

Spis treści

Wykaz podstawowych skrótów użytych w skrypcie	5
Przedmowa	7
1. Informacja w telekomunikacji	9
1.1. Miara informacji	9
1.2. Informacja w przekazie słownym	12
1.3. Informacja w przekazie obrazowym	14
Zadania	16
2. Podstawowe pojęcia telekomunikacji	17
2.1. Sygnały w systemie telekomunikacyjnym	17
2.2. Przepustowość systemu telekomunikacyjnego	19
2.3. Transmisja w pasmie naturalnym i transmisja pasmowa	22
2.4. Zwielokrotnienie kanału telekomunikacyjnego	24
2.5. Ogólne zależności dotyczące sygnałów	27
Zadania	29
3. Modulacja amplitudy	30
3.1. Sygnał zmodulowany amplitudowo	30
3.2. Widmo sygnału AM	32
3.3. Moc sygnału AM	37
3.4. Otrzymywanie sygnałów z modulacją AM	39
3.5. Demodulacja sygnałów AM	42
Zadania	47
4. Modulacja kąta	48
4.1. Modulacja częstotliwości i fazy	48
4.2. Sygnał FM	50
4.3. Sygnał PM	51
4.4. Widmo sygnałów z modulacją kąta	53
4.5. Moc sygnału FM	58
4.6. Otrzymywanie modulacji FM	59
4.7. Demodulacja sygnałów FM	59
4.8. Modulatory i demodulatory fazy	62
Zadania	63

5. Modulacje impulsowe	65
5.1. Twierdzenie o próbkowaniu	65
5.2. Sposoby próbkowania sygnału	70
5.3. Analogowe modulacje impulsowe	72
5.4. Pasma sygnałów PAM, PWM, PPM	73
Zadania	75
6. Modulacja PCM	76
6.1. Wytwarzanie sygnału PCM	76
6.2. Szum kwantyzacji	77
6.3. Modulacja delta	80
6.4. Zabezpieczanie transmisji PCM przed zakłóceniami	82
6.5. Pasma częstotliwości wymagane do transmisji PCM	86
6.6. Sposoby kodowania sygnałów PCM	88
6.7. Modulacja PCM fali nośnej	92
Zadania	99
7. Transmisja sygnałów w obecności zakłóceń	100
7.1. Właściwości szumu	100
7.2. Źródła szumu	103
7.3. Charakterystyka szumowa modulacji AM	107
7.4. Charakterystyka szumowa modulacji FM	110
7.5. Odporność szumowa modulacji impulsowych	114
7.6. Odporność szumowa sygnałów PCM	115
7.7. Porównanie odporności szumowej modulacji PCM i FM	122
Zadania	125
8. Ogólne zagadnienia odbioru sygnałów	126
8.1. Podstawowe parametry odbiornika	126
8.2. Współczynnik szumów	127
8.3. Czułość odbiornika	130
8.4. Odbiornik superheterodynowy	131
8.5. Odbiornik homodynowy	135
8.6. Optymalny odbiornik sygnałów cyfrowych	136
Zadania	140
Literatura	142

Wykaz podstawowych skrótów użytych w skrypcie

AM	modulacja amplitudy
ASK	kluczowanie amplitudy
BPSK	dwuwartościowe kluczowanie fazy
CDM, CDMA	zwielokrotnienie w dziedzinie kodowej
DAT	cyfrowy zapis dźwięku na taśmie magnetycznej
DSB	podwójna wstęga boczna
DSB-SC	podwójna wstęga boczna z wytłumioną falą nośną
FDM, FDMA	zwielokrotnienie w dziedzinie częstotliwości
ΦM	modulacja kąta
FM	modulacja częstotliwości
FSK	kluczowanie częstotliwości
GPS	satelitarny system nawigacji globalnej
GSM	system telefonii komórkowej 2. generacji
ISDN	sieć cyfrowa (telefoniczna) z integracją usług
NRZ	kod cyfrowy bez powrotu do zera
PAM	modulacja amplitudy impulsów
PCM	modulacja kodowo-impulsowa
PM	modulacja fazy
PPM	modulacja położenia impulsów
PSK	kluczowanie fazy
PWM	modulacja szerokości impulsów
QAM	kwadraturowa modulacja amplitudy

QPSK	czterowartościowe kluczowanie fazy
RZ	kod cyfrowy z powrotem do zera
S/N, SNR	stosunek mocy sygnału do mocy szumu
SSB	pojedyncza wstęga boczna
SSB-SC	pojedyncza wstęga boczna z wytłumioną falą nośną
TDM, TDMA	zwielokrotnienie w dziedzinie czasu

Przedmowa

Człowiek jest istotą społeczną – z tego truizmu wynika, że potrzebuje porozumiewania się z innymi osobami, czyli wymiany informacji. Tak więc nadawanie i odbiór informacji stanowi istotny aspekt ludzkiego życia.

Informacje przekazywane są słownie, za pomocą obrazów, gestów, znaków pisanych itd. Jeżeli zachodzi potrzeba przekazania informacji na większą odległość – mamy do czynienia z jakąś formą telekomunikacji.

Rozwój telekomunikacji datuje się od XIX w., kiedy pojawiły się, początkowo bardzo skromne, techniczne możliwości przesyłu informacji na większą odległość. Ważniejsze daty w rozwoju tych możliwości przedstawiono w tabeli.

1837	telegraf Morse'a
1866	funkcjonujący telegraficzny kabel transatlantyczny
1876	telefon (Bell), pierwsza rozmowa telefoniczna
1889	automatyczna centrala telefoniczna (Strowger)
1901	radiowa łączność międzykontynentalna (Marconi)
1927	międzykontynentalna rozmowa telefoniczna
1966	pierwszy satelita telekomunikacyjny
1969	pierwsza sieć komputerowa (ARPANET)
1970	łączność światłowodowa
~1980	początek systemów telefonii komórkowej
1989	międzykontynentalny kabel światłowodowy

Podstawą do opracowania niniejszego skryptu były prowadzone przez autora od kilkunastu lat i uaktualniane na bieżąco wykłady z przedmiotów takich jak „telekomunikacja”, „podstawy telekomunikacji”, „systemy telekomunikacyjne” czy „wybrane zagadnienia telekomunikacji” na studiach zaocznych i dziennych na Wydziale Elektrotechniki i Informatyki Politechniki Rzeszowskiej.

Skrypt przeznaczony jest zasadniczo dla studentów kierunku *elektrotechnika*, odbywających zajęcia z przedmiotu „telekomunikacja”, oraz dla studentów kierunku *informatyka* – z przedmiotu „systemy telekomunikacyjne”. Skrypt może być przydatny dla studentów także innych kierunków, ściślej związanych z telekomunikacją, jako materiał wprowadzający do bardziej dokładnego studiowania poruszanych zagadnień.

Wykłady dla studentów kierunków *elektrotechnika* i *informatyka* są z konieczności tak ułożone, że zawierają zarówno typowe podstawy telekomunika-

cji, jak i wybrane współczesne systemy telekomunikacyjne (telefonii stacjonarna, telefonii komórkowa, systemy światłowodowe...).

Niniejszy skrypt dotyczy tylko *pierwszej* części tych wykładów, tj. ogólnych zagadnień informacji i telekomunikacji, metod kodowania i modulacji, wpływu szumu i zakłóceń na transmisję oraz ogólnych zasad odbioru sygnałów. Materiał ten jest zawarty w 8 rozdziałach. Na końcu każdego rozdziału zamieszczono po kilka prostych zadań, których rozwiązanie niewątpliwie ułatwi opanowanie przedstawionego materiału.

Omówienie wybranych współczesnych systemów telekomunikacyjnych ma – w zamierzeniu autora – stanowić treść następnego skryptu.

Autor

1

Informacja w telekomunikacji

1.1. Miara informacji

Termin telekomunikacja wywodzi się z greckiego *tele* (odległy) oraz *communicare* (dzielenie się z kimś), co łącznie symbolizuje przekazywanie wiadomości na odległość, bez wnikania, w jaki sposób i jakimi środkami zostanie to zrealizowane.

Według definicji ustalonej przez Międzynarodową Unię Telekomunikacyjną (ITU – *International Telecommunication Union*) telekomunikacja to jakiekolwiek przesyłanie, emisja albo odbiór znaków, sygnałów, pisma, informacji wizyjnych albo dźwięków za pomocą przewodów, radia, systemów optycznych lub elektromagnetycznych.

W Polsce definicję telekomunikacji zawiera ustawa „Prawo telekomunikacyjne”, wg której „telekomunikacja to nadawanie, odbiór lub transmisja informacji jakiejkolwiek natury, w szczególności: sygnałów, znaków, pisma, obrazów lub dźwięków, za pomocą przewodów, fal radiowych bądź optycznych lub i innych środków wykorzystujących energię elektromagnetyczną”.

Co to jest informacja? – jest to pojęcie dość trudne do zdefiniowania, czy nawet zaliczane do pojęć pierwotnych, ale można jednak spotkać takie definicje, jak np. „informacja – obiekt abstrakcyjny, który w postaci zakodowanej może być przechowywany, przesyłany, przetwarzany i użyty do sterowania” albo „informacja – miara niepewności zajścia pewnego zdarzenia spośród skończonego zbioru zdarzeń możliwych”.

Inna definicja informacji to „ślad, mniej lub więcej trwały, pozostawiony w umyśle człowieka przez każdy, mniej lub więcej nieoczekiwany, bodziec energetyczny odbierany przez człowieka z otoczenia” [8, s. 26]. W tym aspekcie, jeżeli nie ma odbiorcy, nie ma i informacji.

Skoro telekomunikacja jest to przesyłanie informacji na odległość, przy technicznym podejściu do tego problemu zachodzi konieczność pomiaru ilości tej informacji – czy jest jej dużo czy mało, czy jest ona przekazywana szybko czy powoli. W ten sposób możliwa będzie optymalizacja i ocena sprawności systemów telekomunikacyjnych.

W życiu codziennym do informacji stosuje się przymiotniki typu „ważna, istotna, nieważna, wstrząsająca...” – ale jak to ująć liczbowo? Zgodnie z badaniami Shannona przyjęto, że jeżeli informacja jest bardzo ważna, to jej miara liczbową powinna być duża, jeżeli jest nieważna – mała [5].

* Nowa Encyklopedia Powszechna PWN, Warszawa 1995, t. 3, s. 53-54.

Zwróćmy uwagę na proces percepcji odbieranej informacji. Jeżeli informacja jest zupełnie zaskakująca, może spowodować istotną zmianę postępowania odbiorcy itd. (bo był nieprzygotowany...). Jeżeli jednak informacja była spodziewana, nie stanowi zaskoczenia i reakcja odbiorcy na nią jest umiarkowana czy wręcz jej nie ma (prawdopodobieństwo takiej informacji było duże – bliskie jedności).

To, czy informacja jest zaskakująca (czyli gdy jej miara jest duża) czy nie (miara jest mała), zależy od jej prawdopodobieństwa. Gdy prawdopodobieństwo jakiejś informacji jest równe 1, jest to wiadomość pewna i otrzymanie takiej informacji nie zmienia w niczym stanu wiedzy odbiorcy (bo on już wcześniej wiedział). Można więc przyjąć, że liczbową miarą takiej informacji dla odbiorcy jest po prostu zero.

Podczas transmisji źródło informacji posługuje się pewnym **inwentarzem wiadomości** – zasobem możliwych wiadomości. Inwentarz ten może zawierać ich niewiele, minimum dwie możliwe, np. „tak” i „nie”, ale może ich być znacznie więcej, np. kilkadziesiąt tysięcy. Każda z tych wiadomości, statystycznie biorąc, ma jakieś prawdopodobieństwo (jej pojawienia się i transmisji). Odbiorca oczekuje tych wiadomości z tymi prawdopodobieństwami (często w sposób nieświadomy); odebranie konkretnej wiadomości powoduje, że odbiorca przyjmuje ją za pewną, eliminując pozostałe możliwe wiadomości.

Zgodnie z propozycją Shannona, jeżeli nadawana jest wiadomość a_i z jakiegoś inwentarza i prawdopodobieństwo jej jest równe $p(a_i)$, to miarą zawartości informacji w tej wiadomości jest liczba

$$I = \log_2 \frac{1}{p(a_i)} \quad (1.1)$$

Jak określa się jednostkę informacji, czyli kiedy z powyższego wyrażenia otrzyma się liczbę 1? Wtedy, gdy $p(a_i) = 0.5$. Przekazanie wiadomości o zdarzeniu, którego prawdopodobieństwo jest równe 0.5, jest równoważne przekazaniu jednej jednostki informacji, nazywanej **bitem** (skrót od **binary unit**). Istnieje ścisły związek między „tym” bitem i bitem jako impulsem występującym podczas cyfrowego przesyłania danych, np. w komputerach; będzie o tym mowa nieco później.

Przykład 1.1

Źródło informacji nadaje dwa symbole a_1 i a_2 , np. mogą to być słowa tak i nie (coś oznaczające), przy czym $p(a_1) = p(a_2) = 0.5$. Nie oznacza to wcale, że symbole te nadawane są ściśle na przemian, bo wtedy taki przekaz mógłby być przewidziany dowolnie naprzód i nie byłby wcale potrzebny. Szybkość nadawania tych symboli wynosi 100/s. Ile informacji na sekundę nadaje to źródło?

Każdy z symboli zawiera 1 bit (b) informacji, czyli omawiana transmisja jest równoważna 100 b/s. W ten sposób określona została **przepływność** informacji albo szybkość bitowa (*bit rate*). Jest to liczba bitów wiadomości przekazywanych w jednostce czasu.

Przykład 1.2

Źródło informacji nadaje dwa symbole a_1 i a_2 , ale tym razem jeden z symboli jest bardziej prawdopodobny, np. $p(a_1) = 0,8$ i $p(a_2) = 0,2$. Przykładowo, mógłby to być ciąg symboli: $a_1 a_1 a_2 a_1 a_2 a_1 a_1 a_2 a_1 a_2 a_2 a_1 a_1 a_1 a_2 \dots$. Szybkość nadawania wynosi ponownie 100 symboli na sekundę. Ile informacji na sekundę nadaje to źródło?

W tym przypadku:

- ilość informacji zawarta w symbolu a_1 wynosi $\log_2 1/0,8 = 0,32$ b,
- ilość informacji zawarta w symbolu a_2 wynosi $\log_2 1/0,2 = 2,33$ b.

W ciągu 1 sekundy źródło nadaje 100 symboli, przy czym statystycznie jest wśród nich 80 symboli a_1 i 20 symboli a_2 , zatem ilość przekazanej w czasie 1 sekundy informacji jest równa $80 \cdot 0,32 + 20 \cdot 2,33 = 72,2$ b. Przepływność informacji wynosi więc ok. 72 b/s.

W tym przypadku szybkość przekazywania informacji uległa wyraźnemu zmniejszeniu. Jest tak dlatego, że $p(a_1) = 0,8$ i odbiorca ma prawo oczekiwać częstszego pojawiania się symbolu a_1 (i to się potwierdza), czyli symbol ten nie stanowi dla niego takiego dużego zaskoczenia (zawiera mniej informacji). Symbol a_2 jest odpowiednio mniej prawdopodobny; zawiera więc więcej informacji, ale występuje na tyle rzadko, że nie powoduje to zwiększenia średniej ilości informacji przypadającej na jeden symbol.

Jak określić ilość informacji w przypadku ogólnym, gdy źródło informacji posługuje się bogatszym inwentarzem wiadomości? Wówczas należy określić, ile średnio informacji zawiera jedna wiadomość, z uwzględnieniem tego, jak często może ona się pojawiać w ciągu nadawanych wiadomości.

Wielkość ta nazywana jest entropią i wyraża się następująco:

$$\bar{H} = \frac{p(a_1) \cdot \log_2 \frac{1}{p(a_1)} + p(a_2) \cdot \log_2 \frac{1}{p(a_2)} + \dots}{p(a_1) + p(a_2) + \dots}, \quad (1.2)$$

czyli

$$\bar{H} = \sum_i p(a_i) \cdot \log_2 \frac{1}{p(a_i)}, \quad (1.3)$$

ponieważ $\sum_i p(a_i) = 1$.

Jest to typowa średnia ważona, gdzie współczynnikami wagowymi są prawdopodobieństwa poszczególnych informacji.

W termodynamice entropia jest miarą rozproszenia energii. Rozważana tu entropia jest też miarą pewnego rozproszenia – jeżeli prawdopodobieństwo różnych informacji byłoby jednakowe (byłoby ono równomiernie „rozproszone”), to entropia byłaby największa.

Dysponując wielkością entropii, przepływność informacji I można wyrazić następująco:

$$I = \bar{H} \cdot S_t, \quad (1.4)$$

gdzie S_t jest to **szybkość symbolowa**, tj. szybkość nadawania znaków, symboli, gestów... W każdym z nich jest zawarte średnio \bar{H} bitów informacji.

1.2. Informacja w przekazie słownym

Jak ująć ilość i przepływność informacji w przekazie słownym (rozmowa bezpośrednia, rozmowa telefoniczna, słuchanie audycji...)?

Informacja przekazywana jest za pomocą zdań, a te składają się z wyrazów. Wyrazy w danym języku występują z określoną częstotliwością – czyli prawdopodobieństwem. Oczywiście jest to bardzo zależne od środowiska, w jakim się poruszamy: czy jest to środowisko ludzi bardziej lub mniej wykształconych, humanistów czy techników. Prosty, niewyszukany język to ok. 5000 słów w użyciu, bardziej wyszukany – np. 10 razy więcej, a więc prawdopodobieństwo wystąpienia wyrazów może być rzędu 10^{-4} – 10^{-5} , co daje bardzo przybliżoną entropię ok. 13–20 bitów na wyraz*. Przy wypowiedzianiu np. 5 wyrazów na sekundę daje to przepływność informacji ok. 100 b/s.

Uwaga

☛ Powyższe rozważania są prymitywne, ponieważ należałoby jeszcze zwrócić uwagę na związki międzywyrazowe (ogólnie międzysymbolowe). Pewne wyrazy występują prawie zawsze razem, np. „przede wszystkim”, „za wszelką cenę”, „serdeczne życzenia” itd. – jeżeli słyszy się pierwszy wyraz, można się domyśleć następnych; niekiedy oczywiście nietrafnie. W ten sposób prawdopodobieństwo wiadomości wzrasta, czyli ilość informacji maleje. Na przykład słysząc znaną wszystkim maksymę (przysłowie), już po pierwszych wyrazach odbiorca wie, co będzie dalej aż do jej końca, czyli w tej sytuacji nie odbiera dużo informacji, chociaż sama maksyma może liczyć i kilkanaście słów.

* Częste używanie słów niecenzuralnych w rozmowie świadczy o niskiej kulturze człowieka. Świadczy też, że nieznane mu są podstawowe zagadnienia teorii informacji. Często używany wyraz ma przecież duże prawdopodobieństwo, czyli zawartość „informacji” w nim jest mała, dlatego jego używanie nie ma logicznego uzasadnienia.

Jezeli występują określone związki pomiędzy przekazywanymi wiadomościami, ilość przekazywanej informacji ulega zmniejszeniu.

Ilościowy przekaz informacji słownej można rozważyć dokładniej, schodząc do poziomu liter i zgłosek, z których składają się wyrazy i słowa. Liczba liter jest niewielka i dość łatwo można wyznaczyć prawdopodobieństwo każdej z nich, a co za tym idzie – zawartość informacji w każdej literze (tab. 1.1).

Tab. 1.1. Przybliżona częstość występowania liter w języku polskim [12].

Litera	Prawdopodo- bieństwo, %	Litera	Prawdopodo- bieństwo, %	Litera	Prawdopodo- bieństwo, %	Litera	Prawdopodo- bieństwo, %
a	8.6	l	0.4	ni	3.3	ś	0.8
ą	1.1	g	1.4	n	5.6	t	4.3
b	1.5	h	0.9	ń	0.1	u	2.1
c	5.8	i	8.6	o	7.3	w	4.2
ć	0.6	j	2.8	ó	0.8	y	3.8
d	3.3	k	2.9	p	2.7	z	5.7
e	8.8	ł	2.2	r	3.7	ż	0.06
ę	1.4	ł	1.4	s	4.2	ź	1.2

Jak widać, najczęściej w języku polskim występują samogłoski. Łącznie ich prawdopodobieństwo wynosi blisko 40%, to znaczy, że (średnio!) co drugą, trzecią literą w tekście jest samogłoska. Z zestawienia wynika, że najrzadziej w tekście polskim pojawia się litera „ż”, co oznacza, że zawiera ona najwięcej informacji – ok. 11 b.

Na podstawie danych z tab. 1.1 można policzyć, że zwykła średnia zawartość informacji w literach to ok. 5,7 b na literę, natomiast średnia ważona, czyli entropia, to ok. 4,6 b na literę. Wypowiadając wyraz składający się np. z 10 liter, przekazuje się ok. 46 b informacji, ale należy ponownie pamiętać o związkach – tym razem między literami. Pewne litery występują częściej razem, pewne zestawienia liter są bardziej prawdopodobne, inne nie (np. długi ciąg spółgłosek jest rzadkością w języku polskim), dlatego pierwsze litery wyrazu powodują, że można się domyślić jego zakończenia, a więc ilość informacji w tych końcowych literach będzie zmniejszona.

Uwzględniając, że w szybkiej rozmowie można wypowiadać kilka wyrazów na sekundę, a każdy z nich może zawierać np. 10 liter, przepływność informacji przekazywanej w tej rozmowie można oszacować na 100–200 b/s; odpowiada to mniej więcej oszacowaniu zrobionemu nieco wcześniej na poziomie wyrazów.

Warto zauważyć, że oszacowana przepływność informacji słownej jest bardzo mała w porównaniu z technicznymi możliwościami współczesnych systemów telekomunikacyjnych.

1.3. Informacja w przekazie obrazowym

Obraz składa się z pikseli – elementarnych obszarów o jednolitej barwie i jaskrawości. Każdy z nich z określonym prawdopodobieństwem może przybrać daną barwę, a więc każdy piksel zawiera określoną ilość informacji. Jeżeli np. obraz jest czarno-biały (bez odcieni szarości), to każdy z pikseli może być tylko czarny albo biały, więc prawdopodobieństwo tego, że jest on czarny lub biały można przyjąć za 0,5. Oczywiście, także tutaj mogą występować związki międzysymbolowe, tj. pomiędzy sąsiednimi pikselami (wykorzystywane to jest w systemach TV cyfrowej), ale dla uproszczenia rozważań zostaną tu zignorowane.

Jeżeli obraz jest w skali szarości, np. 8-stopniowej, to prawdopodobieństwo tego, że piksel przybierze akurat dany odcień szarości, jest równe $1/8$, co odpowiada zawartości 3 b informacji w tym pikselu, ponieważ

$$\log_2 \frac{1}{1/8} = 3.$$

Im więcej możliwych kolorów lub odcieni szarości może przybrać piksel, tym więcej bitów informacji będzie on zawierał. W technice komputerowej znane jest zapisywanie tzw. obrazów bitmapowych, jako np. 16-kolorowych, 256-kolorowych itd. – odpowiada to 4 lub 8 bitom zawartym w każdym pikselu. Bardzo dobrej jakości obrazy zapisywane są jako 24-bitowe, co wskazuje, że każdy z pikseli może przybrać jeden z 16777216 kolorów, tj. że prawdopodobieństwo wystąpienia konkretnego koloru jest równe odwrotności tej liczby ($1/16777216$).

Na rysunku 1.1 pokazano przykłady prostych obrazów o różnej zawartości informacji.

Widoczne jest, że przekaz wiadomości za pomocą obrazów jest znacznie bardziej efektywny, bo obrazy zawierają znacznie więcej informacji niż nawet długie teksty czy wypowiedzi słowne.

Jest takie powiedzenie, że „jeden obraz jest wart więcej niż tysiąc słów”, co w przypadku obrazów o dobrej rozdzielczości (jak na rys. 1.1e) wydaje się oczywiste.

Jak ocenić przepływność informacji w przypadku transmisji obrazów telewizyjnych? Przyjmuje się, że obraz TV składa się z 440 tys. pikseli (wynika to

ze zdolności rozdzielczości oka ludzkiego i odległości, z jakiej ogląda się obraz TV). W europejskim standardzie obrazy TV są nadawane 25 razy na sekundę, czyli nadawanych jest 11 mln pikseli na sekundę. Ile bitów zawiera każdy piksel? W najprostszym przypadku, gdy transmitowany jest obraz monochromatyczny w skali szarości i gdy liczba rozróżnianych poziomów szarości jest równa 8 (wartość minimalna dla TV), wtedy każdy piksel zawiera 3 b informacji. Ostatecznie przepływność informacji w sygnale TV to

$$I = 440 \cdot 10^3 \cdot 25 \cdot 3 = 440 \cdot 10^3 \cdot 25 \cdot 3 = 33 \text{ Mb/s.}$$

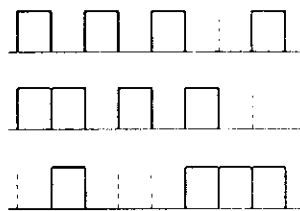
Gdy obraz będzie kolorowy, np. 16-kolorowy, przepływność będzie równa 44 Mb/s itd.



Rys. 1.1. Przykłady obrazów o różnej zawartości informacji: a) 300 bitów (20x30 pikseli, czarno-biały), b) 1200 bitów (10x15 pikseli, 256 poziomów szarości), c) 4800 bitów (20x30 pikseli, 256 poziomów szarości), d) 19200 bitów (40x60 pikseli, 256 poziomów szarości), e) 480 tys. bitów (200x300 pikseli, 256 poziomów szarości).

To, że obrazy są bardziej komunikatywne od słów i innych „nośników” informacji, powoduje, że – jak się ocenia – ok. 80% całej informacji odbieranej przez człowieka w życiu trafia do niego przez wzrok, reszta przez słuch oraz dotyk, powonienie, smak itd. Dlatego oddziaływanie na ludzi za pomocą obrazów jest bardziej skuteczne, co widać chociażby w dziedzinie reklamy.

W systemach z dwuwartościową (binarną) modulacją kodowo-impulsową (PCM) (patrz rozdz. 6.) informacja jest przekazywana za pomocą impulsów (napięcia, prądu, lub natężenia światła) o ustalonej amplitudzie i szerokości, następujących w określony sposób po sobie (rys. 1.2). Najmniejszy zespół takich impulsów, stanowiących jakąś całość informacyjną, to tzw. słowo PCM. Poszczególne składniki tego słowa (tzn. impulsy lub ich brak) określane są bitami; jak wspomniano, mają one ścisły związek z bitami jako jednostkami informacji.



Rys. 1.2. Przykładowe 8-bitowe słowa PCM, zawierające także 8 bitów informacji.

Na rysunku 1.2 pokazano przykłady słów 8-bitowych, tzn. zawierających 8 przedziałów czasowych, w których mogą się pojawić coś oznaczające impulsy. W każdym z tych przedziałów może wystąpić impuls lub może go nie być; jeżeli przyjąć, że prawdopodobieństwo każdego z tych dwóch stanów jest takie samo (0,5), to oczywiste jest, że każdy taki stan zawiera 1 b informacji. Jeżeli stany te są niezależne od siebie, to 8 takich stanów w kolejnych przedziałach czasu zawierać będzie 8 b informacji, ogólnie zaś n -bitowe słowo zawiera n bitów informacji. Dlatego obliczenia przepływności informacji w systemach PCM mogą być bardzo łatwe, gdyż ilość informacji jest równa po prostu liczbie przekazywanych bitów jako składników słów.

Przykład 1.3

W telefonii PCM stosuje się słowa 8-bitowe (w sensie ich długości), nadawane 8000 razy na sekundę. Jaka jest przepływność informacji w takim przypadku? Odpowiedź: 64 kb/s, gdyż $8000/s \cdot 8 \text{ b} = 64000 \text{ b/s}$.

Zadania

1. Źródło informacji posługuje się 4 symbolami: a_1 , a_2 , a_3 , a_4 , przy czym $p(a_1) = 0,5$, $p(a_2) = 0,25$, $p(a_3) = p(a_4) = 0,125$. Obliczyć entropię tego źródła.
2. Na podstawie danych zawartych w tab. 1.1 znaleźć, która litera ma zawartość informacji, równą entropii dla wszystkich liter ($\bar{H} = 4,6 \text{ b}$).
3. Obliczyć ilość informacji zawartą w słowie „telefon”.
4. Obrazek o rozmiarach 100x150 pikseli w 8 odcieniach szarości jest transmitowany z szybkością 2/s. Jaka jest przepływność informacji w tym przypadku?
5. Jaka przepływność informacji wystąpiłaby w przypadku telewizji przekazującej 25 razy na sekundę obrazy o rozdzielczości 24 b?

2

Podstawowe pojęcia telekomunikacji

2.1. Sygnały w systemie telekomunikacyjnym

Informacja w telekomunikacji jest przesyłana za pomocą sygnałów elektrycznych (lub optycznych), wytwarzanych np. w mikrofonie, przetworniku obrazowym lub innym przetworniku wielkości wejściowej, mikroprocesorze, klawiaturze, skanerze itp.

W tym miejscu warto zwrócić uwagę na często używany termin – sygnał. **Sygnał** to jakieś zjawisko (elektryczne, optyczne, akustyczne, zapachowe, dotykowe...), które zawiera informację. Jeżeli to zjawisko nie zawiera użytecznej informacji czy wręcz stanowi zakłócenie – bardziej właściwy jest termin **przebieg** (np. przebieg szumowy, przebieg zakłócający...).

Sygnał analogowy to przebieg ciągły, którego wartość chwilowa jest analogiem (odpowiednikiem – np. jest proporcjonalna) pierwotnej informacji (np. ciśnienie \Rightarrow napięcie). Każda wartość występującego sygnału analogowego stanowi jakąś informację.

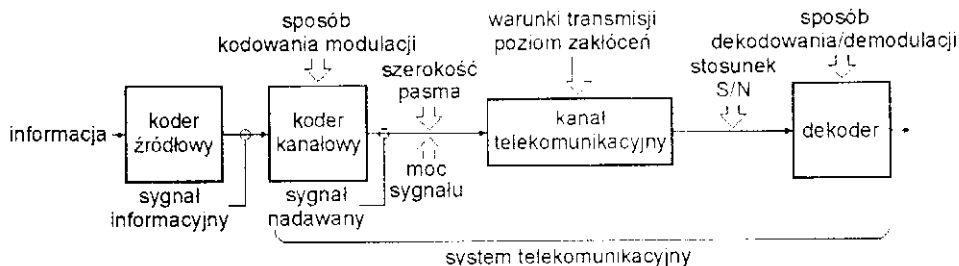
Sygnał cyfrowy jest nieciągły co do wartości (amplitudy) i czasu – ulega zmianom tylko w dyskretnych momentach czasu. Informacja w sygnale cyfrowym jest zawarta w skończonej liczbie jego dozwolonych wartości, najczęściej dwóch wartościach skrajnych, a wartości pośrednie są ignorowane lub wręcz zabronione.

W życiu codziennym spotykamy się z wieloma systemami telekomunikacyjnymi, np. systemem telefonii stacjonarnej, systemem telewizji czy radiofonii, systemem telefonii komórkowej itd. Pojęcie systemu jest raczej intuicyjnie zrozumiałe, ale można je tu sprecyzować. **System** to układ tak powiązanych elementów, że ich wzajemne relacje tworzą pewną całość. System ma wejścia i wyjścia i dokonuje przetwarzania wielkości wejściowych na wyjściowe; stanowi to cel istnienia systemu.

System telekomunikacyjny to zespół urządzeń telekomunikacyjnych, mediów transmisyjnych oraz zasad ich współdziałania (relacji). Obejmuje sam sprzęt oraz metody przetwarzania sygnałów w systemie, metody ich transmisji (np. tzw. protokoły komunikacyjne), sposób odbioru itd. (rys. 2.1).

Najczęściej informacja oryginalna (np. odpowiednie zmiany ciśnienia powietrza związane z mową czy muzyką, zmiany jasności obrazu...) nie nadaje się do dalszej transmisji, dlatego musi być zamieniona na postać elektryczną

(optyczną), która do takiej transmisji już się nadaje lub nadaje się do dalszego przetwarzania.



Rys. 2.1. Ogólny schemat blokowy systemu telekomunikacyjnego z pokazaniem jego niektórych parametrów.

Tę zamianę oryginalnej informacji na postać elektryczną nazywamy **przetwarzaniem** lub **kodowaniem źródłowym** (termin kodowanie nie oznacza tu szyfrowania, jedynie przetworzenie informacji oryginalnej na postać bardziej wygodną do dalszych operacji). Kodowanie źródłowe może być typowym kodowaniem cyfrowym, jeżeli informacja wejściowa jest tego typu lub jeżeli wszystkie dalsze ogniwa systemu są przystosowane do sygnałów cyfrowych; w systemie analogowym za kodowanie źródłowe można uznać samo przetwarzanie informacji oryginalnej na wielkość elektryczną (np. zmian ciśnienia akustycznego na odpowiadające im zmiany napięcia wyjściowego mikrofonu, zmiany jaskrawości punktów obrazu na napięcie wyjściowe przetwornika obrazowego w kamerze TV itd.).

Tak otrzymany **sygnał informacyjny** może być przesłany do odbiorcy, ale najczęściej musi być jeszcze dopasowany do kanału telekomunikacyjnego, tj. medium fizycznego, w którym ma się odbywać transmisja (może to być wolna przestrzeń, przestrzeń kosmiczna, światłowód, przewód współosiowy czy nawet morskie głębiny). Każde z tych mediów ma inne właściwości fizyczne, które mogą być wykorzystane optymalnie tylko przy określonej postaci sygnału transmitowanego. Dlatego sygnał informacyjny przechodzi jeszcze przez **koder kanałowy**, gdzie jest tak przetwarzany, aby jego transmisja przez kanał telekomunikacyjny była w jakimś sensie optymalna czy też w ogóle możliwa. Dlatego w ogólnym przypadku **sygnał nadawany** różni się od sygnału informacyjnego.

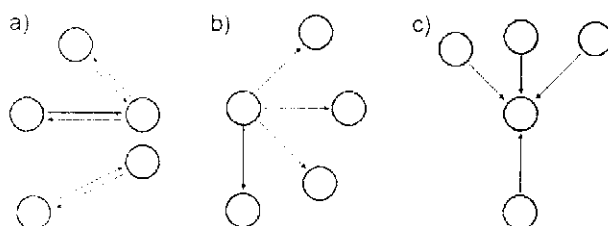
W czasie transmisji przez kanał telekomunikacyjny sygnał podlega tłumieniu, dyspersji i innym szkodliwym zjawiskom; oddziałują na niego zakłócenia zewnętrzne, często nazywane szumem – w wyniku tego sygnał odebrany ma postać inną niż sygnał nadawany. Po stronie odbiorczej sygnał odebrany podlega przekształceniom odwrotnym niż po stronie nadawczej. W końcu, w mniej lub bardziej zniekształconej formie, informacja trafia do jej odbiorcy.

Przetwarzanie sygnału informacyjnego na sygnał nadawany, odbywające się w koderze kanałowym, może zachodzić bez istotnej zmiany charakteru sygnału; często jednak we współczesnej telekomunikacji spotyka się zmianę dość radykalną – najczęściej z postaci analogowej na postać cyfrową; możliwe są tu zresztą wszystkie kombinacje charakterów sygnałów. Zestawiono je w tab. 2.1, podając jednocześnie przykłady systemów, w których występują.

Tab. 2.1. Zestawienie typów rodzajów sygnałów informacyjnych i sygnałów nadawanych.

Sygnał informacyjny	Sygnał nadawany	Przykład zastosowania
Analogowy	analogowy	klasyczna telefonia przewodowa, radiofonia AM, FM
Analogowy	cyfrowy	telefonia PCM, zapis dźwięku CD i DAT, radiofonia cyfrowa
Cyfrowy	analogowy	modemy – łączność cyfrowa po liniach telefonicznych
Cyfrowy	cyfrowy	sieci komputerowe

W systemie telekomunikacyjnym może zachodzić łączność (telekomunikacja) **porozumiewawcza**, **rozsiewcza** (dyfuzyjna) lub **zbiorcza**. Ich istotę pokazano na rys. 2.2.



Rys. 2.2. Istota telekomunikacji: a) porozumiewawczej, b) rozsiewczej, c) zbiorczej.

Przykładem systemów porozumiewawczych mogą być systemy telefonii stacjonarnej i ruchomej; systemy rozsiewcze to radiofonia, telewizja; systemy zbiorcze to systemy monitoringu, czyli nadzoru i zbierania danych ze stacji ostrzegawczych, meteorologicznych itd.

2.2. Przepustowość systemu telekomunikacyjnego

Przeznaczeniem systemu telekomunikacyjnego jest przekazywanie informacji, dlatego bardzo ważne jest, ile tej informacji można przekazywać w da-

nych warunkach i czy informacja ta nie będzie zniekształcona. W ten sposób dochodzi się do określenia bardzo ważnego parametru, jakim jest **przepustowość**. Jest to maksymalna ilość informacji, jaką można przesłać w jednostce czasu w danych warunkach. Od czego zależy ta wielkość?

Badania teoretyczne doprowadziły Shannona do sformułowania bardzo użytecznej zależności [5]:

$$C_t = B \cdot \log_2 \left(1 + \frac{S}{N} \right), \quad (2.1)$$

gdzie C_t jest przepustowością systemu (kanału) wyrażoną w bitach na sekundę, B jest szerokością pasma wykorzystywaną do transmisji, Hz, S jest mocą sygnału, a N jest mocą szumu (zakłóceń) oddziałującego na ten sygnał, przy założeniu że szum ten ma charakter gaussowski (rozdz. 7.).

Należy zaznaczyć, że wzór (2.1) podaje teoretyczną wartość przepustowości systemu, ale nie określa, jak ją można praktycznie uzyskać. Konstruktorom różnych systemów telekomunikacyjnych pozostaje opracowywanie takich sposobów modulacji czy kodowania, aby jak najlepiej wykorzystać możliwości tkwiące w parametrach tych systemów (szerokość pasma i wartość S/N). Jak na razie, w żadnym przypadku nie udało się uzyskać przepustowości równej wartości teoretycznej; najbliższe tej wartości są systemy wykorzystujące tzw. turbokodowanie.

W zależności (2.1) najczęściej nieistotna jest wartość mocy samego sygnału czy mocy szumów, tylko ich stosunek. Dlatego o wyrażeniu S/N mówi się, że jest to „stosunek mocy sygnału do szumu” albo w skrócie „stosunek sygnał-szum” (pamiętając, że dotyczy to mocy!). Stosując anglosaskie oznaczenie SNR (*Signal to Noise Ratio*), wzór (2.1) można zapisać w postaci:

$$C_t = B \cdot \log_2 (1 + \text{SNR}). \quad (2.2)$$

Należy dodać, że stosunek S/N powszechnie wyrażany jest w decybelach wg zależności

$$\frac{S}{N}, \text{ dB} = 10 \log \frac{S}{N}.$$

Zgodnie ze wzorem (2.1), przepustowość jest proporcjonalna do szerokości pasma i rośnie ze wzrostem mocy sygnału, jeżeli szumy pozostają stałe. Nieco żartobliwie wzór można określić „prawem wymiany kilowatów na kiloherce”, ponieważ tę samą przepustowość systemu można uzyskać albo przy stosowaniu słabego sygnału (mała moc nadajnika) i przy szerokim pasmie, albo odwrotnie – przy dużej mocy sygnału i wąskim pasmie. Ta alternatywa znalazła zastosowa-

nie praktyczne, m.in. w systemach łączności kosmicznej, gdzie poziomy sygnałów odbieranych są znikome ($S/N \ll 1$), ale dzięki szerokiemu pasmu częstotliwości jest możliwe uzyskanie wymaganej przepustowości.

Gdy pasmo zajmowane przez transmisję rośnie nieskończenie, przepustowość systemu też rośnie, ale tylko do pewnej granicy. Jest tak dlatego, że ze wzrostem szerokości pasma zwiększa się moc szumów oddziaływających na sygnał – jeżeli szumy mają stałą gęstość widmową mocy (wyrażaną w watach na herce), to ich moc jest wprost proporcjonalna do szerokości pasma (będzie o tym mowa w rozdz. 7.). Dlatego

$$\lim_{B \rightarrow \infty} C_t = \lim_{B \rightarrow \infty} B \cdot \log_2 \left(1 + \frac{S}{N_0 B} \right) = \frac{1}{\ln 2} \cdot \frac{S}{N_0}, \quad (2.3)$$

gdzie N_0 jest widmową gęstością mocy szumów, W/Hz.

W rozdziale 1. była określona przepływność informacji w przypadku rozmowy oraz przekazu telewizyjnego. Obecnie może być interesujące porównanie tych wartości z teoretyczną przepustowością systemów do ich przesyłania.

System telefoniczny (analogowy) charakteryzowany jest przez standardowo przyjmowane pasmo częstotliwości (0,3–3,4 kHz), tj. szerokość tego pasma wynosi 3,1 kHz. Przyjmując nieduży stosunek mocy sygnału do szumu równy 20 dB (co odpowiada stosunkowi 100 razy), z zależności (2.1) można wyznaczyć przepustowość takiego kanału telefonicznego. Będzie ona wynosić

$$C_t(\text{tlf}) = 3,1 \cdot 10^3 \log_2(101) \approx 21 \text{ kb/s}. \quad (2.4)$$

Przepustowość ta jest blisko 2 rzędy większa od przepływności typowej rozmowy telefonicznej. Inaczej mówiąc, prowadząc taką rozmowę, wykorzystuje się zaledwie ok. 1% przepustowości kanału telefonicznego. Czy naprawdę jest tak źle?

Nie aż tak bardzo, bo w przekazie telefonicznym mamy do czynienia z przekazywaniem także informacji pozasłownej (niewerbalnej) – wysokość głosu rozmowy, jego brzmienie, spokój, zdenerwowanie – to wszystko może też być określone z takiej rozmowy, a przecież są to informacje nieobejmowane przez proste prawdopodobieństwo liter czy wyrazów. Niekiedy ta pozasłowna informacja może mieć nawet większe znaczenie dla odbiorcy niż ta przekazywana słowami.

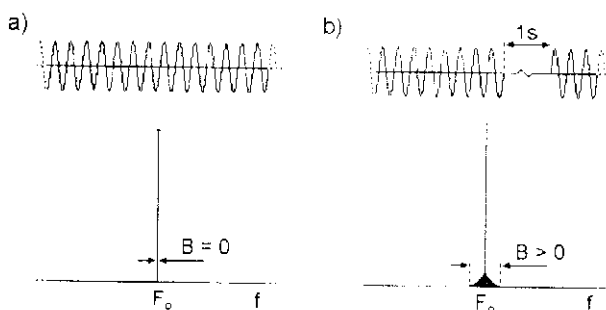
Typowy system telewizji charakteryzuje się zakresem częstotliwości potrzebnym do transmisji obrazu (0–6 MHz), a więc $B = 6 \text{ MHz}$. Przy stosunku $S/N = 30 \text{ dB} \Leftrightarrow 1000$ daje to przepustowość

$$C_t(\text{TV}) = 6 \cdot 10^6 \log_2(1001) \approx 60 \text{ Mb/s.} \quad (2.5)$$

natomiast przy lepszym stosunku SNR przepustowość odpowiednio wzrośnie.

Przepustowość systemu TV jest większa od obliczonej przepływności tylko blisko dwukrotnie. Oznacza to, że system TV nie jest zrobiony tak bardzo „na wyrost”, jak można by to powiedzieć o systemie telefonicznym. Dzięki temu, że przepustowość systemu TV jest większa od obliczonej przepływności, w systemie o takich parametrach jest możliwe przesłanie dodatkowych informacji o większej liczbie poziomów szarości, o kolorze, przesłanie sygnałów telegazety itd.

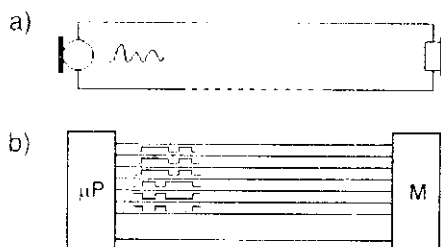
Należy podkreślić, że przepustowość każdego systemu telekomunikacyjnego jest uwarunkowana niezerowym pasmem częstotliwości. W związku z tym niemożliwe jest przesłanie jakiegokolwiek informacji za pomocą przebiegu będącego „czystą” falą sinusoidalną o częstotliwości F_0 , trwającą od $-\infty$ do $+\infty$, ponieważ taki przebieg charakteryzuje się zerową szerokością pasma (rys. 2.3a). Jeżeli natomiast w tym przebiegu wystąpi jakaś niezdeterminowana zmiana, przesłanie prostej informacji będzie już możliwe, ponieważ szerokość pasma takiego przebiegu będzie skończona (rys. 2.3b).



Rys. 2.3. Widmo nieskończonego przebiegu sinusoidalnego: a) przebiegu ciągłego – pasmo jest nieskończenie wąskie, b) przebiegu z 1-sekundową przerwą – pasmo jest skończone (rzędu ułamka herca).

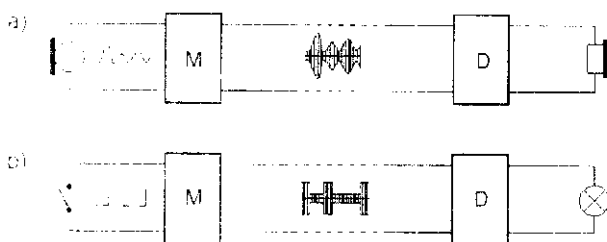
2.3. Transmisja w pasmie naturalnym i transmisja pasmowa

Jeżeli w systemie telekomunikacyjnym sygnał nadawany zajmuje to samo pasmo częstotliwości, jakie zajmuje sygnał informacyjny (ściśle lub w przybliżeniu), występuje wtedy **transmisja w pasmie naturalnym**. Może to być np. pasmo akustyczne (20 Hz – 20 kHz), wizyjne (0 – 6 MHz) lub inne, dowolnej szerokości. Proste przykłady transmisji w pasmie naturalnym pokazano na rys. 2.4.



Rys. 2.4. Przykłady transmisji w pasmie naturalnym: a) najprostsza transmisja telefoniczna, b) transmisja pomiędzy mikroprocesorem (jako źródłem sygnałów) i pamięcią.

Podczas transmisji na większą odległość, szczególnie drogą radiową, wytwarzany sygnał informacyjny jest tak przekształcany w koderze kanałowym, że zajmuje istotnie zmienione pasmo częstotliwości – przesunięte w górę i zwykle rozszerzone. Jest to wtedy **transmisja poza pasmem naturalnym** lub **transmisja pasmowa** (rys. 2.5).



Rys. 2.5. Uproszczony schemat transmisji pasmowej: a) sygnału informacyjnego analogowego, b) sygnału cyfrowego. M – modulator, D – demodulator.

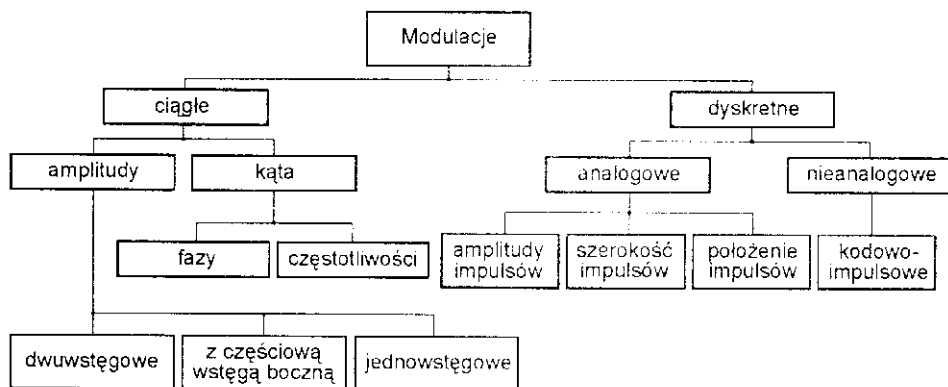
Przykładem takiej transmisji jest transmisja z użyciem modulacji fali nośnej, tj. „nakładania” sygnału informacyjnego na falę nośną (najczęściej sinusoidalną). Falą nośną może też być ciąg impulsów prostokątnych.

Jakie korzyści daje proces modulacji? Jest ich co najmniej kilka:

- jeżeli stosowane są fale nośne o różnej częstotliwości, to możliwa jest selekcja częstotliwościowa (ze względu na częstotliwość fali nośnej) różnych sygnałów, które w oryginalnej postaci zajmują to samo pasmo,
- możliwe jest dopasowanie się do fizycznych właściwości medium transmisyjnego (wolna przestrzeń, przewód miedziany lub światłowód),
- możliwe jest zwiększenie odporności transmisji na zakłócenia,
- możliwe jest zwielokrotnienie kanału telekomunikacyjnego, tj. jednocześnie przesyłanie przez niego wielu niezależnych informacji,

- w przypadku transmisji radiowej możliwe jest efektywne wypromiowanie sygnału w postaci fali elektromagnetycznej przez antenę.

Podstawowe rodzaje modulacji stosowane w telekomunikacji pokazano na rys. 2.6.



Rys. 2.6. Podstawowe modulacje stosowane w telekomunikacji.

Ilość informacji przesyłanej przez system telekomunikacyjny zależy od jego przepustowości (a ta jest proporcjonalna do pasma) i od czasu transmisji -- jest iloczynem tych wielkości. Dlatego można powiedzieć, że czas i pasmo są wielkościami wymiennymi podczas transmisji danej ilości informacji. Informację tę można przesłać w szerokim pasmie i bardzo szybko, można też użyć wąskiego pasma, ale transmisja musi wtedy trwać odpowiednio dłużej.

Wymienność czasu i pasma znalazła wyraz m.in. w sposobach zwielokrotniania kanałów telekomunikacyjnych.

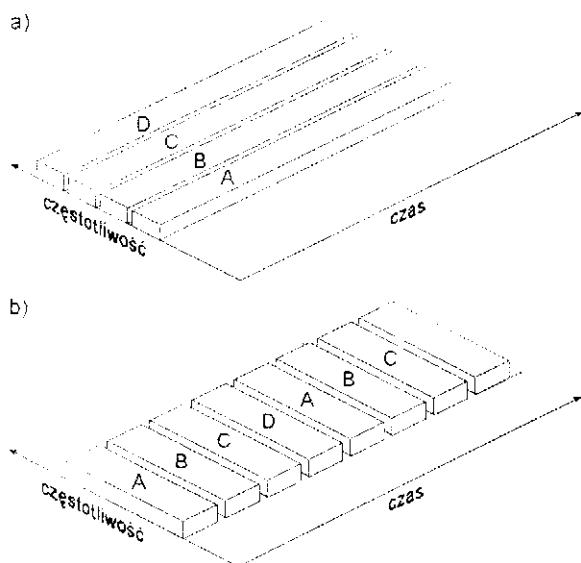
2.4. Zwielokrotnienie kanału telekomunikacyjnego

Zwielokrotnienie kanału telekomunikacyjnego polega na tym, że w jednym kanale (wolnej przestrzeni, światłowodzie, przewodzie...) jednocześnie przesyłane są sygnały różnych użytkowników, w sposób pozwalający na prawidłowe rozdzielenie tych sygnałów w miejscu odbioru. Zwielokrotnienie prowadzi do pewnego skomplikowania systemu telekomunikacyjnego, ale pozwala efektywniej wykorzystać ograniczone zasoby medium transmisyjnego.

Pierwszym stosowanym sposobem zwielokrotnienia było **zwielokrotnienie w dziedzinie częstotliwości**, oznaczane skrótowo FDMA (*Frequency Division Multiple Access*) lub FDM (*Frequency Domain Multiplexing*). Polega ono na tym, że spośród dostępnego pasma częstotliwości każdy użytkownik ma do swojej wyłącznej dyspozycji jedynie pewien wycinek pasma, którego może używać przez cały czas (rys. 2.7a). Każdy taki wycinek jest rozłączny z podobnymi in-

nymi i używając odpowiednich układów filtrujących, użytkownicy mogą bezkolizyjnie i jednocześnie korzystać z przydzielonego im zakresu częstotliwości. Ten system zwielokrotnienia jest typowy dla analogowych systemów telekomunikacyjnych, jest używany także w systemach cyfrowych (także światłowodowych) oprócz jeszcze innych sposobów zwielokrotniania.

Przykładem tego sposobu zwielokrotnienia jest stosowane w radiofonii i telewizji przydzielanie każdej stacji radiowej czy telewizyjnej różnych zakresów częstotliwości pracy, w których te stacje mogą jednocześnie nadawać swoje programy. Można powiedzieć, że w systemie tym zwielokrotnieniu podlega wolna przestrzeń, w której mogą się rozchodzić sygnały różnych użytkowników.



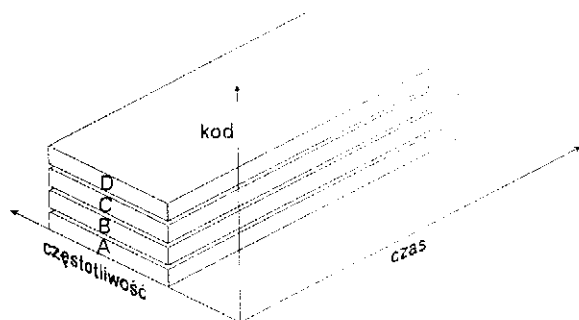
Rys. 2.7. Poglądowe przedstawienie podziału dostępnych zasobów czasu i pasma częstotliwości w systemach: a) FDMA. b) TDMA. Litery A, B, C, D oznaczają różnych użytkowników systemu.

Zwielokrotnienie w dziedzinie czasu – oznaczane jako TDMA (*Time Domain Multiple Access*) lub TDM (*Time Domain Multiplexing*) – opracowano później, na bazie odpowiedniego poziomu rozwoju układów cyfrowych. Przy tym zwielokrotnieniu dostępne pasmo częstotliwości jest używane przez wszystkich użytkowników, ale w sposób naprzemienny – każdy z nich wykorzystuje przyznaną mu tzw. szczelinę czasową (rys. 2.7b). Można by mieć obawy, czy taki przerywany sposób łączności ma sens. Okazuje się, że tak – będzie o tym mowa w rozdz. 5. – przy odpowiednio częstym następowaniu szczelin czasowych po sobie występuje możliwość bezbłędnego przesyłania ciągłych sygnałów informacyjnych o ograniczonym pasmie.

Przykładem powszechnych systemów ze zwielokrotnieniem czasowym są telefonia ISDN i telefonia komórkowa GSM.

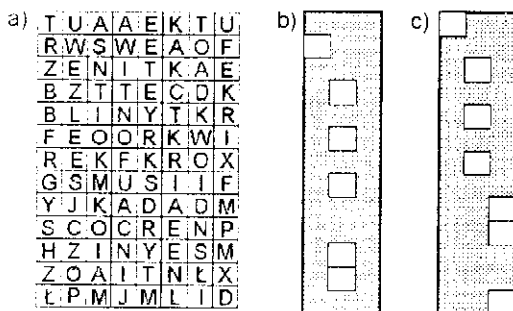
Najnowszym sposobem zwielokrotnienia kanału telekomunikacyjnego jest **zwielokrotnienie w dziedzinie kodowej** – CDMA (*Code Division Multiple Access*). Wszyscy użytkownicy mają dostęp do całego użytecznego pasma przez cały czas, ale posługują się różnymi „kluczami” (kodami), za pomocą których wprowadzają sygnały do kanału telekomunikacyjnego i je odbierają. Dla danego użytkownika zakodowane sygnały innych współużytkowników są nieistotne, najwyżej traktowane jako zakłócenie (szum) i na ogół nie przeszkadzają w transmisji (rys. 2.8).

System jest używany m.in. w systemach bezprzewodowego dostępu do Internetu (Wi-Fi), w telefonii komórkowej 3. generacji (częściowo także w telefonii GSM) i w systemie nawigacji GPS.



Rys. 2.8. Przedstawienie zwielokrotnienia w dziedzinie kodowej – wszyscy użytkownicy korzystają z pełnego pasma przez cały czas, znajdując się w innych „warstwach” kodowych.

Aby przybliżyć zasadę CDMA, można podać przykład, który częściowo oddaje jej istotę. Przedstawiona na rys. 2.9a tablica liter wydaje się bezsensowna – jest bowiem wypełniona pozornie chaotycznym układem liter – jest to odpowiednik całego kanału (i pasma) zajętego przez wszystkich użytkowników. Użytkownik A posługuje się kluczem przedstawionym na rys. 2.9b, który maskuje „niepotrzebne” litery i odsłania właściwe. Przykładając ten klucz do tablicy i przesuwając go poziomo wzdłuż kolejnych kolumn liter, może odczytać pewną informację. Podobnie użytkownik B, używając innego klucza (rys. 2.9c), może odczytać inną, przeznaczoną dla niego informację, natomiast informacji przeznaczonej dla użytkownika A nie odczyta. Obydwie informacje występują jednocześnie, ale dzięki różnym kluczom nie przeszkadzają sobie i trafiają do właściwych odbiorców.



Rys. 2.9. Ilustracja systemu CDMA: a) tablica liter – odpowiednik kanału zajętego przez wszystkich użytkowników, b) klucz do odczytywania informacji przez użytkownika A, c) klucz do odczytywania informacji przez użytkownika B.

2.5. Ogólne zależności dotyczące sygnałów

Podstawowym typem przebiegów i sygnałów stosowanych w telekomunikacji są przebiegi sinusoidalne. Wiadomo, że także inne przebiegi (np. prostokątne) dają się wyrazić jako superpozycja przebiegów sinusoidalnych. Ogólnie przebieg taki można zapisać jako

$$a(t) = A \cdot \cos \Phi(t), \quad (2.6)$$

gdzie A jest amplitudą, $\Phi(t)$ zaś jest fazą chwilową. W najprostszym przypadku

$$a(t) = A \cdot \cos(\omega t + \varphi), \quad (2.7)$$

gdzie ω jest pulsacją przebiegu, a φ jego fazą początkową.

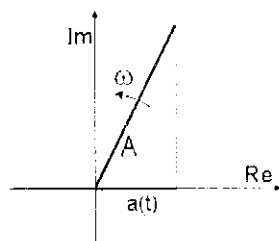
W zapisie można oczywiście także użyć funkcji sinus, ale tradycyjnie w zagadnieniach telekomunikacji używa się cosinusa, ze względu na jego parzystość, co ułatwia zapis i przekształcanie niektórych zależności. Używając zapisu zespolonego, można zapisać, że

$$a(t) = \operatorname{Re} \{ A \cdot e^{j\Phi(t)} \}. \quad (2.8)$$

Interpretację takiego zapisu pokazano na rys. 2.10.

Miarą wartości chwilowej przebiegu (2.8) jest rzut na oś rzeczywistą wektora zespolonego o długości równej A , obracającego się na płaszczyźnie zespolonej. Obrót w lewo (przeciwnie do kierunku ruchu wskazówek zegara) uważa

się za obrót w kierunku dodatnim; odpowiada to dodatniej pulsacji, gdyż faza chwilowa wtedy wzrasta.



Rys. 2.10. Interpretacja przebiegu harmonicznego jako rzutu wektora zespolonego na oś rzeczywistą.

Jeżeli przebieg (2.7) jest sygnałem, tzn. zawiera informację, to jego parametry muszą się jakoś zmieniać, i to w sposób będący pewnym zaskoczeniem dla obserwatora (odbiorcy). Informacja w takim sygnale może być zawarta albo w zmienności amplitudy, albo w zmienności fazy chwilowej.

Przy zapisie typu $a(t) = A \cdot \cos(\omega t + \varphi)$ faza chwilowa $\Phi(t)$ jest zmienna, ale w sposób jednoznacznie ustalony, i można ją wyznaczyć w dowolnej chwili w przyszłości, a więc taka zmienność fazy chwilowej nie przenosi informacji. Aby tak było, faza ta musi być zmienna inaczej – w zapisie (2.7) musi się zmieniać albo pulsacja, albo faza początkowa.

Z postaci zapisu (2.7) wynika bezpośrednio, która wielkość jest pulsacją, a która fazą początkową. Bywa jednak, że zapis sygnału jest na tyle skomplikowany, że bezpośrednie określenie pulsacji może być trudne. W takim przypadku korzysta się z najbardziej ogólnej zależności

$$\Omega(t) = \frac{d}{dt} \Phi(t), \quad (2.9)$$

gdzie $\Omega(t)$ jest pulsacją chwilową (na oznaczenie tej wielkości tradycyjnie stosuje się dużą literę omega). W konsekwencji zależność odwrotna to

$$\Phi(t) = \int_0^t \Omega(\tau) d\tau + \varphi, \quad (2.10)$$

gdzie φ jest stałą całkowania.

Przykład 2.1

Przy zapisie $a(t) = 5 \cdot \cos(1000t)$ oczywiste jest, że $\Omega = 1000/\text{s}$, natomiast $f = 1000/2\pi \text{ Hz}$. Ale w przypadku sygnału o postaci $a(t) = 5 \cdot \cos(1000 \cdot t^2)$ bezpośrednie określenie pulsacji nie jest możliwe, natomiast ze wzoru (2.9) otrzymuje się

$$\Omega(t) = \frac{d}{dt} \Phi(t) = \frac{d}{dt} (1000 \cdot t^2) = 2000 \cdot t.$$

Pulsacja jest więc zmienna – rośnie liniowo w funkcji czasu.

Zadania

1. Obliczyć częstotliwość chwilową przebiegu danego wyrażeniem $y(t) = \cos(\sin 2\pi t^2)$ dla $t = 1 \text{ s}$.
2. Jaka musi być jakość sygnału (S/N) w kanale telefonicznym ($B = 3,1 \text{ kHz}$), aby teoretyczna przepustowość tego kanału wyniosła 56 kb/s ?
3. Wyrazić stosunki mocy równe 0,002, 16, 20, 150 w decybelach.
4. W pewnym systemie telekomunikacyjnym $S/N = 20 \text{ dB}$. Ile razy wzrośnie jego teoretyczna przepustowość, jeżeli moc sygnału wzrośnie o 6 dB (inne parametry nie ulegają zmianie)?
5. Czy w kanale telekomunikacyjnym o szerokości pasma 2 MHz i $S/N = -30 \text{ dB}$ jest teoretycznie możliwa transmisja danych o przepływności 1 kb/s ?

3

Modulacja amplitudy

3.1. Sygnał zmodulowany amplitudowo

Modulacja amplitudy, oznaczana powszechnie jako AM (*Amplitude Modulation*), stanowi jedną z najdawniej stosowanych modulacji fali nośnej.

W liniowej modulacji amplitudy amplituda chwilowa sygnału $A(t)$ jest liniową funkcją sygnału modulującego

$$A(t) = A_0[1 + k \cdot x(t)], \quad (3.1)$$

gdzie współczynnik proporcjonalności k musi być tak dobrany, aby amplituda była zawsze wielkością nieujemną, a więc aby

$$k \cdot \min\{x(t)\} > -1. \quad (3.2)$$

Na podstawie zależności (2.7) i (3.1) zapis sygnału z modulacją amplitudy będzie następujący:

$$a_{AM}(t) = A_0[1 + k \cdot x(t)] \cdot \cos \Omega_0 t. \quad (3.3)$$

Na oznaczenie pulsacji lub częstotliwości nośnej używane będą duże litery Ω i F , w celu podkreślenia, że częstotliwość nośna tradycyjnie jest dużo większa niż częstotliwość sygnałów modulujących, do których stosowane będą oznaczenia małe.

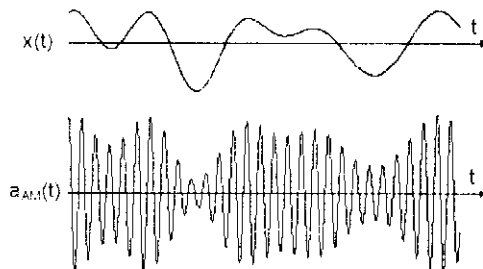
Przykładowy przebieg sygnału modulującego $x(t)$ i odpowiadający mu przebieg sygnału zmodulowanego AM przedstawiono na rys. 3.1.

Najprostszym przypadkiem modulacji amplitudy – i najczęściej rozważanym – jest modulacja jednym sygnałem sinusoidalnym. Niech ten sygnał ma postać

$$x(t) = X_m \cos(\omega_m t + \psi) \quad (3.4)$$

(ω_m jest pulsacją sygnału modulującego), wtedy z zależności (3.3)

$$a_{AM}(t) = A_0[1 + k \cdot X_m \cos(\omega_m t + \psi)] \cdot \cos \Omega_0 t. \quad (3.5)$$



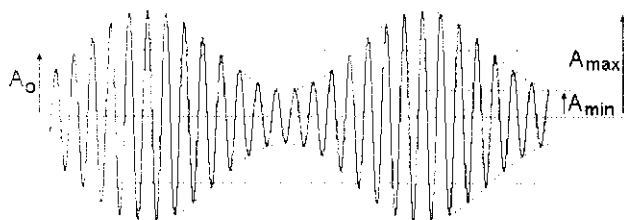
Rys. 3.1. Przykładowy sygnał modulujący i sygnał zmodulowany AM.

Można tu wprowadzić uproszczenie zapisu w postaci

$$a_{AM}(t) = A_o [1 + m \cdot \cos(\omega_{in} t + \psi)] \cdot \cos \Omega_o t, \quad (3.6)$$

gdzie $m = k \cdot X_m$ to **współczynnik głębokości modulacji**, $m \in [0, 1)$; bardzo często jest on wyrażany w procentach, wtedy $m \in [0, 100\%)$.

Przykładową czasową zmienność sygnału (3.6) pokazano na rys. 3.2.



Rys. 3.2. Charakterystyczne parametry sygnału AM zmodulowanego jednym sygnałem sinusoidalnym.

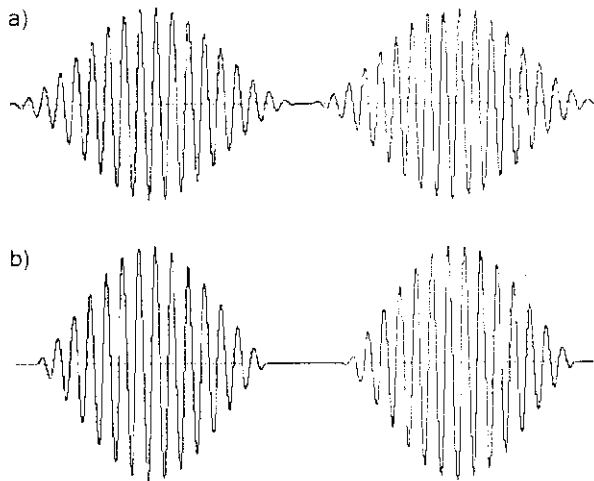
Z rysunku 3.2 i zależności (3.6) wynika, że $A_{max} = A_o(1 + m)$, $A_{min} = A_o(1 - m)$, skąd można wyznaczyć, że

$$m = \frac{A_{max} - A_{min}}{A_{max} + A_{min}}. \quad (3.7)$$

Jest to zależność o dużym znaczeniu praktycznym, umożliwia bowiem wyznaczanie współczynnika głębokości modulacji na podstawie oscyloskopowej obserwacji sygnału zmodulowanego. Warto zauważyć, że zależność (3.7) jest

słuszna także dla podwojonych wartości A_{\max} i A_{\min} , czyli dla odpowiednich wartości międzyszczytowych, które są łatwiejsze do odczytu z ekranu oscyloskopu niż amplitudy.

Graniczną wartością współczynnika głębokości modulacji jest $m = 100\%$ (rys. 3.3a), wtedy $A_{\min} = 0$ i występują chwilowe (punktowe) zaniki fali nośnej. Jeszcze większa głębokość modulacji, tj. naruszenie nierówności (3.2), prowadzi do tzw. przemodulowania, co powoduje powstanie poważnych zniekształceń sygnału (rys. 3.3b), dlatego sytuacji takiej w praktycznych systemach starannie się unika.



Rys. 3.3. Przebiegi AM: a) z głębokością modulacji $m = 100\%$,
b) z głębokością modulacji $m > 100\%$.

3.2. Widmo sygnału AM

Wyznaczenie widma sygnału to określenie, jakie składowe harmoniczne (sinusoidalne) są zawarte w rozważanym sygnale – o jakiej amplitudzie (widmo amplitudowe) i o jakiej fazie początkowej (widmo fazowe). Najbardziej praktyczna jest znajomość widma amplitudowego, bo to od niego zależy, czy np. sygnał „mieści się” w założonym pasmie częstotliwości. Dlatego dalsze rozważania dotyczące widm sygnałów zostaną ograniczone do widm amplitudowych.

Wyznaczenie widma sygnału AM jest dość proste, gdyż rozłożenie na składniki wyrażenia (3.6) daje

$$a_{AM}(t) = A_0 \cos \Omega_0 t + m A_0 \cos(\omega_m t + \psi) \cdot \cos \Omega_0 t. \quad (3.8)$$

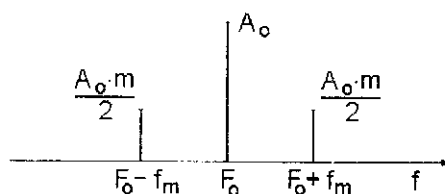
Na podstawie trygonometrycznej tożsamości $\cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2}$ wyrażenie $\cos(\omega_m t + \psi) \cdot \cos \Omega_o t$ można przedstawić w postaci sumy

$$\cos(\omega_m t + \psi) \cdot \cos \Omega_o t = \frac{1}{2} \cos[(\Omega_o + \omega_m)t + \psi] + \frac{1}{2} \cos[(\Omega_o - \omega_m)t - \psi]. \quad (3.9)$$

Dlatego sygnał AM jest sumą trzech przebiegów harmoniczných:

$$a_{AM}(t) = A_o \cos \Omega_o t + \frac{mA_o}{2} \cos[(\Omega_o + \omega_m)t + \phi_o + \psi] + \frac{mA_o}{2} \cos[(\Omega_o - \omega_m)t + \phi_o - \psi], \quad (3.10)$$

co można przedstawić graficznie w sposób pokazany na rys. 3.4.



Rys. 3.4. Amplitudowe widmo sygnału AM (3.10).

Jak widać, widmo sygnału AM zawiera prążek fali nośnej oraz dwa prążki boczne (wstęgi boczne), położone symetrycznie po przeciwnych stronach prążka fali nośnej. Amplituda tych prążków bocznych jest proporcjonalna do współczynnika głębokości modulacji m i nie przekracza połowy amplitudy fali nośnej.

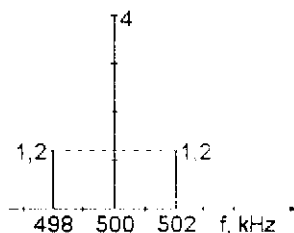
Przykład 3.1

Znaleźć widmo sygnału AM wyrażonego w postaci

$$a(t) = 4 \cdot [1 + 0.6 \cdot \sin(2\pi \cdot 2000 \cdot t)] \cdot \cos \pi \cdot 10^6 t.$$

Przez porównanie z zależnością (3.6) otrzymuje się $A_o = 4$, $m = 0.6$, $f_m = 2$ kHz, $F_o = 500$ kHz. Górny prążek boczny wystąpi przy częstotliwości 502 kHz, natomiast dolny – przy 498 kHz. Wysokość tych prążków będzie

$mA_0/2 = 0,6 \cdot 4/2 = 1,2$. W rezultacie widmo to można przedstawić graficznie jak na rys. 3.5.



Rys. 3.5. Widmo sygnału z przykładu 3.1.

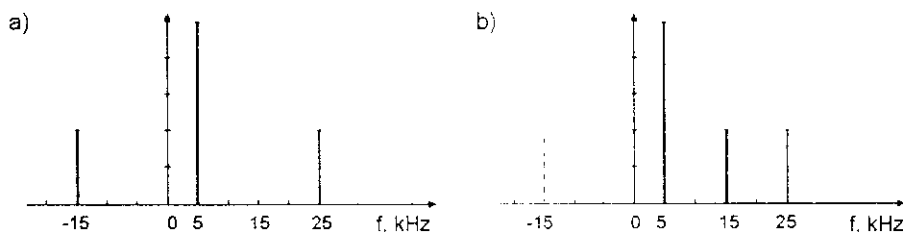
We wzorze (3.6) nie ma znaczenia, czy częstotliwość nośna jest większa od częstotliwości modulującej (tak zwykle jest), czy mniejsza. Jeżeli jest mniejsza, to trzeci składnik wyrażenia (3.10) będzie przedstawiał przebieg o częstotliwości ujemnej. Należy tu zaznaczyć, że o ujemnej częstotliwości można mówić tylko wówczas, gdy stosuje się zespolony zapis sygnału (rys. 2.10). Wtedy ujemna częstotliwość oznacza, że wektor, reprezentujący wielkość zespoloną, obraca się w prawo (w kierunku ruchu wskazówek zegara). Fizycznie istniejąca wielkość odpowiada rzutowi tego wektora na oś poziomą, a ten rzut i jego zmienność czasowa są takie same jak w przypadku obrotu wektora w lewo (co odpowiada częstotliwości dodatniej).

Dlatego ujemna lub dodatnia częstotliwość przebiegów zespolonych oznacza fizycznie to samo, zatem formalnie ujemna częstotliwość zostanie praktycznie odczytana jako „normalna” – dodatnia.

Przykład 3.2

Należy określić widmo amplitudowe sygnału AM o następujących parametrach: $A_0 = 3$, $m = 80\%$, $F_0 = 5$ kHz, $f_m = 20$ kHz.

Formalnie widmo to będzie miało postać jak na rys. 3.6a; będzie więc symetryczne, ale dolny prążek boczny będzie miał częstotliwość ujemną. Natomiast mierząc to widmo za pomocą np. analizatora widma lub innego



Rys. 3.6. Widma sygnału AM z rozważanego przykładu: a) widmo teoretyczne, b) widmo zmierzone praktycznie.

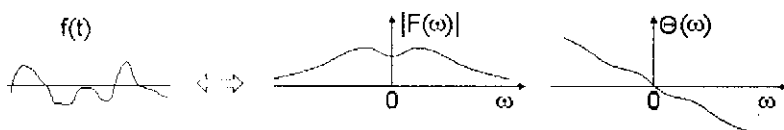
przypadku, stwierdzimy, że w widmie tym są zawarte składowe o dodatniej częstotliwości 5, 15 i 25 kHz (rys. 3.6b), zatem praktycznie widmo to nie będzie już symetryczne.

Niesymetria amplitudowego widma sygnału AM wskazuje na fakt, że częstotliwość nośna jest mniejsza od częstotliwości modulującej.

Należy zwrócić uwagę na to, że rozważane przypadki sygnału AM są przypadkami idealizowanymi. Jeżeli sygnały te są określone w przedziale czasu $(-\infty, +\infty)$, to nie przenoszą żadnej informacji, bo szerokość ich pasma jest równa zeru. Inaczej mówiąc, ściśle określona matematyczna forma sygnału powoduje, że sygnał ten może być odtworzony w dowolnej chwili w przyszłości, czyli nie stanowi dla odbiorcy żadnego zaskoczenia. Dopiero zmiany częstotliwości sygnału modulującego i jego poziomu spowodują, że pasmo sygnału będzie miało skończoną szerokość, co umożliwi przesłanie informacji.

Do określania widma dowolnych przebiegów bardzo przydatna jest transformata Fouriera. Określa ona gęstość widmową danego przebiegu, tzn. zawartość w nim przebiegów harmonicznym w nieskończenie wąskim pasmie wokół dowolnej pulsacji ω ; dana jest za pomocą zależności [2, 5, 6]:

$$F\{f(t)\} = \int_{-\infty}^{+\infty} f(t) \cdot e^{-j\omega t} dt = |F(\omega)| \cdot e^{j\Theta(\omega)}. \quad (3.11)$$



Rys. 3.7. Ilustracja przekształcenia Fouriera (3.11); pokazano przykładowe przebiegi $f(t)$, $|F(\omega)|$ i $\Theta(\omega)$.

Tak otrzymana transformata jest ogólnie zespolona i składa się z części amplitudowej $|F(\omega)|$ oraz fazowej $\Theta(\omega)$. Część amplitudowa dla rzeczywistych funkcji $f(t)$ jest symetryczna względem punktu $\omega = 0$, tj. zawiera składowe o ujemnej i dodatniej pulsacji (rys. 3.7). Należy pamiętać, że jest to podejście formalne; fizycznie składowe te mają częstotliwość dodatnią.

Wykorzystując właściwości przekształcenia Fouriera, można łatwo dojść do transformaty przebiegu zmodulowanego AM, czyli określić także jego widmo – dla dowolnego sygnału modulującego. Wiadomo, że transformata Fouriera iloczynu dwu funkcji czasu jest splotem odpowiadających im transformat, tj.

$$F\{f_1(t) \cdot f_2(t)\} = F\{f_1(t)\} * F\{f_2(t)\}, \quad (3.12)$$

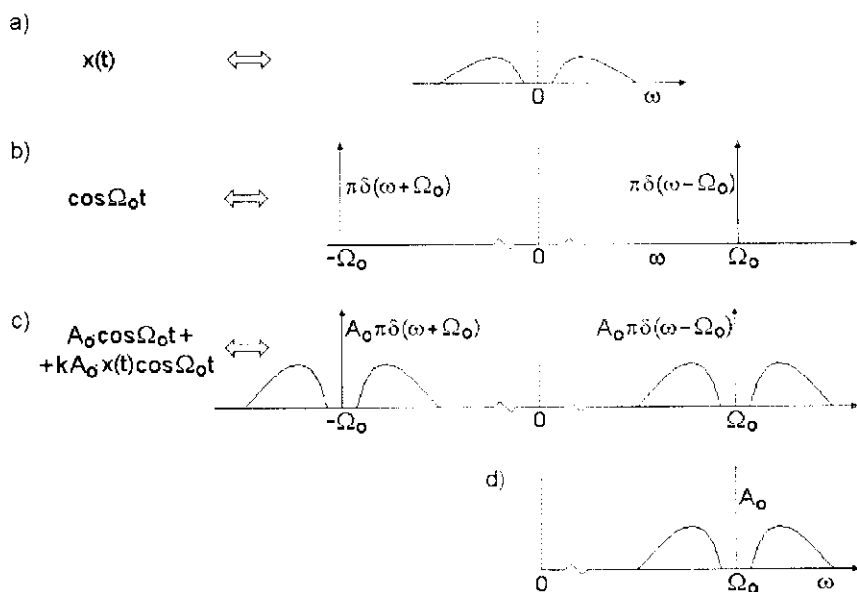
gdzie $*$ jest znakiem mnożenia splotowego. Zatem jeżeli

$$a_{AM}(t) = A_0[1 + kx(t)] \cdot \cos \Omega_0 t = A_0 \cos \Omega_0 t + kx(t) \cdot A_0 \cos \Omega_0 t,$$

to

$$F\{a_{AM}(t)\} = F\{A_0 \cos \Omega_0 t\} + F\{kx(t)\} * F\{A_0 \cos \Omega_0 t\}. \quad (3.13)$$

Na rysunku 3.8 pokazano transformaty Fouriera (ich część amplitudową) odpowiednich komponentów sygnału AM oraz transformatę „końcową”, będącą sumą transformaty z rys. 3.8b i splotu transformat z rys. 3.8a i b, z uwzględnieniem współczynników skalujących k i A_0 . Na rysunku 3.8d natomiast pokazano widmo sygnału AM (w dziedzinie pulsacji), mierzone fizycznie.

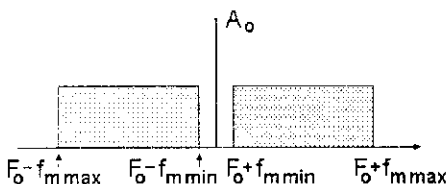


Rys. 3.8. Transformaty Fouriera składowych sygnału AM: a) przykładowego sygnału modulującego, b) fali nośnej, c) sygnału końcowego AM, d) widmo sygnału AM mierzone fizycznie.

Przedstawione transformaty Fouriera zawierają w pewnych miejscach funkcje delta Diraca, co oznacza, że wtedy gęstość widmowa amplitudy wyrażana np. w jednostkach V/Hz lub V/rad/s jest nieskończona! Nie jest to nienormalne, gdyż przecież np. „czysta”, nieskończona w czasie fala nośna ma zerową szerokość pasma (patrz rozdz. 2.), ale ma skończoną amplitudę. Jej gęstość widmowa jako iloraz tej amplitudy i szerokości pasma jest formalnie nieskoń-

czona. Natomiast pomiar widma takiego przebiegu za pomocą przyrządu, o zawsze skończonej szerokości pasma (określającego jego rozdzielczość widmową), to nic innego jak uśrednianie (lub całkowanie) tej gęstości w tym pasmie. W wyniku takiej operacji funkcje delta „znikają” i zmierzone amplitudy wszystkich składowych widma przybierają wartości skończone.

Ogólnie widmo sygnału AM może być przedstawione w sposób pokazany na rys. 3.9, bez wnikania w to, jaki jest konkretny poziom sygnału modulującego i jego częstotliwość (są one przecież zmienne!). Zakreskowanie prostokątów oznacza, że jest to obwiednia widma, a właściwe (lub chwilowe) widmo mieści się „wewnątrz” nich.



Rys. 3.9. Widmo sygnału AM
(umowny sposób przedstawienia).

3.3. Moc sygnału AM

Rozważany sygnał AM (3.10) jest sumą trzech przebiegów harmonicznym o różnej częstotliwości, dlatego moc tego sygnału jest równa sumie mocy poszczególnych składowych.

Aby wyznaczyć tę moc, należałoby wiedzieć, czy przebieg dany wzorem (3.10) przedstawia prąd czy napięcie, i znać rezystancję, w której ta moc jest wydzielana. Dla zachowania ogólności rozważań zakłada się, że moc ta jest wydzielana w jednostkowej rezystancji; wtedy nie ma potrzeby określania, czy rozważany przebieg jest napięciem czy prądem – po prostu jako moc przebiegu przyjmuje się kwadrat jego wartości skutecznej. Takie podejście, zgodne ze stosowanym w literaturze, będzie zachowane w niniejszym skrypcie. W tym przypadku

$$P_{AM} = \frac{A_0^2}{2} + \frac{1}{2} \cdot \left(\frac{mA_0}{2} \right)^2 + \frac{1}{2} \cdot \left(\frac{mA_0}{2} \right)^2, \quad (3.14)$$

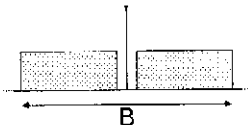
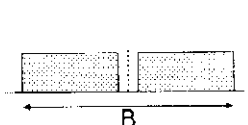
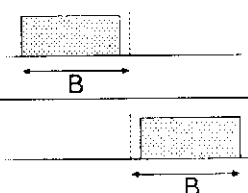
czyli inaczej

$$P_{AM} = \frac{A_0^2}{2} \left(1 + \frac{m^2}{2} \right). \quad (3.15)$$

Moc sygnału zmodulowanego jest większa od mocy fali nośnej o moc wstęg bocznych, w najlepszym razie ($m = 1$) jest ona większa tylko o 50%. Dlaczego tylko? Dlatego, że pożądane jest, aby moc wstęg bocznych była jak największa, gdyż o samej transmisji informacji i o jej zasięgu decydują właśnie wstęgi boczne (fala nośna nie przenosi informacji). Dla $m = 100\%$ zaledwie 33% całej mocy sygnału AM stanowi moc wstęg bocznych, pozostała część to moc „bezużytecznej” fali nośnej.

Z tego względu opracowano też inne, pozbawione tej wady, rodzaje modulacji amplitudy, będące modyfikacją modulacji „klasycznej”. Ich podstawowe cechy przedstawiono w dwóch ostatnich wierszach tab. 3.1.

Tab. 3.1. Zestawienie typów modulacji amplitudy i ich niektórych właściwości.

Rodzaj modulacji AM	Postać widma	Szerokość pasma	Stosunek mocy wstęg bocznych do mocy całego sygnału
Dwuwstęgowa z falą nośną – „klasyczna” AM		$B = 2 \cdot f_{\text{max}}$	$< 0,33$
Dwuwstęgowa bez fali nośnej AM DSB-SC		$B = 2 \cdot f_{\text{max}}$	1,0
Jednowstęgowa bez fali nośnej AM SSB-SC		$B = f_{\text{max}}$	1,0

Odmiany modulacji amplitudy, które przedstawiono w tabeli, mają na celu poprawienie niekorzystnego bilansu energetycznego, jakim cechuje się „klasyczna” dwuwstęgowa modulacja amplitudy z falą nośną.

Modulacja dwuwstęgowa bez fali nośnej DSB-SC (*Double Side Band – Suppressed Carrier*) różni się od modulacji klasycznej tylko brakiem fali nośnej. Modulacja jednowstęgowa bez fali nośnej SSB-SC (*Single Side Band – Suppressed Carrier*) to modulacja DSB-SC z usuniętą jeszcze jedną wstęgą boczną.

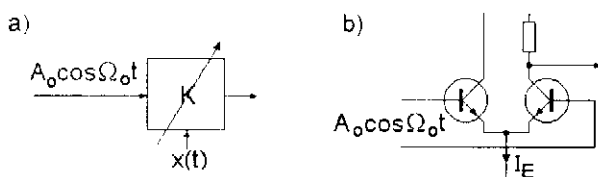
Sens tych modulacji jest taki, że skoro fala nośna nie przenosi informacji, to nie trzeba tej fali umieszczać w sygnale. Jest ona potrzebna tylko na etapie wytwarzania wstęg bocznych, ale potem już nie, dlatego można ją jakoś z pełnego sygnału AM usunąć, pozostawiając tylko jego użyteczną część – wstęgi

boczne. Z tego względu modulacja AM DSB-SC jest lepsza pod względem energetycznym niż modulacja AM.

Modulacja SSB-SC jest jeszcze dalszym „poprawieniem” sygnału zmodulowanego AM. Ze względu na omawianą już symetrię widma sygnału AM informacja zawarta w obydwu wstęgach bocznych jest taka sama. Nadmiar (redundancja) informacji jest korzystny z punktu widzenia niezawodności łączności, ale niepotrzebnie zajmuje pasmo częstotliwości, które mogłoby ulec dwukrotnemu zwężeniu. Węższe pasmo to mniejszy wpływ zakłóceń zewnętrznych, oddziaływających na przekazywany sygnał, i możliwość uzyskania większej czułości odbiornika (patrz rozdz. 7. i 8.). Ponadto węższe pasmo to możliwość zmieszczenia w określonym zakresie częstotliwości większej liczby niezależnych kanałów (w systemie FDMA).

3.4. Otrzymywanie sygnałów z modulacją AM

Jest wiele rozwiązań układowych prowadzących do uzyskania klasycznej modulacji amplitudy. Najbardziej oczywisty i często stosowany to zastosowanie wzmacniacza fali nośnej o wzmacnieniu regulowanym w sposób elektroniczny za pomocą sygnału modulującego (rys. 3.10a). Wzmacniacz taki to najczęściej układ pary różnicowej, gdyż jego wzmacnienie jest proporcjonalne do prądu I_E (rys. 3.10b).



Rys. 3.10. Wzmacniacz o regulowanym wzmacnieniu jako modulator amplitudy: a) ogólne oznaczenie, b) uproszczony schemat ideowy wzmacniacza o regulowanym wzmacnieniu.

Naturalny sposób otrzymywania modulacji DSB-SC wynika z zależności (3.10) – jeżeli usunąć z widma sygnału AM falę nośną, otrzyma się

$$a_{\text{DSB}}(t) = \frac{mA_0}{2} \cos[(\Omega_0 + \omega_m)t + \psi] + \frac{mA_0}{2} \cos[(\Omega_0 - \omega_m)t - \psi], \quad (3.16)$$

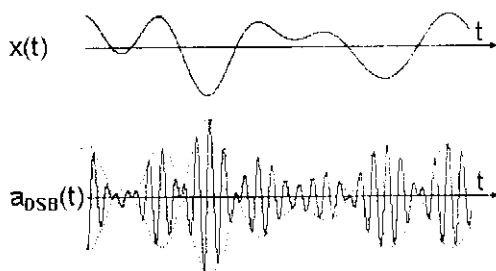
czyli

$$a_{\text{DSB}}(t) = m \cos(\omega_m t + \psi) \cdot A_0 \cos \Omega_0 t. \quad (3.17)$$

Sygnal DSB to po prostu fala nośna pomnożona przez sygnał modulujący; w ogólnym przypadku zaś można napisać, że

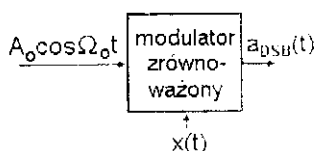
$$a_{\text{DSB}}(t) = x(t) \cdot A_o \cos \Omega_o t. \quad (3.18)$$

Przykładową postać sygnału AM DSB-SC przy danym sygnale $x(t)$ pokazano na rys. 3.11.



Rys. 3.11. Przykładowy sygnał modulujący i odpowiadający mu sygnał AM DSB-SC.

Sygnał zmodulowany AM DSB-SC można uzyskać przez odfiltrowanie fali nośnej z sygnału AM, ale technicznie byłoby to trudne. Istnieje natomiast wiele rozwiązań układowych modulatorów, w których fala nośna „z definicji” nie pojawia się na ich wyjściu; modulatory takie noszą nazwę zrównoważonych (rys. 3.12). Można je zbudować z wykorzystaniem mnożnika analogowego, na parach lub czwórkach identycznych elementów nieliniowych itp. [7, 9].



Rys. 3.12. Otrzymywanie sygnału AM DSB-SC z wykorzystaniem modulatora zrównoważonego (przypadek ogólny).

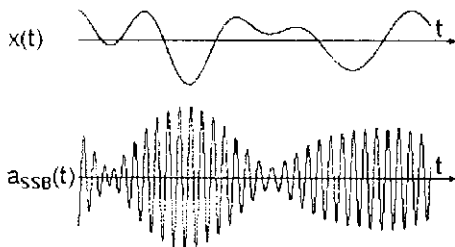
Sygnał SSB to sygnał AM zawierający tylko jedną wstęgę boczną i bez fali nośnej. Wskutek tego przebieg czasowy takiego sygnału niekiedy może wydawać się dziwny, gdyż jego obwódnicia nie zachowuje informacji o sygnale modulującym (rys. 3.13).

Zapis sygnału AM SSB-SC przy sinusoidalnym sygnale modulującym jest prosty. Z postaci sygnału AM DSB-SC (3.16)

$$a_{\text{DSB}}(t) = \frac{mA_o}{2} \cos[(\Omega_o + \omega_m)t + \psi] + \frac{mA_o}{2} \cos[(\Omega_o - \omega_m)t - \psi].$$

Przez usunięcie z niego jednej wstęgi bocznej uzyska się

$$a_{\text{SSB}}(t) = \frac{mA_o}{2} \cos[(\Omega_o + \omega_m)t + \psi]. \quad (3.19)$$



Rys. 3.13. Przykładowy sygnał modulujący i odpowiadający mu sygnał AM SSB-SC.

Zapis sygnału AM SSB-SC w przypadku ogólnym jest nieco trudniejszy. Wyrażenie (3.19) można rozpisać jako cosinus sumy kątów:

$$a_{\text{SSB}}(t) = \frac{mA_o}{2} [\cos \Omega_o t \cdot \cos(\omega_m t + \psi) - \sin \Omega_o t \cdot \sin(\omega_m t + \psi)], \quad (3.20)$$

skąd

$$a_{\text{SSB}}(t) = \frac{A_o}{2} [\cos \Omega_o t \cdot m \cos(\omega_m t + \psi) - \sin \Omega_o t \cdot m \sin(\omega_m t + \psi)]. \quad (3.21)$$

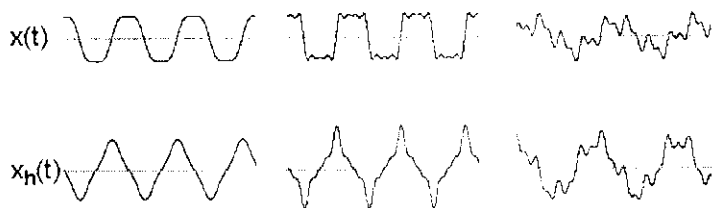
Wyrażenie $m \cdot \cos(\omega_m t + \psi)$ jest tu odpowiednikiem sygnału modulującego – zależność (3.4). Jak natomiast zinterpretować wyrażenie $m \cdot \sin(\omega_m t + \psi)$? Jest to też sygnał modulujący, ale przesunięty w fazie o 90° .

Pomijając tu szczegółowe uzasadnienia, okazuje się, że w ogólnym przypadku sygnał SSB należy zapisać jako

$$a_{\text{SSB}}(t) = A_o [\cos \Omega_o t \cdot x(t) - \sin \Omega_o t \cdot x_h(t)], \quad (3.22)$$

gdzie $x_h(t)$ jest takim przebiegiem czasowym, w jakim wszystkie składowe harmoniczne mają amplitudy identyczne z amplitudami sygnału modulującego $x(t)$,

lecz ich fazy są przesunięte o 90° . Postać czasowa takiego przebiegu może zupełnie nie przypominać sygnału oryginalnego (rys. 3.14).

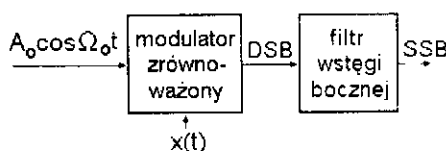


Rys. 3.14. Przykładowe przebiegi $x(t)$ i odpowiadające im przebiegi $x_h(t)$.

Matematycznie przebieg $x_h(t)$ o takich właściwościach można wyrazić za pomocą przekształcenia Hilberta [5]:

$$x_h(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau. \quad (3.23)$$

Modulatory SSB buduje się najczęściej z wykorzystaniem metody filtracji. Generuje się mianowicie sygnał AM DSC-SC, a następnie eliminuje się jedną z jego wstęg bocznych za pomocą filtru pasmowego o dużej selektywności (rys. 3.15).



Rys. 3.15. Ilustracja metody filtracji otrzymywania sygnału AM SSB-SC.

Filtry takie budowane są z użyciem rezonatorów piezoelektrycznych (kvarcowych), zapewniających wymaganą dużą stromość charakterystyki tłumienia w funkcji częstotliwości.

Inna metoda otrzymywania sygnału SSB to tzw. metoda fazowa, oparta na wzorze (3.22). Obecnie ma ona wartość raczej historyczną i nie będzie tu omawiana; można się z nią zapoznać np. w pracach [3, 5, 6].

3.5. Demodulacja sygnałów AM

Demodulacja to wydobycie sygnału oryginalnego z sygnału zmodulowanego.

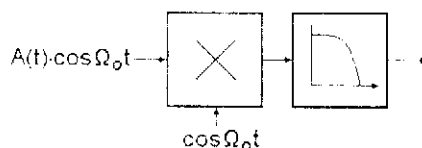
Bardzo dobrym sposobem demodulacji sygnału AM jest pomnożenie go przez falę nośną. Rzeczywiście, wskutek takiej operacji uzyskuje się

$$a_{AM}(t) \cdot \cos \Omega_0 t = \{A_c [1 + k \cdot x(t)] \cdot \cos \Omega_0 t\} \cdot \cos \Omega_0 t. \quad (3.24)$$

Wykonanie działań daje w efekcie

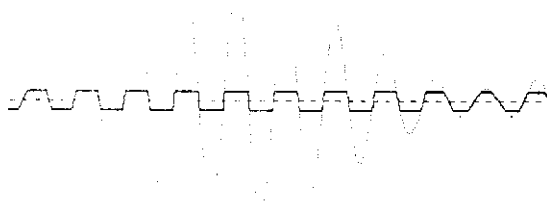
$$a_{AM}(t) \cdot \cos \Omega_0 t = \frac{A_c}{2} + \frac{A_c}{2} x(t) + \frac{A_c}{2} [1 + kx(t)] \cdot \cos 2\Omega_0 t. \quad (3.25)$$

Pierwszy składnik to składowa stała, niezależna od przekazywanej informacji, drugi jest proporcjonalny do pożądanego sygnału $x(t)$, trzeci składnik jest sygnałem zmodulowanym AM o połówkowej amplitudzie i podwójnej częstotliwości nośnej. Składnik ten może być odfiltrowany za pomocą względnie prostego filtra dolnoprzepustowego, jak to pokazano na rys. 3.16.



Rys. 3.16. Układowa realizacja demodulacji AM przez mnożenie przez falę nośną.

Demodulator o takiej zasadzie działania nazywa się **demodulatorem synchronicznym**. Pojawia się tu pytanie, skąd wziąć w miejscu demodulacji (tj. w odbiorniku) przebieg fali nośnej. Falę nośną można odfiltrować z sygnału wejściowego (co nie jest takie łatwe), można ją uzyskać z generatora lokalnego, jednak w razie niedokładnego jej odtworzenia może to doprowadzić do poważnych zniekształceń demodulacji. Na szczęście, istnieje łatwy sposób otrzymania „czystej” fali nośnej w odbiorniku, pokazany na rys. 3.17.



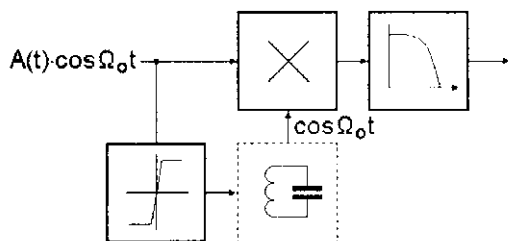
Rys. 3.17. Sygnał zmodulowany AM i przebieg otrzymany przez silne ograniczenie amplitudy.

Jest to fragment sygnału AM oraz przebieg otrzymany z niego po zastosowaniu ograniczania amplitudy na względnie niskim poziomie. Ograniczony przebieg zbliżony jest do przebiegu trapezowego czy wręcz prostokątnego (jeżeli poziom ograniczania będzie dostatecznie mały). Najcenniejszą cechą tego ograniczonego przebiegu jest to, że jego częstotliwość, określona przez punkty przechodzenia przez zero, jest dokładnie równa częstotliwości fali nośnej. Można się o tym przekonać, rozważając zależność (3.3):

$$a_{AM}(t) = A_o[1 + k \cdot x(t)] \cdot \cos \Omega_o t.$$

W jakich momentach przebieg ten przechodzi przez zero? – w momentach odpowiadających zerom czynników $[1 + k \cdot x(t)]$ oraz $\cos \Omega_o t$. Jeżeli modulacja jest liniowa i zachowana jest nierówność (3.2), pierwszy czynnik nie przybiera nigdy wartości zerowej. Zatem zera rozważanej zależności to tylko zera fali nośnej. Otrzymany przez ograniczenie przebieg przechodzi przez zero dokładnie w tych samych momentach czasu, w których przechodziłaby fala nośna, ma więc jej częstotliwość.

Bardziej rozbudowany schemat blokowy demodulatora synchronicznego z takim odtwarzaniem fali nośnej pokazano na rys. 3.18. Układ zawiera ogranicznik amplitudy oraz (niekoniecznie) obwód rezonansowy do odfiltrowania sinusoidalnej fali nośnej z przebiegu wyjściowego ogranicznika.

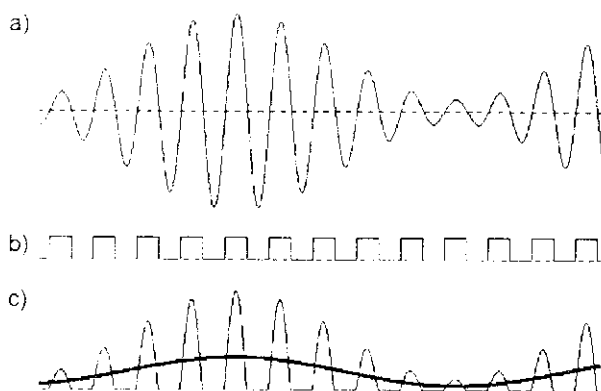


Rys. 3.18. Schemat blokowy demodulatora synchronicznego AM z układem odtwarzania nośnej.

W porównaniu z innymi układami demodulatorów amplitudy demodulator synchroniczny cechuje się większym skomplikowaniem i kosztem, jednak wykazuje dużą odporność na zakłócenia o częstotliwości różniącej się od częstotliwości nośnej. Z tego względu jest chętnie stosowany w bardziej odpowiedzialnych i profesjonalnych układach odbiorczych AM (np. w odbiornikach TV).

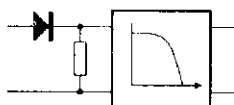
Inne metody demodulacji AM, znacznie prostsze, w istocie wynikają z zasady demodulacji synchronicznej.

Na rysunku 3.19 przedstawiono wynik pomnożenia sygnału zmodulowanego AM przez jednobiegunową falę nośną o przebiegu prostokątnym – w efekcie otrzymuje się przebieg, który odpowiada dokładnie jednopołówkowemu prostowaniu sygnału wejściowego. Można więc stwierdzić, że prostowanie jednopołówkowe jest równoważne mnożeniu przez (prostokątną) falę nośną; analogicznie, mnożenie przez dwubiegunową prostokątną falę nośną odpowiadać będzie prostowaniu dwupołówkowemu. Wyprostowane przebiegi można wygładzić w filtrze dolnoprzepustowym, aby otrzymać przebieg odtwarzający dokładnie obwiednię sygnału AM, czyli sygnał oryginalny.



Rys. 3.19. Związek prostowania przebiegu z demodulacją synchroniczną: a) sygnał oryginalny, b) prostokątna fala nośna, c) iloczyn poprzednich wielkości – grubszą linią pokazano wartość średnią tego iloczynu.

W ten sposób można stwierdzić, że demodulację AM można zrealizować w prostym układzie tzw. demodulatora prostownikowego (rys. 3.20).



Rys. 3.20. Najprostszy prostownikowy demodulator AM.

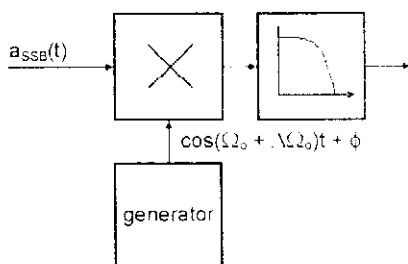
W układzie tym jest ważne, aby obecność filtra nie wpływała na pracę samego prostownika i aby napięcie wejściowe miało dostatecznie dużą wartość, żeby móc pominąć nieliniowość charakterystyki diody na początku zakresu przewodzenia.

Bardzo podobny układowo jest tzw. detektor obwiedni AM (rys. 3.21a). Pomimo tego podobieństwa jego zasada działania jest inna; cechuje się blisko trzykrotnie większym napięciem wyjściowym niż demodulator prostownikowy, ale i możliwością wystąpienia zniekształceń demodulacji w razie zbyt dużej stałej czasowej (będącej iloczynem rezystancji i pojemności) obciążenia.



Rys. 3.21. Demodulator obwiedni AM: a) schemat ideowy, b) przebieg napięcia wyjściowego (zaznaczono grubszą linią) przy zmodulowanym amplitudowo napięciu wejściowym.

Demodulacja sygnałów AM SSB-SC jest trudniejsza (w mniejszym stopniu dotyczy to także sygnałów AM DSB-SC). Można pokazać, że – tak jak poprzednio – pomnożenie tych sygnałów przez falę nośną prowadzi do prawidłowej demodulacji tych sygnałów. Niestety, nie można tu stosować odtwarzania fali nośnej, jak to pokazano na rys. 3.18, gdyż fali nośnej w sygnałach DSB i SSB po prostu nie ma! Pozostaje jedynie wytworzenie jej w specjalnym generatorze lokalnym w odbiorniku. Wbrew pozorom nie jest to łatwe, gdyż taka fala musi być odtworzona dokładnie co do fazy i co do częstotliwości, zgodnie z parametrami fali użytej w nadajniku (niekiedy bardzo odległym). Wszelkie, nawet niewielkie odchyłki częstotliwości i fazy, z którymi trzeba się zawsze liczyć (rys. 3.22), prowadzą do zniekształceń przebiegu zdemodulowanego.

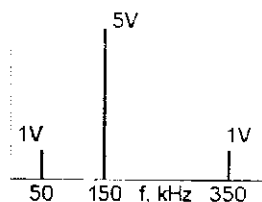


Rys. 3.22. Demodulator DSB i SSB z lokalnym wytwarzaniem fali nośnej z możliwym błędem częstotliwości i fazy.

Zniekształcenia te zmniejszają zrozumiałość np. przekazu fonicznego, zmieniając przede wszystkim widmo sygnału zdemodulowanego, co głosowi ludzkiemu nadaje charakterystyczne brzmienie („efekt kaczoza Donalda”), a w muzyce jest nie do przyjęcia. Dlatego transmisje bez fali nośnej stosowane są do przekazywania jedynie komunikatów słownych (telefonii), nie są natomiast stosowane do przekazywania dźwięku o lepszej jakości.

Zadania

1. Sygnał modulujący jest dany zależnością $x(t) = 3\cos\omega_1 t + 5\sin\omega_2 t$, $\omega_1 \neq \omega_2$.
We wzorze (3.1) tak dobrać współczynnik k , aby sygnał AM nie był przemodulowany (tj. aby $m < 100\%$).
2. Narysować widmo sygnału AM modulowanego sygnałem z poprzedniego zadania, jeżeli $k = 0,1$; $\omega_1 = 2\pi \cdot 5000 \text{ s}^{-1}$; $\omega_2 = 2\pi \cdot 7000 \text{ s}^{-1}$; $A_o = 10$; $F_o = 50 \text{ kHz}$.
3. Wyznaczyć parametry sygnału AM danego w postaci $a(t) = 4 \cdot [1 + 0,95 \cdot \cos(3 \cdot 10^4 t + 0,5)] \cdot \cos(2 \cdot 10^6 t)$.
4. Nadajnik AM dostarcza fali nośnej o mocy 400 W. Jaka może być największa moc wstęp bocznych (przy pełnej modulacji)?
5. Określić parametry (A_o , F_o , f_m , m) sygnału AM o widmie podanym na rysunku.



4

Modulacja kąta

4.1. Modulacja częstotliwości i fazy

Modulacja kąta polega na oddziaływaniu sygnału modulującego fazę chwilową przebiegu nośnego.

Faza chwilowa zależna jest od pulsacji, fazy początkowej i oczywiście od czasu. Sygnał modulujący może oddziaływać na pulsację lub fazę (jak „na razie”, czas nie jest sterowalny). W zależności od tego mówi się odpowiednio o modulacji częstotliwości lub modulacji fazy. Właściwości tych modulacji są bardzo podobne, dlatego określa się je ogólnie mianem modulacji kąta. Stosowane powszechnie skrótowe oznaczenia tych modulacji to odpowiednio FM (*Frequency Modulation*), PM (*Phase Modulation*) oraz ΦM .

Modulacja częstotliwości (liniowa) występuje wtedy, gdy częstotliwość (czyli także pulsacja) chwilowa jest liniową funkcją sygnału modulującego, co można zapisać jako

$$\Omega(t) = \Omega_0 + k_f \cdot x(t), \quad (4.1)$$

gdzie k_f jest odpowiednim współczynnikiem proporcjonalności. Wzór (4.1) można przyjąć jako definicję modulacji FM.

Jak wiadomo, faza chwilowa i pulsacja chwilowa są ze sobą związane wzorem (2.10). Na tej podstawie z zależności (4.1) można określić fazę chwilową sygnału FM. Będzie ona wynosić

$$\Phi(t) = \Omega_0 t + \phi_0 + k_f \int_0^t x(\tau) d\tau. \quad (4.2)$$

Modulacja fazy (liniowa) występuje wtedy, kiedy faza chwilowa jest (liniową) funkcją sygnału modulującego:

$$\Phi(t) = \Omega_0 t + \phi_0 + k_p \cdot x(t); \quad (4.3)$$

powoduje to, że pulsacja chwilowa wyraża się wówczas następująco:

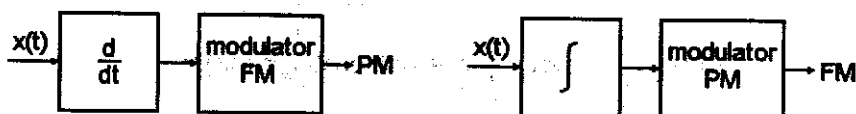
$$\Omega(t) = \Omega_0 + k_p \cdot dx/dt. \quad (4.4)$$

W celu porównania zależności te zestawiono w tab. 4.1.

Tab. 4.1. Podstawowe zależności dla modulacji częstotliwości i modulacji fazy (grubszą ramką zaznaczono wzory definicyjne).

Rodzaj modulacji ΦM	$\Omega(t)$ – pulsacja chwilowa	$\Phi(t)$ – faza chwilowa
FM	$\Omega(t) = \Omega_0 + k_f \cdot x(t)$	$\Phi(t) = \Omega_0 t + \phi_0 + k_f \int_0^t x(\tau) d\tau$
PM	$\Omega(t) = \Omega_0 + k_p \cdot dx/dt$	$\Phi(t) = \Omega_0 t + \phi_0 + k_p \cdot x(t)$

Warto zwrócić uwagę na duże podobieństwo wzorów zamieszczonych w tab. 4.1; w zasadzie różnią się one tylko tym, że zamiast sygnału $x(t)$ w jednej postaci występuje albo pochodna, albo całka tego sygnału. Można powiedzieć, że z dokładnością do całki lub pochodnej modulacje FM i PM są takie same. Dzięki temu można łatwo – przynajmniej teoretycznie – uzyskiwać modulację FM za pomocą modulatora PM i odwrotnie (rys. 4.1) przez uprzednie całkowanie lub różniczkowanie sygnału modulującego.



Rys. 4.1. Sposoby otrzymywania modulacji PM za pomocą modulatora FM i odwrotnie.

Uwaga

- Zależność (4.2) jest jedynym prawidłowym zapisem fazy chwilowej dla sygnału FM. Błędem byłoby tu np. (intuicyjne?) napisanie, że

$$\Phi(t) = [\Omega_0 + k_p x(t)] \cdot t + \phi_0,$$

wtedy bowiem pulsacja chwilowa, zgodnie z zależnością (2.9), wynosiłaby

$$\Omega(t) = \Omega_0 + k_f \cdot x(t) + k_f \cdot t \cdot dx/dt;$$

byłaby więc zależna nie tylko od sygnału $x(t)$, ale także od jego pochodnej.

4.2. Sygnał FM

Podobnie jak w przypadku AM, rozważony zostanie najprostszy przypadek modulacji FM i następnie PM, gdy sygnałem modulującym jest sygnał sinusoidalny:

$$x(t) = X_m \cos(\omega_m t + \psi).$$

Wówczas, zgodnie ze wzorem definicyjnym (4.1),

$$\Omega(t) = \Omega_o + k_f \cdot x(t) = \Omega_o + k_f \cdot X_m \cos(\omega_m t + \psi), \quad (4.5)$$

co można zapisać jako

$$\Omega(t) = \Omega_o + \Delta\Omega_o \cos(\omega_m t + \psi), \quad (4.6)$$

gdzie wielkość $\Delta\Omega_o$ jest dewiacją pulsacji nośnej, tj. maksymalną odchyłką pulsacji od wartości spoczynkowej. Z dewiacją pulsacji związana jest dewiacja częstotliwości – bardzo ważny parametr sygnału FM:

$$\Delta F_o = \Delta\Omega_o / 2\pi. \quad (4.7)$$

Na tej podstawie faza chwilowa sygnału FM wynosi

$$\begin{aligned} \Phi(t) &= \int_0^t \Omega(\tau) d\tau = \int_0^t \Omega_o + \Delta\Omega_o \cos(\omega_m \tau + \psi) d\tau = \\ &= \Omega_o t + \frac{\Delta\Omega_o}{\omega_m} \sin(\omega_m t + \psi) + \phi_o, \end{aligned} \quad (4.8)$$

gdzie bezwymiarowa wielkość

$$\frac{\Delta\Omega_o}{\omega_m} = \frac{k_f \cdot X_m}{\omega_m} = \frac{\Delta F_o}{f_m} = m_f \quad (4.9)$$

nazywa się **indeksem (wskaźnikiem) modulacji FM**; wartość tego indeksu ma duży wpływ na właściwości użytkowe sygnału FM.

Wartość indeksu modulacji FM zależy od dewiacji częstotliwości, czyli od amplitudy sygnału modulującego oraz od częstotliwości modulującej.

Przykład 4.1

W radiofonii UKF FM maksymalna dewiacja częstotliwości jest ograniczona do 75 kHz, natomiast częstotliwość modulująca podczas transmisji monofonicznej zawarta jest w zakresie (30 Hz, 15 kHz). Jakie wartości może przyjmować wtedy indeks modulacji FM?

Jeżeli sygnał modulujący miałby maksymalną amplitudę, to dewiacja będzie równa 75 kHz. Wtedy indeks m_f będzie zawarty w zakresie (5, 25000). Zmniejszenie amplitudy sygnału spowoduje zmniejszenie dewiacji i proporcjonalne zmniejszenie wartości indeksu m_f . Jego minimalna wartość jest bliska zeru – występuje to wówczas, gdy amplituda sygnału modulującego jest znikoma.

Podczas transmisji stereofonicznych wartości częstotliwości modulującej zawarte są w zakresie (30 Hz, 53 kHz), a ograniczenie dewiacji pozostaje. Wtedy indeks m_f teoretycznie może być równy (1,4; 25000). W praktycznym systemie stereofonii ponadakustyczna częstotliwość modulująca ma celowo obniżony poziom, dlatego dolna wartość indeksu m_f zostaje znacznie zmniejszona.

Wykorzystując indeks modulacji w zapisie fazy chwilowej $\Phi(t)$, uzyskuje się

$$\Phi(t) = \Omega_0 t + m_f \cdot \sin(\omega_m t + \psi) + \phi_0, \quad (4.10)$$

natomiast postać sygnału FM będzie wynosić

$$a_{FM}(t) = A_0 \cos[\Omega_0 t + m_f \sin(\omega_m t + \psi) + \phi_0]. \quad (4.11)$$

4.3. Sygnał PM

Jeżeli przyjętą postać sygnału $x(t) = X_m \cos(\omega_m t + \psi)$ podstawić do zależności (4.3), będącej definicją modulacji fazy, otrzyma się

$$\begin{aligned} \Phi(t) &= \Omega_0 t + \phi_0 + k_p \cdot X_m \cos(\omega_m t + \psi) = \\ &= \Omega_0 t + \phi_0 + \Delta\Phi_0 \cos(\omega_m t + \psi). \end{aligned} \quad (4.12)$$

Przez analogię wielkość $\Delta\Phi_0$ nazywa się dewiacją fazy; jest to maksymalna odchyłka fazy od jej „ustalonej” wartości, wynikającej z liniowego narastania z czasem. Wtedy

$$a_{PM}(t) = A_0 \cos[\Omega_0 t + \phi_0 + \Delta\Phi_0 \cos(\omega_m t + \psi)]. \quad (4.13)$$

Modulacji fazy towarzyszy zawsze modulacja częstotliwości. Obliczając pochodną fazy chwilowej (4.12), uzyskuje się

$$\Omega(t) = \Omega_0 - \Delta\Phi_0 \omega_m \sin(\omega_m t + \psi), \quad (4.14)$$

dlatego występująca wtedy dewiacja częstotliwości

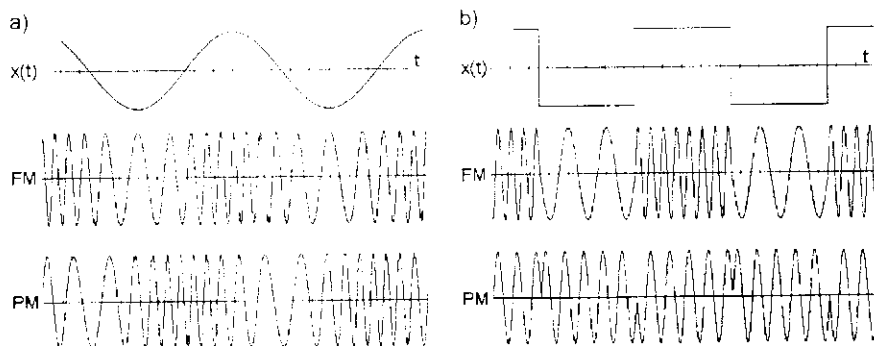
$$\Delta F_0 = \Delta\Phi_0 f_m. \quad (4.15)$$

W tabeli 4.2 zestawiono wyrażenia na parametry sygnałów FM i PM przy sinusoidalnym sygnale modulującym.

Tab. 4.2. Parametry sygnałów PM i FM przy sygnale modulującym $x(t) = X_m \cos(\omega_m t + \psi)$.

Parametr	FM	PM
Faza chwilowa	$\Phi(t) = \Omega_0 t + m_f \sin(\omega_m t + \psi) + \phi_0$	$\Phi(t) = \Omega_0 t + \Delta\Phi_0 \cos(\omega_m t + \psi) + \phi_0$
Pulsacja chwilowa	$\Omega(t) = \Omega_0 + \Delta\Omega_0 \cos(\omega_m t + \psi)$	$\Omega(t) = \Omega_0 - \Delta\Phi_0 \omega_m \sin(\omega_m t + \psi)$
Dewiacja częstotliwości	$\Delta F_0 = \frac{k_f \cdot X_m}{2\pi}$	$\Delta F_0 = \Delta\Phi_0 f_m = k_p X_m f_m$
Dewiacja fazy	$\Delta\Phi_0 = m_f = \frac{k_f \cdot X_m}{\omega_m}$	$\Delta\Phi_0 = k_p \cdot X_m$

Przy modulacji jednym tonem sinusoidalnym sygnały $a_{FM}(t)$ i $a_{PM}(t)$ są prawie nierozróżnialne, np. podczas obserwacji oscyloskopowej (rys. 4.2a), natomiast można zauważyć różnice między nimi, gdy sygnał modulujący zawiera więcej składowych harmonicznych (rys. 4.2b).

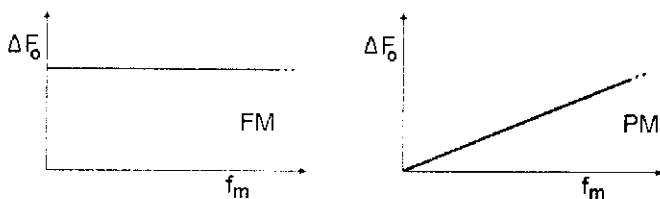


Rys. 4.2. Postać czasowa sygnałów FM i PM: a) podczas modulacji jednym tonem sinusoidalnym, b) podczas modulacji sygnałem złożonym.

Przykład 4.2

Dany jest pewien modulator kąta (zamknięty w „czarnej skrzynce”), ale nie wiadomo, czy to jest modulator częstotliwości, czy modulator fazy. Czy można to ustalić w wyniku pomiarów zewnętrznych?

Jest to możliwe, jeżeli do dyspozycji będzie miernik dewiacji częstotliwości. Należy wówczas zmierzyć dewiację częstotliwości sygnału wyjściowego tego modulatora przy dwóch różnych wartościach częstotliwości sinusoidalnego sygnału modulującego o tej samej amplitudzie. W przypadku modulatora FM dewiacja częstotliwości się nie zmieni, natomiast dla modulatora PM zwiększy się przy wzroście częstotliwości modulującej. Wynika to z zależności podanych w przedostatnim wierszu tab. 4.2 i zilustrowanych na rys. 4.3.



Rys. 4.3. Zależność dewiacji częstotliwości od częstotliwości modulującej dla FM i PM.

Fakt, że w przypadku modulacji FM dewiacja częstotliwości nie zależy od częstotliwości modulującej, pozwala na dość łatwą transmisję sygnałów szerokokopasmowych, pasmo bowiem sygnału FM nie jest liniowo uzależnione od częstotliwości modulującej, jak to było w przypadku modulacji AM. Będzie o tym mowa w następnym punkcie.

4.4. Widmo sygnałów z modulacją kąta

W celu pewnego uproszczenia skomplikowanych zapisów wyrażeń w rozważaniach dotyczących widma pominięte zostaną fazy początkowe fali nośnej i sygnału modulującego. Oczywiście jest, że widmo, czyli zawartość składowych harmonicznych w sygnale (ich częstotliwości i amplitudy), nie może zależeć od przyjętych faz początkowych. Rozważona zostanie modulacja FM jako znacznie częściej spotykana w praktyce.

Sygnał FM o ogólnej postaci (4.11) $a_{FM}(t) = A_o \cos(\Omega_o t + m_f \cdot \sin \omega_m t)$ można przedstawić z wykorzystaniem zapisu zespolonego

$$a_{FM}(t) = \operatorname{Re} \left\{ A_o e^{j(\Omega_o t + m_f \cdot \sin \omega_m t)} \right\} = \operatorname{Re} \left\{ A_o e^{j\Omega_o t} \cdot e^{jm_f \sin(\omega_m t)} \right\}. \quad (4.16)$$

Można powiedzieć, że zapisany w ten sposób sygnał to fala nośna „pomnożona” przez widmo (składniki) wynikające z wyrażenia $e^{jm_f \sin(\omega_m t)}$.

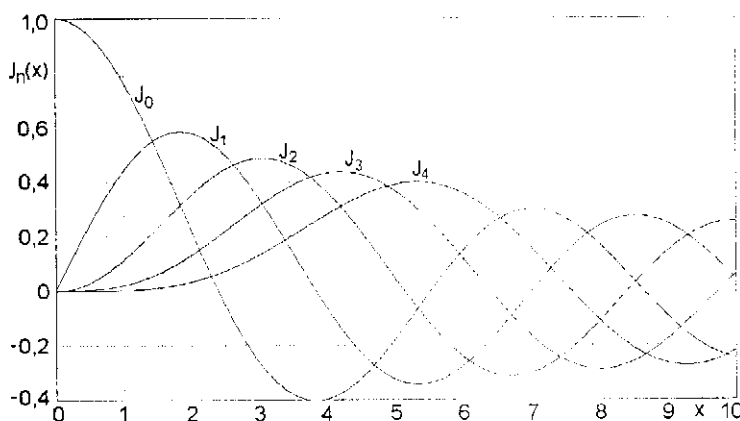
Ostatnie wyrażenie można przedstawić następująco [6]:

$$e^{jm_f \sin \omega_m t} = \sum_{n=-\infty}^{\infty} J_n(m_f) \cdot e^{jn\omega_m t} \quad (4.17)$$

gdzie współczynniki szeregu J_n są wartościami funkcji Bessela pierwszego rodzaju n -tego rzędu argumentu m_f . Całościowo jest to po prostu zespolony szereg Fouriera, w którym współczynniki szeregu są równe tym wartościom funkcji Bessela. Funkcje Bessela to jedne z ważnych funkcji nieelementarnych, spotykanych często w różnych zagadnieniach fizycznych. Ich wartości są tabelaryzowane [1], ponadto można je dość łatwo wyznaczyć numerycznie, szczególnie z zależności [1]:

$$J_n(x) = \frac{1}{\pi} \int_0^{\pi} \cos(x \cdot \sin \Theta - n \cdot \Theta) d\Theta. \quad (4.18)$$

Przebiegi funkcji Bessela pierwszego rodzaju rzędów 0–4 pokazano na rys. 4.4.



Rys. 4.4. Przebieg funkcji Bessela pierwszego rodzaju rzędów 0–4 dla argumentu w zakresie 0–10.

Funkcje Bessela przypominają funkcje trygonometryczne sinus i cosinus, ale ich amplitudy maleją do zera w przypadku dużych argumentów, funkcje wyższych rzędów w przypadku małych argumentów są bliskie zeru; zaczynają się istotnie różnić od zera tym później, im wyższy jest rząd funkcji (widać to też z tab. 4.3).

Sygnał FM można zatem przedstawić jako

$$a_{FM}(t) = \operatorname{Re} \left\{ A_o e^{j\Omega_o t} \cdot \sum_{n=-\infty}^{\infty} J_n(m_f) \cdot e^{jn\omega_m t} \right\}. \quad (4.19)$$

Po wykonaniu mnożenia i obliczeniu części rzeczywistej otrzyma się

$$\begin{aligned} a_{FM}(t) = & A_o J_0(m_f) \cdot \cos \Omega_o t + \\ & + A_o J_1(m_f) \cdot \cos(\Omega_o + \omega_m)t - A_o J_1(m_f) \cdot \cos(\Omega_o - \omega_m)t + \\ & + A_o J_2(m_f) \cdot \cos(\Omega_o + 2\omega_m)t + A_o J_2(m_f) \cdot \cos(\Omega_o - 2\omega_m)t + \\ & + A_o J_3(m_f) \cdot \cos(\Omega_o + 3\omega_m)t - A_o J_3(m_f) \cdot \cos(\Omega_o - 3\omega_m)t + \dots \end{aligned} \quad (4.20)$$

Uwzględniono tu następujący związek dla funkcji Bessela:

$$J_{-n}(m_f) = (-1)^n J_n(m_f). \quad (4.21)$$

Widmo sygnału FM zawiera nieskończenie (!) wiele składników, położonych na osi częstotliwości w odległości od siebie równej f_m . Należy podkreślić, że widmo to jest skomplikowane (i nieskończone) nawet dla najprostszego przypadku pojedynczego tonu modulującego – a co będzie, gdy sygnał modulujący będzie zawierał składowe o różnych wartościach częstotliwości? Wtedy widmo stanie się jeszcze bardziej skomplikowane, wystąpią bowiem składowe o różnych kombinacyjnych wartościach częstotliwości. Pod tym względem modulacja FM jest modulacją nieliniową – jej widmo w przypadku sumy sygnałów modulujących nie jest sumą widm od poszczególnych składowych.

Widmo sygnału FM jest teoretycznie nieskończone, ale w praktyce ograniczone, bo amplitudy funkcji Bessela maleją ze wzrostem argumentu; jednocześnie funkcje wyższych rzędów też przybierają coraz to mniejsze wartości (tab. 4.3).

Tab. 4.3. Wartości funkcji Bessela dla różnych wartości argumentu.

m_f	1	2	3	4	5
$J_1(m_f)$	0,4401	0,5767	0,3391	0,0659	0,3275
$J_2(m_f)$	0,1148	0,3528	0,4860	0,3641	0,0465
$J_3(m_f)$	0,0196	0,1290	0,3091	0,4302	0,3649
$J_4(m_f)$	0,0020	0,0339	0,1320	0,2810	0,3912
$J_5(m_f)$	0,0003	0,0007	0,0430	0,1321	0,2612
$J_6(m_f)$	$<10^{-4}$	0,0017	0,0113	0,0490	0,1310
$J_7(m_f)$	$<10^{-4}$	0,0002	0,0026	0,0152	0,0534
$J_8(m_f)$	$<10^{-4}$	$<10^{-4}$	0,0005	0,0040	0,0184

Można przyjąć, że do „praktycznego” widma zaliczy się te składowe, których amplituda przekracza 1% amplitudy fali nośnej w przypadku bez modulacji, czyli te, których moc jest większa od 10^{-4} mocy tej fali nośnej. Jest to dość przesadne wymaganie, dlatego niekiedy uwzględnia się tylko składowe o amplitudzie większej od 10% fali nośnej, czyli o mocy większej od 1%.

Za pomocą pogrubionych krawędzi w tab. 4.3 pokazano granice, kiedy wartości funkcji Bessela stają się w przybliżeniu mniejsze od 0,01 lub równe 0,01. Można zauważyć, że liczbę koniecznych do uwzględnienia składowych widma można wyznaczyć z empirycznego wzoru

$$n \approx m_f + 2 \quad (4.22a)$$

dla warunku $J_n(m_f) > 0,01$ lub ze wzoru

$$n \approx m_f + 1 \quad (4.22b)$$

dla warunku $J_n(m_f) > 0,1$. Z tych względów widmo sygnału FM można z dobrym przybliżeniem uznać za skończone i określić szerokość pasma sygnału FM jako

$$B = 2(m_f + 2)f_m = 2(\Delta F_o + 2f_m) \quad (4.23a)$$

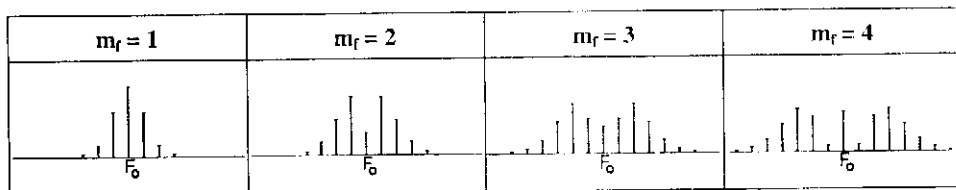
albo

$$B = 2(m_f + 1)f_m = 2(\Delta F_o + f_m). \quad (4.23b)$$

Ostatni wzór nosi nazwę wzoru Carsona [5].

Przykładowe widma sygnału FM w przypadku kilku wartości indeksu m_f pokazano w tab. 4.4.

Tab. 4.4. Widma modulacji FM dla kilku wartości indeksu modulacji.



Przykład 4.3

Ile wynosi szerokość pasma w radiofonicznych transmisjach UKF FM? Jak wiadomo, dewiacja jest ograniczona do 75 kHz. Podczas transmisji monofonicznej $f_{m \max} = 15$ kHz, dlatego

$$B = 2(\Delta F_o + 2f_m) = 2(75 \text{ kHz} + 2 \cdot 15 \text{ kHz}) = 210 \text{ kHz}$$

lub

$$B = 2(\Delta F_o + f_m) = 2(75 \text{ kHz} + 15 \text{ kHz}) = 180 \text{ kHz}$$

(na podstawie wzoru Carsona).

Podczas transmisji stereofonicznej $f_{m \max} = 53 \text{ kHz}$, zatem

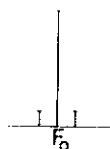
$$B = 2(\Delta F_o + 2f_m) = 2(75 \text{ kHz} + 2 \cdot 53 \text{ kHz}) = 362 \text{ kHz}$$

lub na podstawie zależności (4.23b)

$$B = 2(\Delta F_o + f_m) = 2(75 \text{ kHz} + 53 \text{ kHz}) = 256 \text{ kHz}.$$

Szerokość pasma sygnału stereofonicznego jest wyraźnie większa niż monofonicznego. Powinno to być oczywiste w świetle ogólnych rozważań z rozdz. 2., bo przecież pierwszy sygnał zawiera więcej informacji, czyli do jego przesłania potrzebny jest kanał o większej przepustowości.

Obliczone dotąd pasma sygnału FM są wielokrotnie większe od maksymalnej częstotliwości modulującej: modulacja FM jest więc **modulacją szerokopasmową** ($B \gg f_{m \max}$). Ta szerokopasmowość występuje wtedy, gdy indeks modulacji ma znaczną wartość. Gdy indeks m_f jest mały ($m_f \ll 1$), widmo redukuje się jedynie do fali nośnej i dwu pierwszych prążków bocznych (rys. 4.5), a szerokość pasma staje się równa podwojonej częstotliwości modulującej – wzory (4.22) wtedy nie obowiązują! Taki przypadek nazywa się **wąskopasmową modulacją FM**. Na pierwszy rzut oka można by stwierdzić, że na rys. 4.5 przedstawione jest widmo sygnału AM (dla $m \approx 25\%$). Czy to oznacza, że sygnał FM jest wtedy równoważny sygnałowi AM?

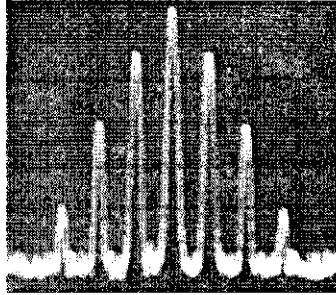


Rys. 4.5. Widmo modulacji FM
dla $m_f = 0,25$.

Odpowiedź jest negatywna – widma AM i FM nie są wtedy identyczne, bo chociaż widma amplitudowe są takie same, to jednak fazowe zależności (które nie były tu rozważane) między składowymi tych widm są inne w każdym przypadku.

Na rysunku 4.6 przedstawiono obraz otrzymany na ekranie analizatora widma, gdy do jego wejścia doprowadzono przebieg FM o indeksie modulacji $m_f = 1$. Widoczny jest największy prążek fali nośnej oraz po 3 ($3 = m_f + 2$) prążki po jego lewej i prawej stronie; ich wysokość względem prążka fali nośnej

wyduje się nie taka jak pokazano w tab. 4.3, ale jest to skutek logarytmicznej podziałki pionowej analizatora.



Rys. 4.6. Widmo sygnału FM dla $m_f = 1$ uzyskane na analizatorze widnia.

Widma modulacji PM są podobne do przedstawionych wyżej widm FM, z tą różnicą że rolę parametru, mającego istotny wpływ na ich postać, ma dewiacja pulsacji $\Delta\Phi_o$ (zamiast m_f).

4.5. Moc sygnału FM

Moc sygnału FM to moc przebiegu o stałej amplitudzie, niezależnej od modulacji, zatem wyraża się wzorem

$$P_{FM} = \frac{A_o^2}{2}. \quad (4.24)$$

Na zagadnienie to można spojrzeć też inaczej – moc sygnału FM to moc wszystkich składowych widma tego sygnału, a więc składowych, których amplitudy są odpowiednio równe $A_o J_n(m_f)$. Dlatego

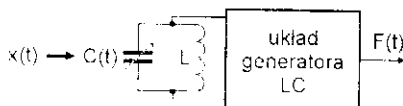
$$P_{FM} = \sum_{-\infty}^{\infty} \frac{[A_o J_n(m_f)]^2}{2} = \frac{A_o^2}{2} \sum_{-\infty}^{\infty} J_n^2(m_f). \quad (4.25)$$

Ze względu na znaną właściwość funkcji Bessela, że suma ich kwadratów jest równa jedności, otrzymuje się wzór (4.24).

Przy dużych indeksach modulacji widmo sygnału jest szerokie, ale amplitudy poszczególnych składowych są małe. Jest to konieczność, bo skoro przybywa składowych sygnału, to ich amplitudy muszą maleć, aby ich łączna moc pozostała stała.

4.6. Otrzymywanie modulacji FM

Najbardziej rozpowszechnionym sposobem uzyskiwania sygnałów FM jest bezpośrednia modulacja generatora fali nośnej. Nie można tu zastosować osobnego generatora fali nośnej i osobnego modulatora FM – modulacja musi się odbywać w samym generatorze. Może to być generator elektroniczny LC (rys. 4.7), w którym o częstotliwości wytwarzanego przebiegu decyduje obwód rezonansowy, zawierający sterowaną elektrycznie pojemność; może to być także generator RC sinusoidalny lub relaksacyjny, przestrzajany w sposób elektryczny (zwykle napięciem). W pokazanym na rys. 4.6 przypadku częstotliwość wyjściowa jest dana w przybliżeniu przez częstotliwość rezonansową obwodu LC:



Rys. 4.7. Bezpośrednia modulacja FM generatora LC.

$$F_o = F_o(t) \approx \frac{1}{2\pi\sqrt{L \cdot C(t)}}. \quad (4.26)$$

Na tej podstawie można pokazać, że

$$\frac{dF_o}{F_o} = -\frac{1}{2} \cdot \frac{dC}{C}. \quad (4.27)$$

Zatem, aby uzyskać np. 1% dewiacji częstotliwości fali nośnej ($\pm 1\%$), pojemność obwodu rezonansowego powinna się zmieniać o $\pm 2\%$. Jako sterowaną elektrycznie pojemność powszechnie stosuje się diody pojemnościowe (warikapy).

Warto dodać, że uzyskiwanie sygnału FM przez wykorzystanie modulatora PM (rys. 4.1) jest możliwe w zasadzie tylko teoretycznie, gdyż do uzyskania dużych dewiacji częstotliwości przy małej częstotliwości modulującej modulator fazy musiałby zapewniać zmiany przesunięcia fazy np. o kilkaset lub jeszcze więcej radianów (!). Wynika to bezpośrednio z zależności (4.15).

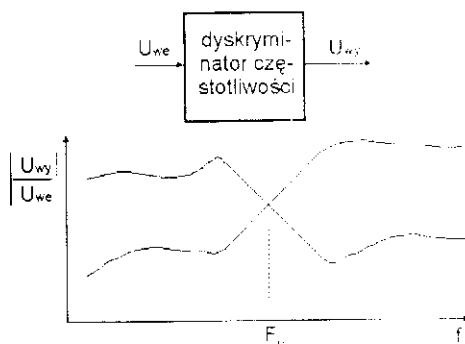
4.7. Demodulacja sygnałów FM

Zastosowanie prostych demodulatorów diodowych do demodulacji FM nie wchodzi w rachubę, bo one reagują na amplitudę sygnału wejściowego, a w przypadku FM jest ona stała. Jak można sprawdzić, także demodulator syn-

chroniczny w rodzaju przedstawionego na rys. (3.14) nie będzie pracował poprawnie.

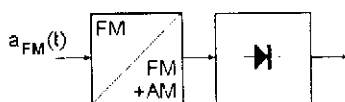
Chęć wykorzystania bardzo prostych demodulatorów diodowych AM (kiedyś tylko takie były dostępne...) do demodulacji sygnałów FM doprowadziła do opracowania demodulatora FM z tzw. dyskryminatorem częstotliwości. Dyskryminator ten to układ zamieniający modulację FM na dokładnie jej odpowiadającą modulację AM (odpowiadającą pod względem zawartości informacji). Wtedy do wydobywania informacji z tak przetworzonego sygnału wystarczy już sam demodulator AM.

Dyskryminator częstotliwości może być wykonany z wykorzystaniem układów zawierających obwody rezonansowe lub inne elementy selektywne. W zasadzie może to być dowolny układ, byle w wymaganym zakresie częstotliwości miał monotonicznie (i liniowo) malejący lub rosnący przebieg modułu transmitancji (rys. 4.8).



Rys. 4.8. Ogólna postać dyskryminatora częstotliwości i jego przykładowe charakterystyki częstotliwościowe.

Należy podkreślić, że dyskryminator nie zmienia częstotliwości sygnału wejściowego, dlatego na wyjściu zostaje zachowana wejściowa modulacja częstotliwości, ale zostaje ona uzupełniona o dodatkową (tutaj bardzo istotną) modulację amplitudy. Dlatego można powiedzieć, że ten dyskryminator to przetwornik FM/FM+AM. Na rysunku 4.9 pokazano schemat blokowy demodulatora FM z wykorzystaniem dyskryminatora częstotliwości i klasycznego demodulatora amplitudy.

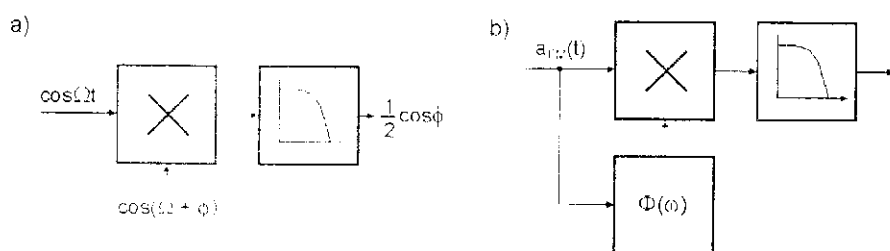


Rys. 4.9. Schemat blokowy demodulatora FM z dyskryminatorem częstotliwości.

Współcześnie powszechnie stosowany jest układ demodulatora FM, nazywany demodulatorem iloczynowym lub koincydencyjnym (niekiedy kwadraturowym). Zasada jego pracy opiera się na schemacie blokowym przedstawionym na rys. 4.10a. Napięcie wyjściowe w takim układzie zależy tylko od wzajemnego przesunięcia obydwu przebiegów wejściowych, a nie zależy od ich częstotliwości (jednakowej). Rzeczywiście, przebieg na wyjściu mnożnika można zapisać jako

$$\cos \Omega t \cdot \cos(\Omega t + \varphi) = \frac{1}{2} \cos \varphi + \frac{1}{2} \cos(2\Omega t + \varphi); \quad (4.28)$$

drugi składnik prawej strony tego wyrażenia przedstawia przebieg o wysokiej częstotliwości, który może być odfiltrowany przez filtr dolnoprzepustowy.



Rys. 4.10. Iloczynowy demodulator FM: a) podstawa działania, b) pełny schemat blokowy demodulatora.

Układ demodulatora iloczynowego (rys. 4.10b) zawiera przesuwnik fazowy o przesunięciu fazowym $\varphi(\omega)$ zależnym od częstotliwości. Dlatego na wyjściu układu otrzymuje się przebieg, którego wartość chwilowa jest zależna od częstotliwości chwilowej sygnału wejściowego (a ta w sygnale FM zależy od sygnału modulującego).

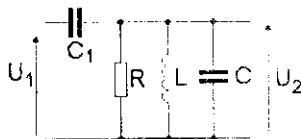
Jeżeli jako przesuwnik fazowy wykorzystać układ o postaci jak na rys. 4.11, przesunięcie fazowe będzie się wyrażało wzorem

$$\varphi \approx \frac{\pi}{2} - \arctg(2Q \frac{F(t) - f_0}{f_0}), \quad (4.29)$$

gdzie $F(t)$ jest częstotliwością wejściową, f_0 jest częstotliwością rezonansową obwodu złożonego z elementów R , L , C , a Q jest dobrocią tego obwodu.

Przy założeniu, że argument funkcji \arctg jest dostatecznie mały,

$$\varphi \approx \frac{\pi}{2} - 2Q \frac{F(t) - f_0}{f_0}, \quad (4.30)$$



Rys. 4.11. Typowy przesuwnik fazowy wykorzystywany w iloczynowych demodulatorach FM.

wtedy na wyjściu filtra dolnoprzepustowego w układzie z rys. 4.10b

$$u_{wy}(t) \approx \cos\left(\frac{\pi}{2} - 2Q \frac{F(t) - f_o}{f_o}\right) = \sin\left(2Q \frac{F(t) - f_o}{f_o}\right) \approx 2Q \frac{F(t) - f_o}{f_o}. \quad (4.31)$$

Jeżeli $f_o = F_o$, tzn. jeśli częstotliwość rezonansowa obwodu z rys. 4.11 będzie równa częstotliwości nośnej sygnału wejściowego, napięcie wyjściowe będzie wynosić (na podstawie „definitywnego” wzoru (4.1))

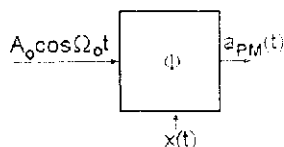
$$u_{wy}(t) \approx 2Q \frac{F(t) - F_o}{F_o} = 2Q \frac{x(t) \cdot k_f}{F_o 2\pi} = \frac{Q}{F_o} \frac{k_f}{\pi} x(t). \quad (4.32)$$

Z uwzględnieniem wprowadzonych przybliżeń napięcie wyjściowe jest proporcjonalne do przebiegu modulującego, zatem układ zachowuje się jak demodulator FM.

Demodulatory iloczynowe, w postaci odpowiednich układów scalonych, stosowane są w odbiornikach radiofonicznych FM, odbiornikach TV, tunerach satelitarnych itd.

4.8. Modulatory i demodulatory fazy

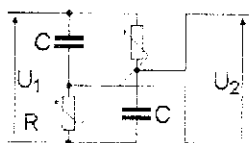
Modulator fazy w sposób funkcjonalny można przedstawić jak na rys. 4.12.



Rys. 4.12. Ogólna postać modulatora fazy.

Blok oznaczony „Φ” jest przesuwnikiem fazy sterowanym przez sygnał modulujący. Przesuwnikiem tym mogą być różne czwórniki RC lub LC, w których wartości niektórych elementów mogą być zmieniane napięciem steru-

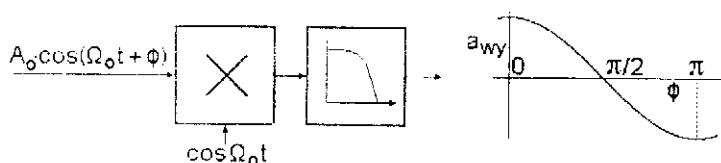
jącym (w sposób elektryczny). Przykład takiego prostego czwornika pokazano na rys. 4.13.



Rys. 4.13. Układ czwornika RC możliwego do wykorzystania jako modulator fazy.

W układzie tym, jak można łatwo sprawdzić, argument transmitancji jest równy $2 \arctg(\omega RC)$, natomiast jej moduł jest równy jedności. Sterując wartością rezystorów lub kondensatorów, można uzyskać odpowiednie zmiany fazy napięcia wyjściowego, niestety są one ograniczone do zakresu $(0, 180^\circ)$ i są nieliniowe, tak jak nieliniowy jest przebieg funkcji $\arctg(\omega RC)$. Lepsze jakościowo modulatory fazy buduje się z użyciem pętli synchronizacji fazowej (PLL) lub układów cyfrowych [7, 9].

Demodulację PM można uzyskać np. w wyniku przemnożenia sygnału PM przez odpowiednio odtworzoną falę nośną (co jednak może wymagać dość złożonych układów) i odfiltrowanie zbędnych produktów mnożenia (rys. 4.14). Uzyskiwana charakterystyka demodulatora jest nieliniowa; można ją zlinearyzować, jeżeli, zamiast sinusoidalnych, mnożeniu podlegać będą odpowiednie przebiegi prostokątne.



Rys. 4.14. Schemat blokowy prostego demodulatora fazy i jego charakterystyka przejściowa.

Podobnie jak w przypadku modulatorów, demodulatory PM o lepszych parametrach budowane są z użyciem pętli synchronizacji fazowej i (lub) układów cyfrowych [7, 9].

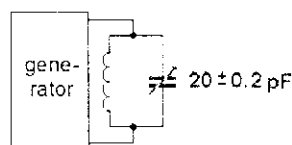
Zadania

1. W pewnym systemie FM częstotliwość nośna zmienia się w zakresie 110–120–130–120–110–120... kHz z szybkością 200 razy na sekundę. Ile wynosi częstotliwość środkowa, dewiacja częstotliwości i indeks modulacji?

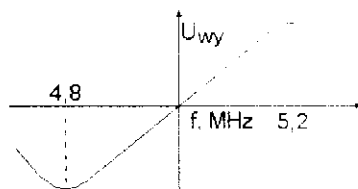
2. Częstotliwość sygnału modulującego wzrosła 2 razy. natomiast amplituda tego sygnału zmalała 2 razy. Co się stanie z dewiacją częstotliwości w wypadku modulatora FM i w wypadku modulatora PM?
3. Dane jest widmo sygnału FM zmodulowanego jednym tonem sinusoidalnym. Ile wynosi indeks modulacji? (Należy się posłużyć rys. 4.4).



4. Na podstawie danych zawartych w tab. 4.3 sprawdzić, że suma kwadratów funkcji Bessela jest stała, niezależna od argumentu.
5. Jak będzie się wyrażać pasmo sygnału FM, gdy $m_f = 1$?
6. Pewien modulator fazy zapewnia maksymalną dewiację fazy równą $\pi/4$. Jaka będzie maksymalna dewiacja częstotliwości sygnału wyjściowego tego modulatora, jeżeli $f_m = 1000$ Hz?
7. W układzie bezpośredniego modulatora FM pojemność kondensatora w obwodzie rezonansowym, określającym częstotliwość wyjściową, zmienia się o $\pm 0,2$ pF (rysunek). Jaka dewiację częstotliwości uzyska się w tym układzie, jeżeli $F_0 = 20$ MHz?



8. Dana jest charakterystyka demodulatora częstotliwości (tzw. krzywa S). Jaka maksymalną dewiację może mieć sygnał doprowadzony do tego demodulatora, aby demodulacja przebiegała bez zniekształceń?



5

Modulacje impulsowe

5.1. Twierdzenie o próbkowaniu

Rozważane wcześniej modulacje były modulacjami ciągłymi – sygnał zmodulowany był ciągły; w sposób ciągły zmieniał swoje wartości np. od minimalnej do maksymalnej. Sygnały ciągłe można także transmitować (zapisywać) w sposób przerywany, nieciągły, bez jakiegokolwiek straty informacji w nich zawartej.

Przykład 5.1

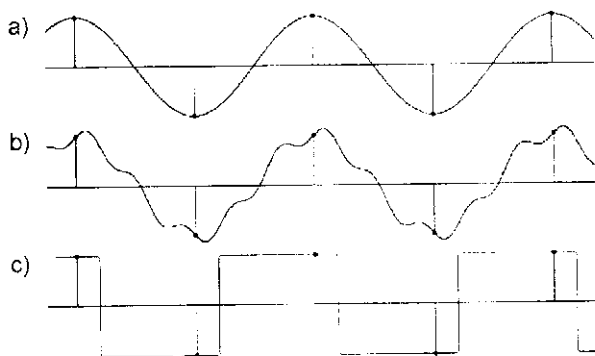
Należy sporządzić wykres zmienności temperatury otoczenia w ciągu dnia. Przygotowaliśmy odpowiednią skalę i przystępujemy do rejestracji wartości temperatury. Jak często należy odczytywać temperaturę i nanosić ją na wykres – co 1 sekundę, co 1 minutę? Nie, byłoby to bezsensowne, gdyż większość kolejnych odczytów byłaby taka sama. W rzeczywistości zapewne wystarczy temperaturę odczytywać co 10 lub nawet 30 minut, nanosić punkty na wykres, potem je połączyć linią ciągłą. Nikt rozsądny nie zakwestionuje takiego sposobu pomiaru i zapisu temperatury. W ten sposób (może podświadomy) stosujemy ważne twierdzenie o próbkowaniu – że wielkość ciągłą (temperatura) można odczytać w dyskretnych momentach czasu (jako próbki tej wielkości), a jej wartość w dowolnym momencie może zostać wyznaczona np. metodą interpolacji na podstawie sąsiednich wartości próbek. Jest to możliwe, gdyż w rozważanym przykładzie temperatura jest wielkością **wolnozmienną** w porównaniu ze stosowanymi okresami odczytu (próbkowania).

Podobne podejście można zastosować do dowolnych wielkości analogowych, nawet szybkozmiennych, byle przyjąć odpowiednie krótkie okresy próbkowania. Mówi o tym twierdzenie o próbkowaniu (Kotelnikowa-Shanona). Brzmi ono następująco:

Sygnał (przebieg) o pasmie ograniczonym do f_{\max} jest całkowicie określony przez swoje wartości (próbki) położone w równych odległościach czasu, mniejszych niż $1/2f_{\max}$.

W brzmieniu twierdzenie jest bardzo proste, ale może wywoływać pewne reakcje niedowierzania. Można je rozproszyć za pomocą następujących rozważań.

Na rysunku 5.1a pokazano prosty przebieg z zaznaczonymi wartościami próbek. Może się tu nasunąć wątpliwość – przecież takie same wartości próbek może mieć nieskończenie wiele innych przebiegów – np. z rys. 5.1b i c; dlaczego nie można ich odtworzyć? dlaczego odtworzony musi być tylko przebieg z rys. 5.1a? – dlatego, że pozostałe przebiegi nie spełniają warunku szerokości pasma. A skąd wiadomo, jaka ma być ta szerokość pasma? – z obserwacji częstotliwości próbkowania, gdyż ona określa, że maksymalna częstotliwość sygnału na pewno musi być mniejsza od połowy częstotliwości próbkowania. Przebiegi z rys. 5.1b i c mają pasmo szersze, dlatego nie należą do „klasy” przebiegów, które mogłyby być odtworzone na podstawie odebranych próbek.



Rys. 5.1. Przykładowe przebiegi z zaznaczonymi ich próbkami: a) przebieg sinusoidalny, b), c) inne przebiegi o takich samych wartościach próbek jak dla przebiegu z rys. a.

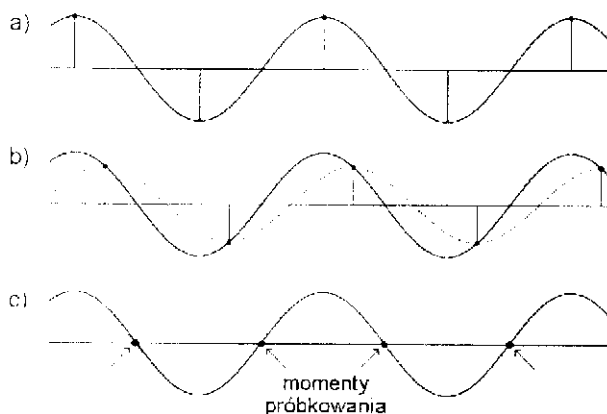
Gdyby stosować próbkowanie dokładnie co pół okresu największej częstotliwości, w szczególnym przypadku mogłoby wystąpić niekorzystne zjawisko – odtworzony sygnał miałby zmniejszoną amplitudę (rys. 5.2b) lub w szczególnym przypadku wręcz zerową (rys. 5.2c)! Aby tego uniknąć, konieczne jest, żeby okres próbkowania był mniejszy od $1/2f_{\max}$, czyli by częstotliwość próbkowania była większa od podwójnej maksymalnej częstotliwości występującej w sygnale. Można to zapisać następująco:

$$F_p > 2f_{\max} \quad (5.1)$$

Graniczna, najmniejsza wartość częstotliwości próbkowania (niestosowana praktycznie) nosi nazwę częstotliwości Nyquista

$$F_N = 2f_{\max}, \text{ a więc } F_p > F_N.$$

Z tego względu występująca w twierdzeniu o próbkowaniu nierówność jest „ostra”. Jak się okaże, „ostrość” tej nierówności wymagana jest także z innego powodu.



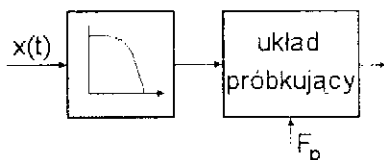
Rys. 5.2. Zjawiska możliwe przy próbkowaniu dokładnie co pół okresu częstotliwości sygnału: a) próbki przypadają w momentach ekstremów sygnału – próbkowanie „prawidłowe”, b) próbki przesunięte względem ekstremów – możliwość odtworzenia sygnału o mniejszej amplitudzie (pokazanego linią kropkowaną), c) próbki w zerach sygnału – odtworzenie sygnału jest niemożliwe.

W praktycznych systemach wykorzystujących próbkowanie sygnałów analogowych częstotliwość maksymalna sygnału i częstotliwość próbkowania są ściśle ustalone i stanowią ważne parametry techniczne. Parametry te, w odniesieniu do niektórych systemów, przedstawiono w tab. 5.1.

Tab. 5.1. Parametry niektórych systemów wykorzystujących próbkowanie sygnałów analogowych.

Rodzaj systemu	Częstotliwość maksymalna	Częstotliwość próbkowania
Telefonia	3,4 kHz	8 kHz
Compact Disc	20 kHz	44,1 kHz
Magnetofon cyfrowy	20 kHz	48 kHz
Radiofonia cyfrowa	15 kHz	32 kHz
Telewizja cyfrowa	6 MHz	13,5 MHz

Przy ustalonej częstotliwości próbkowania, stanowiącej istotny parametr systemu, należy zadbać, aby w sygnale wejściowym, podlegającym przetwarzaniu, nie znalazła się częstotliwość większa od dopuszczalnej. Dlatego zawsze w takich systemach przed układem próbkującym umieszcza się filtr dolnoprzepustowy, wytłumiający składowe, których częstotliwość mogłaby naruszać nierówność (5.1). Filtr taki nosi nazwę **filtru antyaliasingowego** (rys. 5.3).



Rys. 5.3. Miejsce filtra antyaliasingowego w systemie z próbkowaniem sygnału

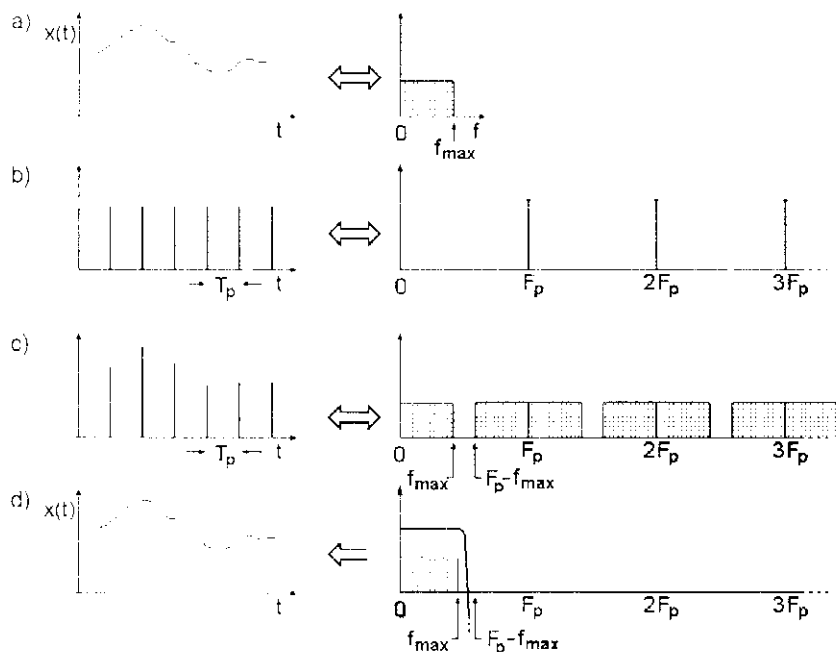
Dowód twierdzenia o próbkowaniu można znaleźć np. w pracy [5]. Tutaj pokazane będzie jego uzasadnienie na gruncie znanej równoważności opisów sygnałów w dziedzinie czasu i częstotliwości. Przebiegi czasowe mogą być w sposób kompletny opisane albo w dziedzinie czasu, albo w dziedzinie częstotliwości. Każdy z tych sposobów przedstawienia odpowiada drugiemu – jeżeli znany jest dokładny opis sygnału w dziedzinie częstotliwości, można na tej podstawie odtworzyć jego dokładną postać czasową, i odwrotnie.

Czasowemu przebiegowi rozważanego sygnału $x(t)$ (rys. 5.4a) odpowiada jego widmo pokazane (w uproszczeniu) po prawej stronie. Widmo to zawiera pełną informację o sygnale $x(t)$, można więc umieścić tu znak równoważności między przebiegiem czasowym sygnału a jego widmem.

Ciąg impulsów próbkujących to ciąg z rys. 5.4b. Są to bardzo wąskie impulsy (w granicy są to impulsy Diraca), powtarzane okresowo; widmo takiego ciągu jest periodyczne i nieskończone.

Próbkowanie sygnału polega na pomnożeniu go przez ten ciąg impulsów próbkujących (rys. 5.4c). W dziedzinie częstotliwości odpowiada temu splot widma sygnału i widma impulsów. W wyniku splotu otrzymuje się widmo pokazane po prawej stronie rys. 5.4c. Widmo to jest nieskończone i zawiera w sobie widmo sygnału oryginalnego oraz składowe położone wokół kolejnych harmonicznych częstotliwości próbkowania.

Usunięcie tych wyższych składowych za pomocą odpowiedniego filtra dolnoprzepustowego (rys. 5.4d) spowoduje, że pozostanie tylko widmo sygnału oryginalnego, które jak wiadomo, odpowiada dokładnie samemu sygnałowi w dziedzinie czasu. Zatem, usuwając z ciągu próbek częstotliwość większą niż f_{max} , można odtworzyć kompletny sygnał oryginalny (jedynie jego amplituda może zostać zmieniona).



Rys. 5.4. Uzasadnienie twierdzenia o próbkowaniu: a) sygnał oryginalny i jego widmo, b) ciąg impulsów próbkujących i ich widmo, c) ciąg próbek sygnału i jego widmo, d) widmo próbek po filtracji odpowiada sygnałowi oryginalnemu.

Z rysunku 5.4d wynika, że wspomniany filtr (tzw. filtr rekonstruujący) musi mieć odpowiednio dobraną charakterystykę selektywności – musi mianowicie przepuszczać częstotliwość f_{\max} , natomiast znacznie tłumić częstotliwość $F_p - f_{\max}$ i wyższą. Realizacja takiego filtra będzie fizycznie możliwa, gdy

$$f_{\max} \leq F_p - f_{\max}. \quad (5.2)$$

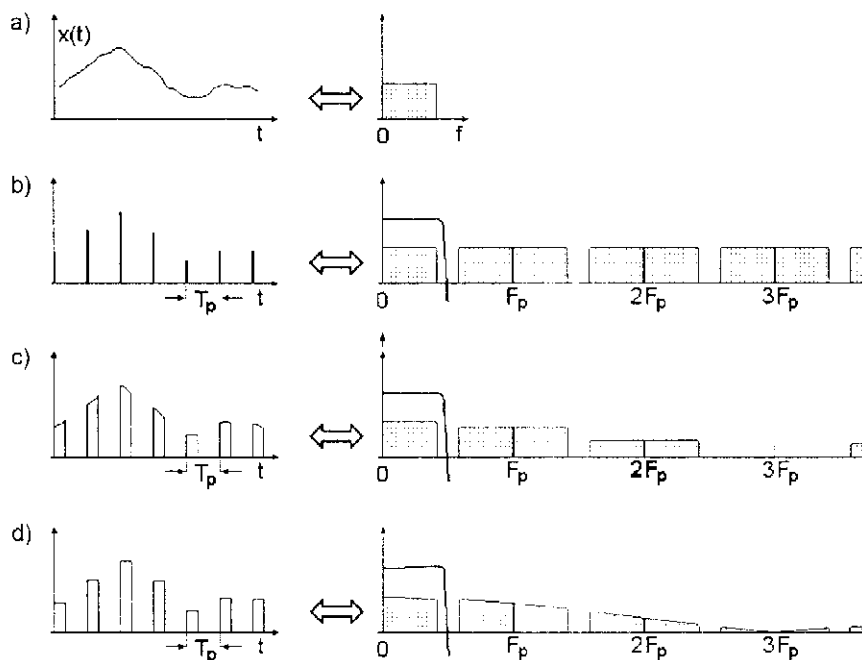
Teoretycznie, w zależności (5.2) mógłby wystąpić znak równości, ale wtedy filtr musiałby mieć nieskończenie wielką stromość charakterystyki przy przechodzeniu ze stanu przepuszczania do tłumienia, co jest niemożliwe do uzyskania. Dlatego $f_{\max} < F_p - f_{\max}$, czyli

$$F_p > 2f_{\max}.$$

Jest to ponownie warunek (5.1), tym razem otrzymany na podstawie innych przesłanek fizycznych. Warunek ten jest konieczny, aby nie następowało nakładanie się (aliasing) widma sygnału oryginalnego i widma położonego wokół pierwszej harmonicznej częstotliwości próbkowania.

5.2. Sposoby próbkowania sygnału

Przytoczone rozważania dotyczyły **próbkowania idealnego**, tj. z użyciem nieskończenie wąskich impulsów próbkujących. W rzeczywistości impulsy te mają skończoną szerokość. Ciąg takich impulsów nie ma widma równomiernego, przez co widmo sygnału po próbkowaniu ulega zmianie w porównaniu z próbkowaniem idealnym. Na rysunku 5.5 zestawiono – w celu porównania – możliwe sposoby próbkowania sygnału i odpowiadające im widma. Na rysunku 5.5b pokazano, omówione wcześniej, próbkowanie idealne, na rys. 5.5c – tzw. **próbkowanie naturalne** – impulsami skończonej szerokości. Istotne jest tu to, że po próbkowaniu wierzchołki tych impulsów zachowują przebieg sygnału oryginalnego. Na rysunku 5.5d przedstawiono często stosowane **próbkowanie idealne z pamięcią**, kiedy wysokość impulsów jest równa wartościom sygnału w momentach próbkowania, potem przez jakiś czas wysokość tych impulsów pozostaje stała.



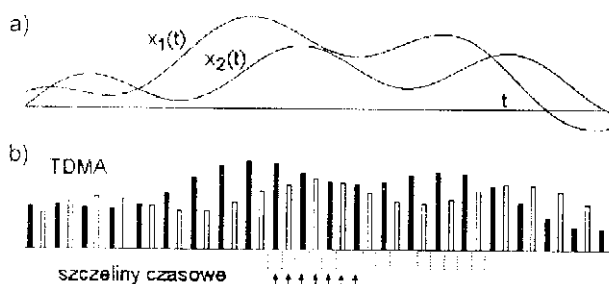
Rys. 5.5. Różne sposoby próbkowania sygnału i uzyskiwane widma: a) sygnał oryginalny i jego widmo, b) próbkowanie idealne, c) próbkowanie naturalne, d) próbkowanie idealne z pamięcią.

Ostatni przypadek jest interesujący, ponieważ otrzymane widmo sygnału po próbkowaniu jest zmienione w ten sposób, że nawet najniższa część tego widma jest zdeformowana w stosunku do widma sygnału oryginalnego. Zatem

takie próbkowanie formalnie nie zachowuje informacji o sygnale oryginalnym! Rada na to jest jednak prosta – do wydzielenia sygnału wystarczy użyć filtru dolnoprzepustowego o odpowiednio ukształtowanej charakterystyce w zakresie przepuszczania (nieco „podbijającego” wyższą częstotliwość) i widmo sygnału oryginalnego zostanie odtworzone prawidłowo, czyli odtworzony zostanie i sam sygnał. Charakterystykę takiego filtru zaznaczono schematycznie na rysunku.

Przedstawiony problem występuje wyraźnie, gdy impulsy próbkujące są dość szerokie, np. rzędu 30–40% długości ich okresu. W praktyce, przy znacznie węższych impulsach, występująca deformacja widma jest nieznaczna (np. poniżej 1%) i może być zaniedbana.

Informacja o wartościach próbek jest transmitowana w sposób przerywany. Przerwy między okresami transmisji mogą być „oddane” do dyspozycji innych użytkowników, czyli można uzyskać zwielokrotnienie jednego kanału telekomunikacyjnego. W ten sposób z jednego kanału (jednego przewodu, światłowodu, fali nośnej...) może (prawie) jednocześnie korzystać np. kilku czy nawet kilkudziesięciu użytkowników. Jest to zwielokrotnienie w dziedzinie czasu, oznaczane TDM lub TDMA – była o tym mowa w rozdz. 2.



Rys. 5.6. Przykład dwukrotnego zwielokrotnienia w dziedzinie czasu: a) różne sygnały oryginalne $x_1(t)$ i $x_2(t)$, b) ciąg impulsów, reprezentujących naprzemiennie próbki obydwu sygnałów.

Najprostsze takie zwielokrotnienie pokazano na rys. 5.6. Dwa różne sygnały $x_1(t)$ i $x_2(t)$ są próbkowane, a informacja o wartościach ich próbek jest przesyłana za pomocą impulsów o zmiennej amplitudzie w występujących na przemian, odpowiadających im tzw. **szczelinach czasowych**, tj. w przedziałach czasu przeznaczonych do transmisji informacji pochodzącej z jednego źródła. Na rysunku impulsy są rozróżnione za pomocą cieniowania, w rzeczywistości mogą być rozróżnione tylko wg szczeliny czasowej, w której występują, co powoduje konieczność utrzymywania synchronizacji między nadajnikiem a odbiornikiem tak przesyłanych informacji.

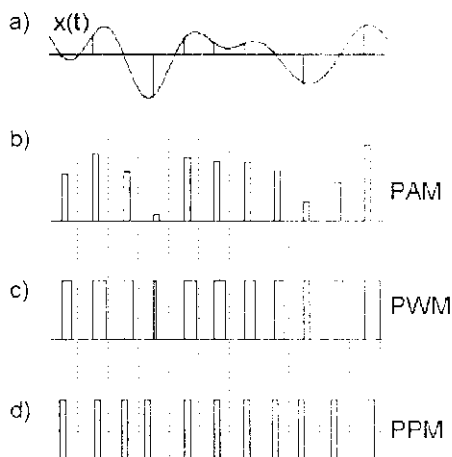
5.3. Analogowe modulacje impulsowe

Wartości próbek są wielkościami analogowymi – informacja o nich może być przesyłana w sposób analogowy, ale w dyskretnych momentach lub przedziałach czasu. Do tego celu są stosowane:

- modulacja amplitudy impulsów – **PAM** (*Pulse Amplitude Modulation*),
- modulacja położenia impulsów – **PPM** (*Pulse Position Modulation*),
- modulacja szerokości impulsów – **PWM** (*Pulse Width Modulation*).

Dwie pierwsze stanowią odpowiedniki ciągłych modulacji AM i PM, natomiast trzecia nie ma takiego ścisłego odpowiednika. Należy pamiętać, że ciągłe modulacje FM i PM są znacznie podobne do siebie, dlatego można powiedzieć, że PPM jest też jakimś odpowiednikiem FM.

W dziedzinie modulacji impulsowych nie stosuje się typowej modulacji częstotliwości (w tym przypadku – częstotliwości powtarzania impulsów), gdyż wtedy niemożliwe byłoby czasowe zwielokrotnianie kanału łączności.



Rys. 5.7. Analogowe modulacje impulsowe: a) sygnał oryginalny – modulujący, b) przebieg z modulacją amplitudy impulsów, c) przebieg z modulacją szerokości impulsów, d) przebieg z modulacją położenia impulsów.

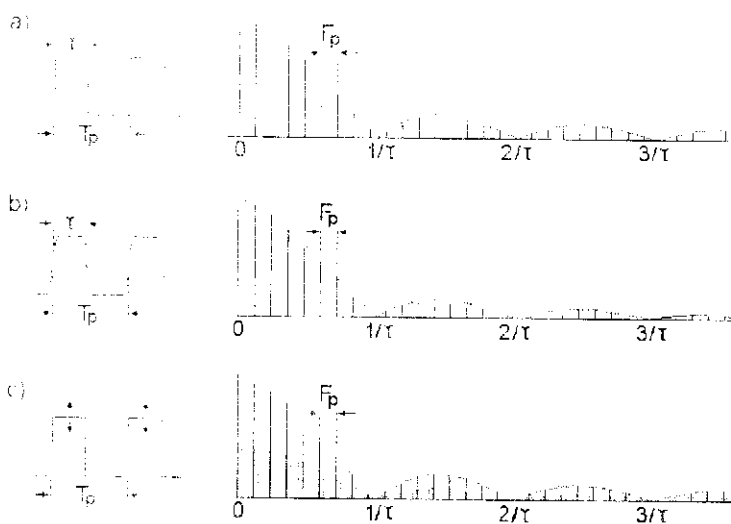
Modulacja PAM (rys. 5.7b) stosowana jest jedynie w bardzo prostych systemach łączności, gdyż jest mało odporna na zakłócenia, natomiast jest powszechnie stosowana jako pośredni etap przetwarzania sygnału w koderach PCM (rozdz. 6.). Sygnał na wyjściu każdego układu próbkującego jest praktycznie sygnałem PAM.

Modulacje PWM i PPM cechują się większą odpornością na zakłócenia, szczególnie modulacja PPM, i dlatego są stosowane we względnie prostych

systemach łączności na niewielkie odległości, np. do transmisji danych pomiarowych, spotykane są też w pewnych systemach światłowodowych. Nie są natomiast stosowane w rozległych, rozsiewczych systemach telekomunikacyjnych, gdzie znalazła zastosowanie m.in. modulacja impulsowo-kodowa PCM, omówiona w następnym rozdziale.

5.4. Pasmo sygnałów PAM, PWM i PPM

Do przesłania sygnałów PAM, PWM czy PPM potrzebne jest określone pasmo częstotliwości. Jak je określić? Widmo ciągu niemodulowanych impulsów prostokątnych szerokości τ , o okresie powtarzania T_p (impulsowej fali nośnej) jest przedstawione na rys. 5.8a.



Rys. 5.8. Widmo ciągu impulsów prostokątnych: a) impulsów idealnych, b) impulsów ze skończonymi wartościami czasu narastania i opadania, c) impulsów o modulowanej amplitudzie.

Widmo to dotyczy idealnych impulsów prostokątnych. W rzeczywistości impulsy te mają zawsze skończony czas narastania i opadania, co objawia się tym, że ich widmo szybciej zanika ze wzrostem częstotliwości (rys. 5.8b). W razie modulacji amplitudy tych impulsów ich widmo ulega „wypełnieniu” – pojawiają się wstęgi boczne przy każdej harmonicznej zawartej w ciągu impulsów niemodulowanych (rys. 5.8c). Podczas modulacji PWM widmo jest podobne; w przypadku PPM jest bardziej skomplikowane, ale nie wykracza poza obwiednię pokazaną na rysunkach.

W uproszczeniu przyjmuje się, że widmo ciągu impulsów prostokątnych kończy się przy częstotliwości $1/\tau$ (pierwsze miejsce zerowe widma) i jest to szerokość pasma niezbędna do przeniesienia sygnałów za pomocą modulacji PAM, PWM lub PPM.

Dla przypadku z rys. 5.8a w pasmie szerokości $1/\tau$ zawarte jest ok. 92% energii sygnału, a więc jej znacząca część, dla przypadku impulsów rzeczywistych (rys. 5.8b) – jeszcze więcej.

Uwaga

- ☛ W razie modulacji PWM, gdy szerokość impulsów jest zmienna, określając szerokość pasma, należy brać pod uwagę impulsy najwęższe.

W celu ilustracji zagadnień związanych z pasmem modulacji impulsowych można rozważyć następujący przykład.

Przykład 5.2

Sygnał telefoniczny jest próbkowany z częstotliwością 8 kHz, czyli co 125 μ s. Jaka jest szerokość pasma potrzebna do transmisji informacji o wartościach tych próbek za pomocą modulacji PAM?

Jeżeli nie jest potrzebne zwielokrotnienie czasowe kanału, informacje o wartościach próbek tego sygnału można przesłać za pomocą dość szerokich impulsów PAM, np. $\tau = 50 \mu$ s. Wtedy wymagana szerokość pasma tego sygnału będzie równa ok. 20 kHz ($1/50 \mu$ s). Jeżeli jednak chcemy ten kanał łączności zwielokrotnić w systemie TDMA, np. 10-krotnie, należy zastosować impulsy 10 razy węższe (co najmniej 10 razy, gdyż należy jeszcze uwzględnić pewien odstęp „bezpieczeństwa” między impulsami); wtedy pasmo każdego z 10 sygnałów będzie równe co najmniej 200 kHz. Tak więc zwielokrotnienie w dziedzinie czasu wiąże się z koniecznością poszerzenia pasma częstotliwości.

Uwaga

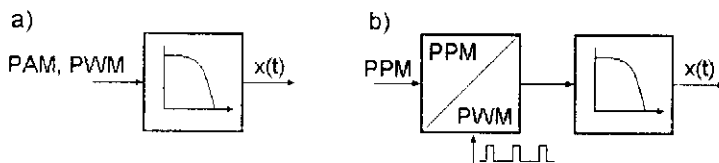
- ☛ Każdy z 10 sygnałów zajmuje pasmo 200 kHz i mimo to pasmo „całkowite”, potrzebne do ich jednoczesnej transmisji, pozostaje takie samo. Wszystkie sygnały wykorzystują to samo pasmo częstotliwości, ale ściśle biorąc, nie jednocześnie, lecz w przydzielonych im szczelinach czasowych. Dlatego sygnały te mogą występować (prawie) jednocześnie, wzajemnie się nie zakłócając.

Realizacja układowa modulatorów PAM, PPM i PWM odbywa się za pomocą typowych układów impulsowych (zawierających układy opóźnień czasowych, przerzutniki monostabilne, generatory przebiegów piłowych i prostokątnych itd.).

Demodulacja sygnałów PAM i PWM jest całkiem prosta – wystarczy do tego celu filtr dolnoprzepustowy (rekonstruuujący) o częstotliwości granicznej

równej $f_{m \max}$ lub większej niż $f_{m \max}$, natomiast dużo mniejszej niż częstotliwość powtarzania impulsów F_p .

Demodulacja sygnałów PPM jest trudniejsza, gdyż wymaga dysponowania przebiegiem impulsowym o fazie odniesienia. Pozwala to na zamianę sygnału PPM na PWM lub PAM, a następnie jego filtrację (rys. 5.9).

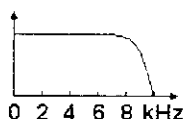


Rys. 5.9. Przykłady demodulacji sygnałów zmodulowanych impulsowo:

a) PAM i PWM, b) PPM.

Zadania

1. Na podstawie rys. 5.4 narysować widmo przebiegu na wyjściu układu próbkującego, do którego doprowadzono przemienny sygnał sinusoidalny o $f = 10$ kHz. Częstotliwość próbkowania jest równa 48 kHz.
2. Narysować wymaganą charakterystykę filtra DP w systemie z próbkowaniem, jeżeli $f_{m \max} = 3.8$ kHz i $F_p = 8$ kHz.
3. W systemie PAM stosowane są impulsy szerokości $8 \mu s$, o częstotliwości powtarzania 40 kHz. Oszacować zajmowaną szerokość pasma.
4. Pewien sygnał wyraża się zależnością $a(t) = 2 \cos \omega t + 0,8 \sin 2\omega t - 0,2 \cos 3\omega t$, gdzie $\omega = 2\pi \cdot 2000/s$. Jaka musi być najmniejsza częstotliwość próbkowania sygnału, aby nie wystąpiły zniekształcenia z powodu nakładania się widm?
5. W odniesieniu do sygnału z zadania 4, założyć, że filtr na wyjściu systemu (rekonstruujący) ma charakterystykę amplitudową pokazaną na rysunku i wyznaczyć w tym przypadku wymaganą częstotliwość próbkowania.





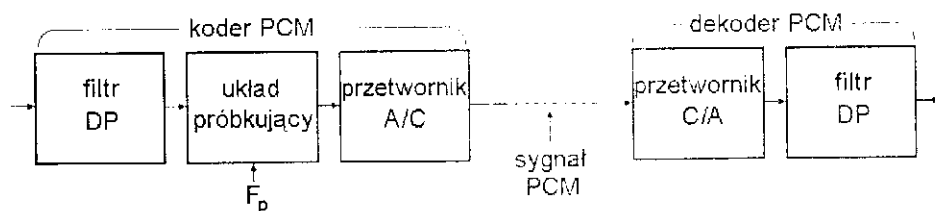
Modulacja PCM

6.1. Wytwarzanie sygnału PCM

Modulacja kodowo-impulsowa (PCM - *Pulse Code Modulation*) jest bardzo ważnym rodzajem modulacji, stosowanym powszechnie w telefonii, telewizji, w zapisie sygnałów multimedialnych itd. Użycie terminu modulacja nie jest tu adekwatne, gdyż nie występuje tutaj typowe nakładanie informacji na fale nośną (przynajmniej podczas transmisji w paśmie naturalnym), ale nazwa PCM przyjęta jest w skali światowej.

Modulacja PCM to modulacja impulsowa nieanalogowa, tzn. parametry fizyczne sygnału, takie jak szerokość impulsów czy amplituda, nie są odpowiednikami (analogami) wartości próbek sygnału oryginalnego.

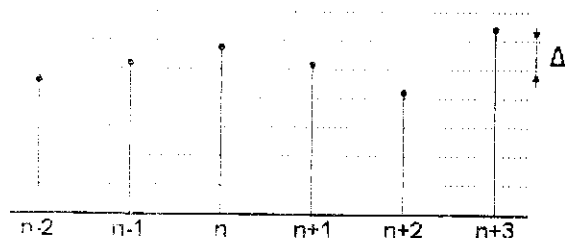
Analogowe wartości próbek nie wpływają na amplitudę czy szerokość impulsów, natomiast zostają zamienione na postać cyfrową i dopiero jako wartości cyfrowe w jakiś sposób są transmitowane. Schemat wytwarzania sygnału PCM oraz jego odbioru pokazano na rys. 6.1.



Rys. 6.1. Sposób wytwarzania sygnału PCM oraz jego odbioru.

Wytwarzanie sygnału PCM odbywa się w koderze, który zawiera elementy typowe dla modulacji impulsowej, a więc filtr antyaliasingowy i układ próbkujący oraz przetwornik A/C, który decyduje o tym, że wyjściowy sygnał jest sygnałem PCM (najczęściej z szeregowym układem bitów).

Zamiana analogowych wartości próbek na wartości cyfrowe odbywa się w przetworniku A/C, który cechuje się skończoną rozdzielczością, tj. zdolnością rozróżniania różnych wartości próbek i przedstawiania ich w postaci różnych wartości cyfrowych. Zilustrowano to na rys. 6.2.



Rys. 6.2. Ilustracja błędów popełnianych podczas kwantyzacji sygnału w przetworniku A/C o szerokości przedziału kwantyzacji równej Δ .

Wartość każdej próbki jest przetwarzana ze skończoną dokładnością, gdyż jest zaokrąglana do najbliższej wartości akceptowanej przez przetwornik A/C, wynikającej z szerokości przedziału kwantyzacji, oznaczonego na rys. 6.2 przez Δ . W tym momencie bezpowrotnie traci się część informacji o sygnale oryginalnym. Na tym rysunku próbki $n-1$, n i $n+1$, pomimo że mają różne wartości, zostaną tak samo zakodowane, gdyż mieszczą się w tym samym przedziale kwantyzacji.

Kwantyzacja jest procesem nieodwracalnym.

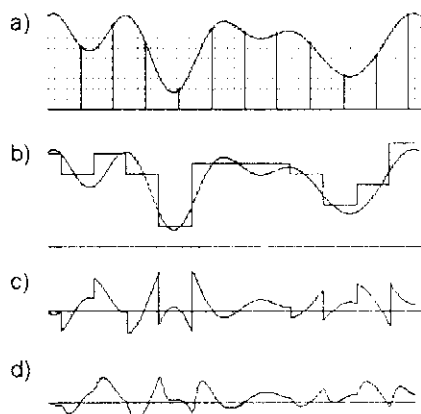
6.2. Szum kwantyzacji

Różnica między sygnałem otrzymanym na wyjściu przetwornika C/A a sygnałem oryginalnym nosi nazwę **szumu kwantyzacji** – jest on nieuchronnym skutkiem procesu kwantyzacji (rys. 6.3).

Przebieg szumu kwantyzacji pokazany na rys. 6.3c cechuje się znacznymi wartościami chwilowymi. W praktyce sytuacja jest lepsza, bo po stronie odbiorczej odtworzony sygnał analogowy przechodzi jeszcze przez filtr dolnoprzepustowy (nazywany filtrem rekonstruuującym), który w pewnym stopniu wygładza „wyskoki” napięcia szumu kwantyzacji (rys. 6.3d).

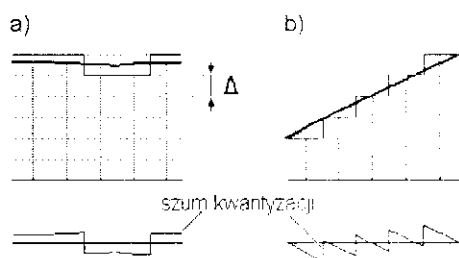
Można przyjąć, że zdekodowany sygnał wyjściowy (po przetworniku C/A) to nic innego jak sygnał oryginalny, zniekształcony przez dodanie szumu kwantyzacji – z tego wynika jasno, że należy dążyć do minimalizacji tego szumu.

Od czego zależy poziom szumu kwantyzacji? – od liczby możliwych poziomów wartości akceptowanych przez przetwornik, czyli od jego rozdzielczości, a ta zależy od liczby wykorzystywanych bitów. W celu otrzymania analitycznej zależności na poziom szumu kwantyzacji można przeprowadzić następujące rozważanie.



Rys. 6.3. Ilustracja powstawania szumu kwantyzacji: a) sygnał oryginalny z zaznaczonymi momentami próbkowania i przedziałami kwantyzacji, b) sygnał skwantowany, c) szum kwantyzacji, d) szum kwantyzacji po filtrze rekonstruującym.

Na rysunku 6.4 pokazano dwa przypadki kwantyzacji sygnału (pokazanego linią pogrubioną).



Rys. 6.4. Różna postać szumu kwantyzacji: a) przy przebiegu wolnozmiennym, b) przy przebiegu zmiennym liniowo.

Jak widać na rys. 6.4, szum kwantyzacji może mieć różny przebieg – gdy sygnał jest wolnozmienny, szum kwantyzacji może być zbliżony do przebiegu prostokątnego; gdy sygnał zmienia się liniowo, szum będzie zbliżony do przebiegu piłokształtnego. Najmniejsza amplituda przebiegu tego szumu (przypadek najkorzystniejszy dla użytkownika) może być równa $\Delta/2$, gdzie Δ jest szerokością przedziału kwantyzacji przetwornika. Dalej zostanie dokładniej rozważony przypadek z rys. 6.4a.

Moc tego szumu kwantyzacji $N = \Delta^2/4$ (podobnie jak wcześniej przyjmuje się jednostkową rezystancję), natomiast chwilowa moc sygnału to kwadrat jego

wartości, a więc $S = A^2$. Stosunek mocy sygnału do mocy szumu kwantyzacji wynosi więc

$$\frac{S}{N} = 4A^2 / \Delta^2. \quad (6.1)$$

Jak duży jest przedział kwantyzacji Δ ? Jeżeli założyć, że przetwarzany sygnał może być dwubiegunowy ($+A/-A$), a jego bieżąca wartość jest równa wartości maksymalnej w danym systemie (najlepszy przypadek), to szerokość przedziału kwantyzacji będzie $\Delta = 2A / L$, gdzie L jest liczbą poziomów kwantyzacji w całym zakresie przetwarzania przetwornika.

Zatem w tym przypadku stosunek mocy sygnału do mocy szumu kwantyzacji wynosi

$$\frac{S}{N} = \frac{1}{4} \cdot L^2. \quad (6.2)$$

W n -bitowym przetworniku A/C , pracującym w systemie dwójkowym, liczba rozróżnialnych poziomów sygnału $L = 2^n$. Zatem

$$\frac{S}{N} = \frac{1}{4} \cdot 2^{2n}. \quad (6.3)$$

Stosunek S/N wyrażany jest bardzo często w decybelach, wtedy

$$\frac{S}{N}, \text{ dB} = 10 \log \left(\frac{1}{4} \cdot 2^{2n} \right) = -6 + 6 \cdot n. \quad (6.4)$$

Podobnie można pokazać, że dla przypadku pokazanego na rys. 6.4.b uzyska się zależność [11]:

$$\frac{S}{N}, \text{ dB} = 4,3 + 6 \cdot n, \quad (6.5)$$

natomiast dla sygnału sinusoidalnego z uwzględnieniem jego mocy czynnej otrzyma się związek [5]:

$$\frac{S}{N}, \text{ dB} = 1,8 + 6 \cdot n. \quad (6.6)$$

W tych rozważaniach, dla ich prostoty, uwzględniono szczególne postaci i amplitudę sygnału. Zwracając uwagę na podobieństwo zależności (6.4)–(6.6) i nie siląc się na dużą dokładność, można w uproszczeniu przyjąć, że

$$\frac{S}{N}, \text{ dB} \approx 6 \cdot n, \quad (6.7)$$

niezależnie od postaci sygnału, jedynie przy założeniu, że ma on maksymalną amplitudę.

Ostatnie zastrzeżenie jest ważne, gdyż każde zmniejszenie amplitudy sygnału powoduje zmniejszenie stosunku S/N (widać to z zależności (6.1)). Przykładowo, przy sygnale o amplitudzie równej 10% amplitudy maksymalnej, moc sygnału zmaleje 100 razy, a szum kwantyzacji się nie zmieni, zatem S/N zmaleje również 100 razy, tj. o 20 dB!

Im więcej bitów ma przetwornik A/C, tym bardziej korzystny stosunek S/N otrzymuje się na jego wyjściu. Niestety, wzrost liczby bitów przetwornika to większa komplikacja systemu i przede wszystkim zwiększenie szerokości pasma potrzebnego do transmisji sygnału PCM, o czym będzie mowa dalej.

6.3. Modulacja delta

Jest to odmiana modulacji PCM, znamienna tym, że – w uproszczeniu – transmitowana jest informacja o przyrostach wartości sygnału, a nie o jego wartościach bezwzględnych. Daje to możliwość uproszczenia kodera (nie jest potrzebny typowy przetwornik A/C) i zmniejszenia liczby bitów potrzebnych do transmisji każdej próbki. Reguła modulacji delta jest następująca:

- sygnał dzielony jest na odpowiednie poziomy kwantyzacji,
- sygnał jest tak kwantowany, aby w momentach próbkowania zmieniał się o jeden i tylko jeden poziom kwantyzacji (w górę lub w dół – nie może pozostać niezmieniony),
- zmiana poziomu kwantyzacji w górę kodowana jest przez 1, zmiana w dół – przez 0.

Do zakodowania informacji o przyroście lub zmniejszeniu się sygnału używany jest więc tylko jeden bit, co powoduje znaczną prostotę tego sposobu kodowania. Niestety, konieczne jest tu stosowanie dużej częstotliwości próbkowania sygnału i dokonywania kwantyzacji – dużo większej niż wynikającej z twierdzenia o próbkowaniu (rys. 6.5). Wynika to z tego, że przedział kwantyzacji powinien być mały, aby dokładnie odwzorowywać przebieg wolnozmiennych partii sygnału, ale te małe przedziały kwantyzacji muszą następować dość szybko po sobie, aby móc śledzić szybko i duże zmiany wartości sygnału. W przeciwnym razie dojdzie do dużych różnic między sygnałem oryginalnym a

W tych rozważaniach, dla ich prostoty, uwzględniono szczególne postaci i amplitudę sygnału. Zwracając uwagę na podobieństwo zależności (6.4)–(6.6) i nie siląc się na dużą dokładność, można w uproszczeniu przyjąć, że

$$\frac{S}{N}, \text{ dB} \approx 6 \cdot n, \quad (6.7)$$

niezależnie od postaci sygnału, jedynie przy założeniu, że ma on maksymalną amplitudę.

Ostatnie zastrzeżenie jest ważne, gdyż każde zmniejszenie amplitudy sygnału powoduje zmniejszenie stosunku S/N (widac to z zależności (6.1)). Przykładowo, przy sygnale o amplitudzie równej 10% amplitudy maksymalnej, moc sygnału zmaleje 100 razy, a szum kwantyzacji się nie zmienia, zatem S/N zmaleje również 100 razy, tj. o 20 dB!

Im więcej bitów ma przetwornik A/C, tym bardziej korzystny stosunek S/N otrzymuje się na jego wyjściu. Niestety, wzrost liczby bitów przetwornika to większa komplikacja systemu i przede wszystkim zwiększenie szerokości pasma potrzebnego do transmisji sygnału PCM, o czym będzie mowa dalej.

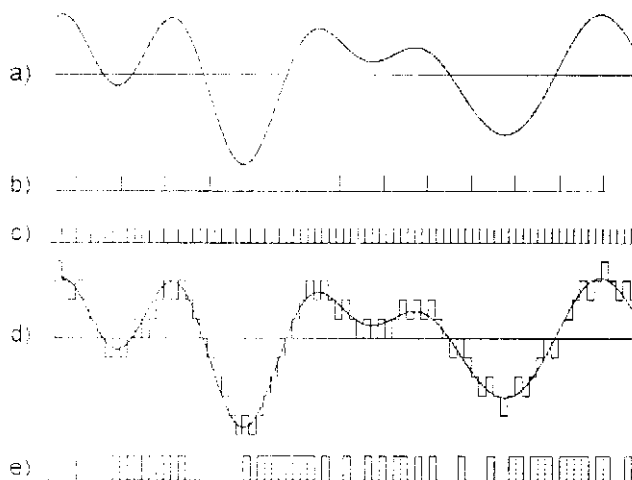
6.3. Modulacja delta

Jest to odmiana modulacji PCM, zmienna tym, że – w uproszczeniu – transmitowana jest informacja o przyrostach wartości sygnału, a nie o jego wartościach bezwzględnych. Daje to możliwość uproszczenia kodera (nie jest potrzebny typowy przetwornik A/C) i zmniejszenia liczby bitów potrzebnych do transmisji każdej próbki. Reguła modulacji delta jest następująca:

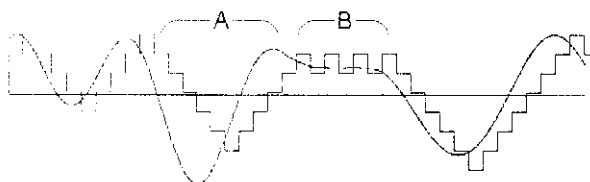
- sygnał dzielony jest na odpowiednie poziomy kwantyzacji
- sygnał jest tak kwantowany, aby w momentach próbkowania zmieniał się o jeden i tylko jeden poziom kwantyzacji (w górę lub w dół – nie może pozostać niezmieniony).
- zmiana poziomu kwantyzacji w górę kodowana jest przez 1, zmiana w dół – przez 0.

Do zakodowania informacji o przyroście lub zmniejszeniu się sygnału używany jest więc tylko jeden bit, co powoduje znaczną prostotę tego sposobu kodowania. Niestety, konieczne jest tu stosowanie dużej częstotliwości próbkowania sygnału i dokonywania kwantyzacji – dużo większej niż wynikającej z twierdzenia o próbkowaniu (rys. 6.5). Wynika to z tego, że przedział kwantyzacji powinien być mały, aby dokładnie odwzorowywać przebieg wolnozmienianych partii sygnału, ale te małe przedziały kwantyzacji muszą następować dość szybko po sobie, aby móc śledzić szybko i duże zmiany wartości sygnału. W przeciwnym razie dojdzie do dużych różnic między sygnałem oryginalnym

nałym a skwantowanym, co da tzw. przeciążenie zbocza lub duży tzw. szum śrutowy (rys. 6.6).



Rys. 6.5. Przebieg sygnału przy modulacji delta: a) analogowy przebieg wejściowy, b) rozmieszczenie momentów próbkowania wynikające z twierdzenia o próbkowaniu, c) rozmieszczenie momentów próbkowania w modulacji delta, d) skwantowany przebieg w koderze delta, e) sygnał wyjściowy kodera delta.

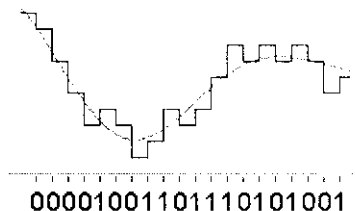


Rys. 6.6. Możliwe błędy przy kodowaniu przyrostowym jak przy modulacji delta: A – obszar występowania przeciążenia zbocza, B – obszar występowania dużego szumu śrutowego.

Konieczność stosowania dużej częstotliwości próbkowania przy tylko jednym bicie „na próbkę” powoduje, że zysk w zmniejszeniu przepływności sygnału w modulacji delta nie jest znaczny lub nie występuje. Spotykane są np. systemy telefoniczne z modulacją delta, w których przepływność wynosi 32 kb/s, podczas gdy w klasycznych telefonicznych systemach PCM wynosi ona zawsze 64 kb/s.

Poprawę może dać zastosowanie tzw. adaptacyjnej modulacji delta. Polega ona na tym, że stosownie do potrzeb wartość przedziału kwantyzacji zostaje zmniejszona lub zwiększona. Gdy zmienność sygnału jest znaczna – przedział

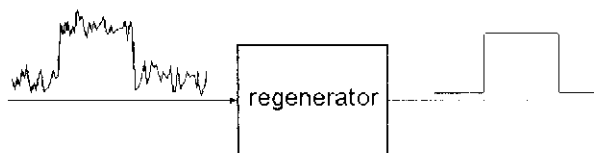
ulega zwiększeniu, np. dwukrotnemu; gdy sygnał wykazuje niewielkie zmiany, przedział ulega dwukrotnej redukcji w stosunku do standardowego (rys. 6.7). Problem w tym, aby o zmianie przedziału kwantyzacji, czynionej w koderze, powiadomić jakoś odbiorcę (tj. dekodery). Wbrew pozorom jest to dość proste – gdy w ciągu zer i jedynek pojawiają się kolejno dwie jedynki (lub dwa zera) – oznacza to konieczność zwiększenia przedziału kwantyzacji, gdy w ciągu występują na przemian 0 i 1 – ulega on zmniejszeniu.



Rys. 6.7. Przykład kwantyzacji przebiegu analogowego w adaptacyjnej modulacji delta.

6.4. Zabezpieczanie transmisji PCM przed zakłóceniami

Sygnały PCM są ze swej natury dość odporne na zakłócenia; wynika z odmiennej postaci sygnału PCM w stosunku do zakłóceń, które mają naturę analogową. Dlatego można dość łatwo przeprowadzać regenerację sygnału PCM bez utraty informacji w nim zawartej (rys. 6.8). Warto zwrócić uwagę na to, że regeneracja sygnału może być przeprowadzana wielokrotnie (nie ma tu ograniczeń fizycznych), co pozwala na osiągnięcie bardzo dużych odległości transmisji.

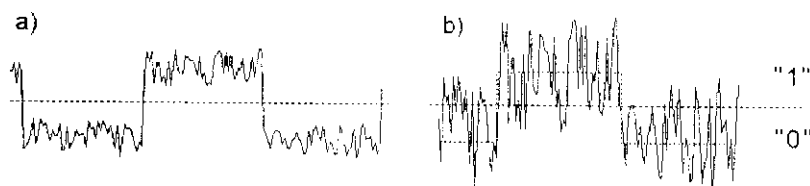


Rys. 6.8. Istota regeneracji sygnału cyfrowego PCM.

Podobna regeneracja zakłóconych sygnałów analogowych jest niemożliwa, gdyż staje się wtedy przed problemem rozróżnienia, która część odebranego przebiegu jest właściwym sygnałem, a która zakłóceniem, które można by usunąć. Usunięcie któregoś z fragmentów odebranego przebiegu (bo przypomina on zakłócenie...) może być w rzeczywistości usunięciem fragmentu sygnału – należy pamiętać, że transmisja informacji jest zawsze związana z niepewnością, co zostało nadane i jaka za chwilę będzie postać sygnału.

Oprócz możliwości regeneracji sygnały PCM mają jeszcze cenną właściwość – mogą być zabezpieczane przed błędami przez pewne algorytmy kodowania i obróbkę sygnału odbieranego.

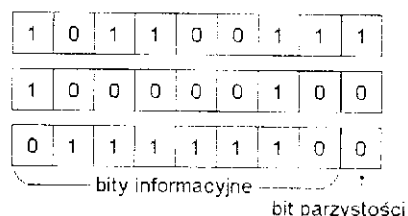
Dopóki szum jest słaby (rys. 6.9a), sygnał PCM jest łatwo odróżnialny od szumu i nie ma trudności z jego dekodowaniem. Gdy szum staje się silny i zaczyna maskować sygnał, jego odtworzenie nie jest takie oczywiste (rys. 6.9b). Jednak wtedy, w wyniku analizy pewnych związków logicznych między odebranymi bitami, można wyceliminować, przynajmniej częściowo, błędy w odbiorze przesyłanej informacji.



Rys. 6.9. Przykład zakłóconego sygnału PCM: a) w niewielkim stopniu, b) w znacznym stopniu, prowadzącym do błędów w odczycie bitów.

Dalej zostaną przedstawione 4 metody zabezpieczania sygnałów PCM przed zakłóceniami.

Bit parzystości jest to jeden bit dodawany na końcu słowa informacyjnego (rys. 6.10); jest on tak dobierany, aby wraz z nim liczba jedynek w sygnale nadawanym była parzysta (lub nieparzysta – jest to do wcześniejszego ustalenia; podobnie może to dotyczyć nie jedynek, lecz liczby zer w sygnale).



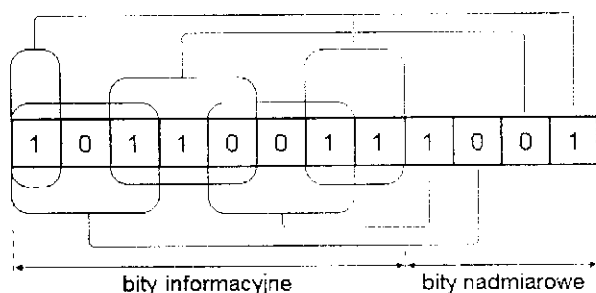
Rys. 6.10. Przykłady bitów parzystości dodanych do bitów informacyjnych.

Sposób ten wydłuża nieco transmitowane słowo albo zmusza do stosowania krótszych bitów, aby w tym samym czasie zmieściło się ich nieco więcej (krótsze bity to szersze pasmo częstotliwości!). System bitu parzystości pozwala na wykrycie, po stronie odbiorczej, że odebrane słowo jest błędne (nastąpiło jego zniekształcenie), ale niestety nie pozwala na określenie miejsca tego błędu. W tej sytuacji należy słowo odrzucić lub zażądać jego powtórzenia przez

nadawcę. Warto zauważyć, że jeżeli w słowie transmitowanym wystąpią dwa błędy jednocześnie, system tak poważnie zniekształconego słowa nie wykryje! Podobnie nie wykryje błędów poczwórnych, gdyby takie były, natomiast wykryje zaistnienie błędu potrójnego, pięciokrotnego itd.

Bit parzystości pozwala na wykrywanie błędów nieparzystego rzędu.

Bity nadmiarowe (kontrolne) to rozwinięcie systemu bitu parzystości: polega ono na zastosowaniu grupy kilku bitów parzystości. Zabezpieczane słowo informacyjne dzielone jest na mniejsze grupy po kilka bitów i do każdej z tych grup dołącza się odpowiedni bit parzystości. Ze względu na krótkość grup niemożliwe lub mało prawdopodobne są w nich błędy wielokrotne (w grupie), natomiast można wykrywać wielokrotne (także parzyste) błędy w całym słowie i je znów odrzucać lub żądać retransmisji.



Rys. 6.11. Przykładowy podział 8-bitowego słowa informacyjnego na grupy i uzupełnienie go bitami nadmiarowymi.

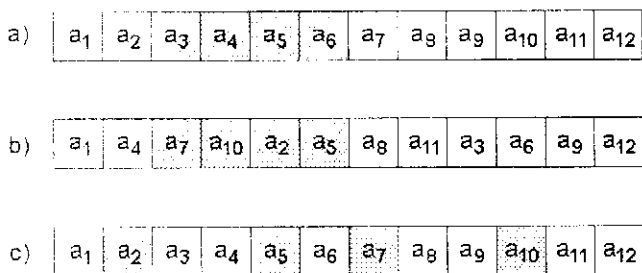
Lepszym rozwiązaniem jest taki podział słowa na mniejsze grupy bitów, aby kontrolowane grupy bitów zachodziły na siebie (rys. 6.11) i do takich grup ustalenie bitu parzystości (grup takich będzie więcej i także więcej bitów dodatkowych). Wtedy, analizując odebrane słowo i weryfikując spełnienie parzystości w każdej z grup, można wyznaczyć miejsce wystąpienia błędu w odebranym słowie. Znajomość tego miejsca pozwala na usunięcie błędu przez zanegowanie błędnego bitu (gdyż nie ma innej możliwości – błąd mógł polegać tylko na tym, że nadane „0” odebrane zostało jako „1” lub odwrotnie).

Taki sposób kodowania informacji cyfrowej, który pozwala na poprawianie błędów bez dodatkowej łączności z nadajnikiem, nazywa się **kodowaniem korekcyjnym**, a umożliwiające to kody – **kodami korekcyjnymi**. Mogą to być kody w rodzaju już przedstawionego, mogą to być bardziej efektywne kody z grupy tzw. kodów blokowych lub splotowych czy najnowszych, tzw. turboko-

dów [10]. Kodowaniu korekcyjnemu poświęconych jest wiele prac; ten sposób kodowania stosowany jest np. w zapisie dźwięku, w transmisji głosu w telefonii komórkowej, przy zapisie danych na nośnikach magnetycznych (dyski komputerowe).

Bity nadmiarowe zwiększają ogólną liczbę transmitowanych bitów, które muszą się zmieścić w tym samym czasie – zatem impulsy reprezentujące te bity muszą być krótsze, a odpowiadające im pasmo częstotliwości szersze. W praktycznych systemach z korekcją błędów pasmo to ulega zwiększeniu np. o kilkanaście procent, a systemy te cechują się dużą złożonością.

Przeplot bitów to jeszcze jeden sposób walki z wpływem zakłóceń na transmisję PCM. Polega on na tym, że przed transmisją naturalna kolejność bitów, w jakiej zostały one wytworzone, zostaje celowo zmieniona – bity zostają „pomieszane” wg określonego algorytmu i tak są transmitowane. Jeżeli na dane słowo podziała silne zakłócenie, trwające przez czas kilku bitów, bity te „wypadną” tak, że będą zupełnie niewiadome. Po odebraniu tej sekwencji i dokonaniu „rozplotu”, tzn. po ustawieniu bitów w prawidłowej, naturalnej kolejności, okaże się, że bity zniekształcone nie sąsiadują z sobą. W pewnych złożonych systemach transmisji ułatwia to zdekodowanie sygnału, gdy bity zniekształcone nie tworzą jednej „paczki”, tylko są przedzielone bitami odebranymi poprawnie (rys. 6.12).

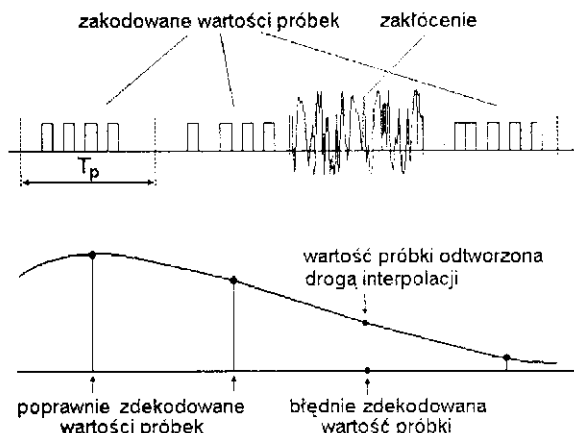


Rys. 6.12. Ilustracja przeplotu bitów: a) naturalna kolejność bitów – zaciemniono „paczkę” bitów zakłóconych, b) słowo z przeplotem bitów i zakłóconą tą samą paczką bitów, c) słowo po rozplotcie bitów – bity zakłócone nie stanowią zwartej paczki.

Interpolacja wartości sygnału to ostatni, tu omawiany, sposób zabezpieczania transmisji PCM. Dotyczy ona (interpolacja) tylko transmisji sygnałów PCM reprezentujących jakiś oryginalny sygnał analogowy (dźwięk, obraz...).

Jeżeli nastąpi całkowite zakłócenie sygnału PCM przenoszącego informację o wartości jakiejś próbki sygnału oryginalnego, to zamiast przyjmować wartość zerową tej próbki albo czekać na powtórzenie tego sygnału, prościej i szybciej będzie obliczyć średnią wartości poprzedniej i następnej próbki. Gdy sygna-

łem oryginalnym jest sygnał analogowy o ograniczonej szybkości zmian (twierdzenie o próbkowaniu!), to jest mało prawdopodobne, aby zmienił się on gwałtownie w czasie 2 lub 3 okresów próbkowania. Wtedy wartość średnia będzie lepszym przybliżeniem rzeczywistej wartości sygnału niż wartość zerowa czy jakaś inna ustalona wcześniej. Zostało to pokazane na rys. 6.13.



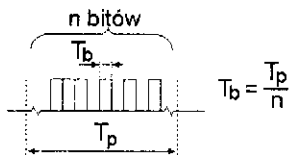
Rys. 6.13. Ilustracja zastosowania interpolacji wartości sygnału do zmniejszania wpływu zakłóceń na transmisję PCM.

Wszystkie sposoby zabezpieczania sygnałów PCM przed błędami powodują, że są one odporne na zakłócenia i przewyższają pod tym względem poznane już sygnały analogowe. Dokładniejsze zestawienie właściwości sygnałów analogowych i cyfrowych pod tym względem będzie dokonane w rozdz. 7.

6.5. Pasmo częstotliwości wymagane do transmisji PCM

Podczas transmisji szeregowej wszystkie bity tworzące słowo wyjściowe kodera PCM oraz ewentualne bity dodatkowe (nadmiarowe) transmitowane są kolejno, w czasie nie większym niż długość jednego okresu próbkowania (bo w następnym okresie przetwornik A/C musi przetwarzać już następną próbkę sygnału). Sytuację tę można przedstawić w sposób pokazany na rys. 6.14.

W niekorzystnym – pod względem szerokości pasma – przypadku bity transmitowane mogą się ułożyć na przemian: 1 0 1 0... i uformować w ten sposób krótkotrwałą falę prostokątną. Częstotliwość powtarzania impulsów w takiej fali będzie wynosić



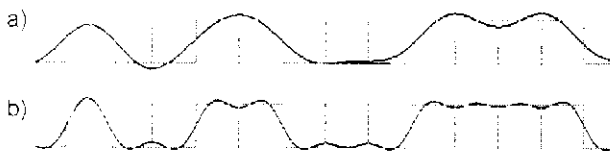
Rys. 6.14. Ilustracja zależności czasowej okresu próbkowania i czasu trwania jednego bitu.

$$f = \frac{1}{2T_b} = \frac{F_p \cdot n}{2}, \quad (6.8)$$

i tę wartość można by przyjąć jako najmniejszą szerokość pasma częstotliwości potrzebnego do poprawnego przesłania sygnału PCM. Okazuje się jednak, że do dobrego przenoszenia dłuższego ciągu jednakowych bitów (jedynek lub zer) pasmo to powinno być co najmniej dwa razy szersze (rys. 6.15). Dlatego jako szerokość pasma sygnału PCM przyjmuje się wartość

$$B_{PCM} = F_p \cdot n = \frac{1}{T_b}, \quad (6.9)$$

czyli pasmo to jest równe odwrotności czasu trwania jednego bitu. Jest to minimalna szerokość pasma; niekiedy (o czym będzie mowa dalej) wymagane pasmo jest jeszcze szersze.



Rys. 6.15. Kształty przenoszonych impulsów przy różnej szerokości pasma (zaznaczono granice między kolejnymi bitami): a) pasmo $B = 1/2T_b$ – zależność (6.8), b) pasmo $B = 1/T_b$ – zależność (6.9).

Jakie przykładowe wartości przybiera to pasmo? Pokazano to w tab. 6.1.

Tab. 6.1. Zestawienie parametrów sygnałów akustycznych i odpowiadających im sygnałów PCM.

Rodzaj sygnału oryginalnego parametry sygnału PCM	Pasma sygnału oryginalnego	Pasma sygnału PCM
Sygnał mowy w telefonii $F_p = 8 \text{ kHz}$, $n = 8$	3.1 kHz	$B \geq 64 \text{ kHz}$
Dźwięk wysokiej jakości $F_p = 44.1 \text{ kHz}$, $n = 16$	20 kHz	$B \geq 706 \text{ kHz}$

Jak widać z podanego zestawienia, pasmo potrzebne do transmisji sygnału PCM (i to jeszcze przy nieidealnej, nieprostokątnej postaci impulsów) jest bardzo szerokie, kilkadziesiąt razy szersze od pasma oryginalnych sygnałów analogowych. Można by więc uznać, że modulacja PCM „marnotrawi” pasmo częstotliwości, które mogłoby być bardziej efektywnie wykorzystane przy innych sposobach transmisji. Początkowo tak właśnie uważano, dlatego, po wynalezieniu, modulacja PCM została na pewien czas „odłożona na półkę” jako nieefektywna.

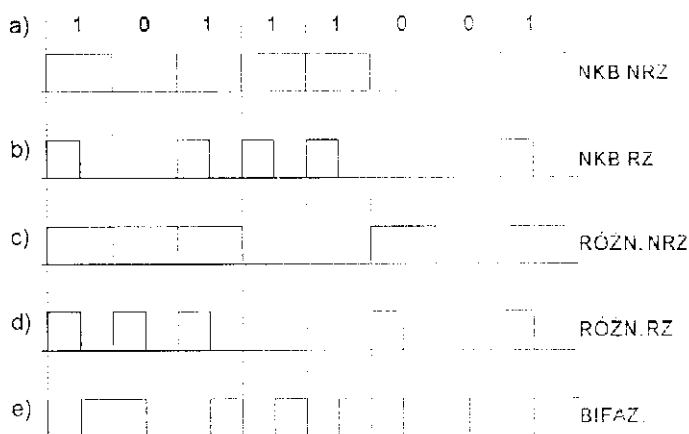
Ponowne zainteresowanie techniką PCM i obecne jej rozpowszechnienie wynikało m.in. z dużej odporności na zakłócenia, znacznie większej niż w innych sposobach transmisji.

6.6. Sposoby kodowania sygnałów PCM

Kod to sposób, w jaki tworzące wiadomość w postaci cyfrowej bity (0 i 1) są reprezentowane za pomocą impulsów napięcia, prądu czy przez natężenie światła.

Przetwornik A/C na rys. 6.1 najczęściej wytwarza słowo wyjściowe w tzw. **naturalnym kodzie binarnym** (NKB), tzn. bitowi „1” odpowiada wysoki poziom napięcia, bitowi „0” zaś odpowiada niski i w takiej postaci sygnał może być transmitowany (rys. 6.16a).

Kod ten może być tzw. **kodem bez powrotu do zera** (NRZ – *NonReturn to Zero*), co oznacza, że w wypadku dłuższej trwającego ciągu bitów 1111... na wyjściu koderu utrzymywana jest ciągle wysoka wartość napięcia, albo **kodem z powrotem do zera** (RZ – *Return to Zero*) (rys. 6.16b).



Rys. 6.16. Istota kodowania bitów „0” i „1” za pomocą kilku prostych unipolarnych kodów binarnych

Możliwy jest też **kod różnicowy** – wtedy wysoki poziom napięcia oznacza, że aktualnie nadawany i poprzedni bit się różnią, natomiast stan niski – że zmiana wartości bitu nie wystąpiła (rys. 6.16c i d).

Przedstawione sposoby kodowania mają tę wadę, że w wypadku specyficznych wartości sygnału oryginalnego mogą wystąpić długie okresy bez żadnego impulsu lub z wieloma impulsami, następującymi bezpośrednio po sobie. Jest to niekorzystne, ponieważ utrudnia zachowanie synchronizacji odbiornika z nadajnikiem. Synchronizacja taka jest konieczna, aby odbiornik „wiedział”, gdzie kończy się informacja o jednym bicie i gdzie się zaczyna o następnym. Specjalne przesyłanie sygnałów synchronizacji byłoby niepraktyczne, gdyż „zabierałoby” część dostępnej przepustowości systemu telekomunikacyjnego.

Podczas transmisji PCM konieczna jest synchronizacja odbiornika z nadajnikiem.

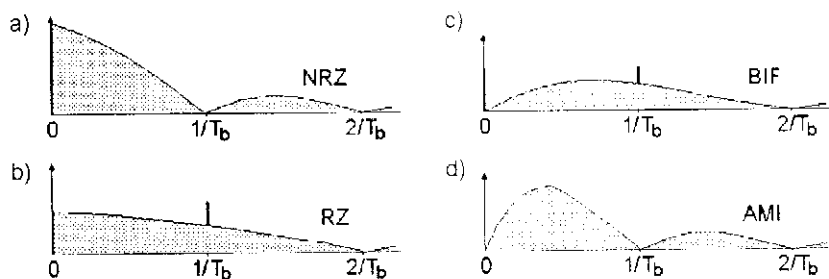
Pod względem możliwości utrzymania synchronizacji kod RZ jest lepszy od kodu NRZ, ale w przypadku długiego ciągu samych zer też „zawodzi”. Z tego względu opracowano tzw. kody bifazowe, w których informacja przekazywana jest poprzez zmiany poziomu napięcia w połowie przedziału czasu przeznaczanego na dany bit, przy czym zmiana ta w jedną stronę oznacza 1, a w drugą 0. Zmiany poziomu, które zachodzą na granicy bitów, nie przenoszą informacji i odbiornik je ignoruje. Można sprawdzić, że w takim systemie kodowania niemożliwe jest, aby sygnał zakodowany zachowywał ustaloną wartość zerową lub wysoką przez czas większy niż czas trwania jednego bitu. Sygnał odbierany w dowolnym przypadku zawiera w sobie silną składową harmoniczną o częstotliwości równej częstotliwości powtarzania bitów. Składowa ta może być wykorzystana do synchronizacji odbiornika z nadajnikiem. Jednym z takich kodów jest popularny kod o nazwie Manchester.

Na rysunku 6.17 pokazano widma sygnałów zakodowanych za pomocą wybranych kodów. Są to obwiednie widm, jakie mogą wystąpić dla różnych możliwych kombinacji ciągów zer i jedynek, stanowiących przesyłaną informację.

Widoczne jest, że kody charakteryzują się różnymi postaciami widm i ich różną szerokością (jako szerokość pasma przyjmuje się częstotliwość pierwszego zera w widmie). Kody RZ cechują się dwa razy szerszym pasmem niż kody NRZ, dlatego te drugie są stosowane znacznie częściej. Również kody bifazowe wymagają dwukrotnie szerszego pasma, ale ze względu na dobre właściwości synchronizacyjne są dość często używane, np. w komputerowych sieciach LAN.

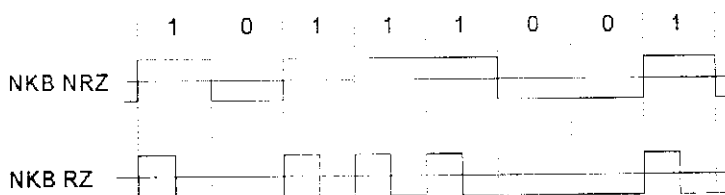
Przedstawione dotychczas kody mają wspólną cechę – wytwarzane sygnały cyfrowe zawierają określoną składową stałą (gdyż zawierają impulsy jednobiegunowe). Często obecność składowej stałej w sygnale PCM jest niepożądana,

bo uniemożliwia stosowanie wygodnych pojemnościowych i transformatorowych sprzężeń w sieciach kablowych.



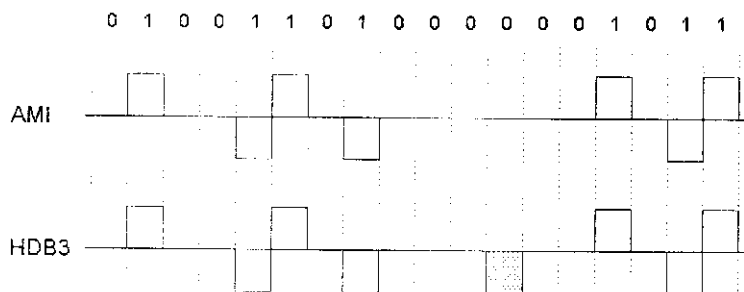
Rys. 6.17. Amplitudowe widma sygnałów zakodowanych za pomocą wybranych kodów: a) widmo kodu NRZ, b) widmo kodu RZ, c) widmo kodu bifazowego, d) widmo kodu AMI (patrz dalej w tekście).

W zasadzie wystarczyłoby zastosować kody dwubiegunowe (bipolarne), które obserwowane długookresowo, nie powinny zawierać składowej stałej w przypadku jednakowej liczby występujących w nich jedynek i zer (rys. 6.18).



Rys. 6.18. Przykładowe kody bipolarne.

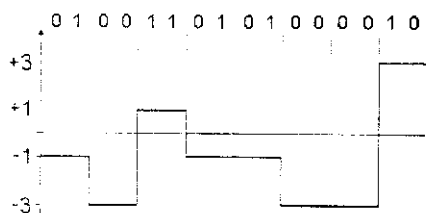
Aby uwolnić się od tego warunku, opracowano kody, które z założenia nie zawierają składowej stałej. Są to kody dwubiegunowe, w których liczba impulsów ujemnych i dodatnich równoważy się w krótkim przedziale czasu. Przykładem takiego kodu może być kod AMI (*Alternate Mark Inversion*), w którym jedynki kodowane są na przemian impulsami $+$ lub $-$, natomiast brak impulsu oznacza zero (rys. 6.19). Kod ten nie zapewnia utrzymania synchronizacji w razie długiego ciągu zer, dlatego modyfikacją kodu AMI jest np. kod HDB3 czy ogólnie HDBn (*High Density Bipolar*), w którym – w przypadku dłużej trwającego ciągu zer – w miejscu n -tego zera dodawany jest impuls, naruszający zasadę naprzemiennych impulsów (w ten sposób może być on odróżniony od „zwykłych” impulsów, kodujących jedynek). Zmodyfikowany kod AMI stosowany jest w sieciach ISDN, natomiast kod HDB3 w międzycentralowych sieciach telefonicznych.



Rys. 6.19. Przykładowe przebiegi sygnałów w kodach AMI oraz HDB3. Zaciemniono impuls naruszający zasadę naprzemiennej biegunowości kolejnych impulsów.

Wszystkie przedstawione kody były kodami dwuwartościowymi (jedynie kody AMI i HDB3 mogą być nazwane pseudotrójwartościowymi – inaczej pseudoternarnymi). Typowe kodowanie wielowartościowe w transmisji PCM w pasmie naturalnym stosowane jest rzadziej. Polega ono na tym, że ustalonych jest kilka (4, 8...) znamionnych (charakterystycznych) wartości napięcia sygnału; więcej wartości to więcej stanów, jakie można zakodować; każdy z nich jest odpowiednio mniej prawdopodobny i może przenosić informację np. o 2, 3... bitach na raz.

W sieciach ISDN stosowany jest np. kod 2B1Q, którego istotę pokazano na rys. 6.20. Jest to kod czterowartościowy, dwubiegunowy. W zależności od wartości tzw. dwubitu (duobitu) – 00, 01, 11, 10 napięcie wyjściowe kodera przybiera jedną z 4 możliwych wartości. Wartości te są z konieczności mniej odległe od siebie niż mogłyby być przy kodowaniu dwuwartościowym, dlatego w czasie przesyłania takich sygnałów konieczny jest mniejszy poziom zakłóceń, aby te 4 wartości mogły być prawidłowo rozróżniane.



Rys. 6.20. Przebieg napięcia w linii w przypadku kodu 2B1Q (skala pionowa podano w jednostkach względnych).

Może to być uzasadnione też na bazie wzoru (2.1) na przepustowość kanału telekomunikacyjnego. Aby przy tym samym pasmie częstotliwości można było

przesłać 2 razy więcej informacji (co w przypadku tego kodu występuje), stosunek S/N musi być odpowiednio większy.

Przedstawione kody można porównać pod względem tzw. **efektywności widmowej**. Jest to przepływność informacji przypadająca na 1 Hz szerokości pasma zajmowanego przez zakodowany sygnał (tab. 6.2).

Tab. 6.2. Efektywność widmowa wybranych kodów.

Kod	Efektywność widmowa, b/s/Hz
NRZ, AMI, HDB3	1
RZ, bifazowy	0,5
2B1Q	2

6.7. Modulacja PCM fali nośnej

Przedstawione sposoby kodowania dotyczyły transmisji cyfrowej w pasmie naturalnym (podstawowym). W razie konieczności transmisji na większe odległości, szczególnie drogą radiową, konieczne jest wykorzystanie fali nośnej i odpowiedniej jej modulacji. Stosowane tu są (lub były) modulacje analogiczne do poprzednio omówionych (AM, FM, PM), jednak ze względu na inny charakter sygnałów modulujących noszą one odpowiednio nazwy:

- **ASK** (*Amplitude Shift Keying*) – kluczkowanie (z przesuwem) amplitudy,
- **FSK** (*Frequency Shift Keying*) – kluczkowanie (z przesuwem) częstotliwości,
- **PSK** (*Phase Shift Keying*) – kluczkowanie (z przesuwem) fazy.

Istota tych modulacji została zilustrowana na rys. 6.21.

Modulacja ASK polega na włączaniu i wyłączaniu (kluczkowaniu) fali nośnej (lub jakiejś pomocniczej – tzw. podnośnej) – w zależności od tego, czy transmisji podlega symbol cyfrowy „1” czy „0”. Ze względu na niewielką odporność na zakłócenia modulacja ASK nie jest obecnie stosowana.

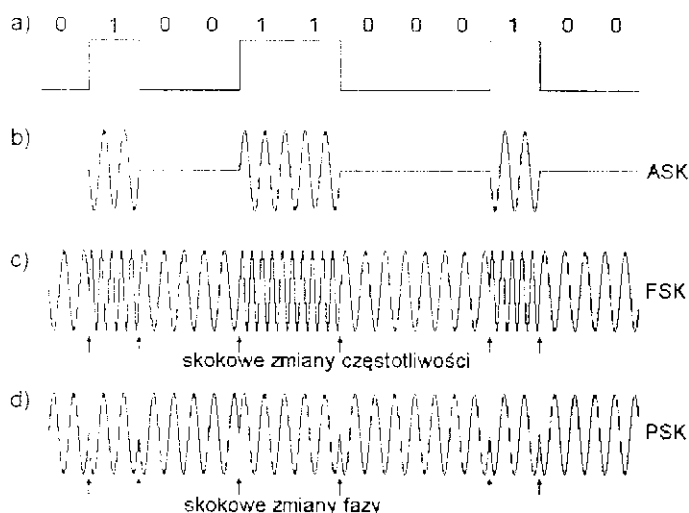
Modulacja FSK polega na skokowych zmianach częstotliwości fali (pod)nośnej, a jej amplituda pozostaje stała. W zależności od nadawanego bitu (symbolu) częstotliwość ta przybiera jedną z wartości znamienych F_1 lub F_2 . Spoczynkowa wartość częstotliwości nośnej w rzeczywistości nie występuje; dla wygody przyjmuje się, że wynosi ona

$$F_0 = \frac{F_1 + F_2}{2}, \quad (6.10)$$

natomiast jako dziewięć częstotliwości można przyjąć

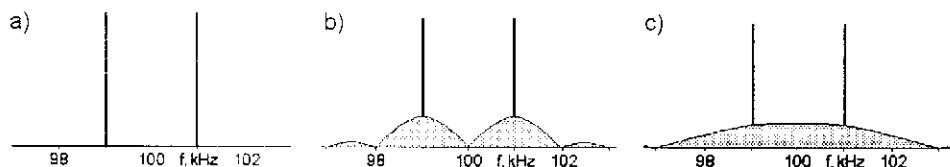
$$\Delta F_0 = \left| \frac{F_1 - F_2}{2} \right|. \quad (6.11)$$

Modulacja ta była używana powszechnie w początkowym okresie rozwoju łączności cyfrowej przez analogowe łącza telefoniczne, była używana do zapisu informacji cyfrowej na taśmę magnetyczną (pierwsze komputery osobiste i domowe z lat 80. ub. wieku), w systemie telefonii komórkowej NMT, obecnie jest używana np. do przesyłania dodatkowych informacji do abonenta w analogowych sieciach telefonicznych (usługa CLIP).



Rys. 6.21. Cyfrowe modulacje pasmowe: a) cyfrowy sygnał modulujący, b) sygnał zmodulowany ASK, c) sygnał zmodulowany FSK, d) sygnał zmodulowany PSK.

Widmo sygnału FSK w razie powolnego kluczowania częstotliwości zawiera praktycznie tylko 2 prążki związane z częstotliwościami znamionowymi F_1 i F_2 (rys. 6.22a).



Rys. 6.22. Przykładowe widma modulacji FSK przy $F_0 = 100$ kHz, $F_2 - F_1 = 2$ kHz: a) przy bardzo powolnym kluczowaniu, b) przy szybkości modulacji 1 kb/s, c) przy szybkości modulacji 2 kb/s.

Przy większej szybkości kluczowania częstotliwości nośnej obserwuje się „zapełnianie” pasma pomiędzy częstotliwościami znamionowymi i zwiększanie szerokości zajętego pasma, ale zawsze w widmie obecne są silne składowe o częstotliwości F_1 i F_2 (rys. 6.22b i c).

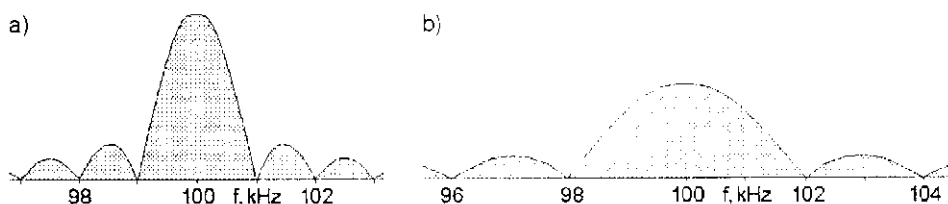
Szerokość pasma sygnału FSK można wyrazić wzorem

$$B_{FSK} \approx 2\Delta F + 2F_{sb}, \quad (6.12)$$

gdzie F_{sb} oznacza częstotliwość równą liczbowo szybkości kluczowania fali nośnej; w typowym kluczowaniu dwuwartościowym

$$F_{sb} = 1/T_b, \quad (6.13)$$

Modulacja PSK polega na skokowych zmianach fazy fali (pod)nośnej o 180° , z zachowaniem jej stałej amplitudy i częstotliwości. Należy jednak pamiętać, że zmianom fazy zawsze towarzyszą zmiany częstotliwości (tym bardziej zmianom skokowym!). Dlatego sygnały PSK zajmują skończone pasmo częstotliwości – tym szersze, im szybciej zachodzi kluczowanie (rys. 6.23).



Rys. 6.23. Przykładowe widna modulacji PSK przy $F_n = 100$ kHz:
a) przy szybkości modulacji 1 kb/s, b) przy szybkości modulacji 2 kb/s.

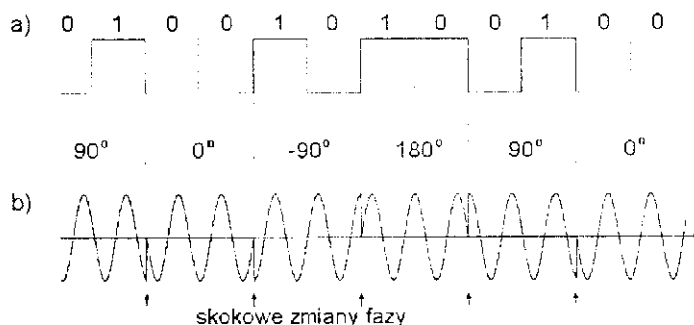
Szerokość pasma sygnału PSK, obejmująca tzw. główny listek widma, wyraża się wzorem

$$B_{PSK} \approx 2F_{sb}, \quad (6.14)$$

gdzie wielkość F_{sb} została określona zależnością (6.13).

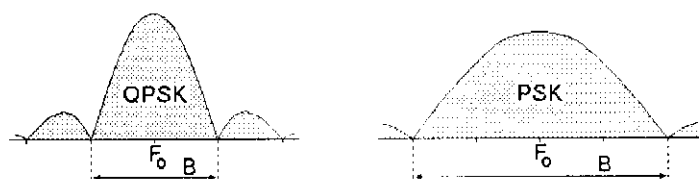
Spośród przedstawionych modulacji modulacja PSK jest najbardziej odporna na zakłócenia i dlatego – w różnych odmianach – jest często stosowana. Popularną odmianą jest czterowartościowa modulacja PSK, oznaczana QPSK (*Quarternary* PSK) lub 4-PSK. Polega ona na tym (rys. 6.24), że skoki fazy fali (pod)nośnej są skwantowane co 90° , co daje 4 możliwe wartości tej fazy (np. -90° , 0° , 90° , 180°); za ich pomocą można przedstawić 4 różne wartości sygna-

łu cyfrowego, zatem można w ten sposób zakodować 4 dwubity (duobity) o postaci 00, 01, 11, 10.



Rys. 6.24. Modulacja QPSK: a) cyfrowy sygnał modulujący, b) sygnał zmodulowany QPSK.

Każdy z tych dwubitów ma prawdopodobieństwo $1/4$, a więc zawiera 2 bity informacji. W ten sposób przy określonej częstotliwości występowania skoków fazy (szybkości symbolowej) uzyskuje się 2 razy większą przepływność informacji albo inaczej taką samą przepływność przy 2 razy mniejszej szybkości nadawania symboli, czyli przy węższym pasmie (rys. 6.25). Na zmniejszenie pasma ma też wpływ fakt, że skoki fazy fali nośnej są tu statystycznie mniejsze, a więc i związana z tym jej nieciągłość, i wynikające stąd skoki częstotliwości chwilowej.



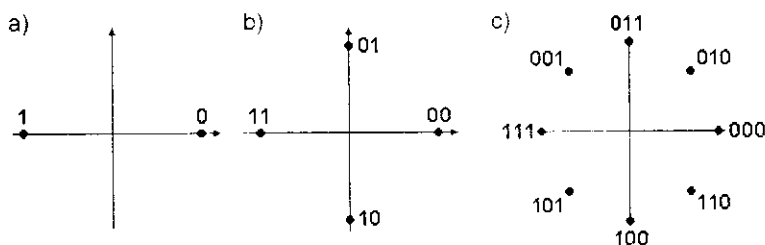
Rys. 6.25. Porównanie widm modulacji QPSK i PSK przy tej samej przepływności transmitowanej informacji.

Ta zaleta modulacji QPSK jest okupiona koniecznością zapewnienia lepszej jakości sygnału przy odbiorze, tj. zapewnieniem większego stosunku S/N (w porównaniu z typowym sygnałem PSK, nazywanym BPSK (*Binary PSK*)).

Modulacja QPSK jest stosowana w transmisjach modemowych, w systemach dostępu radiowego do Internetu oraz w cyfrowych satelitarnych transmisjach TV.

W przypadku pasmowych modulacji cyfrowych, w których zachodzą zmiany fazy fali (pod)nośnej, wygodnie jest prezentować te wartości fazy na płaszczyźnie

czyźnie zespolonej, tak jak to zwykle czyni się dla wielkości harmoniczných (wykresy wskazowe). Sytuację taką w przypadku modulacji BPSK przedstawiono na rys. 6.26a, dla QPSK zaś na rys. 6.26b.



Rys. 6.26. Konstelacje fazowe dla różnych modulacji typu PSK: a) BPSK, b) QPSK, c) 8-PSK.

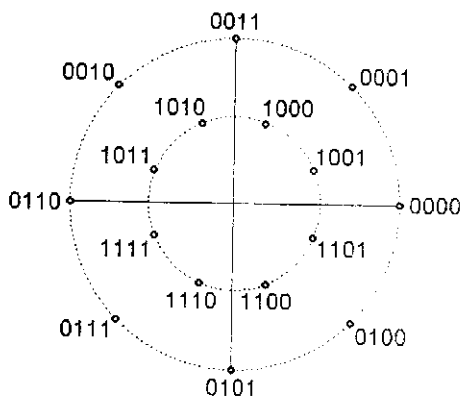
Odległości zaznaczonych punktów od początku układu współrzędnych oznaczają amplitudę fali nośnej, kątowne zaś położenia decydują o fazie początkowej. Nie rysuje się samych wektorów, zaznacza się tylko, gdzie znajdują się ich końce („groty”) przy danej kombinacji bitów.

W każdym przypadku korzystne jest, aby znacznym różnicom bitów czy grup bitów odpowiadały znaczne różnice fazy odpowiednich wektorów. Duże zmiany fazy jest łatwiej dekodować, tym samym maleje prawdopodobieństwo błędów, które prowadziłyby do fałszywego odbioru jednocześnie kilku bitów.

Główna zaleta modulacji QPSK to możliwość zwiększenia przepływności informacji bez zwiększania szerokości pasma. Okazało się to bardzo efektywne, dlatego opracowano następne modulacje typu PSK, o jeszcze większej wartościowości i szybkości transmisji: 8-PSK, 16-PSK itd., w których jednemu skokowi fazy odpowiada kombinacja 3, 4 i więcej bitów ciągu danych (rys. 6.26c). Należy jednak zwrócić uwagę na to, że dobry odbiór takich sygnałów wymaga coraz lepszej ich jakości, gdyż poszczególne znamienne stany sygnału coraz mniej się różnią.

Przy dużej wartościowości rozważanych modulacji (np. 16, 32...), kiedy skoki fazy fali nośnej byłyby zbyt małe, aby zapewnić ich prawidłowe dekodowanie, modulację typu PSK uzupełnia się jeszcze modulacją ASK tak, aby uzyskać możliwie duże różnice pomiędzy różnymi stanami fali nośnej. Modulacje takie oznacza się jako m-PM/AM lub m-PSK/ASK, gdzie m (16, 32, 64...) jest wartościowością modulacji, tzn. liczbą możliwych stanów, jakie mogą przybierać parametry fali nośnej (co do fazy lub co do amplitudy).

Przykładową konstelację położenia wektora fali nośnej sygnału 16-AM/PM pokazano na rys. 6.27.



Rys. 6.27. Przykładowa konstelacja fazowa sygnału modulacji 16-AM/PM.

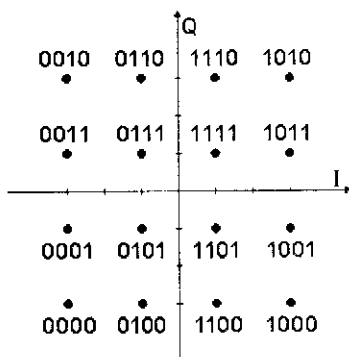
W przedstawionym przypadku punkty konstelacji położone są na dwóch okręgach, odpowiadających dwóm wartościom amplitudy fali nośnej; dla każdej wartości amplitudy faza fali może przybierać po 8 różnych wartości (możliwe są też inne rozkłady punktów konstelacji).

Każdy sygnał harmoniczny można przedstawić jako sumę geometryczną składowej o fazie odniesienia (np. 0°) i składowej do niej ortogonalnej ($\pm 90^\circ$). Ze względu na łatwość wytwarzania i demodulacji sygnału m-ASK/PSK celowe jest, aby jego wszystkie stany dały się wyrazić przez sumę najmniejszej liczby różnych wartości takich składowych. Składowe te noszą odpowiednio nazwy – synfazowej (*In phase*) oraz kwadraturowej (*Quadrature in phase*) i oznaczane są jako I oraz Q; używane są one jako współrzędne w przedstawianych konstelacjach fazowych.

Ze względu na minimalizację liczby potrzebnych wartości składowych I i Q korzystne jest, aby punkty konstelacji fazowej leżały na liniach równoległych do osi współrzędnych, a nie na okręgach. Tego typu modulacje oznacza się jako m-QAM (*Quadrature AM*), bo w istocie można je potraktować jako tylko modulację amplitudy (dyskretną) dwóch składowych: synfazowej i kwadraturowej. Przykład konstelacji fazowej dla modulacji 16-QAM przedstawiono na rys. 6.28. Składowe synfazowa i kwadraturowa mogą przybierać tu po 4 różne wartości względne ($-3, -1, +1, +3$); ich wszystkie możliwe kombinacje dają 16 różnych stanów wypadkowej fali nośnej.

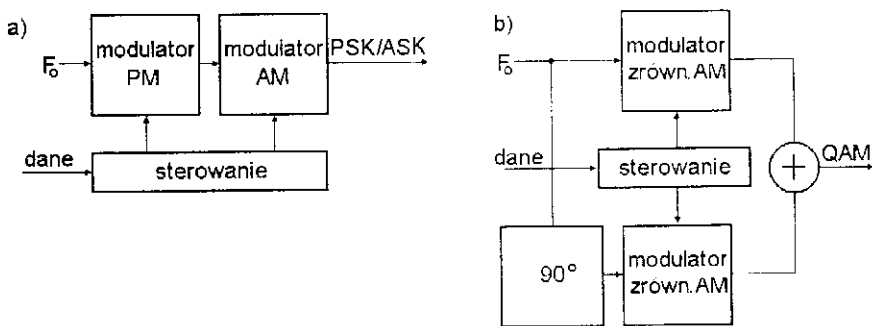
Można zauważyć, że w przedstawionej konstelacji pierwszy bit w każdej ich czwórce określa znak składowej I, trzeci bit – znak składowej Q. Możliwa jest tu duża dowolność w przypisywaniu czwórek bitów poszczególnym stanom fali, dlatego w praktyce warto stosować się do odpowiednich zaleceń [10].

O zaletach modulacji QAM nad modulacjami ASK/PSK może świadczyć rys. 6.29, na którym pokazano sposób otrzymywania sygnałów o konstelacji fazowej z rys. 6.27 oraz z rys. 6.28 – tj. sygnałów o tej samej wartościowości.



Rys. 6.28. Przykładowa konstelacja dla modulacji 16-QAM.

W pierwszym przypadku potrzebny jest dwuwartościowy modulator amplitudy i aż 16-wartościowy modulator fazy (niekorzystne jest to, że są to dwa różne modulatory). W drugim przypadku potrzebne są dwa identyczne czterowartościowe modulatory amplitudy (o prostej budowie).



Rys. 6.29. Przykładowe sposoby otrzymywania: a) modulacji 16-ASK/PSK, b) modulacji 16-QAM.

Modulacja 64-QAM jest stosowana np. do przesyłania sygnału telewizji cyfrowej w sieciach kablowych. Ze względu na takie medium transmisyjne i ograniczoną rozległość tych sieci, jakość sygnału docierającego do odbiorników może być bardzo dobra i możliwe jest zastosowanie modulacji, w której parametry sygnału w poszczególnych stanach tak mało się różnią.

Jeszcze bardziej wartościowa modulacja, 256-QAM, przewidziana jest przy dostępie do Internetu poprzez sieci kablowe TV.

Podobnie jak poprzednio, można podać efektywność widmową pasmowych modulacji PCM (tab. 6.3).

Tab. 6.3. Efektywność widmowa pasmowych modulacji PCM (z falą nośną).

Rodzaj modulacji	Efektywność widmowa, b/s/Hz
ASK	0.5
FSK	< 0.5
PSK	0.5
QPSK	1
m-QAM	$\frac{\log_2 m}{2}$

Zadania

1. Moc szumu kwantyzacji w pewnym 8-bitowym systemie PCM jest równa 2 mW. Ile będzie wynosić ta moc po zmianie przetwornika na 10-bitowy?
2. O ile zmieni się wartość S/N w systemie PCM, jeżeli moc wejściowa sygnału zmaleje 20-krotnie?
3. Jakie pasmo byłoby potrzebne do transmisji sygnału TV o przepływności obliczonej w rozdz. 1. za pomocą kodu NRZ?
4. W pewnym systemie PCM zastosowano przetwornik A/C 8-bitowy o zakresie przetwarzania ± 10 V. Jaka jest szerokość przedziału kwantyzacji?
5. Porównać efektywność widmową transmisji w pasmie naturalnym (tab. 6.2) z efektywnością widmową modulacji pasmowych PCM (tab. 6.3).
6. Ilościowe modulatory amplitudy należałoby zastosować w układzie z rys. 6.29, aby można było wytwarzać sygnał 64-QAM oraz 256-QAM?
7. W nadajniku 16-QAM maksymalna amplituda składowych I oraz Q jest równa 1 V. Jaka maksymalna amplitudę może mieć cały sygnał składowy?
8. Jaka jest efektywność widmowa modulacji 64-QAM i 256-QAM?

Transmisja sygnałów w obecności zakłóceń

7.1. Właściwości szumu

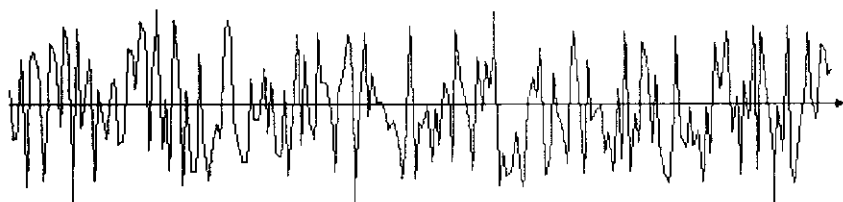
Przetwarzaniu sygnałów w telekomunikacji, a zwłaszcza ich przesyłaniu na większe odległości, zawsze towarzyszą przebiegi niepożądane, pochodzące zarówno z zewnątrz systemu, jak i wytwarzane wewnątrz niego, w jego blokach i urządzeniach składowych.

Przebiegi te ogólnie nazywane są szumem (szumami), ponieważ w zakresie częstotliwości akustycznych, doprowadzone np. do słuchawki czy głośnika, dają efekt podobny do szumu, w którym nie da się wyróżnić żadnej dominującej składowej częstotliwościowej.

Szum pochodzenia zewnętrznego (zewnętrzny) jest skutkiem zjawisk elektrycznych zachodzących w atmosferze (burze, wyładowania) czy w przestrzeni kosmicznej; stanowią one też uboczny lub nawet celowy produkt działalności człowieka (zakłócenia przemysłowe, tzw. walka radioelektroniczna).

Szum pochodzenia wewnętrznego (wewnętrzny) to szum generowany na skutek subtelnych zjawisk związanych z przepływem i ruchem nośników prądu w materiałach przewodzących, rezystancyjnych, półprzewodnikowych itd.

Podstawową cechą szumu elektrycznego jest jego przypadkowość; nie można opisać go żadną funkcją czasową. Niemniej szum ten cechuje się pewnymi prawidłowościami – jest on **stacjonarny**, tzn. jego parametry statystyczne nie zależą od położenia okresu obserwacji na osi czasu; jest on także **ergodyczny**, tzn. jego parametry wyznaczone na podstawie jednego przebiegu czasowego są takie same jak parametry wyznaczone dla jednego momentu czasu, ale na podstawie wielu przebiegów (realizacji) czasowych. Przykładowy przebieg szumowy przedstawiono na rys. 7.1.



Rys. 7.1. Przykładowy przebieg napięcia lub prądu szumu.

Jeżeli jedną z realizacji przebiegu szumowego (stacjonarnego i ergodycznego) oznaczyć formalnie jaką funkcję $n(t)$, to podstawowe parametry statystyczne szumu będą wynosić:

- wartość średnia (oczekiwana)

$$\overline{n(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T n(t) dt, \quad (7.1)$$

- wartość średniokwadratowa

$$\overline{n^2(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T n^2(t) dt, \quad (7.2)$$

- gęstość prawdopodobieństwa

$$g(n) = \lim_{\Delta n \rightarrow 0} \frac{p(n - \Delta n < n < n + \Delta n)}{2\Delta n}. \quad (7.3)$$

Zgodnie z zależnością (7.3) prawdopodobieństwo tego, że chwilowe napięcie szumu w danym momencie będzie zawarte pomiędzy wartościami n_1 i n_2 , jest równe

$$p(n_1 < n < n_2) = \int_{n_1}^{n_2} g(n) \cdot dn. \quad (7.4)$$

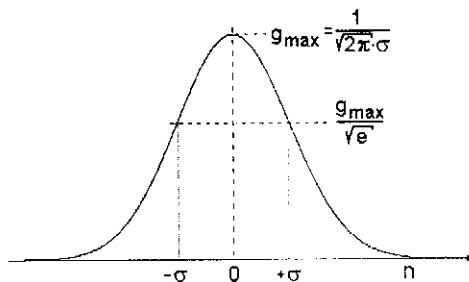
Chwilowa moc szumu jest równa kwadratowi wartości chwilowej, czyli n^2 . Wartość średnia mocy szumu (odpowiednik mocy czynnej) jest to średnia moc obliczona dla wszystkich możliwych chwilowych wartości napięcia szumu, ale z uwzględnieniem ich prawdopodobieństwa (czyli inaczej mówiąc – jak często te wartości występują). Zatem moc średnią szumu N można wyrazić jako

$$N = \int_{-\infty}^{+\infty} n^2 g(n) \cdot dn. \quad (7.5)$$

W rozważaniach teoretycznych bardzo często występuje pojęcie **szumu gaussowskiego**, tj. takiego, dla którego funkcja gęstości prawdopodobieństwa dana jest zależnością [1, 2, 5]:

$$g(n) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{n^2}{2\sigma^2}}. \quad (7.6)$$

gdzie wielkość σ nosi nazwę **dyspersji** albo **odchylenia standardowego**. Wykres funkcji (7.6) został pokazany na rys. 7.2.



Rys. 7.2. Gaussowska funkcja gęstości prawdopodobieństwa.

Należy zwrócić uwagę na to, że – zgodnie z zależnością (7.6) – charakterystyka $g(n)$ pokazana na rys. 7.2 ma po lewej i prawej stronie „ogony” ciągnące się do $\pm\infty$. Oznacza to, że bardzo duże wartości chwilowe szumu są też możliwe, chociaż bardzo mało prawdopodobne. Pokazano to w tab. 7.1, gdzie podano obliczone prawdopodobieństwo tego, że wartość chwilowa szumu przekroczy daną wartość.

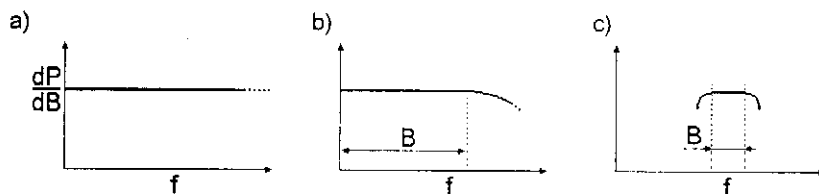
Tab. 7.1. Prawdopodobieństwo przekroczenia założonego poziomu przez wartość chwilową szumu gaussowskiego.

$ n > 0$	$p = 1,0$	$ n > 2,5 \cdot \sigma$	$p = 0,012$
$ n > 0,5 \cdot \sigma$	$p = 0,61$	$ n > 3,0 \cdot \sigma$	$p = 0,0027$
$ n > 1,0 \cdot \sigma$	$p = 0,31$	$ n > 3,5 \cdot \sigma$	$p = 0,0004$
$ n > 1,5 \cdot \sigma$	$p = 0,13$	$ n > 4,0 \cdot \sigma$	$p = 0,00005$
$ n > 2,0 \cdot \sigma$	$p = 0,045$	$ n > 5,0 \cdot \sigma$	$p = 0,0000006$

Dla szumu gaussowskiego $\overline{n(t)} = 0$, $\overline{n^2(t)} = \sigma^2$, zatem dyspersja σ ma sens wartości skutecznej (napięcia lub prądu) szumu. Jeżeli wartość skuteczna szumu wynosi np. 1 V, chwilowa wartość tego szumu w 69% czasu obserwacji nie przekroczy 1 V (tab. 7.1), ale w 4,5% czasu może przekroczyć 2 V, w 0,27% może przekroczyć 3 V itd. Jest to fakt o bardzo doniosłym znaczeniu, jeśli rozważa się wpływ szumu na transmisję sygnałów PCM, o czym będzie mowa dalej.

Szum gaussowski jest tzw. **szumem białym**, tzn. stanowi równomierną mieszaninę przebiegów o różnej częstotliwości („wszystkich” częstotliwości). Nazwa szum biały jest powszechnie stosowana przez analogię ze światłem białym, które jest równomierną mieszaniną różnych barw (częstotliwości fal).

Gęstość widmowa mocy szumu białego (tj. moc przypadająca na jednostkowe pasmo częstotliwości) jest stała; jej graficzne przedstawienie to linia pozioma (rys. 7.3a).



Rys. 7.3. Widma szumu białego: a) widmo idealne – teoretyczne, b) widmo praktyczne, c) widmo szumu białego pasmowego.

Moc szumu białego jest proporcjonalna
do szerokości pasma częstotliwości.

Często za szum biały uznaje się szum o płaskim przebiegu gęstości widmowej w ograniczonym zakresie częstotliwości, np. w pasmie sygnałów odbieranych przez odbiornik. Jest to wtedy **szum biały pasmowy** (rys. 7.3c). Szum biały pasmowy, występujący w pasmie B , można zapisać w postaci przydatnego wyrażenia [2]:

$$n_b(t) = n_c(t) \cdot \cos \omega_s t + n_q(t) \cdot \sin \omega_s t, \quad (7.7)$$

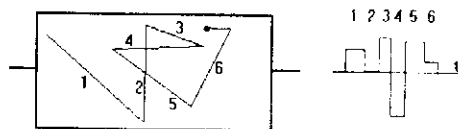
gdzie $n_c(t)$ i $n_q(t)$ (składowa synfazowa i kwadraturowa) są to wzajemnie niezależne gaussowskie przebiegi szumowe o pasmie ograniczonym do $B/2$ i jednakowych wartościach średniokwadratowych, ω_s zaś jest pulsacją środkową rozważanego pasma. Wartość średniokwadratowa przebiegów $\cos \omega_s t$ i $\sin \omega_s t$ jest równa 0,5, dlatego wartość średniokwadratowa szumu pasmowego jest równa tej wartości dla każdej ze składowych $n_c(t)$ i $n_q(t)$:

$$\overline{n_b^2(t)} = \overline{n_c^2(t)} = \overline{n_q^2(t)}. \quad (7.8)$$

7.2. Źródła szumu

Szum wewnętrzny, spotykany w aparaturze elektronicznej i telekomunikacyjnej, to szum termiczny, śrutowy, tzw. szum strukturalny i inne.

Szum termiczny (cieplny) pochodzi od nieustannego, chaotycznego ruchu, w jakim znajdują się elektrony występujące w przewodnikach prądu (rys. 7.4).



Rys. 7.4. Poglądowe przedstawienie bezładnego ruchu jednego elektronu w rezystorze i odpowiadającego mu przebiegu napięcia na rezystorze.

Każdy taki ruch jednego elektronu to „mikroprąd” o bardzo małym natężeniu i czasie trwania, jednak suma (geometryczna – z uwzględnieniem kierunku) takich prądów od bardzo dużej liczby elektronów (np. rzędu $10^{23}/\text{cm}^3$) powoduje powstanie przebiegu zauważalnego w skali makroskopowej. Na jakimkolwiek elemencie rezystancyjnym wywołuje to powstanie napięcia szumu, o przykładowym przebiegu pokazanym wcześniej na rys. 7.1.

Szum termiczny charakteryzowany jest najczęściej wartością średniokwadratową. Dana jest ona wzorem Johnsona [4]:

$$\overline{U_{sz}^2} = 4k \cdot T \cdot B \cdot R, \quad (7.9)$$

gdzie: k – stała Boltzmanna, $k = 1,38 \cdot 10^{-23}$ J/K, T – temperatura bezwzględna, B – szerokość pasma (w którym dokonywana jest obserwacja napięcia), R – rezystancja.

Interesujące jest to, że napięcie szumu zależy od szerokości pasma – np. szerokości pasma woltomierza lub oscyloskopu, za pomocą którego jest mierzone. Jest to prawidłowe, gdyż im szersze pasmo ma ten przyrząd, tym więcej częstotliwościowych składowych szumu on uwzględni, tym samym wskazywane napięcie będzie większe.

Czy oznacza to, że gdyby pasmo, np. woltomierza, było nieskończone, wskazywane napięcie byłoby też nieskończone? Teoretycznie tak, ale:

- nie ma przyrządów pomiarowych o nieskończonej szerokości pasmie,
- zjawisko szumu termicznego jest jednak ograniczone, ale do częstotliwości rzędu kilkuset gigaherców. Typowe przyrządy pomiarowe mają pasmo znacznie węższe, dlatego w praktycznych zastosowaniach nie uwzględnia się ograniczonego pasma tego szumu.

Szum termiczny jest szumem białym, tzn. jest mieszaniną przebiegów o różnych wartościach częstotliwości, z których żaden nie jest wyróżniony ani osłabiony.

Przykład 7.1

Jakie jest napięcie szumów termicznych obwodu wejściowego odbiornika telewizyjnego o parametrach $R = 75 \Omega$, $B = 6 \text{ MHz}$, $T = 300 \text{ K}$?

Obliczenia wg zależności (7.8) dają $\overline{U_{sz}^2} = 7,4 \cdot 10^{-12} \text{ V}^2$, tzn. że wartość skuteczna napięcia szumu wynosi ok. $2,7 \mu\text{V}$.

Przykład 7.2

Oscyloskopy charakteryzują się rezystancją wejściową $R = 1 \text{ M}\Omega$, która jest ustalana zwykle za pomocą rezystora o tej wartości włączonego równolegle do zacisków wejściowych przyrządu. Rezystor ten jest źródłem szumu termicznego, który po przejściu przez odpowiednie bloki oscyloskopu może być widoczny na ekranie i jednocześnie ograniczać czułość tego przyrządu. Jakiego napięcia szumu można oczekiwać, jeżeli pasmo oscyloskopu $B = 0\text{--}50 \text{ MHz}$?

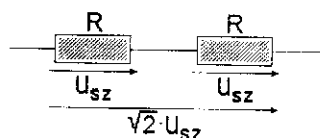
Obliczenia dla $T = 300 \text{ K}$ dają

$$\overline{U_{sz}^2} = 0,83 \cdot 10^{-6} \text{ V}^2.$$

czyli napięcie skuteczne szumu będzie wynosiło ok. $0,9 \text{ mV}$. Jest to dość duża wartość, wskazująca na to, że zjawisko szumu termicznego może ograniczać czułość szerokopasmowych oscyloskopów.

Przykład 7.3

Jeżeli jakiś rezystor charakteryzuje się napięciem skutecznym szumu termicznego równym U_{sz} , to jakie napięcie szumu wystąpi na dwóch takich rezystorach połączonych szeregowo (rys. 7.5)?



Rys. 7.5. Ilustracja sumowania napięć skutecznych szumu.

Intuicyjnie mogłoby się wydawać, że napięcie to powinno być 2 razy większe, ale będzie inaczej. Napięcie skuteczne szumu będzie tylko $\sqrt{2}$ razy większe, bo tak wynika ze wzoru (7.8). Gdyby napięcia skuteczne szumu dodawały się proporcjonalnie, oznaczałoby to, że szum w jednym i drugim rezystorze jest w pełni skorelowany (ma identyczny przebieg czasowy), a tymczasem jest, można powiedzieć, w pełni nieskorelowany. Napięcia szumu pochodzą od chaotycznego ruchu elektronów w dwóch oddzielnych rezystorach i oczywiste jest, że ruchy muszą być wzajemnie niezależne.

Warto podkreślić, że szum termiczny występuje niezależnie od tego, czy przez element rezystancyjny przepływa prąd czy nie.

Szum śrutowy występuje tylko podczas przepływu prądu przez obwód. Bierze się on stąd, że prąd, nawet stały w czasie, to przepływ dyskretnych ładunków elementarnych, a one są niepodzielne. Można przeprowadzić następujące rozumowanie.

Prąd o natężeniu 1 A odpowiada wartości $6,25 \cdot 10^{18}$ elektronów na sekundę, prąd 1 pA odpowiada wartości $6,25 \cdot 10^6$ elektronów na sekundę albo 6,25 elektronów na mikrosekundę... – ostatnia wartość nie ma oczywiście sensu. Oznacza to, że w kolejnych mikrosekundach przez przekrój poprzeczny przewodnika przechodzić będzie przykładowo 6, 7, 6, 6, 7, 6... elektronów, zatem chwilowe natężenie prądu w rozważanej skali czasowej nie będzie stałe.

Wskazuje to na istotę szumu śrutowego – pochodzi on z nierównomierności chwilowego natężenia prądu, spowodowanej tym, że prąd ten to przepływ dyskretnych ładunków, a nie „przelewanie się” jakiegoś ciągłego nośnika tego ładunku.

Jak widać z tego rozważania, szum śrutowy może być problemem, gdy natężenie prądu jest małe. Rzeczywiście – wartość średniokwadratowa prądu szumu śrutowego wyraża się wzorem [4]:

$$\overline{I_{sz}^2} = 2eBI_0, \quad (7.10)$$

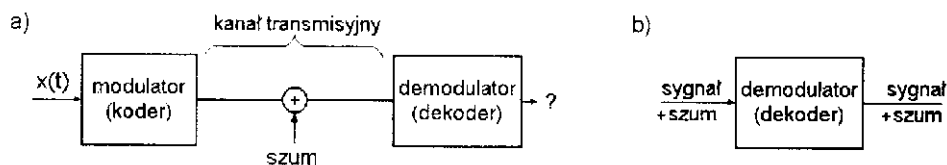
gdzie e jest ładunkiem elementarnym, a I_0 jest wartością średnią prądu. Powoduje to, że względny udział tego szumu w prądzie wzrasta ze spadkiem tego prądu:

$$\frac{\overline{I_{sz}}}{I_0} = \sqrt{\frac{2eB}{I_0}}. \quad (7.11)$$

Szum śrutowy też jest szumem białym, występującym w szerokim, ale skończonym pasmie częstotliwości, sięgającym kilku gigaherców [4].

Transmisję sygnałów w obecności zakłócającego ją szumu rozważa się w sposób następujący. Sygnał oryginalny zostaje zmodulowany lub zakodowany, następnie jest transmitowany przez kanał telekomunikacyjny (przewód współosiowy, światłowód, wolną przestrzeń, falowód...) i przede wszystkim na tym etapie dochodzi do zakłócenia go szumem (rys. 7.6a). W końcu sygnał dochodzi do odbiornika, gdzie po przejściu przez jego stopnie wejściowe trafia na wejście demodulatora lub dekodera. Na jego wyjściu pojawia się sygnał oryginalny, ale w pewien sposób zniekształcony, gdyż sygnałem wejściowym demodulatora jest sygnał zakłócony szumem. Traktując zniekształcenia sygnału wyjściowego jako szum, można przyjąć, że w procesie demodulacji (dekodowania) szum wejściowy demodulatora zostaje jakoś przetworzony na szum wyjściowy.

Wynika z tego, że o odporności szumowej całego systemu będzie decydować zachowanie się demodulatora. Najlepiej, aby demodulator w ogóle nie „przepuszczał” szumu wejściowego na swoje wyjście.



Rys. 7.6. Sposób uwzględniania wpływu szumu na transmisję sygnałów:
a) postać ogólna, b) rozważania ograniczone tylko do demodulatora.

W celu możliwości porównania różnych systemów transmisji pod względem tego, w jakim stopniu szum zawarty w odebranym sygnale przechodzi na wyjście dekodera (demodulatora), wprowadzono pojęcie charakterystyki szumowej modulacji jako zależności

$$\left(\frac{S}{N} \right)_{wy} = f \left(\frac{S}{N} \right)_{we}, \quad (7.12)$$

przy czym dotyczy to demodulatora lub dekodera w danym systemie (rys. 7.6b).

7.3. Charakterystyka szumowa modulacji AM

Charakterystykę tę można dość łatwo wyprowadzić teoretycznie, co zostanie pokazane następująco. Sygnał wejściowy demodulatora dany jest zależnością (3.3):

$$a_{we}(t) = A_o[1 + kx(t)] \cdot \cos \Omega_o t. \quad (7.13)$$

Moc tego sygnału to średniokwadratowa wartość z wielkości (7.13), czyli

$$S_{we} = \frac{A_o^2}{2} [1 + k^2 \overline{x(t)^2}]. \quad (7.14)$$

Na podstawie zależności (7.7) szum wejściowy (pasmowy) można zapisać jako

$$n_{we}(t) = n_c(t) \cdot \cos \Omega_o t + n_q(t) \cdot \sin \Omega_o t, \quad (7.15)$$

a moc tego szumu to

$$N_{we} = \overline{n_{we}^2(t)}. \quad (7.16)$$

Aby rozważyć sytuację na wyjściu demodulatora, założmy, że jest to demodulator synchroniczny. Sygnał wyjściowy jest otrzymywany w wyniku pomnożenia przez falę nośną (rozdz. 3.). Na podstawie wzoru (3.22) można napisać, że

$$S_{wy} = \frac{A_o^2}{4} k^2 \overline{x(t)^2}. \quad (7.17)$$

Szum wejściowy pomnożony przez falę nośną przekształci się do postaci

$$n_i(t) \cdot \cos \Omega_o t = n_c(t) \cdot \cos^2 \Omega_o t + n_q(t) \cdot \sin \Omega_o t \cdot \cos \Omega_o t.$$

Po filtracji w filtrze wyjściowym uzyska się (uwzględniając, że $\cos^2 \Omega_o t = 0,5 + 0,5 \cdot \cos 2\Omega_o t$):

$$n_{wy}(t) = n_c(t)/2. \quad (7.18)$$

Moc tego szumu to

$$N_{wy} = \overline{n_c^2(t)}/4 = \overline{n_{we}^2(t)}/4 \quad (7.19)$$

– na podstawie zależności (7.8). Ostatecznie

$$\frac{S_{wy}}{N_{wy}} = \frac{A_o^2}{\overline{n_{we}^2(t)}} k^2 \overline{x(t)^2} \quad (7.20)$$

oraz

$$\frac{S_{we}}{N_{we}} = \frac{A_o^2}{2 \cdot \overline{n_{we}^2(t)}} [1 + k^2 \overline{x(t)^2}]. \quad (7.21)$$

Zatem

$$\frac{S_{wy}}{N_{wy}} = \frac{S_{we}}{N_{we}} \cdot \frac{2k^2 \overline{x(t)^2}}{1 + k^2 \overline{x(t)^2}}. \quad (7.22)$$

Jeżeli przyjąć, jak to było w rozdz. 3., że $x(t) = X_m \cdot \cos \omega t$, to $X_m k = m$ i wzór (7.22) przybierze postać

$$\frac{S_{wy}}{N_{wy}} = \frac{S_{we}}{N_{we}} \cdot \frac{m^2}{1 + m^2/2}, \quad (7.23)$$

ponieważ $\overline{\cos^2 \omega t} = 1/2$.

W przypadku modulacji AM z falą nośną jakość sygnału po demodulacji, mierzona stosunkiem S/N , jest zawsze gorsza od jakości sygnału przed demodulacją. Wytlumaczenie tego jest proste – przed demodulacją o mocy sygnału decyduje przede wszystkim fala nośna – wiadomo, że jej moc stanowi co najmniej 2/3 mocy sygnału. W trakcie demodulacji moc ta „zamienia się” na moc składowej stałej, która nie ma wpływu na moc sygnału wyjściowego.

Czytelnikowi pozostawia się podobne, ale już znacznie prostsze, wyprowadzenie charakterystyk szumowych modulacji AM DSB-SC i SSB-SC.

Wszystkie te zależności zestawiono w tab. 7.2.

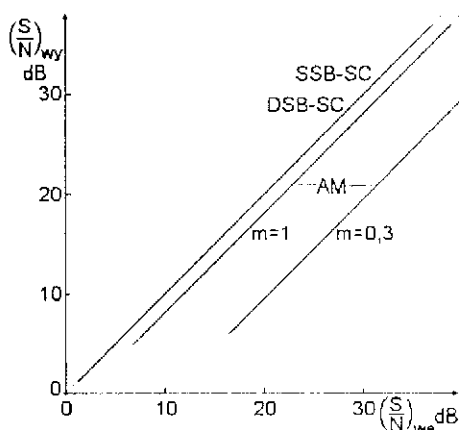
Tab. 7.2. Zestawienie zależności opisujących charakterystyki szumowe modulacji AM.

Rodzaj modulacji	Przypadek ogólny $S_{we} \gg N_{we}$	Ustalona widmowa gęstość mocy szumu wejściowego
AM SSB-SC	$\left(\frac{S}{N}\right)_{wy} = \left(\frac{S}{N}\right)_{we}$	$\left(\frac{S}{N}\right)_{wy} = \frac{S_{we}}{N_{weo}}$
AM DSB-SC	$\left(\frac{S}{N}\right)_{wy} = 2 \cdot \left(\frac{S}{N}\right)_{we}$	$\left(\frac{S}{N}\right)_{wy} = \frac{S_{we}}{N_{weo}}$
AM	$\left(\frac{S}{N}\right)_{wy} = \frac{2m^2}{2 + m^2} \left(\frac{S}{N}\right)_{we}$	$\left(\frac{S}{N}\right)_{wy} = \frac{m^2}{2 + m^2} \cdot \frac{S_{we}}{N_{weo}}$

W tabeli zawarto charakterystyki szumowe dotyczące dwóch przypadków – przypadku ogólnego, przy dowolnej mocy szumu wejściowego i sygnału, oraz przypadku, gdy moc szumu wejściowego jest proporcjonalna do szerokości pasma sygnału (odpowiada to ustaleniu stałej widmowej gęstości mocy szumu wejściowego). Przez N_{weo} oznaczono moc szumu wejściowego dla przypadku SSB (najmniejsza szerokość pasma); dla DSB i AM ta moc będzie równa $2N_{weo}$. Drugi przypadek odpowiada bardziej rzeczywistym warunkom pracy systemów telekomunikacyjnych, jest często uwzględniany w literaturze i dlatego go tu podano.

W wyniku demodulacji sygnału AM SSB-SC nie następuje ani polepszenie, ani pogorszenie stosunku S/N , dlatego ten rodzaj modulacji przyjmowany jest jako odniesienie w rozważaniach odporności szumowej różnych systemów.

Charakterystyki szumowe modulacji amplitudy (drugi przypadek) pokazano na rys. 7.7.



Rys. 7.7. Charakterystyki szumowe modulacji amplitudy.

7.4. Charakterystyka szumowa modulacji FM

Odpowiednie rozważania prowadzące do uzyskania charakterystyki szumowej są bardziej skomplikowane i nie będą tu przytaczane. Wyniki takich rozważań, przedstawionych np. w pracy [2], zamieszczono w tab. 7.3.

Tab. 7.3. Zależności opisujące charakterystykę szumową modulacji FM.

Przypadek ogólny $S_{we} \gg N_{we}$	Ustalona widmowa gęstość mocy szumów wejściowych
$\left(\frac{S}{N}\right)_{wy} = 3m_f^2(m_f + 2)\gamma^2 \left(\frac{S}{N}\right)_{we}$	$\left(\frac{S}{N}\right)_{wy} = \frac{3}{2}m_f^2 \cdot \gamma^2 \cdot \frac{S_{we}}{N_{we0}}$

Wielkość γ , występująca w podanych zależnościach, wynosi

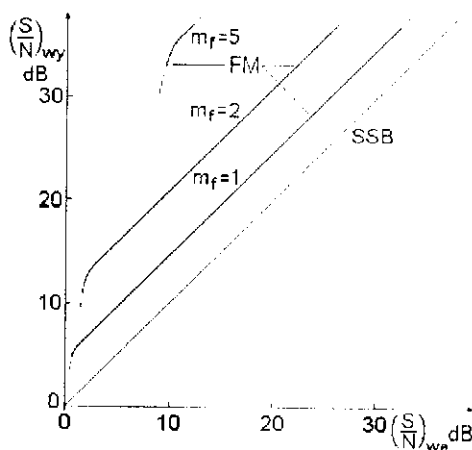
$$\gamma = f_m / f_{mmax} \quad (7.24)$$

dlatego wyrażenie typu $m_f \gamma$ jest po prostu równe $\frac{\Delta F_o}{f_m} \cdot \frac{f_m}{f_{mmax}} = \frac{\Delta F_o}{f_{mmax}}$; często jest ono utożsamiane z wielkością m_f , co jednak nie jest zbyt dokładne.

Podobnie jak poprzednio, podano zależności dwu przypadków – dla dowolnej mocy szumu wejściowego oraz dla ustalonej widmowej gęstości mocy tego szumu; w drugim przypadku moc szumu wejściowego $N_{we} = 2(m_f + 2)N_{we0}$, bo $2(m_f + 2)$ razy szerokość pasma sygnału FM jest większa od szerokości pasma sygnału AM SSB-SC.

Jak wynika z przytoczonych zależności, stosunek S/N na wyjściu demodulatora FM może być wielokrotnie większy niż ten stosunek na jego wejściu, o ile tylko indeks modulacji FM jest duży. Szerokopasmowa modulacja FM cechuje się więc wyraźną odpornością na zakłócenia szumowe; stanowi to jej zaletę, warunkującą jej częste stosowanie.

Charakterystyki szumowe modulacji FM pokazano na rys. 7.8. Załamywanie się tych charakterystyk przy małych stosunkach $(S/N)_{we}$ oznacza, że przy demodulacji FM występuje tzw. **efekt progowy** – aby wystąpiła poprawa jakości sygnału w wyniku demodulacji (aby podane w tab. 7.3 wzory były prawdziwe), sygnał wejściowy musi mieć odpowiednią dużą jakość.



Rys. 7.8. Charakterystyki szumowe modulacji FM; dla porównania podano też charakterystykę szumową modulacji SSB.

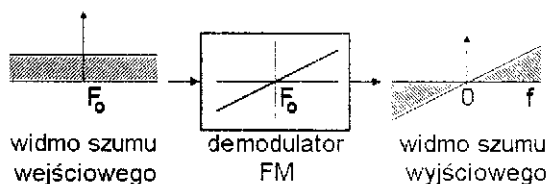
Ze względu na duże podobieństwa modulacji częstotliwości i fazy, właściwości szumowe tej ostatniej są zbliżone do przedstawionych właściwości modulacji FM. Informacje na ten temat można znaleźć np. w pracy [2]. Nie będą one tu omawiane, gdyż modulacja PM – przynajmniej w wersji analogowej – nie jest często stosowana.

Dla pełniejszej oceny właściwości transmisji FM warto rozważyć, jak pracuje demodulator częstotliwości przy szumowym sygnale wejściowym. Działanie idealnego demodulatora można zapisać jako

$$u_{wy}(t) = \alpha \cdot [F(t) - F_0], \quad (7.25)$$

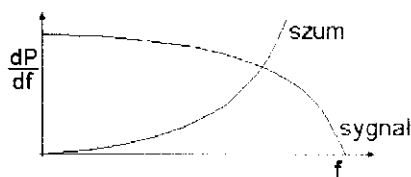
gdzie α jest pewnym współczynnikiem proporcjonalności, $F(t)$ jest częstotliwością chwilową, a F_0 częstotliwością nośną sygnału wejściowego.

Demodulator FM jest więc przetwornikiem częstotliwości na napięcie – także dla częstotliwości zawartych w wejściowym przebiegu szumowym. Jak wiadomo, w przebiegu tym zawarte są „wszystkie” częstotliwości i one są przetwarzane na napięcie wyjściowe, także o charakterze szumu – o tym większej wartości, im bardziej częstotliwość chwilowa szumu wejściowego różni się od wielkości F_0 . Podczas gdy szum wejściowy ma widmo płaskie, szum wyjściowy ma widmo „podbite”, w którym zawartość składowych rośnie liniowo ze wzrostem częstotliwości (rys. 7.9).



Rys. 7.9. Przetwarzanie szumu wejściowego w demodulatorze FM (pokazano widmo amplitudowe szumu).

Z tego względu na wyjściu demodulatora FM występuje niekorzystne zjawisko – moc szumów wyjściowych rośnie ze wzrostem częstotliwości, natomiast moc sygnału zdemodulowanego z reguły maleje ze wzrostem tej częstotliwości (rys. 7.10). Należy zatem się spodziewać zmniejszenia stosunku S/N na wyjściu demodulatora FM w zakresie wyższych wartości częstotliwości modulującej.

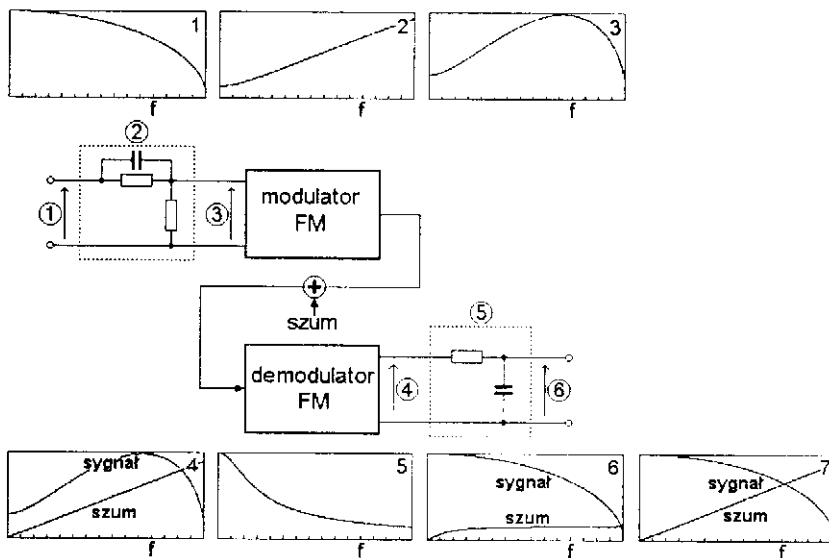


Rys. 7.10. Gęstość widmowa mocy typowego sygnału i szumu na wyjściu demodulatora FM.

Okazuje się, że za pomocą bardzo prostych środków technicznych można tę niedogodność systemów FM znacznie zredukować. Odbywa się to przez zastosowanie tzw. preemfazy po stronie nadawczej i deemfazy po stronie odbiorczej. Schemat tego pokazano na rys. 7.11, gdzie przedstawiono blokowo tor transmisji FM oraz widma i charakterystyki częstotliwościowe niektórych bloków.

Oryginalny sygnał informacyjny (o widmie oznaczonym 1), zanim dojdzie do modulatora FM, przechodzi przez układ preemfazy ②. Jest to układ „podbi-

jający” składowe o wyższej częstotliwości w sygnale modulującym – 2. W wyniku tego widmo sygnału dochodzącego do modulatora jest zdeformowane – składowe sygnału o dużej częstotliwości mają nienaturalnie wysoki poziom – 3. Sygnał o takim widmie jest transmitowany za pomocą modulacji FM do odbiornika, a po drodze dodaje się do niego szum. W odborniku demodulator FM odtwarza taki sygnał, jaki był doprowadzony do modulatora po stronie nadawczej, a więc o zdeformowanym widmie; do tego dodaje jeszcze szum o omówionych wyżej właściwościach, tzn. liniowo rosnący ze wzrostem częstotliwości – 4. Te wielkości trafiają do układu deemfazy ⑤, o transmitancji – 5, będącej odwrotnością charakterystyki układu preemfazy. W wyniku tego sygnał informacyjny odzyskuje swe pierwotne widmo, a szum wyjściowy demodulatora zostaje znacznie stłumiony – 6. Gdyby w systemie nie zastosowano układów preemfazy i deemfazy, relacja między sygnałem i szumem na wyjściu demodulatora wyglądałaby tak, jak pokazano na widmie oznaczonym 7, a więc byłaby bardziej niekorzystna.



Rys. 7.11. Istota preemfazy i deemfazy w systemie FM. Pokazano widma sygnałów w odpowiednich miejscach systemu oraz charakterystyki częstotliwościowe bloków preemfazy i deemfazy.

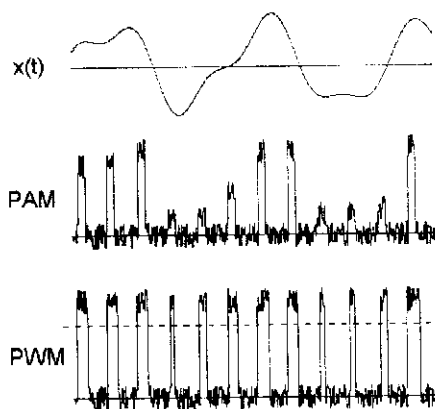
Preemfaza i deemfaza stosowane są powszechnie w radiofonii UKF FM, w systemach fonii TV, systemach satelitarnych itd. Dla radiofonii UKF FM, przy przyjętych tam parametrach preemfazy, zysk w stosunku S/N wynosi ok. 11 dB [3] w porównaniu z systemem bez tych układów. Jest to więc poważny zysk w jakości i warto jeszcze raz zaznaczyć, że uzyskany dzięki bardzo proste-

mu rozwiązaniu (3 elementy biernie po stronie nadawczej i 2 po stronie odbiorczej).

7.5. Odporność szumowa modulacji impulsowych

Modulacje impulsowe typu PAM, PWM, PPM stosowane są rzadko do transmisji informacji na większą odległość, dlatego tutaj przedstawione zostaną tylko ogólne uwagi dotyczące wpływu szumu na takie sygnały impulsowe.

Najbardziej podatna na zakłócenie szumem jest modulacja PAM, bo szum ma charakter addytywny i wyraźnie zmienia amplitudę transmitowanych impulsów, a przecież w takiej modulacji ich amplituda jest miarą sygnału informacyjnego (rys. 7.12).



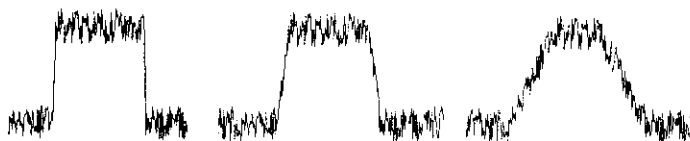
Rys. 7.12. Wpływ addytywnego szumu na sygnały PAM i PWM (o znikomym czasie narastania i opadania impulsów).

Modulacja PWM jest bardziej odporna, gdyż addytywny szum nie zmienia (prawie) szerokości impulsów, natomiast amplituda zakłóconych impulsów może zostać ograniczona (pokazano to linią przerywaną). Jest to dozwolone, gdyż w tym systemie amplituda nie przenosi informacji.

Podobnie odporna jest modulacja PPM, gdyż w jej przypadku ani amplituda, ani szerokość impulsów nie przenoszą informacji (mogą one podlegać regeneracji), natomiast prawdopodobieństwo tego, że pod wpływem szumu nastąpi jakby przesunięcie poziome impulsów, jest niewielkie.

Dokładniej przedstawiono to na rys. 7.13. Przedstawiono na nim trzy impulsy o różnej stromości zboczy, zakłócone takim samym szumem addytywnym. Widoczne jest, że im większa stromość zboczy, czyli mniejszy czas narastania i opadania impulsu, tym słabszy jest wpływ szumu na te zbocza, czyli na szerokość lub położenie impulsu.

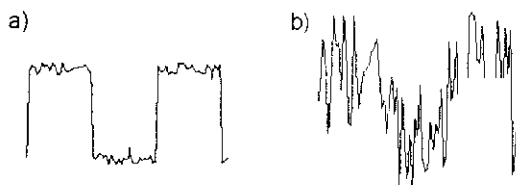
Im impulsy mają bardziej strome zbocza, tym szersze pasmo częstotliwości jest potrzebne do ich przesłania (porównaj rys. 5.8). Zatem im szersze pasmo częstotliwości zajmuje transmisja impulsowa, tym bardziej jest ona odporna na zakłócenia szumowe; jest to wniosek taki jak w przypadku szerokopasmowej modulacji FM.



Rys. 7.13. Ilustracja wpływu zakłóceń szumowych na zbocza impulsu przy różnej stromości tych zboczy.

7.6. Odporność szumowa sygnałów PCM

Jak już zaznaczono w rozdz. 6., sygnały PCM są z samej swojej natury dość odporne na zakłócenia szumowe. Wynika to z odmienności postaci sygnału PCM od szumu i zakłóceń analogowych, przez co można dość łatwo odróżnić go od szumu i przeprowadzać jego pełną regenerację. Drugim czynnikiem, który powoduje odporność tych sygnałów, jest możliwość dodatkowego zabezpieczenia transmisji przed błędami metodą określonych algorytmów kodowania i obróbki sygnału.

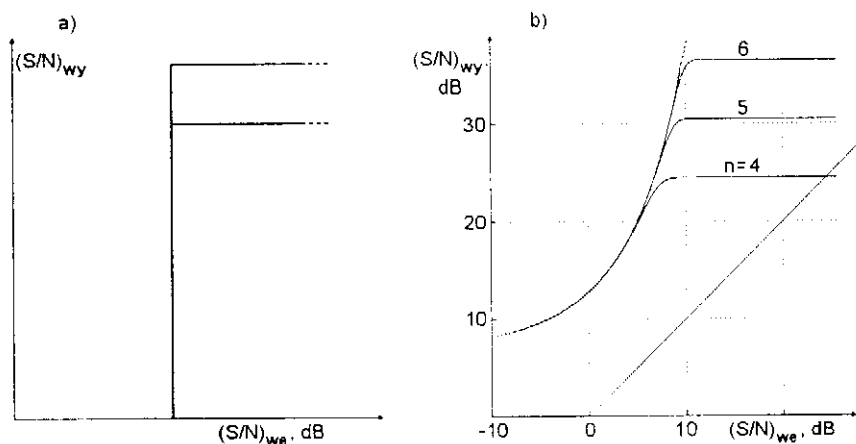


Rys. 7.14. Przykłady sygnałów PCM zakłóconych szumem: a) w stopniu nieutrudniającym dekodowania. b) w stopniu uniemożliwiającym dekodowanie.

Przedstawiając wstępnie zagadnienie wpływu szumu na transmisję PCM, można stwierdzić, że gdy szum jest słaby względem sygnału (rys. 7.14a), odtworzenie tego ostatniego jest łatwe; gdy szum jest dostatecznie silny (rys. 7.14b) – sygnału nie da się poprawnie zdekodować. W tej sytuacji uproszczona charakterystyka szumowa modulacji PCM mogłaby mieć postać przedstawioną na rys. 7.15a.

Silny szum oznacza małą wartość stosunku $(S/N)_{we}$ i wtedy sygnału użytecznego na wyjściu dekodera po prostu nie ma. Gdy szum jest słaby, $(S/N)_{we}$ ma dużą wartość, sygnał jest poprawnie dekodowany, ale zawsze jest „skażony”

szumem kwantyzacji, dlatego $(S/N)_{wy}$ ma skończoną wartość, zależną od tego, z jaką rozdzielczością sygnał był skwantowany i zakodowany po stronie nadawczej.



Rys. 7.15. Charakterystyka szumowa modulacji PCM:
a) idealizowana, b) przykładowa charakterystyka rzeczywista.

W rzeczywistości charakterystyka szumowa modulacji PCM ma postać zbliżoną do pokazanej na rys. 7.15b. Jest to skutkiem tego, że szum zakłócający zawiera przebiegi o określonym rozkładzie amplitud chwilowych. Jak wiadomo, wszystkie wartości amplitud są możliwe, ale im większe, tym z mniejszym prawdopodobieństwem. Przy silnym szumie w najgorszym przypadku może być zakłócona (tj. odebrana błędnie) połowa bitów sygnału. W wyniku tego odtworzony sygnał analogowy będzie mocno zniekształcony, ale jakoś będzie obecny (pomija się tu konieczność utrzymania synchronizacji, niezbędnej w procesie dekodowania sygnału PCM). Zniekształcenia te nie są formalnie szumem, ale można je tak potraktować, gdyż tak samo utrudniają odbiór sygnału jak szum, który byłby zawarty w tym sygnale.

W miarę spadku poziomu szumu jego duże wartości chwilowe stają się coraz mniej prawdopodobne, maleje prawdopodobieństwo zakłócenia bitów sygnału, sygnał jest odtwarzany z coraz mniejszymi zniekształceniami i wartość $(S/N)_{wy}$ wzrasta, ale tylko do momentu, kiedy zniekształcenia te zaczną być maskowane przez szum kwantyzacji. Im mniejszy jest szum kwantyzacji (zależy on od rozdzielczości przetwornika A/C), tym później wystąpi to maskowanie zakłóceń, czyli przy większej wartości wyrażenia $(S/N)_{we}$.

Jak pokazano w rozdz. 6., ze wzrostem rozdzielczości przetwornika A/C o 1 b moc szumu kwantyzacji maleje o 6 dB (4 razy), dlatego kolejne linie poziome odpowiadające różnym wartościom rozdzielczości przetwornika są rozdzielone o 6 dB.

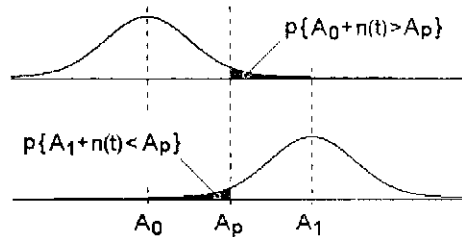
Linia ukośna umieszczona na rys. 7.15b rozdziela obszary, w których $(S/N)_{wy}$ jest większy niż $(S/N)_{we}$ lub odwrotnie. Widać, że zasadniczo charakterystyka szumowa modulacji PCM przebiega powyżej tej linii, tzn. w wyniku dekodowania sygnału PCM następuje poprawa S/N .

Podczas transmisji „czystych” danych cyfrowych, niebędących przetworzonym sygnałem analogowym, operowanie szumem kwantyzacji nie ma sensu – operuje się wtedy prawdopodobieństwem błędu odbioru, często nazywanym **stopą błędów** i powszechnie oznaczanym jako **BER** (*Bit Error Ratio*). Dalej pokazane zostaną przykładowe rozważania teoretyczne, prowadzące do wyznaczenia tego prawdopodobieństwa.

Przy dwuwartościowej (binarnej) transmisji cyfrowej, kiedy nadawanym symbolom 0 i 1 przypisuje się dwie wartości sygnału A_0 i A_1 , prawdopodobieństwo błędu można wyrazić następująco:

$$p_e = p(0) \cdot p\{A_0 + n(t) > A_p\} + p(1) \cdot p\{A_1 + n(t) < A_p\}, \quad (7.26)$$

gdzie $p(0)$ i $p(1)$ są prawdopodobieństwem wystąpienia symboli 0 i 1, A_p jest progiem decyzyjnym, powyżej którego przyjmuje się, że odebrany został symbol 1, a poniżej – 0; $n(t)$ jest przebiegiem szumowym. Można to zilustrować za pomocą rys. 7.16, na którym zacienione pola odpowiadają prawdopodobieństwu błędnego odbioru odpowiednio „zera” i „jedynek”.



Rys. 7.16. Ilustracja prawdopodobieństwa błędnego odbioru symboli 0 i 1 w obecności szumu gaussowskiego.

Prawdopodobieństwa błędów są odpowiednimi całkami z przesuniętej o A_0 lub A_1 funkcji $g(n)$ danej zależnością (7.6):

$$p\{A_0 + n(t) > A_p\} = \int_{A_p}^{\infty} g(n - A_0) \cdot dn$$

oraz

$$p\{A_1 + n(t) < A_p\} = \int_{-\infty}^{A_p} g(n - A_1) \cdot dn. \quad (7.27)$$

Po odpowiedniej zamianie zmiennych $y = n - A_0$ lub $y = n - A_1$ i przy jednakowych prawdopodobieństwach $p(0) = p(1) = 0,5$ uzyskuje się

$$p_e = \frac{1}{2} \left[\int_{A_p - A_0}^{\infty} g(y) \cdot dy + \int_{-\infty}^{A_p - A_1} g(y) \cdot dy \right]. \quad (7.28)$$

Należy tu rozróżnić, w jaki sposób dane cyfrowe są transmitowane – czy kodem dwubiegunowym czy jednobiegunowym, bo od tego zależy średnia moc sygnału.

Dla kodu jednobiegunowego przyjmuje się $A_0 = 0$, $A_1 = A$, $A_p = A/2$, $S = A^2/2$, wtedy

$$p_e = \frac{1}{2} \left[\int_{A/2}^{\infty} g(y) \cdot dy + \int_{-\infty}^{-A/2} g(y) \cdot dy \right]. \quad (7.29)$$

Ze względu na symetrię przebiegu funkcji gęstości prawdopodobieństwa obydwie całki w ostatnim wyrażeniu są sobie równe, dlatego

$$p_e = \int_{A/2}^{\infty} g(y) \cdot dy = \frac{1}{\sqrt{2\pi N}} \int_{A/2}^{\infty} e^{-\frac{y^2}{2N}} \cdot dy. \quad (7.30)$$

Po wprowadzeniu jeszcze nowej zmiennej $u = y/\sqrt{2N}$ dolna granica całkowania przybierze wartość $u_{\min} = A/2\sqrt{2N}$, a całka (7.30) będzie wynosić

$$p_e = \int_{A/2}^{\infty} g(y) \cdot dy = \frac{1}{\sqrt{\pi}} \int_{A/2\sqrt{2N}}^{\infty} e^{-u^2} du. \quad (7.31)$$

Dla kodu jednobiegunowego wyrażenie $A/2\sqrt{2N}$ można zapisać jako $\sqrt{S/N}/2 = \sqrt{SNR}/2$.

Ostatecznie prawdopodobieństwo błędu można wyrazić jako funkcję stosunku mocy sygnału do mocy szumu

$$p_e = \frac{1}{\sqrt{\pi}} \int_{\sqrt{\text{SNR}}/2}^{\infty} e^{-u^2} du. \quad (7.32)$$

Wykorzystując definicję tzw. komplementarnej funkcji błędu [1, 2, 5]

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du. \quad (7.33)$$

można zapisać, że prawdopodobieństwo błędu

$$p_e = \frac{1}{2} \text{erfc} \left(\frac{\sqrt{\text{SNR}}}{2} \right). \quad (7.34)$$

Dla kodu dwubiegunowego o takiej samej „rozpiętości” sygnału $A_0 = -A/2$, $A_1 = A/2$, $A_p = 0$, $S = A^2/4$ oraz

$$p_e = \frac{1}{2} \cdot \left[\int_{A/2}^{\infty} g(y) \cdot dy + \int_{-\infty}^{-A/2} g(y) \cdot dy \right]. \quad (7.35)$$

Ze względu na równość obydwu całek

$$p_e = \int_{A/2}^{\infty} g(y) \cdot dy = \frac{1}{\sqrt{2\pi N}} \int_{A/2}^{\infty} e^{-\frac{y^2}{2N}} \cdot dy. \quad (7.36)$$

Po analogicznej jak poprzednio zmianie zmiennej całkowania otrzyma się

$$p_e = \int_{A/2}^{\infty} g(y) \cdot dy = \frac{1}{\sqrt{\pi}} \int_{A/2\sqrt{2N}}^{\infty} e^{-u^2} du, \quad (7.37)$$

a więc tyle samo co poprzednio (7.31). Jednak tym razem moc sygnału $S = A^2/4$ (jest mniejsza), dlatego dolną granicę całkowania można wyrazić jako

$$\frac{A}{2\sqrt{2N}} = \frac{\sqrt{A^2/4}}{\sqrt{N}} \cdot \frac{1}{\sqrt{2}} = \frac{\sqrt{\text{SNR}}}{\sqrt{2}}. \quad (7.38)$$

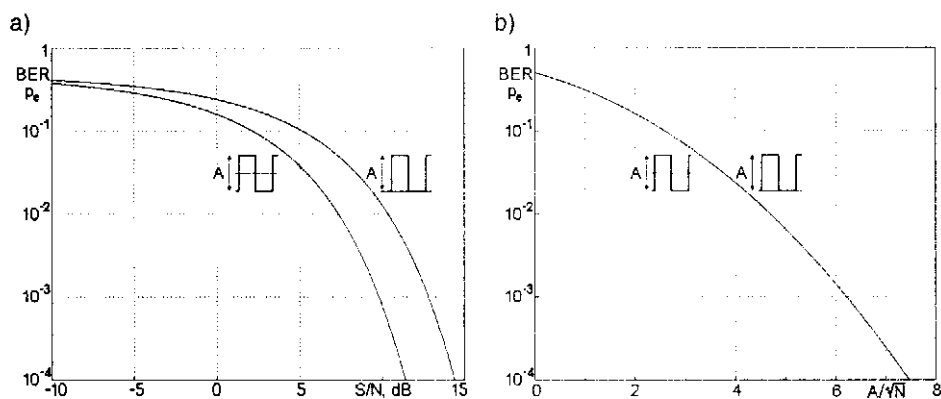
Ostatecznie

$$p_e = \frac{1}{\sqrt{\pi}} \int_{\sqrt{SNR}/\sqrt{2}}^{\infty} e^{-u^2} du, \quad (7.39)$$

czyli inaczej

$$p_e = \frac{1}{2} \operatorname{erfc} \left(\frac{\sqrt{SNR}}{\sqrt{2}} \right). \quad (7.40)$$

Rezultaty tych obliczeń przedstawiono na rys. 7.17a i b. Wynika z nich, że kod dwubiegunowy odznacza się lepszą odpornością szumową – tę samą wartość stopy błędów uzyskuje się przy stosunku S/N mniejszym o 3 dB niż dla kodu jednobiegunowego.



Rys. 7.17. Prawdopodobieństwo błędu binarnej transmisji cyfrowej w pasmie naturalnym dla kodu bipolarnego i dla kodu unipolarnego: a) w funkcji stosunku mocy sygnału do mocy szumu.

b) w funkcji stosunku międzyszczytowej amplitudy sygnału do wartości skutecznej szumu.

Jak to wytłumaczyć? Jak wynika z obliczeń, przy danym poziomie szumów stopa błędów zależy tylko od „rozpiętości” amplitud sygnału – od tego, jak oddalone są od siebie poziomy sygnału A_0 i A_1 . Kody dwubiegunowy i jednobiegunowy o takiej samej rozpiętości dają taką samą stopę błędów (zależności (7.31) i (7.37), ale kod dwubiegunowy ma 2 razy mniejszą moc – to znaczy, że kod ten zapewnia uzyskanie tej samej stopy błędów przy stosunku S/N mniejszym o 3 dB (obydwie krzywe na rys. 7.17a są identyczne, tylko przesunięte w poziomie o 3 dB).

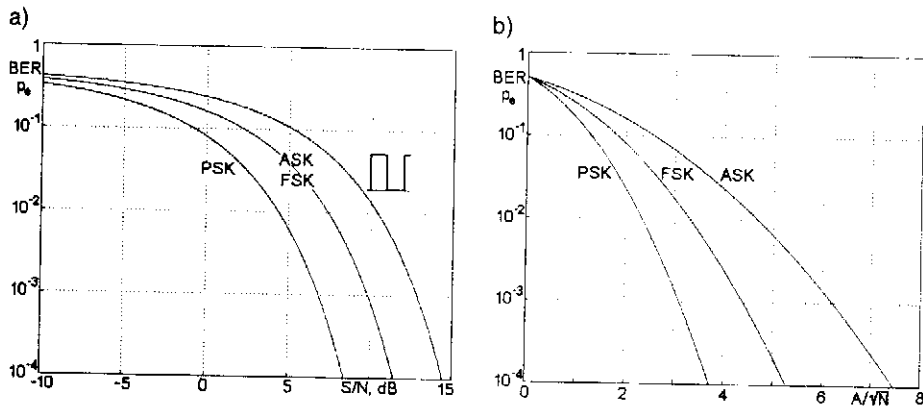
Na rysunku 7.17b pokazano zależność stopy błędów dla obydwu kodów w funkcji międzyszczytowej wartości sygnału kodowego (odniesionej do warto-

ści skutecznej szumu). Jest oczywiste, że w funkcji takiej zmiennej prawdopodobieństwo błędu dla obydwu kodów jest takie samo.

Najbardziej niekorzystnym przypadkiem odbioru jest sytuacja, gdy $p_e = 0,5$ i do tej wartości dąży prawdopodobieństwo błędu, gdy jakość sygnału znacznie się pogarsza. Gdyby prawdopodobieństwo to dążyło do jedności, byłoby to bardzo korzystne, bo proste zanegowanie odebranych bitów dawałoby informację bezbłędną, natomiast przy $p_e = 0,5$ negowanie odebranych bitów jest zupełnie bezcelowe.

Czy z punktu widzenia odporności szumowej „opłacalne” jest stosowanie poznanych wcześniej modulacji typu ASK, FSK lub PSK?

Wyniki odpowiednich obliczeń, opartych na pracy [2], przedstawiono w postaci wykresów na rys. 7.18a i b. Założono tu dla prostoty niezmiennosć progu decyzyjnego A_p , podczas gdy – jak wskazują dokładne rozważania – w celu optymalizacji systemu powinien on być dobierany w zależności od S/N .



Rys. 7.18. Prawdopodobieństwo błędu podczas pasmowej transmisji cyfrowej przy modulacjach ASK, FSK i PSK: a) w funkcji stosunku mocy sygnału do mocy szumu, b) w funkcji stosunku amplitudy sygnału do wartości skutecznej szumu.

Z rysunku 7.18a wynika, że sygnały ASK i FSK zapewniają taką samą stopę błędów transmisji jak w przypadku naturalnego kodu jednobiegunowego (pokazanego tu dla porównania), jednak przy mocy sygnału mniejszej o 3 dB, natomiast moc sygnału PSK w tych samych warunkach może być aż o 6 dB mniejsza od mocy kodu jednobiegunowego. Dlatego odpowiedź na postawione wyżej pytanie jest pozytywna – stosowanie modulacji ASK, FSK i PSK w porównywalnych warunkach daje poprawę jakości transmisji cyfrowej.

Na rysunku 7.18b pokazano zależność stopy błędów dla tych modulacji w funkcji stosunku amplitudy sygnału do wartości skutecznej szumu. Pod tym względem najlepsza jest modulacja PSK, modulacja FSK wymaga sygnału o amplitudzie 1,41 razy większej (o 3 dB), a modulacja ASK wymaga sygnału o 2 razy większej amplitudzie (o 6 dB), aby uzyskać tę samą wartość stopy błędów.

dów. Wynika to z rezultatów pokazanych na rysunku poprzednim i z tego, że moc sygnałów FSK i PSK jest równa:

$$S_{\text{FSK}} = S_{\text{PSK}} = \frac{A^2}{2}, \quad (7.41a)$$

natomiast moc średnia sygnału ASK, przy jednakowym prawdopodobieństwie nadawania jedynek i zer, jest dwa razy mniejsza:

$$S_{\text{ASK}} = \frac{A^2}{4}. \quad (7.41b)$$

Ze względu na przedstawione właściwości modulacja ASK ma obecnie znaczenie tylko historyczne, modulacja FSK jest jeszcze używana w transmisjach o niewielkiej przepływności, np. telefonia komórkowa NMT, telefonia GSM (z dość znaczną modyfikacją), w niektórych bezprzewodowych sieciach komputerowych; modulacje zaś z grupy PSK (QPSK, 8PSK... QAM) są we współczesnej telekomunikacji bardzo popularne (zdecydowały o tym także inne względy).

7.7. Porównanie odporności szumowej modulacji PCM i FM

Z przedstawionych danych wynika, że modulacja PCM cechuje się dobrą odpornością szumową. Z modulacji analogowych można to samo powiedzieć o modulacji FM. Jej odporność szumowa wynika z jej szerokopasmowości – im szersze pasmo zajmuje sygnał FM, tym większy jest parametr m_f , tym bardziej zwiększa się wartość S/N na wyjściu demodulatora częstotliwości.

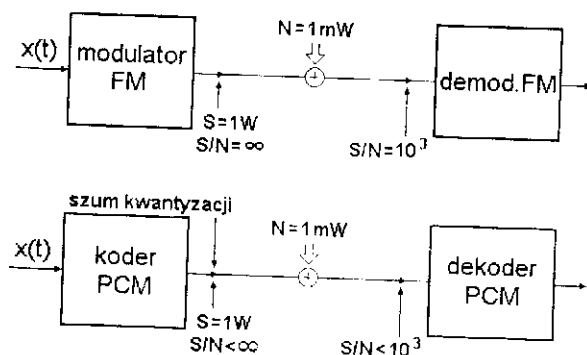
Modulacja PCM jest też modulacją szerokopasmową (była o tym mowa w rozdz. 6.), ale jej odporność szumowa nie wynika bezpośrednio z dużej szerokości pasma, lecz z odmienności postaci samego sygnału od analogowych zakłóceń.

Czy można te modulacje jakoś porównać pod względem odporności szumowej i orzec, która jest lepsza? W celu sprawiedliwego porównania należy przyjąć jednakowe warunki transmisji. Załóżmy mianowicie:

- taką samą moc sygnałów PCM i FM (np. 1 W),
- taką samą szerokość pasma częstotliwości,
- taką samą moc szumu zewnętrznego (np. 1 mW),
- taki sam przesyłany sygnał w postaci przebiegu sinusoidalnego o częstotliwości 3 kHz.

Sytuację tę pokazano na rys. 7.19a i b. Zaznaczono tu poziomy sygnałów i szumu, w odniesieniu do systemu PCM zaznaczono obecność szumu kwanty-

zacji i tylko on będzie tu uwzględniany, gdyż – jak wiadomo – przed innymi rodzajami zakłóceń sygnały PCM mogą być skutecznie zabezpieczone.



Rys. 7.19. Układy do porównania jakości transmisji:
a) dla FM, b) dla PCM.

Wyniki odpowiednich obliczeń dotyczących szerokości pasma 50, 100 i 120 kHz zamieszczono w tab. 7.4. Przy transmisji PCM przyjęto, że częstotliwość próbkowania wynosi 8 kHz.

Tab. 7.4. Porównawcze obliczenia jakości sygnału przy transmisji FM i PCM.

Transmisja	Wielkość	Wzór obliczeniowy	B = 50 kHz	B = 100 kHz	B = 120 kHz
FM	dewiacja częst.	$\Delta F_o = (B - 4f_m)/2$	19 kHz	44 kHz	54 kHz
	indeks mod.	$m_f = \Delta F_o / f_m$	6,3	14,7	18
	$\left(\frac{S}{N}\right)_{wy}$	przypadek ogólny z tab. 7.3 $\gamma = 1$	$\left(\frac{S}{N}\right) = 10^6$ $\Leftrightarrow 60 \text{ dB}$	$\left(\frac{S}{N}\right) = 10,8 \cdot 10^6$ $\Leftrightarrow 70,3 \text{ dB}$	$\left(\frac{S}{N}\right) = 19 \cdot 10^6$ $\Leftrightarrow 73 \text{ dB}$
PCM $F_p = 8 \text{ kHz}$	liczba bitów	$n = B / F_p$	$n = 6$	$n = 12$	$n = 15$
	$\left(\frac{S}{N}\right)_{wy}$	$\left(\frac{S}{N}\right) = 6 \cdot n, \text{ dB}$	$\left(\frac{S}{N}\right) = 36 \text{ dB}$ $\Leftrightarrow 4 \cdot 10^3$	$\left(\frac{S}{N}\right) = 72 \text{ dB}$ $\Leftrightarrow 15,8 \cdot 10^6$	$\left(\frac{S}{N}\right) = 90 \text{ dB}$ $\Leftrightarrow 10^9$

Z porównania wynika, że w przypadku szerokości pasma 50 kHz w przyjętych warunkach lepszą jakość sygnału zapewnia transmisja FM, przy 100 kHz

jakość sygnału jest podobna, natomiast przy szerokości pasma 120 kHz – transmisja PCM jest już wyraźnie lepsza. Okazuje się, że zawsze można znaleźć taką szerokość pasma, przy której jakość transmisji PCM będzie lepsza niż transmisji FM. Wytlumaczenie tego jest następujące.

Dla modulacji FM na podstawie tab. 7.3 (pierwszy przypadek), przy założeniu $m_f \gg 1$ i $\gamma \approx 1$, otrzymuje się zależność

$$\left(\frac{S}{N}\right)_{wy} \approx 3m_f^3 \left(\frac{S}{N}\right)_{we}.$$

Indeks modulacji FM jest proporcjonalny do dziewięci potęgi częstotliwości, a ta w przybliżeniu – do szerokości pasma sygnału, zatem

$$\left(\frac{S}{N}\right)_{wy} \sim B^3 \left(\frac{S}{N}\right)_{we}.$$

W przypadku modulacji PCM stosunek mocy sygnału do mocy szumu kwantyzacji wyrażony w decybelach dany jest zależnością (6.7):

$$\frac{S}{N}, \text{ dB} = 6 \cdot n,$$

to znaczy, że w liczbach bezwymiarowych wyraża się on następująco:

$$\frac{S}{N} = 10^{0,6 \cdot n}.$$

Na podstawie zależności (6.9)

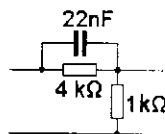
$$\frac{S}{N} = 10^{0,6 \cdot B/F_p}.$$

Ze wzrostem szerokości pasma transmisji stosunek S/N w przypadku transmisji FM rośnie w przybliżeniu z sześciannym (lub z kwadratem, jeżeli uwzględnić wzrost mocy szumów wejściowych) tej szerokości, natomiast w przypadku PCM rośnie wykładniczo; wiadomo, że przy odpowiednio dużej wartości argumentu funkcja wykładnicza rośnie zawsze szybciej niż dowolna funkcja potęgowa.

Fakt ten jest jedną z przyczyn powodujących szerokie rozpowszechnienie transmisji PCM we współczesnych systemach telekomunikacyjnych.

Zadania

1. Rezystor wytwarza szum cieplny o napięciu skutecznym 1 mV. Ile razy zmieni się to napięcie, jeżeli pasmo wzrośnie pięciokrotnie?
2. Na rysunku dany jest schemat ideowy układu preemfazy. Jaki jest moduł transmitancji tego układu dla $f = 1$ kHz i $f = 10$ kHz?



3. Jaka musi być amplituda impulsów jednobiegunowego kodu NRZ, aby stopa błędów wyniosła 10^{-3} . Wartość skuteczna zakłócającego szumu jest równa 0,1 V.
4. Przez analogię do obliczeń pokazanych w rozdz. 7. wyprowadzić zależność na odporność szumową modulacji AM SSB-SC i DSB-SC.
5. Obliczyć stosunek S/N na wejściu odbiornika, jeżeli gęstość widmowa mocy szumu wejściowego wynosi 10 nW/Hz, pasmo odbiornika jest równe 9 kHz, a sygnał użyteczny ma moc 12 mW.
6. Obliczyć S/N na wyjściu demodulatora synchronicznego AM przy danych jak w zadaniu 5., jeżeli sygnał użyteczny jest dany wzorem

$$a(t) = A_o \cdot [1 + 0,5 \cdot \cos(2\pi \cdot 10^3)t] \cdot \cos 2 \cdot 10^6 t.$$

7. Obliczyć S/N na wyjściu idealnego demodulatora FM przy mocy szumu i sygnału jak w zadaniu 5., jeżeli sygnał użyteczny ma następującą postać:

$$a(t) = A_o \cdot \cos[2\pi \cdot 10^6 t + 8 \cdot \sin(2\pi \cdot 10^3 t)].$$

Wskazówka. Należy wyznaczyć parametry ΔF_o i f_m , określić szerokość pasma sygnału i na tej podstawie moc szumu w tym pasmie.

8

Ogólne zagadnienia odbioru sygnałów

8.1. Podstawowe parametry odbiornika

Końcowym elementem każdego systemu telekomunikacyjnego jest odbiornik, „dopasowujący” przesylaną w postaci sygnałów informację do użytkownika.

Każdy „dobry” odbiornik powinien się cechować odpowiednią czułością i selektywnością. Nie powinien zniekształcać odbieranej informacji.

Czułość odbiornika w sensie ogólnym to jego zdolność odbioru słabych sygnałów. Im słabszy sygnał odbiornik może odebrać, tym jego czułość jest większa (lepsz).

Czułość wyrażona liczbowo to najmniejszy poziom sygnału odbieranego (napięcia, prądu lub mocy), który zapewnia uzyskanie na wyjściu odbiornika sygnału o założonej jakości.

Im mniejszy poziom sygnału jest potrzebny do tego celu, tym większą czułość ma odbiornik. Na przykład z dwóch odbiorników o czułości odpowiednio $20\ \mu\text{V}$ i $50\ \mu\text{V}$ ten pierwszy cechuje się większą czułością.

Selektywność to zdolność odbiornika do wydzielenia, spośród wielu sygnałów docierających na jego wejście, tylko sygnału użytecznego. Inne sygnały jednocześnie dochodzące do odbiornika stanowią wtedy zakłócenie i odbiornik powinien je zignorować – powinien na nie być „nieczuły”.

Selektywność wyrażona liczbowo oznacza, ile razy sygnał zakłócający musiałby być silniejszy od sygnału użytecznego, aby był odbierany z taką samą mocą co sygnał użyteczny.

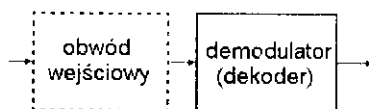
Prosta refleksja prowadzi do wniosku, że czułość odbiornika i jego selektywność muszą „iść w parze” – im większa czułość, tym większa musi być selektywność, w przeciwnym razie sygnał wyjściowy czulego odbiornika byłby zakłócony sygnałami niepożądanymi i przydatność takiego odbiornika byłaby niewielka.

Od czego zależy czułość odbiornika? Kiedy odbiornik będzie mógł poprawnie odebrać słaby sygnał? Można tu podać dwa warunki:

1) sygnał musi być dostatecznie silny względem dochodzących do odbiornika szumu i zakłóceń oraz także względem szumu generowanego w samym odbiorniku. Inaczej mówiąc, stosunek S/N na wejściu odbiornika powinien być odpowiednio duży (niekoniecznie większy od jedności!),

2) sygnał musi być dostatecznie silny, aby można go było w odbiorniku zdekodować lub zdemodulować – nawet gdyby nie zawierał żadnych zakłóceń.

Najprostszy układ odbiornika zawiera sam demodulator, poprzedzony ewentualnym obwodem wejściowym, dopasowującym go do kanału telekomunikacyjnego (rys. 8.1). Przykładem takich prostych odbiorników były odbiorniki radiofoniczne detektorowe, popularne w samych początkach radiofonii.



Rys. 8.1. Najprostszy odbiornik sygnałów.

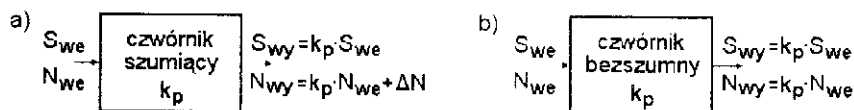
8.2. Współczynnik szumów

W danych warunkach sygnał dochodzący do odbiornika jest „skażony” szumem wejściowym odbieranym razem z sygnałem; w trakcie przetwarzania sygnału dodaje się jeszcze do tego szum generowany wewnątrz układu odbiornika. Szum ten to omówiony już szum termiczny, śrutowy i przede wszystkim szum generowany w elementach aktywnych (tranzystory, układy scalone). Powoduje to, że na wyjściu odbiornika pojawia się przebieg szumowy, będący superpozycją jakoś przetworzonego szumu wejściowego (np. wzmacnionego) i szumu wewnętrznego. Na szum wejściowy konstruktor nie ma dużego wpływu, natomiast szum wytwarzany wewnątrz odbiornika może być zredukowany przez jego odpowiednią budowę.

W celu liczbowego ujęcia szumu wewnętrznego wprowadzono pojęcie współczynnika szumów czwórnika (w szczególnym przypadku także odbiornika).

Jeżeli rozważany czwórnik charakteryzuje się wzmacnieniem mocy k_p (rys. 8.2a), to moc sygnału wyjściowego

$$S_{wy} = k_p \cdot S_{we} \quad (8.1)$$



Rys. 8.2. Bilans mocy sygnału i szumu dla czwórników: a) rzeczywistego (szumiącego), b) idealnego (bezszybnego).

Jeżeli do wejścia czwórnika doprowadzony zostanie przebieg szumowy (np. celowo lub pochodzący ze źródła sygnału), zostanie on tak samo wzmocniony w czwórniku (w procesie wzmacniania czwórnik nie „odróżnia” sygnału od szumu!), ale moc szumu na wyjściu będzie większa niż k_p razy – właśnie ze względu na to, że do wzmocnionego szumu doda się szum własny czwórnika. Można to wyrazić następująco:

$$N_{wy} = k_p N_{we} + \Delta N, \quad (8.2)$$

gdzie ΔN jest mocą szumu własnego czwórnika. W czwórniku idealnym pod względem szumowym (bezszybnym) moc szumu wyjściowego

$$N_{wy} = k_p N_{we}, \quad (8.3)$$

Współczynnik szumów czwórnika to stosunek mocy szumu wyjściowego w rozważanym przypadku do mocy szumu wyjściowego w przypadku idealnym (gdyby czwórnik nie dodawał własnego szumu):

$$F = \frac{k_p N_{we} + \Delta N}{k_p N_{we}}. \quad (8.4)$$

Inna definicja współczynnika szumów to stosunek stosunku mocy sygnału do mocy szumu na wejściu do stosunku mocy sygnału do szumu na wyjściu czwórnika:

$$F = \frac{(S/N)_{we}}{(S/N)_{wy}}. \quad (8.5)$$

Proste przekształcenie pokazuje, że taka definicja jest równoważna poprzedniej:

$$F = \frac{(S/N)_{we}}{(S/N)_{wy}} = \frac{S_{we}}{N_{we}} \cdot \frac{N_{wy}}{S_{wy}} = \frac{S_{we}}{N_{we}} \cdot \frac{k_p N_{we} + \Delta N}{k_p S_{we}} = \frac{k_p N_{we} + \Delta N}{k_p N_{we}}. \quad (8.6)$$

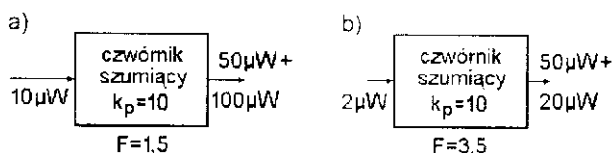
Uwaga

☞ Poziom szumu dodawanego przez czwórnik w danych warunkach jest stały, niezależny od poziomu wzmacnianego sygnału (jeżeli tylko zachowana jest liniowość pracy czwórnika). Moc szumu wejściowego może być dowolna, co powoduje, że tak określony współczynnik szumów jest wielkością zależną od mocy tego szumu. Z tego względu zawsze trzeba podawać, przy jakim pozio-

mie szumu wejściowego został wyznaczony (lub, w sposób pośredni, podaje się pasmo oraz rezystancję, temperaturę...).

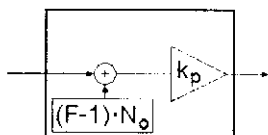
Przykład 8.1

Niech czwórnik szumiący, o wzmacnieniu mocy równym 10, charakteryzuje się tym, że w danych warunkach dodaje „od siebie” $50 \mu\text{W}$ mocy szumu wyjściowego. Jeżeli przeprowadzić pomiar współczynnika szumów tego czwórnika i zastosować szum wejściowy o mocy $10 \mu\text{W}$, stwierdzi się, że $F = 1,5$ (rys. 8.3a). Jeżeli natomiast w czasie pomiaru doprowadzić do czwórnika moc szumu wejściowego równą $2 \mu\text{W}$ (rys. 8.3b), otrzyma się $F = 3,5$. Wydawałoby się, że czwórnik o $F = 3,5$ jest gorszy pod względem szumowym od czwórnika o $F = 1,5$, ale to jest ten sam czwórnik, o niezmienionych właściwościach.



Rys. 8.3. Przykład wpływu mocy szumu wejściowego na wartość współczynnika szumów.

Współczynnik szumów nie stanowi więc jednoznacznej miary właściwości czwórnika, dlatego można się posługiwać wartością mocy lub napięcia szumu własnego odniesioną do wejścia. W rozważanym w przykładzie czwórniku wartość ta wynosi $5 \mu\text{W}$. Przy obliczeniach można też się posługiwać schematem blokowym „wnętrza” czwórnika (rys. 8.4), w którym uwidoczniło źródło szumu własnego o mocy $(F-1) \cdot N_o$, gdzie N_o jest mocą szumów wejściowych, przy której podany współczynnik szumów F został wyznaczony.



Rys. 8.4. Schemat zastępczy czwórnika o współczynniku szumów równym F .

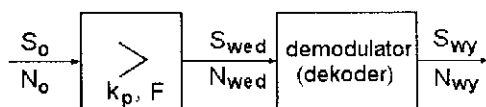
Współczynnik szumów jest wielkością bezwymiarową, ale bardzo często wyrażany jest w decybelach wg zależności

$$F, \text{ dB} = 10 \log F. \quad (8.7)$$

Wartości współczynnika szumów różnych urządzeń wzmacniających czy odbiorczych zależą od zakresu częstotliwości pracy, przenoszonego pasma itd., i typowo są zawarte w zakresie od ułamka decybelu (np. konwerter satelitarny) do kilku-kilkunastu decybeli, ale zawsze to zależy od warunków pomiaru.

8.3. Czulość odbiornika

Jaki związek występuje między współczynnikiem szumów a czulością odbiornika? Na wyjściu odbiornika, czyli także na wyjściu demodulatora, sygnał użyteczny powinien mieć określoną jakość wyrażoną przez $(S/N)_{wy}$ lub odpowiednią wartość stopy błędów BER. Przy danej modulacji lub danym sposobie kodowania oznacza to, że na wejściu demodulatora stosunek S_{wed}/N_{wed} musi mieć pewną wartość, wynikającą z charakterystyki szumowej danej modulacji (rys. 8.5).



Rys. 8.5. Schemat blokowy odbiornika pomocny w wyznaczaniu jego czulości.

Jeżeli demodulator jest poprzedzony wzmacniaczem o wzmocnieniu mocy k_p i współczynniku szumów F (wyznaczonym przy mocy szumów wejściowych N_o), to bezpośrednio z drugiej definicji współczynnika szumów wynika, że

$$F = \frac{S_o/N_o}{S_{wed}/N_{wed}}, \quad (8.8)$$

skąd

$$S_o = N_o \cdot F \cdot \left(\frac{S}{N} \right)_{wed}. \quad (8.9)$$

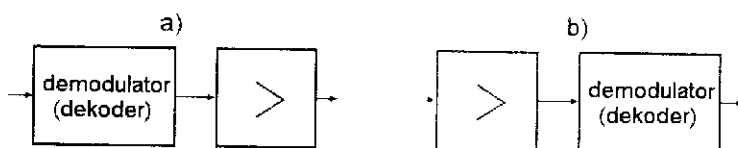
Jest to zależność określająca czulość odbiornika, tj. moc sygnału wejściowego zapewniającą uzyskanie odpowiedniej jakości sygnału wejściowego demodulatora, czyli też na wyjściu całego odbiornika.

Im silniejszy jest szum wejściowy, im większy jest współczynnik szumów, tym wymagana moc sygnału jest większa, czyli odbiornik ma gorszą czulość.

Im większy współczynnik szumów, tym czulość odbiornika jest mniejsza.

Rozważmy teraz drugi warunek dużej czułości odbiornika, który wymaga, aby sygnał był dostatecznie silny.

Działanie wielu układów demodulatorów amplitudy i częstotliwości czy dekoderek sygnałów cyfrowych pasmowych uwarunkowane jest ich nieliniowością. Układ demodulatora musi być nieliniowy, aby demodulacja była możliwa! Niestety, przy małej amplitudzie sygnału każdy nieliniowy element można uznać za liniowy; demodulator zawierający takie elementy przestanie wtedy działać. Włączenie po demodulatorze wzmacniacza nawet o dużym wzmocnieniu nie poprawi czułości odbiornika, bo wzmacniacz ten nie będzie miał po prostu co wzmacniać (rys. 8.6a).



Rys. 8.6. Możliwa kolejność wzmocnienia i demodulacji sygnału w odbiorniku:
a) wzmocnienie po demodulacji, b) wzmocnienie przed demodulacją

Poprawę czułości może dać jedynie wzmacniacz przed demodulatorem (rys. 8.6b) – wtedy nawet słaby sygnał po wzmocnieniu stanie się na tyle silny, że może już podlegać skutecznej demodulacji. Wynika stąd ważny wniosek:

Czułość odbiornika zależy od wzmocnienia
przed demodulatorem (dekoderem)

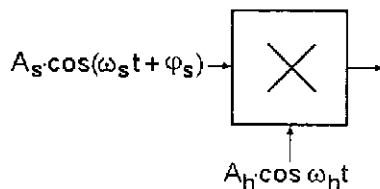
Zbudowanie odbiornika o dużej czułości na podstawie schematu z rys. 8.6b (tzw. odbiornika o bezpośrednim wzmocnieniu) jest jednak trudne z dwu powodów:

- 1) im większe wzmocnienie wzmacniacza, tym trudniej zapewnić jego stabilną pracę. Należy pamiętać, że często są to wzmacniacze wysokiej i bardzo wysokiej częstotliwości,
- 2) jeżeli odbiornik ma być dostrajany do różnych kanałów (w systemie FDMA), to wzmacniacz ten musi być selektywny i przestrajany. Uzyskanie dużej i stałej selektywności wzmacniacza przestrajanego jest trudne, bo ze zmianą częstotliwości zmieniają się parametry wielu jego elementów składowych.

8.4. Odbiornik superheterodynowy

Dobrym rozwiązaniem tych dylematów stał się układ odbiornika superheterodynowego wynaleziony przez E. Armstronga w 1916 r. Podstawą jego działania jest tzw. przemiana częstotliwości, przy której następuje zachowanie pier-

wotnej informacji, zawartej w sygnale. Najprostszy koncepcyjnie stopień przemiany częstotliwości pokazano na rys. 8.7.



Rys. 8.7. Przykładowy stopień przemiany częstotliwości na mnożniku analogowym.

Zawiera on mnożnik analogowy, pełniący tu funkcję tzw. mieszacza, do którego dochodzi sygnał użyteczny oraz przebieg sinusoidalny z generatora lokalnego, tzw. heterodyny. Przy założeniu jednostkowego współczynnika mnożnika na jego wyjściu uzyskuje się

$$a_{wy} = A_s A_h \cos(\omega_s t + \varphi_s) \cdot \cos \omega_h t = \frac{A_s A_h}{2} \cos[(\omega_s - \omega_h)t + \varphi_s] + \frac{A_s A_h}{2} \cos[(\omega_s + \omega_h)t + \varphi_s] \quad (8.10)$$

Najbardziej interesujący jest pierwszy składnik widma wyjściowego, tj. składnik różnicowy o **pośredniej częstotliwości** (p.cz.) $f_p = f_h - f_s$:

$$a_p(t) = \frac{A_s A_h}{2} \cos[(\omega_s - \omega_h)t + \varphi_s]. \quad (8.11)$$

Składnik ten można nazwać sygnałem, gdyż zawiera taką samą informację jak sygnał oryginalny (wejściowy). Rozważając, że sygnał wejściowy jest modulowany np. amplitudowo, tzn. $A_s = A_s(t)$, można zauważyć, że taką samą modulację będzie zawierał sygnał (8.11); podobnie będzie dla modulacji FM czy PM oraz innych, bardziej złożonych, np. QAM.

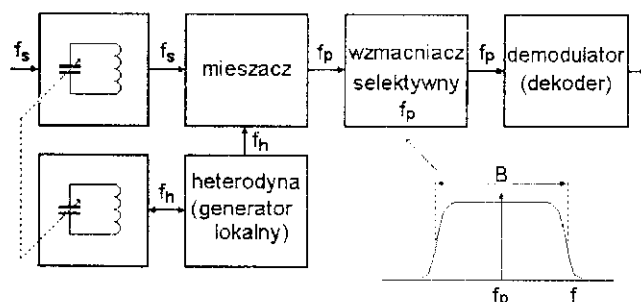
Zatem, konstruując odbiornik sygnału nadawanego na częstotliwości f_s , można go zastąpić odbiornikiem skonstruowanym na częstotliwość pośrednią f_p , poprzedzonym odpowiednim stopniem przemiany częstotliwości. Jest to korzystne, ponieważ:

1) sygnał o częstotliwości pośredniej ma z reguły dużo mniejszą częstotliwość niż sygnał wejściowy i dlatego jest łatwiejszy do wzmacniania,

2) przestrajając odbiornik, można tak dobierać częstotliwość heterodyny, aby przy zmianach częstotliwości sygnału różnica $f_h - f_s$ pozostawała stała –

ułatwia to znakomicie budowę selektywnego wzmacniacza sygnału pośredniej częstotliwości, który wtedy nie musi być przestrajany.

Schemat blokowy odbiornika z przemianą częstotliwości, nazywanego **odbiornikiem superheterodynowym**, pokazano na rys. 8.8.

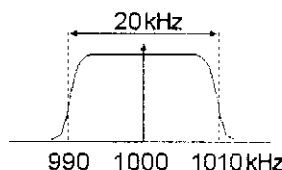


Rys. 8.8. Schemat blokowy odbiornika superheterodynowego.

Blok umieszczony przed mieszaczem to obwód wejściowy, którego zadaniem jest wstępna filtracja przychodzących na wejście sygnałów (do tego celu stosuje się najczęściej obwody rezonansowe). Drugi taki blok współpracuje z układem heterodyny i służy do ustalania jej częstotliwości. Na rysunku zaznaczono, że heterodyna jest przestrajana jednocześnie z obwodem wejściowym, ale jak już wspomniano, to jednocześnie przestrajanie musi polegać na zachowaniu stałej różnicy częstotliwości heterodyny i sygnału.

Układ odbiornika superheterodynowego ze względu na obecność wzmacniacza przed demodulatorem pozwala na uzyskanie dużej czułości oraz, jak się okazuje, także dużej selektywności.

Niech odbiornik zawiera selektywny wzmacniacz częstotliwości pośredniej o charakterystyce amplitudowej pokazanej na rys. 8.9. Jest to charakterystyka łatwa do uzyskania w przypadku wzmacniacza selektywnego nieprzestrajanego ($f_p = 1000 \text{ kHz} = \text{const!}$).



Rys. 8.9. Charakterystyka selektywności wzmacniacza pośredniej częstotliwości w rozważanym przykładzie.

Niech odbiornik ma odbierać sygnał nadawany na częstotliwości 5 MHz. Oznacza to, że częstotliwość heterodyny musi być ustawiona na wartość 6 MHz.

wtedy bowiem w mieszaczu z tych dwóch częstotliwości powstanie sygnał różnicowy 1 MHz, który następnie zostanie wzmoconiony we wzmacniaczu pośredniej częstotliwości i w końcu zdemodulowany.

Niech do odbiornika, oprócz sygnału użytecznego, dochodzi też przebieg zakłócający o znacznie zbliżonej częstotliwości, np. 5,05 MHz. Częstotliwość tego przebiegu różni się zaledwie o 1% od częstotliwości sygnału użytecznego i odfiltrowanie tego pierwszego w obwodzie wejściowym odbiornika (przed mieszaczem) może być bardzo trudne.

Przebieg zakłócający w mieszaczu „spotyka” przebieg heterodyny 6 MHz i powstaje z tego częstotliwość różnicowa 0,95 MHz. Częstotliwość ta wypada daleko poza pasmem przepustowym wzmacniacza p.cz., dlatego przebieg zakłócający przez ten wzmacniacz nie przejdzie i nie wywoła żadnego skutku na wyjściu odbiornika.

Odbiornik superheterodynowy ma więc dobrą zdolność tłumienia sygnałów niepożądanych, czyli cechuje się dobrą selektywnością, niemożliwą do uzyskania w układzie odbiornika o bezpośrednim wzmożeniu (rys. 8.6b).

Układ odbiornika superheterodynowego, z racji swych zalet, jest stosowany w wielu systemach i urządzeniach telekomunikacyjnych. W takiej postaci są wykonywane odbiorniki radiofoniczne, telewizyjne, radarowe, odbiorniki w telefonach komórkowych (nie wszystkich) itd. Prowadzi się też próby odbioru superheterodynowego w dziedzinie łączności optycznej.

Odbiornik superheterodynowy nie jest jednak pozbawiony wad. Jedną jest obecność tzw. kanału lustrzanego. Zauważmy, że w rozważanym przykładzie, jeżeli do odbiornika dotrze przebieg zakłócający o częstotliwości 7 MHz i jeżeli przejdzie on przez obwód wejściowy, to w mieszaczu ulegnie przemianie na częstotliwość $6\text{MHz} - 7\text{MHz} = -1\text{ MHz}$, co jest równoważne 1 MHz, i wywoła zakłócenie. Częstotliwość 7 MHz nosi w tym przypadku nazwę częstotliwości lustrzanej.

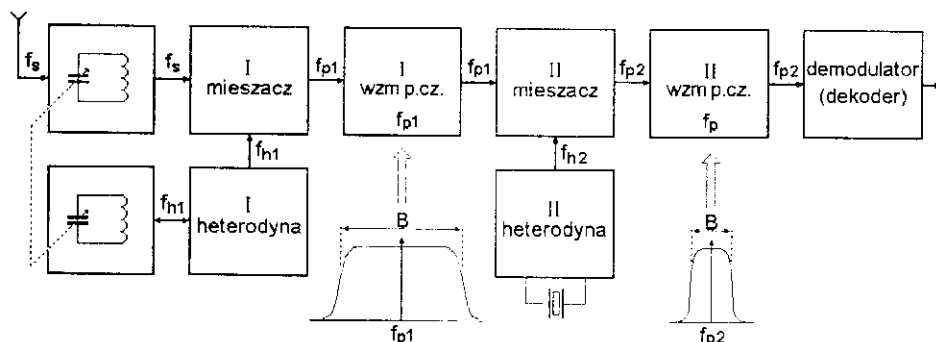
Aby nie dopuścić do tego zjawiska, konieczne jest stłumienie ewentualnego przebiegu lustrzanego już w obwodzie wejściowym odbiornika. Jest to dość łatwe, gdyż częstotliwość sygnału użytecznego i lustrzanego różni się znacznie (w rozważanym przykładzie aż o 40%); tym bardziej, im większą wartość ma częstotliwość pośrednia, zgodnie ze wzorem

$$f_L = f_s + 2f_p. \quad (8.12)$$

Dla dobrego tłumienia sygnałów lustrzanych korzystne jest, aby częstotliwość pośrednia była jak największa. Wtedy względna różnica między częstotliwością f_s i f_L jest duża. Z drugiej strony dla dobrego tłumienia sygnałów zakłócających położonych blisko częstotliwości sygnału użytecznego korzystne jest, aby częstotliwość pośrednia była mała (łatwiej wtedy o dobrą selektywność wzmacniacza p.cz.). Konieczny jest zatem albo kompromis w wyborze często-

tliwości pośredniej, albo zastosowanie w odbiorniku tzw. **podwójnej przemiany częstotliwości**.

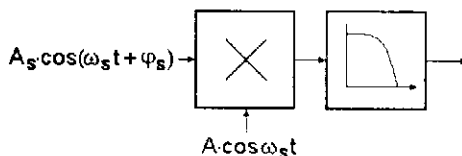
Odbiornik z podwójną przemianą częstotliwości jest to kaskadowe połączenie dwóch odbiorników superheterodynowych: pierwszego o dużej częstotliwości pośredniej i drugiego o małej (rys. 8.10). Ten dość złożony układ odbiornika odznacza się dobrą selektywnością dla tzw. dużych odstrojeń (częstotliwość lustrzana) i dla tzw. małych odstrojeń. Stosowany jest w bardziej odpowiedzialnych urządzeniach odbiorczych, natomiast w urządzeniach powszechnego użytku (radioodbiorniki, telewizory) – tylko sporadycznie (ze względu na większy koszt).



Rys. 8.10. Schemat blokowy odbiornika superheterodynowego z podwójną przemianą częstotliwości.

8.5. Odbiornik homodynowy

Rzadziej spotykanym układem odbiornika zapewniającym uzyskanie dużej czułości i selektywności jest odbiornik homodynowy (zasadę działania pokazano na rys. 8.11), inaczej odbiornik o bezpośredniej przemianie częstotliwości (*direct conversion receiver*). Jest to w istocie odpowiednik demodulatora synchronicznego omówionego w rozdz. 3.



Rys. 8.11. Zasada odbioru homodynowego.

Pomnożenie sygnału o określonej częstotliwości przez przebieg o częstotliwości nośnej daje (najczęściej) demodulację sygnału, przy czym poziom uzyskanego sygnału użytecznego zależy od poziomu zastosowanego przebiegu no-

śnego. W przypadku odpowiedniego typu mieszacza, przez zastosowanie dużego napięcia tego przebiegu, można uzyskać dużą czułość odbiornika. Selektywność takiego odbiornika zależy natomiast tylko od filtru dolnoprzepustowego, włączanego po mieszaczu.

Rozważmy podobny do poprzedniego przykład liczbowy. W celu odbioru sygnału o częstotliwości 5 MHz potrzeba, aby częstotliwość generatora lokalnego wynosiła dokładnie 5 MHz. Dokładne odtworzenie tej częstotliwości stanowi określoną trudność, obecnie rozwiązywaną przez realizację dość złożonych algorytmów i stosowanie zaawansowanych układów scalonych.

Jeżeli do odbiornika dotrze przebieg zakłócający o częstotliwości 5,05 MHz, zmiesza się on z przebiegiem generatora i powstanie z tego przebieg o częstotliwości 50 kHz (składową sumacyjną o częstotliwości 10,05 MHz tu zaniebujemy). Przebieg ten może zostać całkowicie wytłumiony w filtrze wyjściowym, czyli nie zakłóci odbioru sygnału użytecznego. Pasma odbiornika homodynowego jest równe podwójnemu pasmu filtru wyjściowego. Filtr ten, o względnie niskiej częstotliwości granicznej, może być zrealizowany jako selektywny filtr aktywny, bez użycia indukcyjności, co ułatwi jego ewentualne scalenie z resztą układu odbiornika.

Zaletą odbiornika homodynowego – w porównaniu z superheterodynowym – jest brak lustrzanego kanału odbioru i mniejsza złożoność budowy.

Odbiorniki homodynowe spotyka się w niektórych telefonach komórkowych oraz w bezprzewodowych sieciach komputerowych (Wi-Fi i innych) i np. odbiornikach Bluetooth.

8.6. Optymalny odbiornik sygnałów cyfrowych

Obecnie rozważona zostanie trzecia cecha dobrego odbiornika, polegająca na niezniekształcaniu odbieranej informacji. Skąd mogą się wziąć te zniekształcenia?

W przypadku odbiorników sygnałów analogowych zniekształcenia mogą pochodzić z nieliniowości stopni odbiornika (nasycenie, przesterowanie wzmacniaczy, generowanie sygnałów niepożądanych – harmoniczne...) oraz z obecności szumu zakłócającego odbiór. Szum ten powoduje, że sygnał wyjściowy odbiornika jest zniekształcony, przez co zawartość informacji w nim jest zmniejszona. W przypadku odbiornika sygnałów cyfrowych szum ten powoduje, że od czasu do czasu występuje błędny odczyt nadanej informacji cyfrowej. Jak wiadomo, to prawdopodobieństwo może być bardzo małe, ale nigdy (teoretycznie) nie maleje do zera.

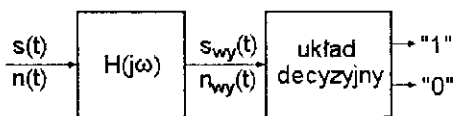
Oczywiste, że chcemy, aby prawdopodobieństwo błędnego odczytu bitów było jak najmniejsze. W odniesieniu do odbiorników sygnałów cyfrowych wprowadza się pojęcie **odbiornika optymalnego**. Jest to odbiornik, w którym prawdopodobieństwo błędnego odbioru informacji cyfrowej jest najmniejsze.

Odbiornik optymalny zapewnia najmniejsze
prawdopodobieństwo błędnego odbioru.

Istotną funkcją każdego odbiornika cyfrowego jest podejmowanie decyzji na podstawie odebranego sygnału, czy odebrany został sygnał 1 czy 0 (w przypadku sygnałów binarnych). W tym aspekcie należy zwrócić uwagę, że odbiornik wcześniej „wie”, jakie sygnały mogą się pojawić, jaka jest ich postać (czas, kształt). Dlatego inna nazwa odbiornika optymalnego to **odbiornik dopasowany**, jest on bowiem dopasowany do przewidywanej postaci sygnału.

Podejmowanie decyzji w odbiorniku odbywa się na podstawie określonego kryterium; na pewno w tym przeszkadza obecność szumu. Odbiornik optymalny musi zatem zapewniać maksymalizację stosunku S/N przed procesem podejmowania decyzji, aby ta następowała z najmniejszym prawdopodobieństwem błędu. Jak to uzyskać?

Budowę odbiornika sygnałów cyfrowych (szczególnie nadawanych w paśmie naturalnym) można przedstawić w uproszczonej postaci jak na rys. 8.12.



Rys. 8.12. Ogólna budowa odbiornika sygnałów
cyfrowych (binarnych).

Odbiornik zawiera przede wszystkim blok decyzyjny „rozpoznający”, czy odebrany został sygnał „0” czy „1”, oraz jakiś stopień wejściowy, oznaczony przez $H(j\omega)$, w pewnym sensie dopasowujący układ decyzyjny do kanału telekomunikacyjnego i do sygnałów w nim występujących.

Jakie właściwości powinien mieć ten stopień, aby na wejściu układu decyzyjnego $S/N = \max$?

Niech $H(j\omega)$ oznacza transmitancję tego stopnia (liniowego!), natomiast jego odpowiedź impulsowa będzie równa $h(t)$. Wiadomo, że istnieje wzajemnie jednoznaczny związek między tymi parametrami, co można zapisać jako $h(t) \Leftrightarrow H(j\omega)$, przy czym transmitancja $H(j\omega)$ jest transformatą Fouriera odpowiedzi impulsowej:

$$H(j\omega) = \int_{-\infty}^{\infty} h(t) \cdot e^{-j\omega t} dt \quad (8.13)$$

Jeżeli sygnał wejściowy $s(t)$ ma transformatę Fouriera równą $S(j\omega)$, to transformatę Fouriera sygnału po stopniu wejściowym można wyrazić jako

$$S_{wy}(j\omega) = H(j\omega) \cdot S(j\omega), \quad (8.14)$$

natomiast w dziedzinie czasu sygnał wyjściowy jest splotem sygnału wejściowego i odpowiedzi impulsowej stopnia wejściowego:

$$s_{wy}(t) = \int_{-\infty}^{\infty} s(\tau) \cdot h(t - \tau) \cdot d\tau. \quad (8.15)$$

W warunkach rzeczywistych sygnałem wejściowym odbiornika jest sygnał użyteczny oraz szum (addytywny), zatem sygnał na wyjściu pierwszego stopnia będzie

$$s'_{wy}(t) = \int_{-\infty}^{\infty} [s(\tau) + n(\tau)] \cdot h(t - \tau) \cdot d\tau, \quad (8.16)$$

gdzie $n(t)$ jest przebiegiem szumu wejściowego. Rozdzielając wyrażenie (8.16) na dwa składniki, uzyskuje się

$$s'_{wy}(t) = \int_{-\infty}^{\infty} s(\tau) \cdot h(t - \tau) \cdot d\tau + \int_{-\infty}^{\infty} n(\tau) \cdot h(t - \tau) \cdot d\tau. \quad (8.17)$$

Widoczne jest, że pierwszy składnik równania (8.17) to jakoś przekształcony (można powiedzieć – przefiltrowany) sygnał użyteczny, drugi składnik to też jakoś przekształcony szum wejściowy.

Maksymalizacja S/N na wyjściu układu decyzyjnego oznacza maksymalizację wyrażenia

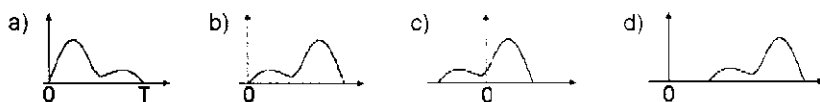
$$\frac{\int_{-\infty}^{\infty} s(\tau) \cdot h(t_0 - \tau) \cdot d\tau}{\int_{-\infty}^{\infty} n(\tau) \cdot h(t_0 - \tau) \cdot d\tau} \quad (8.18)$$

gdzie t_0 jest momentem, w którym ma nastąpić podjęcie decyzji o odebranych sygnale.

Na podstawie nieco dłuższych rozważań można pokazać [5, 6, 10], że optymalną postacią funkcji $h(t)$, prowadzącą do maksymalizacji wyrażenia (8.18), jest

$$h(t) = s(t_0 - t). \quad (8.19)$$

Odpowiedź impulsowa stopnia wejściowego odbiornika optymalnego powinna być równa odwróconemu w czasie i odpowiednio przesuniętemu przebiegowi przewidywanego sygnału wejściowego (rys. 8.13). To przesunięcie (w praktyce tylko opóźnienie, ze względu na realizowalność układu) powinno być jak najmniejsze, aby nie „zwlekać” z podejmowaniem decyzji.



Rys. 8.13. Przebiegi w przykładowym odbiorniku optymalnym: a) sygnał wejściowy, b) odpowiedź impulsowa odbiornika optymalnego, $t_0 = T$, c) odpowiedź impulsowa odbiornika optymalnego, $t_0 = T/2$ – układ nierealizowalny fizycznie, d) odpowiedź impulsowa odbiornika optymalnego, $t_0 = 1,5 T$ – występuje zbyt duża zwłoka w reakcji odbiornika.

Dla danej postaci przewidywanego sygnału stopień wejściowy odbiornika ma ściśle określoną odpowiedź impulsową, czyli także ma określoną transmitancję (w dziedzinie częstotliwości). Zatem stopień wejściowy odbiornika optymalnego to inaczej **filtr dopasowany** (dopasowany do postaci sygnału).

W tabeli 8.1 pokazano niektóre przykłady postaci sygnałów cyfrowych i przebiegi, jakie wystąpią na wyjściach filtrów dopasowanych, oraz transmitancje tych filtrów.

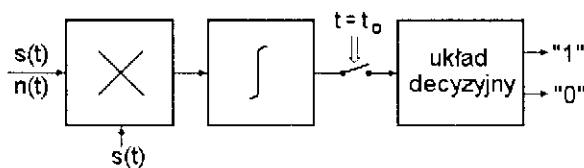
Tab. 8.1. Przebiegi w odbiorniku optymalnym dostosowanym do niektórych postaci sygnału wejściowego oraz transmitancje odpowiadających filtrów dopasowanych.

$s(t) \Rightarrow$ $\int_{-\infty}^{\infty} s(\tau) \cdot h(t - \tau) \cdot d\tau \Rightarrow$			
$ H(j\omega) $			

Widoczne jest, że optymalnym momentem czasu t_0 , kiedy sygnał na wyjściu filtru dopasowanego osiąga maksimum, jest moment zakończenia sygnału wejściowego – gdy cały sygnał wejściowy zostanie „przetworzony” przez odbiornik. Wartość sygnału otrzymana na wyjściu filtru dopasowanego dla $t = t_0$ jest w istocie równa energii sygnału wejściowego (wynika to z zestawienia zależności (8.15) i (8.19)).

Z analizy przypadków pokazanych w tabeli wynika, że przy dość typowych przebiegach impulsów transmitancje filtrów dopasowanych są podobne, ale dla impulsów typu „podniesiony cosinus” (trzeci przypadek) pasmo filtru jest najmniejsze (przedstawiony na rys. 6.15b przebieg cyfrowy składa się w istocie z takich właśnie impulsów).

Jak już wspomniano, wartość sygnału na wyjściu filtru dopasowanego w momencie $t = t_0$ jest równa energii sygnału wejściowego. Dlatego możliwa jest też inna realizacja odbiornika optymalnego, polegająca na dosłownym wyznaczaniu tej energii, tj. na całkowaniu kwadratu sygnału (rys. 8.14).



Rys. 8.14. Inny sposób realizacji odbiornika optymalnego.

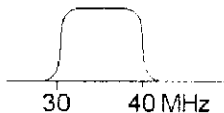
Podjęcie decyzji powinno nastąpić w momencie, gdy cały impuls wejściowy dotrze do odbiornika i zostanie scałkowany, bo wtedy S/N na wyjściu integratora będzie największe.

Konstruowanie odbiornika sygnałów cyfrowych musi się zatem opierać na znajomości dokładnej postaci sygnału nadawanego, bo tylko wtedy można zapewnić najmniejsze prawdopodobieństwo błędnego odbioru sygnałów. Jeżeli jest taka potrzeba, optymalny odbiornik może zawierać na wejściu także bloki wzmacniające, stopnie przemiany częstotliwości itd. Ważne jest, aby przed „rozpoznanie” tak przetworzonego sygnału przepuścić go przez filtr dopasowany, co zapewni najmniejszą stopę błędów na wyjściu układu decyzyjnego (dekodera).

Zadania

1. Współczynnik szumów odbiornika, zmierzony dla mocy szumów wejściowych równej $10 \mu W$, jest równy 3 dB. Ile będzie wynosić współczynnik szumów tego samego odbiornika, zmierzony dla mocy szumów wejściowych równej $5 \mu W$?

2. Korzystając ze schematu zastępczego czwórnika z rys. 8.4, obliczyć współczynnik szumów dwóch kaskadowo połączonych czwórników, o współczynnikach szumów równych $F_1 = 2$ i $F_2 = 5$, wyznaczonych dla mocy szumu wejściowego $N_0 = 10 \mu\text{W}$.
3. Dla odbiornika z zadania 1. obliczyć, jaka wartość mocy sygnału zapewni na wejściu demodulatora stosunek $S/N = 20 \text{ dB}$.
4. Wzmacniacz pośredniej częstotliwości odbiornika superheterodynowego ma charakterystykę częstotliwościową w pasmie 30–40 MHz (rysunek). Częstotliwość heterodyny jest równa 250 MHz. Jakie pasmo będzie odbierane przez ten odbiornik?



5. Zakres odbioru odbiornika radiofonicznego długofalowego to 150 – 285 kHz. W jakim zakresie musi być przestrajana heterodyna tego odbiornika, jeżeli częstotliwość pośrednia $f_p = 465 \text{ kHz}$?
6. W jakim zakresie zmieniać się będzie częstotliwość sygnałów lustrzanych dla odbiornika z poprzedniego zadania?
7. Pokazać, że dla momentu $t = t_0$ wartość sygnału na wyjściu filtra dopasowanego jest równa energii sygnału wejściowego.

