) obróbce wielu podobnych sygnałów pochodzących z różnych źródeł. Np. rejestracja danych z kilkunastu czujników pomiarowych, przy użyciu jednej aparatury rejestrującej, może być realizowana tak, aby rejestrator przyłączać kolejno do czujników, zapisując wartość sygnału z danego czujnika w chwili połączenia tego czujnika z rejestratorem. W ten sposób z każdego czujnika zapisywany jest sygnał w postaci ciągu wartości występujących w momentach połączeń, a wartości sygnału w pozostałych chwilach są pomijane, czyli sygnał podlega próbkowaniu.

Jest oczywiste, że zbyt duże przerwy pomiędzy kolejnymi pobieraniem próbek sygnału zagrażają utratą części informacji przekazywanej tym sygnałem. Dlatego należy stosować dostatecznie częste pobieranie próbek, czyli dostatecznie dużą częstotliwość próbkowania. Według twierdzenia Shannona o próbkowaniu, jednoznaczność ciągu próbek i sygnału oryginalnego występuje, jeżeli częstotliwość pobierania próbek jest nie mniejsza aniżeli podwójna wartość najwyższej częstotliwości widma tego sygnału. Ponieważ ściśle oznaczenie najwyższej częstotliwości widma sygnału nie zawsze jest łatwe, a często jest oparte na przybliżonych oszacowaniach, to w praktyce stosuje się raczej zawyżone częstotliwości próbkowania w stosunku do teoretycznych.

Kwantowanie sygnału analogowego polega na określeniu z góry wartości, jakie może on przybierać i bieżącym przyporządkowywaniu aktualnej wartości sygnału do odpowiedniej wartości z uprzednio wybranego zbioru.

Odpowiednikiem kwantowania jest, przyjęta w metrologii, zasada podawania wyników pomiaru z określoną dokładnością. Jeżeli np. zakres pomiarowy suwmiarki wynosi 100 mm, a jej dokładność 0,1 mm, to każdy wynik pomiaru jest wybrany jako jeden z 1001 możliwych. Najmniejszy możliwy wynik wynosi 0, kolejne są 0,1; 0,2; 0,3 aż po największe 99,8; 99,9; 100. Wynik pomiaru (nie późniejszych obliczeń uśredniających) nie może wynosić np. 2,83. Powodem kwantowania jest, m.in. świadomość ograniczonej dokładności sygnałów analogowych, a także jest to zabieg niezbędny przy zamianie sygnałów analogowych na cyfrowe.



Rys. 4.3. Przebieg analogowy oryginalny $u(t)$ i zdyskretyzowany $u^*(t)$

Efektom kwantowania jest przekształcenie ciągłego sygnału analogowego w sygnał "schodkowy". Kwantowanie jest częściej przeprowadzane na próbkach sygnałów analogowych, aniżeli na sygnałach oryginalnych.

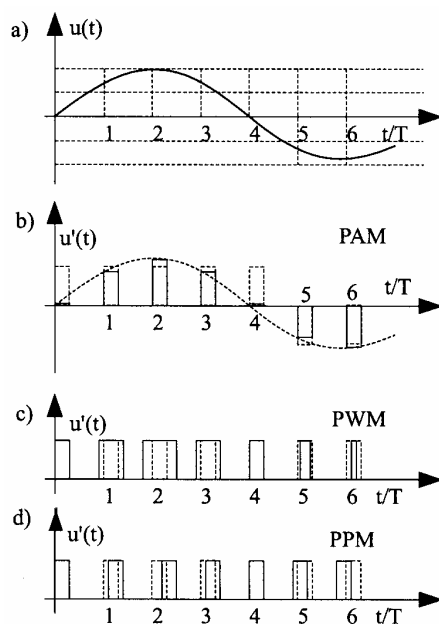
Na rys. 4.3 pokazano przykład sygnału analogowego poddanego próbkowaniu i kwantowaniu. Realizacja obydwu tych procesów nazywa się dyskretyzacją sygnału analogowego.

4.4.3. Modulacja, demodulacja

Oryginalny sygnał elektryczny jest często niedogodny technicznie, szczególnie przy przesyłaniu (transmisji) sygnałów. W takich przypadkach wytwarza się sygnał pomocniczy, zwany nośnym, o parametrach odpowiednich dla przesyłu i sygnał oryginalny, który ma być transmitowany, wiąże się z sygnałem nośnym. Taki proces nazywa się **modulacją**, a układ realizujący to zadanie nazywa się modulatorem.

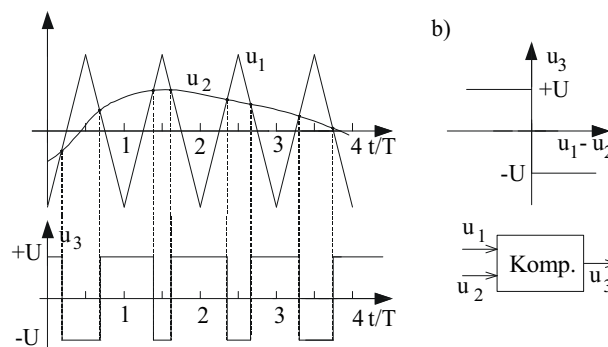
Proces odwrotny, oddzielanie sygnału oryginalnego z sygnału zmodulowanego nazywa się **demodulacją**, którą realizuje demodulator częściej zwany detektorem. Modulację stosowano od dawna, przy transmisji sygnałów akustycznych w radiofonii, używając jako fali nośnej sygnału sinusoidalnego. Ponieważ sygnał sinusoidalny ma postać: $u(t) = A \sin(\omega t + \varphi)$, to parametrami, które można wykorzystać do związania z sygnałem modulującym (zapisywanym) są: amplituda A , częstotliwość ω lub faza φ . Odpowiednio można stosować modulację amplitudy (AM – Amplitude Modulation), częstotliwości (FM – Frequency Modulation) lub fazy (PM – Phase Modulation).

W technice pomiarowej i sterowania częściej korzysta się z ciągu impulsów prostokątnych jako fali nośnej. Opis parametrów impulsu jest bardziej rozbudowany, co pociąga za sobą więcej możliwości modulacyjnych. Podstawowymi parametrami ciągu impulsów prostokątnych, o stałym okresie powtarzania, są: amplituda, szerokość i położenie. Każdy z tych parametrów może być modulowany, co prowadzi do trzech odmian modulacji: modulacja amplitudy impulsów (PAM – Pulse Amplitude Modulation), modulacja szerokości impulsów (PWM – Pulse Width Modulation), modulacja położenia impulsów (PPM – Pulse Position Modulation). Nadto stosowana jest modulacja częstotliwości impulsów (PFM – Pulse Frequency Modulation), często łącznie z inną odmianą modulacji. Na rys. 4.4 pokazano przykłady modulacji ciągu impulsów prostokątnych.



Rys. 4.4. Sygnał oryginalny (a) i przebiegi zmodulowanych impulsów prostokątnych przy modulacji: amplitudy (b), szerokości (c), położenia (d) impulsów

Przy modulacji impulsów prostokątnych, inaczej aniżeli przy modulacji sinusoid, przebieg nośny, wprowadzany do modulatora nie musi być przebiegiem prostokątnym. Na rys. 4.5 pokazano zasadę modulacji szerokości impulsów z zastosowaniem komparatora, czyli układu o dwu wejściach i jednym wyjściu. Napięcie wyjściowe komparatora przyjmuje tylko dwie wartości, np. $+U$ albo $-U$, zależnie od znaku różnicy dwu sygnałów wejściowych. Wprowadzając do jednego wejścia komparatora falę nośną w postaci sygnału trójkątnego, o stałej amplitudzie i okresie, a do drugiego wejścia sygnał modulujący, o dowolnym przebiegu (z ograniczeniami, których omawianie przekracza ramy tego skryptu), otrzymuje się na wyjściu sygnał prostokątny, o zmodulowanych szerokościach impulsów.



Rys. 4.5. Zasada modulacji szerokości impulsów przy zastosowaniu komparatora (a), charakterystyka komparatora (b)

Widmo zmodulowanych sygnałów impulsowych teoretycznie sięga od podstawowej częstotliwości powtarzania impulsów do nieskończoności. W praktyce przyjmuje się, że wystarczające jest uwzględnienie pierwszych kilku do kilkunastu harmonicznych (co i tak zazwyczaj narzuca konieczność stosowania urządzeń szerokopasmowych).

4.4.4. Kodowanie i transmisja sygnałów cyfrowych

Sygnał cyfrowy przenosi informację w postaci liczby. Do tego celu możliwe byłoby użycie sygnału skwantowanego, np. napięcie o wartości 1 wolta odpowiadałoby liczbie 1, a napięcie o wartości 13 woltów liczbie 13. W takim przypadku pojawiłby się problem dokładności wytwarzania, przekazywania i pomiaru napięć. Z tego powodu powszechnie stosuje się sygnał tylko dwuwartościowy, o poziomym napięcia albo niskim (L – Low) albo wysokim (H – High).

Pojedynczy sygnał dwuwartościowy może przekazać tylko dwie liczby, np. 0 lub 1. Aby przedstawić więcej niż 2 liczby trzeba użyć albo większej ilości sygnałów dwuwartościowych, albo dokonać odpowiedniej interpretacji pojedynczego sygnału. To prowadzi do dwu sposobów przesyłania i obróbki sygnałów cyfrowych: równoległe lub szeregowe. W każdym przypadku pojawia się konieczność wyrażenia określonego zbioru liczb przy

zastosowaniu tylko dwu wartości L oraz H. W tablicy 4.1 przedstawiono możliwości zakodowania liczb całkowitych od 0 do 15 za pomocą czterech sygnałów dwuwartościowych, oznaczonych a, b, c, d, a także przykłady innych kodów.

Tablica 4.1

Kody binarne

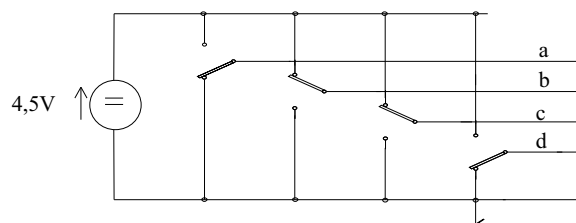
d	c	b	a	Liczba dziesiętna	Kod BCD	Kod HEX
L	L	L	L	0	0	0
L	L	L	H	1	1	1
L	L	H	L	2	2	2
L	L	H	H	3	3	3
L	H	L	L	4	4	4
L	H	L	H	5	5	5
L	H	H	L	6	6	6
L	H	H	H	7	7	7
H	L	L	L	8	8	8
H	L	L	H	9	9	9
H	L	H	L	10	–	A
H	L	H	H	11	–	B
H	H	L	L	12	–	C
H	H	L	H	13	–	D
H	H	H	L	14	–	E
H	H	H	H	15	–	F

W tablicy podano tzw. kod BCD (Binary Coded Decimal) oraz kod HEX (hexadecimal). Kod BCD jest używany do wyrażania liczb od 0 do 9, które są stosowane jako cyfry w liczbach dziesiętnych. Część możliwych kombinacji sygnałów abcd nie jest wykorzystywana w kodzie BCD. Kod hexadecymalny odpowiada szesnastkowemu systemowi liczenia i jest on tak stworzony, aby każdej kombinacji czterech sygnałów abcd odpowiadała jedna cyfra, wobec zbyt małej ilości cyfr w dziesiętnym systemie liczenia wykorzystano jako cyfry litery od A do F. Liczba w tym systemie ma postać np. 3D ($3 \times 16 + 13 = 61$) lub 7F5 ($7 \times 16 \times 16 + 15 \times 16 + 5 = 2037$) lub 24 ($2 \times 16 + 4 = 36$).

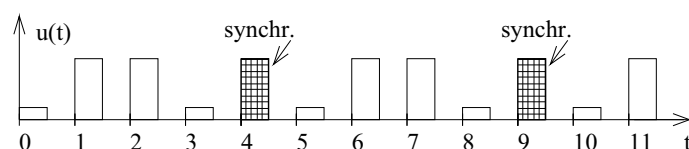
Jeżeli sygnały dwuwartościowe a, b, c, d występują jednocześnie w czterech różnych obwodach elektrycznych, to określa się to jako równoległe przekazywanie sygnałów cyfrowych. Na rys. 4.6 pokazano szkic schematu wytwarzającego cztery sygnały o wartościach 0 V lub 4,5 V, schemat narysowano przy takim ustawieniu przełączników, aby wytworzony sygnał cyfrowy (złożony z czterech sygnałów elementarnych a, b, c, d i dlatego nazywany czterobitowym) odpowiadał liczbie dziesiętnej 6. Równoległa transmisja sygnałów cyfrowych jest stosowana w komputerach; zespół przewodów nazywa się szyną lub magistralą (ang. bus), uzupełniając określenie o ilość przesyłanych sygnałów elementarnych, np. szyna ośmiobitowa lub szesnastobitowa.

Liczbę, np. 6, można także przekazać pojedynczym sygnałem dwuwartościowym, jeżeli ustali się następujące warunki: czas będzie traktowany dyskretnie i sygnał będzie przekazywany w określonych momentach, wyznaczonych przez zegar (np. na początku każdej

sekundy zegarowej); w sekundach o numerach 0; 5; 10; 15... będzie przekazywany pierwszy bit (poziom odpowiadający sygnałowi "a" przy transmisji równoległej); w sekundach o numerach 1; 6; 11; 16... będzie przekazywany drugi bit (odpowiednik sygnału "b") itd. Na rys. 4.7 pokazano sygnał szeregowy przesyłu liczby 6.



Rys. 4.6. Przesył równoległy sygnału cyfrowego



Rys. 4.7. Przesył szeregowy sygnału cyfrowego

Warunki techniczne realnego przesyłu szeregowego są zwykle bardziej złożone aniżeli podany wyżej zarys tej koncepcji. Transmisja szeregowo dyskretnych sygnałów dwuwartościowych stosowana jest od dawna, np. przy "wybieraniu" numeru tarczą telefonu, a obecnie wykorzystywana przy współpracy komputerów w sieciach, o różnym zasięgu, od sieci lokalnych po międzykontynentalne. Dla przetwarzania zbioru sygnałów występujących równoległe, w jeden sygnał niosący informację szeregowo lub odwrotnie, stosuje się urządzenia zwane, odpowiednio, interfejsami równoległo-szeregowymi i szeregowo-równoległymi.

4.5. Przekształcanie energii elektrycznej

Największa część energii elektrycznej jest wytwarzana w generatorach synchronicznych w postaci napięcia przemiennego, trójfazowego. To napięcie ma kształt bardzo zbliżony do przebiegu sinusoidalnego i jest zwykle traktowane jako realizacja funkcji sinus, o ustalonym okresie 20 ms (częstotliwość 50 Hz) i określonej amplitudzie. Tradycyjne przekształcanie polega na zmianie (transformacji) amplitudy napięcia. Przy przesyłaniu energii stosuje się wysokie napięcie, dla zmniejszenia strat przesyłowych, natomiast przy jej użytkowaniu dogodne są niskie wartości napięcia.

Różne procesy techniczne wywołały potrzebę stosowania jeszcze innych postaci energii elektrycznej. Np. przy elektrolizie konieczne jest stosowanie prądu stałego, w trakcji elektrycznej (tramwaj, elektrowóz pociągu) ze względów ruchowych stosuje się szeregowo

silniki prądu stałego itd. Dlatego jednym z rozpowszechnionych typów przekształcenia jest prostowanie prądu przemiennego, czyli zamiana prądu przemiennego (o zerowej wartości średniej) na prąd jednokierunkowy, zawierający składową stałą.

Niektóre obiekty wymagają całkowitej pewności zasilania elektrycznego, należą tu np. szpitalne sale operacyjne lub intensywnej terapii, systemy komputerowe lotnicze, bankowe i wojskowe, systemy zabezpieczeń energetyki jądrowej i in. W takich przypadkach tworzy się dodatkowe systemy zasilania awaryjnego, które powinno włączać się samoczynnie przy uszkodzeniu w normalnym systemie zasilania. Często stawie się warunek, aby przełączenie na zasilanie awaryjne nastąpiło bardzo szybko (np. w czasie 0,01 s), tak aby cała zasilana aparatura nie przerwała swojego normalnego działania (UPS – Uninterruptible Power System). Awaryjnym źródłem energii elektrycznej, utrzymywanym w stanie gotowości, są zwykle akumulatory, dostarczające napięcia stałego, podczas gdy nowoczesna aparatura jest dostosowana do zasilania prądem przemiennym. Z tego powodu konieczne są przekształtniki zamieniające prąd stały w przemienny, sinusoidalny, o częstotliwości 50 Hz. Takie elektroniczne przekształtniki nazywa się falownikami niezależnymi lub autonomicznymi, akcentując w ten sposób ich zdolność do działania bez zasilania z sieci elektrycznej.

W niektórych przypadkach wymaga się płynnej regulacji napięcia przemiennego jak np. przy ściemnianiu oświetlenia teatralnego, dokładnej regulacji podgrzewania elektrycznego w niektórych urządzeniach gospodarstwa domowego itp. Istotnym warunkiem jest często wymóg minimalnych gabarytów oraz kosztu urządzenia regulacyjnego. W podobnych przypadkach stosuje się przekształtniki zwane sterownikami napięcia lub prądu przemiennego. Podobnie, warunek płynnej regulacji wartości, może dotyczyć napięcia stałego. W takim przypadku stosuje się sterowniki napięcia lub prądu stałego, np. dla łagodnego rozruchu silnika prądu stałego, czy stabilizacji napięcia pobieranego z niestabilnej baterii lub akumulatora.

Przytoczone przykłady wskazują na cztery podstawowe typy przekształtników: prostownik, falownik, sterownik napięcia przemiennego, sterownik napięcia stałego. Oczywiście w każdym typie istnieje wiele odmian i modyfikacji niezbędnych do spełnienia różnorodnych wymagań stawianych przekształtnikom.

Często warunki techniczne wymuszają konieczność stosowania dwu przekształtników, działających kaskadowo. Np. napięcie przemiennie o częstotliwości 50 Hz zostaje przetworzone w napięcie stałe, a z kolei to napięcie stałe podlega przetworzeniu w napięcie przemiennie o częstotliwości innej aniżeli 50 Hz.

Przekształtnik energii, lub przekształtniki, są instalowane pomiędzy podstawowym źródłem energii elektrycznej (prądnica, sieć energetyczna, bateria) a odbiornikiem tej energii. Dla uzyskania dużej sprawności całego układu konieczna jest minimalizacja strat energii w przekształtniku. Z tego powodu w przekształtniku powinny być stosowane elementy możliwie mało stratne.

Wśród typowych elementów elektrycznych kondensatory i dławiki (łącznie z transformatorami) zalicza się do elementów konserwatywnych, tj. zdolnych do pobierania, przechowania i zwrotu energii elektrycznej (taką cechą mają przynajmniej idealizowane modele tych elementów). Rezystory natomiast są elementami dyssypatywnymi, tj. pobierają energię elektryczną i zamieniają ją w ciepło, bez możliwości odwrócenia tego zjawiska.

Dlatego w konstrukcji przekształtników unik się wprowadzania rezystorów oraz elementów zachowujących się podobnie jak rezystory, natomiast mogą być stosowane elementy konserwatywne. Zbudowanie przekształtnika z samych dławików i kondensatorów jest niemożliwe, muszą tam być stosowane i inne elementy. Elementem bezstratnym jest np. idealny łącznik, w stanie otwarcia utrzymujący napięcie, ale bez przepływu prądu, w stanie zamknięcia umożliwiający przepływ prądu, ale bez spadku napięcia. W obydwu przypadkach strata mocy i energii w łączniku ($p = u i$) jest zerowa. Dlatego w przekształtnikach stosuje się elementy zachowujące się podobnie do idealnego łącznika, czyli działające dwustanowo: dobre zamknięcie obwodu z możliwością przewodzenia prądu przy minimalnym (najlepiej zerowym) spadku napięcia lub dobre rozwarcie obwodu przy minimalnym (najlepiej zerowym) prądzie. Podobne właściwości mają diody prostownicze i tyrystory, a także można uzyskać taki sposób działania w tranzystorach, przy odpowiednim ich sterowaniu. Te właśnie grupy elementów elektronicznych są obecnie stosowane w przekształtnikach energoelektronicznych.

Dla uzyskania kontroli nad przepływem energii ze źródła do odbiornika, przy korzystaniu z elementów działających podobnie do łączników, trzeba stosować metodę "dawkowania". Metoda polega na tym, aby łączyć źródło z odbiornikiem na określony, krótki czas, potem rozłączać, też na określony czas, znowu łączyć i rozłączać i powtarzać to cyklicznie. Na przykład prostownik można traktować jako urządzenie łączące źródło napięcia przemiennego, sinusoidalnego, z odbiornikiem energii tylko wtedy, gdy chwilowa wartość napięcia źródła jest dodatnia, a odłączające odbiornik wtedy, gdy napięcie źródła jest ujemne; takie czynności są powtarzane 50 razy w ciągu sekundy.

Sposób pracy przekształtników energoelektronicznych można więc przedstawiać jako realizację jednej z metod modulacji impulsowych, często jest to PWM. Można też doszukiwać się analogii pomiędzy pracą przekształtnika a pracą silnika spalinowego, gdzie energia mechaniczna jest dostarczana do wału też metodą "dawkowania" z kolejnych cylindrów.

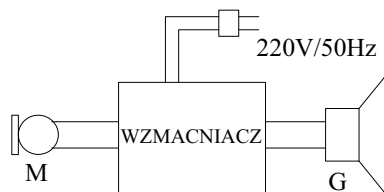
Dokładniej zagadnienia przekształcania energii elektrycznej są rozważane w energoelektronice.

5. WZMACNIACZE

Nazwą wzmacniacz określa się jeden z najważniejszych podzespołów elektronicznych, o bardzo szerokim i różnorodnym zastosowaniu. Ze względu na wielość zastosowań wzmacniacze różnią się między sobą konstrukcją, parametrami, technologią wykonania itp., ale wykazują też ważne cechy wspólne. Na początku niniejszego rozdziału będą podane wiadomości ogólne o wzmacniaczach, a w dalszej części wybrane przykłady wzmacniaczy.

5.1. Schemat blokowy

Powszechnie znany "wzmacniacz" stanowi jeden z elementów wyposażenia profesjonalnych zespołów muzycznych. Do "wzmacniacza" podłącza się mikrofon, lub mikrofony oraz głośnik, lub głośniki, uzyskując w ten sposób pożądane natężenie dźwięku. Oczywiście, aby aparatura działała, to sam "wzmacniacz" musi być włączony do sieci zasilającej,



Rys. 5.1. Wzmacniacz w elektroakustyce,
M – mikrofon, G – głośnik

wzmacniacza (ang. input, symbole wielkości wejściowych mają indeks "i"). Napięcie $u_i(t)$, wytworzone pod wpływem dźwięków w mikrofonie, powoduje przepływ prądu $i_i(t)$ w obwodzie wejściowym wzmacniacza (często ten prąd jest znikomo mały). Pod wpływem sygnału wejściowego, wzmacniacz wytwarza na swoim wyjściu (ang. output, symbole wielkości wyjściowych mają indeks "o") napięcie $u_o(t)$, które jest proporcjonalne do napięcia wejściowego $u_i(t)$:

$$u_o(t) = K_u u_i(t) \quad (5.1)$$

Współczynnik proporcjonalności K_u nazywa się współczynnikiem wzmocnienia napięciowego. W niektórych zastosowaniach, szczególnie elektroakustycznych, stosuje się inną miarę wzmocnienia, obliczaną dla sygnałów sinusoidalnych o wartościach skutecznych U_i , U_o , według zależności logarytmicznej:

$$K_u[\text{dB}] = 20 \lg \frac{U_o}{U_i} \quad (5.1a)$$

zwykle 220 V. Najprostszą strukturę połączeń pokazano na rys. 5.1, przedstawiając wzmacniacz w postaci umownego bloku. Takie przedstawienie (black box – czarna skrzynka) stosuje się wtedy, gdy analizuje się działanie urządzenia, na podstawie właściwości podzespołów, nie wnikając jednak w wewnętrzną budowę samych podzespołów.

W układzie przedstawionym na rys. 5.1 mikrofon jest źródłem analogowego sygnału elektrycznego doprowadzonego do wejścia

Otrzymany wynik opatruje się mianem "decybel"; np. współczynnikowi wzmocnienia $K_u = 100 \text{ V/V}$ odpowiada 40 dB, a wartość 60 dB odpowiada współczynnikowi $K_u = 1000 \text{ V/V}$.

Jeżeli do wyjścia wzmacniacza jest dołączone obciążenie (np. głośnik), to pod wpływem napięcia $u_o(t)$ następuje przepływ prądu wyjściowego $i_o(t)$. Prąd wyjściowy można opisać przy użyciu współczynnika wzmocnienia prądowego K_i zależnością

$$i_o(t) = K_i i_i(t) \quad (5.2)$$

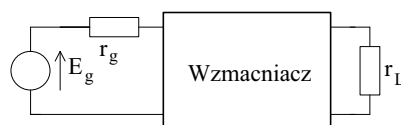
W sytuacji, kiedy wzmacniacz jest połączony ze źródłem sygnału oraz obciążeniem można określić moc elektryczną (mnożąc prąd przez napięcie) doprowadzaną do wejścia wzmacniacza $p_i(t)$ i moc elektryczną odprowadzaną z wyjścia wzmacniacza do obciążenia $p_o(t)$. Podstawową cechą wzmacniacza jest to, że:

$$p_o(t) \gg p_i(t) \quad (5.3)$$

czyli moc i energia elektryczna oddawane do obciążenia przez wzmacniacz są większe od mocy i energii doprowadzonej do wzmacniacza ze źródła sygnału.

Ażeby taka sytuacja była fizycznie realizowalna konieczne jest doprowadzanie dodatkowej mocy (i energii) do wzmacniacza. Dodatkowa energia, pobierana np. z sieci 220 V i odpowiednio przekształcona, jest doprowadzana przez tzw. zasilanie wzmacniacza. Zazwyczaj wzmacniacz wymaga zasilania ze źródła napięcia stałego, które tworzy się, korzystając z prostownika napięcia przemiennego sieci energetycznej lub baterii. Moc zasilania i moc sygnału wejściowego w sumie są większe od mocy wyjściowej, ponieważ w elementach wewnętrznych wzmacniacza część mocy zostaje zamieniona w ciepło, czyli stracona w sensie procesu wzmacniania sygnału. Ponadto zwykle moc sygnału wejściowego jest pomijalnie mała względem mocy zasilania i dlatego do wyznaczania sprawności wzmacniacza (jeżeli jest to konieczne) uwzględnia się tylko moc wyjściową i moc zasilania.

Analiza wzmacniacza najczęściej ma na celu określenie co się dzieje z sygnałem doprowadzonym do jego wejścia, czyli powinna doprowadzić do opisu procesu wzmacniania sygnału, przez wyznaczenie głównych parametrów wzmacniacza. Zazwyczaj zasilanie wzmacniacza jest traktowane jako oczywista konieczność techniczna i jest pomijane w podstawowych rozważaniach. W ten sposób najprostszy układ z wzmacniaczem zostaje sprowadzony do postaci pokazanej na rys. 5.2.



Rys. 5.2. Wzmacniacz jako czwórnik

Na tym rysunku występuje czwórnik elektryczny, reprezentujący wzmacniacz oraz dwójniki elektryczne, elementy E_g oraz r_g symbolizują siłę elektromotoryczną i rezystancję wewnętrzną źródła sygnału wejściowego, wielkość r_L przedstawia obciążenie wzmacniacza.

5.2. Opis wzmacniacza

5.2.1. Parametry czwórnikowe wzmacniacza

Dla opisu właściwości wzmacniacza, traktowanego jako czwórnik, dogodnie jest zastosowanie równań o postaci:

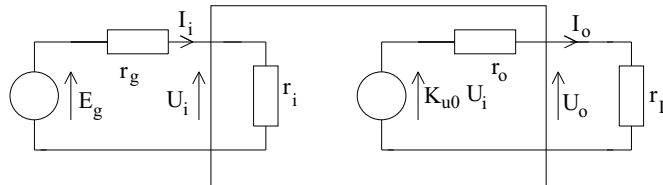
$$\begin{aligned} I_i &= a_{11}U_i + a_{12}I_o \\ U_o &= a_{21}U_i + a_{22}I_o \end{aligned} \quad (5.4)$$

Sens fizyczny parametrów a_{ij} czwornika jest niejednorodny. Parametr a_{11} opisuje wpływ napięcia wejściowego na prąd wejściowy i ma wymiar konduktancji (odwrotność rezystancji), nazywa się go przewodnością wejściową lub odwrotnością rezystancji wejściowej r_i .

Parametr a_{12} opisuje wpływ prądu wyjściowego na prąd wejściowy czyli przedstawia oddziaływanie zwrotne we wzmacniaczu. Zwykle we wzmacniaczach elektronicznych takie oddziaływanie jest pomijalnie małe i z reguły przyjmuje się $a_{12} = 0$. Czwórnik o takiej właściwości nazywa się unilateralnym.

Parametr a_{21} przedstawia wpływ napięcia wejściowego na napięcie wyjściowe, czyli pokazuje jak wzmacniacz zmienia napięcie sygnału wejściowego. Jest to współczynnik wzmocnienia napięciowego, rozwarciowy (ze względu na sposób jego definiowania, wyznacza się go bowiem jako stosunek napięć przy $I_o = 0$, zob. równanie (5.4) przy takim warunku), oznaczany symbolem K_{u0} , wielkość bezwymiarowa, niekiedy opatrywana pseudo-mianem V/V lub V/mV dla podkreślenia jego sensu.

Parametr a_{22} przedstawia wpływ prądu wyjściowego na napięcie wyjściowe, ma wymiar rezystancji i nazywa się go rezystancją wyjściową r_o . Oddziaływanie rezystancji wyjściowej objawia się tym, że przy zwiększaniu obciążenia wzmacniacza (wzrost I_o) następuje zmniejszenia napięcia U_o , co prowadzi do znaku minus w równaniach.



Rys. 5.3. Model czwórnikowy wzmacniacza

Uwzględniając powyższe uwagi i określenia można zapisać uproszczone równania czwórnikowe wzmacniacza w postaci:

$$\begin{aligned} I_i &= \frac{U_i}{r_i} \\ U_o &= K_{u0}U_i - r_o I_o \end{aligned} \quad (5.5)$$

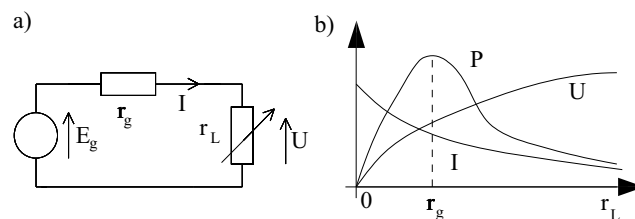
Czwórnikowy model wzmacniacza odpowiadający takiemu opisowi pokazano na rys. 5.3. Do modelu wzmacniacza dołączono model rzeczywistego źródła sygnału złożony ze źródła siły elektromotorycznej E_g i rezystancji wewnętrznej tego źródła r_g oraz rezystancję r_L modelującą obciążenie wzmacniacza. Elementy narysowane wewnątrz bloku czwórnika modelują parametry czwórnikowe wzmacniacza.

Źródło napięciowe umieszczone wewnątrz czwórnika jest źródłem sterowanym, jego SEM zależy od napięcia wejściowego. W takim układzie, jeżeli parametry są znane, to można wyliczyć prądy i napięcia występujące przy pracy wzmacniacza.

Jeżeli np. $r_g = 1 \text{ k}\Omega$, $r_i = 1 \text{ k}\Omega$, $r_o = 1 \text{ k}\Omega$, $r_L = 1 \text{ k}\Omega$, $K_{u0} = 100$, to dla sygnału np. $E_g = 1 \text{ V}$ można wyznaczyć $U_i = 0,5 \text{ V}$, $I_i = 0,5 \text{ mA}$, $U_o = 25 \text{ V}$, $I_o = 25 \text{ mA}$. Znaczący to, że w tym przypadku wzmocnienie napięciowe wynosi 50, wzmocnienie prądowe też 50, a wzmocnienie mocy sygnału 2500. Jak można zauważyć, skuteczność wzmacniania sygnału zależy nie tylko od parametrów wzmacniacza, ale też od rezystancji źródła sygnału i obciążenia wzmacniacza. Taka zależność prowadzi do tzw. problemu dopasowania wzmacniacza do źródła i obciążenia.

Na schemacie pokazanym na rys. 5.3 można zauważyć dwa podobne oczka, jedno w obwodzie wejściowym a drugie w obwodzie wyjściowym. Każde z oczek prezentuje model rzeczywistego źródła, z rezystancją wewnętrzną i jego obciążenie. Analizując wpływ obciążenia na wartości prądu, napięcia na obciążeniu i mocy wydzielanej w obciążeniu znajduje się zależności jak pokazano na rys. 5.4.

Jeżeli rezystancja obciążenia jest znacznie większa od rezystancji wewnętrznej źródła to otrzymuje się duże napięcie na obciążeniu, jeżeli jest odwrotnie występuje duży prąd w obciążeniu a jeżeli te dwie rezystancje są równe, to w obciążeniu wystąpi maksymalna moc.



Rys. 5.4. Dopasowanie obciążenia do źródła: schemat (a), zależność prądu, napięcia i mocy od rezystancji obciążenia r_L (b)

Przenosząc te wnioski na wzmacniacz można znaleźć, że dla $r_i \gg r_g$ jest $U_i \approx E_g$ jeżeli także $r_L \gg r_o$, to $U_o \approx K_{u0} E_g$. W takim przypadku otrzymuje się największe możliwe wzmocnienie SEM źródła sygnału bo $U_o / E_g \approx U_o / U_i \approx K_{u0}$ i wzmacniacz nazywa się napięciowym. Jeżeli natomiast $r_g = r_i$, to $U_i = E_g/2$, i jeżeli także $r_o = r_L$, to $U_o = K_{u0} U_i/2$, czyli łącznie $U_o/E_g = K_{u0}/4$; skuteczne wzmocnienie SEM źródła sygnału jest cztery razy mniejsze niż w wzmacniaczu napięciowym, ale otrzymuje się maksymalne wzmocnienie mocy sygnału (sprawdzenie tego wniosku pozostawia się czytelnikowi), czyli otrzymuje się wzmacniacz mocy. Podobnie można określić wzmacniacz prądu wejściowego. Jeżeli dopa-

sowanie rezystancji na wejściu i wyjściu wzmacniacza odbiega od wartości teoretycznych, to wzmocnienie napięcia, mocy czy prądu jest mniejsze od maksymalnego, możliwego do uzyskania z danego wzmacniacza.

Bardzo często stawia się warunek, aby wzmacniacz był napięciowy tj. silnie wzmacniał napięcie sygnału. Spełnienie takiego warunku wymaga, aby rezystancja wejściowa wzmacniacza r_i była możliwie duża a rezystancja wyjściowa r_o możliwie mała. W idealnym wzmacniaczu napięciowym rezystancja wejściowa $r_i = \infty$, natomiast rezystancja wyjściowa $r_o = 0$.

W konkretnym wzmacniaczu wartość parametrów czwórnikowych zależy od zastosowanych elementów i sposobu ich połączeń i nie może być zmieniana w dowolnie szerokich granicach. Sprzężenie zwrotne, obok innych zalet, umożliwia modyfikację parametrów wzmacniacza.

5.2.2. Sprzężenie zwrotne

Sprzężenie zwrotne polega na tym, że część sygnału wyjściowego wzmacniacza wprowadza się z powrotem na jego wejście. W obwodzie wejściowym może nastąpić dodawanie sygnału sprzężenia zwrotnego do sygnału pochodzącego ze źródła sygnału lub odejmowanie sygnału sprzężenia od sygnału źródła. W pierwszym przypadku określa się sprzężenie jako dodatnie, w drugim jako ujemne. We wzmacniaczach stosuje się sprzężenie ujemne i tylko takie jest rozpatrywane w niniejszym punkcie. Tor powrotny sygnału sprzężenia zwrotnego prowadzi przez drugi czwórnik, często o bardzo prostej budowie wewnętrznej.

Na rys. 5.5 pokazano przykład połączenia wzmacniacza K (parametry z indeksem K) z członem sprzężenia zwrotnego B (parametry z indeksem B). Jest to sprzężenie ujemne – co wynika z wzajemnego ukierunkowania strzałek E_g oraz U_{oB} – napięciowe, bo sygnałem wejściowym czwórnika sprzężenia zwrotnego jest napięcie wyjściowe wzmacniacza oraz szeregowo, bo sygnał sprzężenia zwrotnego jest połączony szeregowo ze źródłem sygnału wejściowego.

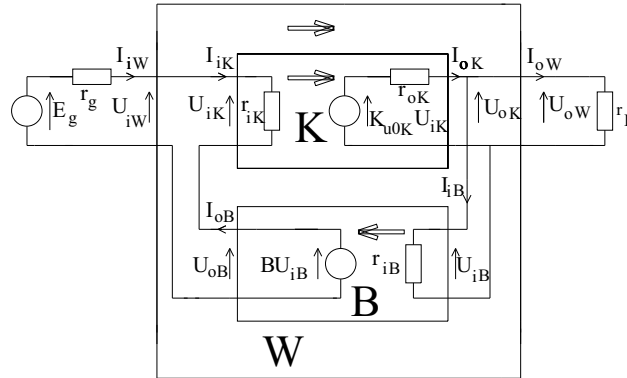
Wzmacniacz K opisują równania, uproszczone podobnie jak równania (5.5):

$$\begin{aligned} I_{iK} &= \frac{U_{iK}}{r_{iK}} \\ U_{oK} &= K_{u0} U_{iK} - r_{oK} I_{oK} \end{aligned} \quad (5.6)$$

a czwórnik sprzężenia zwrotnego B opiszemy równaniami jeszcze bardziej uproszczonymi, aby uniknąć komplikacji formalnych:

$$\begin{aligned} I_{iB} &= \frac{U_{iB}}{r_{iB}} \\ U_{oB} &= B U_{iB} \end{aligned} \quad (5.7)$$

Układ dwu czwórników, pokazany na rys. 5.5, można potraktować jako pewien nowy, wypadkowy czwórnik, którego parametry (z indeksem W) są określone przez parametry obydwu czwórników składowych oraz sposób ich połączenia.



Rys. 5.5. Wzmacniacz K z członem sprzężenia zwrotnego B tworzą nowy czwórnik W

Na podstawie schematu można zapisać poniższe równania, dotyczące wielkości elektrycznych tego połączenia:

$$\begin{aligned}
 U_{iW} &= U_{iK} + U_{oB} \\
 I_{iW} &= I_{iK} = -I_{oB} \\
 U_{oW} &= U_{oK} = U_{iB} \\
 I_{oW} &= I_{oK} - I_{iB}
 \end{aligned}
 \tag{5.8}$$

Na podstawie tych równań, uwzględniając parametry czwórników K oraz B, można zapisać pełne równania czwórnikowe dla wypadkowego układu W, czyli wzmacniacza ze sprzężeniem zwrotnym. Po przekształceniach otrzymuje się:

$$\begin{aligned}
 I_{iW} &= \frac{1}{r_{iK}} \frac{1 + \frac{r_{oK}}{r_{iB}}}{1 + \frac{r_{oK}}{r_{iB}} + K_{u0K} B} U_{iW} + \frac{r_{oK}}{r_{iK}} \frac{B}{1 + \frac{r_{oK}}{r_{iB}} + K_{u0K} B} I_{oW} \\
 U_{oW} &= \frac{K_{u0K}}{1 + \frac{r_{oK}}{r_{iB}} + K_{u0K} B} U_{iW} - \frac{r_{oK}}{1 + \frac{r_{oK}}{r_{iB}} + K_{u0K} B} I_{oW}
 \end{aligned}
 \tag{5.9}$$

Powyższe wzory zawierają parametry nowego, "wypadkowego" czwórnika, jaki powstał przez połączenie wzmacniacza z członem sprzężenia zwrotnego. Postać wzorów wskazuje, że prąd wejściowy zależy od napięcia wejściowego, ale także od prądu wyjściowego, co jest skutkiem wprowadzonego sprzężenia. Napięcie wyjściowe zależy od napięcia wejściowego i prądu obciążenia, podobnie jak w innych wzmacniaczach.

Aby oszacować skutki ilościowe sprzężenia zwrotnego założmy, że wzmacniacz K jest dobrym wzmacniaczem napięciowym i jego rezystancja wyjściowa r_{oK} jest znacznie mniejsza od rezystancji wejściowych r_{iK} oraz r_{iB} . Wówczas wzory (5.9) upraszczają się do postaci:

$$\begin{aligned}
 I_{iW} &\approx \frac{1}{r_{iK}(1+K_{u0K}B)} U_{iW} \\
 U_{oW} &\approx \frac{K_{u0K}}{1+K_{u0K}B} U_{iW} - \frac{r_{oK}}{1+K_{u0K}B} I_{oW}
 \end{aligned}
 \tag{5.10}$$

Te wzory dobitniej pokazują, że nowy czwórnik jest wzmacniaczem, którego rezystancja wejściowa wzrosła $(1+K_{u0K}B)$ razy w stosunku do rezystancji wejściowej wzmacniacza K , natomiast wzmocnienie napięciowe, rozwarciowe zmniejszyło się $(1+K_{u0K}B)$ -krotnie, podobnie jak jego rezystancja wyjściowa. W ten sposób, dobierając parametry członu sprzężenia zwrotnego, można modyfikować parametry "wypadkowego" wzmacniacza. Interesującą cechą wzmacniacza ze sprzężeniem zwrotnym można zauważyć, przyjmując że wzmocnienie rozwarciowe wzmacniacza K_{u0K} jest bardzo duże, zmierzające do nieskończoności (ściślej warunek dotyczy iloczynu $K_{u0K}B$, który powinien być znacznie większy od jedności). W takim przypadku wzory przybliżone (5.10) redukują się do postaci:

$$\begin{aligned}
 I_{iW} &\approx 0 \\
 U_{oW} &\approx \frac{1}{B} U_{iW}
 \end{aligned}
 \tag{5.11}$$

Ten zapis oznacza, że jeżeli wzmacniacz realny ma bardzo duże wzmocnienie napięciowe i zostanie zaopatrzone w ujemne sprzężenie zwrotne, to powstający czwórnik ma cechy wzmacniacza napięciowego idealnego. Jego rezystancja wejściowa jest nieskończenie duża (prąd wejściowy wynosi zero) natomiast rezystancja wyjściowa nieskończenie mała (napięcie wyjściowe nie zależy od prądu obciążenia). Wzmocnienie napięciowe powstałego czwornika zależy tylko od parametru B członu sprzężenia zwrotnego. Zatem, jeżeli tylko wzmocnienie napięciowe realnego wzmacniacza jest bardzo duże, to jego pozostałe parametry nie wpływają na parametry układu ze sprzężeniem zwrotnym. To zjawisko umożliwiło szerokie zastosowanie scalonych wzmacniaczy operacyjnych.

Na koniec należy zaznaczyć, że pokazany na rys. 5.5 układ połączeń wzmacniacza z członem sprzężenia zwrotnego jest tylko jednym z kilku możliwych. W literaturze np. [1] można znaleźć wyczerpujące opisy i porównania różnych układów.

5.2.3. Charakterystyki wzmacniaczy

Dogodną formą przedstawiania pewnych właściwości wzmacniaczy są charakterystyki, czyli zależności prezentowane graficznie. Do ważniejszych charakterystyk wzmacniaczy należą:

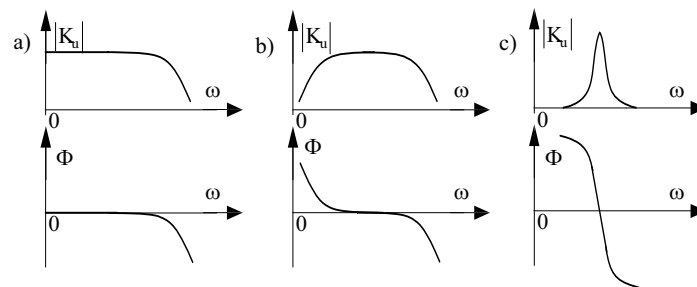
- charakterystyki częstotliwościowe,
- charakterystyka przenoszenia.

Charakterystyki częstotliwościowe pokazują, w jakim stopniu jest wzmacniany sygnał o kształcie sinusoidalnym, o różnych częstotliwościach, czyli przedstawiają zależność współczynnika wzmocnienia napięciowego od częstotliwości sygnału ω . Charakterystyka przenoszenia pokazuje, jak są wzmacniane sygnały o różnych wartościach (amplitudach, jeżeli sygnał jest sinusoidalny).

Zależność pomiędzy działaniem wzmacniacza a częstotliwością wzmacnianego sygnału, którą prezentują charakterystyki częstotliwościowe, wynika z tego, że w konstrukcji wzmacniaczy są stosowane elementy o charakterze pojemnościowym (niekiedy i indukcyjnym), których reaktancja zależy od częstotliwości. Takimi elementami są stosowane w układach wzmacniaczy kondensatory oraz tranzystory, diody itp., w których występują pojemności (pasożytnicze).

Te elementy powodują, że stosunek amplitud, lub wartości skutecznych, sinusoidalnego napięcia wyjściowego do również sinusoidalnego napięcia wejściowego (współczynnik wzmocnienia napięciowego) jest różny przy różnych częstotliwościach sygnału, a ponadto występuje przesunięcie fazowe pomiędzy sinusoidą napięcia wejściowego a sinusoidą napięcia wyjściowego. Z tego powodu stosuje się dwie charakterystyki częstotliwościowe: amplitudową i fazową.

Charakterystyka częstotliwościowa: amplitudowa pokazuje zależność modułu współczynnika wzmocnienia $|U_o/U_i| = |K_u(\omega)|$ od częstotliwości sygnału a charakterystyka fazowa zależność fazy $\Phi = \Phi(\omega)$ od częstotliwości (zob. też p. 4.4.1 – filtracja). Na rys. 5.6 podano przykłady typowych charakterystyk częstotliwościowych wzmacniaczy.

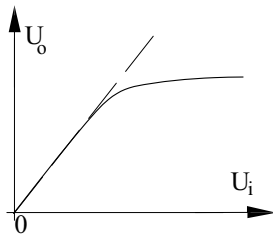


Rys. 5.6. Charakterystyki częstotliwościowe wzmacniacza: prądu stałego (a), pasmowego (b), selektywnego (c)

Jedną z konsekwencji właściwości częstotliwościowych wzmacniaczy są tzw. zniekształcenia liniowe, pojawiające się przy wzmacnianiu sygnałów niesinusoidalnych. Sygnał niesinusoidalny, zgodnie z zależnościami Fouriera (szeregiem lub całką), można traktować jako nieskończoną sumę sygnałów sinusoidalnych. W procesie wzmacniania różne składniki niesinusoidalnego sygnału wejściowego są wzmacniane w różnym stopniu i różnie przesuwane fazowo. Skutkiem tego jest, że sygnał wyjściowy ma inny kształt aniżeli wejściowy. Stopień zniekształcenia zależy od relacji pomiędzy składowymi sygnału a charakterystykami wzmacniacza.

Przedstawiony w p. 5.2.1. opis wzmacniacza jako czwórnika liniowego jest zazwyczaj słuszny tylko w ograniczonym zakresie wartości napięć wejściowego i wyjściowego. Dla opisanego zjawiska wyznacza się charakterystykę przenoszenia, jako zależność wartości napięcia wyjściowego od napięcia wejściowego $U_o(U_i)$. We wzmacniaczu liniowym powinna to być prosta, o nachyleniu proporcjonalnym do współczynnika wzmocnienia napię-

ciowego, zgodnie z zależnością $U_o = K_u U_i$. We wzmacniaczu rzeczywistym zawsze występuje odchylenie od prostej, czyli pojawia się nieliniowość charakterystyki przenoszenia. Powodem nieliniowości charakterystyki jest nieliniowość samych elementów wzmacniających (tranzystorów) i ograniczenia w ich warunkach pracy. Nieliniowości są szczególnie wydatne przy dużych sygnałach, ale niekiedy także i dla sygnałów bliskich zera.



Rys. 5.7. Charakterystyka przenoszenia sygnału sinusoidalnego

Łatwo sobie uzmystwić, że np. uzyskanie dowolnie dużej wartości napięcia wyjściowego jest technicznie niemożliwe (ograniczenie wnosi zwykle wartość zastosowanego napięcia zasilania, dostosowana z kolei do wytrzymałości napięciowej użytych elementów).

Nieliniowość charakterystyki przenoszenia oznacza zmienność (niestałość) współczynnika wzmocnienia napięciowego. Skutkiem tego efektu jest, że sygnały o małych wartościach są wzmacniane inaczej niż sygnały o dużych wartościach. Powoduje to odkształcenie sygnału wyjściowego np. po wzmocnieniu sygnału sinusoidalnego otrzymuje się napięcie wyjściowe o kształcie prawie trapezowym. Te zniekształcenia nazywa się nieliniowymi.

Niekiedy zamiast charakterystyki przenoszenia można podać zakres napięć (wejściowych lub wyjściowych), w którym współczynnik wzmocnienia zachowuje stałą wartość albo charakterystyka pokrywa się z linią prostą. Na rys. 5.7 podano przykład rzeczywistej charakterystyki przenoszenia, wyznaczonej z pomiaru wartości skutecznej napięć przy sinusoidalnym sygnale wejściowym.

5.3. Wzmacniacz różnicowy

W automatyce i metrologii najszerzej obecnie używanym wzmacniaczem sygnałów, zarówno przeniennopiędowych jak i stałoprądowych, jest scalony wzmacniacz operacyjny. Taki scalony wzmacniacz zawiera kaskadowe połączenie kolejnych stopni wzmacniających, z których pierwszy jest z reguły wzmacniaczem różnicowym. Właściwości wzmacniacza operacyjnego w znacznym stopniu wynikają z cech wzmacniacza różnicowego.

W niniejszym punkcie wzmacniacz różnicowy jest przedstawiony w dwu etapach, na początku opisuje się ogólne właściwości wzmacniacza różnicowego i jego działanie pod wpływem wybranych kombinacji sygnałów wejściowych. W dalszym ciągu przedstawia się parametry i model nieliniowy tego wzmacniacza. Właściwości dynamiczne wzmacniacza różnicowego nie są opisane, uzupełniające informacje można znaleźć w literaturze, np. [1].

5.3.1. Budowa i działanie wzmacniacza różnicowego

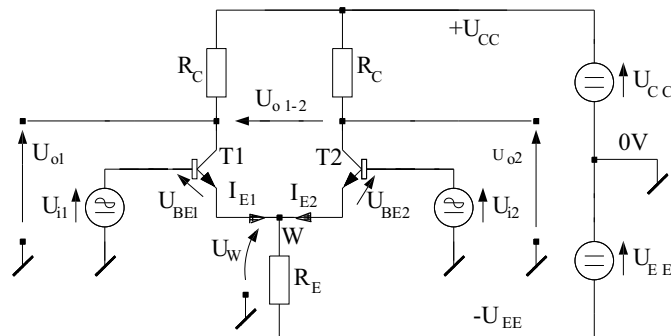
Prosty schemat elektryczny wzmacniacza różnicowego, wykonanego z tranzystorów bipolarnych, pokazano na rys. 5.8. Układ zawiera dwie gałęzie, z których każda jest złożona z rezystora R_C i tranzystora bipolarnego a ponadto jest włączony jeden rezystor R_E ,

wspólny dla obydwu tranzystorów. Zasilanie układu stanowią dwa źródła napięcia stałego połączone szeregowo, zwykle wartości napięć tych źródeł są jednakowe np. po 6 V czy 12 V. Środkowy punkt połączenia źródeł zasilania jest traktowany jako punkt potencjału odniesienia dla całego wzmacniacza i nazywany "masą" lub "ziemią elektryczną", a oznaczany umownie jako "zero woltów – 0 V". W stosunku do masy jedno z napięć zasilania jest więc dodatnie a drugie ujemne, dlatego często zasilanie wzmacniacza opisuje się skrótowo jako np. ± 6 V albo ± 12 V.

Wzmacniacz różnicowy ma dwa wejścia: pierwsze z nich to baza tranzystora T1, a drugie to baza tranzystora T2. Sygnały wejściowe, wprowadzane pomiędzy wejście a masę, oznaczono U_{i1} oraz U_{i2} . Wyjściami wzmacniacza, czyli punktami gdzie można przyłączyć obciążenie, są: kolektor tranzystora T1 – masa i/lub kolektor tranzystora T2 – masa, są to dwa tzw. wyjścia asymetryczne, sygnały na tych wyjściach oznaczono U_{o1} oraz U_{o2} ; określa się też trzecie wyjście, zwane symetrycznym lub różnicowym, są to obydwa kolektory tranzystorów T1 i T2, sygnał wyjściowy na tym wyjściu jest oznaczony U_{o1-2} .

Aby tranzystory mogły działać jako elementy wzmacniające ich złącza muszą być odpowiednio spolaryzowane, czyli wszystkie elektrody tranzystora muszą być połączone z odpowiednimi napięciami. Dla zapewnienia takiej polaryzacji każda z baz tranzystorów T1 i T2 musi być połączona ze wspólną masą albo przez źródło sygnału wejściowego, albo przez inny element przewodzący prąd stały albo wprost przez zwarcie.

Analizę układu z dwoma źródłami sygnałów wejściowych, którego schemat pokazano na rysunku 5.8, przeprowadzimy najpierw opisując ten układ ogólnymi równaniami, a następnie będziemy rozpatrywać wybrane przypadki szczególne, modyfikując odpowiednio do nich uprzednio zapisane równania.



Rys. 5.8. Wzmacniacz różnicowy z zasilaniem i dwoma źródłami sygnałów wejściowych

Przy uproszczonej analizie wzmacniacza różnicowego przyjmuje się idealną symetrię układu, tj. zakłada się, że rezystory kolektorowe są jednakowe, a obydwa tranzystory są identyczne. Ponadto zakłada się, że wzmacnienie prądowe tranzystorów jest bardzo duże, czyli sterujący prąd bazy jest znacznie mniejszy od prądu kolektora $I_B \ll I_C$, czego konsekwencją jest prąd kolektora prawie równy prądowi emitera $I_C \approx I_E$.

Obwód wejściowy wzmacniacza opisuje równanie

$$U_{i1} - U_{i2} = U_{BE1} - U_{BE2} \quad (5.12)$$

Napięcie pomiędzy punktem W połączenia emiterów a masą wynosi:

$$U_W = (I_{E1} + I_{E2}) R_E - U_{EE} \quad (5.13)$$

Wielkości wejściowe można też zapisać w uzależnieniu od potencjału punktu W:

$$U_W = U_{i1} - U_{BE1} = U_{i2} - U_{BE2} \quad (5.14)$$

Napięcia wyjściowe asymetryczne wynoszą:

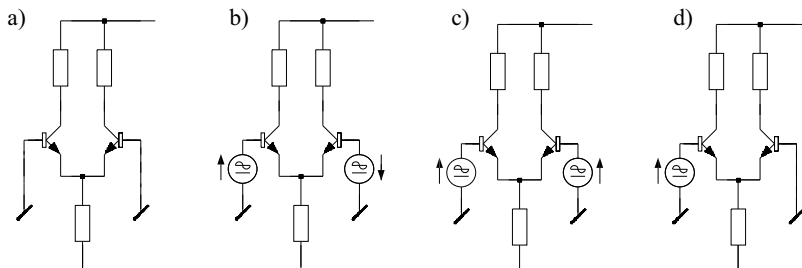
$$U_{o1} = U_{CC} - R_C I_{C1}; \quad U_{o2} = U_{CC} - R_C I_{C2} \quad (5.15)$$

czyli ich wartość zależy zarówno od parametrów układu (R_C , U_{CC}) jak i stanu tranzystorów (I_C).

Napięcie wyjściowe symetryczne jest:

$$U_{o1-2} = U_{o1} - U_{o2} = -R_C(I_{C1} - I_{C2}) \quad (5.16)$$

Warto zauważyć, że wpływ napięcia zasilającego U_{CC} na napięcia wyjściowe asymetryczne jest jawny, natomiast napięcie wyjściowe symetryczne nie zależy (przynajmniej w sposób jawny) od napięcia zasilania.



Rys. 5.9. Wzmacniacz różnicowy przy różnych kombinacjach sygnałów wejściowych

Dla określenia jak wzmacniacz różnicowy wzmacnia sygnały wejściowe rozpatrzmy cztery przypadki szczególne, których uproszczone schematy pokazano na rys. 5.9, kolejno zakładając że:

- obydwa sygnały wejściowe wynoszą zero, jest to sytuacja "gotowości do pracy wzmacniacza", w tym stanie prądy i napięcia tranzystorów określają ich punkt pracy;
- do jednego z wejść jest doprowadzony mały sygnał o niezerowej wartości, a do drugiego wejścia sygnał o tej samej wartości, ale o odwrotnym znaku, wzmacniacz jest sterowany "różnicowo";
- do obydwu wejść doprowadzone są małe, niezerowe sygnały o takich samych wartościach i znakach, jest to sytuacja równoważna zwarceniu obydwu wejść i wprowadzeniu sygnału "wspólnego" dla nich;

d) do jednego z wejść jest doprowadzony mały sygnał o niezerowej wartości, a do drugiego wejścia sygnał zerowy (drugie wejście zwarte z masą), wzmacniacz jest sterowany asymetrycznie.

Traktując sytuację w przypadku **a** jako stan odniesienia, będziemy określać zmiany prądów i napięć we wzmacniaczu powstające w sytuacjach **b**, **c** i **d**. Dla ujednoznacznienia zapisu wartości prądów i napięć w różnych przypadkach będziemy je opatrywać górnymi indeksami a, b, c lub d, odpowiednio do sytuacji której dotyczą.

Przypadek a

$$U_{i1} = U_{i2} = 0$$

zatem:

$$U_{BE1}^a = U_{BE2}^a = U_{BE}^a \quad (5.17)$$

Przy jednakowych napięciach baza – emiter w jednakowych tranzystorach płyną jednakowe prądy:

$$I_{C1}^a = I_{C2}^a = I_C^a; \quad I_{E1}^a = I_{E2}^a = I_E^a \quad (5.18)$$

Napięcie w punkcie W wynosi

$$U_W^a = -U_{BE}^a \quad (5.19)$$

albo

$$U_W^a = 2I_E^a R_E - U_{EE} \quad (5.20)$$

a zatem prąd emiterowy każdego z tranzystorów:

$$I_E^a = \frac{U_{EE} - U_{BE}^a}{2R_E} \quad (5.21)$$

Zakładając wartość napięcia baza – emiter (np. 0,6 V), napięcie zasilanie i wartość rezystora emiterowego można obliczyć prąd emiterowy, a w kolejności i pozostałe wielkości opisujące punkt pracy tranzystorów. Jak założono poprzednio, prądy kolektorowe są w przybliżeniu równe emiterowym $I_C \approx I_E$, co pozwala wyznaczyć napięcia wyjściowe jako:

$$U_{o1}^a = U_{o2}^a = U_{CC} - R_C I_C^a > 0 \quad (5.22)$$

oraz

$$U_{o1-2}^a = 0 \quad (5.23)$$

W opisanym stanie pracy, z trzech napięć wyjściowych wzmacniacza różnicowego tylko wyjściowe napięcie różnicowe, na wyjściu symetrycznym, odpowiada ściśle najprostszemu obrazowi działania wzmacniacza, polegającemu na przyjęciu, że jeżeli sygnał wejściowy jest zero (jak w rozpatrywanej sytuacji), to sygnał wyjściowy też ma być zero.

W odniesieniu do napięć na wyjściach asymetrycznych taki stan nie występuje, pomimo zerowych napięć wejściowych napięcia wyjściowe są niezerowe, co wymaga następującego komentarza: jako sygnał wyjściowy na wyjściu asymetrycznym należy traktować nie bezwzględną wartość napięcia na tym wyjściu, ale zmianę tego napięcia, obliczoną

przez porównanie do wartości, jaka występowała bez sygnałów wejściowych, czyli w stanie oznaczonym jako przypadek **a**.

Przypadek b

$$U_{i1} = U_i^b; \quad U_{i2} = -U_i^b$$

zatem:

$$U_{BE1}^b - U_{BE2}^b = 2U_i^b \quad (5.24)$$

Doprowadzone sygnały wejściowe spowodowały zmianę wysterowania tranzystorów T1 i T2. Porównując wartości napięć baza – emiter w przypadku **b** ze stanem w przypadku **a** zapisujemy zmiany tych napięć jako:

$$\Delta U_{BE1}^b = U_{BE1}^b - U_{BE}^a; \quad \Delta U_{BE2}^b = U_{BE2}^b - U_{BE}^a \quad (5.25)$$

Można wykazać, że dla małych wartości sygnału wejściowego (pojęcie "mały sygnał" będzie sprecyzowane w następnym punkcie) jest:

$$\Delta U_{BE1}^b = U_i^b; \quad \Delta U_{BE2}^b = -U_i^b \quad (5.26)$$

co oznacza, że napięcie baza – emiter tranzystora T1 wzrasta, a tranzystora T2 maleje, w porównaniu do stanu w przypadku **a**.

Dla wyznaczenia skutków zmiany wielkości U_{BE} w tranzystorze bipolarnym posłużymy się równaniami (3.18), opisującymi liniowy model tego tranzystora. Te równania jeszcze bardziej uprościmy, odrzucając człony, które zazwyczaj są znacznie mniejsze od pozostałych, a pozostawiając tylko zależności:

$$\Delta U_{BE} \approx h_{11e} \Delta I_B; \quad \Delta I_C \approx h_{21e} \Delta I_B \quad (5.27)$$

z których otrzymuje się:

$$\Delta I_C \approx \frac{h_{21e}}{h_{11e}} \Delta U_{BE} \approx \Delta I_E \quad (5.28)$$

Zmiana napięć baza – emiter wywoła więc zmiany prądów tranzystorów. W porównaniu do stanu w przypadku **a** zmiany prądów kolektorowych wynoszą:

$$\begin{aligned} \Delta I_{C1}^b &= I_{C1}^b - I_{C1}^a \approx \frac{h_{21e}}{h_{11e}} \Delta U_{BE1}^b \approx \frac{h_{21e}}{h_{11e}} U_i^b \\ \Delta I_{C2}^b &= I_{C2}^b - I_{C2}^a \approx \frac{h_{21e}}{h_{11e}} \Delta U_{BE2}^b \approx -\frac{h_{21e}}{h_{11e}} U_i^b \end{aligned} \quad (5.30)$$

Odnotujmy, że przyrost prądu w tranzystorze T1 jest równy zmniejszeniu prądu w tranzystorze T2. Skutkiem tego jest stałość sumy prądów kolektorowych:

$$I_{C1}^b + I_{C2}^b = I_{C1}^a + I_{C2}^a \quad (5.31)$$

Pamiętając o założeniu $I_C \approx I_E$, można stwierdzić, że w przypadku **b** prąd płynący przez rezystor emiterowy pozostaje taki sam jak w przypadku **a** i dlatego potencjał punktu W dla obydwu tych przypadków jest taki sam. Wartości sygnałów wyjściowych asymetrycznych określimy jako:

$$\begin{aligned}\Delta U_{o1}^b &= U_{o1}^b - U_{o1}^a = -R_C \Delta I_{C1}^b \approx -R_C \frac{h_{21e}}{h_{11e}} U_i^b \\ \Delta U_{o2}^b &= U_{o2}^b - U_{o2}^a = -R_C \Delta I_{C2}^b \approx R_C \frac{h_{21e}}{h_{11e}} U_i^b\end{aligned}\quad (5.32)$$

Wprowadzone sygnały wejściowe spowodowały zmniejszenie napięcia na wyjściu asymetrycznym z tranzystora T1 i wzrost napięcia na wyjściu z tranzystora T2. Zauważmy, że zmiana napięcia wyjściowego jest proporcjonalna do sygnału wejściowego, współczynnik proporcjonalności jest to współczynnik wzmocnienia napięciowego i w tym przypadku wynosi on $h_{21e}R_C/h_{11e}$. Skutkiem przeciwnych zmian napięć na obydwu wyjściach asymetrycznych jest pojawienie się dwukrotnie większego sygnału na wyjściu symetrycznym:

$$U_{o1-2}^b = U_{o1}^b - U_{o2}^b \approx -2R_C \frac{h_{21e}}{h_{11e}} U_i^b \quad (5.33)$$

Przypadek c

$$U_{i1} = U_{i2} = U_i^c > 0$$

Symetryczne wystawienie obydwu tranzystorów powoduje jednakowe zmiany napięć baza-emiter zatem:

$$U_{BE1}^c = U_{BE2}^c = U_{BE}^c \quad (5.34)$$

Podobnie jak w przypadku **a**, przy jednakowych napięciach baza-emiter w obydwu tranzystorach, także ich prądy będą jednakowe:

$$I_{C1}^c = I_{C2}^c = I_C^c; \quad I_{E1}^c = I_{E2}^c = I_E^c \quad (5.35)$$

Napięcie w punkcie W wynosi

$$U_W^c = U_i^c - U_{BE}^c \quad (5.36)$$

albo może być wyrażone jako:

$$U_W^c = 2I_E^c R_E - U_{EE} \quad (5.37)$$

Porównując wartość tego napięcia w przypadku **c** z wartością tegoż napięcia w przypadku **a** znajduje się:

$$U_W^c - U_W^a = U_i^c - U_{BE}^c + U_{BE}^a = U_i^c - \Delta U_{BE}^c \quad (5.38)$$

albo

$$U_W^c - U_W^a = 2R_E (I_E^c - I_E^a) = 2R_E \Delta I_E^c \quad (5.39)$$

Przyrównując dwie powyższe wielkości i korzystając z zależności (5.28): znajduje się:

$$\Delta I_C^c \approx \frac{h_{21e}}{h_{11e} + 2h_{21e}R_E} U_i^c \quad (5.40)$$

co pozwala obliczyć zmiany napięć na wyjściach asymetrycznych, czyli wartość sygnału wyjściowego asymetrycznego:

$$\Delta U_{o1}^c = \Delta U_{o2}^c = -R_C \Delta I_C^c \approx -\frac{h_{21e} R_C}{h_{11e} + 2h_{21e} R_E} U_i^c \approx -\frac{R_C}{2R_E} U_i^c \quad (5.41)$$

Napięcia na obydwu wyjściach zmieniły się, ale w tym przypadku współczynnik wzmocnienia wynosi w przybliżeniu $-R_C/2R_E$, czyli jest inny aniżeli był wyznaczony w przypadku **b**. Napięcie wyjściowe różnicowe wynosi:

$$U_{o1-2}^c = 0 \quad (5.42)$$

bo napięcia na obydwu wyjściach asymetrycznych zmieniły się jednakowo.

Przypadek d

$$U_{i1} = U_i^d > 0; \quad U_{i2} = 0.$$

Taka sytuacja występuje często w praktyce, gdy sygnał, określony jako napięcie względem wspólnej masy całego układu elektronicznego, ma być wzmocniony we wzmacniaczu różnicowym (lub operacyjnym). Pomimo sterowania tylko jednego wejścia należy założyć, że zmieniają się wielkości elektryczne w obydwu tranzystorach. Napięcia baza – emiter tranzystorów spełniają równość:

$$U_{BE1}^d = U_{BE2}^d + U_i^d \quad (5.43)$$

Napięcie pomiędzy punktem W połączenia emiterów a masą wynosi:

$$U_W^d = U_i^d - U_{BE1}^d = -U_{BE2}^d \quad (5.44)$$

Wielkość zmian napięć baza – emiter w tranzystorach określimy porównując przypadek **d** z przypadkiem **a**:

$$\Delta U_{BE1}^d = U_{BE1}^d - U_{BE1}^a = U_i^d - (U_W^d - U_W^a) \quad (5.45)$$

oraz

$$\Delta U_{BE2}^d = U_{BE2}^d - U_{BE2}^a = U_i^d - (U_W^d - U_W^a) \quad (5.46)$$

Zapisując wartość napięcia U_W w uzależnieniu od prądów emiterowych, otrzymujemy:

$$U_W^d = (I_{E1}^d + I_{E2}^d) R_E - U_{EE} \quad (5.47)$$

Zmianę wartości tego napięcia, spowodowaną sygnałem wejściowym, określamy przez porównanie przypadków **d** i **a**:

$$U_W^d - U_W^a = (I_{E1}^d - I_{E1}^a + I_{E2}^d - I_{E2}^a) R_E = (\Delta I_{E1}^d + \Delta I_{E2}^d) R_E \quad (5.48)$$

Dla wyznaczenia zmian wartości prądów emiterowych posłużymy się równaniem przyrostowym (5.28), otrzymując po przekształceniach:

$$U_W^d - U_W^a \approx \frac{h_{21e} R_E}{h_{11e}} (\Delta U_{BE1}^d + \Delta U_{BE2}^d) \quad (5.49)$$

Podstawiając tę wartość do wzorów (5.45), (5.46) i odpowiednio je przekształcając, otrzymujemy:

$$\Delta U_{BE1}^d \approx \frac{h_{11e} + h_{21e}R_E}{h_{11e} + 2h_{21e}R_E} U_i^d \approx \frac{1}{2} U_i^d \quad (5.50)$$

oraz

$$\Delta U_{BE2}^d \approx -\frac{h_{21e}R_E}{h_{11e} + 2h_{21e}R_E} U_i^d \approx -\frac{1}{2} U_i^d \quad (5.51)$$

Powyższe wzory potwierdzają podaną uprzednio uwagę, że pomimo iż sterowane jest tylko jedno wejście wzmacniacza, to w obydwu tranzystorach napięcia baza – emiter uległy zmianie. Przy znacznym uproszczeniu można też przyjmować, że w rozpatrywanej sytuacji sygnał wejściowy dzieli się na połowy, czyli rozkłada się równo na obydwa tranzystory, działając na nie z odwrotnymi znakami. To powoduje, że w jednej gałęzi wzmacniacza prąd wzrasta a w drugiej maleje. Porównując nieco mniej uproszczone zapisy można jednak dostrzec, że zmiany napięć sterujących obydwa tranzystory różnią się nie tylko znakiem ale i wartością. Jeżeli spełniony jest warunek: $h_{11e} \ll h_{21e}R_E$, to ta różnica wartości napięć jest pomijalnie mała.

Sygnały wyjściowe asymetryczne, rozumiane jako zmiana napięć na wyjściach asymetrycznych, wynoszą:

$$\Delta U_{o1}^d = -R_C \Delta I_{C1}^d \approx -R_C \frac{h_{21e}}{h_{11e}} \frac{h_{11e} + h_{21e}R_E}{h_{11e} + 2h_{21e}R_E} U_i^d \approx -\frac{1}{2} \frac{h_{21e}R_C}{h_{11e}} U_i^d \quad (5.52)$$

$$\Delta U_{o2}^d = -R_C \Delta I_{C2}^d \approx R_C \frac{h_{21e}}{h_{11e}} \frac{h_{21e}R_E}{h_{11e} + 2h_{21e}R_E} U_i^d \approx \frac{1}{2} \frac{h_{21e}R_C}{h_{11e}} U_i^d \quad (5.53)$$

Wartości sygnałów wyjściowych asymetrycznych, podobnie jak napięcia baza – emiter, różnią się znakami i wartościami, i podobnie różnica wartości może być pominięta przy dostatecznie dużej wartości rezystancji R_E . Sygnał wyjściowy różnicowy, czyli napięcie na wyjściu symetrycznym wynosi:

$$U_{o1-2}^d = -R_C (\Delta I_{C1}^d - \Delta I_{C2}^d) \approx -\frac{h_{21e}R_C}{h_{11e}} U_i^d \quad (5.54)$$

Uproszczenie tego wzoru nie wymaga założenia odpowiednio dużej wartości rezystancji R_E . Podobnie jak w przypadku **a** i **b**, także w przypadku **d**, napięcie na wyjściu symetrycznym odpowiada ściśle najprostszemu obrazowi działania wzmacniacza, tj. potwierdza proporcjonalność sygnałów wyjściowego i wejściowego ze współczynnikiem wzmocnienia napięciowego $h_{21e}R_C/h_{11e}$.

Rozpatrzone powyżej przypadki wskazują na pewne cechy wzmacniacza różnicowego:

- napięcia na wyjściach asymetrycznych zmieniają się przy każdej kombinacji sygnałów wejściowych, aczkolwiek z różnymi współczynnikami wzmocnienia napięciowego (por. np. wzory (5.32); (5.41); (5.52)),
- napięcie na wyjściu symetrycznym pojawia się tylko przy niejednakowym wysterowaniu obydwu tranzystorów, ale zawsze z takim samym współczynnikiem wzmocnienia napięciowego (wzory (5.54); (5.33)).

5.3.2. Parametry wzmacniacza różnicowego

Sposób reakcji wzmacniacza różnicowego na sygnały wejściowe narzuca pewien typ opisu jego właściwości. Zamiast rozważania fizycznie przyłączanych sygnałów wejściowych U_{i1} oraz U_{i2} wprowadza się dwie wielkości pomocnicze: wejściowe napięcie różnicowe U_{idiff} oraz wejściowe napięcie wspólne (sumacyjne) U_{icom} , definiowane jak niżej:

$$\begin{aligned} U_{idiff} &= U_{i1} - U_{i2} \\ U_{icom} &= \frac{U_{i1} + U_{i2}}{2} \end{aligned} \quad (5.55)$$

a napięcia wyjściowe uzależnia się od tych sygnałów następująco:

$$\begin{aligned} \Delta U_{o1} &= -\frac{1}{2} K_{udiff} U_{idiff} + K_{ucom} U_{icom} \\ \Delta U_{o2} &= \frac{1}{2} K_{udiff} U_{idiff} + K_{ucom} U_{icom} \\ U_{o1-2} &= -K_{udiff} U_{idiff} \end{aligned} \quad (5.56)$$

Wielkość K_{udiff} nazywa się współczynnikiem wzmocnienia napięciowego wejściowego sygnału różnicowego, a wielkość K_{ucom} współczynnikiem wzmocnienia napięciowego wejściowego sygnału wspólnego.

Sygnały wejściowe, zastosowane w uprzednio opisanych i pokazanych na rys. 5.9 przypadkach szczególnych, odpowiadają następującym wartościom sygnału różnicowego i wspólnego:

przypadek b): $U_{idiff} = 2U_i^b; \quad U_{icom} = 0$

przypadek c): $U_{idiff} = 0; \quad U_{icom} = U_i^c$

przypadek d): $U_{idiff} = U_i^d; \quad U_{icom} = \frac{U_i^d}{2}$

Porównując odpowiednie wzory, wyprowadzone dla przypadków szczególnych, można znaleźć wartości współczynników wzmocnienia napięciowego jako:

$$K_{udiff} \approx h_{21e} \frac{R_C}{h_{11e}} \quad (5.57)$$

oraz

$$K_{ucom} \approx -\frac{h_{21e} R_C}{h_{11e} + 2h_{21e} R_E} \approx -\frac{R_C}{2R_E} \quad (5.58)$$

Idealny wzmacniacz różnicowy powinien dobrze wzmacniać sygnał różnicowy U_{idiff} oraz nie powinien wzmacniać sygnału wspólnego U_{icom} , czyli powinien mieć duży współczynnik K_{udiff} i mały współczynnik K_{ucom} . Taki warunek stawia się, ponieważ reakcja wzmacniacza (zmiana napięć wyjściowych) na sygnał wspólny jest taka sama jak reakcja na następujące, niekorzystne, czynniki:

- działanie temperatury: zmiana temperatury powoduje zmiany prądów i napięć w tranzystorze, jeżeli obydwa tranzystory będą zmieniać temperaturę tak samo, to można to interpretować jako wprowadzenie sygnału wspólnego do nich;
- działanie zakłóceń: jeżeli w przewodach prowadzących do obydwu wejść wzmacniacza będą indukowane takie same napięcia zakłóceń, to będą one stanowiły sygnał wspólny dla wzmacniacza;
- niestabilność zasilania: zmiany wartości napięć zasilających, działające tak samo na obydwa tranzystory, można interpretować jako wprowadzenie sygnału wspólnego do wzmacniacza.

Jeżeli wzmacniacz reaguje słabo, lub wcale, na jednakowe sygnały, których miarą jest wielkość sygnału wspólnego U_{icom} , to również jego reakcja będzie słaba lub żadna na zmiany temperatury, symetryczne zakłócenia i niestabilności źródeł zasilania. Jest to bardzo ważna cecha wzmacniacza różnicowego, decydująca o jego szerokim zastosowaniu.

Wzory (5.56) można przekształcić do postaci:

$$\begin{aligned}\Delta U_{o1} &= -\frac{1}{2}K_{udiff} \left(U_{idiff} + \frac{U_{icom}}{CMRR} \right) \\ \Delta U_{o2} &= \frac{1}{2}K_{udiff} \left(U_{idiff} - \frac{U_{icom}}{CMRR} \right)\end{aligned}\quad (5.59)$$

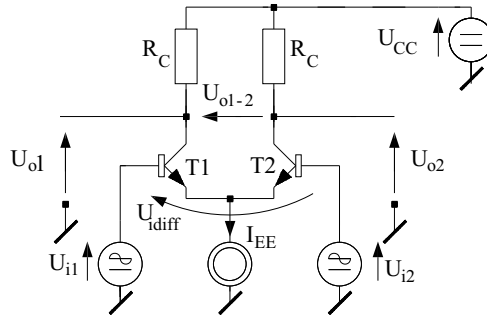
podkreślającej, że sygnał wejściowy wzmacniacza różnicowego należy traktować jako złożony z dwu składników: "pożytecznego" sygnału różnicowego i "szkodliwego" sygnału wspólnego, przyczym sygnał wspólny jest osłabiony przez podzielenie jego wartości przez CMRR (Common Mode Rejection Ratio – współczynnik tłumienia sygnału wspólnego). Wartość CMRR powinna być możliwie duża, w rozpatrywanym układzie wzmacniacza wynosi ona:

$$CMRR = \frac{-K_{udiff}}{2K_{ucom}} = \frac{h_{21e}R_C}{h_{11e}} \frac{h_{11e} + 2h_{21e}R_E}{2h_{21e}R_C} \approx \frac{h_{21e}R_E}{h_{11e}} \quad (5.60)$$

Praktycznie stosowanym sposobem zwiększenia CMRR jest zastąpienie rezystora emiterowego źródłem prądowym. Idealne, teoretyczne źródło prądowe, jak pokazane na schemacie rys. 5.10, ma rezystancję wewnętrzną nieskończenie dużą i zapewnia stałą wartość prądu. Praktyczny układ tranzystorowy, działający podobnie jak źródło prądowe, ma rezystancję wewnętrzną dużą, ale nie nieskończenie dużą, dlatego w praktycznych układach wzmacniaczy różnicowych wartość współczynnika tłumienia sygnału wspólnego CMRR jest duża ale skończona.

5.3.3. Model nieliniowy wzmacniacza różnicowego

Na rys. 5.10 pokazano schemat wzmacniacza zasilanego przez źródło napięciowe U_{CC} i źródło prądowe I_{EE} . Pozostałe oznaczenia i założenia są takie same jak używane w poprzednich punktach. Analizę układu przeprowadza się opierając się na modelach tranzystorów przedstawionych w p. 3.4.2.



Rys. 5.10. Zasilanie prądowe emiterów we wzmacniaczu różnicowym

Tranzystory bipolarne opisuje się wzorem (3.14) w uproszczonej postaci:

$$I_{C1} \approx \alpha_F I_{E0} \exp \frac{U_{BE1}}{V_T} \quad (5.61)$$

$$I_{C2} \approx \alpha_F I_{E0} \exp \frac{U_{BE2}}{V_T}$$

Napięcie wejściowe różnicowe wynosi:

$$U_{idiff} = U_{BE1} - U_{BE2} \quad (5.62)$$

Stosunek prądów tranzystorów można więc uzależnić od napięcia wejściowego wzmacniacza:

$$\frac{I_{C1}}{I_{C2}} = \frac{\alpha_F I_{E0} \exp \frac{U_{BE1}}{V_T}}{\alpha_F I_{E0} \exp \frac{U_{BE2}}{V_T}} = \exp \frac{U_{idiff}}{V_T} \quad (5.63)$$

uwzględniając ponadto stałość prądu zasilania:

$$I_{C1} + I_{C2} \approx I_{E1} + I_{E2} = I_{EE} \quad (5.64)$$

można zapisać:

$$I_{C1} = I_{EE} \frac{\exp \frac{U_{idiff}}{V_T}}{1 + \exp \frac{U_{idiff}}{V_T}} \quad (5.65)$$

$$I_{C2} = I_{EE} \frac{1}{1 + \exp \frac{U_{idiff}}{V_T}}$$

Korzystając ze wzoru definicyjnego funkcji tangens hiperboliczny, otrzymujemy:

$$\frac{1}{2}(1 + \operatorname{tgh} x) = \frac{e^{2x}}{e^{2x} + 1}$$

co pozwala zapisać prądy tranzystorów jako:

$$\begin{aligned} I_{C1} &= \frac{1}{2} I_{EE} \left(1 + \operatorname{tgh} \frac{U_{\text{idiff}}}{2V_T} \right) \\ I_{C2} &= \frac{1}{2} I_{EE} \left(1 - \operatorname{tgh} \frac{U_{\text{idiff}}}{2V_T} \right) \end{aligned} \quad (5.66)$$

Znając zależność prądów kolektorów od różnicowego sygnału wejściowego i korzystając z zależności (5.15) i (5.16), można określić napięcia wyjściowe wzmacniacza:

$$\begin{aligned} U_{o1} &= U_{CC} - \frac{1}{2} I_{EE} \left(1 + \operatorname{tgh} \frac{U_{\text{idiff}}}{2V_T} \right) R_C \\ U_{o2} &= U_{CC} - \frac{1}{2} I_{EE} \left(1 - \operatorname{tgh} \frac{U_{\text{idiff}}}{2V_T} \right) R_C \\ U_{o1-2} &= U_{o1} - U_{o2} = -I_{EE} R_C \operatorname{tgh} \frac{U_{\text{idiff}}}{2V_T} \end{aligned} \quad (5.67)$$

Zależności trzech napięć wyjściowych od wejściowego napięcia różnicowego, opisane wzorami (5.67), są to charakterystyki przenoszenia wzmacniacza różnicowego, pokazane na rys. 5.11. Jak wynika z wzorów i wykresów, charakterystyki przenoszenia wzmacniacza różnicowego są nieliniowe. Stosując konwencjonalne techniki linearyzacji można znaleźć, że jeżeli założy się $|U_{\text{idiff}}| < 0,5 V_T$ (około 13 mV przy 300 K), to tangens hiperboliczny we wzorach można zastąpić funkcją liniową z błędem nieprzekraczającym 2%. Otrzymuje się wtedy m.in. zależność:

$$U_{o1-2} \approx -\frac{I_{EE} R_C}{2V_T} U_{\text{idiff}} \quad (5.68)$$

wyrażającą, w jawnej formie, podstawowy parametr wzmacniacza różnicowego – współczynnik wzmocnienia napięciowego sygnałów różnicowych:

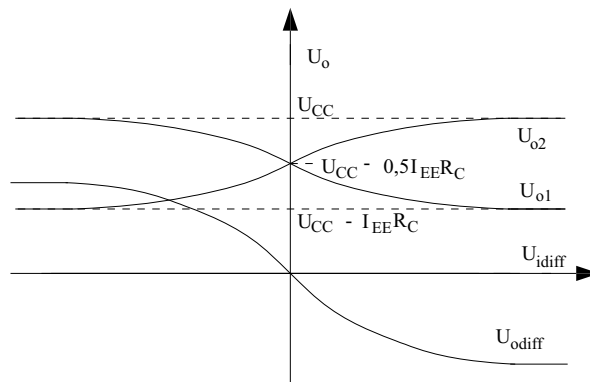
$$K_{\text{udiff}} \approx -\frac{I_{EE} R_C}{2V_T} \quad (5.69)$$

Wzór (5.68) wskazuje na interesującą cechę wzmacniacza różnicowego – napięcie wyjściowe jest proporcjonalne do iloczynu prądu zasilania i sygnału wejściowego. Jeżeli stałoprądowe źródło zasilające I_{EE} zastąpić źródłem prądowym sterowanym, to ten wzmacniacz umożliwi dokonywanie operacji mnożenia sygnałów.

Warunek przeprowadzenia linearyzacji wzorów (5.67) z określonym błędem (np. 2%) może służyć jako kryterium określające "małe sygnały" jako takie, przy których U_{idiff} jest mniejsze od określonej wartości (np. 13 mV).

Warto dodać, że nawet przy stosunkowo niewielkiej różnicy napięć sterujących obydwu tranzystory, U_{diff} około 0,1 V, jeden z nich zostaje zatkany, a cały prąd I_{EE} przepływa przez drugi tranzystor. W takim stanie wzmacniacz nie reaguje na małe sygnały, czyli przestaje wzmacniać. Ten stan nazywa się nasyceniem wzmacniacza, na charakterystyce przenoszenia odpowiada mu odcinek płaski.

Analizując pozostałe wzory (5.67) znajduje się, że w wartościach napięć na wyjściach asymetrycznych są dwie składowe, z których jedna zależy tylko od zasilania (U_{CC} , I_{EE}), a druga od zasilania i sygnału wejściowego. Tylko tę drugą składową traktuje się jako sygnał wyjściowy wzmacniacza, który w poprzednim punkcie był oznaczany jako ΔU_{01} i ΔU_{02} . Dokonując odpowiednich przekształceń, znajduje się, że na wyjściach asymetrycznych współczynniki wzmocnienia napięciowego dla sygnału wejściowego różnicowego są o połowę mniejsze od K_{diff} i mają różne znaki.



Rys. 5.11. Charakterystyki przenoszenia wzmacniacza różnicowego

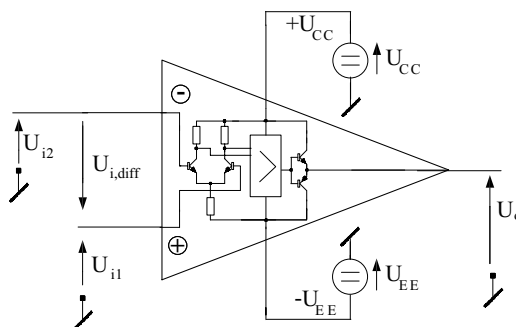
Dwa wejścia wzmacniacza różnicowego umożliwiają swoistą "manipulację" znakami sygnałów. Przy przyjętych oznaczeniach dodatniemu sygnałowi wejściowemu, różnicowemu odpowiada ujemny sygnał wyjściowy, różnicowy. Dodatni sygnał wejściowy U_{diff} uzyskuje się np. zwierając wejście 2 z masą ($U_{i2} = 0$) i podając dodatnie napięcie do wejścia 1 ($U_{i1} > 0$) – napięcie wyjściowe różnicowe będzie ujemne. Ale dodatni sygnał wejściowy U_{diff} można też uzyskać zwierając wejście 1 z masą ($U_{i1} = 0$) i podając ujemne napięcie do wejścia 2 ($U_{i2} < 0$) – napięcie wyjściowe różnicowe także będzie ujemne. W pierwszym przypadku wzmacniacz, oprócz wzmocnienia, powoduje też zmianę znaku sygnału i dlatego, działające w tej sytuacji, wejście 1 nazywa się odwracającym i oznacza się znakiem minus. W drugim przypadku znak sygnału nie zmienia się a działające w tej sytuacji wejście 2 nazywa się nieodwracającym i oznacza znakiem plus.

Na koniec trzeba dodać, iż przedstawione analizy są niekompletne, ponieważ nie uwzględniają efektów dynamicznych. Aby je uwzględnić, należałoby wprowadzić np. nieliniowe pojemności do modelu tranzystora. Przeprowadzane dokładniejsze analizy, często metodą symulacji komputerowych, potwierdzają intuicyjny wniosek, że wzmacniacz różni-

cowy dobrze wzmacnia sygnały wolnozmiennie oraz sygnały sinusoidalne o ograniczonych częstotliwościach, czyli jego charakterystyka częstotliwościowa jest typową charakterystyką wzmacniacza prądu stałego, jak pokazano na rys. 5.6.

5.4. Wzmacniacz operacyjny (WO)

Wzmacniacz operacyjny, którego nazwa wywodzi się od możliwości realizacji operacji matematycznych, zawiera stopnie wzmacniające w połączeniu kaskadowym. Stopniem wejściowym jest zwykle wzmacniacz różnicowy, stopniem końcowym jest wzmacniacz mocy. Całość jest wykonywana jako jeden układ scalony, wyposażony w stosunkowo rozbudowane podzespoły, łącznie zawierający kilkanaście do kilkudziesięciu tranzystorów, diod, rezystorów. Zasilanie wzmacniacza jest zorganizowane tak, jak w układzie różnicowym, to jest z dwu źródeł napięcia stałego.



Rys. 5.13. Wzmacniacz operacyjny

Wejścia wzmacniacza operacyjnego to wejścia pierwszego stopnia różnicowego, czyli dwa zaciski ("+" oraz "-"), do których można przyłączać źródła sygnałów, pamiętając o konieczności zapewnienia połączenia, pośredniego lub bezpośredniego, każdego z wejść z ogólną masą. Wyjście wzmacniacza jest zwykle asymetryczne, napięcie wyjściowe U_o jest określane względem masy. Na rys. 5.13 pokazano szkic przedstawiający symbol wzmacniacza, uzupełniony o uproszczone elementy schematu wewnętrznego i oznaczenia napięć. Producenci umieszczają w jednej obudowie jeden lub dwa, a niekiedy i cztery, wzmacniacze o wspólnym zasilaniu, pracujące niezależnie lub z ustalonym współdziałaniem.

5.4.1. Parametry wzmacniacza operacyjnego

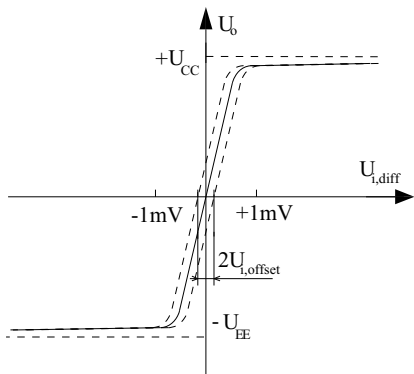
Parametry techniczne wzmacniacza obejmują kilkadziesiąt pozycji i charakterystyk. W dalszym tekście przedstawia się wybrane, ważniejsze parametry techniczne scalonego wzmacniacza operacyjnego.

Wzmocnienie napięciowe (różnicowe) wzmacniacza operacyjnego jest duże, wynosi kilkadziesiąt tysięcy do miliona V/V. Wynika to z przemnażenia współczynników wzmoc-

nienia kolejnych stopni (kilkadziesiąt do kilkuset) przy ich połączeniu kaskadowym. Warto zauważyć, że przy zasilaniu np. ± 12 V maksymalna wartość napięcia wyjściowego jest ok. ± 11 V, aby uzyskać taką wartość przy wzmacnieniu na przykład 100 000 V/V wystarczy doprowadzić do wejścia sygnał różnicowy o wartości rzędu 0,1 mV. Tak małe wartości napięcia często, w pierwszym przybliżeniu, traktuje się jako zero, co przy skończonej wartości napięcia wyjściowego, oznacza przyjęcie nieskończonego dużego wzmacnienia wzmacniacza operacyjnego.

Prąd wejściowy pierwszego stopnia jest bardzo mały, zwykle ma wartość poniżej jednego mikroampera, co pozwala w uproszczeniu traktować go jako wartość zerową. Innymi słowy oznacza to przyjęcie, że rezystancja wejściowa wzmacniacza operacyjnego jest nieskończenie wielka. Przy dokładniejszych obliczeniach trzeba uwzględnić prądy wejściowe opisane przez ich średnią wartość, zwaną prądem polaryzacji oraz różnicę, zwaną prądem niezrównoważenia i jej zmienność z temperaturą.

Charakterystyka przenoszenia wzmacniacza operacyjnego jest pokazana na rys. 5.14. Pracy liniowej odpowiada odcinek nachylony, odcinki poziome oznaczają wprowadzenie wzmacniacza w nasycenie. Przy wzmacnianiu sygnałów analogowych korzysta się z zakresu pracy liniowej, któremu odpowiada przedział wejściowych napięć różnicowych rzędu ułamka miliwolta. Aby zapewnić tak małe napięcia różnicowe wejściowe wzmacniacza operacyjnego, zwykle zaopatruje się go w silne ujemne sprzężenie zwrotne. Sygnał sprzężenia zwrotnego, odejmując się od sygnału wejściowego doprowadzonego z zewnątrz, częściowo kompensuje ten sygnał, tak aby pozostała różnica nie powodowała nasycenia wzmacniacza.



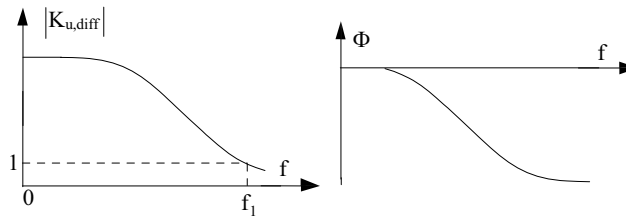
Rys. 5.14. Charakterystyka przenoszenia wzmacniacza operacyjnego

wzmacniacza. Napięcie to może być określone jako wyjściowe napięcie niezrównoważenia przez podanie możliwej wartości napięcia wyjściowego wzmacniacza przy zerowym napięciu wejściowym, ale częściej określa się wejściowe napięcie niezrównoważenia wzmacniacza (input offset) jako maksymalną wartość napięcia wejściowego, jaka może być niezbędna, aby napięcie wyjściowe sprowadzić do zera.

Charakterystyki częstotliwościowe: amplitudową i fazową, wzmacniacza operacyjnego pokazano na rys. 5.15. Warto zauważyć, że te charakterystyki dotyczą "gołego" wzmacniacza; z otwartą pętlą sprzężenia zwrotnego. Stosowane z reguły ujemne sprzężenie zwrotne modyfikuje silnie przebieg charakterystyk zespołu złożonego ze wzmacniacza połączonego z członem sprzężenia zwrotnego.

Dość często, ze względu na rozrzuty produkcyjne, charakterystyka przenoszenia wzmacniacza jest przesunięta na osi odciętych (nie przechodzi przez punkt 0,0), jak pokazano na rys. 5.14. W takich przypadkach określa się napięcie niezrównoważenia

Dość często, ze względu na rozrzuty produkcyjne, charakterystyka przenoszenia wzmacniacza jest przesunięta na osi odciętych (nie przechodzi przez punkt 0,0), jak pokazano na rys. 5.14. W takich przypadkach określa się napięcie niezrównoważenia



Rys. 5.15. Charakterystyki częstotliwościowe

Często stosowaną miarą właściwości dynamicznych wzmacniacza operacyjnego jest podawanie częstotliwości granicznej f_1 , przy której moduł współczynnika wzmocnienia napięciowego zmniejsza się do jedności.

5.4.2. Zastosowania liniowe WO

W niniejszym punkcie podaje się tylko pięć (z pośród setek) przykładów zastosowań wzmacniaczy operacyjnych w układach liniowych. Określenie "liniowy" oznacza tutaj, że w tych układach korzysta się z liniowego odcinka charakterystyki przenoszenia WO, stosując właściwe ujemne sprzężenie zwrotne. W analizie wszystkich przykładów zastosowano jednakowe uproszczenia przyjmując, że wzmacniacz operacyjny jest idealnym wzmacniaczem napięciowym o podstawowych parametrach:

- rezystancja wejściowa nieskończenie duża ($I_i = 0$ także przy $U_i \neq 0$),
- zerowa rezystancja wyjściowa,
- wzmocnienie napięciowe nieskończenie duże ($K_u = \infty$ zatem $U_{idiff} = 0$),
- całkowicie płaska charakterystyka częstotliwościowa (niezależna od f).

Przyjęte założenie nieskończenie dużego wzmocnienia napięciowego powoduje, że o właściwościach układu zawierającego wzmacniacz ze sprzężeniem zwrotnym, decyduje właśnie to sprzężenie a nie wzmacniacz (por. wzór 5.11), przytaczane poniżej wzory, opisujące analizowane przykłady, potwierdzają ten wniosek.

W praktyce projektowej, tak znaczne uproszczenie jak powyższa idealizacja wzmacniacza operacyjnego, może być stosowane wyłącznie na etapie wstępnych projektów koncepcyjnych. Projekty techniczne muszą uwzględniać rzeczywiste parametry rzeczywistych układów scalonych, co wymaga znacznie bardziej szczegółowego opisu tych układów.

Odmienny sposób wykorzystania scalonych wzmacniaczy operacyjnych, z wykorzystaniem ich właściwości nieliniowych, przedstawiono w następnym rozdziale, przy opisie układu z dodatnim sprzężeniem zwrotnym.

Wzmacniacz sygnału odwracający (znak)

Schemat układu pokazano na rys. 5.16, gdzie oznaczono też wielkości elektryczne. Sygnał wejściowy U_i (mierzony względem wspólnej masy) jest doprowadzony do wejścia "-" wzmacniacza operacyjnego przez rezystor R_1 . Wzmocniony sygnał wyjściowy U_o (też mierzony względem wspólnej masy, a "wytworzony" przez wzmacniacz), otrzymuje się

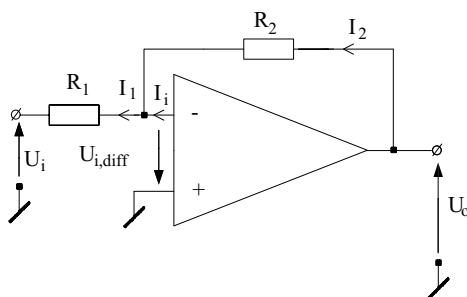
z wyjścia wzmacniacza operacyjnego. Ten sygnał może być doprowadzony do obciążenia wzmacniacza, co na rysunku oznaczono zaciskiem, a także jest wykorzystany do utworzenia ujemnego sprzężenia zwrotnego. Rezystory R_1 i R_2 tworzą czwórnik ujemnego sprzężenia zwrotnego. Równania opisujące dwa oczka układu i jeden węzeł mają postać:

$$\begin{aligned} U_i + I_1 R_1 + U_{\text{idiff}} &= 0 \\ U_o + I_2 R_2 + U_{\text{idiff}} &= 0 \\ I_1 &= I_1 + I_2 \end{aligned} \quad (5.70)$$

Zakładając dla uproszczenia, że $U_{\text{idiff}} = 0$ oraz $I_1 = 0$ otrzymuje się:

$$U_o = -\frac{R_2}{R_1} U_i \quad (5.71)$$

co oznacza że współczynnik wzmocnienia układu, tj. wzmacniacza operacyjnego z członem sprzężenia zwrotnego, wynosi $-R_2/R_1$. Znak minus oznacza "odwrócenie" sygnału tj. dodatniemu sygnałowi wejściowemu odpowiada ujemny sygnał wyjściowy i odwrotnie. Dobierając stosunek rezystancji R_2 do R_1 uzyskuje się potrzebne wzmocnienie napięciowe, większe lub mniejsze od jedności.



Rys. 5.16. Wzmacniacz odwracający

Rezystancja wejściowa układu może być wyznaczona z definicji, jako stosunek napięcia doprowadzanego do wejścia U_i do prądu pobieranego ze źródła sygnału czyli I_1 . W tym układzie jest to wartość rezystancji R_1 .

Dobór konkretnych wartości rezystancji jest ograniczony parametrami technicznymi rzeczywistego wzmacniacza. W bardzo zgrubnym podejściu projektowym można wskazać, że równania (5.70), oraz wnioski z nich wypływające, będą słuszne pomimo uproszczeń, gdy wszystkie napięcia: U_i , U_o , $I_1 R_1$, $I_2 R_2$ są znacznie większe od pominiętej wartości U_{idiff} , a prądy: I_1 , I_2 są znacznie większe od pominiętej wartości prądu wejściowego I_i . Przy dokładniejszym projektowaniu należy korzystać z bardziej szczegółowych analiz takich układów.

Wzmacniacz sygnału nieodwracający (zachowujący znak)

Na rys. 5.17 pokazano schemat wzmacniacza nieodwracającego, czyli zachowującego znak sygnału. Czwórnik sprzężenia zwrotnego w tym schemacie jest to dzielnik oporowy

utworzony z rezystorów R_1 i R_2 . Wydzielona część napięcia wyjściowego jest doprowadzona do wejścia "-" wzmacniacza operacyjnego, a sygnał wejściowy do wejścia "+", tak więc sygnał różnicowy U_{diff} , realnie oddziaływujący na wzmacniacz operacyjny, jest różnicą sygnału wejściowego i sygnału sprzężenia zwrotnego (sprzężenie ujemne). Przyjmując dla uproszczenia $U_{\text{diff}} = 0$, tym samym przyjmuje się, że napięcia wejść "+" i "-" są jednakowe, czyli:

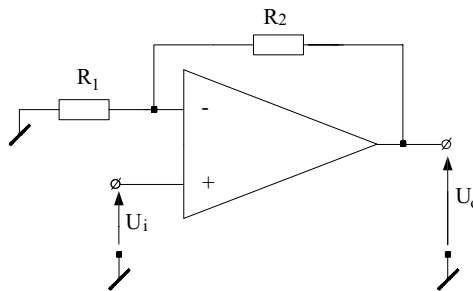
$$U_i = U_o \frac{R_1}{R_1 + R_2} \quad (5.72)$$

skąd otrzymuje się:

$$U_o = \left(1 + \frac{R_2}{R_1}\right) U_i \quad (5.73)$$

Współczynnik wzmocnienia wzmacniacza nieodwracającego, wynosi więc:

$$K_u = \frac{U_o}{U_i} = 1 + \frac{R_2}{R_1} \quad (5.74)$$



Rys. 5.17. Wzmacniacz sygnału nieodwracający

W rozpatrywanym przykładzie można uzyskać potrzebny współczynnik wzmocnienia (nie mniejszy od 1) dobierając odpowiednio stosunek wartości rezystancji R_1 i R_2 . W szczególności można zastosować zwarcie wyjścia z wejściem "-" i przerwę zamiast rezystora R_1 , otrzymując wzmacniacz o wzmocnieniu jeden, czyli wtórnik.

Rezystancja wejściowa, czyli stosunek napięcia do prądu wejściowego, przy przyjętych uproszczeniach, jest nieskończenie wielka.

Wzmacniacz różnicy sygnałów

Schemat układu pokazano na rys. 5.18. Zadaniem tego układu jest wytworzenie sygnału wyjściowego U_o , o wartości równej różnicy ważonej (tj. z ustalonymi współczynnikami przy odjemnej i odjemniku) dwu sygnałów wejściowych U_{i1} oraz U_{i2} . Równania opisujące oczka (dla idealnego WO) mają postać:

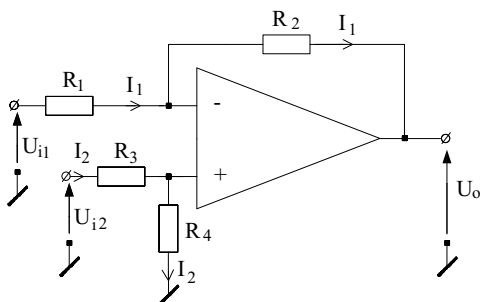
$$\begin{aligned} U_{i1} - I_1 R_1 - I_1 R_2 - U_o &= 0 \\ U_{i2} - I_2 R_3 - I_2 R_4 &= 0 \\ U_{i1} - I_1 R_1 - I_2 R_4 &= 0 \end{aligned} \quad (5.75)$$

skąd, po przekształceniach, otrzymuje się:

$$U_o = \frac{R_2}{R_1} \left(\frac{R_1 + R_2}{R_2} \frac{R_4}{R_3 + R_4} U_{i2} - U_{i1} \right) \quad (5.76)$$

a jeżeli zastosuje się rezystory $R_1 = R_3$ oraz $R_2 = R_4$, to powstaje wzmacniacz różnicy sygnałów:

$$U_o = \frac{R_2}{R_1} (U_{i2} - U_{i1}) \quad (5.77)$$

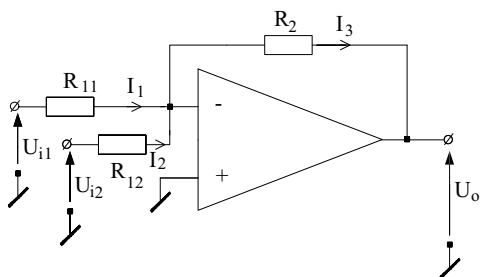


Rys. 5.18. Wzmacniacz różnicy sygnałów

Układ może być stosowany jako człon odejmujący dwa analogowe sygnały w obwodach automatyki, czyli spełniający funkcję tzw. "węzła sumującego" w układach automatycznej regulacji.

Sumator

Schemat układu pokazano na rys. 5.19. Zadaniem tego układu jest wytworzenie sygnału wyjściowego U_o , o wartości równej sumie ważonej (tj. z ustalonymi współczynnikami przy kolejnych składnikach) dwu sygnałów wejściowych.



Rys. 5.19. Wzmacniacz sumy sygnałów

Sumowaniu można też poddać więcej niż dwa składniki. Rozbudowa sumatora polega na zwiększaniu ilości rezystorów łączących wejście "-" wzmacniacza z kolejnymi źródłami sygnałów. Równania opisujące oczka (przy założeniu idealności WO) mają postać:

$$\begin{aligned} U_{i1} - I_1 R_{11} &= 0 \\ U_{i2} - I_2 R_{12} &= 0 \\ U_o + I_3 R_2 &= 0 \end{aligned} \quad (5.78)$$

a równaniem węzła jest:

$$I_1 + I_2 = I_3 \quad (5.79)$$

skąd otrzymuje się:

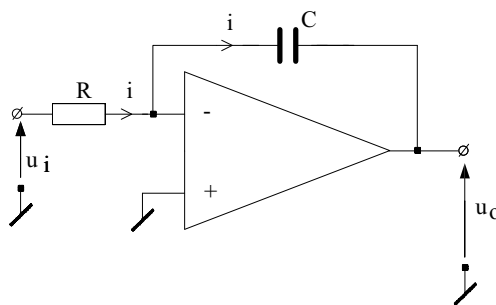
$$U_o = - \left(\frac{R_2}{R_{11}} U_{i1} + \frac{R_2}{R_{12}} U_{i2} \right) \quad (5.80)$$

Przy jednakowych rezystorach wejściowych $R_{11} = R_{12} = R_1$ układ wzmacnia sumę sygnałów, nadając jej odwrotny znak, zgodnie z zależnością:

$$U_o = - \frac{R_2}{R_1} (U_{i1} + U_{i2}) \quad (5.81)$$

Wzmacniacz całkujący – integrator

Schemat układu pokazano na rys. 5.20, gdzie oznaczono też wielkości elektryczne. Rezystor R i kondensator C tworzą człon ujemnego sprzężenia zwrotnego. Przez kondensator C sygnał z wyjścia wzmacniacza trafia do jego wejścia.



Rys. 5.20. Integrator

Formalnie opis działania układu otrzymuje się rozważając proces ładowania kondensatora C przy napięciu na nim $u_c(t)$. Prąd ładujący kondensator $i(t)$ jest wymuszony przez przyłożony sygnał wejściowy $u_i(t)$. Przy założeniu zerowej wartości napięcia pomiędzy wejściami wzmacniacza można opisać dwa oczka układu równaniami:

$$\begin{aligned} u_i(t) &= R i(t) \\ u_o(t) &= -u_c(t) \end{aligned} \quad (5.82)$$

pamiętając o podstawowej zależności dla kondensatora:

$$i(t) = C \frac{d}{dt} [u_c(t)] \quad (5.83)$$

wyprowadza się wzór:

$$u_o = -\frac{1}{RC} \int_0^t u_i(t) dt + u_o(0) \quad (5.84)$$

Przebieg sygnału wyjściowego integratora jest proporcjonalny do całki z sygnału wejściowego ze znakiem minus, z uwzględnieniem początkowej wartości napięcia na kondensatorze (oraz na wyjściu układu zgodnie z (5.82)).

Czytelnikowi pozostawia się ocenę ścisłości i prawidłowości tego zapisu oraz interpretację całkowania funkcji czasu $u(t)$, po czasie dt , przy zmiennej z czasem t granicy całkowania.

5.5. Wzmacniacze mocy

Wzmacniacze mocy są konstruowane dla dostarczenia odpowiednio dużej mocy do obciążenia. Wartość mocy wyjściowej jest zwykle podstawowym założeniem konstrukcyjnym, obok którego stawia się wymagania dotyczące sprawności, minimalnych zniekształceń sygnału i inne, natomiast wzmocnienie napięcia jest zwykle osiągane w stopniach wzmacniających poprzedzających stopień mocy.

Dla uzyskania dużej mocy wyjściowej niezbędne jest wytwarzanie dużych wartości napięcia i prądu wyjściowego. Rzutuje to na dobór tranzystorów, które muszą być odporne na przepływ dużych prądów i wytrzymujące wystąpienie znacznych napięć. Takie tranzystory nazywa się ogólnie tranzystorami mocy.

Praca tranzystora przy dużych wartościach napięcia i prądu wiąże się zwykle z dużymi stratami mocy elektrycznej, zamieniającej się na ciepło w tranzystorze. To z kolei komplikuje konstrukcję, ponieważ trzeba stosować radiatory (chłodnice) skutecznie odprowadzające wytworzone ciepło do otoczenia, przy zachowaniu ograniczonego przyrostu temperatury tranzystora.

Innym problemem technicznym jest to, że przy dużych zmianach prądów i napięć w tranzystorach wyraźnie objawiają się nieliniowości ich charakterystyk elektrycznych, co powoduje nieliniowość charakterystyki przenoszenia wzmacniacza i w efekcie prowadzi do zniekształceń wzmacnianych sygnałów.

Te dwa zadania, tj. zaprojektowanie wzmacniacza o małych stratach (dużej sprawności) i jednocześnie liniowej charakterystyce przenoszenia (małych zniekształceniach), prowadzą najczęściej do sprzecznych rozwiązań. Z tego powodu spotyka się wiele różnych układów, opartych na różnych koncepcjach, próbujących znaleźć rozwiązania kompromisowe.

Dla zilustrowania problematyki wzmacniaczy mocy rozpatruje się przykład wzmacniacza przeciwobnego, wykonanego z tranzystorów komplementarnych. Schemat tego wzmacniacza pokazano na rys. 5.21. Układ składa się z dwu bipolarnych tranzystorów mocy, komplementarnych, to znaczy o możliwie jednakowych właściwościach elektrycznych,

a różniących się budową, jeden typu pnp a drugi typu npn. Zasilanie wzmacniacza jest podwójne: $\pm U$ względem masy, wejście pomiędzy zwartymi bazami a masą, wyjście pomiędzy zwartymi emiterami a masą.

Napięcie pomiędzy bazą a emiterem jest jednakowe w obydwu tranzystorach, ale ze względu na odmienne budowy tylko jeden z nich może zostać wprowadzony w stan aktywny. Jeżeli np. napięcie U_{BE} jest dodatnie i przekracza wartość ok. 0,6 V, to tranzystor npn zacznie przewodzić, przy takim napięciu złącze baza – emiter tranzystora pnp będzie spolaryzowane wstecznie i ten tranzystor będzie zatkany. Przy napięciu U_{BE} mniejszym od ok. 0,6 V obydwa tranzystory są zatkane.

Przy przewodzącym na przykład tranzystorze npn, prąd ze źródła zasilania $+U$ przepływa przez ten tranzystor i obciążenie. Wartość prądu jest taka, aby spadek napięcia na obciążeniu, czyli napięcie wyjściowe U_o , spełniało równanie oczka:

$$U_i = U_{BE} + U_o \quad (5.85)$$

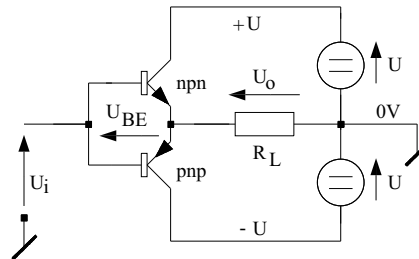
przy $U_i > U_{BE}$. Podobne równanie otrzymuje się zakładając, że sygnał wyjściowy jest ujemny i przewodzi tranzystor pnp. Te równania przedstawiają charakterystykę przenoszenia wzmacniacza $U_o(U_i)$, pokazaną na rys. 5.22.

Maksymalne napięcie wyjściowe otrzymuje się wprowadzając odpowiedni tranzystor w nasycenie, czemu odpowiada napięcie kolektor – emiter poniżej 1 V. W uproszczeniu można przyjąć, że maksymalne napięcie wyjściowe jest równe napięciu zasilania $U_{omax} = U$. Przy $U_i \gg U_{BE}$ równanie (5.85) można uprościć do postaci:

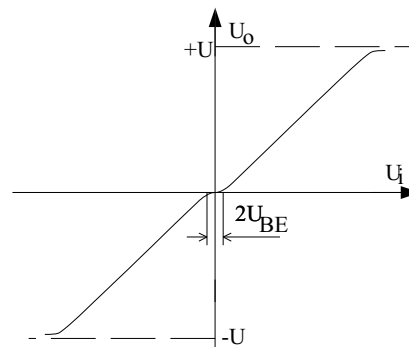
$$U_i \approx U_o \quad (5.86)$$

z której wynika, że ten wzmacniacz nie wzmacnia napięcia, (współczynnik wzmocnienia napięciowego jest w praktyce nieco mniejszy od jedności) ale prąd wejściowy jest prądem bazy a prąd wyjściowy prądem emitera, zatem następuje wzmocnienie prądu tylokrotnie, ile wynosi współczynnik wzmocnienia prądowego tranzystora $h_{21E} + 1$. Wzmocnienie mocy we wzmacniaczu można więc oszacować jako $K_p \approx h_{21E}$.

Maksymalna moc sygnału stałoprądowego jest wydzielana w obciążeniu wzmacniacza przy maksymalnym napięciu wyjściowym i wynosi w przybliżeniu:



Rys. 5.21. Przeciwsobny wzmacniacz mocy



Rys. 5.22. Charakterystyka przenoszenia wzmacniacza przeciwsobnego

$$P_{o\max} \approx \frac{U^2}{R_L} \quad (5.87)$$

Sprawność wzmacniacza jest stosunkiem mocy oddanej do obciążenia do mocy pobranej z zasilania i wynosi:

$$\eta \approx \frac{I_o U_o}{I_o U} = \frac{U_o}{U} \quad (5.88)$$

czyli jest zmienna, zależnie od stosunku wartości sygnału wyjściowego U_o do wartości napięcia zasilania U . Sprawność zbliża się do jedności dla sygnału bliskiego do maksymalnego (ten szacunek jest zbyt optymistyczny z powodu przyjętych założeń upraszczających) albo zmniejsza się do zera, gdy prąd wyjściowy maleje do zera.

Maksymalna moc strat w tranzystorze występuje przy sygnale wyjściowym równym połowie napięcia zasilania i wynosi:

$$P_{C\max} \approx \frac{(0,5U)^2}{R_L} = 0,25P_{o\max} \quad (5.89)$$

czyli każdy z zastosowanych we wzmacniaczu tranzystorów powinien być zdolny do pracy przy wydzielaniu w nim mocy strat zamienianych na ciepło, tylko czterokrotnie mniejszej od maksymalnej mocy oddawanej ze wzmacniacza.

Zarówno nieliniowość charakterystyki przenoszenia, jak i niekorzystnie duże straty w tranzystorach uzasadniają poszukiwania lepszych układów wzmacniaczy.

Konwencjonalna droga poprawy liniowości polega na zmianie warunków pracy tranzystorów np. przez zapewnienie pracy wyłącznie w stanie aktywnym (bez stanów zatkania lub nasycenia) ale prowadzi to do znacznego pogorszenia sprawności. Innym sposobem poprawy liniowości jest wykorzystanie właściwości ujemnego sprzężenia zwrotnego, wymaga to jednak rozbudowy układu o stopnie poprzedzające i elementy sprzężenia. Na takiej drodze konstruowane są scalone wzmacniacze mocy stosowane w urządzeniach elektroakustycznych.

Poprawę sprawności wzmacniaczy mocy nawet do ponad 95% uzyskuje się w układach impulsowych, ale kosztem komplikacji układowych i zawyżenia wymagań co do elementów.

Uwzględniając znaczne różnorodności rozwiązań wzmacniaczy mocy dzieli się je na tzw. klasy. Pomijając formalne kryteria podziału przyjmuje się, że:

- klasa A oznacza pracę tranzystora wyłącznie w stanie aktywnym,
- klasa B to dwa tranzystory, z których jeden jest w stanie aktywnym przy jednej biegunowości sygnału wejściowego, a drugi jest aktywny przy przeciwnej biegunowości sygnału, jak w opisanym powyżej układzie,
- klasa C to tranzystor (lub tranzystory) przewodzący prąd w postaci różnie odkształconych impulsów i współpracujący z obwodami rezonansowymi LC dla odtworzenia sinusoidalnych napięć i prądów,
- klasa D to tranzystory sterowane dwustanowo, zwykle z zastosowaniem modulacji szerokości impulsów, i współpracujące z odpowiednimi demodulatorami.

Bardziej szczegółowe przedstawienie problemów wzmacniaczy mocy można znaleźć w literaturze [9, 10, 14, 27].

6. GENERATORY ELEKTRONICZNE

Generatorem nazywa się układ elektroniczny wytwarzający określony sygnał elektryczny, tj. układ mający dwa zaciski (wyjściowe), pomiędzy którymi pojawia się napięcie $u(t)$, o znanym, zaplanowanym przez konstruktora, przebiegu. Generowany sygnał może mieć parametry ustalone lub nastawiane przez użytkownika, to drugie dotyczy najczęściej aparatury laboratoryjnej, ale nie tylko. Szczególne wymagania techniczne mogą określać np. stabilność częstotliwości generowanego sygnału okresowego, dokładność odtworzenia określonej funkcji matematycznej, wartość mocy oddawanej przez generator do obciążenia lub jeszcze inne parametry.

Wśród różnych klasyfikacji generatorów ważniejszy jest podział na generatory częstotliwości i mocy. W pierwszym przypadku generator spełnia funkcję wzorcowego zegara, narzucając rytm pracy komputera lub długość fali nadajnika radiowego. W drugim przypadku zadaniem generatora jest dostarczyć do obciążenia odpowiednio dużą moc przy określonej częstotliwości jak np. w kuchence mikrofalowej. W praktyce stosuje się różne układy generatorów, z których nieliczne będą opisane poniżej.

6.1. Sprzężenie zwrotne dodatnie

Znaczna większość generatorów jest budowana z wzmacniaczy zaopatrzonych w dodatnie sprzężenie zwrotne. Przypomnijmy, że analizując w p. 5.2.2 połączenie wzmacniacza K z członem sprzężenia zwrotnego B zapisano m.in. równanie (5.8): $U_{iW} = U_{iK} + U_{oB}$ oznaczające, że przy sprzężeniu ujemnym napięcie wejściowe wzmacniacza U_{iK} jest różnicą napięcia wejściowego układu U_{iW} i sygnału sprzężenia zwrotnego U_{oB} .

Jeżeli sprzężenie zwrotne jest dodatnie to te sygnały sumują się, jak pokazano na rys. 6.1, czyli obowiązuje:

$$U_{iW} = U_{iK} - U_{oB} \quad (6.1)$$

Prowadząc dalej analizę, podobnie jak w p. 5.2.2, dochodzi się do uproszczonego wzoru, podającego związek pomiędzy sygnałem wejściowym U_{iW} a sygnałem wyjściowym U_{oW} , dla wzmacniacza z dodatnim sprzężeniem zwrotnym:

$$U_{oW} \approx \frac{K_{u0K}}{1 - K_{u0K}B} U_{iW} \quad (6.2)$$

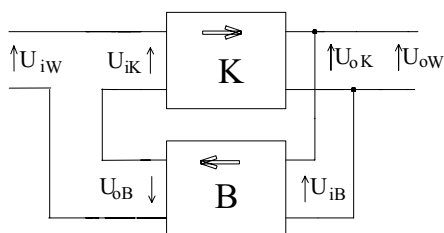
Interesujący, z punktu widzenia teorii generatorów, przypadek ma miejsce, gdy mianownik ułamka przyjmuje wartość zero, czyli gdy: $K_{u0K}B = 1$, tj. iloczyn rozwarciowego współczynnika wzmocnienia wzmacniacza K_{u0K} i współczynnika sprzężenia B jest równy jedności.

Taka sytuacja, niedopuszczalna w matematyce, może być łatwo zrealizowana technicznie, bo oznacza tylko dobór współczynnika dodatniego sprzężenia zwrotnego B o war-

tości równej $1/K_{u0K}$ (np. dla wzmacniacza o wzmocnieniu 50 V/V należy zastosować sprzężenie o współczynniku B równym 0,02). Pewną interpretację takiej sytuacji można uzyskać zapisując zależność (6.2) w postaci:

$$\frac{U_{oW}}{U_{iW}} \approx \frac{K_{u0K}}{1 - K_{u0K}B} \quad (6.3)$$

i odczytując ją następująco: Jeżeli $K_{u0K}B = 1$ to znaczy, że przy $U_{iW} = 0$ wystąpi skończona wartość U_{oW} , różna od zera. Innymi słowy: wzmacniacz z dodatnim sprzężeniem zwrotnym spełniającym warunek $K_{u0K}B = 1$ wytwarza sygnał wyjściowy bez sygnału wejściowego, czyli staje się generatorem.



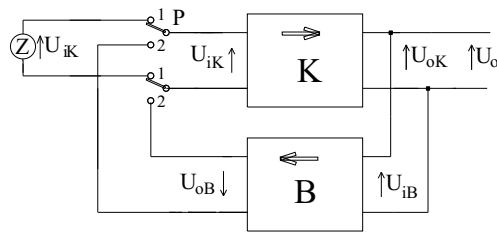
Rys. 6.1. Wzmacniacz z dodatnim sprzężeniem zwrotnym

Podobną interpretację warunku generacji $K_{u0K}B = 1$ można uzyskać przeprowadzając eksperyment myślowy w układzie, którego schemat pokazano na rys. 6.2. Jeżeli przełącznik P jest ustawiony w pozycji 1, to obwód sprzężenia zwrotnego jest rozarty, do wejścia wzmacniacza trafia napięcie U_{iK} pochodzące z obcego, pomocnego tylko w tym eksperymencie, źródła Z.

Wzmacniacz wzmacnia sygnał wejściowy tak, że na jego wyjściu otrzymuje się: $U_{oK} = K_{u0K} U_{iK}$. Wzmocniony sygnał trafia do członu B i, po przejściu przez ten człon, przyjmuje wartość: $U_{oB} = B U_{oK} = B K_{u0K} U_{iK}$, które będzie równe U_{iK} wówczas, gdy $K_{u0K} B = 1$, wtedy z wyjścia czwórnika B otrzymuje się takie samo napięcie jak z pomocniczego źródła Z.

Wystarczy teraz wyobrazić sobie błyskawiczne przerzucenie przełącznika w pozycję 2, co spowoduje zamknięcie obwodu sprzężenia zwrotnego, ale nie zmieni warunków pracy wzmacniacza K, który nadal otrzymuje sygnał wejściowy U_{iK} tylko już nie z pomocniczego źródła Z, ale z toru sprzężenia zwrotnego. Jeżeli nie zmienia się sygnał wejściowy, to i sygnał wyjściowy wzmacniacza pozostanie bez zmian, pomimo odrzucenia źródła Z. Wzmacniacz ze sprzężeniem zwrotnym dodatnim, które spełnia warunek formalny $B K_{u0K} = 1$ wytwarza więc sygnał wyjściowy, bez pobudzenia wejścia z zewnętrznego źródła, czyli staje się generatorem.

W układach praktycznych rolę pomocniczego źródła, pobudzającego generację, często spełniają szумы i zakłócenia inicjujące tzw. wzbudzenie generatora, podtrzymywanie generacji następuje dzięki działaniu obwodu dodatniego sprzężenia zwrotnego.



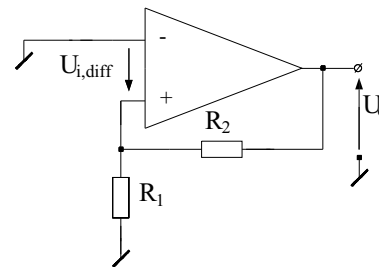
Rys. 6.2. Model wzbudzenia generatora

Na koniec warto odnotować, że dokładniejszy opis matematyczny generatorów wymaga przeprowadzania analizy układów nieliniowych. Dlatego powyższe uwagi, oparte na opisach (modelach) czwórnika liniowego, nie mogą być kompletne i ścisłe. W szczególności należy odnotować, że wielokrotnie przytaczany powyżej warunek $B K_{u0K} = 1$ jest tylko przypadkiem szczególnym obowiązującego ogólnie warunku generacji:

$$K_{u0K} B \geq 1 \quad (6.4)$$

Przykład realizacji praktycznej dodatniego sprzężenia zwrotnego pokazano na rys. 6.3. Taki układ stanowi istotną część tzw. przerzutnika Schmitta i wielu układów pochodnych.

Napięcie wejściowe różnicowe wzmacniacza operacyjnego jest tu równe spadkowi napięcia na rezystorze R_1 , umieszczonym w dzielniku napięcia R_1/R_2 . W przypadku idealizowanym można by założyć, że napięcie wyjściowe wzmacniacza jest równe zero, zatem i napięcie na rezystorze R_1 jest zerowe czyli i napięcie $U_{i,diff}$ jest zero, co potwierdzałoby założenie zerowego napięcia wyjściowego. Jednak w rzeczywistości zawsze wystąpi na wyjściu napięcie nierównoważenia wzmacniacza (niedoskonałość technologiczna) lub napięcie zakłóceń (wpływ otoczenia) czy szumu.



Rys. 6.3. Dodatnie sprzężenie zwrotne

Bardziej realnym jest więc założenie, że na wyjściu układu pojawia się przypadkowe, niewielkie napięcie, np. dodatnie względem masy, o pewnej wartości U . To napięcie, doprowadzone do dzielnika oporowego, powoduje, że na wejściu "+" wzmacniacza pojawi się napięcie $UR_1/(R_1 + R_2)$, też dodatnie. Napięcie wejściowe zostaje wzmacnione, zatem na wyjściu wzmacniacza wzrasta napięcie dodatnie, które poprzez dzielnik oddziałuje na jego wejście. W ten sposób można sobie wyobrazić przebieg lawinowy narastania napięcia wyjściowego, zainicjowany dowolnie małym, przypadkowym napięciem jakie pojawiło się w dowolnym punkcie układu.

Taki lawinowy proces doprowadza do nasycenia wzmacniacza tj. do stanu, w którym na wyjściu pojawia się największe, możliwe technicznie, napięcie. To napięcie, zwane napięciem nasycenia U_{osat} , jest zwykle o ok. pół wolta mniejsze od napięcia zasilania.

Układ wzmacniacza z dodatnim sprzężeniem zwrotnym, przedstawiony na rys. 6.3, ma więc taką cechę, że sam wprowadza się w stan nasycenia, z maksymalnym, stałym napięciem wyjściowym. Znak napięcia wyjściowego jest przypadkowy, zależny od wielu czynników działających w momencie załączenia układu. Ten układ można formalnie potraktować jako generator napięcia stałego o wartości sygnału $+U_{\text{osat}}$ lub $-U_{\text{osat}}$. Warto zauważyć, że napięcie wejściowe wzmacniacza operacyjnego będzie wynosić:

$$U_{\text{idiff}} = \pm U_{\text{osat}} \frac{R_1}{R_1 + R_2} \quad (6.5)$$

i jest zależne od doboru rezystancji dzielnika. W odróżnieniu od układów z ujemnym sprzężeniem zwrotnym, gdzie działanie sprzężenia zwrotnego sprowadzało wartość U_{idiff} praktycznie do zera, w układach z dodatnim sprzężeniem zwrotnym, to napięcie może przyjmować względnie duże wartości i powinno być uwzględniane w analizach układów.

Jeżeli zadaniem generatora jest wytworzenie sygnału okresowego, to konieczne jest wprowadzenie do układu elementów dynamicznych tj. takich, których działanie narzuci określone zmiany napięcia czy prądu powtarzające się cyklicznie (okresowo) w czasie, czyli z określoną częstotliwością.

Dobór i sposób działania elementów dynamicznych zależy odżądanego kształtu generowanego sygnału. Ogólnie dzieli się generatory wg kształtu sygnału na sinusoidalne i impulsowe. W tym drugim przypadku zwykle konieczne są uściślenia jak np. impulsów prostokątnych, trójkątnych, piłokształtnych itd.

Elementami narzucającymi kształt i częstotliwość sygnałów sinusoidalnych są obwody rezonansowe LC albo filtry RC. Wykorzystanie obwodów rezonansowych wynika z tego, że opisuje się je równaniem różniczkowym drugiego rzędu, którego rozwiązanie (przebieg napięcia lub prądu) może zawierać funkcję sinus lub cosinus, jeżeli spełnione są odpowiednie warunki. Zasady wykorzystania filtrów RC są mniej oczywiste i będą przedstawione w dalszej części.

W generatorach impulsowych najczęściej rolę elementu narzucającego przebieg czasowy spełnia kondensator ładowany prądem jednokierunkowym ze źródła napięcia lub prądu. Skutkiem ładowania kondensatora jest stopniowe narastanie jego napięcia, proporcjonalnie do gromadzonego ładunku. To narastające w czasie napięcie jest wykorzystywane w układach generatorów impulsowych do opóźniania zjawisk elektrycznych tak, aby uzyskać okresowo powtarzające się efekty

Oczywiście w generatorze, niezależnie od kształtu sygnału, stosuje się z reguły dodatnie sprzężenie zwrotne jako nadrzędny warunek generacji.

6.2. Generacja sygnałów sinusoidalnych

W elektronice jest znanych i stosowanych wiele różnych układów generatorów sygnałów sinusoidalnych. W ramach tego skryptu ograniczamy się do przytoczenia tylko jednego przykładu generatora LC i jednego przykładu generatora RC, w literaturze przedmiotu, np. [2, 9], można znaleźć wiele innych rozwiązań.

6.2.1. Generator sygnału sinusoidalnego LC

Na rys. 6.4 pokazano schemat połączeń generatora zbudowanego na bazie wzmacniacza operacyjnego i obwodu rezonansowego LC.

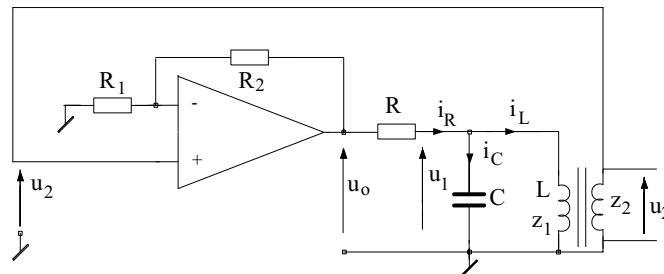
Wzmacniacz pracuje w układzie wzmacniającym, bez odwracania znaku sygnału, ze współczynnikiem wzmocnienia ustalonym rezystorami R_1 i R_2 na wartość:

$$K_u = 1 + \frac{R_2}{R_1} \quad (6.6)$$

Napięcie wyjściowe wzmacniacza, które jest jednocześnie napięciem wyjściowym generatora, wynosi więc:

$$u_o = K_u u_2 \quad (6.7)$$

Napięcie wejściowe wzmacniacza u_2 , jak i w innych generatorach, otrzymuje się z toru dodatniego sprzężenia zwrotnego.



Rys. 6.4. Generator w układzie Meissnera

Obwód rezonansowy jest zbudowany z kondensatora o pojemności C i cewki o indukcyjności L . Oprócz uzwojenia podstawowego, o ilości zwojów z_1 , na tym samym obwodzie magnetycznym wykonane jest drugie uzwojenie, o ilości zwojów z_2 , tak więc ten element jest właściwie transformatorem. Drugie uzwojenie służy do wykonania obwodu dodatniego sprzężenia zwrotnego. Napięcie u_2 , powstające na tym uzwojeniu wynosi:

$$u_2 = u_1 \frac{z_2}{z_1} = u_1 B \quad (6.8)$$

napięcie u_1 występuje na obwodzie rezonansowym. Literą B oznaczono przekładnię zwojową, która w tym przypadku jest też współczynnikiem sprzężenia. Dodatkowy rezystor R odtłumia obwód rezonansowy, czyli umożliwia powstawanie sinusoidalnych drgań napięcia na kondensatorze i cewce, w pewnym stopniu niezależnych od napięcia wyjściowego wzmacniacza.

Dla węzła w obwodzie rezonansowym zapisuje się równanie prądów w postaci: $i_R = i_C + i_L$, gdzie poszczególne prądy wynoszą:

$$i_R = \frac{u_o - u_1}{R}; \quad i_C = C \frac{du_1}{dt}; \quad i_L = \frac{1}{L} \int u_1 dt \quad (6.9)$$

Podstawiając te prądy do równania węzła i przekształcając je z uwzględnieniem parametrów K_u oraz B otrzymuje się równanie:

$$\frac{d^2 u_o}{dt^2} + \frac{1 - K_u B}{RC} \frac{du_o}{dt} + \frac{1}{LC} u_o = 0 \quad (6.10)$$

Wprowadzając pomocnicze parametry:

$$\xi = \frac{1 - K_u B}{2RC}, \quad \omega^2 = \frac{1}{LC} \quad (6.11)$$

przekształca się równanie (6.10) do postaci:

$$\frac{d^2 u_o}{dt^2} + 2\xi \frac{du_o}{dt} + \omega^2 u_o = 0 \quad (6.12)$$

Rozwiązaniem tego równania jest funkcja $u_o(t)$:

$$u_o = U_o \exp(-\xi t) \cos\left(\sqrt{\omega^2 - \xi^2} t\right) \quad (6.13)$$

wartość $U_o = u_o(0)$ oznacza warunek początkowy, czyli napięcie na wyjściu wzmacniacza w chwili $t = 0$. Dla przypadku $\xi = 0$ rozwiązanie upraszcza się do postaci:

$$u_o = U_o \cos(\omega t) \quad (6.14)$$

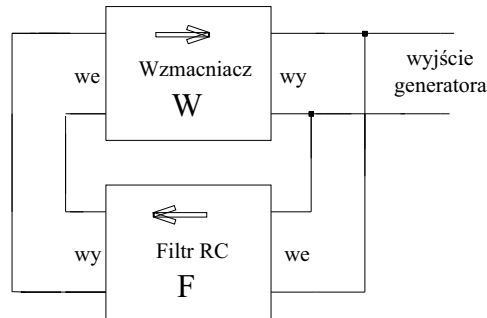
czyli przy spełnionym warunku generacji $K_u B = 1$, na wyjściu wzmacniacza, które jest jednocześnie wyjściem generatora, występuje napięcie kosinusoidalne, o częstotliwości własnej obwodu rezonansowego i amplitudzie zależnej od warunku początkowego.

W układzie praktycznym doprowadza się do tego, aby iloczyn $K_u B$ był nieco większy od jedności, wówczas funkcja $u_o(t)$ przedstawia "sinusoidę" o narastającej z czasem amplitudzie, co powoduje, że przy dowolnie małym, przypadkowym napięciu U_o nastąpi jednak wzbudzenie generatora do dużych wartości napięcia wyjściowego. Aby ustalić wartość amplitudy generowanego napięcia wykorzystuje się naturalną, lub sztucznie powiększoną, nieliniowość charakterystyki przenoszenia. Ponieważ wzrości wartości napięcia towarzyszy zmniejszanie się współczynnika wzmocnienia, więc początkowa nadwyżka $K_u B$ ponad jedność, występująca przy małej wartości generowanego napięcia, maleje ze wzrostem amplitudy aż do wartości granicznej $K_u B = 1$, której odpowiada określona wartość generowanego napięcia.

6.2.2. Generator sygnału sinusoidalnego RC

Ze względu na znacznie prostsze wytwarzanie rezystorów, w porównaniu do technologii cewek indukcyjnych, szczególnie przy generacji sygnałów o niezbyt wysokich częstotliwościach, częściej stosuje się generatory, zwane skrótowo "generatorami RC". Ścisłej tę

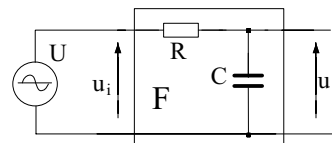
odmianę generatorów należy nazywać "generator z dodatnim sprzężeniem zwrotnym poprzez filtr oporowo-pojemnościowy".



Rys. 6.5. Schemat blokowy generatorów RC

Schemat blokowy generatora RC pokazano na rys. 6.5, składa się on ze wzmacniacza W i filtra F połączonych w układ sprzężony czyli wyjście wzmacniacza jest połączone z wejściem filtra a wyjście filtra jest połączone z wejściem wzmacniacza.

Dla wyjaśnienia działania generatora z filtrem RC należy, choćby pobieżnie, naszkicować zarys analizy filtrów, które zostały wstępnie wzmiankowane w rozdziale 4. Rozpatrzmy układ pokazany na rys. 6.6, przedstawiającym schemat najprostszego filtra RC, do wejścia którego dołączono źródło napięcia sinusoidalnego U.



Rys. 6.6. Filtr RC

Sam filtr stanowi dzielnik napięcia, wartość napięcia wyjściowego u_o można przedstawić jako wynik podziału napięcia wejściowego u_i stosownie do impedancji części filtru:

$$u_o = u_i \frac{Z_2}{Z_1 + Z_2} \quad (6.15)$$

gdzie impedancje wynoszą:

$$Z_1 = R, \quad Z_2 = \frac{1}{j\omega C} \quad (6.16)$$

Po przekształceniach otrzymuje się:

$$\frac{u_o}{u_i} = \frac{1}{\sqrt{(\omega RC)^2 + 1}} \exp[-j(\arctg \omega RC)] \quad (6.17)$$

Zapisując sinusoidalne sygnały: wejściowy i wyjściowy w postaci zespolonej, ich stosunek wyraża się przez:

$$\frac{u_o}{u_i} = \frac{U_o \exp[-j(\omega t + \phi)]}{U_i \exp(j\omega t)} = \frac{U_o}{U_i} \exp(j\phi) \quad (6.18)$$

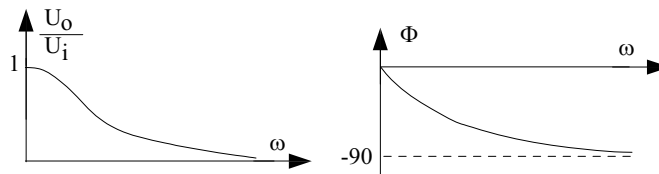
Obydwa powyższe wzory można zinterpretować następująco: jeżeli do wejścia filtra o schemacie pokazanym na rys. 6.6, doprowadza się napięcie sinusoidalne o amplitudzie U_i oraz częstotliwości ω , to na wyjściu tego filtra otrzymuje się napięcie sinusoidalne, też o częstotliwości ω , o amplitudzie U_o określonej zależnością:

$$U_o = \frac{U_i}{\sqrt{(\omega RC)^2 + 1}} \quad (6.19)$$

a ponadto przesunięte w fazie względem napięcia wejściowego o kąt ϕ :

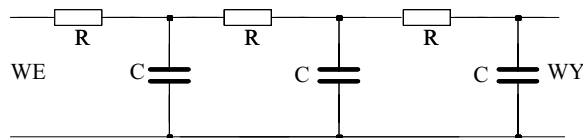
$$\phi = -\arctg \omega RC \quad (6.20)$$

Ponieważ w obydwu powyższych wzorach występuje częstotliwość ω , to dla różnych wartości częstotliwości stosunek amplitud oraz przesunięcie fazowe będą różne. Zależności tych wielkości od ω , pokazane na rys. 6.7 są to właśnie charakterystyki częstotliwościowe, amplitudowa oraz fazowa, analizowanego filtra.



Rys. 6.7. Charakterystyki filtra RC

Tak prosty filtr, jak pokazany na rys. 6.6, nie nadaje się jeszcze do budowy generatora. Podobny filtr, zwany drabinkowym filtrem RC, którego schemat pokazano na rys. 6.8, umożliwia już wykonanie generatora. Niestety wzór opisujący jego transmitancję widmową jest bardziej rozbudowany aniżeli wzór (6.17), aczkolwiek jego sens pozostaje taki sam jak w przypadku opisu pojedynczego członu RC.



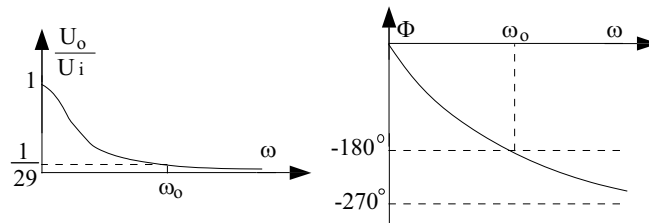
Rys. 6.8. Filtr drabinkowy

Charakterystyki filtra drabinkowego przedstawiono na rys. 6.9. Ważną cechą jest to, że dla pewnej częstotliwości ω_0

$$\omega_0 = \frac{1}{\sqrt{6RC}} \approx \frac{0,41}{RC} \quad (6.21)$$

przesunięcie fazowe tego filtra wynosi -180 stopni elektrycznych, tj. sinusoida napięcia wyjściowego o częstotliwości ω_0 jest przesunięta względem sinusoidy wejściowej o 180

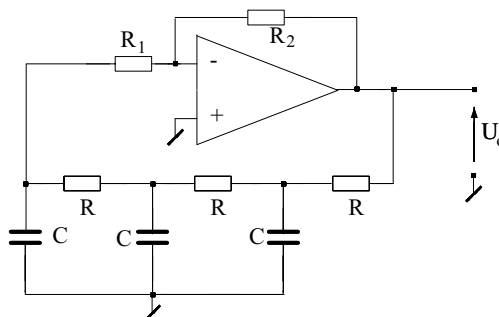
stopni, czyli jest "odwrócona". Dla częstotliwości ω_0 tłumienie amplitudy w tym filtrze wynosi 29, tj. amplituda sinusoidy wyjściowej jest 29 razy mniejsza od amplitudy sinusoidy wejściowej (skutek dzielenia napięcia aż w trzech dzielnikach napięcia).



Rys. 6.9. Charakterystyki filtru drabinkowego

Generator napięcia sinusoidalnego RC powstaje, jeżeli filtr drabinkowy RC połączy się ze wzmacniaczem odwracającym znak sygnału i mającym wzmocnienie 29 V/V, jak pokazano na rys. 6.10.

Dowolny sygnał sinusoidalny będzie odwrócony (przesunięty fazowo o 180 stopni) we wzmacniaczu, ale tylko sygnał sinusoidalny o częstotliwości ω_0 będzie też przesunięty o 180 stopni w filtrze. Zatem tylko sygnał o tej częstotliwości, po dwukrotnym odwróceniu, trafi z powrotem do wejście wzmacniacza bez przesunięcia fazowego. Dwudziestodziewięciokrotny ubytek amplitudy tego sygnału w filtrze będzie skompensowany 29-krotnym wzmocnieniem we wzmacniaczu i dzięki temu tylko dla sygnału o częstotliwości ω_0 będzie spełniony warunek generacji $K_u B = 1$.



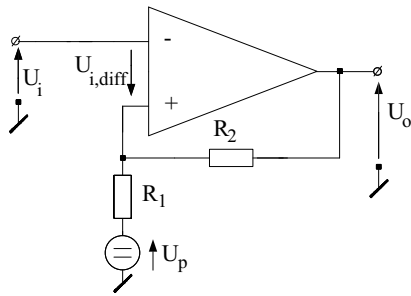
Rys. 6.10. Generator z filtrem drabinkowym

Także i w tym przypadku, dla ułatwienia wzbudzenia się generatora, należy tak uformować charakterystykę przenoszenia wzmacniacza, aby jego wzmocnienie było nieco większe od 29 V/V dla sygnałów o małych amplitudach i wynosiło 29 V/V dla wymaganej amplitudy generowanego sygnału.

6.3. Generacja sygnałów impulsowych

6.3.1. Przerzutnik Schmitta

Przerzutnik Schmitta (zob. p. 6.1 i rys. 6.3) sam nie jest generatorem sygnału okresowego, jednak stosowany jest często jako istotna część generatorów impulsowych. Na rys. 6.11 pokazano schemat jednego z wariantów układowych tego przerzutnika.



Rys. 6.11. Przerzutnik Schmitta

Rezystory R_1 i R_2 stanowią dodatnie sprzężenie zwrotne, zapewniające wprowadzenie wzmacniacza w stan nasycenia, czyli zapewniające, że napięcie wyjściowe przyjmuje tylko wartości $+U_{sat}$ lub $-U_{sat}$. Dodatkowe źródło napięcia stałego U_p służy do przesuwania charakterystyki przerzutnika.

Dla uproszczenia analizy przyjmuje się, że jeżeli napięcie wejściowe różnicowe wzmacniacza U_{idiff} jest dodatnie, to wzmacniacz jest nasycony z dodatnim napięciem wyjściowym $+U_{sat}$, a jeżeli napięcie wejściowe U_{idiff} jest ujemne, to wzmacniacz jest nasycony z ujemnym napięciem wyjściowym

$-U_{sat}$. Wartość napięcia U_{idiff} jest różnicą napięć obydwu wejść wzmacniacza:

$$U_{idiff} = U_{i+} - U_{i-} \quad (6.22)$$

Symbolami U_{i+} ; U_{i-} oznaczono napięcia pomiędzy odpowiednio wejściem "+" albo "-" wzmacniacza a wspólną masą elektryczną całego układu.

Napięcie U_{i-} jest równe napięciu wejściowemu przerzutnika U_i , doprowadzonemu do układu ze źródła sygnału wejściowego.

Napięcie U_{i+} pochodzi z toru sprzężenia zwrotnego i jest określone przez wartość napięcia nasycenia wzmacniacza, napięcie pomocnicze U_p oraz rezystancyjny dzielnik napięcia. Wartość tego napięcia wynosi:

$$U_{i+} = U_o \frac{R_1}{R_1 + R_2} + U_p \frac{R_2}{R_1 + R_2} \quad (6.23)$$

Zakładając na przykład, że wzmacniacz jest w nasyceniu dodatnim, tj. $U_o = +U_{sat}$, zakładamy tym samym, że $U_{idiff} > 0$. Ponieważ napięcie różnicowe zależy też od napięcia wejściowego U_i , to trzeba sprawdzić, przy jakim napięciu wejściowym U_i będzie $U_{idiff} > 0$. Na podstawie przytoczonych wyżej wzorów można znaleźć, że taki stan wystąpi jeżeli:

$$U_i < U_{sat} \frac{R_1}{R_1 + R_2} + U_p \frac{R_2}{R_1 + R_2} \quad (6.24)$$

Podobnie można wyznaczyć warunek wystąpienia ujemnego nasycenia wzmacniacza:

$$U_i > -U_{sat} \frac{R_1}{R_1 + R_2} + U_p \frac{R_2}{R_1 + R_2} \quad (6.25)$$

Na rys. 6.12 pokazano zależności pomiędzy napięciami: wyjściowym U_o i wejściowym U_i , opisane wzorami (6.24) i (6.25), czyli tak zwaną charakterystykę przejściową przerzutnika Schmitta. Ta charakterystyka nie jest funkcją (w sensie matematycznym) ponieważ ma przedział niejednoznaczności zwany pętlą histerezy.

Zewnętrzny sygnał wejściowy, o wartościach przekraczających wyznaczone powyżej granice tj. $U_i > U_1$ lub $U_i < U_2$, wymusza określony znak wyjściowego napięcia nasycenia, tym samym ustawiając przerzutnik jednoznacznie w określony stan. W przedziale środkowym, $U_1 > U_i > U_2$, wewnątrz pętli histerezy, znak napięcia wyjściowego nie zależy od aktualnej wartości sygnału wejściowego, ale pozostaje taki, jaki był wymuszony poprzednio. Szerokość pętli histerezy wynosi:

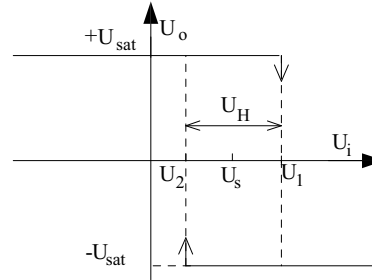
$$U_H = U_1 - U_2 = +U_{\text{sat}} \frac{2R_1}{R_1 + R_2} \quad (6.26)$$

a położenie jej środka:

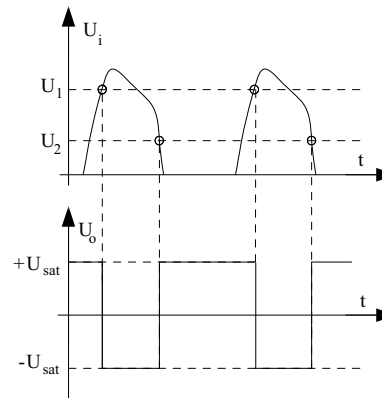
$$U_S = \frac{U_1 + U_2}{2} = U_p \frac{R_2}{R_1 + R_2} \quad (6.27)$$

Zarówno szerokość pętli histerezy, jak i jej przesunięcie wzdłuż osi U_i można dobierać niezależnie, stosownie do potrzeb, operując podziałem w dzielniku oporowym oraz wartością dodatkowego źródła polaryzacji U_p .

Przykład zastosowania przerzutnika Schmitta do przeformowywania impulsów "rozmytych" w prostokątne pokazano na rys. 6.13. Takie zabiegi stosowane są szczególnie często przy współpracy podzespołów analogowych z podzespołami cyfrowymi. Niekiedy nawet nazywa się przerzutnik Schmitta jednobitowym przetwornikiem analogowo-cyfrowym.



Rys. 6.12. Charakterystyka przerzutnika Schmitta



Rys. 6.13. Działanie przerzutnika Schmitta jako układu formującego impulsy prostokątne

6.3.2. Multiwibrator astabilny

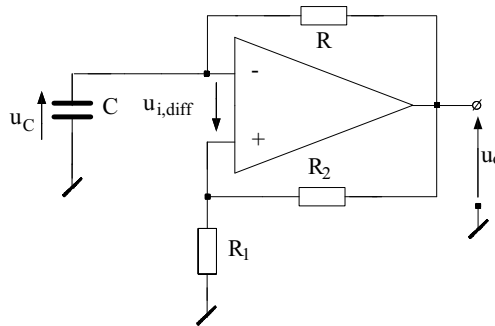
Multiwibrator astabilny jest generatorem ciągu impulsów prostokątnych o określonej częstotliwości i amplitudzie. Schemat tego generatora, wykorzystujący przerzutnik Schmitta, pokazano na rys. 6.14.

Aby wyjaśnić działanie układu założmy, że po włączeniu kondensator C jest nienaładowany ($u_C = 0$) przez co napięcie wejścia "-" względem masy wynosi zero. Działanie

dodatniego sprzężenia zwrotnego – rezystory R_1 i R_2 – powoduje, że wzmacniacz wprowadza się w nasycenie i napięcie na jego wyjściu wynosi np. $+U_{\text{sat}}$, zatem napięcie na wejściu "+" względem masy wynosi $+U_{\text{sat}}R_1/(R_1 + R_2)$. Pod wpływem napięcia wyjściowego $+U_{\text{sat}}$ następuje ładowanie kondensatora C przez rezystor R . Napięcie na kondensatorze narasta wykładniczo: zdążając do wartości $+U_{\text{sat}}$.

$$u_C(t) = +U_{\text{sat}} \left[1 - \exp\left(-\frac{t}{RC}\right) \right] \quad (6.28)$$

Zmiana napięcia na kondensatorze, od przyjętej wartości zero, w kierunku wartości $+U_{\text{sat}}$, powoduje, że w pewnym momencie napięcia na wejściach "-" oraz "+" wzmacniacza przyjmują takie same wartości. W takiej sytuacji wejściowe napięcie różnicowe maleje do zera, a następnie narastałoby z przeciwnym znakiem. Ale zmniejszenie napięcia $u_{i,\text{diff}}$ do wartości bliskiej zeru powoduje wyprowadzenie wzmacniacza ze stanu nasycenia, w związku z czym jego napięcie wyjściowe zmniejsza się.



Rys. 6.14. Multiwibrator astabilny

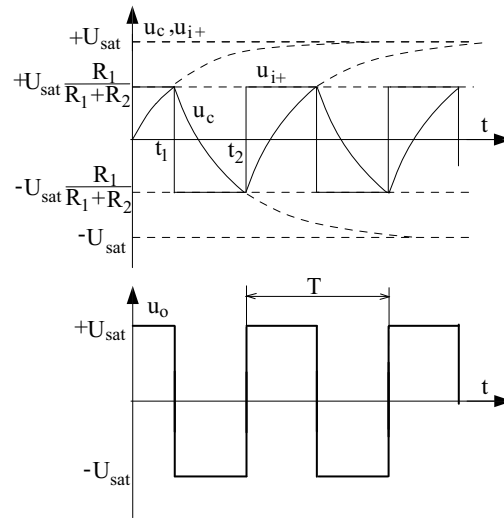
Przyjmując dla uproszczenia, że wyprowadzanie wzmacniacza ze stanu nasycenia rozpoczyna się przy zerowym napięciu różnicowym w chwili t_1 , można wyznaczyć ten moment z warunku:

$$+U_{\text{sat}} \frac{R_1}{R_1 + R_2} = +U_{\text{sat}} \left[1 - \exp\left(-\frac{t_1}{RC}\right) \right] \quad (6.29)$$

skąd:

$$t_1 = RC \ln\left(\frac{R_1 + R_2}{R_2}\right) \quad (6.30)$$

Zmniejszenie napięcia wyjściowego poniżej wartości $+U_{\text{sat}}$, poprzez działanie dzielnika oporowego powoduje bezzwłoczne obniżenie napięcia na wejściu "+". Ponieważ napięcie na kondensatorze nie może zmieniać się bardzo szybko (zmiana zgromadzonej energii), to na wejściu "-" napięcie zmienia się wolniej aniżeli napięcie na wejściu "+". W ten sposób pojawia się narastający sygnał wejściowy różnicowy, powodujący wysterowanie wzmacniacza tak, że jego napięcie wyjściowe zmierza do wartości ujemnych.



Rys. 6.15. Przebiegi napięć w multiwibratorze astabilnym

Dodatnie sprzężenie zwrotne powoduje, że proces przesterowania wyjścia wzmacniacza od wartości $+U_{\text{sat}}$ do wartości $-U_{\text{sat}}$ przebiega bardzo szybko, przy niezbyt dokładnej obserwacji wydaje się to przeskokiem z dodatniego napięcia nasycenia do ujemnego napięcia nasycenia.

Po "przeskoku" napięcie wejścia "-" pozostaje początkowo przy poprzedniej wartości $+U_{\text{sat}}[1 - \exp(-t_1/RC)]$, natomiast napięcie wejścia "+" będzie wynosić teraz $-U_{\text{sat}}R_1/(R_1 + R_2)$. Kondensator C będzie teraz ładowany przez opornik R do wartości $-U_{\text{sat}}$, tj. początkowo będzie rozładowywany od pozostałej wartości dodatniej napięcia do zera, a następnie będzie narastało na nim napięcie ujemne. Ten proces opisuje wzór:

$$u_C(t) = -U_{\text{sat}} \left[1 - \exp\left(\frac{-t}{RC}\right) \right] + U_{\text{sat}} \frac{R_1}{R_1 + R_2} \exp\left(\frac{-t}{RC}\right) \quad (6.31)$$

W chwili t_2 napięcie na przełączającym się kondensatorze (czyli i napięcie na wejściu "-"), przyjmie wartość taką, jaka występuje na wejściu "+", co zainicjuje kolejny "przeskok" wzmacniacza od nasycenia ujemnego do dodatniego. Zapisując równość obydwu napięć w chwili t_2 :

$$-U_{\text{sat}} \left[1 - \exp\left(\frac{-t_2}{RC}\right) \right] + U_{\text{sat}} \frac{R_1}{R_1 + R_2} \exp\left(\frac{-t_2}{RC}\right) = -U_{\text{sat}} \frac{R_1}{R_1 + R_2} \quad (6.32)$$

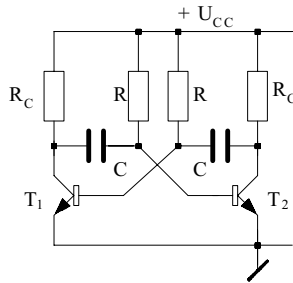
wyznaczą się:

$$t_2 = RC \ln\left(1 + \frac{2R_1}{R_2}\right) \quad (6.33)$$

Wyznaczony czas t_2 odpowiada czasowi występowania napięcia ujemnego na wyjściu wzmacniacza. Identyczna procedura może być wykorzystana do wyznaczenia czasu występowania napięcia dodatniego. Procesy przeskoków i przeładowań kondensatora będą powtarzać się jak pokazano na rys. 6.15. Na wyjściu wzmacniacza otrzymuje się więc ciąg prostokątnych impulsów napięcia o wartościach na przemian $+U_{\text{sat}}$ i $-U_{\text{sat}}$ okres powtarzania impulsów T wynosi dwukrotną wartość t_2 , czyli:

$$T = 2RC \ln \left(1 + \frac{2R_1}{R_2} \right) \quad (6.34)$$

Dobierając wartości rezystancji R , R_1, R_2 oraz pojemność kondensatora C można uzyskać potrzebną wartość częstotliwości $f = 1/T$ powtarzania impulsów prostokątnych na wyjściu multiwibratora, amplituda tych impulsów jest równa napięciu nasycenia wzmacniacza.



Rys. 6.16. Tranzystorowy multiwibrator astabilny

Multiwibrator astabilny może być wykonywany i w innych układach, niekoniecznie z zastosowaniem wzmacniacza operacyjnego.

Konwencjonalne rozwiązanie, stanowiące tranzystorową modyfikację lampowego układu Eccles–Jordana, znane jest w licznych modyfikacjach układowych już od dziesiątków lat. Schemat najprostszego wariantu pokazano na rys. 6.16. Stosowane są tu dwa elementarne wzmacniacze w układzie wspólnego emitera, z których każdy zawiera jeden tranzystor oraz dwa rezystory a ponadto dwa kondensatory sprzęgające, które są odpowiednio przeładowywane w czasie pracy.

Działanie generacyjne zapewnia dodatnie sprzężenie zwrotne, polegające na tym, że wyjście pierwszego wzmacniacza jest połączone przez kondensator sprzęgający z wejściem drugiego wzmacniacza (kaskadowe połączenie dwu wzmacniaczy), a sygnał z wyjścia drugiego stopnia, też przez kondensator, powraca na wejście pierwszego wzmacniacza.

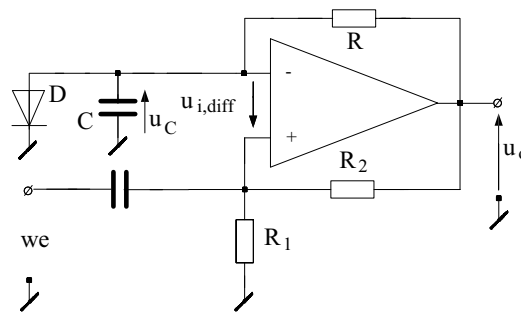
6.3.3. Multiwibrator monostabilny

Multiwibrator monostabilny jest generatorem pojedynczego impulsu prostokątnego o określonej amplitudzie i czasie trwania. Multiwibrator monostabilny jest generatorem wyzwalanym, to znaczy normalnie jest w stanie oczekiwania na impuls pobudzający, który może mieć dość dowolny kształt. Po pobudzeniu układ monostabilny wytwarza jeden standardowy impuls prostokątny i powraca do stanu oczekiwania na następne pobudzenie. Schemat tego generatora, wykorzystujący przerzutnik Schmitta, pokazano na rys. 6.17.

Stabilny stan oczekiwania polega na wprowadzeniu wzmacniacza w nasycenie z dodatnią wartością napięcia wyjściowego. W tym stanie w układzie płyną dwa prądy: od wyjścia wzmacniacza przez dzielnik rezystancyjny R_1/R_2 oraz przez rezystor R i diodę D . Napięcie na wejściu "+" względem masy wynosi $+U_{\text{sat}}R_1/(R_1 + R_2)$, napięcie na wejściu "-"

jest równe spadkowi napięcia na przewodzącej diodzie, czyli ok. 0,6 V. Aby taki stan był stabilny, to napięcie wejściowe różnicowe powinno być dodatnie, zatem powinno być:

$$U_{\text{sat}} \frac{R_1}{R_1 + R_2} > 0,6 \text{ V} \quad (6.35)$$



Rys. 6.17. Multiwibrator monostabilny (uniwibrator)

Pobudzenie multiwibratora powoduje się wprowadzając do wejścia dowolny impuls, który obniży chwilowo napięcie wejścia "+" poniżej wartości 0,6 V, aby spowodować chwilową zmianę znaku wejściowego napięcia różnicowego. Wtedy nastąpi przejście wzmacniacza do stanu nasycenia ujemnego i wymuszona chwilowym pobudzeniem ujemna wartość U_{idiff} zostanie podtrzymana przez zmienioną wartość napięcia z dzielnika oporowego. Jednocześnie rozpocznie się ładowanie kondensatora przez rezystor R od początkowej wartości +0,6 V do wartości $-U_{\text{sat}}$. Na rys. 6.18 pokazano przebiegi napięć w układzie.

Przebieg napięcia na kondensatorze opisuje wzór:

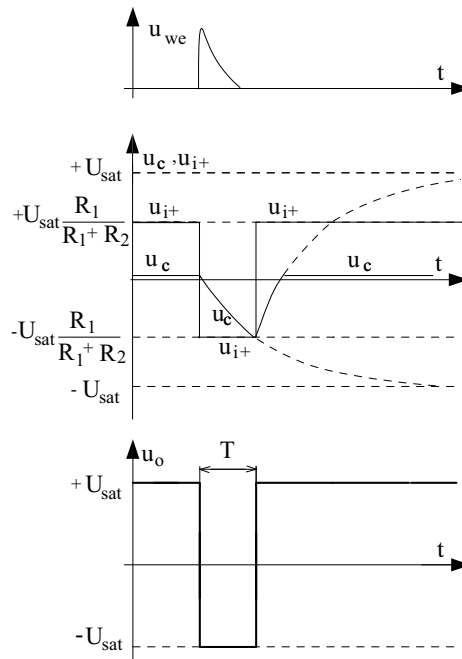
$$u_C(t) = -U_{\text{sat}} \left[1 - \exp\left(\frac{-t}{RC}\right) \right] + 0,6 \text{ [V]} \exp\left(\frac{-t}{RC}\right) \quad (6.36)$$

Po czasie T napięcie na przeładowywanym kondensatorze osiągnie wartość napięcia, jakie jest ustalone dzielnikiem na wejściu "+". W tym momencie zmienia się znak napięcia wejściowego różnicowego, co powoduje prawie skokowe przejście wzmacniacza do stanu nasycenia dodatniego czyli zakończenie generacji pojedynczego impulsu. Czas trwania impulsu wyznacza się przez porównanie napięcia na wejściu "-" z napięciem na wejściu "+":

$$-U_{\text{sat}} \left[1 - \exp\left(\frac{-T}{RC}\right) \right] + 0,6 \text{ [V]} \exp\left(\frac{-T}{RC}\right) = -U_{\text{sat}} \frac{R_1}{R_1 + R_2} \quad (6.37)$$

przy uproszczeniach wyznacza się

$$T \approx RC \ln\left(\frac{R_1 + R_2}{R_2}\right) \quad (6.38)$$

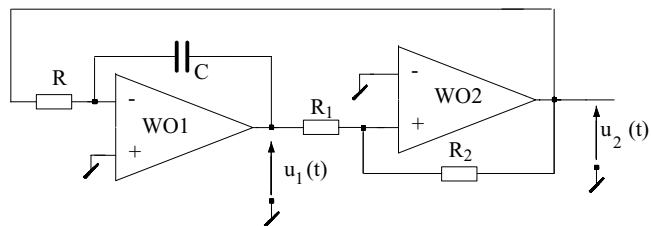


Rys. 6.18. Przebiegi sygnałów w multiwibratorze monostabilnym

Po wygenerowaniu impulsu w układzie musi jeszcze nastąpić powtórne przeładowanie kondensatora do początkowej wartości napięcia, tj. 0,6 V. W tym czasie należy unikać pobudzenia generacji następnego impulsu, bo jego czas trwania będzie odbiegał od standardu zapisanego wzorem (6.38).

6.3.4. Generator sygnału trójkątnego

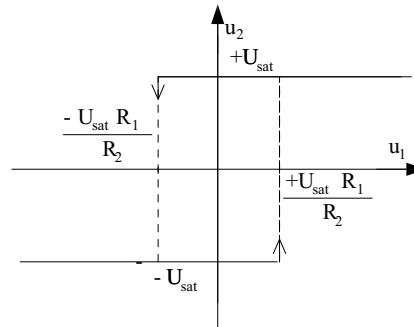
Na rys. 6.19 pokazano schemat układu generatora wytwarzającego dwa różne sygnały na dwu wyjściach. Na jednym wyjściu jest generowany sygnał trójkątny $u_1(t)$, a na drugim wyjściu sygnał prostokątny $u_2(t)$.



Rys. 6.19. Generator impulsów prostokątnych i trójkątnych

W układzie generatora można zauważyć dwa podzespoły: przerzutnik Schmitta oraz integrator (zob. p. 5.4.2). Przerzutnik Schmitta wykonany jest na wzmacniaczu WO2 z dodatnim sprzężeniem zwrotnym przez rezystor R_2 , a integratorem jest wzmacniacz WO1 zaopatrzony w ujemne sprzężenie zwrotne przez kondensator C.

W rozpatrywanym generatorze zastosowano nieco inny wariant przerzutnika Schmitta aniżeli przedstawione poprzednio. Niezmienną cechą jest dodatnie sprzężenie zwrotne (rezystory R_1 i R_2), różnica polega na wprowadzeniu sygnału wejściowego przerzutnika przez rezystor R_1 i połączeniu wejścia "-" z masą elektryczną. Charakterystykę tego przerzutnika czyli zależność jego napięcia wyjściowego u_2 od wejściowego u_1 pokazano na rys. 6.20. Wartości napięć progowych oblicza się podobnie jak pokazano w p. 6.3.1, aby te napięcia były mniejsze od wartości U_{sat} musi być spełniony warunek $R_1 < R_2$ (niespełnienie tego warunku spowoduje przesterowanie wzmacniacza WO1 w integratorze poza zakres pracy liniowej).



Rys. 6.20. Charakterystyka przerzutnika Schmitta wg fragmentu schematu z rys. 6.19

Przerzutnik może znajdować się w jednym z dwu stanów nasycenia, zatem z jego wyjścia do wejścia integratora zostaje doprowadzone napięcie o ustalonej wartości U_{sat} o znaku dodatnim lub ujemnym.

Integrator na swoim wyjściu wytwarza napięcie proporcjonalne do całki po czasie z sygnału wejściowego. Jeżeli sygnałem wejściowym jest napięcie stałe U_{sat} , to napięcie wyjściowe integratora jest liniowo zmienne w czasie:

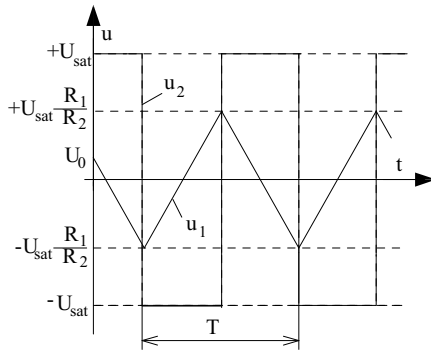
$$u_1(t) = -U_{sat} \frac{t}{RC} + U_0 \quad (6.39)$$

od początkowej wartości napięcia na kondensatorze U_0 . Przy np. dodatnim napięciu wejściowym integratora na jego wyjściu napięcie maleje liniowo od wartości dodatnich do ujemnych z szybkością zależną od wartości U_{sat} oraz R i C .

Działanie całego układu generatora rozpatrzmy zakładając, że napięcie na wyjściu przerzutnika jest $u_2 = +U_{sat}$ i na wyjściu integratora też jest dodatnie $u_1(0) = U_0 > 0$. W takim stanie napięcie u_1 będzie się zmniejszać, zmierzając do wartości $-U_{sat}R_1/R_2$, a przerzutnik będzie pozostawał w stanie dodatniego nasycenia.

Po zmniejszeniu napięcia u_1 do wartości progowej nastąpi przesterowanie przerzutnika, który przechodzi prawie skokowo do stanu nasycenia ujemnego. Ujemny sygnał wejściowy integratora spowoduje narastanie jego napięcia wyjściowego od wartości $-U_{sat}R_1/R_2$ w kierunku napięcia dodatniego. Przy osiągnięciu wartości $+U_{sat}R_1/R_2$ nastąpi kolejne przesterowanie przerzutnika ze stanu nasycenia ujemnego do dodatniego.

Takie zmiany napięć będą powtarzały się cyklicznie. Przebiegi napięć pokazano na rys. 6.21.



Rys. 6.21. Sygnały w generatorze trójkąta i prostokąta

Uwzględniając szybkość narastania napięcia i wartości progowe wyznacza się okres T powtarzania cyklu pracy generatora:

$$T = 4RC \frac{R_1}{R_2} \quad (6.40)$$

Na wyjściu wzmacniacza WO1 otrzymuje się więc napięcie trójkątne, o częstotliwości $f = 1/T$ i amplitudzie $U_{\text{sat}}R_1/R_2$, a na wyjściu WO2 występuje napięcie prostokątne o amplitudzie U_{sat} i tej samej częstotliwości f .

Na koniec warto zaznaczyć, że generatory, szczególnie impulsowe, ale także i sygnałów sinusoidalnych, są obecnie produkowane jako układy scalone w różnych odmianach.

Do takiego układu scalonego zwykle dołącza się tylko kondensator (niekiedy dwa kondensatory lub kondensator i dokładne rezystory) otrzymując generator sygnału o określonych właściwościach.

6.4. Stabilizacja częstotliwości – "generator kwarcowy"

Stołość częstotliwości generatorów, zarówno sygnałów impulsowych jak i sinusoidalnych, zależy w pierwszym rzędzie od stałości parametrów elementów określających częstotliwość tj. od stałości wartości indukcyjności L i pojemności C w obwodach rezonansowych lub stałości rezystancji R i pojemności C w filtrach RC .

Niestety wartości parametrów podlegają zmianom, wywoływanych przez zmiany temperatury elementów lub ich starzenie (zmiany składu chemicznego). Najbardziej kłopotliwe są zmiany temperaturowe, które powodują, że w przeciętnych warunkach użytkowania niestabilność częstotliwości generatorów wynosi od dziesiątych części do pojedynczych procentów częstotliwości projektowanej. Dla wielu zastosowań, szczególnie w aparaturze pomiarowej, zegarach, radiostacjach itp. są to zbyt duże wahania i od dawna poszukiwano i stosowano środki poprawiające stabilność częstotliwości generowanych sygnałów.

Popularnym środkiem stabilizacji częstotliwości generatorów sygnałów sinusoidalnych jest wykorzystanie efektu piezoelektrycznego w rezonatorach kwarcowych, zwanych "kwarcami".

Efekt piezoelektryczny, występujący w niektórych kryształach, polega na wzajemnym oddziaływaniu zjawisk: elektrycznego i mechanicznego. Kryształ umieszczony w polu elektrycznym podlega odkształceniu i odwrotnie, kryształ odkształcony skutkiem naprężeń mechanicznych wytwarza pole elektryczne, powodujące powstanie napięcia na jego ściankach. Oczywiście te efekty zależą też od kierunku działania naprężeń względem osi krystalicznych.

Rezonator kwarcowy to odpowiednio wycięty fragment kryształu kwarcu zaopatrzonego w dwie okładziny metalowe z wyprowadzeniami elektrycznymi, szkic budowy

pokazano na rys. 6.22. Z punktu widzenia elektrotechniki jest to kondensator płaski o dielektryku kwarcowym.

Przykładając napięcie przemiennie do takiego elementu wywołuje się naprężenia zmienne, powodujące cykliczne ściskanie i rozciąganie kryształu, czyli pobudza się drgania mechaniczne tego elementu. Płytkę kwarcową ma określoną częstotliwość drgań własnych (częstotliwość rezonansową) zależną od jej wymiarów, masy i sposobu zamocowania. Jeżeli częstotliwość napięcia pobudzającego drgania zostanie odpowiednio dobrana, to można pobudzić i podtrzymywać w elemencie kwarcowym drgania mechaniczne o jego własnej częstotliwości rezonansowej.

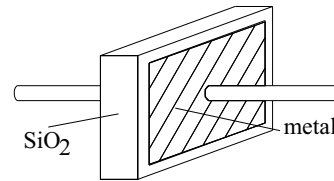
Przy rezonansie mechanicznym tj. przy częstotliwości rezonansowej, a także w bardzo wąskim przedziale wokół tej częstotliwości, występuje szczególne zjawisko elektryczne. O ile dla częstotliwości odległych od rezonansowej rezonator zachowuje się jak kondensator, tj. prąd przemienny wyprzedza w fazie napięcie, to przy rezonansie mechanicznym prąd opóźnia się względem napięcia. Oznacza to, że dla tej częstotliwości rezonator zachowuje się jak indukcyjność a nie jak kondensator. Wyjaśnienia tego zjawiska można poszukiwać w analizie wzajemnych przemian energii elektrycznej i mechanicznej w stanie rezonansu mechanicznego.

Zastosowanie rezonatora kwarcowego w prostym przypadku polega na budowie układu rezonansowego elektrycznego LC, w którym zamiast cewki indukcyjnej montuje się rezonator kwarcowy. Taki układ może oscylować, czyli podtrzymywać drgania elektryczne, tylko łącznie z mechanicznymi i w układzie generatora zachowuje się jak obwód rezonansowy LC tylko dla częstotliwości rezonansowej i w niewielkim przedziale wokół niej. Dla częstotliwości silnie odbiegającej od rezonansowej układ zachowuje się elektrycznie jak dwa kondensatory, co uniemożliwia generację drgań elektrycznych. Zastosowanie rezonatora kwarcowego zapewnia utrzymywanie stałej częstotliwości generatora z uchybem rzędu promila bez szczególnych zabiegów, a po np. termostabilizacji rezonatora niestałość częstotliwości może wynosić 10^{-6} .

6.5. Generator mocy

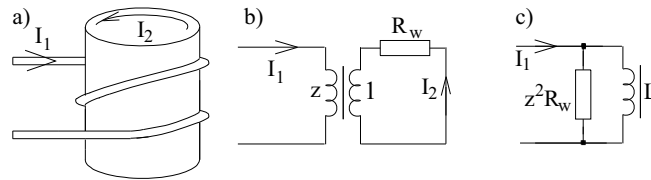
Jednym z zastosowań technologicznych generatorów mocy jest grzejnictwo indukcyjne metali. Nagrzewany wsad jest umieszczony wewnątrz uzwojenia, zwanego wzbudnikiem, które jest zasilane prądem przemiennym. Przemienne pole elektromagnetyczne wzbudza prąd w metalowym wsadzie, a ponieważ każdy metal ma określoną rezystancję R_w , to następuje w nim wydzielanie mocy elektrycznej $I^2 R_w$ zamieniającej się w ciepło.

Na rys. 6.23a pokazano szkic konstrukcji wzbudnika ze wsadami. Wzbudniki wykonywane są w postaci jednego lub więcej zwojów, o kształcie dopasowanym do kształtu wsadu. Dla topienia metali stosuje się prądy o stosunkowo niskich częstotliwościach (50 Hz), ponieważ wtedy wydzielanie ciepła następuje prawie równomiernie w całej objętości wsadu. Prądy o wyższych częstotliwościach, wywołujące silniejszy efekt naskórkowy,



Rys. 6.22. Rezonator kwarcowy

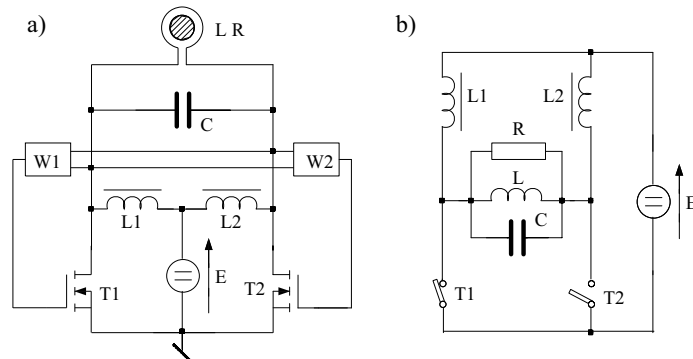
powodują nagrzewanie głównie warstwy przypowierzchniowej wsadu, umożliwiając np. hartowanie powierzchniowe stali.



Rys. 6.23. Wzbudnik i wsad (a), schematy zastępcze (b) i (c)

Prądy o podwyższonych częstotliwościach wytwarza się w elektronicznych generatorach mocy, nazywanych tak, ponieważ istotnymi ich cechami są stosunkowo duże moce oddawane do wsadu i wysoka sprawność, a wymagania odnośnie do np. stabilności częstotliwości generowanej czy kształtu prądu zwykle nie są ostre.

Z punktu widzenia elektrotechniki zespół wzbudnik-wsąd można traktować jak transformator, którego uzwojeniem pierwotnym jest wzbudnik o liczbie zwojów z , natomiast uzwojeniem wtórnym jest jeden zwój wsadu. Ten jeden zwój jest zwarty (zamknięty), a jego rezystancję reprezentuje rezystor R_w . Na rys. 6.23b pokazano schemat elektryczny przedstawiający zespół wsad-wzbudnik, a na rys. 6.23c schemat zastępczy, jaki uzyskuje się przeliczając obciążenie strony wtórnej transformatora na stronę pierwotną z uwzględnieniem przekładni zwojowej transformatora, w tym przypadku $z:1$, i uwzględniając także indukcyjność wzбудnika L .



Rys. 6.24. Schemat generatora przeciwsobnego (a) i schemat zastępczy (b)

Na rys. 6.24a pokazano uproszczony schemat generatora mocy w układzie przeciwsobnym. Elementy L i R reprezentują wzbudnik i wsad, kondensator C zapewnia utworzenie obwodu rezonansowego. Dławiki L_1 i L_2 umożliwiają zasilanie prądowe obwodu RLC ze źródła napięcia stałego E .

W obwodzie rezonansowym pobudza się i podtrzymuje drgania elektryczne tj. wytwarza się i utrzymuje napięcie prawie sinusoidalne na elementach R, L, C, co wymusza prąd przemienny we wzбудniku. Pracę generatora zapewniają tranzystory T1 i T2, które są sterowane dwustanowo, tj. gdy jeden z nich jest zatkany (nie przewodzi prądu), drugi zostaje wysterowany tak silnie, aby przewodził duży prąd przy możliwie małym spadku napięcia. Po czasie jednego półokresu drgań w generatorze następuje zamiana funkcji i tranzystor uprzednio przewodzący zostaje zatkany, a uprzednio zatkany zostaje wysterowany. Synchronizację kluczowania i odpowiednie sterowanie tranzystorami zapewniają wzmacniacze W1 i W2.

Działanie układu rozważymy na podstawie uproszczonego schematu rys. 6.24b zakładając, że został wysterowany tranzystor T1, co zamyka obwód, w którym ze źródła E przez dławik L1 i tranzystor T1 zaczyna płynąć prąd. Ten prąd będzie stopniowo narastał z prędkością $di/dt = E/L1$ i dlatego, po pewnym czasie, trzeba ten obwód przerwać, zatykając tranzystor T1, zanim prąd osiągnie zbyt dużą, niszczącą, wartość; jednocześnie z wyłączeniem T1 zostaje załączony tranzystor T2. Wywołanie prądu w dławiku L1 oznacza zgromadzenie w nim energii pola magnetycznego, próba przerwania tego prądu wywołuje zjawisko samoindukcji elektromagnetycznej (rozładowanie zgromadzonej energii), polegające tutaj na wytworzeniu siły elektromotorycznej, podtrzymującej przepływ prądu teraz od źródła E, przez L1, obwód rezonansowy i tranzystor T2.

Po włączeniu tranzystora T2 zamyka się też drugi obwód, obejmujący E, L2 i T2. W tym obwodzie przepływa narastający prąd, to znaczy, że w indukcyjności L2 następuje gromadzenie energii.

Po czasie jednego półokresu drgań rezonansowych wyłącza się tranzystor T2 a jednocześnie ponownie włącza się tranzystor T1. Przerwanie prądu przez T2 powoduje teraz przekazywanie energii z dławika L2 przez wymuszenie przepływu prądu w obwodzie E – L2 – obwód rezonansowy – T1. Jednocześnie następuje gromadzenie energii w dławiku L1 przez przepływ narastającego prądu w obwodzie E – L1 – T1.

Cykliczne powtarzanie tych procesów powoduje okresowe zasilanie obwodu rezonansowego prądami o zmiennych kierunkach, co podtrzymuje drgania elektryczne w tym obwodzie. Właściwą synchronizację drgań własnych obwodu z impulsami prądu zasilania zapewnia sprzężenie zwrotne, polegające na przełączaniu tranzystorów w takt drgań rezonansowych.

Prąd wpływający do obwodu rezonansowego powoduje ładowanie kondensatora, przepływając przez uzwojenie wytwarza pole elektromagnetyczne wzbudnika, a część tego prądu, zaindukowana we wsadzie, powoduje wydzielanie w nim ciepła. Aby pobudzić drgania w obwodzie rezonansowym konieczne jest spełnienie warunku, znanego z analizy równań różniczkowych:

$$4R^2 > \frac{L}{C} \quad (6.41)$$

oraz zapewnienie, aby w momencie przełączania tranzystorów prąd zasilający (prąd dławika L1 lub L2) był większy od prądu płynącego przez indukcyjność obwodu L o wartość ΔI . Jeżeli spełnione są te dwa warunki, to nastąpi wzbudzenie drgań rezonansowych w obwodzie RLC, a napięcie na elementach tego obwodu jest opisane wzorem:

$$u(t) = -\frac{2\xi R \Delta I}{\sqrt{\omega^2 - \xi^2}} \exp(-\xi t) \sin\left(\sqrt{\omega^2 - \xi^2} t\right) \quad (6.42)$$

Parametry RLC obwodu rezonansowego są wyrażone przez współczynnik tłumienia ξ i częstotliwość własną ω , zdefiniowane wzorami:

$$\xi = \frac{1}{2RC}, \quad \omega^2 = \frac{1}{LC} \quad (6.43)$$

Prąd płynący przez rezystancję, dostarczający mocy grzewczej do wsadu wyraża się zależnością $i(t) = u(t)/R$. Zarówno napięcie jak i prąd mają kształt bliski do sinusoidy, o częstości $\sqrt{\omega^2 - \xi^2}$.

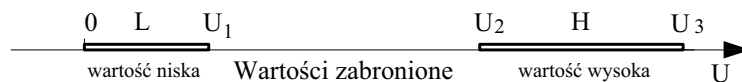
Przedstawiony układ jest tylko jednym ze stosowanych generatorów. Warto zauważyć, że w generatorach mocy unika się stosowania pomocniczych rezystorów, a także wymusza się dwustanową pracę tranzystorów, aby zminimalizować straty, czyli uzyskać wysoką sprawność, niezbędną przy przetwarzaniu mocy na poziomie do kilkudziesięciu kilowatów.

7. ELEKTRONIKA CYFROWA

Wiele układów scalonych jest konstruowanych i produkowanych dla przetwarzania sygnałów dwuwartościowych. Takie układy nazywa się popularnie układami cyfrowymi lub logicznymi, chociaż nie zawsze odpowiada to ściśle ich działaniu. Analizą i syntezą urządzeń przetwarzających sygnały dwuwartościowe zajmują się takie dziedziny jak technika cyfrowa i informatyka. W niniejszym skrypcie przedstawia się tylko rozwiązania techniczne niektórych podzespołów układów cyfrowych i logicznych i, tylko w minimalnym zakresie, wprowadza się pojęcia specjalizowanych dziedzin projektowania urządzeń cyfrowych.

7.1. Podstawowe pojęcia elektroniki cyfrowej

Każdy układ logiczny lub cyfrowy posiada wejście (lub kilka wejść) oraz wyjście (lub kilka wyjść), a wszystkie sygnały wejściowe i wyjściowe są dwuwartościowe, spełniające określone warunki techniczne.



Rys. 7.1. Dozwolone i zabronione wartości sygnałów

Jak podano w rozdziale 4, elektryczne sygnały dwuwartościowe są stosowane w układach przetwarzania informacji ze względu m.in. na dużą odporność na zakłócenia oraz na niewygórowane wymagania co do dokładności wykonania elementów i układów, co ma wpływ na ich cenę. Warunkiem wystąpienia odporności na zakłócenia jest m.in. właściwy dobór przedziałów wartości sygnałów:

- zakres dopuszczalnych wartości sygnału, które są traktowane jako sygnał "niski" (L – Low),
- zakres dopuszczalnych wartości sygnału, które są traktowane jako sygnał "wysoki" (H – High),
- zakres wartości niedopuszczalnych (umownie zabronionych), rozdzielający przedział L od H.

Na rys. 7.1 pokazano oś liczbową z oznaczonymi przedziałami przykładowych wartości sygnału dwuwartościowego.

Wartości sygnału: niska i wysoka mogą być umownie nazwane liczbami zero (0) i jeden (1). Takie przyporządkowanie może być zrobione albo w tzw. konwencji pozytywnej, gdzie wartość niską przedstawia zero a wartość wysoką jeden, albo w konwencji negatywnej stanowiącej odwrotnie.

Liczbowy, umowny, zapis wartości sygnału pozwala na stosowanie terminologii i reguł zaczerpniętych z logiki matematycznej do opisu operacji wykonywanych na sygnałach dwuwartościowych, w takim przypadku do elementów i układów powinno się stosować określenie "logiczne".

Liczbowy, zerojedynkowy, opis sygnałów można też interpretować jako zapis liczb w systemie dwójkowym i przekształcanie sygnałów traktować jako wykonywanie operacji arytmetycznych na tych liczbach, w takim przypadku powinno się stosować określenie "cyfrowe".

Pojedynczy sygnał dwuwartościowy, przyjmujący wartości 0 lub 1, przenosi minimalną dawkę informacji i jest nazywany bitem (bit – binary digit). Zespół kilku sygnałów dwuwartościowych, przenoszących bity składające się na jedną informację nazywa się sygnałem wielobitowym lub słowem binarnym. Zespół ośmiu sygnałów, czyli sygnał ośmio-bitowy, jest szeroko stosowany w technice komputerowej i nazywany bajtem (byte).

Na pojedynczym sygnale dwuwartościowym można wykonać dwie różne operacje:

- 1) zachowanie wartości (wzmocnienie),
- 2) zmiana wartości (inwersja).

W pierwszym przypadku, doprowadzając do wejścia układu sygnał o wartości 0 powodujemy, że na wyjściu tego układu występuje sygnał o wartości 0, i podobnie sygnał wejściowy 1 wywołuje sygnał wyjściowy też 1. Takiemu działaniu nie przypisuje się żadnej funkcji logicznej, a jedynie techniczną funkcję zwiększenia mocy sygnału (wzmocnienia), niezbędną w bardziej rozbudowanych urządzeniach logicznych.

W drugim przypadku sygnał wejściowy 0 wywołuje sygnał wyjściowy 1, a sygnał wejściowy 1 wywołuje sygnał wyjściowy 0. Jest to działanie odpowiadające logicznemu zaprzeczeniu, czyli negacji lub inwersji. Układ logiczny realizujący taką funkcję nazywa się negatorem, inwertorem lub funktorem NOT.

Dla dwu sygnałów wejściowych i jednego wyjściowego istnieje wiele możliwych uzależnień, czyli różnych funkcji logicznych uzależniających sygnał wyjściowy Y od sygnałów wejściowych A i B. Układy o dwu wejściach i jednym wyjściu nazywa się ogólnie bramkami, dodając w nazwie określenie realizowanej funkcji. Najważniejszymi funkcjami logicznymi dwu zmiennych są:

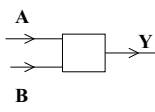
- 1) koniunkcja (iloczyn logiczny),
- 2) alternatywa (suma logiczna).

Bramkę realizującą koniunkcję nazywa się bramką AND, a bramkę realizującą alternatywę bramką OR. Na rys. 7.2 pokazano niektóre z 16 możliwych funkcji logicznych dwu zmiennych, podając nazwy odpowiednich bramek.

Bramki mogą być wykonywane także jako trzywejściowe, czterowejściowe aż po ośmiowejściowe z jednym wyjściem. Wszystkie wejścia w bramce są zazwyczaj "równoprawne", w bramce np. AND podanie zera na dowolne wejście spowoduje pojawienie się zera na wyjściu, a w bramce OR jedynka na dowolnym wejściu wymusza jedynkę na wyjściu. Niektóre bramki mają dwa wyjścia, o różnych uzależnieniach od wejść.

Łącząc odpowiednio wyjścia bramek z wejściami innych bramek buduje się tzw. układy kombinacyjne, charakterystyczne tym, że wartości sygnału czy sygnałów wyjściowych całego układu są w każdej chwili jednoznacznie określone wartościami sygnałów wejścio-

wych tego układu, aczkolwiek to przyporządkowanie może mieć postać dość złożonych funkcji logicznych lub operacji arytmetycznych.



Wej.		AND	OR	NOR	NAND	EXOR
A	B	Y	Y	Y	Y	Y
0	0	0	0	1	1	0
0	1	0	1	0	1	1
1	0	0	1	0	1	1
1	1	1	1	0	0	0

Rys. 7.2. Niektóre funkcje logiczne dwu zmiennych

Niektóre z tych dość rozbudowanych układów wykonują typowe operacje na wielobitowych sygnałach, potrzebne w różnych urządzeniach cyfrowych, i z tego powodu są one produkowane w postaci układów scalonych (zawierających kilka do kilkudziesięciu bramek w jednej strukturze półprzewodnikowej) o określonym przeznaczeniu. Tak wytwarzane są np. sumatory, w których sygnały wyjściowe przedstawiają sumę arytmetyczną dwu sygnałów wejściowych, czy komparatory porównujące dwa sygnały wejściowe i wytwarzające sygnał wyjściowy określający, która z liczb wyrażonych sygnałami wejściowymi jest większa i inne układy. Takie układy nazywa się cyfrowymi blokami funkcjonalnymi, niektóre z nich są przedstawione dalej.

Jednostką arytmetyczno-logiczną (ALU – Arithmetic Logic Unit) nazywa się blok funkcjonalny o zmiennej, sterowanej funkcji przetwarzania wielobitowych sygnałów wejściowych na wyjściowe. Taki układ przy określonym sygnale sterującym działa np. jak sumator arytmetyczny, a przy innym sygnale sterującym działa jak funktor iloczynu logicznego itd. Jednostka arytmetyczno-logiczna jest podstawowym składnikiem procesora komputerowego.

Warto zauważyć, że przy stosowaniu sygnałów wielobitowych każde z pojedynczych wejść czy wyjść bloku funkcjonalnego ma ściśle określone przeznaczenia. Jeżeli są np. cztery wejścia jednobitowe A0, A1, A2, A3, tworzące razem czterobitowe wejście A cyfrowego bloku funkcjonalnego, to podanie sygnału A = 1000 (A0 = 1; pozostałe zero) ma zupełnie inny sens aniżeli podanie sygnału A = 0001 (pierwsze trzy zera; A3 = 1). W bramce czterowejściowej sygnały wejściowe 1000 i 0001 są z reguły równoważne.

Odmianą grupę układów logicznych i cyfrowych stanowią tzw. układy sekwencyjne, budowane z wykorzystaniem przerzutników (por. przerzutnik opisany w p. 6.3.1). W tych układach wartości sygnałów wyjściowych w każdej chwili zależą nie tylko od sygnałów wejściowych w tej chwili, ale także od stanu układu, jaki był wywołany uprzednio. W skrócie opisuje się to, mówiąc o "pamięci" układu, zachowującej informację o przeszłości.

Przerzutniki są często stosowane do zapamiętywania wartości sygnałów wielobitowych. W takim przypadku konieczny jest odpowiedni zespół przerzutników nazywany rejestrem. Produkowane są więc nie tylko pojedyncze przerzutniki ale i rejestry o różnych właściwościach.

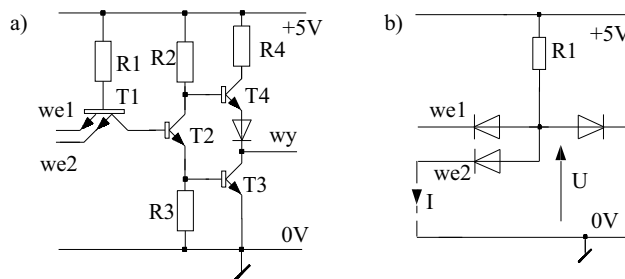
Ważną dziedziną zastosowań przerzutników są liczniki, zespół przerzutników zmieniających odpowiednio swoje stany pod wpływem każdego impulsu wprowadzonego do tzw. wejścia liczącego. Sprawdzając stan przerzutników licznika można otrzymać informację, ile impulsów było poprzednio doprowadzonych do jego wejścia.

Większe zespoły przerzutników, przeznaczone do pamiętania większej ilości liczb binarnych nazywa się pamięciami półprzewodnikowymi. Warto odnotować, że duże pamięci wykonywane są także na podstawie innych koncepcji, bez stosowania przerzutników.

7.2. Funktory TTL, CMOS i ECL

Bramki, przerzutniki i inne układy wykonywane są w różnych technikach, z tranzystorów bipolarnych i unipolarnych, zwykle w postaci zestawu elementów scalonych o różnych funkcjach, ale o jednakowych określeniach poziomów sygnału, napięciach zasilających i innych parametrach technicznych. Takie zestawy zwykle noszą nazwę pochodzącą od jakiejś istotnej ich cechy konstrukcyjnej. Obecnie dość szeroko stosowana jest rodzina TTL (Transistor–Transistor–Logic), rodzina CMOS (Complementary MOS) i rodzina ECL (Emitter Coupled Logic). Każda z rodzin składa się z kilkudziesięciu do kilkuset różnych bramek, przerzutników, cyfrowych bloków funkcjonalnych i innych elementów. Konstruktorzy gwarantują poprawną współpracę wszystkich elementów jednej rodziny między sobą, natomiast współpraca elementów z różnych rodzin wymaga zwykle szczegółowej analizy warunków technicznych i niekiedy wymaga stosowania elementów pośredniczących.

Na rys. 7.3a pokazano uproszczony schemat dwuwejściowej bramki z serii TTL. Taka bramka, podobnie jak inne elementy TTL, jest zasilana napięciem $+5\text{ V}$ z tolerancją 5% , sygnał wyjściowy w stanie niskim wynosi od 0 do $0,4\text{ V}$, a w stanie wysokim od $2,8$ do $5,0\text{ V}$.

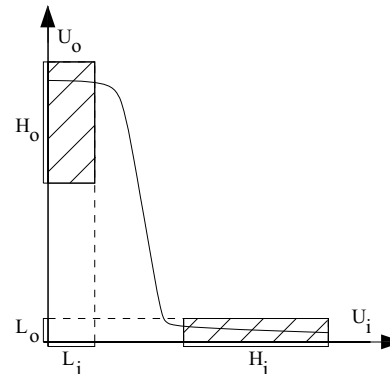


Rys. 7.3. Bramka TT: schemat (a) i model wejść (b)

Elementem wejściowym bramki TTL jest tzw. tranzystor wieloemiterowy, który tutaj działa podobnie jak zestaw diod pokazany na rys. 7.3b. Działanie całego układu rozpatruje się przy dwu wartościach sygnału wejściowego

Podanie sygnału niskiego – np. przez zwarcie wejścia z masą elektryczną – powoduje przepływ prądu przez rezystor R1 oraz diodę wejściową połączoną z masą. Napięcie U

przyjmuje wówczas wartość spadku napięcia na przewodzącej diodzie (około 0,6 V), co nie wystarcza do wysterowania dwu dalszych tranzystorów T2 i T3. W takiej sytuacji wyjście układu jest odłączone od masy przez nieprzewodzący tranzystor T3. Jednocześnie możliwy jest przepływ prądu przez rezystor R2, czyli możliwe jest wysterowanie tranzystora T4. Ten tranzystor zamyka połączenie pomiędzy dodatnim biegunem zasilania (+5 V) a wyjściem. Dzięki temu na wyjściu pojawia się napięcie dodatnie o poziomie wysokim. Warto też zwrócić uwagę na to, że w stanie niskim prąd wejściowy wpływa z bramki, zatem układ sterujący wejściem musi być zdolny do przyjęcia prądu (ok. 1,5 mA) i odprowadzenia go do masy.



Rys. 7.4. Charakterystyka przejściowa bramki TTL

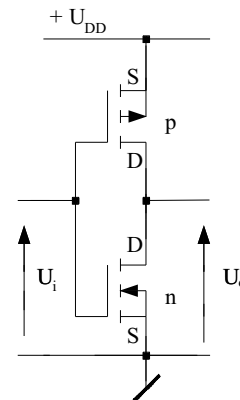
Gdy do obydwu wejść bramki, pokazanej na rys. 7.3, doprowadzone są sygnały o poziomie wysokim (np. przez zwarcie obydwu wejść z zasilaniem +5 V) następuje inny rozptyw prądów. Prąd przepływa przez rezystor R1, złącze baza-emiter tranzystora T2 i złącze baza-emiter tranzystora T3. Dzięki temu te obydwa tranzystory są wysterowane i wyjście układu przez wysterowany tranzystor T3 zostaje połączone z masą, co zapewnia niski poziom sygnału wyjściowego. Jednocześnie prąd, doprowadzany z zasilania przez rezystor R2, przepływa teraz przez T2 i R3, omijając tranzystor T4. Nieprzewodzący tranzystor T4 odcina wyjście układu od poziomu wysokiego. W takim stanie do wejścia bramki wpływa prąd spolaryzowanego wstecznie złącza (ok. 40 μ A). Jak wynika z podanych uwag w bramce TTL mogą wystąpić dwa stany:

- 1) wszystkie wejścia na poziomie wysokim – wyjście na poziomie niskim,
- 2) jedno lub więcej wejść na poziomie niskim – wyjście na poziomie wysokim.

Zapisując te związki w konwencji pozytywnej (niski – 0; wysoki – 1) można znaleźć, że ta bramka realizuje funkcję logiczną NAND. Można też sprawdzić, że w konwencji negatywnej ta sama bramka realizuje inną funkcję logiczną.

Można też używać bramki TTL niestandardowo, tj. doprowadzając do jednego lub dwu wejść sygnał analogowy, ciągły. W takim przypadku sygnał wyjściowy też będzie niestandardowy, analogowy. Charakterystykę przenoszenia, czyli zależność sygnału wyjściowego od wejściowego, pokazano na rys. 7.4, oznaczając zakresy napięć odpowiadające typowemu, dwustanowemu sterowaniu bramki.

Na rys. 7.5 pokazano schemat inwertora w technice CMOS. Taki inwertor składa się tylko z dwu tranzystorów MOS typu wzbogacanego, o różnych kanałach, których bramki oraz dreny są zwarte między sobą. Połączone bramki stanowią wejście inwertora, a połączone dreny jego wyjście.



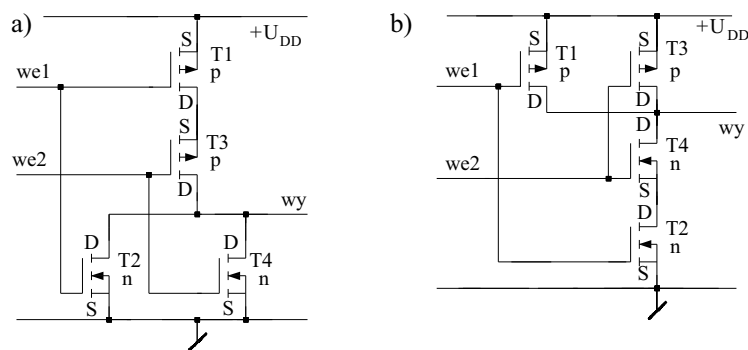
Rys. 7.5. Inwertor CMOS

Sygnal wejściowy o poziomie niskim (np. zwarcie wejścia z masą) powoduje, że napięcie bramka–źródło tranzystora z kanałem n wynosi zero, tak więc ten tranzystor jest w stanie zatkania i nie przewodziąc prądu odłącza wyjście układu od masy elektrycznej. Ten sam sygnał wejściowy dla tranzystora z kanałem p oznacza napięcie bramka–źródło o wartości równej napięciu zasilania. To napięcie jest większe od progowego i powoduje, że tranzystor z kanałem p zostaje wysterowany do stanu dobrego przewodzenia prądu, zapewniając połączenia wyjścia układu z dodatnim biegunem zasilania. Dzięki temu na wyjściu otrzymuje się sygnał o poziomie wysokim.

Przy sygnale wejściowym o poziomie wysokim (np. zwarcie wejścia z dodatnim biegunem zasilania) następuje odwrócenie sytuacji. Teraz napięcie bramka–źródło tranzystora z kanałem p wynosi zero i ten nieprzewodzący tranzystor odłącza wyjście od napięcia zasilania, a napięcie bramka–źródło tranzystora z kanałem n jest duże, większe od progowego, co zapewnia dobre przewodzenie tego tranzystora, czyli połączenie wyjścia z poziomem masy elektrycznej. Dzięki temu na wyjściu otrzymuje się sygnał o poziomie niskim.

Przy sygnałach niestandardowych charakterystyka przenoszenia inwertora CMOS jest podobna do charakterystyki bramki TTL pokazanej na rys. 7.4.

W technice CMOS buduje się i inne elementy. Na rys. 7.6 pokazano przykładowe schematy dwuwejściowych bramek CMOS. Na tych schematach można zauważyć, że bramki są zbudowane z elementów dwu inwertorów z dodatkowymi połączeniami wewnętrznymi. Pokazane na rys. 7.6a tranzystory T1 i T2 stanowią typowy inwertor, podobnie jak tranzystory T3 i T4. Wprowadzenie sygnału wysokiego na dowolne wejście lub na obydwa wejścia powoduje, że jeden z tranzystorów T2, T4 (lub obydwa) przechodzą w stan dobrego przewodzenia zapewniając niski sygnał wyjściowy. Jednocześnie odpowiedni tranzystor T1 czy T3 (albo obydwa) odłączają wyjście od zasilania napięciem wysokim. Tylko podanie obydwu sygnałów wejściowych o poziomie niskim zapewnia jednoczesne przewodzenie tranzystorów T1, T3 oraz nieprzewodzenie T2, T4, co gwarantuje poziom wysoki na wyjściu. Taka bramka w konwencji pozytywnej realizuje funkcję NOR. Odmienna konfiguracja połączeń, pokazana na rys. 7.6b, odpowiada bramce NAND.

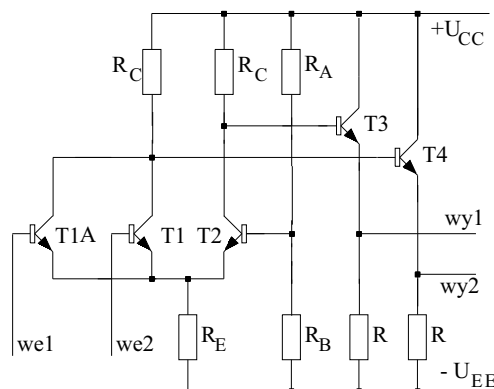


Rys. 7.6. Dwuwejściowe bramki CMOS

Inwertory, bramki, jak i inne elementy, CMOS można zasilac napieciem z zakresu od 3 do 18 V (inaczej niz w rodzinie TTL ktora wymaga zasilania 5 V). Z tego powodu okreczenie przedzialow napieci niskich i wysokich jest dla ukladow CMOS odmienne niz dla TTL i np. dla sygnalow wyjsciowych bramki poziom niski to napiecie od 0 do 50 mV, a poziom wysoki to wartosc mieszczaca sie pomiedzy napieciem zasilania a napieciem zasilania pomniejszonym o 50 mV.

W funktozach ECL wykorzystuje sie wlasciwosci wzmacniacza roznicowego, (patrz p. 5.3). Schemat bramki ECL pokazano na rys. 7.7. Wzmacniacz roznicowy zbudowany jest niesymetrycznie, w jednym ramieniu ukladu wystepuja dwa tranzystory, T1 i T1A, ktorych obwody wyjsciowe sa polaczone rownolegle, a w drugim ramieniu tranzystor T2. Napiecie wejsciowe wzmacniacza od strony tranzystora T2 jest ustalone dzielnikiem rezystancyjnym RA, RB. Napiecia dwuwartosciowych sygnalow wejsciowych, wprowadzanych do tranzystorow T1 i T1A, sa znacznie wieksze lub znacznie mniejsze od napiecia panujacego na bazie tranzystora T2.

Taki dobór wartosci napieci wejsciowych powoduje, ze wzmacniacz zawsze jest wprowadzany w stan nasycenia. Jezeli obydwa sygnaly wejsciowe sa niskie, to nasycenie wzmacniacza polega na tym, ze tranzystory T1 i T1A sa zatkane a tranzystor T2 przewodzi prąd (ale to nie znaczy ze jest on w stanie nasycenia, wręcz odwrotnie, elementy ukladu sa tak dobrane, aby tranzystor przewodzący pozostal w stanie aktywnym). Jezeli do jednego lub obydwu wejści doprowadzi się sygnal wysoki to jeden lub obydwa tranzystory T1 przejdą do stanu aktywnego a tranzystor T2 zostaje zatkaany.



Rys. 7.7. Bramka ECL

Napiecie na kolektorze tranzystora zatkaanego jest bliskie do wartosci U_{CC} , napiecie na kolektorze tranzystora przewodzącego jest znacznie nizsze. Napiecia wyjsciowe wzmacniacza roznicowego sa doprowadzone do dwu wzmacniaczy, ktorych wyjscia sa wyjsciami bramki. Zadaniem tych wzmacniaczy jest m.in. wytworzenie napieci wyjsciowych o wartosciach odpowiednich do dwustanowego wysterowania kolejnej bramki. W konwencji pozy-

tywnej taka bramka na jednym wyjściu realizuje funkcję OR, a na drugim wyjściu jej negację, czyli NOR.

Funktory ECL też dopuszczają różne wartości napięć zasilających, aczkolwiek producent zaleca stosowanie wartości U_{CC} równej zero (zwarcie zacisku $+U_{CC}$ z masą el.) natomiast U_{EE} może wynosić od -3 do $-10V$. Poziomy sygnałów w układach ECL też zależą od obranych napięć zasilania i wynoszą przykładowo poziom wysoki $-0,75V$, a poziom niski $-1,6V$.

Jak wynika z powyższego przeglądu, elementy z różnych rodzin, nawet realizujące identyczne funkcje logiczne, różnią się między sobą budową i działaniem wewnętrznym. Konsekwencją tego są też różnice w parametrach technicznych, obejmujące nie tylko wartości napięć zasilania, czy wartości (poziomy) sygnałów logicznych, ale także i inne cechy, jak szybkość działania elementu, pobór mocy ze źródła zasilania itd.

Wybór elementów do realizacji określonego urządzenia cyfrowego powinien być poprzedzony nie tylko analizą stawianych wymagań i planowanych warunków pracy projektowanego obiektu, ale powinien też uwzględniać dostępność elementów w różnych rodzinach funkatorów. Obecnie najszerszy wybór różnych typów elementów można znaleźć w rodzinie TTL, nieco "młodsza" rodzina CMOS jest aktualnie wzbogacana o nowe typy układów scalonych.

Śród syntetycznych wskaźników, charakteryzujących ważniejsze cechy rodzin elementów, poniżej przytacza się niektóre.

Cecha	ECL	TTL	CMOS
gęstość upakowania ilość standardowych bramek jaką można ulokować na powierzchni 1 mm^2 płytki krzemowej	10	20	300
czas propagacji [ns] opóźnienie z jakim wyjście bramki reaguje na zmianę sygnału wejściowego	0,3 – 2,0	2,0 – 30	3,0 – 100
moc strat [mW] moc tracona na ciepło w standardowej bramce w określonych warunkach pracy	25–60	1,0–20	0,01

Jak wynika z tych wskaźników układy ECL są najszybsze, ale zajmują najwięcej miejsca i zużywają najwięcej mocy, ich przeciwieństwem są układy CMOS, najwolniejsze, ale najbardziej podatne na dużą skalę scalenia, układy TTL mają właściwości pośrednie.

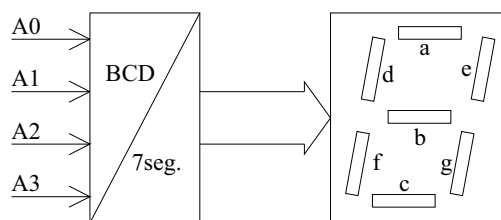
7.3. Cyfrowe bloki funkcjonalne

W niniejszym punkcie przedstawia się kilka przykładów cyfrowych bloków funkcjonalnych, mając na celu pokazanie typowych operacji dokonywanych na sygnałach cyfrowych. W opisie przyjęto zasadę "końcówkową", tj. ograniczono się do pokazania, dostępnych dla użytkownika, wyprowadzeń układu scalonego i opisano ich główne przeznaczenie, nie wnikając ani w konstrukcję wewnętrzną, ani w podstawy teoretyczne działania tych układów.

7.3.1. Kodery

Koder (dekoder, transkoder) to układ przetwarzający wielobitowy sygnał wejściowy w określonym kodzie na, też wielobitowy, sygnał w innym kodzie. Przedrostki de- lub trans- są stosowane dla określenia, które kody uważa się za pierwotne a które za wynik przekształceń.

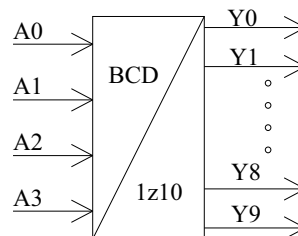
Często stosowane są dekodery sygnału BCD (zob. tabl. 4.1) na kod wskaźnika siedmiosegmentowego. Na rys. 7.8 pokazano układ blokowy kodera ze wskaźnikiem. Na rysunku zastosowano uproszczenie graficzne polegające na tym, że zamiast rysowania 7 linii przedstawiających 7 połączeń elektrycznych dekodera z wskaźnikiem narysowano jedną pogrubioną strzałkę, wskazującą jednocześnie kierunek przepływu sygnałów.



Rys. 7.8. Dekoder BCD na 7 segmentów i wskaźnik

Wskaźnik siedmiosegmentowy zawiera 7 diod świecących, odpowiednio rozmieszczonych, tak aby załączenie odpowiednich z nich powodowało wyświetlenie cyfry ze zbioru od 0 do 9. Na przykład włączenie diod eg powoduje świecenie cyfry 1 a włączenie diod adbgc cyfry 5. Dekoder przetwarza czterobitowy sygnał wejściowy tak, aby liczba reprezentowana tym sygnałem spowodowała wyświetlenie odpowiadającej jej cyfry. Na przykład sygnał wejściowy kodera $A_0 = 1$; $A_1 = A_2 = A_3 = 0$ ($A = 1000$) powoduje zaświecenia cyfry 1, a sygnał $A = 1010$ cyfry 5.

Do celów sygnalizacyjnych stosuje się też dekodek kodu BCD na kod "jeden z dziesięciu", którego schemat blokowy pokazano na rys. 7.9. Czterobitowy sygnał wejściowy A, o wartościach zawartych w kodzie BCD, powoduje wystąpienie sygnału 1 na wyjściu Y o numerze odpowiadającym wartości A, na pozostałych dziewięciu wyjściach Y występują wtedy sygnały o wartości zero. Sygnały wyjściowe mogą być wykorzystane do włączania podświetlenia odpowiednich cyfr.



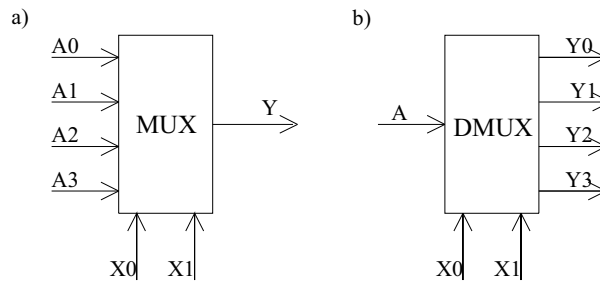
Rys. 7.9. Dekoder BCD na 1 z 10

7.3.2. Multipleksery, demultipleksery

Multipleksery i demultipleksery spełniają funkcję bezstykowych przełączników sygnałów binarnych. Na rys. 7.10 pokazano ich schematy blokowe.

Multipleksery pozwala skierować jeden z czterech jednobitowych sygnałów, doprowadzonych do wejść A_0 , A_1 , A_2 i A_3 na wyjście Y. O tym, który sygnał zostanie skiero-

wany do wyjścia Y decyduje stan wejść X_1 i X_0 zwanych wejściami adresowymi. Np. adres $X_0 = 0; X_1 = 0$ ($X = 00$) powoduje wewnątrz multipleksera proste połączenie wejścia A_0 z wyjściem Y. Podobnie sygnał adresowy np. $X = 10$ połączy wejście A_1 z wyjściem Y, a sygnał $X = 01$ spowoduje przekazywanie sygnału z wejścia A_2 na wyjście Y.



Rys. 7.10. Schematy blokowe: multipleksera (a) i demultipleksera (b)

Istnieją różne warianty multipleksorów, przeznaczone do przełączania sygnałów wejściowych jedno- lub wielobitowych, a ilość wejść A może wynosić 2, 4 lub 8. Oczywiście liczba bitów sygnałów A oraz Y musi być taka sama, natomiast liczba bitów sygnału adresowego X zależy od liczby wejść A . Przy dwu wejściach A wystarcza jednobitowy sygnał adresowy tak, aby przy $X = 0$ połączyć wejścia A_0 z wyjściami Y , a przy $X = 1$ połączyć A_1 z Y . Przy ośmiu wejściach A potrzebny jest trzybitowy sygnał adresowy X .

Odwrotną funkcję względem multipleksera spełnia **demultipleksor** – schemat blokowy na rys. 7.10b. Demultipleksor zapewnia przekazanie sygnału z jednego wejścia A na jedno z czterech wyjść Y_0 lub Y_1 lub Y_2 lub Y_3 . O wyborze wyjścia decyduje sygnał adresowy X . Na wybranym przez adres wyjściu występuje sygnał wejściowy A , na pozostałych wyjściach jest wówczas stan ustalony przez producenta, najczęściej na pozostałych wyjściach występuje zero. Podobnie jak multipleksery produkuje się też różne warianty demultipleksorów, zazwyczaj wytwarzane są pary układów o pokrewnych właściwościach np. multipleksor "4 na 1" oraz demultipleksor "1 na 4"

Multipleksor pozwala na przyłączenie kilku źródeł sygnałów binarnych do jednej linii przesyłowej, może to być np. zbieranie danych z kilku czujników pomiarowych do jednego urządzenia pomiarowego. Demultipleksor pozwala np. na sterowanie kilkoma obiektami z jednego punktu. W każdym przypadku będzie obowiązywać zasada "obsługa po kolei" czyli szeregowego przesyłu informacji.

7.3.3. Sumator

Sumator jest typowym blokiem cyfrowym, dokonującym operacji arytmetycznej dodawania dwu liczb, zwykle wyrażonych w kodzie binarnym naturalnym. Schemat blokowy sumatora pokazano na rys. 7.11.

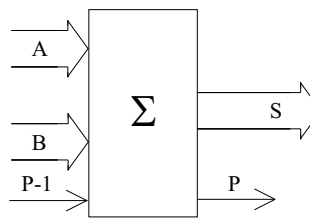
Do sumatora wprowadza się dwa sygnały wejściowe A oraz B , zwykle czterobitowe, przedstawiające składniki, a na wyjściu S otrzymuje się sygnał czterobitowy wyrażający ich sumę.

Dodatkowe wyjście P dostarcza sygnału przeniesienia. Ten sygnał przyjmuje wartość 1 jeżeli wynik sumowania przekracza maksymalną liczbę wyrażoną przez cztery bity, tj. 15. Przy dodawaniu binarnym jest podobnie jak w dziesiętnym, gdzie dodanie dwu liczb jednocyfrowych może dać w wyniku liczbę jednocyfrową (0...9) lub dwucyfrową (10...18). W tym drugim przypadku pojawiająca się jedynka jest przeniesiona do wyższego rzędu. Sygnał przeniesienia P odpowiada takiej "jedynce" w dodawaniu binarnym.

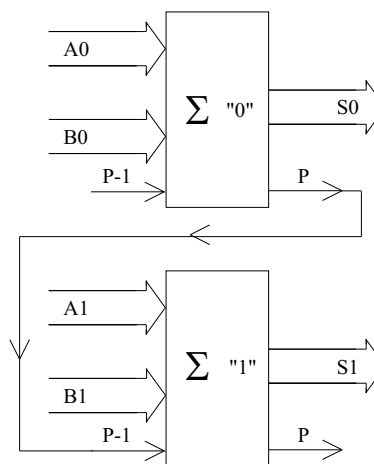
Dla umożliwienia dodawania liczb większych, aniżeli wyrażone przez cztery bity, przewiduje się łączenie sumatorów w zespoły, jak pokazano na rys. 7.12. W takim przypadku przeniesienie powstające w niższym rzędzie musi być dodane w wyższym rzędzie. Dla tego celu sumatory są zaopatrzone w wejście P-1, do którego wprowadza się sygnał przeniesienia z niższego rzędu.

7.3.4. Komparator

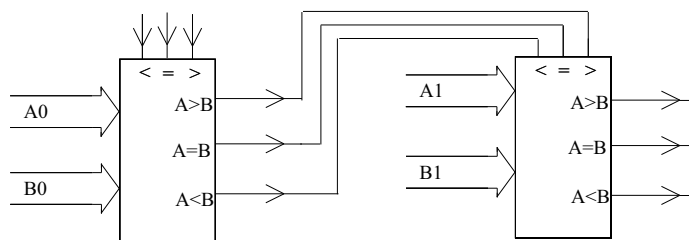
Komparator służy do porównania dwu liczb binarnych A i B. Wynikiem porównania jest stwierdzenie czy są one równe, a jeżeli nie, to która jest większa. Dlatego komparator ma trzy wyjścia: A = B; A < B; A > B oczywiście tylko na jednym z nich może pojawić się sygnał o wartości jeden a na dwu pozostałych będzie zero.



Rys. 7.11. Sumator



Rys. 7.12. Łączenie sumatorów



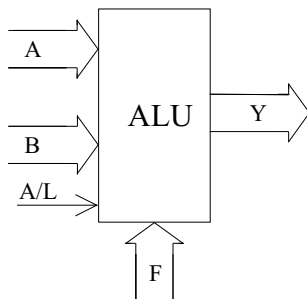
Rys. 7.13. Połączenie komparatorów

Sygnaly wejściowe A i B komparatora są zwykle czterobitowe, jeżeli trzeba porównywać większe liczby to stosuje się odpowiednią ilość komparatorów, jak pokazano na rys. 7.13. W takim przypadku mogą zaistnieć różne sytuacje, np. liczba z niższego rzędu A0

jest większa od B0, ale w wyższym rzędzie A1 jest równe B1 itp. Dla uwzględnienia tego typu zdarzeń i wyznaczenia relacji między kompletnymi liczbami A i B, komparator w wyższym rzędzie musi mieć informację o wyniku komparacji w niższym rzędzie i dlatego komparatory wyposażone są w wejścia " =; <; > " przyjmujące odpowiednią informację z członu poprzedniego.

7.3.5. Jednostka arytmetyczno–logiczna

Jednostka arytmetyczno–logiczna (ALU – Arithmetic Logic Unit) jest blokiem wykonującym różne operacje na dwu wielobitowych sygnałach wejściowych A i B lub na jednym z nich. Rodzaj wykonywanej operacji zależy od sygnału sterującego F, też wielobitowego. Dodatkowe wejście A/L pozwala przestawiać jednostkę na tryb pracy arytmetyczny lub logiczny. Na rys. 7.14 pokazano blok ALU z oznaczeniem wejść i wyjść.



Rys. 7.14. Jednostka arytmetyczno–logiczna

Przy czterobitowym sygnale sterującym F jednostka może wykonać jedną, wybraną z 16 różnych, operację arytmetyczną lub jedną z 16 różnych operacji logicznych. Zmieniając wartości sygnałów F oraz A/L można więc, w przypadku tej jednostki, realizować 32 różne operacje na sygnałach A i B.

Przy dokonywaniu operacji arytmetycznych sygnały A i B są traktowane jako liczby zapisane w kodzie binarnym. Operacjami arytmetycznymi, wykonywanymi przy odpowiednich wartościach sygnału F, są np.: $A + B$; $A + A$; $A - 1$; $A - B - 1$ itd. Warto wiedzieć, że nie wszystkie operacje arytmetyczne, które wydawałyby się podstawowymi, są wykonywane przez ALU. Z reguły jednostka nie wykonuje mnożenia dwu liczb przez siebie, aczkolwiek wykonuje operację $A + A$, która odpo-

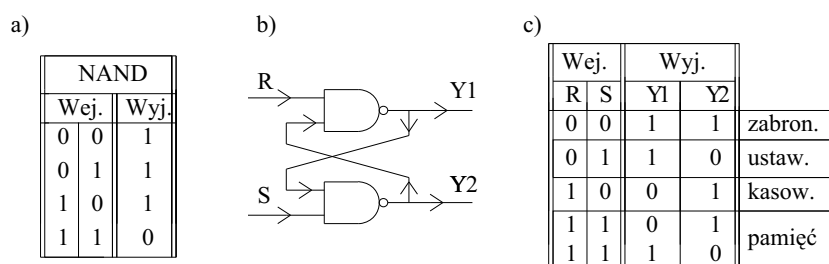
wiada mnożeniu przez 2. Rodzaj wykonywanych operacji jest kompromisem pomiędzy wymaganiami użytkowników a możliwościami technicznymi wykonania przy zachowaniu rozsądnego stopnia komplikacji konstrukcji.

Przy wykonywaniu operacji logicznych sygnały A i B są traktowane jako złożenie pojedynczych bitów. Wykonanie operacji np. AND na sygnałach A i B polega na obliczeniu iloczynu logicznego wartości pierwszych bitów A0 i B0 oraz przesłaniu wyniku do Y0, podobnie traktowane są pary A1 i B1, A2 i B2 itd czyli w tym przypadku ALU działa jak zespół niezależnych bramek AND. Operacjami logicznymi wykonywanymi przez ALU są: NOT(A); AND(A,B); NAND(A,B); NOR(A,B) itp.

Dla umożliwienia zestawiania pojedynczych, scalonych jednostek arytmetyczno–logicznych w zespoły, przetwarzające sygnały A i B o większej liczbie bitów, ALU jest wyposażone w dodatkowe wejścia i wyjścia jednobitowe pozwalające na przekazywanie sygnału przeniesienia przy dodawaniu i inne, podobnego typu informacje.

7.4. Przerzutniki, rejestry

W układach logicznych przerzutniki są wykonywane z bramek typu NOR lub NAND przez połączenie, co najmniej dwu bramek, w układ z pełnym sprzężeniem zwrotnym. Na rys. 7.15 pokazano schemat połączeń najprostszego przerzutnika oraz tablicę zero-jedynkową wartości sygnałów wejściowych i wyjściowych tego układu. Tablicę wyznacza się, analizując działanie każdej z dwu bramek, z uwzględnieniem wzajemnego ich oddziaływania – sprawdzenie prawdziwości tej tablicy pozostawia się czytelnikowi.



Rys. 7.15. Tablica prawdy bramki NAND (a), połączenie dwu bramek NAND w przerzutnik typu RS (b), tablica prawdy przerzutnika RS (c)

W przerzutnikach przyjmuje się, że jeżeli istnieją dwa wyjścia z przerzutnika, to ich sygnały powinny być wzajemnie zanegowane. Dlatego na rys. 7.15 zapisano jako zabronioną kombinację sygnałów wejściowych $R = 0$ i $S = 0$, powodujących wystąpienie jednakowych sygnałów na obydwu wyjściach Y1 i Y2.

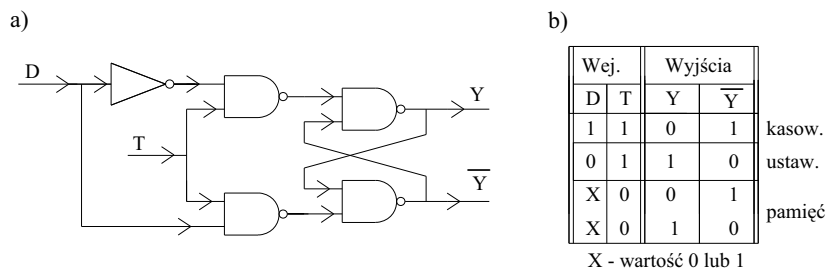
Kombinacje sygnałów wejściowych 0 1 i 1 0 wymuszają sygnał wyjściowy przerzutnika Y1 o wartości odpowiednio 1 lub 0 (Y2 odpowiednio 0 lub 1). Opisuje się to jako ustawienie przerzutnika w stan 1 (operacja S – set) lub kasowanie przerzutnika czyli ustawienie w stan 0 (operacja R – reset).

Kombinacja sygnałów wejściowych 1,1 nie wymusza określonego stanu przerzutnika, pozostaje on w stanie jaki był wymuszony poprzednio, czyli pamięta stan poprzedni. Po wprowadzeniu sekwencji sygnałów wejściowych np. 0 1, a następnie 1 1, przerzutnik pozostaje z sygnałem o wartości 1 na wyjściu Y1, opisuje się to mówiąc, że przerzutnik pamięta, zapisaną w nim poprzednio, jedynkę.

Na rys. 7.16 pokazano jedną z możliwości rozbudowy przerzutnika, w tym przypadku rozbudowano część poprzedzającą właściwy przerzutnik, tak aby wyeliminować możliwość wprowadzenia zabronionej kombinacji sygnałów wejściowych 0,0 do dwu bramek tworzących przerzutnik. Sygnał D pozwala wpisać do przerzutnika jedynkę lub zero przy sygnale T o wartości 1, sygnał $T = 0$ wprowadza przerzutnik w stan pamiętania. Inne odmiany przerzutników, o różnych sposobach ustawiania, kasowania i przechowywania wartości sygnału opisane są w literaturze [3].

Przytoczone powyżej dwa układy reprezentują też dwie koncepcje działania układów cyfrowych: działanie asynchroniczne i działanie synchroniczne. W układach asynchronicz-

nych, takich jak przerzutnik RS, nie czyni się żadnych założeń co do momentów zmian sygnałów wejściowych i wyjściowych, zwykle zmiana sygnału wejściowego powoduje natychmiastową zmianę sygnału wyjściowego. Inaczej jest w układach synchronicznych, w których wprowadza się generator impulsów taktujących zwany zegarem, i wszystkie zmiany sygnałów następują w momentach określanych impulsami taktującymi. W opisanym wyżej przerzutniku D sygnał D można wykorzystać jako sterujący (ustawiający lub kasujący) a sygnał T jako sygnał synchronizujący (zegarowy). Zmiana sygnału na wyjściu przerzutnika może nastąpić tylko przy sygnale zegarowym $T = 1$, bo tylko wtedy informacja z wejścia D może być wprowadzona do przerzutnika. Przy sygnale $T = 0$ (przerwa pomiędzy impulsami zegarowymi) przerzutnik nie może zmienić stanu, niezależnie od tego, jakie wartości występują w tym czasie na wejściu sterującym D.

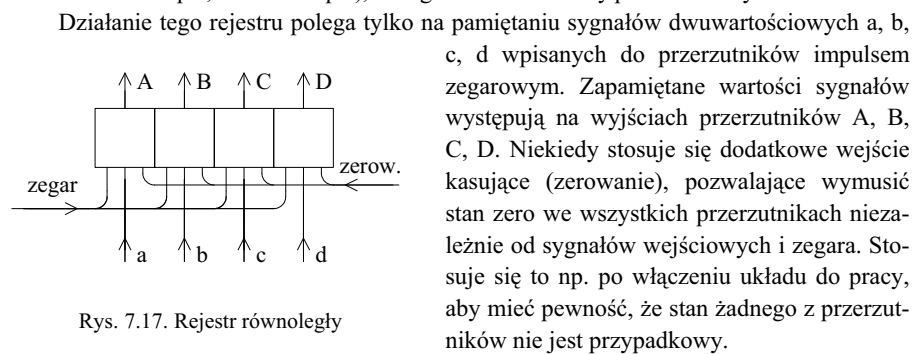


Rys. 7.16. Przerzutnik D: schemat (a) i tablica prawdy (b)

Układy działające synchronicznie są szczególnie dogodnie w złożonych urządzeniach cyfrowych, takich jak komputery, ponieważ ułatwiają współpracę różnych podzespołów między sobą i umożliwiają działanie w określonej kolejności, co jest podstawą wykonywania programów komputerowych.

Rejestrem nazywa się zespół przerzutników spełniających jednakowe, określone zadanie w urządzeniu cyfrowym. Stosowane są zespoły 4, 8, 16 aż po 1024 przerzutniki w jednym rejestrze.

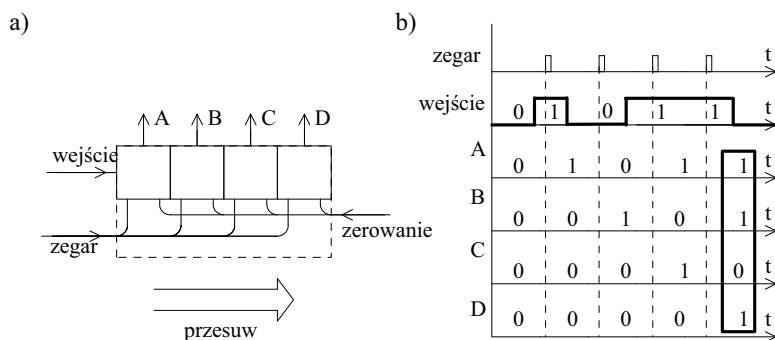
Najprostszym rejestrze jest układ z równoległym wejściem i równoległym wyjściem (zwany PIPO – Parallel Input, Parallel Output), którego schemat blokowy pokazano na rys. 7.17.



Rys. 7.17. Rejestr równoległy

Ważną grupę stanowią rejestry przesuujące, w których można przenieść stan pamiętany w jednym przerzutniku do przerzutnika sąsiedniego. W takich rejestrach często sygnał wejściowy jest doprowadzany do pierwszego, "skrajnego" przerzutnika i stan tego przerzutnika jest przenoszony, w takt impulsów zegarowych, do następnych przerzutników. Na rys. 7.18 pokazano sygnał wejściowy i stany rejestru w trakcie wpisywania informacji, ponieważ informacja wprowadzana jest kolejno, to tego typu wejście nazywa się szeregowym (SI – Serial Input).

Rejestry mogą realizować przesuw w prawo lub przesuw w lewo, lub też kierunek przesuwu może być ustalony dodatkowym sygnałem dwuwartościowym w odpowiednio skonstruowanym rejestrze.

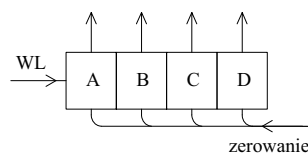


7.18. Rejestr SIPO: budowa (a), przebieg sygnałów (b)

W rejestrach przesuujących złożonych z dużej liczby przerzutników jest niemożliwe zaopatrzenie każdego przerzutnika w oddzielne wyjście (ze względu na ograniczoną technologicznie ilość wyprowadzeń z układu scalonego). W takich przypadkach tylko ostatni, w kolejności wpisywania, przerzutnik ma wyjście zwane szeregowym wyjściem rejestru (SO – Serial Output). Łącząc wyjście SO z wejściem SI tworzy się rejestr zamknięty. W takim rejestrze wpisana informacja krąży w takt impulsów zegarowych. To rozwiązanie jest stosowane jako pamięć dynamiczna o znacznej pojemności.

7.5. Liczniki

Licznikiem nazywa się zespół przerzutników, o odpowiednich połączeniach między nimi, przeznaczony do zliczania ilości impulsów wprowadzonych na jego wejście liczące. Na rys. 7.19 pokazano schemat blokowy licznika złożonego z czterech przerzutników, których wyjścia są jednocześnie wyjściami licznika. Każdy licznik ma wejście liczące, oznaczone na rysunku WL, a ponadto może mieć dodatkowe wejścia np. kasujące (zerujące) wszystkie przerzutniki lub ustawiające je w określony stan i in.



Rys. 7.19. Budowa licznika

W licznikach stosowane są różne kody zliczania (por. tabl. 4.1), najprostszy kod to tzw. binarny naturalny, w którym przerzutnik A zmienia swój stan po każdym impulsie wejściowym, przerzutnik B po każdym dwu impulsach, przerzutnik C po każdym czterech a przerzutnik D po każdym ośmiu impulsach.

Traktując sygnały wyjściowe przerzutników A, B, C, D jako sygnał czterobitowy (słowo binarne) i rozpoczynając od stanu wyzerowania wszystkich przerzutników czyli od stanu 0000, otrzymuje się po kolejnych impulsach wartości: 1000; 0100; 1100; 0010; 1010 itd. Po stanie 1111 wystąpi stan 0000 i licznik będzie działał dalej tak samo. Interpretując te wartości jako liczby zapisane w kodzie binarnym określa się bit A jako najmniej znaczący (LSB – Lowest Significant Bit), a bit D jako najwięcej znaczący (MSB – Most Significant Bit). Przeliczenie na system dziesiętny wykonuje się wg zależności:

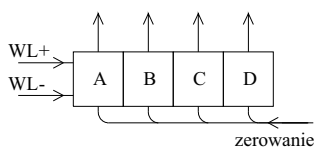
$$L = A + 2B + 4C + 8D$$

obliczając liczbę L, która jest dziesiętnym odpowiednikiem liczby binarnej DCBA. Najmniejsza wartość liczby L wynosi 0 a największa wartość L dla licznika czterobitowego jest 15. Liczniki można budować z dowolnej liczby przerzutników, albo korzystać z typowych liczników, np. czteroprzerzutnikowych w taki sposób, aby sygnał z wyjścia MSB pierwszego licznika wprowadzać do wejścia liczącego drugiego licznika itd. Dla licznika złożonego z liczby n przerzutników (licznika n-bitowego) największa wartość L wynosi $2^n - 1$, np. dwa liczniki czterobitowe pozwalają zliczać 256 impulsów (stan licznika od 0 do 255).

Stosowane są też liczniki zliczające w kodzie BCD, tzw. **liczniki dekadowe**. W tych licznikach następuje wykrycie stanu 1001 (odpowiadającego liczbie 9) przez jego układ wewnętrzny i wyzerowanie licznika, czyli przejście do stanu 0000, po następnym impulsie wejściowym. W ten sposób licznik wraca do stanu początkowego po zliczeniu dziesięciu impulsów, co ułatwia przetwarzanie stanu licznika do postaci czytelnej dla użytkownika.

Dla układów wykorzystywanych przy odliczaniu czasu stosowane są liczniki zliczające w kodzie szóstkowym lub dwunastkowym dla łatwiejszego dostosowania do obowiązującego systemu pomiaru czasu – sekundy, minuty, godziny. W tych licznikach stosuje się, podobnie jak w liczniku dekadowym, układy wewnętrzne wykrywające odpowiednią liczbę i zerujące licznik.

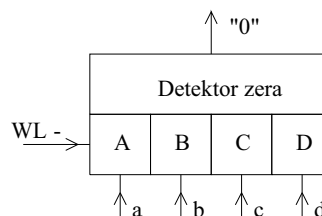
W układach automatyki znajdują zastosowanie **liczniki nawrotne** (rewersyjne) zdolne do dodawania lub odejmowania kolejnego impulsu. W tych licznikach albo występują dwa wejścia liczące, jak pokazano na rys. 7.20, albo jedno wejście liczące i jedno wejście określające kierunek zliczania. Zliczane impulsy, wprowadzane do wejścia WL+ powodują wzrost liczby pamiętanej w liczniku, takie same impulsy doprowadzane do wejścia WL– powodują zmniejszanie tej liczby. Każdy zliczany impuls zwiększa lub zmniejsza stan licznika o jeden. Takie liczniki są stosowane np. do pomiaru współbieżności dwu wałów zaopatrzonych w impulsatory. W liczniku z jednym wejściem liczącym kolejne impulsy są dodawane do stanu licznika jeżeli na wejściu kierunku zliczania jest utrzymywany sygnał 1, albo odejmowane, jeżeli na wejście kierunku zliczania poda się 0.



Rys. 7.20. Licznik nawrotny

Liczniki programowalne pozwalają na wpisanie do nich dowolnej liczby (oczywiście z przedziału obejmowanego tym licznikiem), a impulsy zliczane są odejmowane od wpisanej liczby. Wpisywanie zadanej liczby odbywa się równoległe albo, w licznikach o dużej liczbie przerzutników, liczba jest wpisywana w kilku kolejnych etapach, w każdym etapie do określonej części przerzutników licznika.

Schemat blokowy licznika pokazano na rys. 7.21. Licznik jest zaopatrzony w wewnętrzny układ wykrywania stanu zero (detektor zera) i wyjście z tego układu. Takie liczniki są stosowane na przykład do odliczania potrzebnej zwłoki (opóźnienia czasowego). Wprowadzając do licznika liczbę np. 9 i doprowadzając do wejścia WL impulsy z generatora o częstotliwości np. 1 Hz (okres powtarzania 1 s), otrzymuje się sygnał zera w liczniku po dziewięciu sekundach.



Rys. 7.21. Licznik programowalny

7.6. Mikroprocesor

Mikroprocesor (μP) jest układem scalonym przeznaczonym do przetwarzania sygnałów cyfrowych wg zadanego programu. Ze względu na założoną uniwersalność i znaczny stopień trudności nakładanych zadań, mikroprocesor ma dość skomplikowaną budowę wewnętrzną, jak i rozbudowany system komunikacji z elementami współpracującymi. O stopniu złożoności zadań można się przekonać analizując działanie komputera osobistego (PC), którym właśnie steruje mikroprocesor. W budowie wewnętrznej układu scalonego mikroprocesora używa się do miliona tranzystorów, z odpowiednimi połączeniami, co zajmuje stosunkowo znaczną powierzchnię na płycie krzemowej – około 1 cm². Produkowanych jest kilkadziesiąt typów różnych mikroprocesorów, w niniejszym punkcie przedstawia się tylko ważniejsze właściwości, wspólne dla większości odmian mikroprocesorów.

Efektywne działanie mikroprocesora jest możliwe we współpracy z innymi elementami, tworzącymi razem system mikroprocesorowy czyli mikrokomputer. Takimi elementami są: pamięć programu (ROM – Read Only Memory), pamięć danych (RAM – Random Access Memory), urządzenia wejścia i wyjścia łączące mikroprocesor z operatorem lub sterowanym obiektem technicznym (I/O – Input, Output) oraz elementy pomocnicze jak zasilacze i zegar (generator impulsów), synchronizujący pracę całego mikrokomputera. Systemy mikroprocesorowe są z reguły elastyczne, to znaczy mogą być powiększane o dalsze, potrzebne podzespoły, bez naruszania podstawowej struktury systemu.

Możliwości funkcjonalne mikroprocesora przedstawia jego lista rozkazów, wykaz wielobitowych (8, 16, lub 32 bity) sygnałów, które wprowadzone do mikroprocesora powodują jego określone działanie. Ilość różnych rozkazów wynosi od ok. 30 do ok. 150 w różnych mikroprocesorach. Rozkazy dotyczą różnych czynności, które można podzielić na następujące grupy:

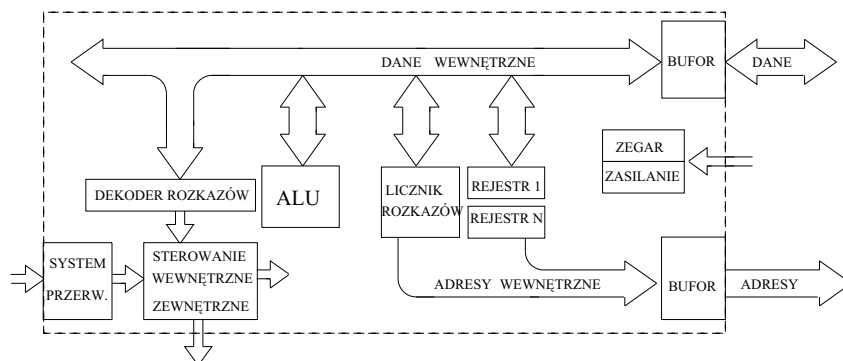
- arytmetyczne (dodaj, odejmij, dodaj 1 itp.),
- logiczne (zaneguj, porównaj i in.),

- przesyłu (prześlij z pamięci do rejestru, z rejestru do wyjścia i in.),
- skoku (skocz z wykonywanego programu do innego programu),
- przerwania (przerwij wykonywany program i wykonuj inny program),
- stanu procesora (zatrzymaj działanie, wyzeruj wszystkie rejestry i in.).

Na rys. 7.22 pokazano uproszczony schemat blokowy budowy mikroprocesora, na schemacie nie oznaczono wszystkich połączeń, aby go nadmiernie nie komplikować. Głównymi podzespołami mikroprocesora są:

- jednostka arytmetyczno-logiczna,
- rejestry, o przeznaczeniu określonym jak i ogólnym,
- licznik rozkazów,
- dekodery rozkazów,
- układ sterowania wewnętrznego i zewnętrznego,
- system przerwania.

Przesył informacji, zarówno wewnątrz procesora, jak i między procesorem a innymi elementami, odbywa się przez tzw. szyny, inaczej zwane magistralami (bus). Sama magistrala jest zespołem równolegle biegnących przewodów (8, 16 lub 32 przewody), do których połączone są wielobitowe wejścia i/lub wyjścia różnych podzespołów. Budowa tych podzespołów musi być taka, aby pomimo zwarcia elektrycznego pomiędzy wyjściami różnych podzespołów oraz ich wejściami, nie powodowało to kolizji w ich pracy. Taki sposób działania osiąga się to przez uaktywnianie (wprowadzanie w stan pracy) tylko określonego podzespołu właściwym sygnałem sterującym, kierowanym do tego podzespołu w odpowiednim momencie.



Rys. 7.22. Mikroprocesor

W niektórych przypadkach kierunek przesyłu informacji przez magistralę jest ustalony, np. tylko mikroprocesor wysyła adresy na szynę adresową i wówczas wystarcza uaktywnienie tylko jednego z podzespołów jako odbiorcy tego adresu. W innych przypadkach, np. przy przesyłaniu danych, kierunek przepływu sygnałów jest różny, albo mikroprocesor otrzymuje dane z innego podzespołu albo mikroprocesor wysyła tam wynik swoich obliczeń. W takim przypadku, oprócz uaktywnienia odpowiednich podzespołów, konieczne jest

jeszcze wysterowanie jednego z nich jako źródła sygnału (połączenie jego wyjść z szyną), a drugiego jako odbiornika tego sygnału (połączenia jego wejść z szyną). Organizację przesyłu informacji realizuje układ sterowania wysyłający właściwe sygnały przez szynę sterującą.

Działanie mikroprocesora polega na pobraniu rozkazu, a następnie na wykonaniu go i przygotowaniu się do pobierania następnego rozkazu. Ciąg rozkazów, czyli program działania, jest utrwalony w pamięci programu, w jej kolejnych komórkach. Aby pobrać rozkaz mikroprocesor wysyła, przez pośredniczący rejestr zwany buforem, adres komórki zawierającej pierwszy rozkaz. Adres ma postać zespołu zer i jedynek, wyrażający umowną liczbę przyporządkowaną określonej pamięci (lub urządzeniu wejścia czy wyjścia) oraz określonej komórce w tej pamięci. W odpowiedzi pamięć przekazuje zawartość swojej wywołanej komórki do szyn danych, poprzez które ten wielobitowy sygnał trafia do mikroprocesora, gdzie zostaje zapisany w buforze i przekazany do dekodera rozkazu w mikroprocesorze. Oczywiście tym operacjom muszą towarzyszyć odpowiednie sygnały sterujące.

Rozkaz może zawierać, oprócz części operacyjnej, np. "dodaj", zapisanej w postaci odpowiedniego ciągu zer i jedynek, także adresy dalszych komórek pamięci, np. przechowywujących składniki, które mają być dodane oraz adres, pod którym ma być ulokowany wynik dodawania. Takie informacje zostają czasowo przechowywane w rejestrach wewnętrznych mikroprocesora.

Dekoder przetwarza otrzymany rozkaz, organizując w ten sposób dalsze niezbędne czynności. Może to polegać np. na przesłaniu, pamiętanego chwilowo w jednym z rejestrów, adresu składnika dodawania na szyny adresowe i przeniesieniu wartości tego składnika z szyn danych do innego, wewnętrznego rejestru. Dalsze działania będą polegały na sprowadzeniu drugiego składnika, przekazaniu składników dodawania do ALU, ustawieniu ALU w tryb dodawania i wysłaniu wyniku dodawania, poprzez szyny danych, do odpowiedniej komórki pamięci danych.

Wykonywanie tych wszystkich działań odbywa się w takt impulsów zegarowych, których częstotliwość (od kilkunastu do kilkuset MHz) określa szybkość działania mikroprocesora. Wykonanie pojedynczej operacji, jak przesył zawartości jednego rejestru do drugiego, nazywa się cyklem maszynowym i zajmuje zwykle kilka taktów zegarowych, wykonanie całego rozkazu, to cykl rozkazowy, trwający kilka cykli maszynowych. W rezultacie wykonanie jednego rozkazu może trwać od ułamka do kilku mikrosekund.

Po wykonaniu wszystkich operacji związanych z odebraniem rozkazem należy przygotować adres komórki pamięci programu zawierającej następny rozkaz. Do tego służy licznik rozkazów, którego zawartość wzrasta o jeden po wykonaniu każdego rozkazu i jest przekazywana na szyny adresowe jako adres kolejnego rozkazu. Ten sposób narzuca umieszczanie kolejnych rozkazów programu w komórkach pamięci o kolejnych, rosnących wartościach adresów.

Dla uelastyczenia możliwości pracy w mikroprocesorach istnieje możliwość dokonywania tzw. skoków (jump) programowych. Układ sterowania, po otrzymaniu odpowiedniego, kolejnego rozkazu z programu podstawowego, powoduje przesłanie na szyny adresowe, zamiast zawartości licznika rozkazów, zawartość rejestru wewnętrznego. Upřednio w tym rejestrze musi zostać zapisany adres, pod którym znajduje się rozkaz do wykonania, zamiast kolejnego rozkazu z podstawowego programu. Oczywiście przejście do tego

innego, nie kolejnego, rozkazu jest początkiem wykonywania tzw. podprogramu, np. ciągu działań, które należy często powtarzać przy wykonywaniu głównego programu. Ostatnim rozkazem podprogramu będzie znowu skok, tym razem do głównego programu, tj. odtworzenie adresu, przy którym został wykonany skok do podprogramu. Dzięki temu, po wykonaniu przez mikroprocesor podprogramu, następuje powrót do wykonywania podstawowego programu.

Podobne nieco działanie następuje przy tzw. przerwaniu (interrupt). System przerwania umożliwia porzucenie głównego programu w dowolnym momencie i przejście do wykonywania innego, określonego podprogramu. W mikroprocesorowych układach sterowania ma to zastosowanie np. w sytuacjach awaryjnych, powodujących wysłanie sygnału przerwania do procesora. W takiej sytuacji mikroprocesor musi przerwać podstawowe działanie jako regulator i rozpocząć wykonywanie podprogramu likwidacji awarii. Nieprzewidziane uprzednio, w głównym programie, przejście do podprogramu, wymaga zapamiętania aktualnie wykonywanego rozkazu z programu głównego łącznie z wartościami danych, zawartych w tej chwili w wewnętrznych rejestrach mikroprocesora. Do tego celu służy tzw. stos, zespół dodatkowych rejestrów, lub wyodrębniony obszar w pamięci danych, do którego zostają przeniesione dane aktualnie znajdujące się w mikroprocesorze tak, aby po zakończeniu realizacji podprogramu wywołanego przerwaniem, był możliwy powrót i kontynuacja programu głównego. Zwykle procesor może przyjmować kilka do kilkudziesięciu sygnałów przerwania, o ustalonych priorytetach ważności, aby reagować na przewidziane przez projektantów sytuacje występujące w nieprzewidywalnych momentach.

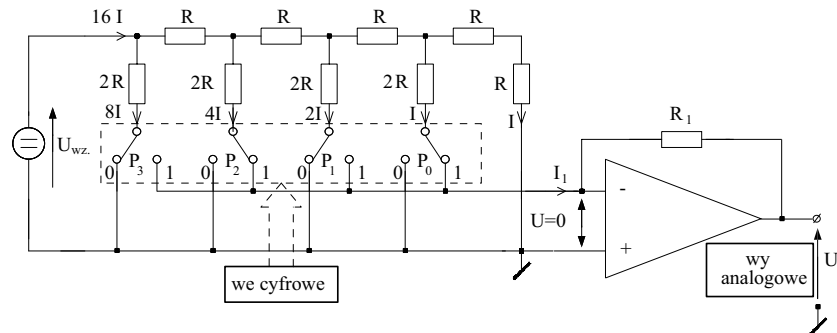
Dobór typu mikroprocesora oraz zestawianie elementów współpracujących dla konkretnego zadania technicznego, a szczególnie układanie programu głównego i podprogramów stanowią zadanie dla specjalistów informatyków i znacznie wykraczają poza zakres niniejszego skryptu.

7.7. Przetworniki c/a i a/c

Przy stosowaniu techniki cyfrowej w automatyce często występuje konieczność korzystania z przetworników pomiarowych dostarczających elektrycznych sygnałów analogowych, bądź wytwarzania sygnału sterującego w postaci analogowej.

Ponieważ wszystkie operacje w systemach mikroprocesorowych przeprowadza się na sygnałach cyfrowych (inaczej może być w tzw. mikroprocesorach sygnałowych), to jest konieczne przetwarzanie sygnału analogowego na cyfrowy lub odwrotnie. Układy elektroniczne używane w tym celu nazywa się przetwornikami analogowo-cyfrowymi czy cyfrowo-analogowymi, oznaczając je w skrócie a/c i c/a.

Na rys. 7.23 pokazano uproszczony schemat jednego z rozwiązań przetworników c/a. Przetwornik zawiera źródło napięcia wzorcowego U_{wz} , złożony dzielnik rezystancyjny, zwany drabinką $R - 2R$, podzespół cyfrowy, pokazany na rysunku jako blok zawierający przełączniki i wzmacniacz operacyjny objęty ujemnym sprzężeniem zwrotnym przez rezystor R_1 .



Rys. 7.23. Przetwornik c/a z drabinką R-2R

Prąd pobierany ze źródła napięcia wzorcowego wpływa do dzielnika rezystorowego, wykonanego z rezystorów o wartościach rezystancji R lub, dwukrotnie większej, $2R$. W dzielniku następuje podział tego prądu w proporcjach 1:2:4:8, jak opisano na rysunku (analityczne uzasadnienie takiego podziału prądu nie jest trywialnie proste). Wartość elementarnego prądu I wynosi w tym przypadku:

$$I = \frac{U_{wz}}{16R} \quad (7.1)$$

Podzespół cyfrowy stanowią układy przełączające, podobne w swej budowie wewnętrznej do funkcyjów ECL, których zadaniem jest kierowanie prądu wpływającego do przełącznika tak, aby wypłynął on jednym z dwu wyjść, stosownie do binarnego sygnału wejściowego. Przy sygnale wejściowym przełącznika równym "0", prąd jest kierowany do masy elektrycznej całego przetwornika i nie wpływa na dalsze jego działanie, przy sygnale "1" prąd jest kierowany do wejścia odwracającego wzmacniacza.

Ponieważ do wejścia wzmacniacza mogą być kierowane prądy z wszystkich przełączników, to następuje tam ich sumowanie. Wartość prądu płynącego w kierunku wejścia wzmacniacza wynosi:

$$I_1 = P_0I + P_12I + P_24I + P_38I = P I \quad (7.2)$$

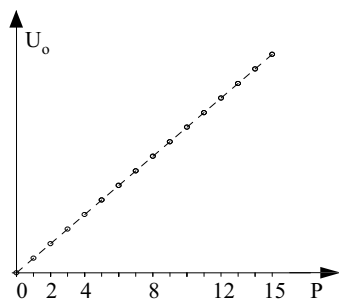
Wielkościami P_0 do P_3 oznaczono stany przełączników, gdy i -ty przełącznik jest w pozycji 0 ("na lewo") wartość P_i wynosi 0, gdy jest w pozycji 1 wartość P_i wynosi 1. Ponieważ położenie przełącznika zależy od cyfrowego sygnału wejściowego, to wielkości P_i można też traktować jako poszczególne bity cztero-bitowego sygnału wejściowego. Wielkość P określona zależnością:

$$P = P_0 + 2P_1 + 4P_2 + 8P_3 \quad (7.3)$$

przedstawia dziesiętny zapis wartości cztero-bitowego sygnału wejściowego. Zależnie od tego sygnału, prąd doprowadzany do wejścia wzmacniacza może zmieniać się skokowo od wartości 0 do wartości $15 I$.

Zapisując równanie napięć dla wzmacniacza otrzymuje się:

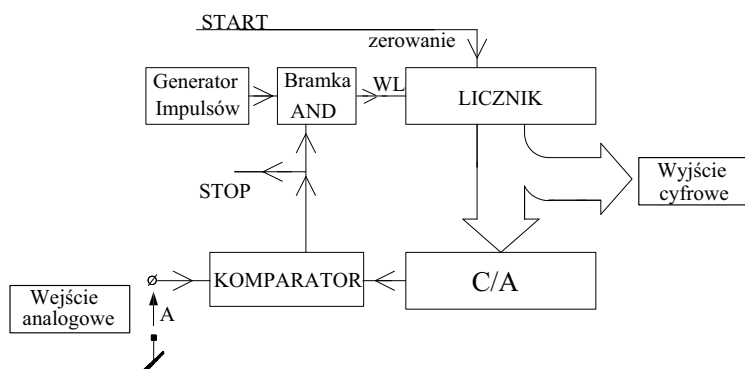
$$U_o = -I_1 R_1 = -U_{wz} \frac{R_1}{16R} P \quad (7.4)$$



Rys. 7.24. Charakterystyka przetwornika cyfrowo-analogowego

co oznacza, że napięcie wyjściowe wzmacniacza, będące jednocześnie wyjściowym sygnałem analogowym przetwornika, jest wprost proporcjonalne do liczbowej wartości cyfrowego sygnału wejściowego. W przedstawionym układzie wyjściowy sygnał analogowy przetwornika przyjmuje jedną z 16 wartości, odpowiednio do 16 wartości cyfrowego sygnału wejściowego. Na rys. 7.24 pokazano charakterystykę przetwornika, charakterystyka jest nieciągła, ponieważ w tym przypadku wejściowy sygnał cyfrowy reprezentuje tylko liczby całkowite od 0 do 15.

W układach praktycznych stosowane są przetworniki c/a co najmniej 8 bitowe, wytwarzające 256 poziomów napięcia, a w miarę potrzeb także 10, 12 i 16 bitowe. Charakterystyki takich przetworników, dzięki zwiększeniu ilości punktów, stają się podobne do linii prostej. W budowie przetworników c/a wykorzystuje się przedstawioną powyżej drabinkę $R - 2R$, ale także i inne koncepcje.



Rys. 7.25. Przetwornik analogowo-cyfrowy

Przetwarzanie odwrotne, to znaczy sygnału analogowego w cyfrowy, jest zwykle bardziej złożone, aniżeli przetwarzanie c/a. Szeroko znanymi przetwornikami a/c są np. woltomierze cyfrowe, o znacznym stopniu komplikacji budowy. Poniżej przedstawia się jedną z prostszych koncepcji przetwornika a/c, z wykorzystaniem przetwornika c/a.

Na rys. 7.25 pokazano schemat blokowy przetwornika a/c, zawierającego przetwornik c/a, komparator sygnałów analogowych, generator impulsów, bramkę i licznik (który może być zastąpiony odpowiednim rejestrem). Przetwarzanie odbywa się przez pewien czas, dlatego proces przetwarzania musi być inicjowany sygnałem "START" (SOC – Start of

Conversion), a zakończenie procesu przetwarzania przetwornik kwituje sygnałem "STOP" (EOC – End of Conversion).

Przetwarzany sygnał analogowy A doprowadzany jest do jednego z wejść komparatora, do drugiego wejścia doprowadzany jest sygnał analogowy z przetwornika c/a. Jeżeli wartość napięcia sygnału wejściowego A jest większa od wartości napięcia z przetwornika c/a, to binarny sygnał wyjściowy komparatora ma wartość 1, jeżeli jest odwrotnie to na wyjściu komparatora występuje 0.

Zewnętrzny sygnał "START", inicjujący przetwarzanie, zeruje licznik. Na wejściu przetwornika c/a pojawiają się więc same zera, co powoduje, że sygnał wyjściowy z tego przetwornika przyjmuje też wartość zero. Jeżeli wejściowy sygnał analogowy A ma wartość większą od zera, to komparator wytworzy sygnał 1, otwierający bramkę pomiędzy generatorem a wejściem liczącym licznika. Już pierwszy impuls zliczony przez licznik zwiększy jego stan, a zatem do przetwornika c/a zostanie przekazany sygnał cyfrowy o wartości odpowiadającej liczbie jeden. Po przetworzeniu na postać analogową przetwornik c/a doprowadzi teraz do komparatora napięcie, o wartości odpowiadającej pierwszemu schodkowi jego charakterystyki (najmniejszy, niezerowy sygnał wyjściowy przetwornika c/a). Jeżeli to napięcie nadal będzie mniejsze od sygnału A, to bramka pozostanie otwarta i licznik zliczy drugi z kolei impuls.

Takie cykle będą się powtarzać aż do momentu, kiedy sygnał wyjściowy przetwornika c/a stanie się większy od sygnału A. Wtedy komparator zmieni swój stan, przekaże sygnał 0 do bramki, zamykając w ten sposób drogę impulsom do licznika. Zmiana stanu komparatora z 1 na 0 stanowi sygnał "STOP", informujący o tym, że w liczniku została już zapisana taka liczba, która przetworzona na postać analogową, jest nieco większa od sygnału A, i może być traktowana jako jego oszacowanie, oczywiście z pewnym błędem. Stan licznika po sygnale "STOP" traktuje się więc jako cyfrową wartość analogowego sygnału A i jest on sygnałem wyjściowym przetwornika a/c.

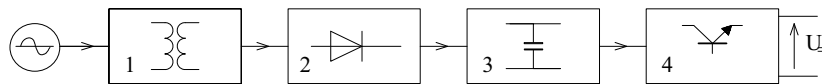
Przetworniki a/c buduje się na podstawie różnych koncepcji, poszukując rozwiązań możliwie dokładnych, aczkolwiek sama zasada kwantowania sygnału ciągłego zawiera w sobie ideę zaokrągleń, oraz rozwiązań możliwie szybko realizujących pojedynczy cykl przetwarzania analogowo-cyfrowego.

8. ZASILACZE

Urządzenia elektroniczne z reguły wymagają zasilania napięciami stałymi. Niewielka część urządzeń, wykonanych jako przenośne (kalkulatory, zegarki) jest zasilana z baterii chemicznych, urządzenia przewoźne (radia, odtwarzacze) są zasilane z akumulatorów ołowiowych pojazdu, ale znaczna większość urządzeń elektronicznych jest wykonywana jako stacjonarne i jest zasilana z sieci elektroenergetycznej za pośrednictwem zasilaczy. Głównym zadaniem zasilaczy jest przetworzenie napięcia przemiennego (220 V, 50 Hz) na napięcie stałe, o wymaganej wartości, zazwyczaj jest to kilka czy kilkanaście woltów.

8.1. Struktury zasilaczy

Zasilacze wykonywane są na podstawie dwóch różnych struktur. Starsze rozwiązanie, stosowane obecnie do układów o mniejszej mocy, zwykle poniżej 25 W, pokazano na rys. 8.1. Podstawowymi podzespołami tego zasilacza są: transformator sieciowy, prostownik, filtr wygładzający napięcie i niekiedy stabilizator napięcia stałego.

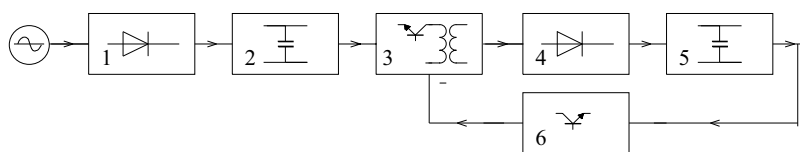


Rys. 8.1. Struktura zasilacza: 1 – transformator, 2 – prostownik, 3 – filtr, 4 – stabilizator

Zadaniem transformatora jest dostosowanie napięcia sieci do potrzeb zasilanego układu i, niekiedy konieczne, odseparowanie galwaniczne układu od sieci prądu przemiennego. Prostownik przetwarza prąd przemienny na prąd jednokierunkowy, nie gwarantując stałości napięcia rozumianej jako wartość stała w czasie. Poprawę kształtu napięcia, w sensie zbliżenia przebiegu napięcia za prostownikiem do linii prostej, powodują filtry wygładzające (filtry prostownikowe nie są identyczne z filtrami liniowymi, rozważanymi w elektrotechnice teoretycznej, gdzie zakłada się, że napięcie wprowadzane do filtru zależy wyłącznie od źródła). W niektórych przypadkach, gdzie niedoskonałość filtracji albo niestałość napięcia sieci elektroenergetycznej przeszkadzają w normalnej pracy zasilanego układu stosuje się stabilizatory, utrzymujące stałą wartość napięcia z dużą dokładnością.

Podstawowym niedostatkim takich zasilaczy są duże gabaryty i masa transformatora sieciowego, a także niewielka sprawność stabilizatorów o działaniu ciągłym. Te czynniki zdecydowały o rozwoju drugiej struktury zasilaczy, zwanej impulsową, pokazanej na rys. 8.2.

Podstawowymi podzespołami zasilacza impulsowego są: prostownik sieciowy, filtr wstępnie wygładzający napięcie, przetwornica zwykle zawierająca transformator, prostownik prądu podwyższonej częstotliwości i filtr końcowy, a także często impulsowa stabilizacja napięcia stałego.



Rys. 8.2. Zasilacz impulsowy: 1 – prostownik, 2 – filtr wstępny, 3 – przetwornica, 4 – prostownik wyższej częstotliwości, 5 – filtr końcowy, 6 – obwody stabilizacji napięcia

Prostownik, wraz z filtrem wstępnym, przetwarzają napięcie sieciowe na napięcie stałe o wartości około 300 V. Takie napięcie zasila przetwornicę, czyli odmianę generatora impulsowego, o dużej sprawności, która wytwarza napięcie w postaci ciągu impulsów prostokątnych. Częstotliwość pracy tego generatora obiera się zwykle powyżej 20 kHz, aby uczynić niesłyszalnym szum wytwarzany w elementach przetwornicy. Napięcie impulsowe doprowadza się do transformatora dla przetworzenia na dogodną wartość. Masa i gabaryty tego transformatora mogą być wydatnie zminiaturyzowane, dzięki pracy z podwyższoną częstotliwością, co jednak zwiększa wymagania jakościowe na materiały magnetyczne.

Po stronie wtórnej transformatora następuje prostowanie i ostateczna filtracja, elementy filtrujące też mogą być zminiaturyzowane dzięki pracy z podwyższoną częstotliwością. W przypadkach koniecznych stosuje się układy stabilizacji napięcia. Są to stabilizatory konwencjonalne, o działaniu ciągłym, ale częściej wprowadza się dodatkowe obwody, pracujące impulsowo i działające na przetwornicę dla skorygowania niepożądanych zmian napięcia wyjściowego.

8.2. Prostowniki

W zasilaczach najczęściej stosowane są trzy układy prostowników, których działanie przedstawia się w dalszych punktach.

Dla uproszczenia analizy prostowników przyjmuje się, że diody prostownicze są idealne, to znaczy, że w kierunku przewodzenia przepływ prądu następuje bez spadku napięcia, a w kierunku wstecznym nie występuje żaden prąd, ponadto w działaniu diod nie ma żadnych opóźnień. Założenie idealności to inaczej przyjęcie, że diody zachowują się jak idealne łączniki zamykające lub otwierające obwody elektryczne.

Chwile przełączeń wyznaczone są przez przebiegi przekształcanych prądów i napięć. Przyjęte uproszczenie jest powszechnie stosowane w energoelektronice przy analizie przekształtników i jest ono dopuszczalne w przeciętnych warunkach ich pracy, jednakże należy o nim pamiętać, aby nie stosować pochopnie wniosków z przedstawionych rozważań do warunków znacznie odbiegających od przeciętnych, w sensie bardzo wielkich lub bardzo małych wartości prądów, napięć czy częstotliwości.

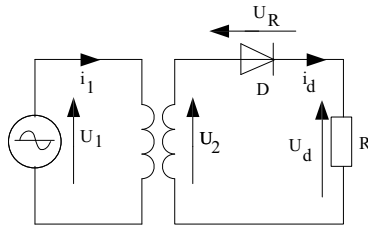
Prostowniki zasilane są napięciem przemiennym, sinusoidalnym i dostarczają do obciążenia prąd jednokierunkowy. Prostownik jest dla sieci prądu przemiennego obciążeniem nieliniowym to znaczy, że kształt prądu pobieranego przez prostownik z sieci często nie jest sinusoidalny, pomimo sinusoidalności napięcia zasilającego ten prostownik. Podobnie

napięcie doprowadzane do obciążenia prostownika też nie jest napięciem stałym, tj. nie jest wartością stałą w czasie.

8.2.1. Prostownik jednopulsowy

Schemat najprostszego układu prostownikowego pokazano na rys. 8.3.

Działanie układu rozpatruje się zakładając, że transformator i dioda są idealne, a obciążeniem prostownika jest tylko rezystor R. Źródło napięcia przemiennego U_1 przedstawia sieć zasilającą prostownik, napięcie tego źródła wynosi:



Rys. 8.3. Prostownik jednopulsowy

$$u_1(t) = U_{m1} \sin(\omega t) \quad (8.1)$$

Transformator sieciowy Tr, o przekładni ϑ , doprowadza napięcie U_2 do obwodu z diodą, przebieg tego napięcia też jest sinusoidalny:

$$u_2(t) = U_{m2} \sin(\omega t); \quad U_{m2} = \frac{U_{m1}}{\vartheta} \quad (8.2)$$

Uwzględniając jednokierunkowe przewodzenie prądu przez diodę łatwo dostrzec, że tylko w jednym półokresie napięcia przemiennego, kiedy jego wartości chwilowe są dodatnie – zgodnie z kierunkiem strzałki na rysunku – może przepływać prąd przez diodę i rezystor. W tym półokresie napięcie wyprostowane u_d , występujące na rezystorze, jest równe napięciu u_2 (dioda idealna – zerowy spadek napięcia przy przepływie prądu). Formalnie zapisuje się to:

$$u_d(t) = u_2(t) \quad \text{dla} \quad 0 < \omega t < \pi \quad (8.3)$$

zatem prąd wynosi:

$$i_d(t) = \frac{U_{m2}}{R} \sin(\omega t) \quad \text{dla} \quad 0 < \omega t < \pi \quad (8.4)$$

W drugim półokresie prąd nie przepływa, a całe napięcie zasilania występuje na diodzie, polaryzując ją wstecznie. Dla tego półokresu zapisuje się

$$u_d(t) = 0, \quad i_d(t) = 0 \quad \text{dla} \quad \pi < \omega t < 2\pi \quad (8.5)$$

natomiast napięcie wsteczne diody wynosi:

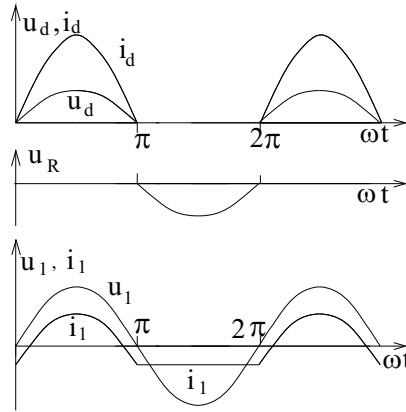
$$u_R(t) = U_{m2} \sin(\omega t) \quad \text{dla} \quad \pi < \omega t < 2\pi \quad (8.6)$$

W kolejnym dodatnim półokresie powtarza się sytuacja z pierwszego półokresu itd. działanie układu jest okresowe, z częstotliwością sieci zasilającej.

Przebiegi prądu i napięć pokazano na rys. 8.4. Ponieważ, w czasie jednego okresu napięcia zasilającego, tylko jeden raz obciążenie jest połączone z tym napięciem (przez połowę okresu), dlatego taki prostownik nazywa się jednopulsowym.

Napięcie wyprostowane (i prąd także) różni się znacznie od idealnego napięcia stałego. Rozkładając je w szereg Fouriera można znaleźć, obok składowej stałej (idealnego na-

pięcia stałego), także nieskończenie wiele wyższych harmonicznych powstałych w procesie prostowania i "psujących" jego skutek.



Rys. 8.4. Prądy i napięcia w prostowniku jednopulsowym

Dla celów technicznych ważne jest określenie wartości składowej stałej napięcia wyprostowanego albo inaczej, wartości średniej tego napięcia. W rozpatrywanym przypadku wynosi ona:

$$U_{d(AV)} = \frac{1}{2\pi} \int_0^{\pi} U_{m2} \sin \omega t \, d\omega t = \frac{U_{m2}}{\pi} \quad (8.7)$$

Oczywiście rozważania na temat napięcia dotyczą też prądu wyprostowanego, którego kształt jest, w tym przypadku, taki sam jak napięcia. Wartość średnia prądu wyprostowanego wynosi:

$$I_{d(AV)} = \frac{U_{m2}}{\pi R} \quad (8.8)$$

Wartość mocy składowych stałych (tylko w pewnych warunkach jest to też moc czynna wydzielana w obciążeniu) określa się jako iloczyn wartości średnich prądu i napięcia w obciążeniu:

$$P_d = I_{d(AV)} U_{d(AV)} = \frac{U_{m2}^2}{\pi^2 R} \quad (8.9)$$

Obciążenie transformatora oblicza się, wyznaczając moce pozorne uzwojeń. Wartość skuteczną prądu wyprostowanego, o kształcie półsinusoidy, płynącego przez uzwojenie wtórne oblicza się według definicji, otrzymując:

$$I_{d(RMS)} = \sqrt{\frac{1}{2\pi} \int_0^{\pi} \frac{U_{m2}^2}{R^2} \sin^2 \omega t \, d\omega t} = \frac{U_{m2}}{2R} \quad (8.10)$$

zatem moc pozorna transformatora, obliczana dla uzwojenia wtórnego, wynosi:

$$S_2 = \frac{U_{m2}}{\sqrt{2}} \frac{U_{m2}}{2R} = \frac{U_{m2}^2}{2\sqrt{2}R} \quad (8.11)$$

Prąd uzwojenia pierwotnego wyznacza się zakładając, że wszystkie składowe harmoniczne prądu (sinusoidalne, przemienne) są transformowane z uwzględnieniem przekładni, ale składowa stała prądu nie jest transformowana, czyli prąd uzwojenia pierwotnego wynosi:

$$i_1(t) = \frac{1}{9} [i_d(t) - I_{d(AV)}] \quad (8.12)$$

Wartość skuteczna tego prądu, obliczana z definicji, wynosi:

$$I_{1(RMS)} = \frac{1}{9} \frac{U_{m2}}{R} \sqrt{\frac{1}{4} - \frac{1}{\pi^2}} \quad (8.13)$$

zatem moc pozorna obliczana dla uzwojenia pierwotnego wynosi:

$$S_1 = 9 \frac{U_{m2}}{R} \frac{1}{9} \frac{U_{m2}}{R} \sqrt{\frac{1}{4} - \frac{1}{\pi^2}} \quad (8.14)$$

Dla ilustracji obliczmy wielkości elektryczne w prostowniku zbudowanym wg schematu rys. 8.3 z transformatorem o przekładni 10, zasilanym z sieci o napięciu 220 V (wartość skuteczna). Napięcie strony wtórnej będzie 22 V, a jego amplituda U_{m2} będzie wynosić około 31,1 V ($1,41 \times 22$), wartość składowej stałej napięcia wyprostowanego $U_{d(AV)}$ będzie około 9,9 V ($31,1/3,14$) i taką właśnie wartość wykaże woltomierz napięcia stałego przyłączony równolegle do obciążenia R. Jeżeli obciążeniem będzie rezystor 100 Ω , to składowa stała prądu obciążenia $I_{d(AV)}$, wskazywana przez miliamperomierz prądu stałego, będzie ok. 99 mA. Iloczyn wskazań przyrządów pomiarowych czyli moc składowych stałych będzie 0,98 W.

Dioda prostownicza będzie obciążona prądem o wartości średniej 99 mA, ale wartości maksymalne, chwilowe, pojawiające się w każdym okresie napięcia zasilającego będą wynosić 311 mA ($31,1/100$) i takie wartości maksymalne powinna dioda wytrzymywać. Dioda powinna wytrzymywać napięcie wsteczne o wartości 31,1 V.

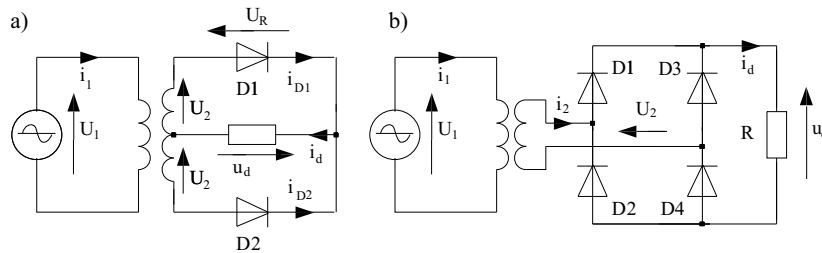
Amperomierz wartości skutecznej (np. elektromagnetyczny) w obwodzie prądu wyprostowanego wskaże ok. 155 mA, a woltomierz wartości skutecznej pokaże napięcie wtórne transformatora 22 V. Iloczyn tych wielkości wyniesie 3,41 VA, co jest ok. 3,5 razy więcej aniżeli moc składowych stałych (0,98 W).

Amperomierz po stronie pierwotnej transformatora będzie wskazywał ok. 12 mA (bez uwzględnienia prądu magnesującego), woltomierz 220 V, zatem moc pozorna doprowadzana do transformatora wynosi ok. 2,64 VA, to jest ok. 2,7 razy więcej aniżeli moc składowych stałych.

Napięcie wyprostowane w rozpatrywanym układzie ma raczej kiepską jakość, transformator wymaga około 3-krotnego przewymiarowania i dlatego ten układ stosuje się wyjątkowo, do zasilania odbiorników o małych wymaganiach i małej mocy, dodając jeszcze elementy filtrujące.

8.2.2. Przetworniki dwupulsowe

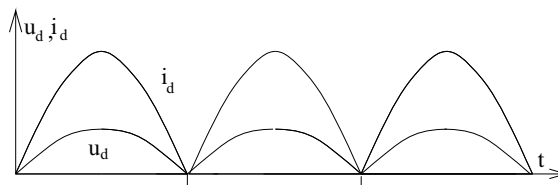
W praktyce częściej stosowane są układy pokazane na rys. 8.5. Pierwszy z nich wymaga konieczności zastosowania transformatora, drugi może być stosowany także bez transformatora. Obydwa układy dwukrotnie w ciągu jednego okresu napięcia zasilającego doprowadzają do obciążenia takie same fragmenty tego napięcia i dlatego nazywane są dwupulsowymi.



Rys. 8.5. Przetworniki dwupulsowe

Działanie układu pokazanego na rys. 8.5a jest podobne do działania dwu układów jednopulsowych, pracujących z przesunięciem o pół okresu napięcia zasilającego. W dodatnim półokresie przewodzi dioda D1, doprowadzając napięcie z jednej połowy uzwojenia wtórnego do obciążenia. W drugim półokresie przewodzi dioda D2. Ponieważ przez uzwojenia wtórne transformatora przepływają prądy w jednym kierunku, to ten układ nazywa się jednokierunkowym.

Na rys. 8.6 pokazano przebiegi prądu i napięcia wyprostowanego w idealizowanym układzie dwupulsowym, jednokierunkowym.



Rys. 8.6. Prąd i napięcie wyprostowane w układzie dwupulsowym

Wartość średnia napięcia wyprostowanego, podobnie jak wartość średnia prądu, jest dwa razy większa aniżeli w układzie jednopulsowym, tj. wynosi

$$U_{d(AV)} = \frac{2U_{m2}}{\pi}, \quad I_{d(AV)} = \frac{2U_{m2}}{\pi R} \quad (8.15)$$

gdzie: U_{m2} – amplituda napięcia jednej połowy uzwojenia wtórnego.

Moc składowych stałych wynosi:

$$P_d = \frac{4U_{m2}^2}{\pi^2 R} \quad (8.16)$$

Moc pozorną obydwu części uzwojenia wtórnego podaje wzór:

$$S_2 = 2 \left(\frac{U_{m2}}{\sqrt{2}} \frac{1}{2} \frac{U_{m2}}{R} \right) = \frac{U_{m2}^2}{\sqrt{2} R} \quad (8.17)$$

Ta moc jest około 1,74 razy większa od mocy składowych stałych.

Analizując szczegółowo strumień magnetyczny, związany z prądami uzwojeń wtórnych, można znaleźć, że w rdzeniu następuje wzajemna kompensacja, zarówno składowych stałych strumienia, jak i wyższych harmonicznych związanych z niesinusoidalnymi prądami obydwu części uzwojenia wtórnego. Skutkiem tego jest, że prąd uzwojenia pierwotnego zawiera wyłącznie składową podstawową, czyli jest sinusoidalny, o częstotliwości napięcia sieci zasilającej. Ponieważ wartości skuteczne napięcia i prądu uzwojenia pierwotnego wynoszą:

$$U_1 = \frac{9U_{m2}}{\sqrt{2}}, \quad I_1 = \frac{1}{9} \frac{U_{m2}}{\sqrt{2} R} \quad (8.18)$$

to wartość mocy pozornej uzwojenia pierwotnego jest:

$$S_1 = \frac{U_{m2}^2}{2R} \quad (8.19)$$

Ta wartość jest około 1,23 razy większa od mocy składowych stałych.

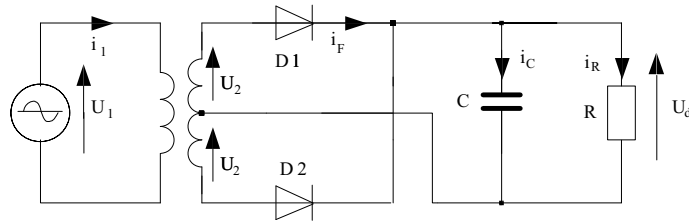
Układ, pokazany na rys. 8.5b, zwany układem mostkowym, także dostarcza napięcia wyprostowanego, dwupulsowego o kształcie pokazanym na rys. 8.6. W układzie z obciążeniem rezystancyjnym zawsze przewodzi dwie diody. Jeżeli w pierwszym półokresie prąd przepływa przez diodę D1, obciążenie i diodę D4, to w drugim półokresie zamyka się obwód: dioda D3, obciążenie, dioda D2.

Wartości średnie prądu i napięcia wyprostowanego opisują wzory (8.15), moc składowych stałych podaje wzór (8.16). Przy obciążeniu rezystancyjnym prąd zarówno uzwojenia wtórnego, jak i pierwotnego, ma kształt sinusoidalny a moc pozorna każdego z uzwojeń wynosi $(U_{m2})^2/2R$, czyli jest 1,23 razy większa od mocy składowych stałych.

Układy prostowników dwupulsowych są częściej stosowane, bo kształt napięcia wyprostowanego jest lepszy, w sensie zbliżenia do idealnego napięcia stałego, aniżeli w układzie jednopulsowym, a stopień przewymiarowania transformatora mniejszy. Niemniej kształt napięcia nie jest zadowalający i zwykle stosuje się różne środki techniczne, ażeby kształt przetwarzanego napięcia bardziej zbliżyć do idealnego napięcia stałego.

8.2.3. Wygładzanie napięcia wyprostowanego

Działanie wygładzające napięcie wyprostowane wykazuje kondensator przyłączony równoległe do obciążenia. Nie jest to działanie ściśle filtrujące, ze względu na to, że kondensator modyfikuje sposób pracy prostownika, aczkolwiek skutki takiego wygładzania można wyrażać w pojęciach właściwych dla teorii filtrów.



Rys. 8.7. Prostownik z obciążeniem RC

Wpływ kondensatora na działanie prostownika rozpatrzmy na przykładzie prostownika dwupulsowego, schemat pokazano na rys. 8.7. Dla uproszczenia rozważań przyjmuje się, że załączenie napięcia zasilającego następuje w momencie przejścia sinusoidy tego napięcia przez oś odciętych, a kondensator nie był naładowany. Odpowiednia dioda, np. D1, rozpoczynając przewodzenie prądu, łączy obciążenie RC z połową uzwojenia wtórnego transformatora. Napięcie na rezystorze R oraz na kondensatorze C są jednakowe i równe napięciu z transformatora czyli:

$$u_R(t) = u_C(t) = U_{m2} \sin(\omega t) \quad \text{dla} \quad 0 < \omega t < \omega t_k \quad (8.20)$$

(sens granicy t_k jest wyjaśniony poniżej). Prądy płynące przez elementy obciążenia oblicza się jako:

$$i_R = \frac{u_R}{R}, \quad i_C = C \frac{du_C}{dt} \quad (8.21)$$

Prąd diody i_F , będący sumą tych dwu prądów, wynosi więc:

$$i_F = U_{m2} \left(\omega C \cos \omega t + \frac{1}{R} \sin \omega t \right) = \frac{U_{m2}}{R} \sqrt{\text{tg}^2 \phi + 1} \sin(\omega t + \phi) \quad (8.22)$$

gdzie:

$$\text{tg} \phi = \omega RC$$

Jak wynika z wzoru (8.22), prąd diody ma też kształt sinusoidalny, ale jest przesunięty fazowo o kąt ϕ względem sinusoidy napięcia i dlatego rozpoczyna się skokowo, jak pokazano na rys. 8.7. Ze wzoru (8.22) można obliczyć, że dodatni prąd diody, który dla $\omega t = \pi/2 - \phi$ przyjmuje wartość maksymalną, następnie maleje. W chwili t_k prąd przyjmuje wartość zero, co oznacza, że dioda przestaje być w stanie przewodzenia i jej prąd, przez pewien czas będzie wynosił zero. Moment zakończenia przewodzenia diody wyznacza się ze wzoru (8.22), rozwiązując równanie $i_F(\omega t_k) = 0$

$$\omega t_k + \phi = \pi \quad (8.23)$$

W momencie zakończenia przepływu prądu przez diodę (który to moment przypada pomiędzy szczytem sinusoidy napięcia przy $\omega t = \pi/2$, a przecięciem tej sinusoidy z osią przy $\omega t = \pi$, zależnie od wartości R i C) napięcie na kondensatorze wynosi:

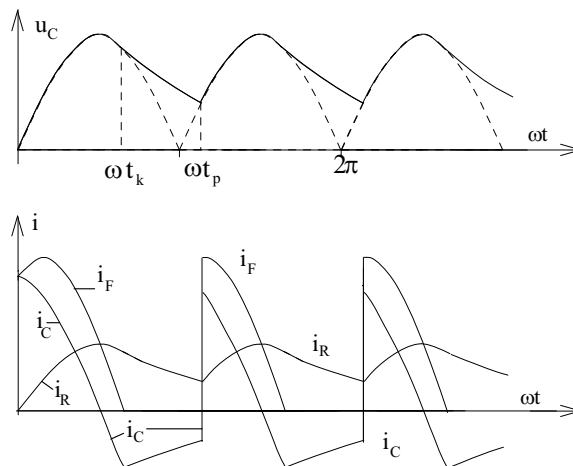
$$U_{Ck} = U_{m2} \sin \phi \quad (8.24)$$

Od momentu t_k nieprzewodząca dioda D1 przerywa połączenie między uzwojeniem wtórnym transformatora a obciążeniem, druga dioda pozostaje nadal w stanie nieprzewodzenia. Dlatego po chwili t_k następuje rozładowanie kondensatora przez równoległy rezystor, czyli napięcie na zespole RC maleje wg krzywej wykładniczej.

$$u_C(\omega t) = U_{Ck} \exp\left(\frac{\omega t - \omega t_k}{\omega RC}\right) \quad (8.25)$$

W tym etapie pracy prostownika prąd płynący przez obciążenie tj. rezystor (i kondensator) też ma przebieg wykładniczy. W pewnym momencie kolejnego półokresu narastające napięcie zasilania zrówna się z malejącym napięciem na kondensatorze, oznacza to zanik wstecznej polaryzacji diody D2, która teraz może połączyć drugie uzwojenie wtórne transformatora z obciążeniem. Niestety wyznaczenie analityczne tego momentu, oznaczonego na rys. 8.8 jako t_p , jest niemożliwe, wiadomo tylko, że przypada on pomiędzy π a $1,5\pi$.

Od tego momentu napięcie na obciążeniu ponownie staje się równe napięciu zasilającemu, jak pokazano na rys. 8.8, a prądy ponownie przyjmują przebieg sinusoidalny i cosinusoidalny na czas mniejszy od połowy okresu napięcia zasilającego.



Rys. 8.8. Prądy i napięcia w prostowniku z obciążeniem RC

Jak wynika z rysunku, napięcie wyprostowane, acz nie jest idealne, to jest jednak bardziej zbliżone do idealnego, aniżeli w przypadku braku kondensatora filtrującego. Skuteczność wygładzania napięcia zależy od iloczynu $\omega RC = \tan\phi$, jeżeli jest on znacznie większy od jedności, to zmniejszanie się napięcia na kondensatorze, a zatem i na obciążeniu R, w czasie kiedy diody nie przewodzą, jest stosunkowo niewielkie. Wartość średnia napięcia wyprostowanego w tym przypadku jest bliska amplitudzie napięcia prostowanego.

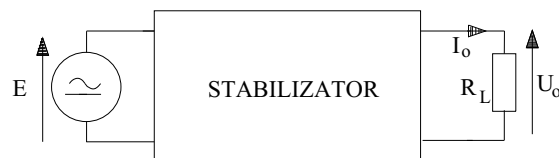
Warto zauważyć, że prądy diod mają kształt tylko fragmentów przebiegów sinusoidalnych, są więc zbliżone bardziej do przebiegów impulsowych. Im bardziej wygładzone

jest napięcie, tym czas przewodzenia diod jest krótszy, a szczytowa wartość prądu większa, co jeszcze bardziej upodabnia przebieg prądu do impulsowego.

Na początku opisu założono, że układ jest włączany do pracy w szczególnym momencie, kiedy chwilowa wartość napięcia zasilającego jest zero. W praktyce taki warunek może być spełniony tylko przypadkowo, zwykle układ zostaje załączony przy niezerowej wartości napięcia zasilającego. Powstaje wówczas sytuacja połączenia nienaładowanego kondensatora ze źródłem napięcia – warunki zbliżone do zwarcia. Analiza takiego stanu wymaga dokładniejszego modelu transformatora i nie będzie tutaj przeprowadzana. Warto jednak wiedzieć, że szczególnie przy większych mocach, w momencie załączenia może pojawić się duży prąd ładowania kondensatora, o charakterze udarowym, stanowiący zagrożenie dla diod i innych elementów układu. W takich przypadkach stosowane są elementy indukcyjne lub rezystancyjne dla ograniczenia tego prądu.

8.3. Stabilizatory napięcia stałego

Stabilizator napięcia stałego jest nieliniowym czwórnikiem, instalowanym pomiędzy niestabilnym źródłem napięcia stałego a obciążeniem wymagającym zasilania napięciem stałym, o ustalonej i stabilnej wartości jak pokazano na rys. 8.9.



Rys. 8.9. Włączenie stabilizatora

Niestabilność źródeł napięcia stałego, realizowanych w praktyce najczęściej jako prostowniki, jest spowodowana niestabilnością napięcia w sieci prądu przemiennego i niedoskonałością filtrów dolnoprzepustowych, wygładzających napięcie wyprostowane. Jeżeli jako źródła napięcia stałego stosowane są źródła elektrochemiczne powodem niestabilności są procesy rozładowania i starzenia.

Drugim powodem niestabilności napięcia na zaciskach źródła rzeczywistego, o niezerowej rezystancji wewnętrznej, są zmiany prądu pobieranego z tego źródła.

Temperatura otoczenia także zwykle wpływa na właściwości źródła napięcia stałego (znane problemy uruchamiania silników spalinowych przy mrozach).

Jeżeli stosuje się stabilizator, to można oczekiwać podobnych oddziaływań czynników zakłócających na napięcie stabilizowane U_o . Traktując U_o jako funkcję wartości napięcia zasilającego E , prądu obciążenia I_o i temperatury otoczenia T ; to jest zapisując funkcję $U_o = f(E, I_o, T)$, można opisać niestałość napięcia stabilizowanego, korzystając z pojęcia różniczki zupełnej, zmodyfikowanej do przyrostów skończonych. Zmianę napięcia stabilizowanego ΔU_o traktuje się więc jako liniową kombinację efektów działania zmian (Δ) argumentów funkcji U_o :

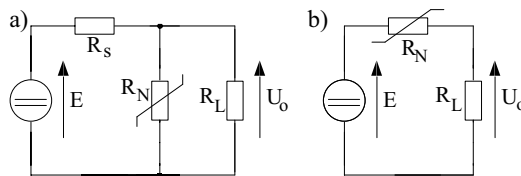
$$\Delta U_o = \frac{\partial U_o}{\partial E} \Delta E + \frac{\partial U_o}{\partial I_o} \Delta I_o + \frac{\partial U_o}{\partial T} \Delta T \quad (8.26)$$

Współczynnikami wrażliwości napięcia stabilizowanego na poszczególne zakłócenia są pochodne cząstkowe obliczane względem tych zakłóceń. Pierwsza z nich przedstawia wrażliwość napięcia stabilizowanego na zmiany napięcia zasilającego i może być nazwana napięciowym współczynnikiem stabilizacji. Druga pochodna opisuje wrażliwość napięcia stabilizowanego na zmiany prądu obciążenia, czyli przedstawia rezystancję wyjściową stabilizatora albo może być nazwana prądowym współczynnikiem stabilizacji. Trzecia pochodna opisuje podobnie wpływ temperatury i może być nazwana temperaturowym współczynnikiem stabilizacji. Całkowita niestabilność napięcia stabilizowanego zależy od wartości zakłóceń i współczynników, które powinny być możliwie małe.

Znane i stosowane są różne układy stabilizatorów napięcia stałego, które można podzielić na:

- stabilizatory szeregowe i równoległe,
- stabilizatory parametryczne i stabilizatory ze sprzężeniem zwrotnym,
- stabilizatory o działaniu ciągłym i impulsowe.

Na rys. 8.10 pokazano zasadę budowy stabilizatorów równoległych i szeregowych.



Rys. 8.10. Stabilizator: równoległy (a) i szeregowy (b)

W stabilizatorze równoległym element nieliniowy jest umieszczony równoległe z obciążeniem. Działanie układu równoległego polega na tym, aby ewentualne zmiany napięcia wyjściowego U_o były kompensowane zmiennym spadkiem napięcia na rezystorze szeregowym R_s . Wartość tego spadku napięcia zależy od zmian prądu przepływającego ze źródła E przez rezystor szeregowy i element nieliniowy. Na przykład wzrost napięcia stabilizowanego, ponad wartość zadaną, powinien wywołać taki przyrost prądu w elemencie nieliniowym, aby zwiększony spadek napięcia na rezystorze szeregowym spowodował obniżenie napięcia wyjściowego do wartości zadanej.

W stabilizatorze szeregowym element nieliniowy, połączony szeregowo z obciążeniem, powinien tak zmieniać swoją rezystancję, aby napięcie wyjściowe U_o pozostawało możliwie stałe.

W stabilizatorach parametrycznych wykorzystuje się naturalną nieliniowość charakterystyki prądowo-napięciowej na przykład diody Zenera lub podobnego elementu.

W stabilizatorach ze sprzężeniem zwrotnym tworzy się zamknięty układ regulacyjny, stosując wzorzec napięcia, człon porównujący napięcie na obciążeniu z wzorcowym oraz

człon wykonawczy, o charakterze regulowanej rezystancji. Zarówno układ ze sprzężeniem, jak i parametryczny, może być typu szeregowego lub równoległego.

Człon wykonawczy w stabilizatorze może pracować w sposób ciągły, jest to wówczas zwykły tranzystor mocy, ewentualnie rozbudowany do postaci zestawnej, czyli układu Darlingtona. Można też zastosować jako człon wykonawczy układ impulsowy, co wymaga rozbudowy stabilizatora o elementy filtracyjne, ale wydatnie zwiększa sprawność układu.

8.3.1. Stabilizator równoległy, parametryczny

Na rys. 8.11. pokazano przykład schematu stabilizatora równoległego, parametrycznego z diodą Zenera.

Warunkiem poprawnej pracy stabilizatora jest zapewnienie przepływu odpowiedniego prądu I_Z przez diodę Zenera, co gwarantuje utrzymanie prawie stałego spadku napięcia U_Z na niej.

Aby uzyskać takie warunki, analizuje się podstawowe równanie stabilizatora:

$$E = R_s(I_o + I_Z) + U_Z \quad (8.27)$$

dla skrajnych wartości parametrów, zakładając zmienność napięcia zasilania E , prądu obciążenia I_o ewentualnie dopuszczalny rozrzut wartości napięcia Zenera U_Z oraz tolerancje rezystora szeregowego R_s .

Wynikiem analizy jest określenie wartości minimalnych i maksymalnych prądu diody i porównanie ich z dopuszczalnymi wartościami katalogowymi. Jeżeli zostanie zapewniona praca diody na stromo opadającym odcinku charakterystyki, odpowiadającym stabilizacji napięcia, to można wprowadzić uproszczony model liniowy diody, ważny tylko dla pracy w tym stanie.

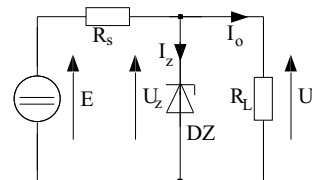
$$U_Z = U_{Z0} + r_Z I_Z \quad (8.28)$$

gdzie: U_{Z0} – umowna wartość napięcia stabilizacji, wyznaczona w punkcie przecięcia prostej aproksymującej charakterystykę z osią odciętych,

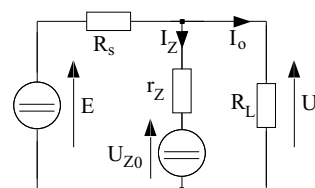
r_Z – nachylenie tej prostej, symbolizujące wartość rezystancji dynamicznej diody.

Wprowadzając schemat, opisany modelem diody, do schematu stabilizatora otrzymuje się schemat zastępczy stabilizatora, pokazany na rys. 8.12.

Na podstawie tego schematu można zapisać równanie, wiążące wartość napięcia wyjściowego, stabilizowanego z wielkości elektrycznymi zakłócającymi tę wartość i z parametrami stabilizatora. To równanie, podobnie jak równanie (8.28), jest ważne tylko w ograniczonym zakresie wartości prądów i napięć.



Rys. 8.11. Stabilizator parametryczny, równoległy



Rys. 8.12. Schemat zastępczy stabilizatora z diodą Zenera

$$U_o = \frac{r_Z}{R_s + r_Z} E - \frac{R_s r_Z}{R_s + r_Z} I_o + \frac{R_s}{R_s + r_Z} U_{Z0} \quad (8.29)$$

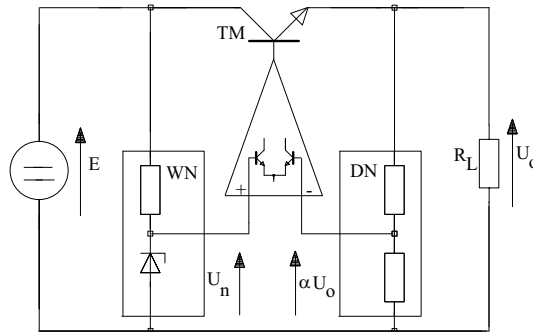
Analiza równania (8.29) pozwala określić główne parametry opisujące cechy jakościowe każdego stabilizatora. Wartości współczynników stabilizacji, dla rozpatrywanego układu równoległego z diodą Zenera, oblicza się więc jako pochodne cząstkowe, różniczkując równanie (8.29):

$$\frac{\partial U_o}{\partial I_o} = -\frac{R_s r_Z}{R_s + r_Z}; \quad \frac{\partial U_o}{\partial E} = \frac{r_Z}{R_s + r_Z}; \quad \frac{\partial U_o}{\partial T} = \frac{R_s}{R_s + r_Z} \frac{\partial U_{Z0}}{\partial T} \quad (8.30)$$

Powyższe wzory można zinterpretować następująco: reakcja napięcia wyjściowego stabilizatora na zmianę prądu obciążenia jest taka, jak gdyby wyjście stabilizatora było zaciskami rzeczywistego źródła napięciowego, o rezystancji wewnętrznej równej rezystancji równoległego połączenia R_s i r_Z . Dla zmian napięcia zasilającego stabilizator zachowuje się jak dzielnik napięcia o rezystorach r_Z i R_s . Wpływ temperatury na napięcie stabilizowane jest nieco mniejszy od wpływu temperatury na wartość napięcia U_{Z0} zastosowanej diody Zenera.

8.3.2. Ciągły stabilizator szeregowy ze sprzężeniem zwrotnym

Na rys. 8.13 pokazano uproszczony schemat stabilizatora szeregowego, ze sprzężeniem zwrotnym, o działaniu ciągłym. Układ zawiera: wzorzec napięcia stałego WN, pomiarowy dzielnik napięcia DN, wzmacniacz błędów i szeregowy tranzystor mocy TM.



Rys. 8.13. Uproszczony schemat stabilizatora szeregowego

Działanie stabilizatora polega na takim sterowaniu tranzystorem TM, aby prąd płynący przez obciążenie osiągnął wartość zapewniającą żądane napięcie wyjściowe. W tym celu porównuje się aktualną wartość napięcia wyjściowego U_o (dokładniej: część tej wartości αU_o określoną w dzielniku oporowym) z wartością wzorcową U_n , wytworzoną w stabilizatorze, różnica tych napięć, stanowiąca błąd stabilizacji, zostaje wzmocniona i powoduje taką korekcję wysterowania tranzystora szeregowego i zmianę wartości prądu płynącego do obciążenia, aby błąd stabilizacji sprowadzić do minimum.

Przerysowując schemat stabilizatora do postaci pokazanej na rys. 8.14 łatwiej zauważyć, że jest to w istocie wzmacniacz prądu stałego, złożony ze stopnia wstępnego, nazwanego uprzednio wzmacniaczem błędów i stopnia mocy z tranzystorem TM w układzie wtórnika emiterowego.

Sygnałem wejściowym wzmacniacza jest napięcie źródła wzorcowego U_n . Zapisując współczynnik wzmocnienia tego wzmacniacza jako:

$$K_u = \frac{U_o}{U_n - U_1} \quad (8.31)$$

oraz zapisując sprzężenie zwrotne jako warunek:

$$U_1 = \alpha U_o \quad (8.32)$$

przy czym $0 < \alpha < 1$, można znaleźć, że:

$$U_o = U_n \frac{K_u}{1 + \alpha K_u} \quad (8.33)$$

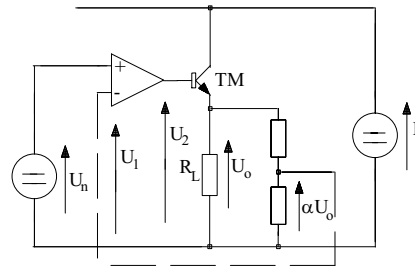
co oznacza, że napięcie wyjściowe, stabilizowane, zależy od napięcia wzorca, podziału w dzielniku oporowym i całkowitego wzmocnienia wzmacniacza. Jeżeli wzmocnienie tego wzmacniacza jest duże, a podział w dzielniku niezbyt mały tak, że $\alpha K_u \gg 1$ to wzór (8.33) przyjmuje postać uproszczoną:

$$U_o \approx \frac{U_n}{\alpha} \quad (8.34)$$

Wzór (8.34) jasno wykazuje celowość stosowania dzielnika oporowego pozwalającego, przez zmianę wartości rezystorów w dzielniku, na zmianę współczynnika α , co z kolei umożliwia regulację wartości napięcia stabilizowanego U_o przy ustalonej wartości napięcia wzorca U_n . Ponadto ze wzoru wynika, że w rozpatrywanym układzie napięcie stabilizowane zawsze jest większe od napięcia wzorcowego ($0 < \alpha < 1$). Górne ograniczenie wartości napięcia U_o stwarza formalny warunek powyższego uproszczenia $\alpha K_u \gg 1$ z czego wynika, że $U_o \ll U_n K_u$. W praktyce jednak, górną granicę napięcia stabilizowanego narzuca warunek zapewnienia pracy tranzystora TM w stanie aktywnym, czyli wyjściowe napięcie stabilizowane musi być mniejsze o co najmniej 1 do 3 woltów od napięcia zasilania E .

Parametry jakościowe stabilizatora wyznacza się na podstawie wzoru (8.26), obliczając pochodne cząstkowe z funkcji (8.33). Wymaga to jednak znacznie dokładniejszego opisu stabilizatora, wskazującego na zależność współczynnika wzmocnienia K_u od napięcia zasilania, prądu obciążenia i temperatury a także podobnych uzależnień dla napięcia wzorca. Jeżeli spełnione są warunki uproszczenia (8.34), to o jakości stabilizatora decyduje jakość wzorca napięcia i dzielnika.

Poważnym zagrożeniem dla stabilizatora szeregowego może być przeciążenie prądowe, a szczególnie zwarcie. Wartość mocy traconej w tranzystorze szeregowym można oszacować na:



Rys. 8.14. Stabilizator szeregowy jako wzmacniacz

$$P \approx I_0 (E - U_0) \quad (8.35)$$

Przy zwarciu wyjścia stabilizatora napięcie na tranzystorze osiąga maksymalną wartość E , a jednocześnie sygnał sprzężenia zwrotnego informuje o zbyt niskim napięciu na wyjściu, co powoduje intensywne sterowanie tranzystora TM w kierunku zwiększania prądu I_0 . Skutkiem tego procesu może być przegrzanie i uszkodzenie tranzystora szeregowego.

Jako zabezpieczenie przed przeciążeniem lub zwarciem wprowadza się dodatkowo nieliniowe sprzężenie zwrotne prądowe, działające po przekroczeniu określonej wartości prądu. Takie sprzężenie, spełniając funkcję nadrzędną nad sprzężeniem napięciowym, powoduje przejście stabilizatora od trybu pracy z utrzymywaniem stałego napięcia do trybu pracy z utrzymywaniem stałego prądu (albo nawet zmniejszaniem prądu).

Ograniczenia prądowe nie gwarantują pełnej ochrony stabilizatora przed przegrzaniem, które może być wywołane niekorzystnym bilansem mocy strat i skuteczności chłodzenia. Dlatego w bardziej rozbudowanych układach stabilizatorów wprowadza się dodatkowo zabezpieczenie termiczne. Wykorzystuje się tu np. różne współczynniki temperaturowe takich elementów jak tranzystory i diody Zenera, tworząc układy scalone ze stabilizatorem w jednym kryształu, w których po nagraniu całości do określonej temperatury, końcowy tranzystor powoduje zmniejszenie prądu stabilizatora.

Stabilizatory szeregowo wytwarzane są jako układy scalone w dwu podstawowych wariantach jako stabilizatory specjalizowane i stabilizatory uniwersalne. W pierwszym przypadku producent umieszcza wszystkie elementy stabilizatora w układzie scalonym, narzucając w ten sposób wartość napięcia stabilizowanego jak np. 5 V, 12 V lub 15 V i pozostaje parametry stabilizatora. Zadaniem użytkownika pozostaje dołączenie filtrów pojemnościowych i zapewnienie odprowadzania ciepła. W drugim przypadku producent scalonego stabilizatora pozostawia użytkownikowi możliwość dostosowania parametrów stabilizatora do indywidualnych potrzeb. Do układu scalonego, oprócz niezbędnych kondensatorów, użytkownik dołącza rezystory dzielnika, ustalając potrzebną wartość napięcia stabilizowanego. Podobnie można rozbudować, istniejący w układzie scalonym, tranzystor szeregowy do postaci zestawnej (układu Darlingtona) powiększając zakres prądu obciążenia. Istotną modyfikacją może być przeniesienie dzielnika napięcia z wyjścia stabilizatora na wewnętrzny wzorzec napięcia i doprowadzenie całkowitej wartości napięcia stabilizowanego jako sygnału sprzężenia zwrotnego. Taki zabieg zmienia postać podstawowego wzoru (8.33) i pozwala budować stabilizatory o bardzo małych wartościach napięcia stabilizowanego.

8.4. Przetwornice napięcia stałego

Jednym z podzespołów zasilaczy impulsowych są układy przetwarzające napięcie stałe, otrzymywane z prostownika sieciowego (ok. 300 V), na napięcie stałe o innej wartości (zwykle znacznie mniejszej np. 5 V lub 12 V). Z różnych znanych układów przekształtników napięcia stałego na napięcie stałe [2] w zasilaczach stosuje się zwykle układy zawierające transformator, który spełnia tutaj dwa zadania: wstępne dostosowanie wartości napięć oraz separacja galwaniczna zasilanych urządzeń od sieci elektroenergetycznej. Od tych

układów zwykle wymaga się też możliwości sterowania wartością napięcia wyjściowego dla uzyskania dokładnej stabilizacji.

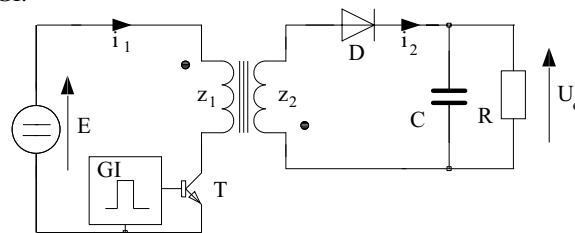
W dalszej części przedstawia się wybrane przykłady przekształtników, zwanych tradycyjnie przetwornicami napięcia stałego, stosowanych w zasilaczach. Należy zwrócić uwagę na pewną osobliwość tych przekształtników, jaką jest współpraca transformatora ze źródłem napięcia stałego. Takie współdziałanie jest możliwe, jednak nie wszystkie wnioski, wynikające z analizy konwencjonalnego sposobu użytkowania transformatora, mogą być tu bezkrytycznie stosowane.

Klasyfikacje przetwornic oparte są na różnych kryteriach. Każda przetwornica, wytwarzając napięcie przemiennie, pracuje cyklicznie czyli okresowo, jeżeli w czasie jednego okresu można wyróżnić podokres pobierania energii ze źródła i drugi podokres, w którym następuje przykazywanie energii do obciążenia, to mówi się o przetwornicy dwutaktowej, w przeciwnym przypadku określa się ją jako jednotaktową.

Inny podział jest oparty o kierunki prądów, pobieranego i oddawanego, odniesione względem uzwojeń transformatora i uwzględniające zmienność strumienia magnetycznego w rdzeniu. Dla dokonywania podziałów wykorzystywane są i inne cechy budowy przetwornic.

8.4.1. Przetwornica dwutaktowa, przeciwbieżna

Na rys. 8.15 pokazano schemat przetwornicy dwutaktowej, przeciwbieżnej (zaporo-wej, przeciwfazowej, flyback DC–DC converter). Obciążenie przetwornicy przedstawia rezystor R , kondensator C spełnia funkcję elementu wygładzającego napięcie wyjściowe. Tranzystor T pracuje jako łącznik (klucz), sterowany z oddzielnego generatora impulsów prostokątnych GI .

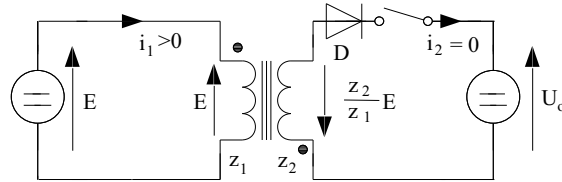


Rys. 8.15. Schemat przetwornicy dwutaktowej

Impuls z generatora, o czasie trwania t_i , wprowadza tranzystor w stan nasycenia, co zamyka obwód, łącząc napięcie E do uzwojenia pierwotnego transformatora. W czasie przerwy między impulsami, na czas t_p , uzwojenie pierwotne zostaje odłączone od źródła E przez nieprzewodzący tranzystor, ale występuje wówczas przepływ prądu w uzwojeniu wtórnym, przez diodę D , uzupełniający ładunek kondensatora. Po przerwie następuje kolejny impuls i cykl się powtarza, okres powtarzania impulsów T wynosi $t_i + t_p$.

W uproszczonej analizie zakłada się, że tranzystor jest idealnym łącznikiem, transformator jest idealnym sprzężeniem magnetycznym dwu cewek i ponadto, że stała czasu RC jest znacznie większa od okresu impulsowania. Z ostatniego założenia wynika, że w czasie

jednego okresu, zmiana napięcia na kondensatorze (i rezystorze reprezentującym obciążenie) jest pomijalnie mała. Można więc, na czas jednego okresu, elementy RC zastąpić źródłem napięcia stałego o wartości U_o . Przy tych założeniach określa się proste schematy zastępcze, odpowiadające obydwu stanom tranzystora, schematy te pokazano na rys. 8.16 i 8.18.



Rys. 8.16. Schemat zastępczy przetwornicy; etap przewodzenia tranzystora

Uproszczoną analizę przeprowadza się dla stanu ustalonego, jaki wystąpi po znacznej liczbie cykli pracy, pomijając procesy przejściowe, towarzyszące załączeniu przetwornicy, zmianom jej obciążenia lub zasilania. Analizuje się kolejno dwa schematy zastępcze, obowiązujące przy tranzystorze przewodzącym i zatkanym.

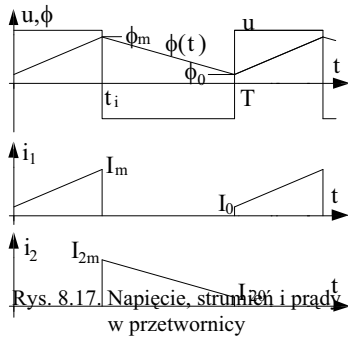
1. Przetwornica przy tranzystorze przewodzącym (rys. 8.16)

Napięcie na uzwojeniu pierwotnym transformatora jest stałe i wynosi E , to znaczy, że pochodna strumienia w rdzeniu jest stała czyli strumień narasta liniowo od wartości początkowej ϕ_0 :

$$E = z_1 \frac{d\phi}{dt}, \quad \phi = \frac{E}{z_1} t + \phi_0 \quad (8.36)$$

napięcie uzwojenia wtórnego wynosi:

$$U_2 = z_2 \frac{d\phi}{dt} = \frac{z_2}{z_1} E \quad (8.37)$$



Rys. 8.17. Napięcie, strumień i prąd w przetwornicy

Zwrot tego napięcia jest przeciwny do kierunku przewodzenia diody, co powoduje, że dioda jest spolaryzowana wstecznie i prąd w uzwojeniu wtórnym nie płynie. W tej części cyklu pracy strumień magnetyczny związany jest wyłącznie z prądem uzwojenia pierwotnego (i tranzystora), który zatem wynosi:

$$i_1(t) = \frac{E}{L_1} t + I_0 \quad (8.38)$$

gdzie: L_1 – indukcyjność uzwojenia pierwotnego,
 I_0 – wartość prądu i_1 odpowiadająca strumieniowi ϕ_0 .

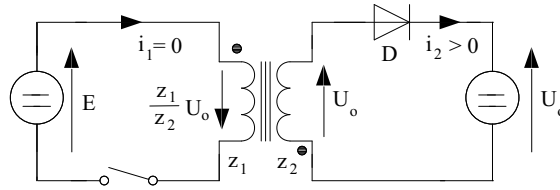
Przy końcu impulsu, dla $t = t_i$, zarówno strumień jak i prąd osiągają wartości maksymalne, φ_m oraz I_m odpowiednio, jak pokazano na rys. 8.17. Wartość maksymalna strumienia wynosi:

$$\varphi_m = \frac{E}{z_1} t_i + \varphi_0 \quad (8.39)$$

a wartość prądu I_m wynosi:

$$i_1(t_i) = I_m = \frac{E}{L_1} t_i + I_0 \quad (8.40)$$

2. Przetwornica przy tranzystorze zatkany (rys. 8.18)



Rys. 8.18. Schemat zastępczy przetwornicy; etap nieprzewodzenia tranzystora

Po przerwaniu obwodu zasilania uzwojenia pierwotnego, wzbudzony uprzednio strumień magnetyczny w rdzeniu będzie zanikał, co powoduje zmianę biegunowości indukowanych napięć. Dioda w obwodzie wtórnym zaczyna przewodzić prąd i_2 . Dla zachowania stałości strumienia, w pierwszym momencie po wyłączeniu tranzystora, prąd i_2 przyjmie wartość maksymalną I_{2m} , wynikającą z zależności:

$$z_1 I_m = z_2 I_{2m} \quad (8.41)$$

Z założeń wynika, że pojemność kondensatora jest na tyle duża, iż ładowanie go prądem i_2 powoduje pomijalnie mały wzrost napięcia na kondensatorze. Skutkiem tego, oraz pominięcia spadku napięcia na przewodzącej diodzie, jest przyjęcie, że napięcie na uzwojeniu wtórnym transformatora jest stałe i wynosi U_o . Oznacza to, że pochodna strumienia w rdzeniu jest stała, czyli strumień maleje liniowo od wartości maksymalnej φ_m :

$$U_o = -z_2 \frac{d\varphi}{dt}; \quad \varphi(t) = -\frac{U_o}{z_2} t + \varphi_m \quad (8.42)$$

Napięcie uzwojenia pierwotnego wynosi:

$$U_1 = -z_1 \frac{d\varphi}{dt} = -\frac{z_1}{z_2} U_o \quad (8.43)$$

W tej części cyklu pracy strumień magnetyczny związany jest wyłącznie z prądem uzwojenia wtórnego (i diody), który wynosi zatem:

$$i_2(t) = -\frac{U_o}{L_2} t + I_{2m} \quad (8.44)$$

L_2 oznacza indukcyjność uzwojenia wtórnego, a I_{2m} jest to wartość prądu odpowiadająca strumieniowi φ_m . Związek pomiędzy indukcyjnościami L_1 a L_2 wynika z budowy transformatora:

$$L_2 = \left(\frac{z_2}{z_1} \right)^2 L_1 \quad (8.45)$$

Przy końcu przerwy między impulsami, dla $t = t_p$, zarówno strumień jak i prąd osiąga wartości minimalne, jak pokazano na rys. 8.17.

W stanie ustalonym pracy układu, wartość minimalna strumienia, występująca na końcu przerwy pomiędzy impulsami, czyli dla $t = t_p$, jest równa wartości początkowej φ_0 , występującej na początku impulsu:

$$\varphi(t_p) = -\frac{U_o}{z_2} t_p + \varphi_m = \varphi_0 \quad (8.46)$$

Porównując ten wzór z wzorem (8.39) znajduje się ważną relację dla rozważanej przetwornicy:

$$\frac{U_o}{E} = \frac{z_2}{z_1} \frac{t_i}{t_p} \quad (8.47)$$

Z zależności tej wynika, że stosunek napięcia zasilającego i wyjściowego, określa zarówno przekładnia zwojowa transformatora, jak i stosunek czasu impulsu do czasu przerwy między impulsami. Oznaczając stosunek czasu impulsu t_i do okresu impulsowania $T = t_i + t_p$ jako współczynnik wypełnienia impulsu D można zapisać wzór (8.47) w wartościach względnych:

$$\frac{U_o}{E} = \frac{z_2}{z_1} \frac{D}{1 - D} \quad (8.48)$$

Przy końcu przerwy między impulsami, dla $t = t_p$, nie tylko strumień ale i prąd i_2 osiąga wartości minimalną I_{20} , jak pokazano na rys. 8.17. W stanie ustalonym pracy układu, wartość minimalna strumienia, występująca na końcu przerwy pomiędzy impulsami przy prądzie I_{20} , jest równa wartości początkowej strumienia, występującej na początku impulsu, przy prądzie uzwojenia pierwotnego I_0 . Z tego wynika zależność:

$$z_1 I_0 = z_2 I_{20} \quad (8.49)$$

wiążąca, podobnie jak wzór (8.40), wartości minimalne obydwu prądów.

W stanie ustalonym, ładunek elektryczny, doprowadzany do kondensatora prądem i_2 w jednym okresie impulsowania, jest równy ładunkowi odprowadzanemu przez rezystor obciążenia w tym okresie:

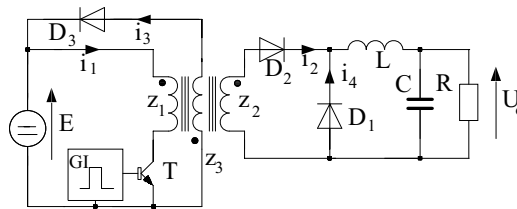
$$\int_0^{t_p} i_2(t) dt = \frac{U_o}{R} T = \frac{1}{2} (I_{20} + I_{2m}) t_p \quad (8.50)$$

Ta ostatnia zależność pozwala wyliczyć wszystkie wartości minimalne, maksymalne i średnie prądów w przetwornicy, dostarczającej napięcia U_o do rezystora obciążenia R .

Zaletą przedstawionej przetwornicy jest korzystna możliwość sterowania wartością napięcia wyjściowego przez zmianę współczynnika wypełnienia impulsów wokół wartości 0,5 (wzór (8.48)). Wadą jest konieczność gromadzenia energii w rdzeniu przez czas impulsu, a oddawanie jej w czasie przerwy między impulsami. Ten sposób pracy zawiąza wymagania stawiane rdzeniowi magnetycznemu.

8.4.2. Przetwornica dwutaktowa współbieżna

Na rys. 8.19 przedstawiono schemat przetwornicy dwutaktowej, współbieżnej (przeprustowej, synfazowej, forward DC – DC converter).



Rys. 8.19. Przetwornica współbieżna

Obciążenie przetwornicy przedstawia rezystor R, natomiast kondensator C, dławik L oraz dioda D_1 spełniają funkcję filtrującą, wygładzającą napięcie wyjściowe. Tranzystor T pracuje jako łącznik (klucz), sterowany z oddzielnego generatora impulsów prostokątnych GI.

Impuls z generatora, o czasie trwania t_i , wprowadza tranzystor w stan nasycenia, co zamyka obwód, łącząc napięcie E do uzwojenia pierwotnego transformatora. W tym stanie następuje też przepływ prądu uzwojenia wtórnego, przez diodę D_2 i dławik L, uzupełniający ładunek kondensatora. W czasie przerwy między impulsami, na czas t_p , uzwojenie pierwotne zostaje odłączone od źródła E przez nieprzewodzący tranzystor, prąd obciążenia przepływa przez dławik i diodę D_1 , a ponadto następuje przepływ prądu przez diodę D_3 , zapewniający demagnetyzację rdzenia. Po przerwie następuje kolejny impuls i cykl się powtarza, okres powtarzania impulsów T wynosi $t_i + t_p$.

Podobnie jak poprzednio, dla uproszczenia zakłada się, że tranzystor jest idealnym łącznikiem, ale wskazuje się na możliwość uwzględnienia prądu magnesowania rdzenia transformatora. Przyjmuje się, że wygładzenie napięcia wyjściowego jest idealne, czyli można, na czas jednego okresu, elementy RC zastąpić źródłem napięcia stałego o wartości U_o .

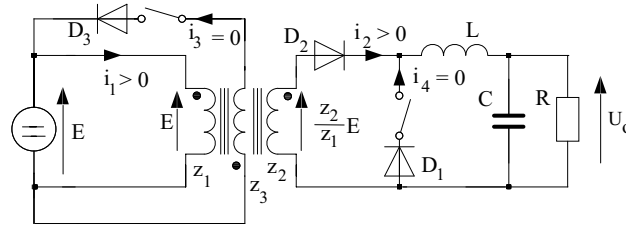
W ten sposób otrzymuje się proste schematy zastępcze, odpowiadające obydwu stanom tranzystora, schematy te pokazano na rys. 8.20 i 8.21.

Ograniczymy się tylko do opisu słownego procesów przebiegających w stanie ustalonym.

1. Przetwornica przy tranzystorze przewodzącym (rys. 8.20)

Po załączeniu tranzystora do uzwojenia pierwotnego zostaje dołączone stałe napięcie zasilające E, co wywołuje liniowy wzrost strumienia magnetycznego w rdzeniu. Kierunki

nawinięcia uzwojeń i kierunek włączenia diody D_2 są takie, że następuje jednoczesny przepływ prądów w uzwojeniu pierwotnym i wtórnym.



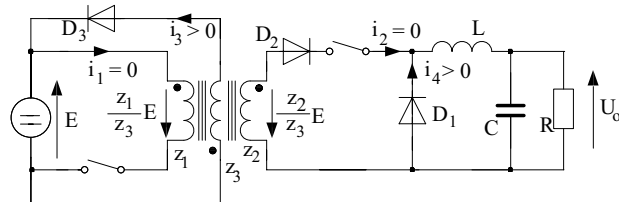
Rys. 8.20. Schemat zastępczy przetwornicy; etap przewodzenia tranzystora

Prąd uzwojenia wtórnego narasta zgodnie z zależnością:

$$L \frac{di_2}{dt} = \frac{z_2}{z_1} E - U_o \quad (8.51)$$

Dzięki równoczesnemu wzrostowi obydwu prądów, wzrost strumienia, który jest związany z różnicą przepływów (i w konsekwencji prądów) uzwojeń pierwotnego i wtórnego, jest mniejszy aniżeli wzrost tych prądów. Pozwala to na minimalizację przekroju rdzenia lub zwiększania mocy przetwornicy. Narastający prąd i_2 powoduje gromadzenie energii w polu magnetycznym dławika L . Szkic przebiegów elektrycznych pokazano na rys. 8.22.

2. Przetwornica przy tranzystorze nieprzewodzącym (rys. 8.21)



Rys. 8.21. Schemat zastępczy przetwornicy; etap nieprzewodzenia tranzystora

Po zakończeniu impulsu sterującego tranzystorem, wzbudzony w rdzeniu transformatora strumień, zaczyna się zmniejszać, co powoduje odwrócenie biegunowości indukowanych napięć. Dioda D_2 przestaje przewodzić, ale rozpoczyna przewodzenie dioda D_3 .

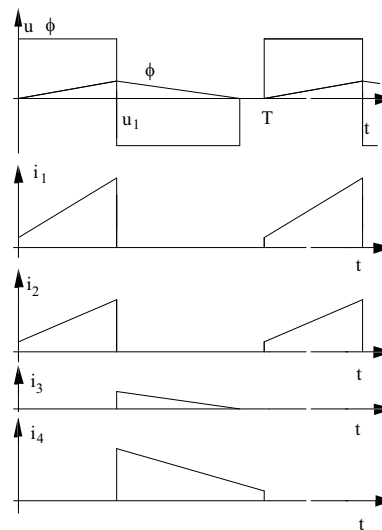
Prąd i_3 , wywołany zanikającym strumieniem w rdzeniu transformatora, przenosi energię zgromadzoną w polu magnetycznym transformatora do źródła zasilającego E . To zjawisko, aczkolwiek nie jest decydujące, to jednak korzystnie wpływa na sprawność przetwornicy. Malejący strumień magnetyczny, uprzednio wzbudzony w rdzeniu dławika L , powoduje powstanie siły elektromotorycznej samoindukcji i wymusza przepływ prądu i_4 przez

diodę D_1 . Ten prąd uzupełnia ładunek kondensatora i przyczynia się do utrzymania prawie stałego napięcia wyjściowego.

Zapisując odpowiednie wzory, podobnie jak to przedstawiano przy opisie przetwornicy przeciwbieżnej, można znaleźć m.in. zależność:

$$\frac{U_o}{E} = \frac{z_2}{z_1} D \quad (8.52)$$

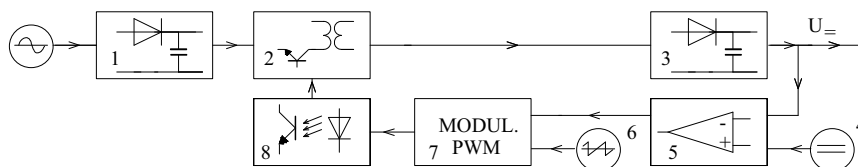
We wzorze wartość D , jak poprzednio, oznacza współczynnik wypełnienia impulsu czyli stosunek czasu impulsu do okresu impulsowania. Ten wzór wskazuje, że i w tej przetwornicy można sterować wartością napięcia wyjściowego U_o poprzez zmianę współczynnika wypełnienia impulsów sterujących tranzystorem.



Rys. 8.22. Przebiegi prądów i napięć w przetwornicy

8.4.3. Stabilizacja, zabezpieczenia, zakłócenia

Powyżej opisane przetwornice można wykorzystać wprost do stabilizacji napięcia stałego w układach ze sprzężeniem zwrotnym. W tym celu należy zastosować wzorzec napięcia zadanego, człon porównujący napięcie wyjściowe z wartością zadaną i układ korygujący współczynnik wypełnienia impulsów – modulator szerokości impulsów (zob. p. 4.4.3.). Przykładowy schemat blokowy zasilacza impulsowego ze stabilizacją napięcia wyjściowego pokazano na rys. 8.23.



Rys. 8.23. Zasilacz impulsowy z impulsową stabilizacją napięcia stałego: 1 – prostownik sieciowy i filtr wstępny, 2 – przetwornica, 3 – prostownik podwyższonej częstotliwości i filtr końcowy, 4 – wzorzec napięcia stałego, 5 – wzmacniacz błędów, 6 – generator fali nośnej, 7 – modulator szerokości impulsów, 8 – separator optoelektroniczny

Na tym przykładzie można zauważyć niektóre zadania techniczne związane ze specyfiką zasilaczy impulsowych. Na przykład wytworzenie napięcia wzorcowego, często w układzie stabilizatora parametrycznego, wymaga zasilania tego układu. Bezpośrednio po włączeniu zasilacza jest dostępne tylko wyprostowane napięcie sieci, co zwykle jest wartością o wiele za dużą do zasilania niskonapięciowej diody Zenera, a użycie tego napięcia

wprowadziłoby potencjał sieci do obwodów wyjściowych zasilacza. Zadanie rozwiązuje się wprowadzając odmienny sposób pracy zasilacza na czas rozruchu układu.

Podobną trudność sprawia uruchomienie przetwornicy przy rozładowanym kondensatorze filtra końcowego, co daje efekt podobny do zwarcia na wyjściu i grozi uszkodzeniem tranzystora. Zadanie jest rozwiązywane przez wprowadzenie tzw. miękkiego startu, to znaczy ograniczenie czasów pierwszych impulsów do bardzo małych wartości, aby uniknąć przetężeń.

Oczywiście w układach musi być stosowane ograniczenie prądowe, na wypadek zwarć lub przeciążeń zasilacza. Dla wygody użytkowników to ograniczenie współpracuje z układem samoczynnego powtórnego załączenia, aby krótkotrwałe przeciążenia nie powodowały całkowitego wyłączenia. Takie rozwiązanie pociąga za sobą konieczność zliczania ilości cykli samoczynnego włączania, aby dla przypadku trwałego przeciążenia lub zwarcia po kilku czy kilkunastu próbach załączeń przejść do stanu trwałego wyłączenia zasilacza.

Zespół przytoczonych wymagań jest na tyle złożony, że całkowicie uzasadnione jest wytwarzanie układów scalonych, zawierających szereg podzespołów niezbędnych do sterowania tranzystorem czy tranzystorami przetwornicy.

Zamknięcie toru sprzężenia zwrotnego stabilizacji napięcia, z wyjścia zasilacza do przetwornicy, połączonej galwanicznie z siecią, wymaga wprowadzenia separacji także w tym torze, inaczej sens stosowania transformatora byłby podważony. Funkcję separacyjną spełnia drugi transformator lub transoptor.

Dodatkowym problemem są zakłócenia radioelektryczne generowane przez zasilacz impulsowy, który jest ich stosunkowo silnym źródłem. Jednocześnie coraz bardziej powszechne stosowanie tych zasilaczy w sprzęcie komputerowym i podobnym, czyli w urządzeniach wrażliwych na zakłócenia obcymi polami, narzuca konieczność obniżania poziomu zakłóceń, generowanych w zasilaczu, przez właściwy montaż, ekranowanie, stosowanie filtrów przeciwzakłóceń i podobne środki.

Odrębnym problemem, wywołanym przez masowe stosowanie zasilaczy impulsowych, o mocach sięgających wieluset watów, jest odkształcenie prądu pobieranego z sieci elektroenergetycznej. To pociąga za sobą zniekształcenia napięcia sieci, przeszkadzające innym użytkownikom energii elektrycznej i dlatego są wprowadzane przepisy narzucające dozwolony poziom odkształceń prądu pobieranego a także współczynnika mocy zasilaczy impulsowych. Szczegółowa prezentacja tych zagadnień jest zadaniem bardzo szerokim i dlatego w tym skrypcie nie będzie przytaczana.

LITERATURA

- [1] Baranowski J., Nosal Z.: Układy elektroniczne cz. I. Układy analogowe liniowe. Warszawa: WNT 1994.
- [2] Baranowski J., Czajkowski G.: Układy elektroniczne cz. II. Układy analogowe nieliniowe i impulsowe. Warszawa: WNT 1994.
- [3] Baranowski J., Kalinowski B., Nosal Z.: Układy elektroniczne cz. III. Układy i systemy cyfrowe. Warszawa: WNT 1994.
- [4] Borkowski A.: Układy scalone w stabilizatorach napięcia stałego. Warszawa: WNT 1985.
- [5] Borkowski A.: Zasilanie urządzeń elektronicznych. Warszawa: WKiŁ 1990.
- [6] Cieślak J.: Półprzewodnikowe elementy optoelektroniczne. Warszawa: Wyd. MON 1984.
- [7] Ferenczi O.: Zasilanie układów elektronicznych. Warszawa: WNT 1988.
- [8] Ferenczi O.: Zasilanie układów elektronicznych – zasilacze impulsowe. Warszawa: WNT 1989.
- [9] Filipkowski A.: Układy elektroniczne analogowe i cyfrowe. Warszawa: WNT 1995.
- [10] Guziński A.: Liniowe elektroniczne układy analogowe. Warszawa: WNT 1992.
- [11] Gniewko B., Klimek C.: Rezonatory i generatory kwarcowe. Warszawa: WKiŁ 1980.
- [12] Golde W.: Wzmacniacze operacyjne i ich zastosowania. Warszawa: WNT 1980.
- [13] Hennel J.: Podstawy elektroniki półprzewodnikowej. Warszawa: WNT 1995.
- [14] Horowitz P., Hill W.: Sztuka elektroniki. Warszawa: WKiŁ 1996.
- [15] Jaczewski J., Opolski A., Stolz J.: Podstawy elektroniki i energoelektroniki. Warszawa: WNT 1981.
- [16] Janke W.: Zjawiska termiczne w elementach i układach półprzewodnikowych. Warszawa: WNT 1992.
- [17] Kalisz J.: Podstawy elektroniki cyfrowej. Warszawa: WKiŁ 1991.
- [18] Kaźmierkowski M., Nowak M., Wójciak A.: Urządzenia elektroniki przemysłowej. Warszawa: WKiŁ 1982.
- [19] Kulka Z., Libura A., Nadachowski M.: Przetworniki analogowo-cyfrowe i cyfrowo-analogowe. Warszawa: WKiŁ 1987.
- [20] Łakomy M., Zabrodzki J.: Scalone przetworniki analogowo-cyfrowe i cyfrowo-analogowe. Warszawa: PWN 1985.
- [21] Marciniak W.: Przyrządy półprzewodnikowe i układy scalone. Warszawa: WNT 1984.
- [22] Nadachowski M., Kulka Z.: Analogowe układy scalone. Warszawa: WKiŁ 1985.
- [23] Nadachowski M., Kulka Z.: Zastosowanie wzmacniaczy operacyjnych. Warszawa: WNT 1986.
- [24] Napieralski A., Napieralska M.: Połowe półprzewodnikowe przyrządy dużej mocy. Warszawa: WNT 1995.
- [25] Nowakowski W.: Układy impulsowe. Warszawa: WKiŁ 1982.
- [26] Pawłowski J.: Podstawowe układy elektroniczne. Wzmacniacze i generatory. Warszawa: WKiŁ 1980.
- [27] Polowczyk M.: Elementy i przyrządy półprzewodnikowe powszechnego zastosowania. WKiŁ, Warszawa, 1986.

- [28] Porębski J., Korohoda P.: SPICE. Program analizy nieliniowej układów elektronicznych. Warszawa: WNT 1992.
- [29] Przeździecki F., Opolski A.: Elektrotechnika i elektronika. Warszawa: PWN 1986.
- [30] Rusek M., Ćwirko R., Marciniak W.: Przewodnik po elektronice. Warszawa: WNT 1986.
- [31] Streetman B.: Przyrządy półprzewodnikowe. Warszawa: WNT 1976.
- [32] Stępień C.: Mikroprocesory firmy Intel. Warszawa: PWN 1992.
- [33] Till W.C., Luxon J.T.: Integrated circuits: materials, devices, and fabrication. Englewood Cliffs N.J.: Prentice Hall Inc. 1982.
- [34] Ziel A.v.d.: Podstawy fizyczne elektroniki ciała stałego. Warszawa: WNT 1980.